Homology 3D modeling and effect of mutations

Determination of protein structure

X-ray crystallography (70,714 in PDB)

- need crystals

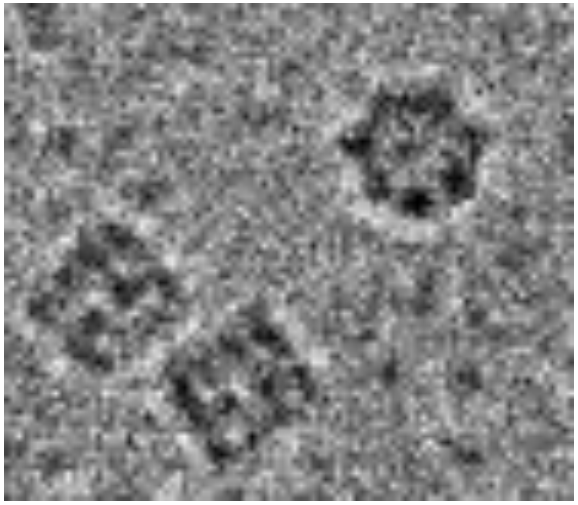
Nuclear Magnetic Resonance (NMR) (9,312)

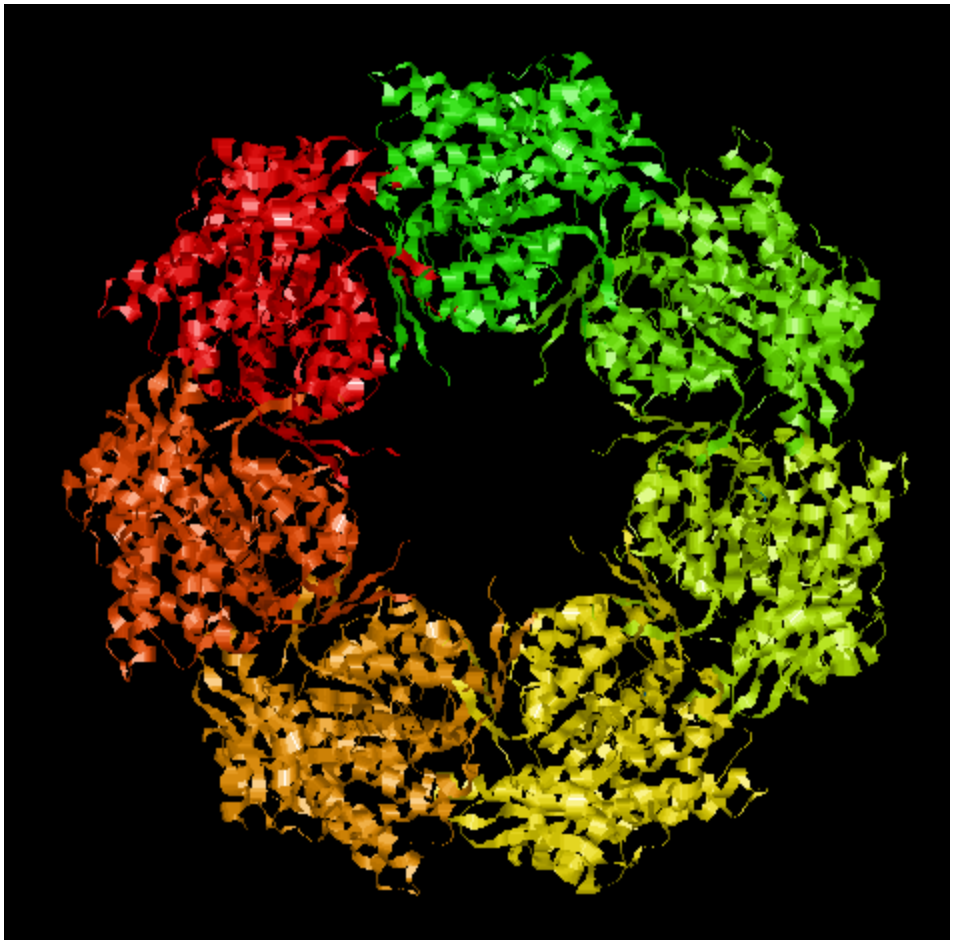
- proteins in solution
- lower size limit (600 aa)

Electron microscopy (422)

- Low resolution (>5A)

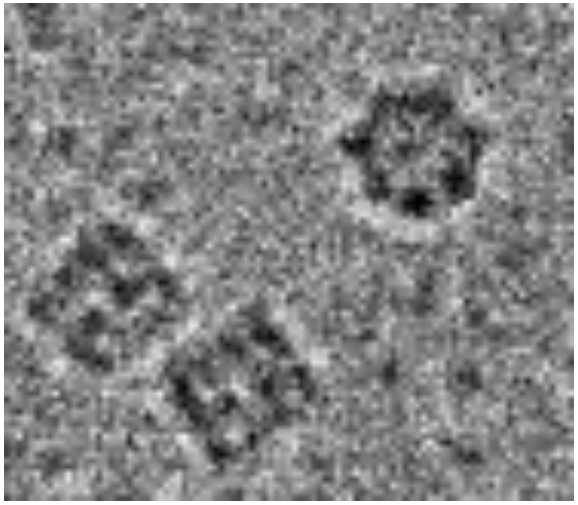
Determination of protein structure

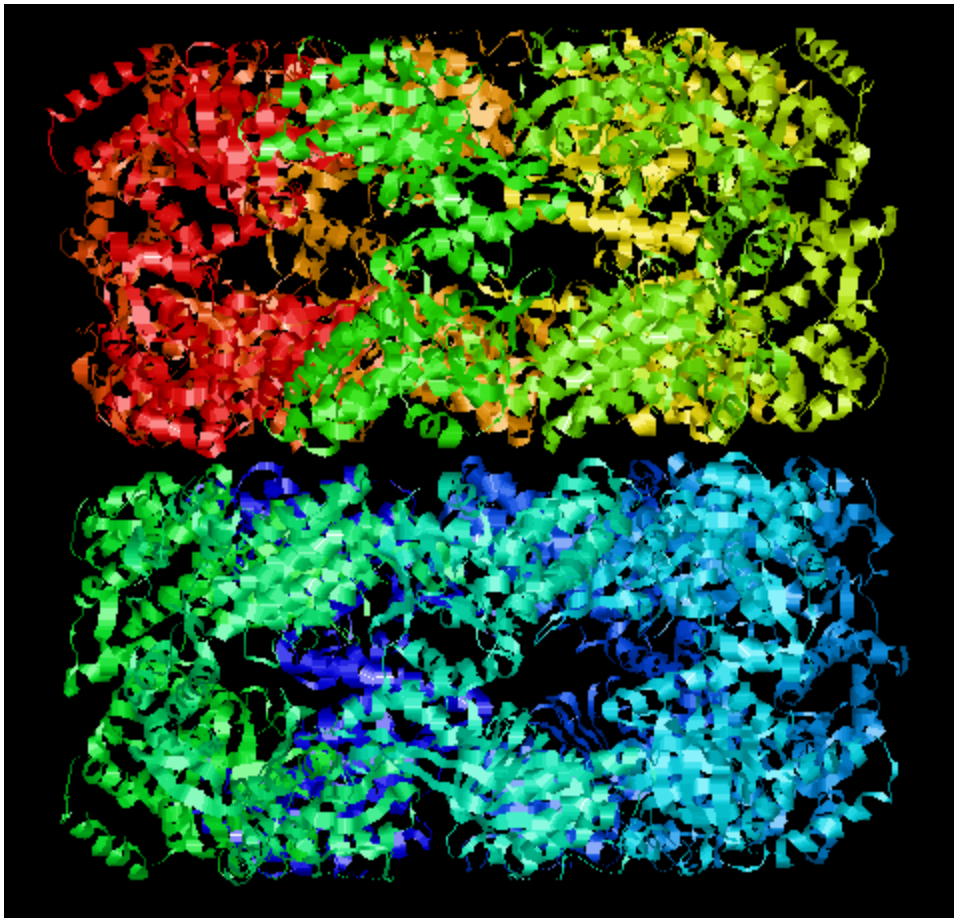




resolution 2.4 A

Determination of protein structure





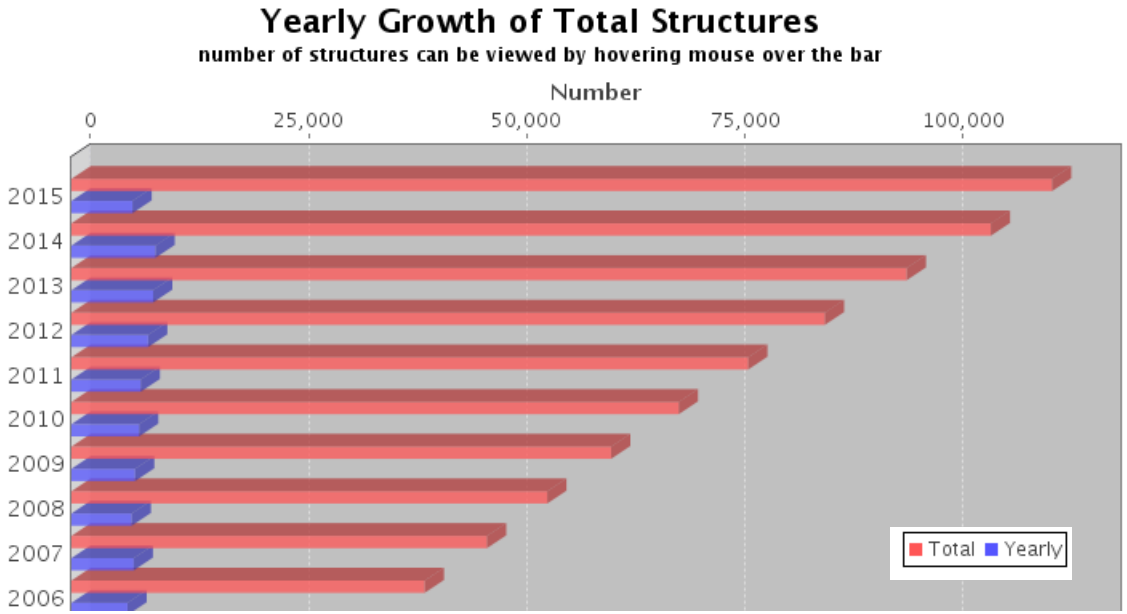
resolution 2.4 A

Structural genomics

PDB at a Glance | 35264 Distinct Protein Sequences | 27724 Structures of Human Sequences | 7550 Nucleic Acid Containing Structures | More Statistics

Currently: 112K 3D structures
from around 36K sequences
46M sequences in UniProt

only 0.08%!

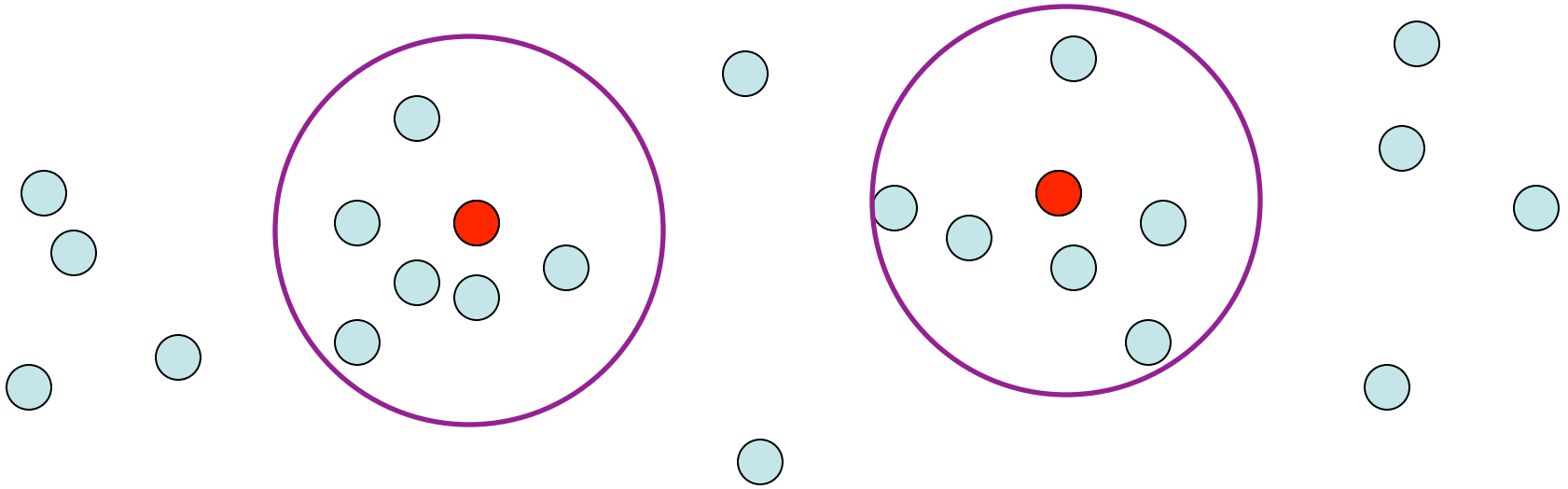


Structural genomics

PDB at a Glance | 35264 Distinct Protein Sequences | 27724 Structures of Human Sequences | 7550 Nucleic Acid Containing Structures | More Statistics

Currently: 112K 3D structures
from around 36K sequences
46M sequences in UniProt

only 0.08%!

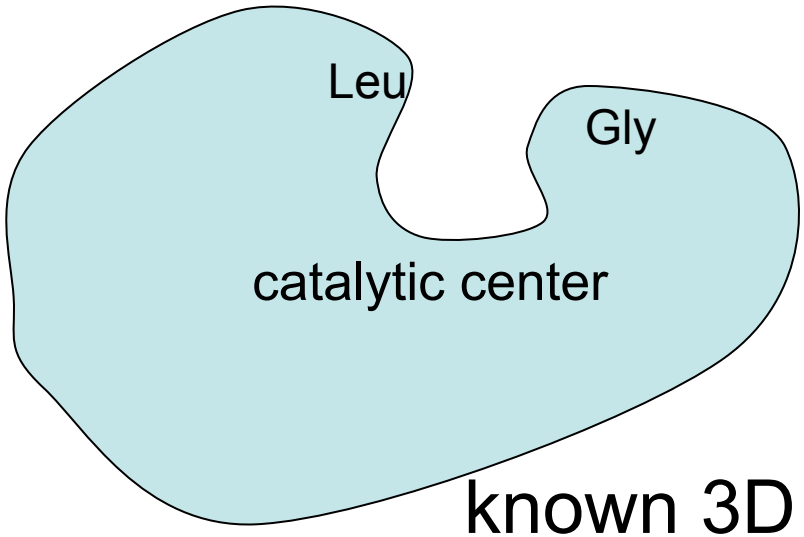


50% sequences covered (25% in 1995)

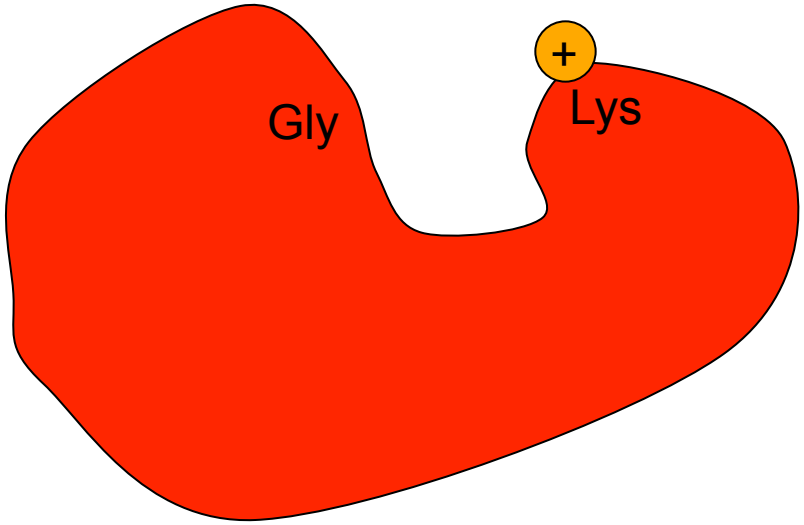
3D structure prediction
Applications: target design

Query sequence

similar to



known 3D

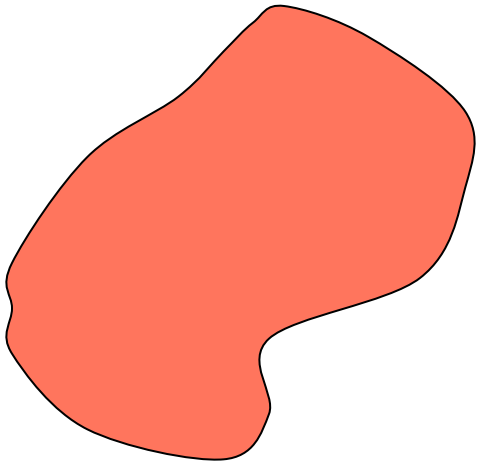


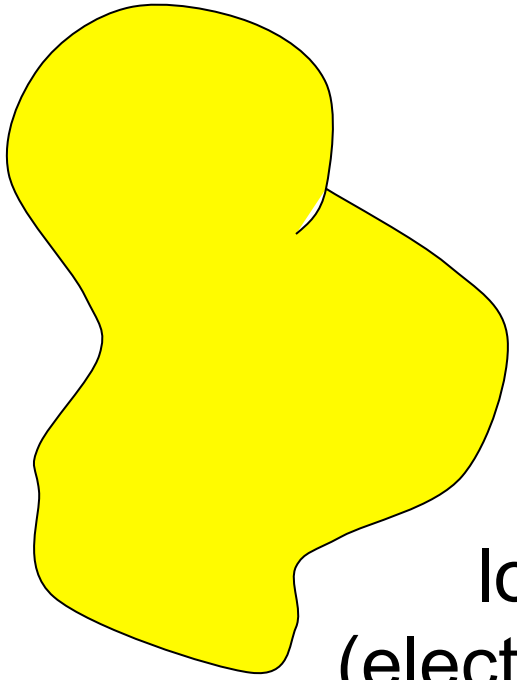
model 3D by homology

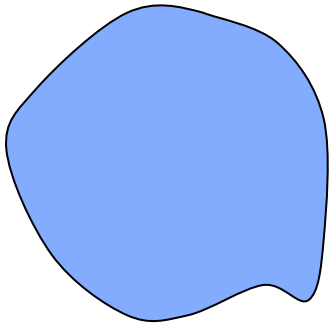
3D structure prediction
Applications: fit to low res 3D

Query sequence 1

Query sequence 2







low resolution 3D
(electron microscopy)

Domains

Protein domains are structural units (average 160 aa) that share:

Function
Folding
Evolution

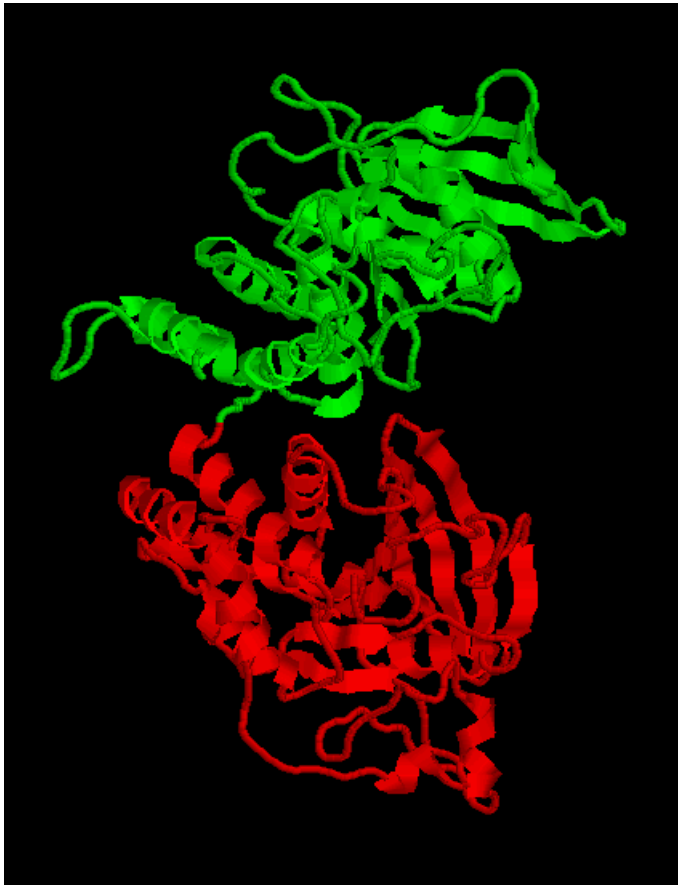
Proteins normally are multidomain (average 300 aa)

Domains

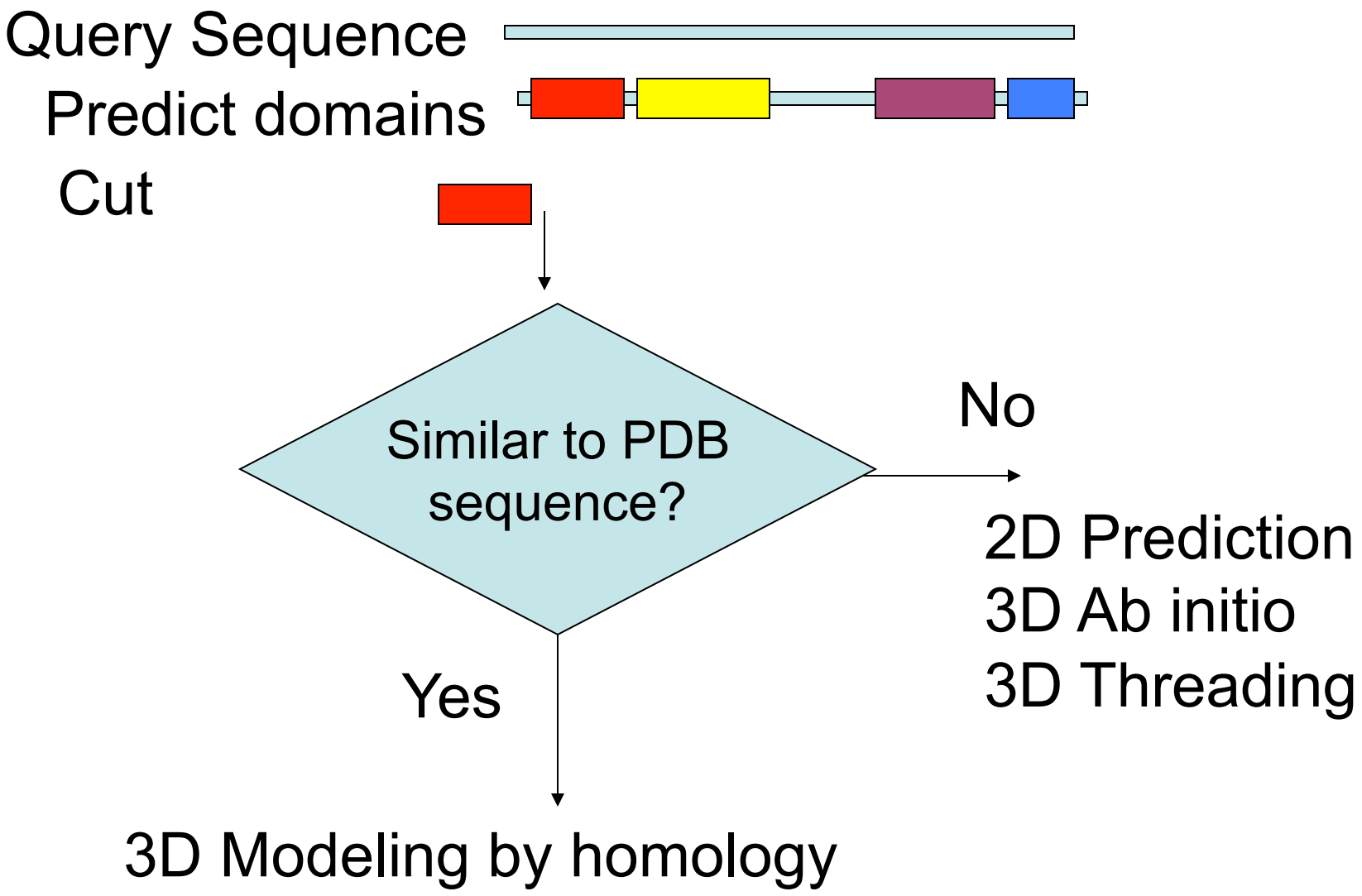
Protein domains are structural units (average 160 aa) that share:

Function
Folding
Evolution

Proteins normally are multidomain (average 300 aa)



Domains


Query Sequence
Predict domains
Cut
Similar to PDB sequence?
No
2D Prediction
3D Ab initio
3D Threading
Yes
3D Modeling by homology

3D structure prediction
Ab initio

Explore conformational space

Limit the number of atoms

Break the problem into fragments of sequence

Optimize hydrophobic residue burial and pairing of beta-strands

Limited success

3D structure prediction
Threading

I-Tasser: Jeffrey Skolnick & Yang Zhang

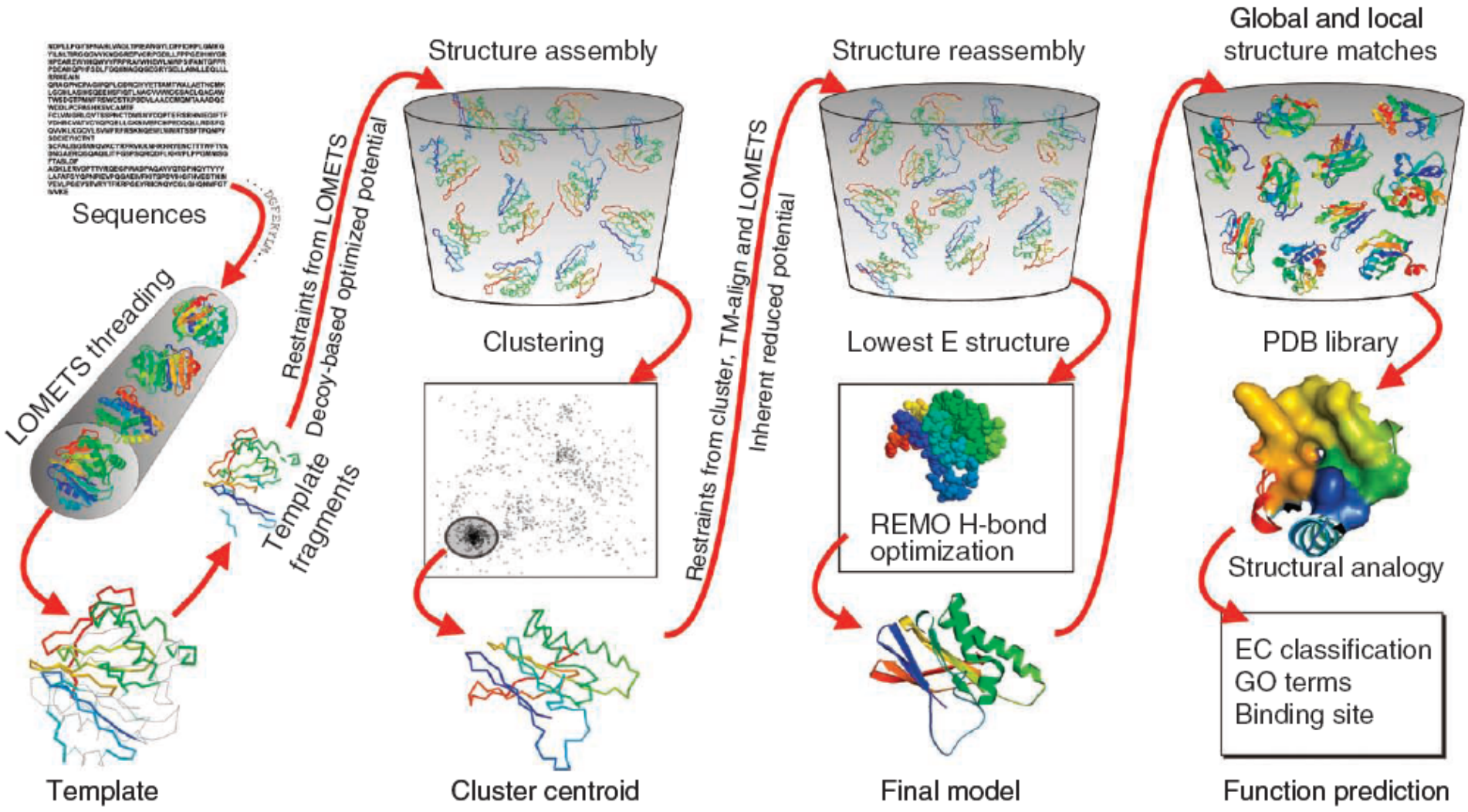
Fold 66% sequences <200 aa long of low homology to PDB

Just submit your sequence and wait… (some days)

Output are predicted structures (PDB format)

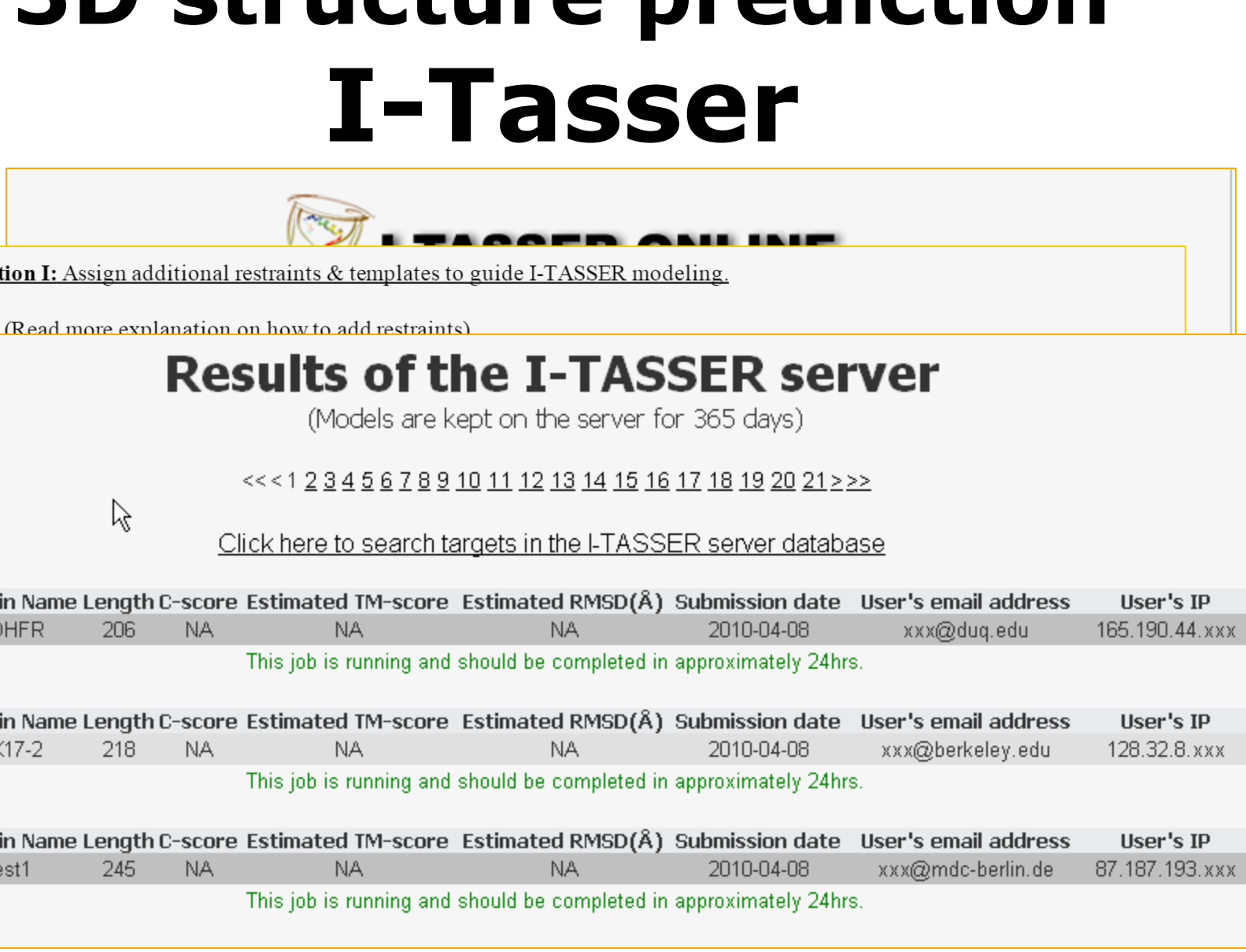
Lee and Skolnick (2008) *Biophysical Journal*
Roy et al (2010) *Nature Methods*
Yang et al (2015) *Nature Methods*

3D structure prediction
I-Tasser



Roy et al (2010) *Nature Methods*

3D structure prediction
I-Tasser



Option I: Assign additional restraints & templates to guide I-TASSER modeling.

(Read more explanation on how to add restraints)

Results of the I-TASSER server

(Models are kept on the server for 365 days)

<< < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 > >>

Click here to search targets in the I-TASSER server database

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S43019	pjDHFR	206	NA	NA	NA	2010-04-08	xxx@duq.edu	165.190.44.xxx
This job is running and should be completed in approximately 24hrs.								

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S43018	FK17-2	218	NA	NA	NA	2010-04-08	xxx@berkeley.edu	128.32.8.xxx
This job is running and should be completed in approximately 24hrs.								

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S43017	test1	245	NA	NA	NA	2010-04-08	xxx@mdc-berlin.de	87.187.193.xxx
This job is running and should be completed in approximately 24hrs.								

http://zhanglab.ccmb.med.umich.edu/I-TASSER/

3D structure prediction
I-Tasser

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S42744	1ijwC	52	0.67	0.80±0.09	1.5±1.4	2010-04-02	xxx@ntu.edu.tw	140.112.94.xxx

Submitted Sequence

>your_protein

Top 5 Models predicted by I-TASSER

Top 10 templates used by I-TASSER

Rank	PDB Hit	Iden1	Iden2	Cov.	Norm. Z-score	Download Align.
1	1b3uA	0.11	0.19	0.94	1.11	Download
2	1b3uA	0.07	0.19	0.98	2.58	Download
3	1b3uA	0.08	0.19	0.97	1.27	Download
4	3dh4A	0.10	0.18	0.87	1.08	Download
5	1qgrA	0.08	0.17	1.00	2.48	Download
6	1pw4A	0.09	0.19	0.80	1.18	Download
7	1jdhA	0.09	0.20	0.92	2.42	Download
8	1ialA	0.09	0.17	0.79	1.15	Download
9	2bkuB	0.09	0.17	1.00	2.35	Download
10	1wa5C	0.08	0.18	1.00	2.27	Download

```
                                                  20                  40                  60
                                                   |                   |                   |
Sec.Str CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCHHHHHHHHHHHHHHHHHHH
Seq     MCNPSTTTTTTTGSENQESRTGLFLDLFSINSFEPTKRNLRHCENRGSPLMAEAVTEAKSLFTLAF
1       ALGVERTRSETIYDEDEVLLGTFTTLVGGPEYVHCL--ESLATVEETVVRDKAVESLRAISHE--
2       AAADGDDSLYLRNEDVQLRLNSIKKLGVERTRSELLPFLTDTIYDEDEVLLALAEQLPEYVHCLL
3       LTDTIYDEDEVLLALAEQLGTFTTLVGGPEYVHCLLPPLESLATVEETVVRDKAVESLRAISHEH
4       ----------------------------------------GGGGGGGGGGGGGGGGWWAVGA
5       MANPGNSTSKDLGTETYRPSCAEIPVNQWPELIPQLVANVTNPNSTEHMKESTLEAIGYICQDID
6       ----------------------FKPAPHKARLPA----------AEIDPTYRRLRWQIFLGIFF
7       ------------------------AVVNLIRAIPELTKLLNDEDQVVVNKAAVMVHQLSKKEA
8       ---------------DEQMLKRRNVSNQGTVNWSVEDIVKGINSNNLESQLQATQAARKLLSRQP
9       MIDENTKLNELVSKDSVKTQQFTGAEQPCESADALVSSSNNGAQSTETSKAVRLAALNALADSKN
10      MTQDGASTNLPWVDENGNHLLPLASRLSNDDMVRPLFRSDELFLEIKLVLDVFTAPFLNLLKTVD
```

Model1 Model2 Model3 Model4 Model5

3D structure prediction
QUARK

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

Go to Job Q12270 to view an example of QUARK output. The description of predicted feature files can be seen in readme.txt.

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

Or upload the sequence from your local computer:

Choose File No file chosen

Email: (mandatory, where results will be sent to)

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

http://zhanglab.ccmb.med.umich.edu/QUARK/

3D structure prediction

GenTHREADER

David Jones http://bioinf.cs.ucl.ac.uk/psipred/

Input sequence or MSA

Typically 30 minutes, up to two hours

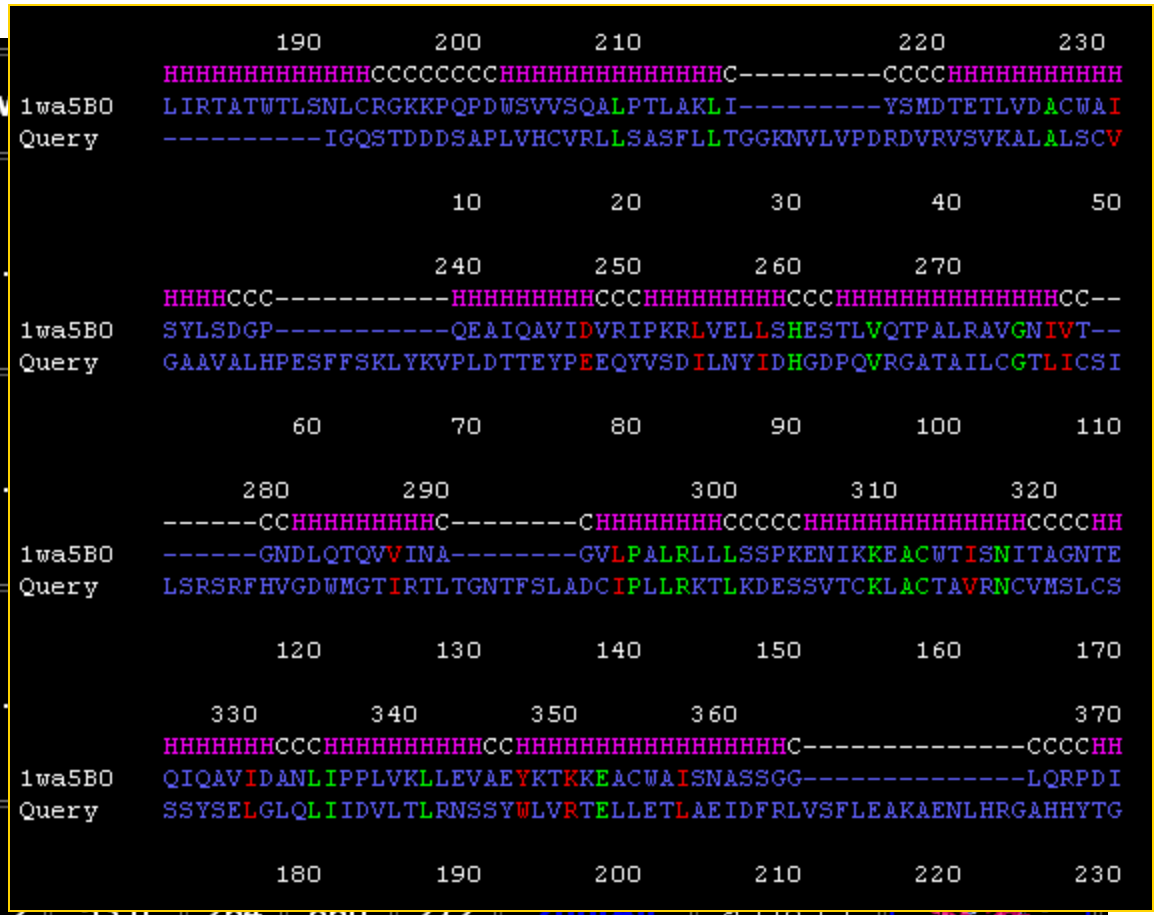
GenTHREADER Jones (1999) *J Mol Biol*

3D structure prediction

GenTHREADER

Output GenTHREADER

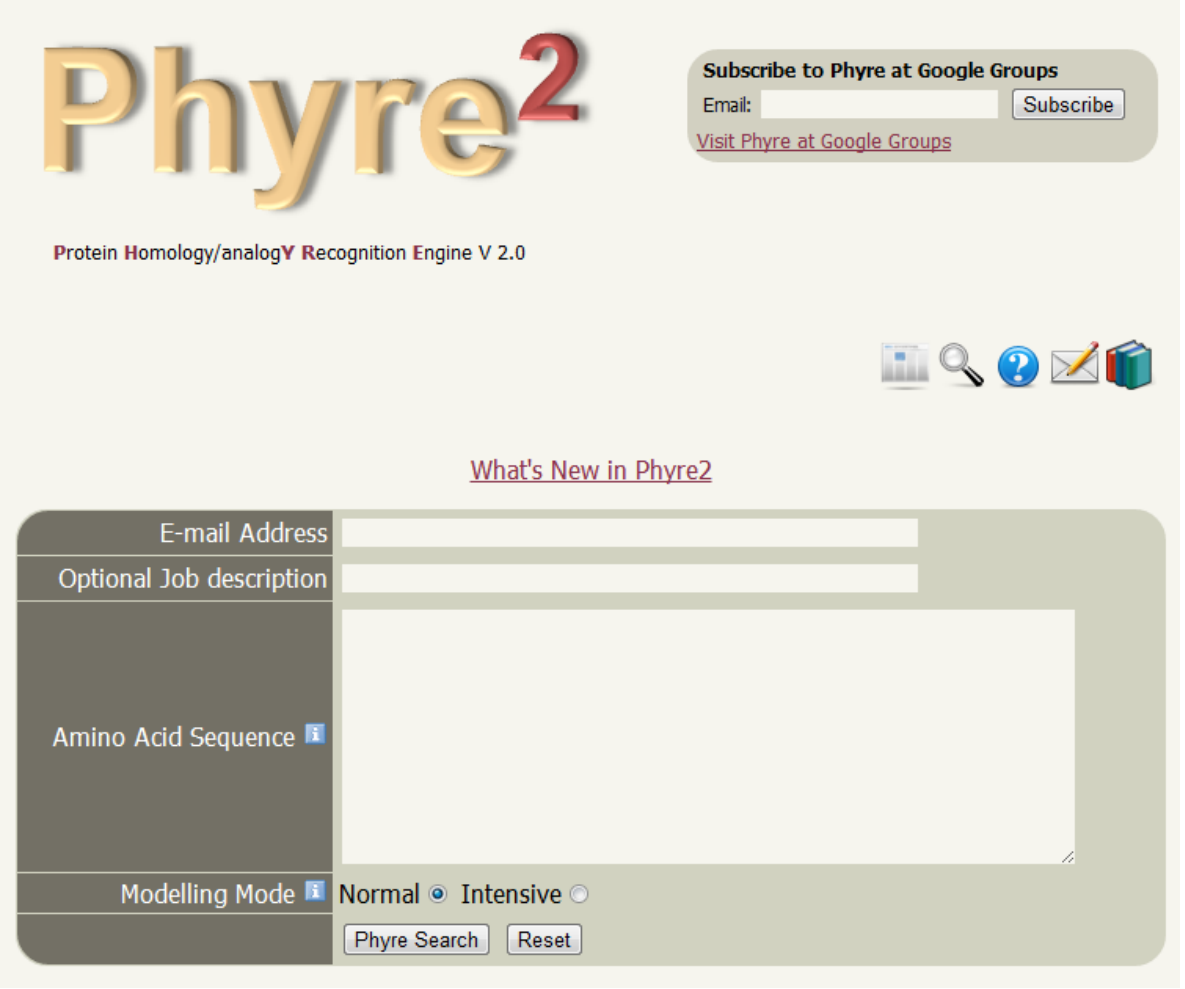
Conf.	Net Score	p-value	PairE	Solv
MEDIUM	45.416	0.001	-410.9	-13.
MEDIUM	45.161	0.001	-466.0	-12.
MEDIUM	44.997	0.001	-390.3	-17.
MEDIUM	43.639	0.002	-379.5	-12.2



3D structure prediction
Phyre

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

Kelley et al (2000) *J Mol Biol*
Kelley and Sternberg (2009) *Nature Protocols*



Processing time can be hours

3D structure prediction
Static solutions

Datasets of precomputed models / computations

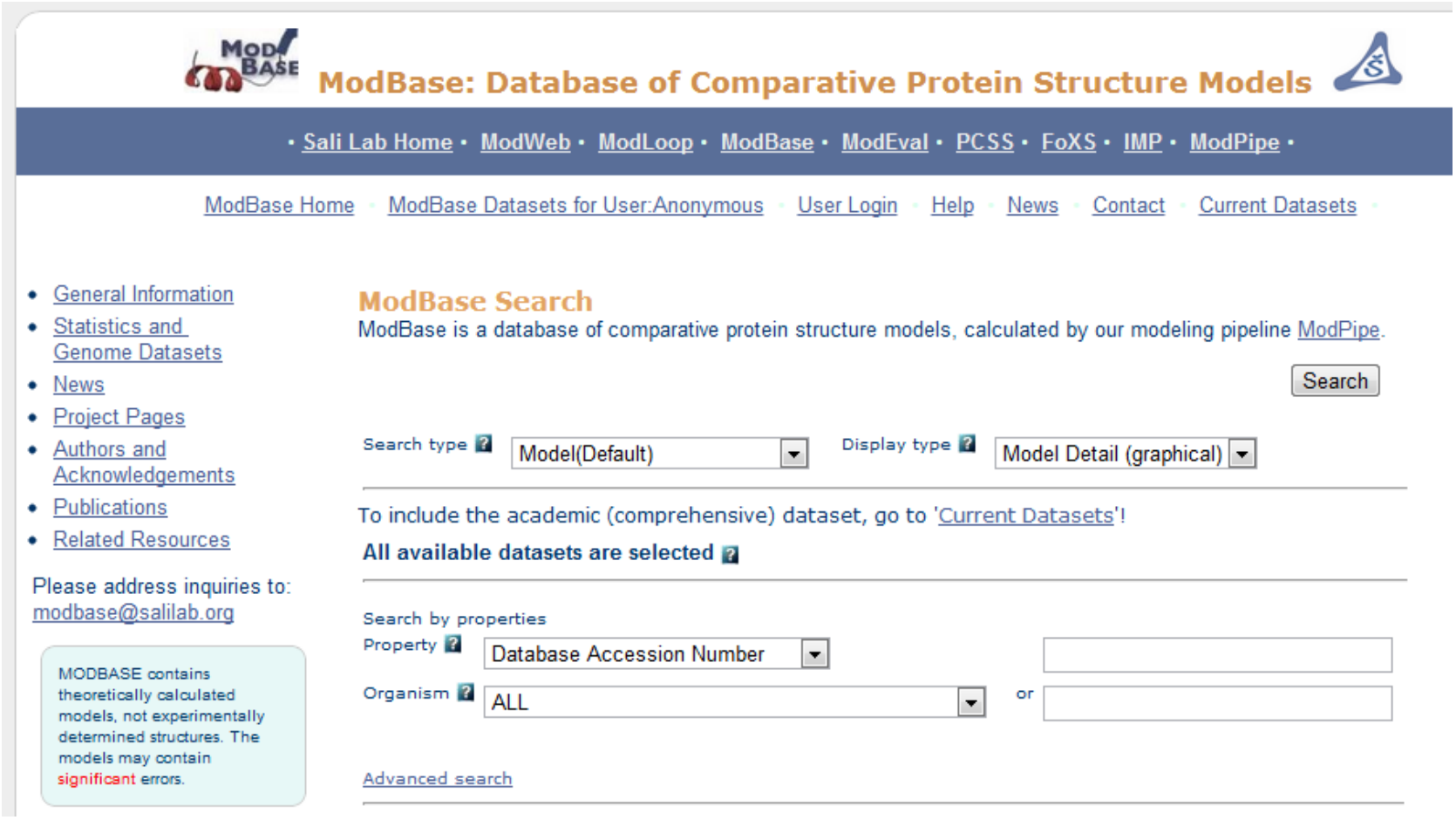
Not flexible

Variable coverage

But you don’t have to wait

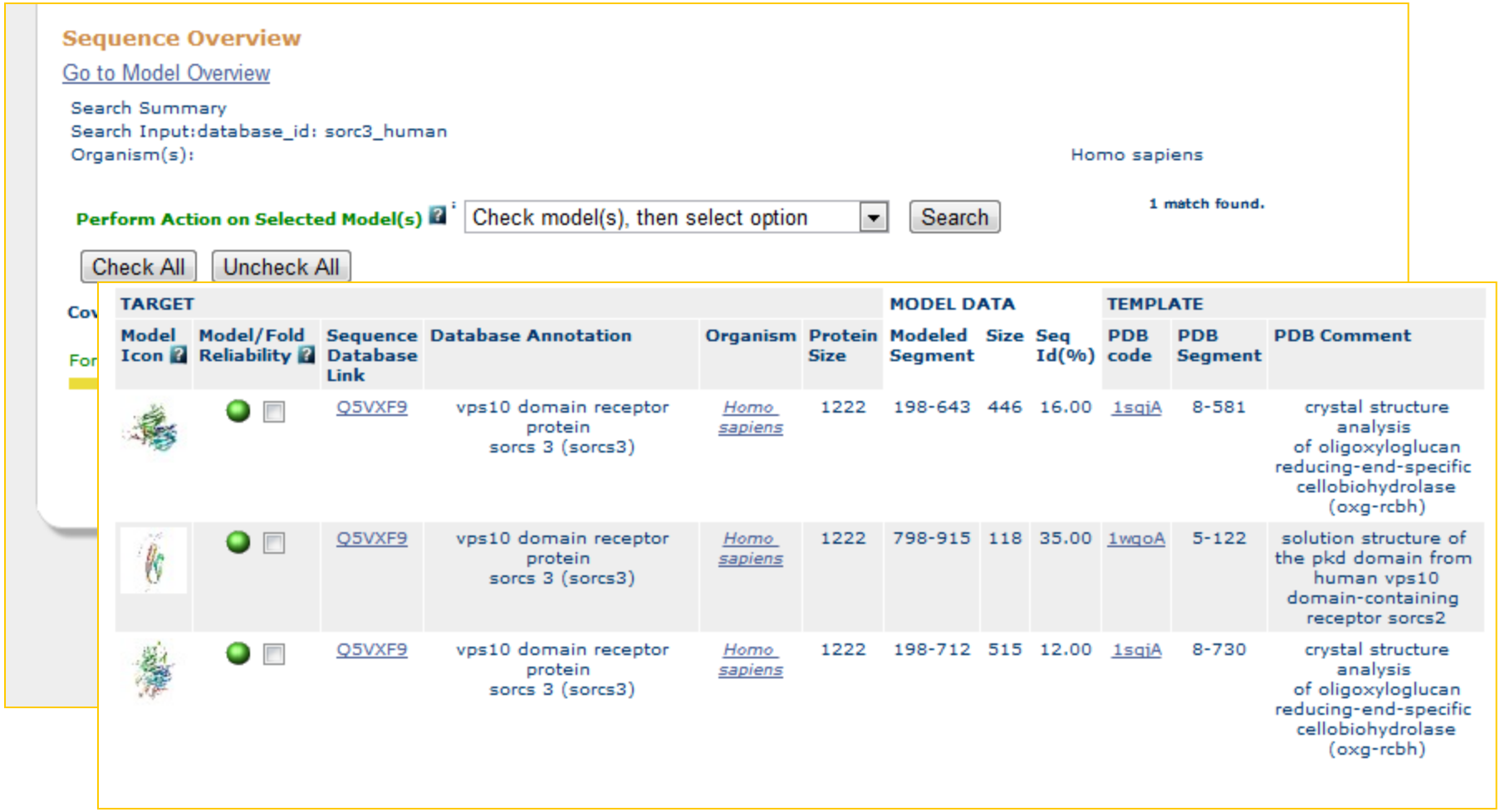
3D structure prediction
MODbase

Andrej Sali http://modbase.compbio.ucsf.edu/



Pieper et al (2011) *Nucleic Acids Research*

3D structure prediction
MODbase



Sequence Overview

Go to Model Overview

Search Summary
Search Input:database_id: sorc3_human
Organism(s): Homo sapiens

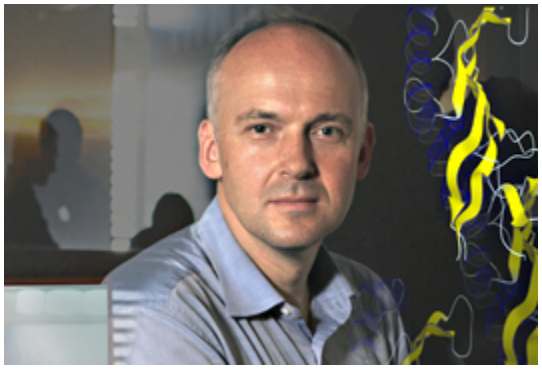
Perform Action on Selected Model(s) : Check model(s), then select option Search

1 match found.

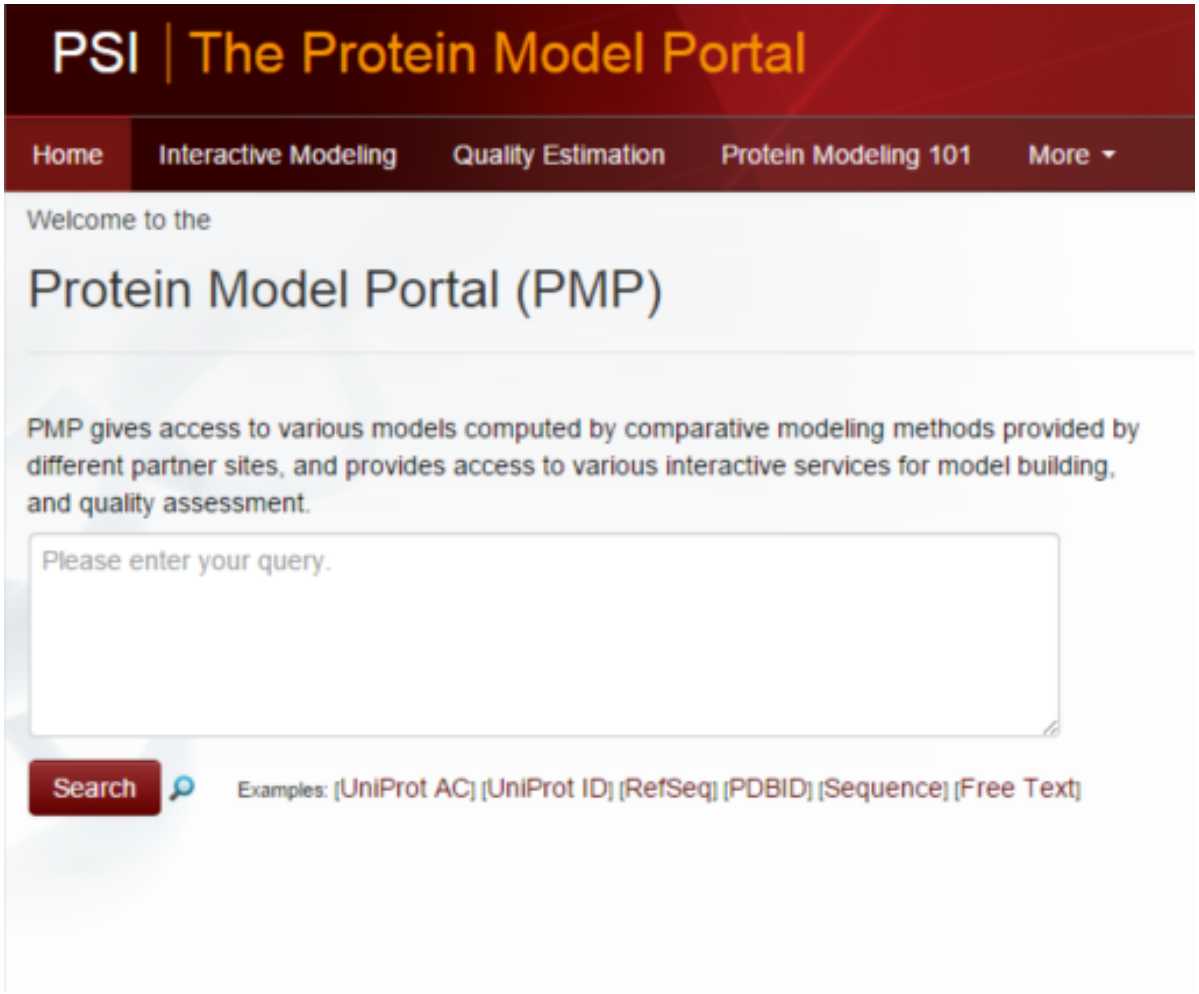
Check All Uncheck All

TARGET						MODEL DATA			TEMPLATE		
Model Icon	Model/Fold Reliability	Sequence Database Link	Database Annotation	Organism	Protein Size	Modeled Segment	Size	Seq Id(%)	PDB code	PDB Segment	PDB Comment
		Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	*Homo sapiens*	1222	198-643	446	16.00	1sqjA	8-581	crystal structure analysis of oligoxyloglucan reducing-end-specific cellobiohydrolase (oxg-rcbh)
		Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	*Homo sapiens*	1222	798-915	118	35.00	1wqoA	5-122	solution structure of the pkd domain from human vps10 domain-containing receptor sorcs2
		Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	*Homo sapiens*	1222	198-712	515	12.00	1sqjA	8-730	crystal structure analysis of oligoxyloglucan reducing-end-specific cellobiohydrolase (oxg-rcbh)

3D structure prediction
Protein Model Portal



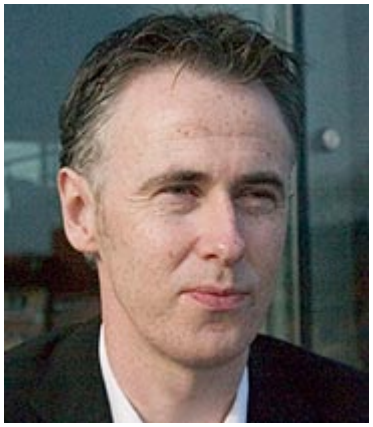
Torsten Schwede

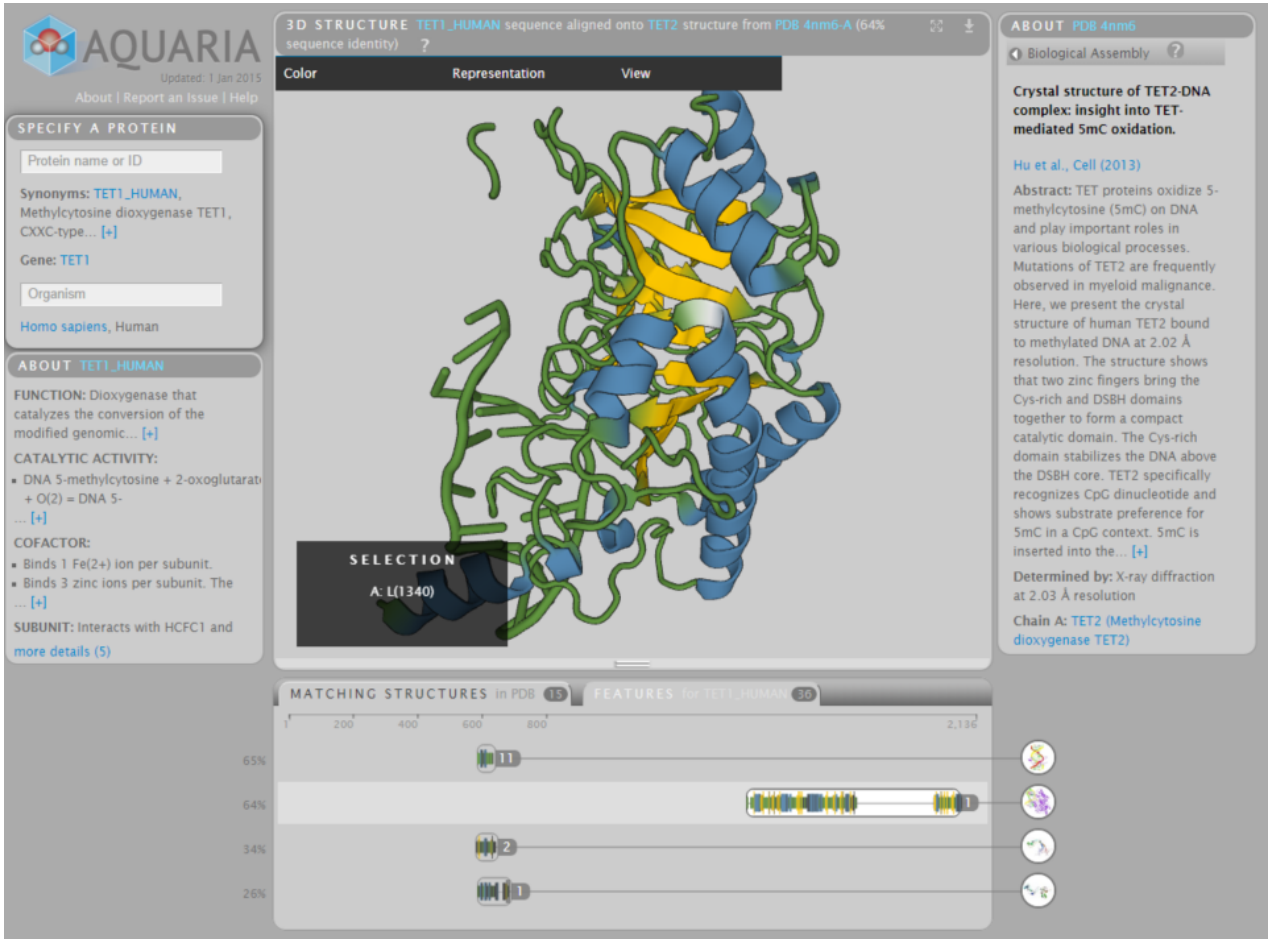


Haas et al. (2013) *Database*

Aquaria

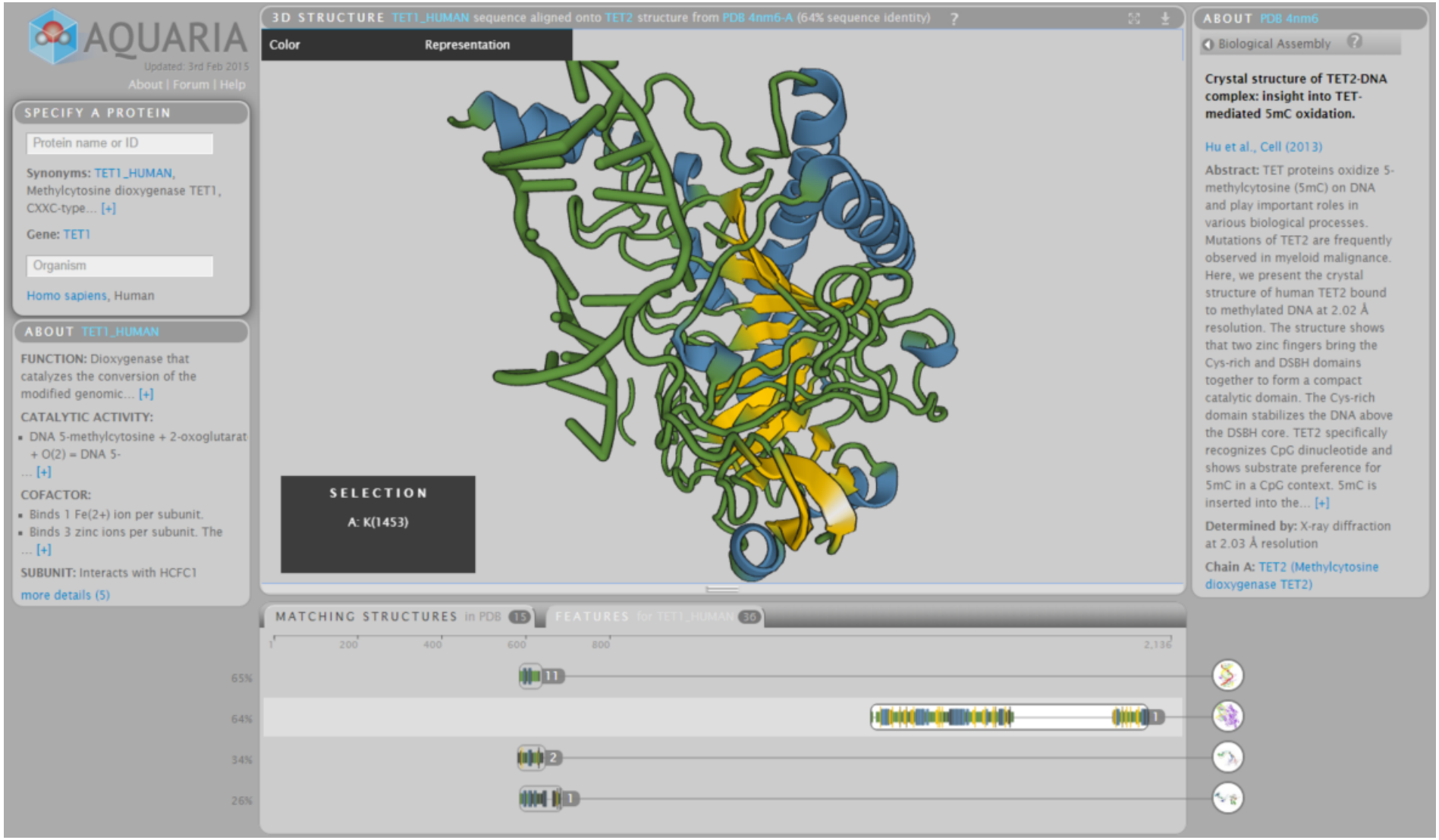
Sean O'Donoghue http://aquaria.ws/



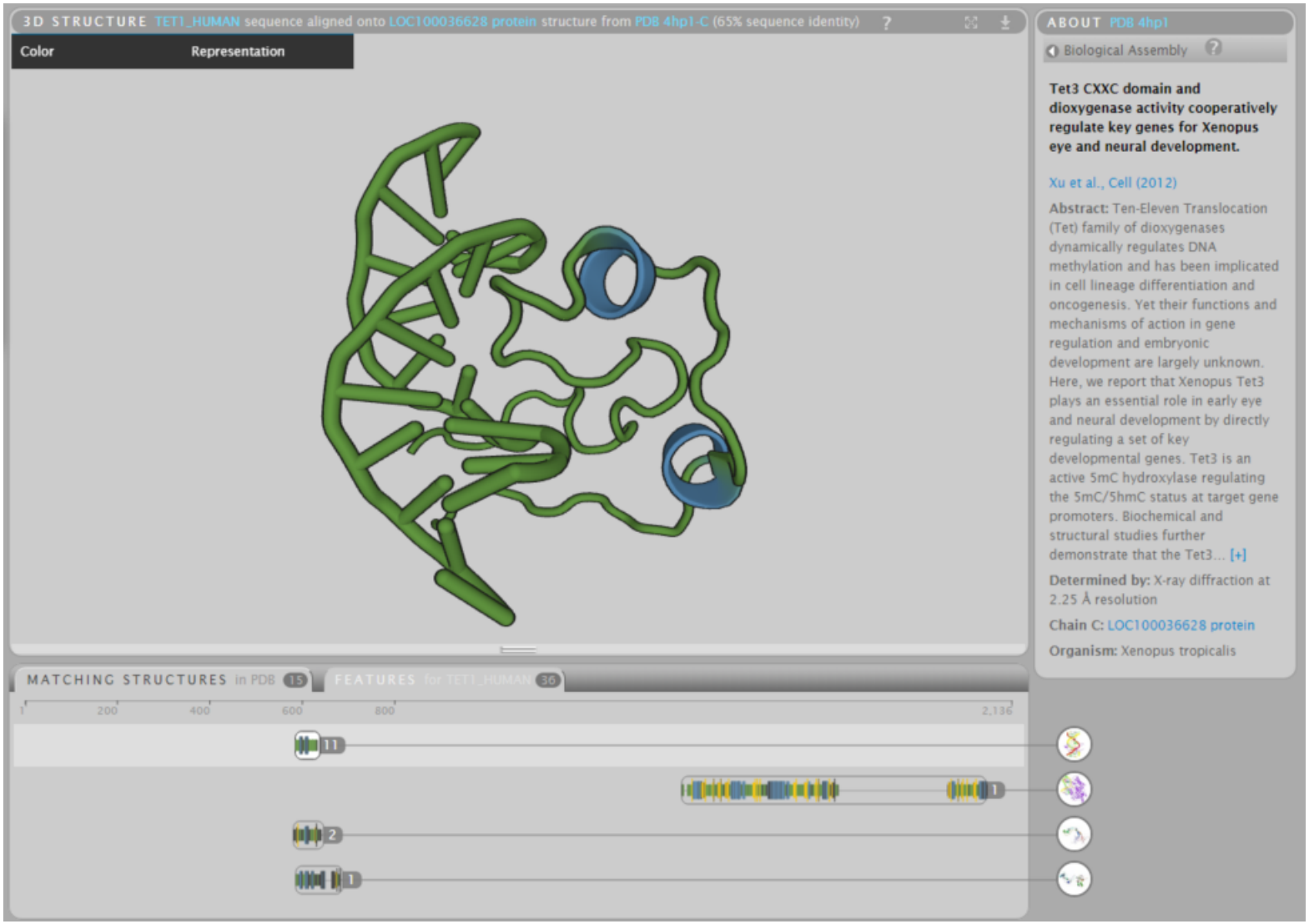


O'Donoghue et al (2015) *Nature Methods*

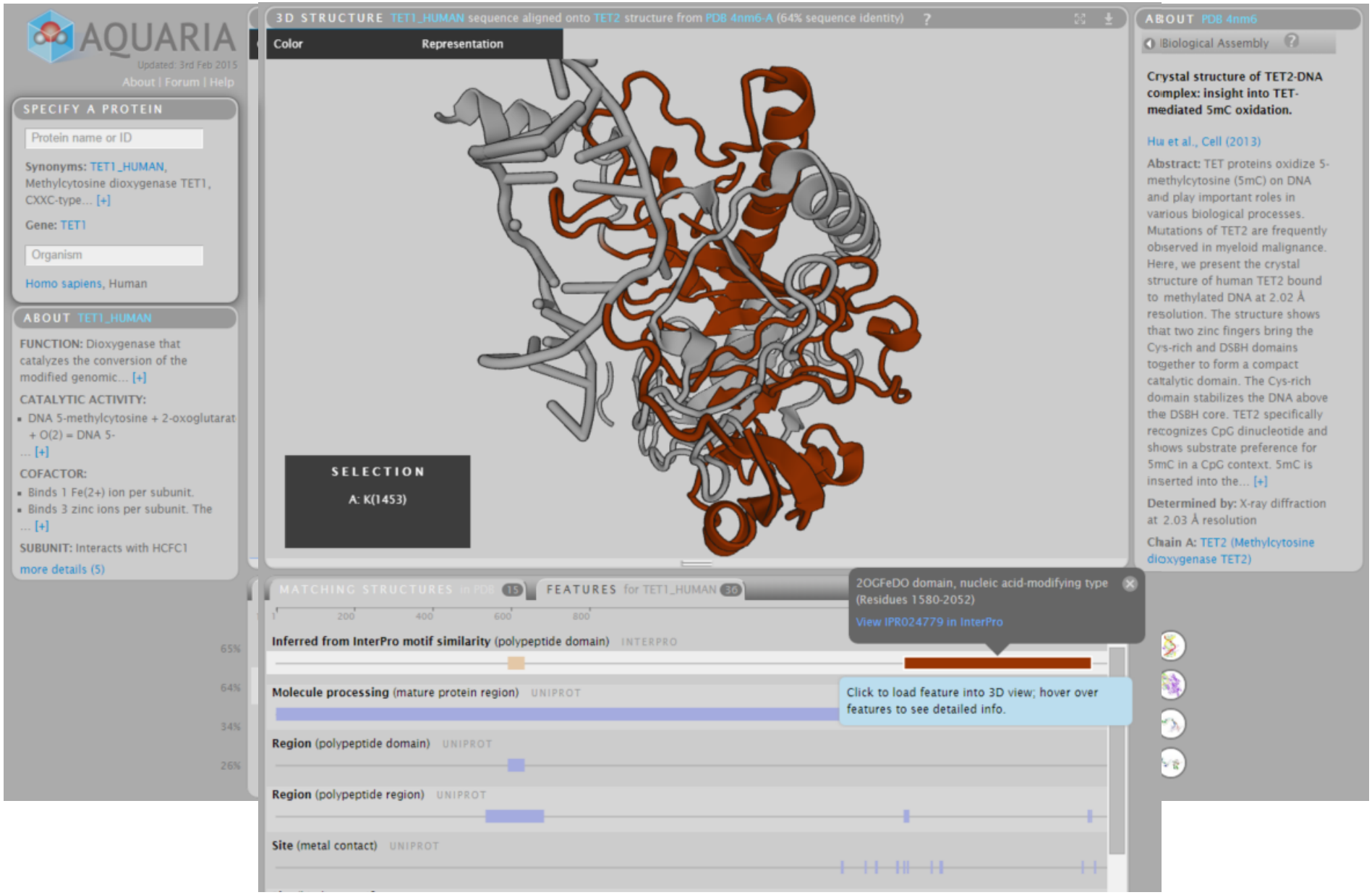
Aquaria


AQUARIA
Updated: 3rd Feb 2015
About | Forum | Help
SPECIFY A PROTEIN
Protein name or ID
Synonyms: TET1_HUMAN, Methylcytosine dioxygenase TET1, CXXC-type... [+]
Gene: TET1
Organism
Homo sapiens, Human
ABOUT TET1_HUMAN
FUNCTION: Dioxygenase that catalyzes the conversion of the modified genomic... [+]
CATALYTIC ACTIVITY:
DNA 5-methylcytosine + 2-oxoglutarat + O(2) = DNA 5- ... [+]
COFACTOR:
Binds 1 Fe(2+) ion per subunit.
Binds 3 zinc ions per subunit. The ... [+]
SUBUNIT: Interacts with HCFC1
more details (5)
3D STRUCTURE TET1_HUMAN sequence aligned onto TET2 structure from PDB 4nm6-A (64% sequence identity)
Color
Representation
SELECTION
A: K(1453)
ABOUT PDB 4nm6
Biological Assembly
Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation.
Hu et al., Cell (2013)
Abstract: TET proteins oxidize 5-methylcytosine (5mC) on DNA and play important roles in various biological processes. Mutations of TET2 are frequently observed in myeloid malignance. Here, we present the crystal structure of human TET2 bound to methylated DNA at 2.02 Å resolution. The structure shows that two zinc fingers bring the Cys-rich and DSBH domains together to form a compact catalytic domain. The Cys-rich domain stabilizes the DNA above the DSBH core. TET2 specifically recognizes CpG dinucleotide and shows substrate preference for 5mC in a CpG context. 5mC is inserted into the... [+]
Determined by: X-ray diffraction at 2.03 Å resolution
Chain A: TET2 (Methylcytosine dioxygenase TET2)
MATCHING STRUCTURES in PDB 15
FEATURES for TET1_HUMAN 36
1
200
400
600
800
2,136
65%
64%
34%
26%
11
1
2
1

Aquaria


3D STRUCTURE TET1_HUMAN sequence aligned onto LOC100036628 protein structure from PDB 4hp1-C (65% sequence identity)
Color
Representation
ABOUT PDB 4hp1
Biological Assembly
Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development.
Xu et al., Cell (2012)
Abstract: Ten-Eleven Translocation (Tet) family of dioxygenases dynamically regulates DNA methylation and has been implicated in cell lineage differentiation and oncogenesis. Yet their functions and mechanisms of action in gene regulation and embryonic development are largely unknown. Here, we report that Xenopus Tet3 plays an essential role in early eye and neural development by directly regulating a set of key developmental genes. Tet3 is an active 5mC hydroxylase regulating the 5mC/5hmC status at target gene promoters. Biochemical and structural studies further demonstrate that the Tet3... [+]
Determined by: X-ray diffraction at 2.25 Å resolution
Chain C: LOC100036628 protein
Organism: Xenopus tropicalis
MATCHING STRUCTURES in PDB 15
FEATURES for TET1_HUMAN 36
1
200
400
600
800
2,136
11
1
2
1

Aquaria


AQUARIA
Updated: 3rd Feb 2015
About | Forum | Help
SPECIFY A PROTEIN
Protein name or ID
Synonyms: TET1_HUMAN, Methylcytosine dioxygenase TET1, CXXC-type... [+]
Gene: TET1
Organism
Homo sapiens, Human
ABOUT TET1_HUMAN
FUNCTION: Dioxygenase that catalyzes the conversion of the modified genomic... [+]
CATALYTIC ACTIVITY:
DNA 5-methylcytosine + 2-oxoglutarat + O(2) = DNA 5- ... [+]
COFACTOR:
Binds 1 Fe(2+) ion per subunit.
Binds 3 zinc ions per subunit. The ... [+]
SUBUNIT: Interacts with HCFC1
more details (5)
3D STRUCTURE TET1_HUMAN sequence aligned onto TET2 structure from PDB 4nm6-A (64% sequence identity)
Color
Representation
SELECTION
A: K(1453)
ABOUT PDB 4nm6
Biological Assembly
Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation.
Hu et al., Cell (2013)
Abstract: TET proteins oxidize 5-methylcytosine (5mC) on DNA and play important roles in various biological processes. Mutations of TET2 are frequently observed in myeloid malignance. Here, we present the crystal structure of human TET2 bound to methylated DNA at 2.02 Å resolution. The structure shows that two zinc fingers bring the Cys-rich and DSBH domains together to form a compact catalytic domain. The Cys-rich domain stabilizes the DNA above the DSBH core. TET2 specifically recognizes CpG dinucleotide and shows substrate preference for 5mC in a CpG context. 5mC is inserted into the... [+]
Determined by: X-ray diffraction at 2.03 Å resolution
Chain A: TET2 (Methylcytosine dioxygenase TET2)
MATCHING STRUCTURES in PDB 15
FEATURES for TET1_HUMAN 36
2OGFeDO domain, nucleic acid-modifying type (Residues 1580-2052)
View IPR024779 in InterPro
Inferred from InterPro motif similarity (polypeptide domain) INTERPRO
Molecule processing (mature protein region) UNIPROT
Click to load feature into 3D view; hover over features to see detailed info.
Region (polypeptide domain) UNIPROT
Region (polypeptide region) UNIPROT
Site (metal contact) UNIPROT

Exercise 1/4

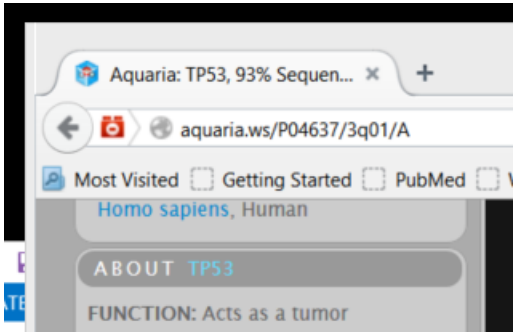
Starting aquaria
(May require a Java update)

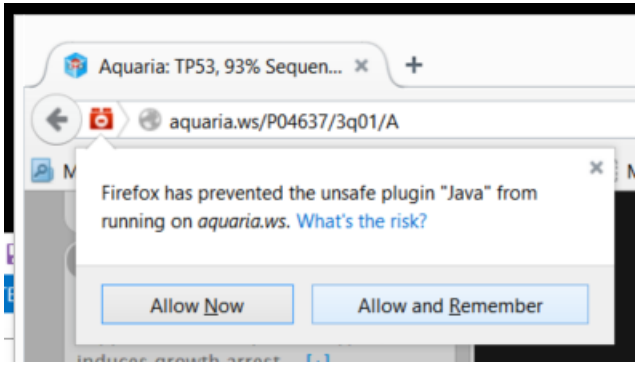
Works best in Firefox (in Chrome with reduced functionality)

Open Firefox mit JRE (from ZDV)

Go to http://aquaria.ws

Run an example. If JAVA blocked unblock it at the plugin icon

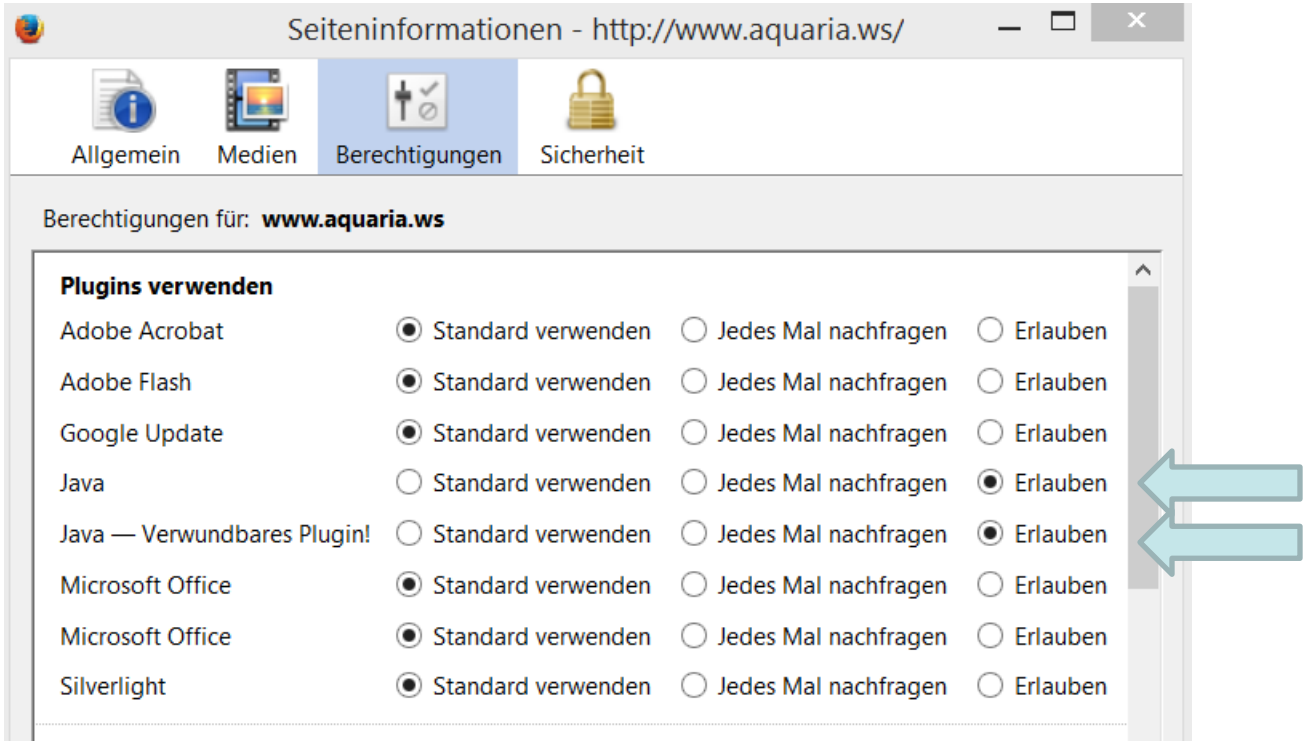




Exercise 1/4

Starting aquaria

Note that aquaria.ws requires that **two** java plug-ins that need to be allowed to run



Exercise 2/4

Comparing different matches in Myosin X

You can load a protein by its UniProt ID

Try Myosin X: http://aquaria.ws/Q9HD67/

Zoom in and out using the mouse wheel (or with shift and drag up and down).

Rotate by click and drag

Click on a residue to select. Shift + Click selects a range. Esc clears the selection.

Double click on a residue centers the molecule on it.

Right click and drag moves the molecule laterally

Compare the different hits with domain annotations using the feature view

Exercise 3/4

Comparing different matches in the human MR

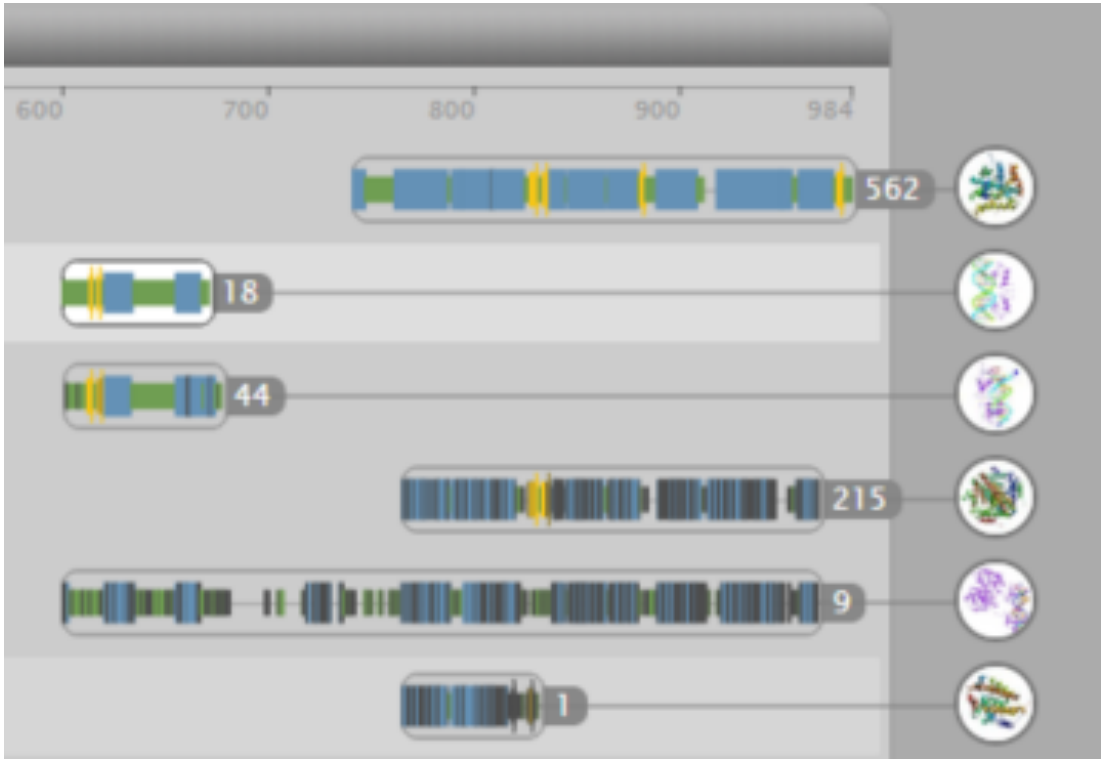
Type NR3C2 in protein name (human mineralocorticoid receptor)
Note and compare the multiple hits.

Which proteins are those?

What do they match in the human mineralocorticoid receptor?

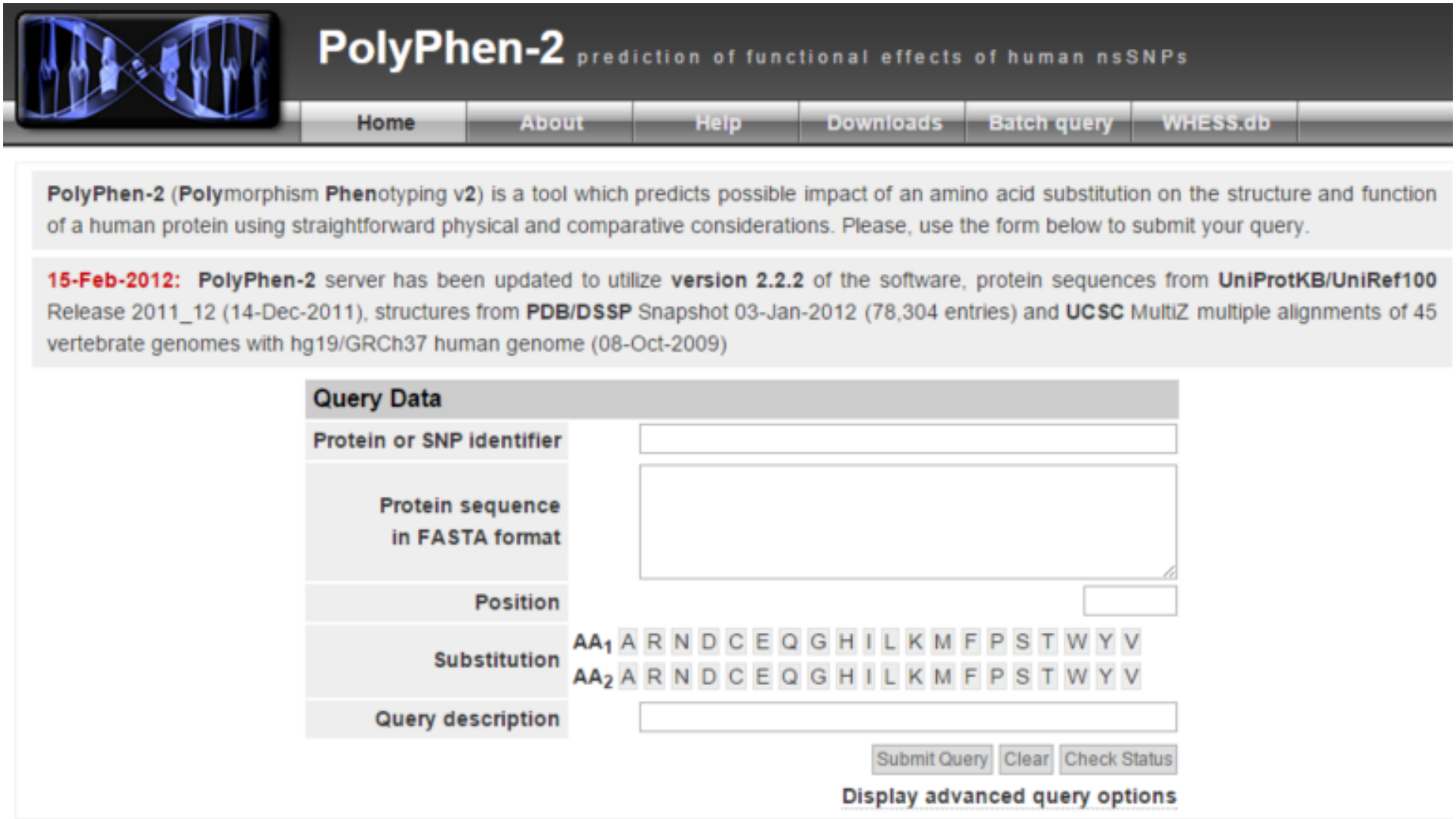
(Use the Features view)

The further down the less similar are the proteins compared. This is represented by a darker color.



Effect of mutations

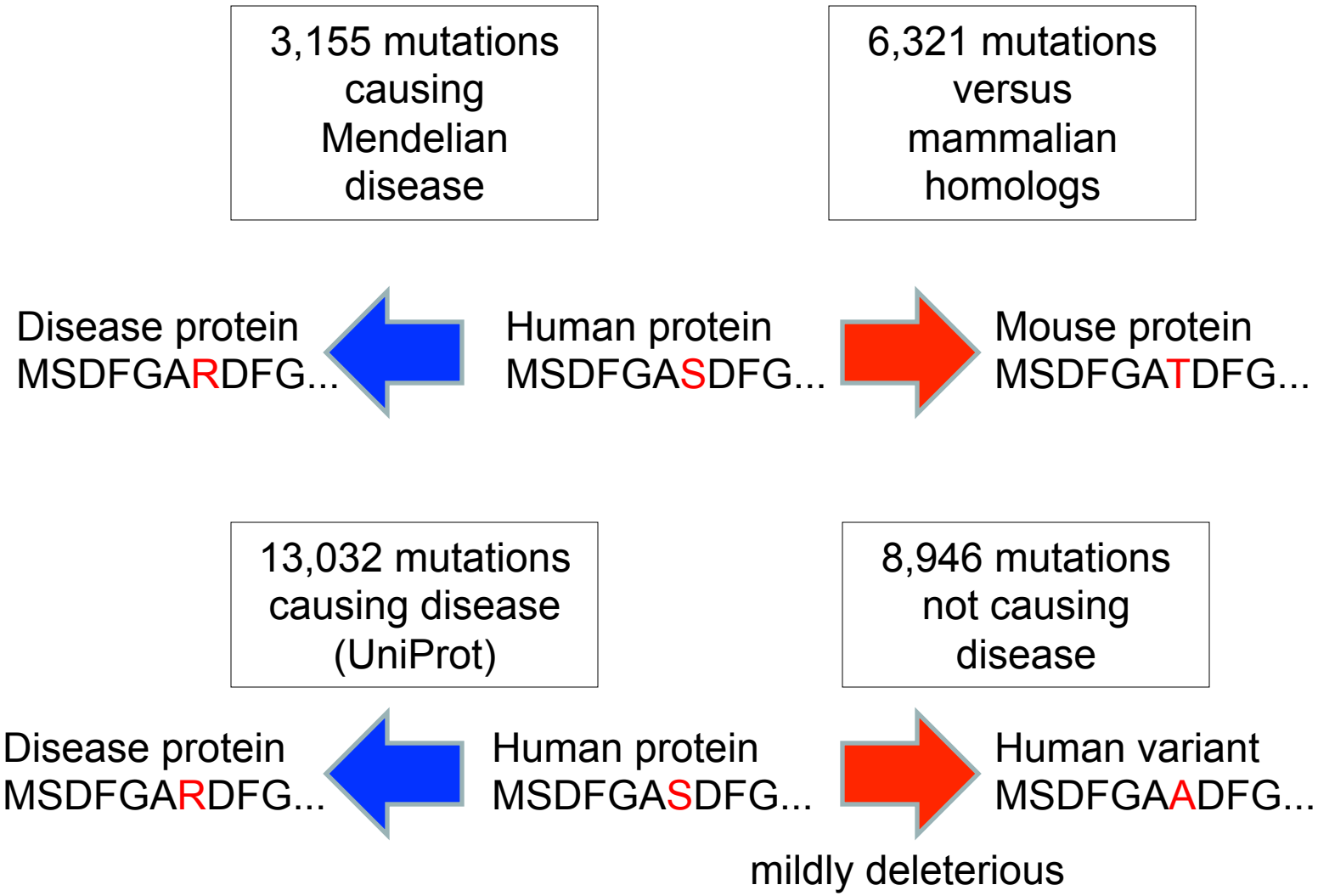
Polyphen2



http://genetics.bwh.harvard.edu/pph2/

Polyphen2

Training



Polyphen2

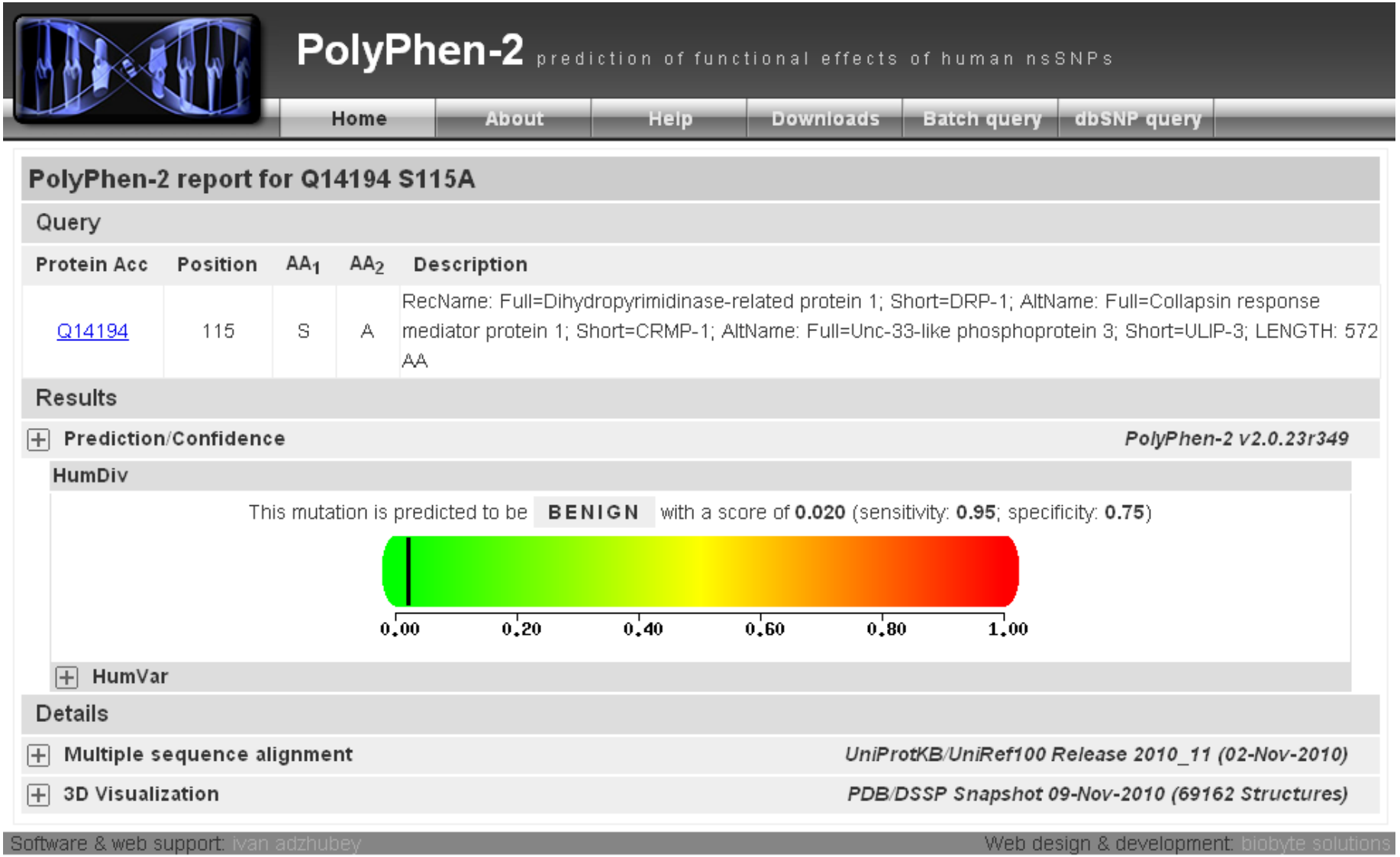
PSIC Score

Likelihood of an amino acid to occupy a specific position in the protein sequence given the pattern of amino acid substitutions observed in the multiple sequence alignment

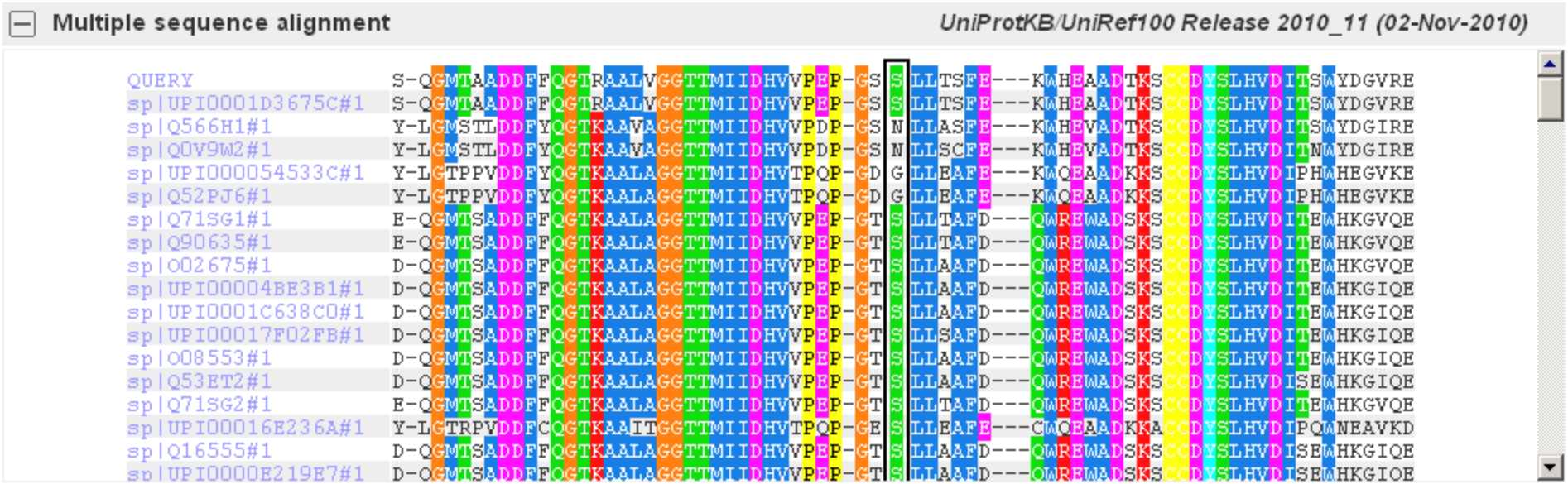
Polyphen2

Usage

Polyphen2


PolyPhen-2 prediction of functional effects of human nsSNPs
Home
About
Help
Downloads
Batch query
dbSNP query
PolyPhen-2 report for Q14194 S115A
Query
Protein Acc
Position
AA1
AA2
Description
Q14194
115
S
A
RecName: Full=Dihydropyrimidinase-related protein 1; Short=DRP-1; AltName: Full=Collapsin response mediator protein 1; Short=CRMP-1; AltName: Full=Unc-33-like phosphoprotein 3; Short=ULIP-3; LENGTH: 572 AA
Results
Prediction/Confidence
PolyPhen-2 v2.0.23r349
HumDiv
This mutation is predicted to be BENIGN with a score of 0.020 (sensitivity: 0.95; specificity: 0.75)
0.00
0.20
0.40
0.60
0.80
1.00
HumVar
Details
Multiple sequence alignment
UniProtKB/UniRef100 Release 2010_11 (02-Nov-2010)
3D Visualization
PDB/DSSP Snapshot 09-Nov-2010 (69162 Structures)
Software & web support: ivan adzhubey
Web design & development: biobyte solutions

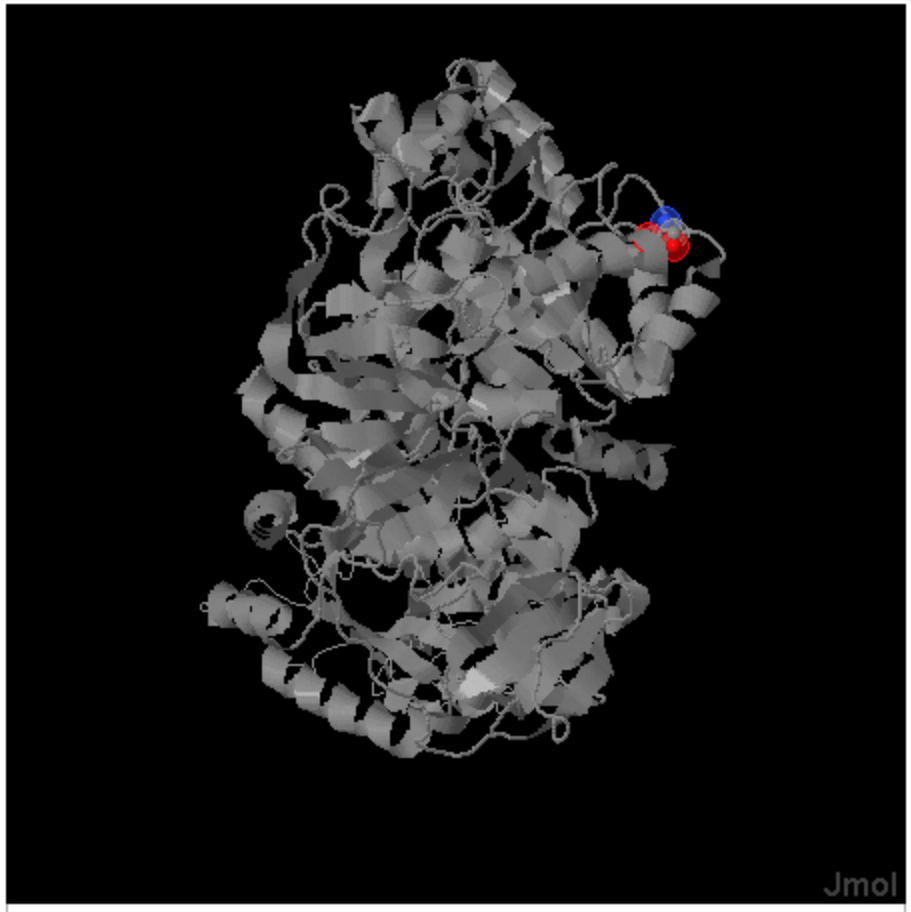
Polyphen2


Multiple sequence alignment
UniProtKB/UniRef100 Release 2010_11 (02-Nov-2010)
QUERY S-QGMTAADDFFQGTRAALVGGTTMIIDHVVPEP-GSSLLTSFE---KAHEAADTKSCCDYSLHVDITSAYDGVRE
sp|UPI0001D3675C#1 S-QGMTAADDFFQGTRAALVGGTTMIIDHVVPEP-GSSLLTSFE---KAHEAADTKSCCDYSLHVDITSAYDGVRE
sp|Q566H1#1 Y-LGMSTLDDFYQGTKAAVAGGTTMIIDHVVPDP-GSNLLASFE---KAHEVADTKSCCDYSLHVDITSAYDGIRE
sp|Q0V9W2#1 Y-LGMSTLDDFYQGTKAAVAGGTTMIIDHVVPDP-GSNLLSCFE---KAHEVADTKSCCDYSLHVDITNAYDGIRE
sp|UPI000054533C#1 Y-LGTPPVDDFYQGTKAALAGGTTMIIDHVTPQP-GDGLLEAFE---KAQEAADKKSCCDYSLHVDIPHAHEGVKE
sp|Q52PJ6#1 Y-LGTPPVDDFYQGTKAALAGGTTMIIDHVTPQP-GDGLLEAFE---KAQEAADKKSCCDYSLHVDIPHAHEGVKE
sp|Q71SG1#1 E-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLTAFD---QWREWADSKSCCDYSLHVDITEAHKGVQE
sp|Q90635#1 E-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLTAFD---QWREWADSKSCCDYSLHVDITEAHKGVQE
sp|O02675#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDITEAHKGVQE
sp|UPI00004BE3B1#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDITEAHKGIQE
sp|UPI0001C638C0#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDITEAHKGIQE
sp|UPI00017F02FB#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLSAFD---QWREWADSKSCCDYSLHVDITEAHKGIQE
sp|O08553#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDITEAHKGIQE
sp|Q53ET2#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDISEAHKGIQE
sp|Q71SG2#1 E-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLTAFD---QWREWADSKSCCDYSLHVDITEAHKGVQE
sp|UPI00016E236A#1 Y-LGTRPVDDFCQGTKAAITGGTTMIIDHVTPQP-GESLLEAFE---CWQEAADKKACCDYSLHVDIPQANEAVKD
sp|Q16555#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDISEAHKGIQE
sp|UPI0000E219E7#1 D-QGMTSADDFFQGTKAALAGGTTMIIDHVVPEP-GTSLLAAFD---QWREWADSKSCCDYSLHVDISEAHKGIQE

Polyphen2

3D Visualization

PDB/DSSP Snapshot 09-Nov-2010 (69162 Structures)



EntryID: 1KCX

ChainID: B

Residue: Ser115

Identity: 97.1%

Overlap: 83.2% (476 aa)

Exercise 4/4

Study the effect of mutants with Polyphen2

•Let's see if you can design a damaging and a benign mutation for human myosin X (open in chimera PDB 3PZD to view and select candidate mutations; pick from chain A).

•Go to the Polyphen2 home page: http://genetics.bwh.harvard.edu/pph2/

•Type the UniProt id of the protein sequence "Q9HD67" in the Protein Identifier window. Type the position of your candidate for a damaging mutation. Select in AA1 the type of amino acid at that position. Now, select an amino acid to mutate to. May be try one with a large side chain, or if the wild type one was hydrophobic, try a hydrophilic one. Be nasty! Then hit Submit Query.
What result did you get? Is it close to one?

•Try your benign mutation in the same way. This time may be choose to mutate to a similar residue to the wild type one. Be gentle! Then hit Submit Query.
What result did you get? Is it close to zero?