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.

The initial work in this thesis explored the use of predicted residue contacts to improve *ab initio* protein structure predictions, which were used to generate ensemble search models for Molecular Replacement in AMPLE. The results proved highly encouraging. Five additional targets were tractable where previous AMPLE attempts would have been unable to achieve structure solution. In particular, the improved decoy quality appeared to be the main reason for the extended target tractability.

Following on from the initial proof-of-concept study, different contact prediction algorithms and ROSETTA energy functions were trialled to identify the optimal strategy to generate the most accurate decoys for unconventional Molecular Replacement in AMPLE. The findings showed supported previous claims that METAPSICOV produces the most precise contact predictions. Furthermore, the ROSETTA FADE energy function outperforms the SIGMOID function. Nevertheless, results also demonstrate that the most accurate structure predictions do not achieve the most Molecular Replacement structure solutions.

Beyond different contact prediction algorithms and ROSETTA energy functions, many alternative fragment-based and fragment-independent protein structure prediction algorithms exist. In this chapter, results highlighted that ROSETTA remains the optimal structure prediction algorithm in combination with AMPLE to maximise structure solutions.

The most accurate protein structure predictions may not be processed optimally in AMPLE. Thus, it is important to explore alternative ensembling strategies when more accurate contact-assisted decoys are used in AMPLE. The findings in this chapter demonstrated the successful application of estimating decoy quality by the satisfcation 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 Molecular Replacement. Preliminary results of this approach strongly hint towards a potential 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.

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