50 Years of Scientific Computing: from Floating Point to Integers

L. Ridgway Scott

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

50 years ago....

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} \cdots$$

This is a very poor algorithm.

It is all about the algorithms.

48 years ago....

The Perkin-Elmer Thermal Analysis

Suite of Codes, including

Differential scanning calorimeter codes

- Purity detection
- Baseline correction

It is all about code correctness.

Some years later .... Worked with Bill Pritchard and Simon Tavener [5]

Used Crouzeix-Raviart nonconforming piecewise linear element

Coercive for the gradient formulation

Not coercive for the strain formulation

Check your math carefully
before designing the code

Thirty years ago .... Fusions

$$x@p = y@q$$

Implementation: the Golden Rule

- First see if you should send data to someone else; if so, do so
- Then recieve any data that this statement implies

Deterministic: avoids race conditions

L. R. Scott, J. M. Boyle, and B. Bagheri. Distributed data structures for scientific computation. In *Hypercube Multiprocessors 1987, M. T. Heath, ed.*, pages 55–66. Philadelphia: SIAM, 1987.

Planguages

PC and Pfortran were used for many projects

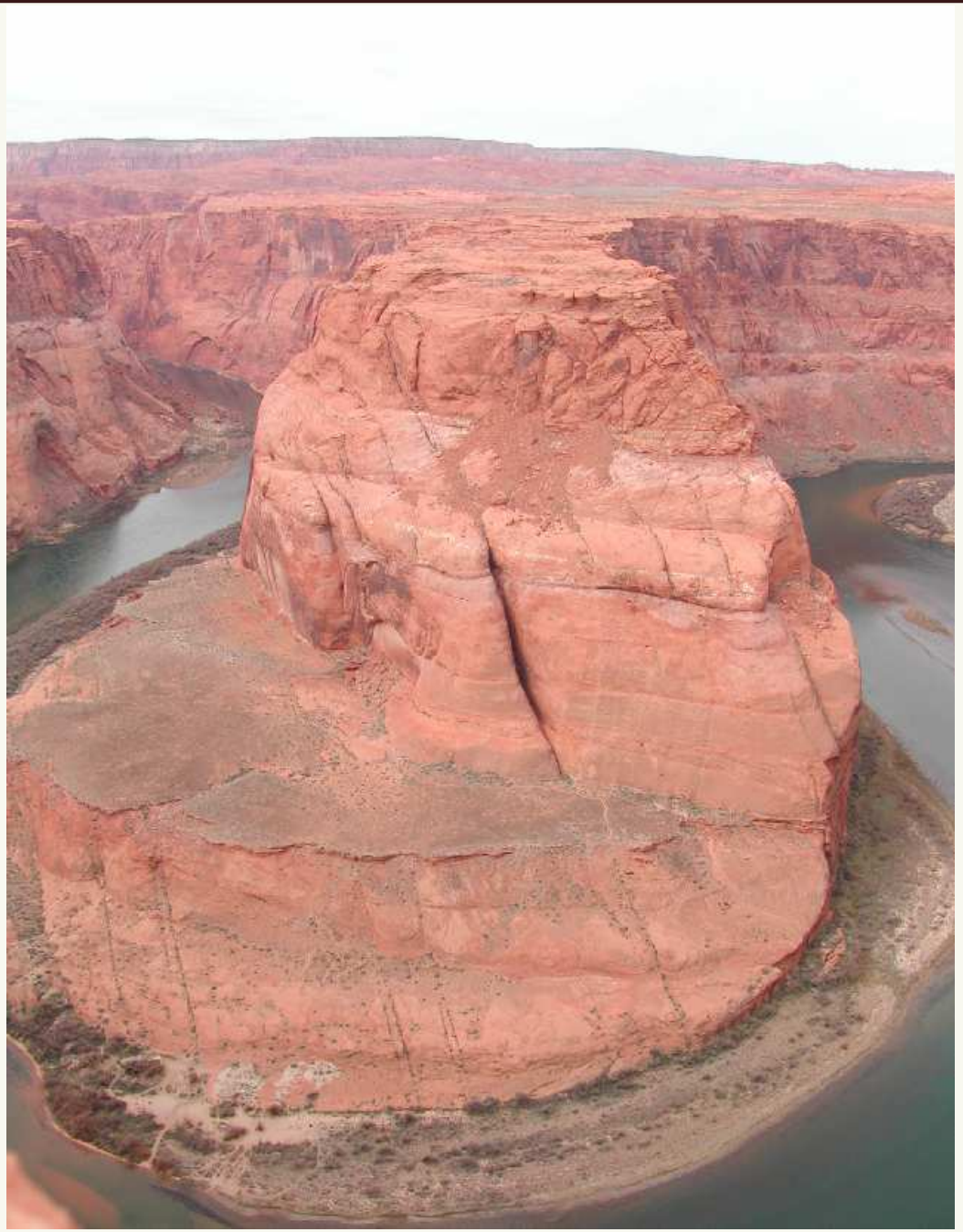
Molecular Dynamics code GROMOS parallelized with Pfortran by Terry Clark

However there were less than 20 parallel lines of code in a 70K deck

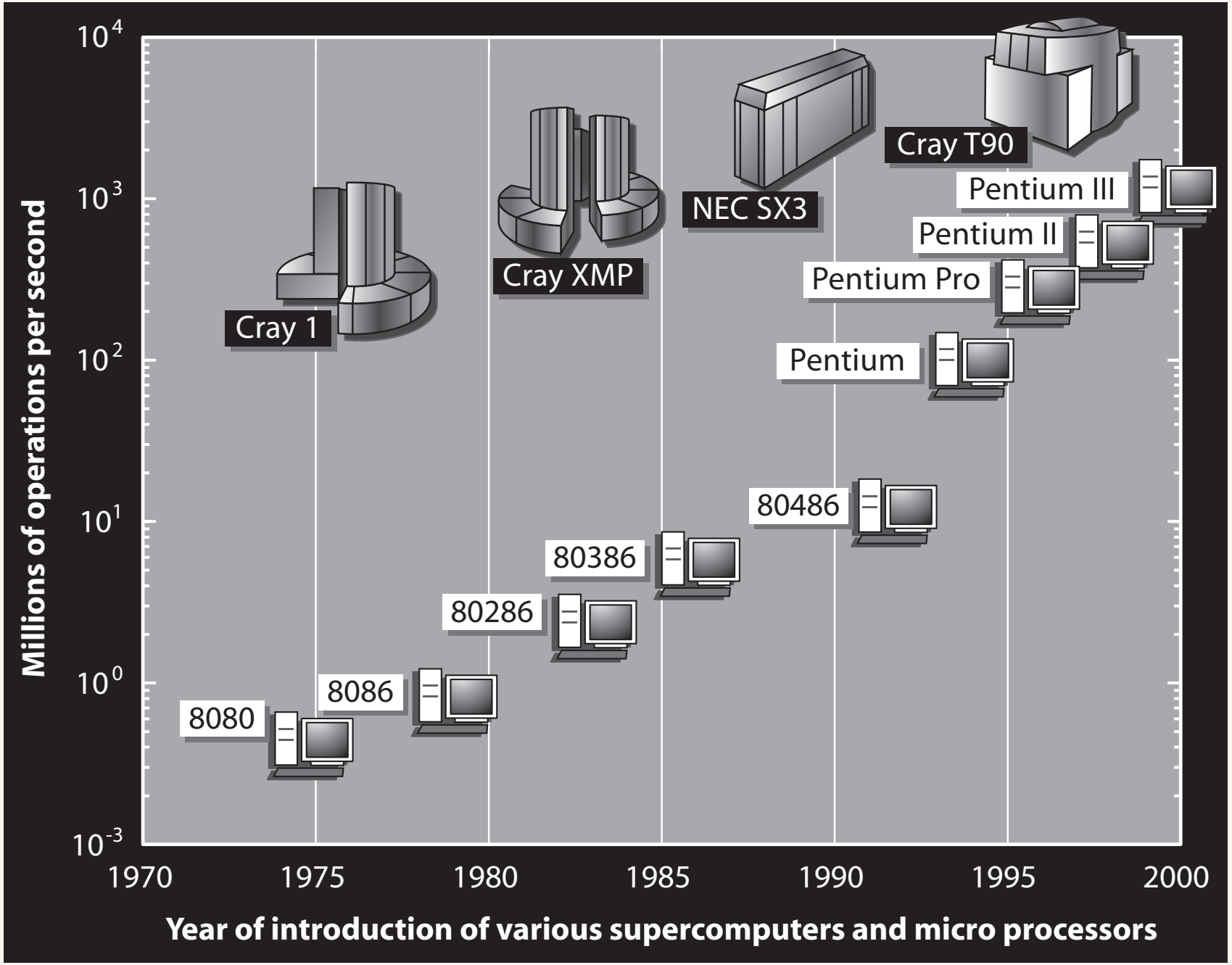
The algorithms are more important than the parallel programming language

But fusions could augment message passing providing deadlock-free code

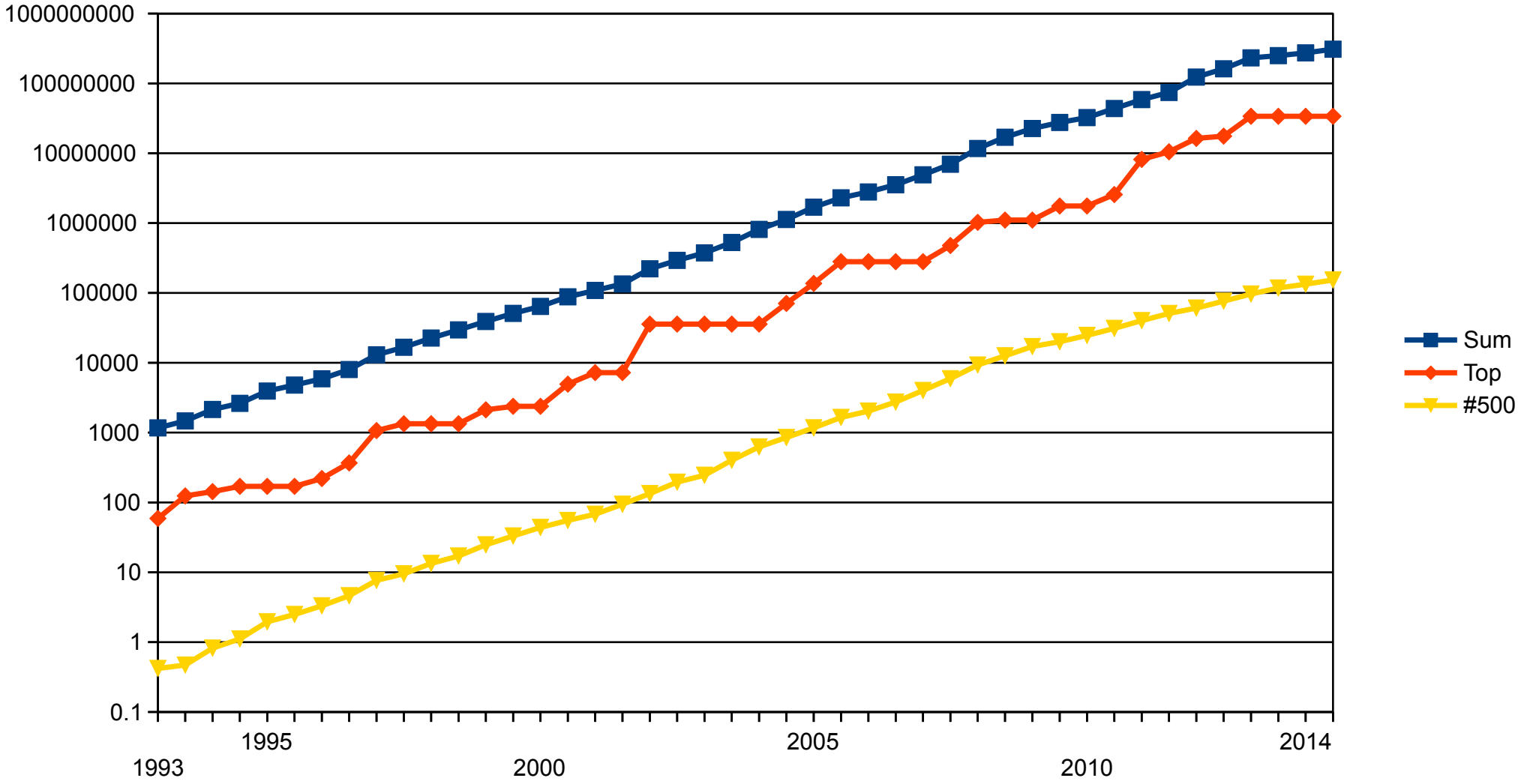
Depiction of a critical section.



Scientific Parallel Computing; Scott, Clark & Bagheri, Princeton 2005



Top 500 supercomputers, Moore's Law



We solved 6-D Schrödinger equations with a finite difference method using tensor contraction techniques

For a given positive integer r, the (simplified) rank-r Tensor-Train format contains all elements $\mathbf{A}$ such that

$$A_{i_1,\ldots,i_d} = A^{(1)}_{i_1} A^{(2)}_{i_2} \cdots A^{(d)}_{i_d},$$

where each $A^{(k)}_{i_k}$ are $r \times r$ matrices.

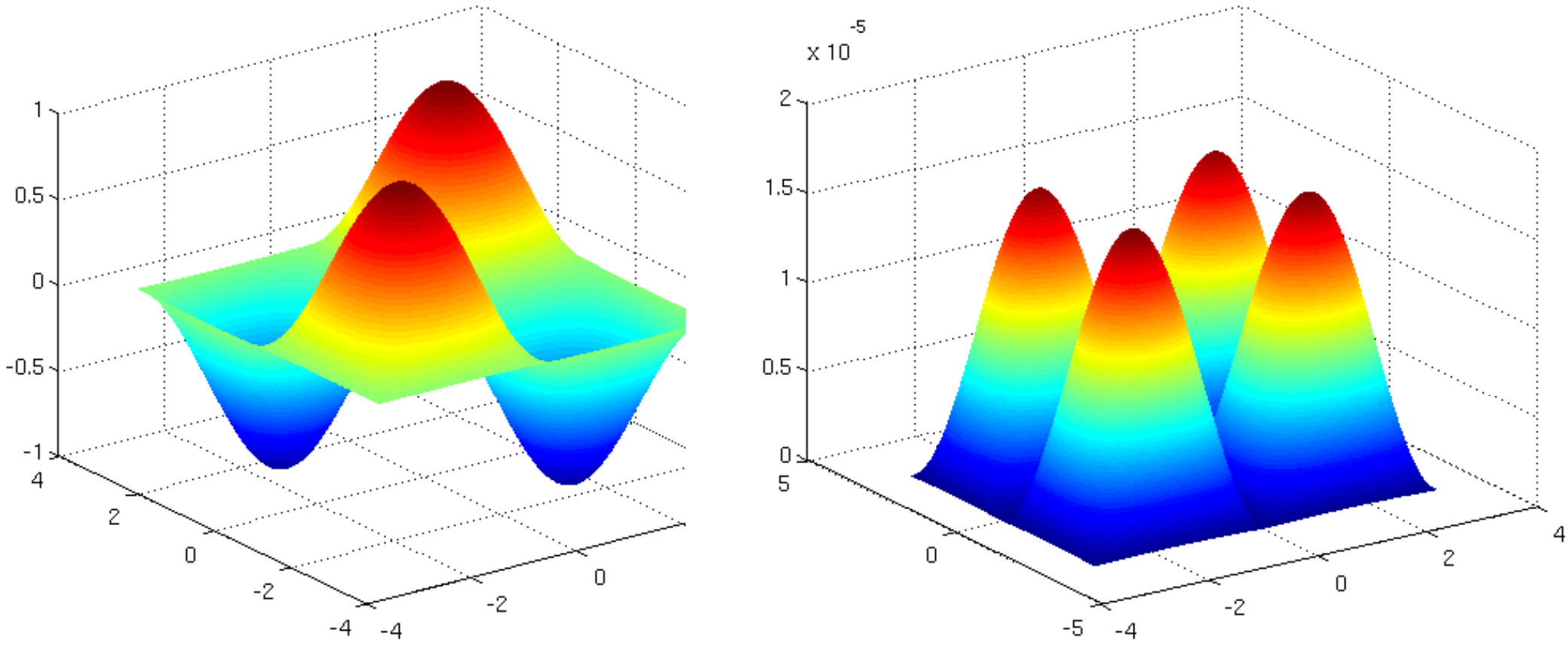
The $A^{(k)}$ can be considered as three-dimensional arrays of size $r^2 \times n$ and are called *cores* of the TT-tensor $\mathbf{A}$.

More generally, r's can be different for each i and

we define an effective rank based on an average.

We used the TT-toolbox 2.2 by I. Oseledets `spring.inm.ras.ru/osel`

Test problem



Test solution Z (left) and pointwise error $Z - Z_h$ (right) evaluated at $\left(\cdot, \frac{\pi}{2}, \frac{\pi}{2}, \cdot, \frac{\pi}{2}, \frac{\pi}{2}\right)$ for $n = 2^{10}$ in each direction of 6 dimensions.

$$(H_0 - \lambda_0)Z = \left(4 - \frac{1}{|x|} - \frac{1}{|y|}\right) \prod_{i=1}^{3} \sin(x_i) \prod_{j=1}^{3} \sin(y_j).$$

Real problem: $n = 2^{12}$

R	$\lambda_R - \lambda_0$	"exact"	rank
$R = 10.0$	$-8.9520 \cdot 10^{-6}$	$-8.754 \cdot 10^{-6}$	27.7
$R = 12.0$	$-2.3335 \cdot 10^{-6}$	$-2.546 \cdot 10^{-6}$	49.5
$R = 14.0$	$-9.5037 \cdot 10^{-7}$		37.4
$R = 16.0$	$-4.1749 \cdot 10^{-7}$		20.8
$R = 18.0$	$-2.0325 \cdot 10^{-7}$		39.2
$R = 20.0$	$-1.0673 \cdot 10^{-7}$		38.7
$R = 22.0$	$-5.9265 \cdot 10^{-8}$		40.4
$R = 24.0$	$-3.5191 \cdot 10^{-8}$		39.7
$R = 26.0$	$-2.1657 \cdot 10^{-8}$		40.7

Number of grid points approaches Avogadro's Number.

Exact values from TABLE II, Relativistic energies of the ground state of the hydrogen molecule, L. Wolniewicz, J. Chem. Phys. 99, 1851 (1993).

First steps towards automation

A finite element code in Ada

- most errors caught at compile time due to type mismatch

fec: C++ library (Babak Bagheri)

- used in theses by several students and post-docs at the U. Houston [2, 3, 4, 6, 7]

L. R. Scott and B. Bagheri. Software environments for the parallel solution of partial differential equations. In Computing Methods in Applied Sciences and Engineering IX, R. Glowinski and A. Lichnewsky, eds., pages 378-392. Philadelphia: SIAM, 1990.

Automatation provides higher-level opportunities for optimization

- Variational formulation provides a language for PDE's
- Finite element simulation can be further automated (fiat, FErari, finat)

Implementations:

- Analysa was written by Babak Bagheri
- FEniCS Project: too many contributors here to list everyone

Models provide abstractions

Two aspects to implementation

- Need to translate correctly to appropriate (e.g., machine) language
- Opportunity to perform optimizations

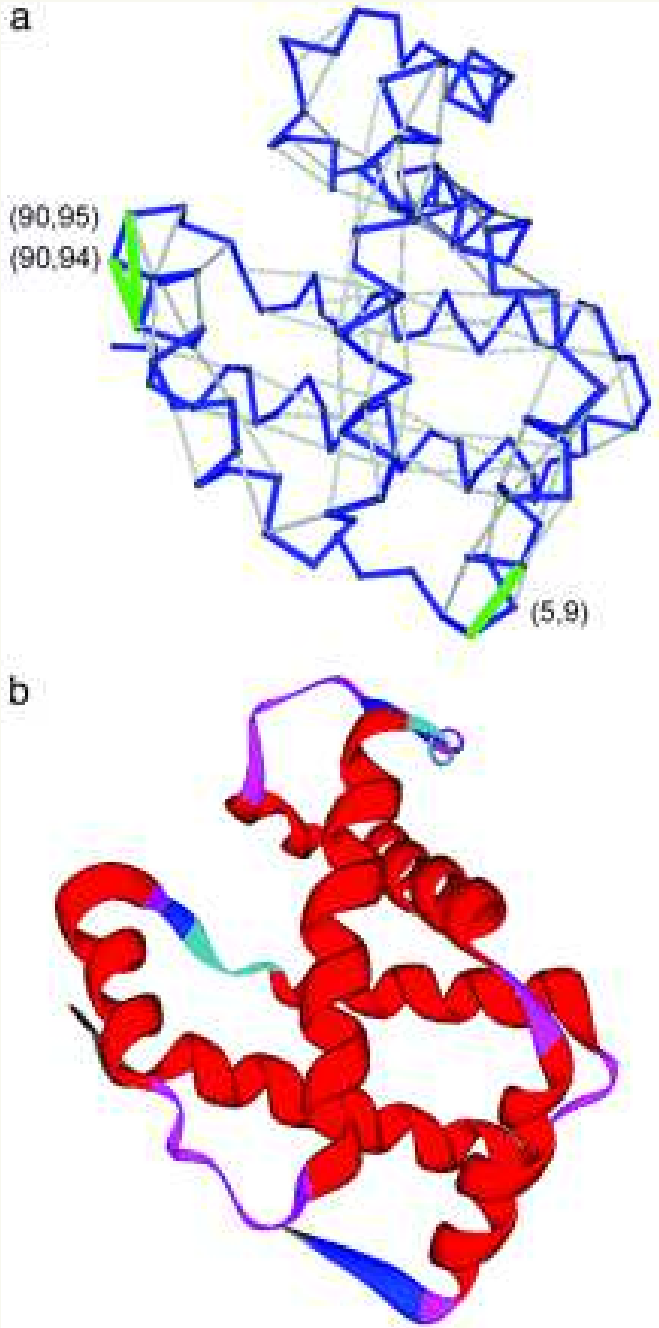
Optimization involves algorithmic challenges

NIC versus DIC

Numeric-intensive versus data intensive computing

Example: extracting information from genome and protein data

Computing with integers is still hard



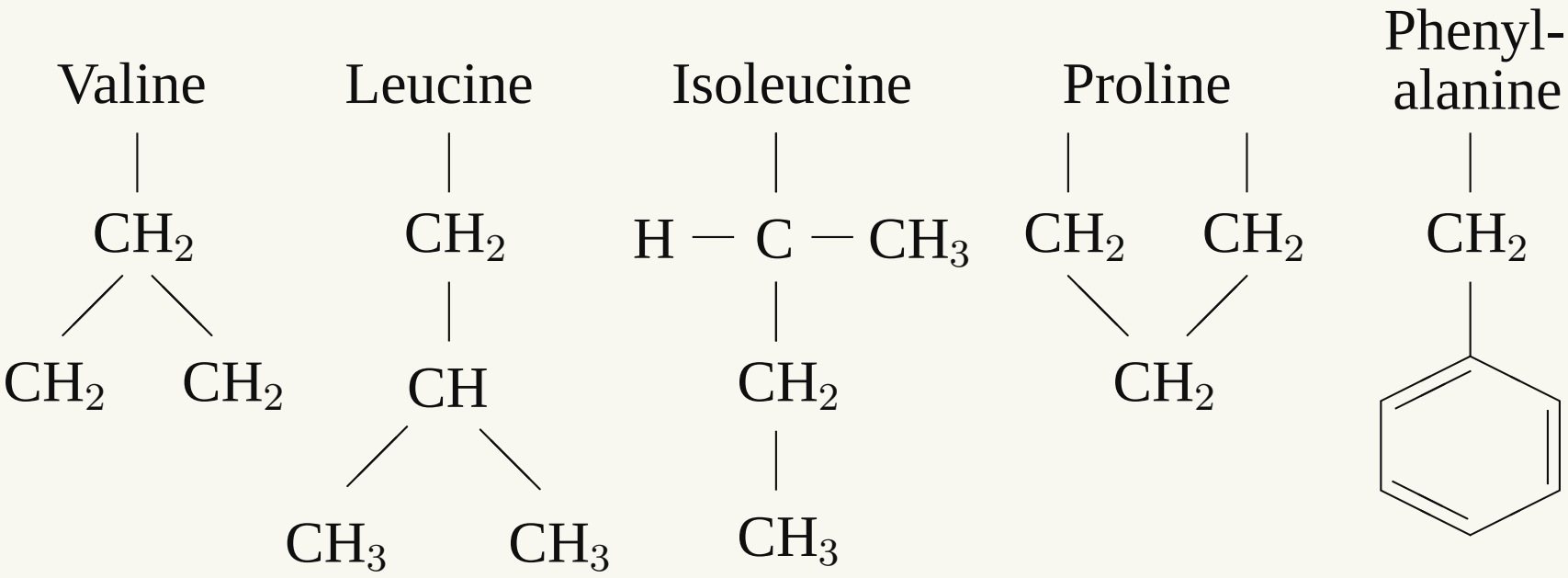
Dehydrons
in human hemoglobin

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 electrostatic environment.

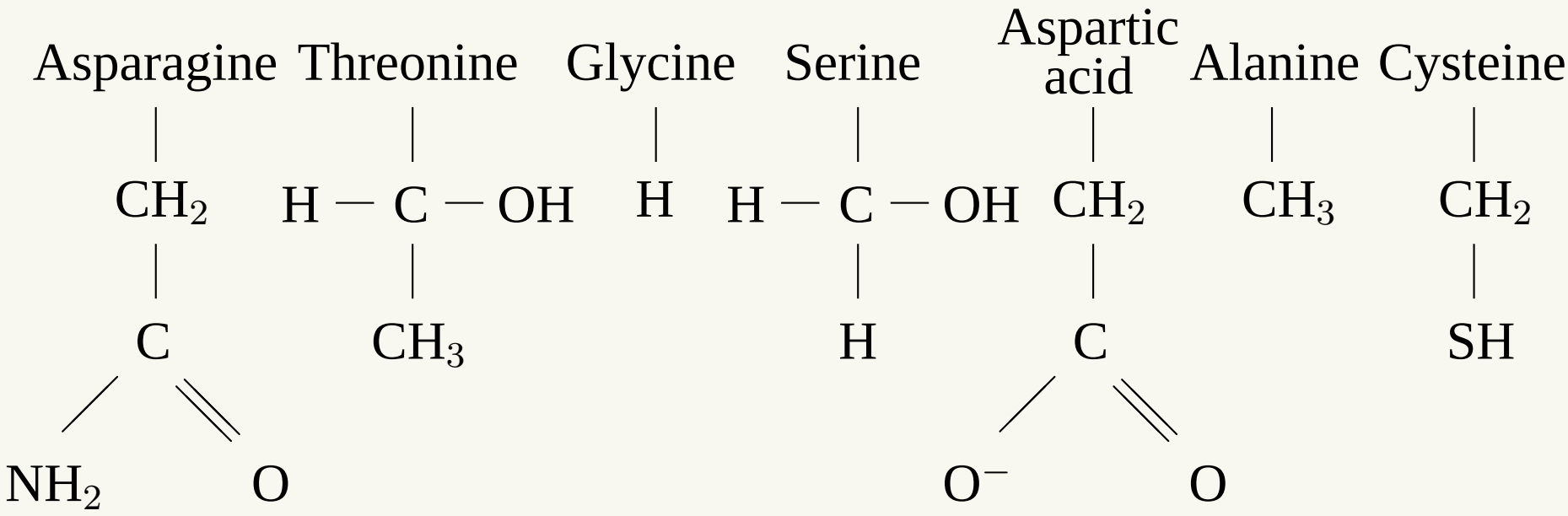
Carbonaceous groups on sidechains are hydrophobic:



Amino acid residues (sidechains only shown) having only carbonaceous groups.

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.

Genetic code

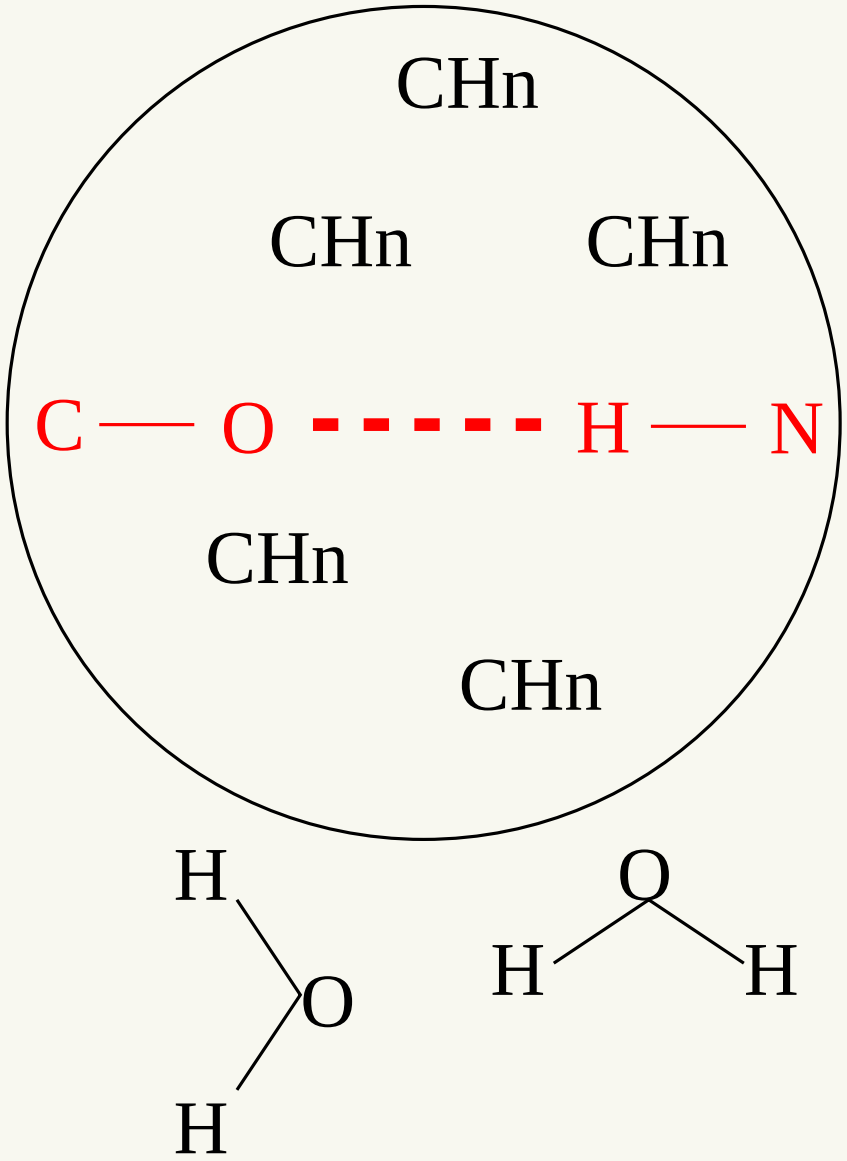
Genetic code minimizes changes of polarity due to single-letter codon mutations, but it facilitates changes in wrapping due to single-letter codon mutations.

Second Position

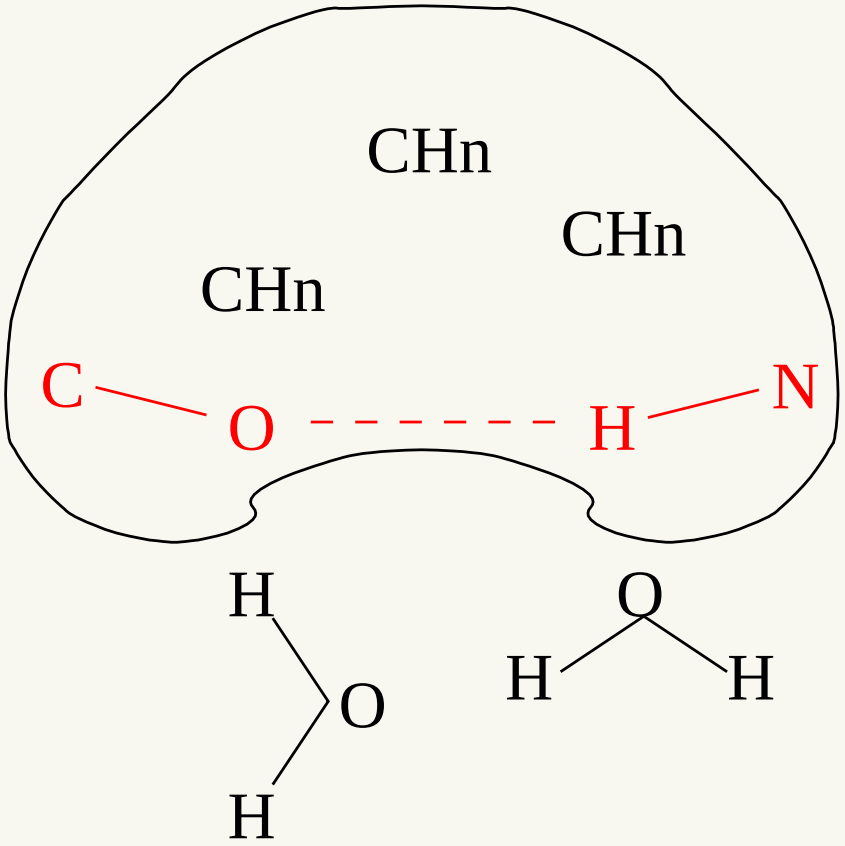
First Position	u	c	a	g	Third Position
u	uuu, uuc: Phe 7 ||	ucu, ucc, uca, ucg: Ser 0 +−	uau, uac: Tyr 6 +−	ugu, ugc: Cys 0 +−	u, c
	uua, uug: Leu 4 ||		uaa: stop; uag: stop	uga: stop; ugg: Trp 7 +−	a, g
c	cuu, cuc, cua, cug: Leu 4 ||	ccu, ccc, cca, ccg: Pro 2 ||	cau, cac: His 1 +−	cgu, cgc, cga, cgg: Arg 2 ++	u, c, a, g
			caa, cag: Gln 2 +−		
a	auu, auc, aua: Ile 4 ||	acu, acc, aca, acg: Thr 1 +−	aau, aac: Asn 1 +−	agu, agc: Ser 0 +−	u, c
	aug: Met 1 +−		aaa, aag: Lys 3 ++	aga, agg: Arg 2 ++	a, g
g	guu, guc, gua, gug: Val 3 ||	gcu, gcc, gca, gcg: Ala 1 ||	gau, gac: Asp 1 −−	ggu, ggc, gga, ggg: Gly 0 +−	u, c, a, g
			gaa, gag: Glu 2 −−		

First digit after residue name is amount of wrapping. Second indicator is polarity; $||$: nonpolar, $+-$: polar, $--$: negatively charged, $++$: positively charged.

Wrapping protects hydrogen bond from water



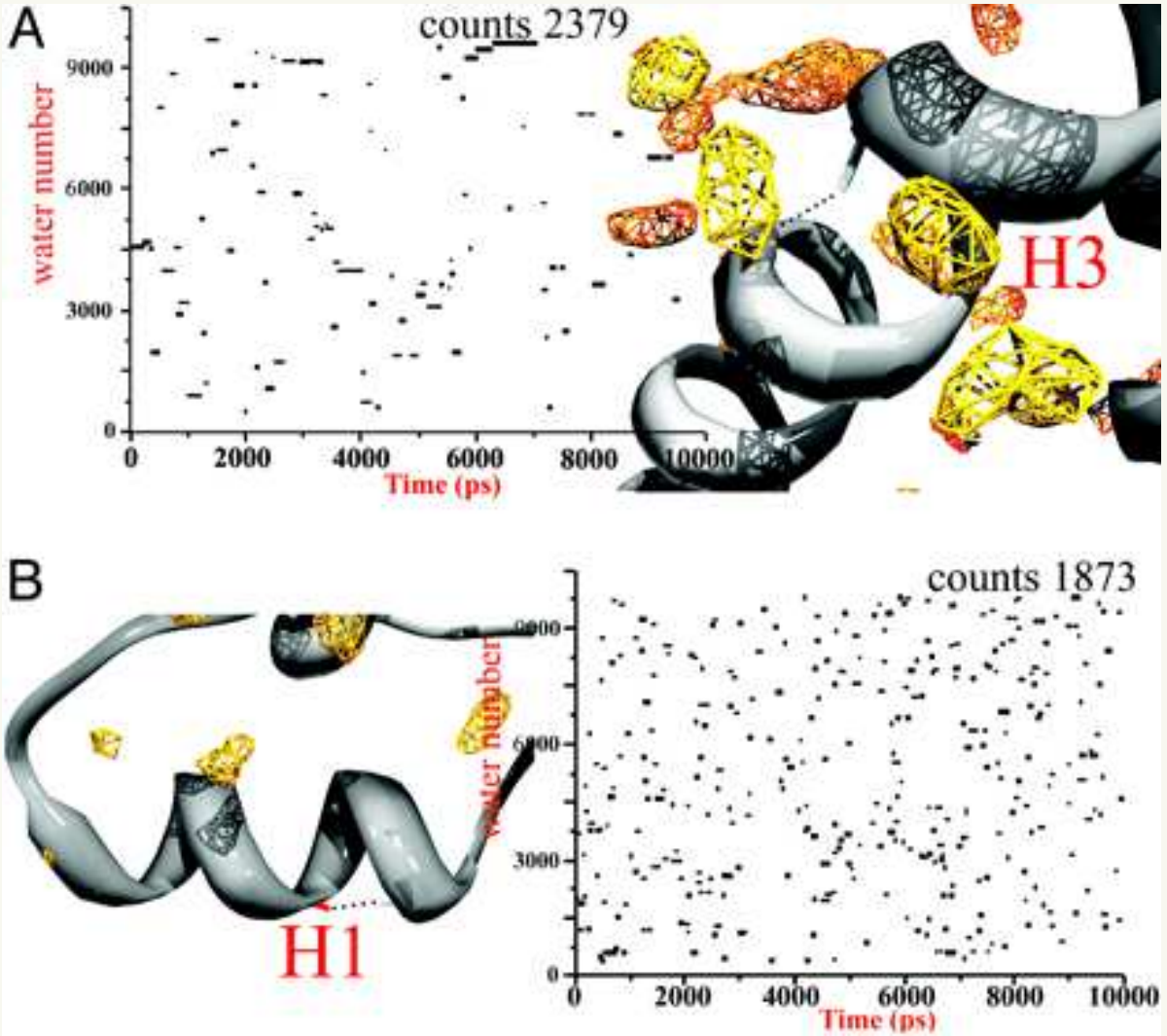
Well wrapped hydrogen bond



Underwrapped hydrogen bond

Extent of wrapping changes nature of hydrogen bond

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



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

Dynamics of hydrogen bonds and wrapping

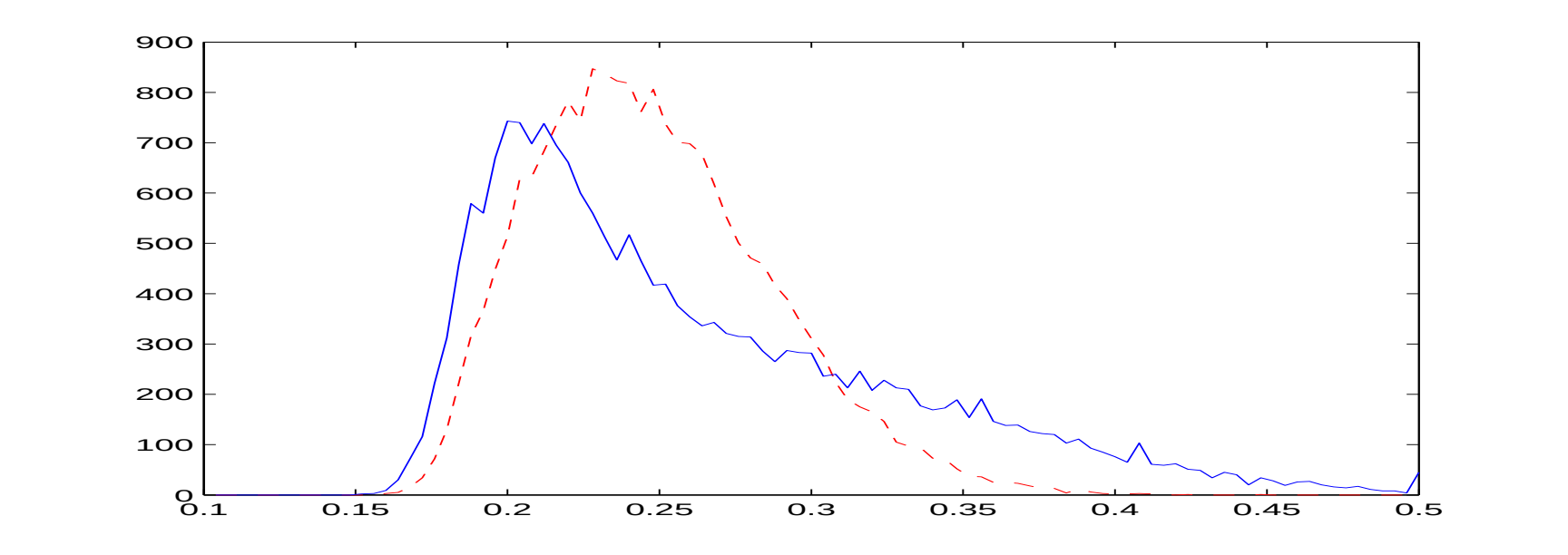
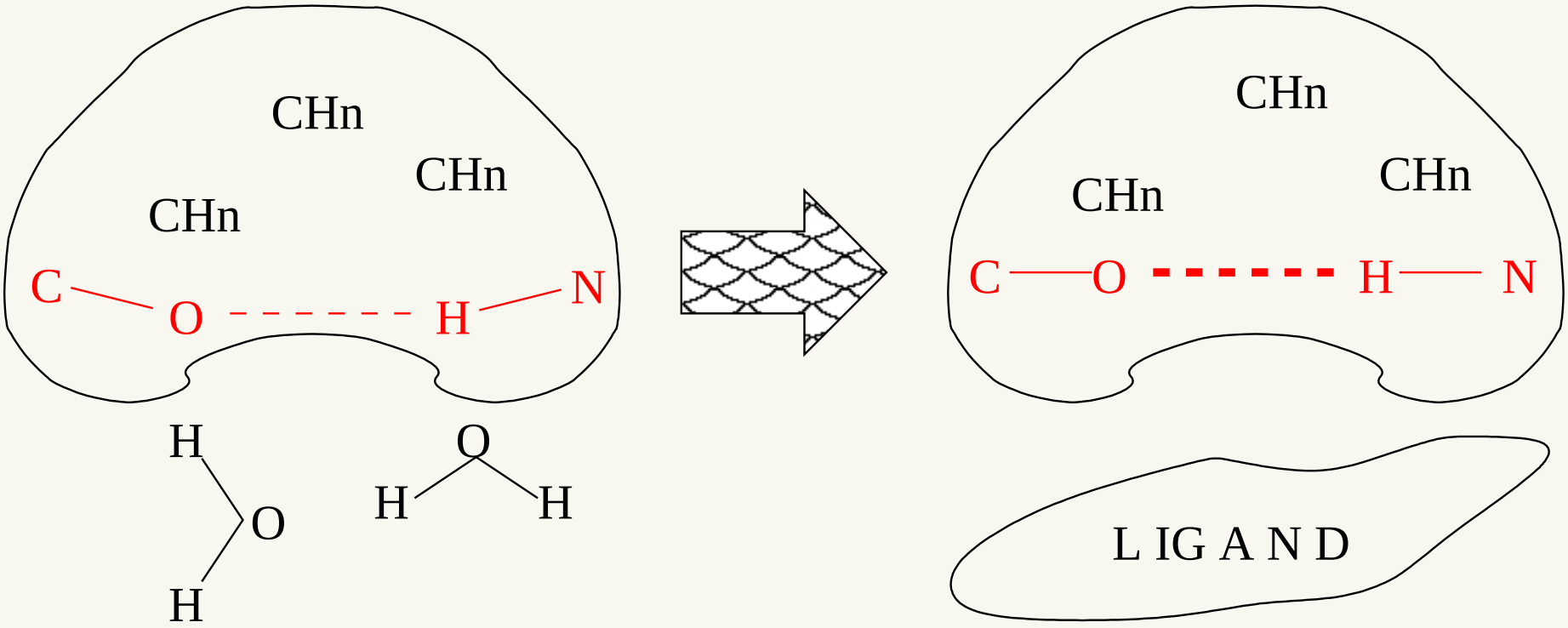


Figure 2: Distribution of bond lengths for two hydrogen bonds formed in a structure of the sheep prion [1]. Horizontal axis measured in nanometers, vertical axis represents numbers of occurrences taken from a simulation with $20,000$ data points with bin widths of 0.1 Ångstrom. Distribution for the well-wrapped hydrogen bond (H3) has smaller mean value but a longer (exponential) tail, whereas distribution for the underwrapped hydrogen bond (H1) has larger mean but Gaussian tail.

Ligand binding removes water

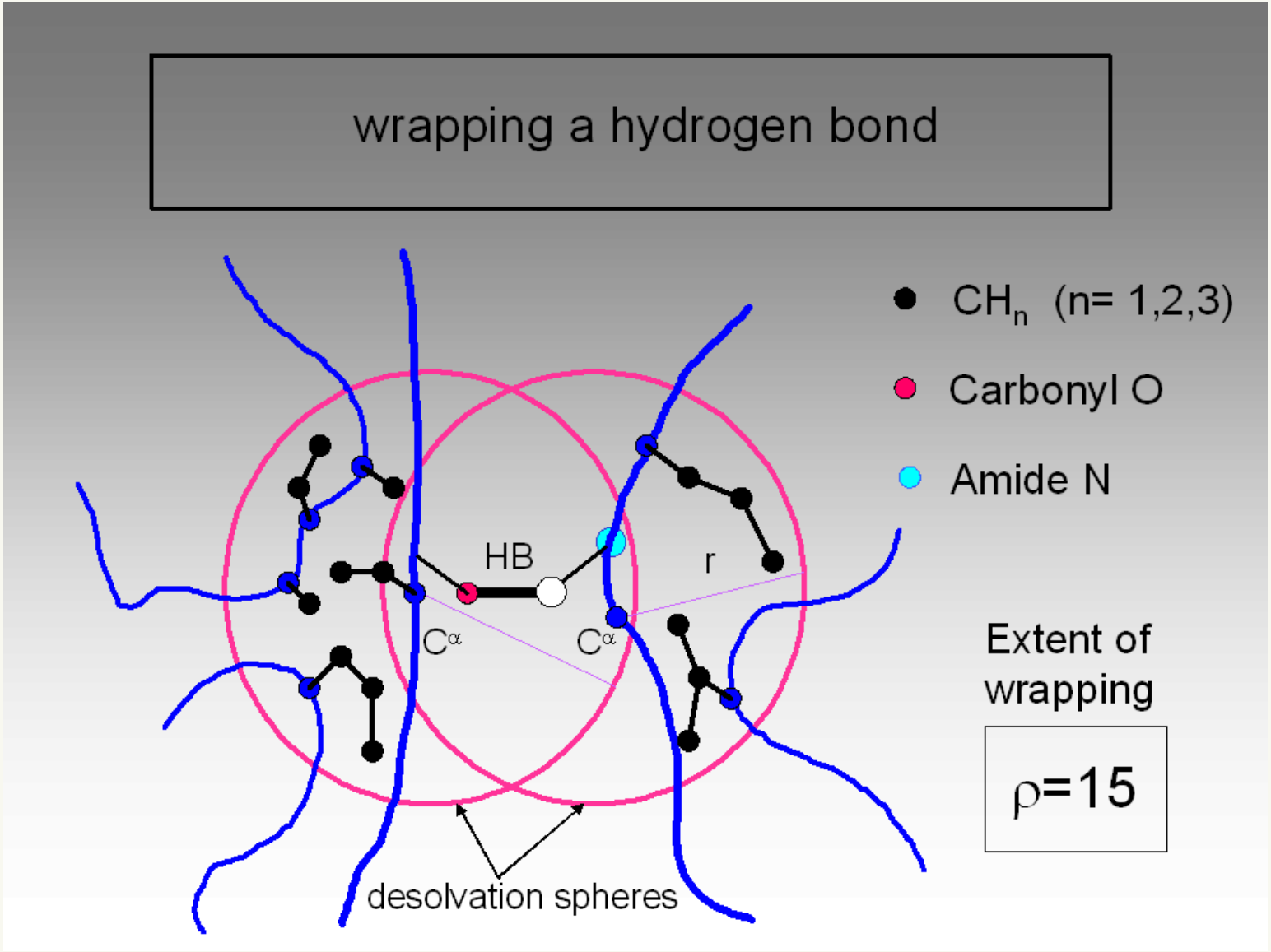


Binding of ligand changes underprotected hydrogen bond (high dielectric) to strong bond (low dielectric)

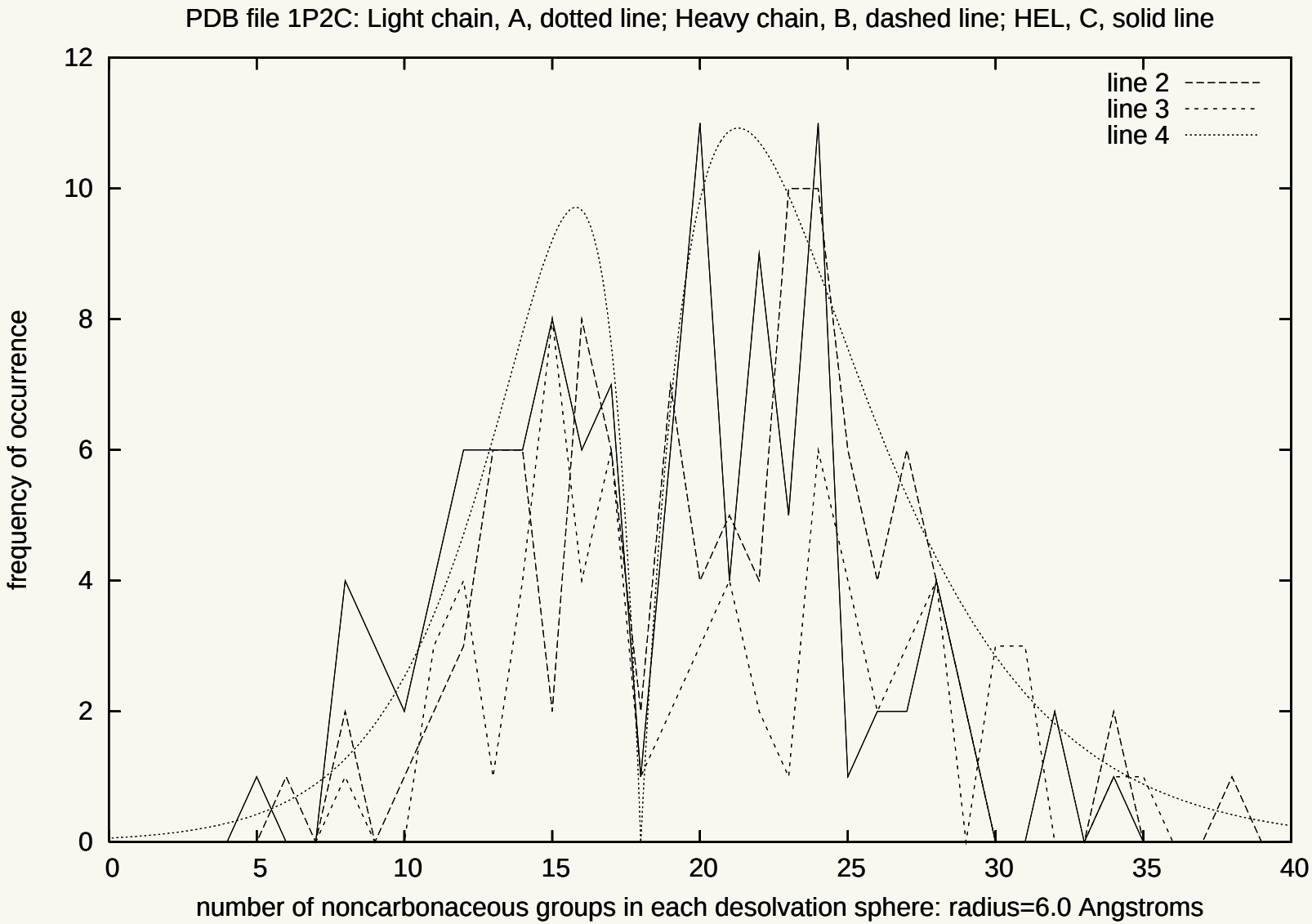
No intermolecular bonds needed!

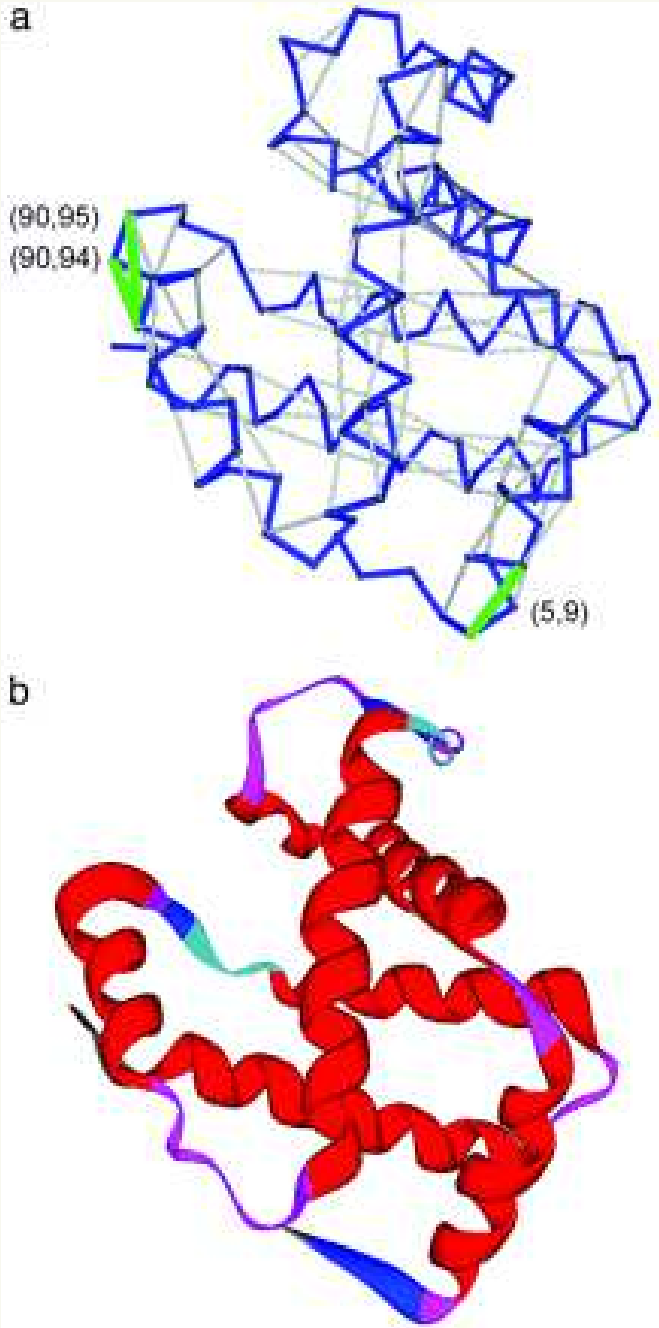
Wrapping made quantitative

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



Distribution of wrapping for an antibody complex.





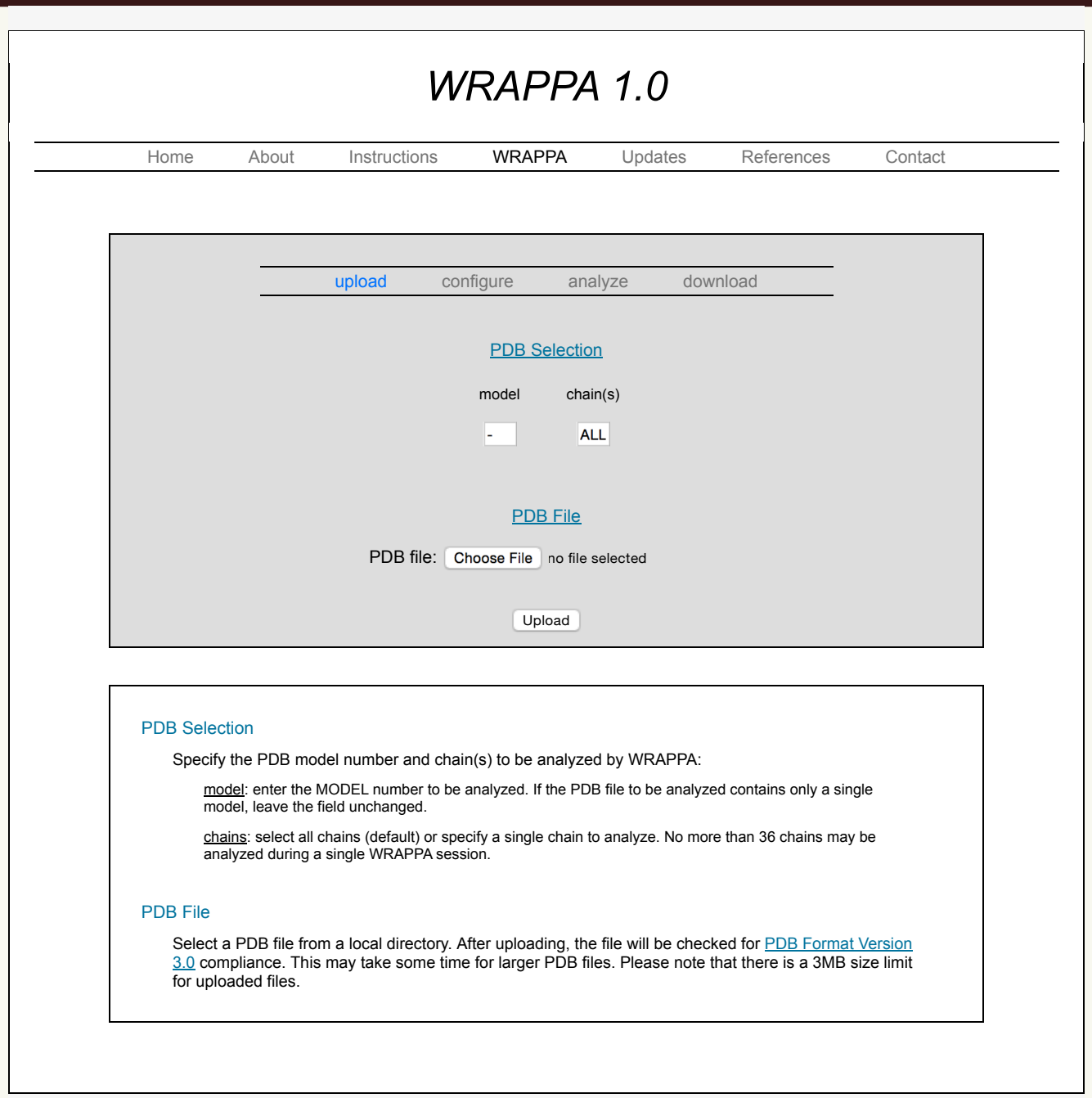
Dehydrons
in human hemoglobin

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 electrostatic environment.

Introducing wrappa: Chris Fraser



What is hard about this?

Protein (and other bio-) data is messy

Some of this is formal in the PDB

- alternate atom locations
- inserted residues

Some of this is not formalized

- missing sidechains
- missing residues

Makes definitive statements difficult

Thanks

This talk featured joint work with

- Babak Bagheri: PC, guarded memory, Pfortran, UHBD, fec, Analysa
- Terry Clark: Pfortran, Euler Gromos, NPACI data-intensive system under linux, SAGE
- Reinhard von Hanxleden: Euler Gromos
- Ernesto Gomez: PC, SOS
- Alexander Veit: compressed Schrodinger
- Chris Fraser: Wrappa, various Python tools

We acknowledge partial support by NSF grant DMS-1226019.

References

[1] Alfonso De Simone, Guy G. Dodson, Chandra S. Verma, Adriana Zagari, and Franca Fraternali. Prion and water: Tight and dynamical hydration sites have a key role in structural stability. *Proceedings of the National Academy of Sciences, USA*, 102:7535–7540, 2005.

[2] A. Ilin, B. Bagheri, R. Metcalfe, and L. R. Scott. Error control and mesh optimization for high-order finite element approximation of incompressible viscous flow. *Comp. methods appl. Mechanics and Eng.*, 150:313–325, 1997.

[3] Hector Juarez, L. R. Scott, R. Metcalfe, and B. Bagheri. Direct simulation of freely rotating cylinders in viscous flows by high–order finite element methods. Research Report UH/MD 241, Dept. Math., Univ. Houston, 1998.

[4] Hector Juarez, L. R. Scott, R. Metcalfe, and B. Bagheri. Direct simulation of freely rotating cylinders in viscous flows by high–order finite element methods. *Computers & Fluids*, 29:547–582, 2000.

[5] W. G. Pritchard, S. J. Tavener, and L. R. Scott. Numerical and asymptotic methods for certain viscous free-surface flows. *Philos. Trans. Roy. Soc. London A*, 340:1–45, 1992.

[6] Anna-Karin Tornberg, R. Metcalfe, L. R. Scott, and B. Bagheri. A fluid particle simulation method. In eds. M-O. Bristeau et al., editor, *Computational Science for the 21st century*, pages 312–321. John Wiley & Sons, New York, 1997.

[7] Anna-Karin Tornberg, R. Metcalfe, L. R. Scott, and B. Bagheri. A front-tracking method for simulating bubble motion in a viscous incompressible fluid using high–order finite element methods. In *1997 ASME Fluids Engineering Division Summer Meeting FEDSM97-3493*, 1997.