Wintersemester 2016/2017
Biomolecular Engineering/Nanobiophysics Module

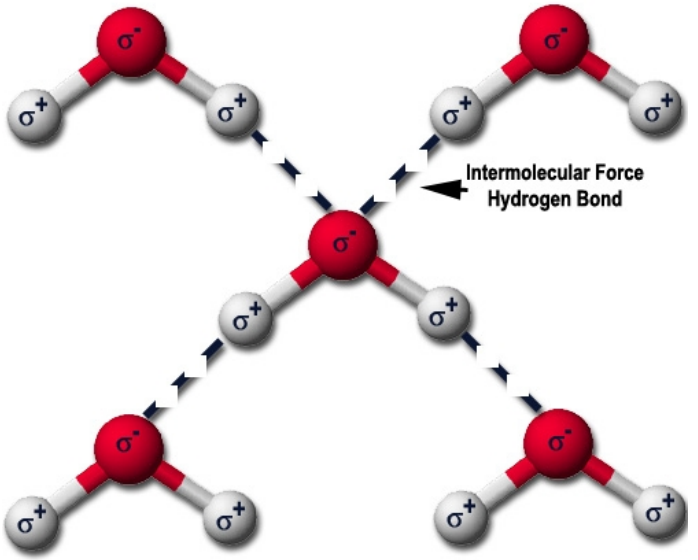
LECTURE 3: MODELLING SOLVENT

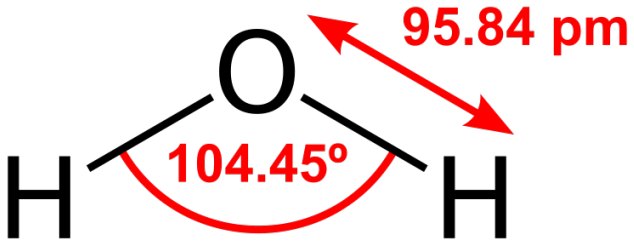
LECTURE 3: MODELLING SOLVENT

- Water unique properties
- Water and biomolecular systems
- Implicit solvent and Poisson-Boltzmann methodology
- Explicit solvent models
- Grid Inhomogeneous Solvation Theory (GIST)
- Solvent challenge in docking
- Water in protein-protein interfaces
- Case studies:
  - MD study of the role of water in protein-protein interfaces
  - Introduction of solvent information for protein contacts prediction
  - Inclusion of water in GAGs docking to proteins

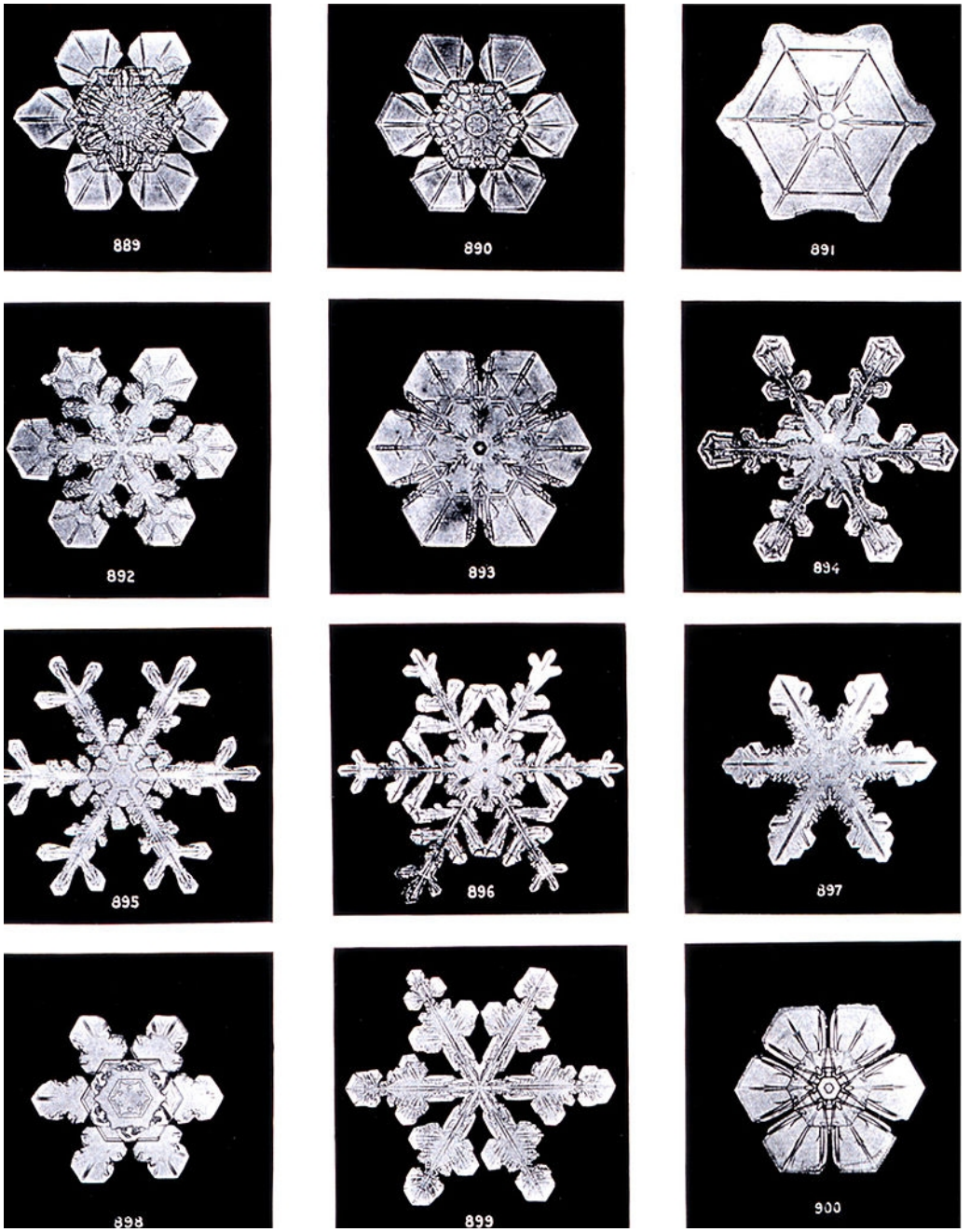
WATER UNIQUE PROPERTIES

- Three-dimensional tetrahedral H-bonding networks
- High boiling and freezing temperatures, vaporization enthalpy, surface tension
- Fluidity increases with increased pressure
- High dielectric constant (~80)
- Diverse crystal forms
- Volumetric anomalities (ice density < liquid water density)
- $2H_2O \leftrightarrow H_3O^+ + OH^-$; $K_w = 10^{-14}$ at 25C
- 1.52% of Earth mass, 90% of human body mass





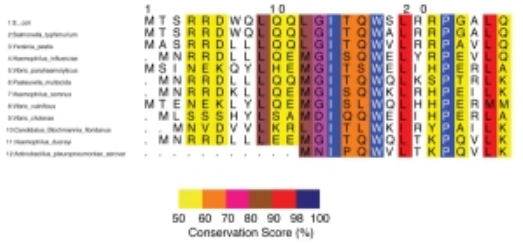
WATER CRYSTALS

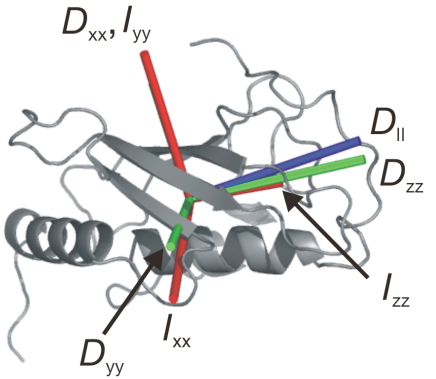


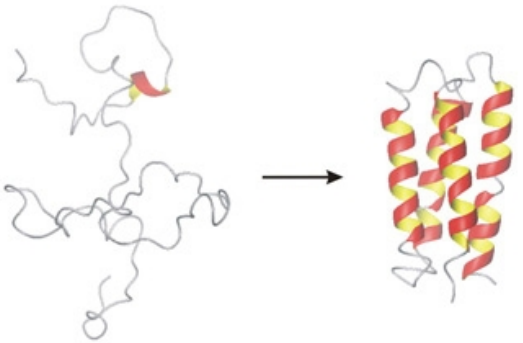
Wilson Bentley, 1902

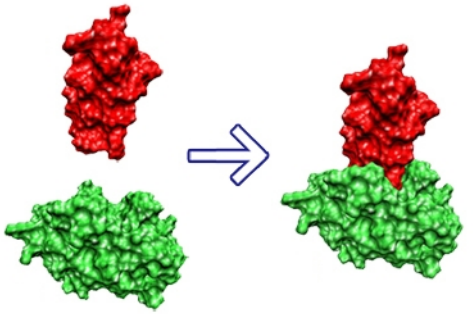
WATER AND BIOMOLECULAR SYSTEMS

- Structural conservation
- Dynamics
- Folding
- Molecular recognition
- Catalytic activity

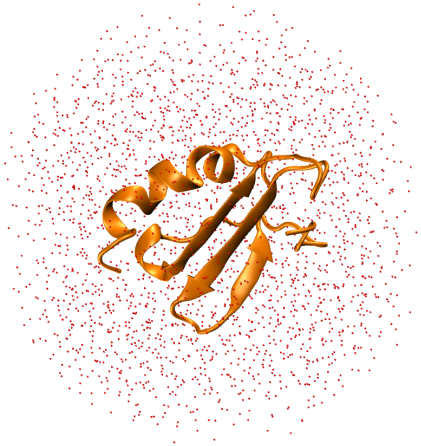


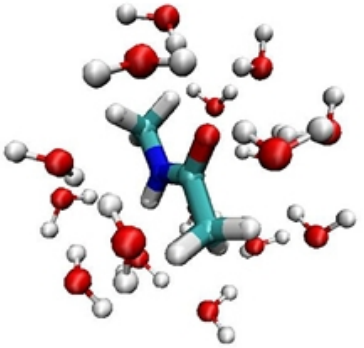






COMPUTATIONAL TREATMENT OF SOLVENT





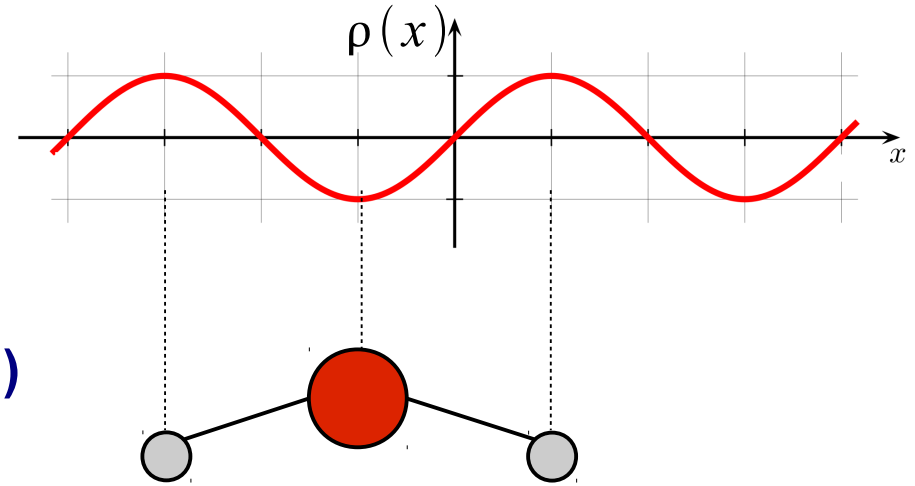
- *In vacuo* – no solvent
- Implicit solvent – continuous solvent with averaged macroscopic properties
- Explicit solvent – each solvent molecule is given explicitly

IMPLICIT SOLVENT

- No individual water molecules, but the space has macroscopic properties, which, on average, reproduce effect of solvation

➢ Dielectric constant ε is the same for the whole space

- ε = const
- ε = A/r, r – distance from solute
- ε = A/r^2, r – distance from solute



➢ Speeding up calculations

➢ No PBC (periodic boundary conditions)

➢ Continuous model

➢ Does not well reproduce local properties

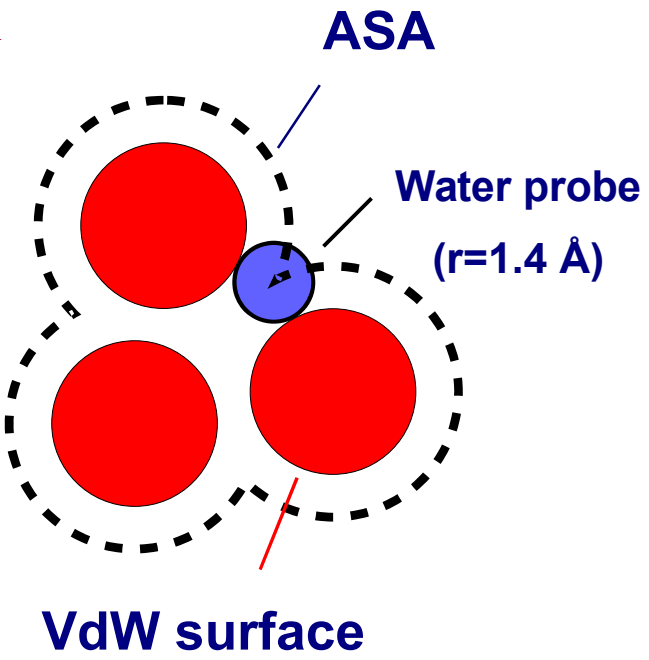
➢ No explicit hydrogen bonding network

MM-PBSA

- **Molecular Mechanics-Poisson-Bolzmann Surface Area**

$$\delta G = \delta G_{vac} + \delta G_{solv}\,;\ \delta G_{vac} = \delta G_{ele} + \delta G_{vdw}$$

Molecular Mechanics (force field)



$$\delta G_{solv} = \delta G_{el} + \delta G_{nonel}$$

$$\delta G_{nonel} \sim ASA$$

$$\delta G_{el} = \frac{1}{2}\int (\rho(r)\phi(r))$$

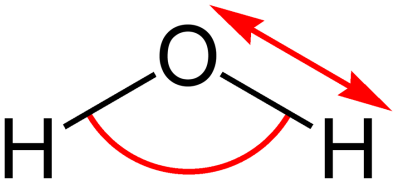
$$\nabla \epsilon(r) \nabla \phi(r) = -4\pi \rho(r) + \kappa^2 \epsilon(r)\, \phi(r) \quad - \quad Poisson - Boltzmann\ equation$$

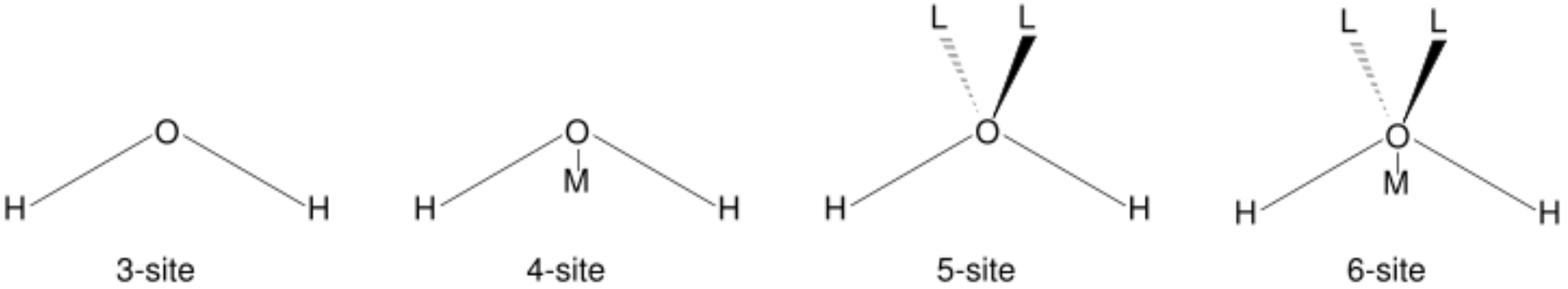
$$G_{el,GB} = \frac{1}{8\pi}\left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon}\right)\sum_{i<j}^{N} \frac{q_i q_j}{f_{GB}} \quad - \quad Generalized\ Born\ approximation$$

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \qquad D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2,\ a_{ij} = \sqrt{a_i a_j}$$

EXPLICIT SOLVENT MODELS

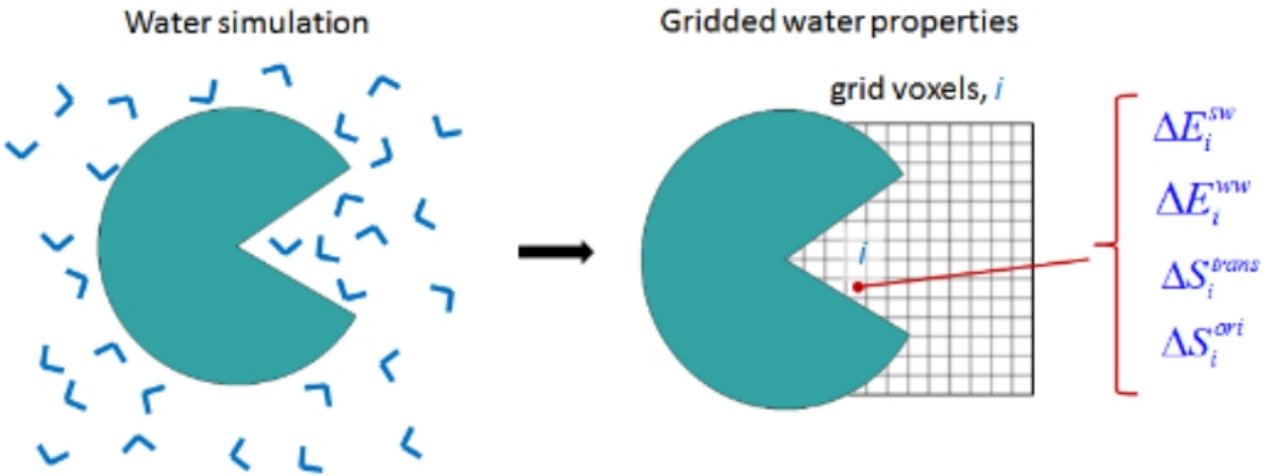
- Geometry: 2 internal parameters (O-H bond length and H-O-H angle)
- Dimensionality
- Number of points used (SPC, TIP3, TIP4, TIP5, TIP6)
- Flexibility
- Ability to reproduce H-bonding networks
- Ability to reproduce certain macroscopic properties
- Polarization





$$V_{ab} = \sum_{i}^{on\,a} \sum_{i}^{on\,b} \frac{k_c q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}}; \qquad V_{pol} = \frac{1}{2} \sum_{i} \frac{(\vec{d} - \vec{d}_0)^2}{\alpha_i} \quad - \quad Polarization\ \ term$$

GRID INHOMOGENEOUS SOLVATION THEORY

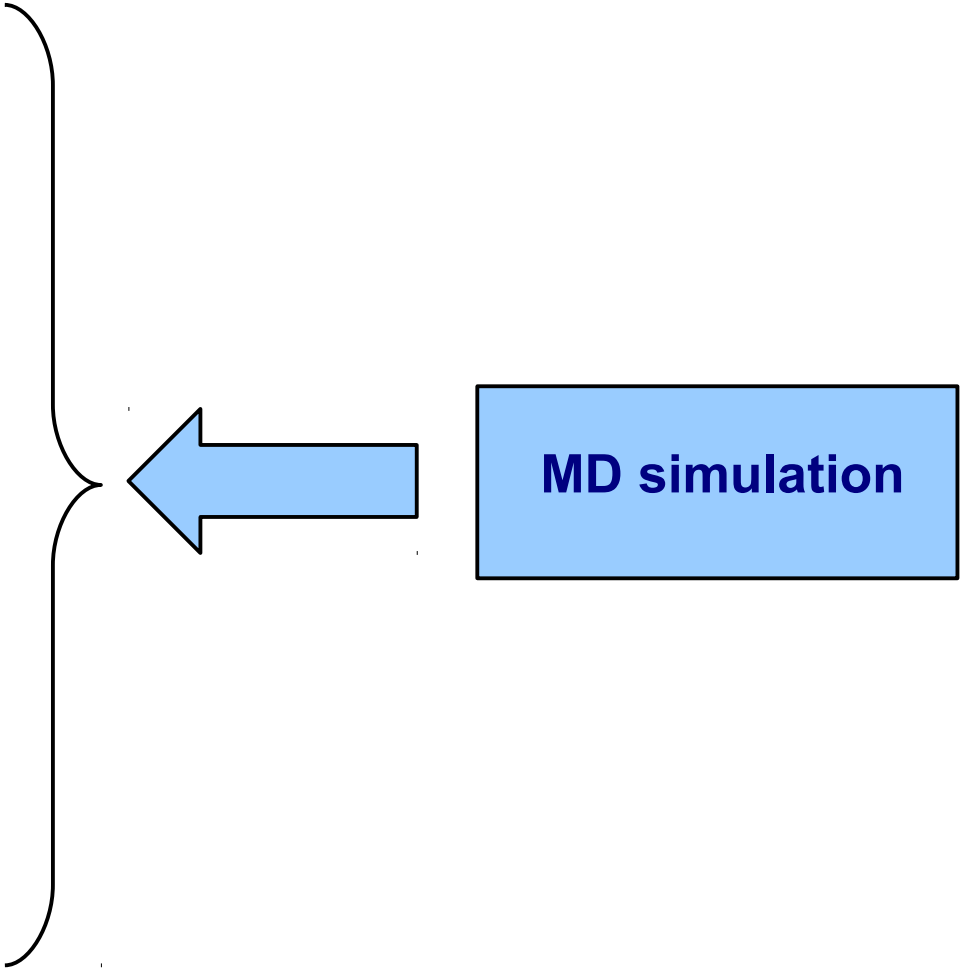


GIST's gridded water properties in a binding site.

- **Explicit solvent**
- **Enthalpy: potential at each analyzed frame**
- **Entropy: directly from the probabilities**
- **Reference: bulk solvent at normal conditions**
- **Challenge for convergence**

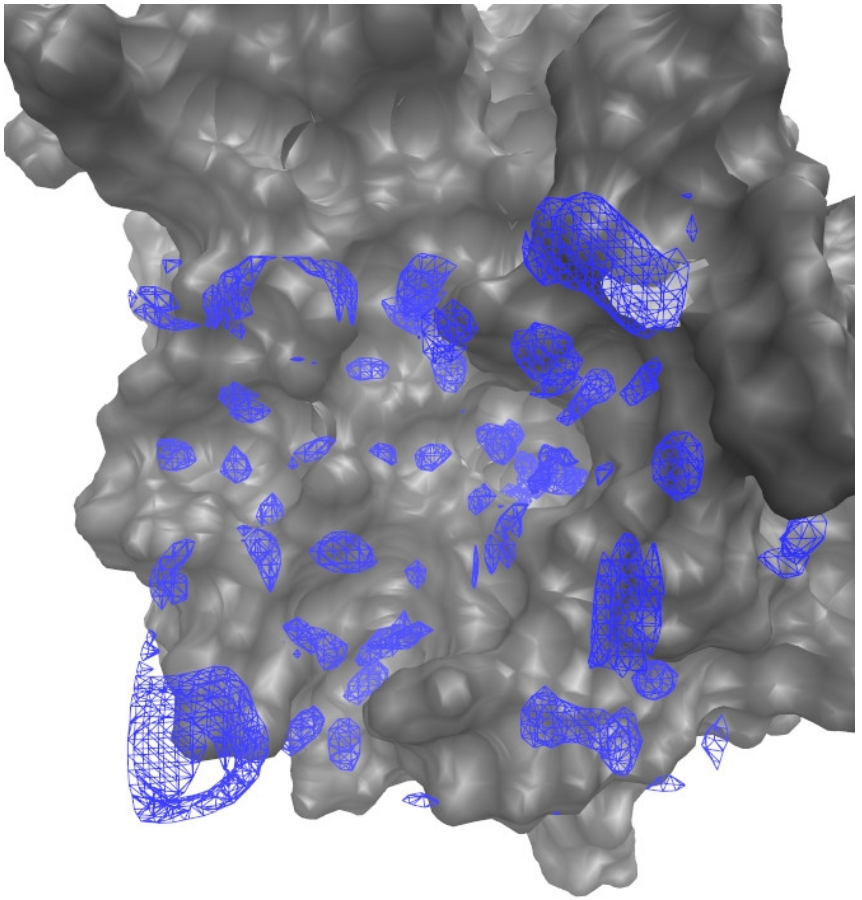
GRID INHOMOGENEOUS SOLVATION THEORY: OUTPUT PARAMETERS

- Water oxygen distribution g(O)
- Water hydrogen distribution g(H)
- E(solute-water)
- E(water-water)
- S(translational)
- S(orientational)
- Water induced dipoles
- Number of neighbouring waters
- Average tetrahedral order parameters

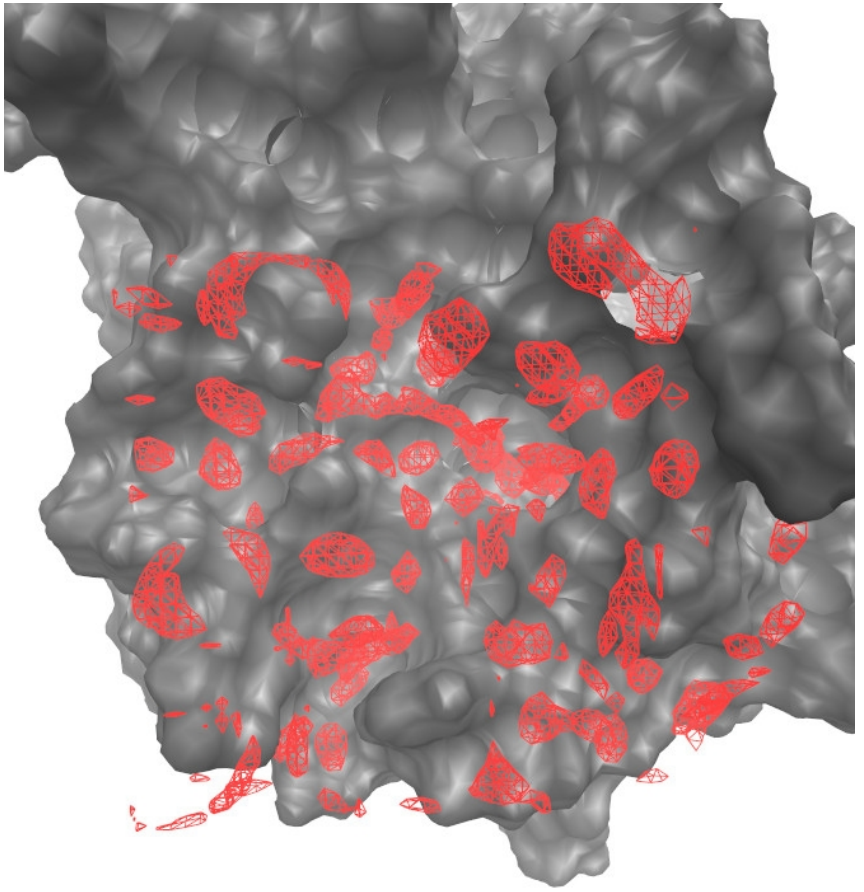


$$q_{tet}(i) = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4} cos(\phi_{ijk} + \frac{1}{3})^2$$

GRID INHOMOGENEOUS SOLVATION THEORY: EXAMPLES

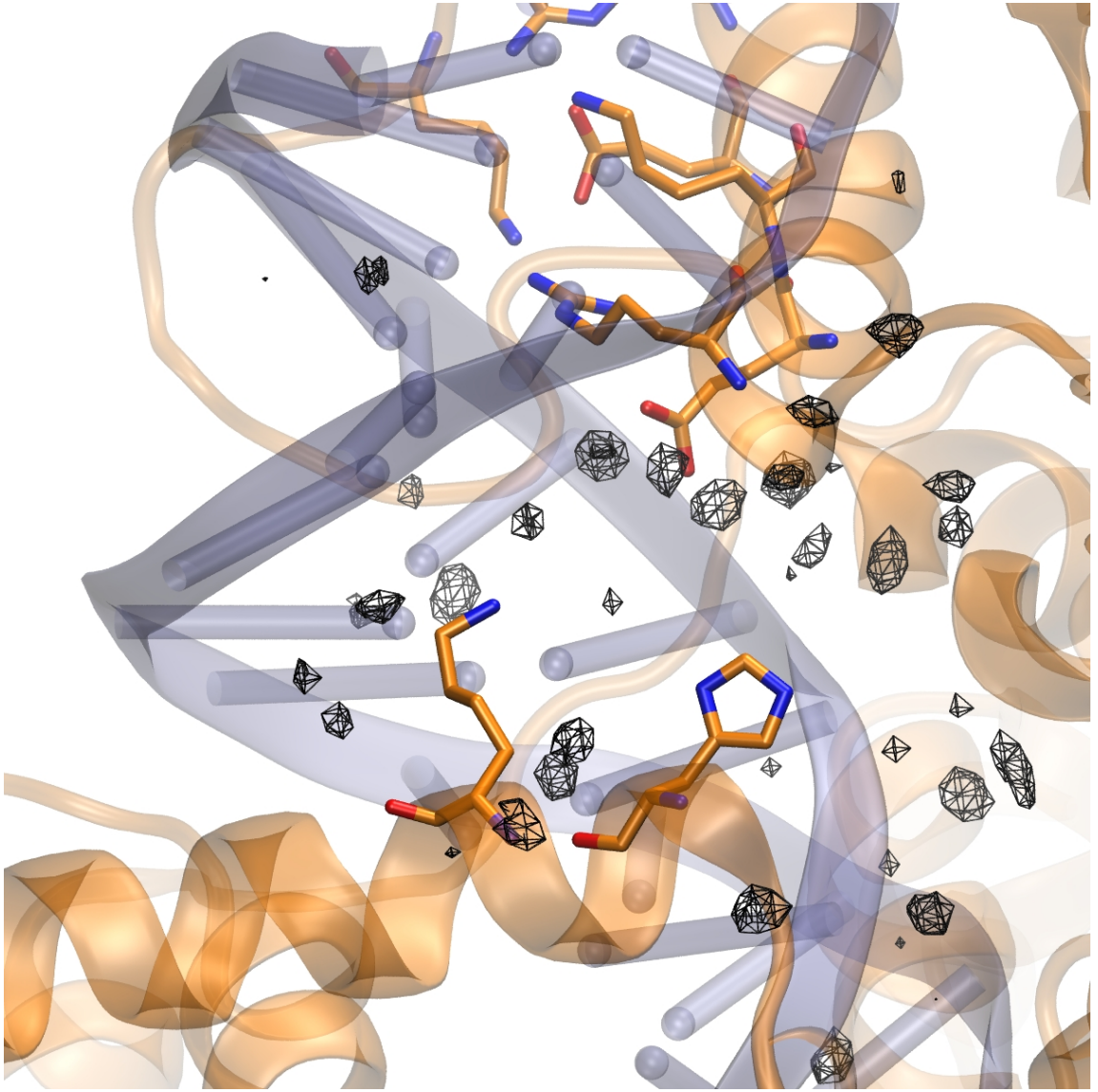


E(solute-water)



E(water-water)

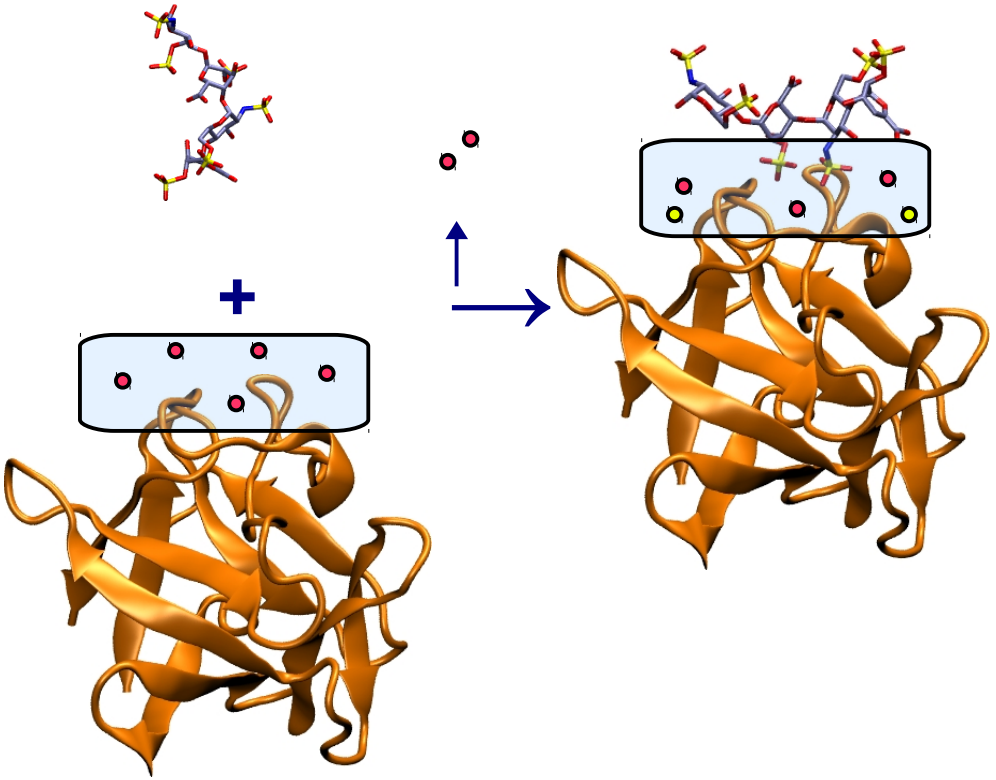
GRID INHOMOGENEOUS SOLVATION THEORY: EXAMPLES



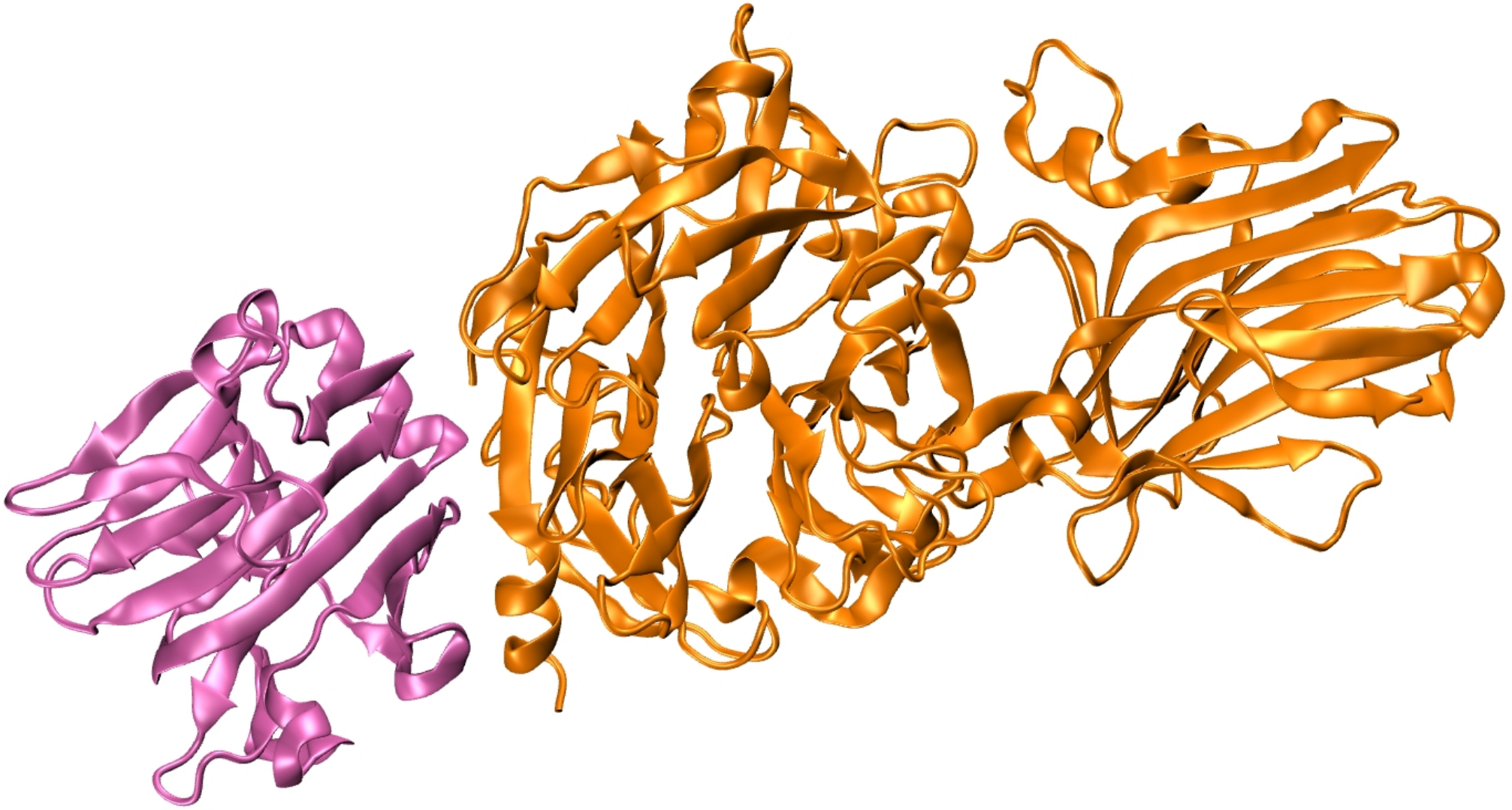
Full free energy (E+S)

SOLVENT CHALLENGE IN DOCKING

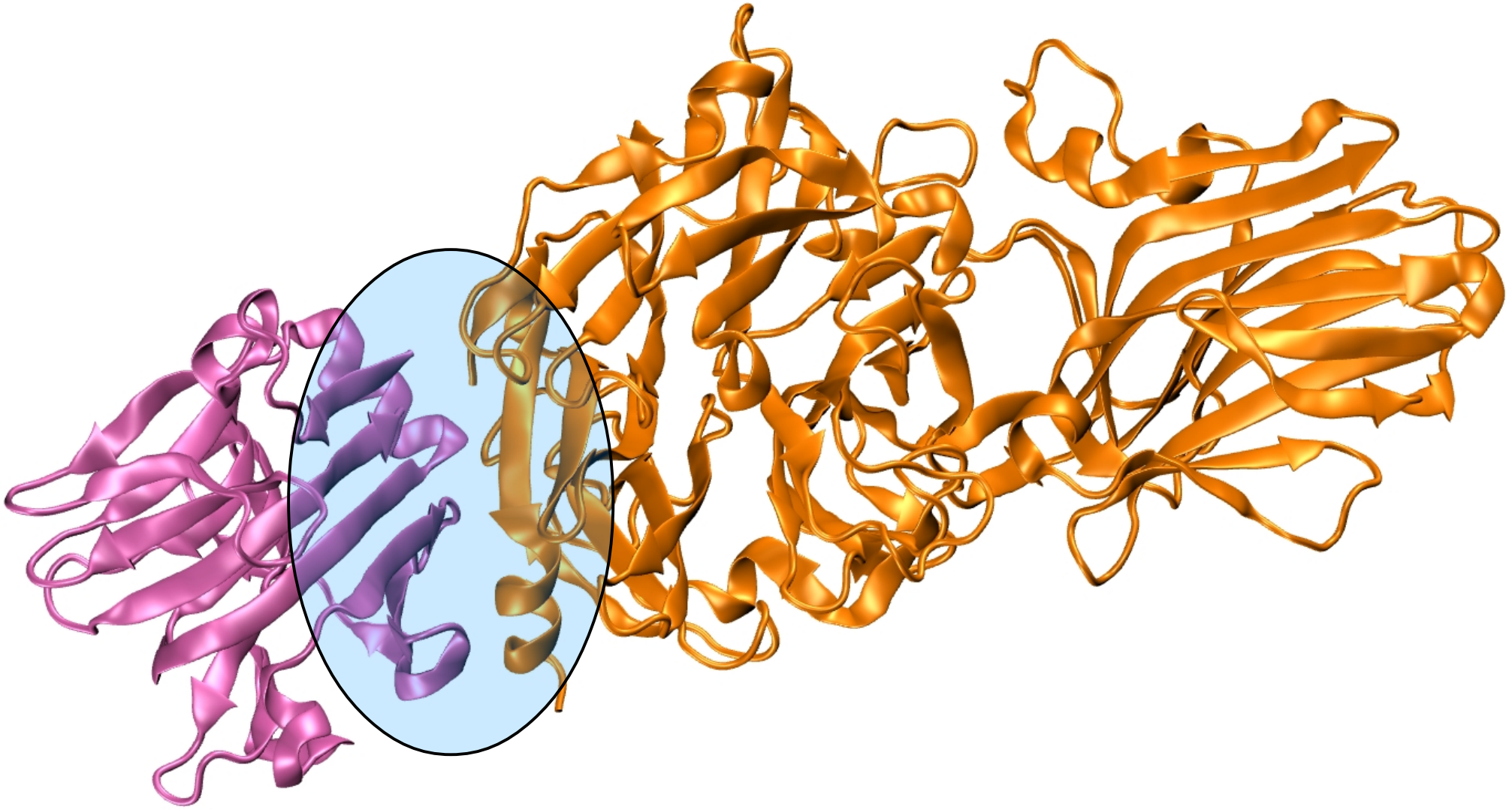
- Implicit solvation
- Explicit solvation:
  - receptor
  - ligand
- Crystal water molecules
- Calculated water molecules:
  - displaced water molecules
  - 'new' water molecules
- Approaches:
  - Monte Carlo
  - Systematic search



PROTEIN INTERFACES

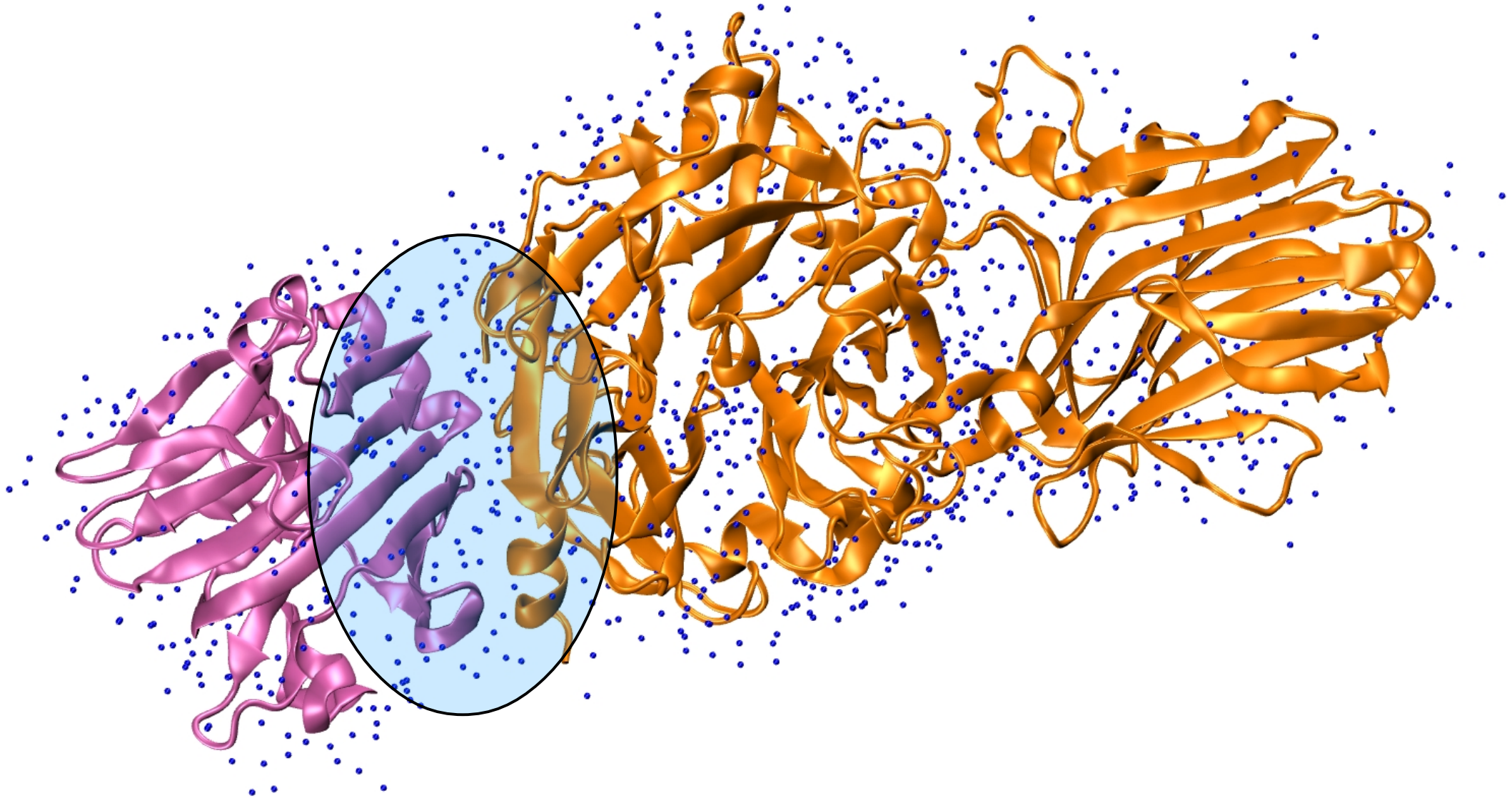


PROTEIN INTERFACES

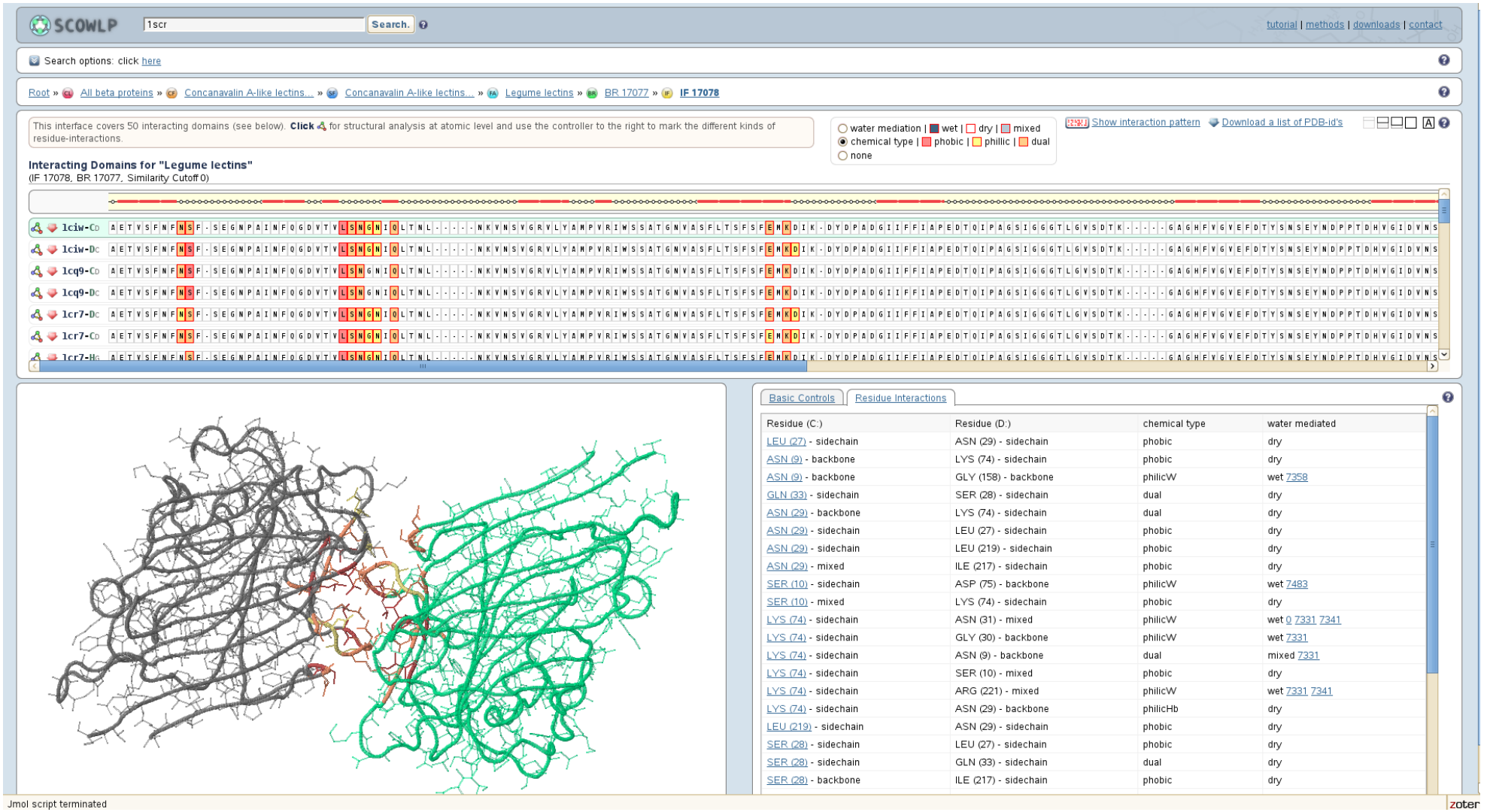


- Protein-protein interface is the part of the space, where protein-protein interaction occurs

SOLVENT IN PROTEIN INTERFACES



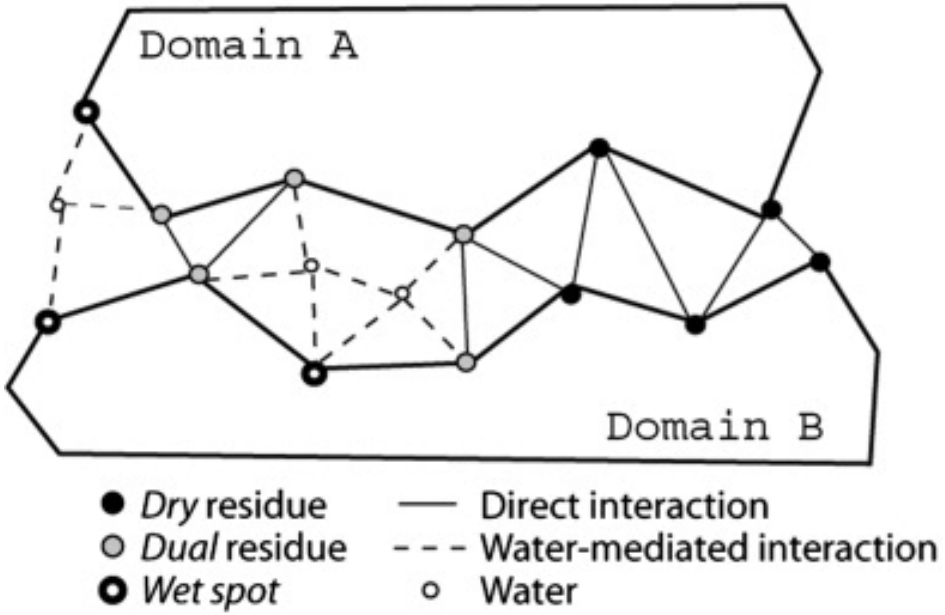
SCOWLP.ORG

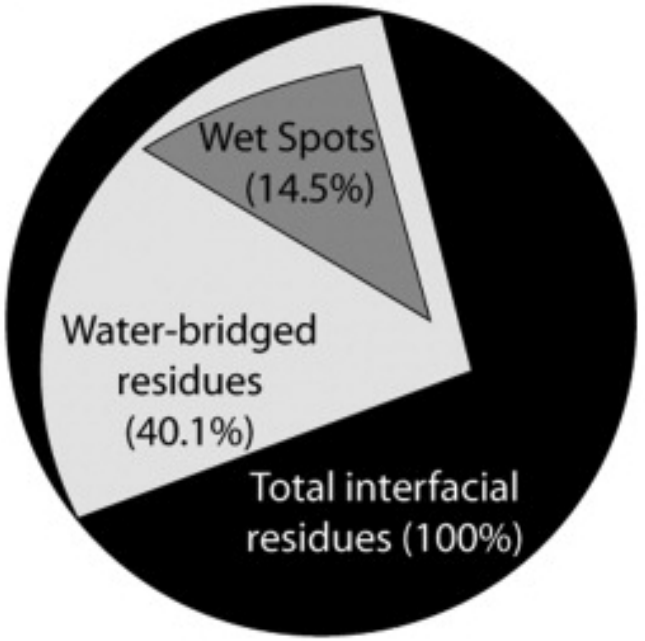


- **SCOWLP is a structural classification of protein binding regions at family level based on the structural classification of proteins, SCOP.**

INTERFACE DEFINITIONS IN SCOWLP

- Structural Characterization of Water, Ligands, and Proteins

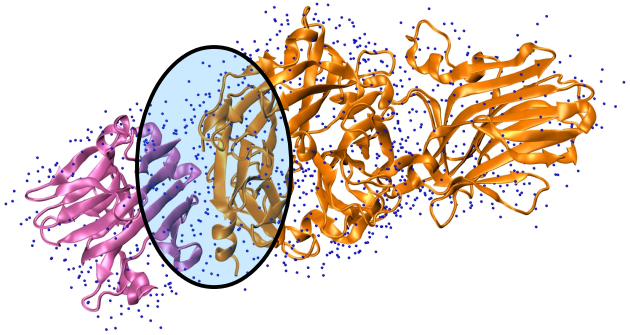




Interactions definition:

- **H-bond: 3.2 Å**
- **Salt bridge: 4.0 Å**
- **VDW: $R_{1\ VDW}+R_{2\ VDW}$**

Water-mediated interactions are important



CASE STUDY I

- MD study of the role of water in protein-protein interfaces

- Introduction of solvent information for protein contacts prediction

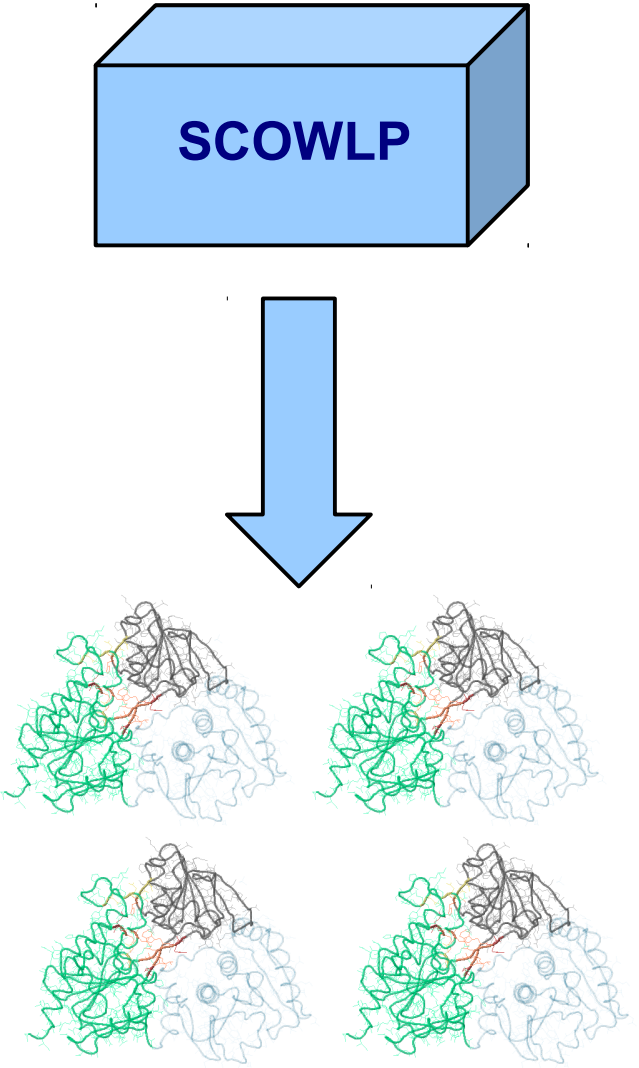
Goals:

- To analyze dynamics and energetics of interfacial residues and interfacial solvent
- To analyze of water role in conservation of protein interfaces
- To improve protein contacts prediction by taking into account solvent data from the PDB

CHOOSING DATASET

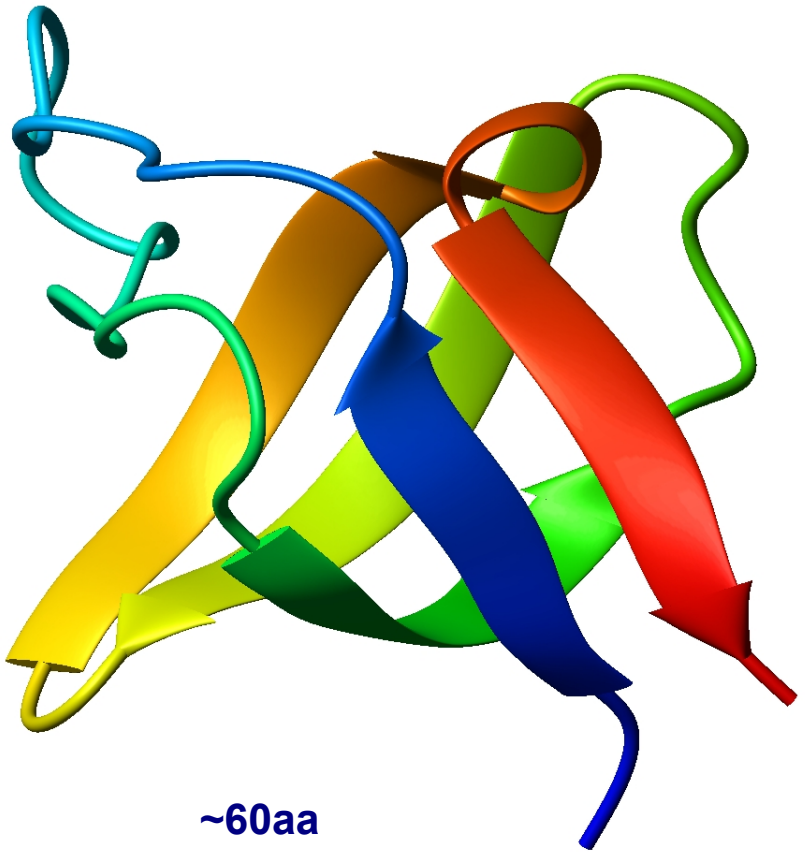
Criteria:

- Representativity:
  - Many members in the family
  - Families with different interfaces
- High resolution (X-Ray structures < 2.5 Å)



MD STUDY DATASET

SH3

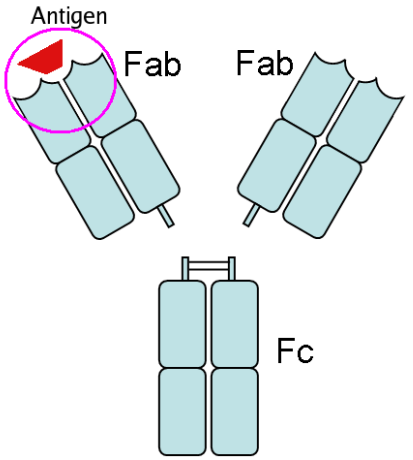


~60aa

ΔASA= (733 ± 195) Å²

7 protein-peptide complexes
4 protein-protein complexes

Immunoglobulin



H, L chains ~220aa

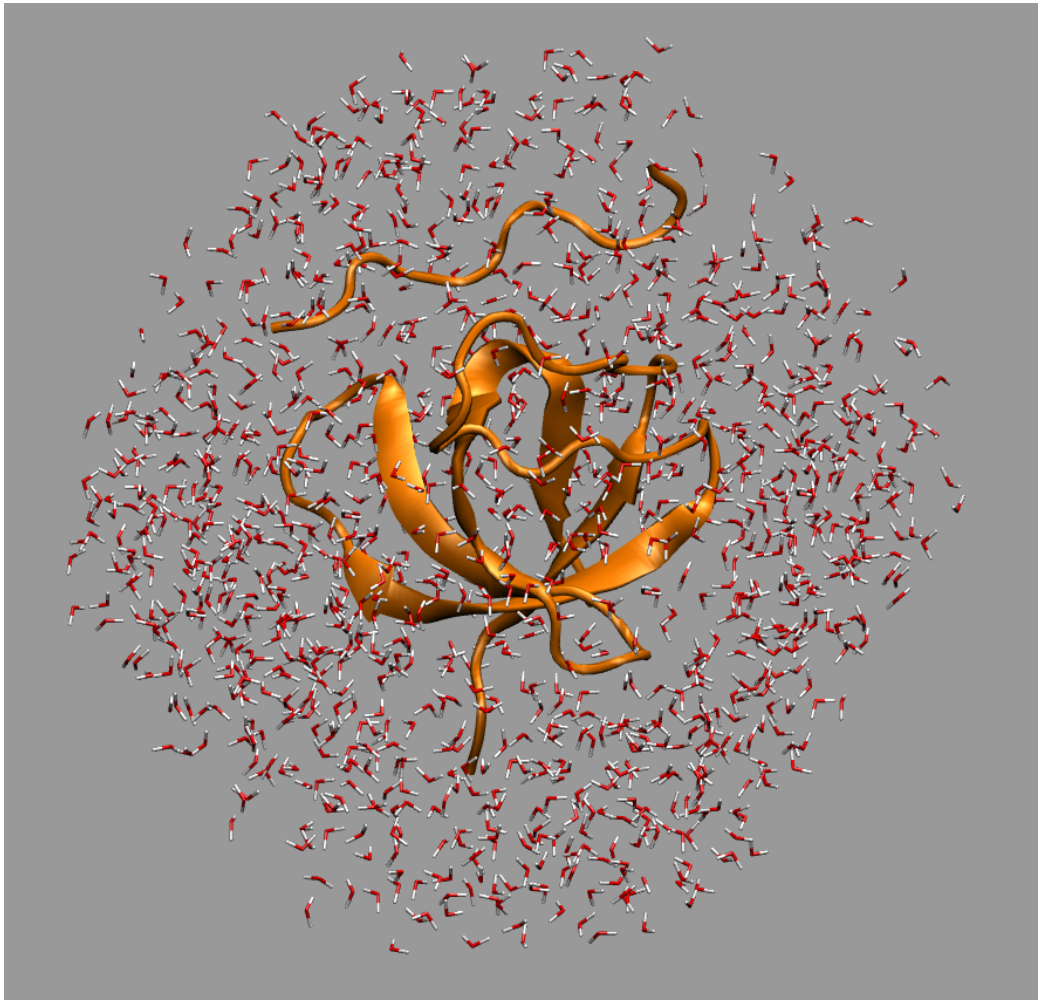
ΔASA= (1291 ± 471) Å²

3 protein-peptide complexes
3 protein-protein complexes

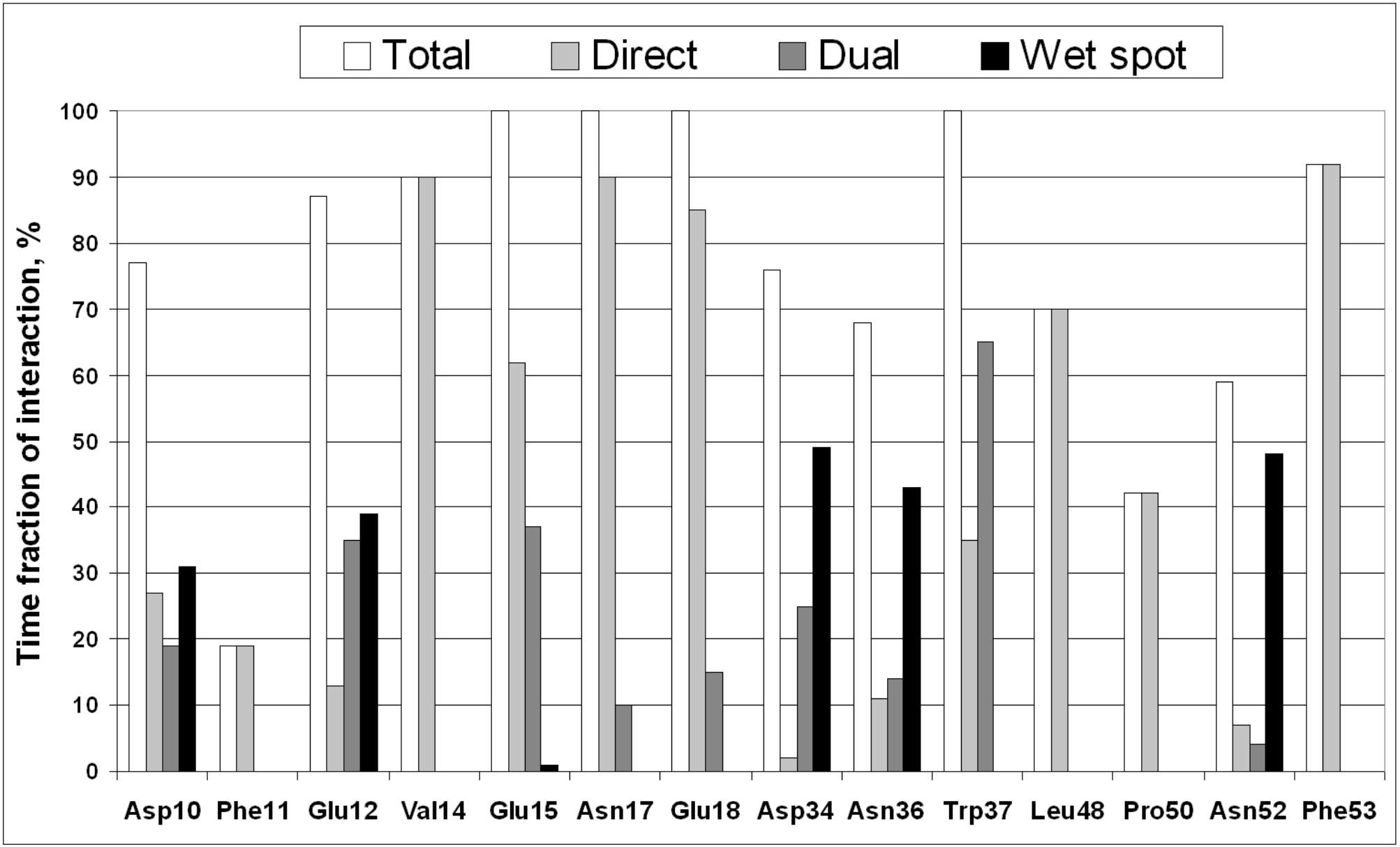
$$\delta\, ASA = \frac{1}{2}\left(ASA(molecule\,1) + ASA(molecule\,2) - ASA(complex) \right)$$

MD SIMULATIONS

- AMBER 8.0
- 10 ns
- Explicit solvent (TIP3P)
- PBC



RELATIVE TIME FRACTIONS (TFS) OF INTERACTIONS



SCOWLP wet spots: Glu12, Asp34, Asn52

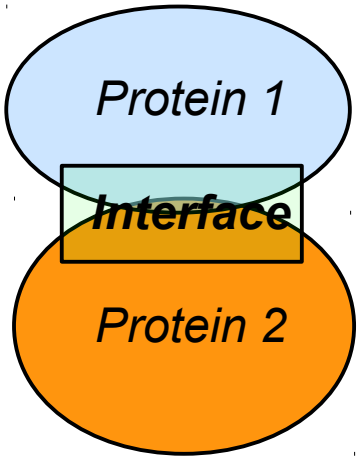
RELATIVE TIME FRACTIONS (TFS) OF INTERACTIONS

In MD analysis each residue is described by TFs and does not belong disambiguously to one of the interfacial classes

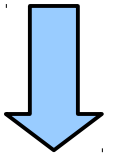
GEOMETRIC SIZES OF INTERFACES

$$\Delta ASA_d = \sum_i \Delta ASA_i \left(TF_{D,i} + \frac{1}{2} TF_{d,i}\right)$$

$$\Delta ASA_w = \sum_i \Delta ASA_i \left(TF_{ws,i} + \frac{1}{2} TF_{d,i}\right)$$



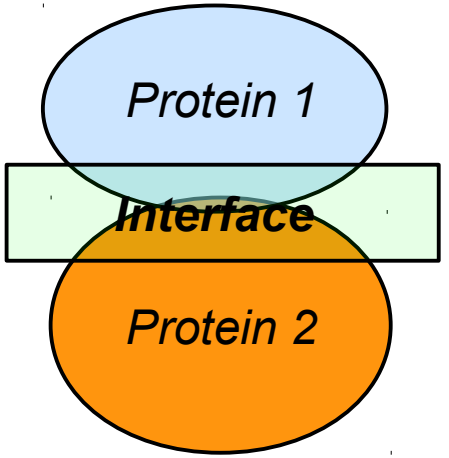
Relative increase of the interface sizes are:



+ water

$\Delta ASA_w/\Delta ASA_d$ = 0.28 ± 0.07 for SH3

$\Delta ASA_w/\Delta ASA_d$ = 0.39 ± 0.13 for Ig



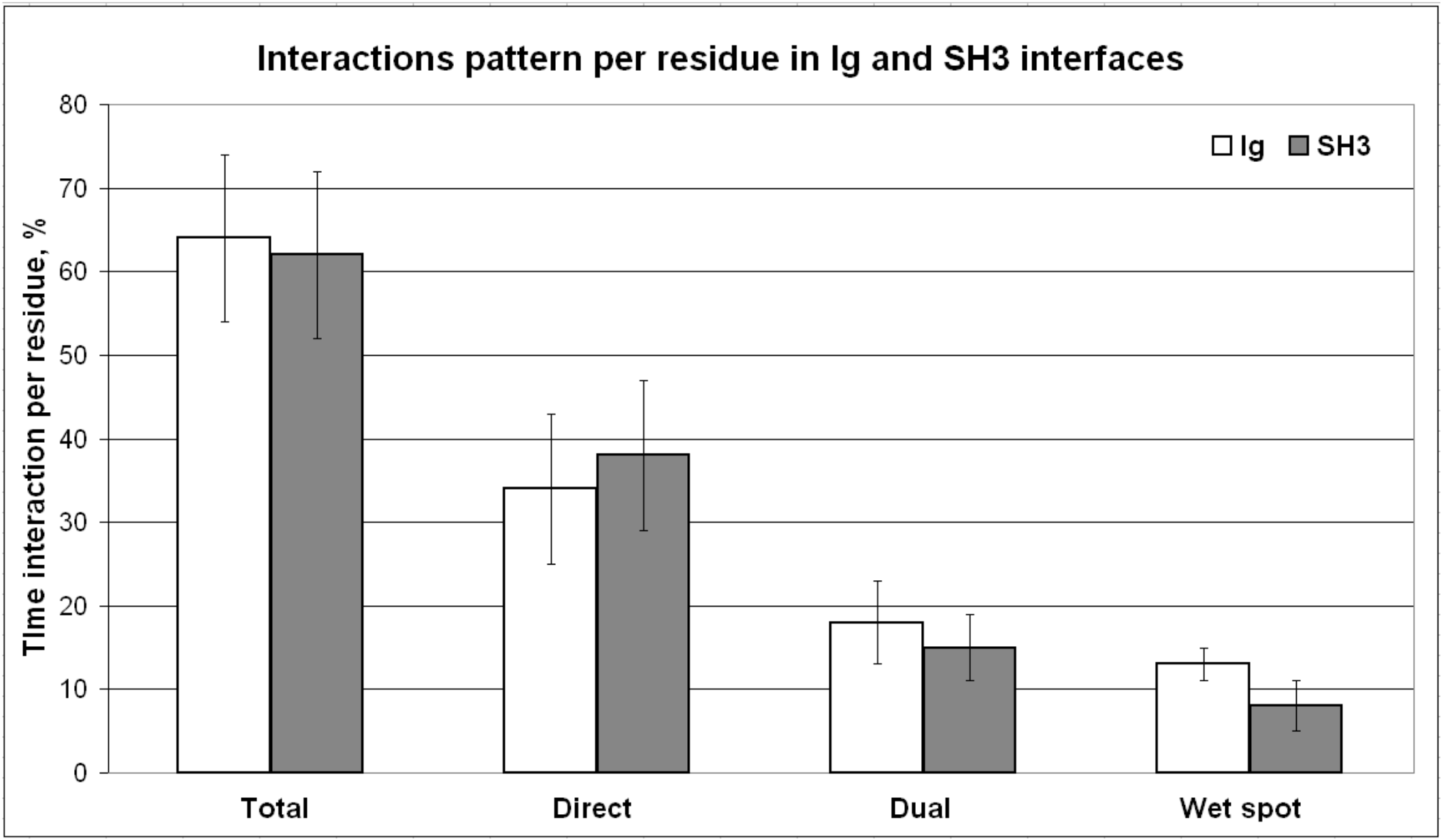
GEOMETRIC SIZES OF INTERFACES

Inclusion of water-mediated interactions in the

interface definition essentially

increases interface size

INTERACTIONS PATTERNS OF IG AND SH3



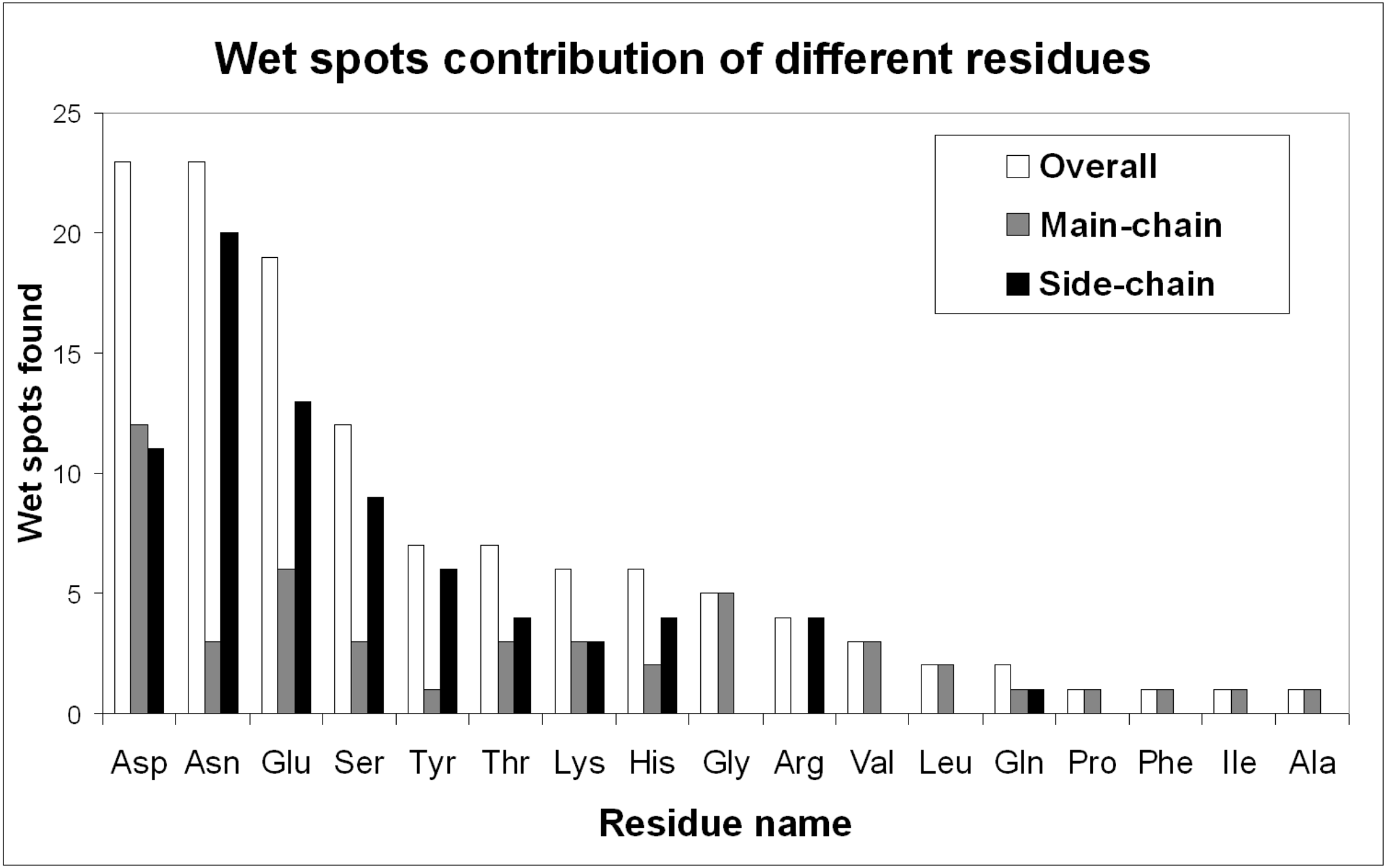
INTERACTIONS PATTERNS OF IG AND SH3

Amount of water-mediated

interactions is comparable with

amount of direct interactions

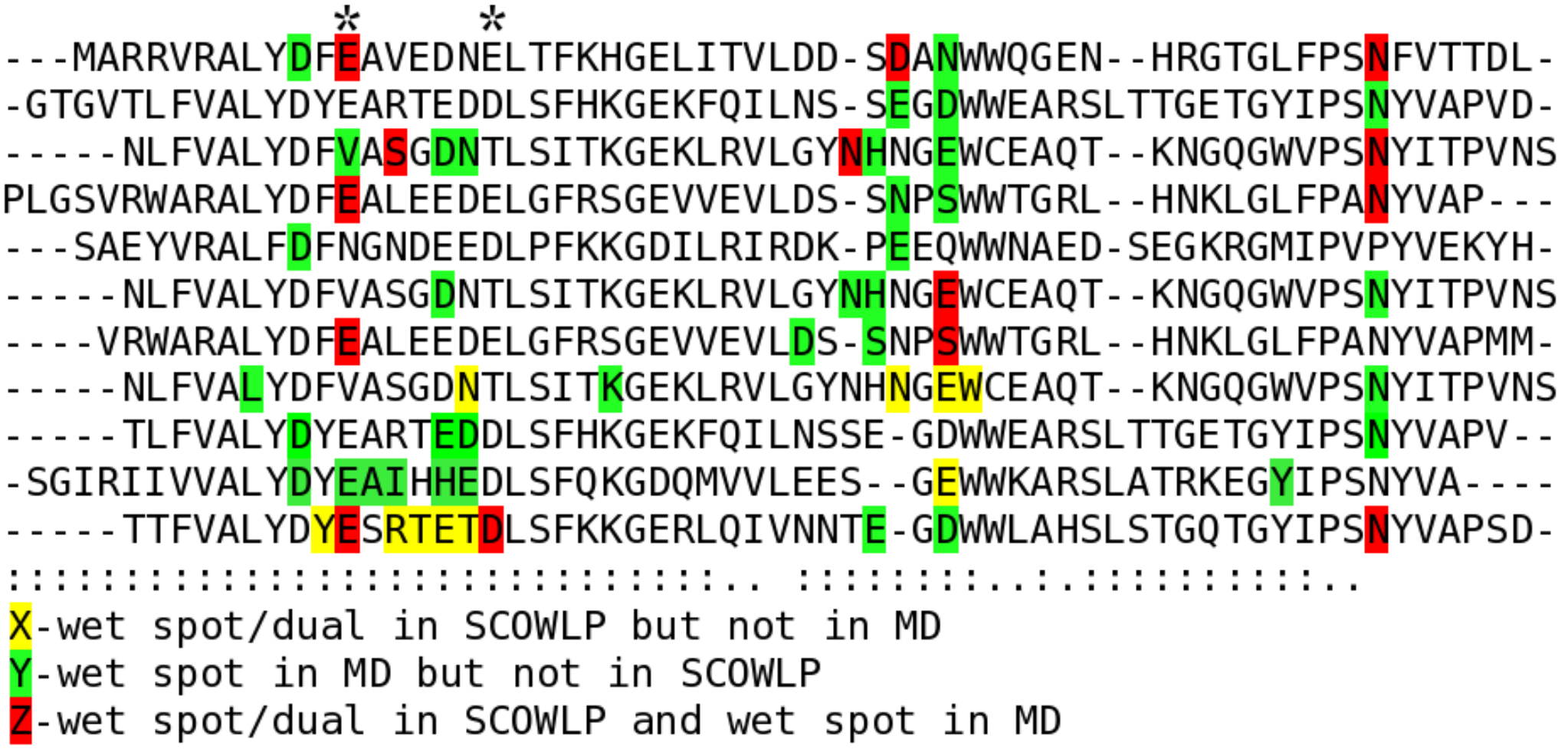
PATTERN OF WET SPOTS INTERACTIONS



PATTERN OF WET SPOTS INTERACTIONS

Water-mediated interactions increase the probability of hydrophobic residues to be an active part of hydrophilic interfaces

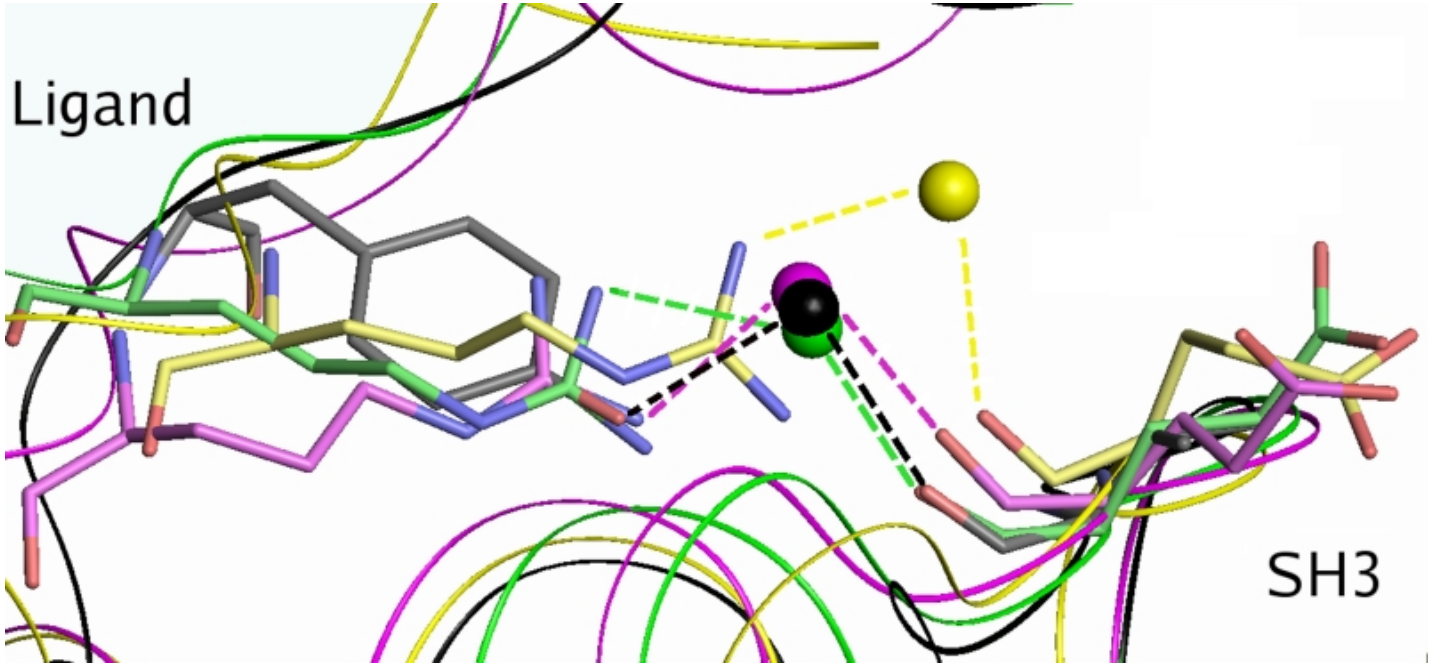
CONSERVATION OF WATER-MEDIATED INTERACTIONS OF SH3 DOMAINS



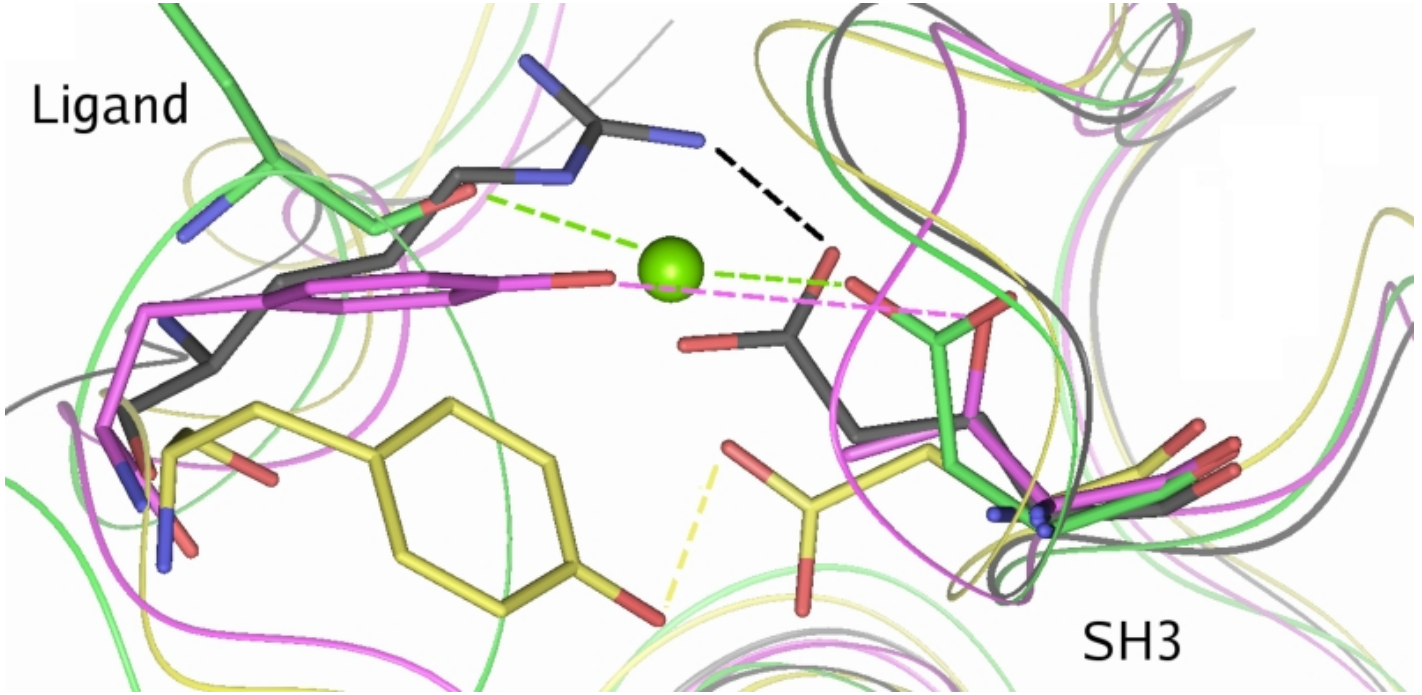
CONSERVATION OF WATER-MEDIATED INTERACTIONS OF SH3 DOMAINS

Interaction conservations vs. sequence/structural conservation

INTERACTIONS CONSERVATION



SH3	Ligand
E12	R64
V10	Y63
E15	R65
E11	R66



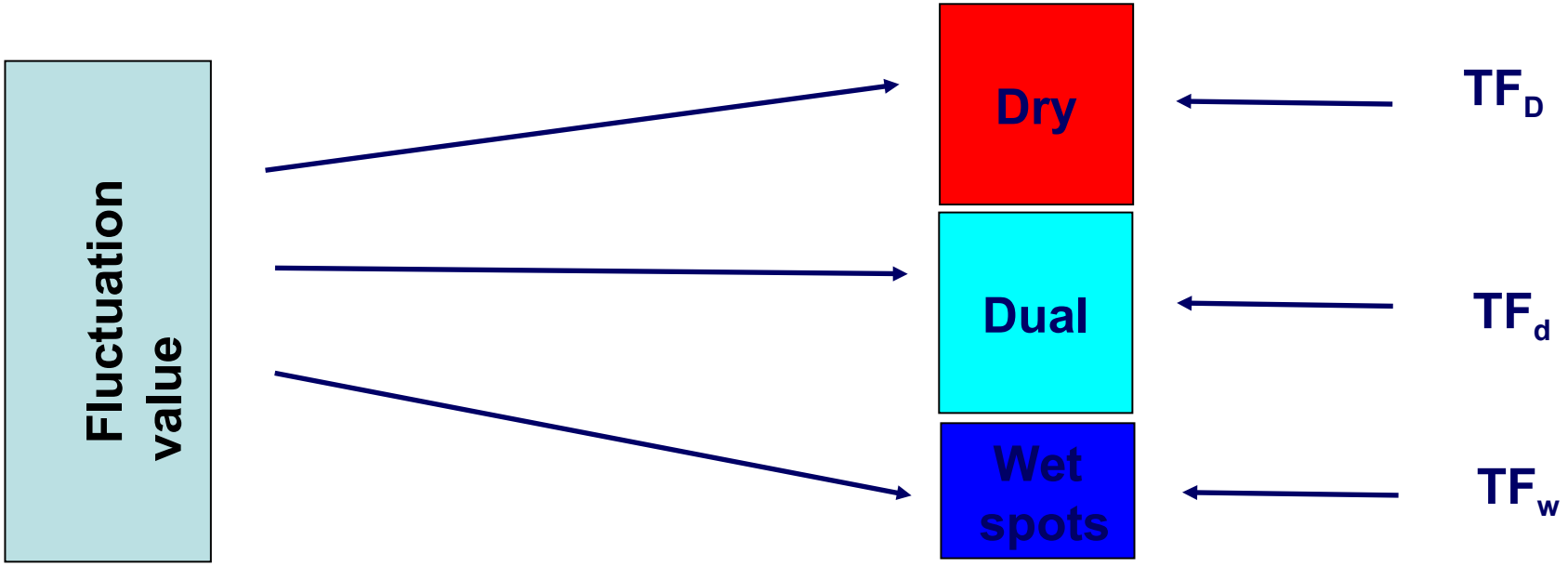
SH3	Ligand
D16	I172
E18	R64
D20	Y66
T16	Y63

INTERACTIONS CONSERVATION

Water molecules as a part of interfaces contribute to the conservation of protein-protein interactions

FLUCTUATIONS OF INTERFACIAL RESIDUES

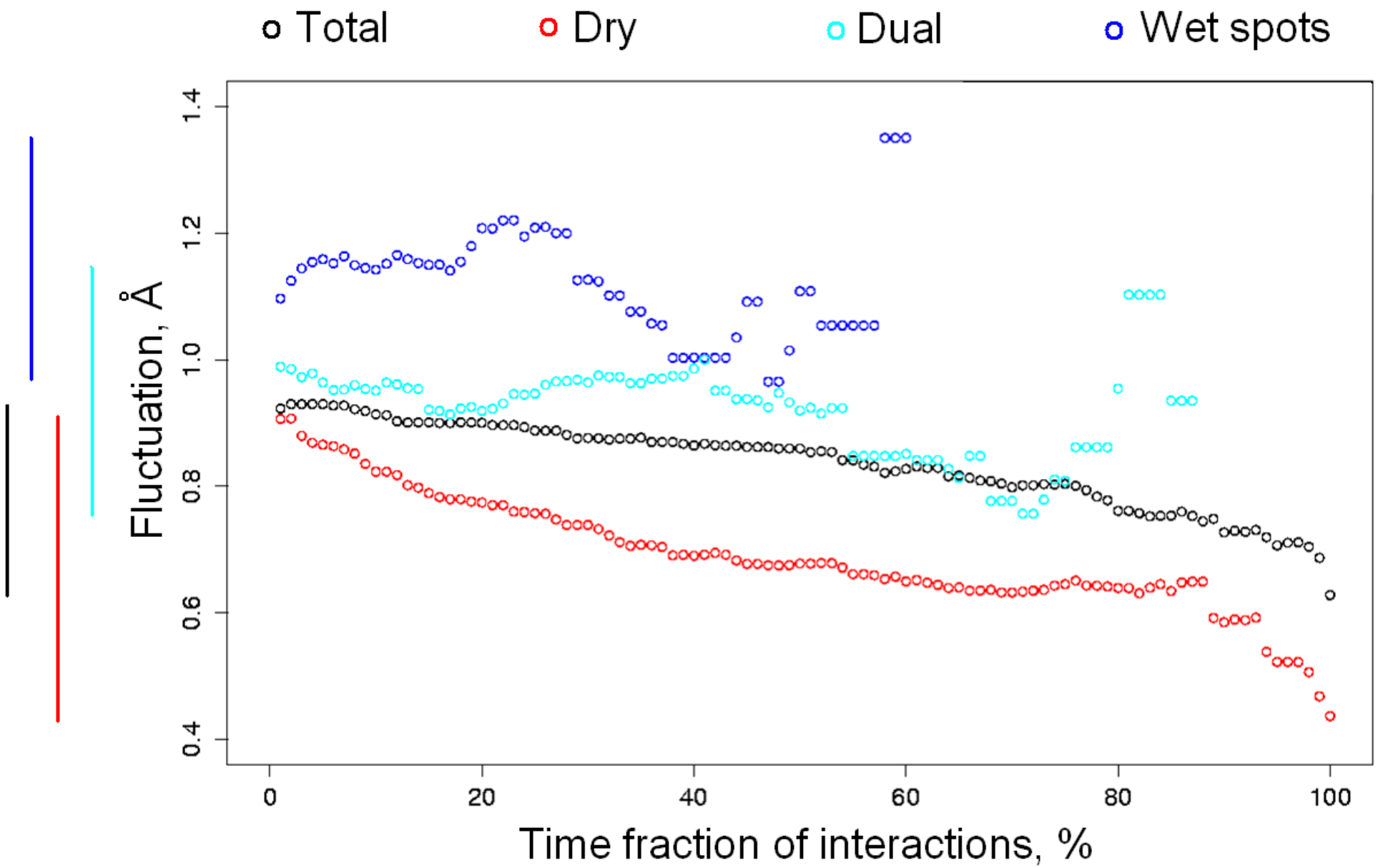
$$F_i^2 = < R_i^2 > - < R_i >^2$$



$F\left(TF_t, TF_D, TF_d, TF_w\right)$ **is analytically unknown fluctuation function**

$$< F(TF_k > x) >_{i,j} = function(TF_k); i \neq k; x \in [0; 100]$$

FLUCTUATIONS OF INTERFACIAL RESIDUES



FLUCTUATIONS OF INTERFACIAL RESIDUES

Wet spots are less mobile than protein surface residues

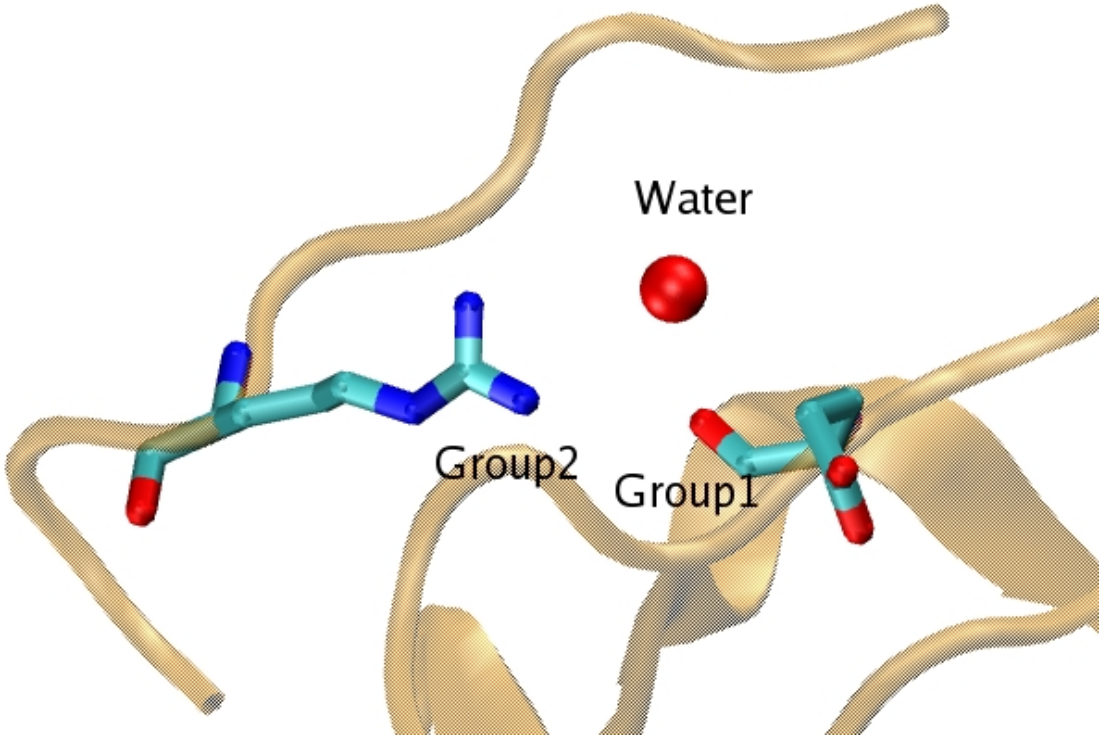
but more mobile than dry residues

MM-PBSA ANALYSIS OF INTERFACIAL RESIDUES

Wet spots, dual and dry interfacial residues are

energetically comparable

RESIDENCE TIME OF INTERFACTIAL WATER



$d_{Group1\text{-}O(H_2O)} < 3.6$ Å

$d_{Group2\text{-}O(H_2O)} < 3.6$ Å

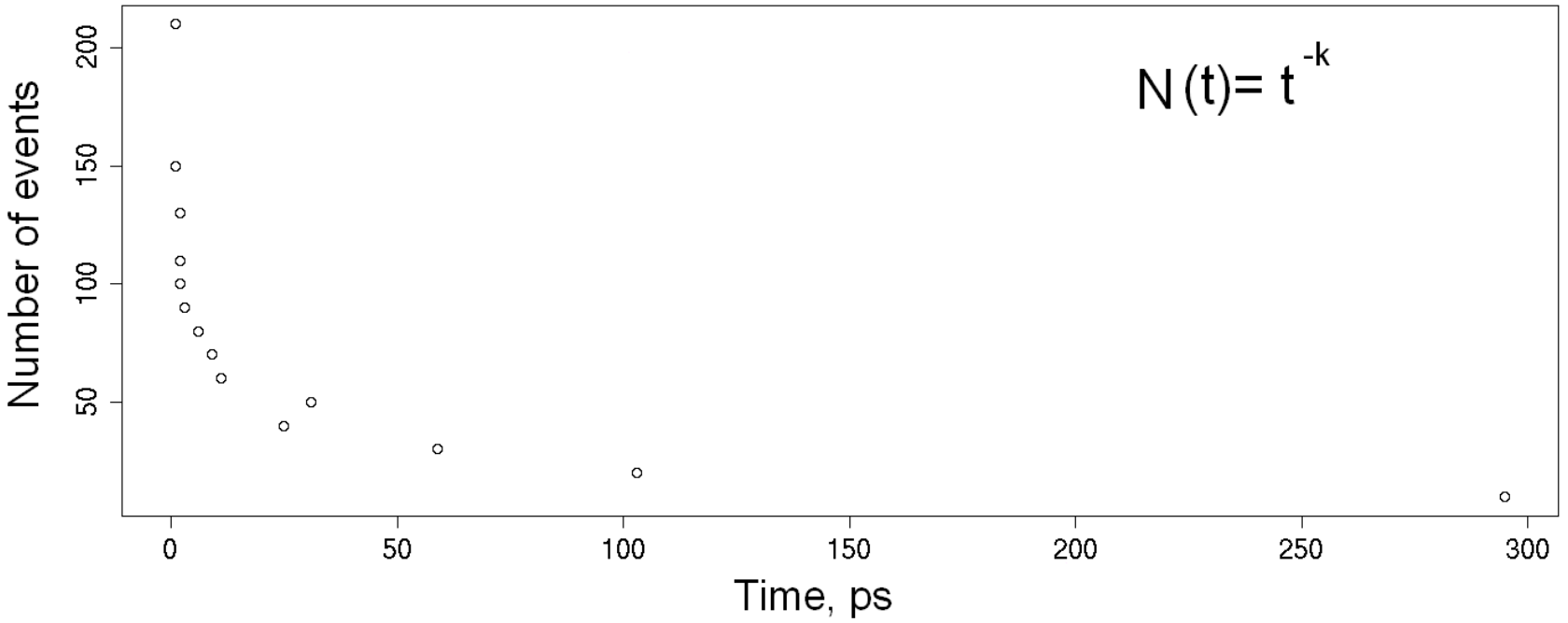
Site is occupied

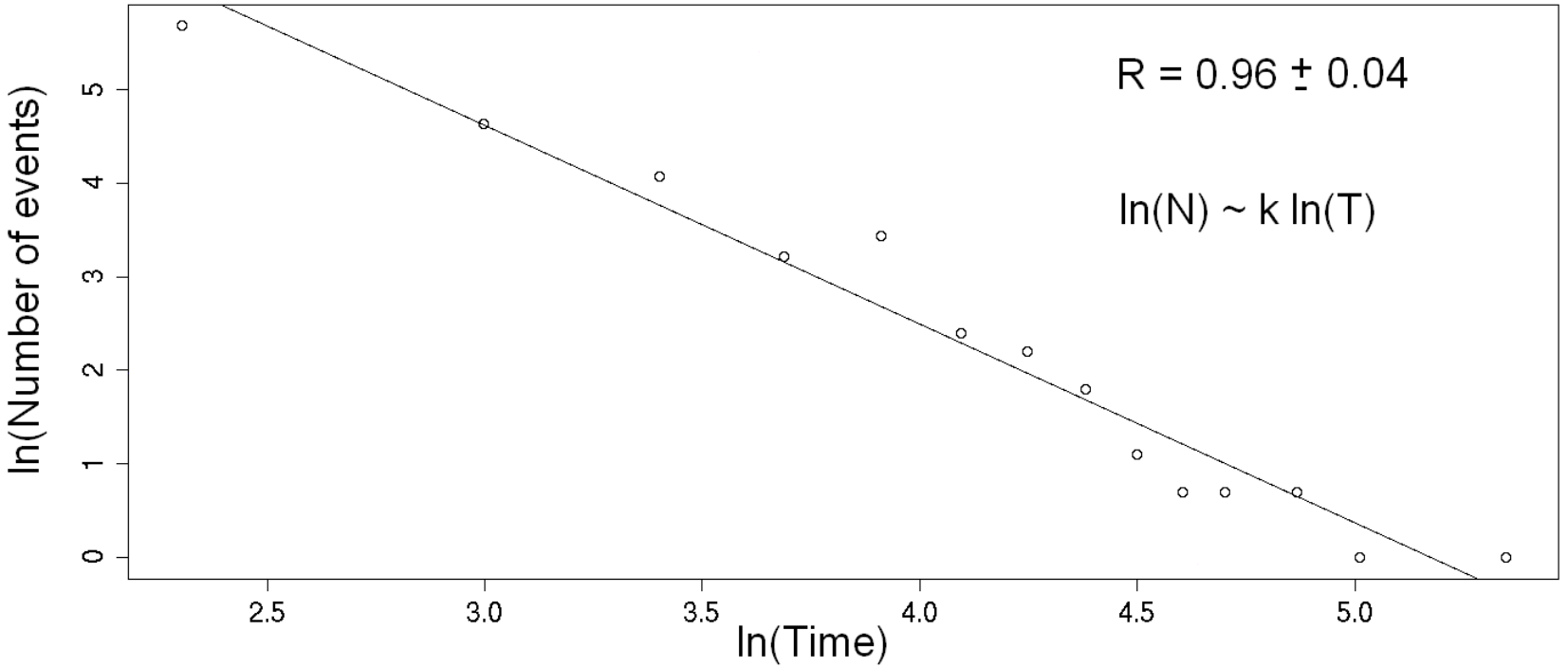
Maximum residence time

Total occupancy

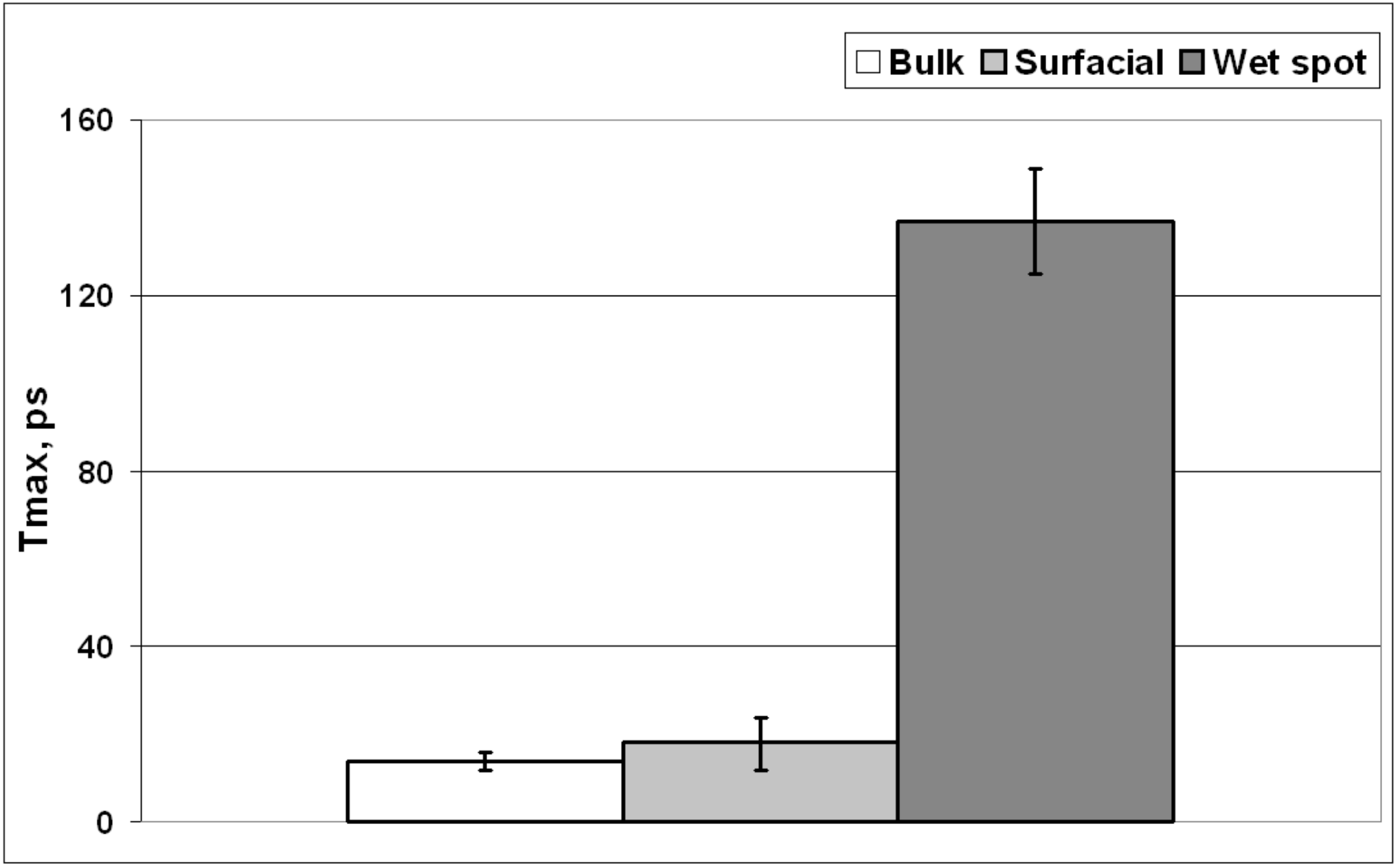
Residence time distribution

RESIDENCE TIME OF INTERFACTIAL WATER





MAXIMUM RESIDENCE TIME



RESIDENCE TIME OF INTERFACTIAL WATER

Water molecules in wet spots have longer residence time

than the ones on protein surface

FREE ENERGY PERTURBATION DOUBLE DECOUPLING METHOD

$$AB(sol) \longrightarrow A(sol) + B(gas) \qquad \Delta G_1^\circ$$

$$B(sol) \longrightarrow B(gas) \qquad \Delta G_2^\circ$$

$$A(sol) + B(sol) \longrightarrow AB(sol) \qquad \Delta G_{AB}^\circ = \Delta G_2^\circ - \Delta G_1^\circ$$

A- protein B - solvent (water)

Double decoupling method

1. Simulation with disappearing charge
2. Simulation with disappearing VDW radius while charge is 0

FREE ENERGY PERTURBATION (1UJ0 EXAMPLE)

Site / E, kcal/mol	Water site type	Elect	VDW	-RT* $\ln(S_a S_b / S_{a*b})$	RT *$\ln(C_0 V_1)$	$\Delta G^0_{1(2)}$	ΔG^0
E12-R64	Wet spot	12.9	-1.5	0.4	-4.4	7.4	-1.4
D34-N66	Wet spot	8.3	0.1	0.4	-4.1	4.7	1.3
D34-N66, 2 H_2O	Wet spot	22.9	-3.7	0.8	-8.2	12.6	-6.6
N52-M61	Wet spot	8.9	0.1	0.4	-4.2	5.2	0.8
N52-M61, 2 H_2O	Wet spot	18.1	0.1	0.8	-7.2	11.2	0.1
L58- R6	Surface	9.8	0.2	0.4	-3.8	6.4	-0.4
D31-S33	Surface	7.6	-0.6	0.4	-3.7	3.7	2.3
Control: lysozyme	Cavity	13.5	0.0	0.4	-3.9	10.0	-4.0
Bulk→vacuo		8.2	-2.2	-	-	6.0	-
Bulk → vacuo (McCammon, 2004)		8.2	-2.2	-	-	6.0	-
Bulk → vacuo		8.3	-2.4	-	-	5.9	-

FREE ENERGY PERTURBATION (1UJ0 EXAMPLE)

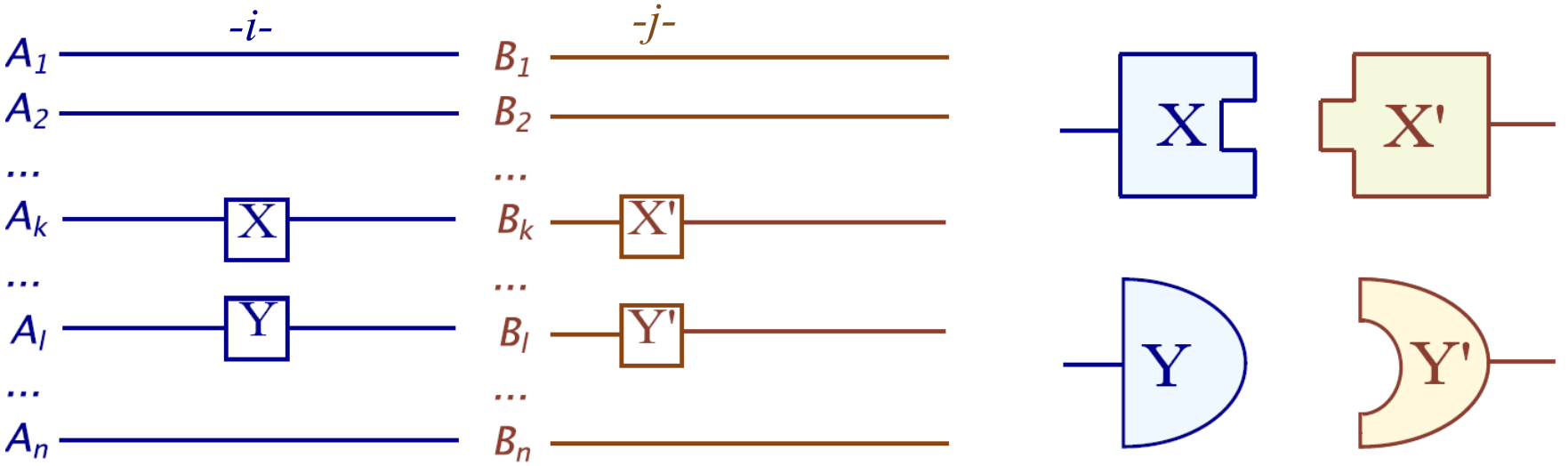
In terms of free energy, interfacial water molecules are very diverse, but significantly affect the free energy of complex formation

CORRELATED MUTATIONS CONCEPT

$$\{A_{k=1..n}\}, \{B_{l=1..n}\} - \textit{domain families}$$

$$r_{ij} = \frac{1}{N^2} \sum_{k,l}^{N} \frac{W_{kl}(S_{ikl} - < S >_i)(S_{jkl} - < S >_j)}{\sigma_i \sigma_j}$$

$$W_{kl} = 1 - \frac{1}{L} \sum_{i=1}^{L} \delta(R_{ik}, R_{il})$$



- **Interacting protein residues coevolve, so that a mutation in one of the interacting counterparts is compensated by a mutation in the other**

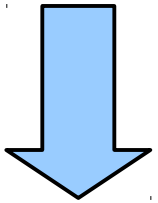
SIMILARITY MATRIX STRUCTURE

	Ala	**Val**	**Ile**	**...**
Ala	1	X(Val-Ala)	X(Ile-Ala)	...
Val	X(Ala-Val)	1	X(Ile-Val)	...
Ile	X(Ala-Ile)	X(Val-Ile)	1	...
...	...	...	...	...

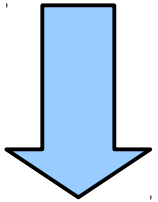
$$X(i\text{-}j) = X(j\text{-}i)$$

OBTAINING WET MATRIX

SCOWLP (PDB)



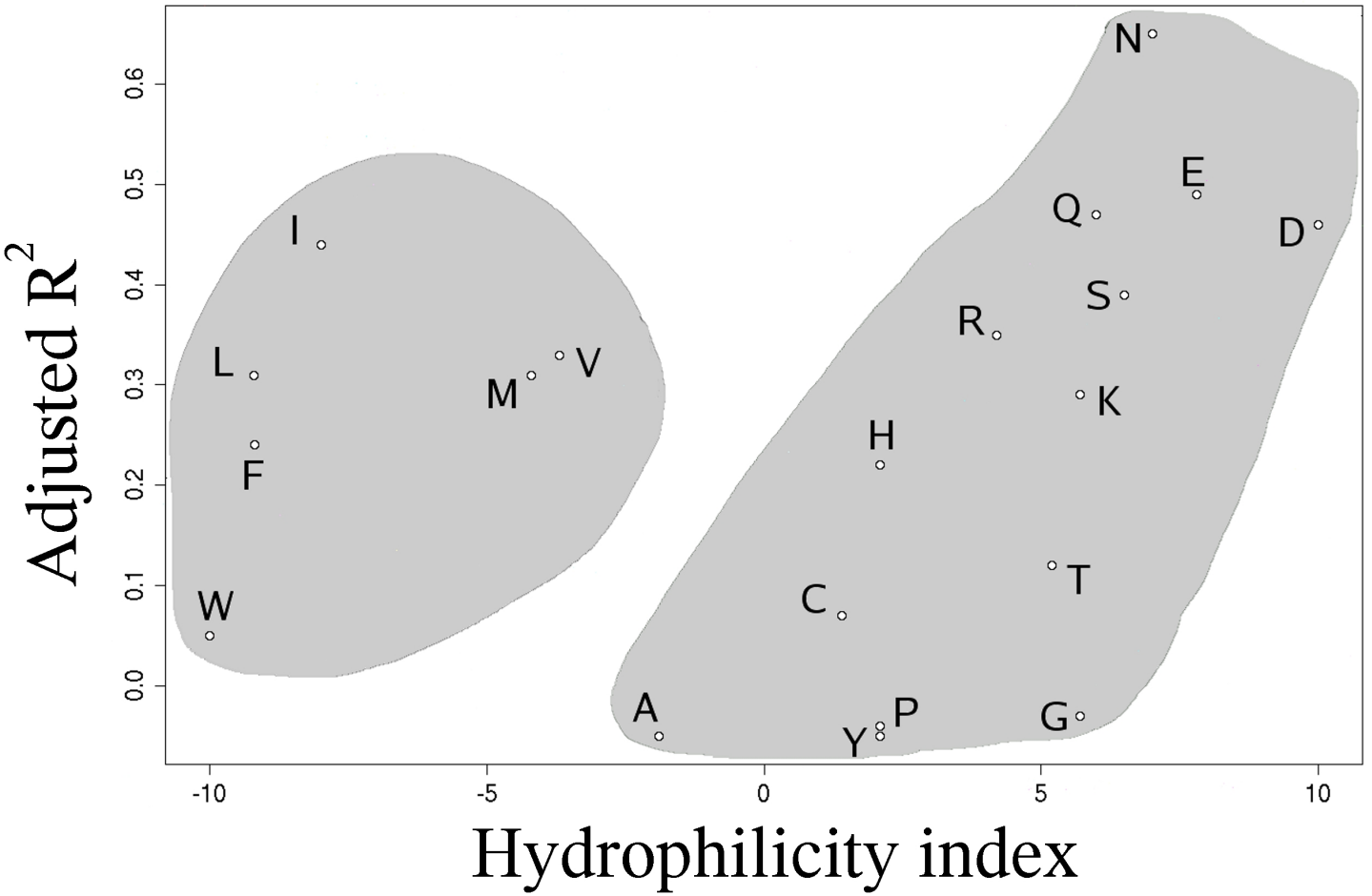
$$p_i = \frac{N_{i,\,water\,contact}}{N_{i,\,total}}$$



$$WET = \{w\}_{ij} = 1 - |p_i - p_j|$$

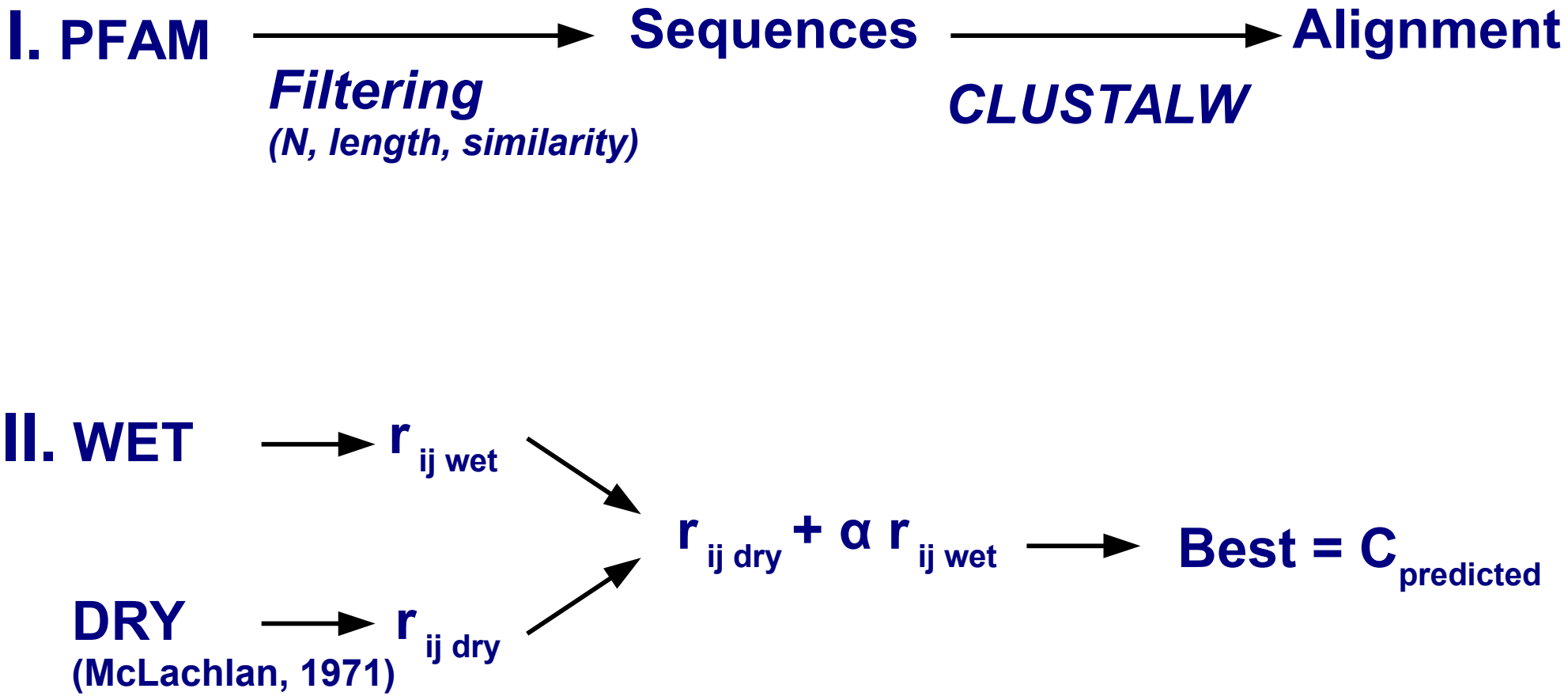
WET VS DRY

Residue	Adjusted R^2
Ala	-0.05
Arg	0.35
Asn	0.65
Asp	0.46
Cys	0.07
Gln	0.47
Glu	0.49
Gly	-0.03
His	0.22
Ile	0.44
Leu	0.31
Lys	0.29
Met	0.31
Phe	0.24
Pro	-0.04
Ser	0.39
Thr	0.12
Trp	0.05
Tyr	-0.05
Val	0.33

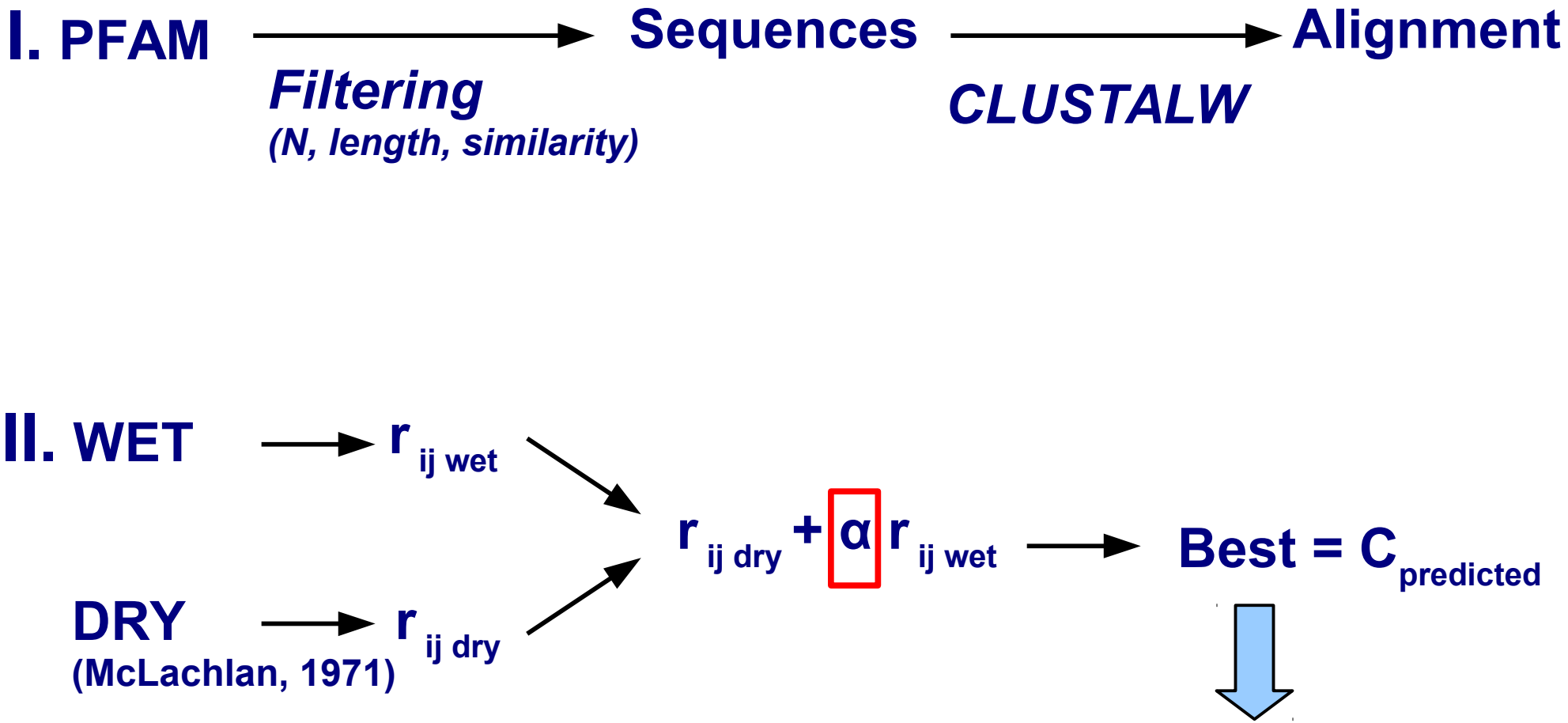


DRY and WET similarity matrices are not completely independent

PREDICTIONS PIPELINE



PREDICTIONS PIPELINE



PREDICTIONS PIPELINE

III. Accuracy = $C_{corr}/C_{predicted}$

Random accuracy = $C_{observed}/C_{max}$

Improvement over random = Accuracy/Random accuracy

$$Wet\ prediction\ ratio(\alpha) = \frac{Accuracy(\alpha)}{Accuracy(\alpha = 0)}$$

$$X_d = \sum_{i=1}^{n} \frac{P_{ic} - P_{ia}}{d_i n}$$

$d_i - distance\ bin\ ;\ n - number\ of\ bins$

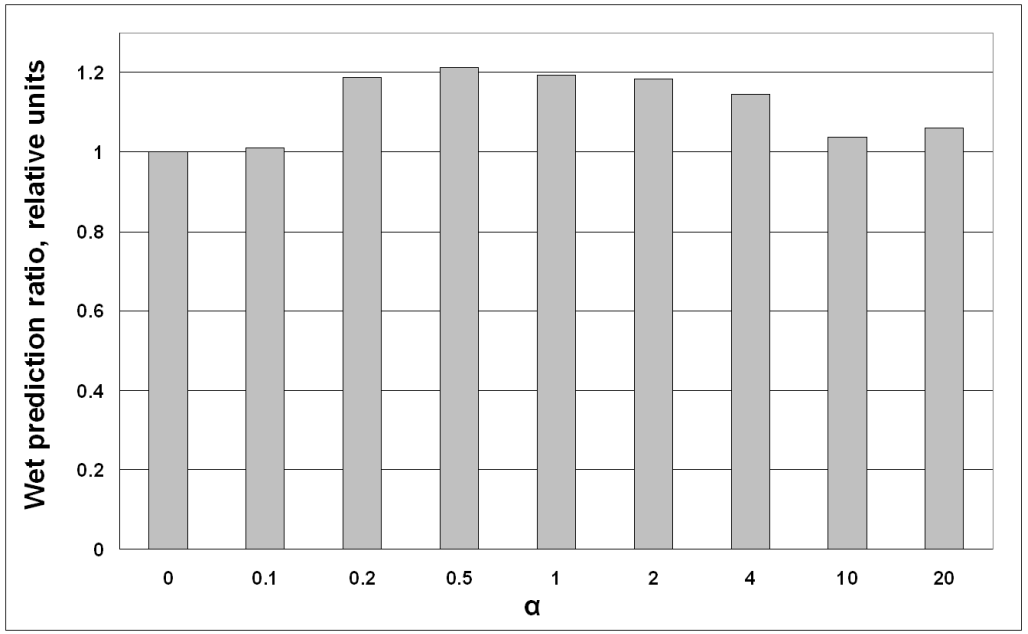
$P_{ic} - correlated\ pairs\ ;\ P_{ia} - all\ pairs$

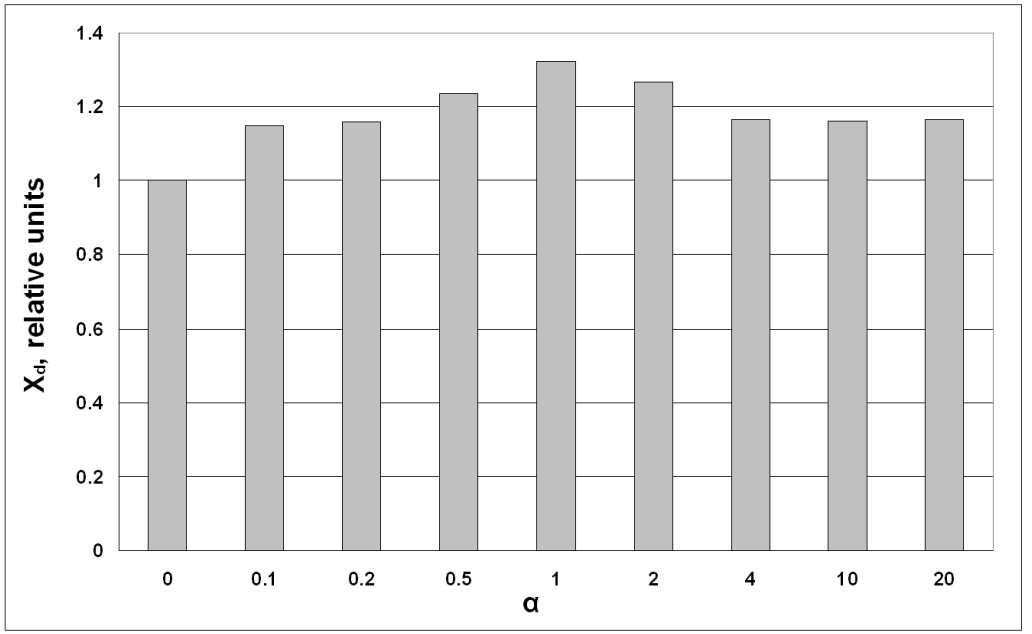
Contacts

Intradomain **Interdomain**

INTRADOMAIN CONTACTS

50 PFAM families
Alignment length: 30-195
Alignment size: 20-295 sequences



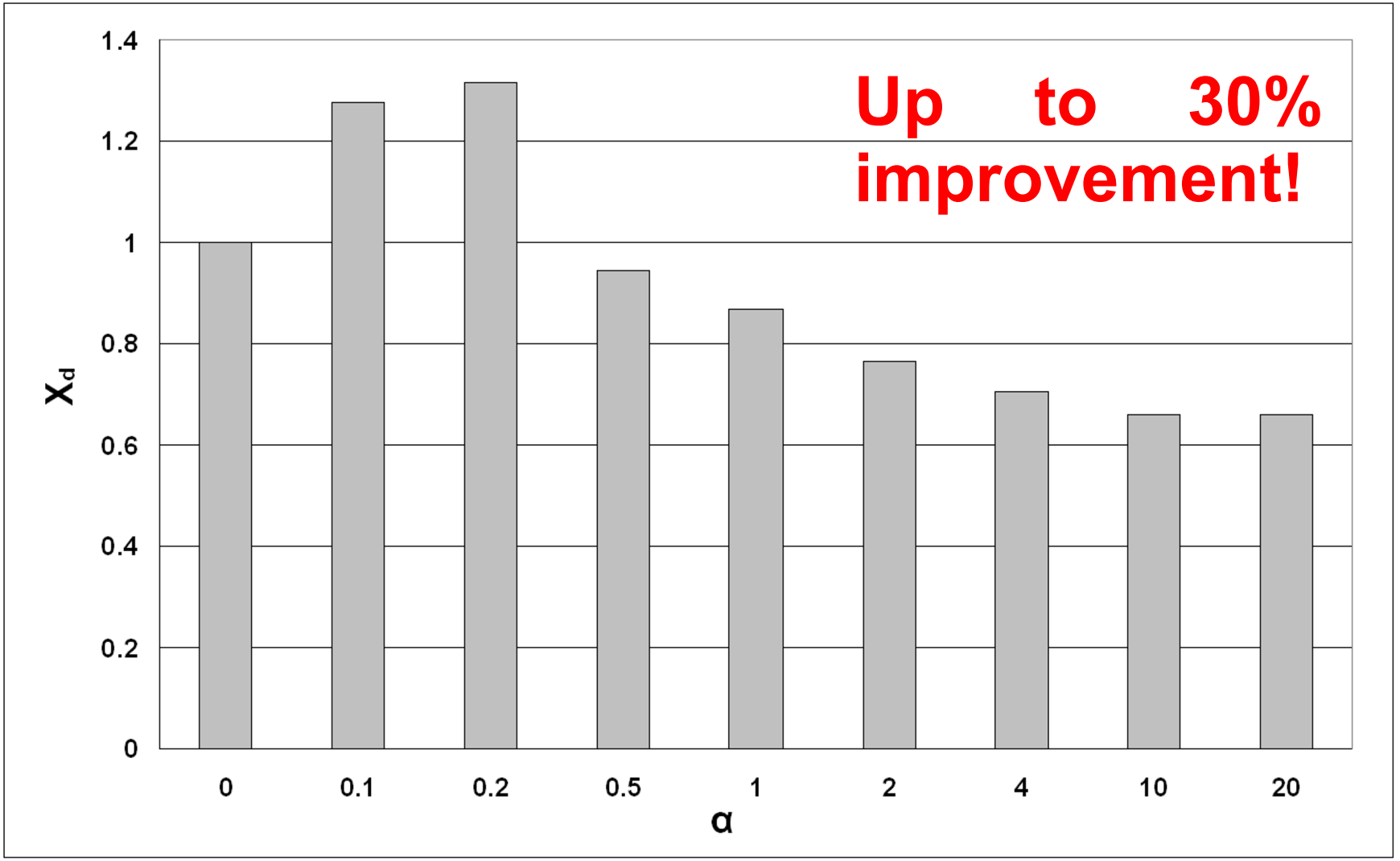


Up to 20% improvement!

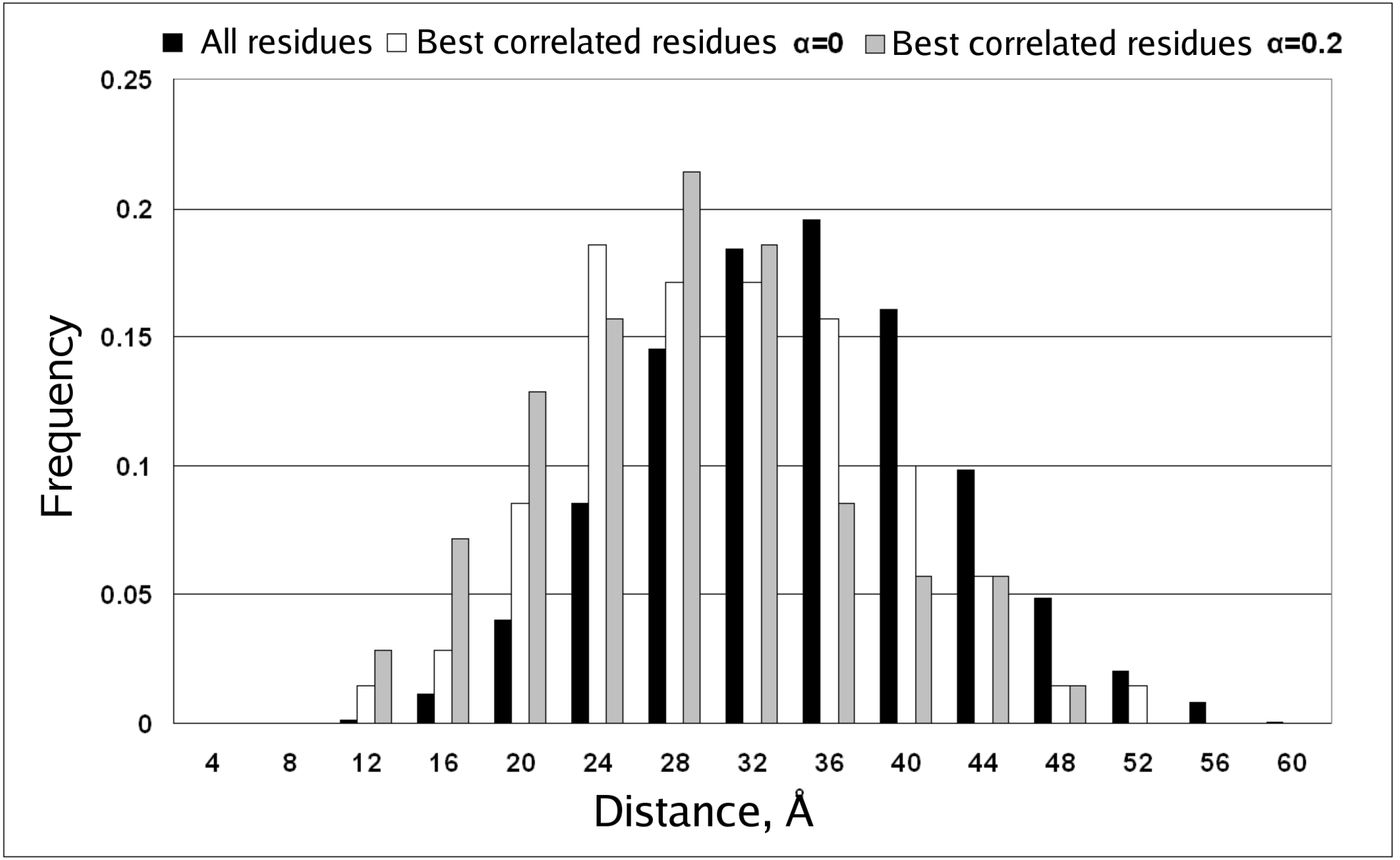
Up to 30% improvement!

INTERDOMAIN CONTACTS

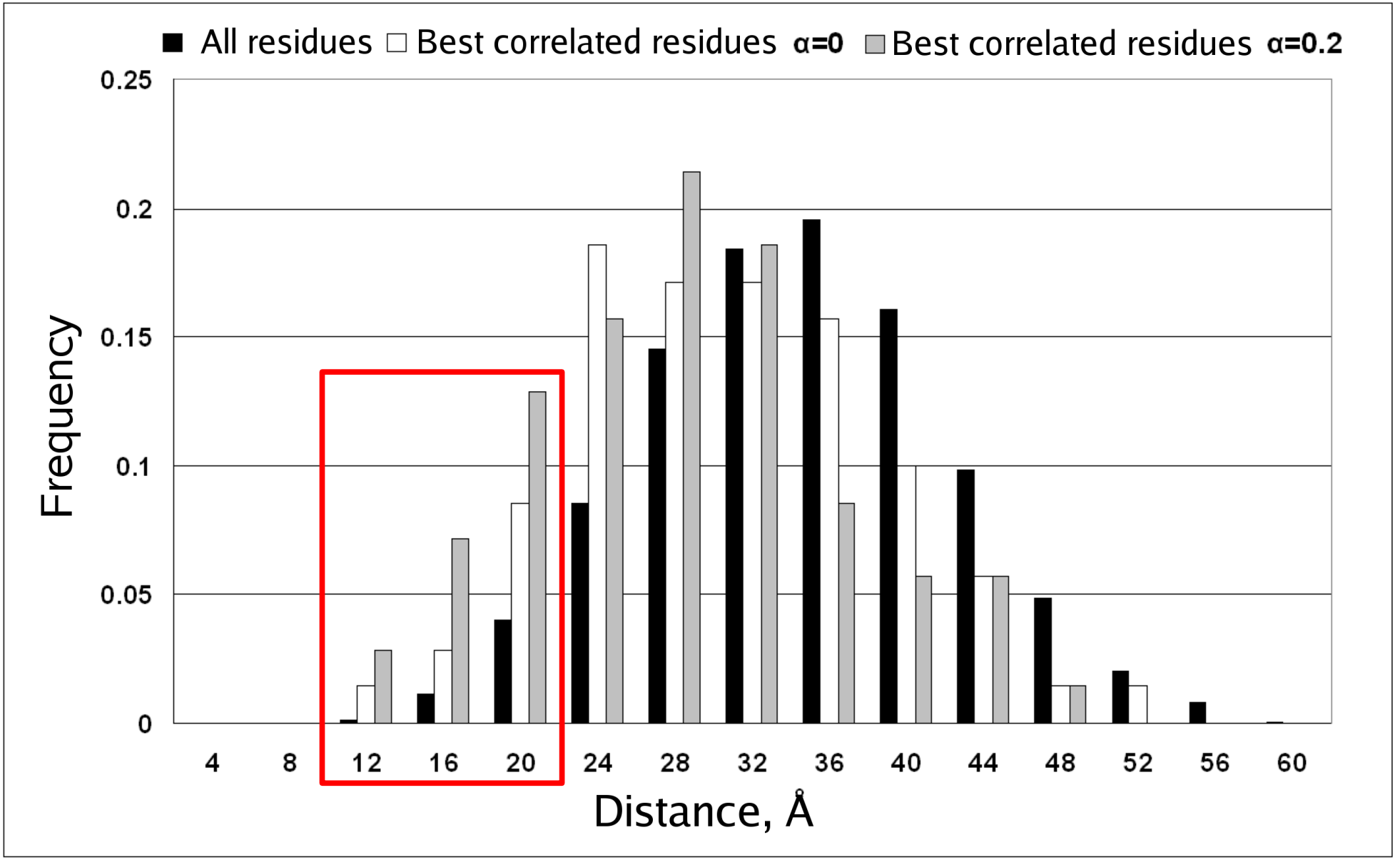
10 PFAM families domain pairs



EXAMPLE: SH3-SH2



EXAMPLE: SH3-SH2

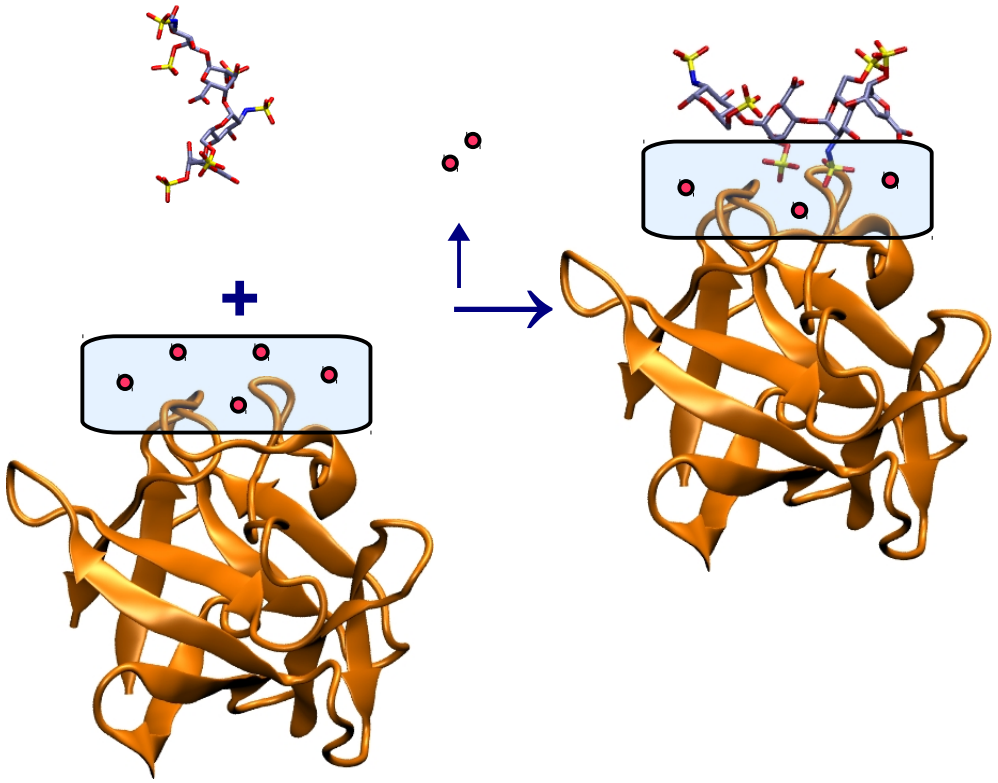


CASE STUDY I: CONCLUSIONS

- All interfacial residue types are quantitatively comparable in terms of their contribution to the energy of complex formation.
- Interfacial water contributes to the conservation of protein-protein interactions and has higher residence time than water at surfaces.
- The introduction of the WET similarity matrix into the concept of correlated mutations significantly improves protein contacts prediction.

CASE STUDY II: SOLVENT ROLE FOR GAG DOCKING

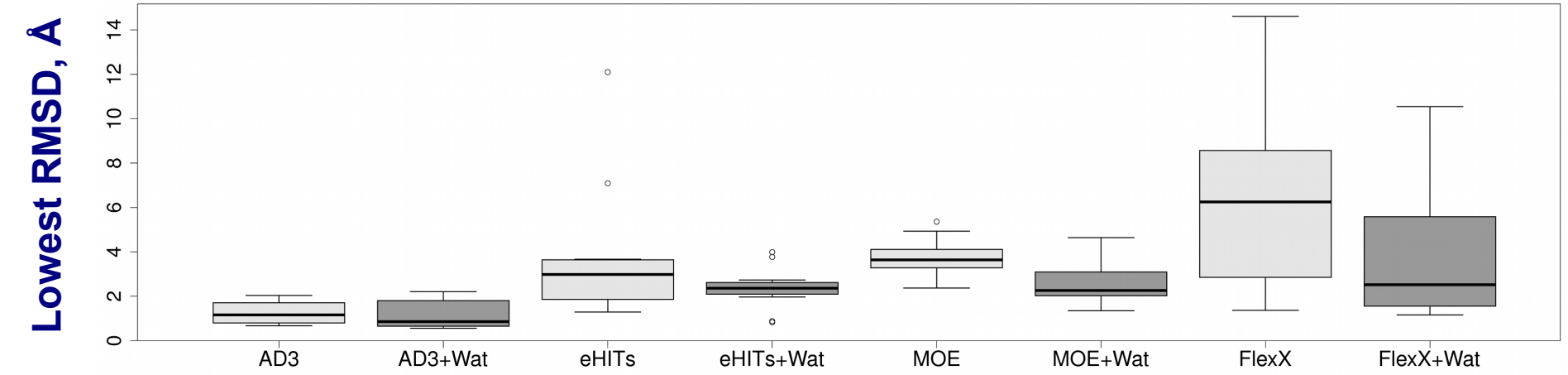
- **Solvent role:**
  - Bridging water molecules
  - Displaced water molecules

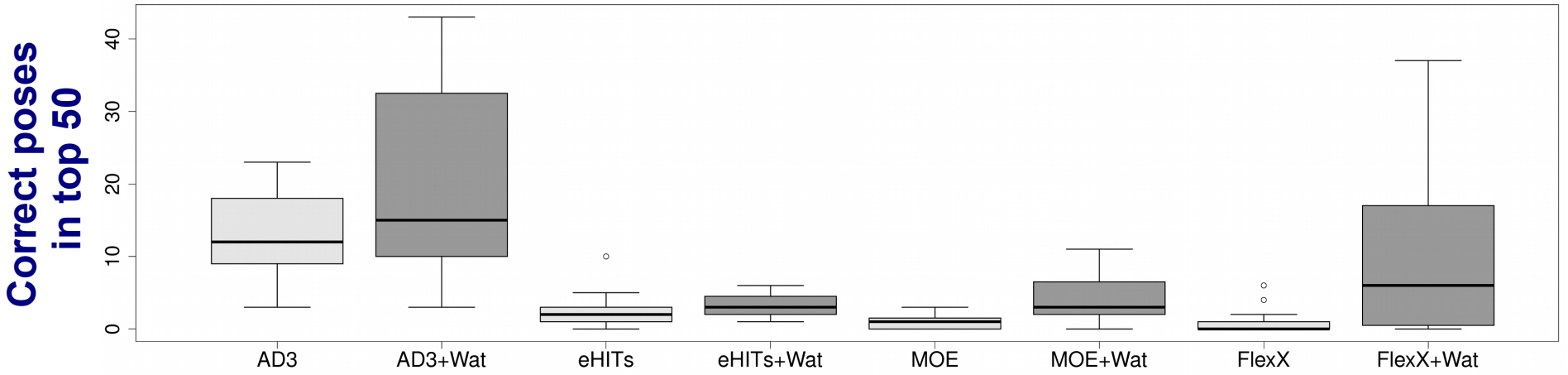


- **Objectives:**
  - To place solvent into the binding site *de novo*
  - To study how much solvent inclusion can improve docking

GAG DOCKING WITHOUT AND WITH EXPLICIT SOLVENT

11 GAG-protein complexes, 4 docking methods





Addition of explicit solvent can significantly improve docking results

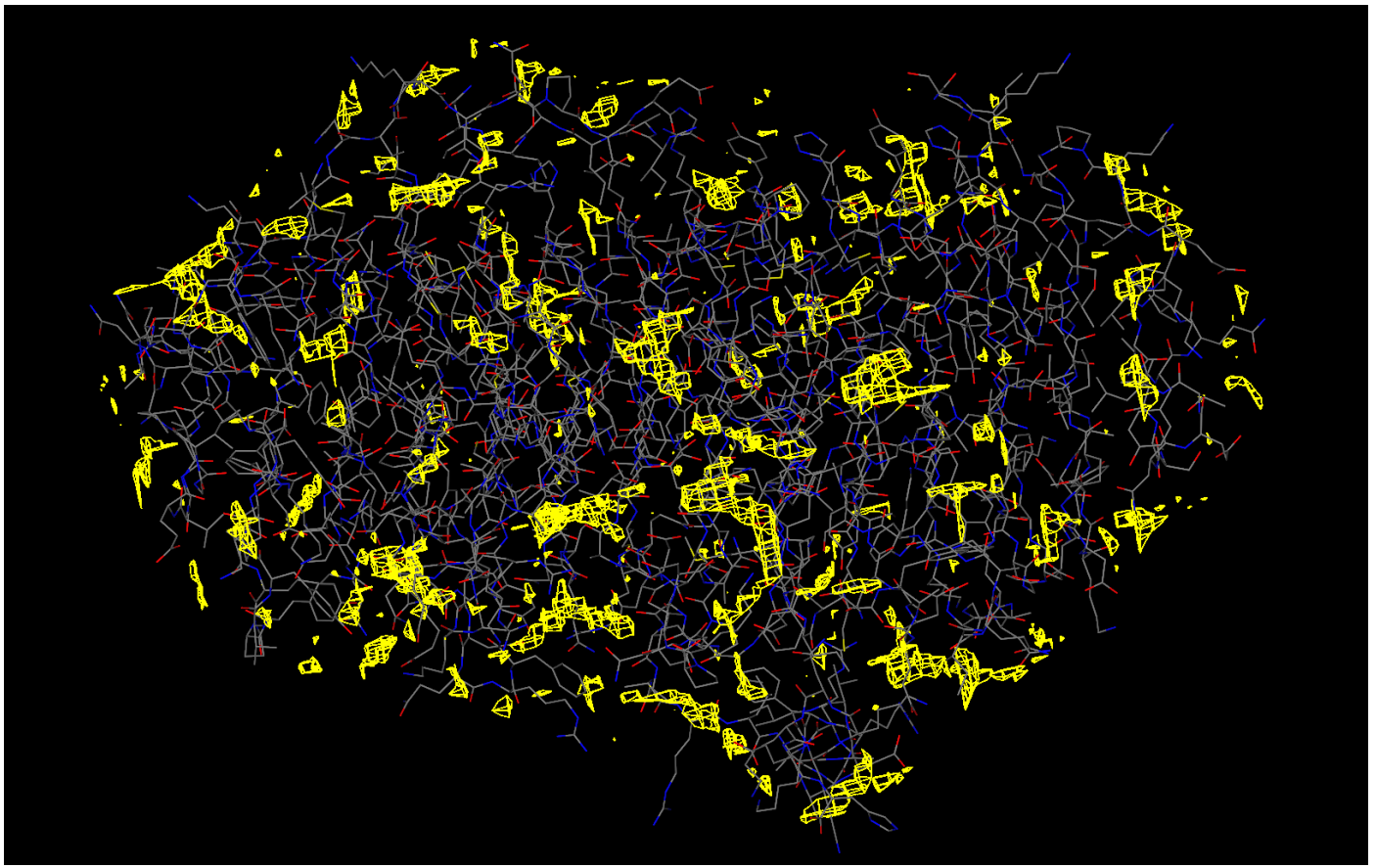
PROBE-BASED MAPPING OF PROTEIN INTERACTIONS

GRID: determines energetically favourable positions for chemical probes in proteins

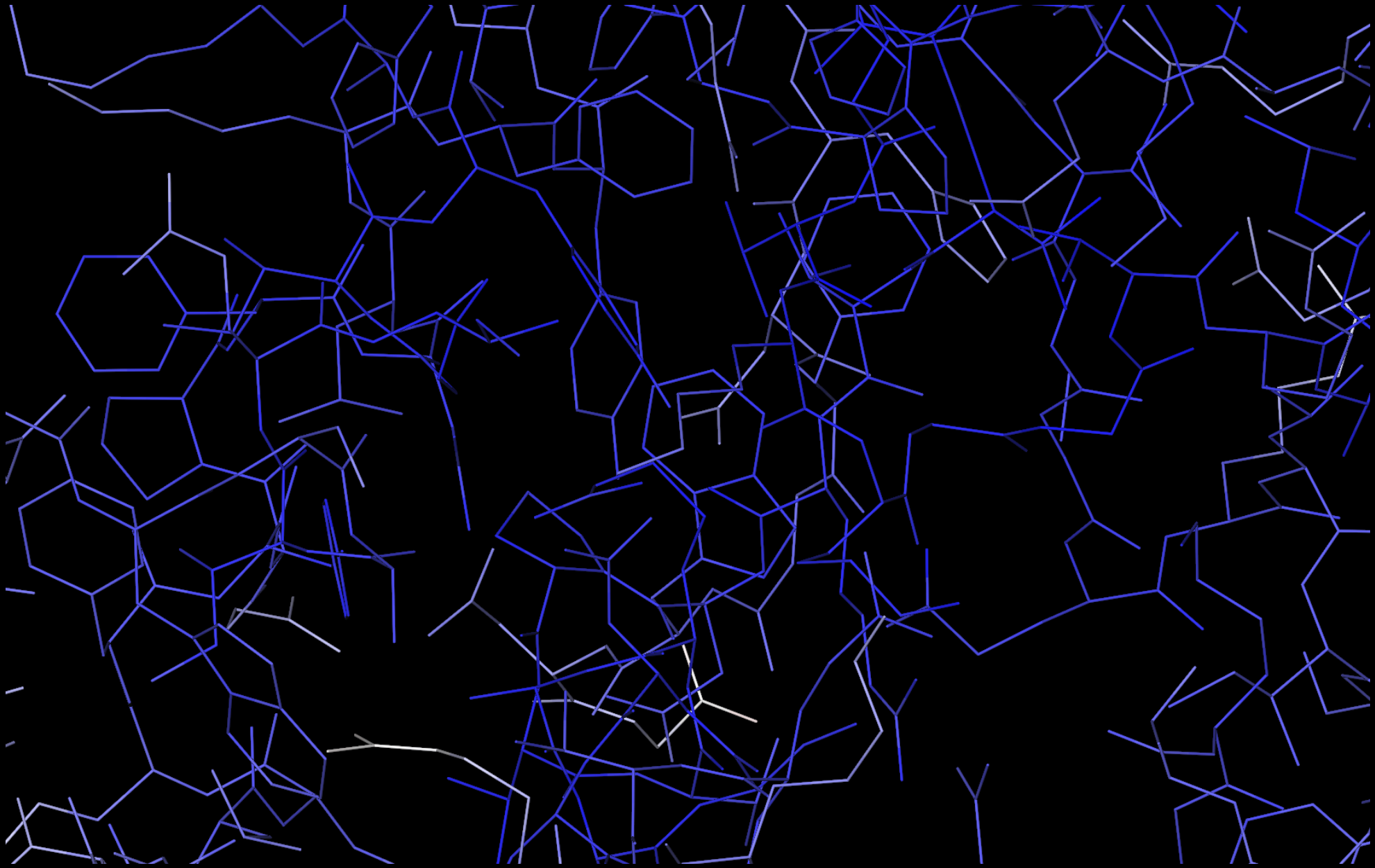
$$E(x_i,y_i,z_i)=E_{el}(x_i,y_i,z_i) + E_{vdw}(x_i,y_i,z_i) + E_{hb}(x_i,y_i,z_i)$$

Probes:

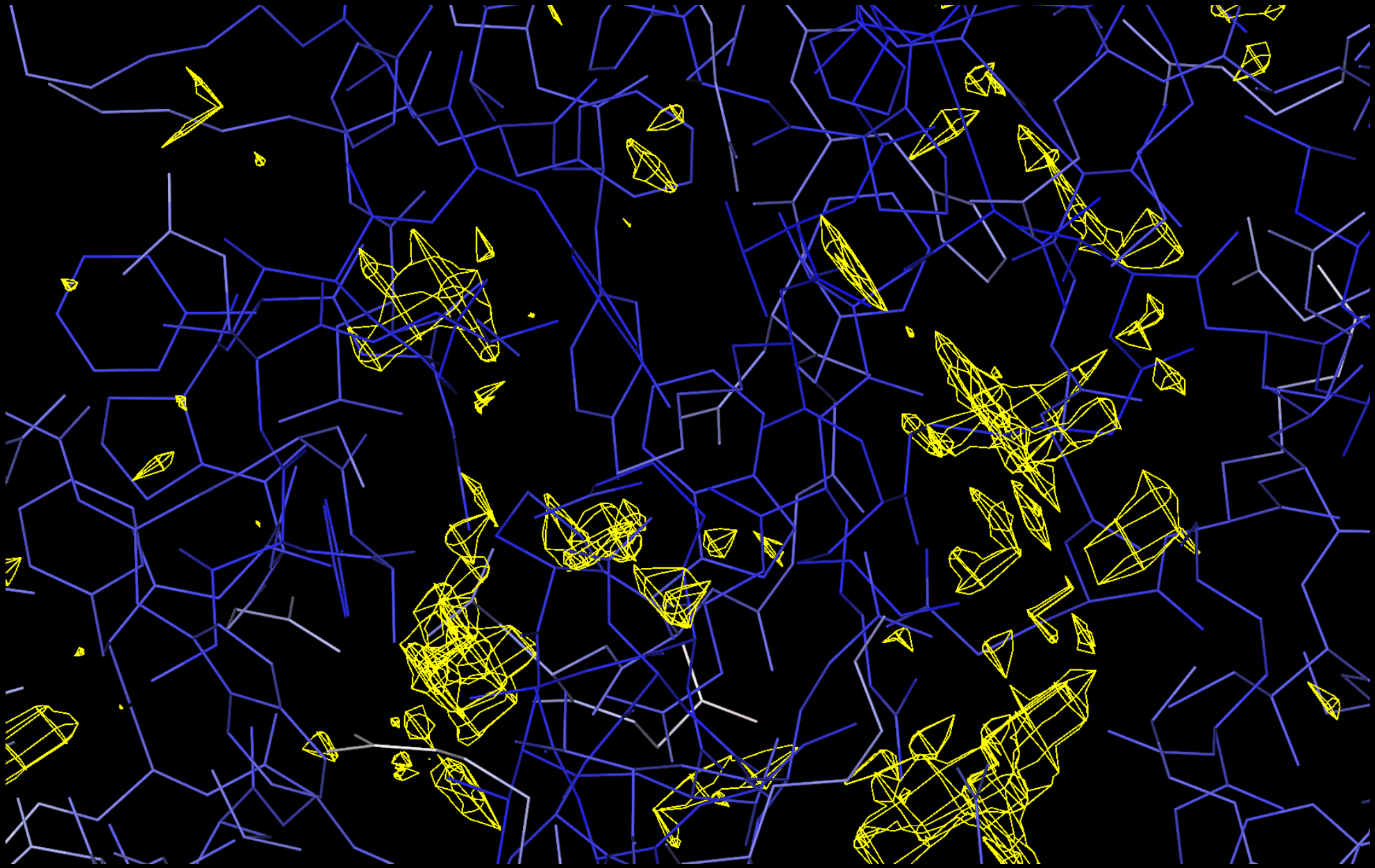
- H_2O
- OH
- COO^-
- $C_{sp3,sp2,sp}$
- –S=O
- others …



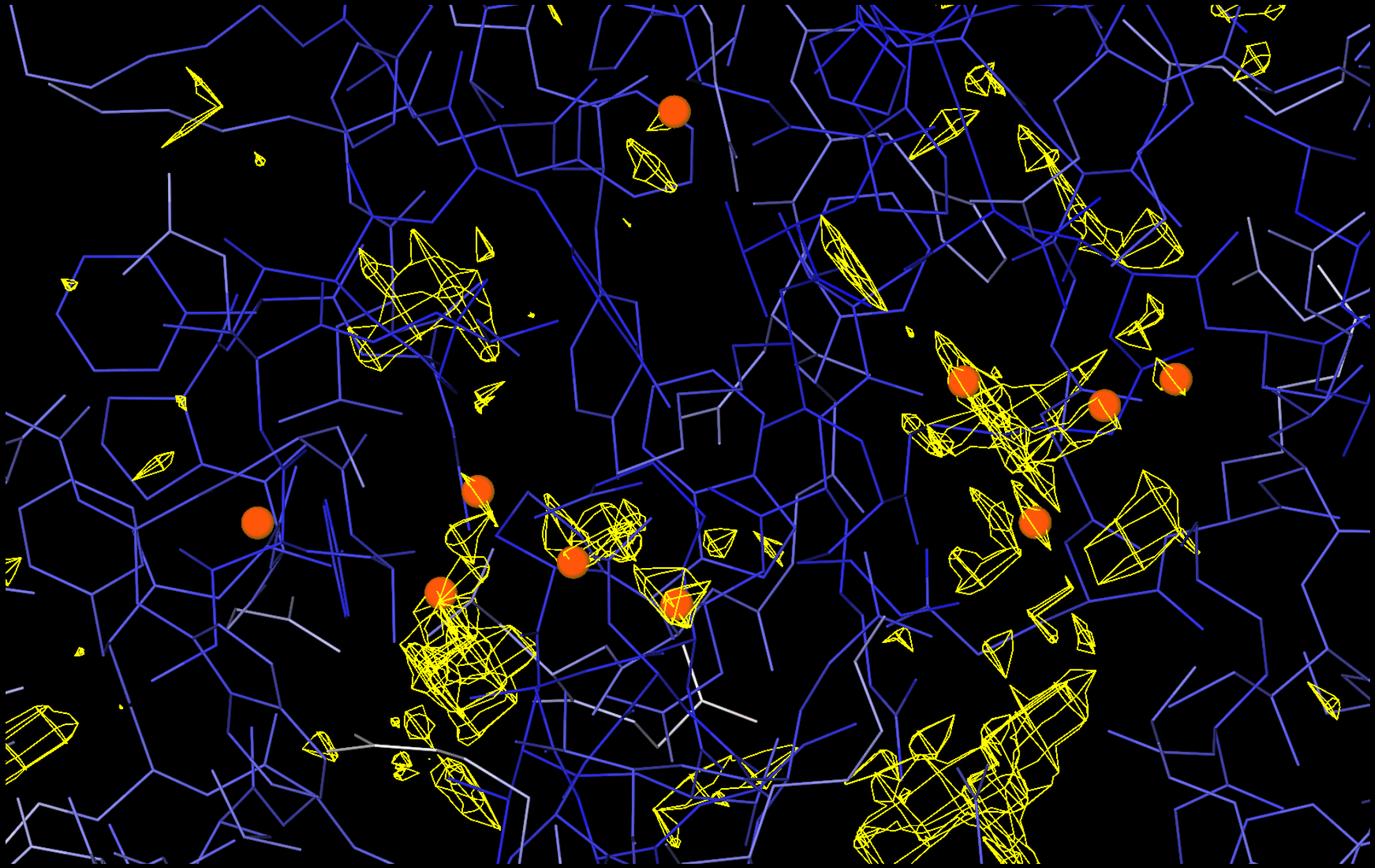
I. BINDING SITE MINIMIZATION



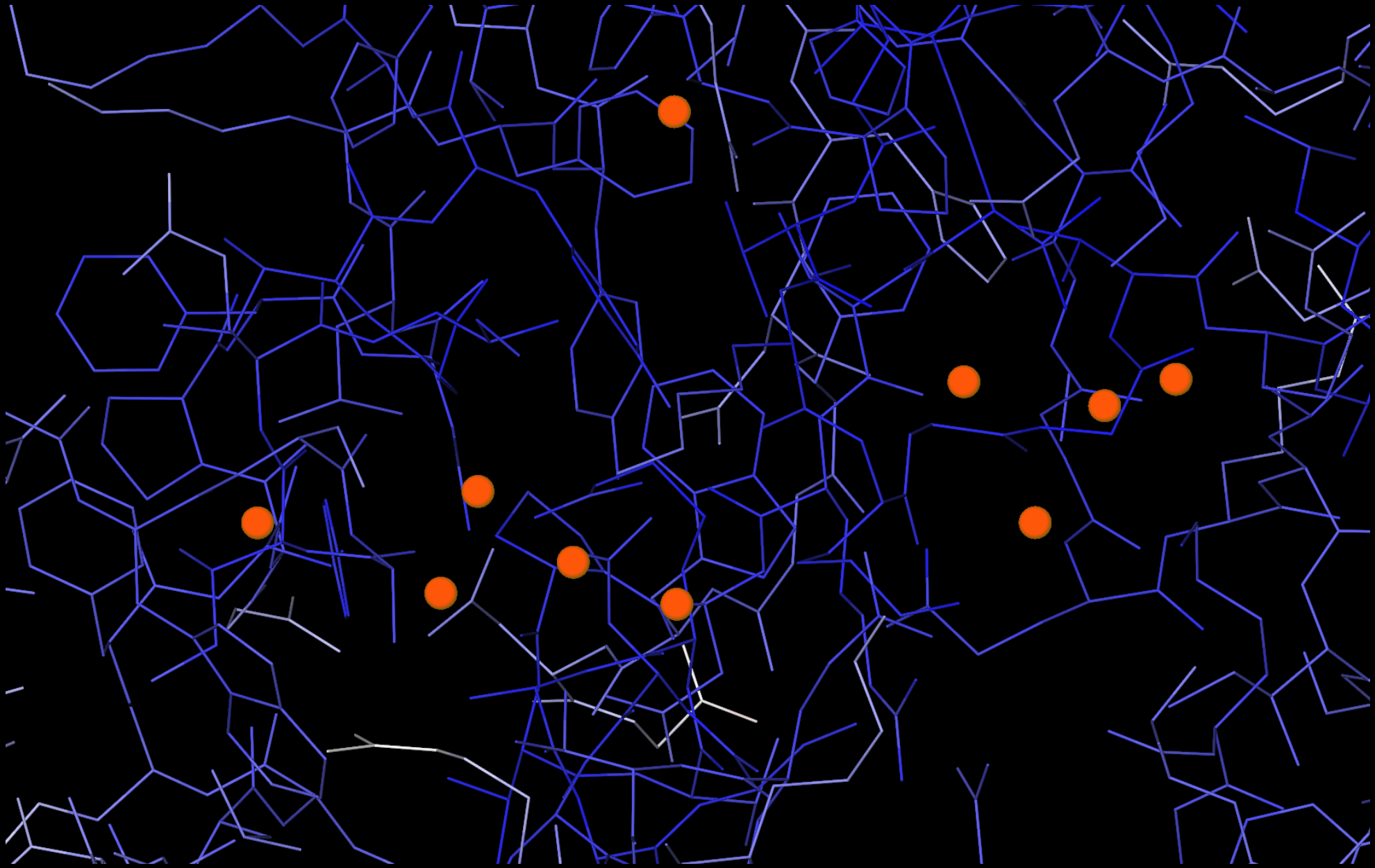
II. PREDICTION OF SOLVENT POSITIONS



II. PREDICTION OF SOLVENT POSITIONS

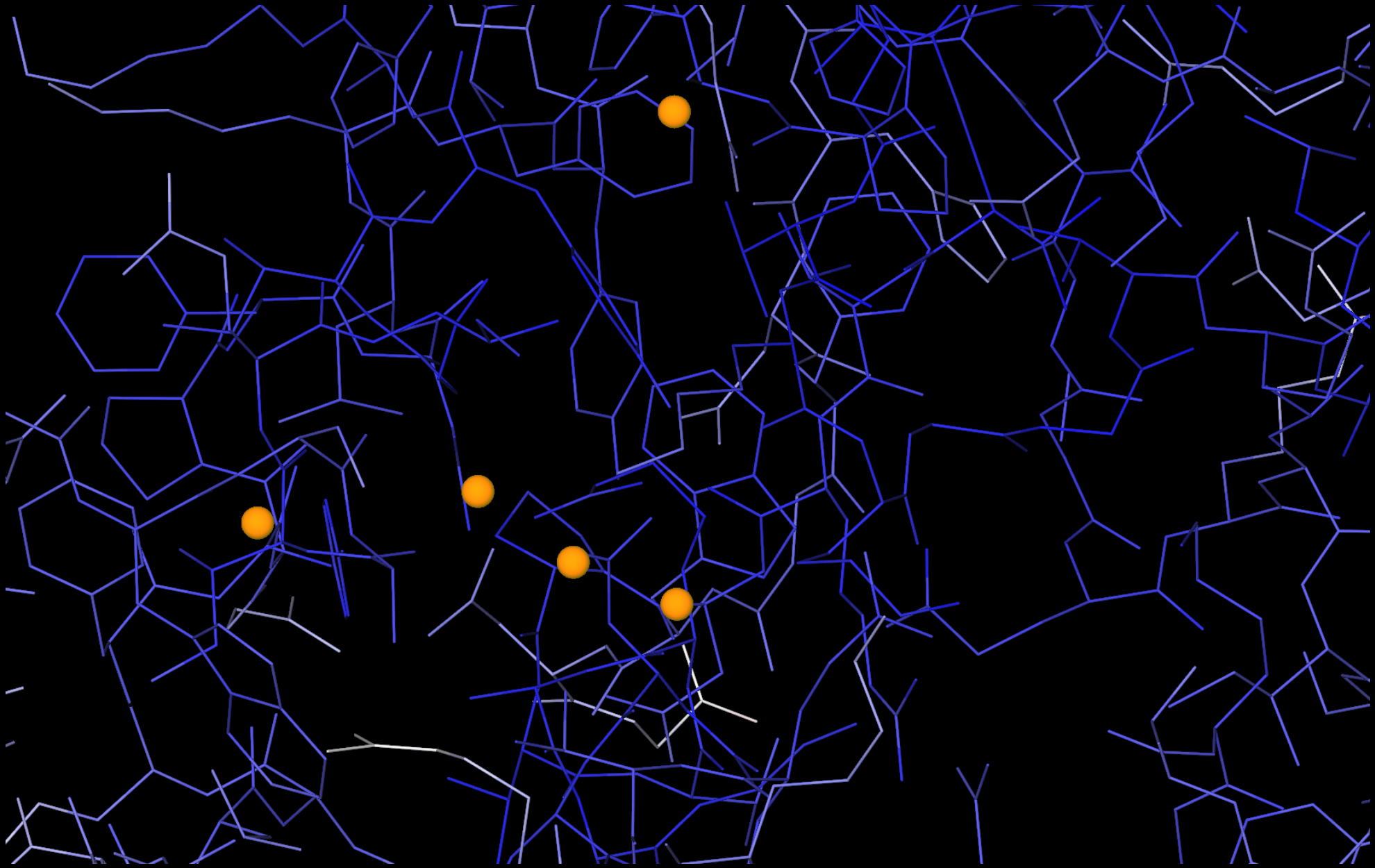


II. PREDICTION OF SOLVENT POSITIONS



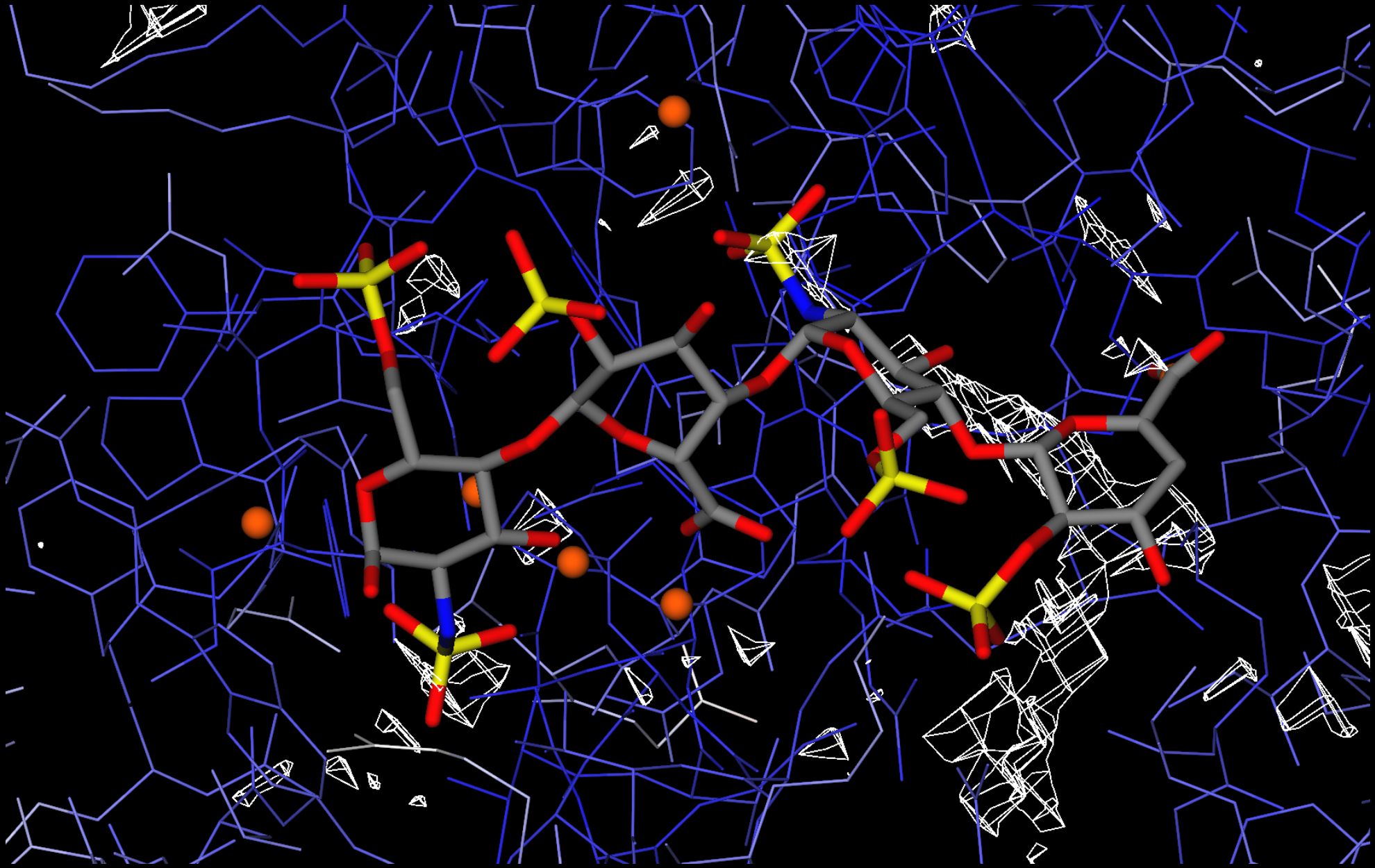
III. PREDICTION OF DISPLACED SOLVENT

III. PREDICTION OF DISPLACED SOLVENT



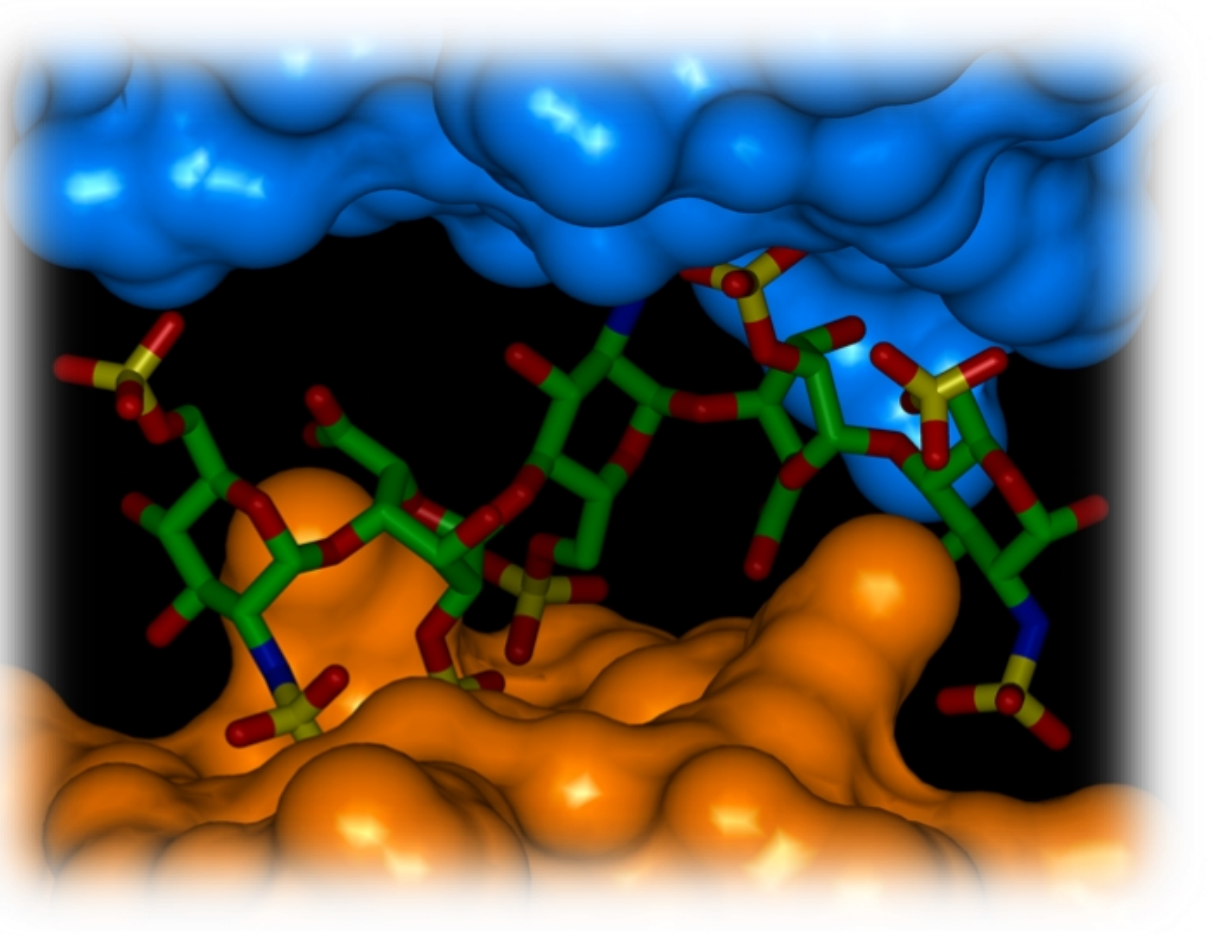
PROOF OF CONCEPT

T8U (1.95 Å):
sulfotransferase+HS (tetra)



SUMMARY

- We *de novo* predict explicit solvent positions in the binding site
- Docking results are improved when explicit solvent is used
- Novel docking approaches are needed to take solvent into account



LECTURE 3: MODELLING SOLVENT

- Water unique properties
- Water and biomolecular systems
- Implicit solvent and Poisson-Boltzmann methodology
- Explicit solvent models
- Grid Inhomogeneous Solvation Theory (GIST)
- Solvent challenge in docking
- Water in protein-protein interfaces
- Case studies:
  - MD study of the role of water in protein-protein interfaces
  - Introduction of solvent information for protein contacts prediction
  - Inclusion of water in GAGs docking to proteins