

The modulation of hydrophobicity in protein-ligand interactions

L. Ridgway Scott

The Institute for Biophysical Dynamics, the Computation Institute, and the Departments of Computer Science and Mathematics, The University of Chicago

This talk is based on joint work with Ariel Fernandez (Rice Univ.), Harold Scheraga (Cornell), and Kristina Rogale Plazonic (Princeton); and at U. Chicago: Steve Berry, John Goldsmith and Jing Liu.

1 Protein basics

Proteins are sequences of amino acids which are covalently bonded along a “backbone.”

Proteins of biological significance fold into a three-dimensional structure by adding hydrogen bonds between carbonyl and amide groups on the backbone of different amino acids.

In addition, other bonds, such as a salt bridge or a disulfide bond can form between particular amino acids (Cysteine has sulfur atoms in its sidechain).

However, the hydrogen bond is the primary mode of structure formation in proteins.

1.1 Chains of amino acid residues

Proteins are chains of amino acid residues whose basic unit is the peptide group.

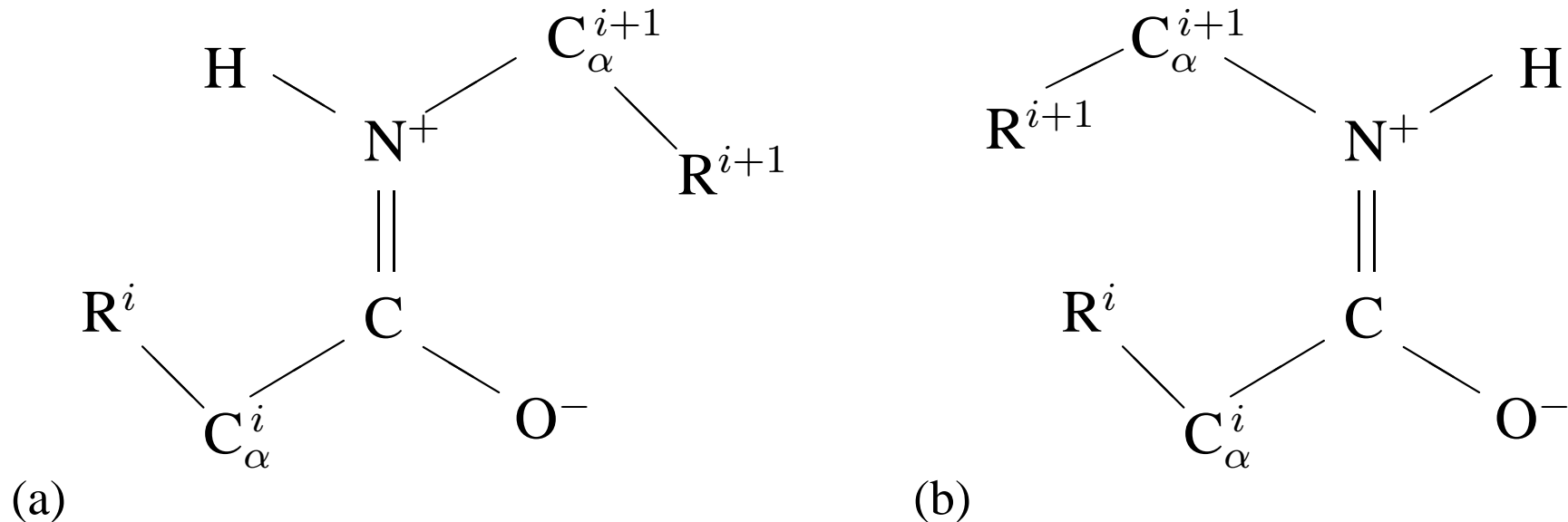


Figure 1: The rigid state of the peptide bond: (a) trans form, (b) cis form. The double bond between the central carbon and nitrogen keeps the peptide bond planar.

1.2 Linear (primary) structure of proteins

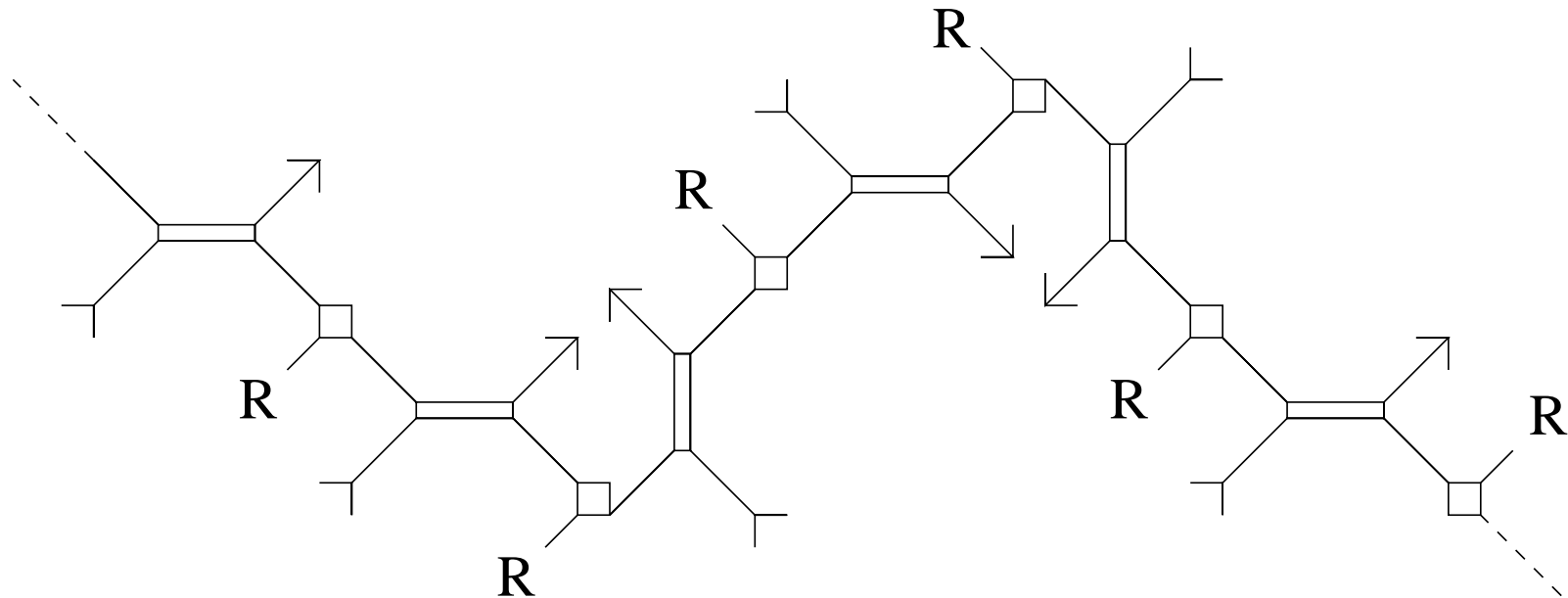
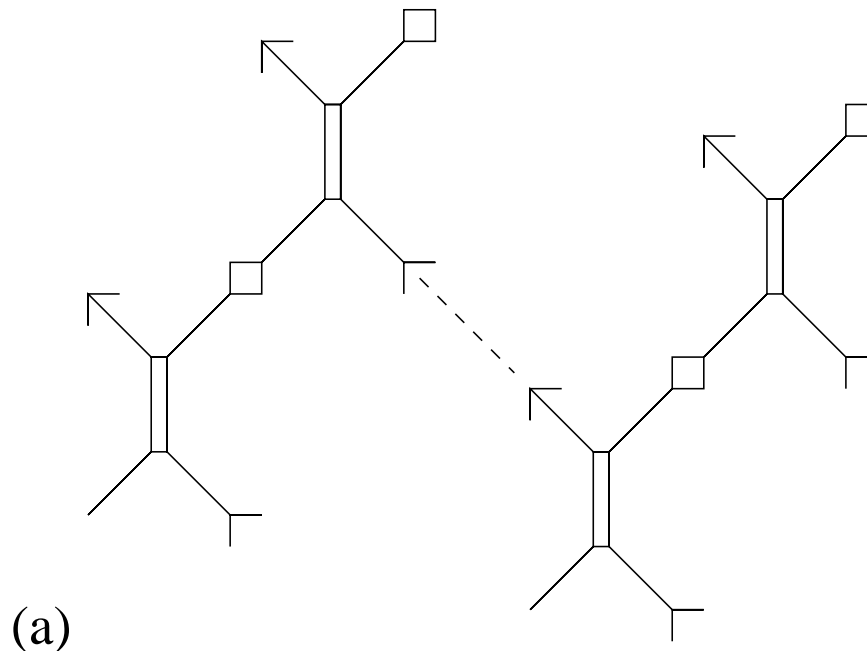


Figure 2: Cartoon of peptide sequence where all peptides are in trans form (cf. Figure 1). Small boxes represent C-alpha carbons, arrow heads represent amide groups NH, arrow tails represent carbonyl groups CO, and thin rectangular boxes are double bond between backbone C and N. The different residues are indicated by R's. The numbering scheme is increasing from left to right, so that the arrow formed by the carbonyl-amide pair points in the direction of increasing residue number. The three-dimensional nature of the protein is left to the imagination.

1.3 Hydrogen bonds and secondary structure

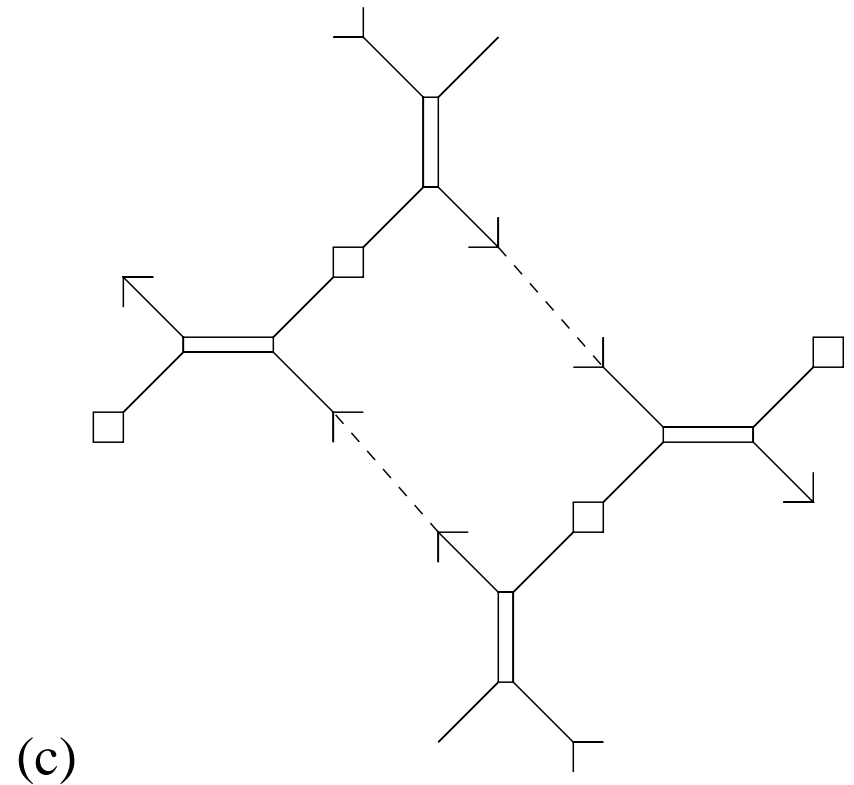
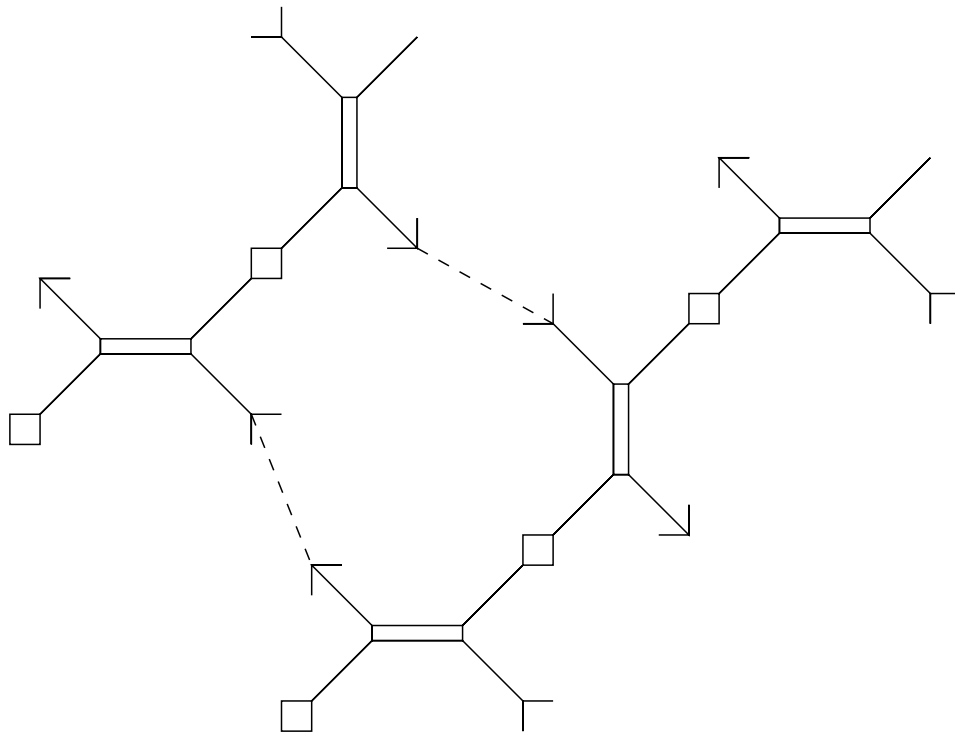
Proteins have a hierarchy of structure, the next being **secondary structure** consisting of two primary types: alpha-helices and beta-sheets (a.k.a., α -helices and β -sheets).

Alpha helices are helical arrangements of the subsequent peptide complexes with a distinctive hydrogen bond arrangement between the amide (NH) and carbonyl (OC) groups in peptides separated by k steps in the sequence, where primarily $k = 4$ but with $k = 3$ and $k = 5$ also occurring less frequently:



1.4 Beta sheets

Beta sheets represent different hydrogen bond arrangements: (b) is the anti-parallel arrangement and (c) is the parallel.



Both structures are essentially flat, in contrast to the helical structure in (a).

Proteins as digital components

Proteins are the essential components of life:

- used to build complexes, e.g., viruses (bricks and mortar)
- involved in signalling (information transmission)
- enzymes essential in catalysis (chemical machines)

In all these cases, protein-ligand interaction is essential.

These interactions are deterministic (always the same).

Proteins function as discrete components not as analog devices.

The hydrophobic effect

Hydrophobic effect crucial in protein-ligand association.

Water is essential to life as we know it, but hostile to proteins.

The role of water in protein biophysics: to modulate electric forces via the dielectric effect.

Hydrophobicity fosters water removal and supports protein-ligand interaction, but it also modulates the dielectric effect.

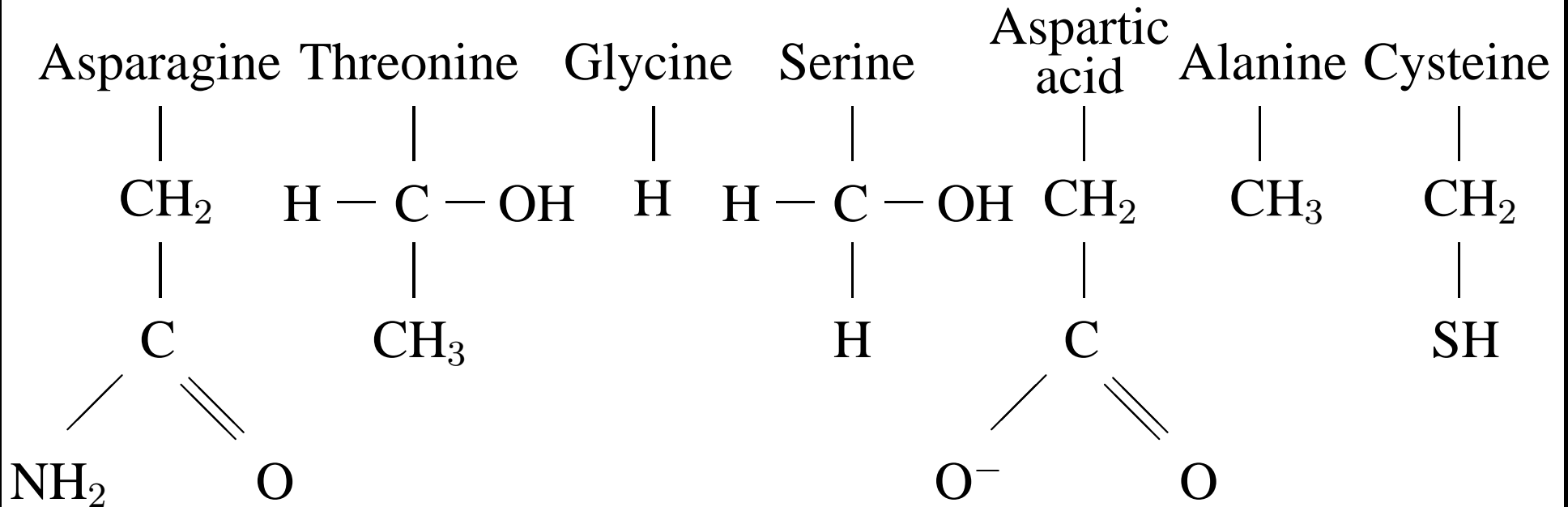
Water is a strong dielectric, and protein sidechains are a complex mix of charged, polar, and hydrophobic parts.

But the hydrophobic effect is non-specific in action.

What makes proteins interact in a repeatable way?

What sidechains are found at interfaces?

By examining interfaces in PDB structures, we can see which residues are most likely to be found at interfaces.



Sidechains most likely to be involved in interactions, ordered from the left (asparagine), are not hydrophobic.

Electronic forces

The only force of significance in biochemistry is the electric force.

— But often modulated by indirection or induction.

In terrestrial biology, water plays a significant role as a dielectric which mediates non-covalent interactions (hydrogen bonds, salt bridges, cation-pi interactions).

But the dielectric effect of water is modulated by hydrophobic components of proteins.

Moreover, a ligand can change the hydrophobic environment upon binding.

In protein-ligand interactions, this makes intramolecular bonds as important as intermolecular interactions.

Our technology

Interaction between physical chemistry and data mining in biophysical data bases.

- Data mining can lead to new **results in physical chemistry** that are significant in biology.
- Using physical chemistry to look at data **provides insights regarding function.**

We review some recent results regarding protein-ligand interaction that are based on novel insights about hydrophobic effects.

We show that sidechain configurations modulate dielectric effect.

We discuss how these can be used to understand a novel factor that supports protein-ligand binding.

A quote

from Nature's Robots

The exact and definite determination of life phenomena which are common to plants and animals is only one side of the physiological problem of today. The other side is the **construction of a mental picture of the constitution of living matter** from these general qualities. In this portion of our work **we need the aid of physical chemistry.**

Jacques Loeb, The biological problems of today: physiology. Science 7, 154-156 (1897).

so our theme is not so new

Data mining definition

WHATIS.COM: Data mining is sorting through data to identify patterns and establish relationships.

Data mining parameters include:

- Association - **looking for** patterns where one event is connected to another event
- Sequence or path analysis - **looking for** patterns where one event leads to another later event
- Classification - **looking for** new patterns (May result in a change in the way the data is organized but that's ok)
- Clustering - finding and **visually documenting** groups of facts not previously known

Conclusion: Data mining involves looking at data.

Data mining lens

If data mining is **looking at data** then

What type of lens do we use?

- All of these have **chemical representations**, e.g.,



- **Alphabetic sequences** describe much of biology: DNA, RNA, proteins.
- All of these have **three-dimensional structure**.
- But structure alone does not explain **how they function**.

Physical chemistry clarifies the picture and allows function to be more easily interpreted.

Sequences can tell a story

Protein sequences

aardvarkateatavisticallyacademicianaccelerative
acetylglycineachievementacidimetricallyacridity
actressadamantadhesivenessadministrativelyadmit
afflictiveafterdinneragrypniaaimlessnessairlift

and DNA sequences

actcatatactagagtacttagacttatactagagcattacttagat

can be studied using automatically determined lexicons.

Joint work with John Goldsmith, Terry Clark, Jing Liu.

Sequences can tell a story (a linguistic lens)

Protein sequences

aardvarkateatavisticallyacademicianaccelerative
acetylglycineachievementacidimetricallyacridity
actressadamantadhesivenessadministrativelyadmit
afflictiveafterdinneragrypniaaimlessnessairlift

and DNA sequences

actcatatactagagtacttagacttatactagagcattacttagat

can be studied using **automatically determined lexicons.**

Joint work with John Goldsmith, Terry Clark, Jing Liu. **But that is another talk**

Talk Outline

Physical chemistry provides **powerful lens** to look at protein data

- Tutorial on hydrophobic wrapping
 - hydrophobic protection (desolvation) of hydrogen bonds
 - new motif: **dehydron=insufficiently desolvated hydrogen bond**
 - dehydrons are involved in protein interaction (**they are sticky**)
- Using dehydrons in bioinformatics
 - the tails of the distribution: **extreme stickiness**
 - number of dehydrons correlates with protein interactivity
 - number of dehydrons differentiates proteins with similar structure
- Using wrapping technology in drug design
- Requires more precise understanding of dielectrics
 - Review of dielectrics
 - Poisson-Debye equation

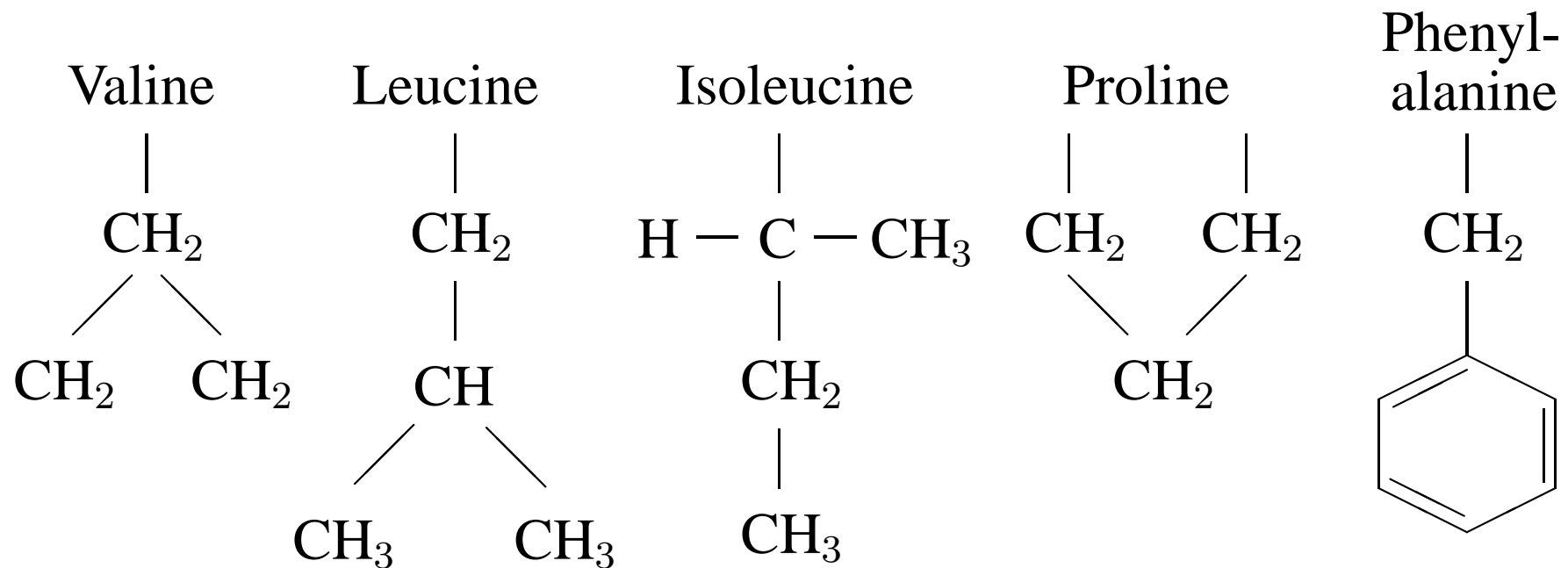
2 Tutorial on hydrophobic wrapping

Effect of modulation of dielectric by hydrophobic groups.

- Amino acid side chains have different properties
- Tutorials on
 - hydrophobicity: carbonaceous groups
 - dielectrics: water screens charges
- Extent of wrapping changes nature of hydrogen bond
- Dehydrons: Under-wrapped hydrogen bonds
 - Antibody binding: dehydrons can guide the way
 - Virus capsid: a model for protein-protein interaction
 - Stickiness of dehydrons

2.1 Amino acid side chains have different properties

Carbonaceous groups on certain side chains are hydrophobic:



Amino acids (side chains only shown) with carbonaceous groups.

2.2 Tutorial on hydrophobicity

Carbonaceous groups (CH, CH₂, CH₃) are hydrophobic because

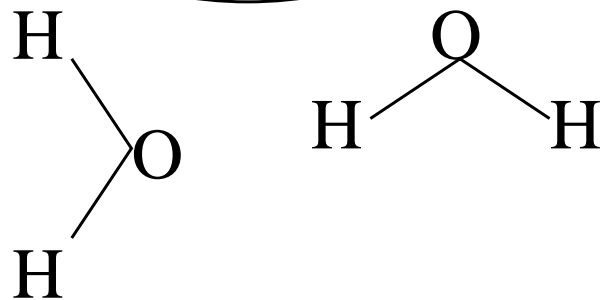
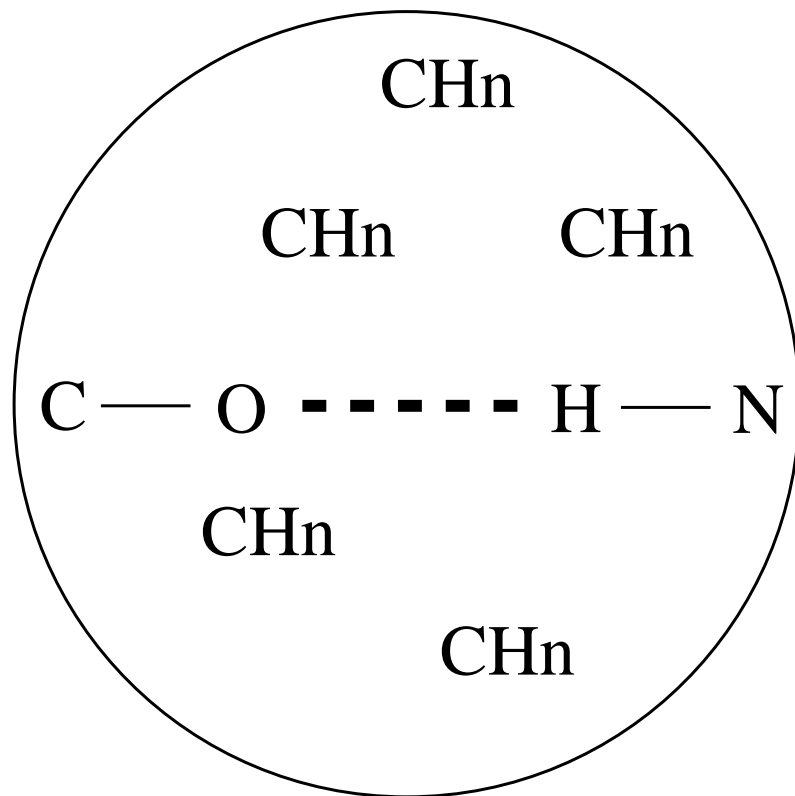
- they are non-polar and thus do not attract water strongly (enthalpic barrier)
- they restrict the mobility of water (entropic barrier)

2.3 Tutorial on dielectrics

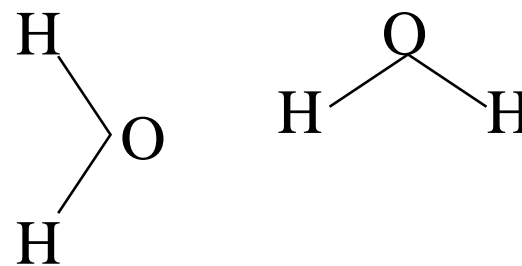
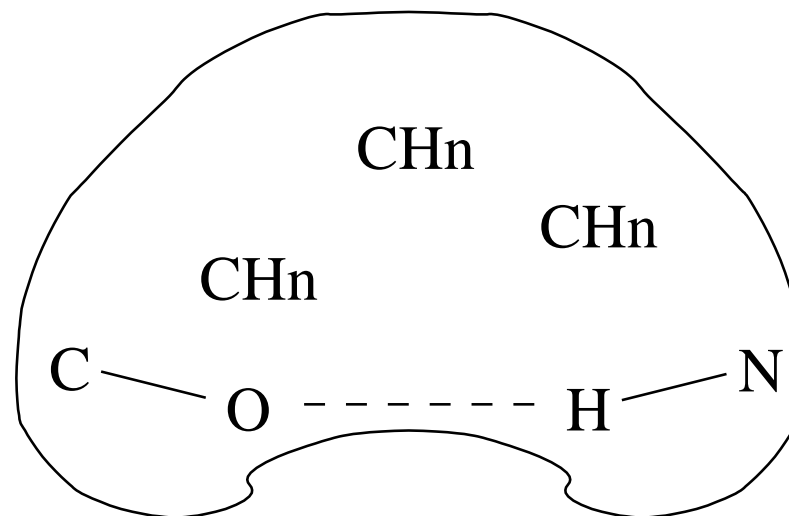
Water removal reduces the dielectric effect and makes electronic bonds stronger.

Number of carbonaceous groups in a region determine extent of water removal and strength of electronic bonds.

2.4 Wrapping protects hydrogen bond from water



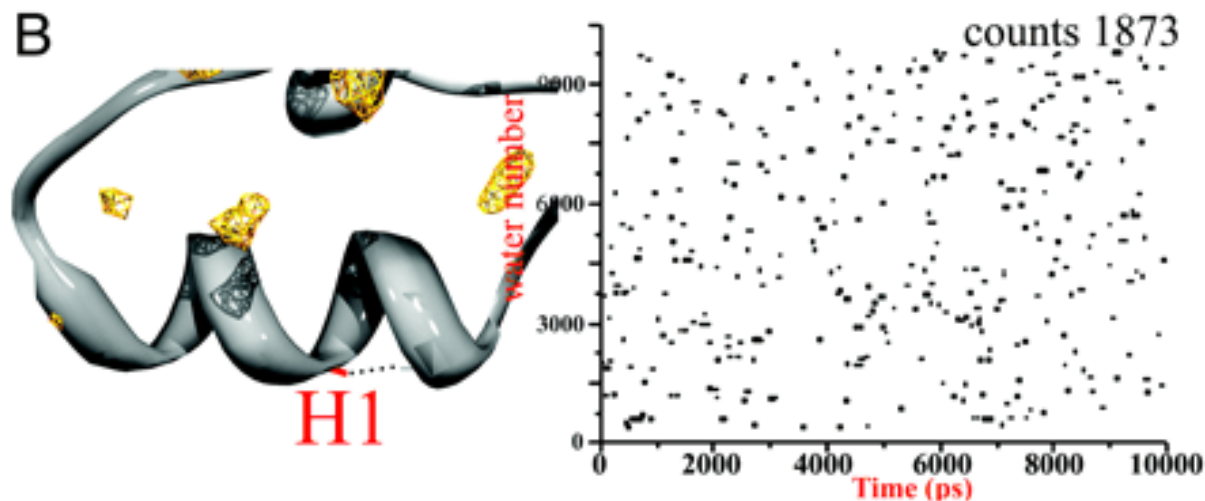
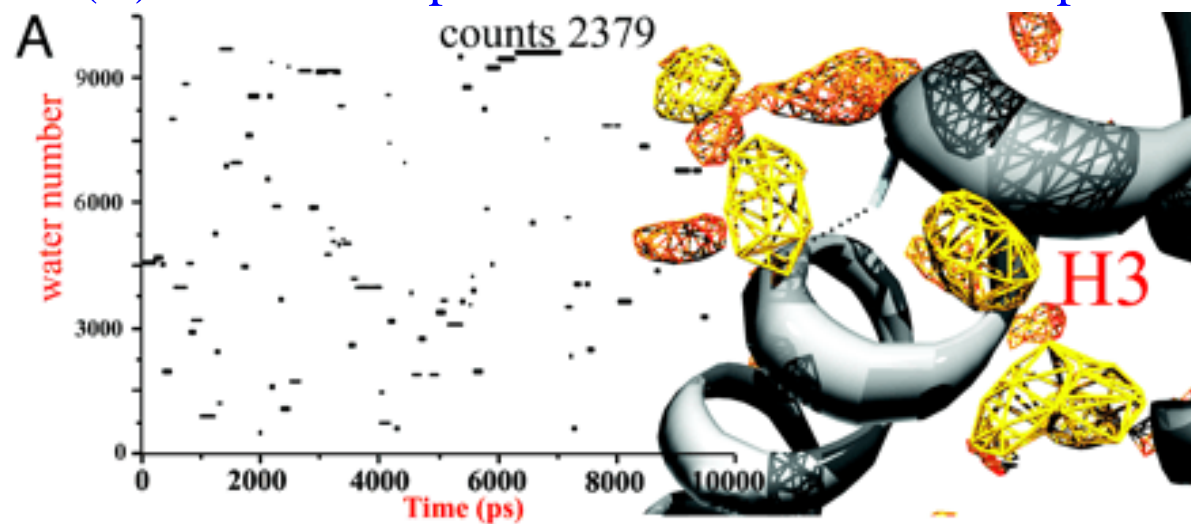
Well wrapped hydrogen bond



Underwrapped hydrogen bond

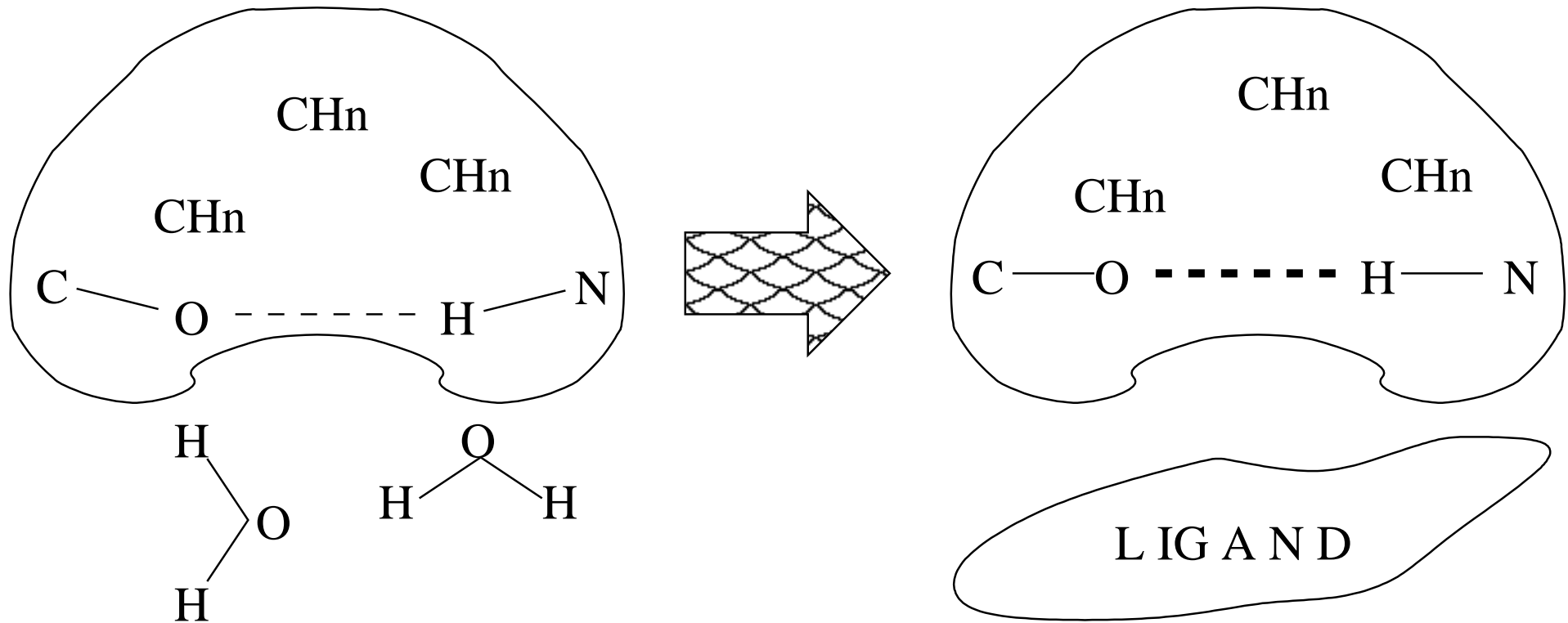
2.5 Extent of wrapping changes nature of hydrogen bond

Hydrogen bonds (B) that are not protected from water do not persist.



From De Simone, et al., PNAS 102 no 21 7535-7540 (2005)

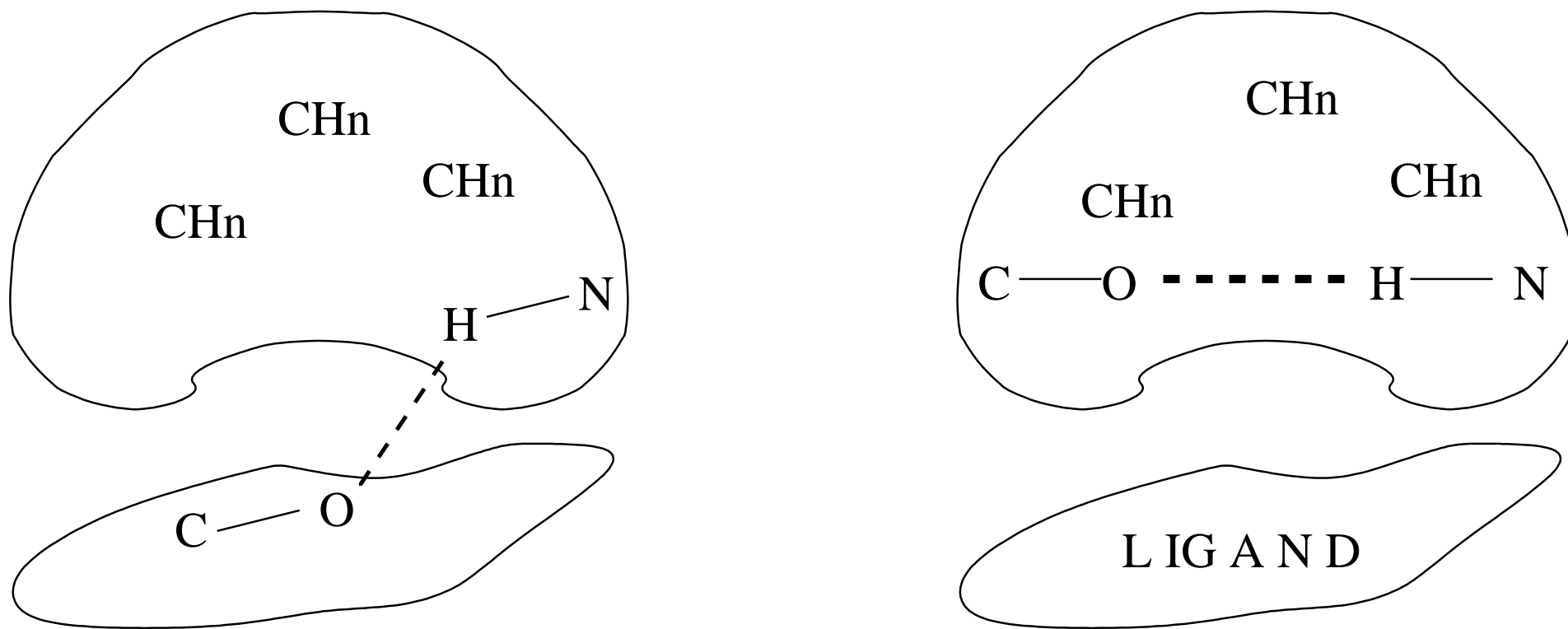
2.6 Ligand binding removes water



Binding of ligand changes underprotected hydrogen bond (high dielectric) to strong bond (low dielectric)

No intermolecular bonds needed!

2.7 Intermolecular versus intramolecular



Intermolecular versus intramolecular hydrogen bonds.

Energetic contribution to binding comparable.

Desolvation cost for intermolecular bond may be higher.

Intermolecular bonds are like the power cord on my computer.

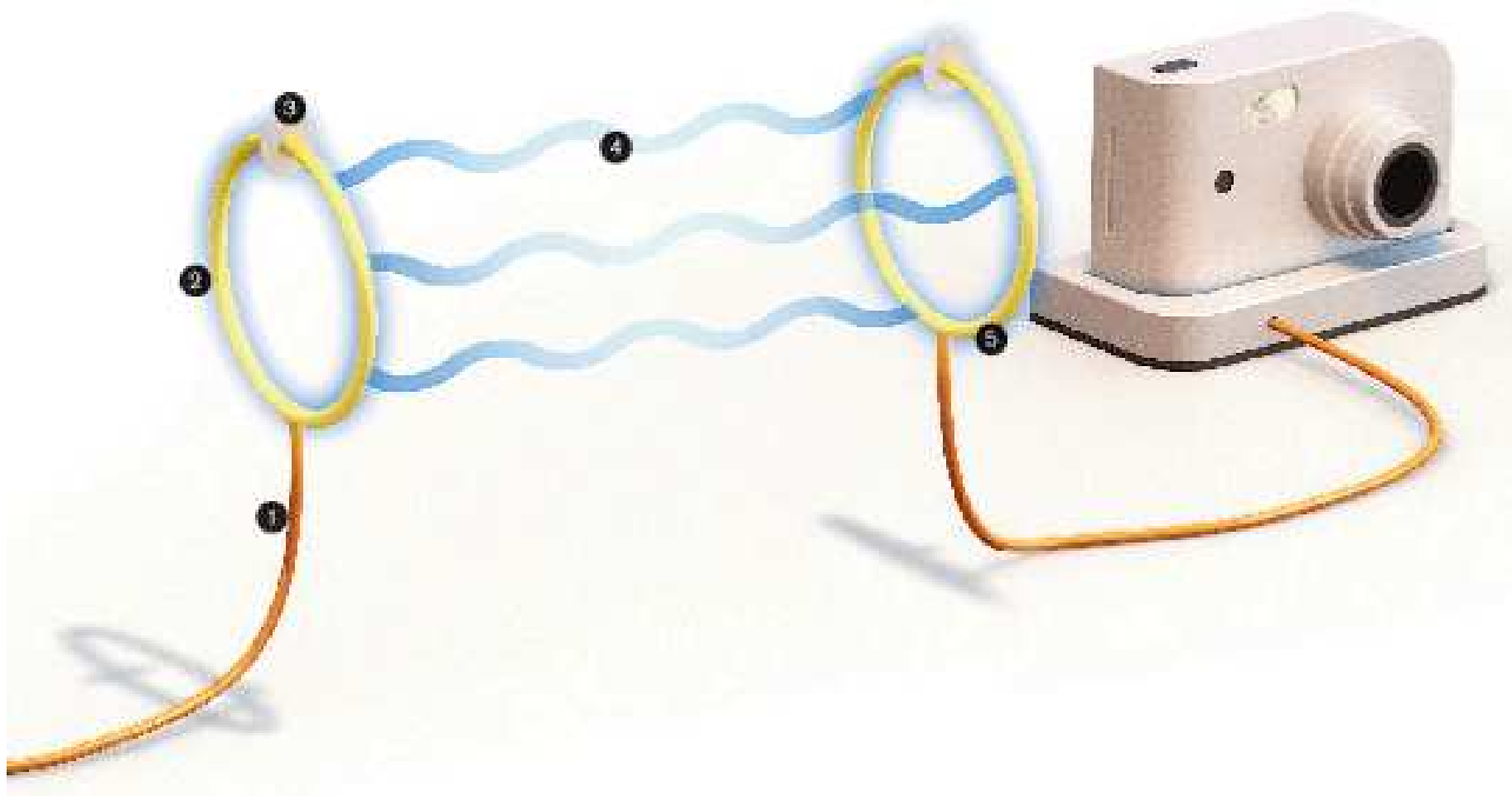
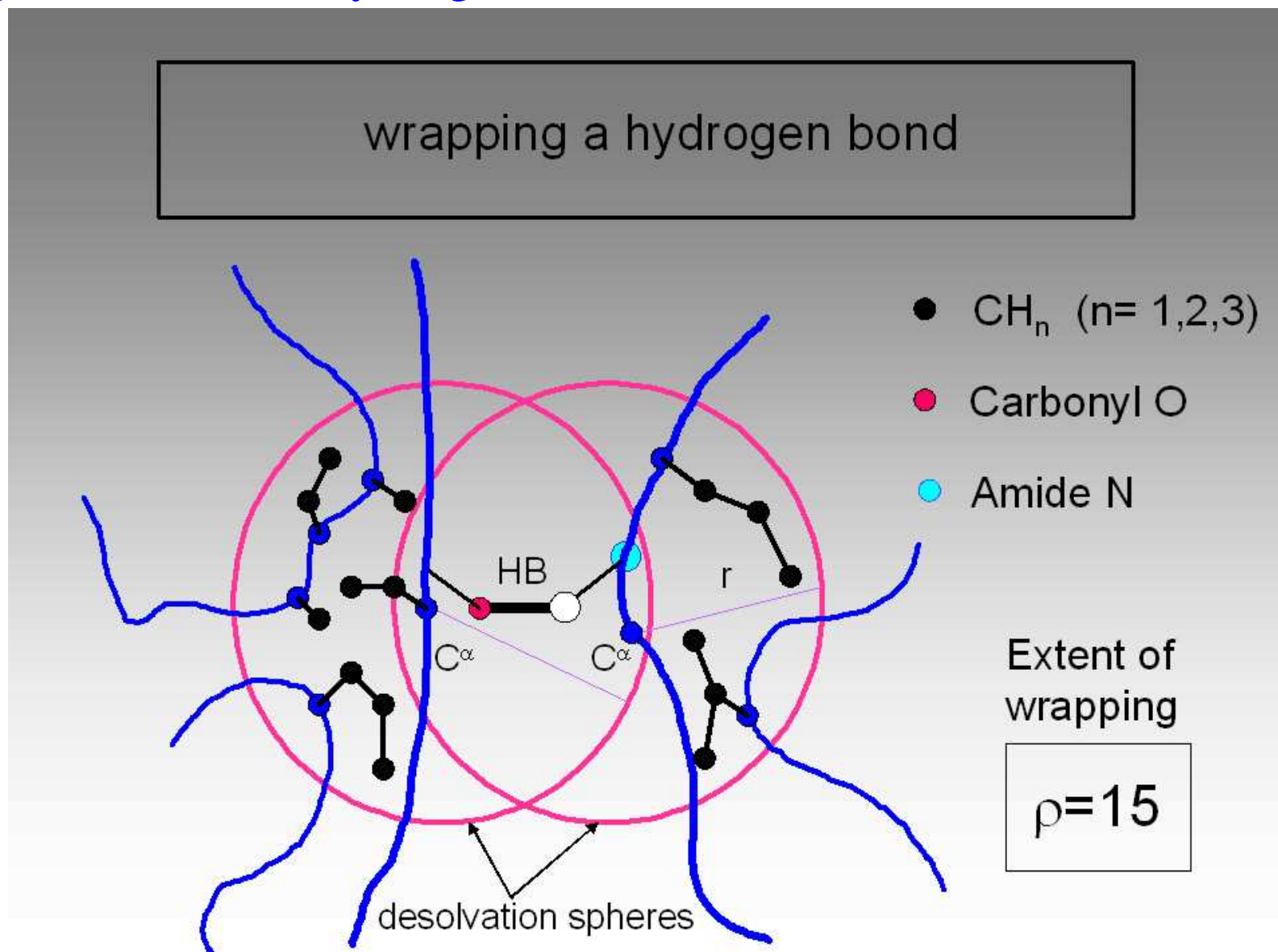


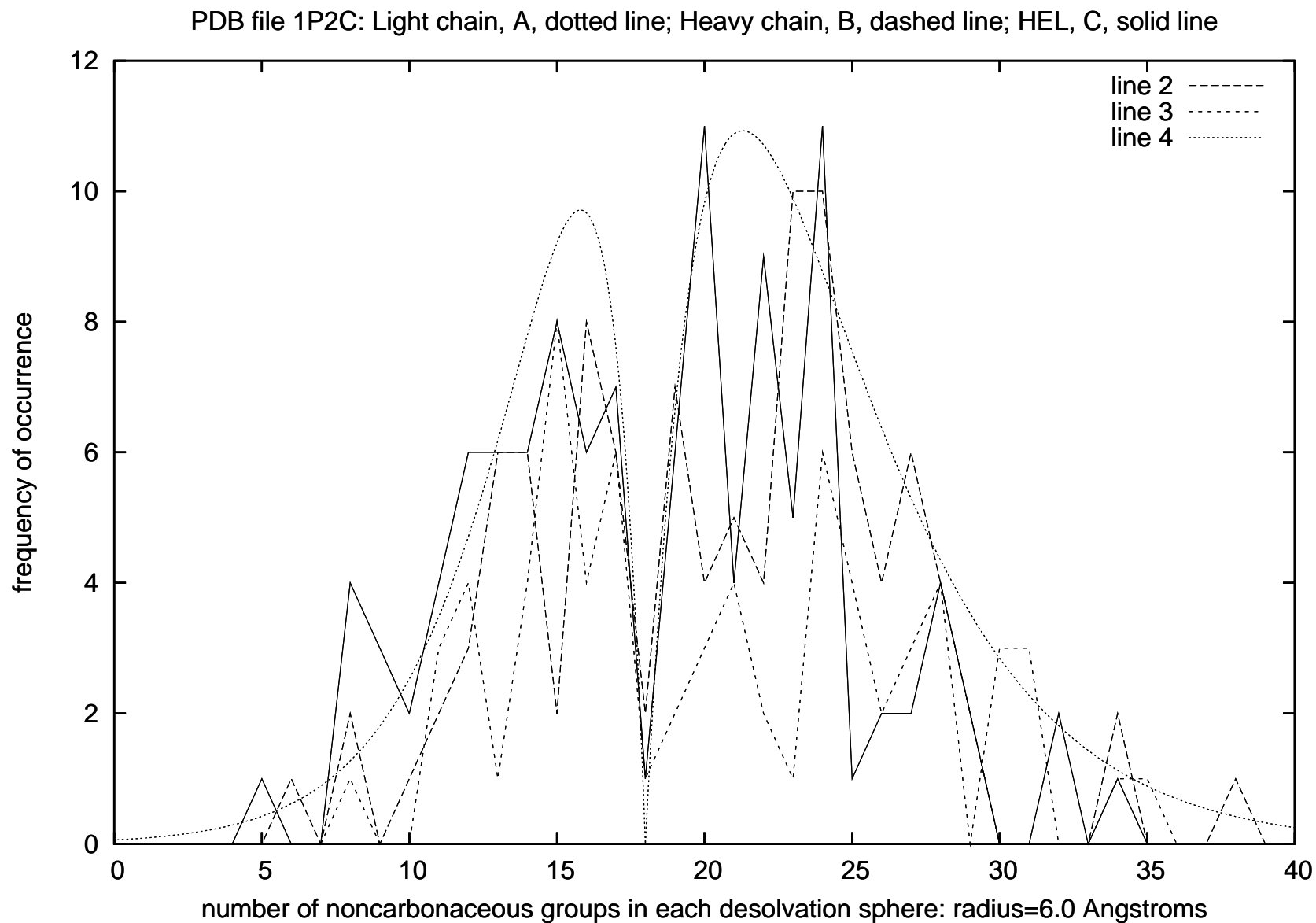
Figure 3: Wireless Charging (from recent Technology Review).

Intramolecular bonds are like the charger on my toothbrush.

Wrapping made quantitative by counting carbonaceous groups in the neighborhood of a hydrogen bond.



Distribution of wrapping for an antibody complex.



2.8 Under-wrapped hydrogen bonds

Hydrogen bonds with insufficient wrapping in one context can become well wrapped by a partner.

The hydrogen bond is much stronger when wrapped.

The change in energy makes these hydrogen bonds sticky.

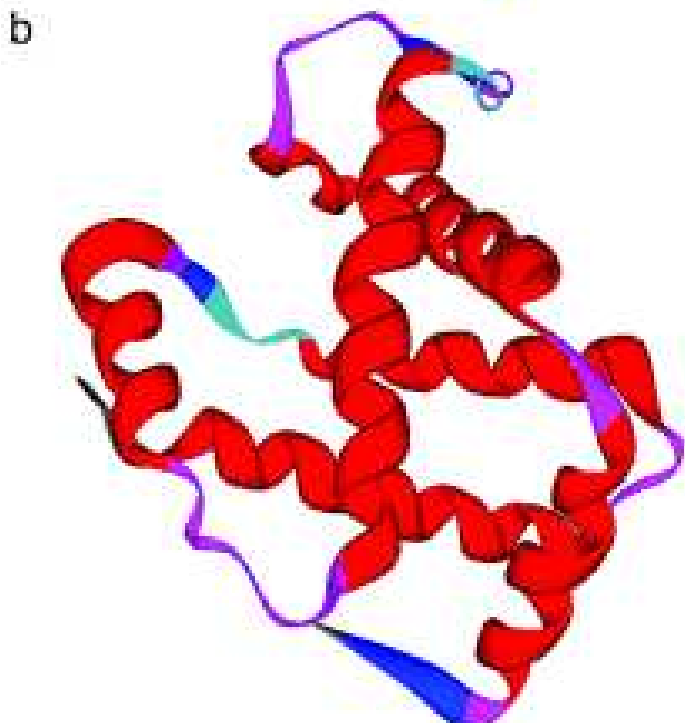
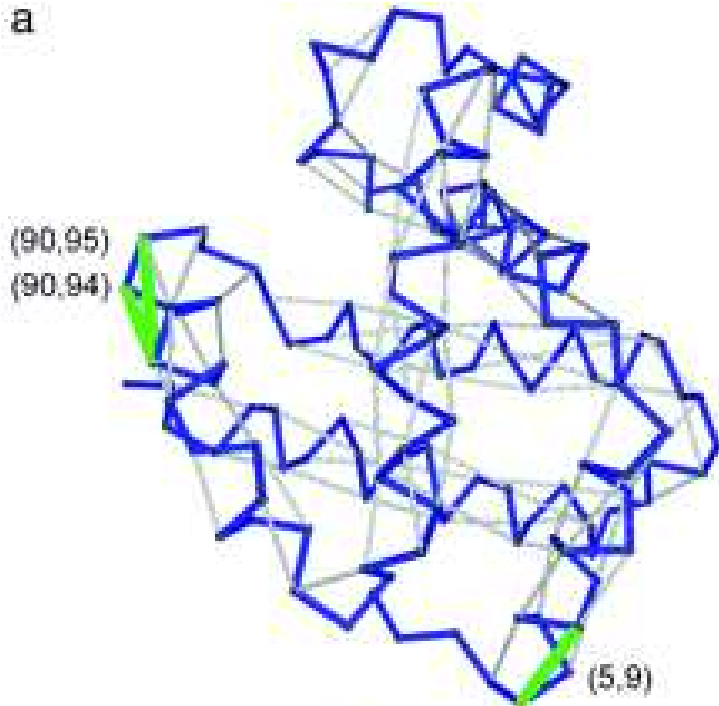
We call such under-wrapped hydrogen bonds

dehydrons

because they can benefit from becoming dehydrated.

The force associated with dehydrons is not huge, but they can act as a guide in protein-protein association.

In our pictures, **our new lens colors dehydrons GREEN** to distinguish from ordinary hydrogen bonds.



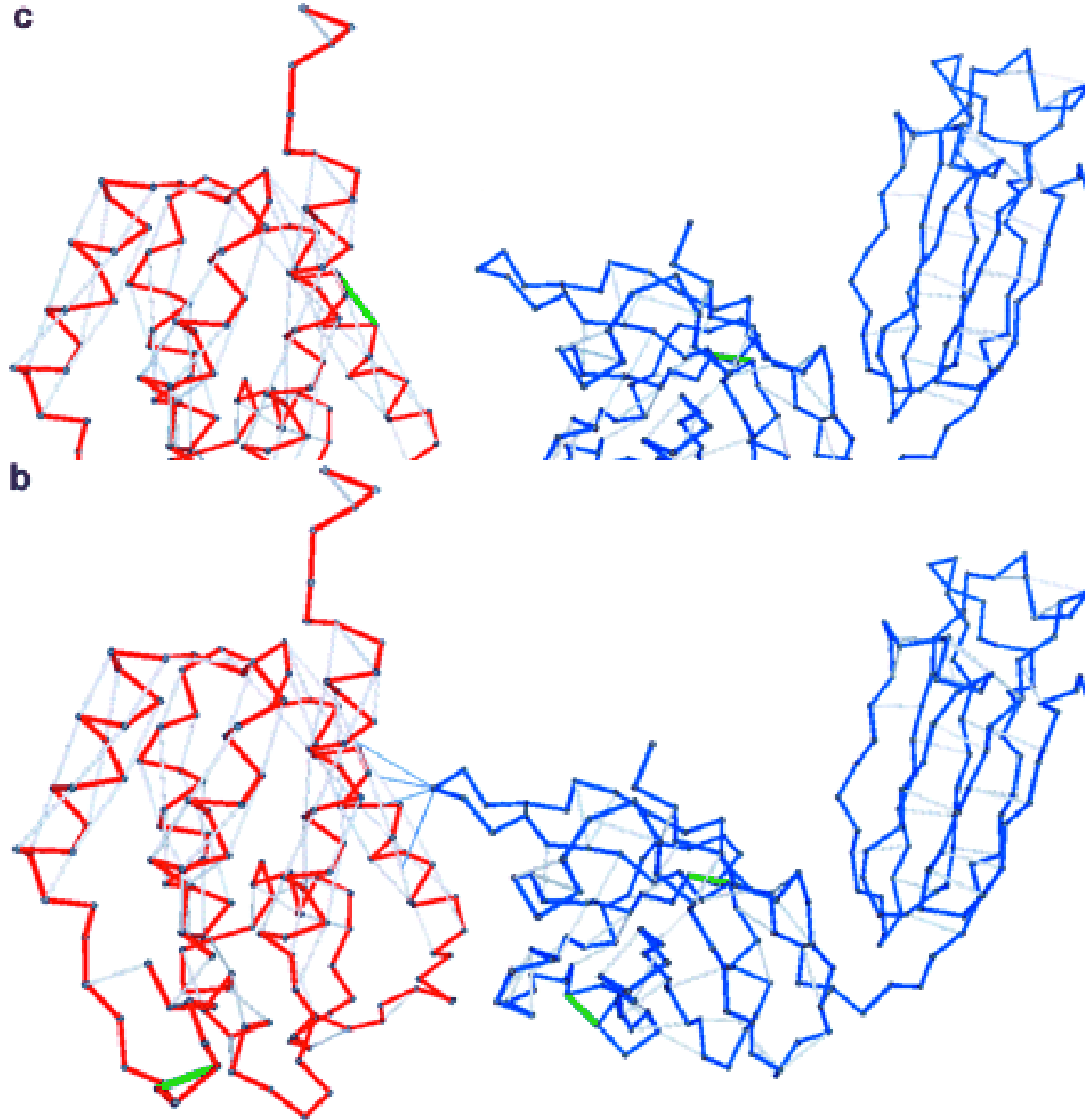
Dehydrons

in human hemoglobin, From PNAS

100: 6446-6451 (2003) Ariel Fernandez,
Jozsef Kardos, L. Ridgway Scott, Yuji Goto,
and R. Stephen Berry. Structural defects and
the diagnosis of amyloidogenic propensity.

Well-wrapped
hydrogen bonds are
grey, and dehydrons are green.

The standard ribbon model
of “structure” lacks indicators
of electronic environment.

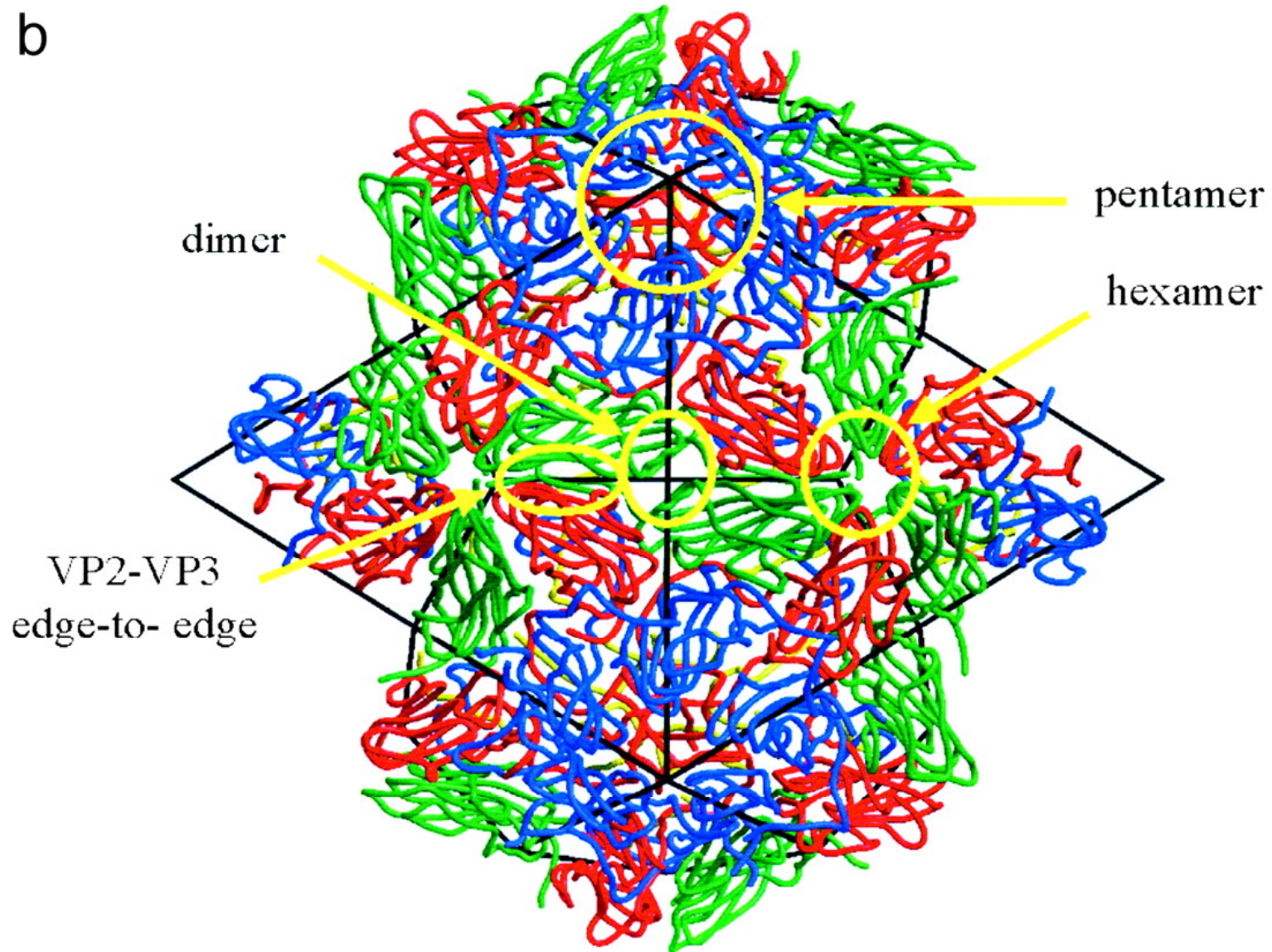


The HIV protease has a dehydrated at an antibody binding site.

When the antibody binds at the dehydrated, it wraps it with hydrophobic groups.

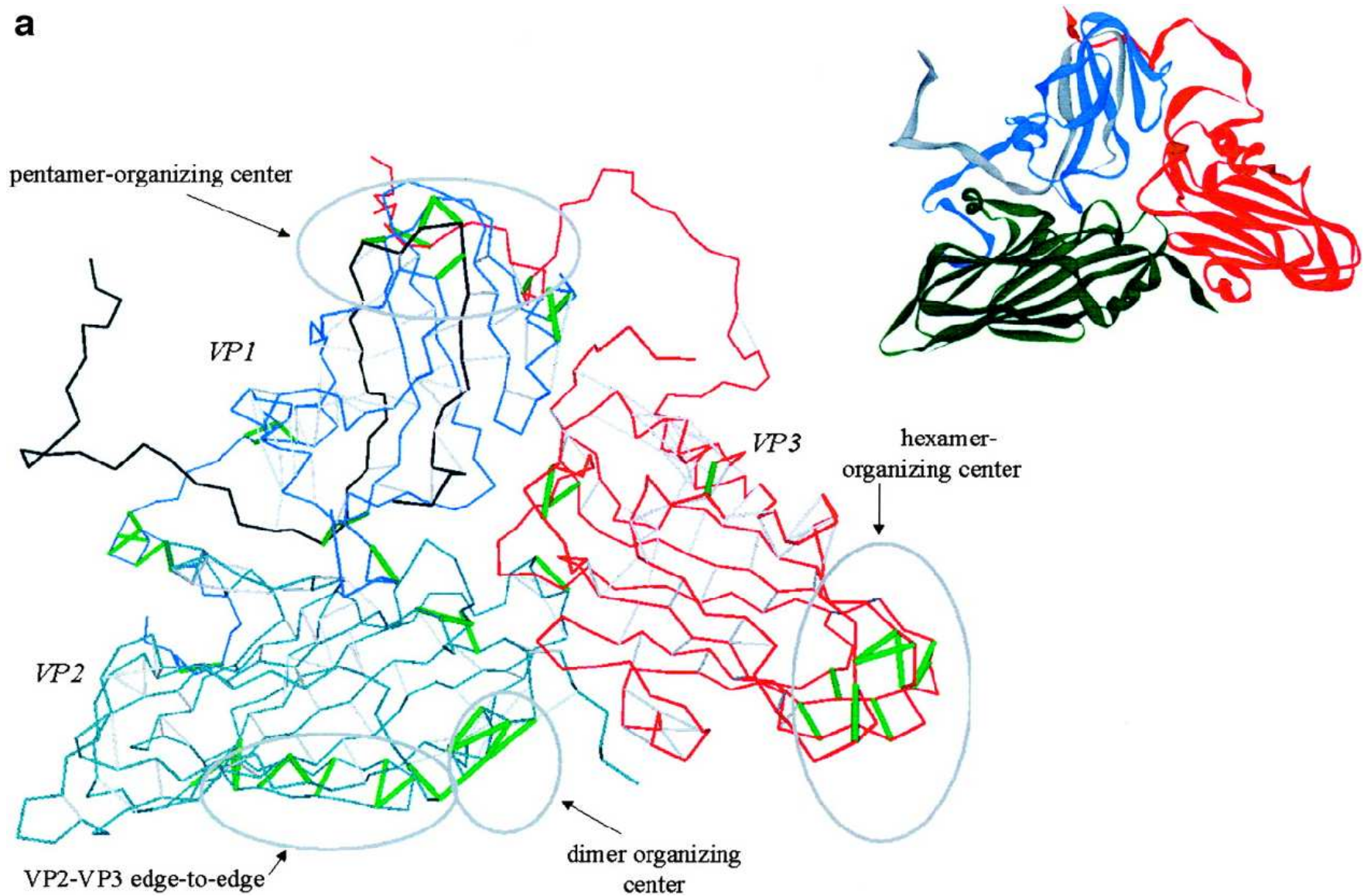
2.9 A model for protein-protein interaction

b



Foot-and-mouth disease virus assembly from small proteins.

a



Dehydrons guide binding of component proteins **VP1, VP2 and VP3** of foot-and-mouth disease virus.

2.10 Stickiness of dehydrons

Attractive force of dehydrons predicted and measured in

Ariel Fernandez and L. Ridgway Scott. Adherence of packing defects in soluble proteins. Phys. Rev. Lett. 2003 91:18102(4)

by considering rates of adhesion to phospholipid (DLPC) bilayer.

Deformation of phospholipid bilayer by dehydrons measured in

Ariel Fernandez and L. Ridgway Scott. Under-wrapped soluble proteins as signals triggering membrane morphology. Journal of Chemical Physics 119(13), 6911-6915 (2003).

Single molecule measurement of dehydronic force in

Ariel Fernandez. Direct nanoscale dehydration of hydrogen bonds. Journal of Physics D: Applied Physics 38, 2928-2932, 2005.

Fine print: careful definition of dehydron requires assessing **modification of dielectric environment by test hydrophobe**. That is, geometry of carbon groups matters, although counting gets it right $\approx 90\%$ of the time [?].

Talk Outline — **Where we are now**

Physical chemistry provides powerful lens to look at protein data

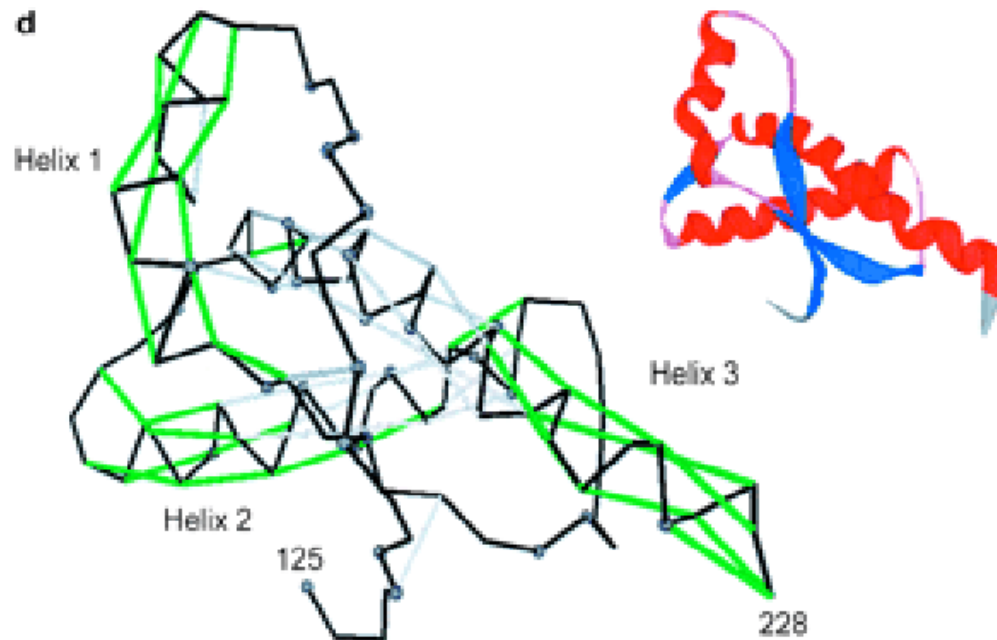
- Tutorial on hydrophobic wrapping
 - hydrophobic protection (desolvation) of hydrogen bonds
 - new motif: dehydron=insufficiently desolvated hydrogen bond
 - dehydrons are involved in protein interaction (they are sticky)
- **Using dehydrons in bioinformatics**
 - the tails of the distribution: **extreme stickiness**
 - number of dehydrons correlates with protein interactivity
 - number of dehydrons differentiates proteins with similar structure
- Using wrapping technology in drug design
- Requires more precise understanding of dielectrics
 - Review of dielectrics
 - Poisson-Debye equation

3 Extreme interaction: amyloid formation

Standard application of bioinformatics: [look at distribution tails](#).

If some is good, more may be better, but too many may be bad.

Too many dehydrons signals trouble: [the human prion](#).



From PNAS 100: 6446-6451 (2003) Ariel Fernandez, Jozsef Kardos, L. Ridgway Scott, Yuji Goto, and R. Stephen Berry. Structural defects and the diagnosis of amyloidogenic propensity.

4 Dehydrons as indicators of protein interactivity

If dehydrons provide mechanism for proteins to interact, then more interactive proteins should have more dehydrons, and vice versa.

We only expect a correlation since there are (presumably) other ways for proteins to interact.

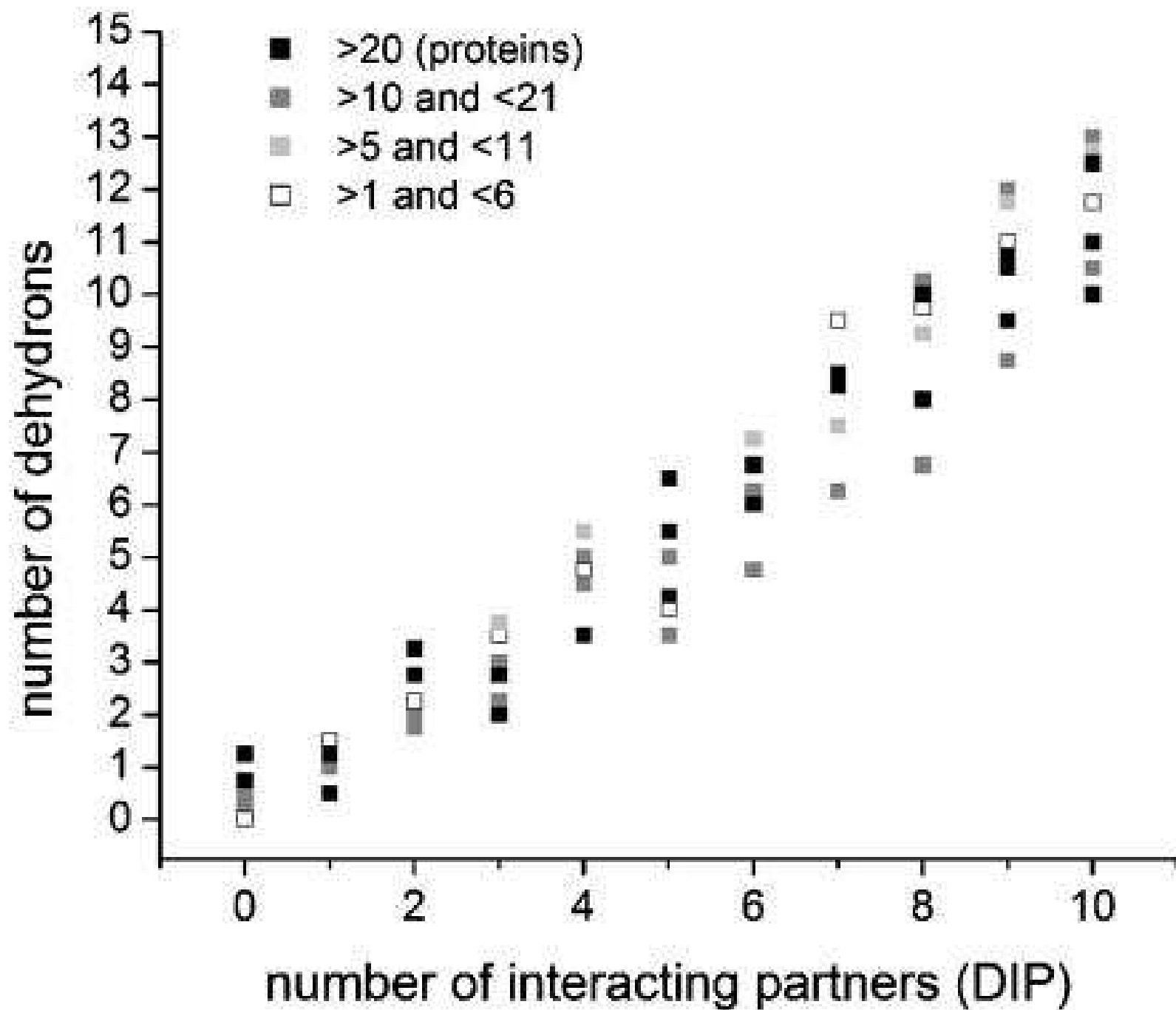
The DIP database collects information about protein interactions, based on individual protein domains: can measure interactivity of different regions of a given protein.

Result: Interactivity of proteins correlates strongly with number of dehydrons.

PNAS 101(9):2823-7 (2004)

The nonconserved wrapping of conserved protein folds reveals a trend toward increasing connectivity in proteomic networks.

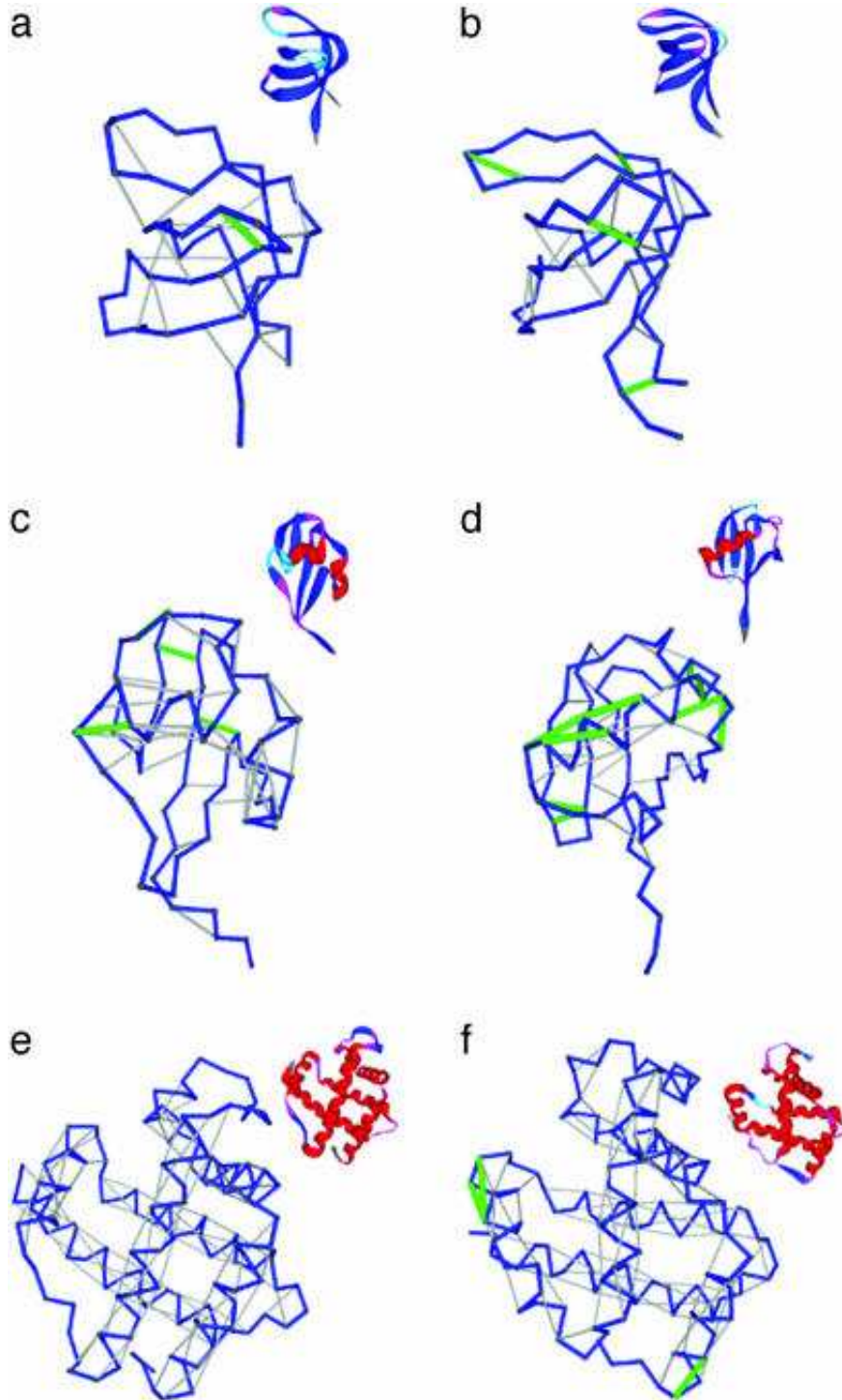
Ariel Fernández, L. R. Scott and R. Steve Berry



4.1 Dehydron variation over different species

Species (common name)	peptides	H bonds	dehydrons
<i>Aplysia limacina</i> (mollusc)	146	106	0
<i>Chironomus thummi thummi</i> (insect)	136	101	3
<i>Thunnus albacares</i> (tuna)	146	110	8
<i>Caretta caretta</i> (sea turtle)	153	110	11
<i>Physeter catodon</i> (whale)	153	113	11
<i>Sus scrofa</i> (pig)	153	113	12
<i>Equus caballus</i> (horse)	152	112	14
<i>Elephas maximus</i> (Asian elephant)	153	115	15
<i>Phoca vitulina</i> (seal)	153	109	16
<i>H. sapiens</i> (human)	146	102	16

Number of dehydrons in Myoglobin of different species



Anecdotal evidence:
the basic
structure is similar, just the
number of dehydrons increases.

SH3 domains are from
nematode *C. elegans* (a)
H. sapiens (b);

ubiquitin is from
E. coli (c) and *H. sapiens* (d);

hemoglobin
is from *Paramecium*
(e). and *H. sapiens*-subunit (f).

4.2 Dehydrons as indicator of complexity?

Is this interactivity an indicator of complexity?

Is this complexity an indicator of evolution?

In any case, the number of dehydrons differentiates homologous proteins found in different species.

We can imagine that protein interactivity became a dominant way in evolution to explore biological space, once genome complexity stabilized.

But regardless, we can exploit dehydron differences in drug design.

4.3 Determinants of folding rates

Contact order determines folding rates for proteins.

Journal of Molecular Biology 277, 985-994 (1998)

Contact order, transition state placement and the refolding rates of single domain proteins

Kevin W. Plaxco, Kim T. Simons and David Baker

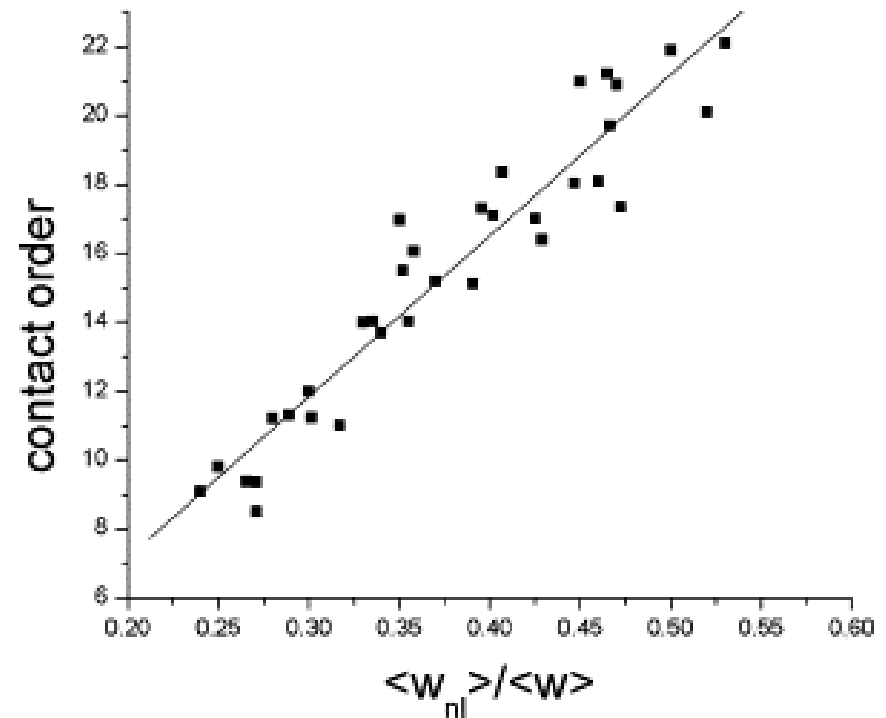
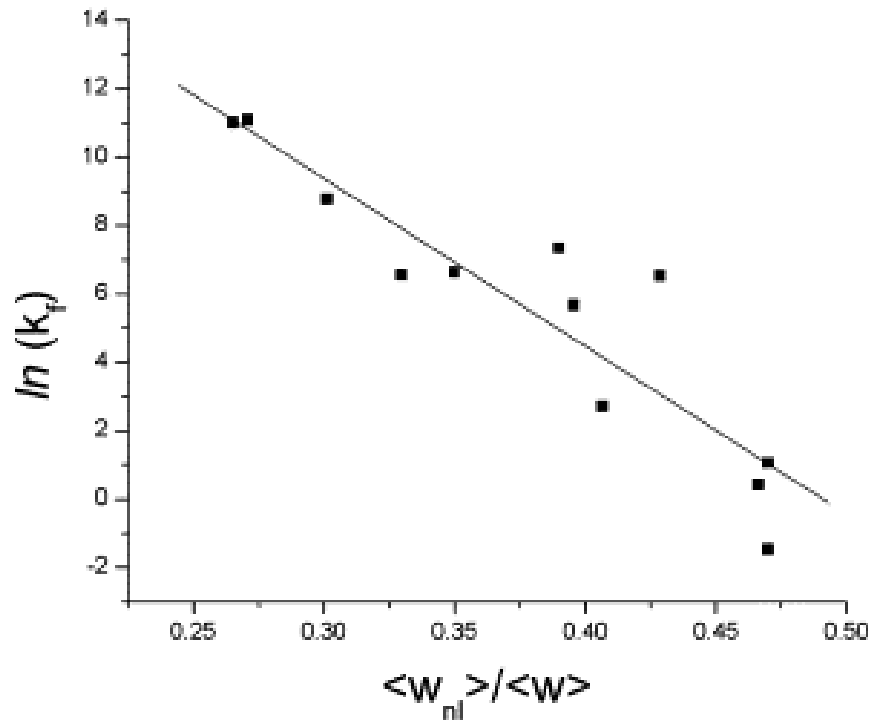
Non-local wrapping of hydrogen bonds gives a similar correlation.

Physics Letters A 321, 263-266 (2004)

Protein folding: a good structure protector is also a good structure seeker

Kristina Rogale and Ariel Fernandez.

From Physics Letters A 321, 263-266 (2004)



Correlation between the logarithm of the unimolecular **folding rate** and the average fraction of **nonlocal** contribution to the **wrapping** of native hydrogen bonds.

Talk Outline — **Where we are now**

Physical chemistry provides powerful lens to look at protein data

- Tutorial on hydrophobic wrapping
 - hydrophobic protection (desolvation) of hydrogen bonds
 - new motif: dehydron=insufficiently desolvated hydrogen bond
 - dehydrons are involved in protein interaction (they are sticky)
- Using dehydrons in bioinformatics
 - the tails of the distribution: extreme stickiness
 - number of dehydrons correlates with protein interactivity
 - number of dehydrons differentiates proteins with similar structure
- **Using wrapping technology in drug design**
- Requires more precise understanding of dielectrics
 - Review of dielectrics
 - Poisson-Debye equation

5 Wrapping technology in drug design

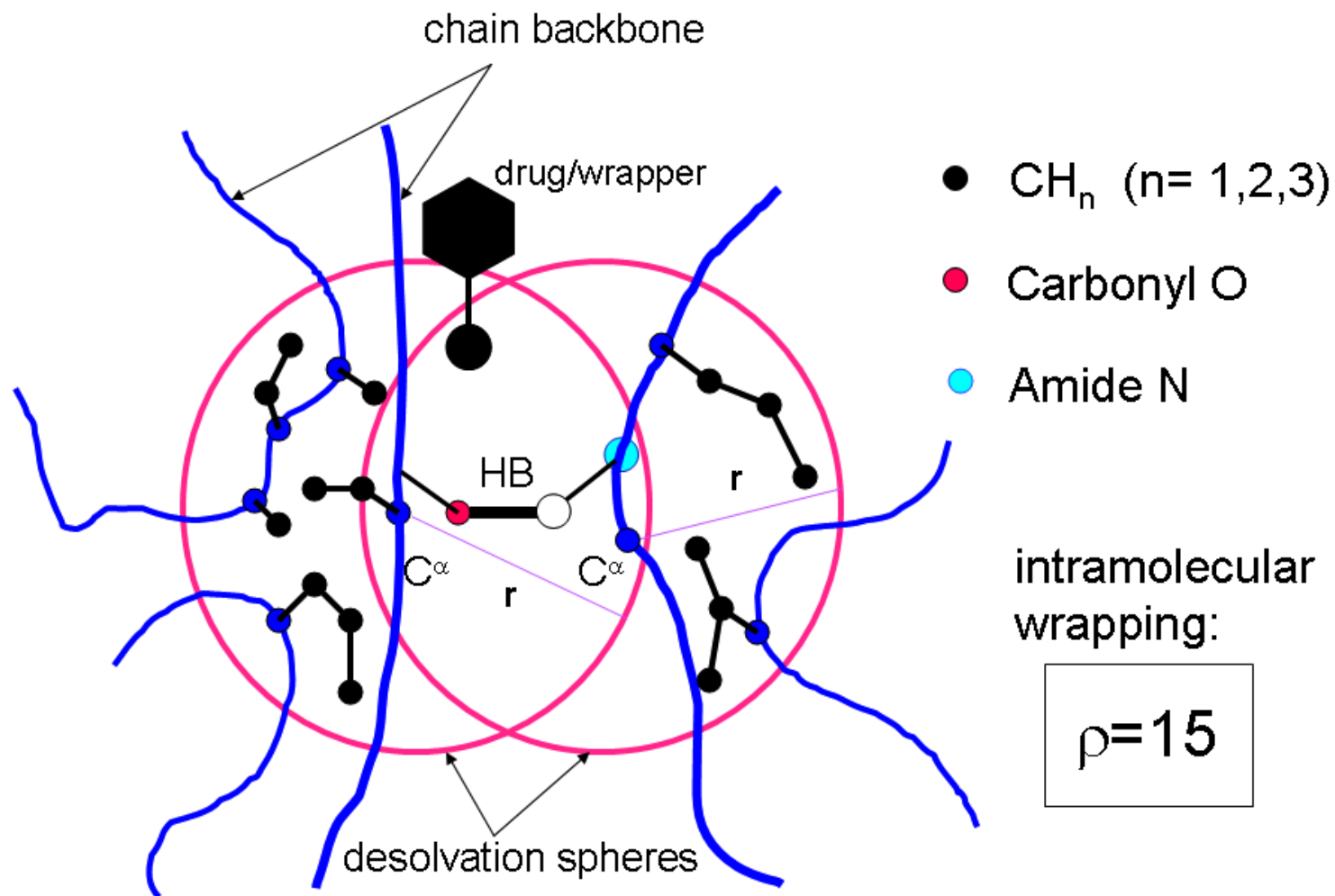
Synopsis of “Modulating drug impact by wrapping target proteins” by Ariel Fernández and L. Ridgway Scott, *Expert Opinion on Drug Discovery* 2007.

Drug ligands often bind to proteins near dehydrons, enhancing their wrapping upon attachment.

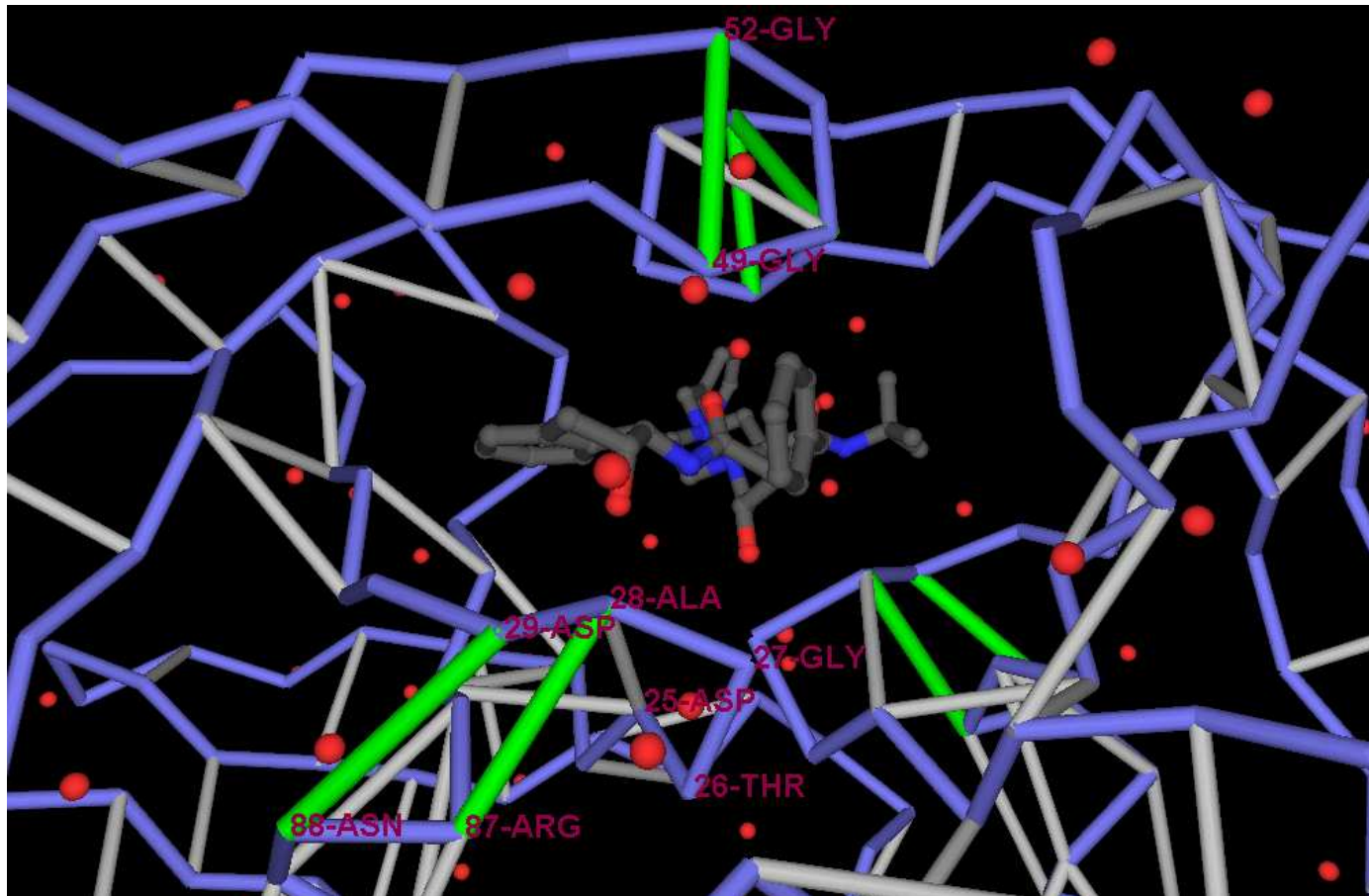
Drug side effects often caused by binding to proteins with structure similar to target.

We can exploit the differences in dehydron patterns in homologous proteins to make drugs more specific.

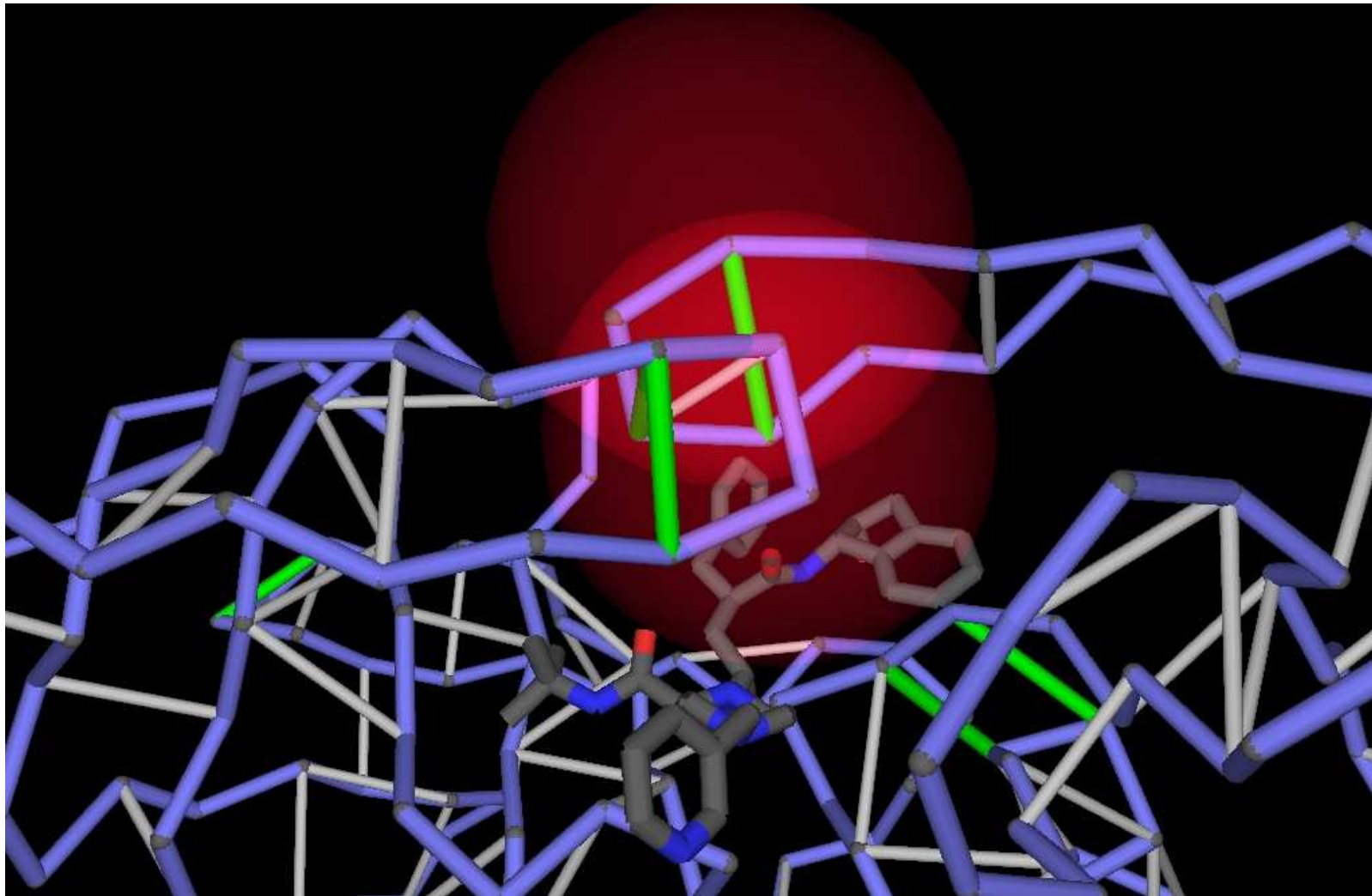
Drug ligand provides additional non-polar carbonaceous group(s) in the desolvation domain, enhancing the wrapping of a hydrogen bond.



HIV-1 protease with 'dehydron wrapper' inhibitor



Detail of the protease cavity, pattern of packing defects, and inhibitor positioned as dehydron wrapper.



Desolvation spheres for flap Gly-49–Gly-52 dehydron containing nonpolar groups of the wrapping inhibitor.

Drug specificity

Tyrosine kinases form a family of proteins with very similar structure.

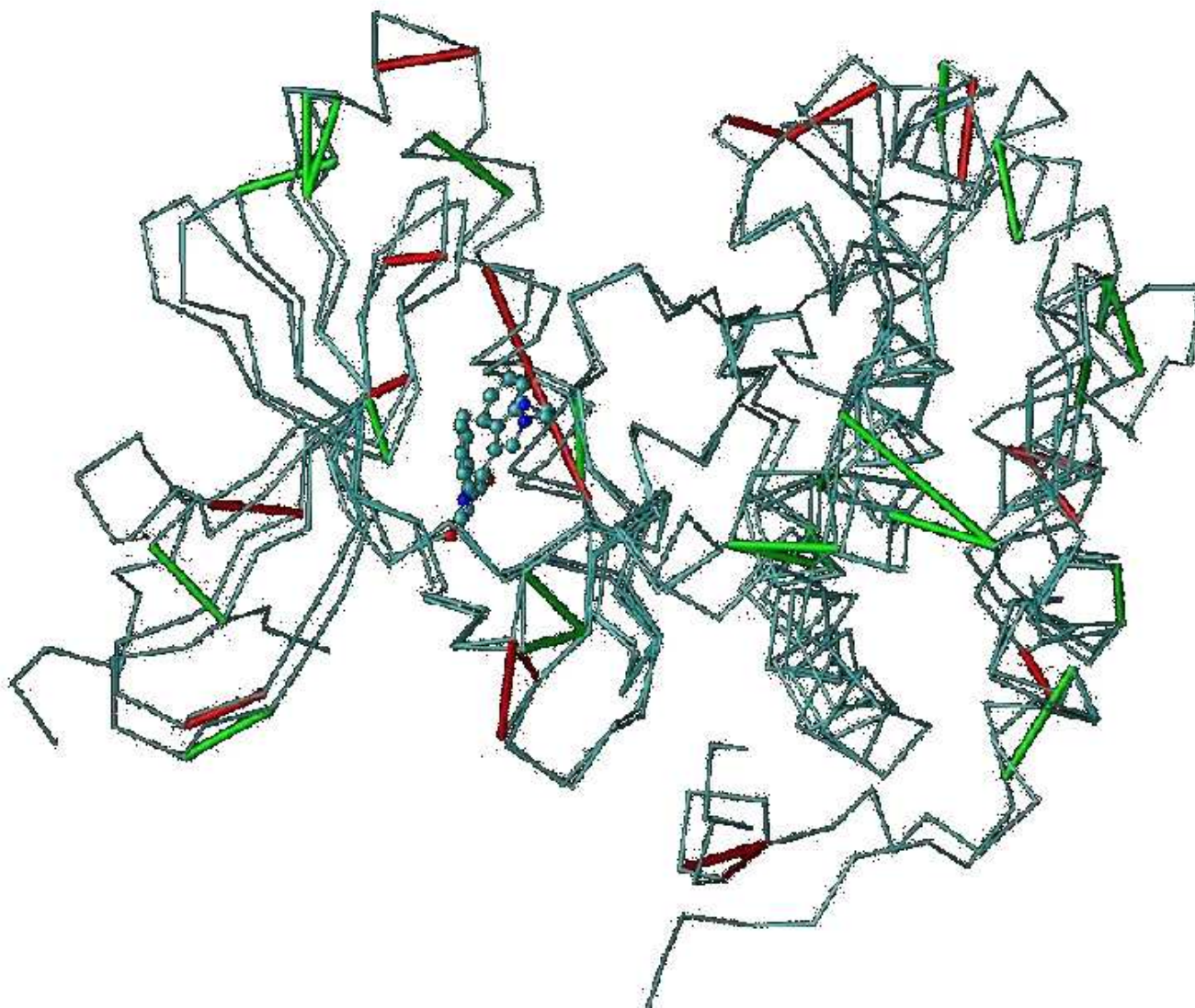
- They are called paralogous because they are similar proteins within a given species.
- These are presumed to have evolved from a common source.
- They are a crucial target of cancer drug therapy.

Gleevec targets particular tyrosine kinases and has been one of the most successful cancer drugs.

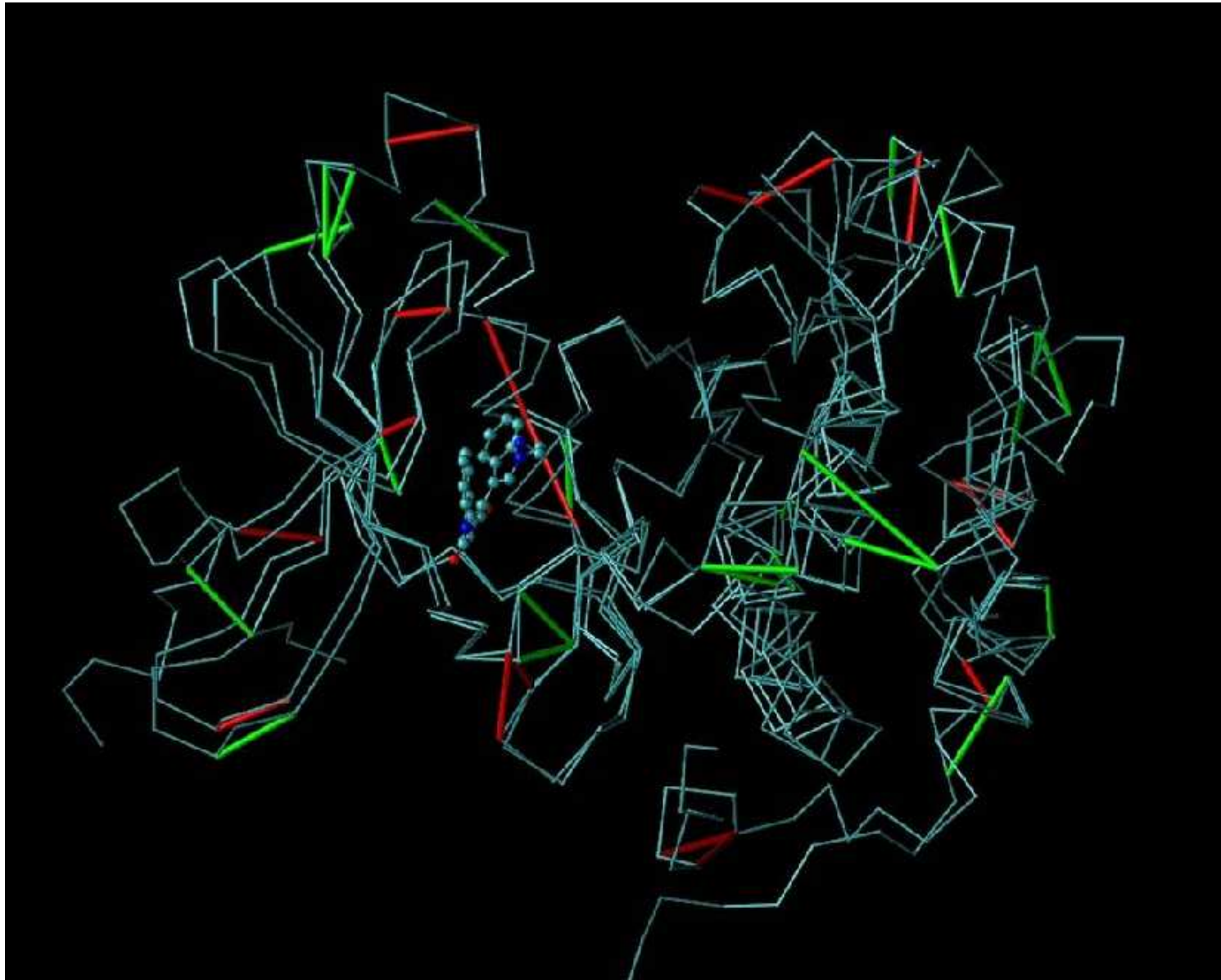
However, it also targets similar proteins and can cause unwanted side effects (it is cardiotoxic).

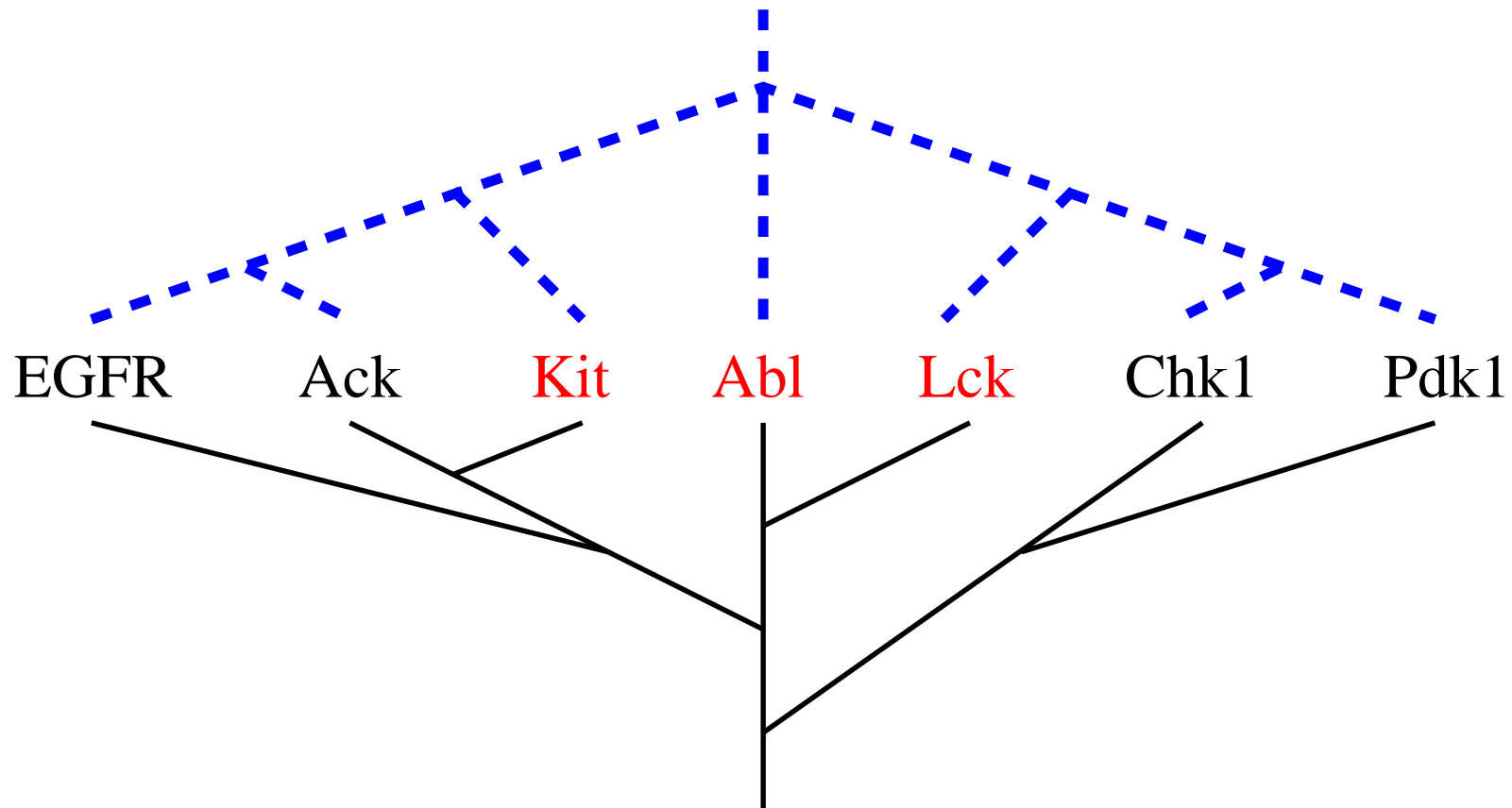
Differences between the dehydron patterns in similar proteins can be used to differentiate them and guide the re-design of drug ligands.

Aligned backbones for two paralog kinases; dehydrons for Chk1 are marked in green and those for Pdk1 are in red.

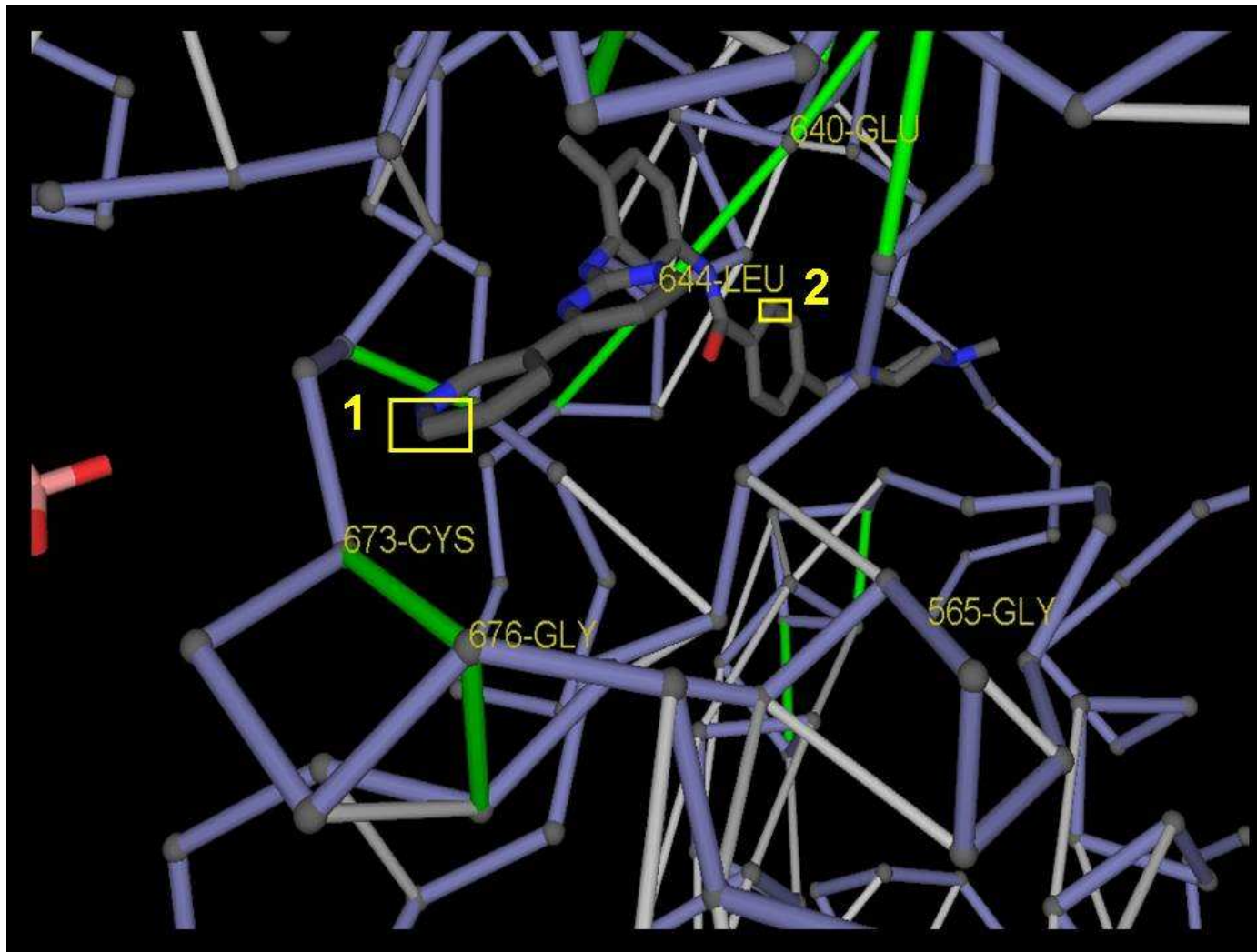


Aligned backbones for two paralog kinases; dehydrons for Chk1 are marked in green and those for Pdk1 are in red.

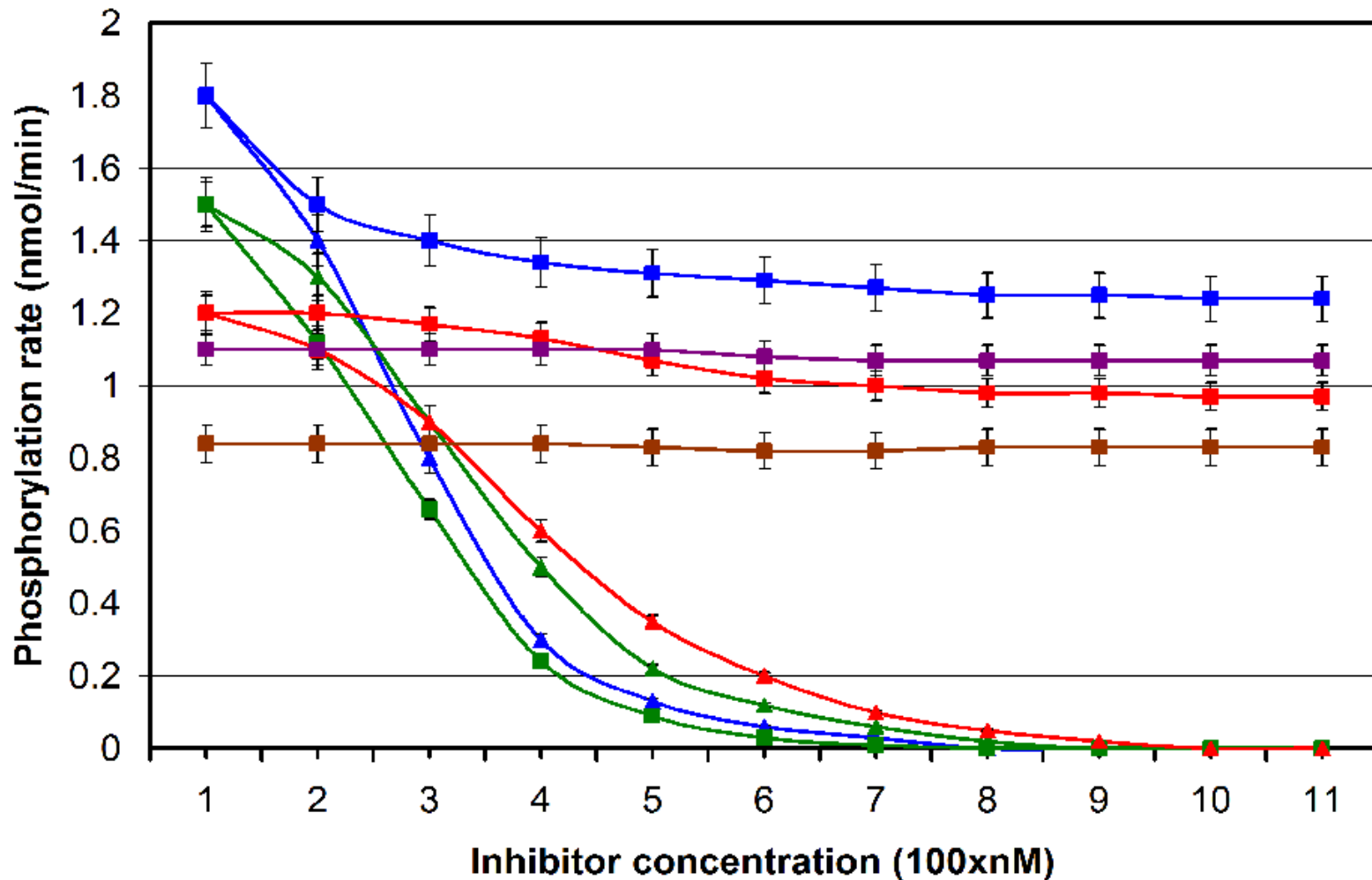




Packing similarity tree (PST, bottom in black) for the seven structurally aligned paralogs of Bcr-Abl. The PST restricted to the alignments of the Gleevec wrapped region in Bcr-Abl is shown (top) with blue dashed lines. The paralogs in red have the most similar packing in the region that aligns with the Gleevec wrapped region in Bcr-Abl and are also primary targets of this inhibitor.



Dehydron Cys673-Gly676 in C-Kit is not conserved in its paralogs Bcr-Abl, Lck, Chk1 and Pdk1. By methylating Gleevec at the para position (1), the inhibitor becomes a selective wrapper of the packing defect in C-Kit.



Phosphorylation rates from spectrophotometric assay on the five kinases Bcr-Abl (blue), C-Kit (green), Lck (red), Chk1 (purple), and Pdk1 (brown) with Gleevac (triangles) and modified Gleevac methylated at positions (1) and (2) (squares). Notice the selective and enhanced inhibition of C-Kit.

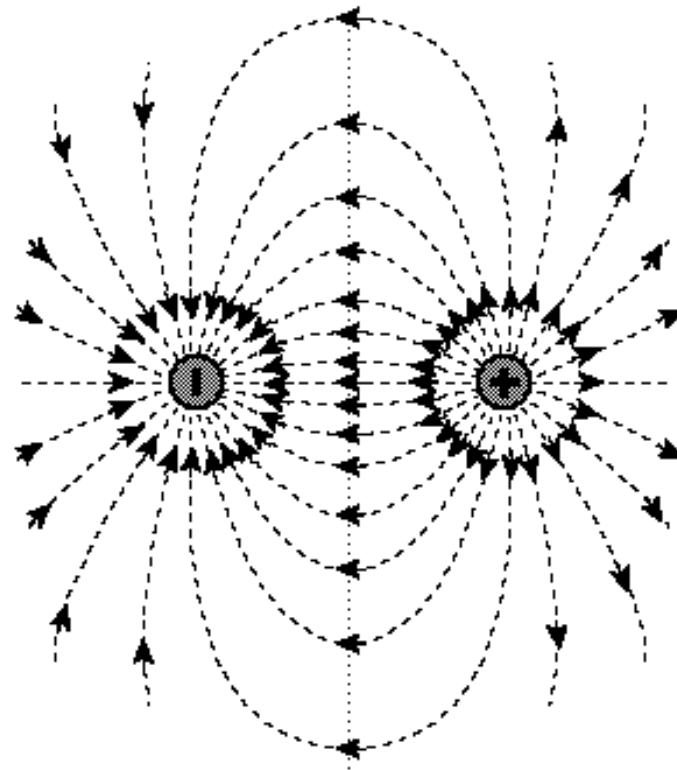
Talk Outline — **Where we are now**

Physical chemistry provides powerful lens to look at protein data

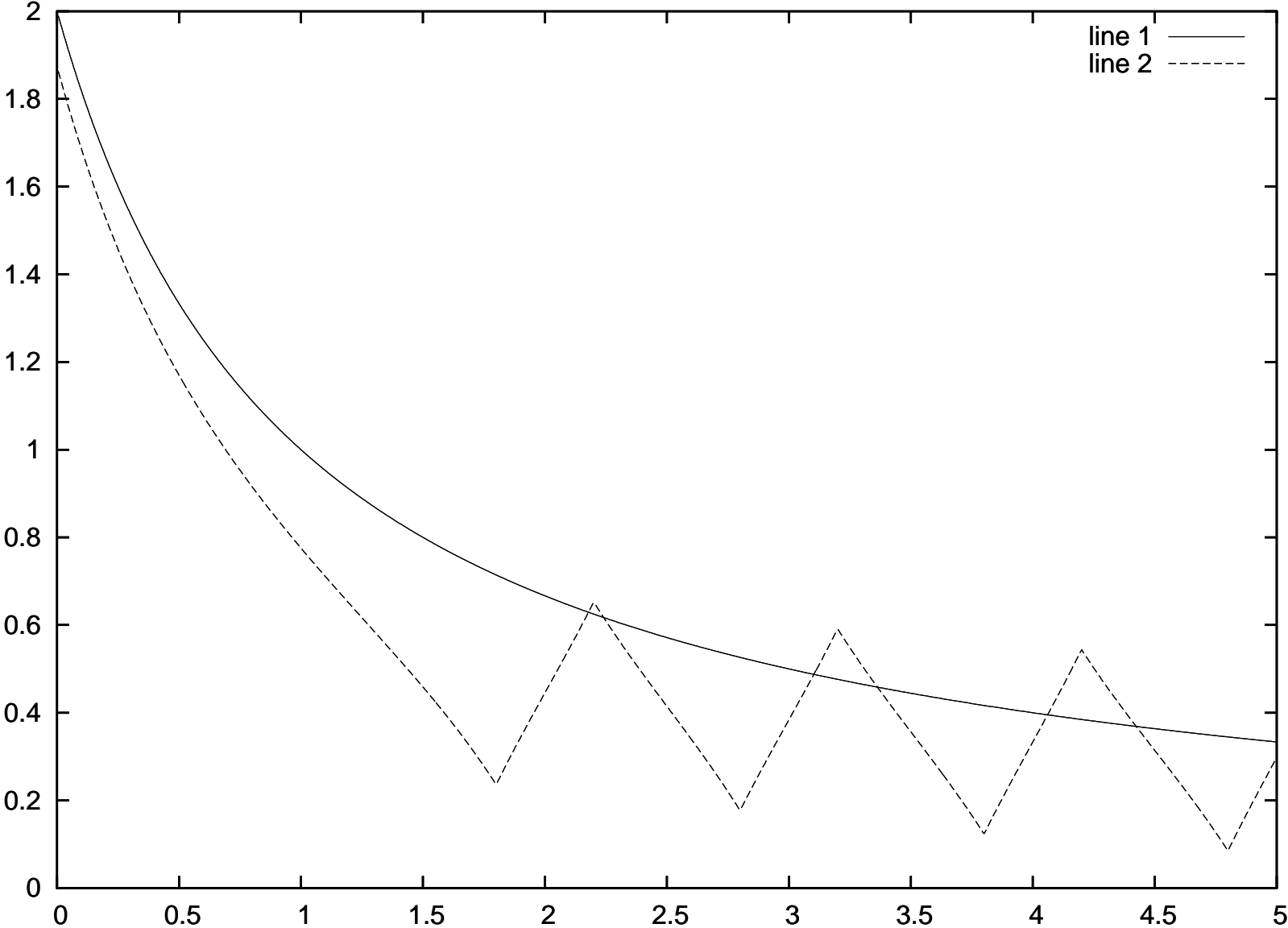
- Tutorial on hydrophobic wrapping
 - hydrophobic protection (desolvation) of hydrogen bonds
 - new motif: dehydron=insufficiently desolvated hydrogen bond
 - dehydrons are involved in protein interaction (they are sticky)
- Using dehydrons in bioinformatics
 - the tails of the distribution: **extreme stickiness**
 - number of dehydrons correlates with protein interactivity
 - number of dehydrons differentiates proteins with similar structure
- Using wrapping technology in drug design
- **Requires more precise understanding of dielectrics**
 - Review of dielectrics
 - Poisson-Debye equation

6 Understanding dielectrics

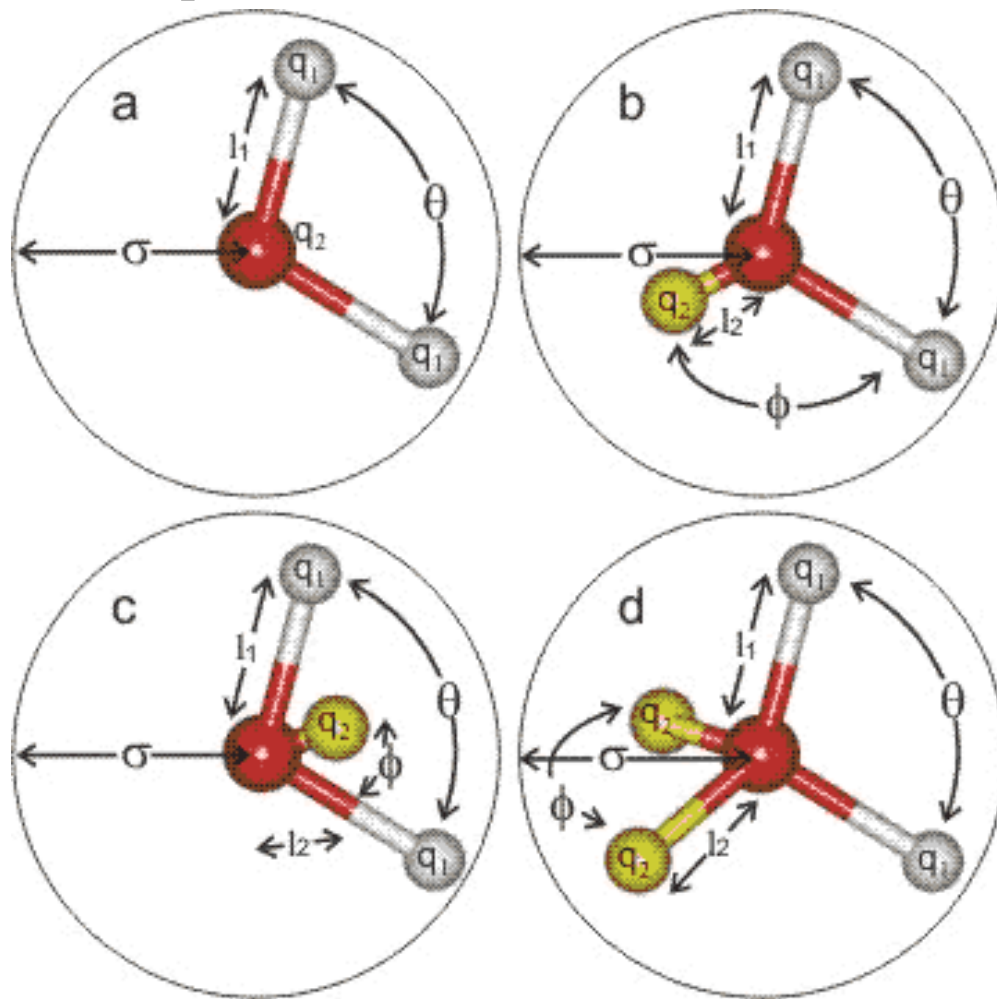
Dipole molecules can moderate charge. Induced dipole interactions lead to van der Waals forces.



Single charge (+2) at zero and dipoles at 2,3,...,10 of width 0.4 and strength 1.



Water is really a double dipole:



In Tip5P, see (d), $q_1 = -q_2 = 0.241$, $l_1 = 0.9572$, $l_2 = 0.70$, $\theta = 104.52$,
 $\phi = 109.47$.

7 Continuum equations for electrostatics

The basic equations of electrostatics for a collection of charges of strength q_i at positions \mathbf{r}_i can be derived from the simple expression

$$\nabla \cdot (\epsilon_0 \mathbf{e}) = \tilde{\gamma}(\mathbf{r}) := \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (7.1)$$

where ϵ_0 is the permittivity of the vacuum. Here $\tilde{\gamma}$ is the charge density and \mathbf{e} is the induced electric field.

We will split the charge density into two parts: $\tilde{\gamma} = \rho + \gamma$ where

- One part ρ comes from fixed charges (e.g., in a protein),
- the other part γ comes from molecules (e.g., water), with net charge zero that are free to rotate. Their configuration is induced by the fixed charges ρ .

We focus on the latter charges first: the dielectric effect.

You may fast-forward to (??) for the main point: $\gamma = -\nabla \cdot \mathbf{p}$.

7.1 Dielectric materials

A dielectric medium is characterized by the fact that the charges are organized in ‘polar’ groups with net charge zero.

Specifically we assume that (a designated part of) the \mathbf{r}_i and q_i can be enumerated as $i = (j, k)$, where

- j is the index for the group and
- k is the index within each group, with
- $\mathbf{r}_i = \mathbf{r}_{(j,k)} := \mathbf{r}_j - \mathbf{r}_{jk}$ and $q_i := q_{jk}$ where

the j -th group of charges q_{jk} sums to zero for all j :

$$\sum_k q_{jk} = 0. \quad (7.2)$$

Then the expression for the charge density can be simplified as

$$\gamma(\mathbf{r}) = \sum_j \sum_k q_{jk} \delta(\mathbf{r} - \mathbf{r}_j - \mathbf{r}_{jk}) \quad (7.3)$$

7.2 Homogeneous charge groups

Let us suppose that the charge groups are homogeneous in the sense that

$$\mathbf{r}_{jk} = \mathcal{R}(\theta_j)\rho_k \quad (7.4)$$

for fixed vectors ρ_k and for some angle $\theta_j \in S_2$ (where S_2 denotes the unit 2-sphere), and further that $q_{jk} = q_k$ independent of j , as is true for certain models of water, such as Tip5P [?]. Then

$$\sum_k q_{jk} \delta(\mathbf{r} - \mathbf{r}_j - \mathbf{r}_{jk}) = \sum_k q_k \delta(\mathbf{r} - \mathbf{r}_j - \mathcal{R}(\theta_j)\rho_k) = \mathcal{F}(\theta_j, \mathbf{r} - \mathbf{r}_j) \quad (7.5)$$

where \mathcal{F} is defined by

$$\mathcal{F}(\theta, \mathbf{r}) = \sum_k q_k \delta(\mathbf{r} - \mathcal{R}(\theta)\rho_k). \quad (7.6)$$

Note that we have now expressed the dielectric charge field γ as

$$\gamma(\mathbf{r}) = \sum_j \mathcal{F}(\theta_j, \mathbf{r} - \mathbf{r}_j) \quad (7.7)$$

7.3 Rotationally invariant δ

If δ is rotationally invariant, then

$$\mathcal{F}(\theta, \mathcal{R}(\theta)\mathbf{r}) = \sum_k q_k \delta(\mathcal{R}(\theta)\mathbf{r} - \mathcal{R}(\theta)\rho_k) = \sum_k q_k \delta(\mathbf{r} - \rho_k) = \nabla \cdot \mathcal{W}(\mathbf{r}), \quad (7.8)$$

where $\mathcal{W}(\mathbf{r}) = \nabla\psi(\mathbf{r})$ and ψ solves $\Delta\psi = \sum_k q_k \delta(\mathbf{r} - \rho_k)$. If $\delta = \text{Dirac } \delta\text{-function}$, then $\mathcal{W} = \text{generalized multipole expression}$

$$\mathcal{W}(\mathbf{r}) = - \sum_k q_k \frac{\mathbf{r} - \rho_k}{|\mathbf{r} - \rho_k|^3}. \quad (7.9)$$

Then

$$\begin{aligned} \sum_k q_{jk} \delta(\mathbf{r} - \mathbf{r}_j - \mathbf{r}_{jk}) &= \mathcal{F}(\theta_j, \mathbf{r} - \mathbf{r}_j) \\ &= \nabla \cdot \mathcal{W}(\mathcal{R}(\theta_j)^t(\mathbf{r} - \mathbf{r}_j)). \end{aligned} \quad (7.10)$$

Therefore

$$\gamma(\mathbf{r}) = \sum_j \nabla \cdot \mathcal{W}(\mathcal{R}(\theta_j)^t(\mathbf{r} - \mathbf{r}_j)) := -\nabla \cdot \mathbf{p}(\mathbf{r}), \quad (7.11)$$

where \mathbf{p} is called the **polarization vector field**.

7.4 Polarization field

With additional charge density ρ independent of the dielectric, (??) becomes

$$\nabla \cdot (\epsilon_0 \mathbf{e}) = \rho + \gamma = \rho - \nabla \cdot \mathbf{p}, \quad (7.12)$$

where the **polarization vector field** \mathbf{p} is defined by

$$\mathbf{p}(\mathbf{r}) = - \sum_j \mathcal{W}(\mathcal{R}(\theta_j)^t(\mathbf{r} - \mathbf{r}_j)). \quad (7.13)$$

We now define a modified electric field \mathbf{v} defined by

$$\mathbf{v} = \mathbf{e} + \frac{1}{\epsilon_0} \mathbf{p}.$$

Then

$$\nabla \cdot (\epsilon_0 \mathbf{v}) = \nabla \cdot (\epsilon_0 \mathbf{e}) + \nabla \cdot \mathbf{p} = \rho. \quad (7.14)$$

The field \mathbf{v} can be interpreted as the field associated with ρ in free space.

7.5 Debye's Ansatz

Directions θ_j that determine the polarization tend on average to cause \mathbf{p} to line up with the induced field: $\mathbf{p} \parallel \mathbf{e}$.

In a thermalized system, there will be fluctuations in the angles θ_j , and we can only talk about mean angles. Debye [?] suggested that

$$\tilde{\mathbf{p}} = (\epsilon - \epsilon_0)\tilde{\mathbf{e}} \quad (7.15)$$

where ϵ denotes an effective permittivity.

Here, $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{e}}$ denote temporal averages over timescale long with respect to the basic thermal motions. We drop tildes and remember that everything represents temporal averages. This gives

$$\nabla \cdot (\epsilon \mathbf{e}) = \nabla \cdot (\mathbf{p} + \epsilon_0 \mathbf{e}) = \nabla \cdot (\epsilon_0 \mathbf{v}) = \rho \quad (7.16)$$

Remarkably, ϵ can be quite large (about eighty times ϵ_0) for water.

7.6 Debye's Ansatz as projection

Uncomfortable with the Debye ansatz? Define $\tilde{\mathbf{p}}$ to be the projection of \mathbf{p} onto the direction \mathbf{e} , with ϵ defined by requiring $\tilde{\mathbf{p}} = (\epsilon - \epsilon_0)\mathbf{e}$.

That is, we write $\mathbf{p} = (\epsilon - \epsilon_0)\mathbf{e} + \zeta\mathbf{e}^\perp$, so that

$$\epsilon = \epsilon_0 + \frac{\mathbf{p} \cdot \mathbf{e}}{\mathbf{e} \cdot \mathbf{e}},$$

with the appropriate optimism that $\mathbf{p} = 0$ when $\mathbf{e} = 0$.

That is, $\epsilon - \epsilon_0$ reflects the correlation between \mathbf{p} and \mathbf{e} .

As defined, ϵ is a function of \mathbf{r} and t , and potentially singular.

Debye postulated that a suitable average $\tilde{\epsilon}$ should be well behaved:

$$\tilde{\epsilon} = \epsilon_0 + \left\langle \frac{\mathbf{p} \cdot \mathbf{e}}{\mathbf{e} \cdot \mathbf{e}} \right\rangle.$$

8 Models for solvated proteins

Now we apply this to a protein surrounded by water.

The electric field \mathbf{e} is determined by the charge field ρ representing the (partial) charges of the sidechains in the protein:

$$\nabla \cdot (\epsilon \mathbf{e}) = \rho \quad (8.17)$$

where the dielectric field ϵ represents some estimation of the resulting correlation between the resulting electric field \mathbf{e} and the induced polarization field \mathbf{p} :

$$\epsilon = \epsilon_0 + \left\langle \frac{\mathbf{p} \cdot \mathbf{e}}{\mathbf{e} \cdot \mathbf{e}} \right\rangle.$$

In bulk water (far away from the protein), we expect $\epsilon \approx 78\epsilon_0$.

‘Inside’ the protein, we expect ϵ to be much closer to ϵ_0 .

But can we just assume that ϵ jumps from one value to another near the ‘surface’ of the protein?

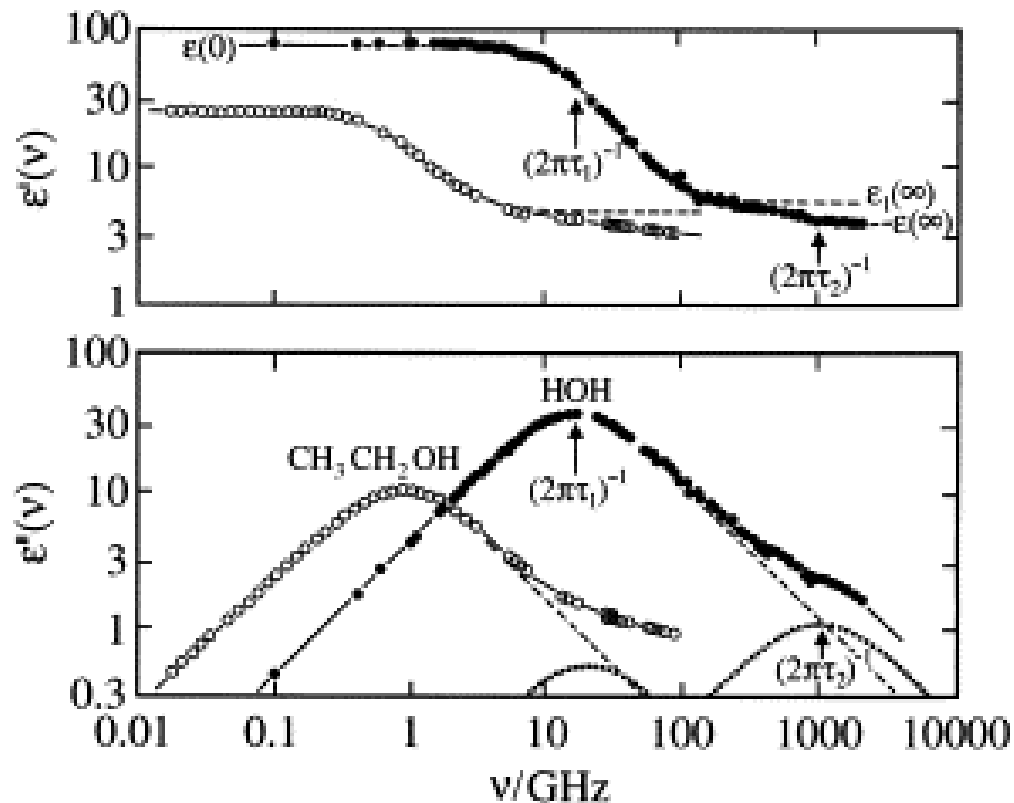
We have suggested [?] that a Fourier Integral Operator must be used to get accurate results near the protein.

8.1 Frequency dependence of dielectric constant

Debye observed that the effective permittivity is frequency dependent:

$$\epsilon(\nu) = \epsilon_0 + \frac{\epsilon_1 - \epsilon_0}{1 + \tau_D^2 \nu^2} \quad (8.18)$$

where τ_D is a characteristic time associated with the dielectric material and ν is the temporal wave number. **Many experiments have verified this [?]:**



8.2 Spatial dependence of electric potential

We are interested in electric fields which are *not* time varying (i.e., $\nu = 0$) but rather spatially varying. Typical charge variations on the surface of a protein: a salt bridge is a pair of sidechains forming a dipole with a separation of a few Å.

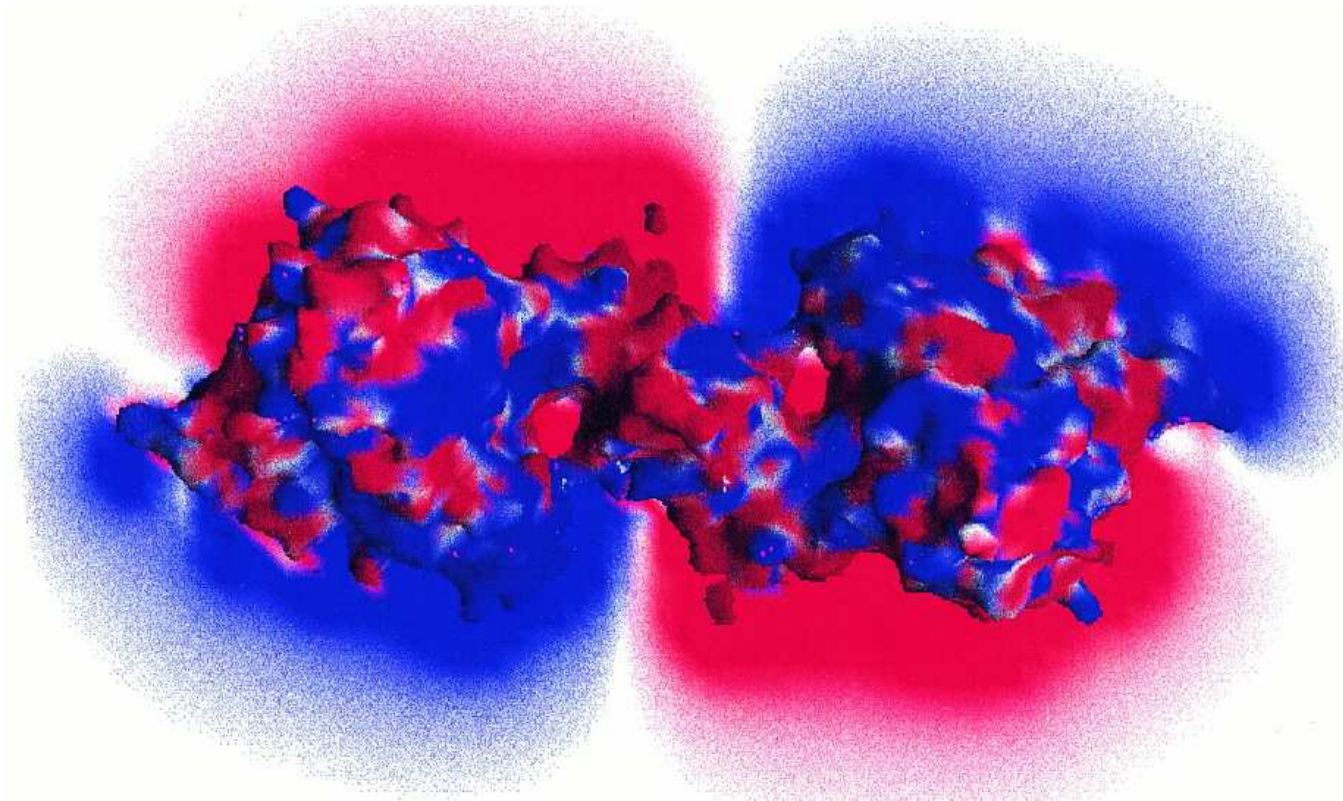


Figure 4: A snapshot [?] of electrostatic potential around the AChE dimer. The value of the potential is given in $(\text{kcal/mol})e^{-1}$ and ranges over $[-0.5, 0.5]$.

8.3 Non-local, frequency-dependent relationship between \mathbf{p} and \mathbf{e}

A model has been proposed by many people [?, ?, ?] to account for the spatial frequency dependence of the (zero temporal frequency) dielectric relationship.

It takes the form

$$\mathbf{p} = \int K(\mathbf{r}, \mathbf{r}') \mathbf{e}(\mathbf{r}') d\mathbf{r} \quad (8.19)$$

where the averaging kernel K satisfies

$$K(\mathbf{r}, \mathbf{r}') = \Gamma(\mathbf{r}, \mathbf{r}') \int e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} L(\mathbf{k}) d\mathbf{k} \quad (8.20)$$

with the expression L representing the **Debye-like frequency dependence**

$$L(\mathbf{k}) = \epsilon_0 + \frac{\epsilon_1 - \epsilon_0}{1 + |\mathbf{k}|^2 \lambda^2} \quad (8.21)$$

and $\Gamma(\mathbf{r}, \mathbf{r}')$ depends on the distribution of hydrophobes.

We use empirical evidence to determine the length scale λ , which we find to be approximately 1.7 Ångstroms [?] in bulk water.

9 The Poisson-Debye equation

The non-local relationship between \mathbf{p} and \mathbf{e} leads to the following:

$$\nabla \cdot \mathcal{P} \nabla \phi = \rho \quad (9.22)$$

where the operator \mathcal{P} operates component-wise and is defined by

$$\mathcal{P}v(\mathbf{r}) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \Gamma(\mathbf{r}, \mathbf{r}') L(\mathbf{k}) v(\mathbf{r}') d\xi d\mathbf{r}' \quad (9.23)$$

and we recall that the symbol L was given by the Debye-like relationship

$$L(\mathbf{k}) = \epsilon_0 + \frac{\epsilon_1 - \epsilon_0}{1 + |\mathbf{k}|^2 \lambda^2} \quad (9.24)$$

and with hydrophobes located at positions \mathbf{r}_j we might take [?, ?]

$$\Gamma(\mathbf{r}, \mathbf{r}') = C \left(1 + \sum_j e^{-(|\mathbf{r} - \mathbf{r}_j| + |\mathbf{r}' - \mathbf{r}_j|) / \lambda} \right) \quad (9.25)$$

Clearly, \mathcal{P} is a symmetric Fourier Integral Operator with symbol $\Gamma(\mathbf{r}, \mathbf{r}') L(\mathbf{k})$.

9.1 Bulk behavior

When Γ is constant ($= (2\pi)^{-3}$), which holds well away from any hydrophobes, there is a simple relation

$$\hat{\phi}(\mathbf{k}) = \frac{\hat{\rho}(\mathbf{k})}{|\mathbf{k}|^2 L(\mathbf{k})} \quad (9.26)$$

which can be used to compute ϕ (and thus \mathbf{e}) from ρ .

We can look at the Green's Function for this system, i.e., $\rho = \delta$ which corresponds to an isolated charge.

This defines a family of potentials $\phi_{\epsilon_r, \lambda}$ for any given ϵ_r and λ , where $\epsilon_r = \epsilon_1/\epsilon_0$. Thus a simple change of variables implies that

$$\phi(\mathbf{r}) = \frac{\epsilon_0}{\lambda} \phi_{\epsilon_r, 1}(\mathbf{r}/\lambda) \quad (9.27)$$

where $\phi_{\epsilon_r, 1}$ is defined using the kernel

$$\tilde{L} = 1 + \frac{\epsilon_r - 1}{1 + \mathbf{k}^2}.$$

In the limit $\lambda \rightarrow \infty$ we find $\phi_\infty = c/\epsilon_0 r$.

9.2 Green's function computation

The computation of $\phi_{\epsilon_r,1}$ requires us to consider the Fourier transform of a radially symmetric function u (which is itself radially symmetric). The following formula holds:

$$|\xi|\widehat{u}(\xi) := 4\pi \int_0^\infty ru(r) \sin(|\xi|r) dr \quad (9.28)$$

where $u(r)$ means $u(\mathbf{r})$ with $r = |\mathbf{r}|$.

Thus the expression $\epsilon_0\phi_\infty/\phi_\lambda = c/r\phi_\lambda$ is

9.3 Green's function picture

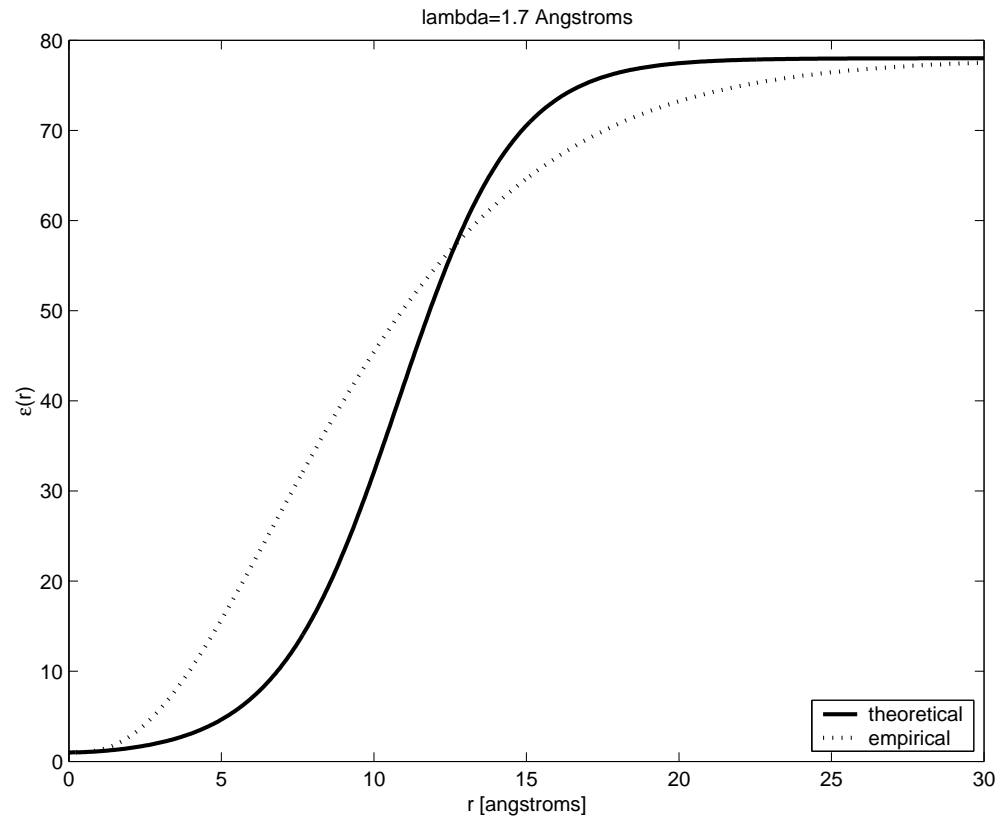


Figure 5: Comparison of empirical permittivity formula in [?] (dotted line) with the model (??) (solid line) where L is defined in (??) with $\lambda = 1.7$. Plotted are the ratios of the effective permittivity experienced by a dipole to the vacuum permittivity as a function of the separation distance in \AA , where the bulk permittivity is that of water.

References

- [1] P. Debye. *Polar molecules*. Dover, New York, 1945.
- [2] Ariel Fernández. Intramolecular modulation of electric fields in folding proteins. *Phys. Lett. A*, 299:217–220, 2002.
- [3] Ariel Fernández and L. Ridgway Scott. Dehydron: a structurally encoded signal for protein interaction. *Biophysical Journal*, 85:1914–1928, 2003.
- [4] U. Kaatze, R. Behrends, and R. Pottel. Hydrogen network fluctuations and dielectric spectrometry of liquids. *Journal of Non-Crystalline Solids*, 305(1):19–28, 2002.
- [5] A. A. Kornyshev and A. Nitzan. Effect of overscreening on the localization of hydrated electrons. *Zeitschrift für Physikalische Chemie*, 215(6):701–715, 2001.
- [6] S. Leikin and A. A. Kornyshev. Theory of hydration forces. nonlocal electrostatic interaction of neutral surfaces. *J. Chem. Phys.*, 92(6):6890–6898, 1990.

- [7] Michael W. Mahoney and William L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112:8910–8922, 2000.
- [8] L. Ridgway Scott, Mercedes Boland, Kristina Rogale, and Ariel Fernández. Continuum equations for dielectric response to macro-molecular assemblies at the nano scale. *Journal of Physics A: Math. Gen.*, 37:9791–9803, 2004.
- [9] Min-Yi Shen and Karl F. Freed. Long Time Dynamics of Met-Enkephalin: Comparison of Explicit and Implicit Solvent Models. *Biophysical Journal*, 82(4):1791–1808, 2002.
- [10] Stanislaw T. Wlodek, Tongye Shen, and J.A.McCammon. Electrostatic steering of substrate to acetylcholinesterase: Analysis of field fluctuations. *Biopolymers*, 53(3):265–271, 2000.