

M
C
M

From Molecular Interaction Fields to Enzyme Inhibitors and Biochemical Networks

Rebecca C. Wade

**Molecular & Cellular Modeling Group,
EML Research
Heidelberg**

rebecca.wade@eml-r.villa-bosch.de

<http://projects.villa-bosch.de/mcm/>

Questions:

pipsa.eml.org

- What are the binding properties of a set of structurally related proteins?
 - ◆ Classification w.r.t. binding properties
 - ◆ Detection of similarities/differences, e.g. for design of selective agents
 - ◆ Identification of regions of similarity/difference

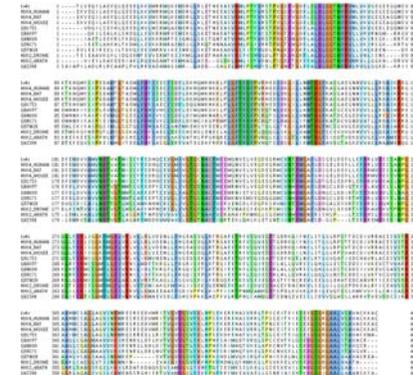
→ PIPSA

- If experimental data on binding affinities or kinetics of a ligand are available for 1 or several proteins, can this be used to estimate values for a similar protein of interest?
 - ◆ Correlate properties from protein structures with binding and kinetic data.

→ qPIPSA

Levels of Protein Comparison

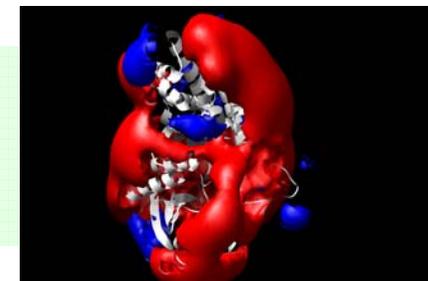
- Amino Acid Sequence Identity



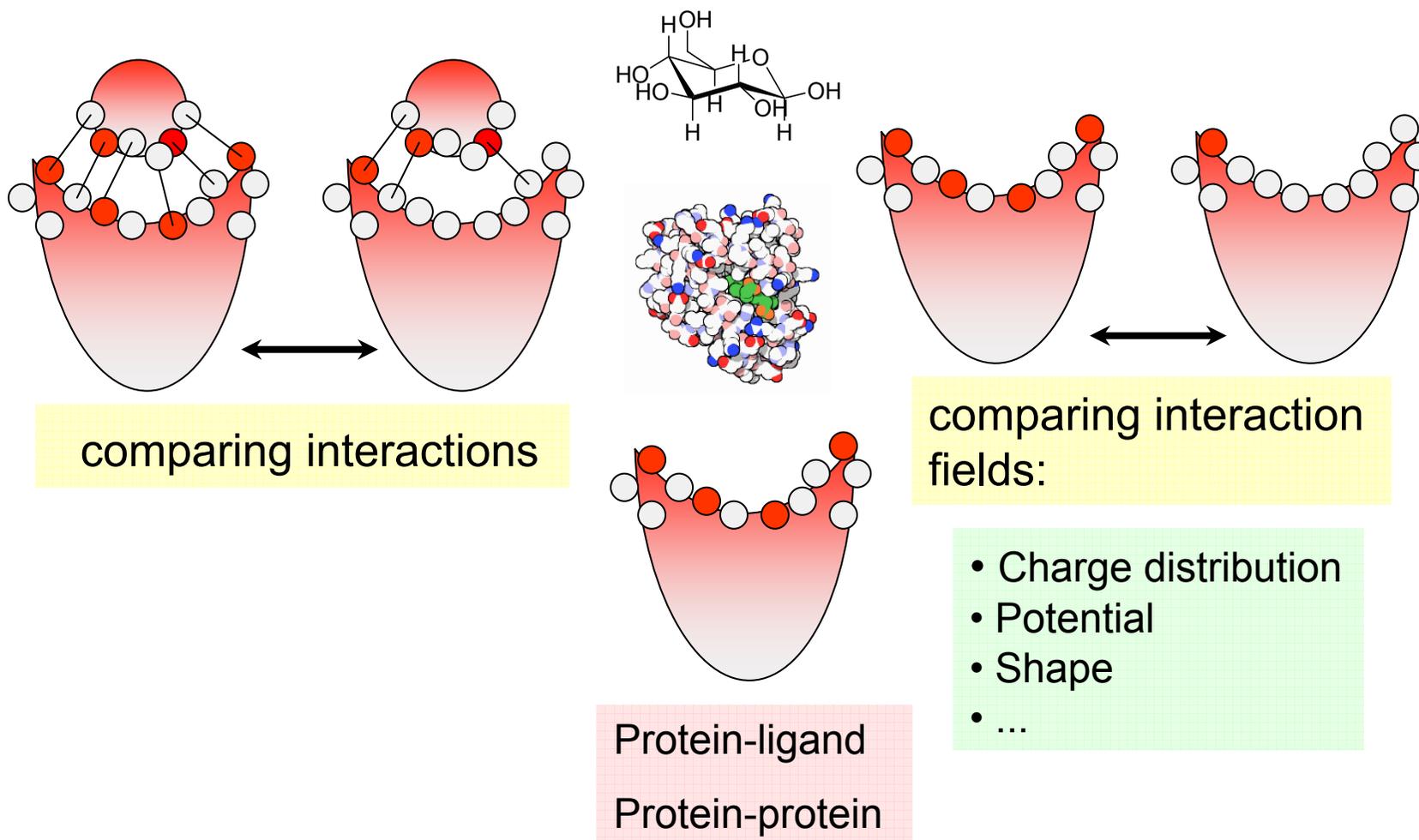
- Protein Structure (NMR, X-Ray)



- Protein Structure/Function Relationship - MIF

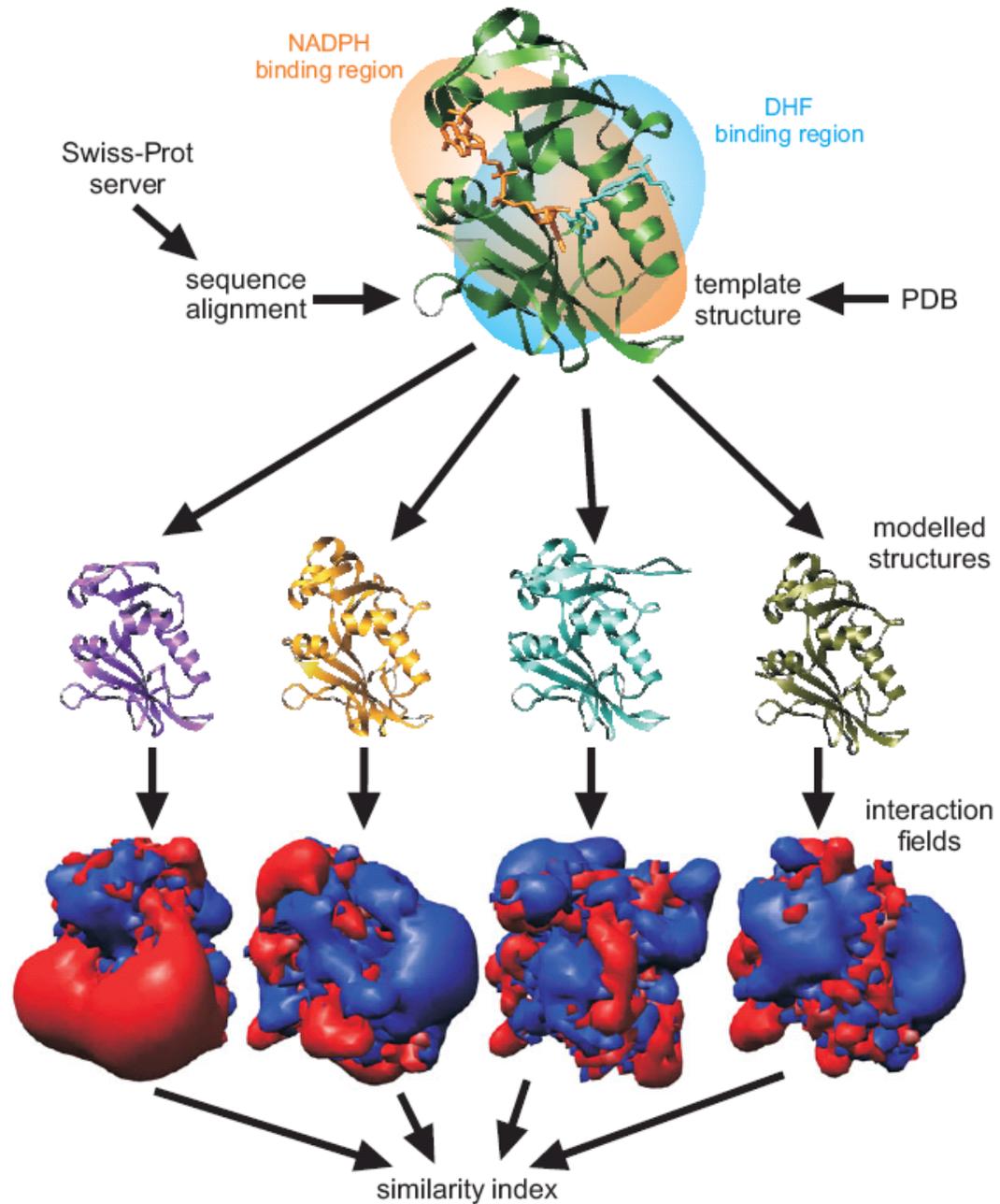


Comparing Molecular Interaction Fields

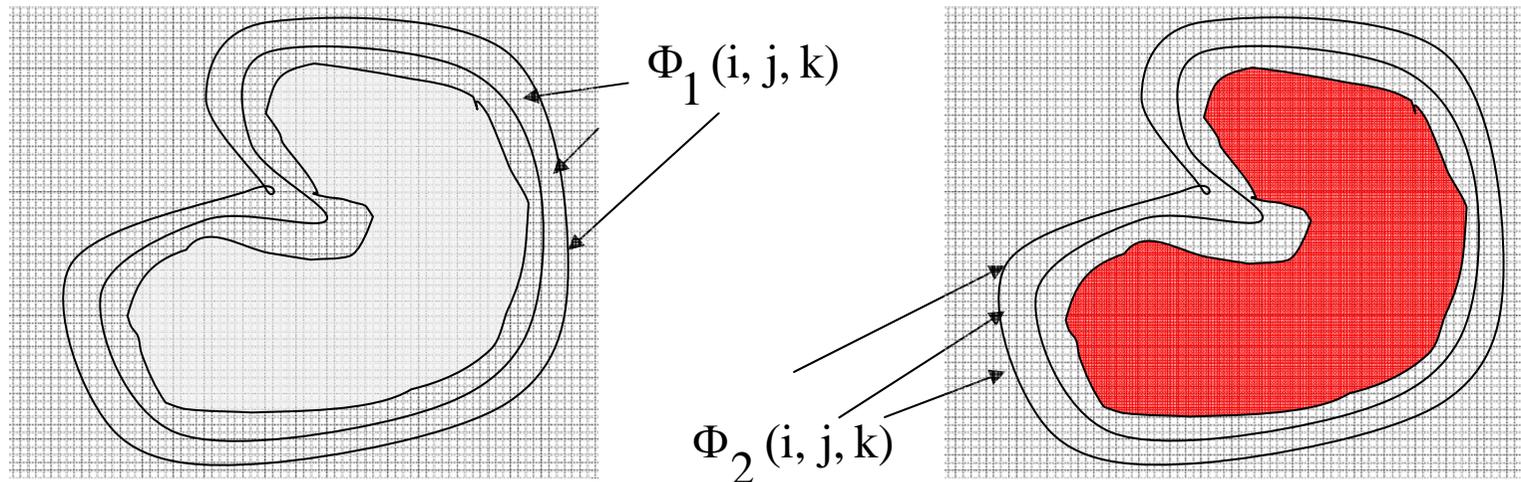


PIPSA :

e.g. dihydrofolate reductases (DHFR) from 20 species



PIPSA: Protein Interaction Property Similarity Analysis

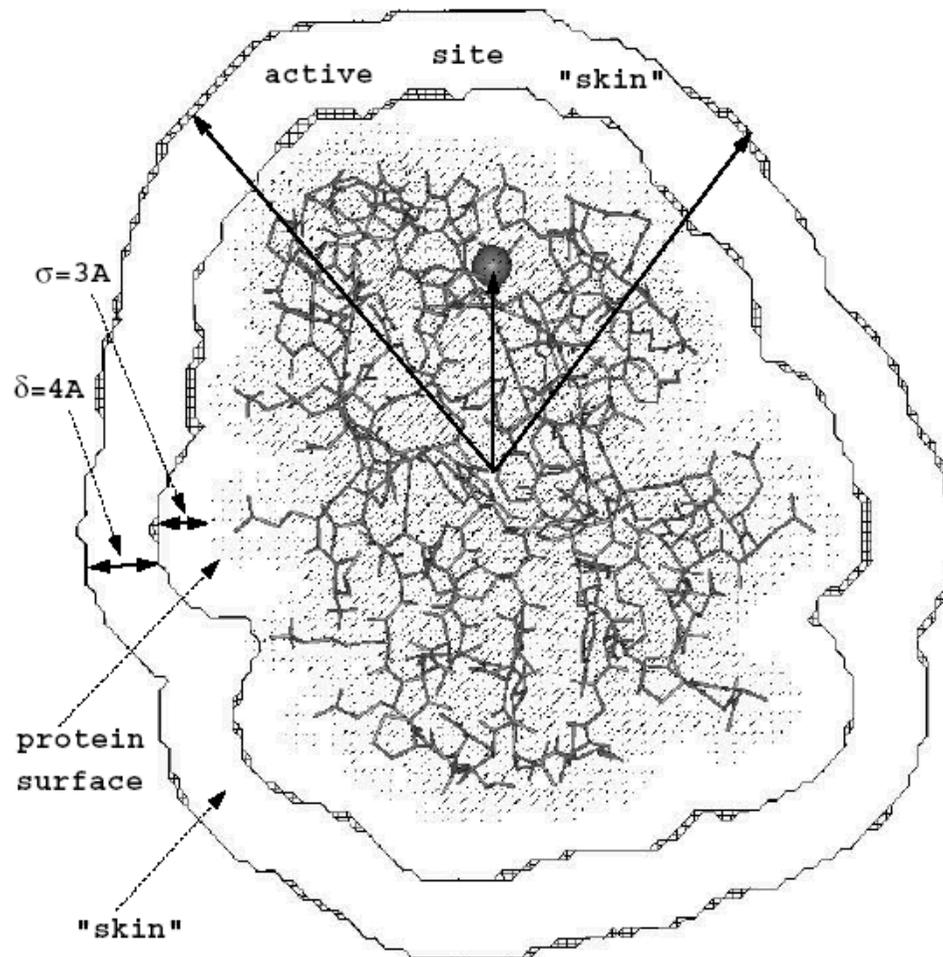


Protein 1

Protein 2

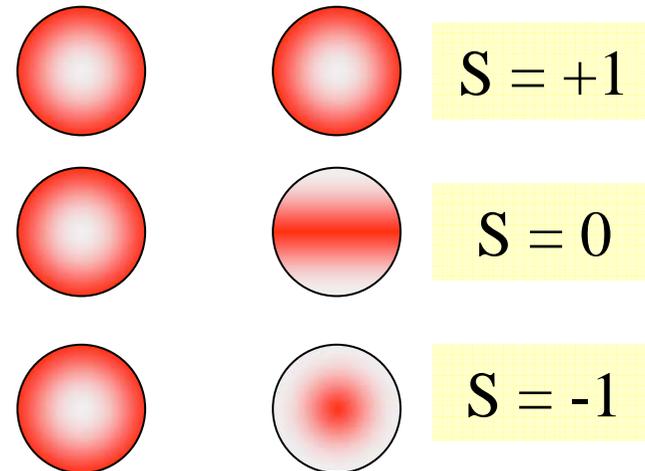
- Interaction fields are calculated on a set of points
- Field values on corresponding points are compared
- Φ = electrostatic potential, shape, probe interaction field, ...

PIPSA: Protein Interaction Property Similarity Analysis



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

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



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

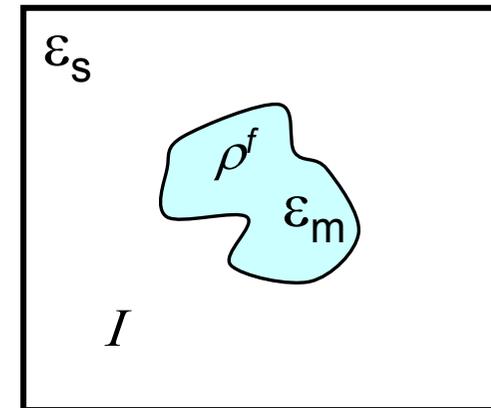
Continuum Electrostatics

- Linearized Poisson-Boltzmann equation

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

$$\kappa^2(r) = \frac{\beta}{\varepsilon_0 \varepsilon_r} \sum_1^N c_{i,bulk} q_i^2 = \frac{2e^2 N_A}{\varepsilon_0 \varepsilon_r kT} I$$

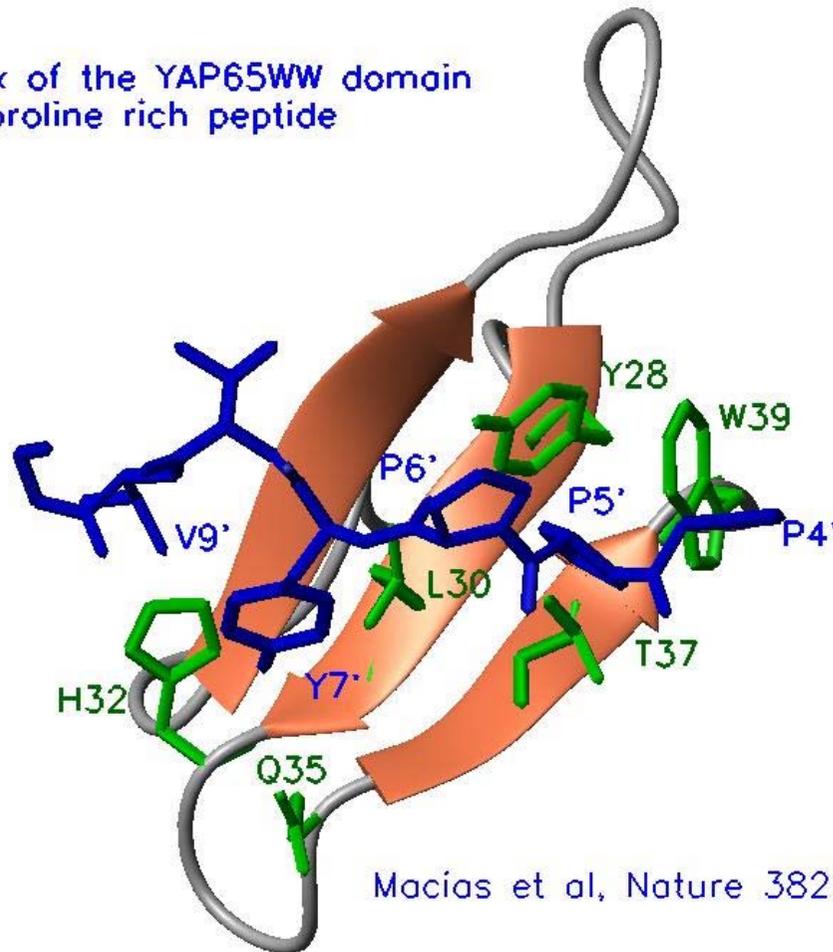
- Finite-difference
- Numerical solution



WW domain/peptide complexes

Binding specificity and affinity determinants?

Complex of the YAP65WW domain
and a proline rich peptide



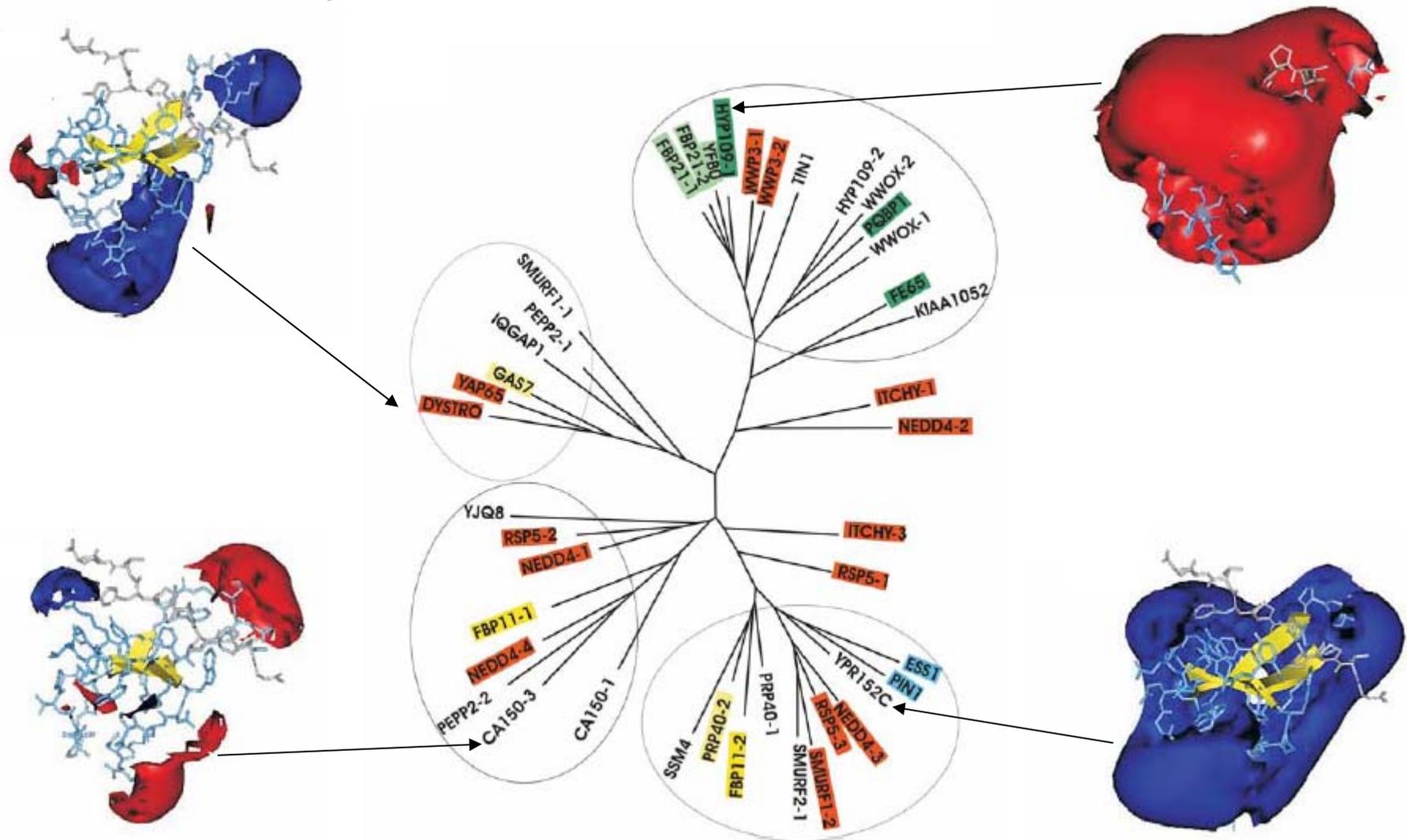
WW domain peptide
binding preferences:

- xPPx(Y/poY)
- (p/Φ)P(p,g)PpR
- (p/Φ)PPRgpPp
- PPLPp
- (p/Ψ)PPPPP
- (poS/poT)P

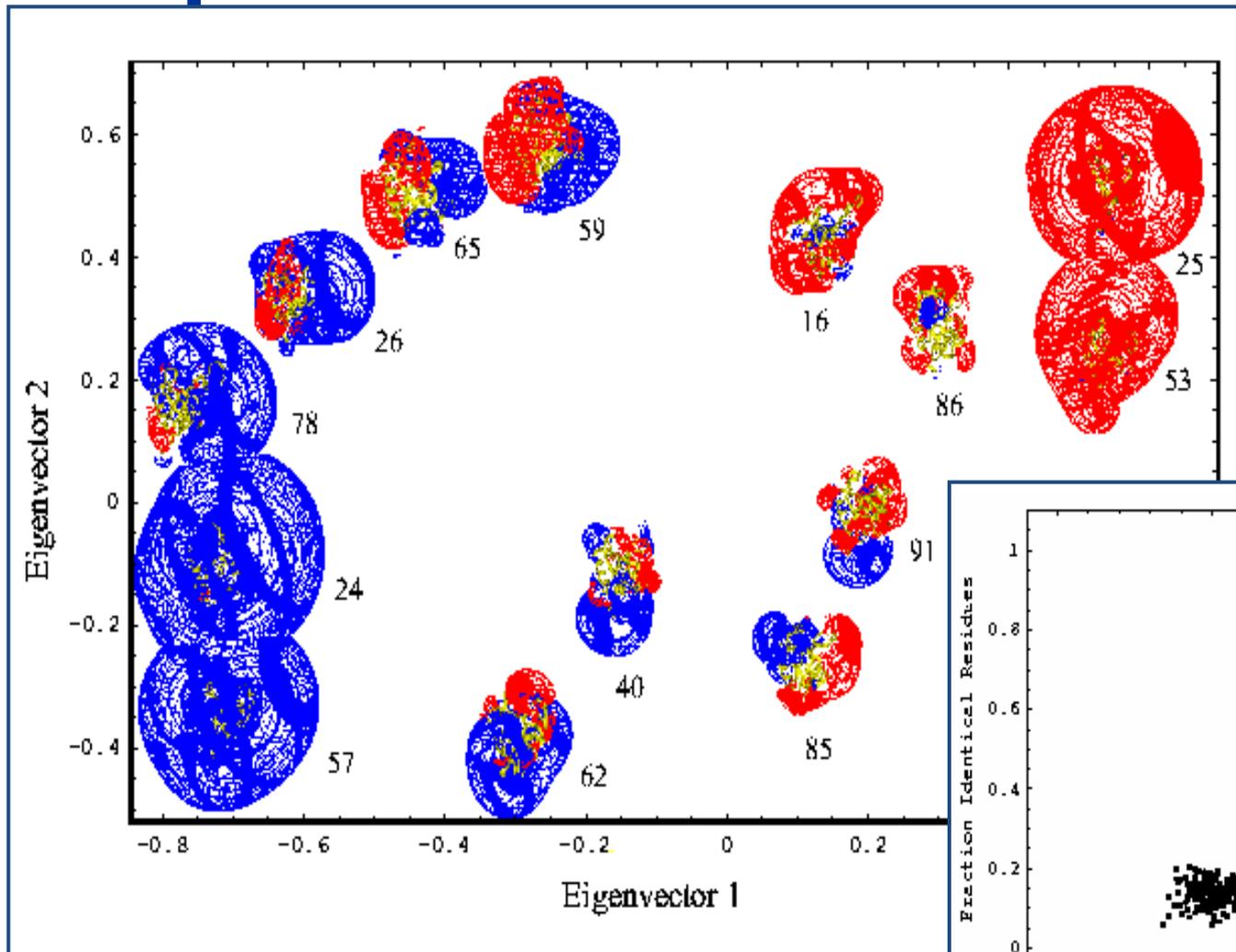
Otte et al. (2003) *Protein Sci.* 12, 491

42 WW Domains: PIPSA epogram for Molecular electrostatic potential

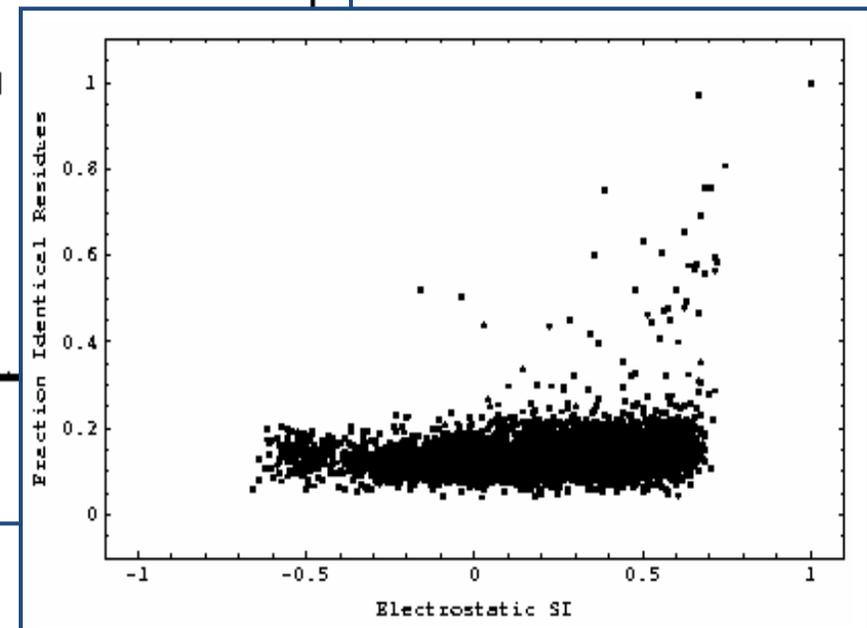
isopotential contours: -0.4 / +0.4 kcal/mol/e



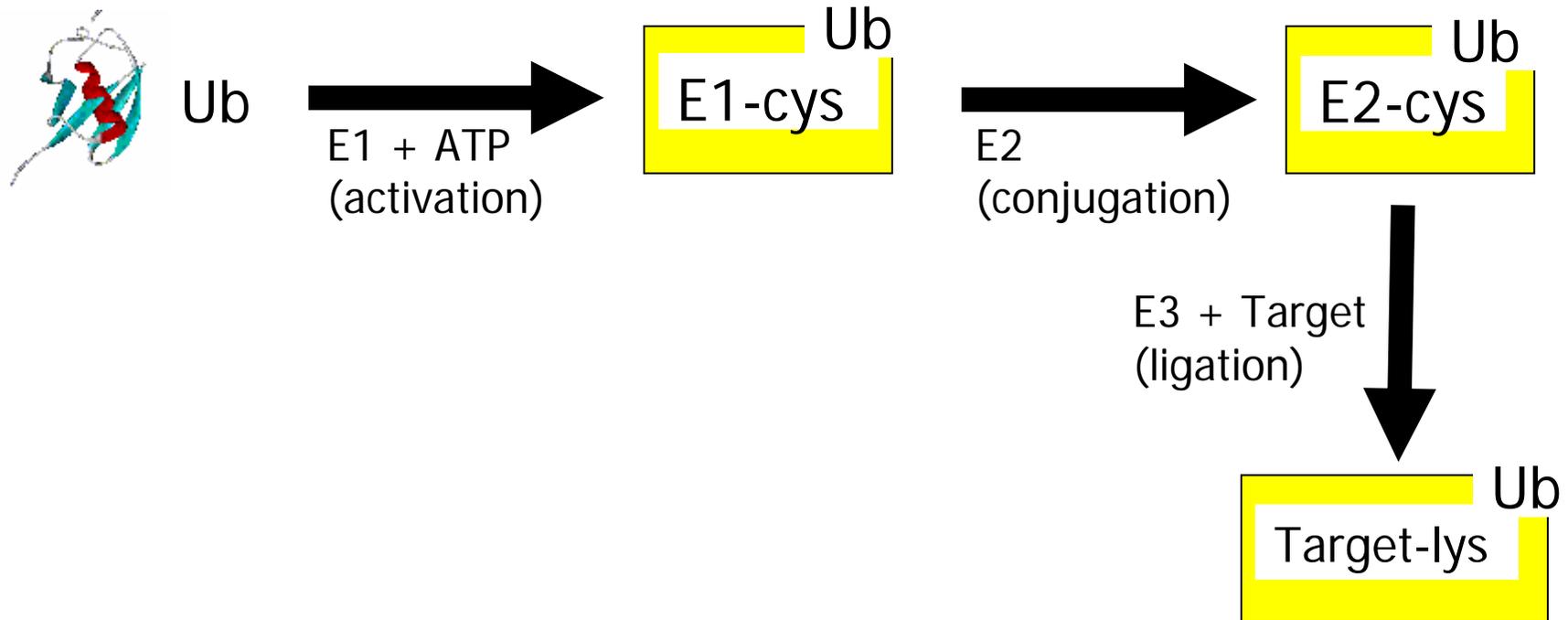
PH domains - distribution in electrostatic potential similarity space



Distribution for DH-linked and internal PH repeat domains



Ubiquitination Machinery



Mono ubiquitination

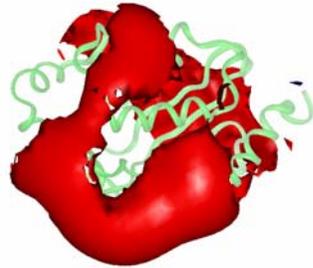
- endocytosis
- histone modification (epigenetic code)



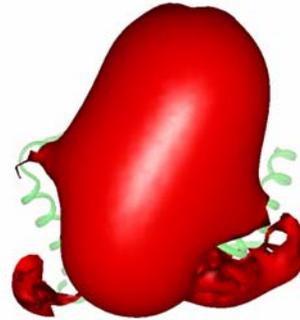
Multi-ubiquitination

- K48 linked chains: proteasomal degradation (cell-cycle, apoptosis ...)
- K63 linked: postreplicative DNA repair

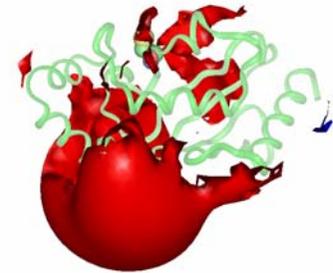
A Sample Of E2 potentials



ScUbc8



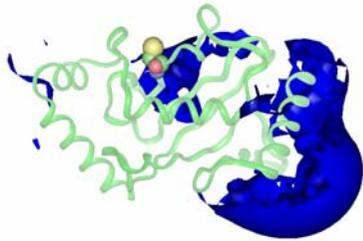
ScUbc3/CDC34



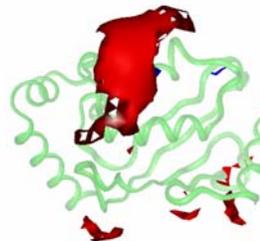
ScUbc2/RAD6

Histone ubiquitination *in vitro*

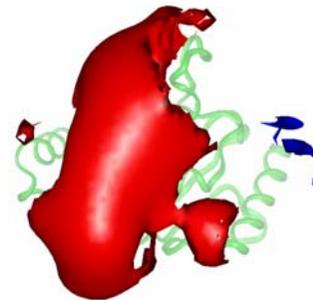
Histone ubiquitination *in vivo*



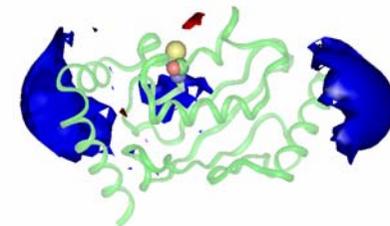
ScUbc9



ScUbc4



ScUbc7



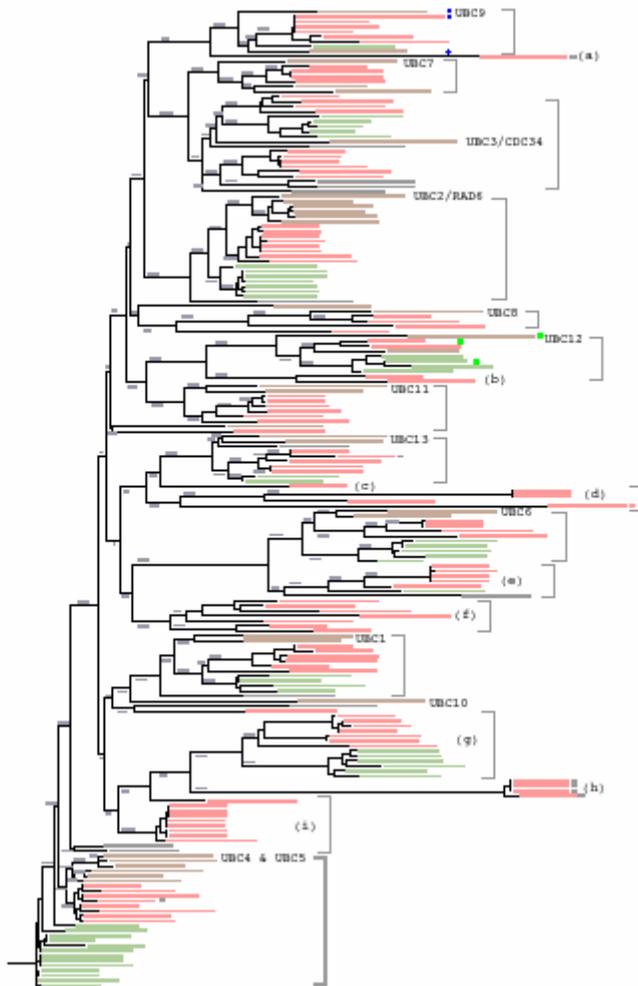
ScUbc6

SUMO

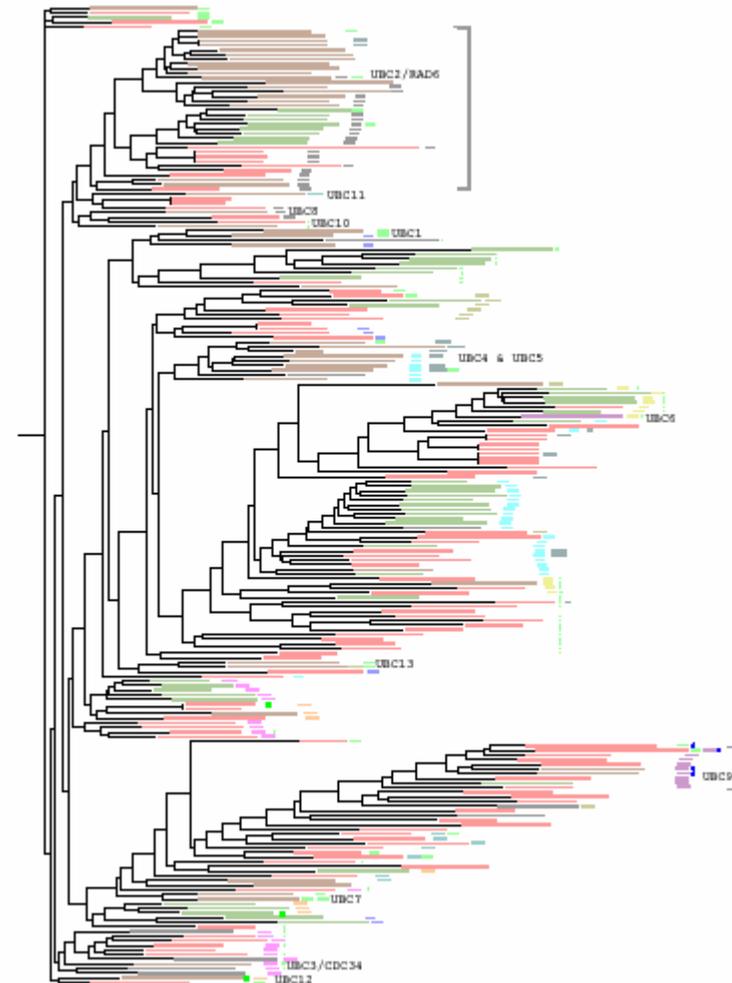
General

Endoplasmic Reticulum

196 E2 proteins:

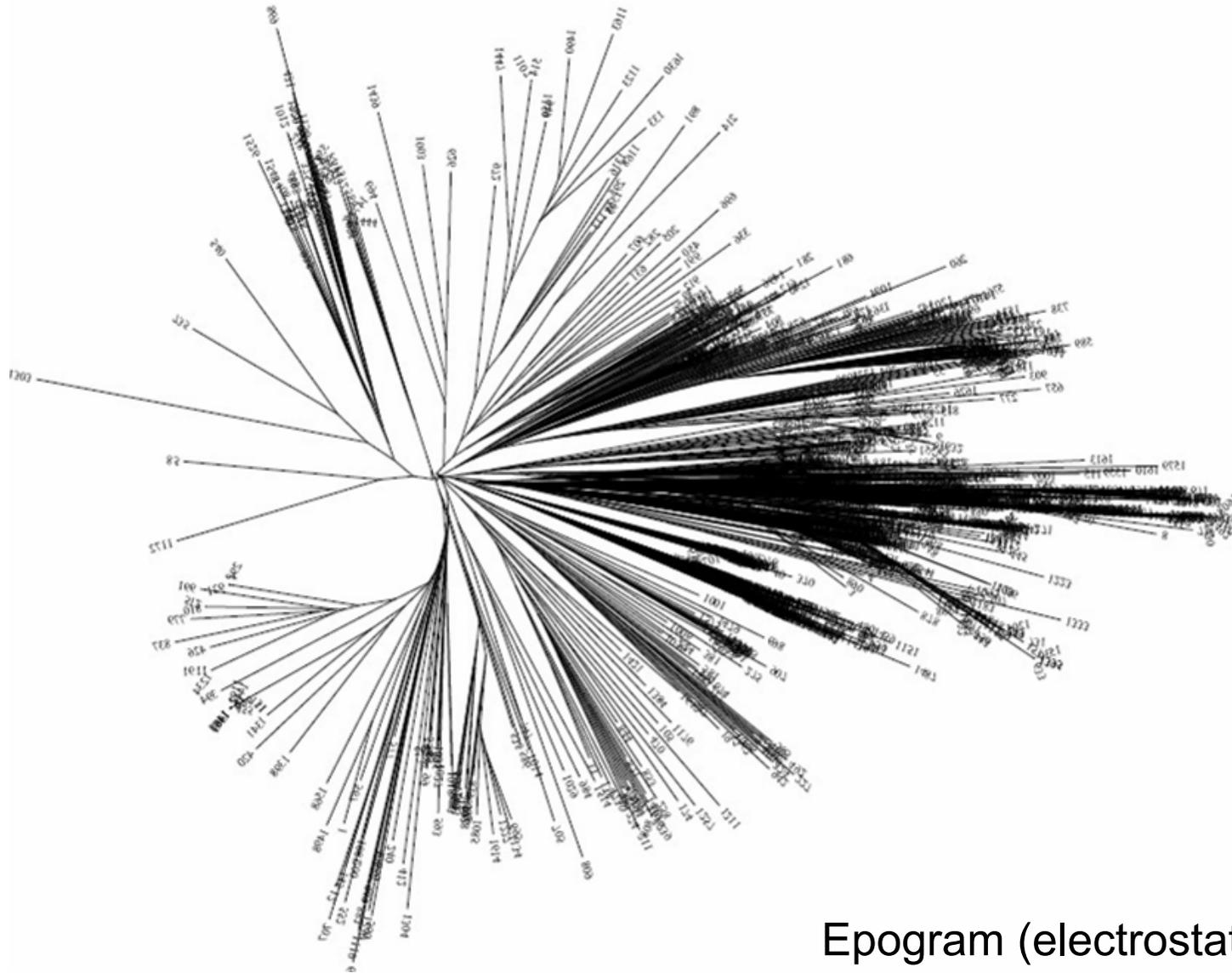


Phylogenetic tree (sequence)



Epogram (electrostatic potential)

1600 E2 proteins:



Epogram (electrostatic potential)

Query proteins by electrostatic similarity:

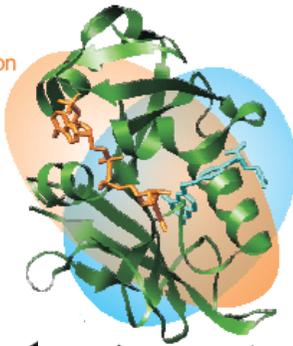
The screenshot displays a web browser window with the URL "Ubiquitin and Ubiquitin-like Protein Resource - Home". The page features logos for EMBL, EML Research, and Wayne State University. The main heading is "Ubiquitin and Ubiquitin-like Protein Web Resource". Below this, it states "The 10 E2s electrostatically most similar to Q9Y385".

The interface shows a "View 1" section with a link to "rotate view by 180 degrees". The results are presented in a grid with three columns: "Backbone with cysteines", "Surface potential +kT (blue) to -kT (red)", and "isopotential +0.5kT (blue) and -0.5kT (red)".

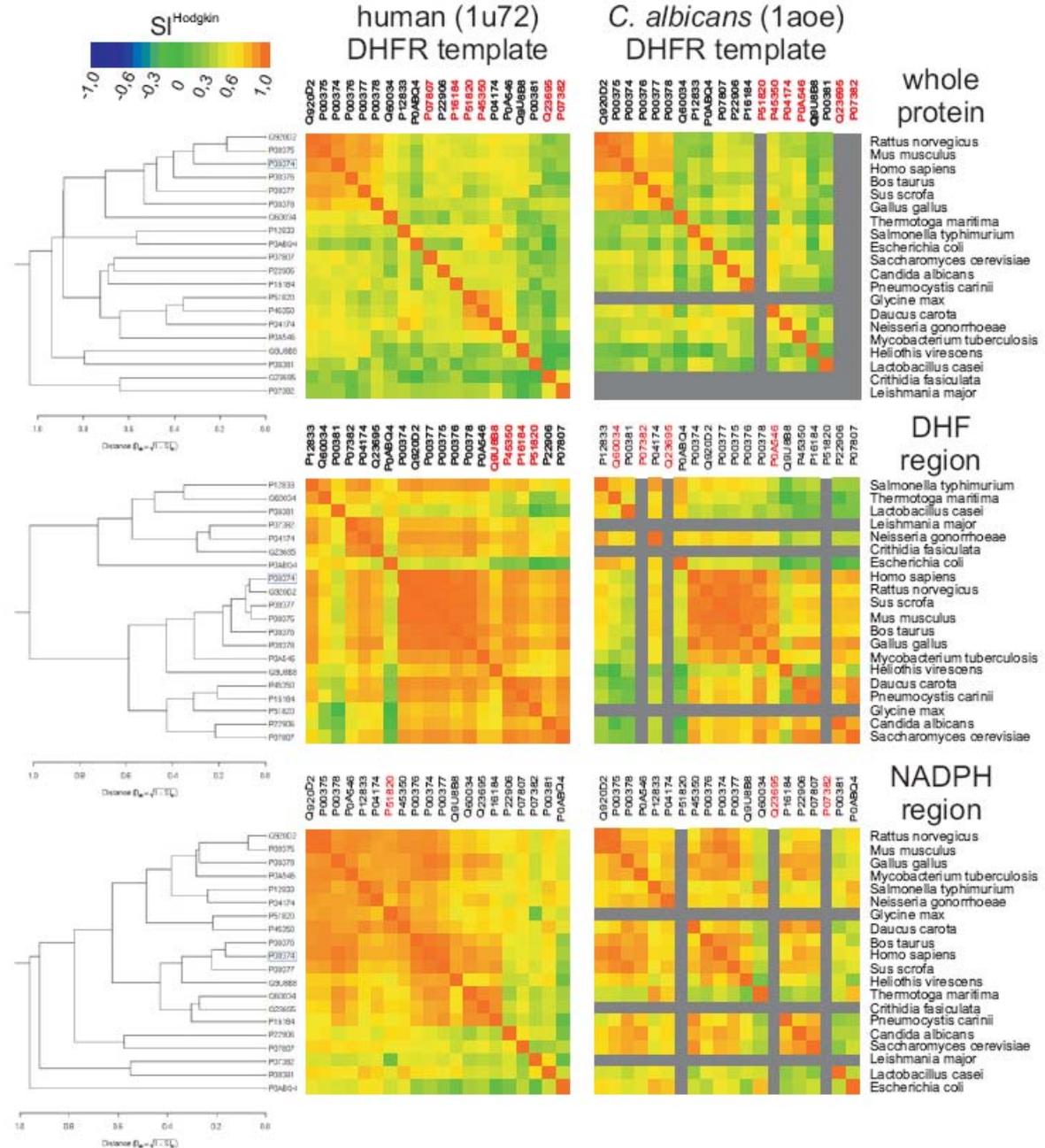
Query:	Backbone with cysteines	Surface potential +kT (blue) to -kT (red)	isopotential +0.5kT (blue) and -0.5kT (red)	Model Information
Q9Y385				Ramachandran Model quality Download Protein Information Uniprot DIP Mutation Information ProSat2 help
Q9JJZ4 Hodgkin Similarity = 1.000 Carbo Similarity = 1.000				Model Information Ramachandran Model quality Download Protein Information Uniprot DIP Mutation Information ProSat2 help
ENSP00000353502 Hodgkin Similarity = 1.000 Carbo Similarity = 1.000				Model Information Ramachandran Model quality Download Protein Information Ensembl Mutation Information ProSat2 help
				Model Information

PIPSA for DHFR from 20 species

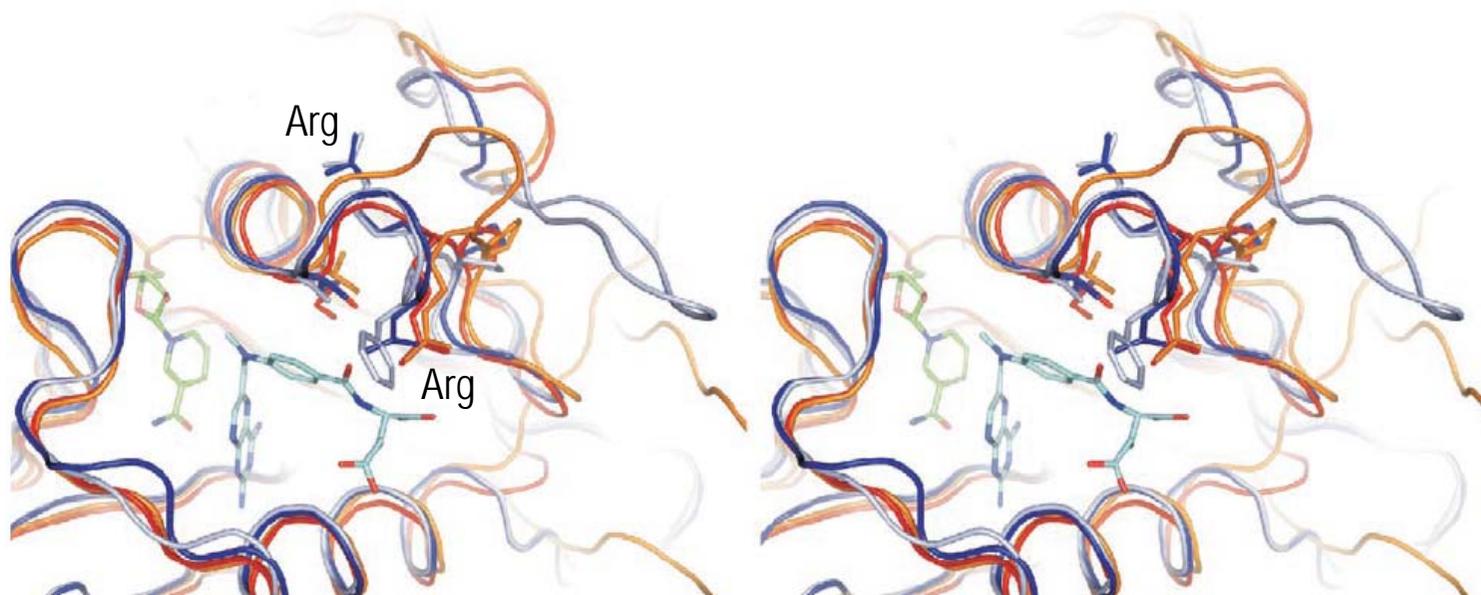
NADPH
binding region



DHF
binding region



Structural modeling issues

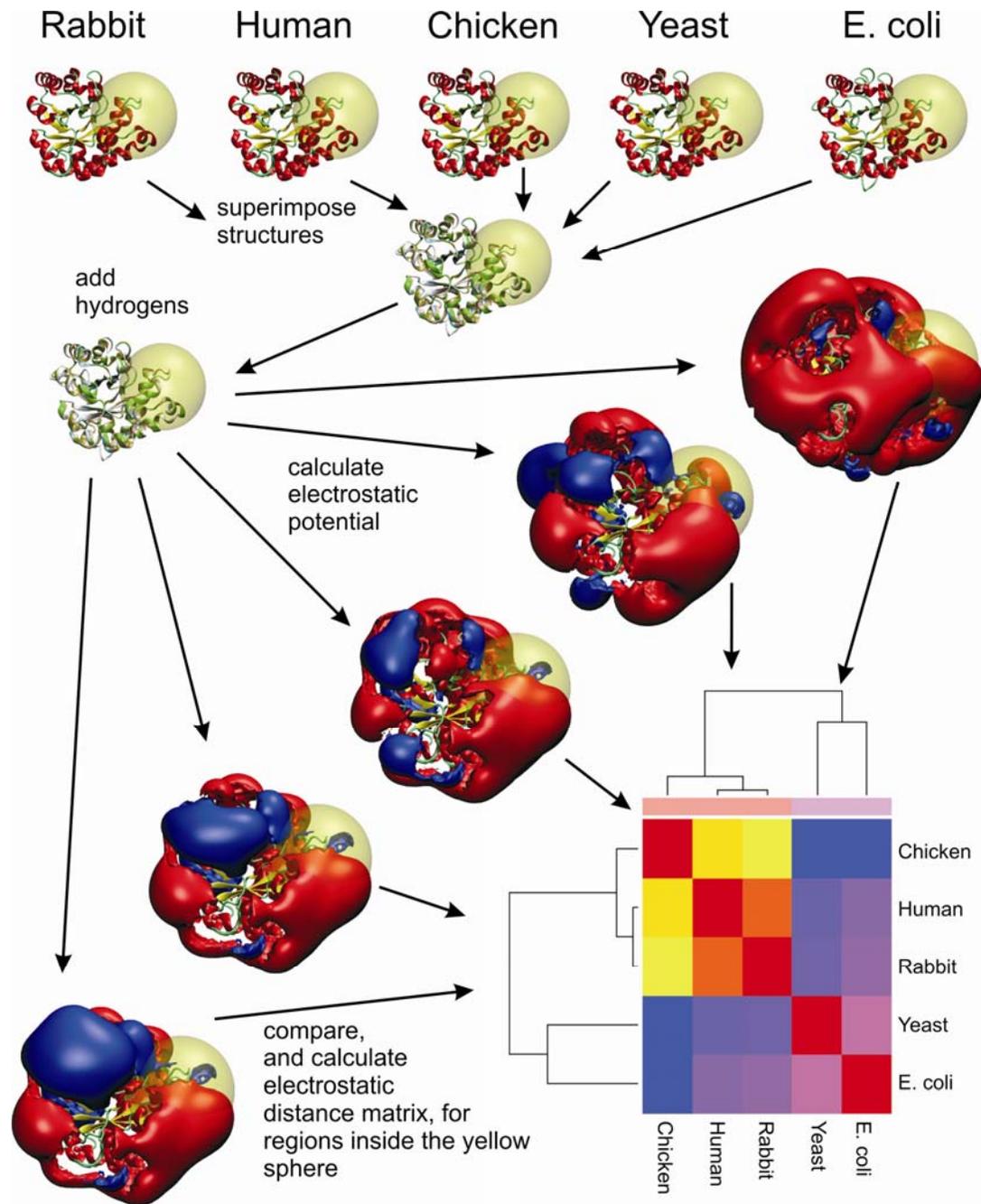


Crystal structures: 1u72 (human; blue) and 1rh3 (*E. coli*; red),
Modelled *Leishmania major* DHFR (P07382) structures:
1u72-based (light blue) and 1rh3-based (orange)

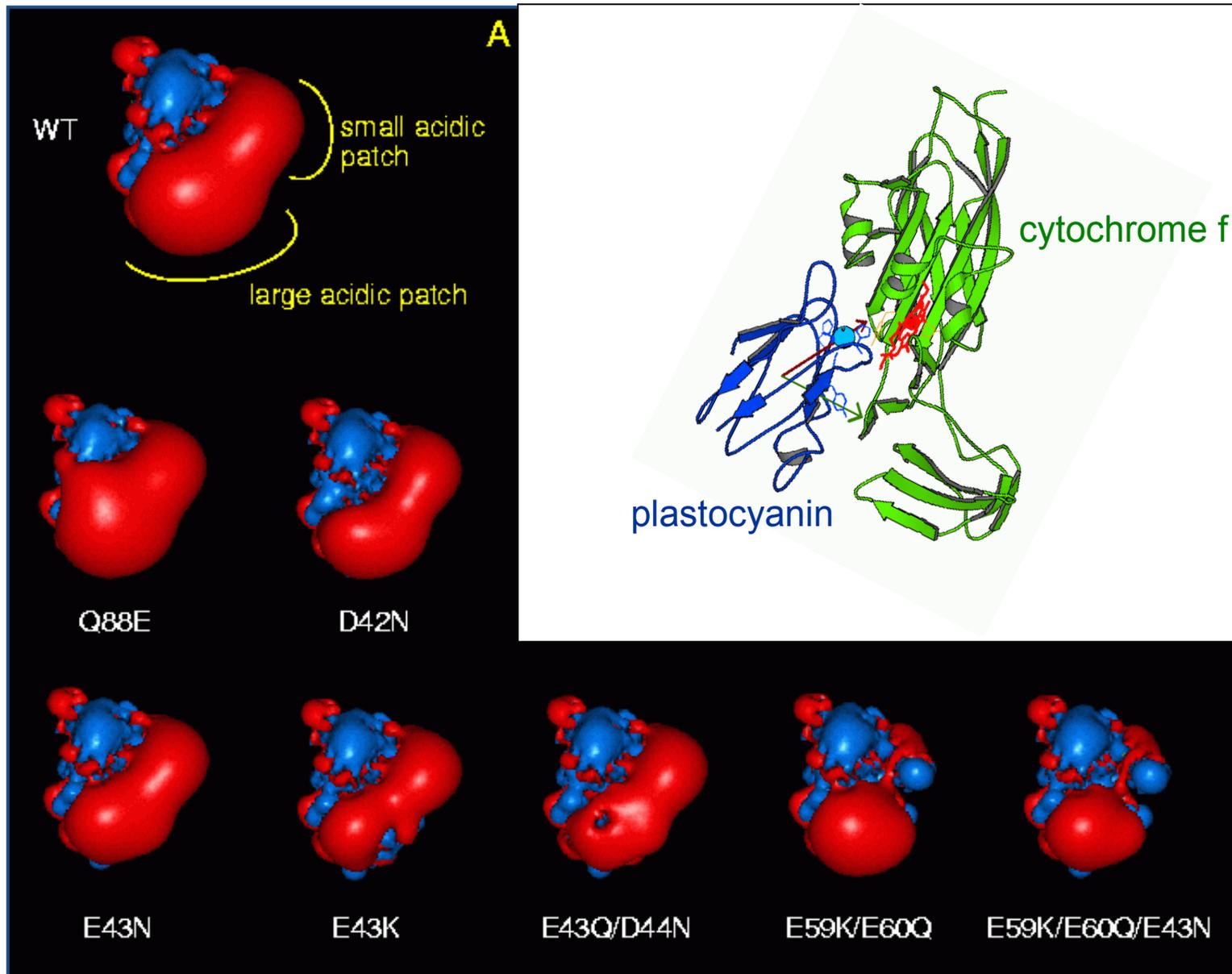
cofactor, NADPH (green); inhibitor, methotrexate (cyan)

webPIPSA:

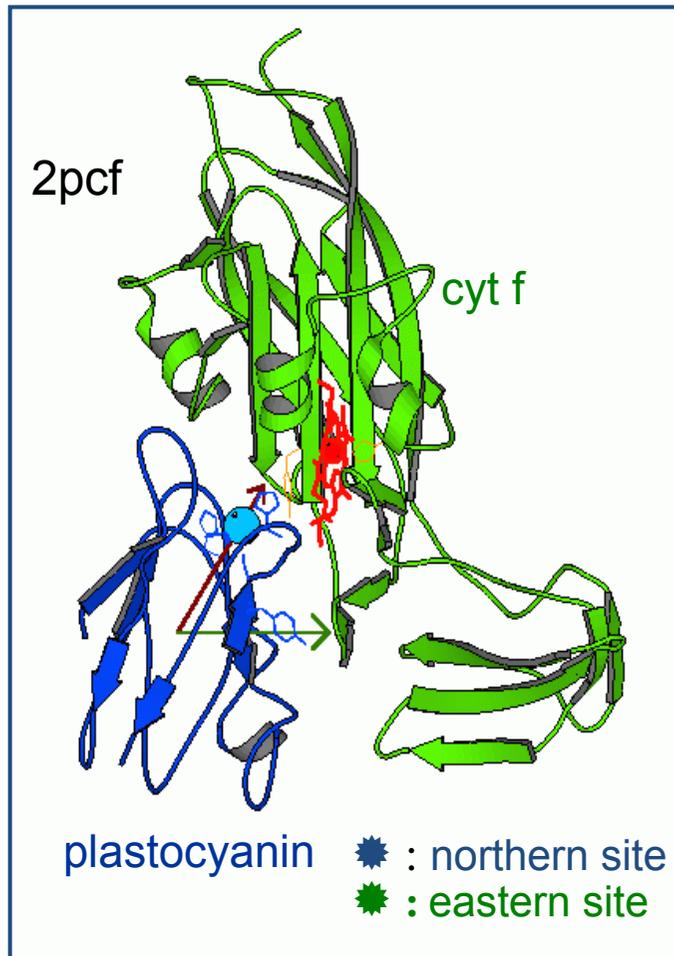
webPIPSA:
pipsa.embl.org



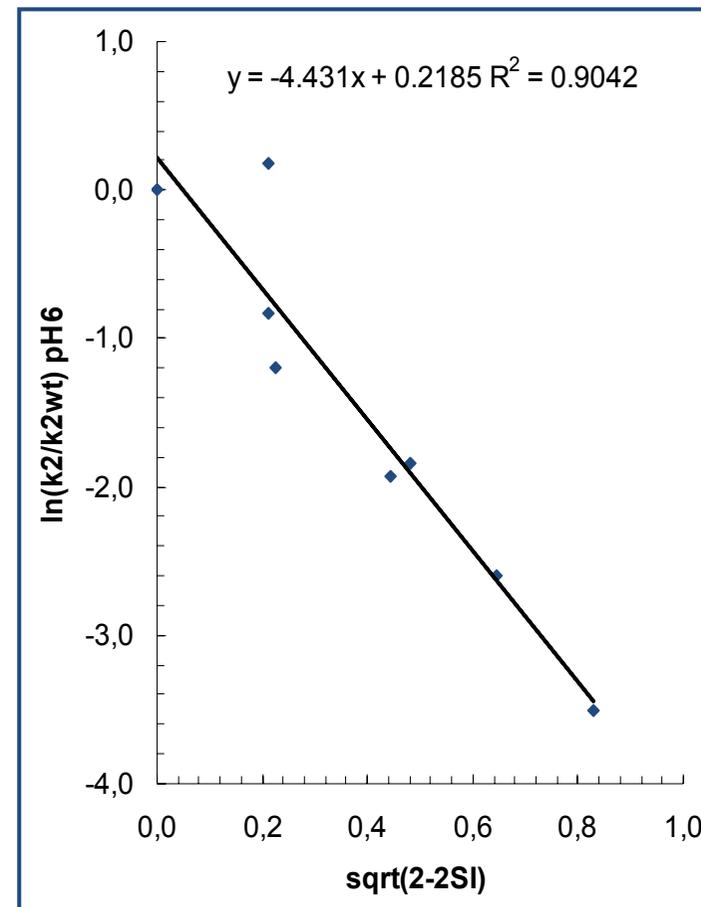
Cupredoxin electron transfer rate



Cupredoxin electron transfer rate



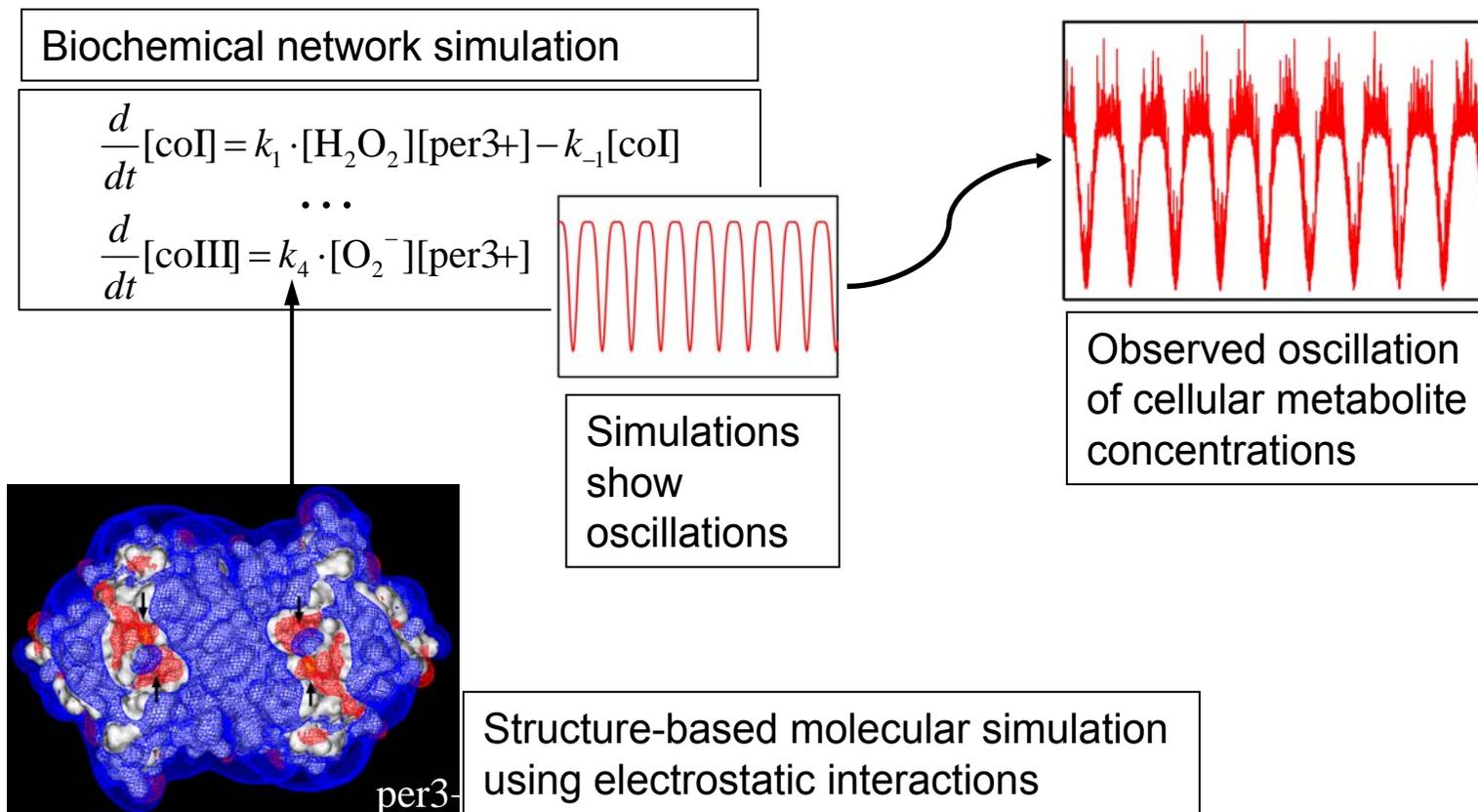
plastocyanin/cytochrome f



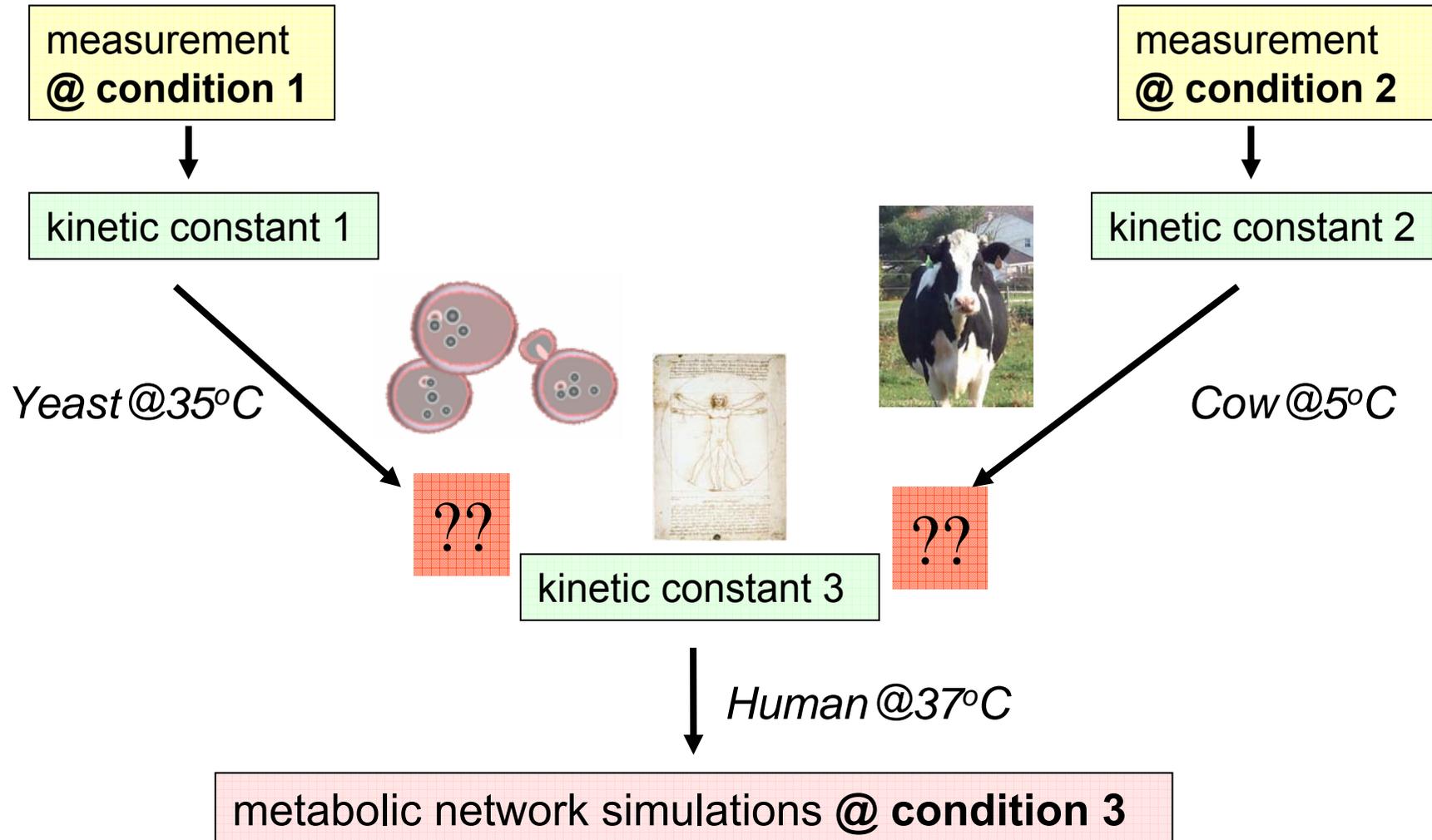
PIPSA: electrostatic potential

De Rienzo et al, *Biophys J.* (2001) 81, 3090-3104,
 Wade et al. *Int J Quant Chem* (2001) 83, 122-7

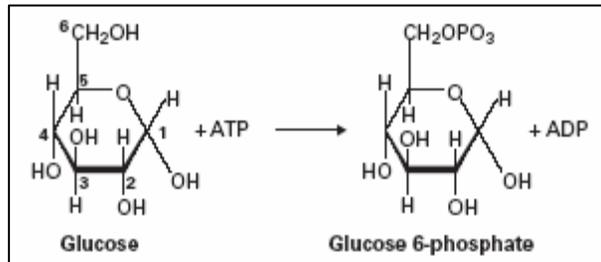
Bridging from molecular simulation to biochemical networks:



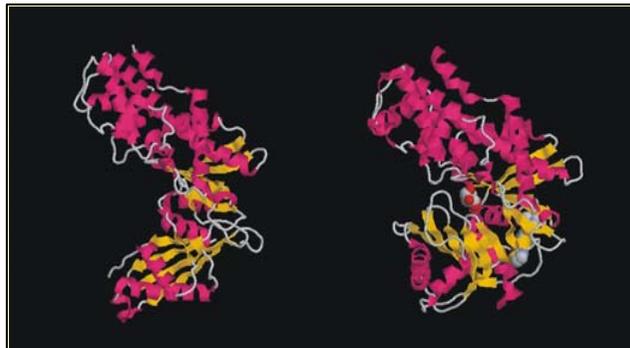
Similarity-Based Estimate of Kinetic Constants



Example qPIPSA: Hexokinases



1. Identify Chemical Reaction of Interest



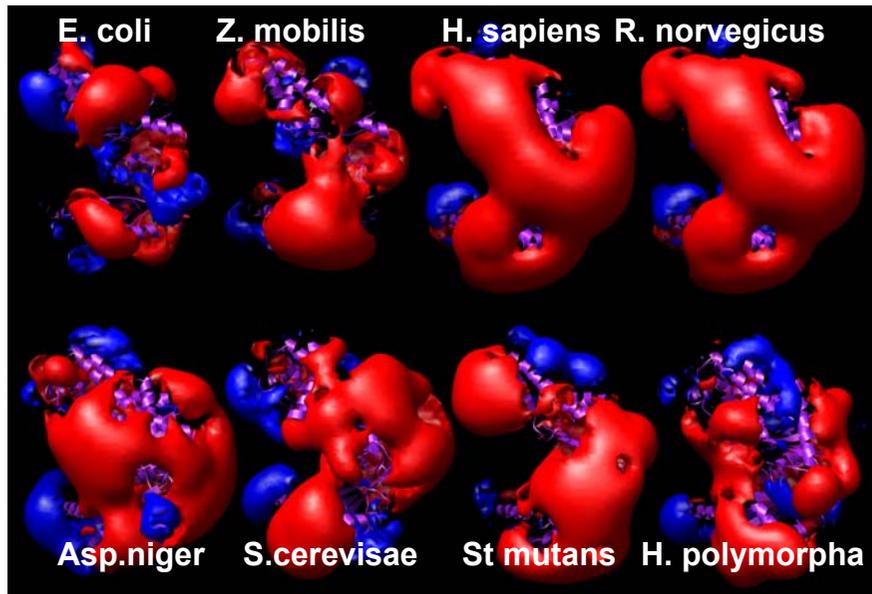
3. Choice of Appropriate Template

	10	20	30	40	50	60	70	80	90	100
HK4_Human_Iso_3	--MPPRSQIPQNS	-QVEQ	LAERFQLEEDLKKVMR	RMQKEMDRGURLETH	----	EEASVKMLRTYVRST	PEGSVGDG	ESLGLDGT	HFVLMVKVGGEE	EQ
HK4_Human_Iso_2	MAMDVTRSQDQALT	-LVQD	LAERFQLEEDLKKVMR	RMQKEMDRGURLETH	----	EEASVKMLRTYVRST	PEGSVGDG	ESLGLDGT	HFVLMVKVGGEE	EQ
HK4_Human_Iso_1	--MLDDBARMEAAKRYE	VEQ	LAERFQLEEDLKKVMR	RMQKEMDRGURLETH	----	EEASVKMLRTYVRST	PEGSVGDG	ESLGLDGT	HFVLMVKVGGEE	EQ
HK4_Rat_Iso_1	--MLDDBARMEATKRYE	VEQ	LAERFQLEEDLKKVMR	RMQKEMDRGURLETH	----	EEASVKMLRTYVRST	PEGSVGDG	ESLGLDGT	HFVLMVKVGGEE	EQ
HK4_Rat_Iso_3	--MAMDTRCGAOLLT	LVQD	LAERFQLEEDLKKVMR	RMQKEMDRGURLETH	----	EEASVKMLRTYVRST	PEGSVGDG	ESLGLDGT	HFVLMVKVGGEE	EQ
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	MSA	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	MSLD	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	MSFDLHKAT	ERAVIQAVQD	CCD	EVTPK	EDLTYA	FI	QMK	GL	APPK	KEHT
	110	120	130	140	150	160	170	180	190	200
HK4_Human_Iso_3	WSVYTKHQMYS	IPE	DAMTG	-----	FAEMERDY	ISECCIS	DF	LKHQMKHKKL	-----	PLDGT
HK4_Human_Iso_2	WSVYTKHQMYS	IPE	DAMTG	-----	FAEMERDY	ISECCIS	DF	LKHQMKHKKL	-----	PLDGT
HK4_Human_Iso_1	WSVYTKHQMYS	IPE	DAMTG	-----	FAEMERDY	ISECCIS	DF	LKHQMKHKKL	-----	PLDGT
HK4_Rat_Iso_1	WSVYTKHQMYS	IPE	DAMTG	-----	FAEMERDY	ISECCIS	DF	LKHQMKHKKL	-----	PLDGT
HK4_Rat_Iso_3	WSVYTKHQMYS	IPE	DAMTG	-----	FAEMERDY	ISECCIS	DF	LKHQMKHKKL	-----	PLDGT
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
	210	220	230	240	250	260	270	280	290	300
HK4_Human_Iso_3	LLRDAIKRR	----	GF	FEMDVVAMVND	VATMIS	CT	EDHQ	-----	CEVGM	IV
HK4_Human_Iso_2	LLRDAIKRR	----	GF	FEMDVVAMVND	VATMIS	CT	EDHQ	-----	CEVGM	IV
HK4_Human_Iso_1	LLRDAIKRR	----	GF	FEMDVVAMVND	VATMIS	CT	EDHQ	-----	CEVGM	IV
HK4_Rat_Iso_1	LLRDAIKRR	----	GF	FEMDVVAMVND	VATMIS	CT	EDHQ	-----	CEVGM	IV
HK4_Rat_Iso_3	LLRDAIKRR	----	GF	FEMDVVAMVND	VATMIS	CT	EDHQ	-----	CEVGM	IV
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
	310	320	330	340	350	360	370	380	390	400
HK4_Human_Iso_3	DE	F	LL	EYDR	LV	DESS	AN	PQ	QLY	EL
HK4_Human_Iso_2	DE	F	LL	EYDR	LV	DESS	AN	PQ	QLY	EL
HK4_Human_Iso_1	DE	F	LL	EYDR	LV	DESS	AN	PQ	QLY	EL
HK4_Rat_Iso_1	DE	F	LL	EYDR	LV	DESS	AN	PQ	QLY	EL
HK4_Rat_Iso_3	DE	F	LL	EYDR	LV	DESS	AN	PQ	QLY	EL
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
	410	420	430	440	450	460	470	480	490	500
HK4_Human_Iso_3	P	ST	TD	CD	V	R	AC	ES	V	T
HK4_Human_Iso_2	P	ST	TD	CD	V	R	AC	ES	V	T
HK4_Human_Iso_1	P	ST	TD	CD	V	R	AC	ES	V	T
HK4_Rat_Iso_1	P	ST	TD	CD	V	R	AC	ES	V	T
HK4_Rat_Iso_3	P	ST	TD	CD	V	R	AC	ES	V	T
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
	510	520	530	540	550	560	570	580	590	600
HK4_Human_Iso_3	CKK	AK	ME	GG	-----	-----	-----	-----	-----	-----
HK4_Human_Iso_2	CKK	AK	ME	GG	-----	-----	-----	-----	-----	-----
HK4_Human_Iso_1	CKK	AK	ME	GG	-----	-----	-----	-----	-----	-----
HK4_Rat_Iso_1	CKK	AK	ME	GG	-----	-----	-----	-----	-----	-----
HK4_Rat_Iso_3	CKK	AK	ME	GG	-----	-----	-----	-----	-----	-----
HK4_Zymomonas	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Ecoli	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Streptococcus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aspargillus	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Aeropyrum	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Hansenula	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
HK4_Yeast	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

2. Multiple Sequence Alignment

Example qPIPSA: Hexokinases

4. Comparative Protein Structural Modelling and Assessment



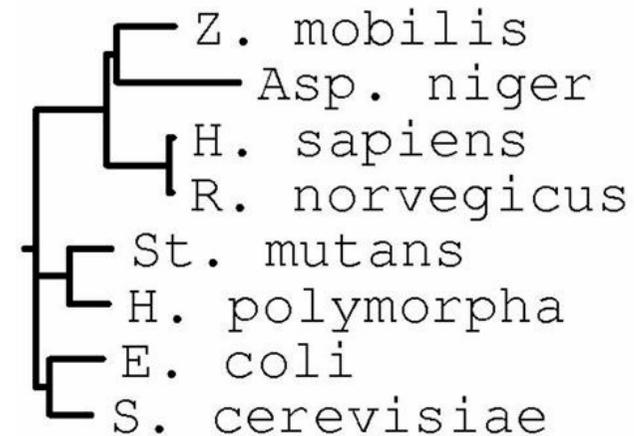
5. Calculation of Interaction Fields

E.Coli hexokinase:

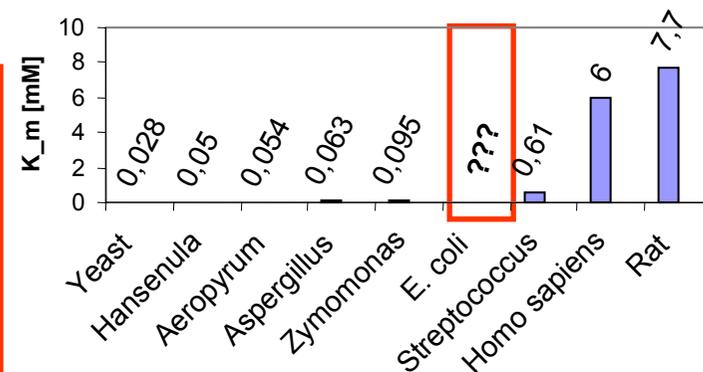
Brenda: Exptal values:
0.15-0.78mM

0.076mM (Raines & Millar (2004))

6. PIPSA



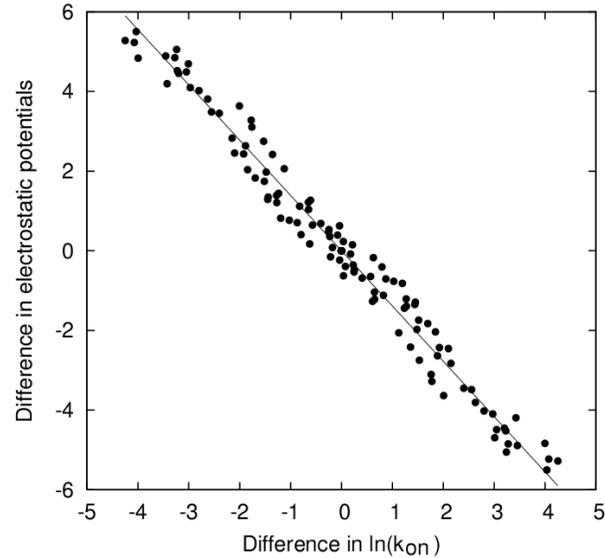
7. qPIPSA rate estimation



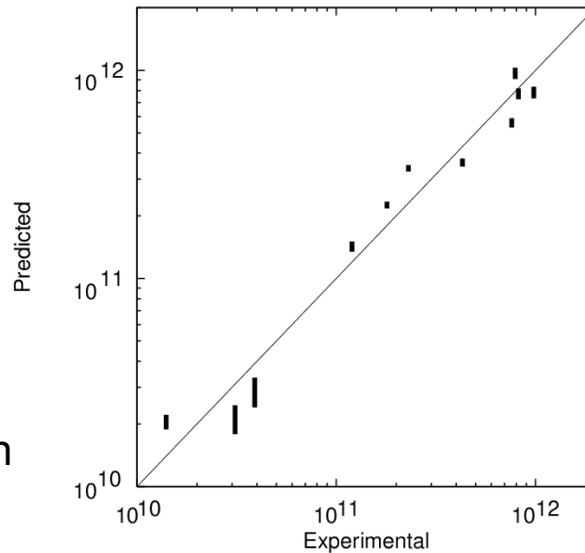
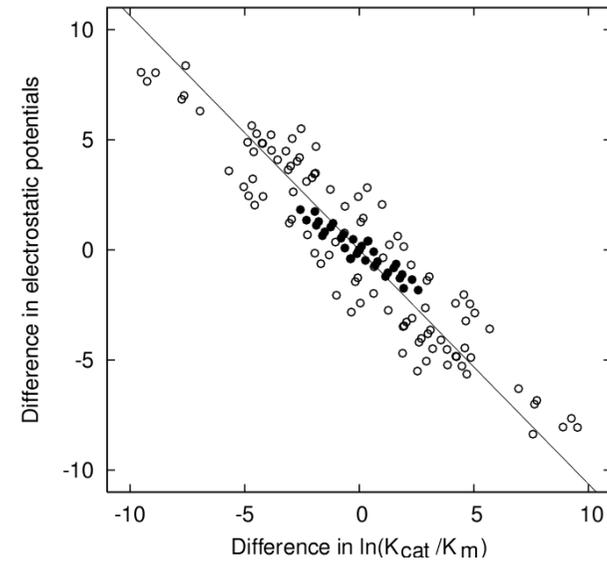
Acetylcholinesterase mutants:

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

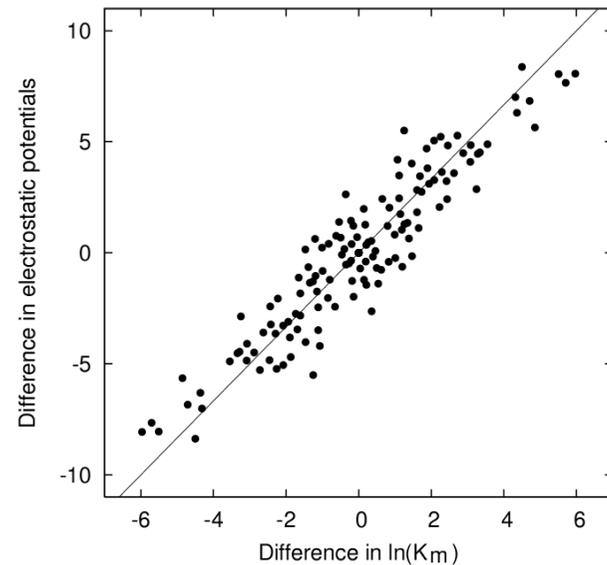
Inhibitor (TFK+)



Substrate (ACTh)

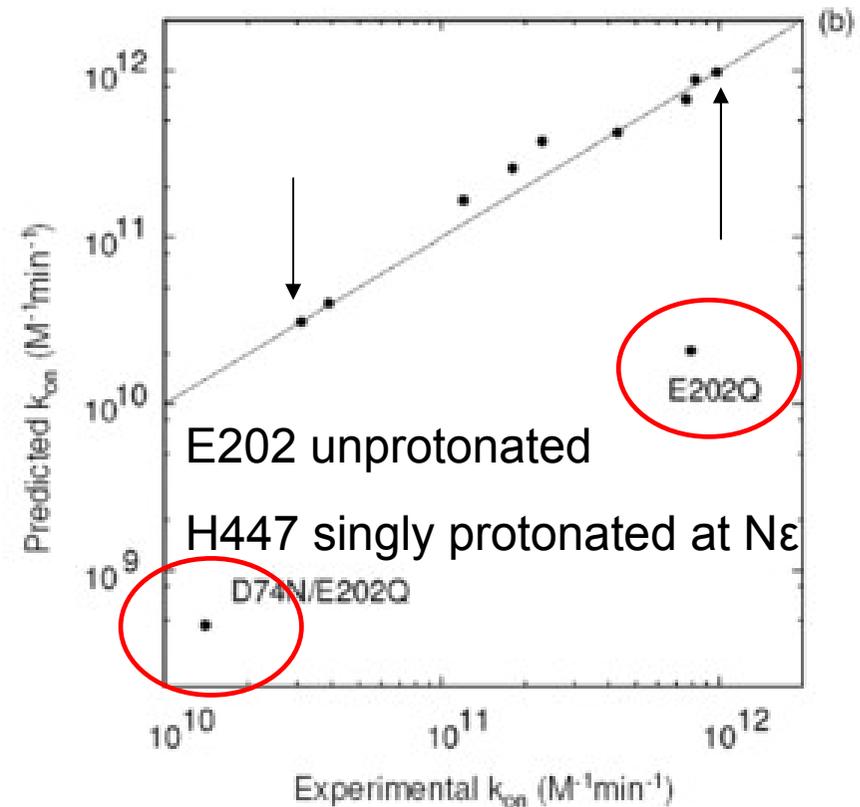
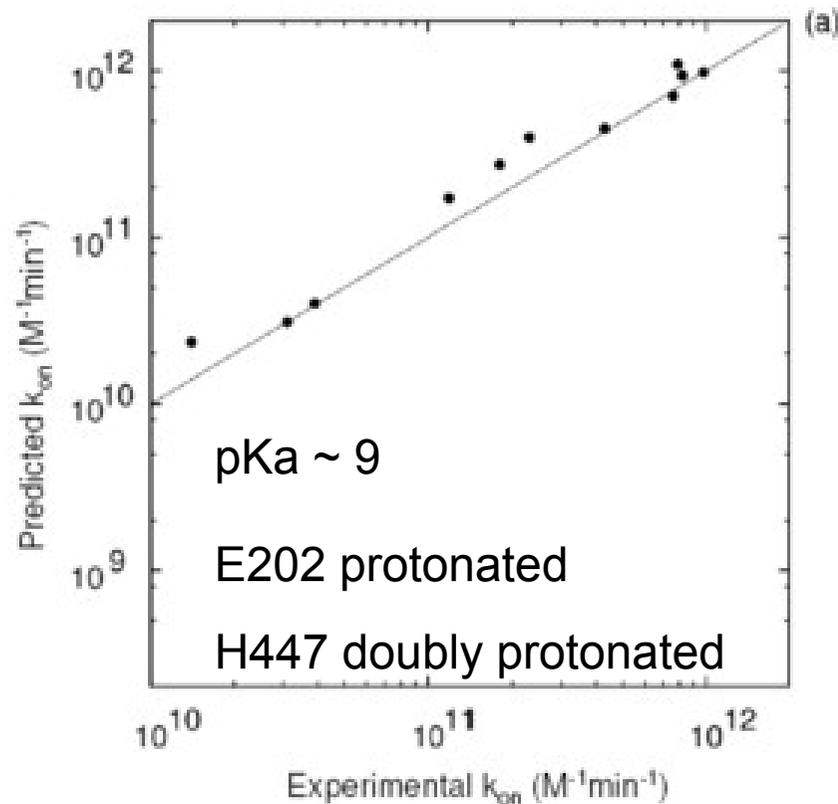


LTO cross-validation



AChE – Computation of TFK+ k_{on} for nine mutants

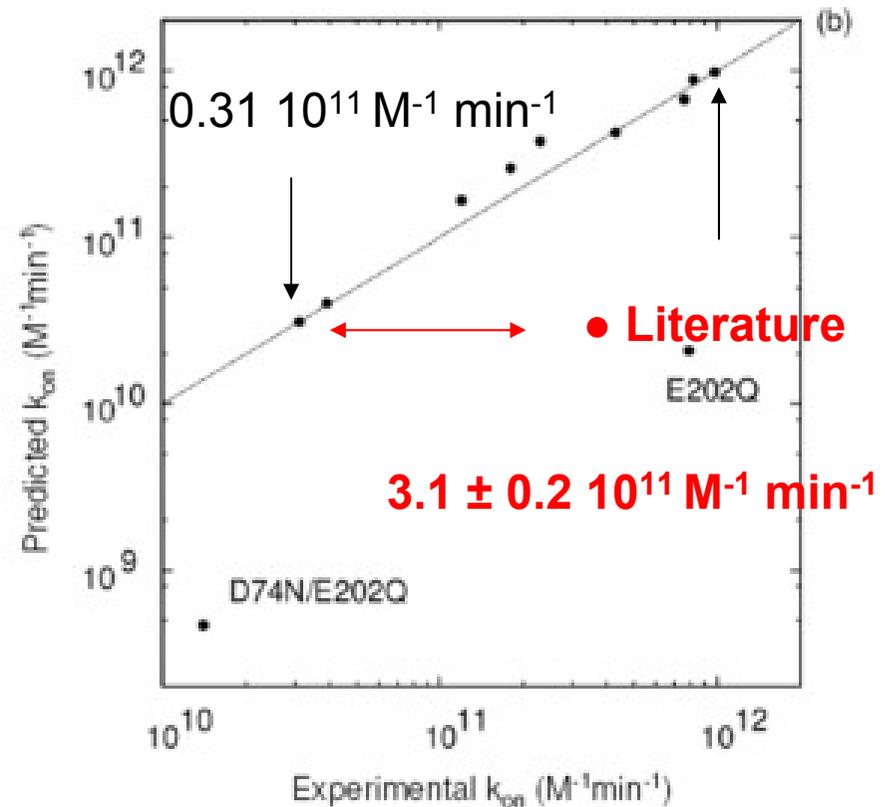
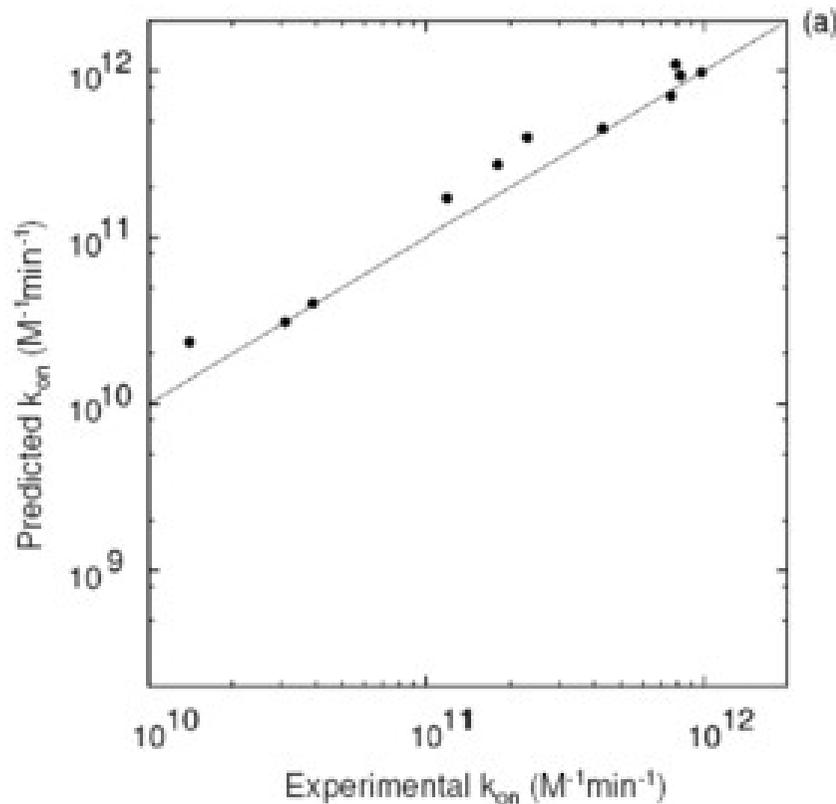
Correct modelling of protonation states is critical !



pH 7: Based on WT and one mutant D74N/D280V/D283N
(0.31 ± 0.02) $10^{11} M^{-1} min^{-1}$ Radic et al. JBC 1997.

AChE – Computation of TFK+ k_{on} for nine Mutants

Carefully check literature values!



pH 7: Based on WT (9.8 ± 0.04) $10^{11} \text{ M}^{-1} \text{ min}^{-1}$) and one mutant D74N/D280V/D283N (0.31 ± 0.02) $10^{11} \text{ M}^{-1} \text{ min}^{-1}$ Radic et al. JBC 1997.

Superoxide dismutase:

Ionic strength, pH and species dependence of k_{cat}/K_m

$$\ln(k_a / k_b) \sim \alpha \cdot \sum_R (\Phi_a - \Phi_b) / \sum_R 1$$

R: <10 Å from catalytic Cu

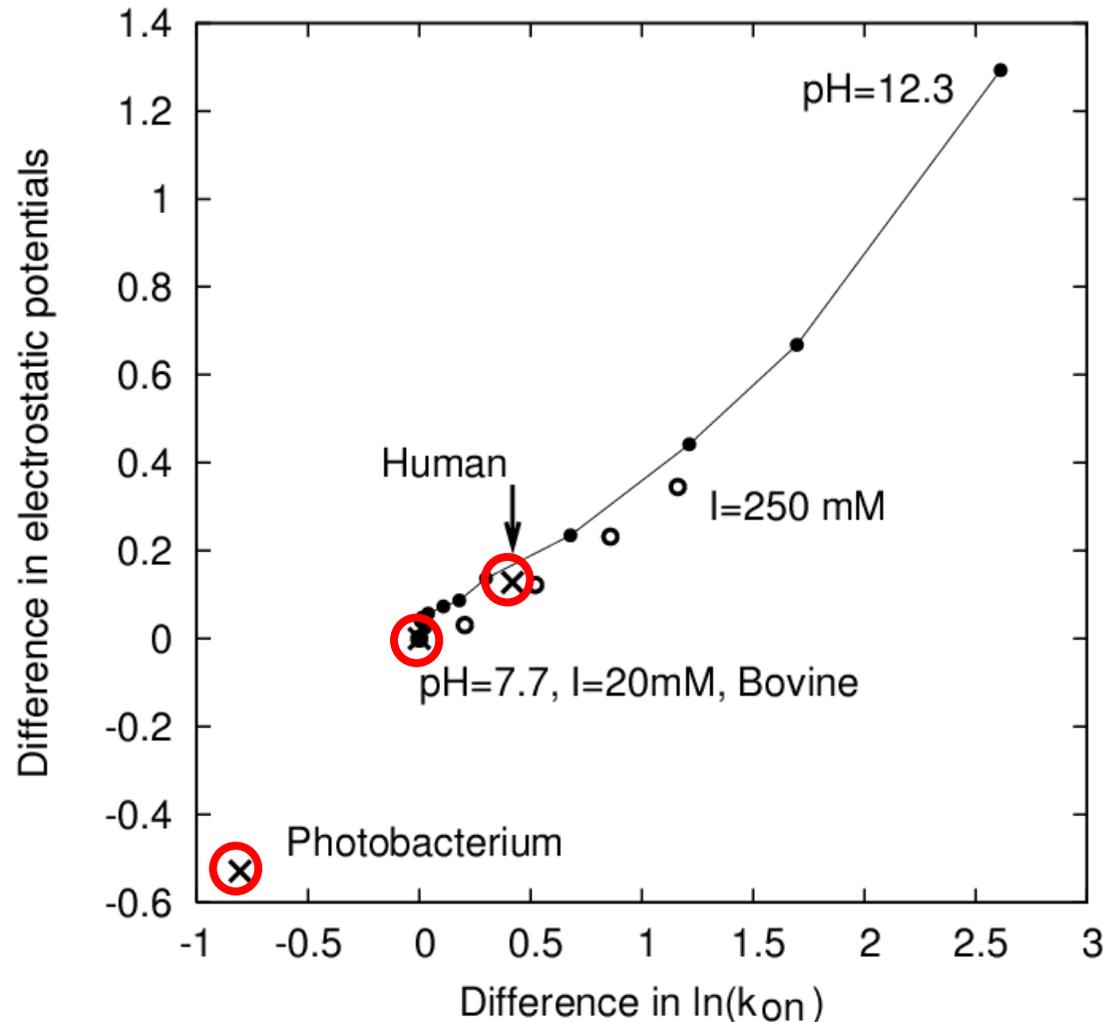
$$y = 0.45 \cdot x$$

(Rcoeff:0.97)

Ionic strength only:

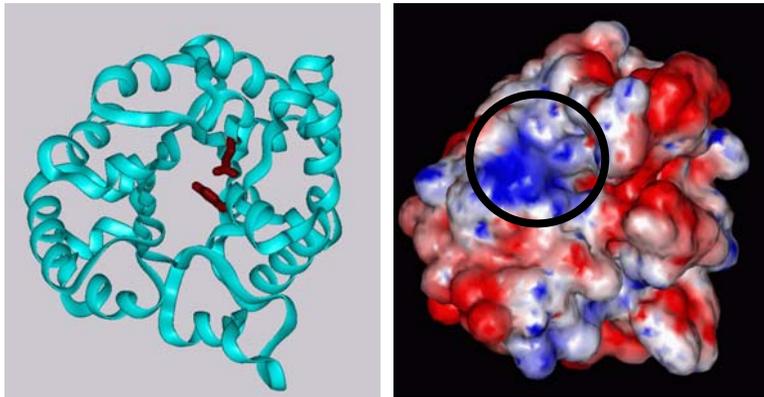
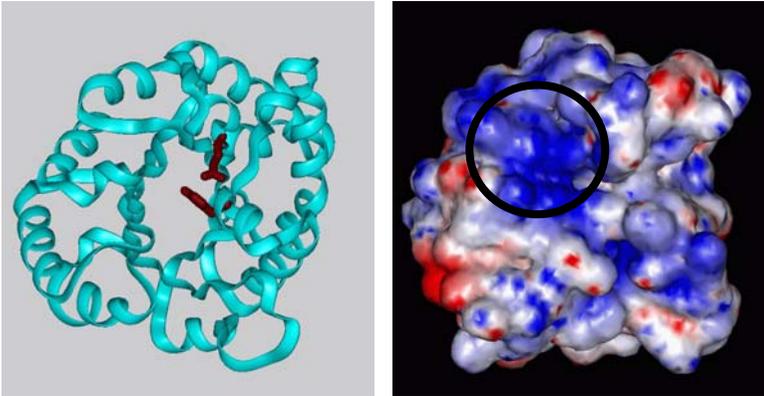
$$y = 0.3 \cdot x$$

(Rcoeff:0.99)



Triose Phosphate Isomerase

T. brucei



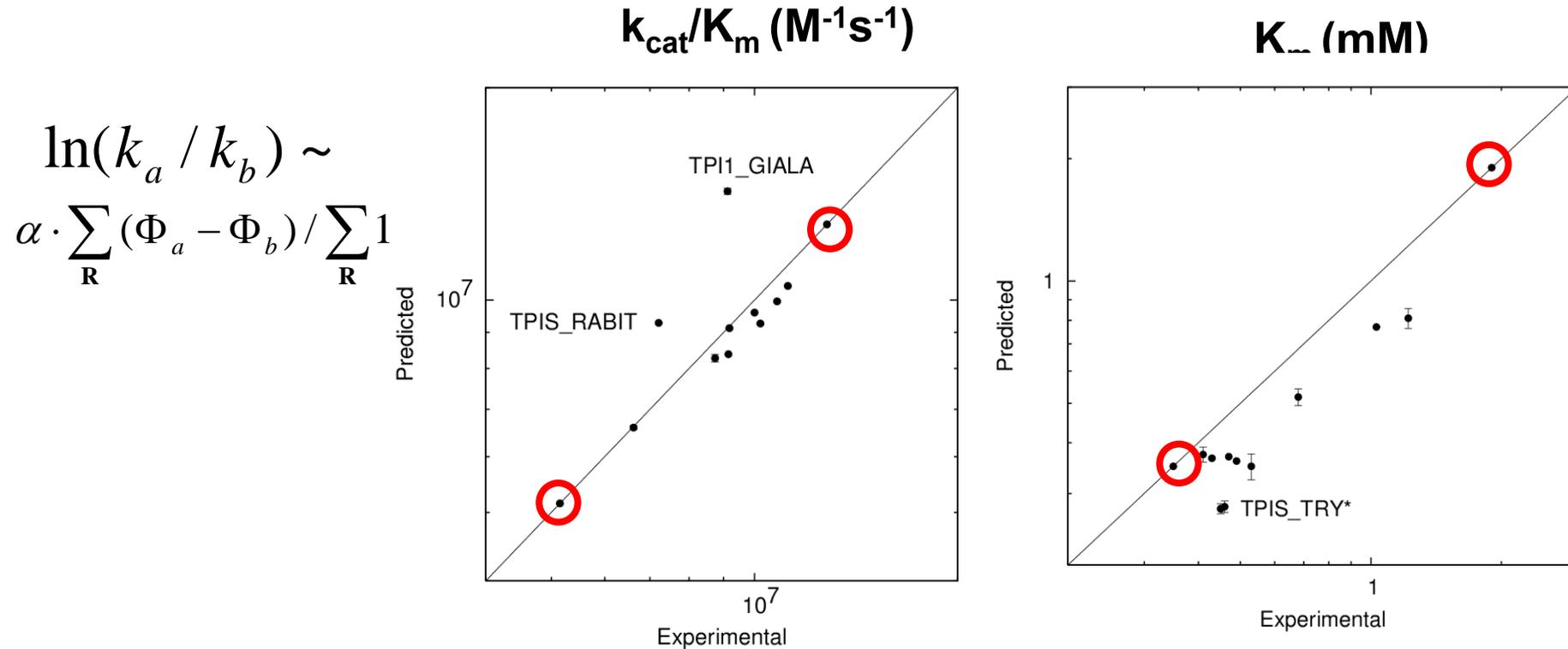
V. marinus

- 40/55% sequence identity/homology
- same fold
- very similar active site
- factor of 3 difference in k_{cat}/K_m

12 species:

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

Triose Phosphate Isomerase



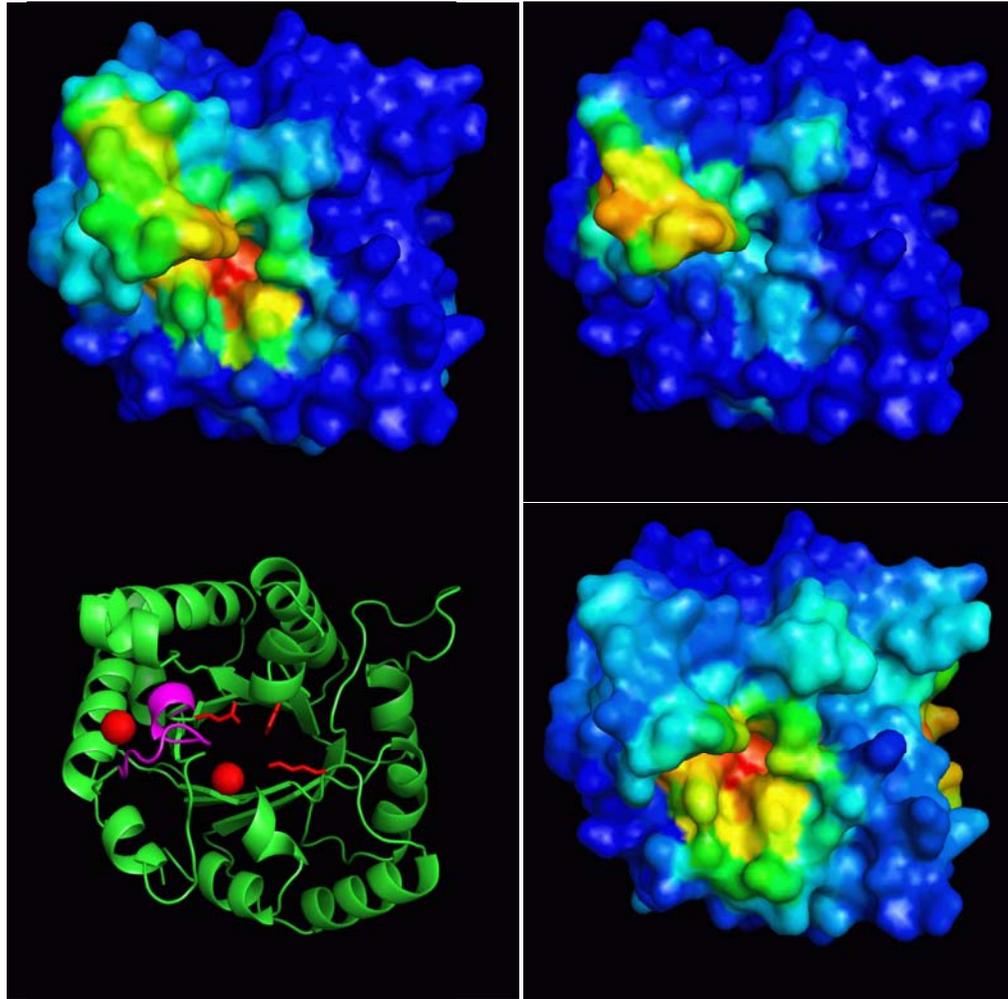
Predictions for 10 TPIs for the substrate glyceraldehyde-3-phosphate based on experimental measurements for the two TPIs from *V. marinus* (TPIS_VIBMA) and *P. falciparum* (TPIS_PLAFA)

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

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

Triose Phosphate Isomerase: Different Sites of Action

$$k_{cat} / K_m$$



$$K_m$$

*Electrostatic
potential
conservation*

blue → red
better predictions/conservation



SYCAMORE: Systems biology's Computational Analysis and Modeling Research Environment sycamore.eml.org

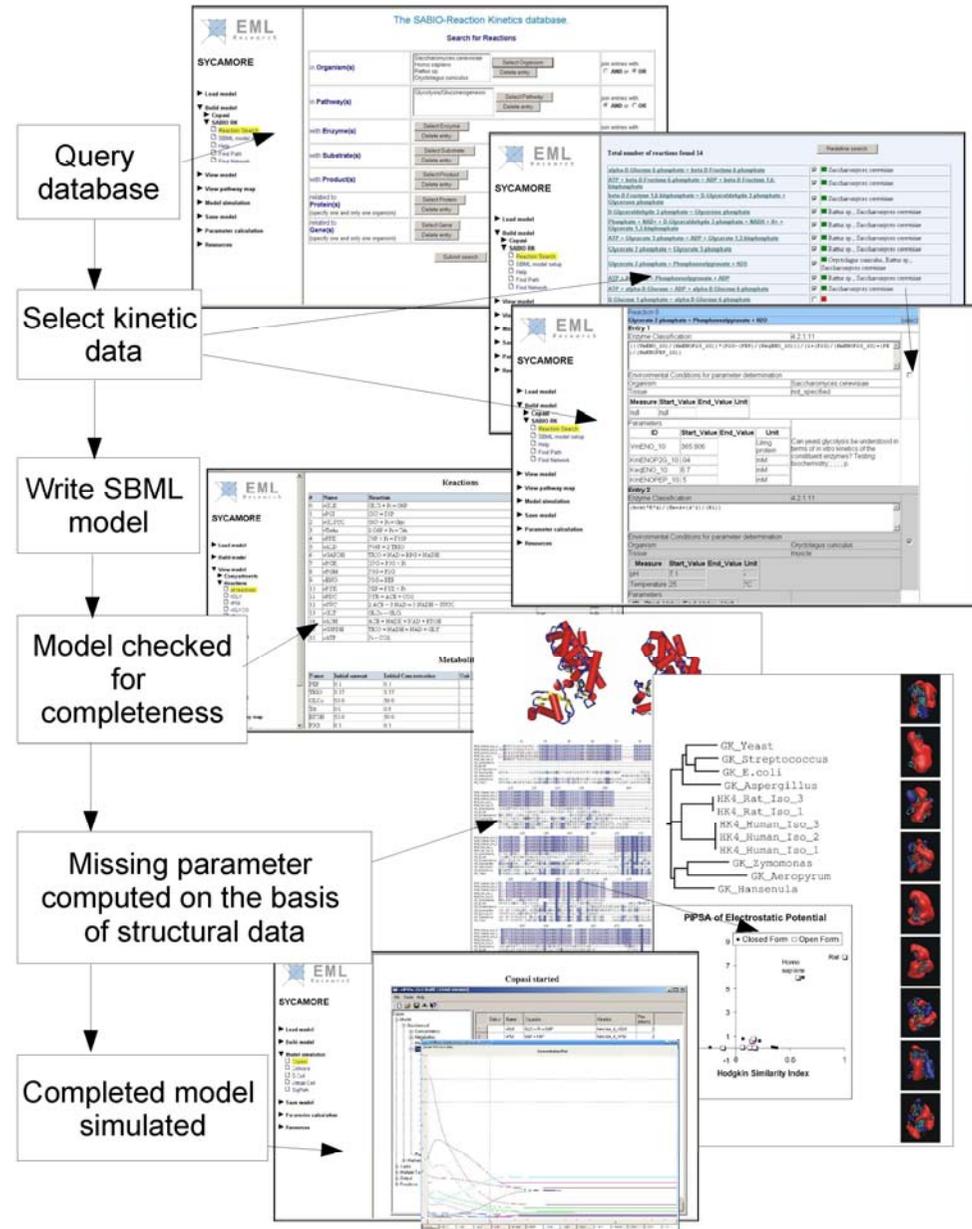


Fig 1: Example of using the available features in SYCAMORE. The case study is modeling and simulating glycolysis in hepatocytes. First, the database is queried and the relevant kinetic data selected. Then the SBML model file is created. This is checked for completeness (this is not implemented for automatic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

Questions:

pipsa.embl.org

- What are the binding properties of a set of structurally related proteins?
 - ◆ Classification w.r.t. binding properties
 - ◆ Detection of similarities/differences, e.g. for design of selective agents
 - ◆ Identification of regions of similarity/difference

→ PIPSA

- If experimental data on binding affinities or kinetics of a ligand are available for 1 or several proteins, can this be used to estimate values for a similar protein of interest?
 - ◆ Correlate properties from protein structures with binding and kinetic data.

→ qPIPSA



The MCMers



L → R:

Divita Garg

Outi Salo-Ahen

Stefan Richter

Stefan Henrich

Erika Nerini

Rebecca Wade

Vlad Cojocaru

Georgi Pachov

Anne Wenzel

Matthias Stein

Michael Martinez

Razif Gabdoulline

+ Oumarou Samna

+ Manohar Pilli

+ Anna Feldman-Salit

+ Bingding Huang

+ Daria Kokh

+ Matthias Janke

\$\$\$: Klaus Tschira Foundation, EU, DAAD, AVH, BMBF, BIOMS, DFG