BCHS 6229
Protein Structure and Function

Lecture 7 (November, 2011)

From Sequence to Function (II):
Sequences and Topology
Structural Biology Knowledgebase

Protein diversity has accumulated over a looonngg time

3D structures more conserved in evolution than aa sequence!

Alignment of protein sequences reveals conserved closely and distantly related families

Alignment of protein structures reveals convergence of fold and function, or extreme divergence of sequence

mechanisms accounting for structural irregularities?

Mutation can create altered function of proteins; gene duplication; recombination

Circular permutation of genes (N and C-termini are close); inteins

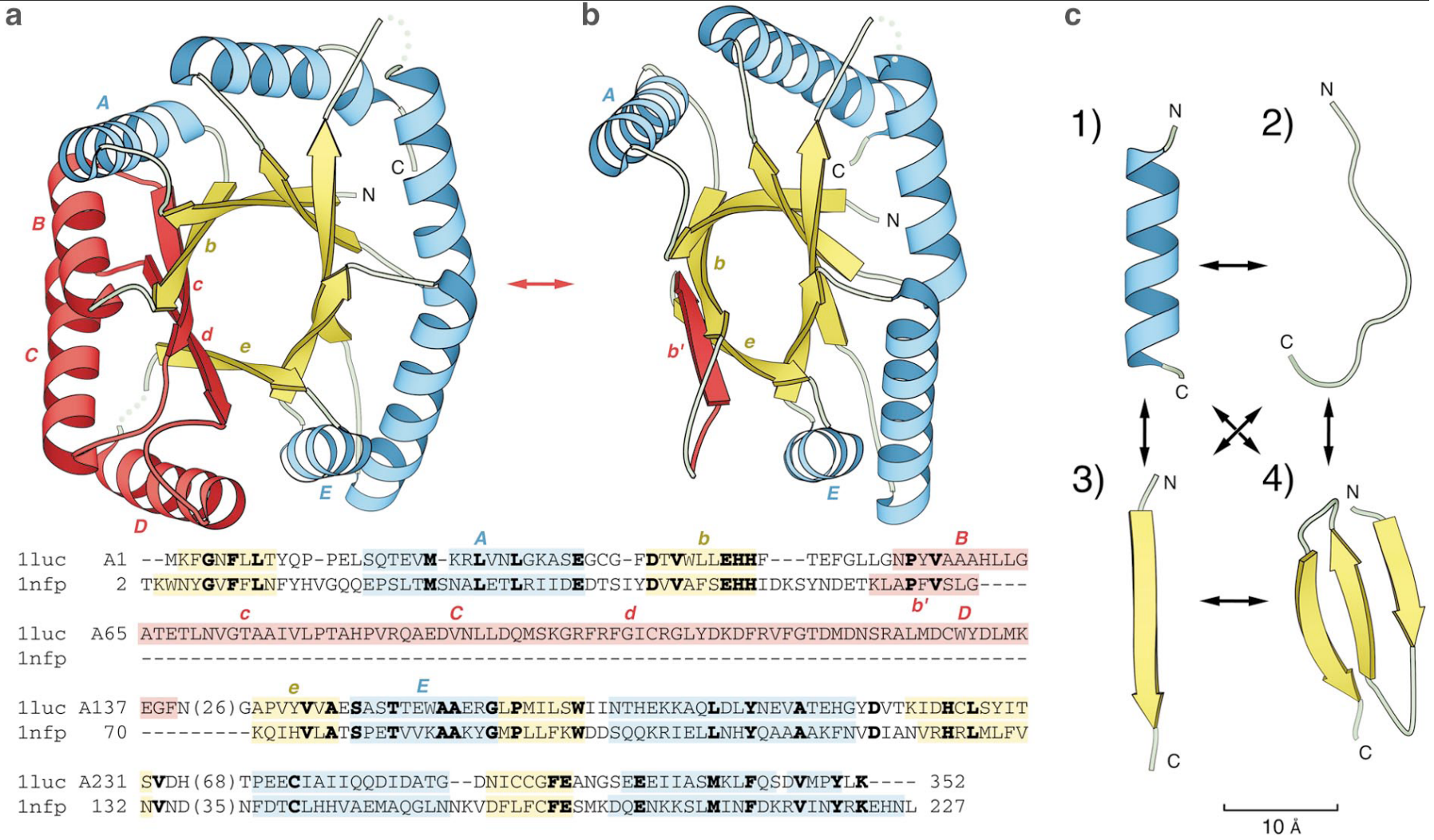
Deletions and insertions

Mutations that result in loss of function cannot accumulate unless the gene/protein is non-essential (or duplicated); may later provide a selective advantage

Fold Change in Evolution of Protein Structures

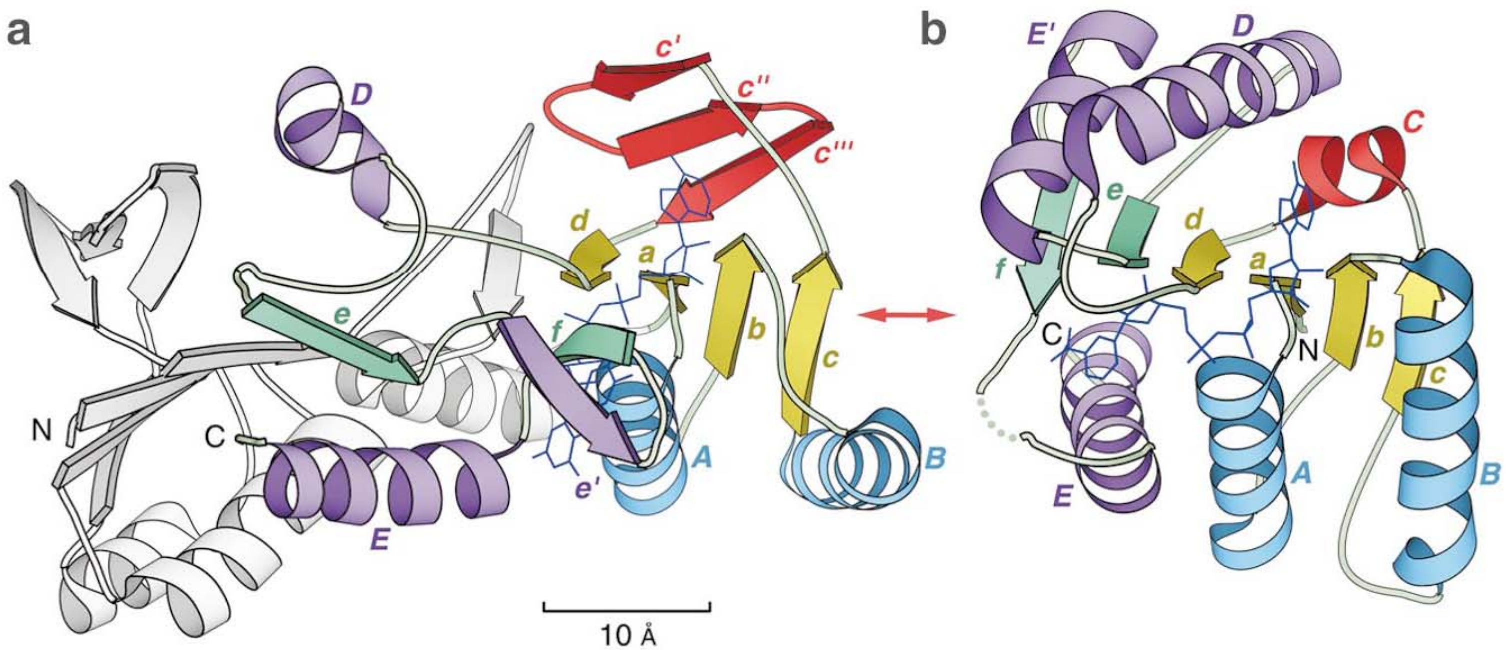
Nick V. Grishin

[Journal of Structural Biology 134, 167–185 (2001)]



1luc - Bacterial Luciferase

1nfp - Nonfluorescent Flavoprotein



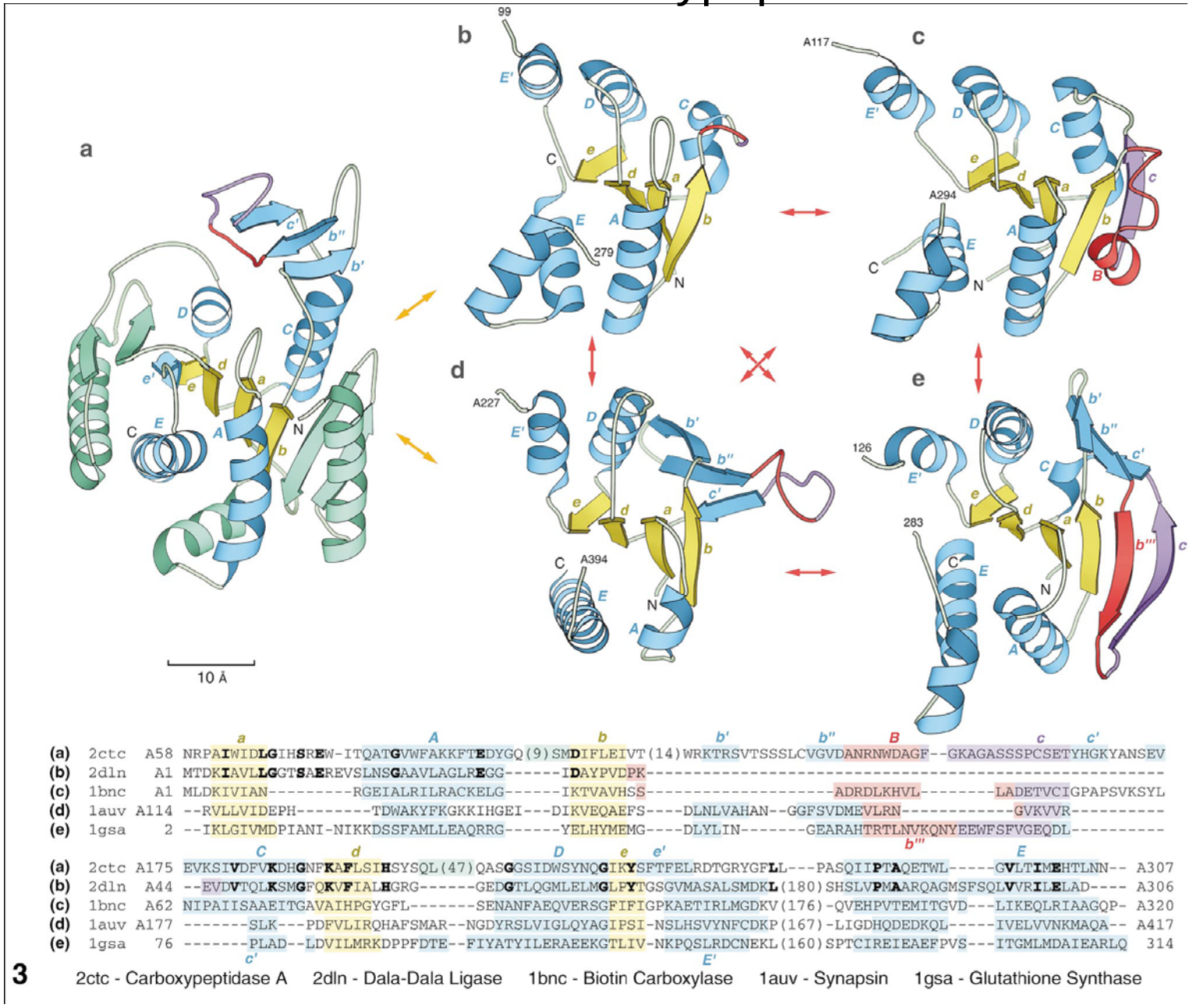

```
               a             A               b                    B                 c
(a) 1npx  149 VNNVVVIGSGYIGIEAAEAFAKAG--KKVTVIDILDRPLGVYLDKEFTDVLTEEMEA--------NNIT
(b) 1ldn  A20 GARVVVIGAGFVGASYVFALMNQGIADEIVLIDA---------NESKAIGDAMDFNHGKVFAPKPVDIW

                     c'       c''      c'''  d                              D
(a) 1npx  208 IATGETVERYEGDGRVQKVVTDKNAYDADLVVVAVGVRP------------NTAWLKGTLELHPNGLIK
(b) 1ldn  A80 HGDYD-----------------DCRDADLVVICAGANQKPGETRLDLVDKNIAIFRSIVES-------
                C          e        e'             f              E
(a) 1npx  265 TDEYMRTSEPDVFAVGD--ATLIKYN-------PADTEVNIA----LATNARKQGRFAVKNLEE   315
(b) 1ldn A124 ---VMASGFQGLFLVATNPVDILTYATWKFSGLPHERVIGSG(85)YYGIAMGLARVTRAILHN A265
                                   E'
```

2 1npx - NADH Peroxidase 1ldn Lactate Dehydrogenase

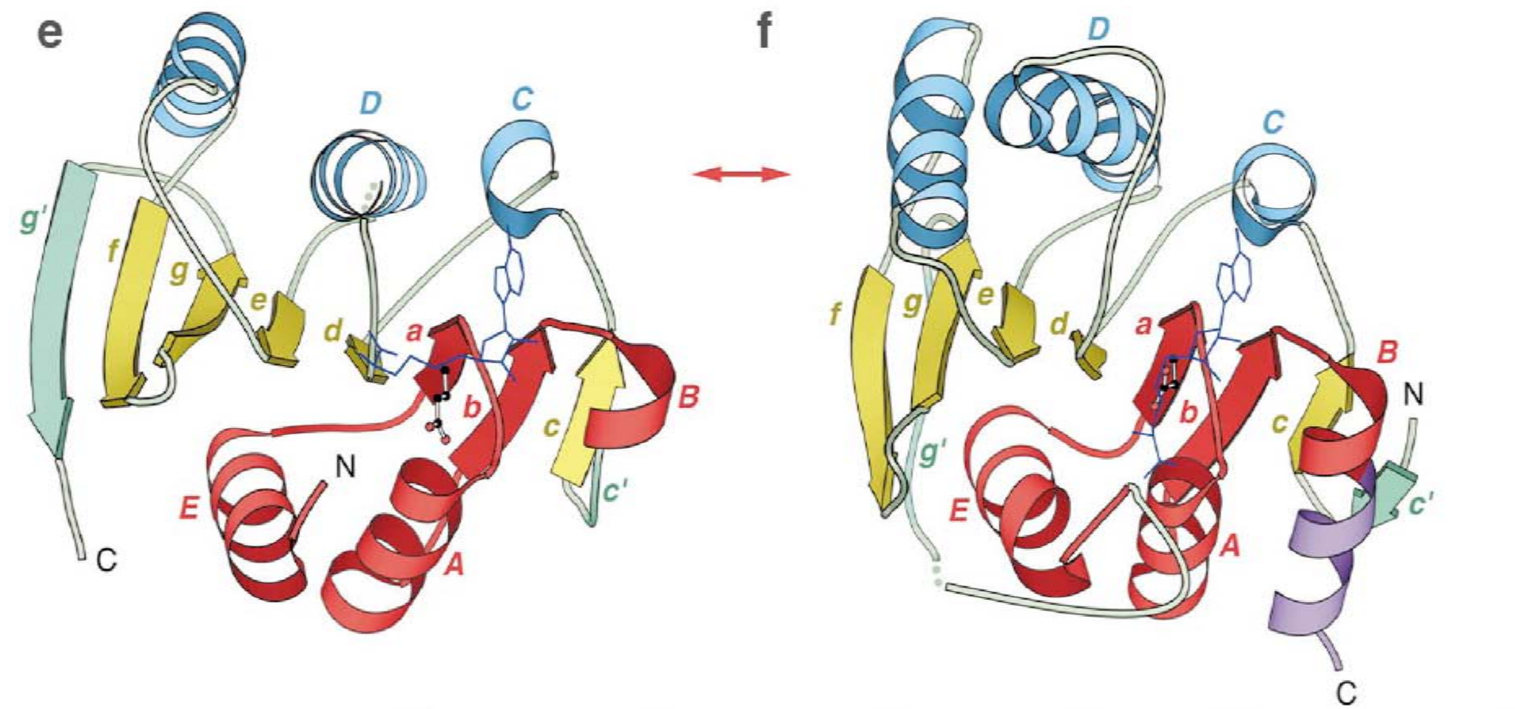
Rossmann fold-like proteins

Rossmann fold-like domains of ATP-grasp proteins and zinc-carboxypeptidase.



3 2ctc - Carboxypeptidase A 2dln - Dala-Dala Ligase 1bnc - Biotin Carboxylase 1auv - Synapsin 1gsa - Glutathione Synthase

Circular permutations




```
                        E                    a              A            b      B            c'
(e) 1aqi  A21  ETPPEVVDFMVSLAEAPRGGRVLEPACAHGPFLRAFREAHGTGYRFVGVEIDPKAL------------DL
(f) 1boo  249-RFPAKLPEFFIRML-TEPDDLVVDIFGGSNTTGLVAER---ESRKWISFEMKPEYV(8)-318 22-AYT

                  c       C              d                D                 e
(e) 1aqi  A80  PPWAEGILADFLLW-EPG--EAFDLILGNPPYG(31)NLYGAFLEKAVRLLKPGGVLVFVVPATWL----
(f) 1boo   25  TSNGSMYIGDSLELLESFPEESISLVMTSPPFA(16)DWFLSFAKVVNKKLKPDGSFVVDFGGAYMKGVP

                                      f                    g          g'
(e) 1aqi A170  --VLEDFALLREFLAREGK--TSVYYLGEVF-----PQKKVSAVVIRFQKSGKGLSLWDTQE--- A222
(f) 1boo  107  ARSIYNFRVLIRMIDEVGFFLAEDFYWFNPS(13)IRVKDAVNTVWWFSKTEWPKSDITK(72)A- 248
```

1aqi - Adenine-N6-Dna-Methyltransferase Taqi

1boo - PvuII Dna Methyltransferase

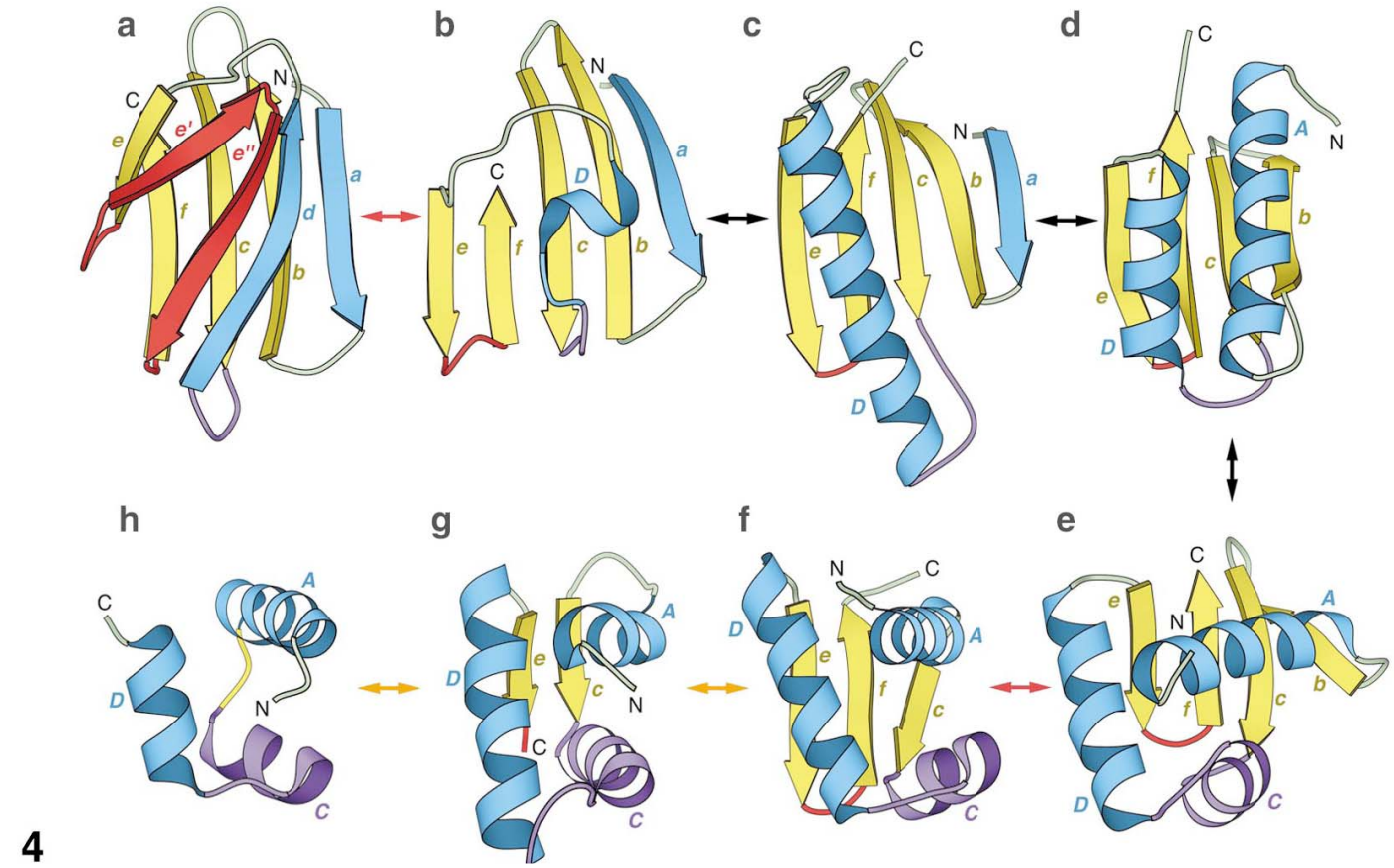


FIG. 4. A path from all-β to all-α proteins.
(a) *Bacillus licheniformis* α-amylases, C terminal domain (1bpl);
(b) *Pseudomonas stutzeri* G4-amylase C-terminal domain (2amg);
(c) γ-subunit of glycogen phosphorylase kinase N-terminal domain (1phk);
(d) sonic hedgehog N-terminal signaling domain (1vhh);
(e) catabolite gene activator protein (CAP), C-terminal domain (1cgp);
(f) biotin repressor N-terminal domain (1bia);
(g) ribosomal protein L11 C-terminal domain (1fow);
(h) HIN recombinase DNA-binding domain (1hcr).

Protein Design Principles

Need a stable hydrophobic core with constrained rotamer conformation for individual side chains (high stability can be achieved without a well-ordered core)

Core depends on three-dimensional arrangement of secondary structure elements and vice versa (breaking α helix, over/underpacking a core)

Natural deviations in core and surface can be associated with disease states resulted from misassociated to misfolded proteins

Amino acid diversity is ultimately required to reflect naturally occurring proteins

Solution experiments on designed proteins can test the computational methods

Topological variation in functionally diverse enzyme superfamily

(Current Opinion in Structural Biology 2011, 21:391–397)

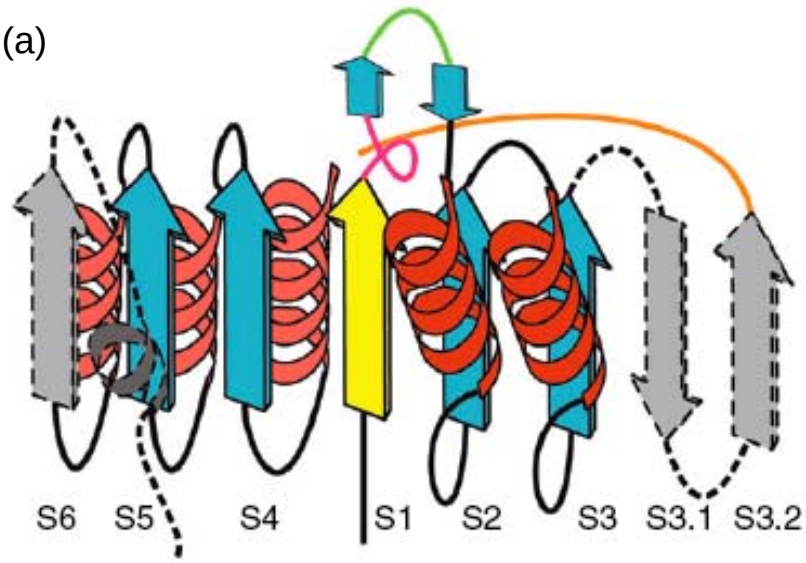
1. The haloalkanoic acid dehalogenase (HAD) SF: cap domain variations enable divergent evolution of many different reaction and substrate specificities.

2. The vicinal oxygen chelate (VOC) SF: mixing and matching subdomains for functional versatility.

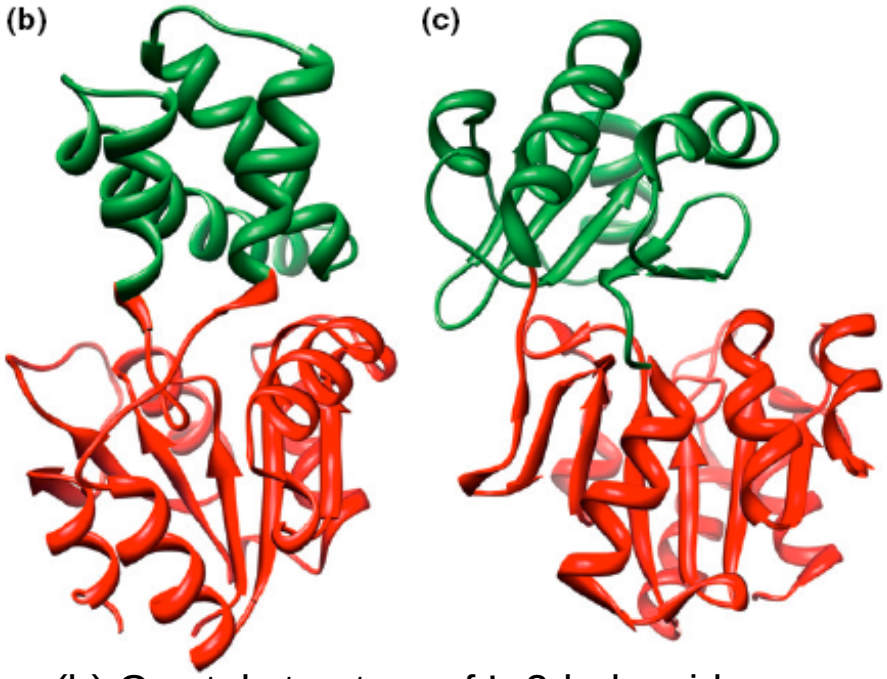
3. The thioredoxin (Trx)-fold like SFs: varied inserts and domain additions extend the redox repertoire of the canonical Trx-fold

Topologies of HAD and cap domains

(a)



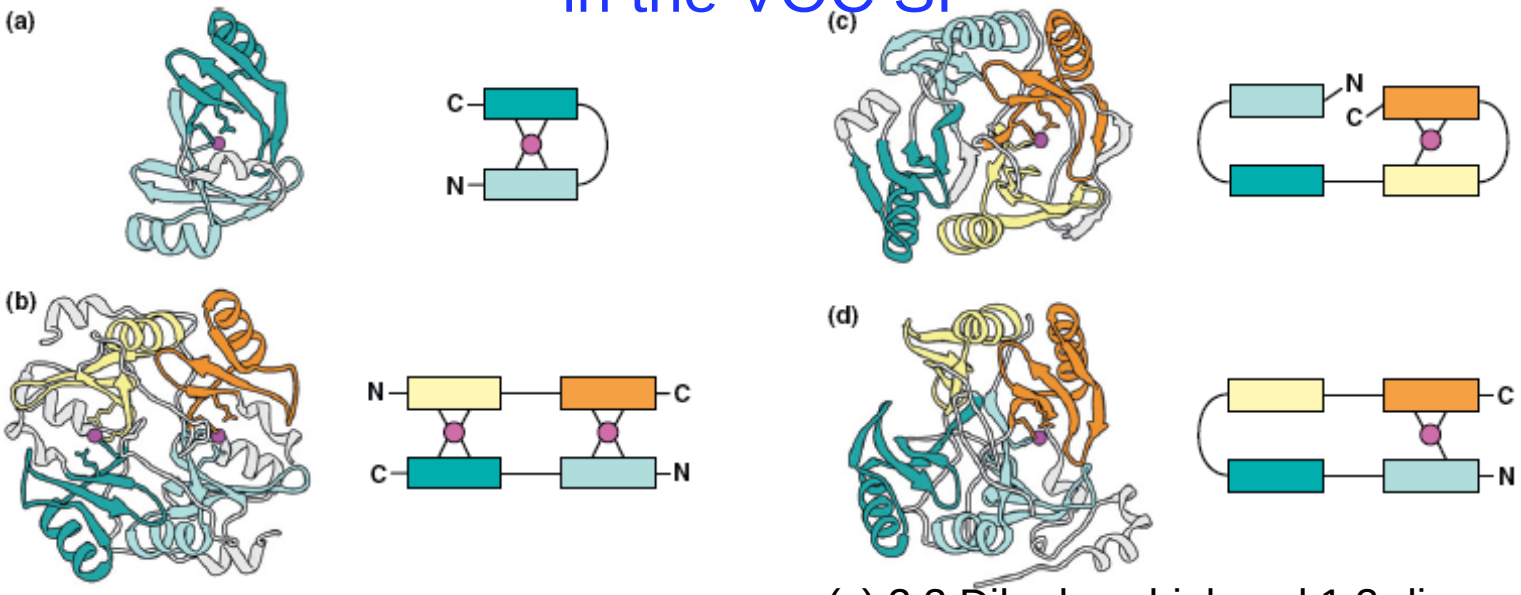
(a) Schematic diagram of the classic HAD domain: yellow, strand containing the catalytic Asp residue; blue, core strands conserved in all HAD SF members; gray, structural elements that may not have occurred in the ancestral structure; green line shows the insertion point for C1 caps; orange line shows the insertion point for C2 caps. Broken lines indicate secondary structure elements not present in all members containing the HAD domain.



(b) Crystal structure of L-2-haloacid dehalogenase with 2-chloro-N-butyrate (ligand not shown) (PDB 1ZRM) showing the HAD domain in red and the C1 cap domain in green.

(c) Crystal structure of the HAD subclass IIB sugar phosphatases (PDB 1YMQ) showing the HAD domain in red and the C2 cap domain in green.

Examples of alternate arrangements of paired βαβββ modules in the VOC SF



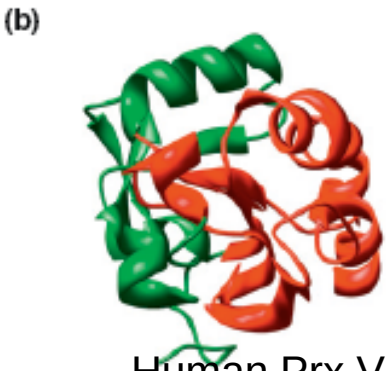
(a) Glyoxalase I from *Clostridium* (PDB 3HDP), in which two modules from a single chain pair to form a metal site.
(b) Human glyoxalase I (PDB 1QIN), in which two chains each containing two modules pair in head-to-tail fashion to form two metal sites.

(c) 2,3 Dihydroxybiphenyl 1,2-dioxygenase from Burkholderia (PDB 1KMY): the four modules in a single chain pair in the order 1–2 and 3–4, and only the latter pair forms a metal site.
(d) Protein of unknown function from Bacillus (PDB 1ZSW): the four modules in a single chain pair in the order 1–4 and 2–3, and only the former pair forms a metal site.

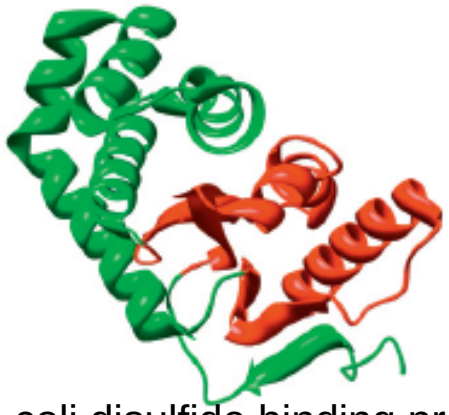
metal ions are magenta and different modules are shown in different colors to highlight the repeating babbb unit. Metal-ligating residues may occur in the first and/or last beta-strands of a module. Parts of the structures not within βαβββ modules are shown in light gray.

Examples of topological variations for Trx-fold SFs

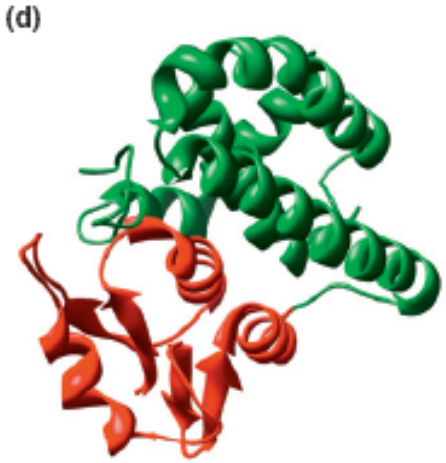
Human Trx
(PDB 1AUC)



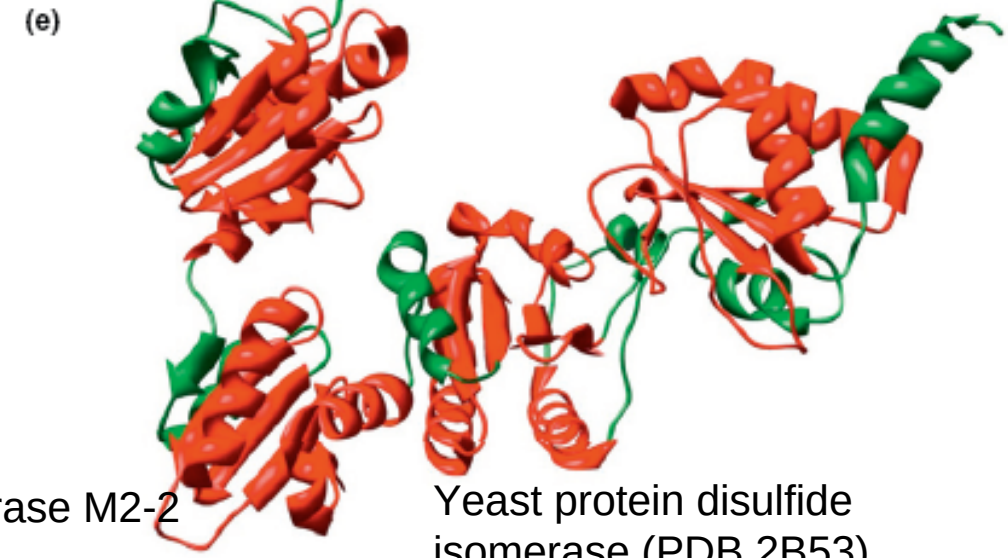
Human Prx V
(PDB 1HD2)



E. coli disulfide binding protein A
(PDB 1FVK)



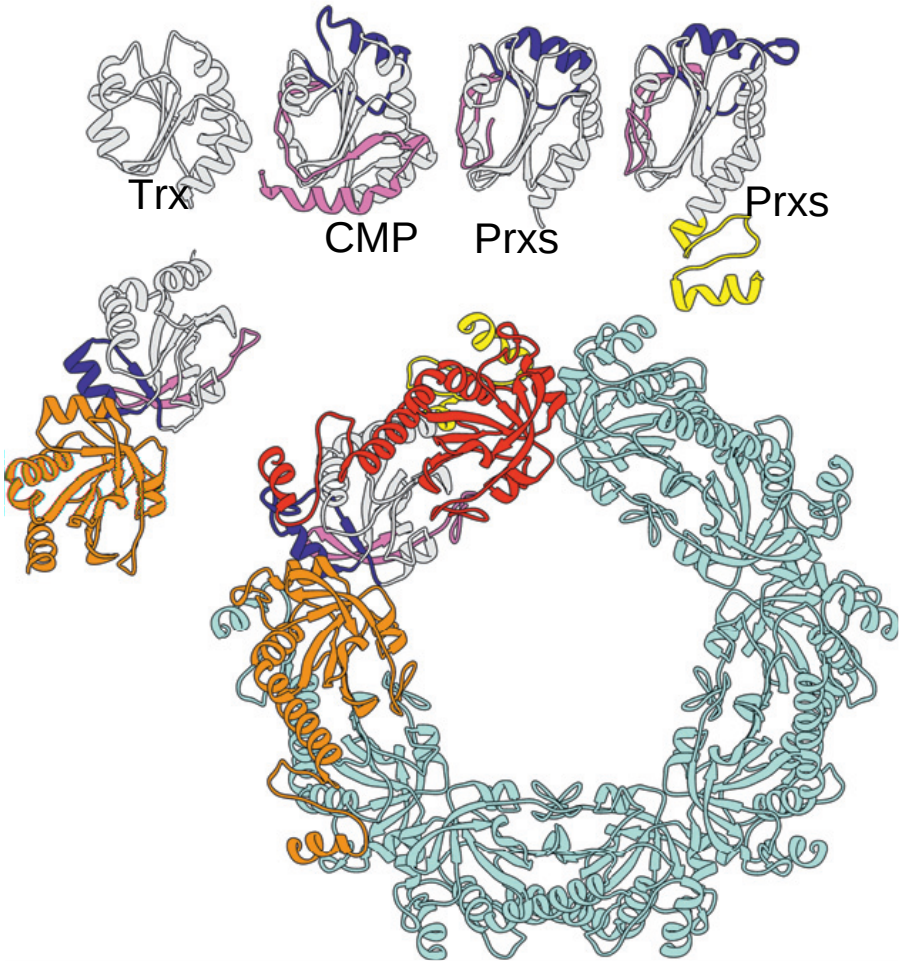
Human glutathione transferase M2-2
(PDB 1XW5).



Yeast protein disulfide
isomerase (PDB 2B53).

Red: Trx-fold, green: inserts or domain pairing additions to the Trx-fold.

Structures of Trx, CMP, and Prxs are consistent with their connections in sequence space, in which CMPs link the other two SFs



The canonical Trx-fold is shown in light gray, an N-terminal extension in darkpink, an insert that includes a helix in blue, and a C-terminal extension in yellow. The bottom row shows the same two Prxs as in the top row, but in homo multimeric assemblies. Bottom left, Prx dimer (PDB 1XXU chains A and B) with one subunit colored as in the top row, the other subunit in orange. This dimer uses one type of interface (marked I) that primarily involves the insert helix (blue). Right, Prx decamer (PDB 1QMV) with one monomer colored as in the top row, a second monomer in orange interacting with the first through the type I interface, a third monomer in red interacting with the first through the type II interface, and the remaining seven monomers in aqua. The type II interface is primarily a concatenation of the central beta-sheet of the Trx-fold, but in this decamer, the C-terminal extension (yellow) is also involved.

PSI SBKB

The Structural Biology Knowledgebase

by Protein Structure Initiative (PSI) and Nature Publishing Group (NPG)

Adopted from Materials prepared by Jennifer Williams, Ph.D.

(Updated: Q1 2011)

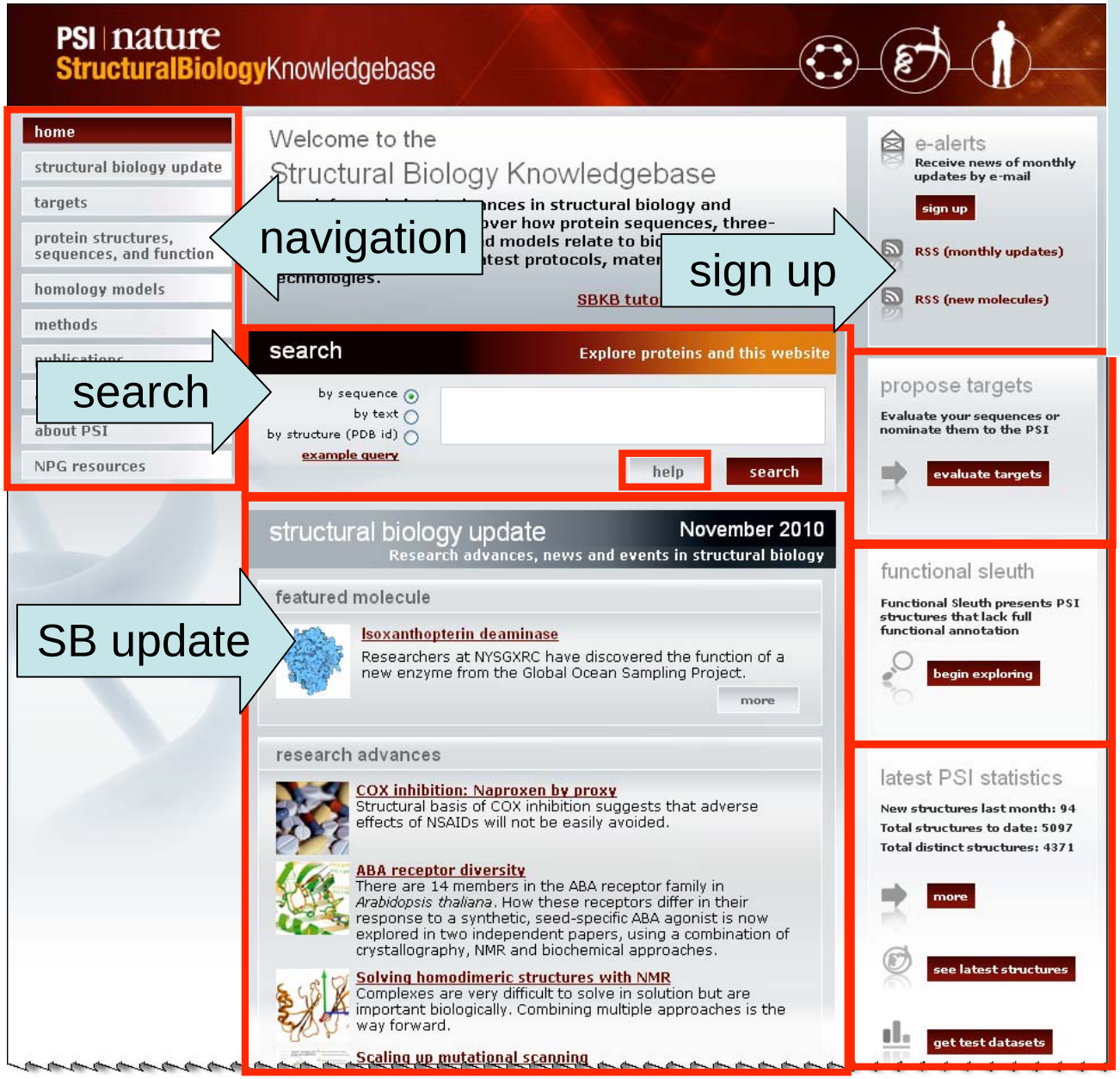
www.openhelix.com

PSI SBKB

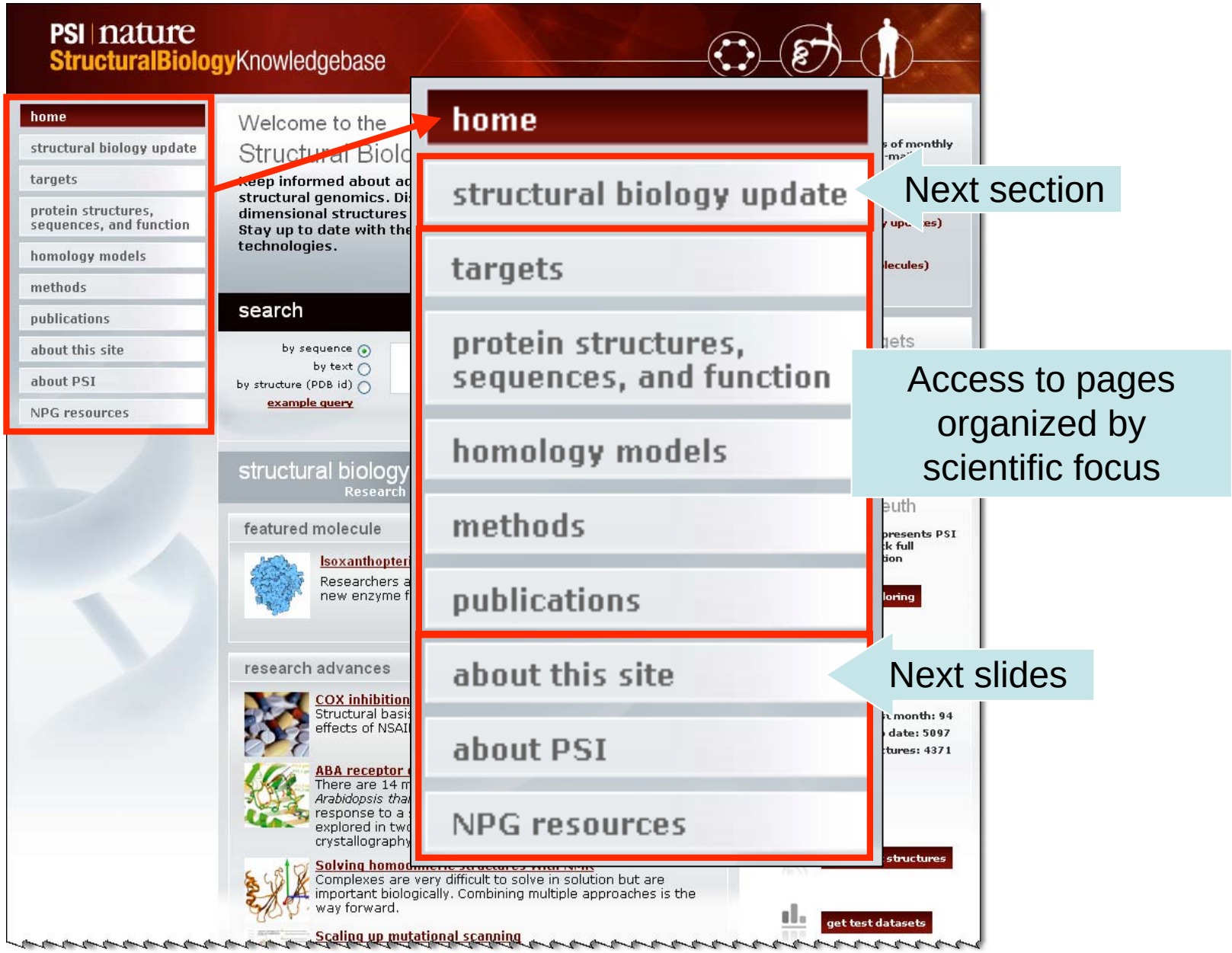
1. Introduction and Credits
2. Structural Biology Update
3. Sequence or Structure Search
4. Text Searches
5. Additional Features
6. Summary
7. Exercises

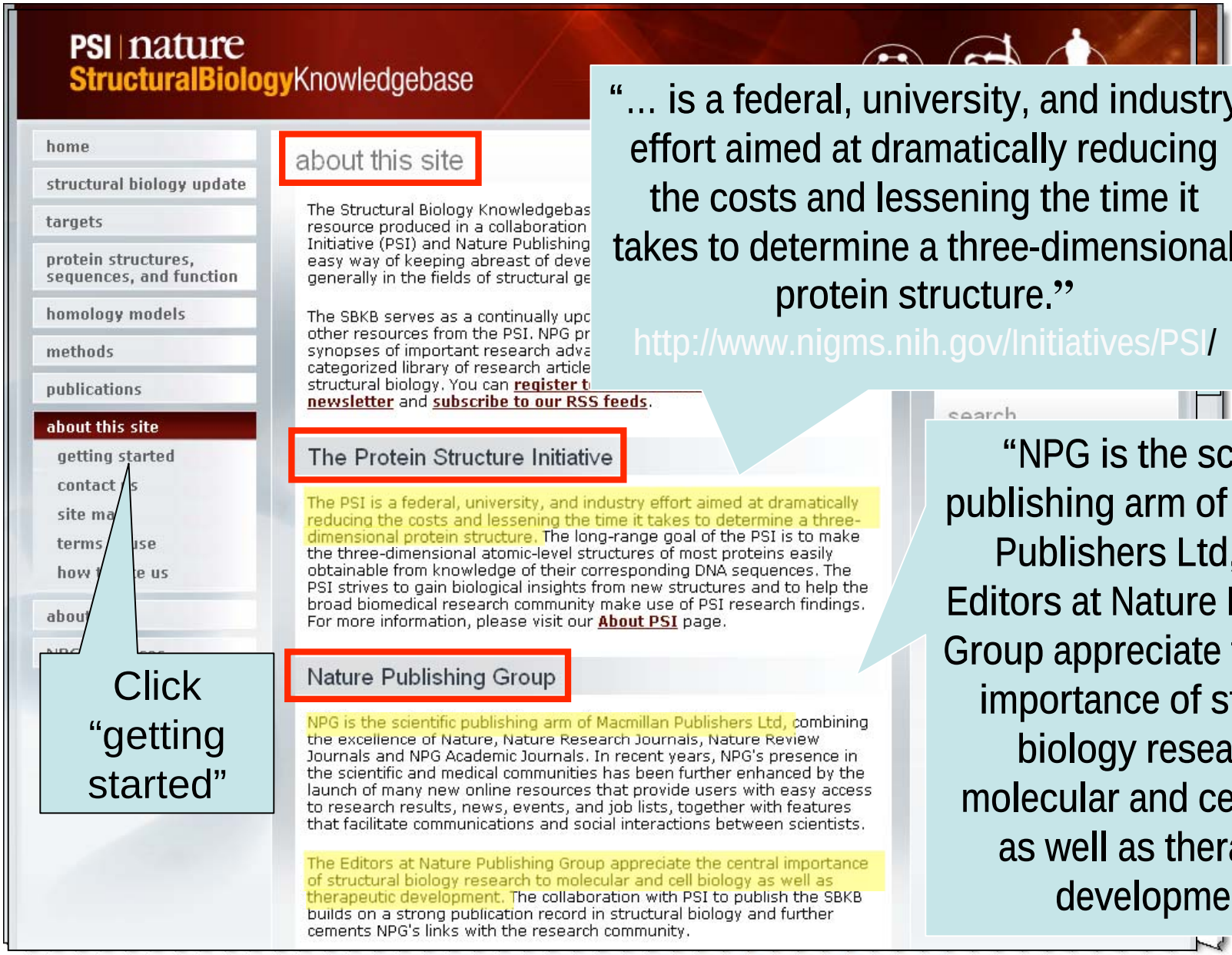
PSI SBKB: http://www.sbkb.org/

1. PSI SBKB Homepage



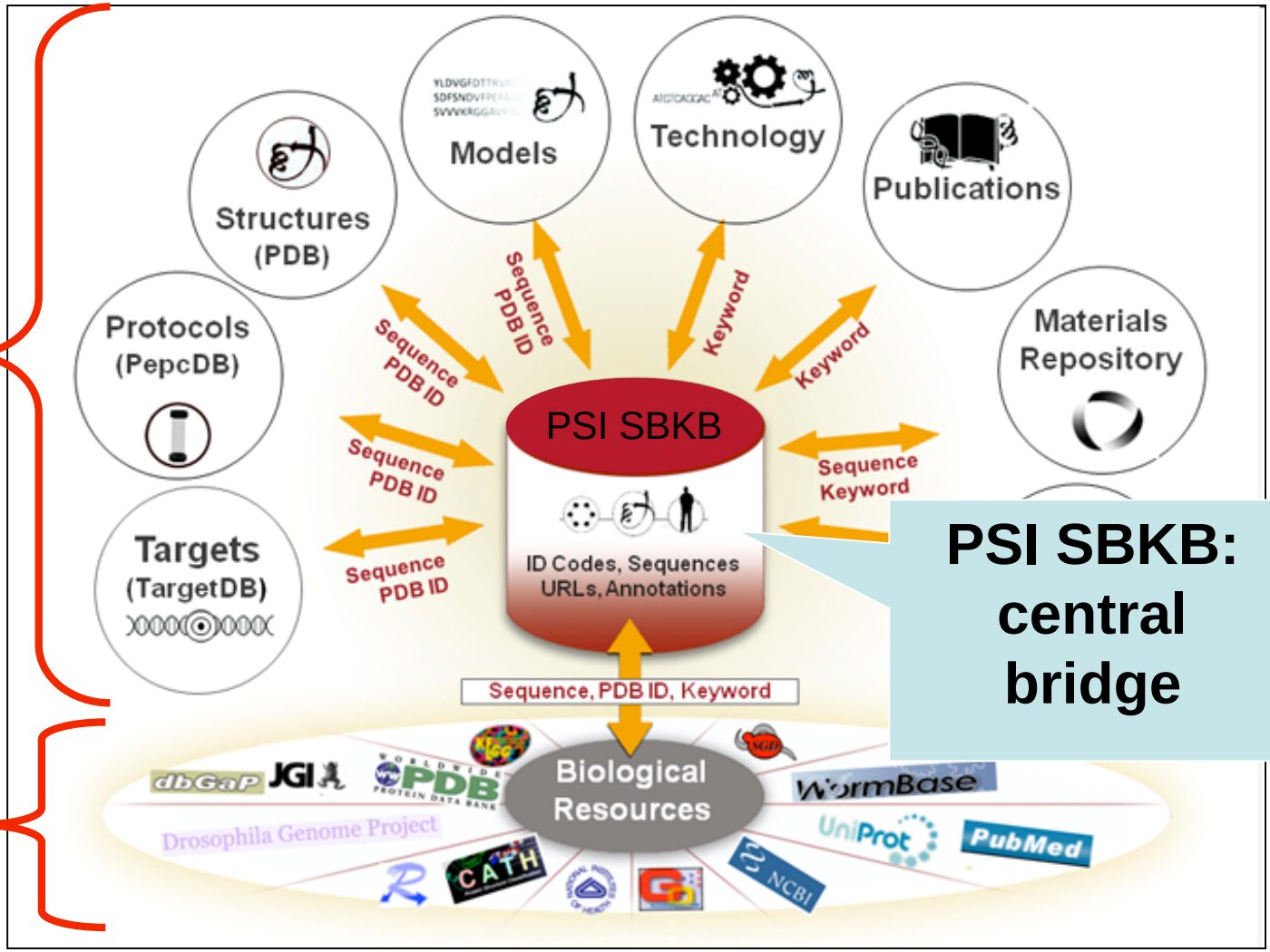
"This '<u>one-stop shop</u>' provides users with the available genetic, structural, functional and experimental information about a particular protein of interest. "


PSI | nature
StructuralBiologyKnowledgebase
home
structural biology update
targets
protein structures, sequences, and function
homology models
methods
publications
about this site
about PSI
NPG resources
Welcome to the
search
by sequence
by text
by structure (PDB id)
example query
structural biology
Research
featured molecule
research advances
COX inhibition
ABA receptor
Solving homo
Complexes are very difficult to solve in solution but are important biologically. Combining multiple approaches is the way forward.
Scaling up mutational scanning
get test datasets
Next section
Access to pages organized by scientific focus
Next slides


PSI | nature
StructuralBiologyKnowledgebase
home
structural biology update
targets
protein structures, sequences, and function
homology models
methods
publications
about this site
getting started
about this site
The Structural Biology Knowledgebas
resource produced in a collaboration
Initiative (PSI) and Nature Publishing
easy way of keeping abreast of deve
generally in the fields of structural ge
The SBKB serves as a continually upd
other resources from the PSI. NPG pr
synopses of important research adva
categorized library of research article
structural biology. You can register t
newsletter and subscribe to our RSS feeds.
The Protein Structure Initiative
The PSI is a federal, university, and industry effort aimed at dramatically reducing the costs and lessening the time it takes to determine a three-dimensional protein structure. The long-range goal of the PSI is to make the three-dimensional atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences. The PSI strives to gain biological insights from new structures and to help the broad biomedical research community make use of PSI research findings. For more information, please visit our About PSI page.
Nature Publishing Group
NPG is the scientific publishing arm of Macmillan Publishers Ltd, combining the excellence of Nature, Nature Research Journals, Nature Review Journals and NPG Academic Journals. In recent years, NPG's presence in the scientific and medical communities has been further enhanced by the launch of many new online resources that provide users with easy access to research results, news, events, and job lists, together with features that facilitate communications and social interactions between scientists.
The Editors at Nature Publishing Group appreciate the central importance of structural biology research to molecular and cell biology as well as therapeutic development. The collaboration with PSI to publish the SBKB builds on a strong publication record in structural biology and further cements NPG's links with the research community.
Click "getting started"
"... is a federal, university, and industry effort aimed at dramatically reducing the costs and lessening the time it takes to determine a three-dimensional protein structure."
http://www.nigms.nih.gov/Initiatives/PSI/
"NPG is the scientific publishing arm of Macmillan Publishers Ltd, ... The Editors at Nature Publishing Group appreciate the central importance of structural biology research to molecular and cell biology as well as therapeutic development."

PSI Portals & Resources

Biological Resources



PSI SBKB: central bridge

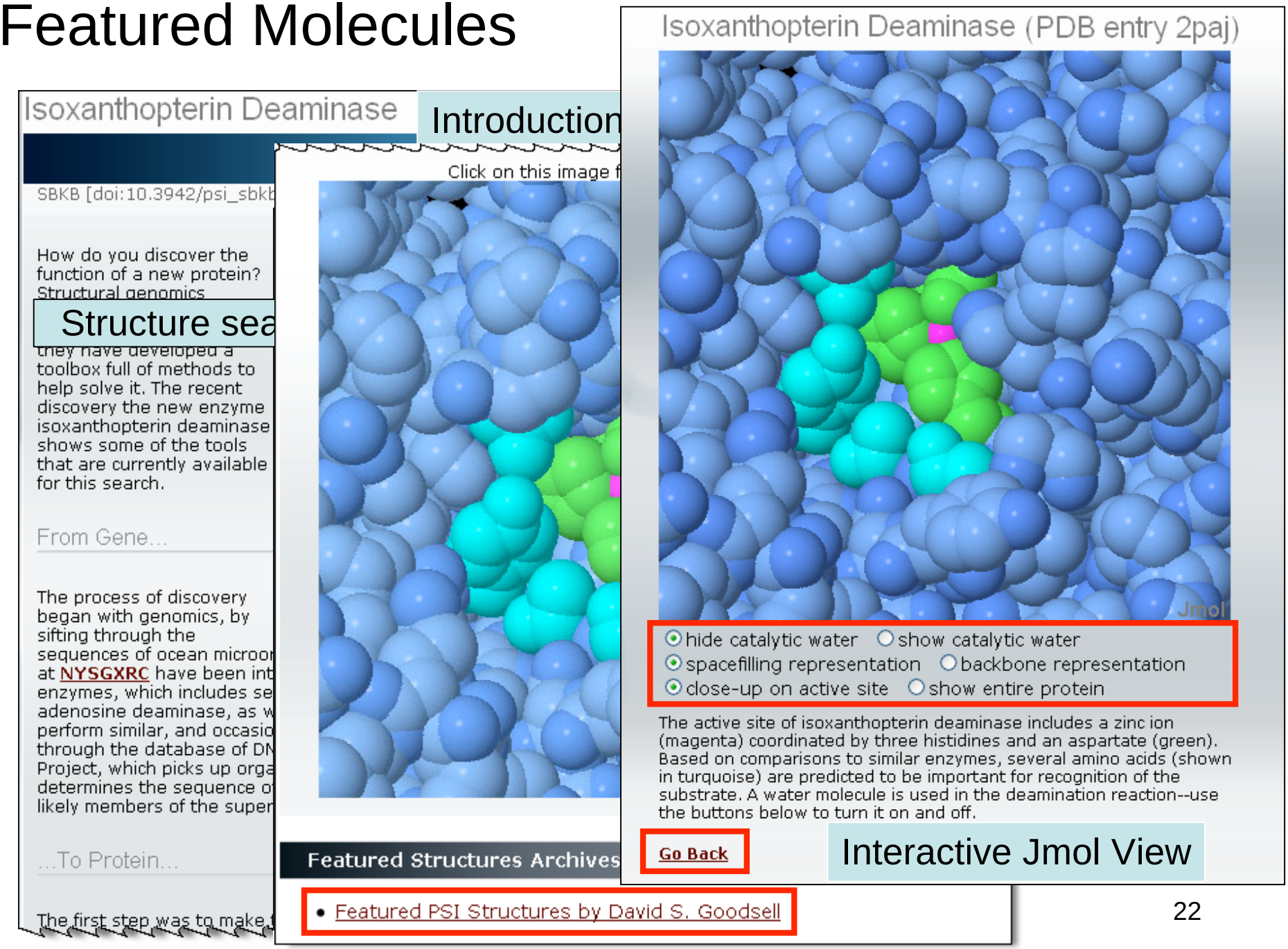
Image - **http://www.sbkb.org/about/getting_started.html**

Resource list - **http://www.sbkb.org/KB/seqstrucfunchub.html**

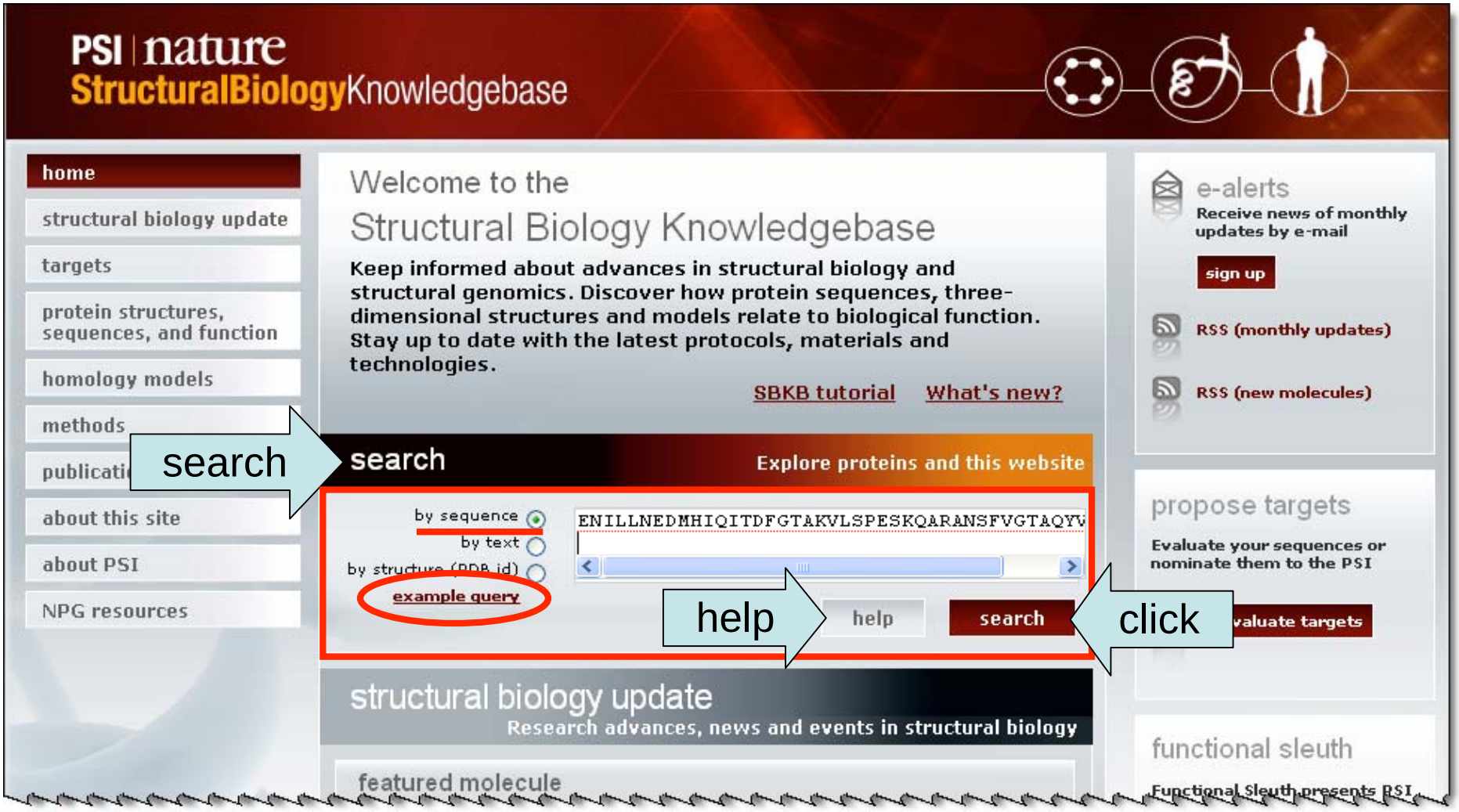
2. Structural Biology Update


PSI | nature
StructuralBiologyKnowledgebase
home
structural biology update
targets
protein structures, sequences, and function
homology models
methods
publications
about this site
about PSI
NPG resources
Welcome to the
Structural Biology Knowledg
Keep informed about advances in structura
structural genomics. Discover how protein
dimensional structures and models relate to biological
function. Stay up to date with the latest protocols, materials
and technologies.
SBKB tutorial
What's new?
search
Explore proteins and this website
by sequence
by text
by structure (PDB id)
example query
help
search
e-alerts
RSS (monthly updates)
RSS (new molecules)
propose targets
Evaluate your sequences or nominate them to the PSI
evaluate targets
functional sleuth
Functional Sleuth presents
latest PSI statistics
get test datasets
structural biology update
Research advances, news and events in structural biology
featured molecule
Isoxanthopterin deaminase
Researchers at NYSGXRC have disco
a new enzyme from the Global Ocean
research advances
COX inhibition: Naproxen by proxy
Structural basis of COX inhibition suggests that adverse
the way forward.
more
news
PSI:Biology in the spotlight
PSI:Biology awards, jobs and clones available, PSI activities in Acta Crystallographica F, and more...
Consortium solves its 1,000th protein structure
more
see all this month's content
research advances
news
featured molecule
research library
see all this month's content
structural biology update
Each month we offer a fresh collection of research, news and events in structural biology and structural genomics, produced by Nature Publishing Group in collaboration with the Protein Structure Initiative.
research advances
COX inhibition: Naproxen by proxy
Structural basis of COX inhibition suggests that adverse effects of NSAIDs will not be easily avoided.
ABA receptor diversity
There are 14 members in the ABA receptor family in Arabidopsis thaliana. How these receptors differ in their response to a synthetic, seed-specific ABA agonist is now explored in two independent papers, using a combination of crystallography, NMR and biochemical approaches.
Solving homodimeric structures with NMR
Complexes are very difficult to solve in solution but are important biologically. Combining multiple approaches is the way forward.
Scaling up mutational scanning
Mutating a protein's sequence is a useful way of uncovering functionally important residues. A large-scale method tracking up to 600,000 variants at once will speed up this analysis.
more this month
featured molecule
Isoxanthopterin deaminase
Researchers at NYSGXRC have discovered the function of a new enzyme from the Global Ocean Sampling Project.
more
news
PSI:Biology in the spotlight
Consortium s
1,000th protein structure
Clever coupling catalysts lauded by chemistry Nobel
Virus-like particles speed bacterial evolution
more
research library
Structural Basis for Substrate Placement by an Archaeal Box C/D Ribonucleoprotein Particle
Novel α-glucosidase from human gut microbiome: substrate specificities and their switch
Engineering of a wheat germ expression system to provide compatibility with a high throughput pET-based cloning platform
more
events
American Society
Human Genetic
60th Annual
November 2, 201
November 6, 2010
Washington, DC, USA
Advances in Protein Crystallography
November 9, 2010 -
November 10, 2010
Florence, Italy
more
research advances
featured molecule
news
events
research library

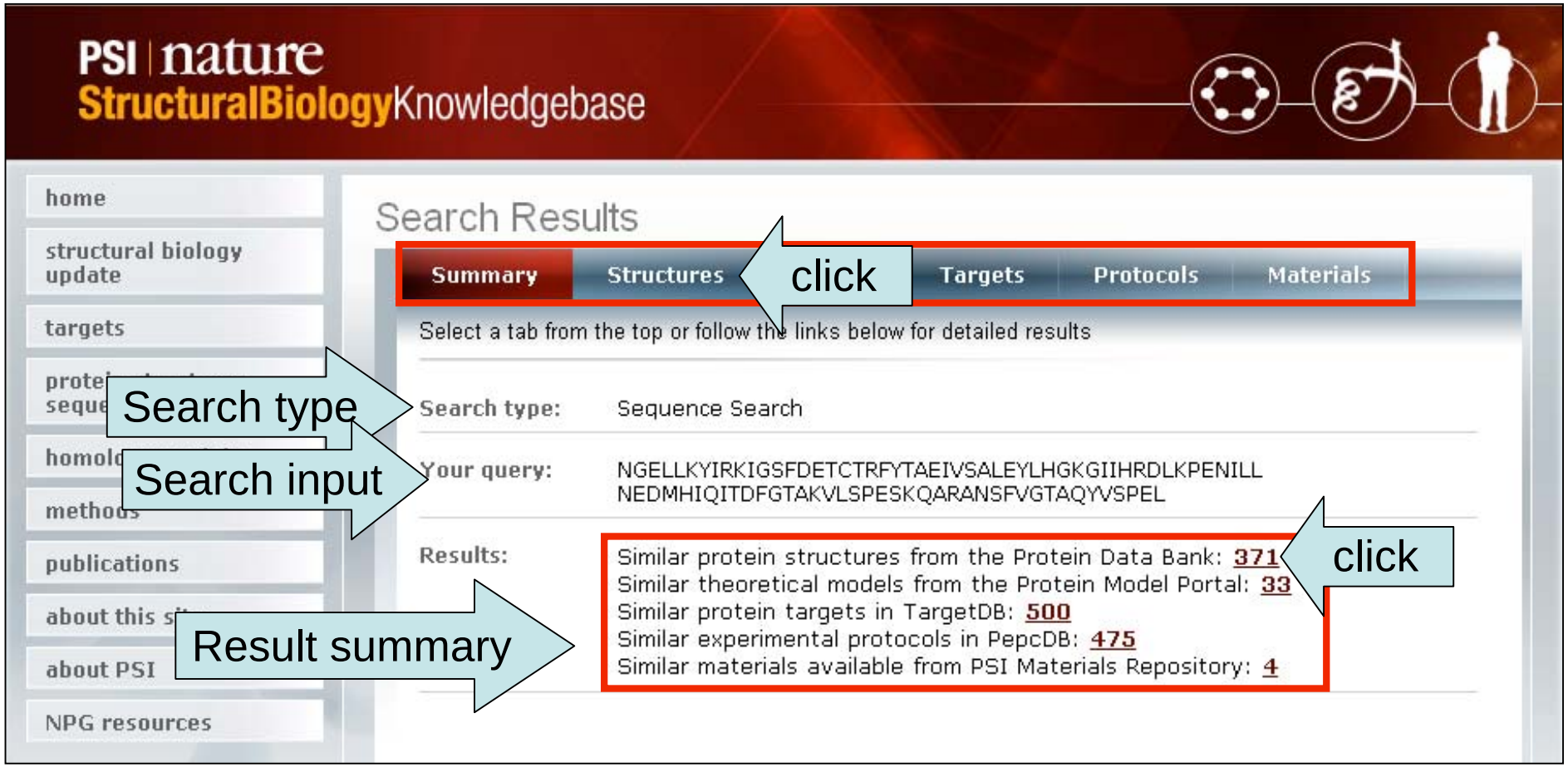
Featured Molecules



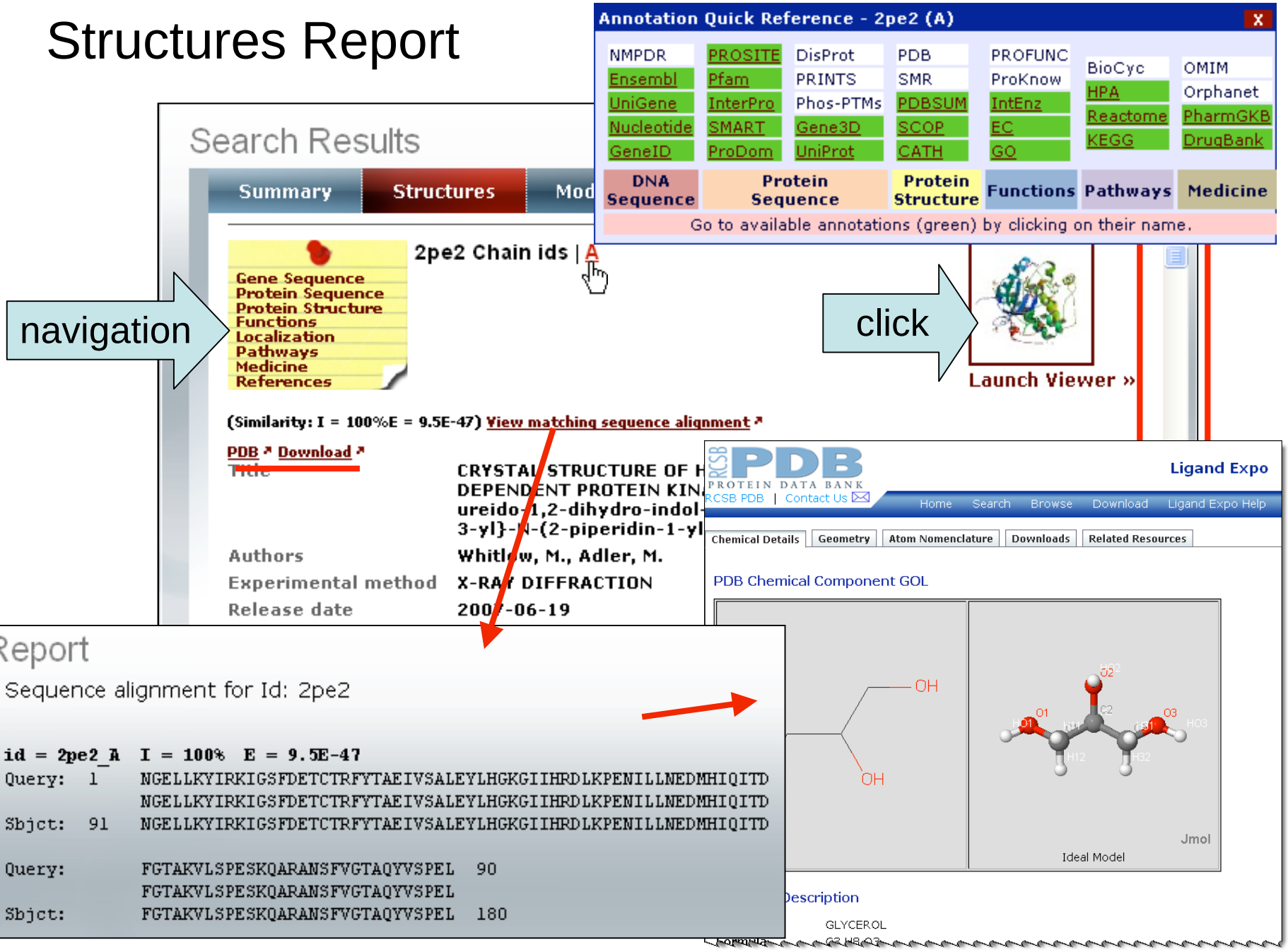
3. PSI SBKB Search Form



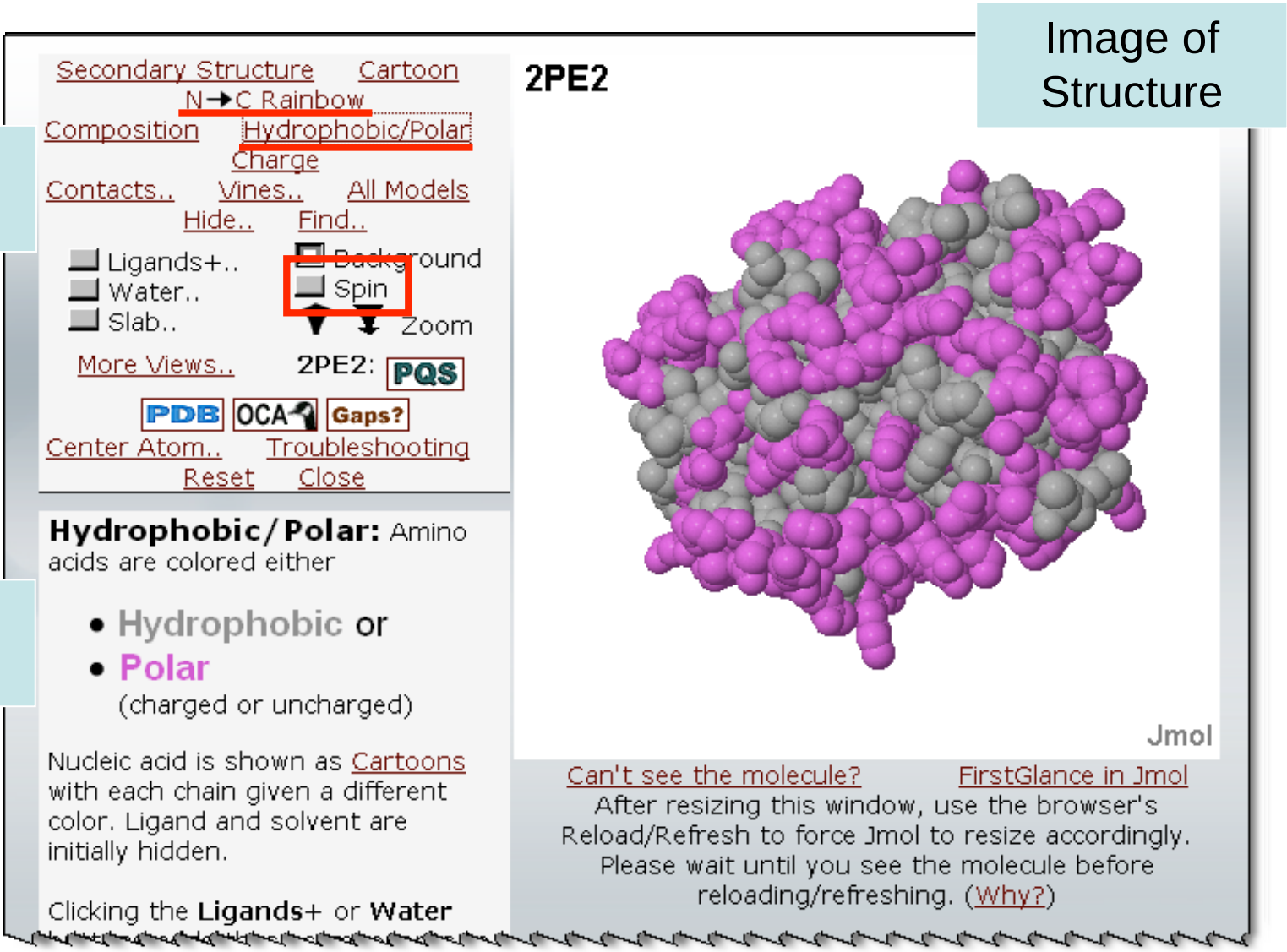
Sequence Search Results



Structures Report

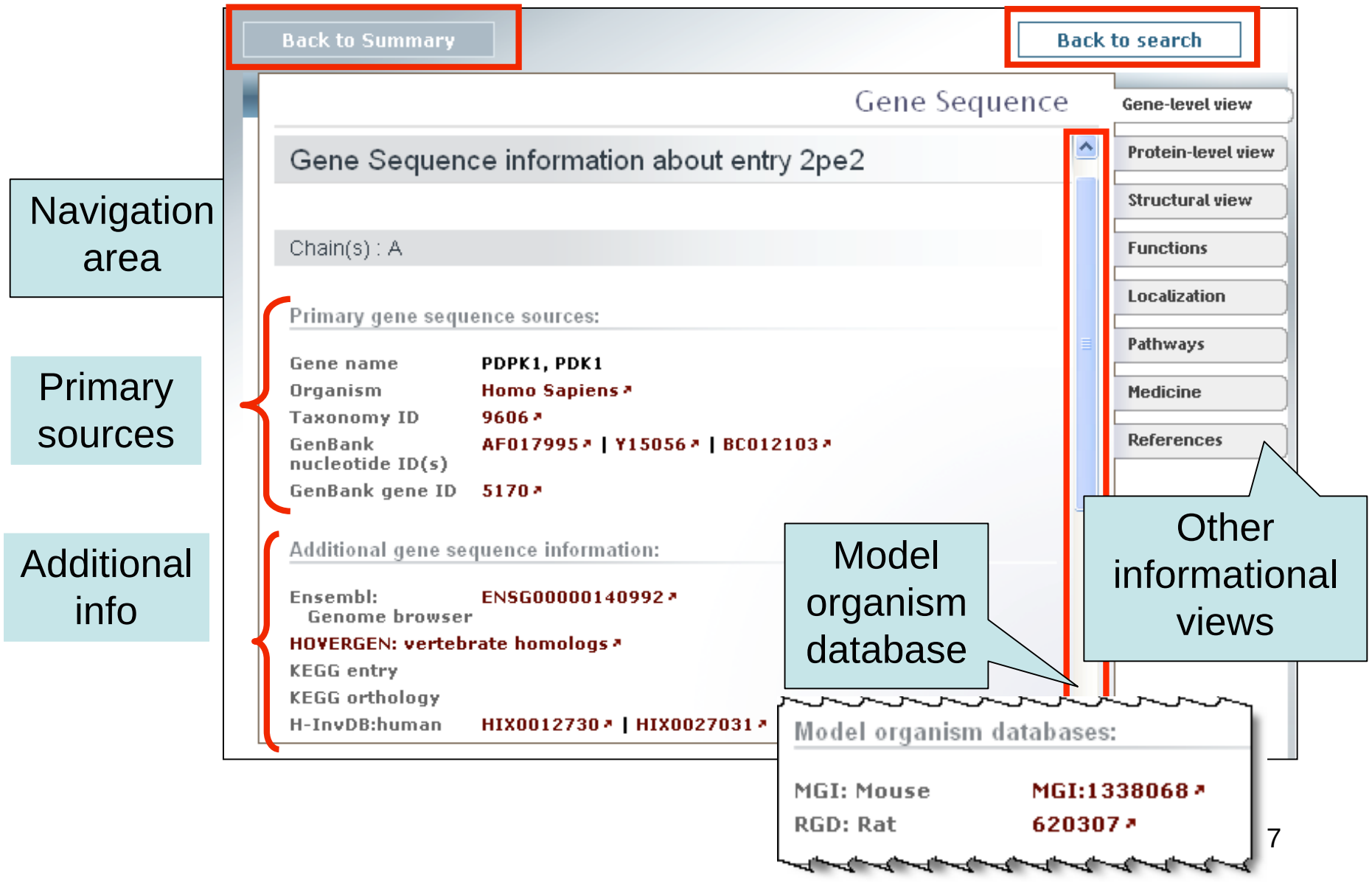


Structure Viewer

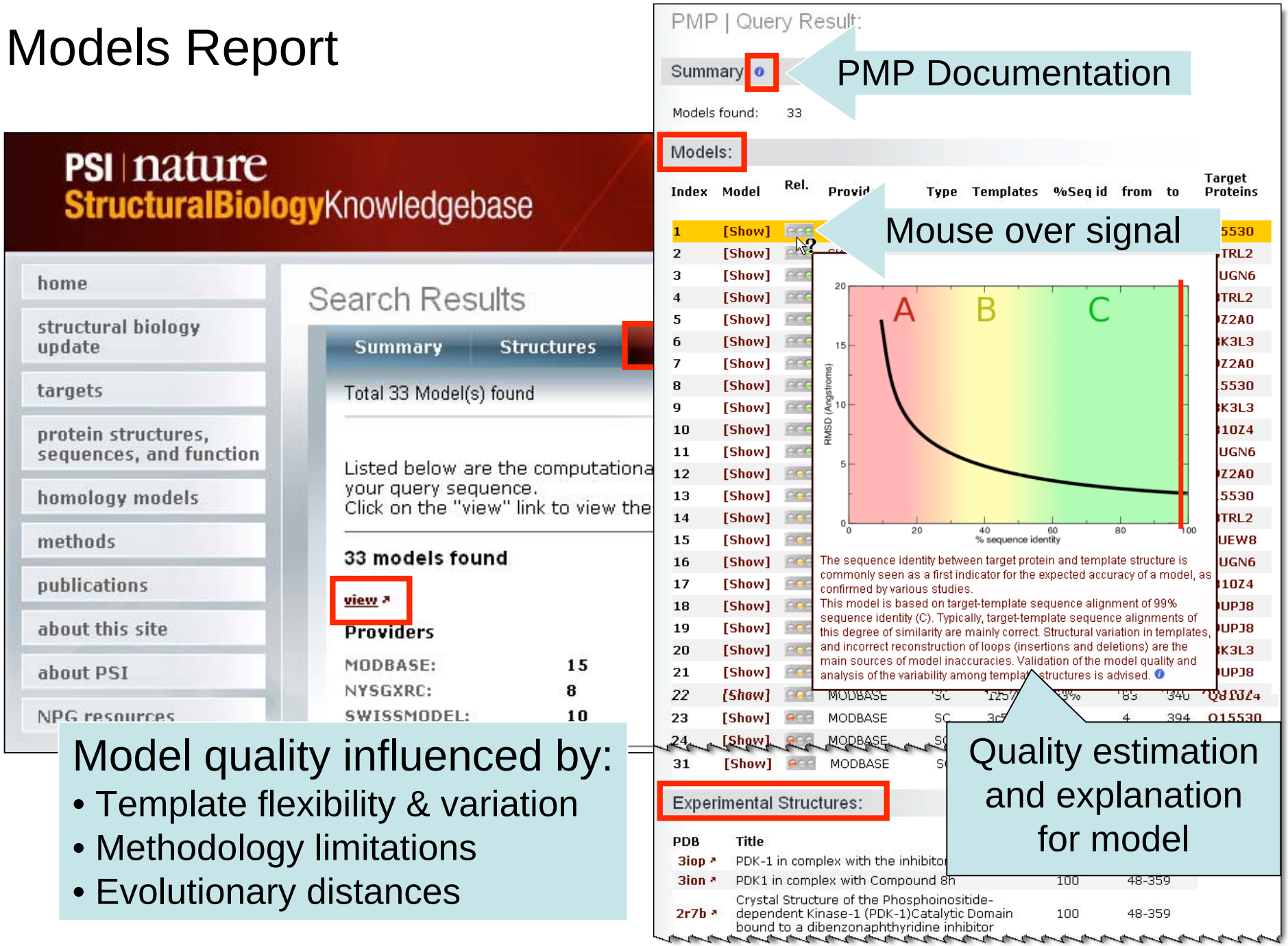


Many visualization options for exploring your protein!

Structure Annotation Reports, Gene-Level View



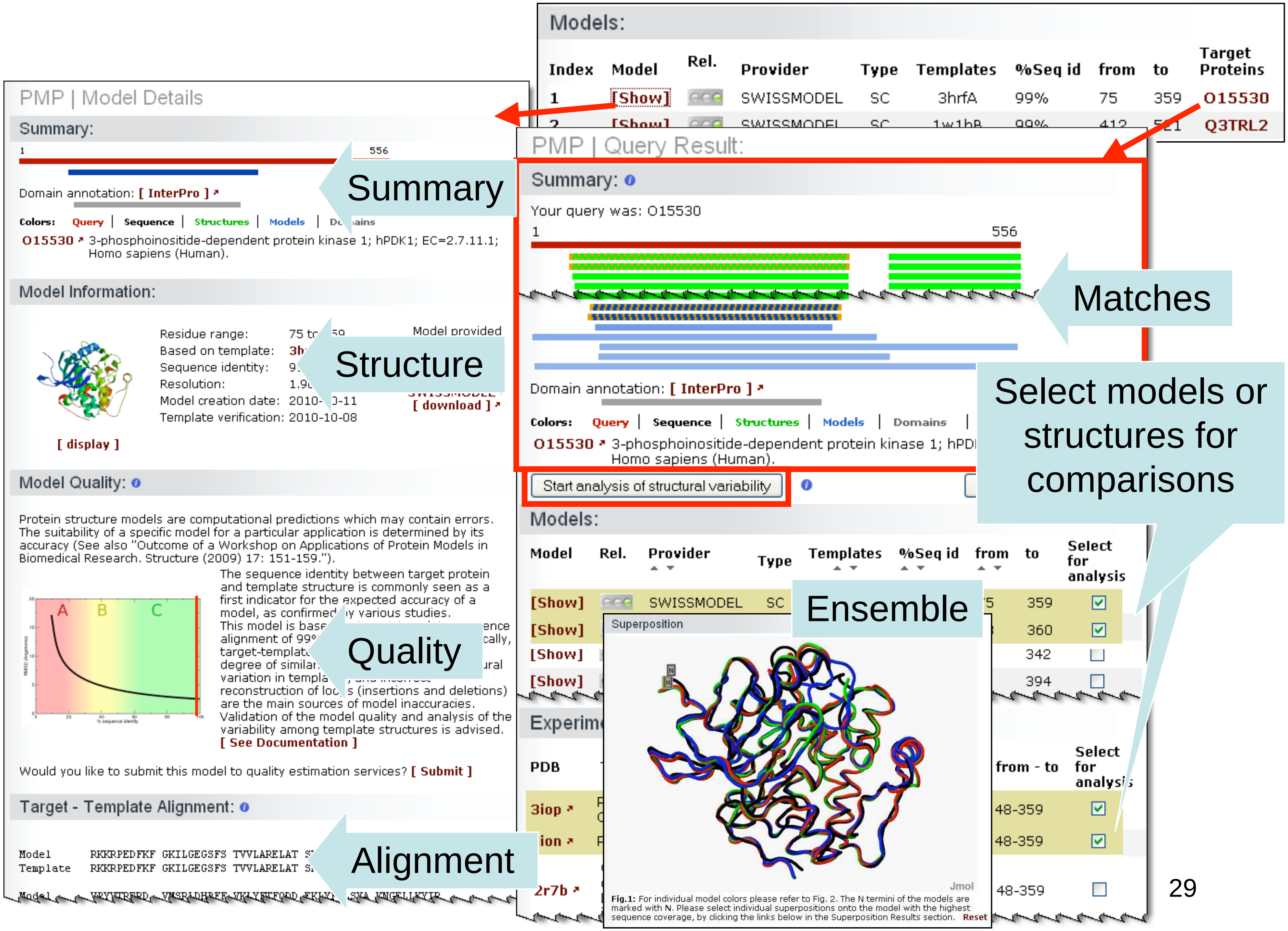
Models Report



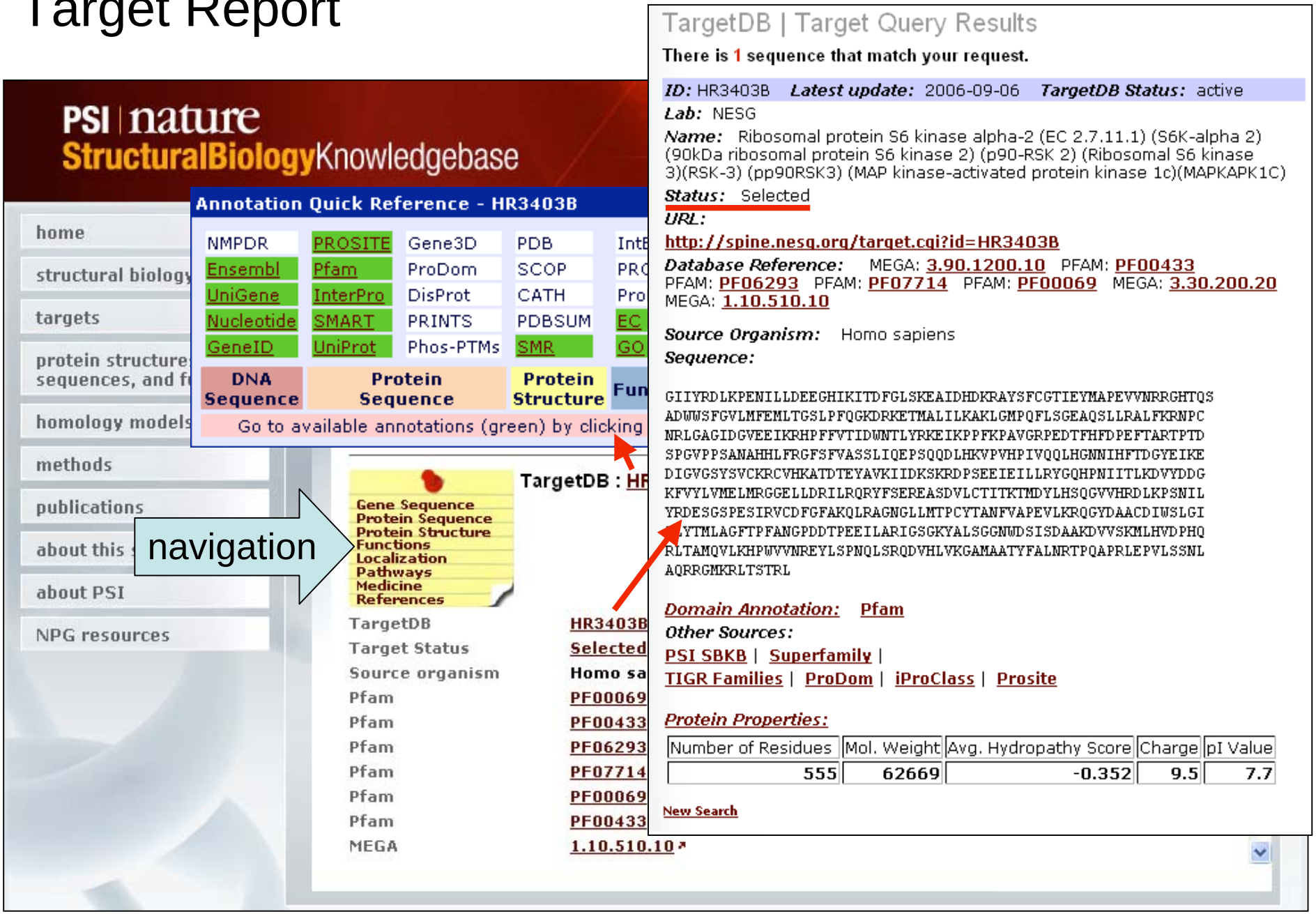
Model quality influenced by:

- Template flexibility & variation
- Methodology limitations
- Evolutionary distances

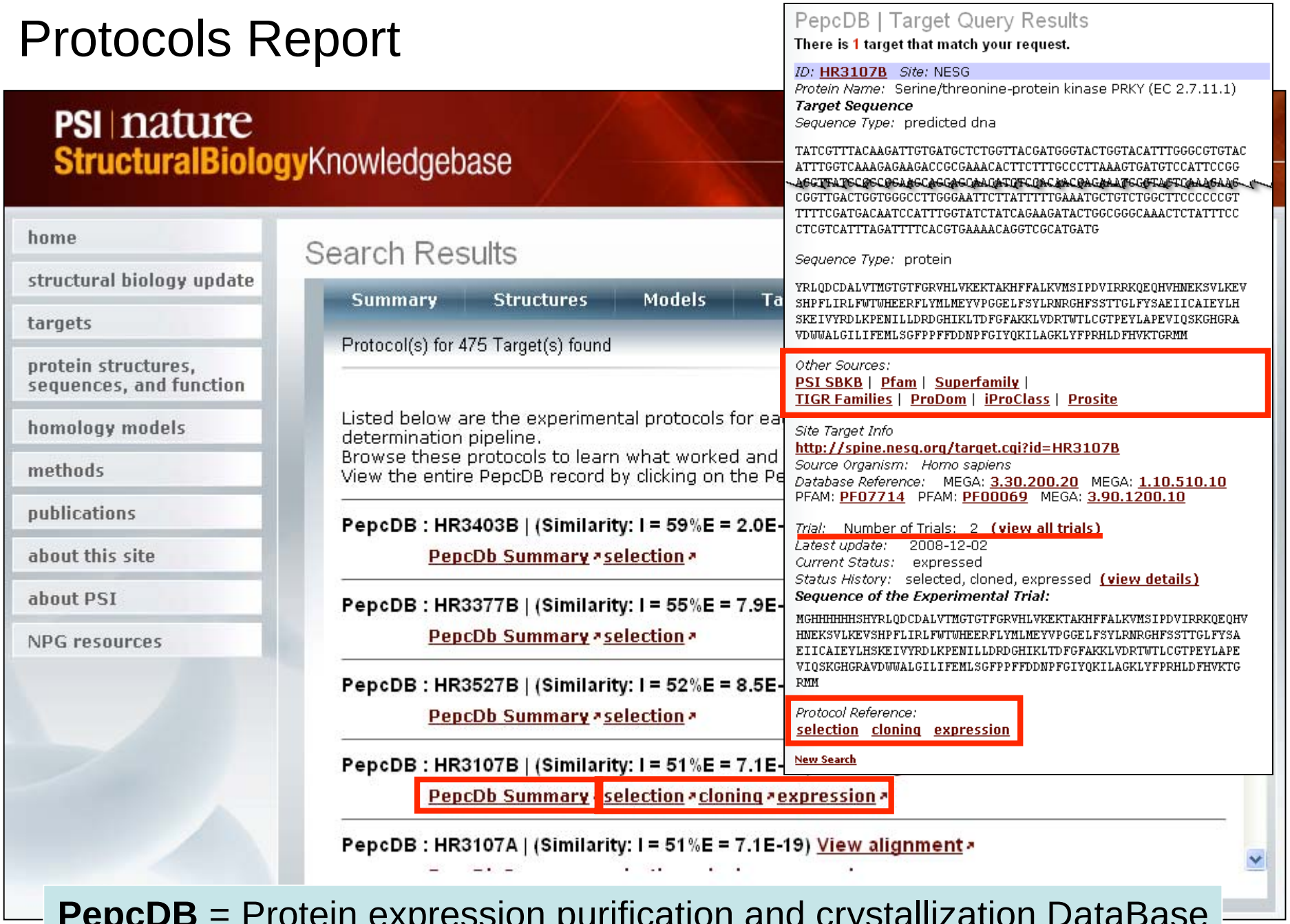
Quality estimation and explanation for model


Summary
Structure
Quality
Alignment
Matches
Ensemble
Select models or structures for comparisons
PMP | Model Details
Summary:
Model Information:
Model Quality:
Target - Template Alignment:
PMP | Query Result:
Your query was: O15530
Start analysis of structural variability
Models:
Superposition

Target Report

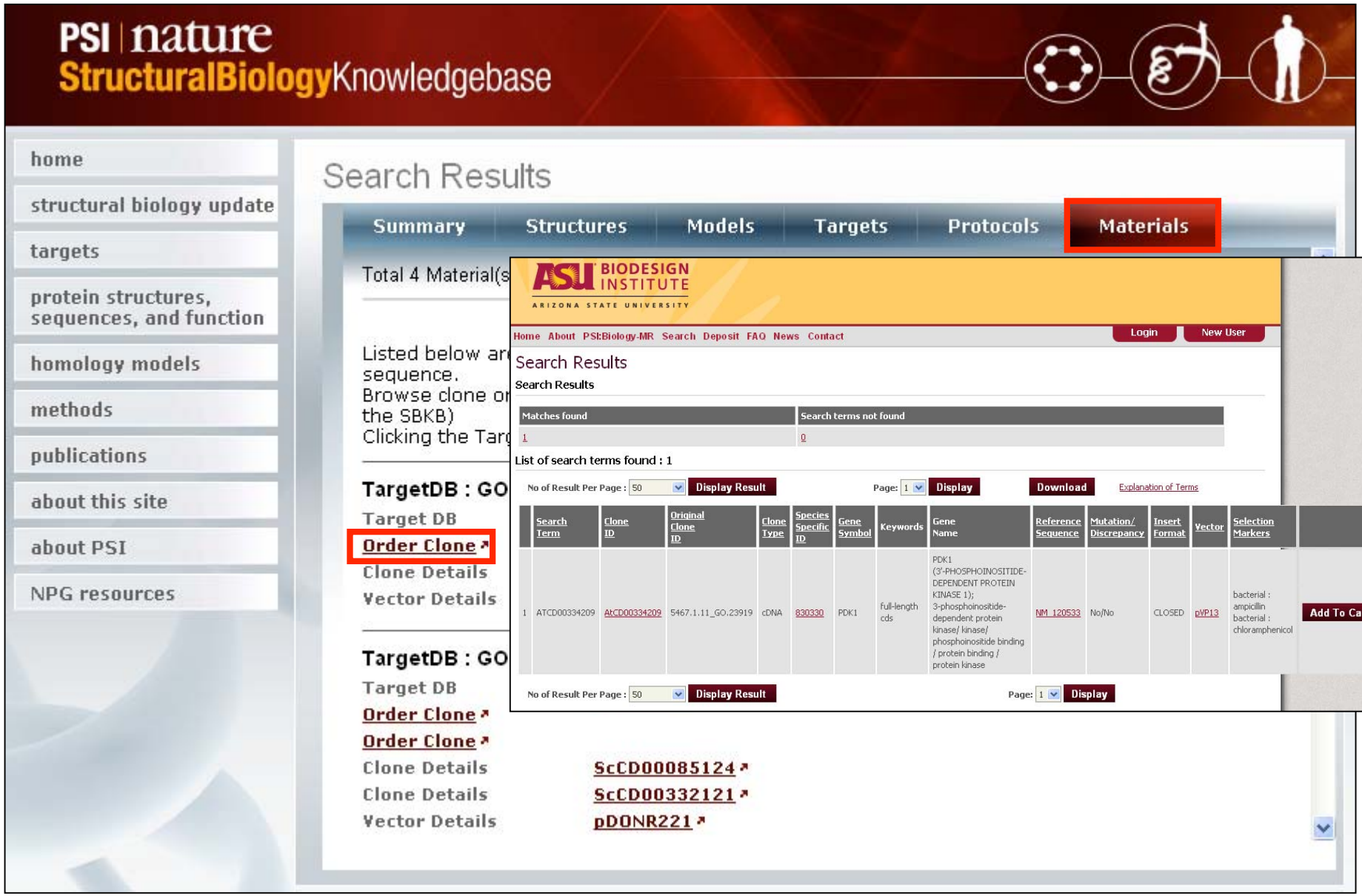


Protocols Report

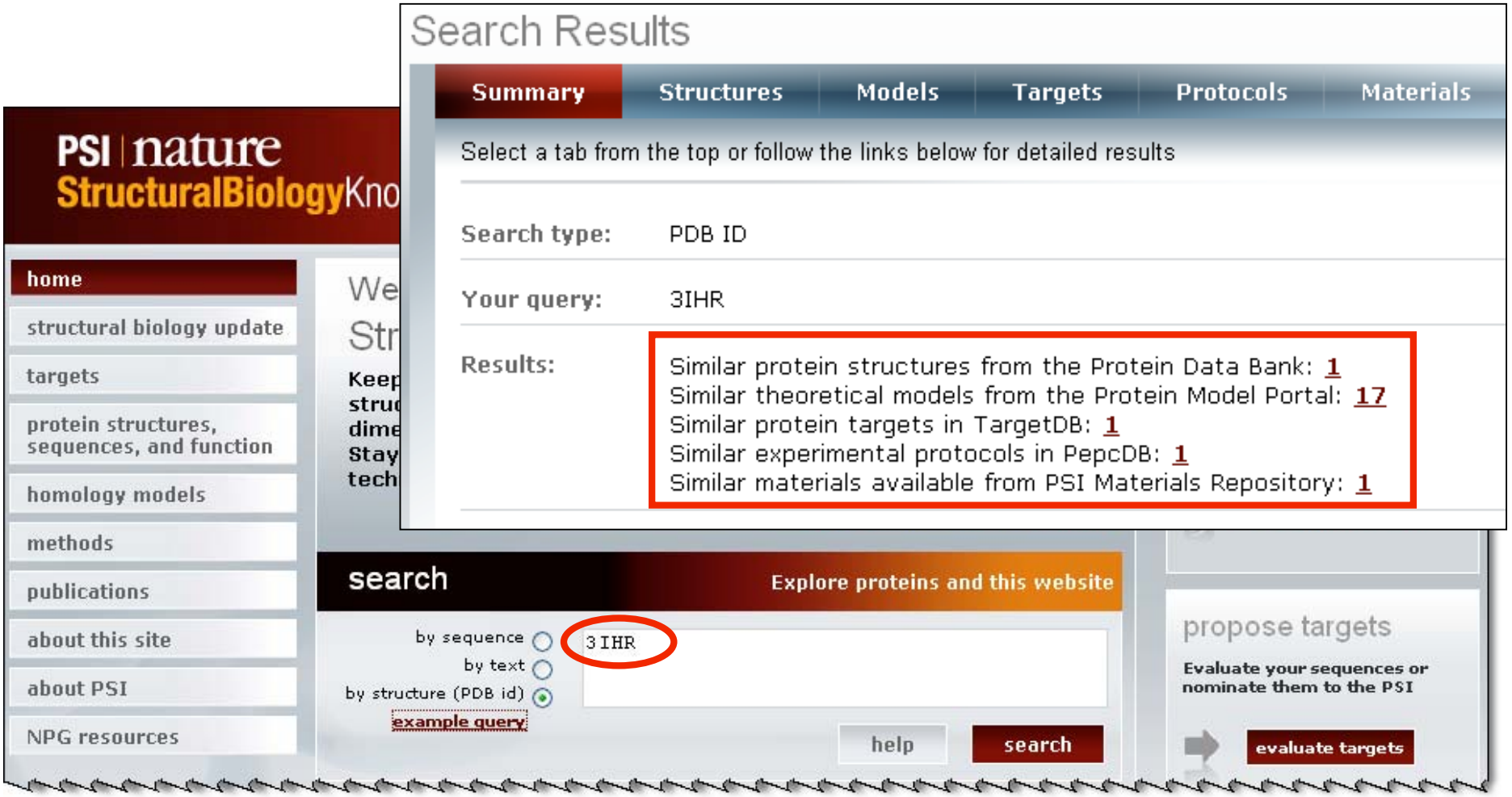


PepcDB = Protein expression purification and crystallization DataBase

Materials Report

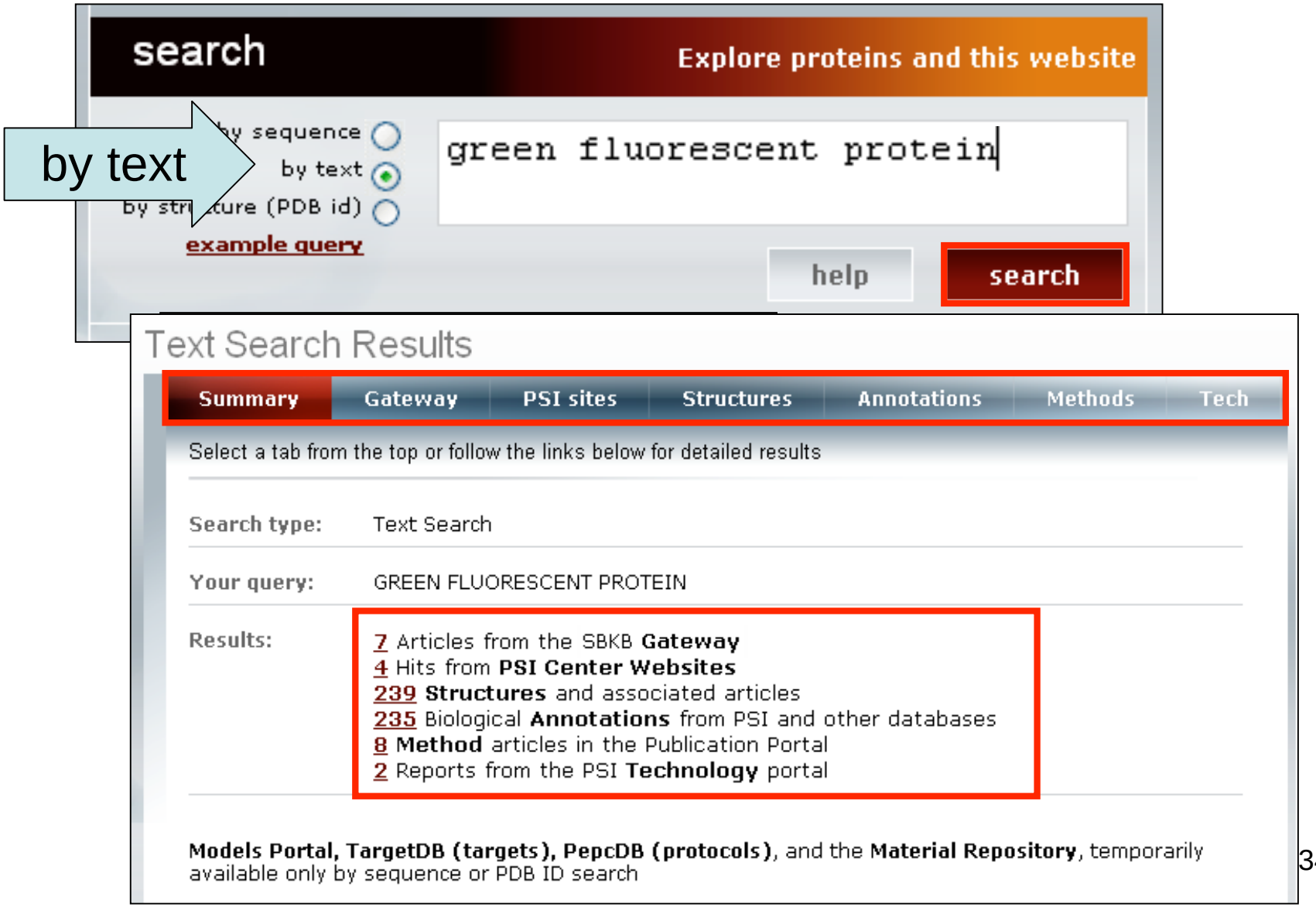


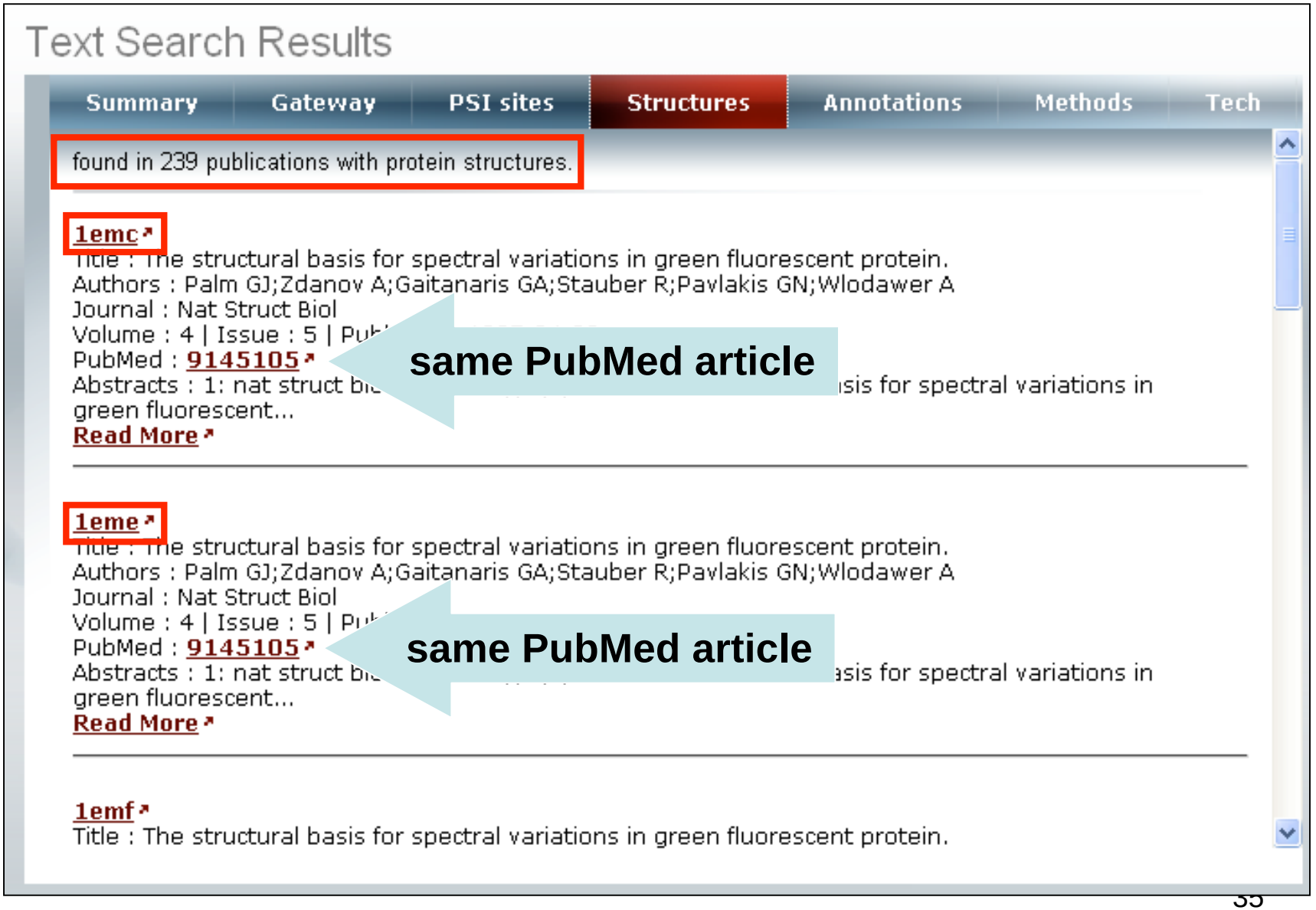
Structure (PDB ID) Search



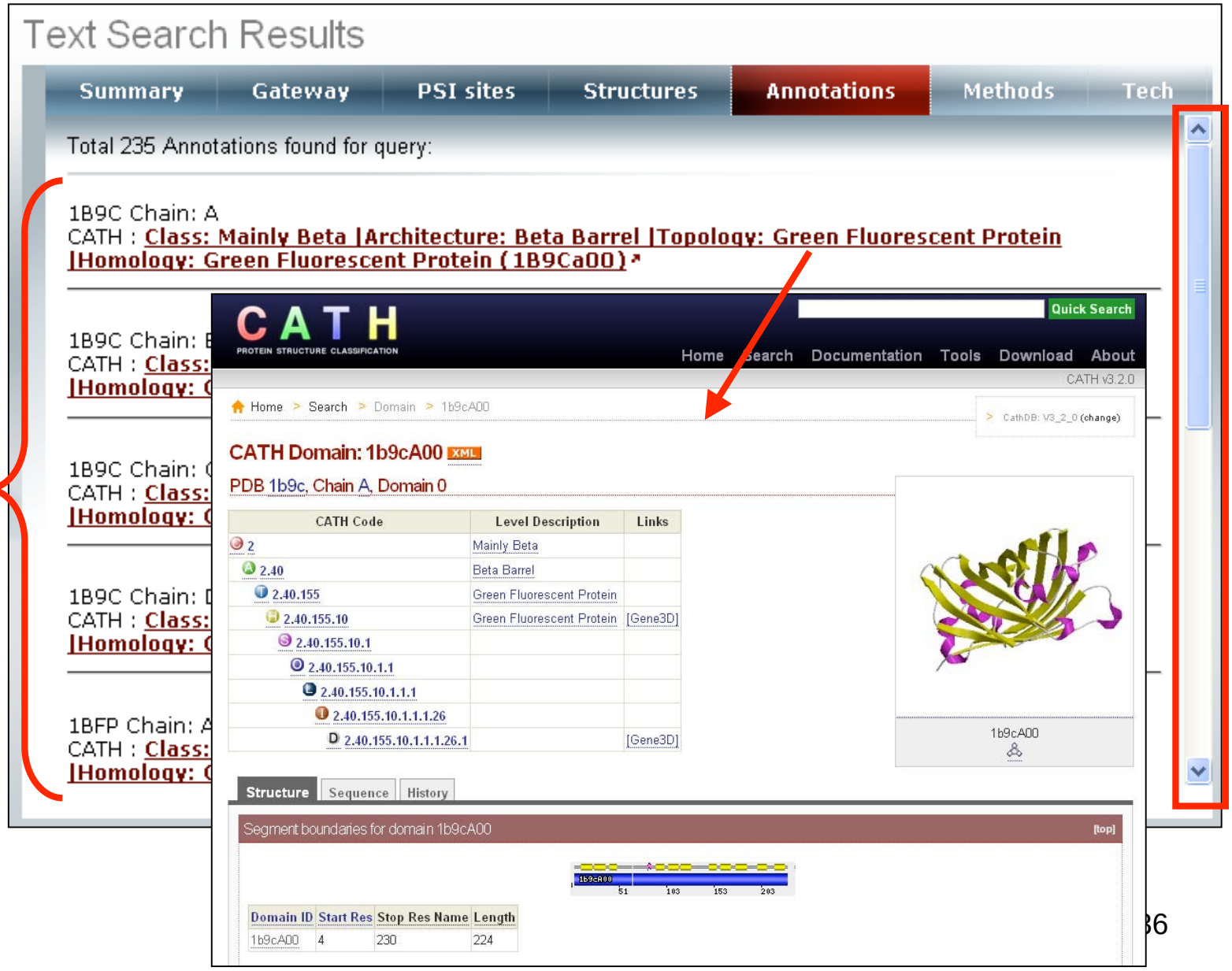
PDB ID = very specific, only single result!

4. Text Searches

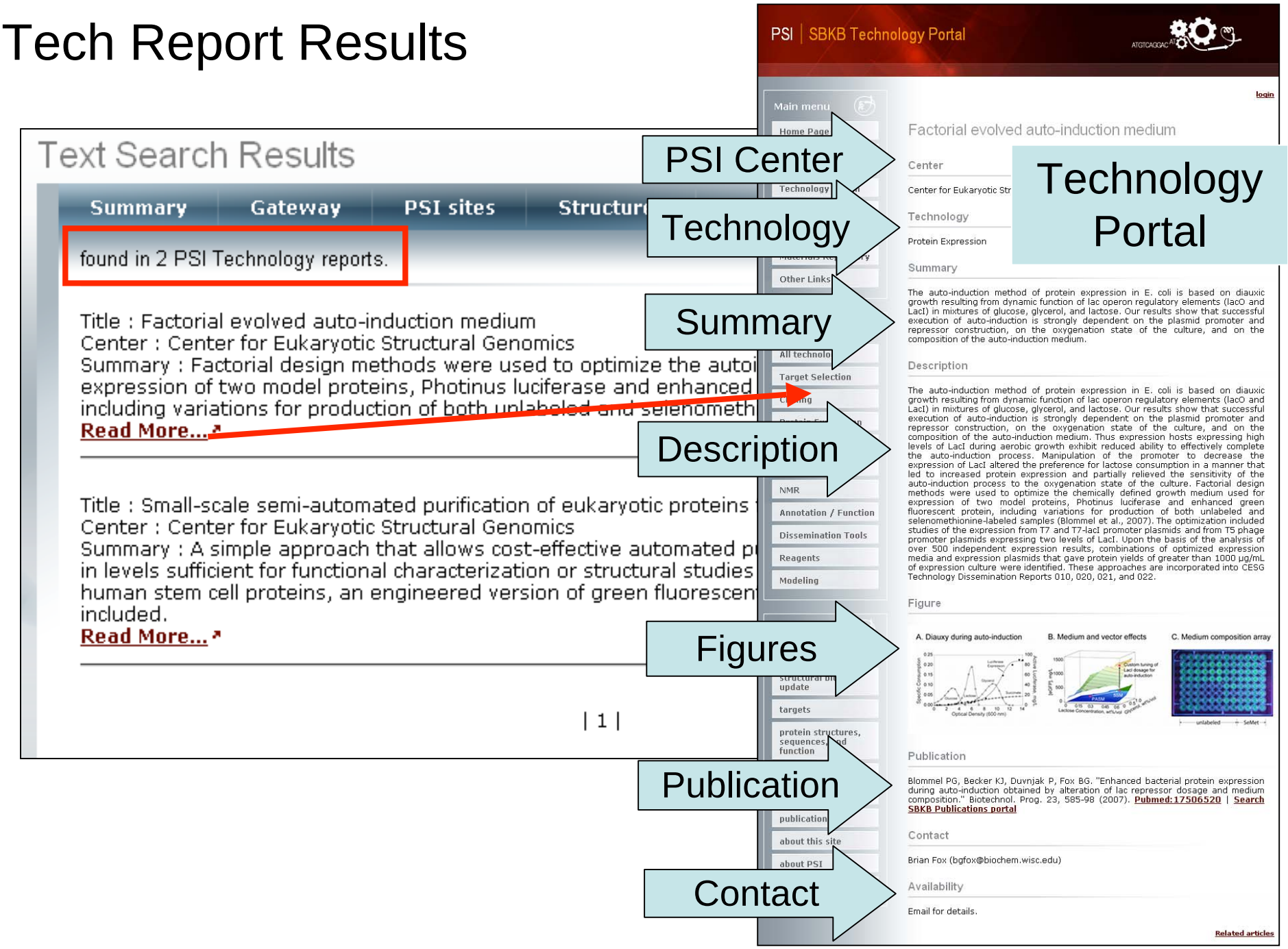



Text Search Results
Summary
Gateway
PSI sites
Structures
Annotations
Methods
Tech
found in 239 publications with protein structures.
1emc
Title : The structural basis for spectral variations in green fluorescent protein.
Authors : Palm GJ;Zdanov A;Gaitanaris GA;Stauber R;Pavlakis GN;Wlodawer A
Journal : Nat Struct Biol
Volume : 4 | Issue : 5 |
PubMed : 9145105
Abstracts : 1: nat struct
sis for spectral variations in
green fluorescent...
Read More
same PubMed article
1eme
Title : The structural basis for spectral variations in green fluorescent protein.
Authors : Palm GJ;Zdanov A;Gaitanaris GA;Stauber R;Pavlakis GN;Wlodawer A
Journal : Nat Struct Biol
Volume : 4 | Issue : 5 |
PubMed : 9145105
Abstracts : 1: nat struct
asis for spectral variations in
green fluorescent...
Read More
same PubMed article
1emf
Title : The structural basis for spectral variations in green fluorescent protein.

CATH results



Tech Report Results



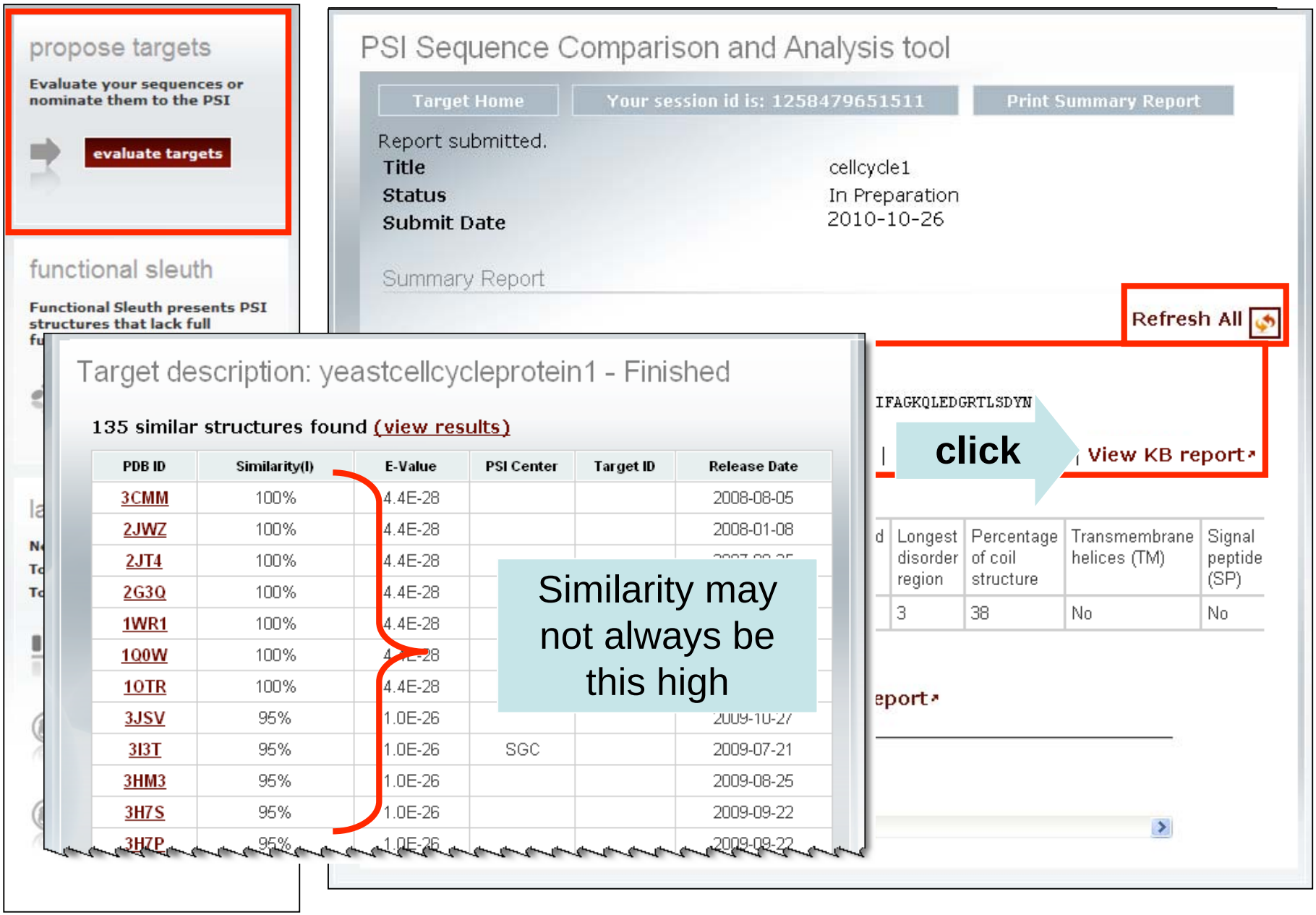
5. Additional Features

sources of additional information

evaluate or propose your amino acid sequences

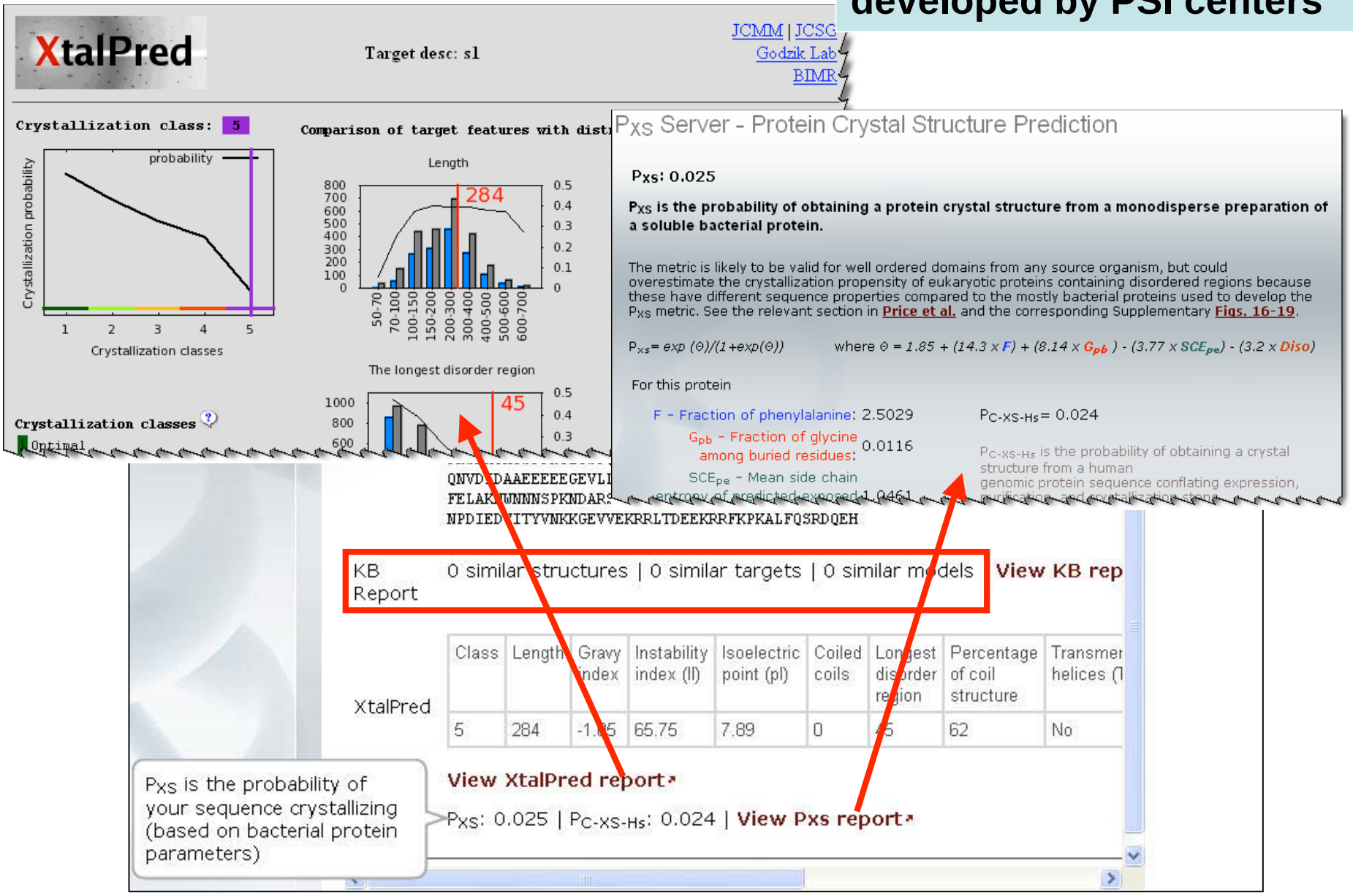
help characterize novel protein sequences

see stats on PSI progress


propose targets
Evaluate your sequences or nominate them to the PSI
evaluate targets
functional sleuth
Functional Sleuth presents PSI structures that lack full
PSI Sequence Comparison and Analysis tool
Target Home
Your session id is: 1258479651511
Print Summary Report
Report submitted.
Title
cellcycle1
Status
In Preparation
Submit Date
2010-10-26
Summary Report
Refresh All
Target description: yeastcellcycleprotein1 - Finished
135 similar structures found (view results)
PDB ID
Similarity(I)
E-Value
PSI Center
Target ID
Release Date
3CMM 100% 4.4E-28 2008-08-05
2JWZ 100% 4.4E-28 2008-01-08
2JT4 100% 4.4E-28
2G3Q 100% 4.4E-28
1WR1 100% 4.4E-28
1Q0W 100%
1OTR 100% 4.4E-28
3JSV 95% 1.0E-26
3I3T 95% 1.0E-26 SGC 2009-07-21
3HM3 95% 1.0E-26 2009-08-25
3H7S 95% 1.0E-26 2009-09-22
3H7P 95% 1.0E-26 2009-09-22
Similarity may not always be this high
IFAGKQLEDGRTLSDYN
click
View KB report
Longest disorder region
Percentage of coil structure
Transmembrane helices (TM)
Signal peptide (SP)
3
38
No
No
eport

Proposed Targets, Example 2

XtalPred & Pxs: Two Sequence Analysis Tools developed by PSI centers



propose targets

Evaluate your sequences or nominate them to the PSI

evaluate targets

functional sleuth

Functional Sleuth presents PSI structures that lack full functional annotation

begin exploring

latest PSI statistics

New structures last month: 42
Total structures to date: 4378
Total distinct structures: 3714

more

see latest structures

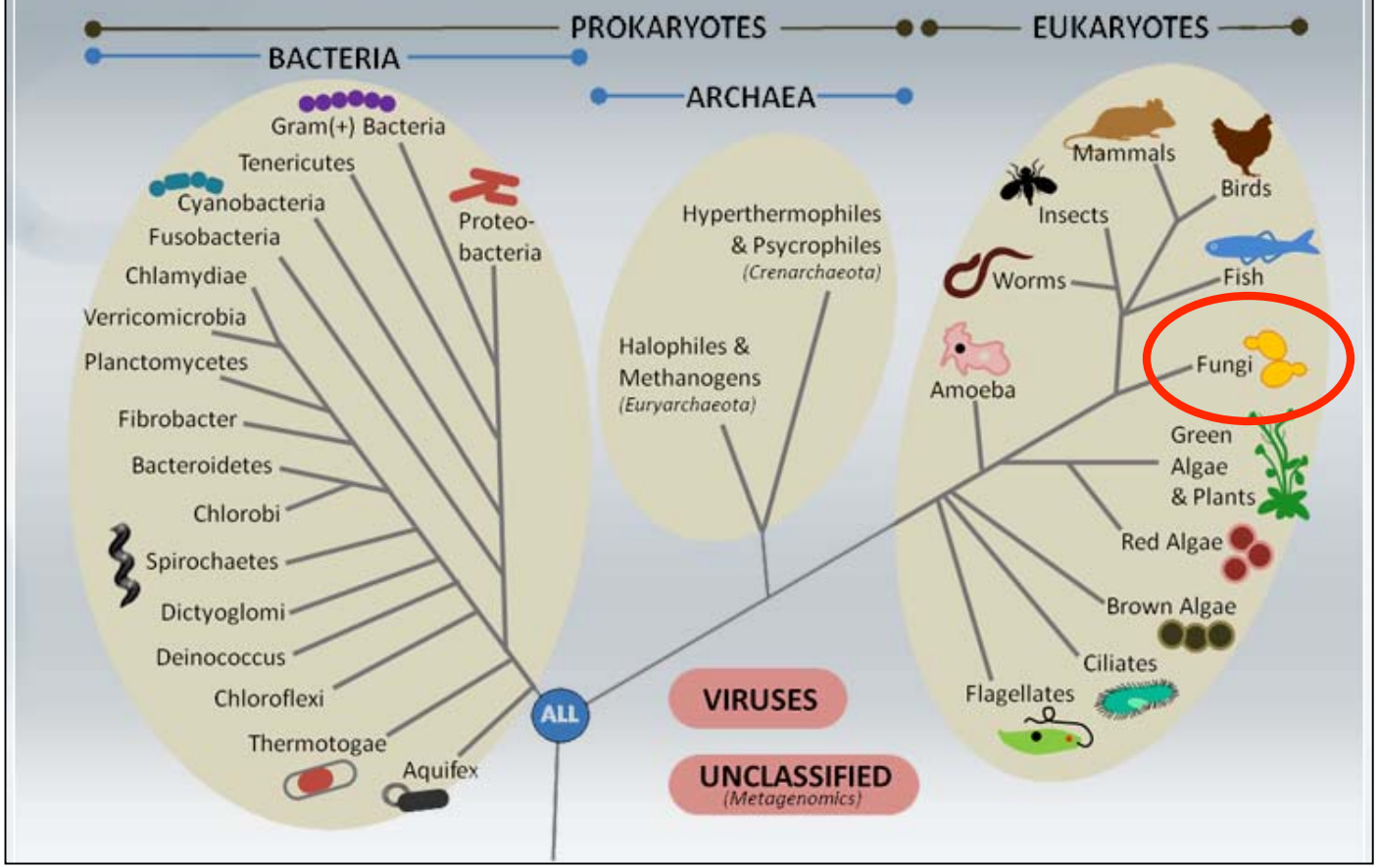
get test datasets

Functional Sleuth

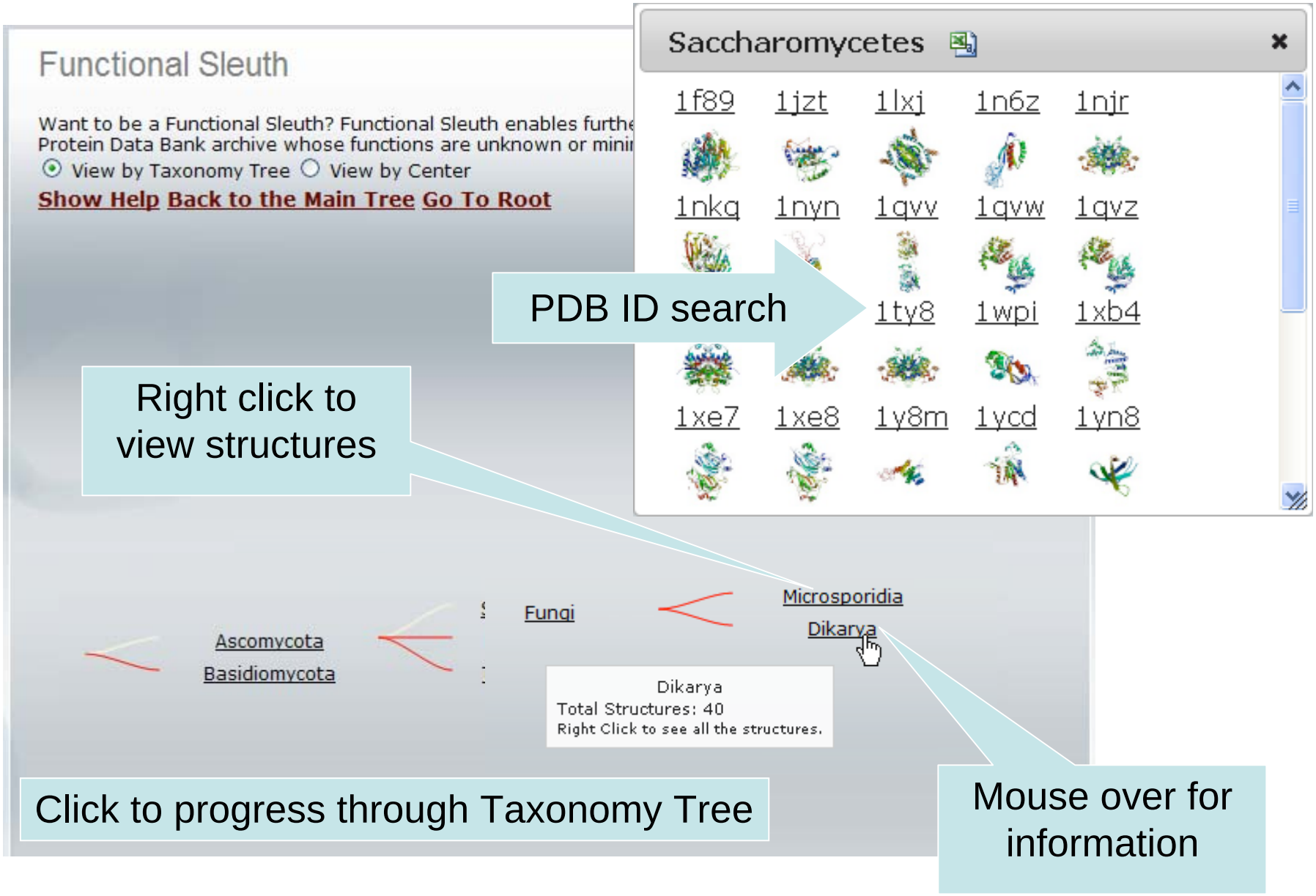
Want to be a Functional Sleuth? Functional Sleuth enables further research for proteins in the Protein Data Bank archive whose functions are unknown or minimally characterized.

(•) View by Taxonomy Tree ( ) View by Center

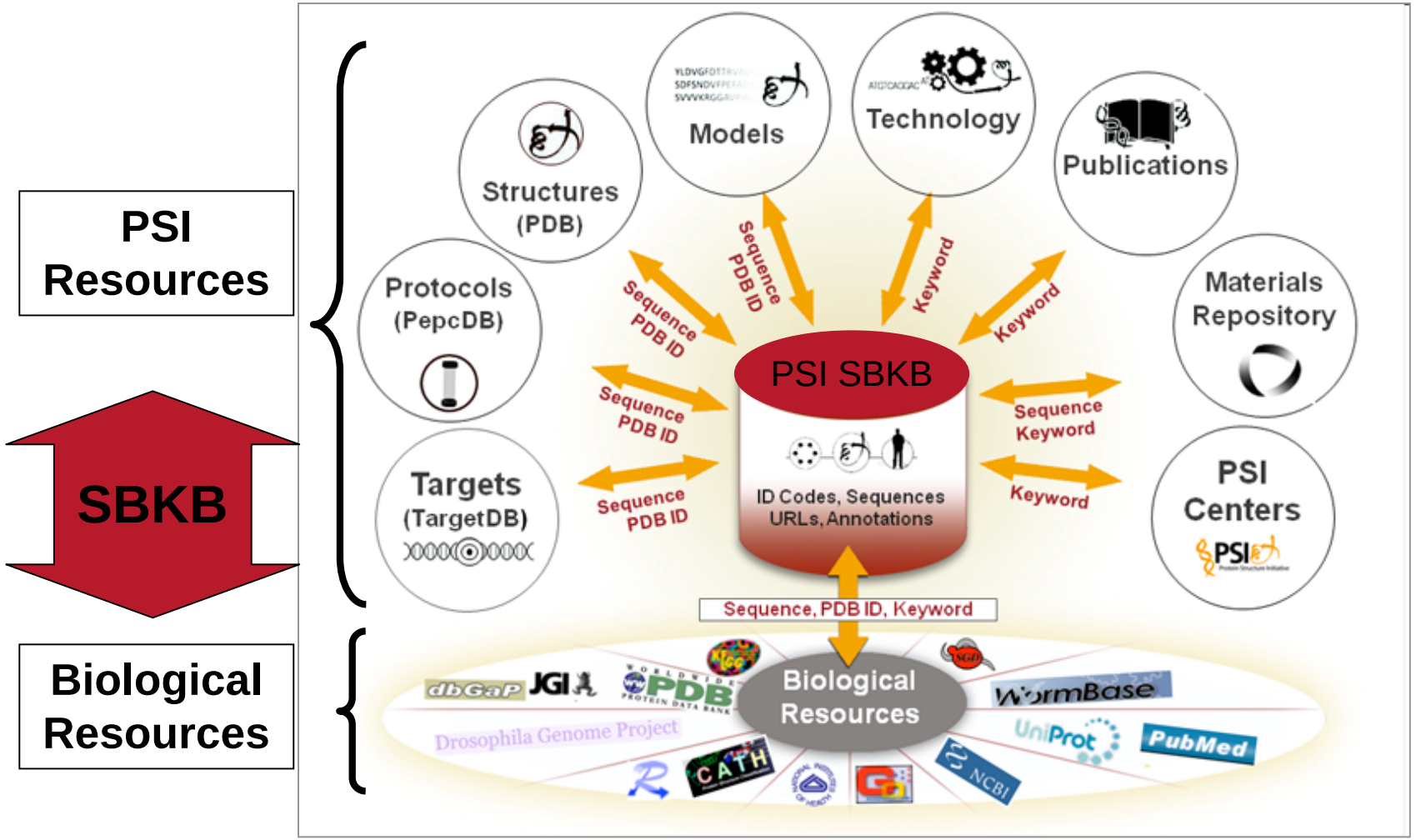
These "structures of unknown function" (SUFs) are organized by source organism. Select a domain/superkingdom or group to begin; or jump to popular phyla or classes using our tree of life. Making a selection will launch our interactive tree browser. You can also **download a text file** of all SUF PDB IDs, updated weekly. Only organisms with structures are displayed.



Functional Sleuth



6. Summary



SBKB: a Bridge between Biological and PSI Structural Information

7. Exercises

1. You recently read an interesting paper that mentioned the PDB structure xxx. Search the PSI SBKB to learn more about this structure.

2. Imagine that you've joined your first rotation lab and have been offered the project of characterizing the protein product from domestic pig given below. You decide to begin your analysis of the sequence at the PSI SBKB to determine anything you can about this protein sequence. Are there any structures with similarity to your sequence? Are there any targets in other species with greater than 90% similarity?

```
>gi|259420073|emb|CBF63208.1| unnamed protein product [Sus scrofa domestica]
MDPETCPCPTGGSCTCAGSCKCEGCKCTSCKKSCCSCCPAECEKCAKDCVCKGGEGAEAEEEKCSCCQ
```

3. You are working on a project trying to crystallize a membrane protein, but you are having trouble with your protocols. What helpful information can you find from the SBKB Research Library as well as the rest of the PSI SBKB?