PIPSA
Comparison of Protein Interaction Properties

Dr. Stefan Henrich
EML Research gGmbH
Molecular and Cellular Modeling
Heidelberg

stefan.henrich@eml-r.villa-bosch.de

http://www.eml-research.de/mcm

EML Research

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

http://sycamore.eml.org

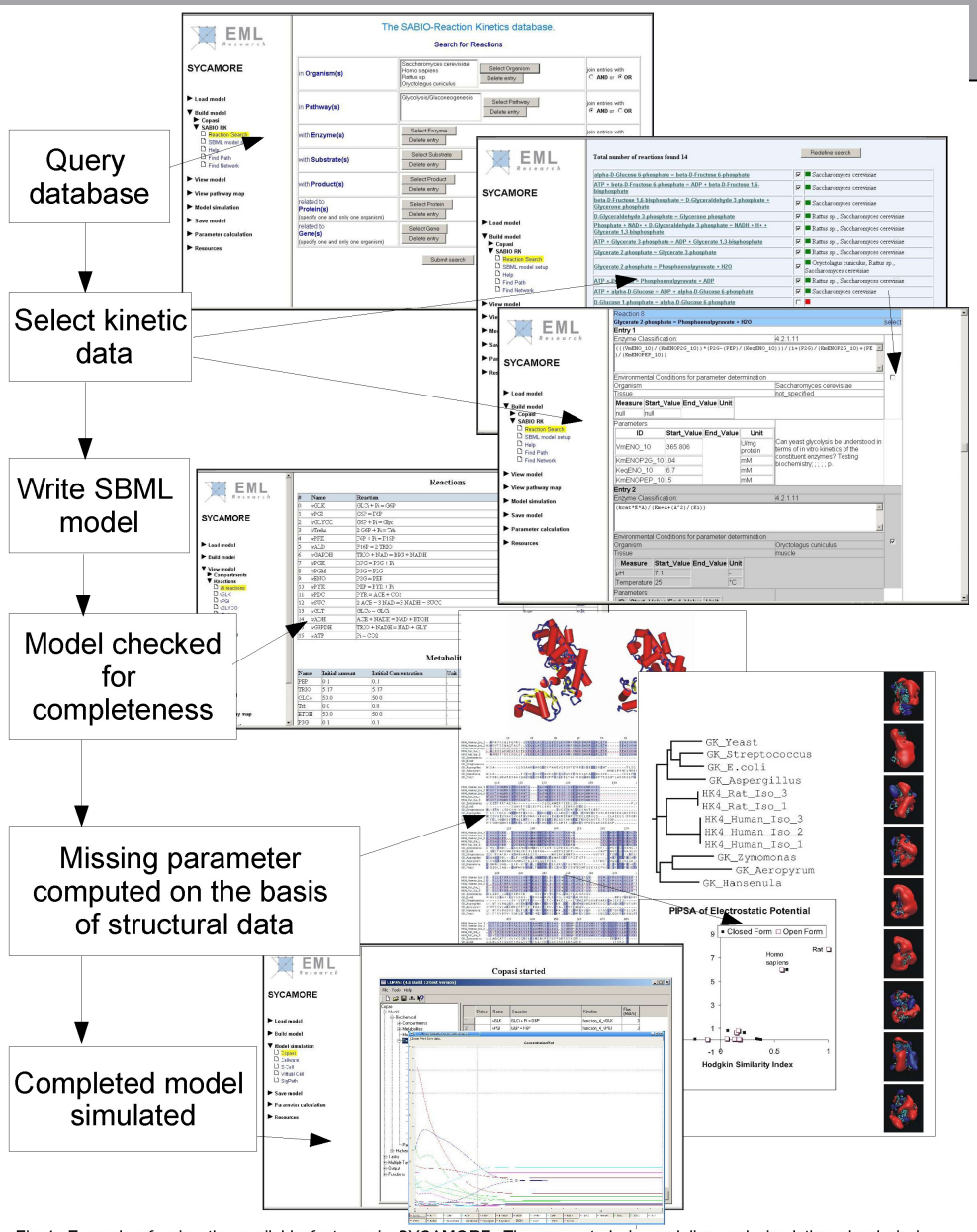


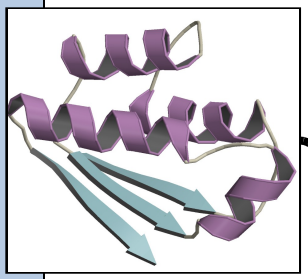
Fig 1: Example of using the available features in SYCAMORE. The case study is modeling and simulating glycolysis in hepatocytes. First, the database is queried and the relevant kinetic data selected. Then the SBML model file is created. This is checked for completeness (this is not implemented for automatic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

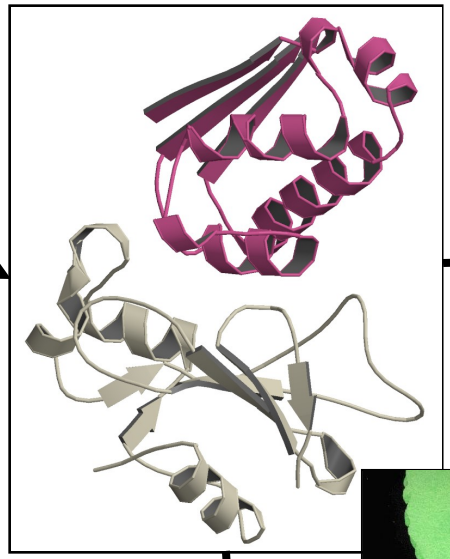
outlines

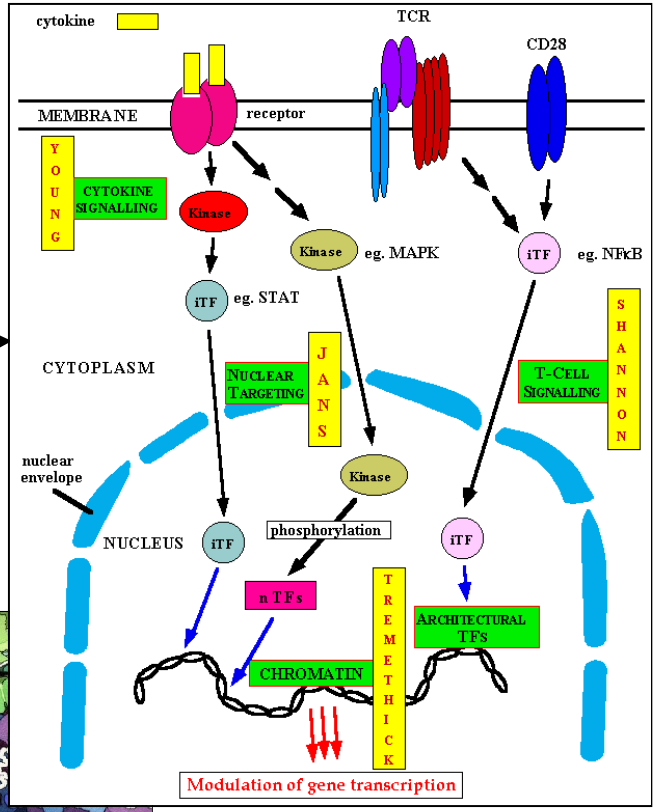
- introduction
- protein structures
- molecular interaction fields
- electrostatic potentials
- PIPSA
- webPIPSA
- visualisation and analysis
- qPIPSA
- tutorial with examples

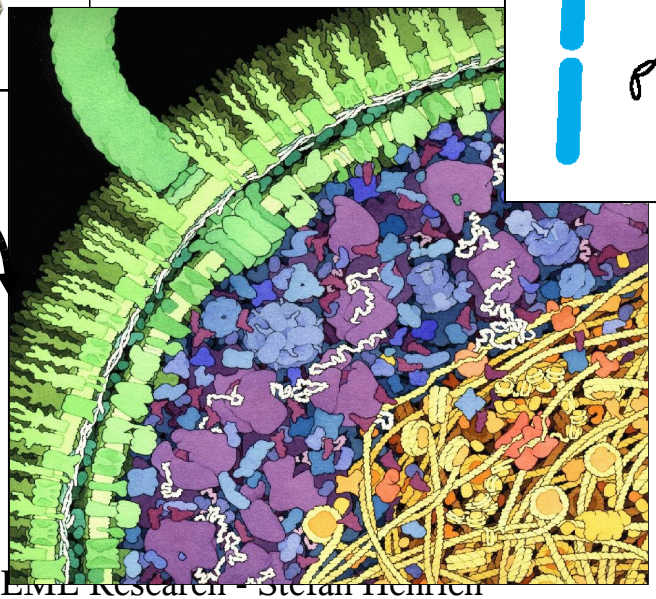
Introduction

From Individual Molecules to Cellular Systems



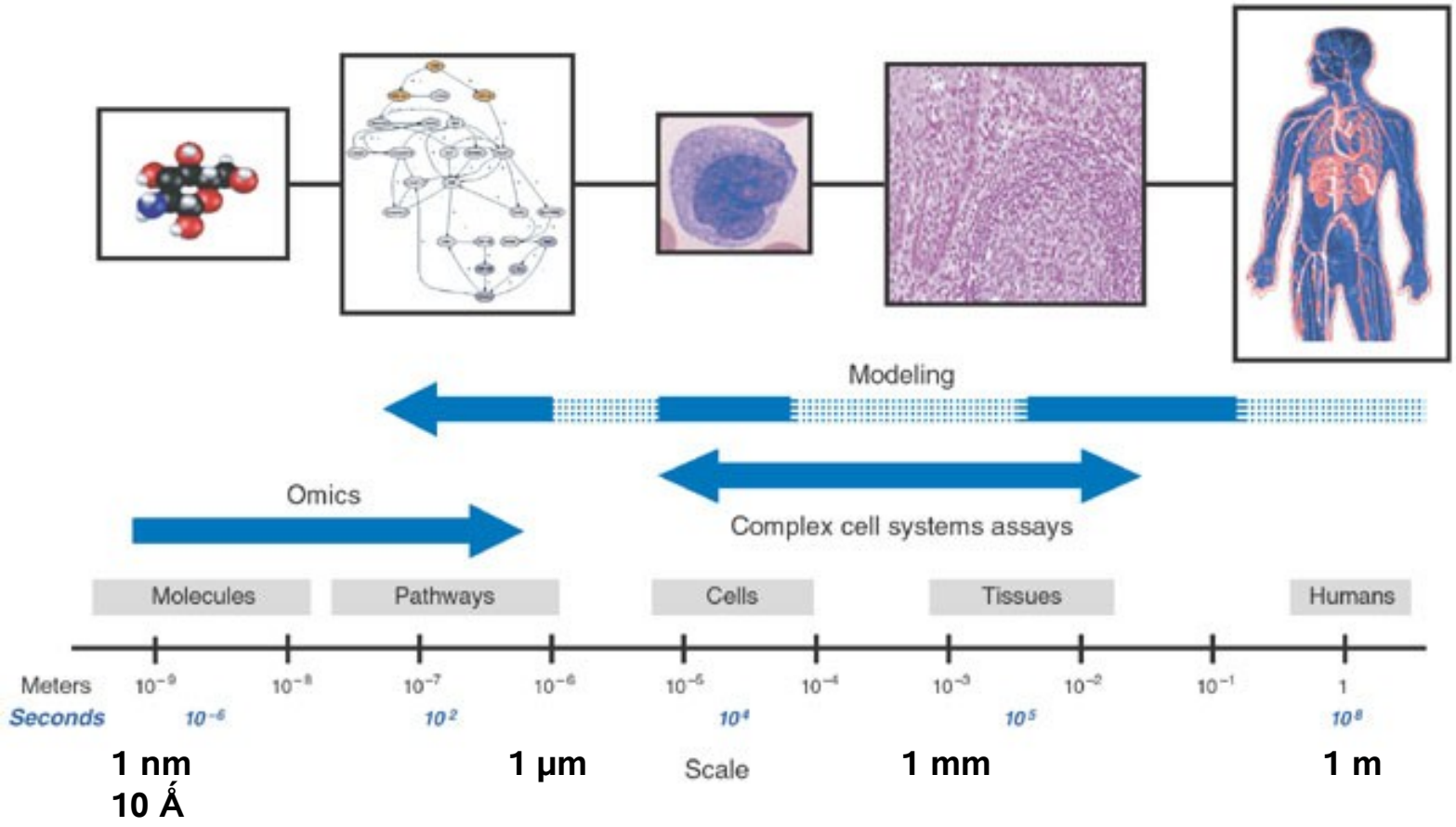






David Goodsell
in Voet,
Biochemie, 1998

dimensions



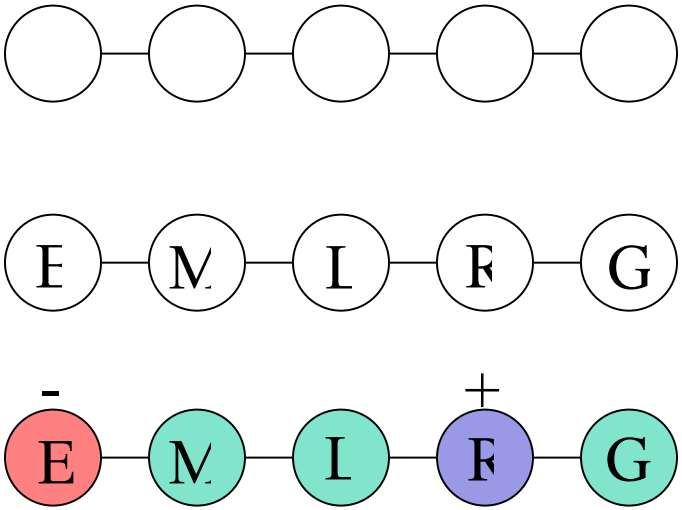
20 natural amino acids

Alanine	Ala	A	(Alanine)
Cysteine*	Cys	C	(Cysteine)
Aspartic Acid	Asp	D	(aciD)
Glutamic Acid	Glu	E	(E comes after D)
Phenylalanine	Phe	F	(Ph=F)
Glycine**	Gly	G	(Glycine)
Histidine***	His	H	(Histidine)
Isoleucine	Ile	I	(Isoleucine)
Lysine	Lys	K	(L follows K)
Leucine	Leu	L	(Leucine)
Methionine	Met	M	(Methionine)
Asparagine	Asn	N	(AsparagiNe)
Proline	Pro	P	(Proline)
Glutamine	Gln	Q	(Qlutamine)
Arginine	Arg	R	(aRginine)
Serine	Ser	S	(Serine)
Threonine	Thr	T	(Threonine)
Valine	Val	V	(Valine)
Tryptophan	Trp	W	(Double ring - W)
Tyrosine	Tyr	Y	(tYrosine)

Sidechain polarity:

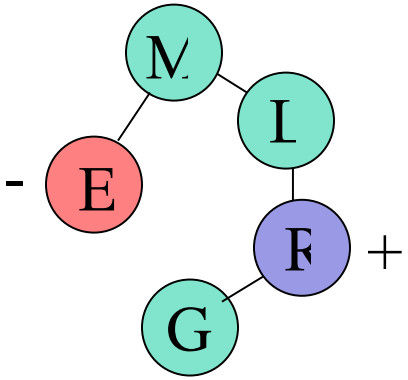
- Non-polar
- Polar
- negatively charged
- positively charged

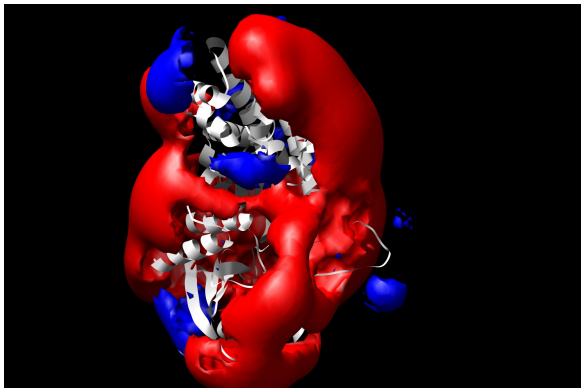
protein folding



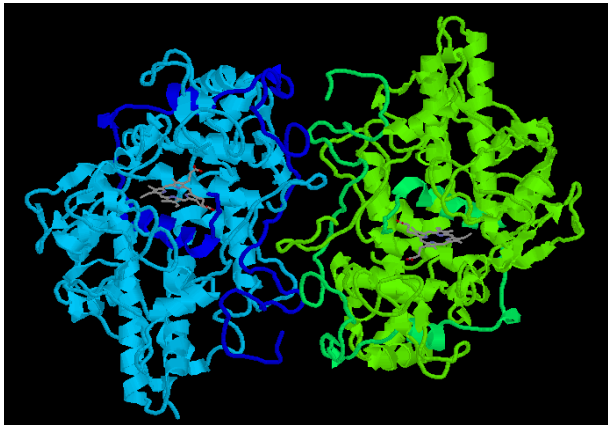
20
natural
amino
acids

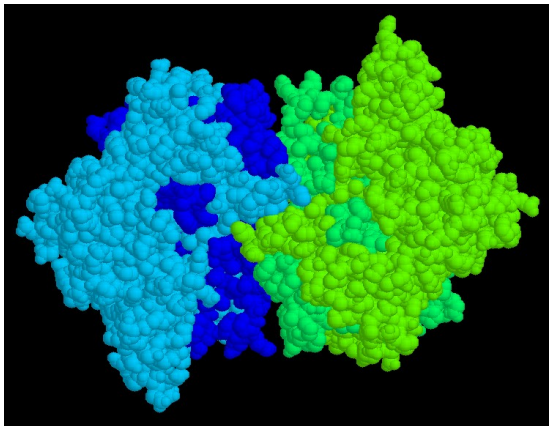
E = Glutamic Acid
M = Methionine
L = Lysine
R = Arginine
G = Glycine

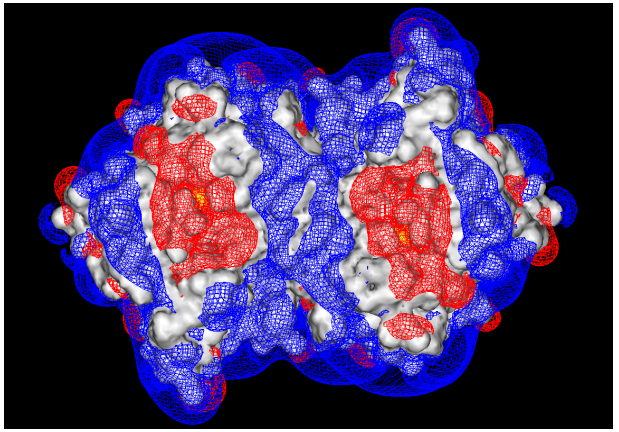


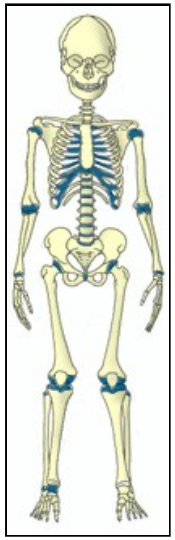


protein flexibility and biomolecular recognition



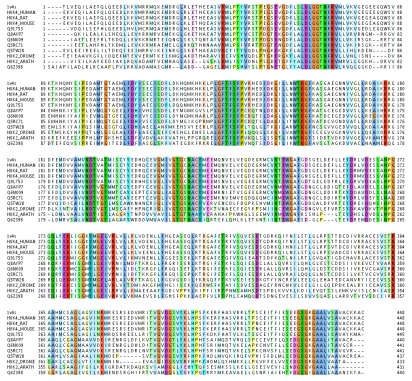


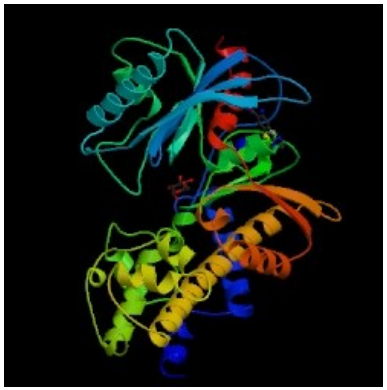


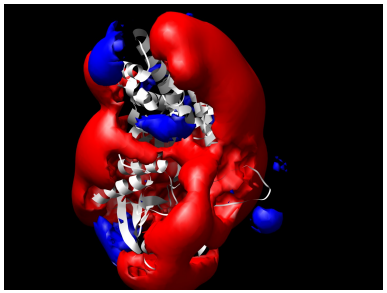


levels of protein comparison

- properties
  - mass, isoelectric point, charge, fingerprints,...
- amino acid sequence
  - bioinformatics (sequence alignment, prediction of motives)
- protein structures
  - X-ray, NMR, EM
- protein structure/function relationship
  - QSAR
  - molecular interaction fields (MIFs)







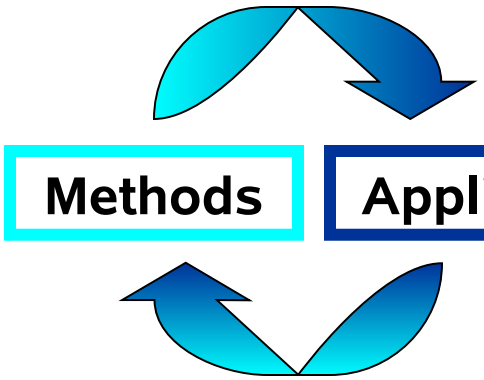
Exploring Protein Interactions via the Computer

- Small molecules
- Proteins
- Nucleic acids
- ….

Protein Interactions

- Stability
- Folding
- Catalysis
- Electron transfer
- ….

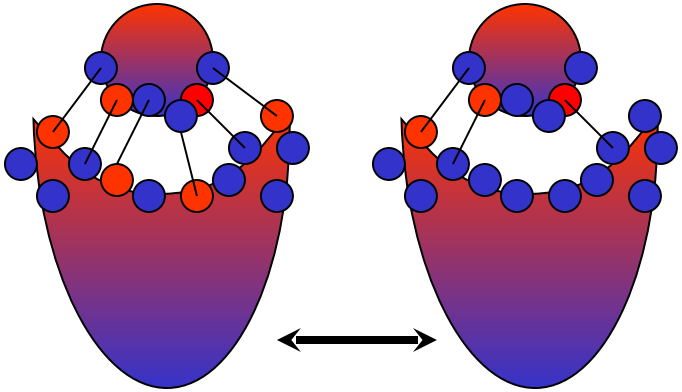
- Physics-based
- Bio/Chemo-Informatics



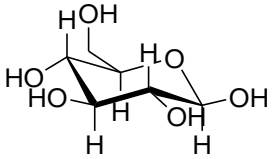
- Individual proteins – in-depth
- Proteins “in context” – biochemical pathways
- Comparison across large protein families

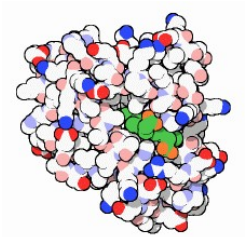
Molecular Interaction Fields

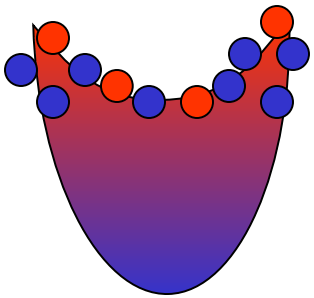
Comparing Molecular Interaction Fields



comparing interactions

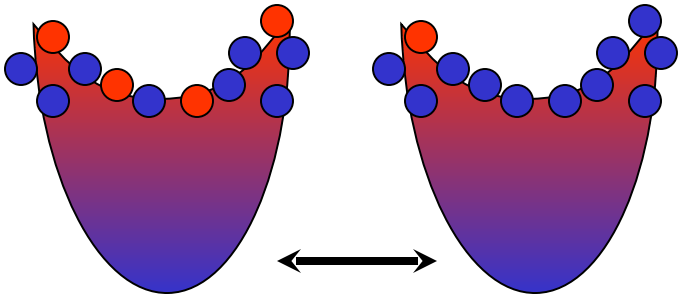






protein-ligand

protein-protein

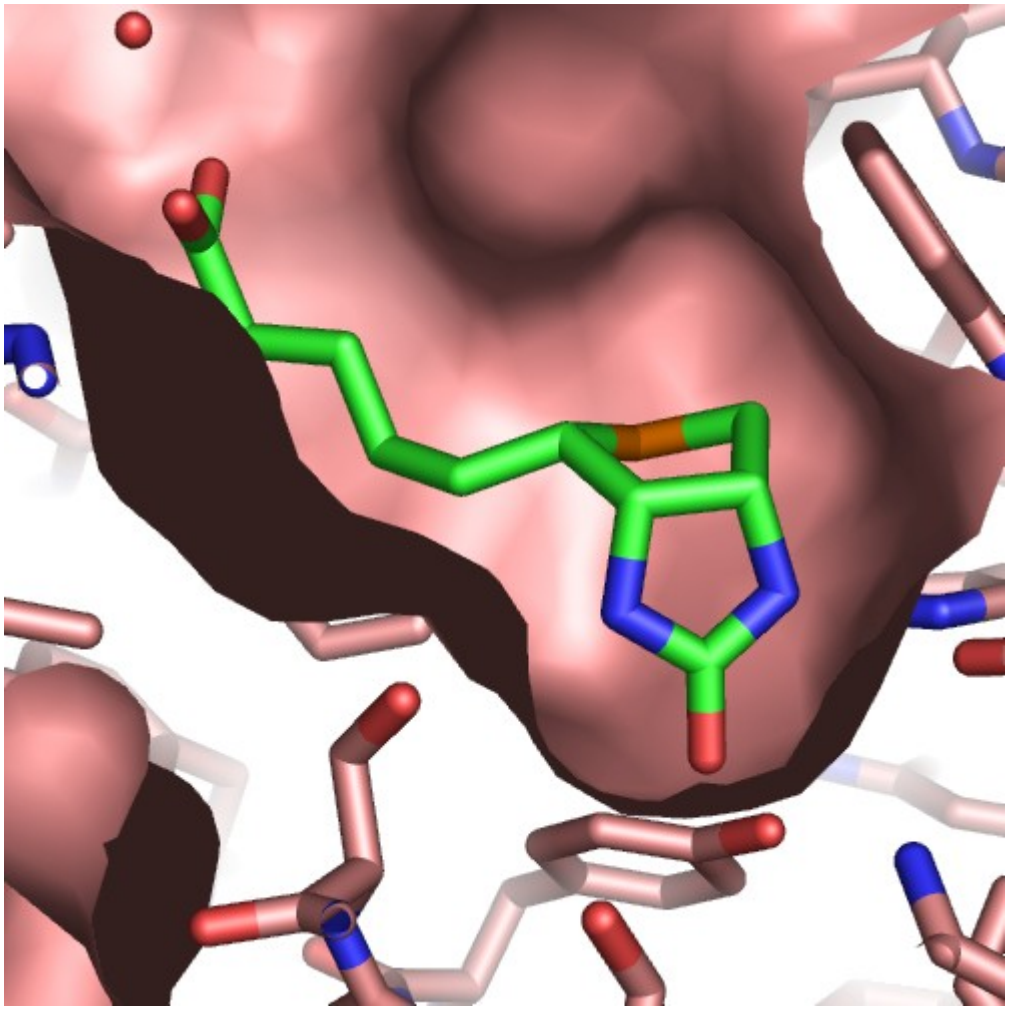


comparing interaction fields

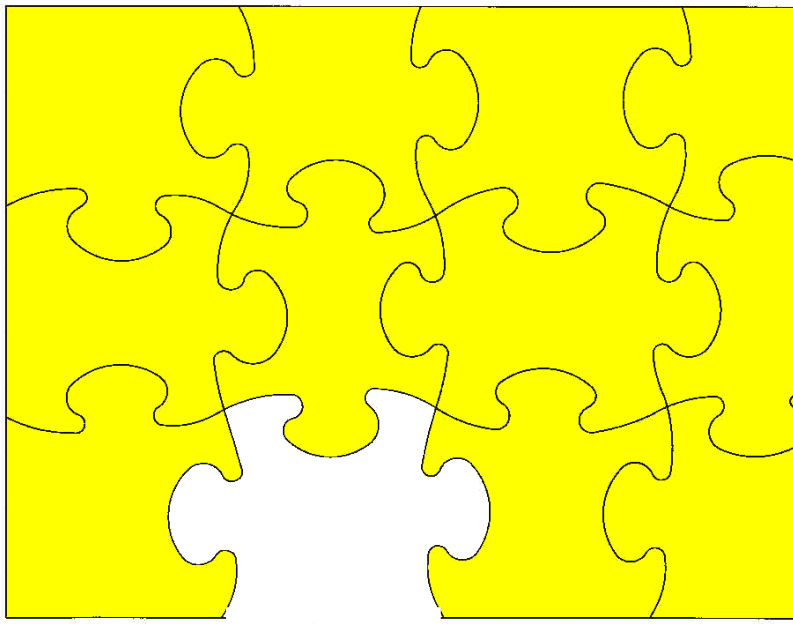
- Charge distribution
- Potential
- Shape
- ...

EML Research

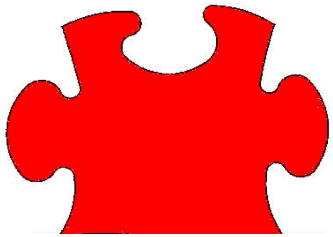
receptor-ligand binding



Streptavidin complex with biotin (1stp)

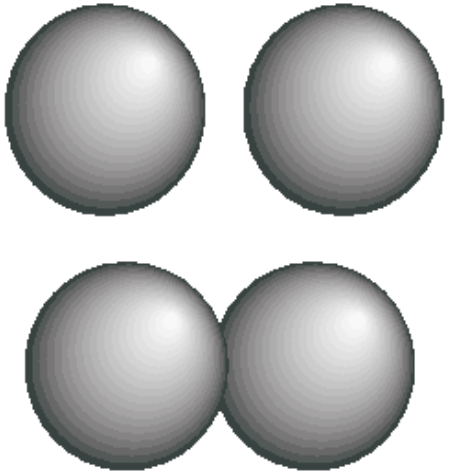


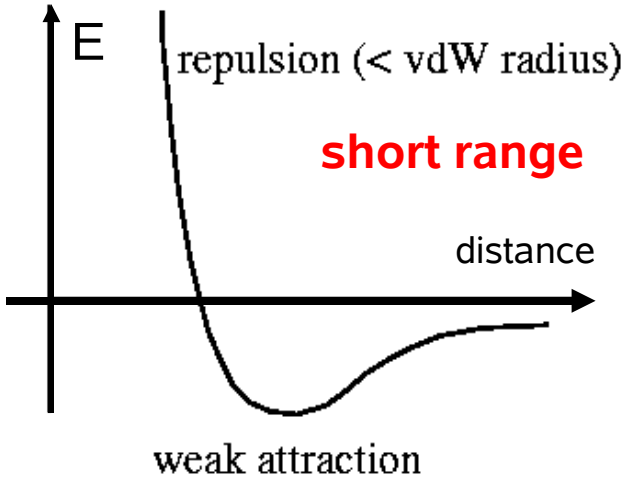
shape complementarity



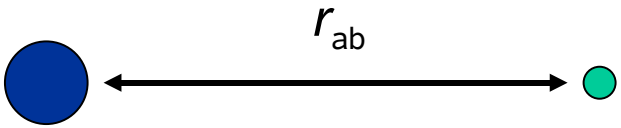
Lennard-Jones (van der Waals)

$$E_{LJ} = \frac{A}{r_{ab}^{12}} - \frac{B}{r_{ab}^{6}}$$



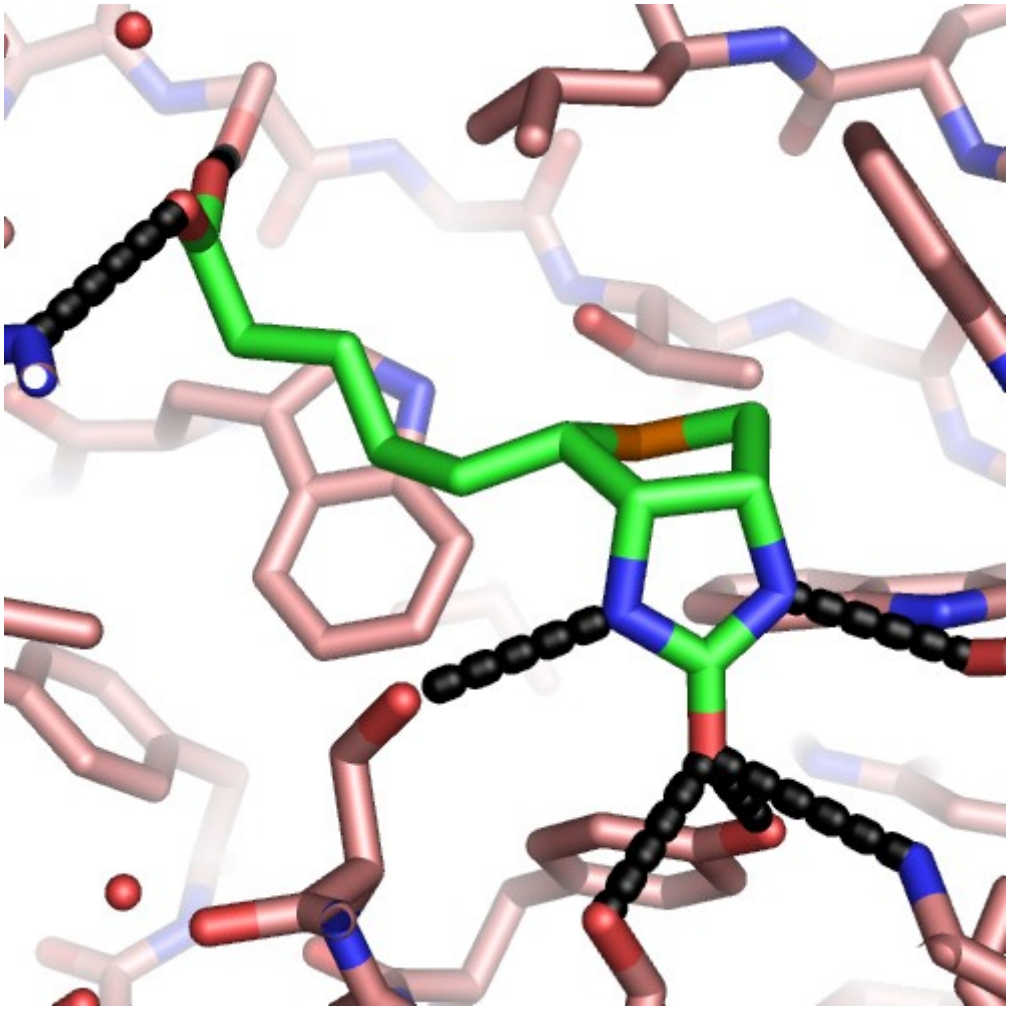


A = 0.5 C $(R_a + R_b)^6$
C: Slater-Kirkwood formula
R_a and R_b: van der Waals radii



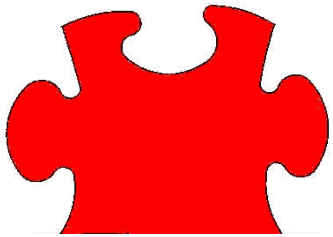
- van der Waals excluded volume repulsion
- London dispersive induced-dipole induced-dipole attraction
- **short-range** (8-10 Å cutoff)

receptor-ligand binding



Streptavidin complex with biotin (1stp)

electrostatic complementarity



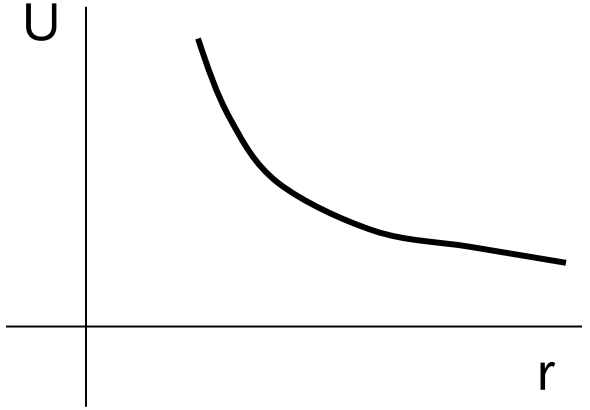
Coulomb's Law

- interaction energy of two point-charges *in vacuo*
- solve Poisson equation
- in SI units:

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

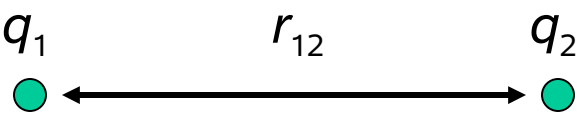
ε_0 = permittivity of free space

(ε_0 = 8.85 10^{-12} F m^{-1})



- in "Biomolecular" units:

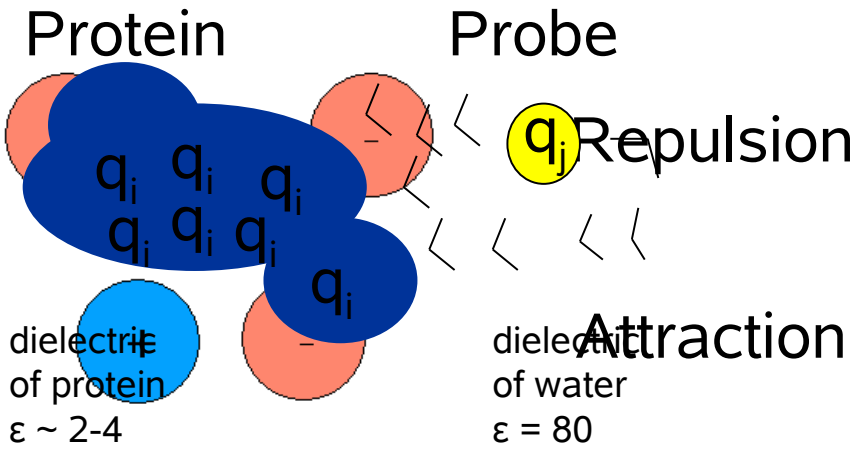
$$U = \frac{332 q_1 q_2}{r_{12}}$$

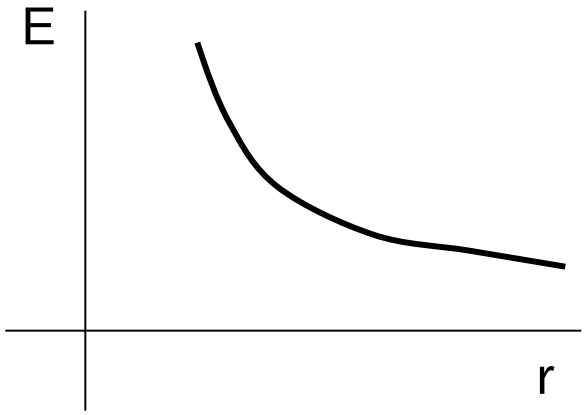


Energy, U: kcal/mol
Charge, q: electron charges
Distance, r: Ångstroms

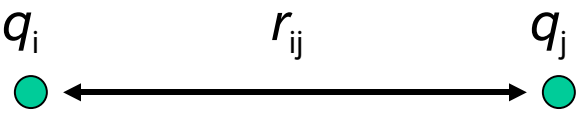
EML Research

Non-bonded interactions: Electrostatic





$$E_{ele} = \frac{q_i q_j}{\varepsilon \pi \varepsilon_0 \varepsilon_r r_{ij}}$$



- Coulombic forces
- Each atom has partial atomic charge q_1, q_2
- Compared to vacuum, the uniform medium weakens (damps/screens/shields) electrostatic interactions between charges
- ε_0: permittivity of free space,
  ε_r: relative dielectric constant of the surrounding medium
- **Long range** (~1/r) (No cutoff)

Polarization

- Electronic
  - Electrons and nuclei of an atom are shifted in opposite directions by the external field
  - Linearly proportional to applied field

    ε_r ~1.5-2.5

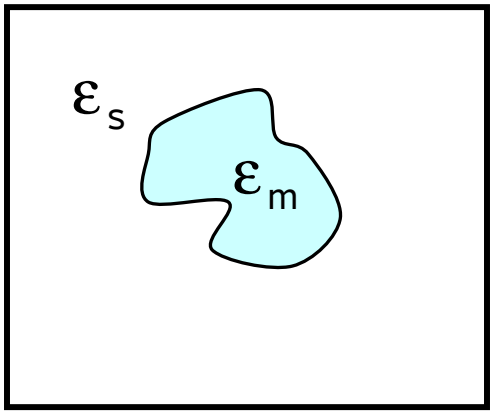
- Orientational
  - Molecules that have a permanent dipole moment, e.g. water (μ_D=1.9 D), align to oppose the external field
  - Average dipole moment rises linearly at weak fields, saturates at high fields

    ε_r up to ca. 80 (water)

Continuum electrostatics for molecules

- dielectric constant is a macroscopic property
- nevertheless, it can be applied microscopically
  - Molecule
  - Part of molecule
- for small molecules, e.g. ethanol, polarizability is primarily electronic
  - Molecule of ε_M~1.5-2.5
    surrounded by vacuum of ε_s~1 or water of ε_s~80 etc
- for macromolecules, orientational polarizability is also important

  ε_M~2-80, dependent on
  - location
  - theoretical model
- non-uniform dielectric
  - Dielectric boundaries



Effective dielectric constant

- non-uniform dielectric leads to distortion of field due to a point charge
- due to total dielectric environment
- effects such as electrostatic focusing

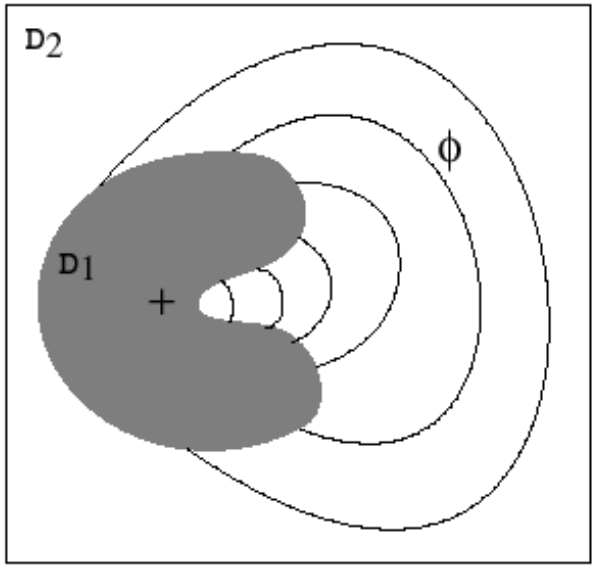


Figure: M. K. Gilson

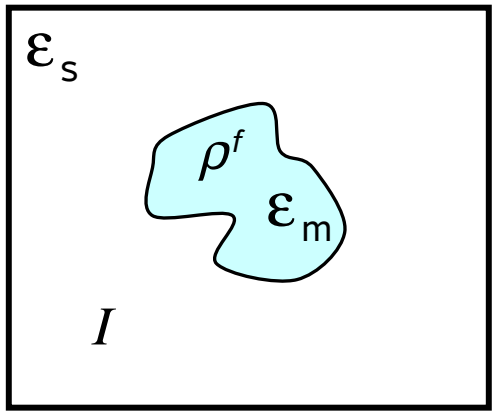
- Linearized Poisson-Boltzmann equation

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

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

- Finite-difference
- Numerical solution

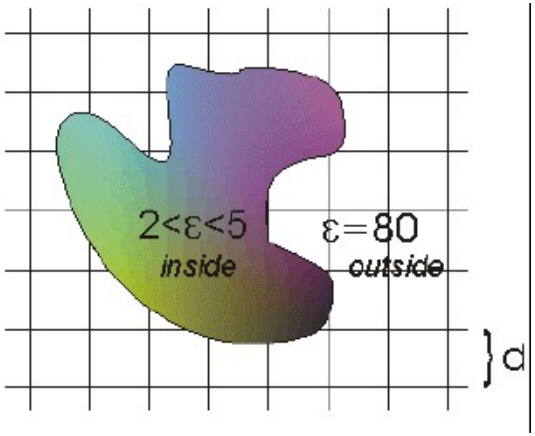
$\phi(\mathbf{r})$=Elec. Potential
$\rho(\mathbf{r})$=Charge density



Solving the Poisson-Boltzmann equation

- Analytic solutions
  - Only for simple geometries: spheres, planar boundaries

- Numerical solutions
  - Convert continuous partial differential equation into discretized problem with systems of linear equations that can be solved by matrix methods



Finite difference method

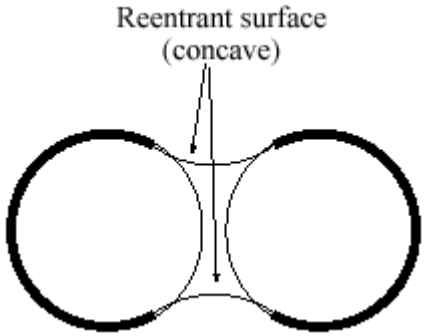
- Discretize space into a cubic lattice, solve iteratively to convergence for each grid point by finite differences

Solving the finite difference linearized Poisson-Boltzmann equation: Practical aspects

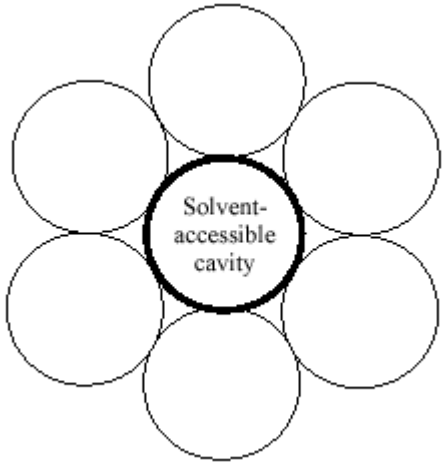
- Need to assign:
- Atomic radii and charges
  - From “standard” force field e.g. CHARMM, AMBER
  - Set specifically for continuum electrostatics, parameterized to reproduce small molecule solvation energies
- Grid size and spacing
  - Large enough for boundary potentials to be sufficiently accurate
    - Focusing using nested grids is possible
  - Spacing of 0.5-1 Å for viewing potentials or Brownian dynamics forces
  - Spacing of 0.2-0.3 Å for energies, pK_as

Solving the finite difference linearized Poisson-Boltzmann equation: Practical aspects

- Need to assign:
- Dielectric boundaries
  - Van der Waals surface
  - Molecular surface

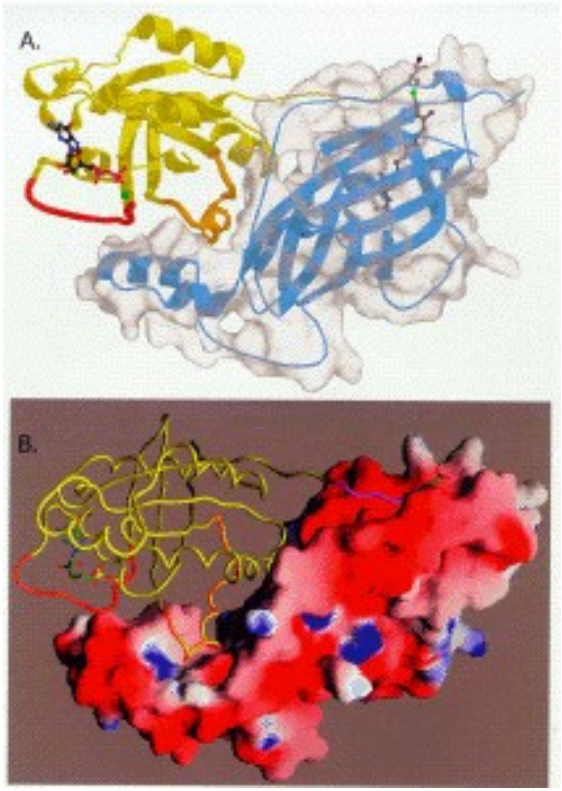


  - Smooth dielectric constant over points adjacent to the boundary

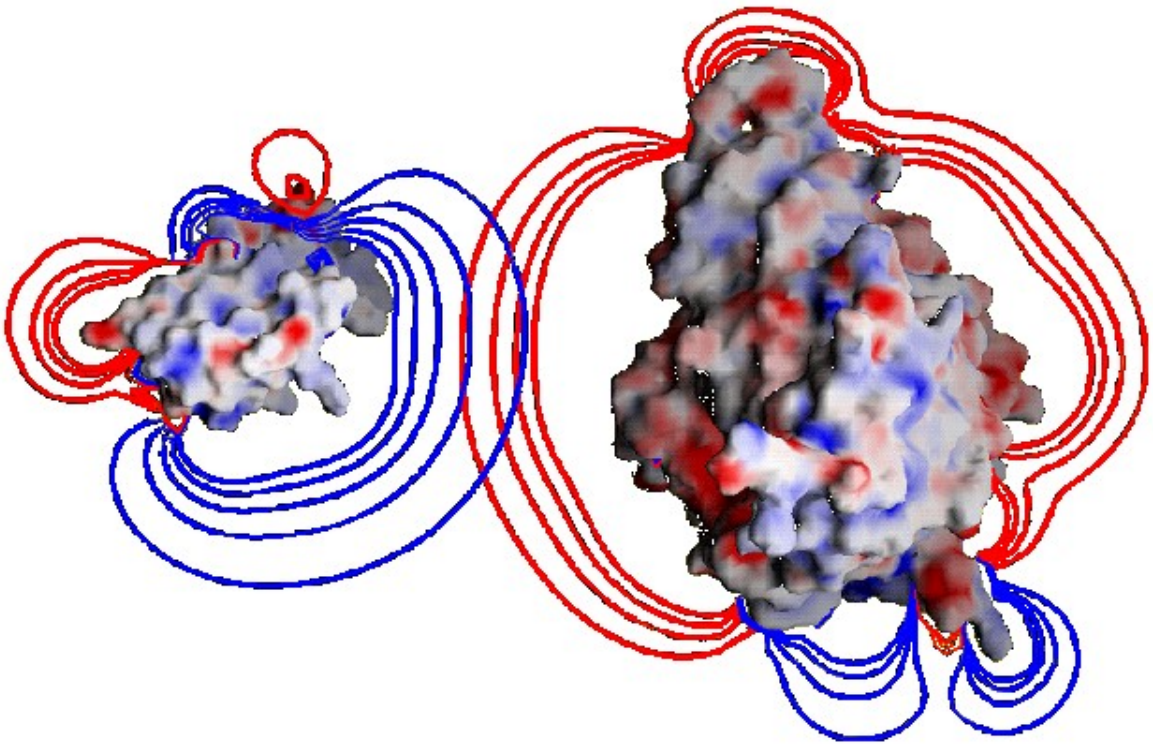


Figures: M. K. Gilson

- Electrostatic potentials
  - Binding complementarity
  - Molecular multipoles
  - Visualization
  - Protein Interaction Property Similarity Analysis (PIPSA)

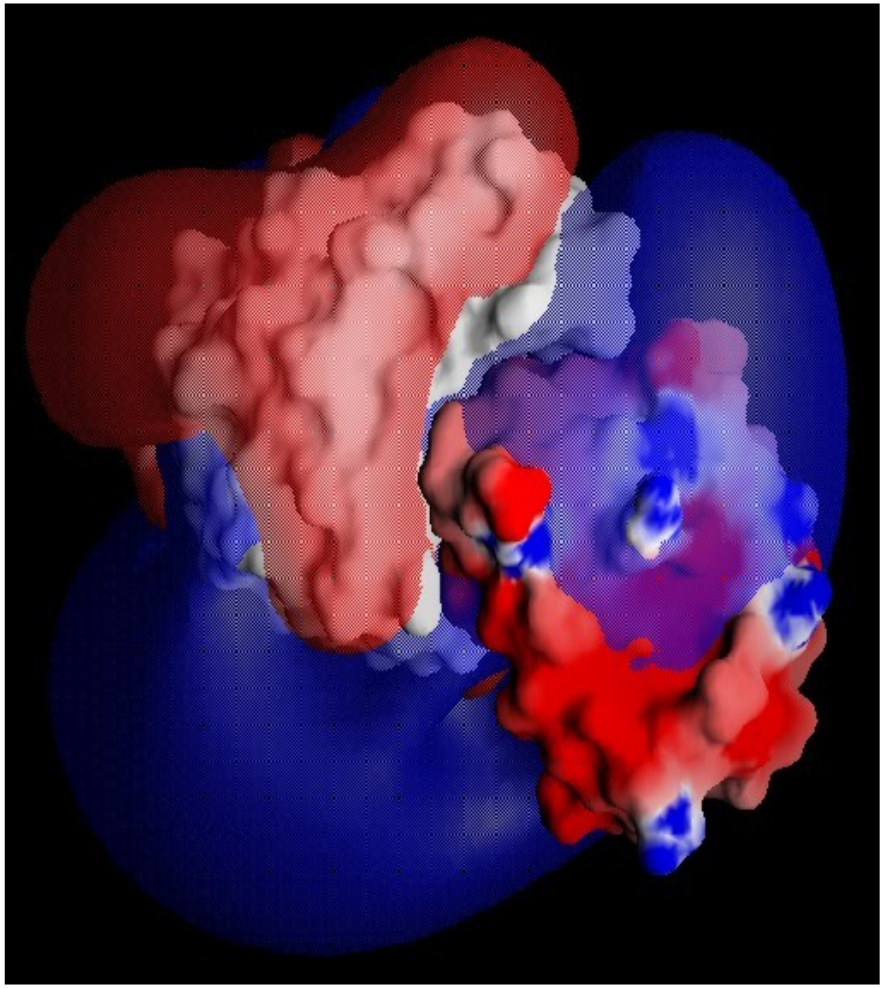


Fasciculin-Acetylcholinesterase: Electrostatic Complementarity



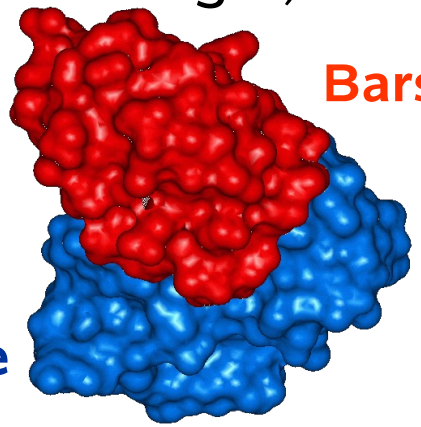
- AChE: -6e
  - neurotransmitter hydrolysis
  - cholinergic synapses
- Fasciculin: +4e
  - neurotoxin inhibitor
- K_d ~ 10^{-13} M
- K_{on} ~ 10^{9} $M^{-1}s^{-1}$ (zero ionic strength)

Barnase-Barstar: Electrostatic Complementarity

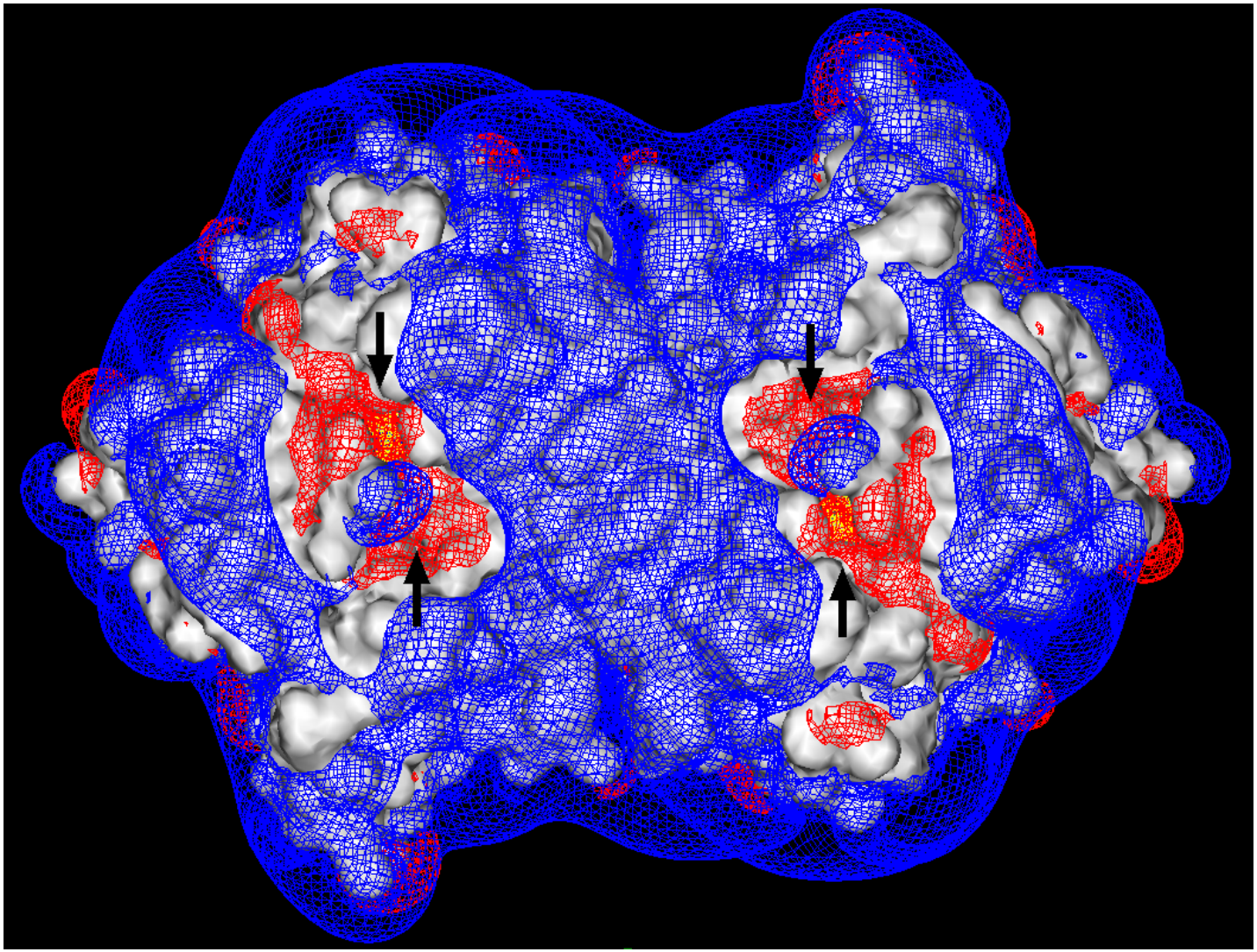


- Barnase : +2e
  - extracellular ribonuclease
  - Bacillus amyloliquefaciens
- Barstar : – 6e
  - intracellular inhibitor

- K_d ~ 10^{-14} M
- K_{on} ~ 10^{10} $M^{-1}s^{-1}$ (zero ionic strength)



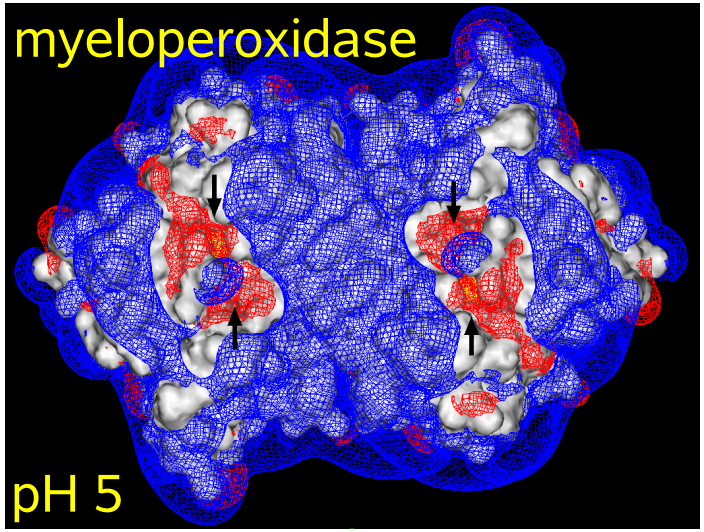
Myeloperoxidase: Active site access

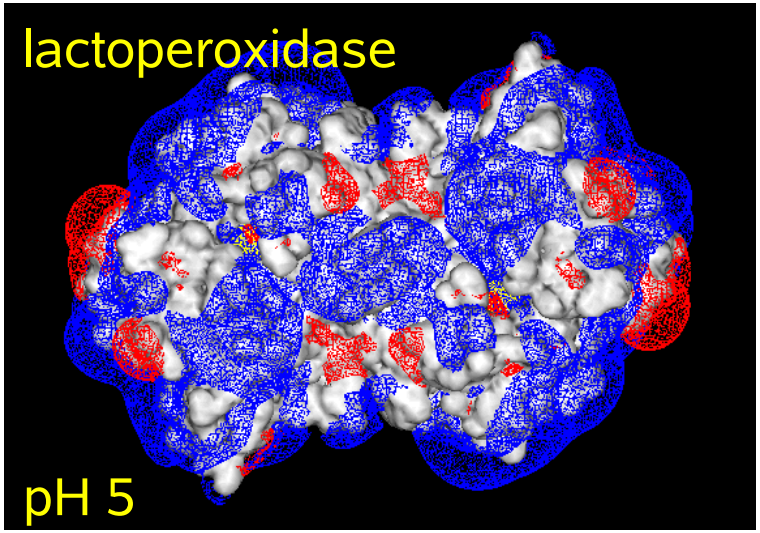


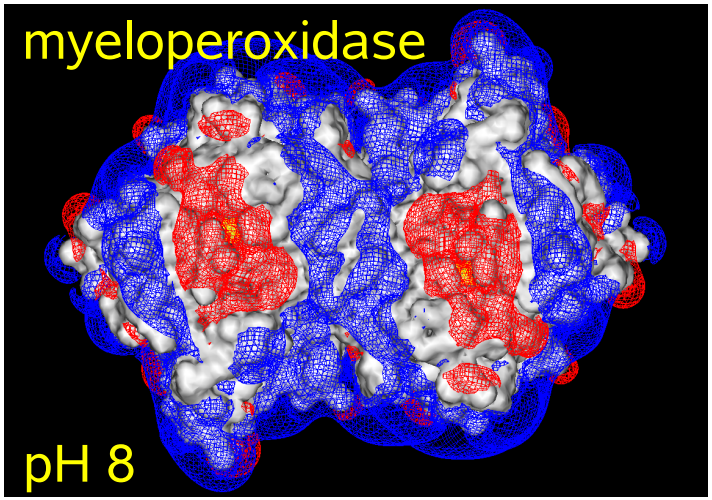
Gabdoulline, Kummer, Olsen, Wade, Biophys. J (2003) 85, 1421-1428

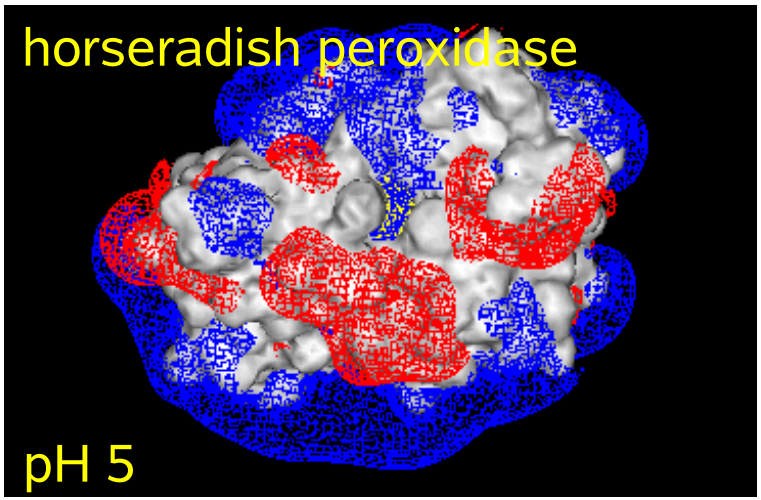
EML Research

Electrostatics of 3 Peroxidases









University of **H**ouston **B**rownian **D**ynamics (UHBD)

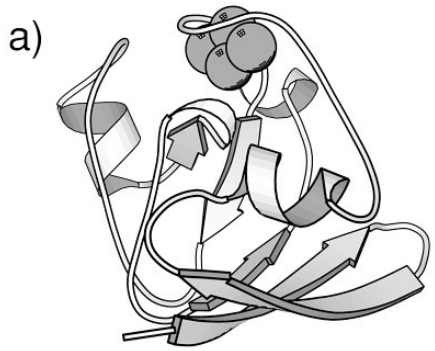
- solves the linearized and non-linearized Poisson-Boltzmann equation

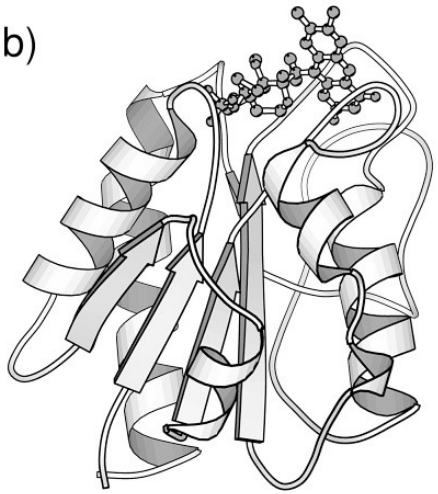
http://adrik.bchs.uh.edu/uhbd.html

Adaptive **P**oisson-**B**oltzmann **S**olver (APBS)

http://apbs.sourceforge.net

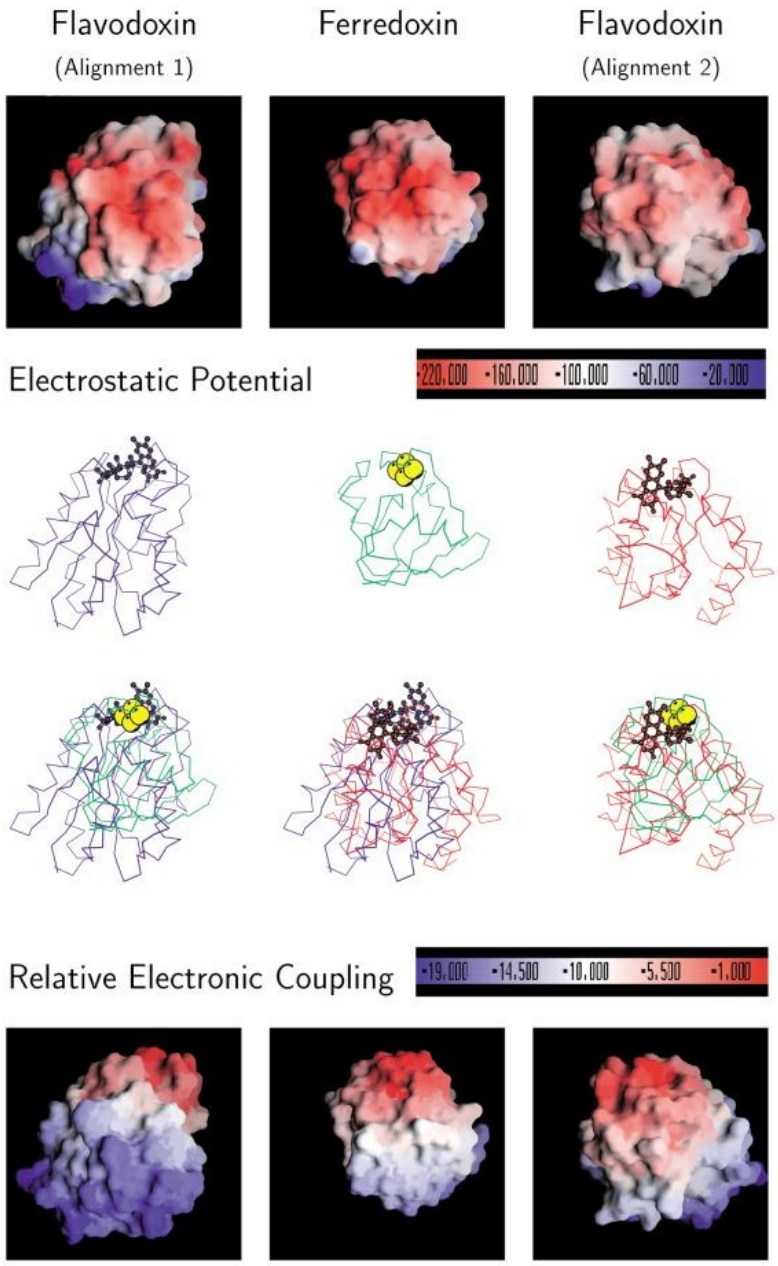
Structural Alignment based on Electrostatic Potential





G. M. Ullmann,
M. Hauswald,
A. Jensen,
E. W. Knapp

Proteins 2000

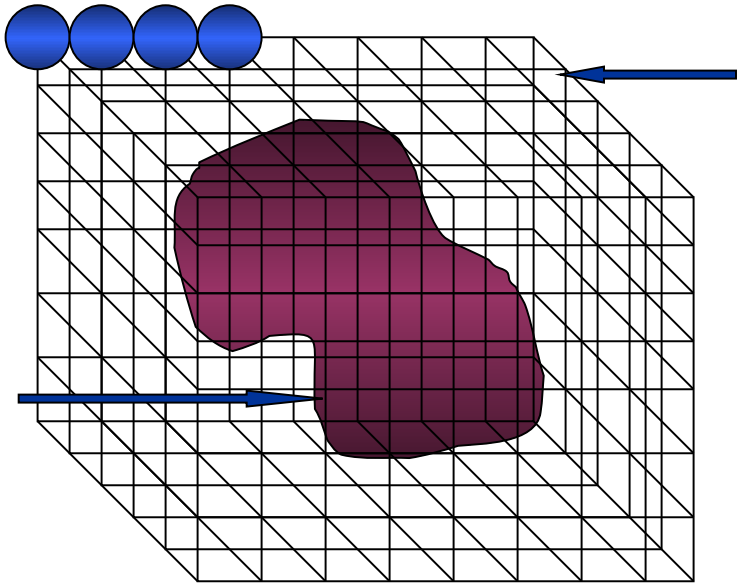


Structures of ferredoxin and flavodoxin. Although both proteins differ in structure and size, they perform the same physiological function.

GRID

other examples for
molecular interaction fields (MIFs)

– Detect energetically favorable binding sites on molecules of known structure

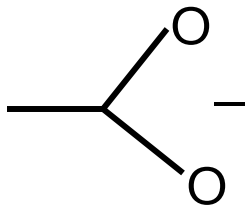


$$\Delta E = \sum_i E_{LJ} + \sum_i E_{EL} + \sum_i E_{HB} + S$$

Goodford, PJ *J. Med. Chem.* (1985) 28, 849-857.

GRID probes

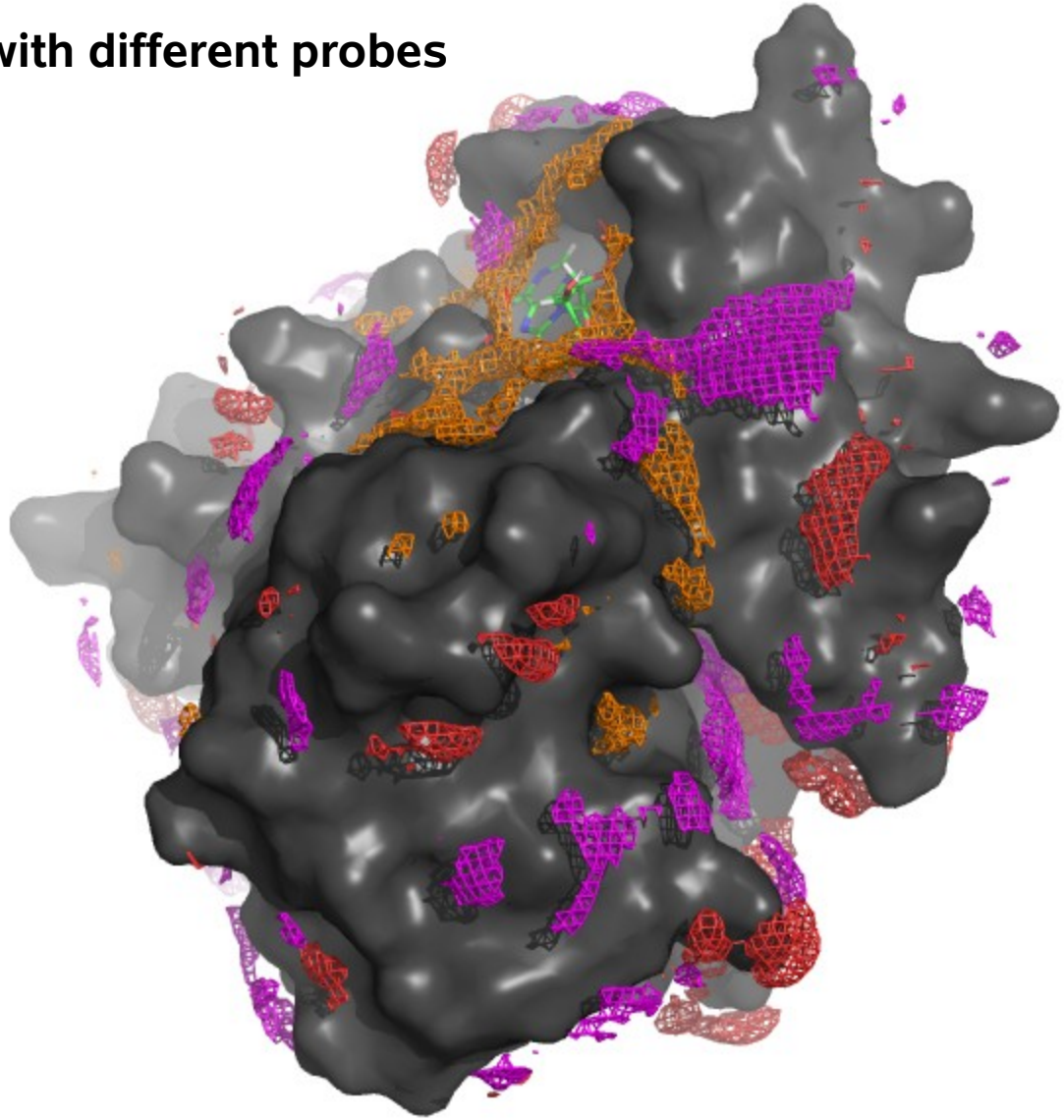
- Most are appropriately parameterized single-point spheres
  - E.g. carbonyl oxygen, hydroxyl, amino, water, dry, sulfate
- Some are multiatom
  - E.g. carboxylate, amide



- Water Probe
  - Hydrophilic protein surface
- Dry Probe
  - Hydrophobic protein surface

GRID: dry / ar-amidine / COO

using Grid with different probes

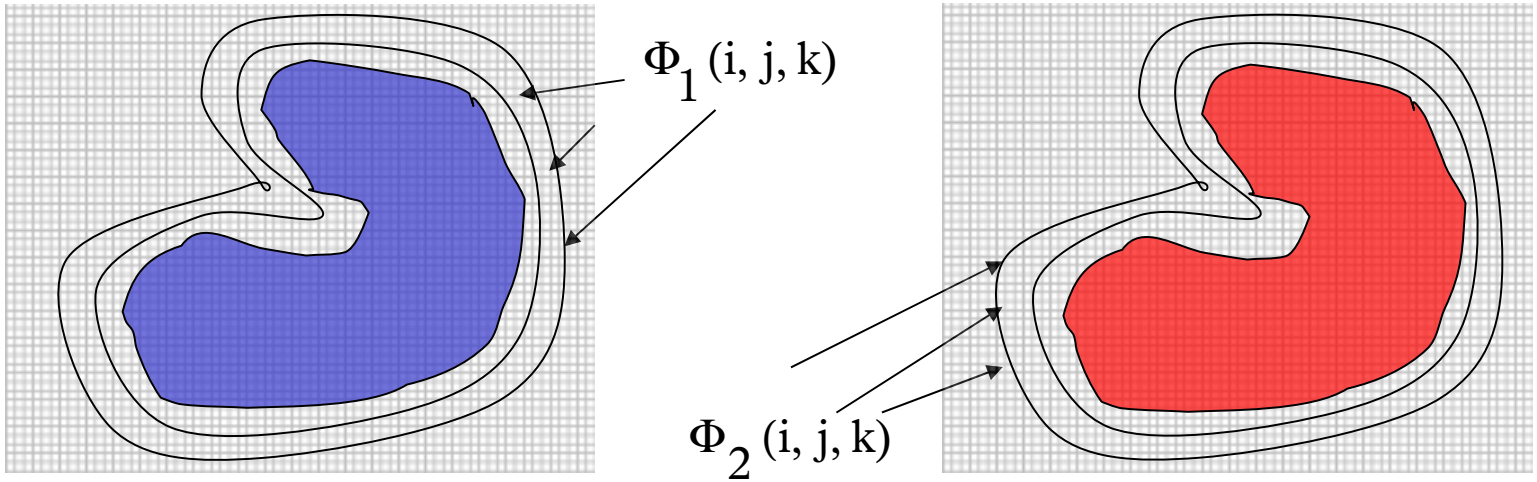


EML Research

PIPSA

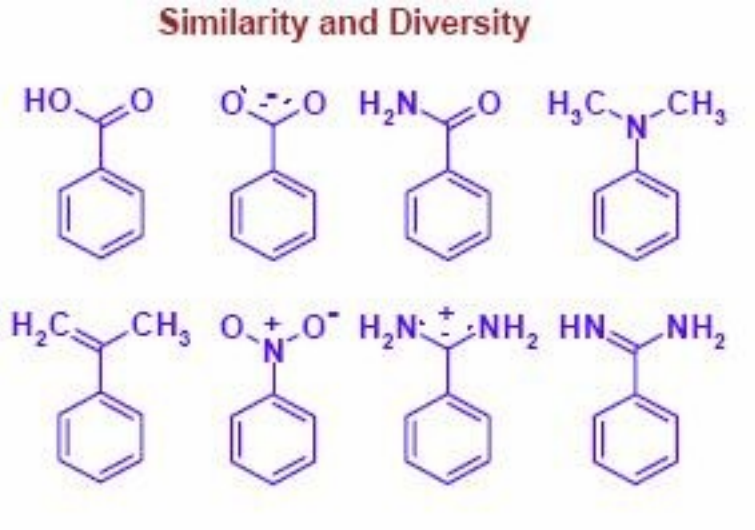
PIPSA: Protein Interaction Property Similarity Analysis

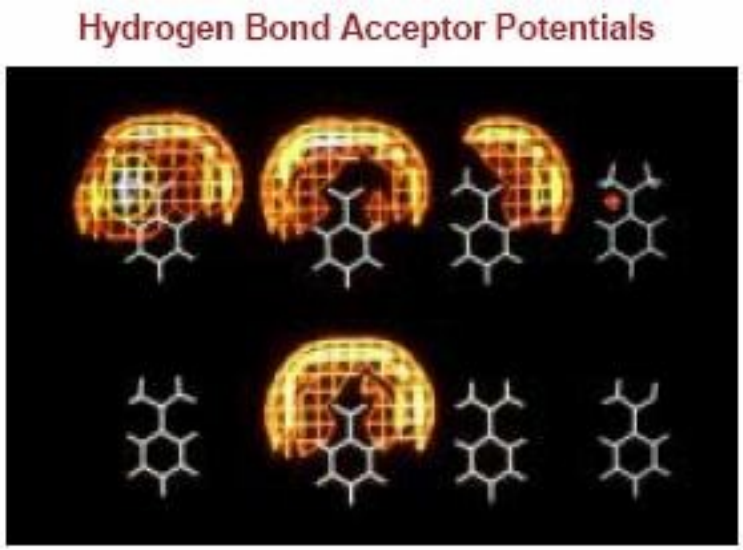
- Classification of the proteins in a protein family by interaction properties
- Detection of regions of similarities and differences in molecular interaction fields (MIFs)
- Correlation of MIFs with functional parameters
- Interaction properties: electrostatic, hydrophobic etc. potentials
- Superimposed protein structures needed

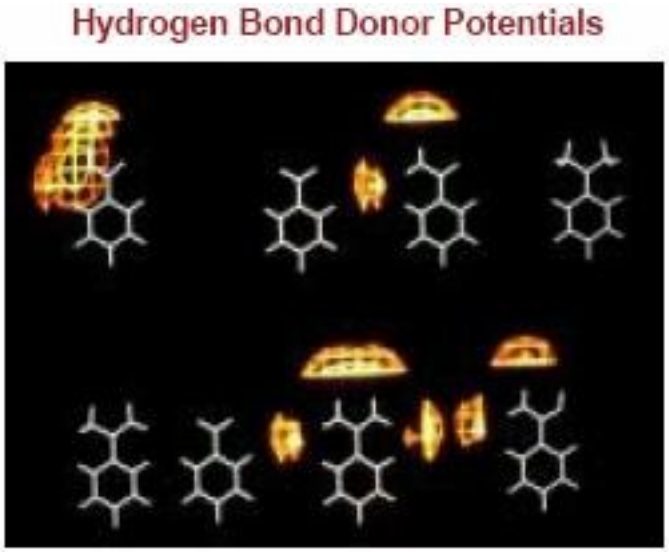


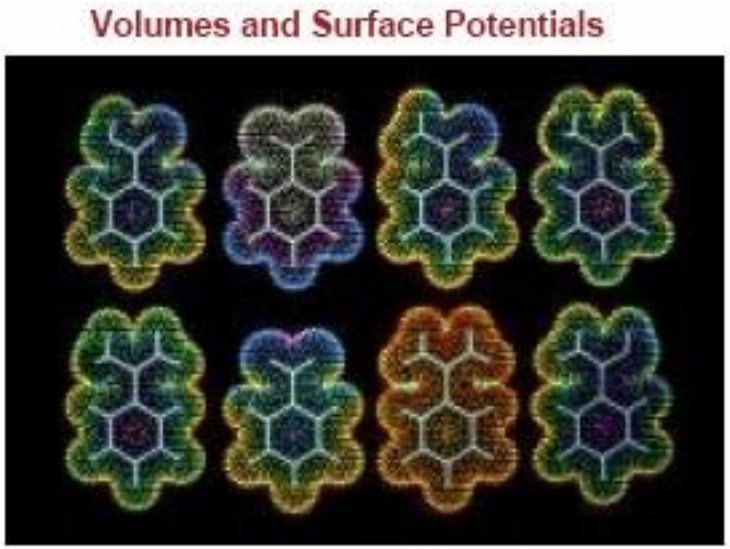
- Interaction fields are calculated on a set of points
- Field values on corresponding points are compared
- Φ = electrostatic potential, shape, probe interaction field, ...

similarity of small molecules









Comparitive Molecular Field Analysis (CoMFA)
Comparitive Molecular Similarity Indices Analysis (ComSIA)

Similarity index (SI)

property X of A and B with N bits

- Euclidiean

$$Euclidean_{AB} = \sqrt{\sum_{i=1}^{N} (X_{iA} - X_{iB})^2}$$

- Tanimoto

$$Tanimoto_{AB} = \frac{\sum_{i=1}^{N} X_{iA}\, X_{iB}}{\sum_{i=1}^{N} (X_{iA})^2 + \sum_{i=1}^{N} (X_{iB})^2 - \sum_{i=1}^{N} X_{iA}\, X_{iB}}$$

- Hodgkin

$$Hodgkin_{AB} = \frac{2\sum_{i=1}^{N} X_{iA}\, X_{iB}}{\sum_{i=1}^{N} (X_{iA})^2 \sum_{i=1}^{N} (X_{iB})^2}$$

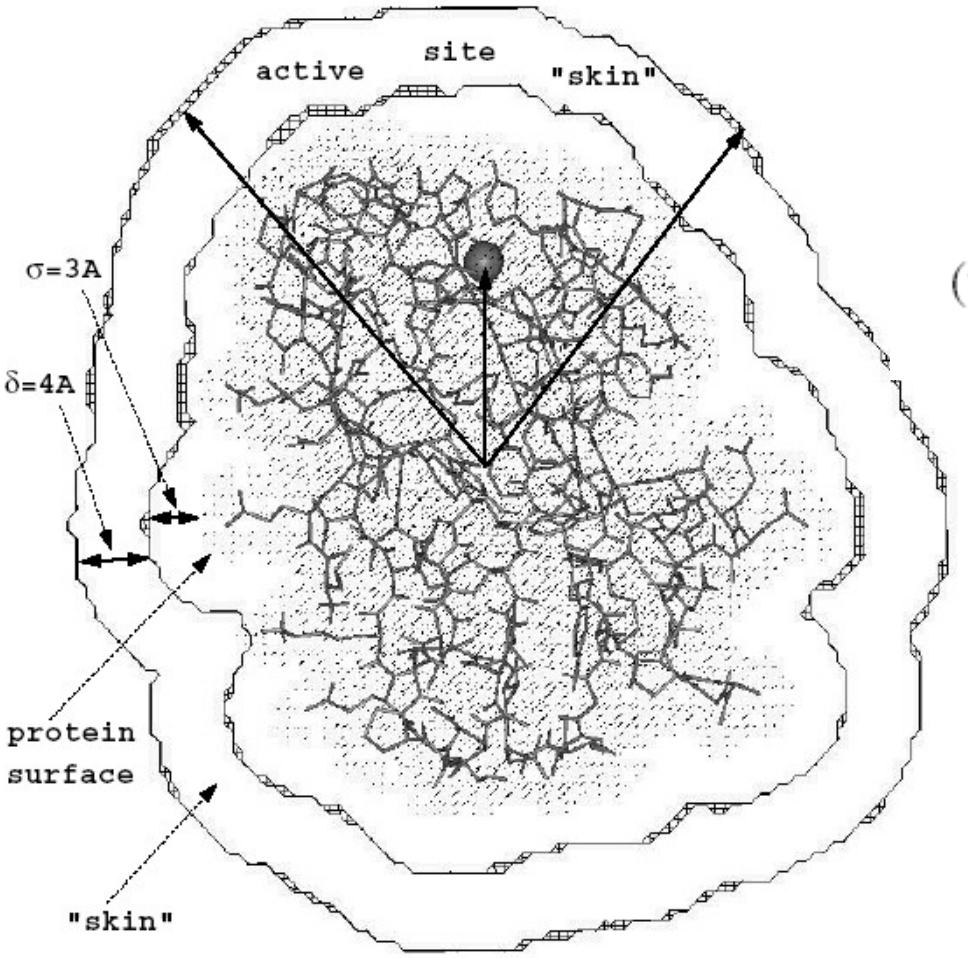
- Carbo

$$Carbo_{AB} = \frac{\sum_{i=1}^{N} X_{iA}\, X_{iB}}{\sqrt{\sum_{i=1}^{N} (X_{iA})^2 + \sum_{i=1}^{N} (X_{iB})^2}}$$

- electrostatic distance

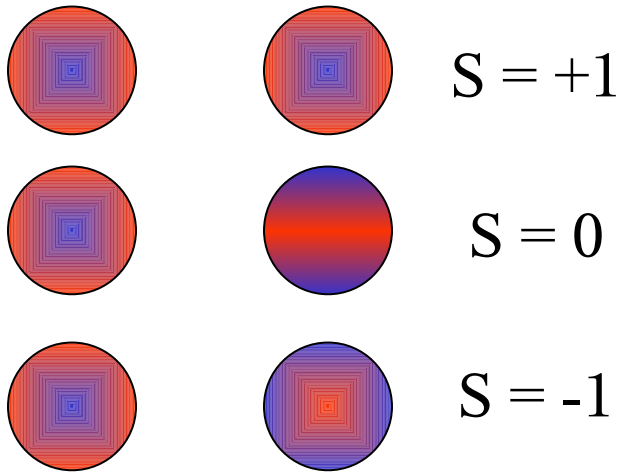
$$D_{ab} = \sqrt{1 - SI_{ab}}$$

PIPSA: Protein Interaction Property Similarity Analysis



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

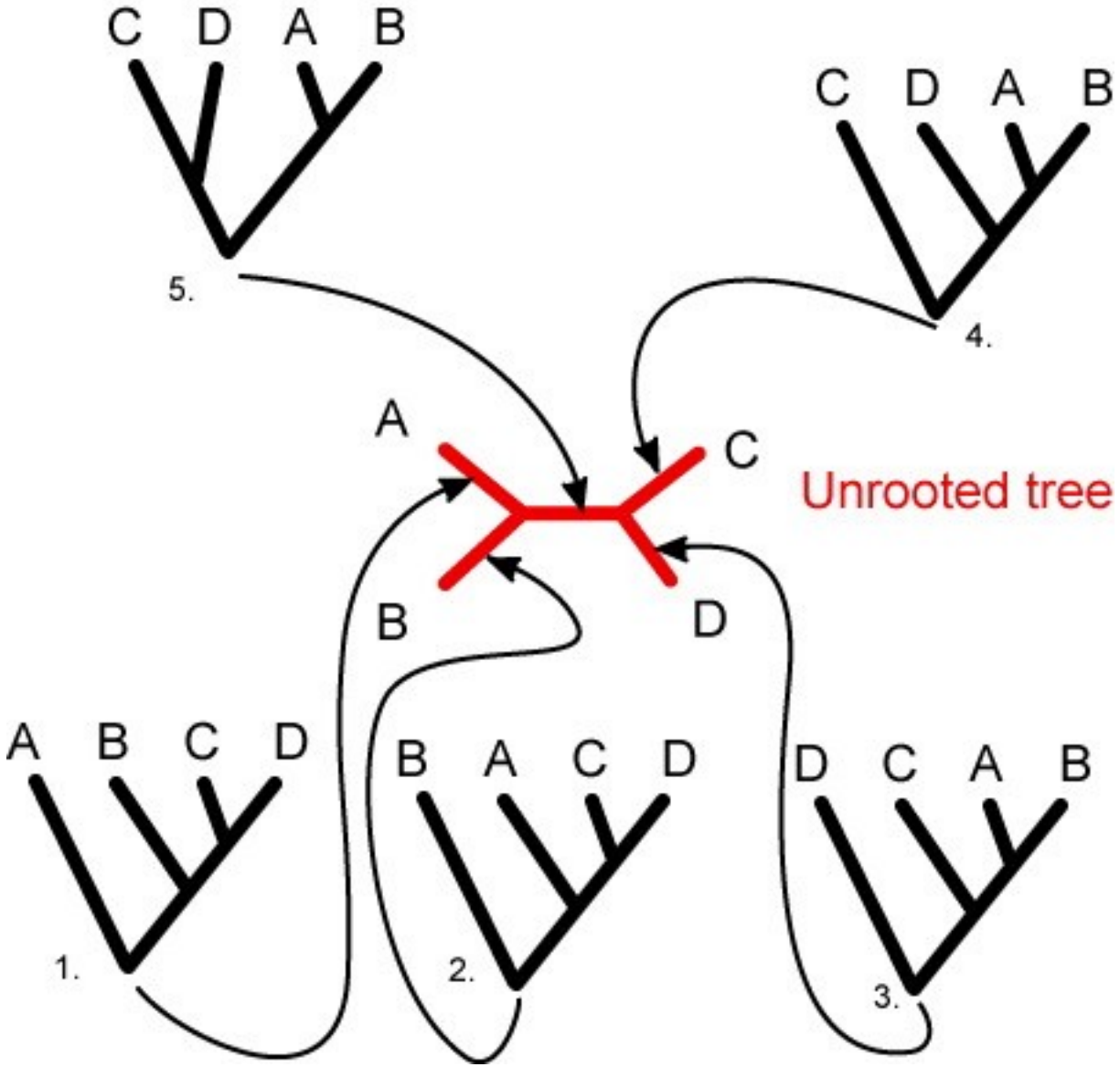
$$(\mathbf{p}_1, \mathbf{p}_2) = \sum_{i,j,k} \phi_1(i, j, k)\phi_2(i, j, k)$$



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

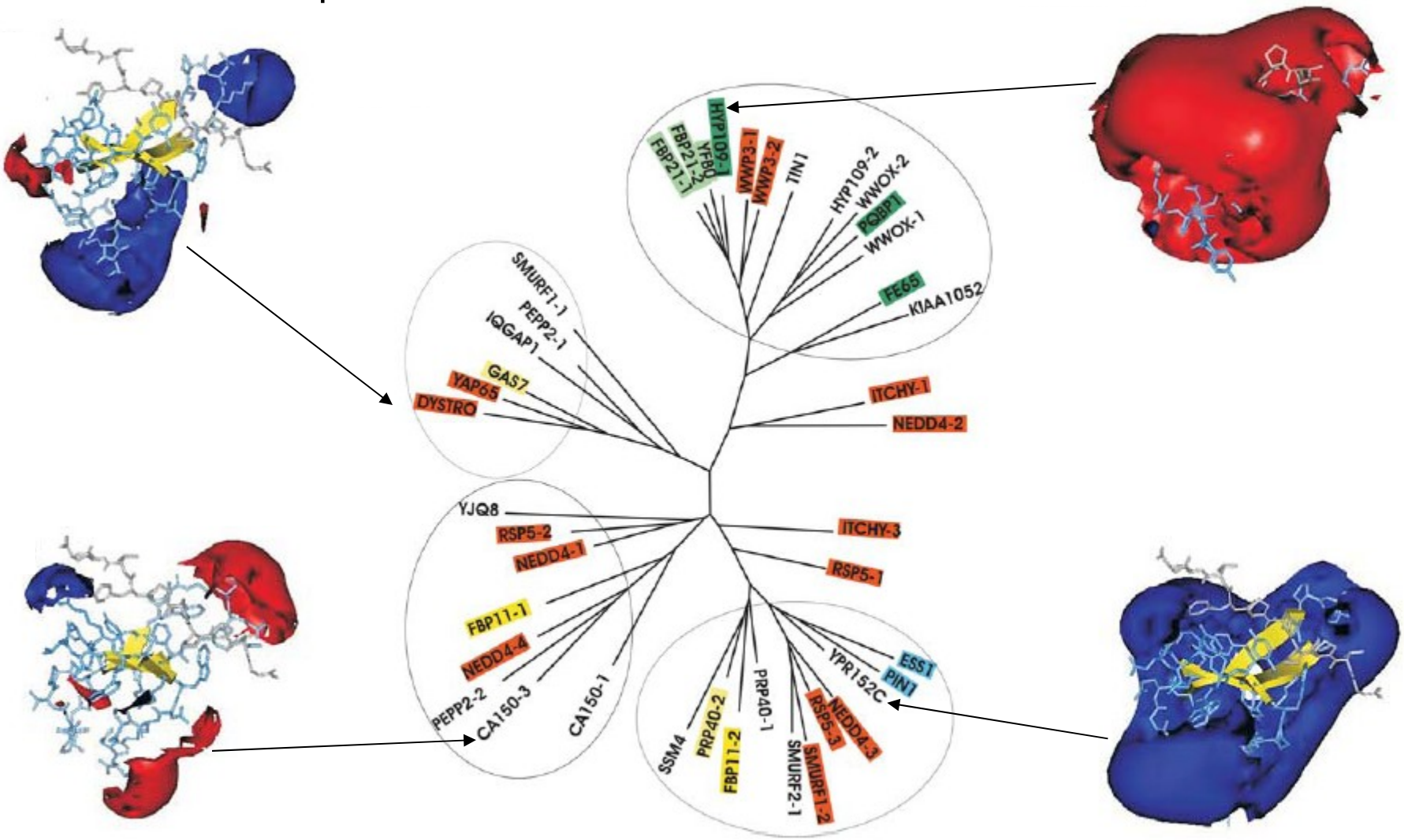
EML Research

rooted and unrooted trees



42 WW Domains:
PIPSA epogram for Molecular electrostatic potential

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

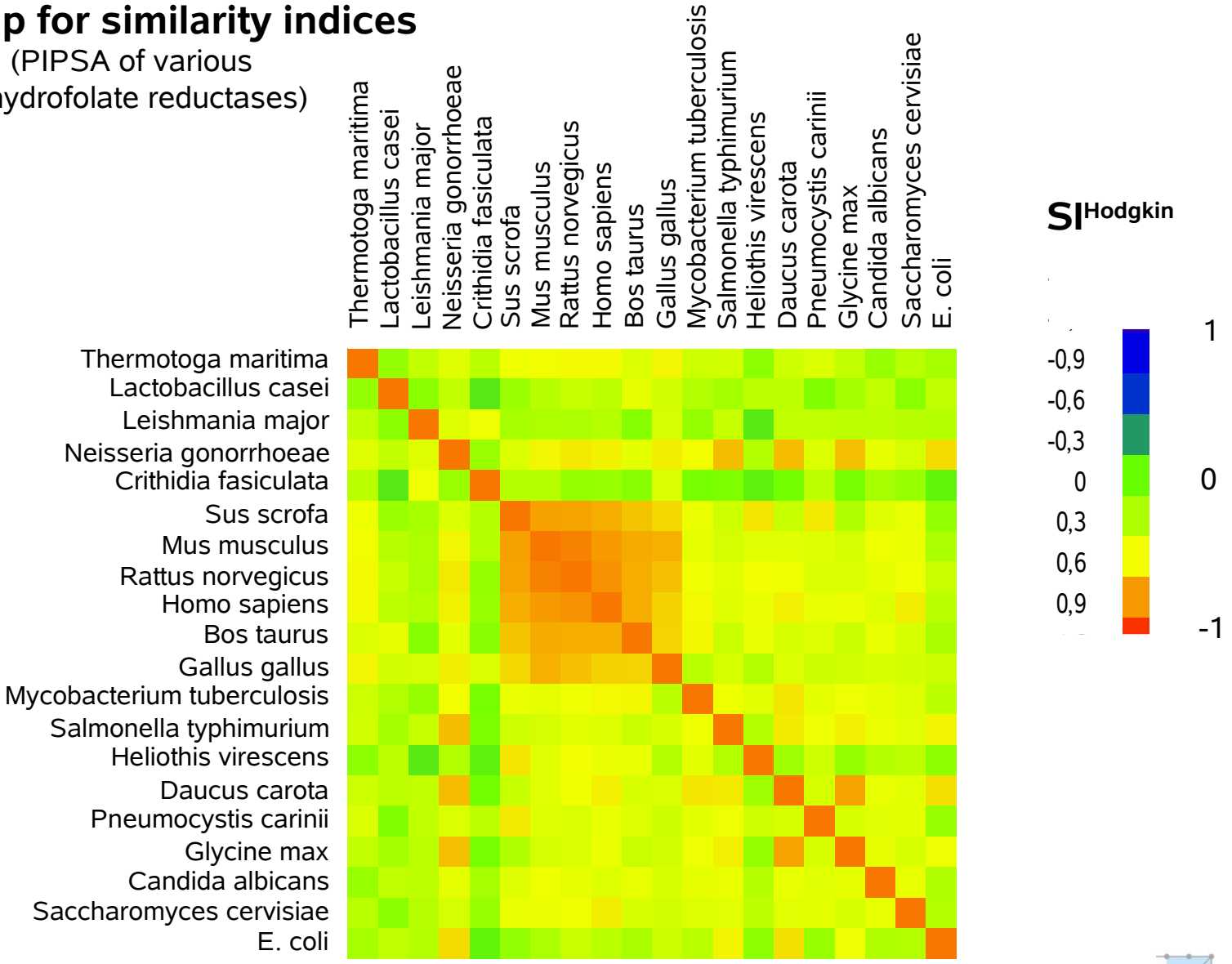


Schleinkofer, Wiedemann et al, JMB (2004) 344, 865-881

EML Research

heat map for similarity indices

(PIPSA of various dihydrofolate reductases)

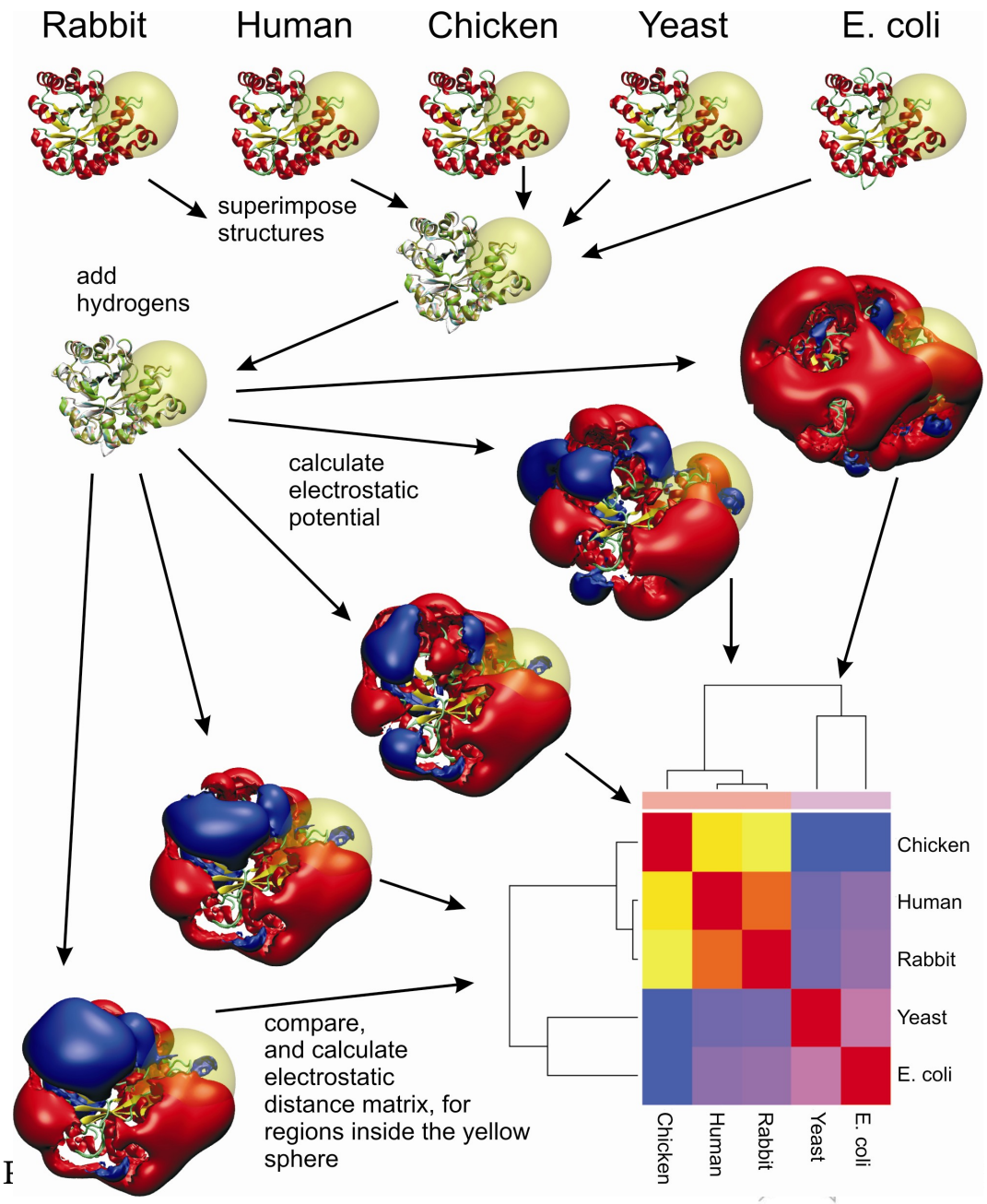


webPIPSA

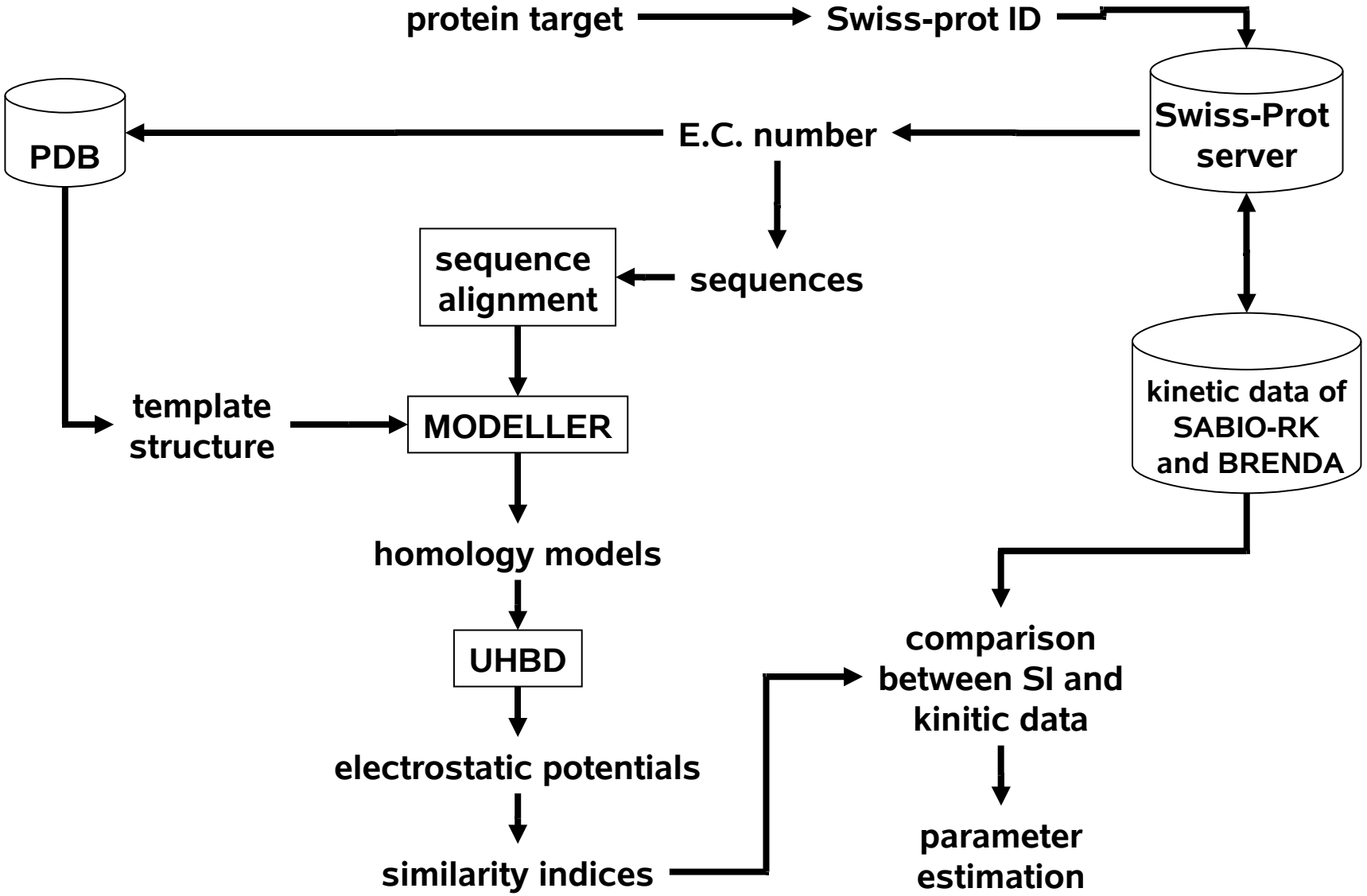
webPIPSA: pipsa.eml.org

triosephosphate isomerase

Richter et al.,
Nucleic Acids Research, 2008



webPIPSA: workflow with homology modeling

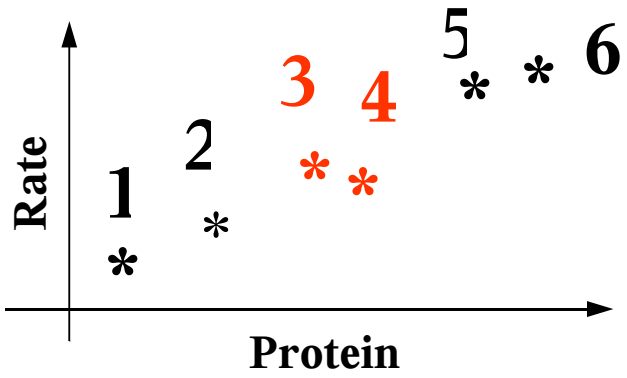


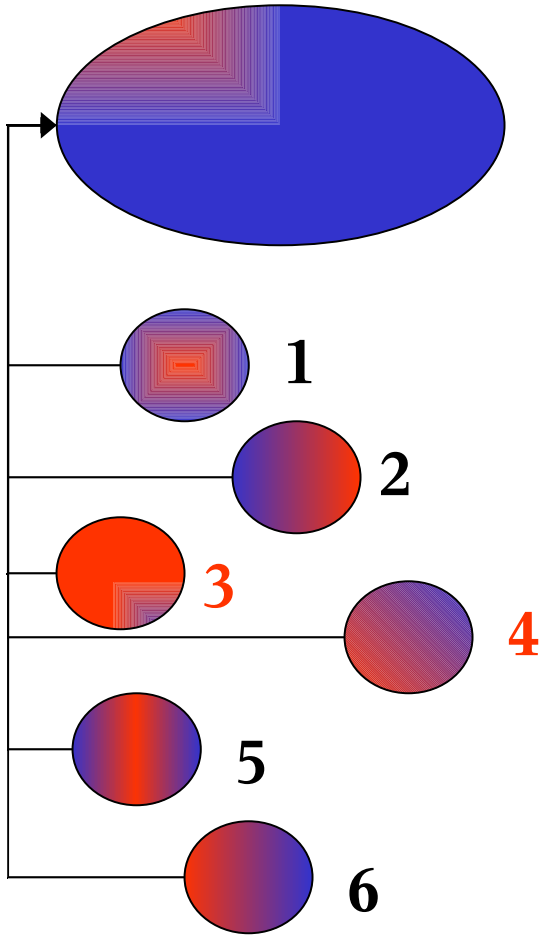
EML Research

qPIPSA

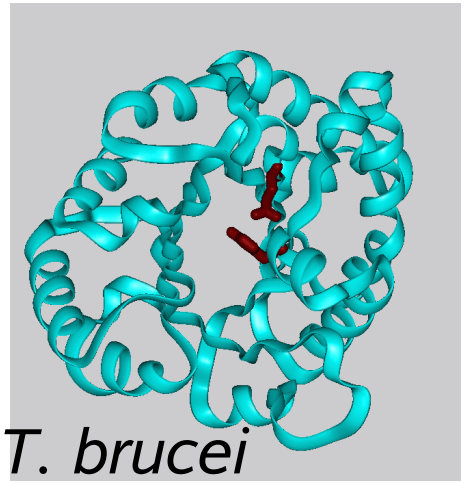
Quantitative PIPSA (qPIPSA)

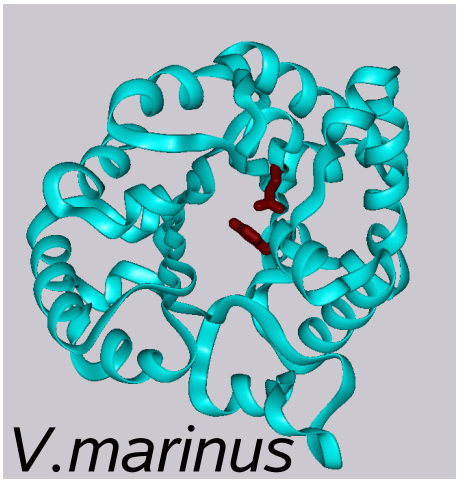
- Rates relative to one Partner
- Training Set Required with Experimental Information
- Predict Relative Ordering and Trends



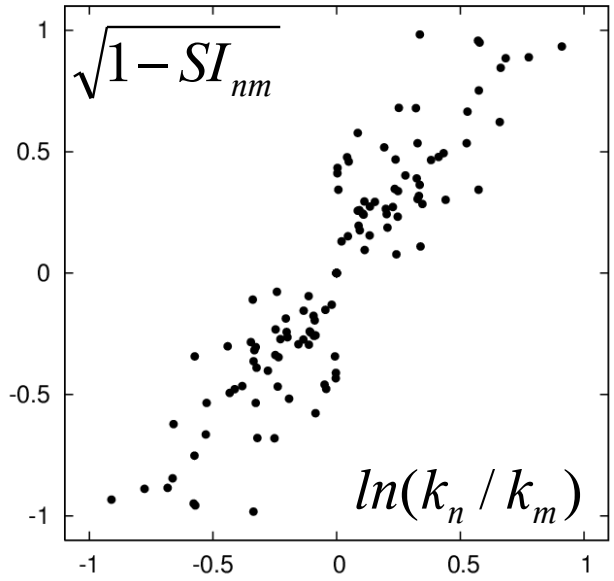


triose phosphate isomerases from different species





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

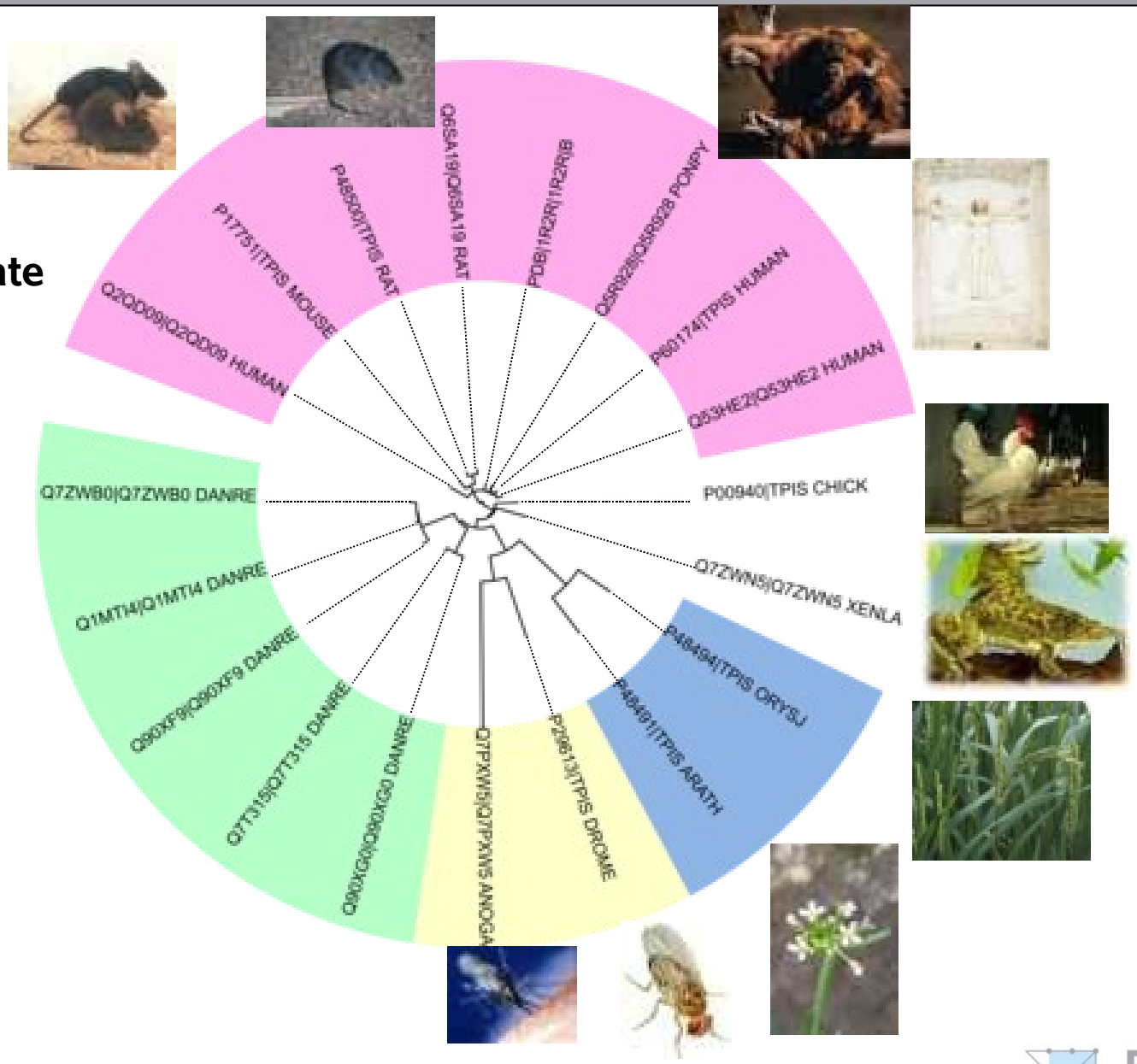


12 species:
Giardia lambli
Spinach
Chicken
E. coli
Human
L. mexicana
P. falciparum
Rabbit
T. brucei
T. cruzi
V. marinus
Yeast

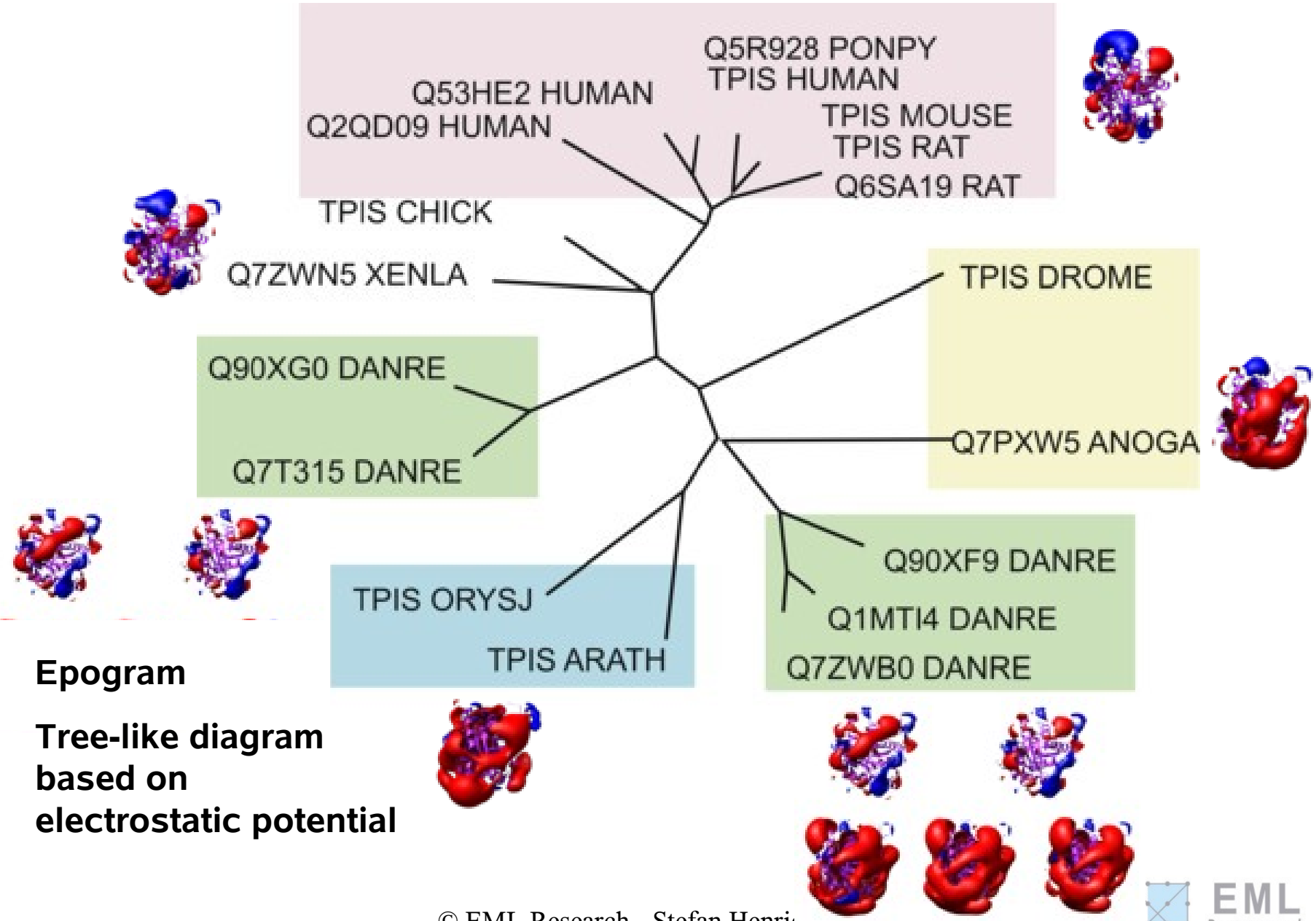
$$ln(k_1/k_2) \approx \Delta G_2^{\ddagger} - \Delta G_1^{\ddagger} \approx \sqrt{\left\langle (\mathbf{p}_1 - \mathbf{p}_2)^2 \right\rangle} \approx \alpha\sqrt{(1-\mathrm{SI}_{12})}$$

Functional Annotation of Enzymes by PIPSA

Triosephosphate isomerases



Functional Annotation of Enzymes by PIPSA



Epogram

Tree-like diagram based on electrostatic potential

EML Research

acknowledgement

Molecular and Cellular Modeling Group

- **Rebecca Wade**
- Divita Garg
- Outi Salo-Ahen
- **Stefan Richter**
- Vlad Cojocaru
- Georgi Pachov
- **Matthias Stein**
- Michael Martinez
- Anna Feldman-Salit
- Bingding Huang
- Daria Kokh
- Matthias Janke

former group members

- **Razif Gabdoulline**
- Ting Wang
- Domantas Motiejunas
- Sulaiman Faisal
- Anne Wenzel
- Erika Nerini

Scientific Databases and Visualization Group

http://www.eml-research.de/english/research/mcm
http://projects.villa-bosch.de/mcm
http://sycamore.eml.org
http://pipsa.eml.org

€€€: *Klaus Tschira Foundation, BMBF, DAAD, NIH, EU, DFG*

EML Research

MCM

http://www.eml-research.de/english/research/mcm http://projects.villa-bosch.de/mcm/software

PIPSA

webPIPSA:

http://pipsa.eml.org

http://sycamore.eml.org

download software:

http://projects.villa-bosch.de/mcm/software/pipsa

PIPSA

Blomberg, N., Gabdoulline, R.R., Nilges, M. and Wade, R.C. (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins*, **37**, 379-387.

Wade, R.C., Gabdoulline, R.R. and De Rienzo, F. (2001) Protein interaction property similarity analysis. *International Journal of Quantum Chemistry*, **83**, 122-127.

Gabdoulline, R.R., Stein, M. and Wade, R.C. (2007) qPIPSA: relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics*, **8**, 373.

Henrich, S., Richter, S. and Wade, R.C. (2008) On the use of PIPSA to Guide Target-Selective Drug Design. *ChemMedChem*, **3**, 413-417.

Richter, S., Wenzel, A., Stein, M., Gabdoulline, R.R. and Wade, R.C. (2008) webPIPSA: a web server for the comparison of protein interaction properties. *Nucl. Acids Res.*