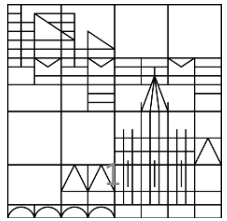
AMPLE, a pipeline for unconventional MR

Dan Rigden (with Ronan Keegan, Jens Thomas, Felix Simkovic, Adam Simpkin, Martyn Winn and Olga Mayans)

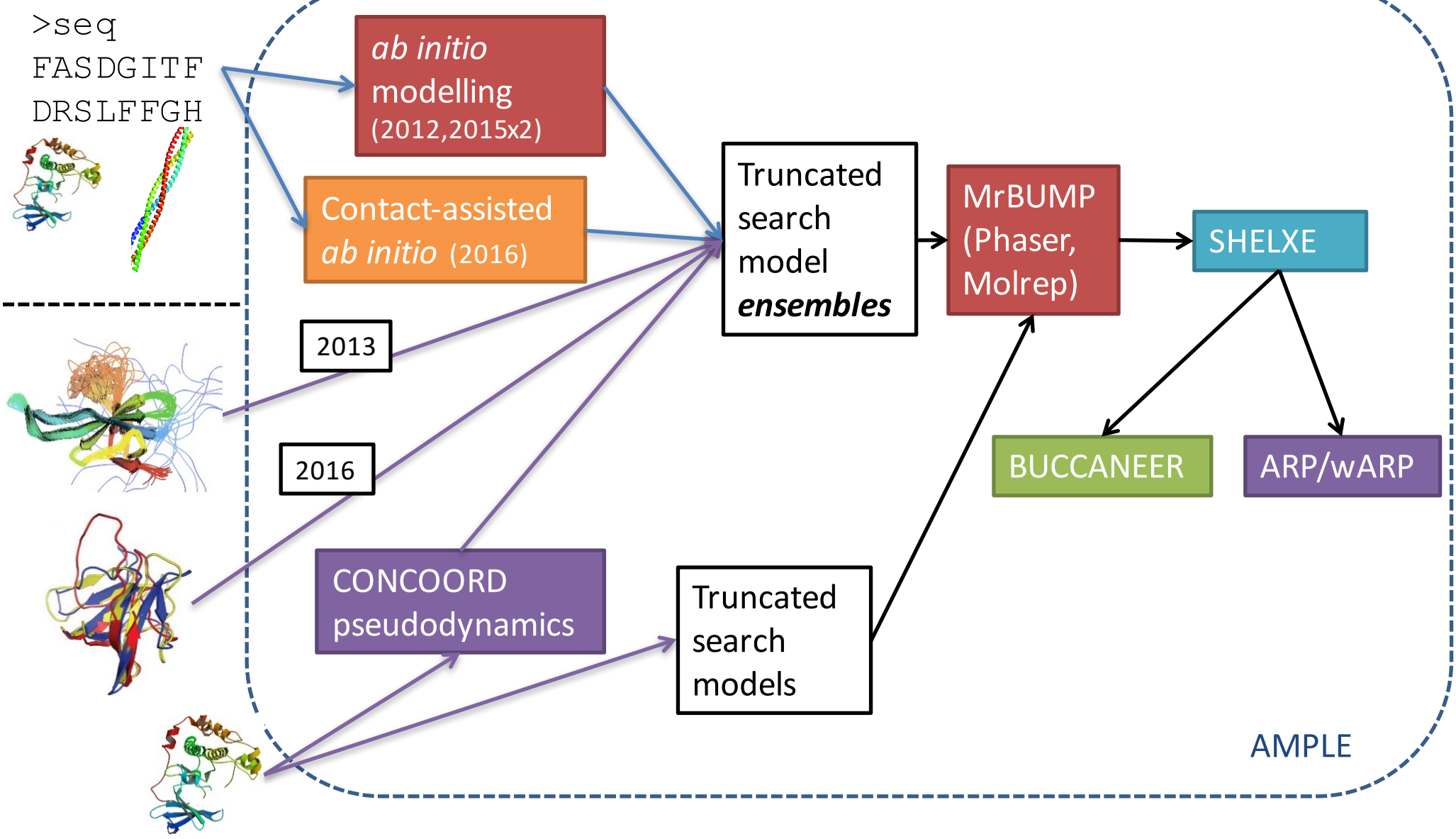
Science & Technology Facilities Council

UNIVERSITY OF LIVERPOOL

Universität Konstanz



Modelling

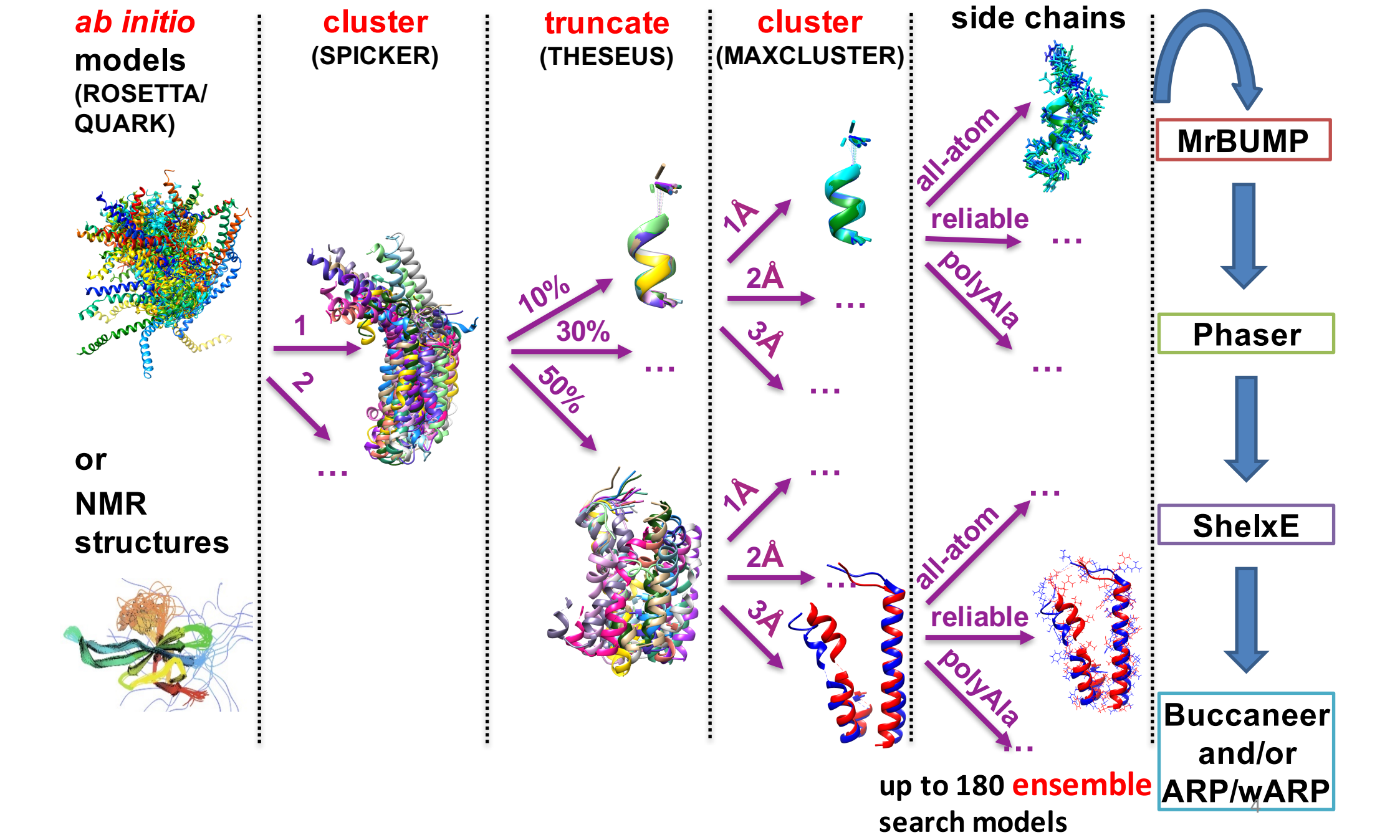


Transforming and processing experimental structures

The original AMPLE conception

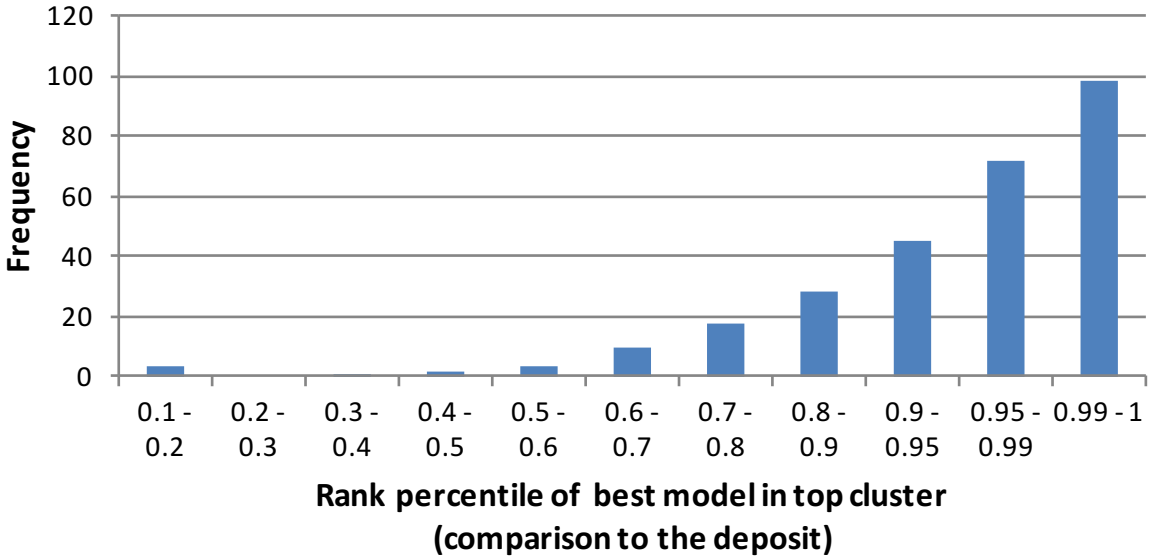
Making MR search models from *ab initio* models by clustering and truncating

Clustering and truncating to make ensemble search models

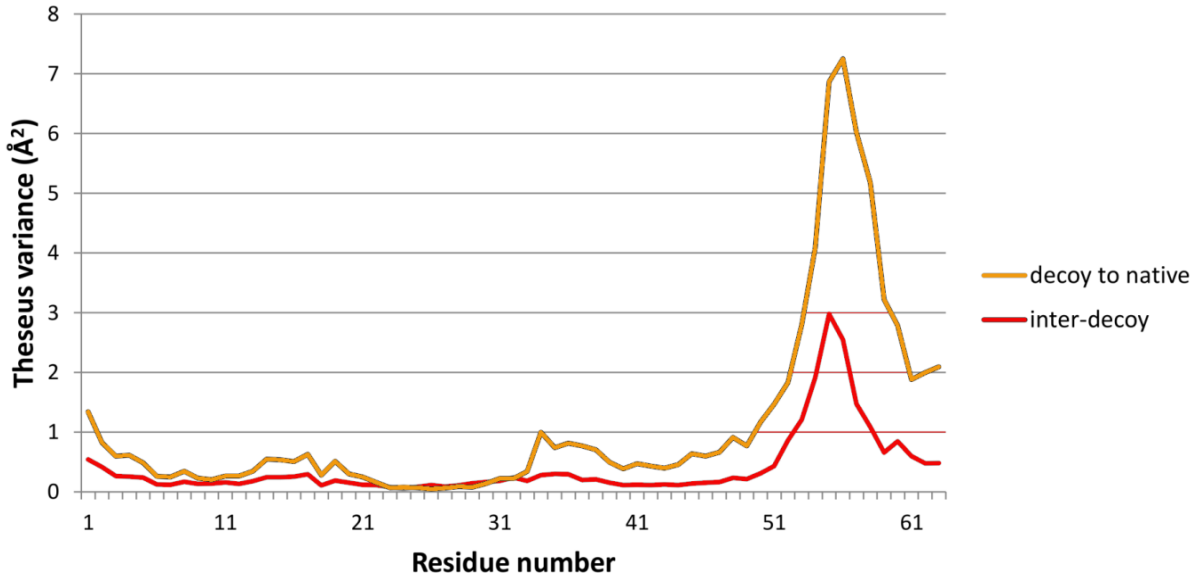


Clustering and truncation work as expected

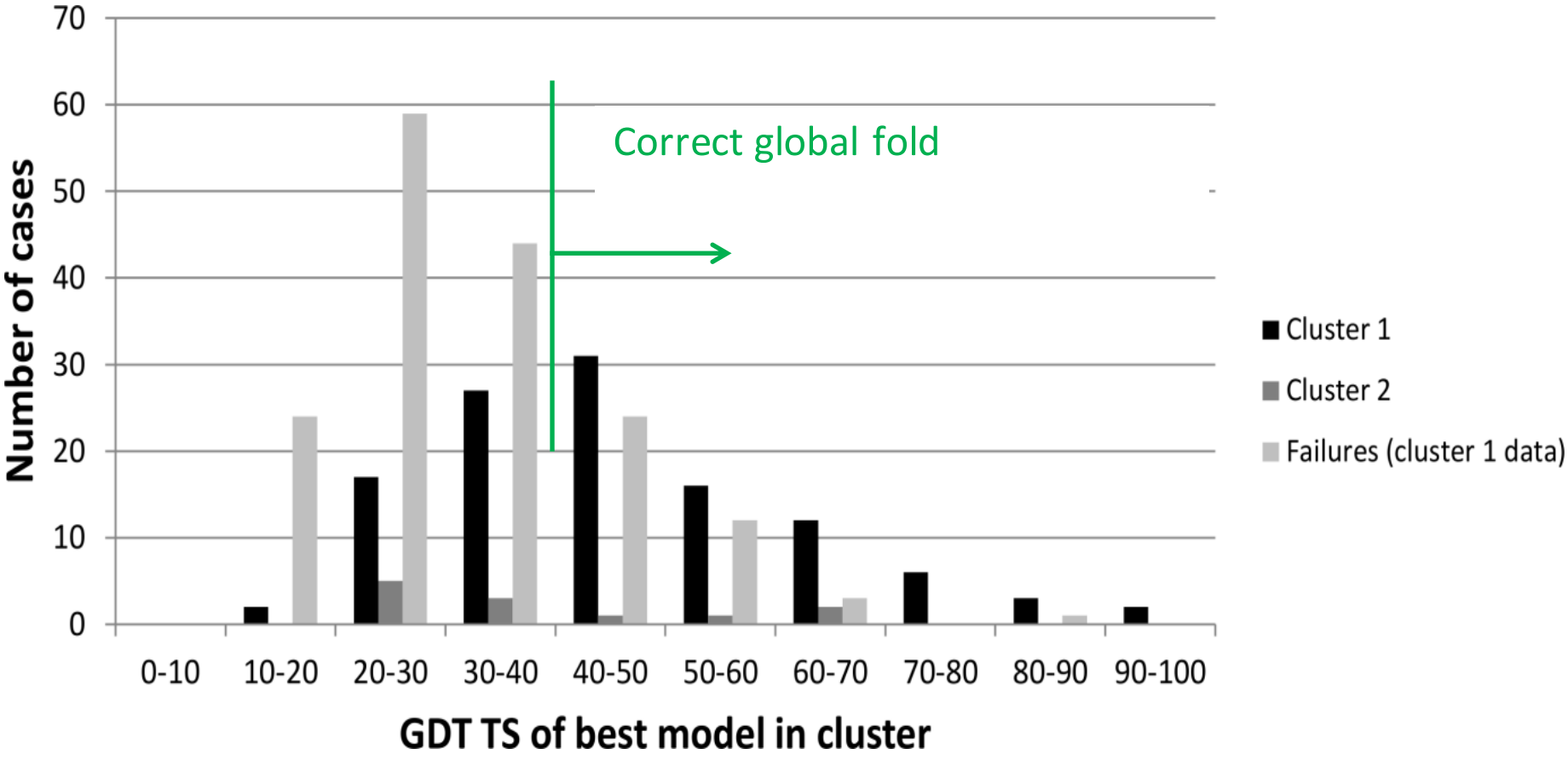
Largest cluster usually contains near-top models



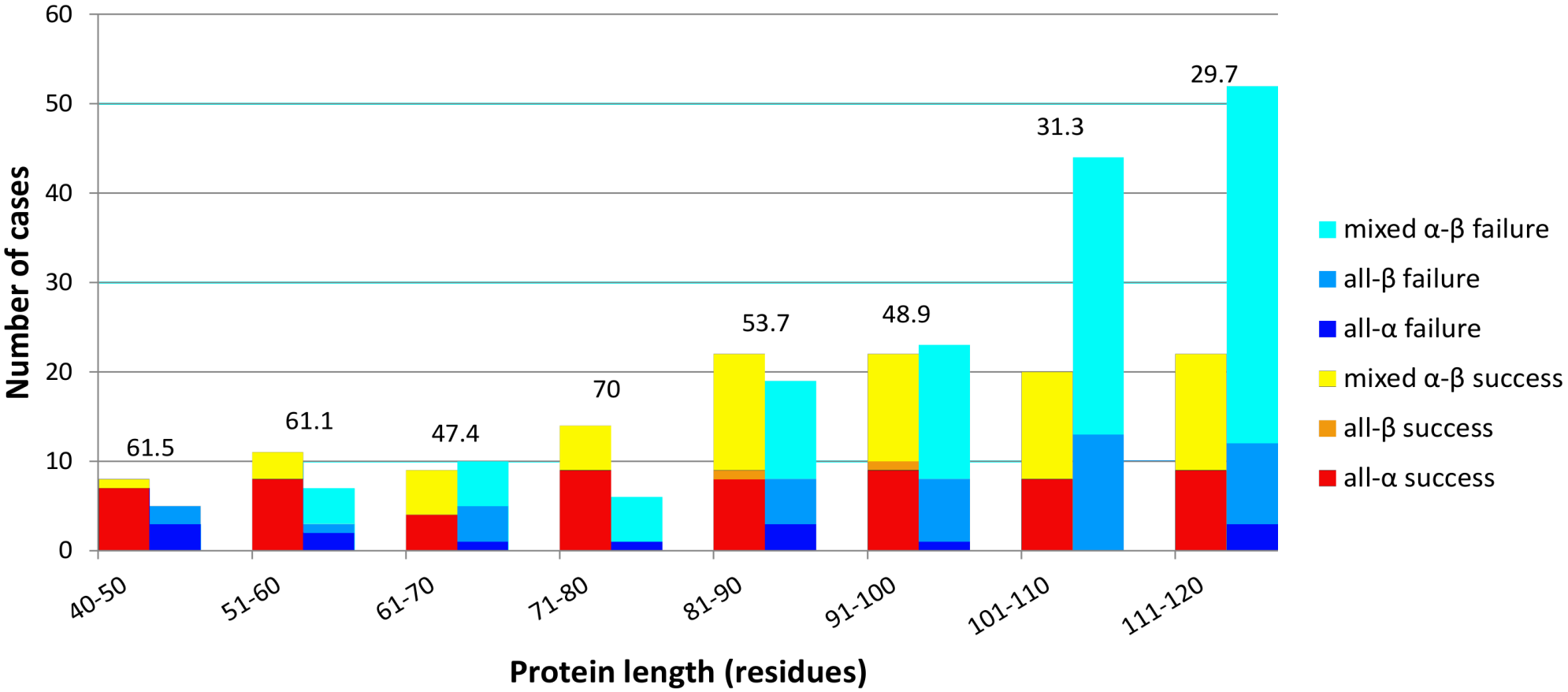
Variance between models often indicates divergence from native structure



Better modelled targets are more successful, but success also with poor models via truncation



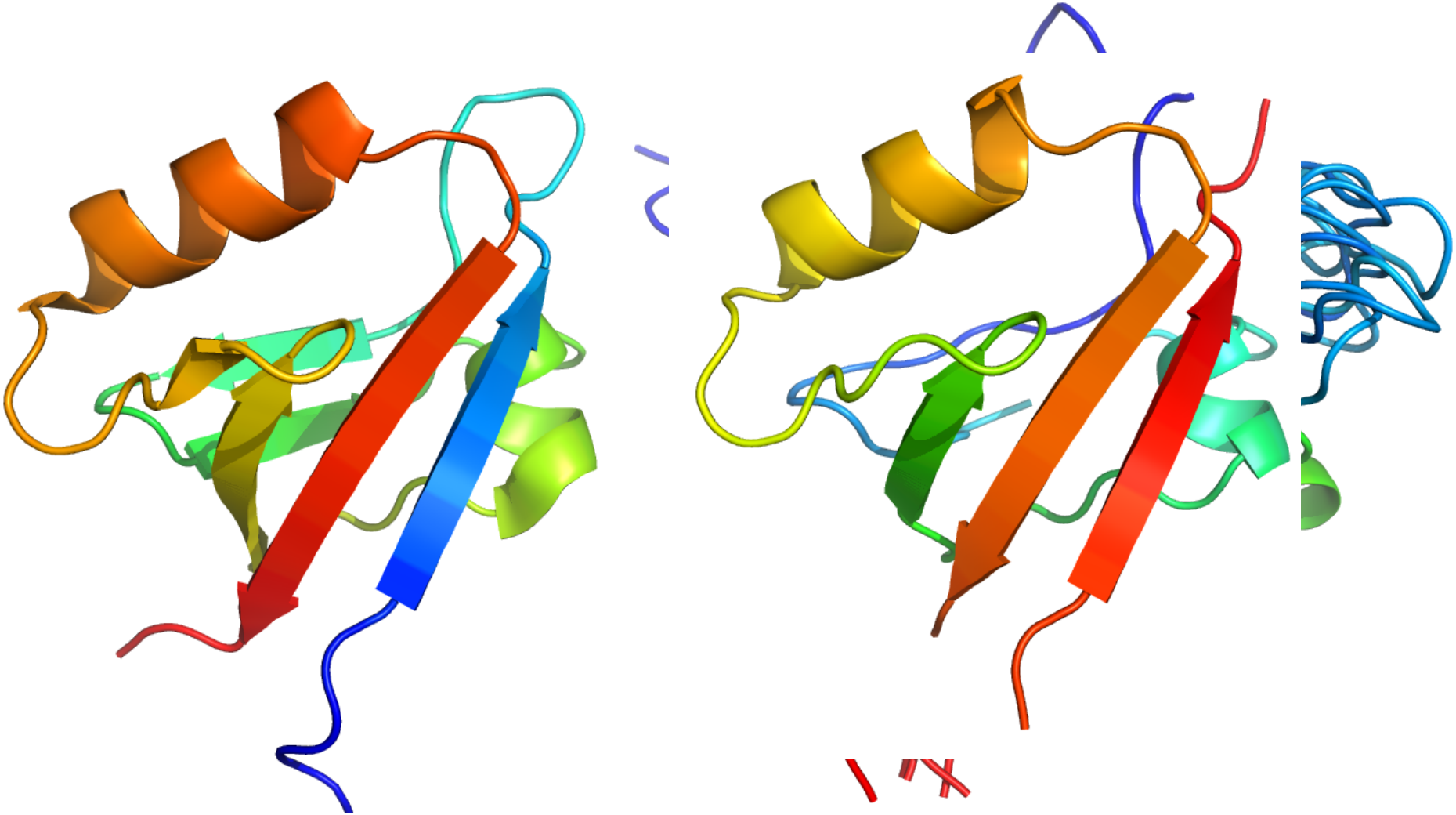
Success with large set of small proteins



- 126 out of 295 (43%) solved. Treated as novel folds.
- all-α, 80%; all-β, 2%; mixed α_β, 37%

Bibby *et al.* (2012) Acta Cryst D**68**, 1622

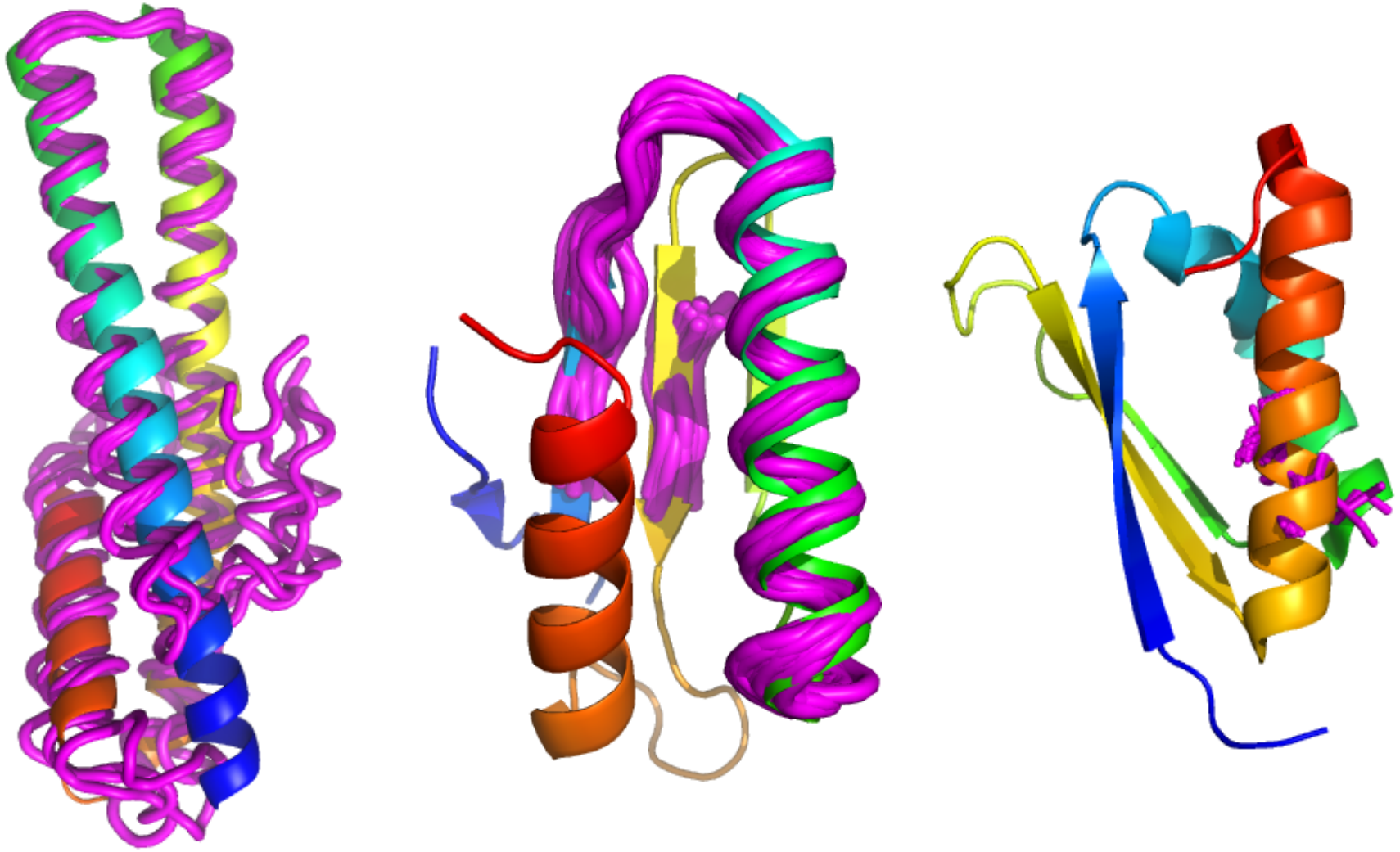
Example success (1r6j, a PDZ domain)



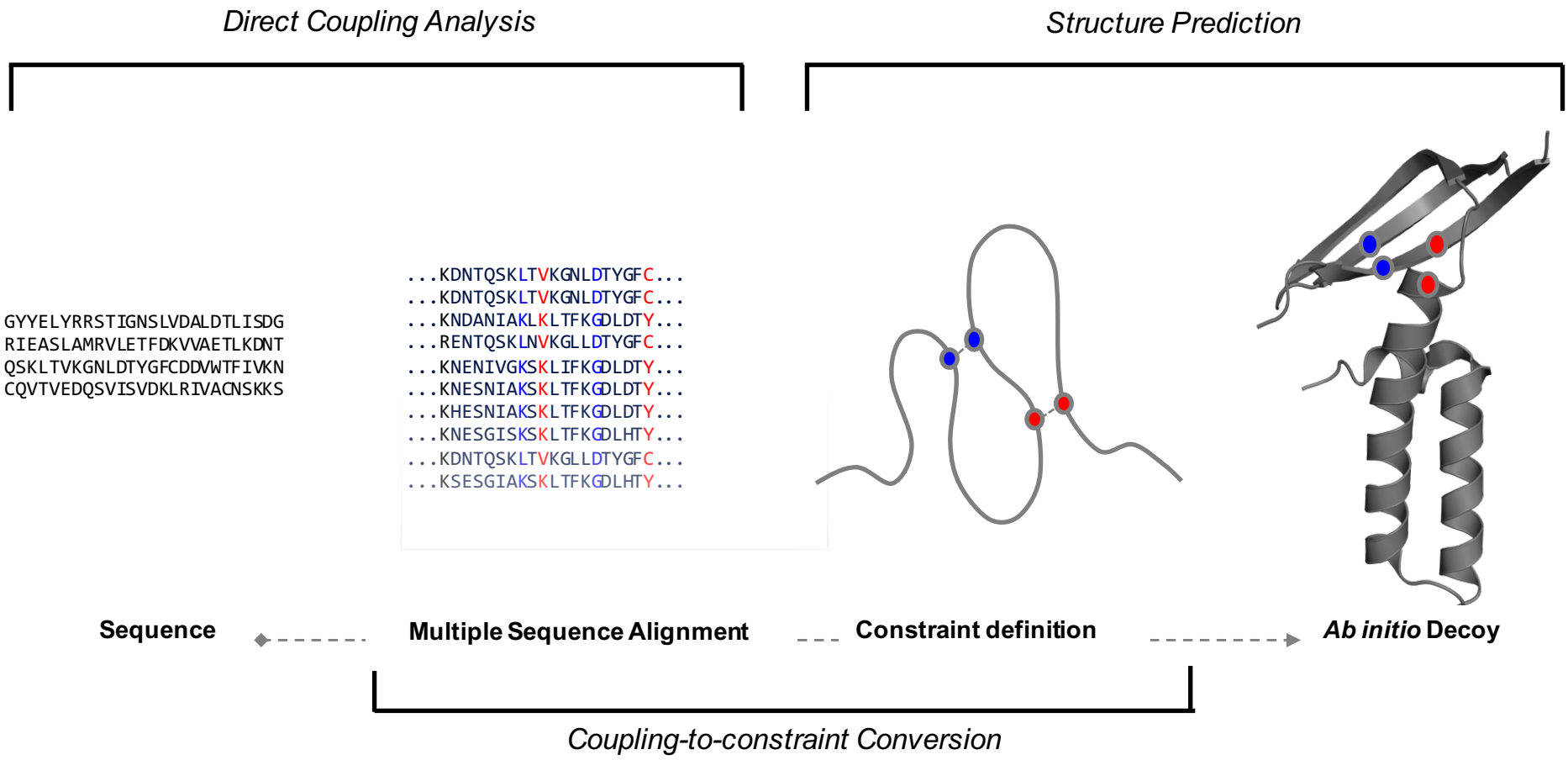
crystal structure

unreleased Rosetta cluster

Success across search model size range

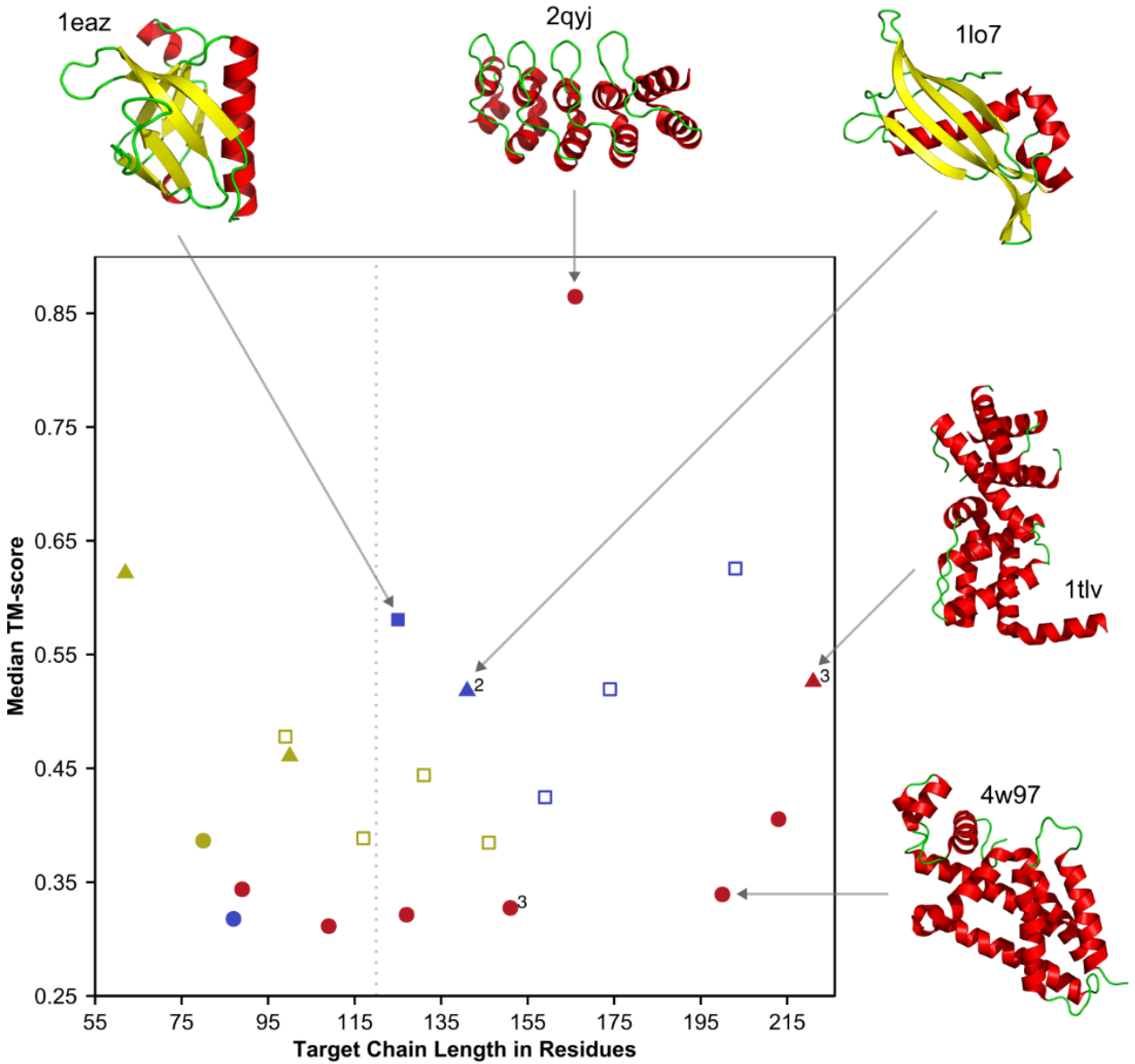


Contact-aided *ab initio* modelling can address larger and harder globular targets



We also integrated a β-structure specific contact method to address β-proteins

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



- Pcons2+bbcontacts
- Pcons2
- No contacts

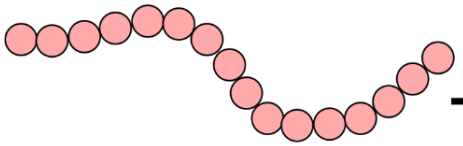
Protein Fold

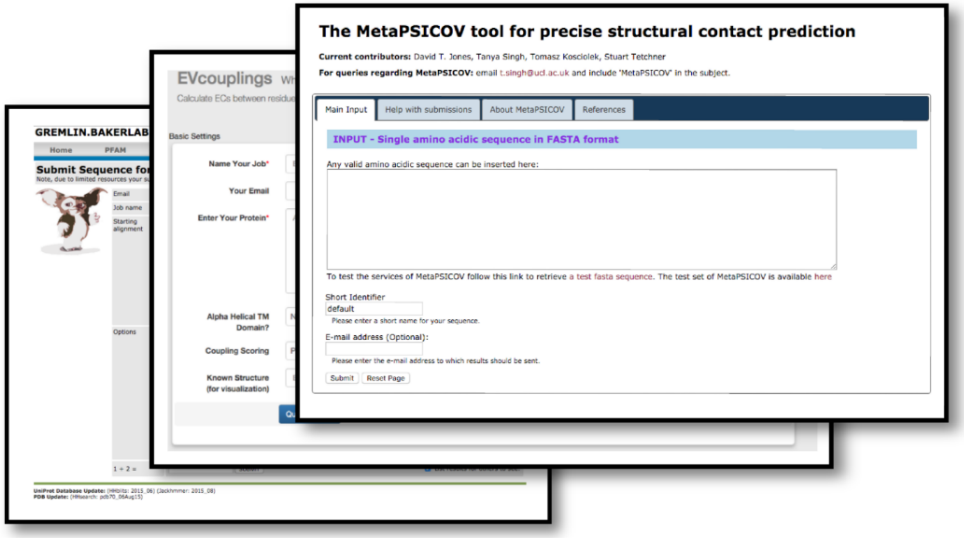
- all-α
- mixed α/ß
- all-ß

Simkovic *et al.* (2016) IUCrJ. submitted

Providing contact info to AMPLE

Sequence





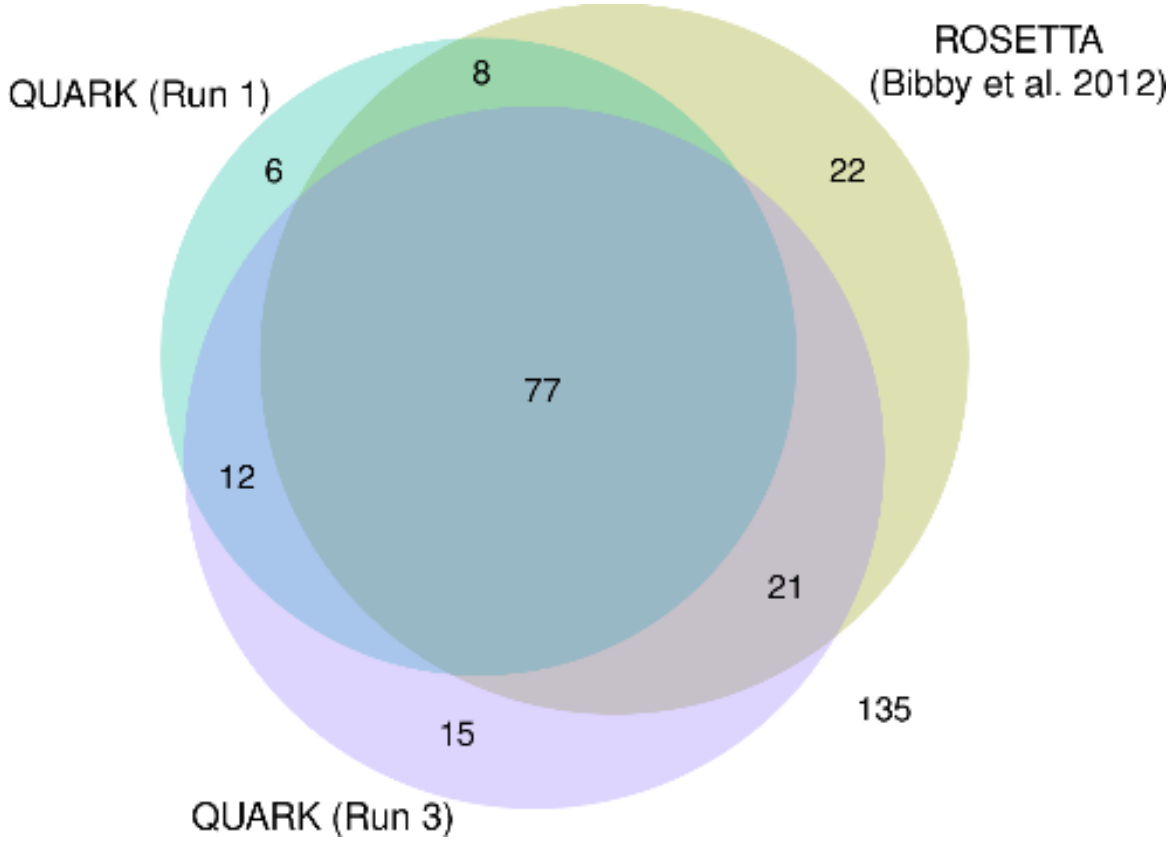
Contact conversion
(if required)

Other tools provided in Ample
- #effective sequences estimate (requires CD-HIT installation)
- Contact map plotting

Contact Restraints Options
-bbcontacts_file
-contact_file
-distance_to_neighbour
-energy_function
-native_cutoff
-restraints_factor
-restraints_file
-restraints_weight

QUARK vs ROSETTA for the *ab initio* modelling

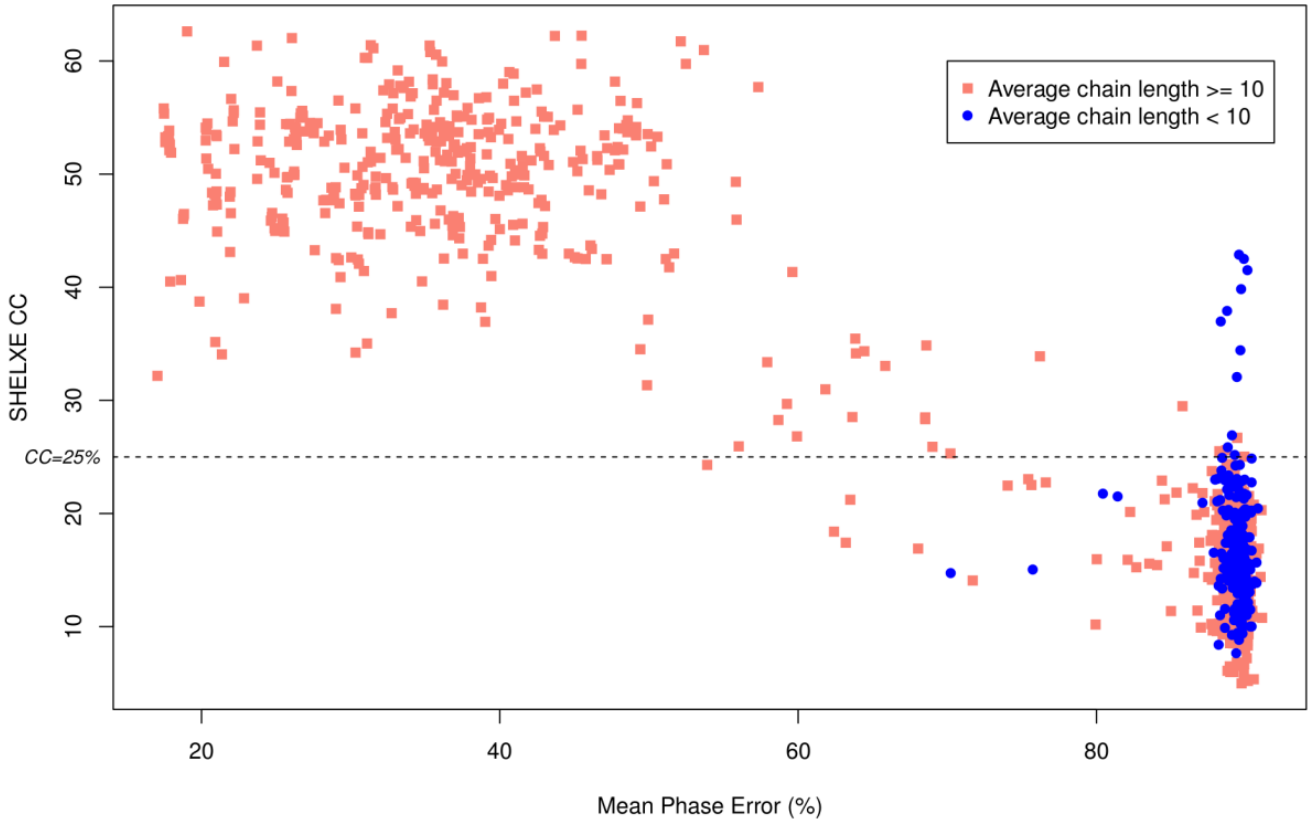
- Re-did testing using QUARK-generated decoy models
- Results are similar to ROSETTA but there are some targets that could only be solved by one or the other
- There is a server for QUARK!



Keegan *et al.* (2015) Acta Cryst. **D71**, 338

QUARK results and SHELXE

- SHELXE following MR
- Criteria for success:
  - SHELXE CC of better than 25%
  - Average fragment chain length of 10 or more
- Result for > 1000 AMPLE jobs using QUARK models

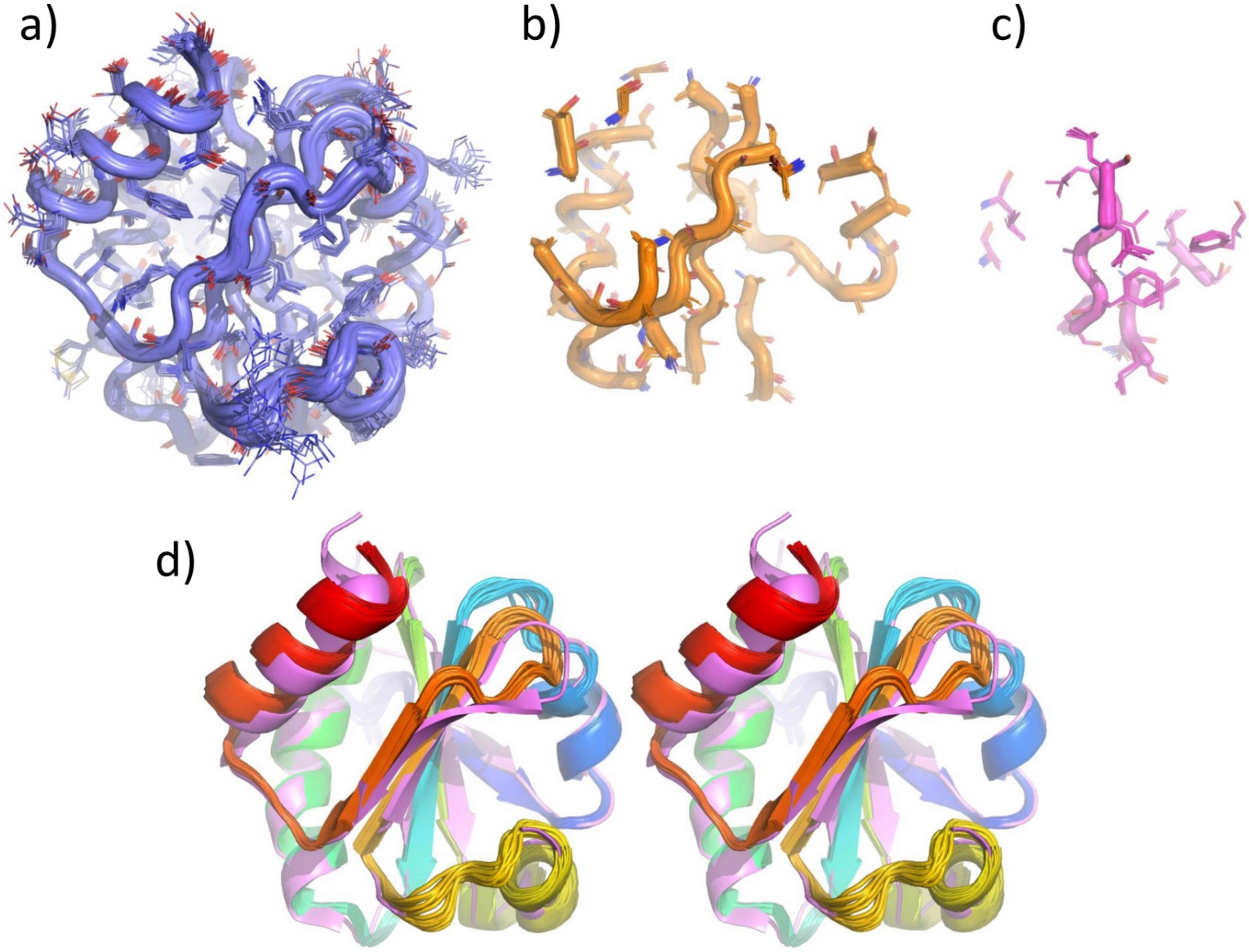


Keegan *et al.* (2015) Acta Cryst. **D71**, 338

AMPLE and NMR structures

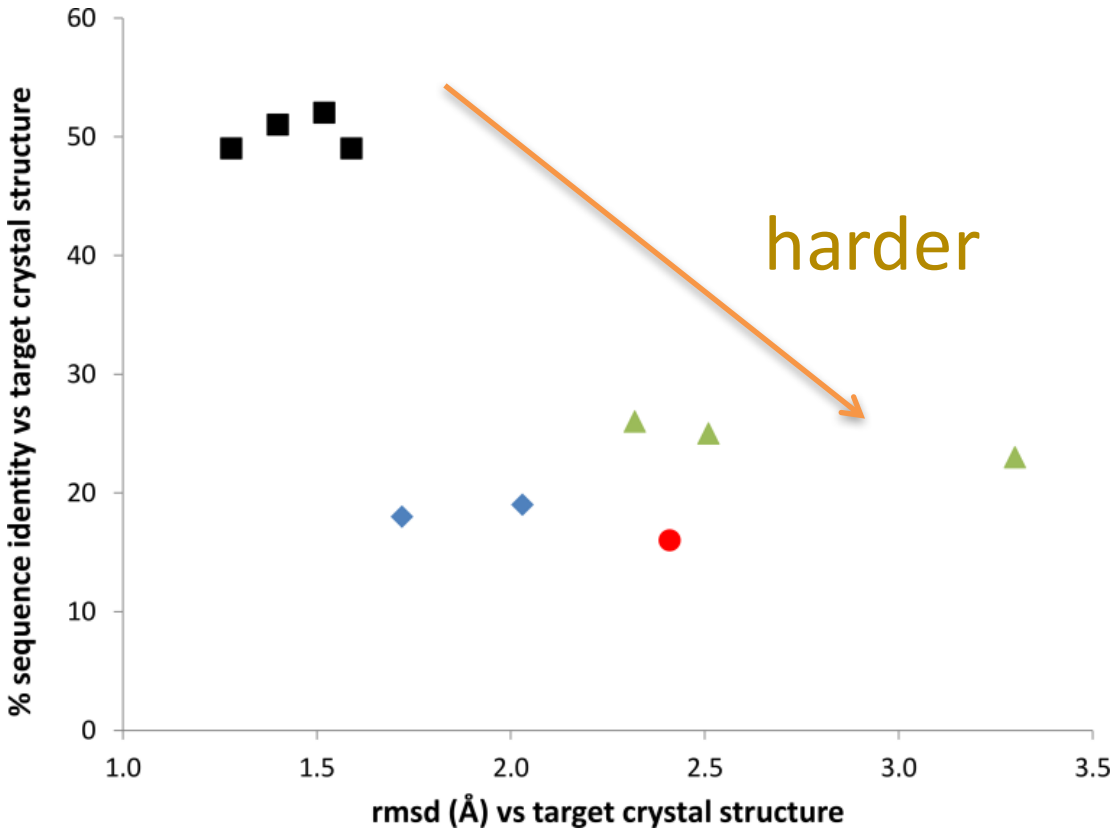
- In *ab initio* modelling variance in a structure cluster indicates unreliability.
- For MR with NMR structures variable regions provide least phasing information
- Can AMPLE's cluster and truncate protocols help with NMR structures (traditionally tricky for NMR)?
- Benchmark vs FindCore

Mao et al. (2009) Structure **19**, 757


a)
b)
c)
d)

AMPLE and NMR structures

- AMPLE solves 24/25 100% sequence identity cases. Similar to FindCore.
- In harder cases, AMPLE succeeds down to 18%, perhaps because of greater sampling. Better than FindCore



FindCore and AMPLE succeed

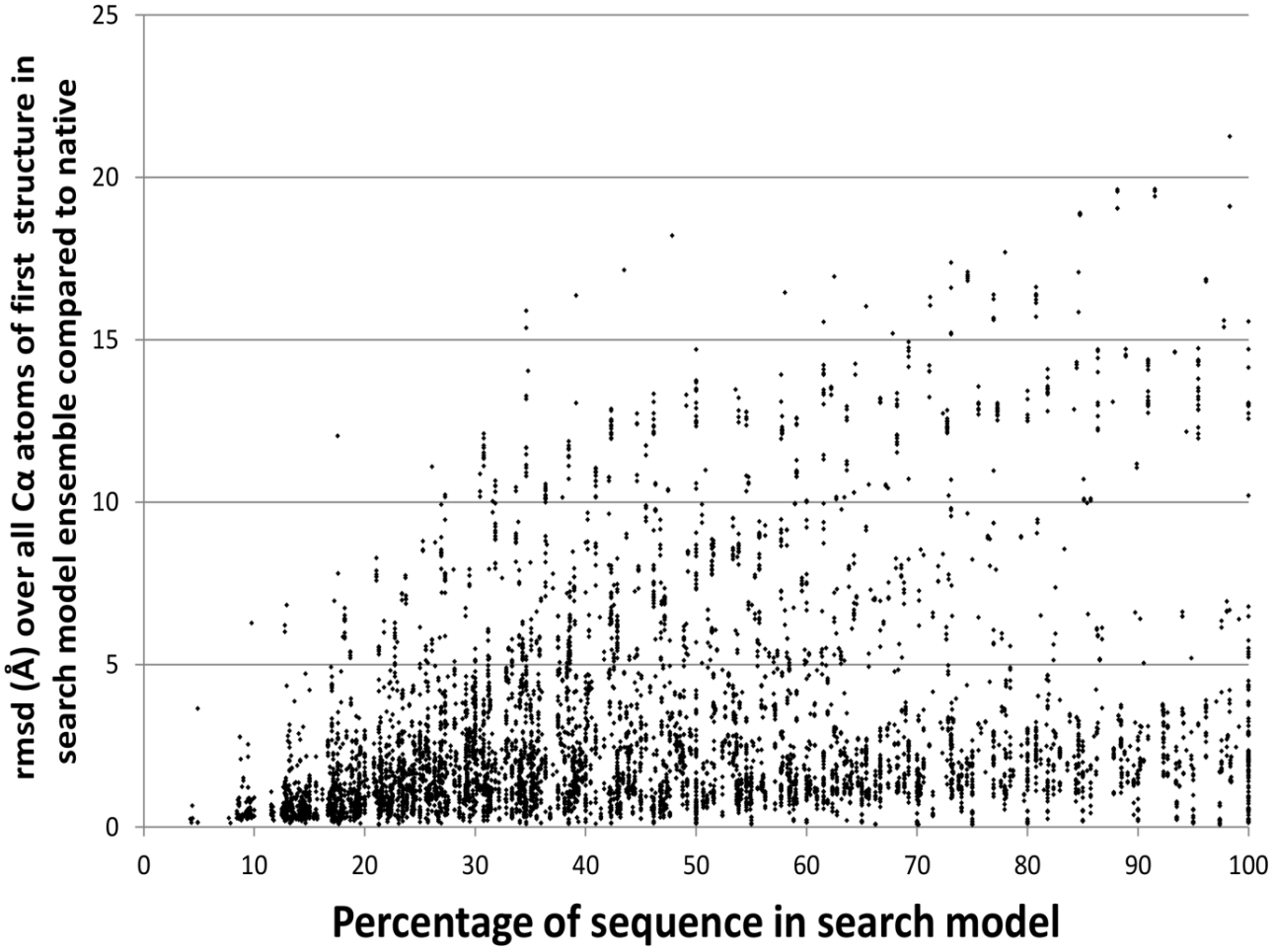
FindCore fails AMPLE succeeds

Both fail

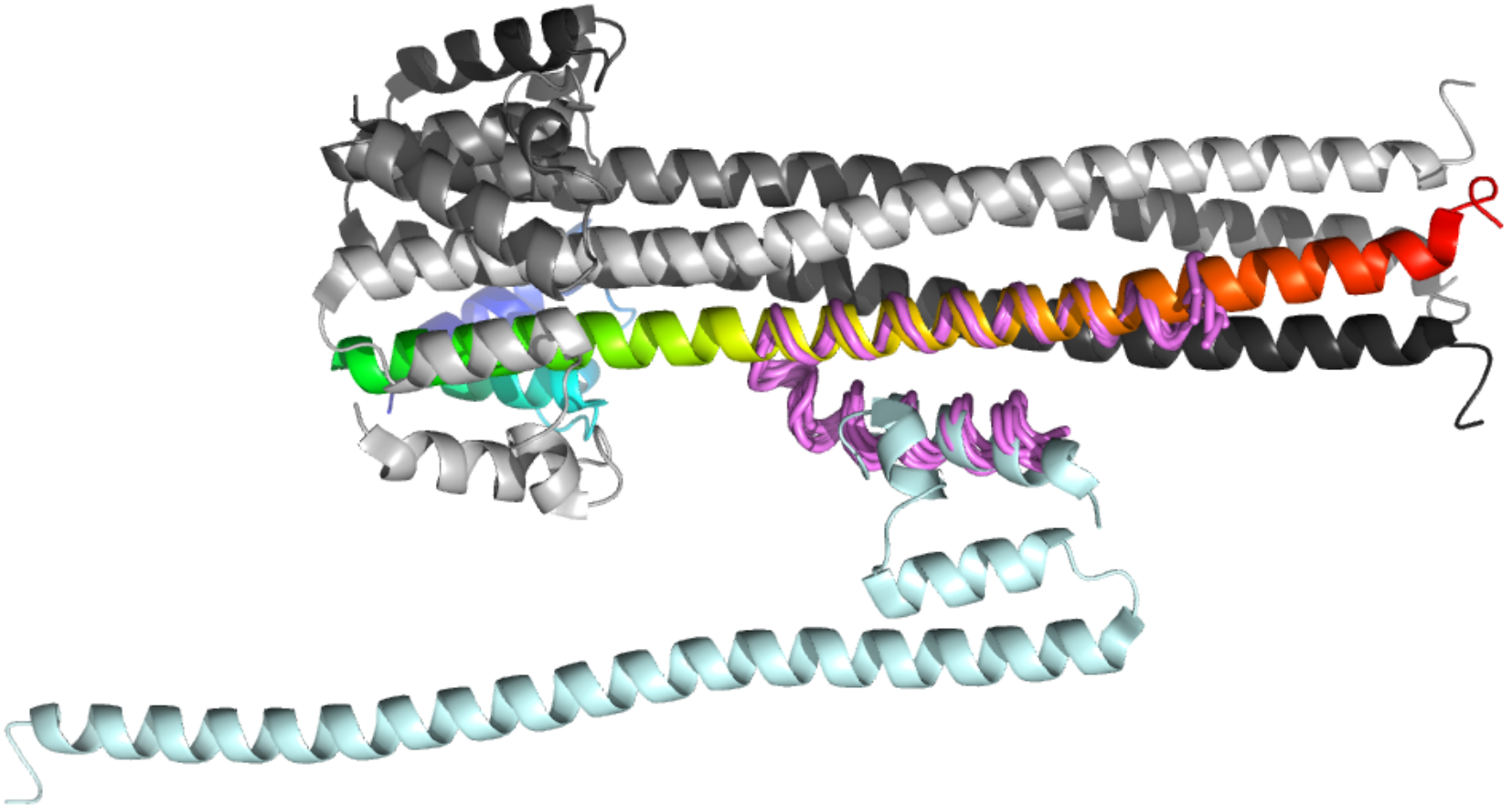
Bibby *et al.* (2013) Acta Cryst D**69**, 2194

With original protein set, inaccurate models sometimes succeeded

>10Å successful ensembles (~5% of total) mainly for coiled-coil or other extended helical targets



Inaccurate models sometimes succeeded



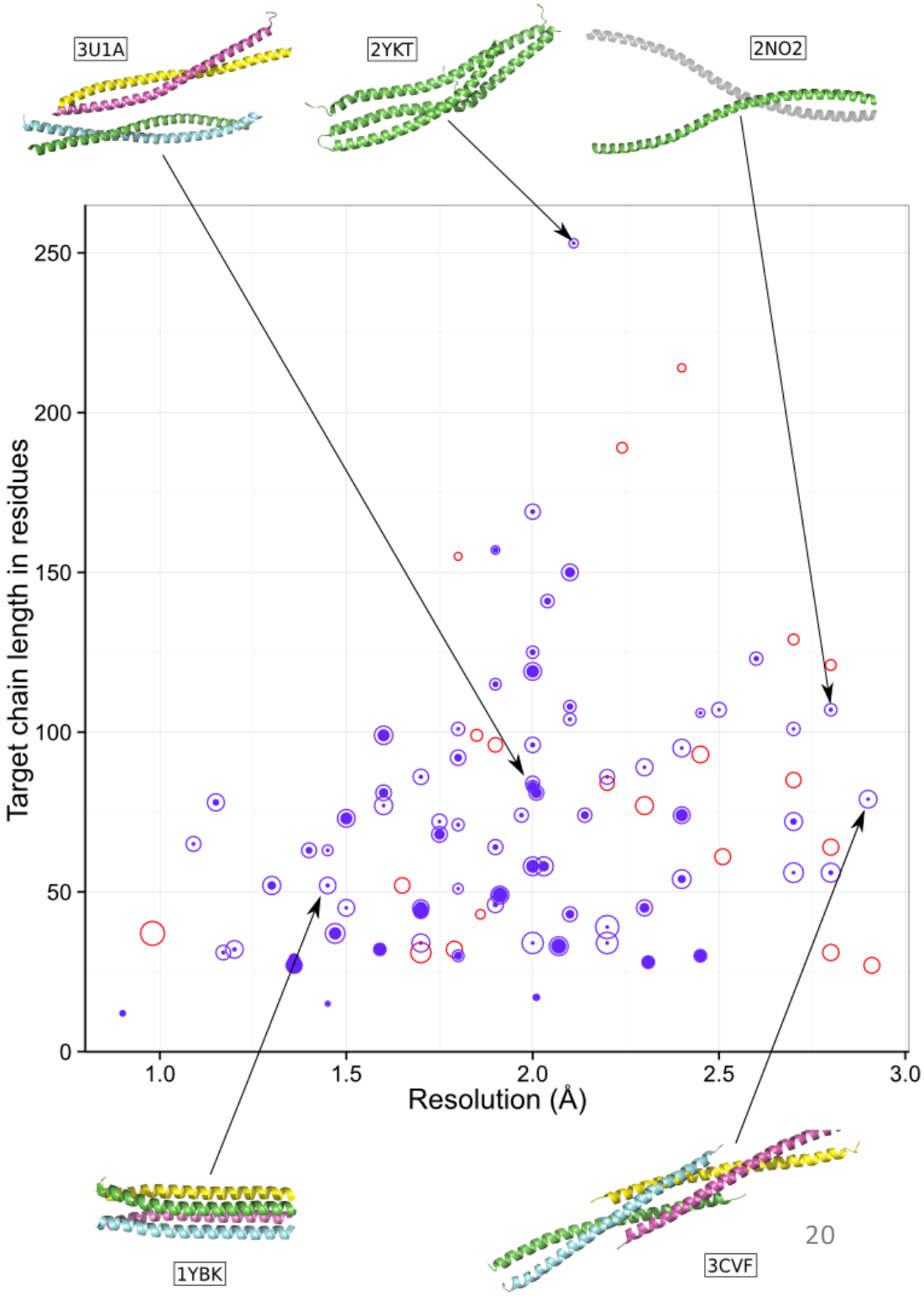
Coiled-coils

Coiled-coils generally considered awkward for MR

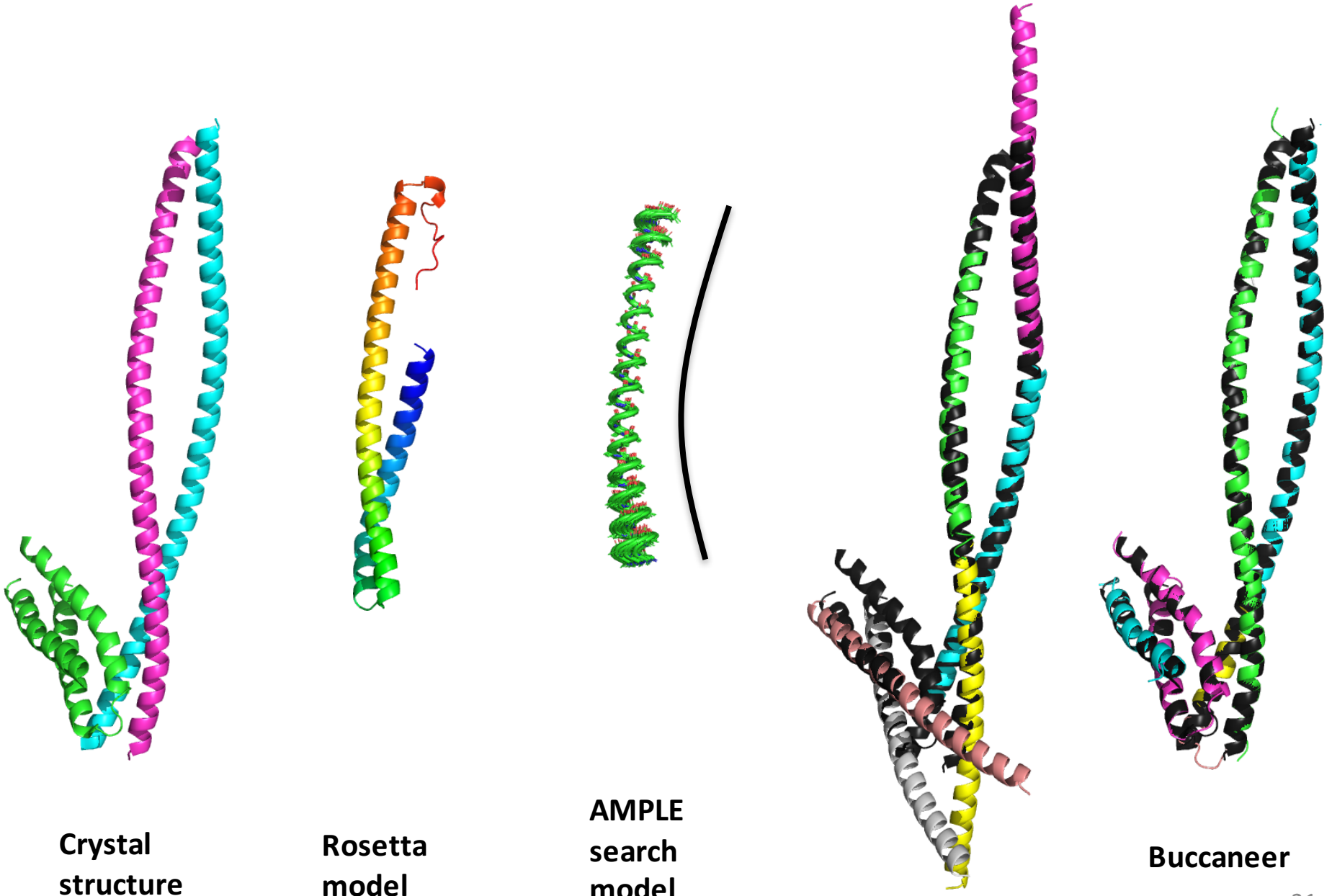
Nevertheless, AMPLE solved more than **~80%. No knowledge of oligomeric state required**

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



Solving coiled-coil *complexes*: 1x79



Crystal structure

Rosetta model

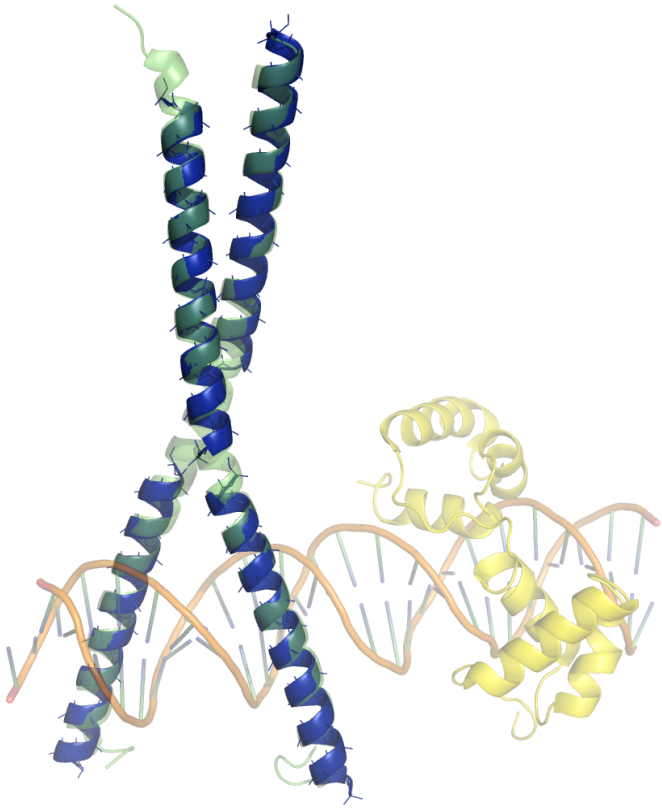
AMPLE search model

Phaser

Buccaneer

Coiled-coils: from bad guys to helpful friends?

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



- **1H8A**: MyB/DNA complex of 278 residues coiled-coil ~50% of scatterers @ 2.23Å.

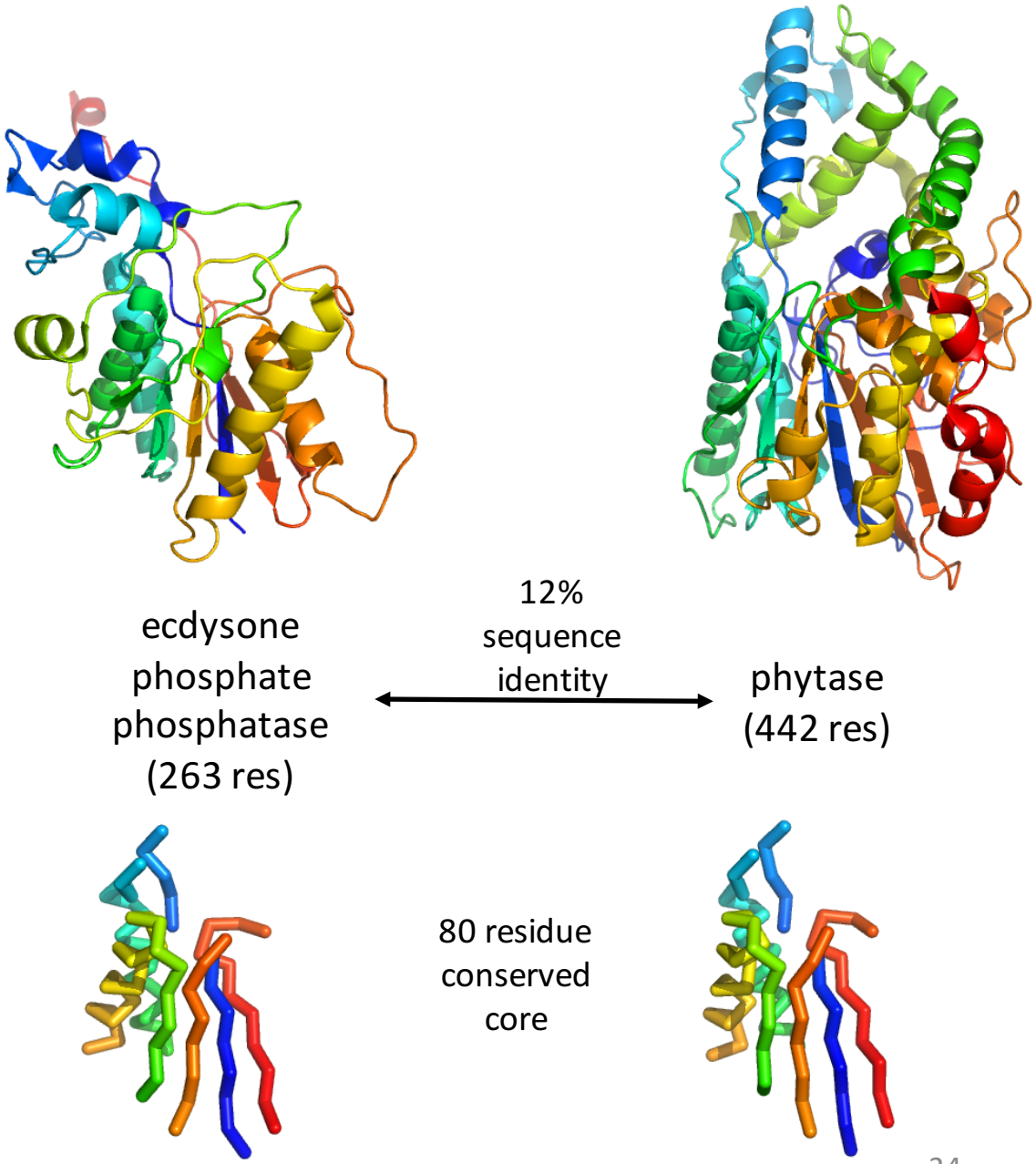
Thomas *et al.* (2015) IUCrJ. **2**, 198

AMPLE and crystal structures

Exploiting the clustering and truncating ideas with distantly homologous crystal structures

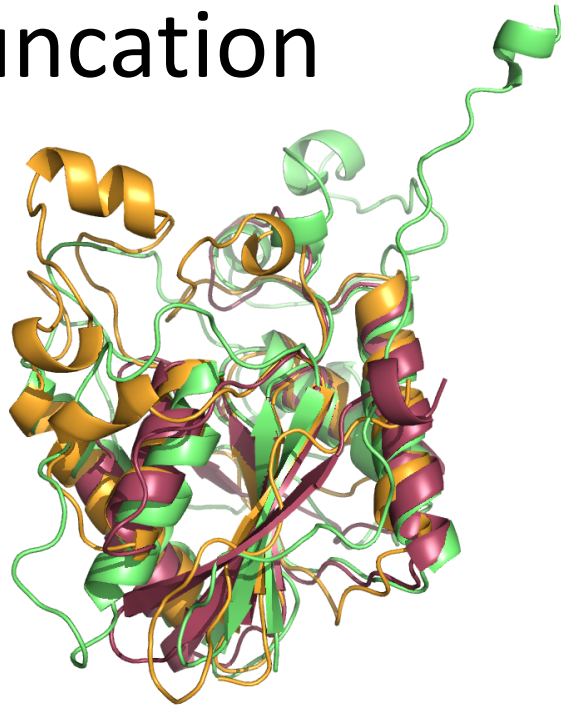
AMPLE for processing crystal structures

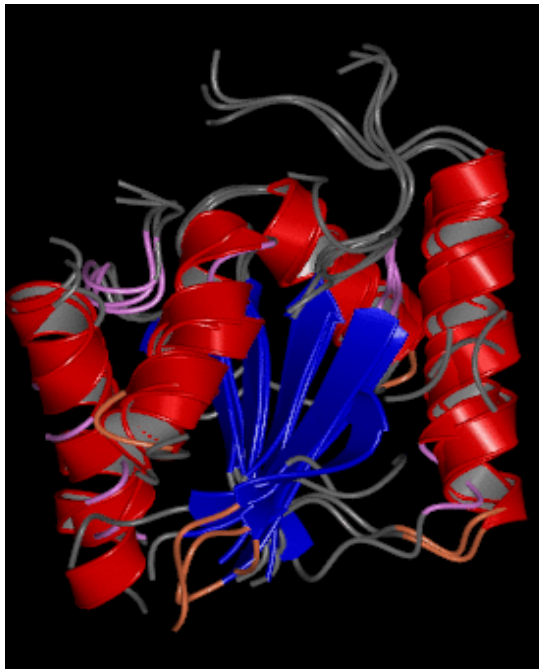
- Often have homologous structures but they are too divergent to solve the target
- AMPLE can help find small, better conserved core for MR



Multi-homologue truncation

- 7 distantly homologous superfamily members (7-28% identity)
- Can 2-3 in superposition solve other 4?
- `-homologs` flag to AMPLE
- `-ampt` flag to MrBUMP



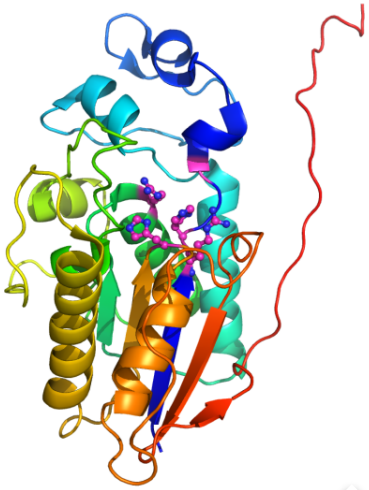


	Structures in superposition					
PDB	1UJB/2A6P/3C7T	1UJB/2A6P	2A6P/3C7T	1UJB/3C7T	MrBUMP	MrBUMP no 1UJB
1E59	0/57	-	-	-	0/10	-
1EBB	11/57	0/57	1/57	14/57	3/10	0/7
2QNI	34/57	14/57	19/57	27/57	0/10	-
3DCY	45/57	41/57	40/57	45/57	2/10	0/7

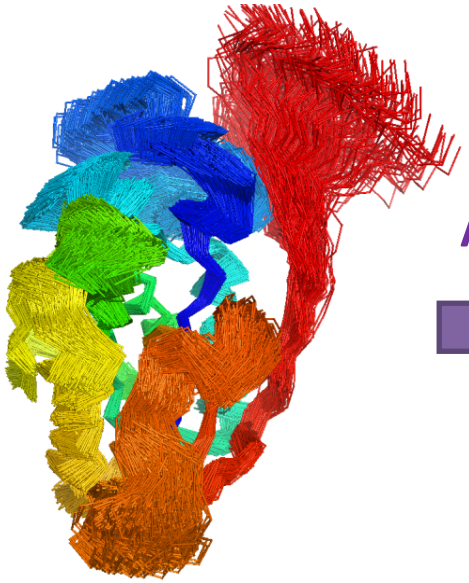
Getting the most out of a single structure

- Highly truncated search models can be sufficient, even required for MR success. How to drive the truncation of a single structure to reach a conserved core?
- Sequence conservation is the most direct measurement but
  - may not be many homologues available
  - can be very slow for large families
- However, packing density/rigidity is typically well-correlated with evolutionary conservation. Can use these measurements, proxies for conservation, to drive truncation
- CONCOORD results reflect rigidity **and** produce ensembles
- Other methods reflect rigidity but without making ensembles

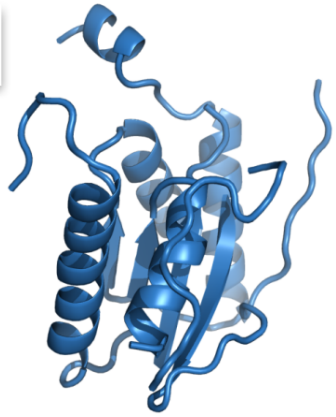
Manual *vs* CONCOORD-derived search models (3c7t)



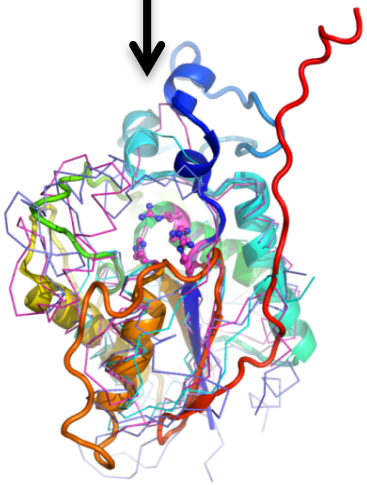
CONCOORD

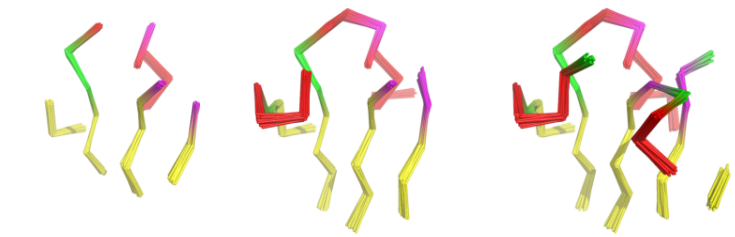


AMPLE



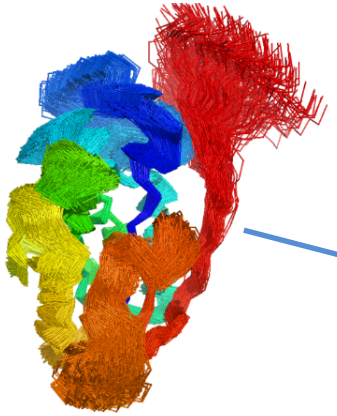
	%id *vs* 3c7t	length	res (Å)	CONCOORD/ AMPLE solved?	MrBUMP with manual edits?
1ujb	22	156	2.1	Yes	No
2qni	13	194	1.8	Yes	Yes
1e59	19	239	1.8	No	No
4e09	18	240	2.45	Yes (3rd cluster)	No
1ebb	23	202	2.3	Yes (2nd cluster)	No
3dcy	20	269	1.75	Yes	No

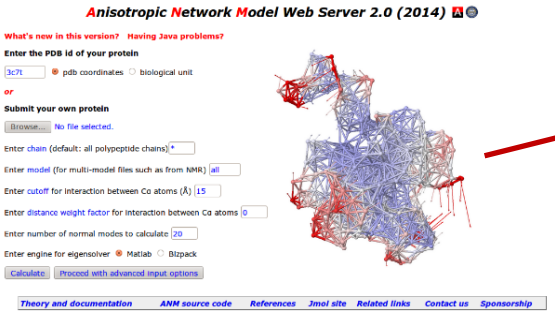


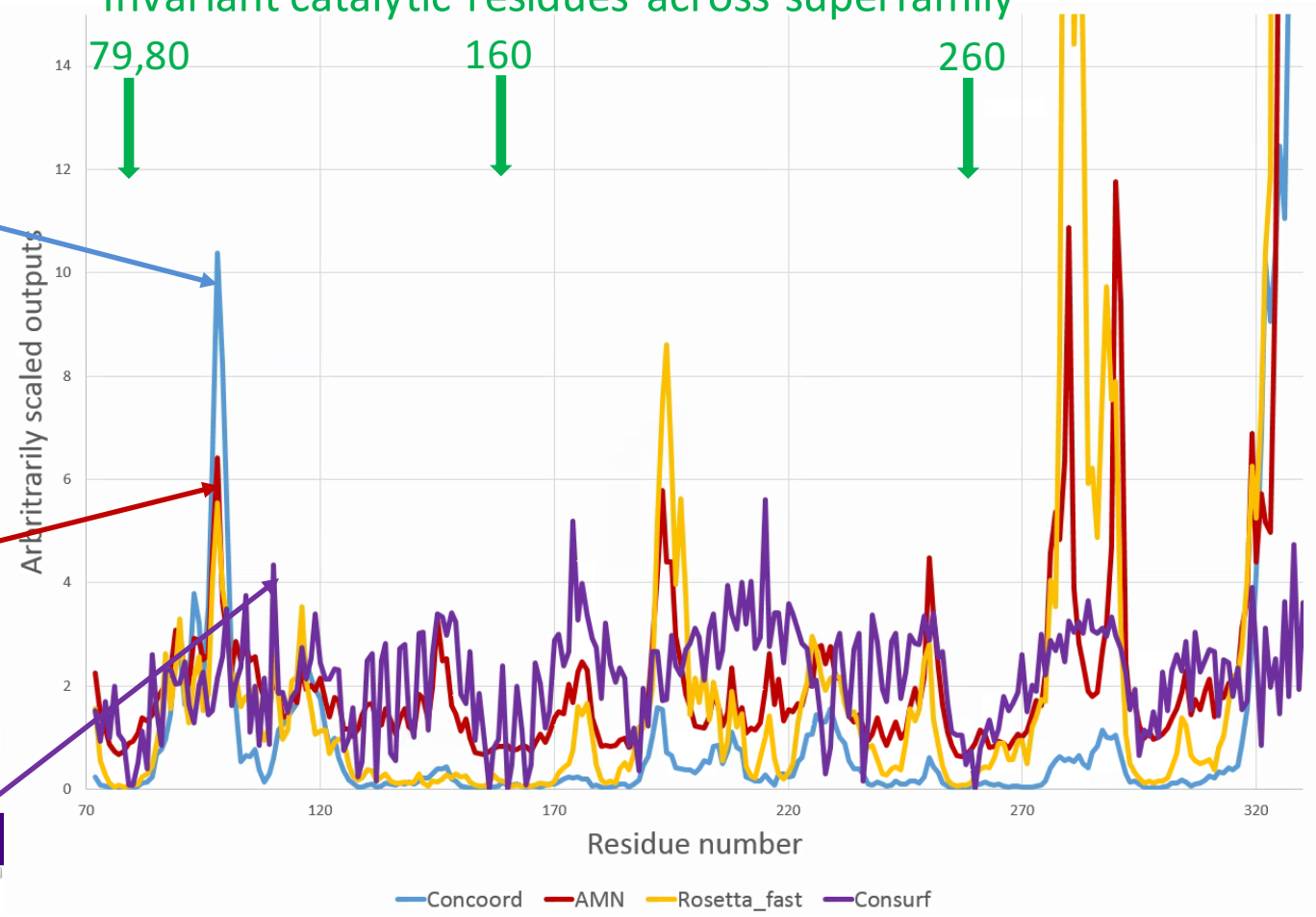


Brutally truncated search model ensembles capture best conserved catalytic core

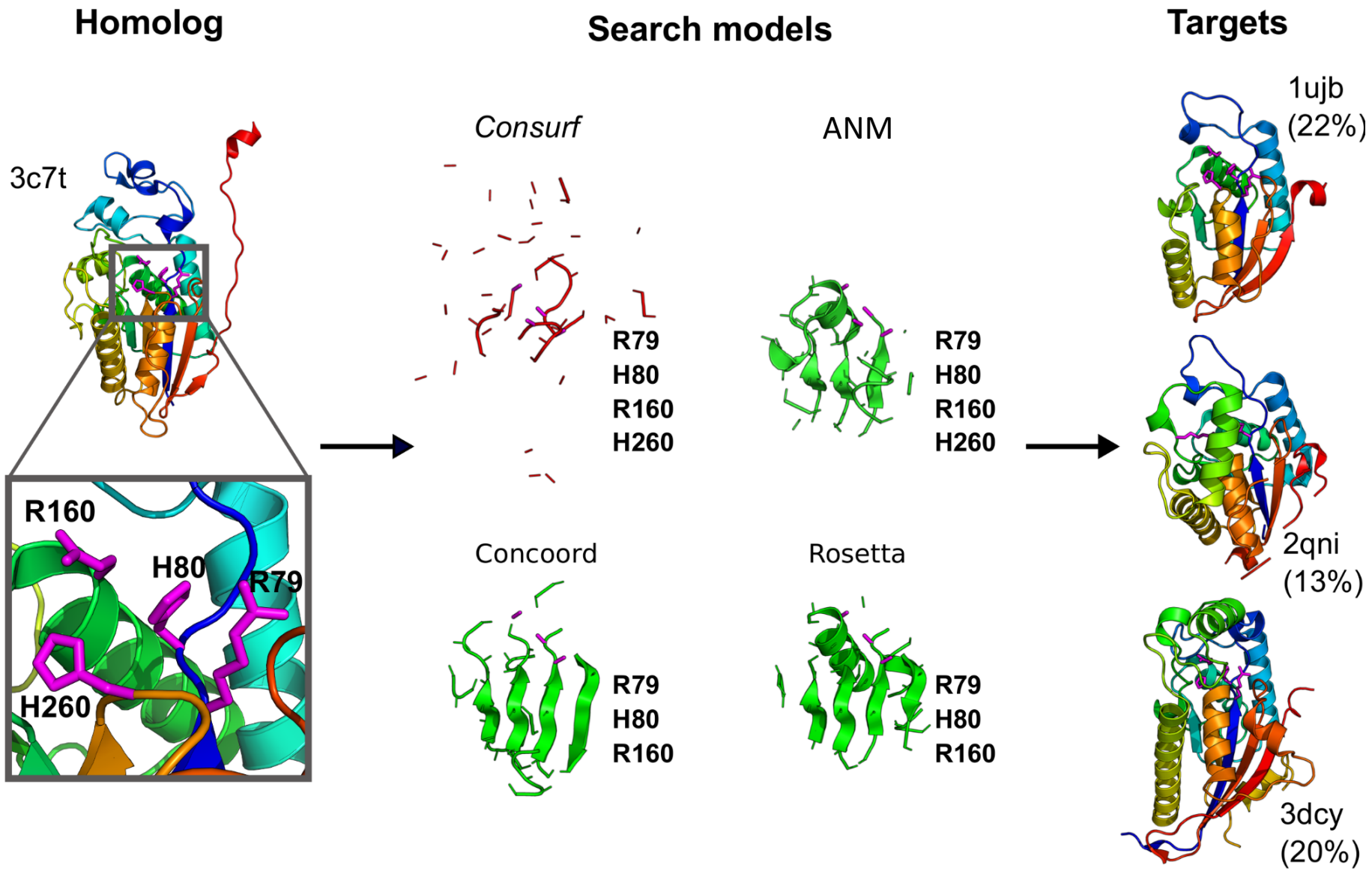
Alternative ways of predicting conservation/rigidity







Conservation/rigidity profiles driving truncation of a single structure



`-single_model, -truncation_scorefile` flags to AMPLE

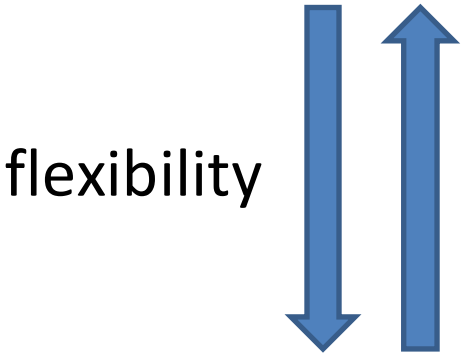
Practicalities

How and when to use AMPLE

How long it takes

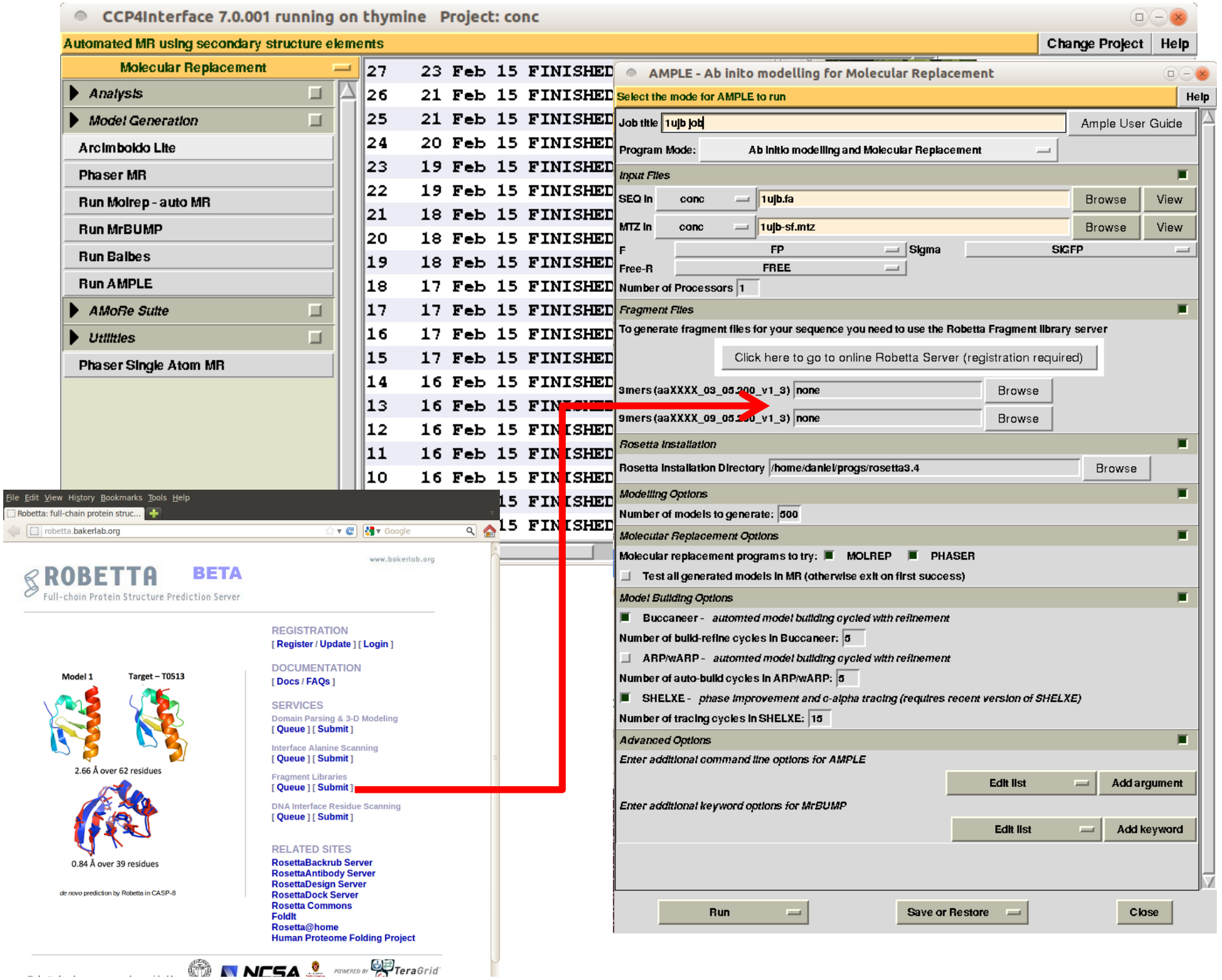
Three ways to use AMPLE

- Server at CCP4online
- Via CCP4i
- Command line



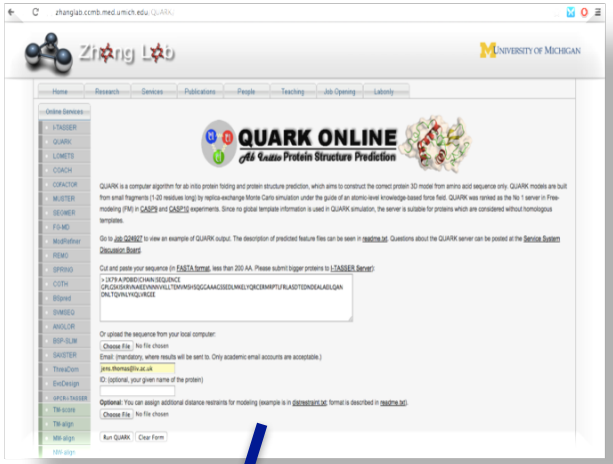

```
daniel@thymine:~/felix$ ample.py -
usage: AMPLE [-h] [-alignment_file ALIGNMENT_FILE] [-all_atom True/False]
             [-arpwarp_cycles ARPWARP_CYCLES] [-blast_dir BLAST_DIR]
             [-buccaneer_cycles BUCCANEER_CYCLES] [-cluster_dir CLUSTER_DIR]
             [-cluster_method CLUSTER_METHOD] [-ccp4_jobid CCP4_JOBID]
             [-constraints_file CONSTRAINTS_FILE] [-debug True/False]
             [-domain_all_chains_pdb DOMAIN_ALL_CHAINS_PDB]
             [-domain_termini_distance DOMAIN_TERMINI_DISTANCE]
             [-dry_run True/False] [-early_terminate True/False]
             [-ensembles_dir ENSEMBLES_DIR] -fasta FASTA
             [-fast_protein_cluster_exe FAST_PROTEIN_CLUSTER_EXE]
             [-F flag for F] [-frags_3mers frags_3mers]
             [-frags_9mers frags_9mers] [-FREE flag for FREE]
             [-gesamt_exe gesamt_exe] [-homologs True/False]
             [-homolog_aligner homolog_aligner] [-ideal_helices True/False]
             [-improve_template improve_template] [-LGA path_to_LGA dir]
             [-make_frags True/False] [-make_models True/False]
             [-maxcluster_exe MAXCLUSTER_EXE] [-max_array_jobs MAX_ARRAY_JOBS]
             [-max_ensemble_models MAX_ENSEMBLE_MODELS]
             [-missing_domain True/False] [-models models] [-mr_keys -mr_keys]
             [-mr_sequence MR_SEQUENCE] [-mustang_exe mustang_exe]
             [-name job_name] [-native_pdb native_pdb] [-nmasu NMASU]
             [-nmodels number of models] [-nr nr] [-nmr_model_in nmr_model_in]
             [-nmr_process nmr_process] [-nmr_remodel True/False]
             [-nmr_remodel_fasta nmr_remodel_fasta] [-no_gui True/False]
             [-nproc Number of Processors] [-num_clusters NUM_CLUSTERS]
             [-output_pdb OUTPUT_PDB] [-purge True/False]
             [-percent percent_truncation] [-psipred_ss2 psipred file]
             [-phaser_kill phaser_kill] [-phaser_rms phaser_rms]
             [-phenix_exe phenix_exe]
             [-rg_reweight radius of gyration reweight]
             [-rosetta_AbinitioRelax rosetta_AbinitioRelax]
             [-ROSETTA_cluster path to Rosettas cluster]
             [-rosetta_db rosetta_db] [-rosetta_dir rosetta_dir]
             [-rosetta_fragments_exe rosetta_fragments_exe]
```

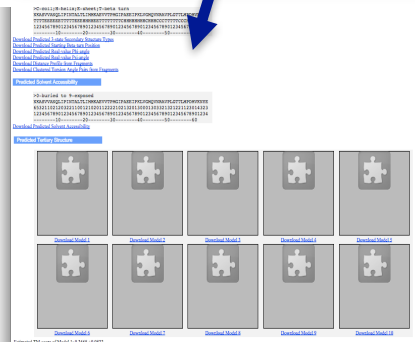
```
> ample.py -fasta my.fasta \
-mtz my.mtz -models          \
/home/me/models/
```


CCP4Interface 7.0.001 running on thymine Project: conc
Automated MR using secondary structure elements
Change Project
Help
Molecular Replacement
Analysis
Model Generation
Arcimboldo Lite
Phaser MR
Run Molrep - auto MR
Run MrBUMP
Run Balbes
Run AMPLE
AMoRe Suite
Utilities
Phaser Single Atom MR
27 23 Feb 15 FINISHED
26 21 Feb 15 FINISHED
25 21 Feb 15 FINISHED
24 20 Feb 15 FINISHED
23 19 Feb 15 FINISHED
22 19 Feb 15 FINISHED
21 18 Feb 15 FINISHED
20 18 Feb 15 FINISHED
19 18 Feb 15 FINISHED
18 17 Feb 15 FINISHED
17 17 Feb 15 FINISHED
16 17 Feb 15 FINISHED
15 17 Feb 15 FINISHED
14 16 Feb 15 FINISHED
13 16 Feb 15 FINISHED
12 16 Feb 15 FINISHED
11 16 Feb 15 FINISHED
10 16 Feb 15 FINISHED
AMPLE - Ab inito modelling for Molecular Replacement
Select the mode for AMPLE to run
Help
Job title 1ujb job
Ample User Guide
Program Mode: Ab initio modelling and Molecular Replacement
Input Files
SEQ In conc 1ujb.fa Browse View
MTZ In conc 1ujb-sf.mtz Browse View
F FP Sigma SIGFP
Free-R FREE
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 (aaXXXX_03_05.200_v1_3) none Browse
9mers (aaXXXX_09_05.200_v1_3) none Browse
Rosetta Installation
Rosetta Installation Directory /home/daniel/progs/rosetta3.4 Browse
Modelling Options
Number of models to generate: 500
Molecular Replacement Options
Molecular replacement programs to try: MOLREP PHASER
Test all generated models in MR (otherwise exit on first success)
Model Building Options
Buccaneer - automted model building cycled with refinement
Number of build-refine cycles in Buccaneer: 5
ARP/wARP - automted model building cycled with refinement
Number of auto-build cycles in ARP/wARP: 5
SHELXE - phase improvement and c-alpha tracing (requires recent version of SHELXE)
Number of tracing cycles in SHELXE: 15
Advanced Options
Enter additional command line options for AMPLE
Edit list
Add argument
Enter additional keyword options for MrBUMP
Edit list
Add keyword
Run
Save or Restore
Close
File Edit View History Bookmarks Tools Help
Robetta: full-chain protein struc...
robetta.bakerlab.org
www.bakerlab.org
ROBETTA BETA
Full-chain Protein Structure Prediction Server
REGISTRATION
[Register / Update] [Login]
DOCUMENTATION
[Docs / FAQs]
SERVICES
Domain Parsing & 3-D Modeling
[Queue] [Submit]
Interface Alanine Scanning
[Queue] [Submit]
Fragment Libraries
[Queue] [Submit]
DNA Interface Residue Scanning
[Queue] [Submit]
Model 1
Target – T0513
2.66 Å over 62 residues
0.84 Å over 39 residues
de novo prediction by Robetta in CASP-8
RELATED SITES
RosettaBackrub Server
RosettaAntibody Server
RosettaDesign Server
RosettaDock Server
Rosetta Commons
Foldit
Rosetta@home
Human Proteome Folding Project
NCSA
TeraGrid

QUARK server

AMPLE server

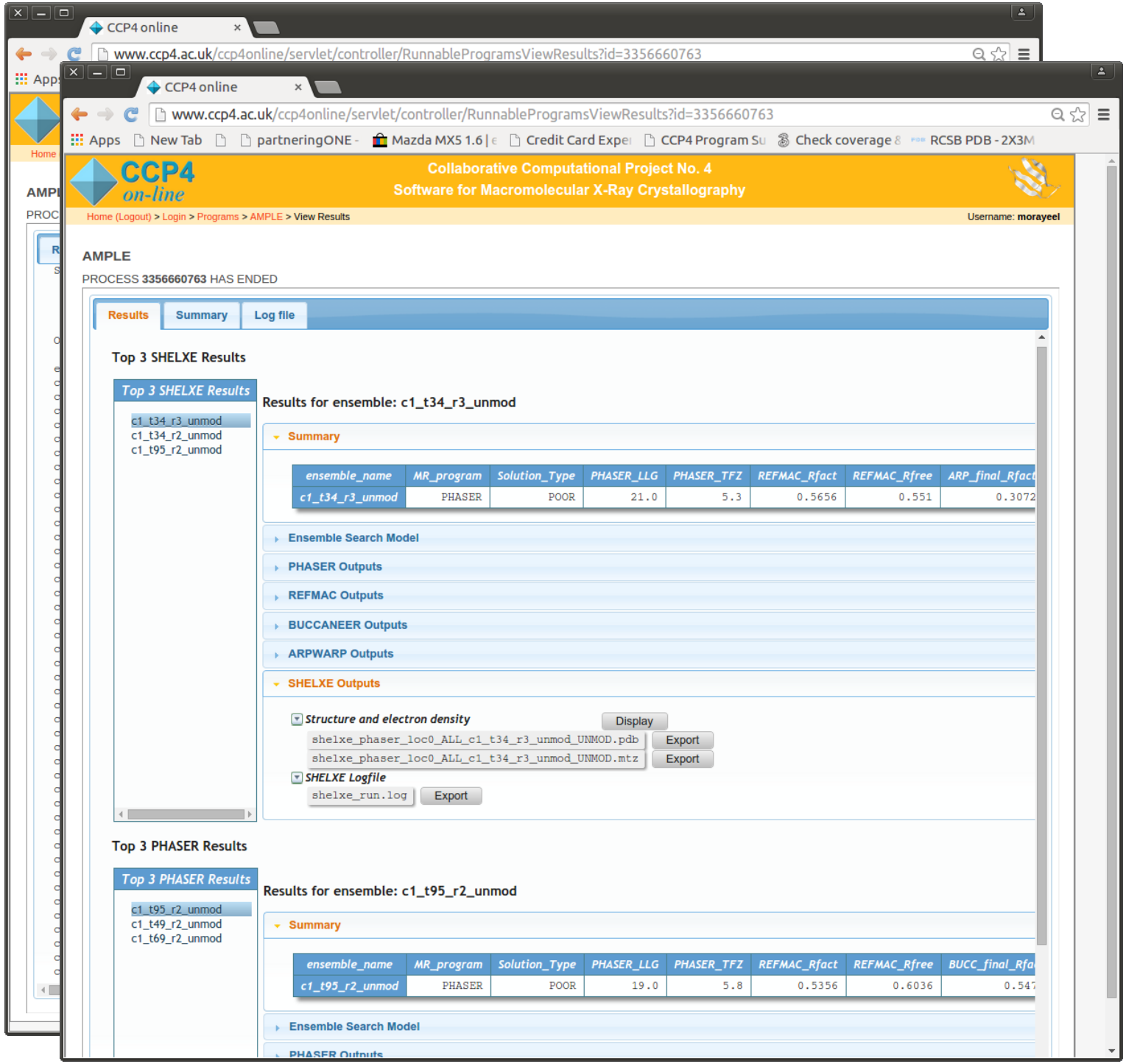




Estimated TM-score of Model 1: 0.3668 ±0.0833
Estimated TM-score of the Best of Top 10 Model: 0.4183 ±0.0764
Download the Decoy Set Used for Clustering

decoys.tar.gz

AMPLE server results



Conclusions: when to consider AMPLE

- If your target is a novel or divergent globular fold and not too large
- If your target contains a coiled-coil protein
- If you have one or more distant homologues available, but they cannot solve your target by conventional means
- If you have an NMR structure for a homologue of your target

Conclusions: how to use AMPLE

- The server is easiest for
  - QUARK *ab initio* models
  - Other model sets eg NMR structure, CONCOORD structures
- Maximum flexibility requires command line or GUI use
  - ROSETTA *ab initio* modelling with extra information eg contacts, disulphides etc
  - Single homologue truncation approaches
- Timings
  - Server 1-2 hours with pre-calculated models
  - Quark server typically takes 12 hours
  - Local running time depends on many factors! With fragments from ROBETTA server (~30 mins) a local job typically takes 12-24 hours on a multicore workstation.

Questions, perguntas?