Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Snakin-1: a Case Study in Racemic Protein Crystallography

Ho Yeung

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Biological Sciences

The University of Auckland

2016
Abstract

X-ray crystallography experiments have well-known bottlenecks that include the ability to crystallise any given protein, and the so-called crystallographic phase problem that hampers the process of structure solution from experimentally measured diffraction intensities. Racemic protein crystallography has the potential to address both of these issues, as racemic mixtures of proteins: a) appear to crystallise more readily than enantiopure solutions; and b) structure determination in the centrosymmetric space groups sampled by racemic mixtures is simplified by restriction of the crystallographic phases to two possible values. The overall aim of this work was to explore the general utility of this technique in de novo protein structure determination.

Snakin-1 is a 63 residue antimicrobial protein originally isolated from potato (*Solanum tuberosum*) which is active against a number of bacterial and fungal phytopathogens such as *Clavibacter michiganensis*, *Pseudomonas syringae* and *Fusarium solani*. It is a member of the GASA (gibberellic acid stimulated in Arabidopsis)/snakin family and the mature protein consists of a GASA domain incorporating six intramolecular disulfide bonds. These proteins are found in a variety of plant species and appear to be involved in a range of functions including cell elongation and cell division. It has also been speculated that the 12 conserved cysteines in these proteins perform a role in redox regulation. The structure of these GASA/snakin proteins is not known and their amino acid sequences do not correspond to any known structural motifs.

In this work, the de novo structure determination of snakin-1 was carried out as a case study in racemic protein crystallography. Total chemical synthesis of the native L-snakin-1 and its enantiomer D-snakin-1 was accomplished using contemporary techniques for protein synthesis. Crystallisation of racemic snakin-1 was readily achieved using macromolecular crystallisation techniques, but crystals of racemic snakin-1 alone were insufficient for structure determination. Iodinated quasi-racemic snakin-1 crystals were subsequently prepared and provided essential phase information for successful structure elucidation. The sensitivity of the C-I bond to X-ray irradiation was instrumental in affording a radiation damage-induced signal used in phasing. This study has shown that despite some difficulties encountered in solving the snakin-1 structure, racemic protein crystallography provided a clear benefit in structure solution; the crystal structure of this antimicrobial protein would
Abstract

not have been solved by conventional methods and lays the way towards the rational design of novel antimicrobial agents.
Preface

All synthetic and crystallographic work described in this thesis was carried out by the author under the supervision of Distinguished Professor Margaret A. Brimble in the School of Chemical Sciences, and Distinguished Professor Edward N. Baker, Dr. Christopher J. Squire and Dr. Paul W. R. Harris in the School of Biological Sciences at the University of Auckland. Biological assays were undertaken by Professor Antonio Molina and Gemma López García at the Centro de Biotecnología y Genómica de Plantas (UPM-INIA), Universidad Politécnica de Madrid.

Parts of this work have been previously published:

Acknowledgements

First and foremost, I would like to thank my supervisor, Distinguished Professor Margaret A. Brimble, for providing me with the opportunity to pursue this stimulating and rewarding research topic in the course of my PhD. The knowledge and experience I gained under your supervision have been invaluable, and I am confident they will continue to serve me well in the future.

I would also like to thank my co-supervisor, Distinguished Professor Ted Baker, for his encouragement and support on the crystallography side of things. Your enthusiasm and knowledge of the field are astounding, and I am very grateful for your advice on my thesis.

Big thanks must go to my second and third co-supervisors, Dr. Christopher J. Squire and Dr. Paul W. R. Harris. Thank you, Chris, for being there when I needed help, and for inspiring me to go further to achieve my goals. Your skills in crystallography and writing have been of great help in the completion of this work, and our discussions on crystallography have been essential in aiding my understanding of the field. Paul, thank you for all that you have taught me in the years I have been in the peptide group. You have been an extremely patient and supportive mentor in the lab, and are a constant inspiration in both your ability and attitude. Without your mentorship this PhD would not have been possible.

Thank you to the Marsden Fund of the Royal Society of New Zealand for providing the funding for this project through a Marsden Grant.

Thank you to the Australian Synchrotron staff, especially Dr. Tom Caradoc-Davies, and Dr. Santosh Panjikar for their assistance in X-ray diffraction data collection. Many thanks also go to Professor Antonio Molina, and Gemma López García at the Centro de Biotecnología y Genómica de Plantas (UPM-INIA), Universidad Politécnica de Madrid, for their assistance in biological assays.

Lia Yosaatmadja, thank you for all your mentorship on the ways of the dark side. Without your help, I probably would not know anything about proteins. Thanks for also teaching me the most important part: to “say ‘no’ to protein”.

Thank you to all the members of the Brimble group whom I have been fortunate enough to know over the many years in this group. Many of you have inspired me throughout my time
here, and I am grateful to you all for such a great environment. Some special mentions must be made: Dr. Renata Kowalczyk, Dr. Sunghyun Yang, Dr. Meder Kamalov, Dr. Harveen Kaur, Dr. Geoff Williams, Dr. Greg Hung, Dr. Stefanie Papst, Dr. Tsz Ying Yuen, Dr. Luis De Leon Rodriguez, Dr. Iman Kavianinia, Dr. Andrew Wadsworth, Dr. Amanda Heapy, Dr. Briar Naysmith, Dr. Paul Hume, Dr. Jonathan Hubert, Dr. Paul Haseler, Dianna Truong, Harry Aitken, Rachelle Quach, Victor Yim, and Lauren Yule. Thank you all for making the Brimble group a great place to be.

Special mentions also to Dr. Janice Choi and Tim Layt. Your work is what keeps the group functioning, and it is very much appreciated.

Thanks also go to my other big science family, the members of the LSB. You all helped to ease my transition from chemistry to biology, and it is a pleasure to have met you all: Dr. Ghader Bashiri, Dr. Paul Young, Dr. Richard Kingston, Dr. Jeremy Keown, Dr. James Jung, Dr. Thomas Lagautrière, Amy Yewdall, Melissa Webby, Ehab Jirgis, Eugene Sun, Matthew Sullivan, and Joy Yang. Special thanks goes to Dr. David Goldstone for always helping out and for providing useful discussions. Go PTDA!

Dr. Karima Medini, Dr. Aimee Son, Zaid Amso and Tom Wright. You guys are more like a second family than a part of the science family. These past few years have been unforgettable, and I am glad for all the fun times we shared.

Finally, I would like to thank my family, for putting up with a PhD student for four years, and for supporting me throughout. I love you all.

Ho Yeung

January 2016
Table of Contents

Abstract ............................................................................................................................. i
Preface ............................................................................................................................... iii
Acknowledgements ........................................................................................................ iv
Table of Contents ........................................................................................................... vi
Abbreviations ................................................................................................................... x

Chapter 1: Racemic Protein Crystallography: .............................................................. 1
1.1 X-ray crystallography and the crystallographic phase problem ....................... 3
1.1.1 Heavy-atom phasing methods ............................................................................ 5
1.1.1.1 Isomorphous replacement ............................................................................. 6
1.1.1.2 Anomalous dispersion ................................................................................... 8
1.1.1.3 Locating heavy-atom sites via Patterson methods ....................................... 12
1.1.1.4 Locating heavy-atom sites via direct methods ............................................. 14
1.1.2 Molecular replacement structure solution .................................................... 15
1.1.3 Racemic protein crystallography ................................................................. 18
1.2 Chemical synthesis of peptides and proteins ..................................................... 21
1.2.1 Solid phase peptide synthesis ....................................................................... 21
1.2.2 Native chemical ligation .............................................................................. 25
1.2.3 NCL at sites without cysteine ...................................................................... 28
1.2.3.1 Reductive desulfurisation of thiol-derived amino acids .............................. 30
1.2.3.2 Alkylation of cysteine ............................................................................... 33
1.2.4 Preparation of peptide α-thioesters ............................................................... 35
1.2.4.1 Alternative Fmoc deprotection reagents ...................................................... 36
1.2.4.2 Late-stage thioester installation ................................................................... 37
1.2.5 Coupling reagents used in SPPS .................................................................. 42
1.3 Protein (re)folding ........................................................................................... 47
1.3.1 Circular dichroism spectroscopy .................................................................. 48
1.4 Scope of this thesis ........................................................................................... 49

Chapter 2: Preparation and Crystallographic Investigation of Racemic Snakin-1 . 51
2.1 Overview ............................................................................................................ 53
2.2 Antimicrobial Peptides ..................................................................................... 53
2.2.1 Cysteine-rich AMPs in plants ...................................................................... 55
2.2.1.1 Plant defensins ............................................................................................. 56
2.2.1.2 Thionins ....................................................................................................... 58
2.2.1.3 α-helical hairpins ......................................................................................... 59
2.2.2 Snakin-1 ....................................................................................................... 60
2.2.3 The GASA/snakin family ............................................................................ 65
2.2.4 Aims of present study .................................................................................. 67
2.3 Synthesis of L-snakin-1 ................................................................................... 68
2.3.1 Synthesis of snakin-1 peptide fragments 2.1 and 2.2 ................................. 69
2.3.2 Native chemical ligation of 2.1 and 2.2 ....................................................... 72
2.3.3 Oxidative folding ........................................................................................... 74
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.2</td>
<td>Crystal structure of true racemic snakin-1</td>
<td>160</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Structural alignment</td>
<td>168</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Structural similarity to crAMPs</td>
<td>170</td>
</tr>
<tr>
<td>3.6</td>
<td>Antimicrobial activity of D-snakin-1</td>
<td>173</td>
</tr>
<tr>
<td>3.7</td>
<td>Discussion</td>
<td>174</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Iodo-RIP</td>
<td>174</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Amphiphilic AMPs and membrane disruption</td>
<td>175</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Modulation of antimicrobial activity</td>
<td>176</td>
</tr>
<tr>
<td>3.7.4</td>
<td>Iodo-L-snakin 1 Gly-254-iodo-Phe-COSR thioester</td>
<td>180</td>
</tr>
<tr>
<td>3.7.5</td>
<td>Reduced full length iodo-L-snakin 1 Gly-254-iodo-Phe-Pro</td>
<td>180</td>
</tr>
<tr>
<td>3.7.6</td>
<td>Iodo-L-snakin 1 Gly-254-iodo-Phe-Pro-OH</td>
<td>181</td>
</tr>
<tr>
<td>3.7.7</td>
<td>Circular dichroism spectrometry</td>
<td>181</td>
</tr>
<tr>
<td>3.7.8</td>
<td>Initial crystallisation screen of iodo-snakin-1 quasi-racemate</td>
<td>181</td>
</tr>
<tr>
<td>3.7.9</td>
<td>Initial crystallisation screen of iodo-snakin-1 quasi-racemate</td>
<td>182</td>
</tr>
<tr>
<td>3.7.10</td>
<td>Determination of antifungal activity</td>
<td>183</td>
</tr>
</tbody>
</table>

**Chapter 4: General Experimental Methods**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Materials and methods</td>
<td>187</td>
</tr>
<tr>
<td>4.2</td>
<td>General Methods for Peptide Synthesis</td>
<td>188</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Ninhydrin test</td>
<td>188</td>
</tr>
<tr>
<td>4.2.2</td>
<td>(N)-phthalamidomethyl polystyrene resin (1 mmol/g loading)</td>
<td>188</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Aminomethyl polystyrene resin (1 mmol/g loading)</td>
<td>188</td>
</tr>
<tr>
<td>4.2.4</td>
<td>(In) situ neutralisation Boc-SPPS</td>
<td>189</td>
</tr>
<tr>
<td>4.2.5</td>
<td>HF cleavage</td>
<td>189</td>
</tr>
<tr>
<td>4.3</td>
<td>Protein Crystallisation</td>
<td>190</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Robot crystallisation screening</td>
<td>190</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Crystallisation optimisation</td>
<td>190</td>
</tr>
<tr>
<td>4.4</td>
<td>X-Ray Crystallography</td>
<td>191</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Crystal preparation</td>
<td>191</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Crystal diffraction data collection</td>
<td>191</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Diffraction data processing</td>
<td>192</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Analysis of unit cell</td>
<td>192</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Structure solution</td>
<td>192</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Model building and refinement</td>
<td>194</td>
</tr>
<tr>
<td>4.4.7</td>
<td>Structure validation tools</td>
<td>194</td>
</tr>
<tr>
<td>4.4.7.1</td>
<td>(R)-factors</td>
<td>194</td>
</tr>
<tr>
<td>4.4.7.2</td>
<td>Structure validation servers</td>
<td>195</td>
</tr>
<tr>
<td>4.4.8</td>
<td>Structural alignment</td>
<td>195</td>
</tr>
<tr>
<td>4.4.9</td>
<td>Crystallography programs</td>
<td>196</td>
</tr>
<tr>
<td>Appendix: Preparation of Truncated Orf Virus Entry Fusion Complex Proteins by Chemical Synthesis</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>211</td>
<td></td>
</tr>
</tbody>
</table>
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-iodo-Phe</td>
<td>4-Iodophenylalanine</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acid</td>
</tr>
<tr>
<td>AcOH</td>
<td>Acetic acid</td>
</tr>
<tr>
<td>ACP</td>
<td>Acyl carrier protein</td>
</tr>
<tr>
<td>AM</td>
<td>Aminomethyl</td>
</tr>
<tr>
<td>AMP</td>
<td>Antimicrobial peptide</td>
</tr>
<tr>
<td>AMPSO</td>
<td>N-(1,1-Dimethyl-2-hydroxyethyl)-3-amino-2-hydroxypropanesulfonic acid</td>
</tr>
<tr>
<td>Boc</td>
<td>tert-Butyloxycarbonyl</td>
</tr>
<tr>
<td>BOP</td>
<td>1-Benzotriazolyloxytris(dimethylamino)phosphonium hexafluorophosphate</td>
</tr>
<tr>
<td>Bzl</td>
<td>Benzyl</td>
</tr>
<tr>
<td>CC</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-coupled device</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>crAMP</td>
<td>Cysteine-rich antimicrobial peptide</td>
</tr>
<tr>
<td>DBU</td>
<td>1,8-Diazabicyclo[5.4.0]undec-7-ene</td>
</tr>
<tr>
<td>Dbz</td>
<td>3,4-Diaminobenzoic acid</td>
</tr>
<tr>
<td>DIC</td>
<td>Diisopropylcarbodiimide</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-N,N-dimethylaminopyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DNP</td>
<td>2,4-Dinitrophenyl</td>
</tr>
<tr>
<td>DVB-PS</td>
<td>Divinylbenzene-poly styrene</td>
</tr>
<tr>
<td>EMBL</td>
<td>European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-Fluorenylethoxycarbonyl</td>
</tr>
<tr>
<td>FT-IR</td>
<td>Fourier-transform infrared</td>
</tr>
<tr>
<td>GASA</td>
<td>Gibberellic acid-stimulated in Arabidopsis</td>
</tr>
<tr>
<td>GnHCl</td>
<td>Guanidine hydrochloride</td>
</tr>
<tr>
<td>HATU</td>
<td>O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>HBTU</td>
<td>O-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HMPA</td>
<td>Hexamethylphosphoramide</td>
</tr>
<tr>
<td>HOAt</td>
<td>1-Hydroxy-7-azabenzotriazole</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-Hydroxybenzotriazole</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>HREM</td>
<td>High-energy remote</td>
</tr>
<tr>
<td>HTH</td>
<td>Helix-turn-helix</td>
</tr>
<tr>
<td>INFL</td>
<td>Inflection point</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LREM</td>
<td>Low-energy remote</td>
</tr>
<tr>
<td>LSQ</td>
<td>Least-squares</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MESNa</td>
<td>Sodium 2-mercaptoethansulfonate</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MPAA</td>
<td>4-Mercaptoacetic acid</td>
</tr>
<tr>
<td>MPEG</td>
<td>Methoxypolyethylene glycol</td>
</tr>
<tr>
<td>MQ</td>
<td>Milli-Q</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td>NCL</td>
<td>Native chemical ligation</td>
</tr>
<tr>
<td>NCS</td>
<td>Non-crystallographic symmetry</td>
</tr>
<tr>
<td>NMP</td>
<td>1-Methyl-2-pyrrolidone</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PAM</td>
<td>Phenylacetamide</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PG</td>
<td>Protecting group</td>
</tr>
<tr>
<td>phi</td>
<td>Phenyl</td>
</tr>
<tr>
<td>PTH1</td>
<td>Pseudothionin-1</td>
</tr>
<tr>
<td>PyAOP</td>
<td>7-Azabenzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate</td>
</tr>
<tr>
<td>PyBOP</td>
<td>1-Benzotriazoloyloxytris(pyrrolidino)phosphonium hexafluorophosphate</td>
</tr>
<tr>
<td>RIP</td>
<td>Radiation damage-induced phasing</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean square difference</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse phase high-performance liquid chromatography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>$R_t$</td>
<td>Retention time</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>Sec</td>
<td>Selenocysteine</td>
</tr>
<tr>
<td>sfAFP</td>
<td>Snow-flea antifreeze protein</td>
</tr>
<tr>
<td>SIR</td>
<td>Single isomorphous replacement</td>
</tr>
<tr>
<td>SN1</td>
<td>Snakin-1</td>
</tr>
<tr>
<td>SPPS</td>
<td>Solid phase peptide synthesis</td>
</tr>
<tr>
<td>StSUT1</td>
<td>Potato sucrose transporter 1</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>TFZ</td>
<td>Translation function Z-score</td>
</tr>
<tr>
<td>TIPS</td>
<td>Triisopropylsilane</td>
</tr>
<tr>
<td>TMS</td>
<td>Tetramethylsilane</td>
</tr>
<tr>
<td>tNCS</td>
<td>Translational non-crystallographic symmetry</td>
</tr>
<tr>
<td>Trt</td>
<td>Trityl</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>Xaa</td>
<td>Any amino acid</td>
</tr>
</tbody>
</table>
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Appendix: Preparation of Truncated Orf Virus Entry Fusion Complex Proteins by Chemical Synthesis

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Synthesis and write-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>90</td>
</tr>
</tbody>
</table>

### CO-AUTHORS

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret A. Brimble</td>
<td>Supervision, advice and editing</td>
</tr>
<tr>
<td>Edward N. Baker</td>
<td>Supervision, advice and editing</td>
</tr>
<tr>
<td>Christopher J. Squire</td>
<td>Supervision, advice and editing</td>
</tr>
<tr>
<td>Paul W. R. Harris</td>
<td>Supervision, advice and editing</td>
</tr>
</tbody>
</table>

### Certification by Co-Authors

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret A. Brimble</td>
<td></td>
<td>14/01/2016</td>
</tr>
<tr>
<td>Edward N. Baker</td>
<td></td>
<td>14/01/2016</td>
</tr>
<tr>
<td>Christopher J. Squire</td>
<td></td>
<td>14/01/2016</td>
</tr>
<tr>
<td>Paul W. R. Harris</td>
<td></td>
<td>14/01/2016</td>
</tr>
</tbody>
</table>

Last updated: 19 October 2015
This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.


| Nature of contribution by PhD candidate | Synthesis, crystallography and write-up |
| Extent of contribution by PhD candidate (%) | 90 |

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret A. Brimble</td>
<td>Supervision, advice, and editing</td>
</tr>
<tr>
<td>Edward N. Baker</td>
<td>Supervision, advice, and editing</td>
</tr>
<tr>
<td>Christopher J. Squire</td>
<td>Supervision, advice, and editing</td>
</tr>
<tr>
<td>Paul W. R. Harris</td>
<td>Supervision, advice, and editing</td>
</tr>
<tr>
<td>Santosh Panjikar</td>
<td>X-ray diffraction data collection</td>
</tr>
<tr>
<td>Antonio Molina</td>
<td>Biological activity testing</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>DProf Margaret Brimble</td>
<td></td>
<td>20/7/16</td>
</tr>
<tr>
<td>Dr. Christopher Squire</td>
<td></td>
<td>20/7/2016</td>
</tr>
<tr>
<td>DProf Edward Baker</td>
<td></td>
<td>20/7/2016</td>
</tr>
<tr>
<td>Dr. Santosh Panjikar</td>
<td></td>
<td>20/7/2016</td>
</tr>
<tr>
<td>Dr. Antonio Molina</td>
<td></td>
<td>20/7/2016</td>
</tr>
</tbody>
</table>

Last updated: 19 October 2015
Garland Science - Books LICENSE
TERMS AND CONDITIONS

Jul 19, 2016

This is a License Agreement between Ho Yeung ("You") and Garland Science - Books ("Garland Science - Books") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Garland Science - Books, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3805650210635</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Dec 15, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Garland Science - Books</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Biomolecular crystallography: principles, practice, and application to structural biology</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2009</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Academic institution</td>
</tr>
<tr>
<td>Format</td>
<td>Print, Electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>chart/graph/table/figure</td>
</tr>
<tr>
<td>Number of charts/graphs/tables/figures</td>
<td>1</td>
</tr>
<tr>
<td>Title or numeric reference of the portion(s)</td>
<td>Figure 6-30</td>
</tr>
<tr>
<td>Title of the article or chapter the portion is from</td>
<td>Diffraction Basics</td>
</tr>
<tr>
<td>Editor of portion(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Author of portion(s)</td>
<td>Bernhard Rupp</td>
</tr>
<tr>
<td>Volume of serial or monograph.</td>
<td>N/A</td>
</tr>
<tr>
<td>Page range of the portion</td>
<td>247-312</td>
</tr>
<tr>
<td>Publication date of portion</td>
<td>2010</td>
</tr>
<tr>
<td>Rights for</td>
<td>Main product</td>
</tr>
<tr>
<td>Duration of use</td>
<td>Current edition and up to 5 years</td>
</tr>
<tr>
<td>Creation of copies for the disabled</td>
<td>no</td>
</tr>
<tr>
<td>With minor editing privileges</td>
<td>no</td>
</tr>
<tr>
<td>For distribution to</td>
<td>Worldwide</td>
</tr>
<tr>
<td>In the following language(s)</td>
<td>Original language of publication</td>
</tr>
<tr>
<td>With incidental promotional use</td>
<td>no</td>
</tr>
<tr>
<td>The lifetime unit quantity of new product</td>
<td>Up to 499</td>
</tr>
<tr>
<td>Made available in the following markets</td>
<td>education</td>
</tr>
<tr>
<td>The requesting person/organization is:</td>
<td>Ho Yeung</td>
</tr>
</tbody>
</table>

https://s100.copyright.com//CustomerAdmin/PLF.jsp?ref=3bf4ea5c-49a4-49f3-8d12-6425f191b796 1/5
Order reference number
Author/Editor Ho Yeung
The standard identifier of New Work hyeu016
Title of New Work Investigation of Biologically Relevant Proteins by Racemic Protein Crystallography
Publisher of New Work University of Auckland
Expected publication date Jun 2016
Estimated size (pages) 200
Total (may include CCC user fee) 0.00 USD

TERMS AND CONDITIONS

The following terms are individual to this publisher:

None

Other Terms and Conditions:

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the “Work(s)”). Copyright Clearance Center, Inc. (“CCC”) grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the “Rightsholder”). “Republication”, as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. “User”, as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a “freelancer” or other third party independent of User and CCC, such party shall be deemed jointly a “User” for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on “net 30” terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.
3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is “one-time” (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: “Republished with permission of [Rightsholder’s name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc.” Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED “AS IS”. CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR
PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.
8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.
8.2 Use of User-related information collected through the Service is governed by CCC’s privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.
8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User’s rights in the new material which includes the Work(s) licensed under this Service.
8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.
8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Ho Yeung ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3611050194048</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 16, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Oxford University Press</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Molecular Biology and Evolution</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms:</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Attila Miseta, Peter Csutora</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>08/01/2000</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Title of your work</td>
<td>Investigation into racemic protein crystallography</td>
</tr>
<tr>
<td>Publisher of your work</td>
<td>n/a</td>
</tr>
<tr>
<td>Expected publication date</td>
<td>Sep 2015</td>
</tr>
<tr>
<td>Permissions cost</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Value added tax</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Ho Yeung 3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td>Attn: Ho Yeung</td>
<td></td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Billing Address</td>
<td>Ho Yeung 3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td>Attn: Ho Yeung</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL**

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources.
which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC’s Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC’s Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press’s written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Ho Yeung ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3757461415165</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 28, 2015</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Nature Protocols</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Using circular dichroism spectra to estimate protein secondary structure</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Norma J Greenfield</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jan 25, 2007</td>
</tr>
<tr>
<td>Licensed Content Volume Number</td>
<td>1</td>
</tr>
<tr>
<td>Licensed Content Issue Number</td>
<td>6</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Investigation into racemic protein crystallography</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Sep 2015</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>200</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Ho Yeung</td>
</tr>
<tr>
<td></td>
<td>3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Billing Address</td>
<td>Ho Yeung</td>
</tr>
<tr>
<td></td>
<td>3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td></td>
<td>Auckland, New Zealand 1010</td>
</tr>
<tr>
<td></td>
<td>Attn: Ho Yeung</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the ‘Use on a Web Site’ option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   **Note: For republication from the British Journal of Cancer, the following credit lines apply.**
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)
For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:
   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   **Note: For adaptation from the British Journal of Cancer, the following credit line applies.**
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:
   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   **Note: For translation from the British Journal of Cancer, the following credit line applies.**
We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Ho Yeung ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3685021222563</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Aug 09, 2015</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Peptides</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Theoretical structural insights into the snakin/GASA family</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>William F. Porto, Octavio L. Franco</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>June 2013</td>
</tr>
<tr>
<td>Licensed Content Volume Number</td>
<td>44</td>
</tr>
<tr>
<td>Licensed Content Issue Number</td>
<td>n/a</td>
</tr>
<tr>
<td>Start Page</td>
<td>163</td>
</tr>
<tr>
<td>End Page</td>
<td>167</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>2</td>
</tr>
<tr>
<td>Format</td>
<td>both print and electronic</td>
</tr>
<tr>
<td>Are you the author of this Elsevier article?</td>
<td>No</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
<tr>
<td>Original figure numbers</td>
<td>Figures 1, 2</td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>Investigation into racemic protein crystallography</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Sep 2015</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>200</td>
</tr>
<tr>
<td>Elsevier VAT number</td>
<td>GB 494 6272 12</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Ho Yeung</td>
</tr>
<tr>
<td></td>
<td>3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td></td>
<td>Auckland, 1010</td>
</tr>
<tr>
<td></td>
<td>New Zealand</td>
</tr>
<tr>
<td></td>
<td>Attn: Ho Yeung</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Billing Address</td>
<td>Ho Yeung</td>
</tr>
<tr>
<td></td>
<td>3A Symonds St, Auckland Central</td>
</tr>
<tr>
<td></td>
<td>Auckland, New Zealand 1010</td>
</tr>
</tbody>
</table>
INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed,
assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

**LIMITED LICENSE**

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homeland of the journal from which you are licensing at [http://www.sciencedirect.com/science/journal/xxxxx](http://www.sciencedirect.com/science/journal/xxxxx) or the Elsevier homepage for books at [http://www.elsevier.com](http://www.elsevier.com); Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at [http://www.elsevier.com](http://www.elsevier.com). All content posted to the web site must maintain the copyright information line on the bottom of each image.

**Posting licensed content on Electronic reserve**: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors**: the following clauses are applicable in addition to the above:

**Preprints**:  
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.). Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications. Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment. Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on
demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

**Elsevier Open Access Terms and Conditions**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

**Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated. The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

**Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions:**

v1.7

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This is a License Agreement between Ho Yeung ("You") and American Society of Plant Biologists ("American Society of Plant Biologists") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Society of Plant Biologists, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3755880603692
License date Nov 25, 2015
Licensed content publisher American Society of Plant Biologists
Licensed content title Plant physiology
Licensed content date Jan 1, 1998
Type of Use Thesis/Dissertation
Requestor type Academic institution
Format Print, Electronic
Portion image/photo
Number of images/photos requested 2
Title or numeric reference of the portion(s) Figures 1 and 2
Title of the article or chapter the portion is from Potato Snakin-1 Gene Silencing Affects Cell Division, Primary Metabolism, and Cell Wall Composition
Editor of portion(s) Donald R. Ort
Author of portion(s) Vanesa Nahiriñak, Natalia InÁ©s Almasia, Paula Virginia Fernandez, Horacio Esteban Hopp, JosÁ© Manuel Estevez, Fernando Carrari and Cecilia Vazquez-Rovere
Volume of serial or monograph. 158
Issue, if republishing an article from a serial 1
Page range of the portion 252-263
Publication date of portion January 2012
Rights for Main product
Duration of use Life of current edition
Creation of copies for the disabled no
With minor editing privileges no
For distribution to Worldwide
In the following language(s) Original language of publication
With incidental promotional use no
The lifetime unit quantity of new product Up to 499
Made available in the education
following markets

The requesting person/organization is: Ho Yeung

Order reference number

Author/Editor Ho Yeung

The standard identifier of New Work HYEU016

Title of New Work Investigation of Biologically Relevant Proteins by Racemic Protein Crystallography

Publisher of New Work University of Auckland

Expected publication date Jun 2016

Estimated size (pages) 250

Total (may include CCC user fee) 0.00 USD

Terms and Conditions

TERMS AND CONDITIONS
The following terms are individual to this publisher:

Permission is granted for the life of the current edition and all future editions, in all languages and in all media.

Other Terms and Conditions: STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the “Work(s)”). Copyright Clearance Center, Inc. (“CCC”) grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the “Rightsholder”). “Republication”, as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. “User”, as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a “freelancer” or other third party independent of User and CCC, such party shall be deemed jointly a “User” for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on “net 30” terms. While User may exercise the rights
licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is “one-time” (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: “Republished with permission of [Rightsholder’s name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc.” Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED “AS IS”, CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER
CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC’s privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User’s rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC’s standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC’s sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court. If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1
Questions? customeercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This is a License Agreement between Ho Yeung ("You") and Springer Science and Bus Media B V ("Springer Science and Bus Media B V") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer Science and Bus Media B V, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

<table>
<thead>
<tr>
<th>License Number</th>
<th>3767540594780</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Dec 14, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Springer Science and Bus Media B V</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Principles of protein x-ray crystallography</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2007</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Academic institution</td>
</tr>
<tr>
<td>Format</td>
<td>Print, Electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>chart/graph/table/figure</td>
</tr>
<tr>
<td>Number of charts/graphs/tables/figures</td>
<td>1</td>
</tr>
<tr>
<td>Title or numeric reference of the portion(s)</td>
<td>Figure 7.1</td>
</tr>
<tr>
<td>Title of the article or chapter the portion is from</td>
<td>The Solution of the Phase Problem by the Isomorphous Replacement Method</td>
</tr>
<tr>
<td>Editor of portion(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Author of portion(s)</td>
<td>Jan Drenth</td>
</tr>
<tr>
<td>Volume of serial or monograph.</td>
<td>N/A</td>
</tr>
<tr>
<td>Page range of the portion</td>
<td>123-171</td>
</tr>
<tr>
<td>Publication date of portion</td>
<td>2007</td>
</tr>
<tr>
<td>Rights for</td>
<td>Main product</td>
</tr>
<tr>
<td>Duration of use</td>
<td>Current edition and up to 5 years</td>
</tr>
<tr>
<td>Creation of copies for the disabled</td>
<td>no</td>
</tr>
<tr>
<td>With minor editing privileges</td>
<td>yes</td>
</tr>
<tr>
<td>For distribution to</td>
<td>Worldwide</td>
</tr>
<tr>
<td>In the following language(s)</td>
<td>Original language of publication</td>
</tr>
<tr>
<td>With incidental promotional use</td>
<td>no</td>
</tr>
<tr>
<td>The lifetime unit quantity of new product</td>
<td>Up to 499</td>
</tr>
<tr>
<td>Made available in the following markets</td>
<td>education</td>
</tr>
<tr>
<td>The requesting person/organization is:</td>
<td>Ho Yeung</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
</tbody>
</table>
TERMS AND CONDITIONS

The following terms are individual to this publisher:

A maximum of 10% of the content may be licensed for republication. The user is responsible for identifying and seeking separate licenses for any third party materials that are identified anywhere in the work. Without a separate license, such third party materials may not be reused.

Other Terms and Conditions:

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the “Work(s)”). Copyright Clearance Center, Inc. (“CCC”) grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the “Rightsholder”). “Republication”, as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. “User”, as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a “freelancer” or other third party independent of User and CCC, such party shall be deemed jointly a “User” for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on “net 30” terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the
license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is “one-time” (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User’s stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: “Republished with permission of [Rightsholder’s name], from [Work’s title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. ” Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED “AS IS”. CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER
EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC’s privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User’s rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court. If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or
This Agreement between Ho Yeung ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number: 3662791104830
License date: Jul 05, 2015
Licensed Content Publisher: Elsevier
Licensed Content Publication: Protein Expression and Purification
Licensed Content Title: Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins
Licensed Content Author: Natalia Kovalskaya, Rosemarie W. Hammond
Licensed Content Date: January 2009
Licensed Content Volume Number: 63
Licensed Content Issue Number: 1
Start Page: 12
End Page: 17
Type of Use: reuse in a thesis/dissertation
Portion: figures/tables/illustrations
Number of figures/tables/illustrations: 2
Format: both print and electronic
Are you the author of this Elsevier article? No
Will you be translating? No
Order reference number
Original figure numbers: Figures 4 and 5
Title of your thesis/dissertation: Investigation into racemic protein crystallography
Expected completion date: Sep 2015
Estimated size (number of pages): 200
Elsevier VAT number: GB 494 6272 12
Requestor Location: Ho Yeung
3A Symonds St, Auckland Central

Billing Type: Invoice
Billing Address: Ho Yeung
3A Symonds St, Auckland Central
INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. Translation: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. For journal authors: the following clauses are applicable in addition to the above:

Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.). Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications. Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment. Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect. Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of
the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

**Elsevier Open Access Terms and Conditions**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](https://creativecommons.org/licenses) for more information.

**Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated. The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

**Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at [http://creativecommons.org/licenses/by/4.0](http://creativecommons.org/licenses/by/4.0).

**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at [http://creativecommons.org/licenses/by-nc-sa/4.0](http://creativecommons.org/licenses/by-nc-sa/4.0).

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at [http://creativecommons.org/licenses/by-nc-nd/4.0](http://creativecommons.org/licenses/by-nc-nd/4.0). Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions:**

v1.7

Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Ho Yeung ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number: 3763520432878
License date: Dec 07, 2015
Licensed Content Publisher: Elsevier
Licensed Content Publication: Trends in Biotechnology
Licensed Content Title: The expanding scope of antimicrobial peptide structures and their modes of action
Licensed Content Author: Leonard T. Nguyen, Evan F. Haney, Hans J. Vogel
Licensed Content Date: September 2011
Licensed Content Volume: 29
Licensed Content Issue: 9
Start Page: 464
End Page: 472
Type of Use: reuse in a thesis/dissertation
Intended publisher of new work: other
Portion: figures/tables/illustrations
Number of figures/tables/illustrations: 1
Format: both print and electronic
Are you the author of this Elsevier article?: No
Will you be translating?: No
Order reference number: 
Original figure numbers: Figure 2
Title of your thesis/dissertation: Investigation into racemic protein crystallography
Expected completion date: Sep 2015
Estimated size (number of pages): 200
Elsevier VAT number: GB 494 6272 12
Requestor Location: Ho Yeung
3A Symonds St, Auckland Central
Auckland, 1010
New Zealand
Attn: Ho Yeung
Billing Type: Invoice
Billing Address: Ho Yeung
INTRODUCTION
1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS
2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all
claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. Translation: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:

Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.). Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications. Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment. Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include...
permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

**Elsevier Open Access Terms and Conditions**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

**Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated. The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

**Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at [http://creativecommons.org/licenses/by/4.0](http://creativecommons.org/licenses/by/4.0).

**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at [http://creativecommons.org/licenses/by-nc-sa/4.0](http://creativecommons.org/licenses/by-nc-sa/4.0).

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at [http://creativecommons.org/licenses/by-nc-nd/4.0](http://creativecommons.org/licenses/by-nc-nd/4.0).

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.8

Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or
+1-978-646-2777.
Chapter 1

Racemic Protein Crystallography:
an Introduction
1.1 X-ray crystallography and the crystallographic phase problem

Since the determination of the first three-dimensional protein structure, myoglobin, by Kendrew et al. in the early 1960s,1-3 X-ray crystallography has become the most widely used method to determine three-dimensional macromolecular structures. To date, this method is the means by which 89% of macromolecular structures in the Protein Data Bank (PDB) have been determined.

The use of X-ray crystallography to determine the structure of a crystal involves bombarding the crystal with X-rays. X-rays are a form of electromagnetic radiation, and are scattered when they encounter regions of high electron density. When an X-ray beam is fired on an ordered periodic arrangement of molecules, i.e. a crystal, diffraction from the electron dense atoms produces a distinctive pattern that can be captured as an image. Measurement of this diffraction pattern provides the data that can then be used to determine the underlying crystal structure (Figure 1.1).

![Figure 1.1. X-ray crystallography utilises the diffraction of X-rays by molecular crystals to determine an atomic model of the underlying molecular motif. Diffraction frame shown was kindly provided by Mr. Matthew Sullivan (University of Auckland), and the protein structure is of hen egg-white lysozyme (PDB code: 2vb1).](image)

Despite its widespread use, macromolecular structure solution by X-ray crystallography is not a straightforward process. The diffraction patterns measured are collections of intensities corresponding to individual X-ray waves scattered from the crystal. These collected intensities can be used to visualise the electron density present in the minimum repeating unit of the crystal, the unit cell, and the basic concepts required to understand this transformation will be discussed briefly (For comprehensive introductory texts, the reader is directed to Rupp,4 and Rhodes5).
An X-ray wave scattered by a crystal can be expressed in the form of a Fourier series called a structure factor $F_{hkl}$:

Equation 1.1. The structure factor equation. $F_{hkl}$ is a structure factor representing diffraction of an X-ray wave from a set of Bragg planes $h,k,l$ and the atoms in those planes. The equation is the form of a Fourier synthesis of the electron density over the entire unit cell, for units of electron density at coordinates $(x, y, z)$.

$$F_{hkl} = V \int \int \int \rho(x, y, z) e^{2\pi i (hx + ky + lz)} \, dx \, dy \, dz$$

where $\rho(x,y,z)$ is a unit of electron density at fractional coordinates $(x, y, z)$ in the unit cell. The intensity of each diffraction spot observed, $I_{hkl}$, is related to $F_{hkl}$.

The structure factor is the integral of the electron density over the entire unit cell, and the equation above shows that every structure factor is influenced by the total electron density in the unit cell. The exponential $e^{2\pi i (hx + ky + lz)}$ indicates that each term in the Fourier series is a three-dimensional function and can be expressed as $\cos 2\pi(hx + ky + lz) + i \sin 2\pi(hx + ky + lz)$. According to Bragg’s model of diffraction, a structure factor with this exponential term can be interpreted as X-ray diffraction originating from the $hkl$ set of Bragg diffraction planes passing through the crystal.6

The reverse Fourier transform of this series, the electron density equation:

Equation 1.2. The electron density equation. $\rho(x, y, z)$ is the reverse Fourier transform of the Fourier series consisting of the structure factors $F_{hkl}$ representing the entire unit cell.

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

shows that given a complete collection of structure factors $F_{hkl}$ measured over the entire unit cell, the electron density $\rho(x,y,z)$ at coordinates $(x, y, z)$ can be determined.

Each $F_{hkl}$ is a representation of a scattered X-ray wave, and so is a function with a frequency, amplitude, and phase. The frequency is the same as that of the incident X-rays. To visualise the amplitude and phase of the structure factor, $F_{hkl}$ will be represented as a complex vector $\mathbf{F}_{hkl}$ in the form:

Equation 1.3. Representation of a structure factor as a complex vector consisting of real $(a)$ and imaginary $(ib)$ components.

$$\mathbf{F}_{hkl} = a + ib$$
where \( a \) is the real component and \( ib \) is the imaginary component.

The wave can be represented as a complex vector whose length represents the \textit{amplitude} of a wave and whose direction represents its \textit{phase}:

\[
\begin{align*}
\mathbf{F}_{hkl} &= |A| e^{i\phi_{hkl}} \\
\mathbf{A} &= |A| e^{i\phi_{hkl}} \\
\mathbf{B} &= i|B|
\end{align*}
\]

\( \phi_{hkl} \) is the phase of the vector \( \mathbf{F}_{hkl} \). This representation is known as an Argand diagram.

As the phase \( \phi_{hkl} \) varies between a range of 0 and \( 2\pi \), the projections of \( \mathbf{F}_{hkl} \) on the real axis, \( |A| \), and on the imaginary axis, \( i|B| \), vary (Figure 1.2). The magnitude of the vector, its amplitude \( |\mathbf{F}_{hkl}| \), is directly proportional to the square root of \( I_{hkl} \), the measured intensity of a diffraction spot (also known as a \textit{reflection}; Equation 1.4):

\[
I_{hkl} = K |\mathbf{F}_{hkl}|^2
\]

To use the reverse Fourier transform above (Equation 1.2) to determine \( \rho(x,y,z) \) from the measured intensities \( I_{hkl} \), the phase \( \phi_{hkl} \) for each \( \mathbf{F}_{hkl} \) is required. The inability to directly determine \( \phi_{hkl} \) from a measurement of \( I_{hkl} \) is known as the crystallographic \textit{phase problem}.

\subsection*{1.1.1 Heavy-atom phasing methods}

A common method of estimating the crystallographic phases, and the first to be used successfully for proteins, requires crystals incorporating heavy atoms. For crystals of
organic molecules containing atoms of carbon, nitrogen, oxygen, and hydrogen, the inclusion of an electron-rich atom such as a metal means that for certain reflections, the contribution from this highly-scattering source will dominate all those from the light atoms. In larger molecules, this effect is increased, as although there is a greater number of scattering contributions from light molecules, interference effects are also increased.

Given the heavy-atom positions, the phase of the substructure $F_H$ is known, and with a measure of the total structure factor of the crystal $F_{PH}$, it is possible to determine the phase of a structure factor corresponding to the remaining atoms $F_P$ (Equation 1.5, Figure 1.3):

Equation 1.5. Representation of a structure factor $F_{PH}$ as the sum of the structure factors of a substructure $F_H$ and the remaining structure $F_P$.

$$F_P = F_{PH} - F_H$$

Figure 1.3. Representation of a structure factor $F_{PH}$ as the sum of structure factors $F_H$ and $F_P$.

Common methods used to measure a heavy-atom signal include isomorphous replacement (§1.1.1.1), and anomalous dispersion from an added or intrinsically-present heavy-atom (§1.1.1.2), and methods typically used to determine the positions of the heavy-atom substructure are Patterson methods (§1.1.1.3) and direct methods (§1.1.1.4).

1.1.1.1 Isomorphous replacement

Isomorphous replacement requires one or more heavy-atom derivative crystals, and one native protein crystal. The subtraction of the native data $F_P$ from a derivative data set $F_{PH}$ can be used to approximate the heavy-atom substructure $F_H$ contribution in the derivative. This is the method first used in the successful solution of the crystal structure of myoglobin.3
Historically, isomorphous replacement required multiple isomorphous derivatives (multiple isomorphous replacement; MIR), as when only a single derivative is used (single isomorphous replacement; SIR), an ambiguity remains between two phases (Figures 1.4, 1.5).

Figure 1.4. Harker construction representing phase determination by SIR. The Harker construction visually represents Equation 1.5, where the phase of $F_P$ and $F_{PH}$ are unknown. The green and yellow circles represent the possible values for $F_{PH}$ and $F_P$, respectively, with the circle of $F_{PH}$ centred at the end of a vector corresponding to the structure factor of the heavy-atom substructure.
Chapter 1

Figure 1.5. Harker construction representing phase determination by MIR. Green and red circles represent the possible values for $F_{PH1}$ and $F_{PH2}$, respectively, and each is centred at the end of a vector corresponding to the structure factor of the respective heavy-atom substructure ($F_{H1}$ or $F_{H2}$). If more than one derivative is available and the $F_H$ of all derivatives are known, the phase of $F_P$ can be determined unambiguously.

Modern methods have enabled the routine determination of crystal structures via SIR, through the use of computational density modification and structure refinement techniques to resolve the phase ambiguity (for introductions to these methods, see Zhang, Cowtan and Main,\textsuperscript{7} and Ten Eyck and Watenpaugh\textsuperscript{8}).

In practise, Equation 1.5 does not hold completely in isomorphous replacement experiments, due to errors in assigning metal atom positions, non-isomorphism between crystals and measurement errors.\textsuperscript{9}

1.1.1.2 Anomalous dispersion

Structure solution by isomorphous replacement is highly dependent on isomorphism between the native crystal and heavy-atom derivatives. This is often difficult to achieve, as heavy-atom derivatisation of native crystals can displace native atoms or result in small movements of the molecules in the crystal. This is sometimes accompanied by changes in the unit cell dimensions. Contemporary methods of heavy-atom phasing commonly exploit wavelength-dependent anomalous dispersion (also known as anomalous scattering) effects,
negating the requirement for multiple isomorphous crystals (for a comprehensive review, see Hendrickson\textsuperscript{10}).

In X-ray crystallography, anomalous dispersion refers to a wavelength-dependent phenomenon that is observed when the energy of an incident photon is sufficient to release an electron from an orbital of an atom, forming a positively-charged ion (Figure 1.6).

<table>
<thead>
<tr>
<th>Shell</th>
<th>n</th>
<th>l</th>
<th>j = l + m_j</th>
<th>m_j = (j... j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>3</td>
<td>2</td>
<td>5/2</td>
<td>±1/2 ±3/2 ±5/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3/2</td>
<td>±1/2 ±3/2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>3/2</td>
<td>±1/2 ±3/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1/2</td>
<td>±1/2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1/2</td>
<td>±1/2</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>1</td>
<td>3/2</td>
<td>±1/2 ±3/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1/2</td>
<td>±1/2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1/2</td>
<td>±1/2</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>0</td>
<td>1/2</td>
<td>±1/2</td>
</tr>
</tbody>
</table>

Figure 1.6. The X-ray absorption process. \textit{Left:} tabulated representation of the different energy levels in an atom to the level of the Dirac fine structure. \( n \) is the principal energy level, \( l \) is the orbital angular momentum, \( j \) is the total angular momentum, and \( m\_j \) is the secondary total angular momentum quantum number. \textit{Right:} the absorption of a high-energy photon (red wiggly arrow) by an electron, resulting in its ejection (blue arrow) from the atom. The ‘hole’ leftover can be filled by a higher-order electron, and this process results in a characteristic X-ray emission. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

Filling of the electron ‘hole’ by an electron from a higher shell results in emission of a wave, which is out of phase by \( \frac{\pi}{2} \) with standard diffracted X-rays.\textsuperscript{10} The phase differences of these emitted photons produce a measurable difference in the Friedel mates \( F\_hkl \) and \( F\_-hkl \) (Figure 1.7), which is exploited in anomalous dispersion phasing methods, further discussed below.
Figure 1.7. Visual representation of the Friedel mates $F_{hkl}$ and $F_{\bar{h}\bar{k}\bar{l}}$ in the absence of anomalous dispersion. The Friedel mates are mirror images in the real axis, and have the phases $\varphi$ and $-\varphi$.

As the liberation of electrons from orbitals is dependent on excitation with photons of sufficient energy, anomalous dispersion effects are dependent on X-ray energy, and thus wavelength. Importantly, most of the lighter elements, with the exception of sulfur and phosphorus, do not show significant anomalous dispersion at the wavelengths accessible for X-ray diffraction. These wavelength-dependent anomalous dispersion effects can be incorporated into an expression for the atomic scattering factor $f$ of an element, and can be separated into the anomalous scattering factor contributions $f'$ and $f''$, corresponding to real and imaginary components, respectively (Equation 1.6).\(^{11}\)

Equation 1.6. Cromer-Mann formula expressing an atomic scattering factor as the sum of a wavelength-independent contribution $f_0$, and wavelength-dependent anomalous scattering contributions $f'$ and $f''$. The wavelength-dependent terms can be expressed as $\Delta f$.

$$f = f_0 + f' + i \cdot f'' = f_0 + \Delta f$$

The differences due to the wavelength dependent contributions ($\Delta f$) can then be used to provide phase information. Usable information can be in the form of the dispersive differences ($\lambda \Delta F_R$) or Bijvoet differences ($\lambda \Delta F_i$), products of the respective $f'$ and $f''$ contributions (Figure 1.8). Where significant anomalous dispersion effects exist, dispersive differences can be determined from $F_{hkl}$ measured at two wavelengths, $\lambda_1F_{hkl}$ and $\lambda_2F_{hkl}$, and
Bijvoet differences can be measured between the Friedel mates $F_{hkl}$ and $F_{hkl}$ at each wavelength.

Figure 1.8. Visual representation of anomalous dispersion effects $\Delta F_R$ and $\Delta F_i$ between $F_{hkl}$ measured at wavelengths $\lambda_1$ and $\lambda_2$. $\Delta F_R$ is the difference between the real components of anomalous dispersion effects between wavelengths, and $\Delta F_i$ is difference between the imaginary components.

According to Friedel’s law, Friedel mates $F_{hkl}$ and $F_{hkl}$, with phases $\phi_{hkl} = -\phi_{hkl}$, have identical magnitudes (Figure 1.7). When significant anomalous dispersion effects are present, however, the addition of the anomalous scattering contributions $f'$ and $f''$ to Friedel mates is such that the magnitudes are no longer equal, and the phase relationship no longer holds (Figure 1.9). The differences between Friedel mates, the Bijvoet differences, are measurable, and can be used to determine the coordinates of the anomalous scatterer substructure in the unit cell and its corresponding phase contribution. This information is then used to solve the phase problem for the entire crystal structure.
Multiple-wavelength anomalous dispersion (MAD) methods exploit both the dispersive differences and Bijvoet differences for the unambiguous determination of phases. In analogy to isomorphous methods, single-wavelength anomalous dispersion (SAD) methods, which utilise only the Bijvoet differences measured from data collected at a single wavelength, result in a phase ambiguity that can again be resolved with modern refinement methods.

1.1.1.3 Locating heavy-atom sites via Patterson methods

Isomorphous and anomalous phasing methods can be used to provide phases for structure solution, but for this, the positions of the heavy-atoms or anomalous scatterers must be known. Two methods are used for the location of heavy-atom sites in a derivative: Patterson methods and direct methods.

The Patterson function $P(u, v, w)$ (Equation 1.7) is an inverse Fourier transform of the squared structure factor amplitudes, $|F_{hkl}|^2$: 

$P(u, v, w) = \sum_{h,k,l} F_{hkl}^* F_{hkl}$
Equation 1.7. The Patterson function, using the coefficient $|F_{hkl}|^2$. $P(u, v, w)$ is an inverse Fourier transform of the squared structure factor amplitudes, $|F_{hkl}|^2$ for the entire unit cell.

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos 2\pi(hu + kv + lw)$$

It was first described by Patterson as a weighted distribution around any element $r$, and represents a function where the maxima correspond to an interatomic vector $u$ between any two atoms at $r$ and $r + u$, in the unit cell (Equation 1.8, Figure 1.10). The Patterson function is an auto-correlation function, and as such, can be determined directly from the measured intensities (proportional to $|F_{hkl}|^2$) without phase information.

Equation 1.8. Representation of the Patterson function as a function correlating density at $r$ and $r + u$. 

$$A(u) = \frac{1}{a} \int_0^a \rho(r)\rho(r + u)dx$$

Figure 1.10. The relationship between: a) atoms 1 and 2 in the unit cell and their inter-atomic distance $u$; and b) the corresponding Patterson function. Reproduced from The Principles of Protein X-ray Crystallography (3rd Ed.; Springer-Verlag, New York), Drenth, J., The Solution of the Phase Problem by the Isomorphous Replacement Method, 123-171, Copyright (2007).

The Patterson function was first used to determine the coordinates of atoms in small inorganic and small organic molecule crystals, but this approach is impractical for the solution of larger structures, as the function contains $N(N-1)$ non-origin peaks for $N$ atoms. Harker introduced a simplification to the Patterson method by exploiting known crystal symmetry, but this alone is not useful in de novo structure determination of large macromolecular structures. Where it is routinely useful in contemporary macromolecular
crystallographic methods is in the determination of heavy atom sites via difference Patterson functions.

Difference Patterson functions are calculated from coefficients representative of the underlying substructure, and thus represent the auto-correlation of the scattering contributed by these atoms. The peaks in difference Patterson functions will then correspond only to the vectors between these atoms. An isomorphous difference Patterson uses $|\Delta F|^2$ difference structure factor coefficients (Equation 1.9) where:

\begin{equation}
|\Delta F|^2 = (|F_{PH}| - |F_P|)^2
\end{equation}

Similarly, the anomalous difference Patterson can be calculated from the $|\Delta F_{zh}|^2$ coefficients (Equation 1.10), for which $|\Delta F_{zh}|$ is proportional to the wavelength-independent contribution of the heavy atom substructure, $|\theta F_{H}(h)|$:

\begin{equation}
|\Delta F_{zh}|^2 = (|F_{hkl}| - |F_{h\bar{k}\bar{l}}|)^2
\end{equation}

The resulting difference Patterson functions can then be used to locate the sites of the corresponding substructure and its phase, and thus the phase of the entire crystal structure.

1.1.1.4 Locating heavy-atom sites via direct methods

The other predominant method for determining the heavy-atom substructure is direct methods. Direct methods rely on mathematical relationships between phases, or through random placement of atoms to generate a correct structure. While it is possible to use direct methods to solve complete crystal structures with data at a resolution better than 1.2 Å, the calculation of phases from large macromolecular structures can be difficult and computationally expensive. Hence, instead of complete structure solution, many modern direct methods approaches are applied to the determination of heavy-atom substructures followed by phasing from these. In analogy to Patterson methods, which use the difference structure factors $|\Delta F|^2$ and $|\Delta F_{zh}|^2$ as coefficients (Equations 1.9, 1.10), substructure solution with direct methods uses normalised difference structure factors $|\Delta E|$.17-19

Direct methods approaches are based on relationships between the phases of reflections. These relationships take the form of structure invariants, linear expressions of these
relationships. The structure invariant used most for direct methods in crystallography is the triplet invariant $\Phi_{HK}$ (Equation 1.11):

Equation 1.11. The triple phase relationship between reflections.

$$
\Phi_{HK} = \varphi_H + \varphi_K + \varphi_{-H-K}
$$

where $\varphi_H$ is the phase of a reflection $F_{hkl}$, $\varphi_K$ is the phase of a reflection $F_{h'k'l'}$, and $\varphi_{-H-K}$ is the phase of a reflection $F_{(-h-h')(k-k')(l-l')}$. In practise, where $F_H$, $F_K$, and $F_{-H-K}$ correspond to large values, the triplet invariant $\Phi_{HK} \approx 0$, and hence $\varphi_H + \varphi_K + \varphi_{-H-K} = 0$. To determine the remaining phases and to refine the phases found via structure invariants, the tangent formula by Karle and Hauptman (Equation 1.12) is commonly used:

Equation 1.12. The Karle-Hauptman tangent formula for phase extension and refinement.$^{18}$

$$
\tan(\varphi_H) = \frac{-\sum_K |E_{K}E_{-H-K}| \sin(\varphi_K + \varphi_{-H-K})}{\sum_K |E_{K}E_{-H-K}| \cos(\varphi_K + \varphi_{-H-K})}
$$

Modern implementations of direct methods structure solution commonly incorporate powerful real-space constraints and dual-space methods (described in detail by Sheldrick et al.$^{19}$).

1.1.2 Molecular replacement structure solution

One method of phasing that does not require heavy-atom derivative crystals is molecular replacement (MR).$^{20}$ Molecular replacement is a method of phasing that uses a model similar to the unknown target structure to provide phases. The placement of a group of known atoms into an unknown crystal provides initial phases, which can be improved by extension of partial models and phase improvement by refinement.

Molecular replacement is carried out as either a brute force six-dimension (position and orientation; $x$, $y$, $z$, $\alpha$, $\beta$, $\gamma$) search to place a search model in an unknown crystal, or more commonly as a rotation and translation search based on Patterson methods. Contemporary methods make use of statistically-based maximum-likelihood analyses$^{21-23}$ in combination with advanced rotation and translation functions, similar to the traditional Patterson-based approaches, which are briefly described herein.
The rotation function, introduced by Rossman and Blow in 1962, is a means of determining the relative orientation of identical subunits in a crystal\(^{24}\). The Patterson function of one such crystal will contain:

1. Peaks corresponding to the self-Patterson of one subunit.
2. Peaks corresponding to the self-Patterson of the other subunits, rotated in agreement with their relative orientation.
3. Cross-Patterson peaks corresponding to the vectors between subunits.

Of these, 1 and 2 will be within a space corresponding to the size of the subunit, whereas 3 may have peaks outside of this. Thus, for a Patterson function superposed onto a copy of itself, at some angle of rotation \(\theta_1\), \(\theta_2\), \(\theta_3\), there will be a high level of overlap corresponding to the rotation relating one subunit to another. This overlap is represented by the rotation function \(R\) (Equation 1.13):

\[
R(\theta_1, \theta_2, \theta_3) = \int_U P_1(x) P_2(y) \, dx
\]

where \(P_1(x)\) and \(P_2(y)\) represent two Patterson functions, and \(U\) is a unit volume in both functions\(^{24}\). The original method of Rossman and Blow, which calculated \(R\) by using the Fourier expansions of \(P_1(x)\) and \(P_2(x)\) (Equation 1.7), has since been improved on by Crowther in the fast-rotation function\(^{25}\).
Figure 1.11. Superposition of a rotated Patterson function on itself. A) A dimer, where each monomer is related by rotation and translation to the other; B) The vector set representing the Patterson function of this dimer. Self-Patterson peaks for each monomer are shown as filled circles linked by lines, and cross-Patterson peaks are shown as open circles; C) Superposition of a rotated copy of the Patterson function upon itself. Reproduced from The International Tables for Crystallography Volume F (International Union of Crystallography, http://journals.iucr.org/), Blow, D.M., Chapter 13.1: Noncrystallographic symmetry, 263-268, Copyright (2006).
The translation function $T$ (Equation 1.14) performs a similar role, but in this case is a function of the translation between cross-Patterson peaks:

Equation 1.14. An expression for the translation function. The translation function is represented as a convolution of the cross-Patterson peaks between a reference molecule and a symmetry-related equivalent ($P_{01}(u,t)$) and the Patterson function of the crystal ($P(u)$).

$$T(t) = \int_V P_{01}(u,t)P(u) \, du$$

where $P_{01}(u,t)$ represents the cross-Patterson peaks between a reference molecule at $u$ and a translation of $t$, $P(u)$ represents the Patterson function of the crystal at $u$, and $V$ is the volume of the unit cell.\textsuperscript{26} One early implementation of this is the modified $T_1$ translation function introduced by Crowther and Blow,\textsuperscript{26} which simplifies the search by subtracting the self-Patterson peaks (Equation 1.15).

Equation 1.15. The modified $T_1$ translation function. $F_{\text{obs}}$ represents the structure factors of the search model, $F_{\text{obs}}$ represents the structure factors of the crystal to be searched, $A_i$ is a rotational symmetry operator (expressed as a matrix) relating the $i$th of $n$ molecules in the crystal to the known molecule.

$$T_1(t) = \sum_h \left( |F_{\text{obs}}(h)|^2 - \sum_{i=0}^{n-1} |F_M(hA_i)|^2 \right) \times F_M(h)F_M^*(hA)e^{-2\pi ih\cdot t}$$

1.1.3 Racemic protein crystallography

In nature, proteins exist as chiral species built exclusively from L-amino acids, except for glycine, which is achiral. D-amino acids are, however, sometimes incorporated into biomolecules containing predominantly L-amino acids, e.g. bacterial peptidoglycan layers. Because of their chiral nature, the arrangement of naturally occurring L-proteins into crystals may only occur in 66 of the 230 possible crystallographic space groups which do not contain symmetry operations that would require inversion of their stereochemistry (Figure 1.12).\textsuperscript{27}
Racemic protein crystallography is a method where crystals are grown from racemic mixtures containing equal amounts of L- and D- protein enantiomers to form a racemic crystal. In a racemic crystal, all of the 230 space groups are permitted, as inversion and mirror symmetry operations can be satisfied by the presence of inversion/mirror-related enantiomers. Moreover, the use of racemic crystals is proposed to have an advantage over conventional protein crystallography, as in centrosymmetric space groups the crystallographic phase problem is simplified. Every reflection in a centrosymmetric space group is centric, and is restricted to a phase of 0 or $\pi$. Centrosymmetric space groups are those where a point inversion centre is located at the unit cell origin, and as such can only occur for protein crystals containing racemic mixtures.

Despite the fact that racemic crystallisation has long been observed for small organic molecules, the absence of methods for generating unnatural D- proteins has meant that prior to the report of racemic rubredoxin in 1993, there had been no examples of racemic protein crystallisation. It is only in recent decades that developments in the chemical synthesis of peptides and small proteins have enabled the routine preparation of the corresponding D- enantiomers for crystallography (discussed in §1.2).

In addition to simplifying the phase problem, racemic crystallography appears to have benefits for protein crystallisation, a major limiting factor in protein crystallography. In 1995, Wukovitz and Yeates proposed a statistical model to explain the choice of space group in protein crystals. Crystallisation of small organic molecules favours the formation of close packing contacts between molecules, and Wilson’s model, based on Kitajgoroskij’s categories, provides a satisfactory explanation of the choice of space group for a
molecule. In contrast, protein crystals favour maintaining a large solvent content over formation of close contacts, with crystals rarely deviating from a solvent content of 40-60% by volume. The statistical model presented by Wukovitz and Yeates explains protein crystallisation based on the statistical likelihood of fulfilling the symmetry requirements imposed by space groups, represented by rigid-body degrees of freedom. Importantly, it was predicted that the centrosymmetric space group \( P1\bar{1} \), which cannot be accessed by the \( L \)-protein alone, has the greatest degrees of freedom and thus the highest likelihood of occurring for proteins (Figure 1.13). By this model, the centrosymmetric space groups \( P2_1/c \) and \( C2/c \) have also been predicted to show significant likelihoods of occurring.

Since the first case of racemic protein crystallography by Zawadzke and Berg, many further examples of this technique have been reported, including a glycoprotein, and a complex of a homodimeric VEGF construct with a peptide antagonist. These studies have supported Wukovitz and Yeates’ predictions by showing: i) multiple cases where crystallisation was successful from racemic mixtures but not from a single enantiomer alone; and ii) that from 11 cases of racemic protein crystals, six were in space group \( P1\bar{1} \).
1.2 Chemical synthesis of peptides and proteins

Contemporary methods to chemically synthesise peptides allows for the facile preparation of peptides up to 50 residues long. Moreover, through the application of ligation methods, the preparation of small proteins > 100 residues has become a matter of routine. The most significant advantage of chemical synthesis, over other methods, is the ability to incorporate unnatural amino acids or to undertake difficult chemical modifications. Site-specific phosphorylation, glycosylation, replacement of disulfide bonds, incorporation of fluorophores, and the inclusion of click chemistry handles, are all commonly and readily achieved by chemical synthesis. The tools by which these syntheses are undertaken, namely solid phase peptide synthesis (SPPS) and native chemical ligation (NCL), are reviewed herein.

1.2.1 Solid phase peptide synthesis

Solid phase peptide synthesis (SPPS) is a technique of synthesising a peptide through the addition of protected amino acids to a growing peptide chain immobilized on a resin solid support. Repetitive coupling cycles are employed, consisting of a coupling step and a deprotection step. In the coupling step, an activated amino acid containing a protected N-terminal amine is coupled to the free amine on the peptidyl resin (Scheme 1.1A). The protected N-terminal amine is deprotected, whereby the protecting group is removed to unmask the free amine (Scheme 1.1B). These two reactions are repeated until the desired peptide is assembled on the resin.

![Scheme 1.1. General reactions for on-resin peptide elongation (SPPS). A: Coupling of a resin-immobilised free amine to a protected, activated amino acid ester. B: Removal of amine protecting group to reveal a reactive free amine.](image)

In-between each coupling and deprotection step, the resin is washed to remove excess reagents and to remove non-bound by-products that would otherwise interfere with the subsequent step. To ensure non-specific couplings to do not take place, all reactive amino acid side-chains are usually protected during synthesis (Albericio and co-workers have...
comprehensively reviewed common side-chain protecting groups used\textsuperscript{39}). Once the peptide is fully assembled on the resin, concomitant cleavage of the linker immobilising the peptide to the resin and any side chain protecting groups affords the crude peptide. A general scheme is shown below (Scheme 1.2).

These methods have almost completely eclipsed classical solution phase synthesis methods, the two primary reasons for this being: \textit{a}) the difficult purification of protected peptide fragments which display limited solubility; and \textit{b}) the inefficiency of multiple purification steps for each coupling and deprotection cycle (for an in-depth discussion see Kent\textsuperscript{40}). In contrast, SPPS allows the peptide to remain soluble \textit{via} anchoring to the insoluble solid support, and purification of the resin bound intermediates is by a simple washing step. Two predominant protecting group strategies exist: the Boc/Bzl strategy\textsuperscript{41} (Boc = tert-butyloxycarbonyl) and the more widespread Fmoc/O\textsubscript{t}Bu strategy\textsuperscript{42-44} (Fmoc = 9-fluorenylmethyl-oxycarbonyl; Figure 1.14).

The Fmoc/O\textsubscript{t}Bu strategy uses ‘orthogonal’\textsuperscript{45, 46} Fmoc (base labile) and tBu-based (acid labile) groups for the respective protection of the amine and reactive side-chain groups, allowing selective removal of one over the other. For the Boc/Bzl protecting group strategy,
the Boc (amine) and benzyl-based (side chain) masking motifs are not truly orthogonal, but differ in their degree of acid lability. These differences in protecting group lability for the two strategies have consequences for the reagents used for amine deprotection, side chain deprotection and resin cleavage. In Fmoc-SPPS, the removal of the N-Fmoc group is normally carried out using a solution of the 2° amine piperidine, while the cleavage of side-chain protecting groups and release of peptide from the resin uses a mild acid, TFA. In contrast, as the protecting groups are not orthogonal in Boc-SPPS, cleavage of the Boc group is carried out with TFA, and the release of peptide from resin uses anhydrous HF, a highly toxic and corrosive acid.

The harsh conditions required for Boc-SPPS entail a number of limitations. Glycosylation of proteins is a common post-translational modification, but due to the HF-lability of the glycosidic bond, glycosylated peptides cannot be prepared via this strategy. Automation is also difficult due to the highly corrosive nature of TFA and HF, unlike Fmoc-SPPS which is routinely automated. Most obviously, the anhydrous HF required for the final cleavage step can cause difficulties due to practical considerations, such as strict safety requirements that require the use of specialised apparatus, supply issues due to country restrictions, and a general desire to avoid using such a hazardous reagent. Despite this, Boc-SPPS can be advantageous in several areas: α-peptidethioesters, the building blocks for native chemical ligation (§1.2.2) used in protein synthesis can be prepared directly on resin (§1.2.4), and in situ neutralisation methods can be easily employed to overcome difficult syntheses.

It is well-known that certain peptides, such as the model acyl carrier protein (ACP) [65-74, Val-Gln-Ala-Ala-Ile-Asp-Tyr-Ile-Asn-Gly], are inherently difficult to synthesise. So called “difficult syntheses” are challenging with both Fmoc and Boc strategies, and appear to be a sequence-dependent effect. The cause of this is believed to be the formation of secondary structure within the growing resin-bound peptides, reducing the accessibility of the reactive amine to coupling with activated amino acids during synthesis. To overcome the difficulties in these syntheses, both Fmoc and Boc strategies have undergone extensive modifications. A particularly useful mode of Boc-SPPS, named in situ neutralisation, has been developed, whereby the amine neutralisation step following a Boc deprotection is replaced with a combined in situ neutralisation/coupling step with excess base (Scheme 1.3).
The development of in situ neutralisation stemmed from observations suggesting that TFA treatment of peptides is able to disrupt aggregation, namely that: i) Boc deprotection is always quantitative compared to the Fmoc deprotection steps, which can be difficult if the peptide is aggregating; and ii) the protonated peptides show greater swelling properties.\textsuperscript{53, 54} These in situ neutralisation methods for Boc-SPPS have been found to improve the syntheses of numerous difficult sequences including that of the model peptide ACP [65-74].\textsuperscript{54} In addition to enhancing the success of difficult couplings, the omission of a separate neutralisation and washing step has the benefit of using less solvent and allowing for shorter coupling cycles.\textsuperscript{55}

In situ neutralisation methods are more difficult to implement in Fmoc-SPPS, as this requires the addition of a separate protonation step following base-mediated Fmoc removal.\textsuperscript{53} In addition, the acid-labile nature of the linkers and side-chain protecting groups normally employed in Fmoc-SPPS precludes the use of strong acids such as TFA. Alewood
et al. have shown that it is possible to perform a mild protonation using the Fmoc-protected amino acids prior to coupling.\textsuperscript{55} This is carried out simply by modifying the order in which reagents are added, with addition of Fmoc-AA/HBTU/DMF to the aminopeptidylresin 5 min before activation with iPr\textsubscript{2}NEt.\textsuperscript{55} Although reported to be useful in general for difficult couplings, it was also mentioned that protonation could be ineffective for sterically hindered Fmoc-AA and lead to guanidino-terminated products when using uronium-aminium coupling reagents like HBTU (§1.2.5).\textsuperscript{56}

Regardless of the synthetic strategy and methods employed, there is an inherent limit to SPPS. It is generally accepted that the synthesis of peptides longer than \(~\)50 residues would be extremely difficult, if not impossible. As the peptide is elongated, the accumulation of undesired side products for each coupling step, generated from incomplete couplings or deprotections, racemisation, and other low level side-products, results in an increasingly complex crude product.\textsuperscript{57} Exceptions to this general rule do exist, however, and advances in microwave-assisted SPPS methods have enabled the successful preparation of a 109-residue exon 1 huntingtin peptide (microwave-assisted SPPS has been reviewed by Jensen and co-workers\textsuperscript{58}).\textsuperscript{59} Another limiting factor for the preparation of long peptides is that high-performance liquid chromatography (HPLC) techniques used for peptide purification suffer from decreased resolution as peptide length increases.\textsuperscript{60}

As an excellent illustration of this general limitation in SPPS, Johnson and Kent reported an exploratory study to synthesise the 372-residue δ-opioid receptor, a G-protein coupled receptor, by in situ neutralisation Boc-SPPS and NCL methods. A synthetic strategy employing seven fragments of consisting of 37-63 residues each was undertaken.\textsuperscript{60} For a number of these peptides, the progression from a 30-residue peptide to the full length fragment showed significant deterioration in the quality of product. The authors ascribed these observations to reaching this limit of SPPS.

1.2.2 Native chemical ligation

Native chemical ligation (NCL) was developed in 1994 by Dawson et al., and circumvents the problems associated with the preparation of long peptides by linear SPPS. The key aspect to NCL is the reaction of a peptide with an N-terminal cysteine to one with a C-terminal thioester to form a native amide bond.\textsuperscript{61} The reaction is performed most commonly in aqueous conditions buffered to neutral pH, as this avoids the problems of cross reactivity of the thioester with basic functionalities at high pH, and the reduced reactivity of the
cysteine thiol moiety at low pH. The highly selective nature of the reaction at neutral pH provides the added benefit of enabling fully unprotected peptides to be used, avoiding the problems of solubility, aggregation, and purification often associated with protected peptides. This reaction has also been shown to be successful in organic solvents such as DMSO and DMF, thereby extending the applicability of NCL to hydrophobic peptides.

NCL is initiated by the transthioesterification of the activated C-terminal thioester with the sulfhydryl group of the N-terminal cysteine of the other fragment. A fast, spontaneous, irreversible $S \rightarrow N$ shift then occurs within the bridged intermediate to form the desired amide linkage (Scheme 1.4).

Scheme 1.4. Proposed mechanism of native chemical ligation.

Internal cysteine residues within the peptide fragments do not need to be masked, as although they may participate in the formation of thioester/thiolactone species, without a free proximal amine present the formation of a stable amide bond cannot occur. The addition of excess thiol can help drive completion of the reaction by reducing disulfides, thiolysing non-productive thiolactones/thioesters, as well as fulfilling a critical role in accelerating ligations by forming more reactive thioester species through thioesterification.
Johnson and Kent showed that the transthioesterification of the peptide thioester with the cysteinyl peptide is the key rate limiting step in native chemical ligation (Scheme 1.4). To improve the speed of reaction by providing a more reactive species, an exogenous thiol is introduced, which undergoes transthioesterification with the peptide α-thioester. At pH 7, the most often used ligation condition, addition of a thiol additive with a pKa > 6.5 results in near quantitative thiol exchange with the existing thioester moiety thereby promoting the ligation. Aryl thiols have been deemed much better catalysts for NCL than alkyl thiols as they have superior leaving group abilities, but are in most cases extremely insoluble in aqueous solution. The most commonly used thiol additives are depicted in Figure 1.15. Thiophenol is sparingly soluble in water, and although sodium 2-mercaptoethansulfonate (MESNa) possesses more favourable solubility properties, ligation is much slower. 4-mercaptoacetic acid (MPAA), an aryl thiol, is water-soluble and is used widely in NCL. MPAA has been shown to promote otherwise sluggish ligation reactions and has now superseded other thiols as the reagent of choice for native chemical ligation.

Aside from the thiol moiety of the thioester, an additional consideration in the rate of ligation is the C-terminal residue in the thioester polypeptide. In general, sterically demanding side-chain branched residues (e.g. Ile, Val) result in prolonged reaction times, whereas sterically unencumbered residues (e.g. Gly, Ala) result in the shortest ones. This has been shown in a study by the Dawson et al. using model peptide α-thioesters where only the C-terminal residue was varied (Figure 1.16). Cysteine and histidine thioesters were also shown to react at an equal rate as that of a glycine thioester and faster than an alanine thioester, suggesting that the side chain functionalities of cysteine and histidine may play a role in the rate-limiting transthioesterification (Figure 1.16).
1.2.3 NCL at sites without cysteine

The presence of a cysteine residue in a protein at a site suitable for ligation is not always possible. Cysteine is one of the more uncommon amino acids found in proteins, and is observed to be present in a 1-2% relative abundance in eukaryotes and in about a 1% relative abundance in prokaryotes (Figure 1.17).\footnote{For Humans, Drosophila, \textit{Saccharomyces cerevisiae}, and \textit{Escherichia coli}, it is predicted that 8-20% of their protein sequences do not contain any cysteine residues at all (Figure 1.17). Furthermore, the presence of a cysteine in a protein does not guarantee that residue is positioned such that it can be effectively utilised for NCL.} For Humans, Drosophila, \textit{Saccharomyces cerevisiae}, and \textit{Escherichia coli}, it is predicted that 8-20% of their protein sequences do not contain any cysteine residues at all (Figure 1.17). Furthermore, the presence of a cysteine in a protein does not guarantee that residue is positioned such that it can be effectively utilised for NCL.
Thus, since NCL was first reported, a number of methods have been investigated to circumvent the requirement for a cysteine at the ligation site. These efforts have focused primarily on the transformation of the ligation site cysteine to other native amino acid residues via desulfurisation or chemical modifications via alkylation. This allows the thiol to be utilised as the ligation handle.
1.2.3.1 Reductive desulfurisation of thiol-derived amino acids

Desulfurisation of cysteine to alanine in a separate post-ligation step is one of the most exploited methodologies (Scheme 1.5). The application of this method to the desulfurisation of ligation-site cysteines was first reported by Dawson, who used a Pd or Raney Ni-catalysed reaction. Alternative conditions using a radical process have recently been disclosed and have been extended to a one-pot ligation-desulfurisation process.

This desulfurisation approach has been successfully extended to general amino acid residues in order to obtain residues other than alanine (Scheme 1.5). At present, chemically modified residues such as Arg, Asp, Glu, Gln, Leu, Lys, Phe, Pro, Thr, Trp, and Val can be used in place of the cysteine at the ligation site and then be desulfurised to form the native amino acid (Figure 1.18). All of these amino acid derivatives must be prepared via multistep synthesis, with exception of commercially available penicillamine ((2S)-2-amino-3-methyl-3-sulfanyl-butanoic acid), selenocysteine (Sec), and γ-thiol proline, and in the case of δ-thiol tryptophan, where the thiol can be directly installed onto tryptophan during SPPS (Scheme 1.6).
Figure 1.18. Amino acids used to generate the corresponding native residue at the ligation site in ligation-desulfurisation strategies. Penicillamine is shown in red, and selenocysteine in blue.
Scheme 1.6. Preparation of a $\delta$-thiol tryptophan peptide by Fmoc-SPPS. Synthetic steps: i) installation of rink amide linker; ii) coupling of first coupling residue; iii) Fmoc-SPPS; iv) sulfenylation: DNPS-Cl, AcOH, DMF, r.t. 2 h; v) peptide cleavage: TFA/iPr3SiH/H2O (90/5/5 v/v/v), r.t. 2 h; vi) thiolysis: thiophenol, 6 M GnHCl, 0.1 M Na2HPO4, pH = 8.0, r.t. 3 h.

The generality of the NCL reaction is illustrated by the fact that many of the amino acids in Figure 1.18 can be successfully used, despite the increased distance between the thiol and the $N_\alpha$-amine in amino acids containing a $\beta$, $\gamma$, or $\delta$-thiol. Each amino acid undergoes ligation/desulfurisation satisfactorily, but in the case of penicillamine, which is desulfurised to valine, the presence of the $\beta$-dimethyl groups significantly retards the ligation reaction.86, 87

In addition, selenocysteine, which contains a selenium atom in place of the sulfur in cysteine, can also be used in NCL.88-91 Selenocysteine is more nucleophilic than sulfur, and ligations using selenocysteine have been reported to occur at a rate 10$^3$ times greater than with cysteine at pH 5, thus allowing a method for the chemoselective ligation of a selenocysteine peptide.89 Notably, NCL with selenocysteine has been reported to occur slowly or not at all when carried out in the absence of a reducing agent.88, 90 This is a result of the low pKa and hence propensity of selenol groups to form diselenide or selenosulfide
bonds, even under acidic conditions, and the difficulty in reducing these to form the active free selenol.\textsuperscript{90}

One major advantage of using selenocysteine at the ligation site is that selenocysteine can rapidly undergo chemoselective deselenisation to form alanine in the presence of free thiols.\textsuperscript{92} The selenol in selenocysteine can form a radical much more easily than thiols, and in the presence of TCEP alone is converted to a native alanine (Scheme 1.7).\textsuperscript{92, 93} This enables the use of ligation/deselenisation strategies for peptides containing disulfides or free cysteines.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{selenocysteine_deselenisation.png}
\caption{Scheme 1.7. Reductive deselenisation of selenocysteine. Deselenisation of Sec requires only TCEP, compared to TCEP and a radical initiator in the case of thiol-derived amino acids.}
\end{figure}

\subsection*{1.2.3.2 Alkylation of cysteine}

Another often exploited post-ligation transformation is the S-alkylation of the ligation site cysteine to form analogues of the naturally occurring amino acids that retain the side-chain functionality, albeit a carbon homologue in the case of pseudo-Glu/Gln (Scheme 1.8). The alkylation is highly selective for cysteine as no other reactive functional side chains such as amines, amides or hydroxyl are affected. Pseudo-glutamate/glutamine residues have been prepared at ligation site cysteines by alkylation with bromoacetic acid or bromoacetamide, respectively.\textsuperscript{94, 95} Modification of the two cysteine residues by alkylation with iodoacetamide to afford pseudo-glutamic acid in an analogue of erythropoietin did not appear to have any effect on biological activity, and a noted increase in \textit{in vivo} activity was attributed to other modifications.\textsuperscript{94} An analogue of lysine, pseudo-lysines can be formed from the alkylation of cysteine with bromoethylamine, and have restored activity to proteins where essential lysines were mutated to cysteines.\textsuperscript{96} Alkylation of homocysteine in a post-ligation manner has also been applied to generate a native methionine residue. Thus, homocysteine, installed to enable the ligation, was subsequently selectively S-methylated to afford methionine (Scheme 1.8).\textsuperscript{97}
A method for generating a serine at a ligation-site cysteine has also been reported by Okamoto and Kajihara. Following NCL, the ensuing cysteine is first S-methylated, then undergoes an intramolecular rearrangement after activation with CNBr, followed by an $O\rightarrow N$ acyl shift affording the native serine with full retention of configuration (Scheme 1.9). Activation of the methylated cysteine with CNBr in formic acid can cause the undesired formylation of unprotected hydroxyl groups, and an additional step is required to remove these. The yield of conversion for this method appears to show some sequence dependence.
1.2.4 Preparation of peptide α-thioesters

To enable NCL, two mutually reactive polypeptide partners are required. N-terminal cysteine peptides can be prepared routinely by either Fmoc or Boc-SPPS, but polypeptides containing the sensitive C-terminal α-thioester require a more robust strategy. Using the Boc/Bzl strategy, peptide α-thioesters can conveniently be prepared by SPPS, which delivers the thioester polypeptide 1.4 directly following cleavage from the resin (Scheme 1.10)
Thus, a carboxylic acid containing a masked sulphydryl group is immobilised to a solid support via the cleavable PAM linker. After the removal of protecting group of the thiol, the C-terminal residue of the desired peptide can be directly coupled to afford the thioester and the remaining residues assembled using standard methods (Scheme 1.10).48 This method can be shortened by omission of the HF-labile linker, enabling the NCL of resin-bound unprotected peptide thioesters.99

Preparing α-thioester peptides on resin cannot be directly undertaken by Fmoc-SPPS, as the thioester functionality is sensitive to the nucleophilic conditions employed in Fmoc removal. However, the popularity and widespread acceptance of Fmoc-SPPS has resulted in the development of numerous Fmoc-based methods providing an alternative preparation of peptide-α-thioesters. A selection of the most utilised methods will be briefly discussed herein. Detailed descriptions and a comparison of these methods are presented elsewhere.100

1.2.4.1 Alternative Fmoc deprotection reagents

The limiting factor in peptide thioester synthesis by Fmoc-SPPS is the nucleophilic cleavage of the thioester by piperidine, the most common reagent used for Fmoc removal. Thus, alternative methods have focused on the use of modified deprotection conditions. Li et al. employed a deblocking mixture consisting of 1-methylpyrrolidine (25% v/v),
hexamethylene imine (2% v/v), and HOBT (2% w/v) in NMP/DMSO (1/1 v/v), in conjunction with the use of a sterically demanding tertiary thiol, 3-mercapto-3-methylbutanoic acid for the installation of the thioester, thus providing some shielding from nucleophilic attack. Alternatively, Clippingdale et al. used the primary thiol 3-mercaptopropionic acid, employed in Boc-SPPS, and a deblocking mixture of DBU (1% v/v), and 1% HOBT (w/v) in DMF, but found that aminolysis of the thioester could not be completely suppressed early in the synthesis.

1.2.4.2 Late-stage thioester installation

The esterification of the C-terminal carboxylic acid with a thiol is another strategy of preparing peptide thioesters, although this requires the use of maximally protected peptides to avoid unwanted side reactions. The direct coupling of a thiol (Scheme 1.11a) or a thioester containing motif (Scheme 1.11b) to the protected peptide acid can be undertaken in solution, or on resin.

Scheme 1.11. Direct thioesterification of a protected peptide acid via coupling of: a) a thiol; b) a Thr-thioester as an example of a thioester containing motif.

Many methods employ the use of an $O\rightarrow S$ or $N\rightarrow S$ acyl transfers to establish the thioester moiety and will not be described in detail. One example of these methods is shown in Scheme 1.12.

Scheme 1.12. $N\rightarrow S$ shift to generating a peptide $\alpha$-thioester, which can be trapped by an exogenous thiol. This method is not completely selective for AA = His, Cys and Gly, but the $N\rightarrow S$ shift and subsequent thioester trapping occurs much faster for these residues. An example of the conditions employed in this method is shown.
A related method that relies on an $N\rightarrow S$ acyl transfer to generate a peptide thioester from a recombinant protein has been widely used in the literature.\textsuperscript{120} It requires a protein construct that contains a C-terminal intein domain.\textsuperscript{121} During protein splicing, an internal intein domain separating two lateral extein domains is excluded, and the extein domains joined. This process progresses through an intramolecular transthioesterification ($N\rightarrow S$ acyl transfer) at the C-terminus of the N-extein (Scheme 1.13).\textsuperscript{121} By expressing a mutant splicing protein, which is unable to undergo transthioesterification between the N- and C-extein regions to ultimately form the ligated extein, the N-extein is trapped as a thioester.\textsuperscript{122} Thus, the thioester can then be intercepted via thiol-thioester exchange with excess thiol (Scheme 1.13). The generation of a desired recombinant protein thioester and subsequent one-pot native chemical ligation with an added N-terminal cysteine peptide has been termed \textit{expressed protein ligation}.\textsuperscript{122}
Scheme 1.13. Protein splicing in a non-functional splicing protein. For a non-functional protein where intramolecular transthioesterification (or esterification) cannot occur, the N-extein thioester can be trapped via an exogenous thiol-thioester exchange.

A further recently disclosed method describes the oxidation of an unprotected C-terminal peptide hydrazide to the azide, followed by substitution with an added thiol to generate the C-terminal thioester (Scheme 1.14). In an extension to this, a phenyl thioester can also be generated in situ prior to native chemical ligation in a one-pot manner. The primary advantage to this methodology is that peptide hydrazides can be readily prepared by standard Fmoc-SPPS methods.
Specialised SPPS linkers, which can be activated to provide peptide thioesters, have also been proven to be reliable ways to generate C-terminal thioester peptides using Fmoc-SPPS. Pessi and co-workers introduced a modified safety-catch linker which immobilises peptide to resin via a sulfonamide linkage.\(^{125}\) Once the on-resin assembly of the peptide is complete, the sulfonamide can be activated by alkylation of the nitrogen, and the peptide thioester obtained by thiolysis, although a separate side chain deprotection step is then required (Scheme 1.15). Furthermore, the acylation of the unreactive sulfonamide linker with the first amino acid is non-trivial and requires specialised reaction conditions to avoid racemisation.

Blanco-Canosa and Dawson have reported an alternative safety-catch linker, based on 3,4-diaminobenzoic acid (Dbz) that can be immobilised to the resin via a standard TFA-labile linker.\(^{126}\) The 3,4-diaminoaryl linker is activated following chain assembly to generate an N-acylurea (Nbz) peptide which is stable to TFA cleavage, HPLC purification, and general handling conditions (Scheme 1.16). At neutral pH, the acylurea moiety can rapidly undergo irreversible thiolysis to generate the peptide thioester, although this is normally performed in situ during the ligation reaction.
In syntheses using highly basic conditions, or those of Gly-rich sequences, the free amine of the 3,4-diaminoaryl linker can undergo acylation to form a branched peptide that cannot be cyclised to generate the Nbz peptide.\textsuperscript{127} Recently, a second-generation linker with a methylated amine was reported by Blanco-Canosa and Dawson, which does not show the undesired acylation (Scheme 1.16).\textsuperscript{128} An alternative solution to this problem was also presented by Ottesen and co-workers, employing reversible Alloc protection of the free amine (Scheme 1.17).\textsuperscript{127}

![Scheme 1.16. Preparation of peptide aryl-thioesters using Nbz linkers. Synthetic steps: i) coupling of Fmoc-Dbz to a resin-immobilised Fmoc-SPPS linker; ii) Fmoc-SPPS; iii) acylation with \(p\)-nitrophenylchloroformate followed by intramolecular cyclisation; iv) cleavage of the Rink linker and thiolysis of the Nbz linker with an aromatic thiol.](image-url)
Chapter 1

Scheme 1.17. Preparation of peptide aryl-thioesters using the first-generation Nbz linker with reversible Alloc protection.

Synthetic steps: i) coupling of Fmoc-Dbz to a resin-immobilised Fmoc-SPPS linker; ii) Alloc protection of the free amine; iii) Fmoc-SPPS; iv) Alloc deprotection; v) acylation with p-nitrophenylchloroformate followed by intramolecular cyclisation; vi) cleavage of the Rink linker and thiolysis of the Nbz linker with an aromatic thiol.

Examination of these Fmoc-based methods for generating peptide thioesters shows that in general, they require either multiple steps, the chemical synthesis of specific starting materials, tedious and sometimes unsuccessful linker activation, or involve unsatisfactory yields. Thus, the direct preparation of peptide thioesters via Boc-SPPS is considered superior where the target peptides are amenable to Boc synthesis.

1.2.5 Coupling reagents used in SPPS

As described above, the general process of coupling protected amino acid residues in SPPS involves the acylation of an amine by the activated acid of an N-protected amino acid (Scheme 1.1). There are many coupling reagents available, which are the subject of extensive reviews, thus the most commonly encountered classes, the carbodiimides and the benzotriazole-derived phosphonium, and uronium-aminium coupling reagents, will be described herein.\textsuperscript{129}

The first use of carbodiimides to activate a carboxylic acid for amide coupling was in the mid-1950s, prior to the development of SPPS methods.\textsuperscript{130} Reaction of the carboxylate ion with a protonated carbodiimide produces an O-acylisourea $1.5$ that can then either react
directly with an amine, or in the presence of excess acid form a symmetric anhydride \( \text{1.6} \), which also reacts with the amine to give the amidated product (Scheme 1.18).

The \( O \)-acylisourea can also undergo an irreversible \( O \rightarrow N \) acyl transfer to form an unreactive \( N \)-acylisourea \( \text{1.7} \), thus consuming a full equivalent of the starting material (Scheme 1.19). \( O \)-acylisourea intermediates derived from carbamate esters such as \( N \)-Fmoc or Boc protected amino acids can additionally undergo an intramolecular cyclisation to form a \( 5(4H) \)-oxazolone species \( \text{1.8} \) (Scheme 1.19) with the expulsion of a urea. Under basic conditions, this species can undergo epimerisation at the amino acid side chain prior to aminolysis to form the amide bond, thus affording a racemic product \( \text{1.9} \). This can be a
significant problem in SPPS,\textsuperscript{132} especially when using racemisation prone amino acids such as His and Cys.

Scheme 1.19. Side reactions of O-acylisourea compounds resulting in either formation of an unreactive N-acylisourea 1.7 or racemisation at the chiral centre of an amino acid.

To alleviate the problem of racemisation, additives can be added to suppress the formation of oxazolone 1.8. Examples of these include \(N\)-hydroxybenzotriazoles, such as \(N\)-hydroxybenzotriazole (HOBt) and 1-hydroxy-7-azabenzotriazole (HOAt).\textsuperscript{133} They
intercept the O-acylurea, thus forming a more reactive N-hydroxybenzotriazole ester 1.10 which reacts more readily (Scheme 1.20) with the amine nucleophile. N-hydroxybenzotriazoles also suppress the $O\rightarrow N$ shift by protonating the secondary amine in the O-acylisourea.\textsuperscript{133}

![Scheme 1.20. Reaction of a protonated O-acylisourea with HOAt forming an activated ester 1.10. The protonated secondary amine is shown in red.](image)

The respective phosphonium salts of these benzotriazole reagents are used, not as additives, but as stand-alone coupling agents. 1-Benzotriazolyloxytris(dimethylamino)-phosphonium hexafluorophosphate (BOP) has been used as an amide coupling agent since 1975 (Figure 1.19).\textsuperscript{134} Base-catalysed formation of a reactive ester with BOP occurs through attack of a carboxylate at the positively-charged phosphonium ion, followed by a substitution with the liberated HOBt to form the reactive benzotriazole ester (Scheme 1.21). Aminolysis then affords the amide bond. One drawback to using BOP is the formation of a highly toxic hexamethylphosphoramide (HMPA) by-product 1.11. 1-Benzotriazolyloxytris(pyrrolidino)phosphonium hexafluorophosphate (PyBOP), another phosphonium coupling agent, was developed to avoid the release of toxic HMPA, and is often used in place of BOP (Figure 1.19).\textsuperscript{135, 136} The HOAt equivalent of PyBOP, 7-azabenzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate (PyAOP), has been found to be more effective than the HOBt reagent (Figure 1.19).\textsuperscript{137}

![Figure 1.19. The phosphonium coupling reagents BOP, PyBOP, and PyAOP.](image)
The respective uronium-aminium salts of HOAt and HOBt, $O$-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and $O$-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU; Figure 1.20),\cite{138} are also widely used as stand-alone coupling agents (Scheme 1.22). It was first thought that the salts existed as the uronium form 1.12, analogous to the phosphonium reagents, but it has since been shown the aminium form 1.13 is the thermodynamically favoured product.\cite{139} Preparation of the more active uronium form 1.12 is possible, but in the presence of base this species rapidly undergoes isomerisation to the aminium species 1.13.\cite{140}

![Diagram](image-url)

**Figure 1.20.** Uronium 1.12 and aminium 1.13 forms of HATU and HBTU.
Scheme 1.22. Activation of a carboxylic acid with uronium and aminium forms of HATU, forming an activated ester.

The base-catalysed activation of carboxylic acids with these reagents occurs analogous to the phosphonium ions, with electrophilic attack at a positively-charged carbon (Scheme 1.22). The activated O-ester is rapidly converted to the N-form, and so coupling using these reagents should employ only a short preactivation period of 30 s – 1 min. Furthermore, in the presence of excess coupling reagent the amino group can react directly, leading to a tetramethylguanidine capping event that effectively blocks the N-terminus and terminates the synthesis. To avoid this, it is important to use a slight excess of the carboxylic acid to the coupling reagent. The phosphonium salts mentioned above do not suffer from this limitation as they do not react directly with the N-terminus.

1.3 Protein (re)folding

Protein (re)folding is a poorly understood process, and the diverse nature of proteins means that (re)folding conditions must be determined empirically. Most proteins exist in a specific three-dimensional conformation that is directly related to their function, so for synthetic and expressed proteins it is often a requirement to ‘(re)fold’ them into their native
conformation before any characterisation can be performed. (Re)folding typically involves
the dilution of a protein that has been prepared in denaturing conditions, e.g. in 8 M urea
or 6 M guanidine hydrochloride, into a folding buffer followed by determination of protein
foldedness. Dilution of a denatured protein stock with a folding buffer allows the protein
to adopt a favoured conformation while avoiding the problems of preformed protein
aggregates and misfolded species. Dilution can be carried out by straight addition of the
diluent (rapid dilution), or by slow dialysis methods. One method commonly used for the
qualitative determination of protein foldedness is circular dichroism spectroscopy,
discussed further below.

The problem of protein (re)folding can be simplified by the presence of native
intramolecular disulfide bridges, which stabilise the native protein conformation. This is
most relevant in the case of proteins amenable to chemical synthesis and with a known
disulfide arrangement. Numerous methods are available for regioselective disulfide
formation, and protein fold stabilisation by the correct disulfide arrangement would in
theory negate the need for comprehensive buffer optimisation prior to crystallisation
screening. For proteins in which the native disulfide arrangement is not known, refolding
conditions will still need to be empirically determined.

1.3.1 Circular dichroism spectroscopy
Circular dichroism (CD) spectroscopy is a semi-quantitative means of determining protein
foldedness. A beam of light that is polarised such that its electric field oscillates in a single
plane can be considered as the resultant of two vectors, one tracing a circle in a clockwise
direction (E_R) and one anticlockwise (E_L) in the direction of the beam. These two vectors
represent circularly polarised light, and the unequal absorption of the two is defined as
circular dichroism. Asymmetric molecules can interact differently with the E_R and E_L,
giving rise to CD spectra.

The circular dichroism spectra of different secondary structure elements are distinct, and
can be a means of identifying what elements a protein contains or if it contains any at all. It is generally accepted that the CD spectrum of a protein is the linear combination of
the contribution from individual secondary structure elements (Figure 1.21).
1.4 Scope of this thesis

Since the development of native chemical ligation methods, numerous protein molecules have been chemically prepared. With the introduction of racemic protein crystallography, chemical synthesis has become vital as the sole method to prepare the required mirror image D- proteins. To date, the technique of racemic protein crystallography has been pioneered by the Kent group. Almost all of the examples in the literature have applied racemic protein crystallography to proteins for which the structure has previously been determined, with the exception of snow flea antifreeze protein, omwaprin, and Rv1738.

This thesis explores the general utility of racemic protein crystallography to the *de novo* structure determination of biologically relevant proteins for which no protein structures are currently available. The application of racemic protein crystallography to the structural determination of a small antimicrobial protein from potato, snakin-1, is described herein.
Chapter 2

Preparation and Crystallographic Investigation of Racemic Snakin-1
2.1 Overview

This chapter describes the preparation of L- and D-snakin-1, and the three-dimensional structure determination of snakin-1 by racemic X-ray crystallography. It is divided into four major sections: the biological relevance of snakin-1 (§2.2), the synthesis of L-snakin-1 (§2.3) and D-snakin-1 (§2.4), and racemic crystallography of snakin-1 (§2.5).

2.2 Antimicrobial Peptides

In recent years, bacterial resistance to antibiotics has become increasingly common worldwide. In early 2014, the World Health Organisation (WHO) released a comprehensive report on the current state of antimicrobial resistance, citing multiple cases of widespread resistance to current treatment regimes.\textsuperscript{163} Examples of the most concerning multi-drug resistant bacteria include strains of \textit{Escherichia coli}, which cause urinary tract infections and bloodstream infections, and \textit{Klebsiella pneumoniae}, which can cause pneumonia.\textsuperscript{164, 165} Both bacteria are commonly encountered as healthcare-related and community-acquired infections. Most concerning is their increasing resistance to current antimicrobial treatments, requiring the use of last-resort antibiotics, such as carbapenems.\textsuperscript{166} Increasing resistance to even these has been observed in \textit{K. pneumoniae} worldwide.\textsuperscript{167} Unless new antimicrobial agents are developed, there is a real risk that in the foreseeable future, commonplace bacterial infections like these will become uncontrollable. In line with this, the WHO has stated that the search for new, effective antimicrobials is of the highest priority.

One group of promising leads for new antimicrobials is that of antimicrobial peptides (AMPs).\textsuperscript{168} Antimicrobial peptides are widely distributed throughout animal and plant species, and are typically cationic, though a small proportion (\textit{ca.} 10\%) has been found to be anionic.\textsuperscript{168} The mechanisms by which AMPs act vary, and although many are supposed to act by disruption of the plasma membrane resulting in cellular leakage, there have been multiple reports of AMPs which target the interior of target cells as well as others that cause effects through signalling events in the extracellular environment.

Compared to the plasma membranes of animal and plant cells, which are typically neutral in charge, those of bacteria and fungi are negatively charged due to the presence of phospholipids with anionic head motifs. Cationic AMPs are thought to selectively target these negatively charged membranes through electrostatic interactions. A number of AMPs
have also been shown to target cancer cells, in which the diseased state results in the accumulation of anionic lipids at the plasma membrane.\textsuperscript{169}

AMPs are typically classified by their structure into one of three categories\textsuperscript{170} (Figure 2.1):

1) $\alpha$-helical peptides
2) $\beta$-sheet peptides
3) extended peptides

Figure 2.1. Major structural classes of AMPs. \textit{A}: $\alpha$-helical peptides; \textit{B}: $\beta$-sheet peptides; \textit{C}: extended peptides. Positively-charged side chains are in blue, negatively-charged ones in red, and other ones in grey. Reprinted from Trends in Biotechnology, 29, Nguyen, L.T.; Haney, E. F.; Vogel, H. J., The expanding scope of antimicrobial peptide structures and their modes of action, 464-472, Copyright (2011), with permission from Elsevier.

$\alpha$-helical AMPs are some of the most well studied AMPs, and include magainin and LL-37. The helices of these peptides are generally amphiphilic, with a hydrophobic face that can bind and disrupt target membranes.\textsuperscript{170} $\beta$-sheet AMPs act by inserting into membranes and forming toroidal pores, one example of which is Protegrin 1. Individual molecules of these AMPs aggregate through formation of hydrogen bonds between $\beta$-strands, resulting in a transmembrane $\beta$-barrel that functions as a pore.\textsuperscript{171} Peptides of the third category, extended AMPs, do not take on any regular secondary structure, and often contain high
numbers of Arg, Trp, or Pro residues. These peptides are not usually membrane disrupting, and tend to have intracellular targets.\textsuperscript{170}

One group of AMPs that is not included in these major classifications is that of the cysteine-rich AMPs found in plants. These peptides act via diverse modes of action and form highly constrained structures, which can contain both $\alpha$-helix and $\beta$-sheet structures. This class of AMPs will be discussed herein.

2.2.1 Cysteine-rich AMPs in plants

Cysteine-rich antimicrobial peptides (henceforth referred to as crAMPs), which contain specific disulfide-rich configurations, make up a large proportion of the currently known plant antimicrobial peptides.\textsuperscript{172} The disulfide connectivity of crAMPs can be used to classify them, as different classes have distinct, highly-conserved disulfide ‘signatures’ (Figure 2.2). These antimicrobial proteins are primarily active against fungi, but activity against bacteria and insects is not uncommon.\textsuperscript{173} Known classes of crAMPs, each with their own distinct disulfide arrangement, include the plant defensins, thionins, hevein-like and knottin-like peptides, non-specific lipid transfer peptides, and $\alpha$-helical hairpins (reviewed by Odintsova and Egorov\textsuperscript{172}). Compared with classic examples of AMPs like magainin 2,\textsuperscript{174} alamethicin,\textsuperscript{175} and the cecropins,\textsuperscript{176} which are known to disrupt or form pores in target plasma membranes, many crAMPs appear to act by specific binding interactions. Three of the known classes of crAMPs, the plant defensins, thionins, and $\alpha$-helical hairpins, will be briefly discussed.
2.2.1.1 Plant defensins

Plant defensins are found in all plant families, and are expressed in many organs but most significantly in the seed. These proteins are 45-54 residues in length and their structures contain a cysteine-stabilised αβ-motif consisting of an α-helix and a three-stranded β-sheet. Plant defensins are structurally related to the insect and mammalian defensins, with the former lacking the N-terminal β-strand, and the latter missing the α-helix present in the plant proteins. In all of these proteins, the structural elements are held in place by disulfide bonds. Plant defensins are held together by four disulfides, although one subclass from *Petunia hybrida* contains an additional disulfide bond that confers additional stability.
Plant defensins have been found to act on different bacterial cell targets, and there appear to be diverse mechanisms by which they can act. They have been shown to interact with microbial membrane components, such as bacterial lipid II, which is essential in bacterial cell wall formation, and fungal sphingolipids and phospholipids.\textsuperscript{184} Though primarily active against plant bacterial and fungal pathogens, activity against the human fungal pathogens \textit{Candida albicans} and \textit{Candida parapsilosis} has been reported.\textsuperscript{185, 186} Once bound to a target membrane, these proteins can undergo cellular uptake and proceed to bind a specific target within the cell,\textsuperscript{187} or remain on the surface of the cell. Known modes of action of surface-bound plant defensins include the initiation of a signalling cascade leading to programmed cell death, and in bacteria, the irreversible binding of lipid II, inhibiting cell wall biosynthesis.\textsuperscript{184} A process of cell death induced by specific signalling pathways appears to be mediated by the binding of defensins to sphingolipids, but the specific signalling process is unknown.\textsuperscript{184} It has also been shown that MsDef1 binds L-type calcium channels in mammalian cells, and this may be responsible for its antifungal activity.\textsuperscript{188}
2.2.1.2 Thionins

Thionins include the α and β thionins, small, typically basic proteins toxic to a variety of organisms including bacteria and yeasts, found in monocots, eudicots, and rosids. They are generally 45–48 residues in length and are expressed as inactive proproteins with a short ~20 residue N-terminal signalling domain and a ~60 residue C-terminal acidic domain. The signalling domains are excised by co-translational cleavage and the C-terminal domains are cleaved by proteinases in vacuoles, resulting in the accumulation of the active, mature thionins in the vacuolar content. All thionins have a highly conserved structure consisting of an anti-parallel two-helix bundle, and a two-stranded β-sheet. These structural elements are held together tightly by three or four disulfide bridges. The toxic properties of thionins appear to depend on specific residues, such as 13Tyr and 8Trp. Remarkably, a truncated construct of a four-disulfide thionin PpTH, containing only the two-helix bundle (residues 7-32), was found to be completely active. The non-toxic thionin crambin has been structurally characterised extensively, and unlike other members of this family, is neutral and hydrophobic in character, and does not contain 13Tyr.

Figure 2.4. The structure of crambin, a thionin with three disulfides (PDB code: 1ejg).

The mechanism of action of these proteins is not known, but evidence suggests that they act by physical interaction with the plasma membrane, which is followed by rapid cell
NMR studies have confirmed that thionins are capable of binding to phosphate, and this is thought to reflect a specific interaction of the proteins with phospholipid heads in a target membrane. Crambin does not interact with membranes and contains an acidic glutamate at position 23, a position where a charged basic residue is located in most thionins, suggesting that this site may be important for membrane interaction. Moreover, membrane interaction alone may be insufficient for the cell-killing effects of thionins, as reduced Viscotoxin A3 was able to bind membranes, but was unable to cause cell death in vivo.

### 2.2.1.3 α-helical hairpins

The α-helical hairpins are small, hydrophilic proteins containing a single, disulfide-bridged, anti-parallel two-helix bundle. Members of this family have shown antimicrobial, trypsin-inhibiting, and ribosome-inhibiting properties. They are normally ca. 30 residues in length, containing two conserved disulfides, with variable lengths of the loop and the α-helices. Interestingly, the truncated PpTH (7-32) thionin construct mentioned above shows the same arrangement of disulfide-forming cysteines, possibly indicating that the α-helical motif may be the functional determinant of the thionins. Alternatively, it may indicate either a common ancestor or convergent evolution, as mechanistic studies on one α-helical hairpin, EcAMP1, found it to be non-membrane disruptive. The specifics of how these proteins carry out their antimicrobial functions are unknown.
2.2.2 Snakin-1

Snakin-1 (SN1) is a 63 amino acid cysteine-rich protein demonstrated to have antimicrobial activity against fungal and bacterial plant pathogens. It contains 12 cysteines, and mass spectrometric analysis indicates that these are all involved in disulfide bridges. The arrangement of cysteines in the sequence is distinct from those of known crAMPs, and may indicate a novel protein fold.

First isolated from potato (*Solanum tuberosum*), snakin-1 inhibits the growth of a number of fungi and bacterial species (Figure 2.6). Examples include the bacterial species *Clavibacter michiganensis* subsp. *sepedonicus*, and the fungal species *Botrytis cinerea*, *Bipolarius maydis*, *Collectotrichum lagenarium*, and *Aspergillus flavus*. Of these, the first two are potato pathogens while the remaining three are pathogens of other plant species, with the IC₅₀ against each of these in the range of 1-10 μM. Snakin-1 acts synergistically with pseudothionin-1 (PTH1), a plant defensin also produced in potato. Treatment against *Pseudomonas syringae* pv. *syringae* and *C. michiganensis* with a
combination of both produced inhibitory effects greater than the sum effect of both individual antimicrobials. Against *Pseudomonas syringae pv. tabaci* and *B. cinerea*, only an additive effect of the two proteins was observed, and may indicate independent modes of action. In addition, snakin-1 was shown to have an aggregating effect on bacteria, but it is not known if this is related to its antimicrobial activity. All that is known about the mechanism of snakin-1 is that it does not cause leaking or aggregation of artificial liposomes that mimic bacterial membranes, suggesting that it has a specific binding target.

In the potato plant, steady-state expression of the *SN1* gene, which encodes snakin-1, occurs at low levels in the tuber, sepal, stamen, and carpel, and at high levels in the stem, axillary bud, young floral bud, and petal (Figure 2.7). The gene corresponds to an 88-residue transcription product, consisting of a 25-residue N-terminal signal domain and the mature 63-residue protein. Expression of snakin-1 in agroinfiltrated *Nicotinia benthamiana* leaf cells indicates that it is localised to the apoplasm/cell wall, in agreement with the initial isolation of the protein from the cell wall fraction of potato cells. As snakin-1 does
not contain a transmembrane domain, its localisation to this compartment may be due to electrostatic interactions or some specific interaction with proteins/lipids.213

Figure 2.7. Snakin-1 expression in the potato plant. Blue tissues show snakin-1 expression, with darker shades of blue corresponding to greater expression. High snakin-1 expression is found in the axillary buds and the young floral buds, moderate expression in the stem, and low expression in the tissues of the tuber. Reproduced with minor changes from PI-005: Seed Potato Inspection Manual (Canadian Food Inspection Agency; Ottawa, ON), Chapter 3 - Plant Morphology, Copyright (2013) (http://www.inspection.gc.ca/plants/potatoes/guidance-documents/pi-005/chapter-3/eng/1381190037846/1381190038643).

Expression of the gene promoter for snakin-1, PStSN1, in transgenic Arabidopsis thaliana has confirmed that it is active in the vasculature, vegetative and reproductive tissues of the plant.214 The promoter is most active in young plants, with a gradual decrease in activity observed with age, indicative of a development role.214 In addition to this temporal expression profile, induction of the promoter was observed when plants were exposed to a
sudden increase or decrease in temperature and when wounded. No induction was observed in response to treatment with the plant hormones abscisic acid, indol acetic acid, and gibberellic acid, or in response to changes in light, UV exposure, or infection with *P. syringae.*

In support of additional roles for snakin-1 other than in defence mechanisms, a recent study by Nahirñak et al. found that *SNI*-silenced potato plants present with small malformed leaves (Figure 2.8). Inspection of the leaves in the silenced plant lines showed that cells were enlarged but fewer in number, indicating a block in cell division and suggesting a role for snakin-1 in regulating this process. Further investigation also showed that metabolism and cell wall composition in leaves of the silenced plants was altered. The lowered levels of ascorbate, a molecule of importance in cell division, in *SNI*-silenced plants suggests that the role of SN1 in plant development is in regulating ascorbate levels.
Another study by Krügel et al.\textsuperscript{216} on the potato sucrose transporter 1 (StSUT1) found, using a split ubiquitin screen, that it interacts with snakin-1 \textit{in vitro}.\textsuperscript{217} StSUT1 is found in the detergent-resistant membrane fraction of plant cells, and is thought to be located in the plasma membrane of cells in the phloem.\textsuperscript{218} Whether this interaction is real \textit{in vivo} is yet to be revealed. However, ascorbate metabolism in broccoli is affected by sucrose supply and may support a role for SN1 in sucrose regulation.\textsuperscript{219}

Snakin-1 shows a significant degree of sequence similarity with a number of known proteins. Since snakin-1 was discovered, a second snakin protein with similar antimicrobial
properties, snakin-2, has also been found in potato.\textsuperscript{212} Sequence alignment of the mature proteins shows that they share 42\% sequence identity. Snakin-like proteins have also been found in pepper (CaSn; \textit{Capsicum annuum})\textsuperscript{220} and \textit{Zizyphus jujuba} fruits (Snakin-Z; Figure 2.9).\textsuperscript{221} Most significantly, sequence homology with proteins of the gibberellic acid-stimulated in 
Arabidopsis (GASA) family has resulted in the classification of snakin-1 within a wider GASA/snakin superfamily.

Figure 2.9. Sequence alignment of full-length mature snakin-1 and snakin-like proteins.

\subsection{2.2.3 The GASA/snakin family}

Members of the GASA/snakin superfamily all have three common features: an N-terminal signal domain of 18–29 amino acids; a middle, variable region with a large degree of variation among family members; and a C-terminal ‘GASA domain’ of approximately 60 residues, including 12 conserved cysteines (Figure 2.10A).\textsuperscript{222} Members of the family have been found in a number of plant species, including tomatoes,\textsuperscript{223} strawberries,\textsuperscript{224} rice,\textsuperscript{225} petunia,\textsuperscript{226} \textit{Gerbera hybrida},\textsuperscript{227} and \textit{Arabidopsis thaliana}.\textsuperscript{228} Despite their name, up-regulation in response to gibberellic acid is not common to all members of this family. Furthermore, although differences in tissue expression exist for proteins in this superfamily, there appears to be a commonality in their functions, with a number of members involved in cell elongation and division.\textsuperscript{222} The often-observed high levels of expression in development-related tissues, and temporal expression profiles, suggest roles in development for these proteins. It has also been reported that some of these proteins have roles in redox regulation, a property that may extend to all members of this superfamily, as the determinant of this function was found to be the 12 conserved cysteines.\textsuperscript{222, 229}

Of the known members of this family, only the few identified snakins\textsuperscript{208, 212, 220} and snakin-like\textsuperscript{221} proteins have proven defence-related functions. Snakin-2 exhibits an activity profile
similar to snakin-1, but its expression profile is markedly different. High levels of steady-state expression of snakin-2 occur in the tuber, leaf, petal, and carpel, all of which show a low level of snakin-1 expression except for the petal, and may indicate complementary roles.\textsuperscript{212} Unlike snakin-1, the expression of snakin-2 can be induced by infection.\textsuperscript{212} Interestingly, the \textit{in vitro} antimicrobial activity of snakin-2 is inhibited by the presence of KCl and CaCl\textsubscript{2}, but whether this occurs with snakin-1 has yet to be reported.\textsuperscript{208, 212}

Little is known about how GASA/snakin proteins function, and their sequences do not reveal homology with any known motifs. Despite the large number of known members of this superfamily, no structure has been determined for any one of these proteins. A computational study by Porto and Franco\textsuperscript{230} used an \textit{ab initio} method to predict the structure of SN1. This model consists of two $\alpha$-helices formed by the N-terminal half of the protein, with the C-terminal half comprising two large loops joined by a 3_10-helix. The predicted structure shows structural similarities with the thionins and $\alpha$-helical hairpins (Figure 2.10B).

The disulfide connectivity was also predicted by Porto and Franco;\textsuperscript{230} however, a study by Harris et. al.\textsuperscript{231} on biologically-active synthetic SN2 found a different arrangement of disulfides. Mass spectrometry confirmed that all the cysteine residues in this synthetic protein were involved in intramolecular disulfide bonds, but only three could be unequivocally characterised \textit{via} enzymatic digestion of folded SN2 by trypsin or chymotrypsin and mass spectrometry analysis.
Investigation of Racemic Snakin-1

2.2.4 **Aims of present study**

The work presented in the following sections of this chapter describes the preparation of racemic crystals of the crAMP snakin-1 to facilitate structural determination by racemic protein crystallography. A crystal structure would unambiguously identify the arrangement of disulfide bonds in this cysteine-rich molecule and identify regions of secondary structure, enabling rational structure-activity-relationship studies to be undertaken, and providing insight into the function of this molecule.

The aims of this study were to:

1. Prepare L-snakin-1 by chemical synthesis.
2. Carry out oxidative folding to form the native protein.
3. Prepare D-snakin-1 and carry out oxidative folding.
4. Obtain racemic crystals of snakin-1.
5. Solve the crystal structure of snakin-1.
2.3 Synthesis of L-snakin-1

Snakin-1, being 63 residues in length, is approaching the limits for an efficient synthesis by linear SPPS. The synthetic strategy described by Harris et al.\textsuperscript{231} exploits a cysteine residue conveniently located near the midpoint of the protein as a site for native chemical ligation (Scheme 2.1). This strategy requires the preparation of two peptide fragments, a peptide α-thioester fragment of \textsuperscript{1}Gly\textsuperscript{-29}Cys (2.1), and a peptide fragment of \textsuperscript{30}Cys\textsuperscript{-63}Pro (2.2). As outlined in the published synthesis,\textsuperscript{231} whereby fragments were prepared by Boc-SPPS, this synthetic strategy was adopted in the current work.

Scheme 2.1. Native chemical ligation strategy for the synthesis of snakin-1.
2.3.1 Synthesis of snakin-1 peptide fragments 2.1 and 2.2

The synthesis of the thioester fragment, ¹Gly-²⁹Cys-COSR 2.1, was carried out on a 0.2 mmol scale using Boc-SPPS methodology with in situ neutralisation (Scheme 2.2). Coupling of Boc protected amino acids was effected with the coupling reagent HATU employing iPr₂NEt, and TFA was used as the Boc deblocking agent.

\[
\begin{align*}
\text{H}_2\text{N} & \quad \text{O} \\
\text{O} & \quad \text{N} \\
\text{i)} & \quad \text{H}_2\text{N} \quad \text{-} \quad \text{O} \quad \text{-} \quad \text{O} \quad \text{N} \\
\text{ii)} & \quad \text{HS} \quad \text{-} \quad \text{N} \quad \text{-} \quad \text{O} \quad \text{-} \quad \text{O} \quad \text{N} \\
\text{iii)} & \quad \text{BocHN} \quad \text{GSNFCDKLCRCSKAGLRCYGCIC} \\
\text{iv)} & \quad \text{H}_2\text{N} \quad \text{-} \quad \text{GSNFCDKLCRCSKAGLRCYGCIC} \\
\end{align*}
\]

²⁹Cys-thioester 2.1

Scheme 2.2. Synthesis of ¹Gly-²⁹Cys-thioester 2.1 by Boc-SPPS requiring: i) functionalisation of aminomethyl polystyrene resin with Boc-Gly-O-CH₂-phi-CH₂-CO₂H followed by Boc deprotection; ii) coupling of S(Trt)-mercaptopropionic acid followed by trityl removal; iii) in situ neutralisation Boc-SPPS; iv) HF cleavage. PG denotes a side-chain protecting group.

S(Trt)-mercaptopropionic acid was used to provide a thiol and was coupled onto Gly-PAM-aminomethyl polystyrene resin using HATU/iPr₂NEt as coupling reagents. The trityl group was removed with a 95:2.5:2.5 mixture of TFA:TIPS:H₂O, which effectively scavenges the trityl cation, revealing the free thiol. Coupling of the C-terminal cysteine residue then completed the installation of the thioester, and the remainder of the sequence was assembled by standard in situ neutralisation Boc-SPPS methods. The desired peptide was obtained after anhydrous HF cleavage and slow gradient RP-HPLC purification, and the purity of isolated peptide was assessed by analytical RP-HPLC and mass spectrometry. Peptide 2.1 was isolated in 3.1% yield and in 95% purity, and the measured mass corresponded to the expected mass ([M+4H]⁺ m/z calculated 823.2, observed 823.1).
Figure 2.11. RP-HPLC chromatogram of purified L-snakin-1 thioester 2.1. Inset: mass spectrum of the peak in the region of 400-1000 m/z corresponding to the desired thioester 2.1 ([M+4H]+ m/z calculated 823.2, observed 823.1). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).

The synthesis of the cysteinylic fragment, 30Cys-63Pro 2.2, was carried out on a 0.2 mmol scale using Boc-SPPS. As Boc-His(DNP)-OH is used in the synthesis of this fragment, the HF-stable 2,4-dinitrophenyl (DNP) group needed to be removed prior to HF cleavage (Scheme 2.3).39 To prevent migration of the dinitrophenyl protecting group to the N-terminal amine, a well-known side reaction, the final Boc protecting group was retained during DNP removal.234
DNP removal is effected by thiolysis,\textsuperscript{235, 236} and was carried out by treatment of the DMF-swelled peptidoresin with a 10-fold excess of thiophenol overnight.\textsuperscript{237} It has previously been reported that thiolysis of DNP with thiophenol produces a yellow by-product, which presumably is 2,4-dinitrophenyl phenyl sulfide (Scheme 2.4).\textsuperscript{235, 237} This indicative colour change was used in the present work as a qualitative measure of DNP cleavage and complete removal was inferred when a fresh addition of thiophenol did not change colour. Using this measure, a single treatment of the DMF-swelled resin with thiophenol overnight was sufficient to remove the sole DNP group in the cysteiny fragment 2.1.

---

Scheme 2.4. Thiolysis of His(DNP) with thiophenol.
Chapter 2

The final Boc group was then removed and the peptide cleaved under standard HF conditions. Purification of the crude product by slow gradient RP-HPLC\textsuperscript{232,233} afforded the desired peptide 2.2 in a yield of 3.4%, and the molecular weight was confirmed by mass spectrometry ([M+5H]\textsuperscript{5+} m/z calculated 762.7, observed 762.7; 98% pure).

![Figure 2.12. RP-HPLC chromatogram of purified L-snakin-1 cysteinyl peptide 3.2. Inset: mass spectrum of the peak in the region of 400-1000 m/z corresponding to the desired peptide 3.2 ([M+5H]\textsuperscript{5+} m/z calculated 762.7, observed 762.7). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H\textsubscript{2}O; buffer B = 0.1% TFA in acetonitrile).](image)

2.3.2 Native chemical ligation of 2.1 and 2.2

The ligation of 2.1 and 2.2 to afford the full-length, linear SN1 was undertaken using the previously determined conditions\textsuperscript{231} with concentrations of 1 mM for each peptide in a solubilising buffer of 6 M guanidine hydrochloride (GnHCl) and 200 mM Na\textsubscript{2}HPO\textsubscript{4} at pH 7.0. 100 mM MPAA was included as a thiol additive to promote the ligation reaction via transthioesterification, and 20 mM TCEP was added to maintain a reducing environment and reverse premature disulfide formation. The reaction appeared complete after three hours, as no further change was observed by LC-MS, and the ligation product was identified by mass spectrometry (Figure 3.3 C; [M+9H]\textsuperscript{9+} m/z calculated 771.5, observed 771.0)
Investigation of Racemic Snakin-1

Figure 2.13. LC-MS chromatograms for the native chemical ligation of L-snakin-1 thioester 2.1 with cysteiny1 peptide 2.2. Reagents and conditions: 100 mM MPAA, 20 mM TCEP, 6 M GnHCl/0.2 M Na2HPO4, pH 6.9. Chromatograms: A) t = 0 h; B) t = 1 h; C) t = 3 h. No further change was observed after 3 h. At t = 0 h, the thioester 2.1 is unresolved from the MPAA peak, which shows high absorbance. Mass spectrum of: D) the broad peptide peak at t = 3 h (Rt = 14.0-16.0) corresponding to the desired product 2.2 at t = 3 h ([M+9H]9+ m/z calculated 771.5, observed 771.3). LC-MS was carried out using an analytical column (Agilent Zorbax 300-SB C3, 300 Å, 3.0 x 150 mm, 5 µm) with a gradient of 1-61% buffer B over 20 min (buffer A = 0.1% formic acid in H2O; buffer B = 0.1% formic acid in acetonitrile).

The crude product was isolated by desalting on a reversed phase column which conveniently removed excess MPAA and TCEP. Thus, the ligation mixture was loaded onto a semi-preparative RP-HPLC column, the column washed with 5% aq. MeCN + 0.1%
Chapter 2

TFA until UV absorbance at 210 nm remained low (<0.1 AU), and the ligation product eluted with 65% aq. MeCN + 0.1% TFA. After lyophilisation, analysis of 2.3 by RP-HPLC (Figure 3.4) indicated that this material was sufficiently pure to proceed directly to the disulfide formation step.

![Figure 2.14. RP-HPLC chromatogram of crude reduced L-snakin-1 2.3. Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).](image)

2.3.3 Oxidative folding

Oxidative folding to form the native disulfide bonds was undertaken using the previously described conditions, involving rapid dilution into a folding buffer containing a cysteine-cystine redox pair. By using a redox pair, unproductive or mismatched disulfides can be reversed, leaving only the thermodynamically favoured disulfides (Scheme 2.5).

![Scheme 2.5. Mechanism of disulfide formation.](image)
Crude linear SN1 2.3 was dissolved in a folding buffer of 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl, and 0.1 M Tris pH 8.0 at room temperature, achieving a final concentration of 0.1 mg/ml of reduced protein. Monitoring of disulfide formation by LC-MS revealed several unresolved peaks that all corresponded to a mass 12 Da less than the starting material, consistent with the formation of six disulfide bonds from 12 cysteine residues (Figure 2.16).
The peak observed in the LC-MS spectrum showed a consistent mass for folded Sn1 (Figure 2.15), although the elution profile was complex. This raised a question whether the broad peak shape in the HPLC chromatogram and the peak splitting in LC-MS chromatogram was a result of misfolded species, or a phenomenon resulting from the chromatographic conditions.

LC-MS analysis of the fractions collected from the RP-HPLC purification found that unresolved elution was observed over all these pure fractions (data not shown). This is not expected to occur if the fractions corresponded to species containing different disulfide...
arrangements, and so it would seem that the phenomenon is a result of the chromatography conditions.

In order to determine the cause of any possible chromatography-related phenomena, different separation conditions were considered. It has been found that heating TFA-containing chromatographic systems for the separation of peptides enhances the peak profiles for a range of peptides, an improvement that has been attributed to improved sorption-desorption kinetics and a temperature-dependent decrease in dissociation.\textsuperscript{239} Heating of the column to 50 °C for RP HPLC analysis appeared to enhance the peak profile and this was interpreted to mean that a single major species was present (Figure 2.16, bottom). Curiously, heating of the column to 50 °C for LC-MS analysis, which was performed using a formic acid buffer, did not improve the elution.

Although an explanation involving sorption-desorption kinetics may be satisfactory for the broad peak observed with TFA containing buffers, it does not explain the chromatography for LC-MS with formic acid containing buffers, suggesting that this was not the cause of the unresolved elution. Moreover, the two sharp maxima at either end of the broad peak observed with formic acid suggests that two distinct, discrete species are formed under these conditions.

The most obvious difference between formic acid-containing and TFA-containing buffers is that formic acid has a higher $pK_a$ of 3.75, whereas the more acidic TFA has a $pK_a$ of 0.5. This corresponds to pH values of 2.68 and 1.90 for aqueous solutions containing 0.1% $v/v$ of formic acid and TFA, respectively. It is possible that this difference in pH between the buffers is the cause of the observed peak shapes. At these pH values, ionisable groups that would likely be of importance are carboxylic acids, such as those on the side chains of aspartate and glutamate.

Although the respective $pK_a$ values of the acid side chains in the acidic residues aspartate and glutamate are between 3.8-4.1 and 4.1-4.6 in model peptides, it is well known that the equivalent $pK_a$ values in folded proteins can show significant deviation from these.\textsuperscript{240} Notably, it has been found that proteins with pI values of $> 8$ tend to exhibit lower aspartate side chain $pK_a$ values (mean 3.1 ± 0.9).\textsuperscript{240} Snakin-1 is a basic protein with a predicted pI of 8.97, and with three residues of both aspartate and glutamate. It is conceivable that if one or more of these aspartate residues showed a decrease in side chain $pK_a$ in the folded protein,
a population of proteins with non-protonated acid side chains could exist at a pH of 2.68 as in 0.1% v/v aq. formic acid. This subpopulation could form ionic interactions with protonated amines of the 14 arginine and lysine residues and lead to oligomerisation between protein molecules, resulting in multiple peaks in chromatography. In order to test this hypothesis, acid titration or NMR characterisation would need to be undertaken. These were considered outside the scope of this thesis, and so chromatographic conditions employing 0.1% TFA containing solvents at 50 °C were used without further characterisation of this phenomenon.

Under optimised chromatographic conditions employing 0.1% TFA-containing solvents at 50 °C, the reaction appeared complete within 17 hours (Figure 2.17). Purification of the oxidised protein was carried out in the same conditions with semi-preparative RP-HPLC to afford the L-protein 2.4 in a yield of 18.5% over two steps (97% purity; Figure 2.18).

Figure 2.17. RP-HPLC chromatograms for the oxidative folding of crude reduced L-snakin-1 2.3 with chromatography performed at 50 °C. Traces shown are from samples taken at t = 0 h (black) and t = 17 h (blue). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).
Investigation of Racemic Snakin-1

Figure 2.18. RP-HPLC chromatogram of purified native L-snakin-1. Inset: mass spectrum of the peak in the region of 600-1400 m/z corresponding to the desired native L-snakin-1 ([(M+9H)+] m/z calculated 770.1, observed 770.1; deconvoluted mass 6921.2 ± 0.8 Da, calculated mass 6922.0). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).
2.4 Synthesis of D-snakin-1

2.4.1 Synthesis of D-snakin-1 fragments 2.5 and 2.6

With a robust synthesis of L-snakin-1 in hand, the preparation of the mirror image D- protein, which was expected to be uneventful, was carried out.

Synthesis of the D- peptide fragments was carried out in the same manner as for the L-peptides (§2.3) but with substitution of the Boc-protected L- amino acids with the corresponding D- amino acid building blocks. Synthesis of the D-[1Gly-29Cys]-thioester 2.5 followed a methodology analogous to that used for 1Gly-29Cys-thioester 2.1; HF cleavage and purification of the full length peptide afforded 24 mg of the peptide thioester with a minor unresolvable cysteine deletion product (3.6% yield, 97% pure; Figure 2.19).

Synthesis of the cysteinyl peptide D-[30Cys-63Pro] 2.6 required additional steps to functionalise aminomethyl polystyrene resin, firstly with the PAM linker, and then esterification with the C-terminal residue, Boc-D-Pro to afford Boc-D-Pro-PAM-resin, as the Boc-D-Pro-PAM linker precursor was not commercially available.

Figure 2.19. RP-HPLC chromatogram of pure D-snakin-1 thioester 2.5. Inset: mass spectrum of the peak in the region of 400-1000 m/z corresponding to the desired thioester 2.5 ([M+5H]+ m/z calculated 658.8, observed 658.7) and showing an unresolved cysteine deletion denoted by asterisks ([M+5H]+ m/z calculated 638.2, observed 638.1). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).
Investigation of Racemic Snakin-1

Scheme 2.6. i) Installation of PAM linker on aminomethyl polystyrene resin, followed by ii) coupling of Boc-D-Pro-OH.

The first step was achieved by reacting aminomethyl resin with (4-hydroxy-methyl)phenylacetic acid in the presence of the carbodiimide coupling reagent diisopropyl-carbodiimide (DIC) to afford the phenylamidomethyl (PAM) resin. The resin-bound alcohol was then esterified twice with Boc-D-Pro-OH, DIC and a stoichiometric amount of 4-N,N-dimethylaminopyridine (DMAP) to give Boc-D-Pro-PAM resin (Scheme 2.6). Coupling of Boc-D-Pro-OH required DIC as a coupling reagent, as it involved condensation of the carboxylic acid to a free hydroxyl group, a reaction that is sluggish when using aminium-uronium-type coupling reagents such as HBTU.

Figure 2.20. Structure of DMAP

DMAP is used as an additive in carbodiimide couplings to accelerate acylation, but due to its nucleophilic nature, it can cause racemisation. In order to limit racemisation in couplings, it is usually included in catalytic quantities. Proline has been found to be generally racemisation-resistant, a feature that has been perhaps incorrectly attributed to it being a secondary amine, as N-alkylated amino acids have previously been found to readily undergo racemisation (Scheme 2.7). A more likely reason appears to be its rigid geometry that restricts formation of an oxazolinium species. As it appeared that proline would not be subject to racemisation under the general reaction conditions employed, stoichiometric quantities of DMAP were included to promote rapid acylation in this first coupling.

Scheme 2.7. Racemisation of N-alkyl carbamate-protected amino acids via an oxazolinium ion.
Once the first residue and linker were installed, assembly of the peptide was performed analogous to 2.2 (§2.3.1), affording 82 mg of the desired product after HF cleavage and purification (10.8% yield, 87% pure). A minor unresolved serine deletion by-product was observed in the mass spectrum of the pure peptide, as well as a minor peak (15%) in the HPLC spectrum with a retention time of 10 min. Presumably this minor peak is not a peptide by-product as it did not ionise in the mass spectrometer and showed a high absorbance at 254 nm, suggesting it may be derived from (or is an adduct of) the p-cresol scavenger used in the HF cleavage reaction. With this in mind, native chemical ligation was attempted without further purification.

Figure 2.21. RP-HPLC chromatogram of pure D-snakin-1 cysteiny1 peptide 2.6 with a non-peptide peak at \( R_t = 10 \) min. Inset: mass spectrum of the peak in the region of 400-1000 m/z corresponding to the desired peptide 2.6 ([M+5H]^5+ m/z calculated 762.7, observed 762.5) and showing an unresolved serine deletion denoted by asterisks ([M+6H]^6+ m/z calculated 621.2, observed 621.2). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H₂O; buffer B = 0.1% TFA in acetonitrile).

2.4.2 Native chemical ligation and oxidative folding

Ligation of the two fragments to afford the full length linear D-protein molecule was carried out under the same conditions as for the L-protein (§2.3.2). The ligation proceeded as expected and was complete within 3 h (Figure 2.22). Isolation of the product 2.7 was achieved by desalting on a reversed phase column as described for the L-protein. 18.0 mg of the crude reduced protein 2.7 (Figure 2.23) was obtained, and used for subsequent refolding without any further manipulation.
Investigation of Racemic Snakin-1

Figure 2.22. LC-MS chromatograms for the native chemical ligation of D-snakin-1 thioester 2.5 with cysteiny peptide 2.6. Reagents and conditions: 100 mM MPAA, 20 mM TCEP, 6 M GnHCl/0.2 M Na2HPO4, pH 7.0. Chromatograms: A) t = 0 h; B) t = 1 h; C) t = 3 h. At t = 0 h, the thioester 2.5 is unresolved from the MPAA peak, which shows high absorbance. No further change was observed after 3 h. Mass spectrum: D) peak corresponding to the desired product 2.7 at t = 3 h ([M+9H]+ m/z calculated 771.5, observed 771.4). LC-MS was carried out using an analytical column (Agilent Zorbax 300-SB C3, 300 Å, 3.0 x 150 mm, 5 µm) with a gradient of 1-61% buffer B over 20 min (buffer A = 0.1% formic acid in H2O; buffer B = 0.1% formic acid in acetonitrile).
Oxidative folding was then carried out in the same folding buffer used for L-snakin-1. Thus, crude 2.7 was dissolved in 150 mM NaCl, 1 mM cysteine, 0.1 mM cystine, 0.1 M Tris pH 8 achieving a final concentration of 0.1 mg/ml of 3.7. The reaction was left overnight, after which time RP-HPLC at 50 °C indicated completion of the reaction (Figure 2.24). The mixture was acidified and purified by RP-HPLC at 50 °C using a gradient of 1-61% aq. MeCN + 0.1% TFA, affording the D- protein (3.8 mg, 19.5% yield over two steps, 96% purity; Figure 2.25).
Investigation of Racemic Snakin-1

Figure 2.24. RP-HPLC chromatograms for the oxidative folding of crude reduced D-snakin-1 2.7, with chromatography performed at 50 °C. Traces shown are for samples taken at t = 0 h (black) and t = 17 h (blue). Reagents and conditions: 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl, 0.1 M Tris pH 8.0. Analytical RP-HPLC was carried out using and analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) using a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).

Figure 2.25. RP-HPLC chromatogram of purified folded D-snakin-1 2.8. Inset: mass spectrum of the peak corresponding to the desired product 2.8 ([M+9H]9+ m/z calculated 770.1, observed 770.1; deconvoluted mass 6921.5 ± 0.3 Da, calculated mass 6922.0) and showing an unresolved serine deletion denoted by asterisks ([M+7H]7+ m/z calculated 977.4, observed 977.3). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) at 50 °C with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).
Comparison of the HPLC chromatograms and mass spectra of the purified L- (2.4) and D- (2.8) snakin-1 proteins shows that they have identical retention times and mass spectra. However, the CD spectra clearly identify the D- protein, as the circular dichroism is of the opposite polarity (Figure 2.26). It is well documented that D- proteins, the mirror image of the L- protein, result in inverted CD spectra compared to the L- proteins. Both proteins show partial α-helical character in 10 mM sodium phosphate buffer (pH 6.8), suggesting that the correct disulfide connectivity was formed (Figure 2.26).

Figure 2.26. CD spectra of L-snakin-1 (blue) and D-snakin-1 (orange). Spectra were measured on protein samples at 10 nM in 10 mM phosphate buffer (pH = 6.8).
2.5 Snakin-1 structure determination

With both L- and D- proteins in hand, the crystallisation of both L-snakin-1 alone and as a racemic mixture of L- and D- protein was carried out.

2.5.1 Crystallisation of L-snakin-1

A stock solution of L-snakin-1 was prepared in MQ H$_2$O and the protein concentration was measured as 22.8 mg/ml. Initial crystallisation screens were prepared using 384 of the conditions described by Moreland et. al. together with those described by Gorrec. Crystallisation experiments were then set up in a sitting drop vapour diffusion format, with 0.15/0.15 μl or 0.20/0.20 μl drops equilibrated against 100 μl reservoirs (see experimental section §2.6.10).

After one week, of the 480 conditions tested in this initial screen only one appeared promising (Figure 2.27). Attempts were made to reproduce this condition by hand as hanging drops in a 24-well plate. The crystallisation condition (2 M LiSO$_4$, 2% w/v PEG 400, 0.1 M Tris pH 8.5) was replicated in a 24-well plate in three different drops (1 μl/1 μl; 1 μl/2 μl; 2 μl/1 μl protein stock/crystallisation buffer; 500 μl reservoir) but none were successful in producing crystals.

![Initial hit from crystallisation screen with L-snakin-1. Crystallisation condition: 2 M LiSO$_4$, 2% w/v PEG 400, 0.1 M Tris pH 8.5. Sitting drops were prepared in Intelli-Plates (Hampton Research) with a 0.15 μl/0.15 μl mixture of protein stock/crystallisation condition against a reservoir of 100 μl.](image)

After four months, a few more promising conditions were observed with similar appearance to Figure 2.27. Only one produced crystals of a suitable size for X-ray diffraction experiments (Figure 2.28). Diffraction images obtained from one of these crystals on the
MX2 beamline at the Australian Synchrotron suggested disorder in crystal packing, making the crystals unusable (Figure 2.29).

Figure 2.28. Crystals obtained from initial screen of L-snakin-1. Crystallisation condition: 3.2 M ammonium sulfate pH 7.0. Sitting drops were prepared in Intelli-Plates with a 0.2 µl/0.2 µl mixture of protein stock/crystallisation condition against a reservoir of 100 µl. Crystals were observed after four months.
The poor diffraction was observed at three different orientations of the crystal. At this point the crystallisation of racemic snakin-1 was prioritised in preference to any further trials with L-snakin-1 alone.
2.5.2 X-ray crystallography of racemic snakin-1

Racemic protein mixtures have been known to crystallise at concentrations lower than that required for the L- protein alone, and initial crystal screening for racemic snakin-1 was therefore begun at a lower concentration. A racemic protein stock was prepared from stock solutions of L- and D-snakin-1 in MQ H₂O at a concentration of 10.6 mg/ml protein, and an initial crystallisation screen using the vapour diffusion method was prepared. All conditions described by Moreland et. al. and Gorrec were used, and crystallisation conditions set up with 0.30/0.30 μl drops against 100 μl reservoirs (see experimental section §2.6.11).

Despite the lower concentration than in our initial screens with L-snakin-1, many different conditions gave crystals within a week, some of which are shown in Figure 2.30.

![Figure 2.30. Crystals grown from racemic snakin-1 initial screens. Crystallisation conditions: left: 10% w/v PEG 20,000, 20% w/v PEG 550, 0.03M NPS, pH 8.5; right: 1.6 M sodium citrate pH 6.5](image)

Crystals from a number of conditions were flash-cooled in liquid nitrogen (see general experimental methods §4.4.1) and sent to the Australian Synchrotron (MX1 beamline) for X-ray diffraction data collection. Of these, three crystals appeared to diffract well and diffraction data were collected from these (Figure 2.31).

The first crystal (Figure 2.31A) showed a high resolution diffraction limit of 1.6 Å, and data were collected over 360° (1° oscillation/1 s exposure). However, this one data set appeared to show an additional lattice and could not be reliably indexed (Figure 2.32). In
addition, the diffraction spots appeared poorly defined, and ice rings were also observed in the diffraction.

Figure 2.31. Crystals of racemic snakin-1. A: 1.6 Å crystal (additional lattice); B: Phcn crystal; C: P1̅ crystal.

The second crystal (Figure 2.31B) showed well-defined diffraction up to a high resolution of 2.0 Å, and data were collected over a 400º range (1º oscillation/1 s exposure; Figure 2.33). The final crystal (Figure 2.31C) also diffracted well (Figure 2.34), and data were collected over a 200º range (1º oscillation/1 s exposure). However, this crystal additionally showed anisotropic diffraction and diffracted extremely poorly along one crystal orientation (Figure 2.35).

Data collected on the two latter crystals indicated the centrosymmetric space groups Phcn and P1̅ respectively, by space group prediction in POINTLESS\textsuperscript{250,251} (Table 2.1, Phcn and P1̅ entries).
Figure 2.32. Representative diffraction frame from the 1.6 Å crystal. The diffraction spots are poorly defined and there appears to be an additional lattice present.
Figure 2.33. Representative diffraction frame from *Pbcn* crystal. The diffraction appeared well-defined for this crystal.
Figure 2.34. A diffraction frame from the well-diffracting orientation of the $P\bar{1}$ crystal. Spots are clear and well-defined.
Figure 2.35. Diffraction frame representative of the poorly diffracting crystal orientation in the $P\bar{1}$ crystal. Diffraction spots are smeary and poorly defined.
Table 2.1 X-ray data collection and processing statistics for racemic snakin-1 crystals.

<table>
<thead>
<tr>
<th>Crystallisation condition</th>
<th>3.2 M ammonium sulfate pH 7.0</th>
<th>20% w/v PEG 3350, 0.2 M Na₂HPO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MX1</td>
<td>MX1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9537</td>
<td>0.9537</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 210r</td>
<td>ADSC Quantum 210r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>Pbcn</td>
<td>P 1̅</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>113.50, 64.59, 68.62</td>
<td>31.70, 43.33, 50.48</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90, 90, 90</td>
<td>93.60, 93.69, 103.18</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.18</td>
<td>0.61</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>19.57-2.00 (2.05-2.00)</td>
<td>42.06 (1.94-1.90)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>34631 (2506)</td>
<td>20004 (5564)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.5 (99.0)</td>
<td>97.5 (95.5)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>8.2 (8.3)</td>
<td>4.3 (4.4)</td>
</tr>
<tr>
<td>$R_{\text{merge}}^a$ on intensity (%)</td>
<td>37.9 (201.0)</td>
<td>17.6 (124.0)</td>
</tr>
<tr>
<td>$R_{\text{pim}}^b$ on intensity (%)</td>
<td>14.0 (73.4)</td>
<td>9.5 (66.9)</td>
</tr>
<tr>
<td>$\langle I/\sigma I \rangle$</td>
<td>5.6 (1.3)</td>
<td>5.8 (1.3)</td>
</tr>
<tr>
<td>$CC_{1/2}^c$</td>
<td>0.992 (0.621)</td>
<td>0.995 (0.787)</td>
</tr>
<tr>
<td>Overall $B$ factor from Wilson plot (Å²)</td>
<td>19.3</td>
<td>21.7</td>
</tr>
</tbody>
</table>

Estimated no. of molecules in asymmetric unit
Estimated solvent content (%)

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_M$ (Å³/Da)</td>
<td>2.27</td>
<td>2.43</td>
</tr>
</tbody>
</table>

[values of the outermost resolution shell are in parentheses]

\[ a \) $R_{\text{merge}} = \frac{\sum_{hkl} \left| \sum_i I_i(hkli) - \overline{I(hkli)} \right|}{\sum_{hkl} \sum_i I_i(hkli)}$  
\[ b \) $R_{\text{pim}} = \frac{\sum_{hkl}[1/(N-1)]^{1/2} \left| \sum_i I_i(hkli) - \overline{I(hkli)} \right|}{\sum_{hkl} \sum_i I_i(hkli)}$  
\[ c \) Pearson correlation coefficient between random half sets of the data
Investigation of Racemic Snakin-1

The choice of a centrosymmetric space group over a chiral space group by space group prediction algorithms is governed by the intensity statistics for the data. When reindexed in $P1$, the cumulative intensity distributions of the reflections (“acentric observed”) in these data are highly consistent with the theoretical distributions for completely centric data, and confirm that the crystals are truly in centrosymmetric space groups, which only racemic crystals can sample (Figure 2.36). These data sets were used for molecular replacement (MR) and *ab initio* structure determination attempts.

![Figure 2.36. Cumulative intensity distribution of: A: Pbcn; B: P1 (Table 2.1). The distribution of the ‘acentric reflections’ observed is displayed in green, and the respective theoretical distributions for acentric and centric are shown in red and black.](image)

The data obtained from these racemic snakin-1 crystals were not at resolutions suitable for direct methods structure solution, with a high resolution cut off of 1.6 Å in the best case. The native snakin-1 protein does not contain any heavy-atoms suitable for heavy-atom phasing but does contain a large number of sulfur atoms, which have been successfully used for anomalous phasing. However, data obtained from centrosymmetric crystals cannot exhibit anomalous dispersion effects, hence structure solution of these data by experimental phasing with sulfur is not possible. Thus, strategies for structure solution with these data were limited to molecular replacement and *ab initio* phasing methods.
2.5.3 Structure solution attempts via molecular replacement

For molecular replacement, search models of the ab initio predicted snakin-1 structure prepared by Porto et al.,230 and models prepared from EcAMP1 and Viscotoxin A3 were used. EcAMP1, an α-helical hairpin protein, and Viscotoxin A3, a thionin, were chosen as search models, as these are crAMPs thought to have similar structures to snakin-1 (Figure 2.10).230 Molecular replacement was carried out using PHASER, both as part of the CCP4 suite252 and as part of the PHENIX suite (see general experimental methods §4.4.5).253

Both search models were unsuccessful in providing a useful MR solution. MR trials included trimmed search models, in which the models were reduced to polyAla chains. PHASER uses maximum-likelihood rotation and translation searches, and employs translation function Z-scores (TFZ) and log-likelihood gain (LLG) values as indicators of the quality of the MR solution.254 A TFZ value > 8.0 and a high LLG value generally signify a correct solution. For all MR trials with PHASER, TFZ values obtained were less than this, and density modification and refinement in PHENIX.REFINE did not result in $R_{free} < 0.50$.

One notable feature is that during MR trials with the Pbcn data, PHASER finds a non-origin peak > 20% of the height of the origin peak and located more than 15 Å from the origin, suggesting that molecules related by translational non-crystallographic symmetry (tNCS) are present in the asymmetric unit. tNCS related molecules, which exist in a similar orientation and are related by translation symmetry, complicate structure solution, as their contributions to the structure factors are modified by constructive and destructive interference between them.255 The additional issue of tNCS may be a contributing factor for the failure to solve the structure by MR.

2.5.4 Structure solution attempts via fragment-based ab initio methods

Unsuccessful MR trials led to attempts at structure solution via iterative ab initio methods (see general experimental methods §4.4.5). This approach used automated iterative rounds of density modification, autobuilding, and refinement, and was previously used in the phasing of another racemic structure, Rv1738.161

In this method, an MR search for short ideal α-helices is first carried out. α-helices are generally invariant in their geometry over short segments of 10-14 residues, a feature that can be exploited by searching for ideal α-helices of this size. However, the poor scores obtained from MR searches when using small fragments as search models leads to a problem in discriminating real solutions from false solutions. The original use of the
fragment-based *ab initio* procedure by Bunker et. al. appears not to have dealt with this problem and instead used only the ‘best’ MR solution from ideal helices of different lengths.\textsuperscript{161}

A best MR solution is first subjected to density modification in *PARROT*.\textsuperscript{256} Additional cycles are then carried out incorporating chain building in *BUCCANEER*,\textsuperscript{257, 258} refinement in *REFMAC*,\textsuperscript{259, 260} followed by another round of density modification in *PARROT* (Figure 2.37). In each cycle, the density-modified map from the end of the previous cycle is used for chain building, with the MR model as a starting model. Using this approach, incorrect models do not bias successive cycles.

This method was applied to the snakin-1 data in *Pbcn* and *P\textilde*1, but was ultimately unsuccessful. In addition, in our hands, the structure of Rv1738 could not be solved using the original data and method (§2.6.12). Bunker has observed that the newer versions of the *BUCCANEER* chain tracing program are not compatible with this method (personal correspondence), and so an older version of the program (version 1.5.1; the method was originally successful with version 1.1.9, which could not be used in our hands) was employed for additional trials. In addition, to confirm whether this method could be used to generate a complete model from a correct, truncated version of the final refined model, truncated models of Rv1738 (PDB code: 4wpy) incorporating residues 69Ser-77Val, 63Val-77Val, and 58Gly-77Val were used as ‘MR solutions’. All attempts at reproducing the structure solution of Rv1738 with these were unsuccessful.
Figure 2.37. Workflow for iterative *ab initio* structure solution. Each iterative cycle takes density modified maps from the previous cycle for chain tracing with the initial MR model.

2.5.5 Structure solution attempts via *ARCIMBOLDO*

The last method of structure solution attempted was *ab initio* phasing via *ARCIMBOLDO* (see general experimental methods §4.4.5).\textsuperscript{261, 262} Like the iterative *ab initio* method described above, *ARCIMBOLDO* is a program which performs *ab initio* structure solution by molecular replacement with short ideal helices. There are many possible non-redundant ways to place a small helix correctly in a MR search, and for these small search models it is not possible to discriminate between successful and unsuccessful MR solutions. In *ARCIMBOLDO*, this problem is treated by not discriminating between solutions – multiple *PHASER* MR solutions are taken forward to density modification and chain tracing in *SHELXE* after successful location of a helix.\textsuperscript{263} Restrictions are placed on the number of forwarded partial solutions only to limit the computational requirements, and these are chosen from those with the top TFZ scores.\textsuperscript{261}
For *ARCIMBOLDO* structure solution trials carried out in this work, *ARCIMBOLDO LITE* was used, a stand-alone program that is optimised for use on single machines. *ARCIMBOLDO* employs *SHELXE*, which cannot work with achiral space groups and thus the data were reduced to lower symmetry chiral space groups. The $P\overline{1}$ data set was reduced to $P1$, while the $Pbcn$ data (with four molecules in the asymmetric unit) was not used, as the high symmetry in the $Pbcn$ space group meant that a large number of molecules would need to be found if this data set was re-indexed in lower symmetry space groups. In the $P1$ data reduced from $P\overline{1}$, *MATTHEWS_COEF* estimates four molecules of snakin-1, corresponding to eight helices according to previous structure predictions (Figure 2.10). *ARCIMBOLDO* structure solution was carried out to search for this number of helices, but structure solution with this dataset was ultimately unsuccessful.

2.5.6 Crystal optimisation and X-ray crystallography

With the inability to solve the structure of snakin-1 by molecular replacement and *ab initio* structure solution with the data in hand, efforts were turned towards optimising crystallisation. The previous data collection had been performed on crystals harvested directly from the initial crystallisation screen, and it was hoped that through optimisation of promising crystallisation conditions, higher resolution data would be obtained, thus enabling direct methods structure solution.

Of the conditions that produced crystals in the initial screen, only one (20% w/v PEG 3350, 0.2 M Na$_2$HPO$_4$) was found to be reproducible, affording long, thin plates (Figure 2.38). The optimal drop setup for growing crystals with this crystallisation buffer were drops of 1.0 µl/0.5 µl protein stock/crystallisation buffer against a reservoir of 100 µl. No further improvement in the crystals could be obtained by fine screening of the crystallisation conditions or by streak seeding, however (see general experimental methods §4.3.2).
Chapter 2

Figure 2.38. Optimised racemic snakin-1 crystals. Crystals are long thin plates, grown in 20% PEG3350, 0.2 M Na$_2$HPO$_4$.

One crystal from this condition was flash-cooled and sent for data collection at the Australian Synchrotron. This crystal showed high quality diffraction throughout, to a maximum resolution of 1.6 Å (Table 2.2), with well-defined spots (Figure 2.40). Space group prediction in POINTLESS indicated the centrosymmetric space group $P2_1/c$.

Figure 2.39. Optimised racemic snakin-1 crystal used for data collection.
Table 2.2. X-ray collection and processing statistics for optimised $P2_1/c$ racemic snakin-1 crystal.

<table>
<thead>
<tr>
<th>Crystallisation condition</th>
<th>20% w/v PEG 3350, 0.2 M Na$_2$HPO$_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MX2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.4586</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 315r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>0.5</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>300</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>0.5</td>
</tr>
<tr>
<td>Space group</td>
<td>$P2_1/c$</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>65.15, 27.74, 73.35</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90, 100.19, 90</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.28</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>34.14-1.60 (1.63-1.60)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>33736 (1631)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.7 (95.1)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.8 (6.0)</td>
</tr>
<tr>
<td>$R_{\text{merge}}^a$ on intensity (%)</td>
<td>8.6 (113.6)</td>
</tr>
<tr>
<td>$R_{\text{pim}}^b$ on intensity (%)</td>
<td>3.9 (50.3)</td>
</tr>
<tr>
<td>$&lt;I/\sigma I&gt;$</td>
<td>10.5 (1.5)</td>
</tr>
<tr>
<td>$CC_{1/2}^c$</td>
<td>0.997 (0.729)</td>
</tr>
<tr>
<td>Overall $B$ factor from Wilson plot (Å$^2$)</td>
<td>21.0</td>
</tr>
<tr>
<td>Estimated no. of molecules in asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Estimated solvent content (%)</td>
<td>47.83</td>
</tr>
<tr>
<td>$V_M$ (Å$^3$/Da)</td>
<td>2.36</td>
</tr>
</tbody>
</table>

[values of the outermost resolution shell are in parentheses]

$R_{\text{merge}}^a = \sum_{hkl} \left| \sum_i l_i(hkl) - \bar{I}(hkl) \right| / \sum_{hkl} \sum_i l_i(hkl)$

$R_{\text{pim}}^b = \sum_{hkl} \left[ 1/(N-1) \right]^{1/2} \sum_i l_i(hkl) - \bar{I}(hkl) / \sum_{hkl} \sum_i l_i(hkl)$

$CC_{1/2}^c$ Pearson correlation coefficient between random half sets of the data
MR and ab initio structure solution methods were carried out as in the case of the Pbcn and P1 data, but again did not result in a correct solution. Phaser indicated that tNCS may be present in this data as well, an additional complication to structure solution. For ARCIMBOLDO, the data were reduced to P1, where eight snakin-1 molecules and hence 16 helices were expected, but a correct solution could not be found. The structure of snakin-1 appeared intractable to these methods, and so we devised alternative methods of structure solution by experimental phasing as described in Chapter 3.
2.6 Experimental

2.6.1 L-snakin-1 \(^{29}\)Cys-thioester 2.1

Boc-Gly-O-CH\(_2\)-phi-CH\(_2\)CO\(_2\)H (0.4 mmol) was coupled to aminomethyl polystyrene resin (0.2 g, 0.2 mmol, 1 mmol/g loading) with DIC (0.4 mmol) in CH\(_2\)Cl\(_2\)/DMF (4/1, \(v/v\), 2 ml) overnight, drained and washed with CH\(_2\)Cl\(_2\). The ninhydrin test was negative. The Boc group was removed with neat TFA for 2 min, washed with DMF and coupled to \(S\)-tritylmercaptopropanionic acid (5 eq.), HATU (4.75 eq.) and \(i\)Pr\(_2\)NEt (12 eq.) in DMF for 1 h. The resin was washed with DMF, the trityl group was removed with a mixture of TFA:H\(_2\)O:TIPS 95/2.5/2.5 \(v/v/v\) for 3 x 2 min, and solid phase peptide synthesis was then performed manually according to the general \(in\ situ\) neutralisation Boc-SPPS methods with Boc-L-amino acids (see general experimental methods §4.2.4). Neat TFA was used as the Boc deblocking reagent and HATU/\(i\)Pr\(_2\)NEt as the coupling reagents. The dried peptidyl resin (0.95 g) was cleaved with anhydrous HF/p-cresol/p-thiocresol (18/1/1, \(v/v/v\), 10 ml) for 1 h at 0 °C. Following evaporation of HF, the peptide was precipitated with cold diethyl ether, isolated by centrifugation, washed with cold diethyl ether, dissolved in 1/1 \(v/v\) acetonitrile/water containing 0.1% TFA, filtered and lyophilized to afford the crude peptide (350 mg). Purification by slow gradient RP-HPLC afforded the title compound (20.6 mg, 3.1% yield; \([M+4H]^{4+}\) \(m/z\) calculated 823.2, observed 823.1).

2.6.2 L-snakin-1 \(^{63}\)Pro 2.2

Boc-Pro-O-CH\(_2\)-phi-CH\(_2\)CO\(_2\)H (0.4 mmol) was coupled to aminomethyl polystyrene resin (0.2 g, 0.2 mmol, 1 mmol/g loading) with DIC (0.4 mmol) in CH\(_2\)Cl\(_2\)/DMF (9/1, \(v/v\), 2 ml) overnight, drained and washed with CH\(_2\)Cl\(_2\). The ninhydrin test was negative. Solid phase peptide synthesis was then performed manually with Boc-L-amino acids using the Boc \(in\ situ\) neutralization method (see general experimental methods §4.2.4), with neat TFA as the Boc de-blocking reagent and HATU/\(i\)Pr\(_2\)NEt as the coupling reagents. After the final amino acid coupling, the DNP group on \(^{45}\)His was removed by thiolysis by the addition of neat thiophenol (200 µl) to the DMF-swelled resin overnight. The resin was drained, washed with DMF, and the final Boc group was removed with neat TFA for 2 mins. The dried peptidyl resin (1.18 g) was cleaved with anhydrous HF/p-cresol/p-thiocresol (18/1/1, \(v/v/v\), 10 ml) for 1 h at 0 °C. Following evaporation of HF, the peptide was precipitated with cold diethyl ether, isolated by centrifugation, washed with cold diethyl ether, dissolved in 1/1 \(v/v\) acetonitrile/water containing 0.1% TFA, filtered and lyophilized to afford the crude peptide.
peptide (553 mg). Purification by slow gradient RP-HPLC afforded the title compound (26.1 mg, 3.4% yield; [M+5H]^{5+} m/z calculated 762.7, observed 762.7).

2.6.3 Reduced full length L-snakin 1 ^{1}\text{Gly}^{-63}\text{Pro} 2.3
Native chemical ligation of ^{1}\text{Gly}^{-29}\text{Cys-thioester} 2.1 (9.2 mg, 2.81 µmol) and ^{30}\text{Cys}^{-63}\text{Pro} 2.2 (10.7 mg, 2.81 µmol) was performed in a buffer solution consisting of 6 M GnHCl, 0.2 M Na_{2}HPO_{4}, 100 mM MPAA and 20 mM TCEP (2.80 ml, pH = 6.9) under an argon atmosphere. The ligation reaction was completed in 3 h at room temperature as judged by HPLC. The reaction mixture was acidified to pH 2 with 5 M HCl, recovered by RP-HPLC and lyophilized. LC-MS analysis of the crude product showed the desired linear ^{1}\text{Gly}^{-63}\text{Pro} product: [M+9H]^{9+} m/z calculated 771.5, observed 771.0 (16.5 mg crude, 84.6% crude yield).

2.6.4 Native L-snakin-1 ^{1}\text{Gly}^{-63}\text{Pro} 2.4
Crude reduced ^{1}\text{Gly}^{-63}\text{Pro} 2.3 (16.5 mg, 2.37 µmol) was dissolved in a buffer solution of 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl and 0.1 M Tris (165 ml, pH = 8.0), and this reaction mixture stood at room temperature. The reaction was monitored by RP-HPLC (50 °C) and LC-MS and was completed within 22 h. Purification of the reaction mixture by RP-HPLC at 50 °C over a gradient of 1-61% B over 120 min (buffer A = 0.1% aq. TFA; B = 0.1% TFA in MeCN) afforded the synthetic L-snakin-1 (4.2 mg, 21.5% yield over two steps; [M+9H]^{9+} m/z calculated 770.1, observed 769.7, deconvoluted mass 6921.2 ± 0.8 Da).

2.6.5 D-snakin-1 D-[^{1}\text{Gly}^{-29}\text{Cys}]\text{-thioester} 2.5
Boc-Gly-O-CH_{2}-\text{phi}-CH_{2}CO_{2}H (0.4 mmol) was coupled to aminomethyl polystyrene resin (0.2 g, 0.2 mmol, 1 mmol/g loading) with DIC (0.4 mmol) in CH_{2}Cl_{2}/DMF (4/1, v/v, 2 ml) overnight, drained and washed with CH_{2}Cl_{2}. The ninhydrin test was negative. The Boc group was removed with neat TFA for 2 min, washed with DMF and coupled to S-tritylmercaptopropionic acid (5 eq.), HATU (4.75 eq.) and iPr_{2}NEt (12 eq.) in DMF for 1 h. The resin was washed with DMF, the trityl group was removed with a mixture of TFA:H_{2}O:TIPS 95/2.5/2.5 v/v/v for 3 x 2 min, and solid phase peptide synthesis was then performed manually according to the general in situ neutralisation Boc-SPPS methods with Boc-D-amino acids (see general experimental methods §4.2.4). Neat TFA was used as the Boc deblocking reagent and HATU/ iPr_{2}NEt as the coupling reagents. The dried peptidyl resin (1.02 g) was cleaved with anhydrous HF/p-cresol/p-thiocresol (18/1/1, v/v/v, 10 ml)
for 1 h at 0 °C. Following evaporation of HF, the peptide was precipitated with cold diethyl ether, isolated by centrifugation, washed with cold diethyl ether, dissolved in 1/1 v/v acetonitrile/water containing 0.1% TFA, filtered and lyophilized to afford the crude peptide. Purification by slow gradient RP-HPLC afforded the title compound (18.2 mg, 2.8% yield; [M+5H]^{5+} m/z calculated 658.8, observed 658.7).

### 2.6.6 D-snakin-1 D-[{^{30}Cys-{^{63}Pro}] peptide 2.6

4-(Hydroxymethyl)phenylacetic acid (0.4 mmol) was coupled to aminomethyl polystyrene resin (0.2 g, 0.2 mmol, 1 mmol/g loading) with DIC (0.4 mmol) in CH₂Cl₂ (2 ml) for 1 h, drained and washed with CH₂Cl₂ and repeated. Boc-D-Pro-OH (5 eq.) was coupled to the resin with DIC (5 eq.) and DMAP (5 eq.) for 1 h, twice. Solid phase peptide synthesis was then performed manually with Boc-d-amino acids using the Boc in situ neutralization method (see general experimental methods §4.2.4), with neat TFA as the Boc deblocking reagent and HATU/ iPr₂NEt as the coupling reagents. After the final amino acid coupling, the DNP group on 45His was removed by thiolysis with addition of neat thiophenol (200 µl) to the DMF-swelled resin overnight. The resin was drained, washed with DMF, and the final Boc group was removed with neat TFA. The dried peptidyl resin (1.22 g) was cleaved with anhydrous HF/p-cresol/p-thiocresol (18/1/1, v/v/v, 10 ml) for 1 h at 0 °C. Following evaporation of HF, the peptide was precipitated with cold diethyl ether, isolated by centrifugation, washed with cold diethyl ether, dissolved in 1/1 v/v acetonitrile/water containing 0.1% TFA, filtered and lyophilized to afford the crude peptide. Purification by slow gradient RP-HPLC afforded the title compound (85% purity; 82 mg, 10.8% yield; [M+5H]^{5+} m/z calculated 762.7, observed 762.5).

### 2.6.7 Reduced full length D-snakin-1 D-[{^{1}Gly-{^{63}Pro}] 2.7

Native chemical ligation of D-[{^{1}Gly-{^{29}Cys}-COSR 2.5 (9.1 mg, 2.77 µmol) and D-[{^{30}Cys-{^{63}Pro}] 2.6 (10.5 mg, 2.77 µmol) was performed in a buffer solution consisting of 6 M GnHCl, 0.2 M Na₂HPO₄, 100 mM MPAA and 20 mM TCEP (2.77 ml, pH = 7.0) under an argon atmosphere. The ligation reaction was completed in 3 h at room temperature as judged by HPLC. The reaction mixture was acidified to pH 2 with 5 M HCl, recovered by RP-HPLC and lyophilized. LC-MS of the crude product showed the desired linear D-[{^{1}Gly-{^{63}Pro}] product: [M+9H]^{9+} m/z calculated 771.5, observed 771.4 (18.0 mg crude, 93.8% crude yield).
2.6.8  Folded D-snakin-1 D-[\textsuperscript{1}Gly\textsuperscript{-63}Pro] 2.8

Crude reduced D-[\textsuperscript{1}Gly\textsuperscript{-63}Pro] 2.7 (18.0 mg, 2.60 \(\mu\)mol) was dissolved in a buffer solution of 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl and 0.1 M Tris (180 ml, pH = 8.0), and this reaction mixture stood at room temperature. The reaction was monitored by RP-HPLC (50 °C) and LC-MS and was completed within 24 h. Purification of the reaction mixture by RP-HPLC at 50 °C over a gradient of 1-61% B over 120 min (buffer A = 0.1% aq. TFA; B = 0.1% TFA in MeCN) afforded the folded synthetic D-snakin-1 (3.8 mg, 19.5% yield over two steps; [M+9H]\textsuperscript{9+} \textit{m/z} calculated 770.1, observed 769.5, deconvoluted mass 6921.5 ± 0.3 Da).

2.6.9  Circular dichroism spectrometry

CD spectra were obtained on a PiStar spectrometer (Applied Photophysics; Leatherhead, UK). Samples were placed in a 1 mm quartz cuvette for measurement. Spectra were collected at 6 °C with a 2 nm optical bandwidth over the range of 180-300 nm in 0.5 nm increments. Measurements at each wavelength were performed over 2.0 s. The measurement of each spectrum was repeated seven times and the final spectrum taken as the mean of the measurements. The averaged baseline spectrum of the buffer was subtracted from the average measured with protein in buffer to obtain the solvent subtracted spectrum. Samples were prepared at 9-10 nM concentration.

2.6.10  Initial crystallisation screen of L-snakin-1

A protein stock solution of L-snakin-1 2.4 was prepared in MQ H\textsubscript{2}O concentrated to 22.8 mg/ml using a Vivaspin 500 concentrator (3,000 MWCO; Vivaproducts, Littleton, MA). The stock solution was centrifuged through a Nanosep MF 0.2 µm filter (Pall Life Sciences) and used in crystallisation without further manipulation.

An estimation of the extinction coefficient using the method of Pace \textit{et al.}\textsuperscript{264} found a value of 0.7541 (mg/ml\textsuperscript{-1} cm\textsuperscript{-1}), which when applied to the measured absorbance of the L-snakin-1 stock solution, resulted in a concentration estimate of 22.8 mg/ml. Protein solutions were used with the assumption that the measured estimate was correct.

Precipitant solutions for the initial crystallisation screens were transferred from 96-well deep well blocks to the reservoir of 96-well Intelli-Plates (Hampton Research, 100 µl per reservoir) using a MultiPROBE II HT/EX liquid-handling robot (Perkin-Elmer). The precipitation solutions used are those described by Moreland \textit{et al.}\textsuperscript{248} and Gorrec\textsuperscript{249} and
were distributed over six Intelli-Plates. Crystallisation was carried out at 18 °C. Sitting drops containing 0.15 µl of quasi-racemic protein stock and 0.15 µl of reservoir solution, or 0.20 µl and 0.20 µl respectively, were mixed and dispensed using an Oryx4 Protein Crystallization Robot.

Upon set up of drops, Intelli-Plates were immediately sealed with HR4-521 ClearSeal Film (Hampton Research). Using light microscopy, plates were inspected daily for the first week, every second day for the second week, every three-to-five days for the third and fourth weeks, and infrequently thereafter.

2.6.11 Initial crystallisation screen of DL-snakin-1
A racemic protein stock solution was prepared from individual stock solutions of L-snakin-1 2.4 and D-snakin-1 2.8 in MQ H2O. Stock solutions of L-snakin-1 (14.1 mg/ml) and D-snakin-1 (13.8 mg/ml) were centrifuged through Nanosep MF 0.2 µm filters (Pall Life Sciences) then mixed in a ratio of 72/72/56 v/v/v of L-snakin-1/D-snakin-1/MQ H2O. A protein concentration of 12.2 mg/ml was determined using a Nanodrop ND-1000 spectrophotometer.

Precipitant solutions for the initial crystallisation screens were transferred from 96-well deep well blocks to the reservoir of 96-well Intelli-Plates (Hampton Research, 100 µl per reservoir) using a MultiPROBE II HT/EX liquid-handling robot (Perkin-Elmer). The precipitant solutions used are those described by Moreland et al.248 and Gorrec,249 and were distributed over six Intelli-Plates. Crystallisation was carried out at 18 °C. Sitting drops containing 0.30 µl of quasi-racemic protein stock and 0.30 µl of reservoir solution were mixed and dispensed using an Oryx4 Protein Crystallization Robot.

Upon set up of drops, Intelli-Plates were immediately sealed with HR4-521 ClearSeal Film (Hampton Research). Using light microscopy, plates were inspected daily for the first week, every second day for the second week, every three-to-five days for the third and fourth weeks, and infrequently thereafter.

2.6.12 Attempts to reproduce structure solution of Rv1738
Fragment-based ab initio structure solution was carried out as described in the general experimental methods (§4.4.5) using an identical copy of the script used in the solution of Rv1738, and the original data and MR solutions used to generate the solution.161 The PHASER MR solution used was obtained from an MR search for a 10-residue ideal helix,
known to be able to generate the complete structure of Rv1738. The process was run with 10 iterative cycles (sufficient to generate the solution; Bunker, personal correspondence) using current crystallography software (see general experimental methods §4.4.9), but failed to generate the correct solution. The process was repeated using an older version of BUCCANEER as it was noted that current versions could not generate the solution (Bunker, personal correspondence), but this was also unsuccessful (PARROT 1.0.3, BUCCANEER 1.5.1, REFMAC 5.8.0107; the version of BUCCANEER used to generate the original solution was 1.1.9, but we were unable to use this version). Finally, to determine whether the fragment-based ab initio structure solution method can generate a complete model with a correct ‘MR solution’, truncated models of the α-helix in Rv1738 consisting of residues 69Ser–77Val, 63Val–77Val, and 58Gly–77Val were used as initial search solutions to be extended. The fragment-based ab initio method was run with each truncated model using both versions of BUCCANEER described here for 30 iterative cycles for each. All attempts were unsuccessful.
Chapter 3

Structure Determination of Snakin-1
3.1 Overview

This chapter details the preparation of a modified L-snakin-1 protein incorporating a 4-iodophenylalanine residue, and the determination of the structure of snakin-1 using quasi-racemic crystals incorporating this modified protein, using experimental phasing techniques. It is divided into seven sections: a brief consideration of anomalous dispersion effects in quasi-racemic crystals (§3.2), the synthesis of iodo-L-snakin-1 (§3.3), structure determination by experimental phasing of quasi-racemic crystals (§3.4), a discussion of the structure of snakin-1 (§3.5), evaluation of the antimicrobial activity of D-snakin-1 (§3.6), a discussion of the results from this work (§3.7), and concluding remarks (§3.8).

3.2 Anomalous dispersion in quasi-racemic crystals

Despite obtaining high resolution diffraction data from crystals grown with racemic protein stock, the structure of snakin-1 could not be determined by molecular replacement or ab initio methods (see the previous chapter, §2.5). In order to determine the structure, an experimental phasing method was pursued.

Two techniques commonly used for experimental phasing are single-wavelength anomalous dispersion (SAD) and multiple-wavelength anomalous dispersion (MAD; introduced in §1.1.1.2). Both are based on the use of anomalous dispersion effects, a wavelength-dependent contribution to the diffraction of atoms. These anomalous dispersion effects produce Bijvoet differences in the Friedel mates ($F_{hkl}$ and $F_{h\bar{k}\bar{l}}$), and dispersive differences between reflections measured at different wavelengths, both of which can be measured and used to provide the positions and phases for the atoms contributing significantly to this effect. Symmetry-related Friedel pairs, i.e. centric reflections, must be equal, and thus will be devoid of Bijvoet differences. Importantly, for crystals in centrosymmetric space groups, each reflection is centric (related through inversion symmetry to its Friedel mate), and so Bijvoet differences are not observed.

In 2008, Pentelute et al.\(^9\) showed that by destroying the perfect inversion symmetry relationship of L- and D-enantiomers of snow flea antifreeze protein (sfAFP), experimental phase information could be obtained. The symmetry was destroyed by incorporation of a selenium containing Gln-analogue, which also provided a source of signal for MAD phasing. Although the perfect symmetry was destroyed, the crystals still retained the improved crystallisation offered by racemic mixtures, providing a method for accessing experimental phasing techniques in these crystal systems.
Given the successful application of anomalous phasing methods in this previous example, quasi-racemic crystallisation of snakin-1 for SAD/MAD methods was chosen for phasing. Typical incorporation of a heavy atom into synthetic peptides/proteins is in the form of selenium as selenocysteine or selenomethionine. Snakin-1 contains no methionine residues but an abundance of cysteines, thus the obvious choice of substitution was selenocysteine for cysteine. However, the true disulfide arrangement in snakin-1 was not known, and due to potential complications in forming a specific seleno-sulfide bond, or in selecting a true disulfide bond for substitution with a diselenide bond, a synthesis incorporating selenocysteine was not considered ideal.

4-iodophenylalanine (4-iodo-Phe; Figure 3.1) has previously been reported to successfully provide phase information in crystals of protein incorporating this unnatural amino acid.\textsuperscript{265} It was envisaged that this residue would be a suitable analogue of tyrosine or phenylalanine that would be stable to the reagents and conditions typically employed in protein chemical synthesis methods. Thus, the synthesis of an analogue of L-snakin-1 incorporating L-4-iodophenylalanine was carried out.

![Chemical structures of phenylalanine, tyrosine, and 4-iodophenylalanine.](image)

**Figure 3.1.** Chemical structures of phenylalanine, tyrosine, and 4-iodophenylalanine.

### 3.3 Synthesis of iodo-L-snakin-1

#### 3.3.1 Choice of substitution position

For an anomalous signal that is useful for phase information, it is important that the anomalous scatterer be within a structured segment of the protein. The JPred 4 secondary structure prediction server\textsuperscript{266, 267} and the i-Tasser structure prediction server\textsuperscript{268-270} both predicted helical segments in the N-terminal end of the protein, but although there is significant overlap in the predictions, the length of the helices differ (Figure 3.2).
From these predictions, it was anticipated that the $^4$Phe and $^{25}$Tyr residues would make good substitution sites for 4-iodophenylalanine, as they are either close to or within the predicted helices. Substitution of $^{25}$Tyr of was chosen for the preparation of an analogue of L-snakin-1, due to the similarity between the structures of 4-iodophenylalanine and tyrosine. As such, it was hoped that any potential impact on crystal packing would be minimised.

The synthetic strategy to prepare this iodo-L-snakin-1 was unchanged from that used to prepare the native protein (Scheme 3.1). Substitution of Boc-Tyr(2-Br-Z)-OH in SPPS with Boc-4-iodo-Phe-OH for coupling at the 25th residue within the thioester fragment 3.1 would afford an iodinated thioester, which can undergo native chemical ligation with the native cysteinyl fragment to form the linear iodinated protein. Disulfide formation under redox conditions would then generate the folded protein.
3.3.2 Synthesis of iodo-L-snakin $^1$Gly-$^{29}$Cys-thioester fragment ($^{254}$-iodo-Phe) 3.1

The synthesis of the iodinated thioester fragment, $^1$Gly-$^{29}$Cys-COSR ($^{254}$-iodo-Phe) 3.1, was carried out on a 0.2 mmol scale by standard in situ neutralisation Boc-SPPS methods with HATU/iPr$_2$NEt as the coupling reagents and neat TFA as the deblocking agent. The only deviation from the synthesis of the native thioester was the replacement of Boc-Tyr(2-Br-Z)-OH with commercially available Boc-4-iodo-Phe-OH (Scheme 3.2). Cleavage of the peptide from resin at 0 °C in anhydrous HF afforded crude product with a major component corresponding to the desired iodinated peptide 3.1 ([M+4H]$^{4+}$ m/z calculated 850.7, observed 850.5). After slow gradient RP-HPLC purification, purified 3.1 was obtained in a 10.7% yield (73 mg, 93% pure; Figure 3.3).

Scheme 3.2. Synthesis of iodo-L-snakin-1 $^1$Gly-$^{254}$-iodo-Phe-$^{29}$Cys-thioester 3.1 by Boc-SPPS. Synthetic steps: i) functionalisation of aminomethyl polystyrene resin with Boc-Gly-O-CH$_2$-phi-CH$_2$-CO$_2$H followed by Boc deprotection; ii) coupling of S(Trt)-mercaptopropionic acid followed by trityl removal; iii) in situ neutralisation Boc-SPPS; iv) coupling of Boc-4-iodo-Phe-OH residue with HATU/iPr$_2$NEt; v) in situ neutralisation Boc-SPPS; vi) HF cleavage. PG denotes a side-chain protecting group.
3.3.3 Native chemical ligation and oxidative folding

The iodo-L-snakin-1 thioester 3.1 was then ligated with the previously prepared native L-snakin cysteinyl fragment 2.2 (§2.3) under the same conditions used for the native L-protein (§2.3.2). The ligation was carried out in a mixture containing 1 mM of each peptide in a ligation buffer comprising 100 mM MPAA, 20 mM TCEP, 6 M GnHCl, and 200 mM Na₂HPO₄ (pH = 6.9). The reaction was monitored by LC-MS, which initially showed only the thioester and not the cysteinyl peptide in the conditions employed. The mass spectrum of the peak corresponding to this thioester changed with time, and after three hours corresponded only to the desired reduced iodo-L-snakin-1 3.2 product (Figure 3.4). The ligation product was recovered by loading on a C4 (Grace) reversed phase HPLC column, elution of the buffer components with 5% aq. acetonitrile followed by elution of the linear protein by 65% aq. acetonitrile, affording crude reduced iodo-L-snakin-1.
Figure 3.4. LC-MS chromatograms for the native chemical ligation of iodo-L-snakin-1 thioester 3.1 with cysteiny peptide 2.2. Reagents and conditions: 100 mM MPAA, 20 mM TCEP, 6 M GnHCl/0.2 M Na2HPO4, pH 6.9. Chromatograms: A) t = 0 h; B) t = 1 h; C) t = 3 h. No further change was observed after 3 h. The highly absorbing peak at R_t = 2.0-3.5 min corresponds to MPAA. Mass spectrum of: D) the broad peptide peak at t = 3 h (R_t = 12.0-14.0) corresponding to the desired product 3.2 at t = 3 h ([M+9H]^{9+} m/z calculated 783.7, observed 782.3). LC-MS was carried out using an analytical column (Agilent Zorbax 300-SB C3, 300 Å, 3.0 x 150 mm, 5 µm) with a gradient of 5-65% buffer B over 20 min (buffer A = 0.1% formic acid in H2O; buffer B = 0.1% formic acid in acetonitrile).

Oxidative folding of the reduced protein was carried out in the same conditions as described for the native snakin-1. Crude 3.2 was dissolved in an oxidative folding buffer comprising 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl, and 0.1 M Tris (pH = 8.0), achieving a concentration of 0.1 mg/ml crude linear protein. Within 24 hours, the reaction appeared
complete by analytical RP-HPLC and LC-MS (Figure 3.5). The reaction mixture was purified by semi-preparative RP-HPLC at 50 °C to afford the desired native iodo-L-snakin-1 3.3 in a 26.3% yield over two steps (93% pure; [M+9H]9+ m/z calculated 782.3, observed 782.3; deconvoluted mass 7031.3 ± 0.6 Da, calculated mass 7031.9 Da; Figure 3.6).

Figure 3.5. RP-HPLC chromatograms for the oxidative folding of crude reduced iodo-L-snakin-1 3.2 with chromatography performed at 50 °C. Traces shown are from samples taken at t = 0 h (black) and t = 21 h (blue). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 μm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H2O; buffer B = 0.1% TFA in acetonitrile).
Figure 3.6. RP-HPLC chromatogram of purified iodo-L-snakin-1 3.3. Inset: mass spectrum of the peak in the region of 780-1420 $m/z$ corresponding to the desired iodo-L-snakin-1 3.3 ([M+9H]$^9+$ $m/z$ calculated 782.3, observed 782.3; deconvoluted mass 7031.3 ± 0.6 Da, calculated mass 7031.9 Da). Analytical RP-HPLC was carried out using an analytical column (Grace Gracesmart C18, 120 Å, 2.1 x 50 mm, 3 µm) with a gradient of 1-61% buffer B over 60 min (buffer A = 0.1% TFA in H$_2$O; buffer B = 0.1% TFA in acetonitrile).

The CD spectrum of the iodo-protein 3.3 was similar to that of the native L-protein 2.4 and also showed partial $\alpha$-helical character (Figure 3.7). Thus, the oxidative folding proceeded similarly to the native protein under the conditions used.
Figure 3.7. CD spectra of L-snakin-1 (blue) and iodo-L-snakin-1 (green). Spectra were measured on protein samples at 10 nM in 10 mM phosphate buffer (pH = 6.8). The spectra show distinct α-helical character, but are not as defined as ideal polyAla helices.
3.4 Structure Determination by Experimental Phasing

3.4.1 Crystallisation of quasi-racemic snakin-1

With both iodo-L-snakin-1 and D-snakin-1 in hand, efforts were directed towards crystallisation of quasi-racemic mixtures of snakin-1.

Crystallisation of quasi-racemic snakin-1 was first undertaken by replication of the most reproducible condition for racemic snakin-1 crystallisation (see the previous chapter, §2.5.6). Drops of 0.5 µl/0.5 µl and 1.0 µl/0.5 µl of protein stock/crystallisation condition 0.2 M Na₂HPO₄, 20% w/v PEG 3350 were prepared against a reservoir volume of 100 µl. A quasi-racemic protein stock solution was prepared from a 1:1 mixture of 12.0 mg/ml iodo-L-snakin-1 3.3 stock to 12.5 mg/ml D-snakin-1 2.8 stock. Disappointingly, crystals did not form from these conditions, and so a full robotic crystallisation screen was performed.

The quasi-racemic protein stock was used to screen against a total of 576 conditions described by Moreland et al.,²⁴⁸ and Gorrec²⁴⁹ (see experimental section §3.9.5). Drops were dispensed as 0.3 µl/0.3 µl mixtures of protein stock/crystallisation conditions against 100 µl reservoirs. Within a week multiple promising conditions were found (Figure 3.8) but diffraction quality crystals were not produced.
This initial crystallisation screen for the quasi-racemate was repeated with a matrix microseeding approach (see experimental section §3.9.6).271, 272 A racemic snakin-1 seed stock was prepared from racemic crystals grown from a crystallisation condition of 0.2 M Na$_2$HPO$_4$, 20% w/v PEG 3350, and diluted to a volume of 200 µl with crystallisation buffer (seed stock A). Drops were dispensed as 0.2 µl/0.2 µl/0.05 µl mixtures of quasi-racemic protein stock/crystallisation condition/seed stock A against all six 96-condition screens. Within two days a number of conditions were found, which gave clusters of well-defined crystals (Figure 3.9).
Chapter 3

Figure 3.9. Dark field (A, D) and bright field (B, C) images of crystals from drop microseeded initial crystallisation screen of quasi-racemic snakin-1. Crystallisation conditions: A: 0.2 M tri-sodium citrate dihydrate, 0.1 M Tris HCl pH 8.5, 30% w/v PEG 400; B: 28% w/v MPEG 5000, 0.2 M AMPSO/KOH pH 9.1; C: 30% w/v PEG 1500, 8% v/v MPD; D: 0.2 M lithium sulfate monohydrate, 0.1 M Tris HCl pH 8.5, 30% w/v PEG 4000

Of these initial hits, one condition could be reproduced by hand (28% w/v MPEG 5000, 0.2 M AMPSO/KOH pH 9.1; 0.4 µl/0.4 µl/0.1 µl drops of protein stock/crystallisation buffer/seed stock A prepared against 100 µl reservoirs; see general experimental methods §4.3.2). From this buffer, the crystallisation condition was then optimised.

3.4.2 Optimisation of crystallisation conditions

Fine screening of the microseeded crystallisation condition was undertaken by adjustment of the quantities of AMPSO/KOH and MPEG 5000 in the crystallisation buffer (see general experimental methods §4.3.2). Drops were again set up with a mixture of 0.4 µl/0.4 µl/0.1 µl of quasi-racemic protein stock/crystallisation buffer/seed stock A (100 µl/reservoir). The
The concentration of AMPSO/KOH was varied between 0.10 M, 0.15 M, 0.20 M and 0.25 M in combination with the variation of MPEG content between 26%, 27%, 28%, 29%, 30%, 31% w/v. It was observed during set up that seed stock A contained some precipitate which disappeared over time. This may be phosphate as there is a high concentration of phosphate in the seed stock buffer.

Within a week, crystallisation was observed in a number of conditions but without any clear pattern with regards to the variation of the buffer contents. The condition with the best crystals (26% w/v MPEG 5000, 0.25 M AMPSO/KOH pH 9.1) produced large thin plate crystals similar to those in Figure 3.9c, and a test exposure on the in-house X-Ray source revealed it to proteinaceous. The predicted P1 unit cell was similar to one found for the initial racemic snakin-1 crystals (A = 32.8, B = 43.2, C = 51, α = 94.2, β = 93.7, γ = 100.4) although different enough to suggest isomorphous replacement would not be a phasing option. Additional crystals in this condition were flash-cooled in liquid N2 with a cryoprotectant of 20% w/v MPEG 5000, 0.19 M AMPSO/KOH pH 9.1, 25% v/v ethylene glycol, and sent for data collection at the Australian Synchrotron (Melbourne, Australia).

### 3.4.3 Iodo-SAD data collection

Data collection using quasi-racemic snakin-1 crystals was aimed at collecting iodo-SAD data for experimental phasing. To maximise the anomalous scattering contribution ($f''$) from iodine, data collection was carried out with a beam energy of 8.5 keV ($\lambda = 1.4586$ Å; $f'' = 6.250108$ for iodine) on the MX1 beamline at the Australian Synchrotron. Data were collected from three points on the crystal by translating the X-ray beam along the long edge (Figure 3.10). At each point, 360 frames were collected, each at 1° oscillation and 1 second exposure. After data were collected over all three points, collection was repeated again at each of those same points, for a total of 2160 frames. Over the course of this data collection, it was found that data collected in the second round of collection showed a distinct loss of diffraction, reflected in the value of the scales used in data processing and in $R_{merge}$ values, and consistent with radiation damage (Figures 3.11, 3.12). Scales are applied to symmetry-related and duplicate reflections for internal consistency of the data set, and $R_{merge}$ values are a measure of the deviation of these reflections from their mean intensity ($R_{merge} = \Sigma_{hkl} \Sigma_{i} |I_i(hkl) - \bar{I}(hkl)|/\Sigma_{hkl} \Sigma_{i} I_i(hkl)$). Inspection of diffraction frames from the well-diffracting orientation of the crystal before and after one round of data collection (Frames from the first crystal point and at the same angle in the first and second 360°
collections) find that diffraction spots at resolutions greater than 2.75 Å are visibly weaker (Figures 3.13, 3.14). Moreover, the presence of a poorly diffracting crystal direction (anisotropic diffraction), similar to that observed in previous racemic snakin-1 crystals (§2.5.2), was visually apparent (Figure 3.15). The poor quality diffraction in these images was also reflected in $R_{\text{merge}}$ values for these frames. These low-quality frames ($N = 21-120, 201-300$ for each set) could not be removed without a significant loss in the completeness of the data (Table 3.1, “Truncated data”).

Figure 3.10. Crystal used for iodo-SAD data collection. Crystal was grown from 26% w/v MPEG 5000, 0.25 M AMPSO/KOH pH 9.1.
Figure 3.11. $R_{merge}$ values for symmetry related reflections within each diffraction frame ($N$) for: $A$: data collected at the first crystal position over $2 \times 360^\circ$ collections; $B$: full data measured at three crystal positions over $2 \times 360^\circ$ collections at each. The vertical line at $N = 361$ in $A$ indicates the cut-off between the two collections.

Figure 3.12. Data scales for each frame for: $A$: data collected at the first crystal position over $2 \times 360^\circ$ collections; $B$: full data measured at three crystal positions over $2 \times 360^\circ$ collections at each. The vertical line at $N = 361$ in $A$ indicates the cut-off between the two collections. Mn($\kappa$) is the mean scale factor applied to symmetry-related or duplicate reflections, and is applied for internal consistency of the data. $0\kappa$ is the scale factor without any influence from an input scale.
Figure 3.13. A diffraction frame ($N = 150$) from the first 360° collection. The red ring shown is at 2.75 Å resolution.
Figure 3.14. A diffraction frame ($N = 150$) from the second 360° collection, equivalent to the frame in Figure 3.13. The red ring shown is at 2.75 Å resolution. A number of diffraction spots outside of the red ring are visibly weaker than their counterparts in Figure 3.13.
Figure 3.15. Data frame \((N = 265)\) representative of diffraction along the poorly diffracting crystal direction \((N = 21-120, 201-300\) across each data collection). Diffraction spots are slightly smeary and are poorly defined.
Table 3.1. Data collection and processing statistics for iodo-SAD experiment.

<table>
<thead>
<tr>
<th></th>
<th>Full data</th>
<th>Truncated data&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beaml ine</td>
<td>MX1</td>
<td>MX1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9537</td>
<td>0.9537</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 210r</td>
<td>ADSC Quantum 210r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>2160</td>
<td>960</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td><strong>Unit cell</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>31.23, 43.33, 51.11</td>
<td>31.23, 43.33, 51.11</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>94.07, 93.79, 102.10</td>
<td>94.07, 93.79, 102.10</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>19.41-2.30 (2.38-2.30)</td>
<td>19.41-2.30 (2.38-2.30)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>256406 (25313)</td>
<td>113220 (10891)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.3 (94.8)</td>
<td>62.9 (66.2)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>23.0 (23.2)</td>
<td>15.5 (14.3)</td>
</tr>
<tr>
<td>&lt;I&gt;/σ(I)&lt;</td>
<td>21.0 (7.3)</td>
<td>14.3 (6.6)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt;&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.999 (0.992)</td>
<td>0.990 (0.985)</td>
</tr>
<tr>
<td>Overall B factor from Wilson plot (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>20.8</td>
<td>-57.1</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>95.0 (93.8)</td>
<td>61.4 (64.8)</td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>11.3 (11.6)</td>
<td>7.6 (7.2)</td>
</tr>
<tr>
<td>DelAnom correlation</td>
<td>0.796 (0.279)</td>
<td>0.152 (0.124)</td>
</tr>
<tr>
<td>Estimated no. of molecules in asymmetric unit</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Estimated solvent content (%)</td>
<td>49.4</td>
<td></td>
</tr>
<tr>
<td>V&lt;sub&gt;m&lt;/sub&gt; (Å&lt;sup&gt;3&lt;/sup&gt;/Da)</td>
<td>2.43</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Truncated data refers to a data set including diffraction frames N = 1-20,121-200, 301-360 for every 360° collection.

Despite an anomalous multiplicity of 11.3 and strong indication of an anomalous signal in the DelAnom correlation (Table 3.1, “Full data”) which indicated a solution should be achievable, structure determination by SAD methods using SHELX C/D/E and the AUTORICKSHAW pipeline were unsuccessful. Poor data quality may be to blame for this,
and the poorly diffracting direction of the crystal \((N = 21-120, 201-300\ across\ all\ collections)\) may have compromised the data collection. However, the radiation damage observed inspired a different approach towards experimental phase information: radiation damage-induced phasing (RIP).

### 3.4.4 Radiation damage in macromolecular crystals

X-rays are ionising radiation and as such, the exposure of protein crystals to them commonly causes damage in the protein molecules. When a crystal is exposed to an X-ray beam, only a small proportion of the incident photons contribute to the elastic Rayleigh scattering that produces useful diffraction information for X-ray crystallography. The remainder of photons contribute to inelastic Compton scattering and photo-ionisation effects, both of which can generate highly reactive free-radicals that cause damage to the protein molecules.\(^{274, 275}\)

Radiation damage-induced changes to crystals on the molecular level can result in observable changes of the X-ray diffraction, and these can be easily followed by differences in the data collection statistics between data collected later and those collected earlier in a single collection. These differences arise from disruption of the lattice, through radiation-induced heating effects, and from physical and chemical phenomena.\(^{275}\) The loss of crystal order then results in loss of diffracting power of the crystal, changes in unit cell dimensions, and increases in mosaicity and \(R_{\text{merge}}\) values.\(^{276}\)

In addition to these non-specific changes to overall crystal order, radiation damage manifests as chemical changes in protein molecules. This kind of radiation damage can be categorised as either primary or secondary. Primary damage is the result of direct ionisation of molecules, while secondary damage is reliant on intermediate reactive species in the solvent of the crystal such as solvated electrons, and hydroxyl and hydrogen radicals. Thus, primary damage is a stochastic process dependent on the radiation dose alone, while secondary effects are also dependent on diffusion of the reactive species, and hence temperature. Interestingly, Ravelli and McSweeney\(^{277}\) observed that the frequency of specific chemical changes to disulfides, side-chain carboxyl groups and metal ions was disproportional to the photoabsorption of the elements involved, suggesting that these changes are the consequences primarily of secondary damage.

One standard method which minimises secondary radiation damage is cryo-cooling, a technique that has been used throughout this work when preparing crystals for data
collection (see general experimental methods §4.4.1). Cryo-cooling involves the rapid cooling of crystals in liquid nitrogen to ensure that water molecules in the solvent channels of the crystal remain in a vitreous state and do not crystallise.\textsuperscript{277, 278} Cooling in liquid nitrogen alone is often insufficient to achieve this and a cryo-protectant such as glycerol or ethylene glycol must be added to the mother liquor to suppress formation of ice crystals, which can destroy crystal order and interfere with diffraction.\textsuperscript{277}

Despite these cryo-cooling methods to minimise damage, the use of intense third-generation synchrotron sources in data collection has caused radiation damage to become a routinely observed complication. This can be severely limiting for cases where the potential phasing signal observed is small, and precise measurements are required over a large collection i.e. as in the measurement of an anomalous dispersion signal. Thus, careful collection and beam attenuation is often required for anomalous data collection.

### 3.4.4.1 Radiation damage-induced phasing

Remarkably, in 2003 Ravelli \textit{et al.} showed that the specific chemical changes over-represented in radiation damage, that is loss of disulfide bonds and carboxylic acids, can be exploited to solve the phase problem in a technique called radiation damage-induced phasing (RIP).\textsuperscript{276} RIP requires two data sets, one set prior to radiation damage, and one after specific damage has occurred in the protein. The data are typically treated as a single isomorphous replacement experiment, where the before-damage set is used as pseudo-native data. Nanao, Sheldrick and Ravelli have further expanded on this approach and have proposed that treating the data as a multiple isomorphous replacement experiment without a native set ($F_{\text{const}}$), where the before ($F_{\text{before}}$) and after-damage sets ($F_{\text{after}}$) are treated as derivative data, may provide more accurate RIP substructures (Figure 3.16).\textsuperscript{279}
Currently, there is no general method to determine when sufficient radiation damage has been achieved for successful phasing, but changes in unit cell dimensions, increases in mosaicity and $R_{\text{merge}}$ values, can be useful indicators of radiation damage.\textsuperscript{276} Like traditional isomorphous replacement methods, success using this technique requires isomorphism between the data sets. Thus, a balance is required between generating sufficient radiation damage to susceptible sites for phasing, and causing excessive radiation damage resulting in extensive crystal disorder, something which may not be possible in practice. Lastly, an empirically determined scale factor $K$ is often required to compensate for the loss of diffracting power that accompanies radiation damage, and fine tuning this parameter can be critical to success.\textsuperscript{279}

Building on this technique, Nanao and Ravelli found that UV irradiation can likewise be used to damage specific sites in macromolecular crystals.\textsuperscript{280} UV-radiation cannot cause the ionisation of water molecules, but is known to cause photoexcitation of aromatic residues, releasing solvated electrons and resulting in photo-reduction of proximal disulfides.\textsuperscript{279, 281, 282} As such, UV-irradiation has been suggested as a method by which RIP can be methodically utilised, allowing intense irradiation to modify sites without excessive damage to crystals. UV irradiation has also been found to cleave the C-Se bond in selenomethionine, providing an experimental phasing technique for untunable X-ray sources where a UV beam is available.\textsuperscript{283}
3.4.5 UV-RIP experiments on racemic snakin-1 crystals

Failure of attempts to use SAD phasing to solve the crystal structure of quasi-racemic snakin-1 led to the consideration of UV-RIP as an alternative for structure solution, as it was an attractive alternative to anomalous dispersion methods that require high redundancy data. Success with RIP methods has been reported on data sets with a multiplicity as low as 0.5, particularly desirable for crystal forms in $P1/P\bar{1}$, where the lack of crystallographic symmetry operations necessitates collection across large angles for a complete data set.\textsuperscript{284} Another advantage to using UV-RIP is that it does not require an anomalous signal, so the optimised true racemic snakin-1 crystals can be used.

Thus, a UV-RIP experiment was carried out on a previously prepared crystal of true racemic snakin-1 (§2.5.6). Two data sets were collected at a wavelength of 0.9537 Å, separated by a 30 min UV exposure or “burn” (MX1 beamline, Australian Synchrotron). Data collection was performed with the intention of obtaining at least a crude model that could be used to solve our previous high-resolution data sets. Toward this aim, data were collected slightly over the minimum 180° range that would produce a complete data set, to limit non-specific radiation damage. Both the before and after UV burn data sets were collected at the same point on the crystal and over a 200° rotation range. Each diffraction image collected was over a 1° oscillation with a one second exposure time (Table 3.2).
Table 3.2. UV-RIP data collection and UV burn statistics.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Before UV burn</th>
<th>After UV burn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MX1</td>
<td>MX1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9537</td>
<td>0.9537</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 210r</td>
<td>ADSC Quantum 210r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>31.18, 43.28, 51.02</td>
<td>31.17, 43.27, 51.01</td>
</tr>
<tr>
<td>$a, \beta, \gamma$ (°)</td>
<td>93.65, 94.29, 104.16</td>
<td>93.65, 94.36, 104.29</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>33.69-2.20 (2.27-2.20)</td>
<td>33.68-2.20 (2.27-2.20)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>12833 (1095)</td>
<td>12820 (1100)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.4 (97.5)</td>
<td>98.4 (97.4)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.2 (2.2)</td>
<td>2.2 (2.2)</td>
</tr>
<tr>
<td>$R_{\text{merge}}^a$ on intensity (%)</td>
<td>13.5 (44.4)</td>
<td>14.8 (48.6)</td>
</tr>
<tr>
<td>$&lt;</td>
<td>I/\sigma I</td>
<td>&gt;$</td>
</tr>
<tr>
<td>$CC_{1/2}^b$</td>
<td>0.985 (0.821)</td>
<td>0.984 (0.764)</td>
</tr>
</tbody>
</table>

UV burn

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (nm)</td>
<td>266</td>
</tr>
<tr>
<td>Pulse repetition rate (kHz)</td>
<td>22</td>
</tr>
<tr>
<td>Average power (mW)</td>
<td>30</td>
</tr>
<tr>
<td>Beam divergence (mrad)</td>
<td>7</td>
</tr>
</tbody>
</table>

[values of the outermost resolution shell are in parentheses]

\[R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl)| - \bar{I}(hkl)| \sum_{hkl} \sum_i I_i(hkl)\]

\[CC_{1/2} = \text{Pearson correlation coefficient between random half sets of the data}\]

The crystal used for UV-RIP again showed a poorly diffracting direction, apparent in frames $N = 65-140$ (Figure 3.19). The data processing statistics of the two data sets do not differ greatly, but there are apparent differences in the $R_{\text{merge}}, CC_{1/2}$, and $<|I/\sigma I|>$ (Table 3.2), suggesting the possibility of radiation damage. This possibility is supported by visual differences in the intensity of diffraction spots in frames before and after the UV burn, mainly in the higher resolution reflections (Figures 3.17, 3.18).

Structure solution by UV-RIP phasing was undertaken using the AUTORICKSHAW RIP module, with variation of the scale factor $K$ between 0.92 and 0.99. All of these trials resulted in values for pseudo-free $CC$ greater than 65%, indicating that a real and usable RIP signal exists, but at the same time, all trials were unsuccessful in building believable
models in *ARP/wARP* (Table 3.3). These phasing trials were extended to systematic trials in *SHELXC/D/E* using parameters based on those in *AUTORICKSHAW*, and were carried out with variation of the scale factor between 0.920 and 0.980 in increments of 0.001. All trials were unsuccessful. A possible reason for the failure to obtain a correct solution is that the relatively small signal from the radiation-susceptible sites could not be accurately determined with the low multiplicity data obtained here. These errors would then be carried over into an inaccurate substructure, providing poor phases.
Chapter 3

Figure 3.17. Diffraction frame ($N = 1$) before the UV-burn data. A selection of diffraction spots that show visual differences after the UV-burn are circled in red.
Figure 3.18. Diffraction frame ($N = 1$) after the UV-burn. A selection of diffraction spots that show visual differences after the UV-burn are circled in red.
Figure 3.19. Diffraction frame from the poorly ordered region of the crystal ($N = 65-140$). Diffraction spots are smeary and not well defined.
Table 3.3. Phasing statistics from AUTORICKSHAW UV-RIP trials employing SHELX/C/D/E and ARP/wARP.

<table>
<thead>
<tr>
<th>Scale factor $K$</th>
<th>0.919</th>
<th>0.958</th>
<th>0.962</th>
<th>0.966</th>
<th>0.970</th>
<th>0.974</th>
<th>0.978</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIND</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Resolution limit for SHELXD</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>$CC_{\text{best}}$†</td>
<td>36.27</td>
<td>36.65</td>
<td>36.01</td>
<td>36.03</td>
<td>36.35</td>
<td>35.98</td>
<td>35.88</td>
<td>35.32</td>
</tr>
<tr>
<td>$CC_{\text{weak}}$†</td>
<td>13.19</td>
<td>11.57</td>
<td>10.77</td>
<td>9.19</td>
<td>10.32</td>
<td>10.18</td>
<td>11.24</td>
<td>10.77</td>
</tr>
<tr>
<td>PATFOM†</td>
<td>24.31</td>
<td>29.46</td>
<td>27.06</td>
<td>28.83</td>
<td>26.60</td>
<td>28.71</td>
<td>27.51</td>
<td>27.12</td>
</tr>
<tr>
<td>Solvent content</td>
<td>0.47</td>
<td>0.382</td>
<td>0.373</td>
<td>0.359</td>
<td>0.356</td>
<td>0.366</td>
<td>0.367</td>
<td>0.375</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.627</td>
<td>0.635</td>
<td>0.637</td>
<td>0.637</td>
<td>0.632</td>
<td>0.638</td>
<td>0.638</td>
<td>0.641</td>
</tr>
<tr>
<td>Pseudo-free $CC$ (%)</td>
<td>65.12</td>
<td>67.16</td>
<td>69.22</td>
<td>69.53</td>
<td>69.53</td>
<td>69.81</td>
<td>69.46</td>
<td>68.47</td>
</tr>
<tr>
<td>Residues built by ARP/wARP</td>
<td>113</td>
<td>83</td>
<td>93</td>
<td>91</td>
<td>85</td>
<td>83</td>
<td>101</td>
<td>89</td>
</tr>
</tbody>
</table>

[†Solution from SHELXD with highest $CC_{\text{best}}$]
3.4.6 Preparation of bromide-soaked crystals for MAD phasing

As UV-RIP appeared unsuccessful, efforts were directed towards another option for experimental phasing with quasi-racemic snakin-1 crystals. MAD experiments require a much lower data redundancy compared to SAD, desirable for crystals in P1. However, the X-ray absorption edges of iodine are not within the 8.5 – 17.5 keV range accessible on the Australian Synchrotron beamlines. The K edge of bromine is within this range, and many examples of successful phasing by Br-MAD from bromide soaked crystals have been reported (Figure 3.20). Bromide ions do not require specific chemical moieties for ordered arrangement in protein crystals, and bromide ions can occupy sites normally occupied by water molecules in non-soaked crystals. Moreover, short cryo-soaks with a bromide-containing cryo-protectant immediately prior to cryo-cooling, are sufficient for the incorporation of bromide ions for MAD phasing. A strategy of using quasi-racemic snakin-1 crystals with a short bromide soak to provide a MAD signal was attractive, and this method was pursued for the next phasing strategy.

Figure 3.20. Bromine $f'$ and $f''$ contributions plotted as functions of wavelength. Plot generated using the Edgeplots web tool (http://skuld.bmsc.washington.edu/scatter) with data calculated using the subroutine library by Brennan and Cowan. The lines shown are on either side of the wavelengths accessible at the Australian Synchrotron.

Initially, further fine screening of quasi-racemic crystallisation conditions was carried out by variation of the crystallisation buffer condition. The MPEG content was varied between 24%, 25%, 26% w/v, and the AMPSO/KOH pH 9.1 concentration between 0.15, 0.20, 0.25, 0.30, 0.40 M. Crystallisation was undertaken using both seed stock A used previously, and fresh seed stock (seed stock B) prepared from crystals in the first drop microseeded fine
screen described previously (§3.4.2). Two of the conditions produced poor crystals using seed stock \( A \), and these were not suitable for X-Ray diffraction. Conditions employing seed stock \( B \) did not produce any crystals.

Due to the apparent difficulty in producing crystals with this fine screen, the reproducibility of the crystallisation conditions was assessed. The optimised condition (26% w/v MPEG 5000, 0.25 M AMPSO/KOH pH 9.1) was repeated again with both seed stocks \( A \) and \( B \), but did not produce any crystals. Possible reasons for the failure with both seed stocks are that seed stock \( A \) may have degraded over time, and that seed stock \( B \) was prepared with potentially sub-optimal crystals grown from seed stock \( A \). The lack of reproducibility of this condition led to re-optimisation with non-seeded crystallisation conditions.

The conditions considered for optimisation at this point were those from the initial crystallisation screen with quasi-racemic snakin-1 (Figure 3.8). Reproducibility of these conditions was first assessed, by setting up drops containing 0.5 µl/0.5 µl, 0.5 µl/1.0 µl, and 1.0 µl/0.5 µl of protein stock/crystallisation buffer (100 µl reservoir). Crystals were formed with 1.0 µl/0.5 µl drops against 4.0 M sodium formate and with 0.5 µl/0.5 µl and 1.0 µl/0.5 µl drops against 1.6 M sodium citrate at pH 6.5. Further fine screening of different sodium formate conditions found that drops set up in 24-well plates (1.0 µl/2.0 µl protein stock/crystallisation buffer, 500 µl reservoir; crystallisation buffer: 3.9 M sodium formate) produced better-defined crystal forms (Figure 3.21). These crystals were used for bromide soaks and MAD data collection.

Figure 3.21. Bright field image of quasi-racemic snakin-1 crystals. Crystals were grown against a 3.9 M sodium formate buffer.
Quasi-racemic crystals were flash-cooled in liquid N\textsubscript{2} after a 10 second quick soak\textsuperscript{286} in:

\begin{itemize}
  \item [a)] a cryo-protectant of 2.9 M sodium formate, 25\% \textit{v/v} ethylene glycol and 1 M NaBr;
  \item [b)] a cryo-protectant of 2.9 M sodium formate, 25\% \textit{v/v} ethylene glycol and 0.5 M NaBr;
  \item [c)] a cryo-protectant of 2.9 M sodium formate, 25\% \textit{v/v} ethylene glycol and 1.0 M NaBr, followed by a 10 second back-soak in 2.9 M sodium formate, 25\% \textit{v/v} ethylene glycol.
\end{itemize}

These cryo-cooled crystals were sent to the Australian Synchrotron for data collection.

### 3.4.7 MAD data collection

For MAD data collection, a crystal soaked in bromide followed by a back-soak (§3.4.6, soaking condition \textit{c}; Figure 3.23) was chosen. To confirm the presence of bromide in the back-soaked crystal, an X-Ray fluorescence scan was carried out. The fluorescence scan showed the characteristic fluorescence peaks for bromine and a small peak corresponding to iodine, indicating successful soaking. From this fluorescence spectrum, three wavelengths corresponding to the Br inflection point (0.9199 Å), low-energy remote (0.9919 Å), and high-energy remote (0.9184 Å), were chosen for X-ray data collection (Figure 3.22).

![Figure 3.22. Br-MAD-wavelengths (black lines) on a plot of $f'$ and $f''$ contributions as functions of wavelength. Plot generated using the Edgeplots web tool (http://skuld.bmsc.washington.edu/scatter) with data calculated using the subroutine library by Brennan and Cowan.\textsuperscript{287} The lines shown are on either side of the wavelengths accessible at the Australian Synchrotron.](http://skuld.bmsc.washington.edu/scatter)

\[ \text{Figure 3.22. Br-MAD-wavelengths (black lines) on a plot of } f' \text{ and } f'' \text{ contributions as functions of wavelength. Plot generated using the Edgeplots web tool (http://skuld.bmsc.washington.edu/scatter) with data calculated using the subroutine library by Brennan and Cowan.} \]
Data at the inflection point were collected first (INFL data; 400 frames, 1°/1 second), followed by the low-energy remote with a starting angle offset of 0.5° (LREM data; 300 frames, 1°/1 second). The beam was then translated along the crystal for collection of the high-energy remote set, with a starting angle offset by -39.75° in comparison to the inflection point data collection (HREM data; 453 frames, 1°/1 second). All data sets were collected with a high level of beam attenuation (98%) to limit radiation damage (Table 3.4). A representative diffraction frame (Figure 3.24) shows distinct diffraction spots and also the fine structure of the CCD detector, indicating that the signal is at the detection limit for the detector. Thus, the level of attenuation appears appropriate for minimising radiation damage while maintaining a good diffraction signal.

Figure 3.23. Crystal used for MAD data collection. Crystal was grown from 3.9 M sodium formate.
Table 3.4. MAD data collection and processing statistics.

<table>
<thead>
<tr>
<th></th>
<th>Inflection point</th>
<th>Low-energy remote</th>
<th>High-energy remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MX2</td>
<td>MX2</td>
<td>MX2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9199</td>
<td>0.9919</td>
<td>0.9184</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 315r</td>
<td>ADSC Quantum 315r</td>
<td>ADSC Quantum 315r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>400</td>
<td>300</td>
<td>453</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>31.23, 37.42, 50.30</td>
<td>31.26, 37.44, 50.39</td>
<td>31.24, 37.43, 50.27</td>
</tr>
<tr>
<td>$a$, $\beta$, $\gamma$ (°)</td>
<td>93.00, 90.58, 102.57</td>
<td>92.96, 90.55, 102.57</td>
<td>92.99, 90.59, 102.64</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.12</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.46-1.51 (1.53-1.51)</td>
<td>36.46-1.57 (1.60-1.57)</td>
<td>36.47-1.63 (1.66-1.63)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>34139 (1635)</td>
<td>29972 (1408)</td>
<td>26741 (1281)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.0 (93.2)</td>
<td>96.6 (93.7)</td>
<td>97.4 (93.7)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.3 (4.1)</td>
<td>3.3 (3.2)</td>
<td>4.9 (4.7)</td>
</tr>
<tr>
<td>$R_{\text{merge}}$ on intensity (%)</td>
<td>10.4 (66.5)</td>
<td>8.0 (64.8)</td>
<td>9.8 (66.0)</td>
</tr>
<tr>
<td>$&lt;I/\sigma I&gt;$</td>
<td>9.5 (2.0)</td>
<td>9.4 (1.8)</td>
<td>10.3 (2.2)</td>
</tr>
<tr>
<td>$CC_{1/2}$</td>
<td>0.997 (0.842)</td>
<td>0.998 (0.811)</td>
<td>0.998 (0.853)</td>
</tr>
<tr>
<td>Overall $B$ factor from Wilson plot (Å$^2$)</td>
<td>7.7</td>
<td>9.9</td>
<td>10.0</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>89.2 (78.6)</td>
<td>90.4 (83.0)</td>
<td>89.9 (82.8)</td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>1.1 (1.2)</td>
<td>1.1 (1.2)</td>
<td>1.1 (1.2)</td>
</tr>
<tr>
<td>DelAnom correlation</td>
<td>0.123 (-0.233)</td>
<td>-0.065 (0.000)</td>
<td>-0.090 (-0.102)</td>
</tr>
</tbody>
</table>

[values of the outermost resolution shell are in parentheses]

a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \bar{I}(hkl)|/\sum_{hkl} \sum_i I_i(hkl)$

b Pearson correlation coefficient between random half sets of the data
Figure 3.24. Data frame representative of diffraction for MAD data collection. The fine structure of the detector is clearly visible as four distinct panels in each section of the 3 x 3 array.
Visual inspection of the diffraction data showed that the crystal appeared well-ordered throughout the collection range without the anisotropic diffraction seen in many of the previous crystals. Structure solution was carried out using both the two- and three-wavelength MAD protocols of the EMBL AUTORICKSHAW pipeline, and SHELXC/D/E suite independently. The output from SHELXC is shown below (Tables 3.5, 3.6, 3.7). A $<d''/\text{sig}>$ value of 0.8 corresponds to zero anomalous signal, suggesting that signal is absent in all data sets except for the inflection data, for which there is some signal at resolutions of 3.75 Å and below. Whether this anomalous signal is real can be disputed, as data processing by AIMLESS shows a poor correlation of the anomalous signal in the data (Table 3.4).

Table 3.5. SHELXC analysis of inflection data set. A $<d'/\text{sig}>$ or $<d''/\text{sig}>$ value of 0.8 corresponds to zero signal.

| 34033 Unique reflections, highest resolution 1.506 Ångstroms |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Resl. | Inf. | 8.09 | 4.98 | 3.75 | 3.07 | 2.62 | 2.31 | 2.07 | 1.99 | 1.74 | 1.61 | 1.51 |
| N(data) | 219 | 736 | 1293 | 1841 | 2473 | 3023 | 3663 | 4147 | 4809 | 5743 | 6086 |
| Chi-sq | 0.50 | 1.16 | 1.93 | 2.04 | 1.23 | 1.49 | 1.44 | 1.51 | 1.58 | 1.62 | 1.65 |
| $<I/\text{sig}>$ | 19.4 | 17.5 | 19.0 | 17.5 | 13.3 | 10.8 | 8.9 | 6.4 | 3.9 | 2.6 | 1.8 |
| %Complete | 97.8 | 99.2 | 98.6 | 98.6 | 98.2 | 98.1 | 97.5 | 97.1 | 96.7 | 95.7 | 94.8 |
| $<d''/\text{sig}>$ | 0.90 | 0.94 | 0.80 | 0.82 | 0.81 | 0.84 | 0.85 | 0.80 | 0.83 | 0.80 | 0.72 |
| CC(1/2) | 47.0 | 9.5 | 8.0 | 8.2 | 11.0 | -4.9 | 1.9 | 4.6 | 11.4 | -4.7 | -18.6 |

Table 3.6. SHELXC analysis of low-energy remote data set.

| 29896 Unique reflections, highest resolution 1.569 Ångstroms |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Resl. | Inf. | 8.41 | 5.18 | 3.90 | 3.19 | 2.73 | 2.40 | 2.15 | 1.96 | 1.81 | 1.68 | 1.57 |
| N(data) | 194 | 650 | 1129 | 1650 | 2157 | 2699 | 3319 | 3708 | 4169 | 4846 | 5375 |
| Chi-sq | 0.86 | 1.02 | 1.12 | 1.05 | 1.52 | 1.40 | 1.32 | 1.38 | 1.68 | 1.61 | 1.62 |
| $<I/\text{sig}>$ | 21.4 | 19.4 | 21.1 | 20.1 | 15.5 | 12.5 | 10.5 | 7.8 | 4.9 | 3.0 | 2.0 |
| %Complete | 96.0 | 98.3 | 97.9 | 97.8 | 97.9 | 97.7 | 97.4 | 96.7 | 96.4 | 95.6 | 94.0 |
| $<d''/\text{sig}>$ | 0.76 | 0.69 | 0.69 | 0.73 | 0.75 | 0.77 | 0.78 | 0.74 | 0.75 | 0.76 | 0.71 |
| CC(1/2) | 36.0 | -6.9 | -29.8 | -30.7 | -15.9 | 2.0 | -3.7 | -6.7 | -16.4 | 11.7 | -37.7 |

Table 3.7. SHELXC analysis of high-energy remote data set.
All MAD trials were unsuccessful in solving the structure of snakin-1. Moreover, the apparent lack of anomalous signal indicated that alternative methods would be required to solve the structure.

### 3.4.8 Structure solution by RIP

An observation on further inspection of the MAD data that the data processing statistics for each data set were slightly different suggested phasing could be attempted using radiation damage. In comparing the low-energy remote data to the inflection point data, increases in the unit cell dimensions and mosaicity were observed (although very modest), features associated with radiation damage (§3.4.4, Table 3.8). Encouraged by these observations, structure determination attempts via X-ray RIP methods were carried out with these data.

#### Table 3.8. Differences in processing statistics for INFL and LREM data, with a high-resolution cut-off of 1.57 Å in both data sets for comparison.

<table>
<thead>
<tr>
<th></th>
<th>Inflection point</th>
<th>Low-energy remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation range per image (°)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>220</td>
<td>220</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>31.23, 37.42, 50.30</td>
<td>31.26, 37.44, 50.39</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma) (°)</td>
<td>93.00, 90.58, 102.57</td>
<td>92.96, 90.55, 102.57</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.46-1.57 (1.60-1.57)</td>
<td>36.49-1.57 (1.60-1.57)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>30086 (1431)</td>
<td>29896 (1450)</td>
</tr>
<tr>
<td>(R_{merge})^a on intensity (%)</td>
<td>8.0 (47.0)</td>
<td>6.8 (51.3)</td>
</tr>
<tr>
<td>(&lt;I/\sigma I&gt;)</td>
<td>7.9 (2.0)</td>
<td>8.4 (1.7)</td>
</tr>
</tbody>
</table>

(values of the outermost resolution shell are in parentheses)

\[ R_{merge} = \sum_{hkl} \sum_i |I_i(hk\ell) - \bar{I}(hk\ell)| / \sum_{hkl} \sum_i I_i(hk\ell) \]

The EMBL AUTORICKSHAW RIP protocol was used for structure solution, with the inflection data as the before-damage set, and the low-energy remote as the after-damage set. To minimise the radiation damage within each data set, thus maximising the signal difference between sets, only the first 220 frames for both the before-damage and after-damage data were used. This data cut off was evaluated by the change in \(R_{merge}\) values for each frame over the complete data (Figure 3.25).
The *AUTORICKSHAW* pipeline successfully built a model of 235 residues traced by *ARP/wARP* (Table 3.9). Examination of these initial maps and models showed that the Cα backbones of all four monomers were modelled, with the D- monomers modelled as correctly-oriented chains of poly-Gly. The density of each disulfide bond was well defined, as well as that of each iodophenylalanine (Figure 3.26).

Table 3.9. Phasing statistics from *AUTORICKSHAW* RIP trials employing *SHELXC/D/E* and *ARP/wARP*, with data at two wavelengths.

<table>
<thead>
<tr>
<th>FIND</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale factor $K$</td>
<td>0.980</td>
</tr>
<tr>
<td>Resolution limit for <em>SHELXD</em></td>
<td>1.90</td>
</tr>
<tr>
<td>$CC_{\text{best}}^\dagger$</td>
<td>26.30</td>
</tr>
<tr>
<td>$CC_{\text{weak}}^\dagger$</td>
<td>11.93</td>
</tr>
<tr>
<td>PATFOM$^\dagger$</td>
<td>12.39</td>
</tr>
<tr>
<td>Solvent content</td>
<td>0.39</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.318</td>
</tr>
<tr>
<td>Connectivity</td>
<td>0.710</td>
</tr>
<tr>
<td>Pseudo-free $CC$ (%)</td>
<td>76.74</td>
</tr>
<tr>
<td>Residues built in <em>ARP/wARP</em></td>
<td>235</td>
</tr>
</tbody>
</table>

[$^\dagger$Solution from *SHELXD* with highest $CC_{\text{best}}$]
<table>
<thead>
<tr>
<th>Data processing</th>
<th>Before-damage (inflection point data)</th>
<th>After-damage (low-energy remote data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MX2</td>
<td>MX2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9199</td>
<td>0.9919</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 315r</td>
<td>ADSC Quantum 315r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>220</td>
<td>220</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell</td>
<td>a, b, c (Å)</td>
<td>31.23, 37.42, 50.30</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.46-1.51 (1.53-1.51)</td>
<td>36.49-1.57 (1.60-1.57)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>34033 (1632)</td>
<td>29896 (1450)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.8 (93.0)</td>
<td>96.5 (93.2)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.4 (2.3)</td>
<td>2.4 (2.4)</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;a&lt;/sup&gt; on intensity (%)</td>
<td>8.9 (56.6)</td>
<td>6.8 (51.3)</td>
</tr>
<tr>
<td>&lt;I/σI&gt;</td>
<td>7.1 (1.6)</td>
<td>8.4 (1.7)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.995 (0.770)</td>
<td>0.997 (0.789)</td>
</tr>
<tr>
<td>Overall B factor from Wilson plot (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>9.3</td>
<td>11.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th>Before-damage (inflection point data)</th>
<th>After-damage (low-energy remote data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>36.46-1.51</td>
<td>36.49-1.57</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.7</td>
<td>96.4</td>
</tr>
<tr>
<td>No. of reflections, working set</td>
<td>32341</td>
<td>28393</td>
</tr>
<tr>
<td>No. of reflections, test set</td>
<td>1667</td>
<td>1479</td>
</tr>
<tr>
<td>Final R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>0.179</td>
<td>0.207</td>
</tr>
<tr>
<td>Final R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.226</td>
<td>0.243</td>
</tr>
<tr>
<td>No. of non-H atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1917</td>
<td>1925</td>
</tr>
<tr>
<td>Ligand</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>Water</td>
<td>265</td>
<td>247</td>
</tr>
<tr>
<td>Total</td>
<td>2220</td>
<td>2196</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.04</td>
<td>1.05</td>
</tr>
<tr>
<td>Average B factors (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>12.3</td>
<td>13.7</td>
</tr>
<tr>
<td>Ligand</td>
<td>17.9</td>
<td>17.5</td>
</tr>
<tr>
<td>Water</td>
<td>18.9</td>
<td>21.0</td>
</tr>
<tr>
<td>Ramachandran plot&lt;sup&gt;†&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured (%)</td>
<td>97.6</td>
<td>98.4</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Molprobity clash score</td>
<td>1.34</td>
<td>0.54</td>
</tr>
<tr>
<td>PDB code</td>
<td>5e5t</td>
<td>5e5y</td>
</tr>
</tbody>
</table>

<sup>a</sup> Rmerge = Σ<sub>hkI</sub> Σ<sub>I</sub>[I(hkI) − I(hkI)] / Σ<sub>hkI</sub> Σ<sub>I</sub>I(hkI)

<sup>b</sup> Pearson correlation coefficient between random half sets of the data

(values of the outermost resolution shell are in parentheses)

[†For models with D- monomers omitted]
Figure 3.26. Snakin-1 model with $2F_o - F_c$ (blue) and $F_o - F_c$ (green and red) electron density at 1.0 $\sigma$. The side chain of the iodophenylalanine residue is displayed, showing distinct fourier ripples as difference density around the iodine.
Iterative cycles of manual rebuilding in COOT and refinement against the after-damage data in PHENIX.REFINE up to a resolution of 1.57 Å was carried out, after which a final model was obtained with $R$-values of $R_{\text{work}}/R_{\text{free}} = 0.21/0.24$ (Table 3.10, after-damage data; PDB code: 5e5y). Model validation using MOLPROBITY gave a clash score of 0.54 (100th percentile for structures at 1.57 ± 0.25 Å, Table 3.10). Ramachandran statistics of only the L- monomers in the model showed no outliers and indicated that 98.4% of the residues adopted the most favoured torsion angles.

Visual inspection of the final refined model and maps to identify the radiation damage sites used for phasing showed that 20 of the 39 sites corresponded to features in the model. Interestingly, the top four sites were at the iodine atoms of the two iodophenylalanine residues in the L- monomers, and the corresponding tyrosine hydroxyl groups in the D- monomers. The intensities of the sites were comparable, with the sites at iodine positions slightly stronger (35 and 31 σ) than those at the hydroxyl positions (31 and 27 σ). Radiation damage at a carbon-iodine bond would be expected to give a sizeable signal, but damage at the tyrosine hydroxyls is unexpected. One possible explanation is that the crystals may display positional disorder, with a proportion of D- monomers replaced in the lattice by iodo-L-snakin-1 monomers, which are then susceptible to radiation damage. The remaining identifiable sites correspond to disulfide bonds and the carboxyl group of C-terminal prolines, both of which are sites that can be damaged through radiation exposure.

Using these 39 sites, a simple assessment of the minimum requirement for a correct solution was carried out. Multiple runs of SHELXE and ARP/wARP autobuilding were carried out, in which the sites used for phasing were varied. The sites chosen for these runs are summarised in Table 3.11. From these trials, it appears that the two iodine and corresponding hydroxyl sites are essential for successful phasing and autobuilding (Table 3.11, “Top 4 sites” and “All sites”). The initial density-modified maps obtained with the successful runs were well defined, and rebuilding and refinement could be readily completed to give the final refined structure. In comparison, all trials that did not include all four sites could not provide good starting phases for model building, demonstrating the phasing power of the iodine atoms. This is not a comprehensive attempt to analyse the data in this way, and it is possible that other solutions may be obtained through advanced phasing and refinement strategies.
Table 3.11. Phasing statistics from SHELXE and ARP/wARP, using different numbers of sites located by AUTORICKSHAW.

<table>
<thead>
<tr>
<th>Sites†</th>
<th>Top site (iodine)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast</td>
<td>0.328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.681</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>64.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.304/0.621</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites†</td>
<td>Top 2 sites (iodine)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>0.315</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.676</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>62.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.318/0.624</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites†</td>
<td>Top 4 sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>0.391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.743</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>75.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.292/0.386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites†</td>
<td>Top 20 sites, excluding top 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>0.301</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>62.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.3385/0.601</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites†</td>
<td>All sites, excluding top 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>0.291</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.678</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>64.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.315/0.619</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sites†</td>
<td>All 39 sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>0.376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity</td>
<td>0.740</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudo-free CC (%)</td>
<td>74.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residues built in ARP/wARP</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{\text{work}}/R_{\text{free}}$ of model</td>
<td>0.296/0.383</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[†Sites used were chosen in order of occupancy]
Finally, to assess whether the difference in the wavelengths between data sets was important for structure solution, RIP was carried out with the individual data sets. One issue with using multi-wavelength data for RIP solution is in determining whether the experiment was truly a RIP experiment, or if the signal was provided by wavelength-dependent scattering (Figure 3.27). To explore this question, RIP experiments were carried out with single-wavelength data. Each data set was divided into two sets containing an equal number of frames collected at the start and end of collection (Table 3.12). Structure solution on these data was then carried out with AUTORICKSHAW.

![Figure 3.27. Comparison of the $f'$ and $f''$ contributions of iodine at the wavelengths used for RIP, plotted as functions of wavelength. A significant increase in the predicted $f''$ is observed with an increase in the wavelength, but there is an insignificant change in $f'$. Plot generated using the Edgeplots web tool (http://skuld.bmsc.washington.edu/scatter) with data calculated using the subroutine library by Brennan and Cowan.\textsuperscript{287} Lines indicate the wavelength of the inflection point and low-energy remote data.](image)

Of the trials carried out, only one (Table 3.12, entry 6) was clearly successful, with 232 residues traced in ARP/WARP. Remarkably, the only difference between entries 4, which was unsuccessful, and the successful entry 6, is a gap of 50 diffraction frames between the before and after set. This illustrates the very sensitive nature of the RIP experiment, and the difficulty in determining a point where there is sufficient damage for successful phasing.

Overall, this successful entry shows that multiple wavelength data was not required for the solution of the quasi-racemic snakin-1 structure, and strongly suggests that the susceptibility of this crystal to radiation is solely responsible for providing the relevant phase information.
Table 3.12. Phasing statistics from AUTORICKSHAW RIP trials employing SHELXC/D/E and ARP/wARP, with single-wavelength data. Data used are from single wavelength collections that have been divided into before (early collection) and after-damage (late collection) data.

<table>
<thead>
<tr>
<th>Entry</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before data</td>
<td>INFL N = 1-200</td>
<td>INFL N = 1-150</td>
<td>LREM N = 1-150</td>
<td>HREM N = 1-200</td>
<td>HREM N = 1-225</td>
<td>HREM N = 1-200</td>
</tr>
<tr>
<td>After data</td>
<td>INFL N = 201-400</td>
<td>INFL N = 251-400</td>
<td>LREM N = 151-300</td>
<td>HREM N = 201-400</td>
<td>HREM N = 226-450</td>
<td>HREM N = 251-450</td>
</tr>
<tr>
<td>FIND</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Scale factor $K$</td>
<td>0.988</td>
<td>0.988</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>Resolution for SHELXD</td>
<td>2.01</td>
<td>2.005</td>
<td>2.07</td>
<td>1.80</td>
<td>1.90</td>
<td>1.90</td>
</tr>
<tr>
<td>$CC_{\text{best}}$†</td>
<td>19.64</td>
<td>20.11</td>
<td>21.17</td>
<td>18.96</td>
<td>19.54</td>
<td>19.66</td>
</tr>
<tr>
<td>$CC_{\text{weak}}$†</td>
<td>5.30</td>
<td>3.80</td>
<td>5.11</td>
<td>4.57</td>
<td>6.56</td>
<td>5.66</td>
</tr>
<tr>
<td>PATFOM†</td>
<td>9.65</td>
<td>11.02</td>
<td>12.95</td>
<td>8.38</td>
<td>10.43</td>
<td>10.42</td>
</tr>
<tr>
<td>Solvent content</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.279</td>
<td>-</td>
<td>0.343</td>
<td>0.347</td>
<td>0.307</td>
<td>0.332</td>
</tr>
<tr>
<td>Connectivity</td>
<td>0.654</td>
<td>-</td>
<td>0.655</td>
<td>0.661</td>
<td>0.666</td>
<td>0.709</td>
</tr>
<tr>
<td>Pseudo-free CC (%)</td>
<td>62.17</td>
<td>-</td>
<td>59.54</td>
<td>62.20</td>
<td>64.85</td>
<td>74.56</td>
</tr>
<tr>
<td>Residues built in</td>
<td>ARP/wARP 31</td>
<td>-</td>
<td>34</td>
<td>21</td>
<td>37</td>
<td>232</td>
</tr>
</tbody>
</table>

[†Solution from SHELXD with highest $CC_{\text{best}}$]
3.5 Crystal structure of snakin-1

3.5.1 Crystal structure of quasi-racemic snakin-1

The quasi-racemic snakin-1 crystal has space group \( P1 \), with two molecules of iodo-L-snakin-1 and two of D-snakin-1 in the asymmetric unit. Within the unit cell, each L-monomer is related through a pseudo-inversion centre to one of the two D-monomers.

Figure 3.28. Crystal structure of quasi-racemic snakin-1 in \( P1 \). Iodo-L-snakin-1 molecules are shown in green and D-snakin-1 molecules in orange. Pseudo-inversion centres are represented by grey spheres, and iodine atoms by purple spheres.
By shifting the unit cell origin to coincide with one of these pseudo-inversion centres, it becomes apparent that the space group is pseudo-$P\bar{1}$, and that the crystal deviates from true inversion symmetry only in the iodine of the iodophenylalanine residues (Figure 3.28). Supporting the idea that the space group is truly pseudo-centrosymmetric is the observation that the cumulative intensity distribution of the full data strongly resembles that of a centric data set (Figure 3.29).

![Cumulative intensity distribution](image)

Figure 3.29. Cumulative intensity distribution for the before-damage (IP) data of quasi-racemic snakin-1 in $P\bar{1}$.

The similarity between the structures of the different monomers was assessed by performing a least-squares superposition with *LSQ SUPERPOSE*, as implemented in *COOT*. Root-mean square difference (RMSD) values are given as a measure of structural deviation between the $C_\alpha$ atoms of the two aligned structures. The RMSD values between the pseudo-inversion related D-/L- monomers in the unit cell are 0.16 and 0.22 Å, lower than the value obtained between the NCS-related L- monomers at 0.41 Å. These low values indicate that the pseudo-inversion symmetry is close to perfect, and that there are no overall topological differences between the structures of iodo-L-snakin-1 and D-snakin-1.

### 3.5.2 Crystal structure of true racemic snakin-1

With a model of iodo-L-snakin-1 in hand, the previously obtained data sets were solved by molecular replacement. The structures of racemic snakin-1 in the centrosymmetric space...
groups $P\bar{1}$, $P2_1/c$ (PDB code: 5e5q) and $Pbcn$ could be solved and confirmed that these crystals are racemic, containing equal amounts of L- and D- protein (Figure 3.30, Table 3.14). For the $P\bar{1}$ structure, the data from the before-damage set from the UV-RIP experiment was used, as the previously collected $P\bar{1}$ data (§2.5.2) could not be solved. Interestingly, molecular replacement could not be used to solve this data in $P1$. Excellent MOLEPROBITY clash scores (100th percentile for resolution) and Ramachandran statistics were obtained for all the refined models.

The backbone RMSD values between iodo-L-snakin-1 and most of the native L-snakin-1 molecules are in the range of 0.3–0.8 Å, confirming that incorporation of the iodine did not perturb the protein structure significantly. The backbone RMSD values of L-snakin-1 molecules from the different crystals show greater deviation and are in the range of 0.2-2.0 Å (Table 3.13). The greatest deviations occur between the $P2_1/c$ molecules and those in the other crystals, a possible reason being that a slightly distorted conformation may be required to fulfil the symmetry of the $P2_1/c$ space group. Interestingly, the remaining molecules of the $Pbcn$ and $P\bar{1}$ crystals do not appear to show significant deviations, a surprising observation when considering that $Pbcn$ contains an additional crystallographic symmetry operator compared to $P2_1/c$. 

Figure 3.30. Structure of true racemic snakin-1 from a $P2_1/c$ crystal. L-snakin-1 is in green and D-snakin-1 is in orange. Secondary structure elements are labelled.
Table 3.13. Backbone RMSD values (Å) between L-snakin-1 chains in true racemic crystals compared with an iodo-L-snakin-1 molecule.

<table>
<thead>
<tr>
<th></th>
<th>$P_2_1/c$</th>
<th>$Pbcn$</th>
<th>$P\bar{1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chain A</td>
<td>Chain B</td>
<td>Chain A</td>
</tr>
<tr>
<td>$P_2_1/c$</td>
<td>Chain A</td>
<td>0.66</td>
<td>2.03</td>
</tr>
<tr>
<td>$Pbcn$</td>
<td>Chain A</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td>$P\bar{1}$</td>
<td>Chain A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.14. Data collection statistics for racemic snakin-1 data.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>( P \bar{1} )</th>
<th>( P2_1/C )</th>
<th>( Pbcn )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>( P \bar{1} )</td>
<td>( P2_1/C )</td>
<td>( Pbcn )</td>
</tr>
<tr>
<td>Crystallization condition</td>
<td>22% w/v PEG 3350, 0.1 M Na(_2)HPO(_4)</td>
<td>20% w/v PEG 3350, 0.2 M Na(_2)HPO(_4)</td>
<td>3.2 M ammonium sulfate pH 7.0</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>MX1</td>
<td>MX2</td>
<td>MX1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9537</td>
<td>1.4586</td>
<td>0.9537</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Quantum 210r</td>
<td>ADSC Quantum 315r</td>
<td>ADSC Quantum 210r</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>200</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Unit cell ( a, b, c ) (Å)</td>
<td>31.18, 43.28, 51.02</td>
<td>65.15, 27.74, 73.35</td>
<td>113.50, 64.59, 68.62</td>
</tr>
<tr>
<td>( a, \beta, \gamma ) (°)</td>
<td>93.65, 94.29, 104.16</td>
<td>90, 100.19, 90</td>
<td>90</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.32</td>
<td>0.28</td>
<td>0.18</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>33.69-2.20 (2.27-2.20)</td>
<td>34.14-1.60 (1.63-1.60)</td>
<td>19.57-2.00 (2.05-2.00)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>12833 (1095)</td>
<td>33736 (1631)</td>
<td>34590 (2509)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.4 (97.5)</td>
<td>97.7 (95.1)</td>
<td>99.4 (99.1)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.2 (2.2)</td>
<td>5.8 (6.0)</td>
<td>8.1 (8.3)</td>
</tr>
<tr>
<td>( R_{\text{merge}} ) on intensity</td>
<td>0.135 (0.444)</td>
<td>0.086 (1.13)</td>
<td>0.316 (1.37)</td>
</tr>
<tr>
<td>( R_{\text{pim}} ) on intensity</td>
<td>0.123 (0.405)</td>
<td>0.039 (0.503)</td>
<td>0.116 (0.499)</td>
</tr>
<tr>
<td>( &lt;I/\sigma(I)&gt; )</td>
<td>4.7 (1.9)</td>
<td>10.5 (1.5)</td>
<td>6.3 (1.8)</td>
</tr>
<tr>
<td>( CC_{1/2} )</td>
<td>0.985 (0.819)</td>
<td>0.997 (0.729)</td>
<td>0.994 (0.765)</td>
</tr>
<tr>
<td>Overall ( B ) factor from Wilson plot (Å(^2))</td>
<td>12.5</td>
<td>21.0</td>
<td>17.7</td>
</tr>
</tbody>
</table>

[values of the outermost resolution shell are in parentheses]

\( R_{\text{merge}} = \frac{\sum hkl \sum_i |I_i(hkl) - \langle I(hkl)\rangle|}{\sum hkl \sum_i I_i(hkl)} \)

\( R_{\text{pim}} = \frac{\sum hkl [1/(N - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|}{\sum hkl \sum_i I_i(hkl)} \)

\( CC_{1/2} \) Pearson correlation coefficient between random half sets of the data
<table>
<thead>
<tr>
<th>Space group</th>
<th>P1</th>
<th>P21/C</th>
<th>Pbcn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystallization condition</td>
<td>22% w/v PEG 3350, 0.1 M Na₂HPO₄</td>
<td>20% w/v PEG 3350, 0.2 M Na₂HPO₄</td>
<td>3.2 M ammonium sulfate pH 7.0</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>33.69-2.20</th>
<th>34.14-1.34</th>
<th>19.32-1.34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness (%)</td>
<td>98.4</td>
<td>97.41</td>
<td>99.4</td>
</tr>
<tr>
<td>No. of reflections, working set</td>
<td>12189</td>
<td>30883</td>
<td>31161</td>
</tr>
<tr>
<td>No. of reflections, test set</td>
<td>635</td>
<td>1618</td>
<td>1551</td>
</tr>
<tr>
<td>Final $R_{work}$</td>
<td>0.325</td>
<td>0.287</td>
<td>0.293</td>
</tr>
<tr>
<td>Final $R_{free}$</td>
<td>0.369</td>
<td>0.307</td>
<td>0.326</td>
</tr>
<tr>
<td>No. of non-H atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>932</td>
<td>987</td>
<td>1426</td>
</tr>
<tr>
<td>Ion</td>
<td>-</td>
<td>-</td>
<td>25</td>
</tr>
<tr>
<td>Water</td>
<td>65</td>
<td>154</td>
<td>231</td>
</tr>
<tr>
<td>Total</td>
<td>997</td>
<td>1141</td>
<td>1682</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.007</td>
<td>0.006</td>
<td>0.007</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.95</td>
<td>0.90</td>
<td>0.98</td>
</tr>
<tr>
<td>Average $B$ factors (Å²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>18.8</td>
<td>24.3</td>
<td>18.0</td>
</tr>
<tr>
<td>Ion</td>
<td>-</td>
<td>-</td>
<td>42.6</td>
</tr>
<tr>
<td>Water</td>
<td>21.4</td>
<td>34.2</td>
<td>27.2</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured (%)</td>
<td>98.4</td>
<td>98.5</td>
<td>99.5</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>1.6</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Molprobity score</td>
<td>0 ($100^{th}$ percentile for 2.2 Å)</td>
<td>0.52 ($100^{th}$ percentile for 1.6 Å)</td>
<td>0.71 ($100^{th}$ percentile at 2.0 Å)</td>
</tr>
<tr>
<td>PDB code</td>
<td>-</td>
<td>5e5q</td>
<td>-</td>
</tr>
</tbody>
</table>

[Values of the outermost resolution shell are in parentheses]
The structure of snakin-1 contains an N-terminal helix-turn-helix (HTH) domain (α1, α2), in agreement with secondary structure predictions (§3.3.1), with the helices held together by three disulfide bridges (Figure 3.31). The C-terminal half of the protein contains two large loops separated in the backbone by a short 3_10-helix (η1) and a five-residue α-helix (α3). These two helices are tethered to the C-terminal end of the helix-turn-helix domain by a disulfide bond linking Cys, which follows α2, with Cys, situated between η1 and α3 Figure 3.32. The loops appear to be rigidly held by this and two additional disulfide bonds, preserving the direction in which these loops are projected.

The two loops project into opposing regions of space, with one forming a highly positively charged electrostatic surface on one face of the HTH domain (Figure 3.31).

Figure 3.31. Structural features of snakin-1. A: cartoon representation of snakin-1 showing the two large loops in silver and dark grey; B: surface electrostatics of snakin-1 from the same view, showing a highly positively charged cleft formed between one side of the grey loop and one face of the helix-turn-helix domain.
Figure 3.32. Annotated sequence of snakin-1 and topology diagram. *Top:* annotated sequence of snakin-1 showing α-helices in blue and the 3₁₀-helix in red. The green numbers below the sequence indicate paired cysteines in disulfide bonds 1-6. *Bottom:* Topology diagram showing the connectivity of the structural elements in snakin-1. Black numbers indicate the residues included in secondary structure elements, and white numbers indicate the cysteine residues for each disulfide bond.
The features present in the model of snakin-1 show similarity to the *ab initio*-predicted model by Porto and Franco,\textsuperscript{230} despite the differences in disulfide connectivity (Figure 2.10). However, the overall topologies of the models differ as reflected by an RMSD value of 4.4 Å, calculated from the least-squares superposition of the model structure with that of L-snakin-1 in $P2_1/c$. The high RMSD value explains the previous failed attempts at structure solution using molecular replacement (see the previous chapter, §2.5).\textsuperscript{294} Least-squares superposition of the helical portion of the predicted model (V Gly-32Glu) similarly found an RMSD value of 2.6 Å, again explaining the failure to obtain a correct solution using this as the search model.

Within all of these crystal structures, there is a common arrangement containing two L-snakin-1 molecules, and this may be the minimum crystallisation unit (Figure 3.33). This unit forms an interface with a buried area of 260-490 Å\textsuperscript{2}, including a single hydrogen bond. No covalent or ionic bonds are present at this interface, and this “dimer” may be a crystallisation artefact although it is seen in three different space groups. Further biophysical characterisation will be required to conclude whether it holds any physiological relevance. Interestingly, Nahirñak *et al.* observed the self-association of snakin-1 in the leaves of agroinfiltrated *Nicotiana benthamiana*, though whether this is relevant to the protein in its native environment, or whether this interaction in *N. benthamiana* is of the same nature as observed here, is not known.\textsuperscript{213}

![Visualisation of snakin-1 “dimers”. Alignment of L-snakin-1 “dimers” from quasi-racemic snakin-1 in $P1$ (green), true racemic snakin-1 in $P1$ (cyan), $P2_1/c$ (magenta), and $Pbcn$ (yellow).](image-url)
3.5.3 Structural alignment

Searches for structural homologues of snakin-1 in the entire PDB archive were carried out using the EMBL PDBeFold server. In PDBeFold, the most prominent results in the top 100 hits ordered by their Q-score, a measure of the quality of alignment based on the RMSD value and length of the alignment (see general experimental methods §4.4.8), correspond to the ubiquitin-binding and structurally related UBA and CUE domains.

UBA and CUE domains are predominantly helical, and visual comparison of these domains with snakin-1 shows that they are not particularly similar, suggesting that the alignment identified is due to the high helical content of these domains and not true homology. The residues required for binding ubiquitin in the UBA domain of p47, namely \(34\text{Leu}\) and \(38\text{Leu}\), are replaced by Cys and Arg respectively in snakin-1, and one other residue, \(42\text{Tyr}\), is not present at all (Figure 3.34b). Similarly, residues involved in hydrophobic interactions (\(467\text{Met}\), \(468\text{Phe}\), \(482\text{Leu}\), and \(494\text{Leu}\)) and ionic interactions (\(466\text{Glu}\), \(491\text{Glu}\), and \(495\text{Asp}\)) between the CUE domain of E3 kinase and ubiquitin, are matched with dissimilar residues in the aligned snakin-1, with exception of \(52\text{Asp}\) (Figure 3.34c). Overall, it appears that snakin-1 is not homologous to these aligned structures.
Figure 3.34. Comparison of snakin-1 with ubiquitin-binding domains. Structures shown are: \( A \): snakin-1 (green); the ubiquitin binding \( B \): UBA domain from p47 (purple; PDB code: 1v92); and \( C \): CUE domain of gp78 E3 kinase (blue; PDB code: 2lvq). Residues involved in binding to ubiquitin are shown in black (hydrophobic residues) and orange (charged residues).

One hit that shows some visual similarity is a conserved protein from *Pseudomonas syringae pv.* Tomato str. DC3000 of unknown function (Figure 3.35, PDB code: 3erm). The visual similarities between the models may indicate similar roles or distant homology between snakin-1 and this protein.
Chapter 3

3.5.4 Structural similarity to crAMPs

The structure of snakin-1 shows similarity to other known classes of crAMPs, and in particular the α-helical hairpins and the thionins, which also contain a HTH domain held by disulfide bonds (see §2.2.1; Figure 3.36). Intriguingly, two of the disulfide bonds joining the HTH domain in snakin-1 are arranged identically to the ones in the α-helical hairpin protein EcAMP1 (Figures 3.37, 3.38).181

Figure 3.36. Structures of the cysteine-rich antimicrobial proteins: L-snakin-1 (green), EcAMP1 (blue, α-helical hairpin; PDB code 2l2r), TPP3 (silver, plant defensin; PDB code 4ujo), and viscotoxin A3 (magenta, thionin; PDB code 1okh).

Figure 3.35. Structural alignment of snakin-1 with the structure in PDB entry 3erm. Snakin-1 (green) shows some topological similarities to this conserved protein (silver), and the chain connectivity appears to be the same.
It was previously shown that a truncated construct of the thionin PpTH consisting of only the HTH domain was the minimum functional unit. This thionin also contains the same
arrangement of two disulfides in the HTH domain as in EcAMP1 (Figure 3.38). It is possible that this disulfide motif represents an evolutionarily conserved motif or structural convergence, but whether activity is associated with the presence of this domain alone is not known. A range of functions have been found in α-helical hairpin proteins (§2.2.1.3), suggesting that this domain is not in itself a determinant of antimicrobial activity, but may help define a more general feature of this structure which enables biological activity.

Although these sequences clearly have the same arrangement for two of the disulfide bonds, sequence alignment of snakin-1 with EcAMP1 in BLASTP failed to find any alignment, indicating that these sequences are dissimilar aside from the disulfide connectivity. However, structural alignment of the two structures in PDBFold returns an alignment with a Q-value of 0.25 and Z score of 4.4, suggesting that the structures are indeed similar. EcAMP1 was not returned in the initial structural alignment search, as the percentage of structural similarity is 50%, and the cut-off was previously set at 75%. Lowering the cut-off to 50% places EcAMP1 87th in rank according to Q-score, still a surprisingly low rank, indicating the dominance of the helices in structural alignment.
3.6 Antimicrobial activity of D-snakin-1

To supplement the structural information provided by the crystal structure of snakin-1, characterisation of the biological activity of synthetic D-snakin-1 was carried out.

These experiments were carried out using synthetic protein prepared for the present study by Professor Antonio Molina at the Centro de Biotecnología y Genómica de Plantas (UPM-INIA), Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid, Spain.

The fungal species *Plectosphaerella cucumerina BBM* and *Fusarium oxysporum* f. sp. *conglutinans* 699 were subjected to inhibition assays with the D- protein. Remarkably, the D- protein showed full activity compared to the L- protein, suggesting that the antimicrobial function of snakin-1 does not involve interaction with any chiral species (Figure 3.39).

![Figure 3.39. Inhibition activity of linear L-snakin-1 (L-SN1), native L-snakin-1 (SN1 folded), and D-snakin-1 on: top: *P. cucumerina BBM*, bottom: *F. oxysporum* f. sp. *conglutinans* 699, determined after incubation for 52 h.](image)
3.7 Discussion

3.7.1 Iodo-RIP

The method of iodo-RIP used here in the solution of the snakin-1 structure was demonstrated in a study by Zwart on the incorporation of anomalous signal in RIP.\textsuperscript{301} In this case, the iodine was incorporated via iodination of tyrosines, a method that may show utility in the general use of iodo-RIP. Other examples of iodo-RIP in the literature have been less deliberate, much like this study. Evans \textit{et al.} saw the loss of iodine density from triiodide ions in triiodide-soaked crystals,\textsuperscript{302} while Ennifar \textit{et al.} observed radiation damage to the iodinated uridines in DNA.\textsuperscript{303} In the latter study, comparison of crystals containing iodinated and brominated uridines suggested that the C-I bond showed greater lability than the C-Br bond with X-ray irradiation. Interestingly, in this case the electron density maps showed a shift of the iodide position with cleavage rather than a loss of iodine density, and to our knowledge this is the only case of such a shift. Dumas and co-workers have suggested that the halide products from C-I and C-Br scission are iodide and bromide ions, respectively, and it is conceivable that given a favourable binding site, these ions would remain ordered and contribute to the electron density.\textsuperscript{304, 305}

Iodo-RIP holds clear advantages over conventional X-ray RIP methods. It is generally accepted that successful application of RIP is handicapped by the lack of an indicator for sufficient radiation damage. However, in more specific cases of RIP, namely iodo-RIP, bromo-RIP and seleno-RIP, it is possible to monitor damage by the loss in intensity of characteristic absorbance peaks.\textsuperscript{304, 306, 307} Another advantage is the phasing power of iodine. The work described in this thesis shows that UV-RIP of the native snakin-1 crystals and RIP phasing experiments where the iodine sites have been omitted from the phasing substructure are unable to solve the snakin-1 structure. In comparison, the two iodine sites, together with sites at the two unmodified Tyr hydroxyl groups in the D- protein, are sufficient for high quality initial phases after density modification. The latter two sites may be, we believe, the manifestation of some positional disorder in the crystal lattice that results in substitution of the L- iodo-protein at some sites (probably randomly distributed) in the crystal lattice. Thus, iodo-RIP, in combination with the general methods of incorporating iodine above, represents a highly pragmatic approach to targeted RIP methods.
3.7.2 Amphiphilic AMPs and membrane disruption

Snakin-1 has an overall positively charged surface potential with fourteen positively charged residues (Arg, Lys), and six negatively charged ones (Asp, Glu). The positively charged residues are clustered between the second large loop and α2 of the HTH domain (Figure 3.31, Figure 3.40). The highly positive electrostatic potential in this region could be involved in targeting the protein to negatively-charged bacterial cell membranes, as many AMPs are thought to do. The structure of snakin-1 also appears to be amphiphilic, with hydrophobic faces present on one side of the HTH domain, and on one face of Loop 1, and may indicate that snakin-1 acts by disrupting membranes (Figure 3.40).

![Figure 3.40. Distribution of hydrophobic residues in snakin-1. Left: cartoon representation of snakin-1 with disulfide bonds in yellow, hydrophobic residues (Ala, Gly, Val, Ile, Leu, Phe) in orange, and other residues in grey; right: predicted electrostatic potential of snakin-1 in the same orientation. A highly positively charged region (blue) is predicted on one face of the HTH domain, with a hydrophobic face (white) on the other.]

A role for snakin-1 in membrane disruption is supported by the full activity of the D- protein reported here. For peptides/proteins that act by direct membrane disruption, the interactions formed with membranes are purely physical in nature. As such, the chirality of the peptides/proteins is irrelevant to their activity, as shown by Merrifield and co-workers in their work on the D- enantiomers of the well-characterised AMPs cecropin A, melittin, magainin 2 amide, and hybrids of cecropin A and melittin (see §2.2). An early study of snakin-1 that involved assessing whether it disrupted anionic micelles composed of the phospholipids phosphatidylcholine and phosphatidyglycerol, found it to be inactive. Taken in combination with the results presented here, as well as the target-dependent antimicrobial activity of snakin-1, it seems likely that snakin-1 acts by disruption of membrane components specific to bacterial and fungal membranes. To determine whether
this is the case, studies of snakin-1’s activity on more detailed models of bacterial and fungal membranes need to be carried out.

3.7.3 Modulation of antimicrobial activity

crAMPs are known to be primarily active against phytopathogens and for snakin-1, all current activity assays reported in the literature have focused on these. It is not currently known if snakin-1 is active against any human pathogens, and to determine whether native snakin-1 has use as a potential human therapeutic, broad activity assays must be undertaken. With a robust chemical synthesis in hand, the preparation of large quantities of snakin-1 for activity assays can be performed. Moreover, with the knowledge of the structural arrangement of snakin-1 provided by the crystal structure, analogues can be prepared in order to identify the functional determinants of snakin-1. This may include analogues where the positively charged residues in the protein, or the residues in the two large loops, have been replaced. Knowledge of the functional determinants in snakin-1 will enable modulation of its activity, and could potentially result in a modified a crAMP that is highly active and “broad spectrum” against human pathogens.

In addition, such studies may be directed towards the discovery of the minimum functional unit of snakin-1, allowing a more efficient pathway to an active AMP, an important consideration for manufacturing potential. Starting points for studies towards a minimum functional unit include an analogue containing only the HTH domain, $^1$Gly-$^{30}$Cys, similar to the truncated PpTH that has been found to be the minimum functional unit in this thionin. One point of interest is the additional disulfide ($^5$Cys-$^{30}$Cys) in the HTH domain of snakin-1, not present in the structurally similar EcAMP1 and truncated PpTH, and it will be interesting to see whether the HTH domain of snakin-1 is active, with and without this additional disulfide.

Lastly, the large number of disulfide bonds in GASA/snakin proteins is unusual, and studies on members of this family have found that these disulfides play a part in additional redox functions of these proteins. To determine any functional role of these disulfides, analogues wherein disulfides, singly and multiply in combination, are replaced by stable covalent linkages should be prepared and their activities compared to native snakin-1. This is made possible by the structural knowledge of the disulfide arrangement provided by the crystal structure. Replacement of these disulfides with bridged diamino diacids, thioethers,
alkynes via olefin metathesis,\textsuperscript{308} or with copper(I)-catalysed alkyne-azide cycloaddition click chemistry,\textsuperscript{309, 310} may be suitable methods for this approach (Figure 3.41).

Figure 3.41. Isosteres for disulfide bonds. A: disulfide bond; B: thioether bridge; C: hydrocarbon-based bridge; D: alkene; E: triazole.
3.8 Conclusion

In this work, the potential of racemic protein crystallography in de novo protein structure determination has been explored. Though a comprehensive comparison of crystallisation success with a racemic stock and protein stock containing only L- protein was not undertaken as in examples in the literature, it does appear that crystallisation with the racemic mixture occurs more readily. Importantly, this work has shown that difficulties can still arise in structure solution using racemic protein crystals.

The strategies used towards experimental phasing for the cysteine-rich snakin-1, avoided the synthesis of analogues incorporating selenium, due to potential complications in the formation of mixed seleno-sulfide linked species. Instead, quasi-racemic snakin-1 was crystallised from a mixture of D- and L-protein, with the L-protein incorporating an analogue of tyrosine, 4-iodophenylalanine. Initial quasi-racemic crystals showed poor crystal order along one direction of the crystal, and this may have prevented de novo structure determination. Additionally, crystals were radiation sensitive, and were not suitable for the collection of high redundancy single-wavelength anomalous dispersion experiments. Quick-soaks of better quasi-racemic crystals in bromide-containing solutions appeared to result in uptake of bromide into quasi-racemic crystals, but these showed no anomalous dispersion at X-ray wavelengths close to the bromide $K$ edge, suggesting a lack of ordered bromide ions.

Surprisingly, despite the high level of X-ray beam attenuation used in the bromide MAD data collection, a significant radiation damage-induced difference between data sets was observed. This difference appears to be from the loss of iodine from 4-iodophenylalanine, and provided excellent initial phases owing to the high phasing power of iodine. The high lability of iodine to X-ray irradiation has previously been exploited for RIP, and may be more significant than that of bromine or selenium. This work reinforces this potential of iodine for de novo structure determination via iodo-RIP.

The structural and functional insights that have been provided by the crystal structure of snakin-1 and the antimicrobial assays of the D-protein will be valuable in studies towards the design of new antimicrobial peptides with snakin-1 as a template. This structural information provides starting points for modifying the activity of snakin-1 by the preparation of rationally designed analogues. Finally, the insights gained from using
racemic protein crystallography in this work will hopefully aid future studies involving this technique.
3.9 Experimental

3.9.1 Iodo-L-snakin \(^{1}\)Gly-\(^{254}\)iodo-Phe-\(^{29}\)Cys-COSR thioester 3.1

Boc-Gly-O-CH\(_2\)-phi-CH\(_2\)CO\(_2\)H (0.4 mmol) was coupled to aminomethyl polystyrene resin (0.2 g, 0.2 mmol, 1 mmol/g loading) with DIC (0.4 mmol) in CH\(_2\)Cl\(_2\)/DMF (4/1, v/v, 2 ml) overnight, drained and washed with CH\(_2\)Cl\(_2\). The ninhydrin test was negative. The Boc group was removed with neat TFA for 2 min, washed with DMF and S-tritylmercaptopropionic acid (5 eq.), HATU (4.75 eq.) and \(iPr\_2NEt\) (12 eq.) in DMF was added to resin and the mixture left to stand for 1 h. The resin was washed with DMF, the trityl group was removed with a solution of TFA:H\(_2\)O:TIPS (95/2.5/2.5, v/v/v) for 3 x 2 min. Solid phase peptide synthesis was then performed manually according to the general \textit{in situ} neutralisation Boc-SPPS methods with Boc-L-amino acids (see general experimental methods §4.2.4). Neat TFA was used as the Boc deblocking reagent and HATU/\(iPr\_2NEt\) as the coupling reagents. Coupling of the \(^{25}\)4-iodo-Phe residue was carried out in the same way using Boc-4-iodophenylalanine-OH with no changes to the coupling reagents or conditions. The dried peptidyl resin (1.10 g) was cleaved with anhydrous HF/p-cresol/p-thiocresol (18/1/1, v/v/v, 10 ml) for 1 h at 0 °C. Following evaporation of HF, the peptide was precipitated with cold diethyl ether, isolated by centrifugation, washed with cold diethyl ether, dissolved in 1/1 v/v acetonitrile/water containing 0.1% TFA, filtered and lyophilized to afford the crude peptide (460 mg). Purification by slow gradient RP-HPLC afforded the title compound (73.0 mg, 10.7% yield; [M+4H]\(^{4+}\) m/z calculated 850.7, observed 850.5).

3.9.2 Reduced full length iodo-L-snakin \(^{1}\)Gly-\(^{254}\)iodo-Phe-\(^{63}\)Pro 3.2

Native chemical ligation of \(^{1}\)Gly-\(^{254}\)iodo-Phe-\(^{29}\)Cys-thioester 3.1 (30.1 mg, 8.86 \(\mu\)mol) and \(^{30}\)Cys-\(^{63}\)Pro 2.2 (34.4 mg, 9.03 \(\mu\)mol) was performed in a buffer solution consisting of 6 M GnHCl, 0.2 M Na\(_2\)HPO\(_4\), 100 mM MPAA and 20 mM TCEP (8.93 ml, pH = 6.9) under an argon atmosphere. The ligation reaction was complete in 3 h at room temperature as judged by HPLC. The reaction mixture was acidified to pH 2 with 5 M HCl, recovered by RP-HPLC and lyophilized. LC-MS analysis of the crude product showed the desired linear \(^{1}\)Gly-\(^{254}\)Iodo-Phe-\(^{63}\)Pro product: [M+9H]\(^{9+}\) m/z calculated 783.7, observed 782.3 (58.3 mg crude, 92.7% crude yield).
3.9.3  **Iodo-L-snakin 1**  \(^1\)Gly-\(^{254}\)-ido-Phe-\(^{63}\)Pro-OH 3.3
Crude reduced \(^1\)Gly-\(^{254}\)-ido-Phe-\(^{63}\)Pro 3.3 (58.3 mg, 8.28 µmol) was dissolved in a buffer solution of 1 mM cysteine, 0.1 mM cystine, 150 mM NaCl and 0.1 M Tris (600 ml, pH = 8.0), and this reaction mixture stood at room temperature. The reaction was monitored by RP-HPLC (50 °C) and LC-MS and was completed within 21 h. Purification of the reaction mixture by RP-HPLC at 50 °C using a gradient of 1-61% B over 120 min (buffer A = 0.1% aq. TFA; B = 0.1% TFA in MeCN) afforded the synthetic iodo-L-snakin-1 (16.5 mg, 26.3% yield over two steps; \([M+9H]^9+ \ m/z\) calculated 782.3, observed 782.3; deconvoluted mass 7031.3 ± 0.6 Da, calculated mass 7031.9 Da).

3.9.4  **Circular dichroism spectrometry**
CD spectra were obtained on a PiStar spectrometer (Applied Photophysics; Leatherhead, UK). Samples were placed in a 1 mm quartz cuvette for measurement. Spectra were collected at 6 °C with a 2 nm optical bandwidth over the range of 180-300 nm in 0.5 nm increments. Measurements at each wavelength were performed over 2.0 s. The measurement of each spectra was repeated seven times and the final spectrum taken as the mean of the measurements. The averaged baseline spectrum of the buffer was subtracted from the average measured with protein in buffer to obtain the solvent subtracted spectrum. Samples were prepared at 9-10 nM concentration.

3.9.5  **Initial crystallisation screen of iodo-snakin-1 quasi-racemate**
A protein stock solution of iodo-L-snakin-1 3.3 was prepared in MQ H<sub>2</sub>O and centrifuged through a Nanosep MF 0.2 µm filter (Pall Life Sciences). The UV absorbance of 4-iodophenylalanine at 280 nm has previously been estimated at 260 M<sup>-1</sup> cm<sup>-1</sup> by De Filippis et al. Thus, the extinction coefficient of iodo-L-snakin-1 was estimated to be 0.5674 (mg/ml)<sup>-1</sup> cm<sup>-1</sup>. Thus, the concentration of the iodo-L-snakin-1 stock solution was estimated to be 12.0 mg/ml.

A quasi-racemic stock solution was then prepared from individual stock solutions of iodo-L-snakin-1 3.3 and D-snakin-1 2.8 in MQ H<sub>2</sub>O. The iodo-L-snakin-1 (12.0 mg/ml) stock solution was mixed with a stock solution of D-snakin-1 (12.5 mg/ml) in a 1:1 ratio, and used without further determination of protein concentration.
Buffer conditions for initial crystallisation screens were transferred from 96-deep well blocks to the reservoir of 96 well Intelli-Plates (Hampton Research, 100 µl per reservoir) using a MultiPROBE II HT/EX liquid-handling robot (Perkin-Elmer). The buffer conditions used are those described by Moreland et al. and Gorrec, and were distributed over six Intelli-Plates. Sitting drops containing 0.30 µl of quasi-racemic protein stock and 0.30 µl of reservoir solution were prepared using an Oryx4 Protein Crystallization Robot.

Upon set up of drops, Intelli-Plates were immediately sealed with HR4-521 ClearSeal Film (Hampton Research). Plates were inspected daily for the first week, every second day for the second week, every three-to-five days for the third and fourth weeks, and infrequently thereafter.

### 3.9.6 Microseeded initial crystallisation screen of iodo-snakin-1 quasi-racemate

A protein stock solution was prepared from individual stock solutions of iodo-L-snakin-1 and D-snakin-1 in MQ H₂O. Stock solutions of iodo-L-snakin-1 (12.0 mg/ml) and D-snakin-1 (12.5 mg/ml) were centrifuged through a Nanosep MF 0.2 µm filter (Pall Life Sciences) then mixed in a 1:1 ratio, and used without further determination of protein concentration.

A seed stock of a racemic snakin-1 crystal grown from 0.2 M Na₂HPO₄/20% PEG 3350 was prepared by physically grinding the crystals, followed by dilution with crystallisation buffer (200 µl).

Buffer conditions for initial crystallisation screens were transferred from 96-deep well blocks to the reservoir of 96 well Intelli-Plates (Hampton Research, 100 µl per reservoir) using a MultiPROBE II HT/EX liquid-handling robot (Perkin-Elmer). The buffer conditions used are those described by Moreland et al. and Gorrec, and were distributed over six Intelli-Plates. Sitting drops containing 0.20 µl of quasi-racemic protein stock, 0.20 µl of reservoir solution and 0.05 µl of seed stock were prepared using an Oryx4 Protein Crystallization Robot.

Upon set up of drops, Intelli-Plates were immediately sealed with HR4-521 ClearSeal Film (Hampton Research). Plates were inspected daily for the first week, every second day for the second week, every three-to-five days for the third and fourth weeks, and infrequently thereafter.
3.9.7 Determination of antifungal activity

This experiment was carried out by Professor Antonio Molina and Gemma López García at the Centro de Biotecnología y Genómica de Plantas (UPM-INIA), Universidad Politécnica de Madrid.

The *in vitro* activity of synthetic L-snakin-1 2.4 and D-snakin-1 2.8 against the fungal plant pathogens *Plectosphaerella cucumerina* BMM and *Fusarium oxysporium* f. sp. *conglutinans* 699212 was tested as described previously. Different concentrations (0, 0.001, 1.25, 3, 5, 10 and 20 µM) of the peptides were tested and the equivalent concentrations causing 50% of fungal inhibition (EC₅₀) were determined at different time points (0 to 52 h). At least two replicates per concentration were performed in each experiment, which were repeated at least four times.
Chapter 4

General Experimental Methods
4.1 Materials and methods

Commercially available starting materials were purchased from Polypeptides Group, ChemPep, Peptides International, Chem-Impex, Acros Organics, AK Scientific, Alfa Aesar, GL-Biochem, Merck Millipore, Sigma-Aldrich, Bio-Rad, Pure Science, MP Biomedicals, ECP and were used as supplied. MQ H$_2$O used to prepare buffer solutions and reagents was obtained from a Milli-Q purification system and sterilised through filtration (0.22 μm).

Unless otherwise specified, the following $N$-Boc amino acids and the equivalent $N$-Boc-D amino acids were used in synthesis:


Analytical RP-HPLC was performed with a Dionex (Sunnyvale, CA) UltiMate 3000 system equipped with a four-channel UV detector using a Gracesmart C18 (5 μm; 2.0 x 50 mm) column (Grace Discovery Science; Columbia, MD) and a linear gradient of 1% to 61% B over 60 min (1% B per minute) at a flow rate of 0.2 ml/min unless otherwise indicated. The solvent system used was A (0.1% trifluoroacetic acid in H$_2$O) and B (0.1% trifluoroacetic acid in acetonitrile). Peptide and protein masses were confirmed by LC-MS using an Agilent (Santa Clara, CA) 1260 infinity system equipped with an Agilent 6120 quadruple mass spectrometer using ESI in the positive mode. LC-MS was performed using a Zorbax 300-SB C3 (5 μm; 3.0 x 150 mm) column (Agilent) and a linear gradient of 1% to 61% B over 21 min (3% B per minute) at a flow rate of 0.3 ml/min. The solvent system used was A (0.1% formic acid in H$_2$O) and B (0.1% formic acid in acetonitrile).

Peptides and proteins were purified using a Dionex UltiMate 3000 system equipped with a Foxy Jr fraction collector (Teledyne Isco) using a Gemini C18 (5 μm; 10.0 x 250 mm) column (Phenomenex) at a flow rate of 5 ml/min and eluted using a one-step slow gradient protocol unless otherwise indicated. The solvent system used was A (0.1% trifluoroacetic acid in H$_2$O) and B (0.1% trifluoroacetic acid in acetonitrile). Fractions were collected, analyzed by RP-HPLC and LC-MS, pooled and lyophilized.
Infrared (IR) spectra were obtained on a Perkin Elmer Spectrum 100 Fourier Transform Infrared (FT-IR) Spectrometer using a universal ATR sampling accessory.

Sequence alignment was carried out using BLASTP\textsuperscript{299, 300} and COBALT.\textsuperscript{312} Structure predictions were carried out using the JPred 4 secondary structure prediction server,\textsuperscript{266, 267} and i-Tasser structure prediction server.\textsuperscript{268-270} Figures were generated using the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. Surface electrostatics were predicted using the APBS plugin for PyMOL\textsuperscript{313, 314} from PQD files generated using PDB files containing atomic coordinates.\textsuperscript{315}

4.2 General Methods for Peptide Synthesis

4.2.1 Ninhydrin test
A small portion of resin was washed with CH$_2$Cl$_2$ and allowed to dry. A drop each of 5% w/v ninhydrin in EtOH, 80% w/v phenol in EtOH and 0.2 mM aq. KCN in pyridine solutions were added to the dried resin and the mixture heated at 100°C for 5 min. Blue coloured beads and a deep blue solution indicated the presence of free primary amine groups, and a yellow solution indicated no amino groups were present.

4.2.2 $N$-phthalamidomethyl polystyrene resin (1 mmol/g loading)
Bio-beads® S-X1 1% DVB-PS resin beads (5.00 g; Bio-Rad; Hercules, CA) were swelled in CH$_2$Cl$_2$ (90 ml). $N$-(hydroxymethyl)phthalimide (1.12 eq., 0.985 g, 5.6 mmol) was added and the suspension stirred for 10 min, followed by addition of methanesulfonic acid (10 ml). The resultant light-brown suspension was stirred for six hours at room temperature. The resin was filtered, washed with CH$_2$Cl$_2$ (500 ml), EtOH (500 ml), and dried in vacuo. An FT-IR spectrum of the dry resin showed an intense peak at 1713 cm$^{-1}$ confirming successful incorporation of the phthalimide.

4.2.3 Aminomethyl polystyrene resin (1 mmol/g loading)
A solution of 20% v/v ethanolamine in EtOH (100 ml) was added to $N$-phthalamidomethyl resin (prepared from 5.00 g 1% DVB-PS resin beads), and the resulting suspension refluxed for eight hours. The mixture was then allowed to cool, after which the resin was filtered, washed with EtOH (500 ml), DMF (500 ml), CH$_2$Cl$_2$ (500 ml), EtOH and dried in vacuo. FT-IR spectrum of the dry resin showed the disappearance of the absorbance peak at 1713 cm$^{-1}$ confirming successful aminolysis.
Estimation of the resin loading was carried out by first coupling a weighed sample of dry resin (11.0 mg) with a large excess of Fmoc-Gly-OH (59 mg, 0.2 mmol) and DIC (31 μl, 0.2 mmol) in CH₂Cl₂:DMF (10:1, 1.1 ml) for 2 hr. The resin was washed with CH₂Cl₂, DMF and MeOH and reweighed (15.0 mg).

Two samples of Fmoc-Gly-AM-resin (1.30 and 1.31 mg, were agitated in 20% v/v piperidine in DMF (3.0 ml for each) and allowed to settle for 10 min. The UV absorbance of both solutions were measured using a Jenway 7135 spectrophotometer (Bibby-Scientific; Staffordshire, UK) at a wavelength of 290 nm and the mean absorbance (1.71 Abs) confirmed a resin loading of 1.02 mmol/g (for the dibenzofulvene-piperidine adduct at 290 nm, ε = 5253 M⁻¹ cm⁻¹)

4.2.4 In situ neutralisation Boc-SPPS

In situ neutralisation Boc-SPPS was carried out by repetitive cycles of deprotection and coupling steps as follows.

Deprotection:

The drained resin was subjected to a 10 second neat TFA flow wash followed by a 2 min batch wash. The resin was drained and washed with DMF (30 s flow wash).

Coupling:

The drained resin was coupled with a mixture of Boc-AA-OH (5 eq.), iPr₂NEt (12 eq.) in 0.475 M HATU/DMF (2 ml) for 5 min, with a pre-activation of 1 min. The resin was washed with DMF (30 s flow wash) and drained.

This cycle was repeated for each residue in target sequences. After the coupling of the final residue in a target sequence, resin was washed with DMF, and a final deprotection carried out. The peptidoresin was then washed with DMF, EtOH and dried in vacuo.

4.2.5 HF cleavage

p-Cresol (0.5 ml) was added to dry peptidoresin (up to ~1.0 g) in a specialised Teflon reaction vessel attached to a Teflon cleavage apparatus (Peptide Institute, Inc., Osaka, Japan). The reaction vessel was cooled to -78°C (using a dry ice/ethanol bath) and HF (10 ml) was condensed into the vessel under vacuum. The reaction was stirred for 1 h at 0°C, after which the HF was removed in vacuo and quenched through a CaO trap. Ice-cold Et₂O
was added to the residue and the resulting suspension was recovered by centrifugation. The supernatant was removed, ether was added, and the peptide recovered by centrifugation. The crude peptide was dissolved in 50% aq. MeCN containing 0.1% v/v TFA and lyophilised.

### 4.3 Protein Crystallisation

#### 4.3.1 Robot crystallisation screening

Initial protein crystallisation screens were set up as sitting drop vapour diffusion experiments. Screening was carried out against six sets of 96-conditions (total of 576 conditions) prepared in-house using and expanding on precipitant solutions from a number of screens (Hampton Sparse Matrix I and II, Hampton MPD, Precipitant Synergy, Clear Strategy, Top67, PEG/pH, PEG/ion, Stura Footprint, Morpheus).\(^{248, 249}\) In-house screens were dispensed into 96-well Intelli-Plates (Hampton Research; 100 µl/reservoir) from 96-well deep well blocks containing the screens via a MultiPROBE II HT/EX liquid-handling robot (Perkin-Elmer). Intelli-Plates were then transferred by hand to an Oryx4 Protein Crystallisation Robot (Douglas Instruments) which carried out the dispensing and mixing of protein with the reservoir solutions (1:1 ratio) in the sample wells. Plates were sealed and allowed to incubate. Preparation of plates and incubation was carried out at 18 °C with a relative humidity of 85%.

#### 4.3.2 Crystallisation optimisation

Once a promising condition was found from initial screening, optimisation was carried out. Promising conditions were first checked for reproducibility by repeating the sitting drop vapour diffusion experiments manually in CrystalClear P™ strips. Both crystallisation buffer taken from the 96-deep-well plates containing the initial screens, and crystallisation buffer prepared by hand were used.

Crystallisation conditions were also checked by sitting drop and hanging drop vapour diffusion in 24-well VDX plates (Hampton Research). Drops were initially set up as 1:1 protein:crystallisation buffer (2 µl), to equilibrate against 500 µl of reservoir solution, and wells were sealed with paraffin grease (Shell).

Once a reproducible condition was identified, the crystallisation condition was optimised by systematic variation of the buffer composition and adjustment of the ratios of protein to crystallisation buffer in drops, and of the volume of drops to reservoir.
Preparation of plates and incubation was carried out at 18 °C with a relative humidity of 85%.

4.4 X-Ray Crystallography

4.4.1 Crystal preparation
In order to limit radiation damage during data collection, crystals were flash-cooled in liquid N\textsubscript{2} prior to X-ray diffraction experiments. Flash-cooling can cause the formation of ice crystals within the crystals, disrupting crystal order, and so some form of cryo-protection must be used.

For this work, cryo-protection was used in the form of agents that penetrate solvent channels (e.g. ethylene glycol), and agents that form a barrier at the air-crystal interface (e.g. oil). To ensure satisfactory cryo-protection, agents that penetrate solvent channels were tested by inclusion in crystallisation buffers in 20-30% v/v proportions. Suppression of ice crystals was judged by test diffraction of flash-cooled cryo-protectant-containing solutions on an in-house Micromax 007 HF rotating anode source (Rigaku) with a mar345 dtb robotic mounting system and image plate detector (Mar research). Cryo-protectants that form a barrier were not tested prior to use.

For crystal flash-cooling, crystals were harvested using a nylon cryoloop (Hampton Research) attached to a cryocap (Hampton Research), and either coated with a 70:30 mixture of Paratone-N:mineral oil, or briefly immersed in a cryo-protectant-containing solution. Crystals were then flash-cooled in liquid N\textsubscript{2} and maintained at -173 °C thereafter.

4.4.2 Crystal diffraction data collection
X-ray diffraction experiments were conducted at the Australian Synchrotron on either the MX1 macromolecular crystallography beamline or MX2 micro crystallography beamline. MX1 is equipped with an ADSC Quantum 210r Detector, and MX2 is equipped with an ADSC Quantum 315r Detector. Both beamlines are controlled using the Blu-Ice interface and are equipped with SAM loading robots, which maintains crystals in a liquid N\textsubscript{2} environment, and mounts crystals onto the goniometer for exposure to the X-ray beam.\textsuperscript{316}

The wavelength and attenuation of the X-ray beam, and the distance of the crystal to the detector were adjusted as appropriate for each experiment.
4.4.3 Diffraction data processing

X-Ray data collected as oscillation images were indexed, integrated, and scaled with the \textit{XDS} suite. The following programs distributed with the \textit{CCP4} suite were used. \textit{POINTLESS} was used for unit cell and space group prediction. \textit{AIMLESS} provided scaling and merging of equivalent measurements. High resolution cut-offs of the X-ray data were chosen as the point where $\langle I/\sigma(I) \rangle = 1.5$ and $CC_{1/2} = 0.3$. Old \textit{TRUNCATE} or \textit{CTRUNCATE} were used to convert intensities to structure factor amplitudes.

The diffraction frames to be included in each data set were decided based on visual inspection of the variation in $R_{\text{merge}}$, $M_n(k)$, and satisfactory values for each frame over a data collection on a crystal.

\[ R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_i I_i(hkl)} \]

4.4.4 Analysis of unit cell

Estimation of the solvent content in a crystal and the likely number of molecules present in the asymmetric unit was carried out by analysis of unit cell using \textit{MATTHEWS_COEF} of the \textit{CCP4} suite. \textit{MATTHEWS_COEF} determines the likelihood for a number of molecules in an asymmetric unit by evaluating the Matthews coefficient:

\[ V_M = \frac{V_{AU}}{(n \times M_r)} \]

where $V_M$ is the Matthews coefficient, $V_{AU}$ is the volume of the asymmetric unit, $n$ is the number of molecules in the asymmetric unit, and $M_r$ is the mass of one molecule in Daltons.

4.4.5 Structure solution

Phasing by molecular replacement was performed using \textit{PHASER} via the GUI of the \textit{PHENIX} suite. Identification of correct solutions involved evaluation of the TFZ (> 8.0 is a good indicator of a correct solution) and LLG values, and crystallographic R-factors ($R_{\text{work}}/R_{\text{free}}$) after maximum-likelihood refinement of the MR solution. The search models employed in attempts to solve the snakin-1 structure by MR were atomic coordinate (PDB) files for EcAMP1 (PDB code: 2l2r), Viscotoxin A3 (PDB code: 1okh) and the model of snakin-1 predicted by Porto and Franco using \textit{ab initio} prediction methods. Models were used without hydrogen atoms, as the native sequences, and as models that had been truncated to polyAla chains. The search model eventually employed in solving the structure
of racemic snakin-1 crystals was the iodo-L-snakin-1 structure solved in this work (PDB code: 5e5y).

Fragment-based ab initio phasing employed PHASER in PHENIX GUI, in combination with the CCP4 programmes PARROT, BUCCANEER, and REFMAC. The workflow of the fragment-based ab initio method used is as follows. A PHASER search was carried out for ideal helices 8-12 residues in length based on the expected number of helices in the predicted protein secondary structure from the JPred 4 and i-Tasser structure prediction servers. Density modification was carried out on the best MR solution in PARROT for 5 cycles. Iterative cycles were then carried out, consisting of chain extension in BUCCANEER starting from the initial MR model (10 cycles), maximum-likelihood refinement in REFMAC (10 cycles), and density modification in PARROT (5 cycles). The chain extension in BUCCANEER for each successive iterative cycle returned to the initial MR model as the template for extension, but used the density modified electron density maps from PARROT in the previous iterative cycle. The workflow was controlled by a bash script identical to that used in the solution of the structure of Rv1738 (kindly provided by Dr Richard Bunker, Friedrich Miescher Institute for Biomedical Research, University of Basel). Identification of correct solutions was determined by the number of residues built in chains, and by the crystallographic R-factors after refinement in REFMAC.

Ab initio structure solution was also attempted using ARCIMBOLDO LITE which employs PHASER and SHELXE. ARCIMBOLDO LITE carries out PHASER searches for ideal helices 8-12 residues in length, followed by SHELXE density modification and chain tracing for the top solutions after successful placement of each helix. The number of helices searched for was based on the expected number of helices in the predicted protein secondary structure from the JPred 4 and i-Tasser structure prediction servers. Identification of correct solutions was by inspecting the SHELXE correlation coefficient of the electron density maps of partial solutions against the experimental data (> 25% indicates a correct solution).

Experimental phasing and model building was performed using SHELXC/D/E and ARP/WARP via the EMBL AUTORICKSHAW server, and as stand-alone programmes. For the AUTORICKSHAW server, solutions were identified by the number of residues built by ARP/WARP, and for the stand-alone programmes, by the
Chapter 4

SHELXE correlation coefficient of the electron density maps of partial solutions against the experimental data (> 25% indicates a correct solution).263

4.4.6 Model building and refinement
Once a correct solution was identified, model building was carried out in ARP/WARP,326-328 followed by iterative cycles of manual rebuilding and rigid-body refinement in COOT,289 and maximum-likelihood refinement in PHENIX.REFINE.253, 290-292, 329 Manual rebuilding of the protein chain was carried out by exploiting knowledge of the specific sequence of the synthetic proteins, and visual examination of the $2F_o - F_c$ and $F_o - F_c$ electron density maps (> 0.5 $\sigma$). Water molecules were initially added using the PHENIX.REFINE ‘update waters’ function, and subsequently refined by manual addition, and removal of waters that appeared weak or did not show appropriate hydrogen bonding. Ligand molecules were added where positive difference density ($F_o - F_c$) inconsistent with waters was overlaid with density in the $2F_o - F_c$ maps. Ligands were chosen from those in the crystallisation condition based on the fit to the electron density, and were removed if subsequent cycles of PHENIX.REFINE showed a loss of the density.

Iterative cycles of rebuilding and refinement were continued until all electron density that could be explained confidently had been modelled, and no significant reduction (> 0.010) was afforded in both $R_{work}$ and $R_{free}$ by subsequent cycles.

4.4.7 Structure validation tools

4.4.7.1 R-factors
During structure solution and model refinement, the R-factors, $R_{work}$ and $R_{free}$ were used as measures of the model quality:

$$R_{work} = \frac{\sum_{(hkl) \in T} |F_{obs}| - k|F_{calc}|}{\sum_{(hkl) \in T} |F_{obs}|}$$

$$R_{free} = \frac{\sum_{(hkl) \in T} |F_{obs}| - k|F_{calc}|}{\sum_{(hkl) \in T} |F_{obs}|}$$

where $T$ is a validation set incorporating a subset of the total reflections in a data set (usually 5%), $F_{obs}$ is the experimental structure factor for a reflection $F_{hkl}$, and $F_{calc}$ is the calculated structure factor for a reflection $F_{hkl}$ given the refined model. The validation set of reflections
$T$ is excluded from the refinement and modelling process, and is used for unbiased validation of the model.\textsuperscript{330}

4.4.7.2 Structure validation servers

The quality of refined models was assessed using the structure-validation web server \textit{MOLPROBITY}.\textsuperscript{293} For the quasi-racemic crystal structures, the model was assessed with the full set of atomic coordinates, and with the coordinates of the L- molecules only, as \textit{MOLPROBITY} is unable to provide appropriate Ramachandran and rotamer statistics for D- molecules. The final models deposited in the Protein Data Bank (PDB) were validated using the PDB validation server (http://www.wwpdb.org/validation/validation-servers).\textsuperscript{331}

4.4.8 Structural alignment

Searches for similar structures were carried out using the EMBL \textit{PDBeFold} structure alignment server. \textit{PDBeFold} uses Secondary Structure Matching to align structures, and provides multiple measures to assess the quality of alignments including the Q-value (a measure of the RMSD of an alignment weighted by the length of the alignment), RMSD, $N_{sse}$ (number of matched secondary structure elements), and $N_{align}$ (number of residues aligned).\textsuperscript{295}

\[
Q = \frac{N_{align}^2}{[[1 + (\text{RMSD}/R_0)^2]N_1N_2]}
\]

where $N_1$ and $N_2$ are the number of aligned residues in structures 1 and 2 respectively, and $R_0$ is an empirically measure of the relative significance of RMSD and $N_{align}$. 

195
4.4.9 Crystallography programs

Unless otherwise specified, the versions of the crystallography programs used in this work are as follows:

<table>
<thead>
<tr>
<th>Program</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>POINTLESS</td>
<td>1.9.16</td>
</tr>
<tr>
<td>AIMLESS</td>
<td>0.3.11</td>
</tr>
<tr>
<td>Phaser</td>
<td>2.6.0</td>
</tr>
<tr>
<td>PHENIX.REFINE</td>
<td>1.10.2155</td>
</tr>
<tr>
<td>PARROT</td>
<td>1.0.3</td>
</tr>
<tr>
<td>BUCCANEER</td>
<td>1.5.1</td>
</tr>
<tr>
<td>REFMAC</td>
<td>5.8.0107</td>
</tr>
<tr>
<td>ARCIMBOLDO LITE</td>
<td>3.00</td>
</tr>
<tr>
<td>SHELXC</td>
<td>2014/2</td>
</tr>
<tr>
<td>SHELXD</td>
<td>2013/2</td>
</tr>
<tr>
<td>SHELXE</td>
<td>2014/4</td>
</tr>
<tr>
<td>ARP/wARP</td>
<td>7.5 patch 2</td>
</tr>
</tbody>
</table>
Appendix

Preparation of Truncated Orf Virus Entry Fusion Complex Proteins by Chemical Synthesis
Preparation of truncated orf virus entry fusion complex proteins by chemical synthesis

Ho Yeung,* Paul W. R. Harris,* Christopher J. Squire,* Edward N. Baker* and Margaret A. Brimble*a,b,c

Members of the Chordopoxvirinae subfamily possess an unusual 11 protein entry-fusion complex (EFC) that is highly conserved and present in all species. The mode of action of this EFC is unknown, and the interactions of the constituent proteins are uncharacterised. Here, we present the chemical synthesis of membrane domain truncated linear constructs of two EFC proteins in orf virus, ORFV036 and 049. By using Boc solid phase peptide synthesis and native chemical ligation methods, these truncated proteins have been readily prepared in milligram quantities. These robust synthetic protocols allow ready access to these polypeptides to facilitate biological studies. Copyright © 2014 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: chordopoxviridae; entry fusion complex; orf virus; native chemical ligation

Introduction

Contagious pustular dermatitis is a zoonotic skin condition primarily affecting goats and sheep, which can lead to mortality amongst young infected animals. It is highly prevalent and hence is of particular importance to countries where goats and sheep are major sources of export, but the economic cost has been difficult to quantify (1). The causative agent of this disease is orf virus, a poxvirus belonging to the Chordopoxvirinae subfamily of the Parapoxviridae genus.

Like all chordopoxviruses, orf virus has a highly conserved genome (2) and is thought to enter cells by mechanisms similar to other species. In vaccinia virus, the almost exclusively studied chordopoxvirus species for entry into host cells, viral entry occurs through either fusion of the viral envelope with the host membrane or by macropinocytosis and subsequent escape from vesicles. In both cases, fusion of the viral envelope with a host membrane occurs, requiring the action of viral entry fusion proteins.

For most viruses, one or two proteins are sufficient for fusion (3). Chordopoxviruses represent an unusual case where a large number of fusion proteins are essential for viral infectivity. The large multiprotein entry-fusion complex embodied in the orf virus envelope is a highly conserved feature of chordopoxviruses (2). It is made up of at least 11 proteins, but to date, the mode of action of this complex and the interactions of the constituent proteins are uncharacterised (3). Thus, investigation and identification of the specific interactions of these proteins would aid in understanding how this complex functions and in developing methods to neutralise infectivity in chordopoxviruses.

Three pairs of entry–fusion complex proteins have been found to interact non-covalently in vaccinia virus: A27L, A22, A16L and G3L (4). The orf virus homologs of G3 and L5 are ORFV036 and 049, respectively (2). Previous attempts at recombinant expression of these two orf proteins in Escherichia coli with omission of putative transmembrane domains failed, and no protein was obtained. (Christopher Squire, personal communication). Here, we present the preparation of transmembrane domain truncated constructs of these proteins by chemical synthesis by using the techniques of solid phase peptide synthesis (SPPS) and native chemical ligation (NCL). Through the development of these synthetic protocols, ORFV036 and 049 are now accessible for further biological studies on their structure and function.

Methods and Materials

General

All solvents and reagents were used as supplied. (0-7-azabenzotriazole-1-y1)-N,N,N,N-tetramethyluroniumhexafluorophosphate (HATU) and 5-ethyl-3-(mesancopropanoic acid were purchased from GL Biochem (Shanghai), China). Dimethylformamide (DMF, analytical reagent grade) and acetonitrile (HPLC grade) were purchased from Scharlau (Barcelona, Spain). Acetonitrile (gradient grade) was purchased from Merck (Darmstadt, Germany). Dipea, N,N-diisopropylethylamine, piperidine, methoxyamine hydrochloride, tolybutyl/oxymonium iodide (TBAT), dimethyl sulfoxide and 4-methylphenylacetic acid (MPAA) were purchased from Aldrich (St Louis, MO, USA). (3-Chlorotriazinyl)phosphonic acid (CTEP–HCl) was purchased from Alfa Aesar (Lanarkshire, UK). TFA was purchased from Oxford Chemicals (West Columbia, SC, USA). Guanidino hydrochloride (Gn–HCl) was purchased from MP Biomedicals (Solon, OH, USA). Anthranilic diacid hydrogen phosphate (analytical reagent grade) was purchased from BCP (Auckland, New Zealand). Anthranil hydroxy fluorido was purchased from Matheson Tristate (Basking Ridge, NJ, USA). Aminomethyl polyethylene resin was synthesized in "
house as described previously on S-1X Biorad beads purchased from Bio-Rad (Hercules, CA, USA). 35S amino acids were purchased from Polypeptide (Strasbourg, France) with the following side chain protection: Boc-Arg(Tos)-OH (Tos = p-toluenesulfonyl), Boc-Glu(p-bzl)-OH, Boc-Cys(3-MeCys)-OH (Bn = benzyl), Boc-Asp(OtBu)-OH, Boc-Lys(NCA)-OH (Xan = xanthylethyl), Boc-Glu(Alloc)-OH, Boc-His(Tos)-OH, DDOCH(1-arginyl), DDOCH(1-lysyl), DDOCH(1-cysteinyl), and Boc-Gly-CH2CH2CH2CO2H.

Peptide synthesis

For ORFv036(25-110) Cys-Lys-S6 bearing a C-terminal acid,aminomethyl polystyrene resin was functionalized with Boc-Lys (G2-Cl)-OH, Boc-Arg(Tos)-OH, and Boc-Glu(Alloc)-OH (2 eq) and DIC (2 eq) in CH2Cl2 for 1 h. For peptides ORFv049(46-128) Cys-Lys-S6 and Cys-Lys-S6,aminomethyl polystyrene resin was functionalized with Boc-Lys (G2-Cl)-OH, Boc-Arg(Tos)-OH, and Boc-Glu(Alloc)-OH (2 eq) and DIC (2 eq) in CH2Cl2 for 1 h. For peptides ORFv049(46-128) Cys-Lys-S6 and Cys-Lys-S6, amino acid incorporation was performed directly following a Boc deprotection of this resin. For ORFv036(25-110) Thr-Ser-COSR, the resin was eluted with two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate. For ORFv049(46-128) Thr-Ser-COSR, an additional two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate. For all cases, the resin was eluted with two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate. For all cases, the resin was eluted with two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate. For all cases, the resin was eluted with two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate. For all cases, the resin was eluted with two 1-M NaOH standard elution protocols by using Boc-Cys2(2Cl)-OH prior to incorporation of the thiolate.

Native Chemical Ligation

Ligation and formyl removal

ORFv036(25-110) Thr-Ser-COSR (4.5 mg, 7.9 pmol) and ORFv036(25-110) Cys-Lys-S6 (1.7 mg, 8.7 pmol) were dissolved in a solution of 0.2 M phosphate/5 M guanidine hydrochloride containing 4 M AAPA (200 mM) and TEAP- HCl (20 mM) at pH 7.0. The final concentrations were 3.3 mM (Thr-Lys-COSR) and 3.6 mM (Cys-Lys-S6) (final pH 6.9). After 2 h at room temperature, the solution was diluted to 30 ml by addition of 0.2 M phosphate/5 M guanidine hydrochloride and pipetted (20 ml) onto Boc-protected alanine acids (1.5 mM, 7.4 mg/2 ml) 2 min to effect a one-step reduction of methionine sulfide and Boc deprotection. The resin was then washed with CH2Cl2, DMF, MeOH, and dithiothreitol.
Preparation of Orf Virus Entry Fusion Proteins

Purification by semi-preparative RP-HPLC (Vydac 218TP diphenyl, Grace, 10 μ, 10 × 250 mm, 5 μm/m with a 1-450 mM gradient over 64 cm (100 μl/min) afforded 430 mg of 3 (theoretical 75.0 mg, yield 56.7% with respect to Thr-Ser-Cys(5)OSR m/z [M + H]+ calc., 1198.7, found, 1199.3.

Reductive desulfuration OFRV03625-110 Thr3-Ser5-Cys14-Lys19 (12.1 mg, 13.3 μmol) was dissolved in a solution of 0.2 M phosphate, 6 M guanidine hydrochloride containing 2.2-z-azobis(2-methylpropionylo)propanehydrochloride (VA-044) (125 μM), glutathione reduced (50 mM) and TCEP. HCl (250 mM) (1.21 ml) at pH 7.0 [9]. The pH of the solution was readjusted to 6.8 as it had been lowered by peptide addition; the reaction was heated to 37°C and progress monitored by LC-MS. After 8 h, the pH was adjusted to 2 with 5 M HCl, diluted to 4 ml with 0.1% acq. TFA and stored at 4°C overnight. Purification by semi-preparative RP-HPLC (Gemini C18, Phenomenex, 5 μ, 10 × 250 mm, 5 μm/m) with a linear gradient 0.1% B (min) to 40% B (min) over 40 min at 1 ml/min afforded 45.5 mg of 4 (theoretical 120.5 mg, yield 37.5%; m/z [M + H]+ calc., 1194.7, found, 1194.3.

ORFV049 (46-128) The 1-Glu3-Cys14-COSR (50 mg, 87.7 μmol) and ORFV049 (46-128) Cys14-Arg22 (22.1 mg, 87.7 μmol) were dissolved in a solution of 0.2 M phosphate, 6 M guanidine hydrochloride containing 4 M TFA (100 mM) and TCEP. HCl (20 mM) (8.71 ml) at pH 7.0. The pH was readjusted to 6.8. The reaction was kept at room temperature for 4 h. MeOH/H2O (145 μl, 0.2 M) was then added and pH adjusted to 7.0 with 5 M HCl. After 9 h, the pH was readjusted to 2 with 5 M HCl, diluted to 4 ml with 0.1% acq. CH3CN (0.1% TFA), and purified by semi-preparative RP-HPLC (Vydac C4, 15 μ, 10 × 250 mm, 5 μm/m) with a 1-40% B gradient over 60 min (1% B/min) to afford 12 (10.7 mg) (theoretical 12.5 mg, 87.7 μmol). The required ligation site cysteine could be transformed to the native residue via post-ligation reactions. Examination of the truncated ORFV036-25-110 sequence [6,34] (Figure 1) for a suitable ligation site revealed a Ser4-Thr motif. This disconnection would result in two fragments of 43 amino acids accessible by Boc SPPS. By using cysteine as a surrogate for Ala44, we envisaged that a fast radical-based deamidation to convert Cys to Ala would afford the desired native protein and so set out to synthesize these peptides (1 and 2, Scheme 1 [12]).

Synthesis of peptides Thr3-Ser5-Cys14-COSR (1) and Cys14-Arg22 (2) was undertaken using Boc SPPS on in-house prepared amimomethyl functionalised polystyrene resin with an HF-labelable PMI linker [67]. The crude HPLC profile of 1a (R = CH3CH2COO− glycyl-CH2-NH2) was found to be poor; incorporation of an additional two lysine residues into the thioester tag 1b, R = CH3CH2COO− Lys44-Lys54-CH2-NH2 resulted in an improved HPLC [13]. Additionally, LC-MS of both crude 1b and 2 showed significant oxidation of methionine residues to methionine sulfone, which complicated the synthesis of Orf Virus Entry Fusion Proteins.

Results and Discussion

Synthesis of ORFV036-25-110

The orfV036 ORF encodes a 110 AA protein predicted by the GCG Wisconsin Package: Program TtomoMlt (Accelrys, San Diego, CA) to contain an N-terminal membrane domain. With the eventual aim of studying the interactions between ORFV035 and 409 in mind, we decided to omit this portion of the protein, as hydrophobicity of the domain could impede fusion on such studies. In NCL, a peptide with an N-terminal cysteine residue can be joined to one with a C-terminal cysteine residue by a native amide bond [10,11]. ORFV036 (46-128) does not contain any cysteine residues, thus we sought a synthetic strategy where

![Figure 1](source_url)

**Figure 1.** Sequences of native ORFV035 and 409. Putative transmembrane domains are shown in red.
HPLC purification. However, by performing a one-step final Boc deprotection and on-resin methionine reduction using TFA, TBAI and Me$_3$S prior to cleavage of the peptides, reduced methionine residues were exclusively obtained [14]. RP-HPLC purification of crude polypeptides using the slow gradient protocol (B) afforded the pure peptides in reasonable yields: 1b: 5.7% m/z [M + H]$^+$ calc. 876.1, found, 875.7; 2: 14.2% m/z [M + H]$^+$ calc. 956.7, found, 956.5.

Peptides 1b and 2 were subjected to standard NCL conditions and dissolved in ligation buffer consisting of 6 M Gn-HCl, 200 mM Na$_2$HPO$_4$, 200 mM 4-MPA and 20 mM TCEP (pH = 7) at room temperature to achieve final concentrations of ca. 3 mM [15]. Ligation was complete within 24 h as determined by LC-MS, and subsequent solid-phase extraction afforded the crude ligated product (Figure 2). Attempts to effect the desulfurisation reaction on this crude material were then performed but were unsuccessful.

Figure 2. A-C: Native chemical ligation (NCL) of 1b and 2, followed by formyl removal of the ligation product in a one-pot procedure (A, NCL, t = 0 min; B, NCL, t = 24 h; and C, formyl deprotection, t = 30 h). The ligation was monitored by LC-MS at 210 nm (Grace C4, 2.1 x 50 mm, 5% B to 65% B, 3% B per min, 0.2 mL/min). The large early-eluting peaks at *7 min* and *10 min* (A and B) are the buffer components and the thiol catalyst, respectively. (C, LC-MS of the final ORF20(S2) - 110 product following a separate radical desulfurisation step (Agilent C3, 3.0 x 150 mm, 5% B to 65% B, 3% B per min, 0.3 mL/min). The solvent system used was A (0.1% TFA/w) and B (0.1% TFA/w) in CH$_2$Cl$_2$.

Scheme 2. Initial synthetic strategy for ORFV049(46–128) with ligation sites at Gln<sup>126</sup>-Ala<sup>127</sup> and Cys<sup>129</sup>-pGln<sup>130</sup>. 

Scheme 3. Revised one-pot synthesis of ORFV049(46–128) using ligation sites at native cysteine residues.
Figure 3. A–D: sequential one-pot native chemical ligation (NCL) of 10 and 11. Thz to Cys conversion and NCL of 9 to [10–11]A: 1st NCL, t = 0 h; B: t = 3.5 h; C: Thz to Cys conversion, t = 9 h; D: 2nd NCL, t = 24 h). The reactions were monitored by LC-MS (A: Grace C4, 2.1 × 50 mm, 5% B to 65% B, 3% B per min, 0.2 mL/min; B: C, D: Agilent C3, 3.0 × 150 mm, 5% B to 65% B, 3% B per min, 0.5 mL/min). The large eluting peak at * 1 min (A) and * 3 min (B, C and D) is the buffer component, and the peak at * 2 min (A) and * 12 min (B), C and D) is the third catalyst. We believe the complex peaks at * 15–19 min (A) to be a mixture of the ketones, mercaptophenylacetic acid exchanged peptide thioester, non-exchanged peptide thioester and product (f). LC-MS of the pure ORV9966-128 product (Agilent C3, 3.0 × 150 mm, 5% B to 65% B, 3% B per min, 0.5 mL/min). The solvent system used was A (0.1% eq. TFA/DCM) and B (0.1% TFA/DCM) in CH3CN.)

Preparation of Orf Virus Entry Fusion Proteins

Preparation of Orf Virus Entry Fusion Proteins

References


Appendix

CHEMICAL SYNTHESIS OF TRUNCATED ORF VIRUS ENTRY PROTEINS

References

References

References


References

References


