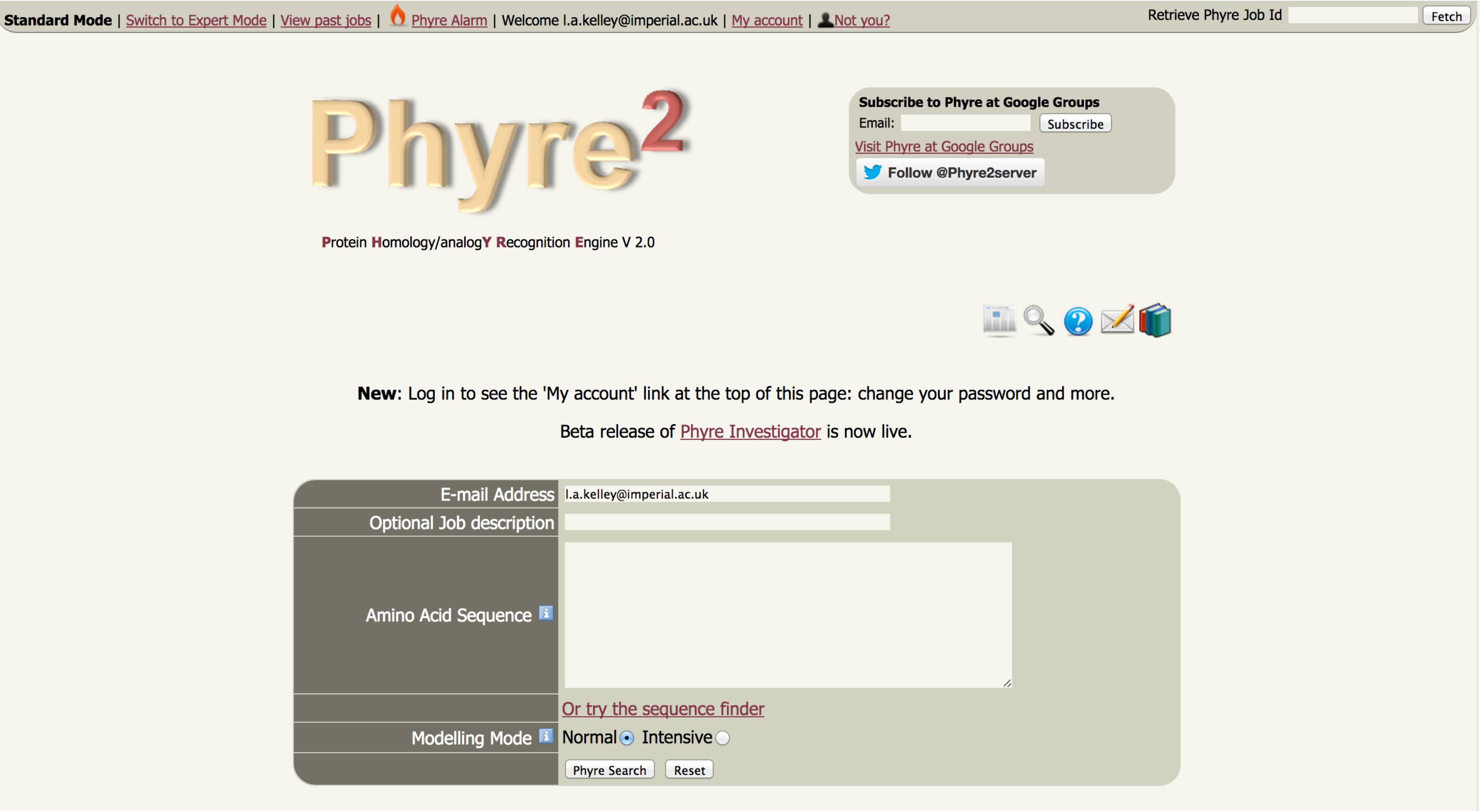
Phyre2

SVYDAAAQLTADVKKDLRDSW
KVIGSDKKGNGVALMTTLFAD
NQETIGYFKRLGNVSQGMAND
KLRGHSITLMYALQNFIDQLD
NPDSLDLVCS…….

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



Protein Homology/analogY Recognition Engine V 2.0

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
Optional Job description	
Amino Acid Sequence	
	Or try the sequence finder
Modelling Mode	Normal ● Intensive ○
	Phyre Search Reset

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

How does Phyre2 work?

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

Phyre2

ARDLVIPMIYCGHGY

User sequence

PSI-Blast

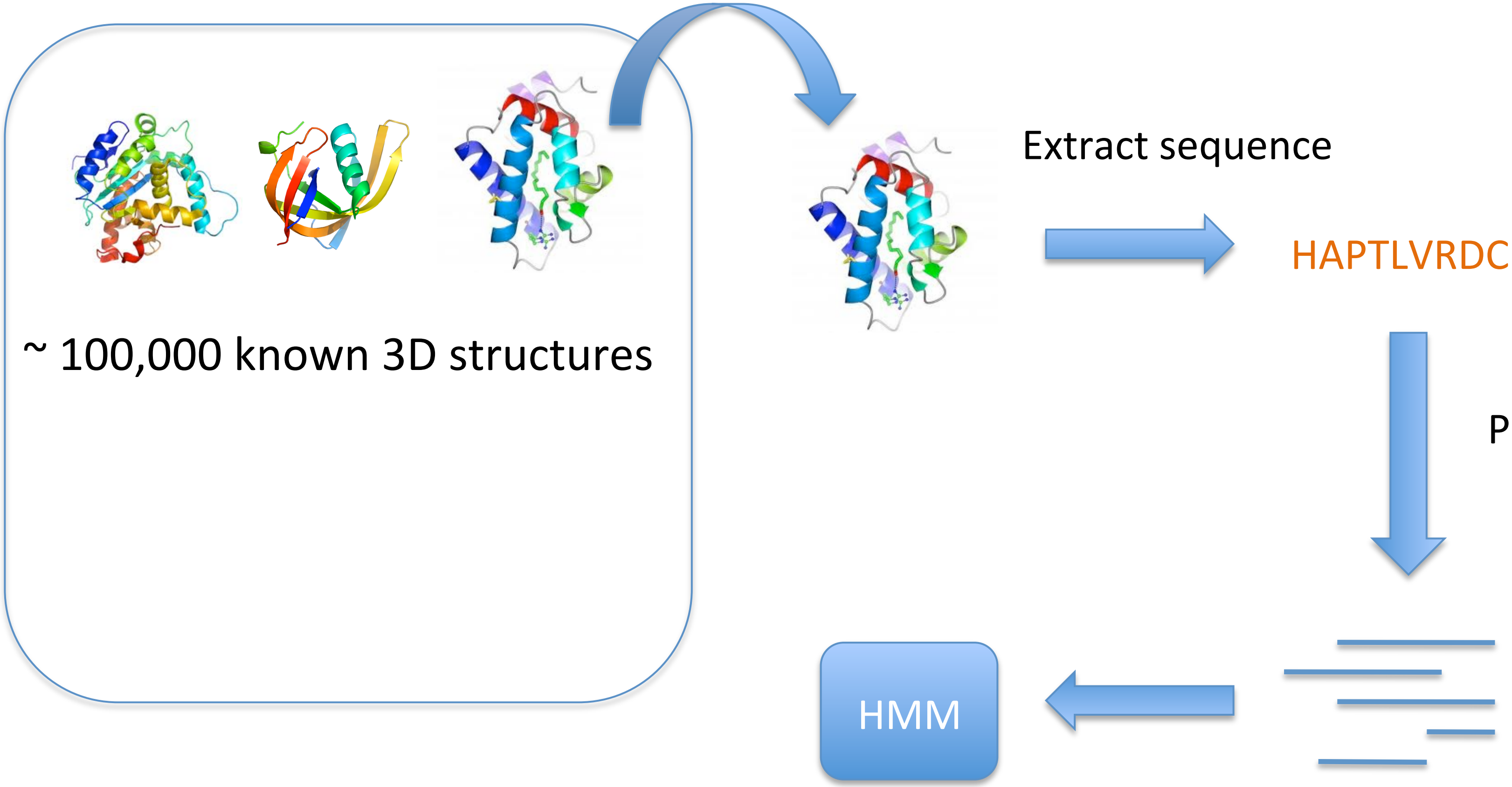
HMM

Hidden Markov model

Capture the mutational propensities at each position in the protein

An evolutionary fingerprint

Phyre2



~ 100,000 known 3D structures

Extract sequence

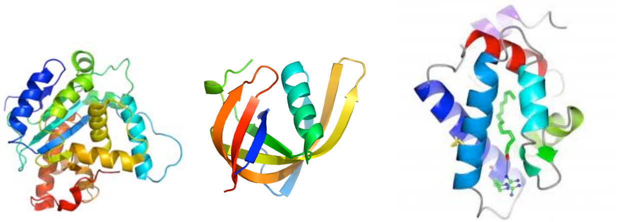
HAPTLVRDC…….

PSI-Blast

HMM

Hidden Markov model
for sequence of KNOWN structure

Phyre2



~ 100,000 known 3D structures

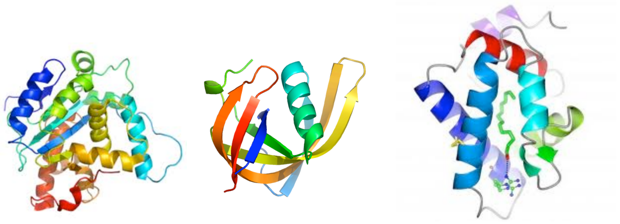
HMM

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

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

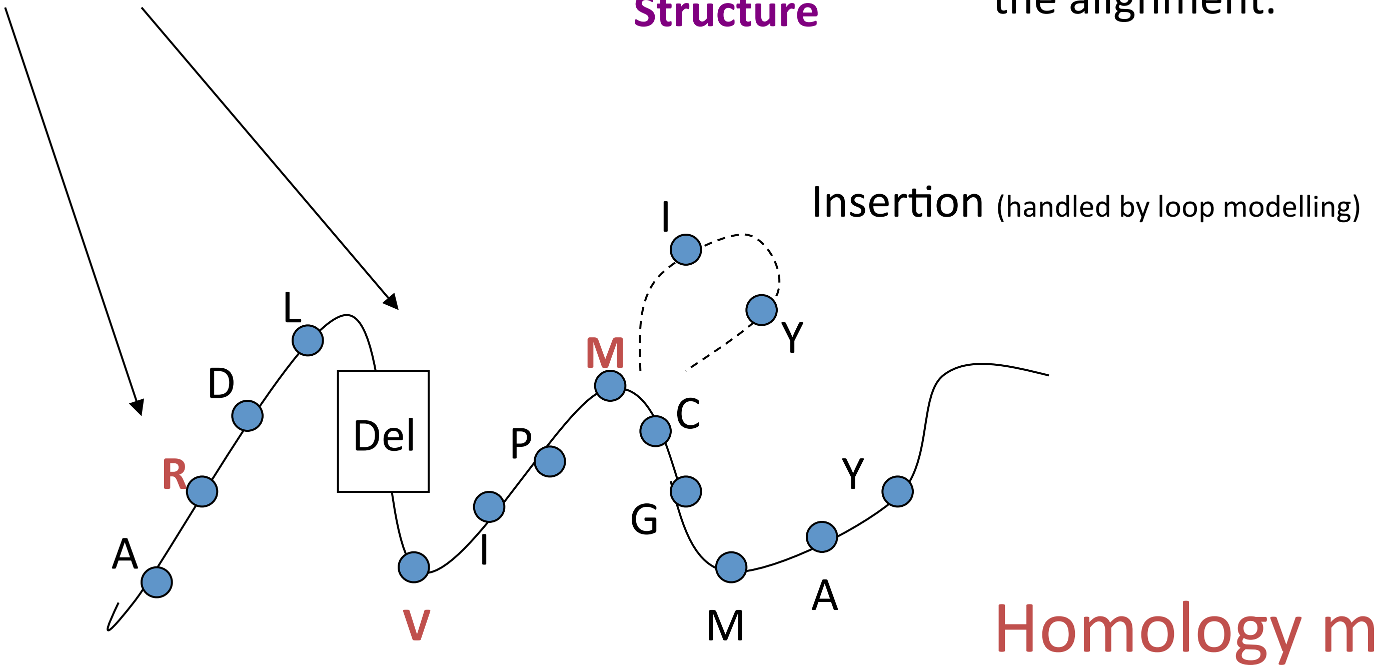
Sequence of known structure

3D-Model

From alignment to crude model

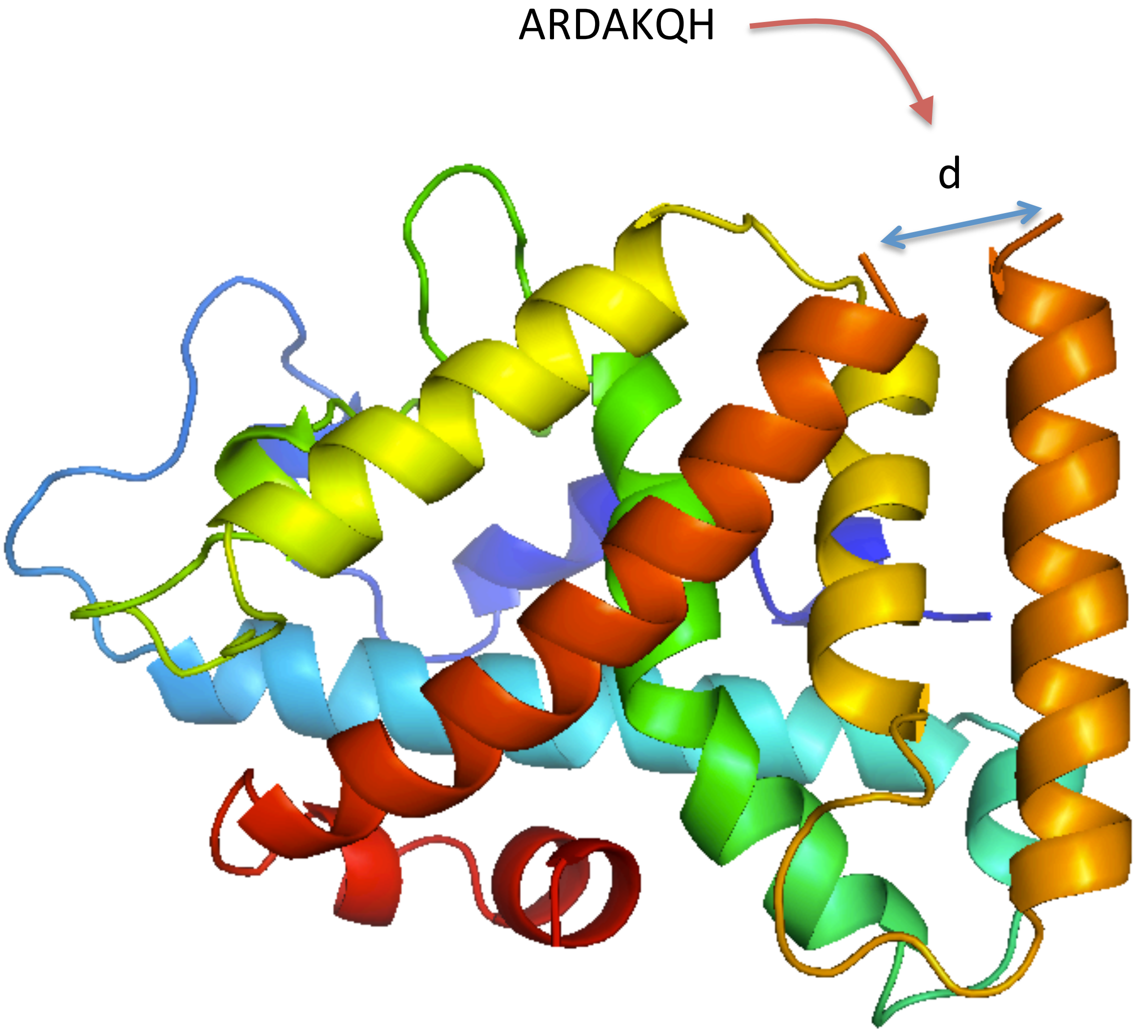
ARDL--VIPMIYCGHGY **Query**
AFDLCDLIPV--CGMAY **Known Structure**

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



Homology model

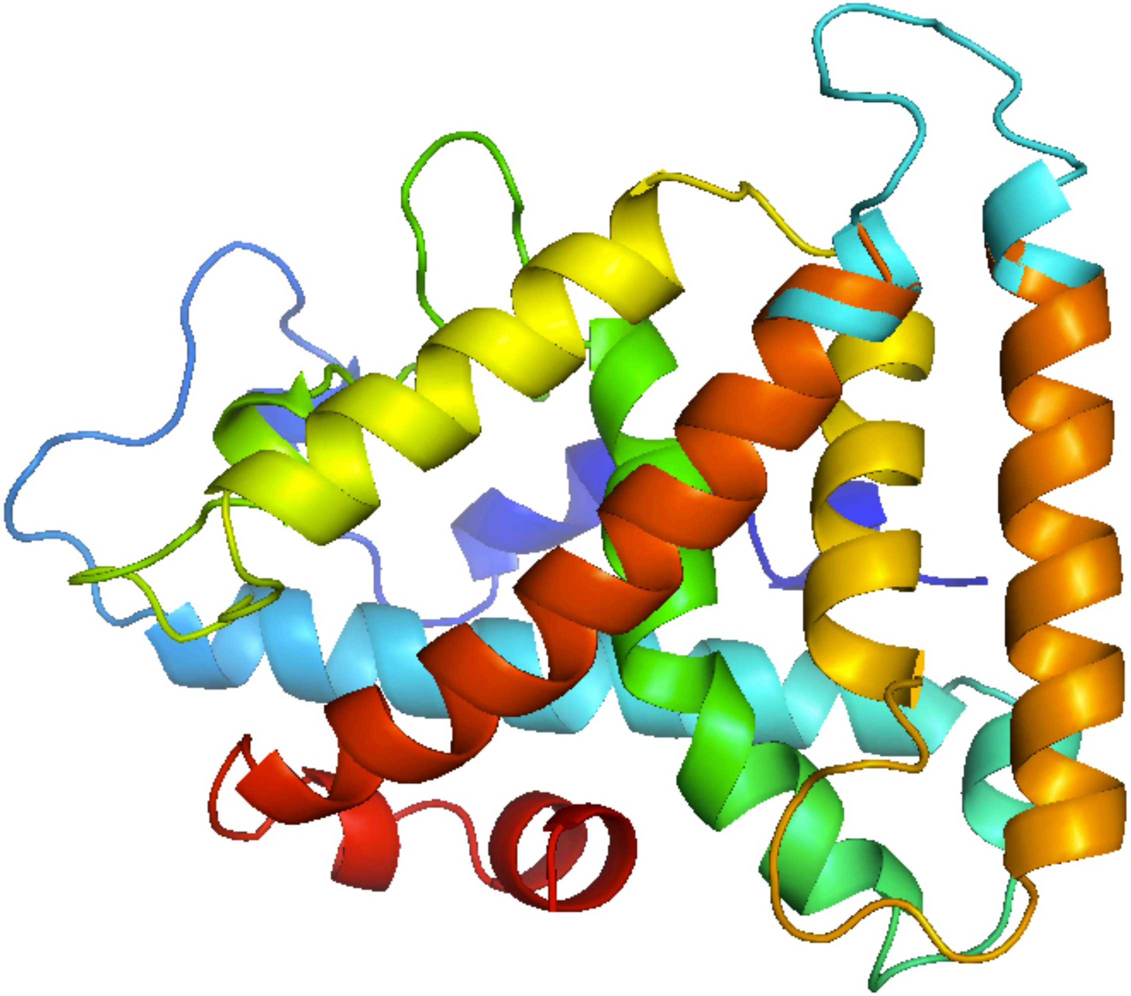
Loop modelling



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

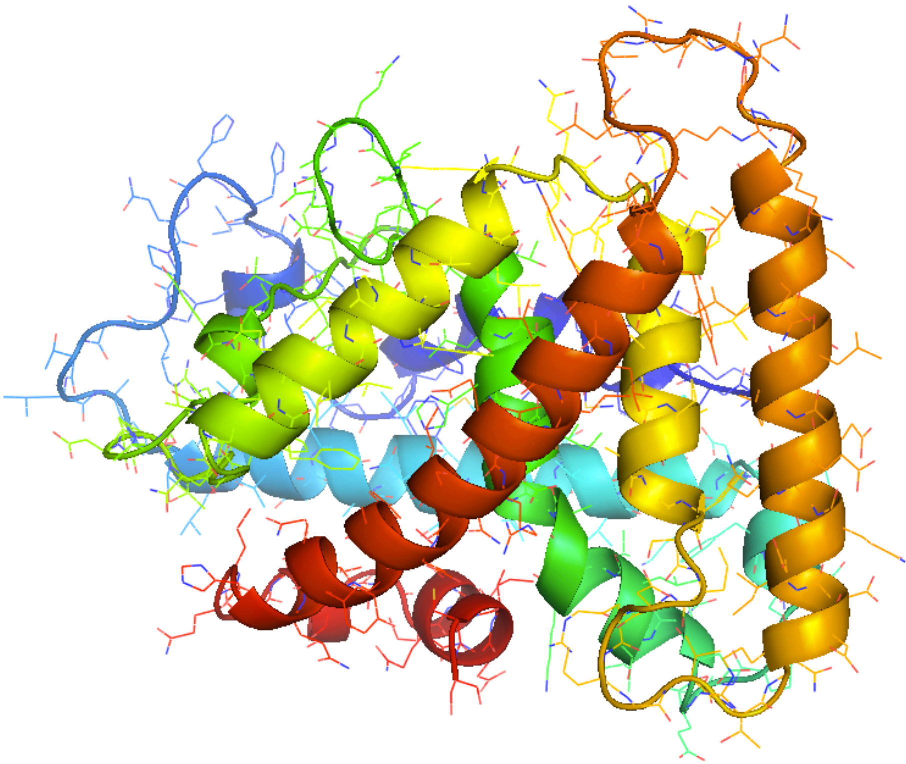
Sidechain modelling



Sidechain modelling

Optimisation problem

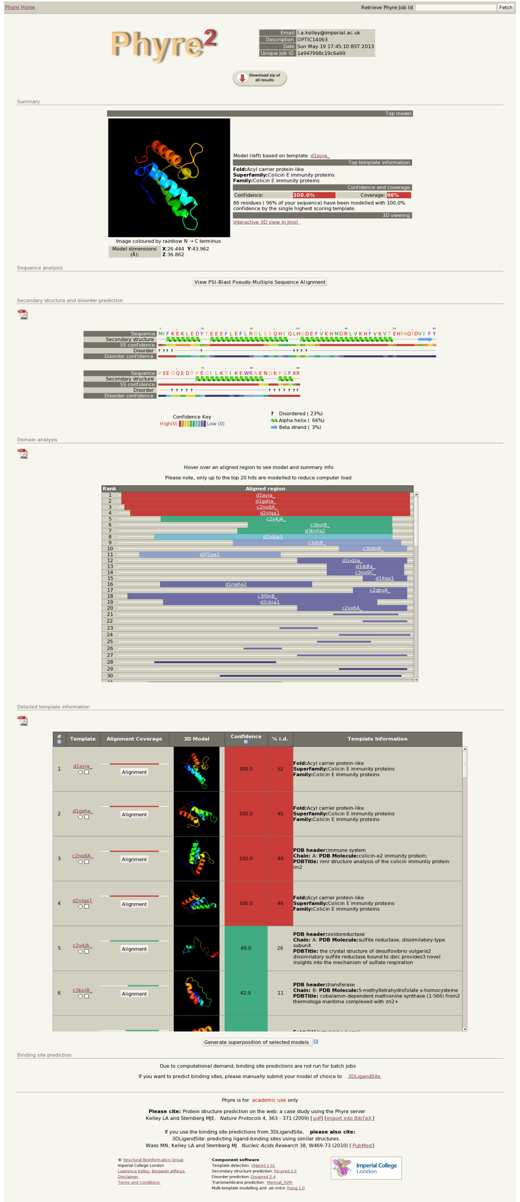
- Fit most probable rotamer at each position
- According to given backbone angles
- Whilst avoiding clashes



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



Phyre Home

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

Download zip of all results

Summary

Top model

Image coloured by rainbow N → C terminus

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

Model (left) based on template d1ayia_

Top template information

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

Confidence and coverage

Confidence: 100.0% Coverage: 96%

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

3D viewing

Interactive 3D view in Jmol

Sequence analysis

View PSI-Blast Pseudo-Multiple Sequence Alignment

Secondary structure and disorder prediction

Sequence MIFKEKLEDYTEEEFLEFLRGLSSQHIQLHGDEFVKHMDRLVKHFVKVTEHPAQTDVIFY
Secondary structure
SS confidence
Disorder ? ?? ? ? ????
Disorder confidence

Sequence PEEGQEDTPEGILKTIKEWRAENGKPGFKR
Secondary structure
SS confidence
Disorder ????? ????????
Disorder confidence

Confidence Key
High(9) Low (0)

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

Domain analysis

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

Rank	Aligned region
1	d1ayia_
2	d1gxha_
3	c2no8A_
4	d2vlqa1
5	c2v4jA_
6	c3bolB_
7	d3bofa2
8	d2v4ja3
9	c3dljB_
10	c3b9nB_
11	d2f2ga1
12	d1odza_
13	d1ddfa_
14	c3oq9C_
15	d1ltqa1
16	d1na6a2
17	c2qbvA_
18	c3f0nB_
19	d2cbia1
20	c2vx6A_
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1ayia_ ○ ☐	Alignment		100.0	52	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
2	d1gxha_ ○ ☐	Alignment		100.0	45	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
3	c2no8A_ ○ ☐	Alignment		100.0	48	**PDB header:**immune system **Chain:** A: **PDB Molecule:**colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immuntiy protein im2
4	d2vlqa1 ○ ☐	Alignment		100.0	46	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
5	c2v4jA_ ○ ☐	Alignment		49.0	26	**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	c3bolB_ ○ ☐	Alignment		42.9	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

Binding site prediction

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]

© Structural Bioinformatics Group
Imperial College London
Lawrence Kelley, Benjamin Jefferys
Disclaimer
Terms and Conditions

Component software
Template detection: HHpred 1.51
Secondary structure prediction: Psi-pred 2.5
Disorder prediction: Disopred 2.4
Transmembrane prediction: Memsat_SVM
Multi-template modelling and *ab initio*: Poing 1.0

Imperial College London

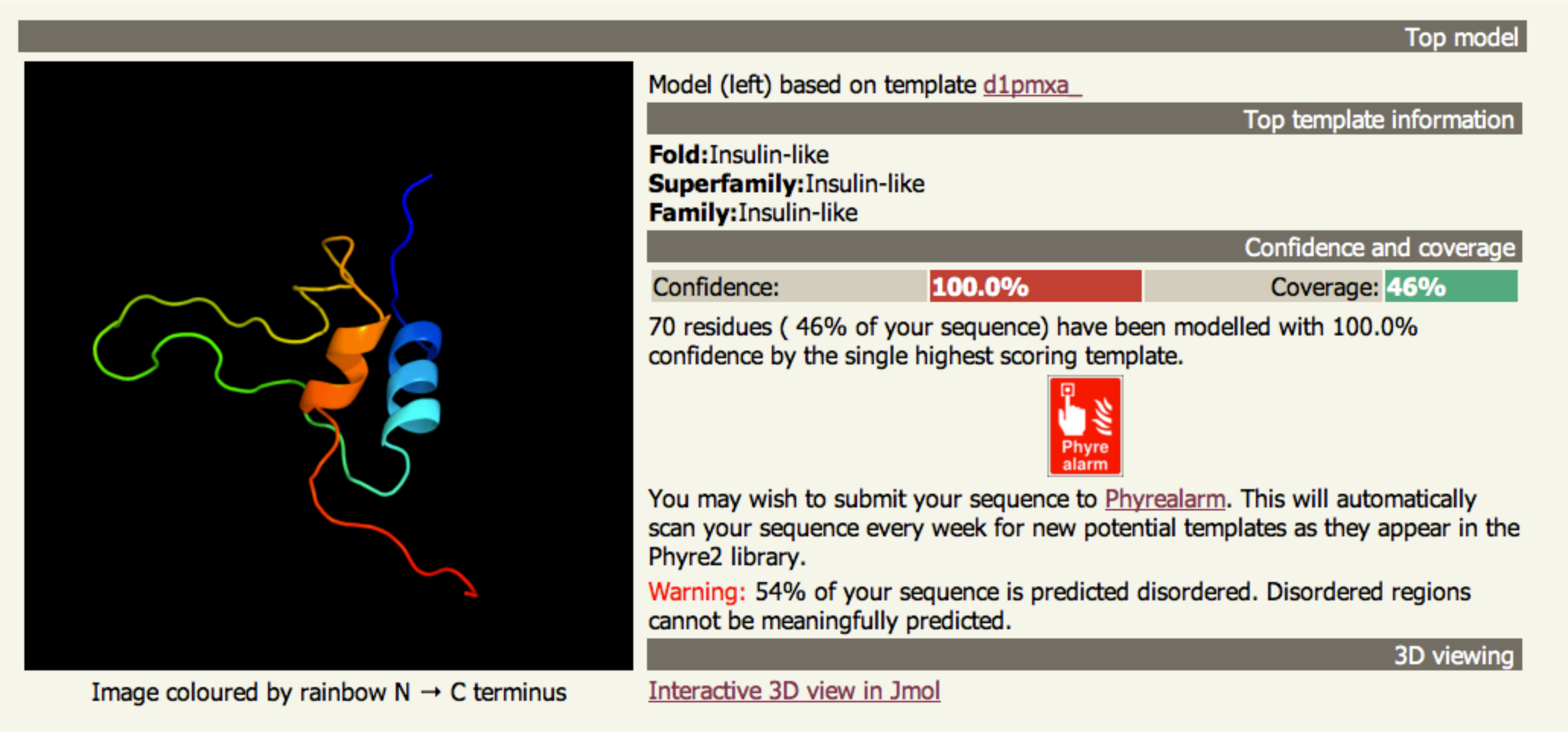
Top model info

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.

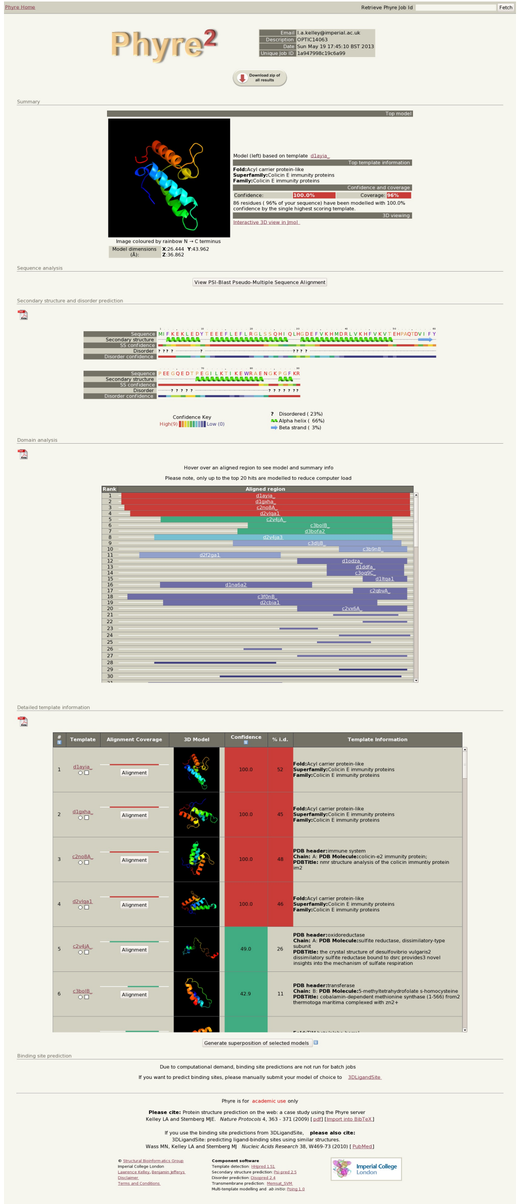
Warning: 54% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

3D viewing

Interactive 3D view in Jmol

Image coloured by rainbow N → C terminus

Example results



Phyre Home

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

Download zip of all results

Summary

Top model

Model (left) based on template d1ayia_

Top template information

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

Confidence and coverage

Confidence: 100.0% Coverage: 96%

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

3D viewing

Interactive 3D view in Jmol

Image coloured by rainbow N → C terminus

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

Sequence analysis

View PSI-Blast Pseudo-Multiple Sequence Alignment

Secondary structure and disorder prediction

Sequence M I F K E K L E D Y T E E E F L E F L R G L S S Q H I Q L H G D E F V K H M D R L V K H F V K V T E H P A Q T D V I F Y
Secondary structure
SS confidence
Disorder ? ? ? ? ? ? ? ? ?
Disorder confidence

Sequence P E E G Q E D T P E G I L K T I K E W R A E N G K P G F K R
Secondary structure
SS confidence
Disorder ? ? ? ? ? ? ? ? ? ? ? ?
Disorder confidence

Confidence Key
High(9) Low (0)

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

Domain analysis

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

Rank	Aligned region
1	d1ayia_
2	d1gxha_
3	c2no8A_
4	d2vlqa1
5	c2v4jA_
6	c3bolB_
7	d3bofa2
8	d2v4ja3
9	c3dljB_
10	c3b9nB_
11	d2f2ga1
12	d1odza_
13	d1ddfa_
14	c3oq9C_
15	d1ltqa1
16	d1na6a2
17	c2qbvA_
18	c3f0nB_
19	d2cbia1
20	c2vx6A_
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1ayia_ ○ ☐	Alignment		100.0	52	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
2	d1gxha_ ○ ☐	Alignment		100.0	45	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
3	c2no8A_ ○ ☐	Alignment		100.0	48	**PDB header:**immune system **Chain:** A: **PDB Molecule:**colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immuntiy protein im2
4	d2vlqa1 ○ ☐	Alignment		100.0	46	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
5	c2v4jA_ ○ ☐	Alignment		49.0	26	**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	c3bolB_ ○ ☐	Alignment		42.9	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

Binding site prediction

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]

© Structural Bioinformatics Group
Imperial College London
Lawrence Kelley, Benjamin Jefferys
Disclaimer
Terms and Conditions

Component software
Template detection: HHpred 1.51
Secondary structure prediction: Psi-pred 2.5
Disorder prediction: Disopred 2.4
Transmembrane prediction: Memsat_SVM
Multi-template modelling and *ab initio*: Poing 1.0

Imperial College London

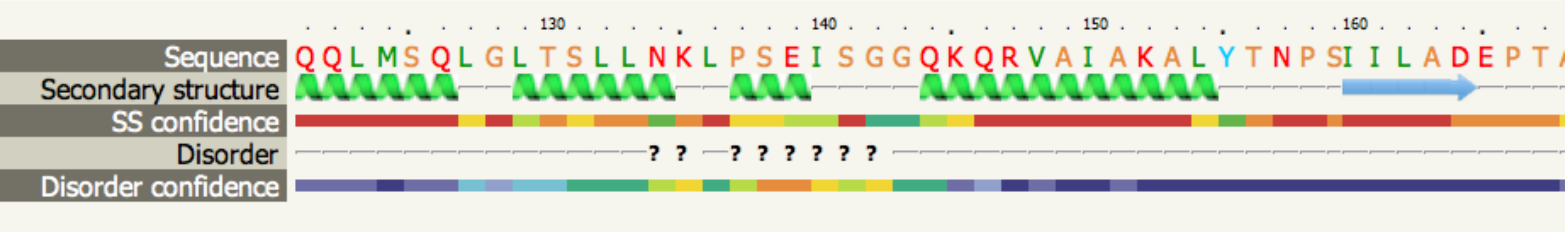
Top model info

Secondary structure/disorder

Domain analysis

Detailed template information

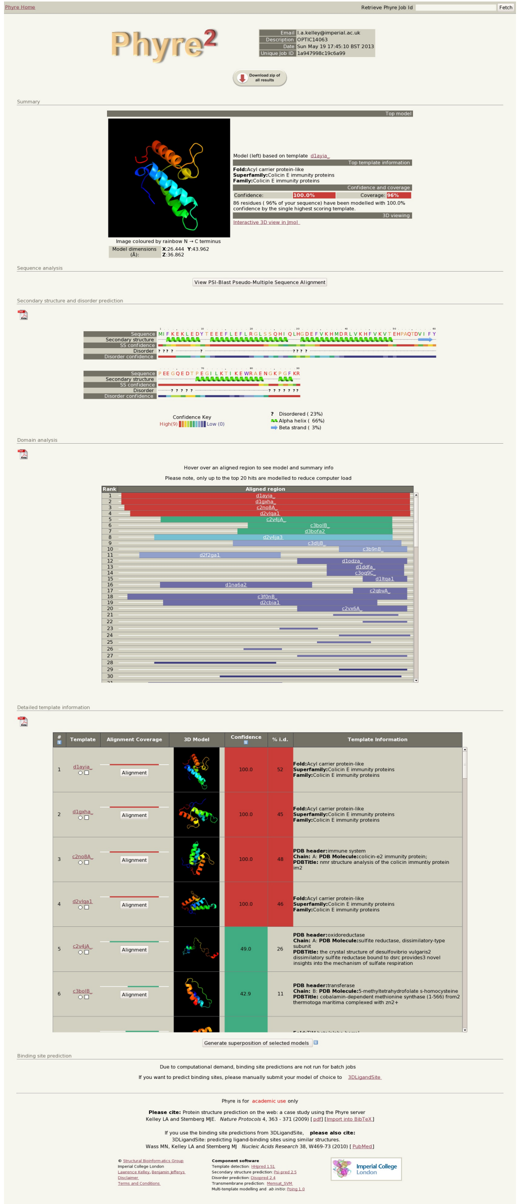
Example SS/disorder prediction



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



Phyre Home
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

Download zip of all results

Summary

Top model

Image coloured by rainbow N → C terminus

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

Model (left) based on template d1ayia_

Top template information

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

Confidence and coverage

Confidence: 100.0% Coverage: 96%

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

3D viewing

Interactive 3D view in Jmol

Sequence analysis

View PSI-Blast Pseudo-Multiple Sequence Alignment

Secondary structure and disorder prediction

Sequence MIFKEKLEDYTEEEFLEFLRGLSSQHIQLHGDEFVKHMDRLVKHFVKVTEHPAQTDVIFY
Secondary structure
SS confidence
Disorder ???? ? ????
Disorder confidence

Sequence PEEGQEDTPEGILKTIKEWRAENGKPGFKR
Secondary structure
SS confidence
Disorder ????? ????????
Disorder confidence

Confidence Key
High(9) Low (0)

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

Domain analysis

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

Rank	Aligned region
1	d1ayia_
2	d1gxha_
3	c2no8A_
4	d2vlqa1
5	c2v4jA_
6	c3bolB_
7	d3bofa2
8	d2v4ja3
9	c3dljB_
10	c3b9nB_
11	d2f2ga1
12	d1odza_
13	d1ddfa_
14	c3oq9C_
15	d1ltqa1
16	d1na6a2
17	c2qbvA_
18	c3f0nB_
19	d2cbia1
20	c2vx6A_
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1ayia_ ○ ☐	Alignment		100.0	52	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
2	d1gxha_ ○ ☐	Alignment		100.0	45	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
3	c2no8A_ ○ ☐	Alignment		100.0	48	**PDB header:**immune system **Chain:** A: **PDB Molecule:**colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immuntiy protein im2
4	d2vlqa1 ○ ☐	Alignment		100.0	46	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
5	c2v4jA_ ○ ☐	Alignment		49.0	26	**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	c3bolB_ ○ ☐	Alignment		42.9	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

Binding site prediction

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]

© Structural Bioinformatics Group
Imperial College London
Lawrence Kelley, Benjamin Jefferys
Disclaimer
Terms and Conditions

Component software
Template detection: HHpred 1.51
Secondary structure prediction: Psi-pred 2.5
Disorder prediction: Disopred 2.4
Transmembrane prediction: Memsat_SVM
Multi-template modelling and *ab initio*: Poing 1.0

Imperial College London

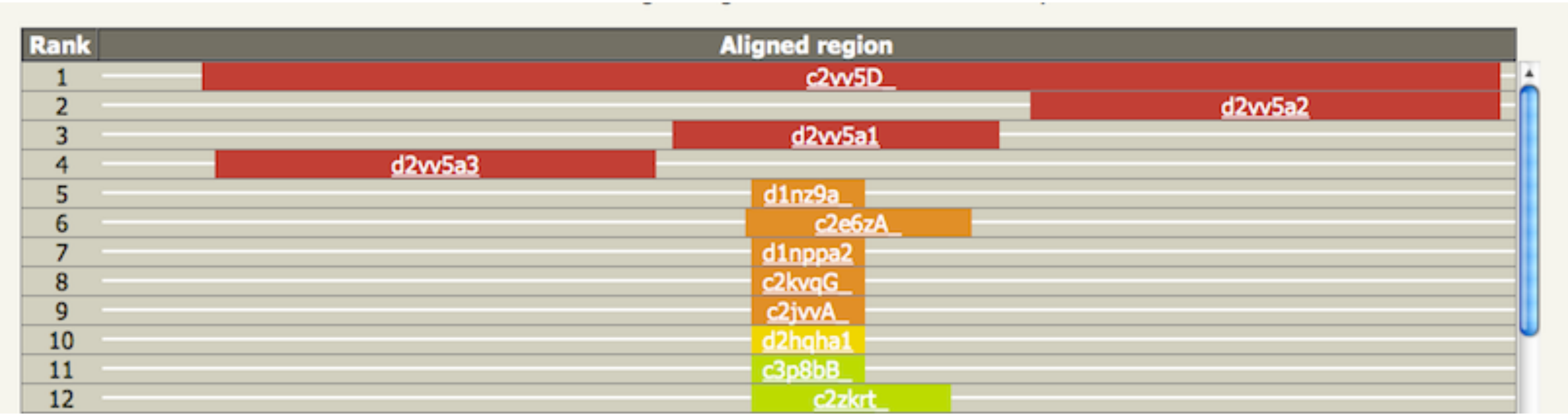
Top model info

Secondary structure/disorder

Domain analysis

Detailed template information

Example domain analysis

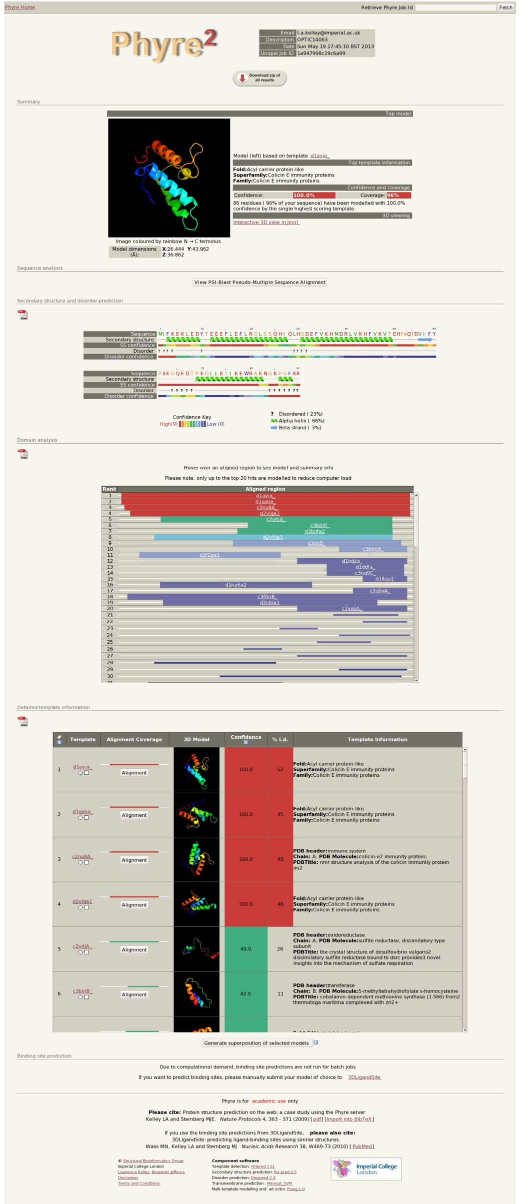


Rank	Aligned region
1	c2vv5D_
2	d2vv5a2
3	d2vv5a1
4	d2vv5a3
5	d1nz9a_
6	c2e6zA_
7	d1nppa2
8	c2kvqG_
9	c2jvvA_
10	d2hqha1
11	c3p8bB_
12	c2zkrt_

Domain analysis

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

Example results



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

Summary

Top model

Model (left) based on template d1ayia_

Top template information

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

Confidence and coverage

Confidence: 100.0% Coverage: 96%

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

3D viewing

Interactive 3D view in Jmol

Image coloured by rainbow N → C terminus

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

Sequence analysis

View PSI-Blast Pseudo-Multiple Sequence Alignment

Secondary structure and disorder prediction

Sequence MIFKEKLEDYTEEEFLEFLRGLSSQHIQLHGDEFVKHMDRLVKHFVKVTEHPAQTDVIFY
Secondary structure
SS confidence
Disorder ? ? ? ? ? ? ? ? ?
Disorder confidence

Sequence PEEGQEDTPEGILKTIKEWRAENGKPGFKR
Secondary structure
SS confidence
Disorder ? ? ? ? ? ? ? ? ? ? ? ? ?
Disorder confidence

Confidence Key
High(9) Low (0)

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

Domain analysis

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

Rank	Aligned region
1	d1ayia_
2	d1gxha_
3	c2no8A_
4	d2vlqa1
5	c2v4jA_
6	c3bolB_
7	d3bofa2
8	d2v4ja3
9	c3dljB_
10	c3b9nB_
11	d2f2ga1
12	d1odza_
13	d1ddfa_
14	c3oq9C_
15	d1ltqa1
16	d1na6a2
17	c2qbvA_
18	c3f0nB_
19	d2cbia1
20	c2vx6A_
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	d1ayia_ ○ ☐	Alignment		100.0	52	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
2	d1gxha_ ○ ☐	Alignment		100.0	45	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
3	c2no8A_ ○ ☐	Alignment		100.0	48	**PDB header:**immune system **Chain:** A: **PDB Molecule:**colicin-e2 immunity protein; **PDBTitle:** nmr structure analysis of the colicin immuntiy protein im2
4	d2vlqa1 ○ ☐	Alignment		100.0	46	**Fold:**Acyl carrier protein-like **Superfamily:**Colicin E immunity proteins **Family:**Colicin E immunity proteins
5	c2v4jA_ ○ ☐	Alignment		49.0	26	**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	c3bolB_ ○ ☐	Alignment		42.9	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

Binding site prediction

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]

© Structural Bioinformatics Group
Imperial College London
Lawrence Kelley, Benjamin Jefferys
Disclaimer
Terms and Conditions

Component software
Template detection: HHpred 1.51
Secondary structure prediction: Psi-pred 2.5
Disorder prediction: Disopred 2.4
Transmembrane prediction: Memsat_SVM
Multi-template modelling and *ab initio*: Poing 1.0

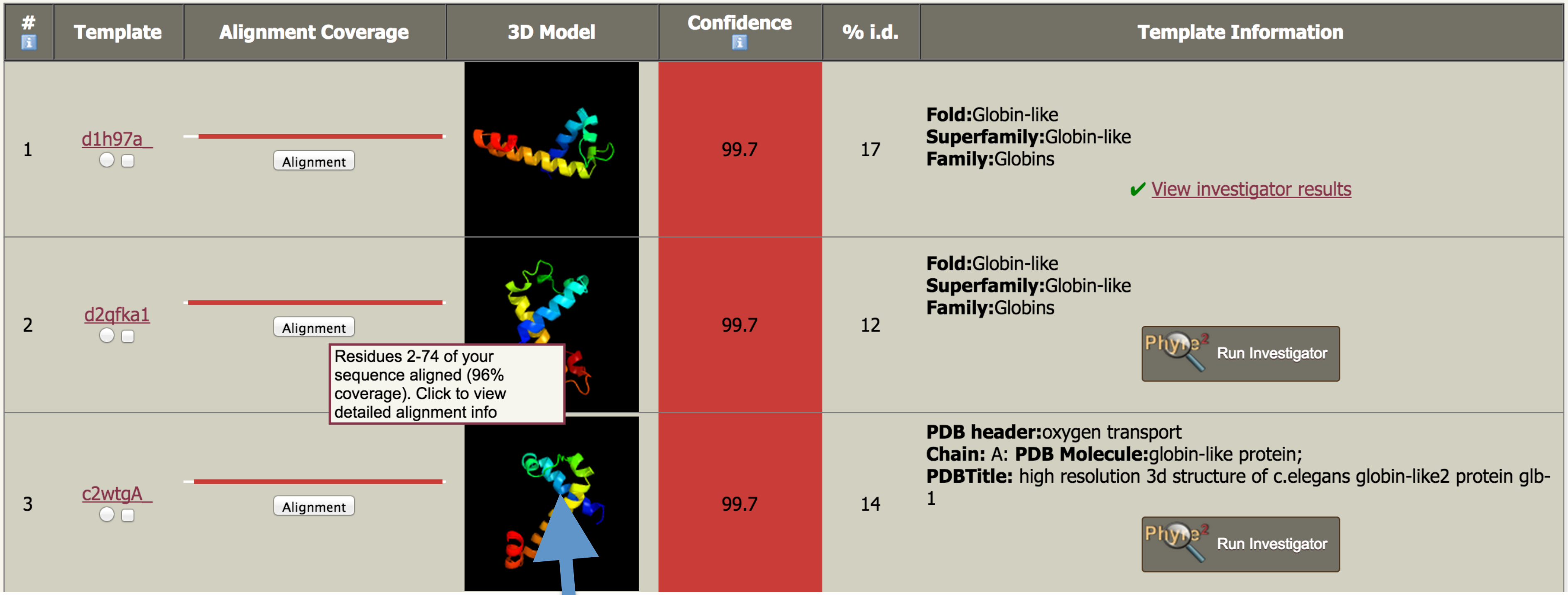
Top model info

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 Phyre2 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 Phyre2 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

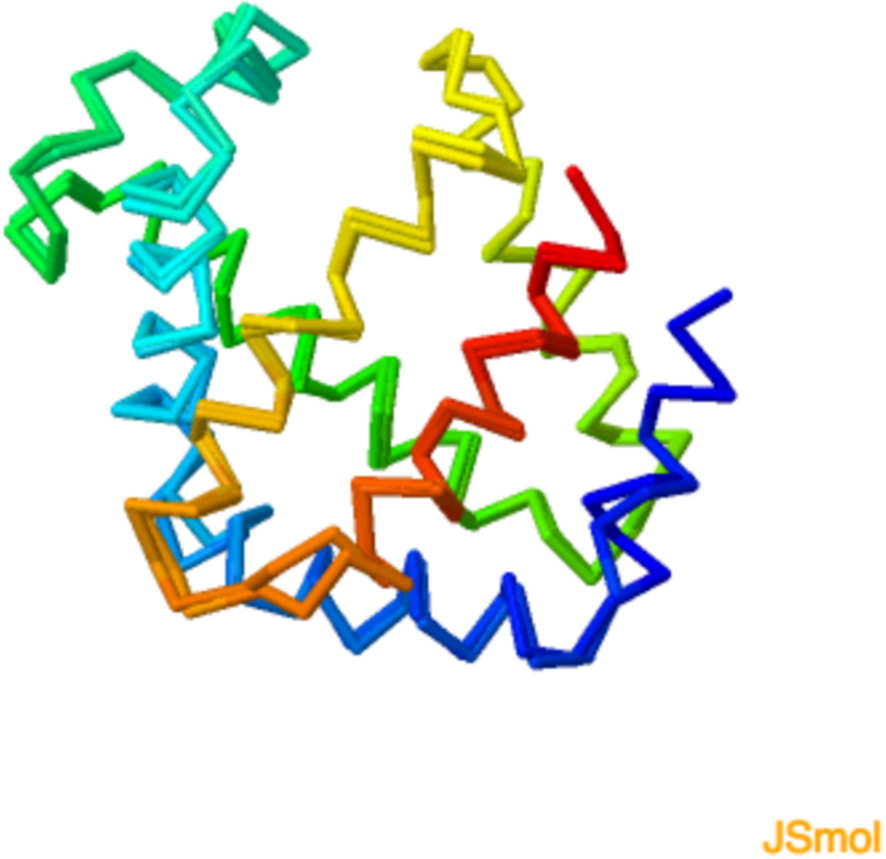
Sequence identity and model accuracy

- 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

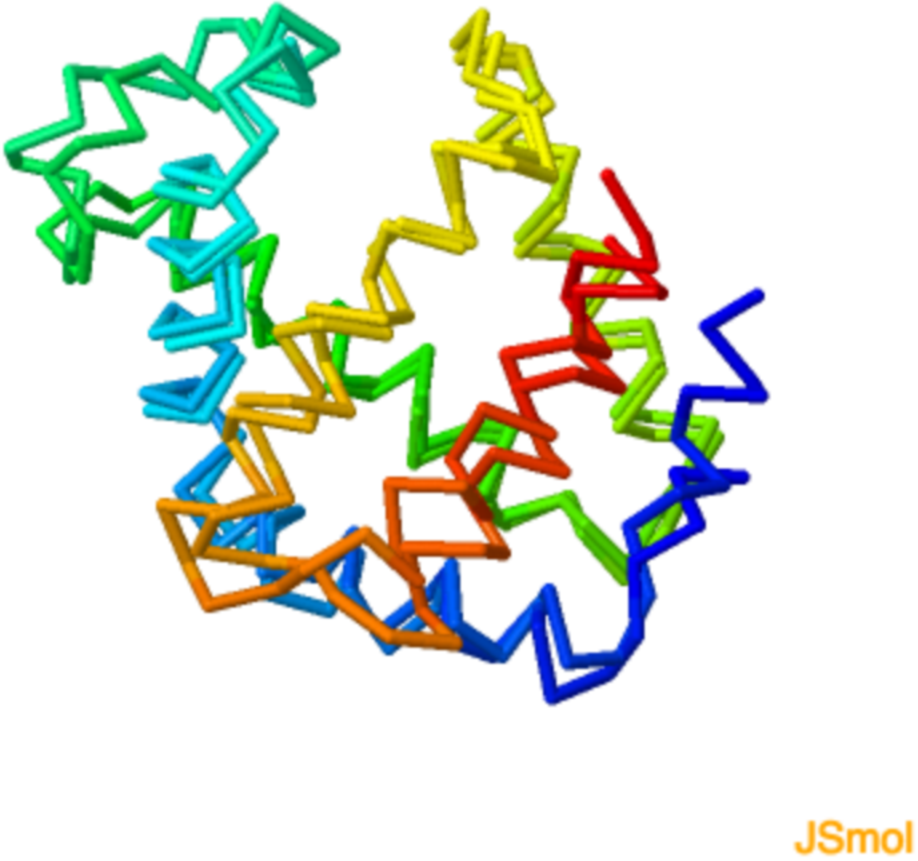
100% confidence,

56% sequence identity, TM-score 0.9



Interpreting results

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

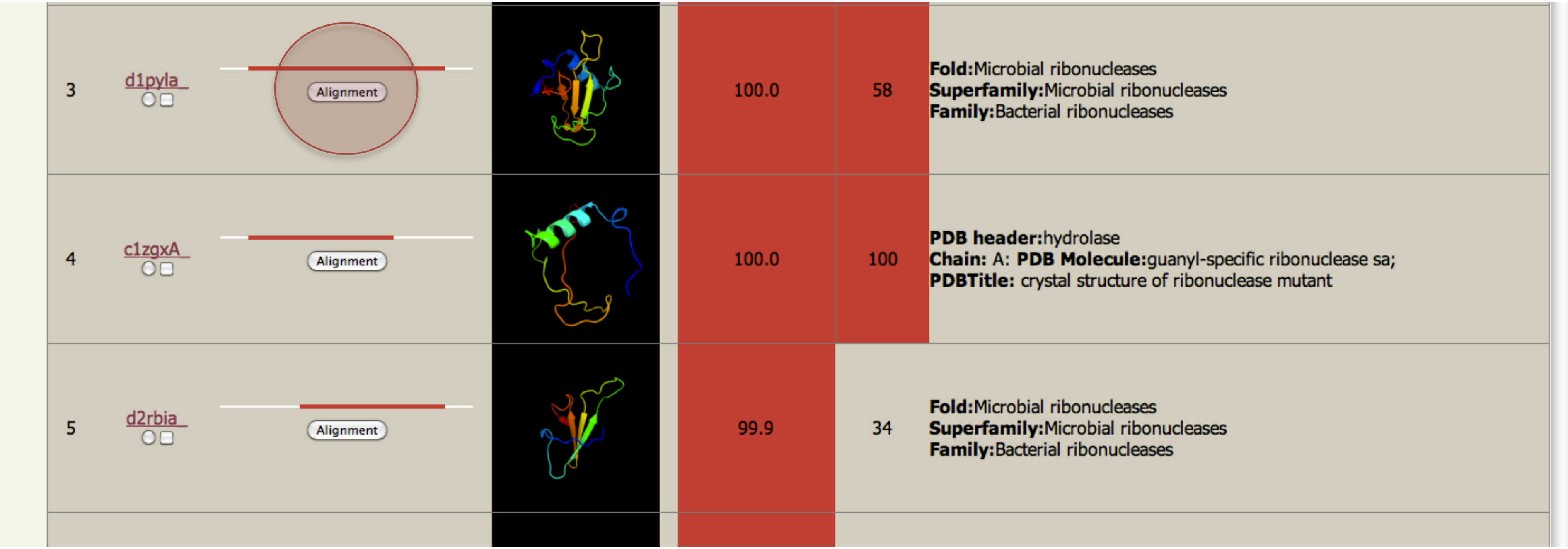


Interpreting results

Checklist

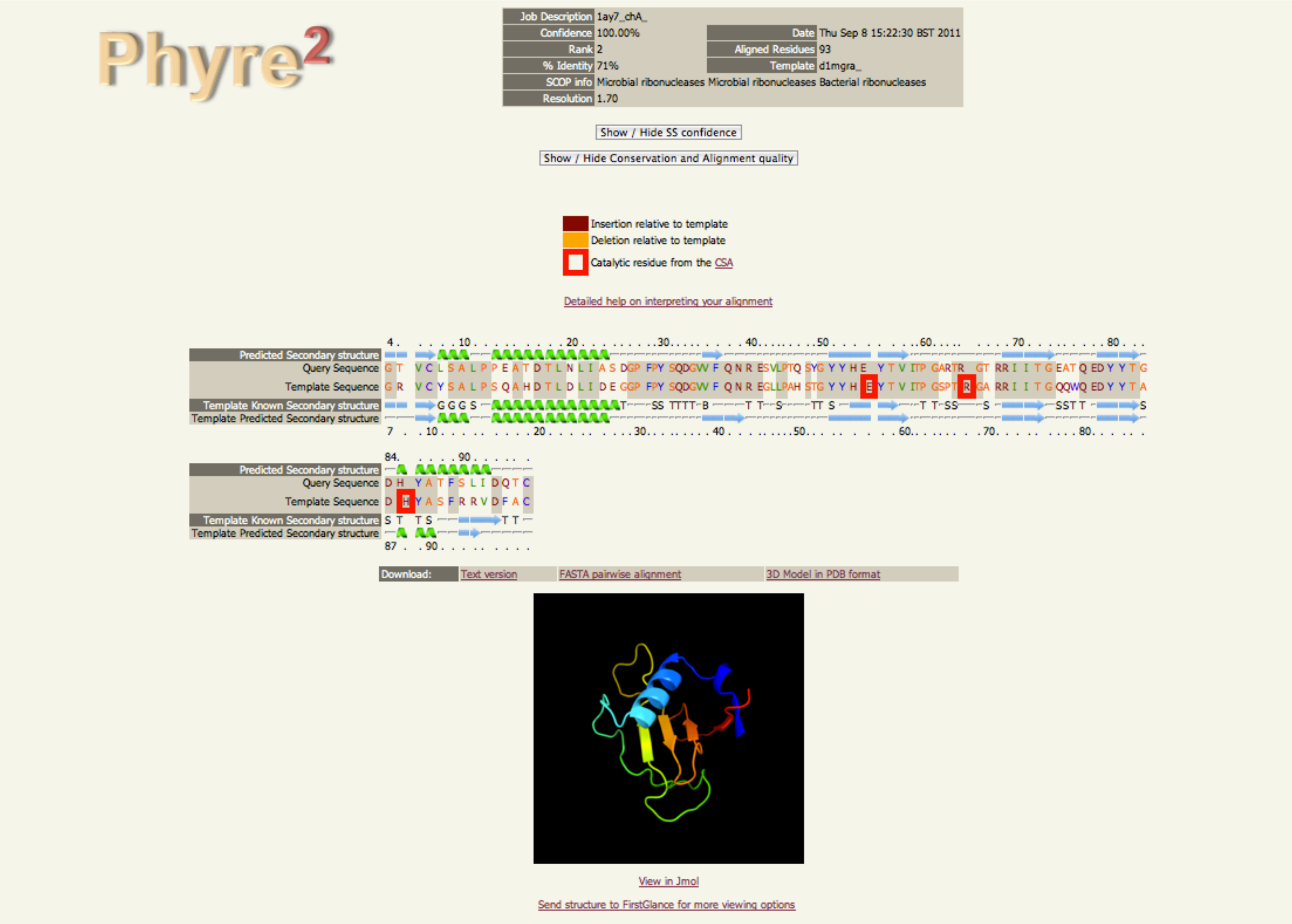
- Look at confidence
- Given multiple high confidence hits, look at % sequence identity
- Biological knowledge relating function of template to sequence of interest
- Structural superpositions to compare models – many similar models increase confidence
- Examine sequence alignment

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



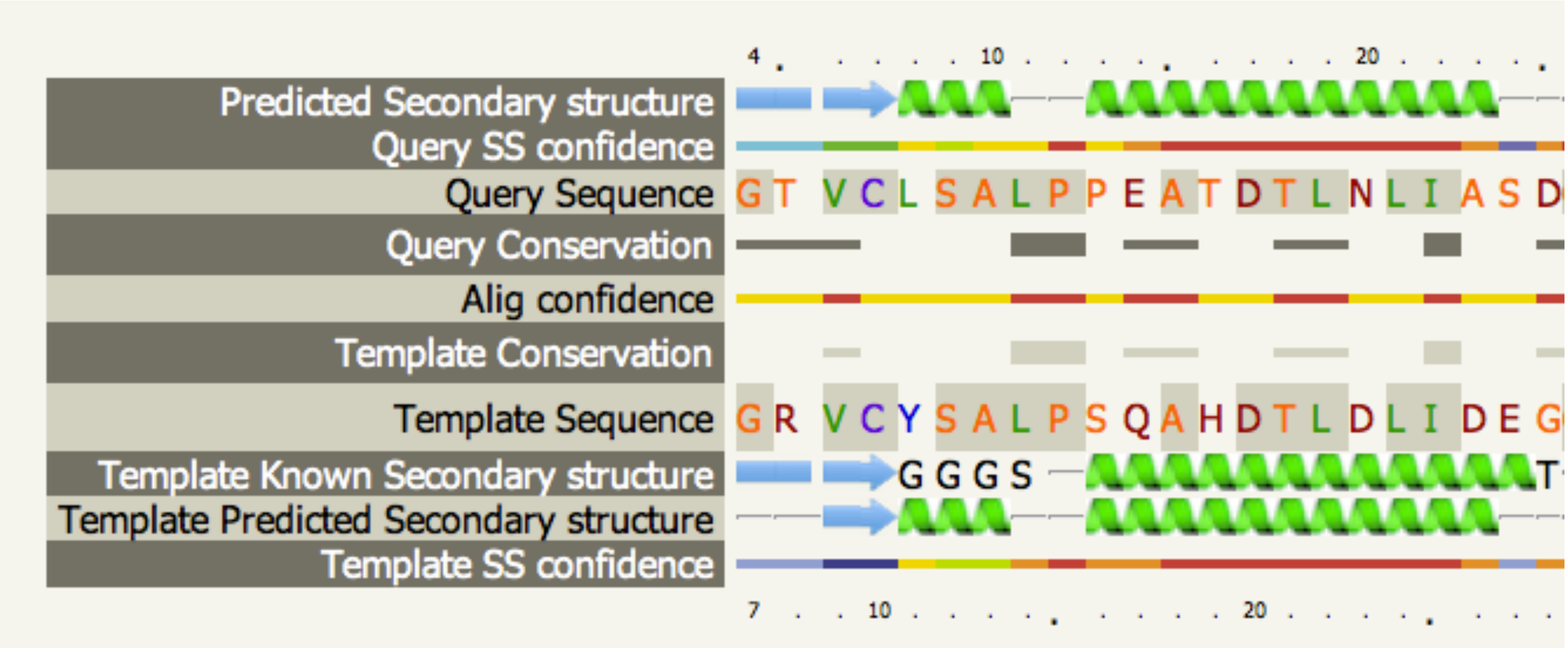
Job Description	1ay7_chA_		
Confidence	100.00%	Date	Thu Sep 8 15:22:30 BST 2011
Rank	2	Aligned Residues	93
% Identity	71%	Template	d1mgra_
SCOP info	Microbial ribonucleases Microbial ribonucleases Bacterial ribonucleases		
Resolution	1.70		

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

Query Sequence: GT VCLSALPPEATDTLNLIASDGP FPY SQDGW F QNR ESVLPTQ SYG YYHE YTV ITP GARTR GT RRIITG EAT Q EDYYTG
Template Sequence: GR VCYSALPSQAHDTLDLIDEGGP FPY SQDGW F QNR EGLLPAH STG YYHE YTV ITP GSPTR GA RRIITG QQWQ EDYYTA
Template Known Secondary structure: GGGS T SS TTTT B T T S TT S T T SS S SST T S

Query Sequence: DH YATFSLIDQTC
Template Sequence: DHYASFRRVDFAC
Template Known Secondary structure: S T T S T T

Alignment view



Predicted Secondary structure
Query SS confidence
Query Sequence: G T V C L S A L P P E A T D T L N L I A S D
Query Conservation
Alig confidence
Template Conservation
Template Sequence: G R V C Y S A L P S Q A H D T L D L I D E G
Template Known Secondary structure: G G G S T
Template Predicted Secondary structure
Template SS confidence

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

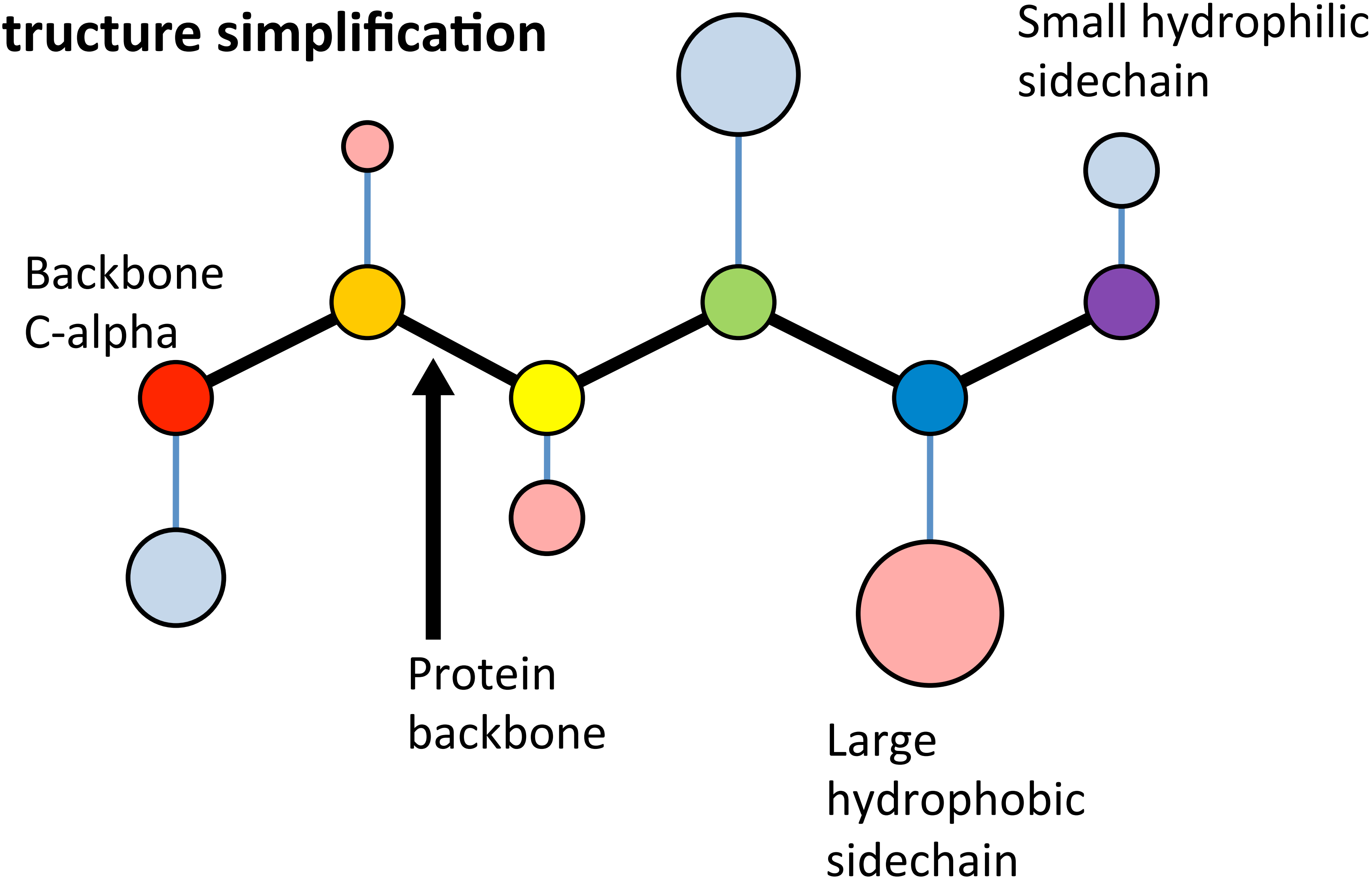
Shortcomings of ‘normal’ Mode

- 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

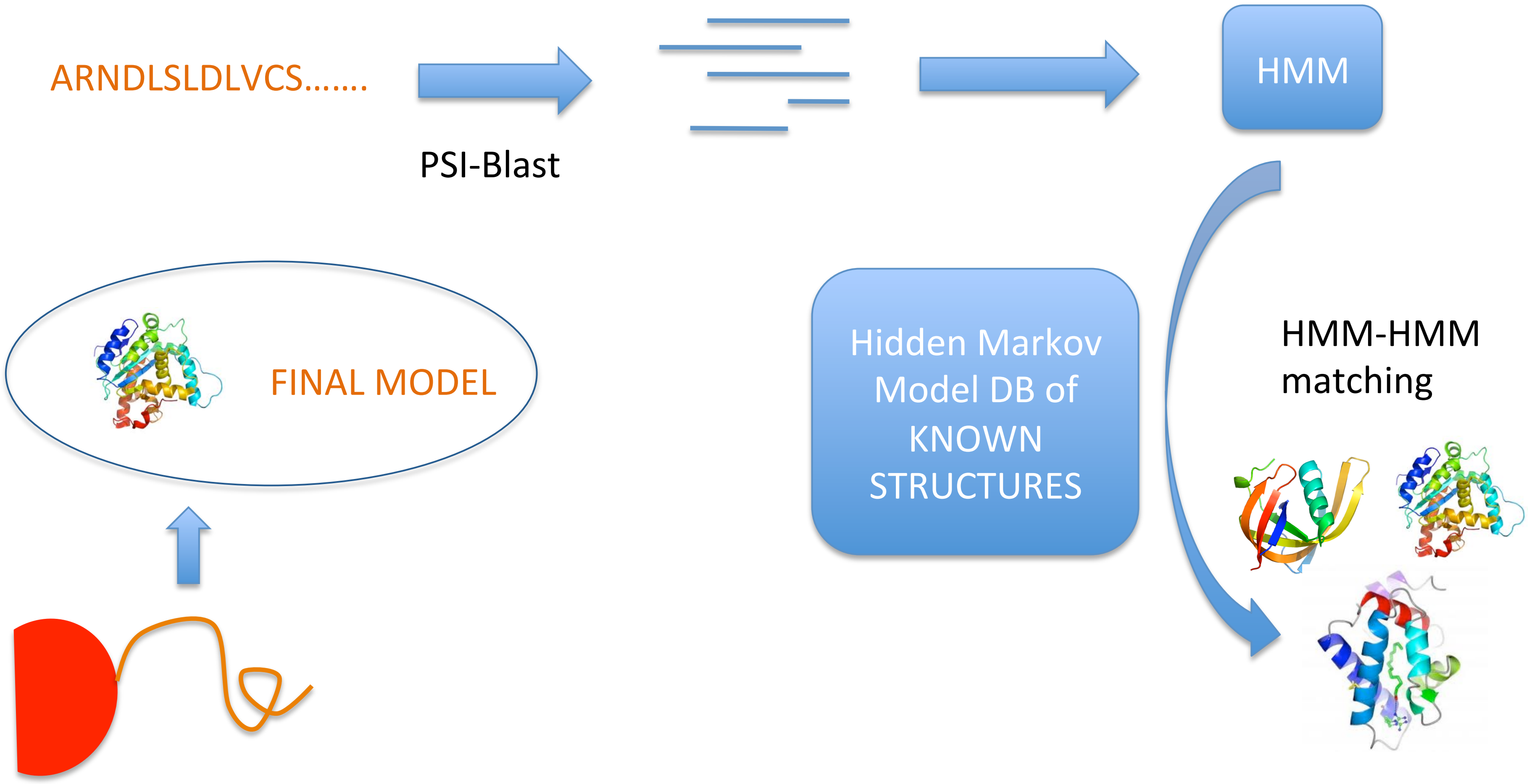
Thus need a system to fold a protein without templates and combine templates when we have them

Poing – simplified folding model

structure simplification



Phyre + Poing



ARNDLSLDLVCS…….

PSI-Blast

HMM

HMM-HMM matching

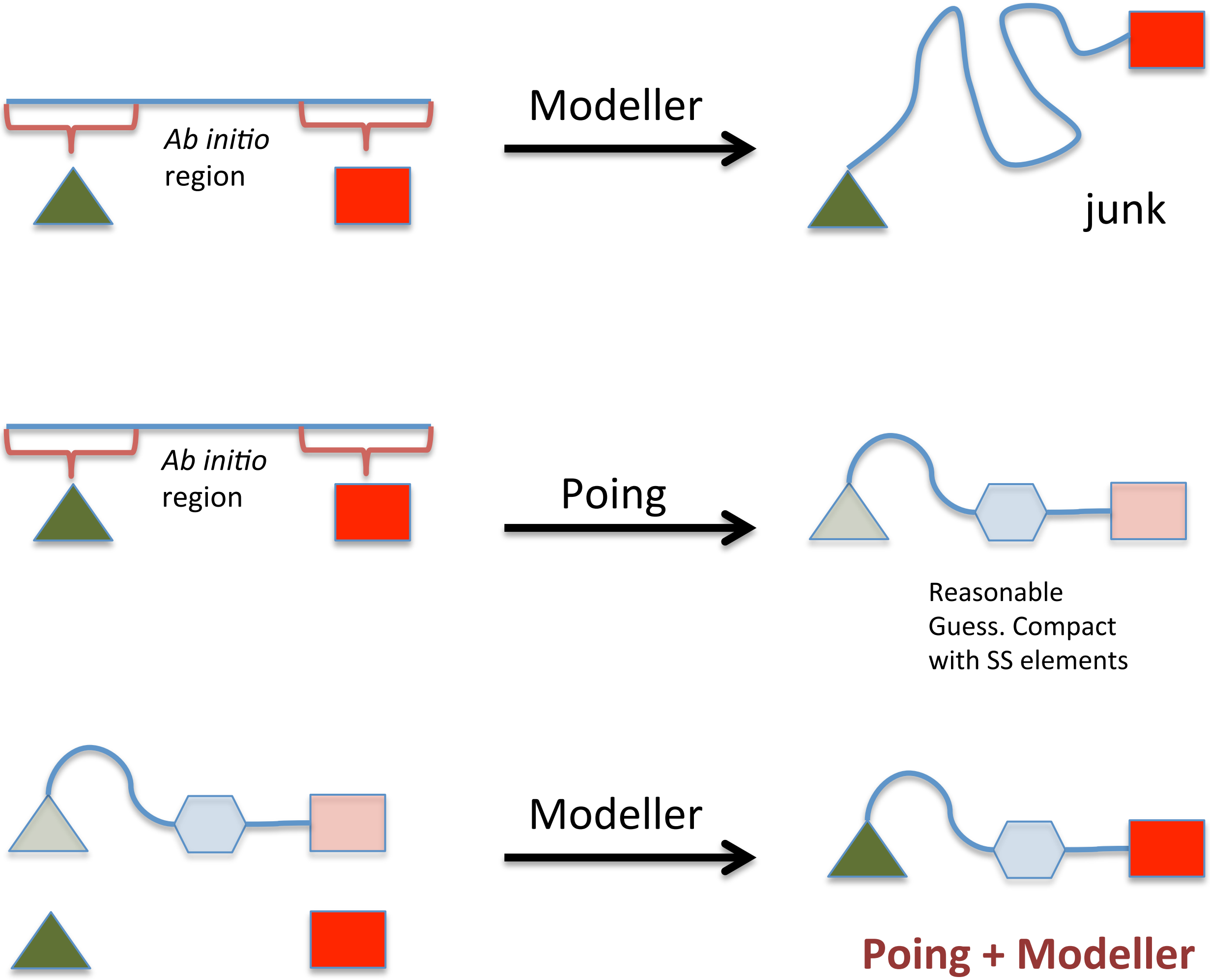
Hidden Markov Model DB of KNOWN STRUCTURES

FINAL MODEL

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

Extract pairwise distance constraints

Phyre + Poing + Modeller



Ab initio region

Modeller

junk

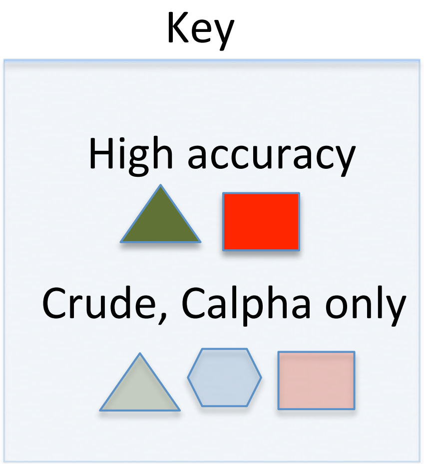
Ab initio region

Poing

Reasonable Guess. Compact with SS elements

Modeller

Poing + Modeller



Key

High accuracy

Crude, Calpha only

Maintains detail in confident regions whilst creating 'reasonable' *ab initio* regions

Intensive mode

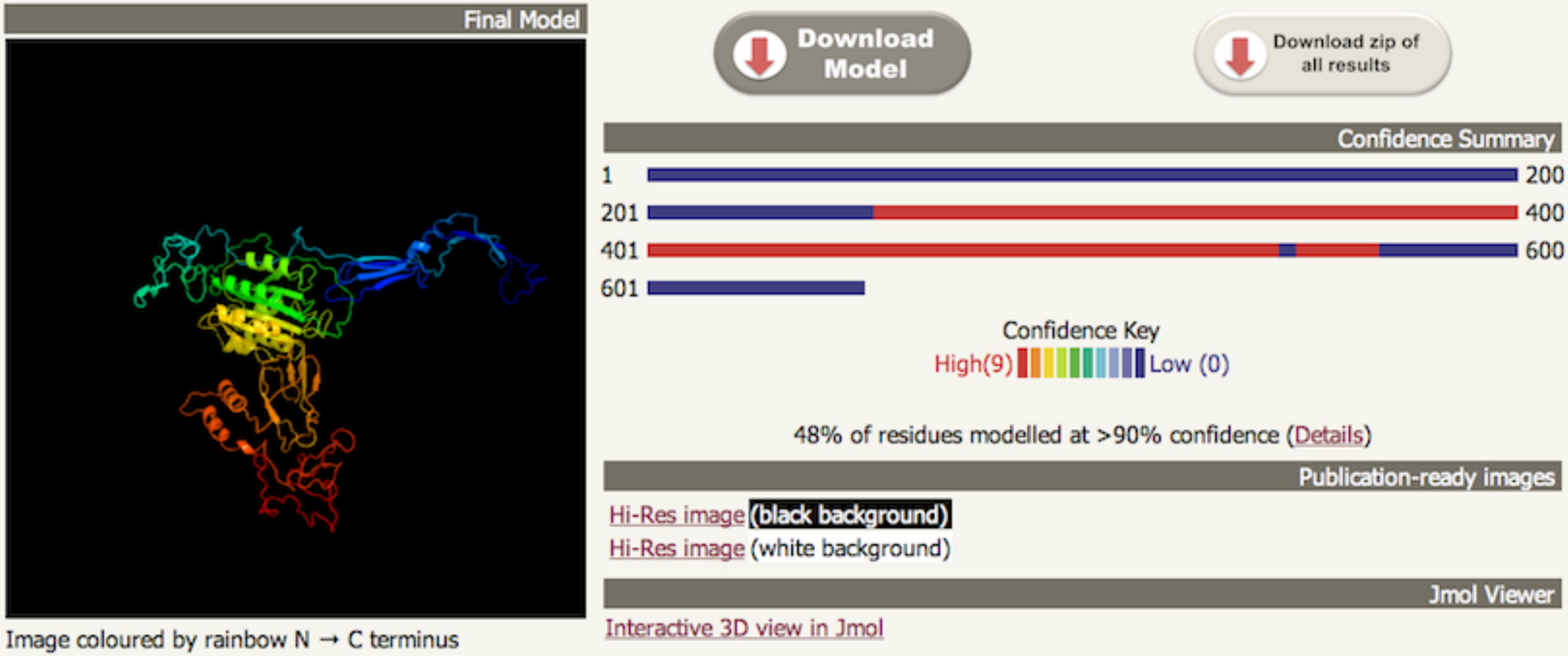


Image coloured by rainbow N → C terminus

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

Hi-Res image (black background)

Hi-Res image (white background)

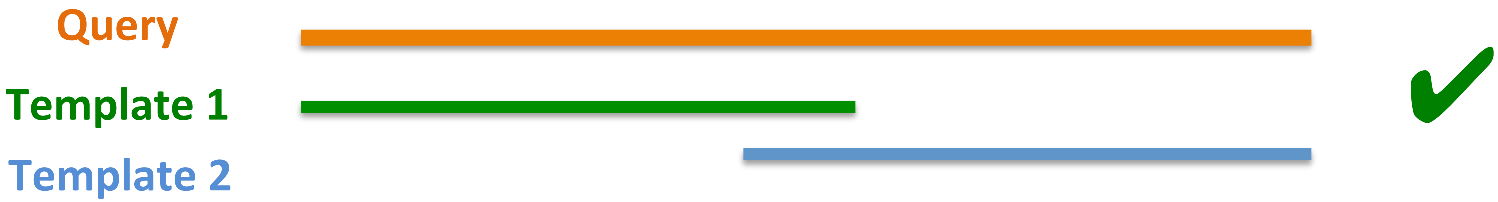
Interactive 3D view in Jmol

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.

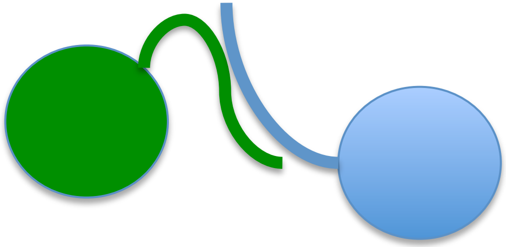


Query

Template 1

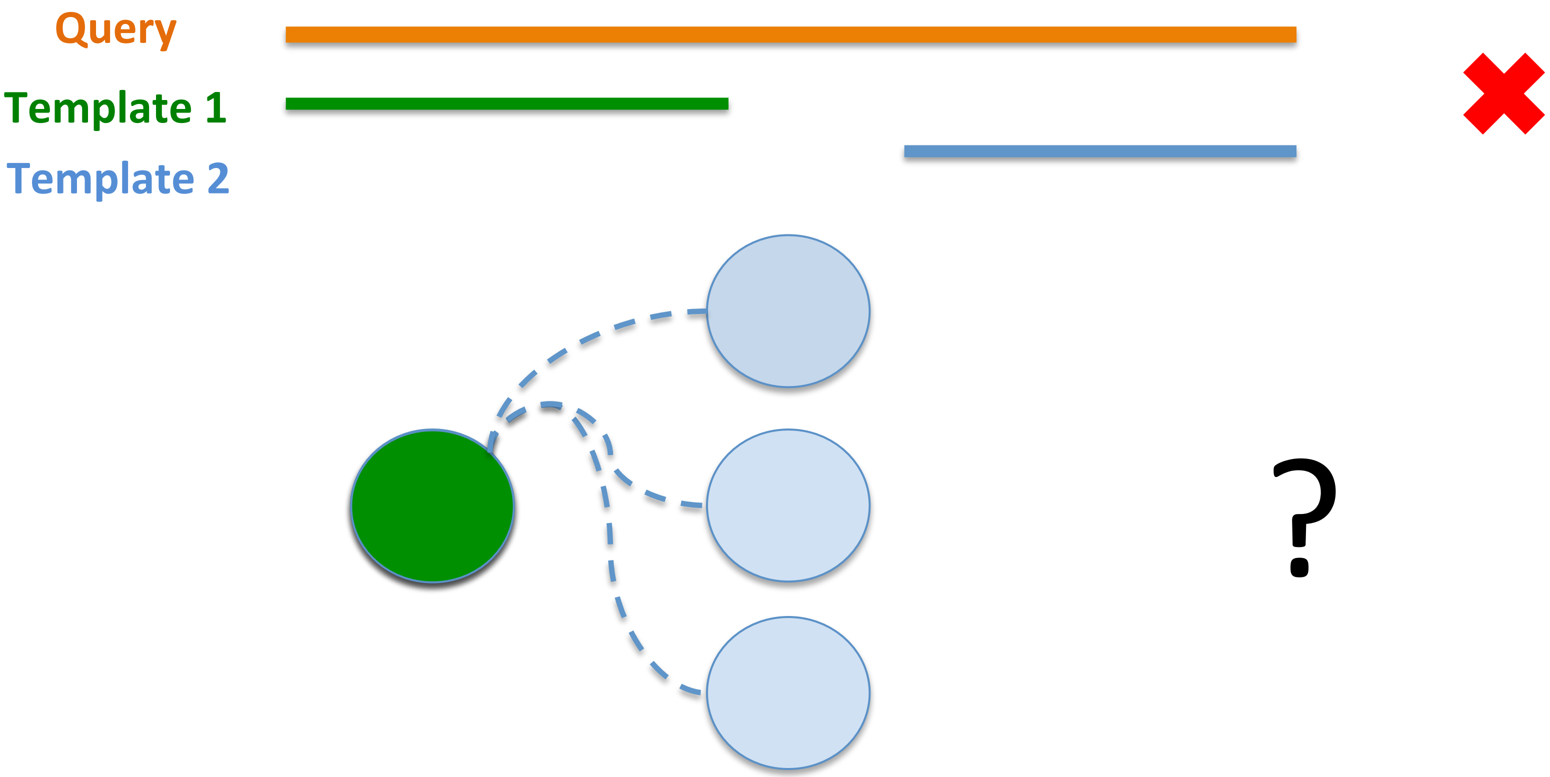
Template 2

✓



Intensive mode

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



Intensive mode

- B-factors in final model indicate ab initio vs. template-based modelled regions
- Can be slow on large proteins and limited to 1,000 residues.
- Under active development – improved version available in a few months.

Intensive mode

“Intensive” does not always equal “Better”!

Checklist

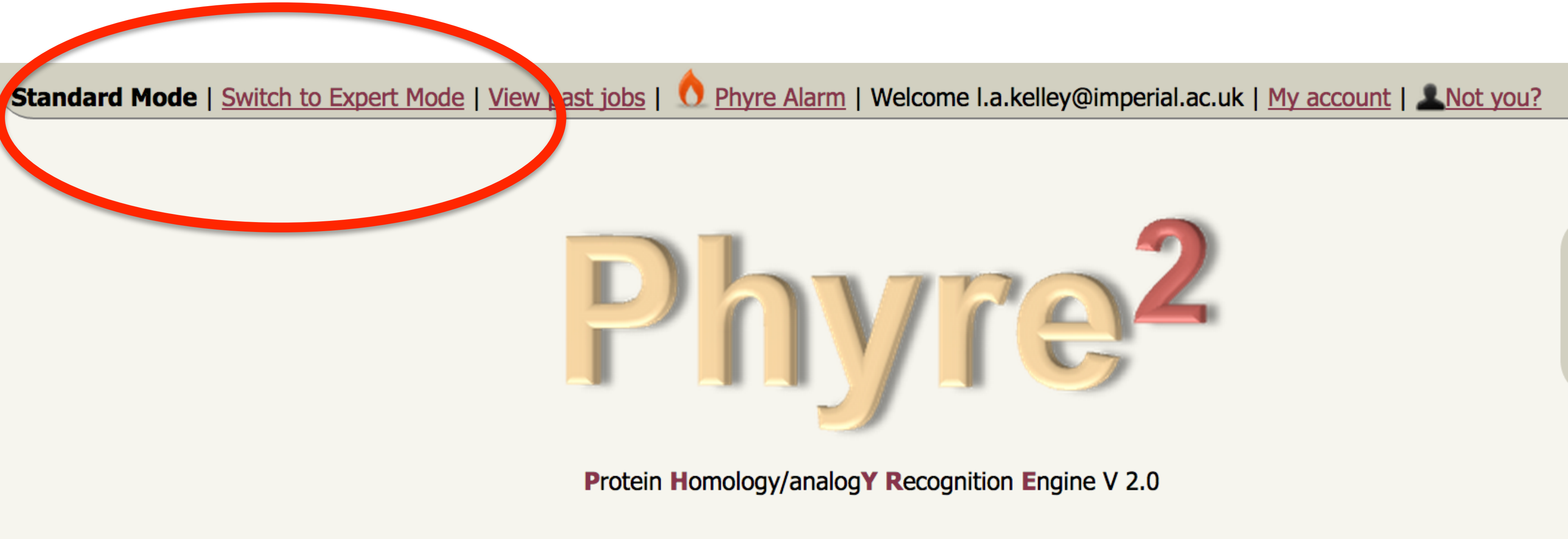
- Always use normal mode first to understand what regions can be well modelled
- Multiple overlapping high confidence domains? Good, try intensive. Otherwise skip it.
- Danger of “spaghettification”
- Active development, new version ‘soon’

How does Phyre2 work?

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

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?

Protein **H**omology/analog**Y** **R**ecognition **E**ngine V 2.0

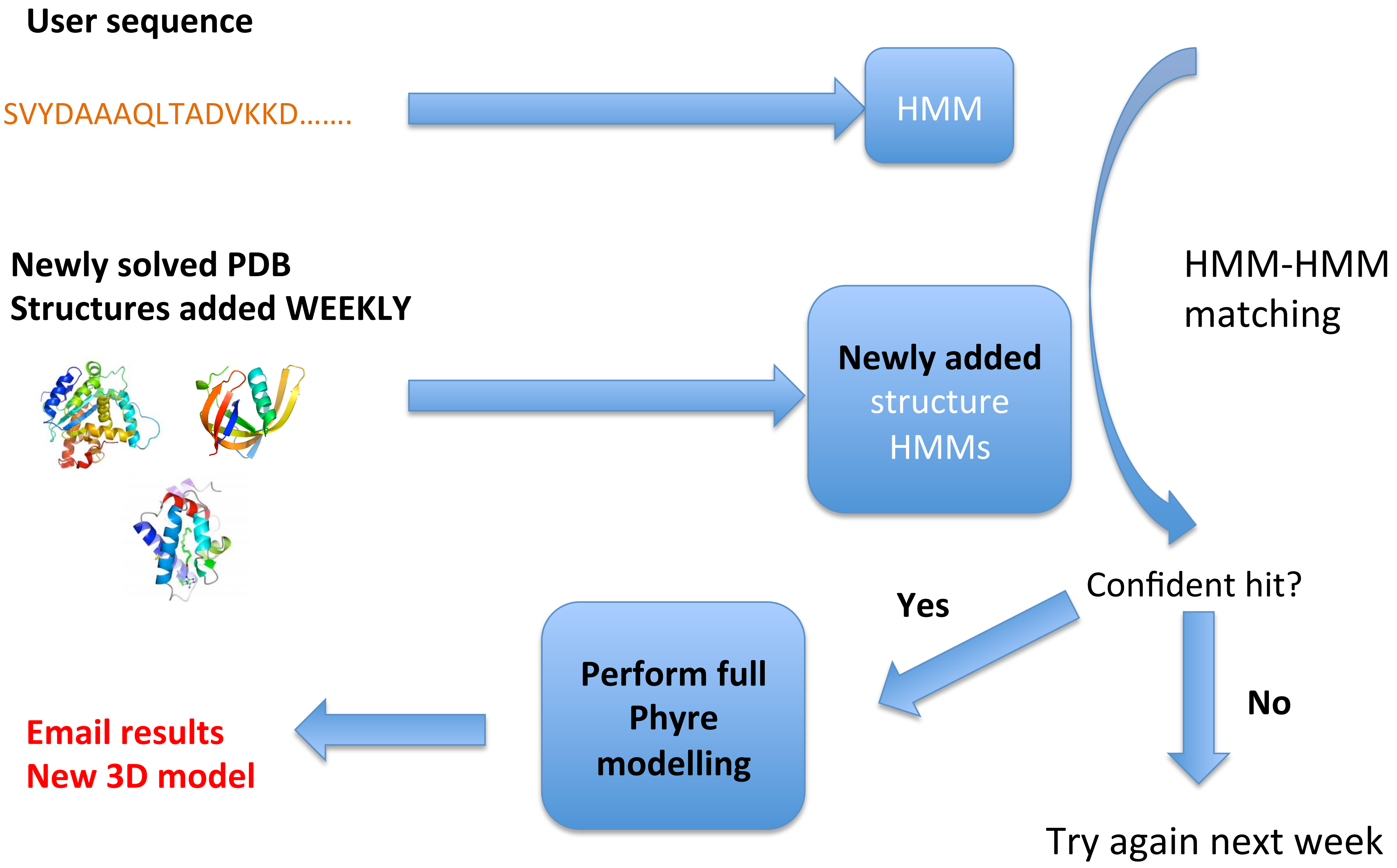
Advanced functions

- 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

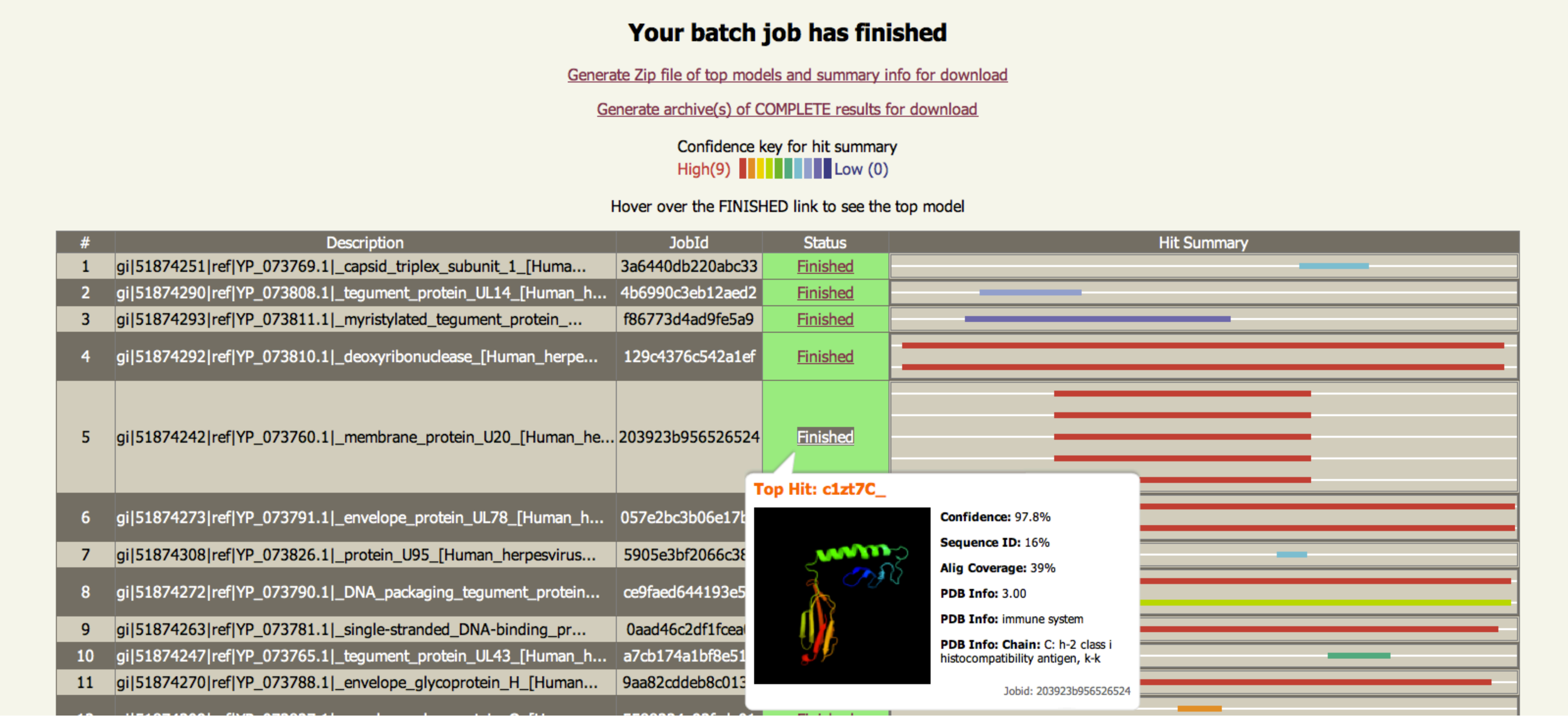
Advanced functions

- 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

Batch Jobs

- Only Normal mode for speed considerations
- Sequences are processed more slowly than individual submissions to maintain user experience
- Batch job progress can be monitored, view intermediate results
- 100 Sequences by default. Use the My Account page to request an increase (up to 1-2k)

Batch Jobs



Your batch job has finished

Generate Zip file of top models and summary info for download

Generate archive(s) of COMPLETE results for download

Confidence key for hit summary

High(9) 🟥🟧🟨🟨🟩🟩🟦🟦🟪🟪 Low (0)

Hover over the FINISHED link to see the top model

#	Description	JobId	Status	Hit Summary
1	gi|51874251|ref|YP_073769.1|_capsid_triplex_subunit_1_[Huma...	3a6440db220abc33	Finished	
2	gi|51874290|ref|YP_073808.1|_tegument_protein_UL14_[Human_h...	4b6990c3eb12aed2	Finished	
3	gi|51874293|ref|YP_073811.1|_myristylated_tegument_protein_...	f86773d4ad9fe5a9	Finished	
4	gi|51874292|ref|YP_073810.1|_deoxyribonuclease_[Human_herpe...	129c4376c542a1ef	Finished	
5	gi|51874242|ref|YP_073760.1|_membrane_protein_U20_[Human_he...	203923b956526524	Finished	
6	gi|51874273|ref|YP_073791.1|_envelope_protein_UL78_[Human_h...	057e2bc3b06e17		
7	gi|51874308|ref|YP_073826.1|_protein_U95_[Human_herpesvirus...	5905e3bf2066c3		
8	gi|51874272|ref|YP_073790.1|_DNA_packaging_tegument_protein...	ce9faed644193e5		
9	gi|51874263|ref|YP_073781.1|_single-stranded_DNA-binding_pr...	0aad46c2df1fcea		
10	gi|51874247|ref|YP_073765.1|_tegument_protein_UL43_[Human_h...	a7cb174a1bf8e51		
11	gi|51874270|ref|YP_073788.1|_envelope_glycoprotein_H_[Human...	9aa82cddeb8c013		

Top Hit: c1zt7C_

Confidence: 97.8%

Sequence ID: 16%

Alig Coverage: 39%

PDB Info: 3.00

PDB Info: immune system

PDB Info: Chain: C: h-2 class i histocompatibility antigen, k-k

Jobid: 203923b956526524

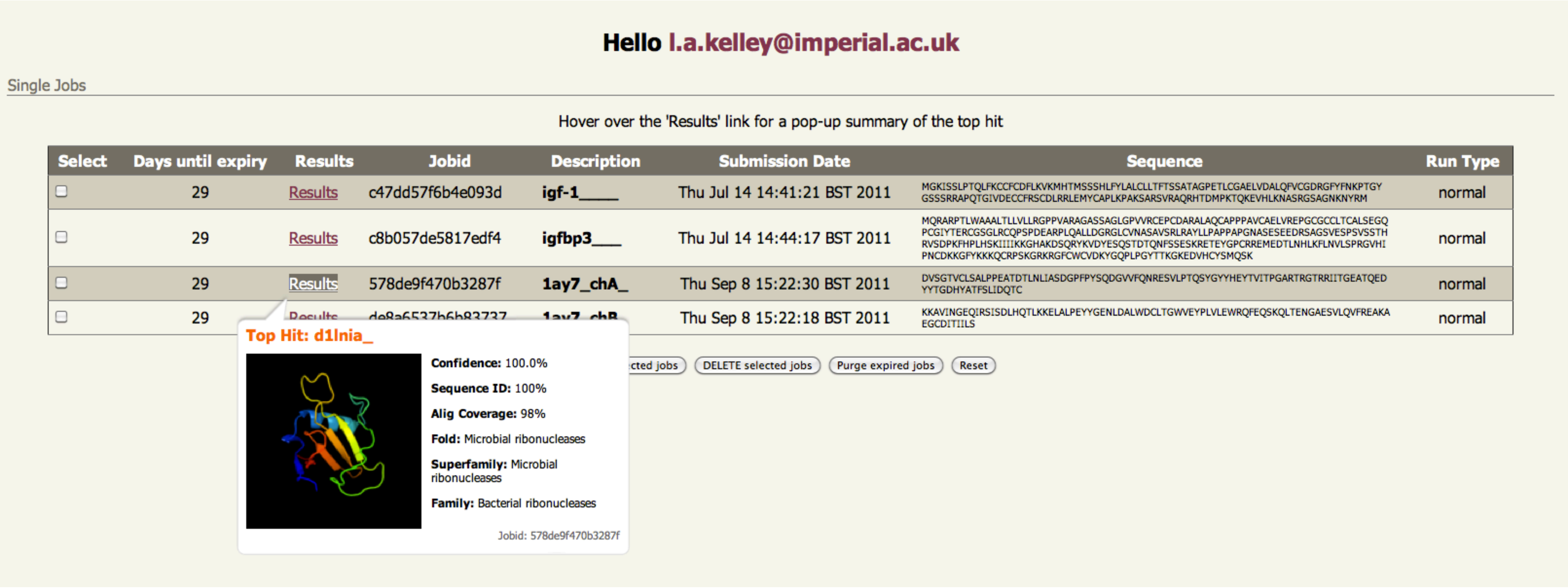
Advanced functions

- 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

Job manager

- Only available if **logged in**
- Shows complete job history, sequence, top hit summary
- Allows deletion and renewal of jobs (they expire after 30 days)

Job manager



Hello l.a.kelley@imperial.ac.uk

Single Jobs

Hover over the 'Results' link for a pop-up summary of the top hit

Select	Days until expiry	Results	Jobid	Description	Submission Date	Sequence	Run Type
☐	29	Results	c47dd57f6b4e093d	**igf-1_____**	Thu Jul 14 14:41:21 BST 2011	MGKISSLPTQLFKCCFCDFLKVKMHTMSSSHLFYLALCLLTFTSSATAGPETLCGAELVDALQFVCGDRGFYFNKPTGY GSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSARSVRAQRHTDMPKTQKEVHLKNASRGSAGNKNYRM	normal
☐	29	Results	c8b057de5817edf4	**igfbp3____**	Thu Jul 14 14:44:17 BST 2011	MQRARPTLWAAALTLLVLLRGPPVARAGASSAGLGPVVRCEPCDARALAQCAPPPAVCAELVREPGCGCCLTCALSEGQ PCGIYTERCGSGLRCQPSPDEARPLQALLDGRGLCVNASAVSRLRAYLLPAPPAPGNASESEEDRSAGSVESPSVSSTH RVSDPKFHPLHSKIIIIKKGHAKDSQRYKVDYESQSTDTQNFSSESKRETEYGPCRREMEDTLNHLKFLNVLSPRGVHI PNCDKKGFYKKKQCRPSKGRKRGFCWCVDKYGQPLPGYTTKGKEDVHCYSMQSK	normal
☐	29	Results	578de9f470b3287f	**1ay7_chA_**	Thu Sep 8 15:22:30 BST 2011	DVSGTVCLSALPPEATDTLNLIASDGPFPYSQDGVVFQNRESVLPTQSYGYYHEYTVITPGARTRGTRRIITGEATQED YYTGDHYATFSLIDQTC	normal
☐	29	Results	de8a6537b6b83737	**1ay7_chB**	Thu Sep 8 15:22:18 BST 2011	KKAVINGEQIRSISDLHQTLKKELALPEYYGENLDALWDCLTGWVEYPLVLEWRQFEQSKQLTENGAESVLQVFREAKA EGCDITIILS	normal

Top Hit: d1lnia_

Confidence: 100.0%

Sequence ID: 100%

Alig Coverage: 98%

Fold: Microbial ribonucleases

Superfamily: Microbial ribonucleases

Family: Bacterial ribonucleases

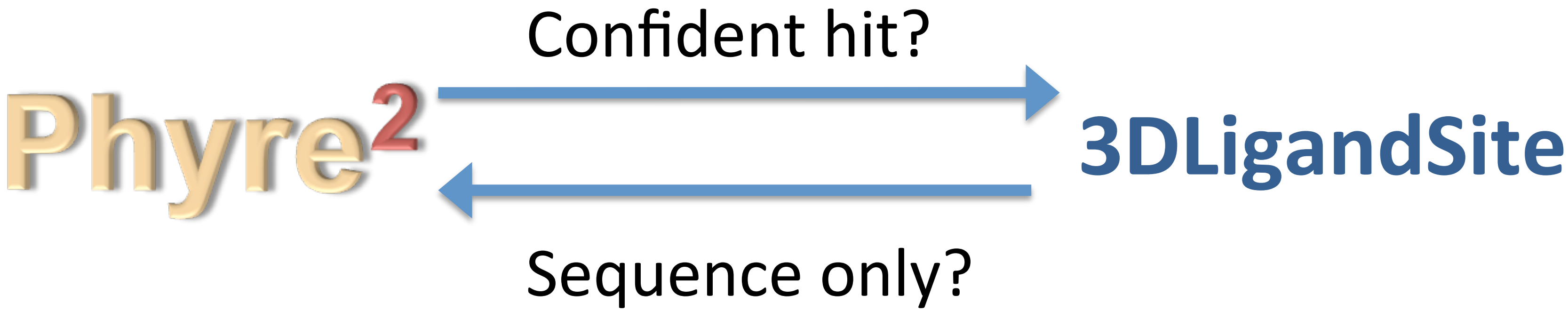
Jobid: 578de9f470b3287f

3DLigandSite

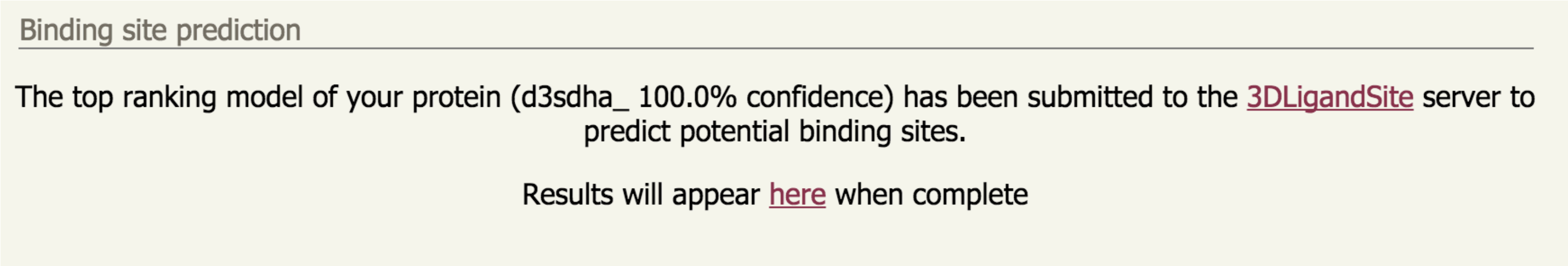
Predict ligand and binding sites from structure

- If a Phyre2 hit is sufficiently confident....
- Top model sent automatically to 3DLigandSite for ligand binding site prediction
- Sequences submitted to 3DLigandSite automatically sent to Phyre2 for modelling

3DLigandSite



Currently a link near bottom of Phyre2 Results



Binding site prediction

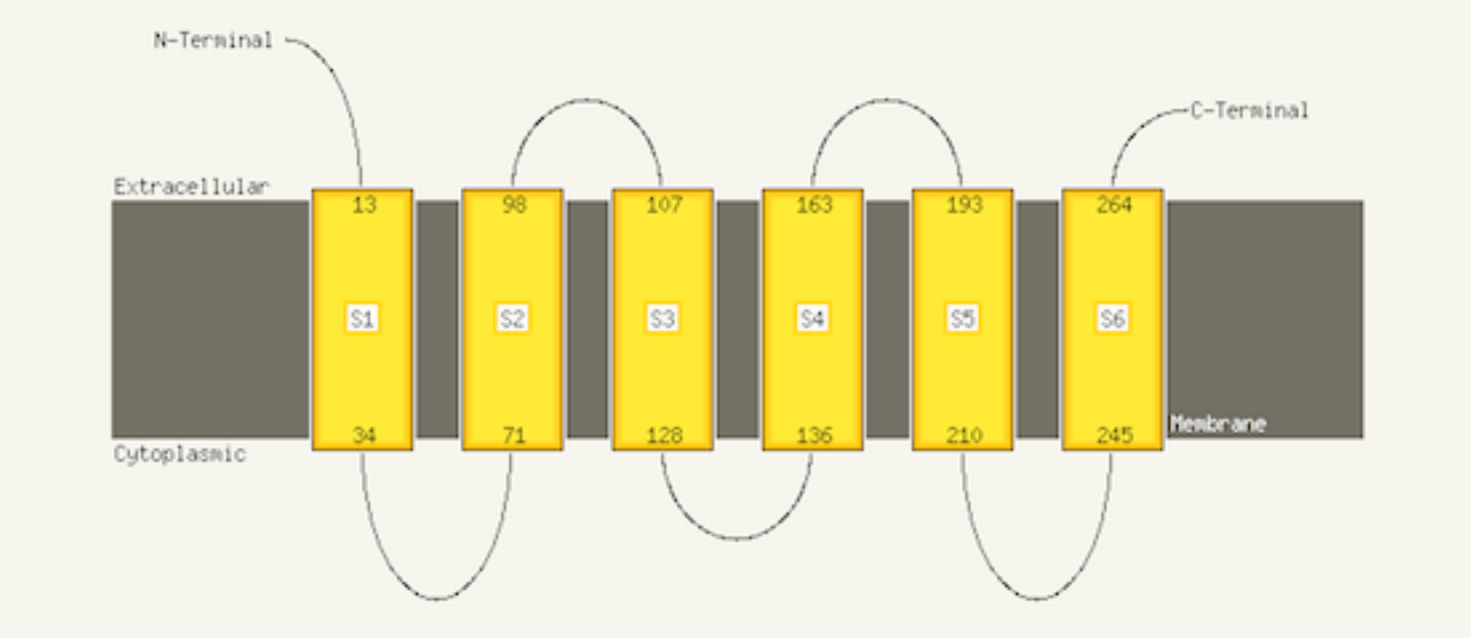
The top ranking model of your protein (d3sdha_ 100.0% confidence) has been submitted to the 3DLigandSite server to predict potential binding sites.

Results will appear here when complete

New version will have an embedded JSMol viewer, summary table of results within Phyre2 web page

Transmembrane helices

If confidently predicted to contain membrane helices, topology prediction run. (Machine learning approach estimated to be 85% accurate)



Memsat-SVM, Nugent & Jones, 2009

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

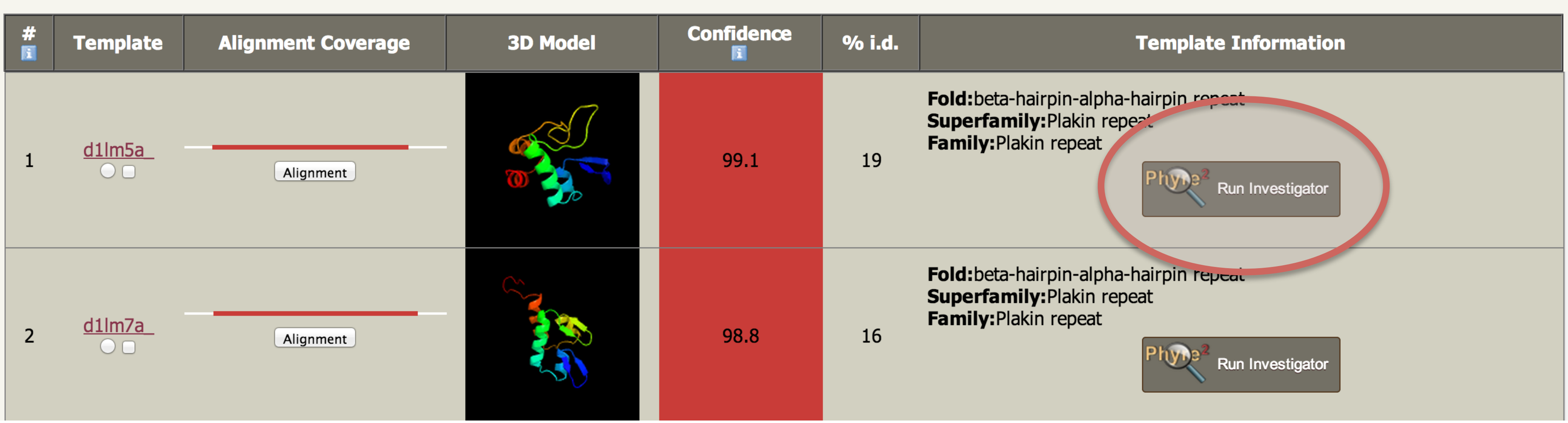
- Clashes
- Rotamer outliers
- Ramachandran outliers
- ProQ2 model quality assessment
- Alignment confidence (HHsearch)
- Conservation/evolutionary trace (Jenson-Shannon divergence –far faster and just as accurate as ET)
- Catalytic Site Atlas
- Disorder
- Pocket detection (Fpocket)
- Protein interface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

Phyre Investigator

Effect of Mutations?

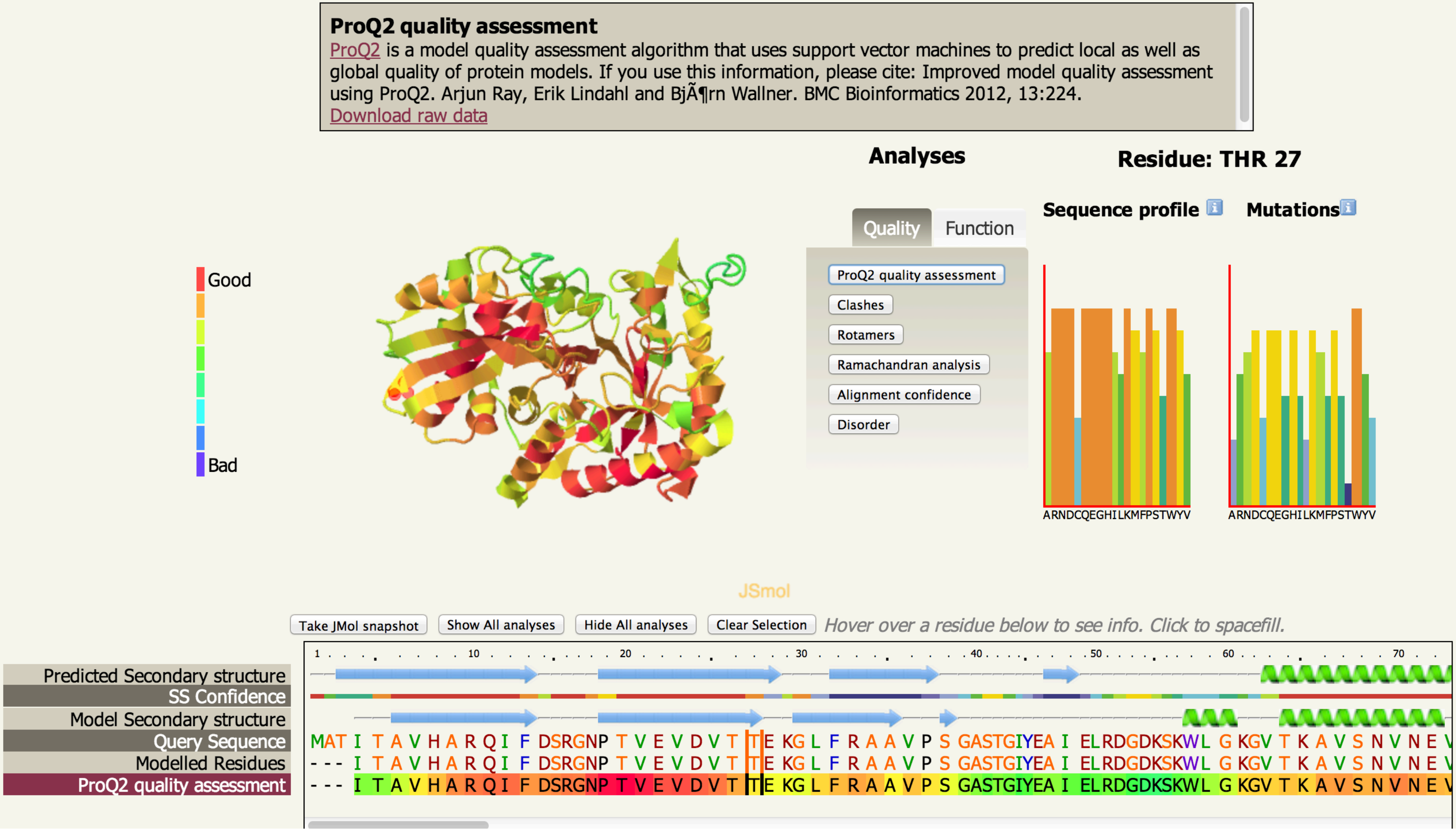
- Will a SNP effect my protein’s function?
- New method: SuSPect by Chris Yates
- Integrated into Phyre Investigator
- Also standalone server

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 Phyre² Run Investigator
2	d1lm7a_ ○ ☐	Alignment		98.8	16	**Fold:**beta-hairpin-alpha-hairpin repeat **Superfamily:**Plakin repeat **Family:**Plakin repeat Phyre² Run Investigator

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.

Predicted Secondary structure
SS Confidence
Model Secondary structure
Query Sequence: MAT I T A V H A R Q I F DSRGNP T V E V D V T T E KG L F R A A V P S GASTGIYEA I ELRDGDKSKWL G KGV T K A V S N V N E V
Modelled Residues: - - - I T A V H A R Q I F DSRGNP T V E V D V T T E KG L F R A A V P S GASTGIYEA I ELRDGDKSKWL G KGV T K A V S N V N E V
ProQ2 quality assessment: - - - I T A V H A R Q I F DSRGNP T V E V D V T T E KG L F R A A V P S GASTGIYEA I ELRDGDKSKWL G KGV T K A V S N V N E V

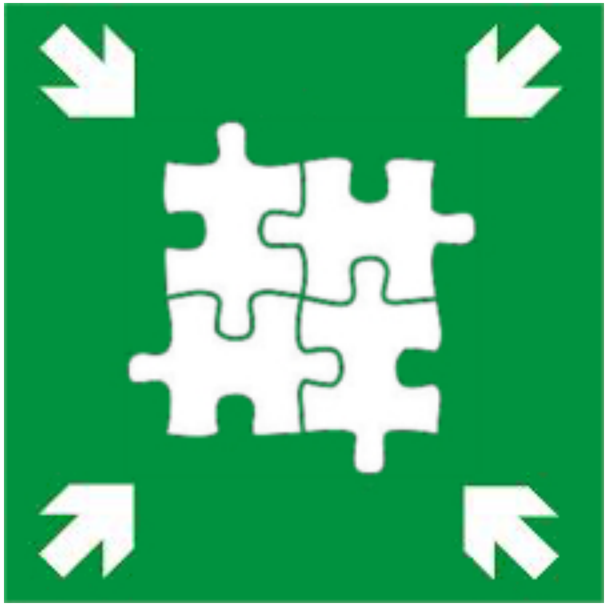
Future

New paper in Nature Protocols:
The Phyre2 web portal for protein modelling, prediction and analysis

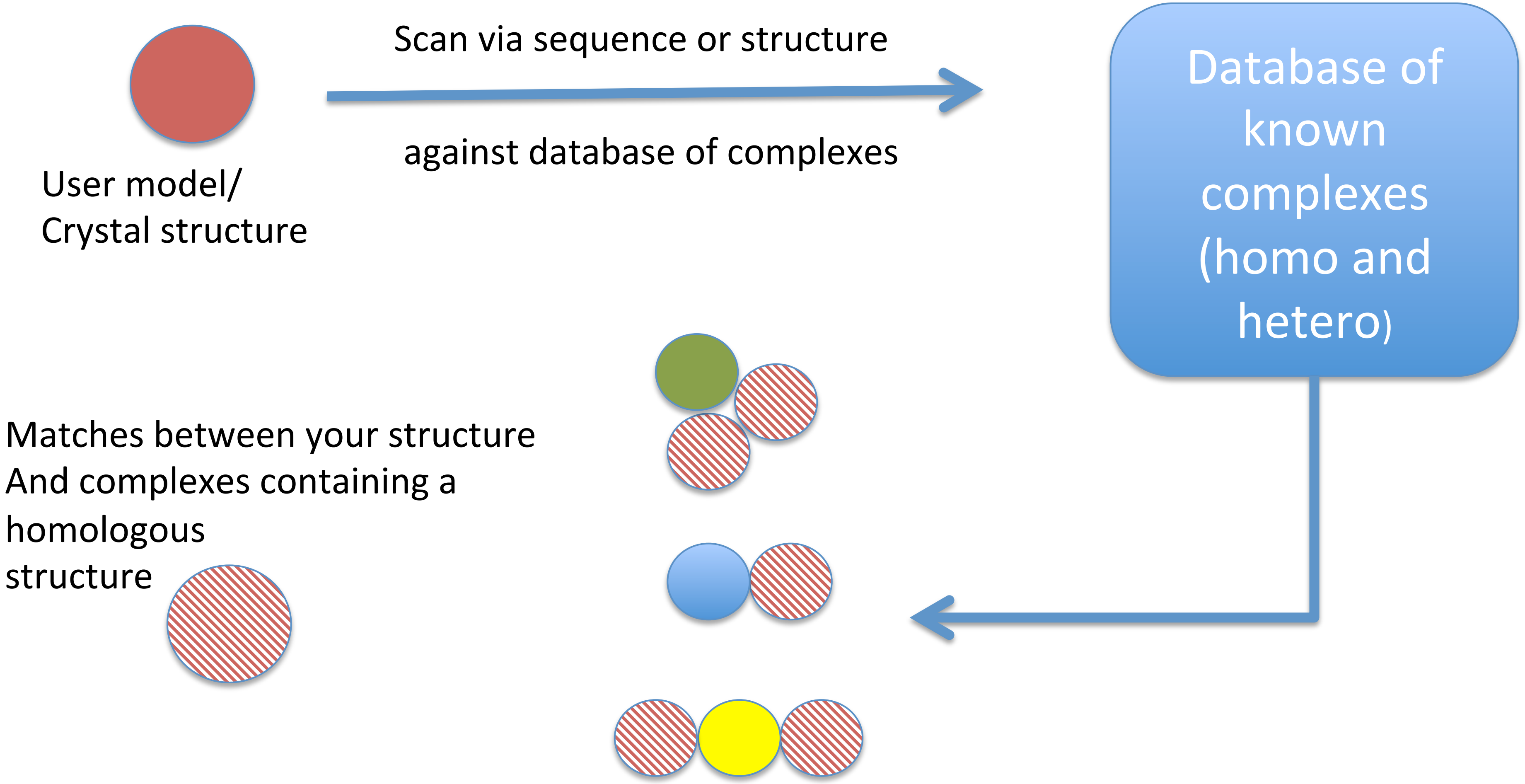
Awaiting proofs, should be out in a few weeks at most

2015: Complex building

Phyre Assembly Point



Complex building



User model/
Crystal structure

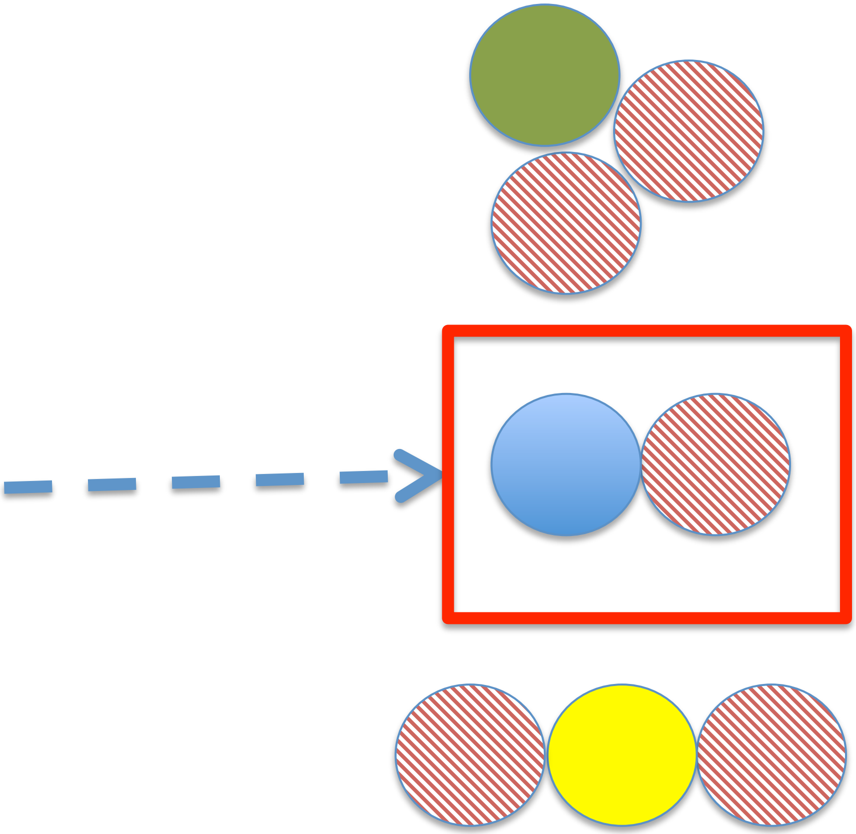
Scan via sequence or structure

against database of complexes

Database of known complexes (homo and hetero)

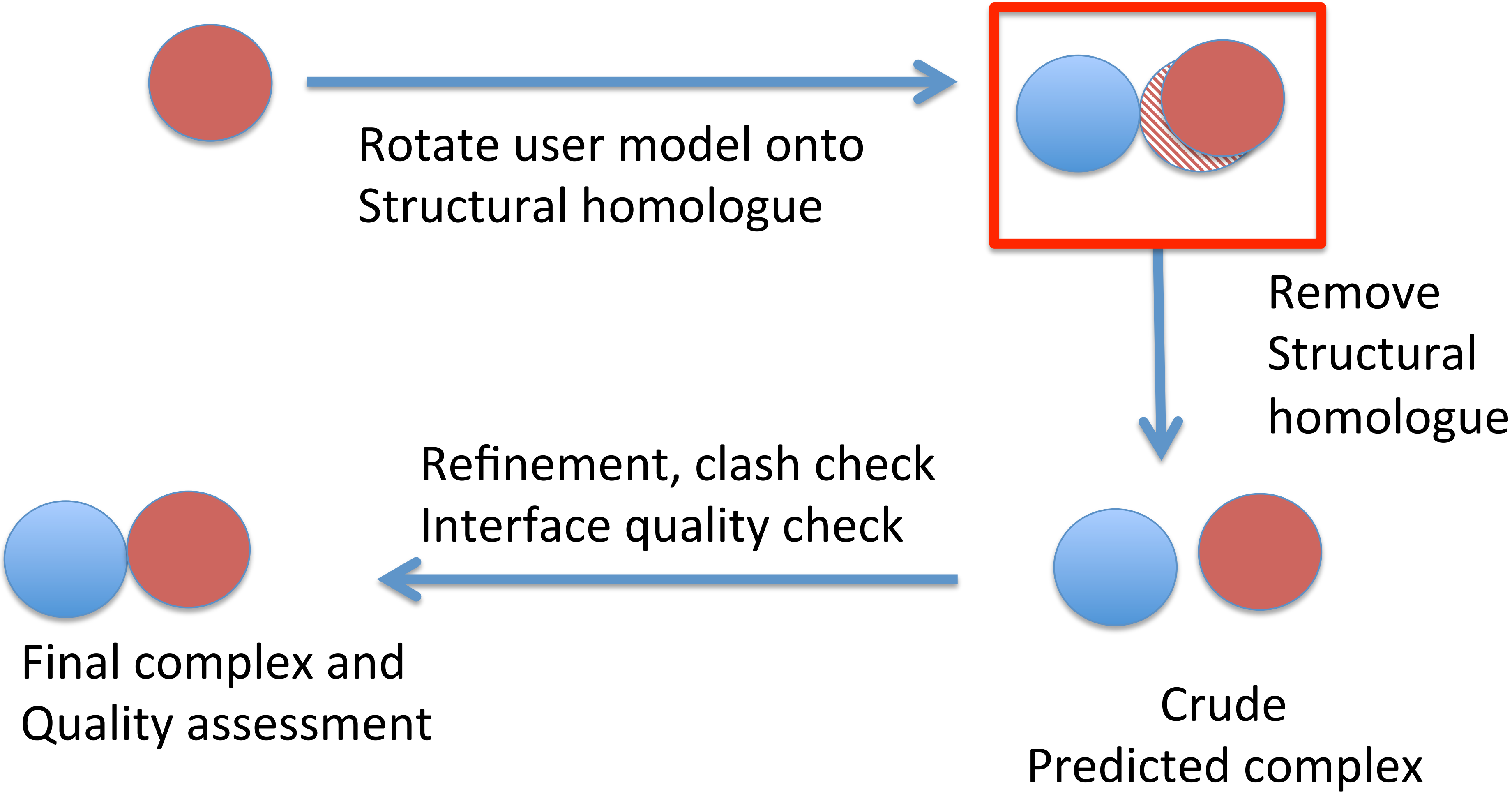
Matches between your structure
And complexes containing a
homologous
structure

Complex building



User chooses which complex
Is desired and which chains
to substitute

Complex building



Rotate user model onto
Structural homologue

Remove
Structural
homologue

Crude
Predicted complex

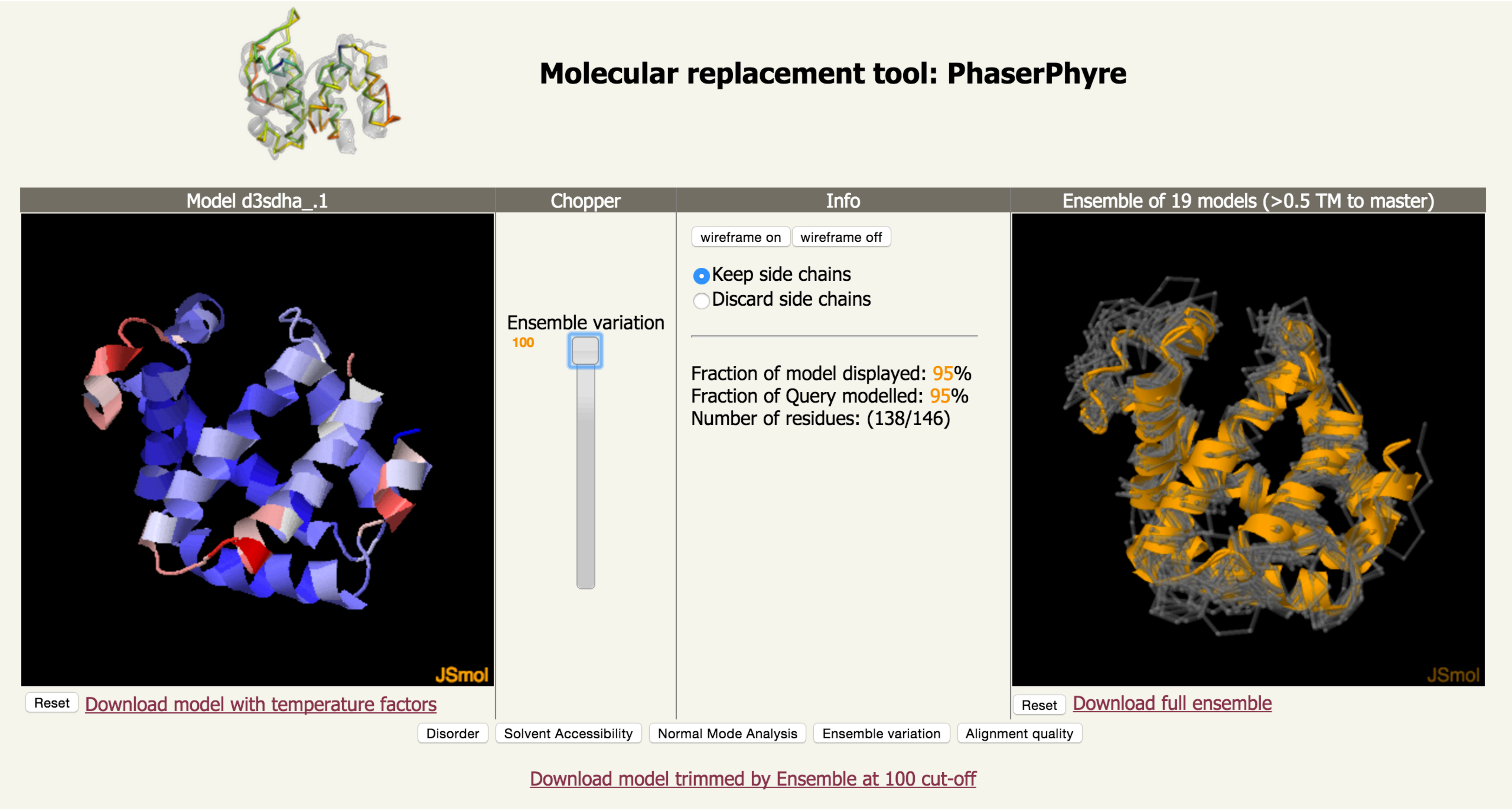
Refinement, clash check
Interface quality check

Final complex and
Quality assessment

Complex building

- Include Rosetta relax protocol for the complex
- Report clashes
- Report quality of interface – investigate existing scoring schemes
- Generalise to permit users to upload multiple structures
- Scan and find existing complexes that contain structural homologues of the input models in different combinations

PhaserPhyre



Molecular replacement tool: PhaserPhyre

Model d3sdha_.1	Chopper	Info	Ensemble of 19 models (>0.5 TM to master)
	Ensemble variation 100	● Keep side chains ○ Discard side chains Fraction of model displayed: 95% Fraction of Query modelled: 95% Number of residues: (138/146)	

Download model with temperature factors

Download full ensemble

Download model trimmed by Ensemble at 100 cut-off

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 release in 1-2 months. Interface under developement

Jalview Integration

- Jalview users can submit their sequence to Phyre2 directly within the Jalview desktop app or web applet
- Should be in place in 1-2 months
- Later, allow Phyre2 users to directly load their model and alignments into Jalview
- Work with Geoff Barton and Jim Procter, Dundee

Feedback welcome

- Ease of use – what's clear/unclear on the site?
- Functionality – is there something you wish it could do that it doesn't?
- Contact:
  - Anonymous feedback on Phyre2 Worshop page
  - Twitter (@phyre2server)
  - Google Groups (groups.google.com/group/phyre)
  - Email (l.a.kelley@imperial.ac.uk)