Harvard-MIT Division of Health Sciences and Technology
HST.508: Quantitative Genomics, Fall 2005
Instructors: Leonid Mirny, Robert Berwick, Alvin Kho, Isaac Kohane

Protein structure forces, and folding

Thermodynamics

Let probability to be in state “a” = P(a), at fixed volume V

	$P(a)/P(b)=\exp(-(F(a)-F(b)/kT)$
Helmholtz free energy	$F=E-TS$ (at V=const)
Entropy	$S=k \log$(Number of accessible states)
Gibbs free energy	$G=H-TS$ (at P=const)
Enthalpy	$H=E+PV$

for molecular systems (liquid) at 1 atm $PV<<kT$

=> $H \approx E$ hence F=G=“Free energy”

INTERACTIONS

Figure removed due to copyright considerations.

SOLVENT: Hydrogen bonds

Figure removed due to copyright considerations.

SOLVENT: Hydrophobic effect

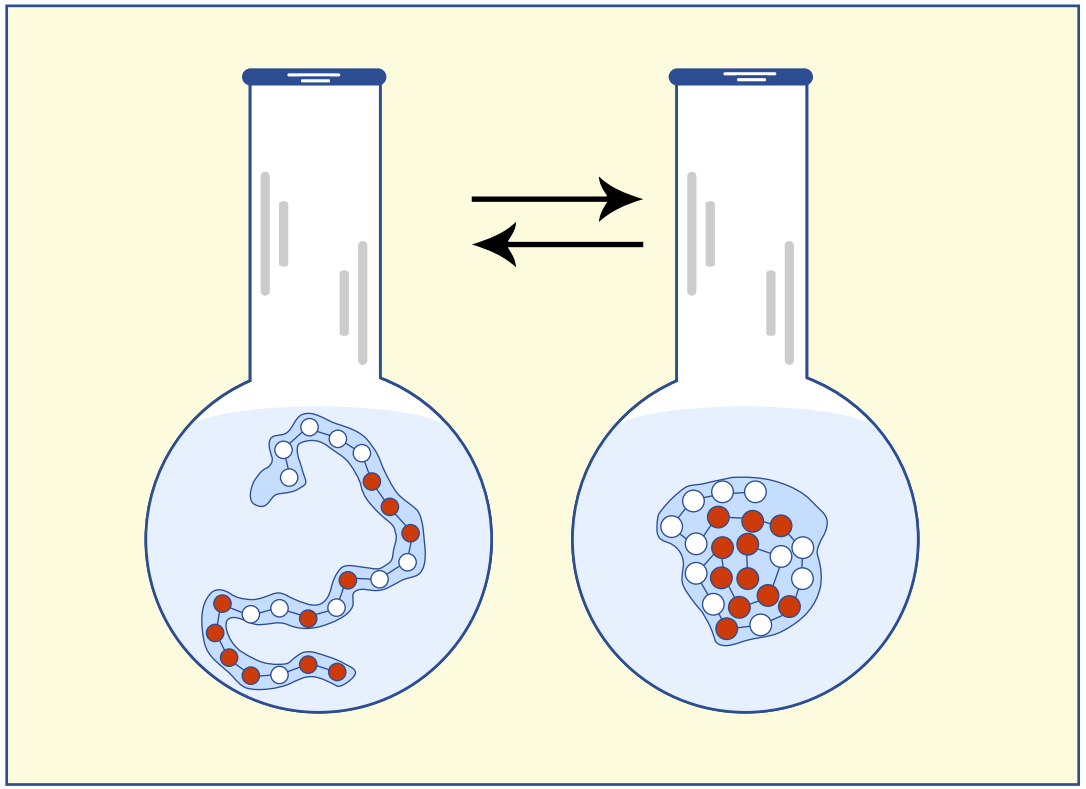
Hydrophobic Interactions
("Hydrophobic Effect")

Frank & Evans 1945

- Water molecules form hydrogen bonds
- Polar groups do not disturb the network of water-water interactions.
- Non-polar (hydrophobic) groups disrupt the network leading to formation of "local ordering" of water.
- Local ordering reduces the entropy

Figure removed due to copyright reasons.

Please see Figure 2 in:
Laidig, Keith E., and Valerie Daggett. "Testing the Modified Hydration-Shell Hydrogen-Bond Model of Hydrophobic EffectsUsing Molecular Dynamics Simulation." *J Phys Chem* 100 (1996): 5616-5619.



Schematic of protein-folding equilibrium. The red and white circles represent hydrophobic and hydrophilic residues, respectively. The shaded region depicts aqueous solution.

Figure by MIT OCW.

Forces

- Hydrophobic interactions
  - Walter Kauzmann
  - energetic (<1nm) and
  - entopic (>1nm)

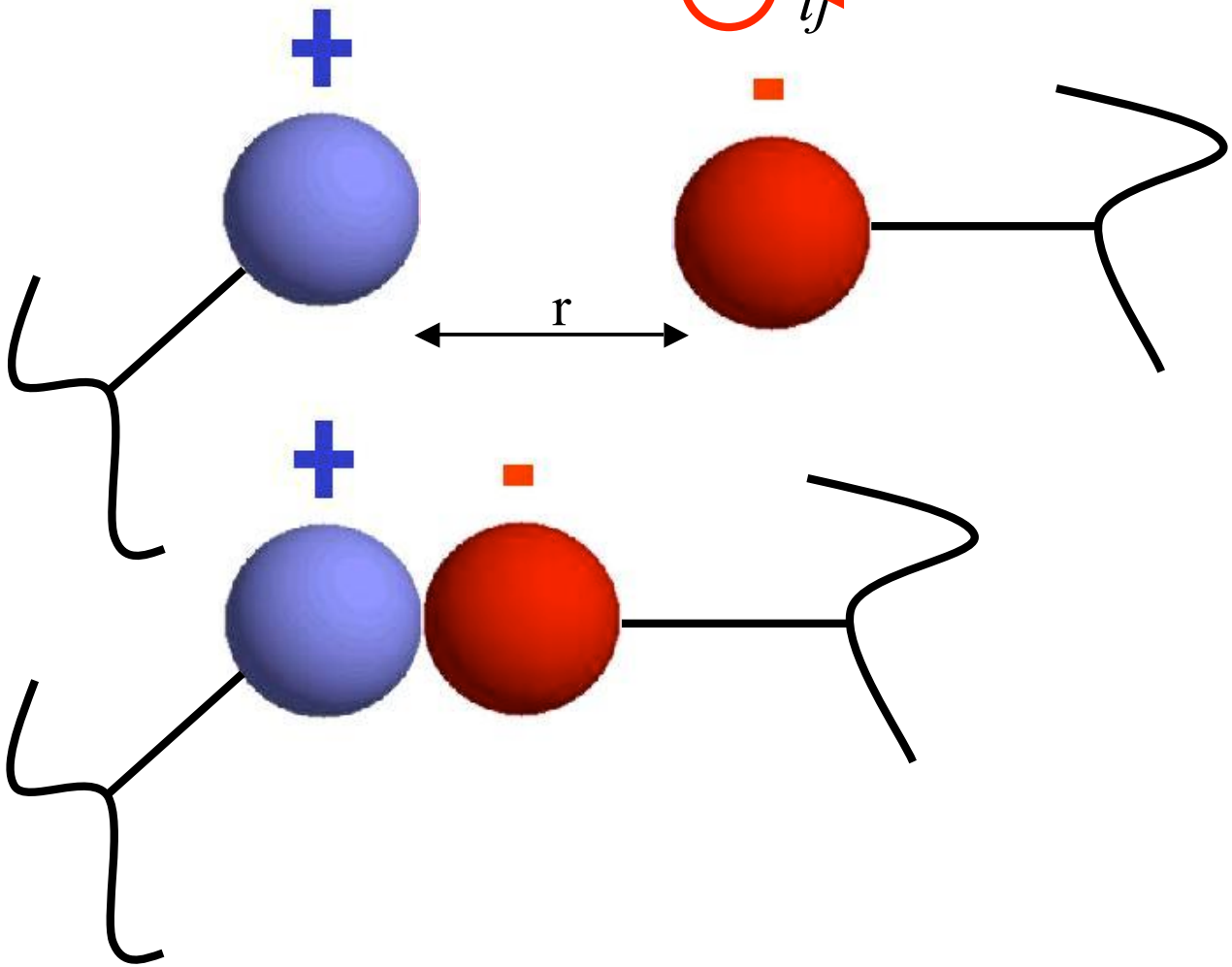
Substitution	Number of examples	ΔΔG (kcal/mol)			ΔG_{tr} (kcal/mol)
		Low	High	Average	
Ile → Val	9	0.5	1.8	1.3 ± 0.4	0.80
Ile → Ala	9	1.1	5.1	3.8 ± 0.7	2.04
Leu → Ala	17	1.7	6.2	3.5 ± 1.1	1.90
Val → Ala	11	0.0	4.7	2.5 ± 0.9	1.24
$-CH_2-$	46	0.0	2.3	1.2 ± 0.4	0.68
Met → Ala	4	2.1	4.6	3.0 ± 0.9	1.26
Phe → Ala	4	3.5	4.4	3.8 ± 0.3	2.02

Figure by MIT OCW.

~10 cal/mol/A^2

ELECTROSTATICS

$$V = \frac{q_i q_j}{4\pi \varepsilon r_{ij}}$$

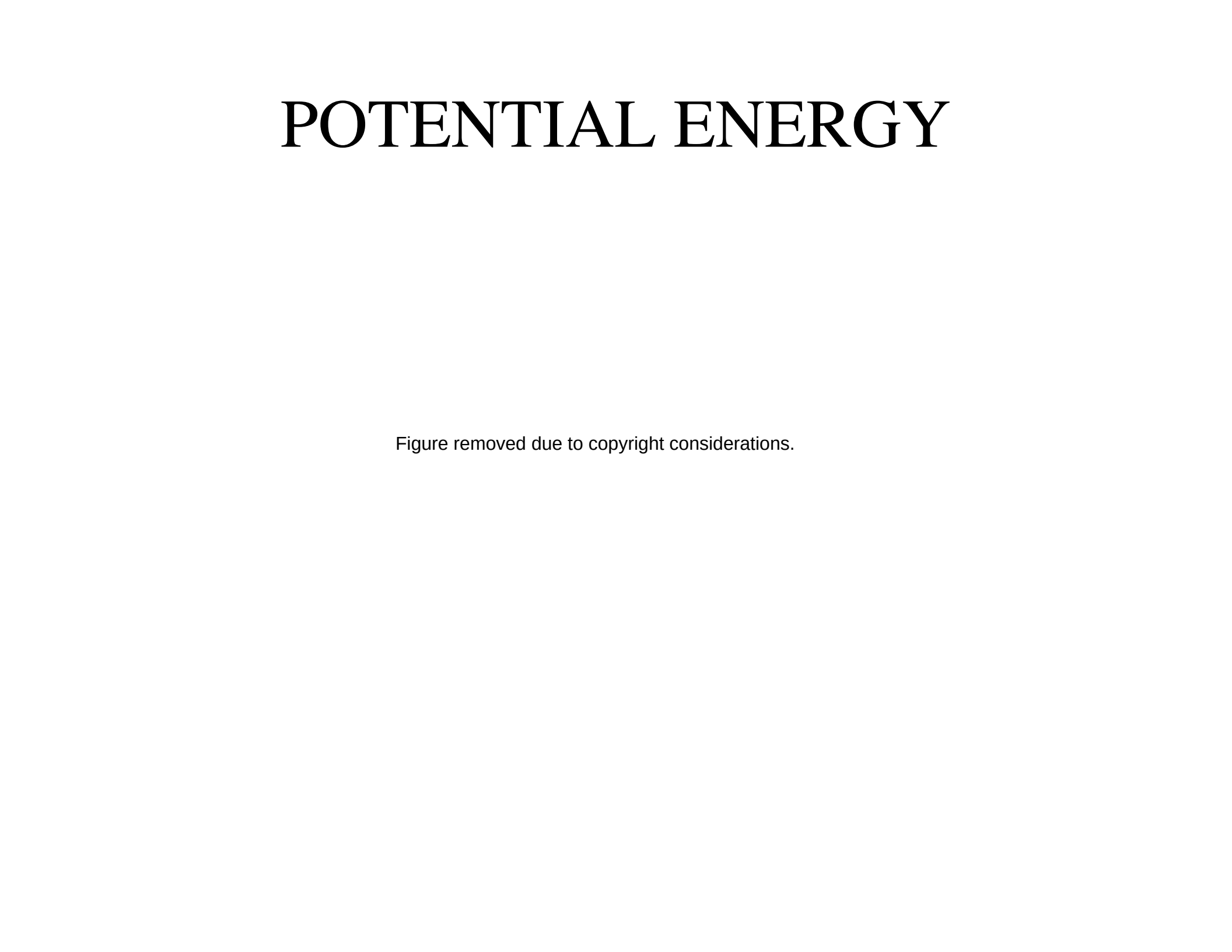


ELECTRO + SOLVENT : Dielectric effect

$$V = \frac{q_i q_j}{4\pi\ \varepsilon\ r_{ij}}; \qquad \varepsilon = 80$$

Figure removed due to copyright considerations.

POTENTIAL ENERGY



In proteins only: Disulfide bonds (S-S bonds)

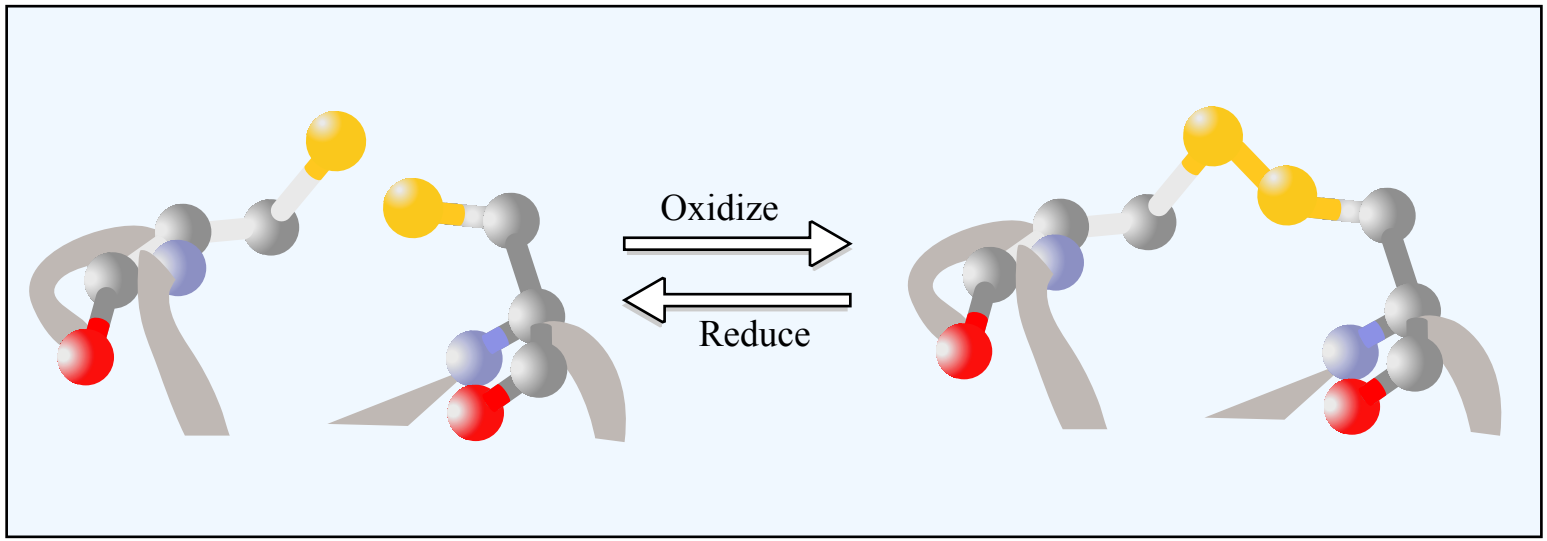


Figure by MIT OCW.

CYS side chain : $-CH_2-SH$

Glutathione

$-CH_2-SH + -CH_2-SH +$ **GS-SG** ⇆

$-CH_2-SH + -CH_2-S-$**SG** + **GSH** ⇆

$-CH_2-S-S-H_2C-$ + 2**GSH**

SUMMARY: Biomolecular forces

Rotation ϕ,ψ	*quantum*	1 Kcal/mol
H-bonds	*entropic*	5 Kcal/mol

Figures removed due to copyright reasons.

VdWaals	*quantum*	0.2 Kcal/mol
Hydrophobic	*entropic*	1.5 Kcal/mol ~10 cal/mol/A^2
Electrostatic	*entropic!*	2-3 Kcal/mol

Protein Folding Problem

- **HOW DOES A PROTEIN FOLD?**

  Levinthal Paradox:

  A protein of 100 amino acids has ~ 4^{100} ~ 10^{62} possible conformations. Folding by trying each conformation in
  10^{-12} sec will take 10^{44} years!

  BUT it takes a protein only 10^{-1}..10^{-2} seconds to fold...

- **PREDICT PROTEIN STRUCTURE FROM IT SEQUENCE.**

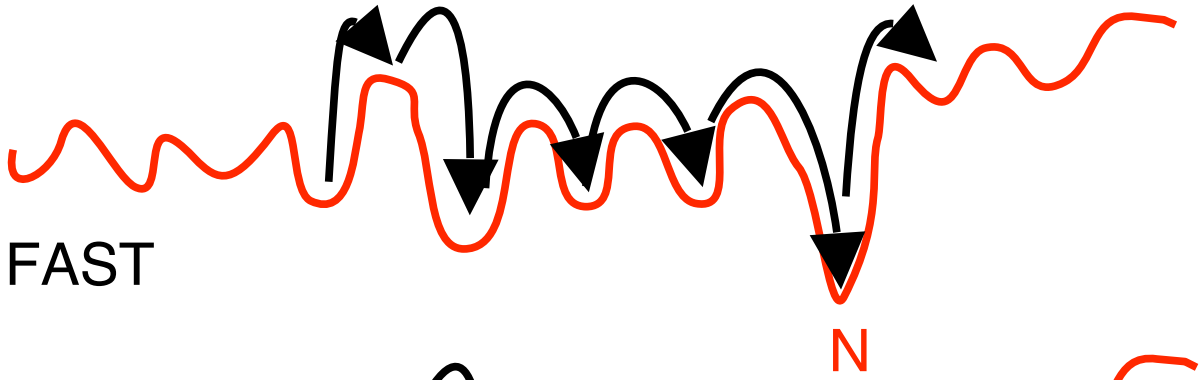
  Is information contained in protein sequence sufficient to determine protein structure?

  Anfinsen Experiment

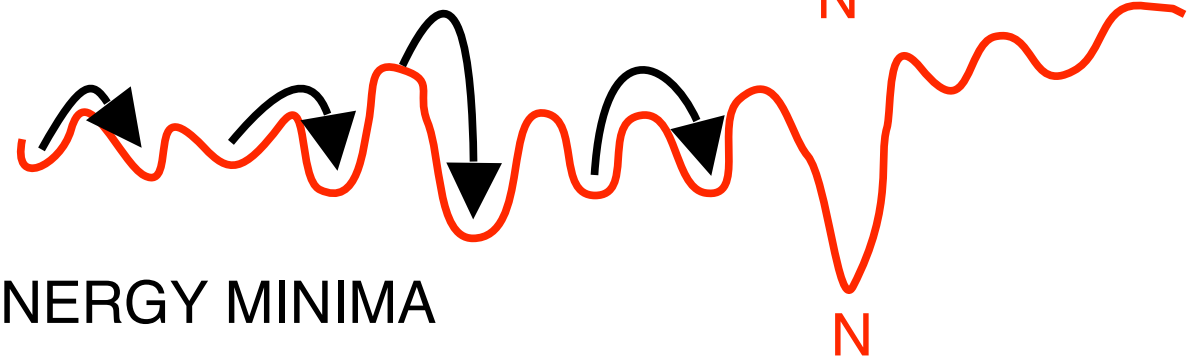
Levinthal paradox

RANDOM PROTEIN

High T
UNSTABLE, i.e.
UNFOLDS VERY FAST



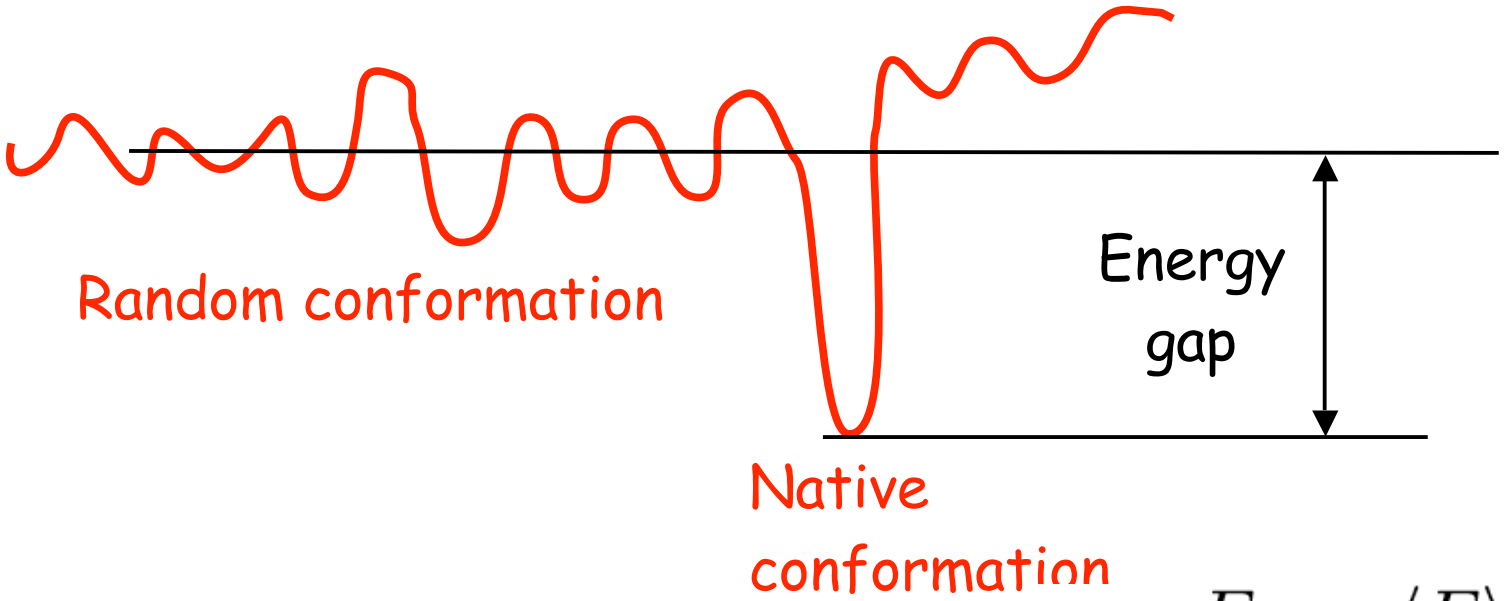
Low T
STABLE, but
CAN'T FOLD
DUE TO LOCAL ENERGY MINIMA



HOW DO NATIVE PROTEINS FOLD???

THEY EVOLVED TO FOLD!

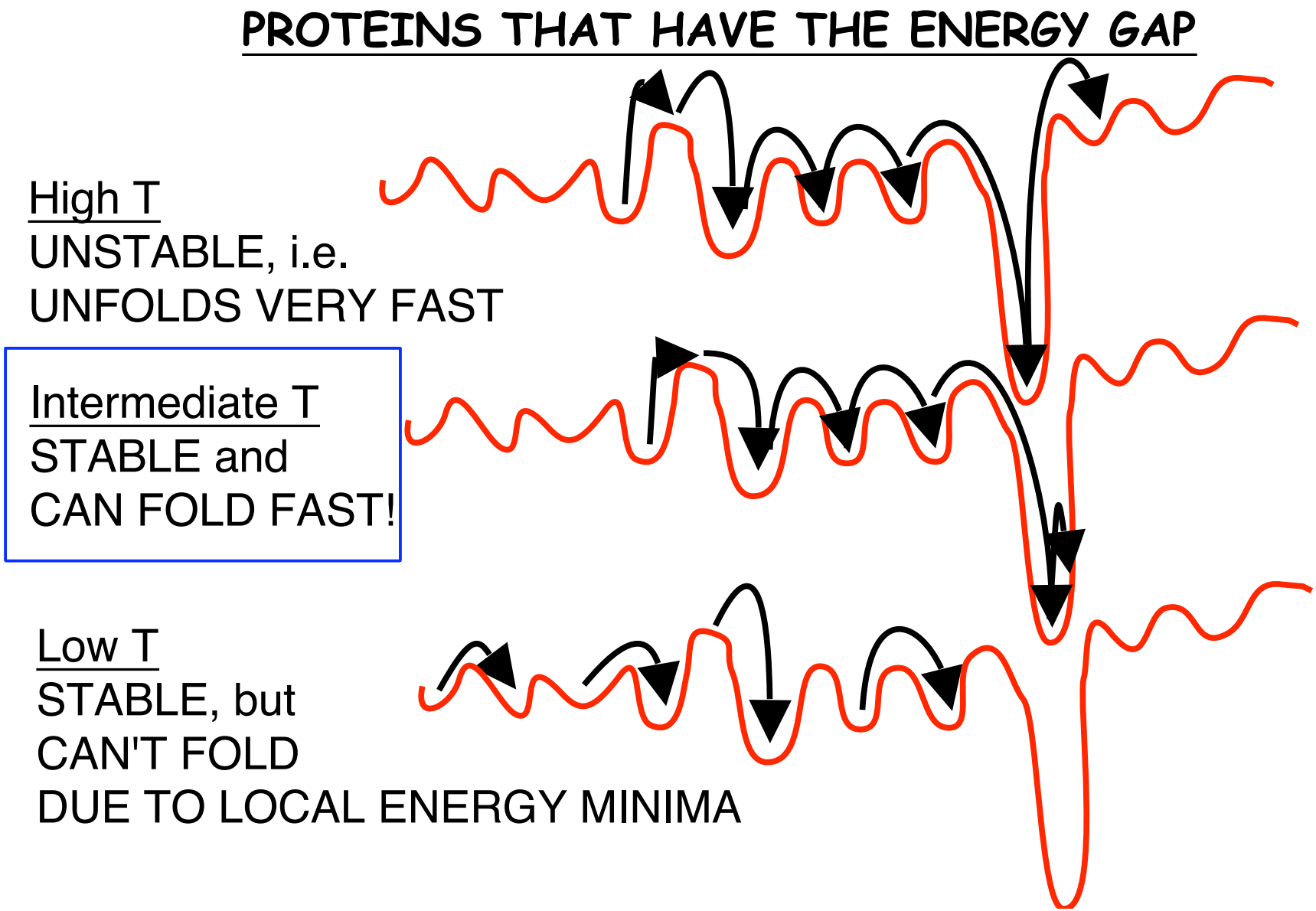
All foldable proteins have large energy gap!



$$Z = \frac{E_N - \langle E \rangle}{\sigma(E)}$$

Why does it help to have the large energy gap?

Solving the Levinthal paradox



Anfinsen Experiment

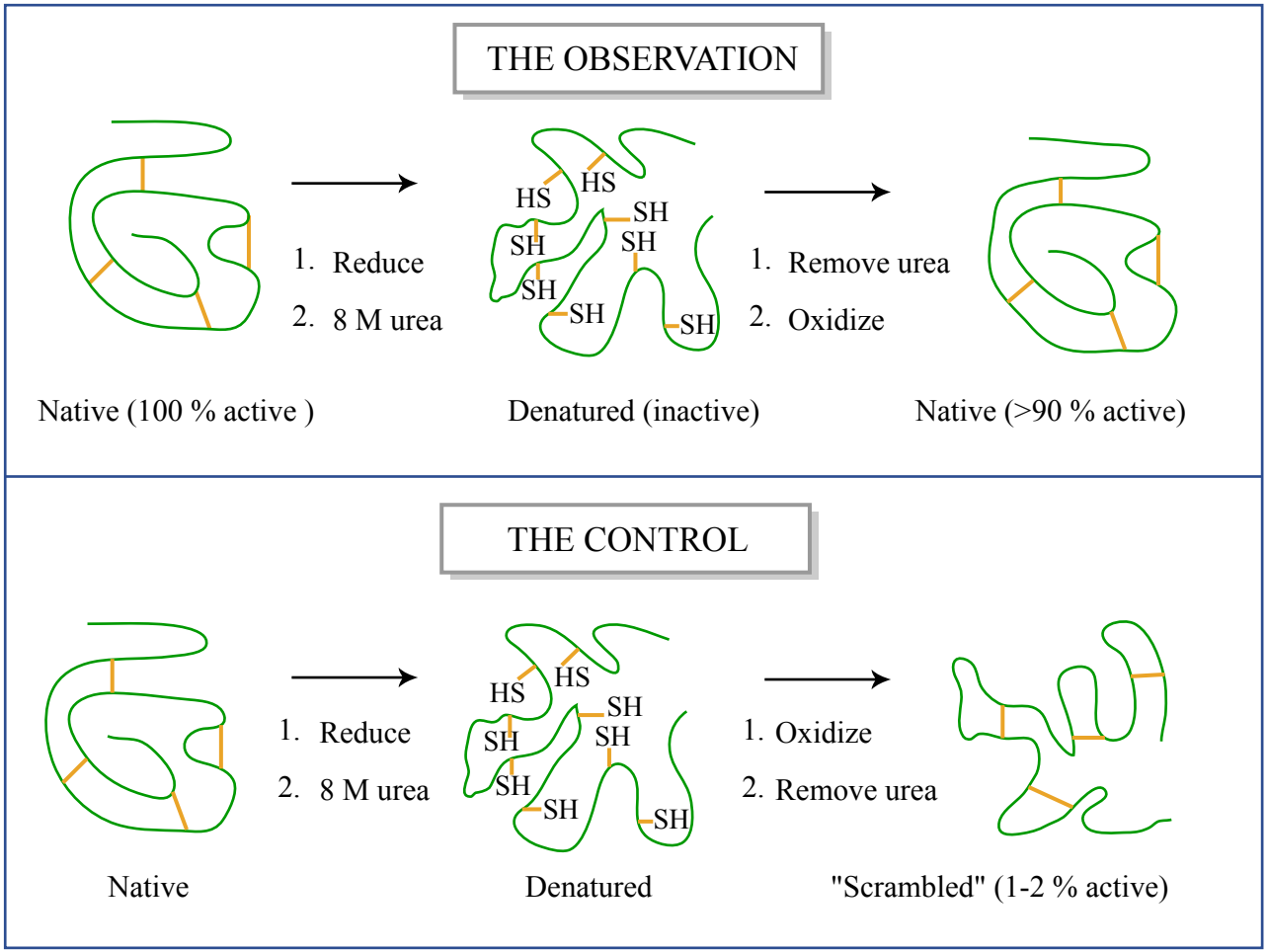
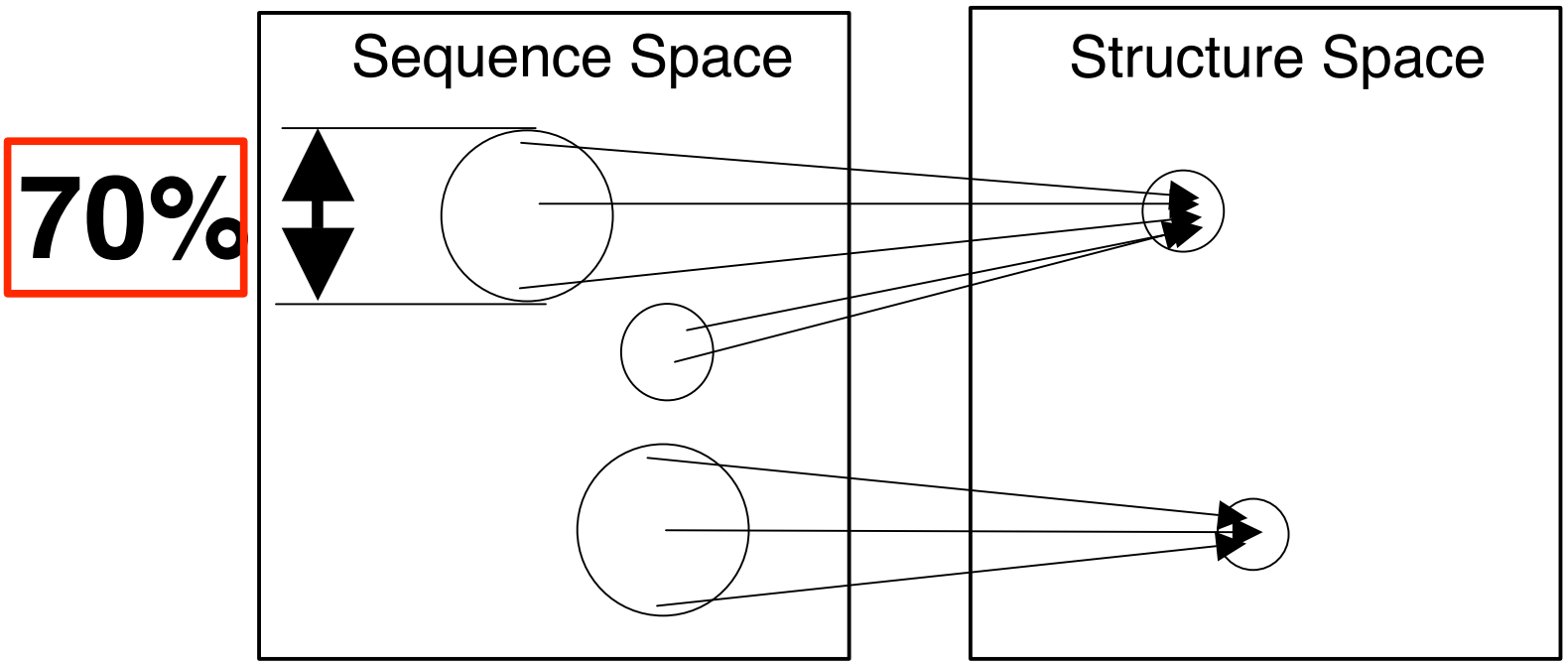


Figure by MIT OCW.

- Information contained in the protein sequence is sufficient to determine protein structure!
- THERMODYNAMIC HYPOTHESIS:
  The native structure is the GLOBAL minimum of free energy.

Sequence–Structure Mapping

- Similar sequences always have similar structures.
- Different sequences have different structures, **but**
- Different sequences may have similar structures.



Protein structure prediction

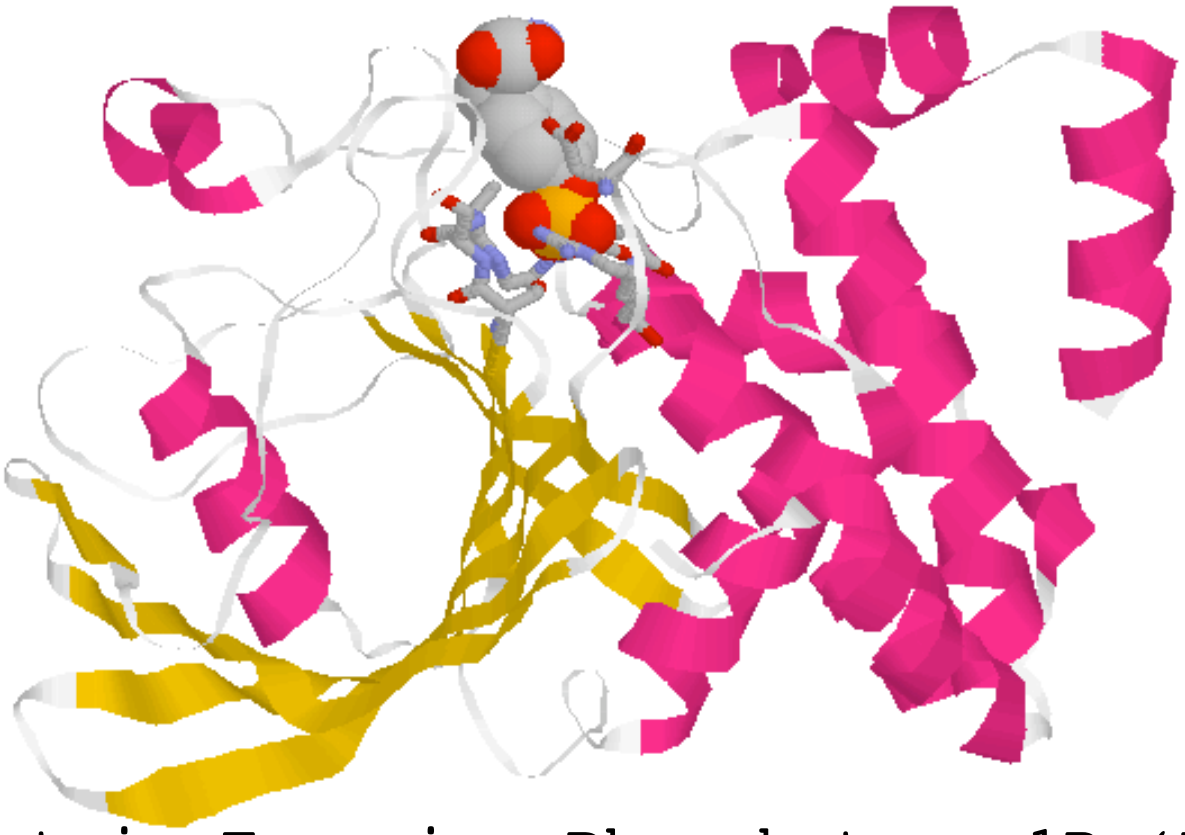
- Homology modeling
- Fold recognition/ Threading
- Ab initio

NEED:

1. Scoring/Energy
2. Sampling/Minimzation

Protein Function: catalysis and binding

Active site



Protein Tyrosine Phosphatase 1B (PDB entry: 1pty) complexed with a phosphotyrosine molecule.

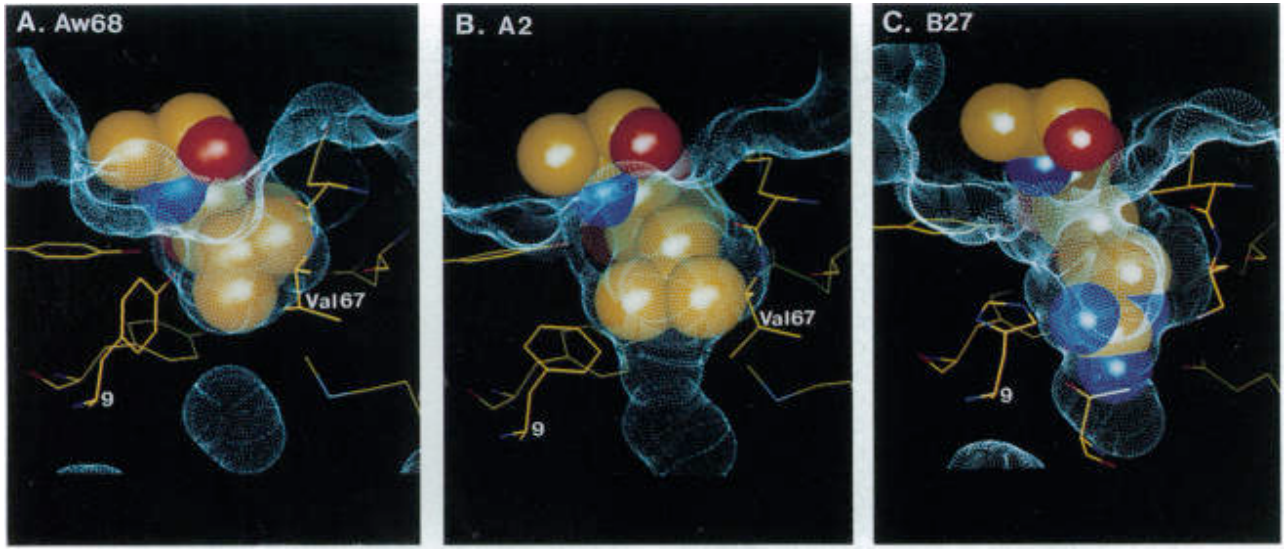
PDB JRNL REFERENCE for PDB ID=1pty: Puius, Y. A., Y. Zhao, M. Sullivan, D. S. Lawrence, S. C. Almo, and Z. Y. Zhang. "Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design." *Proc Natl Acad Sci USA* v94 (1997): 13420-13425.

The Protein Data Bank (PDB - http://www.pdb.org/) is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. "The Protein Data Bank." *Nucleic Acids Research* 28 (2000): 235-242.

(PDB Advisory Notice on using materials available in the archive: http://www.rcsb.org/pdb/advisory.html)

Specificity pocket



Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8053–8057, September 1993
Immunology

Comparison of the P2 specificity pocket in three human histocompatibility antigens: HLA-A*6801, HLA-A*0201, and HLA-B*2705

(protein crystal structure/allelic specificity/peptide binding/antigen presentation)

H.-C. GUO[†], D. R. MADDEN[†], M. L. SILVER[†‡], T. S. JARDETZKY[†], J. C. GORGA[†§], J. L. STROMINGER[†], AND D. C. WILEY[†¶‖]

[†]Department of Biochemistry and Molecular Biology, Harvard University, and [¶]Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge MA 02138

Contributed by D. C. Wiley, June 4, 1993

Figure 2 in: Guo HC et al. "Comparison of the P2 specificity pocket in three human histocompatibility antigens: HLA-A*6801, HLA-A*0201, and HLA-B*2705." *Proc Natl Acad Sci U.S.A.* 90, no. 17 (Sep 1, 1993): 8053-7.