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, 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 original study, it was essential to investigate different contact prediction algorithms and ROSETTA distance-restraints energy functions to optimise decoy quality. Results in this study supported previous findings, which claim METAPSICOV to produce the most precise contact predictions. Furthermore, contact predictions introduced to ROSETTA using the `FADE` energy function outperforms the `SIGMOID` in decoy quality. However, findings demonstrate that the latter produces decoys more suitable for MR structure solutions in AMPLE.

Beyond different contact prediction algorithms and ROSETTA distance-restraint energy functions, alternative protein structure prediction algorithms exist. 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, however resulting AMPLE ensembles did not translate into MR structure solutions.

AMPLE's cluster-and-truncate routine was originally developed to process contact-unassisted decoys. However, more accurate starting decoys may require alternative processing to generate ensemble search models. The findings in this chapter demonstrated the successful application of estimating decoy quality by the satisfaction of long-range contact predictions used initially to restrain the folding procedure. Excluding the decoys that satisfy the least long-range contacts provided further structure solutions previously intractable.

Lastly, contact-driven selection of supersecondary structure elements or subfolds during fragment picking was explored to identify suitable search models for unconventional MR. Preliminary results of this approach strongly hinted towards a promising new approach. Two out of four protein targets were solved with fragments extracted from sequence-independent protein targets, which crucially satisfied many predicted residue contacts.

Acknowledgements

Contents

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