united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR/1499 - 23

INTERNATIONAL WORKSHOP ON PROTEOMICS:
PROTEIN STRUCTURE, FUNCTION AND INTERACTIONS
(5 - 16 May 2003)

" Recognition in biomolecular energy landscapes: protein association vs protein folding "

presented by:

G. Papoian
University of California at San Diego
United States of America

Recognition in Biomolecular Energy Landscapes: Protein Association vs Protein Folding

Garegin A. Papoian

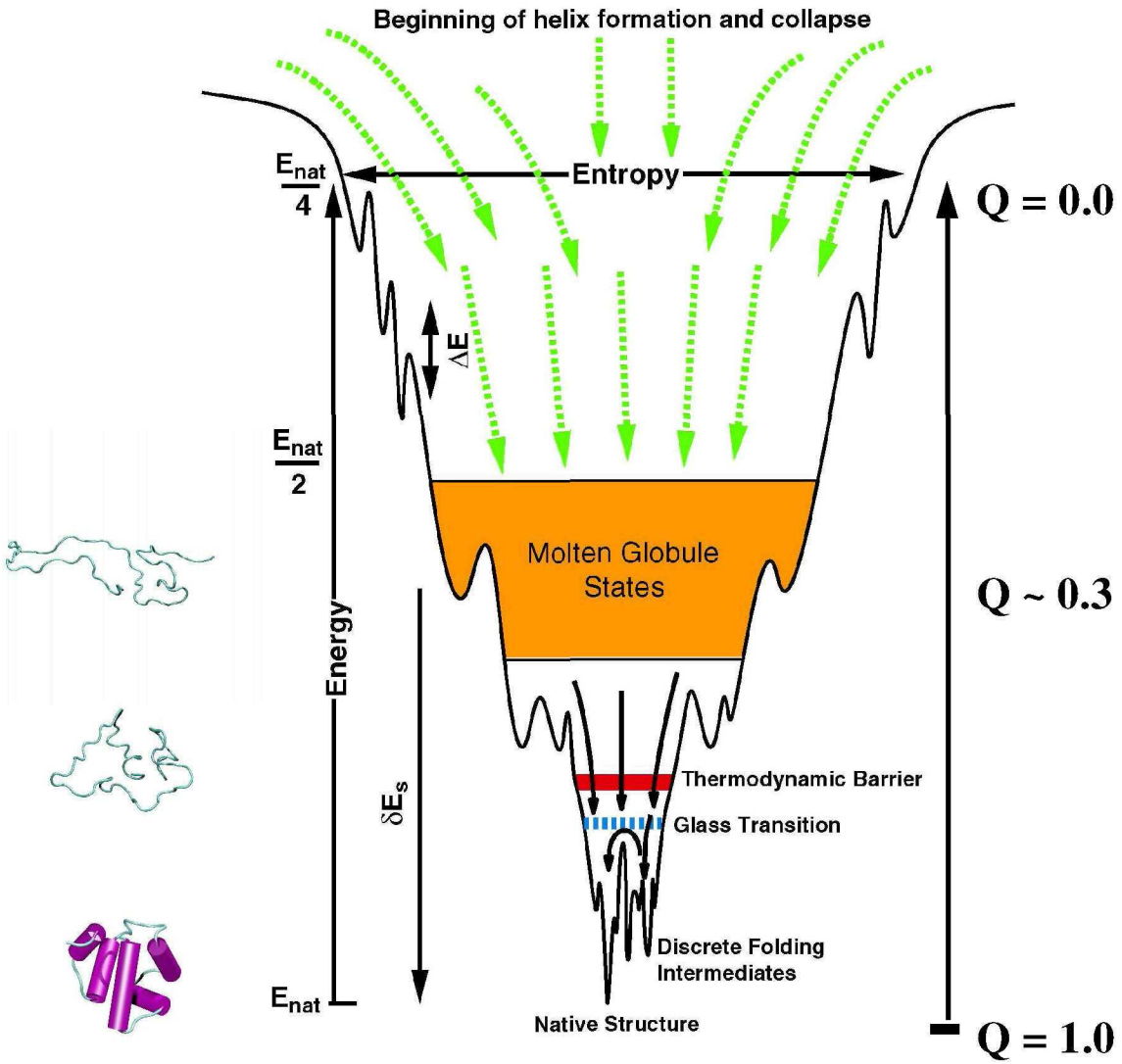
University of California at San Diego

Talk Outline

Johan Ulander
Peter G. Wolynes

- Protein folding funnel & the nature of trap states.
- Coupling of binding and folding processes.
- Optimization of Association Potentials.
- Completing the circle: Using what we have learned in association to improve protein structure prediction algorithms

Energy Landscape Theory of Protein Folding



- **Denaturated Ensemble:**
  - Large Structural Entropy (+),
  - Ruggedness of Energies (+),
  - Energetically Poor (-).
- **Native Ensemble:**
  - Energetically Stable (+),
  - Only Few Configurations (-).
- +/- Legend:
  - (+) stabilizes Free Energy
  - (-) destabilizes Free Energy
- **Folding Order Parameter:**
  - Q – overlap with the native state.

J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, N. D. Socci, PNAS, 92, (1995), 3626

Paradigms of Protein Association

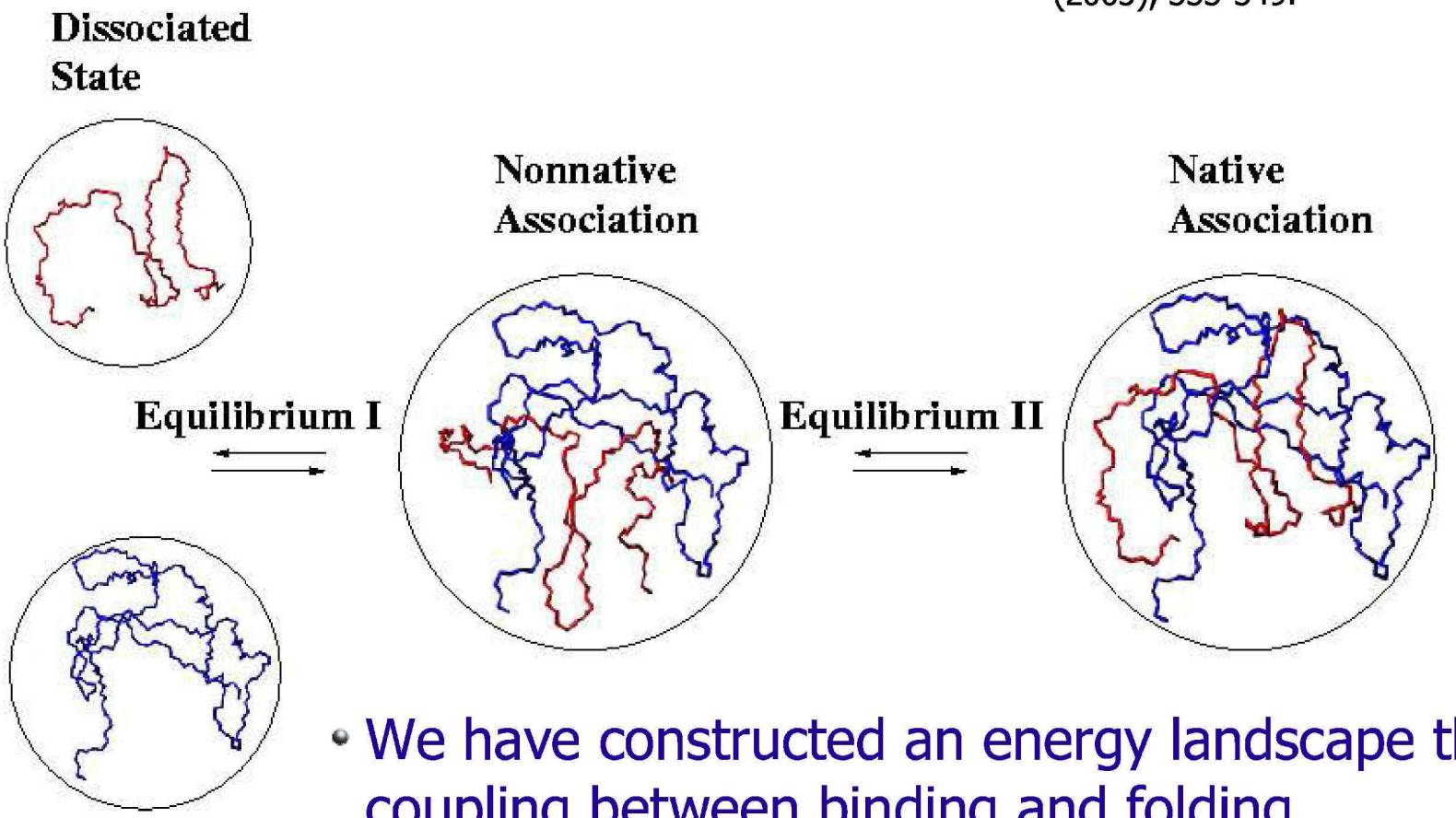
- **Lock-and-Key Mechanism:**
  - E. Fischer (**1890's**),
  - Interact as rigid bodies: Steric and electrostatic complementarity
  - Theoretical Modeling: Docking; Brownian Dynamics of rigid bodies.

- **Induced Fit:**
  - D. E. Koshland Jr (**1960's**).
  - Proteins adjust to each other during association.
  - Theoretical Modeling: Docking with some plasticity allowed.

- **Association between highly disordered proteins:**
  - 1990's.
  - Theoretical Modeling – one of the objectives of the current work.

Coupling of Folding and Binding Funnels

G. A. Papoian and P. G. Wolynes, Biopolymes, 68, (2003), 333-349.



- We have constructed an energy landscape theory of coupling between binding and folding.
- We have surveyed a structural database to extract model parameters.
- We have developed a theory for the entropy change during the binding/folding process.

Free Energy for Simultaneous Binding and Folding

$$F(Q_f, Q_b) = N_f \langle E \rangle^f Q_f + N_b \langle E \rangle^b Q_b$$
$$+ N_f \langle E \rangle^{nn} (1 - Q_f) + N_b \langle E \rangle^{nn} (1 - Q_b)$$
$$- S^0(Q_f, Q_b) T$$
$$- \frac{N_f (1 - Q_f)(1 + \gamma_f Q_f) \langle \Delta E_{nn}^2 \rangle}{2 k_B T}$$
$$- \frac{N_b (1 - Q_b)(1 + \gamma_b Q_b) \langle \Delta E_{nn}^2 \rangle}{2 k_B T}$$

Native Energy

Nonnative Energy

Configurational Entropy

Stabilization due to Ruggedness

N_f, N_b – total number of folding/binding contacts,

$\langle E \rangle^f, \langle E \rangle^b$ – **native** folding/binding average energy,

$\langle E \rangle^{nn}$ – **nonnative** folding/binding average energy,

$\langle \Delta E_{nn}^2 \rangle$ – **nonnative** folding/binding energy variance,

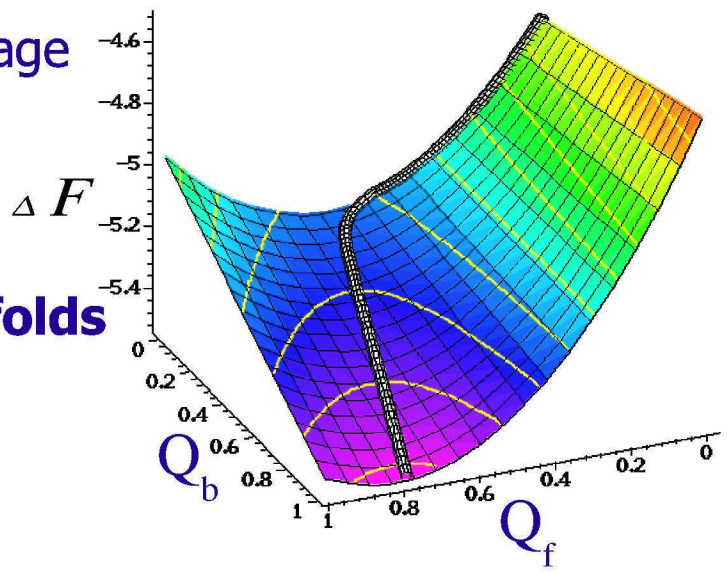
γ_f, γ_b – heterogeneiety of native folding/binding contacts,

Q_f, Q_b – folding/binding order parameters

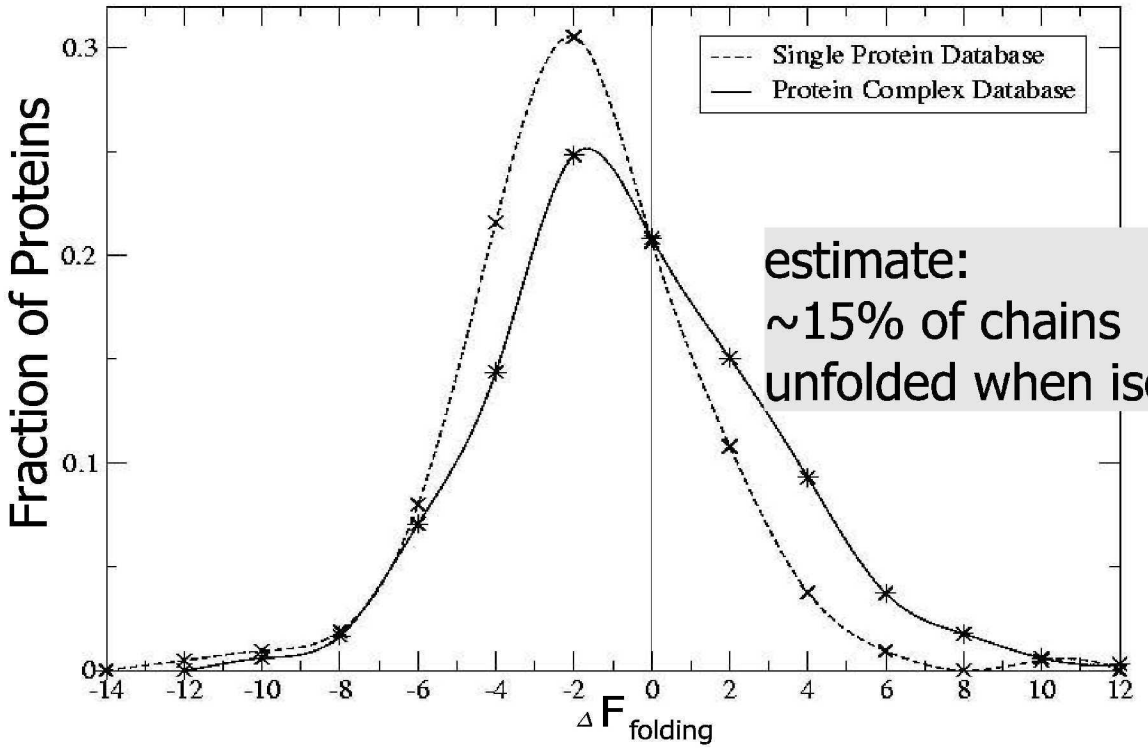
- We have surveyed a structural database to extract parameters listed on the left.
- We have developed a theory for $S^0(Q_f, Q_b)$.

Phase Diagrams for Binding and Folding

Database average parameters:

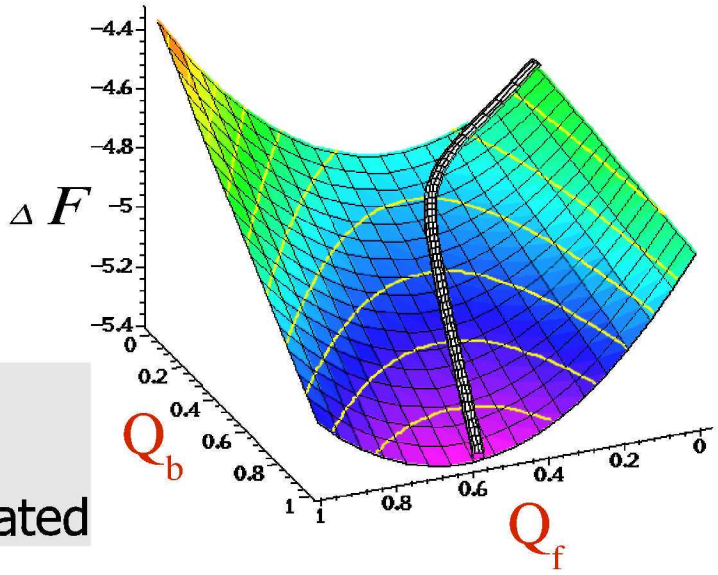


- ~500 protein complexes
- **The protein first folds then binds.**



estimate:
~15% of chains unfolded when isolated

One standard deviation from the database average parameters:

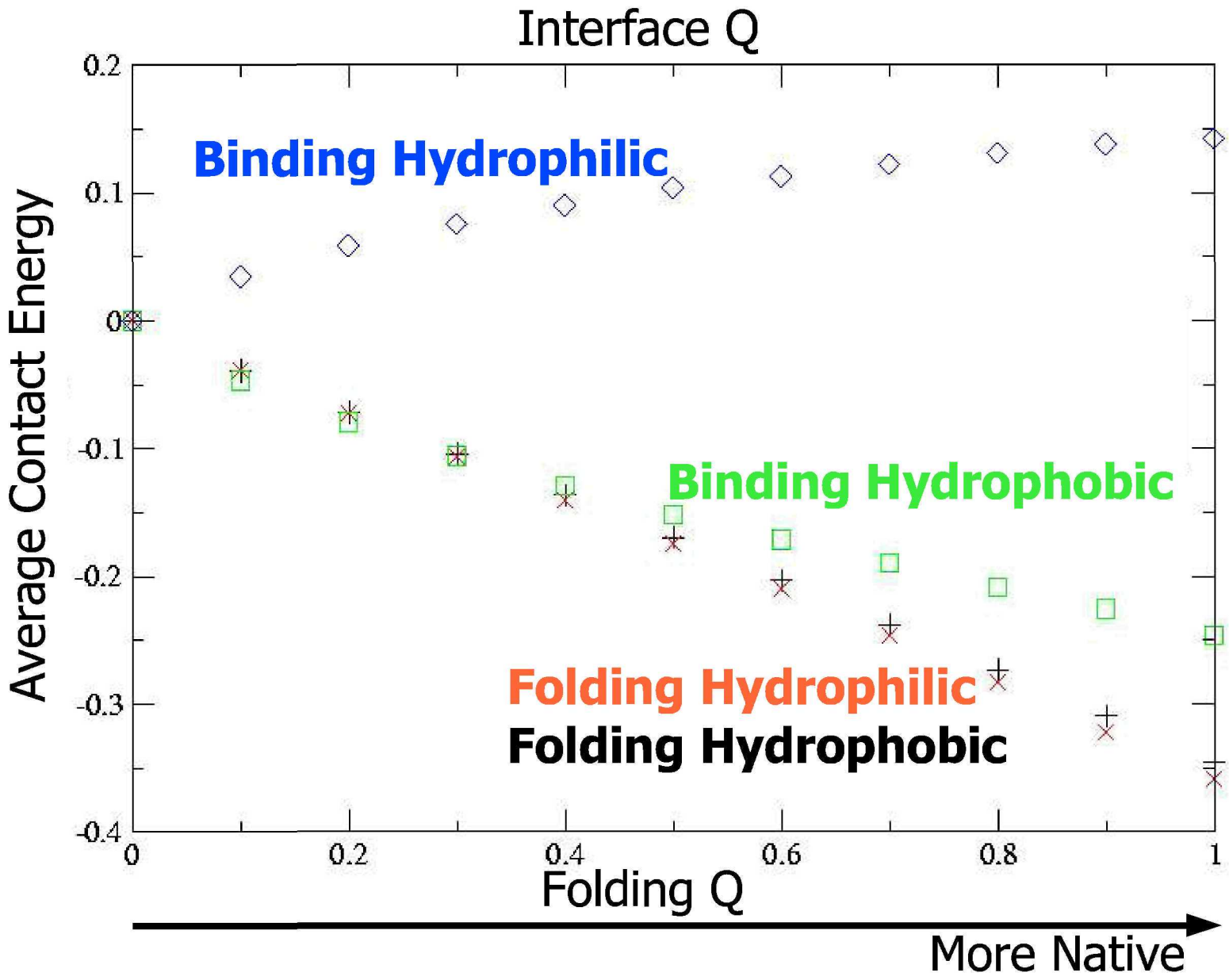


- **On its own the protein is unfolded.**
- **Binding interactions initiate folding.**

Most Unstable Monomeric Chains in the Database

Protein	Chain	$\Delta F_{\text{folding}}$	Short Description
2bpa	3	7.23	Bacteriophage phix174 coat proteins
1una	A	7.32	Unassembled virus coat protein dimer
1lta	C	7.40	Heat-labile enterotoxin (lt) complex with galactose
1mec	4	7.48	Cardio picornavirus coat protein
1cdc	A	7.52	Cd2, N-terminal domain (1–99), truncated form
1tgx	B	7.60	Toxin gamma (cardiotoxin)
1mhl	A	7.75	Human myeloperoxidase isoform c
1cdc	B	7.88	Cd2, N-terminal domain (1–99), truncated form
1bbt	4	7.94	Foot-and-mouth disease virus
1mhl	B	8.25	Human myeloperoxidase isoform c
1tvx	B	8.45	Neutrophil activating peptide-2 variant form m6l
1tgx	A	8.87	Toxin gamma (cardiotoxin)
1fos	G	9.24	Two human c-fos : c-jun : dna complexes
2zta	B	9.67	Leucine zipper monomer
2zta	A	10.26	Leucine zipper monomer
1got	G	10.77	gt-alpha/gi-alpha chimera and the gt-beta-gamma subunits
1lya	A	11.04	Lysosomal aspartic protease, cathepsin d
1fle	I	11.42	Elafin complexed with porcine pancreatic elastase
1tmf	4	12.44	Theiler's murine encephalomyelitis virus coat protein
1lpb	A	13.32	Lipase complexed with colipase

Native Hydrophilic Interfaces are not Recognized by the Standard (Miyazawa-Jernigan) Folding Potentials

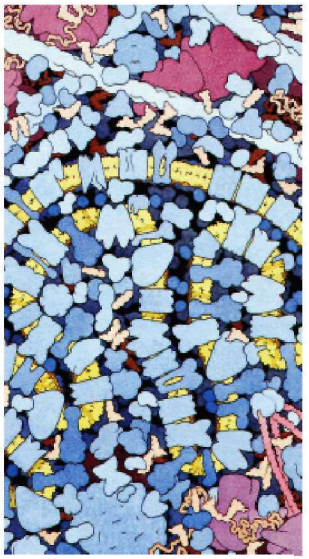


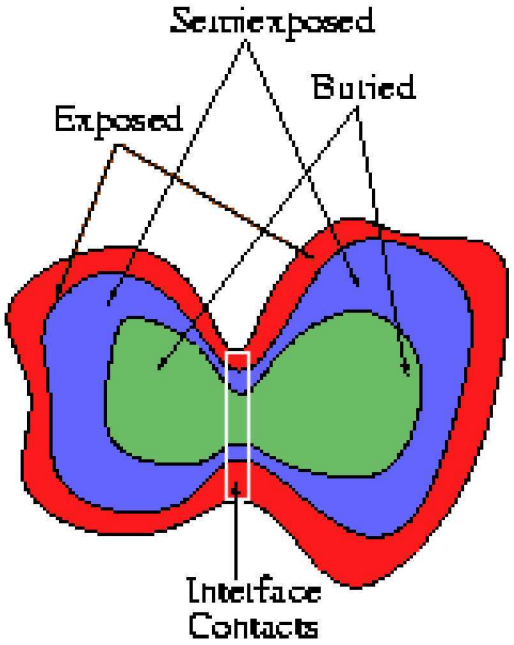
Context-Dependent Nature of Association Potentials

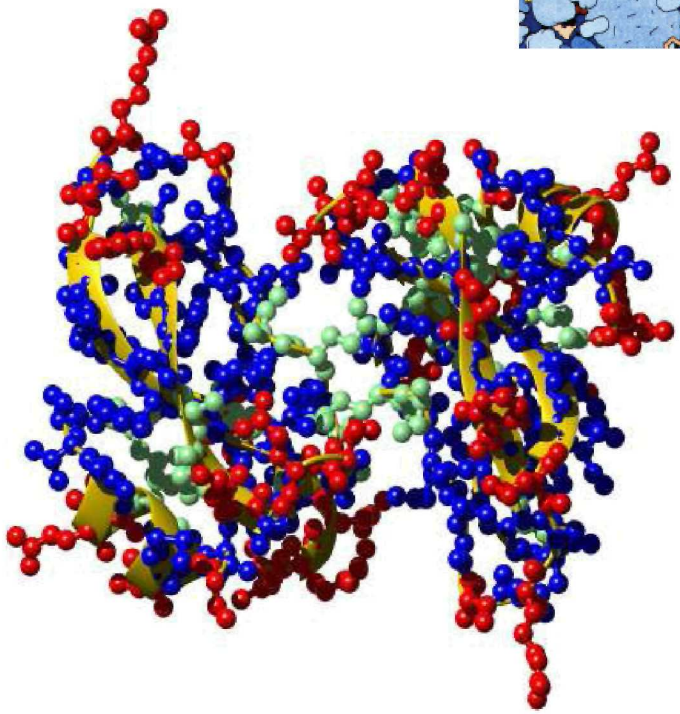
- Protein density: ~300mg/ml
  - Average Protein Concentration: ~5mM

- Need to avoid **nonnative** association with **other** cell proteins that are:
  - Highly Disordered (**Flexible**) – all three layers,
  - Partially Disordered (**Semi-Flexible**) – outer two layers,
  - Natively Ordered (**Rigid**) – outer layer.

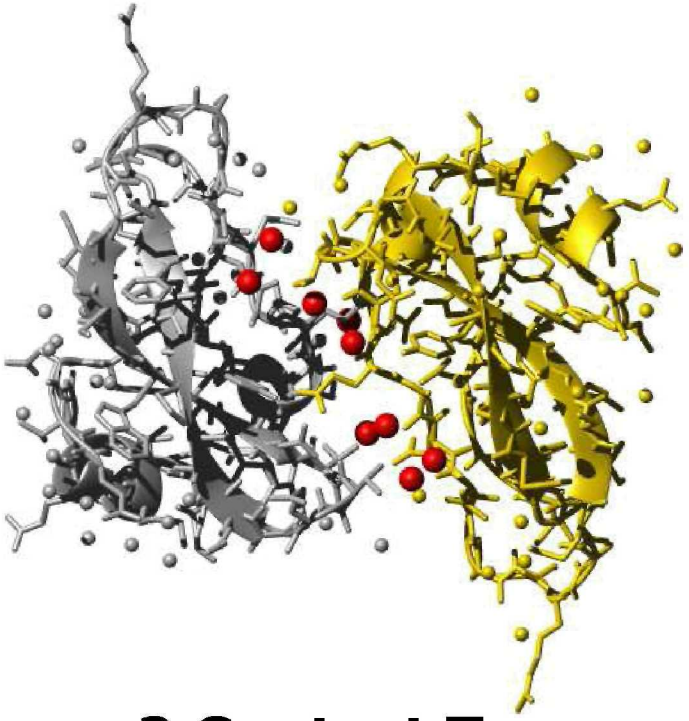
Inside a Eukaryotic Cell. Watercolor on Arches paper. David S. Goodsell. ©1994 Neil Patterson Publishers.







Knowledge-Based Optimization of Direct and Water-Mediated Binding Pair-Potentials



2 Contact Types
3 Trap Models

Defining Contacts:

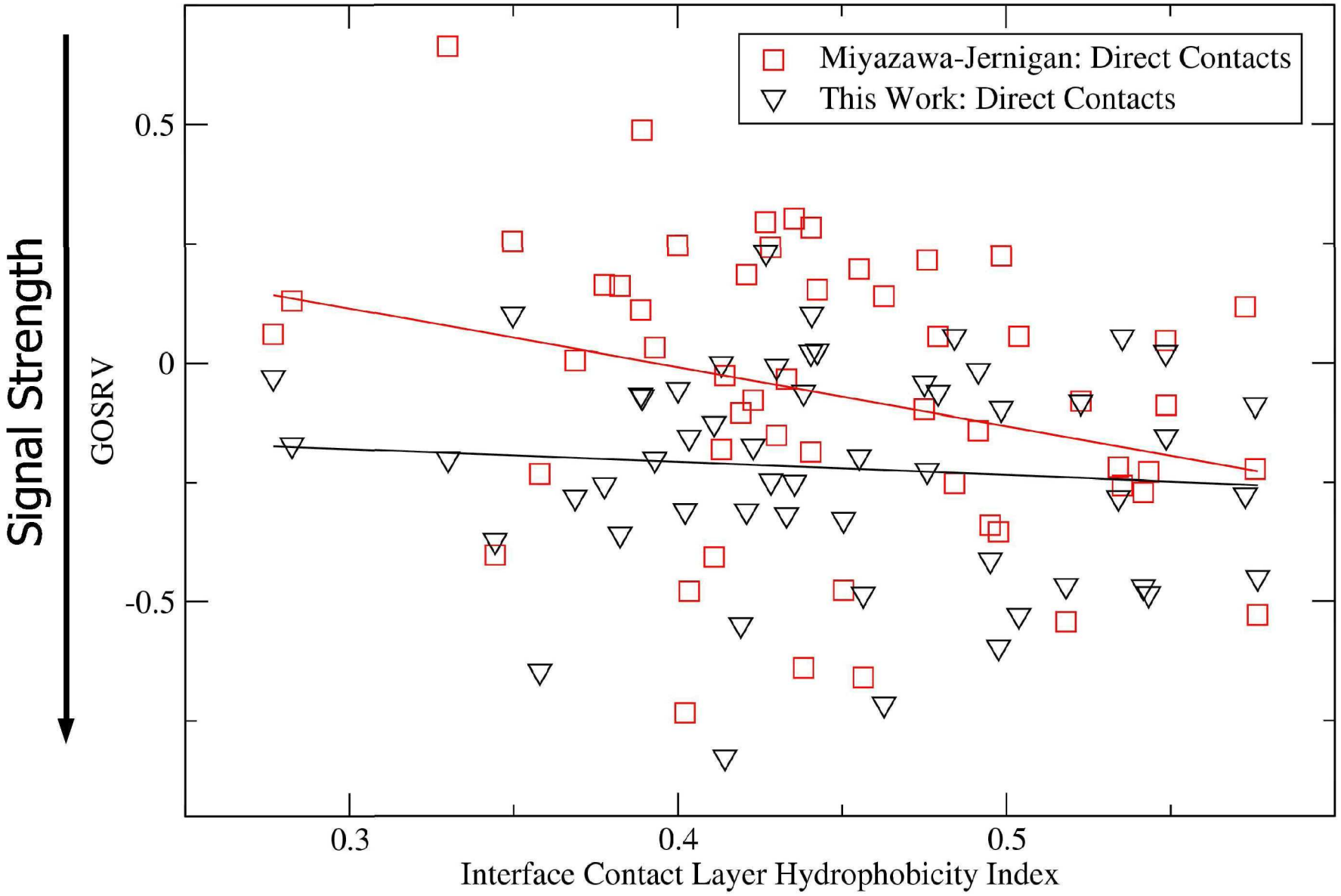
- **Direct** – d < 6.5 Å between C-β atoms.
- **Water-Mediated** – 7.8 Å < d < 9.5 Å between C-β atoms, with the constraint that both residues are at least partially water-exposed.

→ **6 (20x20) pair-potential matrices**

- Optimization Strategy:
  - Maximize Binding Energy Gap while constraining Energy Variance.
  - 222 protein complexes to **train** the potentials.
- Testing:
  - 54 unrelated protein complexes to test for **recognition power**.

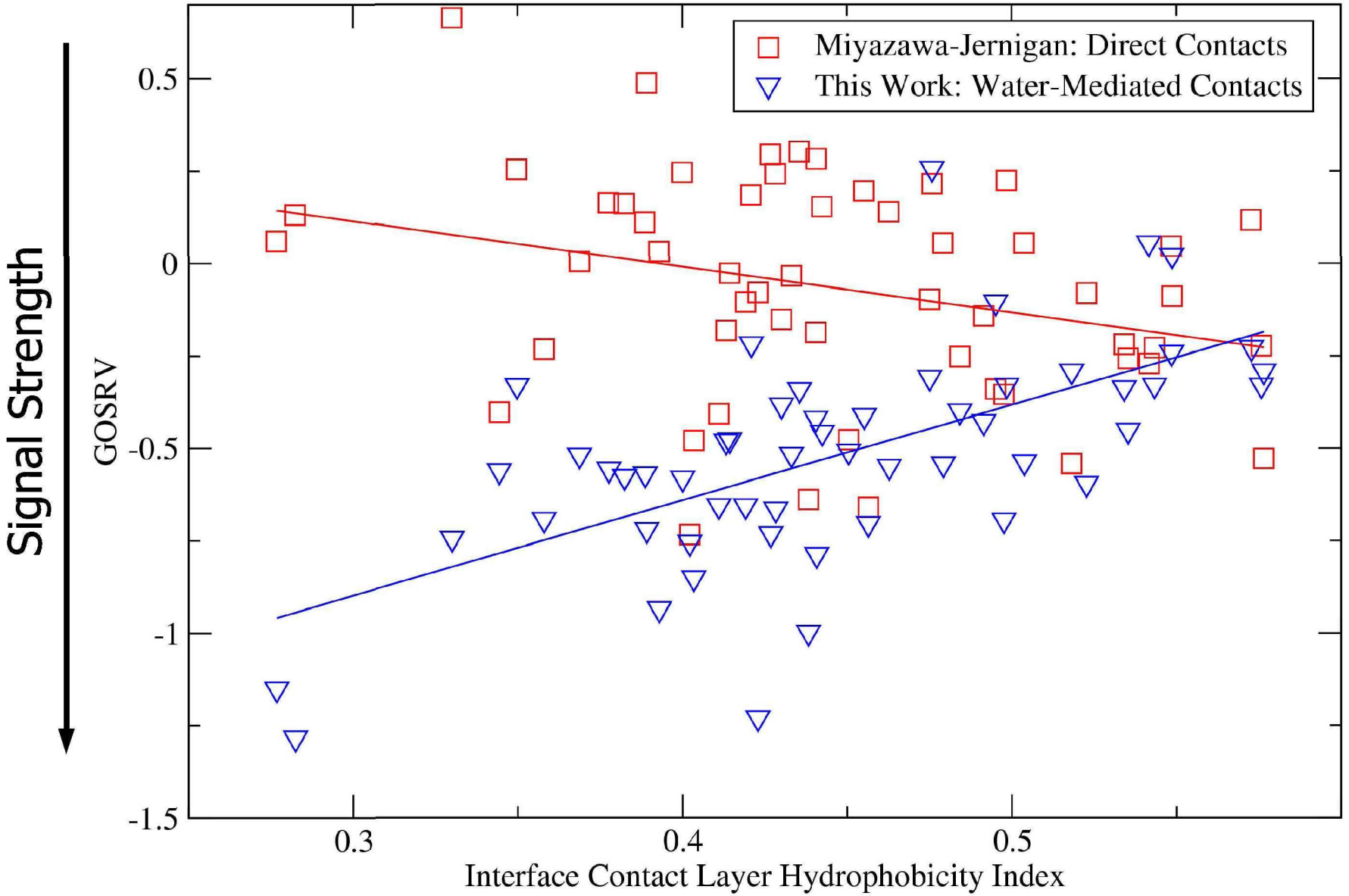
Discrimination against Flexible trap states

(blind test set: 54 proteins)

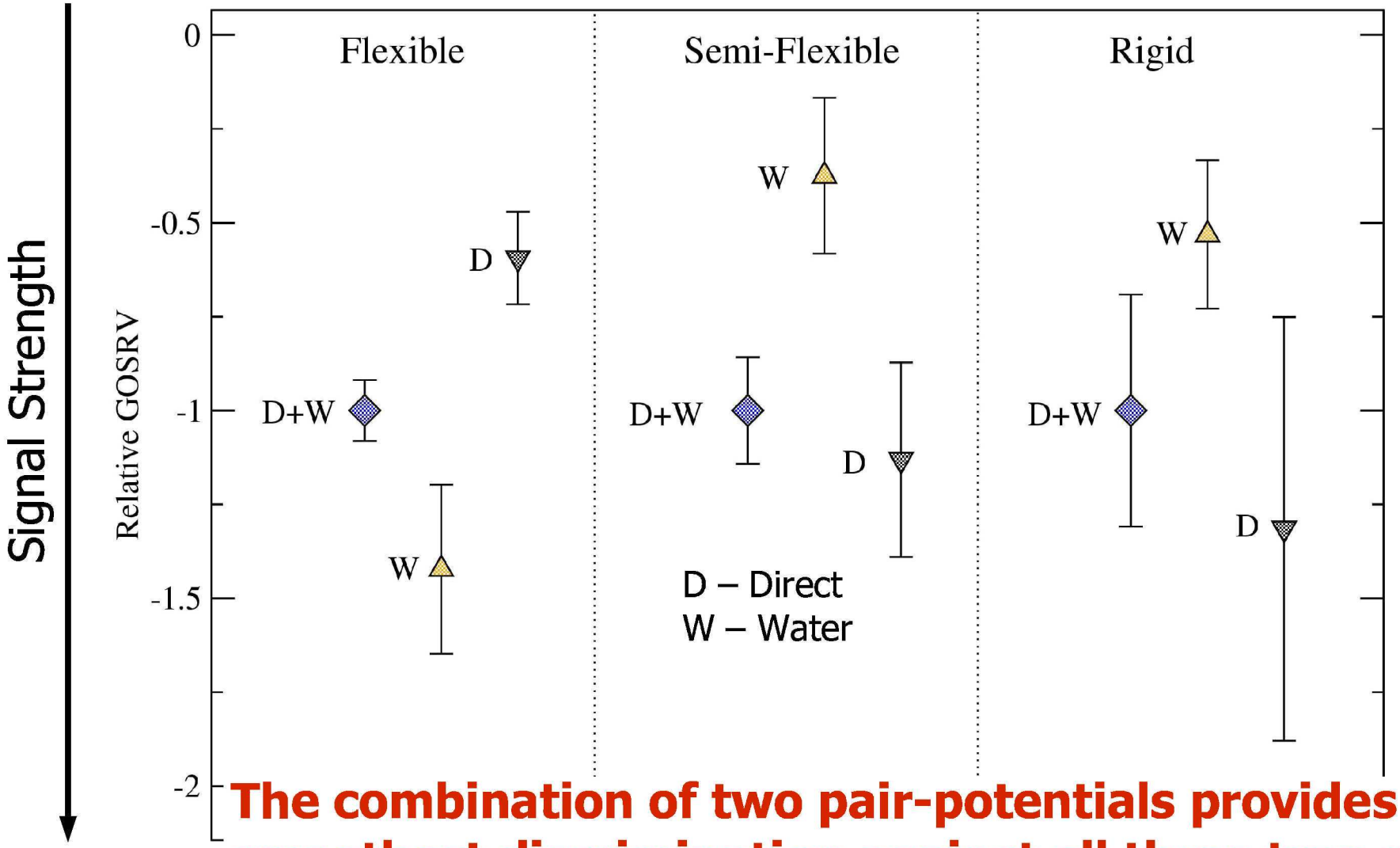


Discrimination against Flexible trap states

(blind test set: 54 proteins)

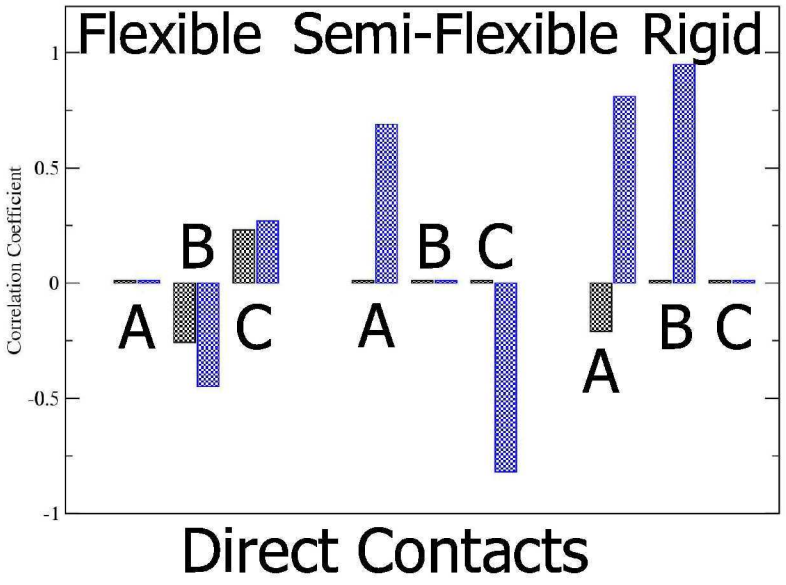


Recognition of Native Interfaces From Trap States: The Relative Performance of Direct and Water-Mediated Pair-Potentials

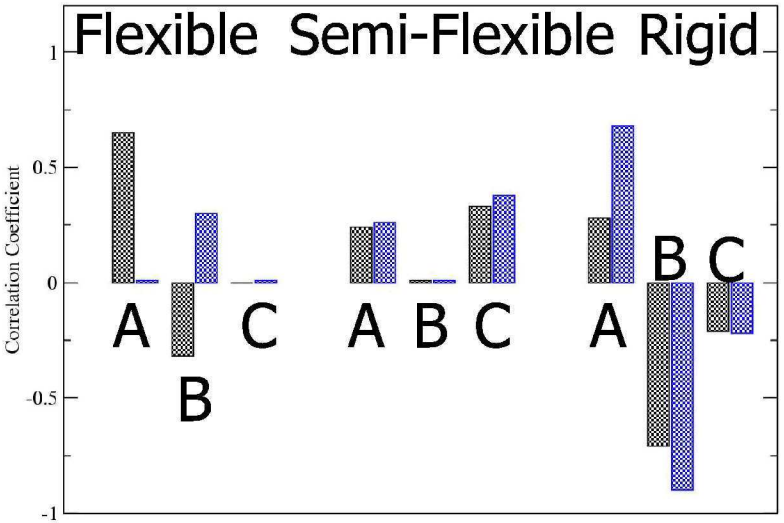


The combination of two pair-potentials provides the smoothest discrimination against all three trap classes.

Coarse-Graining Of Interface Interactions With Canonical Aminoacids



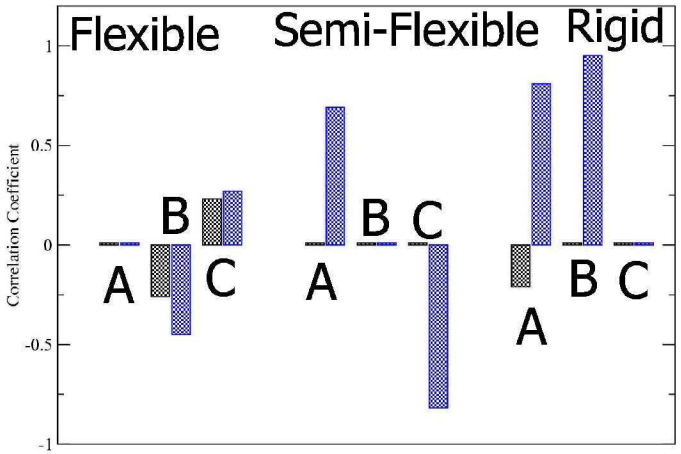
Direct Contacts



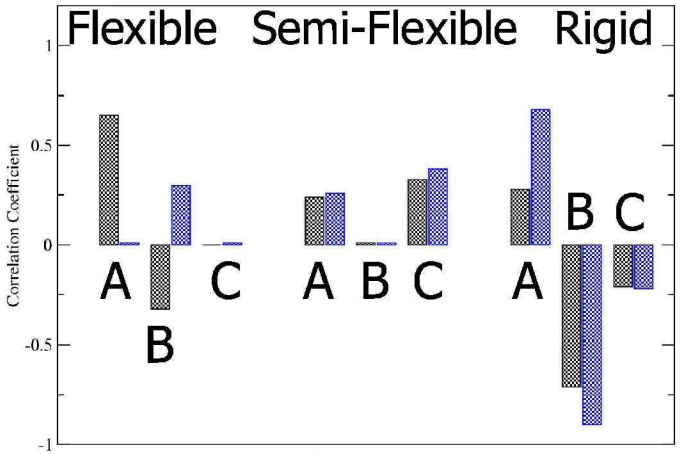
Water-Mediated Contacts

- Leading eigenvectors of (20x20) pair interaction matrices:
  - **A, B, C, ... – new canonical aminoacids.**
- Gray Histograms:
  - Correlations among the **native** interface aminoacid population vectors and **A**, **B**, and **C**.
- Blue Histograms:
  - Correlations among the binding **trap** aminoacid population vectors and **A**, **B**, and **C**.

A Large Differential In Canonical Aminoacid Composition Among Native And Trap States Leads To A Large Recognition Signal

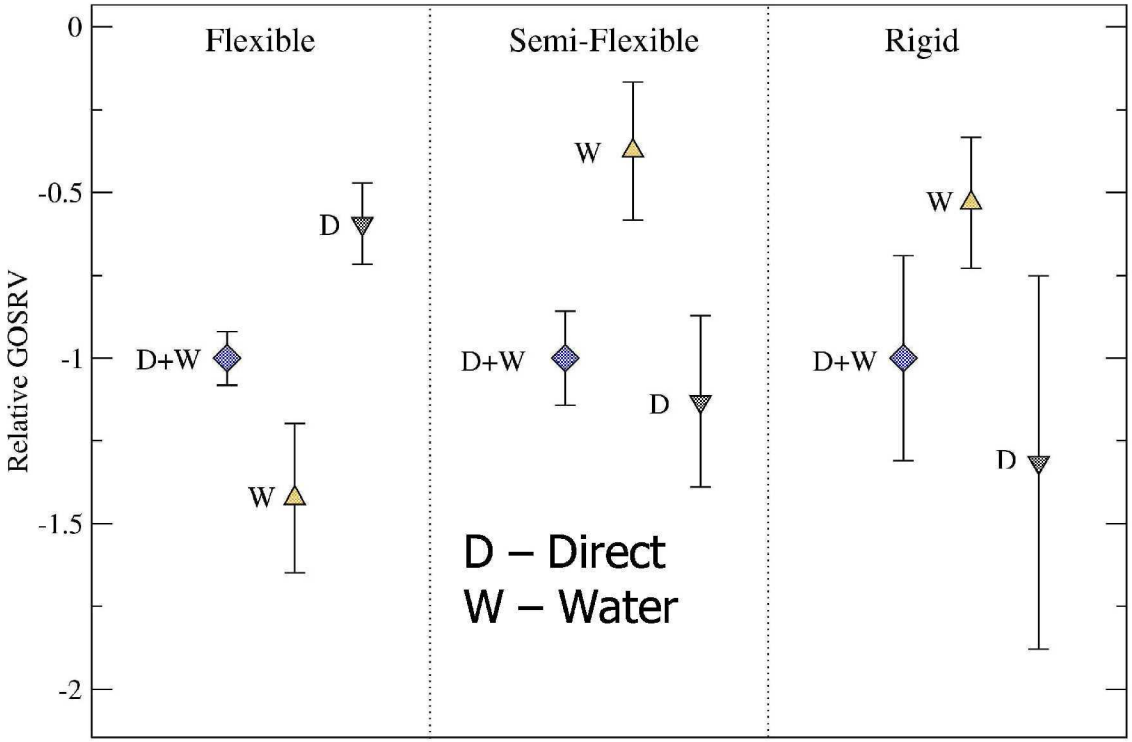


Direct Contacts



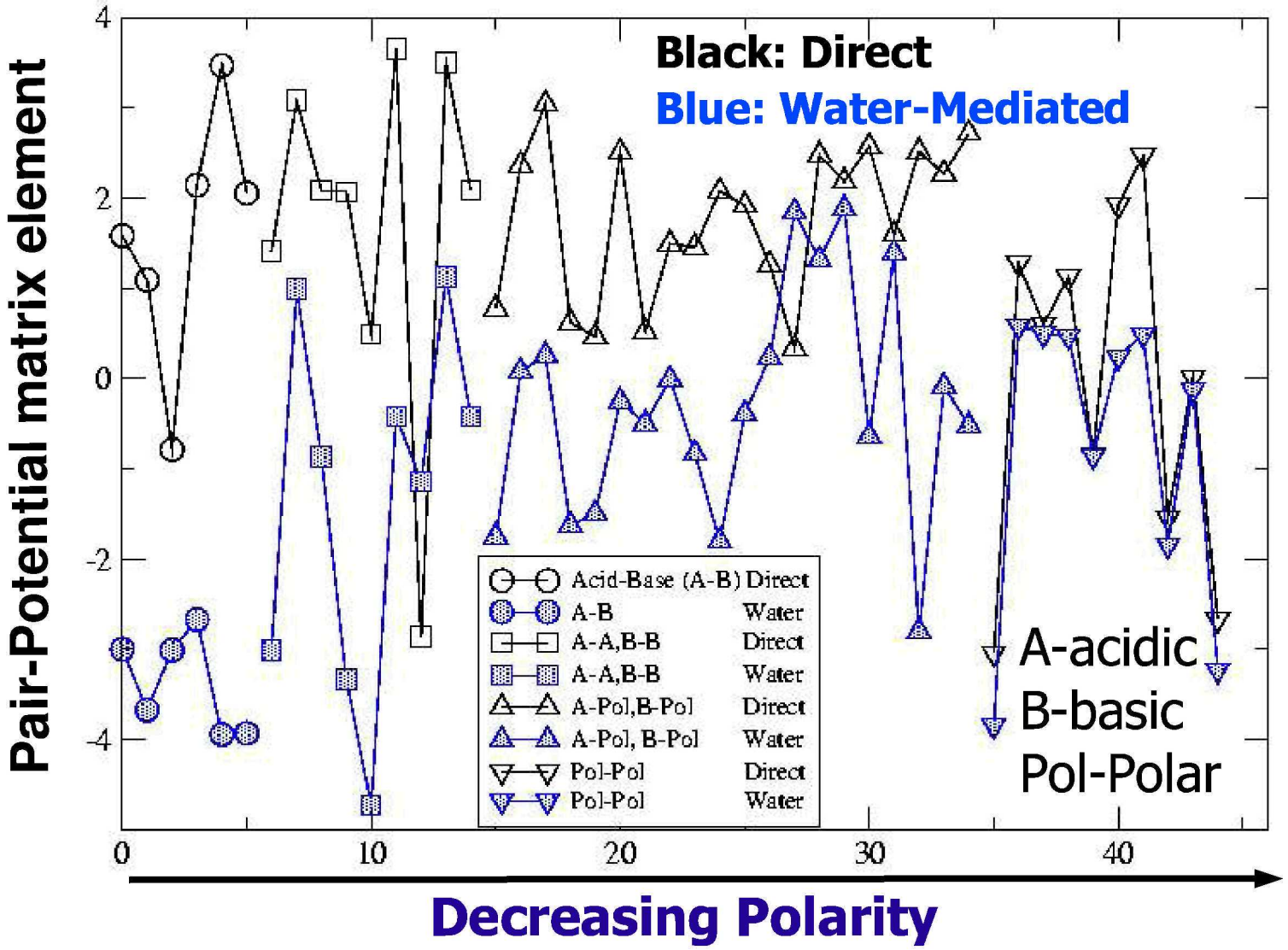
Water-Mediated Contacts

Recognition of Native Interfaces from Traps (negative is better)



Pair-Potentials Among Non-Hydrophobic Residue Pairs

Semi-Flexible Binding Trap Model



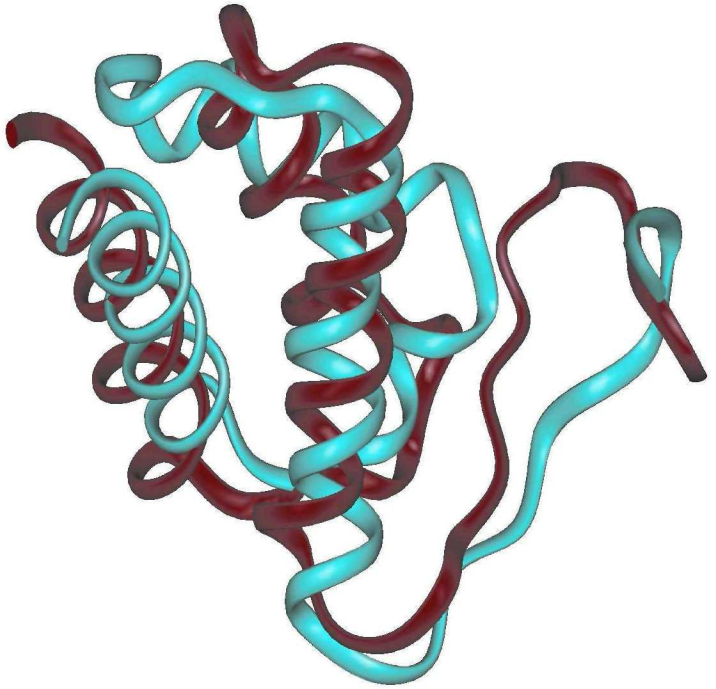
Reduced desolvation penalty favors water-mediated contacts among charged groups.

Protein Structure Prediction Potential

- **Coarse-Grained Modeling.**
- **Learn From Known Protein Structures (Memories):**
  - Pairwise contact potential among local residues (less than 12 residues apart in sequence).
- **Backbone:**
  - **Chain Connectivity Potential.**
  - **Chirality Potential.**
  - **Ramachandran Potential.**
- **Excluded Volume Potential.**
- **Long-Range Contact Potential.**

[Friedrichs and Wolynes 1989]

Example:
Blind Prediction in CASP5: 1H40



Q=0.45, RMDS=5.5 Å

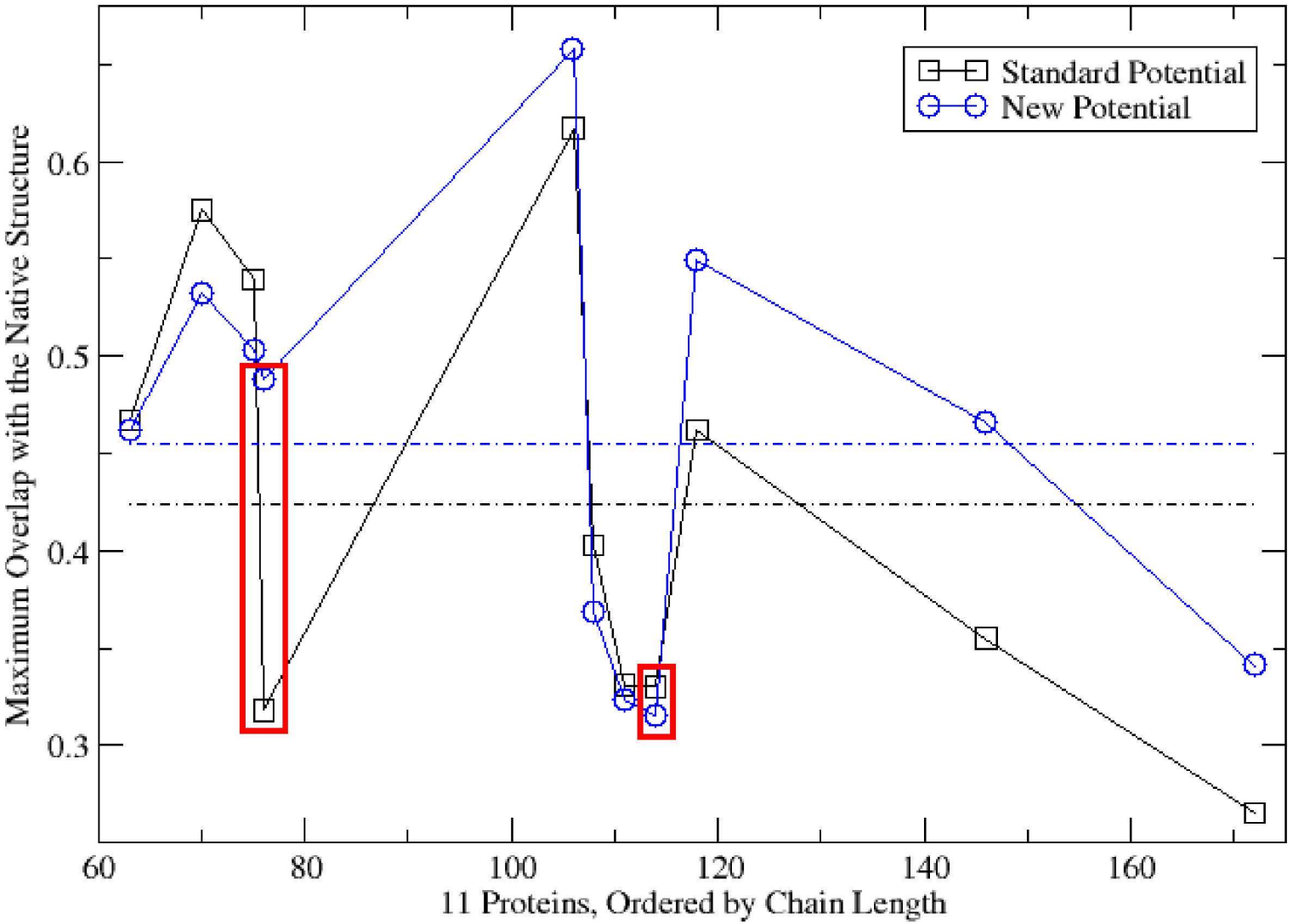
Q is a similarity measure to the native structure.

$$0 \leqslant Q \leqslant 1$$

Simulation Details

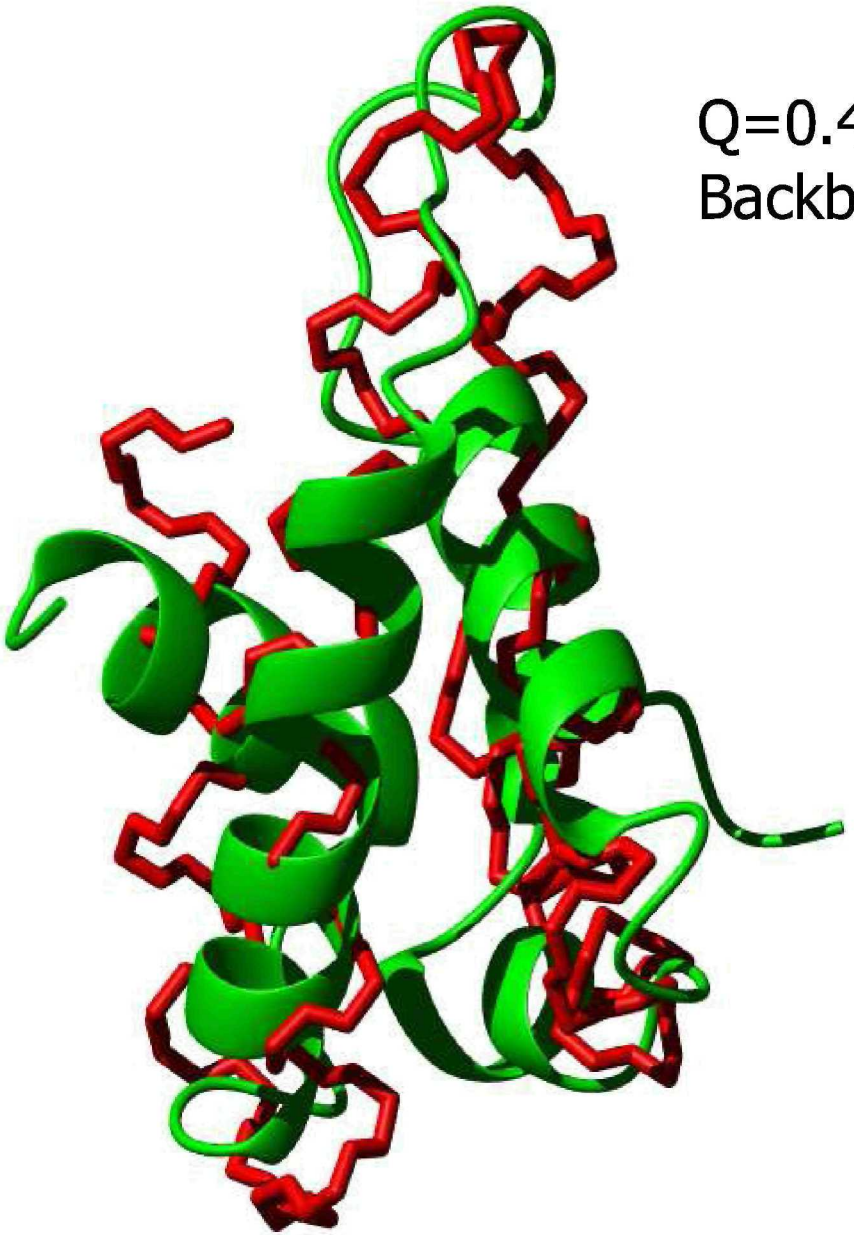
- 1-st contact well (4.5 to 6.5 Å):
  - "Folding" contact potential derived from a set of ~200 monomeric proteins.
- 2-nd contact well (6.5 to 9.5 Å):
  - Low-density regions – Flexible Water Potential
  - High-density regions - "Folding" potential
- 20-letter code for both 1^{st} and 2^{nd} wells.
- A radius of gyration constraint potential to keep the protein collapsed.
- 5 annealing runs for each protein:
  - 9 AMH training proteins and 2 test proteins (1bg8a & 1jwez)

Comparing the standard AMH contact potential with the new potential



*Proteins in red rectangles **are not training** proteins for the standard AMH potential.

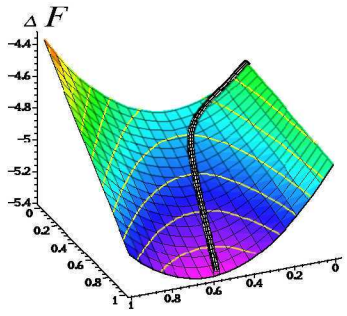
1bg8a: Best predicted structure vs X-ray structure



Q=0.49
Backbone RMSD: 5.0 Å

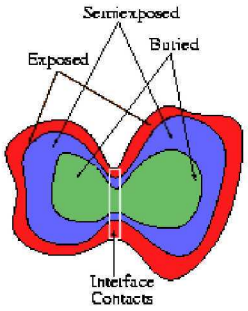
Summary

Theory of Binding and Folding:



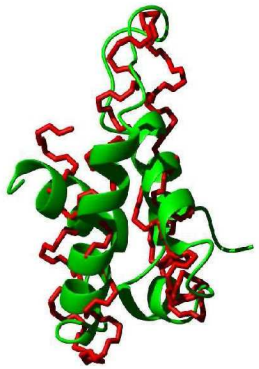
- ~15% of monomers (in the Protein Complex Database) need a partner to fold.
- For these proteins, native binding interactions pull down the folding funnel.

Knowledge-Based Binding Pair-Potentials:



- The combination of direct and water-mediated potentials provides the smoothest recognition.
- Reduced desolvation penalty favors water-mediated contacts among charged groups.

Water-Mediated Potential for Protein Structure Prediction:



- Water-mediated potentials improve significantly protein structure prediction.

1. G. A. Papoian and P. G. Wolynes, Biopolymes, 68, (2003), 333-349.
2. G. A. Papoian, J. Ulander, and P. G. Wolynes, J. Am. Chem. Soc., under review.

Acknowledgments

Peter G. Wolynes

Current & Former Wolynes

Group Members:

Michael Eastwood

Joachim Lätzer

Vassiliy Lubchenko

Steve Plotkin

Michael Prentiss

Johan Ulander

UCSD Colleagues:

Jose N. Onuchic

$$$ - NIH Fellowship in Quantitative Biology