

# Digital biology: 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.

# 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)

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

Hydrophobic effect is considered a dominant effect in protein-ligand association but is non-specific (analog) in behavior.

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

What makes proteins interact in a repeatable way?

## Our thesis

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

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.**

In particular, we review some recent results regarding protein-protein interaction that are based on novel insights about hydrophobic effects. We discuss how these can be used to understand signalling using proteins.

## 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

aardvarkateata vistically academician accelerative  
acetyl glycine achievement acidimetrically acridity  
actress adamant adhesiveness administratively admit  
afflictive after dinner agrypnia aimlessness airlift

and DNA sequences

act catatactagagtagtagacttatatactagagcattacttagat

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 **new 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

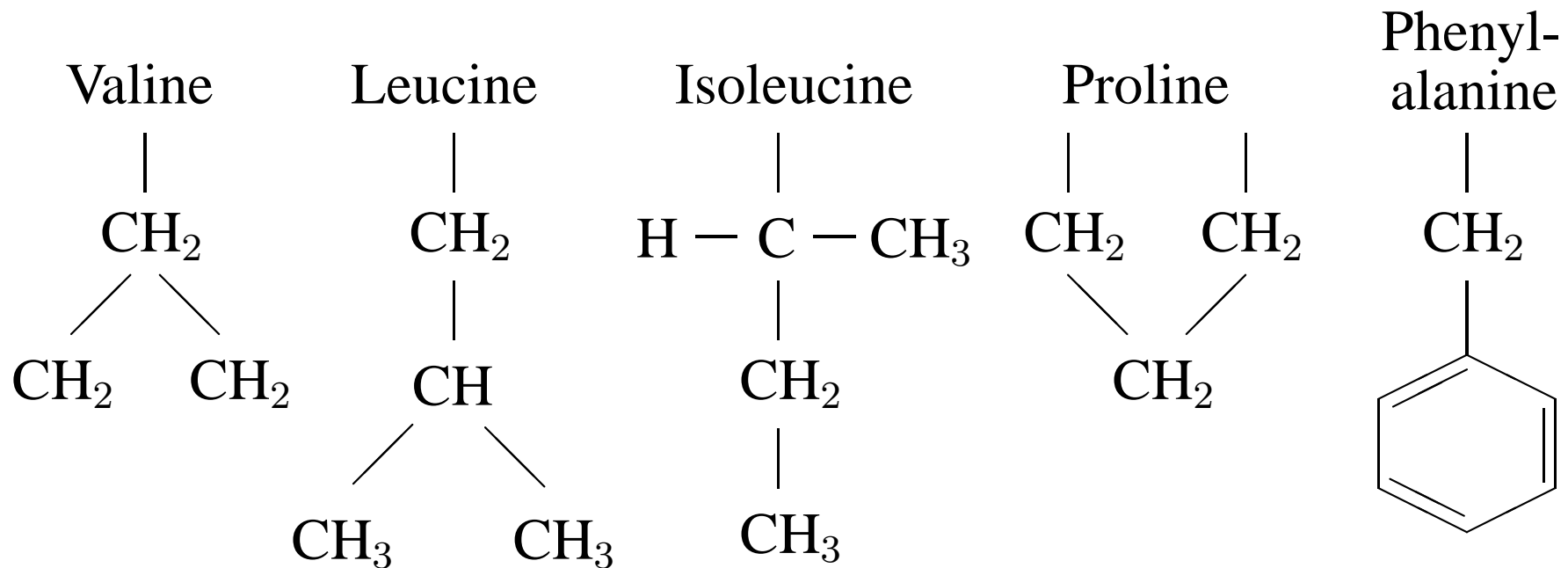
# 1 Tutorial on hydrophobic wrapping

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

## 1.1 Amino acid side chains have different properties

Carbonaceous groups on certain side chains are hydrophobic:



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

## 1.2 Tutorial on hydrophobicity

Carbonaceous groups (CH, CH<sub>2</sub>, CH<sub>3</sub>) are hydrophobic because

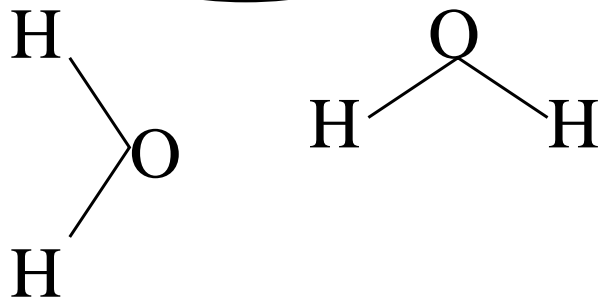
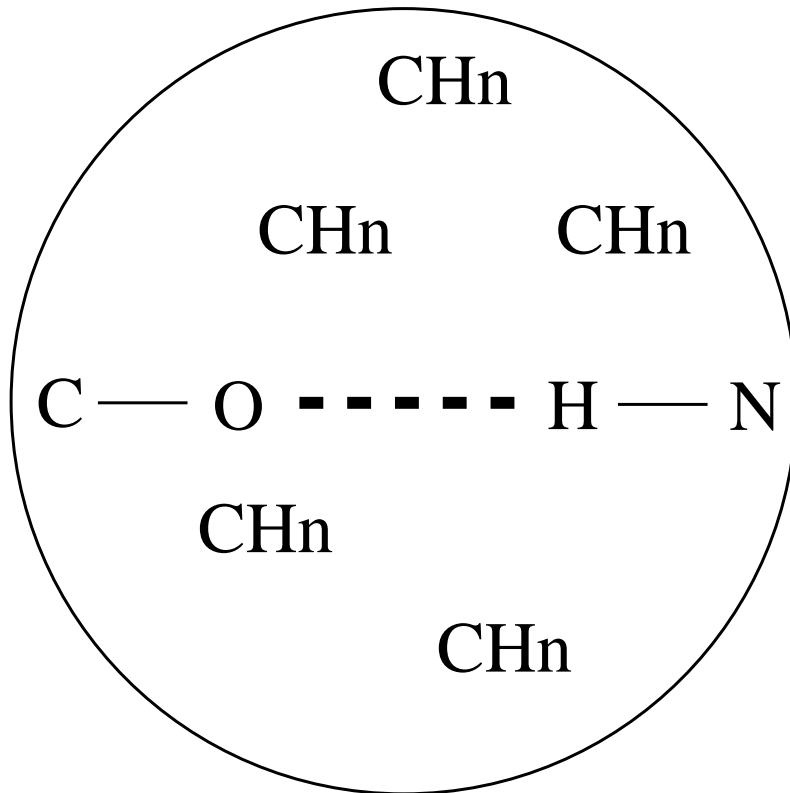
- they are non-polar and thus do not attract water strongly
- they are polarizable and thus damp nearby water fluctuations

## 1.3 Tutorial on dielectrics

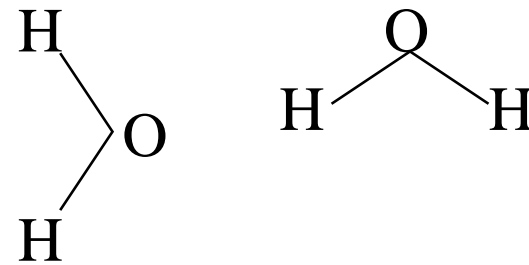
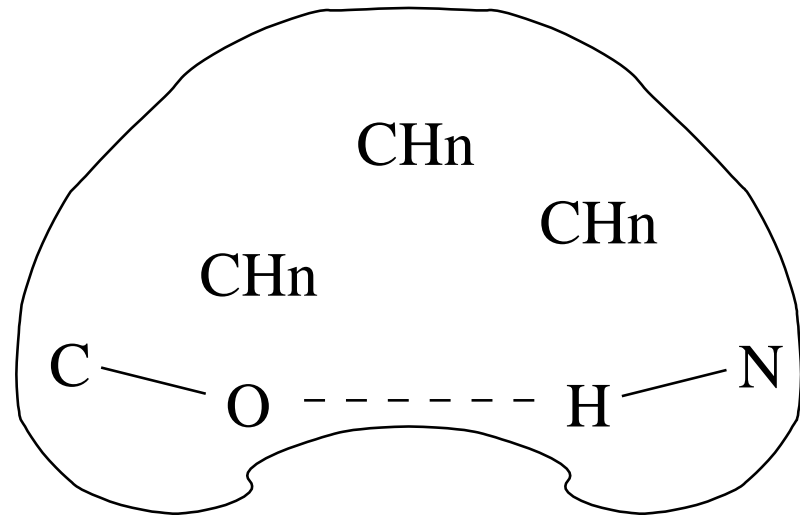
Water removal reduces the dielectric effect and makes electronic bonds stronger.

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

## 1.4 Wrapping protects hydrogen bond from water



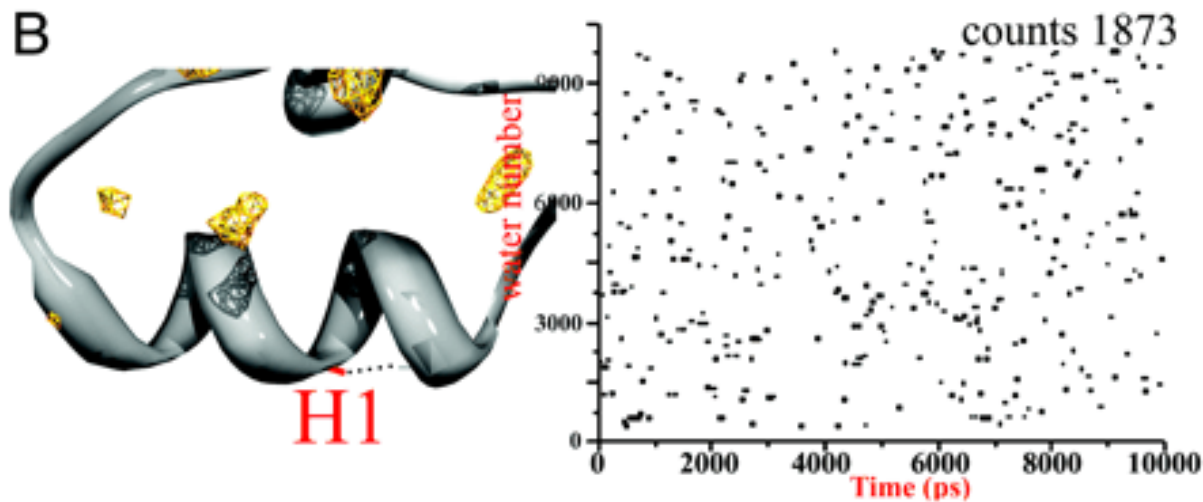
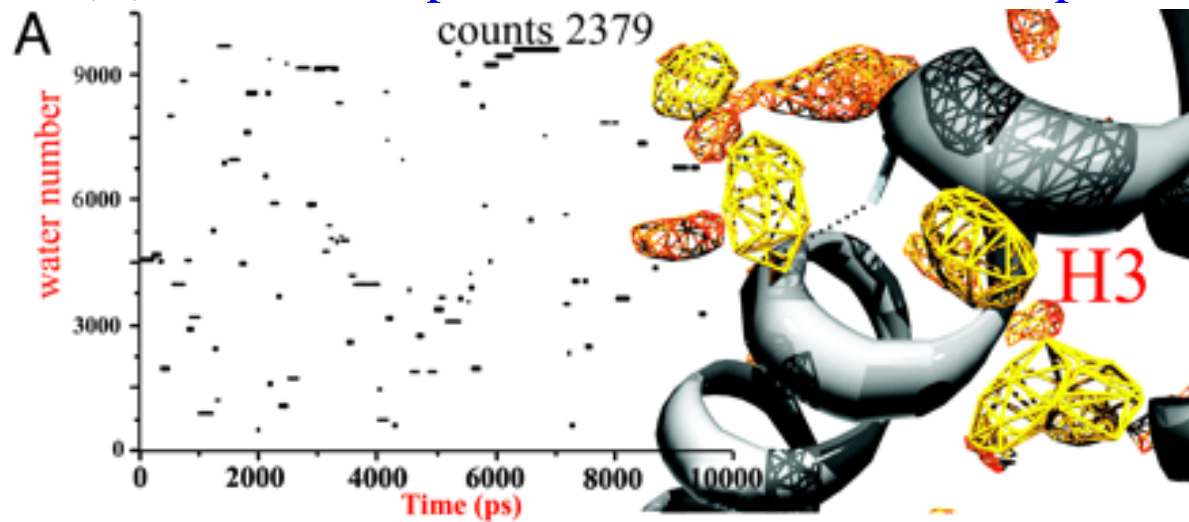
Well wrapped hydrogen bond



Underwrapped hydrogen bond

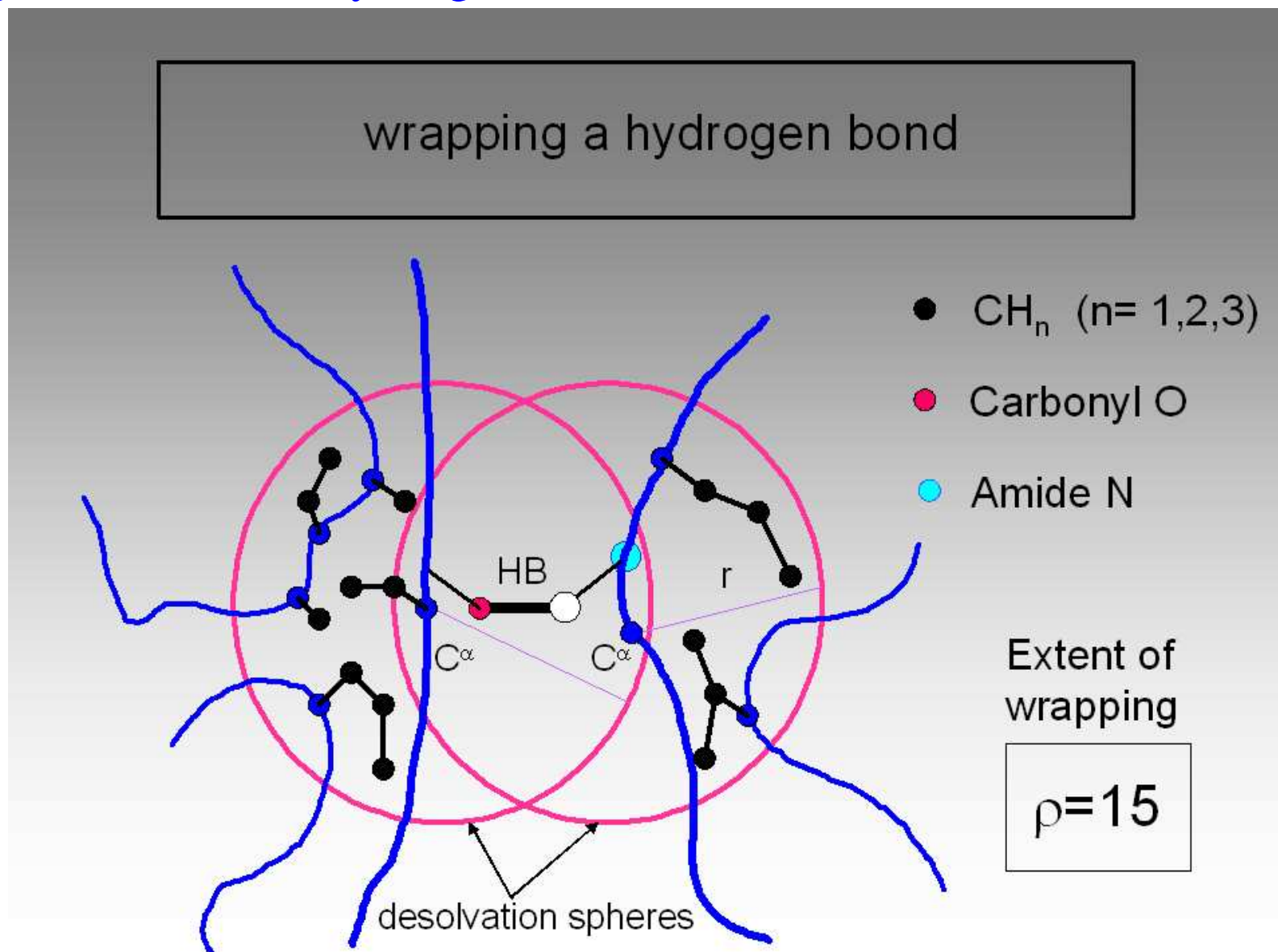
## 1.5 Extent of wrapping changes nature of hydrogen bond

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

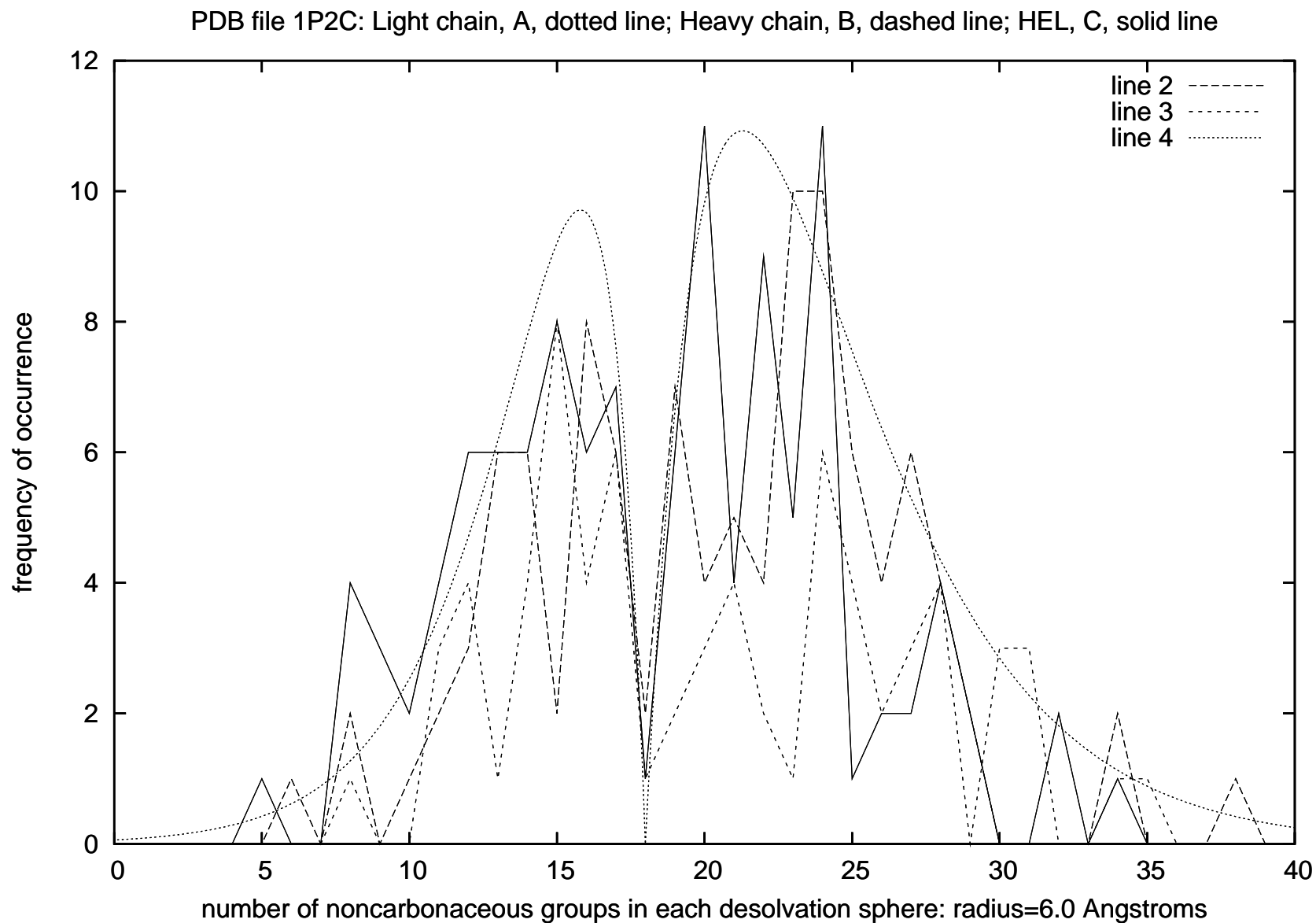


From De Simone, et al., PNAS 102 no 21 7535-7540 (2005)

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



# Distribution of wrapping for an antibody complex.



## 1.6 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.

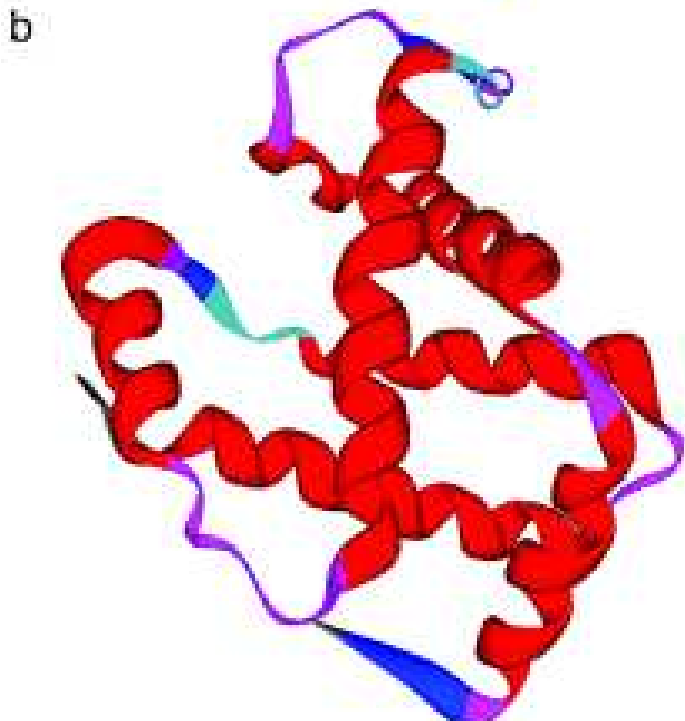
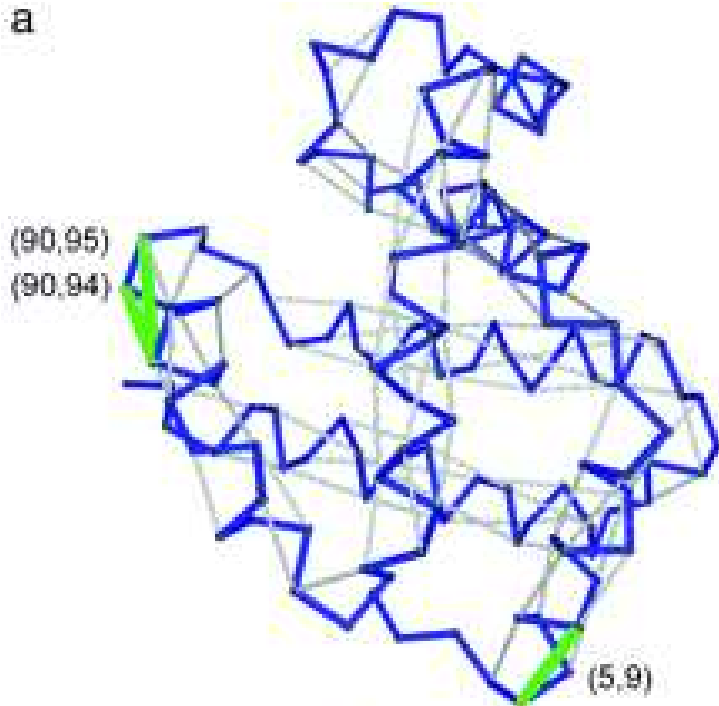
We call such under-wrapped hydrogen bonds

**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

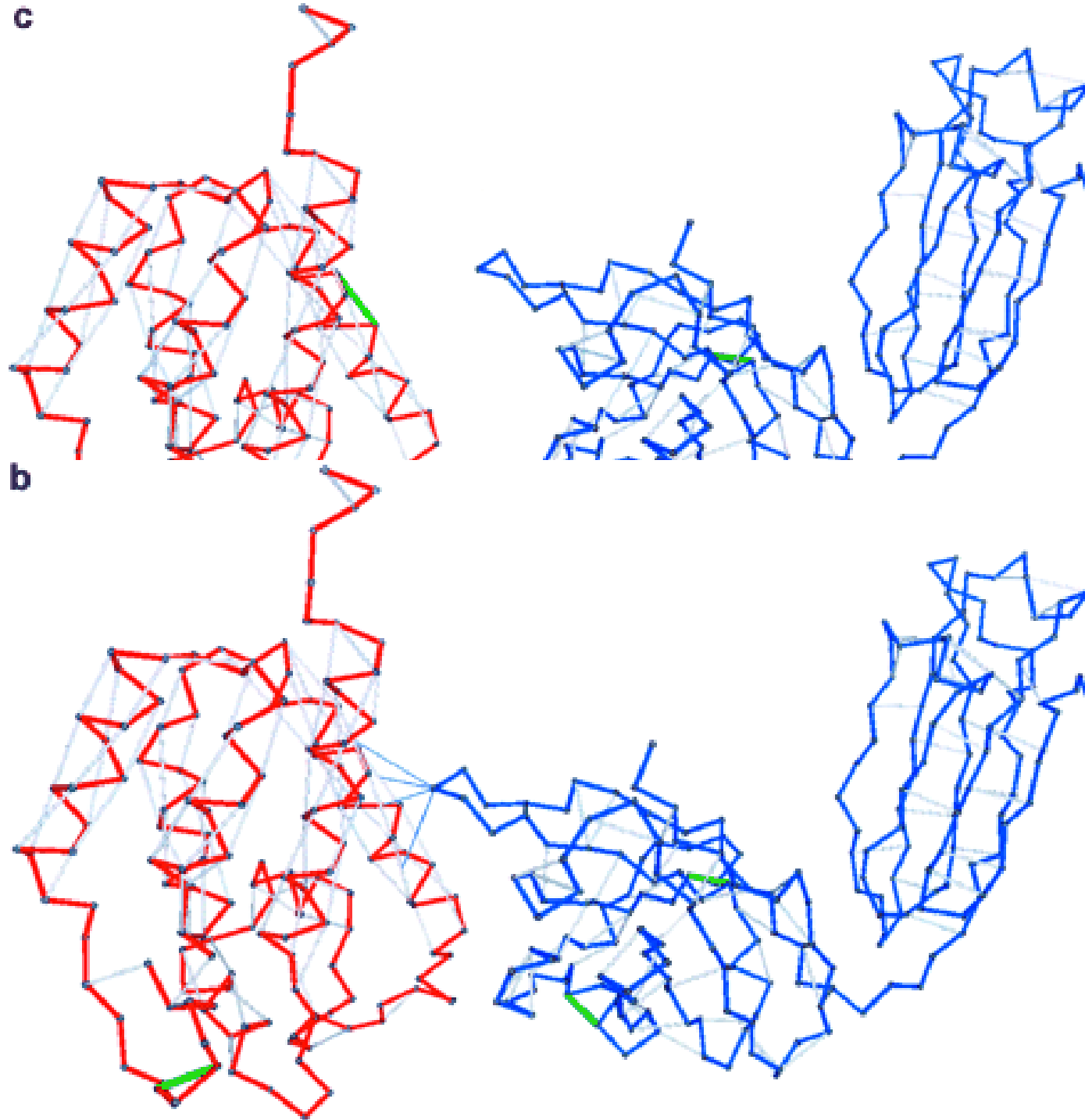
100: 6446-6451 (2003) Ariel Fernandez,  
Jozsef Kardos, L. Ridgway Scott, Yuji Goto,  
and R. Stephen Berry. Structural defects and  
the diagnosis of amyloidogenic propensity.

Well-wrapped

hydrogen bonds are

grey, and dehydrons are green.

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

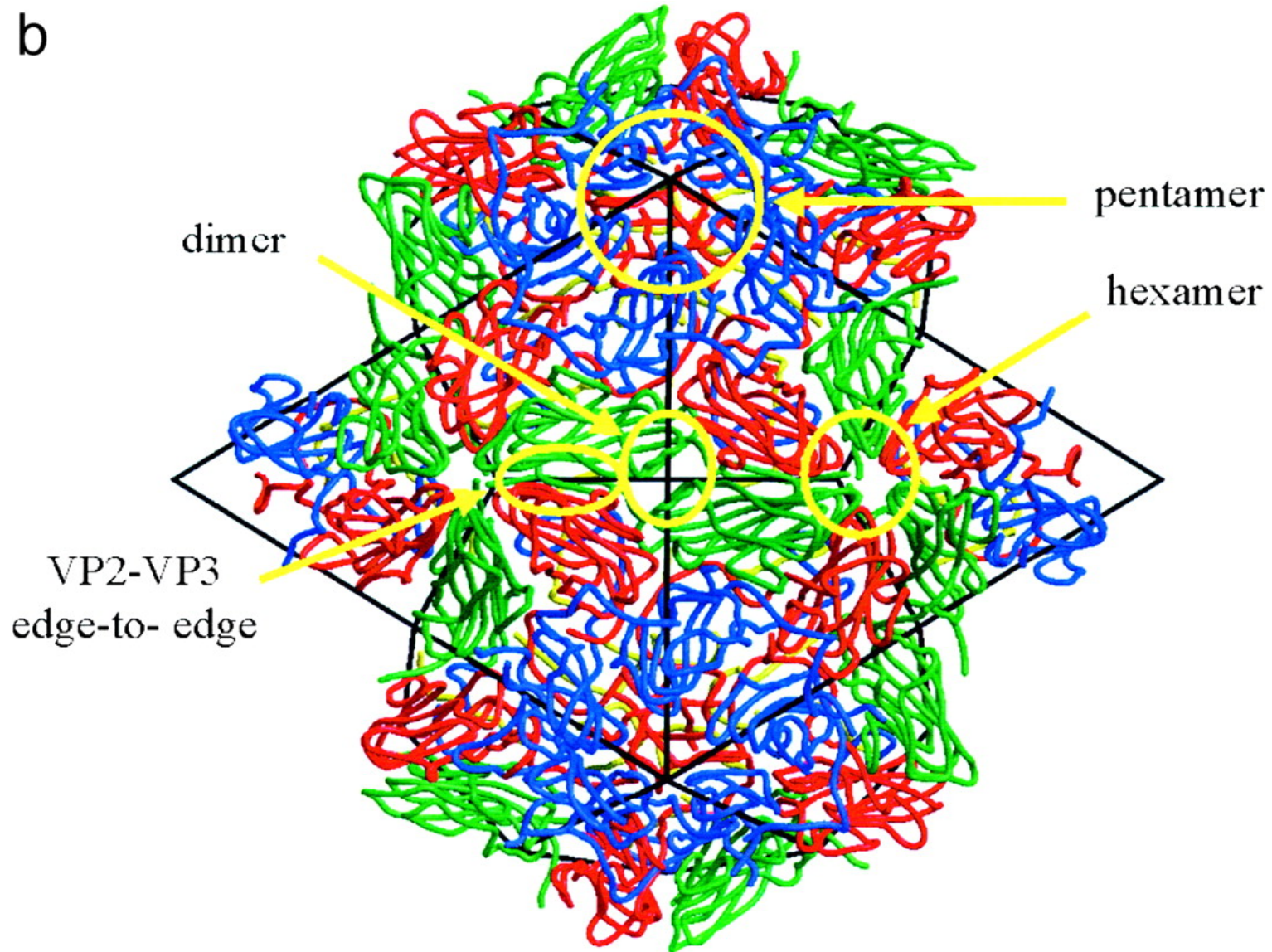


The HIV protease has a dehydron at an antibody binding site.

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

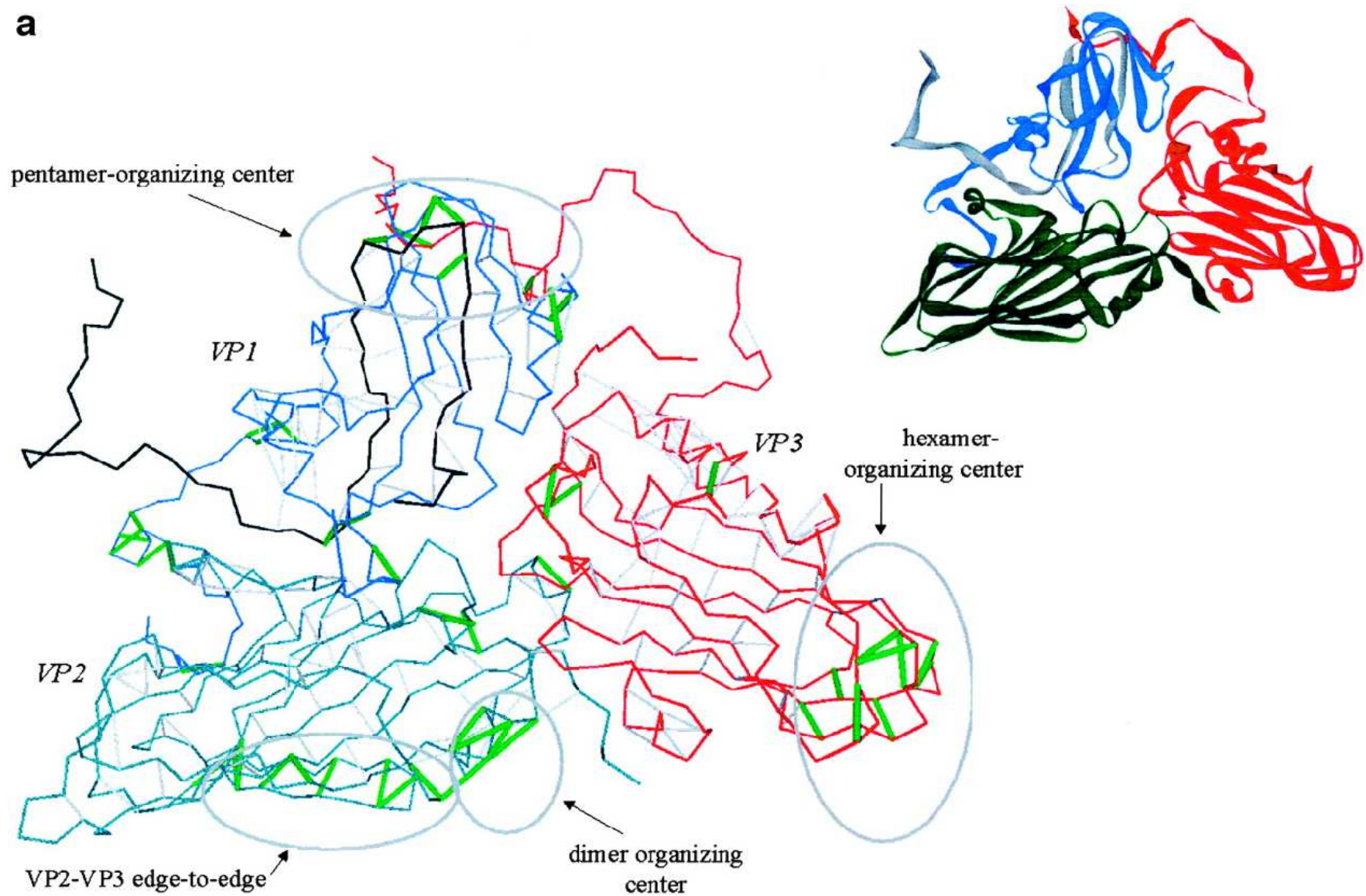
## 1.7 A model for protein-protein interaction

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.

## 1.8 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 [1].

# Talk Outline — **Where we are now**

Physical chemistry provides new lens to look at protein data

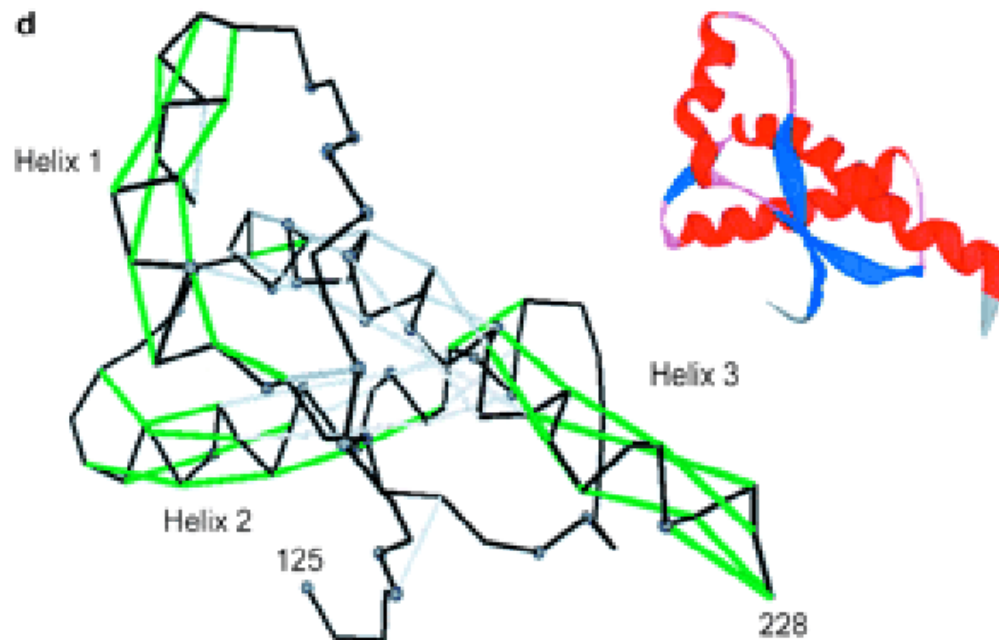
- Tutorial on hydrophobic wrapping
  - hydrophobic protection (desolvation) of hydrogen bonds
  - new motif: dehydron=insufficiently desolvated hydrogen bond
  - dehydrons are involved in protein interaction (they are sticky)
- **Using dehydrons in bioinformatics**
  - the tails of the distribution: **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

## 1.9 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.

## 2 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.

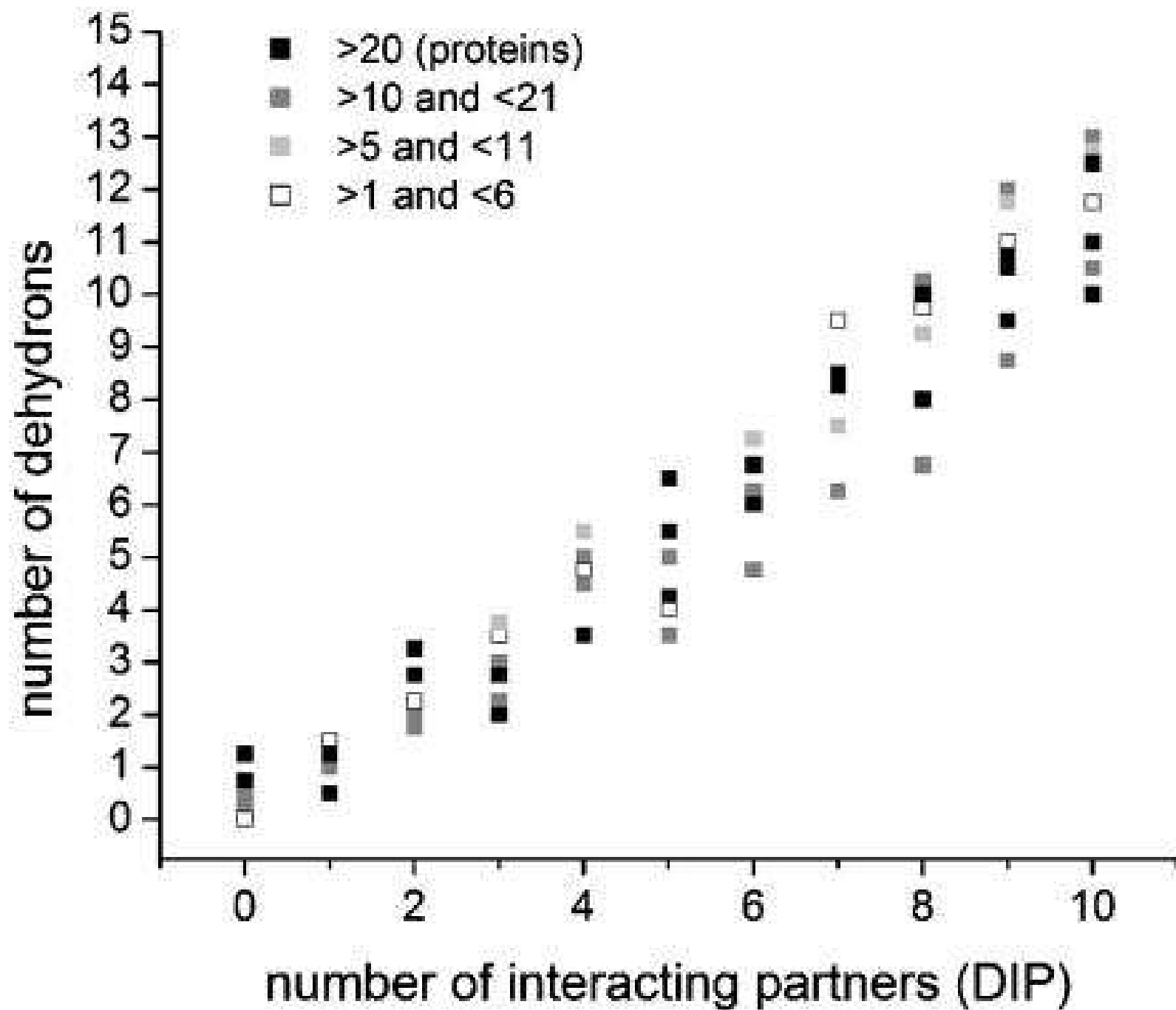
The DIP database collects information about protein interactions, based on individual protein domains: can measure interactivity of different regions of a given protein.

**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.

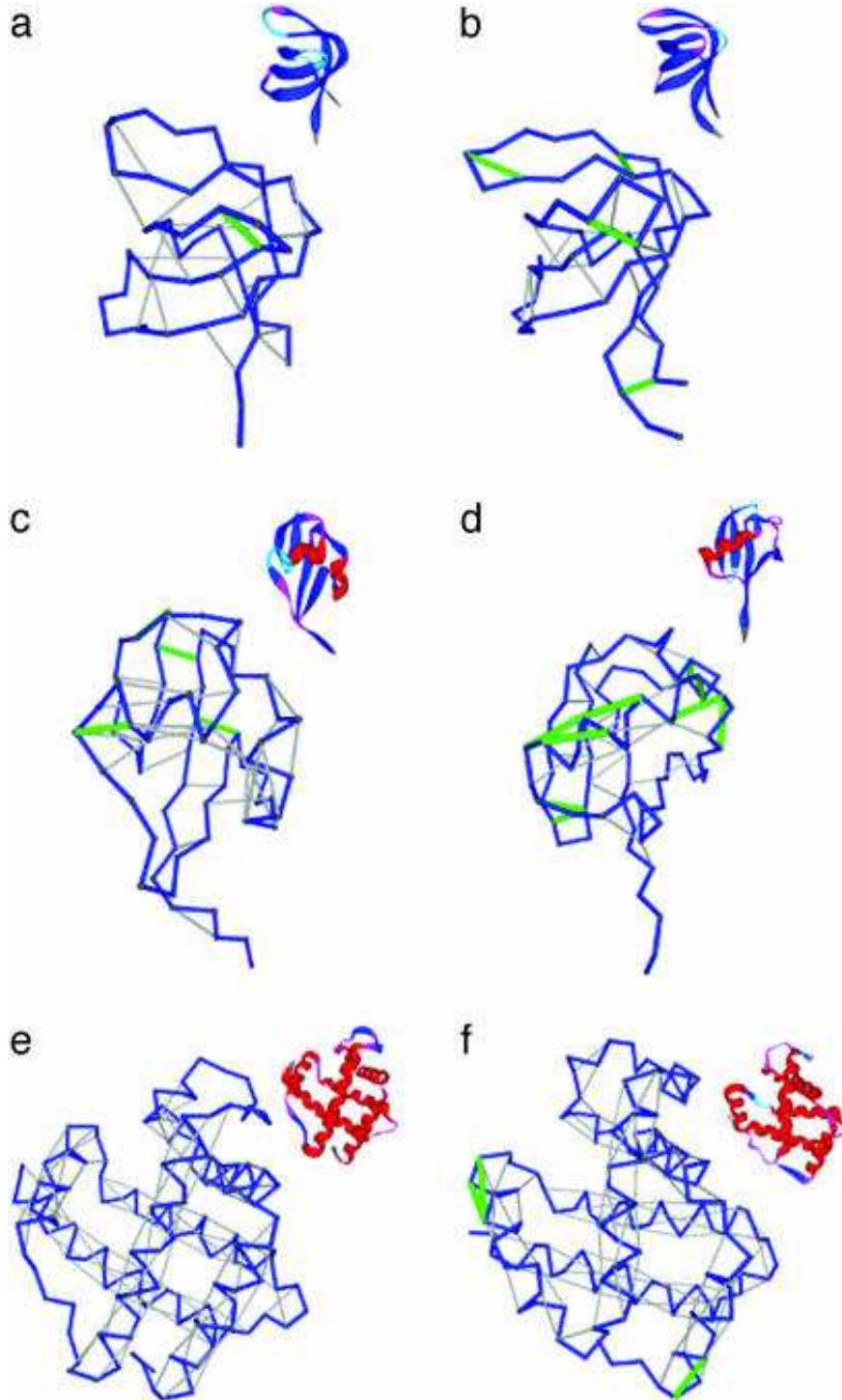
Ariel Fernández, L. R. Scott and R. Steve Berry



## 2.1 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).

## 2.2 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.

## References

- [1] Ariel Fernández and L. Ridgway Scott. Dehydron: a structurally encoded signal for protein interaction. *Biophysical Journal*, 85:1914–1928, 2003.