Structural chemistry and structural biology of anti-cancer agents binding to proteins with reference to a model protein and to heparanase

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2014

Simon W M Tanley

School of Chemistry
Faculty of Engineering and Physical Sciences
# Table of Contents

List of Figures ................................................................................................................................. 9
List of Tables ................................................................................................................................. 16
Abstract ......................................................................................................................................... 20
Declaration ...................................................................................................................................... 21
Copyright and ownership .............................................................................................................. 21
The Author ...................................................................................................................................... 22
Acknowledgements ...................................................................................................................... 22
Abbreviations ............................................................................................................................... 23
General Introduction and Aims ....................................................................................................... 25
Part I Literature Survey ................................................................................................................ 26

1. Fundamentals of protein X-ray crystallography ....................................................................... 26
   1.1. Principles of X-ray structure determination ...................................................................... 26
   1.2. Crystal growth and crystallisation techniques ............................................................... 29
   1.3. Crystal symmetry ............................................................................................................. 30
   1.4. Diffraction of a crystal .................................................................................................... 34
   1.5. Phase determination ....................................................................................................... 37
      1.5.1 Isomorphous replacement (IR) ............................................................................... 38
      1.5.2 Anomalous dispersion ............................................................................................. 38
      1.5.3 Molecular replacement (MR) .................................................................................. 38
   1.6. Electron density maps, interpretation and model building .............................................. 39
   1.7. Refinement and Validation ............................................................................................. 40
   1.8. Limitations of X-ray crystallography ............................................................................ 42

2. Chapter 2: Homology modelling and Molecular Docking ...................................................... 43
   2.1. Homology modelling ....................................................................................................... 43
       2.1.1 Introduction ............................................................................................................ 43
2.1.2 Steps in Homology modelling ................................................................. 43

2.1.2.1 Primary sequence alignment ............................................................... 44

2.1.2.2 Secondary structure prediction ............................................................ 44

2.1.2.3 Protein 3D structure prediction ............................................................ 46

2.1.2.3.1 PHYRE-2 ......................................................................................... 46

2.1.3 Conclusions ......................................................................................... 50

2.2. Molecular Docking .............................................................................. 50

2.2.1 Introduction ......................................................................................... 50

2.2.2 AutoDock4 .......................................................................................... 50

2.2.3 Virtual screening .................................................................................. 53

2.2.4 Conclusions ......................................................................................... 55

3. Chapter 3: Cleavage of heparan sulphate proteoglycans (HSPGs) by the enzyme heparanase (HPSE-1) ................................................................. 56

3.1. The structure and function of heparan sulphate proteoglycans (HSPGs) ................................................................. 56

3.2. Proposed structure and function of heparanase-1 (HPSE-1) ......................................................................................... 57

3.3. HPSE-1 gene regulation ......................................................................... 62

3.4. HPSE-1 in disease .................................................................................. 63

3.4.1 HPSE-1 in tumour metastasis ................................................................ 63

3.4.2 HPSE-1 in tumour angiogenesis ............................................................. 64

3.4.3 HPSE-1 in diabetes ................................................................................. 64

3.4.4. HPSE-1 in inflammation ..................................................................... 65

3.5. Inhibitors of HPSE-1 ............................................................................. 65

3.6. Discovery of HPSE-2 ............................................................................. 69

4. Chapter 4: Binding of the anticancer platinum drugs cisplatin and carboplatin to hen egg white lysozyme (HEWL) as a model protein ......................................................... 72

4.1. Structure and function of lysozyme ........................................................ 72

4.2. Platinum containing anticancer drugs ...................................................... 72

4.3. Platinum anti-cancer drugs binding to HEWL ......................................... 75
Part I References ........................................................................................................................................... 77

Books and online references for Chapter 1-4 ................................................................................................ 85

Part II Structural studies of cisplatin and carboplatin binding to a model protein .......................... 86

Introduction to Part II ................................................................................................................................... 86

5. Chapter 5: Structural Studies of the effect that dimethyl sulfoxide (DMSO) has on cisplatin and carboplatin binding to histidine in a protein ............................................................................. 88

Supplementary Materials ............................................................................................................................... 89

6. Chapter 6: Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals ...................... 90

7. Chapter 7: Room-temperature X-ray diffraction studies of cisplatin and carboplatin binding to His-15 of HEWL after prolonged chemical exposure .............................................................................. 91

Supplementary Materials ............................................................................................................................... 92

8. Chapter 8: The crystal structure analysis of the relative binding of cisplatin and carboplatin in a mixture with histidine in a protein studied at 100 and 300 K with repeated X-ray irradiation 100

Supplementary Materials ............................................................................................................................... 101

9. Chapter 9: Experiences with archived raw diffraction images data: Capturing cisplatin after chemical conversion of carboplatin in high salt conditions for a protein crystal ......................... 107

Supplementary Materials ............................................................................................................................... 108

10. Chapter 10: Carboplatin binding to a model protein in non-NaCl conditions to eliminate partial conversion to cisplatin, and the use of different criteria to choose the resolution limit 109

10.0 Abstract .................................................................................................................................................... 110

10.1 Introduction ............................................................................................................................................. 110

10.2 Methods ................................................................................................................................................ 112

10.2.1 Crystallisation conditions .................................................................................................................. 112

10.2.2 X-ray diffraction data collection, protein structure solution and model refinement .......... 112

10.3 Results .................................................................................................................................................. 114

10.3.1 Data processing ................................................................................................................................. 114

10.3.2 Carboplatin binding to His-15 ........................................................................................................... 117

10.4 Discussion ............................................................................................................................................. 119

10.4.1 Carboplatin binding to His-15 in non-NaCl conditions ................................................................. 119
10.4.2 Determining the diffraction resolution of the data

10.5 Conclusions

10.6 Acknowledgements

10.7 Supplementary Materials

10.8 References

11. Chapter 11: Carboplatin binding to Histidine

11.0 Abstract

11.1 Introduction

11.2 Methods

11.2.1 Crystallisation conditions

11.2.1.2 NaBr

11.2.1.2 Conditions without NaCl or NaBr

11.2.1.3 Other non-NaCl or NaBr conditions

11.2.1.4 Elemental analysis of the HEWL lyophilised powder from Sigma

11.2.2 X-ray diffraction data collection, protein structure solution and model refinement

11.3 Results

11.3.1 Carboplatin crystallisation with HEWL in NaBr crystallisation conditions

11.3.1.1 The formation of bromo transplatin in the Nδ binding site

11.3.1.2 Other bromine binding sites?

11.3.2 Carboplatin binding to His-15 in non-NaCl conditions

11.4 Discussion

11.4.1 Carboplatin binding to His15 in NaBr co-crystallisation conditions

11.4.2 Carboplatin binding to His15 in non-NaCl co-crystallisation conditions

11.4.2.1 The different forms of the His 15 cause differing binding modes

11.4.2.2 The differing levels of structural detail seen of the carboplatin

11.4.2.3 Checks on the possible contamination of chloride ions from the HEWL lyophilised powder

11.5 Conclusions
11.6 Acknowledgements ............................................................................................................. 141
11.7 References .......................................................................................................................... 142
11.8 Supplementary Materials .................................................................................................... 144
12 Chapter 12: X-ray crystal structures of cisplatin binding to histidine in a protein at multiple
temperatures reveals their structural dynamics ........................................................................ 146
12.0 Abstract ................................................................................................................................ 147
12.1 Introduction ........................................................................................................................... 147
12.2 Methods ................................................................................................................................ 148
  12.2.1 Crystallisation conditions ................................................................................................ 148
  12.2.2 X-ray data collection strategy, structure solution and refinement ................................. 148
12.3 Results .................................................................................................................................. 150
  12.3.1 Cisplatin binding to His15 of triclinic HEWL ................................................................. 150
  12.3.2 Structural dynamics of the Pt atoms studied at multiple temperatures bound to triclinic
      HEWL .................................................................................................................................. 152
12.4 Discussion .............................................................................................................................. 154
12.5 Conclusions .......................................................................................................................... 155
12.6 Acknowledgements .............................................................................................................. 155
12.7 Supplementary material ....................................................................................................... 156
12.8 References ............................................................................................................................. 157
12.9 Appendix 1 ............................................................................................................................ 158
Overall conclusions to part II ..................................................................................................... 161
Part III Heparanase ..................................................................................................................... 162
Introduction to part III .................................................................................................................. 162
13. Chapter 13: Heparanase three-dimensional structure solving using pre-existing data ...... 163
13.1 Introduction .......................................................................................................................... 164
13.2 Methods ............................................................................................................................... 165
  13.2.1 Heparanase expression and purification .......................................................................... 165
  13.2.2 Crystallisation conditions ............................................................................................... 165
  13.2.3 Data collection strategy ................................................................................................. 166
13.2.4 Three dimensional structure determination of human heparanase-1 ........................................... 167
13.2.4.1 Using 1BG4 as molecular replacement search model ................................................................. 167
13.2.4.2 Using 3VNY PHYRE-2 homology model as the molecular replacement search model .............. 168

13.3 Results ............................................................................................................................................. 168
13.3.1 HPSE-1 expression, purification and crystallisation ................................................................. 168
13.3.2 Data collection .............................................................................................................................. 168
13.3.3 Solving the structure of HPSE-1 using the Pb derivative raw diffraction data and molecular replacement ................................................................. 169
13.3.3.1 1BG4 Molecular replacement ................................................................................................. 170
13.3.3.2 Using the PHYRE-2 HPSE-1 homology model based on 3VNY as the molecular replacement search model ........................................................................................................ 176
13.3.3.3 Solving the structure of human HPsE-1 in space group P1 ....................................................... 179

13.4 Discussion ........................................................................................................................................ 183
13.4.1 Comparison between 1BG4 and 3VNY results ........................................................................... 184
13.4.2 Solving the structure of HPSE-1 in space group P1 ...................................................................... 185
13.4.3 Future work .................................................................................................................................. 185

13.5 Supplementary material .................................................................................................................. 186

13.6 References ...................................................................................................................................... 187

14. Chapter 14: Expression and purification of human Heparanase-1 ...................................................... 190
14.1 Introduction ..................................................................................................................................... 191

14.2 Methods .......................................................................................................................................... 191
14.2.1 Expression of active and pro-forms of HPSE-1 using the EMBacY method in Grenoble ... 191
14.2.1.1 Transfection ................................................................................................................................. 192
14.2.1.2 Sample processing and YFP measurements ................................................................................ 193
14.2.1.3 Virus amplification ..................................................................................................................... 193
14.2.2 Cloning, insect cell expression and purification of HPSE-1 in Manchester .............................. 194
14.2.2.1 Primer design and PCR ............................................................................................................ 195
14.2.2.2 Transformations into competent cells and Mini Preps ............................................................. 196
14.2.2.3 Virus amplification and expression................................................................. 197
14.2.2.4 Purification........................................................................................................ 197
14.3 Results...................................................................................................................... 198
  14.3.1 HPSE-1 purification using native pro HPSE-1..................................................... 198
  14.3.2 Cloning, expression and purification of Pro native form and Pro TEV cleavable forms of HPSE-1…........................................................................................................ 199
    14.3.2.1 Cloning of Pro native and Pro TEV forms of HPSE-1 into GP67 vector........... 199
    14.3.2.2 HPSE-1 expression in insect cells................................................................. 204
  14.3.2.3 Purification of HPSE-1..................................................................................... 205
    14.3.2.3.1 Purification 1.......................................................................................... 205
    14.3.2.3.2 Purification 2.......................................................................................... 206
    14.3.2.3.3 Purification of a 50mls test expression volume........................................... 208
    14.3.2.3.4 Purification of a 500mls expression in Hi5 cells........................................ 209
14.4 Discussion............................................................................................................... 216
  14.4.1 Insect cell expression system............................................................................. 216
  14.4.2 Future work...................................................................................................... 219
14.5 References.............................................................................................................. 220
15. Chapter 15: Molecular Modelling and Virtual screening studies of human heparanase-1 for the identification of Novel small molecule inhibitors............................................. 221
15. 1 Introduction.......................................................................................................... 222
15.2 Methods.................................................................................................................. 223
  15.2.1 Homology model of HPSE-1.............................................................................. 223
  15.2.2 Virtual screening.............................................................................................. 224
  15.2.3 Design of second generation inhibitors.......................................................... 225
15.3 Results.................................................................................................................... 225
  15.3.1 Homology modelling and evaluation of the model.......................................... 225
  15.3.2 Virtual screening using the NCI diversity 2 fragment set and the HPSE-1 model ...... 231
  15.3.3 Design of second generation inhibitors based on the virtual screening results.... 235
15.4 Discussion .......................................................................................................................... 248
15.4.1 HPSE-1 Homology model .......................................................................................... 249
15.4.2 Virtual screening of the HPSE-1 model using the NCI diversity 2 fragment library ....... 250
15.5 Conclusions ...................................................................................................................... 251
15.6 References ...................................................................................................................... 252

Overall conclusions to Part III ............................................................................................ 255
List of Figures

Figure 1.1 Schematic diagram of a CCD detector ..................................................27

Figure 1.2 Flow chart of the steps involved in producing a three dimensional protein structure ..........................................................28

Figure 1.3 Number of protein structures deposited to the PDB ..............................29

Figure 1.4 The hanging and sitting drop crystallisation methods ............................30

Figure 1.5 Phase diagram of the different states of protein crystallisation ...............30

Figure 1.6 The unit cell of a crystal ..................................................................31

Figure 1.7 The 14 Bravais lattices ..................................................................32

Figure 1.8 Examples of miller indices ................................................................33

Figure 1.9 Example of a raw diffraction data image .............................................34

Figure 1.10 Pictorial representation of Bragg’s law ..............................................35

Figure 1.11a Ewald sphere ........................................................................36

Figure 1.11b The diffracted rays that connect the crystal to the reciprocal lattice points are recorded as diffraction spots on the CCD detector ........................................36

Figure 1.12 Ramachandran plot ........................................................................41

Figure 1.13 Example of the output table from Molprobity .................................41

Figure 2.1 Part of an amino acid sequence submitted to the CONCORD webserver ..................45

Figure 2.2 Results of the secondary structure prediction of HPSE-1 using the PHYRE-2 webserver ........................................................................48

Figure 2.3 The results of the hit search out of PHYRE-2 ....................................49

Figure 2.4 Ribbon diagram produced by PHYRE-2 showing the secondary structure motifs ........................................................................49

Figure 2.5 Grid spacing’s used for searching the conformational space of the protein ......51

Figure 2.6 Docking parameters used ................................................................52

Figure 2.7 An example of a ligand in a docked conformation determined in AutoDock ........................................................................52

Figure 2.8 An example of the binding energy for each conformation of the docked ligand ........................................................................53
Figure S8.3 The difference electron density omit maps for the four 300K X-ray data collection runs after rigid body protein model refinement with all cases having their diffraction data cut at 3.5Å resolution…………………………106

Figure S8.4 The difference electron density omit maps for the first three datasets at 300K data collection after rigid body protein model refinement all cut at 2.9Å resolution……………………………………………………………………...106

Figure S9.1 Anomalous signal to noise ratio for each crystal processed with the different software packages………………………………………………………………………………..108

Figure 10.1 DPI value based on the FreeR factor for each refined model at different resolutions for both the XDS and EVAL processed raw diffraction data images……………………………………………………………………………116

Figure 10.2 Binding of carboplatin to the Nδ atom of His-15 from crystal 1…………......118

Figure 10.3 Binding of carboplatin to the Nδ and Nε atoms of His-15 for crystal 2……………………………………………………………………………………………….118

Figure S10.1 Natural log graph of the DPI value based on Rfree against the resolution limit of the refined model for both the XDS and EVAL processed datasets…………………………………………………………………………………123

Figure S10.2 Comparison between XDS and EVAL for the data of crystal 1……………….124

Figure 11.1 Chemical diagrams of cisplatin and carboplatin………………………………128

Figure 11.2 Carboplatin binding to His-15 in NaBr crystallisation conditions……………….134

Figure 11.3 Distances between atoms in the Nε binding site in the NaBr grown crystals…………………………………………………………………………………………..135

Figure 11.4 Binding of carboplatin to the Nδ and Nε binding sites of His-15 in non-NaCl crystallisation conditions…………………………………………………….137

Figure 11.5 The three different form of Histidine………………………………………………..139

Figure S11.1 Unusual 2Fo-Fc electron density as well as anomalous difference electron density nearby the Cys6-127 disulphide bridge……………………………………..145

Figure 12.1 Cisplatin bound to the Nδ and Nε atoms of His-15……………………………….151

Figure 12.2 Mean u square displacement values for the Pt atoms plotted against the data collection temperature……………………………………………………………….153

Figure S12.1 Cisplatin binding to the Nδ and Nε atoms of His-15 processed via XDS……157

Figure 13.1 HPSE-1 native crystal grown in MES buffer pH 6.0……………………………166
Figure 13.2 Diffraction pattern of the HPSE-1 crystal..................................................169

Figure 13.3 The 3D structure of 1BG4 and 3VNY..........................................................169

Figure 13.4 CLUSTALW alignment file between 1BG4 and the 50kDa subunit of HPSE-1....................................................................................................................171

Figure 13.5 Two different views of the crystal packing of the PHASER output PDB file........................................................................................................................171

Figure 13.6 Omit map electron density around one of the alpha helices.........................172

Figure 13.7 Catalytic Glu residues in the 1BG4 model....................................................172

Figure 13.8 The WHHY amino acid sequence with electron density............................173

Figure 13.9 2Fo-Fc electron density map around the best model, with the crystal packing also shown..........................................................175

Figure 13.10 Overall crystal packing of the final model..............................................177

Figure 13.11 Electron density around the WHHY peptide sequence............................178

Figure 13.12 Catalytic Glu residues in the TIM-β-barrel pointing towards the solvent channel.............................................................................................................178

Figure 13.13 Packing diagram in P1 space group, with two molecules in the asymmetric unit..................................................................................................................180

Figure 13.14 Ramachandran plot of the protein residues in the preferred, allowed and disallowed regions..........................................................181

Figure 13.15 Output from Arp/wArp ..............................................................................182

Figure 13.16 Ramachandran plot of the protein residues in the preferred, allowed and disallowed regions..........................................................183

Figure 14.1 An overview of the steps outlined in the method and the EMBacY cells used......................................................................................................................194

Figure 14.2 Full length sequence of pre-pro HPSE-1 and the full length sequence of the pro form (minus the signal peptide).........................................................194

Figure 14.3 SDS-PAGE gel and western blot analysis...................................................199

Figure 14.4 PCR products ran on a 1% agarose gel......................................................200

Figure 14.5 Ligated PCR products run on a 1% agarose gel..........................................200

Figure 14.6 Sequencing results from the placenta 4 colony.............................................201
Figure 14.7 Sequencing results for the TEV form ........................................................................202
Figure 14.8 TEV PCR products after using the new higher grade primers .........................203
Figure 14.9 After ligation of pro-TEV form DNA into GP67 vector and transformation into cells .................................................................................................................203
Figure 14.10 Sequencing results from clone 3 of the pro-TEV form ........................................204
Figure 14.11 High quality sequence confirmation at the C-terminus ........................................203
Figure 14.12: Microscope pictures of the cells every 24hrs in a 72hr period ..............................205
Figure 14.13 SDS-PAGE gel of the samples from the first expression study ..........................206
Figure 14.14 Western blot after 10 minute exposure .................................................................206
Figure 14.15 Western blot analysis of both the neat and heparin binding samples at the 5 different concentrations of virus .................................................................207
Figure 14.16 Peptide ion mass spectrometry analysis ..............................................................208
Figure 14.17 Western blot analysis of the 150, 300 and 500mM wash steps ............................208
Figure 14.18 Western blot analysis using increasing concentrations of NaCl washes .............209
Figure 14.19a SDS-PAGE gel from the heparin affinity columns ...........................................210
Figure 14.19b SDS-PAGE gel analysis of the remainder of the 500mM elution fractions from the heparin affinity column ..............................................................211
Figure 14.19c SDS-PAGE gel analysis of the remainder of the 1M elution fractions from the heparin affinity column ..............................................................211
Figure 14.20 Gel filtration with light scattering chromatogram ..............................................212
Figure 14.21 SDS-PAGE gel of the 6 peak fractions from the gel filtration chromatogram ........213
Figure 14.22 SDS-PAGE and western blot analysis of fractions 29-37 from gel filtration column .................................................................................................................................213
Figure 14.23 Gel filtration chromatogram ...............................................................................215
Figure 14.24 Gel filtration chromatogram of fractions 25-26 ..................................................215
Figure 14.25 SDS-PAGE gel of the gel filtration fractions from Figure 14.25 .......................216
Figure 15.1 Amino acid sequence of the active form of human heparanase .........................224
Figure 15.2 Homology model of the active from of heparanase from the PHYRE-2 webserver………………………………………………………………..225

Figure 15.3 Secondary structure prediction and disorder confidence of the active heparanase homology model from the PHYRE-2 webserver……………….227

Figure 15.4 Verify 3D results………………………………………………………………..227

Figure 15.5 Ramachandran plot of the heparanase homology model……………………..228

Figure 15.6 ProSA results…………………………………………………………………..228

Figure 15.7 Chemical structures of the known small molecule inhibitors of heparanase.….229

Figure 15.8 Binding of the known small molecule inhibitors of heparanase into the active site of the heparanase homology model……………………………..230

Figure 15.9 The chemical structures of the six potentially novel heparanase scaffolds……233

Figure 15.10 Inhibitor binding into the active site of the 58kDa homology model of heparanase……………………………………………………………………234

Figure 15.11 Second generation inhibitors of the ZINC 13597348 template………………235

**List of published figures**

**Chapter 5**

Figure 1 Chemical structures of cisplatin and carboplatin………………………………….602

Figure 2 Fo-Fc and anomalous difference density maps around His15 for the cases of cisplatin and carboplatin with DMSO……………………………………..606

Figure 3 Overlay of His15 from 2i6z (purple) and our cisplatin_DMSO_Paratone structure (orange) at pH 4.7……………………………………..607

Figure 4 DMSO binding in active site C of HEWL…………………………………………607

Figure 5 The active site of HEWL, showing the catalytic residues Asp52 and Glu35………608

Figure 6 Monosaccharide NAG binding in site D of the active site of HEWL…………….608

Figure 7 Cisplatin binding to the N_ and N” atoms of His15 for HEWL with NAG, cisplatin and DMSO……………………………………………………………609

Figure 8 One molecule of DMSO modelled into active site D of HEWL…………………..609

Figure 9 Removal of the the N-hydrogen from histidine……………………………………610

Figure 10 Tautomeric forms of deprotonated histidine……………………………………610
List of Tables

Table 1.1 The 7 crystal systems with their unit cell restrictions and the essential symmetry of the crystal………………………………………………………………31

Table S5.1 A representative selection of HEWL structures from the PDB, at different pH values, with the overall mean and standard deviation of the B factors calculated as shown…………………………………………………………89

Table S7.1 Data collection strategy for each crystal…………………………………………98

Table S7.2 Significant RMS differences (at 3σ level) between RT and 100K structures for each pair wise comparison with a standard deviation value for each calculated using equations 1 and 2…………………………………99

Table S8.1 Unit cell parameter evaluation of the four 300K datasets, with the standard deviations against the unit cell parameters used given also………………101

Table S8.2 The Pt-N distances in Å are given along with the precisions of these bonds based on the Cruickshank DPI coordinates……………………………..102

Table S8.3 Total absorbed X-ray radiation dose……………………………………………102

Table 10.1 X-ray crystallographic data and final protein model refinement statistics for crystals 1 and 2, processed by 3 different software packages………..113

Table 10.2 Pairwise refinements showing R/Rfree of initial models refined at given resolution (top row), against data as given in left column…………………..115

Table 10.3 The pairwise comparison technique applied to the XDS and EVAL data and final refined models…………………………………………………….117

Table 10.4 Anomalous difference density peak heights for the Pt position at both the Nδ and Ne binding sites for crystals 1 and 2 along with the occupancy values of the Pt atoms, calculated using SHELX……………………………..119

Table 11.1 X-ray crystallographic data and final protein model refinement statistics for all crystals studied…………………………………………………….131

Table 11.2 Anomalous difference electron density peak heights for the Pt positions……………………………………………………………………………….138

Table S11.1 The 48 non-NaCl crystallisation screens used from Hampton research……………………………………………………………………………….143

Table 12.1 X-ray crystallographic data and final protein model refinement statistics for all 3 datasets………………………………………………………….149
Table 12.2 Anomalous difference density and 2Fo-Fc density peak heights for each Pt atoms................................................................................................................152
Table 12.3 Occupancy values of each Pt atom at the Nδ and Nε binding sites of His-15..................................................................................................................152
Table S12.1 Total X-ray absorbed dose for each crystals..................................................156
Table 13.1 Summary of the unit cell parameters, data collection and processing parameters using the c unit cell in space group P2..............................................166
Table 13.2 R and R free values for the HPSE-1 structure model after each refinement step..................................................................................................................174
Table 13.3 Data processing and refinement statistics after using 1BG4 as a search model using the Pb derivative dataset..............................................................175
Table 13.4 Refinement steps, with the R/Rfree values at each stage.................................176
Table 13.5 Data collection, processing and refinement statistics of the PHYRE-2 3VNY MR search model using the Pb derivative dataset........................................179
Table 13.6 Refinement steps, with the R and Rfree values at each step............................180
Table 13.7: Refinement steps, with the R/Rfree values at each step after using ArpWarp and Balbes to solve the model.................................................................182
Table 14.1: Volumes of reagents used for PCR..................................................................195
Table 14.2: Concentration of protein per fraction worked out using a Bradford assay..........................................................214
Table 15.1 Drug-likeliness of the six potentially novel inhibitors from the ZINC database..................................................................................................................232
Table 15.2 ZINC 13597348..............................................................................................236
Table 15.3 ZINC 05180958..............................................................................................238
Table 15.4 ZINC 01573829..............................................................................................241
Table 15.5 ZINC 13143019..............................................................................................243
Table 15.6 ZINC 06576323..............................................................................................245
Table 15.7 ZINC 00350319..............................................................................................247
List of published tables

Chapter 5

Table 1 The co-crystallization conditions used for all crystals……………………………..603

Table 2 Data-collection strategy used for the Bruker PLATINUM135 CCD detector……..603

Table 3 X-ray crystallographic data processed via EVAL15 and refinement statistics for all cisplatin co-crystallizations in aqueous and DMSO media ……604

Table 4 X-ray crystallographic data processed via EVAL15 and refinement statistics for all carboplatin co-crystallizations in aqueous and DMSO media …604

Table 5 X-ray crystallographic data processed via EVAL15 and refinement statistics for HEWL co-crystallized with NAG and HEWL co-crystallized with NAG, cisplatin and 7.5% DMSO using silicone oil as cryoprotectant…………605

Table 6 Occupancies of cisplatin and carboplatin binding to His15 calculated from SHELXTL …………………………………………………………………………………………………………………………………........606

Table 7 The platinum-to-histidine imidazole N distances (Å° )……………………………606

Chapter 6

Table 1 Data collection strategies……………………………………………………………………….109

Table 2 Statistics of the processed data from the Rigaku diffractometer…………………...110

Table 3 Statistics of the processed data from the Bruker diffractometer…………………...111

Table 4 Average errors in reflection positions for EVAL…………………………………….113

Chapter 7

Table 1 X-ray crystallographic data details and protein–ligand model-refinement statistics for all crystals analysed…………………………………………………………………………………………………………………...1301

Table 2 Occupancies of cisplatin and carboplatin binding to the His15 refined using SHELXTL and the Pt—N distances (Å° ) for each binding site……1302

Chapter 8

Table 1 Crystallographic and refinement parameters of all data sets collected using the Bruker internal software program…………………………122

Table 2 The Pt-occupancy values (%) calculated from SHELXTL…………………………123

Chapter 9
Table 1 Anomalous difference density peak heights in the Nδ and Nε binding sites for Cl atoms……………………………………………………………………………881

Table 2 Correlation coefficient of cisplatin or carboplatin binding in the 2Fo – Fc map to His-15 using OVERLAPMAP in CCP4i………………………………882

Table 3 Refined occupancy values obtained using SHELX………………………………883
Abstract

The binding of cisplatin and carboplatin to hen egg white lysozyme, a model protein, has been studied using X-ray crystallography under many different crystallisation conditions (Tanley et al, 2012a; 2012b; 2013a; 2013b and Helliwell & Tanley, 2013). From this work, many new results have been obtained; (1) Two molecules of cisplatin and carboplatin are bound to HEWL in DMSO media using the co-crystallisation method. (2) Two molecules of cisplatin are bound to HEWL in aqueous media after a prolonged chemical exposure of 13 months. (3) Cisplatin is stable up to 1.78MGy of X-ray radiation when bound to a protein. (4) We are the first academic group to make our raw diffraction data images freely available. (5) Carboplatin partially converts to cisplatin due to the high Cl concentration used in the co-crystallisation conditions. (6) We have also seen binding of just one molecule of carboplatin to the His-15 residue as a function of chemistry, pH and elapsed time. (7) Using NaBr crystallisation conditions, we see partial conversion of carboplatin to the trans Br form with a portion of the CBDC moiety still present. (8) Using triclinic HEWL co-crystallisation with cisplatin studied at 3 different data collection temperatures showed a more versatile binding and an overall larger summed occupancy at the Nε binding site.

The human protein, heparanase is over-expressed in many different cancers and its experimental three-dimensional structure is yet to be elucidated. The work presented here includes; (1) Production of a homology model of heparanase and completion of virtual screening in order to identify potential novel small molecule inhibitors. (2) Use of this homology model as a molecular replacement search model (crystals of heparanase and their diffraction data was obtained before my PhD). Due to disorder in the crystal and resulting diffraction pattern as well as heavy metal soaks failing, the crystal structure has been difficult to obtain with this particular crystal system and molecular replacement procedure. (3) Over-expression of heparanase in insect cells and purification has been undertaken in order to produce new crystals.
Declaration

No portion of the work referred to in this thesis has been submitted in support of application for another degree or qualification of this or any other University or other institute of learning.

Copyright and ownership

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, the University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in the University’s policy on the presentation of Theses.
The Author

The author graduated from the University of Manchester in July 2006 with a 2.1 Bachelor of Sciences (Hons) in Biochemistry with industrial experience. Since then, he has studied for a PhD funded by the EPSRC under the supervision of Prof John Helliwell of the University of Manchester, with links to Prof Naomi Chayen at Imperial College London and with Dr Edward McKenzie of the University of Manchester. This project has led to work at the EMBL in Grenoble in Dr Imre Berger’s lab for a total period of four weeks, as well as at the OPPF at the research complex at Harwell with Dr Joanne Nettleship for a week and a half. The author has attended many conferences and workshops, mainly in the UK but also in Italy and Finland and has given two oral presentations as well as 3 posters.

Acknowledgements

I would like to thank Prof John Helliwell for giving me the opportunity to work in his lab on this project and for his help and guidance throughout my studies, as well as teaching me all things crystallography. Also, to thank him for pushing me to publish a lot of my work and for letting me expand my own ideas; these are key skills and are very much appreciated.

I would like to thank Dr James Raftery and Dr Madeline Helliwell for their help with SHELX and general discussions.

I would like to thank Dr Edward McKenzie for providing expertise on the molecular biology side of the experiments.

I would like to thank Dr’s Colin Levy, Patrick Bryant, James Raftery and Chris Muryn for their help and expertise using the home source X-ray diffractometers at the MIB and in the School of Chemistry.

I am very grateful to the EPSRC for funding my studies, to the MBCN and PCUBE for funding my travel and expenses to Grenoble and Oxford.

Lastly, I would like to thank my family for their continuous love and support and would like to thank Jen Gethin for being there for me.
Abbreviations

- Ala  Alanine
- Asn  Asparagine
- BEVS  Baculovirus expression vector system
- BLAST  Basic local alignment search tool
- BM  Basement Membrane
- CBDC  Cyclobutane dicarboxylate
- CCD  Charged coupled device
- Cl  Chloride
- CTR1P  Copper Transporter
- Cys  Cysteine
- DMSO  Dimethyl Sulfoxide
- Dpa  Day after proliferation arrest
- ECM  Extracellular Matrix
- EM  Electron microscopy
- FDA  Food Drug Administration
- Gal  Galactose
- Gent  Gentamycin
- GH  Glycosyl hydrolase
- Glc  Glucuronic Acid
- GlcNac  N-acetylglucosamine
- Gln  Glutamine
- Glu  Glutamate
- HEWL  Hen Egg White Lysozyme
- His  Histidine
- HM  Homology model
- HMG  High mobility group
- HMM  Hidden Markov model
- HS  Heparan Sulfate
- HSPE-1  Heparanase-1
- HSPE-2  Heparanase-2
- HSPG  Heparan Sulfate Proteoglycans
- IdoA  Iduronic Acid
- ICP-OES  Inductively coupled optical emission spectroscopy
- Ile  Isoleucine
- IR  Isomorphous Replacement
- IUCr  International Union of Crystallography
- Kan  Kanamycin
- Ki  Estimated inhibition constants
- MAD: Multiple wavelength anomalous dispersion
- Met: Methionine
- MMPs: Matrix Metalloproteinases
- MR: Molecular replacement
- MS: Mass spectrometry
- N: Nitrogen
- NA: Un-sulphated clusters
- NaCl: Sodium Chloride
- NAG: N-acetyl-D-glucosamine
- NAM: N-acetylmuramic acid
- NMR: Nuclear magnetic resonance
- NS: Highly sulphated clusters
- ON: Overnight
- OPPF: Oxford Protein Production Facility
- PBS: Phosphate buffered saline
- PCR: Polymerase chain reaction
- PDB: Protein Data Bank
- PG: Proteoglycan
- Phe: Phenylalanine
- Pt: Platinum
- RT: Room Temperature
- RMSD: Root mean square deviation
- SDS-PAGE: Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
- Ser: Serine
- SLS: Swiss Light Source
- SN: Soluble part
- SNP: Cell extract
- SP: Signal Peptidase
- SR: Synchrotron radiation
- Tet: Tetamycin
- TIM-barrel: Triosephosphate isomerase-barrel
- Trp: Tryptophan
- Tyr: Tyrosine
- UFS: Urofacial syndrome
- XFEL: X-ray free electron lasers
- XRD: X-ray diffraction
- Xyl: Xylose
- YFP: Yellow fluorescent protein
- 3D: Three dimensional
General Introduction and Aims

The aims of this PhD were to over-express, purify and solve the three-dimensional crystal structure of the sugar binding protein heparanase via X-ray crystallography. Once the three-dimensional structure had been determined the next part was to design active site specific inhibitors for heparanase. This work was carried out throughout the PhD.

Due to difficulties moving the heparanase project to a finalised crystal structure determination led to researching the structural chemistry of the anti-cancer agent’s cisplatin and carboplatin and their binding to sugar binding proteins. This was to see if these drugs bound to the sugar binding site. If these did bind, they could be used in heavy atom soaks or co-crystallisation experiments with sugar binding proteins whose three-dimensional structure is unknown and the anomalous signal of the Pt atom could be used in phasing of the structure.

This thesis is presented in the alternative format as eleven papers/chapters, with five being published in peer-review journals, one published as an e-print on the arXiv webserver, one submitted for publication, one in preparation and three which are unpublished. This work is split into two topics; (1) the binding of cisplatin and carboplatin to hen egg white lysozyme in many different conditions and the results are presented in chapters 5-12. (2) The expression, purification and attempted structure determination of the human protein heparanase-1, which is over-expressed in cancer. Also, virtual screening techniques were used to attempt to identify novel inhibitors, and these results are presented in chapters 13-15.

The first part will present a literature review giving the background to X-ray crystallography, homology modelling and virtual screening. These are the techniques which I have used in my PhD. Next, the literature review introduces heparanase, an important human protein which is over-expressed in many different cancers and explains why the three-dimensional structure of this protein is of importance and why it was studied here. The final part of the literature review introduces hen egg white lysozyme as a model protein and the anti-cancer drugs, cisplatin and carboplatin. Here, it addresses a previous study which our studies are based on and the results of which are presented in chapters 5-12.
Part I Literature Survey

1. Fundamentals of protein X-ray crystallography

1.1. Principles of X-ray structure determination

Before X-rays were discovered in the late 19th century, crystal structure determination was limited to careful measurements of the angles between well-developed crystal faces and plotting normal vectors to the planes in stereographic projections, thereby establishing the point group of a crystal (Rupp, 2010). At this time, the theory of crystallography was ahead of the practical applications due to limitations in the equipment used. X-rays were discovered in 1895 by Wilhelm Roentgen (Ashworth Underwood, 1946). They are high energy electromagnetic radiation waves which interact with matter through their oscillating electric field and magnetic field vector with wavelengths between 0.1 and 100 Å (Rupp, 2010). When X-rays are within the range of 0.5 - 1.6 Å, they can penetrate sufficiently to study samples of up to a millimetre in size and are comparable to bond lengths in proteins, making them perfect for crystallography (Blow, 2002). William Henry Bragg developed the first X-ray spectrometer that used monochromatic X-rays. Nowadays, using home source diffractometers, the X-rays are generated using rotating anodes, consisting of a glass enclosure containing a rhenium-tungsten filament and a disk of either Copper (1.54 Å) or Molybdenum (0.71 Å) which produce the characteristic wavelength of X-rays (Shmueli, 2007). The tungsten filament is heated via an electric current, producing electrons which are accelerated at 40 kV towards the metal disk, causing ionization of inner core electrons of the metal. An electron in thehigher atomic energy level takes the place of the ejected electron, thus emitting a photon with a specific wavelength depending on the metal target (Shmueli, 2007). By using photographic film and now the charge coupled device (CCD), the scattered X-rays can be measured. The photographic film was coated with silver halide, which turns to black metallic silver after X-rays are diffracted onto it, and needs to be developed to produce the diffraction spots (Shmueli 2007). The CCD is an area detector divided into a large number of pixels that measures X-ray intensities over slices of reciprocal diffraction space (Burns, 1988). The CCD detector contains a phosphor screen behind a beryllium window, with the screen converting X-ray photons into hundreds of optical photons (Figure 1.1). The fibre optic taper transmits the optical photons to the CCD computer chip where they are converted to stored electrons (Burns, 1988). The number of stored electrons is directly
proportional to the intensity of the Bragg reflection at each pixel and the number of electrons in each pixel is measured.

![Figure 1.1: Schematic diagram of a CCD detector (Jeffrey 2006)](image)

Most protein X-ray crystallographic data are now collected on 3rd generation synchrotrons such as Diamond Light Source in the UK. These synchrotron sources push the limits to the size of the crystals needed with micrometre size crystals now being used regularly to solve the structure of proteins. As well as this, the wavelength can be tuned to the absorption edge of heavy atoms in the crystal (e.g. the use of the Bromine K edge (see chapter 11) which deliberately enhances the anomalous scattering of one atom versus another. This is the procedure for overcoming the phase problem explained in more detail in section 1.5.

Max Von Laue was the first to discover the diffraction of X-rays by crystals, for which he won the Nobel prize in 1914. William Lawrence Bragg saw these first Laue diffraction patterns and interpreted X-ray diffraction as reflection on discrete lattice planes, which led to Braggs Law (see section 1.4), which relates the diffraction angle to the lattice spacing. The father and son team shared the Nobel prize for Physics in 1915 for their services in the analysis of crystal structure by means of X-ray. The first crystal structure to be determined was that of sodium chloride (NaCl).

X-ray crystallography provides a direct way of forming images of molecules at the atomic level, giving detailed information about how the atoms are linked, their mechanism for recognizing and binding substrates and any conformational change that might occur during protein-substrate binding (Blow, 2002, Glusker and Trueblood, 1972). X-ray crystallography is used for both small molecules and proteins, with protein crystallography described in more detail below.
The preparation of a protein crystal can be difficult and the analysis time consuming before the three-dimensional (3D) structure can be solved (Figure 1.2). Despite these difficulties, over 90,000 protein 3D structures have been deposited in the protein data bank (PDB) using X-ray crystallography (Berman et al., 2000) (Figure 1.3). X-ray crystallography is not the only method used to solve the 3D structure of proteins, with nuclear magnetic resonance (NMR) (~10,000 deposited structures) and electron microscopy (EM) (~650 deposited structures) also used. However, X-ray crystallography is the most powerful tool used for structural biology, with around 90% of the all protein depositions being solved by X-ray crystallography.

Protein over-expression

↓

Protein purification

↓

Crystallisation

↓

Characterisation of crystal symmetry

↓

Full data collection (diffraction patterns)

↓

Phase determination

↓

Calculation of electron density maps

↓

Interpretation and model building

↓

Refinement

↓

Structure validation

**Figure 1.2** Flow chart of the steps involved in producing a 3D protein structure.
1.2. Crystal growth and crystallisation techniques

Before crystallisation, pure protein is needed and can either be purchased commercially or expressed in a variety of different systems (insect cells, mammalian cells or E. Coli). Protein crystallisation is a long, drawn-out process of trial and error. To minimise this time, at least two protein purification steps are needed to achieve a purity of at least 95% as measured by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) or mass spectrometry (MS). The formation of crystals requires the addition of specific salts or polyethylene glycols as precipitants, otherwise the protein will precipitate out of solution and not crystallise. The buffer used; its pH, the temperature and the addition of co-factors or inhibitors determine whether a protein crystallises (McPherson, 2003).

A variety of techniques are used for the crystallisation of proteins, with the vapour diffusion techniques, hanging (Figure 1.4a) and sitting drops (Figure 1.4b) being the most widely used in macromolecule crystallisation (Chayen & Saridakis, 2008). These techniques rely on the presence of a reservoir that absorbs water from the crystallisation drop (protein plus a small amount of the reservoir solution) and drives the solution into a super-saturated state leading to crystallisation (Figure 1.5) (Rupp, 2010). The other methods used are batch crystallisation
and the hot box technique. In the batch method, a high concentration of the protein is dissolved in a low ionic strength buffer; salt is added to increase the ionic strength of the solution, thus obtaining the nucleation point of the protein, forming crystals. The hot box technique, uses a temperature change (high to low) to reach the super-saturation point of the protein. At high temperatures the protein is more soluble, thus changing to low temperatures forces the protein out of solution, forming the crystals.

1.3. Crystal symmetry

A protein crystal presents molecular matter in a highly ordered 3D structure, where every molecule is precisely positioned in relation to its neighbours (Glusker and Trueblood, 1972), displaying a high degree of symmetry (Blow, 2002). The ideal crystal has lattice symmetry and three lattice translations define the unit cell of a crystal, characterized by three vectors (a, b and c) and the angles between them (α, β and γ) (Figure 1.6), giving rise to the seven crystal systems that describe the geometry of the unit cell (Blow, 2002) (Table 1.1).

Figure 1.4: The hanging (a) and sitting drop (b) methods. Crystallisation of proteins brought about via bringing the protein up to super-saturation through the metastable state. Protein solution is green and reservoir is blue.

Figure 1.5: Phase diagram of the different states of protein crystallisation (Rupp, 2010).
The 7 crystal systems with their unit cell restrictions and the essential symmetry of the crystal. The inverse axes mentioned in the essential symmetry of the crystal do not apply to proteins due to their chirality, these only apply to small molecule crystals (Glusker and Trueblood, 1972).

<table>
<thead>
<tr>
<th>Crystal system</th>
<th>Essential symmetry of crystal</th>
<th>Unit Cell restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>One diad axis (2-fold rotation) or mirror plane (inverse diad axis)</td>
<td>(a \neq b \neq c) (\beta \neq \alpha = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>Three orthogonal diad axes or inverse diad axes</td>
<td>(a \neq b \neq c) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>One tetra axis (4-fold rotation) or inverse tetrad axes</td>
<td>(a = b \neq c) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Trigonal/Rhombohedral</td>
<td>One triad (3-fold rotation) axis or inverse triad axis</td>
<td>(a = b = c) (\alpha = \beta = \gamma \neq 90^\circ)</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>One hexad (5-fold rotation) axis or inverse hexad axis</td>
<td>(a = b \neq c) (\alpha = \beta = 90^\circ, \gamma = 120^\circ)</td>
</tr>
<tr>
<td>Cubic</td>
<td>Four triad axes or inverse triad axes</td>
<td>(a = b = c) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
</tbody>
</table>

Introducing translational symmetry to the seven crystal systems form the 14 Bravais lattices (Figure 1.7) which have 4 different ways of centering the lattice points (Blow, 2002; Glusker and Trueblood, 1972):

1. **Primitive (P):** The lattice points are located at the corners of the unit cell
2. **Body-centred (I):** All primitive points included plus an additional point at the centre of the unit cell
3. Face-centred (F): All primitive points included plus an additional point at the centre of each face of the unit cell

4. Base-centred (C): All primitive points included plus an additional point at the centre of each one of the pair of cell faces

![Figure 1.7: The seven crystal systems with the introduction of translational symmetry to form the 14 Bravais lattices (Barron and Smith, 2010).]

Each crystal possesses point symmetry, which is a symmetry operation that passes through a central point and leaves at least one point unchanged and the appearance of the object unaltered (Glusker and Trueblood, 1972). There are four operations associated with point groups to make 32 possible ways of combining the point group symmetry with the Bravais lattices (Blow, 2002).

1. N-fold rotation axes: Rotation through 360°/n (n can only be 1, 2, 3, 4 or 6) which leaves the object unaltered.

2. Mirror planes: designated m and is a reflection which takes place with respect to a mirror plane, where a left handed molecule is converted to a right handed molecule

3. Inversions: Implies that every point x, y, z becomes \(-x, -y, -z\)

4. Improper rotations: Rotation of 360°/n followed by an inversion
Combining the point symmetry with translations gives rise to two space symmetry operations (Blow, 2002).

1. Screw axis: Combination of a rotation followed by an appropriate translation parallel to the axis of rotation to preserve the translational repetition
2. Glide plane: Combination of a reflection in a mirror plane followed by a translation.

The overall symmetry of the crystalline lattice is defined by combining the Bravais lattices and the point group, with translational and space symmetry, forming the 230 crystallographic space groups (Blow, 2002). Proteins only crystallise in 65 of these 230 space groups as the inverse axes does not apply to proteins due to their chirality.

The Miller indices (hkl) (Figure 1.8) define the crystal planes and directions in the Bravais lattices of the crystal and they correspond to the inverse ratio of the intercepts on the a, b and c axes of the unit cell (Rupp, 2010).

\[
\begin{align*}
\frac{1}{1}, \frac{1}{1}, \frac{1}{\infty} &= (110) \\
\frac{1}{1/2}, \frac{1}{1/2}, \frac{1}{1/2} &= (222) = (111)
\end{align*}
\]

**Figure 1.8:** Examples of Miller indices (Barron and Smith, 2010)
1.4. Diffraction of a crystal

A crystal contains many molecules, minimising the radiation damage caused by the X-ray beam and amplifying the signal compared to using a single protein (Blow, 2002). X-rays are penetrating radiation and can be scattered from electrons throughout the crystal and provide a diffraction pattern (Figure 1.9) which is used to build up a 3D structure of the protein (Blow, 2002). The diffraction pattern is related to the lattice and unit cell geometry of the crystal structure, and the intensities of the spots hold the information about the positions of atoms in the unit cell (Clegg, 1998).

![Figure 1.9](image-url) Example of a raw diffraction data image. The crystal is HEWL co-crystallised with carboplatin in DMSO media with paratone as the cryoprotectant.

For waves to interfere constructively and hence amplify the signal, they must fulfil the Bragg equation (equation 1.1):

\[ n\lambda = 2dsin\theta \]

where \( n \) is an integer, \( \lambda \) is the wavelength of the X-rays, \( d \) is the spacing of the planes and \( \theta \) is the glancing angle at which the X-rays are reflected (Bragg L, 1922). In a crystal with many parallel planes, some of the X-ray beam is scattered by the first plane and some passes...
through to each of the other parallel planes; they will only be in phase if the crystal is correctly oriented in the incident beam and the path length between the planes is a whole number of wavelengths (Blow, 2002). The unit cell information can be obtained from the Bragg equation but only the amplitude is known and the phase has been lost (Blow, 2002).

**Figure 1.10:** Pictorial representation of Bragg’s law. The incident rays are in phase and parallel up to the point where beam A hits B and the second beam (A’) continues to the next plane where it is scattered by B’. The extra distance travelled by the second beam is in phase, only if the distance is a whole number of wavelengths (Henry et al http://serc.carleton.edu/research_education/geochemsheets/BraggsLaw.html)

For the diffraction spots to occur they must fulfil the Bragg equation above and each spot must lie on the Ewald sphere (Ewald, 1969). The Ewald sphere takes into consideration the reciprocal lattice of the crystals. Around the centre of the crystal, there is a sphere of radius \(1/\lambda\) and the origin of the reciprocal lattice lies in the direct beam at the edge of the Ewald sphere (Figure 1.11a). A diffraction spot from the crystal lies on the Ewald sphere and the angle between the direct beam and diffracted beam is \(2\theta\) and a diffraction vector \(S\) connects these points, which relates to the lattice spacing of the crystal (Figure 1.11a). When the crystal is rotated the reciprocal lattice rotates around its origin. So when the reciprocal lattice points cross the Ewald sphere they ‘light up’, i.e. diffract (Figure 1.11b) (Jeffrey, 2006).
Figure 1.11a
Graphical representation of the Ewald sphere (Jeffrey, 2006)

Figure 1.11b The diffracted rays that connect the crystal to the reciprocal lattice points are recorded as diffraction spots on the CCD detector (Jeffrey, 2006)
1.5. Phase determination

The crystallographic phase problem arises because it is not possible to record the phase of a Bragg reflection with respect to the incident beam; only the intensity of each reflection can be measured. The phases are needed to correctly solve the structure of a protein by converting the 2D diffraction pattern images into a 3D model of the electron density within the crystal using Fourier transformations (a mathematical model). The diffraction pattern, consisting of well-defined individual spots (Figure 1.9) of the crystal and its electron density, are related by the Fourier series (Read, 1999). The amplitude of a particular structure factor indicates the extent to which the electron density is concentrated on planes parallel to the Bragg plane, while the phase indicates the position of planes of high electron density relative to the Bragg planes. The structure factor equation (Equation 1.2) describes how the X-rays are diffracted by every atom in the crystal and uses the scattering power of each atom ($f_i$ which is the scattering power for the $j^{th}$ atom) and is dependent on the electron density.

$$F(hkl) = \sum_i f_i \exp(2\pi i (hx_i + ky_i + lz_i))$$  \hspace{1cm} \text{Equation 1.2}

The calculation of the electron density of each atom needs to be performed in three dimensions in order for the 3D structure of the protein to be deduced. The electron density equation at a particular point in the crystal (x,y,z) is shown below (equation 1.3) and the unit cell volume (V) of the crystal is needed to recover the electron density on the correct scale (Read, 1999). An inverse Fourier transform is used to regenerate the electron density from the structure factors (a summation of the Miller indices) (Read, 1999).

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{2\pi i \Phi_{hkl}} e^{-2\pi i (hx + ky + lz)}$$  \hspace{1cm} \text{Equation 1.3}

There are three methods used in protein crystallography to overcome the phase problem: isomorphous replacement (IR) (Green et al., 1954), anomalous dispersion (Dauter & Nagem, 2002; Dauter, 2006) and molecular replacement (MR) (Rhodes, 2000).
1.5.1 Isomorphous replacement (IR)

IR compares a native crystal with one or more derivatives that differ due to the addition of heavy metal atoms bound tightly to sites in the asymmetric unit of the crystal (Green et al., 1954). The unit cell and symmetry of the native and heavy atom derivative must be the same and the heavy atom must not cause a significant change to the protein. The heavy atoms scatter X-rays with higher intensities and these are located by Patterson or direct methods in the anomalous difference density maps and isomorphous difference density maps and compared to the native crystal (Green et al., 1954). These heavy metal sites are used to calculate the phases of the reflections and build up the structure.

1.5.2 Anomalous dispersion

Anomalous dispersion can also be used for phase determination and, like IR, depends on changes to the diffraction pattern caused by heavy atoms (Blow, 2002). A big advantage of anomalous dispersion over IR is that only one crystal with heavy atoms is needed, as the wavelength can be varied around the absorption edge of the anomalous scatterers present in the crystal (Hendrickson, 1991; Dauter and Nagem, 2002). This technique is known as multiple wavelength anomalous dispersion (MAD) and two, three, four or five different wavelengths can be used to collect a dataset from one crystal (Hendrickson, 1991; Peterson et al., 1996). However, the more wavelengths used, the longer the exposure times, leading to increased radiation damage.

1.5.3 Molecular replacement (MR)

MR is also used to solve the phase problem and is used when a related structure has been previously solved. This is in contrast to the two previous methods, which are used when the structure of the protein is completely unknown. MR is increasing in prevalence due to the increased number of proteins being deposited in the PDB (~90,000, Figure 1.3) and the relatively short amount of time these calculations take using computer programs such as PHASER (McCoy et al., 2007) in CCP4i. MR is used when a related structure has been previously solved and is typically more than 40% identical in amino acid sequence to the unknown model (Blow, 2002, Stout and Jensen, 1989). In this case, six variables, three rotational and three translational will approximately describe the transformation from one set of coordinates to the other (Giacovazzo et al, 2002). The transformation is described by a matrix that rotates the coordinates of the known molecule into the new orientation followed
by a translation (Giacovazzo et al., 2002). To do this, the Patterson maps for the known and unknown structure are generated and the Patterson can be calculated from the diffraction data without knowing the phases. The Patterson is the Fourier Transform using the intensities as the amplitudes and setting all phases to zero, which gives an atom-atom vector map of the structure (Giacovazzo et al., 2002). Thus using the two Patterson maps one can find the rotation and translation which matches the Patterson vectors of the known structure to the Patterson vectors of the unknown structure (Giacovazzo et al., 2002). The most widely used method for MR is now the maximum likelihood method, which is implemented in PHASER (McCoy et al., 2007) rather than the Patterson method. The maximum likelihood method finds the best model which is the one that is most consistent with the observations. If the model is changed to make the observations more probable, the likelihood increases, indicating that the model agrees with the data more (McCoy et al., 2007). Depending on the sequence similarity of the structures, the model of the protein will either be complete (100% identical sequence) or partially complete. These partial models will only be complete by modelling in the missing residues based on the remaining electron density.

1.6. Electron density maps, interpretation and model building

Once the phases have been obtained, an electron density map is calculated using Fourier transformations and the electron density equation (Equation 1.3) is used to interpret the atomic positions of each atom in the structure (Blow, 2002). The amount of detail that can be extracted from an electron density map depends on the resolution, the quality of the map, and the quality of the phase angles which deteriorate towards the higher resolutions (Rupp, 2010; Blow, 2002). The electron density maps I have used throughout this thesis include the 2Fo-Fc map, the Fo-Fc and OMIT maps. Also, I have used the anomalous difference electron density map to identify metal atoms in the structures I have refined. The anomalous signals are generated using Friedel’s Law and Bijvoet pairs and the anomalous scattering of the different heavier atoms can be compared and such atoms distinguished from one another.

From the electron density map, the disulphide bridges and the free sulphur atoms of methionine (Met) and cysteine (Cys) residues have the highest anomalous difference electron density so are easier to determine first of all. Also, the bulky aromatic side chains of tryptophan (Trp), histidine (His), tyrosine (Tyr) and phenylalanine (Phe) residues are easier to
determine and are used to commence 3D structure determination (Blow, 2002). A molecular graphics programme, (e.g. COOT (Emsley & Cowtan, 2004)) is required to fit the polypeptide chain into the electron density map.

1.7. Refinement and Validation

Once most of the atoms are in their correct position, rigid body (RB) and restrained refinement (RR) (in CCP4i) are used to refine the bond lengths and bond angles of the protein (Rupp, 2010). For protein structure determination the R and Rfree factors are used.

\[ R = \sum \frac{|F_{\text{obs}} - F_{\text{calc}}|}{\sum F_{\text{obs}}} \]  

**Equation 1.4**

The R factor (Equation 1.4) is a measure of the fit of the calculated diffraction to the observed intensity data (Stout and Jensen, 1989). Protein structures usually have an R factor between 16-26\%, which is classed as a well-defined 3D structure due to the disordered movement of the loop and turn regions which cannot be determined accurately. The Rfree factor is calculated from 5\% of the reflections which are excluded from refinement and this measures the accuracy of the structure without bias (Brunger, 1992).

The Ramachandran plot (Figure 1.12) (Ramachandran et al., 1963) is used to validate the 3D structure of proteins. In the main chain of a protein/polypeptide, the N-C\(\alpha\) and C\(\alpha\)-C bonds are relatively free to rotate and the Ramachandran plot takes into consideration the torsion angles (phi and psi) of the amino acids. The phi torsion angle is the rotation between the main chain C\(\alpha\)-N atoms and the psi torsion angle is the rotation between the main chain C\(\alpha\)-C atoms in each amino acid for a protein. If a high percentage of the amino acid residues lie in the forbidden regions of the Ramachandran plot then the structure is unlikely to be correct and will only be valid once the majority of the residues fit the allowed positions (Ramachandran et al., 1963).

Molprobity (Chen et al., 2010) is a web-server where a coordinate file can be uploaded and the structure is checked for its validity (Figure 1.13). It looks at atom clashes, poor rotamers, the Ramachandran plot, C\(\beta\) deviations greater than 0.25\AA, and residues with bad bonds and angles to give an overall molprobity score. These results then determine how good/bad the structure is and whether any improvements are needed before structure deposition.
Figure 1.12: Ramachandran plot used for validation of the 3D protein structure. Pink is favoured regions, yellow are the allowed regions and the grey represents the disallowed regions. Red squares in the grey area show the outliers. The triangles represent Gly residues.

<table>
<thead>
<tr>
<th>Protein Geometry</th>
<th>All-Atom Contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clashscore, all atoms: 14.66</td>
</tr>
<tr>
<td></td>
<td>Clashscore is the number of serious steric overlaps (&gt; 0.4 Å) per 1000 atoms.</td>
</tr>
<tr>
<td></td>
<td>Poor rotamers 2.86%</td>
</tr>
<tr>
<td></td>
<td>Ramachandran outliers 0.00%</td>
</tr>
<tr>
<td></td>
<td>Ramachandran favored 96.43%</td>
</tr>
<tr>
<td></td>
<td>Cβ deviations &gt;0.25Å 0%</td>
</tr>
<tr>
<td></td>
<td>MolProbity score 2.02</td>
</tr>
<tr>
<td></td>
<td>Residues with bad bonds: 0.00%</td>
</tr>
<tr>
<td></td>
<td>Residues with bad angles: 0.00%</td>
</tr>
</tbody>
</table>

Figure 1.13: Example of the output table from Molprobity. Green indicates a high level of accuracy; yellow indicates some improvements could be made; red indicates some problems which need addressing. 100th percentile represents the optimum structure. 1st percentile represents the least optimum.

The final validated 3D structure and the experimentally derived structure factors are deposited in the PDB (Berman et al., 2000) where it goes through checks similar to that of Molprobity and, if accepted, it is stored with all the required information necessary to repeat the experimental methods and can also be used for MR calculations.
1.8. Limitations of X-ray crystallography

Even though X-ray crystallography is the most powerful tool used for 3D structure determination of proteins, the process does have limitations. These include: Requirement of well-ordered crystals, which are obtained through a time consuming trial and error process; requirement of pure protein to obtain good quality crystals; solving the phase problem, which requires anomalous scatterers and is unlikely to give atomic resolution, making it difficult to identify certain atoms and their geometry. The majority of hydrogen atoms cannot be determined by X-ray crystallography due to hydrogen being a weak X-ray scatterer; sub-atomic resolution is needed, with only a few proteins studied at sub-atomic resolution. Hydrogen atoms make up around 50% of all atoms in the protein, and they are important as they form H-bonds in various configurations: between amino acids in the same protein; between amino acids in adjacent proteins to form multi-meric structures; between ligands, co-factors or drugs; between water molecules to form a water hydrogen bond network and they are also involved in catalytic mechanisms (Blakeley et al., 2009). To gain more detail about the hydrogen atoms in a protein, sub-atomic X-ray resolution is needed, but also neutron crystallography. Neutron crystallography is used as the hydrogen atoms diffract neutrons stronger than X-rays, and thus gives information on where the hydrogen atoms are and the important contacts they make (Blakeley et al., 2004). Identifying the hydrogen atoms in and around the biomolecule gives information about the hydrogen bonding, protonation states, and hydration configuration of the biomolecule (Niimura et al., 2006; Niimura & Bau, 2008), all of which are critical features for understanding how the protein actually functions. Also, historically, X-ray crystallography only gave a picture of the structure in a static motion, however proteins are flexible. This is being overcome by using time-resolved Laue diffraction or by speeding up the method of data collection by using synchrotron radiation and the new X-ray free electron lasers (XFELs), so as to measure multiple crystal structures. This then allows researchers to make cartoons of joined structures as frames in a movie, allowing the movement of flexible regions in a protein to be seen. To overcome these disadvantages, other structural methods like MS, NMR, EM and neutron crystallography amongst others are needed in conjunction with X-ray diffraction (XRD) experiments to get the full picture of how the protein is folded, any conformational changes it goes through upon substrate binding and its function within cells.
2. Chapter 2: Homology modelling and Molecular Docking

2.1. Homology modelling

2.1.1 Introduction

X-ray crystallography has its disadvantages (section 1.8) and is a complicated and time-consuming process, starting from protein expression and resulting in a 3D structure of an unknown protein (Figure 1.2). In contrast, homology modelling is a computational method to solve the 3D structure of an unknown protein when experimental techniques fail. The aim of this method is to build a 3D model for an unknown protein on the basis of sequence similarity to proteins of known structure (Marti-Renom et al., 2000). It uses experimentally determined protein structures (i.e. via XRD or NMR) to predict conformation of other proteins with similar amino acid sequences, as a small change in primary amino acid sequence results in a small but not significant change in 3D structure (Sali and Blundell, 1993; Hubbard and Blundell, 1987). The quality of the model produced does however depend on the sequence identity of the known structures to the unknown target protein. A 70% sequence identity gives a structure which has a 1-2Å (Sali & Blundell, 1993) root mean square deviation (r.m.s.d) between the Cα atoms of the protein backbone, whereas at 25% sequence identity the r.m.s.d between the Cα atoms becomes 2-4Å, giving more error in the final structure.

The overall goal of protein homology modelling is to predict the 3D structure of a protein based on its amino acid sequence with an accuracy that is comparable to the best results achieved experimentally (Venselaar et al, 2007). If this could be achieved, then these in silico generated models could be used instead of experimental structures for structure-based drug design, virtual screening (VS), analysis of protein function, protein-protein interactions, antigenic behaviour or rational design of proteins with increased stability or novel functions (Venselaar et al, 2007).

2.1.2 Steps in Homology modelling

Homology modelling requires a number of steps which are fully automated. It requires a number of different computer programs to advance from sequence alignment to full model determination. The steps of this method, plus the programs used for each step will be outlined below, along with an overview of PHYRE-2 (Kelley & Sternberg, 2009) as this server has
been used in Chapter 15 to produce a homology model (HM) of human heparanase-1 (HPSE-1).

2.1.2.1 Primary sequence alignment

Computer algorithms have been designed to compare the primary amino acid sequence of different proteins. The main web-server used is the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). BLAST uses the amino acid sequence database and the PDB (Berman et al., 2000) to search for all known primary sequences and tertiary structures to match with the input amino acid sequence. BLAST uses a heuristic algorithm (Oehmen & Nieplocha, 2006) which compares homologous sequences by locating short matches (words) between the two sequences. Once all the common words have been found, neighbourhood words are searched for and these results are used to build up a full alignment between the sequences and a score is given for the whole alignment (Oehmen & Nieplocha, 2006). The output given after the alignment procedure is a list of sequences or 3D structures from the PDB. The percentage sequence identity between the input sequence and known sequences from the databases and the percentage of the sequence it covers is also given. For example, the structure could have a 30% sequence identity to the input amino acid sequence, but these identical regions could cover 60% of the protein being modelled.

2.1.2.2 Secondary structure prediction

From the output of BLAST (Oehmen & Nieplocha, 2006), CLUSTALW (Thompson et al., 1994) can be used to align the input sequence with the known sequences, giving an alignment output file. There are many secondary protein structure prediction web-servers available, but only CONCORD (Wei et al., 2011) will be discussed. Accurate secondary structure prediction provides better understanding of protein structures when structural homologies are unknown (Wei et al., 2011). CONCORD uses a mixed integer linear optimization method for secondary structure prediction and combines the strengths of seven different secondary structure predictions methods (SSpro (Pollastri et al., 2002), DSC (King & Sternberg, 1996), PROF (Ouali & King, 2000), PROFphd (Rost et al., 2004), PSIPRED (Jones, 1999), Predator (Frishman & Argos, 1996) and GorIV (Garnier et al., 1996)). Combining the seven different servers into one maximizes the number of correctly predicted amino acids to achieve better prediction accuracy then the individual servers alone. A consensus prediction score based on the confidence scores of amino acid predictions of the seven methods is introduced to
evaluate the likelihood of an amino acid being at one of the secondary structure states (Wei et al., 2011). This gives each amino acid a score of 1-10, with 10 being very confident that this amino acid is part of the secondary structure state. However, a score of 1 gives low confidence that this amino acid is correct and is usually given to amino acids which make up loops or turns, which are less conserved when compared to α-helices and β-sheets (Figure 2.1). **CONCORD** has an average prediction accuracy of 83% when tested on a PDB select of 3000 proteins. The seven individual servers on their own, vary in respect to an average prediction accuracy, with GorIV (Garnier et al., 1996) the lowest at 62%, and SSpro (Pollastri et al., 2002) the highest at 82%. Removing the two lowest scoring prediction methods (GorIV and Predator) from **CONCORD**, results in prediction accuracy of 83% but including all seven methods shows slightly better performance consistently then with just the best five methods (Wei et al., 2011).

![Figure 2.1](image_url): Part of an amino acid sequence submitted to the **CONCORD** webserver to determine the secondary structural states of the protein. H signifies that the amino acid is part of a α-helix, E signifies part of a β-strand and C signifies part of a turn or loop. The confidence score (Conf) of each amino acid at that point is also given.
2.1.2.3 Protein 3D structure prediction

Once the primary sequence of the unknown protein has been aligned with homologous sequences and the secondary structure of the protein has been predicted, the next step is to determine the tertiary fold of the protein and then predict its 3D structure. To do this, a variety of different software programs can be used, but PHYRE-2 (Kelley & Sternberg, 2009) will be described in more detail here.

2.1.2.3.1 PHYRE-2

The target amino acid sequence is input to the PHYRE-2 web-server (Kelley and Sternberg, 2009; Jefferys et al., 2010; Söding, 2005) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=help). The sequence is processed by scanning against 10 million unique protein structures (www.ncbi.nlm.nih.gov/refseq) from the SCOP database (Murzin et al., 1995) to detect evolutionary relationships to other homologs using PSI-BLAST (Altschul et al., 1997). These are converted to a hidden Markov model (HMM) which produces an evolutionary fingerprint of the protein. The same is done for the ~90000 structures in the PDB (Berman et al., 2000) using PSI-BLAST (Altschul et al., 1997) again to gather sequence homologs and a HMM (Söding, 2005; Karplus et al., 1998) for each known structure in the database. The HMM which has been generated for the query sequence is then scanned against the ~90000 known HMMs to detect high confidence similarities, and allows the prediction of the fold of the protein and hence the 3D structure. This process is accurate up to 15% sequence similarity. HMMs are a statistical method used in formal foundation for making probabilistic models of linear sequence problems (Rabiner, 1989; Eddy, 2004b) and are used in gene finding, profile searches, multiple sequence alignments and regulatory site identification (Eddy, 2004b). HMMs for protein sequences uses a family of known sequences to build a profile which includes the position-specific probabilities of variation in amino acids, as well as insertions and deletions (Mimouni et al.). The sequence whose structure is not known is then aligned to these profile HMMs to confirm the degree of homology (Mimouni et al.; Haussler et al., 1993; Eddy, 2004)

The output from PHYRE-2 includes the secondary structure prediction (Figure 2.2) and how disordered the residues in this prediction are. The server use PSIPRED (McGuffin et al., 2000) and SSPRO (Pollastri et al., 2002) for secondary structure prediction. Comparing the
output from PHYRE-2 with those from CONCORD prediction gives a slight difference, with PHYRE-2 showing more detail i.e. smaller \( \alpha \)-helices and \( \beta \)-sheets are predicted compared with CONCORD. This could be due to the fact that PHYRE-2 uses the top 100 structures/sequences to determine the 3D structure, giving it more accuracy. The other output files obtained are a domain report, which gives the PDB codes of the top 20 hits with the greatest sequence alignment, and a hit report (Figure 2.3). The hit report gives the top 100 structures from the PDB giving the 3D model for the structure, the confidence level that the target sequence has the same fold as the known protein, the sequence identity between the unknown and known sequence and the name of the known protein (Figure 2.3). The webservice also says how confident it is that the model is correct, what proportion of the model has been built and gives a final model (Figure 2.4). The top ranking model is submitted to the 3DLigand site (Wass et al., 2010), which uses the binding site of the known proteins to determine which residues are most likely to contribute to ligand binding and/or the active site.

Depending on the size of the protein and how many other sequences have been submitted to the server, the 3D structure can be returned within 1 hour. However, there will be errors in this model as the loop and turn regions are difficult to model based on their flexibility, disorder and the fact that these regions are less well conserved compared to the secondary structures. Overall, the accuracy of protein structure prediction depends critically on sequence similarity between the query and template. If a template has greater than 30% sequence identity to the query, then usually most or all of the alignment will be accurate and the resulting relative positions of structural elements in the model will be reliable (Kelley & Sternberg, 2009). Below this level of sequence identity, confidence matches are routinely made by PHYRE-2. Given a high confidence match (>90% confidence), the overall fold of the model will be almost certainly correct and the central core of the model will tend to be accurate, even at sequence identities less than 20% (Kelley & Sternberg, 2009). However, in such cases, greater deviations from the true structure will be observed in more peripheral regions of the protein and in regions neighbouring sequence insertions or deletions in the alignment (Kelley & Sternberg, 2009). Using HPSE-1 as an example, the highest sequence similarity it has to a known protein is 25%, but the confidence of the model was 94%. Thus, relating it to how well PHYRE-2 predicts the structure of unknown proteins, the TIM barrel fold will be correct, but the other parts of the structure will be less well defined. Therefore,
determining the X-ray crystal structure of HPSE-1 and other unknown proteins via experimental methods is still of major interest, as these peripheral regions which are difficult to model could be very important for the function of some proteins (e.g. flexible loops which move upon substrate binding, or linker peptides). Also, with the crystal structure determined, one can look at inhibitor binding using co-crystallisation or soaking experiments, whereas, from the model one can only determine which residues are likely to bind to the substrate.

Figure 2.2: Results of the secondary structure prediction of HPSE-1 using the PHYRE-2 webserver. For the secondary structure confidence, red is highly confident, whereas blue/green is of low confidence that the amino acid is an α-helix (green ribbon), β-sheet (blue arrow) or a loop or turn region (straight line). The disorder confidence uses the same key, with mostly residues in the turn or loop regions being confident of disorder. There is only one predicted helix which is said to be disordered and this is part of the signal peptide of HPSE-1 (Residue 1-36).
Figure 2.3: The results of the hit search from PHYRE-2. Shown are the top 7 hits giving the PDB codes, the sequence identity of the known protein to HPSE-1 and the confidence with which the server has fitted the 3D fold of the known protein to the HPSE-1 amino acid sequence, generating the model.

<table>
<thead>
<tr>
<th>#</th>
<th>Template</th>
<th>Alignment Coverage</th>
<th>3D Model</th>
<th>Confidence</th>
<th>% I.d.</th>
<th>Template Information</th>
</tr>
</thead>
</table>
| 1 | c2g6d_   | Alignment:         | 100.0    | 15         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: D: PDB Molecular alpha-1,6-galactosidase;  
|   |          |                    |          |            |       | PDBTitle: the structure of a family B4 alpha-galactosidase,  
|   |          |                    |          |            |       | arciL, z from clostridium thermocellum in complex with 1,3- 
|   |          |                    |          |            |       | l-lectin arabinosidase  
| 2 | c1w9g_   | Alignment:         | 100.0    | 14         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: E: PDB Molecular alpha-1,6-galactosidase;  
|   |          |                    |          |            |       | PDBTitle: the structure of a family SL alpha-1,6- 
|   |          |                    |          |            |       | galactosidase in complex with 4-nitrophenyl-
| 3 | c2n6b_   | Alignment:         | 100.0    | 12         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: E: PDB Molecular alpha-1,6-galactosidase;  
|   |          |                    |          |            |       | PDBTitle: structure of an inactive mutant of alpha-galactosidase  
|   |          |                    |          |            |       | fenn2 thermococcus xylanlyticus in complex with a3 
|   |          |                    |          |            |       | pentosechardide  
| 4 | c2w8g_   | Alignment:         | 100.0    | 21         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: A: PDB Molecular cellulase;  
|   |          |                    |          |            |       | PDBTitle: crystal structure of a cellulose, a bifunctional glucanase-2 
|   |          |                    |          |            |       | xylanase protein from a metagenome library  
| 5 | c2w8b_   | Alignment:         | 100.0    | 16         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: A: PDB Molecular celluolase144c;  
|   |          |                    |          |            |       | PDBTitle: structure of a pseudomonas polymyx xenoglucoamylace 
|   |          |                    |          |            |       | from g2 family 44 with xylglucan  
| 6 | c192a_   | Alignment:         | 100.0    | 15         |       | PDB header: hydrolase  
|   |          |                    |          |            |       | Chain: A: PDB Molecular endoglucanase a;  
|   |          |                    |          |            |       | PDBTitle: crystal structure of a glycoside hydrolase family 44 
|   |          |                    |          |            |       | endoglucanase2 produced by clostridium acetobutylicum etc 824  
| 7 | c1w9h2_  | Alignment:         | 99.0     | 17         |       | PDB/TIM beta/alpha-barrel  
|   |          |                    |          |            |       | Superfamily: transglycosidases  
|   |          |                    |          |            |       | Family: beta-glucanases  

Figure 2.4: Ribbon diagram produced by PHYRE-2 showing the secondary structure motifs (α-helices in ribbons and β-sheets as arrows). This model of HPSE-1 clearly shows the TIM barrel motif.
2.1.3 Conclusions

Due to the predicted model having errors, X-ray crystallography is still needed for 3D structure determination. However, homology modelling can be used in conjunction with X-ray crystallography as it is a powerful method to gain an insight into the structure and function of unknown proteins before a crystal structure is obtained. Also, a predicted HM can be used in drug design and lead drug optimisation before the experimental structure can be determined. VS techniques using a library of compounds against a predicted model can lead to insights into which types of inhibitors bind to the model and hence these can be tested experimentally to determine the affinity of these inhibitors (see section 2.2).

2.2. Molecular Docking

2.2.1 Introduction

Molecular docking is a computational technique which has been developed mainly for drug discovery purposes. This technique allows the user to dock either known ligands into the binding site of proteins to gather information on which residues are required for substrate binding, or to screen libraries of compounds to get information on potential scaffolds which can then be used to design inhibitors/drugs through fragment based-drug design. Thus, molecular docking is a powerful method to screen many targets at a time, reducing the amount of work needed as the results can then be used to improve on these inhibitors experimentally rather than starting experimental work from scratch.

2.2.2 AutoDock4

AutoDock is the most widely used molecular docking program, which simulates interactions between small flexible ligands and macromolecules of known structure (Morris et al., 1998). It is preferable to use a protein whose 3D structure has been solved and whose co-ordinates have been deposited in the PDB, but it is possible to use HM’s (See section 2.1.2.3.1). Using a HM as the target macromolecule will inherently add errors into the calculations as the structure has not been experimentally determined and depends on the sequence similarity between the unknown and known structures. Thus, docking results using HM’s should be treated with caution until an experimental 3D structure can be obtained. Nonetheless, a HM is
a great starting point to design first generation inhibitors and search through libraries of compounds before a structure is solved.

AutoDock uses a number of different genetic algorithms to search the conformational space of the protein molecule and couples this with energy assessments (Morris et al., 2010; Cosconati et al., 2011). The energy assessments from the docking of the ligands ranks the conformations as a function of energy so as to determine which are the most likely conformations (Morris et al., 1998). The search through the conformational space can either be done through blind docking, where the whole of the protein is searched (Figure 2.5b), or through reducing the conformational space by focusing on the active site (if known), or a reduced area of the protein (Figure 2.5a), which in turn reduces the amount of computation required. Depending on the docking parameters chosen, a number of different conformations of the docked ligands can be obtained (Figure 2.6), especially if the ligand has many rotatable bonds (i.e. flexible small molecules), as it can have many different orientations in the active site. AutoDock lets the ligand have a maximum of 32 rotatable bonds, but it is advised that more than 10 can cause problems in the accuracy of ligand docking (Cosconati et al., 2011). Thus, ligands with a large number of rotatable bonds need to have increased docking parameters to search for more orientations of the ligand.

**Figure 2.5:** Grid spacings used for searching the conformational space of the protein. (a) Grid box centred on the TIM barrel (active site). (b) Grid box searching the whole conformational space of the protein.
Figure 2.6: Docking parameters used. These can be varied due to amount of computing time available, the size of the grid spacings used and how many rotatable bonds the ligand has. The number of GA runs relates to how many times the algorithm is run. The maximum number of evaluations and generations relate to the maximum number of times the program evaluates each GA run.

Once the docking of the ligand has been completed, analyses of the results are relatively simple. The different conformations of the ligand can be easily searched with all van der Waals interactions, H-bonds, pi-pi interactions and all neighbouring amino acid residues highlighted (Figure 2.7). Thus, one can determine whether the ligand interacts with the active site of the protein, thus could be a potential inhibitor. The analysis also gives the binding energy of the docked ligands (Figure 2.8) and the docking log file gives an estimated inhibition (Ki) constant for each conformation (Morris et al., 1998). The error in prediction of these binding energies is in the range of 2-3 kcal/mol (Cosconati et al., 2011). However, these Ki constants are still a good starting point for designing inhibitors as they can narrow down the results to start designing second generation inhibitors.

Figure 2.7: An example of a docked ligand. The neighbouring amino acid residues have van der Waals interactions (white density) and hydrogen bonds (red density around Ser-145 and Glu-142) highlighted at a radius of 1.00Å
Based on the above, AutoDock is a very powerful method to determine the binding site of known ligands into known or unknown binding sites of proteins. However, a major advantage of using AutoDock and its related programs is for VS (see section 2.2.2.1), where one can search through millions of compounds computationally to predict potential novel inhibitors of the desired protein.

2.2.3 Virtual screening

VS is the use of high-performance computing to analyse large databases of chemical compounds in order to identify possible drug candidates (Walters et al., 1998). VS has advantages over high-throughput screening (HTS), in that VS computes binding activity in silico, to test the drug-ability of new targets, whereas HTS does produce false negative results: also chemical synthesis and testing are expensive and not all assays can be automated. However, VS can be used as a complementary technique to HTS, where VS is used to exclude compounds which are predicted not to bind and are non-drug like compounds. This reduces the number of compounds needed in HTS experiments, decreasing the amount of synthesis and testing, ultimately reducing the costs by saving both time and money (Lazarova, 2008). VS does have some disadvantages: it is sometimes inaccurate and is strongly dependent on target, search method and chemical space sampled. However, high-
throughput VS is a great way of selecting starting materials, which can then be tested experimentally once the library of compounds has been reduced in size (Walters et al., 1998).

There are a variety of large databases available (Cosconati et al., 2011), including: NCBI PubChem (pubchem.ncbi.nlm.nih.gov), eMolecules (www.emolecules.com) and ZINC (Irwin & Shoichet, 2006). Most of these databases include commercially available compounds, thus any hits which are generated can be tested experimentally. Other libraries which are available include nutraceuticals (Wishart et al., 2008), natural products (Dunkel et al., 2006) and metabolome (Wishart et al., 2009) libraries which contain compounds expected to have good biological properties. Also, the FDA-approved drugs (www.epa.gov/nct/dsstox) are available as these compounds have already shown biological activity and acceptable safety and toxicity profiles (Cosconati et al., 2011). Therefore these can be searched to see if known drugs can also be used on other targets (i.e. cross reactivity).

As VS uses thousands of compounds to dock onto the target macromolecule, other computer programmes are needed to prepare all the ligands using the same grid and docking file. Raccoon (http://autodock.scripps.edu/resources/raccoon) converts the ligands into AutoDock format and filters the library based on Lipinski’s rules. Lipinski’s rules relate to the molecular properties of the ligand (0-5 H-bond donors, 0-10 H-bond acceptors, 0-500 molecular weight, 0-999 atoms and 0-32 rotatable bonds). Once the ligands have been generated and filtered and the macromolecule identified, both grid files and docking files are generated. As in Figure 2.5, the conformational space used can either be over the whole protein or just centred on the known binding site. The docking files are similar to those in Figure 2.6, but due to the large number of ligands being searched these should be reduced, so as to again reduce the amount of computation time needed. Once these have been defined in Raccoon, it generates grid files and docking files for all ligands.

The main problem after generating the dockings for a large number of compounds is to analyse the results. The program Fox (http://autodock.scripps.edu/resources/fox) can be used to analyse the results. Fox is a relatively new program and is in the experimental stage. Fox filters the docking results by energy (Figure 2.8), clustering, ligand efficiency and interactions. Thus, if the active site is known, amino acid residues can be input and all ligands rejected if they don’t interact with these residues, decreasing the subset of docked ligands.
2.2.4 Conclusions

AutoDock is an excellent computer program which can be used for docking of known ligands into macromolecule targets, when the experimental 3D structure cannot be solved. HM can be used as the target macromolecule, but this does add in errors depending on the sequence identity used for the creation of the HM. A powerful tool in the AutoDock suite of programs which include both Raccoon and Fox is the possibility for VS, where large numbers of compounds can be screened against a target to decrease the amount of subsequent chemical synthesis needed. VS also gives starting points for the creation of inhibitors to reduce the amount of time and money needed to get to first generation inhibitors.
3. Chapter 3: Cleavage of heparan sulphate proteoglycans (HSPGs) by the enzyme heparanase (HPSE-1)

3.1. The structure and function of heparan sulphate proteoglycans (HSPGs)

Heparan sulphate (HS) is a linear carbohydrate polymer synthesised onto a proteoglycan core (PG) in the Golgi apparatus via the sequential action of a HS synthase enzyme (Kreuger et al., 2006; McKenzie, 2007; Vreys & David, 2007). HS synthase builds up HS chains via the addition of alternating glucuronic acid (GlcA) and N-acetylglucosamine (GlcNAc) units in their UDP bound form onto a core protein linkage tetra-saccharide to form HSPGs (Bernfield et al., 1999; Vreys & David, 2007). The HS chains are either attached to the membrane spanning syndecans or the glycosylphosphatidylinositol (GPI) anchored glypicans (Bernfield et al., 1999) which are highly abundant on the cell surface and extracellular matrix (ECM) (Fux et al., 2009). The tetra-saccharide linkage is composed of a specific serine residue localized next to a glycine and flanked by one or more acidic residues from the PG with a xylose, two galactose and a GlcA residue bound (Bernfield et al., 1999; Kreuger et al., 2006) (Figure 3.1).

![Figure 3.1: Structure of a HS chain and the subsequent post-glycosylation modifications (Fux et al., 2009)](image-url)
Post-glycosylation modifications of GlcA and GlcNAc residues occur at various positions along the HS chain (Vlodavsky et al., 2002; Li & Vlodavsky, 2009) (Figure 3.1). Specific GlcNAc residues are modified to GlcNSO₃ by N-deacetylase/N-sulfotransferase enzymes and these can undergo sulphation at C2, C4, C6 and, in the case of heparin at residue C3, catalysed by specific glucosaminy1-O-sulfonyltransferases (Bernfield et al., 1999). Specific GlcA residues undergo epimerisation modifications to create iduronic acid residues which can be sulphated at position C2 (Kreuger et al., 2006; Vreys & David, 2007). These modification steps yield regions of 2-9 highly sulphated clusters (NS domains) separated by larger un-sulphated clusters (NA domains) with the region between containing a mixture of both domains (Vlodavsky et al., 2002; Kreuger et al., 2006; Vreys & David, 2007; Li & Vlodavsky, 2009). The NS domains provide specific docking sites for a variety of cytokines and growth factors (FGF-1, FGF-2, TGF-β, HGF, IL-2, VEGF and anti-thrombin III) in order to sequester them at the cell surface and ECM (Bernfield et al., 1999; Hulett et al., 2000; Kreuger et al., 2006; Vreys & David, 2007; Fux et al., 2009). HSPGs bind to and assemble the ECM proteins fibronectin, laminin and collagen which play key roles in cell-cell and cell-ECM interactions (Parish et al., 2001; Vlodavsky et al., 2002). The ECM is a heterogeneous mixture of proteins and polysaccharides which fills the extracellular space in tissues and provides molecular scaffolding for cells within different organs (Parish et al., 2001; Vlodavsky et al., 2002). The basement membrane (BM) is a specialized ECM structure which provides a physical support for cells and represents a surface on which cells can migrate, proliferate and differentiate (Parish et al., 2001). The BM is composed of HSPGs and plays a role in molecular sieving; it is formed from a hydrated matrix due to the negative sulphate groups on HS (Parish et al., 2001; Vlodavsky et al., 2002). The interaction of HSPGs with the variety of growth factors and cytokines sequesters them at the cell surface and ECM. This controls normal and pathological processes, including: morphogenesis, tissue repair, inflammation, angiogenesis and cancer metastasis (Parish et al., 2001; Vlodavsky et al., 2002). The disruption of the HS chains can thus lead to disease pathologies, and so drugs which inhibit the disruption of HS chains are currently of growing interest.

### 3.2. Proposed structure and function of heparanase-1 (HPSE-1)

HPSE-1 is an endo-β-D-glucuronidase enzyme expressed in the lymphoid organs and placenta and is up-regulated in inflammation, wound healing and diabetic nephropathy.
The primary function of HPSE-1 is to cleave HS chains on the BM of cells and remodel the ECM. This activity is important in embryonic development, and is also associated with tumour metastasis (McKenzie et al., 2003; Simizu et al., 2004): the metastatic potential of a tumour cell is proportional to its HPSE-1 content (Parish et al., 2001). HPSE-1 is over-expressed in a variety of different cancers and leads to increased angiogenesis (Parish et al., 2001; Zetser et al., 2003).

The human HPSE-1 gene is located on chromosome 4q21.3 and encodes a pre-pro polypeptide of 543 amino acids, which spans 50kb and is composed of 14 exons separated by 13 introns (Baker et al., 1999; Dong et al., 2000; Parish et al., 2001). The pre-pro polypeptide contains a C-terminal transmembrane domain (Pro-515-Ile-543) and an N-terminal signal peptidase (SP (Met-1-Ala-35)), which is cleaved within the endoplasmic reticulum to generate the latent 65kDa pro-form (Figure 3.2) (Parish et al., 2001; McKenzie et al., 2003). The pro-form is further processed to form the mature active heterodimer consisting of the N-terminal 8kDa subunit (Gln-36-Glu-109) non-covalently attached to the C-terminal 50kDa subunit (Lys-158-Ile-543) (Levy-Adam et al., 2003; McKenzie et al., 2003; McKenzie, 2007). This processing is bought about via cleavage of the 6kDa linker peptide between Ser-110 (site 1) and Gln-157 (site 2). This requires the bulky hydrophobic amino acid Tyr-156 for accurate cleavage of site 2, brought about by cathepsin L (Abboud-Jarrous et al., 2005; McKenzie, 2007; Vreys & David, 2007). Cathepsin L is a papain-like cysteine proteinase which is synthesised as a pre-pro enzyme and is activated via auto-activation in the acidic environment of the endosome/lysosomes (Abboud-Jarrous et al., 2008). The mature form of cathepsin L is stored within the lysosomes, functions as an endopeptidase and is crucial in producing mature active HPSE-1 through cleavage of the 6kDa linker peptide (Abboud-Jarrous et al., 2005). HPSE-1 has six putative N-glycosylation sites in the 50kDa subunit, and tunicamycin inhibition indicates that these glycosylated residues are key for enzyme secretion (McKenzie et al., 2003; Simizu et al., 2004). HPSE-1 also has at least two heparin binding domains (residues 158-162 and 270-278), and possibly a third domain (residues 411-432), also in the 50kDa subunit (Levy-adam et al., 2005). Levy-Adam et al (2005) noted that residues 158-162 in HPSE-1 had higher affinity binding to heparin and deletion of these residues from the pro-form of the enzyme resulted in loss of HPSE-1 enzymatic activity. In contrast, deletion of residues 270-278 showed no inhibition of HPSE-1 enzymatic activity; therefore these residues have a lower affinity for heparin.
HS cleavage is brought about by two catalytic Glu residues (Glu-225 and Glu-343) in the active site: both are found in the 50kDa subunit (Hulett et al., 2000; Simizu et al., 2004). It is thought that the 8kDa subunit causes a conformational change within the active site of the 50kDa subunit to facilitate the catalytic mechanism (McKenzie, 2007).

**Figure 3.2:** Processing of human HPSE-1. HPSE-1 is synthesised in its pre-pro form which is processed to the inactive pro form by the removal of the SP. This is further processed via removal of the linkage peptide (110-157) by cathepsin L to form the mature active HPSE-1 heterodimer comprised of the 8kDa and 50kDa subunits. The catalytic putative donor (Glu-225) and nucleophile (Glu-343) are shown. Adapted from (Li & Vlodavsky, 2009).
It was concluded that HPSE-1 is homologous to the glycosyl hydrolase (GH) families 10, 39 and 51, with a sequence identity of around 20% with xylanase of which the 3D structure is known (PDB code 1BG4) (Nardella et al., 2004; Li & Vlodavsky, 2009). The low sequence identity is not enough for MR but the high similarity (57%) along with other biological studies suggests that they share similar structures, characterized by the \((\alpha/\beta)_8\)-TIM barrel fold (Zhou et al., 2006; Nardella et al., 2004). However, the 50kDa subunit of HPSE-1 only corresponds to one \(\alpha\)-helix plus 6 \(\beta/\alpha\) units, suggesting a departure from the classical TIM barrel fold. The 8kDa subunit contributes the first \(\beta/\alpha\) and the 6kDa linker connects the second strand and second \(\alpha\)-helix of the barrel (Vreys and David, 2007; Nardella et al., 2004; Dempsey et al., 2000, Zhou et al, 2006). Due to this, it is postulated that the 6kDa linker peptide impedes substrate binding by blocking access to the active site: this means that HPSE-1 processing by cathepsin L is crucial for its activity (Vreys & David, 2007). A typical GH enzyme uses a general acid catalysis mechanism for the hydrolysis of glycosidic bonds, which requires a proton donor and nucleophile: both are conserved in HPSE-1(Nardella et al., 2004). Glu-225 is located at the edge of the TIM barrel and Glu-343 is located in the middle of the TIM barrel, with both sites being on different sides of the active site cleft (Nardella et al., 2004; Vreys & David, 2007). Both residues make contact with HS at specific, infrequent oligosaccharide sequences within the NS domains to generate internal cleavages with GlcA found at the reducing terminal of the digested HS chains (Courtney et al., 2004). HPSE-1 also makes two hydrogen bonds through Lys-139 to an acidic group and Asn-244 to a carbonyl oxygen to stabilise the HS in the active site (Zhou et al., 2006).

The 3D crystal structure of HPSE-1 has yet to be determined due to a number of factors, including: low quantity HPSE-1 expression, purification, extensive glycosylation and poor diffracting crystals (see Chapters 13 and 14). Zhou et al have produced a HM of the 50kDa subunit of HPSE-1 based on the structure of xylanase (1BG4) (Figure 3.3). A second group (Sapay et al., 2012) has also produced a HM of HPSE-1 based on the structure of xylanase (Figure 3.4). However, this model is based on the active 58kDa form using a GS3 construct in place of the 6kDa linker peptide. Both HM’s have the TIM-barrel fold: the two heparin binding domains: a basic binding site with a groove shape with the catalytic sites located in its middle: N-glycosylation and S-cysteinylation sites exposed to the protein surface. Sapay et al propose that His-296 is potentially important for the catalytic mechanism of HPSE-1. The coordinates of the model are only available on request from Sapay.
Figure 3.3: HM of HPSE-1 (red), superimposed onto xylanase (green) and glycanase (blue) to show the characteristic (α/β)₈-TIM barrel fold. The Glu-225 and Glu-343 residues are highlighted in red and they are bound to a xyloligomer (yellow) to show the active site cleavage mechanism (Zhou et al., 2006).

Figure 3.4: Proposed 3D structure of HPSE-1 using homology modelling (Sapay et al., 2012). (a) Shows the overall fold of HPSE-1 with one rotated 180°, with the 50kDa subunit, 8kDa subunit, catalytic site, HS binding domains, N-glycosylation sites and cysteine residues all shown. (b) Shows the catalytic binding domain in more detail, with potentially important residues identified. E225 and E343 are the two catalytic sites, with H296 thought to be involved in the catalytic mechanism. The N residues are the potential N-glycosylation sites of HPSE-1 and the other residues shown are thought to be involved in Heparin/HS binding.
HPSE-1 is a member of the GH 79 family and up until February 2012 no 3D structure of this family was available. Michikawa et al (2012) solved the 3D structure of a beta-glucuronidase enzyme form *acidobacterium capsulatum* (PDB code 3VNY). 3VNY has a sequence identity of 25% when compared to the 50kDa subunit of HPSE-1 with 70% coverage. 3VNY contains the (β/α)₈ TIM barrel fold and has two catalytic Glu residues just like HPSE-1 (Chapter 15 describes the production of a HM of HPSE-1 based on 3VNY).

The HM created by Zhou et al and Sapay et al and the 3D structure of 3VNY (Michikawa et al., 2012), all contribute to the understanding of the 3D structure and mechanism of action of HPSE-1. However, the proteins studied only have 20% (1BG4) and 25% (3VNY) sequence identity to HPSE-1, whereas true HM need around 40% identity in the amino acid sequence.

### 3.3. HPSE-1 gene regulation

The human HPSE-1 gene is located on chromosome 4q21-23 and its promoter consists of 3 Sp1 binding sites, 4 ERE binding sites, 2 GC boxes and a GT box (Jiang et al., 2002). Sp1 binds to these GC boxes (located in two out of the three Sp1 binding sites) and is expressed in most cell types. However, HPSE-1 expression is restricted to a limited number of cell types; the lack of HPSE-1 expression in normal epithelial cells could be partly due to post-transcriptional modifications regulating the inhibition of the HPSE-1 promoter (Jiang et al., 2002). Sp1 is essential for embryogenesis, as Sp1 KO embryos suffer from growth retardation and die prematurely, and mutations of these Sp1 sites lead to impaired HPSE-1 expression, thus Sp1 is the critical transcription factor in regulating basal HPSE-1 promoter activity (Jiang et al., 2002). A second transcription factor, GABP, binds to the ERE binding sites of the HPSE-1 promoters and is responsible for directing transcription initiation by recruiting initiation transcription factors. Hence, Sp1 and GABP co-operatively regulate HPSE-1 promoter activity in normal cells (Jiang et al., 2002).

HPSE-1 expression in normal cells is also regulated at the post-transcriptional level. The 3’ un-translated region (UTR) of HPSE-1 contains a 185-bp region (ARE motif), and alternative splicing of the HPSE-1 gene either results in the full HPSE-1 sequence or a sequence without the ARE motif. ARE motifs have the ability to target mRNA for rapid degradation (Arvatz et al, 2010). The full length HPSE-1 gene sequence results in reduced HPSE-1 levels, but the shorter form with the ARE motif missing causes an increase in HPSE-1 mRNA levels, and
hence HPSE-1 expression, enzymatic activity, cell invasion capacity and the formation of large tumours (Arvatz et al, 2010). In cancer cells, this ARE motif is absent from the HPSE-1 gene sequence, resulting in enhanced HPSE-1 expression and thus could potentially lead to tumour growth and metastasis (Arvatz et al, 2010). The mechanism for HPSE-1 induction is not fully understood as of yet and research is still ongoing.

Also, the HPSE-1 promoter is under regulatory control by p53 (a transcription factor, which has been termed ‘the guardian of the genome’, and which, when mutated, results in cancer) in normal cells by recruiting histone deactylases, which inhibit HPSE-1 expression (Arvatz et al, 2010). However, in cancers, where p53 is mutated or inhibited, its repression of HPSE-1 expression is lost, thus HPSE-1 expression is up-regulated in these types of cancers (Arvatz et al, 2011; Ilan et al, 2006) possibly leading to increase in tumour growth and metastasis.

HPSE-1 expression in cancers is thought to be due to either dysregulation of signalling molecules or transcription factors at a late stage of tumour progression (Jiang et al., 2002) or via aberrant demethylation, as DNA methylation maintains the low levels of HPSE-1 seen in most normal tissues (Simizu et al, 2003).

### 3.4. HPSE-1 in disease

Over-expression of human HPSE-1 has been indicated in many pathologies due to the increased cleavage of the HS chains in both the ECM and BM of cells, where many different growth factors and cytokines are sequestered.

#### 3.4.1 HPSE-1 in tumour metastasis

The cleavage of HS chains in the ECM is believed to be related to the metastatic potential of a tumour cell. It has been identified that HPSE-1 is up-regulated in all human tumours examined using immunohistochemistry, with the most intense HPSE-1 staining localized at the invasive front of the tumour (Vreys & David, 2007; Vlodavsky et al., 2008). This supports a role for HPSE-1 in tumour metastasis and cell invasion due to disruption of the ECM allowing tumour cells to infiltrate into the space which was once occupied by the ECM. Cancer patients who exhibit tumour HPSE-1 positive staining have a significantly higher rate of metastases and reduced post-operative survival compared to patients with tumours with
undetectable or very low levels of HPSE-1 present, again suggesting the key role of HPSE-1 in tumour metastasis (Vreys & David, 2007; Vlodavsky et al., 2008).

3.4.2 HPSE-1 in tumour angiogenesis

In addition to its role in tumour metastasis, HPSE-1 is also involved in tumour angiogenesis. Angiogenesis is the process of developing new blood vessels from existing blood vessels, providing the tumour with increased blood supply. Blood vessels are lined with endothelial cells (ECs) and HPSE-1 is thought to degrade the HS chains on the BM of capillaries (similar to its mechanism for tumour metastasis) which in turn results in EC invasion and migration towards the tumour, increasing the blood vessel network to the tumour (Ilan et al., 2006; Vreys & David, 2007). A second route in which HPSE-1 facilitates angiogenesis is by releasing pro-angiogenic growth factors (bFGF, VEGF) from HS chains in the ECM. This promotes binding to their respective receptors and increased cell signalling, resulting in EC migration and proliferation (Elkin et al., 2001; Vlodavsky et al., 2002, 2008; Ilan et al., 2006; Vreys & David, 2007; Xu et al., 2011)

3.4.3 HPSE-1 in diabetes

The normal expression pattern of HPSE-1 is in tubular epithelial cells but not in the glomerular cells of normal kidney cells. However, patients with diabetic nephropathy have increased HPSE-1 expression in both the tubular epithelial cells and the glomerular cells of the kidneys, leading to damage and dysfunction of the glomerular BM (GBM) (Maxhimer et al., 2005; Szymczak et al., 2010; Shafat et al., 2011). Due to the dysfunction of the GBM, HPSE-1 has been detected in the urine of diabetic patients (Han et al., 2012). This increase of HPSE-1 expression is related to high levels of blood glucose (Maxhimer et al., 2005; Szymczak et al., 2010).
3.4.4. HPSE-1 in inflammation

The inflammatory state serves to eliminate pathogens and dead cells. An inflammatory stimulus typically induces cytokines (e.g. IL-1β or TNF-α), which in turn activate a variety of cells (EC, microglia/macrophages, astrocytes) to release cell-attractant chemokines (CCL2), and up-regulate expression of cell adhesion molecules (ICAM-1) (Zhang et al., 2012).

HS chains in the ECM of cells regulate inflammatory responses in a number of ways: by sequestering cytokines and chemokines into the intracellular space, modulation of leukocytes’ interaction with the ECM through ICAM-1, and initiation of an inflammatory response through interactions with toll-like receptor 4 (TLR4) (Goldberg et al., 2013). HS degradation due to overexpressed HPSE-1 can thus affect leukocyte recruitment, extravasation and migration towards the inflammatory site. This results in release of cytokines and chemokines as well as activation of innate immune cells (neutrophils and T-lymphocytes). HPSE-1 activity was discovered in neutrophils and activated T-lymphocytes which contributed to their ability to extravasate and accumulate in target organs (Vlodavsky et al., 1992, 2011; Goldberg et al., 2013). HPSE-1 expression occurs mainly in epithelial and endothelial compartments in several auto-immune and auto-inflammatory human disorders, such as rheumatoid arthritis, atherosclerosis, inflammatory lung disease, Ulcerative Colitis and Crohn’s disease (Goldberg et al., 2013). Up-regulation of HPSE-1 has also been found in the colonic epithelium of patients with inflammatory bowel disease, which is linked to colon cancer (Vlodavsky et al., 2011). All these diseases have an increased inflammatory signal due to over-expression of HPSE-1 and subsequent degradation of HS chains, leading to an increased inflammatory response.

3.5. Inhibitors of HPSE-1

HPSE-1 is over-expressed in a variety of different cancers and different diseases (see section 3.4) and the 3D structure is crucial in developing active site-specific inhibitors for use in anti-cancer treatments. Small molecule and sugar inhibitors have been discovered through HPSE-1 high throughput assays (McKenzie, 2007; Courtney et al., 2004a; Pan et al., 2006; Xu et al., 2006). The compound, 2,3-dihydro-1,3-dioxo-1H-isoindole-5-carboxylic acid (Figure 3.5a) was found to be both an inhibitor of HPSE-1 activity (IC₅₀ 8µM) and angiogenesis (IC₅₀ 40µM) (Courtney et al., 2004a; McKenzie, 2007). This compound is ideal for drug selection.
due to its low molecular weight, lipophilicity, polar surface and its ability to be modified to select more potent inhibitors (Courtney et al., 2004a; McKenzie, 2007). The central aromatic core and benzoaxole moiety of the inhibitor can be modified to increase its inhibitory effect, with the most potent inhibitor having an IC$_{50}$ of 0.5µM for HPSE-1 inhibition and 0.25µM for angiogenesis inhibition (Courtney et al., 2004a; McKenzie, 2007). Furanythiazole acetic acid (Figure 3.5b) has also been found to inhibit HPSE-1 (IC$_{50}$ 25µM) (Courtney et al., 2005). Modifications focused on replacing the 4-chloro substituent with an amino function which allowed introduction of a wide range of groups (Figure 3.5c) (Courtney et al., 2005). The most potent of these inhibitors had the 4-bromo substituent and had an IC$_{50}$ of 0.4µM for HPSE-1 inhibition and 1µM for angiogenesis inhibition (Courtney et al., 2005; McKenzie, 2007). Other small molecule inhibitors developed include the benzoazol-5-yl acetic acid compounds (Courtney et al., 2005), the symmetrical 1-[4-(1H-Benzimidazol-2-yl)-phenyl]-3-[4-(1H-benzoimidazol-2-yl)-phenyl] urea compounds (Pan et al., 2006) and the N-(4-{[4-(1H-Benzimidazol-2-yl)-arylamino]-methyl}phenyl)-benzamide compounds (Pan et al., 2006). These small molecule inhibitors provide a basis for the design of novel therapeutic agents, however, none of these small molecule inhibitors have reached clinical trials as of yet, thus new inhibitors need to be developed and the experimental 3D structure of HPSE-1 is of great importance to help design more potent small molecule inhibitors (Courtney et al., 2004, 2005).

**Figure 3.5:** Small molecule inhibitors of HPSE-1. (a) The 2,3-dihydro-1,3-dioxo-1H-isooindole-5-carboxylic acid compound (b) the furanythiazole acetic acid compound and (c) the modified furanythiazole acetic acid at the 4-chloro substituent (Zetser et al., 2003).
In 2008, the most advanced inhibitor of HPSE-1 was a highly sulphated pentasaccharide molecule (PI-88 (Figure 3.6)), and at the time was the only HPSE-1 inhibitor to reach clinical trials as a monotherapy or in combination with chemotherapy (Courtney et al., 2004b, 2005; Kudchadkar et al, 2008; McKenzie, 2007). PI-88 inhibits HPSE-1 activity (IC$_{50}$ 1µM) and angiogenesis by blocking the breakdown of the ECM, degradation of the BM and the release of growth factors (Kudchadker et al, 2008). These actions are brought about by either binding HPSE-1 directly, or by acting as a HS mimetic and forming complexes with certain growth factors, inhibiting interactions with their receptors. This leads to a decrease in tumour invasion, new vessel development and inhibition of cell proliferation and survival (Kudchadker et al, 2008). However, phase I and II clinical trials only produced modest results, and PI-88 also causes the development of immune-mediated thrombocytopenia, thus further developments of this inhibitor were halted (Kudchadkar et al., 2008).

A new series of HPSE-1 inhibitors, the PG500 series, are HS mimetics, similar to PI-88 but are fully sulphated oligosaccharides attached to a lipophilic moiety (aglycone) (Figure 3.7). These new compounds have increased potency compared to PI-88 based on in vitro angiogenesis assays and in vivo tumour models (Dredge et al., 2010) due to the addition of this aglycone group. Similarly to PI-88, the PG500 series interfere with angiogenesis via inhibition of VEGF, FGF-1 and FGF-2 growth factors and with metastasis via inhibition of HPSE-1 (Dredge et al., 2010, 2011; Ferro et al., 2012). Out of the PG500 series, PG545 (Figure 3.7) is a potent HPSE-1 inhibitor (Ki = 6nM) and angiogenesis inhibitor (<1µM) and has milder anti-coagulant properties than PI-88 or earlier PG500 compounds (Dredge et al., 2011). Pre-clinical data confirmed that a once or twice weekly dosing of PG545 is anti-
angiogenic: this dosing schedule is much less frequent compared with PI-88 (Dredge et al., 2011; Ostapoff et al., 2013), which was administered daily for four days every week at a dose of 250mg in a 28 day cycle (Gandhi & Mancera, 2010). Thus PG545 is a leading candidate in HPSE-1 inhibition and has recently started phase I and phase II clinical trials.

Recently, two different groups have modified heparin to produce two compounds, sst0001 (Ritchie et al., 2011) and m402 (Zhou et al., 2011), which both have biological activity against HPSE-1. Heparin was used as a starting point, as unmodified heparin is known to possess potent anti-HPSE-1 activity (Bar-Ner et al., 1987). However, using heparin as an anti-cancer agent is undesirable due to its anti-coagulant properties. It is, however, possible to separate the anti-coagulant and anti-heparanase properties of heparin through chemical modifications, thus these non-anti-coagulant heparin molecules could be administered for anti-cancer treatments without the risk of promoting bleeding disorders (Ritchie et al., 2011). sst0001 is a 100% N-acetylated and 25% glycol split, which causes heparin to lose its affinity for anti-thrombin, resulting in the loss of anti-coagulant activity, but maintaining the ability to inhibit HPSE-1 (Ritchie et al., 2011). Pre-clinical data of sst0001 confirms that it is an effective HPSE-1 inhibitor in vivo, exhibiting no side effects (Ritchie et al., 2011). The next stage is to enter clinical trials to determine its potential as a therapeutic agent. sst0001 has been combined with ionizing radiation (IR) and attenuated orthotopic pancreatic tumour spread in vivo (Meirovitz et al., 2011). It is known that pancreatic tumours are resistant to IR alone due to the up-regulation of HPSE-1 expression which involves expression of the Egr-1 transcription factor, but when given along with sst0001, HPSE-1 is inhibited and this prevents tumour resistance and progression (Meirovitz et al., 2011). Clinical trials of this inhibitor started in January 2013.
The HS mimetic, m402 is a heparin-like product which has reduced anti-coagulant activity but still retains the ability to bind to key factors in tumour growth and metastasis (VEGF, FGF and P-selectin); it is a novel HSPG mimetic (Zhou et al., 2011). The difference between m402 and sst0001 is the fact that m402 is not N-acetylated, as it is known that N-acetylation of heparin causes reduced binding to VEGF and FGF. However, sst0001 is 100% N-acetylated but still has anti-angiogenic activity. The pre-clinical data confirms that m402 is a good target drug and clinical trials started in July 2012.

Both the small molecule and sugar inhibitors developed have shown promise in the inhibition of HPSE-1 and angiogenesis. However, these findings have not been replicated in clinical trials as of yet and no HPSE-1 inhibitor is available. There is a pressing need for the 3D structure of HPSE-1 to be elucidated to aid in the discovery of novel and more potent/specific inhibitors.

3.6. Discovery of HPSE-2

A second HPSE enzyme localized on chromosome 10q23-24 was discovered using the published amino acid sequence of HPSE-1 (McKenzie et al., 2000; Vreys & David, 2007), encoding a 592 amino acid protein (Ganesan & Thomas, 2011). Alternative splicing of HPSE-2 gives rise to three different mRNAs, HPSE-2A, HPSE-2B and HPSE-2C, encoding proteins of 480, 534 and 592 amino acids respectively (Figure 3.8) (Vreys and David, 2007; McKenzie et al., 2000; Perreti et al, 2008). The cDNA sequence features two potential ATG-start codons and it is believed that translation starts from the second ATG, resulting in a SP (residues 1-31) (Vreys & David, 2007). Sequence analyses of the three splice variants of HPSE-2 predicts that the proteins are intracellular membrane bound enzymes, with a hydrophobic N-terminus bound to the membrane and a C-terminus retained in the cell (Peretti et al., 2008). HPSE-2C is the only splice variant which is secreted, possibly due to the extra glycosylation sites, which are lost in the HPSE-2B and 2A splice variants (Levy-Adam et al, 2010). HPSE-2 has an overall sequence identity of 35-40% with HPSE-1, which includes conservation of the two catalytic Glu residues. Also, the 50kDa and 8kDa subunits show striking homology between the two enzymes implying that both enzymes are arranged as two subunits (McKenzie et al., 2000; Vreys & David, 2007). However, the 6kDa linker peptides are not conserved, suggesting that HPSE-2 may not undergo post-translational protease
processing like that of HPSE-1 (Figure 3.2) (McKenzie et al., 2000; Vreys & David, 2007). HPSE-2C exhibits higher affinity towards heparin/HS compared to HPSE-1 despite the lack of intrinsic HS-degrading activity (Levy-adam et al., 2010). Thus, HPSE-2C could be important in HPSE-1 inhibition. Indeed, HPSE-2C has been found to be up-regulated in both head and neck carcinomas, which relates to favourable prognosis and prolonged follow-up, and also correlates with reduced tumour metastasis.

The mRNA distribution of each enzyme is considerably different, with HPSE-2 expressed in the brain, testis, uterus and bladder and with no expression in the placenta and lymphoid organs where HPSE-1 is dominant (McKenzie et al., 2000; Vreys & David, 2007). HPSE-2 is over-expressed in some colorectal tumours, whereas HPSE-1 is over-expressed in a variety of different cancers (McKenzie et al., 2000; Vreys & David, 2007). The over-expression of HPSE-2C in these colon cancers relates to a decrease in syndecan-1 from the cell surface (a HSPG in the ECM which binds to growth factors (FGF2, VEGF) sequestering them) which activates proliferation or cell survival of tumour cells (Peretti et al., 2008). HPSE-2C binds to syndecans like HPSE-1, but unlike HPSE-1, it is not internalized with the syndecans, leaving HPSE-2C on the cell surface (Levy-adam et al., 2010; Arvatz et al., 2011). Therefore, the internalisation of HPSE-1 with the syndecans is unique, thus mechanisms for the internalisation of syndecans are of clinical interest, as syndecans mediate the uptake of bacteria and viruses (Arvatz et al., 2011). HPSE-2C could be used against these bacteria and viruses to inhibit their internalization.

Loss of function mutations in HPSE-2 resulting in a truncated protein have been shown to cause urofacial syndrome (UFS), a disease which causes the patient to have a distorted smile, urinary tract infections, incontinence and enuresis which can progress to renal failure and death (Pang et al., 2010; Akl & Momany, 2012). UFS is thought to be caused by a mutation in a gene which has roles within areas of the brain that control facial expression and bladder voiding, or due to lesions in neurones within or near the bladder (Daly et al., 2010). Patients with UFS have shown no HPSE-2 expression in the brain or spinal cord but the facial muscles and urinary bladder show very high levels of HPSE-2 expression (Pang et al., 2010). The mechanism of how HPSE-2 causes UFS is unknown at the moment.
Figure 3.8: Domain structure of the 3 isoforms of HPSE 2. They all contain the 2 putative catalytic residues Glu-225 and Glu-343 (stars). The 5 N-glycosylation positions (lines) are shown, with only H2C containing all 5 N-glycosylation sites. Modified from (Vreys & David, 2007).

The discovery of HPSE-2 could be crucial in determining the 3D structure of HPSE-1. The difficulties in determining HPSE-1 structure include low expression, low purity and disordered crystals. However, if the 3D structure of HPSE-2 is elucidated, MR methods could be used once crystals of HPSE-1 are produced. Otherwise, homology modelling techniques can be used to predict the 3D structure of HPSE-1 based on the structure of HPSE-2 which would be more accurate compared to the methods used by Zhou et al, Sapay et al and the methods outlined in chapter 15, as the sequence similarity of these two enzymes are higher.
Chapter 4: Binding of the anticancer platinum drugs cisplatin and carboplatin to hen egg white lysozyme (HEWL) as a model protein

4.1. Structure and function of lysozyme

Lysozyme is a textbook molecule for which the crystallisation conditions are well known. Also, it has a low molecular weight (14.7kDa) and the crystals are easy to grow (http://lysozyme.co.uk/). Lysozyme is present in mucosal secretion and is part of the innate immune response: it hydrolyses gram positive bacterial cell membranes (Fleming, 1922; Kirby, 2001) at the 1,4-beta-linkages between N-acetylmuramic acid (NAM) and N-acetyl-D-glucosamine (NAG) residues in the peptidoglycan (Fleming, 1922; Kirby, 2001). This reaction occurs in a long, deep cleft which contains a specific active site (containing 6 binding sites (A-F)), which hydrolyses hexasaccharides into disaccharide and tetrasaccharide subunits (Blake et al., 1967; Kirby, 2001). The active site contains a conserved proton donor (Glu-35) and nucleophile (Asp-52), situated between sub-sites D and E, which carry out the hydrolysis reaction (http://lysozyme.co.uk/).

HPSE-1 and lysozyme both use a similar catalytic mechanism to hydrolyse sugar molecules and it is thought that metal-containing inhibitors which bind into the active site of lysozyme could be used as cofactors in the crystallisation of HPSE-1 to form well-ordered diffracting crystals, giving rise to the 3D structure through IR/AD.

4.2. Platinum containing anticancer drugs

The platinum anti-cancer drugs cisplatin and carboplatin (Figure 4.1) have dominated the treatment of a variety of cancers (cervical, head and neck, testicular, bladder, ovarian and non-small-cell lung cancers), with treatment usually given in combination with chemotherapy and radiation therapy (Ivanov et al., 1998; Reedijk, 2003; Vallerga et al., 2004; Wang & Lippard, 2005). Cisplatin and carboplatin are radiation sensitizers (Peters et al., 2000), meaning they are stable under high energy radiation. Therefore, the concurrent use of both radiation therapy and cisplatin administration in cancer treatment causes an increase in cell death compared to using either treatment individually (Seiwert et al., 2007). The combination of cisplatin and radiation therapy uses several mechanisms to enhance cell death, including enhanced formation of toxic platinum intermediates in the presence of radiation-induced free
radicals; inhibition of DNA repair; radiation-induced increase in cellular platinum uptake and cell cycle arrest (Lawrence et al., 2003; Kvols, 2005).

Cisplatin was first discovered in 1845 by Michel Peyrone (Hall, 2010), however, the anti-cancer properties were not revealed until 1965 by Rosenberg et al (Rosenberg et al., 1965), when it was discovered that the formation of cisplatin caused E. coli cell division to arrest (Reedijk, 2003). Cisplatin was approved by the Food and Drug Administration (FDA) in 1978 for administration in conjunction with chemotherapy (Reedijk, 2003; Wang & Lippard, 2005).

DNA is the main biological target of cisplatin and carboplatin, however, 90% of binding is to plasma proteins (Fischer et al., 2008). The protein-platinum anti-cancer drug interactions have attracted interest as they are considered crucial for the pharmacokinetics, biodistribution, resistance processes and toxicity of these metallodrugs (Calderone et al., 2006; Casini et al., 2006, 2007a).

Cisplatin is administered intravenously, encountering a relatively high chloride concentration in the blood plasma, which stabilises the molecule (Alderden et al., 2006). Cisplatin enters the tumour cell either through diffusion or transport by a copper transporter (Ctr1p) (Kostova, 2006). Cisplatin encounters a low chloride concentration in the cell, causing displacement of one chloride ligand for a water molecule, thus becoming positively charged and reactive, whilst being trapped inside the cell (Alderden et al., 2006; Kostova, 2006). A second water molecule displaces the remaining chloride ligand, forming the reactive species ((NH₃)₂Pt). This exerts its anti-cancer effect via interactions with the N7 atoms of guanine bases in DNA (Figure 4.2), forming intrastrand and interstrand crosslinks (Kostova, 2006). Formation of the crosslinks locally unwind and bend double stranded DNA towards the major groove, opening the face of the minor groove and leading to inhibition of DNA replication, transcription and triggering apoptotic cell death (Benedetti et al., 2002; Silverman et al., 2002; Kostova, 2006).
Apoptosis is induced via the binding of a high mobility group to the minor groove of the platinum DNA complex to protect from DNA repair enzymes (Silverman et al., 2002; Kostova, 2006).

Figure 4.2 Cisplatin binding to the N7 atoms of two adjacent guanine bases in a DNA double helix. The 2 Cl atoms have been displaced with the two NH3 groups bound to the Pt centre. PDB code 3LPV (Todd & Lippard, 2010)

Cisplatin is a very effective anti-cancer drug but is rapidly converted to toxic metabolites which have nephrotoxic effects (Zhang, 1996). Cisplatin is thought to bind to the cytoplasmic side of the Na+/K+-ATPase pump, inhibiting this transporter (Huličiak et al., 2012). The Na+/K+-ATPase pump creates the Na+ gradient in the kidney, thus inhibition of this pump causes a collapse of the whole machinery. Cisplatin is thought to bind many Cys residues in the C45 cytoplasmic loop of the Na+/K+-ATPase pump. Binding to Cys 367 is hypothesised to cause a steric hindrance in the phosphorylation and dephosphorylation steps in the Na+/K+-ATPase catalytic cycle. Asp369 is a conserved phosphorylation site, so when cisplatin is bound to Cys 367, this causes inhibition of this phosphorylation mechanism, inhibiting the catalytic cycle, and causing nephrotoxic effects seen in cisplatin treated patients (Huličiak et al., 2012).
Carboplatin, a second generation inhibitor was approved by the FDA in 1988 (Reedijk, 2003) and exerts its effect using the same mechanism as cisplatin. However, due to the addition of the cyclobutanedicarboxylate moiety (CBDC) (Figure 4.1), carboplatin has a slower rate of conversion to toxic metabolites and is tolerated at higher doses than cisplatin (Kostova, 2006).

The toxicity effect caused by the platinum anti-cancer drugs could be partly due to binding extracellular and intracellular proteins, forming drug-inactivation products (Sun et al., 2009). A number of proteins have been studied for their binding affinity to cisplatin, including Ctr1p (Arnesano & Natile, 2008; Crider et al., 2010), superoxide dismutase (Fisher et al., 1991; Calderone et al., 2006), cytochrome c (Casini et al., 2006, 2007a), human albumin (Ivanov et al., 1998) ubiquitin (Hartinger et al., 2007; Zimmermann & Burda, 2010), glutathione (Hartinger et al., 2007; Zhao & King, 2011) and HEWL, a model protein (Casini et al., 2007b; Tanley et al., 2012a; b; Helliwell & Tanley, 2013). Platinum is a soft ion, which makes favourable interactions with sulphur atoms present in proteins through Met and Cys side chains forming strong bonds (Hahn et al., 2001; Zhao & King, 2011). However, HEWL and superoxide dismutase have free His residues which co-ordinate cisplatin through the nitrogen atoms on the imidazole ring (Fisher et al., 1991; Casini et al., 2007b; Tanley et al., 2012a; b; Helliwell & Tanley, 2013).

4.3. Platinum anti-cancer drugs binding to HEWL.

Interactions between HEWL and metal-containing drugs have been analysed, as they are a model for the pharmacokinetics, biodistribution, resistance processes and toxicity of metallodrugs (Casini et al., 2007b). A study by Li (Li, 2006) shows that a number of metal ions (Mn$^{2+}$, Co$^{2+}$ and Ni$^{2+}$) bind to the active site of HEWL, but platinum ions were not tested in this study. Therefore, it is possible that both cisplatin and carboplatin bind to the active site of HEWL and act as sugar binding mimics.

Casini et al (Casini et al., 2007b) investigated the crystallisation of HEWL with both cisplatin and carboplatin via the soaking method. This procedure involved soaking cisplatin and carboplatin at a 10-fold molar excess into pre-grown HEWL crystals at 4°C. Crystallographic studies indicated cisplatin bound with an occupancy of 50% to a nitrogen atom in the imidazole ring of His-15, the only His residue in HEWL. However, in their X-ray crystal structure, carboplatin binding was not seen, in stark contrast to their MS data which showed a
significant mass increase upon carboplatin binding. The MS data also showed binding of a second carboplatin and cisplatin molecule, which was not seen in the X-ray crystal structure. A molecule of dimethyl sulfoxide (DMSO) was seen bound at site C in the active site and therefore, acts as an inhibitor of HEWL. DMSO binding in the active site of HEWL has been seen previously (Jóhannesson et al., 1997), coordinated by Trp-103 and Trp-68.

These studies have shown that the platinum containing drugs bind to HEWL through His-15 which could be a crucial factor in the toxicity of these drugs. Further studies without the addition of DMSO to the crystallisation conditions are needed to confirm if cisplatin and carboplatin bind to the active site of HEWL.

These studies as well as a variety of others looking at the binding of cisplatin and carboplatin to HEWL have been undertaken throughout my PhD and the details of this work and the results obtained are given in Chapters 5-12.
Part I References


Books and online references for Chapter 1-4


Rupp B (2010). Protein crystallisation. In Biomolecular crystallography, principles, practice and application to structural biology pg. 77-134. Published by Garland science


Stout GH and Jensen LH (1989). X-ray structure determination: a practical guide. Published by Wiley and Sons

Part II Structural studies of cisplatin and carboplatin binding to a model protein

Introduction to Part II

The next few chapters document the experimental work I have carried out throughout my PhD, focussing on the binding of cisplatin and carboplatin to HEWL, a model protein. These chapters are presented as papers, with them being either published in peer-review journals (Chapters 5-9), submitted for publication or in preparation (Chapters 11-12), with the exception of one which has been published with arXiv, an e-print journal without peer review (Chapter 10).

I have used X-ray crystallography to study the binding of cisplatin and carboplatin to HEWL. These papers will describe, in detail, the binding of cisplatin and carboplatin to HEWL in DMSO and aqueous medium after 8 days of co-crystallisation (Chapter 5) and after prolonged chemical exposure (between 7-15 months) (Chapter 7), as well as co-crystallising HEWL with cisplatin and carboplatin together to see which one out-competes for binding (Chapter 8). A room temperature study was also carried out on the cisplatin crystals to determine how stable these are during high exposure to X-ray irradiation, as these drugs are used in conjunction with radiation therapy (Chapter 8).

As well as studying the binding of these drugs to HEWL, the raw diffraction data images from all our crystals have been deposited at a webserver maintained by the University of Utrecht, at the raw Tardis archive in Australia and will be available at the University of Manchester through DOI links to the individual datasets (Chapter 6). By making our raw diffraction data images freely available to the community, one can download our data and compare different software packages. With our data, we compared four different software packages (Mosflm, EVAL, d*TREK and Proteum 2) to see how they handle the same data. As a result, collaboration was set up with Kay Diederichs from the University of Konstanz, who downloaded our raw diffraction images and re-processed these with XDS, a different crystallographic software package (Chapter 9). Based on this new processing, we actually see partial conversion of carboplatin to cisplatin due to the high Cl concentration used in the crystallisation conditions. Due to these new findings, non-Cl conditions were sampled in order to co-crystallise HEWL with carboplatin to remove this ability for carboplatin to be
converted to cisplatin (Chapters 10 and 11). A variety of different non-Cl conditions were obtained at different pH’s with carboplatin bound. As well as removing the Cl ions from the crystallisation conditions, NaBr crystallisation conditions were carried out, to see more clearly any chemical transformation of carboplatin to the bromo platinated form. These crystals were measured at beamline I04 at Diamond at an X-ray wavelength of 0.92Å, the Br Kedge, to increase the anomalous signal as well as using the 2Fo-Fc electron density map to identify if the Br ions did indeed substitute onto the Pt centre (Chapter 11).

All of the crystallisation conditions used thus far grew tetragonal crystals which gave the best diffraction of 1.4Å resolution. A new study looked at growing triclinic HEWL crystals as the triclinic form has been studied at atomic resolution of 0.65Å (Chapter 12). The thought was to co-crystallise cisplatin/carboplatin with triclinic HEWL, then one could see atomic resolution of the anti-cancer drugs bound to a protein, and could potentially lead to charge density detail of the ligands. The Pt atoms showed multiple binding sites, thus atomic resolution of these ligands is unattainable in this crystal system. However, the triclinic data did yield structural dynamics of the Pt atoms at different data collection temperatures.
5. Chapter 5: Structural Studies of the effect that dimethyl sulfoxide (DMSO) has on cisplatin and carboplatin binding to histidine in a protein

Simon W.M. Tanley\textsuperscript{a}, Antoine M.M. Schreurs\textsuperscript{b}, Loes M.J. Kroon-Batenburg\textsuperscript{b}, Joanne Meredith\textsuperscript{a}, Richard Prendergast\textsuperscript{a}, Danielle Walsh\textsuperscript{a}, Patrick Bryant\textsuperscript{c}, Colin Levy\textsuperscript{c} and John R. Helliwell\textsuperscript{a*}

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, UK

\textsuperscript{b} Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

\textsuperscript{c} Manchester Interdisciplinary Biocentre (MIB), University of Manchester, 131 Princess Street, Manchester M1 7DN, UK, England.


Keywords: Protein; Hen egg white lysozyme (HEWL); dimethyl sulfoxide (DMSO); Histidine (His); Platinum (Pt); N-acetylglucosamine (NAG); Cisplatin; Carboplatin

JM, RP and DW were all masters’ students in the group and collected one dataset each, which SWMT refined to completion. SWMT set up the other 8 crystal growth experiments, mounted and collected them using the Bruker and R-Axis IV diffractometers. CL and PB run the X-ray facility and helped with collection. LKB and AMMS processed each dataset with EVAL-15. SWMT processed each dataset with Mosflm, d*TREK and Proteum and refined all datasets. The paper was mainly written by SWMT and JRH with comments from all co-authors. JRH designed these experiments, gave advice, expertise and guidance to SWMT.
Structural studies of the effect that dimethyl sulfoxide (DMSO) has on cisplatin and carboplatin binding to histidine in a protein

The anticancer complexes cisplatin and carboplatin target the DNA major groove, forming intrastrand and interstrand cross-links between guanine bases through their N7 atoms, causing distortion of the DNA helix and apoptotic cell death. A major side effect of these drugs is toxicity, which is caused via binding to many proteins in the body. A range of crystallographic studies have been carried out involving the co-crystallization of hen egg-white lysozyme (HEWL) as a test protein with cisplatin and carboplatin in aqueous and dimethyl sulfoxide (DMSO) conditions. Different cryoprotectants, glycerol and Paratone, were used for each of the cisplatin and carboplatin co-crystallization cases, while silicone oil was used for studies involving N-acetylglucosamine (NAG). Both cisplatin and carboplatin do not bind to HEWL in aqueous media on the timescales of the conditions used here, but upon addition of DMSO two molecules of cisplatin or carboplatin bind either side of His15, which is the only His residue in lysozyme and is assumed to be an imidazolyl anion or a chemical resonance moiety, i.e. both imidazole N atoms are chemically reactive. To identify the platinum-peak positions in the ‘with DMSO conditions’, anomalous scattering maps were calculated as a cross-check with the $F_o - F_c$ OMIT maps. Platinum-occupancy values were established using three different software programs in each case. The use of EVAL15 to process all of the diffraction data sets provided a consistent platform for a large ensemble of data sets for the various protein and platinum-compound model refinements with REFMAC5 and then SHELXTL. Overall, this extensive set of crystallization and cryoprotectant conditions allowed a systematic evaluation of cisplatin and carboplatin binding to lysozyme as a test protein via detailed X-ray crystal structure characterizations. DMSO is used as a super-solvent for drug delivery as it is deemed to cause no effect upon drug binding. However, these results show that addition of DMSO causes the platinum anticancer drugs to bind to HEWL. This effect should be considered in toxicity assessments of these drugs and perhaps more widely.

1. Introduction

DNA is the main biological target of the platinum anticancer drugs cisplatin and carboplatin. However, 90% of their reported binding is to plasma proteins (Fischer et al., 2008). Protein–platinum anticancer drug interactions have attracted interest as they are considered to be crucial for the pharmacokinetics, biodistribution, resistance processes and toxicity of these metallodrugs (Calderone et al., 2006; Casini et al., 2006; Casini, Mastrobuoni et al., 2007).
**Cisplatin exerts its anticancer effect by targeting the DNA major groove, forming intrastrand and interstrand cross-links between guanine bases via their N7 atoms (Kostova, 2006) and causing distortion of the DNA helix that leads to inhibition of DNA replication and transcription and triggers apoptotic cell death (Benedetti et al., 2002; Silverman et al., 2002). However, the toxicity caused by cisplatin and carboplatin could partly arise from binding to extracellular and intracellular proteins, forming drug-inactivation products (Sun et al., 2009).** Binding affinity to cisplatin has been studied for a number of proteins, including a copper transporter (Arnesano & Natile, 2008; Crider et al., 2010), a copper chaperone (Boal & Rosenzweig, 2009), superoxide dismutase (Calderone et al., 2006; Casini et al., 2008), cytochrome c (Casini et al., 2006; Casini, Gabbiani et al., 2007), human albumin (Ivanov et al., 1998), ubiquitin (Hartinger et al., 2006), glutathione (Zimmermann & Burda, 2010) and hen egg-white lysozyme (HEWL; Casini, Mastrobuoni et al., 2007). The platinum ion is a soft ligand and makes favourable interactions with S atoms present in proteins through free methionine and cysteine side chains, forming strong bonds (Zimmermann & Burda, 2010; Hahn et al., 2001). However, HEWL and superoxide dismutase have free histidine (His) residues which can coordinate the Pt ion through an N atom on the imidazole ring (Calderone et al., 2006; Casini, Mastrobuoni et al., 2007; Casini et al., 2008).

HEWL is a model protein for which the crystallization conditions and three-dimensional structure have been well documented. A study by Li (2006) reported a number of metal ions (Ca$^{2+}$, Ni$^{2+}$ and Mn$^{2+}$) bound to the active site of HEWL. A major use of DMSO is as a vehicle for drug administration (Dearman et al., 1998; Abedini et al., 2004; Axanova et al., 2005; Peaston & Maddison, 2006; Huang et al., 2007). Cisplatin has previously been studied with DMSO in order to determine whether its anticancer effects are enhanced owing to the penetrant nature of DMSO (Pommier et al., 1988; Baliga et al., 1998). A study by Mickey et al. (1989) showed that using low doses of cisplatin resulted in no cytotoxic effect on a tumour cell; however, on the addition of DMSO it caused severe toxicity. Repeating this with the addition of furosemide resulted in the killing of tumour cells without the toxicity problems. An NMR study by Fischer et al. (2008) revealed the formation of a cisplatin–DMSO adduct within 15 min of dissolving cisplatin in DMSO and complete conversion within 1 h, as observed previously (Kerrison & Sadler, 1977). The DMSO molecule displaces one of the Cl atoms of cisplatin, as the Pt-S interactions are favourable. This adduct can readily cross a membrane and accumulate in cells; however, its activity towards double-stranded DNA is abolished owing to the steric hindrance caused by DMSO, leading to increased toxicity. These findings lead to the conclusion that cisplatin should not be dissolved in DMSO for anticancer treatments owing to the rapid formation of the cisplatin–DMSO adduct (Kerrison & Sadler, 1977; Feng et al., 2004; Fischer et al., 2008).

Casini, Mastrobuoni et al. (2007) published an important communication regarding the structural details of the binding of cisplatin to HEWL at pH 6.5 using the crystal-soaking method and subsequent X-ray diffraction analysis. Our study builds on the work of Casini, Mastrobuoni et al. (2007), but uses cocrystallization rather than the crystal-soaking method. By harnessing the cocrystallization method, we have been able to crystallographically characterize carboplatin binding to HEWL. We have evaluated both aqueous and DMSO crystallization conditions. Furthermore, in order to assess the possible impact of differing cryoprotectants, two cryoprotectants were used in this study (glycerol and Paratone). The results obtained with both glycerol and Paratone in aqueous conditions confirm that they both mimic sugar binding in the active site. To assess whether cisplatin binding to His15 might cause a change in the active site, thus inhibiting the enzymatic reaction, HEWL was cocrystallized with its natural substrate N-acetylglucosamine (NAG) and cisplatin to confirm whether any differences occurred upon binding of cisplatin. Overall, this extensive set of crystallization and cryoprotectant conditions, including the absence and presence of DMSO, allowed a systematic evaluation of the binding behaviour of cisplatin and carboplatin to HEWL as a test protein via detailed X-ray crystal structure characterizations.

![Figure 1](https://example.com/figure1.png)

**Figure 1**
Chemical structures of cisplatin and carboplatin.
Table 1
The cocrytallization conditions used for all crystals are based on the general conditions published by Blundell & Johnson (1976) using the batch crystallization method.

The protein concentration and in two cases temperature were varied to yield more optimal crystals both in number and quality; associated with this, the concentrations of cisplatin or carboplatin were varied to keep the molar ratio of each heavy-atom compound to protein the same.

<table>
<thead>
<tr>
<th>HEWL cocrytallization</th>
<th>HEWL with cisplatin or carboplatin in aqueous medium pH 4.7</th>
<th>HEWL with cisplatin or carboplatin in DMSO medium pH 4.7</th>
<th>HEWL with cisplatin in DMSO medium pH 6.5</th>
<th>HEWL with NAG</th>
<th>HEWL with NAG and cisplatin</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEWL</td>
<td>49 mg (1.6 mM)</td>
<td>49 mg (3.2 mM)</td>
<td>20 mg (0.6 mM)</td>
<td>49 mg (1.6 mM)</td>
<td>20 mg (0.6 mM)</td>
</tr>
<tr>
<td>Cisplatin</td>
<td>3 mg (5 mM)</td>
<td>3 mg (10 mM)</td>
<td>1.1 mg (1.8 mM)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Carboplatin</td>
<td>3.7 mg (5 mM)</td>
<td>3.7 mg (10 mM)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DMSO</td>
<td>—</td>
<td>75 µl (1 mM)</td>
<td>—</td>
<td>75 µl (1 mM)</td>
<td>75 µl (1 mM)</td>
</tr>
<tr>
<td>Sodium acetate (0.04 M)</td>
<td>1 ml</td>
<td>462.5 µl</td>
<td>462.5 µl</td>
<td>1 ml</td>
<td>462.5 µl</td>
</tr>
<tr>
<td>10% sodium chloride</td>
<td>1 ml</td>
<td>462.5 µl</td>
<td>462.5 µl</td>
<td>1 ml</td>
<td>462.5 µl</td>
</tr>
<tr>
<td>NAG</td>
<td>—</td>
<td>—</td>
<td>117 mg (260 mM)</td>
<td>40 mg (90 mM)</td>
<td>—</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>295</td>
<td>295</td>
<td>277</td>
<td>295</td>
<td>277</td>
</tr>
<tr>
<td>Time for crystallization (d)</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2
Data-collection strategy used for the Bruker PLATINUM135 CCD detector.

<table>
<thead>
<tr>
<th></th>
<th>Swing around $\varphi$ (°)</th>
<th>Sweep (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4dd1</td>
<td>20</td>
<td>307.0°</td>
</tr>
<tr>
<td>4dd4</td>
<td>−15.5</td>
<td>202.0°</td>
</tr>
<tr>
<td></td>
<td>−8.0</td>
<td>38.0°</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>35.0°</td>
</tr>
<tr>
<td></td>
<td>22.0</td>
<td>52.0°</td>
</tr>
<tr>
<td></td>
<td>22.0</td>
<td>63.0°</td>
</tr>
<tr>
<td>4dd6</td>
<td>0</td>
<td>360.0°</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>360.0°</td>
</tr>
<tr>
<td>4dd7</td>
<td>0</td>
<td>31.0°</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>180.0°</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>360.5°</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>360.5°</td>
</tr>
<tr>
<td>4ddc</td>
<td>0</td>
<td>360.0°</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>360.0°</td>
</tr>
</tbody>
</table>

2. Materials and methods

2.1. Materials

Cisplatin, HEWL, glycerol and NAG were purchased from Sigma–Aldrich UK. Carboplatin was purchased from Calbiochem USA, DMSO from Cambridge Isotope Laboratory, UK and sodium chloride (NaCl) from Fisher Scientific, UK. Sodium acetate and acetic acid (AnalaR grade) were purchased from BDH, USA. Paratone and silicone oil were purchased from Molecular Dimensions, UK.

2.2. Cocrytallizations

Table 1 summarizes the conditions used in each case; glycerol and Paratone were used as cryoprotectants for all crystals except for HEWL with NAG and HEWL with NAG and cisplatin, for which silicone oil was used.

2.3. X-ray data collection, structure solution and refinement

Of the 11 crystals studied, data for six (4dd0, 4dd2, 4dd3, 4dd9, 4dda and 4ddb; the crystals are named according to the PDB entry for the structure obtained) were collected on a Rigaku R-AXIS IV image-plate detector and data for the remaining five (4dd1, 4dd4, 4dd6, 4dd7 and 4ddc) were collected on a Bruker PLATINUM135 CCD detector both using an X-ray wavelength of 1.5418 Å and a sample temperature of 100 K.

The R-AXIS IV image-plate detector was positioned between 100 and 200 mm away from the crystal (Tables 3, 4 and 5) with an exposure time of between 3 and 10 min for each diffraction image; a 360° rotation range was measured at an angular width of 1°, with the exceptions of 4dd3 and 4dda, for which 272° and 181° rotation ranges were measured, respectively.

For the Bruker PLATINUM135 CCD detector, a strategy program was used to obtain the maximum information from each crystal in the measurement time. Table 2 summarizes the detector swing angles and the sweep of data collected for each data set.

All data sets were processed using EVAL15 (Schreurs et al., 2010) to rule out any software dependencies (Tables 3, 4 and 5) such as on average atomic $B$ factors. Each data set was also processed via MOSFLM (Leslie & Powell, 2007); the R-AXIS data sets via its internal software $d^*$TREK and the Bruker data sets via PROTEUM2.

All structures were solved using molecular replacement with Phaser (McCoy et al., 2007) and restrained refinement with TLS in REFMAC5 (Vagin & Tepljakov, 2010) in CCP4i, using the lysozyme structure 2w1y as the molecular search model (Cianci et al., 2008). Model building, adjustment and refinement were carried out using the Coot (Emsley & Cowtan, 2004) molecular-graphics program and REFMAC5 (Vagin & Tepljakov, 2010) in CCP4i, respectively. Ligand-binding occupancies were calculated using SHELXTL (Sheldrick, 2008). Crystallographic and refinement parameters are summarized in Tables 3, 4 and 5 for all data sets.

It is difficult to offer a consistent resolution-cutoff criterion owing to differing detector apertures and average data $I/\sigma(I)$ effects. Instead, in a companion paper (Tanley et al., 2012) raw diffraction data images will be made available in addition to the PDB depositions listed in this article via their DOIs and our local University data archive. In any case, the resolution
Data reduction

Refinement

Table 3

X-ray crystallographic data processed via EVAL15 and refinement statistics for all cisplatin cocrystallizations in aqueous and DMSO media using glycerol or Paratone as the cryoprotectant at pH 4.7 and 6.5.

Table 4

X-ray crystallographic data processed via EVAL15 and refinement statistics for all carboplatin cocrystallizations in aqueous and DMSO media with glycerol or Paratone as the cryoprotectant at pH 4.7.


limits are reasonably similar, except for two crystals (4dda and 4ddb) for which the resolution is limited by crystal disorder. In the highest resolution shell the average unmerged I/σ(I) varies between 0.9 and 3.5.
Table 5
X-ray crystallographic data processed via EVAL15 and refinement
statistics for HEWL cocryrstallized with NAG and HEWL cocryrstallized
with NAG, cisplatin and 7.5% DMSO using silicone oil as cryoprotectant.

<table>
<thead>
<tr>
<th></th>
<th>HEWL/NAG (4dda)</th>
<th>HEWL, NAG, cisplatin and 7.5% DMSO (4dcd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4 2 2</td>
<td>P2 1 2, 2, 2</td>
</tr>
<tr>
<td>a = b = c (Å)</td>
<td>78.37</td>
<td>77.94</td>
</tr>
<tr>
<td>c = 6 3,6,8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = y = 9 0,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecular mass (Da)</td>
<td>14700</td>
<td>14700</td>
</tr>
<tr>
<td>Detector</td>
<td>R-AXIS IV</td>
<td>Bruker</td>
</tr>
<tr>
<td>Crystal-to-detector distance (mm)</td>
<td>135</td>
<td>60†</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>49543</td>
<td>329619</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>4120</td>
<td>21884</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>19.19–2.40</td>
<td>19.16–1.80</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>12.1(12.8)</td>
<td>15.1(5.6)</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>29.5</td>
<td>16.2</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R factor/Rmerge (%)</td>
<td>20.0/28.5</td>
<td>21.8/25.5</td>
</tr>
<tr>
<td>R factor all (%)</td>
<td>20.4</td>
<td>21.9</td>
</tr>
<tr>
<td>R.m.s.d. bonds (Å)</td>
<td>0.013</td>
<td>0.018</td>
</tr>
<tr>
<td>R.m.s.d. angles (°)</td>
<td>2.38</td>
<td>1.82</td>
</tr>
<tr>
<td>Ramachandran values (%)</td>
<td>97.64</td>
<td>96.85</td>
</tr>
<tr>
<td>Additional allowed</td>
<td>1.57</td>
<td>3.15</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0.79</td>
<td>0</td>
</tr>
</tbody>
</table>

† The distance from the face of the detector to the phosphor plane is an additional 6.6 mm.

3. Results

3.1. Binding of cisplatin and carboplatin in DMSO media

3.1.1. Cisplatin and carboplatin details. In both the cisplatin and carboplatin cases, one molecule was coordinated by the N₈ and N₁ atoms (which we call ‘right-hand’ and ‘left-hand’ sites) of the imidazole ring of His15. This was true for both cryoprotectants (Fig. 2). In the case of carboplatin only the Pt ion and the two N atoms could be modelled based on the electron-density maps (Figs. 2c and 2d); the cyclobutane-
dicarboxylate (CBD) moiety could not be modelled owing to an absence of electron density. The CBD moiety acts as a leaving group, activating carboplatin to bind to DNA and thus giving rise to the possibility of CBD release upon binding of carboplatin to the N atoms of His15. With regard to cisplatin binding (Figs. 2a, 2b and 2e), the number of Cl and N atoms bound to the Pt ion varies based on the electron density observed for each data set.

To compare the crystal-soaking method used by Casini, Mastrobuoni et al. (2007) and the cocystallization method used in this study, HEWL was cocystallized with cisplatin at pH 6.5 in DMSO conditions. However, owing to the differences in crystal-growth time, a direct comparison is not suggested. Fig. 2(e) confirmed that one molecule of cisplatin was again coordinated to both the N₈ and N₁ atoms of the imidazole ring of His15 at pH 6.5, in contrast to Casini, Mastrobuoni et al. (2007) who only observed binding at the N₈ atom of His15 in their crystal structure, although their ESI-MS study indicated a doubly-platinated species. The different pH conditions used (pH 4.7 and 6.5) resulted in no differences in protein conformation in the presence of Pt ions based on using LSQKAB in CCP4i to orient the models in the same space (results not shown). Binding to His15 was confirmed using both Fₒ – Fc and anomalous difference electron-density maps (Fig. 2), with large peaks observed for the Pt ions. The occupancies observed for cisplatin and carboplatin binding to the N₈ and N₁ atoms of His15 were in the range 32–82% for the EVAL15-processed data (Table 6). Each data set was also processed via MOSFELM and either d*TREK (R-AXIS IV detector) or PROTEUM2 (Bruker detector); the occupancies for the Pt ions are also given in Table 6 along with the mean and standard deviations averaged across the results from these processing programs. The mean value range of 47–85% is higher than the results of Casini, Mastrobuoni et al. (2007), who reported 50% occupancy for cisplatin at the N₈ atom of His15 (PDB entry 2i6z) and <15% occupancy for carboplatin binding to His15 (no coordinate file was deposited in the PDB for this case). This would be quite reasonable as the cocystallization time that we have used allows a longer chemical reaction with the histidine, resulting in higher occupancy. Of course, the ‘microenvironment’ of the His15 side chain in the preformed crystal is not the same as that in solution. Comparing our cisplatin/DMSO structure with Paratone with 2i6z (Fig. 3), the N₈ atom is shifted by 0.40 Å, which could facilitate binding of the second Pt ion in the cocystallization method which is not seen in the soaking method. However, only a 2σ confidence level is estimated from the coordinate errors and the Cruickshank DPI values (0.40 ± 0.18 Å), meaning that this shift is not significant at a 3σ level as a reason for the binding of the second Pt ion. The Pt-to-nitrogen distances in Figs. 2 and 7 are shown in Table 7, with the precisions of the bond distances shown in parentheses. The means and standard deviations of these bond lengths have also been calculated using the refined structures based on data from EVAL15, MOSFELM and either d*TREK or PROTEUM2. There is a slight difference between the Pt-to-nitrogen distances of ~2.3–2.4 Å in Table 7 and those of 2.1 Å in Casini et al. (2007) and 2.1 Å in Calderone et al. (2006), but the standard uncertainties evident in Table 7 are too large to confirm a real difference of 0.2–0.3 Å.

Owing to the difference that can be seen in Table 6 for the overall average B factors from each of the four software programs, a selection of ten representative HEWL structures at 100 K from the PDB were made (Supplementary Table S1†). The mean and standard deviation of the overall average B factors from these ten structures were 17.2 ± 3.42 Å². To compare the four processing programs used here, the average B factors were d*TREK, 33.65 ± 4.12 Å²; PROTEUM2, 19.16

† Supplementary material has been deposited in the IUCr electronic archive (Reference: MN5006). Services for accessing this material are described at the back of the journal.
Table 6
Occupancies (shown as percentages) of cisplatin and carboplatin binding to the N\(^{4}\) and N\(^{9}\) atoms of His15 calculated from SHELXTL for all data sets processed via either d*TREK or PROTEUM2, MOSFLM and EVAL15.

The overall average B factor for each structure is given for all three processing programs along with the mean and standard deviation for the occupancy values at each binding site between the different processing programs. The total occupancies for both binding sites and the \(\sigma\) on this are also given (far right column).

\[
\begin{array}{cccccccc}
<table>
<thead>
<tr>
<th>d*TREK/PROTEUM2</th>
<th>MOSFLM</th>
<th>EVAL15</th>
<th>Mean \pm s.d.</th>
<th>Total occupancy/(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N^4)</td>
<td>(N^9)</td>
<td>B factor ((\AA^2))</td>
<td>(N^4)</td>
</tr>
<tr>
<td>4dd4</td>
<td>67</td>
<td>52</td>
<td>23.6</td>
<td>66</td>
</tr>
<tr>
<td>4dd6</td>
<td>57</td>
<td>48</td>
<td>17.9</td>
<td>69</td>
</tr>
<tr>
<td>4dd7</td>
<td>67</td>
<td>54</td>
<td>15.8</td>
<td>64</td>
</tr>
<tr>
<td>4dd9</td>
<td>65</td>
<td>50</td>
<td>29.9</td>
<td>74</td>
</tr>
<tr>
<td>4ddb</td>
<td>88\†</td>
<td>60\†</td>
<td>42.1\†</td>
<td>‡</td>
</tr>
<tr>
<td>4ddc§</td>
<td>65</td>
<td>63</td>
<td>16.2</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\end{array}
\]

\† During processing in d*TREK only 81 images were integrated. \(R_{merge}\) values greater than 0.25 per image were rejected. EVAL15 processing did not have a problem with these images, so all 360\° of data were used for refinement. \† MOSFLM also had a problem with the \(R_{merge}\) values for 4ddb and when removing images based on this criterion SCALA in CCP4 failed owing to too few reflections in the file. § 4ddc was solved in the orthorhombic space group \(P_{2_1}2_12\), via EVAL15 with two molecules in the asymmetric unit and hence with four molecules of cisplatin bound. However, the tetragonal space group \(P_{4_1}2_12\) was given by MOSFLM and PROTEUM2.

Table 7
The platinum-to-histidine imidazole N distances (\(\AA\)) from refinement using the data sets processed via EVAL15, MOSFLM and either d*TREK or PROTEUM2 for the R-AXIS IV or Bruker data sets, respectively.

The precisions of these distances are indicated in parentheses based on the Cruickshank DPI (Cruickshank, 1999) coordinate errors for each pair of atoms (Pt and N) from each model refinement. The means and standard deviations between the software processing programs for the same data set are also supplied.

\[
\begin{array}{cccccccc}
<table>
<thead>
<tr>
<th>EVAL15</th>
<th>MOSFLM</th>
<th>d*TREK/PROTEUM2</th>
<th>Mean \pm s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N^4)</td>
<td>(N^9)</td>
<td>B factor ((\AA^2))</td>
</tr>
<tr>
<td>4dd4</td>
<td>2.37 (0.24)</td>
<td>2.29 (0.23)</td>
<td>2.40 (0.19)</td>
</tr>
<tr>
<td>4dd6</td>
<td>2.36 (0.62)</td>
<td>2.33 (0.59)</td>
<td>2.31 (0.41)</td>
</tr>
<tr>
<td>4dd7</td>
<td>2.28 (0.19)</td>
<td>2.29 (0.19)</td>
<td>2.34 (0.23)</td>
</tr>
<tr>
<td>4dd9</td>
<td>2.33 (0.22)</td>
<td>2.33 (0.21)</td>
<td>2.35 (0.49)</td>
</tr>
<tr>
<td>4db</td>
<td>2.54 (1.33)</td>
<td>2.38 (1.50)</td>
<td>‡</td>
</tr>
<tr>
<td>4ddc_A</td>
<td>2.67 (0.48)</td>
<td>2.26 (0.40)</td>
<td>2.10 (1.06)</td>
</tr>
<tr>
<td>4ddc_B</td>
<td>2.77 (0.42)</td>
<td>2.29 (0.38)</td>
<td>—</td>
</tr>
</tbody>
</table>
\]

\† MOSFLM processing had a problem with the \(R_{merge}\) values for 4ddb and when removing images based on this criterion SCALA in CCP4 failed owing to too few reflections in the file.

Figure 2
\(F_o - F_c\) and anomalous difference density maps around His15 for the cases of cisplatin and carboplatin with DMSO. Anomalous difference density (orange) is shown at 3\(\sigma\) in (a)–(d) and 2.5\(\sigma\) in (e) and \(F_o - F_c\) density (blue) is shown at 3\(\sigma\). (a, b) Two molecules of cisplatin are bound to His15 at pH 4.7 with glycerol and Paratone used as the cryoprotectant, respectively. (c, d) Two molecules of carboplatin are bound to His15 at pH 4.7 with glycerol and Paratone used as the cryoprotectant, respectively. (e) Two molecules of cisplatin are bound to His15 at pH 6.5. The Pt, N and Cl atoms of cisplatin and carboplatin are labelled. The Pt–N\(^{4}\) and Pt–N\(^{9}\) distances are given in Table 7.
3.58 Å²; MOSFLM, 24.1 ± 10.2 Å²; EVAL15, 18.2 ± 4.7 Å². Obviously, these values may include contributions from crystallization and freezing in addition to the intrinsic flexibility of lysozyme, but they are helpfully indicative.

3.1.2. DMSO-binding details. A DMSO molecule was bound at lysozyme saccharide-binding site C of the active site of the enzyme in all crystals tested (Fig. 4); it was bound by Trp63 and Trp108 and confirmed by the presence of anomalous difference density for the S atom. This binding site for DMSO coincides with the deposited coordinates (PDB entry 2i6z) of Casini, Mastrobuoni et al. (2007).

3.2. Absence of binding of cisplatin and carboplatin in aqueous media

As DMSO was bound in the active site of HEWL, mimicking sugar binding, aqueous conditions were used to remove the DMSO inhibitory effect. The aim was to determine whether carboplatin, with its similar shape to a sugar molecule, might mimic sugar binding in the active site. However, no cisplatin or carboplatin binding was observed at the enzyme active site; moreover, surprisingly, no binding was observed at His15 at these crystallization timescales of 4–8 days. There was appropriately shaped $F_o - F_c$ electron density at the enzyme active site for the two cryoprotectants used: glycerol (Figs. 5a and 5c) and Paratone (Figs. 5b and 5d). As a further check, we noted the absence of anomalous difference density (i.e. that might have arisen from a Pt atom). Therefore, in the case with DMSO the DMSO must somehow facilitate cisplatin and carboplatin binding to the N atoms of His15. Using LSQKAB in CCP4i, it was noted that the protein conformation is basically the same in the ‘with DMSO conditions’ and in the ‘without DMSO conditions’; thus, the addition of DMSO does not cause any changes in protein conformation (results not shown).

3.3. Cocrystallization of HEWL with NAG and cisplatin

The cocrystallization of HEWL with its natural substrate NAG and cisplatin was carried out to assess whether cisplatin binding caused any effect upon NAG binding in the active site. HEWL was first cocrystallized with NAG only to confirm its binding in the enzyme active site. A 150-fold molar excess was used, as monosaccharide NAG usually has a greater affinity for site C rather than site D (Perkins et al., 1981).

NAG binding was observed at enzyme active site D with an occupancy of 77% as calculated in SHELXTL (Fig. 6). Monosaccharide NAG usually has a greater affinity for site C rather than site D (Perkins et al., 1981).

Figure 3 Overlay of His15 from 2i6z (purple) and our cisplatin/DMSO/Paratone structure (orange) at pH 4.7. The left-hand binding site (N) has the Pt ion in the same place and the right-hand binding site has the N’ atom shifted by 0.40 Å.

Figure 4 DMSO binding in active site C of HEWL. (a, b) Cocrystallization of cisplatin with glycerol (a) and Paratone (b) as the cryoprotectant at pH 4.7. (c, d) Cocrystallization of carboplatin with glycerol (c) and Paratone (d) as the cryoprotectant at pH 4.7. (e) Cisplatin cocrystallization at pH 6.5 with Paratone as the cryoprotectant. The two catalytic residues Asp52 and Glu35 in site D and Trp63 and Trp108 which line active site C are shown. The $F_o - F_c$ density maps (blue) at 3σ and anomalous difference density maps (orange) at 2.5σ are also shown. DMSO binding is confirmed by the presence of an anomalous difference density map peak, albeit at 2.5σ.
A separate cocrystallization experiment was conducted using HEWL with NAG together with the addition of 7.5% DMSO and a threefold molar excess of cisplatin. HEWL crystallized in the orthorhombic space group $P2_12_12_1$, with two molecules of HEWL in the asymmetric unit. Two molecules of cisplatin were again observed bound to either side of His15 in chain $A$, with occupancies of 32 and 55% for the $N^2$ and $N^1$ binding sites, respectively, and in chain $B$, with occupancies of 40 and 43% for the $N^2$ and $N^1$ binding sites, respectively (one of the binding sites is shown in Fig. 7). Based on the above results, we may expect either NAG or DMSO, or both, to be bound in the enzyme active site. In this case, the binding of a DMSO molecule is observed (Fig. 8) based on anomalous difference density of 2.5 $\sigma$ in the active site. [In Fig. 8 DMSO is bound in site $D$. It is in a different place than in Fig. 4 (site $C$). However, we are not sure why this is.]

4. Discussion
4.1. The initial experimental concepts for these studies

Our initial impetus and interest in embarking on this study arose from the idea that carboplatin could bind to the active site of HEWL, thus mimicking sugar binding (Li, 2006), and if so it would also be an interesting toxicity side effect of this important drug treatment used in veterinary practice. Binding to the active site could lead to the use of carboplatin as a cofactor in the cocrystallization of sugar-binding proteins and structure determination using isomorphous replacement owing to the anomalous scattering power of the Pt ion. Obviously, these ideas also connected with the use of cisplatin and carboplatin in medicine and veterinary practice, as carboplatin and cisplatin are similarly acting anticancer chemotherapeutics.

We found that Casini, Mastrobuoni et al. (2007) had published an X-ray crystallographic study of HEWL crystals soaked with cisplatin (PDB entry 2i6z) and had also investigated carboplatin via the crystal-soaking method, whilst first using MS to confirm the binding of either platinum complex to the protein. The results obtained from their structural analysis confirmed the binding of cisplatin to His15 of HEWL with an occupancy of around 50% at one $N$ atom of the imidazole ring. No crystal structure was reported for the HEWL carboplatin crystal-soaking experiment, but ICP-OES data showed very weak binding (<15%). However, MS studies indicated a clear mass change upon the binding of carboplatin and indicated a doubly-platinated HEWL structure for both the cisplatin and...
the carboplatin cases. PDB entry 2i6z also showed a DMSO molecule bound in the active site, mimicking sugar binding.

We have been able to build on the important findings of Casini, Mastrobuoni et al. (2007) in the results presented in this article. The method of cocryrstallization of HEWL with cisplatin or carboplatin was used in order to avoid any possible crystal-lattice hindrance that may have affected the carboplatin crystal soaking. The use of DMSO media was carried out for a direct comparison and the use of aqueous conditions was added in a further set of experiments to avoid the possible chemical complications that may have occurred with DMSO in confirming whether carboplatin (or cisplatin) might bind in the enzyme active site. A potential further chemical complication is the choice of cryoprotectant, as it is well known that glycerol has specific binding sites in proteins, including mimicking sugars, which is also of interest here. The cryoprotectant used by Casini, Mastrobuoni et al. (2007) was not specified. In our studies both glycerol and Paratone were used, and finally also silicone oil in the studies of lysozyme with NAG. We included the cocryrstallization of HEWL with its natural substrate NAG in our range of chemical conditions to be varied in order to confirm binding in the active site. These data were used as a reference for comparison with HEWL cocryrstallized with NAG and cisplatin in DMSO conditions to confirm whether cisplatin binding caused any effect on NAG binding to the active site.

4.2. Chemical rationalization of cisplatin or carboplatin binding to both histidine N atoms

In the DMSO crystallization conditions, we found that two molecules of cisplatin or carboplatin were coordinated to one histidine moiety: His15. The two metal centres are linked to the N° and N’ atoms of the imidazole ring (Fig. 2) at pH 4.7. The cisplatin study was repeated at pH 6.5 using cocryrstallization. This structural unit (one histidine and two platinum centres) means that in the imidazole of this histidine the usual N-hydrogen of such a residue is absent and that both N atoms are sp2-hybridized with nitrogen lone pairs in the plane of the imidazole ring: effectively, it is an imidazolyl anion. (The ‘imidazolyl anion’ is the same as ‘imidazolate’; see below.) This then provides two N atoms at which a metal centre can bind. The loss of this H atom (as a proton) is made possible by the crystallization conditions used, which contained chloride and acetate ions that could remove the N-hydrogen. The removal of the N-hydrogen would be facilitated by coordination of the imine N atom of the histidine to platinum. The situation is summarized in Fig. 9. This is a plausible explanation for the high occupancies seen for the binding of both carboplatin and cisplatin to either N atom of His15 at both pH 4.7 and 6.5 (Table 6). The notion of His15 conformational flipping in the structure as a static disorder effect, meaning that only one cisplatin/carboplatin molecule has chemically bound, can be ruled out owing to the fact that a C atom would now have to be coordinated to the Pt ion, which it obviously is not.

In solution at a pH similar to the pK°, 6.0–6.3, about equal amounts of the ionic protonated state and the deprotonated state of histidine occur, the latter possibly in two tautomeric forms a and b (Farr-Jones et al., 1993; Fig. 10). Thus, a further possible explanation for the two observed Pt sites is provided by these two tautomeric forms, allowing either N° or N’ to participate in the interaction with Pt.

The imidazolate ion mentioned above is described in Rhodes (2005) (form D; Fig. 11).

A positively charged platinum moiety would thus be a favourable situation for the histidine to have a charge of −1, thus stabilizing the imidazolate anion.

4.3. Estimations of platinum occupancies and their associated standard uncertainties

The occupancy estimates calculated using SHELXTL at this diffraction resolution are without standard uncertainties. The uncertainty values for these occupancies are generally regarded as probably being around ±5%. Table 6 gives the mean values, the summed occupancies and the standard uncertainties for the N° and N’ binding sites for each crystal grown in DMSO medium based on reprocessing of the same data set using three different programs (d*TREK or PROTEUM2, MOSFLM and EVAL15). The standard uncertainties for the N° and N’ binding sites for each crystal grown in DMSO medium based on reprocessing of the same data set using three different programs (d*TREK or PROTEUM2, MOSFLM and EVAL15). The standard uncertainties for the N° and N’ binding sites for each crystal grown in DMSO medium based on reprocessing of the same data set using three different programs (d*TREK or PROTEUM2, MOSFLM and EVAL15). The standard
uncertainties were recognize the orthorhombic symmetry; all other standard owing to problems in rejecting images based on MOSFLM values were used as this data set was not processed 4ddb also has a larger standard uncertainty, but only two large \( R \) atoms with full occupancy compared against the absolute latter are not expected to be large since the bulk of the protein software include random errors but not systematic errors. The data processing with the different diffraction data-processing uncertainties in the Pt-atom occupancies derived from repeat data processing with the different diffraction data-processing software include random errors but not systematic errors. The latter are not expected to be large since the bulk of the protein atoms with full occupancy compared against the absolute atomic scattering factors in the SHELXL refinement create the required absolute scale. As the summed occupancies are all larger than 1.0 (Table 6) it seems that the imidazolyl-anion hypothesis mentioned above is most likely, rather than a chemical resonance effect. It is only 4ddc which gives standard uncertainties greater than ±5 and this is a consequence of the fact that the MOSFLM and PROTEUM2 processing did not recognize the orthorhombic symmetry; all other standard uncertainties were \( \sim \pm 5\% \). The occupancy at the \( N^o \) atom of 4ddc also has a larger standard uncertainty, but only two values were used as this data set was not processed via MOSFLM owing to problems in rejecting images based on large \( R_{merge} \) values. 4dd4, 4dd9 and 4dc also show a larger change in the overall average \( B \) factor across the three different processing programs; however, this software effect does not cause an effect on the estimation of the occupancy at each binding site in these crystals. The average \( B \) factors for PROTEUM2 (19.16 ± 3.38 Å\(^2\)) and EVAL15 (18.2 ± 4.7 Å\(^2\)) are comparable, whereas those for MOSFLM (24.1 ± 10.2 Å\(^2\)) and \( d^*TREK \) (33.65 ± 4.12 Å\(^2\)) are larger for our 11 diffraction data sets. Comparing the average \( B \) factors between the processing programs shows that there is indeed variation. The processing details from the different programs will be summarized in a second paper (Tanley et al., 2012).

4.4. DMSO, glycerol and Paratone binding details

Besides the cisplatin and carboplatin binding behaviour, the DMSO binding is of interest. A DMSO molecule was indeed observed bound in site \( C \) of the enzyme active site (Fig. 4), thus acting as a competitive inhibitor of HEWL. A second DMSO molecule was coordinated by Trp123, and these findings corroborate those of Jóhannesson et al. (1997). In the presence of NAG, DMSO still seems to bind at the enzyme active site, acting in a competitive manner (Fig. 8).

Aqueous crystallization conditions were used to remove the inhibitory effect seen with DMSO at the enzyme active site, but also to evaluate whether DMSO had an impact on binding of cisplatin or carboplatin to His15. A complication at the enzyme active site is that in both the glycerol and Paratone cases a molecule of cryoprotectant binds to the active site (Fig. 5) in the absence of DMSO. Paratone is a polymeric hydrocarbon, \((C_4H_8)_{17}\), but only the monomeric form \((C_4H_8)\) was observed to bind in the active site. The \( F_o - F_c \) density maps obtained with glycerol contained further unmodelled density corresponding to additional glycerol molecules bound to other sites on HEWL. In contrast, the \( F_o - F_c \) map obtained with Paratone contained less unmodelled density related to bound Paratone molecules, making the refinement steps less complicated; as a consequence of these findings, it would be advantageous to use Paratone over glycerol. However, silicone oil would also be advantageous to use as it was not observed to bind to HEWL when used as the cryoprotectant in the NAG cases.

Using aqueous conditions, within our studied timescales of up to eight days for binding, it was found that neither cisplatin nor carboplatin bound to His15. Therefore, addition of DMSO to the crystallization conditions increased the affinity of cisplatin and carboplatin to bind to His15 of HEWL. DMSO is a super-solvent that is used for targeted drug delivery as it is perceived to have no effect upon drug binding (Dearman et al., 1998; Abedini et al., 2004; Axanova et al., 2005; Peaston & Maddison, 2006; Huang et al., 2007). However, our results led to the conclusion that addition of DMSO causes binding of cisplatin and carboplatin to occur at a His residue. This therefore must be a cause for concern as it may cause an increase in the toxic effects of both cisplatin and carboplatin owing to their binding to proteins that they would not normally bind to in aqueous conditions. Fischer et al. (2008) indeed proposed that cisplatin should not be used in
conjunction with DMSO owing to the rapid reaction that occurs between the two and that results in the formation of a cisplatin–DMSO adduct which loses cytotoxicity towards tumour cells and has increased toxic side effects. In our crystallization conditions both cisplatin and carboplatin did not fully dissolve in 7.5% DMSO immediately, but after 12 h all of the cisplatin and carboplatin had dissolved; this was also observed in the aqueous conditions used. The cisplatin–DMSO adduct was not observed in any of the crystal structures presented here; based on our findings, direct coordination of DMSO to Pt might alter the compound reactivity in solution, not only with respect to HEWL binding.

4.5. Comparison with 2i6z

Compared with Casini, Mastrobuoni et al. (2007), who observed one cisplatin binding to His15 (PDB entry 2i6z), we find that two molecules of cisplatin and carboplatin bind to His15. This may simply result from the longer time afforded to the chemical reaction during the crystallization process. Of course, as remarked above, the ‘microenvironment’ of the His15 side chain in the preformed crystal is not the same as that in solution. As an alternative explanation, it is not obvious what chemical condition is different. Indeed, the MS study of Casini, Mastrobuoni et al. (2007) indicated double platination. It is also not obvious why cocrystallization might allow binding at each N atom of the platinum complex, as the solvent channel in the crystal seems to give open access to His15 in the soaking case (Casini, Mastrobuoni et al., 2007). We nevertheless overlayed our cisplatin–lysozyme crystal structure with 2i6z and observed a His15 side-chain shift. We can evaluate whether this is a significant shift as follows. The His residue is shifted by 0.40 Å for our coordinates versus PDB entry 2i6z (Fig. 3). The coordinate error of each atom is calculated using (1), via which the Crucikshank DPI values (Cruijckshank, 1999; Blow, 2002) can confirm whether the coordinate shift is significant.

\[
\text{coordinate error of atom} = \text{DPI} \times (B_{\text{atom}}/B_{\text{average}})^{1/2}. \quad (1)
\]

We use the coordinate error rather than the positional error of the atom as the shift of the atom is in one direction only. (The positional error takes into account the shift in three directions, with that error being \(3^{1/2}\) times larger than the coordinate error; Cruijckshank, 1999). The standard deviation of the calculated shift is determined using the coordinate error of the N atoms in both structures 1 and 2,

\[
\text{standard deviation} = [(\text{coordinate error 1})^2 + (\text{coordinate error 2})^2]^{1/2}. \quad (2)
\]

The value and standard uncertainty for the shift of the N atoms are 0.40 ± 0.18 Å. This gives a value of only just over 2\(\sigma\), or a 95.4% level of confidence, rather than the usually desirable 3\(\sigma\) or 99.7% level of confidence that this coordinate shift is significant.

4.6. NAG binding

NAG is the natural substrate of HEWL and usually binds in site C of the active site (Perkins et al., 1981). Our crystal structure of HEWL with NAG has NAG bound in the less favourable site D (Fig. 6). Upon cocrysalizing HEWL, NAG and cisplatin in DMSO medium, the observed \(F_o - F_c\) density in the active site covers site D (Fig. 8). However, this density is consistent with one molecule of DMSO binding, which is confirmed by the presence of anomalous difference electron density at 2.5\(\sigma\), meaning that DMSO competes with NAG to bind in the active site.

4.7. Comparison with other proteins binding cisplatin and implications for drug toxicity

Only a handful of other proteins have been studied for their affinity to bind cisplatin, including copper transporter (Arnesano & Natile, 2008; Crider et al., 2010), superoxide dismutase (Calderone et al., 2006; Casini et al., 2008), cytochrome c (Casini et al., 2006; Casini, Gabbiani et al., 2007), human albumin (Ivanov et al., 1998), ubiquitin (Hartinger et al., 2006), human copper chaperone (Boal & Rosenzweig, 2009) and glutathione (Zimmermann & Burda, 2010). These studies mostly used MS to confirm the binding of cisplatin to specific Met/Cys or His residues, with only the X-ray crystal structures of cisplatin bound to HEWL (Casini, Mastrobuoni et al., 2007), to a copper chaperone (Boal & Rosenzweig, 2009) and to superoxide dismutase being available (Calderone et al., 2006; Casini et al., 2008). The free His residue available in HEWL is not involved in the catalytic mechanism; hence, cocrystallizing cisplatin or carboplatin with proteins (e.g. chymotrypsin, trypsin) whose catalytic site requires a His residue would be of benefit to determine whether catalysis is inhibited upon binding of these platinum compounds, which could be important in further assessment of the toxic effects of these drugs. Using cocrystallization techniques with the addition of DMSO could pave the way for increased analysis of protein–platinum anticancer drug interactions. These will lead to a better understanding of the pharmacokinetics, biodistribution, resistance processes and toxicity of these metallo drugs (Calderone et al., 2006; Casini et al., 2006; Casini, Mastrobuoni et al., 2007). However, crystallography can only provide data for a nonphysiological environment, in which the presence of high-salt buffer concentrations, as well as a high concentration of Pt drugs, are not realistic. Therefore, crystallographic data should be handled with care and always integrated with other techniques, i.e. mass spectrometry.

5. Conclusions

Both cisplatin and carboplatin do not bind to HEWL in aqueous media, within our studied timescales of up to eight days for binding, but do upon addition of DMSO. The DMSO causes two molecules of either cisplatin or carboplatin to bind to both imidazole N atoms of His15, the only His residue in HEWL. To confirm the relative identities of the atoms in the binding moieties, anomalous scattering maps were calculated.
and were cross-checked with $F_o - F_c$ OMIT maps to prove where the binding of the Pt atoms had occurred. The use of EVAL15 for processing all of the diffraction data sets provided a consistent platform for our large ensemble of data sets for the various protein and platinum-compound model refinements with REFMACS and then SHELXTL. Platinum-occupancy $\sigma$ values were finally calculated using the results from three different diffraction data-processing programs.

The implications of these results are important as DMSO is a super-solvent that is widely used for the delivery of these and other drugs. Since our results show that addition of DMSO accentuates the binding of platinum anticancer drugs to HEWL, this effect must be allowed for in toxicity assessments of these drugs. Further studies involving proteins with histidine in the active site and an assessment of cisplatin or carboplatin binding will lead to the potential side effects of these important anticancer agents being further understood.

JRH is grateful to the University of Manchester for general support, to the ESPRC for a PhD studentship to SWMT, to the School of Chemistry for crystallization and computing facilities and to the Faculty of Life Sciences for X-ray diffraction facilities (the Rigaku system has recently relocated to the Manchester Interdisciplinary Biocentre, but was previously at the Michael Smith Building Macromolecular Facility instigated by colleagues Dr Jordi Bella and Dr Lydia Tabernero, who are also thanked). Thanks also go to Professor Peter Halling, Dr Michele Cianci, Professor John Joule and Dr Alfons Haedener for discussions.

References

Supplementary Materials

Table S5.1

A representative selection of HEWL structures from the PDB, at different pH values, with the overall mean and standard deviation of the B factors calculated as shown.

<table>
<thead>
<tr>
<th>PDB Code</th>
<th>B factor</th>
<th>Temperature data collected at</th>
<th>pH used</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GXR</td>
<td>21</td>
<td>100</td>
<td>6.5</td>
</tr>
<tr>
<td>3AB6</td>
<td>17</td>
<td>100</td>
<td>5.5</td>
</tr>
<tr>
<td>3RU5</td>
<td>15</td>
<td>100</td>
<td>4.0</td>
</tr>
<tr>
<td>2YDG</td>
<td>12</td>
<td>100</td>
<td>4.7</td>
</tr>
<tr>
<td>3RNX</td>
<td>21</td>
<td>100</td>
<td>3.8</td>
</tr>
<tr>
<td>3A8Z</td>
<td>15</td>
<td>100</td>
<td>4.5</td>
</tr>
<tr>
<td>2W1Y</td>
<td>17</td>
<td>100</td>
<td>4.2</td>
</tr>
<tr>
<td>2W1X</td>
<td>16</td>
<td>100</td>
<td>4.2</td>
</tr>
<tr>
<td>2W1M</td>
<td>23</td>
<td>100</td>
<td>4.2</td>
</tr>
<tr>
<td>2W1L</td>
<td>15</td>
<td>100</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Mean/standard deviation: 17.2 +/- 3.42
6. Chapter 6: Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals

Simon W.M. Tanley, Antoine M.M. Schreurs, John R. Helliwell and Loes M.J. Kroon-Batenburg

School of Chemistry, Faculty of Engineering and Physical Sciences, University of Manchester, Brunswick Street, Manchester, M13 9PL, England,

Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, Utrecht, 3584 CH, The Netherlands, and

Correspondence email: l.m.j.kroon-batenburg@uu.nl


Keywords: Data exchange; Data archiving; Metadata

SWMT grew all 11 crystals, collected datasets on all of them, processed these via MOSFLM, TREK and Proteum 2. AMMS processed the datasets via EVAL-15. SWMT refined all structures to an end point and created the table of statistics in the paper. JRH instigated the collaboration between the two groups and pushed the idea of this second paper as a continuation to the first paper. LKB wrote the paper with editing by JRH and SWMT.
Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals

Simon W. M. Tanley,* Antoine M. M. Schreurs,b John R. Helliwella and Loes M. J. Kroon-Batenburgb*

*School of Chemistry, Faculty of Engineering and Physical Sciences, University of Manchester, Bruswick Street, Manchester M14 9PT, UK, and bCrystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, Utrecht, 3584 CH, The Netherlands. Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

The International Union of Crystallography has for many years been advocating archiving of raw data to accompany structural papers. Recently, it initiated the formation of the Diffraction Data Deposition Working Group with the aim of developing standards for the representation of these data. A means of studying this issue is to submit exemplar publications with associated raw data and metadata. A recent study on the effects of dimethyl sulfoxide on the binding of cisplatin and carboplatin to histidine in 11 different lysozyme crystals from two diffractometers led to an investigation of the possible effects of the equipment and X-ray diffraction data processing software on the calculated occupancies and $B$ factors of the bound Pt compounds. 35.3 Gb of data were transferred from Manchester to Utrecht to be processed with EVAL. A systematic comparison shows that the largest differences in the occupancies and $B$ factors of the bound Pt compounds are due to the software, but the equipment also has a noticeable effect. A detailed description of and discussion on the availability of metadata is given. By making these raw diffraction data sets available via a local depository, it is possible for the diffraction community to make their own evaluation as they may wish.

1. Introduction

There is increasing interest in depositing or archiving raw data of scattering experiments with publication of structural papers. This interest in archiving raw data is common to all scientific fields, as highlighted in the ICSU SSCID Report (2011). International Union of Crystallography (IUCr) journals are leaders in the archiving of derived and processed data with crystal structure papers either with articles in Acta Crystallographica Sections B, C and E or in close linking with the PDB (Protein Data Bank; Berman et al., 2000) in the case of Acta Crystallographica Sections D and F. Detailed consideration is now being given to the benefits, and extra costs, of extending the data archiving paradigm to also now include raw data such as diffraction data images. The Diffraction Data Deposition Working Group has been set up by the IUCr, and a variety of reports and ongoing community feedback can be found at the IUCr forum devoted to this matter. Reasons for archiving raw data include to improve the record of science, to ensure the reproducibility and allow detailed checks of scientific data, to safeguard against fraud, and to allow reanalysis with future improved techniques. The digital object identifier for each data set underpinning a published paper at an archive local to where the data were measured is a plausible model to move these developments forward; at the University of Manchester this is being launched in September 2012 and will be available in 2013. As an interim measure a link is provided to a personally maintained web site (http://rawdata.chem.uu.nl). Cost requirements of the long-term professional stewardship of digital data storage and large bandwidth access are important issues, but a further requirement is the provision for a sufficient level of metadata, to allow future use of the data. This paper addresses the challenges and possibilities of exchanging raw data with associated metadata for data processing with non-native software, i.e. not associated directly with a given piece of measuring equipment, for example, from a given manufacturer.

In protein crystallography, X-ray diffraction data are often obtained from synchrotron beamlines that provide high-brilliance beams and rapid data collections. While the synchrotron installations gradually gained in performance, the development of home sources also continued. Currently, microfocus X-ray sources with matching multilayer optics and high-performance detectors are available and can compete with second-generation synchrotron beamlines. They can also be
technically appropriate and indeed, being local, highly convenient. Manufacturers of diffraction equipment for applications in macromolecular and chemical crystallography provide integrated software for designing data collection strategies and data processing. The end result of a diffraction experiment is a series of recorded diffraction data images. Metadata are contained in the headers of these image files or internally on a server computer. The manufacturer’s internal software usually takes care of necessary corrections, such as for detector non-uniformities by using a flood-field image or for distortion due to the fibre optic taper in CCDs. Detaching images from the server computer for processing with ‘alien software’ requires finding all necessary metadata. Several independent data processing software packages have been developed for protein crystallography during the past two to three decades, e.g. MADNES (Messerschmidt, 1986; Messerschmidt & Pflugrath, 1987), DENZO/HKL-2000 (Otwinowski & Minor, 1997), Mosflm (Leslie, 1999), XDS (Kabsch, 1988) and, most recently, EVAL (Schreurs et al., 2010). The first three have been largely optimized for synchrotron data and later adapted to read several commercial detector formats, and (partially) also the possible goniometer geometries. EVAL, by contrast, was developed primarily for use with Nonius equipment with a four-circle Kappa goniometer, while specific efforts were made to implement a large range of detector formats and goniometer geometries.

In a preceding paper (Tanley et al., 2012) the binding of cisplatin or carboplatin to histidine in a protein (lysozyme) was described, based on diffraction data of 11 crystals of different chemical compositions. These were measured on two different X-ray diffractometers, namely a Rigaku Micromax-007 generator with Cu rotating anode equipped with an R-AXIS IV image plate and Oscim confocal mirrors, and a Bruker MICROSTAR Cu rotating anode equipped with a CCD ‘Pt detector’ and HELIOS confocal optics (X8 PROTEUM2). Initially, data processing was done with the equipment’s internal software. After protein model refinement of the 11 crystal structures, the average isotropic atomic B factors seemed to vary systematically between the two diffractometers, suggesting a systematic deviation in the treatment of weak intensity reflections. Since the scientific goal was to determine metal-binding occupancies, a possible influence of systematic errors in the B factors needed to be ruled out and/or corrected for. This led to a joint undertaking of the Utrecht and Manchester structural chemistry research groups to yield processed quality data from just one suite of software, EVAL. The X-ray diffraction data obtained from the 11 lysozyme crystals had been measured using the two diffractometers and processed with four software packages (Tanley et al., 2012). Most of the crystals diffracted to 1.7 Å. In total the diffraction data images (the ‘raw’ data) transferred between Manchester and Utrecht occupied 35.3 Gb of disk space and the process took a few days. In this paper we will address the availability and use of metadata, specifically in the case of processing with EVAL. The results and observations obtained allow detailed informed comment on the archiving of such raw diffraction data and their reprocessing after and beyond their initial study with a variety of software programs.

### 2. Materials and methods

The 11 lysozyme crystals (hen egg white lysozyme, HEWL) co-crystallized with cisplatin or carboplatin, with dimethyl sulfoxide or under aqueous conditions including N-acetylglucosamine, along with different cryosolvents, were prepared as described in the previous paper (Tanley et al., 2012). Out of the 11 crystals studied, six (labelled 1–4, 10 and 11) were collected on a Rigaku R-AXIS IV image plate and the remaining five (5–9) were collected on a Bruker PLATINUM 135 CCD detector, both using an X-ray wavelength of 1.5418 Å and at a temperature at the crystal sample set to 100 K. The data collection strategies (Table 1) were automatically chosen by the integrated strategy software PROTEUM2 on the Bruker equipment, and data were collected simply by a 360° φ scan on the R-AXIS, while requiring a redundancy of at least 8.0.

### 3. Results

Each sequence of X-ray diffraction data images was processed with three different packages, namely the equipment’s software [either d*Trek (Pflugrath, 1999) or PROTEUM2 (Bruker, 2006)], Mosflm (Leslie, 1999) [using SCALA (Evans, 2006) for scaling] and EVAL [using SADABS (Sheldrick, 1996) for scaling]. Tables 2 and 3 list the diffraction data reflection intensity integration and the subsequent protein model refinement statistics. In the cases where the processed X-ray data and three-dimensional coordinate sets were deposited with the PDB, a PDB code is given.

---

#### Table 1

<table>
<thead>
<tr>
<th>Detector</th>
<th>Swing</th>
<th>Sweep</th>
<th>Rotation/frame</th>
<th>Generator filament current (mA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigaku R-AXIS IV</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>4dd0 (3txb)</td>
<td>100</td>
<td>0</td>
<td>360Φ</td>
<td>1</td>
</tr>
<tr>
<td>4dd2 (3txd)</td>
<td>120</td>
<td>0</td>
<td>360Φ</td>
<td>1</td>
</tr>
<tr>
<td>4dd3 (3txe)</td>
<td>120</td>
<td>0</td>
<td>272Φ</td>
<td>1</td>
</tr>
<tr>
<td>4dd9 (3txi)</td>
<td>120</td>
<td>0</td>
<td>361Φ</td>
<td>1</td>
</tr>
<tr>
<td>4dda (3txj)</td>
<td>135</td>
<td>0</td>
<td>181Φ</td>
<td>1</td>
</tr>
<tr>
<td>4ddb (3txk)</td>
<td>200</td>
<td>0</td>
<td>360Φ</td>
<td>1</td>
</tr>
<tr>
<td>Bruker PLATINUM&lt;sup&gt;135&lt;/sup&gt;</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>4dd1</td>
<td>60†</td>
<td>20</td>
<td>307.0Φ</td>
<td>0.5</td>
</tr>
<tr>
<td>4dd4 (3txf)</td>
<td>50†</td>
<td>−15.5</td>
<td>202.0Φ</td>
<td>0.5</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>−8.0</td>
<td>38.0Φ</td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>12.0</td>
<td>35.0Φ</td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>22.0</td>
<td>52.0Φ</td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>22.0</td>
<td>63.0Φ</td>
<td> </td>
</tr>
<tr>
<td>4dd6 (3txg)</td>
<td>60†</td>
<td>20</td>
<td>360.0Φ</td>
<td>0.5</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>20</td>
<td>360.0Φ</td>
<td> </td>
</tr>
<tr>
<td>4dd7 (3txh)</td>
<td>50†</td>
<td>0</td>
<td>31.0Φ</td>
<td>0.5</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>0</td>
<td>180.0Φ</td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>0</td>
<td>360.5Φ</td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>20</td>
<td>360.5Φ</td>
<td> </td>
</tr>
<tr>
<td>4ddc</td>
<td>50†</td>
<td>0</td>
<td>360.0Φ</td>
<td>0.5</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td>20</td>
<td>360.0Φ</td>
<td> </td>
</tr>
</tbody>
</table>

† Distance from face of the detector to the phosphor plane is an additional 6.6 mm.
3.1. Comparison of hardware

3.1.1. Image formats and image headers. An example of an R-AXIS IV image plate header (Fig. 1) shows the information EVAL extracts. After the arrow, EVAL's interpretation is given. The starting and ending rotation angles of the spindle axis were read (3Å/Å 0.0 0.0 1.0), as well as the goniometer starting angles (3Å/Å 0.0 0.0 0.0, for ω, χ and swing θ); the rotation direction was known to us a priori (see next section). The number of pixels (3000 × 3000) and pixel size (100 μm) for each diffraction image were read. The direct beam position on the detector is indicated by the beam centre

<table>
<thead>
<tr>
<th>Crystal</th>
<th>PDB code</th>
<th>Software</th>
<th>3txi</th>
<th>4ddh</th>
<th>3txd</th>
<th>4dd2</th>
<th>3txe</th>
<th>4dd3</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.size</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
</tr>
<tr>
<td>3</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
</tr>
<tr>
<td>4</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
</tr>
</tbody>
</table>

### Table 2

Statistics of the processed data from the Rigaku diffractometer.

Numbers in brackets are the 'outer resolution shell' values.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>PDB code</th>
<th>Software</th>
<th>3txb</th>
<th>4dd0</th>
<th>3txd</th>
<th>4dd2</th>
<th>3txe</th>
<th>4dd3</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.size</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
<td>EVAL</td>
<td>Mosfim</td>
</tr>
<tr>
<td>1</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
</tr>
<tr>
<td>2</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
<td>d*Trek</td>
<td>EVAL</td>
</tr>
</tbody>
</table>

### Notes

† Space group is P41212 for all data. ‡ Reflections contaminated with ice scattering were removed from the data using a de-ice procedure. § Processing with Mosfim did not succeed.

corrections are needed. Every diffraction image takes 18 Mb of disk space. When the images were received in Utrecht after network transfer from Manchester they were immediately compressed, using compress (from the ncompress package), to 3.8 Mb. ncompress is public domain software that uses the LZW (Lempel–Ziv–Welch) algorithm for lossless data compression (Welch, 1984). EVAL uses compress –d to unpack the images on the fly.

The header information of a Bruker PLATINUM CCD detector diffraction data image (Bruker format) is shown in Fig. 2. The starting angles (ANGLES: 0.0 358.75 0.0 0.0) of the goniometer in Eulerian space (swing 2θ, ω, φ and χ) are given and later in the header the crystal sample rotation axis is defined, ψ in this case (AXIS: 3). Pixel intensities are stored as 1 byte integers (IPXEL = 1), and if the number is 255, additional bytes will follow at the end of the image in an overflow table. If not too many overflows occur this is a very

efficient format. The baseline offset (from NEXP), which is needed to store negative numbers in 1 byte integers, has to be subtracted. The gain of the detector is derived from the detector position in pixels is given by 1024 pixels (binned mode) and the pixel size is 89.99μm. In EVAL the Kappa goniometer option is fully implemented, whereas it is not in Mosflm. We had to resort to integration of each scan (rows in Table 1) separately, and scaled them later with SCALA. Each image takes about 1 Mb of disk space, and 800 Kb when compressed, which indicates that the Bruker format is very space efficient.

Diffraction data image formats were kindly provided by Rigaku and Bruker during the development of EVAL in previous years. There is no way to extract data from the image file unless these formats are known. Even once this information is available, header information is often not comprehensive and unambiguous.

### 3.1.2. Metadata

The image format of the Rigaku image plate contains a binary header that did not provide all the information needed. In fact the layout of the goniometer axes information needed. In fact the layout of the goniometer axes is one of the most laborious problems to deal with when implementing data processing for an unknown goniometer. A field in the header for the orientation of the spindle axis is reserved but did not provide all the resolution shell' values.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>5</th>
<th>PROTEUM2</th>
<th>EVAL</th>
<th>Mosflm</th>
<th>6</th>
<th>PROTEUM2</th>
<th>EVAL</th>
<th>Mosflm</th>
<th>7</th>
<th>PROTEUM2</th>
<th>EVAL</th>
<th>Mosflm</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code</td>
<td>Software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4dd1</td>
<td>EVAL</td>
<td>Mosflm</td>
<td>4dd4</td>
<td>EVAL</td>
<td>Mosflm</td>
<td>4dd6</td>
<td>EVAL</td>
<td>Mosflm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit cell†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>78.78</td>
<td>78.88</td>
<td>78.72</td>
<td>78.72</td>
<td>78.83</td>
<td>79.11</td>
<td>78.08</td>
<td>78.01</td>
<td>78.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>37.28</td>
<td>37.29</td>
<td>37.29</td>
<td>37.29</td>
<td>37.02</td>
<td>37.06</td>
<td>37.11</td>
<td>37.07</td>
<td>37.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>97.92</td>
<td>97.92</td>
<td>97.92</td>
<td>97.92</td>
<td>97.89</td>
<td>97.76</td>
<td>97.76</td>
<td>97.76</td>
<td>97.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique reflections</td>
<td>118456</td>
<td>131592</td>
<td>87575</td>
<td>16370</td>
<td>173061</td>
<td>173061</td>
<td>173061</td>
<td>173061</td>
<td>173061</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merged mean I/σ</td>
<td>0.094 (0.278)</td>
<td>0.06 (0.200)</td>
<td>0.108 (0.28)</td>
<td>0.116 (0.357)</td>
<td>0.079 (0.313)</td>
<td>0.076 (0.33)</td>
<td>0.060 (0.286)</td>
<td>0.067 (0.306)</td>
<td>0.068 (0.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>10.4 (7.9)</td>
<td>5.4 (4.4)</td>
<td>9.4 (2.7)</td>
<td>13.1 (5.9)</td>
<td>14.6 (9.1)</td>
<td>9.2 (3.6)</td>
<td>24.3 (16.2)</td>
<td>25.0 (18.0)</td>
<td>26.3 (19.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B Wilson</td>
<td>55.0</td>
<td>28.7</td>
<td>31.7</td>
<td>31.7</td>
<td>31.7</td>
<td>31.7</td>
<td>31.7</td>
<td>31.7</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average atomic B factor (Å²)</td>
<td>18.8 (5.71)</td>
<td>15.7</td>
<td>17.4</td>
<td>14.0</td>
<td>15.2</td>
<td>14.3</td>
<td>15.7</td>
<td>13.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R factor/Rmerge (%)</td>
<td>17.7/23.1</td>
<td>18.8/22.4</td>
<td>19.6/25.9</td>
<td>17.9/23.9</td>
<td>20.2/25.9</td>
<td>22.1/25.8</td>
<td>18.1/23.9</td>
<td>21.4/26.5</td>
<td>19.5/26.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R factor (all)</td>
<td>18.1</td>
<td>19.1</td>
<td>19.9</td>
<td>20.8</td>
<td>20.5</td>
<td>22.3</td>
<td>18.5</td>
<td>21.6</td>
<td>19.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. bond lengths (Å)/angles (°)</td>
<td>0.0267/2.1128</td>
<td>0.020/0.043</td>
<td>0.0191/3.3062</td>
<td>0.0207/2.1302</td>
<td>0.018/1.861</td>
<td>0.0221/1.7157</td>
<td>0.0261/2.2944</td>
<td>0.0183/2.134</td>
<td>0.0176/2.1269</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crystal</th>
<th>8</th>
<th>PROTEUM2</th>
<th>EVAL</th>
<th>Mosflm</th>
<th>9</th>
<th>PROTEUM2</th>
<th>EVAL</th>
<th>Mosflm</th>
</tr>
</thead>
<tbody>
<tr>
<td>4ddc</td>
<td>EVAL</td>
<td>Mosflm</td>
<td>4dd7</td>
<td>EVAL</td>
<td>Mosflm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit cell†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>78.84</td>
<td>78.82</td>
<td>78.80</td>
<td>78.70</td>
<td>78.08</td>
<td>78.01</td>
<td>78.05</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>37.03</td>
<td>37.02</td>
<td>37.00</td>
<td>37.00</td>
<td>37.00</td>
<td>37.00</td>
<td>37.00</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td>99.9 (99.4)</td>
<td></td>
</tr>
<tr>
<td>Unique reflections</td>
<td>361272</td>
<td>500514</td>
<td>323705</td>
<td>51970</td>
<td>12839</td>
<td>3477</td>
<td>21884</td>
<td>11716</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>55.75–1.69</td>
<td>20.67–1.60</td>
<td>19.11–1.72</td>
<td>20.67–1.60</td>
<td>19.11–1.72</td>
<td>20.67–1.60</td>
<td>19.11–1.72</td>
<td>20.67–1.60</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (99.4)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
<td>99.9 (99.9)</td>
</tr>
<tr>
<td>Merged mean I/σ</td>
<td>0.057 (0.156)</td>
<td>0.057 (0.179)</td>
<td>0.059 (0.15)</td>
<td>0.059 (0.15)</td>
<td>0.059 (0.15)</td>
<td>0.059 (0.15)</td>
<td>0.059 (0.15)</td>
<td>0.059 (0.15)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>26.7 (9.8)</td>
<td>31.4 (9.4)</td>
<td>25.2 (9.4)</td>
<td>31.4 (9.4)</td>
<td>25.2 (9.4)</td>
<td>31.4 (9.4)</td>
<td>25.2 (9.4)</td>
<td>31.4 (9.4)</td>
</tr>
<tr>
<td>B Wilson</td>
<td>12.2</td>
<td>12.5</td>
<td>12.2</td>
<td>12.5</td>
<td>12.2</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Average atomic B factor (Å²)</td>
<td>15.8</td>
<td>13.7</td>
<td>13.7</td>
<td>13.7</td>
<td>13.7</td>
<td>13.7</td>
<td>13.7</td>
<td>13.7</td>
</tr>
<tr>
<td>R factor/Rmerge (%)</td>
<td>16.7/23.2</td>
<td>18.3/22.4</td>
<td>17.0/22.7</td>
<td>18.3/22.4</td>
<td>17.0/22.7</td>
<td>18.3/22.4</td>
<td>17.0/22.7</td>
<td>18.3/22.4</td>
</tr>
<tr>
<td>R factor (all)</td>
<td>16.9</td>
<td>18.5</td>
<td>17.3</td>
<td>18.5</td>
<td>17.3</td>
<td>18.5</td>
<td>17.3</td>
<td>18.5</td>
</tr>
<tr>
<td>R.m.s.d. bond lengths (Å)/angles (°)</td>
<td>0.0307/2.5650</td>
<td>0.0200/0.0864</td>
<td>0.0201/0.0594</td>
<td>0.0204/0.0594</td>
<td>0.0201/0.0594</td>
<td>0.0204/0.0594</td>
<td>0.0201/0.0594</td>
<td>0.0204/0.0594</td>
</tr>
</tbody>
</table>

† Space group P412121 for all data except for EVAL data 4dd1 and 4ddc, where it is P212121. ‡ Reflections contaminated with ice scattering were removed from the data using a de-ice software procedure. § The resolution was cut back because at higher resolutions the protein structure refinement gave poor Rmerge statistics.

---


Simon W. M. Tanley et al. • Exchange and archiving of raw data
contain a value in the current data (a4cSpindle in Fig. 1). Also the fastest and slowest running coordinates of the pixel data are not given. Assuming the spindle is perpendicular to the X-ray beam, it is obvious what the sense of rotation is, i.e. clockwise or anticlockwise, by looking at a few consecutive frames. However, looking at the diffraction image means that the horizontal and vertical axes and their direction on the detector are already interpreted (in fact there are eight possible ways of doing so). A consistent interpretation was found previously, also helped by the visibility of the beam stop and cryo nozzle, and used in the current work. We therefore had prior knowledge on how to interpret the data. Rigaku Corporation has developed a new ASCII header type that contains all the definitions for orientations of goniometer axes and for the detector axes in the laboratory frame, so that a comprehensive set of metadata is then provided.

The Bruker image format contains the model of the goniometer and the fixed $\kappa$ angle (MACH3 and KAPPA in Fig. 2) and the goniometer rotation angles defined as Euler angles. Again we learned from previous data that the rotation directions for $2\theta$, $\omega$ and $\chi$ are opposite to that of $\phi$. The Bruker format potentially gives refined detector positions in terms of pitch, roll and yaw, but these are ignored by EVAL, as these will be the result of the PEAKREF (Schreurs, 1999) refinement.

Authors of integration software such as DENZO, Mosflm, MADNES/d*Trek and XDS have done the same tedious unravelling of detector formats. To avoid having to go through such efforts the CBF/imgCIF format was developed (Bernstein & Hammersley, 2005; Bernstein, 2005). It provides a structure in which all metadata can be found in one place. It consists of an ASCII imgCIF header and binary (CBF) or ASCII-based encoded data blocks. The binary format is space efficient owing to the use of compression algorithms, like Byte_offset compression, and is useful for large images and data transfer between collaborating groups, exactly the situation we were engaged with. Three categories of data exist – ARRAY data, AXIS data and DIFFRN data – allowing a unique definition of how to interpret the data, and no prior knowledge would be required if all data items were filled in. This is often not the case, however; e.g. PILATUS detector image files contain all relevant metadata in just a small comment line block, the so-called mini CBF format.

![Figure 1](image1)

**Figure 1** Part of an R-AXIS IV diffraction data image header.

![Figure 2](image2)

**Figure 2** Part of a Bruker diffraction data image header.
3.1.3. Data processing. The strategies chosen in CrystalClear (Rigaku Corporation, Tokyo, Japan) and with the PROTEUM2 processing software were rather different. The Rigaku equipment has only a single spindle axis $\varphi$ and the total rotation range was simply either 180 or 360°, in each case with 1° per frame, the detector being positioned at a distance to exploit the maximum diffraction resolution the crystal offered. The X-ray generator was set to 20 mA and exposure times varied between 3 and 10 min. Image plates have a relatively low quantum gain compared with a CCD (although their performance is very good compared with the photographic X-ray film of yesteryear) and therefore need correspondingly longer exposure times, approximately a factor of 2–3; however, they have a superb uniformity, a very large dynamic range and no spatial distortion. The Bruker X8 PROTEUM2 CCD diffractometer has a smaller detector aperture but can be rotated in the normal plane: and indeed this option was exercised and detector swing angles used in these data collections varied between −15 and 20°. The CCD has a better quantum gain but uses a fibre optic taper, causing non-uniformity and spatial distortion. Furthermore, it is equipped with a Kappa goniometer, allowing scans around $\omega$ and $\varphi$ at different $\kappa$ positions. The X-ray generator was set to 60 mA and exposure times were 10–30 s per 0.5°. The difference between the measuring strategies on the two apparatuses did not have a large effect on the completeness or the redundancies (although the total measuring time was shorter on the PROTEUM2 than on the R-AXIS IV); the data redundancies varied between 12.1 and 25.3 for Rigaku and 14.6 and 31.4 for Bruker (EVAL data in Tables 2 and 3).

We will focus on the EVAL results when comparing the two diffractometers as this has the overriding advantage, since identical processing software is used for the two devices, of giving a consistency of treatment of the diffraction data images. Indexing of all diffraction data was straightforward with DIRAX (Duisenberg, 1992), except for 4dd1 and 4ddc (see later). Each of the crystal unit-cell dimensions and orientation matrix were written to a .rmat file, which the program VIEW (Schreurs, 1998) uses to predict reflection positions and to write corresponding reflection boxes for integration within EVAL15. After EVAL15 processing, a post refinement with PEAKREF gave the final unit cells as listed in Tables 2 and 3. Any disagreement between predicted and refined reflection positions could be a reason to repeat the VIEW/EVAL15 cycle. Table 4 gives the final average errors in the reflected beam directions, in terms of reflection positions on the detector and in rotation angles when using a single unit-cell matrix. The Rigaku data needed refinement of unit-cell orientation for each box file to get acceptable agreement. The positional and angular average errors are still larger with these Rigaku data, typically 1–3 pixels and 0.1–0.2°. Though the latter is well within the rotation range of 1.0°, with the Bruker data it was possible to achieve both sub-pixel and sub-rotation range agreement. The post refinement was carried out by PEAKREF, a very flexible program that allows refinement of the unit-cell matrix, detector position, goniometer offsets and crystal position. No improvement of the predictions of the Rigaku diffraction spots could be obtained without releasing the orientation of the unit cell of each box file (a box file corresponds to a group of approximately 1000 reflections at roughly the same rotation angle). The improvement was mainly in the rotational positions, but no consistent interpretation in terms of crystal or detector movements or goniometer offsets could be found. Thus we introduced additional refinement of the orientation of the unit cell for each box file.

![Figure 3](image-url)

**Figure 3**
Rotational positional errors (in units of 0.01°) in Rigaku data set 1 (4dd0) (a) with a single unit-cell orientation matrix and (b) with different orientation matrices for each box file. (c) Bruker data set 5 (4dd1).
for the Rigaku data. Fig. 3 shows that the final agreement is comparable, given the difference in frame width, to the Bruker data. The case of the Bruker 4dd4 example is an exception. In this case apparently the crystal slips away from its original position as the orientation of the unit-cell axes was rotated approximately 3° between scans 2 and 3.

We cannot determine what causes the larger Rigaku data positional errors. The exchange of the two image plates after every exposure does not seem to affect their exact position, as no correlation with frame number, i.e. odd- and even-numbered images, was observed. The crystal orientation does not change in a systematic way as it is completely different for each experiment, so no correlation with the ϕ motor rotations was found either. We conclude that the crystal was not very well held to its initially placed position, possibly as a result of vibrating instrument parts.

3.1.4. Standard deviations and statistics. Every detector converts the X-ray photons that are absorbed by the phosphor layer into an electronic signal that is read out and stored in the image file. The detective quantum efficiency is a measure of the efficiency with which photons are detected and of the noise performance of the detector. It is defined as the signal-to-noise ratio of the output divided by that of the input. For an ideal detector this ratio would be 1.0. In practice a variety of factors reduce this number, like phosphor absorption efficiency, detector entrance window transmission, a phosphor noise factor, read-out noise, dark current and detector gain (Phillips et al., 2002). An additional quantity of interest is the overall detector gain; we will define gain as the number of analogue-to-digital units (ADU) recorded per X-ray photon. Ideally a pixel intensity should be divided by the gain to obtain the X-ray photon counts, so that standard deviations of pixel intensities can be estimated using Poisson statistics. The R-AXIS header does not contain a value for the gain, so EVAL assumes it to be 1.0. In the Bruker header we found a gain value of 3.83 ADU per photon. Various published papers have shown that the standard deviations of diffraction intensities behave other than according to Poisson distributions; an early description can be found in the text book by Stout & Jensen (1968) and applies even to so-called photon-counting detectors. In the area detector ‘modern era’, Leslie (1999) and Popov & Bourenkov (2003) show that the variance of intensities can be described by a second-order polynomial in I: \( \sigma^2 = k_0 + k_1I + k_2I^2 \). The second term represents the error estimate from Poisson statistics (\( \sigma = I^{1/2} \)) corrected for the gain and Lorentz–polarization factor (Lp).

The expression for \( \sigma^2 \) can be rewritten as \( \left[ \sigma_{\text{dark}}^2 + \sigma_{\text{read}}^2 + \sigma_{\text{g}}^2 + I \right] + (gI)^2 \), where I is the net intensity, g is a factor to be determined during scaling, and the subscripts dark, read and bg denote the dark-current, read-out-noise and background contributions to \( \sigma \). EVAL delivers standard deviations (\( \sigma_{\text{EVAL}} \)) using the first two terms. Scaling programs like SADABS (Sheldrick, 1996) use an error model \( \sigma_{\text{corr}}^2 = K[\sigma_{\text{EVAL}}^2 + (gI)^2] \), in which K and g are refined, to achieve more reliable error estimates from internal standard deviations such that \( \chi^2 = \left( N \sum (I - \langle I \rangle)^2/(N - 1)\sigma^2 \right) \) is close to 1.0. The latter approach is also applied in SCALA (Evans, 2006). If the intensities I are on an absolute scale, i.e. represent actual X-ray photon counts, SADABS typically finds K values in the range 0.7–1.5 and g values in the range 0.02–0.04. An incorrectly estimated gain value will affect the estimated \( I/\sigma \) of reflections, but scaling programs will more or less correct for this, notably via the \( \chi^2 \) analysis. This correction may, however, not be in place when reflections are rejected on the basis of \( (I - \langle I \rangle)/\sigma > 4 \) (in the case of SADABS) and may lead to unwanted rejections. In MOSFLM/SCALA the error model being used is \( \text{sdFac}[\sigma^2 + \text{sdB LpI} + (\text{sdAddI})]^{1/2} \) (as an example, for the processed diffraction data set of crystal 3, \( \text{sdFac} \cong 1.5, \text{sdAdd} \cong 0.02 \) and \( \text{sdB} = 3.17 \) for full reflections). d*Trek also uses a two-term adjustment of the standard deviations to match normal \( \chi^2 \) distributions. In a recent paper, Waterman & Evans (2010) showed that the standard deviations of intensities from profile fitting or summation integrations are indeed underestimated and that simulation of the detection process, taking into account the various sources of error, leads to more realistic error estimates. Because of the similar procedures used for adjusting the standard uncertainties, we believe that comparison of \( I/\sigma \) values remains valid and is a necessary requirement of any physical science, of which crystal structure analysis is but one example.

\( I/\sigma \) values for merged data can be found in Tables 2 and 3, and are summarized in Fig. 4. For the Rigaku data, the numbers for EVAL and Mosflm are in reasonable agreement, while d*Trek produces in general lower values, sometimes markedly so (4dd3, 4ddb). For the Bruker data, EVAL and the PROTEUM2 (SAINT Version 7.68a) processing software are in close agreement even for crystals 5 and 9, which are in fact orthorhombic, leading to a lower redundancy for EVAL (Table 3). Mosflm has lower \( I/\sigma \) values in a majority of cases. The multiple-scan data collection with different detector positions may not be ideal for Mosflm without skilled fine tuning, i.e. which might be possible for the MOSFLM developers themselves. The accessibility of the raw diffraction data images linked to this article thus shows up an immediate advantage of archiving the raw diffraction data relating to a published article.

**Figure 4**
Merged \( I/\sigma \) values for data sets. 1: 4dd0; 2: 4dd2; 3: 4dd3; 4: 4dd9; 5: 4dd1; 6: 4dd4; 7:4dd6; 8: 4dd7; 9: 4ddc; 10: 4dda; 11:4ddb.
Bruker diffraction data images are corrected for the relative sensitivity across the face of the detector by a flood-field image, determined with an isotropically scattering fluorescent sample. Careful inspection of the diffraction data images shows that some moisture had built up (dark variation in the background in the central part of the image) between the front protective screen and the phosphor layer of the detector (see Fig. 5a). A projection of reflections rejected by SADABS onto the detector (4dd7; Fig. 5b) shows that many occur in these areas of moisture, indicating systematic problems with the flood-field correction now not being appropriate as it would have been measured before the moisture build up. This problem can only be removed by a maintenance technician.

For protein-model-refined crystal structures the \( R \) factors are often in the range 20–25%, while the intrinsic measurement errors are around 5%. Vitkup et al. (2002) show that the major contributions to this gap between \( R \) factors and the measurement errors are caused by the lack of a proper description of anisotropic protein motions, which can often not be determined because of the limited resolution of the data. At atomic diffraction resolution, spherical atomic scattering factors are a further inadequate approximation. Indeed \( R_{\text{merge}} \) values of our diffraction data sets range from 5 to 15% for the 1.7 Å crystals, while the \( R_{\text{free}} \) protein model refinement values are 22.3–26.3 (see Fig. 6). For the Rigaku diffraction data all the \( R \) factors agree closely between the data sets. For the Bruker diffraction data sets there is more spread. However the basic \( R_{\text{merge}}/R_{\text{free}} \) gap is the same for each. Again crystals 5 and 9 are exceptions because of their orthorhombic symmetry.

### 3.1.5. Crystal scattering power versus diffraction resolution

The incentive to start this work was an apparent systematic deviation between protein-model-refined \( B \) factors obtained with the diffraction data from different instruments and/or processing software. Thus we undertook data processing of all 11 data sets with the single software package EVAL. EVAL’s diffraction data processing statistics in Tables 2 and 3.

---

**Figure 5**

(a) Bruker diffraction data image of crystal 8. In the central part, left of the backstop shadow towards the solvent ring, the X-ray background scattering is seemingly lower, as shown in the dark stain-like area. (b) Reflections, projected onto the detector, for which the intensity deviates more than 4\( \sigma \) from the expected value.

**Figure 6**

\( R_{\text{free}} \) (upper) and \( R_{\text{merge}} \) (lower) in % for crystals diffracting to 1.7 Å. (a) Rigaku data of crystals 1–4. (b) Bruker data of crystals 5–9.
show that the quality of the crystals varies somewhat, but mostly they diffract to approximately 1.7 Å, except for two crystals (4dda and 4ddb), which simply did not diffract further than 2.4–2.5 Å. In the highest diffraction resolution shells the crystals (4dda and 4ddb), which simply did not diffract further average unmerged $I/\sigma(I)$ varies between 0.9 and 3.5. Despite the difference in beam flux, detector quantum gain and measurement times, average individual $I/\sigma(I)$ values are not grossly different (4.5–11.0; data not shown). Crystals that diffract to 1.7 Å were determined to have similar Wilson $B$ factors for each diffractometer (Fig. 7), but clearly those from the Rigaku-processed diffraction data (crystals 1–4) are significantly larger than those from the Bruker-processed diffraction data (crystals 5–9). However, the difference is much smaller than the average atomic $B$ factors from the protein model refinements (see §3.2 for a discussion on software for the protein model refinements). We can think of two reasons why this diffractometer hardware difference arises. First, Bruker diffraction data images are corrected for non-uniformity by flood-field images. Correction factors can be as much as 10–20%. Any errors in this correction procedure could have a systematic effect on the drop of intensity with 2$\theta$ and thus explicitly on the atomic $B$ factors. However, reflection intensities in the Bruker diffraction data are measured at completely different positions on the detector, because of the various swing angles constituting a complete data set, so that such systematic effects are not likely. Secondly, high-order reflections have higher incidence angles in the case of Rigaku imaging plate data, with the detector set in the usual 2$\theta = 0$ detector position. If the reflections are measured in the thin-phosphor regime (Chupas et al., 2003), the X-ray absorption is proportional to the path length through the phosphor and intensities should be corrected (Zaleski et al., 1998). However, generally, image plates are designed such that, for wavelengths larger than 1 Å, reflected X-ray beams are fully absorbed and such a correction would not be necessary. Still this effect could leave traces that will eventually end up in the protein model atomic $B$ factors.

3.2. Comparison of diffraction data processing software

All the crystal structures were solved using molecular replacement with Phaser (McCoy et al., 2007) and restrained refinement with TLS (translation–libration–screw motion) in REFMAC5 (Vagin & Teplyakov, 2010) in CCP4i (http://www.ccp4.ac.uk/ccp4i_main.php), using the lysozyme structure 2w1y as the molecular search model (Cianci et al., 2008). Model building, adjustment and refinement were carried out using the Coot (Emsley & Cowtan, 2004) molecular-graphics program and REFMAC5 in CCP4i, respectively. Metal ligand binding occupancies and their $B$ factors were finally calculated using SHELXTL (Sheldrick, 2008).

3.2.1. $B$ factors. The Wilson $B$ factors for the processed diffraction data of EVAL and Mosflm agree the closest. In general the agreement between the Wilson and protein-model-refined $B$ factors is very good for EVAL and is somewhat less so for Mosflm (Fig. 8). One would expect a rough correspondence between Wilson $B$ factor and the refined average individual atomic $B$ factors, though the latter tend to be higher in general. Indeed, most numbers in Fig. 8 are negative, but the deviation is significantly larger for $d^\ast$Trek. It appears that the diffraction data processing software may be critical to the published atomic displacement parameters of (protein) structures. Some caution has to be taken here as expert users of Mosflm, $d^\ast$Trek and PROTEUM could find slightly different results. Some specific deviations have an easily explainable cause. For example, crystals 5 (4dd1) and 9 (4ddc) have significantly higher refined $B$ factors with PROTEUM and Mosflm data processing, because loss of

---

Figure 7

Wilson diffraction data $B$ factors (marked ‘Wilson’) and protein-model-refined average isotropic atomic $B$ factors (marked ‘refined’) (Å$^2$) for all 11 crystals. The last two crystals diffract to a lower, i.e. poorer, resolution. 1: 4dd0; 2: 4dd2; 3: 4dd3; 4: 4dd9; 5: 4dd1; 6: 4dd4; 7: 4dd6; 8: 4dd7; 9: 4ddc; 10: 4dda; 11: 4ddb.

Figure 8

$B_{\text{Wilson}} - B_{\text{refined}}$ (Å$^2$) for the 11 crystals. 1: 4dd0; 2: 4dd2; 3: 4dd3; 4: 4dd9; 5: 4dd1; 6: 4dd4; 7: 4dd6; 8: 4dd7; 9: 4ddc; 10: 4dda; 11: 4ddb.
tetragonal symmetry was not recognized (see below), in our hands (JRH and ST), and therefore the difference in conformation between the two independent molecules was modelled as a type of static disorder. Data sets 10 (4dda) and 11 (4ddb) are exceptions because of the low resolution and concomitant TLS refinement only. Therefore the Wilson and refined $B$ factors are not comparable. Wrongly estimated low-order reflection intensities may lead to, most obviously, erroneous Wilson $B$ factors but also atomic $B$ factors. EVAL rejected the lowest resolution reflections that were partly shadowed by the beam stop.

We analysed if significant differences could be found between the final refined structures or the initial electron density maps from which Pt atoms were located. The r.m.s. deviations between atom positions were in the range 0.1–0.6 Å for all pairwise software comparisons, except for crystal 1 where the $d^\#$Trek-refined structure deviated from the results from both EVAL and Mosflm by 0.5–1.0 Å. $B$-factor variations between residues were similar for all data, and $d^\#$Trek always has higher values. Apparently, the larger $B$ factor for $d^\#$Trek is isotropic, i.e. it does not affect one part of the molecule more than others. Densities in $2F_o-F_c$ or $F_o-F_c$ maps at Pt positions are similar in all cases.

The use of EVAL for processing all of the diffraction data sets provided a consistent platform for our large ensemble of data sets for the various protein and platinum compound model refinements with REFMACS and then SHELXTL. Platinum occupancy values, and their standard deviations, were finally calculated using the results from three different diffraction data processing programs. We found that the differences in $B$ factors do not impinge on the occupancies of Pt in cisplatin and carboplatin bound to lysozyme (Tanley et al., 2012) as these agree to within $\pm 5\%$.

### 3.2.2. Contamination with ice

Despite the use of cryoprotectants, some diffraction patterns indicate different levels of ice formation. It was uncertain what influence ice-contaminated reflection intensities would have in the protein model refinements. Diffraction data processing software sometimes has the option to avoid reflection integrations in suspect regions. In EVAL we decided to integrate all data and to decide afterwards which resolution regions we might want to discard. A diffraction data image of crystal 2 (4dd2) and projection of reflections with $|I-\langle I||\sigma > 4$ by SADABS (Fig. 9) show that most of the reflections in ice regions will be rejected, so that no large problems were to be expected in the protein model refinement. Also SADABS rejections are

![Diffraction pattern](image)

**Figure 9**

Diffraction pattern (a) and reflections deviating more than $3\sigma$ from the average of equivalents in SADABS for crystal 2 (4dd2), for both (b) untreated and (c) de-iced data.
shown after our ice-rejection procedure using ANY (Schreurs, 2007). The result is that for crystals 2 (4dd2), 10 (4dda), 11 (4ddb), 6 (4dd4) and 7 (4dd6) the completeness (Tables 2 and 3) dropped to 82–91%. The changes in $R_{\text{merge}}$ were insignificant, whereas the protein model refinement $R$/$R_{\text{free}}$ values were only noticeably different for the lower-resolution crystals: $R$/$R_{\text{free}}$ = 21.4/29.9 versus 20.0/28.5 (de-iced) for crystal 10 (4dda) and 28.5/34.8 versus 21.4/27.9 for de-iced for crystal 11 (4ddb). It would seem that diffraction data scaling programs cannot reject such reflection groups if all diffraction equivalents are equally affected by ice scattering. This, however, is rarely the case and if the diffraction data redundancy is sufficiently large the reflections can actually be rejected, as done by SADABS.

3.2.3. Loss of tetragonal symmetry cases. In the EVAL software, DIRAX finds a first primitive lattice from peaks in a few diffraction images. Our experience is that when a large number of diffraction peaks are chosen the unit cell is sufficiently accurate for integration in EVAL15 without a major difference between observed and predicted peak positions. EVAL15 shifts the diffraction peaks to optimal positions for the profile fit as determined from $\chi^2$. If we are pleased with the agreement we usually do a post refinement to determine the best crystal unit cell for structure determination. If the errors are too large we may refine the unit cell (and orientation), restart the generation of box files at predicted positions and subsequently reintegrate with EVAL15. This may occur, for example, if goniometer offsets or detector positions are not known or not trusted. As mentioned in the Metadata section, we ignore these offsets and refine them with EVAL. However, for two crystals the errors were larger than what we are used to and larger than for the other crystals. Crystal 9 (4ddc), when indexed with $mmm$ symmetry, gave positional errors on the detector of 0.18 mm (corresponding to more than 2 pixels) and a 0.14° error in rotation angle, far too large in our view. In addition the ‘rlaue’ instruction in ANY gives $R_{\text{merge}}/R_{\text{meas}}/R_{\text{pim}}$ of 0.176/0.180/0.038 and 0.127/0.130/0.028 for $mmm$ and $mmm$, respectively; $mmm$ symmetry class was clearly an improvement. We were in fact warned because DIRAX found significant differences for the $a$- and $b$-axis values, with whatever peaks we offered. Of course one can be tempted to assume the well known tetragonal symmetry of HEWL. Release of the constraint between $a$ and $b$ values in PEAKREF and subsequent integration with EVAL15 solved the problem, leading to $a = 77.94$ and $b = 79.09$ Å instead of 78.52 Å, and the diffraction peak agreement became accurate to within 0.07 mm and 0.07° (Fig. 10). Close inspection showed that a similar phenomenon occurred with crystal 5 (4dd1).

4. Conclusions
This joint project needed the network transfer of 35.3 Gb of raw diffraction data images between Manchester and Utrecht. As soon as the images arrived in Utrecht they were compressed to 20 Gb, using the ncompress lossless data compression package, to save disk space and because EVAL can read compressed images. It would have been efficient to compress before file transmission, but d*Trek, PROTEUM and Mosflm only process uncompressed images and therefore they were left untouched in Manchester. It took about 30 h of total transfer time to get the data across. As this was done one data set at a time, constrained by a typical working day, the transfer was spread over several days. In future it may be advisable to use on-the-fly compression (e.g. scp -C in Linux) during file transfer as well as a simple concatenation of the various data sets.

There exists long-term interest in performing comparative studies of hardware and software as exemplified by Helliwell et al. (1981). The present paper, one of the first to be accompanied with archiving of the raw data, may be the start of further comparative studies.

Comparison of diffractometer hardware was achieved by using EVAL processing. The Rigaku diffraction data sets have

![Figure 10](image_url)

**Figure 10**
Distribution of horizontal (in units of 0.01 mm) and rotational (in units of 0.01°) peak shifts of data 4ddc in EVAL15. Horizontal peak shifts using (a) a tetragonal or (b) an orthorhombic unit cell; rotational peak shifts using (c) a tetragonal or (d) an orthorhombic unit cell.
larger positional errors when compared with the Bruker diffraction data sets, which could be due to the crystal not being very well fixed into position, possibly as a result of vibrating instrument parts. The hardware is also partly responsible for the difference in Wilson $B$ factors.

In comparing the software programs, the Wilson $B$ factors are often significantly larger for $d^*Trek$. This would mean that the data processing software may be critical to the published atomic displacement parameters of such protein structures. $Mosflm$ performed worse in the processing of multiple-scan data with different detector positions, as compared with the single-detector-position Rigaku data. Despite differences in $B$ factors of the refined structures derived from data of different processing software, the Pt occupancies were within a $\sigma$ range of $\pm 5\%$ (see Tanley et al., 2012). The availability of the raw diffraction images allows for independent assessment of software packages. The results described here may be biased in favour of EVAL, because we are the experts in this software.

In EVAL we implemented a procedure to avoid reflections affected by ice scattering. However, the results are not much different when no special care was taken, at least in these reasonably highly redundant diffraction data sets. We found that SADABS is capable of rejecting ice-affected reflections when the data have sufficiently high redundancy.

Without prior knowledge we would not have been able to discern sufficient metadata to carry out data processing of both types of diffraction images. This raises concerns with respect to long-term archiving of raw diffraction data. Care has to be taken that in the future unambiguous information is available, although this paper in itself is already a step towards providing the research community with knowledge of such metadata. The raw data will be deposited at Manchester University in 2013, so that software developers are able if they wish to improve on our data processing. Currently, a temporary depository is available at http://rawdata.chem.uu.nl.

Processed and derived data have been deposited with the PDB [PDB codes 3txb, 3txd, 3txe, 3txf, 3txg, 3txh, 3txi, 3txj and 3txk (new to this paper); PDB codes 4dd0, 4dd1, 4dd2, 4dd3, 4dd4, 4dd6, 4dd7, 4dd9, 4dda, 4db, 4ddc (from Tanley et al., 2012)].

We are grateful to Dr Pat Bryant and Dr Colin Levy, who are coauthors on the Tanley et al. (2012) article, for their help in the original diffraction data image measurements for these 11 crystal samples on the R-AXIS IV and Bruker PROTEUM diffractometers installed at the University of Manchester, and indeed we are grateful to all coauthors of Tanley et al. (2012) for this collaboration. JRH is grateful to Brian McMahon of the IUCr for many discussions on raw data archiving. JRH and LKB are members of the IUCr’s Diffraction Data Deposition Working Group and this article is a contribution on the technical aspects related to the policy issues of this group, mentioned only briefly in the introduction to this article. JRH is grateful to Dr Phil Butler, Dr Jonathan Besson and Dr Meik Poschen of the University of Manchester for their raw data archiving expertise and guidance on obtaining digital object identifiers for each diffraction data set.

References
Chapter 7: Room-temperature X-ray diffraction studies of cisplatin and carboplatin binding to His-15 of HEWL after prolonged chemical exposure

Simon W.M. Tanley\textsuperscript{a}, Antoine M.M. Schreurs\textsuperscript{b}, Loes M.J. Kroon-Batenburg\textsuperscript{b} and John R. Helliwell\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, UK

\textsuperscript{b} Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Correspondence email: john.helliwell@manchester.ac.uk


\textbf{Keywords}: Cisplatin; Carboplatin; histidine; aqueous media; DMSO media.

SWMT grew all crystals used in this study, mounted the crystals and collected data on the Bruker APEX II home source X-ray diffractometer. JRH taught SWMT how to mount the crystals in capillary and helped with the collection and integration of data. AMMS processed all data using EVAL-15 and SWMT processed one of the datasets using the Bruker internal software program. SWMT refined all structures presented in this paper. SWMT wrote the article, with JRH and LMJKB editing the manuscript.
Room-temperature X-ray diffraction studies of cisplatin and carboplatin binding to His15 of HEWL after prolonged chemical exposure

The anticancer complexes cisplatin and carboplatin are known to bind to both the N\(^3\) and the N\(^6\) atoms of His15 of hen egg-white lysozyme (HEWL) in the presence of dimethyl sulfoxide (DMSO). However, neither binds in aqueous media after 4 d of crystallization and crystal growth, suggesting that DMSO facilitates cisplatin/carboplatin binding to the N atoms of His15 by an unknown mechanism. Crystals of HEWL cocrySTALLized with cisplatin in both aqueous and DMSO media, of HEWL cocrySTALLized with carboplatin in DMSO medium and of HEWL cocrySTALLized with cisplatin and N-acetylg glucosamine (NAG) in DMSO medium were stored for between seven and 15 months. X-ray diffraction studies of these crystals were carried out on a Bruker APEX II home-source diffractometer at room temperature. Room-temperature X-ray diffraction data collection removed the need for cryoprotectants to be used, ruling out any effect that the cryoprotectants might have had on binding to the protein. Both cisplatin and carboplatin still bind to both the N\(^3\) and N\(^6\) atoms of His15 in DMSO media as expected, but more detail for the cyclobutanedicarboxylate (CBDC) moiety of carboplatin was observed at the N\(^6\) binding site. However, two molecules of cisplatin were now observed to be bound to His15 in aqueous conditions. The platinum peak positions were identified using anomalous difference electron-density maps as a cross-check with F\(_{o}\) – F\(_{c}\) OMIT electron-density maps. The occupancies of each binding site were calculated using SHELXL. These results show that over time cisplatin binds to both N atoms of His15 of HEWL in aqueous media, whereas this binding is speeded up in the presence of DMSO. The implication of cisplatin binding to proteins after a prolonged period of time is an important consideration for the length of treatment in patients who are given cisplatin.

1. Introduction

Cisplatin and carboplatin (Supplementary Fig. S1\(^1\)) are platinum anticancer drugs which target DNA. However, 90% of their reported binding is to plasma proteins (Fischer et al., 2008). A variety of different proteins have been studied for their binding affinity to cisplatin using either mass spectrometry or X-ray diffraction, including a copper transporter (Arnesano & Natile, 2008; Crider et al., 2010), a copper chaperone (Boal & Rosenzweig, 2009), superoxide dismutase (Calderone et al., 2006; Casini et al., 2008), cytochrome c (Casini et al., 2006; Casini, Gabbiani et al., 2007), human albumin (Ivanov et al., 1998), ubiquitin (Hartinger et al., 2007), glutathione reductase (Zimmermann & Burda, 2010), Na\(^+\)/K\(^+\)-ATPase (Huličiak et al., 2012) and hen egg-white lysozyme (HEWL; Casini, Mastrobuoni et al., 2007; Tanley et al., 2012). Carboplatin binding affinity, however, has been less studied, with one study using mass spectrometry (Casini, Mastrobuoni et al., 2007) and another study involving X-ray diffraction (Tanley et al., 2012), both of which looked at binding to HEWL. In our preceding paper (Tanley et al., 2012), HEWL was cocrySTALLized with cisplatin or carboplatin in both aqueous medium and medium with added DMSO, using either glycerol or Paratone as a cryoprotectant, and cisplatin was also cocrySTALLized with HEWL and a ligand (N-acetylg glucosamine; NAG).

\(^{1}\) Supplementary material has been deposited in the IUCr electronic archive (Reference: YTS048).
exposure to these metal compounds (Tanley et al., 2012). Storage in their mother liquor and thereby their prolonged chemical stability in the batch of crystals used for the previous study but after their prolonged storage in their mother liquor and thereby their prolonged stability. The absence of the imidazole ring (an imidazolyl anion), providing two N atoms at which a metal centre can bind. The loss of the N-hydrogen is a feature of this histidine the usual N-hydrogen of such a residue is absent and planar. Histidine bound to two platinum centres means that in the imidazole plane of the imidazole ring (an imidazolyl anion), providing two N atoms of His15 in further detail using the same dataset with cisplatin and HEWL. The 7.5% DMSO condition (1 mM) was observed, giving rise to the conclusion that DMSO somehow increases over time, whether extra detail is observed at the binding sites for all four conditions and whether cisplatin aqueous conditions’ crystals unfortunately prevented their crystal mounting and detailed study.

### Table 1

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Carboplatin/DMSO</th>
<th>Cisplatin/DMSO</th>
<th>Cisplatin/NAG/DMSO</th>
<th>Cisplatin/aqueous</th>
<th>Carboplatin/DMSO, 100 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>4g4c</td>
<td>4g4a</td>
<td>4g4b</td>
<td>4g49</td>
<td>4g4b</td>
<td></td>
</tr>
<tr>
<td>4g4a</td>
<td>4g4b</td>
<td>4g49</td>
<td>4g4b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0.25</td>
<td>0.16</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>300</td>
<td>0.25</td>
<td>0.16</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>300</td>
<td>0.16</td>
<td>0.16</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>300</td>
<td>0.3</td>
<td>0.25</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>300</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>100</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2. Materials and methods

#### 2.1. Materials

Cisplatin, HEWL and NAG were purchased from Sigma–Aldrich, UK. Carboplatin was purchased from Calbiochem, USA. DMSO from Cambridge Isotope Laboratories, UK and sodium chloride (NaCl) from Fisher Scientific, UK. Sodium acetate and acetic acid were purchased from BDH in AnalR grade.

#### 2.2. Crystallization conditions

The cocryostallization conditions for each crystal at pH 4.7 are the same as those previously published (Tanley et al., 2012), apart from the increased crystal storage time (Table 1) and the use of only 0.66 mM DMSO for the cisplatin/DMSO cocryostallization condition compared with the 1 mM DMSO previously used. The crystallization conditions were set up to begin with in 5, 7.5 and 10% DMSO media with cisplatin and HEWL. The 7.5% DMSO condition (1 mM) was used for the cryo study (Tanley et al., 2012). When trying to use these crystals for the room-temperature study, the crystals could not be detached from their pot, but it was possible to obtain some crystals from the 5% (0.66 mM) DMSO condition. The platinum occupancy values for the 5 and 7.5% DMSO conditions are very similar (see below and Tanley et al., 2012).

#### 2.3. X-ray data collection, structure solution and refinement

Crystals were mounted in thin-walled 0.7–1.0 mm glass/quartz capillary tubes. The crystals were then centred on a home-source Bruker APEX II CCD detector diffractometer. Data were collected using the same dataset with cisplatin and HEWL. The 7.5% DMSO condition (1 mM) was used for the cryo study (Tanley et al., 2012). When trying to use these crystals for the room-temperature study, the crystals could not be detached from their pot, but it was possible to obtain some crystals from the 5% (0.66 mM) DMSO condition. The platinum occupancy values for the 5 and 7.5% DMSO conditions are very similar (see below and Tanley et al., 2012).

![Figure](image-url)
Table 2

<table>
<thead>
<tr>
<th>N\textsuperscript{d} binding site</th>
<th>N\textsuperscript{e} binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pt–N distance (Å)</td>
</tr>
<tr>
<td>Occupancy (%)</td>
<td>Occupancy (%)</td>
</tr>
<tr>
<td>Cisplatin/aqueous, RT</td>
<td>50 (1.0)</td>
</tr>
<tr>
<td>Cisplatin/DMSO, RT</td>
<td>73 (0.8)</td>
</tr>
<tr>
<td>Cisplatin/DMSO/NAG, RT</td>
<td>69 (0.4)</td>
</tr>
<tr>
<td>Cisplatin/DMSO/glycerol, 8 d</td>
<td>70 (0.2)</td>
</tr>
<tr>
<td>Carboplatin/DMSO, RT</td>
<td>67 (0.2)</td>
</tr>
<tr>
<td>Carboplatin/DMSO, cryocooled, 15 months</td>
<td>68 (0.4)</td>
</tr>
<tr>
<td>Carboplatin/DMSO, cryocooled, 8 d</td>
<td>66 (0.2)</td>
</tr>
</tbody>
</table>

All structures were solved by molecular replacement with Phaser (McCoy et al., 2007) and restrained refinement with TLS by REFMAC5 (Vagin & Teplyakov, 2010) in CCP4i (McNicholas et al., 2011) using the reported lysozyme structure (PDB entry 2w1y; Cianci et al., 2008) as a molecular search model. Model building, adjustment and refinement were carried out using the Coot molecular-graphics program (Emsley & Cowtan, 2004) and REFMAC5 in CCP4i, respectively. Ligand-binding occupancies were calculated using SHELXTL (Sheldrick, 2008). Crystallographic and refinement parameters are summarized in Table 1.

3. Results

3.1. Cisplatin binding to His15

For the three crystals cocrystallized with cisplatin, cisplatin was found to be bound to the N\textsuperscript{d} atom and the N\textsuperscript{e} atom of the imidazole ring of His15, with the clarity of each of the Cl and N atoms varying in each case (Figs. 1a, 1b and 1c). Comparing the electron density observed in these conditions with that from one of the data sets from our previous cryocooling study (Tanley et al., 2012; Fig. 1d) shows no

Figure 1

Cisplatin binding to HEWL in aqueous medium with 15 months chemical exposure (a), in 5% DMSO medium with 14 months chemical exposure (b), and on cocystallization with 7.5% DMSO and NAG with 7 months chemical exposure (c), with data collection in each case at RT. (d) Cisplatin in 7.5% DMSO medium with Paratone as the cryoprotectant after 8 d cocrystallization studied at cryo temperatures (Tanley et al., 2012) for comparison. These $F_o - F_c$ OMIT electron-density maps are shown at the 1.5$\sigma$ level (in purple) and the anomalous difference electron-density maps (in orange) are shown at the 2.5$\sigma$ level. The atoms bound to each Pt ion differ based on the electron density observed and are labelled, with the N\textsuperscript{d} binding sites on the left-hand side and the N\textsuperscript{e} binding sites on the right-hand side in each figure. The Pt–N distances and the precisions of these distances are given in Table 2.
real difference in the density observed, although the atoms are placed slightly differently. The exception to this is the cisplatin in aqueous medium experiment (Fig. 1a), where we now observe some degree of binding of cisplatin at His15 compared with the absence of binding after 4 d of exposure in the previous study. For this data set, only one N atom in the N<sup>6</sup> binding site can be modelled binding to the Pt ion based on the electron density and only the Pt ion is observed at the N<sup>4</sup> binding site. Fig. 1(c) shows the unusual characteristic of the N<sup>4</sup> atom replacing an amino group, rather than a Cl atom, which acts as a leaving group on cisplatin. Calderone et al. (2006) also observed this in their study of cisplatin binding to His19 of cuprozinc superoxide dismutase. In their study, they only observed electron density for the Pt ion and the two Cl atoms, with both of the amino groups being displaced; hence, this displacement of an amino group has been observed previously. The occupancies of the cisplatin molecule at each binding site are given in Table 2 and were calculated using SHELXLT (Sheldrick, 2008), with a comparison to the cryo-example from the previous study also being given. All occupancy values are expected to have a standard uncertainty of approximately ±5% (Tanley et al., 2012).

3.2. Carboplatin binding to His15

Carboplatin is also found to be coordinated to both the N<sup>4</sup> and N<sup>6</sup> atoms of His15 (Fig. 2). Comparing the binding sites in the RT data study after 13 months of prolonged chemical exposure (Fig. 2a) with those in one of the cryocooled carboplatin data sets from the previous study after 8 d of cocrysallization (Fig. 2b; Tanley et al., 2012) and those in the carboplatin cryocooled data set after 13 months of prolonged chemical exposure (Fig. 2c), extra electron density is observed at the N<sup>4</sup> binding site in the F<sub>c</sub> – F<sub>c</sub> OMIT map for the RT case (Fig. 2a), which is interpretable as a portion of the cyclobutane-dicarboxylate (CBDC) moiety of carboplatin (Supplementary Fig. S1). This extra electron density is not observed in either of the cryocooled data sets (see §3.3 for more detail). For the N<sup>4</sup> binding site, the electron density is very similar for both the RT and cryocooled data sets, with both showing evidence for a portion of the CBDC moiety (Fig. 2). The occupancy values for carboplatin binding to both sites were refined using SHELXLT (Sheldrick, 2008) and are given in Table 2, along with the Pt—N distances, and with the values for the cryocooled example after 8 d crystallization as a comparison.

3.3. Carboplatin binding to the N<sup>6</sup> atom of His15 in more detail

Comparing the F<sub>c</sub> – F<sub>c</sub> OMIT electron-density map at the N<sup>6</sup> binding site (Fig. 2) of the RT 13 months prolonged chemical exposure data set with the cryocooled data set after 8 d crystallization shows a difference, with extra detail being seen in the RT data set. This extra detail corresponds to a downwards shift of the NH1 atom of the Arg14 residue by 0.8 Å (Fig. 3a), meaning that there is space for an extra atom to be accommodated. A cryocooled data set after 13 months prolonged chemical exposure was also collected as a comparison with the RT data after 13 months prolonged chemical exposure. This was to confirm whether the extra detail observed was a consequence of the temperature of data collection or was based on the length of time of crystallization and exposure to carboplatin. The cryocooled data set after 13 months showed the same degree of detail of electron density as the cryocooled 8 d crystallization data set (Figs. 2b and 2c). Comparing the position of the NH1 atom of Arg14 between the RT 13 months data set and the cryocooled 13 months data set, the NH1 atom is shifted both downwards and sideways by 1.8 Å (Fig. 3b). To confirm whether these shifts are significant, the coordinate error of each atom can be calculated using (1) based on the Cruickshank DPI values (Table 1; Cruickshank, 1999; Blow, 2002), taking account of the atomic B (B<sub>atom</sub>) versus the average B (B<sub>average</sub>),

\[ \text{coordinate error of atom} = \text{DPI} \times (B_{\text{atom}}/B_{\text{average}})^{1/2}. \]

(1)

The standard deviation of the calculated shift is determined using the coordinate error of the NH1 atom of the Arg14 residue in both structures (1 and 2),

\[ \text{standard deviation} = [(\text{coordinate error 1})^2 + (\text{coordinate error 2})^2]^{1/2}. \]

(2)

The value and standard uncertainty for the shift between the cryocooled 13 months data set and the RT 13 months data set is 1.8 ± 0.4 Å and the value and standard uncertainty of the shift between the RT 13 months data set and the cryocooled 8 d data set is 0.8 ± 0.2 Å. Thus, the shift of the NH1 atom of Arg14 is significant in both cases and can thus explain the extra detail in the F<sub>c</sub> – F<sub>c</sub> OMIT electron-density map around the N<sup>6</sup> binding site that is observed in the RT 13 months prolonged chemical exposure cocrysallization data set.

4. Discussion

4.1. The initial experimental concepts for these studies

The interest in embarking on this study was to explore in further detail the binding of cisplatin and carboplatin to His15 of HEWL in both DMSO and aqueous media but after a prolonged period of chemical exposure. In our previous study (Tanley et al., 2012), experiments involving HEWL cocrysallization with cisplatin and carboplatin were undertaken with crystal-growth times of between 4

---

**Figure 2**

Carboplatin binding to His15. (a) RT data set for carboplatin binding to HEWL in 7.5% DMSO medium after 13 months of crystal growth, (b) cryocooled data set for carboplatin binding to HEWL in 7.5% DMSO media with Paratone as the cryoprotectant after 8 d cocrysallization (from Tanley et al., 2012) and (c) cryocooled data set for carboplatin binding to HEWL in 7.5% DMSO media with silicone oil as the cryoprotectant after 13 months of crystal growth. The F<sub>c</sub> – F<sub>c</sub> OMIT electron-density maps (in purple) are shown at 1.5σ and the anomalous difference electron-density maps (in orange) are shown at 3σ. The atoms of the carboplatin molecule are labelled, along with the N<sup>6</sup> and N<sup>4</sup> atoms of His15.
and 8 d. These results showed that the addition of DMSO facilitated the binding of these Pt complexes to both the N\(^{3}\) and N\(^{4}\) atoms of His15, with high-occupancy binding being observed for both binding sites (~65% for N\(^{3}\) and ~50% for N\(^{4}\)), whereas in aqueous conditions this binding was not observed.

This new study used four of the previously grown crystal types which had been left on the shelf for between seven and 15 months. HEWL cocrystallized with carboplatin in aqueous medium was not studied here owing to the fragility of the crystals, which prevented their mounting in a capillary. RT X-ray diffraction was used as this removed the need for cryoprotectants: the previous study showed that both glycerol and Paratone can bind to the protein. The use of RT also had the practical benefit of reducing the number of X-ray diffraction data sets needed.

### 4.2. Cisplatin/carboplatin binding to His15 in further detail

Cisplatin binding in this study shows many similarities compared with the previous study (Fig. 1), although the atoms discerned to be bound to the Pt ion do vary somewhat from data set to data set. The occupancy values for the platinum binding to both the N\(^{3}\) atom (69–73%) and the N\(^{4}\) atom (41–47%) are comparable to those observed in the previous study (~65% ± 5% for the N\(^{3}\) binding site and ~50% ± 5% for the N\(^{4}\) binding site), with the exception of the cisplatin aqueous data set, which will be discussed in more detail below (§4.3).

Looking at carboplatin binding to the N\(^{3}\) binding site, there is a difference between this study and the previous results after only 8 d of crystal growth, as only the two N atoms of the complex were observed. Now, however, we can see evidence for a portion of the CBDC moiety and for one N atom bound to the Pt ion (Fig. 2a). In comparing the N\(^{3}\) binding sites in the RT 13 months prolonged chemical exposure study and the cryocooled data set after 8 d crystallization (Fig. 3a), it was noted above that the NH1 atom of Arg14 had shifted downwards by 0.8 Å. Also, the Arg14 NH1 atom had shifted to the left and downwards by 1.8 Å compared with the cryocooled data set after 13 months prolonged chemical exposure (Fig. 3b). These shifts are significant (0.8 ± 0.2 and 1.8 ± 0.4 Å) based on the standard deviations of the coordinate error for the NH1 atom in both structures using (1) and (2). This significant atomic shift observed in the RT data set corresponds to being able to observe extra detail in the F\(_{o}\) − F\(_{c}\) OMIT electron-density map around the N\(^{3}\) binding site, and thus extra atoms can be modelled into the electron density (Fig. 2). The Arg14 residue does not shift downwards between the two cryocooled data sets; it has only shifted sideways, meaning that both the cryocooled data sets have similar electron density and only the Pt ion and two N atoms can be modelled (Fig. 2b and 2c).

At RT we observe that Arg14 moves out of the way, leading to more detail being observed in the electron-density maps at the N\(^{3}\) binding site. This must have taken place in the cryocooled crystals because all the crystals were cocrystallized at RT; it appears that on cryocooling the crystals become disordered or the Arg14 residue is frozen in a different conformation, therefore obscuring details in the electron-density maps. Hence, collecting the data at RT has resulted in more detail being observed at the N\(^{3}\) binding site owing to the conformation of the Arg14 residue not being affected by the data-collection temperature, as was the case in the 100 K structure. The resolution of the RT 13 months prolonged chemical exposure data set extended to 2.0 Å, the 13 months prolonged chemical exposure cryocooled data set extended to 2.08 Å resolution (Table 1) and the 8 d cryocooled data set extended to 1.6 Å resolution (Tanley et al., 2012). Thus, diffraction resolution is not the reason why more detail was observed at RT, as the two cryocooled data sets show very similar electron density at the N\(^{3}\) binding site (Fig. 2) even with these different resolution limits. Another possibility is that the observation of extra detail after 13 months of prolonged chemical exposure could arise from a change in pH over time. However, measuring the pH of each crystallization pot resulted in the same pH being observed (pH 4.7), except for cisplatin in 5% DMSO, which had a pH of 5.0. Also, as the occupancy values and the electron-density maps of cisplatin binding to His15 were very similar for both the short and prolonged chemical exposures, a change in pH can be ruled out as the reason for the observation of extra detail.

We should also consider the possibility of relatively poor solubility, as cisplatin is known to have a low solubility of 0.25 mg ml\(^{-1}\) in water and of 1 mg ml\(^{-1}\) in 0.9% NaCl. The conditions used here were 3 mg ml\(^{-1}\) cisplatin for all cases apart from the cisplatin with NAG cocrystallization, where cisplatin was used at 1.1 mg ml\(^{-1}\), but this is still a threefold molar excess over the protein. A 1 mg ml\(^{-1}\) solution of cisplatin equates to a molar concentration of 3.3 mM, which is the same molar concentration as that of HEWL in all conditions apart from the cryocooled data sets.
from the cocrystallization of cisplatin with NAG. Hence, if only 1 mg of cisplatin was soluble to start with, and the increased chemical exposure caused more cisplatin/carboplatin to dissolve, this could cause the extra detail that was observed compared with 8 d of crystallization and also increased occupancy values. However, we do not observe an increase in occupancy after prolonged chemical exposure, with the exception of the cisplatin aqueous crystallization conditions, which are described in more detail in §4.3. The other crystallizations must have already reached saturation after 8 d of crystallization.

Overall, the temperature of data collection in this study has led to differences being observed in the electron-density maps, with those at RT showing extra detail compared with those at 100 K for the N\textsuperscript{o} binding site.

In the N\textsuperscript{o} binding site (Fig. 2), the OMIT electron-density map detail is comparable between the studies. However, the greater detail observed at the N\textsuperscript{o} binding site in the RT data led us to model in a portion of the CBDC moiety into the N\textsuperscript{o} binding sites of the cryo-cooled data. Again, like cisplatin, the carboplatin Pt occupancies in this study are comparable to the previous results (67% and 51% for the N\textsuperscript{o} and N\textsuperscript{i} binding sites, respectively, compared with 66% and 53% for the cryo-cooled data), meaning that the system had reached chemical equilibrium.

Comparing the RT structures with those of their 100 K counterparts using \textit{LSQKAB} in \textit{CCP4} and plotting the r.m.s. differences between the structures shows some significant changes (Supplementary Table S2). Comparisons made were for carboplatin/DMSO (Supplementary Fig. S2 comparing the RT 13 months prolonged chemical exposure case with the 100 K 13 months prolonged chemical exposure case; Supplementary Fig. S3 compares the RT 13 months prolonged chemical exposure case with the 100 K 8 d crystallization case, cisplatin/DMSO (Supplementary Fig. S4), cisplatin/aqueous (Supplementary Fig. S5) and cisplatin/NAG/DMSO (Supplementary Fig. S6). Each supplementary figure shows the r.m.s. differences between the structures and the average B factors per residue for each structure. The plots of the B factors for each pair of structures shows the same overall trend, with the RT data sets having higher B factors overall, as expected. The significant changes in the r.m.s. difference plots occur at Gly71 in all comparisons; however, there are also some other significant changes which are given in Supplementary Table S2. All of the residues that show these significant changes between the 100 K and RT structures are parts of loop regions in HEWL. To our knowledge, these regions of HEWL are not chemically significant and also are not likely to be involved in the differences observed owing to the extra detail at the N\textsuperscript{o} binding site.

RT data collection, of course, has the disadvantage of potentially increased X-ray radiation damage compared with data collection at 100 K. No radiation damage was observed in our data sets, but the \( R_{merge} \) values (Table 1) for each crystal are high, which is a possible sign of radiation damage. The main reason for this, however, was significant data-scaling problems, most likely because the crystals did not sit completely still in their capillary. Additionally, the RT crystals do not diffract to a very high resolution, so that low \( R_{merge} \) values would not be expected. Also, whilst the apparatus is very good the X-ray source is not a rotating anode and thus the X-ray beam intensity is weaker than what would be considered the usual norms for, for example, synchrotron data.

Another concern could be chemically specific radiation damage, which we now discuss. For example, Ramagopal \textit{et al.} (2005) showed that protein–heavy atom bonds can be severely affected by X-ray radiation damage; thus, not taking the radiation-damage effect into account could cause a problem in the occupancy values of heavy-metal atoms or metal-containing ligands. In fact, we have investigated this in detail in a separate paper (Helliwell & Tanley, 2012) for the case of cisplatin binding to His15 of HEWL after continued X-ray irradiation. In that study, we collected four repeat X-ray diffraction data sets from one crystal; our results show that cisplatin is stable and also that the occupancy values and electron-density maps over the four data sets are the same as the first data set and are comparable to the occupancies presented here and previously (Tanley \textit{et al.}, 2012). Thus, even if radiation damage was the cause of the high \( R_{merge} \) Values observed here, cisplatin itself is not affected by radiation damage and nor are the occupancy values.

### 4.3. Cisplatin binding in aqueous conditions

Cisplatin binding to His15 was not observed in aqueous conditions after 4 d of crystal growth. However, X-ray diffraction studies of these crystals after 15 months of prolonged chemical exposure showed cisplatin binding to both binding sites of His15. The occupancy values obtained for both binding sites (N\textsubscript{i} = 50% and N\textsubscript{o} = 33%) are slightly lower than those obtained for the DMSO conditions. Thus, this system of chemical conditions presumably still has not reached a chemical equilibrium, \textit{i.e.} this is speeded up in the presence of DMSO.

### 4.4. Stability of cisplatin and carboplatin over prolonged storage

A study by Kristjansson \textit{et al.} (1988) looked at the stability of cisplatin in pharmaceutical formulations and whether cisplatin formed toxic oligomers over a period of one year. Pharmaceutical formulations of cisplatin contain 0.9% NaCl solution. The study by Kristjansson and coworkers determined that the potentially toxic oligomeric products were unstable in NaCl and that solutions of these oligomeric compounds were converted back to cisplatin in NaCl conditions. Our study used a 10% NaCl solution in the crystallization conditions; thus, at this high NaCl concentration cisplatin would be stable and would not form these potentially toxic products.

Carboplatin stability has been studied by Sewell \textit{et al.} (1987) and Gust & Schnurr (1999). The study by Sewell \textit{et al.} (1987) suggested that carboplatin reconstituted in water at a concentration of 10 mg ml\textsuperscript{-1} is stable for 14 d. On the other hand, Gust & Schnurr (1999) tested the stability of carboplatin (1 mg ml\textsuperscript{-1}) in 0.9% NaCl solution; after 28 d of storage at RT they found that 10% of the carboplatin was converted to cisplatin, while only 3% was converted after storage at 277 K for 28 d. The carboplatin crystals used in our study were stored at RT for 13 months and had a 10% NaCl solution added to the crystallization conditions. Hence, it would be expected that most of the carboplatin would be converted to cisplatin after such a prolonged length of storage; however, the binding sites after 13 months prolonged chemical exposure with both cryo and RT data collection show evidence for the CBDC moiety; hence, no conversion is seen in this study. This lack of conversion could be a consequence of the binding of carboplatin to His occurring very quickly in solution and of the formation of crystals within days; hence, the carboplatin was ‘captured’ in its starting chemical form.

### 4.5. Impact of these results

The result of cisplatin binding to HEWL in aqueous media after prolonged crystal storage shows that over time cisplatin can bind to HEWL and possibly other proteins which have free histidine residues (Zimmermann & Burda, 2010; Hahn \textit{et al.}, 2001), \textit{i.e.} even in the absence of DMSO. This could be important when considering treatment lengths for patients who are given cisplatin, as over a prolonged...
period of time cisplatin will bind to proteins in the body if it is not broken down and expelled before it accumulates.

5. Conclusions

Over a prolonged period of crystallization and chemical exposure (15 months), cisplatin binds to His15 of HEWL in aqueous media, as opposed to a lack of binding after 4 d of crystallization and crystal growth. The use of RT data collection compared with cryocooling and X-ray diffraction data collection at 100 K led to more detail being observed for the CBDC moiety of carboplatin in the N° binding site electron-density map, thus helping us to discern more detail in the N° binding site electron-density map for the previous cryocooled data sets (Tanley et al., 2012). Comparing carboplatin binding at the N° binding site between the 13 months prolonged chemical exposure RT data set, the 8 d crystallization cryocooled data set and the 13 months prolonged chemical exposure cryocooled data set shows that data collection at RT and not the length of the crystallization time causes increased detail in the electron-density map of the ligand to be observed.

Are our results here widely applicable, notably to human proteins and in a medical context? The main point is the chemical one that the interaction of histidine with these two platinum compounds is neatly observed. The main point is the chemical one that the interaction of histidine with these two platinum compounds is neatly investigated via His15 in HEWL. This is quite generic and we believe that it is indeed, therefore, widely applicable. Of course, the crystallization mixture is chemically specific for HEWL.

JRH is grateful to the University of Manchester for general support, to the ESPRC for a PhD studentship to SWMT and to the School of Chemistry for crystallization, computing and X-ray facilities. We are grateful to Dr Chris Muryn of the School of Chemistry for his stewardship of the APEX II CCD X-ray diffractometer.

References


Supplementary Materials

**Figure S7.1** Chemical structures of cisplatin and carboplatin

![Chemical structures](image)
Figure S7.2 (a) RMS difference plot between the RT and 100K structures for the carboplatin_DMSO case after 13 months of prolonged chemical storage. Asn-46, Try-47, Asp-48, Pro-70 and Gly-71 show the biggest RMS differences (but see Table S7.2 for the significant statistical evaluations) and these residues all lie in loop regions of the protein and are, to our knowledge, not chemically significant. However, the shift between Pro-70 is not significant to a 3 σ level. (b) Average B factors for the 100K crystal structure of carboplatin_DMSO case after 13 months of prolonged chemical storage and (c) the average B factors for the RT crystal structure of carboplatin_DMSO after 13 months prolonged chemical storage. The trend in the B factors for both structures is the same, with the RT crystal structure having overall higher B factors as expected.

![Main Chain and Side chain deviations](image1)

![Average Bfactors (all atoms) Chain A](image2)

![Average Bfactors (all atoms) Chain A](image3)
Figure S7.3 (a) RMS difference plot between the RT carboplatin_DMSO case after 13 months of prolonged chemical storage and the 100K Carboplatin_DMSO_Glycerol case after 8 days of crystallisation. Gly 71 again shows the biggest RMS differences between the structures (labelled). (b) Average B factors for the RT crystal structure of carboplatin_DMSO case after 13 months of prolonged chemical storage and (c) the average B factors for the 100K crystal structure of carboplatin_DMSO_Glycerol after 8 days of co-crystallisation. The trend in the B factors for both structures is the same, with the RT crystal structure having overall higher B factors as expected.
Figure S7.4 (a) RMS difference plot between the RT cisplatin_DMSO case after 13 months of prolonged chemical storage and the 100K cisplatin_DMSO_Glycerol case after 8 days of crystallisation. Gly 71 again shows the biggest RMS differences between the structures (labelled), but this shift is not significant at the 3σ level. (b) Average B factors for the RT crystal structure of cisplatin_DMSO case after 13 months of prolonged chemical storage and (c) the average B factors for the 100K crystal structure of cisplatin_DMSO_Glycerol after 8 days of co-crystallisation. The trend in the B factors for both structures is the same, with the RT crystal structure having overall higher B factors as expected.
Figure S7.5 (a) RMS difference plot between the RT cisplatin_aqueous case after 13 months of prolonged chemical storage and the 100K cisplatin_aqueous_Glycerol case after 8 days of crystallisation. Gly 71, Asn 103, Val 109 and Ala 110 shows the biggest RMS differences between the structures (labelled) and these residues lie in loop regions of the protein and to our knowledge, are not chemically significant. However, these changes are not significant at a 3 $\sigma$ level. (b) Average B factors for the RT crystal structure of cisplatin_aqueous case after 13 months of prolonged chemical storage and (c) the average B factors for the 100K crystal structure of cisplatin_aqueous_Glycerol after 4 days of co-crystallisation. The trend in the B factors for both structures is the same, with the RT crystal structure having overall higher B factors as expected.
Figure S7.6 (a) RMS difference plot between the RT cisplatin_NAG_DMSO case after 13 months of prolonged chemical storage and the 100K cisplatin_NAG_DMSO_Glycerol case after 8 days of crystallisation. Gly 71 and Ser-72 show the biggest RMS differences between the structures (labelled) and these residues lie in loop regions of the protein and to our knowledge, are not chemically significant. (b) Average B factors for the RT crystal structure of cisplatin_NAG_DMSO case after 13 months of prolonged chemical storage and (c) the average B factors for the 100K crystal structure of cisplatin_NAG_DMSO_Glycerol after 8 days of co-crystallisation. The trend in the B factors for both structures is the same, with the RT crystal structure having overall higher B factors as expected.
### Table S7.1

Data collection strategy for each crystal

<table>
<thead>
<tr>
<th></th>
<th>Swing around (degrees)</th>
<th>Sweep (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cisplatin_aq</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>189.0ω</td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>101.0Φ</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>101.0Φ</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>156.5ω</td>
<td></td>
</tr>
<tr>
<td>-15.0</td>
<td>174.0Φ</td>
<td></td>
</tr>
<tr>
<td><strong>Cisplatin_DMSO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-15.0</td>
<td>173.5ω</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>100.0ω</td>
<td></td>
</tr>
<tr>
<td>-15.0</td>
<td>117.5Φ</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>152.0Φ</td>
<td></td>
</tr>
<tr>
<td>-5.0</td>
<td>157.0Φ</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>174.5Φ</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>189.0ω</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>292.0Φ</td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>329.0Φ</td>
<td></td>
</tr>
<tr>
<td><strong>Cisplatin_DMSO_NAG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-25.0</td>
<td>360.0Φ</td>
<td></td>
</tr>
<tr>
<td>-25.0</td>
<td>188.5ω</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>188.5ω</td>
<td></td>
</tr>
<tr>
<td>30.0</td>
<td>103.5ω</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>73.0Φ</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>90.0Φ</td>
<td></td>
</tr>
<tr>
<td>30.0</td>
<td>27.5Φ</td>
<td></td>
</tr>
<tr>
<td><strong>Carboplatin_DMSO_RT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>176.0ω</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>89.5Φ</td>
<td></td>
</tr>
<tr>
<td>-5.0</td>
<td>115.0ω</td>
<td></td>
</tr>
<tr>
<td>-10.0</td>
<td>138.0ω</td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>141.0ω</td>
<td></td>
</tr>
<tr>
<td>-25.0</td>
<td>348.0Φ</td>
<td></td>
</tr>
<tr>
<td>Carboplatin_DMSO_100K</td>
<td>-25.0</td>
<td>331.5 Φ</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>-5.0</td>
<td>124.0 Φ</td>
</tr>
<tr>
<td></td>
<td>-5.0</td>
<td>152.5 Φ</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>359.5 Φ</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>73.0 Φ</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>360.0 Φ</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>275.0 Φ</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>359.5 Φ</td>
</tr>
</tbody>
</table>

Table S7.2: Significant RMS differences (at 3 σ level) between RT and 100K structures for each pair wise comparison with a standard deviation value for each calculated using equations 1 and 2.

<table>
<thead>
<tr>
<th>Pair-wise Comparisons</th>
<th>Significant changes to amino acids</th>
<th>RMS difference (Å) between structures with standard deviations calculated from equations 1 and 2.</th>
</tr>
</thead>
</table>
| RT prolonged chemical exposure carboplatin DMSO with 100K prolonged chemical exposure carboplatin DMSO | Asn 46 0.8 +/- 0.3  
Thr 47 1.2 +/- 0.3  
Asp 48 0.9 +/- 0.3  
Gly 71 1.1 +/- 0.3 |                                                                                                  |
| RT prolonged chemical exposure carboplatin DMSO with 100K 8 day crystallisation carboplatin DMSO Glycerol | Gly 71 0.8 +/- 0.2 |                                                                                                  |
| RT prolonged chemical exposure cisplatin NAG DMSO with 100K 8 day crystallisation cisplatin NAG DMSO Glycerol | Gly 71 0.9 +/- 0.3  
Ser 72 0.8 +/- 0.3 |                                                                                                  |
8. Chapter 8: The crystal structure analysis of the relative binding of cisplatin and carboplatin in a mixture with histidine in a protein studied at 100 and 300 K with repeated X-ray irradiation

John R. Helliwell\textsuperscript{a}\textsuperscript{*} and Simon W.M. Tanley\textsuperscript{a}

\textsuperscript{a}School of Chemistry, University of Manchester, Brunswick street, Manchester, Greater Manchester, M13 9PL, England

Correspondence email: john.helliwell@manchester.ac.uk

Acta Cryst D (2013) 69, 121-125

Keywords: Cisplatin; Carboplatin; Histidine; Relative toxicity; Binding occupancy; X-ray radiation damage; Radiation therapy; Temperature variation and structure.

JRH conceived the study themes. SWMT grew all crystals used in this study. JRH mounted the crystals and collected data on the Bruker APEX II home source diffractometer and set up the four repeated dataset collections on the room temperature crystal. SWMT processed the datasets using the Bruker internal software program and refined all structures presented in this paper. SWMT wrote the article, with JRH editing the manuscript.
The crystal structure analysis of the relative binding of cisplatin and carboplatin in a mixture with histidine in a protein studied at 100 and 300 K with repeated X-ray irradiation

The anticancer agents cisplatin and carboplatin bind to histidine in a protein. This crystal structure study at data-collection temperatures of 100 and 300 K examines their relative binding affinities to a histidine side chain and the effect of a high X-ray radiation dose of up to ∼1.8 MGy on the stability of the subsequent protein–Pt adducts. Cisplatin binding is visible at the histidine residue, but carboplatin binding is not. Five refined X-ray crystal structures are presented: one at 100 K as a reference and four at 300 K. The diffraction resolutions are 1.8, 2.0, 2.8, 2.9 and 3.5 Å.

1. Introduction

Cisplatin and carboplatin (Supplementary Fig. S1) are platinum anticancer drugs which bind to the N7 atoms of guanine bases in DNA, leading to inhibition of DNA replication and transcription and triggering apoptotic cell death (Benedetti et al., 2002; Silverman et al., 2002). However, 90% of their reported binding is to plasma proteins (Fischer et al., 2008). Thus, these drugs cause serious side effects to patients. Cisplatin is rapidly converted to toxic metabolites which cause nephrotoxic effects (Zhang & Lindup, 1996; Hulíčiak et al., 2012), whereas carboplatin has a slower rate of conversion to toxic metabolites owing to the addition of the cyclobutane-dicarboxylate (CBDC) moiety (Supplementary Fig. S1) and hence is tolerated at higher chemical doses compared with cisplatin (Kostova, 2006). Even though these side effects are observed, both cisplatin and carboplatin remain in use for the treatment of testicular, bladder, ovarian and lung cancers, to name but a few (Wang & Lippard, 2005; Ivanov et al., 1998).

Cisplatin treatment is usually given to patients in combination with radiation therapy (Reedijk, 2003; Wang & Lippard, 2005; Vallerga et al., 2004) as this drug is a radiation sensitizer (Peters et al., 2000), with standard treatment being to alternate between chemotherapy and radiation therapy. The concurrent treatment of cancers with both radiation therapy and cisplatin administration causes an above-additive effect, in which interaction with the radiation field leads to increased killing of cancer cells compared with each treatment individually (Seiwert et al., 2007). The combination of cisplatin and radiation therapy uses several mechanisms to enhance cell killing, including the enhanced formation of toxic platinum intermediates in the presence of radiation-induced free radicals, inhibition of DNA repair, a radiation-induced increase in cellular platinum uptake and cell-cycle arrest (Lawrence et al., 2003; Kvols, 2005).

Received 14 August 2012
Accepted 25 October 2012

PDB References: cisplatin/carboplatin mixture binding to HEWL, 4gcb; 4gcc; 4gce; 4gcf

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved
In terms of crystallography, a previous study (Tanley, Schreurs, Kroon-Batenburg, Meredith et al., 2012) has shown that cisplatin and carboplatin bind to both the N\textsuperscript{4} and N\textsuperscript{7} atoms of His15 in hen egg-white lysozyme (HEWL), a model protein, in DMSO medium after 8 d of cocystallization. A further study (Tanley, Schreurs, Kroon-Batenburg & Helliwell, 2012) has used both 300 K X-ray diffraction data collection and 100 K X-ray diffraction data collection for carboplatin/DMSO crystals after 13 months of chemical exposure. More detail was observed at the N\textsuperscript{7} binding site for the CBDC 'tartratoplatinum(II) complexes in methanol/water solutions which increases their activity in killing tumour cells by binding to DNA. The radiation reduction of bis(1-ethylimidazole)-tartratoplatinum(II) complexes in methanol/water solutions has been examined (Kalecinka et al., 1997); this study found that the yields of platinum(I) depended on the presence of O\textsubscript{2} in solution as well as the dose applied (up to 2 kGy). Our study reported here provides direct experimental structural details of the X-ray radiation-sensitive nature of cisplatin when bound to a model protein at 300 K, which is important for directly determining the relative binding affinities of this anticancer drug to a model protein as well as the radiation stability of the histidine and its complexation.

### 2. Methods

#### 2.1. Crystallization conditions

49 mg HEWL (3.2 mmol) was co-crystallized using the batch method with both 3 mg cisplatin (10 mmol) and 3.7 mg carboplatin (10 mmol) as well as 0.04 M sodium acetate buffer pH 4.7 (in 462.5 µl) and 10% NaCl precipitant (in 462.5 µl). Data sets were collected from the same crystal at 300 K in order to understand the relative binding affinities of the mixture of cisplatin and carboplatin under continued X-ray irradiation. It is known that cisplatin and carboplatin are radiation sensitizers, as described above, and thus they are believed to be chemically stable during radiation therapy, which increases their activity in killing tumour cells by binding to DNA.

<table>
<thead>
<tr>
<th>Data-set code</th>
<th>Data set 1</th>
<th>Data set 2</th>
<th>Data set 3</th>
<th>Data set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code</td>
<td>4gcb</td>
<td>4gcc</td>
<td>4gcd</td>
<td>4gce</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data reduction</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>(^{4}\text{P}=2,\text{2},\text{2})</td>
<td>(^{4}\text{P}=2,\text{2},\text{2})</td>
<td>(^{4}\text{P}=2,\text{2})</td>
<td>(^{4}\text{P}=2,\text{2})</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, (^{\circ}))</td>
<td>(a = b = 78.78, c = 36.98, \alpha = \beta = \gamma = 90.0)</td>
<td>(a = b = 79.86, c = 37.80, \alpha = \beta = \gamma = 90.0)</td>
<td>(a = b = 79.86, c = 37.80, \alpha = \beta = \gamma = 90.0)</td>
<td>(a = b = 79.87, c = 37.76, \alpha = \beta = \gamma = 90.0)</td>
</tr>
<tr>
<td>Molecular mass (Da)</td>
<td>14700</td>
<td>14700</td>
<td>14700</td>
<td>14700</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Diffractometer</td>
<td>Bruker PLATINUM 135</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
</tr>
<tr>
<td>Crystal-to-detector distance (mm)</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>1285547</td>
<td>130777</td>
<td>82393</td>
<td>73601</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>11301</td>
<td>8718</td>
<td>3318</td>
<td>2982</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>55.71–1.80 (1.85–1.80)</td>
<td>56.47–2.00 (2.05–2.00)</td>
<td>56.47–2.80 (2.87–2.80)</td>
<td>56.47–2.90 (2.97–2.90)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (99.0)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
<td>99.7 (99.2)</td>
</tr>
<tr>
<td>R(_{merge}) (%)</td>
<td>0.069 (0.305)</td>
<td>0.202 (0.796)</td>
<td>0.195 (0.575)</td>
<td>0.230 (0.627)</td>
</tr>
<tr>
<td>Mean (F_t/F_i) (ε)</td>
<td>113.6 (44.2)</td>
<td>123.3 (1.42)</td>
<td>175.8 (4.5)</td>
<td>147.3 (3.5)</td>
</tr>
<tr>
<td>Cruickshank DPI (Å)</td>
<td>0.14</td>
<td>0.14</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Average B factor (Å(^{2}))</td>
<td>13.2</td>
<td>26.3</td>
<td>34.7</td>
<td>36.9</td>
</tr>
</tbody>
</table>

\(^{\dagger}\) The disallowed residue in the Ramachandran plot is Gly102, which is part of a loop region of the protein.
with 1 mM DMSO (75 ml) and left at 277 K. An overall sixfold molar concentration of the Pt compounds over that of the protein was used.

2.2. X-ray data collection, structure solution and refinement

For the cryo data collection, the Bruker temperature-control device was set to 100 K and a crystal of 0.15 mm in size after 8 d of crystal growth was cryoprotected using silicone oil and centred on an in-house Bruker PLATINUM 135 detector at a distance of 50 mm from the detector. An X-ray exposure time of 10 s per 0.5° crystal rotation range was used with an X-ray wavelength of 1.5418 Å.

For the room-temperature (RT) data collection, the Bruker temperature-control device was set to 300 K and a crystal of 0.15 mm in size after 10 d of crystal growth was mounted in a 1 mm quartz capillary tube and centred on an in-house Bruker APEX II detector at a distance of 40 mm from the detector. Each data set consisted of a full 360° rotation of the crystal comprising 40 s X-ray exposure per 0.5° crystal rotation range for data sets 1 and 2 and 60 s per 0.5° crystal rotation range for data sets 3 and 4 with an X-ray wavelength of 1.5418 Å. φ scans were used for data collection in all cases (Tanley, Schreurs, Kroon-Batenburg & Helliwell, 2012).

The 100 K data set was processed using the PROTEUM II internal software package on the Bruker PLATINUM 135 detector. The 300 K data sets were processed using the APEX II internal software package on the Bruker APEX II detector; all four data sets from the 300 K data collection were processed and integrated separately. The unit-cell parameters were checked for X-ray damage effects, which showed possible significant changes for RT data set 4 (see Supplementary Table S1). All crystal structures were solved using molecular replacement with Phaser (McCoy et al., 2007) and restrained refinement with TLS (except for the 100 K data set, which was solved with restrained refinement only and no TLS) in REFMAC5 (Vagin & Teplyakov, 2010) in CCP4 using the reported lysozyme structure 2w1y (Cianci et al., 2008) as a molecular search model. Model building, adjustment and refinement were carried out using the Coot (Emsley & Cowtan, 2004) molecular-graphics program and REFMAC5 (Vagin & Teplyakov, 2010), respectively, in CCP4. Crystallographic and refinement parameters are summarized in Table 1. Ligand-binding occupancies were calculated using SHELXTL (Sheldrick, 2008) and are given in Table 2.

3. Results

3.1. Binding at His15

Cisplatin rather than carboplatin is observed bound to both the Nδ and Nε atoms of the imidazole ring of His15 in both the 100 and 300 K data sets. Fig. 1 shows the details of this binding of cisplatin at 100 K (Fig. 1a) and for the first data set collected at 300 K (Fig. 1b). In the Nδ binding site, clear anomalous difference electron density is observed for the Cl atoms in the 100 K structure, with each atom being distinct, whereas for the 300 K data set 1 the Cl atoms are less well defined, but some anomalous electron density is also observed for them here. However, the Nε binding site is more difficult to assign as either cisplatin or carboplatin. In the 100 K data set a distinct Cl atom is observed, but the anomalous difference density for this atom is weak. Data set 1 collected at 300 K shows less detail in the Nε binding site; hence, only a Pt ion can be assigned into the density. The other three data sets collected at 300 K were collected to lower resolutions (2.8, 2.9 and 3.5 Å); they also all show clear binding to both binding sites (Supplementary Fig. S2) supported by anomalous difference electron density for the Pt ion, but the data are of poorer quality owing to overall radiation damage. Hence, the data are too poor to provide detailed information as to the coordinated atoms beyond the Pt ion. Hence, it is difficult to assign whether cisplatin or carboplatin is bound to the Nε atom of His15.

Table 2

<table>
<thead>
<tr>
<th>Data set</th>
<th>Nδ binding site</th>
<th>Nε binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 K data collection</td>
<td>53</td>
<td>40</td>
</tr>
<tr>
<td>300 K data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data set 1</td>
<td>65</td>
<td>64</td>
</tr>
<tr>
<td>Data set 2</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Data set 3</td>
<td>67</td>
<td>56</td>
</tr>
<tr>
<td>Data set 4</td>
<td>71</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 2: The Pt-occupancy values (%) calculated from SHELXL (Sheldrick, 2008).

Based on Tanley, Schreurs, Kroon-Batenburg, Meredith et al. (2012), the sigmas on these occupancies are estimated to be ~5%.

Figure 1

Cisplatin binding to the Nδ and Nε atoms of His15. (a) 100 K data collection at 1.8 Å resolution, (b) 300 K data set 1 at 2.0 Å resolution. The 2F₀ − Fc map (in purple) is at the 1.5 r.m.s. cutoff level and the anomalous difference density map (in orange) is at the 3σ cutoff level. The Nδ and Nε atoms in the imidazole ring are labelled, together with the Pt and Cl atoms of the bound cisplatin moiety. The Cl occupancies calculated from SHELXL are 73 and 51% for Cl1 and Cl2 in the 100 K data set and 78 and 68% for Cl1 and Cl2 in the RT data set 1. Cl3 and Cl4 have occupancies of 25 and 36%, respectively, in the 100 K data set.

Figure 1

(a) 100 K data collection at 1.8 Å resolution, (b) 300 K data set 1 at 2.0 Å resolution. The 2F₀ − Fc map (in purple) is at the 1.5 r.m.s. cutoff level and the anomalous difference density map (in orange) is at the 3σ cutoff level. The Nδ and Nε atoms in the imidazole ring are labelled, together with the Pt and Cl atoms of the bound cisplatin moiety. The Cl occupancies calculated from SHELXL are 73 and 51% for Cl1 and Cl2 in the 100 K data set and 78 and 68% for Cl1 and Cl2 in the RT data set 1. Cl3 and Cl4 have occupancies of 25 and 36%, respectively, in the 100 K data set.
present, but owing to the observation of cisplatin binding in the 100 K structure it is assumed that cisplatin rather than carboplatin is also bound in the 300 K structure. A whole cisplatin molecule was initially centred on the Pt position for both the N\textsuperscript{4} and N\textsuperscript{N} binding sites and its coordinated atoms were removed based on electron density not being observed. The Pt–His N distances were not restrained during refinement. The occupancy values of the Pt ions at each binding site in the 100 K and the four 300 K data sets are given in Table 2.

The data set at 100 K has an extra atom visible near the Arg14 residue (Fig. 1a). This is assumed to be likely to be a Cl atom. HEWL has the six usual chloride-binding sites and in this 100 K data set these have anomalous difference electron-density peak heights of between 5σ and 11σ, with the new site having a peak-height value of 9.6σ. However, the 2F\textsubscript{o} − F\textsubscript{c} electron density for this Arg14 residue is weak and thus it is difficult to assign this density to the Arg14 side chain or to a molecule of water. Assuming the presence of an arginine, the B factors are between 24 and 34 Å\textsuperscript{2}.

3.2. Absorbed X-ray radiation dose for the 300 K crystal study

The X-ray diffraction data collected at 300 K were partitioned into four data sets, as described in §2.2. This data-collection strategy was used to confirm how the cisplatin/carboplatin molecule is when bound to a histidine residue in a model protein, as it is known that these drugs are radiation sensitizers and thus are used in conjunction with radiation therapy. This crystal was not ‘back-washed’ before data collection; hence, the solvent channels of the protein crystal contained Pt ions. Taking into account explicitly the solvent content as well as the protein content, and the bound cisplatin and DMSO molecules to the protein, the absorbed X-ray radiation dose is 0.36 MGy for data sets 1 and 2 and 0.53 MGy for data sets 3 and 4. This thus gives an overall absorbed radiation dose of 1.78 MGy for this crystal. The dose calculation and equations used are given in Supplementary Table S3. Supplementary Fig. S3 shows the F\textsubscript{o} − F\textsubscript{c} electron-density maps after rigid-body refinement for all four data-set runs but with data sets 1, 2 and 3 cut at 3.5 Å resolution (the resolution limit of data set 4) and Supplementary Fig. S4 shows data sets 1 and 2 cut at 2.9 Å resolution (the resolution limit of data set 3). These figures confirmed that the electron-density detail for each data-set experiment is exactly the same, which was confirmed by the Pt-occupancy values (Table 2). The occupancies do differ slightly, but they are not greatly different and are in the range ±5% (Tanley, Schreurs, Kroon-Batenburg, Meredith et al., 2012); this is thus in agreement with the principle that cisplatin is a radiation sensitizer and is stable during repeated X-ray irradiation. Furthermore, we observed that its binding to histidine remained throughout this X-ray dose.

4. Discussion

This study uses the same molar concentration of cisplatin and carboplatin to determine which of the anticancer compounds has higher affinity for binding to a histidine residue in a model protein. Additionally, this study also looked at the stability of the bound compound (cisplatin) to continued X-ray irradiation. Finally, the stability of the attachment of cisplatin to His15 is assessed. At the diffraction resolutions used here (i.e. ~2 Å), it is difficult to detect whether any reduction of Pt\textsuperscript{2+} to Pt\textsuperscript{+} occurs in this study; this would have to be assessed in a future study at a higher resolution.

4.1. Cisplatin rather than carboplatin binding is preferred

A threefold molar excess of both cisplatin and carboplatin was used in the cocrystallization with HEWL, giving an overall sixfold molar excess of the Pt compounds over the protein. Both the 300 and 100 K X-ray diffraction data sets showed evidence for binding of cisplatin over carboplatin in the N\textsuperscript{4} binding site, as indicated by the presence of anomalous difference electron density for the Cl atoms bound to the Pt ion (Figs. 1a and 1b). Cisplatin also bound to the N\textsuperscript{N} binding site in the 100 K structure (Fig. 1a), as again indicated by the presence of anomalous difference electron density for a Cl atom. However, for the 300 K X-ray diffraction data sets it was difficult to assign which Pt compound bound to the N\textsuperscript{N} atom owing to the deteriorating quality of the data under repeated X-ray irradiation, meaning that only a Pt ion could be modelled into the density and no specific coordinated species could be assigned. However, owing to the fact that cisplatin rather than carboplatin was observed to be clearly bound in the 100 K structure, it can be deduced that cisplatin does have an overall higher binding affinity for the His side-chain N\textsuperscript{4} and N\textsuperscript{N} atoms compared with carboplatin in this model study. For the binding of two Pt ions to occur to the imidazole ring of this histidine side chain, both of the N atoms are sp\textsuperscript{2}-hybridized with N lone pairs in the plane of the imidazole ring, providing two N atoms to which a metal centre can bind (Tanley, Schreurs, Kroon-Batenburg, Meredith et al., 2012). The usual N-hydrogen of one of the histidine N atoms has to be removed and this could be facilitated by the crystallization conditions used, which contained both chloride and acetate ions.

4.2. Absorbed X-ray radiation dose of the 300 K crystal

This study also looked at the stability of cisplatin towards continued X-ray irradiation once bound to the histidine side chain. Under continued X-ray irradiation in the 300 K study, the crystal absorbed 1.78 MGy of radiation (Supplementary Table S3) over the four X-ray diffraction data sets collected, and the electron-density maps for the bound cisplatin molecules looked very similar even at the differing resolutions of each data set (Supplementary Figs. S3 and S4). The results confirmed that cisplatin is stable over prolonged exposure to X-ray irradiation as it does not dissociate from the histidine residue, meaning that its relative binding affinity to this model protein is fixed. This result was backed up by the Pt-occupancy values for each 300 K X-ray diffraction data set (Table 2), which do not vary significantly (they are within ±5%) between each run. Hence, the relative binding affinity of cisplatin to...
HEWL, our model protein, has not been affected by the quantity of X-ray radiation absorbed.

5. Conclusions
This study showed that cisplatin has a higher relative binding affinity than carboplatin for binding to a histidine side chain in a model protein. Also, cisplatin was stable as a chemical structure upon continued X-ray irradiation; continued binding of cisplatin to the histidine side chain was observed even after prolonged irradiation. This means that the relative binding affinity of cisplatin to a protein is the same at the end of the range of X-ray absorbed dose used in this study. Obviously, it should be mentioned that the chemical mixtures used in the crystallization chemical conditions required for crystallography are not the same as those used when treating a cancer patient; likewise, the precise photon-energy and X-ray dose used in this study are different (see Supplementary Table S3). These differences impose a limitation on the relevance of the conclusions reached here in terms of a cancer treatment scenario.

JRH is grateful to the University of Manchester for general support, to the ESPRC for a PhD studentship to SWMT and to the School of Chemistry for crystallization, computing and X-ray facilities. We are grateful to Dr Chris Muryn of the School of Chemistry for his stewardship of the X-ray CCD APEX II diffractometer.

References
**Supplementary Materials**

**Table S8.1** Unit cell parameter evaluation of the four 300K datasets, with the standard deviations against the unit cell parameters used given also.

<table>
<thead>
<tr>
<th></th>
<th>a (Standard deviation) Å</th>
<th>c (Standard deviation) Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell parameters (Table 1)</td>
<td>79.86</td>
<td>37.80</td>
</tr>
<tr>
<td>Dataset 1</td>
<td>79.62 (0.16)</td>
<td>37.73 (0.04)</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>79.90 (0.02)</td>
<td>37.78 (0.01)</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>79.69 (0.12)</td>
<td>37.68 (0.08)</td>
</tr>
<tr>
<td>Dataset 4</td>
<td>79.25 (0.43)</td>
<td>37.58 (0.15)</td>
</tr>
</tbody>
</table>

Taking into account the standard deviations between the unit cell parameters used in processing the data and the unit cell parameters of each run using the first 20 images, each are within 2 sigma except for dataset 4. Due to this change in dataset 4, dataset 4 was re-integrated and reprocessed using the unit cell parameters of the first 20 images. However, after re-integration and re-processing the unit cell parameters did not differ greatly, as compared with the original unit cell parameters used (a=79.87, c= 37.76 versus other Table 1 values). Hence between the first dataset and the fourth dataset the ‘a’ unit cell parameter was very similar, thus the unit cell parameters used for datasets 2 and 3 must also have been very similar, so that no significant increase in the unit cell was seen at these high radiation doses. The intrinsic variations seen suggest that the quoting of the sigmas to 2 decimal places by the software is over optimistic but are given here for completeness.
Table S8.2 The Pt-N distances in Å are given along with the precisions of these bonds in parentheses based on the Cruickshank DPI coordinates (see Table 1) for each pair of atoms (Pt and N) for each model refinement to give an indication specific to these atoms of the precision of their separation i.e. with their individual B factors taken into account.

<table>
<thead>
<tr>
<th>Model Refinement</th>
<th>Pt to N distance in Nδ binding site</th>
<th>Pt to N distance in Nε binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td>100K data collection</td>
<td>2.1Å (0.3)</td>
<td>2.2Å (0.3)</td>
</tr>
<tr>
<td>300K data collection. Dataset 1</td>
<td>2.3Å (0.4)</td>
<td>2.4Å (0.4)</td>
</tr>
<tr>
<td>300K data collection. Dataset 2</td>
<td>2.6Å (0.5)</td>
<td>2.6Å (0.5)</td>
</tr>
<tr>
<td>300K data collection. Dataset 3</td>
<td>2.3Å (0.7)</td>
<td>2.4Å (0.8)</td>
</tr>
<tr>
<td>300K data collection. Dataset 4</td>
<td>2.4Å (0.8)</td>
<td>2.2Å (0.7)</td>
</tr>
</tbody>
</table>

Note: These Pt to histidine nitrogen distances are unrestrained values from the refinement. The nitrogen atoms within the imidazole ring itself are of course restrained.

Table S8.3 Total absorbed X-ray radiation dose

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Formula used taken into account the solvent contents</td>
<td>C_{61.01} N_{192.01} O_{523.01} S_{11.0003} Na_{10} Cl_{10.014} Pt_{1.27} *</td>
</tr>
<tr>
<td>X-ray Intensity (photons/sec/mm(^2))</td>
<td>5x10(^9) **</td>
</tr>
<tr>
<td>Total exposure time (sec)</td>
<td>Runs 1 and 2 = 28800</td>
</tr>
<tr>
<td></td>
<td>Runs 3 and 4 = 43200</td>
</tr>
<tr>
<td>Sample size (mm)</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Sample cross sectional area presented to the incident X-ray beam (mm(^2))</td>
<td>0.025mm(^2)</td>
</tr>
<tr>
<td>Sample volume (cm(^3))</td>
<td>(0.015 x0.015 x 0.01) cm(^3) = 2.25x10(^{-6}) cm(^3)</td>
</tr>
<tr>
<td>X-ray beam fraction absorbed</td>
<td>$1-e^{-\mu t}$ ($\mu = 1.51 \text{ mm}^{-1}$, $d = 0.1 \text{ mm}$)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>$= 0.14 \ (14%) \ a$</td>
</tr>
<tr>
<td>Photon energy (eV)</td>
<td>8042</td>
</tr>
<tr>
<td>Conversion factor 1 eV to 1 Joule</td>
<td>$1.6 \times 10^{-19} \text{ J}$</td>
</tr>
<tr>
<td>Crystal mass (g) = (density (g/cm$^3$) x volume (cm$^3$))</td>
<td>$0.810 \text{ g/cm}^3 \times 2.25 \times 10^{-6} \text{ cm}^3 = 1.82 \times 10^{-6} \text{ g}$</td>
</tr>
<tr>
<td>Absorbed X-ray dose (J/g); (a) data sets 1 and 2 (b) datasets 3 and 4</td>
<td>(a) $5 \times 10^9 \times 0.025 \times 2.88 \times 10^4 \times 0.14 \times 8042 \times 1.6 \times 10^{-19} / 1.82 \times 10^{-6}$ $= 356.32 \text{ J/g}$ $= 0.36 \text{ MGy}$</td>
</tr>
<tr>
<td></td>
<td>(b) $5 \times 10^9 \times 0.025 \times 4.32 \times 10^4 \times 0.14 \times 8042 \times 1.6 \times 10^{-19} / 1.82 \times 10^{-6}$ $= 534.48 \text{ J/g}$ $= 0.53 \text{ MGy}$</td>
</tr>
<tr>
<td>Individual dataset X-ray absorbed doses</td>
<td>Run 1 = 0.36 MGy</td>
</tr>
<tr>
<td></td>
<td>Run 2 = 0.36 MGy</td>
</tr>
<tr>
<td></td>
<td>Run 3 = 0.53 MGy</td>
</tr>
<tr>
<td></td>
<td>Run 4 = 0.53 MGy</td>
</tr>
<tr>
<td>Total X-ray absorbed Dose</td>
<td>1.78 MGy</td>
</tr>
<tr>
<td></td>
<td>Allowing for errors and inaccurate estimates this sample has absorbed ~1.8 MGy</td>
</tr>
</tbody>
</table>

Absorbed dose = Intensity x exposure time x sample cross sectional area x beam absorbed fraction x photon energy x $1.6 \times 10^{-19} / \text{ Crystal mass}$

The calculation used to obtain the numbers of waters, NaCl ion pairs, DMSO and cisplatin/carboplatin molecules in the solvent portion of the space group’s asymmetric unit was:-

Solvent content fraction x unit cell volume (in Å$^3$) x $10^{-24} \text{ cm}^3$ (to convert the unit cell volume from Å$^3$ to cm$^3$) x Avogadro’s Number x concentration (in Moles) / 1000 (in cm$^3$ i.e. the volume for 1 M).
* The chemical formula for the ordered portion of the asymmetric unit is $C_{609}N_{192}O_{187}S_{11}$ Pt$_{1.2}$

And the chemical formula for the solvent channel derived from the above formula is $C_{2.01}N_{0.01}O_{336.01}S_{0.0003}Na_{10}Cl_{10.014}$ Pt$_{0.07}$

** The incident X-ray intensity estimate is approximately what a modern sealed tube microfocus X-ray source with suitable modern X-ray beam conditioning optics can deliver.

a The linear absorption coefficient was calculated using SHELX for the chemical formula sum of the ordered portion of the unit cell asymmetric unit and the solvent fraction of the asymmetric unit.

The above evaluation and conditions can be compared with those used in cancer patient combination therapy. For curative cancer cases, the typical dose for a solid epithelial tumour ranges from 60 to 80 Gray (Gy), while lymphomas are treated with 20 to 40 Gy. Preventative (adjuvant) doses are typically around 45 – 60 Gy in 1.8 – 2 Gy fractions (for breast, head, and neck cancers.). The photon energy used in radiation therapy is in the voltage range of 4-25 MeV as these penetrate well to deep sites within the body.

![Chemical structures of cisplatin and carboplatin](image)

**Figure S8.1** Chemical structures of cisplatin and carboplatin
Figure S8.2 Cisplatin binding to His-15 of HEWL at RT after continued X-ray irradiation. (a) RT dataset 2, (b) RT dataset 3m and (c) RT dataset 4. The 2Fo-Fc map (in purple) is at the 1.5 rms cut-off level and the anomalous difference density map (in orange) is at the 3σ cut-off level. The Nδ and Nε atoms in the imidazole ring are labelled along with the bound Pt atoms.
**Figure S8.3:** The difference electron density omit maps for the four 300K X-ray data collection runs after rigid body protein model refinement with all cases having their diffraction data cut at 3.5Å resolution. Green is dataset 1, blue dataset 2, red dataset 3 and yellow dataset 4.

**Figure S8.4:** The difference electron density omit maps for the first three datasets at 300K data collection after rigid body protein model refinement all cut at 2.9Å resolution. Green is dataset 1, blue dataset 2 and red dataset 3.
9. Chapter 9: Experiences with archived raw diffraction images data: Capturing cisplatin after chemical conversion of carboplatin in high salt conditions for a protein crystal

Simon W.M. Tanley\textsuperscript{a}, Kay Diederichs\textsuperscript{b}, Loes M.J. Kroon-Batenburg\textsuperscript{c}, Antoine M.M. Schreurs\textsuperscript{c}, John R. Helliwell\textsuperscript{a}\textsuperscript{*}.

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, UK

\textsuperscript{b} Department of Biology, University of Konstanz, Germany

\textsuperscript{c} Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

\textsuperscript{*}Correspondence email: john.helliwell@manchester.ac.uk


Keywords: Cisplatin; carboplatin; conversion; archiving, raw diffraction images data

KD downloaded our freely available raw diffraction data images and processed those using XDS. SWMT refined all the XDS processed datasets. JRH, KD, LKB and AMMS worked together to determine the difference between the different software packages. JRH guided the analyses to determine the percentage of partial chemical conversion. SWMT wrote the manuscript with editing by JRH, LKB and KD.
Experiences with archived raw diffraction images data: capturing cisplatin after chemical conversion of carboplatin in high salt conditions for a protein crystal

Simon W. M. Tanley,* Kay Diederichs,b Loes M. J. Kroon-Batenburg,c Antoine M. M. Schreursc and John R. Helliwell**

*a School of Chemistry, Faculty of Engineering and Physical Sciences, University of Manchester, Brunswick Street, Manchester M13 9P1, UK, bDepartment of Biology, University of Konstanz, Germany, and cCrystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

E-mail: john.helliwell@manchester.ac.uk

The archiving of raw diffraction images data is the focus of an IUCr Diffraction Data Deposition Working Group (see http://forums.iucr.org/). Experience in archiving and sharing of raw diffraction images data in collaboration between Manchester and Utrecht Universities studying the binding of the important anti-cancer agents, cisplatin and carboplatin to histidine in a protein, has recently been published. Subsequently, these studies have been expanded when one of us (KD) joined in the detailed further analyses of each set of raw diffraction images with XDS. The raw diffraction images, measured at Manchester University, are available for download at Utrecht University and now also mirrored at the Tardis Raw Diffraction Data Archive in Australia. Thus a direct comparison of processed diffraction and derived protein model data from XDS with the published results has been made. The issue of conversion of carboplatin to cisplatin under a high chloride salt concentration has been taken up and a detailed crystallographic assessment is provided. Overall, these new structural chemistry research results are presented followed by a short summary of developing raw data archiving policy and practicalities as well as documenting the challenge of making appropriate and detailed recording of the metadata for crystallography.

Keywords: cisplatin; carboplatin; conversion; archiving; raw diffraction images data.

1. Introduction

The archiving of raw diffraction images data is the focus of an IUCr Diffraction Data Deposition Working Group (see http://forums.iucr.org/). Experience in archiving and sharing of diffraction data in collaboration between Manchester and Utrecht Universities studying the binding of the important anti-cancer agents, cisplatin and carboplatin to histidine in a protein has recently been described in detail (Tanley et al., 2013a). Subsequently, these studies have been expanded when one of us (KD) joined in the detailed further analyses of each set of raw diffraction images with XDS (Kabsch, 1988). The motivation for this collaboration was due to the title of the previously published article (Tanley et al., 2013a) which suggested that a rather comprehensive approach was taken to compare software packages, with the implicit intention to correlate some metrics of data quality [\(R_{\text{merge}}\), \(I/\sigma\), \(CC_{1/2}\), completeness and diffraction precision index (DPI) values, all as a function of resolution] with the amount of biological insight that these data allow one to obtain. Thus, a comprehensive investigation becomes even more useful with every addition that broadens its basis, which suggested contributing the processing with yet another widely used software package (XDS) on the existing raw diffraction data images. All of the raw diffraction images data can be found at http://rawdata.chem.uu.nl/#0001, except for 4g4c, which are at http://rawdata.chem.uu.nl/#0002, and now also mirrored at the Tardis Raw Data Archive in Australia (http://vera183.its.monash.edu.au/experiment/view/40/). The raw data sets, being a part of the supplementary materials available with the previously published article (Tanley et al., 2013a), have also led to a Supplementary Materials button being added to the Journal of Applied Crystallography contents page for our article (doi:10.1107/S0021889812044172), the first of its kind. The raw diffraction data were measured on home X-ray sources at Manchester University but our study has implications for
Synchrotron radiation facilities, where a large fraction of macromolecular X-ray data these days are measured. That said, the lead investigator for any given synchrotron radiation beam time proposal presumably would hold the data management responsibilities of their relevant funding organization and would thus finally place the data for any given publication with their preferred raw data archive in addition to the formally required Protein Data Bank (PDB) depositions of processed and derived data. It is noted that various Universities in the UK are setting up research data archives for use by its research staff, and to serve all research fields, so as to satisfy the existing, and increasing, mandates of funding organizations. Specialist needs such as at CERN are being assumed by Universities as to be met centrally, i.e. at CERN itself. Neutron facilities are also in parallel setting up or have set up raw data archives hosted by the neutron facility itself. Synchrotron radiation facilities have a variety of ‘raw data retention policies’ (see the IUCr Forum for details).

This study focuses on a direct comparison of processed diffraction and derived protein model data from XDS (Kabsch, 1988) with the results published previously (Tanley et al., 2013a). The possible issue of partial conversion of carboplatin to cisplatin under a high chloride salt concentration (Tanley et al., 2012a; Gust & Schnurr, 1999) has been taken up based on new evidence coming from the re-processing of our diffraction images as well as re-visiting the already published data and a detailed crystallographic assessment is provided.

Overall we give a mix of research results in a particular study (anti-cancer agents in their chemical behaviours and their binding to histidine in a protein) and also a short summary of developing raw data archives for diffraction data images in terms of policies and practicalities. We will also document the not insignificant challenge of making appropriate and detailed recording of the metadata for crystallography raw diffraction data.

2. Methods

Details of all the crystallization conditions and the X-ray data collection strategy for each crystal has already been previously documented (Tanley et al., 2012a,b, 2013a). PDB IDs: 4dd0, 4dd1, 4dd2, 4dd3, 4dd4, 4dd6, 4dd7, 4dd9, 4dda, 4ddb, 4dce, 3txb, 3txd, 3txe, 3txi, 3txj, 3txk, 3txf, 3txg, 3txh, 4g49, 4g4a, 4g4b, 4g4c, 4g4h.

Graphs of the anomalous signal-to-noise ratio per resolution shell for each crystal processed by the different software packages are given in the supplementary materials.1

3. Results

3.1. Anti-cancer agents in their chemical behaviours and their binding to histidine in a protein

From previous studies (Tanley et al., 2012a,b) it is shown that carboplatin bound to the Nδ and Ne atoms of His-15 of HEWL in DMSO crystallization conditions. It was also shown that cisplatin did likewise but in addition it proved possible to show that its binding also took place over prolonged chemical exposure of around a year; carboplatin-based lysozyme crystals under aqueous conditions produced crystals too fragile after a year’s storage to study successfully. Through the Utrecht archive, one of us (KD) downloaded the data sets diffraction images reanalysed here. Thus, using XDS to process these archived raw diffraction images (PDB ID 4dd7/4dd9/4g4c), clearly significant anomalous difference density peaks in the carboplatin binding site were seen; appropriately placed for where a Cl atom in cisplatin would be expected to be for the Nδ binding site and for two Cl atoms in the Ne binding site. Comparing these anomalous difference density peak heights with the peak heights from the EVAL (Schreurs et al., 2010), MOSFLM (Leslie, 1999) and PROTEUM/D*Trek (Bruker, 2006; Pflugrath, 1999) processing programs (Tanley et al., 2013a) (Table 1), it was noted that for 4dd7, these three Cl atoms were mis-interpreted first of all by us (ST and JRH), with evidence of their presence coming from all four processing programs. Whereas, for 4dd9 and 4g4c, there is no evidence for all of these Cl atoms above our 3σ cut-off level, with only one Cl in the Ne binding site being present above the 3σ cut-off level. This new finding, along with re-modelling the EVAL deposited data, leads to the conclusion that the carboplatin could in fact have been partially or possibly even fully converted to cisplatin. This chemical conversion and the respective percentages were then evaluated for each binding site (Fig. 1). The partial transformation of carboplatin to cisplatin has been seen previously reported on in solution (Gust & Schnurr, 1999). In a 0.9% NaCl solution, 10% of the carboplatin had converted to cisplatin after 28 d storage at room temperature and the concentration of NaCl used in our crystallization conditions (10%) and the storage of our crystals between days and months at room temperature (Tanley et al., 2012a,b) could indeed facilitate this partial conversion. The largest transformation took place in 4dd7, a crystal with only three days of storage. OVER-

Table 1

<table>
<thead>
<tr>
<th>Anomalous peak height</th>
<th>Nδ Cl</th>
<th>Ne Cl</th>
<th>Ne Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEWL co-crystallized with carboplatin in 10%</td>
<td>3txh PROTEUM2 MOSFLM</td>
<td>3.4</td>
<td>4.7</td>
</tr>
<tr>
<td>NaCl and DMSO media with glycerol as the cryoprotectant</td>
<td>4dd7_EVAL XDS</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>HEWL co-crystallized with carboplatin in 10%</td>
<td>3txi a*TREK MOSFLM</td>
<td>2.5</td>
<td>4.2</td>
</tr>
<tr>
<td>NaCl and DMSO media with paratone as the cryoprotectant</td>
<td>4dd9_EVAL XDS</td>
<td>3.0</td>
<td>3.8</td>
</tr>
<tr>
<td>HEWL co-crystallized with carboplatin in 10%</td>
<td>4g4c_EVAL PROTEUM2</td>
<td>2.4</td>
<td>2.5</td>
</tr>
<tr>
<td>NaCl and DMSO media studied at room temperature</td>
<td>XDS</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

1 Supplementary data for this paper are available from the IUCr electronic archives (Reference: YS5072). Services for accessing these data are described at the back of the journal.

Table 2
Correlation coefficient of cisplatin or carboplatin binding in the $2F_o - F_c$ map for both the Nδ and Ne atoms of His-15 using OVERLAPMAP in CCP4i, with the Pt centre removed, i.e the calculation is just for the remaining bound ligands.

Note that in this calculation each ligand is considered separately whereas in the SHELX calculation summarized in Table 3 both of the ligand occupancies are refined in the same calculation.

<table>
<thead>
<tr>
<th>HEWL co-crystallized with carboplatin in 10% NaCl and DMSO media with glycerol as the cryoprotectant</th>
<th>Carboxatin</th>
<th>Cisplatin</th>
<th>Cisplatin</th>
<th>Cisplatin</th>
</tr>
</thead>
<tbody>
<tr>
<td>4g4c_EVAL</td>
<td>4dd7_EVAL</td>
<td>0.09</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>4dd9_EVAL</td>
<td>XDS</td>
<td>0.12</td>
<td>0.36</td>
<td>0.54</td>
</tr>
</tbody>
</table>

LAPMAP (CCP4 package) was used to give the statistical correlation coefficient of which ligand could be bound at each binding site based on the $2F_o - F_c$ maps as well as SHELX (Sheldrick, 2008) being used to obtain the refined occupancies of the Pt atom centre at each binding site, and the occupancy values of each bound ligand remaining atoms at each binding site. This information, using both EVAL and XDS processing programs for each dataset (4dd7, 4dd9 and 4g4c), is given in Tables 2 and 3.

4. Discussion

4.1. Anti-cancer agents in their chemical behaviours and their binding to histidine in a protein

Archiving our raw diffraction images enabled us to re-process them, resulting in biological insight going beyond that of our previously published results (Tanley et al., 2013a; http://rawdata.chem.uu.nl/#0001; http://vera183.its.monash.edu.au/experiment/view/40/). It is now thought that instead of carboplatin binding alone to HEWL in DMSO media (Tanley et al., 2012a,b) we actually see a mixture of carboplatin and cisplatin in the binding sites due to the partial conversion of carboplatin to cisplatin in the high salt conditions (10% NaCl) used for the crystallization procedure. Table 1 shows the anomalous difference density peak heights in the binding sites, attributed to Cl atoms for different software programs. These confirm that, for 4dd7, anomalous difference density for these Cl atoms are seen and should have been noted previously and, also, this dataset shows the most conversion of carboplatin to cisplatin based on the correlation coefficients out of OVERLAPMAP (Table 2). However, we had set up clear chemical conditions and carboplatin only naturally was what we, wrongly as it turns out, expected. Whereas for 4dd9 and 4g4c, the anomalous difference density peaks are weaker, and thus harder to determine that the Cl atoms are in fact present. From the anomalous difference density peak heights, along with correlation coefficient of cisplatin/carboplatin binding to the $2F_o - F_c$ maps in OVERLAPMAP as well as the refined occupancy values from SHELX (Tables 1–3), it can be noted that carboplatin has been partially converted to cisplatin in the high salt conditions used for the crystallization procedure. Also, the correlation coefficient values from Table 2 for each ligand as well as the occupancy values for each ligand in Table 3 generally correlate very well between the XDS and EVAL refined models.

Based on the new finding that carboplatin has partially converted to cisplatin in the high NaCl salt content used for the crystallization conditions, a new study, involving crystallization of HEWL with carboplatin in non-NaCl conditions, is underway (Tanley et al., 2013b).
4.2. Developing raw data archiving policies and practicalities for diffraction data images and the challenges of making metadata available for crystallography

Archiving of raw diffraction data images policy and practicalities for crystallography is at its early stages (see the IUCr Forum documents); very reasonably the issues involve costs versus benefits but within a fairly rapid expansion of technical options such as University archives becoming established and cloud storage commercial options becoming more commonly available (Helliwell et al., 2012). In Tanley et al. (2013a) the very pragmatic method of using a personal web link held at Utrecht University for our various raw data sets was used. From the results shown here, without this web link archiving, we would not have re-examined the previous diffraction data (Tanley et al., 2012a,b, 2013a) and thus would have missed this most interesting chemical effect of partial conversion of carboplatin to cisplatin.

Providing sufficiently rich metadata to fully describe raw diffraction data is important for carrying out data processing of diffraction images with subsequent software. One very specific point seen in this study was that XDS required manual input of the rotation axis direction and the direction of the detector X/Y axes for Bruker CCD diffractometer data, whereas, with the EVAL software, they were readily interpreted from the header information of the diffraction images. Our previously published work (Tanley et al., 2013a) presented a detailed and rich description of metadata for two commercial area detector diffractometers and is thus, we hope, an exemplar of what detail is required.

5. Conclusions

The studies outlined here focused on the advantages of archiving the raw diffraction images for X-ray crystallography, via a relatively simple personal web link method rather than, say, a formal single University or centralized raw data archive. Due to archiving our raw diffraction images data, new findings have been achieved owing to re-processing the images with XDS as well as re-examining the previously published data. Thus, we now see carboplatin partially converting to cisplatin in the crystal structure due to the high NaCl salt concentrations used in the crystallization conditions. This has then led us to try to find new crystallization conditions without NaCl to seek to capture carboplatin binding on its own.

We are grateful for research support from the Universities of Konstanz, Manchester and Utrecht. ST is funded under an EPSRC PhD Research Studentship. EPSRC requires data archiving at a level appropriate to the relevant research community data deposition policies. These policies are in effect for our discipline set by the IUCr and who is reviewing its policies via the Diffraction Data Deposition Working Group; this is Chaired by JRH and who warmly acknowledges many discussions with Brian McMahon and Tom Terwilliger as well as LK-B, all members of the DDD WG and its consultants. The role of Professor Dr Sine Larsen in establishing the DDD WG, when she was President of the IUCr, was obviously a pivotal leadership in these important matters and who we also heartily thank.

References

Supplementary Materials

**Figure S9.1** Anomalous signal to noise ratio for each crystal processed with the different software packages.
10. Chapter 10: Carboplatin binding to a model protein in non-NaCl conditions to eliminate partial conversion to cisplatin, and the use of different criteria to choose the resolution limit

Simon W.M. Tanley\textsuperscript{a}, Kay Diederichs\textsuperscript{b}, Loes M.J. Kroon-Batenburg\textsuperscript{c}, Antoine M.M. Schreurs\textsuperscript{c} and John R. Helliwell\textsuperscript{a*}

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, UK

\textsuperscript{b} Department of Biology, University of Konstanz, Germany

\textsuperscript{c} Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Correspondence email: john.helliwell@manchester.ac.uk

Available at \url{http://arxiv.org/abs/1309.4661}

Keywords: Cisplatin; Carboplatin; histidine; MPD, conversion

JRH conceived the use of non-NaCl conditions. SWMT came up with the idea of the 48 conditions crystal screen approach. SWMT set up all the non-NaCl crystallisation conditions, mounted the crystals and collected the data on a Bruker APEX II home source diffractometer. SWMT processed the data using the Bruker internal software package, AMMS processed all the data using EVAL-15 and KD processed the data using XDS. SWMT refined all the structures presented in this paper. KD, JRH and LKB discussed the differences seen with the software and the lack of anomalous signal for the Pt atom. SWMT wrote the article, with JRH, LKB and KD editing the manuscript.
10.0 Abstract

HEWL co-crystallisation conditions of carboplatin without NaCl have been utilised to eliminate partial conversion of carboplatin to cisplatin observed previously. Tetragonal HEWL crystals were successfully obtained in 65% MPD with 0.1M citric acid buffer at pH 4.0 including DMSO. The X-ray diffraction data resolution to be used for the model refinement was reviewed using several topical criteria together. The CC1/2 criterion implemented in XDS led to data being significant to 2.0Å, compared to the data only being able to be processed to 3.0Å using the Bruker software package (SAINT). Then using paired protein model refinements and DPI values based on the FreeR value, the resolution limit was fine tuned to be 2.3Å. Interestingly this was compared with results from the EVAL software package which gave a resolution limit of 2.2Å solely using <I/sigI> crossing 2, but 2.8Å based on the Rmerge values (60%). The structural results showed that carboplatin bound to only the Nδ binding site of His-15 one week after crystal growth, whereas five weeks after crystal growth, two molecules of carboplatin are bound to the His-15 residue. In summary several new results have emerged: - firstly, non-NaCl conditions showed a carboplatin molecule bound to His-15 of HEWL; secondly, binding of one molecule of carboplatin was seen after one week of crystal growth and two molecules were bound after five weeks of crystal growth; and thirdly, the use of several criteria to determine the diffraction resolution limit led to the successful use of data to higher resolution.

10.1 Introduction

Cisplatin and carboplatin are platinum anti-cancer drugs which have been used for a long time in the fight against cancer by targeting DNA. However, 90% of their reported binding is to plasma proteins (Fischer et al 2008). Thus, these drugs cause toxic side effects. Cisplatin is rapidly converted to toxic metabolites which cause nephrotoxic effects (Zhang, 1996; Huliciak et al, 2012), whereas carboplatin is less toxic due to the addition of the cyclobutanedicarboxylate moiety, which has a slower rate of conversion. Carboplatin can therefore be tolerated by patients at higher doses compared to cisplatin (Kostova, 2006). Previous crystallographic studies of cisplatin and carboplatin with HEWL, a model protein, have shown binding of two molecules to its His-15 residue in DMSO media (Tanley et al, 2012a; 2012b; Helliwell & Tanley 2013) and for cisplatin even after prolonged exposure in
aqueous medium (Tanley et al., 2012b). Subsequently, through public archiving of our raw diffraction images at Utrecht University (Tanley et al., 2013, http://rawdata.chem.uu.nl/#0001;) and now also mirrored at the Tardis Raw data archive in Australia (http://vera183.its.monash.edu.au/experiment/view/40/), a collaboration was set up with one of the authors of this article (KD) who downloaded and re-processed the diffraction images, measured in Manchester, with the XDS software package (Kabsch, 2010) to compare with our previous results. Reviewing these results along with the previous publication of the carboplatin bound structures in DMSO media studied at cryo and room temperatures (Tanley et al., 2012a; Tanley et al., 2012b) (PDB id’s 4dd7, 4dd9 and 4g4c) it was noted that the anomalous difference electron density peaks within the carboplatin binding sites weren’t only for the Pt centres (Tanley et al., 2013b). Thus, this suggested that with the high NaCl concentrations used in our crystallisation conditions (Tanley et al., 2012a; 2012b; Helliwell & Tanley, 2013) the carboplatin could be partially converted to cisplatin, with the extra anomalous difference density thereby being attributed to the Cl atoms of cisplatin. This partial conversion of carboplatin has been seen previously in solution (Gust & Schnurr, 1999). Due to these new findings, the binding sites could thus contain a mixture of carboplatin and cisplatin rather than just the carboplatin molecule, which is what we initially thought was solely bound (Tanley et al., 2013b).

Based on these new findings, this study focuses on crystallising HEWL with carboplatin in non-NaCl conditions to remove the possibility of carboplatin converting to cisplatin. Crystals were successfully grown without NaCl and the results are presented here. The raw diffraction data images for these MPD chemical medium grown crystals were processed using the Bruker software program (SAINT) as well as EVAL (Schreurs et al., 2010) and XDS (Kabsch, 2010).

Determining the high resolution cut off limit for X-ray crystallographic data is a major discussion point (Diederichs & Karplus, 2013; Evans & Murshudov, 2013), with Rmerge values being used for many years (Arndt et al., 1968) and in recent years supplanted by monitoring where \(<I/\sigma(I)\) crosses a value of 2.0. Rmerge measures the relative spread of \(n\) independent measurements of the intensity of a reflection around their average. A relatively new criterion has been implemented, \(CC_{1/2}\), which is based on the Pearson correlation coefficient of two half datasets (Karplus & Diederichs, 2012; Evans, 2011). Our study looks to determine the high resolution cut off of our newly processed datasets using these multiple
criteria and, notably, to see if one can use data to a higher resolution than indicated from \(<\text{I/sigI}>\) crossing 2.0.

Several new results have emerged: - Firstly, non-NaCl conditions show a carboplatin molecule bound to His-15 of HEWL; secondly, binding of one molecule of carboplatin was seen after one week of crystal growth and two molecules were bound after five weeks of crystal growth; and thirdly, the use of several criteria combined allowed us to more carefully assess the diffraction resolution limit than hitherto.

10.2 Methods

10.2.1 Crystallisation conditions.

20mg of HEWL (0.6mM) was dissolved in 1ml distilled water. 1.4mg of carboplatin (1.8mM) was added in a 3-fold molar excess over that of the protein, along with 75µl DMSO and mixed until all the carboplatin had dissolved. 48 non-NaCl crystal screens from Hampton Research were set up; these comprised 2µl protein/carboplatin/DMSO solution aliquots each mixed with 2µl reservoir solution and were set up in hanging drop crystallisations with 1ml of reservoir solution. Crystallisation trays were left at room temperature and crystals grew after ~ 4 weeks in the condition: 65% MPD with 0.1M citric acid buffer at pH 4.0.

10.2.2 X-ray diffraction data collection, protein structure solution and model refinement

A 0.3mm sized crystal was mounted into a loop with silicon oil used as the cryoprotectant and X-ray diffraction data were measured one week after crystal growth (crystals took ~ 4 weeks to grow). The crystal was mounted on a Bruker APEXII home source diffractometer and data collection was carried out at 110K with a data collection strategy used to gain the most information possible. A second crystal (0.1mm sized), five weeks after initial crystal growth, was mounted into a loop, again with silicon oil used as the cryoprotectant. The crystal was again mounted on the same Bruker APEXII home source diffractometer and X-ray diffraction data collection was carried out at 127K with a data collection strategy again used to gain the most information possible. The two crystal datasets were processed using the Bruker software package (SAINT) as well as EVAL (Schreurs et al, 2010) and XDS (Kabsch, 2010). The structures were solved using MR with PHASER (McCoy et al, 2007) and restrained refinement with REFMAC5 (Vagin et al, 2004), using the reported lysozyme
structure 2w1y as molecular search model (Cianci et al 2008). Model building, adjustment and refinement were carried out respectively using the COOT (Emsley & Cowtan, 2004) molecular graphics programme and REFMAC5 (Vagin, 2004) in CCP4i. Ligand binding occupancies were calculated using SHELX (Sheldrick, 2008). Crystallographic and refinement parameters for crystals 1 and 2 processed using the SAINT, EVAL and XDS software packages are summarized in Table 10.1.

**Table 10.1**

X-ray crystallographic data and final protein model refinement statistics for crystals 1 and 2, processed by 3 different software packages.

<table>
<thead>
<tr>
<th>PDB id</th>
<th>Bruker (SAINT)</th>
<th>EVAL</th>
<th>XDS</th>
<th>Bruker (SAINT)</th>
<th>EVAL</th>
<th>XDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not deposited</td>
<td>4t0</td>
<td>4t1</td>
<td>Not deposited</td>
<td>4t2</td>
<td>4t3</td>
</tr>
<tr>
<td>Data collection temperature (K)</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>127</td>
<td>127</td>
<td>127</td>
</tr>
</tbody>
</table>

**Data reduction**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell parameters (Å)/(°)</td>
<td>a=b= 76.97</td>
<td>a=b= 76.77</td>
<td>a=b= 76.82</td>
<td>a=b= 77.12</td>
<td>a=b= 77.29</td>
<td>a=b= 77.12</td>
</tr>
<tr>
<td>c= 36.28</td>
<td>c= 36.36</td>
<td>c= 36.47</td>
<td>c= 36.54</td>
<td>c= 36.63</td>
<td>c= 36.56</td>
<td></td>
</tr>
<tr>
<td>α=β=γ= 90.0</td>
<td>α=β=γ= 90.0</td>
<td>α=β=γ= 90.0</td>
<td>α=β=γ= 90.0</td>
<td>α=β=γ= 90.0</td>
<td>α=β=γ= 90.0</td>
<td></td>
</tr>
</tbody>
</table>

| Protein molecular Mass | 14700 | 14700 | 14700 | 14700 | 14700 | 14700 |
| Molecules per asymmetric unit | 1 | 1 | 1 | 1 | 1 | 1 |
| Detector to crystal distance (mm) | 40.20 | 40.20 | 40.20 | 40.37 | 40.37 | 40.37 |
| Observed reflections | 58539 | 202135 | 231303 | 187183 | 169869 | 206299 |
| Unique reflections | 4211 | 7770 | 7765 | 6323 | 7901 | 7850 |
| Resolution (Å) (last shell) | 36.28 – 3.00 | 19.18 – 2.00 | 39.41 – 2.00 | 36.54 – 2.16 | 25.15 – 2.00 | 39.41 – 2.00 |
| (3.10 – 3.00) | (2.03 – 2.00) | (2.05 – 2.00) | (2.19-2.16) | (2.04 – 2.00) | (2.05 – 2.00) |
| Completeness (%) | 99.8 (99.1) | 99.8 (99.8) | 99.7 (99.8) | 100 (100) | 98.9 (93.7) | 99.4 (32.4) |
| Rmerge (%) | 0.255 (0.563) | 0.340 (2.71) | 0.422 (3.935) | 0.227 (0.738) | 0.193 (1.41) | 0.248 (2.27) |
| (I/σ(I)) | 11.4 (2.8) | 9.2 (0.95a) | 8.9 (0.5)§b | 14.4 (2.01) | 14.2 (1.15c) | 14.6 (0.5)§d |
| Multiplicity | 23.9 (9.8) | 26.1 (12.3) | 29.7 (9.8) | 29.6 (13.3) | 21.6 (5.2) | 26.3 (6.0) |
| CC1/2 | - | 0.38 | 0.16 | - | 0.51 | 0.32 |
$a <I/\text{sig}I> = 2.4 \text{ at } 2.2\text{Å}, <I/\text{sig}I> = 1.7 \text{ at } 2.15\text{Å}$

$b <I/\text{sig}I> = 2.0 \text{ at } 2.24\text{Å}$

$c <I/\text{sig}I> = 2.3 \text{ at } 2.07\text{Å}, <I/\text{sig}I> = 1.7 \text{ at } 2.03\text{Å}$

$d <I/\text{sig}I> = 2.17 \text{ at } 2.17\text{Å}, <I/\text{sig}I> = 1.77 \text{ at } 2.11\text{Å}$

* Final refinement statistics to 2.1Å for crystal 1 processed by EVAL, whereas the data reduction statistics are to 2.0Å; see also Table 11.2 a and b.

** Final refinement statistics to 2.3Å for crystal 1 processed by XDS, whereas the data reduction statistics are to 2.0Å; see also Table 11.2 a and b.

### 10.3 Results

#### 10.3.1 Data processing

Data processing of the raw diffraction images was carried out using SAINT (Bruker internal software), XDS (Kabsch, 1988) and EVAL (Schreurs et al, 2010). The use of these three different processing programs for crystal 1 in this case was vital, with the SAINT software processing the data to 3.0Å resolution ($<I/\text{sig}I> = 2.8$ to 3.0Å), while processing to higher resolutions failed. However, using XDS, we found that based on the CC$_{1/2}$ criterion (Karplus & Diederichs, 2012) there was information in the data up to 2.0Å resolution. Using EVAL, the data would have been cut at 2.2Å based on where $<I/\text{sig}I>$ cross 2, or 2.8Å based on the Rmerge values of this dataset exceeding 60%. However, as the CC$_{1/2}$ criterion from XDS showed data to 2.0Å resolution still carried information, the raw diffraction data images were also processed to 2.0Å resolution using EVAL. Due to the differences in the resolution limit chosen between SAINT, EVAL and XDS, separate pairwise refinements were used for the
XDS and EVAL data to determine which resolution was optimal (Table 10.2), along with the DPI values based on the FreeR value (Figure 10.1). 2.3Å was chosen via either criterion for the XDS processed raw diffraction data images and 2.1Å for the EVAL processed raw diffraction data images. The best final model and data obtained from XDS and EVAL was used for calculation of R-values against both datasets, allowing unambiguous identification of the best data and model (Table 10.3), with the EVAL processed data and model to 2.1Å resolution being marginally better than the XDS processed data and model to 2.3Å.

Table 10.2 Pairwise refinements (Karplus & Diederichs, 2012) showing R/Rfree of initial models refined at given resolution (top row), against data as given in left column. (a) XDS data and (b) EVAL data. Bold numbers indicate improvement of Rfree compared to refinement at lower resolution. The resolution which yields the consistently best model is underlined; this resolution cut-off is hereby determined as optimal for these data, the given protein model, the refinement strategy and refinement program.

<table>
<thead>
<tr>
<th>Resolution used for R-value calculation</th>
<th>3.0 Å</th>
<th>2.5 Å</th>
<th>2.4 Å</th>
<th>2.3 Å</th>
<th>2.2 Å</th>
<th>2.1 Å</th>
<th>2.0 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Å</td>
<td>17.7/25.6</td>
<td>17.6/26.2</td>
<td><strong>17.7/25.4</strong></td>
<td>17.6/25.4</td>
<td>17.6/25.5</td>
<td><strong>17.6/25.4</strong></td>
<td></td>
</tr>
<tr>
<td>2.4 Å</td>
<td>17.9/25.7</td>
<td>17.9/25.7</td>
<td><strong>17.9/25.6</strong></td>
<td>17.9/25.6</td>
<td>17.9/25.5</td>
<td>17.9/25.5</td>
<td></td>
</tr>
<tr>
<td>2.3 Å</td>
<td>18.5/24.8</td>
<td><strong>18.6/24.7</strong></td>
<td><strong>18.5/24.7</strong></td>
<td>18.8/24.7</td>
<td>18.8/24.7</td>
<td>18.8/24.7</td>
<td></td>
</tr>
<tr>
<td>2.2 Å</td>
<td>18.8/25.9</td>
<td>18.8/25.9</td>
<td>18.8/25.9</td>
<td>18.8/25.9</td>
<td>18.8/25.9</td>
<td>18.8/25.9</td>
<td></td>
</tr>
<tr>
<td>2.1 Å</td>
<td>19.7/27.2</td>
<td>19.7/27.3</td>
<td>19.7/27.3</td>
<td>19.7/27.3</td>
<td>19.7/27.3</td>
<td>19.7/27.3</td>
<td></td>
</tr>
<tr>
<td>Resolution that model was refined at</td>
<td>3.0 Å</td>
<td>2.5 Å</td>
<td>2.3 Å</td>
<td>2.1 Å</td>
<td>2.05 Å</td>
<td>2.0 Å</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>3.0 Å</td>
<td>19.2/28.7</td>
<td>19.0/27.4</td>
<td>18.9/28.0</td>
<td>18.9/27.7</td>
<td><strong>18.6/27.4</strong></td>
<td>18.9/27.9</td>
<td>18.9/27.6</td>
</tr>
<tr>
<td>2.5 Å</td>
<td>20.7/29.9</td>
<td>20.7/30.1</td>
<td>20.7/29.8</td>
<td><strong>20.5/29.4</strong></td>
<td>20.7/29.6</td>
<td>20.2/29.6</td>
<td></td>
</tr>
<tr>
<td>2.15 Å</td>
<td></td>
<td>22.1/28.3</td>
<td>22.2/28.3</td>
<td>22.2/28.3</td>
<td>22.2/28.3</td>
<td><strong>22.3/28.3</strong></td>
<td></td>
</tr>
<tr>
<td>2.1 Å</td>
<td></td>
<td></td>
<td>22.3/28.3</td>
<td>22.4/28.3</td>
<td><strong>22.3/28.3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.05 Å</td>
<td></td>
<td></td>
<td></td>
<td>22.5/28.9</td>
<td>22.5/28.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0 Å</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23.1/29.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 10.1** DPI value based on the FreeR factor for each refined model at different resolutions for both the XDS and EVAL processed raw diffraction data images. The FreeR set was kept the same for both the XDS and EVAL processed datasets. The curves are the line of best fit through the data points using a power 5/2 fit (Our data follows the same trends as in Blow, 2002) in OriginPro 8.5.1 with the line showing excellent agreement with the data points.
Table 10.3 The pairwise comparison technique applied to the XDS and EVAL data and final refined models. The best model obtained from the XDS and EVAL data, respectively, was used for calculation of R-values against both datasets, allowing unambiguous identification of the best data and model. Underlining shows the EVAL final model and data at 2.1Å resolution has lower Rfree compared to the XDS model against the EVAL data. Also, the EVAL final model has the same Rfree as the XDS final model against the XDS data.

<table>
<thead>
<tr>
<th>Resolution used for R-value calculation</th>
<th>Resolution that model was refined at</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Å EVAL</td>
<td>2.3 Å XDS</td>
</tr>
<tr>
<td>22.3/28.3</td>
<td>24.4/29.5</td>
</tr>
</tbody>
</table>

10.3.2 Carboplatin binding to His-15

For crystal 1, collected one week after crystal growth shows carboplatin bound to the Nδ binding site of His-15 at 3.0Å from the SAINT software package (Figure 10.2a), 2.3Å from XDS (Figure 10.2b) and 2.1Å from EVAL (Figure 10.2c). The models at 2.3Å and 2.1Å (Figures 10.2b and c) show more detail in the 2Fo-Fc map at the Nδ binding site, as more detail for the atoms bound to the Pt centre can be seen compared to just the Pt atom being visible in the density to 3.0Å. Binding of the carboplatin molecule to the Ne atom however, is not seen. The anomalous difference density is weak as determined from all three software packages, with the highest peak being 5.1σ at the Pt position, when the data was cut to 4.0Å resolution (Table 10.4).

For crystal 2 (Figure 10.3), collected five weeks after crystal growth, one molecule of carboplatin is bound to the Nδ and the Ne atoms of His-15. The same as for crystal 1, the anomalous difference density for the Pt atoms is weak, with the highest peak being 4.6σ for the Nδ binding site and 5σ for the Ne binding site (Table 10.4).
**Figure 10.2** Binding of carboplatin to the Nδ atom of His-15 from crystal 1 at (a) 3.0Å from the SAINT (Bruker) processed data (b) 2.3Å from the XDS processed data and (c) 2.1Å from the EVAL processed data. 2Fo-Fc maps are shown at the 1.2σ cut off level. Anomalous difference electron density (orange) maps are shown at the 3.0σ cut off level. No anomalous difference electron density was present in the Bruker dataset map (a).

**Figure 10.3** Binding of carboplatin to the Nδ and Nε atoms of His-15 for crystal 2 at (a) 2.16Å from the SAINT (Bruker) processed data (b) 2.0Å from the XDS processed data and (c) 2.0Å from the EVAL processed data. 2Fo-Fc maps are shown at the 1.2σ cut off level. Anomalous difference electron density (orange) maps are shown at the 3σ cut off level.
Table 10.4 Anomalous difference density peak heights for the Pt position at both the Nδ and Nε binding sites for crystals 1 and 2 along with the occupancy values of the Pt atoms, calculated using SHELXL (Sheldrick, 2008).

<table>
<thead>
<tr>
<th></th>
<th>Nδ</th>
<th></th>
<th>Nε</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anomalous peak height (σ)</td>
<td>Pt occupancy (%)</td>
<td>Anomalous peak height (σ)</td>
<td>Pt occupancy (%)</td>
</tr>
<tr>
<td>Crystal_1_Bruker</td>
<td>0</td>
<td>47</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Crystal_1_EVAL</td>
<td>5.1</td>
<td>39</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Crystal_1_XDS</td>
<td>3.4</td>
<td>29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Crystal_2_Bruker</td>
<td>4.1</td>
<td>42</td>
<td>4.6</td>
<td>16</td>
</tr>
<tr>
<td>Crystal_2_EVAL</td>
<td>4.6</td>
<td>22</td>
<td>5.0</td>
<td>31</td>
</tr>
<tr>
<td>Crystal_2_XDS</td>
<td>4.4</td>
<td>23</td>
<td>3.1</td>
<td>38</td>
</tr>
</tbody>
</table>

10.4 Discussion

10.4.1 Carboplatin binding to His-15 in non-NaCl conditions

Crystals grew in 65% MPD with 0.1M citric acid buffer at pH 4.0 after ~4 weeks, and one week after crystal growth one molecule of carboplatin was seen bound to the Nδ atom of His-15 (Figure 10.2). However, no binding was seen to the Nε binding site. At the low pH used (pH 4.0) in these crystallisation conditions, the His-15 residue is assumed to be a protonated histidine, which means that both N-hydrogen atoms would need to be removed before binding to the Pt centre can be achieved at these histidine N atoms. In the earlier studies we have observed binding of two molecules of cisplatin and carboplatin to His-15 in HEWL (Tanley et al., 2012a; 2012b; Helliwell & Tanley, 2013), one to each N, and was determined to be due to the histidine residue being an imidazolyl anion which is formed by the removal of the usual N-hydrogen atoms, which could be brought about by Cl ions in the co-crystallisation conditions used in those studies (Tanley et al., 2012a). However, as the MPD crystallisation conditions do not contain Cl ions, this leaves the citrate ions in the crystallisation mixture deemed to be able to extract an N-hydrogen atom. As these crystals took 4 weeks to grow and then were measured both one week and five weeks after crystal growth, the chemistry of the citrate ions removing a proton from the histidine nitrogen(s) would start immediately after the conditions were set up. Thus, the equilibrium in de-
protonation of the His N atoms and binding to cisplatin is reached slower compared to the previous crystallisation conditions as only one binding site for carboplatin is seen one week after the crystals have grown (take ~4 weeks to grow), compared to the two binding sites we previously saw after around 8 days of co-crystallisation (Tanley et al., 2012a; 2012b; Helliwell & Tanley, 2013). For the case of our second crystal, which was measured five weeks after crystal growth, binding of two molecules of carboplatin to the His-15 residue is seen (Figure 10.3). Thus, the extraction of the second N-hydrogen atom took longer due to the relatively low concentration of citrate ions in these MPD medium crystallisation conditions. The occupancy values for the Pt atoms in this study (mean occupancy of ~30% for both the Nδ and Nε binding sites, Table 10.4) are lower compared to the occupancies seen when the crystals were grown in the high NaCl conditions used previously (~70% for the Nδ binding site and ~50% for the Nε binding site) (Tanley et al., 2012a; 2012b; Helliwell & Tanley, 2013).

10.4.2 Determining the diffraction resolution of the data

As well as now seeing just carboplatin bound in the MPD conditions, this study also looked at the use of combining several criteria to determine the diffraction resolution limit of these data, and indeed which led to the successful use of data to higher resolution. The CC₁/₂ criterion in particular (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013) for determining the resolution limit for a given dataset meant that for crystal 1, data to a higher resolution was used compared to using the traditional statistics to determine the resolution limit of the data (Rmerge values or where <I/sigI> crosses 2). Once the data were processed and refined, the pairwise refinements technique (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013) was used to determine which resolution was best for these data. In this case, using CC₁/₂ (Table 10.1), the pairwise refinement technique (Table 10.2) and the DPI value (Cruickshank, 1999) of the refined model using the Rfree factor (Figure 10.1); 2.3Å resolution was determined for the XDS processed diffraction data images and 2.1Å resolution for the EVAL processed diffraction data images for crystal one. These are compared to 3.0Å from SAINT, with attempts to process the data to higher resolution using SAINT failing to produce any results. In Table 10.2, the highlighting in bold represents the improvement in R/Rfree values against the resolution used to refine the model and for the XDS case the
refined model at 2.3Å resolution has the lowest R/Rfree gap (6.3) of all the models, meaning that for the XDS processed data, 2.3Å resolution is most probably the optimum high resolution cut off to use for this dataset, but is not clear cut. However, for the EVAL processed data, it is clear that 2.1Å resolution is the optimum resolution to use for this dataset. The graphs of the DPI values based on the Rfree value (Figure 10.2) agrees with these results from the pair-wise refinement technique as at these resolutions, one can see a flattening off of the curve agreeing with the results in Blow, 2002. Plotting the natural logs of the DPI values and the resolution (Supplementary Figure 10.1), EVAL shows a power law of 2.4 and XDS 2.3, i.e. very close to the theory value of 2.5 (Blow, 2002) which shows that the DPI is proportional to the power 5/2. Interestingly, the XDS processed dataset to 2.3Å resolution has a very similar DPI value to the EVAL processed dataset to 2.1Å resolution. If the CC1/2 criterion had not been available, this dataset would have been processed by EVAL to 2.8Å based on the Rmerge values being higher than 60% and using <I/sigI> crossing 2 this was 2.2Å but had an Rmerge value of 140%. Thus, processing these data with multiple software packages with these different criteria to determine the overall resolution limit led to use of higher resolution.

The best model obtained from the XDS and EVAL data, respectively, was used for calculation of R-values against both datasets, allowing unambiguous identification of the best data and model (Table 10.3). From these pair-wise refinements, it is determined that the EVAL processed data to 2.1Å and the final refined model is marginally better than the XDS processed data to 2.3Å and its final refined model. The reason why the R-values (Rwork/Rfree) of the best model against the XDS data are about 2% lower than the best model against the EVAL data is not known to us, and this question is currently under investigation.

The higher high-resolution limit of the EVAL data compared with the XDS data is consistent with the CC1/2 values reported by EVAL at 2.1Å resolution (80%) and XDS at 2.3Å resolution (75%) respectively.

As a further comparison for crystal 2, this dataset was processed to 2.16Å by SAINT, which was where the <I/sigI> value crossed 2, but with CC1/2 implemented in both XDS and EVAL, this dataset was processable to 2.0Å; the edge of the detector, which therefore limited the high resolution for crystal 2.
Comparing the XDS and EVAL unscaled and scaled data for crystal 1 leads to a number of conclusions about this particular dataset; (i) the statistics of the EVAL data including the CC<sub>1/2</sub> criterion are slightly better than XDS. (ii) The difference between XDS and EVAL is that large intensity reflections are larger in EVAL than XDS (Supplementary Figure 10.2a). (iii) The standard deviations of individual reflections in EVAL are larger than in XDS (Supplementary Figure 10.2b). The difference is largely caused by SADABS and XSCELE applying different error models. SADABS relies heavily on internal standard deviations. (iv) The refinement results are better (~2% lower Rfactors) for XDS. (v) The correlation between the F’s from XDS and EVAL is very good, except for the lowest resolution, where some reflections are partially screened by the beam stop, the definition of the beam stop being slightly different in EVAL from XDS.

10.5 Conclusions

The study outlined here shows carboplatin bound to the Nδ and Nε atoms of His-15 from HEWL in non-NaCl conditions, thus removing the possibility of carboplatin partially converting to cisplatin in high salt conditions seen previously (Tanley et al 2013b).

This study also looked at the use of several criteria to determine the resolution range of the datasets. Using CC<sub>1/2</sub>, the pairwise refinement technique and the DPI value based on FreeR, 2.1Å was confirmed as the best resolution limit to use for the EVAL processed raw diffraction images.

10.6 Acknowledgements

We are grateful for research support from the Universities of Konstanz, Manchester and Utrecht. ST is funded under an EPSRC PhD Research Studentship. We are grateful to Dr Ed Pozharski, University of Maryland for valuable discussions.
10.7 Supplementary Materials

**Figure S10.1** Natural log graph of the DPI value based on Rfree against the resolution limit of the refined model for both the XDS and EVAL processed datasets. One would expect a power law of 2.5 based on equation 9 of Blow, 2002. The EVAL data has a power law of 2.4, and XDS 2.3, but both datasets follow the same trend (Figure 10.1), similar to the results in Blow, 2002.
Figure S10.2 Comparison between XDS and EVAL for the data of crystal 1. (a) Log$_{10}$ (intensity) of reflections, which are larger in EVAL (horizontal) than XDS (vertical) and (b) I/σ comparison of EVAL vs. XDS data. The error model in SADABS, used for the EVAL data, is largely responsible for the observed differences in I/σ.

10.8 References


Chapter 11: Carboplatin binding to Histidine

Simon W.M. Tanley\textsuperscript{a}, Kay Diederichs\textsuperscript{b}, Loes M.J. Kroon-Batenburg\textsuperscript{c}, Colin Levy\textsuperscript{d}, Antoine M.M. Schreurs\textsuperscript{c} and John R. Helliwell\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a}School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, England

\textsuperscript{b}Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

\textsuperscript{c}Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

\textsuperscript{d}Manchester Institute of Biotechnology (MIB), University of Manchester, 131 Princess Street, Manchester, M1 7DN, England

Correspondence email: John.helliwell@manchester.ac.uk

Keywords: Carboplatin; Histidine; Avoid partial conversion to cisplatin; Non-NaCl conditions

To be submitted to Acta Cryst F

SWMT set up and grew the non-NaCl crystallisations, mounted the crystals and measured them on a home source Bruker APEX II diffractometer. SWMT processed 4 of the datasets with the Bruker internal software package. AMMS and LKB processed one of the datasets with EVAL and KD processed one of the datasets with XDS. JRH came up with the idea of NaBr crystallisation conditions to confirm whether the partial conversion of carboplatin to the bromo-platinated form occurred in a similar mechanism to the conversion of carboplatin to cisplatin in the Cl crystallisation conditions. SWMT grew the NaBr crystals and measured one on a home source Bruker APEX II diffractometer. CL collected a dataset on beamline I04 at Diamond and SWMT processed the data using Mosflm. SWMT refined all of the structures presented here. KD came up with the idea of using elemental analysis to check the Cl concentration in the lyophilised HEWL powder. SWMT wrote the publication with JRH with comments from all co-authors.
11.0 Abstract

Carboplatin is a second generation platinum anti-cancer agent used for a variety of cancers. Previous X-ray crystallographic studies of carboplatin binding to histidine in HEWL saw partial conversion of carboplatin to cisplatin due to the high NaCl concentration used in the crystallisation conditions. HEWL co-crystallizations with carboplatin in NaBr conditions have now been carried out to confirm if carboplatin converts to the bromO form, and whether this took place in a similar way to the partial conversion of carboplatin to cisplatin in the NaCl conditions seen previously. We report here, firstly, that a partial chemical transformation took place but to a transplatin form. Thus, in moving forward to try to resolve a purely carboplatin binding at histidine, this study has utilised co-crystallisation of HEWL with carboplatin without NaCl to eliminate this partial chemical conversion of carboplatin. Tetragonal HEWL crystals co-crystallized with carboplatin were successfully obtained in four different conditions, each at different pHs. The structural results obtained show carboplatin bound to either one or both of the nitrogen atoms of His-15 of HEWL and this particular variation was dependent on the concentration of anions in the crystallization mix, and the elapsed time, as well as the pH used. The structural detail of the bound carboplatin molecule also differed between them. Overall the most detailed crystal structure showed the majority of the bound carboplatin atoms to the platinum center; the 4 carbon ring structure of the cyclobutanedicarboxylate moiety (CBDC) remained elusive however.

11.1 Introduction

Cisplatin and carboplatin are platinum anti-cancer drugs which have been used for a long time in the fight against cancer by targeting DNA. However, 90% of their reported binding cases are to plasma proteins (Fischer et al 2008). Thus, these drugs cause toxic side effects. Cisplatin is rapidly converted to toxic metabolites which cause nephrotoxic effects (Zhang, 1996; Huliciak et al, 2012), whereas carboplatin is less toxic due to the addition of the CBDC moiety, which has a slower rate of conversion (Figure 11.1). Carboplatin can therefore be tolerated by patients at higher doses compared to cisplatin (Kostova, 2006).

“The crystal structure of carboplatin on its own has been determined (Beagley et al 1985). It showed a mirror plane containing the platinum and the average cyclobutane ring and
bisecting the molecule whereby one of the cyclobutane carbon atoms showed excessive thermal motion interpreted in terms of a dynamically puckering ring”. We have built on that work and on the study of Casini et al 2007, which reported a crystal structure of cisplatin with HEWL, showing one cisplatin bound to His15. Casini et al 2007 also reported mass spectroscopy data of both cisplatin and carboplatin binding to HEWL. Cisplatin binding to Histidine residues of Superoxide dismutase (Calderon et al, 2006; Casini et al,2008) and Cytochrome C (Casini et al, 2006) have also been determined. Our X-ray crystallographic studies of cisplatin with HEWL, a model protein, have shown binding of two molecules to its His-15 residue in DMSO media (Tanley et al, 2012a; 2012b; Helliwell & Tanley 2013) and even after prolonged exposure in aqueous medium (Tanley et al, 2012b). Subsequently, through public archiving of our raw diffraction images at Utrecht University (Tanley et al, 2013a, http://rawdata.chem.uu.nl/#0001; http://rawdata.chem.uu.nl/#0002) and now also mirrored at the Tardis Raw data archive in Australia (http://vera183.its.monash.edu.au/experiment/view/40/), a collaboration was set up with one of the authors of this article (KD) who downloaded and re-processed the diffraction images, measured at the University of Manchester, with the XDS software package (Kabsch, 2010) to compare with our previous results. Reviewing those results along with our previous publication of the carboplatin bound structures in DMSO media studied at cryo and room temperatures (Tanley et al, 2012a; Tanley et al, 2012b) (PDB id’s 4dd7, 4dd9 and 4g4c) it was noted that there were two small, extra, anomalous difference electron density peaks within the carboplatin binding sites (Tanley et al, 2013b). Thus, this suggested that with the high NaCl concentrations used in our crystallisation conditions (Tanley et al, 2012a; 2012b; Helliwell & Tanley, 2013); carboplatin could be partially converted to cisplatin, with the extra anomalous difference density thereby being attributed to the Cl atoms of cisplatin. This partial conversion of carboplatin had been seen previously in solution (Gust & Schnurr, 1999), and which confirmed the idea as a possibility. Due to these new findings, the His15 binding sites could thus contain a mixture of carboplatin and cisplatin rather than just the pure carboplatin molecule.

Figure 11.1 Chemical diagrams of cisplatin and carboplatin.
Based on those findings (Tanley et al, 2013b), arising from sharing the raw diffraction data images (Tanley et al, 2013a), we can now report co-crystallisation of HEWL and carboplatin in NaBr conditions; this chemical condition was used to try to make the expected two bromines more readily visible than two partially occupied chlorines. We also have investigated crystallising HEWL with carboplatin in non-salt i.e. neither NaCl nor NaBr conditions to completely remove the possibility of carboplatin converting to the chloro or bromo forms. Again we were able to build on a very useful previous study (Weiss et al, 2000) which crystallised HEWL in 75% MPD at pH 8.0. In addition we made an extensive search whereby we surveyed 48 different non-NaCl crystallisation conditions from Hampton research (Supplementary Table 11.1 (S11.1)) to try and find a variety of conditions and pHs to see which gave the best detailed binding site of the carboplatin molecule in the absence of these salt ions in the crystallisation mix. We have also carried out an elemental analysis of the Sigma supplied HEWL to scrutinise any chloride content in particular.

11.2 Methods

11.2.1 Crystallisation conditions

11.2.1.2 NaBr

Co-crystallisation of HEWL with carboplatin in NaBr solution was carried out under similar conditions as published in Dauter & Dauter 1999 and Lim et al 1998 but with 20mg/ml HEWL co-crystallised with 1.4mg carboplatin with 75µl DMSO, 462.5µl 0.1M NaAc and 462.5µl 1M NaBr solution.

11.2.1.2 Conditions without NaCl or NaBr

20mg of HEWL (0.6mM) was dissolved in 1ml distilled water. 1.4mg of carboplatin (1.8mM) was added in a 3-fold molar excess over that of the protein, along with 75µl DMSO and mixed until all the carboplatin had dissolved. 48 crystal screens from Hampton Research (listed in Table S1) were set up; these comprised 2µl protein/carboplatin/DMSO solution aliquots each mixed with 2µl reservoir solution and were set up in hanging drop crystallisations with 1ml of reservoir solution. The crystallisation trays were left at room temperature and the crystals that grew and yielded detailed structural results, described below, were in the conditions: 65% MPD with 0.1M citric acid buffer at pH 4.0; 0.2M
NH₄SO₄, 0.1M NaAc in 25% PEG 4000 at pH 4.6; 0.1M NaCitrate, 20% propanol, 20% PEG 4000 at pH 5.6 and 2M NH₄formate, 0.1M HEPES at pH 7.5.

11.2.1.3 Other non-NaCl or NaBr conditions

Besides those described in section 11.2.1.2 above, crystals also grew in the following conditions: (i) 0.1M Imidazole, 1M NaAc at pH 4.6 (ii) 20% Jeffamine 500, 0.1M HEPES at pH 7.5 and (iii) 0.1M NaHEPES, 0.8M NaK Tartrate. These showed that carboplatin had not bound.

11.2.1.4 Elemental analysis of the HEWL lyophilised powder from Sigma

Elemental analysis scrutinising the chloride content of the lyophilised HEWL powder purchased from Sigma shows that there was 2.6% chlorine present. For crystallisation, 20mg of HEWL was dissolved in 1ml of water and from this, 2μl aliquots were used to set up the hanging drop crystallisations. Thus, from this starting 2.6% Cl, the percentage of Cl in each of our crystallisation droplet conditions was around 0.005%. This is therefore much lower than the 10% (1.4M) NaCl solution used in our previous crystallisation conditions (Tanley et al, 2012a; 2012b; Helliwell & Tanley 2013). Thus, this 0.005% level of Cl ions suggests that there would be no significant conversion to cisplatin.

11.2.2 X-ray diffraction data collection, protein structure solution and model refinement

Crystals were each scooped into a loop with silicon oil used as the cryoprotectant. All XRD data were measured on a Bruker APEXII home source diffractometer at an X-ray wavelength of 1.5418Å, except for one of the NaBr grown crystals whose XRD data were collected on Beamline I04 at Diamond, with an X-ray wavelength of 0.9163Å used. The XRD data collections were carried out at fixed temperatures between 100-127K (Table 1). For the home Lab runs each crystal’s XRD data were processed using the Bruker software package (SAINT) (Bruker 2006) except for the case of the 65% MPD with 0.1M citric acid buffer at pH 4.0 crystallisation mix where crystal one was processed with EVAL (Schreurs et al, 2010) and crystal two by XDS (Kabsch, 2010). For the NaBr grown crystal; XRD data collected at Diamond Light Source Beamline I04, was processed with Mosflm (Leslie, 1999).

The crystal structures were solved using molecular replacement with PHASER (McCoy et al, 2007) using the reported lysozyme structure 2w1y as molecular search model (Cianci et al 2008) and restrained refinement with CCP4i (Winn et al, 2011) REFMAC5 (Vagin et al,
For the NaBr grown crystal measured on Beamline I04, anisotropic B factors were refined, as the high resolution afforded that possibility, whereas all other datasets were refined with isotropic B factors. Model building, adjustment and refinement were carried out respectively using the COOT (Emsley & Cowtan, 2004) molecular graphics programme and REFMAC5 (Vagin, 2004) in CCP4i. Ligand binding occupancies were calculated using SHELXL (Sheldrick, 2008). The crystallographic and molecular model refinement parameters are summarized in Table 11.1. The esd values for the Pt occupancies (Table 11.2) were calculated using the full matrix inversion technique in SHELXL (Sheldrick, 2008). This method does not work when the resolution is worse than 2.5Å.

Table 11.1 X-ray crystallographic data and final protein model refinement statistics for all crystals studied.

<table>
<thead>
<tr>
<th>Crystallisation conditions</th>
<th>0.1M NaAc, 1M NaBr</th>
<th>0.1M NaAc, 1M NaBr</th>
<th>@65% MPD, 0.1M citric acid buffer pH4.0 crystal 1</th>
<th>@65% MPD, 0.1M citric acid buffer pH4.0 crystal 2</th>
<th>0.2M NH₄NO₃</th>
<th>0.1M Nacitrate 20% propanoal</th>
<th>2M NH₄ formate 0.1M HEPES pH 7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffractometer</td>
<td>Diamond I04</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
<td>Bruker APEX II</td>
</tr>
<tr>
<td>Processing program</td>
<td>Mosflm</td>
<td>SAINT</td>
<td>EVAL</td>
<td>XDS</td>
<td>SAINT</td>
<td>SAINT</td>
<td>SAINT</td>
</tr>
<tr>
<td>Time between crystallisation set up and XRD data collection (weeks)</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>5</td>
<td>16</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>PDB id</td>
<td>4nsf</td>
<td>4nsg</td>
<td>4lt0</td>
<td>4lt3</td>
<td>4nsh</td>
<td>4nsi</td>
<td>4nsj</td>
</tr>
<tr>
<td>Data collection temperature (K)</td>
<td>100</td>
<td>100</td>
<td>110</td>
<td>127</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

**Data reduction**

<table>
<thead>
<tr>
<th>Space group</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
<th>P4₁2₁2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell parameters (Å)</td>
<td>a=b= 78.58</td>
<td>a=b= 78.37</td>
<td>a=b= 76.77</td>
<td>a=b= 77.12</td>
<td>a=b= 77.08</td>
<td>a=b= 77.71</td>
<td>a=b= 77.49</td>
</tr>
<tr>
<td>c= 37.29</td>
<td>c= 37.21</td>
<td>c= 36.36</td>
<td>c= 36.56</td>
<td>c= 37.14</td>
<td>c= 36.84</td>
<td>c= 37.12</td>
<td>c= 37.12</td>
</tr>
<tr>
<td>Detector to crystal distance</td>
<td>205.7</td>
<td>40</td>
<td>40.2</td>
<td>40.4</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>(mm)</td>
<td>197352</td>
<td>150324</td>
<td>202135</td>
<td>206299</td>
<td>145058</td>
<td>294049</td>
<td>831022</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>18627</td>
<td>8110</td>
<td>7770</td>
<td>7850</td>
<td>6965</td>
<td>9485</td>
<td>13055</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18.14 – 1.47 (1.50 – 1.47)</td>
<td>30.89 – 2.00 (2.10 – 2.00)</td>
<td>19.18 – 2.00 (2.03 – 2.00)</td>
<td>39.41 – 2.00 (2.05 – 2.00)</td>
<td>30.69 – 2.10 (2.13 – 2.10)</td>
<td>33.29 – 2.30 (2.65 – 2.30)</td>
<td>33.48 – 1.70 (1.80 – 1.70)</td>
</tr>
<tr>
<td>Resolutions (Å)</td>
<td>91.7 (64.2)</td>
<td>97.8 (85.9)</td>
<td>99.8 (99.8)</td>
<td>99.4 (32.4)</td>
<td>99.8 (99.7)</td>
<td>94.8 (100)</td>
<td>99.7 (98.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>0.090 (0.460)</td>
<td>0.224 (0.619)</td>
<td>0.340 (2.71)</td>
<td>0.248 (2.27)</td>
<td>0.126 (0.499)</td>
<td>0.295 (0.505)</td>
<td>0.212 (0.942)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>16.3 (4.2)</td>
<td>10.8 (1.3)</td>
<td>9.2 (0.9)</td>
<td>14.6 (0.5)</td>
<td>19.2 (2.2)</td>
<td>7.5 (1.5)</td>
<td>19.3 (1.4)</td>
</tr>
<tr>
<td>(I/σ (I))</td>
<td>10.6 (8.9)</td>
<td>18.1 (3.6)</td>
<td>26.1 (12.3)</td>
<td>26.3 (6.0)</td>
<td>20.8 (5.8)</td>
<td>7.8 (7.2)</td>
<td>63.5 (30.1)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>17.2</td>
<td>13.7</td>
<td>26.8*</td>
<td>27.6</td>
<td>21.7</td>
<td>29.5</td>
<td>19.2</td>
</tr>
<tr>
<td>CC1/2</td>
<td>-</td>
<td>-</td>
<td>0.38</td>
<td>0.32</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Cruickshank DPI (Å)</td>
<td>0.07</td>
<td>0.27</td>
<td>0.25*</td>
<td>0.19</td>
<td>0.30</td>
<td>0.60</td>
<td>0.13</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>0.02/2.1</td>
<td>0.03/0.7</td>
<td>0.02/1.9*</td>
<td>0.01/1.6</td>
<td>0.01/1.2</td>
<td>0.01/1.4</td>
<td>0.02/2.2</td>
</tr>
</tbody>
</table>

### Refinement

| R factor/ R free | 12.5/18.1 | 21.8/27.0 | 22.3/28.3* | 19.5/25.7 | 19.6/26.1 | 22.0/28.4 | 19.9/25.7 |
| RMSD bonds (Å)/ Angles (°) | 0.02/2.1 | 0.03/0.7 | 0.02/1.9* | 0.01/1.6 | 0.01/1.2 | 0.01/1.4 | 0.02/2.2 |

### Ramachandran values (%)

| Most favoured | 97.6 | 96.1 | 96.1 | 97.6 | 92.1 | 92.1 | 96.1 |
| Additional allowed | 2.4 | 3.9 | 3.9 | 2.4 | 7.9 | 6.3 | 3.9 |
| Disallowed | 0 | 0 | 0 | 0 | 0 | 1.6£ | 0 |

@ These two data sets are described in an arXiv preprint (Chapter 10) (Tanley et al 2013c); this preprint represented the start of the search for non-NaCl crystallisation conditions for carboplatin with HEWL. Tanley et al 2013c also logged ideas for combinations of criteria assessing diffraction data resolution limits and which study is being pursued separately.

$\text{a } <\text{I}/\text{sigI}> = 2.4 \text{ at 2.2Å, } <\text{I}/\text{sigI}> = 1.7 \text{ at 2.15Å}$

$\text{b } <\text{I}/\text{sigI}> = 2.17 \text{ at 2.17Å, } <\text{I}/\text{sigI}> = 1.77 \text{ at 2.11Å}$

* Final refinement statistics to 2.1Å for 65% MPD, 0.1M citric acid buffer pH4.0 crystal 1 processed by EVAL, whereas the data reduction statistics are to 2.0Å.
The two residues in the disallowed regions of the Ramachandran plot are Gly residues (Gly 16 and 102).

11.3 Results

11.3.1 Carboblatin crystallisation with HEWL in NaBr crystallisation conditions

11.3.1.1 The formation of bromo transplatin in the Nδ binding site

This experiment sought to make any conversion to the Bromo platinated form more clearly discernible than the chloro version described in Tanley et al. 2013b. The most surprising result is that the trans bromo form occurred, rather than a cis bromo form (Figure 11.2). Two crystals were studied, each quite a long while after crystallisation was set up, namely after 10 and 11 weeks.

For the crystal from which the CuKα dataset was measured, platinum binding is seen at both the Nδ and Nε binding sites (Figures 11.2a) with anomalous difference electron density peak heights for the Pt positions of 12.1σ and 3.2σ observed in the Nδ and Nε binding sites respectively, which agree closely with Pt occupancy values of 94% (+/- 6) and 28% (+/- 7) calculated from SHELXL respectively (Table 11.2). An occupancy value of 94% is the highest we have observed for a Pt atom in any cisplatin, carboplatin or, now, bromo-transplatin form. The Nε binding site is harder to interpret due to there being two (and not three) anomalous difference density peaks, and both are weak and both are of similar heights. The interpretation (shown in Figures 11.2a and 11.3a) is guided by the expected distances to the His15 Nε of the Pt and of the assigned-to-be-bromine distance to that Pt. A second crystal was used to collect the XRD dataset for the NaBr condition and was collected on beamline I04 at Diamond Light Source using an X-ray wavelength of 0.9163Å, thereby on the short wavelength side of and very close to the Br K edge, thereby has an optimised f" signal (3.9 versus 1.4 electrons at CuKα). Figure 11.2b shows the difference electron density maps. Similarly to the CuKα dataset results described above, a Pt atom is seen bound to both the Nδ and Nε atoms of His-15, again with a strongly occupied Pt in the Nδ binding site (76% +/- 12) and a weak occupancy in the Nε binding site (13% +/- 11). In the Nδ binding site, besides the Pt peak, there are two large anomalous difference density peaks of 13σ and 15σ which are readily assignable to be Br atoms and at distances of 2.5Å (+/-0.1) away from the Pt atom, definitely confirming the carboplatin has become the trans bromo platinated form. These Br
atoms are also confirmed by the presence of strong 2Fo-Fc electron density peaks. In the Nε binding site, where there is weak binding, the interpretation is still difficult but easier due to the strong anomalous difference map peaks compared with the CuKα case. The Pt atom could be assigned based on the closest distance to the His 15 Nε and the Br peak assignment then followed (Figure 11.3b).

A recent study of an iodo form of cisplatin bound to His-15 of HEWL (Messori et al, 2013) showed ‘peculiar features’ involving “the presence of three peaks of anomalous electron density close to a Pt atom suggesting the presence of two alternative modes of binding of [PtI2NH3] moiety”. Thus, scrutinising our 2Fo-Fc density at the extreme left of Nδ Pt centre in each of the CuKα and the I04 Diamond datasets, it has a very similar shape, and there is no anomalous difference electron density. This rules out the possibility of a Br atom substituting also at this third position. This electron density is also too detailed to be a single N atom. Thus, a portion of the CBDC moiety of the carboplatin molecule (Figure 11.1) must still be present at this binding position, suggesting that the carboplatin has ‘only’ partially converted to the trans bromo platinated form.

**Figure 11.2** Carboplatin binding to His-15 with Br atoms as labelled assigned bound to Pt due to the presence of anomalous difference electron density. (a) CuKα dataset and (b) I04 dataset from diamond. 2Fo-Fc map (purple) is shown at the 1.5rms contour level and the anomalous difference electron density map (orange) is shown at the 3σ contour level.
Figure 11.3 The distances between atoms in the Nε binding site of the crystals grown in NaBr conditions. (a) CuKα dataset and (b) I04 dataset from Diamond. (a) The Pt to Nε bond distance is 2.6Å +/- 0.56, the Pt to Br distance is 2.8Å +/- 0.53 and the Br to NH bond distance is 3.6Å +/- 0.35. (b) The Pt to Nε bond distance is 2.6Å +/- 0.11, the Pt to Br distance is 2.5Å +/- 0.11 and the Br to NH bond distance is 3.5Å +/- 0.09. The sigmas given on the bond distances are calculated using the Cruickshank DPI values (Cruickshank, 1999). The distance between the Br atom and the NH group of Ile-88 is a normal halide H-bond distance.

11.3.1.2 Other bromine binding sites?

Based on the anomalous difference electron density map, it is observed that the Br atoms are indeed bound to the protein and are in the same positions as observed by Dauter & Dauter (1999) and Lim et al (1998), and with these sites also being the same as the usual Cl binding sites.

11.3.2 Carboplatin binding to His-15 in non-NaCl conditions

For the crystals grown in 65% MPD, 0.1M citric acid buffer at pH4.0, XRD data were collected from two separate crystals 4 weeks apart (Figure 11.4a and b). Data from crystal 1 was collected one week after crystal growth and shows carboplatin bound to the Nδ binding site of His-15 (Figure 11.4a). However, crystal 2, from the same crystallisation drop was collected 5 weeks after crystal growth was set up and now a molecule of carboplatin is seen
bound to both the Nδ and Nε atoms of His-15 (Figure 11.4b). These two crystals also show differences in the amount of detail observed for the carboplatin molecule at the Nδ binding site. Most importantly, the amount of detail in the Nε binding site of crystal 2 (Figure 11.4b) is the most we have seen at this particular binding site, out of all the crystals studied, where we observe a portion of the CBDC moiety.

The crystal grown in 0.2M NH₄SO₄, 0.1M NaAc, 25% PEG 4000 at pH 4.6 shows a molecule of carboplatin bound to both the Nδ and Nε atoms of His-15 (Figure 11.4c). No extra detail is seen in either binding site compared to the crystals grown at pH 4.0 in 65% MPD described above.

The crystal grown in 0.1M Na citrate, 20% propanol, and 20% PEG4000 at pH 5.6 again shows a molecule of carboplatin bound to both the Nδ and Nε atoms of His-15 (Figure 11.4d). In the Nε binding site, the electron density allows only a Pt atom to be modelled in. However, at the Nδ binding site in this pH 5.6 crystal, one sees the most detail for the carboplatin molecule, with only the four carbon ring of the CBDC moiety missing in the electron density.

The last condition in which crystals were obtained was with 2M NH₄ formate and 0.1M HEPES pH at 7.5. These crystals only show carboplatin bound to the Nδ binding site (Figure 11.4e), and with electron density visible solely for the Pt atom and one N atom. Whereas, in the Nε binding site, 2Fo-Fc electron density is seen at 2.5 rms, but no anomalous difference density is present. However, this electron density is around 3.5Å away from the Nε atom, and thus we cannot interpret it as a Pt atom even if it might be weakly bound i.e. due to this unusually large distance (versus 2.4 Å when properly bound).
Figure 11.4 Binding of carboplatin to the Nδ and/or Nε atom of His-15. (a) 65% MPD, 0.1M citric acid buffer at pH4.0 crystal 1, (b) 65% MPD, 0.1M citric acid buffer at pH4.0 crystal 2, (c) 0.2M NH₄SO₄, 0.1M NaAc, 25% PEG 4000 at pH4.6, (d) 0.1M Na citrate, 20% propanaol, 20% PEG4000 at pH5.6, (e) 2M NH₄ formate, 0.1M HEPES at pH7.5. 2Fo-Fc maps (purple) are shown at the 1.2rms contour level. Anomalous difference electron density (orange) maps are shown at the 3.0σ contour level.

The occupancy values of the Pt atoms at each binding site along with the anomalous difference density peak heights are given in Table 11.2 for each of the ‘non NaCl’ crystallisation conditions. With the exception of the crystal at pH 5.6, all of the occupancy values are lower compared to the average Pt occupancies seen previously in the NaCl crystallisation conditions (~70% for the Nδ binding site and ~50% for the Nε binding site with an estimated +/- 5% standard uncertainty) (Tanley et al, 2012a).
Table 11.2 Anomalous difference electron density peak heights for the Pt position at both the Nδ and Nε binding sites (σ) along with the occupancy values of the Pt atoms (%) calculated using SHELX (Sheldrick, 2008).

<table>
<thead>
<tr>
<th></th>
<th>Nδ</th>
<th>Nε</th>
<th>Summed occupancies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anomalous peak height (σ)</td>
<td>Pt occupancy (%)</td>
<td>Anomalous peak height (σ)</td>
</tr>
<tr>
<td>pH 4.0 Crystal 1</td>
<td>5.1 39 +/- 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pH 4.0 Crystal 2</td>
<td>4.6 22 +/- 7</td>
<td>5.0 31 +/- 4</td>
<td></td>
</tr>
<tr>
<td>pH 4.6</td>
<td>9.5 49 +/- 8</td>
<td>5.2 40 +/- 8</td>
<td></td>
</tr>
<tr>
<td>pH 5.6 Crystal 1</td>
<td>14.2 65 +/- 3</td>
<td>11.6 46 +/- 4</td>
<td></td>
</tr>
<tr>
<td>pH 7.5</td>
<td>5.1 29 +/- 3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NaBr, I04 diamond</td>
<td>42.9 75 +/- 12</td>
<td>6.4 13 +/- 11</td>
<td></td>
</tr>
<tr>
<td>NaBr, CuKα</td>
<td>12.1 94 +/- 6</td>
<td>3.1 28 +/- 7</td>
<td></td>
</tr>
</tbody>
</table>

The pH 7.5 crystal showed two unusual Fo-Fc density and anomalous difference density peaks very near to one of the disulphide bonds (Cys 6-127) (see Supplementary Figure 11.1).

11.4 Discussion

11.4.1 Carboplatin binding to His15 in NaBr co-crystallisation conditions

Using NaBr in the crystallisation conditions as with NaCl indeed confirmed the partial chemical conversion of carboplatin. Here though we see a trans bromo platinated form, rather than a chloro cisplatin. The bromo transplatin, in the NaBr grown crystal studied at the CuKα X-ray wavelength, has the highest occupancy we have seen (94% +/- 6 at the His15 Nδ position).
11.4.2 Carboplatin binding to His15 in non-halide co-crystallisation conditions

11.4.2.1 The different forms of the His 15 cause differing binding modes

From the four crystallisation conditions in which carboplatin is seen bound to the His-15 residue, the percentage of the binding of carboplatin molecules to the Nδ and Nε atoms of His-15 varies considerably. This chemical behaviour can be compared to the NaCl crystallisation conditions, and which always showed one cisplatin bound at each of the Nδ and Nε atoms of His-15 (Tanley et al, 2012a; 2012b, Helliwell & Tanley 2013).

As we have explained previously (Tanley et al 2012a), which we briefly reiterate here, the case of binding to both Nδ and Nε atoms of His-15 could be due to the His residue being an imidazolyl anion, or due to the Histidine being able to exist in two tautomeric forms in solution at a pH similar to the pKₐ (6.0-6.3). Thus, in these two tautomeric forms, either the Nδ or the Nε atom can participate in the interaction with Pt. The imidazolyl anion is formed by the removal of both of the N-hydrogen atoms, and can be brought about by a high concentration of Cl⁻ ions (1.4M) in the co-crystallisation conditions used in those studies (Tanley et al, 2012a). The imidazolyl anion has a lone pair of electrons on both N atoms each capable of bonding with metal atoms. Thus the His residue is known to be able to exist in three different forms; protonated His at pH <6.0, de-protonated His at physiological pHs between 6.5-7.5 and as the imidazolyl anion formed by the extraction of the two N-hydrogen atoms (Figure 11.5). In our previous studies (Tanley et al, 2012a; 2012b; Helliwell & Tanley 2013) the crystallisation was carried out at pH 4.6, thus the His should be in its protonated form, but the cisplatin was in fact seen bound to both the Nδ and Nε atoms of His-15, meaning that both the N atoms must have had their N-Hydrogen atoms extracted. This can be due to the high concentration of Cl⁻ ions in the crystallisation mix. In these studies, the overall summed occupancies of the two Pt atoms was larger than 100% and thus the imidazolyl anion conclusion was drawn.
Figure 11.5 The three different forms of His. (a) Protonated form which exists at pH < 6.0, (b) De-protonated form which exists at physiological pHs between 6.5-7.5 and (c) Imidazolyl anion which exists once both the N-Hydrogen atoms have been extracted.

In the 65% MPD, 0.1M citric acid crystallisation conditions at pH 4.0, one would expect the His to be in its protonated form once again. Indeed, one week after crystal growth, one molecule of carboplatin is seen bound to the Nδ atom (Figure 11.4a). Interestingly, when a second crystal was measured 4 weeks later, a molecule of carboplatin was seen bound to both the Nδ and Nε atoms (Figure 11.4b). Since this crystallisation condition contains citrate ions the imidazolyl anion can again be formed, but since the citrate is at a much lower concentration compared to the Cl⁻ ions used previously (0.1M compared to 1.4M) we can expect the imidazolyl anion to form more slowly. Also, due to the summed occupancy being less than 100% for these two Pt atoms (Table 11.2), the histidine tautomers could be a possibility.

For the crystals grown in 2M NH₄ formate, 0.1M HEPES at pH 7.5, these crystals did grow quickly, after a few days, and the first XRD dataset was also fairly promptly measured, but no binding to His-15 was seen (results not shown). However, when a second crystal was measured around 18 weeks after initial crystallisation set up (Table 11.1), a molecule of carboplatin was seen bound namely to the Nδ atom of His-15 (Figure 11.4e). At pH 7.5, the His residue will already exist in its de-protonated form, thus only the other N-Hydrogen atom needs to be extracted to form the imidazolyl anion. This crystallisation mix contained a high concentration (2M) of formate ions (HCOO⁻), thus one can expect these formate ions to extract the second N-Hydrogen atom, leaving two N atoms with lone pairs of electrons to bind to the Pt centre of carboplatin. Instead, we see a 2Fo-Fc electron density peak 3.4 Å away from the Nε atom, and thus we cannot interpret it as a Pt atom even if it might be weakly bound i.e. due to this unusually large distance. Also, no anomalous difference electron density is seen at this Nε position either.

11.4.2.2 The differing levels of structural detail seen of the carboplatin

From all the crystal structure results, we do see different amounts of detail in the Nδ and Nε binding sites for the carboplatin molecule (Figure 11.4). The crystal at pH 5.6 shows the most detail in the Nδ binding site (Figure 11.4d), with more atoms of the CBDC moiety (Figure
11.1) being modelled in, apart from the four carbon ring structure, which still cannot be modelled due to the lack of electron density seen (Figure 11.4d). A plausible explanation for not seeing this four carbon ring structure is due to the dynamic disorder already indicated in the small molecule crystal structure of carboplatin made by Beagley et al, 1985.

**11.4.2.3 Checks on the possible contamination of chloride ions from the HEWL lyophilised powder**

As the lyophilised HEWL powder does still contain trace quantities of Cl ions, we naturally checked for any evidence of their binding. So, looking at the anomalous difference electron density of all the studies here, down to 2.0 sigma possible signal levels, only the structure at pH7.5 has any sign of such an anomalous peak i.e. with one of 2.5σ for just one of these usual Cl sites, whereas the other usual Cl binding sites did not contain any anomalous peaks.

**11.5 Conclusions**

Co-crystallisation of HEWL and carboplatin in NaBr was carried out, with the XRD results showing that the carboplatin molecules indeed suffered a chemical conversion. This was to the trans bromine form of cisplatin, and which contrasts with the NaCl case where chloro cisplatin was seen (Tanley et al 2013b). Since a portion of the CBDC moiety was still also present this confirms that the conversion of carboplatin to its trans bromo form is partial.

Under non-NaCl crystallisation conditions several of the HEWL crystals obtained led to carboplatin binding alone being seen. Out of the four such crystal conditions obtained, the crystals grown in 0.1M Na citrate, 20% propanol, 20% PEG4000 at pH 5.6 show the most detail we have ever seen for the carboplatin molecule, with only the 4 carbon ring structure not being present in the electron density maps.

**11.6 Acknowledgements**

We are grateful for research support from the Universities of Konstanz, Manchester and Utrecht. ST is funded under an EPSRC PhD Research Studentship at the School of Chemistry, University of Manchester.
11.7 References


### 11.8 Supplementary Materials

#### Table S11.1 The 48 non-NaCl crystallisation screens from Hampton research.

<table>
<thead>
<tr>
<th>Screen Type</th>
<th>Buffer Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% MPD, 0.1M Citric Acid pH 4.0</td>
<td>0.1M NaAc, 2M NH₄SO₄</td>
</tr>
<tr>
<td>10% MPD, 0.1M NaAc pH 5.0</td>
<td>0.2M NH₄Ac, 0.1M Nacitrate, 30% MPD</td>
</tr>
<tr>
<td>10% MPD, 0.1M MES pH 6.0</td>
<td>0.1M NaHEPES, 1.4M Nacitrate</td>
</tr>
<tr>
<td>10% MPD, 0.1M HEPES pH 7.0</td>
<td>0.2M NH₄SO₄, 30% PEG4000</td>
</tr>
<tr>
<td>10% MPD, 0.1M NaAc pH 8.0</td>
<td>0.1M NaHEPES, 0.8M NaK tartrate</td>
</tr>
<tr>
<td>20% MPD, 0.1M NaAc pH 5.0</td>
<td>0.1M NaHEPES, 10% propanol, 20% PEG 4000</td>
</tr>
<tr>
<td>40% MPD, 0.1M Citric Acid pH 4.0</td>
<td>0.1M Nacacodylate, 1.4M NaAc</td>
</tr>
<tr>
<td>40% MPD, 0.1M NaAc pH 5.0</td>
<td>0.5M LiSO₄, 15% PEG 8000</td>
</tr>
<tr>
<td>40% MPD, 0.1M MES pH 6.0</td>
<td>0.2M CaAc, 0.1M Nacacodylate, 18% PEG 8000</td>
</tr>
<tr>
<td>40% MPD, 0.1M HEPES pH 7.0</td>
<td>0.1M NaAc, 2M Naformate</td>
</tr>
<tr>
<td>65% MPD, 0.1M Citric Acid pH 4.0</td>
<td>0.1M NaHEPES, 2% PEG 400, 2M NH₄SO₄</td>
</tr>
<tr>
<td>65% MPD, 0.1M NaAc pH 5.0</td>
<td>0.2M NH₄Ac, 0.1M NaAc, 30% PEG 4000</td>
</tr>
<tr>
<td>65% MPD, 0.1M MES pH 6.0</td>
<td>0.2M NaAc, 0.1M Nacacodylate, 30% PEG 8000</td>
</tr>
<tr>
<td>65% MPD, 0.1M HEPES pH 7.0</td>
<td>0.2M NH₄SO₄, 0.1M NaAc, 25% PEG 4000</td>
</tr>
<tr>
<td>0.1M NaAc, 2.0M NH₄SO₄</td>
<td>10%PEG 6000, 0.1M HEPES pH 7.5, 5% MPD</td>
</tr>
<tr>
<td>0.1M NaAc, 8% PEG 400</td>
<td>0.2M Nacitrate, 0.1M NaHEPES, 30% MPD</td>
</tr>
<tr>
<td>0.2M NH₄Ac, 0.1M NaAc, 30% PEG 4000</td>
<td>0.4M NaK Tartrate</td>
</tr>
<tr>
<td>0.1M NaHEPES, 1.6M NaKPO₄</td>
<td>20% Jeffamine 500, 0.1M HEPES pH 7.5</td>
</tr>
<tr>
<td>0.1M Nacitrate, 1M NH₄PO₄</td>
<td>70% MPD, 0.1M HEPES pH 7.5</td>
</tr>
<tr>
<td>0.1M NaAc, 2M NaFormate</td>
<td>0.2M NH₄SO₄, 0.1M NaAc, 25% PEG 4000</td>
</tr>
<tr>
<td>0.1M imidazole, 1M NaAc</td>
<td>2M NH₄ formate, 0.1M HEPES pH 7.5</td>
</tr>
<tr>
<td>0.2M Nacitrate, 0.1M NaHEPES, 20% propanol</td>
<td>0.2M NH₄SO₄, 0.1M NaAc pH 4.6, 30% PEG 2000</td>
</tr>
<tr>
<td>0.1M Nacitrate, 20% propanol, 20% PEG 4000</td>
<td>10% PEG 8000, 0.1M HEPES pH 7.5, 8% Ethylene Glycol</td>
</tr>
<tr>
<td>0.2M Nacitrate, 0.1M Na cacodylate, 30% propanol</td>
<td>0.1M NaAc, 8% PEG 4000</td>
</tr>
</tbody>
</table>
An unusual feature in the pH7.5 crystal structure at Cys 6-127

The dataset from the pH7.5 crystal showed unusual 2Fo-Fc electron density as well anomalous difference electron density near one of the disulphide bridges (Cys 6-127) (Figure S11.1). The distance between these two peaks of 3.4Å and their close proximity to the disulphide itself (3 to 4 Å) would suggest reduced disulphide due to X-ray radiation damage (Helliwell, 1988). It did not prove possible to retain good polypeptide geometry and place each sulphur into the two peaks. As an additional check the first and last run of the XRD data collection was processed and the protein model refined separately to confirm whether this density was due to an X-ray radiation damage effect. However, the 2Fo-Fc electron density as well the anomalous difference density is seen in the first data collection run as well as the last run. Another possible interpretation could be an alternative starting conformation of the disulphide bond. The crystallisation mix contained HEPES buffer which contains sulphate ions. Thus, based on the anomalous difference electron density, the density might be due to two SO$_4$ ions, each partially occupied, as they would repel each other.

Figure S11.1 2Fo-Fc electron density as well as anomalous difference electron density nearby the Cys6-127 disulphide bond. Sulphur atoms are placed into the density to show the distances between the disulphide bond and these peaks.

Chapter 12: X-ray crystal structures of cisplatin binding to histidine in a protein at multiple temperatures reveals their structural dynamics

Simon W.M. Tanley\textsuperscript{a} and John R. Helliwell\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a}School of Chemistry, University of Manchester, Brunswick street, Manchester, Greater Manchester, M13 9PL, England

Correspondence email: john.helliwell@manchester.ac.uk

In preparation for submission to Acta Cryst F

JRH came up with the experiments. SWMT grew all the crystals, mounted them and measured them on a home source Bruker APEX II diffractometer. SWMT processed all the datasets using the Bruker internal software package and refined all the structures. JRH steered the triclinic study towards structural dynamics as a charge density study of a single occupancy cisplatin was not visible. SWMT and JRH wrote the manuscript.
12.0 Abstract

Cisplatin is a platinum anti-cancer agent whose binding to HEWL, a model protein, has been extensively studied by X-ray crystallography. This new study focuses on co-crystallisation of cisplatin with the triclinic form of HEWL to obtain high resolution structural data on the cisplatin binding sites in the hope of gaining charge density detail. These studies show a versatile binding of cisplatin to the histidine residue, and more versatile than seen in our previous studies in the tetragonal and orthorhombic crystal forms. However, not all of these Pt binding modes are stable. Our efforts to find a crystal form to allow charge density level of details of cisplatin binding to histidine cannot be developed with this triclinic crystallisation condition.

12.1 Introduction

Cisplatin is a platinum anti-cancer agent, used to treat a variety of different cancers. Cisplatin binding to the His-15 residue of HEWL, a model protein has been extensively studied using X-ray crystallography (Casini et al., 2007; Tanley et al. 2012 a; 2012 b; Helliwell and Tanley, 2013). In DMSO co-crystallisation conditions cisplatin binds to the Nδ and Nε atoms of His-15, (Helliwell and Tanley, 2013; Tanley et al., 2012 b) and even after prolonged chemical exposure in aqueous medium (Tanley et al., 2012 a). The impetus for this study looked at growing triclinic HEWL crystals co-crystallised with cisplatin to study the binding of cisplatin to His-15 at atomic resolution. HEWL has been previously studied at atomic resolution of 0.65Å in the triclinic form (Wang et al., 2007) and also via neutron crystallography (Bon et al., 1999). Thus crystallising cisplatin with triclinic HEWL could lead to obtaining atomic resolution of cisplatin bound to His-15 of HEWL, which therefore, could even lead to charge density data for the cisplatin ligands.

This study confirms the binding of cisplatin to His-15 in this triclinic crystal form and the best resolution obtained, was to 0.98Å resolution using a home source Cu Kα CCD diffractometer. The results show a versatile binding of cisplatin to His-15, and more versatile than seen in our previous studies in the tetragonal and orthorhombic crystal forms. The use of multiple temperatures either side of the protein glass transition, (anticipated being 180K), reveals that not all these Pt cisplatin binding modes are stable, which is again different to our results in the tetragonal and orthorhombic crystal forms.
12.2 Methods

12.2.1 Crystallisation conditions

Triclinic HEWL crystals were grown using a similar method to Wang et al (2007), however the batch method was used with 40mg HEWL (2.7mM) co-crystallised with 2.6mg cisplatin (8.1mM), with the platinum compounds being in a 3-fold molar excess to the protein. 462.5μl of a 0.02M NaAc solution along with 462.5μl of a 0.5M NaNO₃ solution was used with 75μl DMSO added. Crystals were kept at 294K.

12.2.2 X-ray data collection strategy, structure solution and refinement

X-ray diffraction data from three cisplatin HEWL triclinic crystals were collected at data collection temperatures of 150K, 200K and room temperature (RT) i.e. approximately 295K. Paratone was used as the cryoprotectant for the datasets collected at 150 and 200K, with a 0.4-0.5mm loop used to mount the crystals. For the RT dataset, a 0.7mm quartz capillary was used to mount the crystal. The three diffraction data sets were collected on a Bruker APEX II home source CCD diffractometer, with the detector placed 40mm away from the crystal using an X-ray wavelength of 1.5418Å. The internal data collection strategy program was used for each crystal to get the most out of the data, with multiple passes used with different swing angles, sample rotation ranges and exposure times used. All data were processed using the internal Bruker APEXII software.

All crystal structures were solved using molecular replacement with PHASER (McCoy et al., 2007) and restrained refinement in REFMAC5 (Vagin and Teplyakov, 2010) in CCP4i, using the lysozyme structure 2w1y as the molecular search model (Cianci et al., 2008). Model building, adjustment and refinement were carried out using the COOT (Emsley & Cowtan, 2004) molecular-graphics program and REFMAC5 (Vagin & Teplyakov, 2010), refining individual anisotropic B factors in CCP4i for the datasets at 150 and 200K and isotropic B factors for the dataset at RT. Cisplatin ligand binding occupancies were calculated using SHELXL (Sheldrick, 2008). Crystallographic and refinement parameters are summarized in Table 12.1. The standard uncertainties on the quoted B factors were derived using SHELXL and the full matrix inversion.
Table 12.1 X-ray crystallographic data and protein model refinement statistics for all three crystals collected at different data collection temperatures

<table>
<thead>
<tr>
<th>Unit cell parameters (Å)/(°)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a = 26.99</td>
<td>a = 26.77</td>
<td>a = 27.34</td>
</tr>
<tr>
<td>b = 31.81</td>
<td>b = 31.38</td>
<td>b = 32.13</td>
</tr>
<tr>
<td>c = 34.07</td>
<td>c = 33.86</td>
<td>c = 34.29</td>
</tr>
<tr>
<td>α = 89.08</td>
<td>α = 88.90</td>
<td>α = 88.04</td>
</tr>
<tr>
<td>β = 72.00</td>
<td>β = 72.31</td>
<td>β = 71.17</td>
</tr>
<tr>
<td>γ = 67.81</td>
<td>γ = 68.46</td>
<td>γ = 68.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDB id’s</th>
<th>4mwk</th>
<th>4mwm</th>
<th>4mwn</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data collection temperature (K)</th>
<th>150</th>
<th>200</th>
<th>294</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal size (mm)</td>
<td>0.6</td>
<td>0.3</td>
<td>0.25</td>
</tr>
<tr>
<td>Total absorbed X-ray dose (MGy)*</td>
<td>0.31</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>Crystal growth time</td>
<td>5 weeks</td>
<td>8 days</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Observed Reflections</td>
<td>202732</td>
<td>118846</td>
<td>112029</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>51605</td>
<td>37288</td>
<td>19160</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>29.28-0.98</td>
<td>32.09-1.12</td>
<td>32.31-1.42</td>
</tr>
<tr>
<td>(1.02-0.98)</td>
<td>(1.15-1.12)</td>
<td>(1.15-1.42)</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>90.6 (51.7)</td>
<td>95.1 (84.6)</td>
<td>99.2 (98.4)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>0.045 (0.209)</td>
<td>0.087 (0.179)</td>
<td>0.145 (0.554)</td>
</tr>
<tr>
<td>Mean I/sig(I)</td>
<td>15.6 (3.2)**</td>
<td>5.9 (2.1)</td>
<td>6.8 (1.1)£</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.5 (0.6)</td>
<td>2.2 (0.9)</td>
<td>5.7 (2.0)</td>
</tr>
<tr>
<td>Cruickshank DPI (Å) for coordinate error</td>
<td>0.022</td>
<td>0.037</td>
<td>0.084</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>8.5</td>
<td>13.0</td>
<td>15.4</td>
</tr>
<tr>
<td>R factor/ R free (%)</td>
<td>11.7 / 14.5</td>
<td>14.7 / 18.7</td>
<td>20.8/ 23.5</td>
</tr>
<tr>
<td>R factor all</td>
<td>11.9</td>
<td>14.9</td>
<td>21.0</td>
</tr>
<tr>
<td>RMSD bond lengths (Å)/ Angles (°)</td>
<td>0.039 / 2.907</td>
<td>0.023 / 2.323</td>
<td>0.021 / 2.247</td>
</tr>
<tr>
<td>Ramachandran Favoured</td>
<td>97.5</td>
<td>97.5</td>
<td>96.0</td>
</tr>
<tr>
<td>Ramachandran Allowed</td>
<td>2.5</td>
<td>2.5</td>
<td>3.2</td>
</tr>
<tr>
<td>Ramachandran Disallowed</td>
<td>0</td>
<td>0</td>
<td>0.8$</td>
</tr>
</tbody>
</table>

* Supplementary Table 1 shows the calculation used to estimate the total absorbed dose per dataset.

** The edge of the detector was at 0.98 Å resolution, thus no more data available to drop the <I/SigI> value to 2.0

£ Resolution where <I/SigI> crosses 2 = 1.61 Å
The residue in the disallowed region of the Ramachandran plot is Asn 103, and which is part of a turn region in the protein.

12.3 Results

12.3.1 Cisplatin binding to His15 of triclinic HEWL

Cisplatin binds to both the N\(\delta\) and N\(\varepsilon\) atoms of His-15 in triclinic HEWL. For each dataset collected at the three different data collection temperatures (150K, 200K and RT), there is evidence for versatile binding (Figure 12.1) due to the presence of multiple anomalous difference density peaks in close proximity to one another as well as large peaks in the 2Fo-Fc density maps (Table 12.2), with each Pt atom having a lower binding occupancy (Table 12.3) than previously observed (Tanley et al., 2012a; 2012b; Helliwell & Tanley 2013). However, the summed occupancies of the multiple Pt atoms in two of the N\(\varepsilon\) binding sites are larger than we have seen before (Tanley et al., 2012a; 2012b; Helliwell & Tanley 2013). The N\(\delta\) binding site only shows a split occupied Pt studied at 150K data collection temperature (Figure 12.1a). However, the N\(\varepsilon\) binding site at each temperature shows multiple Pt positions (Figures 12.1a-c). Due to previously seeing slight differences in the electron density maps from using different processing programs (Tanley et al., 2013a; 2013b), XDS (Kabsch, 1988) was used to check the electron density around the His-15 binding sites. Supplementary Figure 12.1 shows the XDS electron density around the His-15 binding sites for the 200K structure as a representative example. The electron density maps are very similar between the processing programs, thus three Pt atoms are also seen in the XDS processed dataset and the same resolution cut off limit was used for the data.
Figure 12.1 Cisplatin bound to the Nδ and Nε atoms of His-15. (a) 150K, (b) 200K and (c) RT. 2Fo-Fc electron density map (Purple) shown at a cut off level of 1.5rms. Anomalous difference electron density (orange) shown at the 3σ cut off level. The molecular graphics program, COOT was used to prepare the figures (Emsley & Cowtan, 2004)
Table 12.2 Anomalous difference electron density and 2Fo-Fc density peak heights (as a multiple of σ or rms respectively) for each Pt atom per dataset at the Nδ and Nε binding sites.

<table>
<thead>
<tr>
<th></th>
<th>Nδ Binding site</th>
<th>Ne binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pt 1</td>
<td>Pt 2</td>
</tr>
<tr>
<td></td>
<td>Anom 2Fo-Fc</td>
<td>Anom 2Fo-Fc</td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>crystal at 150K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>crystal at 200K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>4.6</td>
<td>9.0</td>
</tr>
<tr>
<td>crystal at RT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12.3 Occupancy values (%) of each Pt atom at the Nδ and Nε binding sites of His-15. Occupancies were calculated using SHELX L (Sheldrick, 2008) and the errors on the occupancies were calculated using the full matrix inversion in SHELXL.

<table>
<thead>
<tr>
<th></th>
<th>Nδ Binding site</th>
<th>Ne binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pt 1</td>
<td>Pt 2</td>
</tr>
<tr>
<td></td>
<td>Anom</td>
<td>Anom</td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>22 +/- 5</td>
<td>13 +/- 12</td>
</tr>
<tr>
<td>crystal at 150K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>29 +/- 10</td>
<td>14 +/- 9</td>
</tr>
<tr>
<td>crystal at 200K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisplatin HEWL</td>
<td>36 +/- 2</td>
<td></td>
</tr>
<tr>
<td>crystal at RT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12.3.2 Structural dynamics of the Pt atoms studied at multiple temperatures bound to triclinic HEWL.

Due to the Pt atoms showing versatile binding across all three temperatures studied (Figure 12.1), the mean u square displacement values (⟨u^2⟩) of each Pt atom were calculated based on their B factors using the formula B=8π^2⟨u^2⟩. Figure 2 shows the plot of the ⟨u^2⟩ against temperature for each Pt atom in the binding sites, which is based on a similar graph to Figure
One of the Pt atoms at the data collection temperature of 150K (Ne Pt 2) has an unusually high B factor (144 Å²) and hence high <u²> (1.82 Å²). This Pt atom has the highest occupancy seen in this study. The Pt atoms at 150K have higher <u²> compared to the overall average protein atom (0.11), which is the same for the RT Pt atoms (<u²> value for average protein atom is 0.18), whereas at 200K, some of the Pt atoms in fact have a lower <u²> compared to the average protein atom (0.13), which is unusual as Pt is the heavier atom. The Nδ Pt 2 atom as well as the Ne Pt 3 and 4 atoms are not shown in this Figure due to them only being present in one or two of the temperatures, thus not enough data points to see the trend with temperature.

Figure 12.2 Mean u square displacement values for the Pt atoms plotted against the data collection temperature. The trends for the Ne Pt 3 atom follow a quadratic curve, , where at 150 and 200K the displacement values are similar but is larger at RT, similar to the trend seen in Doster et al (1989). The other Pt atoms do not show this trend and some of the Pt atoms, Nδ Pt1 and Ne Pt1 at 150K and 200K respectively have a lower mean <u²> displacement value compared to the average protein atom at that data collection temperature (orange line). The Ne Pt 4 atom has a large B factor and hence large <u²> and has been excluded from the comparison.
12.4 Discussion

The initial impetus to undertake this study was to determine cisplatin’s binding to a protein at atomic resolution to seek atomic resolution and even possibly charge density data for these ligands. Triclinic HEWL has been studied previously at 0.65Å (Wang et al., 2007) and with neutron crystallography (Bon et al., 1999), thus co-crystallising cisplatin with the triclinic form of HEWL could have resulted in reaching such diffraction resolutions of the cisplatin molecule in its bound state.

The results for all of the crystals showed cisplatin’s versatile binding at either one or both of the N binding sites of His-15 (Figure 12.1), confirmed by the presence of multiple anomalous difference electron density peaks and 2Fo-Fc density peaks (Table 12.2). Thus multiple binding modes (up to four at the Nε binding site and up to two at the Nδ binding site) are seen. The occupancy values for the Pt atoms are individually lower compared to those previously seen in the tetragonal HEWL crystal cases (Table 12.3), apart from Pt 2 in the Nε binding site of the 150K crystal with occupancy of 60%, which is similar. This Pt does however have a very large B factor (144Å), but a B factor of 144Å² leads to that Pt atom scattering at 1% at 4Å resolution compared with its scattering intensity at the centre of the pattern. This is then adequate to be resolved as a separate atom from its neighbours, which is indeed the case in the 2Fo-Fc and anomalous difference electron density maps as well as 144Å² being computed by REFMAC. However, the summed occupancies for two of the Nε binding sites is larger than what has been seen previously (Tanley et al., 2012 a, 2012b; Helliwell and Tanley, 2013) suggesting that potential toxicity can be higher, and depends on the chemical mix we have used in vitro and possibly therefore in vivo as well. In addition, in combination therapy (chemo and radiation) against cancer of cisplatin binding to cancerous cellular DNA and X-ray radiation therapy targeted specifically onto a tumour, higher occupancy of cisplatin binding to tumour proteins and their histidines would cause stronger X-ray absorption and likely cancer cell killing effect. See appendix 1 (Section 12.9) for a short topical review on this.

In this triclinic system, not all of the Pt cisplatin binding modes are stable i.e. with high B factors developing in the Nε Pt 2 case (Figure 12.2) and multiple Pt atoms are seen bound to the His-15 residue. In a previous study, using tetragonal HEWL (Helliwell & Tanley, 2013), we showed that cisplatin binding was stable up to a dose of 1.7MGy. The total absorbed X-
ray dose of each dataset in this study was 0.31MGy, 0.37Mgy and 0.48MGy (Table 12.1 and Table S12.1), which shows that the cisplatin molecule should be stable at these lower total absorbed X-ray doses for each dataset, but in this triclinic case they are not stable.

Due to seeing this split occupancy of the Pt atom bound to His15 of triclinic HEWL at different data collection temperatures, trying to get to even higher resolutions (0.6Å) using synchrotron radiation sources would also most likely lead to split occupancy for the Pt atoms, thus charge density data on these ligands would be unattainable in this crystal system.

12.5 Conclusions

These studies show a versatile binding of the cisplatin to a histidine in the triclinic form of HEWL, and more versatile than seen in our previous studies in the tetragonal and orthorhombic crystal forms. Thus multiple cisplatin binding modes (up to four at the Nε binding site and up to two at the Nδ binding site) are seen. Not all of these Pt cisplatin binding modes are stable i.e. with extremely high B factors developing in one case. Also, our efforts to find a crystal form to allow charge density level of details of cisplatin binding to histidine cannot be developed with this triclinic crystallisation condition.

12.6 Acknowledgements

JRH is grateful to the University of Manchester for general support, to the EPSRC for a PhD studentship to SWMT and to the School of Chemistry for crystallisation, X-ray diffraction facilities and computing facilities.
### 12.7 Supplementary material

#### Table S12.1 Total X-ray absorbed dose for each crystal

<table>
<thead>
<tr>
<th></th>
<th>150K</th>
<th>200K</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Formula used taking into account the solvent contents</td>
<td>$\text{C}<em>{645.003} \text{N}</em>{143.4081} \text{O}<em>{514.3016} \text{S}</em>{1.0037} \text{Na}<em>{3.58} \text{Cl}</em>{1.026} \text{Pt}_{1.55}$</td>
<td>$\text{C}<em>{635.003} \text{N}</em>{141.4081} \text{O}<em>{513.3016} \text{S}</em>{1.0037} \text{Na}<em>{3.58} \text{Cl}</em>{5.026} \text{Pt}_{1.09}$</td>
<td>$\text{C}<em>{621.004} \text{N}</em>{132.6099} \text{O}<em>{510.802} \text{S}</em>{17.0037} \text{Na}<em>{4.58} \text{Cl}</em>{1.032} \text{Pt}_{1.052}$</td>
</tr>
<tr>
<td>X-ray Intensity (photons/sec/mm$^2$)</td>
<td>$5 \times 10^9$</td>
<td>$5 \times 10^9$</td>
<td>$5 \times 10^9$</td>
</tr>
<tr>
<td>Total exposure time (sec)</td>
<td>40400</td>
<td>39980</td>
<td>55090</td>
</tr>
<tr>
<td>Sample size (mm)</td>
<td>0.60</td>
<td>0.30</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.30</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Sample cross sectional area presented to the incident X-ray beam (mm$^2$)</td>
<td>0.36mm$^2$</td>
<td>0.09mm$^2$</td>
<td>0.0625mm$^2$</td>
</tr>
<tr>
<td>Sample volume (cm$^3$)</td>
<td>$(0.06 \times 0.06 \times 0.05) \times 1.8 \times 10^{-4}$ cm$^3$</td>
<td>$(0.03 \times 0.03 \times 0.02) \times 1.8 \times 10^{-5}$ cm$^3$</td>
<td>$(0.025 \times 0.025 \times 0.02) \times 1.25 \times 10^{-5}$ cm$^3$</td>
</tr>
<tr>
<td>X-ray beam fraction absorbed</td>
<td>$1 - e^{-\mu d}$ ($\mu = 1.76$ mm$^{-1}$, d= 0.5 mm)</td>
<td>$1 - e^{-\mu d}$ ($\mu = 1.70$ mm$^{-1}$, d= 0.2 mm)</td>
<td>$1 - e^{-\mu d}$ ($\mu = 1.48$ mm$^{-1}$, d= 0.2 mm)</td>
</tr>
<tr>
<td></td>
<td>= 0.58 (58%)</td>
<td>= 0.28 (28%)</td>
<td>= 0.25 (25%)</td>
</tr>
<tr>
<td>Photon energy (eV)</td>
<td>8042</td>
<td>8042</td>
<td>8042</td>
</tr>
<tr>
<td>Conversion factor 1 eV to 1 Joule</td>
<td>$1.6 \times 10^{-19}$ J</td>
<td>$1.6 \times 10^{-19}$ J</td>
<td>$1.6 \times 10^{-19}$ J</td>
</tr>
<tr>
<td>Crystal mass (g) = (density (g/cm$^3$) x volume (cm$^3$))</td>
<td>$0.954$ g/cm$^3$ x $1.8 \times 10^{-4}$ cm$^3$ = 1.72x10$^{-5}$ g</td>
<td>$0.973$ g/cm$^3$ x $1.8 \times 10^{-5}$ cm$^3$ = 1.75x10$^{-5}$ g</td>
<td>$0.925$ g/cm$^3$ x $1.25 \times 10^{-5}$ cm$^3$ = 1.15x10$^{-5}$ g</td>
</tr>
<tr>
<td>Absorbed X-ray dose (J/g); (a) data sets 1 and 2 (b) datasets 3 and 4</td>
<td>$5 \times 10^9$ x 0.36 x 4.04x10$^4$ x 0.58 x 8042 x $1.6 \times 10^{-19}$ / $1.72 \times 10^{-4}$</td>
<td>$5 \times 10^9$ x 0.09 x 3.998x10$^4$ x 0.28 x 8042 x $1.6 \times 10^{-19}$ / $1.75 \times 10^{-5}$</td>
<td>$5 \times 10^9$ x 0.0625 x 5.509x10$^4$ x 0.25 x 8042 x $1.6 \times 10^{-19}$ / $1.15 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>= $315.53$ J/g</td>
<td>= $370.39$ J/g</td>
<td>= $481.55$ J/g</td>
</tr>
<tr>
<td></td>
<td>= 0.31 MGy</td>
<td>= 0.37 MGy</td>
<td>= 0.48 MGy</td>
</tr>
<tr>
<td>Total X-ray absorbed Dose</td>
<td>0.31 MGy</td>
<td>0.37 MGy</td>
<td>0.48 MGy</td>
</tr>
</tbody>
</table>
Cisplatin binding to the Nδ and Nε atoms of His-15 processed via XDS. The 2Fo-Fc map (Purple) is shown at the 1.5σ cut-off level. The detail is very similar to that seen in Figure 12.1b which was processed using the Bruker APEXII internal software program.

12.8 References


The relevance to medical treatments of these results is interesting since the summed occupancies of cisplatin binding to the Nε histidine is larger in this study than we have seen before and suggests that potential toxicity can be higher. In combination therapy (chemo and radiation) against cancer, a higher occupancy of any high-Z element binding to tumour proteins and/or DNA would cause a stronger X-ray absorption and likely cancer cell killing effect, obviously a benefit. In particular, gold nanoparticles have been extensively studied in combination with radiation therapy to enhance the killing of tumour cells (Hainfeld et al., 2004, 2010; Liu et al., 2008). Hainfeld et al (2004) showed that intravenous administration of small gold nanoparticles (1.9nm) can deliver high levels of gold to tumours with specificity.
and thereby improve X-ray radiation therapy. This was the first publication to show control of a malignant tumour through preferential absorption of X-rays by high-Z-nanoparticles (Hainfeld et al., 2004). Other high-Z elements have been previously used with X-ray radiation therapy, including; Iododeoxyuridine (IUdR) (Nath et al., 1990) and cisplatin (Biston et al., 2004). IUdR requires 22-45% of the thymine in cellular DNA to be substituted with iodouracil, which is difficult to control to just tumour cells and only gives a small enhancement over radiation therapy on its own (Nath et al., 1990). Iodine contrast media (Mello et al., 1983; Rose et al., 1999) has been used but is of low molecular weight and diffuses out of the normal vascular system, so delivery to the tumour is compromised. Gold has an advantage over Iodine as its K-edge is farther away from bone and tissue absorptions and the gold nanoparticles have a longer blood half-life (Hainfeld et al., 2004).

Cisplatin administration with radiation therapy using a synchrotron source has been studied (Biston et al., 2004) but the non-uniform delivery system meant it wasn’t tumour specific and only 30μM Pt was incorporated. This low Pt concentration and the fact that there was no observable improvement in effect either below or above the K-edge of Pt shows a weak significance for using this high-Z element. However, there was a 3 fold enhancement of effect with cisplatin then just the radiation therapy on its own which can be explained by the radiation improving the penetration of the drug and resulting in improved pharmacological effect (Biston et al., 2004). Carboplatin could be used instead of cisplatin to increase the concentration of the Pt atoms bound to the tumour as it is less toxic than cisplatin so can be administered at a 5 times higher dose (Jandial et al., 2009). However, using cisplatin and radiation therapy in malignant gliomas caused no effect on the tumours, in a study of 285 patients (Garfield et al., 1991). Whereas, in a single case study with a patient with a malignant glioma (Witman et al., 1982), after administration of cisplatin with radiation therapy, the tumour regressed. Thus, the use of the high-Z elements and radiation therapy are cancer specific, but overall the use of these heavy atoms and radiation therapy causes stronger X-ray absorption by the tumours and enhances tumour killing.

By seeing cisplatin binding to a His in a model protein at an overall higher occupancy could thus be related to cisplatin binding to both DNA and tumour proteins. In this case, the cisplatin would be at the heart of the tumour cells' functioning enzymes as well as binding to normal cells, therefore, spatially targeting the X-ray therapy beam to the tumour would thus cause increased X-ray absorption and likely cancer cell killing effect.
References


Overall conclusions to part II

In summary, these chapters have shown a variety of new results; (1) Two molecules of cisplatin are seen bound to HEWL in DMSO media using the co-crystallisation method. (2) Two molecules of cisplatin are bound to HEWL in aqueous media after a prolonged chemical exposure of 13 months. (3) Cisplatin is stable up to 1.78 MGy of radiation when bound to a protein. (4) We are the first academic group to make our raw diffraction data images freely available. (5) Carboplatin partially converts to cisplatin due to the high Cl concentration used in the co-crystallisation conditions. (6) We have found 4 new non-NaCl conditions which grew tetragonal crystals and showed carboplatin binding. (7) We have also seen binding of just one molecule of carboplatin to the His-15 residue as a function of chemistry, pH and elapsed time. (8) Using NaBr crystallisation conditions, we see partial conversion of carboplatin to the Br form with a portion of the CBDC moiety still present. (9) In the NaBr case we see conversion to the trans platinated form and not the cis palatinated form as seen in the Cl crystallisation conditions. (10) Using triclinic HEWL co-crystallisation with cisplatin studied at 3 different data collection temperatures showed a more versatile binding and an overall larger summed occupancy at the Ne binding site.
Part III Heparanase

Introduction to part III

HPSE-1 is a protein over-expressed in the majority of cancers, and its over-expression is likely to cause tumour metastasis and angiogenesis. The experimental 3D structure has not been elucidated. Thus the structure is of importance in the design of active site specific inhibitors of this enzyme.

The purpose of these chapters is: (Chapter 13) To attempt structure determination using previously grown HPSE-1 crystals using the MR technique due to heavy metal soaks failing. (Chapter 14) Describe the over-expression of HPSE-1 using the insect cell expression system, as well as purifying the product. (Chapter 15) Describe the production of a HM of HPSE-1 based on the recently solved structure of a GH 79 family member (the same family as HPSE-1) and to perform a VS study to identify potential novel small molecule inhibitors using this new HM. All chapters are unpublished work.
13. Chapter 13: Heparanase three-dimensional structure solving using pre-existing data

Simon W.M. Tanley\textsuperscript{a}, Naomi Chayen\textsuperscript{b}, Edward McKenzie\textsuperscript{c}, Stuart Fisher\textsuperscript{a} and John R. Helliwell\textsuperscript{a}

\textsuperscript{a}School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, England

\textsuperscript{b}Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, London, SW7 2AZ, England

\textsuperscript{c}Manchester Interdisciplinary Biocentre (MIB), University of Manchester, 131 Princess Street, Manchester M1 7DN, UK, England

Unpublished work.

The protein used in this study was made by EM, then at Oxford Glycosciences and now at the University of Manchester. The crystals were grown by NC at Imperial College London and the diffraction data was collected on a microfocus beamline at the Swiss Light Source (SLS). SF and JRH processed the native and heavy atom derivatives in space group P2\textsubscript{1}. This work was done prior to SWMT starting his PhD.

SWMT has used MR using homologous structures of 20-25\% sequence identity to try solving the structure of this dataset as the crystals were disordered and the heavy atom soaks did not work. SWMT used the PHYRE-2 webserver to create a HM based on the 50kDa subunit of HPSE-1. SWMT wrote the chapter with editing by JRH
13.1 Introduction

HPSE-1 is an endo-β-D glucuronidase enzyme (Zetser et al., 2003) whose primary function is to cleave HS chains in the BM of cells and is responsible for remodelling the ECM (McKenzie et al., 2003; Simizu et al., 2004). This activity is important in embryonic development and the enzyme is up regulated in inflammation and wound healing (McKenzie et al., 2003; Zetser et al., 2003). However, over-expression of this enzyme leads to increased angiogenesis and is associated with tumour metastasis (Parish et al., 2001). More details are given in chapter 3.

HPSE-1 is a 543 amino acid protein (Baker et al., 1999; Dong et al., 2000), comprised of an N-terminal signal peptidase (SP (Met1 – Ala 35)) which is cleaved within the endoplasmic reticulum to generate the latent 65kDa pro-enzyme form (Parish et al., 2001; McKenzie et al., 2003). The pro-form is further processed by cleaving the 6kDa linker peptide between Ser 110 and Gln 157 bought about by Cathepsin L (Abboud-Jarrous et al., 2005; McKenzie, 2007; Vreys & David, 2007), to form the mature active heterodimer consisting of the N-terminal 8kDa subunit (Gln 36 – Glu 109) non-covalently linked to the C-terminal 50kDa subunit (Lys 158 – Ile 543) (Levy-Adam et al., 2003; McKenzie et al., 2003; McKenzie, 2007).

As HPSE-1 is implicated in a variety of cancers, the development of inhibitors is very important. The 3D crystal structure of HPSE-1 is yet to be elucidated, but the 3D structure is important for the design of active site specific inhibitors. HM’s of HPSE-1 have been published (see Chapters 3 and 15 for more detail), which suggest that it is similar to the GH’s in fold, catalytic mechanism and sequence similarity (Zhou et al., 2006; Sapay et al., 2012). The 50kDa subunit of HPSE-1 is 20% sequence identical to xylanase (PDB code 1BG4), whereas it is 25% sequence identical to a beta-glucuronidase enzyme from acidobacterium capsulatum (PDB code 3VNY) (Michikawa et al., 2012).

This chapter focuses on solving the 3D structure of HPSE-1 using MR due to heavy atoms not binding to the protein during heavy metal soaks. Also, the diffraction pattern at certain parts of the crystal had streaky spots and disorder, making structure determination difficult. The expression and purification of HPSE-1 will be given in more details in chapter 14, as the expression of HPSE-1, its crystallisation and data collection outlined in this chapter were carried out prior to the start of my PhD.
13.2 Methods

13.2.1 Heparanase expression and purification

The active form of HPSE-1 was expressed as a dual expression construct, with the 8kDa and 50kDa subunits PCR amplified from their respective templates, cloned into pZero blunt vectors, excised and cloned individually into digested pAcUW 51 vectors. These vectors were co-transfected with baculovirus DNA and grown in insect cells (SF9). Purification of HPSE-1 was achieved using heparin affinity chromatography. The methods used for cloning, expression and purification are exactly the same as in McKenzie et al, 2003.

13.2.2 Crystallisation conditions

The active HPSE-1 form to be crystallised has the sequence:

“8kDa” Subunit:

ADPGQDVVD LDFFTQEPLH LVSPSFLSVTI DANLATDPRFLILLGSPKL RTLARGLSPA YLRFGGTKTD FLIFDPKKE

“50kDa” subunit:

ADPGK KFKNSTYSRS SVDVLYTFANCSGLDLFLG ALLRTADLQ WNNSNAQLLL DYCSSKGYNISWELGNEPN SFLKKADIFINGSQGGRAYKLLRSTFKNALKLYGPDVGQPRRTAKMLKSKLADGGEVIDSVTWXHHYLLNGRTA TRE DFLNPDVLDI FISSQKVFQ VVESTPQKVVWLETSSAY GGGAPLLSDTFAAGFMWLKD LGLSARMGIE VVMRQVFPPGA GNYHVDENSENF DLPDPYWLSSL FKKVLGTGKVMASVQGSKR RKLRVLHLCT NTDPNYKEG DLTYAINGH NVTKYLRLPY PFSNQVDKYLLRPLGPHGL LSKSVQLNGL TLKMDQTL PPLMEKPLRP GSSLGLPAFS YSFFVIRNAKVAACI

With the N-terminus of both subunits containing the ADPG run of amino acids which are added in the cloning step as they are needed for expression of the dual expression vector. HPSE-1 was concentrated to 11mg/ml in 25mM Tris buffer pH 7.5 with 20mM NaCl and crystallisation occurred in MES buffer pH 6.0 (Figure 13.1). Crystals were soaked with PbCl₂, OsCl₂ and Hg for 24 hours.
13.2.3 Data collection strategy

Data collection was carried out on a micro focus beam line at the SLS synchrotron where 180° of the crystals were measured with 1° rotations. 4 native datasets, a Pb soak, a Hg soak and an Os soak dataset were all collected. The best native dataset, the Pb dataset and Hg dataset could all be processed in space group P2 by Dr. S. Fisher. A summary of the data collection and processing parameters using the c unit cell in space group P2 are shown in Table 13.1.

Table 13.1: Summary of the unit cell parameters, data collection and processing parameters using the c unit cell in space group P2. Values shown in parentheses are for the outer resolution bin. These statistics were provided by Dr S. Fisher who processed these datasets in space group P2.

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Lead</th>
<th>Mercury</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit cell parameters (Å/°)</strong></td>
<td>a = 79.21 b = 71.03</td>
<td>a = 79.62 b = 71.23</td>
<td>a = 79.61 b = 71.17</td>
</tr>
<tr>
<td></td>
<td>c = 49.14 β = 91.36</td>
<td>c = 47.73 β = 92.70</td>
<td>c = 47.57 β = 91.99</td>
</tr>
<tr>
<td><strong>Crystal class</strong></td>
<td>Monoclinic</td>
<td>Monoclinic</td>
<td>Monoclinic</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>41.31 – 2.50</td>
<td>20.66 - 2.59</td>
<td>52.88 - 3.20</td>
</tr>
<tr>
<td><strong>Observed Reflections</strong></td>
<td>53023</td>
<td>50960</td>
<td>19204</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>18228</td>
<td>15948</td>
<td>6783</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>95.9 (95.6)</td>
<td>93.2 (66.0)</td>
<td>76.5 (78.3)</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>2.9 (2.9)</td>
<td>3.1 (1.8)</td>
<td>2.8 (2.9)</td>
</tr>
</tbody>
</table>

**Figure 13.1:** HPSE-1 native crystal grown in MES buffer pH 6.0
### 13.2.4 Three dimensional structure determination of human heparanase-1

Out of the three datasets collected, the Pb derivative dataset was used for MR. Both the Pb and Hg derivative datasets did not contain heavy atoms bound to the protein based on both the anomalous difference and Patterson maps.

#### 13.2.4.1 Using 1BG4 as molecular replacement search model

Xylanase (PDB code IBG4) (Schmidt et al., 1998) has 20% amino acid sequence identity to the 50kDa subunit of HPSE-1, contains the TIM-barrel fold and uses the same catalytic mechanism. 1BG4 was used as it has been previously used as a template to produce a HM of HPSE-1 (Zhou et al., 2006; Gandhi & Mancera, 2011; Sapay et al, 2012). 1BG4 was aligned with the 50kDa subunit using CLUSTALW (Thompson et al., 1994) and the alignment file imported into CHAINSAW (Schwarzenbacher et al., 2004; Stein, 2008). CHAINSAW works by modifying 1BG4 by pruning the non-conserved residues. This pruning can be done either, to the gamma atom, the beta atom or retains all atoms common to the target and model residues. CHAINSAW also mutates residues from the target model to the alignment sequence to produce a simplified search model. This simplified search model is then input into PHASER (McCoy et al., 2007) along with the Pb derivative dataset. Rigid body refinement (RB) and restrained refinement (RR) in REFMAC5 (Vagin & Teplyakov, 2010) was used along with amplitude twin refinement. The molecular graphics programmes COOT (Emsley & Cowtan, 2004) and BUCCANEER (Cowtan, 2000) were used to build the missing parts of the structure into the remaining electron density.
13.2.4.2 Using 3VNY PHYRE-2 homology model as the molecular replacement search model

3VNY is a beta-glucuronidase enzyme from *acidobacterium capsulatum* (Michikawa *et al.*, 2012), and is a member of the GH 79 family and is the first structure to be solved of this family of proteins, which includes HPSE-1. 3VNY has a sequence identity of 25%, whilst covering 70% of the 50kDa subunit of HPSE-1. PHYRE-2 (Kelley & Sternberg, 2009), a homology modelling webserver was used to build a HM of HPSE-1 using the 50kDa sequence and 3VNY as the template structure (see Chapters 2.1.2.3.1 and 15 for more details). The PHYRE-2 HM was then used in PHASER (McCoy *et al.*, 2007), with REFMAC5 (Vagin & Teplyakov, 2010), COOT (Emsley & Cowtan, 2004) and BUCCANEER (Cowtan, 2000) being used as before to build the structure.

13.3 Results

13.3.1 HPSE-1 expression, purification and crystallisation

The active form of human HPSE-1 was expressed as a dual construct in insect cells and purified using a heparin binding affinity column. The sample purified as a doublet peak in the chromatogram. The active form crystallised in MES buffer pH 6.0 (Figure 13.1).

13.3.2 Data collection

Native and heavy atom derivative datasets were collected at the SLS synchrotron and the diffraction pattern at different angular rotations of all crystals showed both order and disorder (Figure 13.2). The disorder in the crystal makes the solving of the structure more complicated, as it is difficult to obtain the spacings between spots in the disordered pattern, making space group determination challenging. From the output of PHASER, the crystal was said to be twinned. A reason for the disorder could be due to the protein being purified as a doublet, meaning the crystal could contain both forms of the protein, affecting the crystal packing and hence resulting in the disorder seen in the diffraction pattern (Figure 13.2b).
**Figure 13.2:** Diffraction pattern of the HPSE-1 crystal. (a) The ordered diffraction pattern, with clear rounded spots. (b) The disordered diffraction pattern with streaky spots.

### 13.3.3 Solving the structure of HPSE-1 using the Pb derivative raw diffraction data and molecular replacement

Both 1BG4 and 3VNY were used as starting models, with 1BG4 having sequence identity of 20%, with the 50kDa subunit of HPSE-1, whereas 3VNY has sequence identity of 25%. Both use an acid catalysis mechanism for the hydrolysis of glycosidic bonds and have a proton donor and nucleophile which are conserved in HPSE-1 (Glu-225 and Glu-343). Figure 13.3 shows the similarity between 1BG4 and 3VNY based on their 3D structure. The TIM barrel fold is clearly seen in both structures; however 3VNY is the larger of the two proteins.

**Figure 13.3:** The 3D structure of 1BG4 (a) and 3VNY (b).
13.3.3.1 1BG4 Molecular replacement

*CLUSTALW* and *CHAINSAW* were used to produce a simplified search model based on 1BG4. *PHASER* was used for MR and the translational function score (TFZ) was 3.2. Z scores in X-ray crystallography are a statistical measurement looking at the number of standard deviations (sigmas) over the mean (Z-score) (http://www.phenix-online.org/documentation/phaser.htm). In MR *PHASER*, the rotational function (RF) and TF are both determined by Z scores. An RFZ score will only be found once a translation has been performed in the MR calculation. A TFZ of 3.2 represents a standard deviation of less than 5 and hence relates to the difficulty of solving this dataset, with standard deviations of less than 5 meaning *PHASER* hasn’t solved the structure, whereas standard deviations greater than 8 means *PHASER* has definitely solved the structure (http://www.phenix-online.org/documentation/phaser.htm). However, even though the TFZ score is 3.2, the crystal packing (Figure 13.5) shows gaps between the molecules, which could represent the space for the 8kDa subunit to be modelled in as only the 50kDa subunit was used as the starting search model. Also, as a test, the OMIT map electron density around an α-helix was well defined (Figure 13.6). Another test of how good this model is, is to look at the catalytic residues (Glu-225 and Glu-343, but in the model, they are Glu-69 and Glu-187). Figure 15.7 shows the position of these residues, however, it is clear that these are not in close proximity to each other and neither are they in the TIM barrel domain, which is where the active site is believed to be located.

In *COOT*, manual model building was undertaken to fit the missing side chains/residues based on the remaining electron density. The run of amino acids which include, WHHYY, is an ideal starting point for model building as all five residues have aromatic side chains and are the easiest residues to model, if the data quality is good enough (Figure 13.8), except for Cys and Met residues if anomalous data is obtained. HPSE-1 contains four Cys residues in the 50kDa subunit and one in the 8kDa subunit. Four of these residues are thought to form disulphide bridges, (C127-179 and C437-542), with C127 being in the 8kDa subunit and the other three from the 50kDa subunit (Gandhi & Mancera, 2011). From the 1BG4 simplified search model and the electron density data, there is no way of building in this disulphide bond due to the lack of anomalous difference electron density.
**Figure 13.4:** *CLUSTALW* alignment file between 1BG4 and the 50kDa subunit of HPSE-1. The key corresponds to * meaning the residues are conserved, : and ., meaning the residues are similar and – showing gaps between the sequences.

**Figure 13.5:** Two different views of the crystal packing.
Figure 13.6: Omit map electron density around one of the alpha helices. (a) The view down the axis of the helix, (b) The perpendicular view of the helix fitting the electron density. The electron density is shown at the 1.5 r.m.s contour level.

Figure 13.7: Catalytic Glu residues in the 1BG4 model. (a) The residues are not in close proximity to undertake the catalytic mechanism. (b) The 2Fo-Fc electron density around Glu 69 (i.e. Glu 225). This electron density covers the majority of the side chain. The electron density is shown at the 1.5 r.m.s contour level.
Figure 13.8: The WHHYY amino acid sequence with the 2Fo-Fc electron density map shown. The electron density covers the W, one H and one Y residue, whereas, the other H and Y residue has no electron density, making these residues hard to model in. The electron density is shown at the 1.5 r.m.s contour level.

BUCCANEER in CCP4i, was used to automate model building. Table 13.2 gives the R/Rfree values for each round of refinement and model building. After each round of refinement, the density modification (DM) program in CCP4i was used to improve the quality of maps. DM works by applying real space constraints based on known features of a protein electron density map in order to improve the approximate phasing obtained from experimental source. This was done using solvent flattening, which reduces the noise in the disordered solvent region, sharpening the map around the protein atoms. The best model had R/Rfree values of 38.9%/46.2%, with 363 out of the 386 residues modelled. 16 of these un-modelled residues are from the C-terminus region (residues 370-386), which is highly flexible and hard to model. Figure 13.9 represents the packing of the best model, showing the solvent channels, which one would hypothesise the 8kDa subunit should fit into based on Figure 13.4, but no electron density for this 8kDa subunit is seen. The number of residues in the disallowed regions of the Ramachandran plot (Table 13.3) shows the geometry of the protein is very poor and with relatively high R/Rfree values, the structure cannot be reliably solved.
**Table 13.2:** R and R free values for the HPSE-1 structure model after each refinement step.

<table>
<thead>
<tr>
<th>Refinement Step</th>
<th>R Factor (%)</th>
<th>R free (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB after PHASER</td>
<td>54.2</td>
<td>53.6</td>
</tr>
<tr>
<td>RR after PHASER</td>
<td>45.6</td>
<td>48.2</td>
</tr>
<tr>
<td>RR using DM map</td>
<td>43.5</td>
<td>48.3</td>
</tr>
<tr>
<td>BUCCANEER 1 with RR</td>
<td>42.6</td>
<td>47.9</td>
</tr>
<tr>
<td>BUCCANEER 2 with RR</td>
<td>40.1</td>
<td>47.9</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>39.9</td>
<td>47.9</td>
</tr>
<tr>
<td>Remove unknown atoms from BUCCANEER run. Has residues 1-341, then RR</td>
<td>41.9</td>
<td>46.9</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>41.8</td>
<td>46.5</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>41.6</td>
<td>47.8</td>
</tr>
<tr>
<td>Use last output coordinate file as search model in PHASER, then RR</td>
<td>46.0</td>
<td>50.0</td>
</tr>
<tr>
<td>RR from new PHASER output</td>
<td>41.0</td>
<td>47.0</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>40.0</td>
<td>45.4</td>
</tr>
<tr>
<td>BUCCANEER with RR</td>
<td>39.7</td>
<td>45.4</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>40.0</td>
<td>46.0</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>40.0</td>
<td>46.4</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>40.0</td>
<td>46.7</td>
</tr>
<tr>
<td>BUCCANEER with RR</td>
<td>38.8</td>
<td>46.3</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR</td>
<td>38.9</td>
<td>46.2</td>
</tr>
</tbody>
</table>
**Figure 13.9:** 2Fo-Fc electron density map (Purple) around the best model, with the crystal packing also shown. The solvent channels can be seen (gaps between the density), with no extra density observed for the 8kDa subunit. The electron density is shown at the 1.5 r.m.s contour level.

**Table 13.3:** Data processing and refinement statistics.

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th>1BG4 molecular replacement model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2_1</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td>a = 79.23 \ b = 70.86 \ c = 47.43 \ β = 92.89</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>50960</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15139</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>79.14 – 2.60 (2.66 – 2.60)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.7 (90.2)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.106 (0.234)</td>
</tr>
<tr>
<td>&lt;I&gt;/ &lt;σ(I)&gt;</td>
<td>4.3 (2.8)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.1 (1.8)</td>
</tr>
<tr>
<td>Cruickshank DPI (Å)</td>
<td>0.191</td>
</tr>
<tr>
<td>Wilson B factor</td>
<td>49.8</td>
</tr>
<tr>
<td>Average B factor</td>
<td>51.4</td>
</tr>
<tr>
<td>R factor/R free</td>
<td>38.9/ 46.2</td>
</tr>
<tr>
<td>R.m.s.d bond length (Å)/ Angles (°).</td>
<td>0.013/ 2.063</td>
</tr>
<tr>
<td>Ramachandran favoured (%)</td>
<td>46.1</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>23.6</td>
</tr>
<tr>
<td>Ramachandran disallowed (%)</td>
<td>30.3</td>
</tr>
</tbody>
</table>
13.3.3.2 Using the PHYRE-2 HPSE-1 homology model based on 3VNY as the molecular replacement search model

3VNY is a GH 79 family member whose 3D structure was recently solved (Michikawa et al., 2012) and is 25% sequence identical to the 50kDa subunit of HPSE-1, with 70% coverage. The PHYRE-2 webserver was used to create a HM of the 50kDa subunit based on the structure of 3VNY and this model is now used as the MR search model (see Chapters 2 and 15 for more details).

The PHYRE-2 model has 361 out of 386 residues modelled with 100% confidence and 94% coverage. All side chain residues are also built in the PHYRE-2 model whereas, the side chains are pruned in the 1BG4 CHAINSAW simplified search model.

Once again, PHASER, RB, RR, manual model building in COOT and automated model building in BUCCANEER were used to improve the structure. Table 13.4 gives the refinement steps along with the R/Rfree values at each stage.

Table 13.4: Refinement steps, with the R/Rfree values at each stage.

<table>
<thead>
<tr>
<th>Refinement step</th>
<th>Rfactor (%)</th>
<th>Rfree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB after PHASER</td>
<td>50.9</td>
<td>48.2</td>
</tr>
<tr>
<td>RR after PHASER</td>
<td>47.7</td>
<td>49.7</td>
</tr>
<tr>
<td>RR after PHASER with 30 cycles of refinement</td>
<td>44.5</td>
<td>49.9</td>
</tr>
<tr>
<td>RR after PHASER with 50 cycles of refinement</td>
<td>43.9</td>
<td>50.2</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR with 10 cycles</td>
<td>42.3</td>
<td>50/2</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR with 30 cycles</td>
<td>42.4</td>
<td>49.9</td>
</tr>
<tr>
<td>BUCCANEER 1 with RR</td>
<td>43.9</td>
<td>49.6</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR with 10 cycles</td>
<td>42.6</td>
<td>49.8</td>
</tr>
<tr>
<td>BUCCANEER 2 with RR</td>
<td>42.8</td>
<td>51.1</td>
</tr>
</tbody>
</table>
Residues added/removed or fitted into electron density, then RR with 10 cycles

BUCCANEER 3 and Residues added/removed or fitted into electron density, then RR with 10 cycles

BUCCANEER 4 with RR

Residues added/removed or fitted into electron density, then RR with 10 cycles

Model refine in COOT then RR

Added side chains that fitted into electron density then RR

<table>
<thead>
<tr>
<th>Treatment</th>
<th>R-factor</th>
<th>R-free</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues added/removed or fitted into electron density, then RR with 10 cycles</td>
<td>42.8</td>
<td>47.2</td>
</tr>
<tr>
<td>BUCCANEER 3 and Residues added/removed or fitted into electron density, then RR with 10 cycles</td>
<td>44.8</td>
<td>48.5</td>
</tr>
<tr>
<td>BUCCANEER 4 with RR</td>
<td>44.9</td>
<td>47.8</td>
</tr>
<tr>
<td>Residues added/removed or fitted into electron density, then RR with 10 cycles</td>
<td>44.9</td>
<td>48.3</td>
</tr>
<tr>
<td>Model refine in COOT then RR</td>
<td>42.6</td>
<td>47.2</td>
</tr>
<tr>
<td>Added side chains that fitted into electron density then RR</td>
<td>42.8</td>
<td>46.5</td>
</tr>
</tbody>
</table>

The best structure obtained from this refinement has higher R/Rfree values than the 1BG4 MR final structure. Also, the crystal packing (Figure 13.10), and the electron density around the WHHYY peptide (Figure 13.11) is not as good as the equivalent for the 1BG4 MR final structure (Figures 13.4 and 13.7). The catalytic Glu residues are shown in Figure 13.12, and compared to Figure 13.8, this structure has the catalytic residues in the TIM-barrel pointing towards the solvent channel and they are in close proximity to each other so are in potentially the correct place for the catalytic mechanism to occur. Table 13.5 gives a summary of the data collection statistics and the refinement statistics for the 3VNY PHYRE-2 MR search model. Again, the number of disallowed residues in the Ramachandran plot suggests that the geometry of the structure is again poor, but is better than the final structure when using the 1BG4 simplified search model for MR.

Figure 13.10: Overall crystal packing of the final model, showing residues overlapping between neighbouring molecules.
**Figure 13.11:** Electron density around the WHHYY peptide sequence. The density is not continuous across the protein backbone and overall not as good for the side chains as Figure 13.8 for the 1BG4 solution. 2Fo-Fc electron density is shown at the 1.5 r.m.s contour level.

**Figure 13.12:** Catalytic Glu residues in the TIM-β-barrel pointing towards the solvent channel (a) with the 2Fo-Fc electron density around the two Glu residues also shown (b). 2Fo-Fc electron density is shown at the 1.5 r.m.s contour level.
**Table 13.5:** Data collection, processing and refinement statistics of the PHYRE-2 3VNY MR search model using the Pb derivative dataset.

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th>PHYRE-2 3VNY molecular replacement model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.1 (1.8)</td>
</tr>
<tr>
<td>Unit cell parameters (Å/°)</td>
<td>a = 79.62  b = 71.23  c = 47.73  β = 92.70</td>
</tr>
<tr>
<td>Cruickshank DPI (Å)</td>
<td>1.102</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>50960</td>
</tr>
<tr>
<td>Wilson B factor</td>
<td>49.8</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15524</td>
</tr>
<tr>
<td>Average B factor</td>
<td>47.5</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>39.62 – 2.59 (2.65 – 2.59)</td>
</tr>
<tr>
<td>R factor/R free</td>
<td>42.8/46.5</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.2 (76.4)</td>
</tr>
<tr>
<td>R.m.s.d bond length (Å)/Angles (°)</td>
<td>0.032/ 4.013</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.106 (0.234)</td>
</tr>
<tr>
<td>Ramachandran most favoured (%)</td>
<td>57.8</td>
</tr>
<tr>
<td>&lt;I&gt;/ &lt; σ (I)&gt;</td>
<td>4.3 (2.8)</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>21.1</td>
</tr>
<tr>
<td>Ramachandran disallowed (%)</td>
<td>21.1</td>
</tr>
</tbody>
</table>

**13.3.3.3 Solving the structure of human HPsE-1 in space group P1**

Due to the structures solved in space group P2 having poor protein geometry but reasonable R/Rfree values and electron density for the α-helices, the space group was checked using the Zanuda webserver (http://www.ysbl.york.ac.uk/YSBLPrograms/). From the output of Zanuda, (see section 13.5) it recommends that the original space group was incorrect and the correct space group is P1. The Pb dataset was reprocessed in P1 and MR was again used with the PHYRE-2 3VNY search model. In the P1 space group, there are now 2 protein molecules in the asymmetric unit (Figure 13.13).

Table 13.6 gives the refinement steps and the R/Rfree values at each stage. This structure now gives the best R/Rfree values seen thus far (37.6/39.0) and the geometry of the protein has improved, but still has 5.8% of the residues falling in the disallowed regions of the Ramachandran plot (Figure 13.14).
Table 13.6: Refinement steps, with the R and Rfree values at each step

<table>
<thead>
<tr>
<th>Refinement step</th>
<th>Rfactor (%)</th>
<th>Rfree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB after PHASER</td>
<td>40.9</td>
<td>41.4</td>
</tr>
<tr>
<td>RR after PHASER</td>
<td>39.1</td>
<td>41.3</td>
</tr>
<tr>
<td>BUCCANEER 1 with RR with 10 cycles</td>
<td>37.5</td>
<td>40.4</td>
</tr>
<tr>
<td>Refined the Ramachandran plot and built in missing residues, then RR with 10 cycles</td>
<td>37.2</td>
<td>39.5</td>
</tr>
<tr>
<td>BUCCANEER 2 with RR</td>
<td>37.6</td>
<td>39.0</td>
</tr>
</tbody>
</table>
The structure solved in space group P1 now gives the best refinement statistics. However, this structure still has problems with geometry (Figure 13.14) as well as the electron density maps not showing density for extra residues. Due to this, BALBES (Long et al., 2008)(http://www.ysbl.york.ac.uk/~fei/balbes/) was used for MR and Arp/wArp (Langer et al., 2008)(http://www.embl-hamburg.de/ARP/) for model building, again using the Pb derivative dataset and the PHYRE-2 3VNY HM. This was to compare the different programs, to see if one could achieve better results. From the output of Arp/wArp, 3 chains were found, with chain A being more or less complete, whereas chains B and C were around half complete with these chains containing different residues, i.e. if these two chains could be put together then this would make a nearly complete chain and hence 2 molecules in the asymmetric unit (Figure 13.15).

Overall, this gives the best R/Rfree values seen thus far (Table 13.7, 32.8/37.7), but the model is worse due to these 3 separate chains as well as a large number of outliers in the Ramachandran plot (15.4%) (Figure 13.16) compared to Figure 13.14. Solving the structure in P1 space group has improved the R/Rfree values as well as the model geometry (Figure 13.14). However, the electron density is still poor and trying to manually improve the

---

**Figure 13.14** Ramachandran plot of the protein residues in the preferred, allowed and disallowed regions.
structure by adding in missing residues or missing side chains makes the structure worse, thus, the structure is incomplete.

Figure 13.15 Output from Arp/wArp, where 3 chains have been found in the data.

Table 13.7: Refinement steps, with the R/Rfree values at each step after using Arp/Warp and Balbes to solve the model

<table>
<thead>
<tr>
<th>Refinement step</th>
<th>Rfactor (%)</th>
<th>Rfree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB after Balbes/ArpWarp</td>
<td>37.0</td>
<td>40.1</td>
</tr>
<tr>
<td>RR after Balbes/ArpWarp with 10 cycles</td>
<td>34.5</td>
<td>39.1</td>
</tr>
<tr>
<td>RR after Balbes/ArpWarp with 30 cycles</td>
<td>33.6</td>
<td>40.2</td>
</tr>
<tr>
<td>RR and TLS after Balbes/ArpWarp with 30 cycles</td>
<td>32.8</td>
<td>37.7</td>
</tr>
<tr>
<td>BUCCANEER with RR and TLS</td>
<td>35.4</td>
<td>41.7</td>
</tr>
<tr>
<td>Residues added/ removed or fitted into electron density, then RR and TLS with 30 cycles</td>
<td>45.5</td>
<td>50.0</td>
</tr>
</tbody>
</table>
13.4 Discussion

HPSE-1 is an important enzyme involved in wound healing and development, but overexpression has been implicated in tumour metastasis and angiogenesis. HPSE-1 expression is known to be proportional to the metastatic effect of a given tumour, thus developing drugs to inhibit this enzyme will hopefully lead to anti-cancer treatments.

Using MR to solve the structure of HPSE-1 is difficult as the highest sequence identity to a protein whose 3D structure is known is 25%, which only covers the 50kDa subunit. Either heavy atom soaks, co-crystallisation with heavy metal ligands or producing protein with seleno methionine is the way forward to solve the structure using IR or AD techniques to solve the phase problem (Chapter 1).

The over-expression of HPSE-1 does have difficulties; mainly that it is a secreted glycosylated protein, meaning expression in insect/mammalian cells is needed (see Chapter 14). Glycosylation adds to the problem of crystallisation, as the sugar molecules on the surface of the protein could disrupt the packing of the crystal due to their flexibility. However, it has also been noted that these sugar molecules can in fact help in the packing of
crystals, making them more ordered and less likely to break during mounting and cryo-freezing (Dao-Thi et al., 1996; Mesters & Hilgenfeld, 2007). The active form which was crystallised here did contain all 6 N-glycosylation sites. The crystals produced were disordered, which could be caused by; the sugar molecules, the fact that the protein purified as a doublet, or the crystallisation conditions need optimising.

Overall, it is a difficult protein to express and when heavy atom soaks fail, MR with low sequence identity homologs (20-25%) is difficult as the model is hard to design. The 50kDa subunit contains a TIM-β-barrel fold which is conserved in the majority of GH’s so is easier to model. It is the 8kDa subunit and the 6kDa linker region of the pro-form of HPSE-1 which are less well conserved so difficult to model, hence why just a model based on the 50kDa subunit was used as a starting point.

One of the difficulties with modelling HPSE-1 is due to the formation of two disulphide bridges, but HPSE-1 contains 5 cysteine residues. It is thought that residues 127 and 179 and 437 and 542 form these disulphide bonds based on experimental results (Simizu et al., 2007) and modelling studies (Gandhi & Mancera, 2011). Modelling the disulphide bonds is important as these give a great indication of how the enzyme folds. The experimental results from Simizu et al, (2007) indicate that the disulphide bond between Cys 437-542 is needed for both secretion and activation of the active enzyme; hence this disulphide bond has to be present in the datasets collected. The search models used for MR do not contain disulphide bridges and no anomalous difference electron density was observed for these S atoms, thus this disulphide bond could not be modelled, thus, making structure determination very difficult.

**13.4.1 Comparison between 1BG4 and 3VNY results**

1BG4 and 3VNY were used as the starting models as they have 20% and 25% sequence identity to the 50kDa subunit of HPSE-1 respectively, contain the TIM barrel fold and are GH enzymes with conserved Glu residues in the active site. 1BG4 has been used previously for homology modelling studies (Zhou et al., 2006; Sapay et al., 2012) and the structure of 3VNY was only recently solved (Michikawa et al., 2012). Using MR with 1BG4 as a simplified search model after CHAINSAW did give better statistics for the structure of HPSE-1 (R/Rfree 38.9%/46.2%) compared to using the PHYRE-2 3VNY HM (R/Rfree 42.8%/46.5%). It was not only the R factors which showed the overall quality difference between the
models, the crystal packing (Figure 13.2 and Figure 13.6) and the electron density around the protein backbone and some aromatic side chains (Figure 13.5) was also better. However, the placement of the catalytic Glu residues in the PHYRE-2 3VNY MR final structure (Figure 13.11) correspond to that of the HM produced by Sapay et al, with both residues being in close proximity, pointing towards the solvent channel and located in the TIM-β-Barrel. The 1BG4 catalytic residues however, were positioned further apart and not in the TIM-β-barrel (Figure 13.7). Hence, both models do have their advantages and disadvantages, with neither of them being 100% correct. Both of these structures show poor protein geometry as the Ramachandran plot has a number of disallowed residues. For both MR models, the TFZ scores were below 5 which corresponds to the difficulty in solving the structure. The reasons for this include: the raw diffraction data images showed disorder (Figure 13.2), the space group could be wrong, or because the MR search models are not good enough due to the low sequence identity.

13.4.2 Solving the structure of HPSE-1 in space group P1
Due to the electron density maps showing promise, especially for the α-helices and having an R-factor below 40%, this implies that the space group could be wrong. Importing the data into Zanuda (section 13.5) suggests that the space group for this data is in fact P1 and not P21. Using the Pb derivative dataset processed in P1 along with the PHYRE-2 3VNY HM sees improvements in the R/Rfree values (32.8/37.7) compared to before, but there are still problems with this structure. Trying to add in missing side chains/residues based on the remaining electron density made the model worse and even though the protein geometry improved, it is still poor. None of these structures include all of the residues or the crucial disulphide bond, thus this data and using MR to solve the structure is proving very difficult.

13.4.3 Future work
Another way to tackle this problem would be to use the PHENIX crystallographic suite of software (Adams et al., 2010). The program MR.ROSETTA (Terwilliger et al., 2012) builds up a search model based on a number of already known protein sequences based on sequence similarity and builds an ensemble of structures. Hence, more than one model is searched using PHASER. So far, using the PHENIX programmes hasn’t improved structure determination but more work is needed to try all possibilities. The best way to try and solve the 3D structure of HPSE-1 however, is to obtain more protein and grow new crystals (See chapter 14).
13.5 Supplementary material

zanuda 1.04 (12 Feb 2012)

coordinates HPSE-1_50kda_3VNY_12.3_parrot1_buccaneer4_refmac7.pdb
data pb_p2_a_200307_scaIa1.mtz
readability test passed (Refmac_5.5.0109)
resolution 2.585
space group P 1 2 1
cell 79.620 71.230 47.730 90.00 92.70 90.00

Step 1.
R-factors for the starting model.
Transformation into a supergroup.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Space group</th>
<th>R.m.s.d.</th>
<th>Refinement in tested group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref</td>
<td>from the</td>
<td>starting</td>
<td>Rigid</td>
</tr>
<tr>
<td></td>
<td>model, A</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>&gt;&gt; 2 P 1 2 1</td>
<td>0.0004</td>
<td>--</td>
<td>0.4821</td>
</tr>
<tr>
<td>2 P 1 2 1</td>
<td>0.0004</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Step 2.
Refinements in subgroups.
There are 2 subgroups to test.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Space group</th>
<th>R.m.s.d.</th>
<th>Refinement in tested group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref</td>
<td>from the</td>
<td>starting</td>
<td>Rigid</td>
</tr>
<tr>
<td></td>
<td>model, A</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>&gt;&gt; 2 P 1 2 1</td>
<td>0.0004</td>
<td>--</td>
<td>0.4821</td>
</tr>
<tr>
<td>1 P 1</td>
<td>1.9979</td>
<td>0.4980</td>
<td>0.4524</td>
</tr>
<tr>
<td>2 P 1 2 1</td>
<td>0.8739</td>
<td>0.5039</td>
<td>0.5001</td>
</tr>
<tr>
<td>&lt;&lt; 1 P 1</td>
<td>1.9979</td>
<td>0.4980</td>
<td>0.4524</td>
</tr>
</tbody>
</table>

Step 3.
Refinement of the best model.
Candidate symmetry elements are added one by one.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Space group</th>
<th>R.m.s.d.</th>
<th>Refinement in tested group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref</td>
<td>from the</td>
<td>starting</td>
<td>Rigid</td>
</tr>
<tr>
<td></td>
<td>model, A</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>&gt;&gt; 1 P 1</td>
<td>1.9979</td>
<td>0.4980</td>
<td>0.4524</td>
</tr>
<tr>
<td>1 P 1</td>
<td>1.9075</td>
<td>0.4980</td>
<td>0.4567</td>
</tr>
</tbody>
</table>
R-factor in the original subgroup is NOT the best. The original space group assignment seems to be incorrect.

13.6 References


14. Chapter 14: Expression and purification of human Heparanase-1

Simon W.M. Tanley\textsuperscript{a}, Naomi Chayen\textsuperscript{b}, Edward McKenzie\textsuperscript{c}, Fred Garzoni\textsuperscript{d}, Alice Aubert\textsuperscript{d}, and John R. Helliwell\textsuperscript{a}

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University Of Manchester, Brunswick Street, Manchester, M13 9PL, England

\textsuperscript{b} Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, London, SW7 2AZ, England

\textsuperscript{c} Manchester Interdisciplinary Biocentre (MIB), University of Manchester, 131 Princess Street, Manchester M1 7DN, UK, England

\textsuperscript{d} EMBL Grenoble, BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Unpublished work

SWMT has worked in EM’s laboratory in Manchester and with FG and AA at the EMBL in Grenoble to try and express the HPSE-1 protein using the insect cell expression system. The protein made was sent to NC for crystallisation trials. SWMT wrote the chapter with editing by JRH and EM.
14.1 Introduction

Human HPSE-1 is an enzyme involved in wound healing and inflammation, but when over-expressed, can lead to tumour metastasis by remodelling both the BM and ECM (see Chapter 3). The 3D structure of HPSE-1 is unknown, and elucidating its structure is vital to designing active site inhibitors which could potentially lead to anti-cancer treatments (see Chapters 13 and 15). This chapter focuses on the over-expression and purification of HPSE-1 using the insect cell expression system (Baculovirus expression system (BEVS)).

BEVS is used for large eukaryotic multi-subunit complexes, large recombinant DNA insertions, high quality product for structural analyses, for secreted and glycosylated proteins and does not require specific safety measures (Trowitzsch et al., 2010). The methods used in Grenoble are slightly different to the ones used in Manchester. The multiBac system was pioneered in Dr Imre Berger’s laboratory (Berger et al., 2004) and improved further (Fitzgerald et al., 2006; Bieniossek and Berger, 2009) resulting in the production of the EMBacY virus. MultiBac uses a Tn7 attachment site in a Lacza gene for integrating foreign genes, with blue-white screening on gentamycin (Gent) plates used to determine successful integration (Trowitzsch et al., 2010). White colonies represent the baculovirus genome which has undergone homologous recombination and contains the plasmid, the gene of interest and the Gent gene, whereas the blue colonies do not contain the gene of interest or Gent gene. The EMBacY cells contain a gene coding for yellow fluorescence protein (YFP), which is used to observe the production of the protein of interest in a fluorescence spectrophotometer (Fitzgerald et al., 2006). Once YFP production reaches a plateau, it correlates with the peak production of the protein.

14.2 Methods

14.2.1 Expression of active and pro-forms of HPSE-1 using the EMBacY method in Grenoble

Full length pro HPSE-1 cDNA and the active form were PCR amplified and the PCR products cloned into pZero blunt transfer vectors (Invitrogen), using the method published by McKenzie et al (McKenzie et al., 2003). The DNA template was the same as the one used to produce the protein in chapter 13.
Aliquots (10μl) of pro and active forms of HPSE-1 DNA constructs were transfected with 50μl of chemical competent DH10EMBacY cells, left on ice for 45 mins and then 43°C for 45s. 600μl of LB media was added and left overnight (ON) at 37°C.

15μl of the cells were diluted in 135μl of media in serial dilutions (1/1; 1/10; 1/100 and 1/1000). 135μl of each dilution were streaked out on to antibiotic plates (Kanamycin (Kan)/Tetamycin (Tet)/Gent/IPTG/BluOGal) and incubated for 24hrs at 37°C. Blue and white colonies grew after 24hrs but were too small to differentiate and were left for a further 24hrs. 5 white colonies and 1 blue colony for control was chosen for both the pro and active forms of HPSE-1 and re-streaked on to Kan/Tet/Gent/IPTG/BluOGal plates and incubated at 37°C ON. For each plate, two colonies were chosen (Pro HPSE-1 #3 and #5; active HPSE-1 #3 and #5) for pre-culture and launched with 2 mls of LB media containing 2μl of Kan, Tet and Gent and incubated at 37°C ON.

**14.2.1.1 Transfection**

Details of the methods used were provided by Mr Fred Garzoni and are outlined below. The pre-cultured cells were centrifuged for 10mins at 4000rpm, the supernatant discarded and 300μl of re-suspension buffer added to the pellet and re-suspended gently. 300μl of lysis buffer was added and gently mixed until blue and 300μl of neutralization buffer added and gently mixed until white. The solution was centrifuged for 10mins at 13000rpm, supernatant kept and centrifuged for a further 5mins at 13000rpm. The supernatant was discarded and 700μl of isopropanol added to the pellet and centrifuged for 10mins at 13000rpm. The supernatant was discarded and 200μl of 70% ethanol was added gently to the pellet and centrifuged for 5mins at 5000rpm. All the ethanol was discarded and 50μl of 70% ethanol was added to the pellet.

In a sterile fume hood, the ethanol was discarded and pellet dried for 10 mins. 20μl of sterile water added and the pellet resuspended gently before addition of 200μl media. The transfection reagent was prepared for 4 plasmids by mixing 400μl of media with 40μl of X-treme gene. 100μl of the transfection reagent was added to each resuspended pellet.

2mls of media was pipetted on to each well of a 6-well plate. 1ml of healthy cells (range 0.5-1 x 10⁶) were pipetted onto 5 wells and left for 15 mins. This gives a control well of media and a cell control well. To two wells, 150μl of pro #3 and 150μl of pro #5 transfection
mixture was added. This was repeated for the prepared active HPSE-1 transfection mixture (#3 and #5).

14.2.1.2 Sample processing and YFP measurements

After 48-60 hrs the supernatant was removed from the two pro #3 wells (~6mls) and put in a falcon tube (V₀). This was repeated for all other wells. 3mls of media were added to each well and left for two to three days. The supernatant was removed and the cells re-suspended in 500μl of phosphate buffered saline (PBS). The re-suspended cells were sonicated for 30s. 50μl was added to protein gel loading buffer (cell extract (SNP)). The remaining 450μl was centrifuged for 3 mins at 13000rpm. 50μl was removed and protein gel loading buffer added (soluble part (SN)). YFP measurement was made using the remaining 400μl of solution to give the amount of desired protein in the cells. With the SNP and SN of each colony, an SDS-PAGE and western blot analysis was performed to confirm amount of desired protein.

14.2.1.3 Virus amplification

25 mls of healthy cells (0.5-1x10⁶ cells) were infected with 3 mls of active or pro HPSE-1 V₀, giving a V₁ amplification. Cells were counted everyday and if they exceeded 1x10⁶, dilution was required as the cells would be still dividing every 24 hours. However, if the number of cells was less than 1x10⁶, then the cells had stopped dividing and have reached Dpa (day after proliferation arrest). Once they reach Dpa, a cell suspension containing 1x10⁶ cells was transferred to an eppendorf and the procedure outlined in section 14.2.1.2 was used to perform YFP measurements and produce SNP and SN fractions for SDS-PAGE and western blots for expression analysis. Cells were counted every 12 hours and the procedure from section 14.2.1.2 repeated until YFP measurement reached a plateau. At this point, cells were harvested by centrifugation at 800rpm for 3 mins. The pellet was kept either at -20°C or used to scale up protein expression. The cells never reached Dpa for either the pro or active forms of HPSE-1, thus infection and production of HPSE-1 is unlikely to occur.

The media from each amplification was also kept and used for SDS-PAGE and western blot analysis. An overview of all the steps used is shown in Figure 14.1.
14.2.2 Cloning, insect cell expression and purification of HPSE-1 in Manchester

Full length, native pro HPSE-1 cDNA from brain, placenta and pancreas cDNA libraries (Clontech) were used as these are abundant sources of HPSE-1. Three archived samples of native pro form of HPSE-1 were re-tested to see whether these still contain the HPSE-1 DNA (samples used in section 14.2.1). Full length pro HPSE-1 cDNA with a TEV cleavable linker (Figure 14.2) inserted before and after the 6kDa linker region was also PCR amplified.
**Figure 14.2:** The left hand panel is the full length sequence of pre-pro HPSE-1 and the right hand panel is the full length sequence of the pro TEV form (minus the signal peptide) of HPSE-1, with TEV sites inserted before and after the 6kDa linker region. Signal peptide is blue, 8kDa subunit in green, 6kDa linker peptide in purple, 50kDa subunit in black and the TEV cleavage sites in red.

### 14.2.2.1 Primer design and PCR

High purity primers (IDT) for the TEV cDNA pro form of HPSE-1 were designed against the DNA sequence of HPSE-1, with the forward primer containing a BglIII restriction site (underlined) 5’- CAC AGA TCT CAG GAC GTG GTG GAC CTG GAC TTC -3’ and the reverse containing an EcoRI restriction site (underlined) 5’- CAC GAA TTC TCA GAT GCA GGC GGC CAC CTT GGC -3’.

The primers for the native pro forms of HPSE-1 were designed based on codon optimised HPSE-1 DNA nucleotide sequencing and again contain the BgIII and EcoRI restriction sites for the forward and reverse primers respectively. (HPSE-1 pro native for: 5’- CAC AGA TCT CAG GAC GTG GTG GAC CTG GAC TTC -3’, HPSE-1 pro native rev: 5’- CAC GAA TTC TCA GAT GCA AGC AGC AAC TTT GGC -3’).

The pro native HPSE-1 and pro TEV HPSE-1 were PCR amplified using standard conditions (Table 14.1):- 35 cycles, with initiation at 98°C for 1 min, annealing and extension at 98°C for 10 seconds, 55°C for 30 seconds and 72°C for 1 min and final elongation for 10 mins at 72°C. All PCR products were run on a 0.8% agarose gel (1.6g/200mls TAE, 10μl dye) for 30 mins at 200V.

**Table 14.1:** Volumes of reagents used for PCR

<table>
<thead>
<tr>
<th></th>
<th>TEV</th>
<th>Brain</th>
<th>Pancreas</th>
<th>Placenta</th>
<th>Pro 1</th>
<th>Pro 2</th>
<th>Pro 3</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>1μl</td>
<td>5μl</td>
<td>5μl</td>
<td>5μl</td>
<td>2μl</td>
<td>2μl</td>
<td>2μl</td>
<td>0μl</td>
</tr>
<tr>
<td>5X Buffer</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
</tr>
<tr>
<td>dNTPs (10mM)</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
</tr>
<tr>
<td>For Primer (10pM)</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
<td>1μl</td>
</tr>
</tbody>
</table>
The PCR products were excised from the gel and cleaned using a QIAquick gel extraction kit (Qiagen). Three gel volumes of buffer QG were added (100μl/100mg) and left for 10mins at 50°C. One gel volume of isopropanol was added and centrifuged for 1min for the DNA to bind and flow through discarded. 750μl of PE ethanol wash buffer was added and left to stand for 5mins before centrifuging for 1min. The flow-through was discarded and centrifuged for a further minute to remove residual buffer. 30μl of EB elution buffer was added to the middle of the membrane, left to stand for 1min and centrifuged for a further minute to elute concentrated DNA.

30μl of the concentrated DNA was restriction digested with 1.5μl each of BglII and EcoRI restriction enzymes (fermentas) in a total volume of 50μl with restriction digest buffer. Digests were incubated for 90mins at 37°C. 250μl of PB buffer was added to the digest mix, centrifuged for 1min and flow through discarded. 750μl of PE buffer added, left to stand for 5mins and centrifuged for 1min. The flow-through was discarded, left to stand for 2mins, centrifuged for a further minute to remove all residual ethanol from the DNA then eluted with 30μl of EB buffer as above. 1-3μl of the purified digested insert was ligated to PAcGP67 vector digested with NamHI and EcoRI enzymes and ligated ON at 4°C with T4 DNA ligase in appropriate buffer (NEB).

### 14.2.2.2 Transformations into competent cells and Mini Preps

0.4μl of BME was added to 50μl of competent cells to increase their transformation efficiency. 5μl of ligated DNA was added and mixed gently, left on ice for 30mins, heat shocked at 42°C for 45s and 200μl of SOC rich medium added and left for 1hr shaking to recover (37°C). 200μl of each was added to ampicillin plates and left ON at 37°C for colonies to grow. Five colonies from each plate were inoculated in LB ampicillin broth and incubated ON at 37°C with shaking (220rpm).
Cells were centrifuged for 1min, supernatant discarded and pellet re-suspended in 250μl of cold P1 buffer with RNaseA. 250μl of buffer P2 was added and gently mixed until blue. 350μl of N3 neutralizing buffer was added and a white precipitant formed (debris material), centrifuged for 10mins at 13000rpm at 4°C. Supernatant was added to a QIAprep spin column, centrifuged for a further minute, flow through discarded and 750μl PE wash buffer was added to the column, left to stand for 5mins and centrifuged for 1min. The flow-through was discarded and the column centrifuged for a further minute to remove any residual wash buffer (ethanol can denature downstream enzymes, e.g. during sequencing of the DNA). 50μl of EB elution buffer was added to the centre of the membrane, left to stand for a minute and centrifuged to elute 50μl of pure DNA.

HPSE-1 DNA was digested with BamHI and EcoRI restriction enzymes at 37°C for 1hr 15mins, placed on ice for 30mins and ran on a 1% agarose gel (2g/200mls 1x TAE buffer, 10μl safe view DNA stain). Two constructs were chosen for sequencing by GATC biotech.

14.2.2.3 Virus amplification and expression
1.2x10^6 SF9 cells were plated onto a well in a 6-well plate. 1.4μl (250ng) of pro TEV HPSE-1 DNA (176ng/μl) was added to 1ml of serum free media with pen/strep antibiotics, 5μl lipofection reagent and 2.5μl baculovirus DNA (transfection mix) and left for 30mins for liposomes to form. Media was removed from the cells and transfection reagent added drop-wise, left ON at 28°C. 1ml of media was added and left for a week to infect the cells. Cells were spun down and secreted virus media added to 50mls of SF9 cells to amplify the virus. After a week the cells had swollen and burst. These were spun down, serum added (2% final concentration) and kept at 4°C.

On a 6 well plate, 1.4mls of Hi5 cells were added (1.4x10^6 cells per well). To each well a different crude concentration of the virus was added giving a cell control, 2, 10, 20, 50 and 100μl virus and left for 72 hrs.

14.2.2.4 Purification
The media was taken from all wells (1ml), 20μl was used neat, 200μl of the media was concentrated seven-fold using viva spin columns (30kDa membrane) and 500μl used for affinity purification with heparin sepharose beads. 40μl of the heparin sepharose beads were spun for 5mins at 2500rpm, ethanol supernatant removed, washed with PBS and centrifuged for 5mins. The PBS was removed, the media added, rotated for 2hrs and centrifuged to
remove media and unbound proteins. Beads were then washed with 150mM NaCl buffer (in PBS), spun for 5mins and washed with 300mM NaCl buffer (in PBS) and spun again. A 2x SDS buffer with BME was added to all samples, vortexed, centrifuged and placed on a 95°C heat block for 5mins before being loaded onto a 10% SDS-PAGE gel.

14.3 Results

14.3.1 HPSE-1 purification using native pro HPSE-1

In Grenoble, the expression of the pro and active forms of human HPSE-1 was carried out in SF21 cells. Only the results from the pro form are shown as the active form of the enzyme was not 100% correct based on sequence analysis. The media for both pro form constructs were concentrated five-fold using viva spin columns, around 100-fold using heparin sepharose beads and the samples were analysed by SDS-PAGE (Figure 14.3a) and western blots (Figure 14.3b). After concentration via the heparin sepharose beads (Lanes 5 and 6), some bands had increased in intensity compared to the five-fold concentrated samples (Lanes 3 and 4). However no bands were seen for the media control before concentration. A number of bands are present in the SDS-PAGE gel, meaning the purification strategy needs optimising. From the western blot analysis (Figure 14.3b), no bands were observed except for the HPSE-1 positive control, which showed an excess of bands, thus is likely to be degraded. The main reason for failing to detect expressed protein was thought to be due to the fact that the clone had been frozen for ≥7 years and nicks in the DNA could have formed during the thawing and re-freezing process which produce mutations in the DNA and hence prevents expression of HPSE-1. This result was expected due to the cells not reaching Dpa or YFP being measured, which leads to the conclusion that the cells weren’t infected with the HPSE-1 DNA.
Figure 14.3: SDS-PAGE gel (a) and western blot (b) analysis of the media from pro form #3 and #5 neat (1 and 2), 5 fold concentrated (3 and 4) and after heparin affinity binding (5 and 6). A HPSE-1 band was seen in the +ve control lane with some bands increased in intensity after heparin binding. No bands were seen in the western blot except for the HPSE-1 +ve control.

14.3.2 Cloning, expression and purification of Pro native form and Pro TEV cleavable forms of HPSE-1

14.3.2.1 Cloning of Pro native and Pro TEV forms of HPSE-1 into GP67 vector

As a result of no protein being expressed in section 14.3.1, a new vector (pAcGP67) was used. Samples described in this section include the native pro-form from placenta, brain and pancreas cDNA libraries, 3 archived pro-form clones (one of which was used in section 14.3.1) and a new TEV cleavable pro-form from a pre-existing library in Dr McKenzie’s group.

Each sample underwent PCR amplification (Table 14.1) and the PCR products were cut with the restriction enzymes BamHI and EcoRI and run on a 1% agarose gel (Figure 14.4). From this gel, the TEV (lane 6), placenta (lane 8), pro 2 form (lane 10) and the cut GP67 vector (lane 3) were excised from the gel. No band of the correct weight (1kb) for the brain (lane 7) and pancreas (lane 9) PCR product was seen, with the placenta product having a weak contamination band at lower molecular weights. All three pro forms (lanes 10-12) had contamination bands with pro 3 (lane 12) having a weak band at the correct molecular weight. Pro 1 and pro 2 (lane 10 and 11) had stronger bands at the correct molecular weight but only pro 2 was used as this represents the same clone that was used in section 14.3.1.
Figure 14.4: PCR products ran on a 1% agarose gel. (1) Markers (2) uncut GP67 (3) Cut GP67 (4) Blank (5) –ve (6) TEV (7) Brain (8) Placenta (9) Pancreas (10) Pro 1 (11) Pro 2 (12) Pro 3

The cleaned products were ligated into cut pAcGP67 vectors, transformed into competent cells and five colonies were selected for pre-culture. These were then mini-prepped, the DNA restriction digested by BamH1 and EcoRI restriction enzymes and run on a 1% agarose gel (Figure 14.5). From this gel, all of the 5 colonies picked from the TEV clone contained the insert, only 1 colony form the placenta cDNA library contained the insert and none of the pro-2 forms contained the insert. Thus, this result with the pro-2 form corroborates with the previous result where no HPSE-1 expression was observed as the control plate had a very low background meaning the clones we used to start with did not contain the correct HPSE-1 DNA anymore (presumably due to 7 years on the shelf affect).

Figure 14.5: Ligated PCR products run on a 1% agarose gel. The markers are to the left of the TEV products, with U being uncut and 1-5 being the 5 different colonies chosen.

From Figure 14.5, two of the TEV colonies were chosen along with colony 4 from placenta and sent off for sequencing by GATC. The placenta reverse primer sequence was confirmed as 100% correct, however it did contain a known polymorphism at position 307, where a Lys residue is mutated to an Arg residue (Figure 14.6). The forward primer sequence however had an oligo mutation and was sent for re-synthesis.
Figure 14.6: Sequencing results from the placenta 4 colony, showing the known polymorphism in the reverse primer.

The TEV sequencing results were similar to the placenta form, as the reverse primer was 100% correct, but the forward primer missed the first 140 residues of the N-terminus (Figure 14.7).
Figure 14.7: Sequencing results for the TEV form. The forward primer is on the left, showing the first 140 residues missing and the right hand panel showing the reverse primer sequence which is confirmed as 100% correct.

Using new purer primers, PCR was repeated for the TEV pro form (Figure 14.8). There was a small band in the negative control, suggesting a small contamination, but the intensity was much stronger than previously seen (Figure 14.4). The TEV DNA was ligated with the cut pAcGP67 vector and five colonies were picked after transformation to undergo pre-culture (Figure 14.9). All five colonies contained the insert and again two were chosen for sequencing with the results shown in Figure 14.10. The sequencing results now correctly confirmed the identity of HPSE-1 as both the forward and reverse primers were 100% correct. Figure 14.11 shows the high quality sequence confirmation at the N-terminus.
**Figure 14.8:** TEV PCR products after using the new higher grade primers along with a negative control. Shows the TEV pro-form has been inserted.

![M TEV -VE](image)

**Figure 14.9:** After ligation of pro-TEV form DNA into GP67 vector and transformation into cells. All 5 colonies chosen contain the TEV insert.

![M U 1 2 3 4 5](image)

**Figure 14.11:** High quality sequence confirmation at the N-terminus.

<table>
<thead>
<tr>
<th>Lab1</th>
<th>Sequence Name: 9158189</th>
<th>Run ended: Feb 9, 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>GCCTTTGGCAGGATCATAACG</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>GCCTTTGGCAGGATCATAACG</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>GCCTTTGGCAGGATCATAACG</td>
<td></td>
</tr>
</tbody>
</table>

D L Q D
Figure 14.10: Sequencing results from clone 3 of the pro-TEV form. Forward primer on the left panel and reverse primer of the right hand panel, with both being 100% correct

14.3.2.2 HPSE-1 expression in insect cells

Clone 3 produced the clearest sequencing results and had a concentration of 176ng/μl of DNA with a purity of 1.63 based on the ratio of 260/280nm absorbance. 250ng of this DNA was co-transfected with baculovirus DNA and subsequent recombinant virus infected into 50mls of SF9 cells to amplify the virus. This was repeated twice to get a higher titre of virus. Crude volumes of this virus (2μl, 10μl, 20μl, 50μl and 100μl) were added to 1.4x10^6 Hi5 cells on a 6-well plate and left for 72 hours for infection to occur (Figure 14.12). From Figure 14.12, one can see that even at 2μl of the virus, the cells were infected when compared to the mock. Infection is characterised by much larger swollen cells, gaps in the well, where the cells have lost their adherence properties, and are less rounded. These changes were seen quite clearly in the wells with 20μl, 50μl and 100μl volumes of the virus.
14.3.2.3 Purification of HPSE-1

14.3.2.3.1 Purification 1

After one round of virus amplification, only 4 samples were used for purification of HPSE-1 (Mock, 2μl, 50μl and 100μl) as the cells didn’t look as infected as in Figure 14.12, meaning the virus was of low titre (data not shown). The media from each sample was used either neat, after seven-fold concentration using viva spin columns with a membrane of 5000Da, or after binding to heparin sepharose beads. Only 10 samples were run on the coomassie stained gel (Figure 14.13), with the intense band around the 72kDa mark being serum albumin. However there was a band underneath the serum band (Figure 14.13) which was not present in the mock lane around the 70kDa mark, attributed to the HPSE-1 pro TEV protein. However, very weak bands were seen in lanes 6-9 which contained the heparin binding samples.

In the western blot analysis (Figure 14.14) there are strong clear bands in lanes 12 and 13 (50 and 100μl heparin binding) with weaker bands seen in lanes 8 and 9 (50 and 100μl seven-fold concentrated samples). These preliminary results obtained with a low virus titre were promising and optimisation of the virus titre and purification steps were required.
Figure 14.13: SDS-PAGE gel of the samples from the first expression study. A large serum band is present as expected, with a smaller band (arrow) being present in all samples apart from the mock which could relate to HPSE-1, running at around 70kDa molecular weight. There are faint bands in lanes 6-9, which correspond to this HPSE-1 band; however, it is very difficult to see.

Figure 14.14: Western blot after 10 minute exposure. The only lanes shown are: (8) 50μl virus (7 fold concentrated), (9) 100μl virus (7 fold concentrated), (10) mock (heparin binding), (11) 2μl virus (heparin binding), (12) 50μl virus (heparin binding) and (13) 100μl virus (heparin binding)

14.3.2.3.2 Purification 2
After re-amplification of the virus, the cells were infected at a higher titre as can be seen at the lowest concentration of virus (2μl) in Figure 14.12. Serum was not added to the virus before infection of the cells so the HPSE-1 band could be visualised on the gel. Serum was however, added to the stock virus to preserve the sample while it was stored at -80°C. This
time only neat and heparin binding samples were tested. After binding to the heparin sepharose beads, 150mM, 300mM and 500mM NaCl 1M Tris HCL pH 7.5 solutions were used for the washing the beads with all wash steps retained.

On the SDS-PAGE gel (data not shown as the gel was used for the western blot), the neat samples show a band increasing with the virus concentration, and this band is not seen in the mock lane. However, the heparin binding lanes have very weak intensity again, thus is difficult to see the HPSE-1 band. From the western blot (Figure 14.15), strong bands around the 72kDa molecular weight mark are seen in the neat samples bar the mock, plateauing around the 20μl sample. Degradation bands are seen in the 50 and 100μl samples. However, no HPSE-1 was detected in the heparin binding samples, except the 100μl sample which is of weak intensity when compared to the neat lane.

Figure 14.15: Western blot analysis of both the neat and heparin binding samples at the 5 different concentrations of virus.

The absence of HPSE-1 protein in the western blot analysis after heparin binding affinity, suggests that the HPSE-1 protein was eluting at low concentrations of salt used in the wash steps. It is known that HPSE-1 should be stable up to 0.6 M NaCl (McKenzie et al., 2003). To determine whether the HPSE-1 protein has eluted in the wash steps, the 50 and 100μl washes were analysed along with the mock 150mM wash as a control. 1ml of each of the washes was concentrated 20-fold using viva spin columns with 10μl loaded on to a gel for coomassie staining and western blot analysis. From the Coomassie gel, again most of the lanes showed weak bands, with some strong bands seen in the 50μl 300mM wash. This band was excised for MS analysis. From the peptide ion MS analysis, the top hit was human
HPSE-1. The other hits were serum and keratin, which are usual contaminants. The peptide sequences covered the 8kDa subunit, the 6kDa linker peptide and the 50kDa subunit (Figure 14.16). On the western blot (Figure 14.17), intense bands were seen in the 300mM wash for both the 50 and 100μl sample, with weaker bands present in both the 150 and 500mM wash steps.

![Figure 14.16: Peptide ion MS analysis. The peptide sequences (highlighted in yellow) from the MS analysis covers the 8kDa subunit (green), 6kDa linker region (purple) and the 50kDa subunit (black) of HPSE-1.](image)

![Figure 14.17: Western blot analysis of the 150, 300 and 500mM wash steps form the 50 μl (lanes 1-3) and 100μl (lanes 4-6) virus samples.](image)

14.3.2.3.3 Purification of a 50mls test expression volume

One of the reasons the HPSE-1 protein could be eluting at lower concentrations than expected, is because 1M Tris HCl pH 7.5 was used in the buffer instead of the more usual 25mM Tris HCl pH 7.5. Hence, in this 50mls test expression, 25mM Tris HCl pH 7.5 buffer was used with differing concentrations of NaCl in the wash steps to determine whether this was the reason for early elution of the HPSE-1 protein. 50mls of Hi5 cells were infected with
1ml of the virus, in the same ratio as 20μl/1ml. The 50mls of media were bound to Sigma heparin beads for two hours, washed with 20mls of an increasing gradient of NaCl washes, (i.e.100, 200, 300, 400 and 500mM NaCl). Bound protein was eluted with 0.5mls 2M NaCl (repeated 4 times to ensure maximum extraction). All the washes were retained and analysed using a western blot (Figure 14.18) along with the elution steps. Western blot analysis after a 2 second exposure showed intense bands in all lanes, with the most abundant being in the 300mM and 400mM wash steps (lanes 3 and 4). Hence, HPSE-1 was again expressed, but once again eluted off much earlier than expected. Thus, in future purification steps, low salt concentrations of 50, 100 and 150mM will be used to wash the beads and 500mM or 1M salt concentrations for the elution steps.

**Figure 14.20:** Western blot analysis using increasing concentrations of NaCl washes (100, 200, 300, 400 and 500mM). Intense bands are seen in the 300 and 400mM wash steps (lanes 3 and 4).

### 14.3.2.3.4 Purification of a 500mls expression in Hi5 cells
Sections 14.2.3.3.1, 14.2.3.3.2 and 14.2.3.3.3 confirmed that HPSE-1 was being expressed and secreted in insect cells and the washing steps had been optimised. A 500mls flask of Hi5 cells were infected with 20mls of virus for 72hrs. The cells were visibly infected (swollen cells but not lysed) suggesting that a higher titre of virus could have been used. Purification was carried out using sigma heparin beads in a peristaltic column setup. 2mls of the beads were washed in PBS and loaded onto the column. 500mls of media was added to the column at a flow rate of 3mls/min. 0.5mls of the starting material and the flow through was kept for analysis. The beads were washed with 30mls of 50mM NaCl in 25mM Tris HCL pH 7.5 and this was repeated with 100mM and 150mM concentrations of NaCl. The washes were collected and kept for analysis. Elution was carried out using 20mls of both 500mM and 1M
NaCl. 1ml fractions were collected and kept for analysis. SDS-PAGE was run using all samples (Figure 14.19a-c). Figure 14.19a shows the 3 wash steps and the first four elutions of 500mM and 1M NaCl, showing intense bands for HPSE-1. Figures 14.19b and 14.19c are the remainder of the elution fractions for both the 500mM and 1M NaCl concentrations respectively. A diffuse band was seen around the 72kDa mark and this was lower than the serum band in the flow through and starting material. The diffuse band is potentially an identification of a glycosylated protein and is of the correct molecular weight to be HPSE-1.

**Figure 14.19a:** SDS-PAGE gel from the heparin affinity columns. Markers (M), starting material (SM), flow through (FT), the 3 wash steps (w1-w3) using 50, 100 and 150mM NaCl, the first four elution’s at 500mM and 1M NaCl concentrations.
**Figure 14.19b:** SDS-PAGE gel analysis of the remainder of the 500mM elution fractions from the heparin affinity column.

**Figure 14.19c:** SDS-PAGE gel analysis of the remainder of the 1M elution fractions from the heparin affinity column.
From these gels, the 20 500mM fractions were pooled together and dialysed against 2L of 25mM Tris pH 7.5 with 150mM NaCl. The sample was concentrated down to a volume of 5mls. 0.5mls was loaded on to an SX2000 gel filtration column with light scattering (Figure 14.20). From the chromatogram, six peaks were seen and the molecular weight and monodispersity of the sample could not be determined. All fractions were collected and the peak fractions were run on a gel to determine which peak represented HPSE-1. From the gel (Figure 14.21), peak 5 (fractions 32-33) contained high levels of HPSE-1, with peak 6 (fraction 36) containing low levels of HPSE-1. These fractions did contain smaller bands (one noticeable one being at around 50kDa) but they were cleaner than before gel filtration (Figure 14.19a). Due to a small amount of HPSE-1 being detected in fraction 36, fractions 29-37 were also analysed (Figure 14.22) to determine when the HPSE-1 protein started to elute off the column. Western blot analysis (Figure 14.22) confirmed that HPSE-1 was eluted off between fractions 30-37, with a smaller degradation band at around 50kDa also being detected in fractions 32-34.

**Figure 14.20:** Gel filtration with light scattering chromatogram. The red trace is the light scattering which shows the polydispersity of the sample (the line decreases) due to the sample not being clean enough. The blue trace is the peaks coming of the chromatogram with peaks 5 and 6 being the HPSE-1 peak.
Figure 14.21: SDS-PAGE gel of the 6 peak fractions from the gel filtration chromatogram. Fractions 32 and 33 have large intense bands for HPSE-1 at around 70kDa with fraction 36 showing a weaker band. There is also a band at around 50kDa in fractions 32 and 33, which could represent a degradation product.
**Figure 14.22:** SDS-PAGE and western blot analysis of fractions 29-37 from the gel filtration column. Western blot shows that HPSE-1 is in all fractions, with fractions 31-34 having a slight degradation product around 50kDa, which could represent the 50kDa subunit after processing.

A Bradford assay was used on fractions 31-36 to determine the protein concentration in each fraction (Table 14.2). These fractions were pooled together to get approximately 500μg/3mls. This 3mls volume was concentrated down to 100μl to obtain approximately 5mg/ml of HPSE-1. The buffer was diluted from 150mM NaCl to 100mM NaCl. This sample was then sent to Prof Naomi Chayen in Imperial College London to set up crystallisation trials; Prof Chayen having been the first to crystallise the protein (Chapter 13).

**Table 14.2:** Concentration of protein per fraction calculated using a Bradford assay.

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Protein concentration (mg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>0.11</td>
</tr>
<tr>
<td>32</td>
<td>0.11</td>
</tr>
<tr>
<td>33</td>
<td>0.08</td>
</tr>
<tr>
<td>34</td>
<td>0.08</td>
</tr>
<tr>
<td>34</td>
<td>0.07</td>
</tr>
<tr>
<td>36</td>
<td>0.04</td>
</tr>
</tbody>
</table>

From the 4.5mls of sample left from the 500mM elution from the heparin column, this was concentrated down to 1ml. This 1ml was run through the gel filtration column, with the chromatogram (Figure 14.23) resembling the one seen previously (Figure 14.20). This time the main peak (fraction 29- red arrow) was re-run though the column to determine the molecular weight and monodispersity of the sample, this peak however did not come out as HPSE-1. Fractions 25-26 (Black arrow) were also run through the column again (Figure 14.24) and this peak had a molecular weight of 67kDa, but the sample was determined to be
polydispersed. From Figure 14.23, fractions 20-35 except for fractions 25-26 (these were rerun in the gel filtration column- Figure 14.24) were run on an SDS-PAGE gel (Figure 14.25). Fractions 27-35 contained the HPSE-1 band with fractions 30-35 being the cleanest. These 6 fractions were pooled together and from a Bradford assay, contained 1mg/3mls of protein. The protein was concentrated to remove the low molecular weight non-specific proteins (Figure 14.25). Due to the gel filtration runs being performed at room temperature, the HPSE-1 was thought to have been degraded between runs and hence why it came out as polydispersed.

**Figure 14.23:** Gel filtration chromatogram. Black arrow represents fractions 25-26, whereas the red arrow represents fraction 29.

**Figure 14.24:** Gel filtration chromatogram of fractions 25-26. The main peak is again fractions 25-26; however this peak is polydispersed with a molecular weight of 67kDa.
Figure 14.25 SDS-PAGE gel of the gel filtration fractions from Figure 14.23. Fractions 27-35 have the HPSE-1 band with fractions 30-35 being the cleanest. These fractions were pooled together and concentrated using a 30kDa membrane to remove the low molecular weight bands.

14.4 Discussion
This chapter focused on the cloning, expression and purification of human HPSE-1 in insect cells in the hope of producing purified protein for crystallisation trials.

14.4.1 Insect cell expression system
Previously cloned pro and active HPSE-1 template DNA produced by Dr McKenzie (McKenzie et al., 2003) were used to try and express the protein. The ‘active’ construct taken to Grenoble was sequence correct but had mutations in the promoter region of the vector which resulted in the lack of expression and when trying to express the pro-form, no expression was seen (Figure 14.3). The reason for no expression of the pro-form is that it had been kept for 7 years at -20°C and the thawing and re-freezing process could cause nicks to form in the DNA, which could potentially lead to mutations which inhibit expression of the protein.
As a result, a new vector was used (pAcGP67) to clone the pro-native and the pro-TEV forms of HPSE-1. Based on sequencing results, the placenta pro-native form of HPSE-1 was not 100% sequence correct so only the pro TEV form was carried forward.

The pro-TEV form of HPSE-1 contains TEV cleavable sites either side of the 6kDa linker peptide (Figure 14.2). Removal of these TEV sites using TEV protease should result in the production of the active 58kDa HPSE-1 enzyme (Nardella et al., 2004). Once the TEV form had been cloned and the sequence was 100% correct (Figures 14.8–14.11), virus was produced and amplified using SF9 cells. Expression was tested a number of times, and western blot analysis confirmed that HPSE-1 was expressed.

HPSE-1 contains two main heparin binding sites (Residues 158-162 and 270-278) (Levy-adam et al., 2005) and these are used to purify HPSE-1 using a heparin binding column. Usually native pro-HPSE-1 has a high affinity for heparin with a salt concentration of over 600mM needed to elute HPSE-1 (McKenzie et al., 2003). However, in this study, HPSE-1 was eluting from the column at much lower salt concentrations (200-400mM NaCl). Levy-Adam et al (2005) noted that residues 158-162 in HPSE-1 had higher affinity binding and deletion of these residues from the pro-form resulted in loss of enzymatic activity. Whereas, deletion of residues 270-278 showed no inhibition of HPSE-1 activity, therefore these residues have lower affinity for heparin compared to residues 158-162. Thus, the TEV cleavage site inserted between residue 157 and 158 must cause a change in residues 158-162, losing this high affinity heparin binding domain once it is folded. Hence, the TEV-pro form could just be binding to the heparin column using residues 270-278, which is why this study sees much lower affinity binding than usual. However, Nardella et al (2004) who reported the first expression of the TEV form of HPSE-1, purified it using 600mM NaCl, thus this study sees a difference in the purification step.

Due to the heparin affinity column not removing all of the non-specific proteins (Figure 14.19a-c), gel filtration was also used as a second purification step (Figure 14.20). The gel filtration chromatogram shows the non-specific proteins eluting in the first 4 peaks, with peak 5 being HPSE-1. Running these peaks on a gel and western blot (Figures 14.21-14.22) did however; show some other faint bands in the HPSE-1 peak, meaning it is not 100% pure. The gel filtration column used could only handle samples of 1ml or less due to the size of the column. Thus, a larger gel filtration column is needed to increase the separation of HPSE-1.
from the non-specific proteins, as even after this second purification step, low molecular weight proteins were still present in the sample (Figure 14.22 and 14.25), making crystallisation more difficult. Also, a second band at around 50kDa is present (Figure 14.22), which is likely to be a degradation product, which again could cause problems during crystallisation as more than one form of the protein is present. Despite these problems, 6 fractions containing the HPSE-1 protein were pooled together and concentrated to 5mg/ml and sent for crystallisation trials. While crystallisation trials were being set up, the remaining protein was loaded onto the gel filtration column, but this time fractions 25-26 were re-run to determine the molecular weight and monodispersity of the sample (Figure 14.23-14.24). Unfortunately, the sample came out as polydispersed, but the molecular weight was correct at 67kDa. Being polydispersed will make crystallisation difficult as the sample could contain multiple forms of the protein, which will disrupt the packing of HPSE-1 causing the crystals to be disordered. However, running these samples on a gel resulted in no HPSE-1 being present in the fractions, thus it is postulated, that the protein had degraded due to being left out at room temperature between the two gel filtration runs, hence the reason for the polydispersity of the sample. From the first gel filtration column (Figure 14.23), the six cleanest fractions (Figure 14.25) were pooled together with a concentration of 1mg/3mls. This second batch of HPSE-1 protein was concentrated to approximately 8mg/ml and was sent to Prof Chayen for more crystallisation trials, with this second batch being cleaner than the first.

As of yet, no crystals, crystal hits or even precipitation have been seen in these crystallisation trials. The purification steps need further optimisation to increase the purity of the sample as well as larger amounts of the protein needs to be produced for further crystallisation trials. This work was interrupted due to lack of funding and the lack of lab availability during my PhD so the work required producing purer protein could not be undertaken and the future work needed is outlined in section 14.4.2 below.
14.4.2 Future work

Future work would include: (1) Optimising the purification steps to obtain higher purity protein as the protein produced thus far still contains impurities. The key step here would be to use a larger gel filtration column to increase the separation of HPSE-1 from the low molecular weight proteins, in order to produce a purer sample for crystallisation trials. (2) Use TEV protease to cleave the linker peptide in the hope of obtaining active protein. (3) Use an enzymatic assay to determine the activity of this manually produced active protein. (4) Use deglycosylation agents to partially deglycosylate the protein to test how soluble it is, and if removing a portion of the sugars would lead to increased chance of ordered crystals. (5) Mutagenesis studies, knocking out the N-glycosylation sites sequentially to determine which are needed for secretion of the protein. If protein can be obtained with fewer N-linked sugars: this could again help in producing ordered crystals. (6) More crystallisation trials needed on the purer protein and/or active protein to obtain crystals. (7) Co-crystallisation with heavy metal ligands/compounds (i.e. cisplatin/carboplatin) to try and form ordered crystals which would help in solving the phase problem. (8) Expression of HPSE-1 in mammalian cells has been started and purification of this protein is needed along with TEV cleavage to achieve pure protein to send for crystallisation trials.

Preliminary expression of HPSE-1 has been undertaken using the mammalian cell expression system at the Oxford Protein Production Facility (OPPF) located at the research complex at Harwell. So far, the pro-TEV form of HPSE-1 has been expressed and a 1litre scale up has been undertaken. A nickel column has been used for affinity binding as the protein is His-tagged and gel filtration used to purify the protein. The protein will be cleaved using TEV protease, the activity measured and a heparin affinity column used to clean up the sample as the sample still contains low molecular weight contaminants.
14.5 References


Chapter 15: Molecular Modelling and Virtual screening studies of human heparanase-1 for the identification of Novel small molecule inhibitors

Simon W.M. Tanley\textsuperscript{a} and John R. Helliwell\textsuperscript{a}

\textsuperscript{a} School of Chemistry, Faculty of Engineering and Physical Sciences, University of Manchester, Brunswick Street, Manchester, M13 9PL, England

Unpublished work

SWMT came up with the ideas in this chapter and carried out all of the work. JRH helped edit the chapter and gave advice on the Lipinski rules of drug design.
15.1 Introduction

HPSE-1 is an endo-β-glycosidase enzyme from the GH 79 family which cleaves HS chains in the ECM and BM of cells (McKenzie et al., 2003; Simizu et al., 2004). The ECM and BM supports the cells with the HS chains providing docking sites for many different growth factors and cytokines (Bernfield et al., 1999; Hulett et al., 2000; Kreuger et al., 2006; Vreys & David, 2007; Fux et al., 2009). Thus cleavage of these HS chains by HPSE-1 can lead to over-activation of growth factors and cytokines and lead to many pathologies, e.g., metastatic cancer and angiogenesis to name only a few (Parish et al., 2001; Zetser et al., 2003). HPSE-1 is a 543 amino acid protein, which is processed from its pre-pro form to a latent 65kDa pro-form via cleavage of the N-terminal signal peptide region (Met-1-Ala-35) in the endoplasmic reticulum (Parish et al., 2001; McKenzie et al., 2003). This pro-form is further processed to form the mature active 58kDa heterodimer consisting of the N-terminal 8kDa subunit (Gln-36-Glu-109) non-covalently attached to the C-terminal 50kDa subunit (Lys-158-Ile-543), via cleavage of the 6kDa linker peptide between Ser-110 and Gln-157 brought about by Cathepsin L (Levy-Adam et al., 2003; McKenzie et al., 2003; Abboud-Jarrous et al., 2005; McKenzie, 2007; Vreys & David, 2007). See chapter 3 for more details.

A number of inhibitors have been designed against human HPSE-1, including, small molecule inhibitors and HS mimetics (see chapter 3.5) (Courtney et al., 2004, 2005; Pan et al., 2006; Xu et al., 2006; Kudchadkar et al., 2008; Dredge et al., 2011; Ritchie et al., 2011; Zhou et al., 2011). However, none have been implemented in the clinic as of yet. The HS mimetics have shown greater promise, with PI-88 (Kudchadkar et al., 2008), PG545 (Dredge et al., 2011), m402 (Zhou et al., 2011) and sst0001 (Ritchie et al., 2011) all entering clinical trials as HPSE-1 inhibitors.

Xylanase (Schmidt et al., 1998), has a sequence identity of 20% to the catalytic 50kDa subunit of HPSE-1, but an overall similarity of 57%, which includes conservation of the TIM (α/β)8 barrel fold and the two catalytic Glu residues in the active site (Zhou et al., 2006; Gandhi & Mancera, 2011). Two groups have produced HM’s of HPSE-1 based on the similarity between HPSE-1 and xylanase, with one group using the active 58kDa form of the enzyme (Sapay et al., 2012) and the second group using the 50kDa subunit (see chapter 3.2) (Zhou et al., 2006). The coordinate files of these HM’s are not freely available, as there is no current mechanism for deposition of theoretical models in the PDB (Berman et al., 2006).
This study focuses on generating a new HM of HPSE-1, based on the recently published crystal structure of a GH 79 family member, a β-Glucuronidase from *acidobacterium capsulatum* (Michikawa et al., 2012) (PDB id 3VNY). This is the first 3D crystal structure of a GH 79 family member solved, and thus could give a greater insight into the active site of human HPSE-1 compared to that of xylanase. However, it has only 20% sequence identity to the active 58kDa form of HPSE-1, but a sequence identity of 25% to the catalytic 50kDa subunit. The TIM (α/β)_8 barrel domain and the two catalytic Glu residues in the active site are, however, conserved (Michikawa et al., 2012).

This new HM of HPSE-1 is used for VS studies, using the NCI-diversity 2 fragment set from the Zinc database (http://zinc.docking.org/). The computer program, AutoDock (Morris et al., 1998) is designed for drug discovery purposes and has been used in this study to simulate the interaction between the diverse fragment set and the HPSE-1 HM. Since none of the previous small molecule ligands have yet entered clinical trials, this new study looks to identify novel small molecule inhibitors of HPSE-1, which could then be developed further to hopefully lead to new anti-cancer treatments. Six new suggested inhibitors of HPSE-1 are put forward and each satisfies the Lipinski rules of drug likeliness.

### 15.2 Methods

#### 15.2.1 Homology model of HPSE-1

The amino acid sequence of the active 58kDa form of HPSE-1 (Figure 15.1) was input into the PHYRE-2 webserver (Kelley & Sternberg, 2009). PHYRE-2 performs a PSI-BLAST (Altschul et al., 1997) and aligns the top entries based on amino acid sequence and 3D structural information from the SCOP database (Murzin et al., 1995) and the PDB (Berman et al., 2000) using HMMs to detect evolutionary relationships to homologous structures (see chapter 2 for more details). Based on these searches, 3VNY (Michikawa et al., 2012), was used as the sequence homolog. 455 out of the 460 residues were modelled and this model was evaluated for its quality using the output from PHYRE2 (Kelley & Sternberg, 2009) and VERIFY 3D (Bowie et al., 1990; Luthy et al., 1992) (http://nihserver.mbi.ucla.edu/Verify_3D/). Consistency with the Ramachandran plot was analysed and the quality of the model was validated by ProSA, a webserver for protein
structure analysis (Sippl, 1993; Wiederstein & Sippl, 2007) (https://prosa.services.came.sbg.ac.at/prosa.php).

36-
QDVVLDFFQTQEPLHLVSPSFLSVTIDANLATDPFLULLGSPPKLRTLARGLSPAYLRGGTKTDVLIDFPKKE-
109

158-
KKFKNSTYRSSVVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLDYCSSKGYNISWELGNEPNLSFLKK
ADIFINGSQLGEDFQLHKLRLKSTFKNAKLYGPVDGQPRRKTAKMLKSFLKAGGEVIDSVTWHHYLYNGRTATK
EDFLNPDVLDIFISSVQKVFQVPERPGKKVWVLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVV
MRQVFFGAGNYHLVDENFDLPLDYWLSLFLKKLVGTVKLMASVQGSKRRKLRYVLHCTNDPRLYKEGDLTLY
AINLHNVTYLRPLPYFNSKQDVKYLLRPLGPHGLLSKSQVLNGLTLKMVDQTLPLMEKPLRPGSSLGLPAFS
YSFFVIRNAKVAACI -543

**Figure 15.1:** Amino acid sequence of the active form of HPSE-1. The 8kDa subunit is underlined and in bold (Q36-E109), with the 50kDa subunit not underlined (K158-I543).

### 15.2.2 Virtual screening

For VS, the NCI-Diversity 2 fragment set (a small library of around 2000 compounds representing the broader chemical space of the larger 140,000 compounds) was downloaded from the Zinc database (http://zinc.docking.org/pdbqt/). Raccoon (http://autodock.scripps.edu/resources/raccoon) was used to verify that each ligand passed the Lipinski rules of drug likeliness (Lipinski et al., 1996; Lipinski, 2004), with 1541 ligands being used in the VS procedure. Raccoon was used to produce both the grid and docking files for each ligand. In AutoDock (Morris et al., 2010), the active 58kDa HPSE-1 HM was prepared by adding polar hydrogen atoms and partial charges to the protein. The grid file used to search the conformational space of the protein was chosen to focus just on the TIM(α/β)₈ barrel of the protein, where the active site is predicted to be located. The docking parameters were: - 10 genetic algorithm runs, 2000 evaluations and 200 generations used. Once all files were generated, docking of all 1541 ligands was completed using a batch file on a windows HP laptop, with the overall calculation time taking ~3 days for full completion. Analysis of the results was performed manually with all ligands with an estimated inhibition constant (Ki) greater than 300μM being rejected to reduce the number of ligands which needed to be further analysed.
15.2.3 Design of second generation inhibitors

From the VS study, the top hits which had an estimated $K_i$ of $<300\mu M$ as well as being novel inhibitors where extended to cover more parts of the active site in order to increase their binding affinities. The binding of each ligand was manually inspected using PyMOL (The PyMOL Molecular Graphics System, Version 0.99 Schrödinger, LLC.) and extended. These extended ligands were checked using the molinspiration webservice (http://www.molinspiration.com/cgi-bin/properties) to determine whether they pass the Lipinski Rules criterion and if passed, each new ligand was made using Prodrg (Schüttelkopf & van Aalten, 2004) and then docked into the active site of the HPSE-1 HM using AutoDock (Morris et al., 2010).

15.3 Results

15.3.1 Homology modelling and evaluation of the model

Figure 15.2 shows the overall HM of the active form of HPSE-1 created by the PHYRE-2 webserver, and has 455 residues modelled out of 460.

![Figure 15.2: Model of the active form of HPSE-1 from PHYRE-2. (a) Shows the overall fold of the protein, including the TIM $(\alpha/\beta)_8$-barrel. (b) The surface view of the protein, with the active site residues in blue (the two catalytic Glu residues (E142 and E260) along with His-213, Asn-141, Tyr-215 and Tyr 265 and the hydrophobic pocket which is circled (Gln-300-Asn-307)). HS binding site 1 is highlighted in yellow (K75-D88) and HS binding site 2 highlighted in black (Q187-K197).](image)
To confirm how valid this HM is, the output from PHYRE-2 includes the confidence of disorder and secondary structure prediction (Figure 15.3). This rates each amino acid into how confident the program is, that these residues are part of an α-helix, β-sheet, or loop and turn region, with the majority of the disorder coming from the C-terminus region of the model as well as some loop and turn regions. The VERIFY 3D (http://nihserver.mbi.ucla.edu/Verify_3D/) webserver was also used to validate the model. These results (Figure 15.4) show that the scores for nearly all amino acid residues are within reasonable regions, showing that the model is of good quality, with only the C-terminus region being poorly modelled which corresponds well with the PHYRE-2 results (Figure 15.3). Figure 15.5 shows the Ramachandran plot with the majority of the amino acid residues falling in the most favoured regions and allowed regions (96.5%, 439/455), with 3.5% (16/455) of the residues falling in the disallowed regions. 5 of these disallowed residues are prolines. There are slightly more outliers than would be liked, but based on the VERIFY 3D results, ~7 of these residues fall into the C-terminus region which is not modelled as well as the rest of the protein. However, having 96.5% of the residues in allowed regions and a good VERIFY 3D result, this HM was used for VS. The ProSA webserver checks the overall quality of the protein structure (https://prosa.services.came.sbg.ac.at/Prosa.php) and gives a logz score of -6.71 which falls into the range observed for experimental protein structures of this amino acid sequence length (Figure 15.6).
Figure 15.3: Secondary structure prediction and disorder confidence of the active HPSE-1 HM from the PHYRE2 webserver.

Figure 15.4: The Verify 3D results for the 58kDa active HPSE-1 HM.
Figure 15.5: Ramachandran plot of the HPSE-1 HM. Residues in areas I and II are the favoured and allowed regions. Residues in area III and residues in red are the disallowed residues. The triangles represent Gly residues.

Figure 15.6: ProSA result. The logz score is -6.71 (circled), which falls into the range of experimentally derived structures of the same protein polypeptide length.

Another test of the quality of this HM was to dock the known small molecule inhibitors of HPSE-1 to confirm whether they bind to the active site. This then gives confidence that this HM is of good enough quality to identify potential novel inhibitors of HPSE-1. Four of the known small molecule inhibitors were chosen: the 2,3-dihydro-1,3-dioxo-1H-isoindole-5-carboxylic acids (Courtney et al., 2004), the furanyl acetic acid compounds (Courtney et al., 2005), the benzoaxazol-5-yl acetic acid compounds (Courtney et al., 2005) and the symmetrical 1-[4-(1H-Benzimidazol-2-yl)-phenyl]-3-[4-(1H-benzoimidazol-2-yl)-phenyl] urea compounds (Pan et al., 2006) (Figure 15.7). All four compounds bound in the active site of the HPSE-1 HM near to the catalytic Glu residues whilst also covering the hydrophobic
pocket (Figure 15.8). The best estimated inhibition constant (Ki) for the four compounds was; 53μM, 470nM, 1.7μM and 11.5μM respectively.

Figure 15.7 Chemical structures of the four known small molecule inhibitors of HPSE-1. (a) 2,3-dihydro-1,3-dioxo-1H-isoindole-5-carboxylic acid, (b) the furanyl acetic acid compound (Courtney et al., 2005), (c) the benzoxazol-5-yl acetic acid compound and (d) the symmetrical 1-[4-(1H-Benzimidazol-2-yl)-phenyl]-3-[4-(1H-benzoimidazol-2-yl)-phenyl] urea compound.
Figure 15.8 Binding of the four known small molecule inhibitors of HPSE-1 into the active site of the HM of the active 58kDa form of HPSE-1. (a) 2,3-dihydro-1,3-dioxo-1H-isooindole-5-carboxylic acid, (b) the furanyl acetic acid compound (c) the benzoxazol-5-yl acetic acid compound and (d) the symmetrical 1-[4-(1H-Benzimidazol-2-yl)-phenyl]-3-[4-(1H-benzoimidazol-2-yl)-phenyl] urea compound. The active site residues are highlighted in blue.
15.3.2 Virtual screening using the NCI diversity 2 fragment set and the HPSE-1 model

The NCI-diversity 2 fragment set (http://zinc.docking.org/pdbqt/) was downloaded from the Zinc database. 1541 of the ligands passed the Lipinski-like rules of drug likeliness (Number of hydrogen bond donors (<5), number of hydrogen bond acceptors (<10), molecular weight (<500Da) and number of rotatable bonds (<10). It however, doesn’t use the calculated LogP values (-1 - +5) to filter the ligands). All 1541 ligands were docked into the HPSE-1 HM using just the TIM (α/β)₈-barrel as the search space. Out of these 1541 ligands, 105 had an estimated Ki of ≤ 300μM, with 22 having ≤ 100μM, 12 ≤ 50μM and 4 with ≤ 25μM. Out of the 105 ligands which bound with an estimated Ki of ≤300μM, 6 potentially new scaffolds were found. These 6 potentially novel ligands passed all the Lipinski rules of drug likeliness when calculated using the molinspiration webserver (http://www.molinspiration.com/cgi-bin/properties) (Table 15.1). The other ligands which bound with estimated Ki of less than 300μM, were either similar to the known small molecule inhibitors or didn’t pass the calculated logP value, thus these ligands were rejected. Again, due to seeing similar fragments to the known small molecule inhibitors of HPSE-1 binding in the active site, shows this model is of good enough quality to pick these fragments out during the VS study. The 6 potential novel ligands from the Zinc database are, Zinc13597348, Zinc 05386901, Zinc 01573829, Zinc 13143019, Zinc06576323 and Zinc00350319 (Figure 15.9). Five of the inhibitors are fused ring systems (Figure 15.9) and all bind in the active site, in close proximity to the two catalytic Glu residues (Figure 15.10). Zinc13597348 is similar to the aglycone moiety of the PG545 series of HS mimetics, i.e. the sugar part is missing.
Table 15.1:

Drug likeliness of the novel inhibitors from the Zinc database, with the estimated inhibition constants for all 6 along with the compound name it is most similar to.

<table>
<thead>
<tr>
<th></th>
<th>13597348</th>
<th>05386901</th>
<th>01573829</th>
<th>13143019</th>
<th>06576323</th>
<th>00350319</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight</td>
<td>389</td>
<td>317</td>
<td>359</td>
<td>277</td>
<td>340</td>
<td>240</td>
</tr>
<tr>
<td>Calculated LogP</td>
<td>4.7</td>
<td>4.6</td>
<td>3.2</td>
<td>3.2</td>
<td>4.0</td>
<td>2.4</td>
</tr>
<tr>
<td>H-bond donors</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>H-bond acceptors</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rotatable bonds</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Estimated Ki values</td>
<td>40μM</td>
<td>43μM</td>
<td>96μM</td>
<td>153μM</td>
<td>174μM</td>
<td>226μM</td>
</tr>
<tr>
<td>Compound name</td>
<td>Androstene/Indeno-pyrazole</td>
<td>Phenanthio[9,10-e]-1,2,4-Triazine,3-[(Phenylmethyl)thio]pyridine</td>
<td>Napthopyran-4-one,2,3-dihydro-2-(3-pyridinyl)</td>
<td>Triazino-indole</td>
<td>1,2,4,5-tetrazine, heahydro-3,6-diphenyl</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 15.9:** The chemical structures of the 6 potential novel HPSE-1 scaffolds. (a) Zinc13597348, (b) Zinc05386901, (c) Zinc01573829, (d) Zinc13143019, (e) Zinc06576323 and (f) Zinc00350319.

**Figure 15.10:** Inhibitor binding into the active site of the 58kDa HM of HPSE-1. (a) Zinc13597348, (b) Zinc05386901, (c) Zinc01573829, (d) Zinc13143019, (e) Zinc06576323 and (f) Zinc00350319. Each figure shows the surface view of the protein with the active site highlighted in dark blue. The active site comprises, the catalytic Glu residues (Glu-142 and Glu-260), Asn-141, Tyr-215, His-213, Tyr-265 and the hydrophobic pocket (Gln-300-Asn-307) which is circled. The residues highlighted in black are part of the HS binding domain 2 (Gln-187-Lys-197).
15.3.3 Design of second generation inhibitors based on the virtual screening results

The six potential novel inhibitors of HPSE-1 (Figure 15.9) all make good starting points to develop second generation inhibitors as they are low molecular weight compounds (Table 15.1) and based on the docking results (Figure 15.10) there is room in the active site for these compounds to be extended, thus increasing their binding affinity. From extending these starting ligands, the best estimated Ki achieved was 150nM and 540nM for the extended Zinc13597348 template (Figure 15.11). These two ligands as well as all the other second generation inhibitors designed, bound to the active site in close proximity to the two catalytic Glu residues. The chemical structures of each new ligand along with its estimated Ki values are given in Tables 15.2-15.7.

Figure 15.11 Second generation inhibitors of the Zinc 13597348 template with estimated Ki in the nano molar range (a) 540nM and (b) 150nM
**Table 15.2:** Zinc13597348. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7a) with an estimated Ki value of 40μM.

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>Estimated Ki values</th>
<th>Molecular weight</th>
<th>Calculated LogP</th>
<th>H-bond donors/acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40 μM</td>
<td>389</td>
<td>4.7</td>
<td>1/3</td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 μM</td>
<td>407</td>
<td>4.7</td>
<td>1/3</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 μM</td>
<td>450</td>
<td>2.9</td>
<td>3/5</td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.8 μM</td>
<td>478</td>
<td>3.4</td>
<td>3/5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compound</td>
<td>IC50 (μM)</td>
<td>EC50 (nM)</td>
<td>pIC50</td>
<td>Ref.</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>(d)</td>
<td>0.54</td>
<td>478</td>
<td>3.2</td>
<td>3/6</td>
</tr>
<tr>
<td>(e)</td>
<td>6.9</td>
<td>464</td>
<td>3.8</td>
<td>3/6</td>
</tr>
<tr>
<td>(f)</td>
<td>0.15</td>
<td>490</td>
<td>2.6</td>
<td>4/7</td>
</tr>
</tbody>
</table>
Table 15.3: Zinc05180958. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7b) with an estimated Ki value of 43μM.

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>Estimated Ki values</th>
<th>Molecular weight</th>
<th>Calculated LogP</th>
<th>H-bond donors/acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Chemical structure" /></td>
<td>43μM</td>
<td>317</td>
<td>4.6</td>
<td>1/2</td>
</tr>
<tr>
<td><img src="image2" alt="Chemical structure" /></td>
<td>34μM</td>
<td>463</td>
<td>2.5</td>
<td>4/8</td>
</tr>
<tr>
<td><img src="image3" alt="Chemical structure" /></td>
<td>8.8μM</td>
<td>475</td>
<td>3.2</td>
<td>3/8</td>
</tr>
<tr>
<td>(c)</td>
<td>Conjugate</td>
<td>$11 \mu M$</td>
<td>491</td>
<td>2.8</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>(d)</td>
<td>Conjugate</td>
<td>$17.4 \mu M$</td>
<td>402</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e)</td>
<td>Conjugate</td>
<td>$2.6 \mu M$</td>
<td>445</td>
<td>1.0</td>
</tr>
<tr>
<td>(f)</td>
<td>Conjugate</td>
<td>$9.8 \mu M$</td>
<td>471</td>
<td>1.3</td>
</tr>
</tbody>
</table>
(g) 2.3\textmu M 437 0.7 4/8

(h) 1.35\textmu M 471 3.0 4/8
Table 15.4: Zinc01573829. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7c) with an estimated Ki value of 96μM.

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>Estimated Ki values</th>
<th>Molecular weight</th>
<th>Calculated LogP</th>
<th>H-bond donors/acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>96μM</td>
<td>359</td>
<td>3.2</td>
<td>0/3</td>
</tr>
<tr>
<td>(b)</td>
<td>20μM</td>
<td>403</td>
<td>3.1</td>
<td>1/5</td>
</tr>
<tr>
<td></td>
<td>8μM</td>
<td>475</td>
<td>2.8</td>
<td>2/7</td>
</tr>
<tr>
<td>2.9μM</td>
<td>471</td>
<td>3.3</td>
<td>3/7</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td><img src="c" alt="Chemical Structure" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2.2μM</th>
<th>486</th>
<th>2.4</th>
<th>3/8</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="d" alt="Chemical Structure" /></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 15.5**: Zinc13143019. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7d) with an estimated Ki value of 153μM.

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>Estimated Ki values</th>
<th>Molecular weight</th>
<th>Calculated LogP</th>
<th>H-bond donors/acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Chemical structure" /></td>
<td>153μM 2.6 μM 52 μM 28 μM</td>
<td>277 363 346 394</td>
<td>3.2 1.7 3.6 3.8</td>
<td>1/3 3/6 3/6 3/6</td>
</tr>
</tbody>
</table>

- **(a)**
- **(b)**
- **(c)**
**Table 15.6:** Zinc06576323. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7a) with an estimated Ki value of 174μM.

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>Estimated Ki values</th>
<th>Molecular weight</th>
<th>Calculated LogP</th>
<th>H-bond donors/acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>174μM</td>
<td>340</td>
<td>4</td>
<td>0/4</td>
</tr>
<tr>
<td>(a)</td>
<td>23.2 μM</td>
<td>398</td>
<td>3.8</td>
<td>1/6</td>
</tr>
<tr>
<td>(b)</td>
<td>6.7 μM</td>
<td>412</td>
<td>3.1</td>
<td>1/6</td>
</tr>
<tr>
<td>(c)</td>
<td>11.7 μM</td>
<td>440</td>
<td>5.0</td>
<td>1/6</td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td>IC₅₀ (μM)</td>
<td>IC₅₀ (nM)</td>
<td>IC₅₀ (nM)</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>(d)</td>
<td><img src="image" alt="Structure (d)" /></td>
<td>15.2</td>
<td>445</td>
<td>5.6</td>
</tr>
<tr>
<td>(e)</td>
<td><img src="image" alt="Structure (e)" /></td>
<td>13.8</td>
<td>489</td>
<td>5.5</td>
</tr>
<tr>
<td>(f)</td>
<td><img src="image" alt="Structure (f)" /></td>
<td>10.3</td>
<td>441</td>
<td>3.0</td>
</tr>
<tr>
<td>(g)</td>
<td><img src="image" alt="Structure (g)" /></td>
<td>30</td>
<td>442</td>
<td>3.0</td>
</tr>
<tr>
<td>(h)</td>
<td><img src="image" alt="Structure (h)" /></td>
<td>15.8</td>
<td>456</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Table 15.7: Zinc00350319. The chemical structure of each new ligand, its molecular weight, calculated LogP values, number of H-bond donors/acceptors and its estimated Ki values are given and compared to the parent compound (Figure 7a) with an estimated Ki value of 226μM.
### 15.4 Discussion

HPSE-1 is over-expressed in the majority of cancers and the level of expression is related to the metastatic potential of the tumour. Hence, designing inhibitors of HPSE-1 are important to develop new anti-cancer treatments. The experimental 3D structure of HPSE-1 is yet to be elucidated, thus, inhibitors have been developed based on high throughput assays as well as using HM’s. Only the HS mimetics have shown promise as HPSE-1 inhibitors, with a
number of these entering clinical trials (PI-88, PG545, M402, SST0001) (Kudchadkar et al., 2008; Dredge et al., 2011; Ritchie et al., 2011; Zhou et al., 2011), whereas the small molecule inhibitors have had less success (Courtney et al., 2004, 2005; Pan et al., 2006; Xu et al., 2006). This study has produced a new HM of HPSE-1 based on the recently solved crystal structure of a β-glucuronidase from the GH 79 family (Michikawa et al., 2012). Based on this new HM, VS has been used to identify novel theoretical HPSE-1 inhibitors.

### 15.4.1 HPSE-1 Homology model

The HM presented here was produced using PHYRE-2 (Kelley & Sternberg, 2009), which aligns the input amino acid sequence with homologs using HMMs. The homolog used to create the HM was 3VNY; with 20% sequence identity to the active 58kDa form of HPSE-1 and 25% sequence identity to the catalytic 50kDa subunit. The new HPSE-1 HM presented here does have its disadvantages: - (i) 3.5% of the amino acids fall outside the favoured regions in the Ramachandran plot (Figure 15.5) due to the C-terminus of the protein being disordered. (ii) The main disadvantage with this model and the previous ones published by Zhou et al. (2006) and Sapay et al. (2011), is the absence of disulphide bridges. Human HPSE-1 has 5 Cys residues in the 65kDa pro form, with 4 of these thought to form 2 disulphide bridges, whereas the 58kDa active form of the enzyme has only 1 disulphide bridge thought to form through Cys-542 and Cys-432 (Simizu et al., 2007). Hence the new chemical inhibitors designed against this HM of HPSE-1 may contain errors due to the lack of this disulphide bridge. As can be seen in Figures 15.4 and 15.5, the C-terminus region of this model is the most disordered, which corresponds to the likely formation of this disulphide bridge. Comparing the active 58kDa HM of HPSE-1 produced in this study with the model described by Sapay et al. (2011), shows a lot of similarities. The main one being, the HS binding domains are in similar positions, with the HS binding domain 2 (Q270-K280, (Q187-K187 in this model)) being close to the active site, forming a large basic cluster at the exit of the active site groove. Whereas the HS binding site 1 (K158-D171, (K75-D88 in this model)) is not directly involved in the active site groove shaping and is located away from the modelled active site. Thus, Sapay et al. (2011) proposed that HS binding site 1 could be a sub-site involved in substrate recognition, while HS binding site 2 would be directly involved in the interaction with sugar units located at the cleavage site. This new HPSE-1 HM agrees with this proposal, however, these do disagree with the proposed model of Mosulen et al.
(Mosulén et al., 2011; Gozalbes et al., 2013), who see the two HS binding domains being either side of the two catalytic Glu residues in close proximity, with both making contact to the inhibitors. Even though the C-terminus of the protein is disordered and therefore likely not well modelled, overall this model is of good quality, confirmed by 96.5% of all residues being in allowed regions of the Ramachandran plot (Figure 15.5) and the overall quality of the model is reasonable, based on the verify 3D results and the ProSA score (Figures 15.4 and 15.6). The biggest confidence boost that this model is correct is due to the fact that the known small molecule inhibitors bind in the active site of this model (Figure 15.8). Thus, this HM is of good quality for use in a VS study to detect potential novel theoretical HPSE-1 inhibitors.

15.4.2 Virtual screening of the HPSE-1 model using the NCI diversity 2 fragment library

Overall, out of the 1541 compounds tested in this VS study, 6 were predicted to inhibit HPSE-1 with estimated Ki of less than 230μM as well as being potentially novel HPSE-1 inhibitors (i.e., these compounds are freely available but have not previously been shown to be potential inhibitors of HPSE-1). All six ligands bind in the active site of this HM with either one or both of the catalytic Glu residues in close proximity. As can be seen in Figure 15.9, the ligands do not cover all of the active site, meaning there is room for extending each ligand and critically to extend them to fill the hydrophobic pocket (Gln-300- Asn-307, circled in Figure 15.10) as this site is thought to be important for increased binding affinity of the inhibitors (Zhou et al., 2006). Each of the ligands are also good starting points for being extended as they are low molecular weight compounds, and only contribute a few H-bond donors and acceptors. However careful attention to Lipinski Rules will likely mean that this approach would be limited in scope.

Zinc13597348 is 88% similar to the Androst-5-ene-3,17-diol, 16-(1-piperidinylmethyl)-, (3β,16β,17β)- compounds (CAS number 1287246-60-9) which are in use to treat ovary cancer by acting on membrane transport proteins (Serly et al., 2011). This ligand is also similar to the agylcone group of the PG545 series of HS mimetics, without the sulphated oligosaccharides attached.

Zinc05386901 is 81% similar to the Indeno[1,2-c]pyrazole, 3-(3,4-dichlorophenyl)-4,?-dihydro- (9CI) compound (CAS number 41903-02-0).
Zinc01573829 is the Phenanthro[9,10-e]-1,2,4-triazine, 3-[(phenylmethyl)thio]- compound (CAS number 59851-31-9) and has exhibited inhibition of Mycobacterium tuberculosis Thiamin Phosphate Synthase (Khare et al., 2011).

Zinc13143019 is 79% similar to the naptho-pyran-4-one,2-3-dihydro-2-(3-pytradiny1) compounds (CAS number 1125445-67-1). Zinc06576323 is 80% similar to the triazino-indole compounds (CAS number 905216-08-2).

Zinc00350319 is the 1,2,4,5-tetrazine,hexahydro-3,6-diphenyl compound (CAS number 500359-13-17) and is thought to be an inhibitor of the Nuclear export of Topoisomerase II alpha for the treatment of cancer (Patent numbers: WO 2010/068947 A2/ WO2012/075484 A2).

Some of these ligands have already been shown to treat cancer, thus these inhibitors could also inhibit HPSE-1 as a by-product leading to cross reactivity and more than one protein target for these drugs.

Extending these six parent compounds to fill more pockets of the active site has led to new compounds being designed which exhibit estimated Ki values in the nano molar range (150nM and 540nM, Figure 15.10 and Table 15.2), with all second generation inhibitors, having higher estimated Ki compared to their parent compounds (Tables 15.2-15.7). The molecular weight values for these second generation inhibitors are large (>450Da- <500Da), but they all still meet the Lipinski rules of drug-likeliness.

These 6 new scaffolds have shown promise in the techniques of HM and VS. However, due to the second generation inhibitors having high molecular weights, these inhibitors are very limited to further improvement. Also, the estimated Ki’s are not good enough for lead drug optimisation, with values in the 10’s of nM and lower needed. Hence, VS is a good technique when inhibitors are unknown, but due to many small molecule inhibitors of HPSE-1 already known, the ones found in this study even though they are predicted to inhibit HPSE-1; might not lead to new inhibitors being designed. However, testing these inhibitors experimentally will be of interest to clarify the theoretical results presented here.

15.5 Conclusions

This study delivers a new HM of HPSE-1 based on the known structure of a GH 79 family member, which is of overall good quality and has similarities to the previously published HM’s.
A VS study using the NCI diversity 2 set docked into this new HM was undertaken, with 6 potentially novel inhibitors of HPSE-1 being proposed. Second generation inhibitors of these 6 initial compounds have also been designed and the best estimated Ki obtained thus far was 150nM, thus are thought to be relatively high affinity inhibitors. However, due to the molecular weight of these second generation inhibitors being large (450-500Da) as well as not having estimated Ki in the 10’s of nM range, means these inhibitors cannot be improved further. Two of these potential HPSE-1 inhibitors are known to have anti-cancer effects by inhibiting different enzymes. Hence, their inhibitory activity of HPSE-1 could be a secondary effect which helps in the anti-cancer treatments.

15.6 References


Overall conclusions to Part III

From these three chapters we have shown a variety of results.

Using the PHYRE-2 webserver, a HM based on the crystal structure of 3VNY is reported and shows similarities to the HM produced by Sapay et al. The problem with this model is the poor modelling of the C-terminus region. In this region, a disulphide bond between Cys-542 and Cys-432 should form. In this and the other reported HM’s, this disulphide bond is missing, thus the model of the structure is not 100% correct. To confirm how accurate this model is, docking of the known small molecule inhibitors were carried out and these bound in the active site. Subsequently, a VS study was undertaken to identify potential novel small molecule inhibitors of HPSE-1. From this VS study, 6 potentially novel inhibitors were discovered. However, trying to improve the estimated Ki on these ligands resulted in very large molecules with the best estimated Ki of 150nM. To be a good lead drug, the Ki needs to have low nM affinity (<10nM). Thus, due to the size of these inhibitors, none were taken forward. Testing these compounds experimentally to see if they do inhibit HPSE-1 would be the next step in order to confirm the theoretical results. Some of the compounds identified are already known to be anti-cancer agents, determining whether they do inhibit HPSE-1 could then lead to these drugs binding to more than one target and thus being used for different cancers.

The HM produced was used as a MR search model to try and solve the 3D structure of HPSE-1 due to the fact that heavy atom soaks failing in the crystals produced. The crystals showed both order and disorder in the diffraction pattern making these difficult to solve; also the MR search model only had a sequence identity of 25%. The best structure obtained was in space group P1 with R/Rfree values of 32.8/38.8. However, the geometry of this model was poor with many residues lying in the disallowed regions of the Ramachandran plot. Also, the structure was incomplete, with a lot of missing residues and side chains and due to the lack of anomalous difference electron density, the disulphide bond could not be modelled. Thus, using this data to solve the structure of HPSE-1 is very difficult and new crystals and diffraction data is needed before the structure can be solved.

I have over-expressed HPSE-1 using the insect cell expression system. The pro-TEV form was produced, but crystallisation trials have thus far failed. Future work would include: (1) Optimising the purification steps to obtain higher purity protein as the protein produced thus far still contains impurities. (2) Use TEV protease to cleave the linker peptide in the hope of
obtaining active protein. (3) Use an enzymatic assay to determine the activity of this active HPSE-1 protein. (4) Use deglycosylation agents to partially deglycosylate the protein to test how soluble it is, and if removing a portion of the sugars would lead to increased chance of ordered crystals. (5) Mutagenesis studies, knocking out the N-glycosylation sites sequentially to determine which are needed for secretion of the protein. If protein can be obtained with fewer N-linked sugars: this could again help in producing ordered crystals. (6) More crystallisation trials needed on the purer protein and/or active protein to obtain crystals. (7) Co-crystallisation with heavy metal ligands/compounds (i.e. cisplatin/carboplatin) to try and form ordered crystals which would help in solving the phase problem. (8) Expression of HPSE-1 in mammalian cells has been started and purification of this protein is needed along with TEV cleavage to achieve pure protein to send for crystallisation trials.

The crystal structure of this protein is of the utmost importance to design active site-specific drugs which could eventually lead to new anti-cancer treatments.