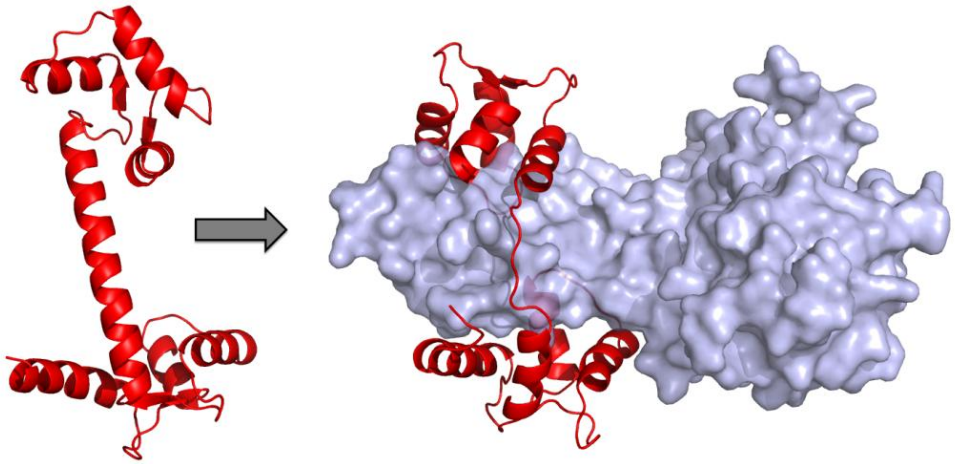
Reviews in Computational Biology:

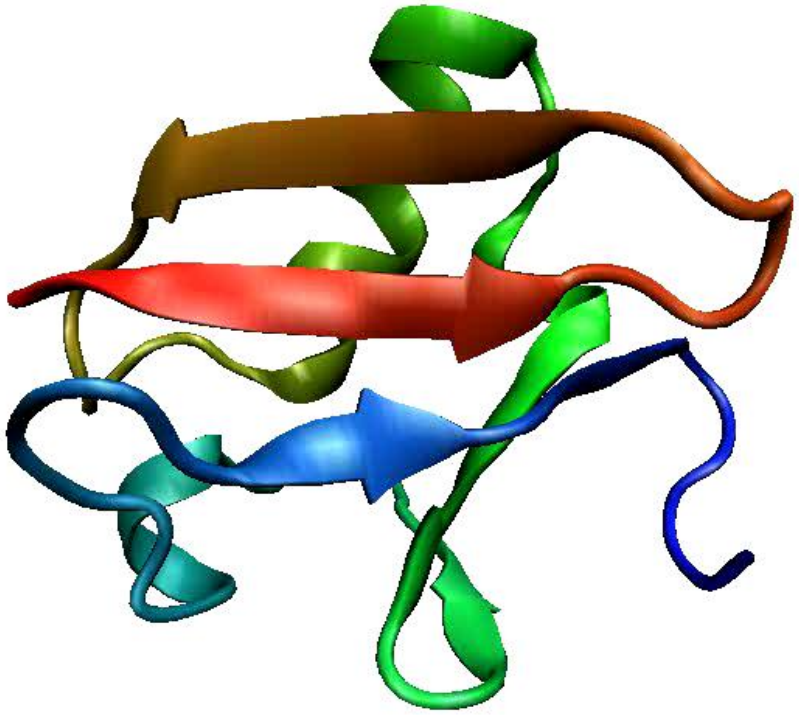
Relating the structural and evolutionary dynamics of proteins



Joseph Marsh

MRC Laboratory of Molecular Biology

15 February 2012

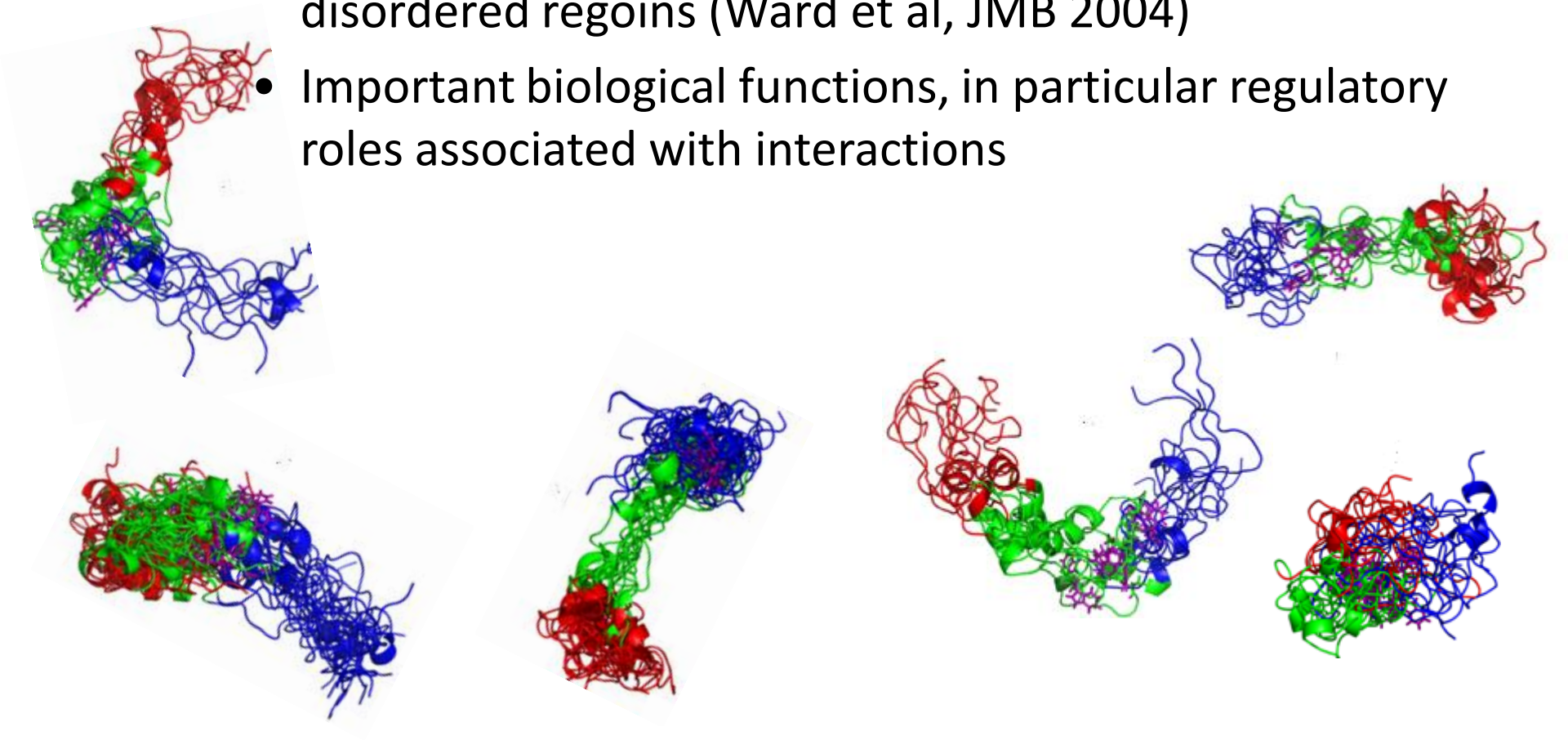


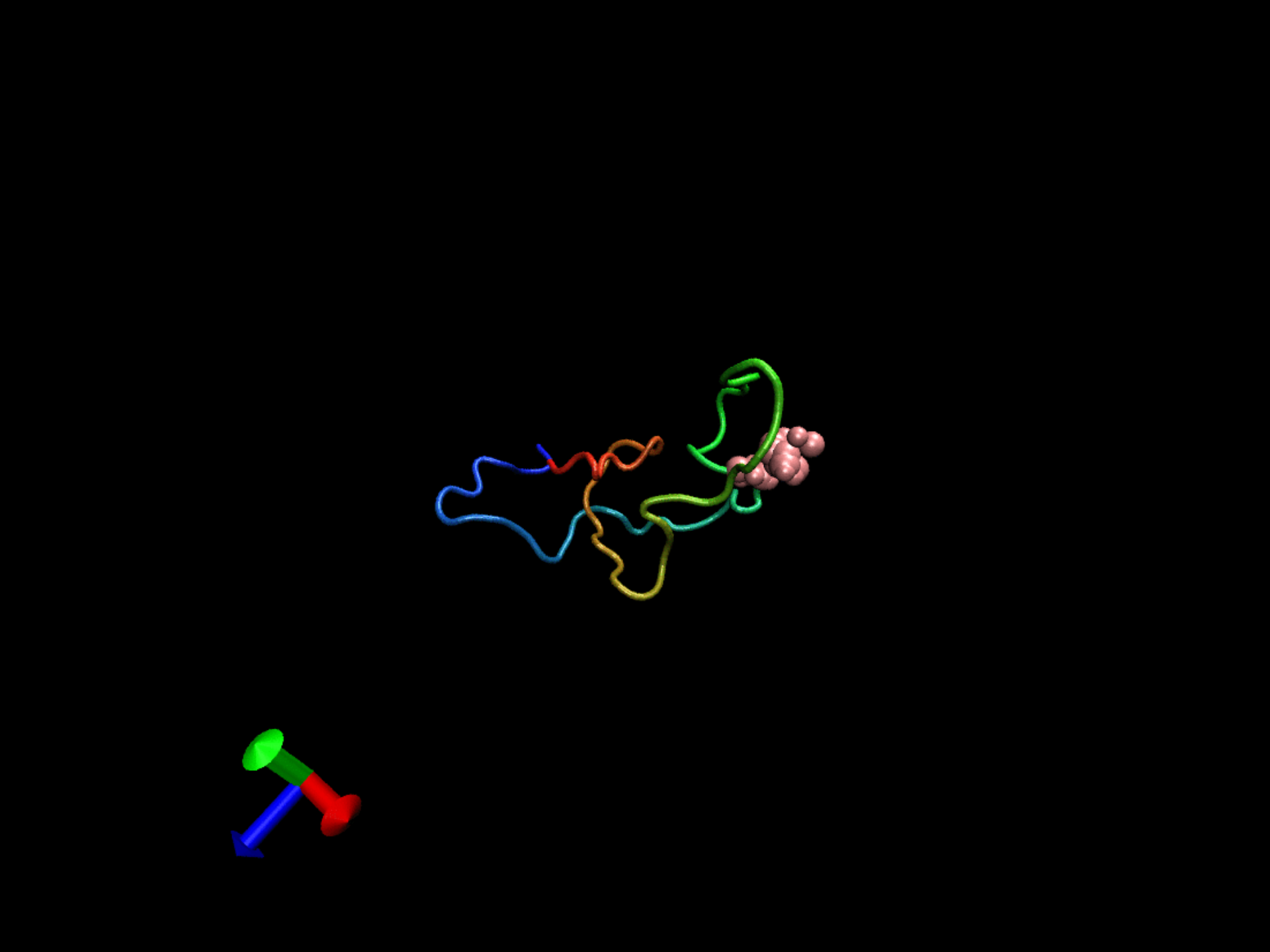
Lange et al. (2008)
Science 320:1471

"Everything that living things do can be understood in terms of the jigglings and wigglings of atoms." – Richard Feynman

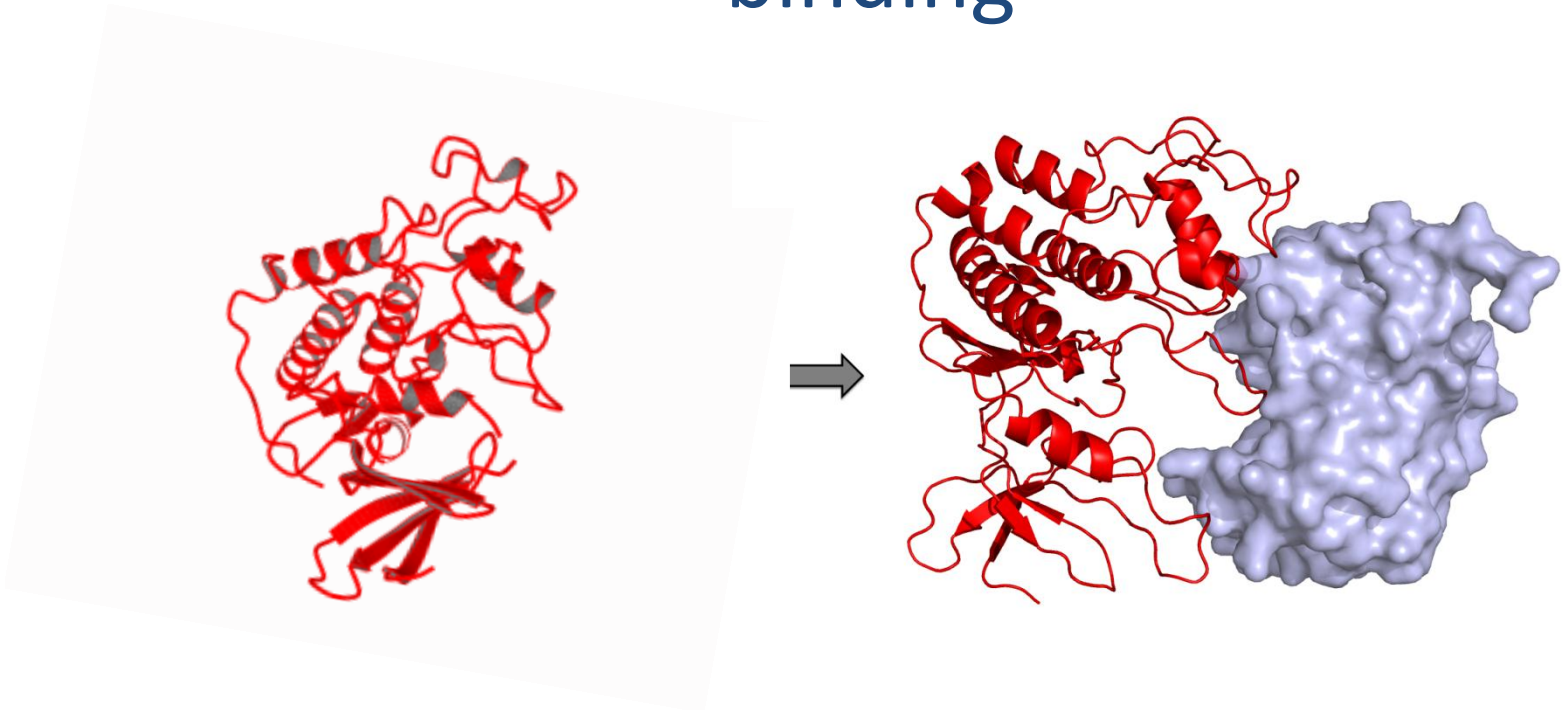
Intrinsically disordered proteins

- Highly flexible – lack stable, folded globular structure
- 1/3 of eukaryotic proteins predicted to contain long disordered regoins (Ward et al, JMB 2004)
- Important biological functions, in particular regulatory roles associated with interactions





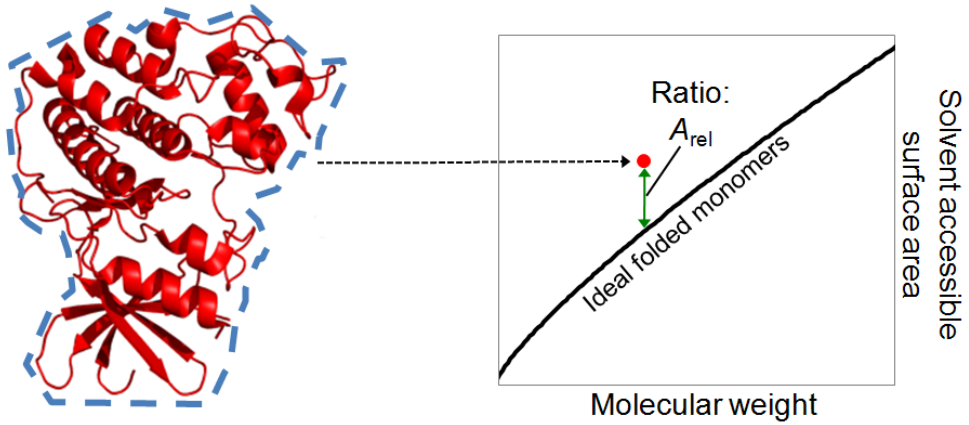
Proteins conformational changes upon binding



Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding

Joseph A. Marsh[1,*] and Sarah A. Teichmann[1]
[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
*Correspondence: jmarsh@mrc-lmb.cam.ac.uk
DOI 10.1016/j.str.2011.03.010

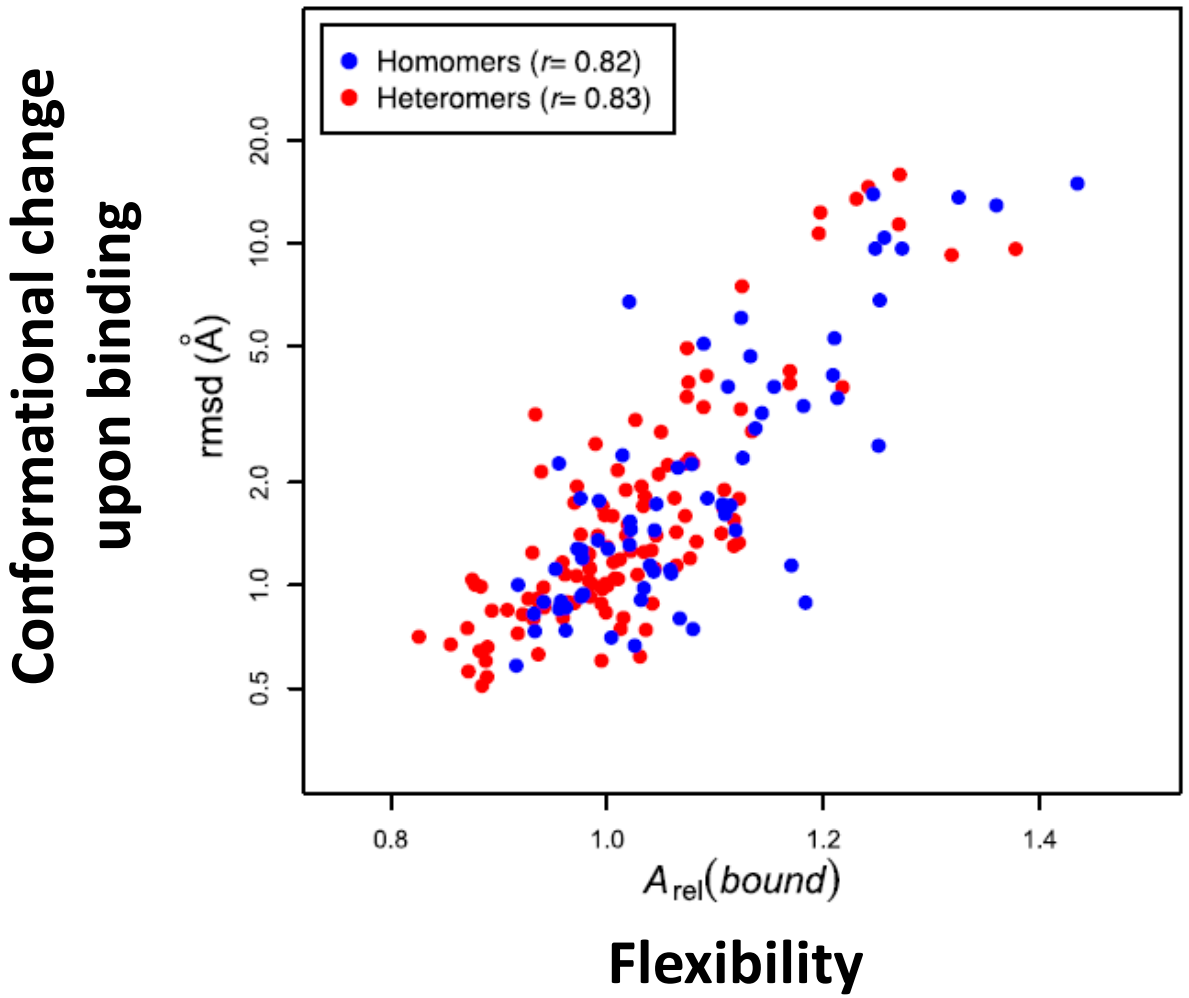
$$A_{rel} = \frac{A_s(observed)}{A_s(predicted)}$$



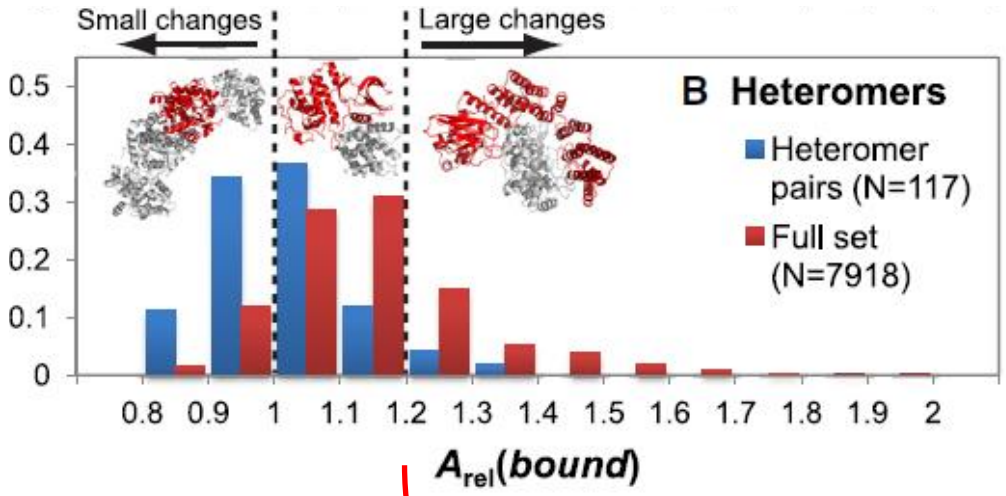
High A_{rel}=> exposing too much surface area for a folded monomer

-correlates strongly with intrinsic flexibility

Intrinsic flexibility strongly correlates with the magnitude of conformational change

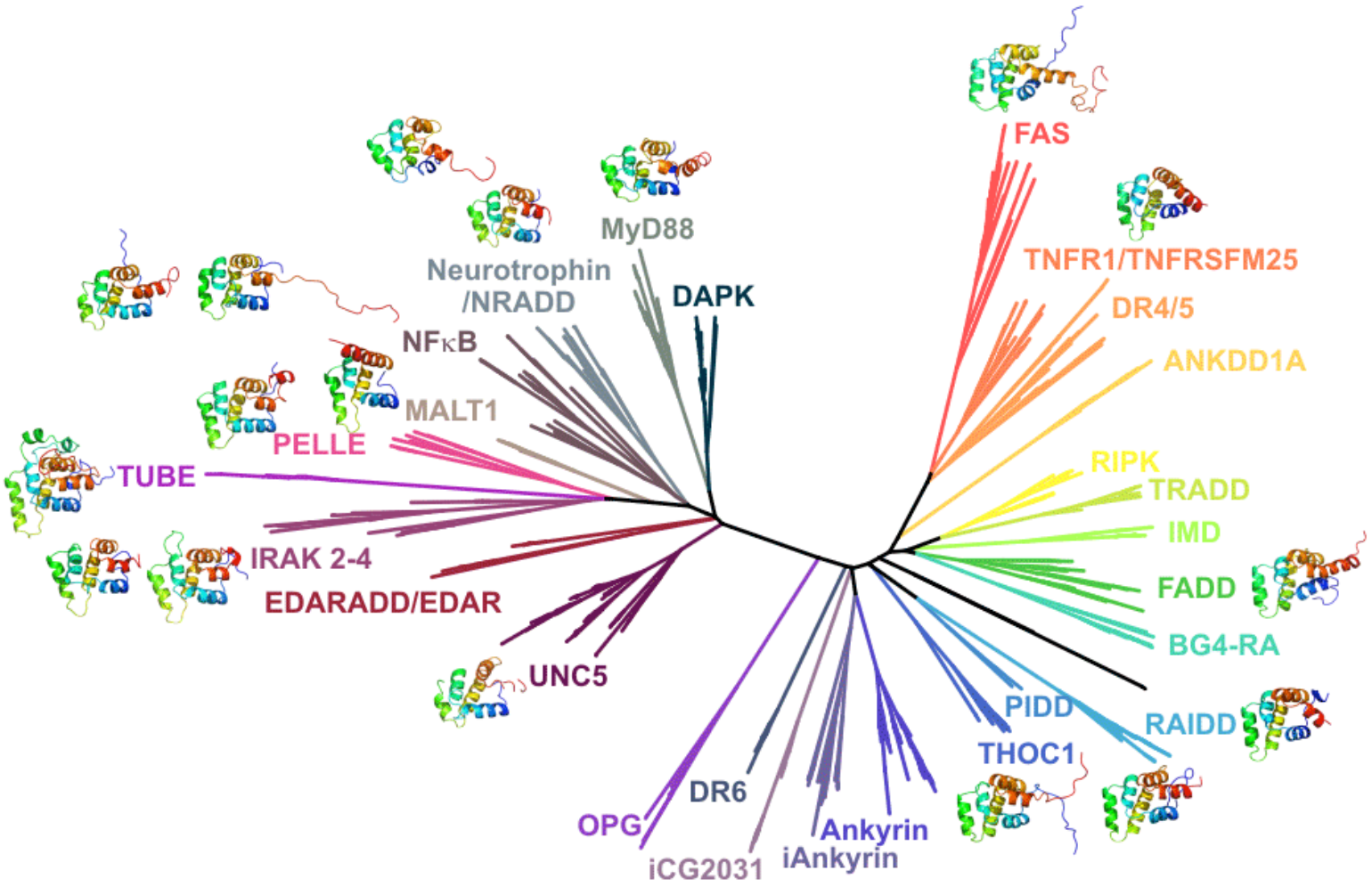


Large conformational changes upon binding are common



30% of subunits have A_{rel} > 1.2 and RMSD > 5 A

Evolutionary dynamics



http://www.nimr.mrc.ac.uk/research/paul-driscoll/

Science (1987) 235:318-321

Multiple Conformational States of Proteins: A Molecular Dynamics Analysis of Myoglobin

R. ELBER AND M. KARPLUS

A molecular dynamics simulation of myoglobin provides the first direct demonstration that the potential energy surface of a protein is characterized by a large number of thermally accessible minima in the neighborhood of the native structure (for example, approximately 2000 minima were sampled in a 300-picosecond trajectory). This is expected to have important consequences for the interpretation of the activity of transport proteins and enzymes. Different minima correspond to changes in the relative orientation of the helices coupled with side-chain rearrangements that preserve the close packing of the protein interior. The conformational space sampled by the simulation is similar to that found in the evolutionary development of the globins. Glasslike behavior is expected at low temperatures. The minima obtained from the trajectory do not satisfy certain criteria for ultrametricity.

"The conformational space sampled by the simulation is similar to that found in the evolutionary development of the globins."

Table 2. The relative motions of helices involved in contacts. The fluctuations in the distances (in angstroms) and angles (in degrees) between pairs of helices are given. Listed are the rms fluctuations, the maximum differences found in the set of minimized conformations, and the differences between structures in the globin series.

Helix contact	Trajectory				X-ray structures*	
	$\langle \Delta r^2 \rangle^{1/2}$	$\langle \Delta\theta^2 \rangle^{1/2}$	Δr_{max}	$\Delta\theta_{max}$	Δr	$\Delta\theta$
A-H	0.7	2	2.2	10	2.0	12
B-E	0.4	14	1.6	39	3.0	25
B-G	0.4	6	1.9	25	6.2†	30
F-H	0.4	1	1.8	5	1.4†	10†
G-H	0.3	2	1.3	7	2.5	15

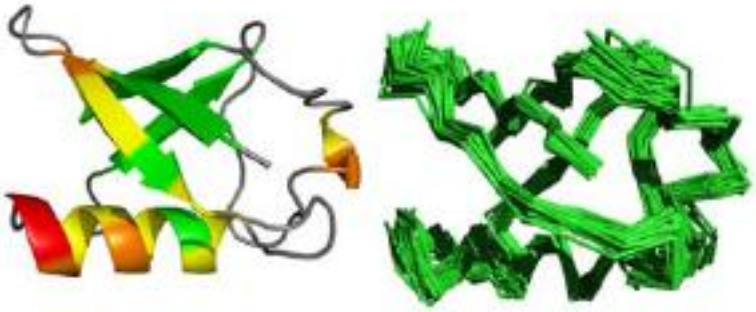
OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

A Correspondence Between Solution-State Dynamics of an Individual Protein and the Sequence and Conformational Diversity of its Family

Gregory D. Friedland[1,2,3], Nils-Alexander Lakomek[4], Christian Griesinger[4], Jens Meiler[5]*, Tanja Kortemme[1,2,3]*

1 Graduate Group in Biophysics, University of California San Francisco, San Francisco, California, United States of America, **2** Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, United States of America, **3** California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America, **4** Department for NMR-based Structural Biology, Max-Planck Institute for Biophysical Chemistry, Goettingen, Germany, **5** Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America

Conformational ensemble describing dynamics of a single sequence



Intrinsic dynamics

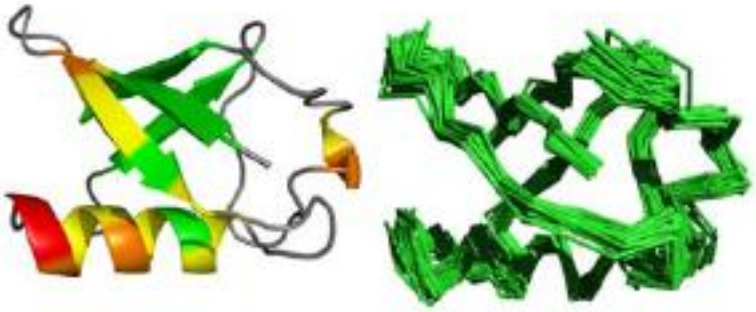
OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

A Correspondence Between Solution-State Dynamics of an Individual Protein and the Sequence and Conformational Diversity of its Family

Gregory D. Friedland[1,2,3], Nils-Alexander Lakomek[4], Christian Griesinger[4], Jens Meiler[5]*, Tanja Kortemme[1,2,3]*

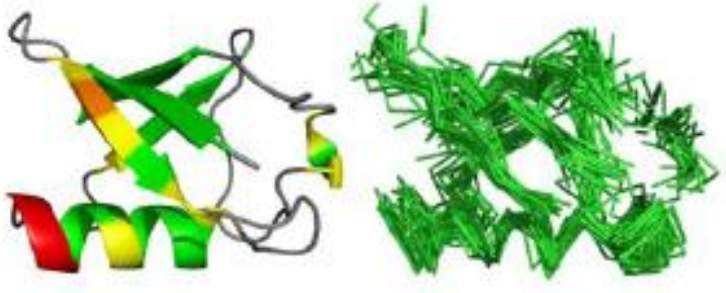
1 Graduate Group in Biophysics, University of California San Francisco, San Francisco, California, United States of America, **2** Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, United States of America, **3** California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America, **4** Department for NMR-based Structural Biology, Max-Planck Institute for Biophysical Chemistry, Goettingen, Germany, **5** Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America

Conformational ensemble describing dynamics of a single sequence



Intrinsic dynamics

Ensemble of X-ray structures describing conformational variability explored by a protein family



Evolutionary dynamics

Evolutionary dynamics resemble intrinsic dynamics!

Evidence from flexible backbone protein design...

doi:10.1016/j.jmb.2004.11.062 J. Mol. Biol. (2005) 346, 631–644

JMB

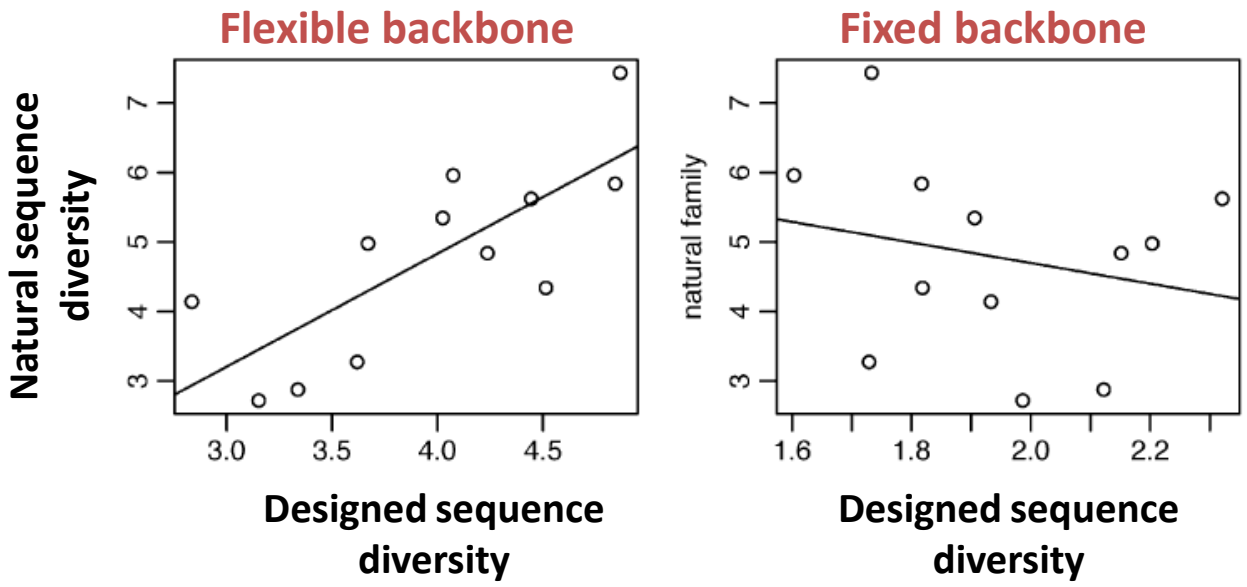
Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Recapitulation of Protein Family Divergence using Flexible Backbone Protein Design

Christopher T. Saunders[1] and David Baker[2*]



Incorporating backbone flexibility into flexible design captures the natural sequence diversity of protein families

doi:10.1016/j.jmb.2004.11.062 J. Mol. Biol. (2005) 346, 631–644

JMB

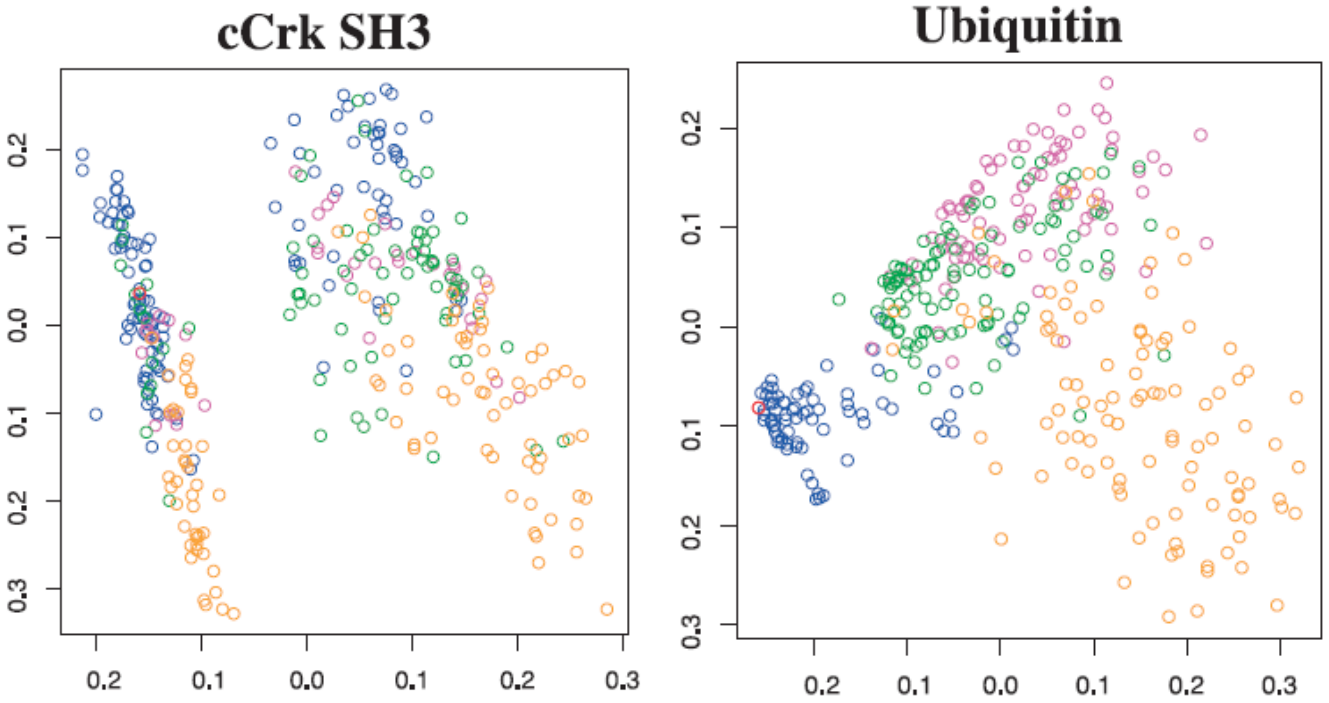
Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Recapitulation of Protein Family Divergence using Flexible Backbone Protein Design

Christopher T. Saunders[1] and David Baker[2]*

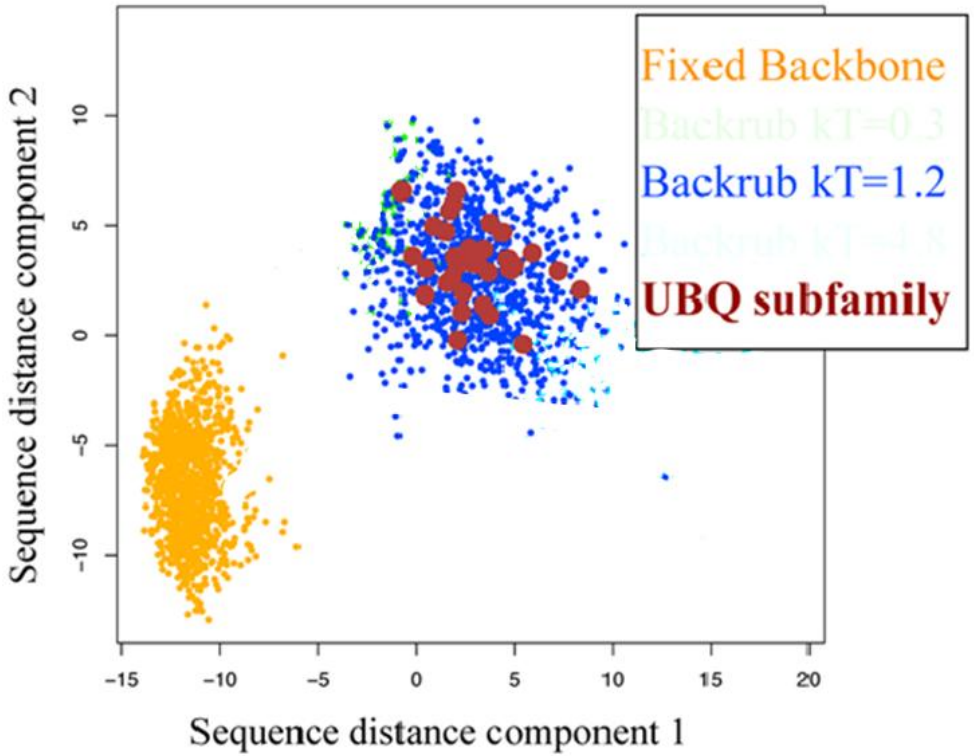


OPEN ACCESS Freely available online PLOS COMPUTATIONAL BIOLOGY

A Correspondence Between Solution-State Dynamics of an Individual Protein and the Sequence and Conformational Diversity of its Family

Gregory D. Friedland[1,2,3], Nils-Alexander Lakomek[4], Christian Griesinger[4], Jens Meiler[5]*, Tanja Kortemme[1,2,3]*

1 Graduate Group in Biophysics, University of California San Francisco, San Francisco, California, United States of America, 2 Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, United States of America, 3 California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America, 4 Department for NMR-based Structural Biology, Max-Planck Institute for Biophysical Chemistry, Goettingen, Germany, 5 Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America



Accurate ensemble models do even better!

Evolution of intrinsically disordered proteins...

J Mol Evol (2002) 55:104–110
DOI: 10.1007/s00239-001-2309-6

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary Rate Heterogeneity in Proteins with Long Disordered Regions

Celeste J. Brown,[1] Sachiko Takayama,[1] Andrew M. Campen,[1] Pam Vise,[1] Thomas W. Marshall,[1] Christopher J. Oldfield,[1] Christopher J. Williams,[2] A. Keith Dunker[1]

[1] School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA
[2] Division of Statistics, University of Idaho, Moscow, ID 83844, USA

Received: 10 September 2001 / Accepted: 7 January 2002

Table 1. Average difference in genetic distance, Δ, between ordered and disordered regions of 26 protein families

Protein family	Reference	Detection method[a]	No. sequences	Δ[b]	*p* value[c]
Replication protein A	Jacobs et al. (1999)	NMR	7	−1.92	0.001
NF-KB p65	Schmitz et al. (1994)	NMR	4	−1.18	0.001
Glycyl-tRNA synthetase	Logan et acl. (1995)	X-Ray	24	−1.69	0.002
Regulator of G-protein signaling 4	Tesmer et al. (1997)	X-Ray	17	−0.96	0.001
Topoisomerase II	Berger et al. (1996)	X-Ray	28	−0.87	0.001
Calcineurin	Kissinger et al. (1995)	X-Ray	23	−0.84	0.001
c-Fos	Campbell et al (2000)	NMR	23	−0.82	0.001
Thyroid transcription factor	Tell et al. (1998)	CD, LP	12	−0.76	0.001
Sulfotransferase	Bidwell et al. (1999)	X-Ray	12	−0.74	0.013
Phenylalanine-tRNA synthetase	Mosyak et al. (1995)	X-Ray	14	−0.69	0.001
Coat protein, tomato bushy stunt virus	Hopper et al. (1984)	X-Ray	7	−0.63	0.001
Gonadotropin	Lapthorn et al. (1994)	X-Ray	9	−0.61	0.001
Coat protein, Sindbis virus	Choi et al. (1991)	X-Ray	6	−0.60	0.025
Histone H5	Aviles et al.(1978)	NMR	9	−0.41	0.001
Small heat shock protein	Kim et al. (1998)	X-Ray	6	−0.36	0.457
Telomere binding protein	Horvath et al. (1998)	X-Ray	8	−0.29	0.001
Cytochrome BC1	Iwata et al. (1998)	X-Ray	7	−0.27	0.034
DNA-lyase	Gorman et al. (1997)	X-Ray[d]	8	−0.18	0.001
Bcl-xL	Muchmore et al. (1996)	X-Ray, NMR	7	−0.13	0.001
Coat protein, southern bean mosaic virus	Silva and Rossmann (1985)	X-Ray	6	−0.09	0.100
α-Tubulin	Jimenez et al. (1999)	NMR	80	−0.06	0.034
Epidermal growth factor	Louie et al. (1997)	X-Ray	10	−0.03	0.736
Prion	Riek et al. (1997)	NMR	72	0.03	0.636
Glycine N-methyltransferase	Huang et al. (2000)	X-Ray	11	0.09	0.095
ssDNA binding protein	Tucker et al. (1994)	X-Ray	20	0.37	0.010
Flagellin	Vonderviszt et al. (1989)	LP	34	0.66	0.023

Disordered regions evolved more rapidly in 22/26 families studies

Genetics (1998) 149:445

Assessing the Impact of Secondary Structure and Solvent Accessibility on Protein Evolution

Nick Goldman,* Jeffrey L. Thorne† and David T. Jones‡

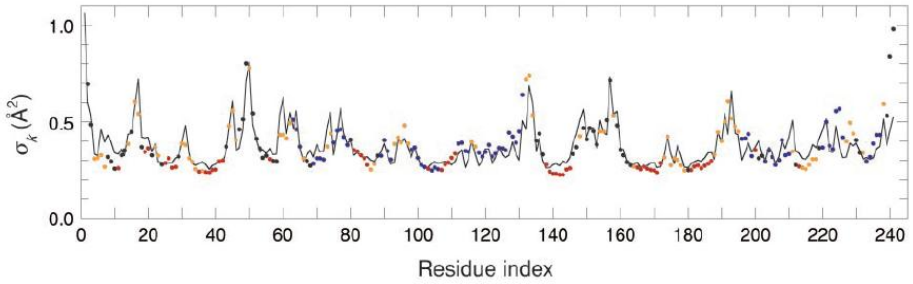
*Department of Genetics, University of Cambridge, Cambridge CB2 3EH, United Kingdom, †Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203 and ‡Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

"…the solvent accessibility status of a site has a particularly strong association with the process of amino acid replacement that it experiences."

Flexibility and packing in proteins

Bertil Halle*

Department of Biophysical Chemistry, Lund University, Box 124, SE-22100 Lund, Sweden

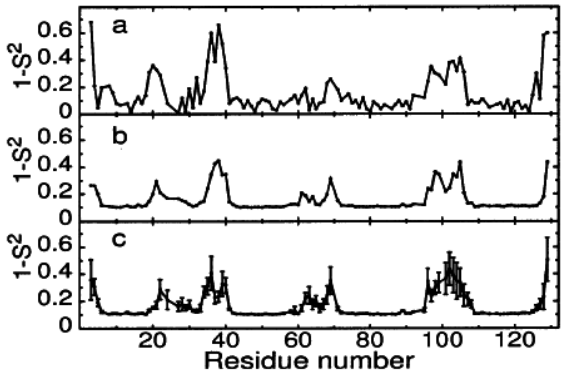


Contact Model for the Prediction of NMR N−H Order Parameters in Globular Proteins

Fengli Zhang and Rafael Brüschweiler*

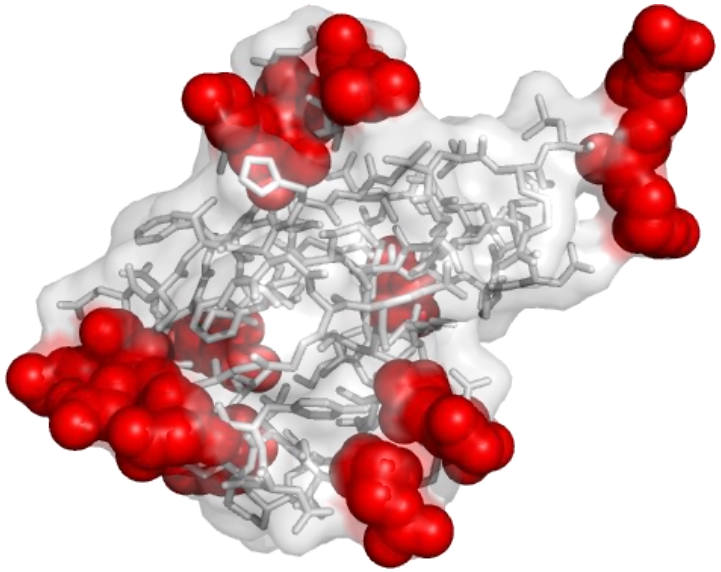
Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts 01610

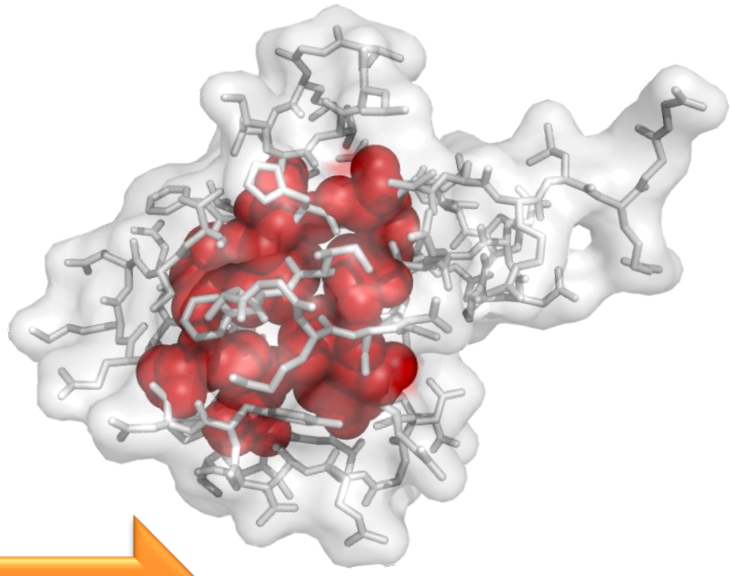
Received July 24, 2002



Flexibility is largely determined by local contact density!

A simple explanation for the correspondence between structural and evolutionary dynamics?





Contact density

Less constrained:
-more flexible
-faster evolution

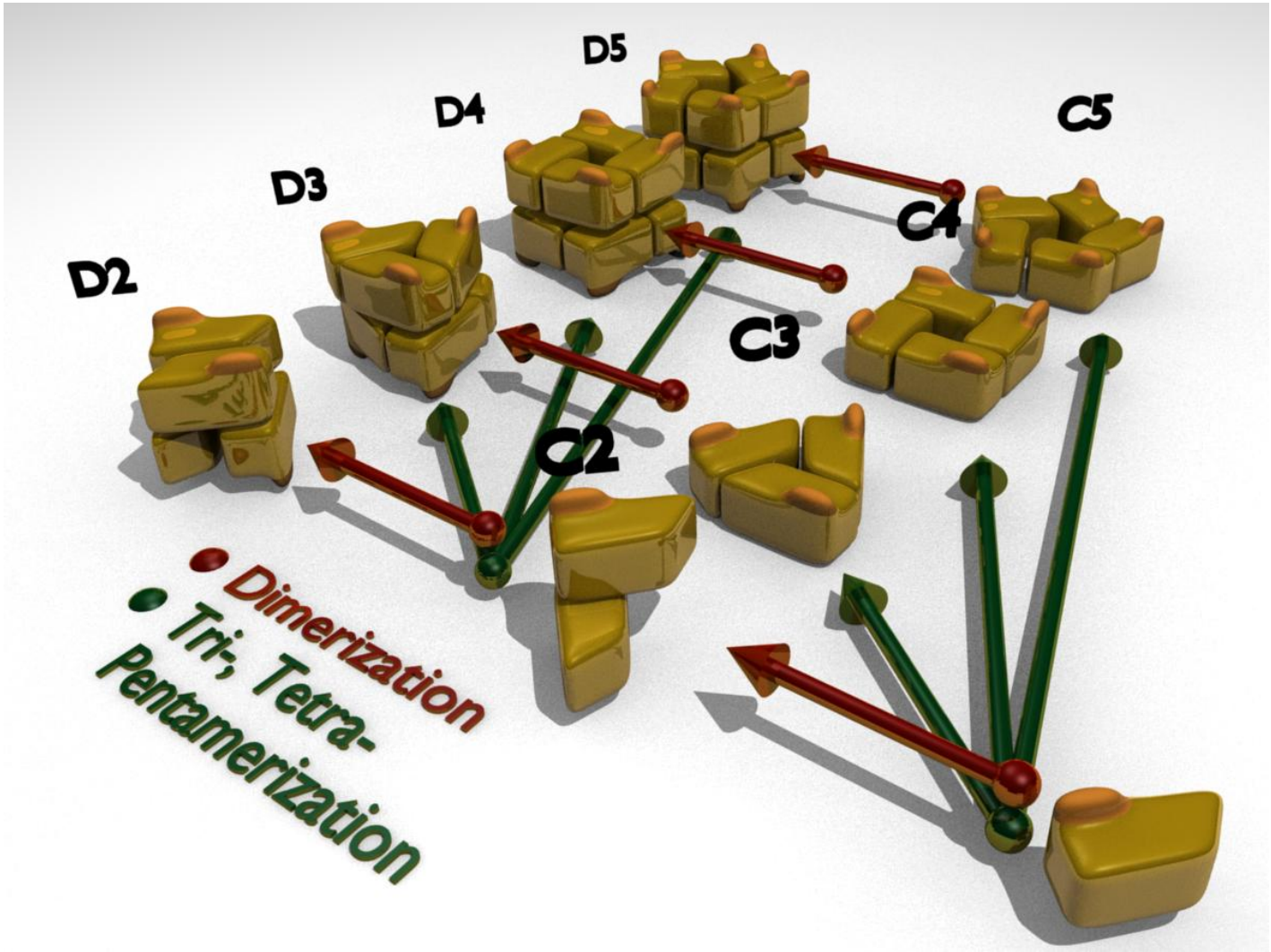
More constrained:
-less flexible
-slower evolution

nature Vol 453|26 June 2008|doi:10.1038/nature06942

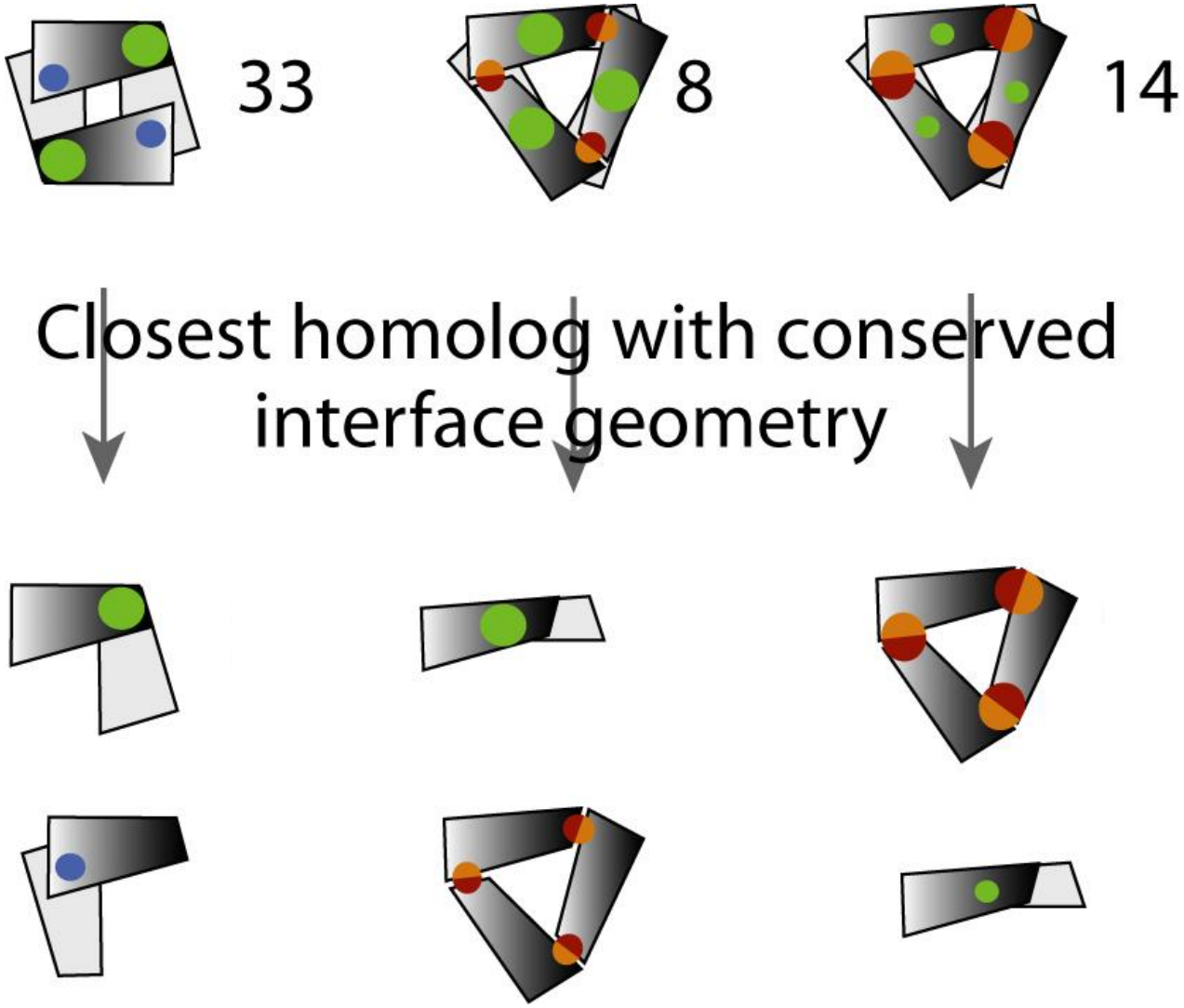
LETTERS

Assembly reflects evolution of protein complexes

Emmanuel D. Levy[1], Elisabetta Boeri Erba[2], Carol V. Robinson[2] & Sarah A. Teichmann[1]

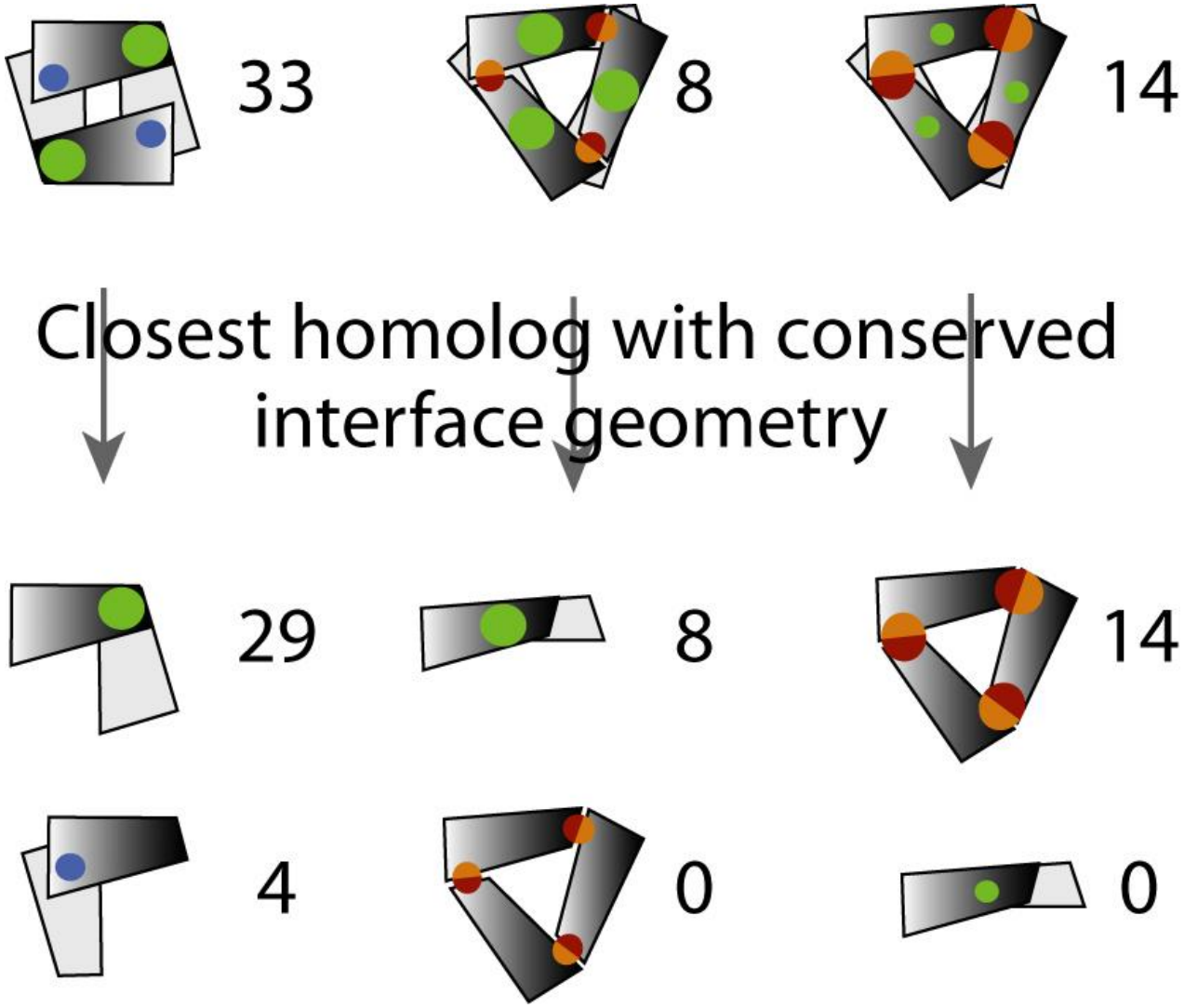


Homology to the largest interface is common

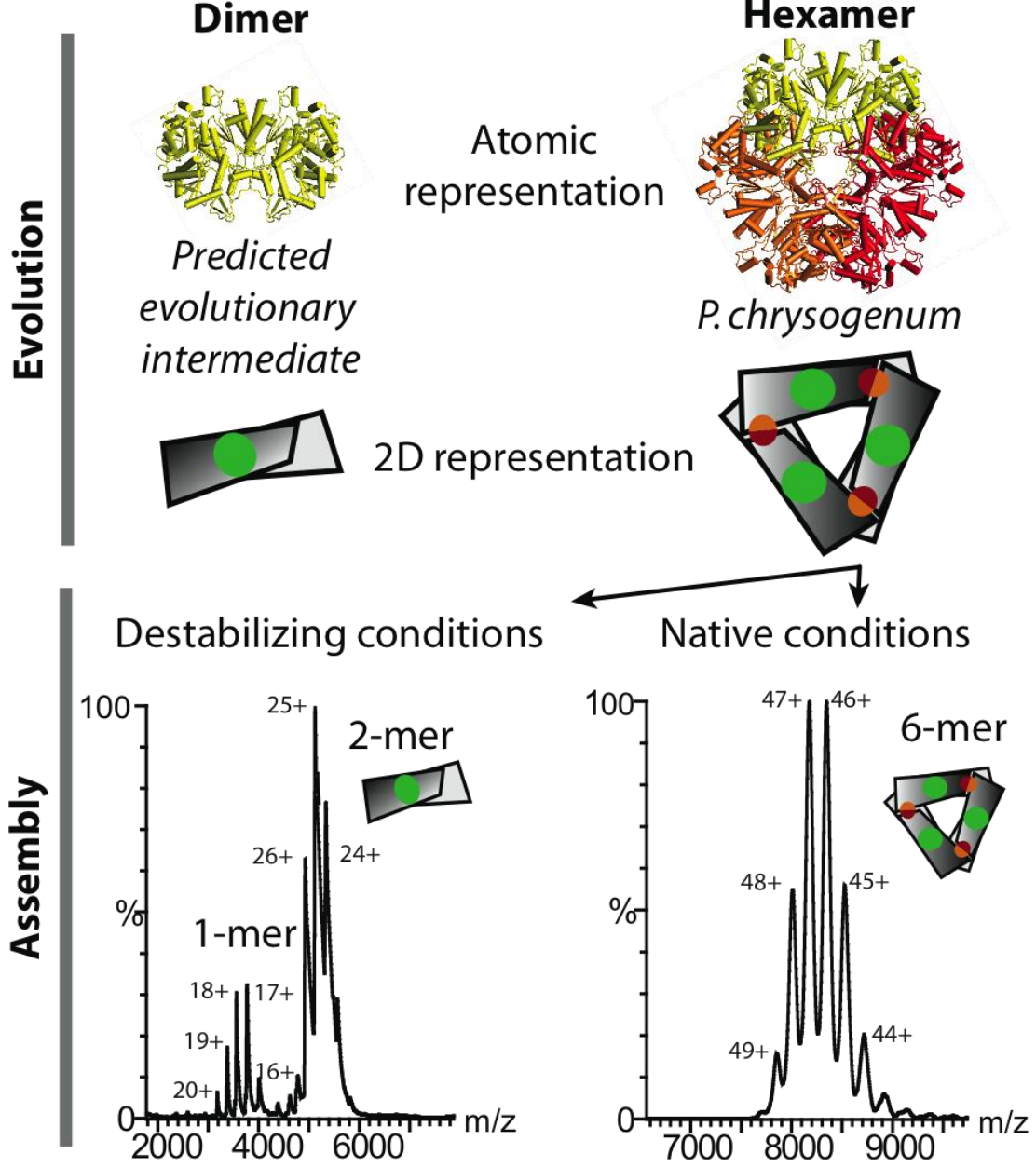


Evolutionary pathways can be predicted!

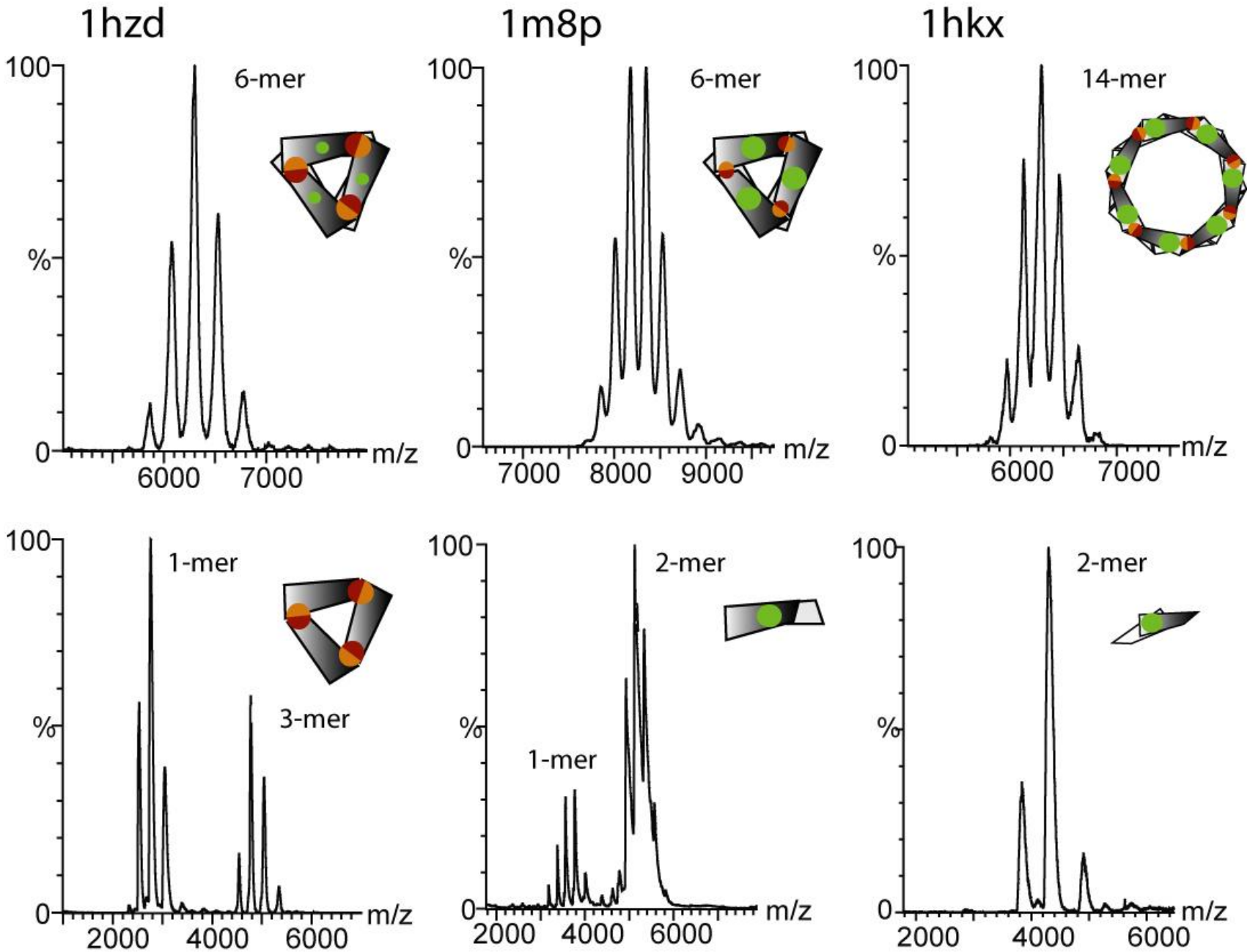
Homology to the largest interface is common



Evolutionary pathways can be predicted!


Dimer
Hexamer
Evolution
Atomic representation
Predicted evolutionary intermediate
P. chrysogenum
2D representation
Assembly
Destabilizing conditions
Native conditions
100
%
0
2000
4000
6000
m/z
25+
26+
24+
2-mer
1-mer
18+
17+
19+
16+
20+
7000
8000
9000
47+
46+
6-mer
48+
45+
49+
44+

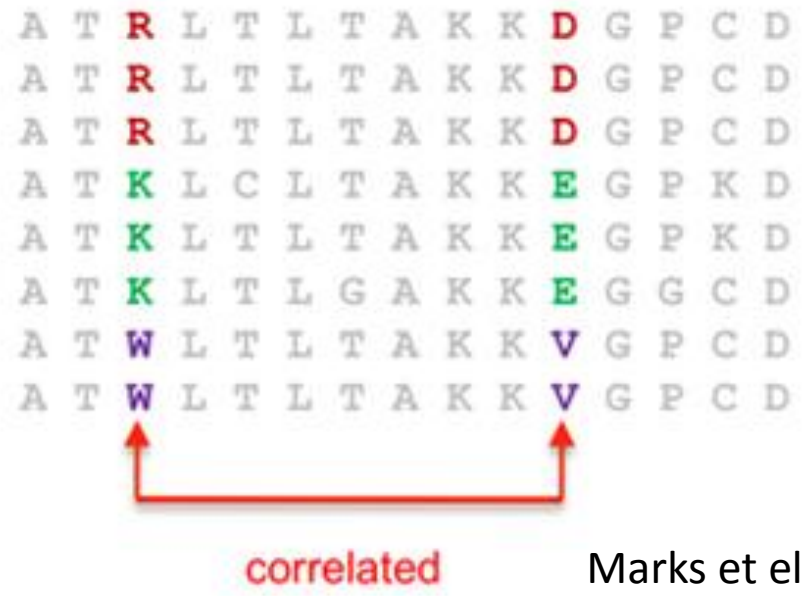
Macromolecular mass spectrometry



Complexes with characterised assembly pathways - agreement with predictions

	PDB id	Description	Number of chains	Predicted intermediate	Observed intermediate	Agreement with prediction
a	1m8p	ATP sulfurylase	6	2	2	✓
	1vea	RNA Binding AP	6	3	-	-
	1hzd	AUH protein	6	3	3	✓
	1j2p	Proteasome α-ring	7	-	-	✓
	1umg	Fructose 1,6 biPase	8	2	2	✓
	1m3u	KHMase	10	2	2	✓
	1pvv	OCTase	12	3	**3**;6	✓
	1hkx	Ca2+ dept. kinase	14	2	2	✓
b	1pfk	PFK I	4	2	2	✓
	1nhk	NDK	4	2	-	-
	1t3d	Serine AT	6	3	3	✓
	1aon	GroES	7	-	-	✓
	1di0	Lumazine Synth.	10	5	5	✓
	1ogf	rbsD	10	5	5	✓

Residue co-evolution: protein dynamics and allostery



Marks et el. PLoS ONE, 2011

Residue co-evolution: protein dynamics and allostery

Science (1999) 286:295

Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families

Steve W. Lockless and Rama Ranganathan*

From Wikipedia:

Statistical coupling analysis or **SCA** is a technique used in bioinformatics to measure covariation between pairs of amino acids in a protein multiple sequence alignment (MSA). More specifically, it quantifies how much the amino acid distribution at some position *i* changes upon a perturbation of the amino acid distribution at another position *j*. The resulting **statistical coupling energy** indicates the degree of evolutionary dependence between the residues, with higher coupling energy corresponding to increased dependence.

$$\Delta\Delta G_{i,j}^{stat} = \sqrt{\sum_{x} (ln P_{i|\delta j}^{x} - ln P_{i}^{x})^2}$$

Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families

Steve W. Lockless and Rama Ranganathan*

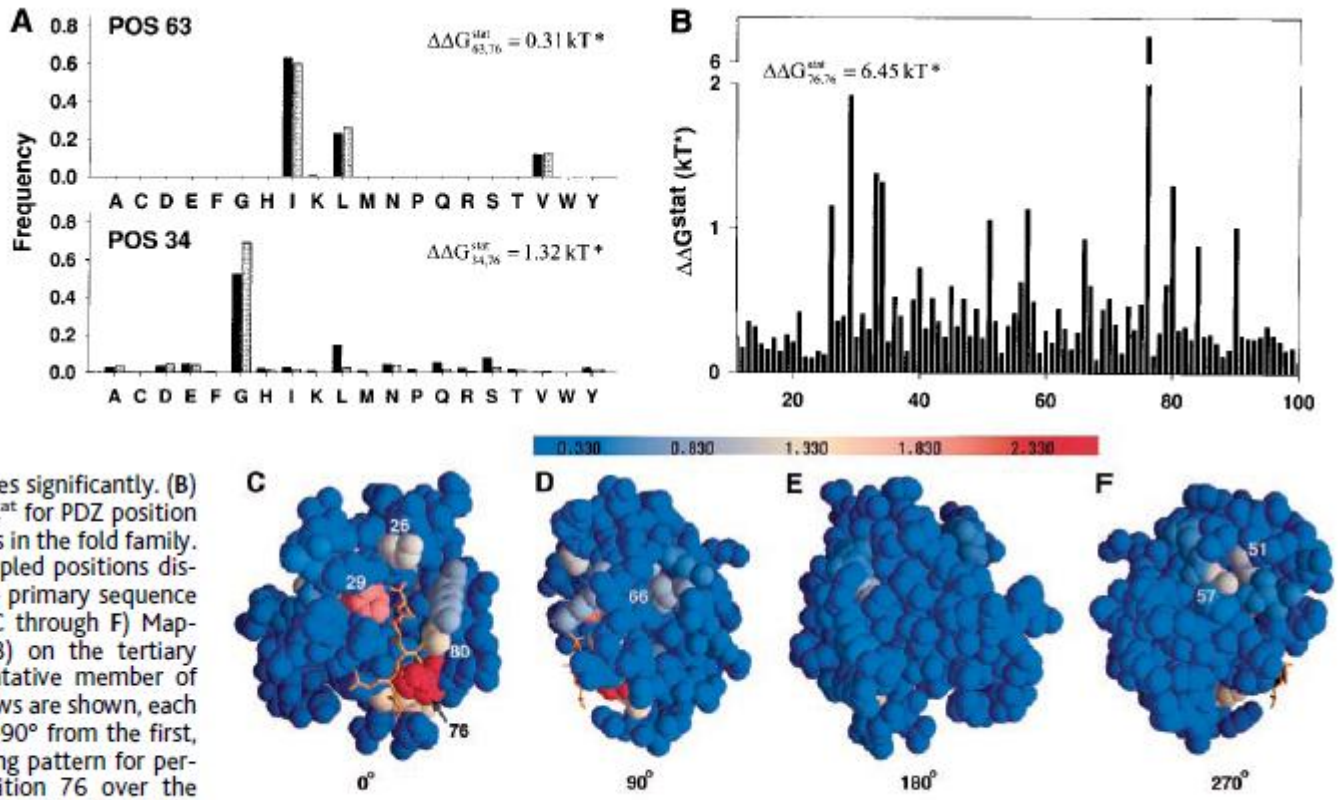


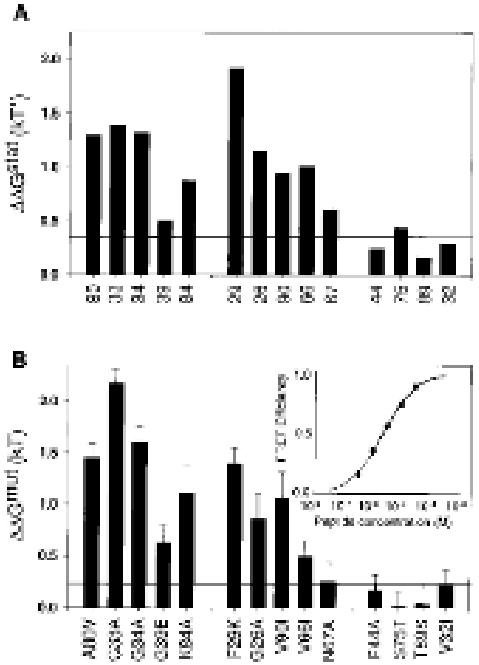
Fig. 2. Statistical coupling for a single site in the PDZ domain family. (A) Examples of amino acid distributions for two PDZ domain sites before (black bars) and after (gray bars) a 6.45 kT^* perturbation at position 76. The distribution at position 63 changes very little upon perturbation at position 76, despite high overall conservation, and the distribution at position 34 changes significantly. (B) A full mapping of $\Delta\Delta G^{stat}_{i,j}$ for PDZ position 76 for all other positions in the fold family. Only a small set of coupled positions distributed throughout the primary sequence emerge above noise. (C through F) Mapping of the data in (B) on the tertiary structure of a representative member of the fold family. Four views are shown, each successively rotated by 90° from the first, of the statistical coupling pattern for perturbations at PDZ position 76 over the entire PDZ domain. Coupled positions describe energetic interactions at sites spatially close to and distant from the point of perturbation. The color scale ranges from blue (0.330 kT^*) to red (2.330 kT^*).

Tightly coupled energetic network between distant residues

Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families

Steve W. Lockless and Rama Ranganathan*

Statistical coupling



Double mutant cycle coupling

Support from NMR experiments…

doi:10.1016/j.jmb.2003.11.010 *J. Mol. Biol.* (2004) **335**, 1105–1115

JMB

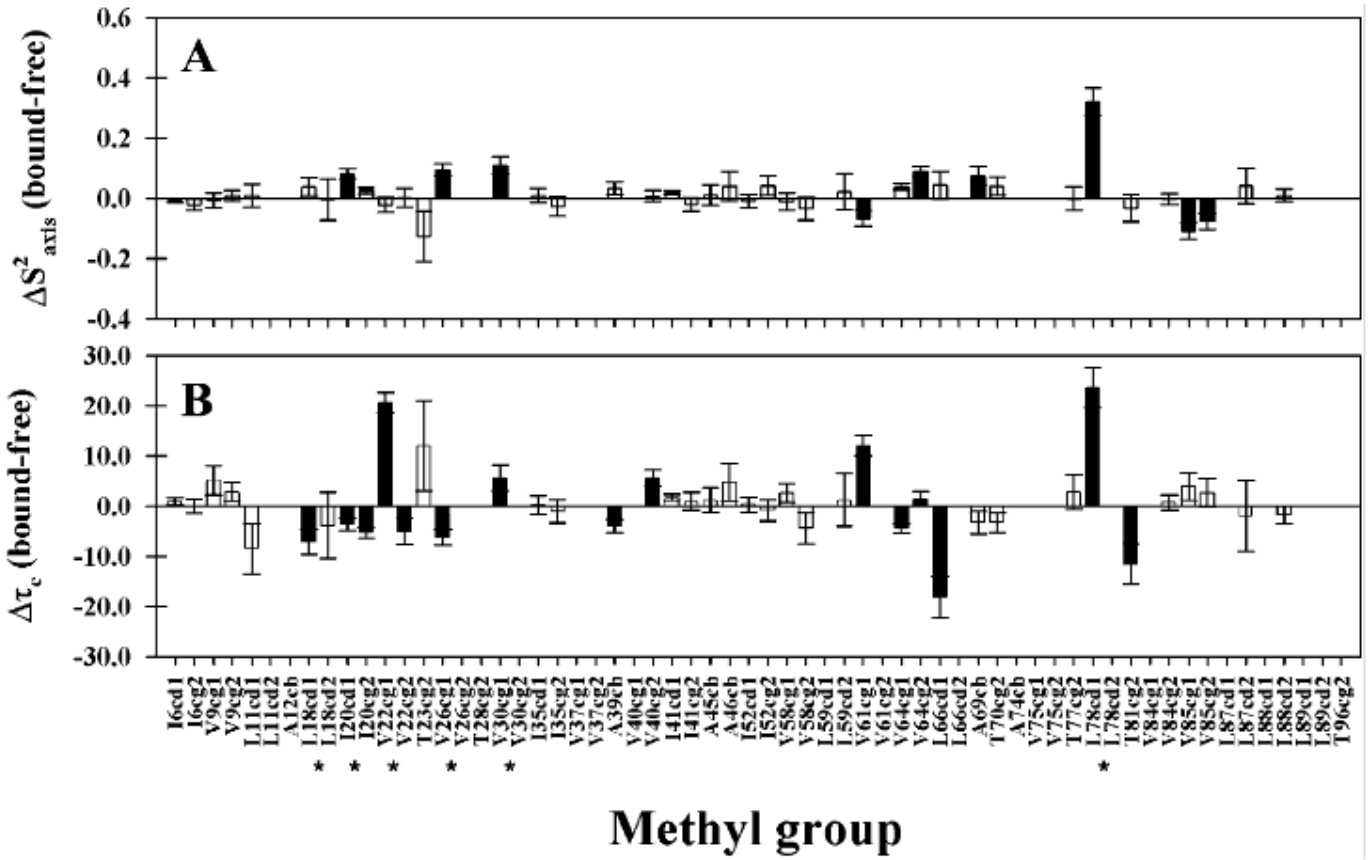
Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain

Ernesto J. Fuentes[1,4], Channing J. Der[2,4] and Andrew L. Lee[1,3,4]*



doi:10.1016/j.jmb.2003.11.010 J. Mol. Biol. (2004) 335, 1105–1115

JMB

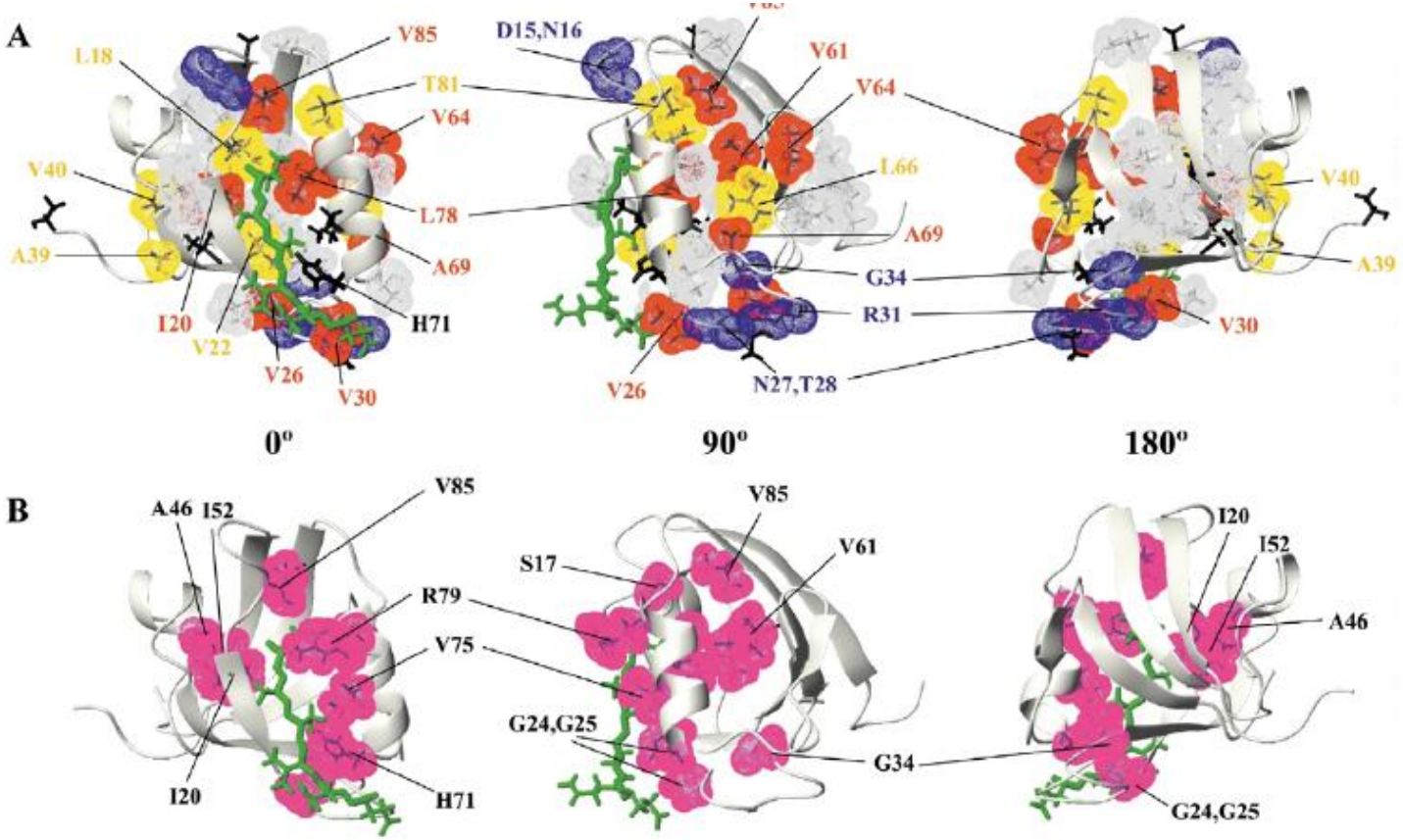
Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain

Ernesto J. Fuentes[1,4], Channing J. Der[2,4] and Andrew L. Lee[1,3,4]*



Correlated evolutionary dynamics correspond to internal protein motions

Further studies

Statistical coevolution analysis and molecular dynamics: Identification of amino acid pairs essential for catalysis

R. August Estabrook, Jia Luo, Matthew M. Purdy, Vyas Sharma, Paul Weakliem, Thomas C. Bruice, and Norbert O. Reich†

Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106

Identification of functional motions in the adenylate kinase (ADK) protein family by computational hybrid approaches

Dagoberto Armenta-Medina,* Ernesto Pérez-Rueda, and Lorenzo Segovia

Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología,
Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

"Fluctuograms" Reveal the Intermittent Intra-Protein Communication in Subtilisin Carlsberg and Correlate Mechanical Coupling with Co-Evolution

Jordi Silvestre-Ryan[1], Yuchun Lin[2], Jhih-Wei Chu[2]*

1 Department of Bioengineering, University of California, Berkeley, Berkeley, California, United States of America, **2** Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, California, United States of America

Molecular Evolution of Protein Conformational Changes Revealed by a Network of Evolutionarily Coupled Residues

Jouhyun Jeon,[1] Hyun-Jun Nam,[2] Yoon Sup Choi,[2] Jae-Seong Yang,[2] Jihye Hwang,[1] and Sanguk Kim*[,1,2,3]

[1]Division of Molecular and Life Science, Pohang University of Science and Technology, Pohang, Korea

[2]School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, Korea

[3]Division of IT Convergence Engineering, Pohang University of Science and Technology, Pohang, Korea

Evolutionary dynamics reflect structural dynamics

- Correspondence between intrinsic dynamics and “evolutionary ensembles”
- Protein design with flexible backbones shows similar sequence diversity to natural families
- Intrinsically disordered proteins evolve faster than folded proteins
- Homeric complex assembly pathways follow evolutionary pathways
- Residue co-evolution identifies long-range correlated motions