



Covariation-derived residue contacts in *ab initio* modelling and Molecular Replacement

Felix Šimkovic

Thesis submitted in accordance with the requirements of the
University of Liverpool
for the degree of
Doctor in Philosophy

September 2018

Institute of Integrative Biology
University of Liverpool
United Kingdom

Covariation-derived residue contacts in *ab initio* modelling and Molecular Replacement

Felix Šimkovic

This thesis is concerned with the application of predicted residue contacts in *ab initio* protein structure prediction and Molecular Replacement (MR).

Initially, in Chapter 3, research explored the use of predicted residue contacts to improve *ab initio* protein structure predictions, which were used to generate AMPLE ensemble search models for MR. The results proved highly encouraging: four additional targets were tractable where previous AMPLE attempts were unable to achieve structure solution. Furthermore, a novel approach to enhance β -rich decoy quality proved critical for an additional structure solution.

Leading on from the work in Chapter 3, it was essential to investigate different contact prediction algorithms and ROSETTA distance-restraints energy functions to optimise decoy quality. Results presented in Chapter 4 supported previous findings, which claimed that METAPSICOV produced the most precise contact predictions. Furthermore, results showed that target-specific decoy quality may be affected by the ROSETTA distance-restraint energy function used, which also translated into MR structure solutions in AMPLE.

Beyond different contact prediction algorithms and ROSETTA distance-restraint energy functions, alternative protein structure prediction algorithms exist. In Chapter 5, a study to compare the most promising alternatives to ROSETTA was conducted to explore potential alternatives for AMPLE. However, ROSETTA remained the optimal structure prediction algorithm to maximise structure solutions in AMPLE. A promising fragment-independent alternative, CONFOLD2, generated similarly accurate decoys, but the resulting AMPLE ensembles did not produce successful MR structure solutions.

AMPLE's cluster-and-truncate routine was originally developed to process contact-unassisted decoys. However, more accurate starting decoys, such as those deriving from contact-assisted modelling, may require processing differently to generate the ensemble search models. The findings in Chapter 6 demonstrated that decoy quality could be reliably predicted by measuring the satisfaction of the long-range contact predictions used initially to restrain the folding procedure. Excluding the decoys that satisfied the fewest long-range contacts enabled further structure solutions of targets that were previously intractable.

Lastly, in Chapter 7, contact-driven selection of supersecondary structure elements or subfolds identified by fragment picking software was explored as a novel route to search models for unconventional MR. Preliminary results of this approach showed promise. Two out of four protein targets were solved with fragments extracted from unrelated protein targets which, crucially, satisfied many predicted residue contacts.

Acknowledgements

First and foremost, I would like to ...

To Joanna, grandma and my family.

Contents

List of Figures	vi
List of Tables	vii
List of Equations	viii
1 Introduction	1
2 Materials & Methods	3
3 Evolutionary covariance in <i>ab initio</i> structure prediction-based Molecular Replacement	5
4 Evaluation of ROSETTA distance-restraint energy functions on contact-guided <i>ab initio</i> structure prediction	7
5 Alternative <i>ab initio</i> structure prediction algorithms for AMPLE	9
6 Decoy subselection using contact information to enhance MR search model creation	11
7 Protein fragments as search models in Molecular Replacement	13
8 Conclusion & Outlook	15
A Appendix	17

List of Figures

List of Tables

List of Equations

Chapter 1

Introduction

Chapter 2

Materials & Methods

Chapter 3

Evolutionary covariance in *ab initio* structure prediction-based Molecular Replacement

Chapter 4

Evaluation of ROSETTA distance-restraint energy functions on contact-guided *ab initio* structure prediction

Chapter 5

Alternative *ab initio* structure prediction algorithms for AMPLE

Chapter 6

Decoy subselection using contact information to enhance MR search model creation

Chapter 7

Protein fragments as search models in Molecular Replacement

Chapter 8

Conclusion & Outlook

Appendix A

Appendix