AMPLE
***A**b Initio* **M**odelling of **P**roteins for Mo**le**cular Replacement

Jens Thomas

Institute of Integrative Biology

December 2015

UNIVERSITY OF LIVERPOOL

Outline

- What is AMPLE?
- How it works
- When it works
- Why it works
- How to use it
- What we're doing now

AMPLE

***A**b initio* **M**odelling of **P**roteins for mo**LE**cular replacement

- Joint development by the University of Liverpool and CCP4

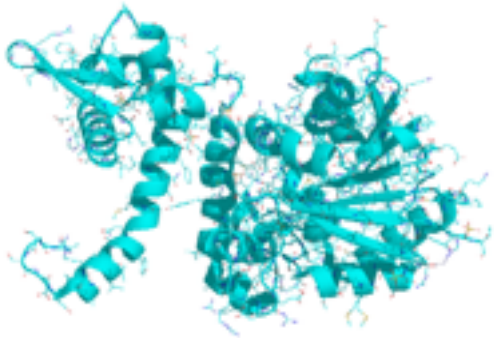
UNIVERSITY OF LIVERPOOL

- AMPLE is a comprehensive project to assess the suitability of unconventional models in molecular replacement

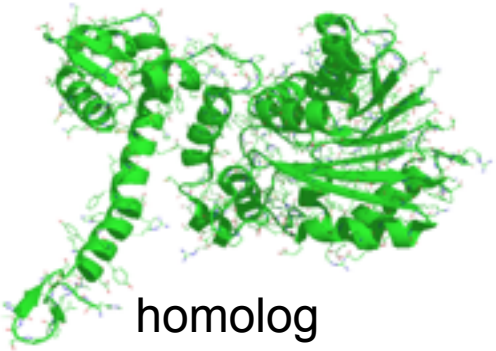
- An additional goal of the project is to make AMPLE into an automated software tool that can be made generally available to potential users through CCP4

Molecular Replacement

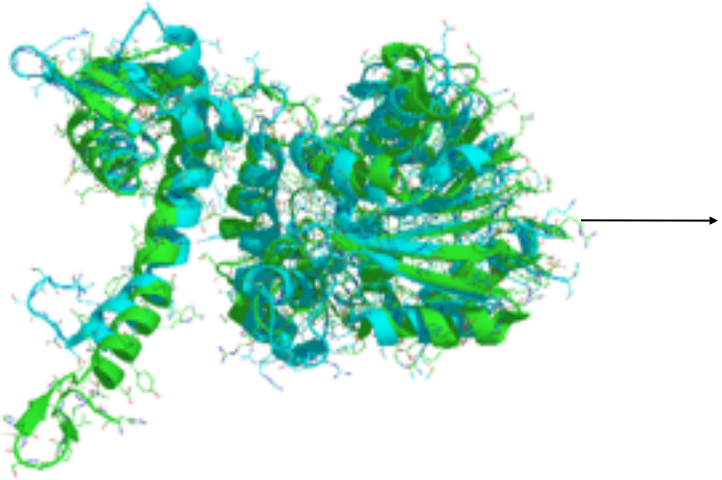
Sequence search for a suitable homolog

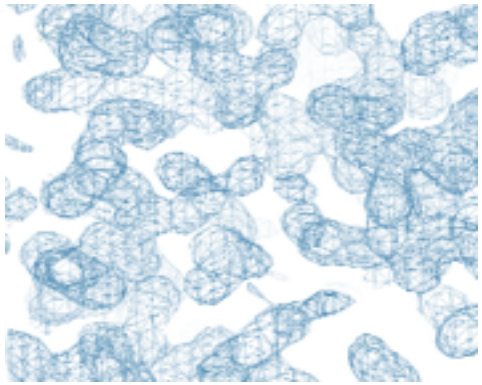


target



homolog





electron density map

- Phases from homolog
- Intensities/positions from target

UNIVERSITY OF LIVERPOOL

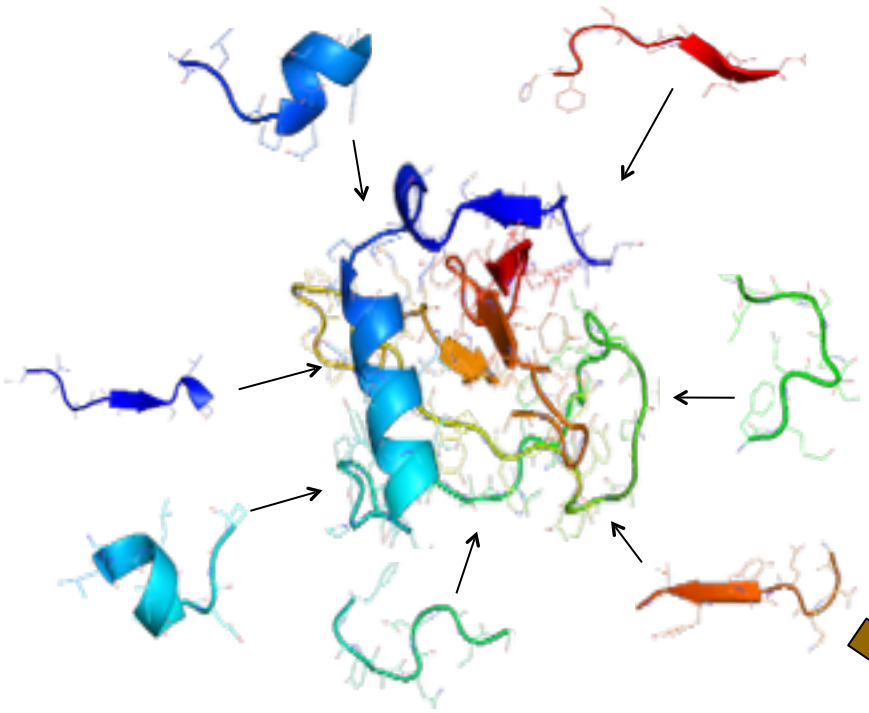
Replacing Molecular Replacement

- Globular proteins with new folds or distant homologies
- Coiled-coils (small changes in sequence or crystallisation conditions can lead to different polymorphs crystallising)
- Helical membrane proteins
- NMR structures as search models
- Not sure exactly what is in the crystal

What to do?

- **AMPLE**
- **ARCIMBOLDO**

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

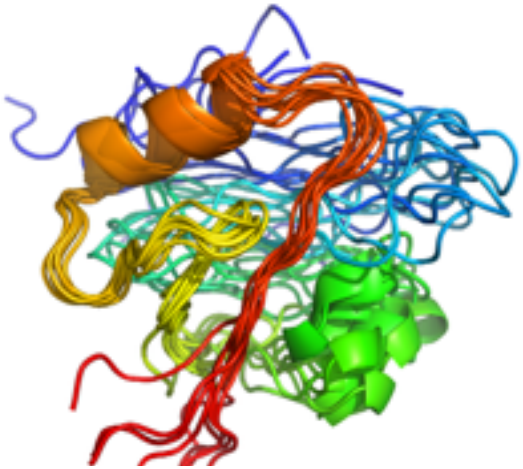
This is beyond the capabilities of most crystallography groups.

But:

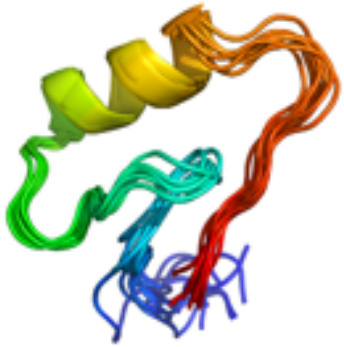
- creating 1000 decoys ~ 1/2 day on a single CPU with ROSETTA
- QUARK modelling is available online and takes ~ 1 day

Ab Initio Models and Molecular Replacement

Simple *ab initio* modelling	Molecular Replacement
Clusters of similar structures	Works well with superposed **ensembles** approximating the target
Within/between clusters, similarity indicates accuracy ∴ trim inaccurate regions leaving more reliable **core**	May only require a **partial** model

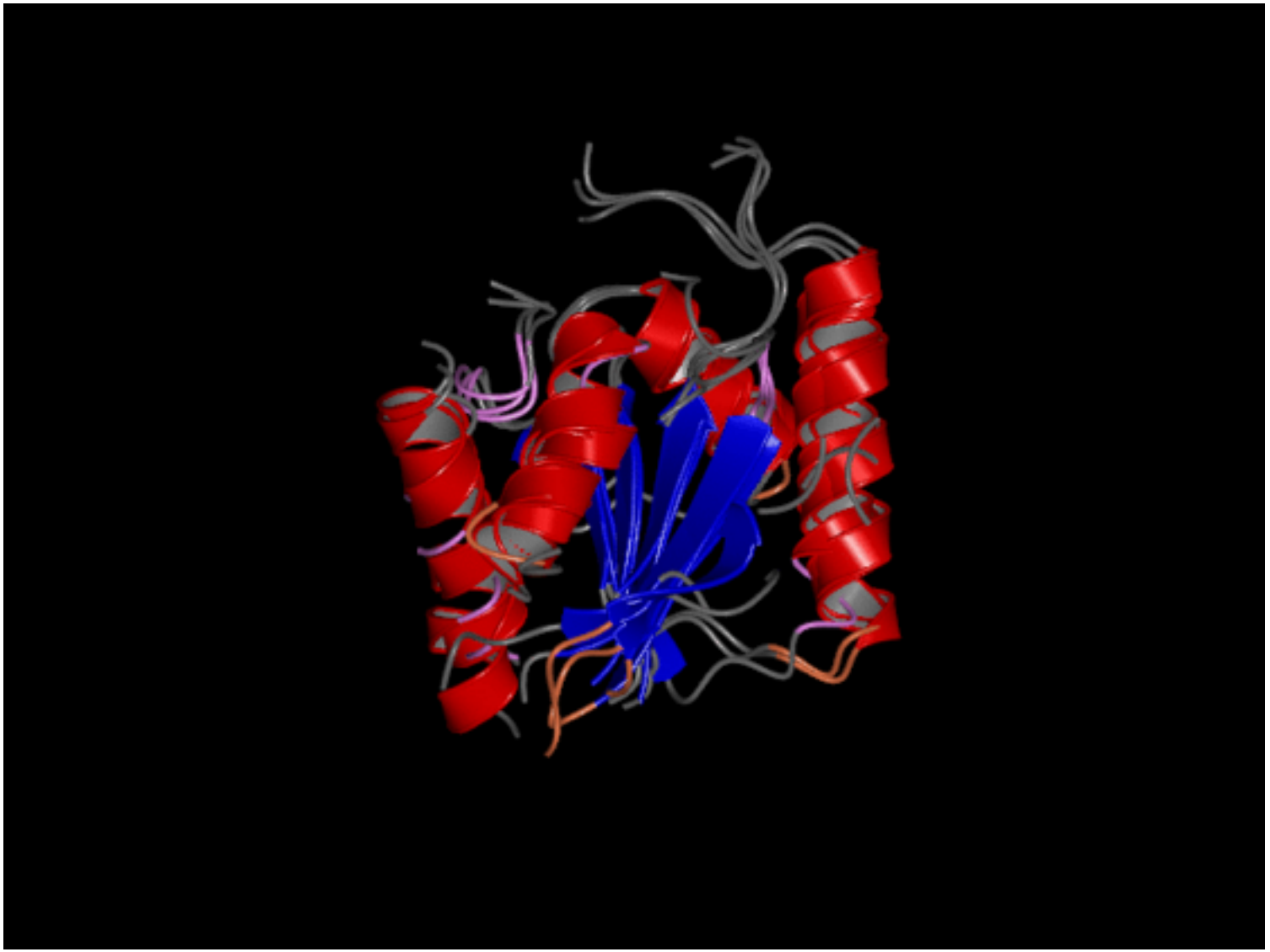


trim to accurate core

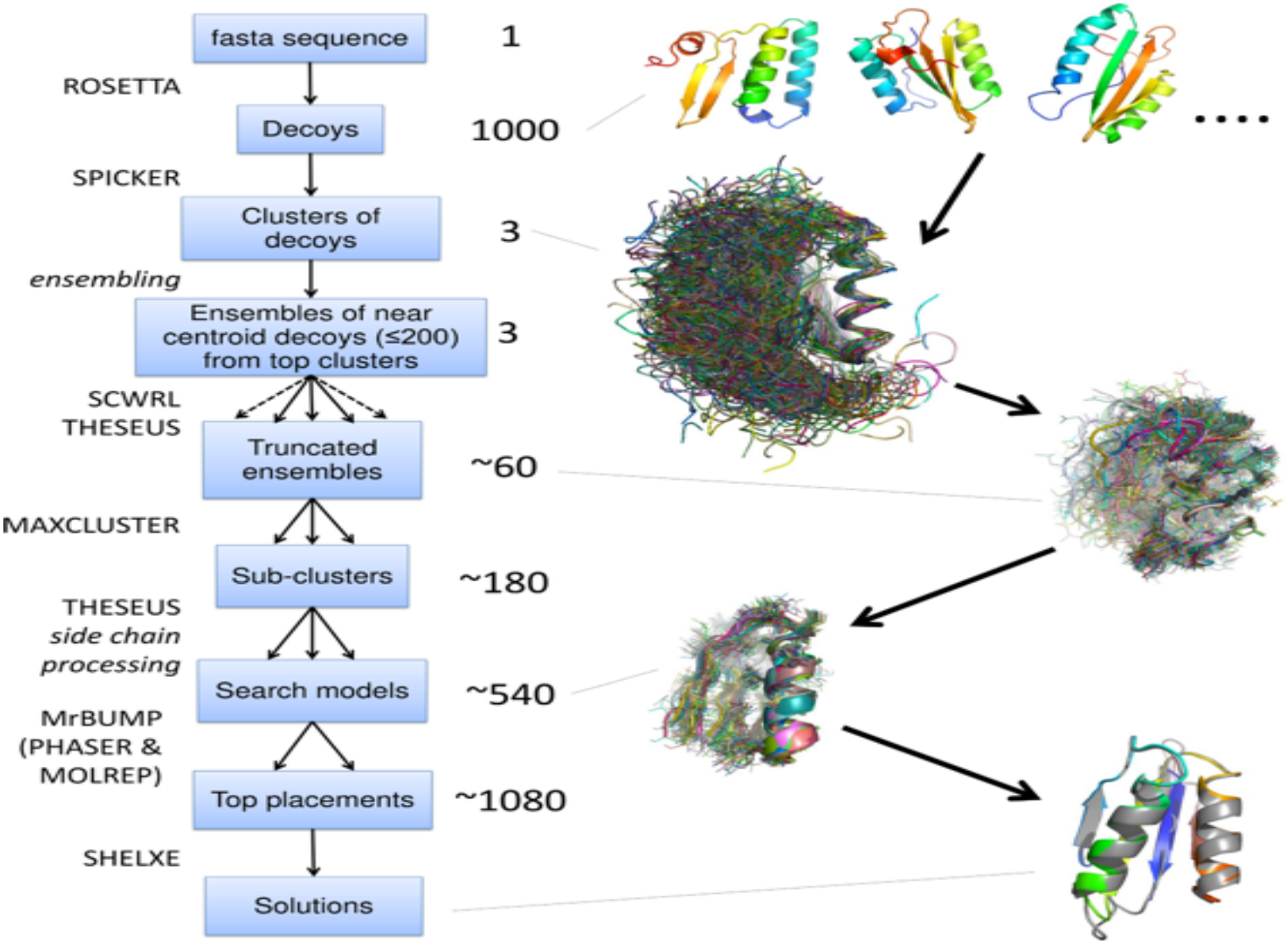


UNIVERSITY OF LIVERPOOL

Truncation in action



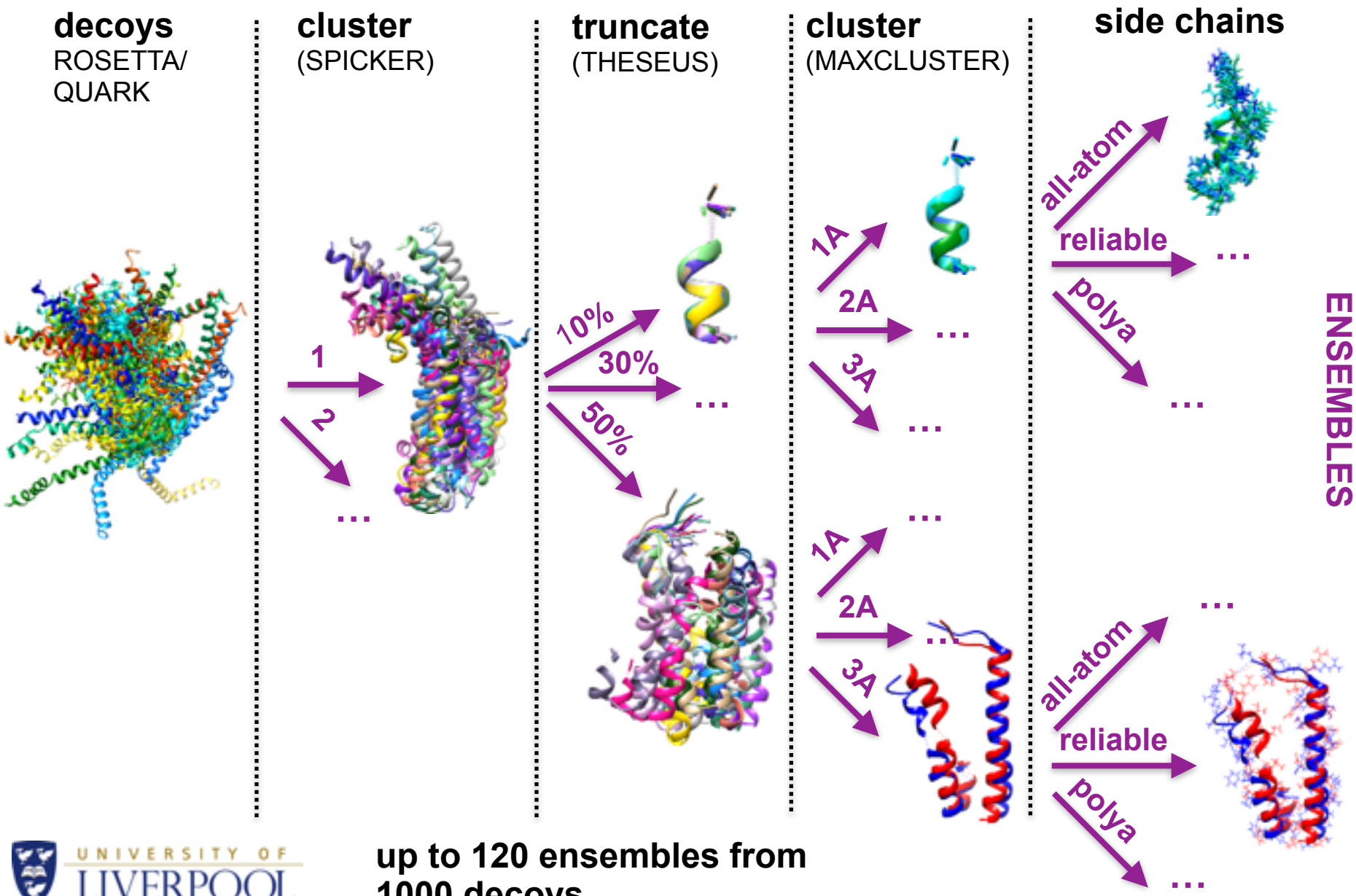
The AMPLE pipeline



UNIVERSITY OF LIVERPOOL

Bibby, J., Keegan, R.M., Mayans, O., Winn, M.D., Rigden, D.J., 2012. Acta Crystallographica Section D Biological Crystallography 68, 1622–1631.

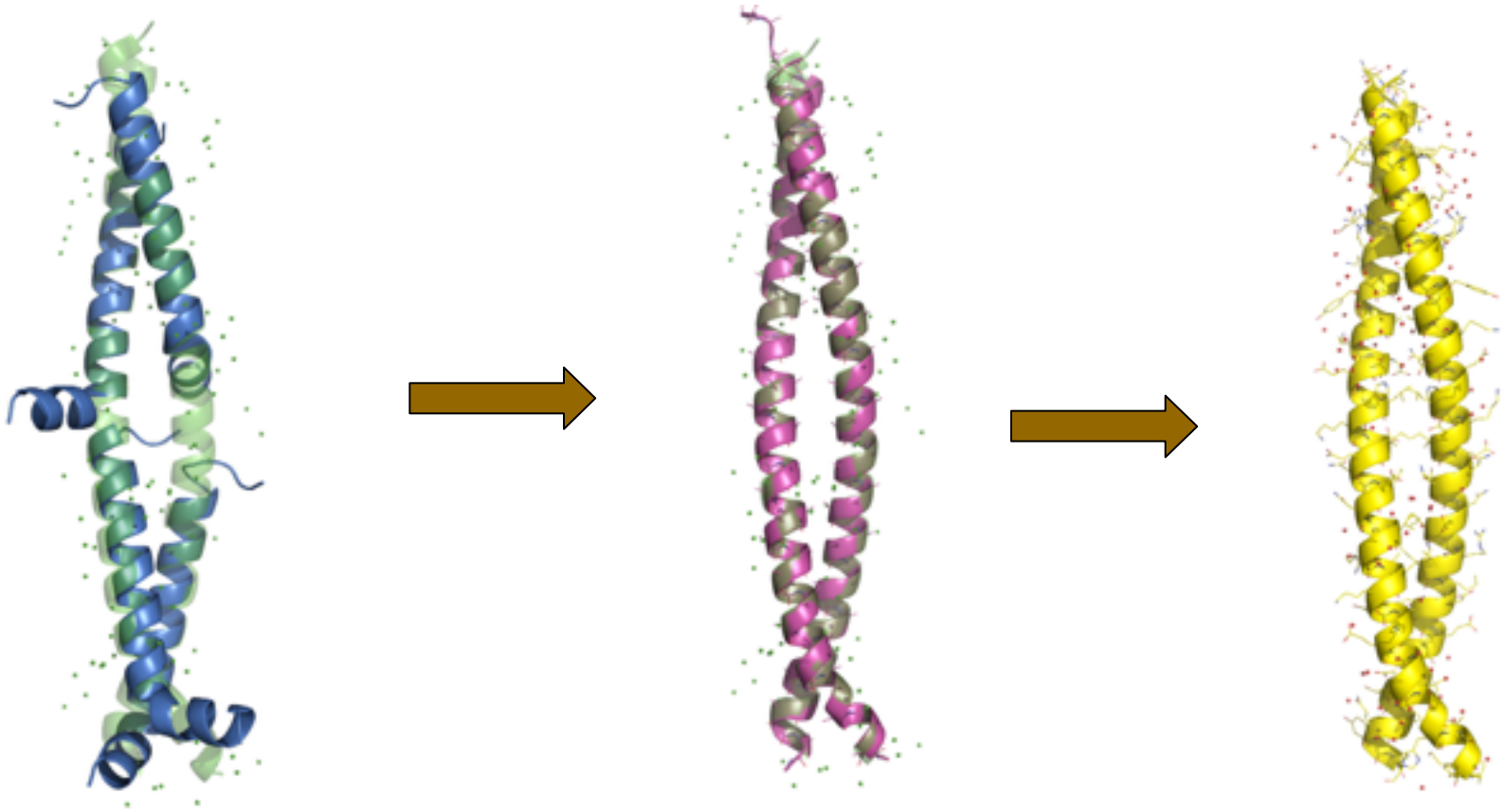
Pathways to ensembles



UNIVERSITY OF LIVERPOOL

up to 120 ensembles from 1000 decoys

MRBUMP MR pipeline



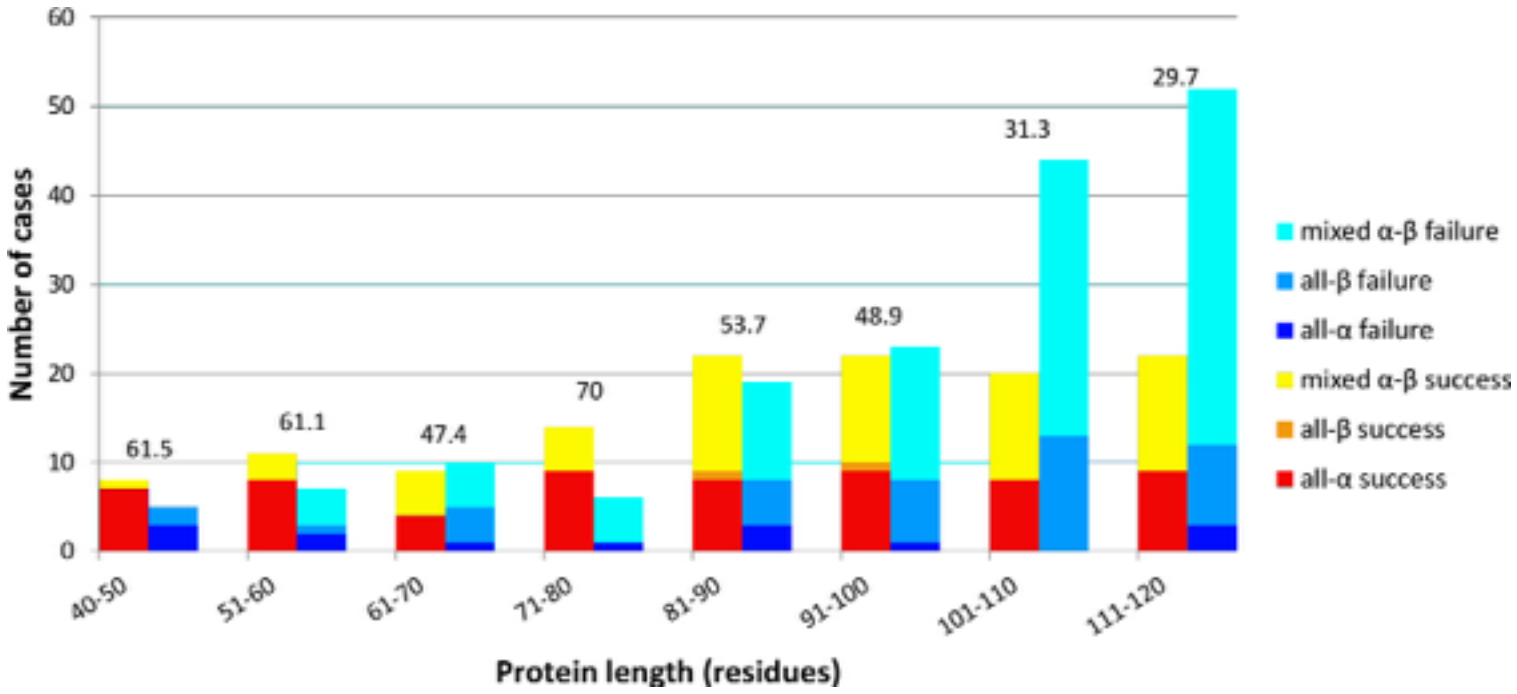
PHASER: places search model

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

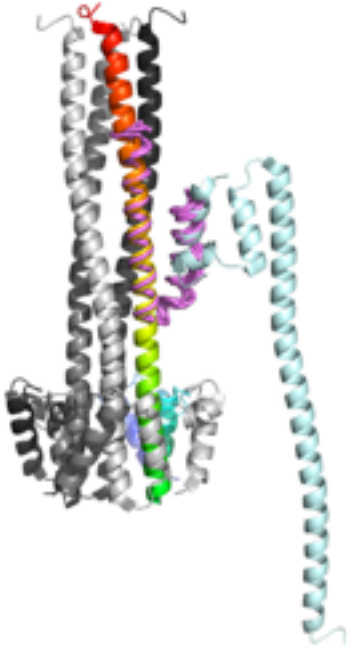


Initial success rate: all-α: 80%, all-β:2%, mixed α-β: 37%

UNIVERSITY OF LIVERPOOL

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 < residues < 253
  - 0.9 < resolution Å < 2.91



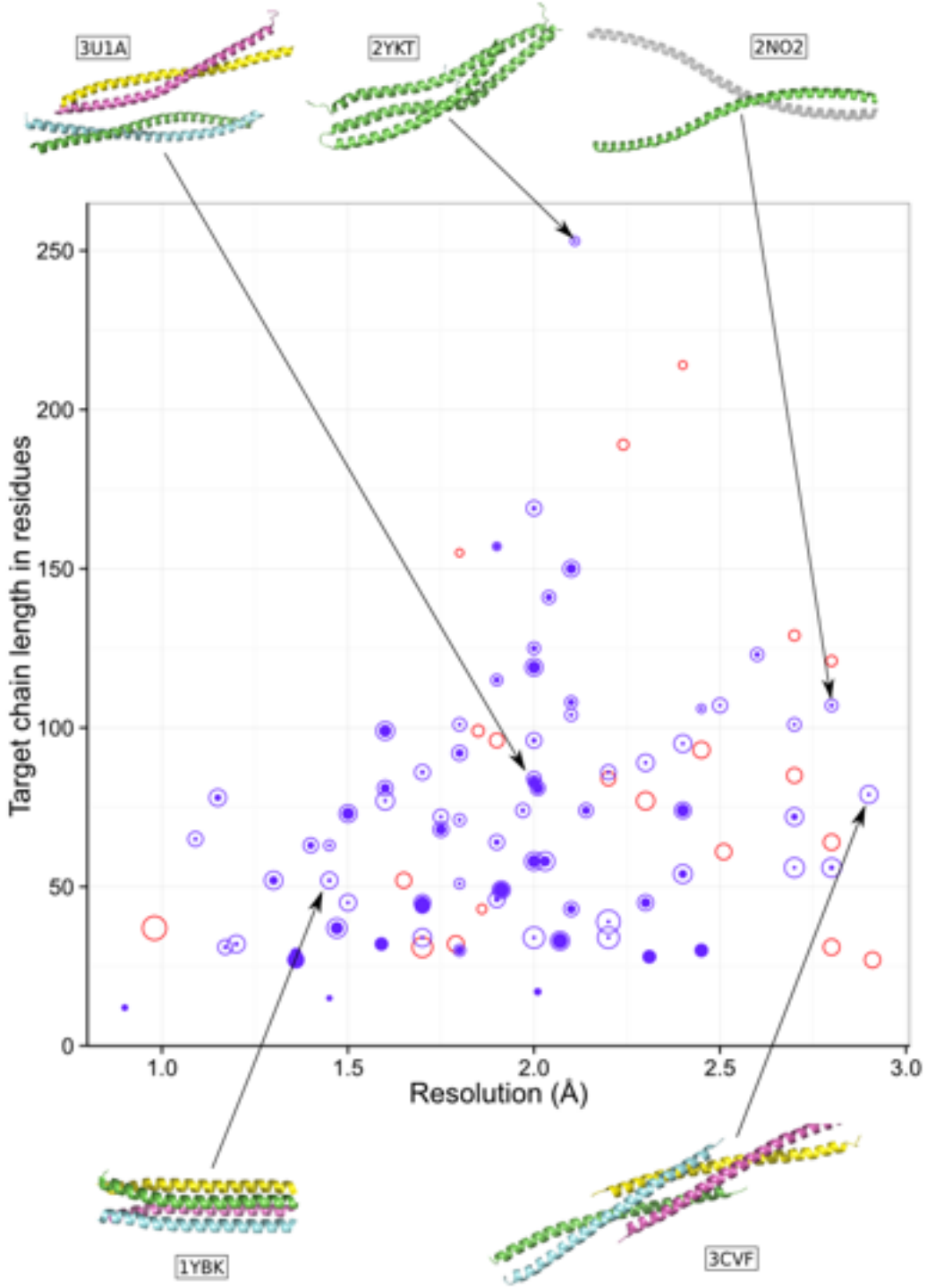
UNIVERSITY OF LIVERPOOL

Coiled-coils

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

Successes included:

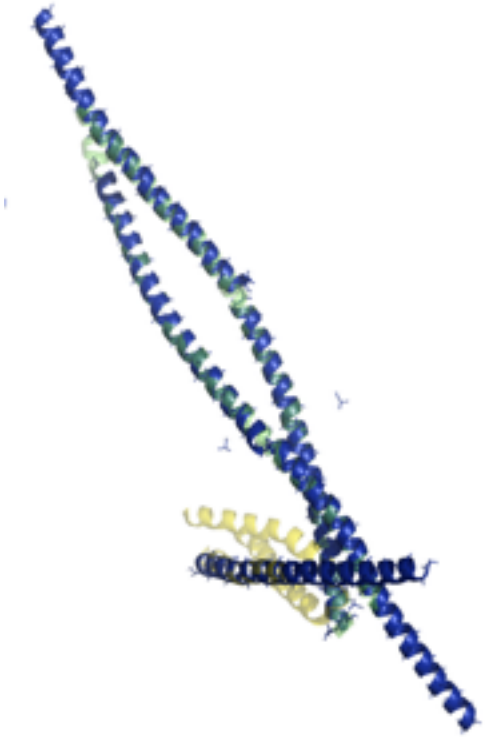
- **3U1A:** 334 residues
- **3CVF:** resolution of 2.8A
- **2NO2:** a domain of Huntingtin-interacting protein 1, that contains a long, unconventional coiled-coil-like assembly originally phased experimentally using MAD
- **1YBK** right-handed coiled coil

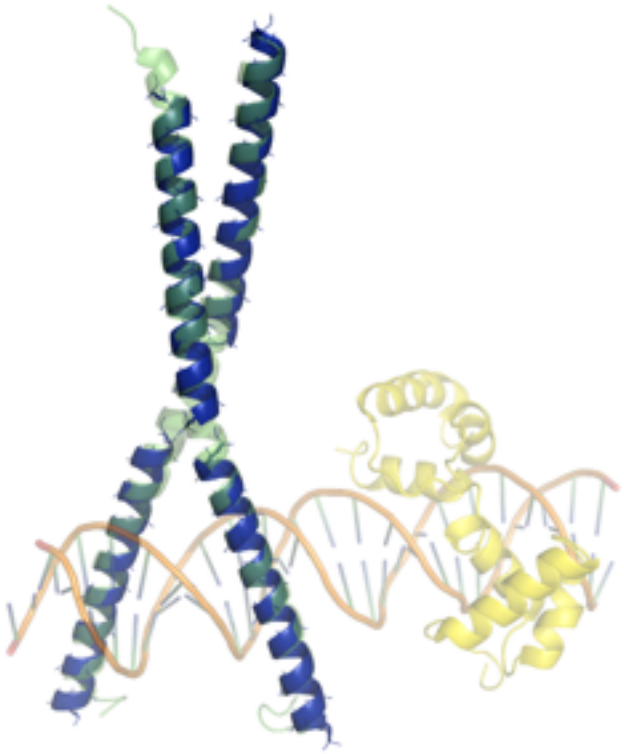


UNIVERSITY OF LIVERPOOL

Complex Ambitions

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



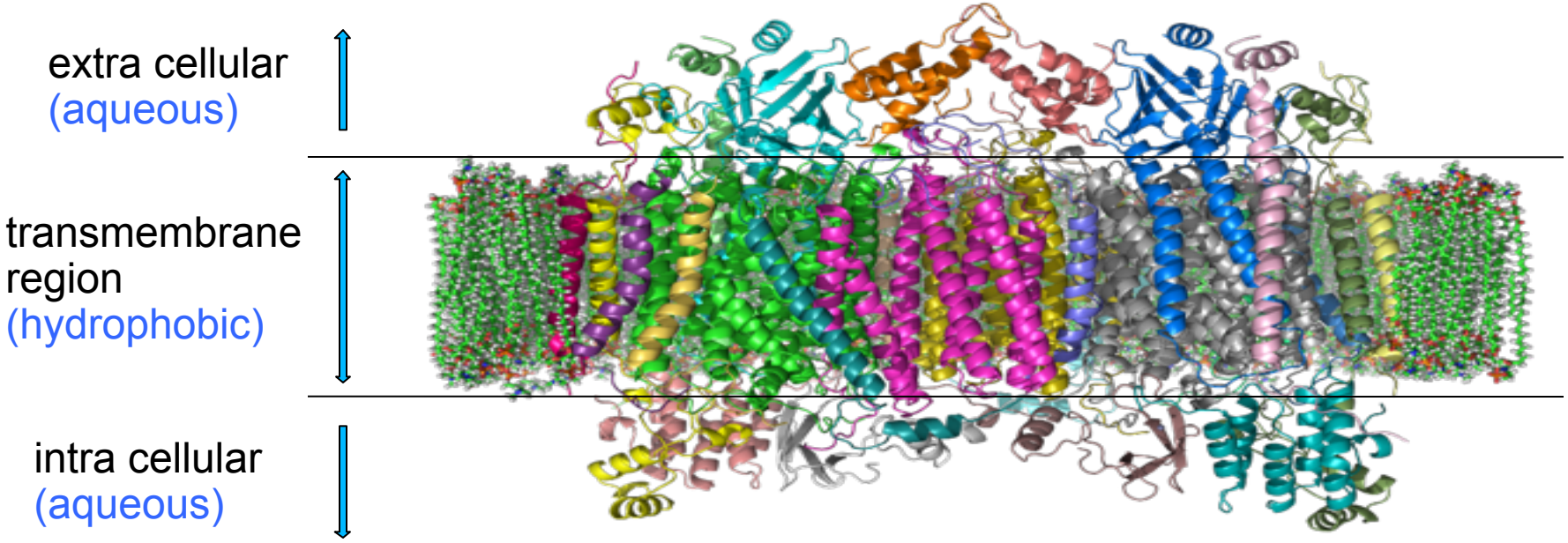


- **1X79**: three chains, 322 residues, ~70% coiled-coil @ 2.41Å
- **1H8A**: MyB/DNA complex of 278 residues ~ 60% coiled-coil @ 2.23Å.

UNIVERSITY OF LIVERPOOL

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

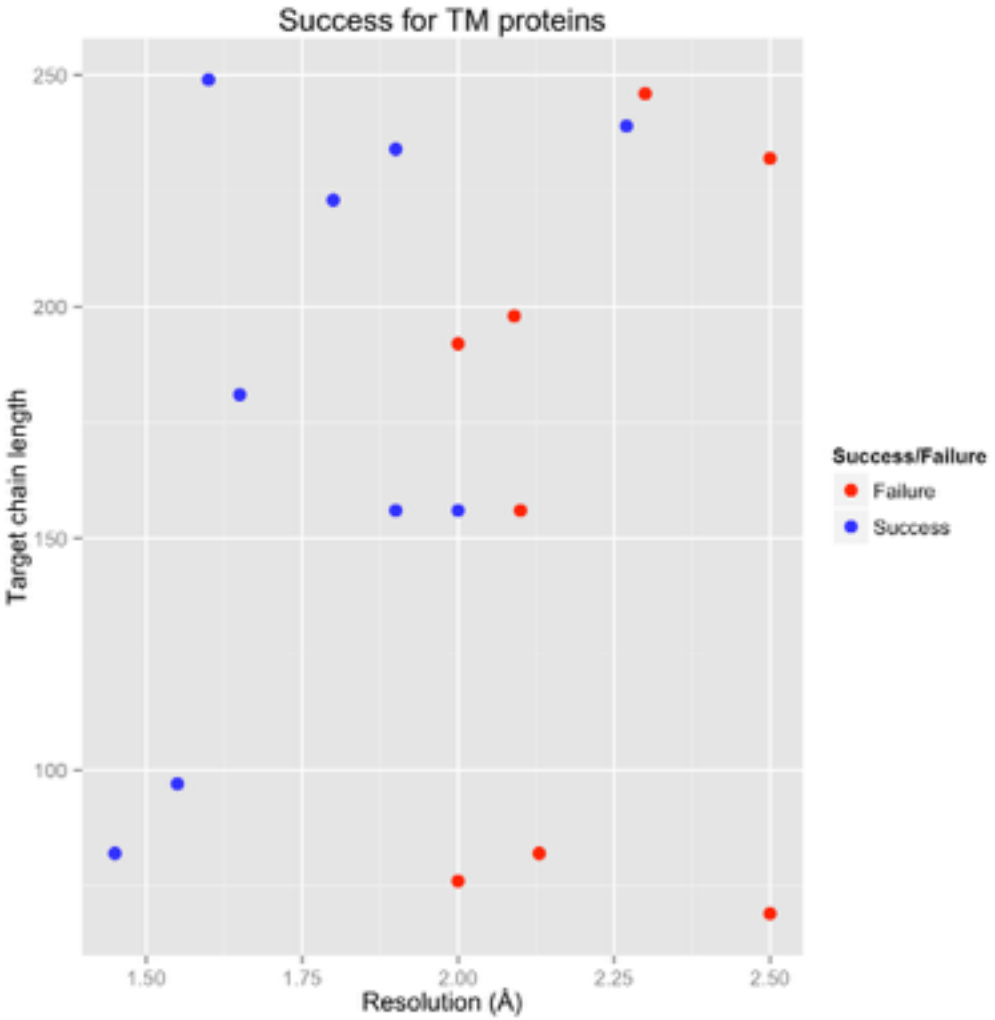


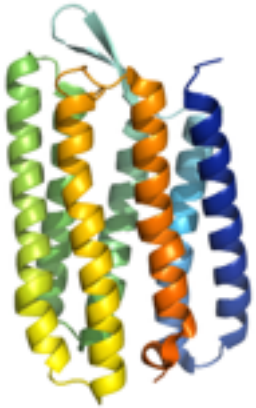
UNIVERSITY OF LIVERPOOL

Image: http://en.wikiversity.org/wiki/File:Cytochrome_C_Oxidase_1OCC_in_Membrane_2.png

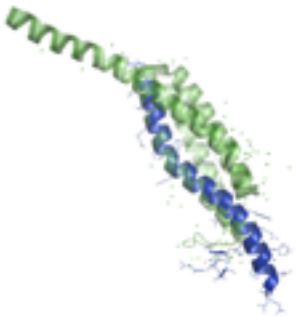
Transmembrane Results

- Selected 18 alpha-helical transmembrane proteins:
  - 23 – 249 residues
  - 1.45 – 2.5A resolution
- 9 clear successes
- Successes extended to 249 residues (3HAP).





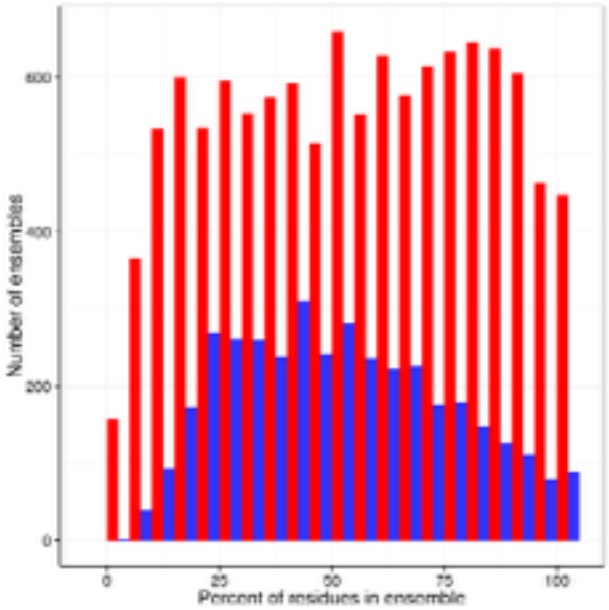
3HAP

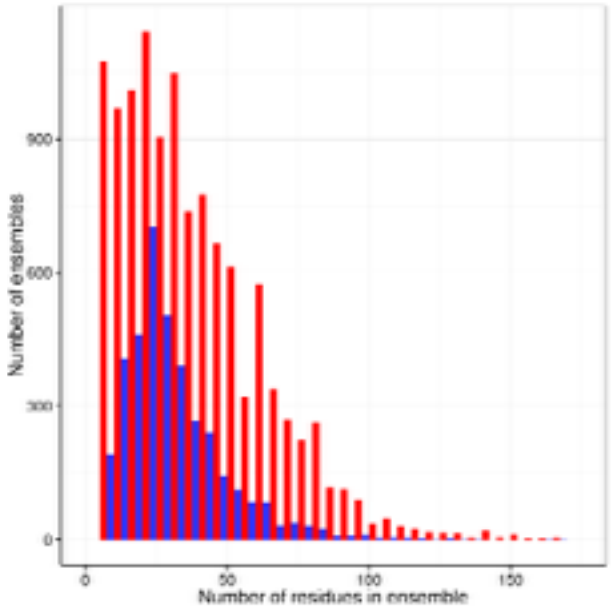


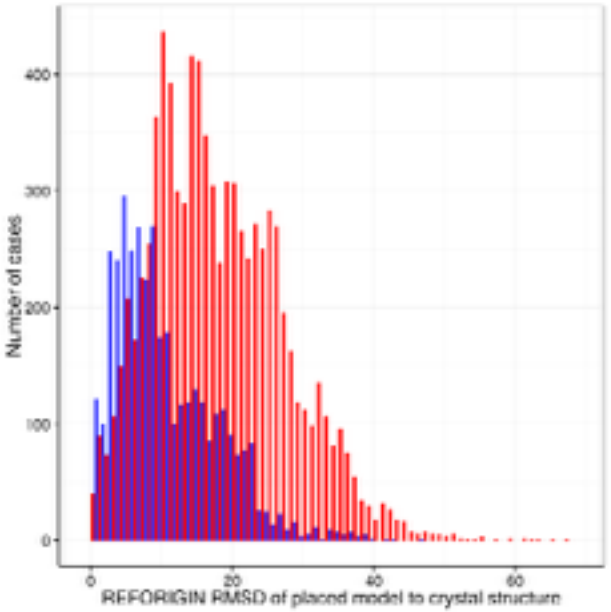
3PCV

UNIVERSITY OF LIVERPOOL

What is solving these structures?

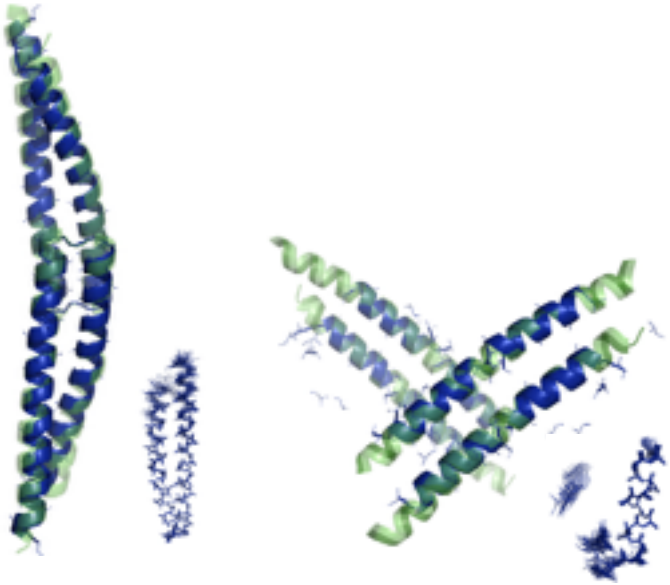


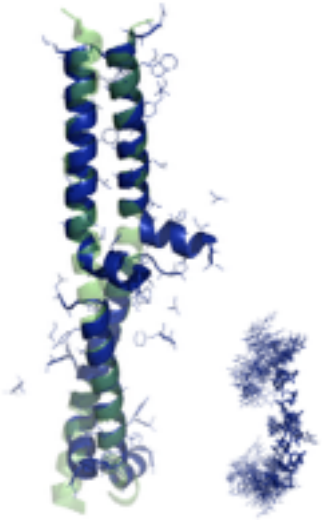


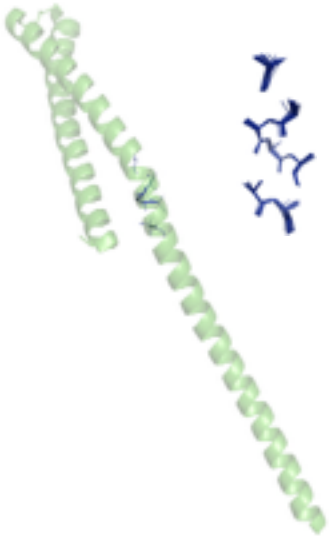


- Sample all proportions of the model but usually less than half
- Short (~25-30 residues) search models
- Search model seems to be in the wrong place!

What is solving these structures?







2IC9: 2 copies
RIO=125
RIO_in=125

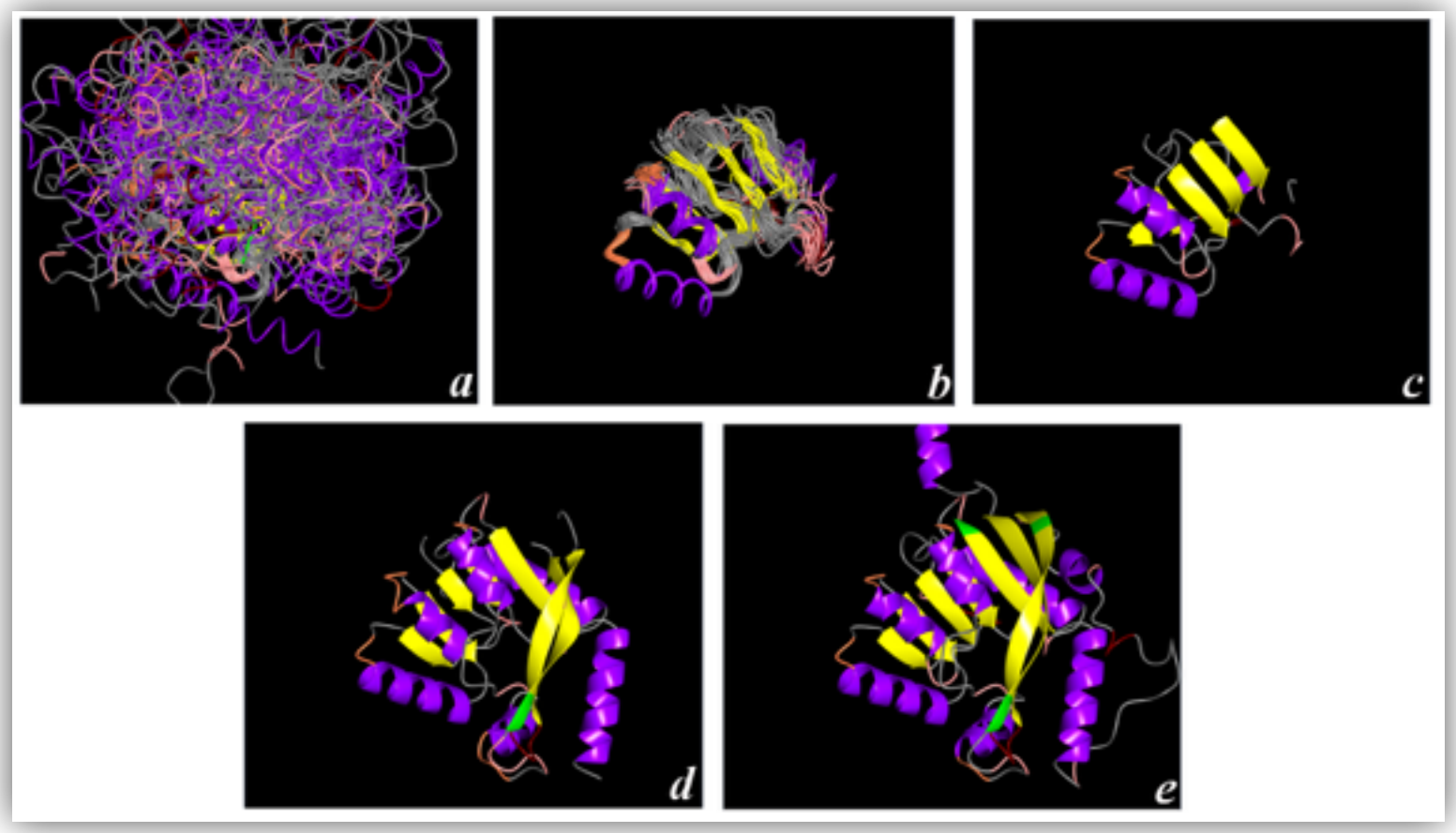
3H00: 8 copies
RIO=110
RIO_out=110

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

3OKQ: 1 copy
RIO=4
RIO_out=4

A mixture of in-register and out-of-register placements

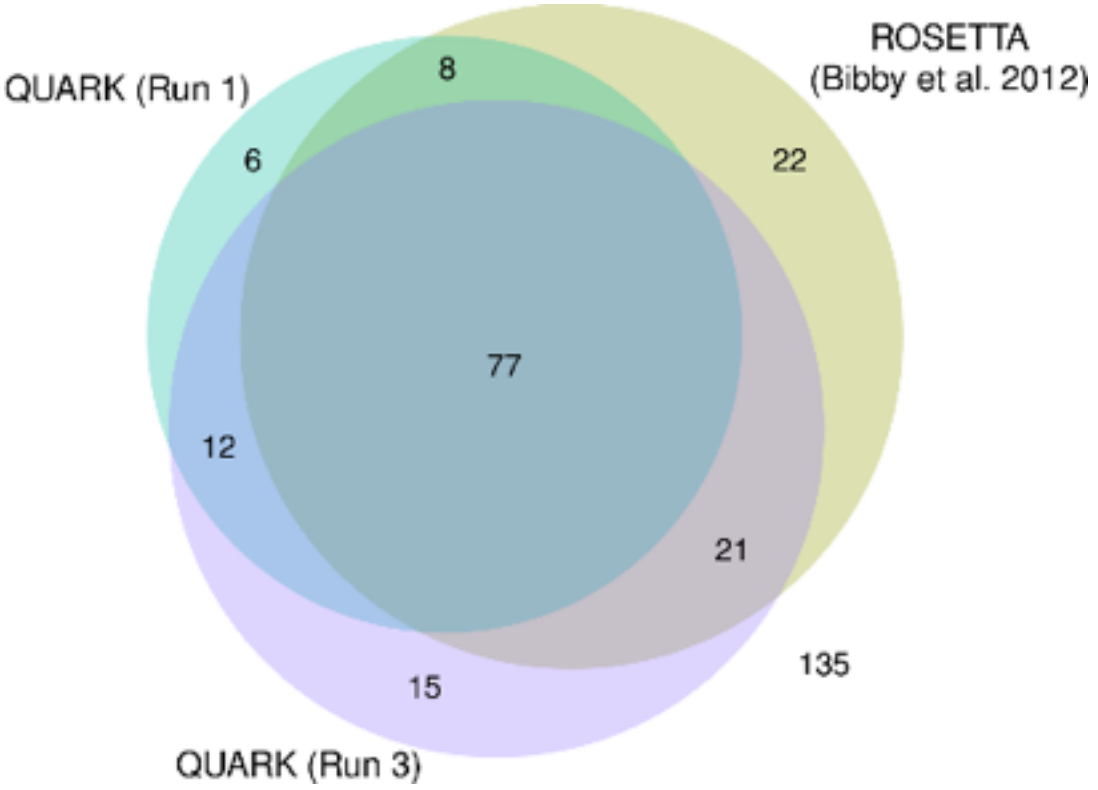
Exploiting distant homologues



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

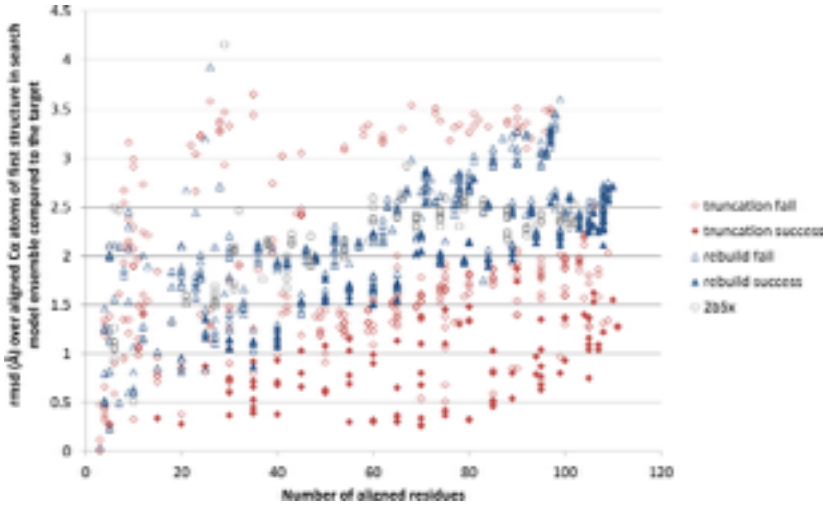


Remodelling related NMR structures or distant homologues

- **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)

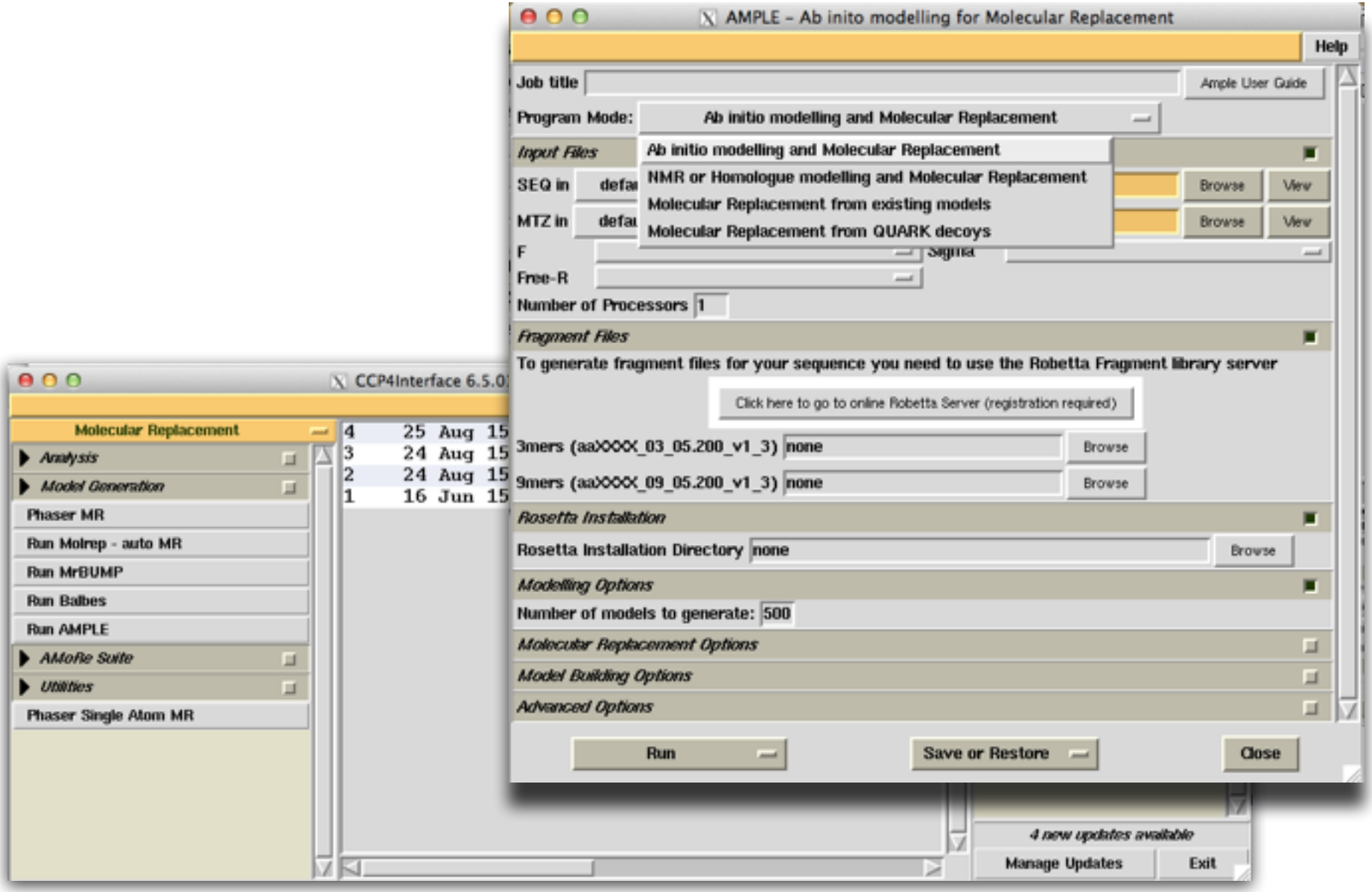
- 19 by truncation alone
- 3 with remodelling



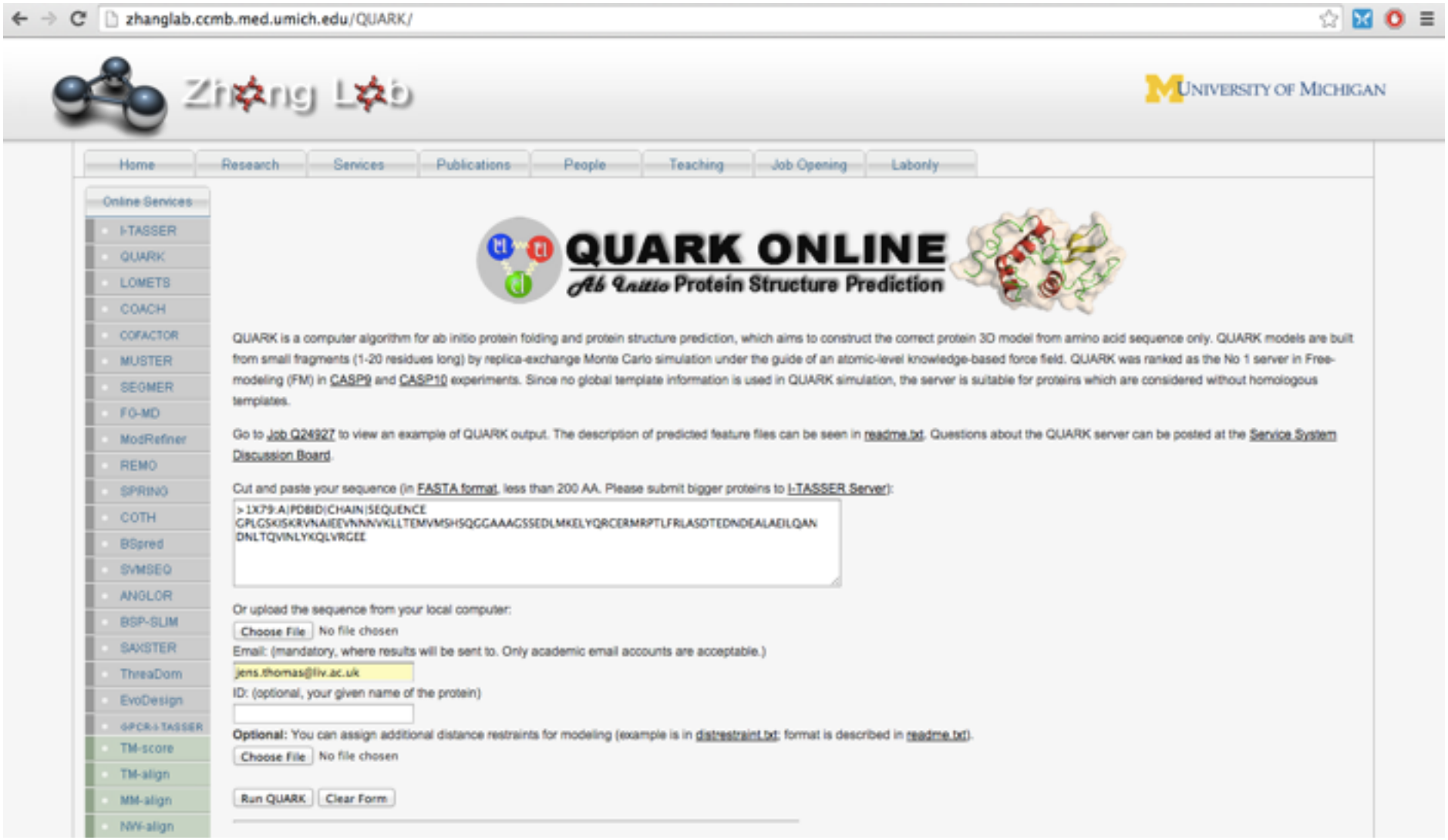
Practicalities

- Available since CCP4 6.3.0.
- Requires installation of several non-CCP4 packages (all freely available to academics):
  - **SHELXE** (now in CCP4 7.0)
  - **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

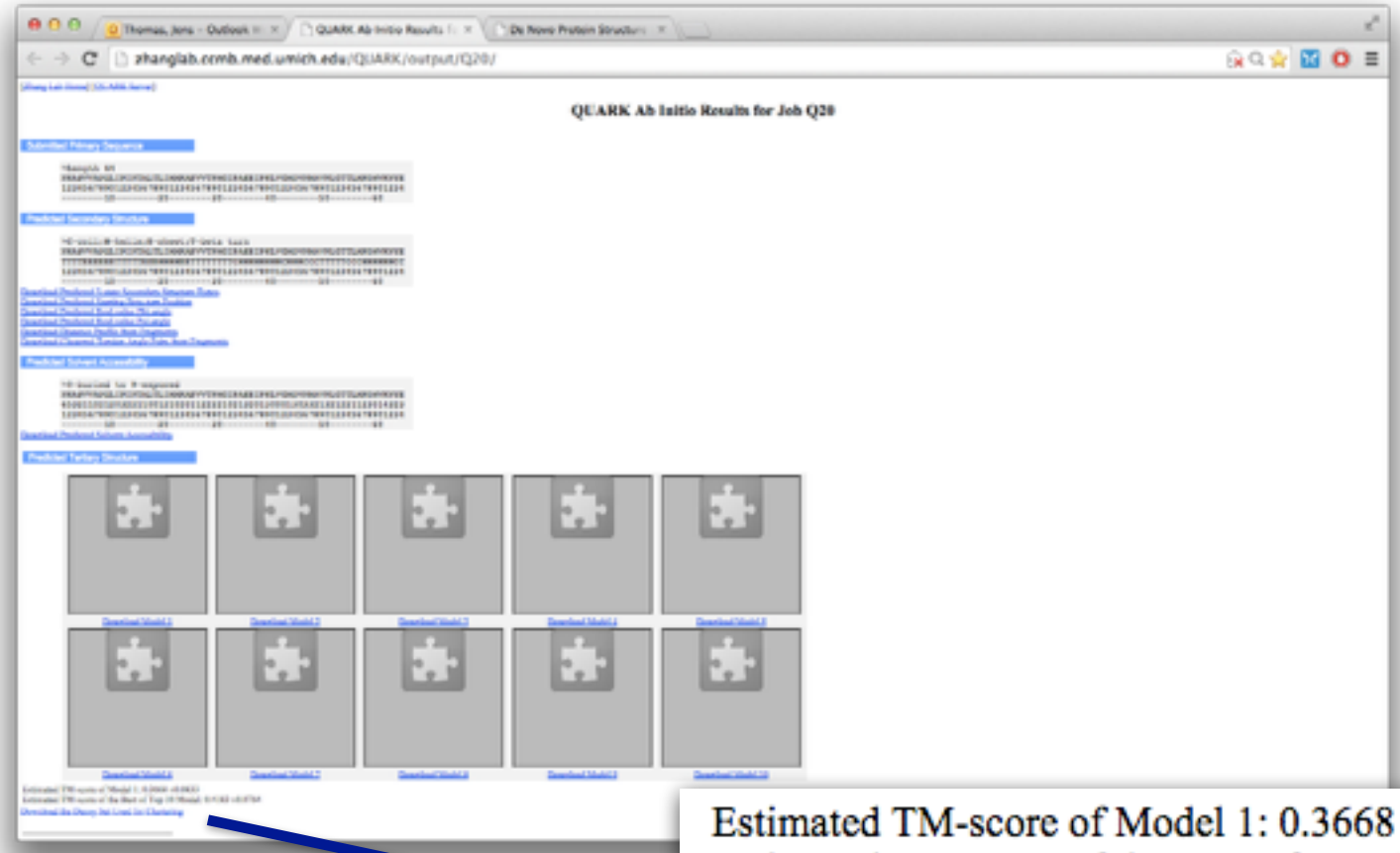


Create models with Quark



University of Liverpool

Quark Decoys

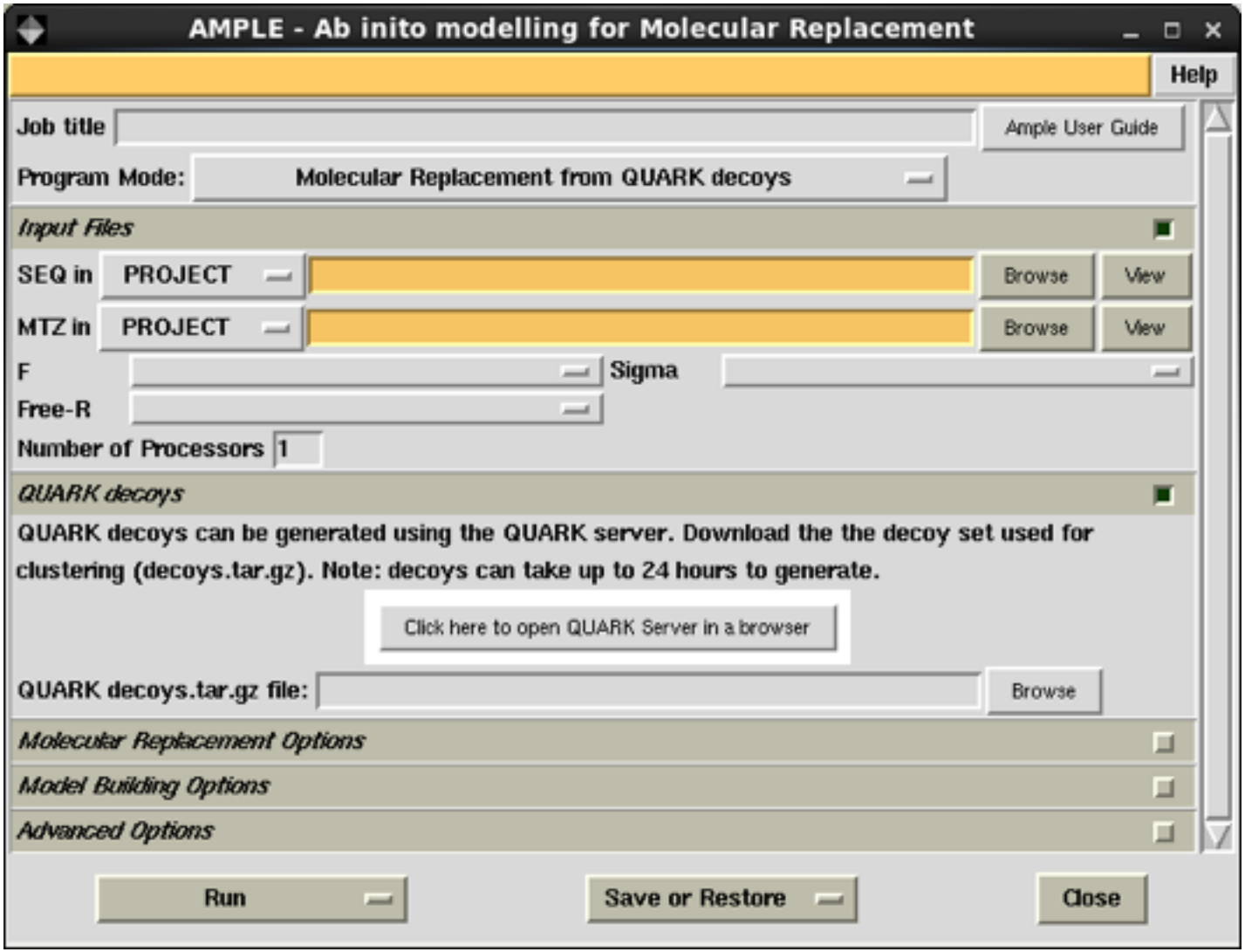


Estimated TM-score of Model 1: 0.3668 ±0.0833
Estimated TM-score of the Best of Top 10 Model: 0.4183 ±0.0764
Download the Decoy Set Used for Clustering

decoys.tar.gz

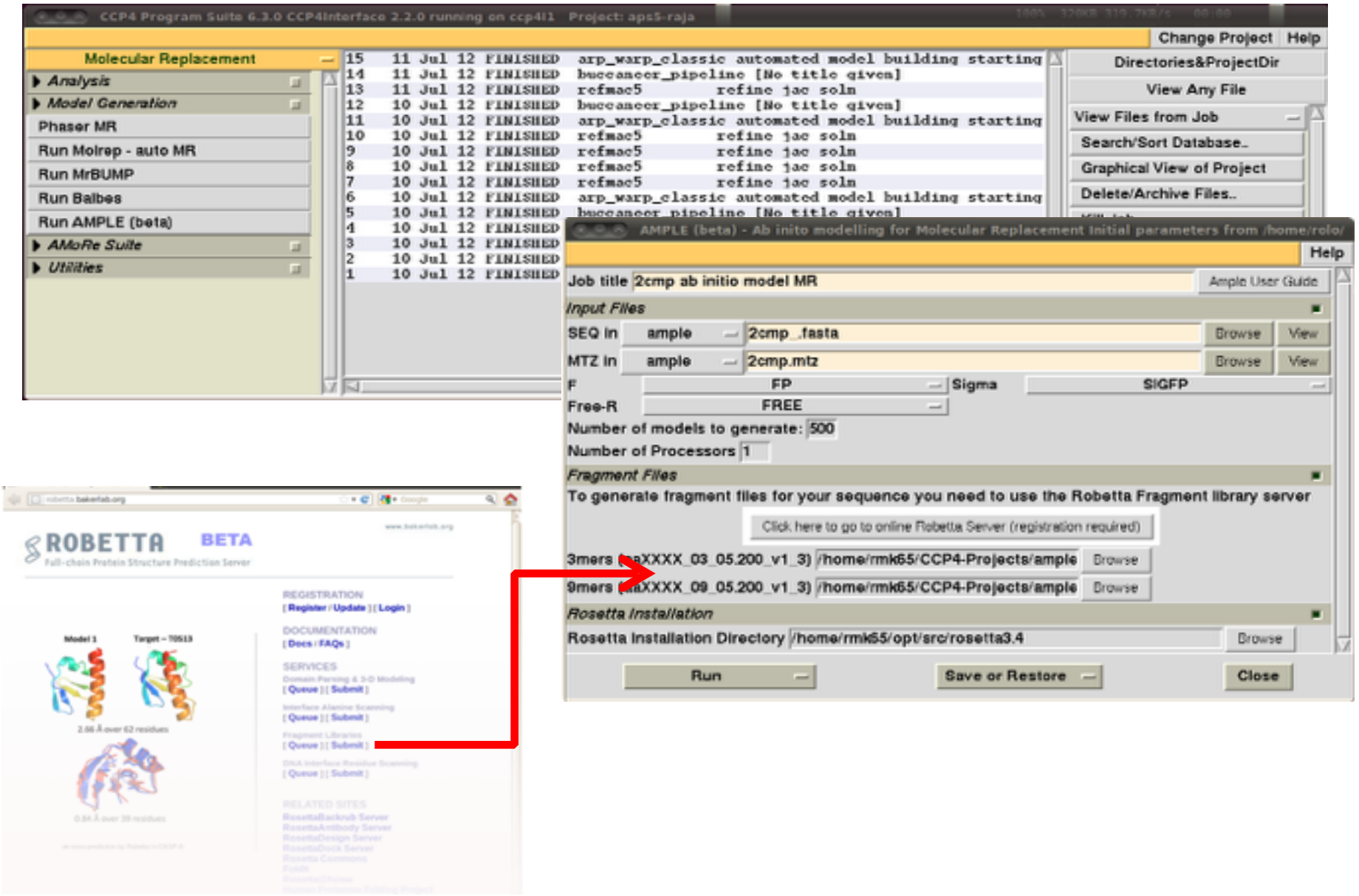
UNIVERSITY OF LIVERPOOL

Quark Decoys

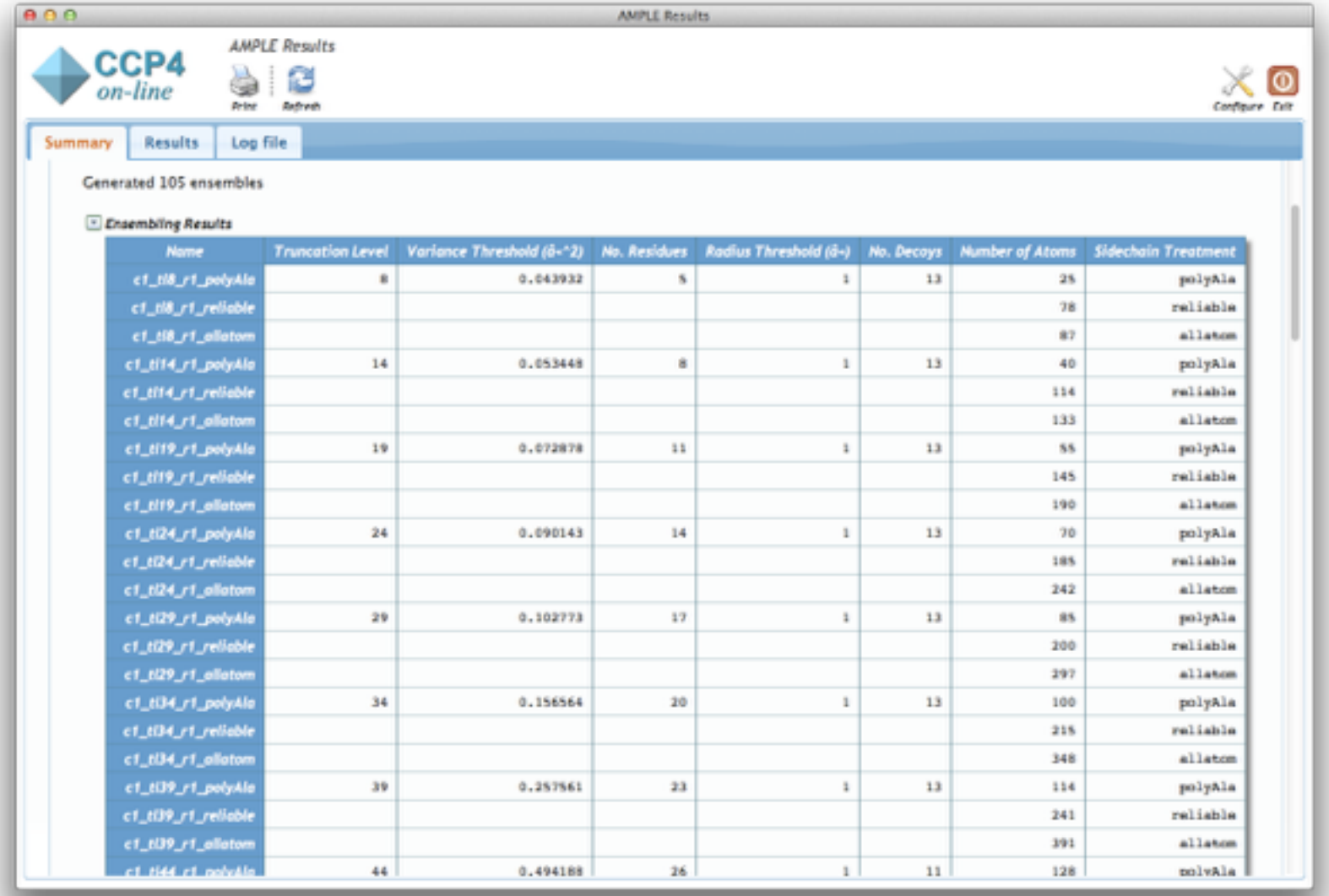


UNIVERSITY OF LIVERPOOL

Running Rosetta



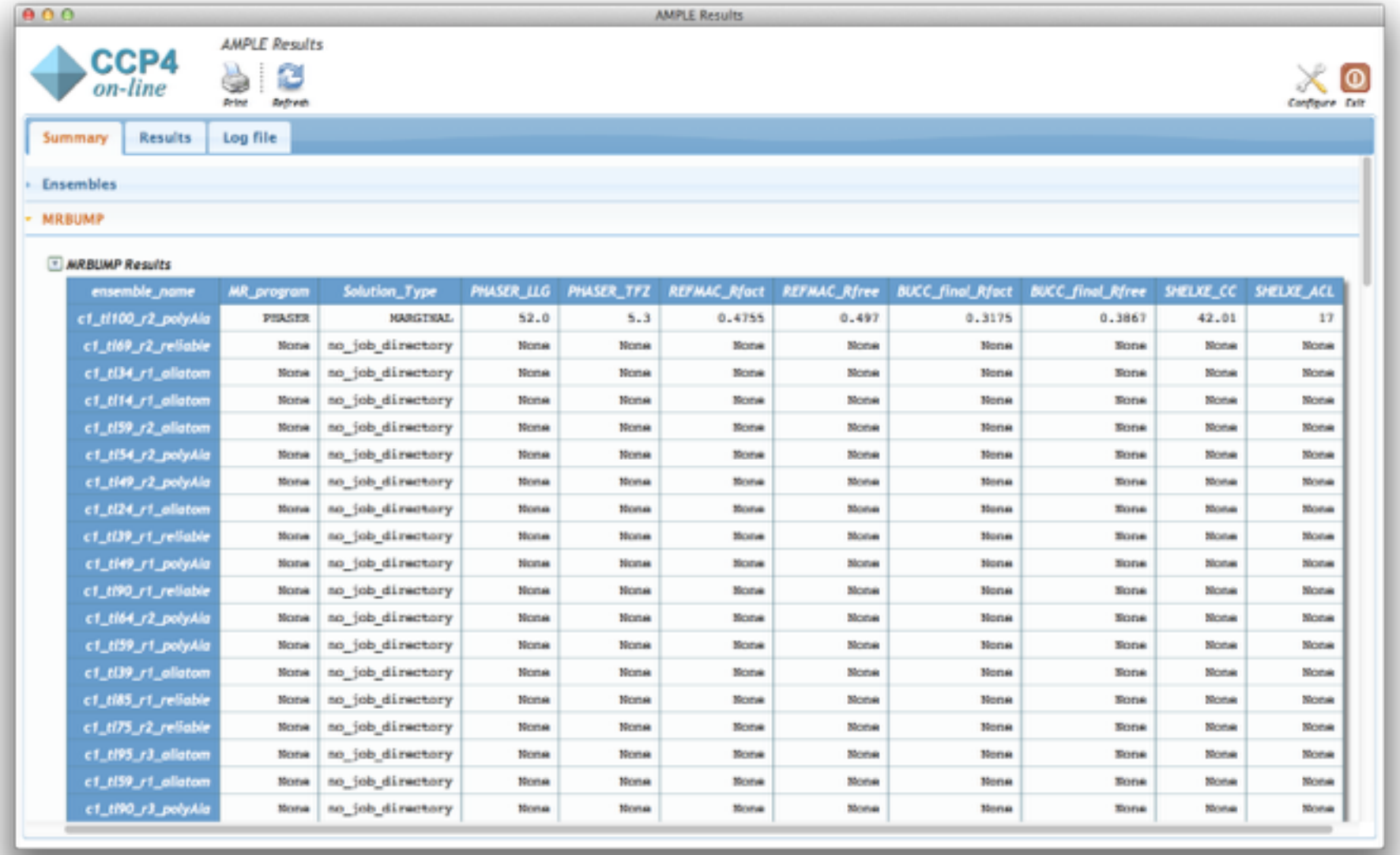
Results: Ensembles



Name	Truncation Level	Variance Threshold (Å^2)	No. Residues	Radius Threshold (Å)	No. Decoys	Number of Atoms	Sidechain Treatment
c1_t18_r1_polyAla	8	0.043932	5	1	13	25	polyAla
c1_t18_r1_reliable						78	reliable
c1_t18_r1_allatom						87	allatom
c1_t14_r1_polyAla	14	0.053448	8	1	13	40	polyAla
c1_t14_r1_reliable						114	reliable
c1_t14_r1_allatom						133	allatom
c1_t19_r1_polyAla	19	0.072878	11	1	13	55	polyAla
c1_t19_r1_reliable						145	reliable
c1_t19_r1_allatom						190	allatom
c1_t24_r1_polyAla	24	0.090143	14	1	13	70	polyAla
c1_t24_r1_reliable						185	reliable
c1_t24_r1_allatom						242	allatom
c1_t29_r1_polyAla	29	0.102773	17	1	13	85	polyAla
c1_t29_r1_reliable						200	reliable
c1_t29_r1_allatom						297	allatom
c1_t34_r1_polyAla	34	0.156564	20	1	13	100	polyAla
c1_t34_r1_reliable						215	reliable
c1_t34_r1_allatom						348	allatom
c1_t39_r1_polyAla	39	0.257561	23	1	13	114	polyAla
c1_t39_r1_reliable						241	reliable
c1_t39_r1_allatom						391	allatom
c1_t44_r1_polyAla	44	0.494188	26	1	11	128	polyAla

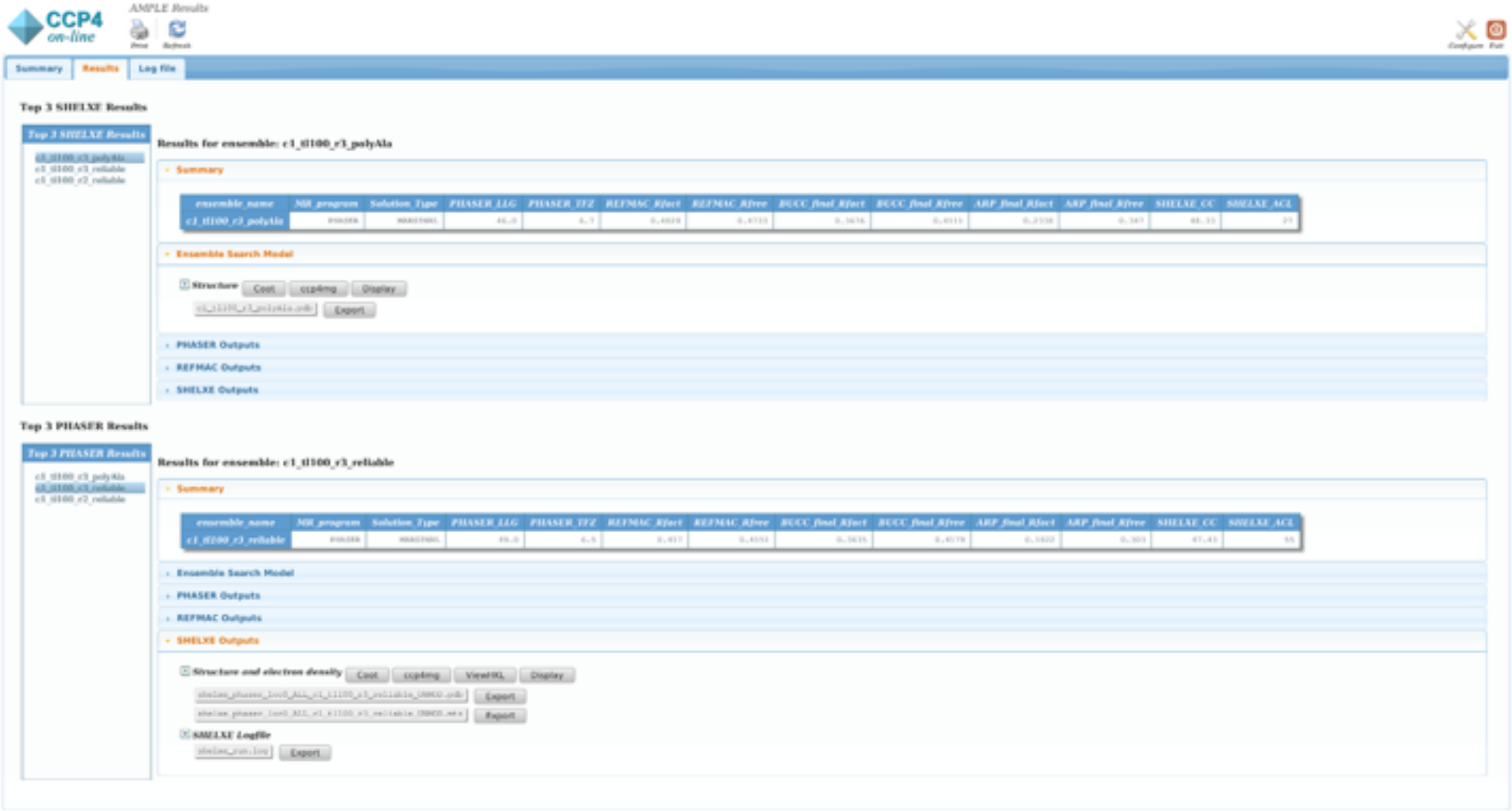
UNIVERSITY OF LIVERPOOL

Results: MrBUMP



ensemble_name	MR_program	Solution_Type	PHASER_LLG	PHASER_TFZ	REFMAC_Rfact	REFMAC_Rfree	BUCC_final_Rfact	BUCC_final_Rfree	SHELXE_CC	SHELXE_ACL
c1_tl100_r2_polyAla	PHASER	MARGINAL	52.0	5.3	0.4755	0.497	0.3175	0.3867	42.01	17
c1_tl69_r2_reliable	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl34_r1_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl14_r1_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl59_r2_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl54_r2_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl49_r2_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl24_r1_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl39_r1_reliable	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl49_r1_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl90_r1_reliable	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl64_r2_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl59_r1_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl39_r1_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl85_r1_reliable	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl75_r2_reliable	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl95_r3_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl59_r1_allatom	None	no_job_directory	None	None	None	None	None	None	None	None
c1_tl90_r3_polyAla	None	no_job_directory	None	None	None	None	None	None	None	None

Results: Files



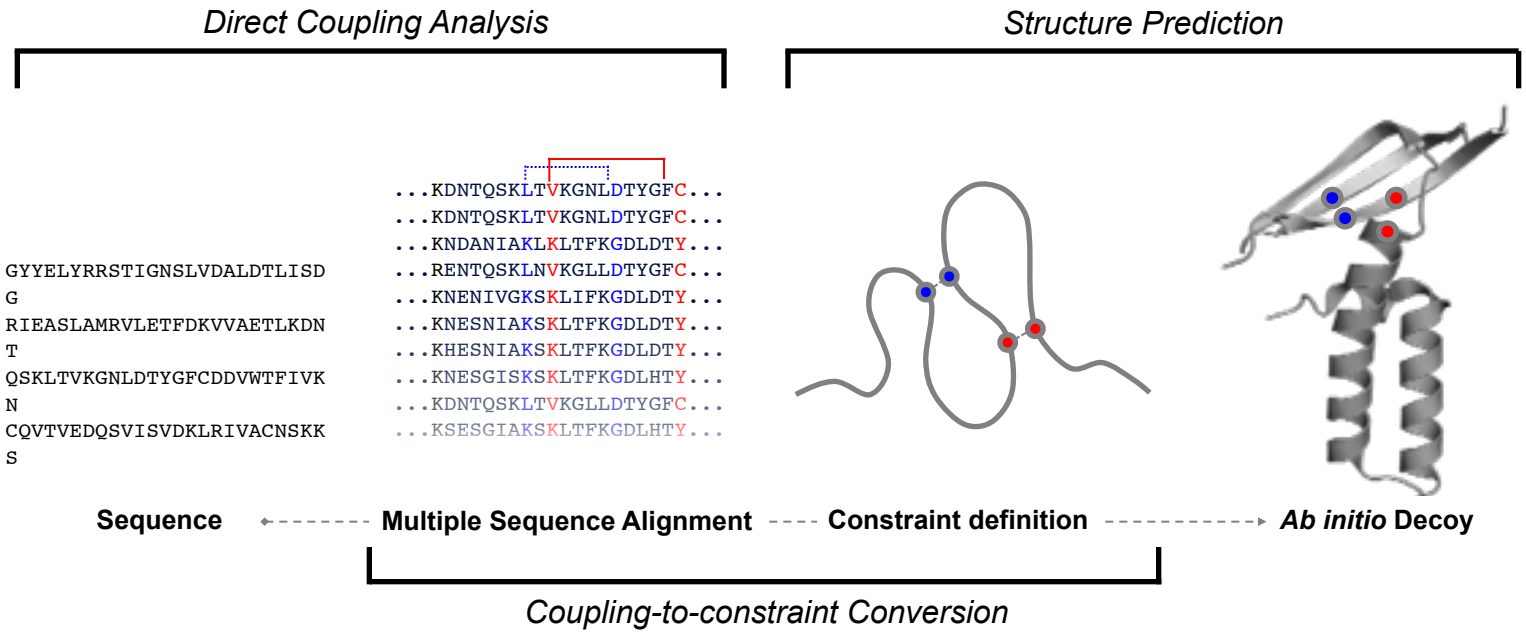
UNIVERSITY OF LIVERPOOL

AMPLE server - coming soon!

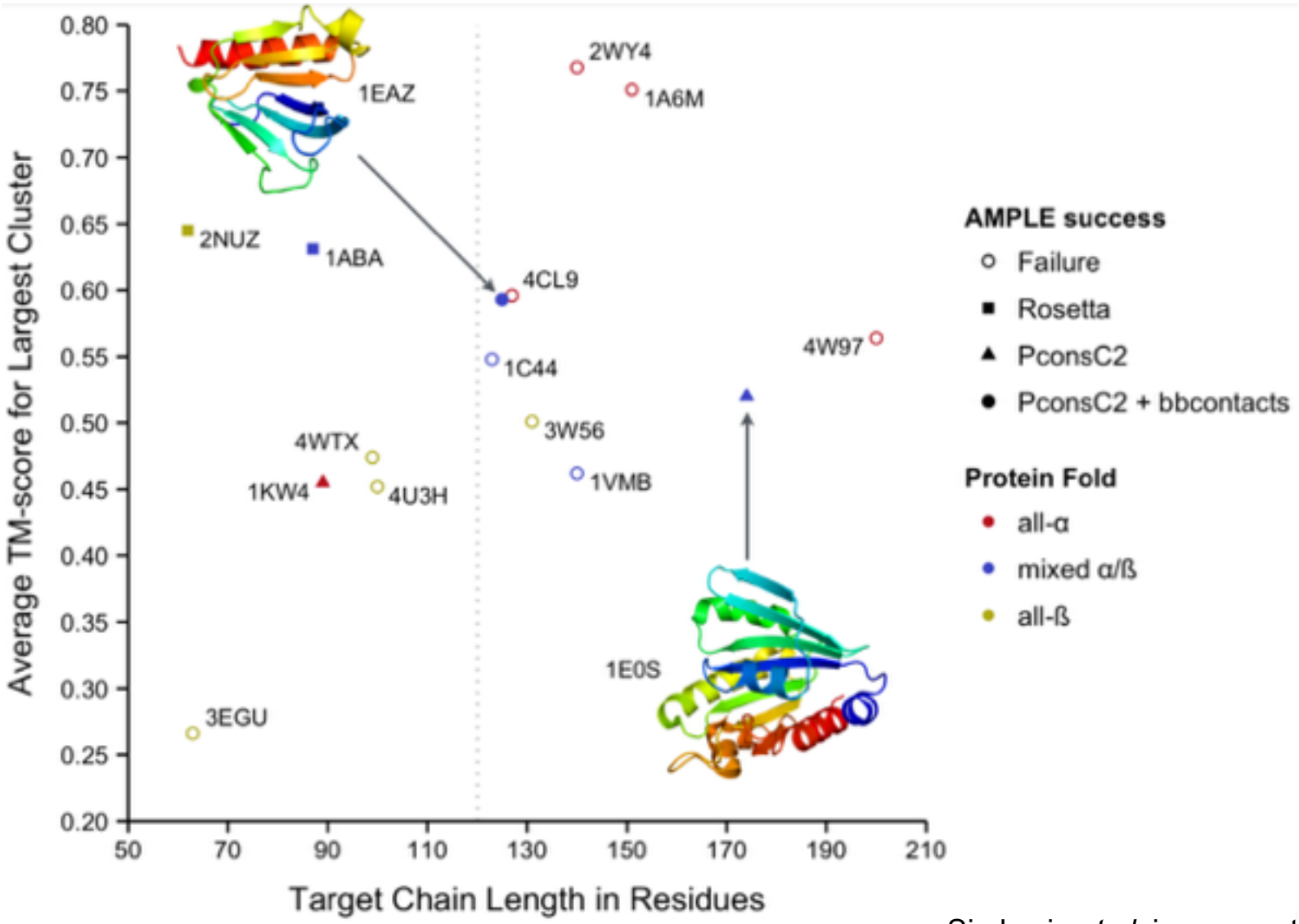
UNIVERSITY OF LIVERPOOL

Ongoing developments

Residue-residue contact prediction



With contact predictions AMPLE succeeds with larger and more β-rich proteins



Simkovic *et al.* in preparation

Summary

- AMPLE – a pipeline for Molecular Replacement with unconventional models.
- 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.
- very easy to use.

Credits

- Dan Rigden, Olga Mayans – IIB, Liverpool University
- Felix Simkovic, Adam Simpkin and Jaclyn Bibby – Liverpool University
- Ronan Keegan, Martin Wynn – CCP4 (STFC)
- BBSRC
- Hartree Center
- **Phaser:** Randy Read, Airlie McCoy & Gabor Bunkozci
- **SHELX:** Andrea Thorn, Tim Gruene & George Sheldrick
- **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**)