Mercury User Guide and Tutorials

2020.0 CSD Release

Copyright © 2020 Cambridge Crystallographic Data Centre

Registered Charity No 800579

Conditions of Use

The Cambridge Structural Database System (CSD System) comprising all or some of the following: ConQuest, CSD-Editor, Decifer, Mercury, Mercury CSD, CSD-Materials (formerly known as the Solid Form or Materials module of Mercury), Mogul, IsoStar, DASH, SuperStar, web accessible CSD tools and services, WebCSD, CSD sketchers, CSD data files, CSD data updates, the CSD database, sub-files derived from the foregoing data files, documentation and command procedures, test versions of any existing or new program, code, tool, data files, sub-files, documentation or command procedures which may be available from time to time (each individually a Component) is a database and copyright work belonging to the Cambridge Crystallographic Data Centre (CCDC) and its licensors and all rights are protected.

Use of the CSD System is permitted solely in accordance with a valid Licence of Access Agreement or Products Licence and Support Agreement and all Components included are proprietary. When a Component is supplied independently of the CSD System its use is subject to the conditions of the separate licence. All persons accessing the CSD System or its Components should make themselves aware of the conditions contained in the Licence of Access Agreement or Products Licence and Support Agreement or the relevant licence.

In particular:

- The CSD System and its Components are licensed subject to a time limit for use by a specified organisation at a specified location.
- The CSD System and its Components are to be treated as confidential and may NOT be disclosed or re-distributed in any form, in whole or in part, to any third party.
- Software or data derived from or developed using the CSD System may not be distributed without prior written approval of the CCDC. Such prior approval is also needed for joint projects between academic and for-profit organisations involving use of the CSD System.
- The CSD System and its Components may be used for scientific research, including the design of novel compounds. Results may be published in the scientific literature, but each such publication must include an appropriate citation as indicated in the Schedule to the Licence of Access Agreement or Products Licence and Support Agreement and on the CCDC website.
- No representations, warranties, or liabilities are expressed or implied in the supply of the CSD System or its Components by CCDC, its servants or agents, except where such exclusion or limitation is prohibited, void or unenforceable under governing law.

Licences may be obtained from:

Cambridge Crystallographic Data Centre
12 Union Road
Cambridge CB2 1EZ, United Kingdom

Web: http://www.ccdc.cam.ac.uk
Telephone: +44-1223-336408
Email: admin@ccdc.cam.ac.uk

Contents

1 Overview of Functionality 1
2 Introduction to using Mercury 5
2.1 Overview of the Interface 5
2.2 Top-Level Menu 6
2.3 Picking Modes 9
2.4 Right-Clicking in the Display Area 9
2.5 Undoing Commands and Resetting the Display 10
2.5.1 Undoing Commands 10
2.5.2 Resetting the Display to the Default View Direction, Scale and Perspective 10
2.5.3 Removing Packing and Nonbonded Contacts 10
2.6 Selecting and Deselecting Atoms and Molecules 10
2.6.1 The Purpose of Selecting Atoms and Molecules 10
2.6.2 Ways of Selecting Atoms and Molecules 10
2.6.3 Selecting by SMARTS string 11
2.6.4 Identifying Selected Atoms 12
2.6.5 Saving Atom Selections 13
2.7 Several Ways of Doing Things 13
2.8 Command Line Options 13
3 Loading Structures 16
3.1 Loading the Cambridge Structural Database or Other CSD-Format Databases 16
3.2 Loading Structures Found by a ConQuest Search 17
3.2.1 Loading Search Results from ConQuest 17
3.2.2 Controlling ConQuest Hit Highlighting 18
3.3 Loading a Structure from a File 20
3.3.1 Overview 20
3.3.2 Auto Edit Structure on Load 21
3.4 Sketching a Molecule 21
4 Selecting and Displaying Structures 23
4.1 Selecting and Displaying Structures: Using the Structure Navigator 23
4.2 Displaying Multiple Structures Simultaneously 24
4.3 Overlaying Crystal Structures 25
4.4 Overlaying Molecules 26
5 Moving the Display and Setting the View Direction 28
5.1 Translating the Display 28
5.2 Rotating the Display 28
5.3 Animating the Display 29

5.4 Flipping the Display by 90 Degrees 29
5.5 Scaling the Display 30
5.6 Viewing Along Crystallographic Axes 30
5.7 Viewing Along and Perpendicular to bonds 31
5.8 Setting a Rotation centre 31
5.9 Viewing Along and Perpendicular to Planes 32
5.10 Changing the Display Perspective 32
6 Customising the Display 33
6.1 Setting General Display Properties 33
6.1.1 Setting Background Colours 33
6.1.2 Depth Cueing 33
6.1.3 Labels 34
6.1.4 Setting Lighting Effects 34
6.1.5 Z-Clipping 34
6.1.6 Box-Clipping 35
6.1.7 Viewing Structures in Stereo 35
6.2 Setting Global Structure Display Properties 38
6.2.1 Setting a Global Display Style 38
6.2.2 Setting Sticks Display Options 38
6.2.3 Setting Ellipsoid Display Options 39
6.2.4 Setting Ball and Stick Display Options 40
6.2.5 Setting Spacefill Display Options 41
6.2.6 Setting Polyhedral Display Options 42
6.2.7 Setting Contact Display Options 43
6.2.8 Setting Measurement Display Options 44
6.2.9 Setting a Global Colouring Scheme 45
6.2.10 Changing Element Colour Settings 45
6.2.11 Custom Carbon Colouring 46
6.2.12 Switching Hydrogens On and Off 46
6.2.13 Turning All Atom Labels On or Off 46
6.2.14 Atom Labelling options 47
6.2.15 Colouring and Resizing All Labels 48
6.2.16 Customising the Display of Unit Cell Edges and Labels 48
6.3 Setting Display Properties for Particular Atom(s) or Bond(s) 49
6.4 Using Display-Property Dialogue Boxes 49
6.4.1 Using the Atom Styles, Atom Labels, Atom Colours and Show/hide Dialogue Boxes 49
6.4.2 Using the Select Color Dialogue Box 51
6.5 Atom and Molecule Colouring Conventions 52
6.5.1 Standard Element Colours 52

6.5.2 Symmetry-Equivalence Colouring 53
6.5.3 Atomic Displacement Colouring 55
6.5.4 Symmetry Operation Colouring 55
6.5.5 Gasteiger Charge Colouring 57
6.5.6 Colouring by Element or Suppression 57
6.6 Bond-Type Display Conventions 58
6.7 Managing Display Styles 58
6.7.1 Overview 58
6.7.2 Loading and Applying Display Styles 59
6.7.3 Creating and Editing Display Styles 60
6.7.4 Style Settings 61
6.7.5 Sharing, Deleting and Renaming Display Styles 62
7 Centroids and Planes 63
7.1 Creating Centroids 63
7.2 Creating Least-Squares Mean Planes 63
7.3 Creating Miller Planes 64
7.4 Setting the Display Properties of Centroids and Planes 64
7.5 Deleting Centroids and Planes 66
8 Measurement 67
8.1 Measuring Distances, Angles and Torsions 67
8.2 Clearing Measurements 67
9 Crystal Structure Display Options 69
9.1 Displaying and Labelling Unit Cell Axes 69
9.2 Colouring Conventions for Unit Cell Axes 69
9.3 Displaying Packing Diagrams (Assemblies of Unit Cells) 69
9.3.1 Packing Ranges 69
9.3.2 Displaying the Contents of a Single Unit Cell 69
9.3.3 Displaying Several Unit Cells and Fractions of Unit Cells 70
9.3.4 Switching Packing On and Off 71
9.3.5 Automatically placing contents of the asymmetric unit within the unit cell 71
9.4 Displaying Slices Through Crystals 71
9.4.1 What a Slice Is 71
9.4.2 Defining and Displaying a Slice 71
9.5 Displaying Symmetry Related Molecules 73
9.6 Displaying a Molecular Shell 73
9.7 Displaying Graph Sets 74
9.7.1 Customising the definitions of the Graph Sets 75
9.7.2 Customising hydrogen bond definitions for Graph Sets 76
9.7.3 Customising display of hydrogen bonds in Graph Sets 76

9.7.4 Saving Graph Sets 77
9.8 Displaying Voids 77
9.9 Displaying Space-Group Symmetry Elements 78
10 Finding Hydrogen Bonds and other Nonbonded Contacts 81
10.1 Overview of Nonbonded Contact Functionality 81
10.2 Basic Steps in Defining and Displaying Nonbonded Contacts 81
10.3 Default Nonbonded Contact Definitions 81
10.3.1 Default Definition of Hydrogen Bonds 81
10.3.2 Default Definition of Short Nonbonded Contacts 82
10.4 Defining New Types of Nonbonded Contacts 83
10.4.1 Specifying Distance Limits for Nonbonded Contacts 83
10.4.2 Intramolecular Nonbonded Contacts 83
10.4.3 Defining a New Type of Hydrogen Bond; The Define H-bonds Dialogue Box 84
10.4.4 Adding a New Short-Contact Definition; The Define Short Contacts Dialogue Box 86
10.4.5 Defining a New Type of Specific Contact; The Define Specific Contacts Dialogue Box 86
10.5 Colouring Contacts Based on Distance 87
10.6 Editing and Copying Nonbonded Contact Definitions 87
10.7 Deleting Nonbonded Contact Definitions 88
10.8 Viewing Nonbonded Contact Definitions 88
10.9 Switching the Display of Nonbonded Contacts On and Off 89
10.10 Finding Out What Symmetry Operations Were Used to Generate Contacts 90
10.11 Listing a Summary of the Contacts on View 90
11 Building Networks of Nonbonded Contacts 91
11.1 The Purpose of Building a Network 91
11.2 Basic Steps in Building a Network 91
11.3 Colouring Conventions for Contacts 94
11.4 Setting Display Properties of Contacts 95
11.5 What Network Expansion Is 97
11.6 Methods for Network Expansion 98
11.7 Deleting Contacts from a Network 99
11.8 Deleting Molecules from a Network 99
11.9 Resetting a Network to Its Starting Point 99
12 Editing a Structure 101
12.1 Manually Editing a Structure 101
12.2 Auto-Editing Options 102
12.3 Normalising Hydrogen Atom Positions 104
12.4 Polymer Expansion 105

12.5 Transforming, Translating or Inverting Molecules 106
12.6 Changing the Setting of a Space Group 107
12.7 Inverting a Crystal Structure 108
12.8 Reducing the Symmetry of a Crystal Structure 108
13 Visualising CSD- and PDB-based Intermolecular Contact Distributions 110
13.1 Overview 110
13.2 Setting Up Mercury For Use With IsoStar 111
13.3 Analysing Molecule(s) for Available Contact Distributions 111
13.4 Displaying a Distribution for a Specific Contact 113
14 Geometry Validation 114
14.1 Overview 114
14.2 Performing a Geometry Check 114
14.2.1 Applying Search Filters 115
14.2.2 Finding More Hits 116
14.2.3 Configuring Results Display Options 117
14.3 Viewing and Analysing Results 118
15 Displaying Additional Information about the Structure 121
15.1 Additional Data Items Available for Display 121
15.1.1 Display of Text Information 121
15.1.2 Chemical Diagram Display 122
15.1.3 Spreadsheet Display of Atoms and Their Properties 123
15.1.4 Spreadsheet Display of Bonds and Their Properties 125
15.1.5 Spreadsheet Display of Contacts and Their Properties 126
15.1.6 Spreadsheet Display of Centroids and Their Properties 127
15.1.7 Spreadsheet Display of Planes and Their Properties 127
15.1.8 Spreadsheet Display of Space Group Symmetry Operators 128
15.1.9 Spreadsheet Display of Distances, Angles and Torsion Angles 129
15.1.10 Spreadsheet Display of All Angles and All Torsions 130
15.2 Controlling which Data Items are Displayed 131
15.3 Manipulating Spreadsheets 132
15.4 Saving Spreadsheets 133
16 Powder Diffraction Patterns 135
16.1 Displaying and Saving Powder Diffraction Patterns 135
16.2 Customising the Powder Pattern Display 136
16.3 Technical Details of the Powder Pattern Simulation 138
17 Saving Files and Copying to the Clipboard 140
17.1 Saving Mercury Displays as Images 140
17.1.1 Saving Images from Mercury 140
17.1.2 POV-Ray Image Rendering from Mercury 140
17.1.3 POV-Ray Movie Rendering from Mercury 142

17.1.4 Advanced Rendering with POV-Ray 143
17.2 Saving Structures in Mercury Format 143
17.3 Saving Structures in External (3rd Party) Formats 144
17.4 Saving Mercury Displays as 3D Printing Files 144
17.4.1 Overview 144
17.4.2 Output of 3D Printing Files 144
17.4.3 Advice for 3D Printing File Output 146
17.5 Copying Images to the Clipboard 146
18 Data Analysis 147
18.1 Purpose of the Data Analysis Module 147
18.2 Definition of Data Analysis Terms 147
18.2.1 Data-Analysis Terms: Data Set 147
18.2.2 Data-Analysis Terms: Row, Column, Entry, Descriptor 147
18.2.3 Data-Analysis Terms: Current Entry 148
18.2.4 Data-Analysis Terms: Selection, Current Selection, Named Selection 148
18.2.5 Data-Analysis Terms: Data Views 149
18.2.6 Data-Analysis Terms: Visible, Hidden 149
18.2.7 Data-Analysis Terms: Locked, Unlocked 150
18.3 Rules When Using the Data Analysis Module 152
18.4 Starting a Data Analysis 153
18.5 Reloading a Saved Data Analysis 153
18.6 The Data-Analysis Window and Menus 154
18.7 Restoring the Data-Analysis Window 155
18.8 Docking and Undocking Plots and Other Data Views 155
18.9 Simultaneously Viewing All Spreadsheets, Plots and Tables 156
18.10 Using Drag and Drop between Data Views 156
18.11 Locking and Unlocking Spreadsheets, Plots and Tables 157
18.12 Guidelines on Using the Mouse to Interact with Data Views 157
18.13 Data-Analysis Spreadsheet Operations 158
18.13.1 Overview of Spreadsheet Functionality 158
18.13.2 Spreadsheet Contents 158
18.13.3 Finding Spreadsheet Entries 159
18.13.4 Visualising Spreadsheet Entries 159
18.13.5 Browsing Spreadsheet Entries with the Keyboard 159
18.13.6 Hiding and Unhiding Spreadsheet Rows 160
18.13.7 Hiding, Unhiding and Reordering Spreadsheet Columns 160
18.13.8 Sorting on Spreadsheet Columns 161
18.13.9 Combining Spreadsheet Columns: Accounting for Topological Symmetry 162

18.13.10 Colouring on Spreadsheet Columns 163
18.13.11 Grouping Entries by Descriptor Values: Tree Views 164
18.13.12 Copying Spreadsheets 167
18.13.13 Restoring a Lost Spreadsheet 167
18.14 Handling Multiple Data Sets 167
18.15 Adding New Descriptors to a Data-Analysis 168
18.15.1 Creating New Descriptors by Arithmetical Operations 168
18.15.2 Creating a Text Descriptor from an Integer Descriptor 169
18.15.3 Creating a Discrete Descriptor from a Continuous Descriptor; Categorising by Range 170
18.15.4 Importing Additional Descriptors from the CSD 171
18.16 Making and Using Data Selections 171
18.16.1 The Role of Selections in Data Analysis 171
18.16.2 Summary of Methods for Setting the Current Selection 171
18.16.3 Setting the Current Selection by Use of the Menu 172
18.16.4 Modifying the Current Selection from the Spreadsheet 172
18.16.5 Modifying the Current Selection from a Scatter Plot 173
18.16.6 Modifying the Current Selection from a Heat Plot 173
18.16.7 Setting the Current Selection by Filtering on Descriptor Values 174
18.16.8 Setting the Current Selection by Filtering on Descriptor Percentiles 175
18.16.9 Setting the Current Selection from a Descriptive Statistics Table 176
18.16.10 Creating a Named Selection 176
18.16.11 Recalling a Named Selection 176
18.16.12 Listing, Removing and Combining Named Selections 177
18.16.13 Selecting Entire Data-Sets 178
18.17 Creating and Manipulating Data Plots 178
18.17.1 Creating Plots 178
18.17.2 Plot Mouse Modes 179
18.17.3 Zooming Plots 179
18.17.4 Panning Plots 180
18.17.5 Setting the Current Selection from a Plot 180
18.17.6 Changing Plot Axes 180
18.17.7 Printing and Saving Plots 180
18.17.8 Deleting Plots 180
18.17.9 Histogram Features 180
18.17.10 Cone Angle Correction 181
18.17.11 Polar Histogram Features 183
18.17.12 Scatter Plot Features 184
18.17.13 Polar Scatter Plot Features 186
18.17.14 Heat Plot Features 187

18.18 Statistical Calculations 188
18.18.1 Calculating Descriptive Statistics 188
18.18.2 Calculating Circular Statistics 190
18.18.3 Calculating Correlation Matrices 190
18.18.4 Principal Component Analysis 191
18.18.5 Testing Whether Two Means are Significantly Different 194
18.18.6 Adding Descriptors to a Statistics Table 195
18.18.7 Removing Rows from Statistics Tables 195
18.18.8 Saving Statistics Tables 195
18.19 Saving Data Analyses 195
19 CSD-Community 197
19.1.1 Windows 197
19.1.2 Linux 197
19.1.3 MacOS 197
19.2 Overview 197
19.3 Deposit and Access of Structures 198
19.4 Teaching Database 198
19.5 Syntax Checking for CIFs 198
19.6 Reduced Cell Checking 199
20 CSD-System 200
20.1 Overview 200
20.2 Accessing CSD-System 200
20.3 Searching and Data Analysis 200
20.4 Intramolecular Geometry Analysis 201
20.5 Intermolecular Interaction Analysis 201
21 CSD-Materials 203
21.1 Overview 203
21.2 Accessing CSD-Materials 204
21.3 Searching on Motifs 204
21.3.1 Motif Selection 205
21.3.2 Structure Selection 209
21.3.3 Display of Motif Search Results 211
21.3.4 Filtering Motif Search Results 213
21.4 Searching on Crystal Packing Features 214
21.4.1 Running a Search 214
21.4.2 Available Database Search Filters 218
21.4.3 Viewing and Analysing Results 219
21.5 Determining Crystal Packing Similarity 221
21.5.1 Running a Crystal Packing Similarity Calculation 221
21.5.2 Setting Comparison Options 222

21.5.3 Viewing and Analysing Results 224
21.6 Calculating Energies and Structural Properties 228
21.6.1 Introduction 228
21.6.2 Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology 228
21.6.3 Theoretical Calculations using MOPAC 230
21.6.4 Calculating Intermolecular Energies 235
21.7 Determining Hydrogen Bond Propensities 238
21.7.1 Introduction 238
21.7.2 Starting the Propensity Prediction Wizard 239
21.7.3 Selecting a Target Molecule 240
21.7.4 Defining Functional Groups 241
21.7.5 Generating Fitting Data 242
21.7.6 Running a Hydrogen Bond Analysis 243
21.7.7 Fitting the Statistical Model 244
21.7.8 Assessing a Statistical Model 246
21.7.9 Visualising the results of a Hydrogen Bond Propensity calculation ..248
21.7.10 Using the Hydrogen Bond Propensity Results Table 248
21.7.11 Recalculating Hydrogen Bond Propensities 250
21.7.12 Charting Possible Hydrogen Bond Combinations 250
21.8 Quick Assessment of H-Bond Coordination 253
21.9 Assessing Molecular Interaction Preferences Using the CSD 254
21.9.1 Introduction 254
21.9.2 Setting up a Calculation 255
21.9.3 Visualisation and Analysis 256
21.10 Co-Crystal Design 259
21.10.1 Introduction 259
21.10.2 Screening by Molecular Complementarity 260
21.11 Analysing Hydrate Structures 263
21.11.1 Introduction 263
21.11.2 Using the Hydrate Analyser 264
21.11.3 Structure Overview 264
21.11.4 Water H-Bonding 266
21.11.5 Water Space 267
21.11.6 Water Interaction Maps 268
21.11.7 Coordination Polymer 269
21.12 Analysing Solvate Structures 269
21.12.1 Introduction 269
21.12.2 Using the Solvate Analyser 270
21.12.3 Solvent Selection and Space Calculation 270

21.12.4 Solvent H-bonding 271
21.12.5 Structure Summary 272
21.13 Generating Conformers Using the CSD 273
21.13.1 Introduction 273
21.13.2 Performing a Calculation 273
21.13.3 Advanced Options 274
21.14 Structure Solution from Powder Data 275
21.15 Post-Search Options 276
21.16 Controlling Hit Highlighting 276
21.17 Saving Searches and Results 277
21.18 Managing Searches 278
21.19 Managing Motifs 279
22 CSD-Discovery 281
22.1 Overview 281
22.2 Accessing CSD-Discovery 281
22.3 Mapping Molecular Interaction Preferences 281
22.4 Generating Conformers 282
22.5 Protein-Ligand Docking 282
22.6 Analysing Protein-Ligand Interactions 282
22.7 Overlaying Multiple Ligands 282
23 The CSD Python API 283
23.1 Overview 283
23.2 Setting Up the CSD Python API 283
23.3 Using the CSD Python API Menu 283
23.3.1 CSD Python API Sub-menus 283
23.3.2 Running Scripts from the CSD Python API Menu 284
23.3.3 CCDC Built-In Scripts 284
23.3.4 User-Created Scripts 285
23.4 The CSD Python API Options Dialogue 285
23.5 The CSD Python API Documentation 287
23.6 The CSD Python API Forum 287
23.7 Troubleshooting 287
24 References 288
25 Acknowledgements 289
26 Appendix A: Glossary and Technical Details 290
27 Appendix B: Keyboard Modifiers 319
28 Appendix C: Tutorials 320
28.1 Tutorial 1: Exploring Nonbonded Interactions 320
28.2 Tutorial 2: Rationalising Crystal Morphology 332
28.3 Tutorial 3: Investigating Pseudosymmetry 338

28.4 Tutorial 4: Looking for Possible Missing Water Molecules 344
28.5 Tutorial 5: Investigating Polymorphism 351
28.6 Tutorial 6: Analysing interactions, planes and ADPs 359
28.7 Tutorial 7: Assessing the Risk of Polymorphism via H-Bonding Propensities ...370
28.8 Tutorial 8: Searching on Hydrogen-Bond Motifs 385
28.9 Tutorial 9: Studying Crystal Packing Features 394
28.10 Tutorial 10: Searching on Packing Similarity 403
28.11 Tutorial 11: Generating High Resolution Figures 411
28.12 Tutorial 12: Using Mercury's Statistical Analysis Tools 418
29 Appendix D: What's New 427

1 Overview of Functionality

- Mercury is a program for visualising crystal structures in three dimensions, and for analysing the results of ConQuest substructure searches, or to investigate other data sets, which may be read in from *.csv* (comma-separated value) or *.tsv* (tab-separated value) files. Its features include:
- The ability to read in crystal structures in various formats (see Loading Structures).
- The ability to show all the files that have been loaded in the current session, and navigate between them.
- The ability to display multiple structures simultaneously (see Displaying Multiple Structures Simultaneously) and to perform least-squares overlay of pairs of structures (see Overlaying Crystal Structures) and pairs of chemically identical molecules (see Overlaying Molecules).
- The ability to rotate and translate the 3D crystal-structure display, and to view down cell axes, reciprocal cell axes, and the normals to least-squares and Miller planes (see Moving the Display and Setting the View Direction).
- A range of 3D visualisation options, including choice of display styles (Wireframe, Capped Stick, Ball and Stick, Spacefill, Ellipsoid and Polyhedral), colours, labelling schemes, and the ability to hide and then re-display atoms or molecules (see Customising the Display).
- A display style manager that contains pre-defined styles (for presentations, publication etc.), and provides a mechanism for creating and modifying sets of Mercury display styles, which are automatically saved between sessions.
- The ability to measure and display distances, angles and torsion angles involving atoms, centroids and planes (see Measurement).
- The ability to create and display centroids, least-squares mean planes and Miller planes (see Centroids and Planes).
- The ability to display unit cell axes, the contents of any number of unit cells in any direction, or a slice through a crystal in any direction (see Crystal Structure Display Options).
- Free space in crystal structures, or voids, can be calculated and displayed as coloured surfaces. Control over both the calculation parameters and display options are also available (see Displaying Voids).
- The ability to display space-group symmetry elements (see Displaying Space-Group Symmetry Elements).
- Location and display of intermolecular and/or intramolecular hydrogen bonds, short nonbonded contacts, and user-specified types of contacts (e.g. between a specific pair of elements in a given distance range, which may be expressed in terms of absolute or van der Waals corrected distances) (see Finding Hydrogen Bonds and other Nonbonded Contacts).
- The ability to build and visualise a network of intermolecular contacts (see Building Networks of Nonbonded Contacts).
- The ability to perform graph set analysis of hydrogen-bonding patterns (see Displaying Graph Sets).

- The ability to show extra information about the structure on display, such as the chemical diagram (if available), atomic coordinates and the symmetry operations used to generate contacts (see Displaying Additional Information about the Structure).
- The ability to manually edit structures including the addition of hydrogen atoms, the addition of bonds, the editing of bond types, the removal of atoms, bonds and molecules, the addition of atoms and the editing of atom types or labels (see Manually Editing a Structure).
- The ability to automatically assign bond-types, standardise bond-types to Cambridge Structural Database conventions and to add missing hydrogen atoms (see Auto-Editing Options).
- The ability to calculate, display and save the powder diffraction pattern for the structure on view (see Powder Diffraction Patterns).
- The ability to transform, invert or translate molecules within a structure (see Transforming, Translating or Inverting Molecules) and the ability to change the space group setting (see Changing the Setting of a Space Group).
- The ability to invert crystal structures (see Inverting a Crystal Structure) or reduce the symmetry of a structure from the current space group to one of the available subgroups (see Reducing the Symmetry of a Crystal Structure).
- The ability to generate structure representations as high-resolution graphics, movies or 3D printable model files (see Saving Files and Copying to the Clipboard).
- CSD-Community allows:
    - The ability to deposit experimental structures and access data from the CSD online (see Deposit and Access of Structures)
    - Access to the CSD Educational Collection of structures containing 700+ structures carefully selected to enhance chemistry learning (see Teaching Database)
    - The ability to validate CIFs and ensure experimental structures files are correctly formatted for submission to journals and database (see Syntax Checking for CIFs)
    - Functionality for performing crystal structure reduced cell checks against the CSD before collecting a full data set (see Reduced Cell Checking)
- CSD-System allows:
    - The ability to perform searches using WebCSD and ConQuest as well as analyse results using interactive charting, plotting and statistics (see Searching and Data Analysis).
    - The ability to analyse molecular conformations using Mogul in order to validate the intramolecular geometries of newly determined crystal structures and identify unusual geometric features (see Intramolecular Geometry Analysis).
    - The ability to analyse intermolecular interaction geometries using IsoStar, for example to validate the geometry of an H-bond or short-range interaction within a crystal structure (see Intermolecular Interaction Analysis).
    - Advanced powder pattern simulation (see Customising the Powder Pattern Display).

- CSD-Materials allows:
    - Structural searches to be carried out to identify interaction motifs that occur between particular functional groups and determine their relative abundance (see Searching on Motifs).
    - Searches can also be performed by selecting atoms in a currently loaded crystal structure and searching for structures that contain a similar packing arrangement (see Searching on Crystal Packing Features).
    - Calculation of crystal packing similarity between structures containing the same (or similar) compound (see Determining Crystal Packing Similarity).
    - The ability to display Bravais, Friedel, Donnay and Harker (BFDH) theoretical crystal morphologies (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology).
    - A range of molecule-based gas phase theoretical calculations using an interface to MOPAC (see Theoretical Calculations using MOPAC).
    - Force field intermolecular energy calculations using the UNI intermolecular potentials (see Calculating Intermolecular Energies).
    - Determination of hydrogen bond propensities to assess the potential risk of polymorphism (see Determining Hydrogen Bond Propensities).
    - Quick assessment of H-bond coordination likelihoods (see Quick Assessment of H-Bond Coordination).
    - Mapping of interaction preferences around complete molecules in a crystal structure (see Assessing Molecular Interaction Preferences Using the CSD).
    - Co-crystal design based on assessment of molecular complementarity (see Co-Crystal Design).
    - Analysis of hydrate crystal structures (see Analysing Hydrate Structures).
    - Analysis of solvate crystal structures (see Analysing Solvate Structures).
    - Generation of conformers based on geometrical statistics from the CSD (see Generating Conformers Using the CSD).
    - Solution of crystal structures from powder diffraction data using DASH (see Structure Solution from Powder Data).
- CSD-Discovery allows:
    - Mapping of interaction preferences around molecules (see Mapping Molecular Interaction Preferences).
    - Generation of conformers using knowledge from experimentally-determined crystal structures (see Generating Conformers).
    - Flexible docking of ligands into protein binding sites (see Protein-Ligand Docking).
    - Identification of interaction hotspots in protein binding sites (see Analysing Protein-Ligand Interactions).
    - Flexible overlay of multiple ligands of the same target (see Overlaying Multiple Ligands).
- The ability to save displays (see Saving Files and Copying to the Clipboard).
- The ability to undo (reverse) previous action(s), reapply these actions, and reset the display to default views (see Undoing Commands and Resetting the Display).

Note: Some of the above functionality is only available to CSD-System, CSD-Materials, CSD-Discovery or CSD-Enterprise users. Please see the *Release and Installation Notes* for details on activating Mercury for access to the full features.

2 Introduction to using Mercury

2.1 Overview of the Interface

The Mercury interface will vary slightly depending on whether the software is activated or not. You will need to activate Mercury to access CSD-System features - please see the *Release and Installation Notes* for details. A small CCDC icon will be shown in the interface against all features that are not available in your version.

The Mercury interface consists of the following:

- *Top-level menu* (see Top-Level Menu).
- *Tool bar* containing common, basic options, e.g. for setting display styles, colours, picking modes, view directions.
- *Display area* for showing 3D structures.
- *Structure Navigator*, containing the files and databases that are loaded into the program. Clicking on a structure in this list causes it to be displayed in the display area (see Selecting and Displaying Structures).
- *Options area* containing some common options.

- *Contacts list* containing a list of various types of nonbonded contacts that can be displayed or un-displayed by using the tick-boxes (see Switching the Display of Nonbonded Contacts On and Off).
- *Status bar* for displaying help messages.

2.2 Top-Level Menu

- **File** allows you to read in structures for display (see Loading a Structure from a File), or lists of structures obtained from a search of the Cambridge Structural Database or other databases in CSD format (see Loading Structures Found by a ConQuest Search). It also provides facilities for saving Mercury displays (see Saving Files and Copying to the Clipboard).
- **Edit** provides facilities for undoing and redoing actions (see Undoing Commands and Resetting the Display) and copying to the clipboard (see Copying Images to the Clipboard). Also available is:
    - The ability to manually edit structures including the addition of hydrogen atoms, the addition of bonds, the editing of bond types and the removal of atoms, bonds and molecules (see Manually Editing a Structure).
    - The ability to automatically assign bond-types, standardise bond-types to Cambridge Structural Database conventions and to add missing hydrogen atoms (see Auto-Editing Options).
    - The ability to normalise the position of hydrogen atoms (see Normalising Hydrogen Atom Positions).
    - The ability to perform polymer expansion (see Polymer Expansion).
    - The ability to perform symmetry transformations on molecules or structures (see Transforming, Translating or Inverting Molecules).
    - The ability to change the space group setting (see Changing the Setting of a Space Group).
    - The ability to invert a structure (see Inverting a Crystal Structure).
    - The ability to reduce the symmetry of a crystal structure from the current space group to one of the available subgroups (see Reducing the Symmetry of a Crystal Structure).
- **Selection** allows you to select sets of atoms in order for you to perform an action on them, such as colouring (see Selecting and Deselecting Atoms and Molecules).
- **Display** provides access to the following functionality:
    - Setting **Styles, Colours, Labels and Show/Hide** status (i.e. whether objects are displayed or hidden) (see Customising the Display).
    - Listing, in spreadsheet form, the atoms, bonds and contacts on view, and calculated objects and geometries such as planes, centroids, distances, angles and torsions. Also, depending on file format, displaying information such as the chemical diagram and literature reference (see Displaying Additional Information about the Structure).
    - Display of symmetry elements (see Displaying Space-Group Symmetry Elements).
    - Calculation and display of voids and control over the calculation parameters and display options (see Displaying Voids).

- Setting lighting, background, depth cueing and labelling display options (see Setting General Display Properties).
- A display style manager that contains pre-defined styles (for presentations, publication etc.), and provides a mechanism for creating and modifying sets of mercury display styles, which are automatically saved between sessions.
- Facilities for moving the display and viewing along crystallographic axes (see Moving the Display and Setting the View Direction)

- **Calculate** provides access to the following functionality:
  - Creating and displaying **Centroids** and **Planes** (see Centroids and Planes).
  - Controlling displays of crystal **Packing** and use of the **Slicing** facility to display a section through the crystal structure (see Crystal Structure Display Options).
  - Calculating nonbonded **Contacts**, including H-bonds (see Finding Hydrogen Bonds and other Nonbonded Contacts), and creating molecular networks (see Building Networks of Nonbonded Contacts).
  - Calculate and display Graph Sets for hydrogen bond patterns (see Displaying Graph Sets).
  - Calculating, displaying and saving the powder diffraction profile of the structure on view (see Displaying and Saving Powder Diffraction Patterns).
  - Performing least-squares overlay of pairs of structures (see Overlaying Crystal Structures).
- **CSD-Community** provides access to the following CSD-Community functionality:
  - The ability to deposit experimental structures and access data from the CSD online (see Deposit and Access of Structures)
  - Access to the CSD Educational Collection of structures containing 700+ structures carefully selected to enhance chemistry learning (see Teaching Database)
  - The ability to validate CIFs and ensure experimental structures files are correctly formatted for submission to journals and database (see Syntax Checking for CIFs)
  - Functionality for performing crystal structure reduced cell checks against the CSD before collecting a full data set (see Reduced Cell Checking)
- **CSD-System** provides access to the following CSD-System functionality:
  - The ability to perform searches using WebCSD and ConQuest as well as analyse results using interactive charting, plotting and statistics (see Searching and Data Analysis).
  - The ability to analyse molecular conformations using Mogul in order to validate the intramolecular geometries of newly determined crystal structures and identify unusual geometric features (see Intramolecular Geometry Analysis).
  - The ability to analyse intermolecular interaction geometries using IsoStar, for example to validate the geometry of an H-bond or short-range interaction within a crystal structure (see Intermolecular Interaction Analysis).

- Selection of the Cambridge Structural Database and/or an in-house database in CSD format (see Loading the Cambridge Structural Database or Other CSD-Format Databases).

- **CSD-Materials** provides access to the following CSD-Materials functionality:
  - Structural searches can be carried out to identify interaction motifs that occur between particular functional groups and determine their relative abundance (see Searching on Motifs).
  - Searches can also be performed by selecting atoms in a currently loaded crystal structure and searching for structures that contain a similar packing arrangement (see Searching on Crystal Packing Features).
  - Crystal packing similarity calculations can now be performed between structures containing the same compound (see Determining Crystal Packing Similarity).
  - Simulation of morphology using the Bravais Friedal Donnay Harker (BFDH) method (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology).
  - A range of molecule-based gas phase theoretical calculations using an interface to MOPAC (see Theoretical Calculations using MOPAC).
  - Force field intermolecular energy calculations using the UNI intermolecular potentials (see Calculating Intermolecular Energies).
  - The ability to calculate hydrogen bond propensities in order to assess the potential risk of polymorphism (see Determining Hydrogen Bond Propensities).
  - The ability to quickly calculate H-bond coordination likelihoods in order to assess the interactions observed in a given crystal structure (see Quick Assessment of H-Bond Coordination).
  - Mapping of interaction preferences around complete molecules in a crystal structure (see Assessing Molecular Interaction Preferences Using the CSD).
  - Co-crystal design based on assessment of molecular complementarity (see Co-Crystal Design).
  - Analysis of hydrate structures (see Analysing Hydrate Structures).
  - Analysis of solvate crystal structures (see Analysing Solvate Structures).
  - Generation of conformers based on geometrical statistics from the CSD (see Generating Conformers Using the CSD).
  - Solution of crystal structures from powder diffraction data using DASH (see Structure Solution from Powder Data).
- **CSD-Discovery** provides access to the following CSD-Discovery functionality:
  - Mapping of interaction preferences around complete molecules in a crystal structure (see Mapping Molecular Interaction Preferences).
  - Generation of conformers based on geometrical statistics from the CSD (see Generating Conformers).
  - Flexible docking of ligands into protein binding sites (see Protein-Ligand Docking).
  - Identification of interaction hotspots in protein binding sites (see Analysing Protein-Ligand Interactions).

- Flexible overlay of multiple ligands of the same target (see Overlaying Multiple Ligands).

- **CSD Python API** enables you to run supplied or user-written scripts created with the ***CSD Python API*** (see The CSD Python API).
- **Help** provides access to on-line help, tutorials and activation information.

2.3 Picking Modes

The *Picking Mode* box in the tool bar controls what happens when you left-click on items in the display area. Some of the below options include specialist mouse cursor images as a reminder to the user:

- Use of **Pick Atoms** mode permits the selection of atoms by clicking on them in the display area. Once selected, a set of atoms can be subjected to an operation, e.g. colouring, or used for a calculation, e.g. of a centroid.
- The **Lasso Atoms** mode allows selection of a range of atoms by drawing a perimeter around those atoms (this option includes a change of the mouse cursor to a lasso icon). You can also engage **Lasso Atoms** picking mode temporarily by holding down the *Alt* key.
- The **Expand Contacts** mode enables you to pick on a nonbonded contact (specifically, a *hanging contact*) in order to display in full, the molecule at the end of the contact (see What Network Expansion Is).
- Picking atoms in **Toggle Labels** mode will display their atom labels - pick again to switch off the label.
- The **Move Labels** mode allows manual repositioning of text labels.
- Use of **Measure Distance***,* **Measure Angle** or **Measure Torsion** modes permits the measurement of geometrical parameters by picking two, three or four objects (e.g. atoms, centroids or planes), respectively (these options include a change of the mouse cursor to a crosshairs icon).
- Clicking in the Mercury visualiser whilst in **Reveal Symmetry-Generated Molecules** mode will reveal the closest symmetry generated molecule to the area clicked.

2.4 Right-Clicking in the Display Area

- Right-clicking in the display area generates menus which provide access to many display options.
- The options available will depend on whether you right-click on:
  - A blank area in the display window, away from objects such as atoms, bonds, centroids, planes, unit cell axis or nonbonded contacts.
  - Specific objects in the display, i.e. an atom, bond, centroid, plane, unit cell axis or nonbonded contact.
- Menu items will be greyed out if they are inapplicable given the current state of the display, e.g. you cannot expand a network of nonbonded contacts until you have found and displayed at least one contact from which to begin.

- To perform operations using the keyboard (see **Appendix B: Keyboard Modifiers**)

2.5 Undoing Commands and Resetting the Display

2.5.1 Undoing Commands

To undo or redo instructions, hit **Edit** in the top-level menu and then:

- Hit **Undo** to reverse the last action.
- Hit **Undo** repeatedly to reverse previous actions one by one (up to 40 instructions can be undone).
- Hit **Redo** to re-apply an instruction that has just been undone.
- To remove all distances, angles and torsion angles select **Clear Measurements** from the *Picking Toolbar*.

2.5.2 Resetting the Display to the Default View Direction, Scale and Perspective

- The display area can be returned to the default view direction (*viz*. looking down the crystallographic *b* axis) and default scale and perspective in two ways:
  - Open the dial box by hitting **View** in the top-level menu followed by **Dial box...** in the resulting pull-down menu. Hit **Reset** in the dial box.
  - Right-click in the display background, select **View** from the resulting menu and then hit **Reset View** in the next menu.
- Any crystal packing or intermolecular contact information will be retained in the resulting default view, but this can be removed by using the **Reset** button in the main Mercury window.

2.5.3 Removing Packing and Nonbonded Contacts

To remove all crystal packing and nonbonded contacts from the display, and reset the view to the default scale and view direction (*viz*. looking down the crystallographic *b* axis), hit the **Reset** button at the bottom-left of the main Mercury window.

2.6 Selecting and Deselecting Atoms and Molecules

2.6.1 The Purpose of Selecting Atoms and Molecules

Selection of atoms or molecules is useful for changing properties such as display style, calculating objects such as mean planes, etc. Once a set of atoms has been *selected*, then one or more operations can be performed on those atoms.

2.6.2 Ways of Selecting Atoms and Molecules

Atoms may be selected or deselected in several ways:

- Hit **Selection** in the top-level menu to get access to **Select All, Deselect All** and **Invert Selection** buttons.
- Hit **Selection** in the top-level menu and then **Select** or **Deselect** in the pull-down menu. The resulting menu allows you to select or deselect:
  - All atoms of a particular type, e.g. **Non-Hydrogen**.

- Individual **Molecules** or **Atoms** by left-clicking on them**.**

- Right-click anywhere in the display area (atom, bond or background) and choose **Selection** from the resulting menu. This gives the same pull-down menus as described above.
- Set the *Picking Mode* (in the tool bar near the top of the main Mercury window) to **Select Atoms** and click on individual atoms with the left mouse button to select them. Once selected, an atom can be deselected by clicking on it again.
- Use the select by SMARTS toolbar to select atoms or groups of atoms based on a SMARTS string (for more details, see Selecting by SMARTS string).
- To select a group of atoms, enable the *Lasso Selection Mode* by clicking on the lasso icon in the picking mode toolbar.

then with the left mouse button depressed draw around the atoms you wish select. Release the left mouse button to make the selection. The mouse cursor will change to a lasso icon as a reminder whilst you are in **Lasso Atoms** picking mode. You can also engage the **Lasso Atoms** mode temporarily by holding down the *Alt* key.

- To select any atoms that are directly bonded to any other selected atoms right-click anywhere in the display area and choose **Selection** followed by **Expand Selection** from the resulting menu.
- When the *Picking Mode* is set to **Select Atoms** or **Expand Contacts**, all atoms become deselected if you left-click anywhere in the display-area background.
- Press the *shift* key and then click on any atom to select or deselect the entire molecule containing that atom (when in **Select Atoms** mode).
- Open the **Atoms** spreadsheet, e.g. by clicking on the **More Info** button (located at the bottom centre of the main Mercury window) and selecting **Atom List...** from the resulting menu. Clicking on a row of the spreadsheet then selects the corresponding atom in the main Mercury window. Clicking while holding the *Control* key down will cause previous selections to be retained; otherwise, they will be lost.
- Once the **Atoms** spreadsheet is open, all atoms of a particular element type may be selected by a two-step process. First, sort the spreadsheet on element, label or Sybyl type (click on the **Element**, **Label** or **SybylType** buttons at the head of the spreadsheet); this will put all atoms of a given element type into adjacent rows of the spreadsheet. Then, click on the first of the required atoms, keep the mouse button depressed, move the cursor down to the last of the required atoms and release the mouse button. The first, last and all intervening atoms (i.e. all atoms of the chosen element type) should become selected.
- To perform operations using the keyboard (see **Appendix B: Keyboard Modifiers**).

2.6.3 Selecting by SMARTS string

- To access the **Select by SMARTS** toolbar, hit **Display** in the top-level menu, and then under **Toolbars**, ensure that **Select by SMARTS** is enabled. If that option is enabled, you will see the **Select by SMARTS** highlighting toolbar at the top of the Mercury visualiser.

- Atoms or groups of atoms in the Mercury visualiser can be selected based on any given SMARTS string – simply type in the SMARTS string and hit Enter to select the relevant atoms.

- The SMARTS format (SMiles ARbitrary Target Specification) is simply a language for describing substructure patterns in molecules as a simple string of ASCII characters.
- If you would like to learn more about how SMARTS patterns work, see the web application SMARTSviewer by the University of Hamburg: http://smartsview.zbh.uni-hamburg.de/.

2.6.4 Identifying Selected Atoms

Atoms which are selected are indicated by small yellow dots, e.g.

2.6.5 Saving Atom Selections

- Those atoms currently selected in the display can be saved by selecting Selection from the top-level menu and **Save Current Selection...** from the resulting pull-down menu. Enter a name for the atom selection when prompted.
- To highlight the atoms belonging to any saved selection pick the required selection name from the **Atom Selections** pull-down menu, which is situated above the visualiser display area.
- Atom selections are retained when saving in Mercury *.mryx* file format (see Saving Structures in Mercury Format).

2.7 Several Ways of Doing Things

There are often several ways of doing the same thing. For example, you may set the colour of centroids via the **Calculate** menu's **Centroids** button, or via the **Display** menu's **Colours** sub-menu, or by right-clicking on the centroid and using the resulting pull-down menu.

2.8 Command Line Options

When opened from the command line, Mercury recognises the following options:

Controlling debug output

```
[-loglevel <0, 1, 2, 3, 4>]
```

Determines the level of detail for the output log. The default is 0 (no output). 4 gives maximum output.

```
[-logfile <filename>]
```

Specifies the filename to which debug log output should be written. By default, on Unix, logging information is written to the console. On Windows PCs, the logging output can only be seen when a filename is supplied.

Controlling 3D-drawing

These options all control the requests that Mercury makes of the OpenGL drawing system. It is important to note that there is no guarantee requesting a feature will mean that it is available.

```
[-doublebuffer | -nodoublebuffer]
```

Enable double-buffering to give a flicker-free image.

```
[-depth | -nodepth]
```

Enables/disables the depth-buffer

```
[-rgb | -norgb]
```

Enables OpenGL RGBA mode. This is the required mode. Disabling will cause colours to be shown incorrectly in the 3D display

`[-alpha | -noalpha]`

Enables/disables the alpha channel

`[-accum | -noaccum]`

Enable/disables the accumulation buffer

`[-stencil | -nostencil]`

Enables/disables the stencil buffer

`[-stereo | -nostereo]`

Enables/disables stereo buffering

`[-directrendering | -nodirectrendering]`

Enables/disables direct rendering (i.e. 'hardware acceleration')

`[-overlay | -nooverlay]`

Enables/disables the overlay plane

Help

`[-h | -help ]`

Displays a brief help message, showing the main command-line options

Communication with other Mercury instances

`[-client]`

The effect of a mercury `-client [filename]` command will depend on whether a `-client` Mercury session is already running. If one is, the named file will be opened in the existing Mercury session. Otherwise, Mercury will be started with the named file loaded and the new session will then listen for, and respond to, any further mercury `-client` commands.

Controlling the Graphical User Interface

`[-style=windows | -style=motif | -style=plastique | -style=cde | -style=cleanlooks]`

Changes the user interface style

Depending on the platform, the styles "windowsxp", "aqua" or "macintosh" may also be available.

`[-reverse]`

Draw all user interface components, from Right-to-left, instead of from Left-to-right.

Unix-specific options

`[-d]`

Displays the values that have been assigned to key environment variables by the Mercury start-up script, and then exits.

`[-mesa]`

Request that Mercury uses the Mesa libraries distributed by CCDC, in preference to the system OpenGL installation. This is sometimes used to work around 3D display problems on some hardware.

`[-verify]`

Prints a message confirming whether the mercury executable can be found, and then exits.

3 Loading Structures

3.1 Loading the Cambridge Structural Database or Other CSD-Format Databases

- When Mercury is launched, it tries to detect whether or not the Cambridge Structural Database (CSD) is installed (see Cambridge Structural Database. If the CSD can be found, it is opened within the *Structure Navigator* on the right-hand side of the main Mercury window:

- To find out more about any of the databases loaded in the *Structure Navigator*, simply hover the mouse over the name of the database and you will see a tooltip showing the file location of the database and the number of entries within it.
- If no other databases are available (e.g. you are using the free version of Mercury) you can access a free teaching subset of the CSD by hitting **CSD-Community** in the top-level menu, and then selecting **Open Teaching Database**.
- Mercury will also try to find any other CSD-format databases you may have created (see PreQuest). Any such databases are added to the *Structure Navigator*, which can then be used to switch between databases for viewing.
- If you have a CSD-format database that is not detected automatically by Mercury, you can load it by hitting **CSD-System** in the top-level menu, and then selecting **Select Databases…**. Any CSD-format databases opened in this way will be remembered between sessions.
- To remove a specific CSD-format database, right-click on the database name within the *Structure Navigator* and select **Remove** from the resulting pull-down menu. To remove all CSD-format databases right-click on *Databases* within the *Structure Navigator* and select **Remove All Databases** from the resulting pull-down menu.
- It is possible to "cut" a list of refcodes from a file, for example a text file or a spreadsheet, and paste it into the *Structure Navigator*. These structures will then appear as a **Refcode list** in the *Structure Navigator.*

3.2 Loading Structures Found by a ConQuest Search

3.2.1 Loading Search Results from ConQuest

- If you have performed a search in ConQuest (see ConQuest), and that program has detected that Mercury is installed, you will be able to transfer the structures and the values of any geometric parameters found in your search to Mercury. This can be done using ConQuest's *View Results* pane in several ways:
    - Hit **File** in the ConQuest top-level menu followed by **View Entries in Mercury.** Alternatively, you have defined some parameters (e.g. bond lengths, angles) as part of the search query. The values of these parameters in the hits can be tabulated and explored in Mercury by hitting **File** in the ConQuest top-level menu followed by **Analyse Data in Mercury.**
    - Hit the **Analyse Hitlist** button (located above the scrolling list of refcodes in the ConQuest *View Results* pane), and then hit **Visualise Structures.** Alternatively, if you have defined some parameters (e.g. bond lengths, angles) as part of the search query. The values of these parameters in the hits can be tabulated and explored in Mercury by hitting the **Analyse Hitlist** followed by **Analyse Data.**
    - Click the right mouse button on the scrolling list of refcodes in ConQuest, and then hit **Visualise Structures.** If you have defined some parameters as part of the search query then the values of these parameters in the hits can be tabulated and explored in Mercury by clicking the right mouse button on the scrolling list of refcodes in ConQuest, and then selecting **Analyse Data.**
- ConQuest hits (*.c2m* files) can be loaded from within Mercury by hitting **File** in the top-level menu, and then selecting **Open.** Alternatively, right-click anywhere in the *Structure Navigator* and select **Open...** from the resulting pull-down menu. You can

also drag-and-drop files into a running Mercury window. Once selected the ConQuest hits will be added to the *Structure Navigator*.

- To remove a specific ConQuest hit file right-click on the file name within the *Structure Navigator* and select **Remove** from the resulting pull-down menu. To remove all ConQuest hit files right-click on *ConQuest Hits* within the *Structure Navigator* and select **Remove All ConQuest Hits** from the resulting pull-down menu.
- Mercury can also read text files containing lists of refcodes (see Cambridge Structural Database
- Refcode (entry ID)), one per line, provided that the database containing the corresponding structures has already been loaded. Refcode lists can be loaded in various ways (see Loading a Structure from a File).

3.2.2 Controlling ConQuest Hit Highlighting

- The highlighting of hit fragments in ConQuest is transferred to Mercury when Mercury is called from ConQuest (see Loading Search Results from ConQuest). Mercury will start together with the *Hit Fragment Display Options* dialogue:

- This dialogue allows users to navigate hit fragments and to customise fragment highlighting. The *Hit Fragments Display Options* dialog can also be opened by selecting **CSD-System** from the top-level menu, followed by **ConQuest Hit Highlighting...** from the resulting menu.
- The *Fragment Selection* pane displays the refcode(s) of the query result(s) together with the fragment(s) in the crystal structure containing the search substructure and

any 2D/3D parameters. By default only the first refcode in the ConQuest search results will be displayed. A query can contain several fragments; these are listed below the refcode in the tree view:

- The Tree view can be switched off by deselecting the **Tree view** tick box. This will produce a list of the selected query(s) together with a table of the parameter value(s). If a query contains several fragments, the parameter values for each fragment will be shown on a separate line.
- Selecting **Load multiple structures** allows for the simultaneous display of several search results with the search fragment overlaid in the Mercury display. The selected queries will also be shown in the *Hit Fragment Display Options* dialogue. *Note*: Displaying Packing features will disrupt the overlay of the search fragments.
- Select **Show only hit atoms** to hide those atoms not in the search fragment (i.e. Mercury will only show atoms in the search substructure).
- The **Show parameters** tick box controls which geometric parameters to display in Mercury: individual distances, angles or torsions can be selected. The default is to display geometric parameters defined in the search.
  Note: Distances, angles and torsion defined between or to objects such as centroids or planes will not be displayed.
- The *Fragment Highlighting* pane provides options for of hit atoms and contacts:

- Select the required display style (Capped Stick, Ball and Stick, etc.) and the colour of the atoms in the hit fragment.
- The colour of non-bonded contacts can also be customised. This is only relevant when a search for a non-bonded contact has been performed in ConQuest.
  Note: In order to see the changed colour, **Show parameters** under the *Fragment Selection* tab must be switched off.

3.3 Loading a Structure from a File

3.3.1 Overview

- Mercury has its own file format (filename extension *.mryx*). The *.mryx* file is binary, so cannot be read using a text editor.
- In addition, Mercury can read crystal structures from the following types of files:
  - *cif* (see cif Format).
  - *mol* (see mol Format).
  - *mol2* (see mol2 Format).
  - *pdb* (see pdb Format).
  - *res* (see res Format).
  - *sd* (see sd Format).
  - *xyz* (see xyz Format).
- A file in any format known to Mercury, including a file of CSD refcodes, can be loaded in several ways:

- Hit **File** in the top-level menu, and then select **Open**. Choose the file type (i.e. format) you want and then select the file, or files you wish to open.
- When typing the `mercury` command (typically on Unix), you can supply one or more filenames on the command line, and Mercury will open the file(s) when it starts.
- If you are using a computer that supports *Drag and Drop* (such as Windows 98 or later), you can drag a file icon and drop it onto the Mercury program (or a shortcut to it). Mercury will launch and open the dropped file or files.
- A file (or multiple files) can also be opened by dragging and dropping them into a running Mercury window.
- Files can be loaded by right-clicking anywhere in the *Structure Navigator* and selecting **Open...** from the resulting pull-down menu. Once selected the file will be added to the *Structure Navigator*.
  Note: To remove a specific file, right-click on the file name within the *Structure Navigator* and select **Remove** from the resulting pull-down menu. To remove all files of a particular type right-click on either *Structures*, *Refcode Lists*, or *Mercury Files* within the *Structure Navigator* and select **Remove All** from the resulting pull-down menu.

- A list of the ten most recently opened files is available by selecting **File** from the top-level menu, then **Recent Files** from the resulting drop down menu. Click on a listed filename to load the file into Mercury.

3.3.2 Auto Edit Structure on Load

- This feature allows structures read into Mercury in non-CSD format files (i.e. *res*, *cif*, *pdb*) to be automatically edited using the settings defined in the Edit > Auto-Edit Structure dialogue (see Auto-Editing Options).
- This feature is switched on and off by hitting **File** in the top-level menu then selecting **Auto Edit Structure on Load**. This selection is preserved between Mercury sessions
- Auto edits can be reversed by clicking **Edit** in the top-level menu then clicking **Undo: Auto-edit structure**.

3.4 Sketching a Molecule

- Mercury allows the user to draw a molecule in a sketcher window then load that molecule into the Mercury visualiser
- Open the sketcher by hitting **File** in the top-level menu then selecting **Sketch Molecule…**
- Full instructions on use of the sketcher can be found in the Mogul documentation (see http://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/#mogul)
- Once a drawn molecule has been loaded into the Mercury visualiser, the CSD Conformer Generator can be used to create likely solid state conformations of the molecule (see Generating Conformers Using the CSD)
- When sketching molecules, users should note that:
  - If using Groups in the sketcher, functional groups should be expanded by right-clicking the group and selecting **Expand Group** from the resulting menu

- Use of the “Any” atom type in the sketcher will result in an “Unknown Atom” being displayed in the Mercury visualiser for the sketched molecule
- Hydrogen atoms will only be generated in the 3D molecule if they are drawn explicitly in the sketcher, if they are shown implicitly they will not be transferred to the 3D molecule

4 Selecting and Displaying Structures

4.1 Selecting and Displaying Structures: Using the Structure Navigator

- Each structure, file, and database loaded into Mercury will be displayed in the *Structure Navigator* on the right-hand side of the main Mercury window.
- The *Structure Navigator* is a dockable window, which means it can be placed in various positions in the main Mercury window, or kept as an entirely separate window. To do this, move the mouse cursor to the top of the *Structure Navigator* then drag the window with the mouse, keeping the left-button depressed, and put it down where you want by releasing the mouse button. The *Structure Navigator* can also be switched on or off by clicking on the **Display** button in the top-level menu, selecting **Toolbars**, then enabling or disabling the *Structure Navigator*.
- The *Structure Navigator* shows all the files that have been loaded in the current session, and allows easy navigation between them. Loaded crystal structures are presented in a hierarchical tree organised first by file type (Databases, Structures, Refcode Lists, ConQuest Hits, or Mercury Files) and subsequently by filename and then by individual structure identifier.
- Arrows are used to expand and collapse sections within the *Structure Navigator*. In the example below, clicking on the arrow that points to the right next to *Structures* will show all structure files loaded in the current Mercury session viz. *873083-873083.cif*, *921390.cif* and t*erpyridyl.cif*. Clicking on the arrow next to t*erpyridyl.cif* will then list the structure identifiers for each crystal structure within that file:

- To view a structure, select the corresponding identifier in the *Structure Navigator*.
- The arrow keys can be used for quick step-wise traversal through the tree. Use the up and down keyboard keys to move between nodes and the left and right arrow key to collapse and expand nodes respectively.
- Structures can also be browsed by clicking the >> button to display the next structure or << to display the previous structure listed.

- To find a specific structure, enter the structure identifier into the text box just above the *Structure Navigator*. The structure will be displayed automatically. Hit the **Find** button to locate successive crystal structures beginning with the same characters. This can be useful when locating all entries within a refcode family (*ACETYL01, ACETYL02, ACETYL03*, etc.).
- The *Structure Navigator* lists the space group for each loaded structure. To sort structures by space group click on the **Spacegroup** button.
- To view all loaded structures as flat list of identifiers switch off the **Tree View** check box. This will disable the hierarchical tree view of crystal structures.
- By default, only one structure can be viewed at a time and the structure currently on display will be highlighted. However, it is possible to display multiple structures simultaneously (see Displaying Multiple Structures Simultaneously).

4.2 Displaying Multiple Structures Simultaneously

- Mercury allows two or more structures to be displayed simultaneously.
- To display multiple crystal structures, switch on the **Multiple Structures** tick box located under the list of structures on the right-hand side of the main Mercury window. Select the structures you wish to display by clicking on their corresponding identifiers within the *Structure Navigator*. Each structure selected for viewing will be shown in the display area.
- Various options are available for controlling the display of multiple structures. To access these, hit the **Structures...** button located under the list of structures on the right-hand side of the main Mercury window. This will open the *Multiple Structures* dialogue:

- This dialogue lists all the structures currently being displayed. Within this window you can:
    - Control whether movable structures will rotate around a single global rotation centre, or whether structures individually rotate around their own

rotation centre. Select either the **Global rotation centre** or **Local rotation centre** radio button to set the required behaviour.

- Remove individual structures from the display. Hit the **Delete** button next to the structure you wish to remove.
- Choose to make individual structures Visible. Turning this off is useful for hiding particular structures.
- Choose to make individual structures *Active*. Subsequent actions e.g. packing will be applied to active structures only.
- Choose to make individual structures *Movable*. Subsequent rotation and translation of the display will be applied to movable structures only.
- Set the colour of individual structures by selecting the required option from the appropriate **Colour** pull-down menu.

4.3 Overlaying Crystal Structures

- To perform least-squares overlay of pairs of structures, display the two structures you wish to overlay (see Displaying Multiple Structures Simultaneously) then select **Calculate** from the top-level menu followed by **Structure Overlay...** from the resulting pull-down menu. This will open the *Structure Overlay* dialogue.
- In the main Mercury window, select pairs of atoms from the two crystal structures you wish to overlay (these pairs of atoms will be used to calculate the least-squares overlay). Broken yellow lines will be displayed between each selected pair of atoms:

- As pairs of atoms are selected they will be listed in the *Structure Overlay* dialogue:

- Once the desired number of pairs are selected, hit the **Overlay** button.
- Overlay is performed by moving the second crystal structure relative to the first crystal structure. During overlay crystal structures will be kept rigid, e.g. it is not possible to rotate torsions to find better fits or perform inversions.
- Crystal structure overlays can be reversed by selecting **Edit** from the top-level menu followed by **Undo: Overlay** from the pull-down menu.
- The root-mean-square (RMS) of the pairs of atoms used for the overlay is displayed in the *Structure Overlay* dialogue.
- To refine the orientation, you can continue to select further pairs of atoms. These will be added to the pairs already listed in the *Structure Overlay* dialogue. To remove a pair of atoms hit the **Delete** button next to the pair you wish to remove. Hit the **Overlay** button in order to re-calculate the overlay using the current selection.

4.4 Overlaying Molecules

- If you are interested in comparing the chirality or stereochemistry (see Chirality) of two molecules, use the *Molecule Overlay* tool which employs both inversion and flexibility to calculate molecule overlays.
- To overlay two molecules of the same chemical structure, either within the same crystal structure or belonging to different crystal structures, display the two molecules you wish to overlay then select **Calculate** from the top-level menu followed by **Molecule Overlay...** from the resulting pull-down menu. This will open the *Automatic Molecule Overlay* dialogue.
- Select an atom from each molecule and hit **Overlay.** To find the best overlay of molecules the tool overlays the molecules directly (with no changes), inverts one molecule prior to overlay, rotates flexible torsion angles of one molecule prior to overlay, and applies an inversion and allows flexible torsion angles to rotate prior to overlaying the molecules.
- The results of the different methods for overlaying are presented in the dialogue along with the RMSD for the overlay and the maximum distance between two equivalent atoms in the overlay (Max. D). Click on the radio buttons to display the calculated overlay.

- If the overlay of the molecules is not improved by the application of an inversion, the inversion is not applied and the RMSD reported is the same as the RMSD calculated for the direct overlay.
- If an overlay is improved by the application of inversion symmetry the pair of molecules may be enantiomers.
- If the reported RMSD for the overlay is quite low, but the Max. D value is large, it may be an indication that the structures are quite similar, except for a localised area. For example a comparison of *cis* and *trans* geometries may result in a low RMSD but high Max. D.
- Click on **Reset** before repeating an overlay or selecting different molecules
- *Note*: Rings are not flexed during the overlay. Hydrogen atoms are ignored. The RMSDs reported for multiple overlays of the same pair of molecules may vary. The algorithm is not exhaustive and between overlays, may find different, closely related solutions.

5 Moving the Display and Setting the View Direction

5.1 Translating the Display

The contents of the display area can be translated in several ways:

- By holding the middle mouse button down while moving the cursor in the display area (three-button mouse only).
- By scrolling the mouse wheel to translate up/down or pushing the mouse wheel to either side (tilt click) to translate left/right (scroll wheel mouse only).
- By moving the cursor in the display area while keeping both the left-hand mouse button and the keyboard *Ctrl* key pressed down (or on macOS, the *Cmd* key).
- By using the arrow buttons in the tool bar:

- By using the dial box. Open the box by hitting **Display** in the top-level menu followed by **Dial box...** in the resulting pull-down menu. Then use the buttons labelled **+x**, etc., in the right-hand area of the dial box:

- The speed of translation can be changed by altering the translate/scale *Increment* value in the dial box.
- To perform operations using the keyboard see **Appendix B: Keyboard Modifiers**.

5.2 Rotating the Display

The contents of the display area can be rotated in several ways:

- By moving the cursor in the display area while keeping the left-hand mouse button pressed down (*x* and *y* rotation), or while keeping both the left-hand mouse button and the *Shift* key pressed down (*z* rotation).
- By using the buttons labelled **x-, y-**, etc. in the tool bar:

- By using the dial box. Open the box by hitting **Display** in the top-level menu followed by **Dial box...** in the resulting pull-down menu. Then use the buttons labelled **+x,** etc., in the left-hand area of the dial box:

- The speed of rotation can be changed by altering the rotate *Increment* value in the dial box.
- To perform operations using the keyboard see Appendix B: Keyboard Modifiers.

5.3 Animating the Display

Animating the contents of the display can be toggled on and off by clicking the check-box beside the **Animate…** button in the Mercury tool bar:

Clicking on the **Animate…** button will open the *Animate Structure* dialogue:

The following options are available:

- The nature of the animation, either rotation or oscillation, can be selected using the **Rotate** and **Oscillate** check-boxes.
- The axis of rotation can be selected using the ***x***, ***y*** or ***z*** radio buttons.
- The rate of rotation or oscillation can be set using the **Rate** slider.
- The amplitude of the oscillation can be set using the **Amplitude** slider.

Use of the *Animate Structure* tool bar can be helpful when assessing what movies will look like as rendered directly from Mercury using POV-Ray – for more information see POV-Ray Movie Rendering from Mercury.

5.4 Flipping the Display by 90 Degrees

The contents of the display area can be rotated by 90 degrees about any of three orthogonal directions by:

- Using the buttons labelled **x-90, x+90,** etc. in the tool bar:

- Using the dial box. Open the box by hitting **Display** in the top-level menu followed by **Dial box...** in the resulting pull-down menu. Then use the buttons labelled **-90, +90**, etc., in the left-hand area of the dial box:

5.5 Scaling the Display

The contents of the display area can be scaled (i.e. zoomed in or out) in several ways:

- By moving the cursor up and down in the display area while keeping the right-hand mouse button pressed down.
- By using the buttons labelled **zoom-** and **zoom+** in the tool bar:

- By using the dial box. Open the box by hitting **Display** in the top-level menu followed by **Dial box...** in the resulting pull-down menu. Then use the buttons labelled **zoom-** and **zoom+** in the right-hand area of the dial box:

- The zoom rate can be changed by altering the translate/scale *Increment* value in the dial box.
- To perform operations using the keyboard see Appendix B: Keyboard Modifiers.

5.6 Viewing Along Crystallographic Axes

The contents of the display area can be viewed along real cell axes (see Unit Cell) or reciprocal cell axes (see Reciprocal Cell Axes) in several ways:

- Using the buttons labelled **a, b** and **c** (real axes) or **a*, b*** and **c*** (reciprocal axes) in the tool bar:

- Selecting **Display** from the top-level menu, **View along** from the resulting pull-down menu, and then **a, b, c, a*, b* or c*** as appropriate.
- More specific viewing directions between either real cell or reciprocal cell axes can be viewed by selecting **Display** from the top-level menu, **View along** from the resulting pull-down menu, and then **Unit Cell Direction [uvw]** or **Miller Direction [hkl]** as appropriate. Viewing along a **Unit Cell Direction [uvw]** will allow you to

choose a specific view that is any combination of the unit cell axes a, b and c. Viewing along a **Miller Direction [hkl]** will allow you to choose a specific view that is any combination of the reciprocal cell axes a*, b* and c* - these directions are perpendicular to the equivalent Miller planes. The dialogue box will remain open with **Apply** as the default option to allow repeated use, to exit press the **Close** button.

- Right-clicking in the background, selecting **View** from the resulting menu, and then choosing the view direction from the resulting menu.
- The default viewing direction can be chosen using the **Default view**: drop-down menu on the main toolbar. You can choose any of the cell axes or reciprocal cell axes as the default direction. Alternatively, the **Custom** view option allows the user to define an arbitrary direction in relation to the cell axes down which to view the structure. By default, we allow the user to choose the standard crystallographic axes as options to view the cell, but the **Custom** option allows the user to pick a direction (by rotating the crystal structure in the viewer) and then applying that view to all subsequently loaded structures. So, if you want to look down the body diagonal of a set of related structures, you can align the first crystal structure with that view, choose **Custom** and the structures that are loaded subsequently will be displayed with the view down the body diagonal.

5.7 Viewing Along and Perpendicular to bonds

The view direction of a molecule can be changed to either view down a particular bond or perpendicular to a bond.

- To view down a bond, right-click on the bond and select **View along bond**.
- To view perpendicular to a bond, right-click on the bond and select **View perpendicular to bond**.
- Hit **Reset** (towards the bottom left of the Mercury window) to return the view to the default view.

5.8 Setting a Rotation centre

To enable rotations about a particular atom of a displayed structure:

- Right-click on an atom and choose the option **Rotation Centre**, followed by **Set Rotation Centre**.
- To return the rotation centre to the default location (roughly the centre of the display) right-click on either the backdrop or on an atom and choose the option **Rotation Centre** followed by **Reset Rotation Centre**.

5.9 Viewing Along and Perpendicular to Planes

If a Miller plane (see Creating Miller Planes) or least-squares plane (see Creating Least-Squares Mean Planes) has been defined, the view direction can be changed so that you are looking down the normal to the plane or down a direction that lies in the plane.

- To view down the normal, right-click on the plane and select **View perpendicular to plane** from the resulting pull-down menu.
- To view down a vector lying in the plane (i.e. edge-on to the plane), right-click on the plane and select **View along plane** from the resulting pull-down menu.

5.10 Changing the Display Perspective

The display perspective can be changed by moving the cursor up and down in the display area while keeping the right-hand mouse button and the *Shift* key pressed down.

To perform operations using the keyboard see **Appendix B: Keyboard Modifiers**.

6 Customising the Display

6.1 Setting General Display Properties

Note: It is possible to create and modify sets of Mercury display styles. This will allow multiple display set-ups to be tailored for different purposes, e.g. work, publication etc. (see Managing Display Styles).

6.1.1 Setting Background Colours

- You can switch between the default black background and an alternative colour by right-clicking in the background area and hitting **Draw Backdrop**. The alternative colour will be a blue gradient.
- For more options, hit **Display** in the top-level menu followed by **Display options...** in the next menu. Select *Background* from the resulting *Display Options* dialogue.
- To set a customised, uniform background colour, select the **Single Colour** option, click on the adjacent panel (which will be black by default) and choose a colour from the resulting palette (see Using the Select Color Dialogue Box).
- To create a background which changes gradually in colour, select the **Gradient** option, click on each of the four panels in turn and select colours from the resulting palette to achieve the effect of your choice.
- Select the **None (Black)** option or hit the **Defaults** button to revert to the default black background.

6.1.2 Depth Cueing

- Depth cueing aims to improve the 3D effect by displaying objects at the front more brightly than those at the back. The extent to which this improves 3D perception varies according to what is on display and it may not always be effective.
- Use the *Depth cue* check box underneath the main 3D display area to switch depth cueing on or off.
- To customise the depth-cue settings, hit **Display** in the top-level menu followed by **Display options...** in the next menu. Select *Depth Cueing* in the resulting *Display Options* dialogue.
- Type in values in the *Front brightness* and *Back brightness* boxes to control how bright an object at the front is compared to one at the back. The numbers must fall in the range 0.0 to 1.0.
- Select the *Equation* to be used for calculating the depth cueing. **Linear** or **Exponential2** give better results when dealing with large structures.
- The *Enabled* check box can be used to switch depth cueing on and off.
- Hit the **Defaults** button to revert to default depth-cue settings.

6.1.3 Labels

- To set the colour of all labels (i.e. all atom, centroid, plane and unit cell axes labels), hit **Display** in the top-level menu, then **Display Options...** in the pull-down menu. In the resulting *Display Options* dialogue select *Labels*, click on the **Label colour** panel (which will be white by default) and choose a colour from the resulting palette (see Using the Select Color Dialogue Box).
- To set the size of all labels (i.e. all atom, centroid, plane and unit cell axes labels), hit **Display** in the top-level menu, then **Display Options...** in the pull-down menu. In the resulting *Display Options* dialogue select *Labels*, and enter a new label size.
- The positions of atom labels can be adjusted manually to avoid them overlapping nearby objects in the display. Set the *Picking Mode* status (near the top of the Mercury window) to **Move Labels** then reposition each of the atom labels.
- The new position of labels will be honoured even if the display is rotated.
- Hit the **Defaults** button to revert to default label settings.

6.1.4 Setting Lighting Effects

- The display area can be thought of as being illuminated by ambient lighting (i.e. low lighting from all directions) and up to four directional lights. Ambient lighting and two of the directional lights are switched on by default. (The lighting has no effect if you are displaying everything in wireframe style.)
- To control the lighting, hit **Display** in the top-level menu followed by **Display options...** in the next menu. Select *Lighting* in the resulting *Display Options* dialogue.
- Settings for each of the four directional lights can be adjusted independently by selecting the appropriate panel *Light 1*, *Light 2*, etc.
- Switch each light on or off by use of the *Enabled* check boxes.
- Each directional light lies along a vector direction which can be altered by typing in new values in the *Position x*, *y* and *z* entry boxes. (1,1,1) lies above the top-right of the screen, (-1,-1,-1) lies below the bottom-left.
- Any colour can be altered by clicking on the colour panel and selecting from the resulting palette (see Using the Select Color Dialogue Box). *Diffuse colour* refers to the general background lighting. *Specular colour* controls the colour of the reflections off atoms when displayed in space-filling mode. You can set the specular colour to black to remove these reflections completely.
- In the case of ambient lighting, this can be switched on or off by use of the *Enabled* check box. The colour may be changed by clicking on the **Diffuse ambient colour:** panel and selecting from the colour palette (see Using the Select Color Dialogue Box).
- Hit the **Defaults** button to revert to default settings.

6.1.5 Z-Clipping

- Z-clipping restricts the contents of the rendered image based on the Z coordinate of individual objects. There is both a near clipping plane and a far clipping plane. Objects closer to the viewer than the front plane are not rendered, and objects further from the viewer than the rear plane are not rendered.

- The near and far clipping planes are measured from the global rotation centre, which is normally at the centre of the display, but can be moved by right-clicking on an individual atom and selecting **Rotation Centre** followed by **Set Rotation Centre**.
- To customise the positions of the near and far clipping planes, hit **Display** in the top-level menu followed by **Display Options...** in the next menu. Select *Z-Clipping* from the resulting Display Options dialogue.
- The **Enabled** check box can be used to switch Z-Clipping on and off.
- The Far Plane and Near Plane sliders allow interactive control over the position of the clipping planes. At any given slider setting, the top and bottom white boxes show the limits of the current display. Slider settings can take the values from -100 to +100.
- You can type distance limits directly into the white boxes, but you must terminate each number with a carriage return (i.e. Enter) for it to take effect.
- Moving the middle grey rectangle with the mouse alters both sliders simultaneously. By setting a narrow range, this therefore allows you to view slices through the rendered image.
- Hit the **Defaults** button to revert to default Z-Clipping settings.

6.1.6 Box-Clipping

- Box-clipping restricts the contents of the rendered image based on the size of a 3D box, fixed relative to the centre of the structure. The dimensions of the box are given by six coordinates along the X, Y and Z axes. Any objects outside of the given dimensions are not rendered.
- To view and customise the box-clipping view, hit **Display** in the top-level menu followed by **Display Options...** in the next menu. Select **Box-Clipping** from the resulting **Display Options** dialogue.
- The **Enabled** check box can be used to switch Box-Clipping on and off.
- In order to customise the box view, sliders allow interactive control over the position of the X, Y and Z distances. At any given slider setting, the six white boxes show the limits (X, -X, Y, -Y, Z, -Z) of the current display. Slider settings can take the values from -1000 to +1000.
- Distance limits can be entered directly into the white boxes, or by dragging the grey rectangle slider control button with the cursor. Left-clicking the slider to the left or right of the grey rectangle slider control button will increase or decrease the box size in small increments for fine control of the image.
- Hit the **Defaults** button to disable the Box-Clipping view and revert to default settings.

6.1.7 Viewing Structures in Stereo

6.1.7.1 Overview

- Stereo viewing works on the premise that most humans use binocular vision to perceive depth and see the world in 3D. Binocular vision relies on the fact that we have two eyes, which are approximately three inches apart. This separation causes each eye to see the world from a slightly different perspective, and the different views are then pieced together by the brain. The brain understands the differences

in the two views and uses them to calculate distance, creating our sense of depth and ability to gauge distance.

- To view structures in stereo it is necessary to have the appropriate set-up on your computer, whether this is an emitter with associated stereoviewing eye-wear, or a monitor converter with stereoviewing eye-wear. Providers of this hardware include the solution offered by NVidia which requires the following components:
  - NVidia 3D Vision Kit (glasses & IR emitter), *http://www.nvidia.co.uk/object/product_geforce_3D_VisionBundle_uk.html #geforce_specs*
  - Compatible NVidia graphics card, see *http://www.nvidia.com/object/3d-vision-requirements.html#* We recommend a mid-high end Geforce 500/400/300 series.
  - A 120Hz *3D-ready* display.
- There is also a "pro" solution offered by NVidia that is more expensive than the options listed above, however this may be unnecessary unless you are looking at a corporate solution and/or the ability for multiple people to view the 3D screen at once.
- Only if your computer is correctly configured for stereo viewing will the **Stereo** tickbox, beneath the main menu, be active. If your computer is not configured for stereo viewing or if the settings are not correctly configured, the **Stereo** tickbox will be greyed out.

6.1.7.2 Making Use of Stereo Viewing

- If your computer is correctly set-up for stereo viewing, and assuming the hardware required for viewing in stereo has been switched on, stereoscopic viewing can be activated using the **Stereo** tickbox beneath the main menu options in Mercury.
- Alternatively stereo viewing can be switched on using the **Enabled** tickbox in the *Stereo* pane of the *Display Options* window, obtained via **Display** followed by **Display Options** from the main toolbar.

6.1.7.3 Adjusting Stereo Viewing Settings

- The settings for stereo viewing are accessed via **Display** followed by **Display Options** from the main toolbar, and by clicking on *Stereo* in the resultant window.

- From within this window it is possible to control settings for:
  - *Separation*: use the slider to control the amount of separation between the left and right eye views.
    Note: Although larger values for *Separation* enhance the stereo view effect, these are more likely to give rise to eye strain and/or headaches.
  - *Focal Point*: use the slider to control how close the focal point is to the front or back of the structure(s) being displayed:
    -1 - places all structure(s) on display behind the monitor screen
    0 - places the middle of the structure(s) in the plane of the screen
    1 - places all the structure(s) on display in front of the monitor screen
    Note: Values between -1 and 0 are likely to reduce the effects of eye strain.
- Values for *Separation* and *Focal Point* can also be typed into the boxes to the right of the sliders.
- Default settings can be restored using the **Defaults** button.

6.1.7.4 Stereo Viewing: Troubleshooting

- Problems selecting atoms.
  - This can cause problems when viewing in stereo as what you are seeing is the average of two images. To overcome this problem, close one eye (we would recommend the left eye) and select the atom(s) this way.
- The stereo button is disabled in Mercury: stereo can't be turned on.
  - For some reason, the computer thinks that the stereo viewing hardware capability is not available. Check your hardware settings and set-up to ensure the hardware is correctly configured.
- Although stereo settings are turned on and I'm looking through the stereo glasses, I can still see two separate images and not one.
  - Check the emitter is turned on.

- Check the glasses are turned on.
- Are the glasses within range of the emitter?
- Do the batteries in the glasses need to be replaced?

6.2 Setting Global Structure Display Properties

6.2.1 Setting a Global Display Style

- To set all atoms to a new display style (Wireframe, Stick, Ball and Stick, Spacefill, Ellipsoid or Polyhedral), you must first ensure that no atoms are selected (see 2.6.4 Identifying Selected Atoms and Appendix B: Keyboard Modifiers.
- You can then use any of the following methods:
  - Set the required style in the tool-bar *Style* box, located near the top of the main Mercury window, e.g.

  - Right-click in the display-area background, pick **Styles** from the pull-down menu, and select the required style (**Wireframe, Capped sticks, Ball and stick, Spacefill, Ellipsoid, Polyhedral**).
  - Hit **Display** in the top-level menu, **Styles** in the resulting pull-down menu, then pick the required display style.
- To switch the display of bond-types on or off, use one of the following methods:
  - Right-click in the display-area background and select **Styles** followed by **Display Bond Types** from the resulting pull-down menu.
  - Hit **Display** in the top-level menu, **Styles** in the resulting pull-down menu, then click on **Display Bond Types** from the resulting pull-down menu.
- To identify aromatic rings by displaying a circle within each ring, use one of the following methods:
  - Right-click anywhere in the display-area background and select **Styles** followed by **Display Aromatic Rings** from the resulting menu.
  - Hit **Display** in the top-level menu, **Styles** in the resulting pull-down menu, then click on **Display Aromatic Rings** from the resulting pull-down menu.
- Display styles can also be set for individual atoms, bonds and molecules (see Setting Display Properties for Particular Atom(s) or Bond(s)).
- It is possible to create and modify sets of Mercury display styles, which are automatically saved between sessions (see Managing Display Styles).

6.2.2 Setting Sticks Display Options

To customise the stick display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Stick Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background and select **Styles** followed by **Stick Settings...** from the resulting pull-down menu.
- This will open the *Stick Display Options* window:

- The stick radii (in Angstroms) used for both the Capped Sticks display style and the unit cell edges (if shown as Capped Sticks) can be controlled interactively using the *Stick Radius* slider.
- To revert back to the original settings, click **Defaults**.

6.2.3 Setting Ellipsoid Display Options

Displacement ellipsoids can be displayed for structures where the CIF or SHELX *.res* file contains Atomic Displacement Parameter values (Uequiv and Uij values or Bequiv and Bij values) (see Setting a Global Display Style):

To customise the Ellipsoid display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Ellipsoid Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background and select **Styles** followed by **Ellipsoid Settings** from the resulting pull-down menu.
- This will open the *Ellipsoid Display Options* window which can be used to customise various settings.

- *Probability level (%)*: use this to alter the probability level (thus size) the ellipsoids are displayed at. For crystal structures determined from neutron diffraction data or at low temperature, you may want to increase the probability level; for structures determined at high temperature, you may want to reduce the probability levels.
- *Draw principal ellipses, in colour*: use this tick box to enable or disable the display of the principal ellipses as lines around the atomic ellipsoids. You can also choose the colour to use for displaying the principal ellipses.
- *Draw hydrogens as fixed-size spheres, of radius (Å)*: use this tick box to control the display of H atoms.
- *Draw Non-Positive-Definite atoms as cubes*: when this tickbox is enabled non-positive-definite atoms are shown as cubes, otherwise they are shown as very small spheres. A non-positive-definite atom is termed as such when the corresponding ellipsoid cannot be represented in real space, e.g. the ellipsoid may have a negative volume. Non-positive-definite atoms therefore highlight a problem with the data, e.g. a particular atom may have been determined as a carbon atom but it is in fact a nitrogen atom.
- Bonds can either be represented as *Wireframe* or *Stick* and if shown in *Stick* mode, the radius of the bonds can be modified using the adjacent slider.
- The settings are saved between Mercury sessions. Use the **Defaults** button to return to the standard settings.

Further information about ADPs is provided (see Atomic Displacement Parameters (ADPs)).

6.2.4 Setting Ball and Stick Display Options

To customise the ball and stick display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Ball and Stick Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background and select **Styles** followed by **Ball and Stick Settings...** from the resulting pull-down menu.
- This will open the *Ball and Stick Display Options* window:

- The *Atom Size* sliders allow interactive control over the size of atoms in the Ball and Stick display style. The white boxes show the current atom size (value times van de Waals radius). The atom size range can be set between 0.1 and 0.8 times the van der Waals radius. You can also type values directly into the white boxes, but you must terminate each number with a carriage return (i.e. Enter) for it to take effect.
- The size of metals and non-metals can be controlled independently.
- The bond radii (in Angstroms) used in the Ball and Stick style can also be controlled interactively using the *Bond Radius* slider.
- To revert back to the original settings, click **Defaults**.

6.2.5 Setting Spacefill Display Options

To customise the spacefill display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Spacefill Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background and select **Styles** followed by **Spacefill Settings...** from the resulting pull-down menu.
- This will open the *Spacefill Display Options* window:

- The *Atom Size* sliders allow interactive control over the size of atoms in the Spacefill display style. The white boxes show the current atom size (value times Van de Waals radius). The atom size range can be set between 0.1 and 1.0 times the van der Waals radius for non-metals, and between 0.1 and 1.5 times the van der Waals radius for

metals. You can also type values directly into the white boxes, but you must terminate each number with a carriage return (i.e. Enter) for it to take effect.

- The size of metals and non-metals can be controlled independently.
- The bond radii (in Angstroms) used in the Ball and Stick style can also be controlled interactively using the *Bond Radius* slider.
- To revert back to the original settings, click **Defaults**.

6.2.6 Setting Polyhedral Display Options

In the context of the Polyhedral Display options, 'Central Elements' refers to atoms around which polyhedra are centred (e.g. the silver atom in the above image) and 'Ligand Elements' are the neighbouring atoms which create the corners of the polyhedron (coloured grey, red and yellow in the image).

To customise the polyhedral display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Polyhedral Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background or an atom/polyhedron in the visualised structure and select **Styles** followed by **Polyhedral Settings...** from the resulting pull-down menus.
- This will open the *Polyhedral Display Options* window:

The *Connectivity options* section gives three choices:

- The **Include Pi bonds** checkbox (default is unselected) allows the option to include pi-bonded atoms as vertices. This may be useful for structures with unusual coordination environments e.g. η^1 coordination.
- The **Central Elements…** and **Ligand Elements…** buttons (with 'Central' and 'Ligand' as defined above) allow user control over the construction of polyhedra. By default, central elements include all metals and ligand elements are non-metals, however for certain structures custom settings may be required (e.g. frameworks including phosphate tetrahedra where phosphorus would be chosen as a central element). Both buttons open a window with a periodic table (see below) which allow you to add or remove elements as central or ligand atoms. Clicking OK will automatically update the view in the current Mercury window.

The *Display options* section has two parts:

- The **Draw edges** checkbox (default is selected) draws solid lines around edges of each polyhedron. If the Opacity option (below) is used, the polyhedral edges remain at full opacity.
- The **Opacity** slider allow interactive control over the transparency of the polyhedral display, from 0 (transparent) to 1 (fully opaque). The default value is 1, giving fully opaque polyhedra.

To revert back to the original settings, click **Defaults**.

6.2.7 Setting Contact Display Options

To customise the contact display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Contact Settings...** from the resultant pull-down menus. Alternatively, right-click in the display-area background and select **Styles** followed by **Contact Settings...** from the resulting pull-down menu.
- This will open the *Contact Display Options* window:

- The contact radii (in Angstroms) used for any displayed contacts shown as Capped Sticks can be controlled interactively using the *Contact Radius* slider.
- To revert back to the original settings, click **Defaults**.

6.2.8 Setting Measurement Display Options

To customise the measurements display settings:

- Click on **Display** in the top-level menu, then **Styles** and **Measurement Settings...** from the resultant pull-down menus.
- This will open the *Measurement Settings* window:

- The display settings of distances, angles and torsions can each be customised independently.

- To control the colour of the measurements, click on the relevant colour picker for that measurement type. Similarly, the style of display (wireframe or capped sticks) can be controlled using the relevant radio buttons.
- Finally, the thickness of the measurements (whether in wireframe or capped sticks mode) can be controlled using the **Thickness** sliders.
- To revert back to the original settings, click **Defaults**.

6.2.9 Setting a Global Colouring Scheme

- The following global colouring scheme are available:
  - Colour by element type (see Standard Element Colours).
  - Colour by symmetry equivalence (see Symmetry-Equivalence Colouring).
  - Colour by atomic displacement (see Atomic Displacement Colouring).
  - Colour by symmetry operation (see Symmetry Operation Colouring).
  - Colour by Gasteiger charge (see Gasteiger Charge Colouring).
  - Colour by Element or Suppression (see Colouring by Element or Suppression).
- To apply a colouring scheme you must first ensure that no atoms are selected (see (see **Appendix B: Keyboard Modifiers**).
- Identifying Selected Atoms). You can then use either of the following methods:
  - Set the required option in the tool-bar *Colour* box, located near the top of the main Mercury window, e.g.

  - Right-click in the display-area background, pick **Colours** from the pull-down menu, and select the required option (**Colour By Element, Colour by Symmetry Equivalence, Colour by Atomic Displacement, Colour by Symmetry Operation**).
- Colours can also be set for individual atoms and molecules (see Setting Display Properties for Particular Atom(s) or Bond(s)).
- It is possible to create and modify sets of Mercury display styles, which are automatically saved between sessions (see Managing Display Styles).

6.2.10 Changing Element Colour Settings

- When atoms are coloured by element (see Setting a Global Colouring Scheme), each atom is assigned a colour depending on its element type.
- To change the colour used for particular elements hit **Display** in the top-level menu, select **Colours** from the pull-down menu and then hit **Element Colours...** This opens a dialogue box which you can use to specify the desired settings:

- Colour settings for the full periodic table can be specified. To change the colour of a particular element click on the corresponding coloured panel. This will open the Select Color dialogue box which can be used for customising colours (see Using the Select Color Dialogue Box).

6.2.11 Custom Carbon Colouring

- The Custom Carbon option allows a colour to be specified for carbon atoms via the colour palette, leaving the colours of remaining atoms unchanged.
- To set a custom carbon colour right-click in the display and select **Colours**, and then **Custom Carbon...** from the resulting menus. This will open the *Select Colour* dialogue box (see Using the Select Color Dialogue Box).

6.2.12 Switching Hydrogens On and Off

- The display of hydrogen atoms can be turned on and off by use of the **Show hydrogens** tick box in the main Mercury window.
- If **Show hydrogens** is switched on but no H-atoms are visible, it is likely that hydrogen-atom positions were not located in the determination of the structure that is currently being viewed, i.e. they are missing from the list of 3D atomic coordinates.

6.2.13 Turning All Atom Labels On or Off

- To switch all atom labels on or off, you must first ensure that no atoms are selected (see (see **Appendix B: Keyboard Modifiers**).

Identifying Selected Atoms). You can then use any of the following methods:

- Switch on or off the **Label atoms** tick-box located near the bottom of the main Mercury window.
- Switch on or off the **Show Labels** tick-box located above the visualiser on the left-hand side. Options are available for choosing which type of atom to label and with what information (see Atom Labelling options).

- Right-click in the display-area background, pick **Labels** from the pull-down menu, and select the required option (**Show labels, Hide labels**).
- Hit **Display** in the top-level menu, **Labels** in the resulting pull-down menu, then pick the required option (**Label all**, **Clear all labels**).

Atom labels can also be turned on and off individually (see Setting Display Properties for Particular Atom(s) or Bond(s)). The positions of atom labels can be adjusted manually (see Labels).

6.2.14 Atom Labelling options

There are two drop down menus located above and to the right of the visualiser window that allow highly customisable atom label options.

- The first drop-down menu offers choices of atom selections:
  - All atoms
  - Carbon atoms
  - Hydrogen atoms
  - Non-hydrogen atoms
  - Non C,H
  - Metal atoms
  - Non-metal atoms
  - Contact atoms
  - Stereocentres
- The second drop down menu lists properties with which the atoms can be labelled, and they are:
  - Atom label
  - Element
  - Atomic number
  - Formal charge
  - Gasteiger charge
  - MOPAC partial charge
  - Sybyl type
  - Stereochemistry
  - Position: fractional
  - Position: orthogonal
  - Occupancy

- ADP type
- Uequiv
- Bequiv

- Selections made in the drop down menus will be displayed when the **Show Labels** tickbox or the **Label Atoms** tickbox (in the *Options* pane towards the bottom right of the main Mercury window) are checked.
- *Note*: Labelling of stereocentres and assignment of stereochemistry (see Chirality) will only be performed if bond type information is available. Bond types will be available for CSD or in-house database entries, or if your file loaded contains bond type data (such as MOL2), otherwise you will need to assign bond types. For more details on how to assign bond types in Mercury see Auto-Editing Options.

6.2.15 Colouring and Resizing All Labels

To set the colour of all labels (i.e. all atom, centroid, plane and unit cell axes labels), use one of the following methods:

- Hit **Display** in the top-level menu, then **Display Options...** in the pull-down menu. In the resulting *Display Options* dialogue select the *Label* tab, hit **Label Colour** and select from the resulting colour palette (see Using the Select Color Dialogue Box).
- Hit **Display** in the top-level menu, **Labels** in the resulting pull-down menu, then **Label colour...** and select from the resulting colour palette (see Using the Select Color Dialogue Box).
- Right-click in the display-area background, pick **Labels** from the pull-down menu, select **Label colour...** and choose from the resulting colour palette (see Using the Select Color Dialogue Box).

To set the size of all labels use one of the following methods:

- Hit **Display** in the top-level menu, then **Display Options...** in the pull-down menu. In the resulting *Display Options* dialogue select the *Label* tab and enter a new **Label Size**.
- Hit **Display** in the top-level menu, **Labels** in the resulting pull-down menu, then **Label size...** and enter a new label size in the resulting dialogue.
- Right-click in the display-area background, pick **Labels** from the pull-down menu, select **Label size...** and enter a new label size in the resulting dialogue.

The positions of atom labels can be adjusted manually (see Labels).

6.2.16 Customising the Display of Unit Cell Edges and Labels

- Unit cells can be displayed in several ways (see Displaying and Labelling Unit Cell Axes).
- To set the colour of a unit cell, right-click on a structure's unit cell edge, pick **Colour** from the pull-down menu and select a colour from the resulting menu.
- To make the unit cell axes more prominent, right-click on a unit cell edge, pick **Style** from the pull-down menu and choose **Capped Sticks**.
- The crystallographic a, b and c axes are normally coloured red, green and blue, respectively. To turn off this colouring, right-click on a structure's unit cell edge and click on **Color Unit Cell Axes** from the resulting pull-down menu.

- To switch unit cell axes labels on or off, right-click on a structure's unit cell edge, pick **Labels** from the pull-down menu, and select the required option (**Show labels, Hide labels**). Cell axes labels can also be turned off in the *Packing and Slicing* dialogue box (see Displaying and Labelling Unit Cell Axes). The size and colour of unit cell axes labels can also be set (see Colouring and Resizing All Labels).

6.3 Setting Display Properties for Particular Atom(s) or Bond(s)

There are several methods for setting the display properties of a particular atom, bond, molecule, or set of atoms:

- Select the atom and bonds whose display properties you wish to change (see Ways of Selecting Atoms and Molecules). Right-click in the display-area background, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then select the desired display-property setting from the next menu.
- Select the atoms and bonds whose display properties you wish to change (see Ways of Selecting Atoms and Molecules). Hit **Display** in the top-level menu, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then hit **Atoms...** or **Bonds...** to obtain a dialogue box which you can use to specify the required settings (see Using the Atom Styles, Atom Labels, Atom Colours and Show/hide Dialogue Boxes).
- With no atoms or bonds selected, hit **Display** in the top-level menu, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then hit **Atoms...**, **Selected atoms...** or **Bonds...** This opens a dialogue box which you can use to (*a*) specify the desired settings, and then (*b*) select the atoms, bonds or molecule to which those settings are to be applied (see Using the Atom Styles, Atom Labels, Atom Colours and Show/hide Dialogue Boxes).
- Right-click on an individual atom or bond, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then select the desired display-property setting from the next menu. The chosen setting will be applied to the atom or bond on which you clicked.

In addition:

- You can switch individual atom labels on and off by setting the *Picking Mode* to **Label** in the tool bar near the top of the main Mercury window. Left-clicking on an atom will then toggle its label on and off.

6.4 Using Display-Property Dialogue Boxes

6.4.1 Using the Atom Styles, Atom Labels, Atom Colours and Show/hide Dialogue Boxes

- When setting atom display properties you may be presented with one of the following dialogue boxes:

- When setting bond display properties you may be presented with one of these dialogue boxes:

- If there were atoms selected before the dialogue box was opened (see (see **Appendix B: Keyboard Modifiers**).
- Identifying Selected Atoms), only the top part of the box will be shown and there will be **Cancel** and **OK** buttons at the bottom instead of **Close**.
- All the dialogue boxes work in basically the same way.
- Select the desired display-property setting by switching on the appropriate button in the top area of the box.
- If atoms were selected before the dialogue box was opened, all you then have to do is hit **OK.** This closes the dialogue box and applies the setting you have chosen to the selected atoms.
- If no atoms were selected before the dialogue box was opened, use the buttons in the lower half of the box to apply the chosen display-property setting to:

- **Picked atom(s)** - click on the required atoms.
- **Picked molecule(s)** - click on one atom of the required molecule(s).
- All **Atoms**, all **Non-hydrogen** atoms only, or all **Hydrogen** atoms only (options not present in *Bond Styles* and *Bond Colours* dialogue boxes).

Then hit **Close** to close down the dialogue box.

- The *Atom Colours* dialogue box allows you to colour **By Element** (see Standard Element Colours), **By Symmetry equivalence** (see Symmetry-Equivalence Colouring) or assign a customised **Colour**. If you select the latter, you should then click on the coloured panel to choose the colour you want; this will open the *Select Color* dialogue box (see Using the Select Color Dialogue Box).

6.4.2 Using the Select Color Dialogue Box

- You may need to use the *Select Color* dialogue box when choosing a colour for an atom, molecule or other object:

- There are several ways to select a colour:
  - Click on one of the *Basic colors.*
  - Click on the *colour palette*.
  - Click on one of the *Custom colors* (only possible if you have defined some earlier in the Mercury session).
  - Click in the vertical *lightness slider.*
  - Type your required *Red, Green* and *Blue* values.
  - Type your required *Hue, Saturation* and *Lightness* (i.e. *Val.*) values. The *Hue* defines the colour, e.g. *0* is red, *120* is green, and *240* is blue. The *Saturation* defines the amount of colour; *255* is fully saturated, whereas a colour with a saturation of zero is always white. The *Lightness* can also be controlled; a *Lightness* of zero always gives black.
- Once you have chosen the required colour, hit **OK** to continue.
- To create a customised palette, you can choose a colour and then hit **Add To Custom Colors**. This adds the current colour to the *Custom colors* area for use later in the same session.

6.5 Atom and Molecule Colouring Conventions

6.5.1 Standard Element Colours

- When atoms are coloured by element (see Setting a Global Colouring Scheme), each atom is assigned a colour depending on its element type.
- The default colours used for the common elements are as shown below:

- Unknown atoms types are given in pale green (bottom left of the view above).
- To find this file:
  - If you are on a Windows PC, it will be in the folder *examples\PeriodicTable.mol2* in the folder where Mercury is installed. For Mercury this is most likely to be:

    C:\Program Files\CCDC\CSD_2020\Mercury\examples\PeriodicTable.mol2
  - On Unix, it will be in the directory *examples/mercury/PeriodicTable.mol2* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the file will be in:

    /home/cambridge/examples/mercury/PeriodicTable.mol2
  - On macOS, it will be in the directory *Examples/Mercury* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

    Applications/CCDC/CSD_2020/Examples/Mercury/PeriodicTable.mol2

- It is possible to change the colour used for each individual element (see Changing Element Colour Settings).

6.5.2 Symmetry-Equivalence Colouring

- If all atoms are coloured **by Symmetry equivalence** (see Setting a Global Colouring Scheme), each different molecule (or ion) in the crystal chemical unit is assigned a different colour (the crystal chemical unit is the same as the asymmetric unit in the majority of structures (see Crystal Chemical Unit).
- When a packing diagram is then constructed, any given molecule is assigned the same colour as that of the molecule in the crystal chemical unit to which it is related by crystallographic symmetry.
- This means that molecules of the same colour are crystallographically (and therefore chemically and geometrically) identical to each other.
- It is possible to customise the colours used for *Colour by Symmetry equivalence*. Choose **Display** from the top-level menu, then select **Colours** followed by **Symmetry Equivalence colours...** This will open the **Symmetry Equivalence Colours** dialogue.
- For example:
  - If the crystal chemical unit of a crystal contains only one molecule, and a packing diagram of several unit cells is constructed, all molecules will have the same colour, e.g.

  - If the crystal chemical unit contains two chemically different molecules or ions, each will be assigned a different colour and the molecules in the packing diagram will have two colours, e.g.

- If the crystal chemical unit contains two molecules which are chemically identical but not related by crystallographic symmetry, each will be coloured differently and, again, the packing diagram will contain molecules of two colours, e.g.

6.5.3 Atomic Displacement Colouring

- When atoms are coloured by atomic displacement (see Setting a Global Colouring Scheme), each atom is assigned a colour based on its calculated Uequiv value. *Note:* Uequiv values can be shown in the Atoms spreadsheet (see Spreadsheet Display of Atoms and Their Properties).
- The following colours are used:
  - If there is no atomic displacement info for the atom, the atom is coloured grey.
  - Uequiv < 0.001: Grey.
  - Uequiv = 0.001: Blue.
  - Uequiv between 0.001 and 0.1 - a sliding scale from Blue to Green.
  - Uequiv = 0.1: Green.
  - Uequiv between 0.1 and 0.2 - sliding scale from Green to Red.
  - Uequiv = 0.2: Red.
  - Uequiv > 0.2: Red.

6.5.4 Symmetry Operation Colouring

- If all atoms are coloured **by Symmetry operation** (see Setting a Global Colouring Scheme), each molecule or ion is coloured according to the symmetry relationship it has with the asymmetric unit. If packing is not turned on, the contents of the asymmetric unit will appear in a light grey.

- When a packing diagram is then constructed, the molecules or ions are coloured by their symmetry relationship to the asymmetric unit. The relationships between symmetry and colour are as follows:
    - inversion symmetry - golden yellow.
    - 2-fold screw axes and 2-fold rotation axes are coloured a lighter shade and a darker shade of green, respectively.
    - glide planes – magenta.
    - 3-fold axes (rotation axes, screw axes, and roto-inversion axes) are coloured various shades of yellow.
    - 4-fold axes (rotation axes, screw axes, and roto-inversion axes) are coloured various shades of red.
    - 6-fold axes (rotation axes, screw axes, and roto-inversion axes) are coloured various shades of mid-blue.
    - mirror symmetry - light blue.
- An example of a structure, coloured by symmetry operation, in C2/*c* is shown below. *Note*: The molecules related to the asymmetric unit through a 2-fold screw axis are coloured a darker green (towards the top of the diagram) than the molecules related by the 2-fold rotation axis (towards the bottom of the diagram).

- Further information about the symmetry operators and the colours used to display the symmetry generated molecules is given in the *Symmetry* pane of the **More Info** dialogue (see Display of Text Information).

6.5.5 Gasteiger Charge Colouring

- When atoms are coloured by Gasteiger charge (see Setting a Global Colouring Scheme), each atom is assigned a colour based on the value of the Gasteiger charges.
- Increasing red colour represents increasing positive charge and increasing blue colour represents increasing negative charge. The colour used for a particular charge value is obtained by interpolating between colour values. Red is assigned to charges of +1 or above and blue is assigned to charges of -1 and below. Intermediate charge values are obtained by a linear interpolation between red and blue.

6.5.6 Colouring by Element or Suppression

- When the main menu *Colour* option is set to **by Element or Suppression**, all suppressed atoms in the 3D view are represented in the same (*Unknown*) colour so that they can be easily distinguished from other, non-suppressed atoms. All other atoms are shown in their usual default element colour. In the example below the suppressed, disordered atoms are coloured as light green.

6.6 Bond-Type Display Conventions

- Single, double, triple and quadruple bonds are indicated by one, two, three and four lines, respectively.
- A circle within a ring indicates that all the ring bonds have the aromatic bond type.
- If some, but not all, bonds in a ring have the aromatic type (which is not chemically reasonable but can happen), no circle is displayed in the centre of the ring but each aromatic bond is indicated by a small red circle going through the bond.
- Pi bonds to metals are indicated by single broken lines.
- Delocalised bonds are indicated by double broken lines.
- The display of bond types and aromatic rings can be disabled by clicking on the **Display** button in the top-level menu, selecting **Styles...** and switching off **Display Bond Types** and **Display Aromatic Rings** from the resulting menus. Alternatively, right click in the display background select **Styles...** and switch off **Display Bond Types** and **Display Aromatic Rings** from the resulting menus.

6.7 Managing Display Styles

6.7.1 Overview

- The Mercury *Style Manager* contains a number of pre-defined display styles. Each style has a name, and defines colour and style selections for objects (including atoms, bonds, contacts, cell edges etc.) as well as settings for background, label and element colours. For your convenience, a shortcut to the *Style Manager* is provided via the *Manage Styles* pull-down in the top-level menu of Mercury:

- Three general, pre-defined display styles are provided (*Work*, *Publication*, and *Presentation*). The settings used for these pre-defined display styles are summarised in the table below:

Style	Work	Publication	Presentation
Background Colour	Black	White	Pale yellow
Atom and Bond drawing style	Wireframe	Ball and Stick	Ball and Stick
Contact drawing style	Wireframe	Stick	Stick
Cell edges colour	White	Black	Black
Label colour	White	Black	Black
Element colours	Standard	Colour of elements C, Ca, Cl, D, F, H, N, O, P, S slightly adjusted to be better distinguishable when printed greyscale	Same as Publication

- Two further styles are provided:
  - ORTEP, which aims to generate an ORTEP-style plot. A gray-scale diagram is produced with ellipsoids displayed, if present in the input file.
  - 3D Print, which aims to increase the general default radii for cylinders (including both bonds and intermolecular contacts) to give you a good starting point for a successful 3D print (see Saving Mercury Displays as 3D Printing Files).
- The *Style Manager* also provides a mechanism for creating new display styles and modifying existing styles. Changes are saved between sessions.
- Display styles can be copied from other users or other computers, and loaded into Mercury thus providing a mechanism for sharing display styles. It is hoped that journal publishers may make available for download their recommended publication styles.

6.7.2 Loading and Applying Display Styles

- Click on the **Manage Styles...** button in the *Display* toolbar. Alternatively, hit **Display** in the top-level menu, and then select **Manage Styles...** This will open the *Style Manager* window.
- To Apply one of the default display styles (*Work*, *Publication*, or *Presentation*), or to apply an existing user-defined style, select the required style from those listed in the *Style Manager* and click on the **Apply** button. The main Mercury display will update automatically to reflect the selected style.

- *Mercury Style Definitions* can be copied from other users or other computers, and loaded into Mercury, thus providing a mechanism for sharing display styles. To load a *Mercury Style Definition* click on the Open... button, and specify a filename ending in *.msd*. The loaded style will be listed in the *Style Manager.*

6.7.3 Creating and Editing Display Styles

- Click on the **Manage Styles...** button in the *Display* toolbar. Alternatively, hit **Display** in the top-level menu, and then select **Manage Styles...** This will open the *Style Manager* window:

- To create a new display style based on either the current display settings, or one of the default display styles (*Work*, *Publication*, or *Presentation*):
  - Click on the New... button.
  - In the resulting *New Style* dialog select to use the **Current display settings,** as a starting point for the new display style. Alternatively, select to use one of the default display styles (**Work**, **Publication**, **Presentation**, **ORTEP**, or **3D Print Style**) as a starting point.
  - To create the new style (without making any changes) hit **Create...** and enter a name for the style when prompted. Alternatively, if you wish to modify any style settings click on **Create and Edit...** This will launch the *Style Settings* window: from here you can change the colour and style for Mercury objects, adjust lighting, background, depth cueing and labelling options, and change element colour and ellipsoid display settings (see Style Settings).
  - New display styles are automatically saved between Mercury sessions.
- To edit an existing display style:
  - Select the style you wish to edit from those listed in the *Style Manager*.

- To make a copy of an existing style before editing any settings hit **Copy...** then and enter a name for the style when prompted. The copy can then be selected from the *Style Manager.*
- Click on the **Edit...** button. This will launch the *Style Settings* window, from here you can change the colour and style for Mercury objects, adjust lighting, background, depth cueing and labelling options, and change element colour and ellipsoid display settings (see Style Settings).
- Changes to existing display styles are automatically saved between Mercury sessions.

6.7.4 Style Settings

- When defining or editing a display style the *Style Settings* window is used to change the colour and style for Mercury objects, adjust lighting, background, depth cueing and labelling options, and change element colour and ellipsoid display settings.

- The colour and style of Mercury objects such as atoms, bonds, contacts, cell edges, centroids and planes can be set by selecting from the appropriate drop-down menus.
- Click on the **Display options...** button to change lighting, background, depth cueing and labelling options (see Setting General Display Properties).

- To change the colour of particular elements, click on the **Element Colours...** button. The resulting *Element Colour Options* dialog can be used to specify the element colours for all of the periodic table. Left-click on the element colour in the list that you wish to change, this will open the *Select Color* dialogue box which you can use to choose the required colour (see Using the Select Color Dialogue Box).
- To customise ellipsoid display settings, click on the **Ellipsoid settings...** button. This will launch the *Ellipsoid Display Options* dialog (see Setting Ellipsoid Display Options).
- To customise ellipsoid display settings, click on the **Polyhedral settings...** button. This will launch the *Polyhedral Display Options* dialog (see Setting Polyhedral Display Options).
- When you are happy with the selections made hit either **Apply** (window will remain open and display will update to reflect the settings selected), or **OK** (closes window, changes are not applied). If you are creating a new style you will be prompted to enter a style name. New display styles and changes to existing styles are automatically saved between Mercury sessions.
- If editing one of the default display styles (*Work*, *Publication*, or *Presentation*) hit **Default** to return to the factory settings for that style.
- To close the Style Settings window without saving any changes hit **Cancel**.

6.7.5 Sharing, Deleting and Renaming Display Styles

- Click on the **Manage Styles...** button in the *Display* toolbar. Alternatively, hit **Display** in the top-level menu, and then select **Manage Styles...** This will open the *Style Manager* window.
- A *Mercury Style Definition* may be saved in *.msd* format by selecting the style from those listed in the *Style Manager,* hitting the **Save As...** button and specifying a filename ending in *.msd*. These style files can be copied to other users or other computers and loaded into Mercury thus providing a mechanism for sharing display styles. It is hoped that some journal publishers may make available for download their recommended publication styles.
- To delete a display style, select the style from those listed in the *Style Manager* and hit **Delete...**
  Note: It is not possible to delete any of the pre-defined display styles (*Work*, *Publication*, or *Presentation*).
- To rename an existing display style select the style from those listed in the *Style Manager,* hit **Rename...** and then enter a new name when prompted.
  Note: It is not possible to rename any of the pre-defined display styles (*Work*, *Publication*, or *Presentation*).

7 Centroids and Planes

7.1 Creating Centroids

The centroid of a set of atoms can be created in two ways:

Method 1:

- Choose **Calculate** from the top-level menu and **Centroids...** from the pull-down menu. Hit **New Centroid...** in the resulting dialogue box.
- This opens the *Centroid Properties* dialogue box, which allows you to create the least-squares centroid of:
  - A set of two or more atoms: switch on the **Picked atoms** button, click on the required atoms in the display, then hit **OK**.
  - A ring system: switch on the **Ring** button, click on one atom in the required ring, then hit **OK**.
  - A molecule or a set of molecules: switch on the **Picked molecule** button, click on one atom in each of the required molecules, then hit **OK**.
  - **All** atoms in the display, all **Non-hydrogen** atoms in the display, or all **Hydrogen atoms** in the display.
- You can use the options in the bottom part of the *Centroid Properties* dialogue box to set display properties such as label and colour (see Setting the Display Properties of Centroids and Planes).

Method 2:

- Select two or more atoms (see Ways of Selecting Atoms and Molecules).
- Right-click anywhere in the display-area background, select **New** from the pull-down menu and then **Centroid** in the next menu.
- The new centroid will appear. Its display properties (colour, label, etc.) can be edited if required (see Setting the Display Properties of Centroids and Planes).

7.2 Creating Least-Squares Mean Planes

The least-squares mean plane of a set of atoms can be created in two ways:

Method 1:

- Open the *Plane Properties* dialogue box:
  - by choosing **Calculate** from the top-level menu, **Planes...** from the pull-down menu and then clicking on **New Plane...** in the *Planes* dialogue box;
  - or by choosing **Calculate** from the top-level menu, **Packing/Slicing...** from the pull-down menu, clicking on the **Show slice** checkbox and then clicking on the **New...** button in the *Packing and Slicing* dialogue box. This will automatically show a slice of the crystal structure, and properties such as slice depth, slice area and the displacement of the slice bounds from the plane can be controlled with the slider bars.

- Select the **Mean plane** tab in the *Plane Properties* dialogue box. This allows you to create the least-squares mean plane of:
    - A set of three or more atoms: switch on the **Picked atoms** button, click on the required atoms in the display, then hit **OK**.
    - A ring system: switch on the **Ring** button, click on one atom in the required ring, then hit **OK**.
    - A molecule or a set of molecules: switch on the **Picked molecule** button, click on one atom in each of the required molecules, then hit **OK**.
    - **All** atoms in the display, all **Non-hydrogen** atoms in the display, or all **Hydrogen atoms** in the display.
- You can use the options in the bottom part of the *Plane Properties* dialogue box to set display properties such as label and colour (see Setting the Display Properties of Centroids and Planes).

Method 2:

- Select three or more atoms (see Ways of Selecting Atoms and Molecules).
- Right-click anywhere in the display-area background, select **New** from the pull-down menu and then **Mean Plane** in the next menu.
- The new plane will appear. Its display properties can be edited if required (see Setting the Display Properties of Centroids and Planes).

7.3 Creating Miller Planes

A Miller Plane (see Miller Plane) can be created as follows:

- Open the *Plane Properties* dialogue box:
    - by choosing **Calculate** from the top-level menu, **Planes...** from the pull-down menu and then clicking on **New Plane...** in the *Planes* dialogue box;
    - or by choosing **Display** from the top-level menu, **Packing/Slicing...** from the pull-down menu, clicking on the **Show slice** checkbox and then clicking on the **New...** button in the *Packing and Slicing* dialogue box. This will automatically show a slice of the crystal structure, and properties such as slice depth, slice area and the displacement of the slice bounds from the plane can be controlled with the slider bars (see Displaying Slices Through Crystals).
- Select the **hkl** tab in the *Plane Properties* dialogue box. This allows you to create a new Miller plane by entering the appropriate integer indices *h, k, l*.
- You can use the options in the bottom part of the *Plane Properties* dialogue box to set display properties such as label and colour (see Setting the Display Properties of Centroids and Planes).

7.4 Setting the Display Properties of Centroids and Planes

Centroids and planes have several display properties, viz. colour; transparency (i.e. whether you can see through the object or not); label (i.e. what the object is called); whether the label is displayed or not; whether the centroid or plane itself is displayed or hidden. These properties may be set in several ways:

Method 1:

- Choose **Calculate** from the top-level menu and **Centroids...** or **Planes...** from the next menu.
- The resulting dialogue box lists the current set of centroids or planes, e.g.

- Left-click on the required centroid or plane and hit **Edit...**
- Use the resulting *Properties* dialogue box to set the desired display properties. Clicking on the coloured panel will open the *Select Color* dialogue box which can be used for customising colours (see Using the Select Color Dialogue Box).
- In the example below, the user has set *Show* on (i.e. the object will be displayed, not hidden), the label on (i.e. a label *MY PLANE* will be displayed), the transparency off (i.e. the plane will be opaque) and the colour to green:

- For planes, the *Properties* dialogue box is also accessible by choosing **Calculate** from the top-level menu, **Packing/Slicing...** from the pull-down menu, selecting the required plane by clicking on the down-arrow next to the box labelled *Select Plane* in the *Packing and Slicing* dialogue, and then hitting the **Edit...** button.

Method 2:

- Right-click on the plane or centroid in the display area, select **Colours**, **Labels** or **Show/Hide** as required, then select the desired option.
- For example:
  - Selecting **Label** followed by **Edit label...** will allow you to type in a new name for the object.
  - Selecting **Colours** followed by **Blue** will turn the object blue.
  - Selecting **Show/Hide** followed by **Hide** will undisplay the object.

- Selecting **Colours** followed by **Colour...** will open the *Select Color* dialogue box which can be used for customising colours (see Using the Select Color Dialogue Box).

Method 3:

- Choose **Display** from the top-level menu, then select **Labels**, **Colours** or **Show/hide** in the pull-down menu, depending on which property you wish to alter.
- Select either **Centroids...** or **Planes...** from the next menu.
- The resulting dialogue box lists the current set of centroids or planes:
  - For colours, left-click on the item in the list that you wish to change. This will open the *Select Color* dialogue box which you can use to choose the required colour (see Using the Select Color Dialogue Box).
  - For labels and show/hide, set the *Action* buttons to the required setting and then left-click on the item in the list that you wish to change.
- Alternatively, right-clicking on an item in the list will produce the menus described above under *Method 2.*

Note: Changes made to the colour and/or size of centroid and plane labels will be applied to all labels in the display (i.e. all atom, centroid, plane and unit cell axes labels) (see Colouring and Resizing All Labels).

7.5 Deleting Centroids and Planes

Centroids and planes can be deleted (i.e. completely removed, not hidden) in several ways:

- Right-click on a centroid or plane in the display. In the resulting pull-down menu, select either **Delete Object** to delete the centroid or plane you have clicked on, or **Clear All Objects** to delete all centroids and planes.
- Choose **Calculate** from the top-level menu and select **Centroids...** or **Planes...** as appropriate. In the resulting dialogue box:
  - left-click on an item in the list box and hit **Delete**;
  - or right-click on an item in the list and select either **Delete Object** to delete the centroid or plane you have clicked on, or **Clear All Objects** to delete all centroids and planes.

8 Measurement

8.1 Measuring Distances, Angles and Torsions

- Measurement mode can be set in two ways:
  - By setting the *Picking Mode* in the tool bar (near the top of the main Mercury window) to the required parameter type, viz. **Measure Distance, Measure Angle** or **Measure Torsion.**
  - By right-clicking in the display area (but not on an atom, bond or object), selecting **Measure from the pull-down menu**, and then choosing **Measure Distances, Measure Angles** or **Measure Torsions**.
- Geometrical measurements (intramolecular or intermolecular) can be displayed by clicking on e.g., two atoms for a distance, three atoms for an angle or four atoms for a torsion angle.
- The display settings for geometrical measurements can also be controlled individually (see Setting Measurement Display Options).
- Centroids can be used in distance, angle and torsion angle measurements.
- Planes can be used in distance measurements with other planes, atoms and centroids. When planes are close to parallel, but not exactly parallel, the measured distance between them will always be 0.00 Å. This is due to the fact that the planes effectively extend to infinity (even though they are only shown with a finite plane in the visualiser) and will therefore intersect at some point. An alternative approach when assessing distances involving not-quite-parallel planes calculated from groups of atoms might be to calculate a centroid and measure the distance from a centroid to a plane. Angles between planes can also be measured. To measure the angle between two planes, select the **Measure Angles** option from the main menu *Picking Mode* pulldown.

Once this option has been selected, click on each plane in turn and the angle between the planes will be displayed. Not that you can identify if a plane is selected as the corners of the plane will be highlighted with yellow spheres.

8.2 Clearing Measurements

- All geometrical measurements can be removed from the display by:
  - Clicking on **Clear Measurements** in the tool bar near the top of the main Mercury window.
  - Right-clicking in the display-area background and then hitting **Clear Measurements** in the resulting pull-down menu.
  - Right-clicking in the display-area background, selecting **Measure** in the resulting pull-down menu, and then hitting **Clear Measurements** in the next menu.

- An individual measurement can be removed by right-clicking on any part of the dotted line corresponding to the measurement that does not lie on a chemical bond and selecting **Delete Measurement** from the resulting pull-down menu.

9 Crystal Structure Display Options

9.1 Displaying and Labelling Unit Cell Axes

Unit cell axes (see Unit Cell) can be displayed in several ways:

- Switch on the **Show cell axes** box (situated underneath the display area in the main Mercury window). By default, the axes will be labelled (see Customising the Display of Unit Cell Edges and Labels).
- Hit **Calculate** in the top-level menu and then select **Packing/Slicing...** The resulting dialogue box provides tick boxes to **Show cell axes** and **Label cell axes**.
- There is also a **Show cell axes** box in the **Build Networks** pane of the *Contacts* dialogue box (obtainable by selecting **Calculate** in the top-level menu and then **Contacts....** in the next menu).
- Alternatively, to hide a unit cell, right-click on a structure's unit cell edge, pick Show/Hide from the pull-down menu, and select hide from the resulting menu.

It is also possible to customise the display of unit cells edges and labels (see Customising the Display of Unit Cell Edges and Labels).

9.2 Colouring Conventions for Unit Cell Axes

By default, the crystallographic *a, b* and *c* axes are coloured red, green and blue, respectively. These colouring conventions can be customised (see Customising the Display of Unit Cell Edges and Labels).

9.3 Displaying Packing Diagrams (Assemblies of Unit Cells)

9.3.1 Packing Ranges

- Molecules and ions will be included in a packing diagram (i.e. a display showing one or more unit cells or fractions of unit cells) if their geometrical centres lie within the current *packing ranges*.
- When a new structure is loaded, the packing range along each crystallographic axis is automatically set to 0.0 to 1.0. This means that, by default, packing diagrams will include molecules whose centroids have fractional crystallographic coordinates (see Fractional Coordinates) falling between 0.0 and 1.0 in all three (*a,b,c*) directions.
- Packing ranges can be edited (see Displaying Several Unit Cells and Fractions of Unit Cells) and therefore packing diagrams can be customised to show any number of unit cells in any direction.

9.3.2 Displaying the Contents of a Single Unit Cell

The contents of a single unit cell (see Unit Cell) can be displayed in several ways:

- Turn on the **Packing** tick box (situated underneath the display area in the main Mercury window). This will display one unit cell provided the packing ranges (see Packing Ranges are at their default values (0.0 to 1.0 in all directions)).

- Right-click in the display-area background, select **Packing** from the pull-down menu and **Packing** from the next menu. This will display one unit cell provided the packing ranges are at their default values.
- Hit **Calculate** in the top-level menu followed by **Packing/Slicing...** in the next menu. In the resulting dialogue box, ensure that the packing ranges are set to their default values of 0.0 to 1.0 (by using **Reset** if required) and then turn the **Pack** tick box on.

9.3.3 Displaying Several Unit Cells and Fractions of Unit Cells

- Hit **Calculate** in the top-level menu and then **Packing/Slicing...** to obtain the *Packing and Slicing* dialogue box.
- Turn the **Pack** tick box on, and then set the desired packing ranges (see Packing Ranges) along the *a, b* and *c* axes (if you do this by directly typing into the entry boxes, you will have to hit the *Enter* key on your keyboard, or *Tab* to the next box, before the new settings take effect). For example:

Note: Packing ranges set using the above dialogue box will be applied to all subsequently viewed structures, even if packing has been switched off and switched back on at a later time. To reset the packing ranges to 0 to 1.0 on all axes, hit the **Reset** button.

- Four options are available to customise the choice of atoms included in the defined packing range. The options are to include atoms:
  - **…that fit** Displays only the atoms whose coordinates are at or within the boundaries of the packing range specified. This option may lead to fragments of molecules being displayed.
  - **…in molecules whose centroids fit** Displays only the molecules whose centroids lie at or within the boundaries of the packing range specified. Default option when Mercury starts up.
  - **…in molecules where Any atom fits** Displays whole molecules for which any constituent atom lies at or within the boundaries of the packing range specified.
  - **…in molecules where All atoms fit** Displays whole molecules for which all constituent atoms lie at or within the boundaries of the packing range specified.
- The **Reset** button can be used to return the ranges to their defaults (0.0 to 1.0 in all directions).

9.3.4 Switching Packing On and Off

You can switch between displaying a packing diagram or a single molecule or, more precisely, the molecule(s) in the Crystal Chemical Unit (see Crystal Chemical Unit) by:

- Using the **Pack** on/off tick box in the *Packing and Slicing* dialogue box.
- Using the **Packing** on/off tick box situated underneath the display area in the main Mercury window.
- Right-clicking in the display-area background, selecting **Packing** from the pull-down menu, and then using the **Molecule** or **Packing** options to switch between representations.

9.3.5 Automatically placing contents of the asymmetric unit within the unit cell

You can ensure that the contents of the asymmetric unit are displayed within the unit cell by:

- Using the **Auto Centre** tick box located underneath the **Packing** tick box.
- An algorithm is applied which moves the molecules of the asymmetric unit such that they lie as close as possible to 1/4, 1/4, 1/4.
- Ticking the **Auto Centre** tick box when a structure is already loaded will apply the algorithm to shift the asymmetric unit.

9.4 Displaying Slices Through Crystals

9.4.1 What a Slice Is

Mercury allows you to display molecules that satisfy two requirements:

- Their centroids occur within a specified packing range (see Packing Ranges).
- Their centroids are within a specified perpendicular distance of a user-defined least-squares or Miller plane.

This is a *slice* through the crystal.

9.4.2 Defining and Displaying a Slice

- Hit **Calculate** in the top-level menu and choose **Packing/Slicing...** to obtain the *Packing and Slicing* dialogue box.
- Set packing ranges (see Packing Ranges).
- Specify the plane of interest. This can be done by:
  - Clicking on the **Show Slice** checkbox and selecting a plane you have previously created; a list is obtainable by clicking on the down-arrow next to the box labelled *Select Plane:*

- Creating a new least-squares plane (see Creating Least-Squares Mean Planes) or Miller plane (see Creating Miller Planes) by using the dialogue box obtained by hitting the **New...** button.

- All molecules whose centroids lie in the specified packing ranges and between the grey boundary planes will now be displayed, e.g.

- To change the depth of the slice, use the *Depth* slider bar or type a new value into neighbouring box.
- To change the area of the slice, use the *Area* slider bar or type a new value into the neighbouring box.
- To displace the slice, use the *Displacement* slider bar or type a new value into the neighbouring box. Displacing the slice will alter the depth of molecules shown above and shown below the plane.
- The *Depth, Area* and *Displacement* values can be reset to their default values of 10.0Å, 20.0Å (the length of the side of the slice) and 0.0Å by using the **Reset** button.
- When the **Show Slice** checkbox is ticked, the *Select planes* list is automatically populated with the planes that represent the faces of the calculated BFDH morphology (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology), along with the relative area of these faces within the morphology. The largest and most morphologically important faces are listed first. The properties of the slices can be changed (depth, area and displacement) to explore the chemistry at the morphological faces.

9.5 Displaying Symmetry Related Molecules

- It is possible to display symmetry generated molecules at specified positions within the visualiser.
- To reveal the closest symmetry generated molecule to a chosen position within the visualiser click on the **Reveal symmetry-generated molecules** icon in the *Picking Mode* toolbar:

- Then, left click at the desired location within the visualiser. The closest symmetry generated molecule to that position will be revealed.
- This method can be used to display molecules at specified positions within a packing diagram without having to define contacts in order to locate them.

9.6 Displaying a Molecular Shell

- Generating a molecular shell will display all molecules within a specified distance of a selected atom or atoms. In some fields this would referred to as a coordination sphere.
- To display a molecular shell hit **Calculate** in the top-level menu, then **Molecular Shell...** in the resulting pull-down menu.

- Select an atom (or atoms) or a whole molecule (or molecules) of interest (see Ways of Selecting Atoms and Molecules).
- Specify the contact distance range. Contact distances can be specified in two ways: as absolute values in Angstroms, or as distances relative to the sum of van der Waals Radii (see van der Waals Radii).
- Hit **Apply** to display all molecules that contain an atom within the specified contact distance range of the selected atom(s), or molecule(s).

- Enable the **Show contact lines** check box to display contact lines between selected atoms and those of the molecular shell. Contact distances can be displayed by enabling the **Label distance** check box.

9.7 Displaying Graph Sets

- To display graph sets for hydrogen bond networks (see Graph Sets), select **Calculate** from the top-level menu and choose **Graph Sets...** from the resulting pull-down menu.
- The graph sets calculated for the structure will be displayed in the visualiser and in a dockable window, initially located on the right-hand side of the Mercury window. As a dockable window the *Graph Sets* window can be placed in various positions in the main Mercury window, or kept as an entirely separate window. To do this, move the mouse to the top of the *Graph Sets* window then drag the window with the mouse, keeping the left mouse button depressed, and place it where you want it by releasing the mouse button.

- The *Graph Sets* window presents information about the calculated graph sets:
  - *Descriptor*. Gives the type of graph set using a standard notation involving a code of letters and numbers, for example R4,4(12). The first, capitalised letter indicates what the type of pattern is: D indicates a discrete, finite pattern, C indicates an infinite chain, R indicates a ring and S indicates an intramolecular (or self-interacting) pattern. The next two numbers give the number of acceptors and the number of donors participating in the pattern. The number in brackets (the degree) is the total number of atoms in a path that encompasses the entire pattern. The final string of letters, for example >a>b>a>b, details the ordering of the hydrogen bonds. The letters indicate the symmetry equivalence of the hydrogen bonds (within a pattern, all hydrogen bonds labelled “a” are symmetry equivalent but are not symmetry equivalent to hydrogen bonds labelled with a “b”) and the “<“ and “>” symbols preceding the letters give an indication of the direction of the

hydrogen bond (donor to acceptor or acceptor to donor) as the pattern is traversed.

- *Level*. Indicates the number of symmetry independent hydrogen bonds that compose the pattern. The option to change the level of graph sets displayed is available via the **Options**, **Edit Options...** menu item of the *Graph Sets* window (see Customising the definitions of the Graph Sets).
- *Period*. The size of the repeat unit (number of hydrogen bonds) for ring and chain patterns.
- *# molecules*. The number of molecules that compose the pattern in its entirety.

- The graphic display of the graph sets shows the hydrogen bonds that compose the pattern, coloured and labelled by symmetry equivalence. All atoms that compose the pattern are highlighted, by default, using the ball and stick style. Options for the colour of the hydrogen bonds, the style of highlighting and the display of labels are available via the **Options**, **Highlighting...** menu item of the *Graph Sets* window (see Customising display of hydrogen bonds in Graph Sets).

9.7.1 Customising the definitions of the Graph Sets

Options to customise the calculation and display of graph sets and are available via the **Options**, **Edit Options...** pull-down menu of the *Graph Sets* window.

- *Graph set level pane*. Increase or decrease the level of the graph sets displayed by clicking on the up and down arrows. A minimum level of 1 and a maximum level of 4 is allowed. The level of a graph set is defined as the number of symmetry independent hydrogen bonds that compose the pattern.
- *Size limits pane*. Increase or decrease the number of hydrogen bonds allowed for different graph set types:
  - To alter the allowed size of Ring patterns, increase or decrease the **Maximum ring size** (maximum 10, minimum 0).
  - To alter the allowed size of Chain patterns, increase or decrease the **Maximum chain size** (maximum 6, minimum 0).

- To alter the allowed size of a discrete chain pattern, increase or decrease the **Maximum discrete size** (maximum 6, minimum 0).
- To exclude a graph set type from the calculations and the display, choose a value of 0 for that graph set type.

- Click on the **Default** button to return to the default values for graph set display.

9.7.2 Customising hydrogen bond definitions for Graph Sets

Options available to customise the calculation of hydrogen bonds used to make up graph sets are available via the **Options**, **H-bond definition...** pull-down menu of the *Graph Sets* window.

There are two tabs for customising the definition of H-bonds: one for customising the geometry definitions and one for customising H-bond atom types.

- *H-Bond Geometry* tab
  - Distance constraints can be applied to the donor to acceptor atom distances or the hydrogen to acceptor atom distances by selecting the appropriate choice from the **Constrain distance between** pull-down menu.
  - Set the required upper and lower distance limits (see Specifying Distance Limits for Nonbonded Contacts). These may be specified as absolute distances in Angstroms, or as distances relative to the sum of vdW radii of the atoms involved in the H-bond.
  - The D-H...A angle that defines a hydrogen bond can be changed in the *Define hydrogen bond angle* pane
  - Set the minimum number of covalent bonds between the contact atoms (see Intramolecular Nonbonded Contacts). The bond count always refers to the number of bonds between the *D* and *A* atoms irrespective of whether contacts are defined as D....A or H....A.
- *H-Bond Atom Types* tab
  - Select the required combination of donor and acceptor atoms via the tick-boxes provided in the scrolling windows labelled *Donor atom types* and *Acceptor atom types*. It is difficult to guarantee correct classification of donors and acceptors if one or more H atoms are missing from the structure (see Hydrogen Atoms).
- Click on **Default** followed by **Apply** to return the H-bond definitions to their default values.

9.7.3 Customising display of hydrogen bonds in Graph Sets

The options to highlight graph set hydrogen bonds are available via the **Options**, **Highlighting...** pull-down menu of the *Graph Sets* window.

- *H-bond colour*: Choose to colour *By symmetry equivalence* (all symmetry equivalent hydrogen bonds will be coloured the same) or pick a colour to be applied to all hydrogen bonds.
- *Style*: determines how the atoms that compose the entire graph set are displayed. To display the graph set without highlighting the constituent atoms, select the option to *Use current style*.

- To toggle the display of hydrogen bond labels (a,b,c, etc.) click on the *Show H-bond labels* tickbox.

9.7.4 Saving Graph Sets

To save a list of the calculated graph sets for a structure, click on the **Options**, **Save...** pull-down menu of the *Graph Sets* window.

9.8 Displaying Voids

- The voids function allows any empty spaces (voids) in crystal structures that are big enough to contain a spherical "probe" of a given radius to be found.
- Click on **Display** in the top-level menu and select **Voids....** from the resulting menu. This will open the *Voids* dialogue.

- Use the **Show voids** tick box to toggle the display of voids.
- The size of the **Probe Radius** and the **Approximate Grid spacing** can be adjusted. A smaller value for the probe radius will identify smaller spaces and generate larger voids. A smaller value for the grid spacing will create smoother looking surfaces.
- Voids can be calculated using two different methods:
  - *Calculate voids using Solvent Accessible Surface* gives the volume which can be occupied by the centre of a probe of a given radius.
  - *Calculate voids using Contact Surface* maps the volume that can be occupied by the full probe (including its radius) and thus gives an estimate of the volume that could be filled by solvent or guest molecules.

- A full description of these two different surfaces and the ways in which they can be used is given in: Leonard J. Barbour, *Chem. Commun*., **2006**, 1163-1168.

- The display of the void regions in the crystal can be customised using the following options:
  - **Colour**: click on the coloured panel and select the colour required (see Using the Select Color Dialogue Box).
  - **Transparency**: the transparency value controls the transparency/opaqueness of the surface of the void. The scale used is between 0 and 1. Å value of 0 will result in an opaque surface.
- The void volume is given both in $Å^3$ and as a percentage of the unit cell volume.
- The **Apply** button should be clicked for changes to be made to the display. Changes made to these settings are not saved between Mercury sessions.

9.9 Displaying Space-Group Symmetry Elements

- To display space-group symmetry elements hit **Display** in the top-level menu, then **Symmetry Elements...** in the resulting pull-down menu.
- The symmetry elements present in one unit cell, as shown in the International Tables, calculated from the symmetry operators, will be displayed:

- Inversion centres (centres of symmetry) are displayed as small spheres, rotational axes as lines and mirror planes and glide planes as planes. In the example shown (CSD entry AACRUB01), the space group is *C2/c* and both molecules in the asymmetric unit are on special positions: the ruthenium complex sits on a centre of symmetry, whereas the tetrafluoroborate counter-ion sits on a two-fold rotation axis.
- The display can be customised via the following dialogue:

- The following options are available:
  - *Inversions*: allows the display of inversion centres to be switched on or off. The size of the spheres representing inversion centres can also be altered with respect to the unit cell.
  - *Axes*: allows combinations of rotation axes to be switched on or off, and allows control over the colours of the axes. By default, screw axes are drawn with "arrows" to allow them to be distinguished from the other axes; this can be switched off.
  - *Glide & mirror planes*: the presence of planes can make the display crowded, and the display of planes can be switched off. By default, the planes are semi-transparent and shadows are not included: setting the transparency to 0.0 (i.e. opaque) in combination with shadows gives the planes a more 3 dimensional look.
  - For each set of symmetry elements there is a **Show** tickbox. These tickboxes can be used to toggle the display of groups of symmetry elements. Unchecking all the tickboxes and hitting **Apply** will remove the display of symmetry elements for the currently loaded structure.
- *Note:* The **Apply** button will need to be clicked on for changes to be made to the display. Changes made to these settings are not saved between Mercury sessions.
- Symmetry elements will be displayed for all structures while the *Symmetry elements* window is open. To exit from symmetry elements display, click either **OK** or **Cancel** in the *Symmetry elements* window. Symmetry elements will continue to be displayed for the present crystal structure until a new file is loaded, or an alternative structure is loaded.
- For glide planes, it is not possible to display the direction of the glide. This direction can be obtained from the *Symmetry Operators List...* tab, accessible via the **More Info** button by adding the *Detailed Description* field via the **Customise...** button.

- *Note*: The *Symmetry Operators List...* tab lists the symmetry operators, whereas the symmetry elements display displays the symmetry *elements*: the identity, for example, is a symmetry operator, but no associated symmetry element can be displayed. Likewise, a single inversion centre is listed among the symmetry operators, but when combined with the lattice periodicity, this corresponds to eight symmetry elements per unit cell (27 are displayed, but those residing on corners, edges and planes are shared).
- Different types of axes can coincide, in which case only one of the coinciding axes is drawn. A trivial example is a six-fold rotation axis, which implies the presence of a three-fold and a two-fold rotation axis; indeed, when space group *P6* (no. 168) is displayed and the display of six-fold axes is switched off, three-fold axes are drawn where the six-fold axes used to be. Switching off the three-fold axes as well only leaves the two-fold axes. A less trivial example is space group $P4_2/m$ (no. 84), where the 4_2 screw axis coincides with a four-fold rotoinversion axis: only the rotoinversion axis is shown. Switching off the display of rotoinversion axes makes the 4_2 screw axis appear; switching off the display of screw axes as well reveals the presence of a two-fold axis.
- The display of symmetry elements is highly educational and can for example be used to illustrate the presence of rotoinversion axes in combination with true inversion centres as is the case in space group $P4_2/m$ (no. 84), or species occupying two different special positions.
- In order to study molecules on special positions, we suggest combining the display of symmetry elements with examples taken from some of the higher symmetry space groups or point groups found in CSDSymmetry, see: *http://www.ccdc.cam.ac.uk/free_services/csdsymmetry/*.
  Special positions can only be generated by symmetry elements that do not have an intrinsic translational part: a screw axis, for example, cannot generate a symmetry image of a molecule in such a way that it overlaps with the original molecule because of the translation that is involved (polymers and catena compounds are an exception to this rule). There are only two kinds of symmetry elements that have an intrinsic translation: screw axes and glide planes. We therefore suggest switching off screw axes and glide planes when examining special positions.
- Symmetry-imposed disorder is a special case of a molecule on a special position, namely one where the point group of the molecule is incommensurate with the local point group in the crystal structure, and we therefore suggest switching off screw axes and glide planes when examining symmetry-imposed disorder as well.

10 Finding Hydrogen Bonds and other Nonbonded Contacts

10.1 Overview of Nonbonded Contact Functionality

Mercury will find and display the following types of intermolecular and/or intramolecular nonbonded contacts:

- Hydrogen bonds.
- Short nonbonded contacts, i.e. all contacts between atoms of any type that are shorter than a specified van der Waals corrected distance (see Specifying Distance Limits for Nonbonded Contacts).
- Specific nonbonded contacts (i.e. contacts in a specific distance range between atoms of particular types, e.g. *chlorine...oxygen contacts in the range 2-4*Å).

10.2 Basic Steps in Defining and Displaying Nonbonded Contacts

- Define the types of contacts in which you are interested, e.g. *all H-bonds*, or *H-bonds between amide NH donors and carboxylate oxygens*, or *Cl...Br contacts shorter than the sum of the vdW radii of Cl and Br* (see Default Nonbonded Contact Definitions).
- Instruct Mercury to find and display contacts of the types you have defined (see Switching the Display of Nonbonded Contacts On and Off).
- Switch contacts on and off as desired, e.g. you could switch off one type of contact but continue to display the others (see Switching the Display of Nonbonded Contacts On and Off).
- If desired, build a network of molecules connected by the nonbonded contacts you have found (see Building Networks of Nonbonded Contacts).

10.3 Default Nonbonded Contact Definitions

Two types of nonbonded contacts are defined by default when Mercury is opened. Specifically, there is:

- a default definition of hydrogen bonds (see Default Definition of Hydrogen Bonds);
- a default definition of short nonbonded contacts (see Default Definition of Short Nonbonded Contacts).

In addition, you may add your own, customised definitions of nonbonded contacts (see Defining New Types of Nonbonded Contacts).

10.3.1 Default Definition of Hydrogen Bonds

- The default definition of a hydrogen bond (*D-H...A*) is a contact meeting the following criteria:
    - The donor (*D*) must be a nitrogen, oxygen or sulphur atom covalently bound to at least one hydrogen. It is not necessary for the 3D coordinates of the H atom to be known, i.e. H-bonds will be found even if H atoms have not been located in the crystal structure (see Hydrogen Atoms).

- The acceptor (*A*) must be a nitrogen, oxygen, sulphur or halogen with at least one available lone pair (e.g. pyramidal trigonal nitrogen is regarded as an acceptor but planar trigonal nitrogen is not).
- The *D...A* distance must be less than the sum of van der Waals Radii (see van der Waals Radii) of the *D* and *A* atoms.
- The contact may be (a) intermolecular, or (b) intramolecular involving donor and acceptor atoms separated by at least 3 covalent bonds within the molecule.

- This pre-set default H-bond definition is always associated with an item (viz. *H-Bond Default definition*) in the list box underneath the display area in the main Mercury window:

- The default definition can be edited by double-clicking on it and changing options as required in the resulting *Define H-bonds* dialogue box (see Defining a New Type of Hydrogen Bond; The Define H-bonds Dialogue Box).
- Hydrogen bonds satisfying the default definition can be displayed in the display area by switching on the tick box next to the words *H-Bond Default definition*:

10.3.2 Default Definition of Short Nonbonded Contacts

- The default definition of a short contact in Mercury is any intermolecular contact shorter than the sum of the van der Waals Radii (see van der Waals Radii) of the atoms involved.
- This pre-set default definition is always associated with the entry that reads: *Short Contact <(sum of vdW radii)* in the list box underneath the display area in the main Mercury window:

- The default definition can be edited by double-clicking on it and setting options as required in the resulting *Define Short Contacts* dialogue box (see Adding a New Short-Contact Definition; The Define Short Contacts Dialogue Box).
- Short contacts satisfying the default definition can be displayed in the display area by switching on the tick box next to the words *Short Contact <(sum of vdW radii)*:

10.4 Defining New Types of Nonbonded Contacts

Mercury allows you to add your own definitions of (or edit existing definitions of):

- H-bonds (see Defining a New Type of Hydrogen Bond; The Define H-bonds Dialogue Box).
- Short contacts (see Adding a New Short-Contact Definition; The Define Short Contacts Dialogue Box).
- Specific contacts (see Defining a New Type of Specific Contact; The Define Specific Contacts Dialogue Box).

In each case, you will need to:

- Specify distance limits (see Specifying Distance Limits for Nonbonded Contacts).
- Specify whether the contact is allowed to be intermolecular and/or intramolecular (see Intramolecular Nonbonded Contacts).

10.4.1 Specifying Distance Limits for Nonbonded Contacts

Distance limits for nonbonded contacts can be defined in two ways:

- As absolute values in Angstroms, e.g. $2 < d(X...Y) < 5$Å.
- As distances relative to the sum of van der Waals Radii (see van der Waals Radii), i.e.

  $v(X) + v(Y) - p < d(X...Y) < v(X) + v(Y) - q$

  where $v(X)$ and $v(Y)$ are the vdW radii of X and Y respectively and p and q are user-specified values in Angstrom (e.g. if $p = 0.5$ and $q = 0.1$, the contact must be at least 0.1Å shorter than the sum of vdW radii but no more than 0.5Å shorter).

10.4.2 Intramolecular Nonbonded Contacts

- Whenever you define a nonbonded contact, you must specify whether the contact is allowed to be intermolecular, intramolecular or either.
- For contacts that are allowed to be intramolecular, you must specify the minimum number of covalent bonds that must separate the two contact atoms.
- For example, with the setting below for a Cl...Br contact definition:

the Cl1 and Br1 contact, separated by 4 bonds, would be allowed but the Br2 and Cl1 atoms, separated by 3 bonds, would not:

10.4.3 Defining a New Type of Hydrogen Bond; The Define H-bonds Dialogue Box

A new type of *hydrogen bond* may be defined as follows:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select the top-level **Calculate** option followed by **Contacts....** in the resulting pull-down menu, to bring up the *Contacts* dialogue box. Then hit **Add** in the *Contacts* dialogue box and pick **H-bond...** to bring up the *Define H-bonds* dialogue box.
- Alternatively, you can get straight to the *Define H-bonds* dialogue by right-clicking anywhere in the white list box under the main Mercury display area and selecting **Add H-bond...** from the resulting menu.
- Specify whether it is necessary for the position of the hydrogen atom involved in the H-bond to be known (i.e. for the H atom to have been located in the crystal-structure analysis).
- If the *Require hydrogen atom to be present* button is set on, then a minimum value for the *D-H...A* angle must be specified (in degrees).
- Select the required combination of donor and acceptor atoms via the tick-boxes provided in the scrolling windows labelled *Donor atom types* and *Acceptor atom types*. It is difficult to guarantee correct classification of donors and acceptors if one or more H atoms are missing from the structure (see Hydrogen Atoms). C-H is available as a donor group but is not classified as such by default.
- Set the required upper and lower distance limits (see Specifying Distance Limits for Nonbonded Contacts). These may be specified as absolute distances in Angstroms, or as distances relative to the sum of vdW radii of the atoms involved in the H-bond.

If the *Require hydrogen atom to be present* button is on, the distance constraint refers to the *H...A* distance; otherwise it refers to the *D...A* distance.

- Specify whether the contacts are to be intermolecular only, intramolecular only, or whether both are allowed.
- If intramolecular bonds are allowed, specify the minimum number of covalent bonds between the contact atoms (see Intramolecular Nonbonded Contacts). Even if the *Require hydrogen atom to be present* button is on, the bond count refers to the number of bonds between the *D* and *A* atoms.
- The following settings will find intermolecular and intramolecular H-bonds between all types of donors and 6-membered ring aromatic nitrogen acceptors, subject to the constraints that (*a*) the H atom must have been located in the crystal structure analysis, (*b*) the donor-H...nitrogen angle must be greater than or equal to 150 degrees, (*c*) the hydrogen...nitrogen distance must be at least 0.1Å shorter than the sum of vdW radii of H and N, and (*d*) if the contact is intramolecular, the donor and N atoms must be separated by more than 4 covalent bonds:

- Hit **Default** followed by **Apply** if you want to return all the settings to their default values.

10.4.4 Adding a New Short-Contact Definition; The Define Short Contacts Dialogue Box

A new definition of *short contact* can be created as follows:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select the top-level **Calculate** option followed by **Contacts....** in the resulting pull-down menu, to bring up the *Contacts* dialogue box. Then hit **Add** in the *Contacts* dialogue box and pick **Short...** to bring up the *Define Short Contacts* dialogue box.
- Alternatively, you can get straight to the *Define Short Contacts* dialogue by right-clicking anywhere in the white list box under the main Mercury display area and selecting **Add Short...** from the resulting menu.
- Set the required upper distance limit - this must be a distance expressed relative to the sum of the vdW radii of the atoms (see Specifying Distance Limits for Nonbonded Contacts).
- Specify whether the contacts are to be intermolecular only, intramolecular only, or whether both are allowed.
- If intramolecular bonds are allowed, specify the minimum number of covalent bonds between the contact atoms (see Intramolecular Nonbonded Contacts).
- You cannot specify the element types of the interacting atoms; any element type will be allowed. Create a *specific contact* definition if you are only interested in contacts between particular types of atoms (see Defining a New Type of Specific Contact; The Define Specific Contacts Dialogue Box).
- The following settings will find intermolecular and intramolecular contacts between any pair of atoms, subject to the constraints that (*a*) the contact must be at least 0.3Å shorter than the sum of vdW radii, and (*b*) if the contact is intramolecular, the atoms must be separated by at least 4 covalent bonds:

- Hit **Default** followed by **Apply** if you want to return all the settings to their default values.

10.4.5 Defining a New Type of Specific Contact; The Define Specific Contacts Dialogue Box

A new type of *specific contact* may be defined as follows:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select the top-level **Calculate** option followed by **Contacts....** in the resulting pull-down menu, to bring up the *Contacts* dialogue box. Then hit **Add** in the *Contacts* dialogue box and pick **Specific...** to bring up the *Define Specific Contacts* dialogue box.

- Alternatively, you can get straight to the *Define Specific Contacts* dialogue by right-clicking anywhere in the white list box under the main Mercury display area and selecting **Add Specific...** from the resulting menu.
- Use the pull-down menus next to the entry boxes labelled *Contacts between* to specify the atom types of the contacting atoms. If "Other Elements" is chosen, a periodic table will be shown from which it is possible to select one or more elements, or groups of elements.
- Set the required upper and lower distance limits (see Specifying Distance Limits for Nonbonded Contacts). These may be specified as absolute distances in Angstroms, or as distances relative to the sum of vdW radii of the atoms involved in the contact.
- Specify whether the contacts are to be intermolecular only, intramolecular only, or whether both are allowed.
- If intramolecular bonds are allowed, specify the minimum number of covalent bonds between the contact atoms (see Intramolecular Nonbonded Contacts).
- The following settings will find intermolecular and intramolecular contacts between chlorine and oxygen atoms, subject to the constraints that (*a*) the contact must be at least 0.1Å shorter than the sum of vdW radii, and (*b*) if the contact is intramolecular, the atoms must be separated by at least 5 covalent bonds:

- Hit **Default** followed by **Apply** if you want to return all the settings to their default values.

10.5 Colouring Contacts Based on Distance

- Contacts can be coloured based on distance via **Display**, **Colours**, **Contacts**. When you select this option, the contacts displayed will be coloured on a smooth scale between three settings (short, mid, long) (see Setting Display Properties of Contacts).

10.6 Editing and Copying Nonbonded Contact Definitions

Any nonbonded contact definition (whether it be an *H-bond, short contact* or *specific contact*) can be edited by any of the following methods:

- Double click on the appropriate contact definition in the list box just below the display area, or right-click and pick **Edit...** This will bring up the appropriate dialogue box containing the settings for that contact definition. These can then be edited (see Defining New Types of Nonbonded Contacts).
- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select the top-level **Display** option followed by **Contacts....** in the resulting pull-down menu, to bring up the *Contacts* dialogue box. There are two options:
  - Direct editing: to change the definition of an existing contact, click on the appropriate entry in the list box and then select **Edit...** This will bring up either the *Define H-bonds* (see Defining a New Type of Hydrogen Bond; The Define H-bonds Dialogue Box), the *Define Short Contacts* (see Adding a New Short-Contact Definition; The Define Short Contacts Dialogue Box), or the *Define Specific Contacts* (see Defining a New Type of Specific Contact; The Define Specific Contacts Dialogue Box) dialogue box, depending on the type of contact selected. You may now edit the definitions.
  - Editing a copy: if you wish to preserve an existing definition, but use it as the basis for creating a new definition, then you may make a copy for editing. Right-click on the appropriate entry in the list box, select **Copy**, and a new entry will appear in the list. Select the newly copied entry in the list and hit **Edit...**

10.7 Deleting Nonbonded Contact Definitions

Any nonbonded contact definition, except the two start-up defaults, can be deleted as follows:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select the top-level **Calculate** option followed by **Contacts....** in the resulting pull-down menu, to bring up the *Contacts* dialogue box.
- Right-click on the appropriate entry in the list box and then select **Delete.**

10.8 Viewing Nonbonded Contact Definitions

Mercury builds up a list of H-bond and other nonbonded contact definitions as they are set. This can be viewed:

- In the scrolling window just below the display area, e.g.

- In the dialogue box obtained by hitting **Calculate** in the top-level menu and **Contacts...** in the resulting pull-down menu. Details of each contact definition can be viewed by double clicking on the appropriate item in the list.

10.9 Switching the Display of Nonbonded Contacts On and Off

Before you can display nonbonded contacts, you must define the types of contacts in which you are interested (see Default Nonbonded Contact Definitions). Once this is done, the contacts you have defined can be displayed, undisplayed or deleted completely, as follows:

- Contacts of a particular type (i.e. corresponding to a particular nonbonded-contact definition) can be displayed by switching on the tick box to the left of the corresponding contact definition in the *Contacts List* (the list box underneath the display area, or in the *Contacts* dialogue box).
- For example, the following setting will display all H-bonds (default definition), Cl...O contacts (a user-defined type of contact) and will not display short contacts (other than those that are displayed by virtue of being H-bonds or Cl...O contacts):

- A particular type of contact can be undisplayed by switching its tick box off.
- Individual contacts can be deleted from the display by right-clicking on the contact and selecting **Delete Contact** from the pull-down menu.
- All contacts from an atom can be deleted by right-clicking on the atom and selecting **Contacts** followed by **Delete Contacts from this Atom**.
- All contacts can be switched off by using the **Reset** button at the bottom-left of the main Mercury window.

10.10 Finding Out What Symmetry Operations Were Used to Generate Contacts

- Click on the **More Info** button (centre bottom of main Mercury window) and select **Contacts List** from the resulting menu.
- A spreadsheet will be displayed which lists all the contacts on view for the current structure. For each contact, the spreadsheet shows the symmetry operations used to generate the two atoms involved in the contact from their symmetry-equivalent atoms in the crystallographic asymmetric unit. This information is in the columns headed **Symm. op. 1** and **Symm. op. 2**.

10.11 Listing a Summary of the Contacts on View

- Click on the **More Info** button (centre bottom of main Mercury window) and select **Contacts List** from the resulting menu.
- A spreadsheet will be displayed which lists all the contacts on view for the current structure. Use the **Customise** button to decide what columns you want shown, then save the spreadsheet out to a file (see Spreadsheet Display of Contacts and Their Properties).

11 Building Networks of Nonbonded Contacts

11.1 The Purpose of Building a Network

- A network is an assembly of molecules in a crystal structure connected together by one or more types of nonbonded contacts. For example, here is a network of molecules connected by *OH...O* and *NH...O* hydrogen bonds:

- Building a network enables you to understand the key intermolecular interactions that lead to the observed crystal structure. These interactions will often be hydrogen bonds, but may be other electrostatic interactions, hydrophobic contacts, etc.

11.2 Basic Steps in Building a Network

- Begin by displaying just the molecule(s) that constitute the Crystal Chemical Unit (see Crystal Chemical Unit), i.e. the molecule(s) that will be shown by default when a structure is first viewed with packing switched off, e.g.

- Define the types of nonbonded contacts in which you are interested, e.g. *all hydrogen bonds*, or *OH...O hydrogen bonds*, or *all contacts shorter than the sum of van der Waals radii (see* Default Nonbonded Contact Definitions).
- Find the contacts (of the type(s) you have defined) formed by the displayed molecules to neighbouring molecules in the crystal structure (see Switching the Display of Nonbonded Contacts On and Off). This should create a display such as:

- *Expand* the *hanging contacts (see* Colouring Conventions for Contacts), i.e. show in full the molecule at the end of each contact (see Methods for Network Expansion). This should produce a display such as:

- Continue expanding selected hanging contacts until the pattern of interactions in the crystal structure becomes clear. It may help to view the network down the three crystallographic axes in turn (use the **a, b** and **c** buttons in the tool bar near the top of the Mercury window).
- If the network grows too large, remove unwanted molecules (see Deleting Molecules from a Network) or contacts (see Deleting Contacts from a Network) to get a clearer picture.
- The final network may still contain a number of contacts that hang from its periphery; these can be deleted to tidy the display (see Deleting Contacts from a Network).

11.3 Colouring Conventions for Contacts

- Nonbonded contacts are shown in either cyan (light blue) or red, e.g.

- Contacts shown in red are *hanging contacts*. The molecule at one end of a hanging contact is shown in full. Only one atom (the contacting atom, or *hanging atom*) is shown of the molecule at the other end.
- The red dotted line terminates in a cross in the display, whose colour indicates the element type of the contacting atom from the other molecule, provided you are colouring by element (see Standard Element Colours). Atom labels for these hanging atoms can be displayed.
- Contacts shown in cyan are not hanging contacts. Both molecules at the end of a cyan contact are shown in full.

11.4 Setting Display Properties of Contacts

- To set the colour of contacts, hit the top-level **Display** button followed by **Colours** in the resulting pull-down menu, then **Contacts...** in the next menu. This will bring up the *Colour Contacts* dialogue box:

- Select **By expanded/hanging** to display contacts according to the default colouring conventions (see Colouring Conventions for Contacts).
- Select **Colour** in order to choose your own. You should then click on the coloured panel to choose the colour you want; this will open the Select Color dialogue box (see Using the Select Color Dialogue Box).
- Select **Colour by distance** to colour contacts according to whether they are short, mid, or long range contacts. Distances are based on the van-der-Waals corrected distances of the contacts atoms within the range of (the sum of atomic van der Waals radius -1 Angstrom) to (the sum of atomic van der Waals radius). You should then click on the coloured panels and choose the colour you want for each contact distance. Clicking ona panel will open the Select Color dialogue box (see Using the Select Color Dialogue Box).
- The selected colour can be applied to **All Contacts**, **Expanded Contacts** or **Hanging Contacts** as required.
- To set the display style of contacts, hit the top-level **Display** button followed by **Styles** in the resulting pull-down menu, then **Contacts...** in the next menu. This will bring up the *Contact Style* dialogue box:

- Select either **Wireframe** or **Stick** representation. The chosen display style can then be applied to **All Contacts**, **Expanded Contacts** or **Hanging Contacts** as required.

- To customise the display of an individual contact, right-click on the contact, select the appropriate option from the pull-down menu (**Style, Colour**) and then select the desired display-property setting from the next menu. The chosen setting will be applied only to the contact on which you clicked.
- It is possible to create and modify sets of Mercury display styles, which are automatically saved between sessions (see Managing Display Styles). These styles can be used to modify the colour scheme used for hanging and expanded contacts.

11.5 What Network Expansion Is

- The starting point for a network expansion is one or more molecules with hanging contacts (see Colouring Conventions for Contacts), e.g.

- This starting point is reached by:
  - defining the types of nonbonded contacts in which you are interested (see Default Nonbonded Contact Definitions), e.g. *OH...O hydrogen bonds*;
  - finding the contacts of these types that are formed by the molecule(s) in the crystal chemical unit (see Switching the Display of Nonbonded Contacts On and Off).

- Network expansion involves changing one or more hanging contacts into non-hanging contacts, i.e. showing in full the molecules at the ends of hanging contacts. Network expansion therefore involves adding molecules to the outside of the current assembly.

11.6 Methods for Network Expansion

Molecular networks can be expanded in several ways:

- Left-clicking on a hanging contact or hanging atom.
- Using the *Build Network* dialogue box.
- Right-clicking on a hanging contact.
- Right-clicking on atoms or hanging atoms.
- Right-clicking in the display-area background.

Left-clicking on a hanging contact or hanging atom (see Colouring Conventions for Contacts):

- Ensure that *Picking Mode* (near the top of the main Mercury window) is set to **Expand Contacts**:

- Left-click on a hanging contact (or the atom at the end of the contact) to expand the contact.

Using the *Build Network* dialogue box:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select **Calculate** from the top-level menu followed by **Contacts....** from the resulting pull-down menu. Then select the **Build Network** tab in the *Contacts* dialogue box. Then:
  - **Hit Expand All** to expand all hanging contacts.
  - **Hit Expand Picked** and then left-click on selected hanging contacts to expand them.

Right-clicking on a hanging contact:

- **Hit Expand All** to expand all hanging contacts.
- **Hit Expand Contact** to expand the contact you clicked on.

Right-clicking on atoms or hanging atoms:

- Choose **Contacts from the pull-down menu. Then hit:**
- **Expand All** to expand all hanging contacts.
- **Expand Contacts from this Atom** to expand all contacts involving the picked atom.
- **Expand Contacts from this Molecule** to expand all contacts from the molecule containing the picked atom, which need not itself be involved in any contacts.

Right-clicking in the display-area background:

- Choose **Contacts**, then hit **Expand All** to expand all hanging contacts.

11.7 Deleting Contacts from a Network

Contacts can be deleted from a network in several ways:

- Using the Build Network dialogue box.
- Right-clicking on a contact.
- Right-clicking on atoms or hanging atoms.
- Right-clicking in the display-area background.

Using the *Build Network* dialogue box:

- Hit the **Contacts...** button (located near to the bottom-centre of the main Mercury window), or select **Calculate** from the top-level menu followed by **Contacts....** from the resulting pull-down menu. Then select the **Build Network** tab in the *Contacts* dialogue box. Then:
    - **Hit Delete Hanging** to delete all hanging contacts (see Colouring Conventions for Contacts).
    - **Hit Delete Picked** and then left-click on individual contacts to delete them.

Right-clicking on a contact:

- Hit **Delete Hanging Contacts** to delete all hanging contacts.
- Hit **Delete Contact** to delete the contact you clicked on.

Right-clicking on atoms or hanging atoms:

- Choose **Contacts from the pull-down menu. Then hit:**
    - **Delete Hanging Contacts** to delete all hanging contacts.
    - **Delete Contacts from this Atom** to delete all contacts involving the picked atom.
    - **Delete Contacts from this Molecule** to delete all contacts from the molecule containing the picked atom, which need not itself be involved in any contacts.

Right-clicking in the display-area background:

- Choose **Contacts**, then hit **Delete Hanging Contacts** to delete all hanging contacts.

11.8 Deleting Molecules from a Network

To delete a molecule, right-click on any atom or bond in the molecule and select **Delete this Molecule** from the pull-down menu.

11.9 Resetting a Network to Its Starting Point

There are several ways to reset a network to its starting point:

- Right-click on a contact and select **Contacts** followed by **Reset Contacts** from the pull-down menus.
- Right-click on an atom or in the display-area background, select **Contacts** from the pull-down menu and then hit **Reset Contacts**.

- Hit **Clear** in the *Build Networks* dialogue box (obtainable, e.g., by selecting **Calculate** in the top-level menu, **Contacts...** in the next menu and then selecting the **Build Networks** tab).
- Hit the **Reset** button at the bottom left-hand corner of the main Mercury window (this will remove *all* contacts; the previous methods will leave hanging contacts from the starting molecules).

12 Editing a Structure

When editing a structure, changes are applied to the underlying crystallographic data. Consequently, when multiple copies of a molecule are displayed (e.g. when packing is turned on) edits will be automatically applied to all symmetry-generated molecules.

12.1 Manually Editing a Structure

- To manually edit a structure, hit the **Edit** button in the top-level menu and select **Edit Structure...** from the resulting pull-down menu. This will launch the *Edit Structure* dialogue:

- Within the *Manual Edit* pane of this window you can choose to:
  - Add atoms: click on the **Atoms** button, select the element type and hybridisation required from the pull-down menus, then click on the atom to which the new atom is to be added.
  - Add hydrogen atoms: click on the **Hydrogen Atoms** button, select the number of required hydrogen atoms, then click on the atom(s) to which hydrogens are to be added.

- Add bonds: click on the **Bonds** button, choose the bond type from the pull-down menu, then click on the two atoms that you wish to be bonded.
- Edit an element type: click on the **Set Element Type to** button, select the element type required from the pull-down menu, then click on the atom whose element type you wish to change. Enable the **Update Labels Automatically** check-box to change the atom label to that chosen.
- Edit an atom label: click on the **Set Atom Label...** button, click on the atom whose label you wish to change in the display, then enter the new label in the resulting *Edit atom* dialogue.
- Edit a formal charge: click on the **Set Atom Charge...** button, select the formal charge you wish to apply using the dial box, then click on the atom(s) you wish to modify in the visualiser.
- Edit a bond type: click on the **Set Bond Type to** button, choose the bond type from the pull-down menu, then pick the bond(s) you wish to change.
- Remove atoms and bonds: click on the **Atoms & Bonds** button, then pick on the atoms or bonds you wish to remove.
- Remove all hydrogens: click on the **All Hydrogens** button, then click on an atom or bond in the structure from which you wish to remove all hydrogen atoms.
- Remove complete molecules: click on the **Molecules** button, then pick an atom or bond from within the molecule you wish to remove. It is also possible to only remove a particular copy of a molecule (see Deleting Molecules from a Network).
- Remove all formal charges: click on the **All Formal Charges** button, then pick an atom or bond within the particular molecule you would like to remove all formal charges from. Formal charges on all atoms of that molecule will be reset to zero.
- Remove selected atoms: make a selection of any number of atoms within the visualiser and then click on the **Selected Atoms** button, all of the atoms that were selected will be removed from the structure (including symmetry copies of those atoms if they are present in the visualiser).

- Molecules can be broken apart by the deletion of bonds.
- New bonds may join together two separate molecules in a single structure. However, you cannot add a bond between two different structures (when in multiple structure mode). Nor can you join together symmetry related molecules, for example if packing is turned on.
- When reading structures from the Cambridge Structural Database, all **Suppressed Atoms** (see **Suppressed Atoms**) are treated as being in a single molecule. Using remove **Molecules** therefore provides an easy way to remove all suppressed atoms from within a structure.

12.2 Auto-Editing Options

- There are a number of options available for editing a structure automatically; to access these click the **Edit** button in the top-level menu and select **Auto Edit...** from the resulting drop-down menu. This will launch the *Auto Edit Structure* dialogue:

- The following options are available:
  - Identify polymeric bonds: ensure the **Identify polymeric bonds** check box is checked. This will analyse if atoms are part of a repeating network and enable use of the Polymer Expansion menu. Note that this may result in disordered structures with overlapping atomic positions being treated as a polymeric.
  - Guess bond types: ensure the **Guess bond types** check box is checked, then apply this to either **All** or **Only bonds with unknown types**.
  - Standardise bond types: ensure the **Standardise to Cambridge Structural Database conventions** check box is checked. This standardisation can be optionally applied to **Aromatic bonds**, **Delocalised bonds** or both**.**
  - Add missing hydrogen atoms: ensure the **Add missing H atoms** check box is checked. The number of hydrogens added to each atom will be sufficient to satisfy the atom's unfilled valencies.

- Click the **Apply** button to automatically edit all loaded structures according to the selections made. The **Close** button will exit the *Edit Structure* dialogue without applying edits to the structure; if any options have been changed (e.g. a checkbox has been unchecked) an additional *Changed settings* dialogue will appear asking if you wish to apply the changed settings.

12.3 Normalising Hydrogen Atom Positions

- Hydrogen-atom positions determined by X-ray diffraction can be unreliable because hydrogen has a low scattering power for X-rays. In addition, the lengths of covalent bonds involving hydrogen are almost invariably too short when measured by X-ray diffraction, because the centroid of the hydrogen-atom electron density is displaced towards the heavy atom to which the hydrogen is bonded.
- These errors can be corrected by normalising the hydrogen atom positions. This involves moving the hydrogen atom along the covalent bond vector (X-H) so as to make the X-H distance equal to the average neutron diffraction value (hydrogen atoms are accurately located by neutron diffraction).
- Select **Edit** from the top-level menu, then **Normalise Hydrogens...** from the resulting menu. this will open the *Normalise Hydrogens* dialogue:

- By default, normalisation of hydrogen atom positions is available for C-H, O-H and N-H bonds (the H atom will be moved along the bond vector to make the bond length equal to the average neutron diffraction value). The average C-H, O-H and N-H bond lengths are displayed under *Current Settings*.
- Hit **Apply** to normalise all terminal hydrogen atoms positions to the bond length values displayed under *Current Settings*.
- It is possible to customise the values for C-H, N-H and O-H H-normalisation. Enter the element symbol and element-terminal hydrogen bond length and click on the **Store Value** button. To reset the values for C-H, N-H and O-H H-normalisation hit the **Default** button.
- It is also possible to specify values to normalise the position of H atoms bonded to other elements.

12.4 Polymer Expansion

- Polymeric structures in the CSD can be identified as the polymeric links will be differentiated with the polymeric bond style of long and short dashed sections. To expand polymeric structures, select **Edit** from the top-level menu, then **Polymer Expansion...** from the resulting menu. This will open the *Polymer Expansion* dialogue:

- The *Selected atoms and bonds* section of the menu gives the available options if an atom or bond is selected. To expand a particular polymeric bond, left-click on the bond you wish to expand in the display. Alternatively, click on a polymeric atom to expand all its polymeric bonds. If an area is selected using the lasso picking mode, polymer expansion will occur on all selected atoms and bonds.
  - **Expand at selection** will expand only the selected atom(s) or bond(s) by either the Whole or Sub unit as described below
  - **Expand symmetry related** will expand the selected atom(s) or bond(s) and all other symmetry-equivalents. This also enables multiple expansions of the polymer by repeated clicking of the same atom or bond.
  - **Prune at selection** will remove any previous-expanded section of the polymer at this point. Note that the initial crystal chemical unit cannot be deleted by this method.
- The *Expand by* section of the menu gives two options for how the polymer will be expanded when suitable atoms or bonds are selected.
  - **Whole unit** expansion will add an additional repeat of the initial crystal chemical unit to the structure. All polymeric structures in the CSD will include a full coordination sphere of all symmetry-independent atoms of the structure.
  - **Sub unit** expansion adds all atoms and bonds up to the next polymeric atom or bond in the structure.
- Any hanging polymeric bonds in the structure(s) currently displayed may optionally be labelled “Po” for increased visibility of the polymeric bonds.

- To expand all polymeric bonds in the structure, click on the **Expand All** button.
- To undo a step of polymer expansion, click **Edit** in the top-level menu and choose **Undo Polymeric expansion** or press the *Ctrl* and Z keys.
- Click **reset** to return to the network to its starting point, this will delete any expanded parts of the structure.
- Closing the Polymer Expansion dialogue will remove all “Po” labels.
- The Polymer Expansion dialogue will be inactive (with all options greyed out) if the selected structure is not polymeric or the Asymmetric Unit is selected in the Display Options window. The Auto Edit Structure dialogue can be used to identify polymeric bonds in cases where the bond types of a structure have not been determined e.g. refcode ACOCUS.

12.5 Transforming, Translating or Inverting Molecules

- To transform molecules in a structure, for example to apply symmetry operators of the space group to a molecule or to translate or invert a structure, select **Edit** from the top-level menu, then **Transform Molecules...** from the resulting menu. This will open the *Transform Molecules* dialogue.

- *The Crystal Symmetry Operators* tab of the dialogue allows a molecule or structure be transformed by applying one of the symmetry operators of the space group.

- Select a symmetry operator by clicking on an entry in the list and click on **Transform** to apply the symmetry operation to the selected molecule or the whole structure.
- **Copy** will apply the symmetry operation to a copy of the selected molecule or the whole structure.
- Unit cell translations can be applied by selecting the appropriate integers for translations on a, b and/or c and clicking on **Transform** or **Copy**.

- The *Inversion and Translation* tab allows you to apply translations of +1/2 or -1/2 to the a, b or c axes, or to invert a structure or a molecule, irrespective of whether these transformations are allowed by the space group.
  - Transformations performed in this tab will change the coordinates of the structure that Mercury works with. Therefore, if a molecule is inverted and packing is then switched on, the resulting unit cell will be the inverse of the original structure.
  - The simulated powder pattern is automatically recalculated when the coordinates of the structure are changed. Therefore if you apply a transformation to the structure which is not allowed by the space group, the simulated powder pattern will be different from that of the original structure. For example, if you apply a translation of (1/2, 0, 0) to a structure in *P*4, the simulated powder pattern will change, indicating that the crystal structure has changed. Alternatively, if you invert a structure in $P2_12_12_1$, the simulated powder pattern remains the same, indicating the transformation has not resulted in a new structure.

12.6 Changing the Setting of a Space Group

- To change the setting of a space group, for example standardising a structure from $B2_1/c$ to $P2_1/c$ or switching between hexagonal and rhombohedral axes for a structure in *R*-3, choose **Edit** from the top-level menu followed by **Change Space Group Setting...**

- All possible settings for the space group are listed in the *Spacegroup Setting* drop down menu. Choose the new space group setting for the structure and the change will be applied immediately. Hit **OK** to accept the change and close the dialogue, or click on **Cancel** to revert any changes made and close the dialogue.

- The drop-down menu will include all settings listed in International Tables (*see International Tables for Crystallography, Volume A, Kluwer Academic Publishers, 1983*) as well as some additional settings that are observed in CSD structures, but not explicitly listed in International Tables.
- To standardise the structure to the first setting listed in International Tables, click on the **Standardise** button. This will also remove any origin shifts applied. Note that this will simply apply the standard setting, not necessarily the best setting where there are two equally reasonable options, such as $P2_1/c$ and $P2_1/n$.
- Data presented in the **More Info** spreadsheet (see Display of Text Information) will change to reflect the new space group setting chosen. For example, the space group and cell parameters reported in the Structure window will reflect the new setting, as will the atom coordinates and the symmetry operators shown in the Atoms and Symmetry windows respectively.

12.7 Inverting a Crystal Structure

- To invert the crystal structure currently being displayed, choose **Edit** from the top-level menu followed by **Invert Structure**. This will immediately invert the crystal structure.
- This will only have any effect if the crystal structure is in a Sohncke space group (such as $P1$, $P2_1$, $C2$, $P3_1$) – *i.e.* one of the 65 space groups that only contain the identity, rotation or screw axis symmetry elements. If the structure is in a non-Sohncke space group, or the model currently being visualised is not a crystal structure, this will do nothing.
- Data presented in the **More Info** spreadsheet (see Display of Text Information) will change to reflect the edited structure. For example, the space group name will change if the space group is one of an enantiomorphous pair (such as $P3_1$ and $P3_2$), as will the atom coordinates and the symmetry operators shown in the Atoms and Symmetry windows respectively.

12.8 Reducing the Symmetry of a Crystal Structure

- To reduce the symmetry of the crystal structure currently displayed from its original space group to one of the available subgroups, choose **Edit** from the top-level menu followed by **Change Spacegroup to Subgroup**.

- The current space group and setting are displayed in the dialogue including a link to the online edition of International Tables Volume A: Space-group symmetry. All available maximal subgroups will be displayed are listed in the *Available Subgroups*

drop down menu. Choose the subgroup you wish to reduce the symmetry to and then hit **Apply** to impose the change on the structure. You will see the information listed in the *Current Spacegroup* section change.

- To more easily visualise the resulting change in symmetry, turn on the display of symmetry elements before performing the space group to subgroup change (see Displaying Space-Group Symmetry Elements).
- To revert back to the original symmetry, use Mercury's undo functionality by choosing **Edit** from the top-level menu followed by **Undo**, or by hitting Ctrl+Z.
- This space group to subgroup transformation functionality can be useful for example when comparing crystal structures before and after phase transitions that involve a change in symmetry, or when trying to understand the symmetry relationships between space groups.

13 Visualising CSD- and PDB-based Intermolecular Contact Distributions

13.1 Overview

- IsoStar is a library of intermolecular interactions and contains data on:
  - intermolecular interactions in small-molecule crystal structures taken from the Cambridge Structural Database (CSD);
  - protein-ligand interactions in X-ray structures from the Protein Data Bank (PDB)
- Crystallographic information in IsoStar is presented as 3D scatterplots. Each scatterplot has been calculated by searching the CSD or PDB for nonbonded interactions between a pair of functional groups A and B. The A...B contacts are transformed so that the A groups are least-squared superimposed. The resulting scatterplot shows the experimentally observed distribution of B (the contact group) around A (the central group), e.g.

- Such a scatterplot gives information about the frequencies and directionalities of intermolecular contacts. The above example shows that OH groups (the contact group in this case) strongly hydrogen-bond to the carbonyl oxygen of esters (the central group) but not to the other oxygen atom.
- IsoStar can therefore be used to validate e.g. the geometry of an H-bond or short range interaction within a crystal structure.
- Once selected within Mercury a molecule will be broken down into constituent functional groups (IsoStar central groups). The fragment of interest is selected by the user and a list of available contact groups is presented. The scatterplot is then

loaded into the client window once the contact group is selected (see Analysing Molecule(s) for Available Contact Distributions).

- Alternatively, a scatterplot for a specific contact can be displayed (see Displaying a Distribution for a Specific Contact).
- For more information on IsoStar, please visit *http://www.ccdc.cam.ac.uk/products/csd_system/isostar/*

13.2 Setting Up Mercury For Use With IsoStar

- The location of the IsoStar server must be specified so that the scatterplots can be successfully retrieved.
- Click on **CSD-System** in the top-level menu and then select **IsoStar Settings...** from the resulting menu. Insert either the public IsoStar home page or your own local IsoStar home page URL in the resultant **IsoStar Settings** dialogue, e.g. *http://isostar.ccdc.cam.ac.uk/html/isostar.html* or *http://computer_name/html/isostar.html*.

13.3 Analysing Molecule(s) for Available Contact Distributions

- To analyse a molecule (or molecules) for functional groups around which contacts may exist in the IsoStar database:
  - Select an atom within the molecule (or molecules) of interest (see Ways of Selecting Atoms and Molecules).
  - Once an atom (or atoms) is selected, either click on **CSD-System** in the top-level menu and select **IsoStar Interaction Check...** from the resulting menu, or right-click on one of the selected atoms and select **IsoStar Interaction Check...** from the resultant pull-down menu.
  - A pop-up window will appear listing the central groups that were found in the molecule(s) selected.

- Pick the preferred central group by clicking on the appropriate hyperlink from the central group table, e.g. acetyl above. Picking a central group will take you to a page detailing the available contact groups.

- The contact groups are listed based on their elemental content.

- The numbers in the CSD or PDB columns correspond to the number of observations of the selected contact in the CSD or PDB respectively. Click on the number hyperlink to view the corresponding scatterplot.
- Further information on interacting with IsoStar scatterplots is provided in the IsoStar documentation: *http://www.ccdc.cam.ac.uk/support/documentation/#isostar*

13.4 Displaying a Distribution for a Specific Contact

- To analyse specific contacts:
  - Right-click on the contact of interest and select **IsoStar Interaction Check...** from the resultant pull-down menu.
    Note: Information on displaying contacts is provided elsewhere (see Finding Hydrogen Bonds and other Nonbonded Contacts).
  - A pop-up window will appear listing the available scatterplots that contain the selected contact.

  - Click on **Download** to display the appropriate CSD- or PDB-based scatterplot.
  - Further information on interacting with IsoStar scatterplots is provided in the IsoStar documentation: *http://www.ccdc.cam.ac.uk/support/documentation/#isostar.*

14 Geometry Validation

14.1 Overview

- Mogul is a library of molecular geometry derived from the Cambridge Structural Database (CSD) and provides information on the preferred values of bond lengths, valence angles, acyclic torsion angles and ring conformations.
- Mogul searches, initiated from within Mercury, can be used e.g. to validate the intramolecular geometries of newly determined crystal structures and identify unusual geometric features.
- Searches can be performed on a specific bond, valence angle, torsion angle or ring geometry in a query molecule. Alternatively, all geometric features within the loaded structure(s) can be validated.
- Mogul provides an analysis of the geometry of selected fragments. For example, if you run a search on a bond, valence angle or torsion in your molecule, Mogul will retrieve all entries from the CSD that have the same molecular feature (i.e. bond, angle or torsion fragment within the same structural environment) and return the experimental measurements of those matching fragments, together with the mean and standard deviation of their distribution. For rings, Mogul retrieves relevant ring fragments and reports the minimum and maximum ring RMSD.
- Query fragments will then be classified according to whether Mogul considers their geometry to be *Unusual* or *Not Unusual* based on statistical measures calculated from matching fragments in the CSD.
- For more information, please visit *http://www.ccdc.cam.ac.uk/products/csd_system/mogul/*

14.2 Performing a Geometry Check

- Select the atoms that are needed to define the geometric parameter(s) of interest by clicking on them with the left-hand mouse button (see Selecting and Deselecting Atoms and Molecules). For example, selecting three bonded atoms A-B-C will result in searches being run on the valence angle A-B-C and the two bond lengths A-B and B-C (searches can subsequently be restricted to certain fragment types only).
- If no atoms are selected then a complete analysis of all loaded molecules will be performed.
- *Note:* It is not possible to search for cyclic torsions, i.e. the central bond of the torsion angle must be acyclic. It is not possible to search for bonds, angles or torsions in which one or more defining atoms is a metal or hydrogen atom. It is not possible to search for non-bonded contacts or improper torsions, so chemically-bonded atom sequences must be selected in order to form a valid search fragment.
- Click on **CSD-System** in the top-level menu and then select **Mogul Geometry Check** from the resulting menu. This will open the *Mogul Search Settings* dialogue:

- The following search and display configuration options are available:
  - *Fragment types*: It is possible to restrict searches to certain fragment types. Switch on the check-boxes corresponding to the fragment types you wish to include.
  - *Search filters options*: Filters can be applied e.g. to exclude CSD structures with low experimental precision (see Applying Search Filters).
  - *Selection mode:* If the number of hit fragments that are structurally identical to the query fragment is insufficient then it is possible to find similar fragments and have these included in the results (see Finding More Hits).
  - *Hits Display Mode*: Click on the **Customise fragment classification...** button to customise the classification of fragments in the results display (see Configuring Results Display Options).
- Hit **Search** to accept the current settings and run the search.

14.2.1 Applying Search Filters

- Search results can be filtered e.g. to exclude structures with low experimental precision.
- Switch on the **Apply filters** check-box in the *Mogul Search Settings* dialogue and select the filters you wish to use. Available filters include:

- **R-factor**: Switch on the R-factor check box and then select one of the three options from the drop-down list in order to restrict the results to structures with R-factors less than or equal to 5%, 7.5% or 10%.
- **Exclusion of Solvents or Non-solvents**: Switch on the Exclude check box and then select either Solvents or Non-solvents from the drop-down list. Results will be filtered to exclude fragments from either solvent or non-solvent molecules. *Note*: A molecule is considered to be a solvent if it is found in a CCDC catalog of known solvent molecules. However, if a crystal structure contains only solvent molecules then the largest molecule will be assumed to be a non-solvent.
- **Heaviest Element**: Switch on the Heaviest Element check box and then select an element from the drop-down list. Results will be filtered to exclude fragments from structures that contain elements heavier than the specified element.
- **Exclusion of Organometallics or Organics**: Switch on the Exclude check box and then select either Organometallics or Organics from the drop-down list. Selecting to exclude Organometallics will eliminate from the results any fragments from structures that contain a transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po. Selecting to exclude Organics will eliminate fragments from organic structures (i.e. structures that do not contain any of the elements listed above).

Note: When filter have been applied it will not be possible to retrospectively incorporate excluded entries (without running the search again).

14.2.2 Finding More Hits

- When searching on a particular geometric feature, the number of hit fragments found that are structurally identical to the query fragment may not be sufficient. In such cases, the search can look for fragments that, while not identical to the query, are sufficiently similar as to be relevant. These fragments will then be incorporated into the results, which would therefore consists of observations from fragments that are similar to the query fragment as well as fragments that are identical to the query (if these exist).
- To prohibit the inclusion of similar fragments, select the **Only find fragments that match exactly** radio button in the *Mogul search settings* dialogue.

- If a minimum specified number of fragments that are identical to the query are not found a search for similar fragments can be performed. To enable this, select the radio-button labelled **Find similar fragments if number of exact fragments is less than**. The criteria for initiating a search for similar fragments can be set independently for bond, angle, torsion angle and ring fragments.

14.2.3 Configuring Results Display Options

- Query fragments will be classified (in the *Mogul Results Viewer*) according to whether Mogul considers their geometry to be *Unusual* or *Not Unusual* based on statistical measures calculated from matching fragments in the CSD.
- To customise the statistical criteria used to classify fragments click on the **Customise fragment classification...** button in the *Mogul Search Settings* dialogue. This will open the *Customise Classification of Fragments* window:

- For bonds and valence angles, the following statistical criteria can be used to classify the geometry of a fragment:
  - *| x - Mean |* : indicates how far the fragment query value deviates from the mean of the corresponding distribution.
  - *| z-score |* : is the difference between observed and mean values of a geometric parameter divided by the standard deviation of the distribution.
- For torsion angles, the following statistical criteria can be used:
  - *Local density* : indicates the percentage of experimental measurements that are within *x* degrees of the value in your query molecule.
  - *| d(min) |* : is the difference between the value of the torsion angle in your query and the nearest torsion angle in the corresponding distribution.

- For rings, the following criteria can be used:
  - *Density* : indicates the fraction of experimental measurements that fall within *x* degrees of the torsion RMSD (root mean square deviation) from the query molecule. The torsion RMSD reports the deviation of the ring conformation of each hit fragment from the query ring, in terms of the corresponding intracyclic torsion angles.
  - *| d(min) |* : the value of the torsion RMSD calculated for the nearest experimental structure to the query ring
- A bond length, valence angle, torsion angle or ring conformation in your query molecule will subsequently be classified as either *Unusual* or *Not Unusual* according to the criteria specified.
- Query fragments are further classified according to whether or not there is sufficient data to draw a reliable conclusion, i.e. depending on the number of experimental measurements found by Mogul, the data will be classified as *Enough hits* (reliable conclusions can be drawn) or *Few hits* (not enough data to be confident). The criteria for classifying the number of hits as *Few* can be set independently for bond, angle, torsion angle and ring fragments.
- Hit **OK** to accept the current settings, **Cancel** to close the *Customise Classification of Fragments* dialogue without saving any changes, or **Reset** to return settings to their default values.
- The criteria governing the classification of fragments can also be customised after a search has been run. Click on the **Fragments...** button in the *Mogul Results Viewer* followed by **Customise...** in the resulting *Fragment Display Selection* dialogue. By default, changes to classification settings made when viewing results will apply only to the results currently on display. To make the changes apply to all subsequent geometry checks run during the current Mercury session select **Set as default**.

14.3 Viewing and Analysing Results

- On completion of a search, results are displayed in the *Mogul Results Viewer*.
- Data are presented in the form of a tree organised first by fragment type (bond, angle, torsion, ring) and subsequently by structure identifier. Each search fragment is then listed together with its associated statistical measures, calculated from matching entries in the CSD.
- Fragments are classified and colour coded for easy identification of *Unusual* geometric features (see Configuring Results Display Options). In the example shown below the bond fragment C4-C3 is classified as *Unusual* when compared to matching entries in the CSD:

Type	Molecule	Fragment	Classification	No. of hits	Query value	Mean	Std. dev.	| z-score |	Local density
bond									
:									
		C4 C3	Unusual (enough hits)	466	1.554	1.520	0.014	**2.416**	
		O2 C2	Not unusual (enough hits)	10000	1.454	1.425	0.024	**1.214**	
		C5 C4	Not unusual (enough hits)	305	1.536	1.517	0.025	**0.775**	
		O1 C1	Not unusual (enough hits)	368	1.382	1.389	0.037	**0.190**	
		O5 C5	Not unusual (enough hits)	1285	1.440	1.433	0.017	**0.434**	
		C3 C2	Not unusual (enough hits)	6057	1.523	1.519	0.027	**0.134**	
		O5 C1	Not unusual (enough hits)	42	1.421	1.425	0.015	**0.253**	
		O3 C3	Not unusual (enough hits)	10000	1.426	1.425	0.024	**0.028**	
		O4 C4	Not unusual (enough hits)	5464	1.431	1.429	0.023	**0.055**	
		C1 C2	Not unusual (enough hits)	140	1.524	1.525	0.019	**0.018**	
angle									
:									
		O1 C1 O5	Not unusual (enough hits)	42	113.095	109.878	2.731	**1.178**	
		C5 C4 C3	Unusual (enough hits)	201	107.121	109.708	1.260	**2.053**	
		O2 C2 C1	Not unusual (enough hits)	140	107.328	109.890	2.127	**1.204**	
		O4 C4 C3	Not unusual (enough hits)	466	111.666	110.226	1.942	**0.741**	
		C1 C2 C3	Not unusual (enough hits)	133	109.237	110.535	1.585	**0.819**	
		C4 C3 C2	Not unusual (enough hits)	187	109.668	110.886	1.773	**0.687**	
		O3 C3 C2	Not unusual (enough hits)	10000	109.245	110.353	2.242	**0.494**	
		O5 C5 C4	Not unusual (enough hits)	215	111.930	110.828	1.594	**0.691**	
		O4 C4 C5	Not unusual (enough hits)	305	108.359	109.142	2.070	**0.378**	

- Highlighted in bold is the classification metric for the fragment, specifically *z-score* (bonds and angles) and *Local density* (torsions and ring fragments). The definition of these parameters is provided elsewhere (see Configuring Results Display Options).
- To control which data items are included in the spreadsheet click on the **Columns** button and select the items you to be displayed: the spreadsheet will update automatically to reflect the selections made.
- The order in which spreadsheet columns are displayed can be changed by pressing down the left-hand mouse button on a column header, moving the cursor while keeping the mouse button depressed to the new location, and then releasing the button.
- The rows of a spreadsheet can be sorted according to the values in any of the columns. For example, to sort the rows by **z-score** (in descending order), click on the **z-score** button at the top of the spreadsheet.
- Clicking on a row in the spreadsheet will result in the chosen fragment being highlighted in the main structure display window. Click on a row again to deselect it. To clear all selections hit the **Deselect all fragments** button.
- Double-clicking on a row will result in the fragment being displayed in the native Mogul interface. From here the underlying histogram and relevant entries from the CSD can be examined in detail.
- It is possible to control which fragments are displayed based on their classification. This can be useful when analysing results from a large number of query fragments. Click on the **Fragments...** button. This will open the *Fragment Display Selection* dialogue:

- Select to display either **All fragments**, **Fragments with unusual geometries**, or **Fragments with geometries that are not unusual**.
- To exclude results for fragments which have *few* hits switch on the **Exclude results based on insufficient data** check-box.
- To customise the statistical criteria used to classify fragments click on the **Customise...** button. This will open the *Customise Classification of Fragments* window (see Configuring Results Display Options).
- Hit **Apply** (leaves window open), or **OK** (closes window) to update the *Mogul Results Viewer* according to the selection made. To close the *Fragment Display Selection* window without making any changes hit **Cancel**.

15 Displaying Additional Information about the Structure

15.1 Additional Data Items Available for Display

When a 3D structure is on display in the main Mercury window, the following additional information may be viewed in a separate window:

- Depending on file format, text information such as the literature reference, chemical name, etc. (see Display of Text Information).
- If the structure is from the CSD, its chemical diagram (see Chemical Diagram Display).
- A spreadsheet listing the atoms currently displayed in the 3D window, together with various atom properties such as fractional coordinates (see Spreadsheet Display of Atoms and Their Properties).
- A spreadsheet listing the bonds currently displayed in the 3D window, together with various bond properties such as type and length (see Spreadsheet Display of Bonds and Their Properties).
- A spreadsheet listing the non-covalent contacts currently displayed in the 3D window, together with various contact properties such as interatomic distance and the symmetry operations used to generate the contact (see Spreadsheet Display of Contacts and Their Properties).
- A spreadsheet listing the centroids currently displayed in the 3D window, together with various properties such as orthogonal coordinates (see Spreadsheet Display of Centroids and Their Properties).
- A spreadsheet listing the planes currently displayed in the 3D window, together with various properties (see Spreadsheet Display of Planes and Their Properties).
- A spreadsheet listing the space group symmetry operators (see Spreadsheet Display of Space Group Symmetry Operators).
- Separate spreadsheets listing any measured distances, angles and torsion angles currently displayed in the 3D window (see Spreadsheet Display of Distances, Angles and Torsion Angles).

15.1.1 Display of Text Information

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Structure Information...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Structure Information...**

- Information such as the literature reference, chemical name, etc. for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list) may be available.
- When viewing CSD entries that which have a Digital Object Identifier (DOI) a clickable hyperlink to the original literature source will be available in the *Literature Reference* field.
- When displaying the space group, a clickable hyperlink will be available to connect to International Tables Volume A: Space-group symmetry (online edition) to visualise the symmetry in that space group.
- The specific data items displayed can be changed by clicking on the **Customise...** button (see Controlling which Data Items are Displayed). Data items that can be displayed in the **Structure Information** page are in the following categories:
    - **Cell** (e.g. unit cell lengths, angles, volume, reduced cell information).
    - **Experimental** (e.g. pressure, density, radiation probe, powder study).
    - **CSD Internals** (e.g. modification date, previous refcode, remarks).
    - **Chemical** (e.g. bioactivity, conformer, isomer, source).
    - **Crystal** (e.g. color, habit, melting point, disorder, packing coefficient).
    - **References** (e.g. literature references, cross-references, CCDC number).
- Depending on the file format, some data items may not be available.

15.1.2 Chemical Diagram Display

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Chemical Diagram...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Chemical Diagram...**

- A chemical diagram of the currently selected structure (i.e. the one selected from the **Current structure** drop-down list) may be available.
- No diagram will be displayed if the currently loaded structure is not from the Cambridge Structural Database.
- Complex structures (e.g. cages) may have no diagram, in which case the display area will be empty or will just contain a text-string formula.

15.1.3 Spreadsheet Display of Atoms and Their Properties

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Atom List...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Atom List...**
- The resulting spreadsheet will list all atoms for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list), including those that are hidden (e.g. hydrogens, even if the *Show hydrogens* box is turned off).

	Number	Label	Charge	SybylType	Xfrac	Yfrac	Zfrac	Symm. op.
1	1	Cl1	0	Cl	0.4706	1.0273	0.3358	x,y,z
2	2	C1	0	C.ar	0.4115	0.8258	0.3452	x,y,z
3	3	C2	0	C.ar	0.3797	0.7928	0.4060	x,y,z
4	4	C3	0	C.ar	0.3331	0.6301	0.4102	x,y,z
5	5	C4	0	C.ar	0.3186	0.4990	0.3558	x,y,z
6	6	C5	0	C.ar	0.3525	0.5303	0.2964	x,y,z
7	7	C6	0	C.ar	0.3996	0.6953	0.2905	x,y,z
8	8	C7	0	C.3	0.3969	0.9312	0.4662	x,y,z
9	9	C8	0	C.2	0.3971	0.8172	0.5328	x,y,z
10	10	C9	0	C.2	0.4400	0.6309	0.5441	x,y,z
11	11	N1	0	N.2	0.4800	0.5894	0.4949	x,y,z
12	12	C10	0	C.3	0.4351	0.5033	0.6074	x,y,z
13	13	C11	0	C.2	0.3550	0.9205	0.5737	x,y,z
14	14	C12	0	C.2	0.3233	1.1006	0.5370	x,y,z
15	15	C13	0	C.3	0.2981	0.9940	0.6769	x,y,z
16	16	O1	0	O.3	0.3443	0.8722	0.6385	x,y,z
17	17	O2	0	O.2	0.2836	1.2346	0.5552	x,y,z

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include:
  - The atom number (basically arbitrary, depending on the order in which atoms were loaded into the display area).
  - The atom label. Symmetry-equivalent atoms will usually have the same label.
  - The atom's elemental symbol and atomic number.
  - The formal charge associated with the atom (this can sometimes be rather arbitrary, e.g. in delocalised systems).
  - The Sybyl atom type of the atom (see Sybyl Atom Type).
  - The fractional (see Fractional Coordinates) and orthogonal coordinates (see Orthogonal Coordinates) of the atom. Estimated standard deviations (ESDs) can also be displayed for fractional coordinates.
  - The symmetry operation used to generate the atom; and the "base atom" from which it was generated (i.e. its symmetry-equivalent atom in the asymmetric unit of the crystal structure (see Asymmetric Unit).
  - The atom's occupancy.
  - Various ADP properties (see Atomic Displacement Parameters (ADPs)), such as *ADP type* i.e. isotropic or anisotropic, U_{ij} values and U_{equiv} values, B_{ij} and B_{equiv} values.
  - Various hydrogen-bonding properties such as whether the atom is a donor, an acceptor, what type of donor or acceptor and whether the atom is considered to be capable of accepting a hydrogen bond.
  - Whether the atom is a stereocentre or not.

- The current display status of the atom, i.e. whether or not it is hidden and, if not, its display style.

- Rows in the spreadsheet can be selected (this selects the corresponding atoms in the 3D window) and reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).

15.1.4 Spreadsheet Display of Bonds and Their Properties

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Bond List...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Bond List...**
- The resulting spreadsheet will list all bonds for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list), including those that are hidden (e.g. bonds to hydrogens, if the *Show hydrogens* tick box is turned off).

	Number	Atom1	Atom2	Type	Cyclicity	Length	SybylType
1	1	Cl1	C1	Single	acyclic	1.7299	1
2	2	C1	C2	Aromatic	cyclic	1.3867	ar
3	3	C2	C3	Aromatic	cyclic	1.3756	ar
4	4	C3	C4	Aromatic	cyclic	1.3842	ar
5	5	C4	C5	Aromatic	cyclic	1.3812	ar
6	6	C5	C6	Aromatic	cyclic	1.3952	ar
7	7	C6	C1	Aromatic	cyclic	1.3767	ar
8	8	C7	C2	Single	acyclic	1.5133	1
9	9	C8	C7	Single	cyclic	1.5032	1
10	10	C9	C8	Single	acyclic	1.4542	1
11	11	N1	C9	Double	acyclic	1.2953	2
12	12	C10	C9	Single	acyclic	1.4911	1
13	13	C11	C8	Double	cyclic	1.3378	2
14	14	C12	C11	Single	cyclic	1.4748	1
15	15	C13	O1	Single	acyclic	1.4197	1
16	16	O1	C11	Single	acyclic	1.3442	1
17	17	O2	C12	Double	acyclic	1.2052	2
18	18	O2	C7	Single	cyclic	1.4608	1

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include:
  - The bond number (basically arbitrary, depending on the order in which bonds were loaded into the display area).
  - The labels of the two atoms involved in the bond.
  - The bond type.
  - The bond cyclicity (cyclic or acyclic).
  - The bond length.
  - The Sybyl type of the bond (generally the same as the normal bond type, but, e.g., amide bonds are given the special type *am*).

- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).
- Enable the **Include both ways round** checkbox to include each bond twice in the spreadsheet. The second occurrence will have the *Atom1* and *Atom2* cells reversed. When the spreadsheet is then sorted by either *Atom1* or *Atom2*, all bonds for each atom will be grouped together.

15.1.5 Spreadsheet Display of Contacts and Their Properties

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Contacts List** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Contacts List**.
- The resulting spreadsheet will list all nonbonded contacts for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list).

	Number	Atom1	Atom2	Length	Length-VdW	Symm. op. 1	Symm. op. 2	Expanded
1	1	H7D	Cl1	3.030	0.080	1-x,1-y,1-z	x,-1+y,z	Yes
2	2	Cl1D	H7	3.030	0.080	1-x,1-y,1-z	x,-1+y,z	Yes
3	3	C11	O2D	3.039	-0.181	x,y,z	1/2-x,-1/2+y,z	No
4	4	C12	O2D	3.138	-0.082	x,y,z	1/2-x,-1/2+y,z	No
5	5	H2	C3D	2.984	0.084	x,y,z	1/2-x,-1/2+y,z	No
6	6	H11	O2D	2.685	-0.035	x,y,z	1/2-x,-1/2+y,z	No
7	7	C3	H2D	2.984	0.084	x,y,z	1/2-x,1/2+y,z	Yes
8	8	O2	C11D	3.039	-0.181	x,y,z	1/2-x,1/2+y,z	Yes
9	9	O2	C12D	3.138	-0.082	x,y,z	1/2-x,1/2+y,z	Yes
10	10	O2	H11D	2.685	-0.035	x,y,z	1/2-x,1/2+y,z	Yes
11	11	H10D	C6	2.956	0.056	1-x,1-y,1-z	1-x,-1+y,1/2-z	No
12	12	C6D	H10	2.956	0.056	1-x,1-y,1-z	1-x,-1+y,1.5-z	No

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include:
  - The contact number (basically arbitrary, depending on the order in which contacts were loaded into the display area).
  - The labels of the two atoms involved in the contact and the symmetry operations used to generate them from the asymmetric unit of the structure.
  - The length of the contact; and the length minus the sum of the van der Waals radii of the two atoms.
  - Whether or not the contact has been expanded (see What Network Expansion Is).

- The positions of the two atoms of the contact given in orthogonal (see Orthogonal Coordinates) or fractional coordinates (see Orthogonal Coordinates).

- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).

15.1.6 Spreadsheet Display of Centroids and Their Properties

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Centroids List...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Centroids List...**.
- The resulting spreadsheet will list all centroids calculated for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list).

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include:
  - The centroid number (basically arbitrary, depending on the order in which centroids were defined in the display area).
  - The label of the centroid.
  - The orthogonal coordinates (see Orthogonal Coordinates) or fractional coordinates (see Orthogonal Coordinates) of the centroid.
- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).

15.1.7 Spreadsheet Display of Planes and Their Properties

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select Planes List... from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Planes List...**.

- The resulting spreadsheet will list all planes calculated for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list).

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include:
  - The plane number (basically arbitrary, depending on the order in which planes were defined in the display area).
  - The normal which contains the orthogonal x, y and z components of the vector perpendicular to the plane.
  - The shortest distance from the plane to the origin of the unit cell.
  - The nearest **Miller Plane** (see Miller Plane).
- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).

Note: The *Normal* and *Distance* parameters define the plane through Hesse's normal form of a plane:

`Normal . r = Distance` (where r is the position vector of a point in the plane).

15.1.8 Spreadsheet Display of Space Group Symmetry Operators

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Symmetry Operators List...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Symmetry Operators List...**
- The resulting spreadsheet will list all the symmetry operators available in the space group to which the currently selected structure (i.e. the one selected from the **Current structure** drop-down list) belongs.

	Number	Symm. Op.	Description	Detailed Description	Order	Type
1	1	x,y,z	Identity	Identity	1	1
2	2	1/2-x,1/2-y,1/2+z	Screw axis (2-fold)	2-fold screw axis with direction [0, 0, 1] at 1/4, 1/4, z with screw component [0, 0, 1/2]	2	2
3	3	1/2+x,1/2-y,-z	Screw axis (2-fold)	2-fold screw axis with direction [1, 0, 0] at x, 1/4, 0 with screw component [1/2, 0, 0]	2	2
4	4	-x,y,1/2-z	Rotation axis (2-fold)	2-fold rotation axis with direction [0, 1, 0] at 0, y, 1/4	2	2
5	5	-x,-y,-z	Inversion centre	Inversion at [0, 0, 0]	2	-1
6	6	1/2+x,1/2+y,1/2-z	Glide plane	Glide plane perpendicular to [0, 0, 1] with glide component [1/2, 1/2, 0]	2	-2
7	7	1/2-x,1/2+y,z	Glide plane	Glide plane perpendicular to [1, 0, 0] with glide component [0, 1/2, 0]	2	-2
8	8	x,-y,1/2+z	Glide plane	Glide plane perpendicular to [0, 1, 0] with glide component [0, 0, 1/2]	2	-2

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include the symmetry operator itself, the order and type of the operator and a brief description of the operator, e.g. "Inversion centre" (see Symmetry Operators).
- The colour shown in the Number column corresponds to the colour used to indicate the symmetry relationship between molecules when colouring **by Symmetry Operation** (see Symmetry Operation Colouring).
- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets).
- It is possible to display space-group symmetry elements (see Displaying Space-Group Symmetry Elements).

15.1.9 Spreadsheet Display of Distances, Angles and Torsion Angles

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **Distances List...**, **Angles List...** or **Torsions List...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **Distances List...**, **Angles List...** or **Torsions List...**
- The resulting spreadsheet will list all geometric measurements (distances, angles or torsion angles) displayed for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list).

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include the number of the measured geometric parameter, the objects involved in the measurement and the value of the geometric parameter.
- Rows in the spreadsheet can be reordered (see Manipulating Spreadsheets) and may be written out to a file (see Saving Spreadsheets).

15.1.10 Spreadsheet Display of All Angles and All Torsions

- Either click on the **More Info** button (bottom centre of the main Mercury window) and then select **All Angles...**, **All Torsions...** from the resulting menu; or hit **Display** in the top-level menu, followed by **More Information** and then **All Angles List...** or **All Torsions List...**
- The resulting spreadsheet will list all valence angles or torsion angles for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list). If the original atom coordinates had ESDs, calculated ESDs are displayed for the angles.

	Number	Atom1	Atom2	Atom3	Atom4	Torsion
1	1	Cl1	C1	C2	C3	179.90
2	2	Cl1	C1	C2	C7	-0.94
3	3	C6	C1	C2	C3	-2.70
4	4	C6	C1	C2	C7	176.46
5	5	C1	C2	C3	C4	1.05
6	6	C1	C2	C3	H1	-179.02
7	7	C7	C2	C3	C4	-178.10
8	8	C7	C2	C3	H1	1.83
9	9	C2	C3	C4	C5	1.06
10	10	C2	C3	C4	H2	-176.26
11	11	H1	C3	C4	C5	-178.86
12	12	H1	C3	C4	H2	3.81
13	13	C3	C4	C5	C6	-1.63
14	14	C3	C4	C5	H3	177.71
15	15	H2	C4	C5	C6	175.76
16	16	H2	C4	C5	H3	-4.90

- The data items displayed can be altered by using the **Customise** button (see Controlling which Data Items are Displayed). Available data items include the number of the measured geometric parameter, the objects involved in the measurement and the value of the geometric parameter.

15.2 Controlling which Data Items are Displayed

- The data items displayed in the *Structure* information pane or the *Atoms, Bonds, Contacts*, *Centroids, Planes, Symmetry, Distances, Angles and Torsions* spreadsheets can be changed by clicking on the button in the top-left corner; this button is labelled **Customise**.
- This opens a window showing two lists. The right-hand list shows those data items that are currently being displayed. The left-hand list shows other data items that it is possible to display. If an item occurs in the right-hand list, it will be absent from the left-hand list, and vice versa.

- To add a new item, select the required item by clicking on it in the left-hand list, then hit the **Add** button.
- To remove an item, click on it in the right-hand list and hit the **Remove** button.
- Alternatively, double-click on an item to transfer it from one list to the other.
- To change the order in which the items are arranged, click on an item in the right-hand list and use the **Up** or **Down** button to alter its position in the list.
- Using the **Add, Remove, Up** and **Down** buttons will not change the corresponding display or spreadsheet until the **OK** button is hit to close the *Customise* dialogue window.

15.3 Manipulating Spreadsheets

- The rows of a spreadsheet may be sorted on the values in any column by clicking on the column header. For example, to sort the rows of an *Atoms* spreadsheet by label, click on the **Label** button at the top of the spreadsheet. To reverse the order, click on the **Label** button again.
- Right-clicking on a row in a spreadsheet will produce the same menu as would have been displayed had the corresponding object (atom, bond, contact, centroid, plane etc.) been right-clicked in the main Mercury 3D display window. Thus, for example, an atom's display style can be changed by right-clicking on that atom's row in the spreadsheet, selecting the **Styles** option from the resulting pull-down menu and picking the desired display style:

- Any row of an *Atoms* spreadsheet can be selected by clicking on it. Clicking while holding down the *Control* key will retain row selections that have already been made; otherwise, they will be lost.
- Selecting a row in an *Atoms* spreadsheet causes the corresponding atom in the Mercury 3D display window to also be selected.
- Several contiguous rows of an *Atoms* spreadsheet may be selected by positioning the cursor on the first of the required rows, depressing the mouse button, moving the cursor down to the last of the required rows while keeping the mouse button depressed, and then releasing the mouse button. Performing this sequence of action while holding down the *Control* key will retain row selections that have already been made; otherwise, they will be lost.
- All rows of an *Atoms* spreadsheet may be selected or deselected by use of the **Select All...** and **Deselect All...** buttons.
- If several rows of an *Atoms* spreadsheet are selected, and then one of the selected rows is right-clicked and an option picked from the resulting pull-down menu, the selected action will be applied to all the atoms corresponding to the selected rows.
- Sorting and selecting contiguous rows can be very useful. For example, sorting by label will place symmetry-equivalent atoms next to each other in a spreadsheet, so that they can easily be selected and, say, assigned the same display style.

15.4 Saving Spreadsheets

Any of the available spreadsheets (*Atoms, Bonds, Contacts, Centroids, Planes, Distances, Angles, Torsions*) may be saved to file by hitting the **Save...** button and specifying a filename. Available formats are:

- *.tsv* - data fields on each line separated by the tab character.

- *.csv* - data fields on each line separated by a comma.
- *.txt* - data fields on each line separated by a space.

All rows are written out, irrespective of whether or not the row is selected.

16 Powder Diffraction Patterns

16.1 Displaying and Saving Powder Diffraction Patterns

- A simulated powder diffraction pattern can be calculated and displayed for loaded structures by hitting the **Powder...** button situated at the bottom-centre of the main Mercury window, or by selecting **Calculate** from the top-level menu followed by **Powder Pattern...** from the resulting pull-down menu.
- The powder diffraction pattern for the currently selected structure (i.e. the one selected from the **Current structure** drop-down list) will be shown. The display plots diffraction intensity against 2θ for a given wavelength of radiation.
- All powder patterns are simulated, by default, using the $CuK_{\alpha 1}$ radiation of 1.54056Å. The wavelength used to simulate the pattern can be changed by clicking on **Customise** (see Customising the Powder Pattern Display).

- The wavelength (or wavelengths) used to calculate the pattern is (are) displayed in the bottom left corner of the powder pattern display.
- A number of settings can be altered by hitting the **Customise...** button at the top-left of the powder-pattern display (see Customising the Powder Pattern Display).
- To zoom in on any part of the powder-pattern display, position the cursor, depress the left mouse-button, and move the cursor while keeping the mouse button depressed. The area enclosed by the resulting rectangle will be magnified and displayed when the mouse button is released. To reset the display hit the **Reset zoom** button at the top-left of the dialogue.

- The small vertical bars under the simulated intensity profile indicate the 2θ values at which peaks can occur given the unit cell dimensions and space group. Some of these peaks, of course, may be too weak to show in the profile.
- The simulated powder pattern data may be saved in out in a variety of different formats by hitting the **Save...** button. The format options include the following:
  - ASCII powder pattern format (`.xye` format, see xye Format)
  - Reflection list (`.hkl` format)
  - Tab-separated (`.tsv` format)
  - Tab-separated, with ESDs (`.tsv` format)
  - Bruker RAW format (`.raw` format)
- The current display can also be saved as an image in various file formats by hitting the **Save image...** button.
- The mouse cursor position (the position of the black cross hair cursor) is displayed on-screen to facilitate the identification and comparison of peaks.

16.2 Customising the Powder Pattern Display

- A number of settings can be altered by hitting the **Customise...** button at the top-left of the powder-pattern display. This opens the *Powder Pattern Settings* dialogue. Depending on the type of licence you have, some of the available options here may not be present (see below).

- For *Wavelength, Start (degrees 2 theta), Stop (degrees 2 theta)* and *Step (degrees 2 theta)*, type in the desired values and hit **Apply** (which leaves the dialogue open) or **OK** (closes the dialogue).
- For users with a *CSD-System licence* (see CSD-System) or above (i.e. CSD-Materials, CSD-Discovery or CSD-Enterprise), a second wavelength can also be included. This is

intended to allow simulating non-monochromated laboratory radiation, for example $CuK_{\alpha 1}$ (= 1.54056 Å) includes $CuK_{\alpha 2}$ (= 1.54439 Å) with relative intensities of 1.0 and 0.5 respectively (the *Scale* values in the dialogue).

- The **Include hydrogens** tick box can be switched on or off to include or exclude H atoms from the powder pattern simulation.
- *FWHM (2 theta)*: Full Width at Half Maximum. When using small values for *FWHM (2 theta)*, the value of *Step (degrees theta)* can also be decreased to smooth the appearance of the powder diffraction plot.

- Tick marks indicating peak positions for reflections and systematic absences can be displayed and customised. To show systematic absences enable the **Show systematic absences** check-box.
- The line colour for the displayed powder pattern can be customised using the **Line colour** selection panel.
- Tick marks for systematic absences are coloured differently to those for reflections with a non-zero intensity. To customise the tick mark colour for reflections or systematic absences click on the corresponding coloured panel to choose the colour you want. This will open the *Select Color* dialogue box (see Symmetry-Equivalence Colouring).

- For users with a *CSD-Materials licence* (see CSD-Materials), the presence of preferred orientation can also be simulated using the March-Dollase formula: $(r^2\cos^2(\alpha) + \sin^2(\alpha)/r)^{-3/2}$, where r is the amount of preferred orientation (the March-Dollase parameter field in the dialogue) and α is the angle between the normal to the plane of a reflection and the normal to the hkl plane specified in the dialogue. If $r = 1.0$, there is no preferred orientation; if $r < 1.0$, the intensities of reflections with orientation near hkl is increased, if $r > 1.0$, the intensities of reflections with orientation near hkl is decreased.
  See: Dollase, W. A. (1986). *J. Appl. Cryst.*, **19**, 267-272.
- The BFDH morphology can be used to get an estimate of the hkl direction that is most likely to suffer from preferred orientation.

16.3 Technical Details of the Powder Pattern Simulation

- The Lorentz-polarisation correction assumes a laboratory X-ray source. No absorption is simulated. Fixed slit widths are assumed. No background is included.
- All non-hydrogen atoms are assumed to have isotropic atomic displacement parameters (U_{iso}) of 0.05 Å^2. Hydrogen atoms for which 3D coordinates are available are taken into account and assigned U_{iso} values of 0.06 Å^2.
- The powder pattern simulator takes site occupation factors into account. This corrects the patterns generated for disordered structures read from CIF and SHELX Res files.
- All reflections have a symmetric pseudo-Voight peak shape with a full width half maximum of 0.1 degree 2θ, corresponding to medium resolution laboratory data.
- The (0, 0, 0) reflection is excluded.
- The default 2θ resolution is 50.0 degrees, which, for the default $CuK_{\alpha 1}$ radiation, corresponds to a direct space resolution of 3.0Å.

- Experimental displacement parameters, either isotropic or anisotropic, are taken into account in the calculation if available, i.e. if the crystal structure has been read in from a .cif or .res file.

17 Saving Files and Copying to the Clipboard

Mercury can write:

- Graphics files containing the image currently on display (see Saving Mercury Displays as Images).
- Files containing the 3D coordinates of the molecule(s) on display. You have the choice of saving as:
  - An *.mryx* file, which retains full information about display properties such as atom colours, etc., but can only be read back into Mercury (see Saving Structures in Mercury Format).
  - A file in external formats such as *.mol2* or *.pdb*, which do not hold information about all the Mercury display settings but can be read into other programs (see Saving Structures in External (3rd Party) Formats).

You can also copy images to the clipboard (see Copying Images to the Clipboard).

17.1 Saving Mercury Displays as Images

17.1.1 Saving Images from Mercury

- The current view can be saved as a graphical image in the following file formats:
  - Bitmap file (**.bmp*).
  - Portable Network Graphics file (**.png*).
  - jpeg files (**.jpg* or **.jpeg*).
  - gif files (**.gif*).
  - tif files (**.tif*): uncompressed .tif files written.
- To save the current image, hit **File** in the top-level menu, **Save As...** in the next menu, then choose the required format in the *File Format* dialogue box and hit **OK**. This opens the *Save As* dialogue box which allows you to specify the name and location of the file to be created.
- The *Save Image* dialogue box then appears. A number of default image sizes are provided, as well as multiples of the current window size. User-defined sizes may also be entered. Note that sizes over 1500 by 1500 pixels require large amounts of memory.

17.1.2 POV-Ray Image Rendering from Mercury

- To render an image using POV-Ray directly from Mercury, pick **File** in the top-level menu, then **POV-Ray Image…** in the resulting pull-down menu
- The *POV-Ray Image* control window will be displayed:

- The following options are available:
  - *Width (pixels) and height (pixels)*: Use these settings to control the size and aspect ratio of the image to be rendered. Larger numbers will produce a higher resolution image, but will take longer to render.
  - *Material properties*: Use this to control the way that objects such as atoms and bonds are represented in the image. Options include; Standard Mercury, Shiny, Matt, Metallic and Iridescent.
  - *File format*: The image can be rendered using a PNG or JPG output file format. Note that transparent backgrounds can only be rendered using the PNG output format and JPG is not supported from POV-Ray on Linux operating systems.
  - *Background*: Use this to control the display of the background in the rendered image. Options include; As Mercury, White, Black, Custom Colour and Transparent.

- *Generate Animation Frames*: Use this option to generate a series of frames for an animated movie – for more information see POV-Ray Movie Rendering from Mercury.
- *POV-Ray executable*: Version 3.7 of POV-Ray is included with the Mercury distribution and the path will be included here by default. If the POV-Ray installation has been moved or you wish to use a different installation, then browse to the location of the POV-Ray engine executable here.
- *Output directory*: The output directory path will default to a writable location, but you can also choose to change the output directory path for images and associated files here.

- A lower resolution version of the image to be rendered can be previewed within the *POV-Ray Image* control window by clicking on **Preview**. This allows easy checking of the image aspect ratio, background, lighting and materials settings prior to rendering.
- Once the view and settings have been chosen, click on **Render** to generate the final high resolution image using POV-Ray.

17.1.3 POV-Ray Movie Rendering from Mercury

- To render a set of images using POV-Ray for movie generation, pick **File** in the top-level menu, then **POV-Ray Image…** in the resulting pull-down menu as you would for a normal image.
- Select the view and settings as for a regular image, but ensure that the *Generate Animation Frames* checkbox is ticked.
- The following options are available:
  - *Rotate around*: The animation will consist of a rotation about one axis – use these settings to control which rotation axis will be used. Note that the rotation axis is relative to the screen, not to any crystallographic axis. You can assess what the rotation will look like using the **Animate** tool bar, see Animating the Display for more details on this.
  - *Number of Frames*: Use this to control the number of frames generated for the animation (*n*). The animation produced will be a 360° rotation about the chosen axis, so each frame will relate to a rotation of 360/*n*°. A greater number of frames will produce a smoother movie, but result in a larger movie file size.
- The animation frames will be written to a sub-folder in the chosen output directory with an automatically generated sub-folder name based on the structure identifier.
- Once the view and settings have been chosen, click on **Render** to generate the final set of high resolution images using POV-Ray.
- Now that the animation frames have been rendered, you can collect the frames together into a single movie file using any of a range of third-party software.

17.1.4 Advanced Rendering with POV-Ray

- A POV-Ray input file can also be generated for view and rendering in POV-Ray without using the interface within Mercury. Simply arrange the image as required within the Mercury visualiser, then select **File** from the top-level menu, click on **Save As…** from the pull-down list and choose the POV-Ray File (**.pov*) as the file format. The dialogue will then allow you to name the output file and choose where to save it within your directory structure.
- When a POV-Ray file is produced using Mercury, the program will also generate a file called *ccdc_macro_overrides.inc* in the same directory unless one is already there (N.B. this also happens when using the *POV-Ray Image* dialogue). You can customise your rendered image by adding modified macros (object descriptions) to this file. The new macros will be automatically picked up when you render the image.
- POV-Ray can be used to flexibly generate movies based on a structural view as described in the previous section. For more information about how this is achieved, see the *.pov.ini* file which is saved to the same directory as the *.pov* file.
- For more help on editing your POV-Ray files or adding further advanced features to the rendered image, such as different backgrounds, textures or atmospheric effects, see the POV-Ray online documentation (*http://www.povray.org/documentation/3.7.0/*).
- Current limitations of the POV-Ray output from Mercury:
    - Tests of the image rendering have only been performed using POV-Ray version 3.7, so we cannot guarantee that use with different versions will produce the correct results.
    - Text labels created in Mercury sometimes have incorrect positions with respect to the camera when rendered using POV-Ray. If atom labels are obscured by the atoms, then please try using the Move Labels picking mode in Mercury to reposition the labels.
    - Gradient backgrounds are not implemented - one of the four colours used in Mercury (bottom left corner) is currently applied as a single colour background in POV-Ray
    - Line stippling is not correctly implemented in certain situations - delocalised bonds and measured distances may not have the same sized dashes and gaps as shown in Mercury.
    - Z-clipping and depth cueing are not implemented in POV-Ray output

17.2 Saving Structures in Mercury Format

- Mercury has its own file format (filename extension *.mryx*), which will store complete information about the structure currently on display, including not only the 3D atomic coordinates but also display properties, packing and slicing ranges, nonbonded contact definitions and networks, etc.
- To save the current structure, hit **File** in the top-level menu, **Save As...** in the next menu, the choose **Mercury binary crystal files (*.mryx)** in the *File Format* dialogue box and hit **OK**. This opens the *Save As* dialogue box which allows you to specify the name and location of the file to be created.
- An *.mryx* file can only be read back into Mercury, not into other programs.

- The *.mryx* file is binary, so cannot be read using a text editor.

17.3 Saving Structures in External (3rd Party) Formats

- In addition to CCDC-specific file formats, Mercury can write structures in the following external file formats:
  - *cif* (see cif Format).
  - *mol2* (see mol2 Format).
  - *pdb* (see pdb Format).
  - *res* (see res Format).
  - *xyz (see xyz Format*).
  - *PDF*.
- To save the current structure in one of these formats, hit **File** in the top-level menu, **Save As...** in the next menu, then choose the required format in the *File Format* dialogue box and hit **OK**. This opens the *Save As* dialogue box which allows you to specify the name and location of the file to be created.
- In the case of *mol2* and *pdb* formats, all molecules currently on display are included in the output files. Thus, these formats are useful for transferring assemblies of molecules from Mercury into other programs. However, display properties (e.g. atom colours) are not retained.
- Only the atoms of the crystallographic asymmetric unit are written out when the *res* format is used.

17.4 Saving Mercury Displays as 3D Printing Files

17.4.1 Overview

- Generation of 3D printed structural models is a very valuable additional option to consider for those communicating structural science. The communication of science is frequently seen to be challenging and this is particularly true for structural sciences, where 3D concepts that are difficult to describe in words are important, like conformations and interactions. The right type of visualisation or model can therefore significantly aid structural science communication in research and education.

17.4.2 Output of 3D Printing Files

- Model files for use in 3D printing can be generated from Mercury based on any structural view displayed in the visualiser. Simply arrange the display as required within the Mercury visualiser, then select **File** from the top-level menu and click on **Print in 3D…** from the drop-down list.
- The *3D Printing* control window will be displayed:

- The following options are available:
  - *File format*: The model file can be generated in STL (monochrome) or VRML (colour) format. This choice of file formats should allow printing on most current 3D printing devices. If a choice is available, then we recommend VRML, as this format captures colour and generally results in a smaller file.
  - *Scale (mm/Å)*: Use this to control the size of the model generated. The default scale setting (10.0 mm/Å) should work well for printing single molecules or small numbers of organic molecules. Note that if you significantly increase or decrease the scale, you may want to vary the radii of the atoms and bonds accordingly.
  - *Generate Support Framework*: An optional plinth and supporting columns can be added to the model file generated – this can be helpful for producing a free-standing model that will sit easily on a flat surface. The plinth will be in the same orientation as that of a mean plane calculated for the current structural representation. One column will be added close to the centroid for each molecule, options are provided for controlling the plinth thickness and column radii. If the cell edges are printed then supports for these will also be generated. The support structure can be visualised in any viewer for 3D printable model files.
  - *Output directory*: The output directory path for the generated model files will default to a writable location, but you can also choose to change this. As files are generated, output file names will be automatically created and displayed in the dialogue.
- We strongly recommend use of the specially designed “3D Print” style in Mercury for use when generating model files for 3D printing – this will increase the default radii for cylinders (including both bonds and intermolecular contacts) to give you a good starting point for a successful 3D print. It is also worth referring to the following section - Advice for 3D Printing File Output.

17.4.3 Advice for 3D Printing File Output

- The strength and robustness of 3D printed models varies significantly with the material type and the 3D printer – we recommend that you look up the minimum unsupported wall thickness related to your printer (or printing service) and material type and bear this in mind when generating your model file.
- When the model file is generated the dialogue will update to give an indication of the overall size, thinnest cylinder and smallest sphere diameters, which can be compared against the limits of the 3D printer. Atom and bond radii and the overall model scale can thereby be adjusted to get a suitable model.
- Objects with effectively zero thickness, such as wireframe bonds, calculated planes and any displayed surfaces (such as void surfaces or Full Interaction Maps), will not be represented in the generated model file.
- Any contacts or bond types represented in the visualiser as dashed, solid cylinders (including hydrogen-bonds, delocalised and pi bonds) will be converted into solid cylinders in the 3D model, with the dashes indicated using colour striping. The number of cylinder sections will scale according to the cylinder thickness
- Individual molecules that are not connected by bonds, intermolecular contacts or unit cell edges will be printed as separate, disconnected objects; if this is a problem for your 3D printer or printing service, then you will need to manually add some contacts or bonds to connect the molecules, or use the option to generate a supporting framework
- It is only the 3D objects and styles that are represented in the output file. The orientation of the 3D display is not used when generating the orientation of the model file. It is assumed that the 3D printer itself will orient the model to produce the best output

17.5 Copying Images to the Clipboard

- The display image can be copied to the clipboard by hitting **Edit** in the top-level menu and **Copy** from the resulting pull-down menu.
- The contents of the clipboard (i.e. the copied image from the display area) can then be pasted into other programs.

18 Data Analysis

18.1 Purpose of the Data Analysis Module

- The data-analysis module enables you to analyse the results of ConQuest substructure searches when you have defined some parameters (e.g. bond lengths, angles) as part of the search query. The values of these parameters in the hits can be tabulated and explored.
- Although mainly intended for the analysis of ConQuest searches, the module can be used to investigate other data sets, which may be read in from *.csv* (comma-separated value) or *.tsv* (tab-separated value) files.
- Data can be explored interactively with a spreadsheet linked to the molecular visualiser.
- Data can be plotted, e.g. as histograms and scatter plots.
- Statistical options are available, including calculation of descriptive statistics (mean, standard deviation, etc.), correlation matrices, and hypothesis (significance) tests.
- Any plots and tables you produce are linked to one another and to the spreadsheet. For example, by selecting a point in a scatter plot, you cause the appropriate row of the spreadsheet to be selected.

18.2 Definition of Data Analysis Terms

This section defines some terms that will be used throughout this section on data analysis: *data set; row* (also known as *entry*); *column* (also known as *descriptor*); *current entry; selection*; *current selection*; *named selection*; *data view*; *visible; hidden; locked*; *unlocked*.

18.2.1 Data-Analysis Terms: Data Set

- A *data set* is the data pertaining to one experiment.
- Usually, an experiment will be a ConQuest substructure search with geometric parameters defined.
- However, data sets can arise from other experiments, i.e. any investigation which produced data that can be read in from a comma- or tab-separated value file (in other words, data in tabular form).

18.2.2 Data-Analysis Terms: Row, Column, Entry, Descriptor

- The data to be analysed are held in a spreadsheet, like the one below:

Identifier	NAME	Query	Fragment	ANG1	ANG2	ANO1	DIST1	LAB1	VDW1
search1|ABEFIX|0	ABEFIX	1	1	171.4080	109.1920	17	2.1480	Cl1	1.7500
search1|ABILIG|1	ABILIG	1	1	152.9190	109.4650	53	2.7330	I1	1.9800
search1|ACBUET|2	ACBUET	1	1	149.9960	101.8860	17	2.2320	Cl3	1.7500
search1|ACBUOL|3	ACBUOL	1	1	173.8880	106.8160	17	2.1970	Cl1	1.7500
search1|ACICIY|4	ACICIY	1	1	176.4020	109.4610	17	2.1140	Cl1	1.7500
search1|ACMBPN|5	ACMBPN	1	1	142.8470	104.8860	35	2.4770	Br1	1.8500
search1|ACONTN10|6	ACONTN10	1	1	149.5970	100.2730	53	2.5700	I1	1.9800
search1|ADEHAT|7	ADEHAT	1	1	165.9490	112.4500	17	2.1570	Cl1	1.7500
search1|ADOSHC|8	ADOSHC	1	1	168.4850	109.5010	17	2.4360	Cl1	1.7500
search1|ADOSHC|9	ADOSHC	1	2	167.6400	107.9630	17	2.3410	Cl1	1.7500
search1|ADRENC|10	ADRENC	1	1	147.0150	111.4230	17	2.4710	Cl1	1.7500
search1|ADRENC|11	ADRENC	1	2	170.9940	110.1170	17	2.2260	Cl1	1.7500
search1|ADUXUT|12	ADUXUT	1	1	160.0420	109.4920	35	2.4800	Br2	1.8500
search1|AFMSCY|13	AFMSCY	1	1	140.4280	97.1970	17	2.4150	Cl1	1.7500
search1|AFMSCY|14	AFMSCY	1	2	172.7180	110.9470	17	2.2920	Cl1	1.7500
search1|AGEYAN|15	AGEYAN	1	1	169.2080	109.4820	17	2.3290	Cl2	1.7500
search1|AGEYAN|16	AGEYAN	1	2	172.7360	109.4600	17	2.2100	Cl2	1.7500
search1|AGEYER|17	AGEYER	1	1	165.1590	109.4410	17	2.3120	Cl1	1.7500
search1|AGEYER|18	AGEYER	1	2	163.9110	105.5030	17	2.3380	Cl1	1.7500
search1|AHGLCM|19	AHGLCM	1	1	170.8300	111.7510	17	2.2350	Cl1	1.7500
search1|AHMHPC|20	AHMHPC	1	1	171.0490	105.6930	17	2.2810	Cl1	1.7500
search1|AHUHER|21	AHUHER	1	1	159.2620	104.9520	17	1.6940	Cl1	1.7500
search1|AJEVIV|22	AJEVIV	1	1	173.6860	105.0350	17	2.2610	Cl1	1.7500
search1|AJUROM|23	AJUROM	1	1	177.4300	112.0010	17	2.3200	Cl1	1.7500
search1|AJUROM|24	AJUROM	1	2	178.3660	108.2970	17	2.2390	Cl1	1.7500

- A spreadsheet may contain one or more data sets (see Data-Analysis Terms: Data Set).
- Each *row* of the spreadsheet contains data for one *entry*. If you are analysing results from a ConQuest search, an entry will be the data pertaining to one hit.
- Each *column* of the spreadsheet contains values for one *descriptor*. The first descriptor is always a unique identifier. When analysing a ConQuest search, other columns contain the various parameters you defined for your ConQuest search query.

18.2.3 Data-Analysis Terms: Current Entry

- At any given time, exactly one entry can be the *current entry*.
- In the spreadsheet, the current entry is shown on a blue or light grey background, depending on whether or not the spreadsheet is in focus (i.e. being used).
- When an entry is made the current entry, it is loaded into the Mercury visualiser (assuming the data being analysed relate to a CSD search) and its position is marked on any histograms, scatter plots, etc., that have been created.

18.2.4 Data-Analysis Terms: Selection, Current Selection, Named Selection

- A *selection* is a subset of the entries in the data analysis (in other words, of the rows in the spreadsheet).
- There are two types of selections: the *current selection* and *named selections*.

- At any point in the data analysis, entries (rows) in the spreadsheet that are shown on a grey background belong to the current selection.
- The current selection tends to be changed frequently during a data analysis. It can be set in many ways, e.g. by selecting points on a scatter plot or bars in a histogram.
- At any point, the current selection can be given a name, whereupon it becomes a permanent, named selection that remains constant throughout the remainder of the analysis and can be recalled at any time.

18.2.5 Data-Analysis Terms: Data Views

- Data may be viewed as plots or tables. Different types of plots include histograms, scatter plots and heat plots. Different types of tables include correlation matrices and summaries of descriptive statistics.
- Spreadsheets, plots and tables are generically termed *data views*.

18.2.6 Data-Analysis Terms: Visible, Hidden

- At any given time, each entry in a data view (see Data-Analysis Terms: Data Views) is *visible* or *hidden*.
- For example, note that rows 2 to 5 of the following spreadsheet are highlighted, i.e. are in the current selection (see Data-Analysis Terms: Selection, Current Selection, Named Selection):

Identifier	NAME	Query	Fragment	ANG1	ANG2	ANO1	DIST1	LAB1	VDW1
search1|ABEFIX|0	ABEFIX	1	1	171.4080	109.1920	17	2.1480	Cl1	1.7500
search1|ABILIG|1	ABILIG	1	1	152.9190	109.4650	53	2.7330	I1	1.9800
search1|ACBUET|2	ACBUET	1	1	149.9960	101.8860	17	2.2320	Cl3	1.7500
search1|ACBUOL|3	ACBUOL	1	1	173.8880	106.8160	17	2.1970	Cl1	1.7500
search1|ACICIY|4	ACICIY	1	1	176.4020	109.4610	17	2.1140	Cl1	1.7500
search1|ACMBPN|5	ACMBPN	1	1	142.8470	104.8860	35	2.4770	Br1	1.8500
search1|ACONTN10|6	ACONTN10	1	1	149.5970	100.2730	53	2.5700	I1	1.9800
search1|ADEHAT|7	ADEHAT	1	1	165.9490	112.4500	17	2.1570	Cl1	1.7500
search1|ADOSHC|8	ADOSHC	1	1	168.4850	109.5010	17	2.4360	Cl1	1.7500
search1|ADOSHC|9	ADOSHC	1	2	167.6400	107.9630	17	2.3410	Cl1	1.7500
search1|ADRENC|10	ADRENC	1	1	147.0150	111.4230	17	2.4710	Cl1	1.7500
search1|ADRENC|11	ADRENC	1	2	170.9940	110.1170	17	2.2260	Cl1	1.7500
search1|ADUXUT|12	ADUXUT	1	1	160.0420	109.4920	35	2.4800	Br2	1.8500
search1|AFMSCY|13	AFMSCY	1	1	140.4280	97.1970	17	2.4150	Cl1	1.7500
search1|AFMSCY|14	AFMSCY	1	2	172.7180	110.9470	17	2.2920	Cl1	1.7500
search1|AGEYAN|15	AGEYAN	1	1	169.2080	109.4820	17	2.3290	Cl2	1.7500
search1|AGEYAN|16	AGEYAN	1	2	172.7360	109.4600	17	2.2100	Cl2	1.7500
search1|AGEYER|17	AGEYER	1	1	165.1590	109.4410	17	2.3120	Cl1	1.7500
search1|AGEYER|18	AGEYER	1	2	163.9110	105.5030	17	2.3380	Cl1	1.7500
search1|AHGLCM|19	AHGLCM	1	1	170.8300	111.7510	17	2.2350	Cl1	1.7500
search1|AHMHPC|20	AHMHPC	1	1	171.0490	105.6930	17	2.2810	Cl1	1.7500
search1|AHUHER|21	AHUHER	1	1	159.2620	104.9520	17	1.6940	Cl1	1.7500
search1|AJEVIV|22	AJEVIV	1	1	173.6860	105.0350	17	2.2610	Cl1	1.7500

- If we now hide all entries not in the current selection, only rows 2 to 5 will be visible, the others being hidden:

- Visibility is very important, because if we now create a plot from the spreadsheet, it will show only those entries visible in the spreadsheet (e.g. a scatter plot would only have four points). Similarly, if we calculate a statistical table, the statistical results will be based only on the visible entries.

18.2.7 Data-Analysis Terms: Locked, Unlocked

- At any point during a data analysis, a data view (see Data-Analysis Terms: Data Views) is either *locked* or *unlocked*. If, and only if, it is unlocked, it will automatically be updated when the visibility of entries (see Data-Analysis Terms: Visible, Hidden) is changed in other data views. For example, the scatter plot below was produced when four entries were visible.

- If you do something in the spreadsheet which causes one of the entries to become hidden, the scatter plot will automatically change its appearance provided it is unlocked:

- If it is locked, however, it will remain unchanged until you manually update it.

18.3 Rules When Using the Data Analysis Module

- At any given time, there can be only one current entry (see Data-Analysis Terms: Current Entry).
- The current selection (see Data-Analysis Terms: Selection, Current Selection, Named Selection) is universal, i.e. applies to all data views. Therefore, if you change the current selection in one data view - e.g. by selecting some bars in a histogram - it simultaneously changes in all other data views.
- You can construct a data view (see Data-Analysis Terms: Data Views) from any other data view, e.g. a histogram from a correlation matrix, or a scatter plot from a spreadsheet.
- When you construct one data view from another, the new data view will show only those entries that are visible in the starting data view (or, in the case of a statistical table such as a correlation matrix, only those entries that were included in the calculation of that table).
- If a data view is locked (see Data-Analysis Terms: Locked, Unlocked), the entries that are visible in it (or, in the case of a statistical table, the entries on which the table is based) will remain unchanged if you alter the visibility of entries in other data views.

- If a data view is unlocked, the entries that are visible in it (or, in the case of a statistical table, the entries on which the table is based) will be those that were visible after the last visibility change you made in any data view. In other words, every time you make a visibility change in a data view, it will simultaneously be made in all unlocked data views.

18.4 Starting a Data Analysis

- The data-analysis window can be opened by reading a *.c2m* file (the file produced by exporting data from a ConQuest search) or a *.csv* (comma-separated value) or *.tsv* (tab-separated value) file.
- This may be done by dragging the file into the Mercury window or onto a Mercury desk-top icon.
- Alternatively, hit **File** in the Mercury top-level menu, **Open...** in the next menu, then type in the name of the required file or use the file browser to locate and select it. Recently-used files can be read in by picking **File** in the Mercury top-level menu, then **Recent Files** in the next menu, then selecting the required file name from the resulting pull-down list.
- ConQuest can directly open Mercury and the data-analysis window with its **Analyse Data** option (refer to the ConQuest User Guide for further information).
- If the data are read in from a *.csv* or *.tsv* file, the first line of the file must contain descriptor names, the first of which must be *NAME* (indicating that the first data item on each line will be an entry identifier).
- Any further *.c2m, .csv* or *.tsv* files that are read in will add to the data already in the data-analysis window (rather than overwriting it).

18.5 Reloading a Saved Data Analysis

- A data analysis may be saved as an *.sqlite* file (see Saving Data Analyses). This will save not only the data but also any named selections you have made.
- A data analysis reloaded from an *.sqlite* file will overwrite any data already in the data-analysis window.
- To reload a saved data analysis, drag the *.sqlite* file into the Mercury window or onto a Mercury desk-top icon.
- Alternatively, hit **File** in the Mercury top-level menu, **Open...** in the next menu, then type in the name of the required file or use the file browser to locate and select it. Recently-used files can be read in by picking **File** in the Mercury top-level menu, then **Recent Files** in the next menu, then selecting the required file name from the resulting pull-down list.
- If the data-analysis window is already open, hit **File** in the data-analysis top-level menu, **Load database...** in the next menu, then specify the required file name.

18.6 The Data-Analysis Window and Menus

- Here is an example of the data-analysis window:

Identifier	NAME	Query	Fragment	ANG1	ANG2	ANO1	DIST1	LAB1	VDW1
search1_1|ABEFIX|0	ABEFIX	1	1	171.4080	109.1920	17	2.1480	Cl1	1.7500
search1_1|ABILIG|1	ABILIG	1	1	152.9190	109.4650	53	2.7330	I1	1.9800
search1_1|ACBUET|2	ACBUET	1	1	149.9960	101.8860	17	2.2320	Cl3	1.7500
search1_1|ACBUOL|3	ACBUOL	1	1	173.8880	106.8160	17	2.1970	Cl1	1.7500
search1_1|ACICIY|4	ACICIY	1	1	176.4020	109.4610	17	2.1140	Cl1	1.7500
search1_1|ACMBPN|5	ACMBPN	1	1	142.8470	104.8860	35	2.4770	Br1	1.8500
search1_1|ACONTN10|6	ACONTN10	1	1	149.5970	100.2730	53	2.5700	I1	1.9800
search1_1|ADEHAT|7	ADEHAT	1	1	165.9490	112.4500	17	2.1570	Cl1	1.7500
search1_1|ADOSHC|8	ADOSHC	1	1	168.4850	109.5010	17	2.4360	Cl1	1.7500
search1_1|ADOSHC|9	ADOSHC	1	2	167.6400	107.9630	17	2.3410	Cl1	1.7500
search1_1|ADRENC|10	ADRENC	1	1	147.0150	111.4230	17	2.4710	Cl1	1.7500
search1_1|ADRENC|11	ADRENC	1	2	170.9940	110.1170	17	2.2260	Cl1	1.7500
search1_1|ADUXUT|12	ADUXUT	1	1	160.0420	109.4920	35	2.4800	Br2	1.8500
search1_1|AFMSCY|13	AFMSCY	1	1	140.4280	97.1970	17	2.4150	Cl1	1.7500
search1_1|AFMSCY|14	AFMSCY	1	2	172.7180	110.9470	17	2.2920	Cl1	1.7500
search1_1|AGEYAN|15	AGEYAN	1	1	169.2080	109.4820	17	2.3290	Cl2	1.7500
search1_1|AGEYAN|16	AGEYAN	1	2	172.7360	109.4600	17	2.2100	Cl2	1.7500
search1_1|AGEYER|17	AGEYER	1	1	165.1590	109.4410	17	2.3120	Cl1	1.7500
search1_1|AGEYER|18	AGEYER	1	2	163.9110	105.5030	17	2.3380	Cl1	1.7500
search1_1|AHGLCM|19	AHGLCM	1	1	170.8300	111.7510	17	2.2350	Cl1	1.7500
search1_1|AHMHPC|20	AHMHPC	1	1	171.0490	105.6930	17	2.2810	Cl1	1.7500
search1_1|AHUHER|21	AHUHER	1	1	159.2620	104.9520	17	1.6940	Cl1	1.7500
search1_1|AJEVIV|22	AJEVIV	1	1	173.6860	105.0350	17	2.2610	Cl1	1.7500
search1_1|AJUROM|23	AJUROM	1	1	177.4300	112.0010	17	2.3200	Cl1	1.7500
search1_1|AJUROM|24	AJUROM	1	2	178.3660	108.2970	17	2.2390	Cl1	1.7500
search1_1|AJUROM|25	AJUROM	1	3	162.9600	115.5220	17	2.3270	Cl1	1.7500
search1_1|AJUROM|26	AJUROM	1	4	167.2110	106.1500	17	2.2130	Cl1	1.7500

- The main body of the window shows a data view, in this case a spreadsheet. Any data view can be detached into a separate window by undocking (see Docking and Undocking Plots and Other Data Views).
- Working from the top down, the other components of the window are:
  - The *data-analysis top-level menu* (**File Tools Options Help**).
  - *Data-view tabs*. These allow you to move between the various data views you have created and left docked into the main window. In the above example, the user has a spreadsheet displayed but has also created a scatter plot and correlation matrix, either of which could be displayed by clicking on the appropriate tab.
  - The *data-view title bar*. The first of the two icons at the extreme right of this bar (showing two overlapping squares) enables the data view to be undocked into a separate window. The right-most icon (the cross) deletes the data view.
  - The *data-view menu*. This varies slightly according to what type of data view is on display but usually contains all of the options **File Descriptors Display Selection Plots Statistics**.

- To the right of the data-view menu is a button which controls whether the data view shown in the main body of the window is locked or unlocked; the button shows a closed or open padlock, respectively. The adjacent button (showing two curly arrows) is used to update locked data views (see Locking and Unlocking Spreadsheets, Plots and Tables).

18.7 Restoring the Data-Analysis Window

- The data-analysis window may be closed in the usual way, by clicking the cross icon at the top right of the window.
- To restore it, select **CSD-System** from the Mercury top-level menu, followed by **Data Analysis Module...** from the resulting drop-down menu.
- The window will be restored in the same state as it was when closed.

18.8 Docking and Undocking Plots and Other Data Views

- A data view may be undocked (i.e. detached into a separate window) in several ways:
  - Double-click in the data-view title bar, which looks something like this:

  - Click on the button showing two overlapping squares at the right of this bar.
  - Click on the data-view title bar and, keeping the mouse button depressed, drag the data view to any position where the main window is not shaded in blue.
- To re-dock an undocked data-view window, use any of the following options:
  - Double-click on the window title bar at the top of the undocked window (i.e. the blue bar).
  - Click on the window title bar and, keeping the mouse button depressed, drag it towards the main data-view window until the latter becomes shaded in blue.
- You can control whether new data views are, by default, docked or undocked by selecting **Options** from the data-analysis top-level menu. In the resulting pull-down

menu, **Dock windows** will be ticked or unticked depending on whether the current default is to dock or undock. Click on **Dock windows** to reverse the setting.

18.9 Simultaneously Viewing All Spreadsheets, Plots and Tables

- All data views may be displayed simultaneously in the data-analysis window by selecting **Options** from the data-analysis top-level menu followed by **Show all**.
- The resulting display will look something like this:

- To revert to the normal tabbed display, select **Options** from the data-analysis top-level menu, followed by **Tab all**.

18.10 Using Drag and Drop between Data Views

- If any data view is being displayed at the same time as a spreadsheet or statistics table (i.e. both are visible to you; for example, in separate windows), a convenient drag and drop mechanism may be used to change the nature of the data view.
- For example, if a heat plot and a spreadsheet are on display simultaneously, a new descriptor can be plotted along the x (or y) axis of the heat plot by clicking on the required descriptor name at the head of its spreadsheet column, dragging it onto the relevant axis (i.e. moving the cursor to the axis keeping the mouse button depressed) and dropping it (releasing the mouse button).
- Dropping a descriptor onto a plot changes the plot axis (i.e. the new descriptor is plotted, replacing the old). For histograms, there is only axis that can be changed.

For scatter plots and heat plots, the axis closest to the drop position is changed. In the case of scatter plots, dropping onto the top right corner of the plot changes the third (colour) axis (see Scatter Plot Features).

- Dropping the descriptor onto a statistics table adds the descriptor to the table.

18.11 Locking and Unlocking Spreadsheets, Plots and Tables

- An unlocked data view will automatically change if entries are hidden or made visible in any other data view (i.e. the same entries will be hidden or made visible in the unlocked view). Locked data views do not change.
- If a data view is shown with a closed-padlock button in the data-view title bar, it is locked:

- To unlock it, click on the padlock icon. The button will now show an open padlock:

- Repeat the procedure to lock it again.
- Upon being unlocked, the data view will change so that the entries that are visible in it (or, in the case of a statistical table, the entries on which the table is based) are those that were visible after the last visibility change you made in any data view.
- For a locked data view, clicking on the button displaying curly arrows (next to the padlock button) is equivalent to unlocking the data view and then locking it again (so, effectively updating the data view but leaving it locked). Clicking this button will have no effect if the data view is already unlocked.

18.12 Guidelines on Using the Mouse to Interact with Data Views

- Clicking with the right mouse button on a data view will often produce a menu of context-dependent options.
- Clicking (or clicking and dragging) with the left button can produce a variety of effects, depending on:
  - whether the keyboard *Ctrl* or *Shift* key (or neither) is held down;
  - in the case of plots, the mouse mode that has been set (see Plot Mouse Modes).
- The following bullet points summarise the effects of clicking with the left mouse button on data views (spreadsheets, plots and some types of statistics tables). In the case of plots, it is assumed that the mouse mode has been set to **Select**.
  - Clicking on an element of a data view which necessarily corresponds to exactly one entry (i.e. a row of a spreadsheet or a point on a scatter plot or polar scatter plot) makes that entry the current entry (see Data-Analysis Terms: Current Entry) but has no other effect. In particular, it does not affect whether the entry is in the current selection (see Data-Analysis Terms: Selection, Current Selection, Named Selection).
  - Clicking on an element of a data view which may correspond to more than one data entry (i.e. a bar of a histogram, a cell of a heat plot, or a cell of a

statistics table which represents a count of entries, such as the number of outliers) sets the current selection to contain those entries and nothing else.

- Clicking and dragging over several elements of a data view (bars of a histogram, a region of a scatter plot containing points) sets the current selection to contain the entries represented by those elements, and nothing else.
- Clicking with the *Ctrl* key depressed on an element of a data view that represents one or more entries (spreadsheet row, histogram bar, scatter plot point, etc.) toggles (i.e. reverses) the selection status of all those entries. For example, if the bar of a histogram is *Ctrl*-clicked, all the entries in the bar that are not in the current selection become selected, and vice versa.
- Clicking and dragging with the *Ctrl* key depressed over several elements of a data view toggles the selection status of all entries represented by those elements.
- Clicking with the *Shift* key depressed on an element of a data view that represents one or more entries (spreadsheet row, histogram bar, scatter plot point, etc.) adds those entries to the current selection.
- Clicking and dragging with the *Shift* key depressed over several elements of a data view adds to the current selection all the entries represented by those elements.

18.13 Data-Analysis Spreadsheet Operations

18.13.1 Overview of Spreadsheet Functionality

Spreadsheets in the data-analysis module tabulate the descriptor data of the entries in the data sets being analysed. A range of options are available, including:

- Finding entries by identifier (see Finding Spreadsheet Entries).
- Visualising entries in the molecule display area (see Visualising Spreadsheet Entries) and browsing through them (see Browsing Spreadsheet Entries with the Keyboard).
- Hiding and unhiding rows (see Hiding and Unhiding Spreadsheet Rows) and columns (see Hiding, Unhiding and Reordering Spreadsheet Columns).
- Sorting on (see Sorting on Spreadsheet Columns) and colouring by (see Colouring on Spreadsheet Columns) the values in a descriptor column.
- Combining columns (see Combining Spreadsheet Columns: Accounting for Topological Symmetry).
- Arranging entries in a hierarchical “tree” arrangement (see Grouping Entries by Descriptor Values: Tree Views).
- Adding rows to or removing them from the current selection (see Modifying the Current Selection from the Spreadsheet).
- Adding new descriptor columns (see Adding New Descriptors to a Data-Analysis).

18.13.2 Spreadsheet Contents

- The first column of a data-analysis spreadsheet always contains a unique identifier for each entry.

- The remainder of this section assumes that the spreadsheet was created by transferring into Mercury the results of a ConQuest substructure search:
    - The unique entry identifiers have the form *search-name|refcode|number*. The search name is the name of the search in the ConQuest session that produced the hit list, e.g. *search2*. The numbers run sequentially from 0 to *n-1*, where *n* is the number of hits from the search.
    - By default, the next three columns contain the refcode of the CSD structure containing the hit fragment, the query number (i.e. the ConQuest query that the hit fragment matched), and the fragment number (1 for the first hit fragment in the refcode, 2 for the second, etc.).
    - The remaining columns contain the parameters you defined in ConQuest when you created the search query (for example, any bond lengths and angles you defined for the search substructure).

18.13.3 Finding Spreadsheet Entries

- Entries whose identifiers contain a given string of characters may be found by typing the string in the *Find identifier* box at the top of the spreadsheet. You do not need to start at the beginning of the identifier, e.g. typing *BMHAL* would find *search1|ABMHAL|20*. The search is case sensitive.
- The **Find next** button can then be used to find further entries in the spreadsheet whose identifiers contain the search string.

18.13.4 Visualising Spreadsheet Entries

- Clicking on an entry in the spreadsheet with the left-hand mouse button will make it the current entry (see Data-Analysis Terms: Current Entry). This will cause it to be displayed in the visualiser with the hit fragment highlighted (assuming the spreadsheet refers to a ConQuest substructure search).
- If the visualiser **Multiple Structures** check box is ticked, successive clicks on different entries will cause them all to be loaded into the visualiser, but only the last one will be set to *Visible* in the **Structures** dialogue box (see Displaying Multiple Structures Simultaneously).
- When more than one entry is selected in the spreadsheet (see Making and Using Data Selections) the current selection will be available for display in the visualiser. These structures can be found under *Data analysis most recent selection* within the structure navigator (see Selecting and Displaying Structures: Using the Structure Navigator).

18.13.5 Browsing Spreadsheet Entries with the Keyboard

- Assuming spreadsheet entries refer to CSD structures, you can use the keyboard up and down arrows to browse through them.
- Hitting the keyboard down (or up) arrow switches the current entry (i.e. the entry being displayed in the visualiser) from one row to the next (or previous).
- If some rows of the spreadsheet are selected - i.e. in the current selection (see The Role of Selections in Data Analysis) - you can browse from one to the next (or previous) by hitting the keyboard down (or up) arrow while depressing the *Shift* key. If you use *Ctrl* rather than *Shift*, you will move to the next (or previous) row of the *opposite* selection status.

18.13.6 Hiding and Unhiding Spreadsheet Rows

- Hiding and unhiding spreadsheet rows is of critical importance when using the data analysis module because it controls which entries are included on plots and in statistical calculations. For example, if you hide some rows in the spreadsheet and then ask for a histogram of a particular descriptor, only the currently-visible rows will be included.
- Rows are hidden or unhidden by use of the **Selection** button in the spreadsheet data-view menu:
- To hide all rows in the current selection (i.e. rows that are highlighted in the spreadsheet), pick **Selection** followed by **Hide selected**. Alternatively, hit the keyboard *Delete* key.
- To hide all rows except those in the current selection, pick **Selection** followed by **Hide unselected**.
- To make all rows visible, pick **Selection** followed by **Show all**.
- To show all rows, including those in data sets that have been suppressed, you must first unsuppress those data sets, then proceed as in the preceding bullet point.
- To hide all rows except those in a named selection, first make all rows visible (see preceding two bullet points), then set the current selection to the named selection you require. To do this, pick **Selection**, followed by **Selection** in the next menu, then pick the required named selection from the resulting list. Now complete the operation by picking **Selection** followed by **Hide unselected.**

18.13.7 Hiding, Unhiding and Reordering Spreadsheet Columns

Columns (i.e. descriptors) can be hidden or unhidden in several ways:

- Right-click on the descriptor name at the head of the column and select **Hide** from the resulting pull-down menu.
- Pick **Descriptors** in the spreadsheet data-view menu, then select the data-set name (which will normally be the ConQuest search name) in the next menu. Click on a descriptor name in the resulting list to toggle the display of that descriptor column. A tick alongside a descriptor name in this list means that the column is already displayed.
- Pick **Display** in the spreadsheet data-view menu and **Configure...** in the next menu, which will open a window looking something like this:

- Descriptor columns in the left-hand box of this window are not currently displayed, columns in the right-hand box are.
- You can transfer a column from one box to the other by clicking on it and then hitting **Add>>** (transfers to right-hand box) or **<<Remove** (transfers to left-hand box).
- Hitting **OK** causes columns in the right-hand box (only) to be displayed. **Cancel** will leave the column display unchanged.

- The above window also allows columns to be reordered, because the order in which they are displayed in the spreadsheet will be the same as their order in the right-hand box. The position of a descriptor name in this box may be altered by selecting it and using the **Up** or **Down** button.

18.13.8 Sorting on Spreadsheet Columns

- The rows of a spreadsheet may be sorted on the values in any particular column by right-clicking on the descriptor name at the head of the column and selecting **Sort** from the resulting pull-down menu.
- Repeating this action will reverse the sort order.
- Alternatively, you can do a nested sort, i.e. sort first on one column, then (for those entries whose values in that column are equal) on a second, and so on.
  - Select the columns you want to sort on by clicking on the descriptor names at the column heads.
  - Select **Display** from the spreadsheet data-view menu, then **Sort...** from the next menu. This will open a window looking something like this:

- In the top white box, select from the pull-down list of descriptor names the one you want to sort on first, and specify whether to sort in ascending or descending order.
- Repeat this procedure on the remaining search levels (obviously, you have to pick a different descriptor at each level), then hit **OK**.

18.13.9 Combining Spreadsheet Columns: Accounting for Topological Symmetry

- Two or more descriptors can be combined into a single column. This can be useful when analysing geometrical data from fragments that contain topological symmetry.
- Topological symmetry occurs when there are parameters in a structure which are chemically equivalent in the query, but are usually geometrically different in the crystal structure (unless the topological symmetry of the search fragment is coincident with a crystallographic symmetry element). For example, a query comprising a phenyl ring has six chemically equivalent C-C-C angles, but these are normally geometrically independent in each crystal structure containing a phenyl ring.
- The correct way to handle this problem is to define all of the parameters in the query (e.g. all six C-C-C angles in a phenyl ring), giving rise to multiple columns of data (six in the case of the phenyl ring angles) in the results spreadsheet.
- Data in these columns are chemically equivalent and should be treated as a single distribution. In order to do this we need to combine these multiple descriptors and treat them as a single distribution for all plotting and analysis functions.
- Click on **Tools** in the data-analysis top-level menu, your *search name* in the drop-down menu, followed by **Topological symmetry** in the next menu.
- In the resulting window, select the descriptors you wish to treat as topologically symmetric.

- On hitting **Calculate**, the requested descriptors will be merged and displayed in a single column in a new data spreadsheet.

18.13.10 Colouring on Spreadsheet Columns

- Colouring the cells of columns by the values they contain is an easy way of spotting patterns in the data (for example, whether the values in two columns are correlated).
- Colours are assigned in a spectrum, ranging from dark blue for the lowest values in the column to red for the highest values.
- To colour each cell of a column on its value, right-click on the descriptor name at the head of the column and select **Colour** from the resulting pull-down menu. The column can be uncoloured by repeating the action.
- Alternatively, click on **Display** in the spreadsheet data-view menu, and then **Colours...** in the resulting pull-down menu. This opens a window looking something like this:

- Columns in the left-hand box of this window are not currently coloured, columns in the right-hand box are.
- You can transfer a column from one box to the other by clicking on it and then hitting **Add>>** (transfers to right-hand box) or **<<Remove** (transfers to left-hand box).
- Hitting **OK** causes all columns in the right-hand box (only) to be coloured. **Cancel** will leave the column display unchanged.

18.13.11 Grouping Entries by Descriptor Values: Tree Views

- If the values in a column are integers or text strings, the spreadsheet can be rearranged in a "tree" fashion, such that all entries with the same value in that column are grouped together. (Grouping on a non-integral numerical descriptor is less straightforward but can be done by use of the *Categorise* option (see Creating a Discrete Descriptor from a Continuous Descriptor; Categorising by Range).)
- For example, the spreadsheet below shows results from a ConQuest search for OH...Halide interactions. The column headed *ANO1* contains the atomic number of the halide atom, i.e. 9 (for F), 17 (for Cl), 35 (for Br), or 53 (for I).

Identifier	NAME	Query	Fragment	ANG1	ANG2	ANO1	DIST1	LAB1	VDW1
search1_2|BIVNAV|111	BIVNAV	1	1	177.7880	109.4900	17	2.3550	Cl1	1.7500
search1_2|BIVNAV|112	BIVNAV	1	2	159.2760	118.4930	17	2.4060	Cl2	1.7500
search1_2|BIVNAV|113	BIVNAV	1	3	162.1980	108.6590	17	2.3730	Cl1	1.7500
search1_2|BIVNAV10|114	BIVNAV10	1	1	177.7880	109.4900	17	2.3550	Cl1	1.7500
search1_2|BIVNAV10|115	BIVNAV10	1	2	159.2760	118.4930	17	2.4060	Cl2	1.7500
search1_2|BIVNAV10|116	BIVNAV10	1	3	162.1980	108.6590	17	2.3730	Cl1	1.7500
search1_2|BIZGAS|117	BIZGAS	1	1	165.7660	112.7180	17	2.1870	Cl1	1.7500
search1_2|BIZMUT|118	BIZMUT	1	1	179.1030	111.2090	17	2.2640	Cl2	1.7500
search1_2|BIZNAA|119	BIZNAA	1	1	176.9880	109.4590	35	2.4160	Br2	1.8500
search1_2|BLBCPI01|120	BLBCPI01	1	1	153.0870	116.5200	53	2.7040	I1	1.9800
search1_2|BOBZAT|121	BOBZAT	1	1	165.1730	110.4310	17	1.8350	Cl1	1.7500
search1_2|BOBZAT|122	BOBZAT	1	2	176.6140	121.1050	17	2.2100	Cl1	1.7500
search1_2|BOBZAT|123	BOBZAT	1	3	147.9940	116.3320	17	2.0450	Cl1	1.7500
search1_2|BOCTIW|124	BOCTIW	1	1	175.4610	109.3060	17	2.2360	Cl1	1.7500
search1_2|BODKEK|125	BODKEK	1	1	161.1260	103.2260	35	2.6000	Br1	1.8500
search1_2|BOHHIP|126	BOHHIP	1	1	162.7150	117.2290	17	1.9930	Cl1	1.7500
search1_2|BOHHIP|127	BOHHIP	1	2	163.5620	105.2880	17	2.1480	Cl1	1.7500
search1_2|BONFUG|128	BONFUG	1	1	156.7640	109.4360	53	2.6090	I1	1.9800
search1_2|BOQBUE|129	BOQBUE	1	1	169.2710	105.8490	53	2.7590	I1	1.9800
search1_2|BORROP|130	BORROP	1	1	173.9340	115.6400	35	2.4280	Br1	1.8500
search1_2|BOTLAY|131	BOTLAY	1	1	159.9850	113.7760	17	2.2860	Cl1	1.7500
search1_2|BOTPAB|132	BOTPAB	1	1	164.2700	114.6820	17	2.1160	Cl1	1.7500
search1_2|BOXANC|133	BOXANC	1	1	144.8590	114.0140	17	2.3240	Cl1	1.7500
search1_2|BOXVIU|134	BOXVIU	1	1	169.8380	106.0530	17	2.3290	Cl2	1.7500
search1_2|BOXVIU|135	BOXVIU	1	2	166.1400	109.8050	17	2.2260	Cl2	1.7500
search1_2|BUBYON|136	BUBYON	1	1	175.7420	102.5980	17	2.1360	Cl1	1.7500
search1_2|BUBYON|137	BUBYON	1	2	167.0190	105.6260	17	2.3700	Cl2	1.7500
search1_2|BUBYON|138	BUBYON	1	3	168.0590	111.1100	17	2.3440	Cl1	1.7500
search1_2|BUBYON|139	BUBYON	1	4	166.9360	98.0880	17	2.1680	Cl2	1.7500

- Right-clicking on the descriptor name at the head of the column, followed by selection of the **Group by** option in the resulting menu, changes the spreadsheet so that it looks like this:

- All the F containing hits are now grouped together in the same tree branch, with the Cl, Br, and I containing hits in other separate branches.
- Clicking on the + icons will expand the tree view so that the entry identifiers can be seen. Alternatively, click on the item to be expanded and hit the keyboard right arrow. To collapse the tree view again, use the - icon or keyboard left arrow.
- All the tree branches can be expanded simultaneously by selecting **Display** from the spreadsheet data-view menu, followed by **Groups** and then **Expand all**. They may be collapsed back down again by **Display**, **Groups**, **Collapse all**.
- Grouping can be turned off by hitting **Display** followed by **Groups** and **No grouping**.
- Entries can be grouped at several levels, e.g.
    - This is done by selecting **Display** followed by **Groups** and then **Group by...**, which opens a window looking something like this:

- Descriptor names may be moved from the left-hand to the right-hand box by selecting the name and hitting **Add>>** (the action can be reversed by use of **<<Remove**).
- Within the right-hand box, the position of a descriptor in the list can be altered by selecting it and using the **Up** or **Down** button.
- When the **OK** button is hit, the spreadsheet entries will be grouped by the descriptors in the right-hand box, in the order in which they appear in the box.

18.13.12 Copying Spreadsheets

- A new spreadsheet can be created from any data view, such as an existing spreadsheet or a histogram.
- To do this, hit **Statistics** in the data-view menu and **Spreadsheet** in the next menu.
- All chosen data (both hidden and unhidden) will be transferred to the new spreadsheet, but only the entries and descriptors visible in the data view will be visible in the new spreadsheet.

18.13.13 Restoring a Lost Spreadsheet

- If you accidentally delete your spreadsheet, it can be retrieved by selecting **Tools** from the data-analysis top-level menu followed by your *search name* in the next menu, followed by **Spreadsheet** in the next menu.
- The retrieved spreadsheet will contain all the data you choose to include but will not retain the state of the original spreadsheet, e.g. all rows will be visible even if some were hidden when the spreadsheet was closed.

18.14 Handling Multiple Data Sets

- Each new data set read into the same Mercury session (e.g. each new set of ConQuest search results) will be presented in its own separate data tab.
- You can move between the various data views you have created by clicking on the appropriate data view tab.

18.15 Adding New Descriptors to a Data-Analysis

Additional descriptors can be added to a data analysis (i.e. new columns added to the spreadsheet) in several ways:

- By applying arithmetical operations to existing descriptors (see Creating New Descriptors by Arithmetical Operations).
- By creating a text descriptor from an integer descriptor (see Creating a Text Descriptor from an Integer Descriptor).
- By creating categorical (discrete) descriptors from continuous (floating-point number) descriptors(see Creating a Discrete Descriptor from a Continuous Descriptor; Categorising by Range).
- If the data set is from a ConQuest search, by importing structure-specific information from the CSD (see Importing Additional Descriptors from the CSD).

18.15.1 Creating New Descriptors by Arithmetical Operations

- Pick **Tools** from the data-analysis top-level menu and **Calculator...** from the next menu.
- In the *Calculator* window, type in a name for the new descriptor in the *New descriptor* box.
- Construct the arithmetic expression you want by clicking on the calculator pad and descriptor names. The expression will build up in the white box at the bottom of the window as you proceed. e.g.

- If you are uncertain about the meaning of a calculator function, hover over it to see an explanatory message.
- Click on **Calculate** to create the new descriptor; it will immediately appear as the far-right column in the spreadsheet.
- The calculator is designed to ensure the correct syntax when the mouse is used to select descriptors and operators for most cases. For example, the cursor will move outside descriptor brackets (such as *Min()*) when all necessary arguments have been received. The use of operators such as *Least()* and *Greatest()* will insert commas in normal use. However, it will still be possible to click on operators and descriptors in

the wrong order. If you make a mistake, it will result in a syntax-error message when **Calculate** is hit. You can either correct the expression manually (you can type into the white box and use cut-and-paste) or hit the **Clear** button and start again.

- Some expressions may be declared invalid because they do not correspond to a one-to-one transformation. For example, the expression *Min(search1_ANG1)* is invalid because it simply calculates the minimum value of this descriptor, which is only one number. However, *search1_ANG1 - Min(search_ANG1)* is valid because it will subtract the column minimum from each column value in turn, producing a complete new column of numbers.
- Some functions are invalid for certain descriptor values, e.g. you cannot take the logarithm of a negative number. In this situation, the resulting cell in the new column will be empty.

18.15.2 Creating a Text Descriptor from an Integer Descriptor

- A text descriptor can be created from an integer descriptor. For example, if one of your descriptor columns contains atomic numbers, you might wish to create a column containing the corresponding element symbols.
- Right-click on the descriptor name at the head of the integer-descriptor column and pick **Categorise** from the pull-down menu.
- In the resulting window, enter the descriptor text (in this case the element symbols) for each of the integer values given:

- A descriptor name can be specified by typing into the *Descriptor name* text box. On clicking **OK** a new column will be added to the spreadsheet.

18.15.3 Creating a Discrete Descriptor from a Continuous Descriptor; Categorising by Range

- A categorical (discrete-valued, or text) descriptor can be created from a continuous (floating-point number) descriptor by dividing the latter into ranges.
- In other words, you must divide the observed values of the continuous descriptor into ranges and associate each range with a discrete value, or descriptive text. This may then be used, for example, to group spreadsheet rows together in a tree structure, so that all entries falling into a particular range are in the same tree branch (see Grouping Entries by Descriptor Values: Tree Views).
- Right-click on the descriptor name at the head of the continuous-descriptor column and pick **Categorise** from the pull-down menu.
- In the resulting window, click on the **Add** button. You can then define your first category and its associated range.
- Define the required range for the category by clicking on the green slider arrows and dragging them to the required points, e.g.

- As you move a slider arrow, the descriptor value to which it is pointing will be shown in the nearby white box. It is possible to type the required descriptor value directly into the box.
- Once the upper and lower range limits have been defined enter a category name by typing into the corresponding text box.
- To define subsequent category ranges click on the **Add** button and repeat the steps above.
- The *Default* category will be applied to the remainder of the continuous descriptor range.
- A descriptor name can be specified by typing into the *Descriptor name* text box. On clicking **OK** a new column will be added to the spreadsheet.

18.15.4 Importing Additional Descriptors from the CSD

- If a data set was created from a ConQuest search, so that each entry corresponds to a CSD structure, new descriptor columns can be constructed by importing data from the CSD.
- Click on **Tools** in the data-analysis top-level menu, your *search name* in the drop-down menu, followed by **CSD data...** in the next menu.
- In the resulting window, CSD data items are grouped into categories. Click on the + icons to display the available items. Select the ones you want, keeping the *Control* key depressed to make multiple selections or the *Shift* key to select a range.
- On hitting **OK**, the requested items will be imported and displayed in new columns at the right of the spreadsheet. The importing step may take a minute or two.

18.16 Making and Using Data Selections

18.16.1 The Role of Selections in Data Analysis

- Creating selections (i.e. subsets of entries) enables you to explore data trends. For example, if the hit list contains both amide and thioamide fragments, they could be separated into two different selections which could then be compared.
- Selections fall into two types: the *current selection*, which always exists and typically undergoes many changes during the course of an analysis, and *named selections,* which, once created, are unchanging throughout the data analysis. At any point, you can set the current selection to contain the same entries as a specified named selection (see Recalling a Named Selection).
- Probably the most important use of the current selection is to control visibility because, whenever a plot or statistics table is calculated, it will be based only on entries that are currently visible. Entries are usually hidden by use of the current selection, e.g. by hiding everything outside that selection (see Hiding and Unhiding Spreadsheet Rows).
- Entries (rows) in a spreadsheet that belong to the current selection are shown on a grey background.
- Entries belonging to the current selection are usually assigned a special symbol in scatter plots, e.g. shown as stars rather than circles. Similarly, each bar of a histogram can be divided into two parts, showing how many entries in the bar belong to the current selection and how many do not. Viewing displays like this can be an effective way of spotting systematic differences between entries belonging to the current selection and those outside it.
- This may be addressed more quantitatively by statistical hypothesis testing, e.g. for any given descriptor, an F-test can be applied to test whether its mean value in the current selection is significantly different from that of entries outside the selection (see Testing Whether Two Means are Significantly Different).

18.16.2 Summary of Methods for Setting the Current Selection

The current selection can be set in several ways, including:

- Use of the **Selections** button in any data-view menu (see Setting the Current Selection by Use of the Menu).

- Selecting rows manually in a spreadsheet (see Modifying the Current Selection from the Spreadsheet).
- Selecting bars in a histogram or polar histogram (see Modifying the Current Selection from a Scatter Plot).
- Selecting points in a scatter plot or polar scatter plot (see Modifying the Current Selection from a Scatter Plot).
- Selecting cells in a heat plot (see Modifying the Current Selection from a Heat Plot).
- Filtering on descriptor values, i.e. selecting entries whose descriptors fall in particular ranges (see Setting the Current Selection by Filtering on Descriptor Values).
- Filtering on descriptor percentiles, e.g. selecting entries in the “best” 20% of a descriptor range (see Setting the Current Selection by Filtering on Descriptor Percentiles).
- Selecting certain cells in statistics table (see Setting the Current Selection from a Descriptive Statistics Table).

18.16.3 Setting the Current Selection by Use of the Menu

The current selection may be set by use of the **Selection** button in any data-view menu:

- To remove everything from the current selection, click on **Selection** and then **Select none** in the next menu.
- To add everything to the current selection, click on **Selection**, then hit **Select all** in the next menu.
- To set the current selection so that it contains everything in a particular named selection, but nothing else, click on **Selection**, then on **Selection** in the next menu and pick the required named selection from the resulting list.
- The method described in the preceding bullet point may be used to set the current selection to contain all entries in a particular data set (see Selecting Entire Data-Sets).

18.16.4 Modifying the Current Selection from the Spreadsheet

- To add an entry to the current selection, click on the relevant row of the spreadsheet with the *Ctrl* key depressed (If you forget to use the *Ctrl* key, you will make the row the current entry without altering the current selection.).
- To add a range of entries (i.e. contiguous rows) to the current selection, depress the *Shift* key, click on the first of the required rows and drag the cursor down to the last of the required rows.
- To toggle (i.e. reverse) the selection status of an entry, click on the relevant row of the spreadsheet with the *Ctrl* key depressed.
- To toggle the selection status of a range of entries in a spreadsheet (i.e. contiguous rows), depress the *Ctrl* key, click on the first of the required rows and drag the cursor down to the last of the required rows.
- To set the current selection to contain a range of entries (i.e. two or more contiguous rows), and nothing else, click on the first of the required rows, depress the *Shift* key and then click on the last of the required rows.

- If rows in a spreadsheet have been grouped (see Grouping Entries by Descriptor Values: Tree Views), performing any of the actions described above will apply the relevant selection-status change to everything in that branch.
- The following keyboard shortcuts can also be used:
  - *Ctrl-A* selects all items
  - *Ctrl-D* deselects all items
  - *Ctrl-I* inverts the selection

18.16.5 Modifying the Current Selection from a Scatter Plot

The current selection may be set or modified by interacting with a scatter plot or polar scatter plot:

- Set the mouse mode to **Select** (see Plot Mouse Modes).
- To set the current selection to contain all entries in a particular bar of the histogram, but nothing else, left-click on the bar.
- To set the current selection to contain all entries in two or more adjacent histogram bars, but nothing else, left-click in the first of the required bars and, keeping the mouse button depressed, drag the cursor to enclose the remaining required bars.
- To toggle (i.e. reverse) the selection status of all entries in a bar, left-click in the bar with the *Ctrl* key depressed.
- To toggle the selection status of all entries in two or more adjacent bars, left-click in the first of the required bars with the *Ctrl* key depressed and drag the cursor to enclose the remaining required bars.
- To add all entries in a bar to the current selection, left-click in the bar with the *Shift* key depressed. The selection status of entries lying outside the bar will remain unchanged.
- To add all entries in two or more adjacent bars to the current selection, left-click in the first of the required bars with the *Shift* key depressed and drag the cursor to enclose the remaining required bars.
- The following keyboard shortcuts can also be used:
  - *Ctrl-A* selects all items
  - *Ctrl-D* deselects all items
  - *Ctrl-I* inverts the selection

18.16.6 Modifying the Current Selection from a Heat Plot

- Set the mouse mode to Select (see Plot Mouse Modes).
- To set the current selection to contain all entries in a particular (rectangular) region of the plot, but nothing else, left-click in the display area near to the required region and, keeping the mouse button depressed, drag the cursor to enclose the required points.
- If this proves difficult, it may help to zoom in so as to increase the separation between points (see Zooming Plots).
- To toggle (i.e. reverse) the selection status of a particular entry (i.e. point), left-click on the point with the *Ctrl* key depressed.

- To toggle the selection status of all entries in a particular region of the plot, depress the *Ctrl* key, left-click in the display area near to the required region and, keeping the mouse button depressed, drag the cursor to enclose the required points.
- To add a particular entry (i.e. point) to the current selection, left-click on the point with the *Shift* key depressed.
- To add all entries in a particular region of the plot to the current selection, depress the *Shift* key, left-click in the display area near to the required region and, keeping the mouse button depressed, drag the cursor to enclose the required points.
- The following keyboard shortcuts can also be used:
  - *Ctrl-A* selects all items
  - *Ctrl-D* deselects all items
  - *Ctrl-I* inverts the selection

18.16.7 Setting the Current Selection by Filtering on Descriptor Values

- The current selection can be set to contain entries whose descriptor values fall in particular ranges. For example, you might want to select all entries for which ANG1 is less than 140 and DIST1 is between 2.5 and 3.0.
- One way of doing this is to select entries falling in a particular region of a scatter plot (see Modifying the Current Selection from a Scatter Plot) but a more flexible alternative, capable of handling any number of descriptors simultaneously, is the filtering option.
- To open the *Filter* window, first display the spreadsheet and choose the columns on which you want to filter by clicking on the descriptor names at the column heads (the column backgrounds will turn black). If you omit to do this, it will be assumed that all descriptor columns are relevant.
- Then pick **Selection** from the spreadsheet data-view menu and **Filter...** from the next menu. This will display a *Filter* window, which can be docked with or undocked from the main data-analysis window in the usual ways (see Docking and Undocking Plots and Other Data Views). The window contains a slider bar for each descriptor that you chose. If the *Filter* window is undocked, you can add a new descriptor to it by displaying the spreadsheet in the data-analysis window, clicking on the required descriptor name at the head of its column, and dragging it into the *Filter* window.
- Define the required range for each descriptor by clicking on the green slider arrows and dragging them to the required points, e.g.

- As you move a slider arrow, the descriptor value to which it is pointing will be shown in the nearby white box. It is possible to type the required descriptor value directly into the box.
- The descriptor limits can be set in percentage terms by right-clicking on the slider and selecting **Set by % (see** Setting the Current Selection by Filtering on Descriptor Percentiles).
- The various ranges will be ANDed together in a Boolean sense. For example, the settings shown in the above example will result in a current selection containing the entries visible in the spreadsheet whose *DIST1* value lies between 2.548 and 3.129 and whose *ANG1* value lies between 93.46 and 139.347.
- If the *Update Dynamically* box at the bottom left is ticked, the current selection will be updated as soon as you alter a descriptor limit. Otherwise, it will not be set until you hit the **Apply** button.
- A descriptor range can be reset (i.e. to include all values) by right-clicking in its slider box and hitting **Reset** in the resulting menu. All descriptor ranges can be reset to hitting the **Reset** button at the bottom of the window.
- A descriptor range can be turned off (therefore, not applied as a filter) by unticking the box immediately above its slider bar, or removed entirely from the window by right-clicking on the slider and selecting **Remove**.

18.16.8 Setting the Current Selection by Filtering on Descriptor Percentiles

- In the *Filter* window (see Setting the Current Selection by Filtering on Descriptor Values), the limits for any given descriptor can be set in percentage terms by right-clicking on the descriptor's slider bar and selecting **Set by %**. This will open a window like the one shown below:

- Suppose first that the **hits** radio button is set on, as in the example above:
  - The window allows you to specify that the top (or bottom) x % of hits (i.e. entries) are to be selected, where x is defined by typing into the white box.
  - So, for example, if you have 100 entries visible in your spreadsheet and you set x = 10 (as above), the descriptor range will be set to exclude all but 10 entries, those with the lowest or highest descriptor values, depending on whether you pick the **top** or **bottom** radio button.
- Now suppose that the **structures** radio button is set on:
  - This option is relevant when your data set comprises hits from a ConQuest substructure search, in which you may have more than one hit from any given structure.
  - If, for example, x = 10 and **top** is switched on, entries will be selected in descending order of descriptor value until 10% of structures have at least one entry (i.e. hit fragment) in the selection.
- The selection is not applied until **OK** is hit.

18.16.9 Setting the Current Selection from a Descriptive Statistics Table

Clicking on any cell of a descriptive statistics table which represents a count of entries (e.g. number of outliers) will set the current selection to contain those entries and nothing else.

18.16.10 Creating a Named Selection

- At any time, the current selection may be assigned a name by clicking on **Selection** in the data-view menu and then **Name selection...** in the next menu. This opens a dialogue box in which you can type a (mandatory) name for the selection and (optionally) some comments. The comments may be viewed later by use of the *Selections* window (see Listing, Removing and Combining Named Selections).
- Hitting the **Save** button will save this selection, under the given name, for the duration of the data analysis.
- Named selections can be preserved between Mercury sessions by saving the data analysis in *.sqlite* format (see Saving Data Analyses).

18.16.11 Recalling a Named Selection

- A named selection can be recalled (i.e. made the current selection) by clicking on **Selection** in the data-view menu, followed by **Selection** in the next menu. This displays a list of all named selections known to the data analysis, from which you can pick the one you require.

- Alternatively, if you have an open *Selections* window (see Listing, Removing and Combining Named Selections), you can also recall a named selection by right-clicking on it in this window and picking **Select** from the resulting menu.

18.16.12 Listing, Removing and Combining Named Selections

- Named selections can be listed, removed or combined by picking **Selection** from the spreadsheet data-view menu followed by **Manage...** in the next menu.
- This opens a *Selections* window, which can be docked with or undocked from the main data-analysis window in the usual ways (see Docking and Undocking Plots and Other Data Views). The window will look something like this:

- All named selections will be listed in the white box on the left.
- To delete a named selection, right-click on it in the left-hand white box and hit **Delete** in the pull-down menu. Alternatively, highlight the selection and click on the **Delete** button.

- To view the comments you typed in when you named a selection, right-click on the selection and hit **Details** in the resulting pull-down menu.
- Named selections can be combined by Boolean logic to create new named selections:
    - Build up the required Boolean expression by dragging selections from the left-hand boxes to the boxes on the right. For example, the figure above shows the definition for the Boolean expression *Selection1 AND Selection2 NOT Selection3*.
    - If you make a mistake, you can remove a selection from any of the right-hand boxes by right-clicking on it and hitting **Remove** in the resulting menu. Alternatively, the **Clear** button will empty all three Boolean boxes.
    - Once the Boolean combination is defined, hit **Save**. This will open a *Save Selection* window with which you can name, add comments to, and save the new selection.

18.16.13 Selecting Entire Data-Sets

- Each time a new data set (e.g. *search1*) is read in, a named selection (e.g. *All_search1*) will automatically be created for it, containing all entries in the data set.
- Therefore, the current selection can be set to contain all entries in a particular data set, and nothing else, by clicking on **Selection** in the data-view menu, followed by **Selection** in the next menu, and then picking the required data-set named selection in the resulting list.

18.17 Creating and Manipulating Data Plots

18.17.1 Creating Plots

- If you are in any part of the data-analysis window that displays a data-view menu containing a **Plots** button, the general procedure for creating a plot is as follows:
    - Hide any entries that you do not want included in the plot, e.g. by making a suitable current selection and hiding the unselected entries.
    - Hit the **Plots** button, then pick the type of plot you want from the resulting menu, e.g. **Scatter plot**.
    - This will display a dialogue box in which you can select from pull-down menus the descriptor(s) to be plotted. The number of descriptors to be selected will depend on the nature of the plot, e.g. two or three for a scatter plot:

- If you are displaying a spreadsheet, plots can also be created in the following ways:
  - For any type of plot, you may choose the required column(s) before hitting the **Plots** button by clicking on the descriptor name(s) at the column head(s). Columns thus chosen will be displayed on a black background. If the number of columns chosen is inappropriate for the type of plot you then request, a dialogue box similar to that shown above will be displayed for you to amend your choice.
- If you are displaying a table of statistics, you can create histograms and polar histograms by right-clicking in a row of the table and selecting the appropriate option from the resulting menu. The histogram will plot the descriptor to which the selected row pertains.

18.17.2 Plot Mouse Modes

- When interacting with a plot, left-button clicks with the mouse will have different actions according to the *mouse mode* that has been set.
- To change a mouse mode, select **Mouse** from the plot menu bar options. A list of mouse modes will be displayed, with the current one ticked. Click on an unticked mode to make it current.

18.17.3 Zooming Plots

- You can zoom in on any region of a plot. This is useful, for example, to inspect more closely a crowded region of a scatter plot or to allow easier picking.
- To do this, set the mouse mode (see Plot Mouse Modes) to **Zoom in**. Then left-click in the display area near the edge of the region you want to magnify and, keeping the mouse button depressed, drag the mouse to enclose the required region.
- It is possible to zoom in multiple times.
- To undo zoom-in actions one by one, set the mouse mode to **Zoom out**. Each left-click in the display area will then undo the most recent zoom-in action that has not yet been undone, until there are none left. From that point on, left-clicking will have no effect until you change mouse mode.
- Alternatively, a complete series of zoom-in actions can be undone by right-clicking and selecting **Reset view**.

18.17.4 Panning Plots

To move a plot within its display area, set the mouse mode to **Pan** and move the mouse with the left button depressed. The plot will follow the mouse movement.

18.17.5 Setting the Current Selection from a Plot

The current selection may be set from any type of plot using the **Select** mouse mode (see Summary of Methods for Setting the Current Selection).

18.17.6 Changing Plot Axes

- The descriptors plotted on a plot may be changed by use of the general drag and drop mechanism (see Using Drag and Drop between Data Views).
- Alternatively, if the plot has an **Axes** button in its data-view menu:
- x and y axes can be swapped by clicking on **Axes** followed by **Swap axes** in the next menu.
- A new descriptor can be plotted along the x (or y) axis by hitting **Axes**, followed by **X Axis** (or **Y Axis**), then selecting the required descriptor from the resulting list.

18.17.7 Printing and Saving Plots

- Any plot may be saved to a graphics file. Several formats are supported, including Portable Network Graphics (*.png*), jpeg (*.jpg*), X11 and Portable Pixmap (*.xpm, .ppm*), and tiff (*.tif*).
- To save a plot, click on the **File** button of the plot's data-view menu and select **Export...** from the resulting menu.
- The resulting dialogue box allows selection of format and image size (choose from the available range of standard sizes or click on **Size** and type the required numbers of pixels in the *Width* and *Height* boxes). Type the required file name in the white box to the right of the format specification or use the **...** button to specify the name via a file browser. When you hit **OK**, you will be shown a preview of the image and will be able to cancel if it is not as you expect.
- To print a plot, click on the **File** button of the plot's data-view menu and select **Print...** from the resulting menu.

18.17.8 Deleting Plots

- Any plot which is docked into the main data-analysis window may be deleted by clicking on the x icon at the far right of the bar immediately above the plot area.
- An undocked plot may be deleted by closing its window.

18.17.9 Histogram Features

- Here is an example histogram:

- The proportion of each bar that belongs to the current selection is indicated by use of a different colour (red by default).
- The position in the histogram of the current entry (see Data-Analysis Terms: Current Entry) is indicated by a vertical, annotated line. When the data set is from a CSD search, this will usually be the one being displayed in the Mercury visualiser.
- The appearance of a histogram can be customised by right-clicking anywhere in the histogram display area (or selecting **Display** from the data-view menu) and picking **Configure...**. from the resulting menu. This displays a dialogue box that allows you to:
    - Change the number of bins in the histogram.
    - Set a title for the plot.
    - Modify the annotation of the x and y axes.
    - Change the font.
    - Alter colours (to change a colour, click on the coloured panel and select from the resulting palette).

18.17.10 Cone Angle Correction

- When analysing angles involved in intermolecular interactions, such as D-H...A angles (θ) in hydrogen bonds, it is important to be aware of the difference between Cartesian and spherical polar coordinates.

- To generate an unbiased histogram showing the density of contacts based on a given parameter any bin of the histogram should, in principle, correspond to an equal volume in three-dimensional space.
- When a hydrogen-bond donor (D-H) approaches an acceptor (A) there is essentially only one orientation of the donor that achieves a D-H...A angle of exactly 180°. As θ deviates from linearity the acceptor group sweeps out a possible cone of approach which gets progressively larger as the angle decreases.

- This means that if a straightforward histogram of intermolecular D-H...A hydrogen-bonding angles is plotted, using bins of equal size, the distribution is inherently biased away from 180°.
- By dividing the bin frequencies by sin θ the histogram can be corrected to account for this bias.
- To apply cone correction to a histogram right-click anywhere in the histogram display area (or selecting **Display** from the data-view menu) and pick **Cone angle correction** from the resulting menu. To remove the correction untick the option from the same menu.
- The figure below shows an unmodified histogram (top) and cone-angle corrected histogram (bottom) for a CSD distribution of D-H...A angles observed in a typical hydrogen-bonding interaction.

18.17.11 Polar Histogram Features

- A polar histogram is a histogram in which the bars are arranged on circular axis, e.g.

- Although a polar histogram can be created for any numerical descriptor, it is most appropriate for torsion angles, the distribution of which is inherently circular (since a torsion angle of -180 is the same as +180).
- In other respects, polar histograms operate in the same way as normal histograms (see Histogram Features).

18.17.12 Scatter Plot Features

- Scatter plots can be two or three dimensional. If the former, the two chosen descriptors are plotted along the x and y axes. Three-dimensional scatter plots use colour to represent the value of the third descriptor, e.g.

- Colours are assigned in a spectrum, red indicating a high value and blue a low value. The vertical colour bar on the right of the above plot shows the scale in the third dimension. The display of this bar can be toggled on or off by right-clicking in the display area and hitting **Show colour scale** in the resulting menu.
- Each point on the plot represents one entry in the data set. By default, entries in the current selection are shown as stars, other entries as circles.
- The point on the plot representing the current entry (see Data-Analysis Terms: Current Entry) is labelled. When the data set is from a CSD search, this will usually be the one being displayed in the Mercury visualiser.
- If the data set came from a ConQuest substructure search, clicking on a point with the left mouse button will display the corresponding CSD structure in the Mercury visualiser window, provided you are in **Select** mouse mode (see Plot Mouse Modes).
- A regression line expressing the y-axis variable as a linear function of the x-axis variable can be switched on and off by right-clicking in the plot area and hitting **Show regression line** in the resulting menu. Leave the cursor on the line to get a display of the regression equation and r^2 value.
- The appearance of a scatter plot can be customised by right-clicking in the display area and selecting **Configure...**. from the resulting menu. This opens a dialogue box that allows you to:
    - Set a title for the plot.
    - Modify the annotation of the x, y and “colour” axes.

- Set the minimum and maximum values along the x and y axes.
- Change the font.
- Alter colours (to change a colour, click on the coloured panel and select from the resulting palette).
- Change the symbols used to display selected and unselected points.
- Show selected and unselected points in different colours (obviously not suitable if colour is being used to plot a third descriptor value).
- Change the mouse mode.

18.17.13 Polar Scatter Plot Features

- A polar scatter plot is a scatter plot that uses polar coordinates. One of the descriptors is plotted along a circular axis, the other along the radius of the circle. As for normal scatter plots, colour can be used to represent a third descriptor. Here is an example:

- Although a polar scatter plot can be created for any numerical descriptor, it is most appropriate when the descriptor plotted on the circular axis is a torsion angle, since torsion-angle distributions are inherently circular (a torsion angle of -180 being the same as +180).
- In other respects, polar scatter plots operate in the same way as normal scatter plots (see Scatter Plot Features).

18.17.14 Heat Plot Features

- A heat plot always refers to two descriptors, which are plotted along the x and y axes.
- The plot is divided into a grid. The colour of (at least) the upper triangle of each grid square is dependent on the number of entries falling in that square.
- Colours are assigned in a spectrum, red indicating a high density of points, blue a low density. By default, white indicates that there are no points in the square.
- Here is an example:

- The vertical colour bar on the right of the plot indicates how many entries per grid square a given colour represents. The display of this bar can be toggled on or off by right-clicking in the plot area and hitting **Show colour scale** in the resulting menu. This can also be done in the *Configure heat plot* window (see below).
- The proportion of entries in a cell that belong to the current selection is indicated by use of a triangle of a different colour (dark green by default) in the lower right. The area of the triangle is dependent on the proportion of entries in the cell that are currently selected. If all are selected, the triangle will occupy the entire area below the cell diagonal.
- The display of grid lines can be toggled on or off by right-clicking in the display area and hitting **Show grid** in the resulting pull-down menu. This can also be done in the *Configure heat plot* window (see below).

- The appearance of a heat plot can be customised by right-clicking anywhere in the heat-plot display area (or selecting **Display** from the data-view menu) and selecting **Configure...** from the resulting menu. This displays a dialogue box that allows you to:
    - Change the number of bins (i.e. grid squares) along the axes.
    - Set a title for the plot.
    - Modify the annotation of the x and y axes.
    - Change the font.
    - Alter the background colour (to change, click on the coloured panel and select from the resulting palette).

18.18 Statistical Calculations

18.18.1 Calculating Descriptive Statistics

- If a spreadsheet is on display:
    - To calculate simple statistics (mean, median, standard deviation, etc.) for the visible entries in a descriptor column, choose the column by clicking on the descriptor name at the column head (the column background will turn black) and then hit **Statistics** in the spreadsheet data-view menu, followed by **Descriptive statistics**
    - If several columns are chosen in this way, they will all be included in the *Descriptive Statistics* table. If no columns are chosen, all columns in the spreadsheet will be included.
- If any other data view is on display, e.g. a heat plot:
    - Click on **Statistics** in the data-view menu followed by **Descriptive statistics**. to calculate descriptive statistics for those descriptors involved in the data view you are displaying (e.g. the two descriptors plotted on the heat plot).
- The following descriptive statistics may require some explanation:
    - *Skewness* has a value of zero for a perfectly symmetrical distribution. Negative values imply that the tail of the distribution is longer on the low side than the high side, and vice versa. For example, the following distribution is negatively skewed:

- *Kurtosis* measures the sharpness of a distribution and has a value of zero for a normal distribution. Positive values indicate a distribution with a sharp peak and long, "thick" tails. Negative values indicate a rounded peak with short, "thin" tails.
- *Low quantile* (or *high quantil*e) is, by default, the value below (or above) which 25% of the observations lie.
- *Low outliers* (or *high outliers*) is, by default, the number of observations falling more than 3 standard deviations below (or above) the mean.

• The default presentation of descriptive statistics may be altered by right-clicking in the table area (or selecting **Display** from the data-view menu) and selecting **Configure...** from the resulting menu. The resulting dialogue box allows you to:

- Specify which statistics are to be displayed (select a descriptor name by clicking on it, then hit **<<Remove** to undisplay the descriptor or **Add>>** to display it).
- Set a title.
- Alter the number of standard deviations above or below which an observation is considered an outlier.

- Alter the quantile definition (for example, typing 10 in this box will cause the upper and lower deciles to be displayed rather than the default upper and lower quartiles).

- Left-clicking on any cell which represents a count of entries (e.g. the number of outliers) will set the current selection to those items and nothing else.

18.18.2 Calculating Circular Statistics

- When dealing with data of a circular or periodic nature, such as torsion angles, the distinction between high and low values is arbitrary and the designation of the zero position can vary according to some external convention. By chemical convention, torsion angles are measured in a positive or negative fashion from 0°, with a discontinuity at +/- 180°. The arithmetic averaging of torsions is therefore clearly problematic: at the simplest level, the mean for a bimodal torsion angle distribution with peaks close to +180° and -180° will be close to 0°, which is obviously incorrect.
- In order to treat periodic variables correctly we need to resort to the appropriate statistical model, that of circular statistics, which generates its own specialist descriptors of periodic distributions.
- If a spreadsheet is on display:
  - To calculate circular statistics (mean, median, standard deviation, etc.) for the visible entries in a descriptor column, choose the column by clicking on the descriptor name at the column head (the column background will turn black) and then hit **Statistics** in the spreadsheet data-view menu, followed by **Circular statistics.**
- If any other data view is on display, e.g. a heat plot:

Click on **Statistics** in the data-view menu followed by **Circular statistics** to calculate descriptive statistics for those descriptors involved in the data view you are displaying (e.g. the two descriptors plotted on the heat plot).

18.18.3 Calculating Correlation Matrices

- If a spreadsheet is on display:
  - To calculate a correlation matrix for two or more descriptors, choose the required columns by clicking on the descriptor names at the column heads (the column backgrounds will turn black) and then hit **Statistics** in the spreadsheet data-view menu, followed by **Correlations...**
  - If no columns are chosen, they will all be included.
  - The correlation matrix will be based only on the currently visible rows.
- If any other data view is on display (e.g. a heat plot):
  - Click on **Statistics** in the data-view menu followed by **Correlations..**. to calculate the correlation matrix for those descriptors involved in the data view you are displaying (e.g. the two descriptors plotted on the heat plot).
- By default, the matrix will show Pearson correlation coefficients in the lower triangle, and the significance levels of the coefficients in the upper triangle. Correlations that are statistically significant (significance level < 0.05) will be shown on a coloured background.
- The default contents of the matrix may be altered by clicking on the buttons at the bottom left and top-right of the display area. By default, these will say **Correlation**

and **Significance**, respectively. By clicking on them, you can alter the lower or upper triangle, respectively, to display Pearson correlation coefficient, covariance, significance level, or Spearman correlation coefficient.

- If either Pearson or Spearman correlation coefficients are shown in one of the triangles and significance levels in the other, the significance level will be that of whichever type of correlation coefficient is on display.
- The default presentation may also be modified by right-clicking anywhere in the correlation-matrix display area (or selecting **Display** from the data-view menu) and selecting **Configure...** from the resulting menu. The resulting dialogue box will allow you to:
    - Specify a title.
    - Set the upper or lower triangle to display any of: Pearson correlation coefficient; covariance, significance level; Spearman correlation coefficient.
    - Turn the highlighting of statistically-significant values on or off and alter the colour used for highlighting (click on the coloured panel and select from the resulting colour palette).
    - Alter the level at which a correlation is deemed significant.

18.18.4 Principal Component Analysis

- The method of Principal Components Analysis (PCA) is based on transforming a set of potentially correlated variables into a new, and smaller, set of uncorrelated and mutually orthogonal variables called principal components. This process can make it easier to understand multivariate data and can significantly aid the location and identification of clusters of observations having similar geometry.
- The technique is often used when trying to analyse the variation in a number of correlated molecular or intermolecular geometric parameters within a dataset of related crystal structures.
- To perform a principal component analysis, choose the required descriptors by clicking on the descriptor names at the column heads (the column backgrounds will turn black) and then hit **Statistics** in the spreadsheet data-view menu, followed by **Principal components.** If no columns are chosen, then a window will appear in which you can choose which descriptors you wish to include. This will open a data-view window that looks something like this:

- In the example above we are analysing the geometry of aminofuranoside rings where each of the five internal torsions angles in the ring have been defined *TOR1* to *TOR5*. By applying PCA we can significantly reduce the number of parameters that describe the majority of the conformational variance in the dataset.
- The descriptors that will be included in the calculation can be changed using the tick-boxes next to the descriptors name. Alternatively, you can use the **All**, or **None** buttons to choose descriptors. Right-clicking anywhere in the data-view and selecting **Configure...** from the resulting menu will open a window in which you can customise the output from the calculation and its appearance.
- Hit the **Calculate** button to run the PCA on the selected set of descriptors. The resulting PC scores are displayed on the right-hand side of the data-view:

- Analysis of the principal components indicates that just two components (*PCA1* and *PCA2*) account for 99.99% of the variance in the dataset.
- Select the components you wish to retain using the tick-boxes next to the component names, specify a *Base name* and hit **Save**. Saved components will be added to the list of descriptors on the left and will also appear as new columns in the main spreadsheet data-view. From here a scatterplot of these first two principal components could be generated:

- The plot shows two clusters which correspond to the two main ring conformational types present in the dataset. The cluster on the left comprises C2'-endo rings while those on the right are C3'-endo.

18.18.5 Testing Whether Two Means are Significantly Different

- In any view of the data in which you have some entries selected and some not, and for any descriptor which appears in that data view, you can perform a statistical test to establish whether the mean value of the descriptor for the selected entries is significantly different from the mean of the unselected entries.
- The test only takes account of the visible entries.
- If a spreadsheet is on display:
  - Choose the required descriptor(s) by clicking on the descriptor names at the column heads (the column backgrounds will turn black) and then hit **Statistics** in the data-view menu, followed by **Significance**
  - If no columns are chosen, all descriptors will be included.
- If any other data view is on display (e.g. a heat plot):
  - Click on **Statistics** in the data-view menu followed by **Significance**. The test will be performed on each descriptor involved in the data view you are displaying (e.g. the two descriptors plotted on the heat plot).
- By default, the resulting table will have one row for each descriptor tested. The row will give the mean value of the descriptor for the selected entries, the corresponding value for the unselected entries, and the statistical significance of the difference between the means.

- The statistical test used is an F-test (equivalent in this situation to an unpaired Student t-test), which assumes that the variances of the two distributions being compared are equal.
- By default, rows corresponding to differences that are significant at the 0.05 level are highlighted.
- The default presentation of the results can be altered by right-clicking anywhere in the display area and selecting **Configure...** from the resulting menu. This displays a dialogue box that allows you to:
    - Set a title for the table.
    - Alter the probability level (expressed as a percentage) at which a result will be deemed statistically significant (for example, setting this value to 1 would define the significance level as 0.01).
    - Switch on or off the highlighting of significant results.
    - Set the highlighting colour (click on the coloured panel and choose from the resulting colour palette).

18.18.6 Adding Descriptors to a Statistics Table

A new descriptor can be added to a statistics table by use of the general drag and drop mechanism (see Using Drag and Drop between Data Views).

18.18.7 Removing Rows from Statistics Tables

- To remove a row from any table of statistical results, right-click in the display area and select **Remove rows ...** from the resulting menu.
- To remove several contiguous rows, left-click in the first, depress the *Shift* key, left-click in the last, right-click and select **Remove rows ...**
- Non-contiguous rows can be removed in the same way, but using the *Ctrl* key instead of the *Shift* key.

18.18.8 Saving Statistics Tables

- To save any table of statistical results, click **File** from the data-view menu in the display area and select **Export...** from the resulting menu.
- In the resulting dialogue box, select the required format, which can be one of: comma-separated value (*.csv*); tab-separated value (*.tsv*); space-separated value (*.txt*); or HTML table (*.htm* or *.html*).
- Type the required file name in the right-hand white box, or use the **...** button to specify the file via a file browser.
- If the file you specify already exists and you click on the *Append* box (so it is ticked), the table will be added to the end of the file, leaving the original contents intact. Otherwise, it will overwrite the original contents.

18.19 Saving Data Analyses

- To save a data analysis, select **File** from the data-analysis top-level menu, followed by **Save database.** Use the resulting file browser to specify a name for the output file, which will be stored in *.sqlite* format.

- All descriptors and named selections will be stored, but not plots or other data views. Plots may, however, be saved separately in their own formats (see Printing and Saving Plots), as may statistical tables (see Saving Statistics Tables).

19 CSD-Community

You will need to activate Mercury in order to access features available in CSD-Community, but you can do so for free. To activate CSD-Community in a copy of enCIFer that is not activated, use the CCDC Activation tool and select the CSD-Community tab and then click 'Activate'. The CCDC Activation tool can be started in the following way on each supported platform.

19.1.1 Windows

The CCDC Software Activation tools will be located in the

<INSTALLDIR>\CSD_2020\Software Activation

folder of a CSD System installation, or

<INSTALLDIR>\Software Activation

in other CCDC product installers. To start the graphical user interface, double-click on *ccdc_activator_gui.exe*

19.1.2 Linux

The CCDC Software Activation tools will be located in the

<INSTALLDIR>/CSD_2020/bin

folder of a CSD System installation, or

<INSTALLDIR>\software_activation

in other CCDC product installers. To start the graphical user interface, run *ccdc_activator_gui*

19.1.3 MacOS

The CCDC Software Activation tools will be located in the

<INSTALLDIR>/CSD_2020

folder of a CSD System installation, or

<INSTALLDIR>\Software_Activation

in other CCDC product installers. To start the graphical user interface, double-click the *CCDC-Software-Activation* application.

19.2 Overview

- CSD-Community provides essential software tools to all scientists wanting to utilise structural data for research and education. Capabilities of CSD-Community include:
    - Structure deposit and structure access through the CCDC website
    - Access to an educational collection of structures in the CSD
    - Syntax checking for experimental CIFs
    - Reduced cell checking during data collection

- The following features of CSD-Community are available through the Mercury interface:
    - The ability to deposit experimental structures and access data from the CSD online (see Deposit and Access of Structures)
    - Access to the CSD Educational Collection of structures containing 700+ structures carefully selected to enhance chemistry learning (see Teaching Database)
    - The ability to validate CIFs and ensure experimental structures files are correctly formatted for submission to journals and database (see Syntax Checking for CIFs)
    - Functionality for performing crystal structure reduced cell checks against the CSD before collecting a full data set (see Reduced Cell Checking)

19.3 Deposit and Access of Structures

- The CCDC act as the repository for all small molecule structural data. Data are deposited pre-publication at the point of creation and shared through the Cambridge Structural Database at publication. CSD-Deposit can also be used to publish structures directly through the CSD as a CSD Communication. Click on **CSD-Community** in the top-level menu and select **CSD Deposit** to open the CCDC's online CIF deposition and validation service. Detailed instructions for use of the deposition service are available from the CCDC's website.
- The crystal data for any published structure deposited in the CSD are available from the CCDC's Access Structures service. Click on **CSD-Community** in the top-level menu and select **Access Structures Online** to see CSD entries online and access the associated deposited data files.

19.4 Teaching Database

- The teaching database is a subset of over 700 CSD entries selected to help demonstrate many of the core concepts and principles taught in the undergraduate chemistry curriculum.
- Further information can be obtained by clicking on **CSD-Community** in the top-level menu and selecting **About Teaching Database** to view documentation online.
- The Teaching Database is available through Mercury; this can be accessed by clicking on **CSD-Community** in the top-level menu and selecting **Open Teaching Database**.

19.5 Syntax Checking for CIFs

- EnCIFer is a tool designed to ensure that CIFs are syntactically correct before they are archived, transmitted or used as input to other software, e.g. data validation programs or structure visualisers.
- The enCIFer program operates on single or multi-block CIFs to locate and report any syntax/format violations using the currently enabled CIF dictionary.
- Addition of certain standard additional information is achieved via two data entry wizards:
    - Publication wizard - basic bibliographic information required by most journals and databases that accept CIF deposition documents.

- Data wizard - chemical and physical property information that enhances a CIF for journal or database deposition.

- Visualisation of structure(s) in the CIF is designed to aid users in validating the content and semantics of the CIFs.
- EnCIFer also offers additional crystallographic and publication checks via a link to the checkCIF/PLATON service provided by the IUCr. For more detailed information, see the enCIFer documentation provided within the CSD software.
- EnCIFer can be accessed from the **CSD-Community** top-level menu by selecting the **Launch enCIFer** option.

19.6 Reduced Cell Checking

- CellCheckCSD is a freely available, automated tool, primarily designed for checking unit cells against those deposited in the CSD during data collection.
- CellCheckCSD has been developed for automated use through the Rigaku Oxford Diffraction CrysAlis^{Pro} and Bruker APEX2 software packages.
- The ability to compare the pre-experiment unit cell to all other unit cells in the CSD before collecting a full data set can help identify the compound and avoid an unnecessary data collection.
- Further information is available via the **CellCheckCSD** item in the **CSD-Community** top-level menu.

20 CSD-System

20.1 Overview

- The CSD-System provides direct access to comprehensive 3D chemical structure data combined with powerful tools to exploit this structural knowledge for application across the full range of the physical, chemical and biological sciences including education. Capabilities of the CSD-System include:
  - structure lookup
  - advanced searching functionality
  - data analysis and charting
  - intramolecular geometry analysis
  - intermolecular interaction analysis
- The following features of CSD-System are available through the Mercury interface:
  - The ability to perform searches using WebCSD and ConQuest as well as analyse results using interactive charting, plotting and statistics (see Searching and Data Analysis).
  - The ability to analyse molecular conformations using Mogul in order to validate the intramolecular geometries of newly determined crystal structures and identify unusual geometric features (see Intramolecular Geometry Analysis).
  - The ability to analyse intermolecular interaction geometries using IsoStar, for example to validate the geometry of an H-bond or short range interaction within a crystal structure (see Intermolecular Interaction Analysis).

20.2 Accessing CSD-System

- The CSD-System functionality will be available for all Cambridge Structural Database System subscribers. If you do not already have access, evaluations are normally available for those organisations interested in subscribing.
- To activate CSD-System in a copy of Mercury where it is currently not activated, use the CCDC Activation tool. Activation can be completed both online and offline.
- The **Help, About Mercury...** menu option will display details about your current licensing status, including features available in CSD-System.

20.3 Searching and Data Analysis

- To find the specific chemistry, structural knowledge, textual information or bibliographic data you are looking for, there are a range of search and data analysis features available in the CSD-System.
- The ability to perform fast, cross-platform lookup of structures is provided by WebCSD. Click on **CSD-System** in the top-level menu and select **Launch WebCSD** from the resulting menu. For more detailed information, see the specific WebCSD documentation online.

- When search results are transferred to Mercury from ConQuest, the highlighting of hit fragments is also transferred. For control over exactly how this is displayed in Mercury, click on **CSD-System** in the top-level menu and select **ConQuest Hit Highlighting…** from the resulting menu. For more detailed information, see Controlling ConQuest Hit Highlighting.
- Sophisticated searches and complex combinations of queries can be performed using the program ConQuest. Click on **CSD-System** in the top-level menu and select **Launch ConQuest** from the resulting menu to launch the program. For more detailed information, see the ConQuest documentation provided within the CSD software.
- For any searches that have been performed using ConQuest or the Packing Feature search functionality in Mercury, you can analyse and chart the results using the Data Analysis module. To access this module, click on **CSD-System** in the top-level menu and select **Data Analysis** from the resulting menu. For more detailed information, see Data Analysis.

20.4 Intramolecular Geometry Analysis

- Mogul is a library of molecular geometry derived from the Cambridge Structural Database (CSD) and provides information on the preferred values of bond lengths, valence angles, acyclic torsion angles and ring conformations.
- The intramolecular geometry within the currently visualised structure can be automatically assessed in comparison to known experimentally determined structures using Mogul and each geometric parameter classified as usual or unusual. To perform this automated geometry check, click on **CSD-System** in the top-level menu and select **Mogul Geometry Check…** from the resulting menu.
- To launch an independent instance of the Mogul program, click on **CSD-System** in the top-level menu and select **Launch Mogul** from the resulting menu.
- You can also select which Mogul data libraries will be used for a Mogul Geometry Check. In particular this will allow selection/deselection of CSD update libraries and any in-house database Mogul libraries you have produced. To select the Mogul data libraries to use, click on **CSD-System** in the top-level menu and select **Mogul Settings…** from the resulting menu.
- For more detailed information, see Geometry Validation

20.5 Intermolecular Interaction Analysis

- IsoStar is a library of intermolecular interaction geometry derived from both the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB). The knowledge base provides information on the preferred distances and orientations of interactions such as hydrogen-bonds, halogen-bonds and stacking interactions with the information presented in the form of 3D scatterplots.
- One or molecules within the currently visualised structure can be analysed to find matches in the IsoStar interaction database. To do this you can simply click on **CSD-System** in the top-level menu and select **IsoStar Interaction Check…** in the resulting menu. This will analyse all molecules on display, but you can narrow down the analysis by selecting one or more atoms before performing the interaction check.
- To access the IsoStar server and browse for intermolecular interactions to study, click on **CSD-System** in the top-level menu and select **Launch IsoStar** from the resulting menu.

- The location of the IsoStar server must be specified so that scatterplots can be successfully retrieved. To ensure the settings are correct, click on **CSD-System** in the top-level menu and then select **IsoStar Settings...** from the resulting menu.
- For more detailed information, see Visualising CSD- and PDB-based Intermolecular Contact Distributions

21 CSD-Materials

21.1 Overview

- CSD-Materials provides software for aiding the discovery of new pharmaceutical crystal forms. It utilises the unique information about crystal packing in the Cambridge Structural Database (CSD) as a knowledge base for identifying and analysing crystal-packing motifs. Amongst other functionality, the module enables novel motif searches and analyses of the CSD that are applicable to problems such as:
    - co-crystal design
    - counter ion selection and
    - polymorph prediction
- Specifically, CSD-Materials allows:
    - Structural searches to be carried out to identify interaction motifs that occur between particular functional groups and determine their relative abundance (see Searching on Motifs).
    - Searches can also be performed by selecting atoms in a currently loaded crystal structure and searching for structures that contain a similar packing arrangement (see Searching on Crystal Packing Features).
    - Crystal packing similarity calculations to be performed between structures containing the same (or similar) compound (see Determining Crystal Packing Similarity).
    - The ability to display Bravais, Friedel, Donnay and Harker (BFDH) theoretical crystal morphologies (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology).
    - A range of molecule-based gas phase theoretical calculations using an interface to MOPAC (see Theoretical Calculations using MOPAC).
    - Force field intermolecular energy calculations using the UNI intermolecular potentials (see Calculating Intermolecular Energies).
    - Determination of hydrogen bond propensities to assess the potential risk of polymorphism (see Determining Hydrogen Bond Propensities).
    - Quick assessment of H-bond coordination likelihoods (see Quick Assessment of H-Bond Coordination).
    - Mapping of interaction preferences around complete molecules in a crystal structure (see Assessing Molecular Interaction Preferences Using the CSD).
    - Co-crystal design based on assessment of molecular complementarity (see Co-Crystal Design).
    - Analysis of hydrate crystal structures (see Analysing Hydrate Structures).
    - Analysis of solvate crystal structures (see Analysing Solvate Structures).
    - Generation of conformers based on geometrical statistics from the CSD (see Generating Conformers Using the CSD).
    - Solution of crystal structures from powder diffraction data using DASH (see Structure Solution from Powder Data).

- The following tutorials are available:
  - **Assessing the Risk of Polymorphism via H-Bonding Propensities** (see Tutorial 7: Assessing the Risk of Polymorphism via H-Bonding Propensities)
  - **Searching on Hydrogen-Bond Motifs** (see Tutorial 8: Searching on Hydrogen-Bond Motifs)
  - **Studying Crystal Packing Features** (see Tutorial 9: Studying Crystal Packing Features)
  - **Searching on Packing Similarity** (see Tutorial 10: Searching on Packing Similarity)
  - **Building a Packing Similarity Tree Diagram** (see Tutorial 11: Building a Packing Similarity Tree Diagram)

21.2 Accessing CSD-Materials

- CSD-Materials is usually available at no additional cost to academic Cambridge Structural Database System subscribers. Industrial customers should contact *admin@ccdc.cam.ac.uk* for further information. Evaluations are normally available for those industrial organisations interested in purchasing.
- Licensing of CSD-Materials is handled through the CSD licensing. To activate CSD-Materials in a copy of Mercury where it is currently not activated, use the CCDC Activation tool. The **Help, About Mercury...** menu option will display details about your current licensing status, including features available in CSD-Materials.

21.3 Searching on Motifs

- The motif search option allows the CSD (or any dataset of structures) to be searched for user-defined intermolecular interactions. Once the contacts between functional groups have been specified and the type of motif required has been chosen, the program will generate all possible combinations of the contact groups. Atoms are considered to be in contact if their separation is less than the minimum distance defined. No angular or orientation information is used in the search algorithm. More than one search query can be set at a time and results from motif searches can be manipulated to include or exclude particular motifs.
- To access the hydrogen bond motif searching tool:
  - Click on **CSD-Materials** in the top-level menu and select **Motifs...** from the resulting menu.
- Both of the above methods will open a dialogue from which you can specify the type of motif you wish to use in your search.

- To perform a motif search the following steps need to be taken:
  - Select the motif or motifs to be searched for (see Motif Selection).
  - Select the dataset of structures in which to search for the motif (see Structure Selection).
- The wizard will take you through these two steps in turn.

21.3.1 Motif Selection

- There are three methods of selecting the motif(s) you wish to search for:
  - Selecting a pre-defined motif (see Pre-defined Motifs)
  - Generating a new motif (see Generating a New Motif)
  - Opening a motif query file (see Opening a Motif Query File)
- Select the method you wish to use for motif selection and then click on **Next... >** to proceed to the motif selection area of the wizard.

Pre-defined Motifs

- In the left-hand side of the Wizard select a motif from the expandable lists of homomeric, heteromeric and Water Motifs.
- To see which motifs are available in the expandable lists, click on the "+" to the left of the motif type. The motifs are listed along with a graph set description where appropriate (see Displaying Graph Sets).
- Select the motif by clicking on the name of the motif (for example COOH R2,2(8) in the Homomeric list) and click on **Select Motif >>**. To aid selection, when a motif name is clicked on a diagram of it is displayed.
- More than one motif may be selected.
- Once the motif selection is complete, click on **Next...>** to proceed to selection of the search structures.

Generating a New Motif

- The wizard will take you through the three steps involved in generating a motif query:
  - Defining the functional groups (see Defining Functional Groups)
  - Selecting contacts (see Selecting Contacts)
  - Generating motif search queries (see Generating Motif Search Queries)

Defining Functional Groups

There are three ways to select a functional group to be used in a motif query.

- *Method 1*:
  - Click on **Add** to select a predefined functional group. This will open the *Select substructure* dialogue.
  - Use the **substructure** drop-down list to view the available functional groups.
  - If you wish to filter the list of pre-defined substructures on constituent elements, check the relevant boxes in *Elements*. For example, to include substructures in the drop-down list that contain only carbon and oxygen, uncheck the box labelled all and check the boxes labelled C and O.
  - Alternatively, click on the arrow bar to the right of the display window to view diagrams of all the substructures available. The displayed substructures can be selected by clicking on them.
  - Hit **OK** once a substructure has been selected.
  - Hit **Cancel** to return to the *Motif Search Wizard* dialogue.
- *Method 2*:
  - Click on **Sketch…** to draw a functional group. This will open a sketcher window. For help using the sketcher, please refer to the ConQuest User Guide, see: *http://www.ccdc.cam.ac.uk/support/documentation/#conquest*

- Once satisfied with the motif sketched, click **Store** to proceed. You will be asked to supply a name for the sketched group.

- *Method 3*:
  - Click on **Load…** to load a previously defined functional group, saved as a *.con* file from ConQuest.
- To edit a functional group in the list, select the functional group and click on the **Edit…** button. This will open the sketcher where the motif can be modified. To save the changes made, click on **Store** in the sketcher window.
  Note: The edited functional group will replace the original functional group.
- Click on **Remove** to delete a functional group from the list.
- Once you are happy with the functional groups generated for a motif search then click **Next >** to proceed. The wizard will now prompt you to specify contacts that will make up the motif search queries.

Selecting Contacts

- Select the atoms of the functional groups which define the contact(s) of the motif.
  Note: More than one contact can be defined between the functional groups but each contact has to be defined separately, using the following procedure.
  - Click on **Add…** to define which atoms of the functional group(s) are involved in the motif. This will open the *Contact picker* dialogue:

  - For both the *From* and *To* **Group:** drop-down lists, select the functional group(s) involved in the motif. Then using the **Atom:** drop-down lists, select the atoms between which the desired contact will occur. Alternatively, click on atoms of the functional groups in the display pane of the dialogue. The

contact specified will be shown as a pink dashed line between the relevant atoms.

- The contact distance and H-bond angle used in the search are shown at the bottom of the dialogue. These values can be changed by clicking on the **Modify…** button. This will open the *Contact Definition* window. There are two methods provided for defining a contact distance:
  Method 1: To apply an actual distance, select the **Actual distance** option and enter **Minimum** and **Maximum** values for the range.
  Method 2: To use distances defined by the van der Waals radii, select the **vdW distance** option and add or subtract distances to the sum of the van der Waals radii for the minimum and maximum values of the range.
  To change the default hydrogen bond angle, check the **Define hydrogen bond angle** tickbox and enter a new angle.
- Click on **OK** to proceed with edited values or **Cancel** to discard any changes made.

- Click on **Edit…** to make changes to a selected contact.
- Click on **Remove** to delete a selected contact.
- Once you are happy with the contacts specified click **Next >** to proceed. The wizard will now prompt you to specify the types of motif to be generated.

Generating Motif Search Queries

- The final step in motif generation is the selection of the type of motif (for example, ring, infinite chain, discrete chain) and the number of contacts present in the motif. Using these parameters, along with the functional groups and the contact atoms specified, all possible combinations of the contact atoms that match the requirements of the motif will be generated by the program. For example, selecting a ring motif that contains three contacts (between atoms A and B of the functional groups) will result in two motifs being generated: an AB-AB-AB- motif and an A-B-AB- motif.

- Any number of motif geometries may be included in the motif generation process by checking the appropriate boxes.
- If you do not wish to specify a motif type or the number of contacts involved, you may select motifs that are built from any intramolecular contact, intermolecular contact or mixture of intra- and intermolecular contacts. This is done by checking one of the **intra**, **inter** or **any** boxes in the *Generate simple contacts* area of the dialogue.
- Once all motif types have been selected, click on **Generate >** to begin the motif generation process.
- If a contact between different functional groups has been specified, a dialogue offering the option of placing those functional groups on different molecules is shown.
    - Click on **Add…** to select the functional groups to be placed on different molecules.
    - To delete functional groups from the list, select the functional groups and click on **Remove**.
    - To proceed with the generation of the motifs, click on **Generate >**.
- The wizard will display the progress of the motif generation and list the motifs that are produced. The motif generation may be stopped by clicking on **Stop** and resumed by clicking on **Continue**.
- The nomenclature used to describe the motif starts with a letter indicating the motif type (for example R for ring), followed by the number of contacts in the motif, followed by a string of letters that denotes the sequence of atoms which comprise the motif. The hyphen that separates the letters of the code indicates a contact between molecules.
- If topologically equivalent motifs could be generated from the contacts specified, only a single instance of the motif will appear in the list.
- A default name for the family of motifs generated is given in a text box towards the bottom of the dialogue. This name may be changed or customised.
- To accept the motifs generated, click on **Next >** to proceed to selection of the search structures.
- If you wish to change any of the parameters used in the motif generation click on **< Back** to move back through the wizard windows.

Opening a Motif Query File

- Click on the **Open… >** button on the right-hand side of the wizard and use the file browser to select a previously saved query file or files (with a *.mxml* or *.gmxml* file extension). The motif file(s) opened will be listed in the left-hand side of the dialogue. When all required files have been opened, click on **Next >** to proceed to selection of the search structures.

21.3.2 Structure Selection

- The structures to be searched can now be chosen using the wizard:

- The options available for selecting structures are as follows:
  - The current version of the Cambridge Structural Database and available updates may be selected by highlighting them in the list of *Available Structures* and clicking on **Select >>**. Pre-loaded files of structures are available under *Structures* and pre-loaded refcode lists (.gcd files) are available under *Refcode Lists.*
  - Individual refcodes or families of refcodes can be selected by clicking on **Add Refcode… >**. This will open the *Enter Refcode* dialogue. Type the name of the refcode in the text box. If the refcode belongs to a family of refcodes, all the family members will be listed in the *Refcode Family* pane of the dialogue. To select all members of a refcode family, check the **Enter refcode family** box. To select only a single member of a refcode family type the complete refcode (e.g. AACRUB01) and do not check the **Enter refcode family** box.
  - Files of structures may be loaded by clicking on **Open… >**. The following file formats are available:
    Structure files: *cif, cmf, mol, sdf, sd, mdl, mol2, mol, pdb, ent, res and xyz*
    CSD files: *ind, c2m, gcd, mry*
  - The "Best Representative" refcode lists are composed of the best representative of each unique structure present in the database. There are four lists, each representing a different criterion against which "best" is judged. The lists are `Best_R-factor_list.gcd`, `Best_hydrogens_list.gcd`, `Best_low_temperature_list.gcd`,`Best_room_temperature_list.gcd`. These structures can be loaded into Mercury and used as the basis for CSD-Materials searches. Using these lists will remove bias in the statistics caused by presence of a large number of redetermined structures in a

refcode family. For more information on the creation of the Best Representative lists, see reference 5 (see References). The lists are part of the CSD distribution and can be found in `C:\Program Files\CCDC\CSD_2020\CSD_541\subsets` (if the default installation directory for Windows is chosen) or `/cambridge/cdrom541/subsets` (if the default location for Linux is chosen).

- Completed motif searches, packing feature searches and packing similarity searches are also available to be used as the basis of a new search and are listed under the *Searches* heading.

- *Note*: More than one set of structures may be loaded and searched.
- To remove individual structures or files of structures from the list of Selected Structures, highlight the structure(s) and click on **<< De-select**.
- The total number of structures to be searched is shown at the top of the *Selected Structures* pane of the dialogue.
- To confirm the structure selection step and proceed, click **Next >**.
- Finally, you may enter a name for the search or simply accept the default name provided, then click on **Start Search**.

21.3.3 Display of Motif Search Results

- As the search proceeds, refcodes of hits are displayed in the *Searches* tab of the *Structure Navigator*.

- Information regarding the progress of the search is shown at the bottom of the pane.
- The search can be paused at any time by clicking on the **Pause** button. This stops the search and a new search can be started. The search can also be resumed by clicking on **Continue**.
- To view a hit, click on the refcode in the list. The hit is displayed and the functional groups of the motif are highlighted. Options to customise the highlighting of the motif are given in the *Hit Highlighting* dialogue (see Controlling Hit Highlighting).
- There are two options available for viewing the results of a motif search. Select to view the structures either **by Motif** or **by Structure** from the **View Results** drop-down list in the *View Options* area of the *Searches* tab.
    - **Viewing by Motif:**
      Each motif of the search is listed by name, and the refcodes of the motif hits are given below the motif name in an expandable list.
      Shown alongside the motif name is the running total of the number of hits found and the frequency of occurrence of the motif ((number of hits found / number of structures that contain the functional groups) * 100).

- **Viewing by Structure:**
  The refcodes of the hit structures are listed and the motifs found within the hit are given below the refcode in an expandable list. If the refcode contains more than one motif, all motif names will be listed.
  Shown alongside the refcode is the number of motifs found and the frequency of occurrence for the motif is listed alongside the motif name in the expandable list.

- For either of the above options, the hits can be sorted by clicking on the column headings in the hit display area.
- For example, in *View by Motif* mode, the structures can be sorted on number of hits, with the most popular motif shown first, by clicking on the column heading **number of structures**. Alternatively, the motif with the highest frequency of occurrence can be found by clicking on the column heading **% frequency**. In *View by Structure* mode, the refcode containing the greatest or least number of motifs can be found by clicking on the column heading **number of motifs**.
- Checking the **List all matches** box allows all instances of the motif found within a structure to be viewed. The matches, labelled match 1, match 2 etc. will be shown in the expandable list of hits and can be selected for viewing.
- **Multi-view** mode allows any subset of the motifs found for a refcode to be displayed simultaneously. Clicking on a structure from the list will add the structure to the display.
- You can also view structures that contain the functional groups chosen but do not exhibit the motif(s) searched for by clicking on **Show negative results**. This will automatically switch the viewing mode to *View by structure*.
- All motif searches performed are listed in the drop-down box at the top of the searches tab and can be viewed at any time.

21.3.4 Filtering Motif Search Results

- Motif search results can be filtered to show only those structures that satisfy a logical combination of motif search queries, e.g. to find structures containing both an R2 motif and a C1 infinite chain motif.
  Note: It is not unusual, if you search for more than one motif, to find that one motif belongs to a subset of another.
- Select the motif search you wish to filter from the drop-down list towards the top of the *Searches* tab of the *Structure Navigator*. Click on the **Options...** button and select **Filter Results...** from the resulting menu. This will open the *Filter search results* dialogue:

- The individual motif search queries will be listed under *Queries*. The three boxes on the left-hand side of the *Filter search results* dialogue (labelled *must have*, *must have one of*, *must not have*) represent the logical operations required to define complex combinations of individual search queries.
- Drag motif search queries into the boxes to set up the desired search. A search query can be dragged by clicking on it with the left-hand mouse button and moving the mouse (keeping the button depressed) so that the icon drops into the desired box. Icons may also be dragged from box to box.
- Once the required search filters have been defined, hit **Apply Filters**. The filtered results will be presented in the *Searches* tab of the *Structure Navigator*.
- To clear any search filters that have been set, click on the **Options...** button and select **Clear Filters,** from the resulting menu.

21.4 Searching on Crystal Packing Features

21.4.1 Running a Search

- To access the crystal packing feature search tool:
  - Click on **CSD-Materials** in the top-level menu and select **Search** and then **Crystal Packing Feature...** from the resulting pull-down menus.

- In order to perform a packing feature search, you must have a crystal structure loaded in the visualiser from which you can select a number of atoms from one or more molecules (see Selecting and Deselecting Atoms and Molecules). The selection of atoms can be done either before or after launching the *Crystal Packing Feature* wizard. Once selected, the number of atoms and molecules chosen will be shown in the window.
- Click the **Next >** button to accept the current selection of atoms and continue to the next section which allows modification of the search parameters.

- The wizard lists all the atoms and bonds selected and the current restrictions placed on the atoms for the purposes of the search.
- From the atoms tab of this window you can:
  - Change the element types of the atoms and allow variable element selections.
  - Change the required number of hydrogens for each atom, or set as unspecified.
  - Add restrictions on the atomic charge.
  - Change the required number of bonds for each atom, or set as unspecified.
  - Add restrictions on the cyclicity.
- From the bonds tab you can:
  - Modify the bond types between the atoms selected.
- To make a modification select the atom or bond of interest, and click on the **Modify** button in the *Search Options* pane, then select the required option from the resulting menus. Constraints may also be added or removed from the search by clicking on the appropriate boxes in the *Add Constraints* pane. For example, if you do not wish the number of bonds belonging to an atom to be used as a constraint in the search for similar packing features, uncheck the box Number of bonds.
- Hit **Reset** to return the atom and bond properties to those defined by the original selection of atoms.

- Once you have chosen the required search parameters for the atoms and bonds within the packing fragment, click **Next >** to continue to the next section at which point the geometry tolerances for the search are chosen.

- The wizard will now ask for the required level of geometric similarity to be chosen. The options are as follows:
  - **Very high** (distances within +/- 10 %, angles within +/- 5 degrees)
  - **High** (20 %, 12 degrees)
  - **Medium** (30 %, 20 degrees)
  - **Low** (40 %, 30 degrees)
  - **Custom** (editable tolerance values)
- You can force all matched fragments to belong to the same molecule by clicking the **All matched fragments must belong to same molecule** checkbox
- You can view the geometric parameters within the packing feature fragment that are being used for the search by checking the **Show distances** or **Show angles** checkboxes.
- After choosing the required geometric tolerances, click **Next >** to continue to the next section.
- At this point you can choose to define one or more geometric parameters within the packing feature in order to characterise the deviation from your search geometry. Click on atoms in the visualiser to select distance (two atoms), angle (three atoms)

or torsion angle (four atoms) parameters, then click **Add >** to move the selection to the *Parameter List*.

- Once you are finished defining parameters, or if you do not wish to select any, click on **Next >** to proceed.
- The next stage is to select the structures that are to be used for the search. The wizard allows the selection of the full CSD, specific refcodes (or refcode families), any loaded structures and previously completed searches. It is also possible to read in saved structure files (see Structure Selection).
- At this stage, you will be given the option to apply search filters. Filters are secondary search criteria used to restrict more general searches on the basis of, for example, precision, whether disorder is present, whether there are unresolved errors in the structure, or whether the structure is organic or organometallic (see Available Database Search Filters).
- Finally, you can choose to modify the default values for the *Packing Feature Name* and/or the *Search Name*, which will be used to store the search query and results.
- Click on **Start Search** to begin the packing feature search. The results will be shown in the *Searches* tab of the *Structure Navigator* on the right-hand side of the Mercury window (see Viewing and Analysing Results).

21.4.2 Available Database Search Filters

- The following database search filters are available when performing a packing feature search (see Running a Search).

- **3D coordinates determined**: Switch on to exclude structures for which no 3D information is available.
- **Not disordered**: Switch on to exclude structures which contain crystallographic disorder.
- **No errors**: Switch on to obtain only those structures which are error-free at the 0.05Å level.
- **Not polymeric**: Switch on to exclude polymers and metal catena structures.
- **No powder structures**: Switch on to exclude structures in which the data was obtained from a powder diffraction study.
- **R-factor**: Switch on the checkbox and then click on one of the three radio buttons to restrict the search to structures with R-factors less than or equal to 0.05, 0.075 or 0.1.
- **Only Organics** or **Only Organometallics**: Switch on the checkbox and then click on the appropriate radio button to restrict the search to (i) organic structures only, or (ii) organometallic structures only.

21.4.3 Viewing and Analysing Results

- The packing feature results will appear in the *Searches* tab of the *Structure Navigator* as the search progresses, with each hit being shown as a separate line. The refcode or structure name for each hit will be listed along with:

- the root mean square (RMS) deviation corresponding to the overlay of the hit fragment with the search fragment
- the values of each of the defined geometric parameters for the hit fragment, with each parameter shown in separate column

- A view of the currently selected structure with the hit fragment highlighted is shown in the visualiser.

- The hitlist can be re-ordered based on any column, at any time by clicking on the relevant button at the top of the hit list.
- Multiple hits within the same structure can also be viewed by clicking on the **List all matches** checkbox at the bottom of the *Structure Navigator*.
  Note: If a structure has more than one match, the structure that is displayed when the refcode is selected and the RMS value listed alongside the refcode, represent the best match found within the crystal structure.
- The search fragment can also be overlaid with the hit fragment by clicking on the **Show packing feature** checkbox at the bottom of the window.
- Structures that contain the same chemistry as the search query but do not exhibit a matching packing feature can be viewed by clicking on **Show negative results**.

- Visualisation of the hit fragment and its structure can be customised by clicking on the **Options** button in the *Structure Navigator* and choosing **Hit Highlighting…** in the pull-down menu (see Controlling Hit Highlighting).
- Results from packing feature searches can be saved using the **Options** drop-down menu (see Saving Searches and Results).

21.5 Determining Crystal Packing Similarity

21.5.1 Running a Crystal Packing Similarity Calculation

- To access the crystal packing similarity tool:
  - Click on **CSD-Materials** in the top-level menu and select **Search** and then **Crystal Packing Similarity...** from the resulting pull-down menus.

- The crystal packing similarity calculation is intended to compare multiple structures of the same compound, so the molecular compound must be the same in each structure that is selected for comparison.
- In order to compare the packing of structures you must select structures to use as both reference and comparison structures. Each of the structures listed in the *Reference Structures* list will be compared with each of the structures in the *Comparison Structures* list, so the similarity calculation can be performed using one-to-one, one-to-many, many-to-one or many-to-many. The structures can be selected in two different ways:
  - by clicking on the **Select...** button which will allow you to add structures from loaded databases, refcode lists, ConQuest hits, Mercury files or from saved structures (see Structure Selection).

- or by clicking on the **Add Refcode...** button which allow you to type in a CSD refcode and add this structure, or add the entire family of structures that correspond to the first six letters of this refcode.
- Structures can also be removed from the lists by selecting them within the list and clicking on the **Remove** button.

- Once the structures to have been selected, you can choose to modify the specific settings for the comparison calculation by clicking **Next** in the wizard (see Setting Comparison Options).
- Alternatively, you can simply take the default settings by clicking on **Compare** to start the similarity comparison immediately. The results will be shown in the *Structure Navigator* on the right-hand side of the Mercury window and in the visualiser.

21.5.2 Setting Comparison Options

- There are a number of options available for tailoring the packing comparison to the particular system being looked at.
- In the first page of options within the packing similarity wizard you can modify the **Size of molecular cluster to compare** and the **Geometric tolerances**.

- The crystal packing similarity calculation will compare the central molecule of the structures plus a number of molecules nearby in order to see how similar the packing is. The default size of this molecular cluster is 15 molecules (the central molecule plus 14 others), but this value can be changed in the dialogue box by typing in the required number next to **Size of molecular cluster to compare**.
- You can also choose to filter the results by clicking on **Filter comparisons that do not have all n molecules in common**. If checked, the output will exclude all comparisons

of structures for which the full set of *n* molecules do not match (where *n* is the number of molecules chosen for the cluster comparisons).

- The geometric tolerances for defining whether the packing is similar or not can be modified here as well. The default values are 20% for the **Distance tolerance** and 20 degrees for the **Angle tolerance**, but each of these can be edited by typing in the relevant box.
- Once these options have been selected, click **Next** to proceed to the advanced settings page or click on **Compare** to start the similarity calculation immediately.
- The final screen of options contains more advanced settings for dealing with different types of structures.

- The following options are available here:
  - **Allow molecular differences**: Enable this option to find packing similarity between structures containing similar, but not identical molecules. For example, bromo and chloro analogs.
    Note: This option is computationally demanding and is not suited to large database searches.
  - **Ignore hydrogen positions**: This option is checked by default, due to the fact that hydrogens are not always present in structures from the CSD. If unchecked the positions of hydrogen atoms are taken into account when assessing packing similarity and calculating the root mean square deviation between packing patterns.
  - **Ignore bond types**: If checked the packing comparison does not take into account bond types. This may be particularly important if the compound being studied contains a carboxylate group for example, where the two C to O bonds could be described as either double or single, or a delocalised region with potentially confusing bond orders.

- **Allow structure inversion**: Allow structure inversion: When checked, this option allows the algorithm to invert the structure to find the best overlay between clusters - this can be important when studying Sohncke space groups for example (such as $P1$, $P2_1$, $C2$, $P3_1$). This option is checked by default, due to the fact that for many organic structures in particular the absolute structure may not be determined. If the difference between enantiomers is important in the particular analysis being performed, then the user should uncheck this option.
- **Ignore each atom's hydrogen count**: If checked the packing comparison does not take into account the number of hydrogen atoms bonded to each non-hydrogen atom.
- **Ignore each atom's bond count:** If checked the packing comparison does not take into account the number of bonded atoms.
  Note: By ignoring hydrogen counts and the number of bonded atoms then comparisons are no longer restricted to structures containing the same compound.
- **Ignore smallest molecular components:** If checked, the smaller component/s of the crystal structure (as determined by atom count) is/are ignored. This allows comparison of two different solvates for example. By ignoring the solvent molecules, isostructurality between the packing of the major components of structures can be explored. In a three-component system, where the three components are all different, only the biggest molecule is retained for structure comparisons.
  Note: When using this option, it is worth considering increasing the size of the molecular cluster used, for example to 20 molecules, as ignoring the smaller components can potentially lead to slab-like clusters.
- **Show only the highest similarity result:** When crystal packing similarity calculations are performed on systems with Z' > 1, more than one molecular cluster is generated for the reference crystal (as there is not a single, unique cluster for Z'>1 structures). The use of more than one reference cluster means there will be more than one similarity result obtained. If this checkbox is checked, only the comparison results with the highest similarity will be shown. All results can be seen if this box is unchecked.

- The settings in these wizard pages can also be returned to default values by clicking on the **Default** button. To accept the current settings, click **Next**.
- Finally, the name for the search can be modified or left as the default name. Once this is done, click **Compare** to start the packing similarity analysis.

21.5.3 Viewing and Analysing Results

- There are two ways of viewing the packing similarity results within Mercury. Initially the results are listed as comparisons between pairs of structures with the reference structure name used to order the list.

Searches

crystal_comparison_1 | Options

reference	comparison	molecules in common	RMS	reference group
HXACAN	HXACAN01	1 out of 15	0.091	group1
HXACAN	HXACAN04	1 out of 15	0.09	group1
HXACAN01	HXACAN04	15 out of 15	0.074	group2
HXACAN	HXACAN06	1 out of 15	0.088	group1
HXACAN01	HXACAN06	15 out of 15	0.091	group2
HXACAN04	HXACAN06	15 out of 15	0.02	group2
HXACAN	HXACAN07	1 out of 15	0.09	group1
HXACAN01	HXACAN07	15 out of 15	0.097	group2
HXACAN04	HXACAN07	15 out of 15	0.024	group2
HXACAN06	HXACAN07	15 out of 15	0.011	group2
HXACAN	HXACAN08	15 out of 15	0.075	group1
HXACAN01	HXACAN08	1 out of 15	0.096	group2
HXACAN04	HXACAN08	1 out of 15	0.095	group2
HXACAN06	HXACAN08	1 out of 15	0.094	group2
HXACAN07	HXACAN08	1 out of 15	0.095	group2
HXACAN	HXACAN09	1 out of 15	0.094	group1
HXACAN01	HXACAN09	15 out of 15	0.24	group2
HXACAN04	HXACAN09	15 out of 15	0.17	group2
HXACAN06	HXACAN09	15 out of 15	0.157	group2
HXACAN07	HXACAN09	15 out of 15	0.153	group2
HXACAN08	HXACAN09	1 out of 15	0.098	group1
HXACAN	HXACAN10	1 out of 15	0.099	group1
HXACAN01	HXACAN10	15 out of 15	0.389	group2
HXACAN04	HXACAN10	15 out of 15	0.321	group2
HXACAN06	HXACAN10	15 out of 15	0.309	group2
HXACAN07	HXACAN10	15 out of 15	0.304	group2
HXACAN08	HXACAN10	1 out of 15	0.101	group1
HXACAN09	HXACAN10	15 out of 15	0.153	group2
HXACAN	HXACAN11	1 out of 15	0.103	group1
HXACAN01	HXACAN11	15 out of 15	0.474	group2
HXACAN04	HXACAN11	15 out of 15	0.409	group2
HXACAN06	HXACAN11	15 out of 15	0.399	group2
HXACAN07	HXACAN11	15 out of 15	0.394	group2
HXACAN08	HXACAN11	1 out of 15	0.104	group1
HXACAN09	HXACAN11	15 out of 15	0.245	group2
HXACAN10	HXACAN11	15 out of 15	0.097	group2

Show Structure

Reference Comparison Both

Show missed matches

Group by similarity

- The reference and comparison structure names are shown for each comparison in the *Structure Navigator*. Clicking on any comparison will show a structural overlay of the molecules in common within the visualiser. The number of molecules in common between the two structures is also listed along with the root mean square (RMS) deviation in distance and a similarity measure of the simulated powder patterns of the two structures (see Powder Pattern Similarity Measure). Also, the group to which the comparison structure is assigned based on similarity is displayed (see Group by similarity, below).
- These results can be re-ordered to list the comparisons by comparison structure name, molecules in common, RMS deviation or PXRD similarity simply by clicking on relevant button at the top of the comparison list.

- The visualiser can be customised to show only the reference structure, only the comparison structure or both structures by clicking on the required radio button under *Show Structure* at the bottom of the *Structure Navigator*.
- Missed matches between the molecular cluster overlays can also be displayed in the visualiser by clicking on the **Show missed matches** button. Matching molecules are typically shown in green, with missed matches shown in red. This can be particularly useful for highlighting what the dissimilarity is, for example, when identifying that two polymorphs have a layer in common.

- Different options for displaying the structures can be found by clicking on the **Options** button within the *Structure Navigator* and selecting **Hit Highlighting…** (see Controlling Hit Highlighting**).**
- Clicking on the **Group by similarity** checkbox will enable a second method for listing the comparison results - using this method the structures are formed into groups for which the packing patterns are highly similar. For the purposes of this grouping, structures which have all molecules in the cluster in common are considered similar. A set of structures is placed in the same group if they are each considered to be similar to at least one other member of the group.

Searches

crystal_comparison_1 | Options

group	spacegroup	density	packing coefficient	a	b	c	alpha	beta	gamma
group1									
HXACAN25	Pbca	1.37	0.71	11.7636	7.2863	17.1041	90	90	90
HXACAN31	Pbca	1.383	0.74	11.7652	7.1877	17.1679	90	90	90
HXACAN32	Pcab	1.379	0.72	7.2028	11.7764	17.1677	90	90	90
HXACAN33	Pbca	1.339	0.69	11.8237	7.3971	17.1526	90	90	90
HXACAN37	Pbca	1.393	0.73	11.7552	7.13941	17.1714	90	90	90
HXACAN	Pcab	1.341	0.69	11.805	17.164	7.393	90	90	90
HXACAN08	Pbca	1.377	0.72	17.1657	11.7773	7.212	90	90	90
HXACAN21	Pcab	1.378	0.72	7.1986	11.782	17.183	90	90	90
HXACAN22	Pcab	1.358	0.7	7.2927	11.806	17.169	90	90	90
HXACAN23	Pcab	1.335	0.69	7.4057	11.837	17.162	90	90	90
HXACAN24	Pcab	1.32	0.68	7.4807	11.853	17.16	90	90	90
group2									
HXACAN17	P21/a	1.322	0.71	12.815	9.278	7.084	90	115.57	90
HXACAN18	P21/a	1.314	0.71	12.833	9.315	7.086	90	115.6	90
HXACAN19	P21/a	1.303	0.7	12.872	9.37	7.085	90	115.62	90
HXACAN26	P21/a	1.298	0.69	12.8856	9.38013	7.10096	90	115.7	90
HXACAN27	P21/n	1.337	0.71	7.0915	9.2149	11.6015	90	97.865	90
HXACAN28	P21/n	1.349	0.73	7.077	9.173	11.574	90	97.9	90
HXACAN30	P21/n	1.338	0.72	7.0907	9.2114	11.5972	90	97.844	90
HXACAN34	P21/n	1.297	0.71	11.7144	9.38544	7.10166	90	82.5879	90
HXACAN35	P21/n	1.317	0.68	7.0661	9.3366	11.6508	90	97.41	90
HXACAN36	P21/n	1.351	0.71	7.0842	9.1734	11.547	90	97.988	90
HXACAN01	P21/a	1.293	0.68	12.93	9.4	7.1	90	115.9	90
HXACAN04	P21/n	1.323	0.69	7.0939	9.2625	11.657	90	97.672	90
HXACAN06	P21/a	1.331	0.72	12.763	9.25	7.082	90	115.52	90
HXACAN07	P21/n	1.332	0.7	7.0941	9.2322	11.6196	90	97.821	90
HXACAN09	P21/n	1.411	0.73	6.98	8.915	11.566	90	98.54	90
HXACAN10	P21/n	1.494	0.77	6.885	8.5819	11.519	90	99.12	90
HXACAN11	P21/n	1.541	0.8	6.82	8.374	11.559	90	99.32	90
HXACAN12	P21/n	1.615	0.82	6.625	7.985	11.916	90	99.41	90
HXACAN13	P21/a	1.355	0.73	12.667	9.166	7.073	90	115.51	90
HXACAN14	P21/a	1.35	0.73	12.698	9.173	7.073	90	115.51	90
HXACAN15	P21/a	1.349	0.73	12.709	9.173	7.077	90	115.57	90
HXACAN16	P21/a	1.332	0.72	12.769	9.24	7.079	90	115.54	90
group3									
HXACAN29	Pca21	1.337	0.73	11.8352	8.5601	14.8199	90	90	90

Show Structure

Reference | Comparison | Both

Show missed matches

Group by similarity

- The *Structure Navigator* will now list the analysed structures under the groups into which they have been assigned. This grouping should correspond to the different polymorphic forms for a compound if the system being studied exhibits polymorphism. The space group and unit cell measurements are also shown for each crystal structure, along with calculated density and packing coefficient to aid in the comparison.
- Results from Crystal Packing Similarity searches can be saved using the **Options** drop-down menu (see Saving Searches and Results) or the *Post-Search Options* menu (see Post-Search Options).
- If the results are saved out as a Search Summary File (.txt), this file will include both the rotation matrix and translation vector associated with each reference/comparison pair.

21.6 Calculating Energies and Structural Properties

21.6.1 Introduction

- Mercury is primarily a program for visualising crystal structures in three dimensions, but it also supplies a number of tools designed to help analyse the structures as well.
- Theoretical calculations are commonly used alongside experimental techniques to investigate structures further, whether it is a geometry optimisation, structure prediction or complicated energy calculation.
- Some options for performing theoretical calculations are available through Mercury, including the following:
  - Simulation of crystal morphology using the BFDH method (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology)
  - Semi-empirical theoretical calculations using MOPAC (see Theoretical Calculations using MOPAC)
  - Force field intermolecular energy calculations using the UNI intermolecular potentials (see Calculating Intermolecular Energies)

21.6.2 Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology

- To display BFDH theoretical crystal morphologies hit **CSD-Materials** and then **Calculations** in the top-level menu, then **BFDH Morphology...** in the resulting pull-down menu.
- The crystal morphology, calculated from the unit cell, will be displayed:

- The *BFDH Morphology* control window can be used to customise the settings.

- The following options are available:
  - *Show BFDH Morphology*: Use this to toggle the display of the morphology.
  - *Scale factor*: Use this to alter the size of the morphology with respect to the unit cell.
  - *Auto Scale:* Switch the tick box on to allow auto-rescaling, so that the morphology is in proportion with the unit cell.
  - *Show (hkl) labels*: switch the check box on or off to enable or disable the labelling of the facet edges.
  - *Draw faces*: Switch the check box on or off to view or hide coloured facet faces. The **Colour** button can be used to alter the colour or the faces (see Using the Select Color Dialogue Box). The transparency/opaqueness of the facet faces can be adjusted using the slider.
  - *Shadows*: Switch the tick box on or off to view or hide shadows on the facet faces.
  - *Draw edges*: Switch the tick box on or off to view or hide the facet edges. The colour of the facet edges can be controlled using the **Colour** button.
  - *Fill morphology with molecules*: Switch the tick box on or off to fill morphology shapes with molecules. Molecules will be shown if the molecular centroid is within all BFDH planes/facets. Visualisation of growth surfaces can be used to aid the understanding and control of growth morphology. One aspect of morphology control might e.g. involve the selection of growth inhibitors that inhibit the growth of particular crystal faces. In this case it is useful to quickly see what functional groups were exposed on certain surfaces.
  - **Open:** This button allows an experimental or simulated morphology (not necessarily a BFDH morphology) to be read in from a file in .cif format. The contents of the file must follow the contents of the .cif files written out via the **Save...** button. Reading in morphologies from a file allows the display of morphologies obtained from sources other than Mercury, e.g. morphologies calculated using more sophisticated models. It is also possible to write out the Mercury BFDH morphology to a .cif file, to edit the file in a text editor and then to read the resulting morphology back in to Mercury for display.

When this route is chosen, care must be taken that the facets after editing still form a closed 3-dimensional shape.

- *Save*: This button allows data relating to the BFDH cell to be saved in CIF format. The file name will be of the format *<identifier>-Morphology.cif*, where *<identifier>* is the name, stored in the *data_<identifier>* line of the CIF.

- *Note:* The **Apply** button will need to be clicked on for changes to be made to the display. Changes made to these settings are not saved between Mercury sessions.
- BFDH morphology will be displayed for all structures while the *BFDH Morphology* window is open. To exit from BFDH morphology display, click either **OK** or **Cancel** in the *BFDH Morphology* window. BFDH morphology will continue to be displayed for the present crystal structure until a new file is loaded, or an alternative structure is loaded.
- Technical details regarding BFDH are provided (see Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphologies).

21.6.3 Theoretical Calculations using MOPAC

21.6.3.1 Overview

- MOPAC (Molecular Orbital PACkage) is a semi-empirical quantum chemistry program based on Dewar and Thiel's NDDO approximation. An interface to MOPAC is provided which allows users to perform a limited number of molecule-based gas-phase calculations.
- The interface developed within Mercury only works with MOPAC 2007 and 2009. MOPAC is not distributed with Mercury, so a copy of the software must be obtained separately.
- The MOPAC software is third-party and as such no support is provided by the CCDC. It is currently free for academic not-for-profit use and available at a cost for industrial use. For more information or to obtain a copy of MOPAC, please visit *http://openmopac.net/*.

21.6.3.2 Setting up a Calculation

- To access the MOPAC interface, hit **CSD-Materials** and then **Calculations…** in the top-level menu, then **MOPAC...** in the resulting pull-down menu.
- A MOPAC calculation is performed by choosing the required options from the setup tab (which will be described in this section) and then clicking on **Submit**.
- Six different semi-empirical Hamiltonians are available. The Hamiltonian to be used can be selected from the *Hamiltonian* drop-down menu. You should bear in mind that not all elements in the periodic table are supported and each Hamiltonian supports a different number of elements. For more information on the specific elements supported and the relative benefits of the different Hamiltonians, see the MOPAC manual: *http://openmopac.net/manual/*.

- Each MOPAC calculation will automatically calculate some basic molecular properties, such as heat of formation, symmetry point group and atomic partial charges as well as producing an output file as standard (*AUX* keyword). In addition to these standard options, you can also choose to do any or all of the following by clicking on the checkboxes:
    - Optimise geometry (see Geometry Optimisations).
    - Calculate bond orders (see Calculating Bond Orders).
    - Calculate electrostatic potential (see Electrostatic Potential Calculations).
- For more advanced users of MOPAC there is also a box provided to add extra keywords to the input file. Clicking on the *Preview* tab will display the current input file with the chosen keywords that will be used when the MOPAC calculation is run.
- Choose a location and filename with which to save the MOPAC input file (**.mop*) and this will default to the structure identifier (e.g. the CSD refcode for CSD structures). If you prefer to use a different name for the MOPAC input file, you can type this into the title line.
- If a copy of MOPAC exists in one of the standard directories (such as *C:\Program Files\MOPAC* on Windows) then Mercury will find the executable automatically, otherwise the path to the MOPAC executable must be specified on the line marked *MOPAC Program*.
- Click on **Submit** to run the MOPAC calculation. The progress of the job will be displayed at the bottom of the *Setup* section of the *Calculation* tab and the dialogue box will switch to the *Results* tab once the job is complete.

- Within the *Results* tab you can also read back in the results of previous calculations by clicking on the **Load previous results** button and selecting a MOPAC *.aux* file.

21.6.3.3 Geometry Optimisations

- Gas-phase molecular geometry optimisations can be performed by ticking the **Optimise geometry** checkbox. Pay close attention to crystal structures with missing hydrogens as the interface will not add these automatically. Hydrogens can, however, be added using the structure editing options available (see Editing a Structure).
- MOPAC will produce an output file called **.aux* in the same directory as the input file. The *.aux* file contains the initial and final molecular geometries, for comparison, and can be read by Mercury.

  Note: Once a MOPAC calculation has been run, the crystal structure information will be lost.

21.6.3.4 Calculating Bond Orders

- Applying the option labelled *Calculate bond orders* will determine bond orders for all the bonds within the molecule.
- Clicking on the *Bond Orders* tab will display the bond orders results listed by:
  - the atom labels that define the bond,
  - the CSD bond type,
  - the MOPAC calculated bond order.

- If *Calculate bond orders* was not selected during the set-up of the calculation, the **View** button will remain greyed out.

21.6.3.5 Electrostatic Potential Calculations

- Checking the **Calculate electrostatic potential** tickbox will cause the *ESP* section of the *Results* tab in the *MOPAC* dialogue box to be activated once the MOPAC calculation is complete.

- A molecular van der Waals surface will be displayed with the molecular electrostatic potential (in eV) mapped onto it using a colour scale ranging from red (negative), through white (neutral) to blue (positive).

- The colour range will be automatically scaled in order to show the greatest range of colour for each surface, however, you can choose to type in a colour range in order to compare multiple molecular surfaces directly.

- The ESP surface can be displayed as dots, edges, nodes or triangles with the option to also control the opacity of the surface.
- You can also turn off the display of the ESP surface by unchecking the **Show ESP** tickbox.

21.6.4 Calculating Intermolecular Energies

21.6.4.1 Overview

- This functionality calculates intermolecular potentials using the empirical "UNI" pair potential parameters as published by A. Gavezzotti, *Acc. Chem. Res.*, **27**, 309-314 1994 and *Crystallography Reviews*, **7**, 5-121, 1998. This does include some H-bonding potentials but is not comprehensive: if there are H-bonds present that are not recognised as one of the types in the table, no special potential is applied. Note that if elements are present outside the set C, H, N, O, S, Si, F, Cl, Br, I no potential energy is added for these atoms, giving misleading energy sums. Note also that no geometry optimisation is performed which in some cases can adversely affect the intermolecular potential values.
- There are 3 options for displaying potentials:
  - Show the strongest *N* interactions where the value *N* can be set. This option will show the strongest interactions for all molecules that are currently on display in the visualiser. For example, if you construct a 2x2x2 packing box, this option will show the strongest interactions for all molecules in the 2x2x2 packing box. The program will add molecules to the display if they are required in order to show a particular interaction. You can also choose here whether or not you want to **include symmetry related interactions** here (use the checkbox).
  - Show molecular packing shell. This shows interactions between a central molecule and the closest *N* surrounding molecules.
  - Show interaction between selected molecules.
- The general purpose of this feature is to display those interactions between molecules which are most significant in energetic terms. Considering the approximate nature of the empirical potentials, the calculation should be used in a qualitative rather than a quantitative sense, for the recognition of packing patterns, and relative importance of interactions.
- The method also calculates potentials between a central molecule and the nearest 200 surrounding molecules. The total packing energy (per formula unit), and the convergence with the number of interactions, is shown in the output tab. Note that this energy sum is for the limited set of neighbouring molecules and is not the lattice energy.

21.6.4.2 Setting up a Calculation

- Click on **CSD-Materials** and then **Calculations…** in the top-level menu and select **UNI Intermolecular Potentials...** from the resulting drop-down menu. This will open the *Intermolecular Potentials* dialogue box:

- The following options can be found under the *Calculate* tab:
  - **Show strongest interactions**: Enable this option to display the strongest *n* interactions between the molecules currently on display where *n* is specified by the user. Additional molecules will be displayed automatically in the visualiser if they are required to show an interaction. An option to **include symmetry related interactions** is also available. If this is checked, symmetry-related interactions are considered when selecting the *n* strongest interactions. If this is not checked, only unique interactions are included when selecting the *n* strongest interactions
  - **Molecular packing shell**: Enable this option to show interactions between a central molecule(s) and the surrounding molecular packing shell of *n* molecules, where *n* is specified by the user.
  - **Show interactions between selected molecules**: Enable this option to display interactions between specific molecules. Select two or molecule in the visualiser (see Selecting and Deselecting Atoms and Molecules), then click on the **Calculate** button.
- The following calculation options can additionally be applied:
  - **Normalise Hydrogens**: normalisation of hydrogen atom positions is available for C-H, O-H and N-H bonds (the H atom will be moved along the bond vector to make the bond length equal to the average neutron diffraction value). Enable this option to normalise H atom positions to the following default values: C-H 1.089Å, N-H 1.015Å, O-H 0.993Å. It is possible to customise the values for C-H, N-H and O-H H-normalisation (see Normalising Hydrogen Atom Positions).
  - **Allow unknown atom types**: Potential parameters are available only for a specific set of atom types. Enable this option to allow *unknown* atom types and perform calculations for cases where potentials are not available.

- **Retain molecules when switching options**: Switching between calculation options may remove interaction molecules from the visualiser. Enable this option if you prefer the molecules to be kept on view.

21.6.4.3 Modifying the Display

- The following display options can be customised under the *Highlighting* tab:

- **Label interaction lines**: Interaction energies can be displayed in either kJ/mol or kCal/mol. Alternatively, interaction lines can be labelled according to the symmetry operation that relates one molecule to another.
- **Line width**: Interactions are displayed as a single broken line. The **Single line width** can be specified. Alternatively, the line width can be set according to interaction strength, i.e. the stronger the interaction the thicker the line.
- **Line color**: To set a custom line colour enable the **Single line color** option and choose from the resulting colour palette. This will open the *Select Colour* dialogue box (see Using the Select Color Dialogue Box). Alternatively, lines can be coloured by interaction strength.

21.6.4.4 Viewing Output

- A list of all interatomic distances and their contribution to the total packing energy can be viewed under the *Output* tab in the *Intermolecular Potentials dialogue* box.
- Further calculation details can be viewed by clicking on the **Show more detail** button.
- To save calculation details to a text file click on the **Save...** button and choose a location and filename with which to save the file.

21.6.4.5 Analysing Multiple Structures at Once

- If multiple structures are being displayed at once within the Mercury visualiser (see Displaying Multiple Structures Simultaneously) the *Intermolecular Potentials* dialogue box will rearrange to provide an additional drop-down menu listing the currently available structures to choose from.

- Intermolecular energy calculations will be performed for just one of the displayed structures and the user can select which structure to use for calculations by choosing from the drop-down menu.

21.7 Determining Hydrogen Bond Propensities

21.7.1 Introduction

- The Hydrogen Bond Propensity tool enables knowledge-based polymorphism risk assessment based on a statistical analysis of hydrogen bonding patterns.
- Probabilities for hydrogen bond pairings to form in the target system are calculated from a statistical model built from relevant structures in the CSD. The model encapsulates information regarding the environment of the functional groups, which ensures the prediction is specific to the target molecule.
- Combining probabilities of hydrogen bond formation with a statistical model that captures information regarding how often a functional group participates allows the generation of chemically sensible alternative structures.
- The view of the solid-state landscape of an active ingredient afforded through the combination of propensity and coordination addresses questions such as how likely polymorphism is and whether there is the possibility of a more stable form. Specifically, you can:
    - Predict likely hydrogen bonds for a given molecule
    - Assess crystal forms e.g. by identifying sub-optimal hydrogen bonding

- Calculate hydrogen bond propensities for individual donor and acceptor groups
- Perform a comprehensive analysis of hydrogen bonding on a set of structures

21.7.2 Starting the Propensity Prediction Wizard

- To access the hydrogen bond propensity calculation tool:
  - Click on **CSD-Materials** in the top-level menu and select **Polymorph Assessment** then **Hydrogen Bond Propensities...** from the resulting menus.
- This will open a dialogue which controls the initial steps of the calculation.

- First, specify a working directory to which you have write permissions. Do this by either entering the full path or by clicking on **Browse…** and selecting your working directory.
- Clicking the **Show advanced options** checkbox allows for a number of advanced settings to be specified, although the defaults are recommended:
  - *Functional group library* – A key step of the process is to analyse the prediction molecule and attempt to identify the hydrogen bonding functional groups that are present. A pre-defined library of functional group definitions is distributed with Mercury for this purpose, the default location of which is listed on the settings page. An experienced user may build up their own library of functional group definitions in an alternate location and preferentially use those simply by browsing to that location from the settings page.

- *Selected databases* – It is possible to generate the dataset of relevant structures for the propensity calculation from a specific database or refcode list rather than the whole CSD.
- *Hydrogen bond definition* – The geometric and chemical definition of a hydrogen bond controls when the *Hydrogen Bond Propensity* tool records the presence or absence of hydrogen bonds, and ultimately therefore controls the model predictions. For most applications the default definitions can be used. However, it can be useful to modify the definition of a hydrogen bond (e.g. when dealing with weaker C-H donors). To view and edit the H-Bond definitions click the **Edit**... button (see Default Definition of Hydrogen Bonds).
- *Use existing regression data* – It is possible to use existing regression data from a previous hydrogen bond propensity calculation for the analysis. To load existing data, click the **Load…** button. Existing data should only be used when all functional group definitions for the current structure are identical to those used for the existing data.

21.7.3 Selecting a Target Molecule

- By default, the structure shown in the Mercury visualiser will be loaded automatically by the *Propensity Prediction Wizard* and shown as a 2D chemical diagram.
  Note: Mercury can read structures from a variety of common file types (see Loading Structures).
- To change the current target molecule, navigate to the desired structure using the Mercury visualiser then click the **Update Structure** button. The 2D chemical diagram will then update to the new structure.

Note: When using a crystal structure (i.e. a structure file containing symmetry information), the actual observed hydrogen bonds, along with the calculated propensities, will be listed in the H-Bond Propensity prediction results table (see Using the Hydrogen Bond Propensity Results Table). This allows for the assessment of a crystal form.

- For the purposes of the hydrogen bond propensity calculation, it is important that all bond types for the target molecule are correctly defined and standardised (see Editing a Structure).
- All hydrogen bond donor and acceptor atoms will be automatically identified, and will be listed in the *Donors* and *Acceptors* columns. Clicking on a donor or acceptor in these lists will highlight it in the 2D chemical diagram as well as highlighting the relevant functional group (see Defining Functional Groups).
  Note: For the purposes of the hydrogen bond propensity calculation, it is important that all donors and acceptors are matched to functional groups and that no donor or acceptor atom is matched more than once. A message shown beside either a green tick or a red cross below the 2D chemical diagram provides the user with feedback relating to this.

21.7.4 Defining Functional Groups

- The program will analyse the target molecule and attempt to identify the hydrogen bonding functional groups that are present.
- A pre-defined library of functional group definitions is distributed with Mercury for this purpose. For less common chemical functionality, for which no match can be found in the library, the program will auto-construct functional groups in an attempt to represent these parts of the molecule.

- It is necessary to check these functional group definitions carefully. You can then choose to keep the identified groups or refine the group definitions. Since the group definitions are used to retrieve structures from the CSD that are relevant to the target this step can be critical to obtaining good propensity prediction models.
- Clicking on a functional group will highlight that group in the 2D chemical diagram. It will also highlight the corresponding donors and/or acceptors in the relevant columns.
- To refine the functional group definitions, you can either:
    - Add pre-defined groups from the functional group library by using the **Add…** button.
    - Sketch your own groups using the **Sketch…** button. This will open a sketcher window. For help using the sketcher, please refer to the ConQuest User Guide, see: *http://www.ccdc.cam.ac.uk/support/documentation/#conquest*
    - Load groups from saved *.con* (connser) files using the **Load…** button.
    - Edit a group with the sketcher using the **Edit...** button.
    - Delete a functional group definition using the **Remove** button.
    - Delete all functional group definitions using the **Remove All** button.
- *Note:* There is no right and wrong way to define groups. The aim is to simply capture specific chemical behaviour and the success is often seen after the model fitting step. On the one hand it can be advantageous to be specific in your group definitions, e.g. you may define a hydroxyl group as being specifically an aromatic hydroxyl. However, if the definition is too specific there may be insufficient examples found in the CSD and the model regression step may prove difficult.

21.7.5 Generating Fitting Data

- The propensity tool will derive hydrogen bond propensities by analysing hydrogen bonds formed in existing crystal structures. A set of *fitting data* containing structures with chemistry relevant to the prediction molecule must therefore be generated.
- Click **Next** to proceed to the *Generate Fitting Data* pane of the wizard. This provides two options for generating a fitting data set:
    - You can retrieve CSD structures using a chemical similarity search (Default). To do this, ensure the **Auto-generate structures** option is enabled and then click on the **Generate** button. It is possible to limit the number of structures in the fitting dataset by clicking the **Truncate data generation** checkbox then adjusting the slider to the desired number.
    - Alternatively, you can specify the *fitting data* set manually by loading in an existing set of structures (e.g. from previous work or from ConQuest searches). To do this enable the **Load structures from file** option, click on the **Browse...** button, then in the resulting dialogue specify either a list of CSD refcodes (.gcd file), or a CIF containing multiple structures.
- Following the generation of fitting data, it is possible to start the hydrogen bond analysis immediately by clicking the **Start analysis automatically** checkbox (see Running a Hydrogen Bond Analysis).
- As the training set starts to be populated with structures, an indication of which functional groups are represented can be seen:

- When the generation is complete, a number of structures for each group are listed. This can often be quite uneven and it is better for the model if the numbers of structures for each group is similar.
    - The slider at the bottom of the pane allows you to remove structures from the training set until a more even set of data is obtained.
      Note: Structures are sorted by their similarity to the prediction molecule so moving the slider to the left reduces the number of structures by removing those that are least similar to the prediction molecule.
- Adjust the slider until there are roughly equal numbers of each group, probably at around 1200 structures. The wizard provides guidance as to the sizes of the datasets and whether or not these are suitable.

21.7.6 Running a Hydrogen Bond Analysis

- At this point you have a set of *fitting data* structures from the CSD containing chemistry relevant to the prediction molecule. A statistical model can now be built, first by running an analysis of the hydrogen bonding interactions in the dataset.
- This step automatically analyses the *fitting data* structures, identifies the H-bonds that are present, and collates statistics and descriptors on hydrogen bond formation. This raw data provides the basis of the model regression.
- If the **Start analysis automatically** checkbox has been ticked the hydrogen bond analysis will begin after the generation of the dataset, otherwise click **Analyse** to run the analysis. This step takes a minute or two as it runs through all the atom-atom contacts in the structures of the *fitting data* set.

- It is still possible to adjust the structures used in the dataset using the slider at this stage. Moving the slider to the left will reduce the number of structures included and no further analysis will be required. Moving the slider to the right increases the number of structures and consequently requires further analysis, which is performed by clicking **Analyse**.
- Once the analysis is complete a panel showing statistics will become available. This provides a summary of the number of true and false hydrogen bond observations for each donor and acceptor, i.e. the number of times a hydrogen bond is present (true) and absent (false) in the *fitting data* set.
- Numbers displayed in red show that a low number of observations were found. This might be an indication of potential problems in the model fitting step.
  Note: If problems are encountered during the process then the statistics table can be a good source of information for troubleshooting.
- To avoid the chance of regression failures, Mercury will automatically exclude individual atom types from the model fitting if they contribute a low number of observations. In cases where this happens, the relevant checkbox in the *Ignore* column will be checked. If regression failures are still observed you can exclude individual atom types yourself by enabling the relevant checkbox.
- Click the **Fit Model >** button in the Wizard to fit the hydrogen bond logit model to the data collected (see Fitting the Statistical Model).

21.7.7 Fitting the Statistical Model

- Logistic regression is used to determine the likelihood of hydrogen bond formation and is performed using the statistics program *R*, which is included by default with Mercury.

- In the *Model Fitting* pane of the wizard a model will automatically be generated using all potential model variables. When complete, the tool reports some regression analysis data.

logit_model_1 Coefficients:

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	Significance code	Lower Bound	Upper Bound
(Intercept)	**3.365**	0.191	17.608	2.13697e-69	***	2.994	3.743
Donorother	**0.016**	0.096	0.163	0.870501		-0.174	0.204
Acceptoratom_0_of_imidazole_3	**0.678**	0.157	4.329	1.49772e-05	***	0.373	0.987
Acceptoratom_1_of_cyclic_amide_6	**-0.925**	0.146	-6.341	2.2754e-10	***	-1.213	-0.641
Acceptoratom_2_of_carbamoyl_2	**-0.258**	0.137	-1.882	0.0597722	.	-0.526	0.010
Acceptorother	**0.375**	0.109	3.424	0.000616493	***	0.158	0.588
Competition	**0.020**	0.007	2.904	0.00368501	**	0.007	0.033
Donor_steric_density	**-0.016**	0.002	-8.343	7.22772e-17	***	-0.020	-0.012
Acceptor_steric_density	**-0.034**	0.002	-13.481	2.0232e-41	***	-0.039	-0.029
Donor_aromaticity	**-0.082**	0.232	-0.354	0.723065		-0.538	0.374
Acceptor_aromaticity	**-1.241**	0.205	-6.052	1.42616e-09	***	-1.643	-0.839
Donoratom_3_of_carbamoyl_2	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0(1)_of_cyclic_NdoubleN	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A

- These data can be used to assess the predictive capability of the model and the significance of the model coefficients. It is important that the model is good enough to proceed (see Assessing a Statistical Model).
- The model will, by default, be named *logit_model_1*. Models can be reused by selecting **Use existing regression data** during the initial steps of the calculation (see Starting the Propensity Prediction Wizard).
- To re-fit a propensity model, click the **Refine Model...** button in the *Model Fitting* pane of the wizard. This will open the *Select Model Variables* dialogue which allows you to choose which variables to include when fitting the hydrogen bond model. This means you can try out different variables and test whether they prove significant for prediction of hydrogen bond formation. Only those that prove to be significant should be retained for the final model.

- If you change the model variables it is a good idea to rename your model. This gives you the option of retaining several models which can be used to run a number of propensity calculations without having to start the process again each time.

21.7.8 Assessing a Statistical Model

- A well-defined model function is essential before predictions about likely H-bonds are made for a target structure. This section describes how to check a propensity model regression has successfully converged and provides recommendations for deciding whether to proceed or re-fit a propensity model.
- The model equation and a summary of goodness-of-fit can be found towards the bottom of the *Model Coefficient Statistics* tab within the *Model Fitting* pane of the wizard (shown below). This is the Area under the ROC curve, or AUC. ROC represents receiver-operating characteristics (a term from signal processing).
  Note: Due to the non-deterministic nature of the *Model Fitting* step, the results will vary from model fitting run to model fitting run.

logit_model_1 Coefficients:

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	Significance code	Lower Bound	Upper Bound
(Intercept)	**3.365**	0.191	17.608	2.13697e-69	***	2.994	3.743
Donorother	**0.016**	0.096	0.163	0.870501		-0.174	0.204
Acceptoratom_0_of_imidazole_3	**0.678**	0.157	4.329	1.49772e-05	***	0.373	0.987
Acceptoratom_1_of_cyclic_amide_6	**-0.925**	0.146	-6.341	2.2754e-10	***	-1.213	-0.641
Acceptoratom_2_of_carbamoyl_2	**-0.258**	0.137	-1.882	0.0597722	.	-0.526	0.010
Acceptorother	**0.375**	0.109	3.424	0.000616493	***	0.158	0.588
Competition	**0.020**	0.007	2.904	0.00368501	**	0.007	0.033
Donor_steric_density	**-0.016**	0.002	-8.343	7.22772e-17	***	-0.020	-0.012
Acceptor_steric_density	**-0.034**	0.002	-13.481	2.0232e-41	***	-0.039	-0.029
Donor_aromaticity	**-0.082**	0.232	-0.354	0.723065		-0.538	0.374
Acceptor_aromaticity	**-1.241**	0.205	-6.052	1.42616e-09	***	-1.643	-0.839
Donoratom_3_of_carbamoyl_2	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0(1)_of_cyclic_NdoubleN	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A

- As the prediction is a 2-state true-false outcome, the ROC curve (shown below) provides a convenient measure of correct predictions for both true and false, or discrimination. If you are familiar with linear regression, you will see that there is no linear relationship in logistic regression to obtain usual R^2 fitting statistics, or similar; AUC is a useful equivalent.
- In the ROC curve for a good model, the rate of correct true predictions (the vertical axis) should be steep, while the rate of incorrect false predictions is as low as possible. In this situation, the area under the curve is maximised.

As a general rule:

If ROC = 0.5	no discrimination
If 0.7 ≤ ROC < 0.8	acceptable discrimination
If 0.8 ≤ ROC < 0.9	excellent discrimination
If ROC ≥ 0.9	outstanding discrimination

- It is usually possible to train an H-bonding model with an AUC ≈ 0.8. In this case you are well prepared to continue making predictions for your specified target. If the AUC is 0.7 or lower, it is often best to go back to refine the fitting data. Your sample of structures may be too diverse, i.e. no strong trends in the behaviour of the chemical groups in the set can be determined.
- The model coefficients and fitting statistics for each parameter can be found in the main table of the *Model Coefficient Statistics* tab (shown below). If you have sub-optimal AUC, an important indicator to check is the significance codes represented by *. Three stars (***) is optimal significance, zero () is the lowest significance. It is often the case that one (or two) parameters do not have *** significance. However, if more than one or two of the parameters have a significance code less than two stars it may indicate the training set needs to be refined.

logit_model_1 Coefficients:

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	Significance code	Lower Bound	Upper Bound
(Intercept)	**3.365**	0.191	17.608	2.13697e-69	***	2.994	3.743
Donorother	**0.016**	0.096	0.163	0.870501		-0.174	0.204
Acceptoratom_0_of_imidazole_3	**0.678**	0.157	4.329	1.49772e-05	***	0.373	0.987
Acceptoratom_1_of_cyclic_amide_6	**-0.925**	0.146	-6.341	2.2754e-10	***	-1.213	-0.641
Acceptoratom_2_of_carbamoyl_2	**-0.258**	0.137	-1.882	0.0597722	.	-0.526	0.010
Acceptorother	**0.375**	0.109	3.424	0.000616493	***	0.158	0.588
Competition	**0.020**	0.007	2.904	0.00368501	**	0.007	0.033
Donor_steric_density	**-0.016**	0.002	-8.343	7.22772e-17	***	-0.020	-0.012
Acceptor_steric_density	**-0.034**	0.002	-13.481	2.0232e-41	***	-0.039	-0.029
Donor_aromaticity	**-0.082**	0.232	-0.354	0.723065		-0.538	0.374
Acceptor_aromaticity	**-1.241**	0.205	-6.052	1.42616e-09	***	-1.643	-0.839
Donoratom_3_of_carbamoyl_2	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0(1)_of_cyclic_NdoubleN	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A

- The quantitative parameters (e.g. donor steric density, donor aromaticity, competition) can also be checked. If one of these has low significance, it can be because there is high correlation (e.g. between donor and acceptor aromaticity). In this case, uncheck one from the *Select Model Variables* dialog, and repeat the model fitting.
- To accept the model click **Accept & Calculate >** to perform the propensity calculation and visualise the results (see Using the Hydrogen Bond Propensity Results Table).

21.7.9 Visualising the results of a Hydrogen Bond Propensity calculation

- The hydrogen bond propensity calculation will be performed automatically, and the results will be shown on the *Calculate Propensities* pane of the wizard.
- The *Propensity scores* table is populated with donor-acceptor pairs sorted by their propensity to form a hydrogen bond. Both inter- and intramolecular pairings are shown.
- The *Co-ordination scores* table shows the individual atom likelihoods for different hydrogen bond coordination environments.
- The chart represents all feasible hydrogen bond combinations for the system, plotted with respect to their mean propensity scores and mean co-ordination scores.
- The *Recalculate* section allows you to perform the Hydrogen Bond Propensity calculation again, on a different polymorph of the same system or using a different model.

21.7.10 Using the Hydrogen Bond Propensity Results Table

- The *Propensity scores* table is populated with donor-acceptor pairs sorted by their propensity to form a hydrogen bond. Both inter- and intramolecular pairings are shown.

- A propensity value of 0.8 corresponds to 80% probability of formation for the corresponding interaction.
- In the example below, the most likely intermolecular hydrogen bond is identified as being between N6 of the carbamoyl and N2 of the imidazole. The least likely is between N6 of the carbamoyl and O2 of the cyclic amide.
- Where a hydrogen bond is actually observed in a target crystal structure the word **observed** appears in the column for the corresponding structure, indicated by the target structure's identifier in the column heading. In the example below the second most likely carbamoyl-carbamoyl interaction is observed.

Propensity scores

Hydrogen-bond propensities for DIPGIS12 calculated using 'logit_model_1'

	Donor	Acceptor	Propensity	DIPGIS12
1	INTER-molecular			
2	N6 of carbamo...	N2 of imidazole...	**0.82**	
3	N6 of carbamo...	O1 of carbamo...	**0.75**	observed
4	N6 of carbamo...	N5 of cyclic_Nd...	**0.71**	
5	N6 of carbamo...	N4 of cyclic_Nd...	**0.68**	
6	N6 of carbamo...	O2 of cyclic_a...	**0.58**	
7	INTRA-molecular			
8	N6 of carbamo...	N4 of cyclic_Nd...	**0.90**	
9	N6 of carbamo...	N2 of imidazole...	**0.34**	

Left-clicking table items visualizes H-bond groups/contacts and interacts with chart

Atom Highlighting...

- This table is interactive. Clicking on an atom label in either the *Donor* or *Acceptor* columns will highlight the functional group and label the atom in the Mercury visualiser.
- Clicking on **observed** in the target structure column highlights both the donor and acceptor group.
- The highlighting of functional groups in the visualiser can be customised. To do this click on the **Atom Highlighting...** button and specify the desired settings in the resulting *Hit Highlighting* dialogue (see Controlling Hit Highlighting).
- In the example above the most likely carbamoyl-imidazole interaction (i.e. the pairing with the highest propensity value) is not observed. Why is this? Could it form in place of the observed H-bond, or even alongside it using the second donor hydrogen? Does this indicate there is sub-optimal hydrogen bonding in the structure? We can explore how the combinations of hydrogen bonds fit together using coordination likelihoods (see Charting Possible Hydrogen Bond Combinations).
- If you have applied the propensity model to multiple structures, then you will obtain a corresponding number of *Observed?* columns in the results table.

When you are satisfied that you have completed your hydrogen bond propensity analysis, you can produce a formatted *.html* report by clicking on the **Publish Report…** button. This will open a dialogue where you can name the report file and save it to your desired location. Once you have saved your report, or if you wish to close the hydrogen bond propensity dialogue without saving, click **Close**.

21.7.11 Recalculating Hydrogen Bond Propensities

- You may wish to run the hydrogen bond propensities calculation again, for example for a different polymorph of the same molecule. To do this, use the *Recalculate* section in the *Calculate Propensities* pane of the wizard.

- This section allows you to:
  - Confirm that the correct model is being applied to your prediction molecule. A model saved from a previous session could instead be used by clicking on the **Logit model** drop down menu, selecting **Load from file...**
  - Confirm the prediction molecule against which to run the hydrogen bond propensity calculation. To apply the calculation to a different structure (e.g. an alternative polymorph of the prediction molecule) click on the **Target structure(s)** drop down menu and select **Load from file...**
  - To apply the calculation to multiple structures (i.e. several members of the same polymorph family), click on **Select multiple…** from the **Target structure(s)** pull-down menu. Use the resultant dialogue to select the structures (if loaded) or to browse to a directory where they are stored. *Note:* The chemical composition of all the structures must be the same. Additionally, the structure on which the propensity model was built should be listed first.
- Click **Recalculate** to perform the desired Hydrogen Bond Propensity recalculation.

21.7.12 Charting Possible Hydrogen Bond Combinations

- The **c**hart on the left-hand side of the *Calculate Propensities* pane of the wizard represents all feasible H-bond combinations. We can refer to each data point as a possible structure, with unique H-bonding.

- The points on the chart are colour coded based on the number of H-bond pairs.
- The horizontal (x) axis represents the mean propensity (labelled as *Mean H-Bond Propensity*) for the H-bonds in a structure, directly related to the values in the propensity results table (see Using the Hydrogen Bond Propensity Results Table).
- The vertical axis represents the mean co-ordination likelihoods for the atoms in the structure (labelled as *Mean H-Bond Co-ordination*). These co-ordination likelihoods are shown in the *Co-ordination scores* section of the *Calculate Propensities* pane of the wizard (see below). The scale is relative to a zero base-line so that the most negative values are optimal.
  Note: The values are reported as negative for display purposes. The likelihoods themselves and the overall score can be in the range 0 to 1.
- It is possible to change the number of points on the chart by adjusting the sliders for minimum donor/acceptor coordination likelihoods. Moving the sliders to the right will require H-bond combinations to have higher coordination likelihoods, and consequently fewer points will appear on the chart.

Co-ordination scores

(To refresh table: left-click chart point)

	Atom (D/A)	= 0	= 1	= 2	= 3
1	N6 of carbamo...	0.004	0.055	0.878	0.063
2	N2 of cyclic_n (a)	0.367	0.621	0.012	0.000
3	N4 of cyclic_Nd...	0.744	0.256	0.000	0.000
4	N5 of cyclic_Nd...	0.838	0.162	0.000	0.000
5	O1 of carbamo...	0.118	0.737	0.136	0.009
6	O2 of NH0_ure...	0.589	0.384	0.027	0.000

- For a given data point related to a possible H-bonding outcome, the *Mean H-Bond Co-ordination* is calculated as a mean (average) of the observed co-ordination likelihoods for each donor or acceptor atom in that outcome. Values are coloured green or red for that outcome in the *Co-ordination scores* table to indicate optimal or sub-optimal scores. In the same way, the *Mean H-Bond Propensity* is calculated as a mean of the likelihoods of the observed H-bonds in that possible structure.
- In this way, optimal H-bonding is represented by data points in the lower-right corner of the plot, and sub-optimal H-bonding is represented in the upper-left corner.
- Hovering the mouse over a data point on the chart will show the *Mean H-Bond Propensity* and *Mean H-Bond Co-ordination* values for that data point as a tooltip.
- The *H-bond Assessment Chart* is interactive. You can zoom in on any region of the chart. This is useful, for example, to inspect more closely a crowded region to allow easier selection of a data point. To do this click on the *Magnify* icon underneath the chart then left-click in the display area and, keeping the mouse button depressed, drag the mouse to enclose the required region. It is possible to zoom in multiple times by repeating the above procedure. To return to the original chart click on the **Reset** button.
- Clicking on a data point provides two main features:
    - A breakdown of individual atom likelihoods for H-bond co-ordinations of zero, once, twice etc. This information is displayed in the *Co-ordination scores* table beside the chart and allows you to see which donors or acceptors are behaving optimally in this combination.
    - Highlighting of the corresponding hydrogen bonds in the *Propensity scores* table (see Using the Hydrogen Bond Propensity Results Table).
- The data point relating to the observed structure is always coloured magenta. If you have applied the propensity model to multiple structures, you will see a corresponding number of coloured data points in the chart.
- The selected data point is identified by a **^** symbol. Clicking on the data point corresponding to the observed structure in the above example would display the coordination likelihoods shown in the *Co-ordination scores* table above.
- Cells coloured green or red are the observed outcomes for each atom. Green is the optimal outcome, red is sub-optimal (e.g. O of carbamoyl is most likely to accept once). We can immediately see that the observed structure optimally satisfies most of the donors and acceptors.
- Should you see many more red cells in the coordination table for a particular data point then this indicates the atoms in that structure are not participating in an optimal way.
- All numerical data associated with the analysis including the chart are output to the working directory, defined in the *Settings* page at the start of the wizard. Note the chart data are output to a .tsv file.

21.8 Quick Assessment of H-Bond Coordination

- To access the H-bond Coordination Quick-view component, click on **CSD-Materials** in the top-level menu and select **Polymorph Assessment** then **H-bond Coordination Quick-view...** from the resulting menus.
- This will open a dialog box which enables quick and easy hydrogen-bond likelihood analysis using coordination numbers for the observed structure.
- The feature enables you to quickly assess the likelihood of hydrogen-bond behaviour purely on coordination numbers for the loaded structure.
- Calculated likelihoods are displayed for permitted coordination numbers for each donor and acceptor observed in the current structure. These are computed using CSD-derived models. The coordination likelihood models used here are the same as those used in the Hydrogen Bond Propensity component.
- The coordination numbers for each atom are highlighted based on their likelihood. Green highlighting indicates a maximum likelihood is observed whilst red highlighting indicates there is a more likely alternative coordination number for that atom.
- Results can be saved by hovering over the table and copy-and-pasting into your preferred text editor. Alternatively, you can save the coordination likelihood in HTML format.

21.9 Assessing Molecular Interaction Preferences Using the CSD

21.9.1 Introduction

- Full Interaction Maps provide your molecule's interaction preferences at the click of a button and in the context of the observed crystal structure. This means the molecule's preferred interactions can be visualised in 3D and the intermolecular packing of a crystal structure can be evaluated quickly and easily.
- The tool calculates regions around the molecule (maps) where chemical probe groups are likely to be found, based on pre-extracted IsoStar interaction data from the CSD (Bruno et al., *J. Comput.-Aided Mol. Des.*, 1997, **11**, 525-537). The calculation procedure first identifies distinct functional groups in the molecules viewed, and then finds relevant interaction data in IsoStar. Next it pulls together the group-based interaction data and takes into account the environmental effects of combinative factors and steric exclusion to create a full 3D picture of molecular interaction preferences. More details of the fundamental methodology are given in the following paper: Wood et al., *CrystEngComm*, 2013, **15**, 65-72.
- Full Interaction Mapping allows you to assess how the crystal packing environment can satisfy the interactions of the functional groups of the molecule. It can be used to:
    - Examine and assess the different interactions which are observed in polymorphs.
    - Find a region of a molecule where a co-former interaction may result.
    - Explore the likely interactions of a molecule prior to crystallization.
    - Map the interaction preferences of a slab of molecules from a known crystal to identify where a solvent or an inhibitor may interact.

21.9.2 Setting up a Calculation

- Click on **CSD-Materials** in the top-level menu and select **Full Interaction Maps…** from the resulting menu. This will open the *Full Interaction Maps* dialogue box:

- The following options can be found under the *Options* tab:
  - **Map contour levels**: This will control firstly how many contour surfaces to display (one, two or three) and secondly at what level to generate the contours. For a given probe type, the contour level will indicate the area mapped out by a certain interaction likelihood. A contour level of 6, for example, means the surface will enclose regions where the density of

contacts in the underlying CSD data is 6 times that which you would expect at random.

- **Hotspots**: Enable this option to generate hotspots for each probe type which represent the positions of highest local density for each contour surface (see Use of Hotspots).
- **Probes**: The default setting for FIMs is to generate three sets of contour maps with specific probe types. The probe types have been chosen as proxies to reflect molecular interaction preferences with respect to H-bond donors (Uncharged NH Nitrogen), H-bond acceptors (Carbonyl Oxygen) and hydrophobic groups (Aromatic CH Carbon). There are 17 probe types available in total. The colour bar next to the probe name indicates the colour of the maps that will be generated for that probe – to modify the colour for a given probe, simply click on the colour bar.

- Note that Full Interaction Maps will be calculated around all molecules or fragments that are currently visualised in Mercury unless some atoms have been selected, in which case maps will only be calculated around the selected atoms.
- For accurate interaction preference maps to be generated, it is important for the correct protonation of the molecules to be expressed in the 3D model and for bond types to be assigned for all molecules. The dialogue will prompt the user to edit or assign bond types if any are not known at the point of map calculation.
- Full Interaction Maps will only be calculated (or re-calculated) when the **Calculate Maps** button is clicked. All maps and associated hotspots can be cleared from the display by clicking **Clear Maps & Hotspots**.
- The settings in this dialogue box can also be returned to default values by clicking on the **Defaults** button.
- Finally, Full Interaction Maps can be saved out or loaded using the **Save Maps…** and **Load Maps…** buttons. The **Save Maps…** option requires the selection (or creation) of an empty directory to save into. The **Load Maps…** option allows you to reload a previously saved map and use the original molecular aggregate for the calculation.

21.9.3 Visualisation and Analysis

21.9.3.1 Display of maps

- Once calculated, the appearance of Full Interaction Maps can be controlled from the *Maps* tab of the **Full Interaction Maps** dialogue.
- The contour level, colour, visibility, display type and opacity of the maps can be individually modified for each contour surface relating to each probe type.
- Note that, when used, opacity provides an illustrative guide to show which contour surface relate to the lower propensity levels.
- It is also possible to toggle the visibility of all the maps at once using the **All** checkbox above the **Visible** column.
- One or more contour surfaces can also be deleted from the display by selecting the relevant row(s) in the table (multi-row selection is possible using shift-click or control-click) and clicking on the **Delete** button.

21.9.3.2 Use of Hotspots

- When calculated, hotspots indicate the points of highest local interaction density found for each contour surface and the **Hotspots** tab allows control of the display of these hotspots.
- The minimum propensity level of hotspots to display, as well as colour, visibility, labelling (by propensity value) and opacity can be individually modified for each probe type.
- Note that, when used, opacity provides an illustrative guide to show what the range of propensity values currently visualised (less opaque means lower propensity).
- Visibility, labelling and opacity can also be toggled for all probe types at once using the **All** checkboxes above each of those columns.
- All visualised hotspots can also be converted to dummy atoms using the **Convert Hotspots to Dummy Atoms** button. This can be useful in particular for creating

Crystal Packing Feature searches to find molecules which match a given pattern of interaction preferences (e.g. possible co-formers or surface inhibitors).

21.9.3.3 Log File Output

- The **Log Files** tab contains the output from the maps calculation as well as a report of any errors found during the calculation.
- The *Log Output* can be viewed to understand which underlying central groups have been matched in the IsoStar library from the selected set of molecules.
- The *Error Output* should be routinely checked to assess whether any problems were found during the analysis. There are two main error message types to be aware of:

```
Warning: match3diso: rms match is 0.1918 (Å) for central group 5
(ethenyl)
```

```
Warning: Substructure 0 RES1 has nonmatched atoms - incomplete
residue or erroneous structure. Warning: check_target_atoms: 10:
N1 (substr RES1): not a target atom.
```

- The first of these means that a central group has been matched, but not overlaid in 3D upon the target molecule very effectively. This is not ideal, but the results should still be relatively accurate.
- The second type of error message means that one or more atoms (check target atoms listed) have not been matched to a central group. This is a more serious problem and means that there will be a gap in the Full Interaction Map. Care should be taken when interpreting such a gap in the maps as a negative interaction preference, but instead just as a current lack of data for this type of chemistry.
- Note that metals are not currently treated by Full Interaction Maps, so if there is a metal in your structure, this will not be matched to a central group and will potentially result in a gap in the contour surfaces.

21.10 Co-Crystal Design

21.10.1 Introduction

- Multi-component crystal forms, and co-crystals in particular, are of interest in both academic research and industrial contexts due to their potential for producing materials with modified physical and chemical properties compared to the pure form.
- There are many different ways to approach the design of a co-crystal screen, but the use of knowledge-based methods is well-established. This includes techniques such as the following:
    - Manual searches based on hydrogen-bonding or halogen-bonding motifs (see Searching on Motifs)
    - Co-crystal design by molecular complementarity (see Screening by Molecular Complementarity)

21.10.2 Screening by Molecular Complementarity

21.10.2.1 Overview

- A method for assessing the likelihood of two molecules to form a co-crystal based on the comparison of each of their molecular descriptors has been developed and validated by Fábián. Surveys of observed co-crystals in the CSD were performed to assess any statistical correlations between the properties of the two molecules in a given co-crystal. The observed data indicates that molecules which crystallise together tend to have similar molecular properties.
- For co-crystallisation to be likely, five key molecular descriptors have been identified for which the difference between the values for the two co-crystal components should be small. These descriptors are the fraction of nitrogen and oxygen atoms, the dipole moment and three simple shape descriptors based on a molecular bounding box – the length of the short axis, the short/long axis ratio and the medium/long axis ratio.
- *Note*: This methodology has only been validated for neutral multi-component systems – it is not appropriate for use with salts.
- For more details on the original study and methodology, please see Fábián, *Cryst. Growth Des.*, 2009, **19**, 1436-1443.

21.10.2.2 Setting up a Calculation

- Click on **CSD-Materials** in the top-level menu and select **Co-Crystal Design** then **Screen by Molecular Complementarity…** from the resulting menu. This will open the **Molecular Complementarity Screening Wizard**:

- The following options are available here:
  - **Descriptor Settings**: These options control which molecular descriptors are used for both the calculations and the designation of passes and fails in the virtual co-crystal screen. The default setting is to use all five descriptors. For

advanced uses it is possible to deselect one or more of these descriptors, although care should be taken in such cases.

- **Output Settings**: The settings here control the choice of files that are produced by the calculation. Options are to output files for the **Molecules which pass** the molecular descriptor screen and/or the **Molecules which fail**. Additionally, molecules can be output as multi-MOL2 files or as a folder of individual MOL2 files. All of the output files will be created in a date-stamped folder within the chosen **Working Directory**.

- After clicking the **Next** button users are able to select the target molecule(s) or Active(s) as well as the Co-former(s) to use for virtual co-crystal screening.

- Actives and co-formers can be loaded from any molecular files available such as MOL2 files by clicking on the **Load…** button, or selected from the currently viewed structure by clicking on the **Use Current** button.
- Multiple conformations can be supplied for each given active or co-former. To do this, supply one multi-MOL2 file per compound and name each molecule within that multi-MOL2 file using a fixed starting string, then an underscore and five numbers counting up from 00001, e.g. ‘paracetamol_00001’, ‘paracetamol_00002’, ‘paracetamol_00003’.
- *Note*: If using the output from the CSD Conformer Generator (see Generating Conformers Using the CSD), simply load in the multi-MOL2 file containing the generated conformers. The naming of molecules within the multi-MOL2 file will already be in the correct format for use here.
- 2D chemical diagrams will be generated and displayed for each active or co-former file loaded as a check to confirm which molecules will be used for the calculations.

- Once the actives and co-formers have been selected, to calculate the molecular descriptors and assess the likelihood of co-crystallisation based on these descriptors, click on the **Next** button.

21.10.2.3 Interpreting the Results

- The overall hit rate of each co-former with each active (including assessment of multiple conformations if supplied) will be presented in the form of a simple table of percentage pass rate for that active/co-former pair.

- The values of the molecular descriptors calculated (including multiple conformations where present) and the comparison of the descriptors will be output to the working directory.
- Full details output to the working directory include the following:
  - A summary spreadsheet for the full collection of screening calculations labelled by time stamp. The hit rates are summarised for all actives in a single spreadsheet with a tabulated view of the pass / fail status for each conformer pair. A further summary tab in the spreadsheet is provided for each active.
  - One additional sub-folder per active containing: the MOL2 files chosen for output in the configuration settings; the details of the descriptor values calculated for each conformer; the difference in descriptor values (delta) between each active and coformer pairing. These details can be found in the folder named “[name of active]_results” (e.g. “HXACAN_results”). This information can be used in the tailoring of descriptor settings for advanced uses, although care should be taken in such cases.

21.11 Analysing Hydrate Structures

21.11.1 Introduction

- The Hydrate Analyser provides the user with the ability to quickly and easily analyse the often complex nature of hydrated crystal structures using the wide variety of tools available through Mercury.
- The Hydrate Analyser provides insight into the hydrogen bonding network in the hydrated structure, as well as the spacefilling environment of the water molecules and the interaction preferences of a water molecule in the structure.

- The input structure for analysis is selected from the Mercury Structure Navigator i.e. a CSD entry, in-house database entry, or a structure loaded into Mercury from a 3D crystallographic data file such as a .cif, .res or .mol2 file.

21.11.2 Using the Hydrate Analyser

- To start the Hydrate Analyser, click on **CSD-Materials** in the top-level menu and select **Hydrate Analyser…** from the resulting menu.
- On running the tool, a new window appears containing a set of tabs displaying the various aspects of the analysis and through which the user can interact with the visualised structure:
    - **Structure Overview**: provides a configurable summary for the analysis that can be saved to a report (see Structure Overview).
    - **Water H-Bonding**: displays an automatic analysis of the hydrogen bond motifs involving the water molecules, with an option to visualise these motifs in the crystal lattice (see Water H-Bonding).
    - **Water Space**: functionality to calculate and display the space occupied by the water molecules (see Water Space).
    - **Water Interaction Maps**: a simplified interface to the usual Full Interaction Mapping tool for the calculation and display of the water probe interaction maps (see Water Interaction Maps).
    - **Coordination Polymer**: control the visualisation of metal-coordinated hydrates (see Coordination Polymer).

21.11.3 Structure Overview

- The Structure Overview tab provides the user with a tabulated summary of the analysis. This includes:
    - a labelled chemical diagram for the asymmetric unit
    - the water hydrogen bond geometry and motifs detected (for more details see Water H-Bonding)
    - information on the volume of the structure occupied by water (once this has been calculated through the Water Space tab)
    - a configurable list of additional structural and bibliographic details, the display of which is controlled via the **Configure…** button
- A status bar indicating whether a hydrate has been successfully identified is located at the top of the **Structure Overview** tab (green symbol shown if water has been identified and red symbol shown if not).

- Structures loaded from a .cif are automatically edited to standardise atom and bond types, which is necessary for the water hydrogen bond motif searches and Full Interaction Map calculations to proceed correctly. In the case of automatic editing, a message is shown above the status bar. The user has the option to disable this procedure by clicking the **Disable auto-edit** checkbox at the bottom of the tab.
- The contents of the **Structure Overview** tab can be simplified by clicking on the **Tree View** checkbox:

- The **Save As…** button allows the user to save the displayed table to a .doc (Word) or .html file.

21.11.4 Water H-Bonding

- The **Water Environments** side tab shows the results of an automatic search for the presence of any of 10 common water hydrogen bonding motifs, as defined in this paper: Gillon *et al.*, *Cryst. Growth Des.*, 2003, **3**, 663-673.
- Clicking on a row of this table will highlight the identified motif in the Mercury visualiser. In the case of multiple water molecules displaying the same motif, the hits can be toggled through using the **Browse multiple hits** selector to the right of the table.
- The style of motif highlighting can be controlled via the **Highlighting style…** button.
- The geometric details of any observed motifs are shown in the **Contacts** side tab.
- Input structures where the hydrogen atoms of the water molecules have not been determined will lead to ambiguous results, and the tool will generate a warning to highlight this.

21.11.5 Water Space

- Click on the **Calculate…** button to begin the water space analysis using the default parameters.
- This calculation is performed using the same method as that for calculating void space in crystal structures (see Displaying Voids), and the parameters for such a calculation can be controlled from this tab.
- The display of water molecules, all molecules and water space can be controlled through use of the appropriate **Show**/**Hide** check boxes.
- The display of the water space can be customised using the coloured panels and transparency sliders.
- The settings in this tab can be returned to default values by clicking on the **Defaults** button.
- Finally, the volume occupied by water, both in $Å^3$ and as a percentage of unit cell volume, is shown in the **Results** section at the bottom of this tab. This information is also displayed in the **Structure Overview** tab once the calculation is complete.

21.11.6 Water Interaction Maps

- This provides the user with a simplified version of the Full Interaction Maps dialogue (see Assessing Molecular Interaction Preferences Using the CSD).
- Maps are generated exclusively for the water probe, and as such, the user should first select the whole molecule under study without water molecules before clicking **Calculate…**.
- Usage and control of this dialogue are the same as for the standard Full Interaction Maps tool and clicking on the **Advanced Options** button will open the standard **Full Interaction Maps** dialogue.

21.11.7 Coordination Polymer

- The Coordination Polymer tab allows the user to investigate metal-coordinated hydrates.
- To expand sections of a polymeric structure, click on individual atoms or bonds in the structure (for more details, see Polymer Expansion).

21.12 Analysing Solvate Structures

21.12.1 Introduction

- The Solvate Analyser provides the user with the ability to quickly and easily analyse complex solvate crystal structures, including structures with more than one solvent, a mixture of solvents, co-formers and counter-ions or even disordered solvate structures.
- The Solvate Analyser provides insight into the hydrogen bonding networks in the solvated structures, if present, as well as the spacefilling environment of the solvent molecules in the structure.
- The input structure for analysis is selected from the Mercury Structure Navigator i.e. a CSD entry, in-house database entry, or a structure loaded into Mercury from a 3D crystallographic data file such as a .cif, res or mol2 file.

21.12.2 Using the Solvate Analyser

- To start the Solvate Analyser, click on **CSD-Materials** in the top-level menu and select **Solvate Analyser…** from the resulting menu.
- On running the tool, a new window appears containing a set of tabs displaying the various aspects of the analysis and through which the user can interact with the visualised structure:
    - **Solvent Selection and Space Calculation**: allows the user to add individually selected solvents to a list, calculate and display the space occupied by the solvent molecules (see Solvent Selection and Space Calculation).
    - **Solvent H-Bonding**: lists the hydrogen bonding motifs displayed by the solvent molecules (see Solvent H-bonding).
    - **Structure Summary**: provides a summary of the solvate analysis that can be saved as report (see Structure Summary)

21.12.3 Solvent Selection and Space Calculation

- The **Solvent Selection and Space Calculation** tab allows the user to individually select different solvent molecules within the structure to analyse independently.
- The first step is to select one or more complete molecules in the 3D visualiser, then click on the **Add Solvent From Selected** button to register that group of atoms as a specified solvent. Information will then be displayed in the table for this solvent, additional solvent molecules can also be specified.
- The first row, called *Solvent*, is interactive and allows the user to rename the solvents.
- The display of the solvent molecules, solvent space and solvent selection can be controlled through use of the appropriate **Show Solvent/Show Space/Select Solvent** check boxes.
- The display of the void space occupied by each identified solvent can be customised using the colour panels and *Inside/Outside Opacity* sliders. Solvent can be shown or selected using the interactive checkboxes in the table. In the *Settings* section, the packing, void space probe radius and grid spacing can be easily manipulated by the user.
- The volume occupied by the solvents, both in $Å^3$ and as a percentage of unit cell volume, is shown in the *Results* section.
- Click on the **Calculate Space** button to display the space occupied by the solvents using the selected parameters.
- This calculation is performed using the same method as that for calculating void space in crystal structures (see Displaying Voids). The specific parameters for the calculation can be controlled from within this tab.

21.12.4 Solvent H-bonding

- The **Solvent H-Bonding** tab lists all of the symmetry-unique hydrogen bonds in the structure which involve one or more solvent molecules as either the donor or the acceptor. In each case, the nature of the molecule donating and the molecule accepting will be listed (e.g. non-solvent to methanol, or methanol to non-solvent). If provided, the names given to the solvent molecules in the first tab will be displayed here rather than simply solvent 1 or 2.
- The **Solvent H-Bonding** table lists the donor and acceptor atom labels, the length of the H-bond from H...A in Å, the length minus the sum of the van der Waals radii in Å and the symmetry operators used to generate the donor and acceptor molecules.

21.12.5 Structure Summary

- The **Structure Summary** tab provides a summary of the solvate analysis. This includes:
  - the refcode or structure identifier
  - a labelled chemical diagram of the asymmetric unit
  - a table of the observed H-bonds involving solvent
  - information on the volume of space occupied by each solvent as well as the total if more than one solvent has been identified
  - configurable list of additional structural and bibliographic details, the display of which is controlled via the **Configure…** button

- The **Save As…** button allows the user to save the displayed table to Word document (*.doc) or HTML format (*.html).

21.13 Generating Conformers Using the CSD

21.13.1 Introduction

- The CSD Conformer Generator provides the ability to both minimise molecular conformations and also generate diverse conformer subsets based on CSD data.
- The ability to generate likely conformers of a molecule is particularly useful when assessing the interaction preferences of a molecule in 3D prior to the observation of a solid-state structure of that molecule (see Assessing Molecular Interaction Preferences Using the CSD). A diverse subset of conformations can also be very valuable when performing an *in silico* co-crystal screen based on molecular complementarity (see Co-Crystal Design).
- The methodology starts from an input 3D molecular structure with all hydrogen atoms present, which is optionally minimised in the first step. After that, conformations are sampled based on CSD-derived rotamer distributions and ring templates. A final diverse set of conformers, clustered according to conformer similarity, is returned. Each conformer is locally optimised in torsion space.
- For more detailed information, see the specific CSD Conformer Generator documentation provided within the CSD software. You can find this within the CSD installation directory – on Windows for example this will be something like:

  C:\Program Files\CCDC\CSD_2020\Conformer Generator\docs\conformer_generator.pdf

21.13.2 Performing a Calculation

- Click on **CSD-Materials** in the top-level menu and select **Conformer Generation…** from the resulting menu. This will open the *Conformer Generation* dialogue box:

- A CSD entry, in-house database entry, or other structure loaded into the Mercury visualiser can be used as the input for conformer generation. Alternatively, you can

select a file on disk as the input by selecting **From file** and clicking on the **Browse** button.

- The working directory can be selected under **Working Directory**. A folder will be created in this directory named after the identifier of the input molecule used. Both a copy of the input molecule and the generated conformers will be written to the output folder.
- Conformers may be generated in MOL2 or SDF format, or both, by choosing the relevant checkboxes under **Output format(s)**.
- Click on **Calculate** to generate conformers.

- Generated conformers will be automatically loaded into the **Structure Navigator** tab. One option to assess the range of conformations generated is to create a tailored overlay using the Crystal Packing Feature functionality (see Searching on Crystal Packing Features) in combination with Multiple Structures (see Displaying Multiple Structures Simultaneously).

21.13.3 Advanced Options

- For fine control of the conformer generation process, advanced options can be accessed by clicking on the **Show advanced options** checkbox in the **Conformer Generation** dialogue.

- The following options are available here:
  - **Maximum Number of Conformations**: The default maximum number of conformations is 200, but if you are working with a particularly complex molecule with a large number of rotatable bonds you may want to increase this. Note that the program will only generate distinct conformers, so fewer than this number of conformers may be produced if the search space is exhausted.
  - **Maximum Unusual Rotamers**: The maximum number of unusual rotamers permitted in a molecule. The default value of this is 2.
  - **Minimum Rotamer Probability**: The minimum rotamer probability (how likely a rotamer must be in order to be regarded as probable). The default value is 0.05.
  - **Maximum Conformations to Sample**: The maximum number of conformations to generate for clustering. The default value of this is 1,000,000.
  - **Minimise input molecule**: The default behaviour of the CSD Conformer Generator is to minimise the input molecule using CSD-based information prior to generation of conformers. You can disable this option to skip the initial minimisation step for the input molecule.

21.14 Structure Solution from Powder Data

- In cases where single crystals are unavailable, or analysis of a bulk sample is preferred, structure determination from powder diffraction data is a valuable and accessible alternative option to structure solution from single crystal diffraction data.
- The ability to determine crystal structures from powder diffraction data is provided in CSD-Materials by the program DASH.
- To perform structure solution from powder data, click on **CSD-Materials** in the top-level menu and select **Launch DASH** from the resulting menu.

- The following options will be available:
  - Indexing diffraction patterns
  - Pawley refinement
  - Structure solution using simulated annealing
  - Guidance of solution using molecular geometry information from Mogul
  - Rietveld refinement
- For more detailed information, see the specific DASH documentation provided within the CSD software.

21.15 Post-Search Options

- Once a CSD-Materials search has been completed, the Post-Search Options tab will appear in the Structure Navigator. This tab provides a set of easily accessible options that are typically used after a search has finished.

- From the *Post-Search Options* tab you can do the following:
  - **Save Results:** Allows the saving of the search file (*.msxml*), a search summary file (*.txt*) or a refcode list (*.gcd*) (see Saving Searches and Results).
  - **Edit Search:** If clicked, this button re-opens the Wizard with all the parameters and search information from the search that has just finished. This means that it is very easy to re-run a search with a small change in some settings or search parameters. The **Edit Search** option therefore makes it a lot faster to execute multiple related searches.
  - **Filter Results:** This is only available if a Motif search has been performed. The menu provides options for filtering search results based on motifs using Boolean operators (see Filtering Motif Search Results).
  - **Start New Search:** This re-opens the Wizard which allows you to start any one of the three different search types.

21.16 Controlling Hit Highlighting

- For motif search results visualisation of the hit fragment and the contact motif can be customised by clicking on the **Options** button in the *Searches* tab of the *Structure*

Navigator and choosing **Hit Highlighting...** in the pull-down menu. This will open the *Hit highlighting Style* dialogue:

- Enable the **Apply highlighting** check-box.
- Select the required display style (Capped Stick, Ball and Stick, etc.) and the colour of the atoms in the hit fragment. The **Show only hit atoms** check-box can be used to hide all other parts of the structure(s) being displayed.
- The colour of motif contacts can also be customised.
- For crystal packing feature, crystal packing similarity, and hydrogen bond propensity search results visualisation can be customised in a similar way.

21.17 Saving Searches and Results

- To save the current search, do either of the following:
  - Click on the **Options** button in the *Searches* tab of the *Structure Navigator* and select **Save <search name>** from the resulting pull-down menu.
  - Click on the **Save Results...** button in the *Post-Search Options* tab of the *Structure Navigator*.
- The following types of search can be saved:
  - Motif search (see Searching on Motifs).
  - Crystal packing feature search (see Searching on Crystal Packing Features).
  - Crystal packing similarity (see Determining Crystal Packing Similarity).
- All search types will be saved with the extension *.msxml*.
- Search (*.msxml*) files contain all the information necessary to read the search back into Mercury and display it as if it had just been run. To read a previously saved search, select **File** from the top-level menu followed by **Open** and then select the search file. Alternatively, searches can be opened using the *Motif Searches* manager (see Managing Searches).

- To delete the current search click on the **Options** button in the *Searches* tab of the *Structure Navigator* and select **Delete <search name>** from the resulting pull-down menu. Searches can also be deleted using the *Motif Searches* manager (see Managing Searches).
- To save a summary of the current search as a text (*.txt)* file click on the **Options** button in the *Searches* tab of the *Structure Navigator* and select **Save Results Summary...** from the resulting pull-down menu. In addition, a list of the CSD refcodes of all hit structures can be saved from the same dialogue by choosing the *.gcd* file type.

21.18 Managing Searches

- The following types of search can be managed:
  - Motif search (see Searching on Motifs).
  - Crystal packing feature search (see Searching on Crystal Packing Features).
  - Crystal packing similarity (see Determining Crystal Packing Similarity).
- All search types are saved with the extension *.msxml*.
- Click on the **Options** button in the *Searches* tab of the *Structure Navigator* and select **Manage Searches...** from the resulting pull-down menu. This will open the *Motif Searches* dialogue:

- All the searches performed, along with the number of hits found for each search, are listed in the dialogue. The following options are available:
  - To load the selected search into the *Structure Navigator* for viewing click on the **View** button.
  - To load a previously saved search into the list of available searches click on the **Open** button, then use the file browser to locate the required search (*.msxml*) file.
  - To save a search run during the current session (as a *.msxml* file), click on the **Save** button.
  - To save a search (as a *.msxml* file) under a different name click on the **Save As** button.

- To delete the selected search hit **Delete**.
- To enter a new name for the selected search in the click on the **Rename...** button and enter the new name in the *Rename Motif Search* dialogue.

- Hit **Close** to close the *Motif Searches* dialogue.

21.19 Managing Motifs

- The following types of motif search query can be managed:
  - Single motif search queries. Motif queries are saved with the extension *.mxml*.
  - Generated motif search queries, consisting of multiple motifs. Generated motif queries are saved with the extension *.gmxml*.
- Click on the **Options** button in the *Searches* tab of the *Structure Navigator* and select **Manage Motifs...** from the resulting pull-down menu. This will open the *Search Queries* dialogue:

- All motif search queries are displayed. The following options are available:
  - **Save:** Saves the selected motif query as a *.mxml* file (if it is a single pre-defined motif) or a *.gmxml* file if it is a family of generated motifs.
  - **Save As**: Saves the selected motif query under a new name as a *.mxml* file (if it is a program-defined motif) or a *.gmxml* file if it is a family of generated motifs.
  - **Open:** Loads a previously saved motif query file into the list of available motif queries. Use the file browser to locate the *.mxml* or *.gmxml* file.
  - **Delete**: Deletes the selected motif query.
  - **Rename**: Allows a new name for the selected motif query to be specified.
  - **Generate Motifs...**: Opens the Motif *Search Query Generation* wizard (see Generating a New Motif).

- **Edit Motifs**: Opens the *Motif Search Query Generation* wizard at the point where functional groups for the motif are selected (see Defining Functional Groups).

- Hit **Close** to Close the *Search Query* dialogue.

22 CSD-Discovery

22.1 Overview

- CSD-Discovery provides valuable insights from high quality experimental crystal structure data in the CSD for computational and medicinal chemists. Capabilities of CSD-Discovery include:
  - Ligand-based drug design with conformer generation, ligand overlay and CSD Full Interaction Maps
  - Structure-based drug design including SuperStar intermolecular interaction maps
  - Protein-ligand docking using GOLD
  - CSD Python API functions including conformer generation
- The following features of CSD-Discovery are available through the Mercury interface:
  - Mapping of interaction preferences around molecules (see Mapping Molecular Interaction Preferences)
  - Generation of conformers using knowledge from experimentally-determined crystal structures (see Generating Conformers)
  - Flexible docking of ligands into protein binding sites (see Protein-Ligand Docking)
  - Identification of interaction hotspots in protein binding sites (see Analysing Protein-Ligand Interactions)
  - Flexible overlay of multiple ligands of the same target (see Overlaying Multiple Ligands)

22.2 Accessing CSD-Discovery

- CSD-Discovery is usually available at no additional cost to academic Cambridge Structural Database System subscribers. Industrial customers should contact *admin@ccdc.cam.ac.uk* for further information. Evaluations are normally available for those industrial organisations interested in purchasing.
- Licensing of CSD-Discovery is handled through the CSD licensing. To activate CSD-Discovery in a copy of Mercury where it is currently not activated, use the CCDC Activation tool.
- The **Help, About Mercury...** menu option will display details about your current licensing status, including features available in CSD-Discovery.

22.3 Mapping Molecular Interaction Preferences

- The IsoStar knowledge base is a powerful tool to aid molecular modelling and drug design. It can be used to display interaction preferences around small molecules and ultimately can give a qualitative appreciation of whether two different functional groups are likely to be bioisosteric or not.

- For more details on the interface and how to use the functionality, see Assessing Molecular Interaction Preferences Using the CSD

22.4 Generating Conformers

- Small organic molecules are flexible and can adopt multiple conformations in solution at room temperature. Also, the biologically active conformation can be different from those observed in solution as it may be induced by protein binding. Methods to generate conformational models that capture the bioactive conformation are very valuable for Computer-Aided Drug Design (CADD) studies.
- The CSD Conformer Generator uses the wealth of knowledge in the CSD to explore the conformational space of small molecules.
- For more details on the interface and how to use the functionality, see Generating Conformers Using the CSD.

22.5 Protein-Ligand Docking

- GOLD (Genetic Optimisation for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites.
- To launch an independent instance of the GOLD program, click on **CSD-Discovery** in the top-level menu and select **Launch GOLD Docking** from the resulting menu.
- For more detailed information, see the specific GOLD documentation online.

22.6 Analysing Protein-Ligand Interactions

- SuperStar generates intermolecular interaction maps for detecting binding preferences within a protein cavity.
- 3D maps that highlight hotspots for interaction can be derived from data in the CSD or the Protein Data Bank (PDB). Click on **CSD-Discovery** in the top-level menu and select **Launch Protein Interactions** from the resulting menu to launch the program.
- For more detailed information, see the SuperStar documentation online.

22.7 Overlaying Multiple Ligands

- Ligand Based Design (LBD) strategies require as input a set of molecules known to bind to the same target. These molecules are superimposed in a manner that it is hoped mimics the relative configurations of the ligands when bound in the active site of the target.
- To achieve this, an ensemble of low energy conformations of each ligand is generated, one conformation of each is selected, and the selected conformations are overlaid such that the mapping of common chemical features is optimized.
- To launch an independent instance of the Ligand Overlay program, click **on CSD-Discovery** in the top-level menu and select **Launch Ligand Overlay** from the resulting menu.
- For more detailed information, see the specific Ligand Overlay documentation online.

23 The CSD Python API

23.1 Overview

The **CSD Python API** menu within Mercury allows you to run supplied or user-written Python scripts, especially scripts that make use of the *CSD Python API*. Scripts run from the **CSD Python API** menu will act on the currently displayed structure, whether it is a CSD entry or a user loaded crystallographic file.

23.2 Setting Up the CSD Python API

- The CSD Python API is automatically installed as part of the CSD Installation. For further information, please refer to The CSD Python API Options Dialogue.
- For advanced users who wish to make use of their own Python environment, please refer to the CSD Python API documentation for full installation instructions. It is necessary to have an instance of Python 2.7 installed before setting up the CSD Python API.

23.3 Using the CSD Python API Menu

- Clicking **CSD Python API** in the top-level menu will open the pull-down **CSD Python API** menu. This contains a list of sub-menus containing CCDC Built-In Python scripts and access to the **Options…** menu, in addition to any user-generated sub-menus and scripts (see User-Created Scripts). Also included are links to the CSD Python API Documentation and to the CSD Python API Forum, a community resource for script dissemination and discussion.

23.3.1 CSD Python API Sub-menus

- The **CSD Python API** menu consists of a number of sub-menus relating to sub-directories that contain individual scripts.
- Sub-directories containing Python scripts can be nested within directories in order to create these sub-menus. Users can create their own **CSD Python API** sub-menu architecture using this method. See The CSD Python API Options Dialogue for more information.

23.3.2 Running Scripts from the CSD Python API Menu

- Clicking on a script in the **CSD Python API** menu will run that script for the loaded structure. A dialogue box will appear with a progress bar, indicating that the script is running.
- The dialogue box includes a hyperlink to the folder location where any output files from the script will be written. This location is based upon the Output Directory given in the Settings of the CSD Python API Options Dialogue as reported in section 23.4
- The corner of the dialogue box features a button marked 'Open in Browser…' that will replicate the given script output in your systems' default internet browser.
- It should be noted that the Mercury **CSD Python API** menu is programmed to time out after three minutes of the script running, at which point an error message will be displayed. If this happens, the script will continue running, and the error message will contain useful information to tell you how to find the output, on completion.

23.3.3 CCDC Built-In Scripts

- The following scripts are included with the Mercury **CSD Python API** sub-menus:
  - *conformers_similarity.py* – Generate a number of conformers for the loaded structure and show their RMSD in comparison to the input conformation
  - *generate_conformers.py* – Generate a number of conformers for the loaded structure and show the values of all torsions for each
  - *load_in_conquest.py* – Load the current structure into the *ConQuest* sketch window
  - *crystal_structure_report.py* – Generate a simple report in .html format containing crystallographic and geometric data for the loaded structure, as well as information on hydrogen- and halogen-bonding
  - *molecular_geometry_report.py* – Generate a simple report in .html format describing molecular geometry of the loaded structure, including Mogul geometry analysis
  - *quick_geometry_check.py* – Generate a simple report in .html format containing Mogul geometry analysis for the loaded structure
  - *quick_packing_check.py* – Generate a simple report in .html format containing information about crystal packing descriptors for the loaded structure
  - *simple_report.py* – Generate a simple report in .html format containing basic information about the loaded structure
  - *chemical_name_search.py* – Search the CSD from a pop-up dialogue box in the Mercury interface to find entries relevant to the specified chemical name or synonym
  - *molecular_similarity_search.py* – Search the CSD from the Mercury interface to find entries with a Tanimoto Similarity Coefficient above 0.7 based on the loaded structure
  - *user_support.py* – This script generates useful information for helping to resolve any problems running CSD Python API scripts from within Mercury
  - *welcome.py* – A simple script to help new users write their own scripts. It is recommended to inspect this script in a text editor before getting started

- Please refer to the Troubleshooting section below if any issues are encountered when running these scripts.

23.3.4 User-Created Scripts

- It is possible to run user-created scripts directly from the Mercury **CSD Python API** menu. The locations of user-created scripts can be defined in the **Options…** menu (see below).
- User-created scripts should make use of the *MercuryInterface* utility. This provides information about the structures in Mercury to the user-created script, and also provides information on where to write output files that Mercury will recognise and load automatically.
- The *MercuryInterface* utility is invoked using the following Python commands:

```
from mercury_interface import MercuryInterface
helper = MercuryInterface()
```

- Further examples can be obtained from inspection of the welcome.py script included in the Mercury distribution.
- If you wish your script to be able to be run from the command line as well as from Mercury, you can create a MercuryInterface with:

```
import sys
if sys.stdin.isatty():
    from ccdc.io import MercuryInterface
else:
    from mercury_interface import MercuryInterface
```

The former will create an interface which will parse command line arguments. The only required argument is a refcode or file path from which an entry will be read. There are optional arguments which may be inspected by running the script with the argument '--help'

23.4 The CSD Python API Options Dialogue

- To change any of the Mercury scripting configuration settings, pick **CSD Python API** in the top-level menu, then **Options…** in the resulting pull-down menu.
- The *Mercury Scripting Configuration* control window will be displayed:

- The following options are available:
  - *Location of built-in scripts*: This is the folder where the scripts included in the Mercury distribution are located
  - *Additional Mercury Script Locations*: User defined folders where user written scripts, or folders containing user written scripts, are located. Folders are defined using the **Add Location** button, and any selected folder can be removed using the **Remove Location** button. If a folder is defined in this way, any subfolders containing user written scripts will appear as sub-menus in the **CSD Python API** menu.
  - *Output Directory*: User defined folder where the output from scripts will be stored. This can be modified using the **Browse…** button, and it is recommended that users change this from the default directory. Files generated from scripts run through Mercury are organised by identifier, script name and time stamp, each of which are subfolders of the Output Directory.
  - *Python Installation*: The location of the Python executable. Clicking the **Default** button will set this location to the version of Python included as part of the CSD Installation. Clicking the **Browse…** button will allow advanced

users to make use of their own Python environment. Also shown are Python, Conda and CSD Python API version numbers. Any conda or venv Python environment specified in this location will be activated when a script is run via Mercury.

23.5 The CSD Python API Documentation

- The CSD Python API Documentation contains information on all aspects of the CSD Python API. The Mercury Scripts menu contains a link to these pages. Visit *http://www.ccdc.cam.ac.uk/docs/csd_python_api/* to access the documentation.

23.6 The CSD Python API Forum

- The CSD Python API Forum is a community resource for users to discuss uses and features of the CSD Python API, and to share scripts and ideas. The Mercury Scripts menu contains a link to these pages. Visit *http://www.ccdc.cam.ac.uk/forum/csd_python_api/* to access the Forum.

23.7 Troubleshooting

Some of the supplied scripts require additional Python modules which should be already included in your installation, but may not be in the following situations:

- Users should ensure that they have an instance of matplotlib installed; this is detailed in the CSD Python API installation instructions.
- Some Linux environments do not include Tkinter as standard. If problems occur, Linux users should ensure that they have an instance of Tkinter installed on their system.

24 References

Reference 1 is the most recent and preferred general citation to Mercury.

1. *Mercury CSD 2.0 - new features for the visualisation and investigation of crystal structures*, C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Cryst*., **41**, 466-470, 2008.
2. *Mercury: visualization and analysis of crystal structures*, C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. van de Streek, *J. Appl. Cryst*., **39**, 453-457, 2006.
3. *New software for searching the Cambridge Structural Database and visualising crystal structures*, I. J. Bruno, J. C. Cole, P. R. Edgington, M. K. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr.*, **B58**, 389-397, 2002.
4. *Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data*, F. H. Allen and I. J. Bruno, *Acta Cryst*., **B66**, 380-386, 2010.
   DOI:10.1107/S0108768110012048
5. *Rules governing the crystal packing of mono- and di-alcohols* R. Taylor and C. F. Macrae, *Acta Crystallogr.,* **B57**, 815-827, 2001.
6. *Searching the Cambridge Structural Database for the "best" representative of each unique polymorph*. J. van de Streek. *Acta Crystallogr.*, **B62**, 567-579, 2006.

25 Acknowledgements

The CCDC gratefully acknowledges the following copyright works, which are used under licence in Mercury.

Qt, the platform-independent GUI application framework from Qt Software (Nokia Corporation)
http://www.trolltech.com/

Mesa, Copyright (c) 1999-2001 Brian Paul
http://www.mesa3d.org/

OpenSceneGraph, the high performance 3D graphics toolkit
http://www.openscenegraph.org/

POV-Ray: The Persistence of Vision Ray Tracer
http://www.povray.org/

The Loki library, Copyright (c) 2001 Andrei Alexandrescu
http://loki-lib.sourceforge.net/

tree.hh, Copyright (c) 2001 Kasper Peeters
http://www.aei.mpg.de/~peekas/tree/

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (http://www.openssl.org/).

Portions of this software are copyright © 2012, The FreeType Project (www.freetype.org). All rights reserved.

26 Appendix A: Glossary and Technical Details

This section is a glossary of terminology you may come across when reading the Mercury documentation.

Analogues

Asymmetric Unit

Atomic Coordinates

- Cambridge Structural Database

Refcode (entry ID)

Atomic Labels

Authors

Average Sigma (C-C))

Background

Bioactivity

Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphologies

Cambridge Structural Database

Cell Angles

Cell Lengths

Cell Volume

cif Format

Color

Compound Name

Conformer

ConQuest

Crystal Chemical Unit

Crystal System

Density

Diffraction Details

Disordered Structures

Experimental Notes

Formula

Fractional Coordinates

Graph Sets

Habit

Hydrogen Atoms

Identifier

Literature Reference

Melting Point

Miller Plane

mol Format

mol2 Format

Molecule

Motif

Orthogonal Coordinates

Other Cell Details

Packing Feature

pdb Format

Peptide Sequence

Phase Transitions

Polymorph

pov Format

Powder Study

Powder Pattern Similarity Measure

PreQuest

Pressure

Radiation Probe

Radiation Source

R Factor

Reciprocal Cell Axes

Recrystallisation Solvent

- Cambridge Structural Database

Refcode (entry ID)

Refinement Details

Remarks

res Format

Root Mean Square Deviation (RMSD)

sd Format

Sensitivity

Source

Space Group

Suppressed Atoms

Symmetry Operators

Sybyl Atom Type

Synonym

Temperature

Torsion Angles

Unit Cell

van der Waals Radii

Voids

xye Format

xyz Format

Z (Number of Molecules in Unit Cell)

Z′ (Number of Molecules in the Asymmetric Unit)

Analogues

Where available, isostructural, isomorphous and isotypic analogues will be reported for Cambridge Structural Database (CSD) structures. Details of isoelectronic, analogous or isomeric structures are not included.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Asymmetric Unit

A crystal structure consists of a basic motif that is repeated in 3D space by the symmetry operators of the crystallographic space group. A crystallographer determines the coordinates of the atoms in this basic motif, called the *asymmetric unit.* It is the smallest part of a crystal structure from which the complete structure can be built using space group symmetry.

The asymmetric unit may consist of only one molecule or ion, part of a molecule, or several molecules that are not related by crystallographic symmetry. (This is why you sometimes see more than one molecule in Mercury when you first display a structure from the Cambridge Structural Database.) For example, consider structures of formula C_{12} H_{18} N_4 O_2:

- If the asymmetric unit contains one molecule, the crystallographer must determine the coordinates of 36 atoms.
- If the asymmetric unit contains two molecules, the crystallographer must determine the coordinates of 72 atoms.
- If the asymmetric unit is half a molecule, this implies that the molecule possesses symmetry coincident with a crystallographic symmetry element. For example, the molecule might possess a mirror plane, so that half of the atoms are related to the other half by symmetry. In this case, the crystallographer must determine the coordinates of only 18 atoms.

Related Topics

- Atomic Labels
- Cambridge Structural Database
- Crystal Chemical Unit
- Molecule
- *Z′* (Number of Molecules in the Asymmetric Unit)

Atomic Coordinates

Atomic coordinates are unavailable for about 8% of entries in the Cambridge Structural Database. This situation usually arises for structures published as short communications in certain journals, or if data has not been received from an author. If you try to load a structure that has no coordinates, the text *No structure available* will be displayed in the main display area.

Relevant Program Options

- Listing atomic coordinates (see Spreadsheet Display of Atoms and Their Properties)
- Reading structures (see Loading Structures)
- Saving structures (see Saving Files and Copying to the Clipboard)

Related Topics

- Cambridge Structural Database
- Fractional Coordinates
- Hydrogen Atoms
- Orthogonal Coordinates

Atomic Displacement Parameters (ADPs)

Atomic Displacement Parameters (also known as *atomic vibration parameters* and *thermal parameters*) are used to generate displacement ellipsoids. Displacement ellipsoids represent atomic motion and can be either isotropic or anisotropic. The shape and size of ADPs can be used to highlight potential errors with the data; the smaller and more spherical the ellipses for anisotropic atoms, the better the data (see *Troublesome Crystal Structures: Prevention, Detection and Resolution*, R. L. Harlow, *J. Res. Natl. Inst. Stand. Technol.*, **101(3)**, 327, 1996).

For ellipsoids to be displayed in Mercury, it is necessary for the relevant U_{equiv} and U_{ij} values to be present in the CIF or SHELX *res* file being read into Mercury. U_{equiv} and U_{ij} values are not stored in the CSD, thus it is not possible to view ellipsoids for entries stored in or output from the CSD. CIFs obtained through the CCDC's online request form (*http://www.ccdc.cam.ac.uk/products/csd/request/*) may contain these data where provided by the publishing author. Currently, U_{equiv} and U_{ij} values are written only to SHELX *res* files from Mercury.

Ellipsoids can be displayed in Mercury using the **Style:** drop down menu, or by going to **Display** in the top-level menu, selecting **Styles** from the resultant pull-down menu and clicking on **Ellipsoids**. Settings for the Ellipsoids display style may be controlled using the Ellipsoid settings... window, available from the **Display**, **Styles** menu.

Relevant Program Options

- Displaying ellipsoids (see Setting a Global Display Style)

- Setting ellipsoid display options (see Setting Ellipsoid Display Options)
- Colouring by atomic displacement (see Atomic Displacement Colouring)

Related Topics

- cif Format
- res Format

Atomic Labels

Each atom in a structure from the Cambridge Structural Database has a label, which consists of the element symbol followed by a number and (sometimes) a prime. CSD atom labels are normally identical to, or closely related to, those used by the author. Atoms generated by symmetry (i.e. not belonging to the asymmetric unit) will have an extra letter at the end of their label, which indicates which symmetry operation was used.

Relevant Program Options

- Labelling all atoms (see Turning All Atom Labels On or Off)
- Labelling particular atoms (see Setting Display Properties for Particular Atom(s) or Bond(s))
- Listing atom labels (see Spreadsheet Display of Atoms and Their Properties)
- Customising display of atom labels (see Colouring and Resizing All Labels)

Related Topics

- Cambridge Structural Database
- Crystal Chemical Unit

Authors

Where available, author names will be reported for Cambridge Structural Database (CSD) structures.

Author names are stored exactly as given in the paper, with forenames abbreviated and stored as initials, e.g. F.Allen and F.H.Allen may both occur. Some author names may include dynastic tags, e.g. S.S.Simons Junior, A.J.Arduengo III.

Authors' names are stored in the CSD without umlauts, accents, etc. However, some journals express an umlaut by a following e; thus Müller, Sänger, etc., may sometimes appear as Mueller, Saenger, etc. in the CSD.

Chinese, Korean and Malaysian names are usually stored in full, e.g. Bing Bing Chang, Jung Mi Shin. Occasionally, however, Oriental names are recorded with initials, e.g. H.S.Kim. These variations are due to different journal conventions.

Names from non-Roman alphabets are not always transliterated in the same way in different papers, e.g. Belskii, Belsky, Belskij.

There is some inconsistency in the database in the handling of 2-letter initials, e.g. Yu.T.Struchkov might occasionally be stored as Y.T.Struchkov.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

- Literature Reference

Average Sigma (C-C)

Where available, average estimated standard deviation (e.s.d) [standard uncertainty (s.u.)] of carbon-carbon bond lengths can be displayed for Cambridge Structural Database (CSD) structures.

The average estimated standard deviation (e.s.d.) provides a rough measure of precision and is quoted for most Cambridge Structural Database (CSD) entries. In certain entries, where no carbon-carbon e.s.ds are available, then the value may be derived from C-N, C-O, N-N, N-O, O-O bond-length e.s.d.s.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Background

In Mercury, the term *background* refers to empty parts of the display area. Thus, any point in the display on which there is no displayed object such as an atom, bond, plane or centroid is considered part of the background.

Relevant Program Options

- Effect of right-clicking in background (see Right-Clicking in the Display Area)
- Effect of left-clicking in background (see Ways of Selecting Atoms and Molecules)
- Setting background colour (see Setting Background Colours)

Bioactivity

Where available, bioactivity information can be displayed for Cambridge Structural Database (CSD) structures.

For CSD structures biological activity is reported if the author indicates that the compound, or a near-relative, is of biological interest. Where available related information is also included e.g. not active; possible activity; biologically tested; derivative of or related to a compound with biological activity.

Relevant Program Options

- Displaying text information (see Display of Text Information)
- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- Cambridge Structural Database

Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphologies

BFDH crystal morphology is an approximation based on crystallographic geometrical considerations. For a given structure, the BFDH algorithm will predict the habit or shape of a crystal using the corresponding unit cell and symmetry operator information.

The BFDH theoretical crystal morphology can be viewed by clicking on **Calculate** in the top-level menu, then **BFDH Morphology...** from the resultant pull-down menu.

Relevant bibliographic references:

- A. Bravais, *Ètudes Cristallographiques* (Paris: Gauthier Villars) 1866
- G. Friedel, *Bull. Soc. Franc. Mineral.*, **30**, 326, 1907
- J. D. H. Donnay and D. Harker, *Am. Mineral.*, **22**, 463, 1937

Relevant Program Options:

- Displaying BFDH crystal morphologies (see Displaying Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphology)

Related Topics

- Habit

Cambridge Structural Database

The Cambridge Structural Database (CSD) is the definitive database of small-molecule organic and metallo-organic crystal structures and is maintained and distributed by the Cambridge Crystallographic Data Centre (*http://www.ccdc.cam.ac.uk*).

Relevant Program Options:

- Browsing the CSD (see Loading Structures)
- Displaying chemical diagrams (see Chemical Diagram Display)
- Displaying text information (see Display of Text Information)
- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- ConQuest
- PreQuest

Cell Angles

Where available, cell angles can be displayed for structures. Various unit cell data may be recorded:

- alpha value of interaxial angle alpha (in degrees)
- beta value of interaxial angle beta (in degrees)
- gamma value of interaxial angle gamma (in degrees)

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database
- Cell Lengths
- Cell Volume
- Unit Cell

Cell Lengths

Where available, cell lengths can be displayed for structures. Various unit cell data may be recorded:

- a length of unit cell a-axis (in Angstroms)
- b length of unit cell b-axis (in Angstroms)
- c length of unit cell c-axis (in Angstroms)

Relevant Program Options

- Display of Text Information

Related Topics

- Cambridge Structural Database
- Cell Angles
- Cell Volume
- Unit Cell

Cell Volume

Where available, various unit cell data can be displayed including:

- v volume of unit cell (in cubic Angstroms)

Relevant Program Options

- Display of Text Information
- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- Cambridge Structural Database
- Cell Angles
- Cell Lengths
- Unit Cell

cif Format

The Crystallographic Information File (CIF) format was developed as the standard crystallographic data exchange format (Hall, Allen and Brown, *Acta Cryst.*, **A47**, 655, 1991; *http://www.iucr.org/iucr-top/cif/home.html*).

Relevant Program Options:

- Loading a Structure from a File (see Loading a Structure from a File)
- Saving files in 3rd party formats (see Saving Structures in External (3rd Party) Formats)

Chirality

The term chirality in chemistry refers to a property of some molecules whereby the mirror image of that molecule cannot be directly superposed onto the original molecule. The chirality of molecules can be classified in Mercury based on the stereochemistry at any and all chiral centres within the molecule.

All four (or higher) coordinate atoms will be considered as potential chiral centres (or stereocentres). To be classed as a stereocentre, an atom must have four or more *different* substituents. Any identified stereocentres are assessed using Cahn-Ingold-Prelog (CIP) priority rules (including special CIP rules) to assign an *R* or *S* label to each stereocentre.

In the case of stereocentres with more than four substituents, an *M* label (for 'mixed') may be assigned if the order of CIP priorities does not uniformly increase or decrease. Para-chiral centres are also handled in which the only difference between two of the substituents is the chirality of those substituents.

Note: The chirality will be assigned for any observed molecules in the Mercury visualiser, which by default will only be the crystal chemical unit (see Crystal Chemical Unit). If you are interested in assessing whether both hands (i.e. mirror images) of a given molecule occur within the crystal structure, you can pack a full unit cell to generate all symmetry equivalents and then assess their chirality (see Displaying the Contents of a Single Unit Cell). Assignment of chirality also requires bond type information to be present.

Relevant Program Options:

- Labelling structures by stereochemistry (see Labels)
- Assignment of bond types (see Auto-Editing Options)

Color

Where reported, the colour of the crystal at room temperature in daylight can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Compound Name

Where available, the compound name can be displayed for Cambridge Structural Database (CSD) structures. Compound names in the CSD usually follow the rules of standard chemical nomenclature. Occasionally, a trivial, drug or trade name might be used.

A few conventions are used in constructing compound names:

- Bridging ligands in polymeric metal coordination complexes are identified by the bridging indicator *mu*, with the polymer identified by the prefix *catena*, for example *catena-((mu$_2$-2,5-dihydroxy-p-benzoquinonato)-zinc).*
- Names of hydrates will contain the words *hemihydrate, monohydrate, dihydrate*, etc., otherwise, just *hydrate* if the multiplier is a non-integer value.
- If other solvents are present, the name will contain the word *solvate*; *clathrate* is used for solvates which are clathrated, as in host-guest compounds.
- Deuterated species will always contain the name characters *deuter*.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

- Synonym

Conformer

When available, stereodescriptors describing the shape of the whole molecule or structure can be displayed for Cambridge Structural Database (CSD) structures. Terms such as 1,2-alternative, cone, pinched cone, right-handed, helix, head-tail are used. Such text occurs frequently in the description of calixarenes, peptides and organic polymers, and may relate to configuration.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

ConQuest

ConQuest is the principal program for searching the Cambridge Structural Database. Hit lists from ConQuest searches can be read into Mercury.

Relevant Program Options:

- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- Cambridge Structural Database

Crystal Chemical Unit

In some cases, molecular symmetry coincides with the symmetry elements of a crystallographic space group. For example, a molecule might have a mirror plane which coincides with a crystallographic mirror plane in the space group. In the crystal structure, the molecule has *exact m-symmetry*, the asymmetric unit is half a molecule, and the crystallographer only determines the coordinates of half of the atoms in the molecule.

In this situation, the atoms in the asymmetric unit plus the symmetry-generated atoms forming the other half of the molecule collectively form the *crystal chemical unit*.

If you load such a structure into Mercury from the Cambridge Structural Database, the coordinates of the entire molecule (i.e. the crystal chemical unit) will be read in and displayed, not just the coordinates of the asymmetric unit. The symmetry-generated atoms can be distinguished from the atoms in the asymmetric unit because their atom labels will end in a letter (the letter indicates which symmetry operation was applied to generate the atom).

Relevant Program Options:

- Colouring by symmetry equivalence (see Symmetry-Equivalence Colouring)
- Viewing the crystal chemical unit (see Switching Packing On and Off)

Related Topics

- Asymmetric Unit
- Atomic Labels
- Cambridge Structural Database

- Molecule

Crystal System

Any crystal must belong to one of seven crystal systems, the symmetry of which introduces constraints on the possible values of the unit cell parameters. The seven crystal systems are: triclinic, monoclinic, orthorhombic, tetragonal, trigonal/rhombohedral, hexagonal and cubic.

Related Topics

- Reciprocal Cell Axes
- Space Group
- Unit Cell

Density

Where available, the density can be displayed for structures. This is the density of the crystal, as reported by the author or calculated from the reported chemical formula and unit cell data, using the relationship:

Density = (1.66 x formula weight x Z) / unit cell volume

where Z is the number of molecules in the unit cell.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Diffraction Details

Where available, details describing the diffraction of the crystal can be displayed for Cambridge Structural Database (CSD) structures. These may include details of the experimental set-up, data collection or treatment and details of twinning derived from diffraction.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Disordered Structures

Disordered structures display a lack of regularity. For example, each of the F atoms in a trifluoromethyl group, $-CF_3$, might be randomly distributed between two sites. This means that the crystallographer will report two sets of coordinates for each F atom. In some cases, two alternative sites are occupied equally; in other cases, there is a major site and a minor site. Disorder can involve more than two sites and it can also involve a whole molecule.

Disordered structures in the Cambridge Structural Database may have been treated in one of two ways. In older structures, only one position is kept for each disordered atom. When such a structure is read into Mercury, it will therefore appear as if it is not disordered at all. In more recent structures, all positions of disordered atoms are kept but only one set is connected by bonds to form a complete molecule. When a structure like this is displayed in

Mercury, it will have the appearance of a complete molecule with disconnected atoms at various points to indicate alternative sites for the disordered atoms, e.g.

Relevant Program Options

- Removing disordered atoms from a structure (see Manually Editing a Structure)

Related Topics

- Cambridge Structural Database
- **Suppressed Atoms**

Experimental Notes

Where available, experimental details describing the properties of the crystal before diffraction can be displayed for Cambridge Structural Database (CSD) structures. Typical examples are: irradiated product, superconducting action, humidity, ground state, mesogen, dye, explosive, piezoelectric, triboluminescent, luminescent, monomerization study and thermal decomposition study.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Formula

Where available, the chemical formula can be displayed for Cambridge Structural Database (CSD) structures. Since a crystal structure may contain more than one type of molecule, there are basically two sorts of formulae in the CSD:

- The formula of an individual molecule, for example $C_{10}\ H_{18}\ N_2\ Ni\ O_5$.
- The *sum formula*, i.e. the formula of all the different molecules in a structure added together. For example, the dihydrate $C_{10}\ H_{18}\ N_2\ Ni\ O_5,\ 2(H_2O)$ would have a sum formula of $C_{10}\ H_{22}\ N_2\ Ni\ O_7$.

The letter *n,* which may sometimes be seen in formulae displayed in *View structures***,** indicates that the compound is polymeric*,* for example $(C_3\ H_3\ O_6\ Sc)n$. The formula enclosed in the brackets is that of the monomer repeat unit.

Relevant Program Options

- Displaying text information (see Display of Text Information)
- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- Cambridge Structural Database

Fractional Coordinates

The atomic coordinates of a crystal structure are usually expressed as fractional coordinates, i.e. as fractions of the *a, b* and *c* unit cell vectors. For example, an atom with fractional coordinates *0.5, 0.5, 0.5* would lie half-way along each unit cell edge, i.e. in the middle of the unit cell.

Relevant Program Options:

- Listing atomic coordinates (see Display of Text Information)

Related Topics

- Atomic Coordinates
- Orthogonal Coordinates
- Unit Cell

Graph Sets

Graph set analysis of hydrogen-bonding patterns is based on a set of descriptors which specify the type of pattern along with the numbers of hydrogen-bond donors and acceptors. This type of hydrogen-bonding analysis was popularised by Etter (1990) and subsequently further developed by Bernstein *et al*. (1995). The implementation used within Mercury is based on that developed by Motherwell *et al*. (2000) for the program *RPLUTO*.

A graph-set descriptor is written as $\mathbf{G}^{\mathbf{a}}_{\mathbf{d}}(\mathbf{n})$, in which **G** represents the type of pattern, **a** is the number of hydrogen bond acceptors, **d** is the number of donors and **n** the number of atoms in the pattern. The pattern type, **G**, can be one of four different options: **C** for an infinite chain, **S** for an intramolecular hydrogen bonding pattern, **R** for an intermolecular ring and **D** for a discrete, finite hydrogen-bonding pattern.

Relevant bibliographic references:

- M. C. Etter, *Acc. Chem. Res.*, **23**, 120, 1990

- J. Bernstein, R. E. Davis, L. Shimoni and N.-L. Chang, *Angew. Chem. Int. Ed.*, **34**, 1555, 1995
- W. D. S. Motherwell, G. P. Shields and F. H. Allen, *Acta. Cryst.* B**56**, 466, 2000

Relevant Program Options:

- Displaying Graph Sets (see Displaying Graph Sets)

Habit

When reported by the author, the crystal habit (i.e. the shape of the crystal, such as *needle* or *plate*) can be displayed for Cambridge structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Bravais, Friedel, Donnay and Harker (BFDH) Crystal Morphologies)
- Cambridge Structural Database

Hydrogen Atoms

Hydrogen atoms can be difficult to locate by X-ray crystallography and hence their coordinates are sometimes missing from crystal structures. In such a case, the missing H atoms are not displayed when the structure is read into Mercury.

It is difficult to guarantee correct classification of hydrogen-bond donors and acceptors in the *Define H-bonds* dialogue box when one or more hydrogen atoms are missing from a structure.

X-H covalent bond lengths (*X* = any type of atom) are often underestimated when a structure is determined by X-ray diffraction, and this may lead to unrealistically long *H...O* and *H...N* hydrogen-bond distances.

Hydrogen atoms are accurately located when a structure is determined by neutron diffraction.

Relevant Program Options:

- Switching the display of hydrogen atoms on and off (see Switching Hydrogens On and Off)
- Hydrogen bonds (see Default Definition of Hydrogen Bonds)

Identifier

Each structure in the file or database loaded into Mercury has a unique identifier. When reading structures from a file, the identifier will be closely related to the original name in the input file. Structures from the Cambridge Structural Database (CSD) also have a unique identifier known as a *refcode.*

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database
- Cambridge Structural Database

- Refcode (entry ID)

Literature Reference

Where available, a literature reference can be displayed for Cambridge Structural Database (CSD) structures.

Journal names in the CSD are normally the abbreviations adopted by the International Standards Organisation.

The page number is the starting page of the publication. Most journal pages are numeric but sometimes they contain letters, for example *L25, 123S*.

For about 20 journals, mostly Russian and Chinese, each issue starts at page 1. To provide an unambiguous reference, we store the issue number as well as the page number in the journal page field, separating them by a hyphen, for example *89-3* indicates page 89 of issue 3.

Since 1997, *Acta Cryst.,Sect.C (Cr.Str.Comm.)* has reported so-called *CIF-access papers*. These are represented in the Table of Contents of each issue by a brief abstract and diagram. The actual paper does not appear in the printed issue but can be retrieved from the IUCr archive using the printed journal data validation number. An example of such a number is IUC9900004. In this case the journal page in the CSD would contain 9900004.

J.Chem.Res. is published in two parts - the synopsis (S) and the miniprint (M). Since this journal has no volume number, the page number for part S is stored in the journal volume field and the page number for part M is stored in the journal page field.

The journal volume is normally numeric but may occasionally contain letters, for example *C471*. However, in other cases the alphabetic part of a volume number is absorbed into the journal name, for example *Acta Cryst., Section B, volume 47*.

For most journals, the volume takes a single value for each journal year. However, some journals have more than one volume per year and others have volume numbers that span a year change.

Some journals have no volume number, for example journals of the UK Royal Society of Chemistry, such as *J. Chem.Soc., Dalton Transactions*. Volume numbers are absent from PhD theses and private communications.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database
- Authors

Melting Point

When reported by the authors, the melting point of Cambridge Structural Database (CSD) structures can be displayed (either as a single point or a range, and either in Celsius or Kelvin depending on what the authors report).

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Miller Plane

The Miller plane (*hkl*) is the plane which intercepts the three axes of a unit cell at the points *a/h, b/k, c/l*, where *a, b* and *c* are the unit cell vectors. The indices *h, k* and *l* must be integral. The principal faces of a crystal are usually Miller planes with relatively low (*hkl*) indices, e.g. (100), (111), (2 1 -1). Mercury can calculate and display Miller planes.

Relevant Program Options:

- Viewing along and perpendicular to planes (see Viewing Along and Perpendicular to Planes)
- Creating Miller planes (see Creating Miller Planes)
- Showing slices through crystals (see Displaying Slices Through Crystals)
- Spreadsheet display of planes and their properties (see Spreadsheet Display of Planes and Their Properties)

Related Topics

- Unit Cell

mol Format

Mol is a molecular file format produced by MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577, USA, and is a Mercury input option. Each molfile describes a single molecular structure or query structure; these can contain disjoint fragments, such as would be found in salts or mixtures. Since a mol file cannot store crystallographic unit cell or symmetry information Mercury will not be able to display unit cell diagrams, contact networks, etc. (see sd Format)

Relevant Program Options:

- Reading mol files (see Loading a Structure from a File)
- Saving mol files (see Saving Structures in External (3rd Party) Formats)

mol2 Format

mol2 is a molecular file format produced by Tripos Inc., 1699 South Hanley Road, St Louis, MO 63144-2913, USA (*http://www.tripos.com*), and is a Mercury input and output option.

Relevant Program Options:

- Reading mol2 files (see Loading a Structure from a File)
- Saving mol2 files (see Saving Structures in External (3rd Party) Formats)

Molecule

The word *molecule* in this documentation is used generically to refer to a molecule or ion. For example, sodium acetate monohydrate has 3 molecules: $C_2H_3O_2^-$, Na^+, H_2O.

Motif

A motif is a characteristic pattern of hydrogen bonds within a crystal structure. These interactions may be intermolecular, intramolecular or a combination of the two. Motifs are defined within Mercury as sets of interactions between functional groups with a particular pattern, e.g. rings, chains or discrete contacts. Examples of simply hydrogen bonding motifs

include the common carboxylic acid ring dimer, the hydroxyl OH…O chain, and the amide ring dimer.

Relevant Program Options:

- Performing motif searches (see Searching on Motifs)
- Defining motifs (see Motif Selection)

Orthogonal Coordinates

Although the positions of atoms in crystal structures are usually expressed in fractional coordinates, i.e. relative to the axes of the crystallographic unit cell, it is sometimes convenient to refer them to a set of Cartesian axes (e.g. if the structure is to be written out in a molecular file format). Atomic coordinates expressed in this way are termed orthogonal coordinates.

Relevant Program Options:

- Listing atomic coordinates (see Spreadsheet Display of Atoms and Their Properties).
- Output of structures (see Saving Structures in External (3rd Party) Formats).

Related Topics

- Atomic Coordinates
- Fractional Coordinates
- Unit Cell

Other Cell Details

Where available, other cell parameters for the same compound, polymorph or twin compound can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Packing Feature

A packing feature within a crystal structure is essentially any three-dimensional arrangement of atoms. Usually this term refers to a selection of atoms on multiple molecules within the structure which form a recognisable intermolecular interaction, such as a p…p stacking contact, a carbonyl-carbonyl interaction or a nitro group to halogen interaction.

Relevant Program Options

- Performing packing feature searches (see Searching on Crystal Packing Features)

pdb Format

pdb is the Protein Data Bank format for 3D structures and is a Mercury input and output option. The Protein Data Bank is maintained by the Research Collaboratory for Structural Bioinformatics (*http://www.rcsb.org/pdb/*) and is the single international repository for the processing and distribution of 3D macromolecular structure data, primarily determined experimentally by X-ray crystallography and NMR.

Relevant Program Options:

- Reading pdb files (see Loading a Structure from a File)
- Saving pdb files (see Saving Structures in External (3rd Party) Formats)

Peptide Sequence

Where available, a peptide sequence can be displayed for Cambridge Structural Database (CSD) structures.

The CSD covers peptides of up to 24 residues. Alpha-amino acids and modified alpha-amino acids (also zwitterionic and ionic) are represented by 3-letter codes and symbols; when linked by peptide bonds or non-peptide bonds into a peptide sequence the structure is represented by a peptide sequence of codes, symbols and links reading from the 'N' end of the molecule.

Peptide sequence can represent both cyclic and acyclic (linear) sequences and can contain undefined residues (UND) and branch-point symbols. Bicyclic (or multi- cyclic) peptides, pseudopeptides and retropeptides can also be represented.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Phase Transitions

When reported by the author, information regarding the temperature(s), pressure or other conditions of phase transition(s) can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Polymorph

Where available, polymorph information can be displayed for Cambridge Structural Database (CSD) structures.

Polymorphism is the occurrence of two or more crystalline forms of the same substance. Structures known to be polymorphic contain comments which include the word polymorph (when reported by the author), e.g. non-triboluminescent polymorph.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

pov Format

pov is the scene description format used by the freeware program POV-Ray (*http://www.povray.org*). You can now output Mercury views in pov format in order to generate high-quality rendered images using this ray-tracing program.

Relevant Program Options

- Saving pov files (see Saving Images from Mercury)
- Creating POV-Ray images from Mercury (see POV-Ray Output of Mercury Images)

Powder Study

Where available, a comment is included for Cambridge Structural Database (CSD) structures that have been determined by powder diffraction.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Powder Pattern Similarity Measure

The similarity measure is calculated using the algorithm of De Gelder, Wehrens and Hageman (*J.Comp. Chem.*, (2001) **22**, 273-289) and their new triangle weighting function. See the paper for more details.

Relevant Program Options

- Viewing and Analysing results of a Packing Similarity Search (see Viewing and Analysing Results)

PreQuest

PreQuest is a program that can be used to convert an in-house collection of crystal structures into a database whose format is compatible with that of the Cambridge Structural Database. The resulting database can be viewed with Mercury or searched with ConQuest.

Relevant Program Options:

- Browsing CSD-format databases (see Loading the Cambridge Structural Database or Other CSD-Format Databases)
- Viewing hits from ConQuest searches (see Loading Structures Found by a ConQuest Search)

Related Topics

- Cambridge Structural Database
- ConQuest

Pressure

When reported by the author, details of low or high pressure studies can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Radiation Probe

Where available, details of the radiation probe can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Radiation Source

Where available, data or refinement specifications can be displayed for Cambridge Structural Database (CSD) structures. These indicate:

- synchrotron radiation
- neutron
- neutron and x-ray radiation

References to electron radiation may be found for certain hit structures, though the decision to cease input of these studies was made in January 2001. Mo, Cu or wavelength data (to distinguish more than one study) are only very occasionally available.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

R Factor

The usual measure of experimental precision quoted for a crystal-structure. Roughly speaking, structures with R factors lower than 0.05 (or, equivalently, 5%) are high quality; anything above 0.10 may have significant inaccuracies in bonds lengths and angles.

Relevant Program Options:

- Display of text information (see Display of Text Information)

Reciprocal Cell Axes

The three reciprocal cell axes of a crystallographic unit cell are termed *a*, b** and *c** and lie in the following directions:

- *a** is perpendicular to the real cell axes *b* and *c*;
- *b** is perpendicular to the real cell axes *a* and *c*;
- *c** is perpendicular to the real cell axes *a* and *b*.

Thus, if you view a structure down *a**, you are looking in a direction normal to the *bc* plane of the unit cell. A reciprocal and real cell axis may be coincident, but only if the symmetry of the crystal requires that the relevant cell angles are 90 degrees. For example, in a monoclinic

crystal with the *b* axis unique, the cell angles *alpha* and *gamma* are required by symmetry to be 90 degrees, so the real axis *b* and the reciprocal axis *b** will lie along the same direction.

Relevant Program Options:

- Viewing along reciprocal cell axes (see Viewing Along Crystallographic Axes)

Related Topics

- Crystal System
- Unit Cell

Recrystallisation Solvent

When reported by the author, comments can be displayed for Cambridge Structural Database (CSD) structures to indicate how crystals were prepared; also to indicate the source of natural products, e.g. from the melt, xtal hexane/diethyl ether, Flemingia fruticulose wall (Leguminosae).

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Refcode (entry ID)

Each entry in the Cambridge Structural Database has a unique identifier known as a *refcode*. The refcode contains six letters, e.g. ABACOF. If more than one study of a compound is present in the CSD, the second and subsequent entries will have two numbers after the six letters, e.g. ABACOF01, ABACOF10, ABACOF03.

Deuterated forms of a compound have the same six-letter code as the non-deuterated form. Stereoisomers have different 6-letter codes.

Relevant Program Options:

- Browsing the CSD (see Loading the Cambridge Structural Database or Other CSD-Format Databases)
- Viewing hits from ConQuest searches

Related Topics

- Cambridge Structural Database
- Identifier

Refinement Details

When reported by the author, information concerning the refinement can be displayed for Cambridge Structural Database (CSD) structures e.g. *refinement in centrosymmetric space group*. Other terms recorded include: multipole, high-angle, rigid body, final, 2, IV refinement, isotropic, anisotropic, Frenkel, model 2, full data, kappa, rigid body, high-order.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Remarks

Where available, remarks can be displayed for Cambridge Structural Database (CSD) structures. Remarks include deposition numbers, cross references to identical structures and editorial comment resulting from CCDC validation i.e. errors and discrepancies, a record of correspondence with authors and details of unresolved problems including those resulting from program limitations.

Relevant Program Options

- Displaying text information

Related Topics

- Cambridge Structural Database

res Format

res is the crystal-structure file format used by the program SHELX (*http://shelx.uni-ac.gwdg.de/SHELX/*).

Relevant Program Options:

- Reading res files (see Loading a Structure from a File)
- Saving res files (see Saving Structures in External (3rd Party) Formats)

Root Mean Square Deviation (RMSD)

The root mean square deviation (RMSD) is a commonly used measure of the difference between two sets of values (usually comparing observed data to estimated data). The RMSD is defined as the square root of the mean squared error. In Mercury this is used to measure the geometric difference between packing features or packing patterns in crystal structures.

Relevant Program Options:

- Analysing packing feature searches (see Viewing and Analysing Results)
- Analysing packing similarity searches (see Viewing and Analysing Results)

sd Format

sd is a molecular file format produced by MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577, USA, and is a Mercury input option. SDfiles are structure-data files which contain structural information and associated data items for one or more compounds in mol format (see mol Format).

Relevant Program Options:

- Reading SD files (see Loading a Structure from a File)
- Saving SD files (see Saving Structures in External (3rd Party) Formats)

Sensitivity

When reported by the author, properties of the crystal can be displayed for Cambridge Structural Database (CSD) structures. Typically, comments are included for structures which are:

air- and moisture-sensitive
hygroscopic
efflorescent

deliquescent
heat-sensitive (needs to be stored below room temperature)
oxygen-sensitive
light-sensitive
photo-sensitive
pyrophoric

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Source

When reported by the author, information relating to the natural-product source of the chemical compound or its immediate parent compound can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Space Group

There are 230 possible arrangements of symmetry elements in the solid state. They are called *space groups (see International Tables for Crystallography,* Volume A, Kluwer Academic Publishers, 1983). Any crystal must belong to one (and only one) space group.

The space groups are numbered from 1 to 230 and each is represented by a space group symbol, e.g. space group number 19 has the symbol $P2_12_12_1$. The space group of any structure read into Mercury from the Cambridge Structural Database will be displayed in the bar at the top of the Mercury window, e.g.

Note: For space groups where the name does not specify a setting, a setting will be arbitrarily chosen. For these space groups, an alternate setting can be selected by specifying the symmetry operators, which will always take precedence over a space group name.

Related Topics

- Cambridge Structural Database
- Crystal System
- Unit Cell

Suppressed Atoms

In Cambridge Structural Database (CSD) structures one or more atoms may be suppressed. A common example is when the author reports two sets of atomic coordinates, for major and

minor sites, in a disordered structure. In most cases the atoms of the minor site will be suppressed.

Relevant Program Options

- Removing suppressed atoms (see Display of Text Information)

Related Topics

- Cambridge Structural Database
- Disordered Structures

Symmetry Operators

Each space group (more strictly, each allowable setting of a space group) is characterised by a set of symmetry operators that indicate the relative positions of symmetry-equivalent atoms in the structure. For example, the space group $P2_1$ has the symmetry operators *x*, *y*, *z* and -*x*, *y + 1/2*, -*z*; these mean that for every atom whose fractional coordinates are (*x*, *y*, *z*) there is another symmetry-related atom whose fractional coordinates are (-*x*, *y + 1/2*, -*z*). The order of a symmetry operator refers to the number of atoms it will generate if applied first to an atom in a general position (*x*, *y*, *z*), then to the resulting atom, and so on until the original atom is regenerated; for example, the symmetry operator arising from a 3-fold rotation axis will have order 3. The type of a symmetry operator is an integer number *+n* or -*n*. The sign refers to whether application of the symmetry operator to a chiral atom will result in a symmetry-generated atom of the same or inverted chirality. The number *n* refers to whether application of the operator results in a rotation and, if so, by how many degrees, e.g. if *n* = 3, the operator involves rotation by 360/3 degrees.

Relevant Program Options:

- Listing symmetry operators (see Spreadsheet Display of Space Group Symmetry Operators)

Related Topics

- Chirality
- Space Group

Sybyl Atom Type

Sybyl atom types were devised by Tripos Inc., 1699 South Hanley Road, St Louis, MO 63144-2913, USA (*http://www.tripos.com*) and are used to classify atoms according to their environment, e.g. *C.3*, *C.2* and *C.1* mean, respectively, sp^3, sp^2 and *sp* hybridised carbon; *N.am* means amide nitrogen.

Relevant Program Options:

- Listing atom properties (see Spreadsheet Display of Atoms and Their Properties)

Synonym

When reported by the author, any appropriate synonym(s) for the compound name can be displayed for Cambridge Structural Database (CSD) structures.

Relevant Program Options

- Displaying text information

Related Topics

- Cambridge Structural Database
- Compound Name

Temperature

Where available, a comment can be displayed for Cambridge Structural Database (CSD) structures if the structure has been determined at a temperature other than room temperature. Any structure determined in the range 283-303K is considered to be a room-temperature structure.

Relevant Program Options

- Displaying text information (see Display of Text Information) (see Display of Text Information)

Related Topics

- Cambridge Structural Database

Torsion Angles

Torsion angles are used to describe conformations around rotatable bonds. The torsion angle between 4 atoms A-B-C-D is the angle by which the vector A-B must be rotated in order to eclipse the vector C-D when viewed along the vector B-C. Crystallographers usually express torsion angles in the range -180 to +180 degrees. According to convention (Klyne and Prelog, *Experientia* **16**, 521, 1960) a clockwise rotation is positive and an anti-clockwise rotation is negative.

Torsion angles may be calculated for any set of four atoms.

Relevant Program Options:

- Measuring torsion angles (see Measuring Distances, Angles and Torsions)

Unit Cell

The unit cell is the basic building block of a crystal, repeated infinitely in three dimensions. It is characterised by:

- three vectors (*a, b, c*) that form the edges of a parallelepiped;
- the angles between the vectors (*alpha*, the angle between *b* and *c*; *beta*, the angle between *a* and *c; gamma*, the angle between *a* and *b*).

Depending on the crystal system, there are sometimes restrictions on the values that unit cell parameters can take:

Triclinic/Anorthic	no restrictions	
Monoclinic		*alpha* = *gamma* = 90
Orthorhombic		*alpha* = *beta* = *gamma* = 90
Tetragonal	a = b	*alpha* = *beta* = *gamma* = 90
Trigonal/Hexagonal	a = b	*alpha* = *beta* = 90, *gamma* = 120
Rhombohedral	a = b = c	*alpha* = *beta* = *gamma*
Cubic	a = b = c	*alpha* = *beta* = *gamma* = 90

In the above table, *alpha=gamma*=90 for monoclinic corresponds to the *b-axis unique* setting. Two other settings are possible:

a-axis unique: beta = gamma = 90

c-axis unique: alpha = beta = 90

Relevant Program Options:

- Viewing along unit cell axes (see Viewing Along Crystallographic Axes)
- Creating Miller planes (see Creating Miller Planes)
- Displaying unit cell axes (see Displaying and Labelling Unit Cell Axes)
- Modifying the display of unit cell axes (see Customising the Display of Unit Cell Edges and Labels)
- Colouring conventions for unit cell axes (see Colouring Conventions for Unit Cell Axes)
- Displaying the contents of a unit cell (see Displaying the Contents of a Single Unit Cell)
- Displaying the contents of several unit cells or fractions of a unit cell (see Displaying Several Unit Cells and Fractions of Unit Cells)
- Turning packing on and off (see Switching Packing On and Off)
- Displaying a slice through a crystal (see Defining and Displaying a Slice)

Related Topics

- Crystal System
- Fractional Coordinates
- Miller Plane
- Reciprocal Cell Axes
- Space Group
- *Z* (Number of Molecules in Unit Cell)

van der Waals Radii

Standard van der Waals radii are assigned to the common elements. They are taken from Bondi, *J.Phys.Chem*., **68,** 441, 1964. Other elements are assigned van der Waals radii of 2.0Å.

Standard van der Waals Radii (Å)

Element	Radius
Ag	1.72
Ar	1.88
As	1.85
Au	1.66
Br	1.85
C	1.70
Cd	1.58

Element	Radius
Cl	1.75
Cu	1.40
F	1.47
Ga	1.87
H	1.20
He	1.40
Hg	1.55
I	1.98
In	1,93
K	2.75
Kr	2.02
Li	1.82
Mg	1.73
N	1.55
Na	2.27
Ne	1.54
Ni	1.63
O	1.52
P	1.80
Pb	2.02
Pd	1.63
Pt	1.72
S	1.80
Se	1.90
Si	2.10
Sn	2.17
Te	2.06
Tl	1.96
U	1.86
Xe	2.16
Zn	1.39

Voids

Voids are simply regions of free space within a crystal structure. Depending on the probe size used, this could relate to the solvent accessible volume within the structure, or just the small gaps seen between molecules when the "spacefill" style is used.

The voids in a crystal structure can be calculated and viewed by clicking on Display in the top-level menu and then Voids… from the resulting pull-down menu.

Relevant Program Options:

- Calculating and viewing voids (see Displaying Voids)

xye Format

A format for powder-diffraction patterns. The first line contains the wavelength, in Å. Subsequent lines contain three fields, viz. 2θ, intensity, standard deviation of intensity.

Relevant Program Options:

- Displaying and Saving Powder Diffraction Patterns (see Displaying and Saving Powder Diffraction Patterns)

xyz Format

The xyz format is a general orthogonal coordinates format that is not associated with a particular program, see *http://openbabel.sourceforge.net/wiki/XYZ*.

All atoms that are displayed are written out. No crystallographic information (space group, unit cell) is stored in the file.

Relevant Program Options:

- Saving Structures in External (3rd Party) Formats (see Saving Structures in External (3rd Party) Formats)
- Loading a Structure from a File (see Loading a Structure from a File)

Z (Number of Molecules in Unit Cell)

Although *Z* is normally referred to as the number of *molecules* in the unit cell, it is more strictly the number of *formula units* in the unit cell. For example, if the crystal structure of the compound C_{10} H_{10} Fe_1 N_1 O_2, H_2O had $Z = 2$, this would imply that there were two occurrences of the formula unit C_{10} H_{10} Fe_1 N_1 O_2, H_2O in the unit cell (i.e. strictly speaking, four chemical molecules since the formula unit contains both the metal-containing molecule and a molecule of water).

Relevant Program Options:

- Displaying text information (see Display of Text Information)

Related Topics

- Space Group

Z′ (Number of Molecules in the Asymmetric Unit)

This is best explained with an example. Consider the molecule C_{12} H_{18} N_4 O_2:

- If $Z' = 1$, the asymmetric unit contains 1 molecule and the crystallographer must determine the coordinates of 36 atoms.
- If $Z' = 2$, the asymmetric unit contains 2 molecules and the crystallographer must determine the coordinates of 72 atoms. The two molecules are *crystallographically independent*, i.e. they are not related by symmetry so do not have to have the same geometry.

- If $Z' = 0.5$, the asymmetric unit contains half a molecule and this implies that the molecule possesses symmetry coincident with a crystallographic symmetry element. For example, the molecule might possess mirror symmetry, so that one half of the atoms are related to the other half across a crystallographic mirror plane. In this case, the crystallographer must determine the coordinates of 18 atoms.

In crystal structures of solvates, ion pairs, clathrates, etc., the word *molecule* should be taken to include the entire formula unit, e.g. both the metal-containing moiety and the water molecule for a compound of formula C_{10} H_{10} Fe_1 N_1 O_2, H_2O.

Relevant Program Options:

- Displaying text information (see Display of Text Information)

Related Topics

- Space Group

27 Appendix B: Keyboard Modifiers

Moving Structures

To perform the following actions, click the following combination of mouse and keyboard buttons, drag the mouse, and then release the keys.

Action	Keyboard and Mouse combination
Rotate around X and Y axes	Left mouse-button
Rotate around Z axis	Shift + Left mouse-button
Translate in X and Y planes	Middle mouse button OR Control (PC/Linux)/Command (macOS) + Left mouse-button
Zoom	Right mouse-button, and move mouse up or down
Change perspective ("field of view")	Shift + Right mouse-button, and move mouse up or down

Selection and Context Menus

To perform the following operations, position the cursor over an object (e.g. an atom), then press the following combination of mouse and keyboard buttons, and then release the keys.

Action	Keyboard and Mouse combination
Picking/selecting	Left mouse-button
"Modified" picking/selecting (i.e. selecting whole molecules)	Shift + Left mouse-button
Context menu	Right mouse-button

Note that for macOS, you may need to enable the right-click action on your mouse to be *Secondary Button* in the *Mouse* section of *Settings*. For trackpad users there are additional options for secondary click that can be used, such as allowing a two-finger click to substitute as a right click. These options can be configured in the *Trackpad* section of *Settings* in macOS.

28 Appendix C: Tutorials

28.1 Tutorial 1: Exploring Nonbonded Interactions

Objectives

- To use the *Contacts* functionality in order to find the key intramolecular and intermolecular interactions in a crystal structure.

Steps Required

- Read in the structure.
- Find the hydrogen bonds and display packing along crystallographic axis.
- Find other short nonbonded intramolecular contacts and determine their distances.
- Define a specific type of halogen···halogen interaction, by building up the network.
- Calculate the electrostatic potential using MOPAC (optional).

The Example

- The proportion of pharmaceutical compounds containing fluorine atoms has increased in recent years (J. Wang et al., *Chem. Rev*., 2014, **114**, 2432-2506). The halogen···halogen contacts, along with other nonbonded interactions, can influence the physicochemical properties of pharmaceutical materials including their stability. Therefore, investigating these types of interactions and corelate them with materials behaviour is a common step in solid form analysis.
- This tutorial illustrates how you can investigate the network of intramolecular and intermolecular interactions through contacts for pentafluorobenzoic acid (CSD refcode PFBZAC01; A. Bach, D. Lentz & P. Luger, *J. Phys. Chem. A*, 2001, **105**, 7405):

Menu Commands Required

1. Load the required crystal structure.
   - When Mercury is launched, the Cambridge Structural Database (CSD) should be detected and opened within the *Structure Navigator* on the right-hand side of the main Mercury window. If the CSD database is not detected automatically by

Mercury, you can load it by hitting **CSD-System** in the top-level menu, selecting **Select Databases...** then browsing to the required location.

- Load the structure required for the tutorial by typing *PFBZAC01* in the entry box above the *Structure Navigator* on the right of the Mercury window, or alternatively load the structure from file:
  - Load the structure by hitting **File** in the top-level menu and **Open** in the pull-down menu. You then need to load the file *tutorial1.mol2*. To find this file:
  - If you are using a Windows PC, *tutorial1.mol2* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

    C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

  - On Linux, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the tutorial files will be in:

    /home/cambridge/examples/mercury/tutorial

  - On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

    Applications/CCDC/CSD_2020/Examples/Mercury/tutorial

2. Investigate hydrogen bonding and packing.
   - Change the background colour to white by selecting **Display** from the top-up menu, then **Colours > Background Settings...**. Select white under the **Single colour** option, hit **OK** then **Close**.
   - Click on the tick box next to the words *H-Bond* in the **Display Options** section underneath the 3D visualiser:

   - This finds all the hydrogen bonds formed by the pentafluorobenzoic acid molecule within the crystal structure. The hydrogen bonds are shown as broken red lines for *unexpanded* contacts. Place the cursor anywhere on one of the broken red lines and left-click. This causes the complete molecule at the end of the hydrogen bond to be shown.
   - The display should now look like this:

   - The hydrogen bonds are now displayed in cyan, not red. This is because they have been *expanded*, i.e. the complete molecule is shown at both ends of the hydrogen bond.
   - The carboxylic acid group forms dimers with the hydroxyl group O-H acting as donor and the double bonded oxygen as acceptor.
   - To view how the dimers propagate within the crystal packing, select **Calculate** from the top-level menu and then **Packing/Slicing**. This will open up a dialogue box that enables you to expand the packing along the desired crystallographic axis. Click on the **3x3x3** button and choose the option to Include Atoms **...in molecules where Any atom fits**, hit **OK** then switch on the **Packing** from the *Display Options*.

- You can view the packing along the crystallographic axes by selecting the desired one from the toolbar, select **a** from the top-level options. You can also rotate along the axis by selecting one of the options: **x-, x+, y-, y+, z-, z+**. The options that include **+/-90** rotate the packing 90° in the specified direction.

- Rotate your packing around **z** axis until you obtain a position similar to that shown in the image display below.
- The display of the unit cell axes can also be modified. To ensure that the colour of the unit cell axes shows up on a white background, right click somewhere on the cell axes, select **Colour** and then **Black**. By right clicking on the cell axes you can also select the display as *Wireframe* or *Capped Sticks* from the **Style** option. Select *Capped Sticks*.
- To change the colour of the labels, select **Display** from the top-level menu, then **Colour> Labels**, select **Black** and hit **OK**. Note that the labels of the atoms will be also changed to black. To change the label size, select **Display > Labels > Label Size**. Type 20 and click **OK**. You can also move the labels by selecting **Move Labels** from the **Picking Mode** of the toolbar, then right click and hold on the label you want to move.

- The display should now look like this:

3. Find other short nonbonded contacts and assess how short they are.
   - Reset the display of the 3D visualiser by clicking on the **Reset** button in the Display Options section.
   - Click on the tick box next to the words *Short Contact < (sum of vdW radii)* in the list box underneath the 3D display:

   - This displays all contacts shorter than the sum of van der Waals (vdW) radii. It does not matter whether you leave the H-Bond tick box on or not, since all the H-bonds will also be counted as short contacts, so will be displayed anyway.
   - As you see, there are several short contacts formed between F atoms as well as above and below the planes of the aromatic ring:

- To assess how short these contacts are, we could measure them individually. However, we can colour them by distance, using **Display > Colours > Contacts** from the top-level menu, then **Colour by distance** and hit **All Contacts**.

- To be able to distinguish between the contacts you can display them as Stick by selecting **Display > Style > Contacts…** from the top-level menu. Tick on **Stick** in the *Style* box and then hit **All Contacts**. You can play with the thickness by selecting **Display > Style > Contact settings…**. We used 0.02 **Contact Radius (Angstroms)** for this example:

- As you can see the H-bonds are shown in red as 'medium' short contacts, then the fluorine contacts as 'long' short contacts in blue, with a couple in reddish-purple. The blue contacts are the 'longer' short contacts, followed by purple and then by reddish-purple being the shortest one.

4. Visualise intramolecular interactions.
   - Select **Calculate** from the top-level menu and then **Contacts...** from the resulting menu. This opens up a dialogue box that enables you to tailor the contacts that are found by Mercury.
   - In the *Contacts* dialogue box witch off the *H-Bond* from the *Display Options* box.
   - To view potential non-H-bonded intramolecular interactions, you can customise the definition of Short Contacts, by double-clicking on the words **Short Contact** in the *Display Options* box.
   - In this resulting dialogue box, toggle off the *Intermolecular* checkbox and switch on the *Intramolecular separated by > 3 bonds* option:

- We notice that there are two short intramolecular contacts formed between oxygen atoms of the carboxylic acid group and the F atoms of the aromatic ring.

- We can further investigate these types of contacts by measuring the distances between the atoms involved. Right click in the 3D visualiser and select **Measure > Measure Distances**. Select F1 and O1, then, F5 and O2. The contact distances will appear as labels in the visualiser. To see the errors on these distances for publication purpose, click on **More info** in the *Display Options* box and select **Distances List.** The estimated standard deviations (esds) will be displayed for each measured distance in brackets.

- The observed intramolecular F···O distances are well below the expected van der Waals contact (2.99 Å, ref. Bondi, *A. J. Phys. Chem*., 1964, **68**, 441). A charge density study on this structure has shown that a bond critical point can be found between the F···O atoms, which indicates that there is an interaction between them. In this case, the type of interaction was confirmed as repulsion, with low charge density at the critical points (ref A. Bach et al., *J. Phys. Chem. A,* 2001, **105**, 7405-7412). These repulsive intramolecular interactions might contribute to the slightly tilted orientation of the carboxylic acid.

5. Classifying the halogen···halogen contacts.

- Hit **Default** and **Apply**, then **OK** to close the *Define Short Contacts* box. Click on **Label atoms** in the *Display Options* box. Make sure that the *Short Contacts < (sum of vdW radii)* is still switched on.
- Click on the atom labelled F2 to expand the contact network. Right click and select **Contacts** then **Delete Hanging Contacts**.
- Halogen...halogen contacts are generally observed to have two preferred geometries: type I (θ1 = θ2) and type II (θ1 = 180° and θ2 = 90°); where θ1 = C-X1···X2 and θ2 = X1···X2-C, X1, X2 = halogen atoms. Type I interactions are generally viewed as repulsive interactions, but less repulsive than you will expect due to polar flattening effects, eg: fluorine atoms facing each other exhibiting little polarizability of electron density around the atoms (ref. Hathwar

et al., *Cryst. Growth Des*, 2011, **11**, 1338-1346). Type II interactions are often attractive in nature.

- We can measure the angles of the F···F contacts in the expanded molecular network by right click **Measure > Measure Angles** and select C3-F2···F5 and F2···F5-C6. The two angles are relatively close, and we can confirm this is a type I halogen...halogen contact.

- As mentioned earlier, we would expect that a type I contact will be repulsive in nature, but we could investigate the F5···F2 interaction further using an electrostatic potential calculation.
- Switch off the **Short Contact** from *Display Options* and by right click select **Clear Measurements**.
- If you are interested, and you have access to CSD-Materials, select **CSD-Materials** from top-level menu, then **Calculations** and hit **MOPAC** (Note that MOPAC installation is pre-required for this step). Select **PM3** as the Hamiltonian, tick **Calculate electrostatic potential** and untick all the other boxes then hit **Submit** and **Save**.

- The calculation can take a few minutes. After the run is finished the electrostatic potential surface will be shown in the 3D visualiser. In the *MOPAC* dialogue, under the *Results* tab, then the *ESP* tab, change **Display Type** to **triangle**, set **Opacity** to 0.5
- Change the colour scales for **High** (blue) to 0.06 and **Low** (red) to -0.06 to accentuate the contrast.

- The display should now look like this:

- The surface around the fluorine atoms shows similar electrostatic potential indicating that the F5···F2 interaction is repulsive.
- The repulsive interaction between the O atoms of the carboxylic groups and the neighbouring fluorine atoms can also be spotted from the electrostatic potential map.

This ends the tutorial.

28.2 Tutorial 2: Rationalising Crystal Morphology

Objectives

- To use the *Slicing* functionality in order to understand the morphology of a crystal by examining the functional groups exposed at some {hkl} faces.

Steps Required

- Read in the structure.
- Define the (0 0 2) Miller plane and construct a slice of the crystal either side of this plane.
- Displace the slice to view the chemical groups exposed on the (0 0 2) surface.
- Repeat with the (0 1 1) and (1 0 0) surface.
- Predict the BFDH morphology (optional).

The Example

- Particle shape and size distribution have a direct impact on the materials performance, such as APIs (ref C. Sun, *J. Pharm. Sci.,* 2001, **90**, 569-579). The morphology of a particle upon crystallisation from solution can be dictated by the solvent selection (ref Rosbottom et al. *Cryst. Growth Des.*, 2017, **17**, 4151-4161). Therefore, rationalising solvent selection and it's effect on the morphology represents an important step in obtaining particles of desired shape and size.
- Ibuprofen particles were crystallised in two different morphologies: needle (from less polar and aprotic solvents) and polyhedral (in polar protic solvent).
- The tutorial shows how some aspects of the morphology of ibuprofen can be explained and controlled by displaying slices through the crystal and looking at the polarity of functional groups associated with certain Miller planes.

- This example (CSD refcode IBPRAC) is taken from T. T. H. Nguyen et al*. Cryst. Growth Des.*, 2017, **17**, 3088-3099. The ibuprofen crystal structure is taken from J. F. McConnell, *Cryst. Struct. Commun.*, 1974, **3**, 73.

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure required for the tutorial by typing IBPRAC in the entry box above the Structure Navigator on the right of the Mercury window, or alternatively load the structure by hitting **File** in the top-level menu and **Open** in the pull-down menu. You then need to load the file *tutorial2.mol2*. To find this file:
     - If you are using a Windows PC, *tutorial2.mol2* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

       C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial
     - On Unix, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the tutorial files will be in:

       /home/cambridge/examples/mercury/tutorial
     - On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

       Applications/CCDC/CSD_2020/Examples/Mercury/tutorial
2. Examine Ibuprofen morphology
   - Ibuprofen morphology was calculated using BFDH Morphology as illustrated below:

- You can see that the (1 0 0) face is broader than the other faces. The (0 1 1) and (0 0 2) faces were shown to be similar by anisotropy factor calculations (ref T. T. H. Nguyen et al. *Cryst. Growth Des.*, 2017, **17**, 3088-3099). The anisotropy factor measures the degree of saturation of the intermolecular interactions exposed to a specific crystal surface (hkl).

- To examine the (0 1 1) and (0 0 2) planes select **Calculate** from the top-level menu, then **Packing/Slicing…**, toggle on the **Show slice** checkbox and select the (0 0 2) plane from the **Select plane** drop-down. Toggle off the **Show Bonds** and change the **Depth** to 5.00 and the **Displacement** to -2.2.

- View the structure edge-on to the (0 0 2) plane by right-clicking on the plane and selecting **View along plane** from the pull-down menu. Then rotate the display slightly so that you are looking almost, but not quite, edge-on to the plane. Check the **Short Contact** box on from the *Display Options* box. You can see that the predominant groups on this face are involved in van der Waals interactions:

- A polar protic solvent will bind to this surface and prevent the formation of needle morphology.
- If the plane you want to visualise is not in the **Select plane** list you can select **New** and click on the *hkl* tab, then type in the Miller indices. This will also display the plane labels.
- Change the plane to (0 1 1) and switch on the **H-Bond** from the *Display Option* and toggle off the **Short Contact**. The predominant groups on this face are involved in hydrogen bonding interactions:

- The (1 0 0) face is the slower growing face and exhibits the lowest attachment energy (ref T. T. H. Nguyen et al., *Cryst. Growth Des.*, 2017, **17**, 3088-3099). Select this plane and change the **Style** in the tool-bar (near the top of the main Mercury window) to **Spacefill**.
- Toggle off **H-Bond** from *Display Options* dialogue box to focus on groups.
- The groups exposed to this surface are all hydrophobic and this explains why this face is dominant in the crystal morphology.

3. Calculate the predicted BFDH morphology.
   - If you are interested to predict the morphology and you have access to CSD-Materials, hit **CSD-Materials** and then **Calculations** in the top-level menu, then **BFDH Morphology...** in the resulting pull-down menu.
   - Leave all toggle boxes ticked and hit **OK**.
   - The predicted crystal morphology will be displayed in the correctly aligned orientation with respect to the unit cell. The faces are labelled by their Miller Indices.
   - You can align the morphology with the slice plane. In the *Morphology* dialogue box, toggle the **Auto Scale** and set the **Scale factor** to 200. In the *Packing and Slicing* dialog box you can increase the **Area** to 40.

- As mentioned above the largest face found on the predicted crystal form is the (1 0 0) face, in agreement with experiment.
- To create an image of the morphology as the one at the beginning of the Tutorial toggle off **Show slice** in the *Packing and Slicing* dialog box. Right click on the slice plane and select **Delete Object**, then right click on the molecule and select **Delete molecule**. Hit **Colour** in the *Morphology* dialog box and select **White**, then change the **Draw edges Colour** to **Black**.

This ends the tutorial.

28.3 Tutorial 3: Investigating Pseudosymmetry

Objectives

- To find a chain of hydrogen bonds in a crystal structure.
- To view the chain down various crystallographic axes.
- To establish that the chain is an approximate 3-fold helix.

Steps Required

- Read in the structure.
- Find the chain of hydrogen bonds that links neighbouring molecules together.
- Adjust the colouring scheme so that crystallographically equivalent molecules have the same colour.
- View the H-bond chain from several different directions, including both down and perpendicular to crystallographic axes.
- Measure geometrical parameters.

The Example

- The tutorial shows that the crystal structure of:

contains an approximate 3-fold helix.

- The structure is taken from Harper, J.K., Arif, A.M. & Grant, D.M., *Acta Cryst.* **C56**, 451, 2000.

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure required for the tutorial by typing *KIXDUQ* in the entry box above the *Structure Navigator* on the right of the Mercury window, or alternatively load the structure from file:
     - Load the structure by hitting **File** in the top-level menu and **Open** in the pull-down menu. You then need to load the file *tutorial3.mry*. To find this file:
     - If you are using a Windows PC, *tutorial3.mry* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

       C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

     - On Unix, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the tutorial files will be in:

       /home/cambridge/examples/mercury/tutorial
     - On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

       Applications/CCDC/CSD_2020/Examples/Mercury/tutorial
   - There are three molecules in the asymmetric unit. This is unusual. The structure has been refined in the orthorhombic space group $P2_12_12_1$ (this is displayed in the bar at the very top of the Mercury window).
2. Find the chain of hydrogen bonds that this structure contains.
   - Switch on the tick box next to the words *H-Bond Default definition* in the list box underneath the main display area:

   - This displays the hydrogen bonds between the hydroxyl groups. To add another molecule to the display, click with the left-hand mouse button on either of the two *hanging contacts*, i.e. either of the contacts shown as broken red lines.
   - Repeat this until you have a chain of about six molecules:

3. Hide the hydrogen atoms and colour molecules by symmetry equivalence.
   - Simplify the display by hiding the hydrogen atoms. An easy way to do this is to turn off the **Show hydrogens** tick box near the bottom of the Mercury window:

   - You will see a white box labelled *Colour* in the tool bar near the top of the Mercury window. Change the setting of this box to **by Symmetry equivalence**:

   - The three different molecules in the original crystallographic unit are now assigned different colours. Symmetry related molecules are assigned the same colour. Your display should look something like this:

4. Look at the structure down different crystallographic axes.
   - Click on the letter **b** in the tool bar just above the display area:

   - This alters the view direction so that you are looking down the crystallographic *b* axis. To demonstrate this, you can click on the **Show cell axes** tick box underneath the main display area:

   - Now alter the direction of view by clicking on the first item in the tool bar, labelled **a.**
   - You are now looking down the crystallographic *a* axis and the significance of this view is immediately clear. The three crystallographically independent molecules are arranged in an almost perfect 3-fold helix:

- At this stage, we must consider the possibility that the molecules are not actually crystallographically independent but are related by a 3-fold screw axis. This would mean that the structure had been refined in a space group of lower symmetry than the true one.

5. Inspect the helix carefully.

   - Select **x-90** from the tool bar to view the helix in a different (viz. orthogonal) direction. If necessary, use the arrows towards the right-hand end of the tool bar to translate the image, and the **zoom** buttons at the very end of the tool bar if the scale needs to be altered:

   - If the symmetry is perfect, all three C-O bonds should subtend the same angle to the supposed 3-fold axis. Visual inspection suggests that the angles are similar but not quite the same. Use the tool-bar button **y-90** to rotate around the axis and get a better impression (you may want to turn the cell axes back off).

6. Measure molecular dimensions.

   - You can measure equivalent bond lengths, valence angles and torsion angles in the three different molecules to see how well their values compare (and you can also measure the O...O hydrogen-bond distances). Set the *Picking Mode* status (near the top of the Mercury window) to **Measure Distance, Measure Angle** or **Measure Torsion***,* depending on which you wish to measure, then click on atoms in the display area to make the measurements.
   - The molecules are extremely similar in geometry, though not identical. The three-fold symmetry is close but not exact.
   - This is as far as Mercury can take you - there might be crystallographically exact 3-fold symmetry or it might be only approximate pseudosymmetry.
   - The only way of making sure is to input the cell dimensions to a program such as *Le Page* which can hunt for unit cells that are consistent with higher symmetry

(*http://www.ccp14.ac.uk*). As it happens, no unit cells of higher crystal symmetry can be found - the crystallographers' original choice of space group appears to be right and the apparent 3-fold symmetry is only approximate.

This ends the tutorial.

28.4 Tutorial 4: Looking for Possible Missing Water Molecules

Objectives

- To investigate the hydrogen bonding in a putative crystal structure.
- To decide whether a water molecule has been omitted from the structure.

Steps Required

- Read in the structure.
- Perform an initial inspection of the hydrogen bonds.
- Examine in detail the hydrogen-bonding environment of a particular atom, using a customised H-bond definition.
- Draw the appropriate inference and further investigate by visual inspection of spacefill displays.

The Example

- Structure solution from powder diffraction data is becoming increasingly viable, but it is not as reliable as conventional single-crystal analysis. Any solution based on powder data should be examined closely in a program such as Mercury to ensure that it contains no suspicious features.
- This example presents a structure that we will suppose came from a powder diffraction study. It is of the following molecule:

- We will examine the hydrogen bonding arrangement in this supposed structure to see whether it all makes sense.

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure by hitting **File** in the top-level menu and **Open** in the pull-down menu. You then need to load the file *tutorial4.mry.* To find this file:
     - If you are using a Windows PC, *tutorial4.mry* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

       C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

     - On Unix, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge,* the tutorial files will be in:

       /home/cambridge/examples/mercury/tutorial
     - On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

       Applications/CCDC/CSD_2020/Examples/Mercury/tutorial
2. Inspect the molecule.
   - First look at the molecule. It is an α-amino acid and, although the hydrogen atoms have not been located, it is highly likely that the molecule is a zwitterion (i.e. the acid group is anionic, $-CO_2^-$, and the amino group is cationic, $-NH_3^+$). The molecule also contains an aliphatic alcohol group and a pyridine ring with an -OH *meta* to the ring nitrogen.
3. Find and inspect the hydrogen bonds.
   - Switch on the tick box next to the words *H-Bond Default definition* in the list box underneath the main display area:

   - The hydrogen bonds formed by the molecule are shown as broken lines. If you have not seen this type of display before, you are advised to do Tutorial 1 (see Tutorial 1: Exploring Nonbonded Interactions) before continuing with this exercise.
   - The hydrogen bonds displayed in red are *hanging*, i.e. the complete molecule at one end of the contact is shown, but only one atom from the molecule at the other end. However, one of the hydrogen bonds happens to be intramolecular, which means it is not a hanging contact and is shown in cyan.
   - First, look at the hydroxyl group on the pyridine ring. It forms two hydrogen bonds, as you would expect:

- Display the molecules at the end of these bonds by left-clicking on each of the broken red lines. We see that the pyridine -OH accepts a hydrogen bond from the $-NH_3^+$ group and donates to one of the carboxylate oxygens.
- The carboxylate oxygen that accepts a hydrogen bond from the pyridine hydroxyl group also forms a second H-bond, indicated by the broken red line:

- Click on the broken red line to see what is at the other end. It is the alcohol OH. Since we assume that the acid group is ionised, the donor in this hydrogen bond must be the alcohol oxygen.
- We also see that the alcohol oxygen (labelled O3 below) accepts an intramolecular hydrogen bond from the $-NH_3^+$ group:

4. Review whether the hydrogen-bonding capabilities of each group are fully satisfied.
   - The hydrogen bonds we have found satisfy the full hydrogen-bonding potential of both hydroxyl groups (they form two H-bonds each, as expected) and one of the carboxylate oxygens. However, we have not found any hydrogen bonds to the other carboxylate oxygen or the pyridine nitrogen. Even more surprising, only two of the three $-NH_3^+$ protons appear to be H-bonded. If true, this would be very unusual.
5. Explore the hydrogen bonding of the $-NH_3^+$ group more carefully.
   - We will now concentrate on exploring the environment of the $-NH_3^+$ group and, in particular, finding out what the third proton might be doing.
   - Remove the network we have created so far by hitting the **Reset** button at the bottom left-hand corner of the Mercury window. This takes us back to the original single-molecule display.
   - Hit **Calculate** in the top-level menu and **Contacts...** in the pull-down menu. This opens the *Contacts* dialogue. Click on the **Add** button and select **H-bond** from the pull-down menu. This opens the *Define H-bonds* dialogue, which enables you to define and search for particular types of hydrogen bonds.
   - The tick boxes in the lists headed *Donor atom types* and *Acceptor atom types* enable you to switch on or off particular sorts of donors and acceptors. However, it is sometimes difficult for Mercury to assign donor and acceptor types accurately, especially if hydrogen atoms are missing from the structure, so it is unwise to be too specific. In the present case, a reasonable compromise is to turn all nitrogen donors on and all oxygen donors off. This will ensure that we only find hydrogen bonds involving the $-NH_3^+$ group, not the -OH groups. Ensure, therefore, that the nitrogen tick box in the donor list is on and that the oxygen tick box is off:

- Turn all acceptors on:

- Now hit the **Apply** button. Your new, customised hydrogen-bond definition appears in the list box underneath the Mercury display area:

- Mercury automatically searches for and displays all hydrogen bonds of the type you have specified, i.e. between any sort of NH donor and any type of acceptor. These are displayed as usual in the display area.
- We still see only two hydrogen bonds from the $-NH_3^+$ group:

- However, we can try the effect of increasing the maximum allowed length of hydrogen bonds by changing *Maximum = sum of vdW radii plus 0.0* to *Maximum = sum of vdW radii plus 0.3* in the *Define H-bonds* dialogue and hitting **Apply**:

- There is still no third hydrogen bond to the $-NH_3^+$ group.
- Increase further to *Maximum = sum of vdW radii plus 0.5* and hit **Apply** again.
- At last, a third contact to the $-NH_3^+$ nitrogen appears. Left-click on it to show the molecule at the other end. The interaction is with the second of the carboxylate oxygens - the one that did not seem to be forming any H-bonds in our earlier inspection.

- However, the *hydrogen bond* is suspiciously long: you can measure it as 3.408Å by setting *Picking Mode* (in the tool bar near the top of the main Mercury window) to **Measure Distance** and then selecting the N and O atoms.
- When you measure contact distances, you may find that you inadvertently expand the contact instead, i.e. add a new molecule to the end of a hanging contact. If this happens, it is because you did not click close enough to the atom. You can remove the unwanted molecule by right-clicking on it and hitting **Delete this Molecule** from the pull-down menu.
- Compare with the distances of the other N...O hydrogen bonds, both < 2.9Å.

6. A missing water molecule.
    - An obvious inference at this point is that a water molecule is missing from the structure, an error that might easily occur in a crystal structure solution based on powder diffraction data.
    - Establish if there is any empty space in the crystal structure by displaying Voids. To do this select **Display** then **Voids**. Keep the default settings and hit **Apply**. It is quite obvious there are unoccupied areas in the crystal structure that might accommodate a water molecule.
    - In fact, there *is* a water molecule in the correct structure - we took it out for the purposes of the exercise. Read in the file *correct.mry* (which you will find in the same directory as the file you read in at Step 1) to see the true structure and the correct hydrogen-bond network.
    - The correct structure (which is actually not a powder structure but a single-crystal determination) is from Konig, W.A., Pfaff, K.-P., Bartsch, H.-H., Schmalle, H. & Hagenmaier, H. (*Liebigs Ann. Chem*., **1980**, 1728) to whom we apologise for abusing their data!

This ends the tutorial.

28.5 Tutorial 5: Investigating Polymorphism

Objectives

- To use the overlay functionality in Mercury to investigate the differences in the extended crystal structures of a pair of polymorphs.

Steps Required

- Read in a structure.
- Display additional information about the structure.
- Perform an initial inspection of the hydrogen bonds.
- Display a second crystal structure simultaneously.
- Overlay the two structures and compare their H-bonding motifs.

The Example

- The understanding of polymorphism (i.e. why a given molecule adopts a particular crystal packing arrangement) is of extreme importance in the pharmaceutical industry, particularly for bioavailability and process problems, and could develop knowledge in various fields such as nucleation and crystal growth.
- This example investigates differences in the hydrogen bonding of the orthorhombic and monoclinic polymorphs of naphthalene tetroxide imine:

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure required for the tutorial by typing *BAZYAC* in the entry box above the *Structure Navigator* to the right of the Mercury window. If you do not have access to the full CSD, open the example file *tutorial5a.mry* from the *tutorials* directory of the Mercury installation. To find this file:
     - If you are using a Windows PC, the example file will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

       C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

- On Unix, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge,* the tutorial files will be in:

  /home/cambridge/examples/mercury/tutorial

- On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

  Applications/CCDC/CSD_2020/Examples/Mercury/tutorial

- To display additional information about the structure, click on the **More Info** button (bottom centre of the main Mercury window) and then select **Structure Information...** from the resulting menu. Note that this additional information will not be visible for the structure if you used the example MRY file.

- We can see that the structure of naphthalene tetroxide imine displayed is the orthorhombic polymorph.
  Note: If no polymorph data item is displayed then the list of data items can be changed by clicking on the **Customise...** button in the top-left corner. To add a new item, select the required item by clicking on it in the left-hand list, then hit the **Add** button.

2. Investigate the hydrogen bonding.

- Switch on the tick box next to the words **H-Bond Default definition** in the list box underneath the main display area:

- No hydrogen bonds are found. However the structure does contain an NH donor (the hydrogen atom is missing). We can try the effect of increasing the maximum allowed length of hydrogen bonds by changing the default H-bond definition.
- Double-click on the words **H-Bond Default definition** in the list box underneath the main display area, this will open the *Define H-bonds* dialogue. Select the **Actual Distance** radio button, change the maximum value to 3.2Å then hit **Apply**:

- Close the *Define H-bonds* dialogue.
- Two hanging hydrogen bonds are now displayed. The NH hydrogen atom coordinates are not available for this structure. However, we can infer that one hydrogen bond involves the NH donating its hydrogen. Also, the epoxy oxygen atom can be seen to accept a hydrogen bond.
- Click on the hanging H-bond from the nitrogen (not the oxygen) to add another molecule to the display. Do the same from the hanging contact from the nitrogen of the new molecule.
- We should now have a hydrogen-bonded chain consisting of three molecules:

- Simplify the display by hiding the hydrogen atoms. To do this is to turn off the **Show hydrogens** tick box near the bottom of the Mercury window.

3. Load the second polymorph of naphthalene tetroxide imine.
   - To enable the display of multiple crystal structures, switch on the **Multiple Structures** tick box located under the list of structures on the right-hand side of the main Mercury window.
   - Load the structure by typing *BAZYAC01* in the entry box above the scrolling list at the right of the Mercury window. Alternatively, load *tutorial5b.mry* from the *tutorials* directory of the Mercury installation.
     Note: If more than one study of a compound is present in the CSD, the second and subsequent entries will have two numbers after the six letter refcode, i.e. polymorphic forms of the same structure are included in the same refcode "family".
   - In structure information spreadsheet window notice that *BAZYAC01* is the monoclinic polymorph.
4. Compare the two crystal structures.
   - Hit the **Structures...** button located under the list of structures on the right-hand side of the main Mercury window. This will open the *Multiple Structures* dialogue. The two structures currently being displayed are listed:

- Turn off the display of *BAZYAC* by switching off the **Visible** check-box for that structure. Only *BAZYAC01* should now be visible in the main Mercury display:

- Click on the two hanging H-bonds shown from the molecule currently being displayed. We should now have a hydrogen-bonded chain consisting of three molecules (similar to that constructed for *BAZYAC* in part 3 of this tutorial).
- Make *BAZYAC* visible again by switching on the **Visible** check-box in the *Multiple Structures* dialogue. In order to distinguish *BAZYAC* from *BAZYAC01* we can colour it differently. In the *Multiple Structures* dialogue, click on the **Colour** pull-down list for *BAZYAC* and select yellow.

- Switch off the **Movable** check-box for BAZYAC. Subsequent actions e.g. rotation will be applied to movable structures (*BAZYAC01*) only. Position the chain of *BAZYAC01* molecules so that they are in a similar orientation to *BAZYAC*:

- Make *BAZYAC* movable again by switching on the **Movable** check-box in the *Multiple Structures* dialogue. Rotating the display will now affect both structures.

5. Overlay the structures.

- Select **Calculate** from the top-level menu followed by **Structure Overlay...** from the resulting pull-down menu. This will open the *Manual Crystal Overlay* dialogue.
- Pairs of atoms in *BAZYAC* and *BAZYAC01* must be selected in order to overlay the two crystal structures. These pairs of atoms will be used to perform the least-squares overlay.
- Select the nitrogen atom in the first molecule in the chain of *BAZYAC* (coloured yellow) followed by the corresponding nitrogen atom in *BAZYAC01.* A broken yellow line will be displayed between the selected pair. Repeat this procedure for the two remaining pairs of nitrogen atoms in the display:

- As pairs of atoms are selected they will be listed in the *Manual Crystal Overlay* dialogue:

- Once the three pairs of nitrogen atoms are selected, hit the **Overlay** button.
- Overlay is performed by moving the second crystal structure (*BAZYAC01*) relative to the first crystal structure. During overlay, crystal structures are kept rigid. The root-mean-square (RMS) of the pairs of atoms used for the overlay will be displayed in the *Manual Crystal Overlay* dialogue.
- The display should now look like this:

- Inspect the structures more closely. Notice that both structures show similar but not identical H-bonded chains of molecules along the crystallographic *b* axis. In particular, one of the hydrogen bonds forms to a different oxygen atom in *BAZYAC* than in *BAZYAC01.*
- This is common in polymorphism. Structures often show a common motif, but the chains or ribbons or molecules connected by the motif are packed differently in the third dimension.
- The example used in this tutorial is from Gervais and Coquerel, *Acta Cryst*. (2002). **B58**, 662-672.

This ends the tutorial.

28.6 Tutorial 6: Analysing interactions, planes and ADPs

Objectives

- To calculate planes and interplanar angles between molecules in order to evaluate intermolecular geometries in crystal structures.

Steps Required

- Read in the structure.
- Calculate planes and angles between planes.
- Calculate intermolecular potentials using UNI force field (optional).

The Example

- Functional groups can influence molecular orientation and assembly in multi-component molecular complexes such as co-crystals in the solid-state material (ref I. Sovago et al., *Cryst. Eng. Comm*, 2016, **18**, 5697-5709). The properties of such molecular complexes are therefore influenced by the selection of the co-formers used in the co-crystallisation process. Tuning the physicochemical characteristics to design systems with favourable properties represents a powerful tool in crystal engineering.
- This tutorial explains how to investigate molecular complexes exhibiting different molecular orientations in the 1:1 and 2:1 stoichiometries. The CSD refcodes (CECYIV01 and WUXYUJ02) used in this tutorial are taken from: I. Sovago et al., *Cryst. Eng. Comm*, 2016, **18**, 5697-5709.

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure required for the tutorial by typing WUXYUJ02 in the entry box at the top of the *Structure Navigator* on the right of the Mercury window, or alternatively load the structure from file:
     - Load the structure by hitting **File** in the top-level menu and **Open** in the pull-down menu. You then need to load the file *tutorial6a.mol2*. To find this file:
     - If you are using a Windows PC, *tutorial6a.mol2* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

- On Unix, it will be in the directory examples/mercury/tutorial in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the tutorial files will be in:

  /home/cambridge/exampes/mercury/tutorial

- On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

  Applications/CCDC/CSD_2020/Examples/Mercury/tutorial

2. Examine the H-bonding network.

   - The CSD refcode WUXYUJ02 represents the 2,4-dimethylpyridinium (lutidinium) hydrogen chloranilate structure in a 1:1 stoichiometry.
   - Tick on the *H-Bond Default definition* in the *Display Options* dialog. You can see that the N+-H is acting as a donor and the deprotonated oxygen is acting as an acceptor. The adjacent oxygen atom of the chloranilate anion lying in close proximity presents the possibility of bifurcated hydrogen bonds to be formed.

   - You can measure the distances between the donor (N+) and the acceptors (O) atoms, as well as the N+-H···O angles. The hydrogen bond formed between the N+-H and the deprotonated O is quite short and linear (168°; so likely to be relatively strong), while a much longer, less linear contact can be seen with the adjacent O atom (117°; so typically not considered as a conventional H-bond).

3. Measure the angle between two planes.
   - The lutidinium (donor) and chloranilate anion (acceptor) molecules can be described as two different molecular planes.
   - Hold shift and left-click on the lutidinium molecule.
   - Select **Calculate** from the top-level menu, then **Planes...** from the resultant pull-down menu. In the Planes window, click on the **New Plane...** button, then in the **Planes Properties** dialog select **Picked molecule** and hit **OK**:

- The plane fitted to the lutidinium counterion should now be displayed in the 3D visualiser:

- Repeat the procedure to define the chloranilate anion plane. Ensure that *Pick Atoms* is selected in the *Picking Mode* drop-down tool bar. Select the **Planes Properties** and chose the blue colour as indicated below, then hit **OK**:

- You should now see the two planes in the 3D visualizer:

- To calculate the angle between the planes right click in the 3D visualiser then select **Measure > Measure Angles**. After clicking on both planes, the angle will be displayed as shown in picture above.
- We will now compare the angle between the two planes for the bis(2,4-dimethylpyridinium) chloranilate (2:1 stoichiometry).
- Close the Planes dialog, and type CECYIV01 in the Structure Navigator dialog box. Ensure that the *Pick Atoms* is selected in the *Picking Mode* drop-down tool bar.
- Hold shift and left click on the lutidinium molecule.
- Select **Calculate** from the top-level menu, then **Planes...** from the resultant pull down menu. In the *Planes* window, click on the **New Plane...** button. In the **Planes Properties** dialog select **Picked molecule** and hit **OK**:
- Your display should look like this:

- Repeat the procedure to define the chloranilate dianion plane. You can now measure the angle between the two planes.

- The lutidinium and chloranilate molecules lie in approximately the same plane with the angle between them of 1.6° in the 2:1 stoichiometry, whereas in the 1:1 the rings are distinctly skewed.

4. Intermolecular potential calculations
   - We will now compare the intermolecular interaction strength in the single and double deprotonated chloranilic acid molecular complexes.
   - Close the *Plane Properties* and *Planes dialogs* and hit **Reset** in the *Display Options* dialog to remove the planes.

- Hold Shift and select the two molecules, then select **CSD-Materials > Calculations > UNI intermolecular potentials…** In the *Intermolecular Potentials* dialog enable **Show interactions between selected molecules**. The intermolecular interaction potential is display in the 3D visualiser. The value is in kJ/mol. If you want to change the units to kCal/mol you can do this in the *Highlighting* tab in the *Intermolecular Potentials* dialog. As you can see the intermolecular potential in 2:1 stoichiometry is -9.3 kJ/mol.

- To calculate the intermolecular potential in 1:1 stoichiometry type WUXYUJ02 in the *Structure Navigator* dialog. Hold shift and select the two molecules in the 3D visualiser then hit **Calculate** in the *Intermolecular potentials* dialog. You can see that the interaction between the two molecules is stronger (-19.9 kJ/mol) than in the 2:1 stoichiometry.

- This tool can help correlate the orientation of molecules and the stoichiometries in molecular complexes with the intermolecular interaction strength.

5. Display Ellipsoids.

- Select the WUXYUJ03 refcode in the structure navigator then modify the **Style** to *Ellipsoid* from the top-level toolbar.

- The Anisotropic Displacement Parameters (ADPs) will be displayed for the atoms in the 3D visualiser. Note that in this case the ADPs for H atoms are also displayed as the structure was determined using neutron diffraction.

- Displaying the ADPs can help for instance to interpret crystal structures that exhibit disorder. If you load CSD refcode WUXZAQ03, which is the 2,5-lutidinium chlorinate (1:1), you can see that the Hydrogen atoms of one of the methyl groups are heavily disordered.

- You can modify the display style of ellipsoids by selecting **Display > Manage Styles…** from the top-level menu, select **Edit** and then **Ellipsoid Settings…**You

can change the colour of the principal ellipses, the Probability level (normally displayed as 50%), bond style, radius etc. Note that the Non-Positive-Definite atoms will be displayed by default as cubes.

This ends the tutorial.

28.7 Tutorial 7: Assessing the Risk of Polymorphism via H-Bonding Propensities

Objectives

- Create an H-bond propensity model using CSD structures.
- Perform predictions on a polymorphic target compound.
- Explore the H-bonding landscape via use of coordination likelihoods combined with H-bond propensities and assess whether there is a possibility of a more stable form (i.e. risk of polymorphism).

Steps Required

- Start a H-Bond propensity calculation.
- Specify a target structure (prediction molecule).
- Examine the functional groups that describe the target structure.
- Search the CSD to generate an appropriate data set.
- Build the statistical H-Bond propensity model.
- Assess the model.
- Apply the model to the target.

The Example

- The existence of multiple, different crystalline forms (polymorphs) of a compound can have significant implications on both the physicochemical properties of the material and also the control of that material from a regulatory viewpoint. Finding the most stable polymorph and understanding the polymorphic landscape is an important process when dealing with solid state materials and especially in early stage formulation within the pharmaceutical industry. The Hydrogen Bond Propensity (HBP) tool in Mercury can be used to assess the potential for polymorphism and the relative likelihoods of the H-bonding networks in any observed polymorphs for a specific target system.
- In this example we are going to investigate the stability and polymorphism risk of bicalutamide (Casodex) using HBP. Two polymorphs are known for this structure with CSD refcodes JAYCES (form I) and JAYCES02 (form II). There is a third entry in this CSD refcode family - JAYCES01 - but this is the same polymorph as JAYCES (form I) and the R-factor of the first structure is lower, so we will use this one for the analysis. The relative stability between forms I and II has been determined experimentally at a range of temperatures, with the system identified as monotropic and form I was shown to be the stable polymorph (D. R. Vega et al., *Int. J. Pharm.*, 2007, **328**, 112-118). The crystal structures used in this tutorial were published in: Xiu-Rong Hu et al., *Acta Cryst. E*, 2005, **6**, 3897 and D. R. Vega et al., *Int. J. Pharm.*, 2007, **328**, 112-118.

Menu Commands Required

1. Load the required crystal structure.
   - Load the structure required for the tutorial by typing JAYCES in the entry box above the Structure Navigator on the right of the Mercury window.
2. Examine the H bonding network.
   - The compound bicalutamide exhibits 4 potential H-bond acceptors (three O and one N atoms) and 2 potential donors (O-H and N-H groups):

   - We are now going to compare the H-bond network of the two bicalutamide polymorphs and examine which acceptors and donors are being used.
   - With *JAYCES* (Form I) loaded, toggle on the **H-Bond** checkbox in the *Display Options* dialogue, then right-click on **H-Bond** and select **Edit** to check the definition being used. Ensure that the option to **Require hydrogen atom to be present** is toggled on - this will make sure that H-bonds are measured from the hydrogen, not the heteroatom, which is always a more accurate approach if hydrogens are included in the structure. Click on **OK** to accept the settings.
   - Now expand the hanging contact from the O-H donor by clicking on the hanging atoms red dotted line to obtain a similar picture to that below. A single H-bond

interaction is found between the carbonyl C=O and the hydroxy OH groups, so it looks like just one donor and one acceptor are used in Form I.

- In this case, after expanding the first contact, we can also see that the N-H group is pointing directly towards one of the sulfonyl O atoms. To explore whether this is a significant interaction that we should consider, right-click on **H-Bond** again and choose **Edit**.
- Increase the **Contact distance range** such that the **Maximum = sum of vdW radii** setting is 0.1 Å, rather than the default setting of 0.0 and hit **OK**. You will see that a second H-bond interaction appears with quite a linear angle (D-H...A angle of 155°) - this is likely to be a significant interaction in the structure. It is always worth looking critically at the crystal structure and molecular packing to ensure that all the key H-bond interactions are captured.
- We can therefore see that there are two donors and two acceptors used in Form I of bicalutamide.

- You can follow the same steps to load form II: *JAYCES02* in the structure navigator. In Form II the H-bond interactions are formed between the sulfonyl S=O and amide NH groups, and the nitrile CN and hydroxy OH groups. Note that Form II has two different donors and acceptors.

3. Calculate Hydrogen Bond Propensity for Form II.

- Hit the **Reset** button in the *Display Options* dialogue box to display only one molecule of *JAYCES02* in the 3D visualiser. From the top-level menu select **CSD-Materials > Polymorph Assessment > Hydrogen Bond Propensities…** This will open the *Propensity Prediction Wizard*.

- Select a working directory by entering the full path or by clicking on **Browse…** (a)
- The molecule displayed in the 3D visualiser will be automatically assigned as the prediction molecule (b). At this stage you can still modify the prediction molecule by selecting another entry from structure navigator and then click on **Update Structure** button. The potential hydrogen bond donors and acceptor atoms are automatically identified and linked to their functional groups. For bicalutamide, two donor atoms have been identified: N1 defined as ar_al_trans_amide functional group and O3 as part of acyclic_al_oh group. There are five acceptors identified. Note that O3 is identified as both donor and acceptor as standard for a hydroxy group.
- Toggle on the *Show advanced options* checkbox in the *Propensity Prediction Wizard*, then click on **Edit** button next to **Hydrogen bond definition**. To make

sure that the second hydrogen bond present in form I is taken into account, change the *Maximum = sum of vdW radii* setting to 0.1 Å and hit **OK**.

- The *Donors* and *Acceptors* atoms can be highlighted in the 2D chemical diagram by selecting them from the list.
- You can also highlight a functional group from the *Match from library* list, the corresponding atoms will be automatically highlighted in the *Donors* and *Acceptors* lists (c). The functional group as defined will appear in the second window of the *Functional groups* dialogue box.

- You can adjust the functional groups if needed by either:
  - **Add…** pre-defined groups from a functional group library
  - **Sketch…** your own groups with a sketcher
  - **Load…** groups from saved .con (connser)
  - **Edit...** a group with the sketcher
  - **Remove** a functional group definition
- In general, the functional groups should be correctly assigned, hence we will leave all the settings as default for this example and click on **Next**.
- We can now start to retrieve structures from the CSD with the most relevant chemistry to our target structure based on the functional group definitions.
- In the *Generate Fitting Data* pane of the wizard ensure that the *Start analysis automatically* checkbox is unticked and click on the **Generate** button. This may take a few minutes to run depending on the speed of your PC.

- As the training set (generated fitting data) starts to be populated with CSD structures the functional groups and an indication of their **Count** and **Advice** can be seen.

- After the generation is completed, the total number of structures found for each group is listed. Sometimes, this can be uneven, and is a good practice to aim for a model with groups evenly represented. This can be achieved by using the slider highlighted in blue which allows you to remove or add structures until a more even set of data are obtained.

- Adjust the slider until there are roughly equal numbers of each group. In general, around 300-400 structures per functional group should be enough. In this case, we can use a few more structures as there are only four functional groups involved. Adjust the slider till you have around 2000 structures, then hit **Analyse**.

- Once the analysis is finished the number of the true and false outcomes will be listed.
- The wizard would automatically tick the **Ignore?** checkboxes for any functional groups that had very low numbers of true or false hits in the analysis. These would also be highlighted in red. There are no very low values in this example, click the **Fit Model >** button to continue.

- In this case, you should achieve an Area under the ROC (receiver operating characteristic) curve (AUC) of around 0.75. In general, we are aiming for an AUC of around 0.75 or above to achieve a good H-bond prediction. Further information on assessing the goodness-of-fit is provided elsewhere (see Assessing a Statistical Model).
- Select **Accept & Calculate** to continue.

- The results of the HBP calculation will be displayed. The chart will plot the Mean H-bond Propensity vs the Mean H-Bond Co-ordination for all the putative H-bonding networks possible. The *Propensity* and *Co-ordination scores* are listed in the tables on the right-hand side of the *Calculate Propensities* pane of the wizard.
- If a specific H-bond is observed in your target structure this will be marked as **observed** in the *Propensity scores* table. The table is also interactive and by clicking on **observed** will highlight the donor and acceptor group in the Mercury 3D visualiser. Clicking on an atom label in either the *Donor* or *Acceptor* columns will highlight the functional group and label the atom in the 3D visualiser.

- You can zoom on the chart using the magnifying glass icon in the lower left-hand corner of the wizard, to go back to the default option press **Reset**. When you click on the points in the chart the selected triangle, or circle will have a little black arrow underneath, e.g.:

- You can save the results in HTML format by hitting the **Publish Report…** bottom.
- The observed H-bonding network in the loaded crystal structure is represented as a magenta circle in the chart. If you click on it, you can see that the corresponding H-bonds will be highlighted in blue in the *Propensity score* table. Our targeted polymorph has two hydrogen bond interactions highlighted in blue in the *Propensity score* table. Note that these two interactions do not include the highest propensity interaction in the table. The most likely individual H-bond is identified to be O3 (hydroxy) as donor interacting with O4 (amide) as acceptor.
- In the *Co-ordination scores* table the (a) stands for acceptor and (d) for donor, the number in the first line =0, =1… etc. represents the number of times a particular functional group might donate or accept. If the values are coloured in green the outcome is optimal (e.g. in this observed structure, N1 of acylic_amide (d) donates once and this outcome is optimal - the most likely outcome). The red colour stands for the sub-optimal outcomes, for instance N3 of nitrile (a) accepts once in form II, but based on CSD data this type of atom, in this environment, is more likely to accept zero times.
- You can use the sliders to filter the points in the chart. The default behaviour is to filter out putative H-bonding networks with very low likelihood coordination scenarios (less than 0.1, or 10%, likelihood).

- The most likely putative H-bonding networks should be found in the lower-right corner of the plot and sub-optimal H-bonding will be found in the upper-left corner. Therefore, the chart can be read in diagonal, displaying the most likely networks in the lower-right corner.
- If this was the first polymorph of bicalutamide determined experimentally, we would note that there are other putative H-bonding networks that exhibit both better propensity and better coordination, so there is a reasonable risk of polymorphism here.
- If you click on the lower-right triangle in the chart you can see that this putative network includes the H-bond with the highest propensity score in the table (O3 donor and O4 acceptor) and a second interaction with a lower propensity (N1 donor and N2 acceptor).

- This hydrogen network with the higher propensity is present in Form I. To see where Form I is located in the chart you can load it by clicking **Target structure(s)** drop-down menu in the *Recalculate* section and click on **Select Multiple files…** You can load and compare maximum 10 polymorphs using this functionality.

- In the *Search Structure Section* dialogue click on **Find refcode identifier(s)**, then select *JAYCES02*, hit **OK**, click on **Find refcode identifier(s)** again and select *JAYCES*, then hit **OK**:

- You should now see both refcodes in the right-hand side in the *Search Structure Section* dialogue box. Select **OK** to return to the *Propensity Prediction* wizard.

- You can see in the **Target structure(s)** there are 2 item(s) selected now. Hit **Recalculate.**

- Both polymorphs are now displayed in the propensity chart. The form I structure is represented as green and it has a more likely H-bonding network based on both propensity and coordination. Looking at updated chart, we would now note there is a low risk of finding a further polymorph that is more stable than form I. These results agree with the relative stability calculations based on energy and temperature (D. R. Vega et al., *Int. J. Pharm*., 2007, **328**, 112-118).

- You can see in the co-ordination scores tables that both forms have two functional groups which are interacting in a sub-optimal manner (the numbers are coloured in red). There is no putative H-bonding network displayed on the chart with a better co-ordination outcome though, which indicates that the structure is inherently experiencing some frustration in H-bonding. This imbalance can be a sign that solvates and co-crystals of the target compound are likely, and in fact there are both co-crystals and solvates of bicalutamide in the CSD (e.g. refcodes *KIHZIL* and *FAHFIG*).

Co-ordination scores

(To refresh table: left-click chart point) **Form I**

	Atom (D/A)	= 0	= 1	= 2
1	N1 of acyclic_amide (d)	0.170	0.818	0.011
2	O3 of acyclic_al_oh (d)	0.170	0.802	0.028
3	N2 of nitrile (a)	0.692	0.301	0.007
4	O1 of ar_sulfonyl (a)	0.843	0.157	0.000
5	O2 of ar_sulfonyl (a)	0.734	0.262	0.005
6	O3 of acyclic_al_oh (a)	0.967	0.030	0.003
7	O4 of acyclic_amide (a)	0.542	0.452	0.006

Co-ordination scores

(To refresh table: left-click chart point) Form II

	Atom (D/A)	= 0	= 1	= 2
1	N1 of acyclic_amide (d)	0.237	0.752	0.011
2	O3 of acyclic_al_oh (d)	0.114	0.850	0.036
3	N2 of nitrile (a)	0.692	0.301	0.007
4	O1 of ar_sulfonyl (a)	0.734	0.262	0.005
5	O2 of ar_sulfonyl (a)	0.794	0.205	0.001
6	O3 of acyclic_al_oh (a)	0.927	0.069	0.005
7	O4 of acyclic_amide (a)	0.658	0.339	0.003

- Note that each point on the chart stands for a specific, putative H-bond network. If a system has two polymorphs with the same H-bond network this will be represented only once on the chart.

This ends the tutorial.

28.8 Tutorial 8: Searching on Hydrogen-Bond Motifs

Objectives

- To search for motifs formed by the carbamoyl functional group and to find the most common and least common motif.

Steps Required

- Start a Motif Search
- Choose the carbamoyl functional group
- Generate all possible motif geometries
- Choose structures to search
- Analyse the structures and motifs returned

The Example

- Gaining control of how molecules pack in the solid state is a very desirable goal, for many reasons. Hydrogen bond motifs are commonly used as a way of guiding the interactions between molecules. It is important to understand what a likely outcome is for functional group interactions between molecules. We are going to look at the interactions involving the carbamoyl moiety to learn how these groups like to fit together.

Menu Commands Required

1. Begin the Motif Search
   - Click on **CSD-Materials** in the top-level menu and select **Search** then **Motifs…** from the resulting menus.
   - In this example, we are going to generate our own motifs, rather than using the pre-defined motifs available in the expandable lists. Select the **Create new motifs** radio button and click on **Next** to start the *Motif Search wizard*.
2. Select the functional groups for the motif
   - There are three basic steps in the process of motif generation: Defining functional groups, Selecting contacts and Generating motif search queries. We are going to perform all three.
   - Here you can define functional groups that will be included in the motif search queries. You can either:
     - Add pre-defined groups from a set list using the **Add…** button.
     - Sketch your own groups with a sketcher using the **Sketch…** button.
     - Load groups from saved *.con* (connser) files using the **Load…** button.
   - The carbamoyl group is a pre-defined functional group so click on **Add…** to choose the group. Select the carbamoyl_2 group from the drop-down list in the *Select substructure* dialogue

- Click on **OK** and then click on **Next** to proceed to the contacts page.

3. Selecting the contacts

- Next we have to define the contacts between the functional groups that will be used to generate the motifs. Click on the **Add…** button. In the resulting dialogue, make a contact between the carbamoyl N and the carbamoyl O by selecting the carbamoyl groups from the *from* and *to* drop-down lists.

- The atoms that are involved in the contact can be chosen by clicking on the atoms in the display window, but since in this case, the atoms chosen by default are correct, click on **OK** to proceed.
- *Note*: In this example we do not have the option of selecting a contact that involves hydrogen atoms. This is simply because hydrogens were not explicitly drawn when the carbamoyl group was defined; instead the hydrogens are viewed as a constraint on the nitrogen atom.

- Click on the **Next** button to proceed to the final section of the *Motif Search* wizard.

4. Generating motifs

- We are interested in looking for motifs involving rings and infinite chains. Click on the check boxes for the ring motifs that contain 2, 3, 4, 5 and 6 contacts. Also click on the check boxes for infinite chains that repeat every 1, 2, 3, and 4 contacts.

- Click on **Next** to start the algorithm that takes the contacts and motif geometries requested and generates all the possible combinations of contacts that match the criteria. For example, checking the boxes as described above will generate 33 unique motifs.
- When the motifs have been generated, click on **Next**.

5. Structure selection

- We now need to choose the structures that will be searched using this Motif query. We wish to search the entire Cambridge Structural Database, so from the *Databases* expandable list, click on the latest version of the CSD available. Once this is selected a button with a right-hand arrow will become active and, in this case, coloured blue.

- Click on this button to move the database you have selected to the right-hand Selected Structures window then click **Next** to continue.
- Finally, enter a name for the search, for example “carbamoyl” and begin the search by clicking on **Start Search**.

6. Viewing and analysing the results.

- You can allow the search to run through the entire database (which may take a while) or you can stop the search early by hitting the **Pause** button at the bottom of the *Searches* tab within the *Structure Navigator*.

- There are two ways of looking at the results of a search. Firstly, we are going to use the *View results by motif* mode. Make sure **by motif** is selected from the *View results* drop-down list towards the bottom of the *Structure Navigator*.
- *View by motif* mode lists the motifs searched for, with the refcodes that exhibit the motif listed underneath.
- The selected structure, with the motif fragment of the molecule highlighted is shown in the display. The highlighting of the motif can be customised by clicking on the **Options** button at the top of the *Structure Navigator* and by selecting hit **highlighting**. Highlight the motif using the spacefill option.
- To view which motifs are formed most commonly, click on the **% frequency** column heading in the *Structure Navigator*. This will sort the motifs in ascending order of frequency of occurrence, so the most common motifs are at the end of the list. We see that the most common motif is the R2 dimer motif, followed closely by the C1 infinite chain motif.

- Both of these motifs have a reasonably high propensity to form, indicating that these motifs are potentially useful tools for engineering crystal structures.
- Some of the least common motifs in this case are rings with 5 contacts. Explore the patterns occurring in these structures a bit more by clicking on the refcode names listed below the R5 motifs. An example of a structure with an R5 motif is TAVKUX01 - in this case an R5 motif is achieved through more than one molecule in the asymmetric unit (motif R5_A-B-AB-AB-AB-).

- To find the structures that contain both the R2 dimer motif and the C1 chain motif we can filter the search results. Click on the **Options** button towards the top of the *Structure Navigator* and select **Filter Results…**. this will open the *Filter search results* dialogue.
- Drag and drop the R2 dimer motif and the C1 infinite chain motif to the *Must have* box.

- Hit **Apply Filters** and then **Close**. The results of the filtering are presented in the second way of viewing structures, that of *View by Structure* mode.
- In the *View by Structure* mode the refcodes are listed and the motif(s) found within each structure are listed below the refcodes.

- Click on the column heading **# motifs** to sort the list in ascending order. Refcodes that contain two motifs will only contain the R2 motif and the C1 infinite chain motif. Click the **Multi-view** mode check box towards the bottom of the pane. This will allow motifs within a structure to be viewed simultaneously. Find a structure that contains only 2 motifs and expand the list to view the motifs. Click on each motif and you will be able to view how the motifs fit together in the crystal structure.
- If you look at the refcodes that contain more than 2 motifs you will notice that the R6(AB-AB-AB-AB-AB-AB) motif is very common amongst structures that contain both the R2 motif and the C1 infinite chain motif. It is not unusual, if you search for more than one motif, to find that one motif belongs to a subset of another.
- If we filter the results of the search to remove all structures that contain the R2 dimer motif and the C1 infinite chain motif, we are left with a small percentage of structures. Looking through these structures it is interesting to note that the C2(AB-AB-AB-AB) motif is quite prevalent.
- Thus we are beginning to build an understanding of how carbamoyl functional groups like to come together in crystal structures: dimers and infinite chains are very popular and these motifs dominate the packing. The difference between

the C2 infinite chain motif and the C1 infinite chain is that two molecules, instead of one, define the repeat unit. This motif can occur when there is more than one carbamoyl group present on the molecule, or when there is more than one carbamoyl-bearing molecule in the asymmetric unit. However, it is these two simple motifs that dominate the packing: extended ring motifs often result from a combination of these basic units.

This ends the tutorial.

28.9 Tutorial 9: Studying Crystal Packing Features

Objectives

- To use the Crystal Packing Feature search functionality in CSD-Materials to find and compare structures in the CSD with packing contacts that are similar to a specific interaction.

Steps Required

- Read in a structure.
- Find a specific intermolecular interaction.
- Perform a Crystal Packing Feature search.
- Analyse results.
- Overlay and study closely related structures.

The Example

- This tutorial studies a chlorophenyl stacking contact in the structure of 4-chlorobenzoic acid and how this compares to other halophenyl stacking contacts in the CSD. The structure is taken from Colapietro, M. & Domenicano A. (1982), *Acta Cryst.* **B38**, 1953-1957.

Menu Commands Required

1. Load the required crystal structure.
   - Load the first structure for this tutorial by typing CLBZAP02 in the entry box at the top of the Structures pane within the *Structure Navigator*. This will show the structure of 4-chlorobenzoic acid in the visualiser section of the display.
2. Study the packing contacts in the structure.
   - Switch on the tick box next to H-Bond in the list box underneath the display area. This will display the hydrogen bonding interactions formed by the carboxylic acid group of the molecule.

- Clicking with the left mouse button on one of the dashed red lines will show the molecule at the other end of the hydrogen bond - these interactions form a centrosymmetric hydrogen-bonded dimer between two molecules.

- For the purposes of this tutorial we are interested in stacking contacts between phenyl groups, in order to display these interactions, we need to generate a new specific contact.
- To generate a specific contact for the stacking interactions, click on the **Contacts…** button in the *Display Options* pane, this will launch the *Contacts* dialogue box. Click on the **Add** button and choose **Specific…** from the pull-down menu, this will display the *Define Specific Contacts* dialogue. Display contacts between C and C atoms by selecting C from both the drop-down menus next to **Contacts between**, then select the **Actual distance** radio button and set the **Minimum** and **Maximum** values to 0.0 to 3.7 Å respectively.

- Next click on **OK** and then close the *Contacts* dialogue - the visualiser will now show stacking contacts perpendicular to the hydrogen-bonded ring. Expand the structure through these stacking contacts once in each direction and you will see that a 2D tape is formed by these intermolecular interactions.

3. Perform a Crystal Packing Feature search

- Turn off the display of contacts at this point by clicking on the checkboxes for **H-bond** and **Specific Contact**. Now select the six carbon atoms and one chlorine atom of the chlorophenyl group (atoms C1, C2, C3, C4, C5, C6 and Cl1) in each of two molecules related by a stacking contact.

- These 14 atoms will now be used as the search parameters for similar packing features in the CSD.
- Choose **CSD-Materials** from the top-level menu and select **Search** then **Crystal Packing Feature…** from the pull-down menus. This will launch the *Packing Feature Search* wizard.
- The first page of the dialogue will state that you have selected 14 atoms from 2 molecules from CLBZAP02, click **Next** to proceed.

- We will now relax the search criteria slightly to allow the chlorine atoms to be any halogen. To do this, select the two chlorine atoms in atoms pane (to select multiple atoms, left click whilst holding down the *Ctrl* key), click on the **Modify** button and choose **Element > More > Any Halogen** from the resulting drop-down menus. Ensure that the *Number of hydrogens* and *Number of bonds* checkboxes are checked and then click on **Next** to proceed.

- Choose **Very high** similarity by clicking the relevant radio button under *Level of Geometric Similarity Required* and then click **Next**.
- At this point you can choose one or more geometric parameters to record for each of the matched packing features that are found. Select the two chlorine atoms in the 3D view and click **Add Distance >** to choose this parameter, then click **Next** to continue.
- Here the structures you wish to search for the defined packing feature must be selected. Add the full Cambridge Structural Database to the *Selected Structures* section by choosing the CSD from the list of available structures and clicking on the blue right-hand arrow button.
- Click **Next** and then specify the following database search filters: **3D coordinates determined, Not disordered, No errors, Not polymeric, R-factor <= 0.1, Only organics**.
- Click **Next** followed by **Start Search** to run the packing feature search, hits will now start appearing in the *Searches* tab of the *Structure Navigator*. Once the search has returned 100 or so structures, hit **Pause** in the *Searches* tab – there's no need to run the search to completion.

4. Study list of search results

- Take a look at the list of hits shown in the *Structure Navigator*. There will be over 100 structures found with a similar packing feature in the database. Click on the **RMS** column heading at the top of the results list and the results will be ranked in order of the RMS deviation between the search packing feature and the hit packing feature. You will see that the halogen to halogen distance is also shown in the results list for each match.

- A glance at the structures with low RMS values (i.e. very similar packing features) shows that there are a number of other refcodes starting with CLBZAP, these are the same structure as CLBZAP02, but determined at different experimental conditions. Click on the **Show packing feature** checkbox at the bottom of the *Structure Navigator* pane and the search structure will be shown overlaid with the hit structure.

5. Identify closely related packing features

   - Click on the button at the top of the refcode list to re-order the hits alphabetically by refcode and then select BRBZAP02, which will be listed as having an overlay RMS deviation of 0.06 Å. This crystal structure is of the closely related compound 4-bromobenzoic acid and it is clear to see that the two structures share a very similar halophenyl packing feature. Show the hydrogen bonds in the same way as before and expand the contacts in both structures - this will show that the structures also both contain the centrosymmetric hydrogen-bonded dimer.

6. Overlay crystal structures

   - In order to investigate these structures further, it is necessary to expand out the network of molecules again. To make sure that the structures are easily distinguishable, click on the **Options** button in the *Searches* tab and select **Hit Highlighting…** from the pull-down menu. Change the colour of the packing feature (4-chlorobenzoic acid) to green and then close the *Hit Highlighting Style* dialogue. The 2D tape common to these structures can be built up by expanding the contacts as shown before. Clicking on the tick box next to **Short Contact** in the display options will show that there is also a similar packing feature formed by CH…O contacts between the phenyl and carboxylic acid groups. These

stacking contacts, CH…O interactions and hydrogen bonds form a double-layer of molecules all of which is conserved between the two structures.

- Finally, add a further specific contact in the same way as before, except between atoms of type F, Cl, Br, I and with a distance range specified as 0.0 to 3.9 Å. Expanding these contacts will show that the double layers stack together in different ways between the two structures - the Cl…Cl contacts are approximately linear, whereas the Br…Br interactions have a Br-Br-C angle of around 90 °. The structures therefore contain two entire conserved planes of interacting molecules, but the way that these stack is different. In CLBZAP02 the planes stack in an ABABABAB pattern, whereas in the BRBZAP02 structure the planes follow the stacking pattern ABCDABCD.
- Show the Bravais-Friedel-Donnay-Harker morphology models for the two structures by clicking on **CSD-Materials** in the top-level menu and then selecting **Calculations** and **BFDH Morphology…** from the pull-down lists. Interestingly the double layer orientations identified in the two structures also corresponds to the shape and orientation of the BFDH morphology (both are plates) as calculated by Mercury. BRBZAP02 is specified as having a crystal habit of plate, available under *Structure Information* by clicking on the **More info** button (note that crystal habit information is not actively sought out by CSD editorial staff, it is provided when available in the publication citing the crystal structure). Study of the paper for CLBZAP01 (Miller et al., 1974, *J. Am. Chem. Soc.*, **96**, 6334) shows that these crystals also grow as plates, with (100) being the developed face of the crystal (as suggested by BFDH).

This ends the tutorial.

28.10 Tutorial 10: Searching on Packing Similarity

Objectives

- To examine the packing similarities within a family of polymorphic structures.

Steps Required

- Select structures for a packing similarity search.
- Start the packing similarity search.
- View and analyse the results.

The Example

- Carbamazepine is a compound, which to date, has been found to have five polymorphs. Polymorphism, the ability of a molecule to crystallize in different three-dimensional structures, is an interesting though potentially troublesome phenomenon. Pharmaceutical companies are particularly interested in the polymorphic behaviour of target drug molecules. Inclusion of something other than the most thermodynamically stable polymorph in a tablet, for example, may lead to changes within the tablet over time, as the less stable polymorph converts to the more stable polymorph. Different polymorphs may exhibit different physical properties, such as the dissolution rate, which will affect the drug's efficacy.

Menu Commands Required

1. Select structures for a packing similarity search
   - Click on **CSD-Materials** from the top-level menu then **Search** and **Crystal Packing Similarity…** from the resulting drop-down lists. This will open a dialogue where you can select structures upon which to perform packing similarity searches.

- Click on **Add Refcode**… in the ***Reference Structures*** pane and in the ***Enter Refcode*** dialogue type CBMZPN in the **Refcode** text box. We want to compare all carbamazepine structures with each other to determine which structures belong to which polymorph, so check the box **Enter refcode family**.

- Click **OK**. A warning dialogue will be shown that tells you that no 3D coordinates are available for some of the selected structures (including refcodes CBMZPN and CBMZPN15) and therefore these cannot be included in the search. Click **OK** to continue.
- Now we want to enter structures for comparison. As above, choose the full CBMZPN refcode family.

2. Starting the packing similarity search
   - Click on **Next** to view the options for searching and comparing the structures.
   - There are two pages of options that can be user-defined.

- **Size of molecular cluster to compare**: The crystal packing similarity calculation will compare the central molecule of the structures plus a number of molecules nearby in order to see how similar the packing is. The default size of this molecular cluster is 15 molecules (the central molecule plus 14 others), but this value can be changed in the dialogue box by typing in the required number next to **Size of molecular cluster to compare**.
- You can also choose to filter the results by clicking on **Filter comparisons that do not have all n molecules in common**. If checked, the output will exclude all comparisons of structures for which the full set of *n* molecules do not match (where *n* is the number of molecules chosen for the cluster comparisons).
- The geometric tolerances for defining whether the packing is similar or not can be modified here as well. The default values are 20% for the **Distance tolerance** and 20 degrees for the **Angle tolerance**, but each of these can be edited by typing in the relevant box.

- Hit **Next** to see the other search options

- **Allow molecular differences**: Enable this option to find packing similarity between structures containing similar, but not identical molecules. For example, bromo and chloro analogs.
  Note: This option is computationally demanding and is not suited to large database searches.
- **Ignore hydrogen positions**: This option is checked by default, due to the fact that hydrogens are not always present in structures from the CSD. If unchecked the positions of hydrogen atoms are taken into account when

assessing packing similarity and calculating the root mean square deviation between packing patterns.

- **Ignore bond types**: If checked the packing comparison does not take into account bond types. This may be particularly important if the compound being studied contains a carboxylate group for example, where the two C to O bonds could be described as either double or single, or a delocalised region with potentially confusing bond orders.
- **Allow structure inversion**: Allow structure inversion: When checked, this option allows the algorithm to invert the structure to find the best overlay between clusters - this can be important when studying Sohncke space groups for example (such as $P1$, $P2_1$, $C2$, $P3_1$). This option is checked by default, due to the fact that for many organic structures in particular the absolute structure may not be determined. If the difference between enantiomers is important in the particular analysis being performed, then the user should uncheck this option.
- **Ignore each atom's hydrogen count**: If checked the packing comparison does not take into account the number of hydrogen atoms bonded to each non-hydrogen atom.
- **Ignore each atom's bond count:** If checked the packing comparison does not take into account the number of bonded atoms.
  Note: By ignoring hydrogen counts and the number of bonded atoms then comparisons are no longer restricted to structures containing the same compound.
- **Ignore smallest molecular components:** If checked, the smaller component/s of the crystal structure (as determined by atom count) is/are ignored. This allows comparison of two different solvates for example. By ignoring the solvent molecules, isostructurality between the packing of the major components of structures can be explored. In a three-component system, where the three components are all different, only the biggest molecule is retained for structure comparisons.
  Note: When using this option, it is worth considering increasing the size of the molecular cluster used, for example to 20 molecules, as ignoring the smaller components can potentially lead to slab-like clusters.
- **Show only the highest similarity result:** When crystal packing similarity calculations are performed on systems with Z' > 1, more than one molecular cluster is generated for the reference crystal (as there is not a single, unique cluster for Z'>1 structures). The use of more than one reference cluster means there will be more than one similarity result obtained. If this checkbox is checked, only the comparison results with the highest similarity will be shown. All results can be seen if this box is unchecked.

- The default options are adequate for this search, so click on **Compare** to start the search.

3. Viewing and Analysing the results

   - In the *Structure Navigator*, under the *Searches* tab, the results of the pairwise comparisons are listed.

Searches

crystal_comparison_1 | Options

reference	comparison	molecules in common	RMS	reference group
CBMZPN01	CBMZPN02	15 out of 15	0.011	group2
CBMZPN01	CBMZPN03	2 out of 15	0.288	group2
CBMZPN02	CBMZPN03	2 out of 15	0.291	group2
CBMZPN01	CBMZPN10	15 out of 15	0.008	group2
CBMZPN02	CBMZPN10	15 out of 15	0.012	group2
CBMZPN03	CBMZPN10	2 out of 15	0.287	group1
CBMZPN01	CBMZPN11	2 out of 15	0.306	group2
CBMZPN02	CBMZPN11	2 out of 15	0.309	group2
CBMZPN03	CBMZPN11	15 out of 15	0.154	group1
CBMZPN10	CBMZPN11	2 out of 15	0.305	group2
CBMZPN01	CBMZPN12	2 out of 15	0.127	group2
CBMZPN02	CBMZPN12	2 out of 15	0.13	group2
CBMZPN03	CBMZPN12	2 out of 15	0.238	group1
CBMZPN10	CBMZPN12	2 out of 15	0.127	group2
CBMZPN11	CBMZPN12	2 out of 15	0.256	group1
CBMZPN01	CBMZPN13	2 out of 15	0.274	group2
CBMZPN02	CBMZPN13	2 out of 15	0.278	group2
CBMZPN03	CBMZPN13	15 out of 15	0.2	group1
CBMZPN10	CBMZPN13	2 out of 15	0.274	group2
CBMZPN11	CBMZPN13	15 out of 15	0.104	group1
CBMZPN12	CBMZPN13	2 out of 15	0.253	group3
CBMZPN01	CBMZPN14	15 out of 15	0.028	group2
CBMZPN02	CBMZPN14	15 out of 15	0.024	group2
CBMZPN03	CBMZPN14	2 out of 15	0.292	group1
CBMZPN10	CBMZPN14	15 out of 15	0.022	group2
CBMZPN11	CBMZPN14	2 out of 15	0.311	group1

Show Structure

Reference Comparison Both

Show missed matches

Group by similarity

- The first comparison, that between CBMZPN01 and CBMZPN02, shows that these crystal structures are the same, with 15 out of 15 molecules overlaying with each other almost perfectly. The result of the comparison between CBMZPN01 and CBMZPN03, however, shows that the packing of only two molecules is similar between the structures, those involved in the amide dimer motif. Switch on H-bonds when viewing the CBMZPN01 and CBMZPN03 comparison to see the dimer motif.

- To see which of the structures correspond to the different polymorphic structures, click on the **Group by similarity** check box towards the bottom of the *Structure Navigator*. The structures are now collected into 4 groups, suggesting that there are only 4 polymorphs.

Searches

crystal_comparison_1 | Options

group	spacegroup	density (	a	b	c	alpha	beta	gamma
group1								
CBMZPN13	P-1	1.335	5.1856	20.5758	22.2411	84.19...	87.97...	85.1053
CBMZPN03	R-3	1.235	35.454	35.454	5.253	90	90	120
CBMZPN11	P-1	1.339	5.1705	20.574	22.245	84.124	88.008	85.187
group2								
CBMZPN26	P21/n	1.283	7.68	11.44	13.92	90	91.22	90
CBMZPN20	P21	1.342	7.542	11.155	13.919	90	92.885	90
CBMZPN01	P21/c	1.347	7.529	11.148	15.47	90	116.17	90
CBMZPN02	P21/n	1.344	7.534	11.15	13.917	90	92.94	90
CBMZPN10	P21/n	1.343	7.537	11.156	13.912	90	92.86	90
CBMZPN14	P21/n	1.333	7.55	11.186	13.954	90	92.938	90
CBMZPN17	P21/n	1.38	7.4874	11.0406	13.7754	90	92.939	90
CBMZPN18	P21/n	0	7.4874	11.0406	13.7754	90	92.939	90
CBMZPN19	P21/n	1.38	7.487	11.041	13.775	90	92.9	90
CBMZPN21	P21/n	1.374	7.498	11.058	13.789	90	92.838	90
CBMZPN22	P21/n	1.382	7.4893	11.0323	13.764	90	92.953	90
CBMZPN23	P21/n	1.382	7.4893	11.0323	13.764	90	92.953	90
CBMZPN24	P21/n	1.326	7.578	11.176	13.991	90	93.08	90
CBMZPN25	P21/n	1.327	7.576	11.188	13.967	90	87.03	90
group3								
CBMZPN12	C2/c	1.296	26.609	6.9269	13.957	90	109.7...	90
group4								
CBMZPN16	Pbca	1.326	9.1245	10.4518	24.8224	90	90	90

Show Structure

Reference Comparison Both

Show missed matches

Group by similarity

- This highlights an issue that can occur when comparing solvates, salts, hydrates or co-crystals containing just one component in common - large gaps can occur in the structures leading to slab-like, unrepresentative, clusters. In this case, CBMZPN03 has large voids in the structure due to disordered solvent molecules that were not modelled by the crystallographer, so it could alternatively be classed as a solvate.
- To perform a more accurate analysis, return to the search by clicking on the **Edit Search…** button in the *Post-Search Options* pane and rerun the search. Click **Next** to access the *Select Options* dialogue. This time increase the cluster size to 25, run the analysis and group by similarity again, at which point the structures should now be grouped into 5 sets of structures.
- If we return to the list of pairwise comparisons (uncheck the **Group by similarity** check box) we can see that one of the comparisons, between structures CBMZPN03 and CBMZPN11, shows 24 out of 25 molecules in common. From the list of grouped structures, we can see that these two structures belong to two different space groups, *P*-1 and *R*-3.
- How can it be that structures in *P*-1 and *R*-3 are so similar? To start this analysis, you can click on **Show missed matches** to display the molecule in the cluster of 25 that did not match.
- You can then investigate the broader packing of the two structures using the **Packing/Slicing…** option under the **Calculate** top-level menu.
- Right-click in the display-area background, pick **Colours** from the pull-down menu, and then select **Colour by Symmetry Equivalence**.
- Show the symmetry elements in the structures by choosing **Display** from the top-level menu and then picking **Symmetry Elements…**.

- You can see that the *P*-1 structure has four molecules in the asymmetric unit and these are coloured red, green, blue and yellow. It can be seen that three molecules of the *P*-1 structure, coloured red, yellow and blue overlay with the three molecules related by the three-fold screw axis of the *R*-3 structure.
- Therefore, part of the reason the *P*-1 and *R*-3 structures are so similar is that the symmetry independent molecules of the *P*-1 structure display pseudosymmetry, approximating the 3-fold screw axis.
- As a further exercise, you can look for packing similarities between one structure of the carbamazepine family CMBZPN01 and the rest of the database. You might want to use the option **Ignore smaller molecular components** in the **Select Options** page of the wizard for this. In this search, not only will you find matches to the other polymorphs, but also matches to hydrate and solvate structures. For example, UNEYIV has 5 molecules packed similarly to CBMZPN01 even though UNEYIV is a dimethylsulfoxide solvate of carbamazepine. There are also some structures, for example the co-crystal TAZRAO, where the amide dimer motif has been disrupted. However, TAZRAO and CBMZPN01 do have a stacking interaction between two carbamazepine molecules in common. Can you find other examples where this motif is the common motif?

This ends the tutorial.

28.11 Tutorial 11: Generating High Resolution Figures

Objectives

- To use the functionality in Mercury, and its link to POV-Ray (an included ray-tracing program), to generate high-resolution graphics for publications.

Steps Required

- Read in a structure.
- Generate a particular structural view within the visualiser.
- Export a POV-Ray scene file describing the view.
- Apply user-defined styles.
- Render high-resolution graphics.

The Example

- Images can be output from Mercury in a variety of formats including jpeg, gif, png and tif, but for high resolution, publication-quality figures it can be advantageous to render the image using a ray-tracing program. This tutorial shows how this can be achieved using Mercury's facility to output an instruction file (**.pov*) which fully describes the currently displayed view for image rendering in POV-Ray. The instruction file can be used to render the image immediately in POV-Ray to produce an image equivalent to that shown in Mercury but at any chosen resolution, alternatively edits can be applied to the file to customise the final high-resolution graphic. This tutorial will make use of POV-Ray which is included as part of the Mercury installation, for more information on this functionality see POV-Ray Output of Mercury Images (see POV-Ray Output of Mercury Images).

Menu Commands Required

1. Load the required crystal structures.

- Any crystal structure can be used for this process, but for the purposes of this tutorial we will use the structure of carbamazepine form III. Load the structure by typing *CBMZPN01* in the entry box at the top of the *Structures* pane within the *Structure Navigator* section of the Mercury window (note this structure is included in the free teaching subset of the CSD as well as the full CSD). This will show the structure of carbamazepine form III in the *visualiser* section of the display.

2. Visualise a hydrogen-bonded pair of molecules.

- Switch on the tick box next to *H-Bond* in the list box underneath the *visualiser* area. This will display the hydrogen bonding interactions formed by the carboxamide group of the molecule. Clicking with the left mouse button on one of the dashed red lines will show the molecule at the other end of the hydrogen bond - these interactions form a centrosymmetric hydrogen-bonded dimer between two molecules.

- For the purposes of this tutorial we will be generating a high-resolution image based on this pair of molecules and the hydrogen bonds between them. The principles will apply equally well to other structural views and other types of intermolecular contact.

3. Apply styles to the structural view.

- Any styles that we apply in the Mercury visualiser will be transferred across to the scene description when we save it - this means that the view we render in POV-Ray will initially be equivalent to the one we see in Mercury. Firstly, set the display style for the contacts to wireframe by going to **Display**, **Styles**, **Contacts** through the top-level menu, click on the **Wireframe** radio button and then click

on the **All Contacts** button. Next, change the molecular display style to *Capped Sticks* by clicking on **Display** in the top-level menu and selecting **Styles**, **Stick** from the pull-down menu. Finally change the background colour to white by going to **Display**, **Colours**, **Background Settings**, choose the **Single Colour** radio button, then click on the colour button next to it and choose white from the resulting colour palette options. You should now have a nice clear image of a hydrogen-bonded dimer ready for generation of a publication quality figure.

4. Export scene description for POV-Ray.

   - The easiest way to generate a high-resolution image is to simply go to **File**, **POV-Ray Image…** and click on **Render**. You can also choose here your preferred image dimensions, material properties, background settings and file format as well as generate a preview of the image. For the purposes of this tutorial, however, we are going to output and investigate manual editing of the POV-Ray scene file.

   - To output the current view as a POV-Ray scene file, go to **File**, **Save As** and select *POV-Ray file* (**.pov*) from the drop-down list of file formats. This will generate a file named *CBMZPN01.pov* (the scene file) in your chosen directory as well as two additional files in the same directory (named *CBMZPN01.pov.ini* and *ccdc_macro_overrides.inc*) which will be explained later.

   - To see what the scene file looks like, you can open the file in POV-Ray, or a simple text editor like Notepad. Note that a copy of POV-Ray is included with Mercury and should be located in a directory like this (example shown for Windows):

     C:\Program Files\CCDC\CSD_2020\Mercury\povray-3.7.0.0

- If you view the scene file, you will see that the file consists of a set of macros at the top, followed by all the positions, orientations and colours of the objects in the scene (e.g. atoms and bonds) at the bottom of the file.

5. Render an example image.

- We will now render an example image using the unedited scene output from Mercury. In POV-Ray, go to **Render, Edit Settings/Render** in the top-level menu. This will open the *Render Settings* dialogue window. Under the INI File section, click on the **Default** button (this will select the *QUICKRES.INI* file), select *512x384, AA 0.3* in the **Section** drop-down menu and then click on the **Render** button. This will render a relatively low-resolution image (512x384 pixels) with anti-aliasing applied that should look essentially identical to the view you were looking at in the Mercury visualiser.

6. Using the *include* file.
   - The procedure by which Mercury constructs views in POV-Ray is to define how each type of object is shown in a "macro" (or function) and then each individual atom or bond or other object uses these definitions. We can very simply override these descriptions by defining new "macros" for particular objects. Rather than editing the scene file itself, we will edit an "include" file, which will allow us to re-use the edited definitions for the same structure (if we re-generate the scene file) or for other structures.
   - Open the *include* file in POV-Ray, named *ccdc_macro_overrides.inc*, which was created when you saved the scene file from Mercury. This include file is automatically read by POV-Ray when you make a picture and any definitions placed in the file will take precedence over those in the scene file. The file will generally be empty to start with (except for a short comment at the top) so no overrides are applied to the macros. If there is already a file named *ccdc_macro_overrides.inc* in the directory, however, Mercury will not over-write this, so if we regenerated the scene from Mercury, we would not need to create a new include file.
   - We will apply small modifications to three of the macros and save these in the include file to illustrate how the method works.
7. Applying edits to the include file.
   - We now also need to open in POV-Ray the example file *tutorial11.inc*. To find this file:
     - If you are using a Windows PC, *tutorial11.inc* will be in the folder *examples\tutorial* in the folder where Mercury is installed. For Mercury this is most likely to be:

       C:\Program Files\CCDC\CSD_2020\Mercury\examples\tutorial

     - On Unix, it will be in the directory *examples/mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in */home/cambridge*, the tutorial files will be in:

       /home/cambridge/examples/mercury/tutorial

- On macOS, it will be in the directory *Examples/Mercury/tutorial* in the directory where Mercury is installed. For instance, if you have installed Mercury in *Applications*, the tutorial files will most likely be in:

  Applications/CCDC/CSD_2020/Examples/Mercury/tutorial

- This example file contains modified versions of three macro functions, to apply these you can either copy and paste the functions into your *ccdc_macro_overrides.inc* file or rename the example file as *ccdc_macro_overrides.inc*.
- The first macro is named *ccdc_draw_stippled_line_segment* - this macro defines how "wireframe" style contacts are rendered in POV-Ray. In our example, this function has been modified to show the hydrogen bonds as spheres. The contacts will be shown in the same colour as before, but now they will be represented as four spheres spread out across the contact, rather than a dashed line.
- The bond and atom objects in the Mercury "capped sticks" style are defined using the macros *ccdc_draw_closed_cylinder* and *ccdc_draw_solid_sphere* respectively. These macro functions have been modified in the *tutorial11.inc* file to use different thicknesses for the objects. In this case the atoms and bonds will now be shown at fixed thicknesses of 0.15 instead of the values originally specified by Mercury.
- To see what has been modified in these macros, you can compare the new functions with the unedited ones in the scene file output by Mercury.

8. Render high-resolution image from scene file.
   - Now to see what you have achieved by editing these macros in the "include" file, return to the **.pov* file in POV-Ray and go to **Render**, **Edit Settings/Render** again. This time choose a higher resolution setting (e.g. 1280x1024) and click on the **Render** button. The figure produced now will be a significantly higher resolution image with the hydrogen-bonds shown as spheres and the molecules displayed using thicker capped sticks than before.

9. Further options.
   - The edits to the POV-Ray "include" file described in this tutorial are relatively simple but it is possible to achieve an enormous variety of different effects (*e.g.*

glassy or metallic finishes) or additional objects (*e.g.* backgrounds or floors) using POV-Ray. For further information on editing or writing *.pov* files, please look at the extensive on-line documentation for POV-Ray (http://www.povray.org/documentation/).

- The POV-Ray "include" file can also be used in combination with the regular **File**, **POV-Ray image** dialogue. If you add your edited *ccdc_macro_overrides.inc* file into the chosen *Output Directory*, then this will be used to override macros when you output any high-resolution images to this directory.

This ends the tutorial.

28.12 Tutorial 12: Using Mercury's Statistical Analysis Tools

Objectives

- To analyse the conformation of sulfa drugs.
- To run a geometric search in ConQuest, then transfer the results to Mercury for statistical analysis.

Steps Required

- Set up a ConQuest search for a sulfone substructure, monitoring a defined geometric parameter.
- Transfer the search results to Mercury.
- Plot histograms and interpret the data.

The Example

- The example used in this tutorial is taken from the following publication:
  Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused analysis,
  K. A. Brameld, B. Kuhn, D. C. Reuter, M. Stahl, *J. Chem. Inf. Model.*, **48**, 1-24, 2008.
- Sulfa drugs are commonly used to control high blood pressure (amongst other medicinal applications) thus it is important to consider what sort of conformations these substructures adopt, and most importantly, which is the preferred low energy conformation.
- The first CSD study of arylsulfonamides was carried out in 1986 (R. L. Beddoes *et al.*, *J. Chem. Soc., Perkin Trans.*, **2**, 787-797, 1986) and despite there being only 49 acyclic sulfonamides in the database at that time, the conclusions Beddoes et al drew on the lowest energy conformation at that time have not changed.
- Both the nitrogen lone pair and the aromatic carbon p orbital dissect the O=S=O angle. It is thought that these rotamers are stabilised by lone pair interactions with the sulfur d orbitals:

Menziani et al, *J. Mol. Struct.*, **256**, 217-229, 1992

- In this example, we look at aryl sulfones and examine the effect of ortho substitution on the conformation around the sulfone substituent, i.e.

- It is assumed that the reader has prior knowledge of how to use ConQuest. If this is not the case, then tutorials can be found at the following URL: *http://www.ccdc.cam.ac.uk/support/documentation/#conquest*.

Menu Commands Required

- Open ConQuest.
- Launch the sketcher by clicking on the **Draw** button. Sketch the following substructure:

- where *QA* is either a C or N atom and X is any atom.
- Ensure the sulfone S atom is defined as acyclic by right-clicking on the S atom and picking **Cyclicity** then **Acyclic** from the resultant pull-down menus. You will notice a small superscript *a* is added to the S atom.
- Label the X atom so its atomic number is returned during the ConQuest search. To do this click on **ADD 3D**, click on the **X** atom, hit **Define** adjacent to *Atom Properties* and activate the **Atomic Number** tick box. Click on **OK** to finish the definition.
- Now define the C(X)-C-S-QA torsion by selecting all atoms involved in the torsion, then **Define**, next to *Torsion*:.

- By default, all values of this torsion will be returned, from 0 to 360 degrees. Ensure only absolute values for the torsion are returned (i.e. torsions between 0 and 180 degrees) by clicking on **Options** and picking **Absolute** from the *APPLY FUNCTION* pull-down menu.
- Click **OK** and **Done** to close the windows then hit **Search**.
- In the **Search Setup** window, check the options in the **Filters** tab to only include structures with *R*-factor <= 0.075 and Only Organics, then click on **Start Search**.
- As the search is running you can scroll through the list of hits that ConQuest has returned. In the top right of the 2D diagram you will see the value for our defined torsion and get an idea of what sort of typical values are observed for this angle. However, as the search returns more than 10,000 hits, clearly this is not the most efficient way to analyse the results! Statistical analysis tools are provided within Mercury to facilitate such analyses.

- Once the search has finished transfer the results to Mercury by clicking **Analyse Hitlist**, **Analyse Data** and then, in the resulting window, **Analyse in Mercury**.
- The resultant *Analyse Data* spreadsheet contains all substructure matches for our sulfone substructure, the CSD refcode to which the substructure belongs, the atomic number for the ortho substituent X as well as values for the defined torsion angle.

Identifier	NAME	Query	Fragment	abs(TOR1)	ano(ANO1)
search1|ABADOX|0	ABADOX	1	1	94.5630	1.0000
search1|ABADOY|1	ABADOY	1	1	63.3080	1.0000
search1|ABADUD|2	ABADUD	1	1	88.6560	1.0000
search1|ABADUD|3	ABADUD	1	2	88.9460	1.0000
search1|ABADUD|4	ABADUD	1	3	92.2340	1.0000
search1|ABAFEQ|5	ABAFEQ	1	1	69.5330	1.0000
search1|ABAFIU|6	ABAFIU	1	1	122.0090	1.0000
search1|ABAMIB|7	ABAMIB	1	1	70.0840	1.0000
search1|ABASAY|8	ABASAY	1	1	92.5730	1.0000
search1|ABEBIS|9	ABEBIS	1	1	126.6010	1.0000
search1|ABEBOY|10	ABEBOY	1	1	85.1450	1.0000
search1|ABEDES|11	ABEDES	1	1	133.8660	1.0000
search1|ABEDES|12	ABEDES	1	2	72.9270	1.0000
search1|ABEDIW|13	ABEDIW	1	1	91.0050	1.0000
search1|ABEDOC|14	ABEDOC	1	1	60.0130	1.0000
search1|ABEMOJ|15	ABEMOJ	1	1	76.9120	1.0000
search1|ABEMUP|16	ABEMUP	1	1	94.3710	1.0000
search1|ABIDEW|17	ABIDEW	1	1	84.6260	1.0000
search1|ABUKEN|18	ABUKEN	1	1	82.4000	1.0000
search1|ABURUK|19	ABURUK	1	1	90.5830	1.0000
search1|ABURUK|20	ABURUK	1	2	104.0180	1.0000

- A histogram can be plotted for these data by selecting the grey **TOR1** header and picking **Plots** and then **Histogram**.
- Note the number of bins displayed in your histogram can be modified. To do this click on **Display**, **Configure** then change the value in the *Number of bins* window. Click **OK** to apply the changes and close the window.
- This histogram contains data for both ortho-unsubstituted and ortho-substituted sulfones. By creating selections (i.e. subsets of entries) we can explore data trends. For example, in the case of our hitlist we can use the selection tool to separate the unsubstituted sulfones from the substituted and compare the two. Probably the most important use of the current selection is to control visibility because, whenever a plot or statistics table is calculated, it will be based only on entries that are currently visible. Further details are provided in the *Data Analysis* section.
- Sort the atomic number column by right-clicking on the grey **ANO1** header and selecting **Sort** from the pulldown. Now select only those entries with an atomic number of greater than 1, i.e. substituted sulfones. This is done by left-clicking on the first entry in the spreadsheet whose ANO1 value is greater than 1, and with the *Shift* key depressed, click on the last entry in the spreadsheet whose ANO1 value is greater than 1.
- You will see the substituted entries highlighted (in red) as a subset within the histogram.

- We can view only the data for unsubstituted sulfones to inspect it in greater detail.
- It is important first of all to introduce the concept of locked and unlocked plots in the statistical analysis tool. At any point during a data analysis, a data view is either locked or unlocked. If, and only if, it is unlocked, it will automatically be updated when the visibility of entries is changed in other data views. If a plot is locked, however, it will remain unchanged until you manually update it. Further details are provided in the *Data Analysis* section of the documentation.
- To view only the data for unsubstituted sulfones, unlock the histogram by clicking on the padlock button at the top right of the histogram window. Now go to the histogram menu options and hit **Selection**, then **Hide Selected**. The histogram will update:

- You can view statistics for the histogram by selecting **Statistics**, then **Descriptive Statistics.** From this you will see the histogram mean for the unsubstituted sulfones is 90 degrees.
- We can view entries contributing to a histogram bar by selecting the bar, or else we can view a range by keeping the left mouse button depressed and dragging the cursor over the area of interest. The selected area will be highlighted in red in the data analysis histogram, the refcodes contributing to the selection are listed under the *Data Analysis* section of the *Structure Navigator*, and the selected entries will appear in Mercury's 3D view.

- Pick the histogram bar closest to 90 degrees by clicking on it and scroll through some of the hits in Mercury's 3D view. A typical example of an aryl-sulfonyl is ASEYOM:

- In ASEYOM we can see that the sulfonyl oxygen atoms are located on one side of the aryl ring.
- Returning to the spreadsheet window, view only the substituted sulfones by going to **Selection** and picking **Show only selected**. Because the histogram is unlocked the display will be updated:

- View statistics for the histogram by selecting **Statistics**, then **Descriptive Statistics.** From this you will see the histogram mean for the substituted sulfones is 77 degrees.
- Select a histogram bin around 70-80 degrees and inspect some of the structures that contribute to this bin. A typical structure is PILPIK:

- From this we can see the presence of a substituent at the ortho position on the aromatic ring has forced the sulfonyl group to rotate away from the substituent so that one of the S=O groups is almost co-planar with the aromatic ring.

Conclusions

- Unsubstituted aryl sulfones adopt a conformation whereby the preferred torsion C(aromatic)-C(aromatic)-S(=O, =O)-C/N is 90 degrees.
- When a single substituent is introduced at an ortho position on the aromatic ring, the preferred conformation of the same torsion changes to around 77 degrees.
- Although a difference of 13 degrees does not sound significant, it is worth bearing this in mind e.g. if validating output from a docking package.
    - For small ligands that are not substituted much beyond the sulfone moiety, a difference of +/- 13 degrees may not change the ligand's conformation significantly.
    - However, for longer ligands that have larger substitutions beyond the sulfone moiety, a difference of +/- 13 degrees will introduce a "lever effect" into the ligand and may affect its fit into the protein binding site.

This ends the tutorial.

29 Appendix D: What's New

Mercury 4.3

- A new H-Bond Coordination Quick-View component has been introduced in CSD-Materials enabling fast assessment of the likelihood of the observed hydrogen-bonds in a structure based on the coordination number of each donor and acceptor using CSD statistics.

Mercury 4.2

- Significant improvements have been made to the Polymer Expansion functionality in Mercury making the expansion and exploration of polymeric, or networked, structures much faster and more responsive. There are now further options for more tailored expansion to explore structures the way you want to. It is also easy to prune back structural views to generate the ideal high impact graphics and movies.

Mercury 4.1

- A new Polyhedral display style has been introduced to Mercury, providing the ability to represent each metal coordination centre as a polyhedron, thus generating much more impactful graphics for metal-organic structures. This display style is a very common way of representing such coordination centres, where the corners of the polyhedra are the centres of the coordinating atoms in the ligands. The polyhedra can be customised in various ways including; which elements are considered for the central atom and ligand atoms, whether edges are drawn and the opacity of the polyhedra.

Mercury 4.0

- A new Solvate Analyser component has been introduced in CSD-Materials providing the ability to quickly and easily analyse even the most complex solvate crystal structures.
- The underlying technology behind Mercury and ConQuest has now been completely replaced, including both the database format and now a completely shared C++ search engine. This has resulted in many more accurate and more effective searches, as well as full consistency between CSD software components. An additional benefit to the database work is the direct display in the visualiser of both anistropic displacement parameters (ADPs) and atomic occupanices for over 650,000 CSD entries now.

Mercury 3.10

- New features have been added to Mercury to aid the visualisation, understanding and modification of crystal structure symmetry including:
    - The ability to view along a specific combination of either the unit cell axes (Unit Cell Directions) or the reciprocal cell axes (Miller Directions)
    - Links directly to International Tables Volume A (Space-group symmetry) online from the crystal structure information

    - The ability to reduce symmetry from the current space group to any valid subgroup
- The underlying database of the CSD has been extended, which allows Mercury to provide direct display in the visualiser of the anisotropic displacement parameters (ADPs) for over 220,000 CSD entries
- The Full Interaction Maps application has been expanded to include C-Br bromine and C-I iodine probes, which are particularly useful for analysing halogen bonding interactions
- The Hydrogen-Bond Propensity application has been further improved and streamlined for usability
- The CSD Python API menu has been improved to provide more feedback and information when running Python scripts within Mercury

Mercury 3.9

- A range of enhancements have been made to the core Mercury functionality including:
    - The ability to make use of touch-screens for pinch, pan and zoom
    - Functionality to sketch a molecule and convert it into a 3D model
    - An option to auto-edit structural files by default upon loading into Mercury
    - More flexibility in the selection of database locations in Mercury
    - The ability to perform additional crystallographic transformations in Mercury (such as inversion, switching space group settings or changing origin choice)
- The CSD Python API menu in Mercury has also now been upgraded with the creation of sub-menus as well as even more new scripts
- The CSD-Materials applications have been improved and extended including:
    - Enhancements to the graphical user interface for studying crystal packing similarity and dissimilarity
    - Significant advances in the layout and usability of the Hydrogen Bond Propensity application

Mercury 3.8

- A new CSD-Community menu has been added to Mercury. This includes a range of functionality including:
    - The ability to deposit experimental structures and access data from the CSD online
    - Access to the CSD Educational Collection of structures containing 700+ structures carefully selected to enhance chemistry learning
    - The ability to validate CIFs and ensure experimental structures files are correctly formatted for submission to journals and database using enCIFer
    - Functionality for performing crystal structure reduced cell checks against the CSD before collecting a full data set using CellCheckCSD
- A new CSD-System menu has been added to Mercury. This gathers together existing functionality as well as adding new features including:

- The ability to perform searches using WebCSD and ConQuest as well as analyse results using interactive charting, plotting and statistics
- The ability to analyse molecular conformations using Mogul in order to validate the intramolecular geometries of newly determined crystal structures and identify unusual geometric features
- The ability to analyse intermolecular interaction geometries using IsoStar, for example to validate the geometry of an H-bond or short-range interaction within a crystal structure

- The CSD-Materials menu has been extended to provide direct access to solution of crystal structures from powder diffraction data using DASH.
- The CSD-Discovery menu has been extended to provide direct access to protein-ligand docking using GOLD, analysis of protein-ligand interactions and flexible overlay for multiple ligands.

Mercury 3.7

- A new CSD-Materials menu has been included in Mercury in place of the previous Solid Form menu. This contains all the previous Solid Form options, plus a range of new functionality including:
  - Co-crystal design using the new molecular complementarity tool to help you identify the most likely co-formers to be effective.
  - A new hydrate analyser providing insights into the structure and stability of hydrated materials.
  - New functionality allowing exploration of solid-state conformation based on geometrical statistics from the CSD.
- A new CSD-Discovery menu is also available in Mercury providing access to both full interaction mapping and conformer generation functionality.
- The hydrogen-bond propensity tool within CSD-Materials has been improved through enhancement of the functional group library and coordination models.
  - Updated statistical models capturing the coordination behaviour of functional groups are presented. These models are used in the final stages of the hydrogen bond propensity calculation to generate the landscape chart.
  - The list of functional groups has been expanded to cover a more complete range of chemistry (it now contains 361 definitions).
  - The statistical models are built using a significantly larger training dataset of structures.
  - The combined functional group library and the new coordination models are used by the hydrogen bond propensity tool by default.
- Mercury now also provides the ability to both identify and assign chiral centres (*R* or *S*), when bond type information is available.

Mercury 3.6

- A new menu is available in Mercury which allows you to run built-in or user-created Python scripts utilising the CSD Python API directly from Mercury. This allows the user to construct and share tailored scripts using the whole spectrum of CSD functionality to answer your targeted research questions.

- A collection of features has been added to Mercury to aid scientific communication through simple, intuitive generation of high-impact structural representations.
    - The ability to generate high-resolution graphics directly from Mercury using POV-Ray has been added with a very intuitive interface.
    - A further extension is also included to allow easy generation of high-resolution movies from Mercury using POV-Ray.
    - An interface has also been added to directly produce 3D printable model files from Mercury of any molecule, packing pattern or extended framework.
- The Hydrogen Bond Propensity functionality has been extended to provide the option to browse to a directory containing a user-defined functional group library.
- A new option has been added to Mercury to allow the user to animate the contents of the current display through rotation or oscillation about a chosen viewing axis.

Mercury 3.5

- A new Full Interaction Mapping tool is available as part of CSD-Materials. This tool for visualising CSD interaction data in 3D is built on top of existing CSD interaction knowledge base technology (IsoStar). Full Interaction Mapping allows you to assess your molecule's interaction preferences at the click of a button and in the context of its observed crystal structure.
- A box-clipping display option has now been added which allows the user to restrict the contents of the rendered image based on the size of a 3D box, fixed relative to the centre of the structure. This is particularly useful combination with visualisation of complex structural objects like void surfaces or Full Interaction Maps.
- Hydrogen Bond Propensity Tool enhancements:
    - The 2D chemical diagram in the Hydrogen Bond Propensity tool has been further enhanced to include element colours and better layout.
    - A diagram has been added to illustrate distance & angles tolerances in the contact definition page.
    - The Hydrogen Bond Propensity chart has been improved to handle large number of points better and display of the selected target structure(s).
- A variety of user-requested improvements to Mercury:
    - The layout of dialogues within the main Mercury interface has been modified to better cope with low resolution displays.
    - A range of Data Analysis usability enhancements have been made including optimisation of the function and layout of the calculator tool and improvements in the link with the 3D visualiser.

Mercury 3.3

- A variety of user-requested improvements to Mercury:
    - It is now possible to copy-and-paste from Mercury's More Info 2D Diagram to many packages including Microsoft Office.
    - Mercury will now save user-modified contact definitions and retain these settings between sessions.
    - A range of Data Analysis usability enhancements have been implemented including removal of Apply buttons from dialogues, improving the link

between the spreadsheet and 3D visualiser and providing better control over font sizes in plots.
    - When viewing the results from a Packing Feature search, you can now specify that the atoms in a Packing Feature hit are all in the same molecule (useful for finding co-formers, linking fragments or identifying possible surface inhibitors).
    - Many improvements have been made to the appearance and performance of Mercury's 3D view.
    - Packing Feature and Molecule Overlay bug-fixes have been implemented.
- Hydrogen Bond Propensity Tool enhancements
    - The 2D diagram generation in the Hydrogen Bond Propensity tool has been improved.
    - An option to disable Mercury's auto-editing feature is now provided: this allows users define their own bond typing and protonation states.
    - The number of functional groups in the Hydrogen Bond Propensity tool from has increased from 146 to 179, thus improving the chemical coverage and consistency in group matching for the tool.
    - The speed of both the Fitting Data Generation and Hydrogen Bond Analysis steps has been significantly improved.
    - More robust handling of functional groups resulting in fewer regression failures.
    - Functionality has been added to apply the propensity model to multiple polymorphs at once, including display of multiple structures in the chart.

Mercury 3.1

- Mercury's 3D drawing engine has been completely re-written, for better display quality and performance. Many long-standing user requests have been implemented.
- In response to user requests, a number of improvements have been made to Mercury’s image output, including:
    - The saving of images at multiples of the current window size.
    - Automatic scaling of labels and lines when saving high resolution images.
    - The ability to re-position (atom) labels within the 3D display; new label positions are honoured even when the structure is rotated.
    - The ability to reposition unit cell axis labels.
    - A new output format that enables labels to be edited (PDF).
    - Retained user-defined output image size between sessions.
    - The ability to output images with transparent backgrounds (requires a graphics card that supports transparency).
    - Improved surface transparencies in POV-Ray output.
- Changes to CSD-Materials include:
    - The addition of motif search files to the recent files menu.
    - The ability to calculate and output transformation matrix for packing similarity results.

- Other changes are the following:
  - Mercury's **File**, **Save As** dialogue now remembers the previously used location between sessions.
  - Short contact van der Waals distances can now be edited to two decimal places (instead of the current one).
  - Atoms in the 3D view can be coloured by element or suppression, i.e. to enable easy identification of disordered atoms.
  - Improvements to depth cueing.
  - Faster refresh rate for largest structures.

Please see the *MercuryChanges.txt* file located in `<install dir>/CSD_2020/Mercury/docs` for full details of changes in the latest version of Mercury.