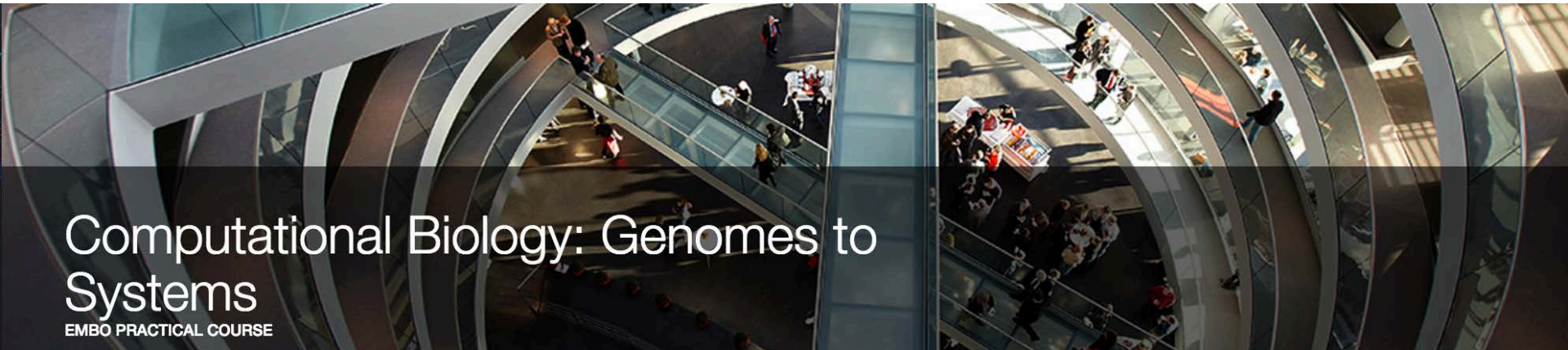
Using protein structures to understand protein function

Rebecca Wade



Heidelberg Institute for Theoretical Studies

HITS

Using protein structures to understand protein function

Rebecca C. Wade

Molecular & Cellular Modeling Group,

Heidelberg Institute for Theoretical Studies (HITS)

&

Zentrum für Molekulare Biologie Heidelberg (ZMBH)
Heidelberg University

rebecca.wade@h-its.org

http://www.h-its.org/mcm/

Heidelberg Institute for Theoretical Studies

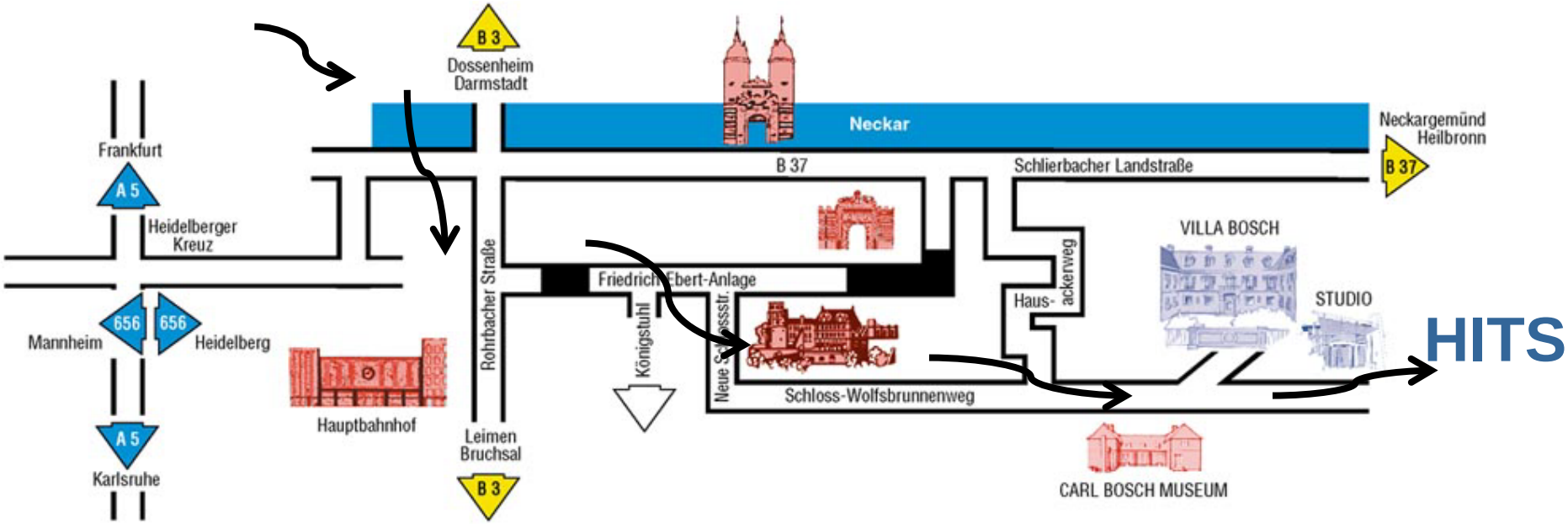
Heidelberg Institute for Theoretical Studies

HITS

Multidisciplinary computational sciences
Bioinformatics to computational linguistics to astrophysics to …..

Klaus Tschira Foundation

KTF



Heidelberg Institute for Theoretical Studies

Multidisciplinary computational sciences:

- Molecular modeling and simulation ; Bioinformatics & databases ; Computational linguistics ; Theoretical astrophysics ; …

Klaus Tschira Stiftung:

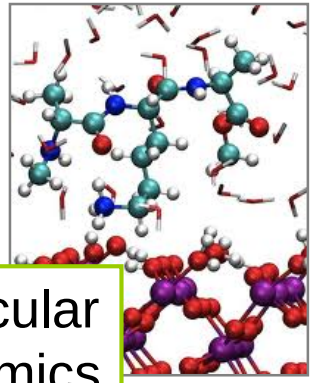
Supports informatics, natural sciences, mathematics

Klaus Tschira (1940-2015)

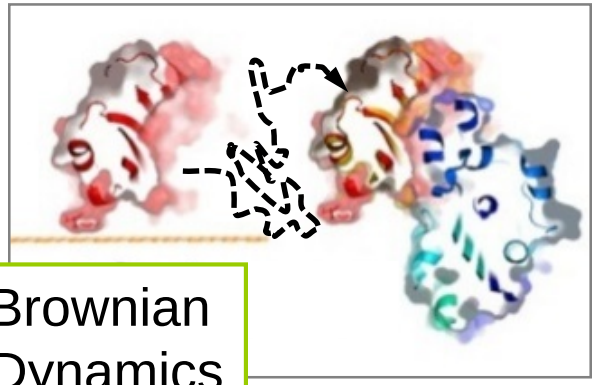
1972 - Co-founded the German software giant SAP AG

1995 – Bought **Villa Bosch**, founded Klaus Tschira Stiftung

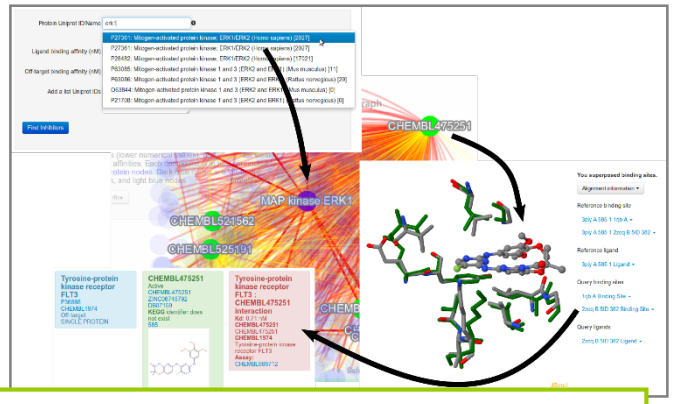
Computing Protein Interactions: Methods ↔ Applications



Molecular Dynamics

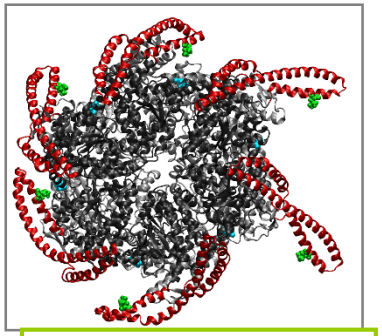


Brownian Dynamics

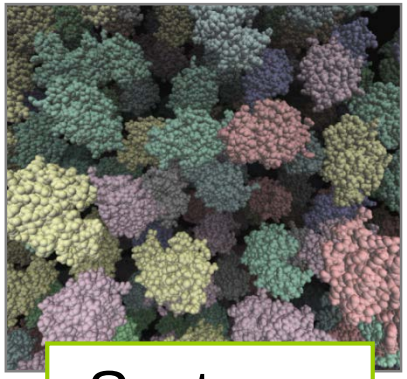


Bio/Chemo-informatics

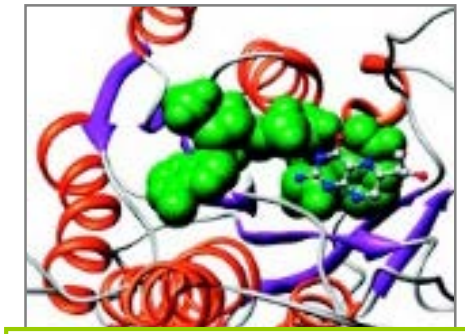
Multiscale Molecular Simulation



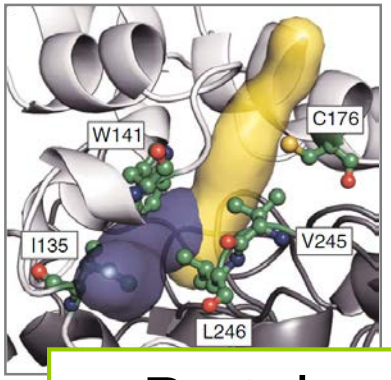
Molecular Biology



Systems Biology



Structure-based Drug Design



Protein Engineering

Using protein structures to learn about protein function: Learning objectives

- Protein structure and function
- Modeling protein structure and dynamics
- Computing interaction properties

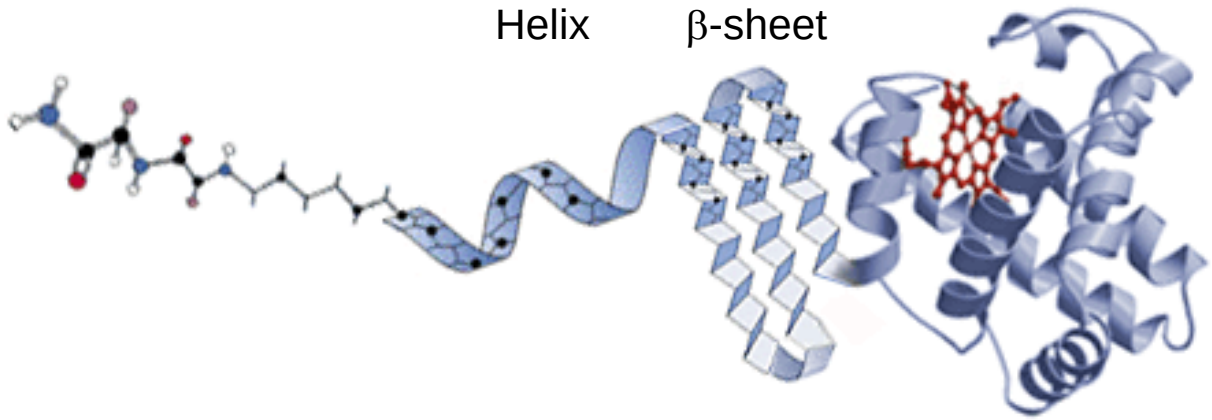
Proteins:

DNA: gatccagctg taccattatg taatataata agacacggac gcac……...

PROTEIN: MFKPVDFSETSPVPPDIDLAPTQSPHHVAPSQDSSYDLLS……
……SMLKNKSFLLHGKDYPNNADNNDNEDIRAKTMNRSQSHV

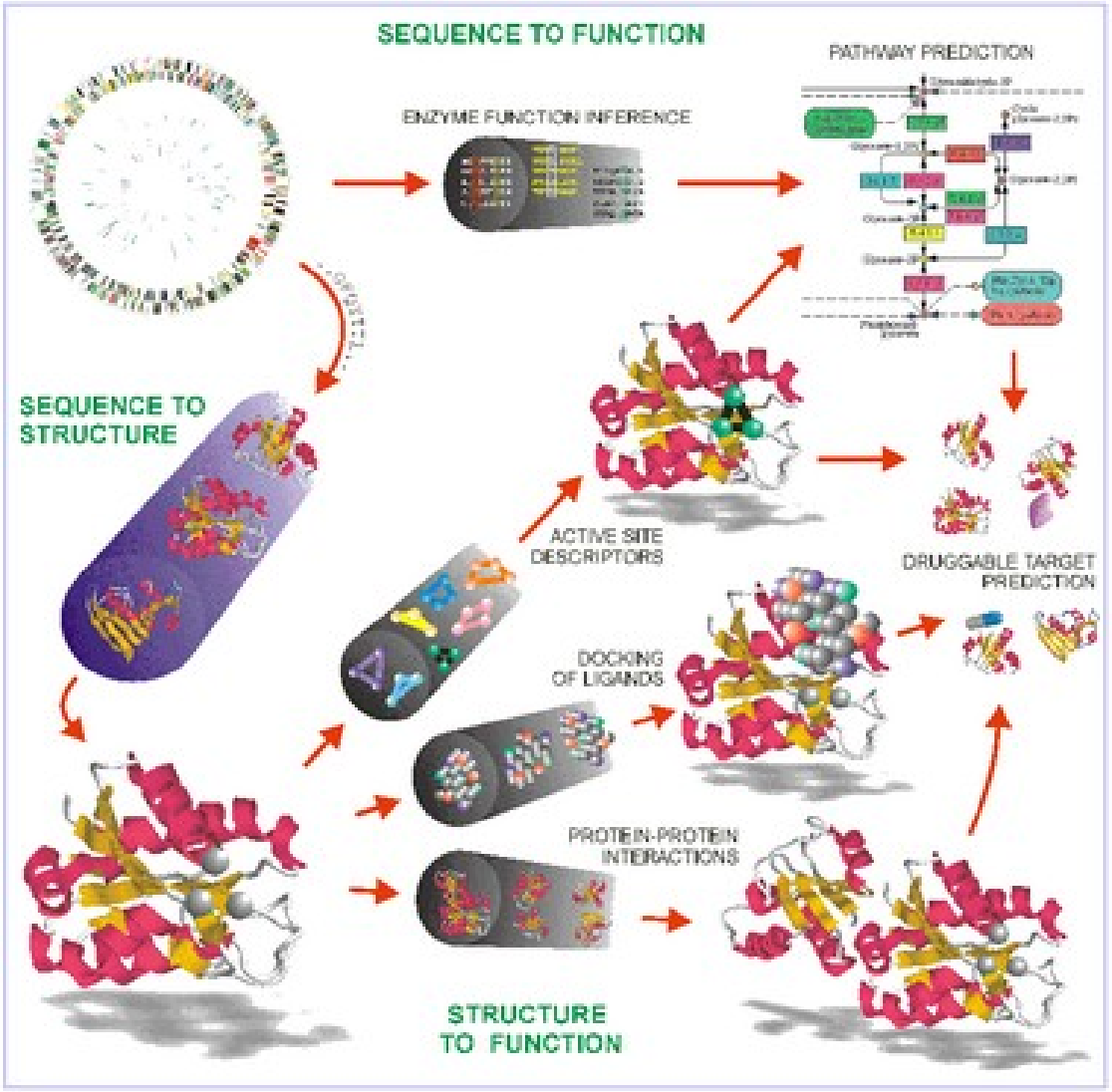
Structure: Primary Secondary Tertiary

Helix β-sheet



Voet, Biochemie: 1998

Sequence → Structure → Function



http://cssb.biology.gatech.edu/skolnick/

The most important question:

- **What do I want to use my protein model(s) (experimental or predicted) for?**

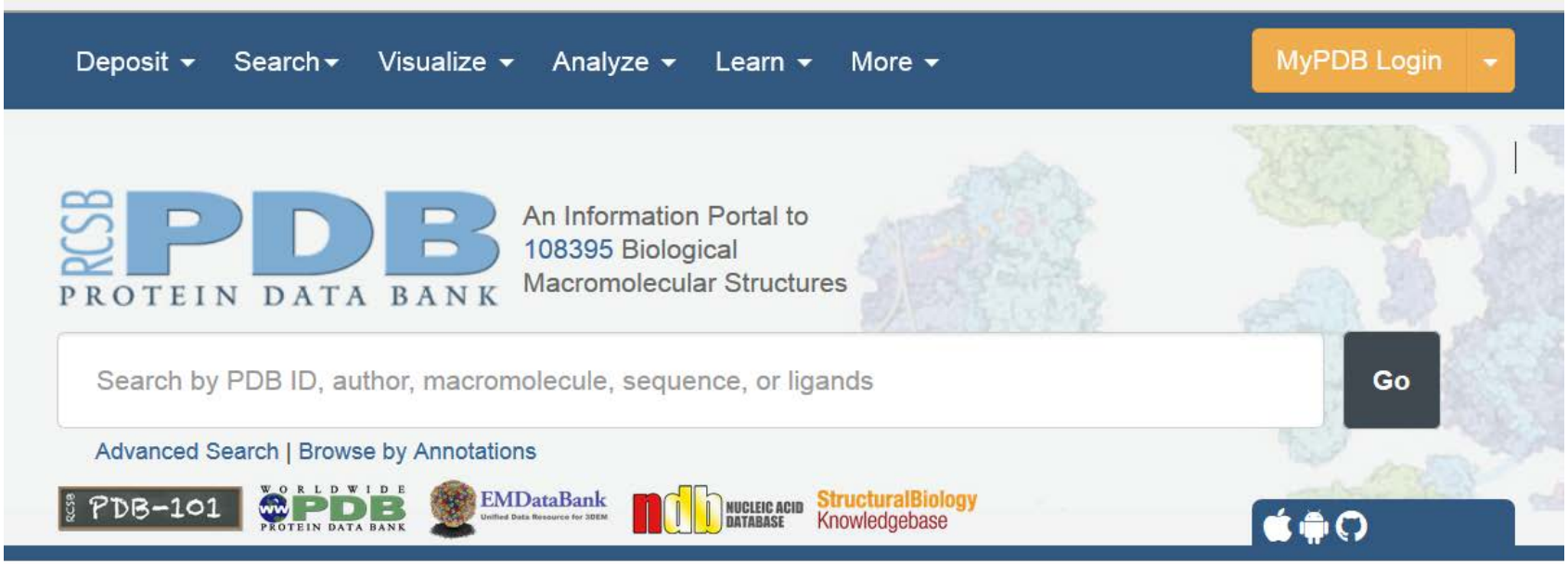
(What effect will errors in the models have?)

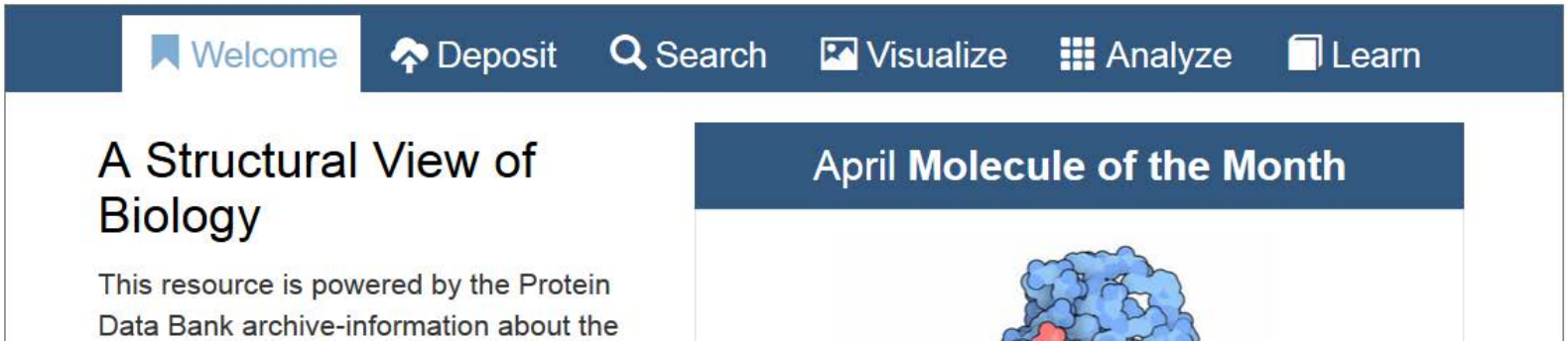
Is the structure of the protein already known from experiments?

If so, does this provide me with a suitable 3D model for answering my question?

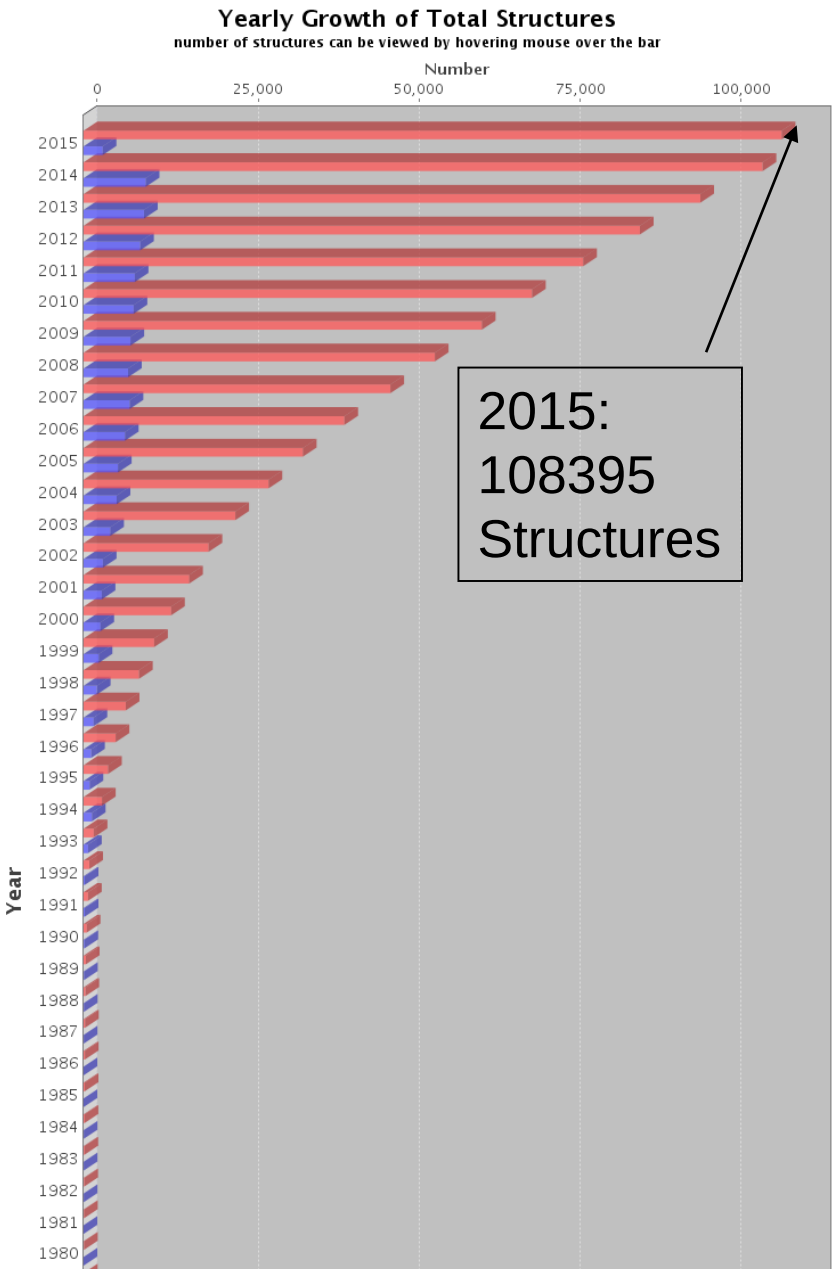
The PDB: Protein Data Bank

- www.rcsb.org





Structures in the PDB:

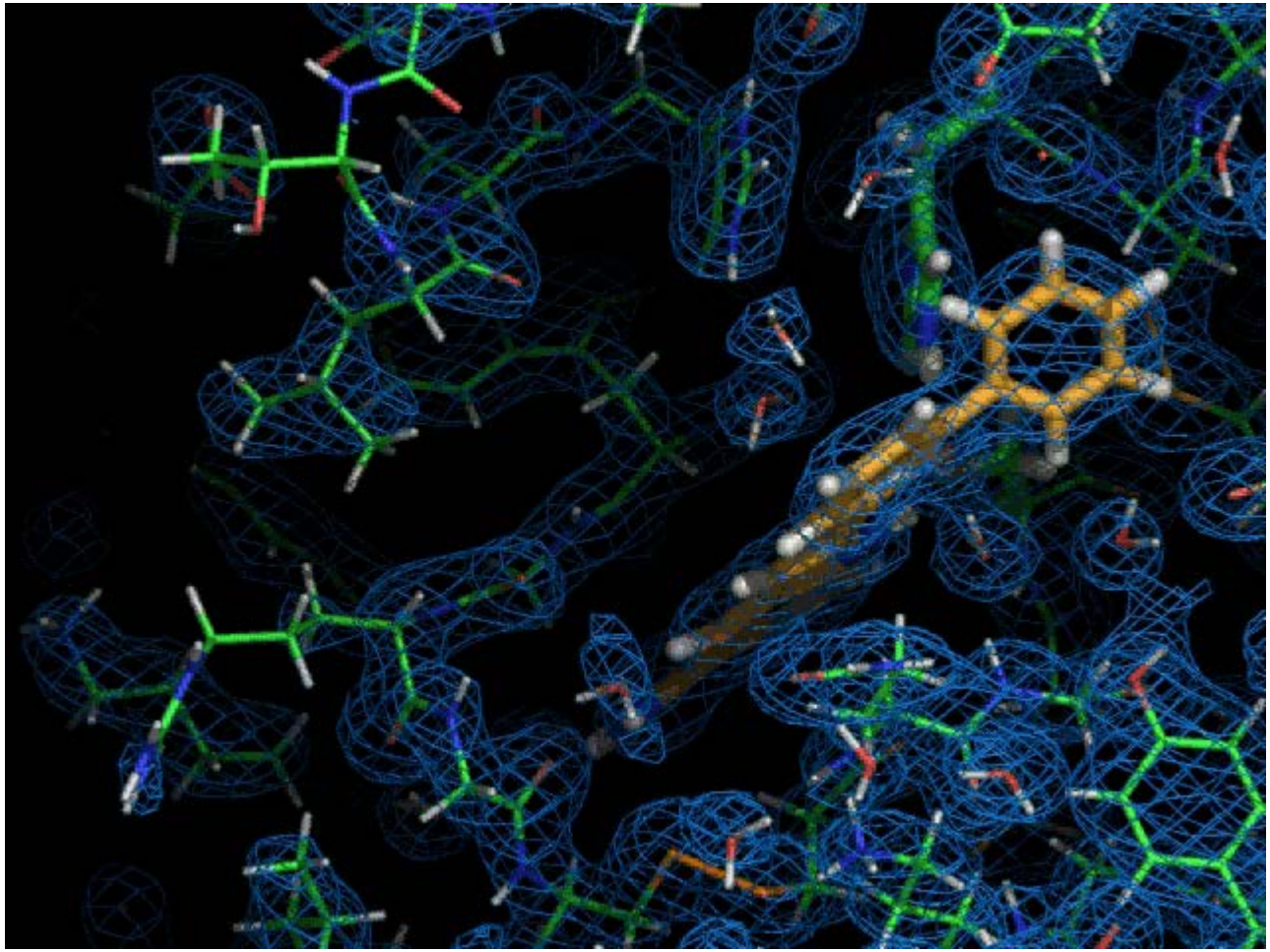


Structures from:

- Xray crystallography
- Nuclear Magnetic Resonance
- Electron microscopy

Crystal structures are models:

Protein-ligand complex with electron density map



human thrombin at a resolution of 1.68 Å (PDB code: 1o5a)

Preparation of a structure:

Structure preparation for modelling:

- downloading PDB structure
- adding **symmetry mates**
- adding **missing parts**

(e.g., loops, side chains)

- adding **hydrogens**

(protonation state!)

- re-orientation of

side chains

(**His, Asn, Gln**)

and **water molecules**

- assigning **parameters**

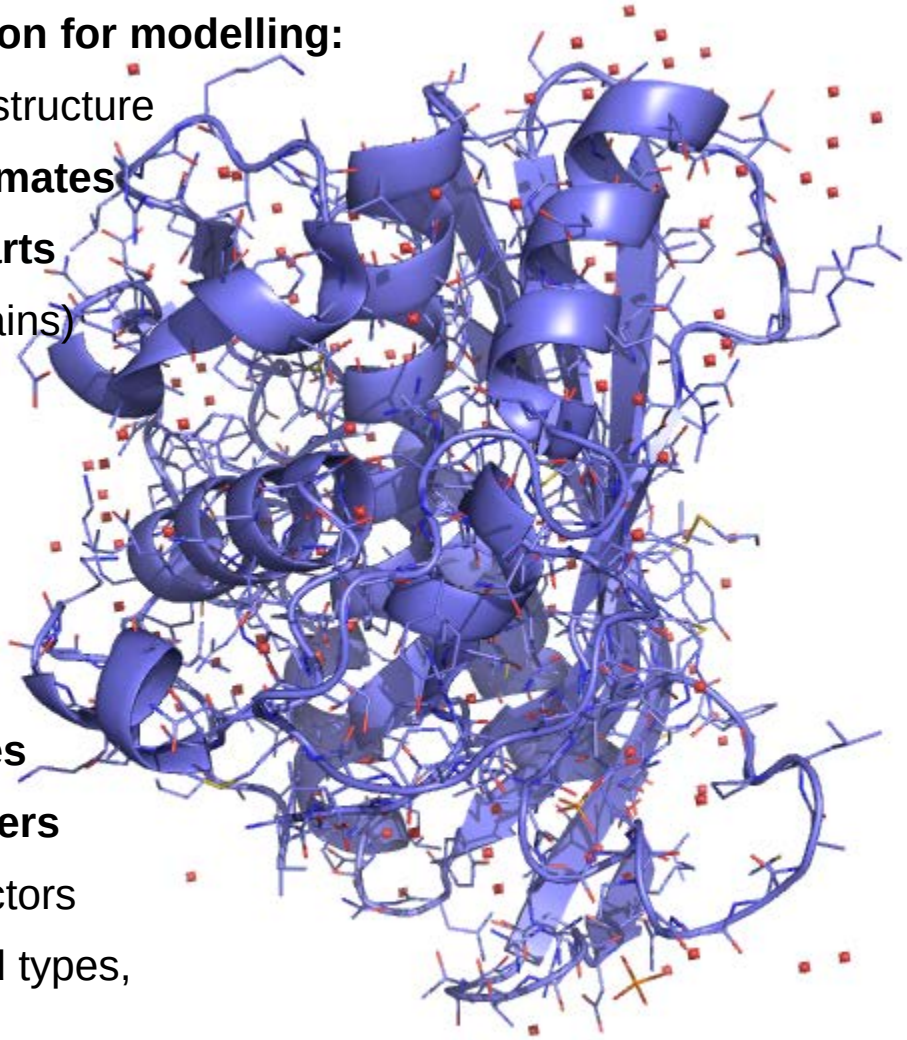
for ligands and cofactors

(e.g., atom and bond types,

angles, charges)

- for MD:

adding water box

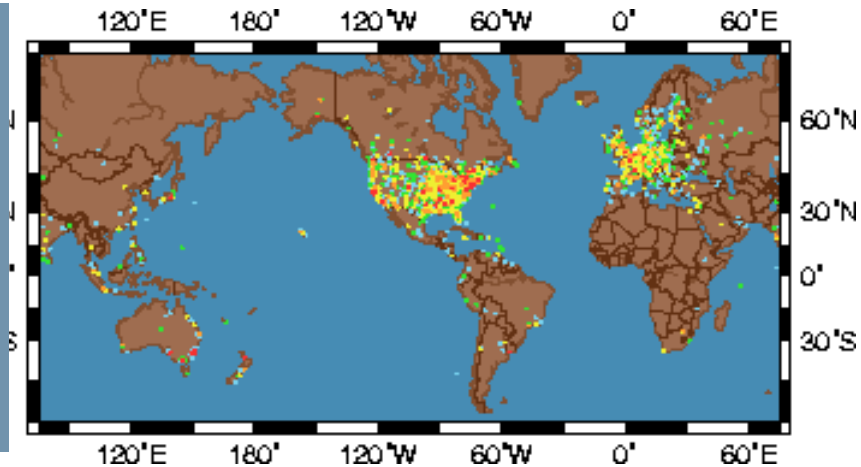


human thymidylate synthase (1ypv)
space group: $P3_121$

If I only know the protein sequence, can I fold the protein and predict the structure?

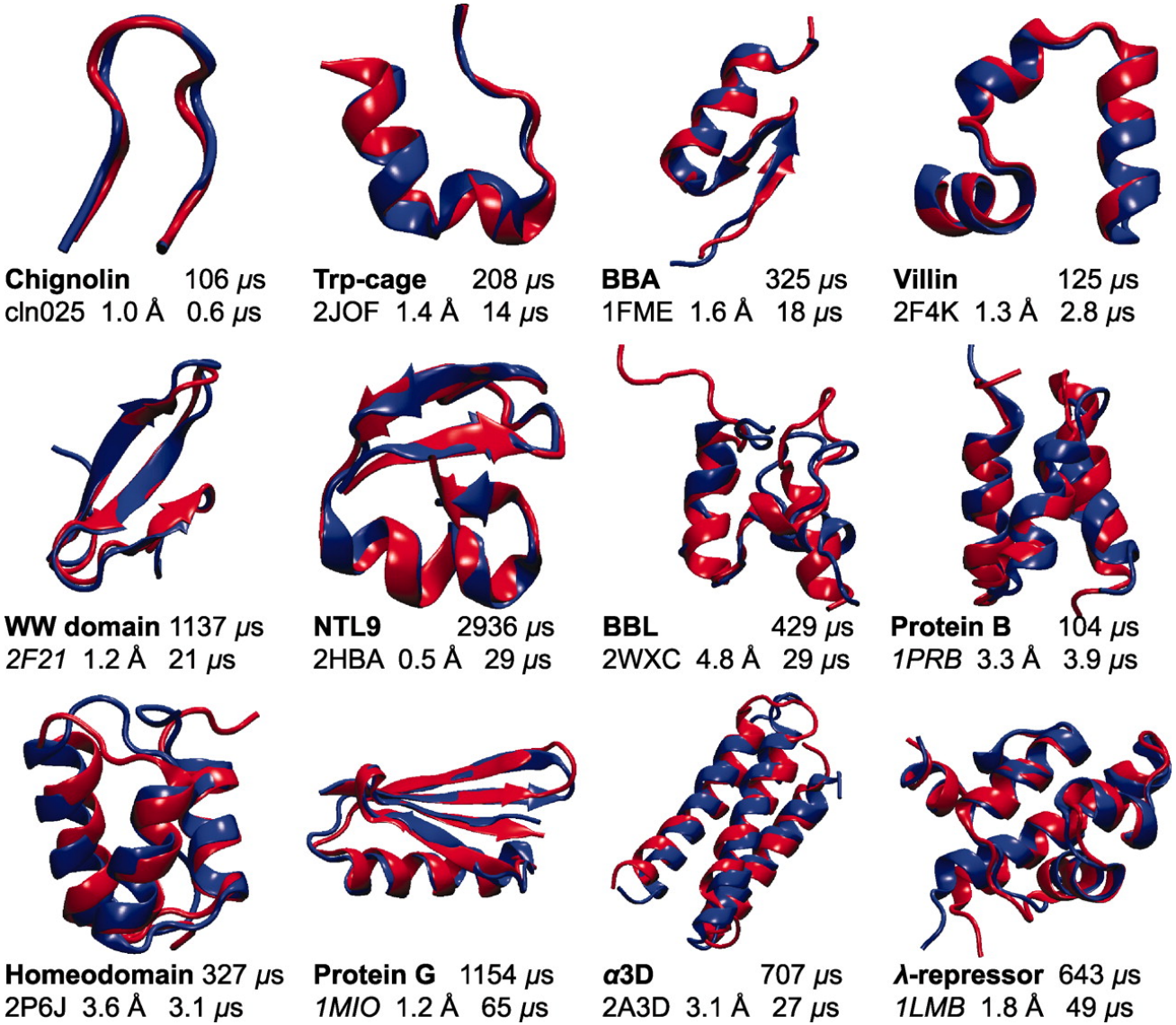
Folding from first principles (ab initio)

- Molecular dynamics simulation
  - CPU
  - Distributed computers
- Villin headpiece (36AA)
  - Fastest folding protein (microsec)
  - Duan&Kollmann (Science 1998)
  - +Many others
- Folding@Home
  - V. Pande, Folding@Home
  - Multiple folding simulations



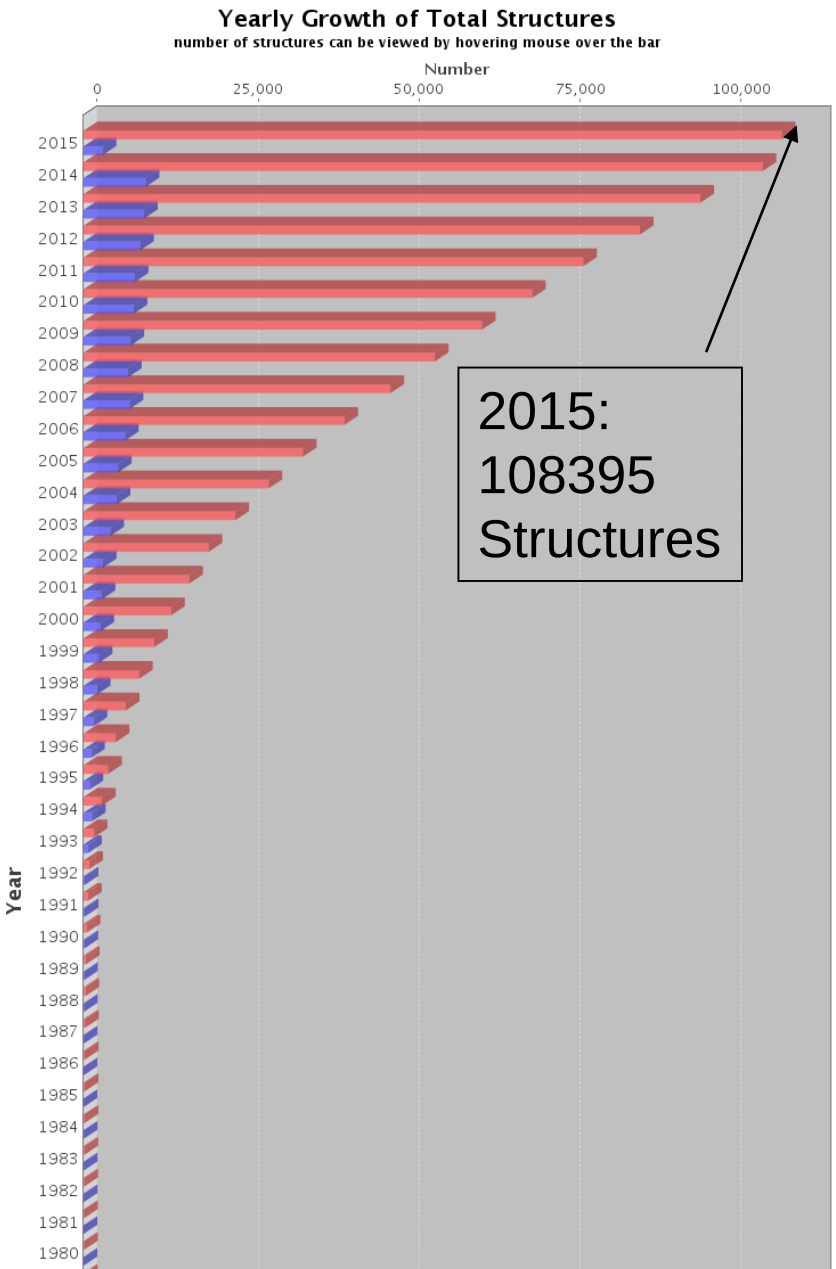
Folding from first principles (ab initio)

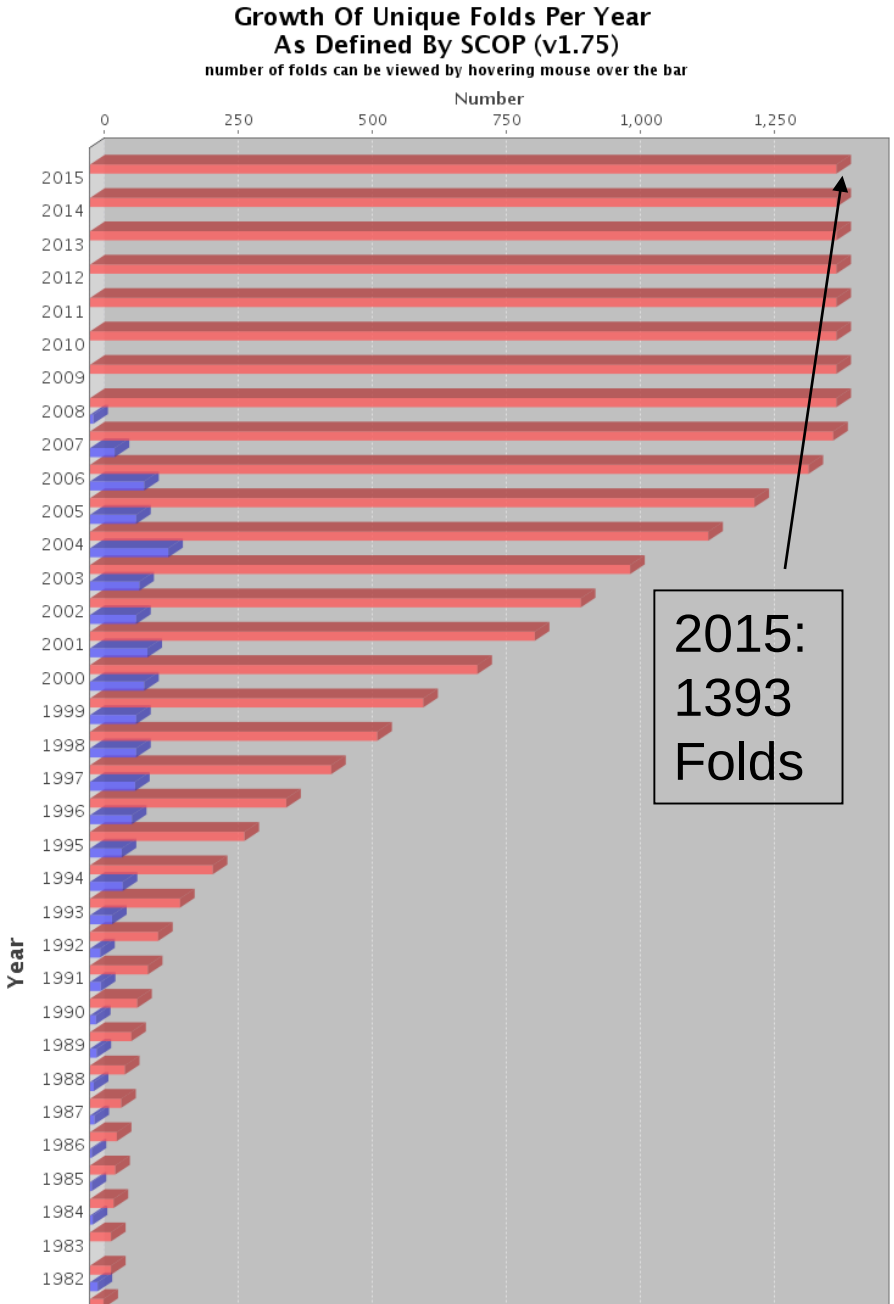
- 12 fast-folding small diverse proteins
  - DE Shaw et al, Science 2011
  - MD simulations: 100 µs - 1 ms



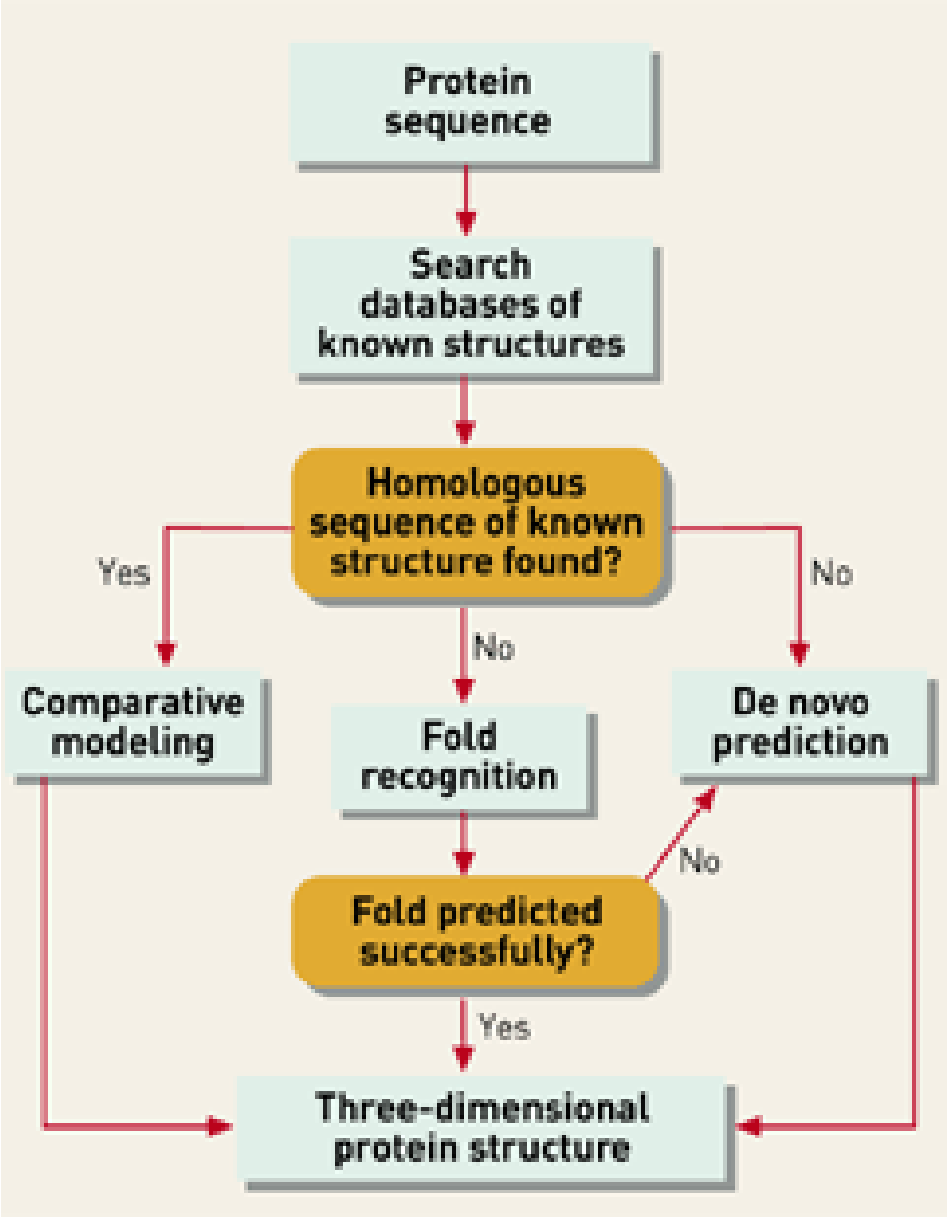
Kresten Lindorff-Larsen et al. Science 2011;334:517-520

Protein 3D Structures in the PDB:





Can I make a 3D model of this protein?



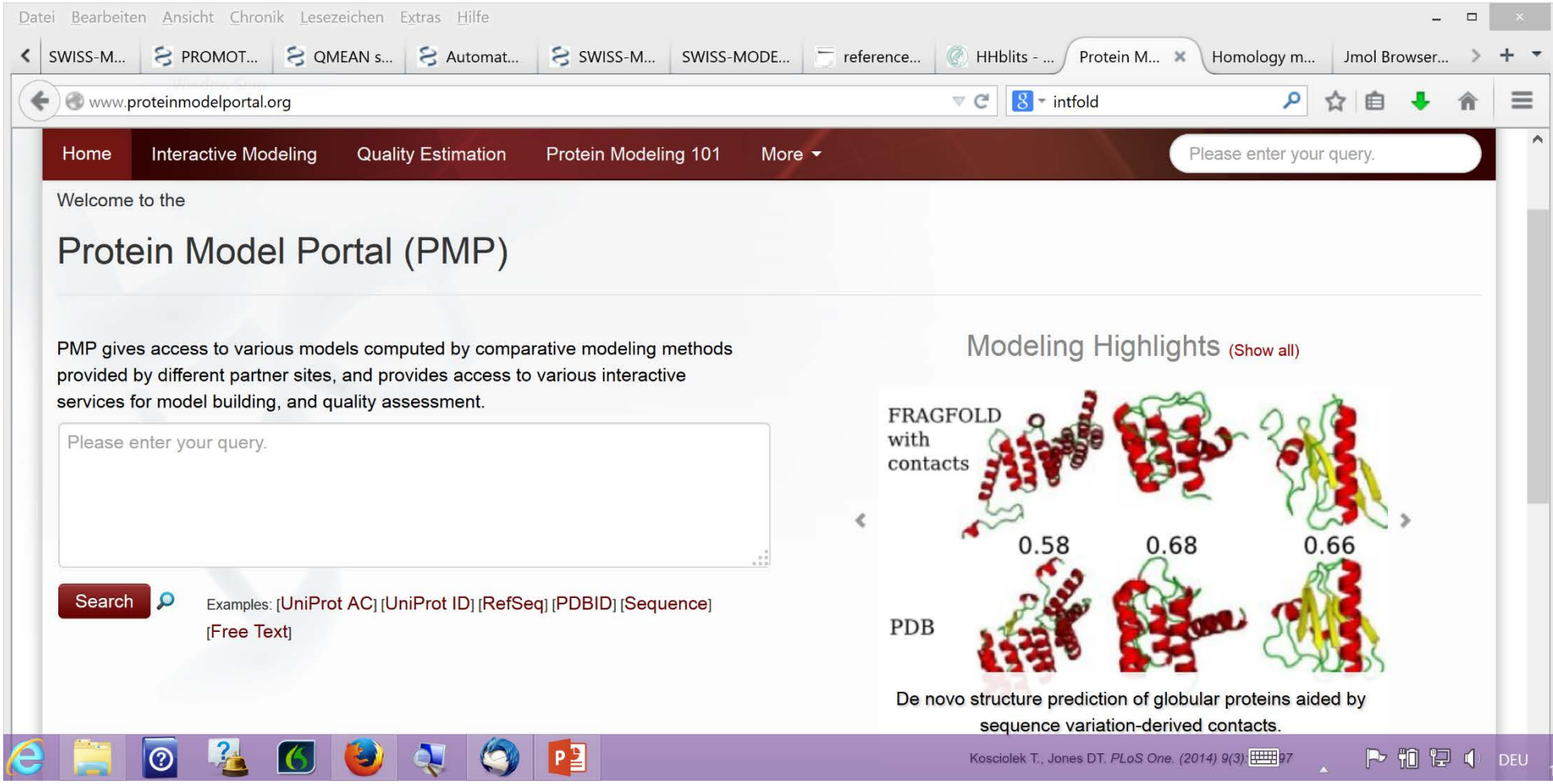
From:
C&EN, 04.08.2003
(R. Russell)

Can I find a 3D model of this protein?

- **Is my protein in a structural database?**
  - Experimentally determined structures:
    - PDB (RCSB)
    - http://www.rcsb.org
  - Comparative model databases, e.g.:
    - ModBase,
    - SwissModel Repository
  - Single point of entry for finding protein structures:
    - Protein Structure Initiative Knowledgebase (PSI KB)
    - http://www.proteinmodelportal.org/

The Protein Model Portal (PMP)

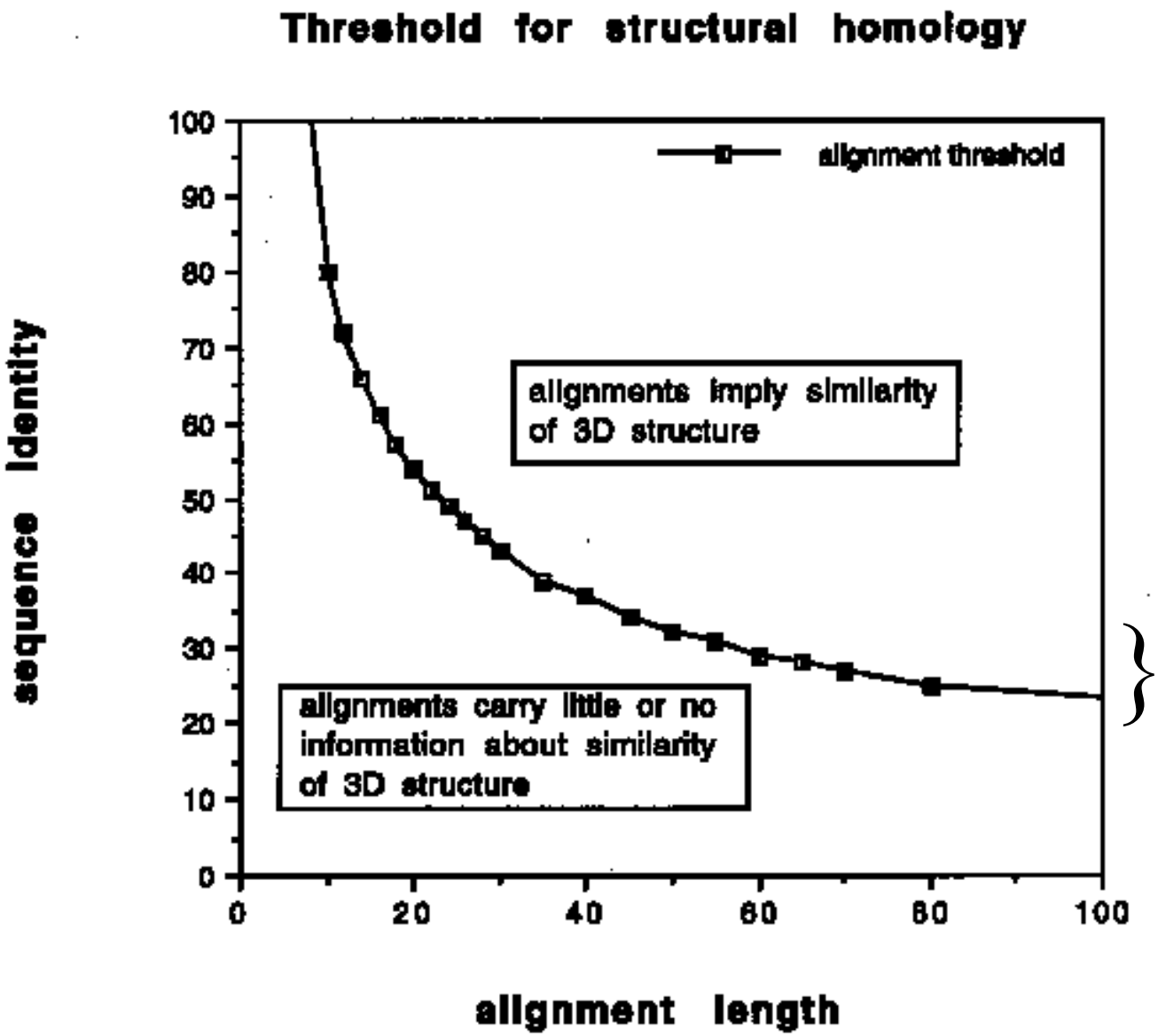
– http://www.proteinmodelportal.org/



Can I make a 3D model of this protein?

- **Is it similar in sequence to proteins with known structure?**
  - Sequence alignment
    - Blast, FASTA, Modeller
    - Multiple-sequence alignment (PSI-BLAST)
    - Structure-based

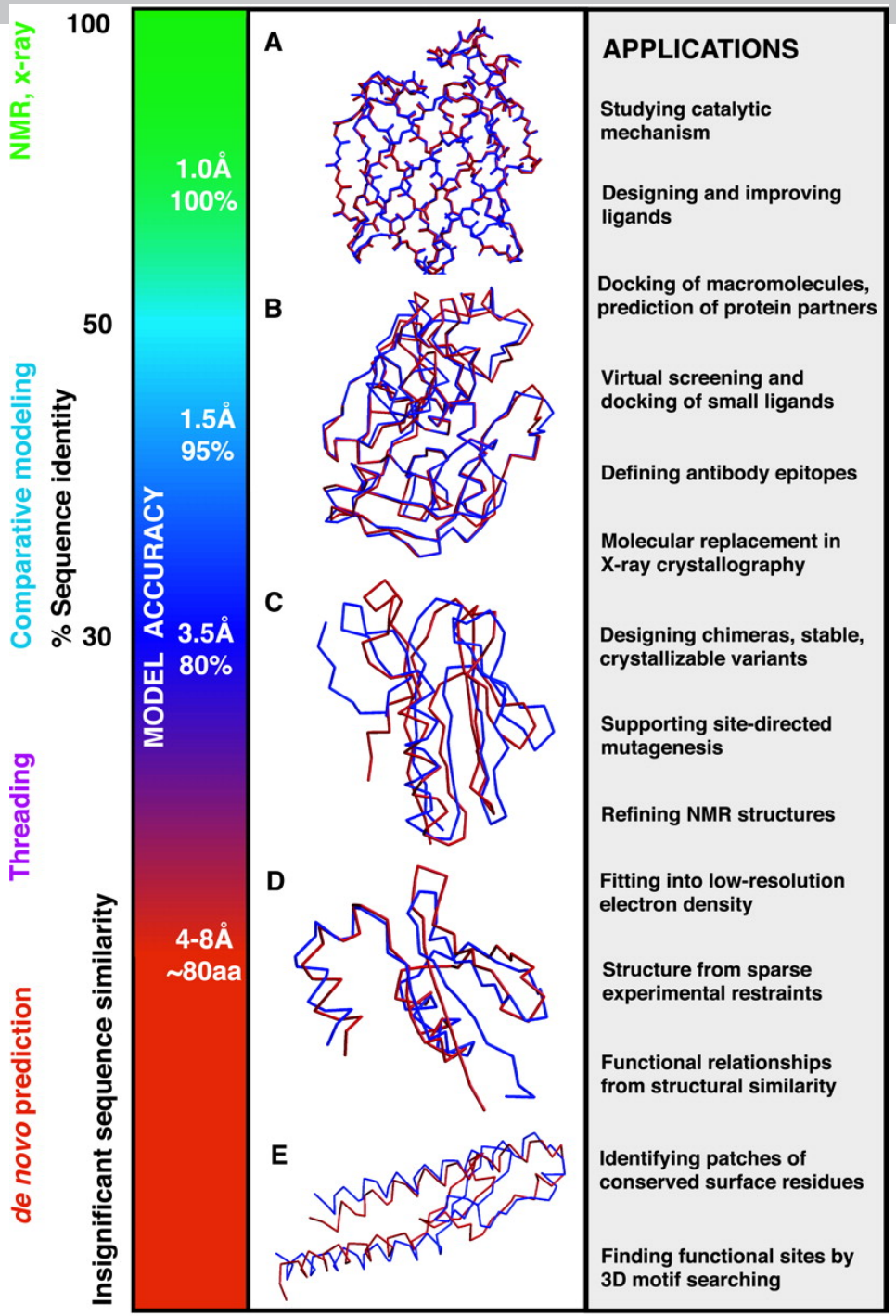
Can I make a 3D model of this protein?



Twilight zone

From: G. Vriend, R. Rodriguez, Based on: C. Sander. R. Schneider Proteins (1991) 9, 56-68.

Can I make a 3D model of this protein?



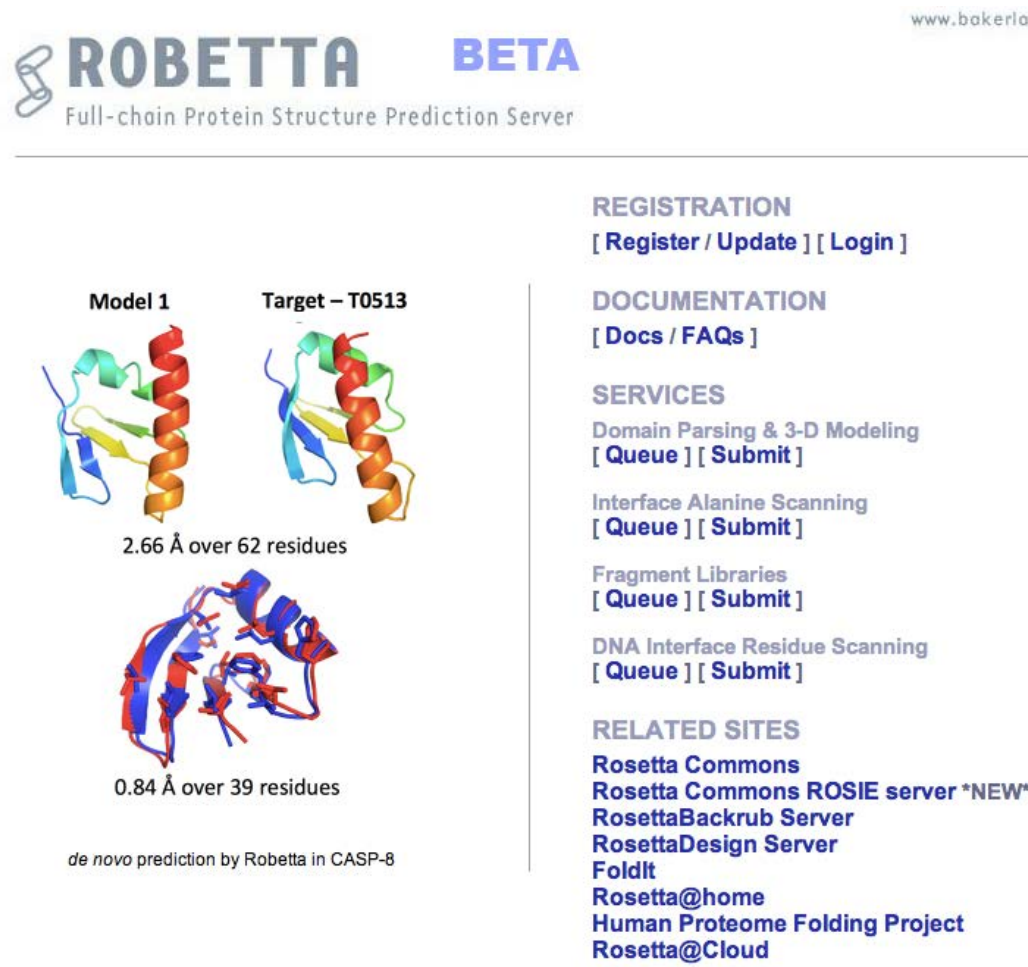
Sali, Baker, Science 2001

Can I make a 3D model of this protein?

- What if my protein is not similar in sequence to any proteins with known structure?

- Fold Recognition
  - Threading of sequence on structures in database

- De Novo Prediction
  - Knowledge-based
  - Simulation
  - *For small proteins*



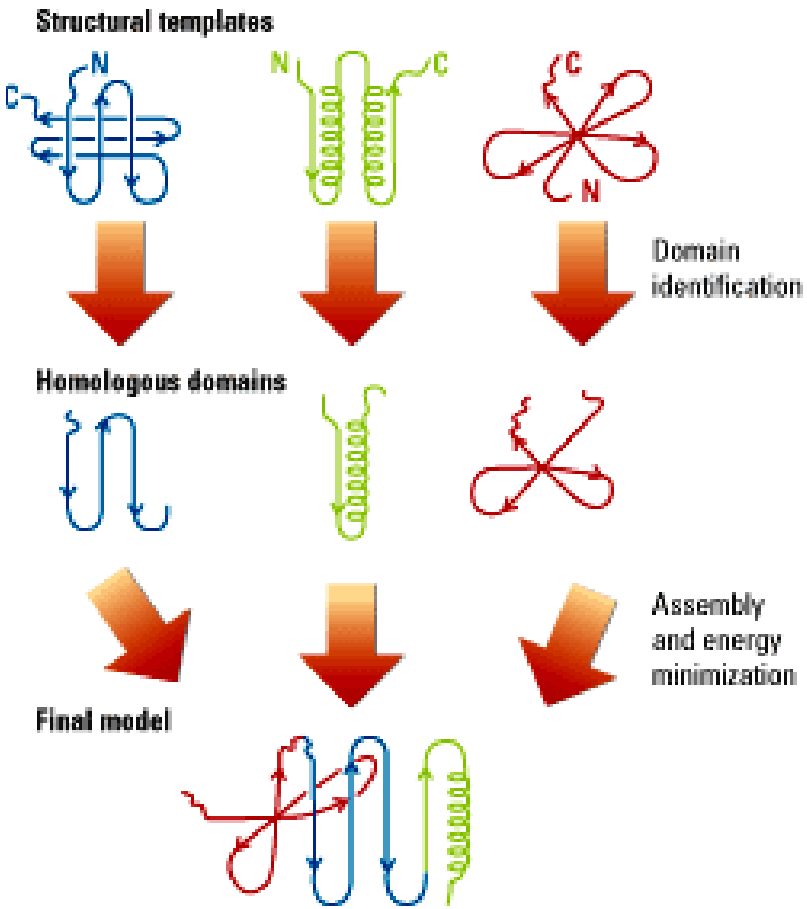
Robetta.bakerlab.org

Can I make a 3D model of this protein?

- What if my protein is not similar in sequence to any proteins with known structure?

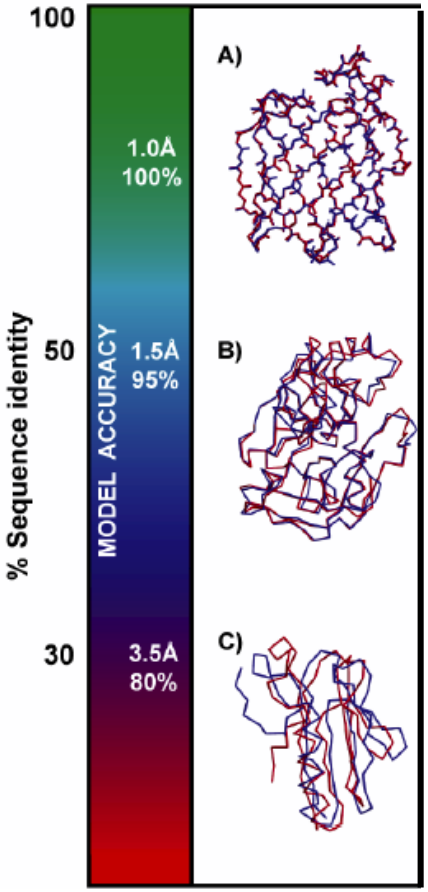
- Fold Recognition
  - Threading of sequence on structures in database

- De Novo Prediction
  - Knowledge-based
  - Simulation
  - *For small proteins*



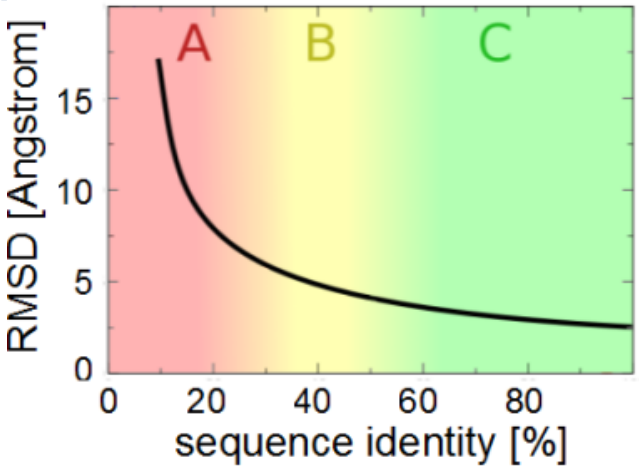
Rosetta:
Simons, K. T., et al. *Proteins Suppl.* **1999,** *3,* 171–176

Can I make a 3D model of this protein?



Marti-Renom, Yerkovitch, Sali Current Prot. Protein Sci.(2002) 2.9.1

- How much sequence identity for comparative (homology) modeling?
  - >50% : "reliable" model
  - >30% : "useful" model
  - <30% : "might be useful" model
- Other metrics for deciding on comparative modeling
  - Multiple sequence alignment
  - Structure-based alignment

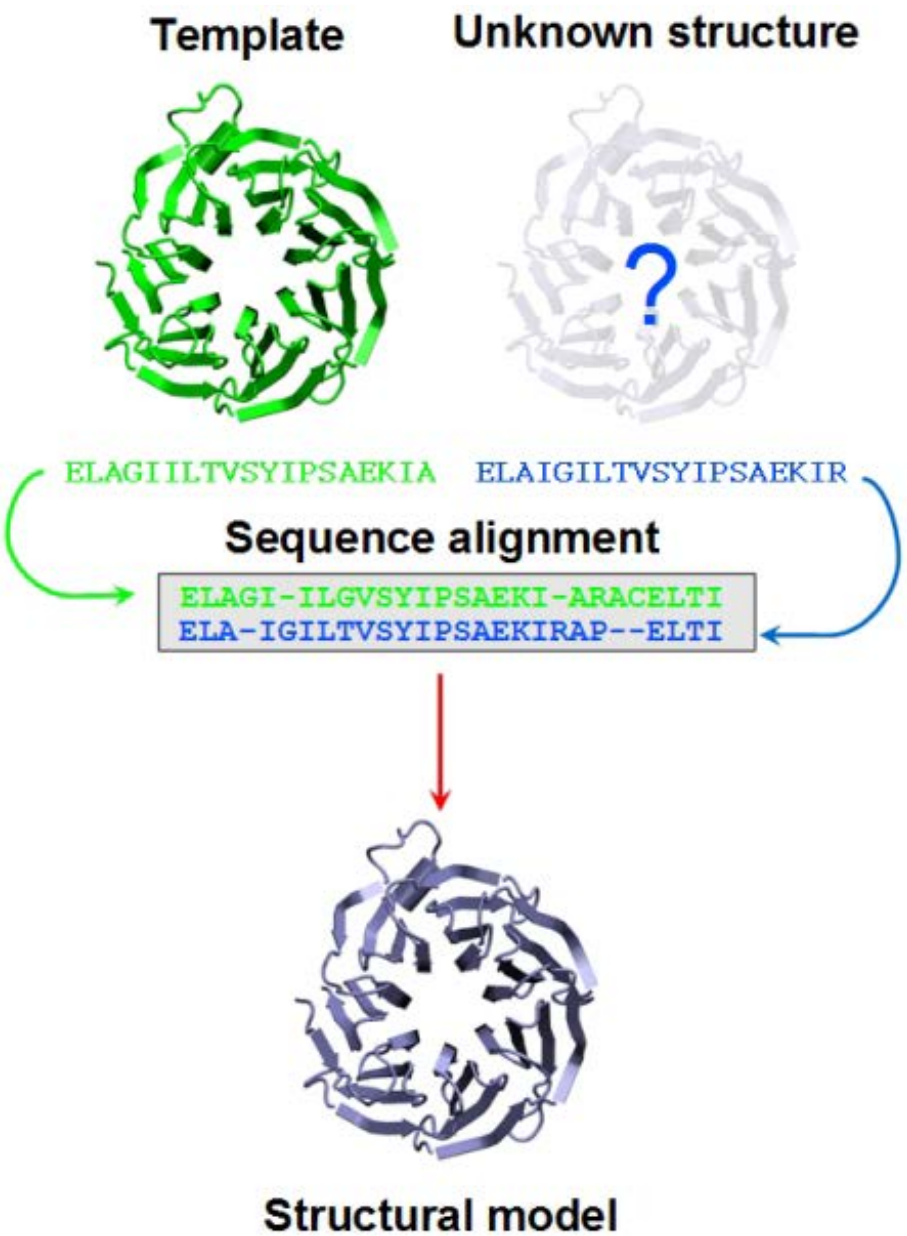


SWISS-MODEL webserver

How can I make a 3D model of this protein?

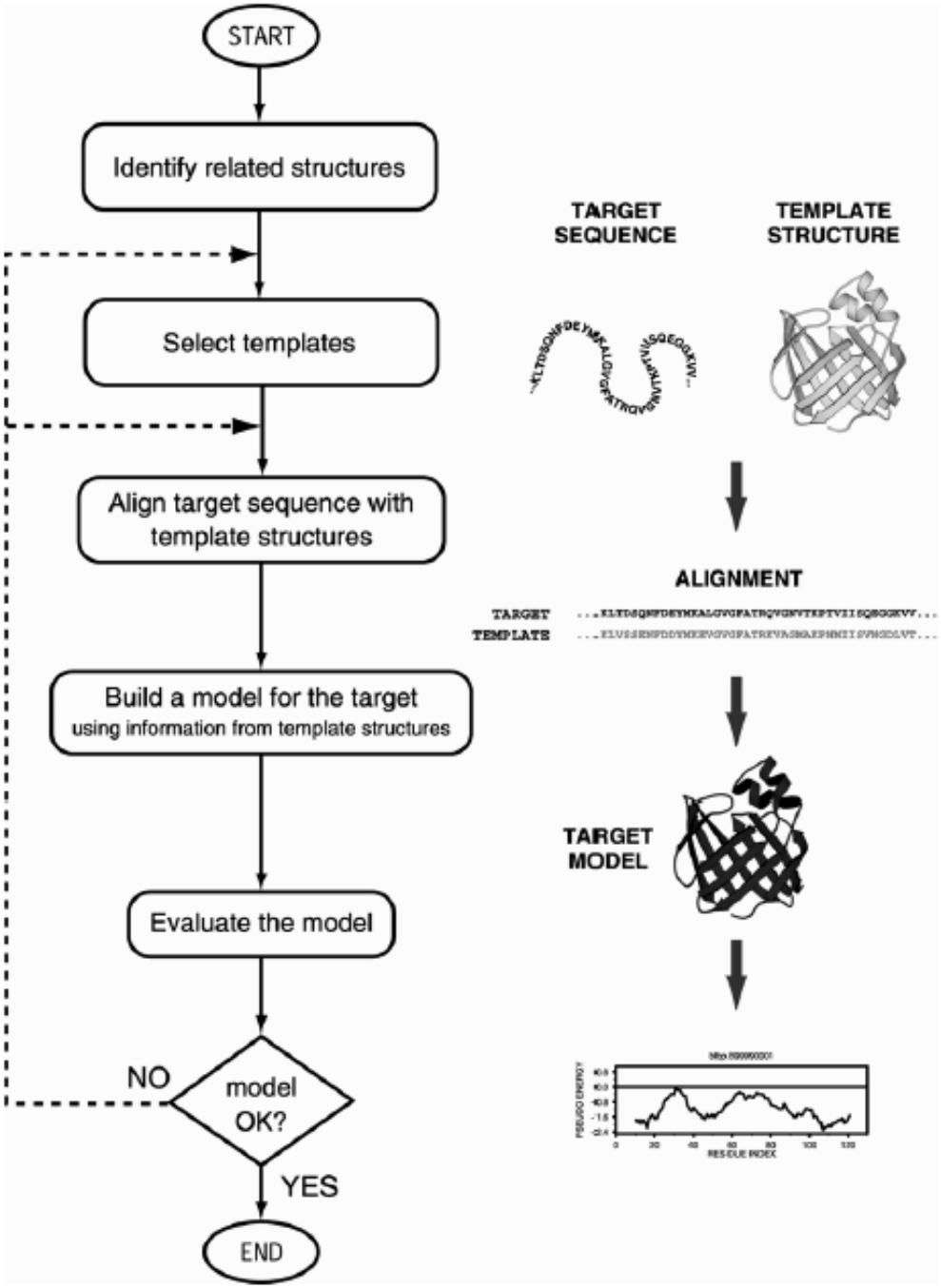
- **Use a server!**
  - http://www.proteinmodelportal.org/?pid=modelling_interactive
  - Enter your aminoacid sequence
  - Run these tools (comparative modeling and/or threading)
    - ModWeb
    - M4T
    - SwissModel
    - I-TASSER
    - Hhpred
    - Phyre2
    - InfFOLD2
    - RaptorX

How can I make a 3D comparative model of my protein?



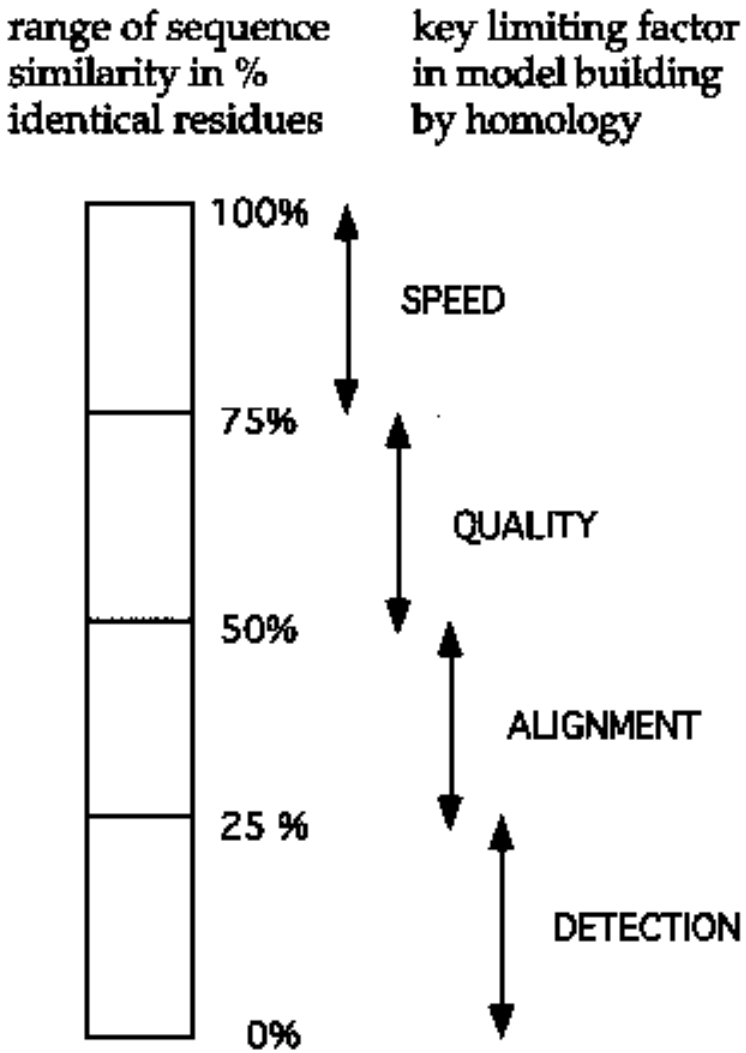
http://www.unil.ch/pmf/en/home/menuinst/technologies/homology-modeling.html

How can I make a 3D comparative model of my protein?



Marti-Renom, Yerkovitch, Sali Current Prot. Protein Sci.(2002) 2.9.1

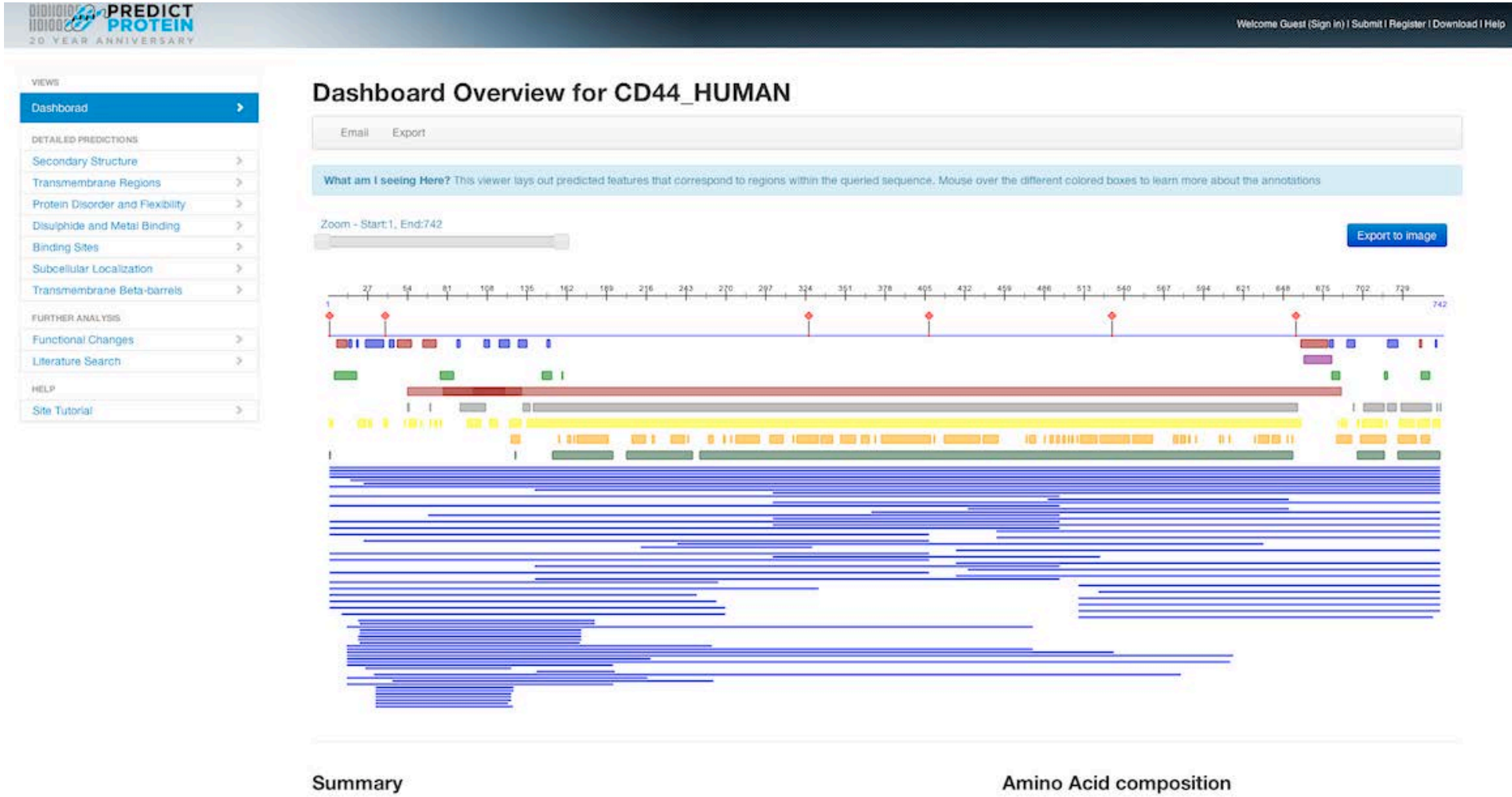
Can I make a 3D comparative model of my protein?



From: G. Vriend

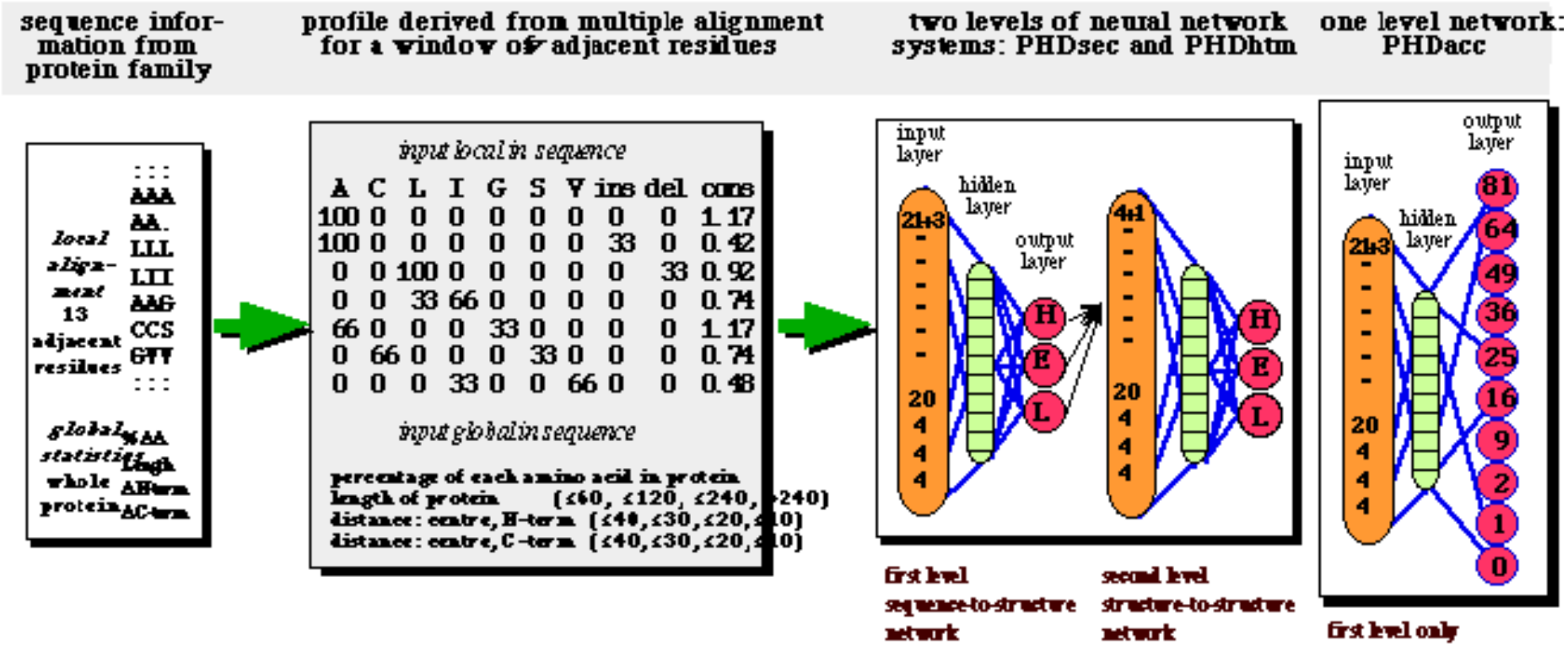
Figure 1. The main limiting steps for model building by homology as function of the percentage sequence identity between the structure and the model.

Local sequence predictions



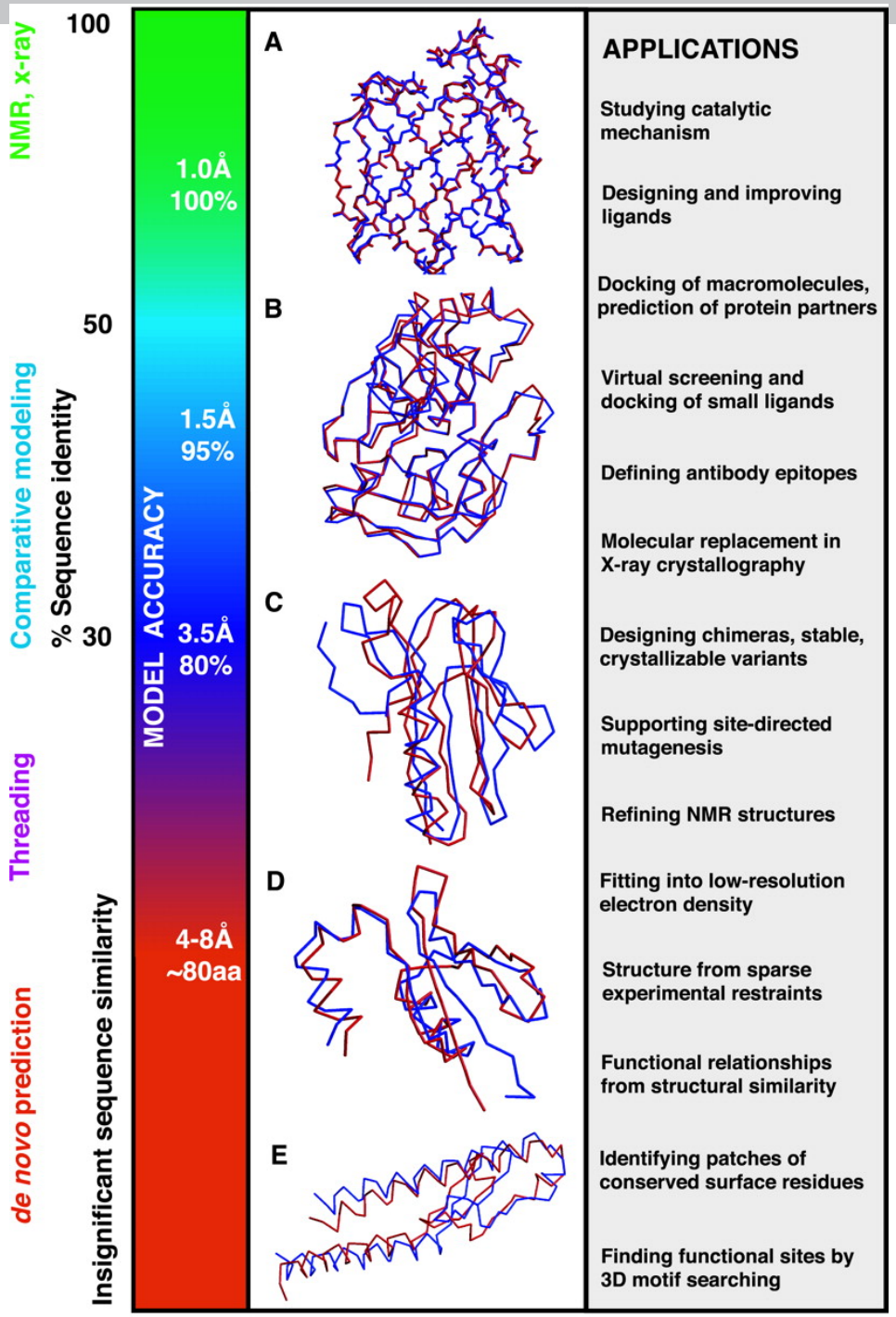
www.predictprotein.org

Local sequence predictions (with global information)



Rost, Burkhard *Meth. in Enzym*. 1996, 266, 525-539

What can I do with my protein model?



Sali, Baker, Science 2001

Challenges for Protein Modeling:

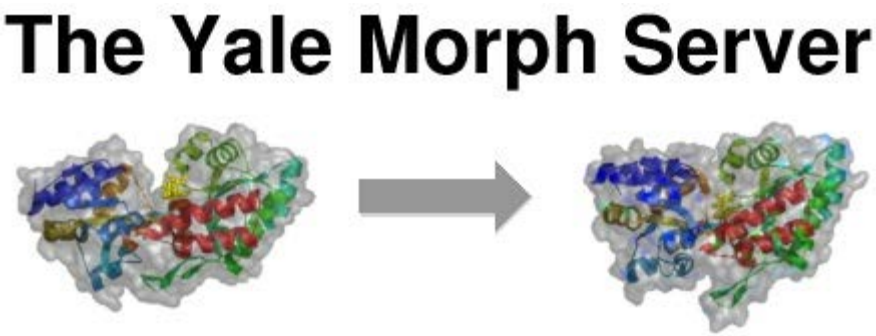
- Sequence alignment
- Loop modeling
- Modeling complexes and oligomers
- Distinguishing the best model(s)

- !! Proteins are dynamic !!

- Predicting bound structure from unbound structure
- Predicting correct oligomeric state and structure
- Protein folding

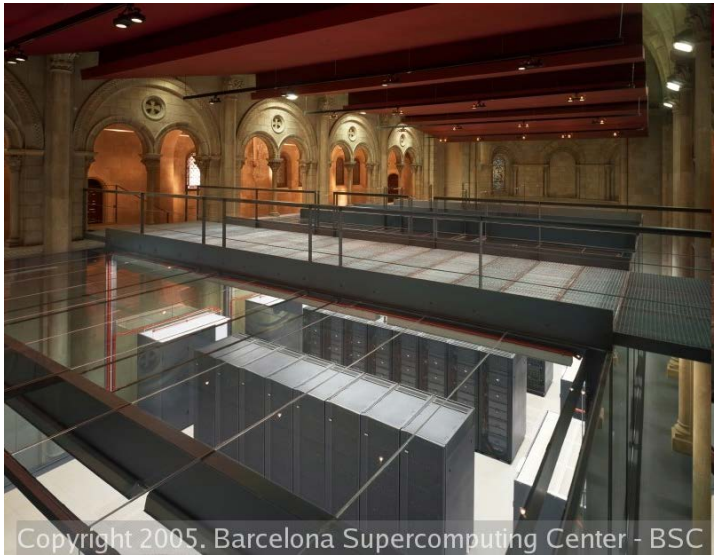
Protein dynamics:

- Morphing:
  - Between two known structures of one protein
  - Morph (Gerstein)
    - http://www.molmovdb.org/molmovdb/morph/
    - Interpolation and energy minimization



Protein dynamics:

- Molecular dynamics simulations
  - Databases
    - MoDEL - Molecular Dynamics Extended Library (Orozco)
      - http://mmb.pcb.ub.es/MODEL/
      - Standard MD simulations
      - Principal components available



    - Dynameomics (Daggett)
      - Protein unfolding using high temperature
      - http://www.dynameomics.org/
    - Etc

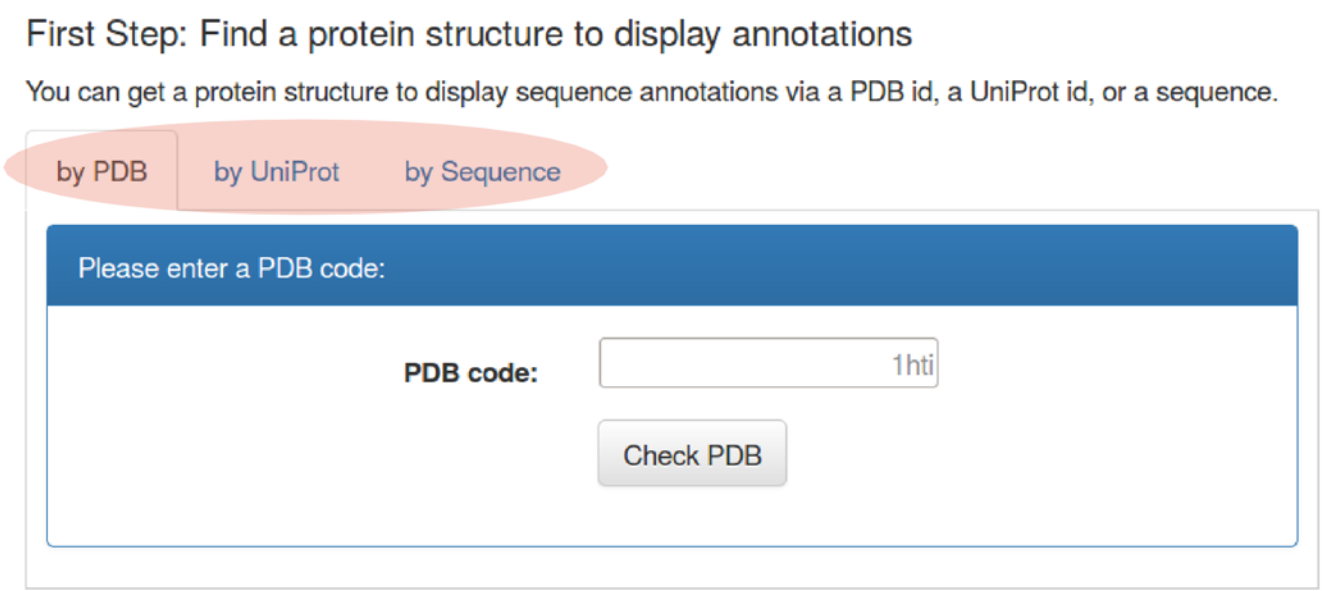
Where can I find out about protein structure prediction tools?

- Collection:
  - http://www.proteinmodelportal.org/?pid=101
- CASP:
  - http://predictioncenter.org/
  - http://www.forcasp.org/
- Tertiary and Secondary structure prediction:
  - http://predictioncenter.org/index.cgi?page=links
  - **http://www.jove.com/video/3259/a-protocol-for-computer-based-protein-structure-function**
- Comparative modeling:
  - http://www.salilab.org/modeller/modeller.html
  - http://swift.cmbi.ru.nl/teach/HOMMOD/
- Folding@Home
  - http://folding.stanford.edu/

What do we already know about mutations in this protein or at this site in the protein ?

ProSAT+:

PROtein Structure Annotation Tool



http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

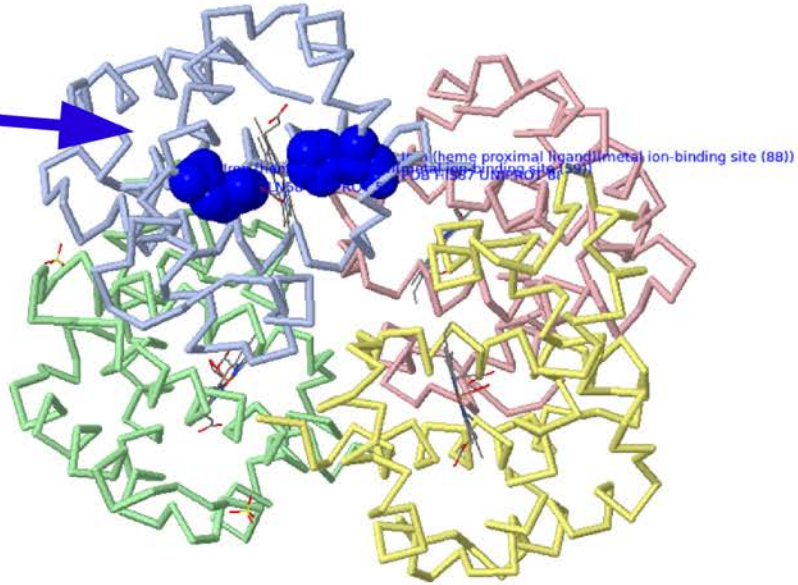
ProSAT+:

PROtein Structure Annotation Tool

Display sequence annotations on 3D structure

Chain A, Source P69905 (Homo sapiens)

- modified residue
- strand
- helix
- variant
- binding
  - Iron (heme distal ligand) 58:A (59)
  - Iron (heme proximal ligand) 87:A (88)
- turn
- kinetics



http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

ProSAT+:

PROtein Structure Annotation Tool

Define sequence annotations and generate URL

ProSAT+ - Generate user URL

User defined sequence annotations

#	UniProt accession	Position(s)	Name	Description	
1	P69905	1-3,5	Name 1	Sequence Description 1	
2	P68871	66	Name 2	Sequence Description 2	
		position(s)	name, do not use ;;| or &	description, do not use ;;| or &	Add

Here you can enter sequence annotations in the following format
P60174:123-129,120,135-140:name1:ThisisOneFeatureDescription|P60174:250,255:name2:AnotherFeature

Convert

You can also transfer all your selected sequence features and amino acids into the table:

Add Selected Features

Get URL

Your Url is:

http://localhost:9000/prosat/prosatexe?pdbcode=1j7y&submit=Check+PDB&user=P69905:1-3,5:Name%201:Sequence%20Description%201|P68871:66:Name%202:Sequence%20Description%202|

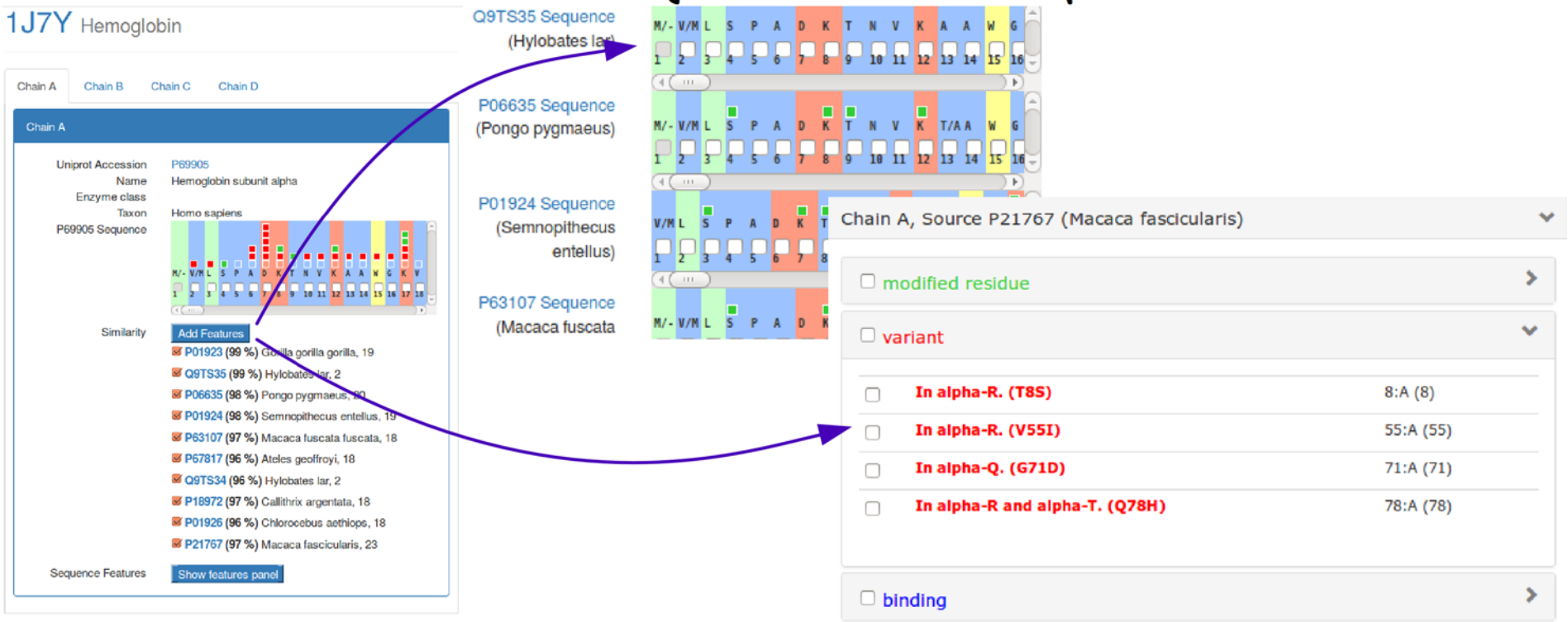
http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

ProSAT+:

PROtein Structure Annotation Tool

Use sequence annotations from evolutionarily related proteins



http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

ProSAT+:

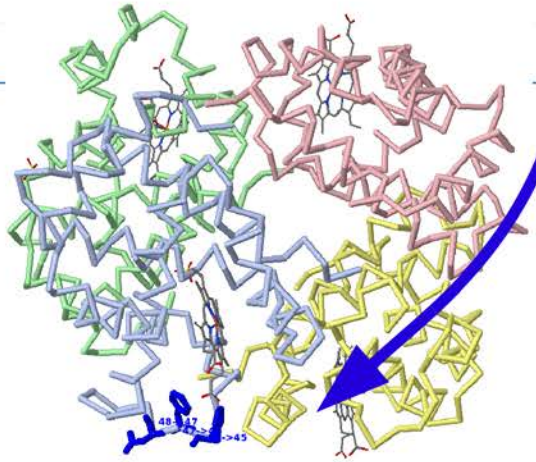
PROtein Structure Annotation Tool

Search for sequence motifs

Motif search: [H][PF][D] Clear Search

Motif found:

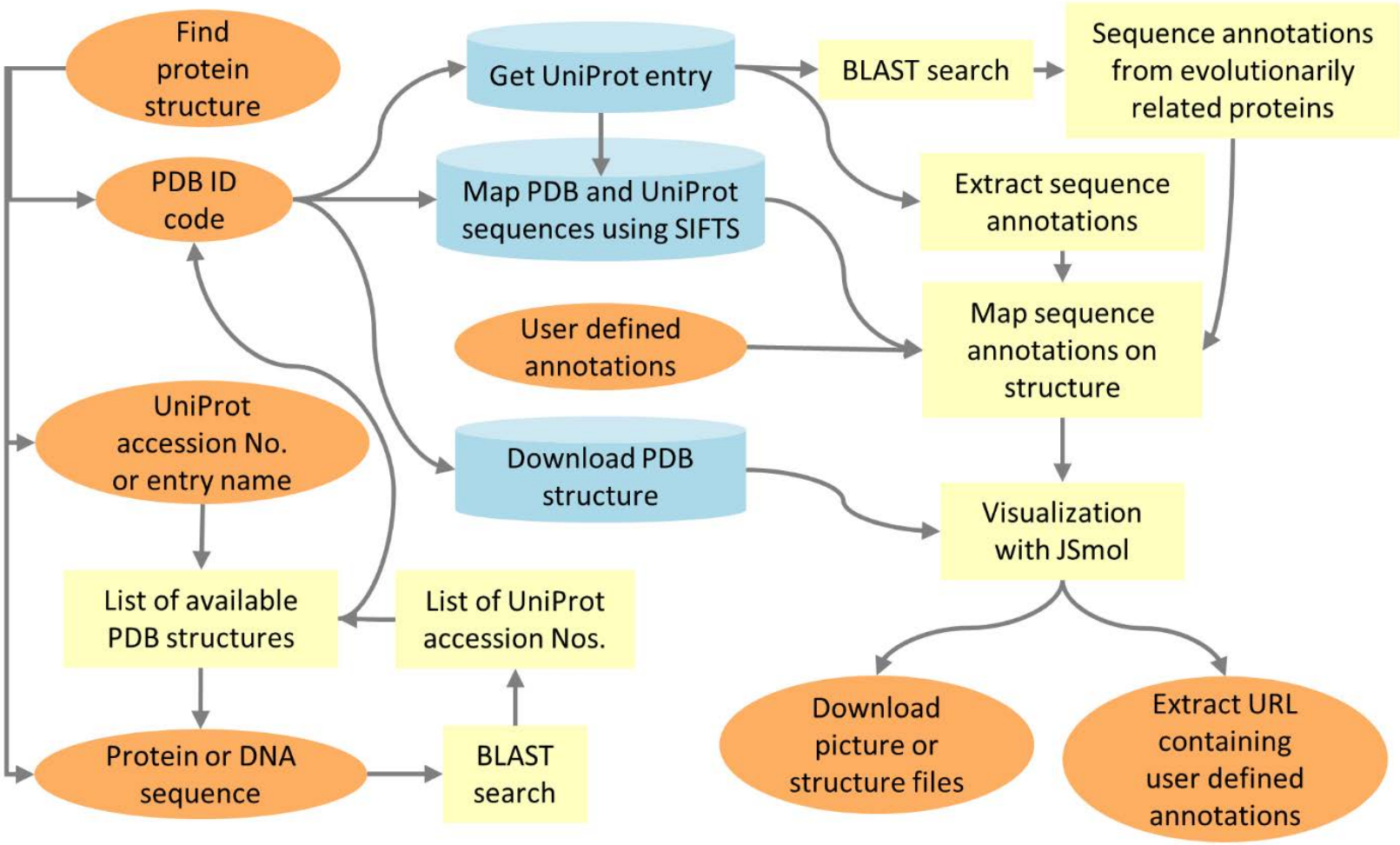
Chain	Match	Position(s)
Chain A	HFD	A 46-48



http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

ProSAT+:
Workflow

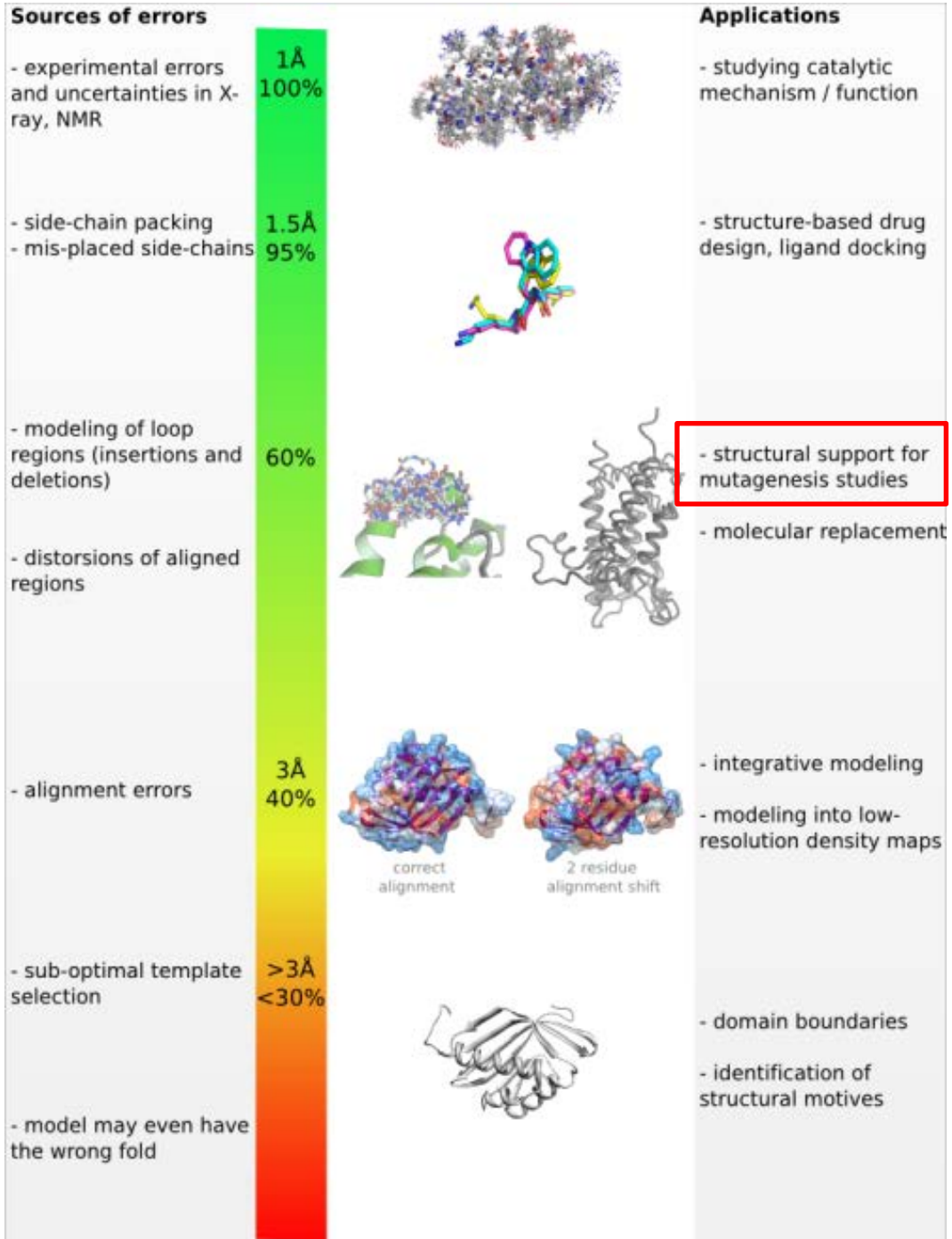


http://prosat.h-its.org

Stank, Richter, Wade PEDS, 2016, in press

What mechanistic effect does a particular mutation have?

Using protein models to predict the effects of mutations

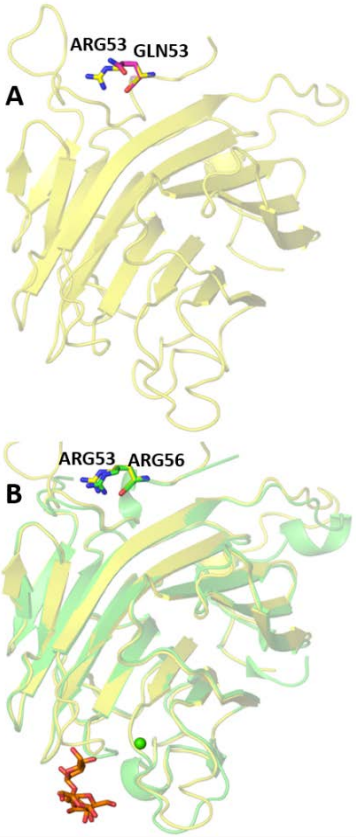


http://www.proteinmodelportal.org/?pid=documentation#modelquality

Missense mutation : mechanism?

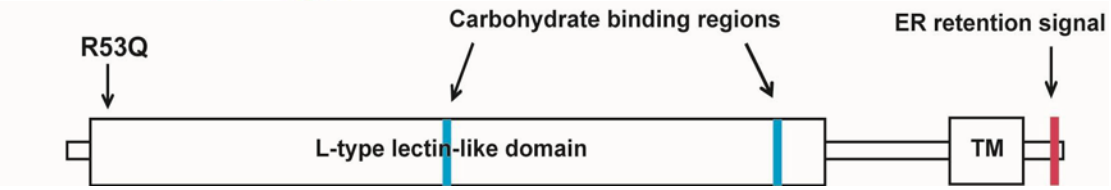
- Homozygous missense mutation in *lectin mannose-binding 2-like* (**LMAN2L**) gene (**R53Q**), identified in whole exome sequencing, segregates with severe **intellectual disability** & **epilepsy** in consanguineous Pakistani family

- **LMAN2L**: ER cargo receptor for glycoprotein transport & quality control, expressed in brain



- **LMAN2L** (Swiss) modeled on Vip36 (*lectin mannose-binding 2, dog*) template (63% SI)

- Mutation on opposite side from Vip36 Ca^{2+} and mannose binding site

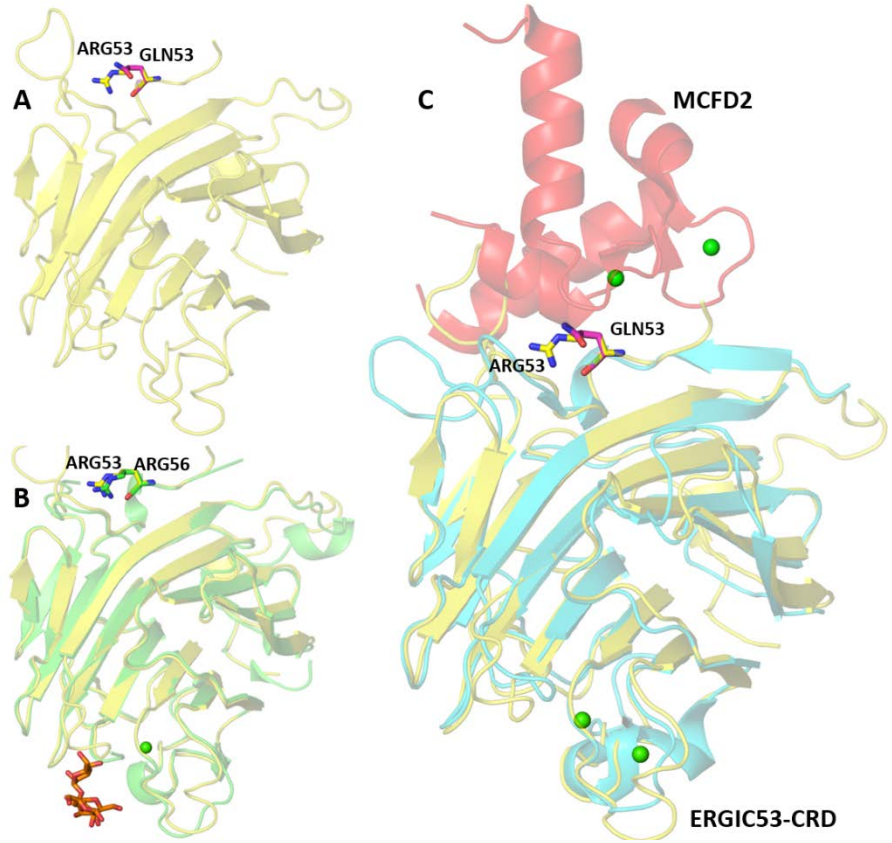


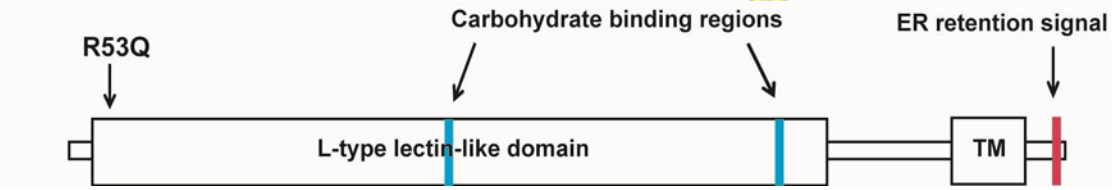
Rafiullah, R., Aslamkhan, M., Paramasivam, N., Thiel, C., Mustafa, G., Wiemann S., Schlesner, M., Wade, R.C., Rappold, G.A., Berkel, S. "*Homozygous missense mutation in the LMAN2L gene segregates with intellectual disability in a large consanguineous Pakistani family*". *J. Med. Genet.* (2016) 53:138-144

Missense mutation : mechanism?

- Superimpose model on ERGIC-53-CRD (*ER Golgi intermediate compartment 53 carbohydrate recognition domain, man*) (35% SI to LMAN2L domain) bound to MCFD2 (*multiple coagulation factor deficiency 2*) to form cargo receptor

- 3D model indicates **R53Q** impairs protein-protein interaction in **LMAN2L**
  - Multimer
  - Unknown protein
  - ERGIC-53





Rafiullah, R., Aslamkhan, M., Paramasivam, N., Thiel, C., Mustafa, G., Wiemann S., Schlesner, M., Wade, R.C., Rappold, G.A., Berkel, S. "*Homozygous missense mutation in the LMAN2L gene segregates with intellectual disability in a large consanguineous Pakistani family*". *J. Med. Genet.* (2016) 53:138-144

Missense mutation : mechanism?

Chain A | Chain B | Chain C | Chain D | Chain E

Chain A

Uniprot Accession	P49256
Name	Vesicular integral-membrane protein VIP36
Enzyme class	
Taxon	Canis lupus
P49256 Sequence	

Positions are given in natural positions.

/- T/- D/- G/- N/- S E H L K R E H S L I K P

16 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63

Caution: The transfer of annotations from similar sequences can be misleading. Lower sequence similarity will result in a lower reliability of the annotations.

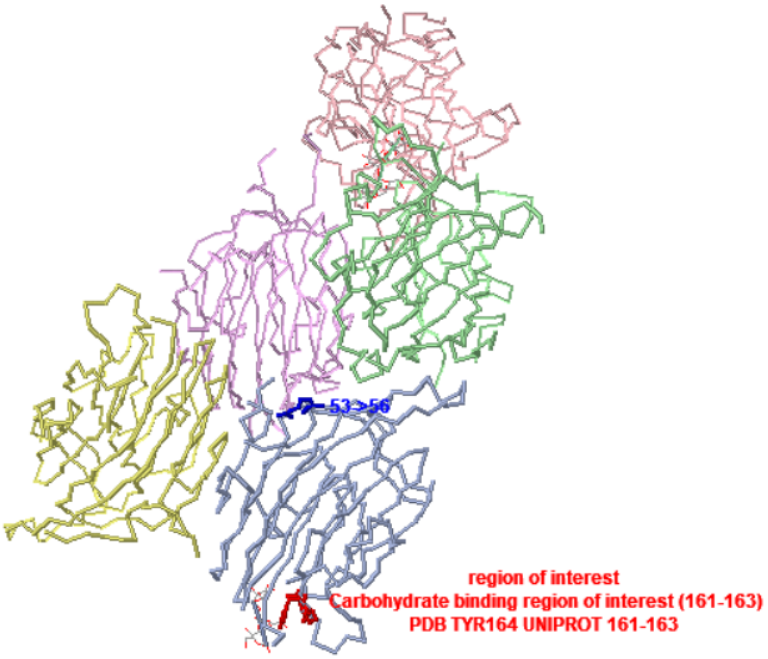
Q9H0V9 Sequence (Homo sapiens)

T/S F/- E Y/H L K R E H S L S/I K P Y Q G V G T/

47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

Sequence Features: Show features panel

Sequence features and sequences taken from Uniprot.



http://prosat.h-its.org/prosat/addfeatures?pdbcode=2E6V&chain=A&submit=&uniprotacc=Q9H0V9&user=P49256:56:56-%3E56:Uniprot%2056%20-%3E%20PDB%2056|Q9H0V9:53:53-%3E56:Uniprot%2053%20-%3E%20PDB%2056|Q9H0V9:258-260:region%20of%20interest%20(258-260):region%20of%20interest,%20Carbohydrate%20binding|

Missense mutation : mechanism?

You searched for UniProt Accession P49257 (LMAN1_HUMAN):

3A4U Protein ERGIC-53

Chain A | Chain B

Chain A

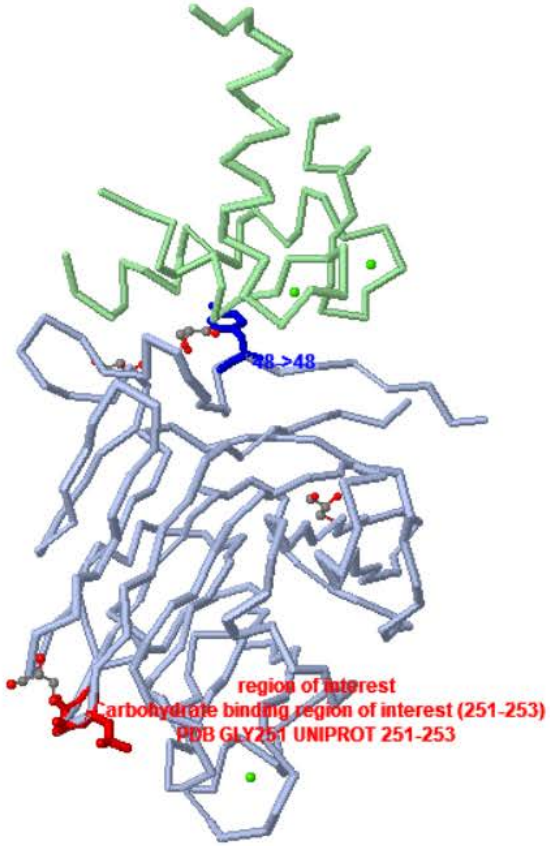
Uniprot Accession	P49257
Name	Protein ERGIC-53
Enzyme class	
Taxon	Homo sapiens
P49257 Sequence	

Positions are given in natural positions.

Similarity: Find similar sequences

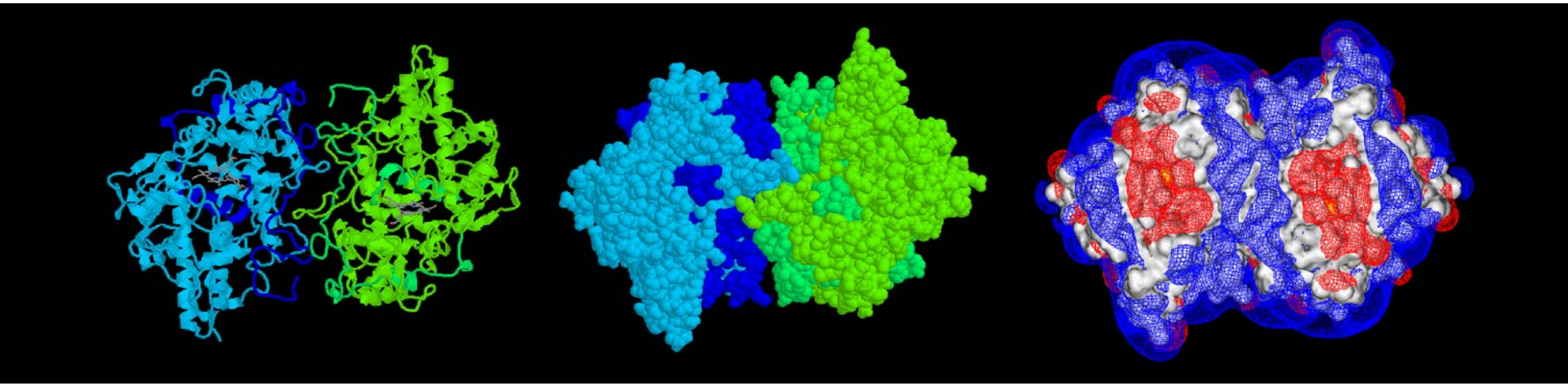
Sequence Features: Show features panel

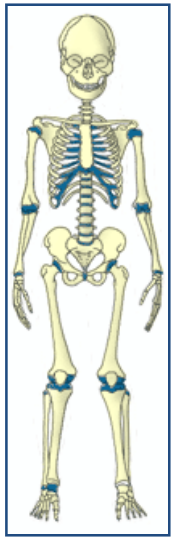
Sequence features and sequences taken from Uniprot.
Structural data retrieved from RCSB.
Mapping data obtained from SIFTS or pairwise alignments using BioJava, additional similar sequences added using Blast.



http://prosat.h-its.org/prosat/prosatexeuniprot?pdbcode=3A4U&submit=Check+PDB&user=P49257:48:48-%3E48:Uniprot%2048%20-%3E%20PDB%2048|P49257:251-253:region%20of%20interest%20(251-253):region%20of%20interest,%20Carbohydrate%20binding|

How do biomolecules recognize each other?



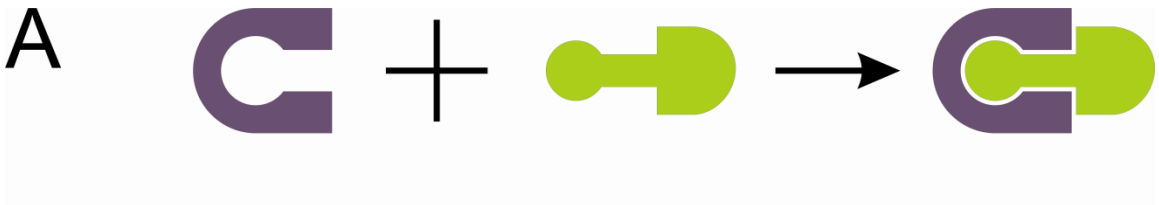


Challenges for predicting receptor-ligand interactions computationally

- **Scoring**
  - Force field, energy function, etc
- **Sampling**
  - Simulation length, space, degrees of freedom, multigraining, etc

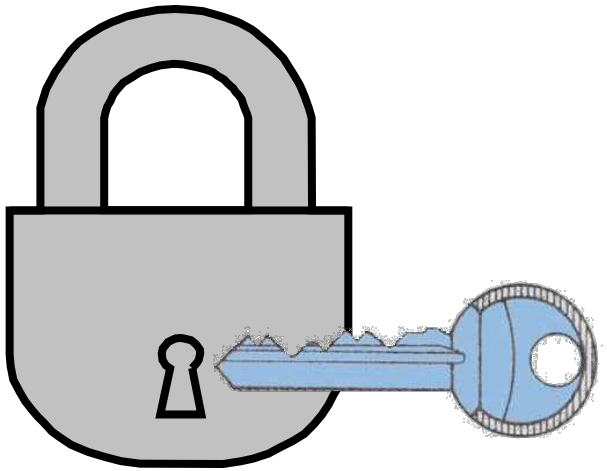
Receptor-ligand binding paradigms

Lock and key

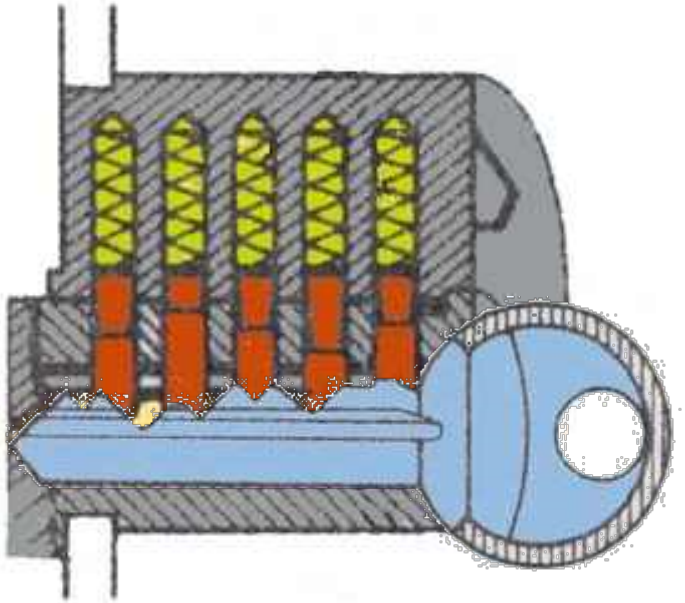


Receptor Ligand

Lock & Key Model



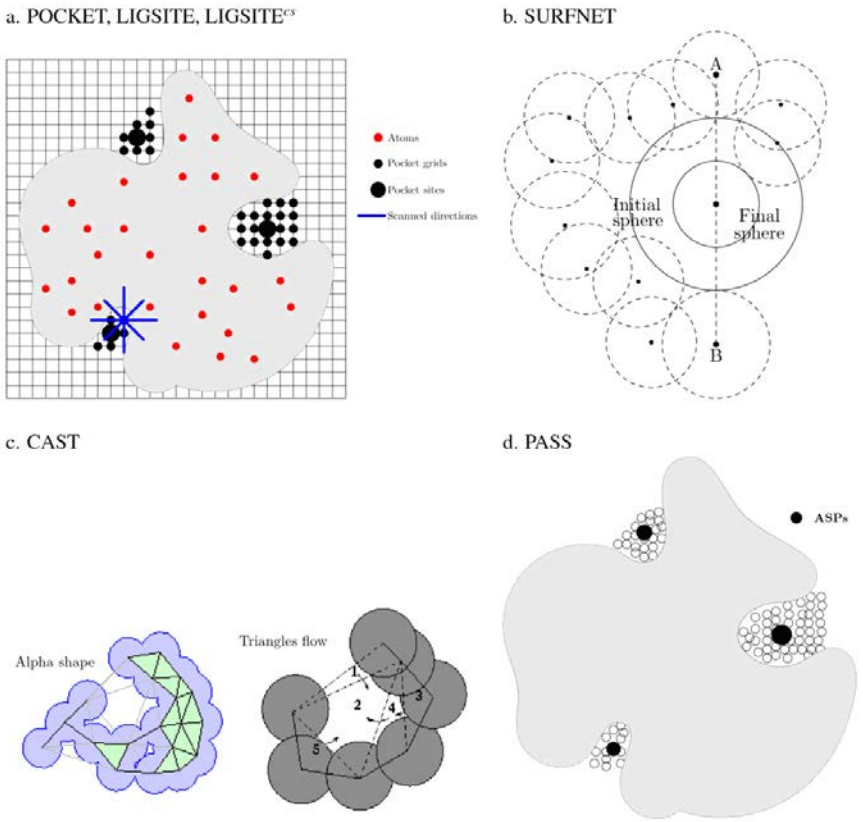
- Fischer, E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme Ber. Deutsch Chem Ges., 27, 2985-2993.
- **„….dass Enzym und Glucosid wie Schloss und Schlüssel zu einander passen müssen….“**



- Laskowski & Thornton (1995): SURFNET: detect 80% of enzyme active sites by looking for crevices on proteins

MetaPocket

- MetaPocket 2.0 **projects.biotec.tu-dresden.de/metapocket/**
  - Consensus method to predict binding sites on protein surfaces
  - 8 methods
    - LIGSITE^{CS}
    - PASS
    - Q-SiteFinder
    - SURFNET
    - Fpocket
    - GHECOM
    - ConCavity
    - POCASA



Huang, B. 'Omics, 2009, 13(4):325-330.

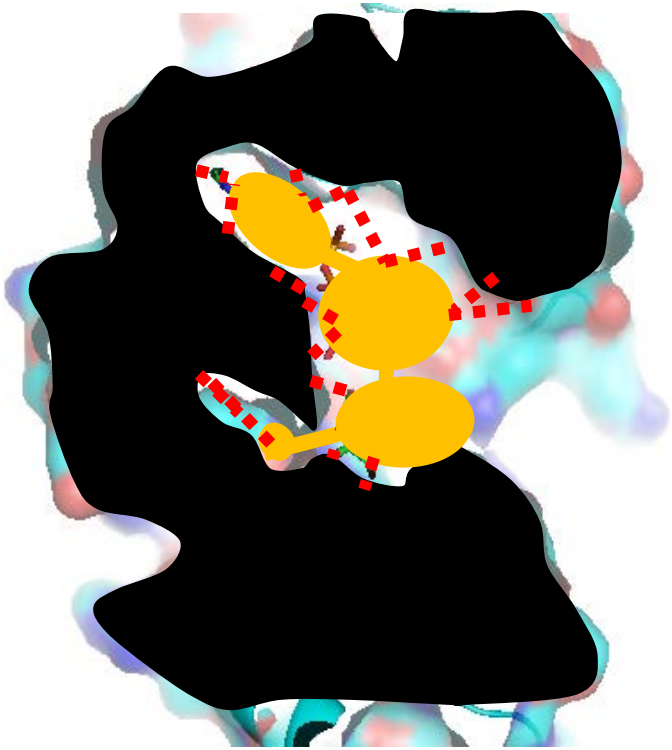
MetaPocket

- MetaPocket 2.0 http://projects.biotec.tu-dresden.de/metapocket/
  - Consensus method to predict binding sites on protein surfaces
  - 8 methods (MPK2) (MPK1: 4 methods)

Dataset	Version	Top1	Top2	Top3	Top4
48(bound)	MPK2	85	92	96	96
	MPK1	83	94	96	96
48(unbound)	MPK2	80	90	94	96
	MPK1	75	85	90	92
210(bound)	MPK2	81	91	95	96
	MPK1	76	89	94	96
Drug-Target 198(bound)	MPK2	61	70	74	76
	MPK1	55	65	68	72

Huang, B. 'Omics, 2009, 13(4):325-330.

Structure-based drug design: the lock & key paradigm

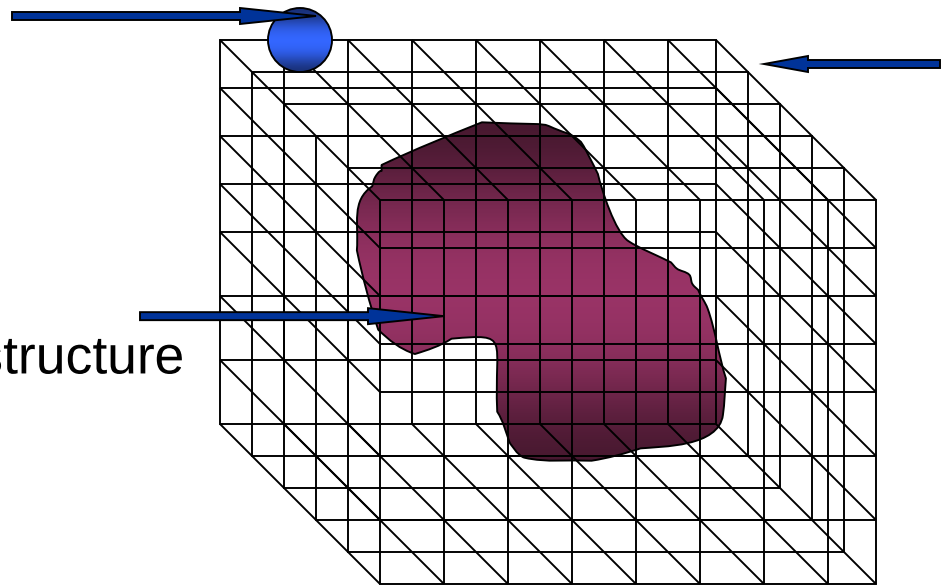


A good compound must:

- Bind strongly
  - many favourable contacts, shape and chemical complementarity
- Bind selectively
  - complex shape, etc

GRID Molecular Interaction Fields

Chemical probe,
E.g. water, amino group, proton



Virtual **grid**

Target,
E.g. protein structure

Detect energetically favorable binding sites on molecules of known structure

$$\Delta E = \sum_i E_{LJ} + \sum_i E_{EL} + \sum_i E_{HB} + S$$

Goodford, PJ *J. Med. Chem.* (1985) 28, 849-857.
Boobbyer et al, *JMC.* 1989, Wade et al, *JMC.* 1993

Structure-based drug design : Influenza

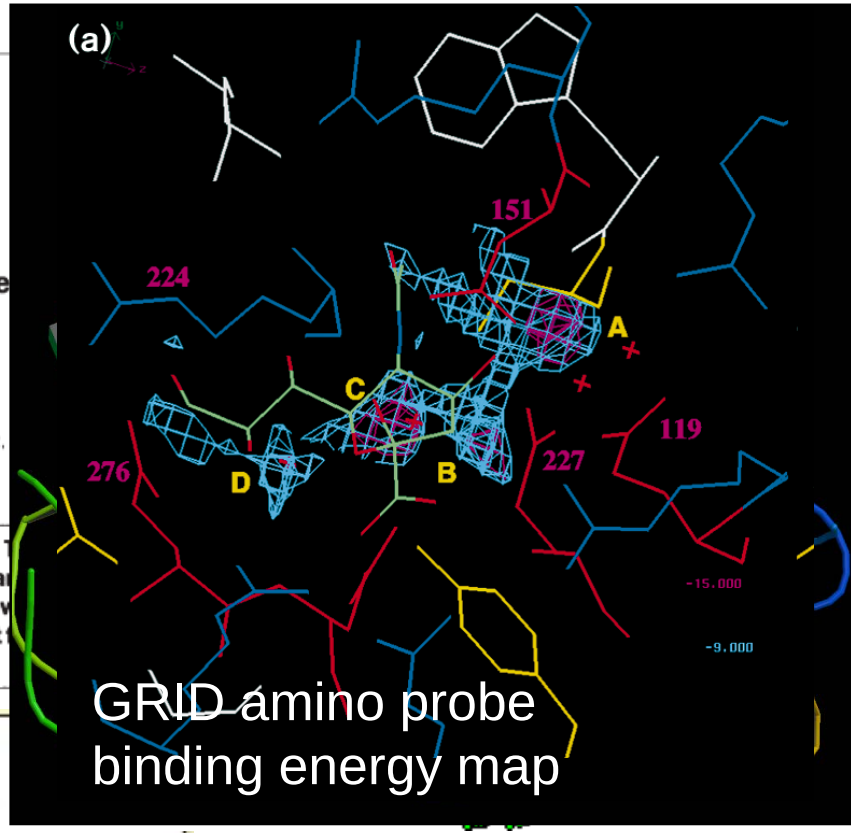
ARTICLES *Nature.* 1993 363:418-23.

Rational design of potent sialidase-based inhibitors of influenza virus replication

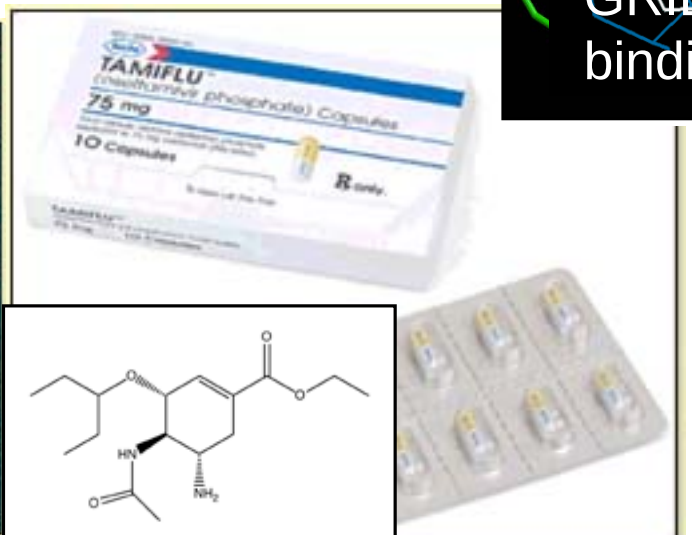
Mark von Itzstein*, Wen-Yang Wu*, Gaik B. Kok*, Michael S. Pegg*, Jeffrey C. Dyason*, Betty Jin*, Tho Van Phan *, Mark L. Smythe*, Hume F. White Stuart W. Oliver*, Peter M. Colman‡, Joseph N. Varghese‡, D. Michael Ryan∮, Jacqueline M. Woods∮, Richard C. Bethell∮, Vanessa J. Hotham∮, Janet M. Cameron∮ & Charles R. Penn∮

*Department of Pharmaceutical Chemistry, Victorian College of Pharmacy, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia
‡ CSIRO Division of Biomolecular Engineering, Parkville, Victoria 3052, Australia
∮ Glaxo Group Research Ltd, Greenford, Middlesex UB6 OHE, UK

Two potent inhibitors based on the crystal structure of influenza virus sialidase have been designed. T compounds are effective inhibitors not only of the enzyme, but also of the virus in cell culture and in a models. The results provide an example of the power of rational, computer-assisted drug design, as v as indicating significant progress in the development of a new therapeutic or prophylactic treatment f influenza infection.



GRID amino probe binding energy map



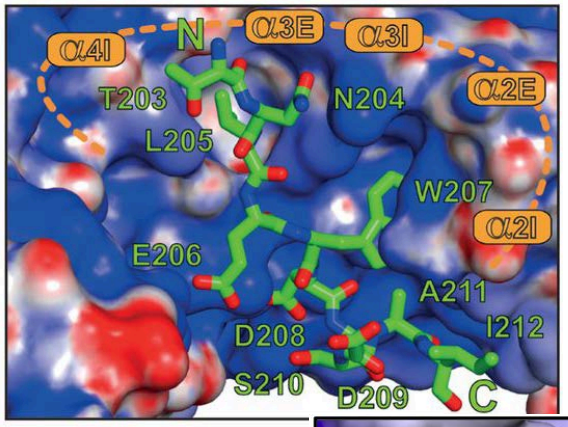
Review:

SBDD against influenza: Wade, *Structure*, 1997, 5, 1139-1145

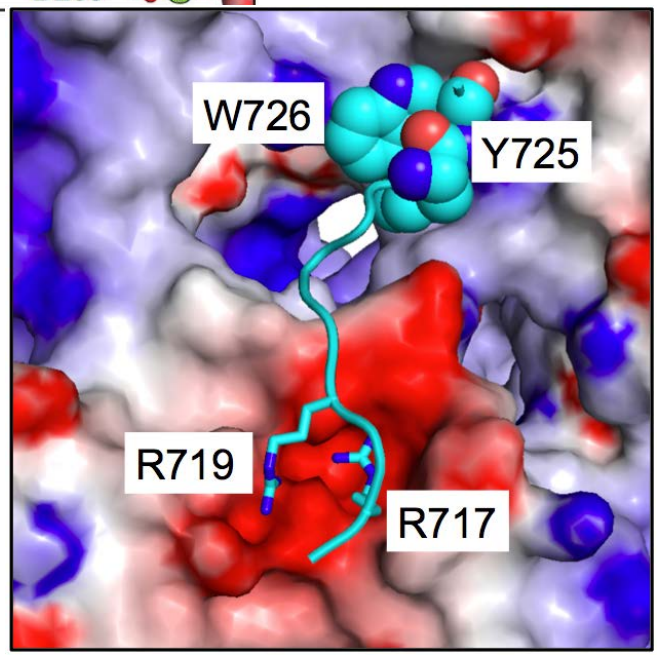
Protein electrostatic properties

…in *Science:*

C

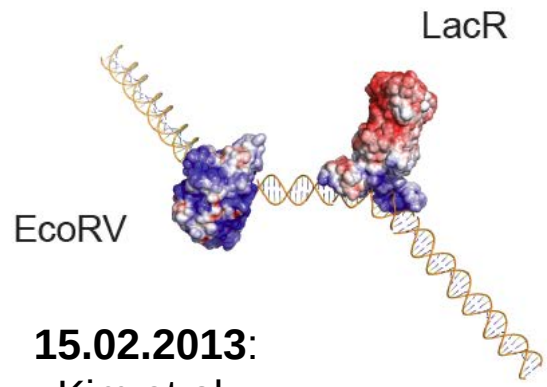


19.04.2013:
Pernigo et al,
(2013) 340, 356



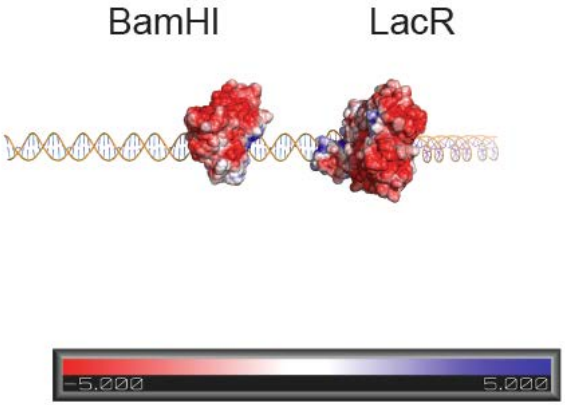
31.05.2013:
Kato et al,
(2013) 340, 1110

A



15.02.2013:
Kim et al,
(2013) 339, 816

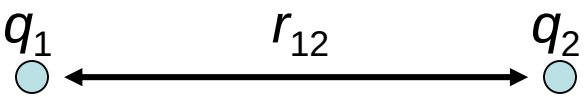
B



Coulomb's Law - Energy

- Interaction energy of two point-charges *in vacuo*
- Solve Poisson equation
- In SI units:

$$U = \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}}$$



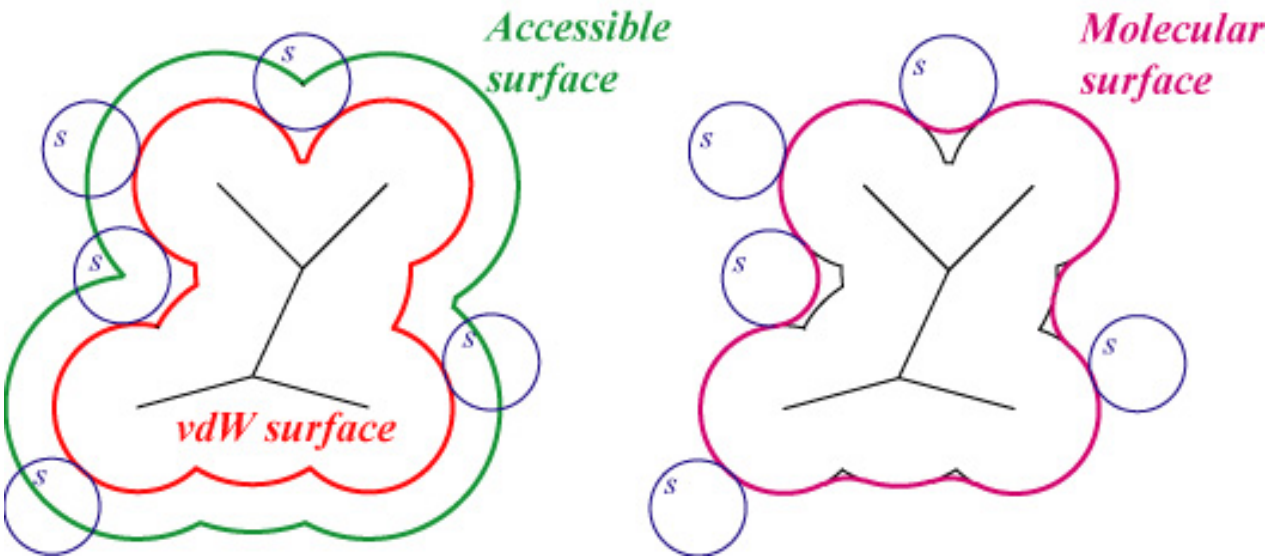
- In "biomolecular" units:

$$U = \frac{332 q_1 q_2}{r_{12}}$$

Energy, U: kcal/mol
Charge, q: electron charges
Distance, r: Angstroms

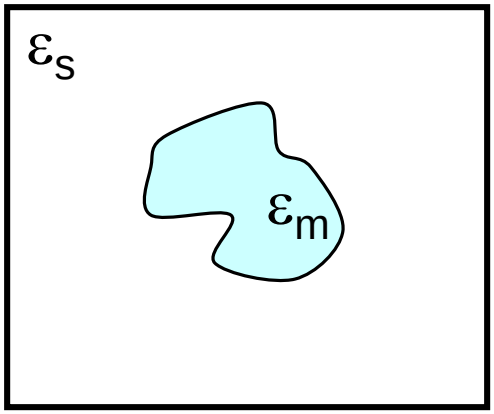
- Electrostatic potential: $\phi(\mathbf{r_2}) = 332 q_1 / r_{12}$

Continuum electrostatics for molecules



http://csb.stanford.edu/koehl/ProShape/

Solute Molecule: ε_m ~2-4
Solvent (water) : ε_s ~80



Mobile ions in the solvent

- Ionic solution with dissolved ions (electrolyte)
  - Ions redistribute in the presence of a molecule with charges to weaken/screen its electrostatic interactions

- Debye-Hueckel theory
  - Implicit model of the ions in the solvent
  - Ions assumed to distribute according to the local potential with a Boltzmann factor

$$c_{ion}(\mathbf{r}) = c_{ion,bulk} e^{-\beta\phi(\mathbf{r})q_{ion}}$$

Continuum Electrostatics

- Poisson-Boltzmann equation

$$-\varepsilon_0 \nabla.\left[\varepsilon_r(\mathbf{r})\nabla\phi(\mathbf{r})\right] = \rho^f(r) + \sum_{i=1}^{N} q_i c_{i,bulk}(\mathbf{r}) e^{-\beta\phi(\mathbf{r})q_i}$$

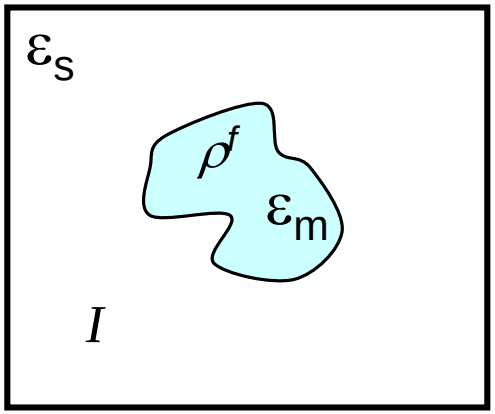
- Linearized Poisson-Boltzmann equation (weak ϕ, low I)

$$-\varepsilon_0 \nabla.\left[\varepsilon_r(\mathbf{r})\nabla\phi(\mathbf{r})\right] = \rho^f(r) - \varepsilon_0\varepsilon_r(r)\kappa^2(r)\phi(r)$$

$$\kappa^2(r) = \frac{\beta}{\varepsilon_0\varepsilon_r}\sum_{1}^{N} c_{i,bulk} q_i^2 = \frac{2e^2 N_A}{\varepsilon_0\varepsilon_r kT} I$$

Debye length = $1/\kappa$
= $3.04/I^{1/2}$ Å at 300K
= 8 Å at I=150mM

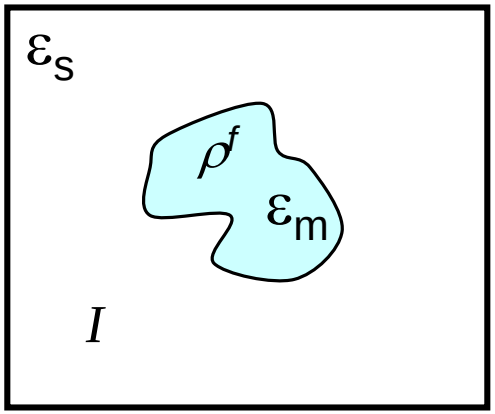
- Analytical solution only for simple shapes
- Numerical solution
- Simple approximations, e.g. $\varepsilon_r = K\, r_{ij}$



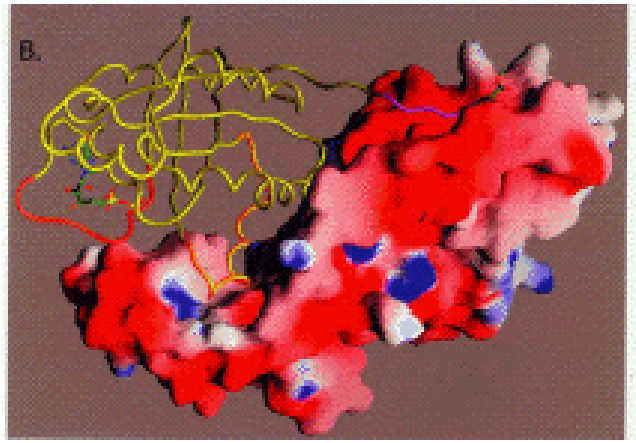
Solving the Poisson-Boltzmann equation for biomolecules

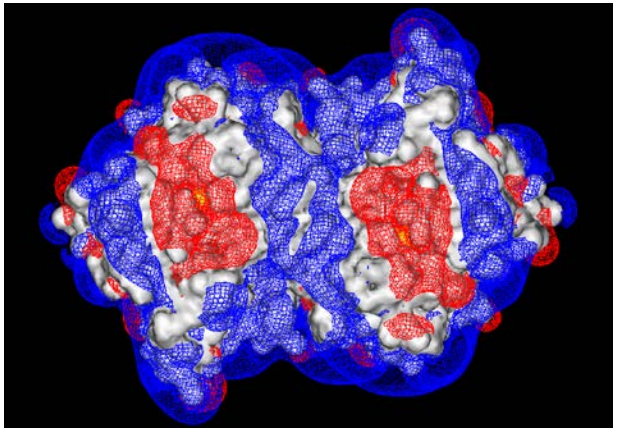
- Need to assign:
  - Atomic partial charges
  - Atomic radii
  - Dielectric constants (solute and solvent)
  - Ionic strength
  - Ion exclusion layer etc

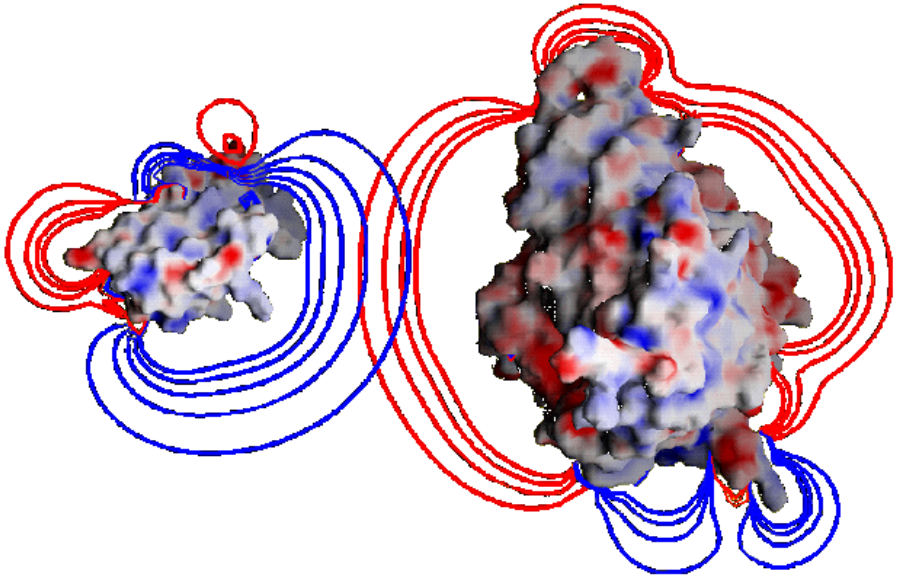
  - Linear vs non-linear
  - Solution method (FD, FE etc)
  - Convergence

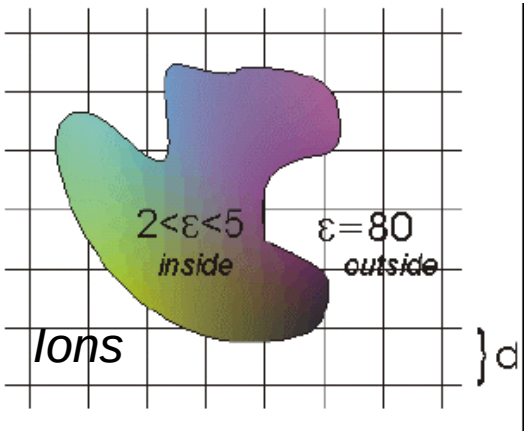


Protein electrostatic potentials





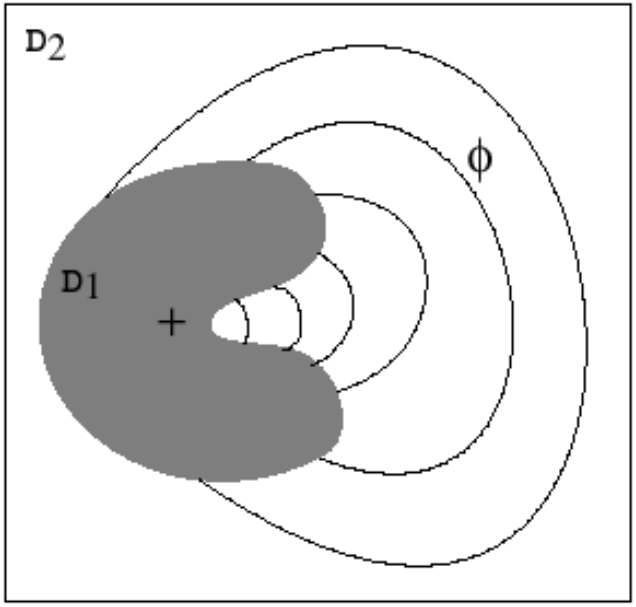


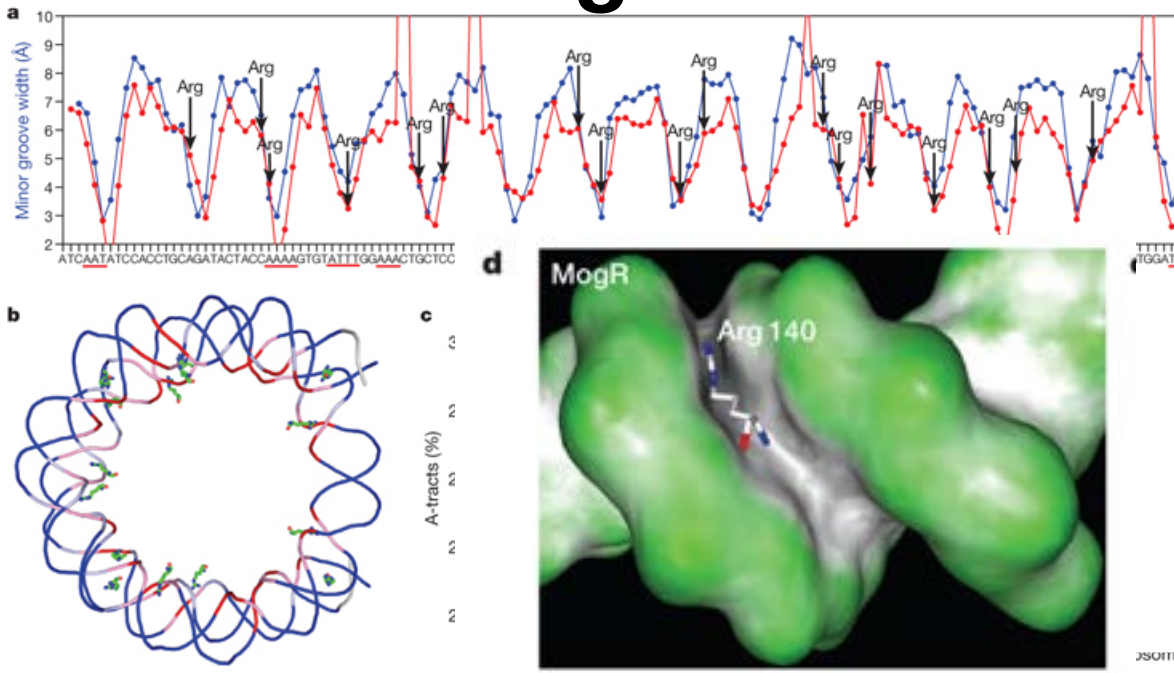


$$\Phi(r) = \frac{q}{4\pi\varepsilon_0 \varepsilon r}$$

Electrostatic focusing

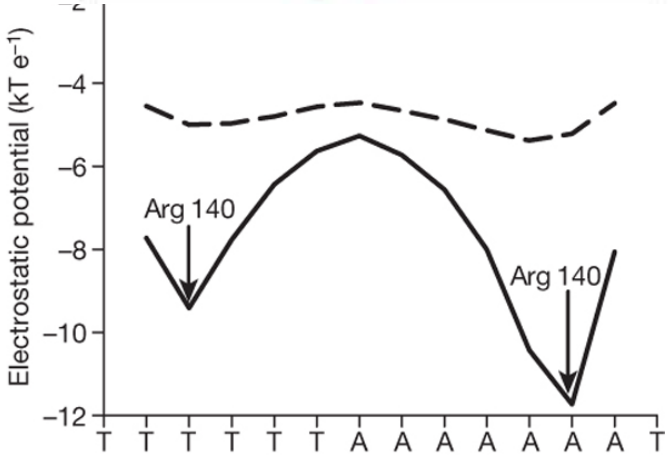
- Due to total dielectric environment
- Non-spherical isocontours around charges





$\varepsilon_m = \varepsilon_s = 80$

$\varepsilon_m = 2;\ \varepsilon_s = 80$

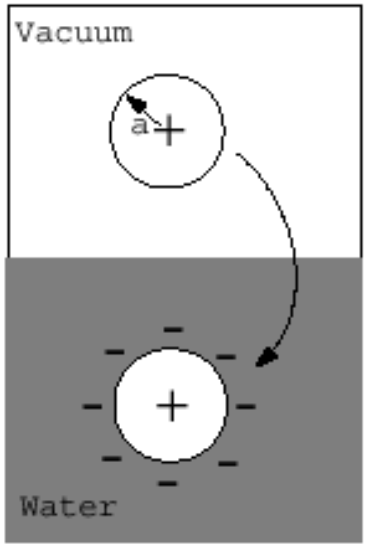


Minor-groove shape recognition by arginines.

R Rohs *et al. Nature* **461**, 1248-1253 (2009) doi:10.1038/nature08473

Electrostatic solvation

- In a high dielectric solvent (e.g. water), charges will tend to be repelled from low dielectric solutes
- Charge polarizes solvent, which produces *reaction field* at charge with which charge interacts



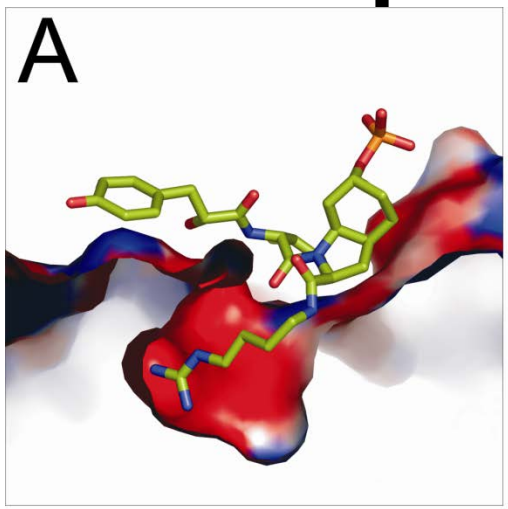
- **Born ion solvation**
  - Work done to transfer Born ion between 2 dielectrics
  - Born ion is point charge in spherical cavity

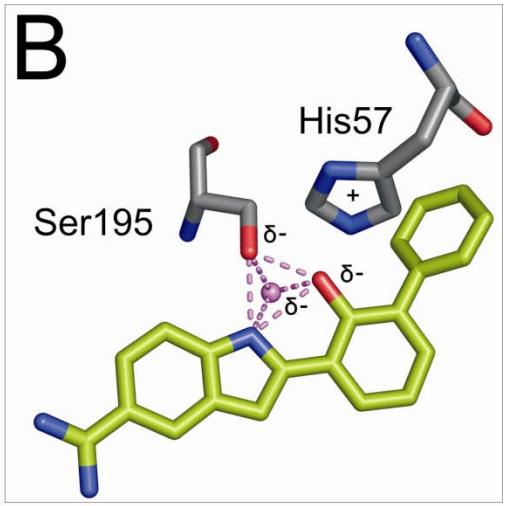
$$U_{Born} = \frac{332q^2}{2a}\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right)$$

  - E.g. Na+, K+, Cl- in water: a ~ 1.5-2.5 Å
  - Free energy of hydration ~ -100- -50 kcal/mol

Electrostatics in ligand-receptor complexes

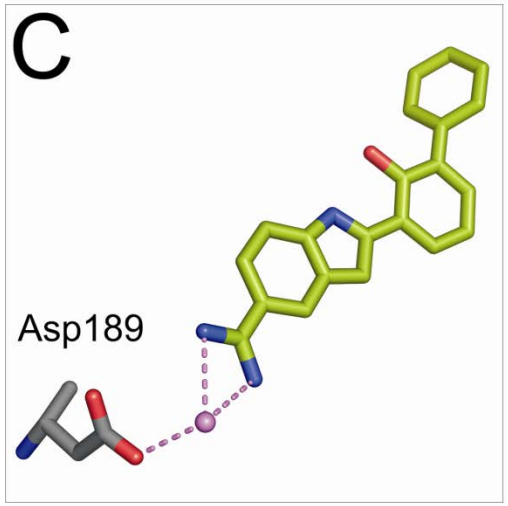
Inhibitor Aeruginosin 98-B
bound to negatively charged pocket of trypsin

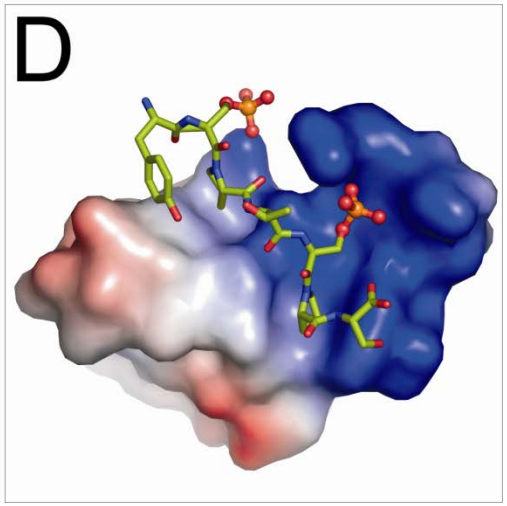




Water mediated short H-bond network in Urokinase-Type Plasminogen activator-inhibitor complex

Water mediated salt-bridge between Urokinase-Type Plasminogen activator Asp189 and inhibitor

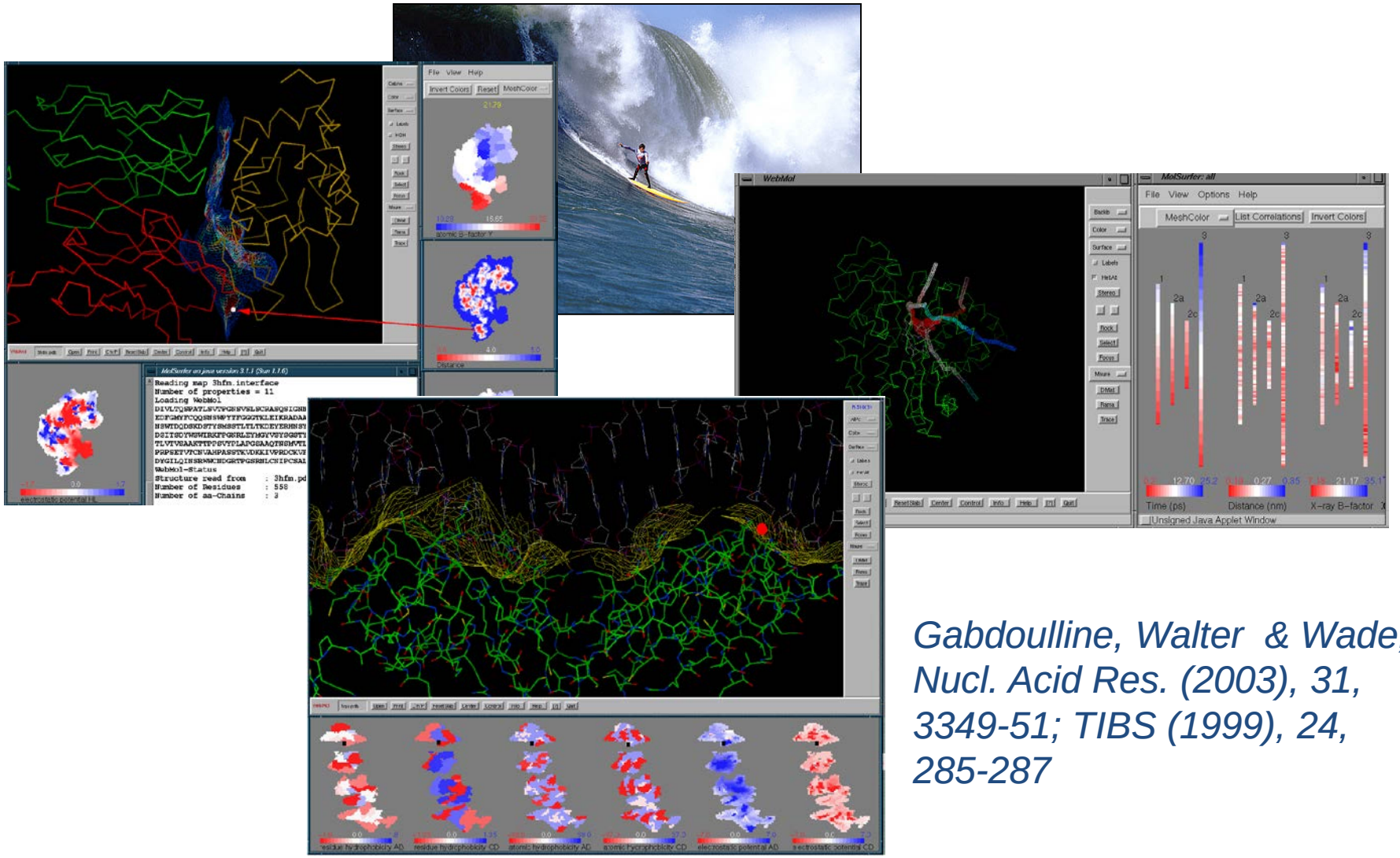




Phosphoserine-proline containing peptide bound to group IV WW domain area with positive electrostatic potential

Motiejunas & Wade, Comprehensive Med. Chem. II, 2007, 193-213

Molsurfer: a Macromolecular Interface Navigator

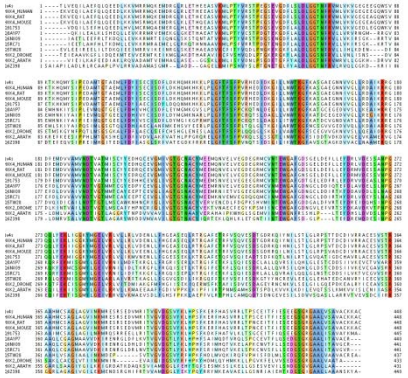


Gabdoulline, Walter & Wade, Nucl. Acid Res. (2003), 31, 3349-51; TIBS (1999), 24, 285-287

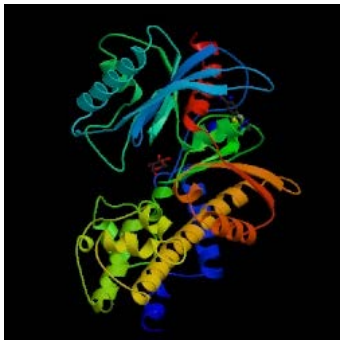
http://projects.h-its.org/dbase/molsurfer/index.html

Levels of Protein Comparison

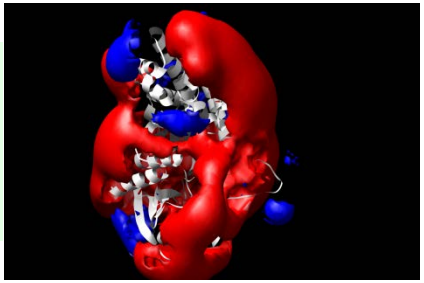
- Amino Acid Sequence Identity



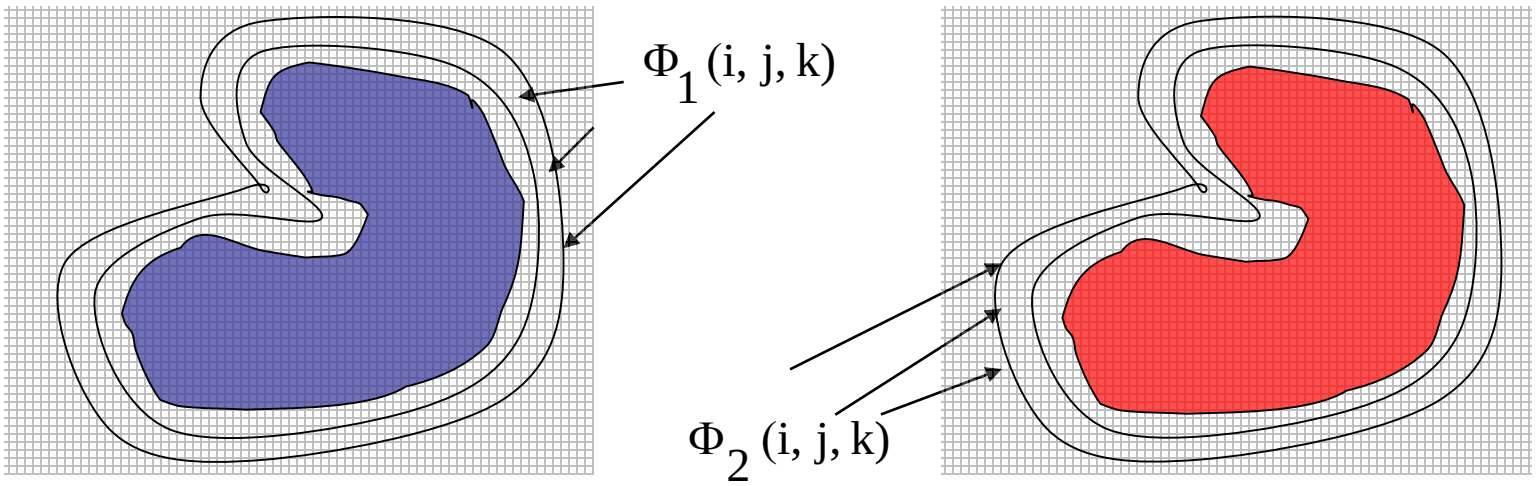
- Protein Structure (NMR, X-Ray)



- Protein Structure/Function Relationship - MIF



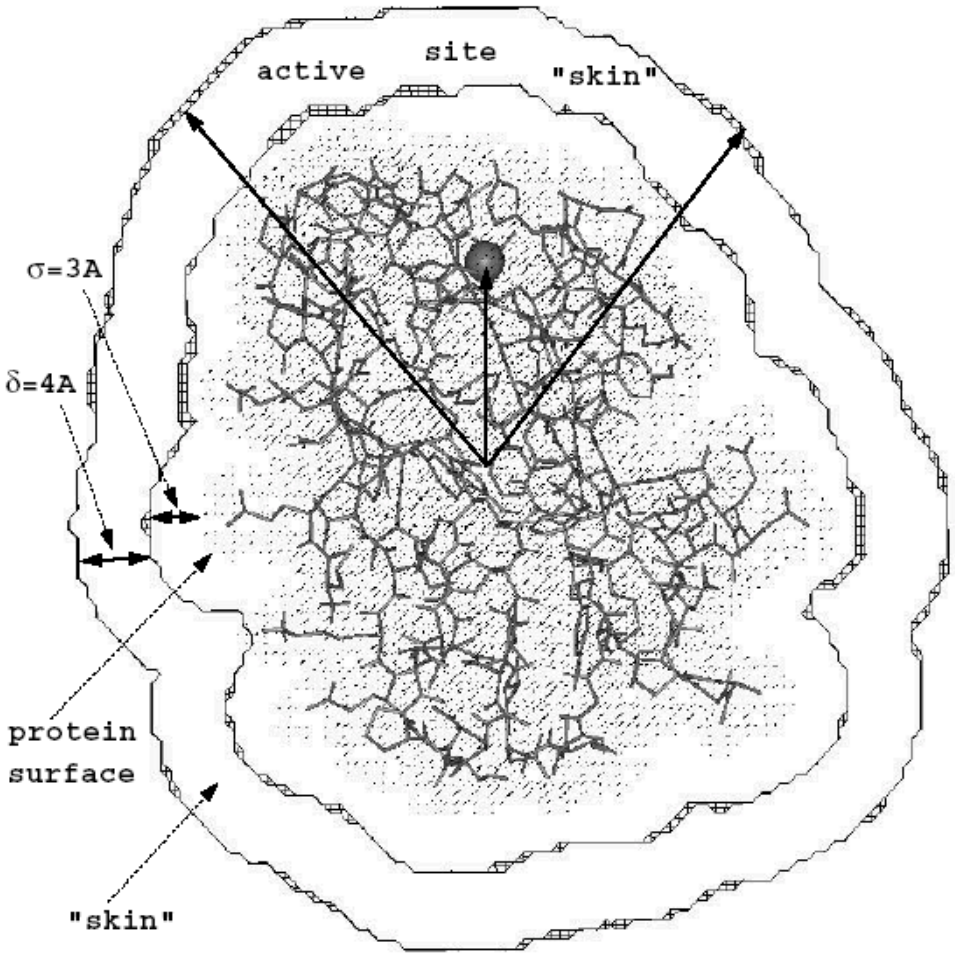
PIPSA:
Protein Interaction Property Similarity Analysis



Protein 1 **Protein 2**

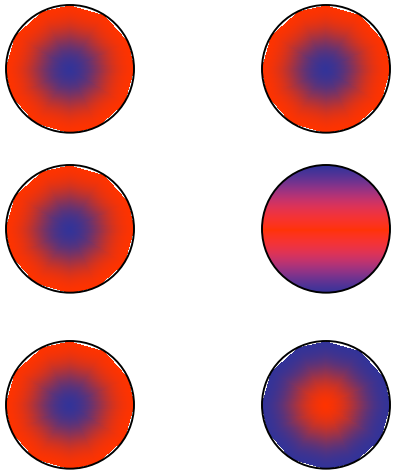
- Interaction fields are calculated on a set of points
- Field values on corresponding points are compared
- Φ = electrostatic potential, shape, probe interaction field, ...

PIPSA: Protein Interaction Property Similarity Analysis



$$SI_{12} = \frac{2(\mathbf{p}_1, \mathbf{p}_2)}{(\mathbf{p}_1, \mathbf{p}_1) + (\mathbf{p}_2, \mathbf{p}_2)}$$

$$(\mathbf{p}_1, \mathbf{p}_2) = \sum_{i,j,k} \phi_1(i, j, k)\phi_2(i, j, k)$$



S = +1

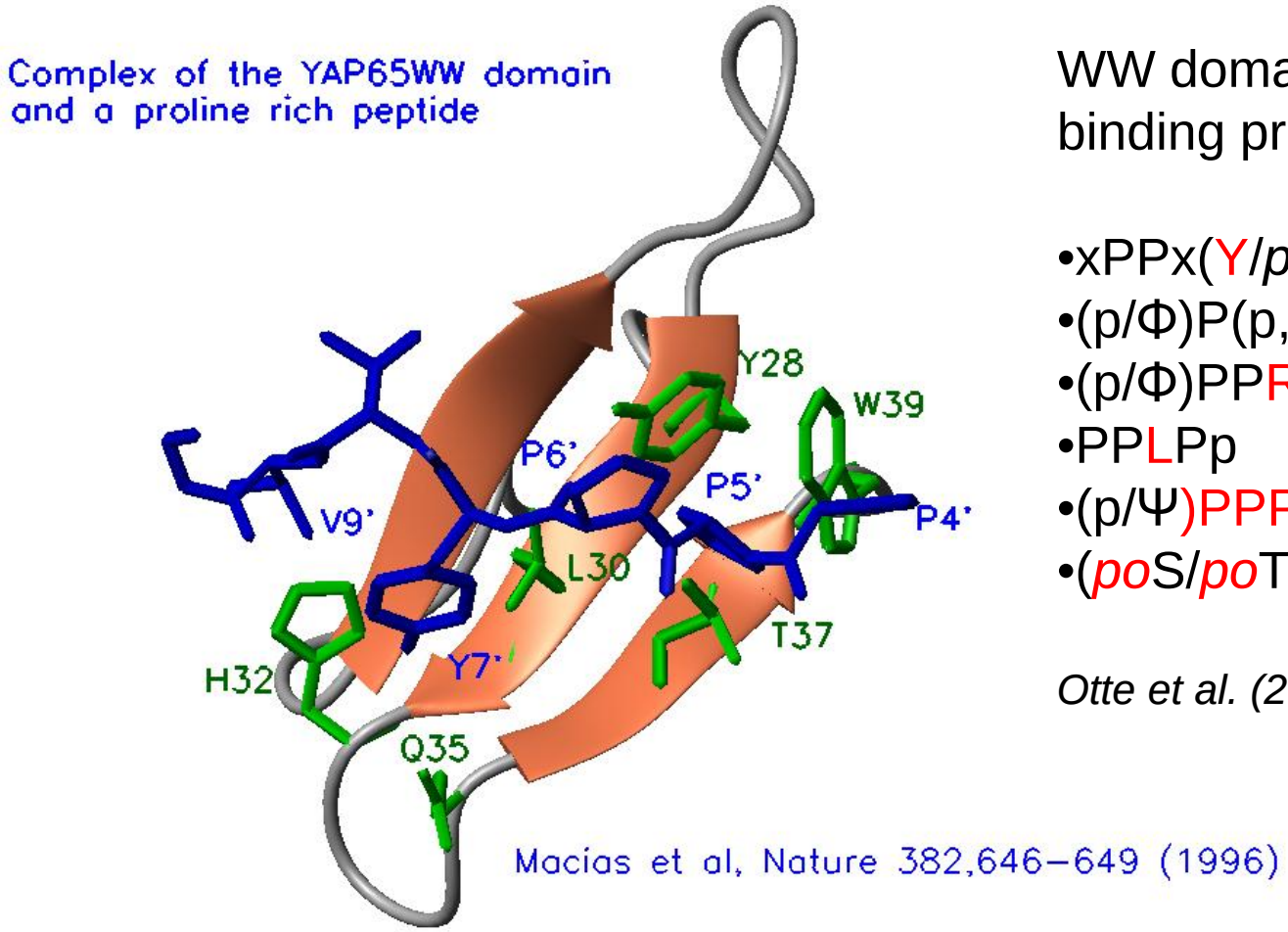
S = 0

S = -1

Wade et al., PNAS, 1998; Blomberg et al. Proteins 1999; De Rienzo et al. Protein Sci. 2000; Wade et al. Intl. J. Quant. Chem. 2001

WW domain/peptide complexes

Binding specificity and affinity determinants?



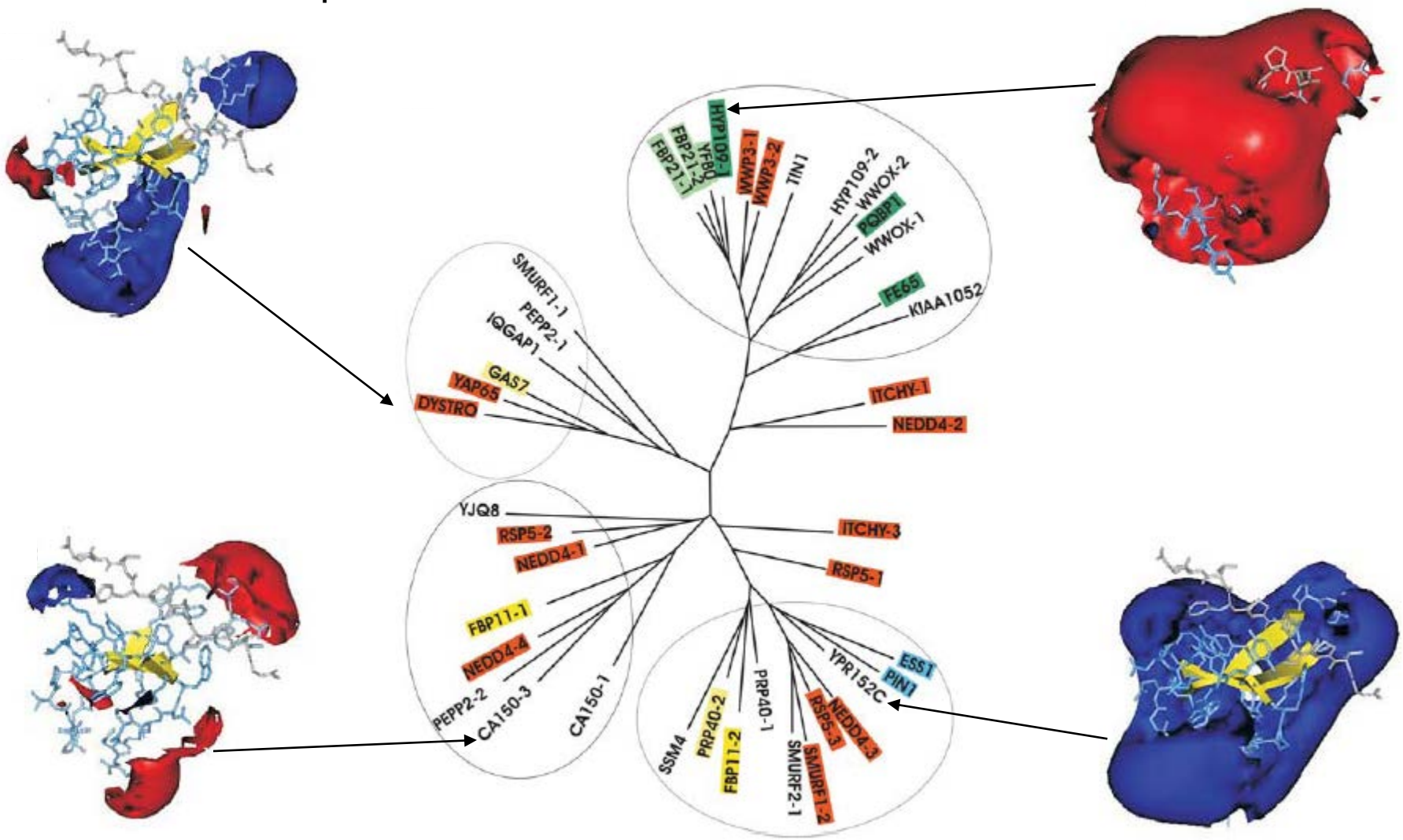
WW domain peptide binding preferences:

- xPPx(Y/*po*Y)
- (p/Φ)P(p,g)PpR
- (p/Φ)PPRgpPp
- PPLPp
- (p/Ψ)PPPPP
- (*po*S/*po*T)P

Otte et al. (2003) Protein Sci. 12, 491

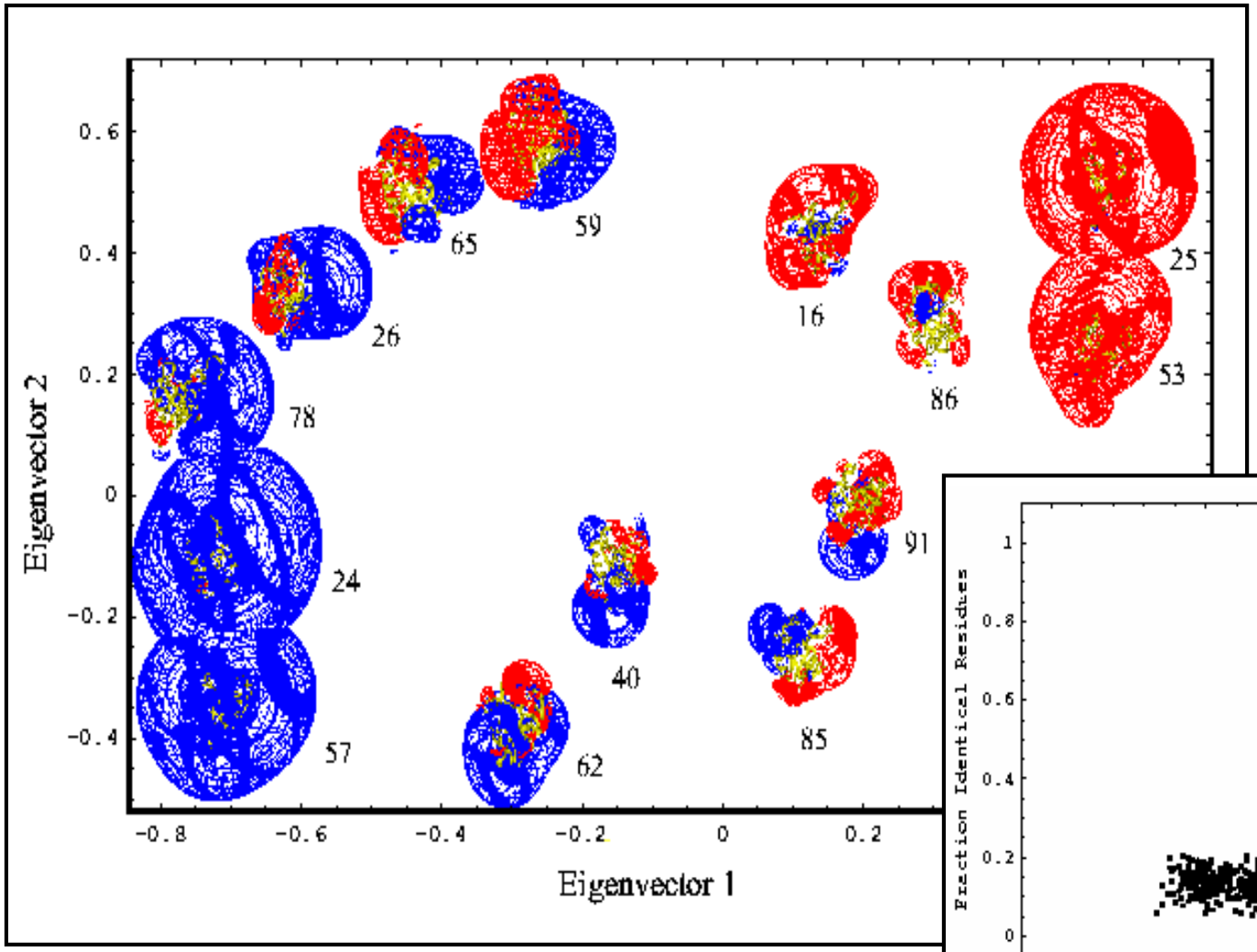
42 WW Domains:
PIPSA epogram for Molecular electrostatic potential

isopotential contours: -0.4 / +0.4 kcal/mol/e

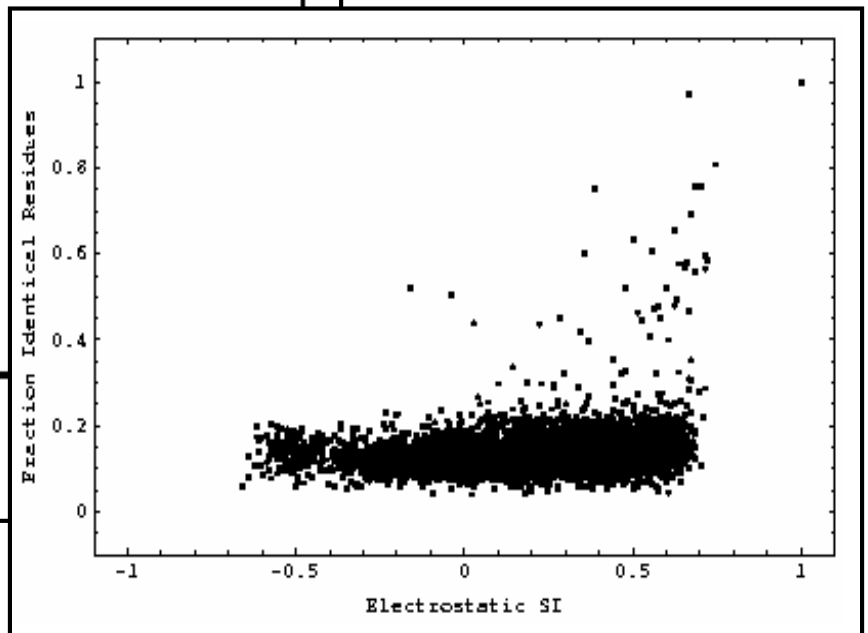


Schleinkofer, Wiedemann et al, JMB (2004) 344, 865-881

PH domains - distribution in electrostatic potential similarity space



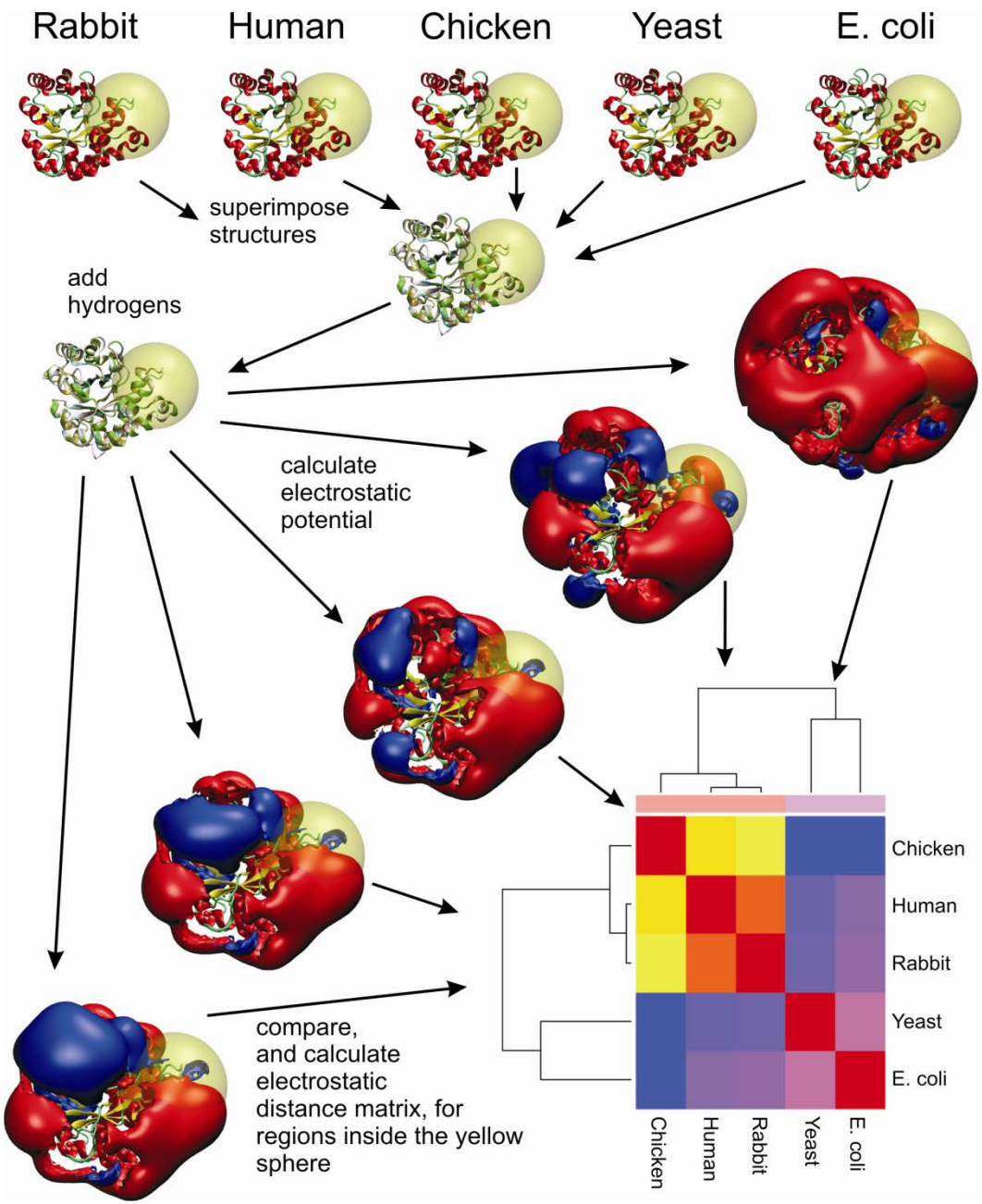
Distribution for DH-linked and internal PH repeat domains



webPIPSA

webPIPSA:
pipsa.h-its.org

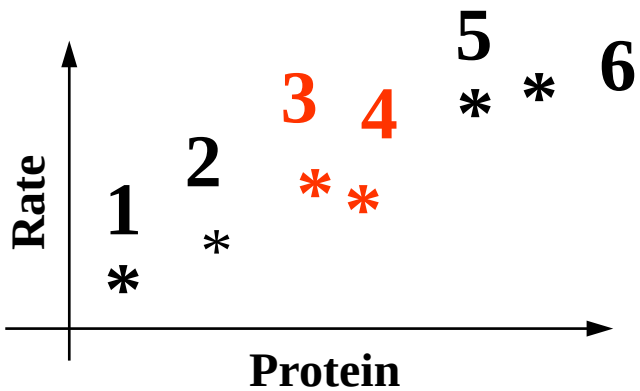
triosephosphate
isomerase



Richter et al.,
Nucleic Acids Research, 2008

quantitative PIPSA (qPIPSA)

- Compare Molecular Interaction Fields
- Quantify similarities and differences
- Training set required with experimental information
- Predict relative ordering and trends

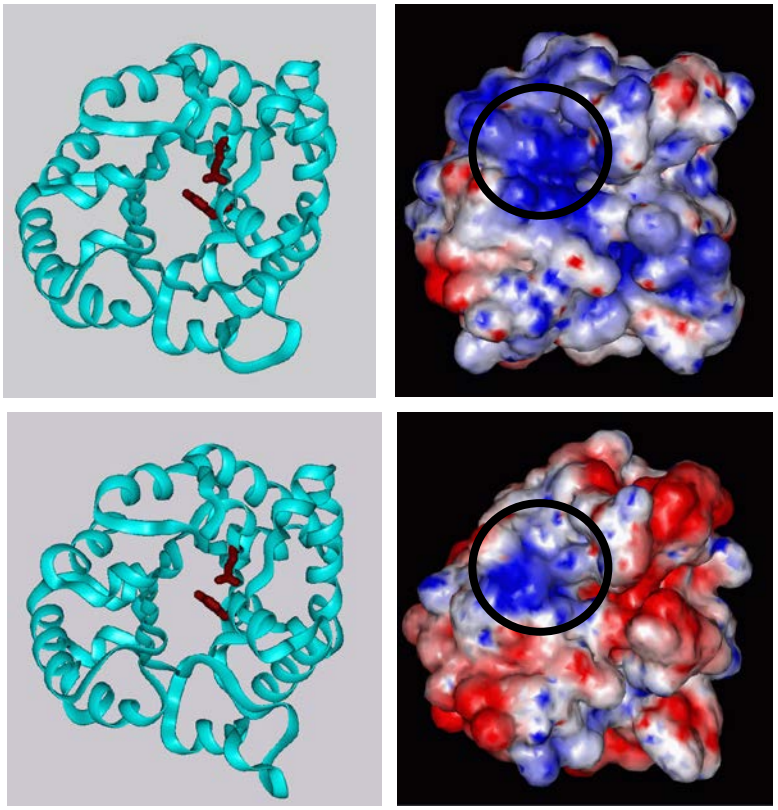


4

Gabdoulline, Stein, Wade, (2007) BMC Bioinformatics 8, 373

Triose Phosphate Isomerase

T. brucei



V.marinus

- 40/55% sequence identity/homology
- same fold
- very similar active site
- factor of 3 difference in kcat/Km

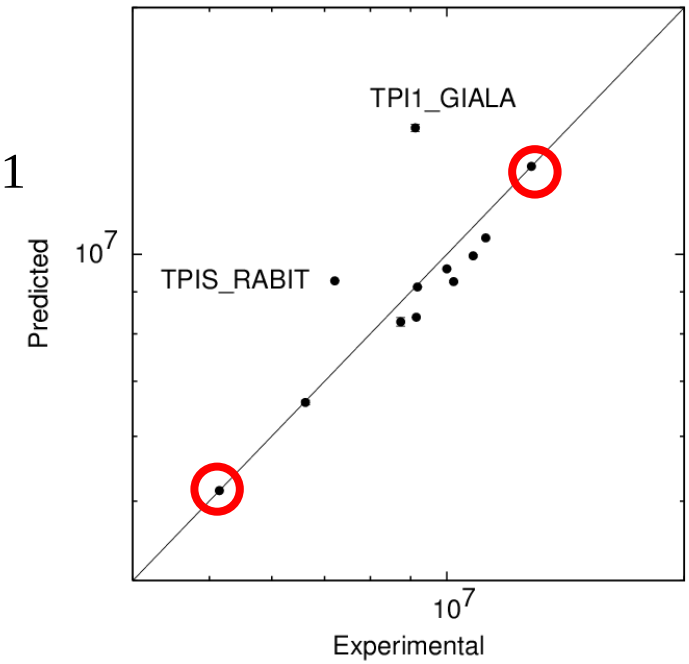
12 species:

Giardia lambli
Spinach
Chicken
E. coli
Human
L. mexicana
P. falciparum
Rabbit
T. brucei
T. cruzi
V. marinus
Yeast

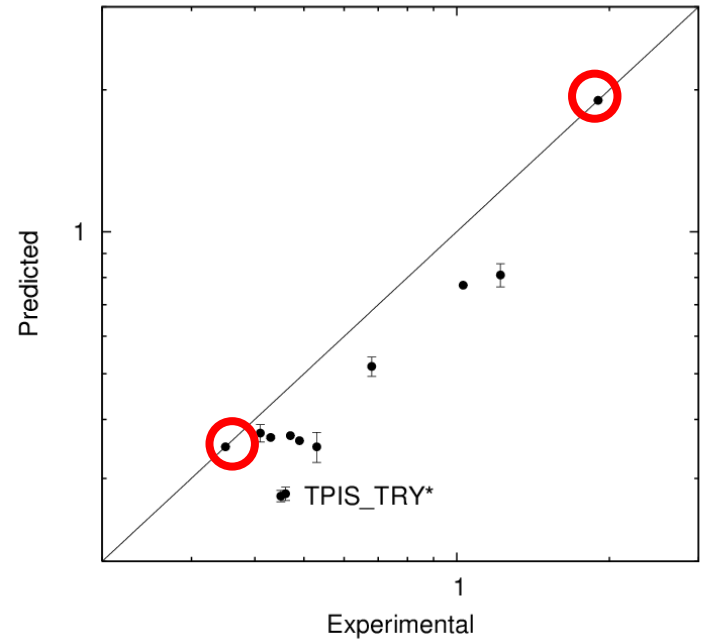
Triose Phosphate Isomerase

$$\ln(k_a / k_b) \sim \alpha \cdot \sum_{\mathbf{R}} (\Phi_a - \Phi_b) / \sum_{\mathbf{R}} 1$$

k_{cat}/K_m ($M^{-1}s^{-1}$)



K_m (mM)



Predictions for 10 TPIs for the substrate glyceraldehyde-3-phosphate based on experimental measurements for the two TPIs from *V. marinus* (TPIS_VIBMA) and *P. falciparum* (TPIS_PLAFA)

1 ln unit increase is related to ca. 1.59 kcal/mol/e increase of av. elec. pot.

1 ln unit decrease is related to ca. 0.85 kcal/mol/e increase of av. elec. pot.

Gabdoulline, Stein, Wade, (2007) BMC Bioinformatics 8, 373

SYCAMORE: Systems biology's Computational Analysis and Modeling Research Environment

sycamore.h-its.org

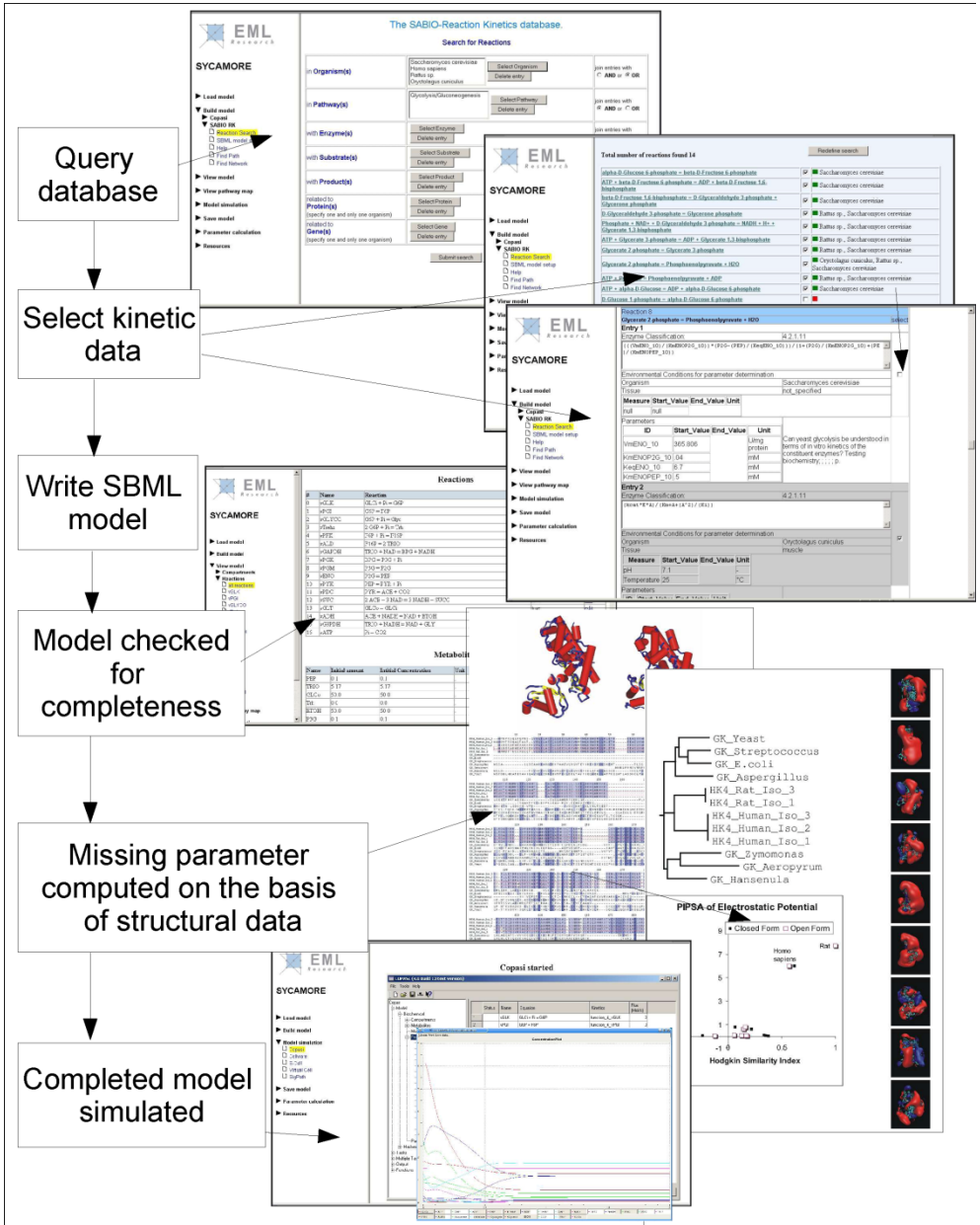


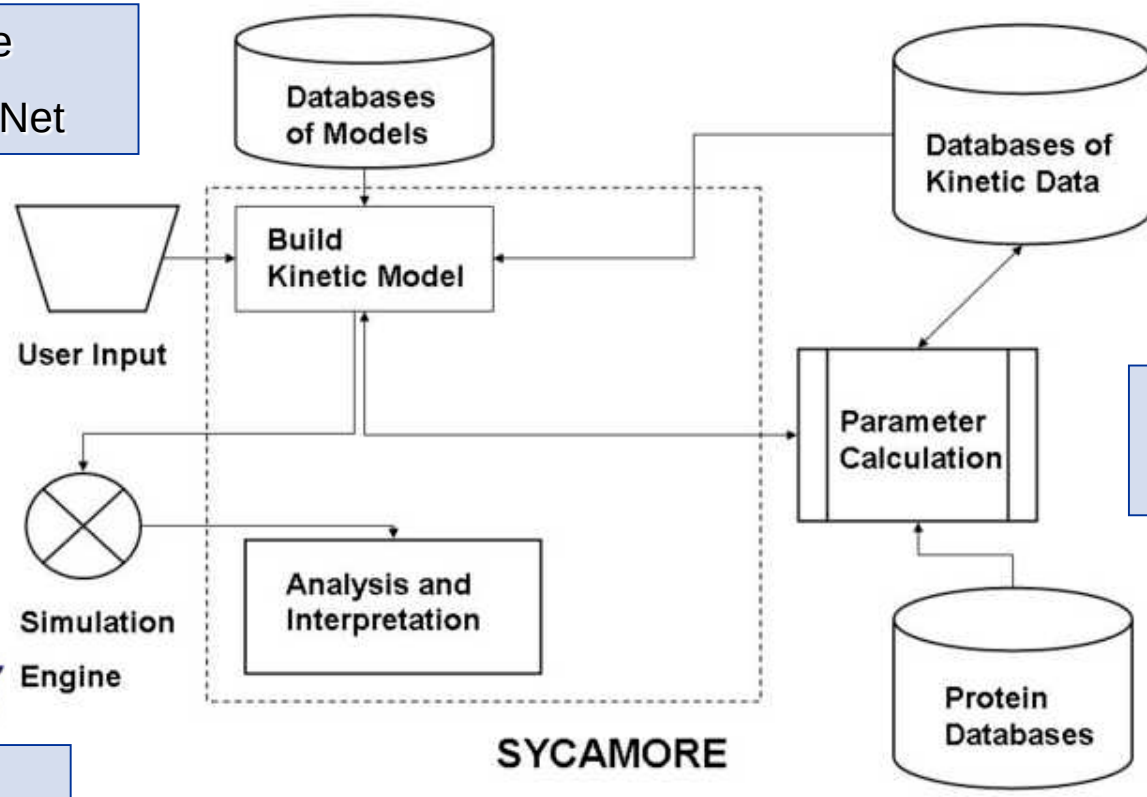
Fig 1: Example of using the available features in SYCAMORE. The case study is modeling and simulating glycolysis in hepatocytes. First, the database is queried and the relevant kinetic data selected. Then the SBML model file is created. This is checked for completeness (this is not implemented for automatic use yet). A missing parameter (here we asume Km for glucokinase to be missing) is then computed using structural data. Finally, the completed model is simulated.

Weidemann, Richter et al (2008) Bioinformatics

SYCAMORE –Systems Biology's Computational and Analysis Modelling Research Environment

JWS Online

Biomodels.Net



BRENDA

SABIO-RK

qPIPSA

etc.

PDB

MODBASE

SwissModel

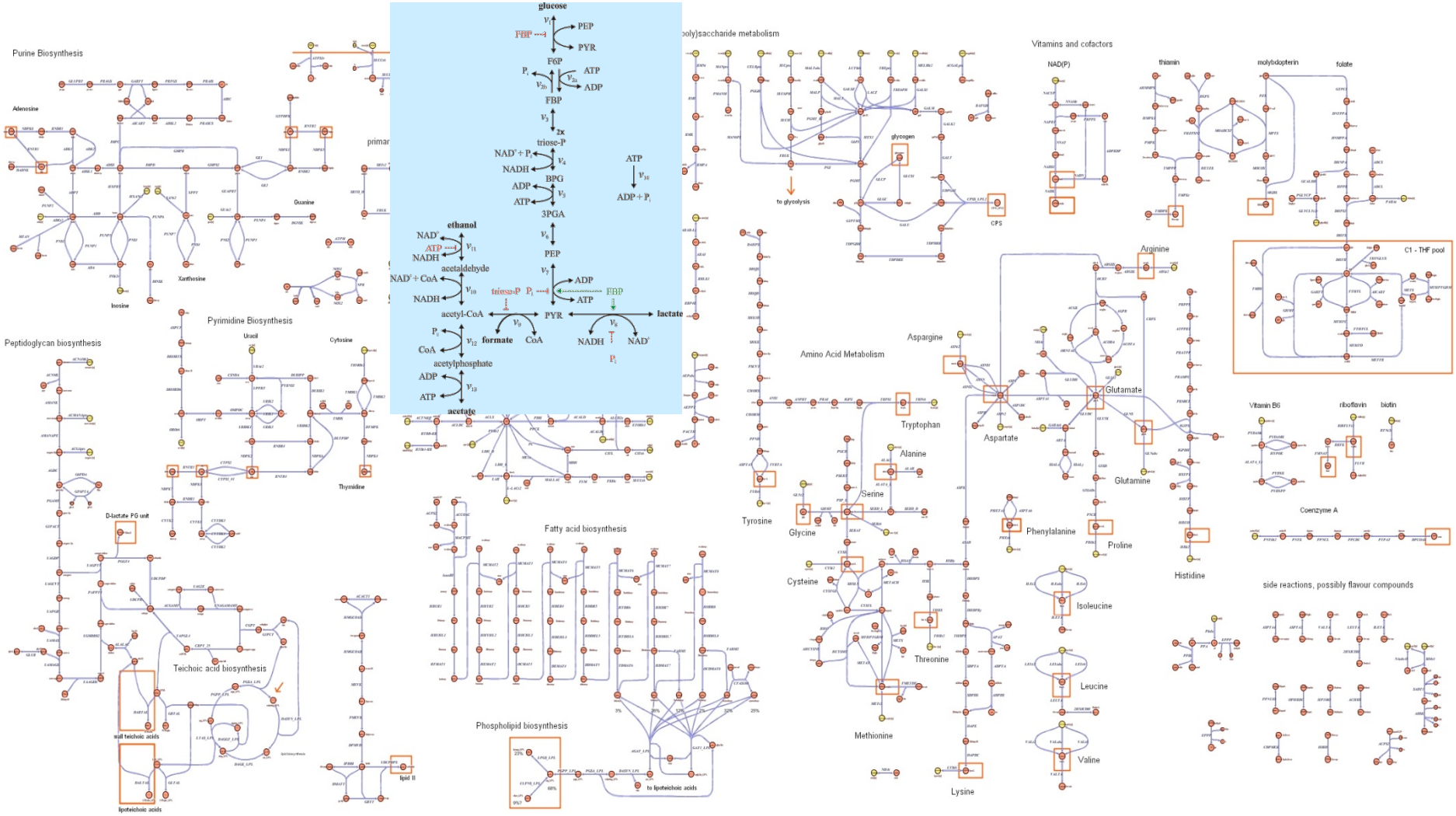
COPASI

etc.

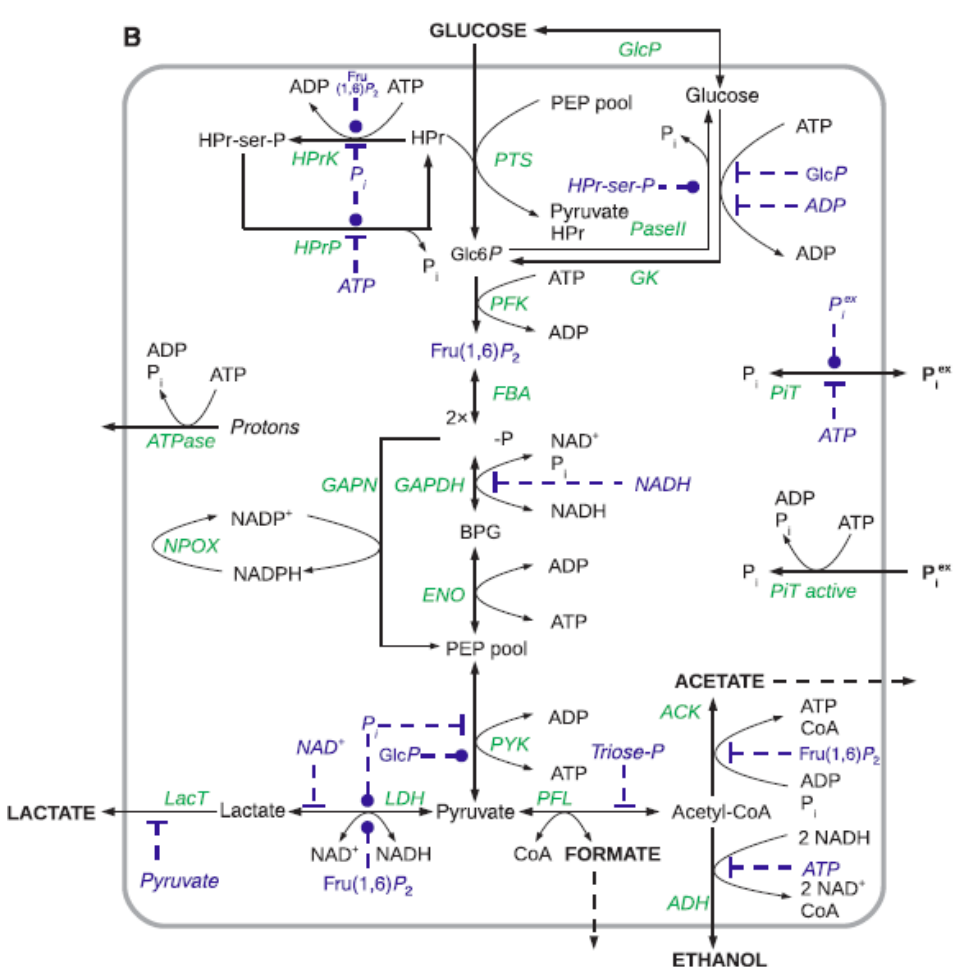
http://sycamore.h-its.org

Weidemann, Richter *et al.* Bioinformatics 2008

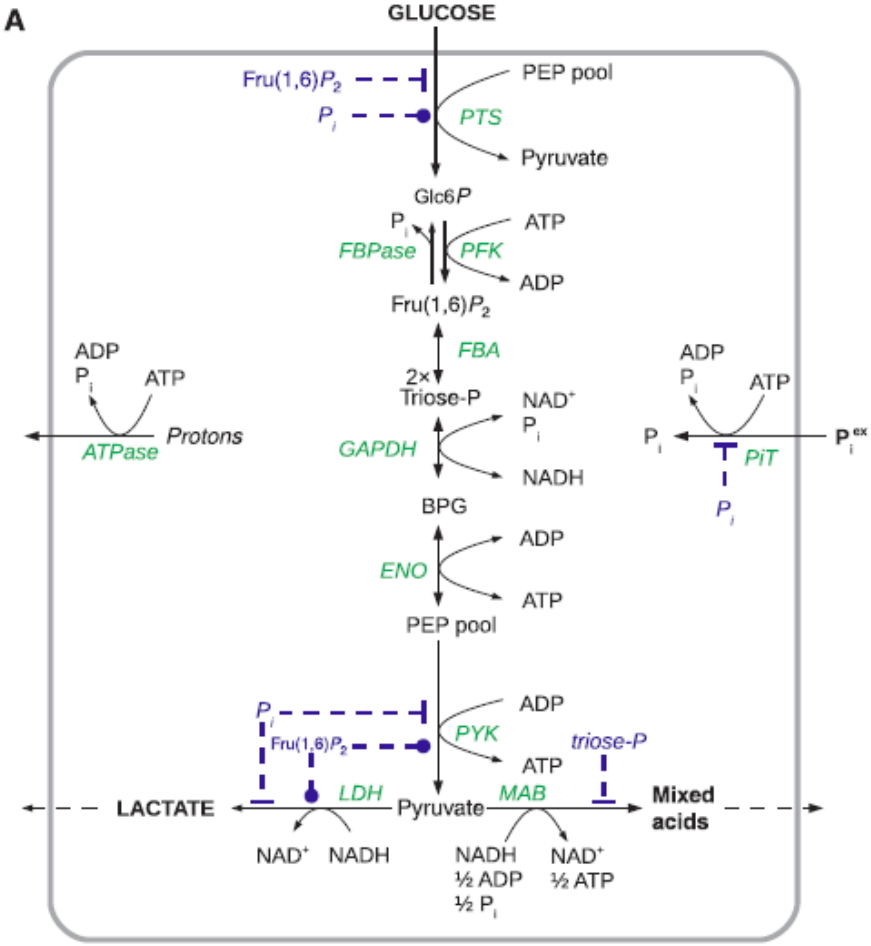
From protein structures to biochemical networks: regulation & cross-talk



Central metabolism of 2 lactic acid bacteria: regulation by phosphate

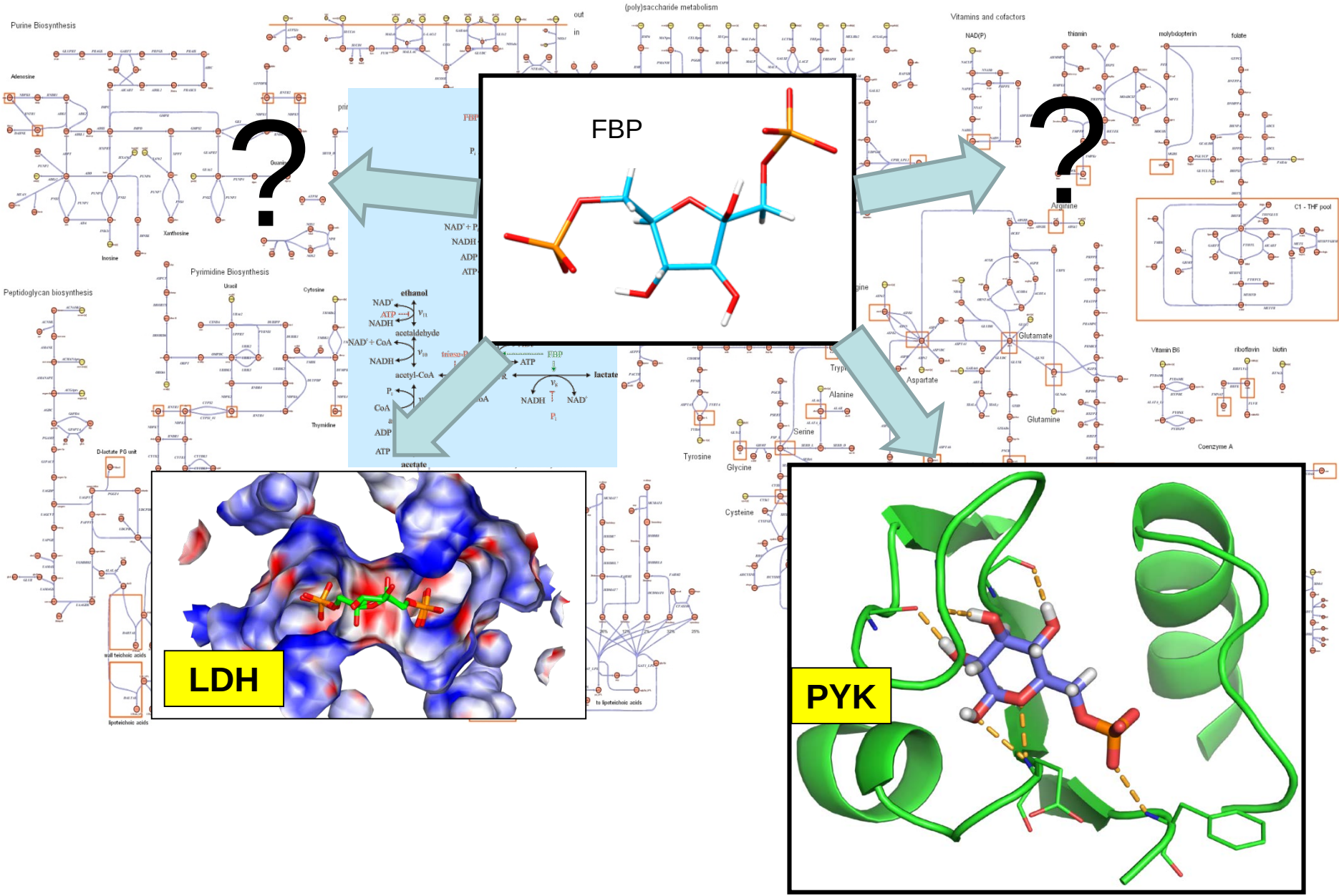


S. pyogenes central metabolism model



L. Lactis central metabolism kinetic model, Levering et al FEBS J. (2012) 279, 1274

Finding cross-talk between reactions



LigDig: a web server to answer ligand-based queries

http://mcm.h-its.org/ligdig

Fuller et al, Bioinformatics, 2015, 31, 1147-49

LigDig: Compound name disambiguation

Find compound by name Search PubChem

You searched for: **fructose 1,6-bisphosphate**

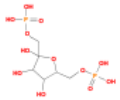
9 possible compounds were found (listed below).

For each compound found, the last column indicates the number of protein 3D structures that are available in the PDB.

Validate the corresponding checkbox(es) and click on the Submit to SearchPDB button at the end of the page. This will display all PDB files that contain your chosen ligand.

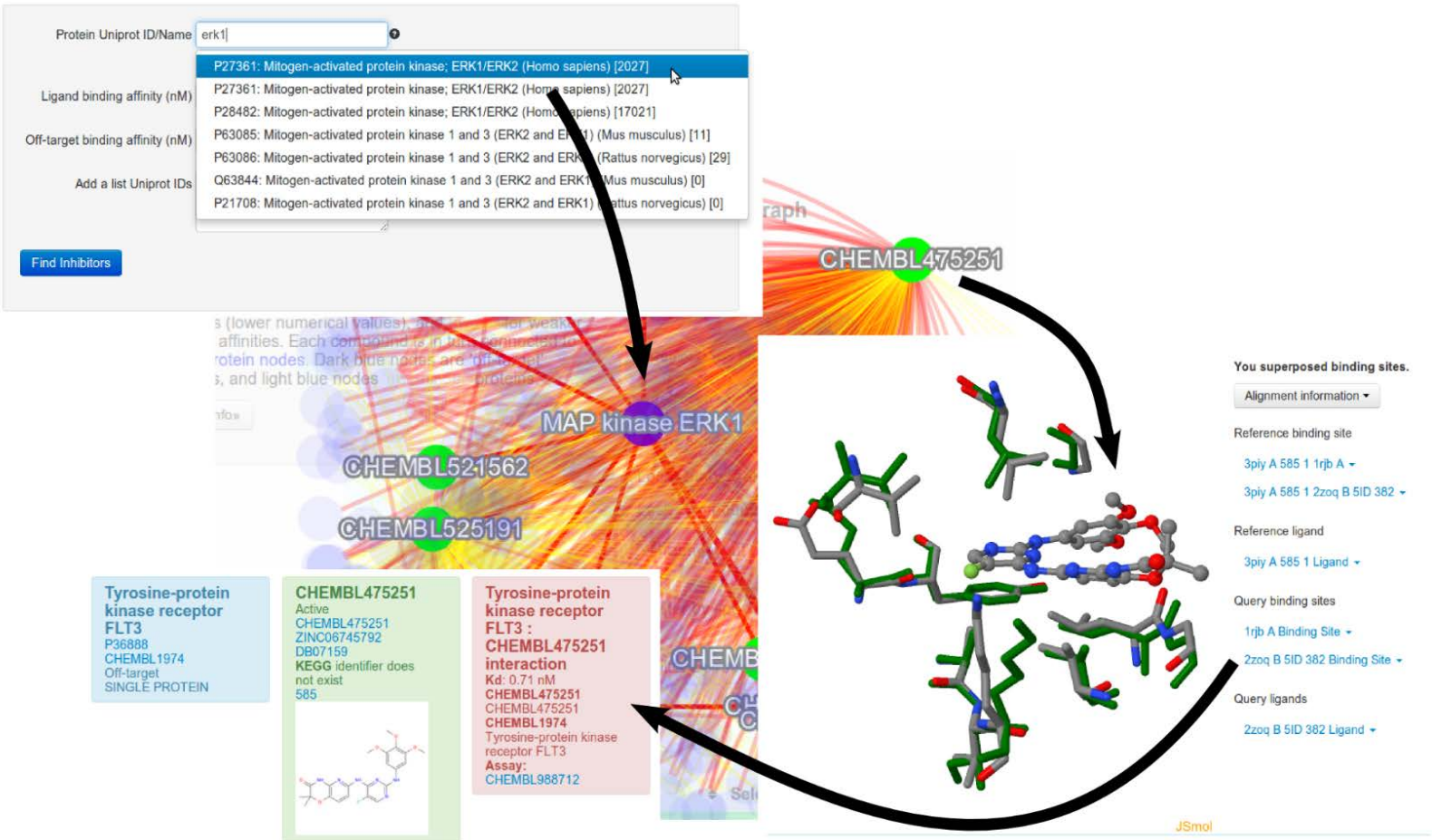
If no checkboxes are selected, or if you select only entries with no protein structures, you will also remain on this page when you click the Submit to SearchPDB button.

Submit to SearchPDB

CID	First Synonym	2D Structure	Additional compound information	# of PDB Structures	Select
172313	D-Fructose 1,6-bisphosphate Find ligand function		More details for **D-Fructose 1,6-bisphosphate** Find binding partners for CHEMBL1089962 in ChEMBL Find binding partners for CHEMBL97893 in ChEMBL More info»	40	☐
10267	fructose-1,6-diphosphate Find ligand function				
445557	CHEMBL1089962 Find ligand function				
718	1,6-di-o-phosphonohex-2-ulofuranose Find ligand function				
2734398	NCGC00166321-01 Find ligand function		More details for **NCGC00166321-01** Find binding partners for CHEMBL2146112 in ChEMBL	0	☐
16219367	F6803_SIGMA				

RESTful → PubChem

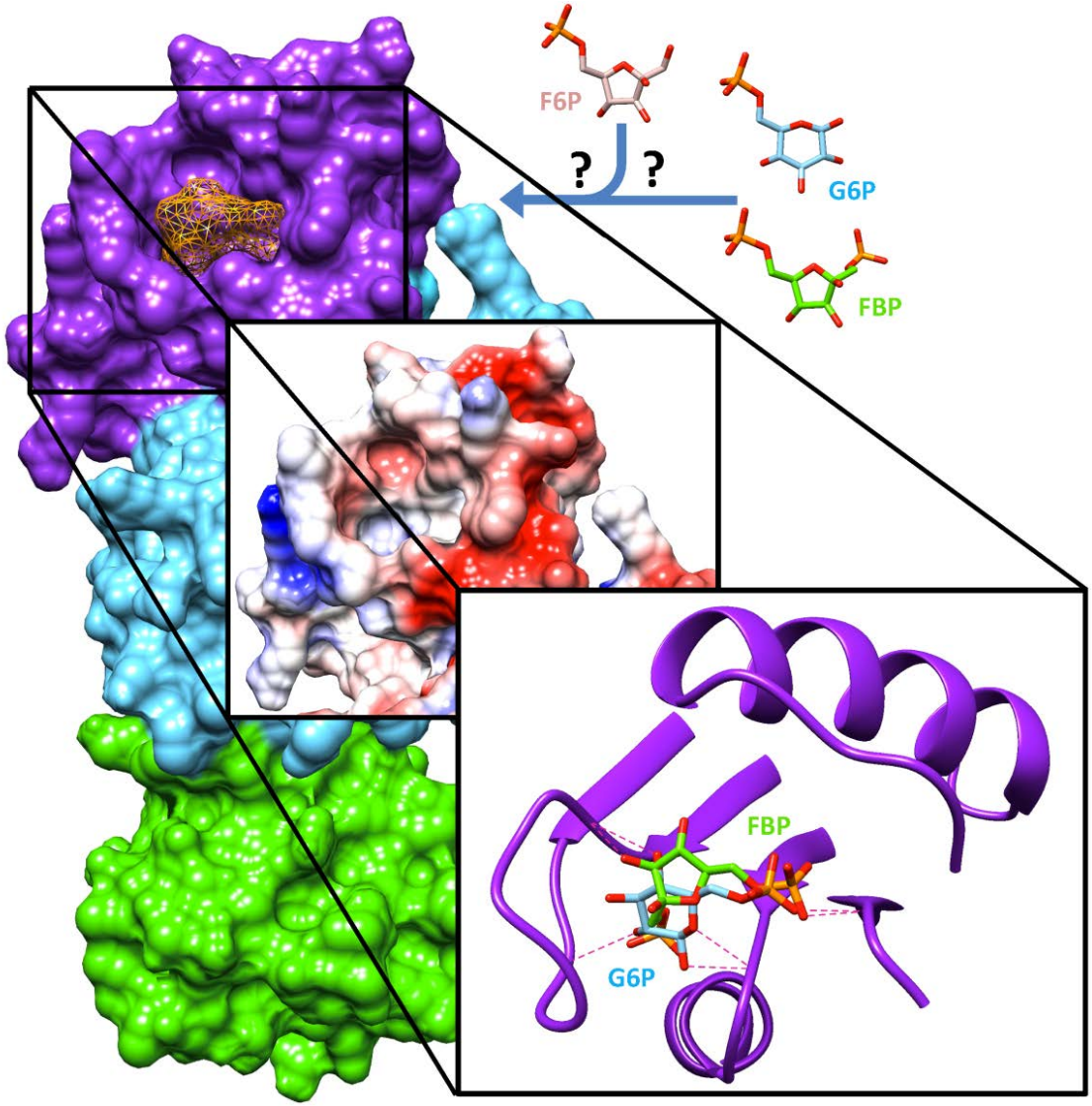
LigDig: Example application to kinase inhibitors



http://mcm.h-its.org/ligdig

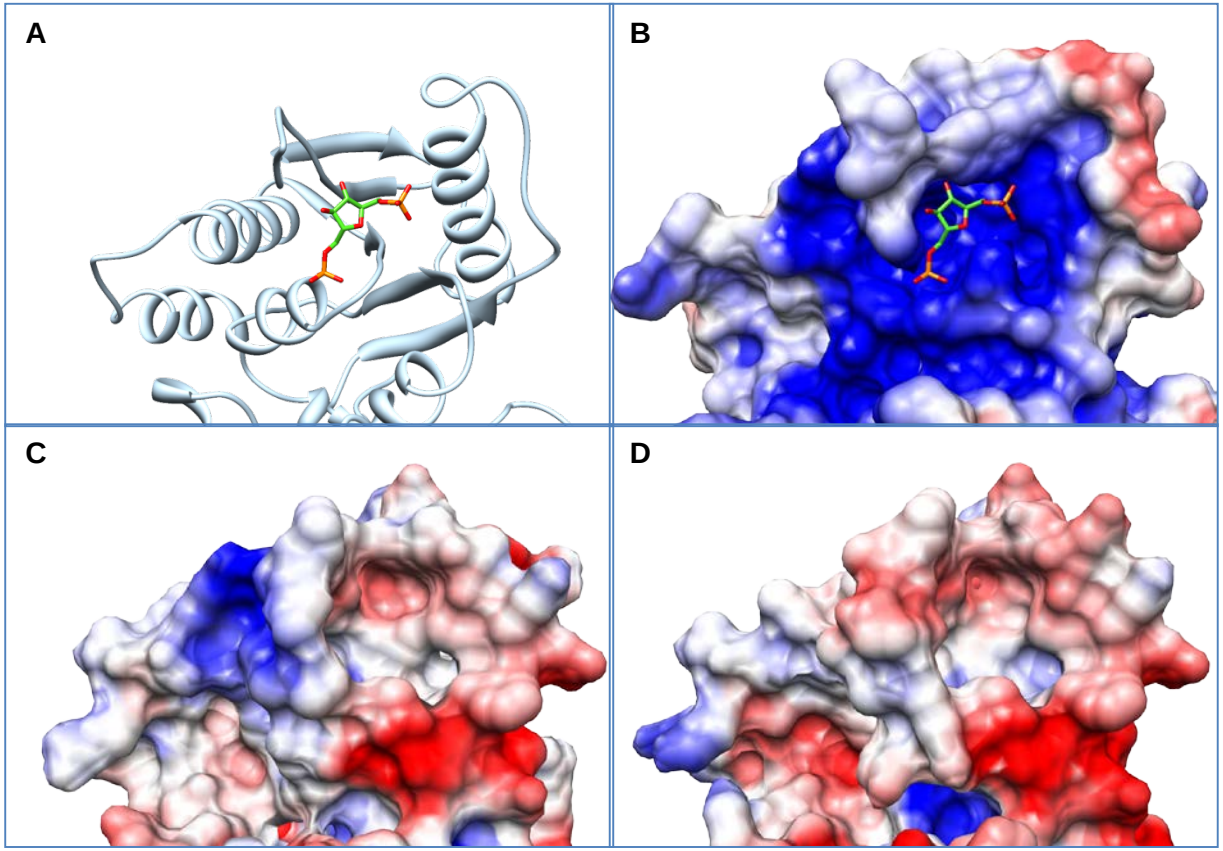
Fuller et al, Bioinformatics, 2015, 31, 1147-49

Pyruvate kinase in lactic acid bacteria: Which are the activators?



Pyruvate kinase – allosteric site: different electrostatic potentials

Crystal Structure of S. cerevisiae PYK with FBP bound



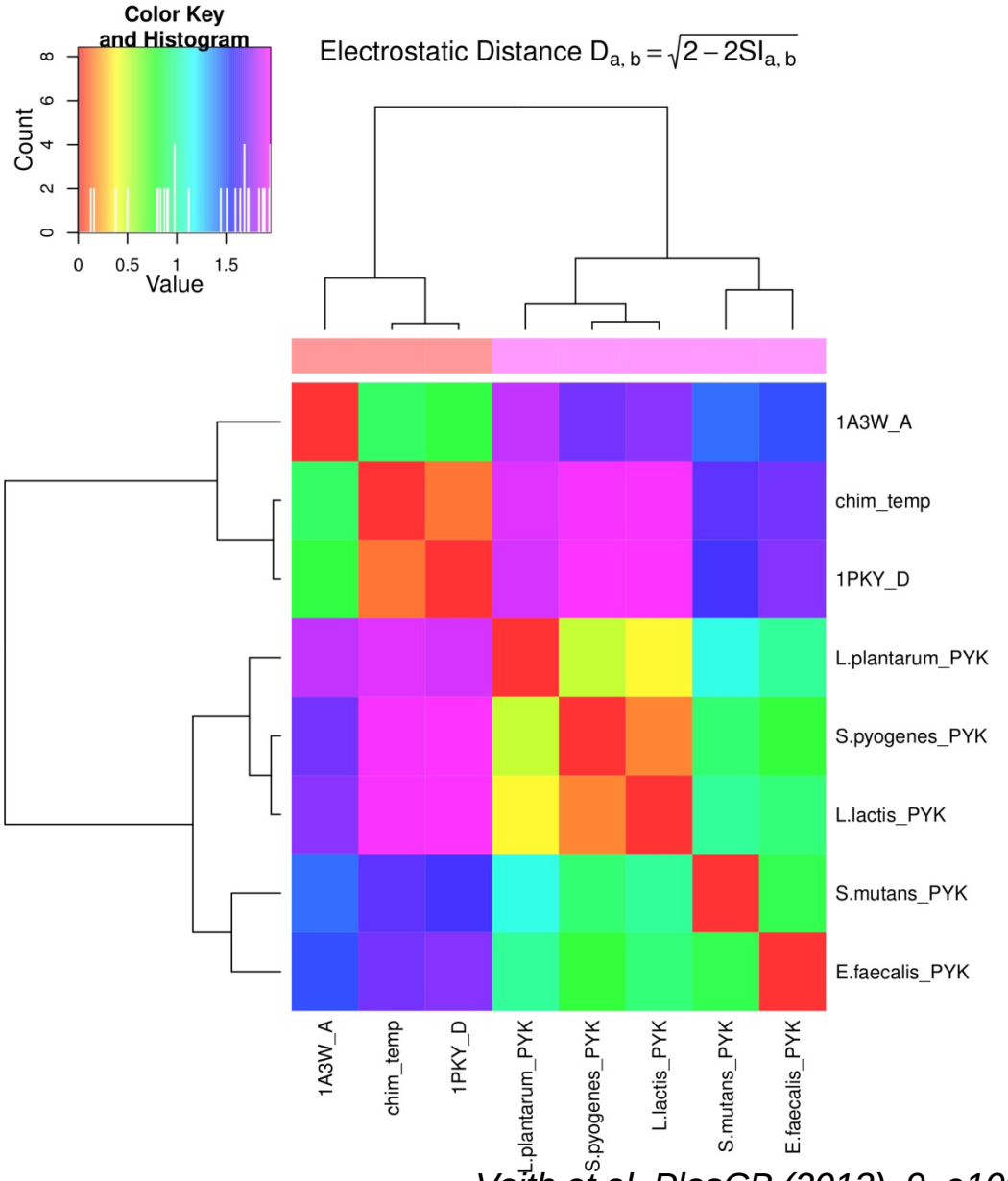
Modelled Structure of S. pyogenes PYK with open allosteric site

Modelled Structure of L. plantarum PYK with open allosteric site

Colored from -2 (red) to +2 (blue) kT/e

Veith et al PlosCB (2013) 9, e1003159

Pyruvate kinase – allosteric site: electrostatic similarity



Cluster: yeast, chimeric template, E. coli

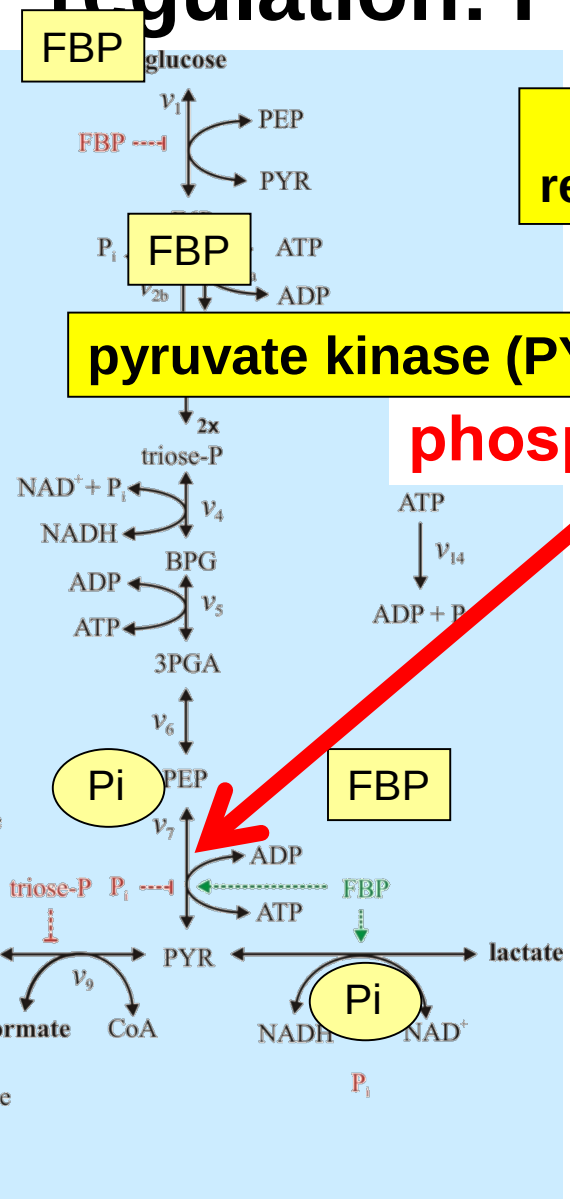
Cluster : L lactis, S. pyogenes, L. plantarum,

S. mutans, E. faecalis

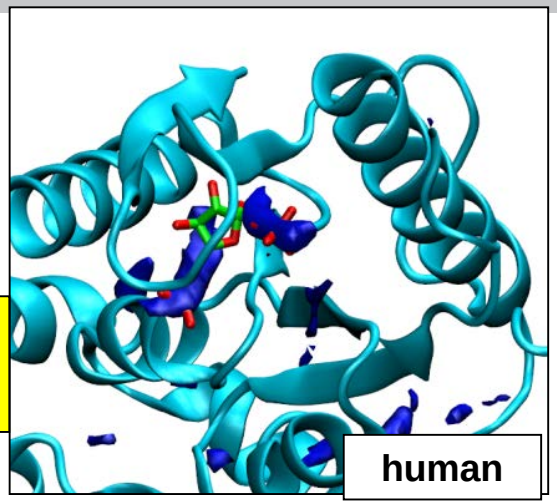
PIPSA analysis

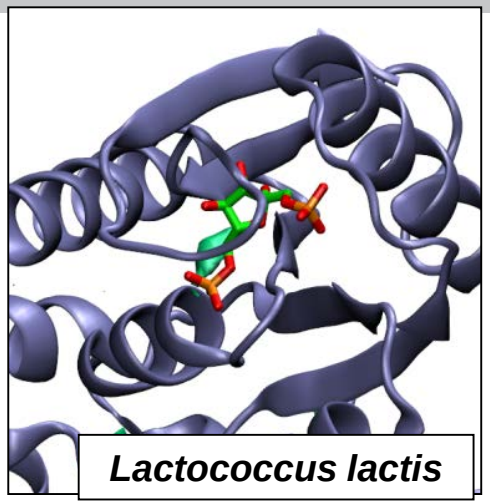
Veith et al PlosCB (2013) 9, e1003159

Allosteric regulation: PYK



FBP-regulation





2 phosphate sites **1 phosphate sites**

phosphoenolpyruvate + ADP ↔ pyruvate + ATP

E. coli	FBP
B. stearoth.	R5P + AMP (G6P)
S. mutans	G6P (R5P)
L. lactis	G6P, R5P, F6P, FBP, Gal6P
S. pyogenes	**G6P, Gal6P, R5P, F6P, FBP**
E. faecalis	**FBP**
L. plantarum	**FBP, G6P, Gal6P**

binding and allosteric regulation of phosphorylated sugars

- Modifier used may depend on environment
- Different LABs adapted to different environments
- Some must survive in more environments

Using protein structures to learn about protein function: Learning objectives

- Protein structure and function
- Modeling protein structure and dynamics
- Computing interaction properties

Thank you for your attention!