

AMPLE

Ab Initio Modelling of Proteins for Molecular Replacement

Jens Thomas

Institute of Integrative Biology

December 2014



UNIVERSITY OF
LIVERPOOL

Outline

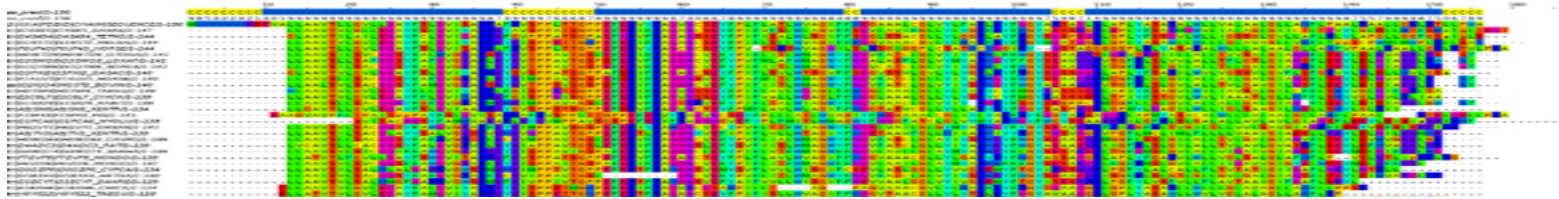
- What is AMPLE?
- How it works
- When it works
- Why it works
- How to use it

AMPLE

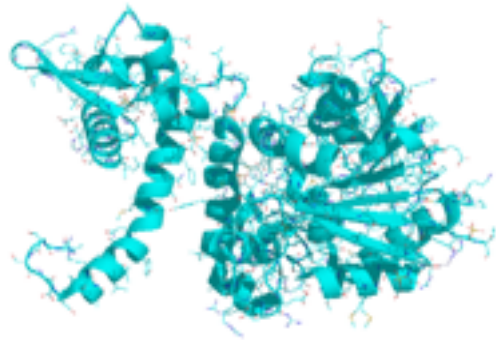
Ab initio **M**odelling of **P**roteins for mo**L**Ecular replacement

- Joint development by CCP4 and the University of Liverpool
- AMPLE is a comprehensive project to assess the suitability of using cheaply obtained *ab initio* models in molecular replacement
- An additional goal of the project is to make AMPLE into an automated software tool that can be made generally available to potential users through CCP4

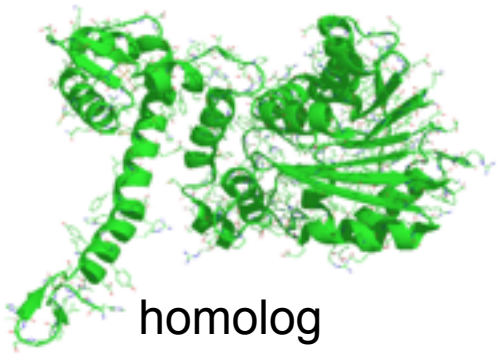
Molecular Replacement



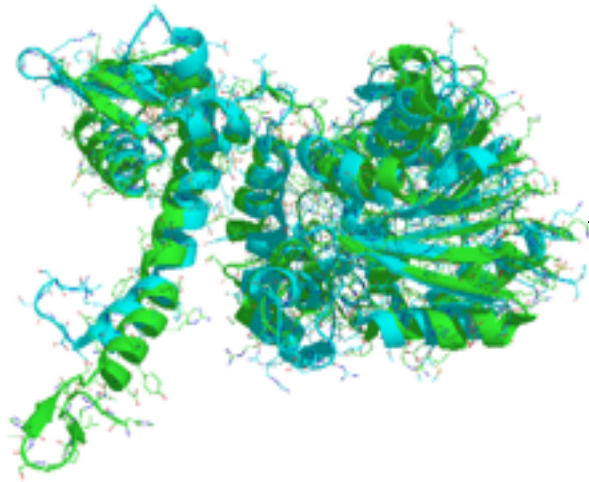
Sequence search for a suitable homolog



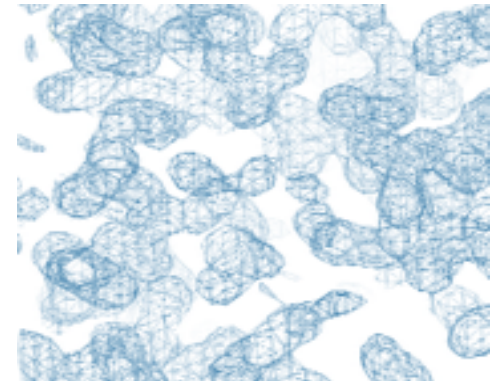
target



homolog



- Phases from homolog
- Intensities/positions from target



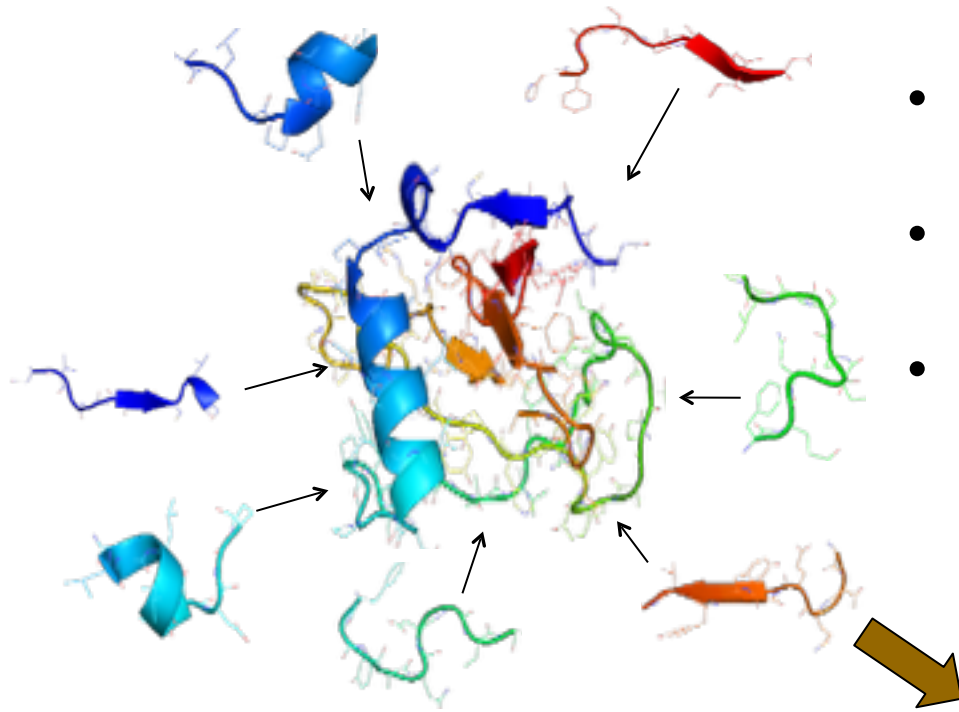
electron density map

A replacement for Molecular Replacement

- No suitably similar homolog
- Small changes in sequence or crystallisation conditions can lead to different polymorphs crystallising (coiled-coils)
- Not sure exactly what is in the crystal



Ab Initio Modelling



- Create database of 3- and 9-residue fragments.
- Start from linear chain and splice in fragment geometry.
- Randomly move the fragments under simplified forcefield
- Score the resulting structures.

- 1000's of “decoys” created.
- Decoys are clustered and centroid representatives of largest cluster are considered candidate fold predictions
- Refinement under a more realistic physics-based force field



Decoys vs. final models

- for normal usage > 20,000 decoys are recommended
- refinement stage can take weeks on a supercomputer

This is beyond the capabilities of most crystallography groups.

But:

- creating 1000 decoys ~ 1/2 day on a single CPU with ROSETTA
- QUARK modelling is available online and takes ~ 1 day

Ab Initio Models and Molecular Replacement

Simple *ab initio* modelling

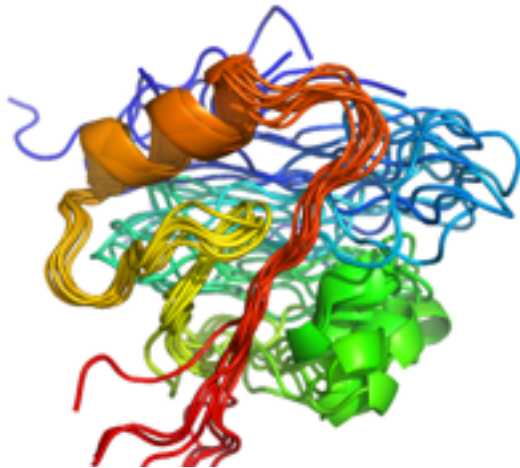
Clusters of similar structures

Within/between clusters, similarity indicates accuracy $\therefore$ trim inaccurate regions leaving more reliable **core**

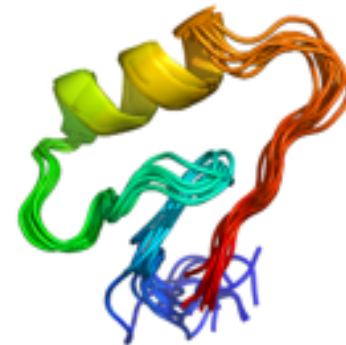
Molecular Replacement

Works well with superposed **ensembles** approximating the target

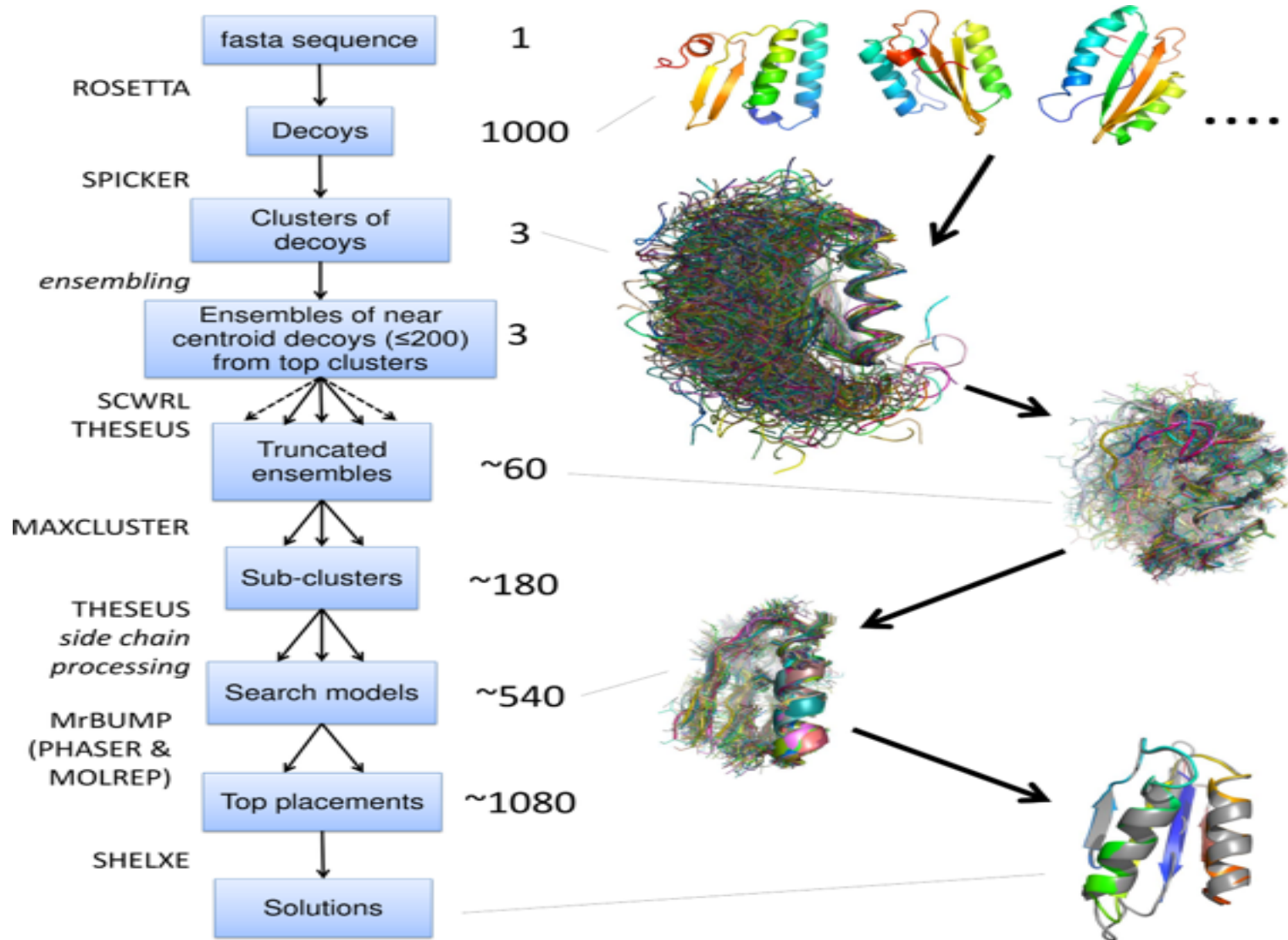
May only require a **partial** model



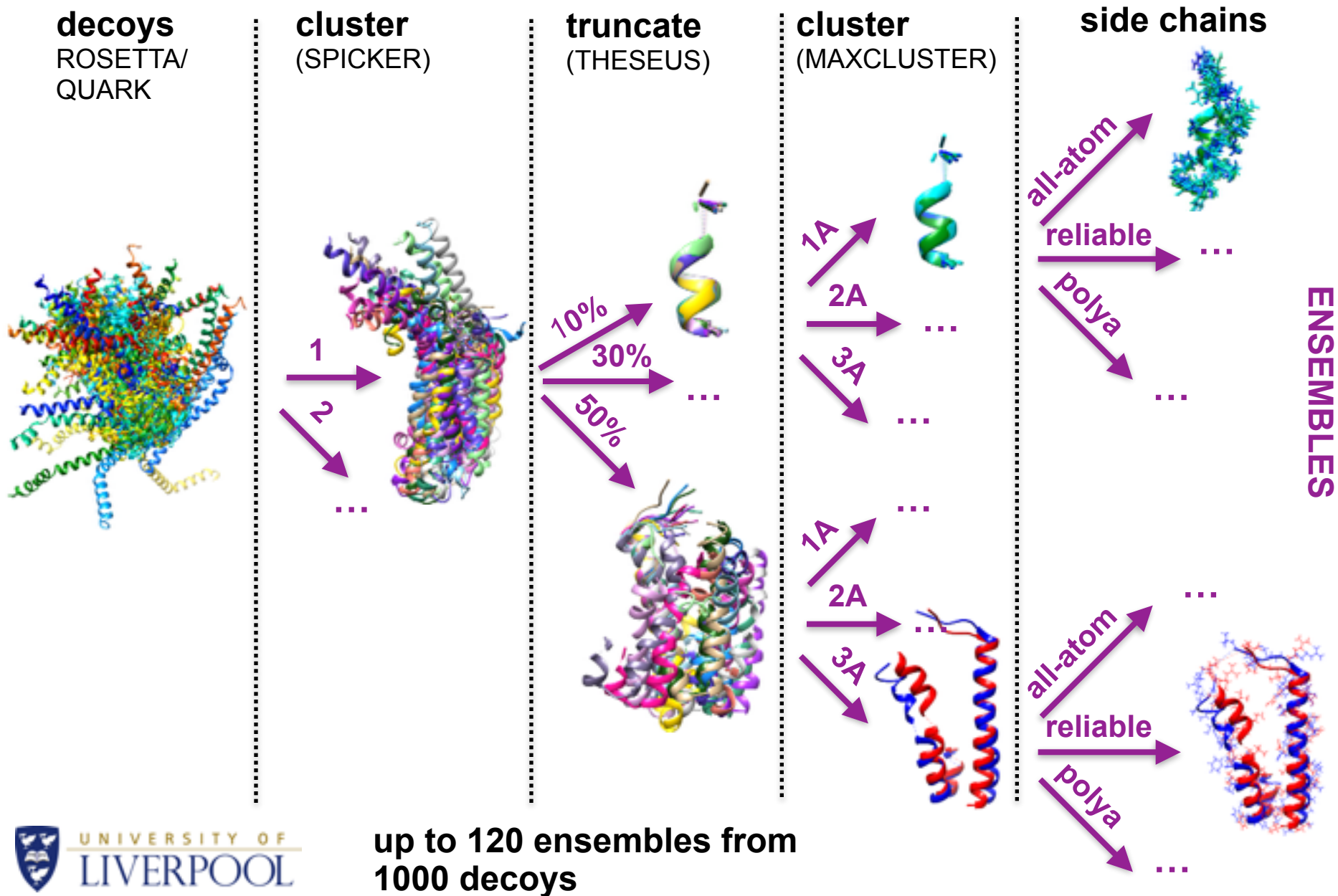
→
trim to accurate core



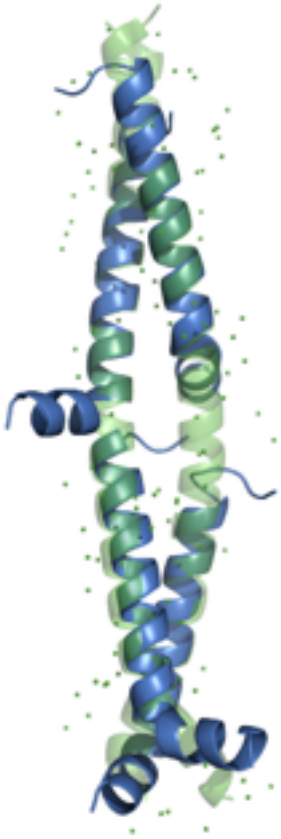
The AMPLE pipeline



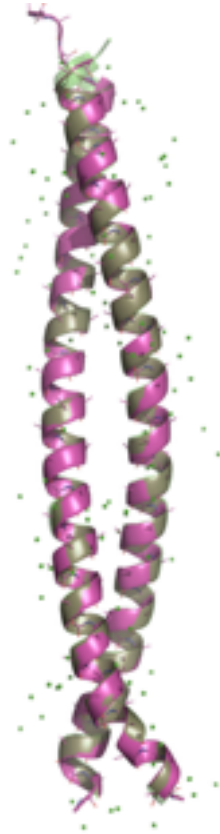
Pathways to ensembles



MRBUMP MR pipeline



PHASER: places search model



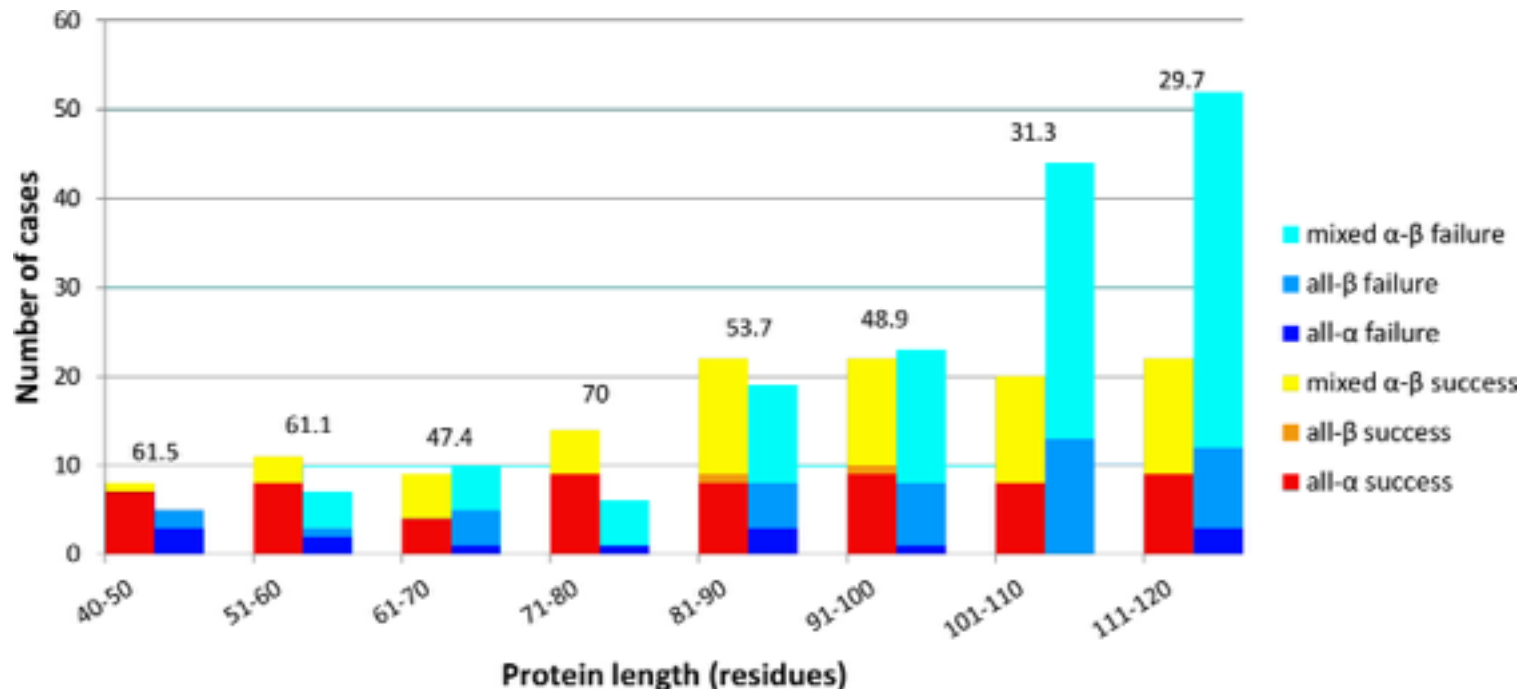
SHELXE: density modification and tracing of the main chain



ArpWARP/BUCCANEER completes the rebuild

Initial AMPLE success

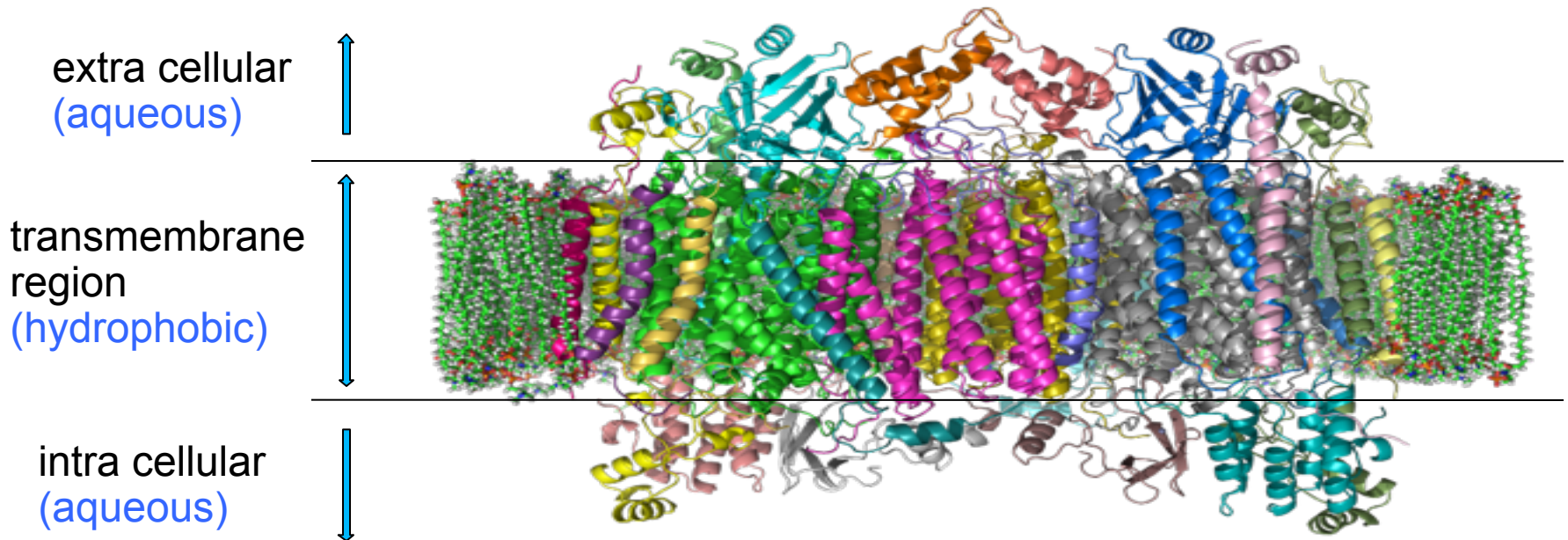
- Test set of 295 small globular proteins (40-120 residues) from the PDB, resolution of 2.2 Å or better
- Single molecule in the asymmetric unit
- Mixture of all- α , all- β and mixed α - β secondary structure



Initial success rate: all- α : 80%, all- β : 2%, mixed α - β : 37%

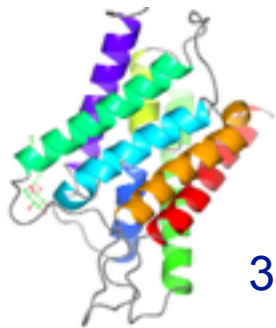
Transmembrane Proteins

- Experimentally very difficult to work with/crystallise
- Represent ~30% of all proteins
- Make up < 3% of structures in the Protein Data Bank
- Presence in the membrane constrains their shape, so they can be easier to model
- MR with *ab initio* transmembrane models hasn't been tried yet

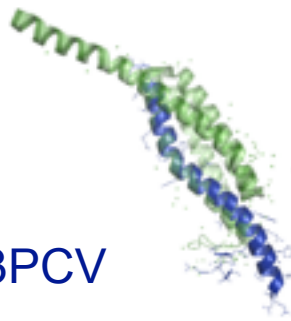


Transmembrane Results

- Selected 18 alpha-helical transmembrane proteins:
 - 23 – 249 residues
 - 1.45 – 2.5A resolution
- 7 clear successes
- 5 possible successes – might work with manual rebuilding
- Successes extended to 223 residues (3GD8).

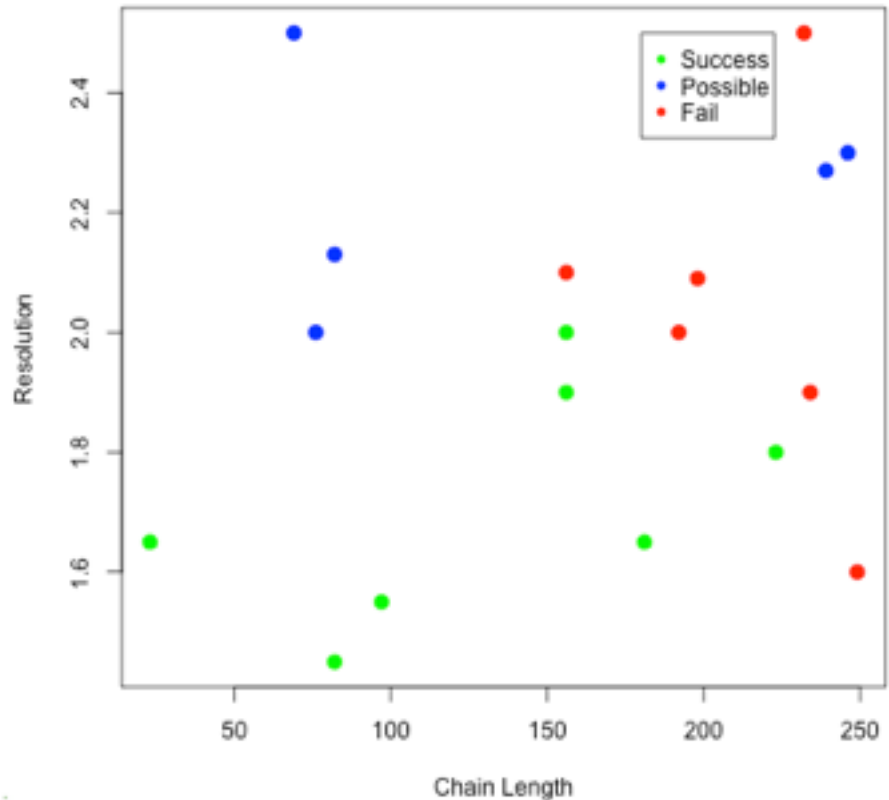


3GD8



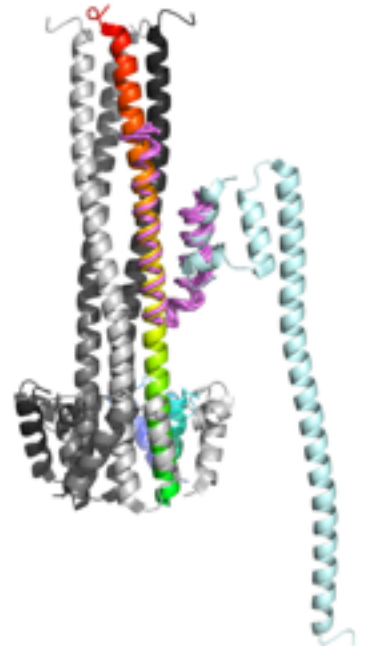
3PCV

Successes as a function of chain length and resolution



Coiled-coils

- Among the most ubiquitous and functionally important class of proteins
- 2-5 amphipathic helices wound around each other in a supercoil with sequence repeats of 7 or 11 residues
- Notoriously difficult to solve with MR techniques
- But, are exclusively α -helices
- Selected 94 cases from the PDB:
 - $12 < \text{residues} < 253$
 - $0.9 < \text{resolution } \text{\AA} < 2.91$

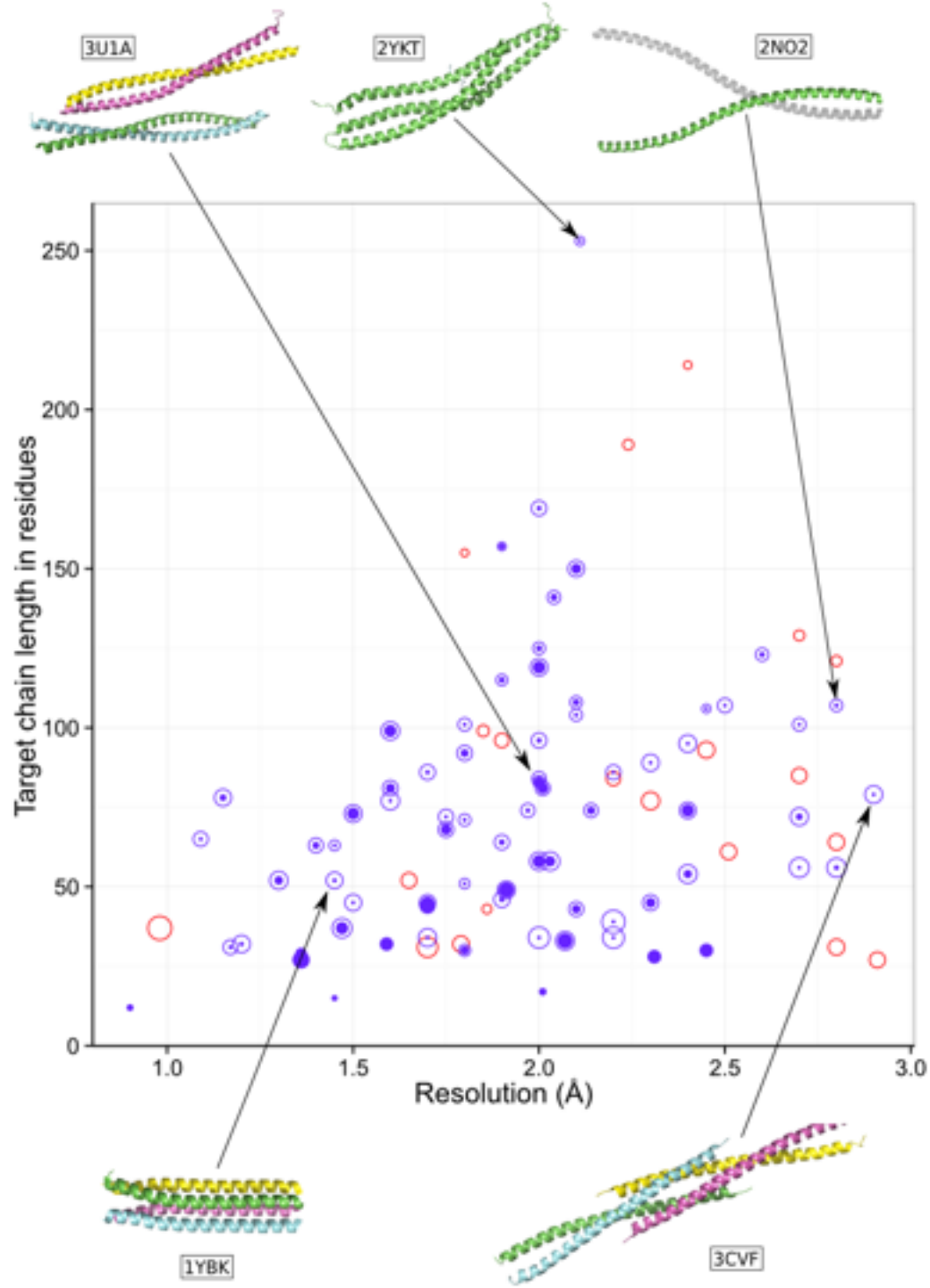


Coiled-coils

Solved more than **78%** with an average CPU time of less than 24 hours.

Successes included:

- **3U1A**: 334 residues
- **3CVF**: resolution of 2.8Å
- **2NO2**: a domain of Huntingtin-interacting protein 1, that contains a long, unconventional coiled-coil-like assembly originally phased experimentally using MAD
- **1YBK** right-handed coiled coil

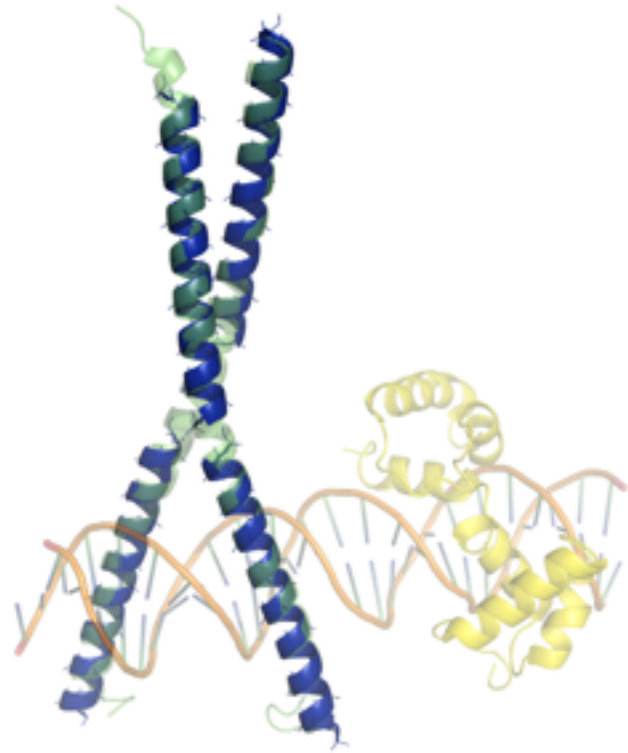


Complex Ambitions

Coiled-coils are often involved in complexes - could AMPLE's ability to solve coiled-coils extend to complexes?

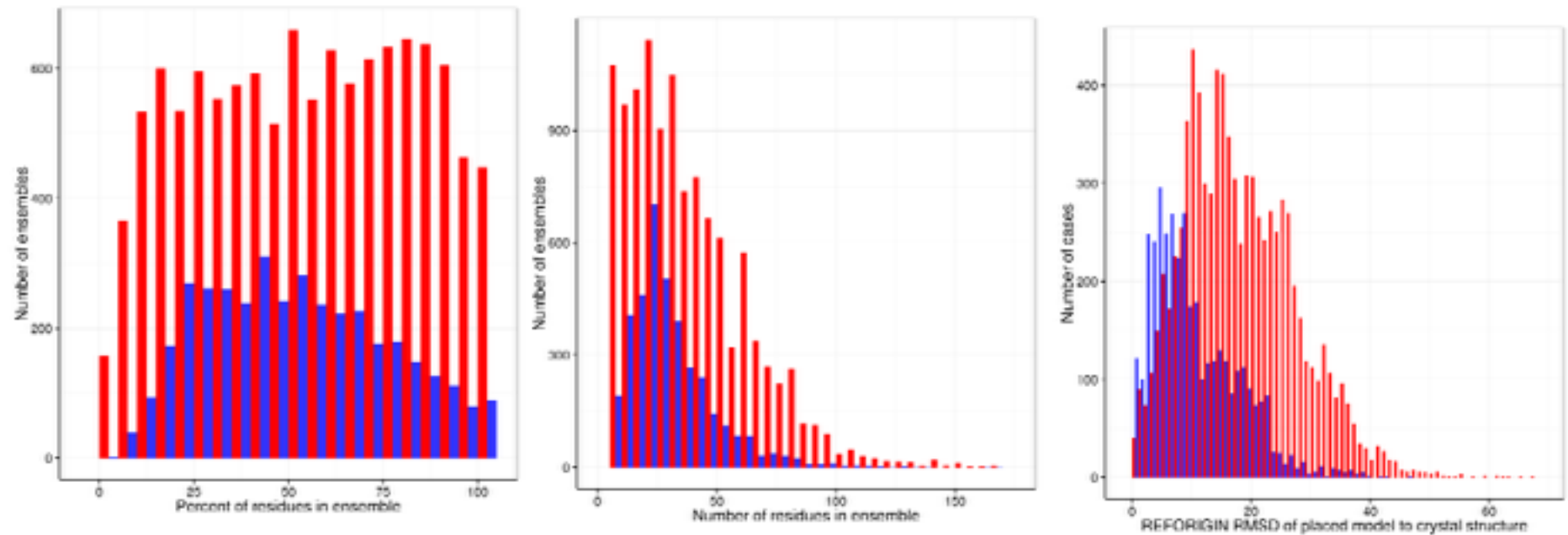


- **1X79**: three chains, 322 residues, ~70% coiled-coil @ 2.41Å



- **1H8A**: MyB/DNA complex of 278 residues ~ 60% coiled-coil @ 2.23Å.

What is solving these structures?



- Sample all proportions of the model but usually less than half
- Short (~25-30 residues) search models
- Search model seems to be in the wrong place!

Quantifying the fragments

Register-Independent Overlay (RIO): indicates how many C α atoms match between crystal structure and search model (<1.5Å) for stretches ≥ 3 C α .

RIO_in: number where sequence of model and structure match

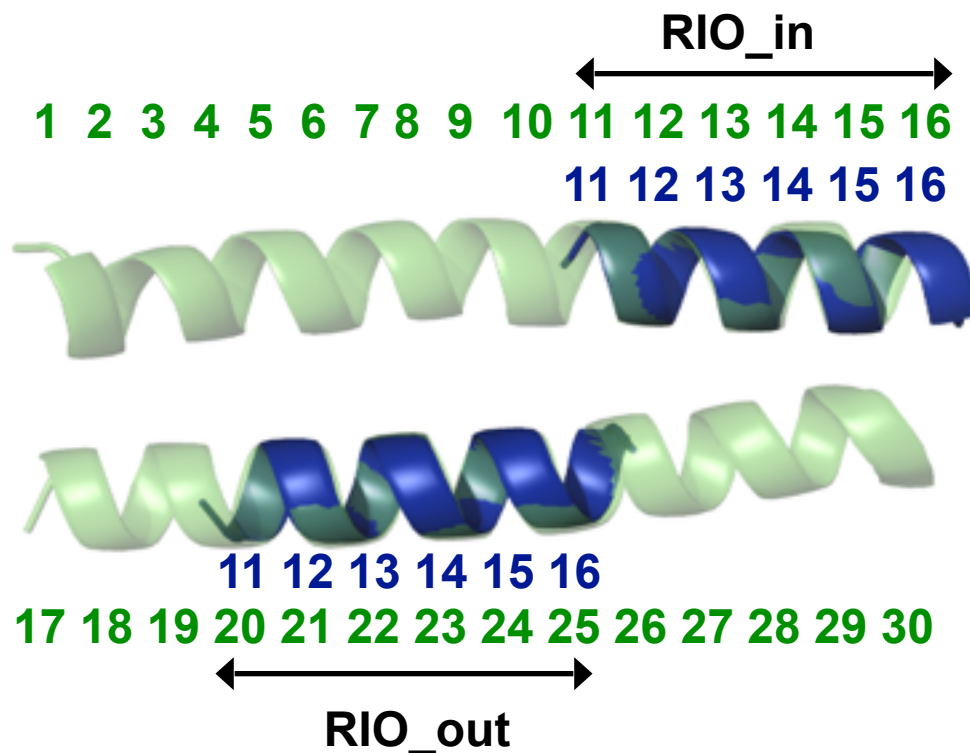
RIO_out: number where sequence of model and structure are different

Model

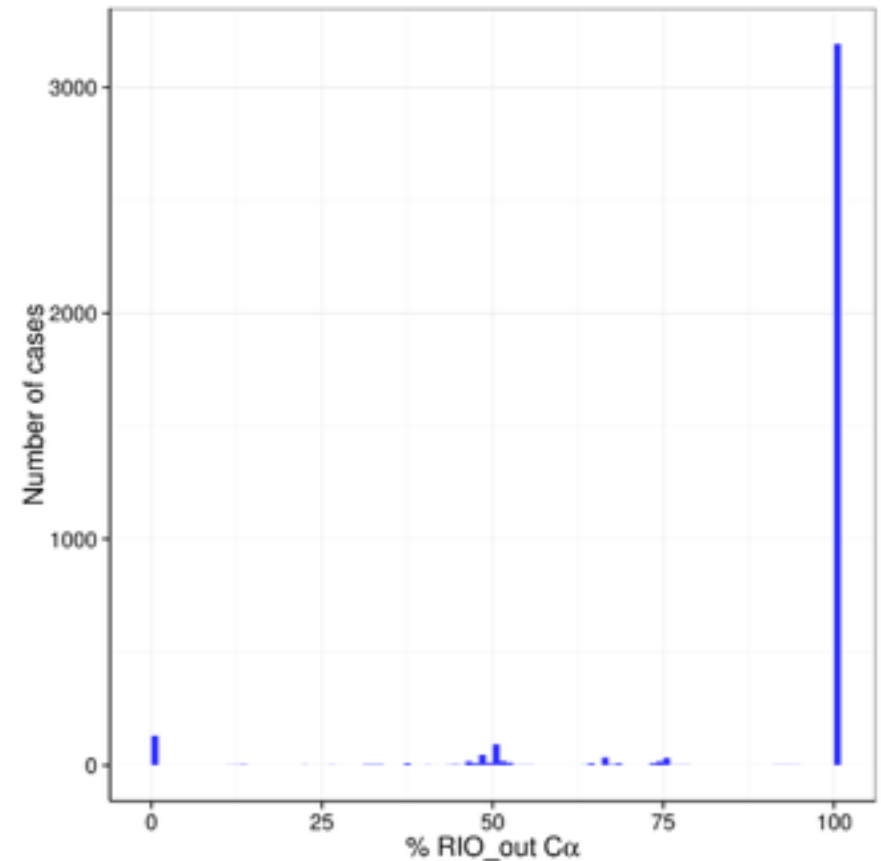
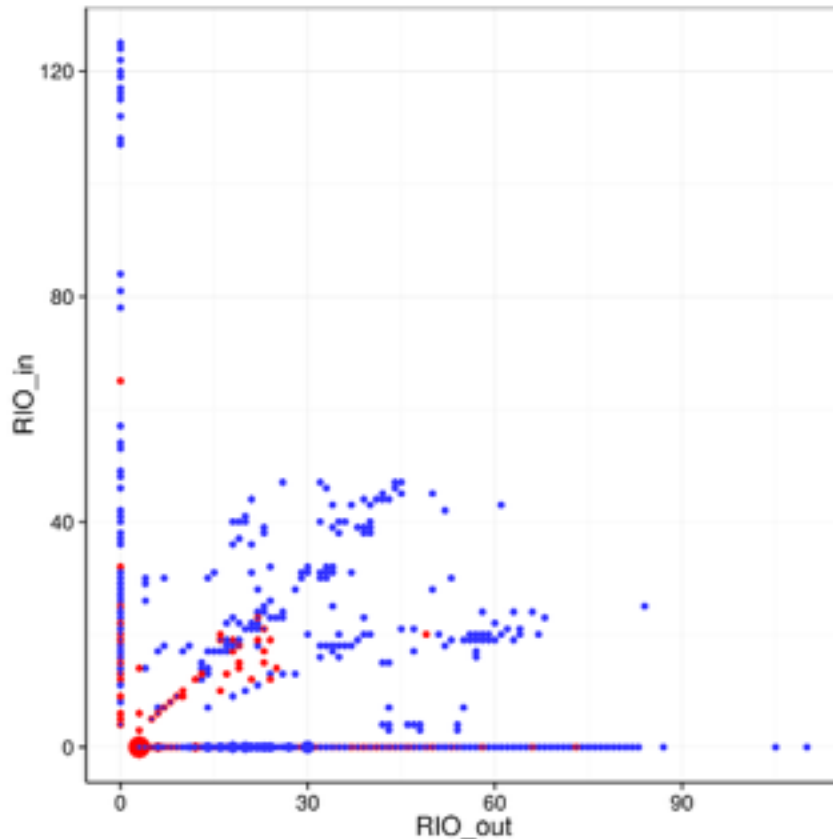


11 12 13 14 15 16

MR result

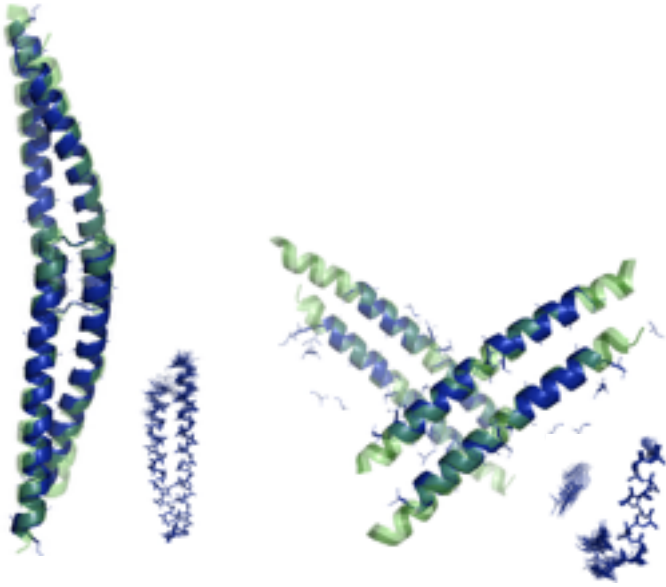


RIO Analyses



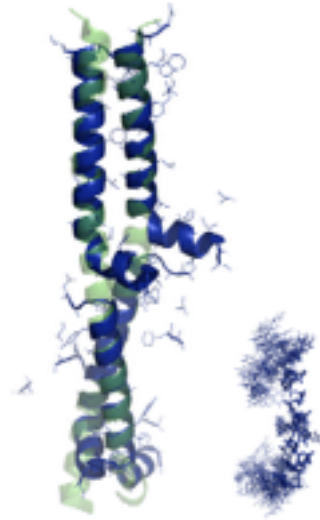
- there are fragments placed correctly and incorrectly wrt sequence
- some solutions contain mixtures of both
- the majority are out-of-register

Actual cases

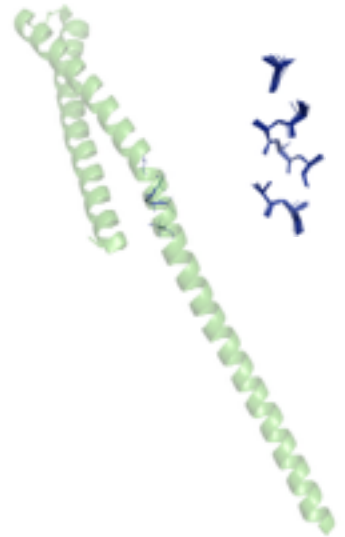


2IC9: 2 copies
RIO=125
RIO_in=125

3H00: 8 copies
RIO=110
RIO_out=110



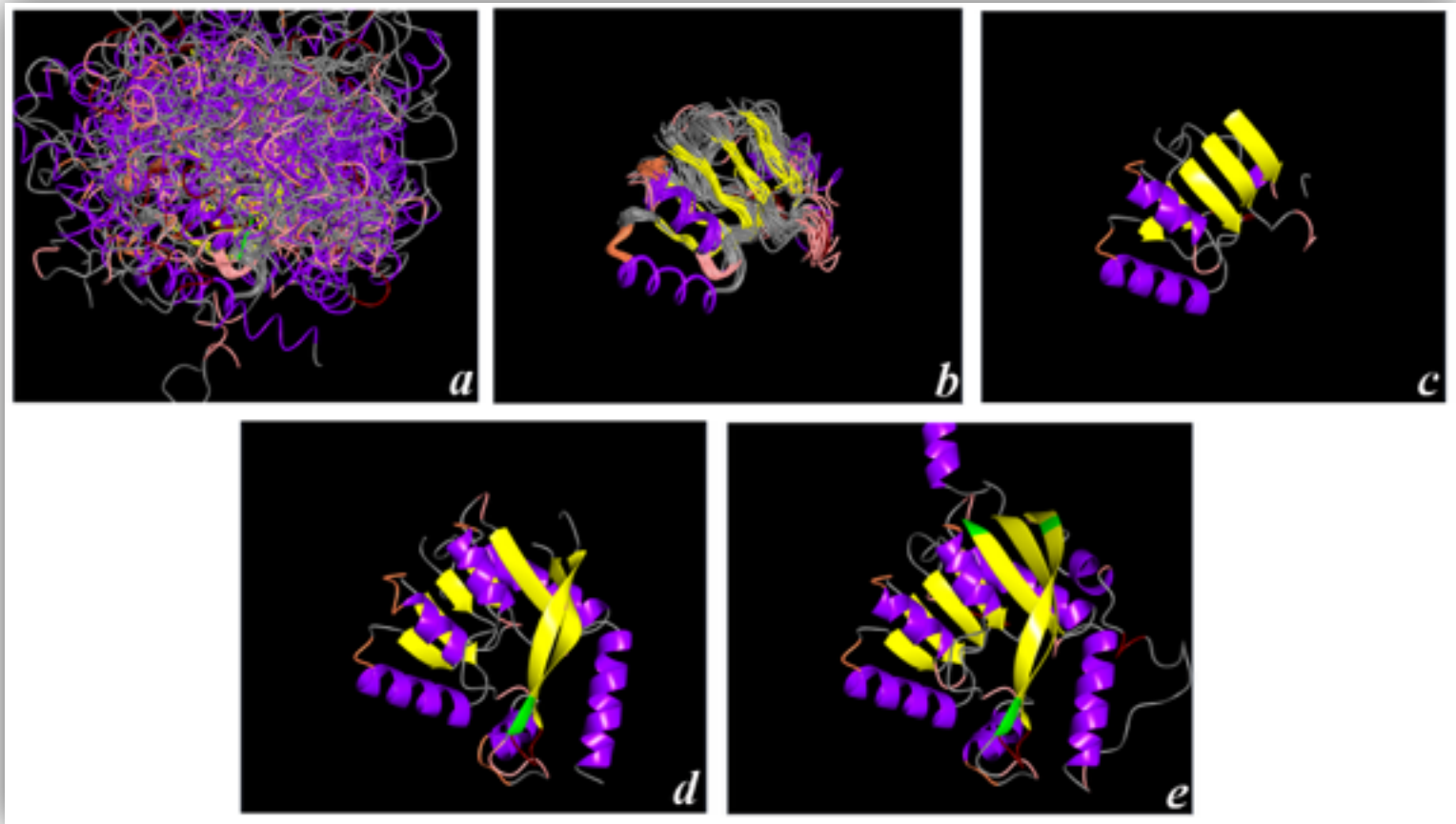
1UII: 4 copies
RIO=91
RIO_in=23
RIO_out=68



3OKQ: 1 copy
RIO=4
RIO_out=4

A mixture of correct and incorrect placements

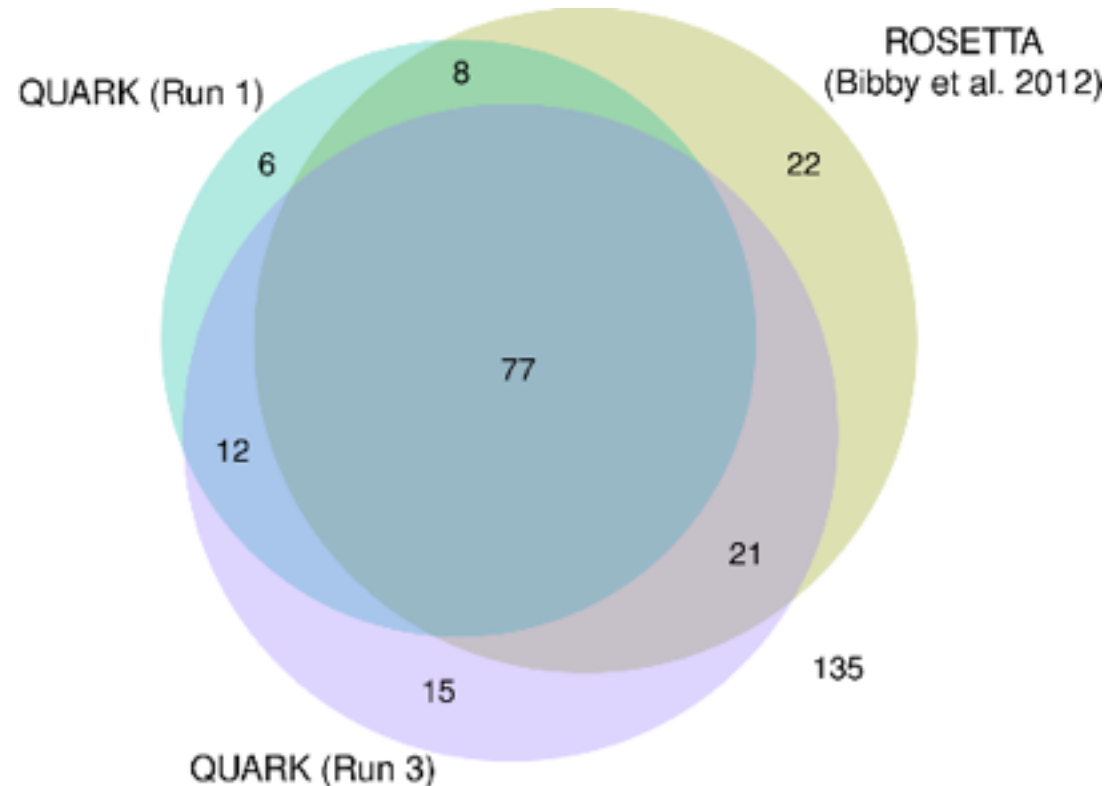
Exploiting distant homologues



a. Clustered decoy models, *b*. Truncated ensemble, *c*. Positioned MR solution, *d*. Shelxe c-alpha trace, *e*. Completed structure

QUARK vs ROSETTA

- Re-did testing using QUARK-generated decoy models
- Results are similar to ROSETTA but there are some targets that could only be solved by one or the other

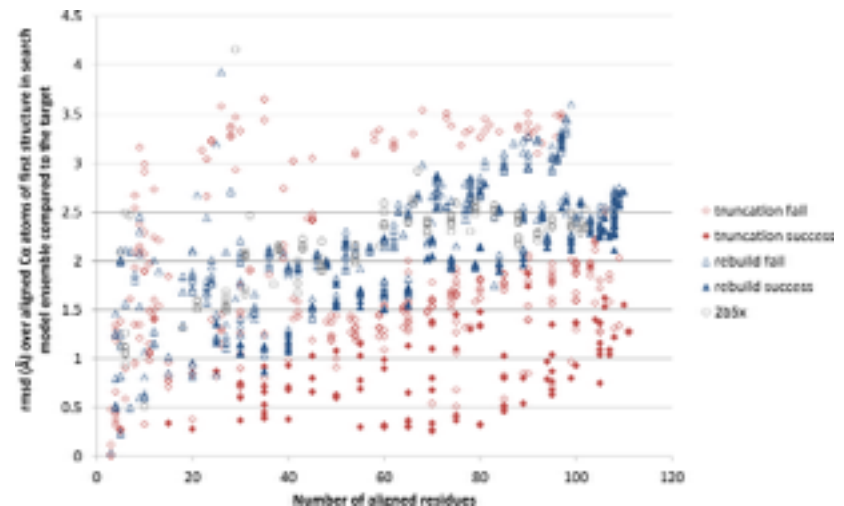


Remodelling related NMR structures or distant homologues

- **Truncation**, where the default AMPLE truncation strategy is applied to the models in the NMR ensemble.
- **Rebuilding**, where the NMR ensemble models are subjected to idealisation, comparative modelling and relaxation with ROSETTA, before being subjected to the standard truncation strategy.

25 test cases (FINDCORE)

- 19 by truncation alone
- 3 with remodelling



Practicalities



- Available since CCP4 6.3.0.
- Requires installation of several non-CCP4 packages (all freely available to academics):
 - **SHELXE**
 - **Rosetta** (optional)
 - **SCWRL** (optional)
- QUARK models now available so can do modelling online.
- Models can come from anywhere – just need a directory of PDB files
- Can exploit multiple processors on a local desktop, but also runs out-of-the-box on a cluster/supercomputer (SGE/LSF).

Available in CCP4i

The screenshot displays the CCP4 Program Suite 6.3.0 CCP4Interface 2.2.0 running on ccp4i1. The main window shows a list of jobs under the 'Molecular Replacement' category. A secondary window, 'AMPLE (beta) - Ab initio modelling for Molecular Replacement initial parameters from /home/rolo/', is open, showing input files and parameters for a job titled '2cmp ab initio model MR'.

CCP4 Program Suite 6.3.0 CCP4Interface 2.2.0 running on ccp4i1 Project: aps5-raj

Molecular Replacement

Job ID	Date	Status	Description
15	11 Jul 12	FINISHED	arp_warp_classic automated model building starting
14	11 Jul 12	FINISHED	buccaneer_pipeline [No title given]
13	11 Jul 12	FINISHED	refmac5 refine jao soln
12	10 Jul 12	FINISHED	buccaneer_pipeline [No title given]
11	10 Jul 12	FINISHED	arp_warp_classic automated model building starting
10	10 Jul 12	FINISHED	refmac5 refine jao soln
9	10 Jul 12	FINISHED	refmac5 refine jao soln
8	10 Jul 12	FINISHED	refmac5 refine jao soln
7	10 Jul 12	FINISHED	refmac5 refine jao soln
6	10 Jul 12	FINISHED	arp_warp_classic automated model building starting
5	10 Jul 12	FINISHED	buccaneer_pipeline [No title given]
4	10 Jul 12	FINISHED	arp_warp_classic automated model building starting
3	10 Jul 12	FINISHED	refmac5 refine jao soln
2	10 Jul 12	FINISHED	refmac5 refine jao soln
1	10 Jul 12	FINISHED	refmac5 refine jao soln

AMPLE (beta) - Ab initio modelling for Molecular Replacement initial parameters from /home/rolo/

Job title: 2cmp ab initio model MR

Input Files

SEQ In: ample 2cmp.fasta Browse View

MTZ In: ample 2cmp.mtz Browse View

F: FP Sigma SIGFP

Free-R: FREE

Number of models to generate: 500

Number of Processors: 1

Fragment Files

To generate fragment files for your sequence you need to use the Robetta Fragment library server

Click here to go to online Robetta Server (registration required)

3mers: 2aXXX_03_05.200_v1_3 /home/rmk65/CCP4-Projects/ample Browse

9mers: 2aXXX_09_05.200_v1_3 /home/rmk65/CCP4-Projects/ample Browse

Rosetta Installation

Rosetta Installation Directory: /home/rmk65/opt/src/rosetta3.4 Browse

Run Save or Restore Close

The screenshot shows the Robetta Beta website, a full-chain protein structure prediction server. It displays a comparison between a model and a target structure, highlighting the accuracy of the prediction.

ROBETTA BETA
Full-chain Protein Structure Prediction Server

Model 1 Target - 70513

2.66 Å over 42 residues

0.84 Å over 39 residues

REGISTRATION
[Register / Update] [Login]

DOCUMENTATION
[Docs / FAQs]

SERVICES
Domain Partitioning & 3-D Modeling [Queue] [Submit]
Interface Amino Acid Scanning [Queue] [Submit]
Fragment Libraries [Queue] [Submit]
DNA Interface Protein Scanning [Queue] [Submit]

RELATED SITES
RosettaBackrub Server
RosettaAntibody Server
RosettaDesign Server
RosettaDock Server
RosettaCommons
Folp
RosettaOnline
RosettaProtein Folding Project

Create models with Quark

← → ↻ zhanglab.ccmb.med.umich.edu/QUARK/ ☆ ⌵ 🔴 ☰

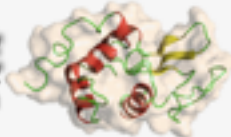
 **Zhang Lab**

UNIVERSITY OF MICHIGAN

Home Research Services Publications People Teaching Job Opening Laborly

Online Services

- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MUSTER
- SEGMENT
- FO-MD
- ModRefiner
- REMO
- SPRING
- COTH
- BSpred
- SVMSEQ
- ANGLO
- BSP-SUM
- SAXSTER
- ThreeDom
- EvoDesign
- gPCR-I-TASSER
- TM-score
- TM-align
- MM-align
- NW-align

 **QUARK ONLINE**
Ab Initio Protein Structure Prediction 

QUARK is a computer algorithm for ab initio protein folding and protein structure prediction, which aims to construct the correct protein 3D model from amino acid sequence only. QUARK models are built from small fragments (1-20 residues long) by replica-exchange Monte Carlo simulation under the guide of an atomic-level knowledge-based force field. QUARK was ranked as the No 1 server in Free-modeling (FM) in [CASP9](#) and [CASP10](#) experiments. Since no global template information is used in QUARK simulation, the server is suitable for proteins which are considered without homologous templates.

Go to [Job Q24927](#) to view an example of QUARK output. The description of predicted feature files can be seen in [readme.txt](#). Questions about the QUARK server can be posted at the [Service System Discussion Board](#).

Cut and paste your sequence (in [FASTA format](#), less than 200 AA. Please submit bigger proteins to [I-TASSER Server](#)):

```
>1X79:AIPO8ID(CHAIN)SEQUENCE
GPLCSKSKRVNAIEEVNNVKKLTENVMSHSQCGAAAGSSDLMKELYQRCERMPTLFRLASOTEDNDEALADLQAN
DNLTQVINLYKQVRCEE
```

Or upload the sequence from your local computer:

No file chosen

Email: (mandatory, where results will be sent to. Only academic email accounts are acceptable.)

ID: (optional, your given name of the protein)

Optional: You can assign additional distance restraints for modeling (example is in [distrestraint.txt](#); format is described in [readme.txt](#)).

No file chosen



Quark Decoys

QUARK Ab Initio Results for Job Q20

Submitted Protein Sequence

Predicted Secondary Structure

Predicted Tertiary Structure

Predicted Solvent Accessibility

Download Model 1

Download Model 2

Download Model 3

Download Model 4

Download Model 5

Download Model 6

Download Model 7

Download Model 8

Download Model 9

Download Model 10

Estimated TM-score of Model 1: 0.3668 $\pm$ 0.0833
Estimated TM-score of the Best of Top 10 Model: 0.4183 $\pm$ 0.0764
[Download the Decoy Set Used for Clustering](#)

Estimated TM-score of Model 1: 0.3668 $\pm$ 0.0833

Estimated TM-score of the Best of Top 10 Model: 0.4183 $\pm$ 0.0764

[Download the Decoy Set Used for Clustering](#)

decoys.tar.gz

Quark Decoys

AMPLE - Ab initio modelling for Molecular Replacement

Help

Job title

Program Mode: **Molecular Replacement from QUARK decoys**

Input Files

SEQ in **PROJECT** Browse View

MTZ in **PROJECT** Browse View

F Sigma

Free-R

Number of Processors **1**

QUARK decoys

QUARK decoys can be generated using the QUARK server. Download the the decoy set used for clustering (decoys.tar.gz). Note: decoys can take up to 24 hours to generate.

[Click here to open QUARK Server in a browser](#)

QUARK decoys.tar.gz file: Browse

Molecular Replacement Options

Model Building Options

Advanced Options

Run Save or Restore Close

Running Rosetta

CCP4 Program Suite 6.3.0 CCP4Interface 2.2.0 running on ccp4i1 Project: aps5-raj

Molecular Replacement

- Analysis
- Model Generation
 - Phaser MR
 - Run Molrep - auto MR
 - Run MrBUMP
 - Run Balbes
 - Run AMPLE (beta)
- AMoRe Suite
- Utilities

15	11 Jul 12	FINISHED	arp_warp_classic	automated model building starting					
14	11 Jul 12	FINISHED	buccaneer_pipeline	[No title given]					
13	11 Jul 12	FINISHED	refmac5	refine jao soln					
12	10 Jul 12	FINISHED	buccaneer_pipeline	[No title given]					
11	10 Jul 12	FINISHED	arp_warp_classic	automated model building starting					
10	10 Jul 12	FINISHED	refmac5	refine jao soln					
9	10 Jul 12	FINISHED	refmac5	refine jao soln					
8	10 Jul 12	FINISHED	refmac5	refine jao soln					
7	10 Jul 12	FINISHED	refmac5	refine jao soln					
6	10 Jul 12	FINISHED	arp_warp_classic	automated model building starting					
5	10 Jul 12	FINISHED	buccaneer_pipeline	[No title given]					
4	10 Jul 12	FINISHED							
3	10 Jul 12	FINISHED							
2	10 Jul 12	FINISHED							
1	10 Jul 12	FINISHED							

AMPLE (beta) - Ab initio modelling for Molecular Replacement initial parameters from /home/rolo/

Job title: 2cmp ab initio model MR

Input Files

SEQ In: ample 2cmp.fasta Browse View

MTZ In: ample 2cmp.mtz Browse View

F: FP Sigma SIGFP

Free-R: FREE

Number of models to generate: 500

Number of Processors: 1

Fragment Files

To generate fragment files for your sequence you need to use the Rosetta Fragment library server

Click here to go to online Rosetta Server (registration required)

3mers: aXXXX_03_05.200_v1_3 /home/rmk65/CCP4-Projects/ample Browse

9mers: aXXXX_09_05.200_v1_3 /home/rmk65/CCP4-Projects/ample Browse

Rosetta Installation

Rosetta Installation Directory: /home/rmk65/opt/src/rosetta3.4 Browse

Run Save or Restore Close

ROBETTA BETA
Full-chain Protein Structure Prediction Server
www.bakerlab.org

REGISTRATION
[Register / Update] [Login]

DOCUMENTATION
[Docs / FAQs]

SERVICES
Domain Partitioning & 3-D Modeling
[Queue] [Submit]

Interface Amino Acid Scanning
[Queue] [Submit]

Fragment Libraries
[Queue] [Submit]

DNA Interface Protein Scanning
[Queue] [Submit]

RELATED SITES
RosettaBackrub Server
RosettaAntibody Server
RosettaDesign Server
RosettaDock Server
Rosetta Commons
Folp
Rosetta@Home
Rosetta Protein Folding Project

Model 1 Target - 70513
2.66 Å over 42 residues
0.84 Å over 29 residues

Summary

- AMPLE – a pipeline for Molecular Replacement with unconventional models
- Initial work with small globular proteins extended to look at transmembrane and coiled-coil proteins
- works particularly well with α -helical proteins
- solves structures with fragments or more complete models
- usually best for small, proteins with reasonable resolution.
- favourable cases can work > 330 residues and > 2.8Å resolution

Credits

- Dan Rigden, Olga Mayans – IIB, Liverpool University
- Jaclyn Bibby – Liverpool University
- Ronan Keegan, Martin Wynn – CCP4 (STFC)
- BBSRC
- Hartree Center
- **Phaser**: Randy Read, Airlie McCoy & Gabor Bunkozci
- Andrea Thorn, Tim Gruene & George Sheldrick (**SHELX**)
- **Refmac**: Garib Mushudov, LMB-MRC Cambridge
- Developers of **Rosetta** and **Quark**
- **Molrep**: Alexei Vagin & Andrey Lebedev
- Thanks to authors of all underlying programs (**Maxcluster**, **Theseus**, **SPICKER**, **SCWRL**)

