Protein structure prediction using Phyre^2 and understanding genetic variants

Prof Michael Sternberg
Dr Lawrence Kelley
Mr Stefans Mezulis
Dr Chris Yates

Imperial College
London

BBSRC
20 Years of Pioneering
Great British Bioscience

MRC Medical Research Council

wellcome trust

Timetable

- **10.00 – 11.00 Lecture**
- **11.00 – 11.30 Tea/Coffee**
- Courtyard, West Medical Building
- **11.30 – 1.00 Hands on workshop using Phyre^2**
- Computer Cluster 515, West Medical Building

Many thanks to Glasgow Polyomics and Amy Cattanach

Glasgow Polyomics
www.polyomics.gla.ac.uk

Overview

- Methods
- Interpretation of results
- Extended functionality
- Proposed developments

- Publications:

  The Phyre2 web portal for protein modeling, prediction and analysis
  Kelley,LA, Mezulis S, Yates CM, Wass MN & Sternberg MJES
  Nature Protocols 10, 845–858 (2015)

  SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. Yates CM, Filippis I, Kelley LA, Sternberg MJE. *Journal of Molecular Biology*.;426, 2692-2701. (2014)

Phyre2

SVYDAAAQLTADVKKDLRDSW
KVIGSDKKGNGVALMTTLFAD
NQETIGYFKRLGNVSQGMAND
KLRGHSITLMYALQNFIDQLD
NPDSLDLVCS…….

Predict the 3D structure adopted by a user-supplied protein sequence

New: Log in to see the 'My account' link at the top of this page: change your password and more.

Beta release of Phyre Investigator is now live.

E-mail Address: l.a.kelley@imperial.ac.uk

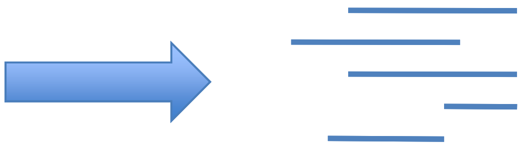
Modelling Mode: Normal ● Intensive ○

http://www.sbg.bio.ic.ac.uk/phyre2

How does Phyre2 work?

- “Normal” Mode
- “Intensive” Mode
- Advanced functions

Phyre2



ARDLVIPMIYCGHGY

User sequence

Homologous sequences

Search the 30 million known sequences for homologues using PSI-Blast.

Phyre2

ARDLVIPMIYCGHGY

User sequence

PSI-Blast

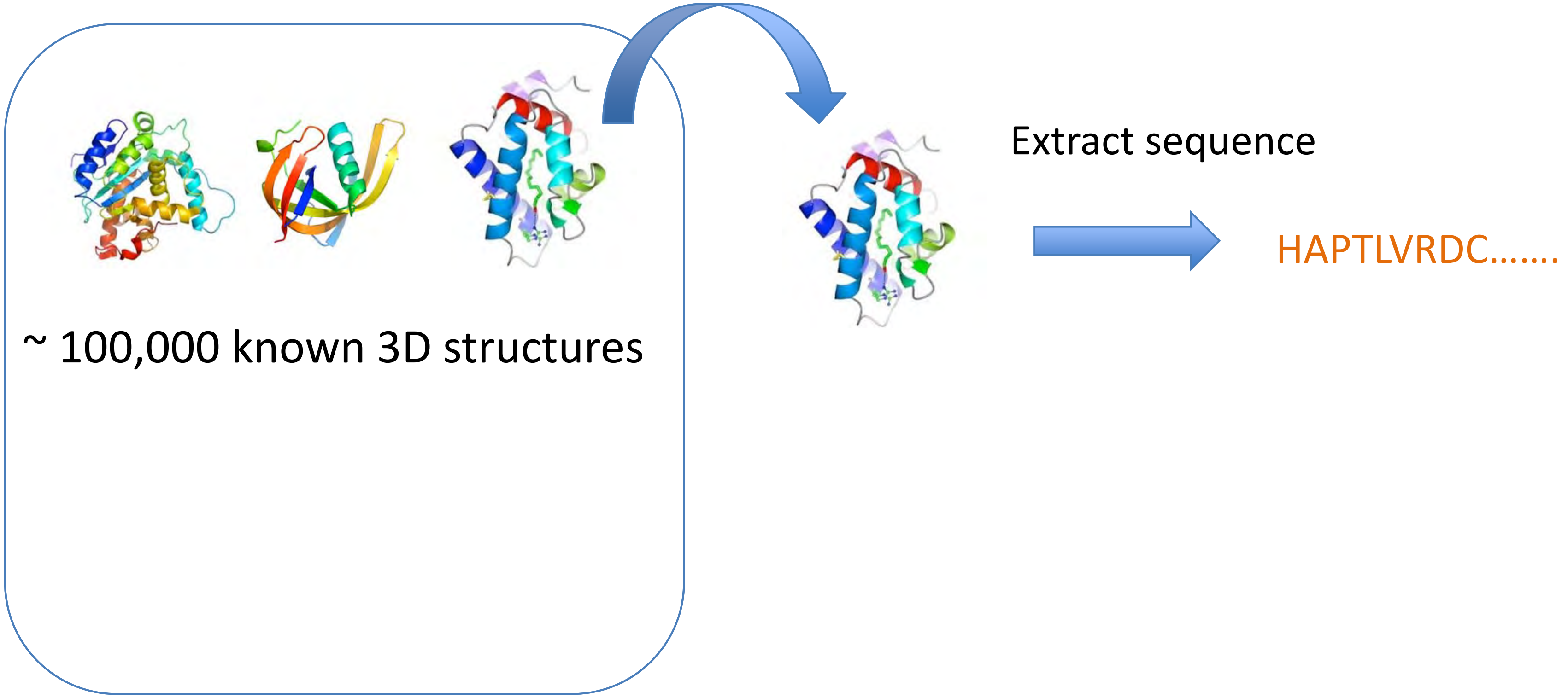
HMM

Hidden Markov model

Capture the mutational propensities at each position in the protein

An evolutionary fingerprint

Phyre2

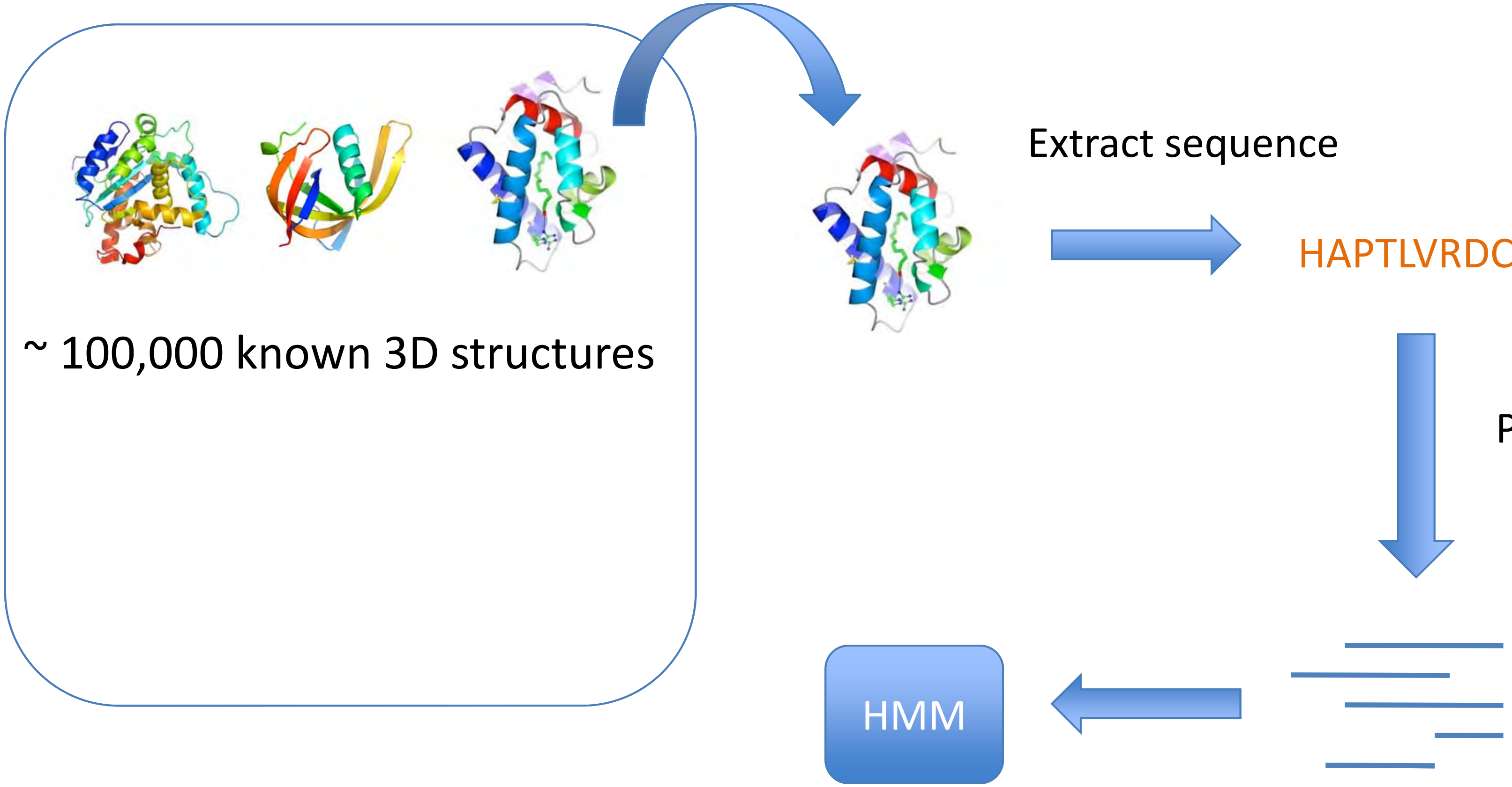


~ 100,000 known 3D structures

Extract sequence

HAPTLVRDC…….

Phyre2



~ 100,000 known 3D structures

Extract sequence

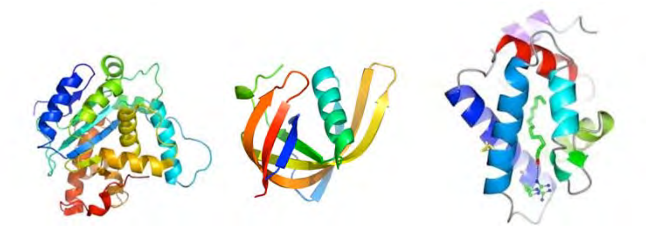
HAPTLVRDC…….

PSI-Blast

HMM

Hidden Markov model
for sequence of KNOWN structure

Phyre2

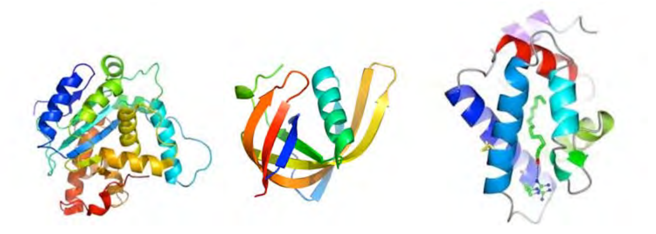


~ 100,000 known 3D structures

HMM HMM HMM

~ 100,000 hidden Markov models

Phyre2



~ 100,000 known 3D structures

Hidden Markov Model
Database of
KNOWN
STRUCTURES

Phyre2

ARDLVIPMIYCGHGY

PSI-Blast

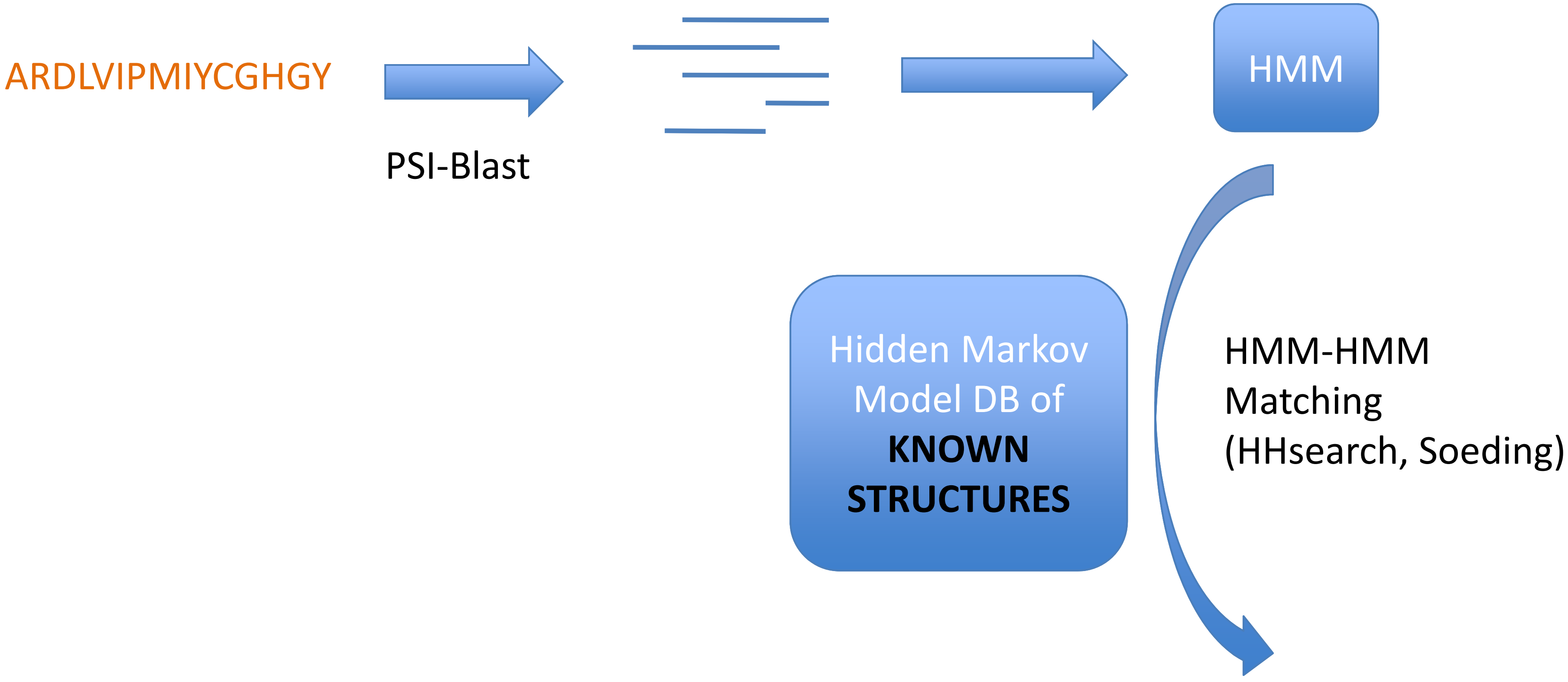
HMM

Hidden Markov model

Capture the mutational propensities at each position in the protein

An evolutionary fingerprint

Phyre2



ARDLVIPMIYCGHGY

PSI-Blast

HMM

Hidden Markov Model DB of **KNOWN STRUCTURES**

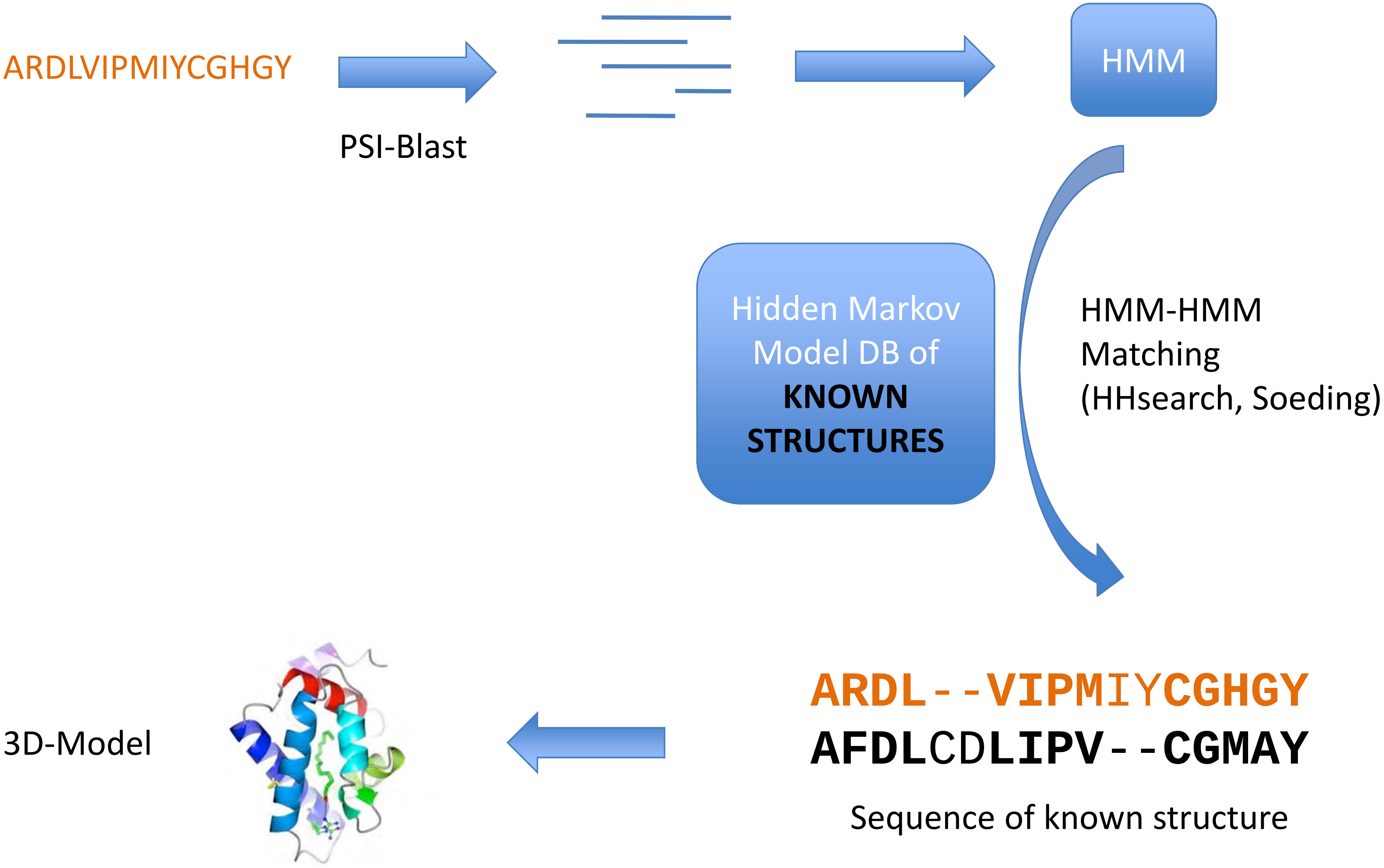
HMM-HMM Matching (HHsearch, Soeding)

Alignments of user sequence to known structures ranked by confidence.

ARDL--VIPMIYCGHGY
AFDLCDLIPV--CGMAY

Sequence of known structure

Phyre2



ARDLVIPMIYCGHGY

PSI-Blast

HMM

HMM-HMM Matching (HHsearch, Soeding)

Hidden Markov Model DB of **KNOWN STRUCTURES**

ARDL--VIPMIYCGHGY
AFDLCDLIPV--CGMAY

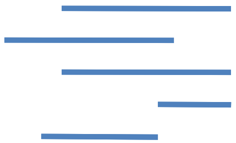
Sequence of known structure

3D-Model

Phyre2

ARDLVIPMIYCGHGY

PSI-Blast



HMM

HMM-HMM
Matching
(HHsearch, Soeding)

Hidden Markov
Model DB of
KNOWN
STRUCTURES

Very powerful –
able to reliably detect extremely
remote homology

Routinely creates accurate models even
when sequence identity is <15%

ARDL--VIPMIYCGHGY
AFDLCDLIPV--CGMAY

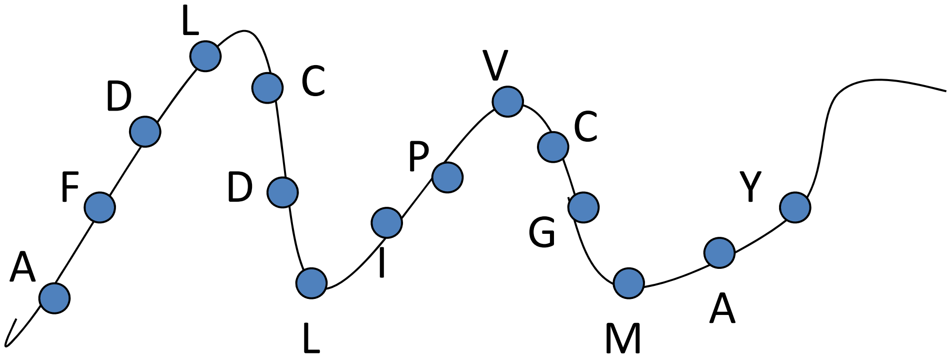
Sequence of known structure

3D-Model

From alignment to crude model

ARDL--VIPMIYCGHGY Query (your sequence)

AFDLCDLIPV--CGMAY Known Structure



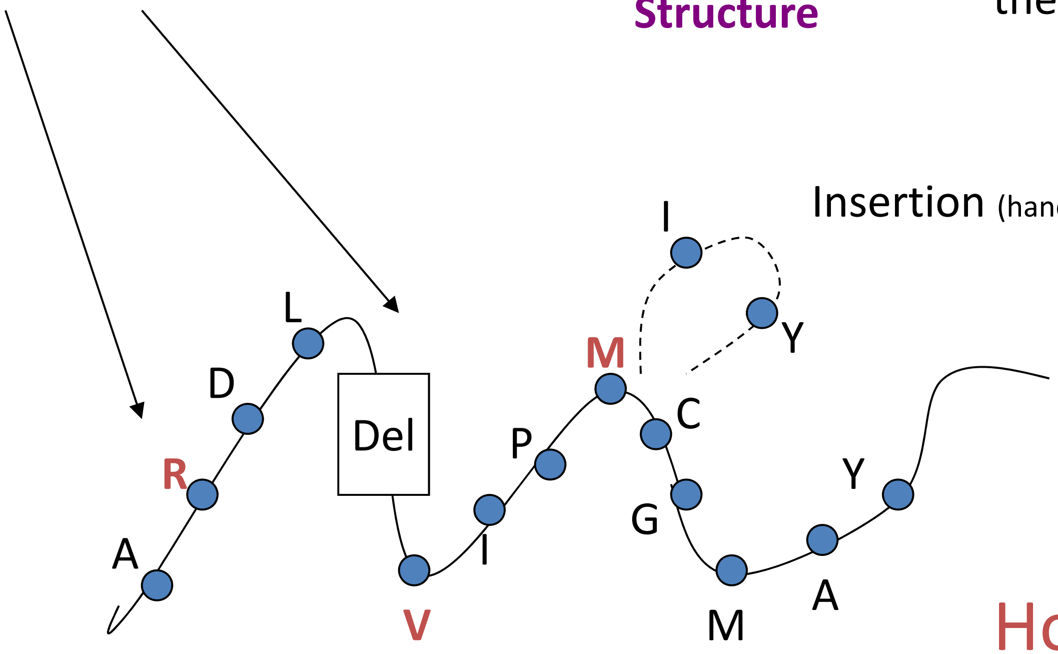
Known 3D
Structure coordinates

From alignment to crude model

ARDL--VIPMIYCGHGY **Query**

AFDLCDLIPV--CGMAY **Known Structure**

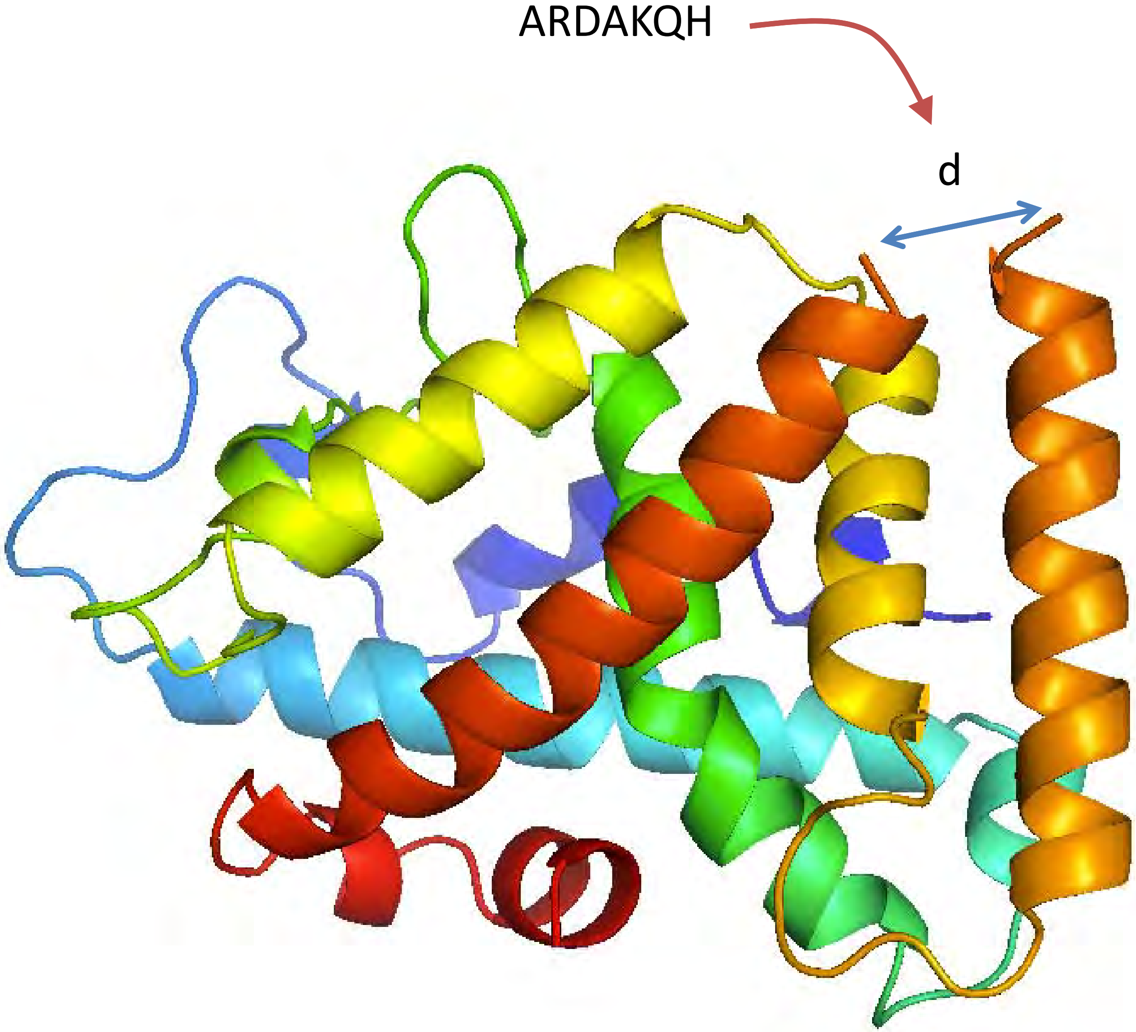
Re-label the known structure according to the mapping from the alignment.



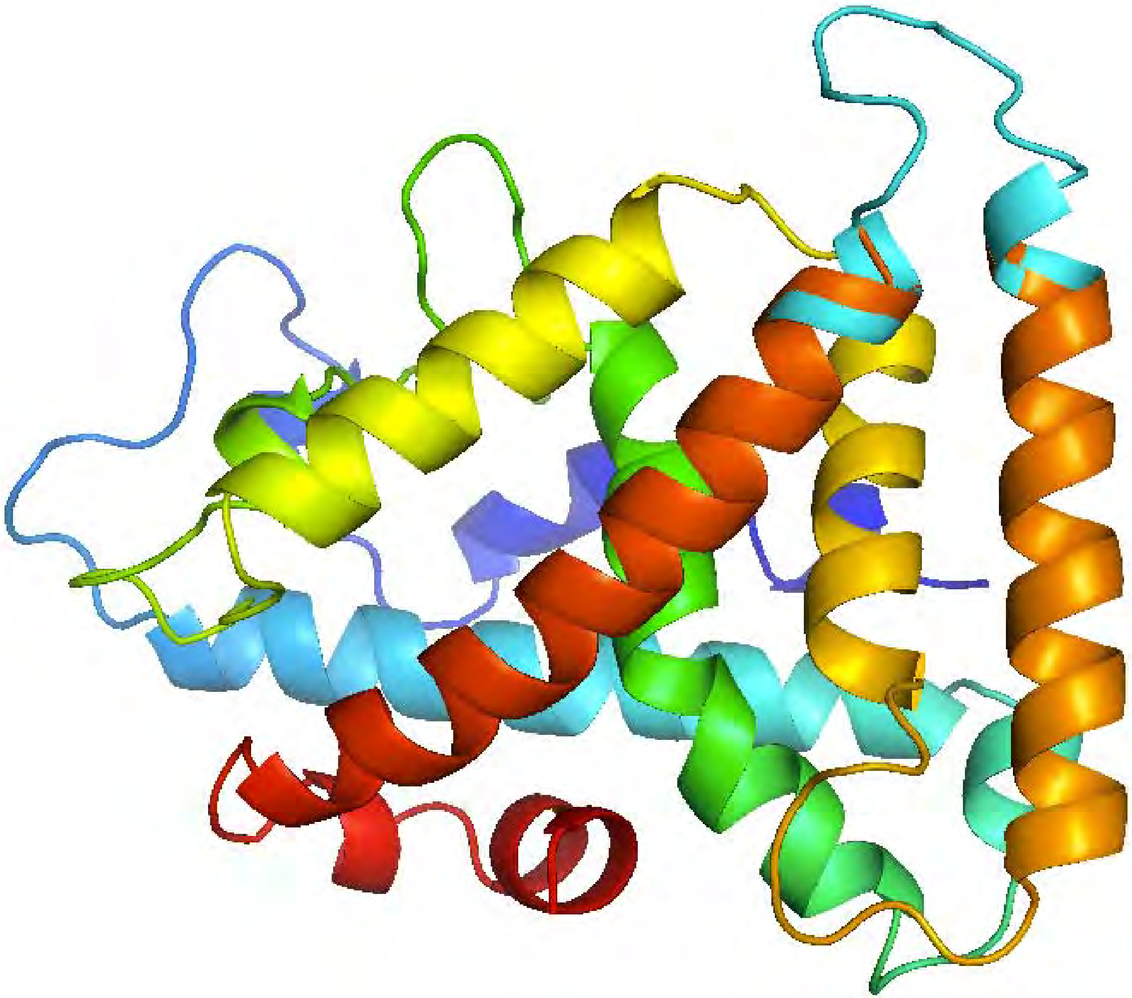
Insertion (handled by loop modelling)

Homology model

Loop modelling



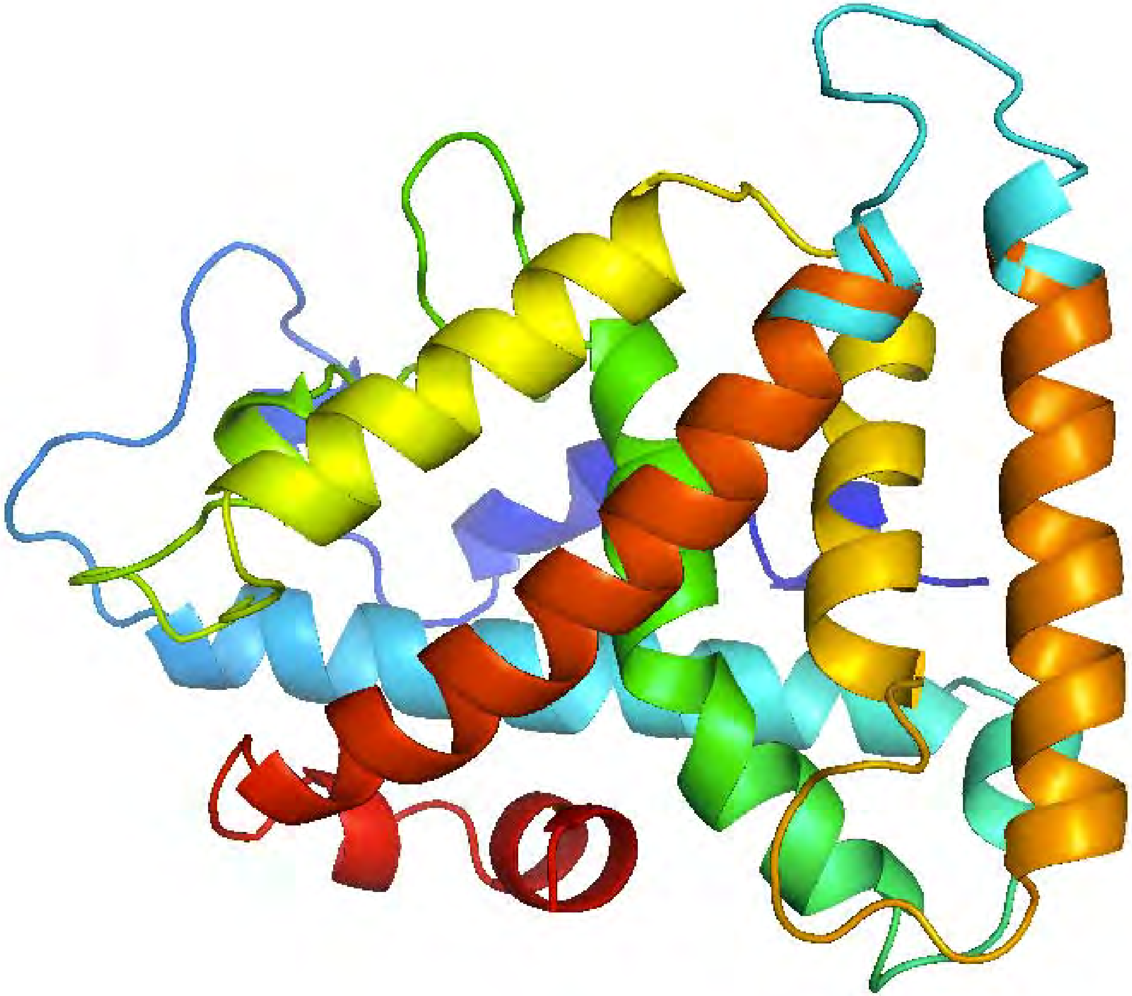
Loop modelling



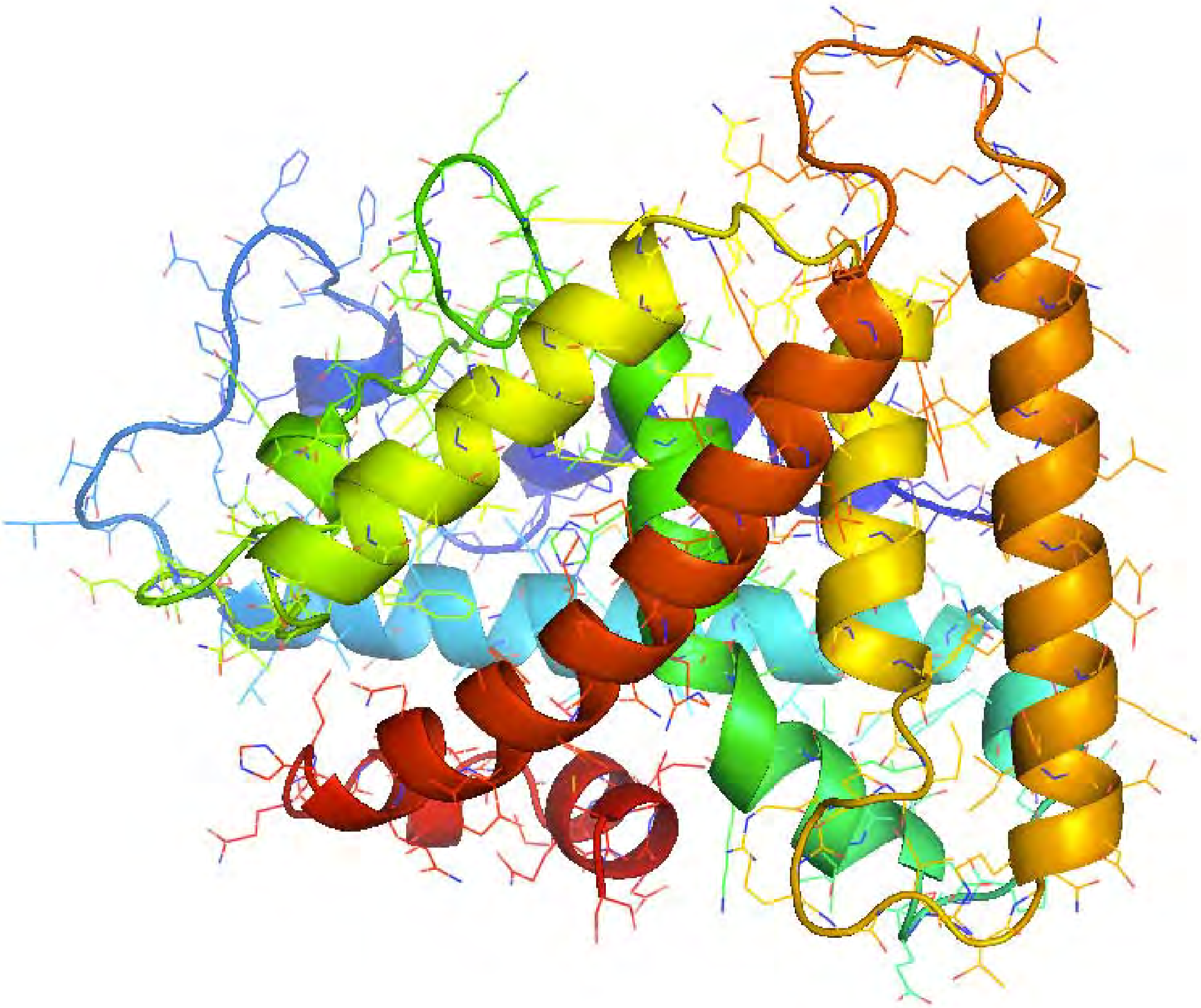
Loop modelling

- Insertions and deletions relative to template modelled by a loop library up to 15 aa’s in length
- Short loops (<=5) good. Longer loops less trustworthy
- Be wary of basing any interpretation of the **structural** effects of point mutations

Sidechain modelling



Sidechain modelling



Sidechain modelling

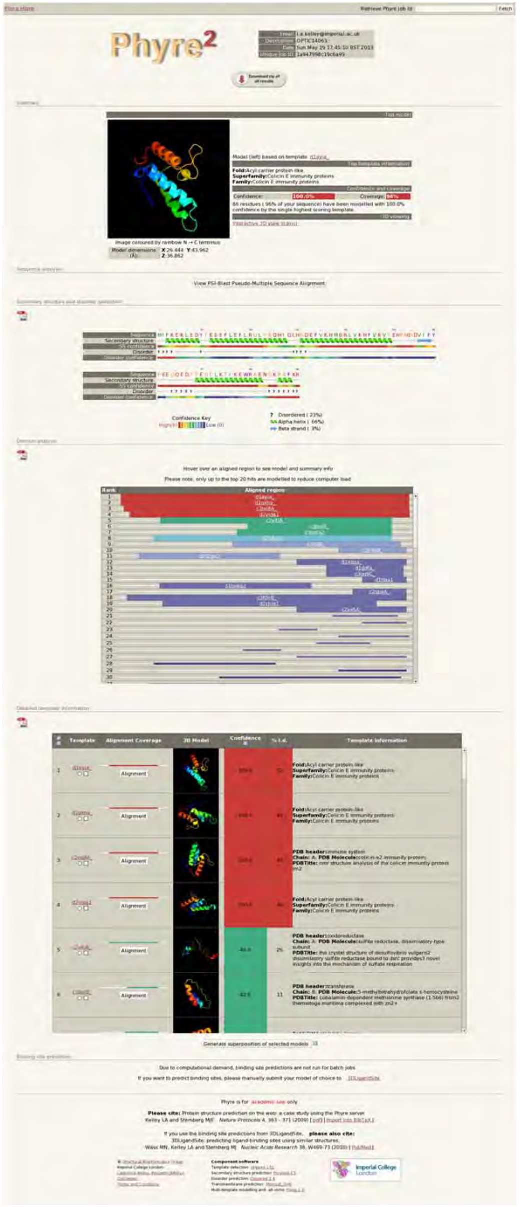
Optimisation problem

- Fit most probable rotamer at each position
- According to given backbone angles
- Whilst avoiding clashes

Sidechain modelling

- Sidechains will be modelled with ~80% accuracy IF……the backbone is correct.
- Clashes *will* sometimes occur and if frequent, indicate probably a wrong alignment or poor template
- Analyse with Phyre Investigator

Example results



Retrieve Phyre job id Fetch

Phyre²

Email l.a.kelley@imperial.ac.uk
Description OPTIC14063
Date Sun May 19 17:45:10 BST 2013
Unique Job ID 1a947998c19c6a99

Model (left) based on template d1ayia_

Fold: Acyl carrier protein-like
Superfamily: Colicin E immunity proteins
Family: Colicin E immunity proteins

Confidence: 100.0% Coverage: 96%

86 residues (96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Image coloured by rainbow N → C terminus

Model dimensions (Å): X:26.444 Y:43.962 Z:36.862

View PSI-Blast Pseudo-Multiple Sequence Alignment

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	[illegible]	Alignment		[illegible]	[illegible]	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
2	[illegible]	Alignment		[illegible]	[illegible]	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
3	[illegible]	Alignment		[illegible]	[illegible]	**PDB header:**immune system **Chain:** A: **PDB Molecule:**colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immunity protein im2
4	[illegible]	Alignment		[illegible]	[illegible]	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
5	[illegible]	Alignment		[illegible]	25	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**sulfite reductase, dissimilatory-type subunit **PDBTitle:** the crystal structure of desulfovibrio vulgaris dissimilatory sulfite reductase bound to dsrc provides novel insights into the mechanism of sulfate respiration
6	[illegible]	Alignment		[illegible]	11	**PDB header:**transferase **Chain:** B: **PDB Molecule:**5-methyltetrahydrofolate s-homocysteine **PDBTitle:** cobalamin-dependent methionine synthase (1-566) from thermotoga maritima complexed with zn2+

Generate superposition of selected models

Due to computational demand, binding site predictions are not run for batch jobs

If you want to predict binding sites, please manually submit your model of choice to 3DLigandSite

Phyre is for academic use only

Please cite: Protein structure prediction on the web: a case study using the Phyre server
Kelley LA and Sternberg MJE *Nature Protocols* 4, 363 - 371 (2009) [pdf] [Import into BibTeX]

If you use the binding site predictions from 3DLigandSite, **please also cite:**
3DLigandSite: predicting ligand-binding sites using similar structures.
Wass MN, Kelley LA and Sternberg MJ *Nucleic Acids Research* 38, W469-73 (2010) [PubMed]

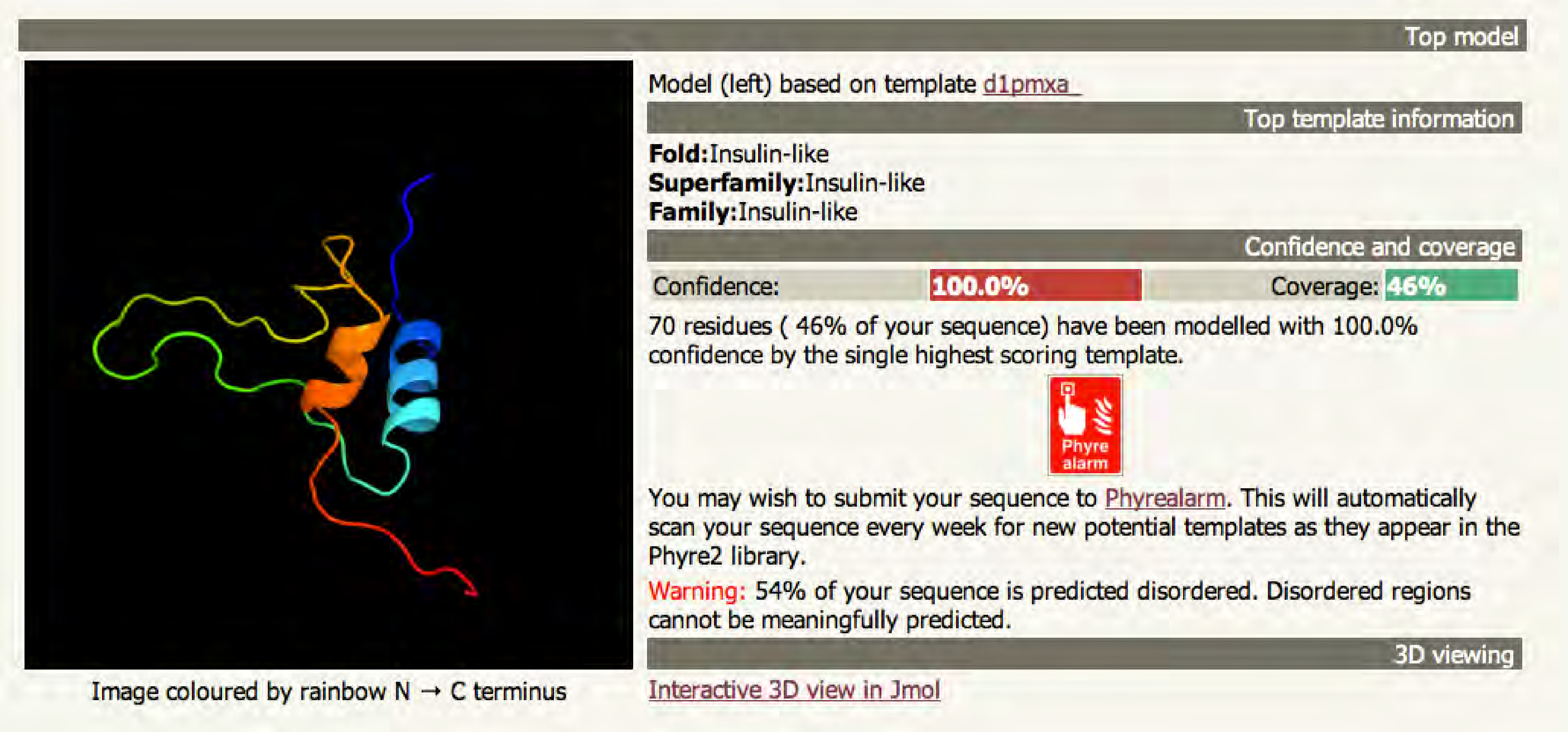
Top model info

Secondary structure/disorder

Domain analysis

Detailed template information

Example results



Top model

Model (left) based on template d1pmxa

Top template information

Fold: Insulin-like
Superfamily: Insulin-like
Family: Insulin-like

Confidence and coverage

Confidence: 100.0% Coverage: 46%

70 residues (46% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

You may wish to submit your sequence to Phyrealarm. This will automatically scan your sequence every week for new potential templates as they appear in the Phyre2 library.

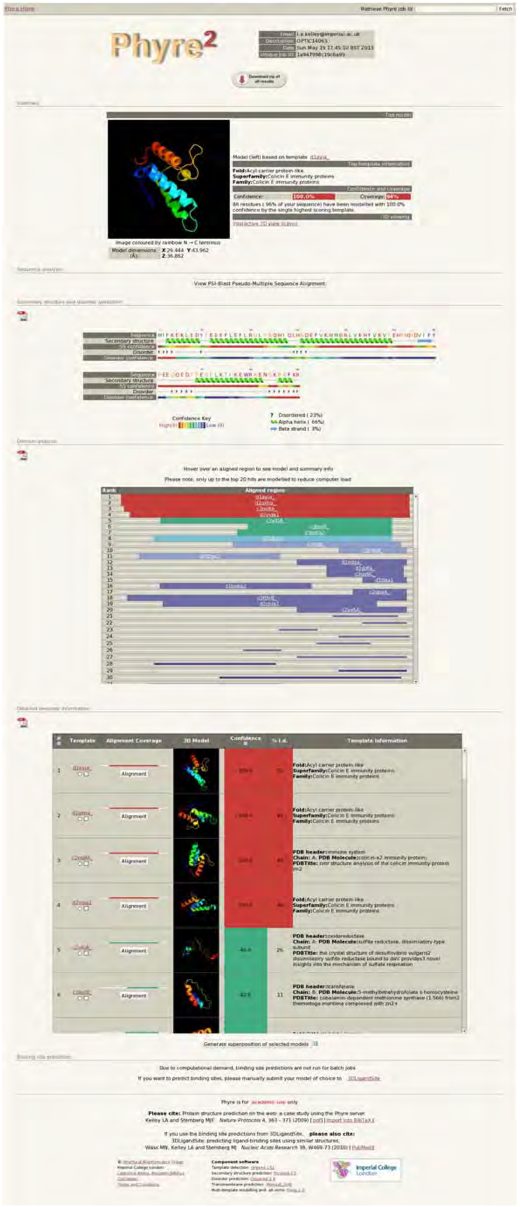
Warning: 54% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

3D viewing

Interactive 3D view in Jmol

Image coloured by rainbow N → C terminus

Example results



Phyre2

Date: Sun May 19 17:45:10 BST 2013

Model (left) based on template d1ayia_

Fold: Acyl carrier protein-like
Superfamily: Colicin E immunity proteins
Family: Colicin E immunity proteins

Confidence: 100.0% Coverage: 96%

86 residues (96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Image coloured by rainbow N → C terminus

Model dimensions (Å): X:29.444 Y:43.962 Z:36.862

View PSI-Blast Pseudo-Multiple Sequence Alignment

Top model info

Secondary structure/disorder

Domain analysis

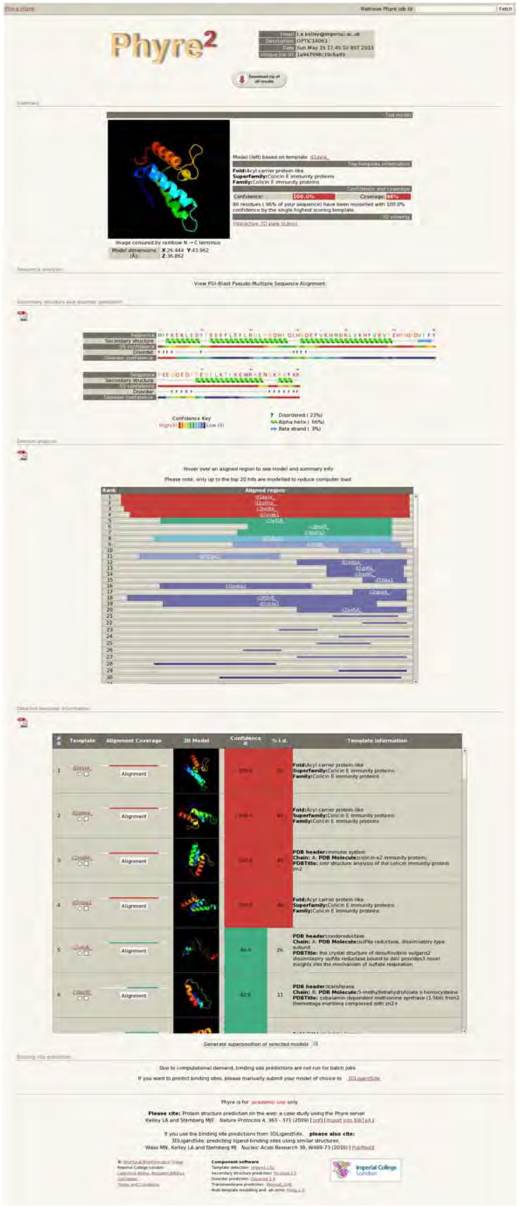
Detailed template information

Example SS/disorder prediction

Secondary structure and disorder

- Based on neural networks trained on known structures.
- *Given a diverse set of homologous sequences*, expect ~75-80% accuracy.
- Few or no homologous sequences? Only 60-62% accuracy

Example results



Date: Sun May 19 17:45:10 BST 2013

Model (left) based on template

Fold: Acyl carrier protein-like
Superfamily: Colicin E immunity proteins
Family: Colicin E immunity proteins

Confidence: 100.0% Coverage: 96%

86 residues (96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Image coloured by rainbow N → C terminus

Model dimensions (Å): X:26.444 Y:43.962 Z:36.862

View PSI-Blast Pseudo-Multiple Sequence Alignment

Confidence Key

? Disordered (23%)
Alpha helix (66%)
Beta strand (3%)

Hover over an aligned region to see model and summary info

Please note, only up to the top 20 hits are modelled to reduce computer load

Generate superposition of selected models

Due to computational demand, binding site predictions are not run for batch jobs

If you want to predict binding sites, please manually submit your model of choice to 3DLigandSite

Phyre is for academic use only

Please cite: Protein structure prediction on the web: a case study using the Phyre server

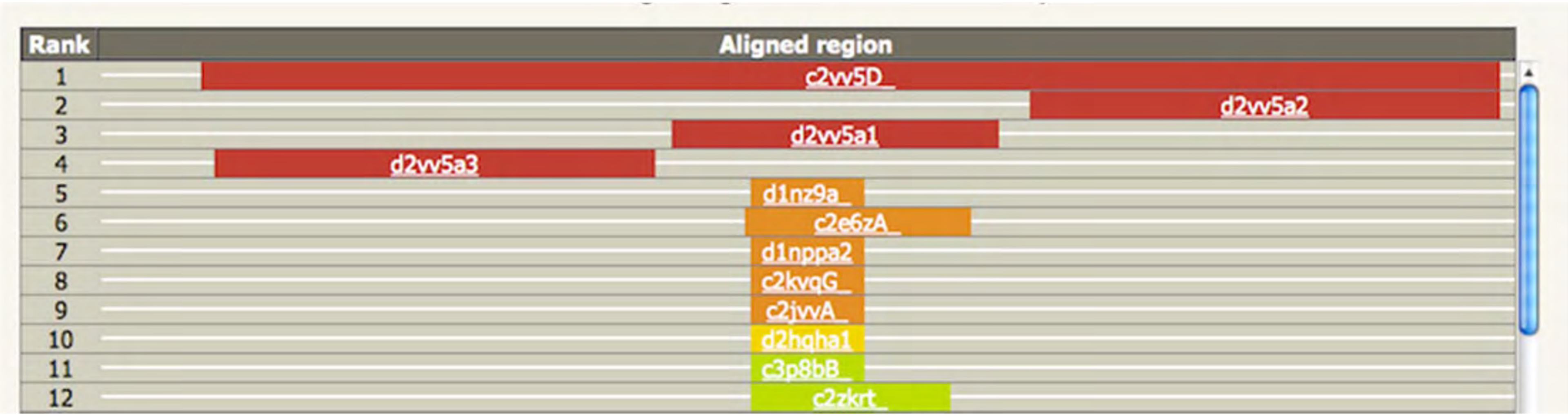
Top model info

Secondary structure/disorder

Domain analysis

Detailed template information

Example domain analysis

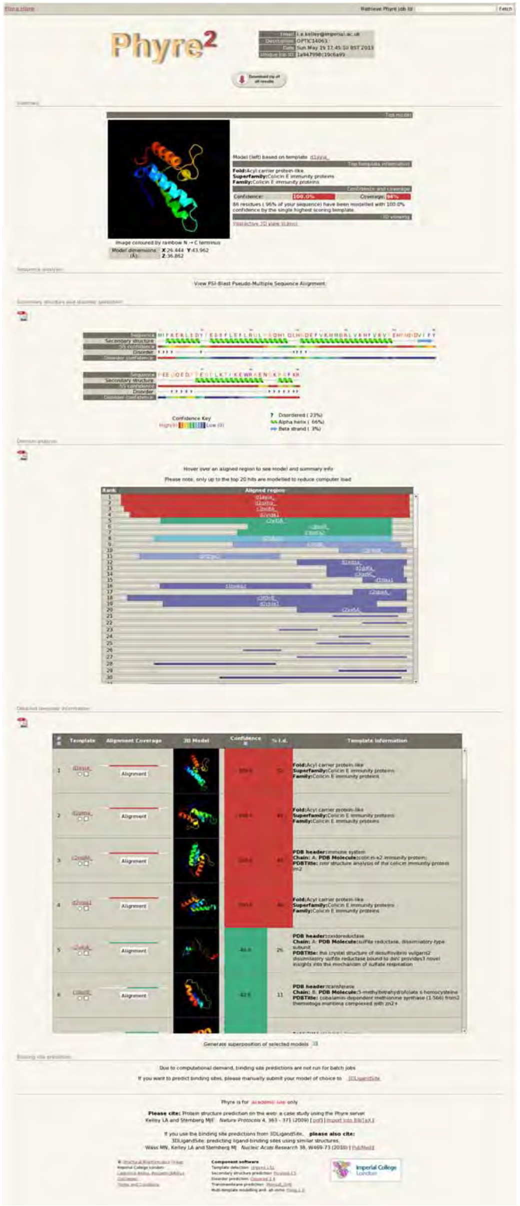


Rank	Aligned region
1	c2vv5D_
2	d2vv5a2
3	d2vv5a1
4	d2vv5a3
5	d1nz9a_
6	c2e6zA_
7	d1nppa2
8	c2kvqG_
9	c2jvvA_
10	d2hqha1
11	c3p8bB_
12	c2zkrt_

Domain analysis

- Local hits to different templates indicate domain structure of your protein
- Multiple domains can be linked using ‘Intensive mode’

Example results



Retrieve Phyre job id Fetch

Email: l.a.kelley@imperial.ac.uk
Description: OPTIC14063
Date: Sun May 19 17:45:10 BST 2013
Unique Job ID: 1a947998c19c6a99

Model (left) based on template d1ayia_

Fold: Acyl carrier protein-like
Superfamily: Colicin E immunity proteins
Family: Colicin E immunity proteins

Confidence: 100.0% Coverage: 96%

86 residues (96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Interactive 3D view in Jmol

Image coloured by rainbow N → C terminus

Model dimensions (Å): X:29.444 Y:43.962 Z:36.862

View PSI-Blast Pseudo-Multiple Sequence Alignment

Confidence Key
High(9) Low (0)

? Disordered (23%)
Alpha helix (66%)
Beta strand (3%)

Hover over an aligned region to see model and summary info

Please note, only up to the top 20 hits are modelled to reduce computer load

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	[illegible]	Alignment		[illegible]	[illegible]	**Fold:** Acyl carrier protein-like **Superfamily:** Colicin E immunity proteins **Family:** Colicin E immunity proteins
2	[illegible]	Alignment		[illegible]	[illegible]	**Fold:** Acyl carrier protein-like **Superfamily:** Colicin E immunity proteins **Family:** Colicin E immunity proteins
3	[illegible]	Alignment		[illegible]	[illegible]	**PDB header:** immune system **Chain:** A: **PDB Molecule:** colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immunity protein im2
4	[illegible]	Alignment		[illegible]	[illegible]	**Fold:** Acyl carrier protein-like **Superfamily:** Colicin E immunity proteins **Family:** Colicin E immunity proteins
5	[illegible]	Alignment		[illegible]	25	**PDB header:** oxidoreductase **Chain:** A: **PDB Molecule:** sulfite reductase, dissimilatory-type subunit **PDBTitle:** the crystal structure of desulfovibrio vulgaris2 dissimilatory sulfite reductase bound to dsrc provides3 novel insights into the mechanism of sulfate respiration
6	[illegible]	Alignment		[illegible]	11	**PDB header:** transferase **Chain:** B: **PDB Molecule:** 5-methyltetrahydrofolate s-homocysteine **PDBTitle:** cobalamin-dependent methionine synthase (1-566) from2 thermotoga maritima complexed with zn2+

Generate superposition of selected models

Due to computational demand, binding site predictions are not run for batch jobs

If you want to predict binding sites, please manually submit your model of choice to 3DLigandSite

Phyre is for academic use only

Please cite: Protein structure prediction on the web: a case study using the Phyre server
Kelley LA and Sternberg MJE. *Nature Protocols* 4, 363 - 371 (2009) [pdf] [Import into BibTeX]

If you use the binding site predictions from 3DLigandSite, **please also cite:**
3DLigandSite: predicting ligand-binding sites using similar structures.
Wass MN, Kelley LA and Sternberg MJ. *Nucleic Acids Research* 38, W469-73 (2010) [PubMed]

Component software

Imperial College London

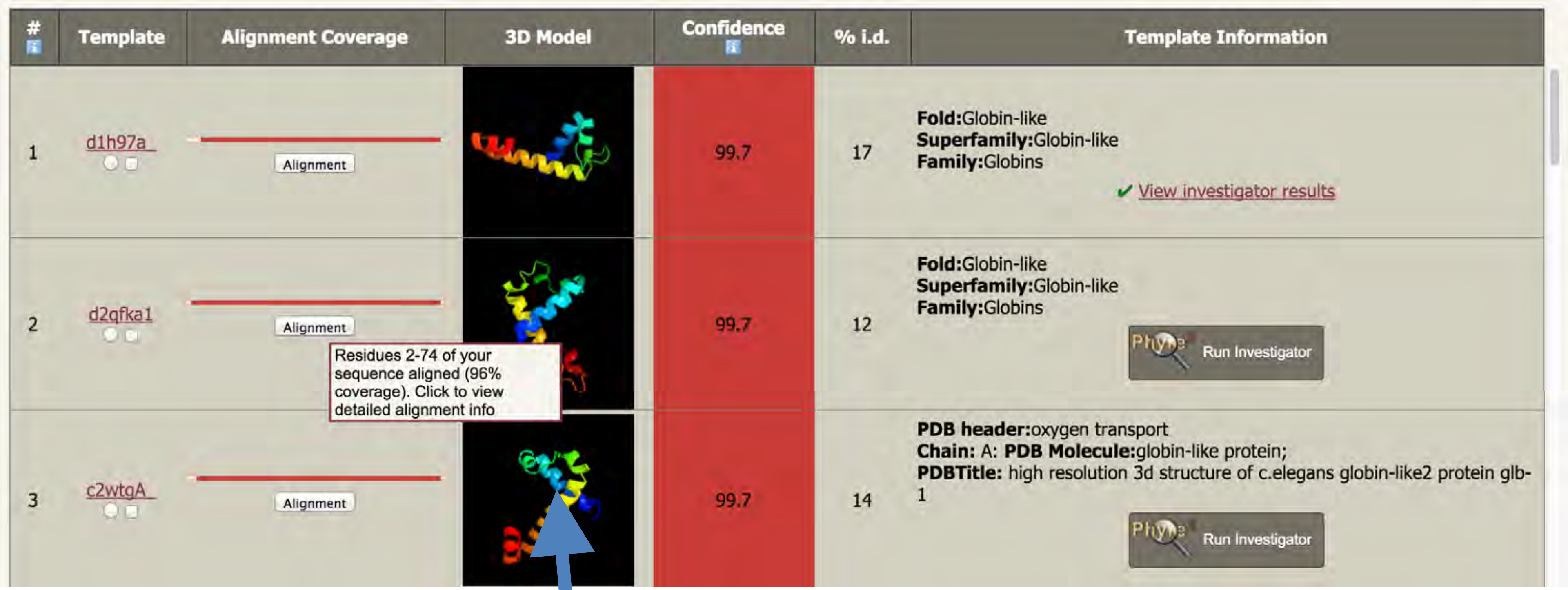
Top model info

Secondary structure/disorder

Domain analysis

Detailed template information

Main results table



#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1h97a_ ○ ☐	Alignment		99.7	17	**Fold:**Globin-like **Superfamily:**Globin-like **Family:**Globins ✔ View investigator results
2	d2qfka1 ○ ☐	Alignment		99.7	12	**Fold:**Globin-like **Superfamily:**Globin-like **Family:**Globins Phyre Run Investigator
3	c2wtgA_ ○ ☐	Alignment		99.7	14	**PDB header:**oxygen transport **Chain:** A: **PDB Molecule:**globin-like protein; **PDBTitle:** high resolution 3d structure of c.elegans globin-like2 protein glb-1 Phyre Run Investigator

Residues 2-74 of your sequence aligned (96% coverage). Click to view detailed alignment info

Actual Model!

Not just a picture of the template – click to download model

Interpreting results

How accurate is my model?

- Simple question with a complicated answer!
- RMSD very commonly used, but often misleading
- Modelling community uses **TM score** for benchmarking: essentially the percentage of alpha carbons superposable on the answer within 3.5Å. Prediction of TM-score coming soon.
- Focused on the protein core, rather than loops and sidechains.

Interpreting results

- MAIN POINT: The confidence estimate provided by Phyre2 is NOT a direct indication of model quality – though it is related…
- It is a measure of the **likelihood of homology**
- Model quality can now be assessed using the new Phyre Investigator (more later)
- New measure of model quality coming soon..

Interpreting results

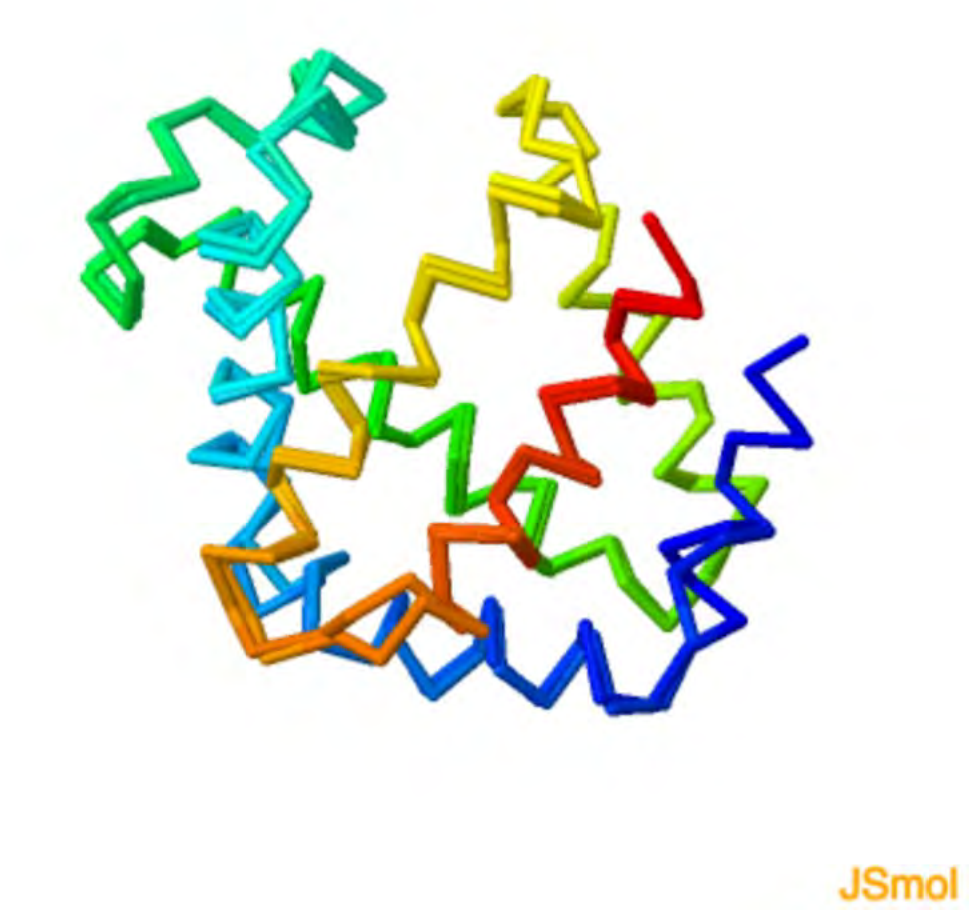
Sequence identity and model accuracy

- High confidence (>90%) and High seq. id. (>35%): almost always very accurate: TM score>0.7, RMSD 1-3Å
- High confidence (>90%) and low seq. id. (<30%) almost certainly the correct fold, accurate in the core (2-4Å) but may show substantial deviations in loops and non-core regions.

Interpreting results

100% confidence,

56% sequence identity, TM-score 0.9



Interpreting results

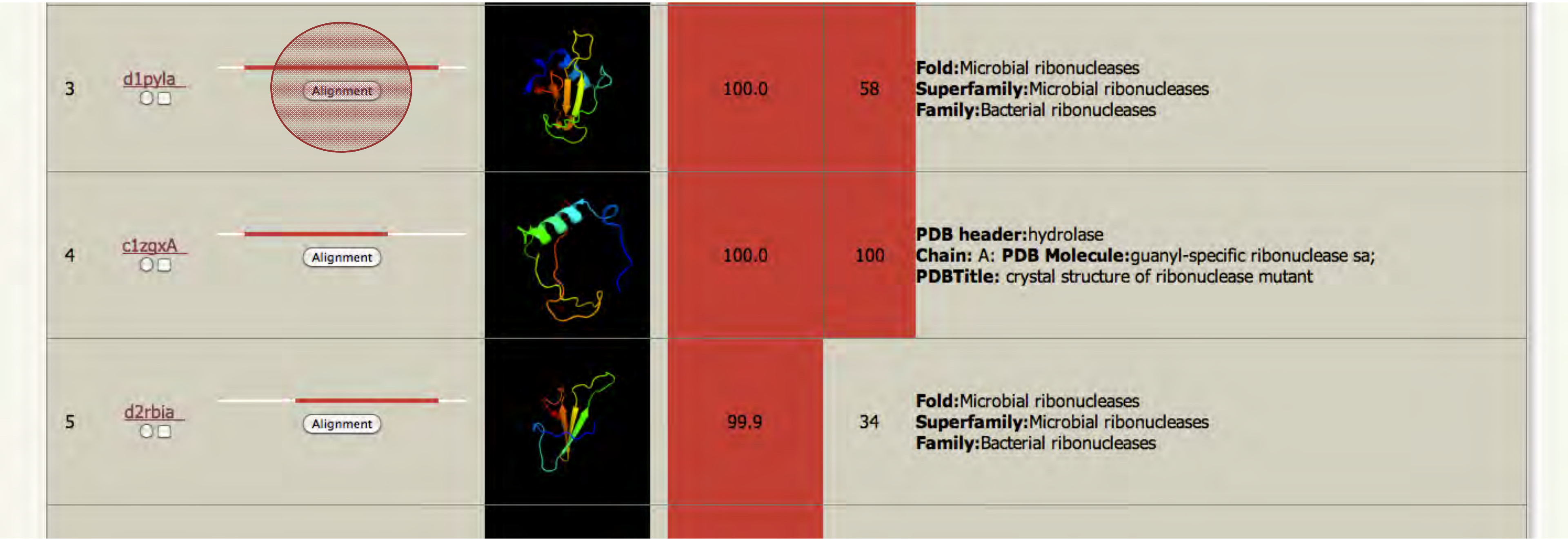
100% confidence,
24% sequence identity, TM-score 0.8

Interpreting results

Checklist

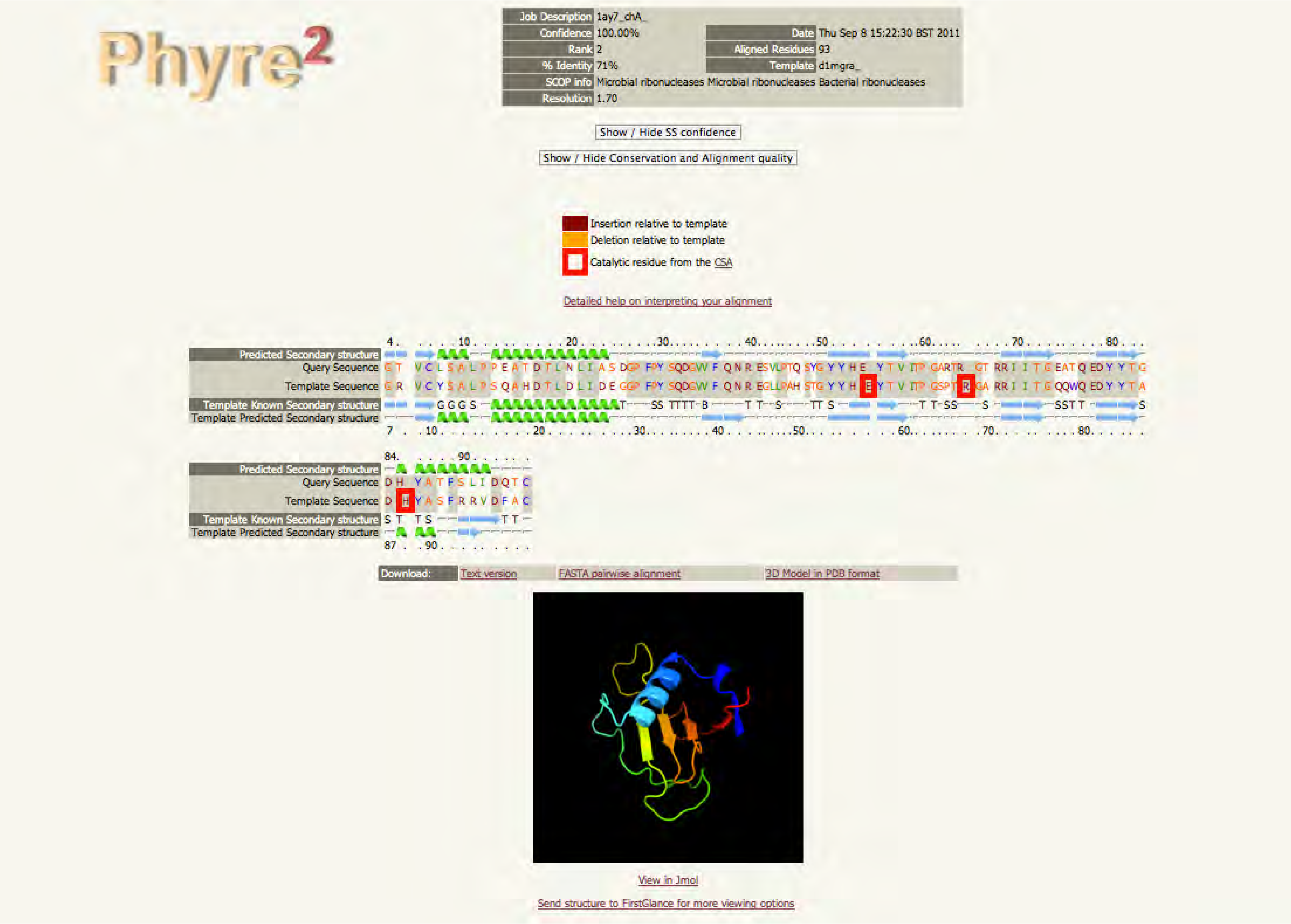
- Look at confidence
- Given multiple high confidence hits, look at % sequence identity
- Biological knowledge relating function of template to sequence of interest
- Structural superpositions to compare models – many similar models increase confidence
- Examine sequence alignment

Main results table



3	d1pyla ○ ☐	Alignment		100.0	58	**Fold:**Microbial ribonucleases **Superfamily:**Microbial ribonucleases **Family:**Bacterial ribonucleases
4	c1zgxA ○ ☐	Alignment		100.0	100	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**guanyl-specific ribonuclease sa; **PDBTitle:** crystal structure of ribonuclease mutant
5	d2rbia ○ ☐	Alignment		99.9	34	**Fold:**Microbial ribonucleases **Superfamily:**Microbial ribonucleases **Family:**Bacterial ribonucleases

Alignment view



Job Description	1ay7_chA_		
Confidence	100.00%	Date	Thu Sep 8 15:22:30 BST 2011
Rank	2	Aligned Residues	93
% Identity	71%	Template	d1mgra_
SCOP info	Microbial ribonucleases Microbial ribonucleases Bacterial ribonucleases		
Resolution	1.70		

🟥 Insertion relative to template
🟧 Deletion relative to template
☐ Catalytic residue from the CSA

Detailed help on interpreting your alignment

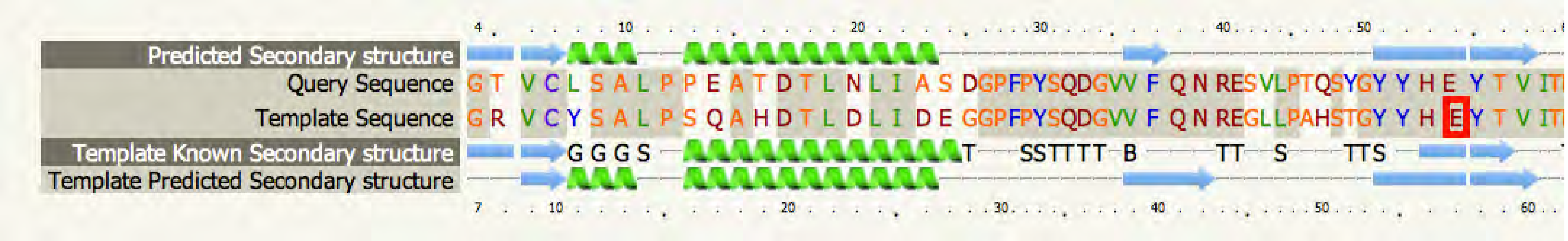
```
                                      4 .   . . . . 10 . . . . . . . . . 20 . . . . . . . . 30. . . . . . . . . 40. . . . . . . . 50 . . . . . . . . 60. . . .   . . . . 70 . . . . . . . . . 80 . . .
     Predicted Secondary structure
                    Query Sequence    G T   V C L S A L P P E A T D T L N L I A S DGP FPY SQDGW F Q N R ESVLPTQ SYG Y Y H E   Y T V ITP GARTR   GT RR I I T G EAT Q ED Y Y T G
                 Template Sequence    G R   V C Y S A L P S Q A H D T L D L I D E GGP FPY SQDGW F Q N R EGLLPAH STG Y Y H E Y T V ITP GSPT R GA RR I I T G QQWQ ED Y Y T A
Template Known Secondary structure            G G G S                          T    SS TTTT B       T T  S     TT S                T T SS    S        SSTT              S
Template Predicted Secondary structure
                                      7 .   . 10 . . . . . . . . . 20 . . . . . . . . . 30. . . . . . . . 40 . . . . . . . . 50. . . . . . . . . 60. . . . . . . . 70. . . . . . . . 80. . . . . .

                                      84.   . . . . 90 . . . . .
     Predicted Secondary structure
                    Query Sequence    D H   Y A T F S L I D Q T C
                 Template Sequence    D H   Y A S F R R V D F A C
Template Known Secondary structure    S T   T S              T T
Template Predicted Secondary structure
                                      87 .  . 90 . . . . . . . . .
```

Download: Text version | FASTA pairwise alignment | 3D Model in PDB format

View in Jmol

Send structure to FirstGlance for more viewing options

Alignment view



Predicted Secondary structure

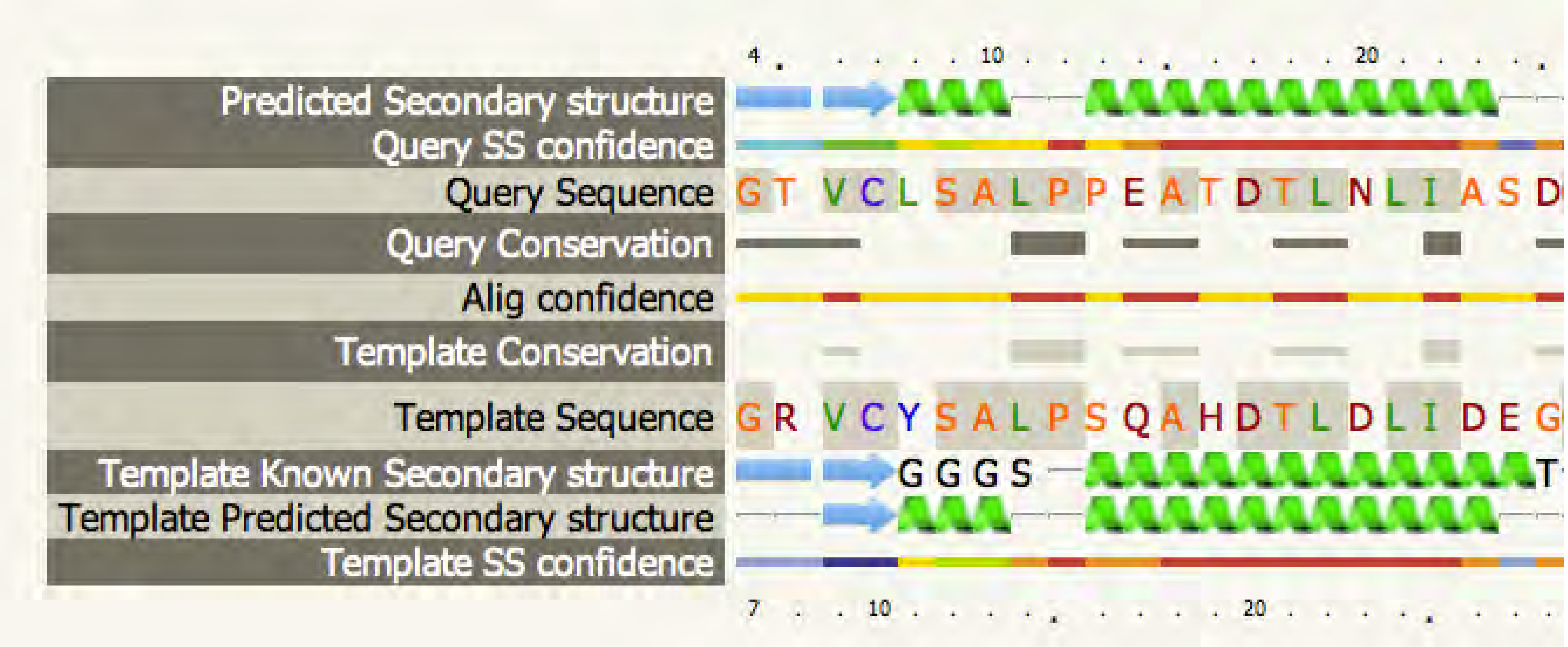
Query Sequence GT VCLSALPPEATDTLNLIASDGPFPYSQDGVVFQNRESVLPTQSYGYYHE YTVIT

Template Sequence GR VCYSALPSQAHDTLDLIDEGGPFPYSQDGVVFQNREGLLPAHSTGYYHEYTVIT

Template Known Secondary structure GGGS T SSTTTT B TT S TTS

Template Predicted Secondary structure

Alignment view



Alignment interpretation

Checklist

- Secondary structure matches
- Gaps in SS elements indicate potentially wrong alignment
- Active sites present in the Catalytic Site Atlas (CSA) for the template highlighted – look for identity or conservative mutations when transferring function
- Alignment confidence per residue

Mutations

- The STRUCTURAL effects of point mutations on structure will NOT be modelled accurately

Checklist

- Is it near the active site?
- Is it a change in the hydrophobic core?
- Is it near a known binding site? (can predict with e.g. 3DLigandSite)
- Phyre Investigator can help (see later)

Is my model good enough?

All depends on your purpose.

- Good enough for drug design? – probably if the sequence identity is very high (>50%)
- Sometimes good enough if far lower seq id but accurate around site of interest.
- High confidence but low seq i.d. still very likely correct fold, useful for a range of tasks.

How does Phyre2 work?

- “Normal” Mode
- “Intensive” Mode
- Advanced functions

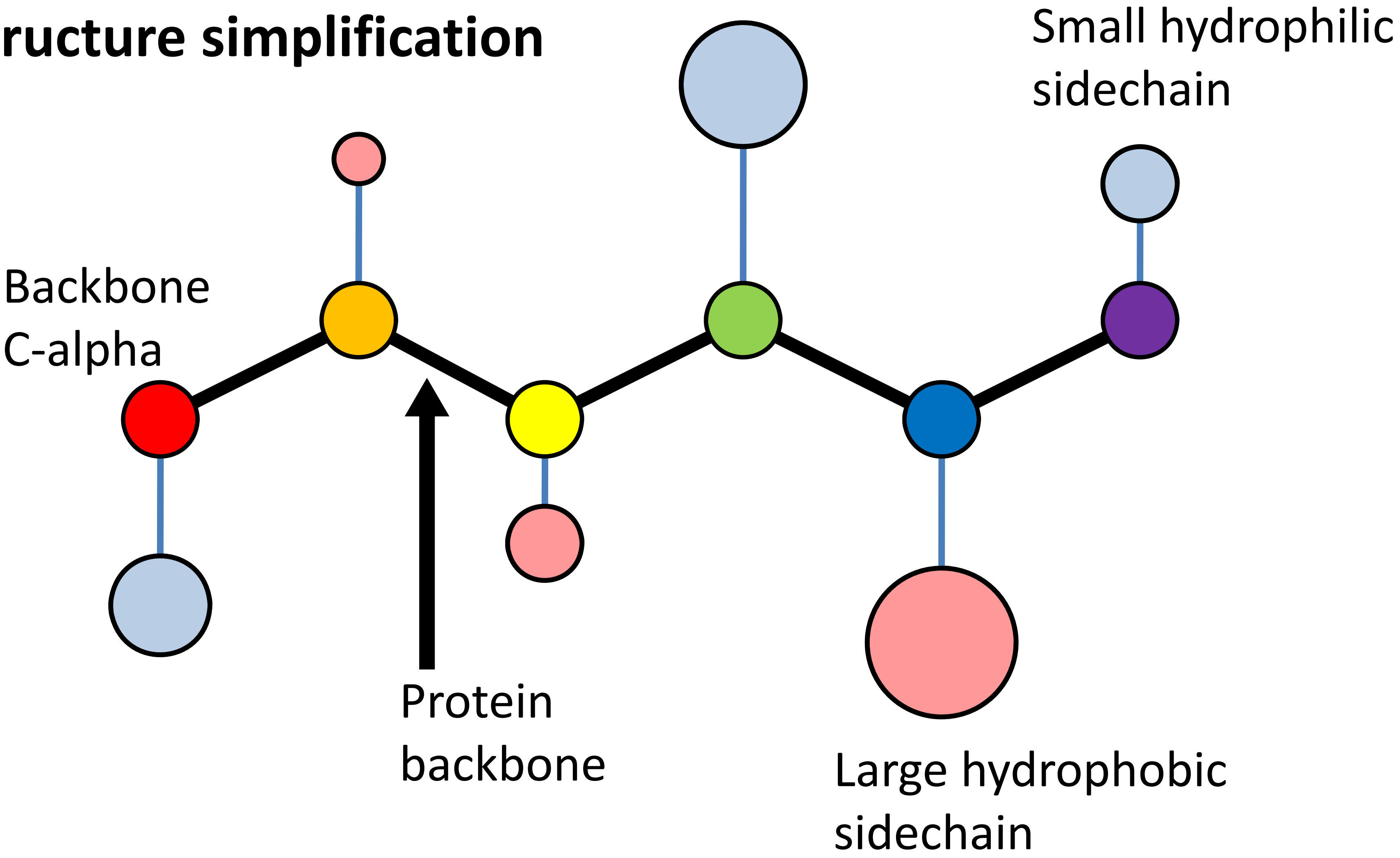
Shortcomings of ‘normal’ Mode

- Individual domains in multi-dom proteins often modelled separately
- Regions with no detectable homology to known structure unmodelled
- Does not use multiple templates which, when combined could result in better coverage

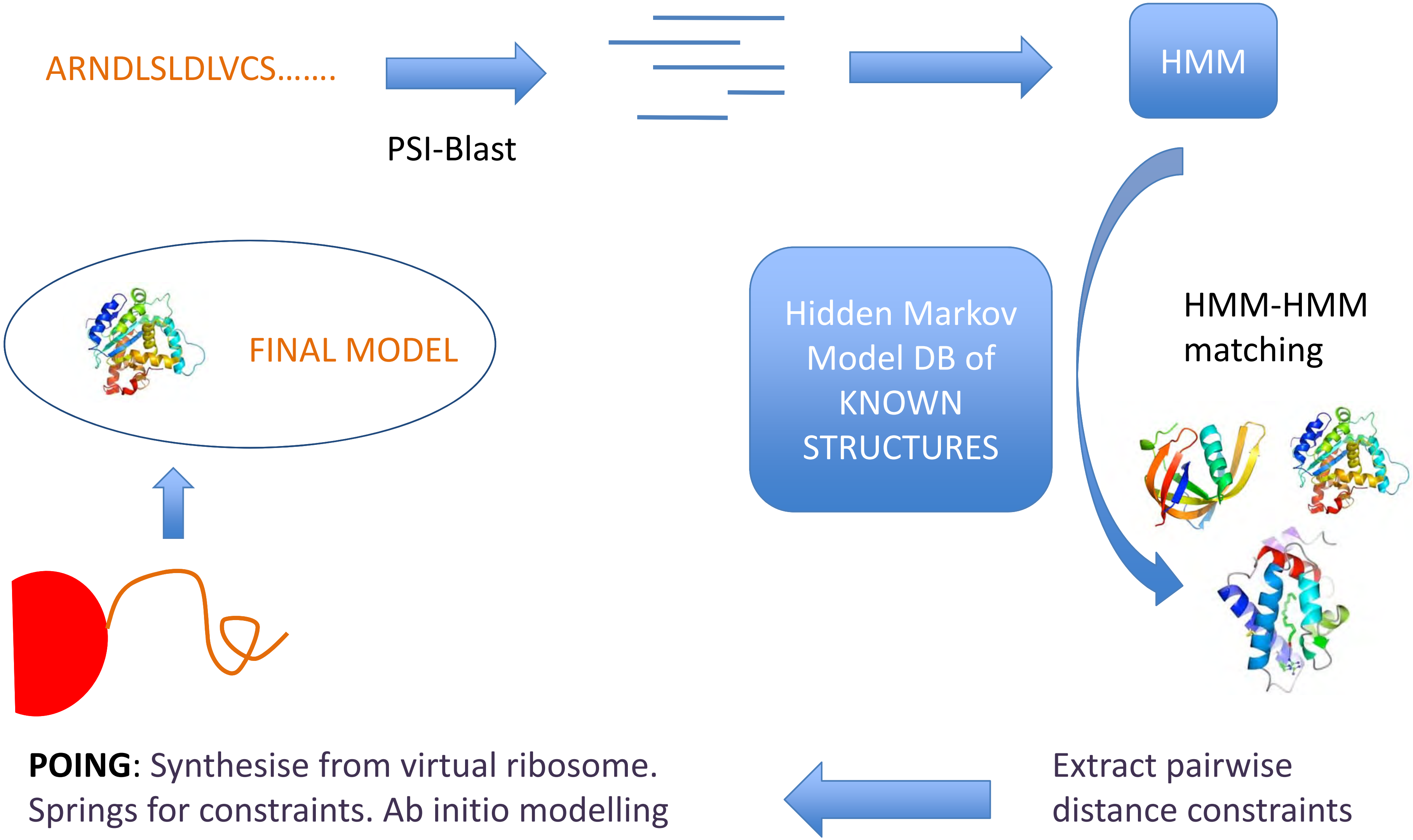
Thus need a system to fold a protein without templates and combine templates when we have them

Poing – simplified folding model

structure simplification



Phyre + Poing



ARNDLSLDLVCS…….

PSI-Blast

HMM

HMM-HMM matching

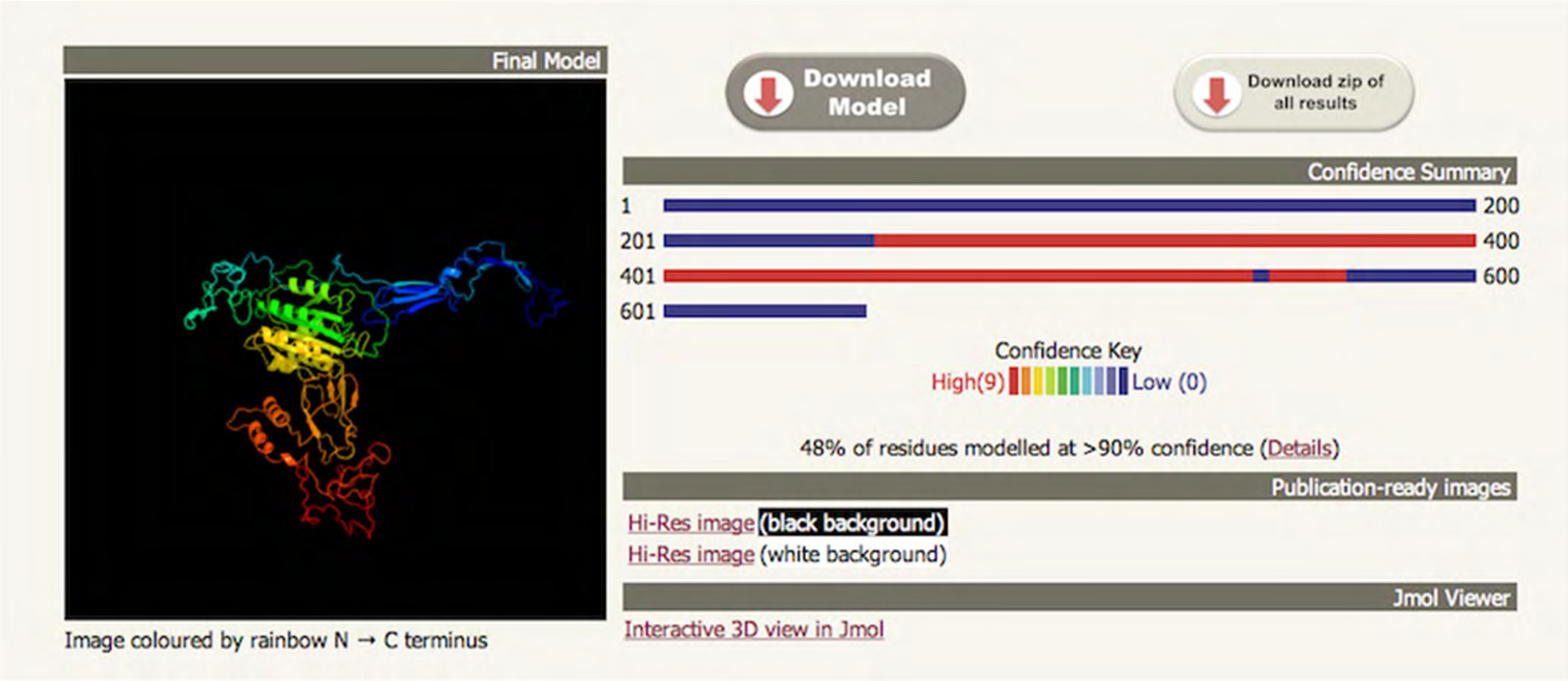
Hidden Markov Model DB of KNOWN STRUCTURES

FINAL MODEL

Extract pairwise distance constraints

POING: Synthesise from virtual ribosome. Springs for constraints. Ab initio modelling of missing regions.

Intensive mode



48% of residues modelled at >90% confidence (Details)

Hi-Res image (black background)

Hi-Res image (white background)

Interactive 3D view in Jmol

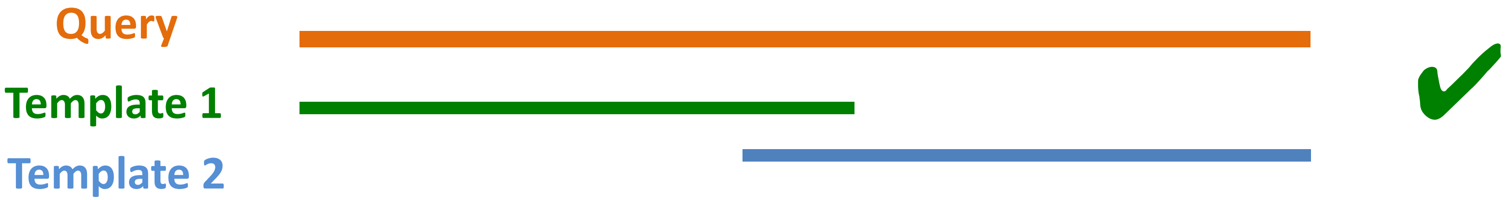
Image coloured by rainbow N → C terminus

Intensive mode

- Designed to handle mutliple domains or proteins with substantial stretches of sequence without detectable homologous structures.
- POOR at ab initio regions
- GOOD at combining multiple templates covering different regions

Intensive mode

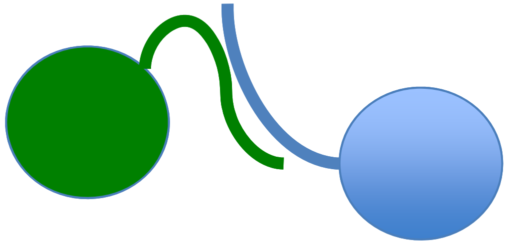
- Relative domain orientation will NOT generally be correct if those domains come from different PDB's with little structural overlap.



Query

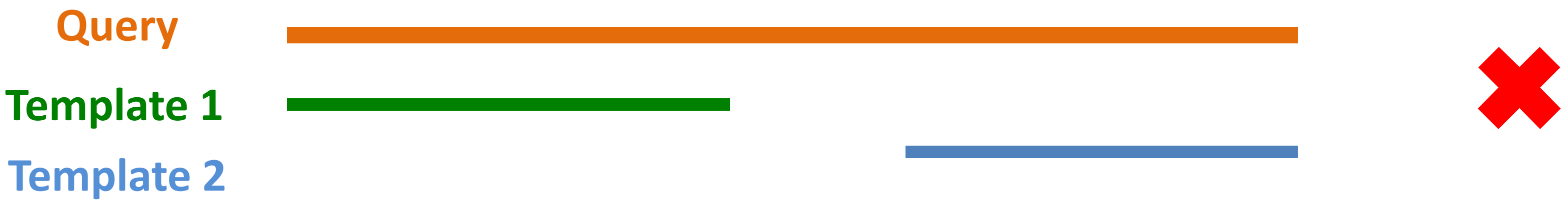
Template 1 ✓

Template 2



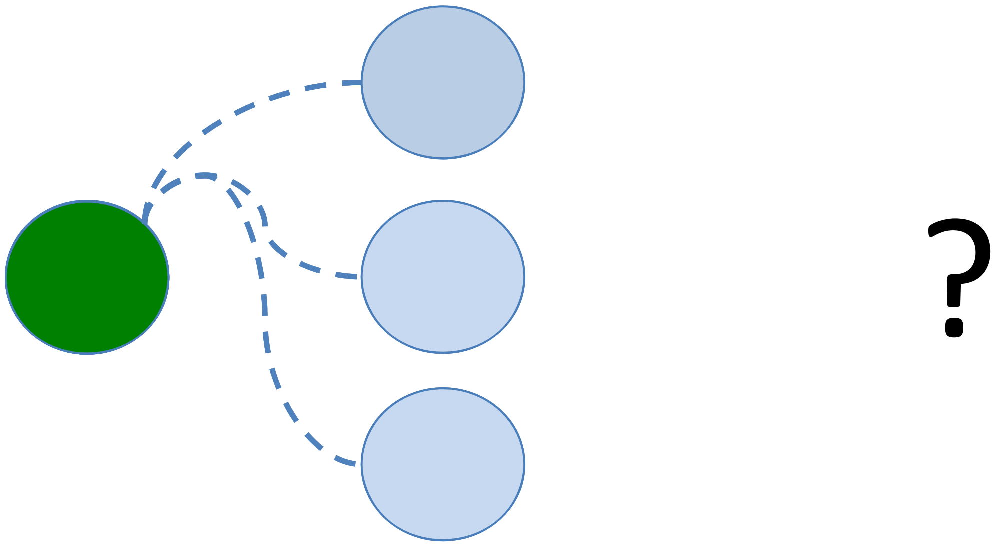
Intensive mode

- Relative domain orientation will NOT generally be correct if those domains come from different PDB's with little structural overlap.


Query

Template 1

Template 2



Intensive mode

“Intensive” does not always equal “Better”!

Checklist

- Always use normal mode first to understand what regions can be well modelled
- Multiple overlapping high confidence domains? Good, try intensive. Otherwise skip it.
- Danger of “spaghettification”
- Active development, new version ‘soon’

How does Phyre2 work?

- “Normal” Mode
- “Intensive” Mode
- Advanced functions
  - Phyre Investigator on web page including mutational analysis by SuSPect
  - Log in to use expert mode

Phyre Investigator

- What parts of a model are reliable?
- What parts may be functionally important? (guide mutagenesis, understand mutants/SNPs)
- What residues are involved in interactions with other proteins?

Phyre Investigator

- Clashes
- Rotamer outliers
- Ramachandran outliers
- ProQ2 model quality assessment
- Alignment confidence (HHsearch)
- Conservation/evolutionary trace (Jenson-Shannon divergence –far faster and just as accurate as ET)
- Catalytic Site Atlas
- Disorder
- Pocket detection (Fpocket)
- Protein interface residues (PI-Site, ProtinDB)
- Conserved Domain Database ‘conserved features’ for NCBI-curated domains

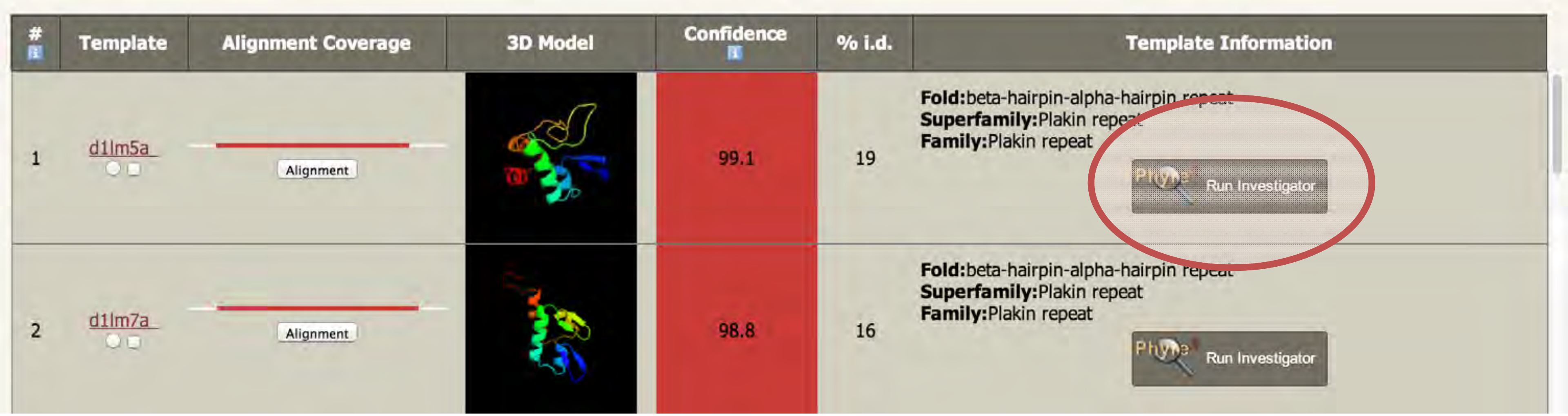
Phyre Investigator

Effect of Mutations?

- Will a SNP effect my protein’s function?
- New method: SuSPect by Chris Yates
- Integrated into Phyre Investigator
- Also standalone server

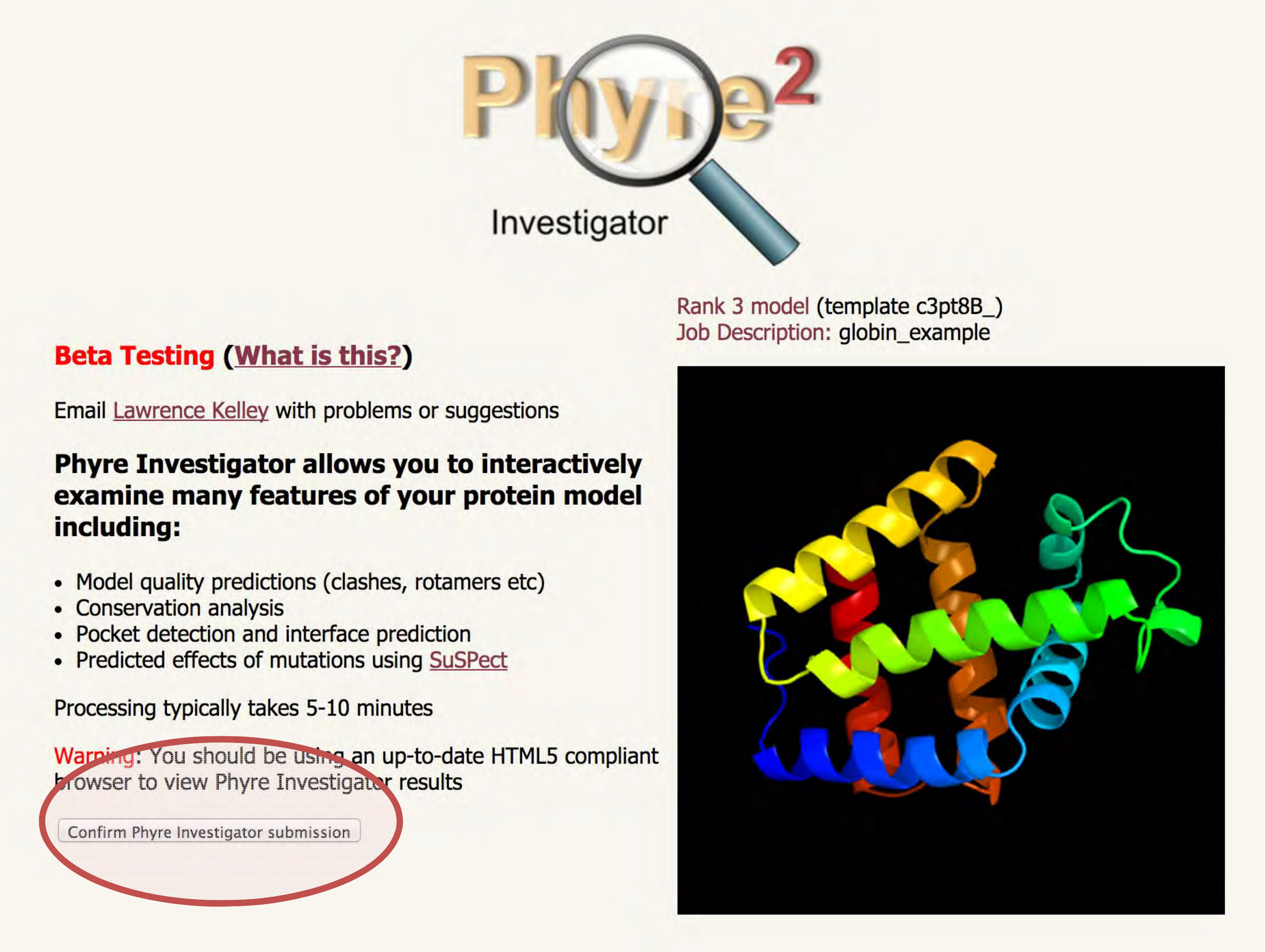
Yates CM, Filippis I, Kelley LA, Sternberg MJE. SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. *Journal of Molecular Biology*. 2014;426(14):2692-2701.

Phyre Investigator



#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1lm5a_ ○ ☐	Alignment		99.1	19	**Fold:**beta-hairpin-alpha-hairpin repeat **Superfamily:**Plakin repeat **Family:**Plakin repeat Phyre Run Investigator
2	d1lm7a_ ○ ☐	Alignment		98.8	16	**Fold:**beta-hairpin-alpha-hairpin repeat **Superfamily:**Plakin repeat **Family:**Plakin repeat Phyre Run Investigator

Phyre Investigator



Rank 3 model (template c3pt8B_)
Job Description: globin_example

Beta Testing (What is this?)

Email Lawrence Kelley with problems or suggestions

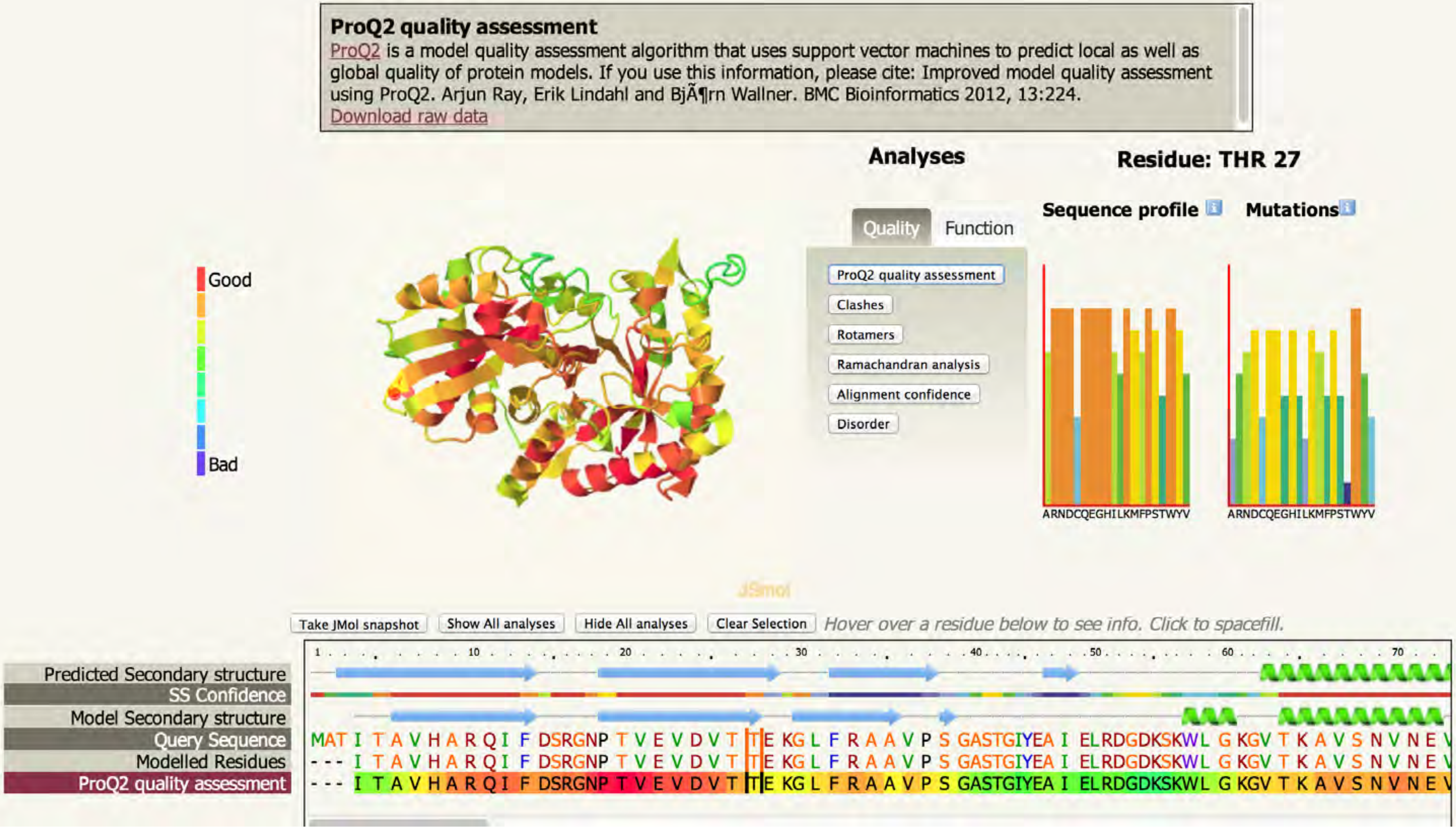
Phyre Investigator allows you to interactively examine many features of your protein model including:

- Model quality predictions (clashes, rotamers etc)
- Conservation analysis
- Pocket detection and interface prediction
- Predicted effects of mutations using SuSPect

Processing typically takes 5-10 minutes

Warning: You should be using an up-to-date HTML5 compliant browser to view Phyre Investigator results

Phyre Investigator



ProQ2 quality assessment

ProQ2 is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. If you use this information, please cite: Improved model quality assessment using ProQ2. Arjun Ray, Erik Lindahl and BjÃ¶rn Wallner. BMC Bioinformatics 2012, 13:224.

Download raw data

Hover over a residue below to see info. Click to spacefill.

SuSPect – Phenotypic effect of amino acid variants

Sequence conservation

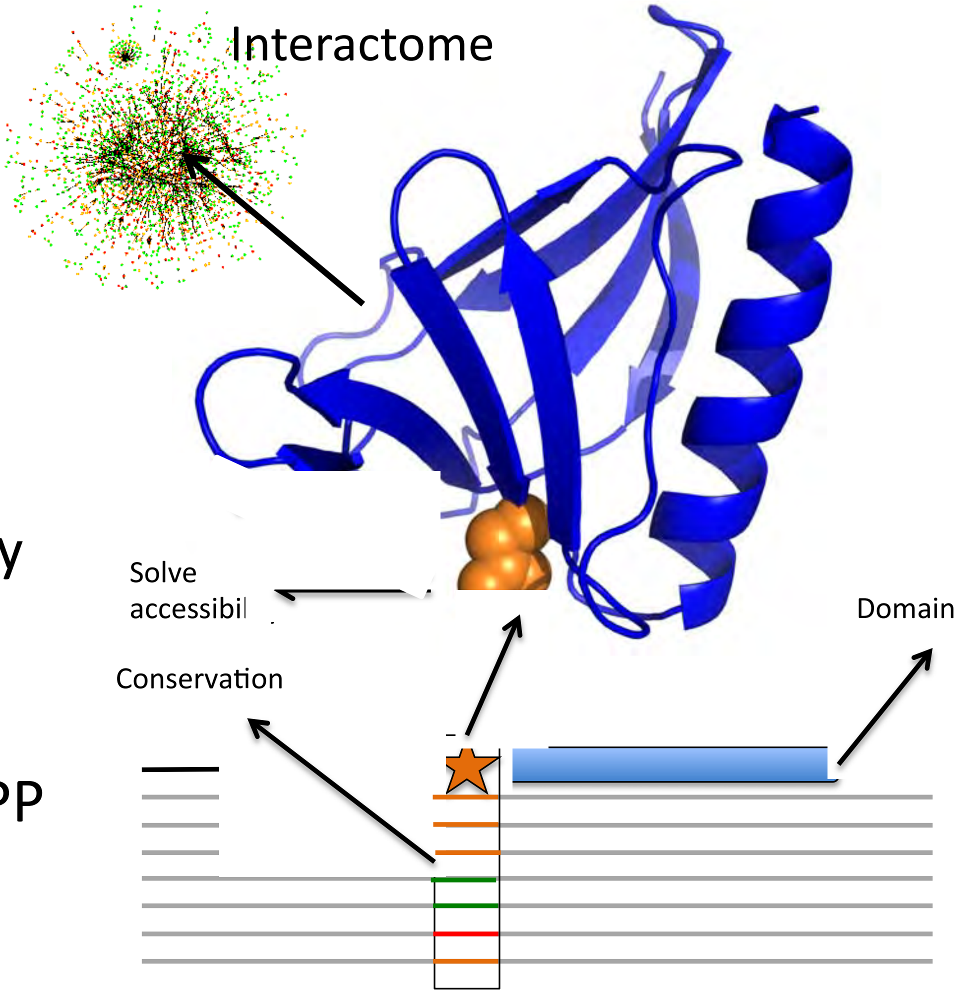
- PSSM
- Pfam domain
- Jensen-Shannon entropy

Structural features

- Predicted solvent accessibility

Network features

- Protein-protein interaction (PP as domain centrality

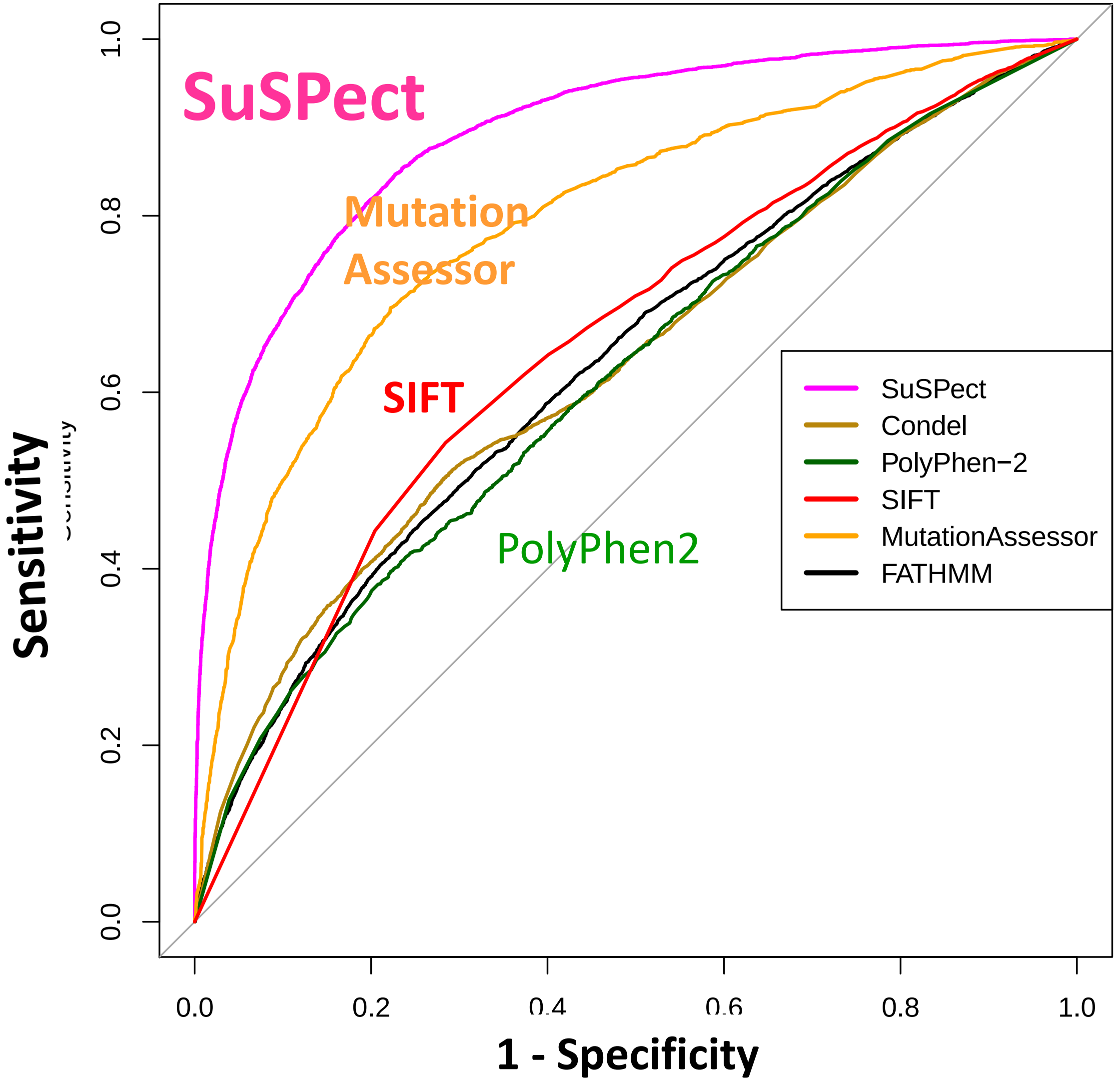


SuSPect – Results on non-training data (VariBench)

$$Specificity = \frac{TP}{TP + TN}$$

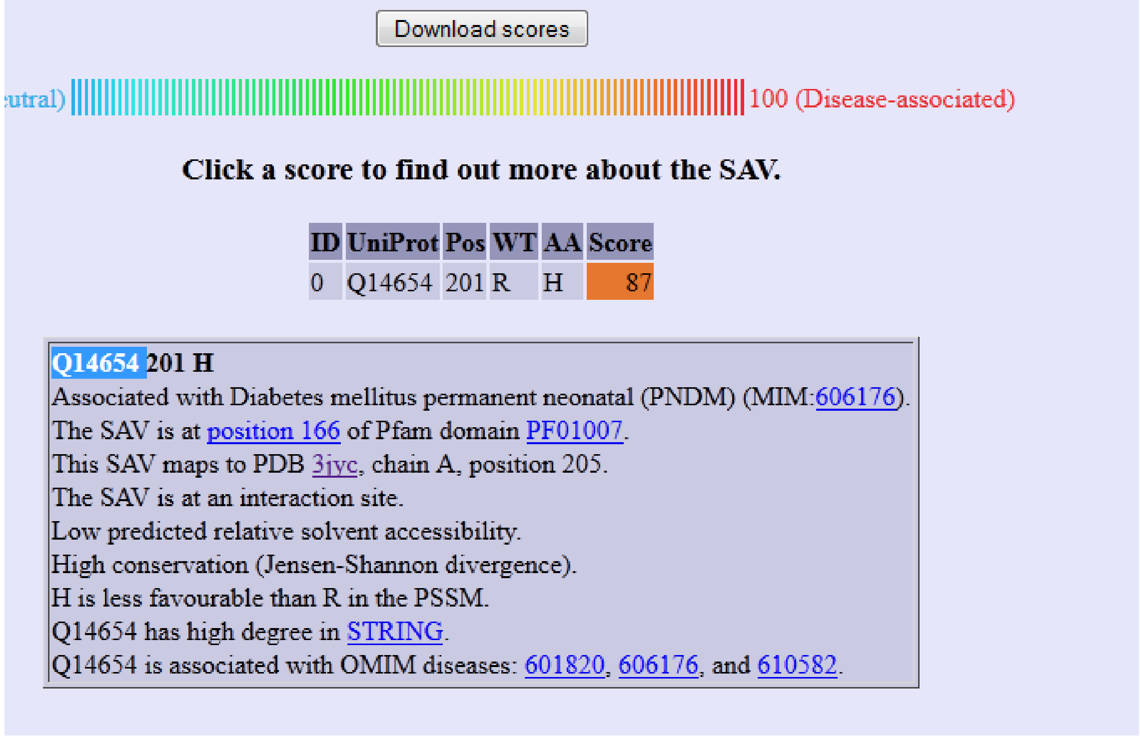
$$Sensitivity = \frac{TP}{TP + FP}$$

Benchmark consists of 20k SNPs
(15k Neutral, 5k pathogenic)



Neonatal diabetes

- Arg 201 His in ATP-sensitive inward rectifier potassium channel 11 (Kir6.6)
- SuSPect gives score of 87/100 – high probability of disease associated

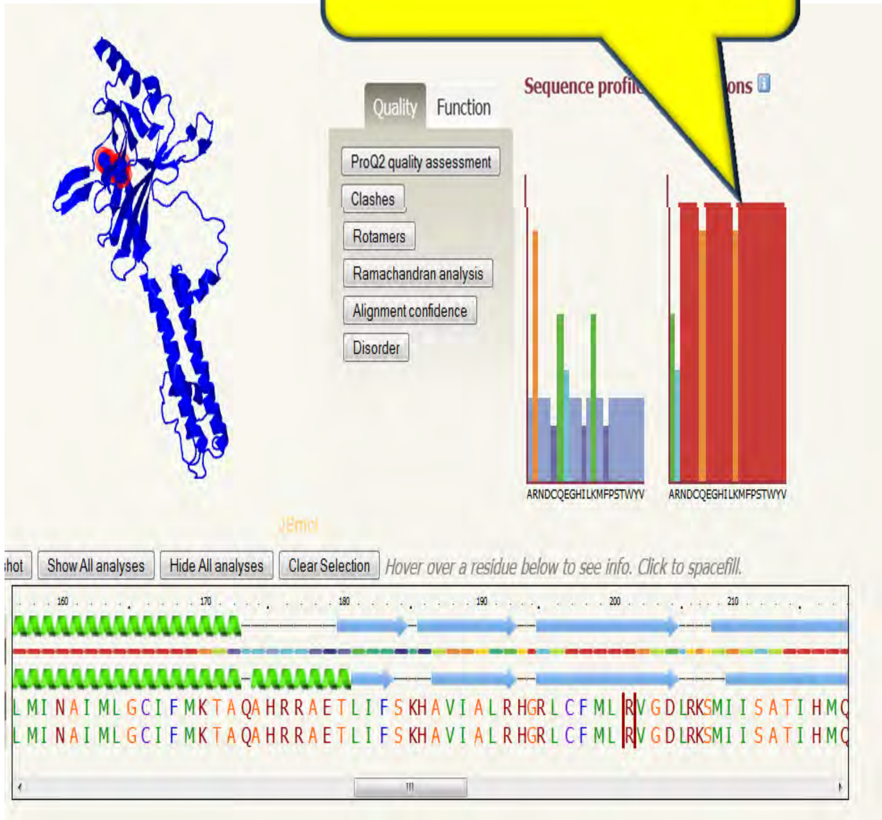


Click a score to find out more about the SAV.

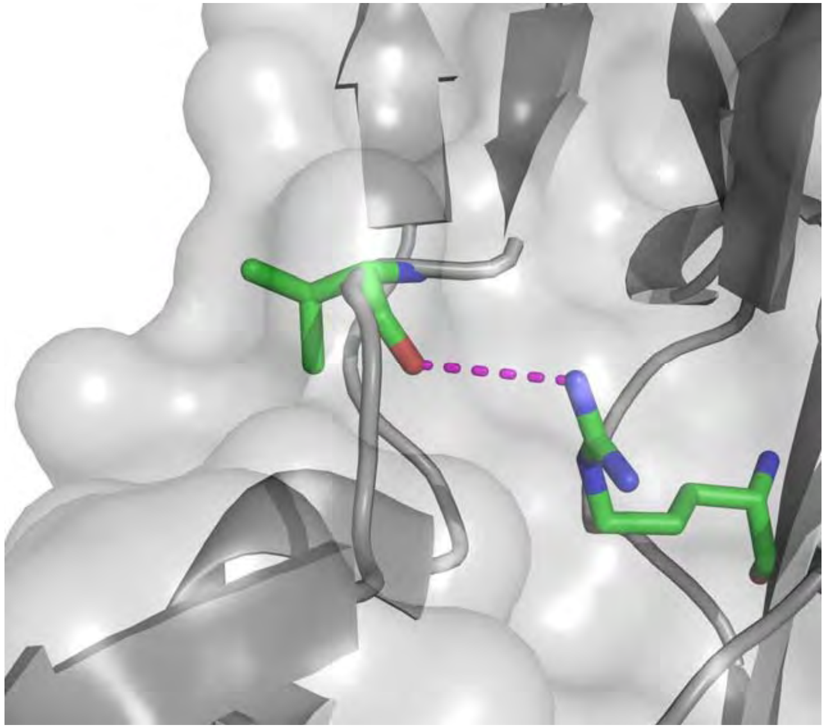
ID	UniProt	Pos	WT	AA	Score
0	Q14654	201	R	H	87

Q14654 201 H
Associated with Diabetes mellitus permanent neonatal (PNDM) (MIM:606176).
The SAV is at position 166 of Pfam domain PF01007.
This SAV maps to PDB 3jyc, chain A, position 205.
The SAV is at an interaction site.
Low predicted relative solvent accessibility.
High conservation (Jensen-Shannon divergence).
H is less favourable than R in the PSSM.
Q14654 has high degree in STRING.
Q14654 is associated with OMIM diseases: 601820, 606176, and 610582.

Phyre2 yields model which suggest structural basis for disease



Most variants predict to be disease associated

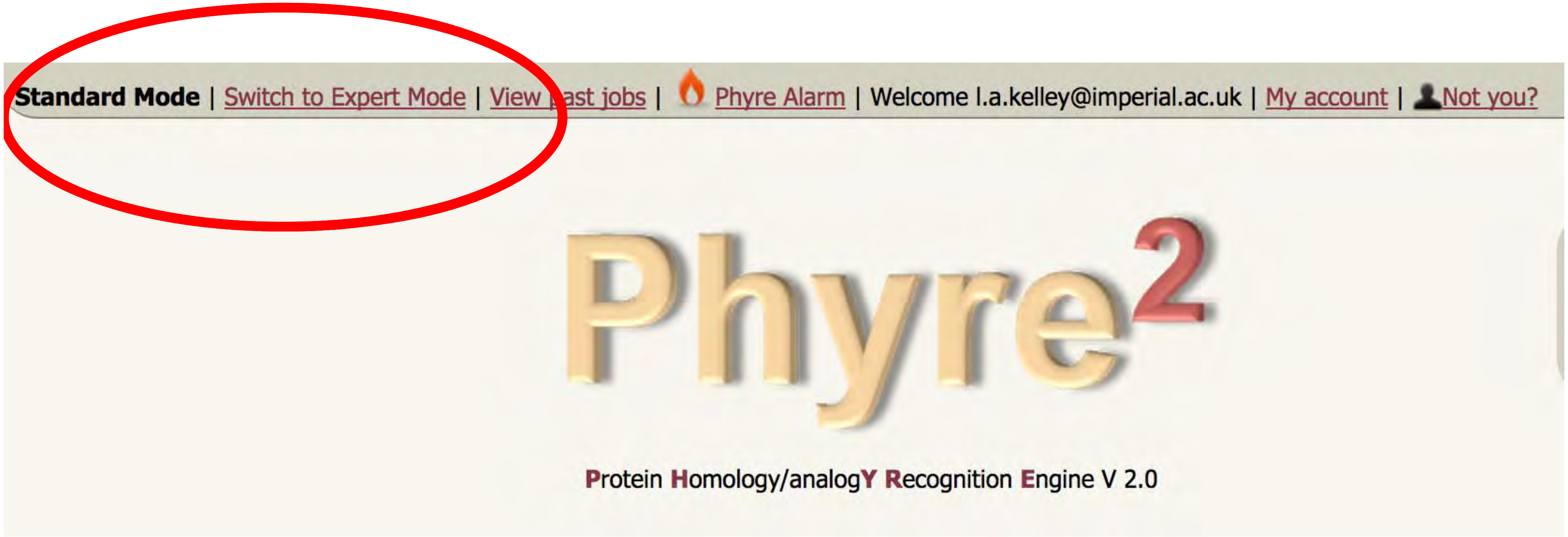


Arg 201 forms H-bond with main chain O

His in variant could not form similar interaction

Advanced functions

Register and Log in to access Expert Mode



Standard Mode | Switch to Expert Mode | View past jobs | Phyre Alarm | Welcome l.a.kelley@imperial.ac.uk | My account | Not you?

Advanced functions

- PhyreAlarm – automatically re-run tricky sequences every week
- BackPhyre – compare a structure to up to 30 genomes
- One-To-One Threading – use specfic PDB for model building
- Batch Jobs – run many sequences at once
- Job Manager – keep track of your jobs and history

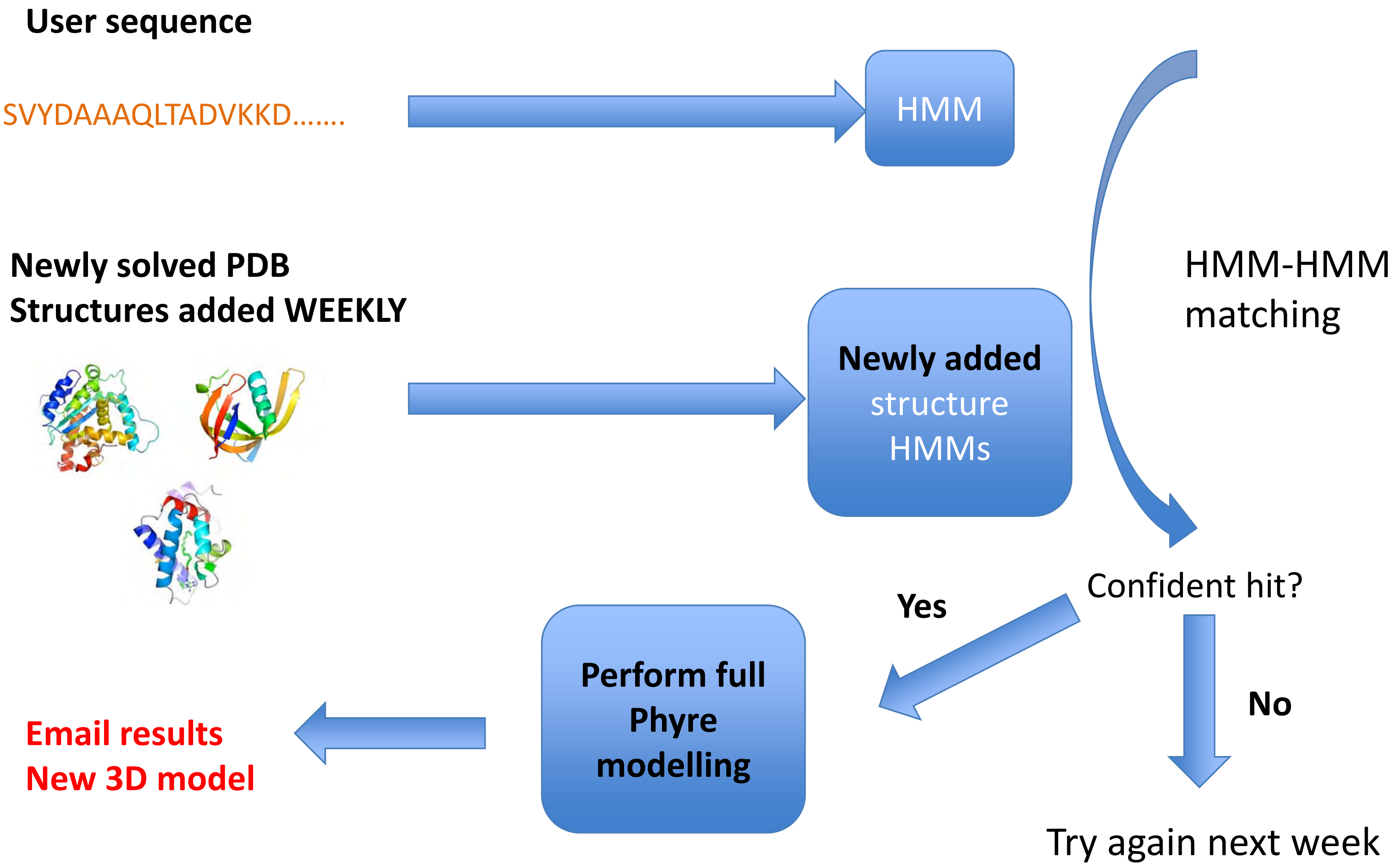
Advanced functions

- **PhyreAlarm –** automatically re-run tricky sequences every week
- **BackPhyre –** compare a structure to up to 30 genomes
- **One-To-One Threading –** use specfic PDB for model building
- **Batch Jobs –** run many sequences at once
- **Job Manager –** keep track of your jobs and history

PhyreAlarm

- Sometimes no confident homology detected
- Automatically try every week as new structures are deposited in the PDB
- Receive an email if hit found
- PhyreAlarm auto-suggested in cases where sequence has low coverage by confident hits
- Two clicks adds your sequence to the alarm queue

PhyreAlarm



User sequence

SVYDAAAQLTADVKKD…….

HMM

Newly solved PDB
Structures added WEEKLY

Newly added structure HMMs

HMM-HMM matching

Confident hit?

Yes

No

Perform full Phyre modelling

Email results
New 3D model

Try again next week

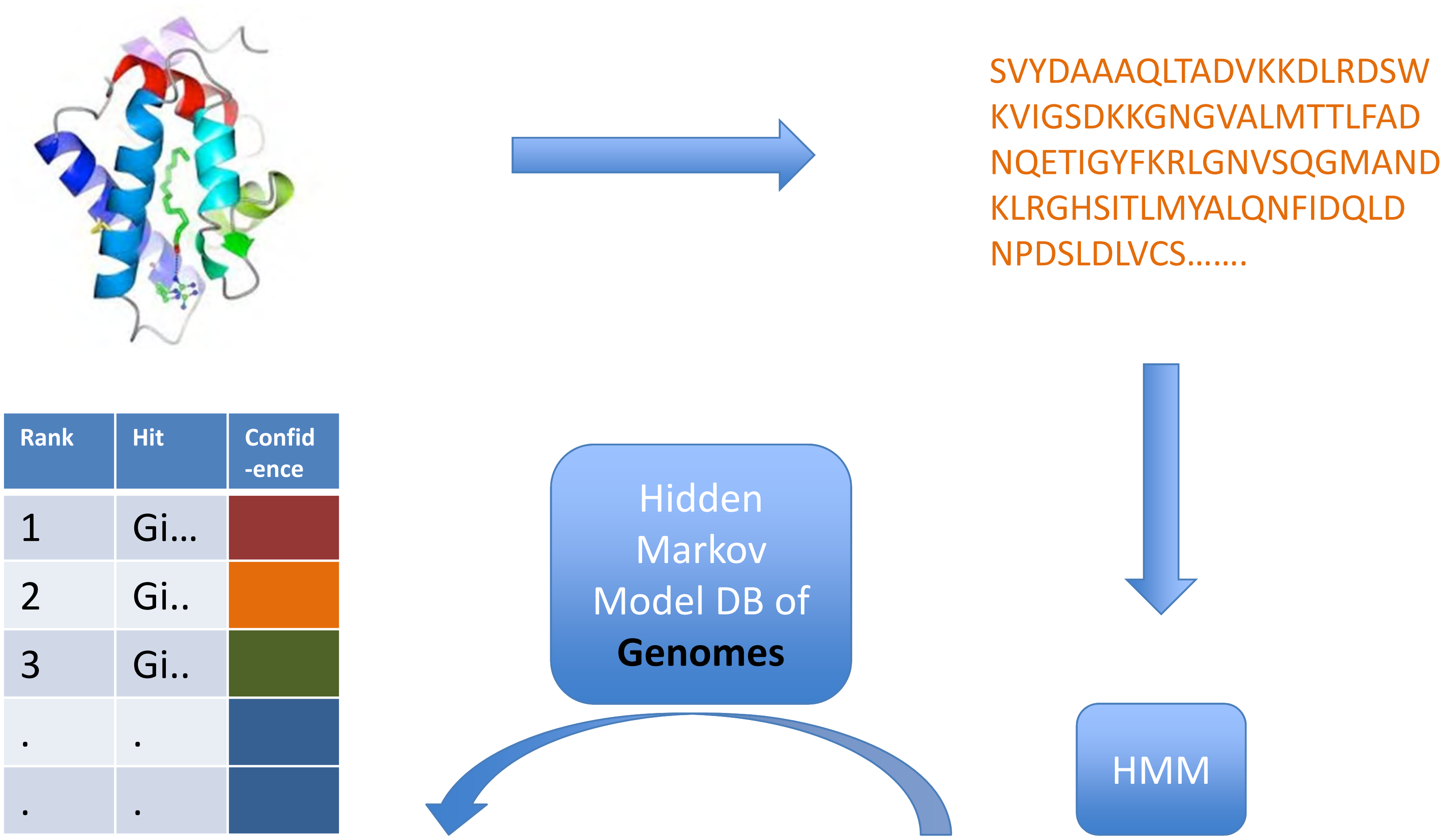
Advanced functions

- PhyreAlarm – automatically re-run tricky sequences every week
- BackPhyre – compare a structure to up to 30 genomes
- One-To-One Threading – use specfic PDB for model building
- Batch Jobs – run many sequences at once
- Job Manager – keep track of your jobs and history

BackPhyre

- Does a structure I’m interested in exist in an organism?
- 30 searchable genomes to-date.
- Scan multiple genomes at a time. Quite fast.
- New version will allow users to upload their own genomes of interest.

BackPhyre



User structure

SVYDAAAQLTADVKKDLRDSW
KVIGSDKKGNGVALMTTLFAD
NQETIGYFKRLGNVSQGMAND
KLRGHSITLMYALQNFIDQLD
NPDSLDLVCS…….

HMM

Hidden Markov Model DB of **Genomes**

HMM-HMM matching

Rank	Hit	Confid-ence
1	Gi…	🟥
2	Gi..	🟧
3	Gi..	🟩
.	.	🟦
.	.	🟦

Ranked list of genome hits

Advanced functions

- PhyreAlarm – automatically re-run tricky sequences every week
- BackPhyre – compare a structure to up to 30 genomes
- One-To-One Threading – use specfic PDB for model building
- Batch Jobs – run many sequences at once
- Job Manager – keep track of your jobs and history

One-to-One Threading

- Useful if you:

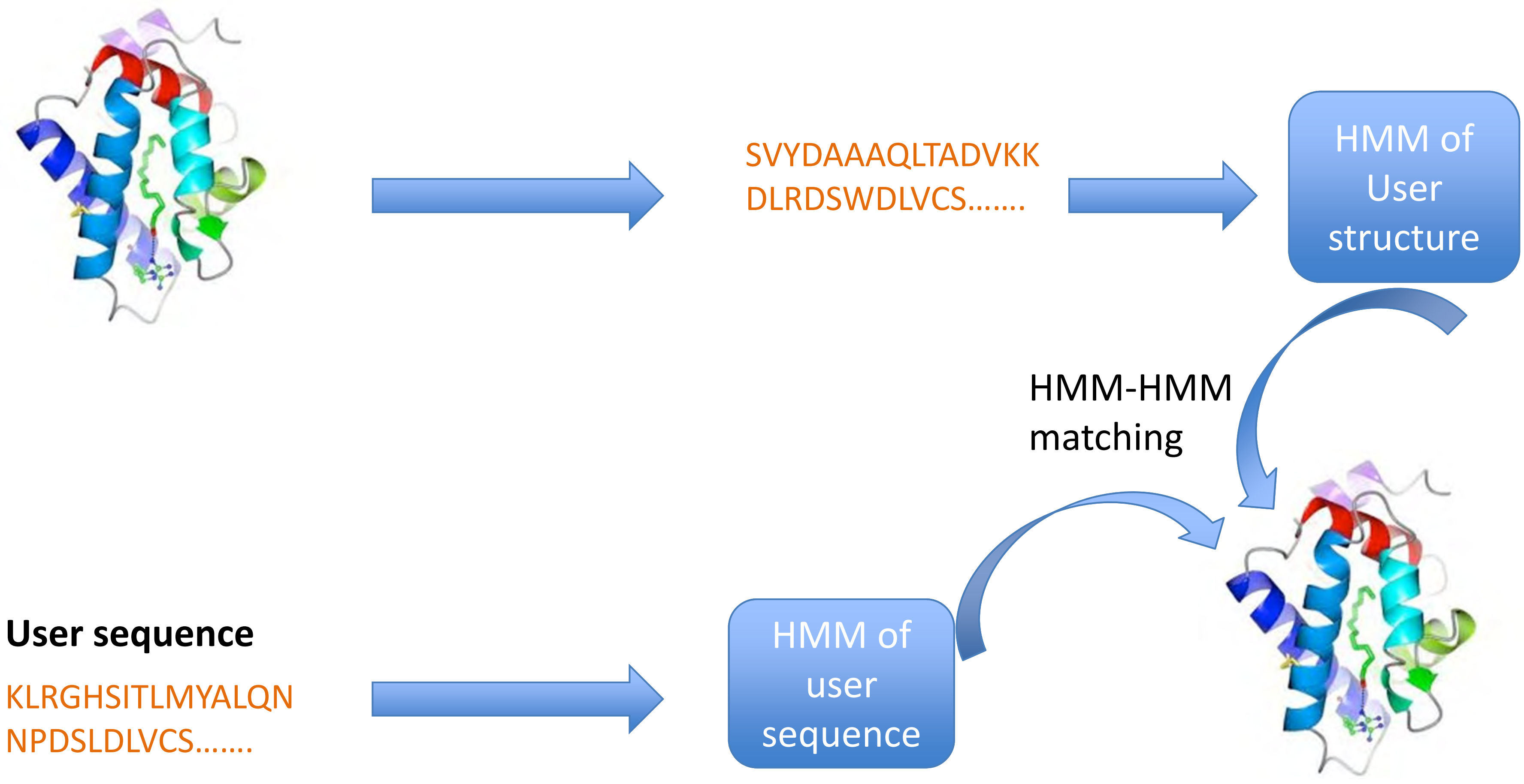
a) Know a better template than found by Phyre2

b) Have your own structure not yet in the PDB

c) Model a a lower-ranked (>20) template

d) Want more expert control over alignment options: local/global, secondary structure weight etc.

One to one threading



User structure

SVYDAAAQLTADVKK
DLRDSWDLVCS.......

HMM of
User
structure

HMM-HMM
matching

User sequence

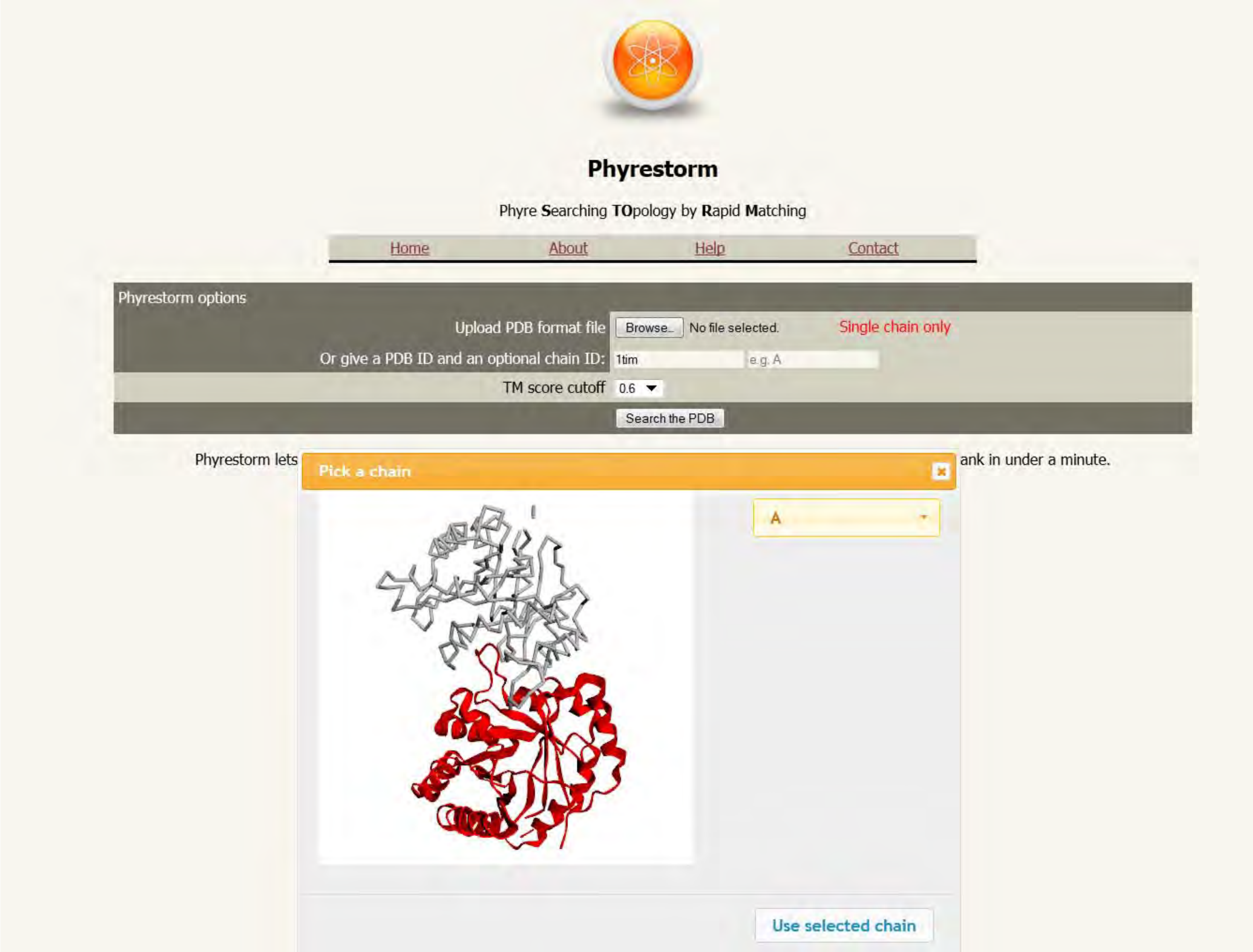
KLRGHSITLMYALQN
NPDSLDLVCS.......

HMM of
user
sequence

Final model

Future

PhyreStorm



Phyrestorm

Phyre **S**earching **TO**pology by **R**apid **M**atching

Phyrestorm options

Upload PDB format file: No file selected. Single chain only

Or give a PDB ID and an optional chain ID: 1tim e.g. A

TM score cutoff: 0.6

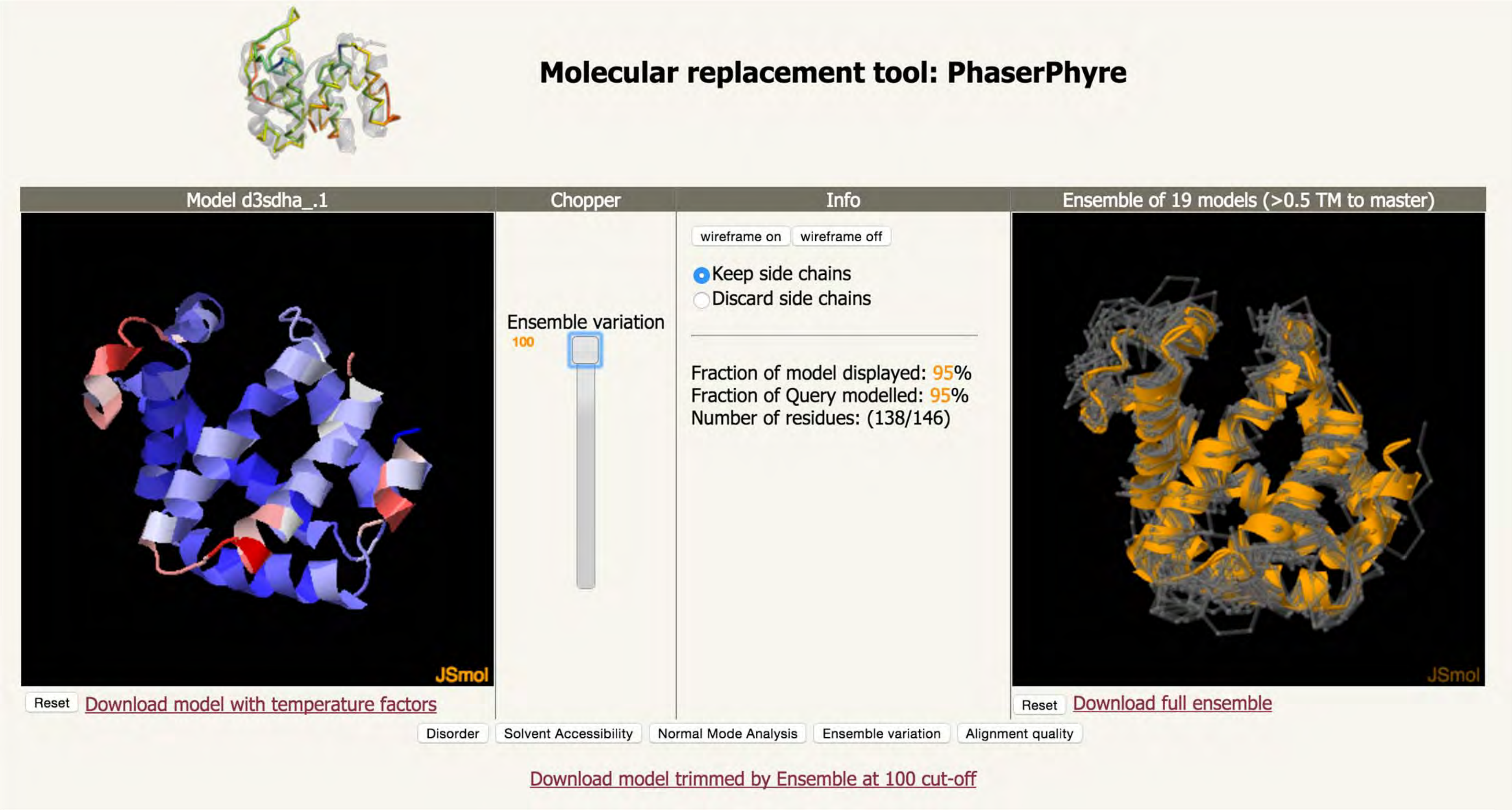
Phyrestorm lets … ank in under a minute.

Pick a chain: A

PhyreStorm

- **S**earching **To**pology with **R**apid **M**atching
- Structural search and alignment of the **entire** PDB in under 1 minute.
- Go directly from a Phyre2 model and find all other similar structures rapidly.
- Beta released

PhaserPhyre



Molecular replacement tool: PhaserPhyre

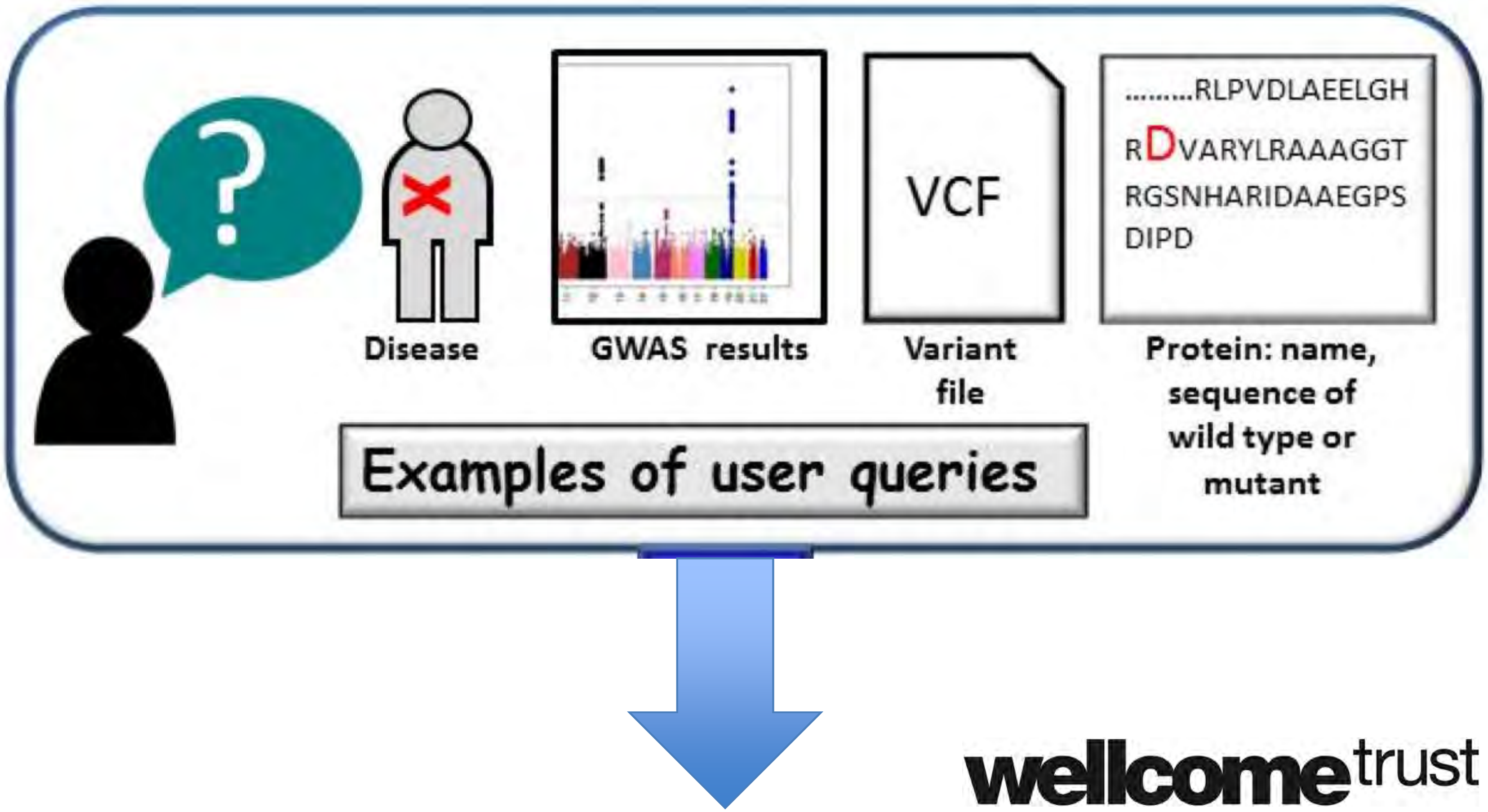
Fraction of model displayed: 95%
Fraction of Query modelled: 95%
Number of residues: (138/146)

● Keep side chains
○ Discard side chains

Ensemble variation: 100

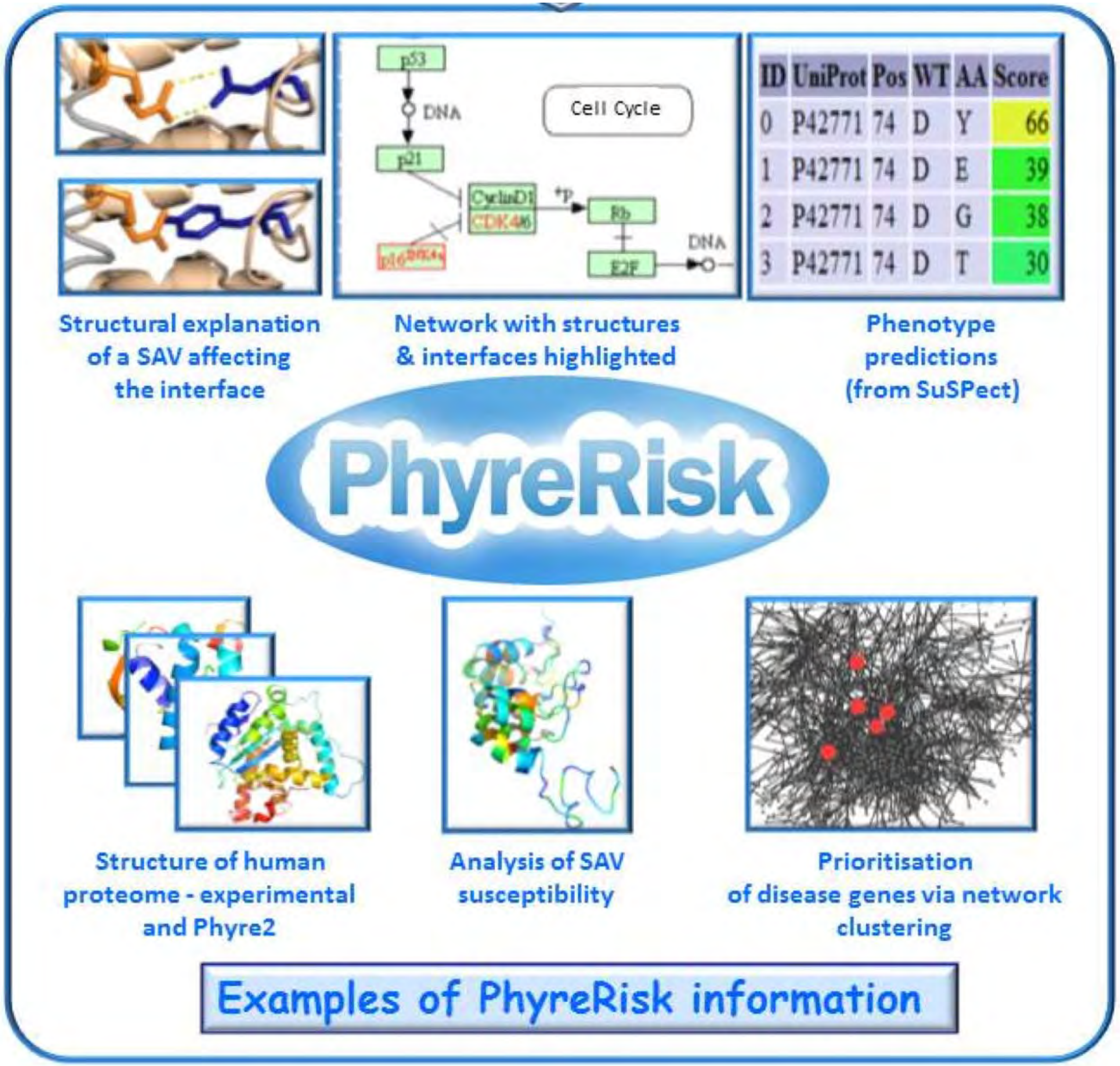
PhyreRisk (with Prof R Houlston ICR)

Integrate disease networks, SNPs, GWAS, protein structure and complexes



wellcome trust

PhyreRisk



ID	UniProt	Pos	WT	AA	Score
0	P42771	74	D	Y	66
1	P42771	74	D	E	39
2	P42771	74	D	G	38
3	P42771	74	D	T	30

Structural explanation of a SAV affecting the interface

Network with structures & interfaces highlighted

Phenotype predictions (from SuSPect)

Structure of human proteome - experimental and Phyre2

Analysis of SAV susceptibility

Prioritisation of disease genes via network clustering

Examples of PhyreRisk information

Protein structure prediction using Phyre^2 and understanding genetic variants.

Prof. Michael Sternberg
Dr. Lawrence Kelley
Mr. Stefans Mezulis
Dr Chris Yates

Imperial College London

BBSRC
20 Years of Pioneering
Great British Bioscience

MRC Medical Research Council

wellcome trust

Timetable

- **10.00 – 11.00 Lecture**
- **11.00 – 11.30 Tea/Coffee**
- Courtyard, West Medical Building
- **11.30 – 1.00 Hands on workshop using Phyre^2**
- Computer Cluster 515, West Medical Building

Many thanks to Glasgow Polyomics and Amy Cattanach

Glasgow Polyomics
www.polyomics.gla.ac.uk