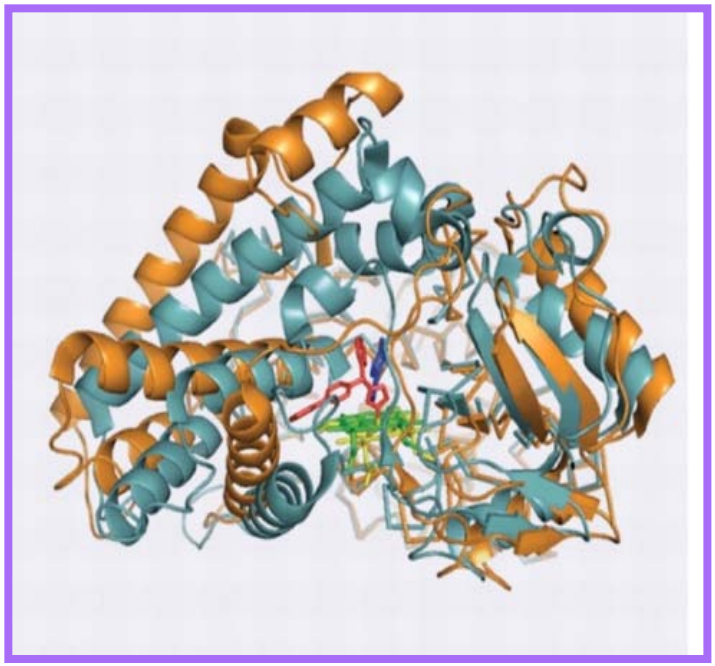
Intrinsically accessible motions enable
Optimal binding of substrate or drugs



Conformational flexibility + sequence variability mediates substrate selectivity

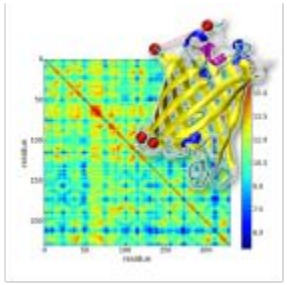
- Two conformations of P450-CYP2B4:

  open (orange) with a large substrate (bifonazole, red), and

  closed (light blue) with the smaller substrate 4-(4-chlorophenyl) imidazole (blue)

See...

N. Tokuriki and D. S. **Tawfik** (2009) *Science* **324**: 203-207



MechStiff

Mechanical Stiffness Calculations

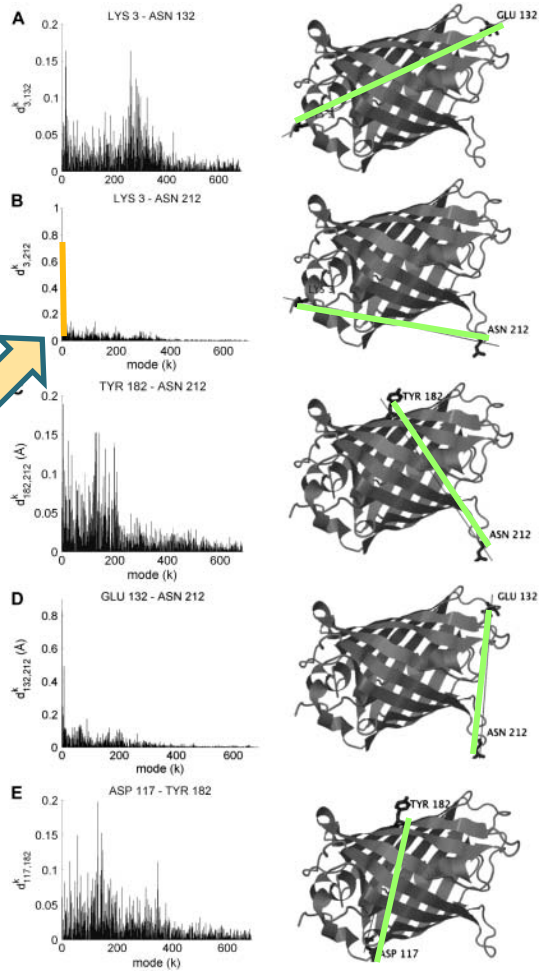
Eyal E, Bahar I Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models*Biophys J* **2008** 94(9):3424-3435.

The mechanical response of proteins

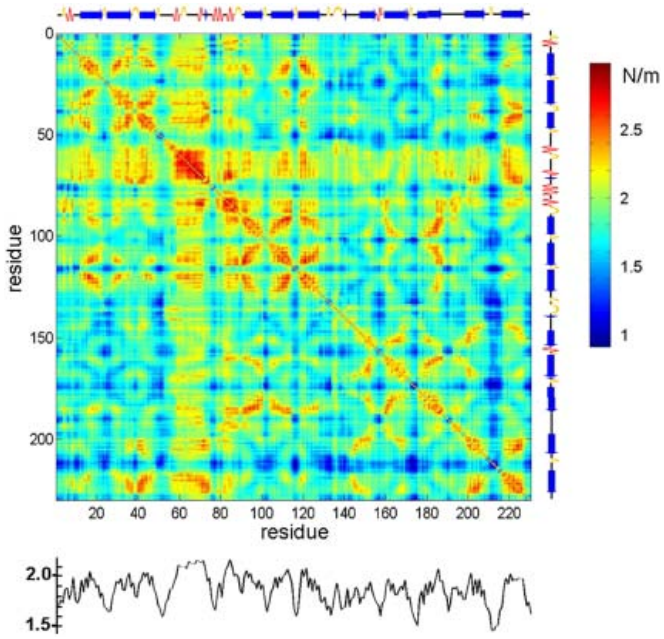
- Probed by AFM experiments
- Usually involves partial or total unfolding
- Depends on
  - the application points of tension
  - or the direction of deformation

-The structural change is accommodated by the collective motions of the protein

- If soft modes can accommodate the change, then the 'effective' resistance to stress, or the effective mechanical stiffness of the protein is smaller.



Constructing a mechanical resistance map for the entire protein



ANM permits us to calculate an effective stiffness/resistance against deformation under uniaxial tension applied to residues i and j.

The idea is simple: i and j already undergo long distance fluctuations by virtue of intrinsically accessible slow modes, the molecule is more 'yielding'.

$< \kappa ij >$ is the 'average force constant'

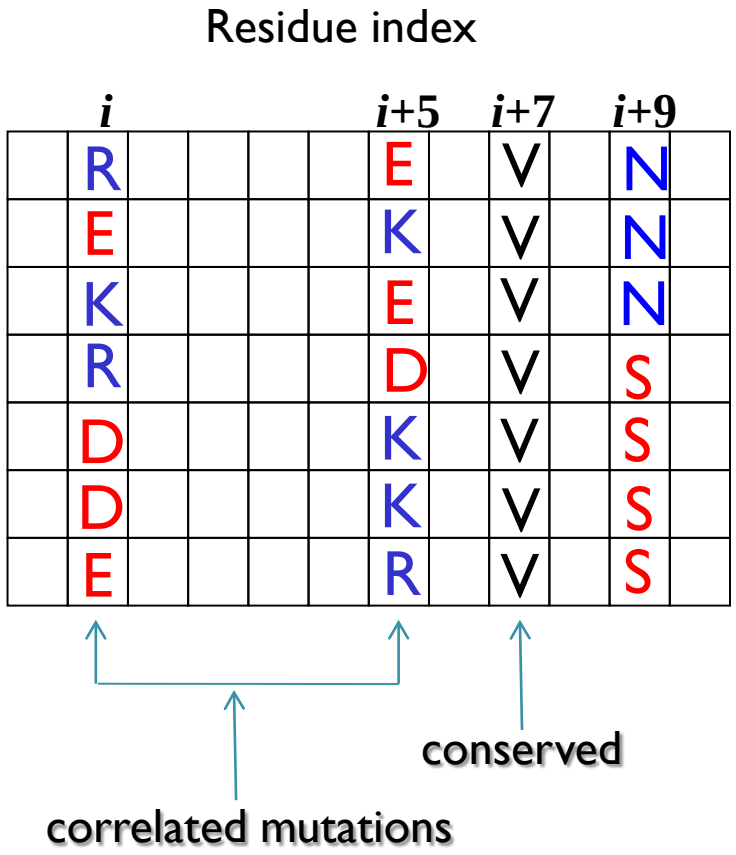
$$< \kappa ij > = \sum_k d_{ij}^{(k)} \lambda_k / \sum_k d_{ij}^{(k)}$$

where $d_{ij}^{(k)}$ is the contribution of mode k to the change $\Delta \mathbf{R}_{ij}$ given by the projection of $\Delta \mathbf{R}_{ij}^{k)}$ onto R_{ij}^{0}, i.e.,

$$d_{ij}^{(k)} = \Delta R_{ij}^{(k)} \; cos(\mathbf{R}_{ij}^{0}, \; \Delta \mathbf{R}_{ij}^{k)})$$

Sequence evolution

an information-theoretic approach



Information entropy (Shannon, 1951)

$$S(i) = \sum_{x_i=1}^{20} P(x_i) \log \frac{1}{P(x_i)}$$

Mutual information (MI)

$$\mathrm{I}(i,j) = \sum_{x_i=1}^{20} \sum_{y_j=1}^{20} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i) P(y_j)}$$

for correlated mutations analysis (CMA)

Mutual Information
without the influence of phylogeny

MIp - to eliminate random noise and phylogenetic components

$$MI_P(i, j) = I(i, j) - APC$$

APC = Average product correction

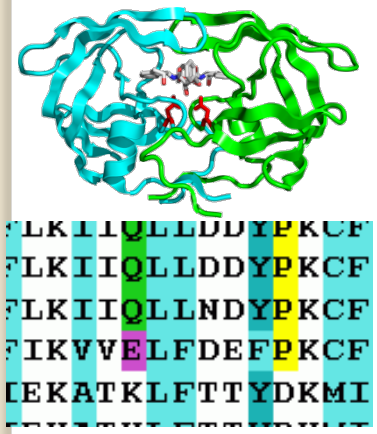
$$= [\, I(i, x)\, I(j, x) \,] \;/\; \langle I(i, j) \rangle$$

	R				E		V		N	
	E				K		V		N	
	K				E		V		N	
	R				D		V		S	
	D				K		V		S	
	D				K		V		S	
	E				R		V		S	

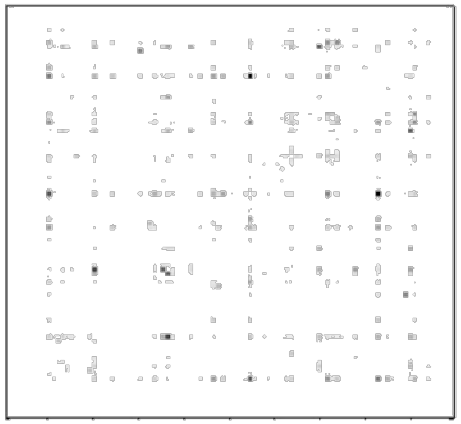
where $I(i, x)$ is the mean mutual information of column $i = \sum_j I(i, j)$

HIV-1 protease correlated mutation analysis (CMA)

MSA of HIV-1 protease



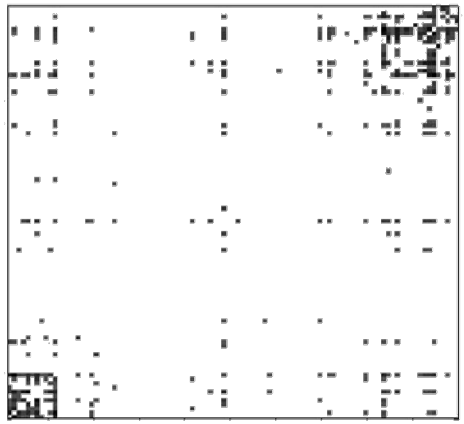
MI matrix $\mathbf{I}_{ij} = \mathrm{I}\ (i, j)$



residue index

Shi and Malik (2000)

spectral clustering



reordered residue index

Dr. Ying Liu

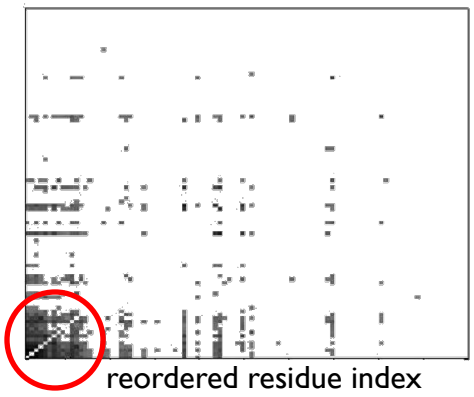
Liu, Eyal & Bahar (2008) *Bioinformatics*

MDR mutations distinguished by CMA

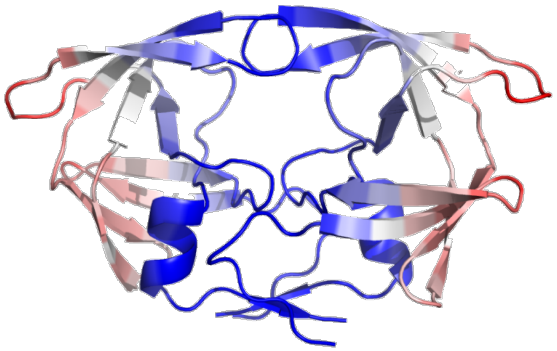
MSA of HIV-1 protease
Stanford HIV Drug Resistance Database
http://hivdb.stanford.edu/

```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR
CDKFGIVVHELGHVVGFWHEHTRPDREDHVV
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL
NFTVGSLIHEIGHAFGLIHEHQRPDRDDYVI
CLTYGTPIHELMHALGFFHEQNRHERDSYVR
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV
CDKFGVVVHELGHVVGFWHEHTRPDRNEFVG
CAYFGTIVHEIGHAIGFHHEQSRPDRDDYIN
CVYHGIIQHELSHALGFYHEHTRSDRNKYVR
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

untreated

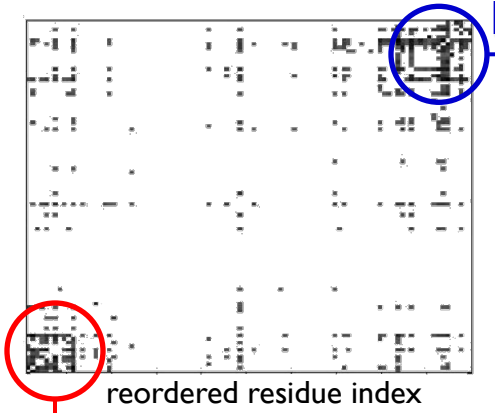


mobility profile

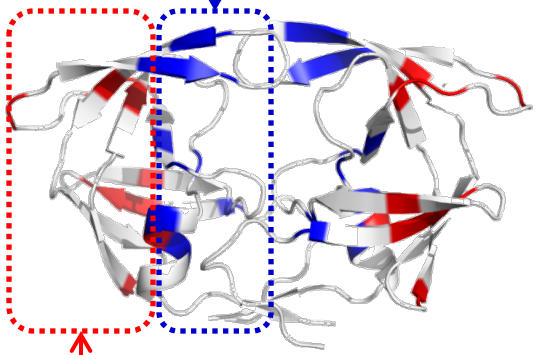



```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR
CDKFGIVVHELGHVVGFWHEHTRPDREDHVV
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL
NFTVGSLIHEIGHAFGLIHEHQRPDRDDYVI
CLTYGTPIHELMHALGFFHEQNRHERDSYVR
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV
CDKFGVVVHELGHVVGFWHEHTRPDRNEFVG
CAYFGTIVHEIGHAIGFHHEQSRPDRDDYIN
CVYHGIIQHELSHALGFYHEHTRSDRNKYVR
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

treated by at least one drug



Drug-resistant cluster



Phylogenetic cluster

Summary

- two groups of correlated mutation sites

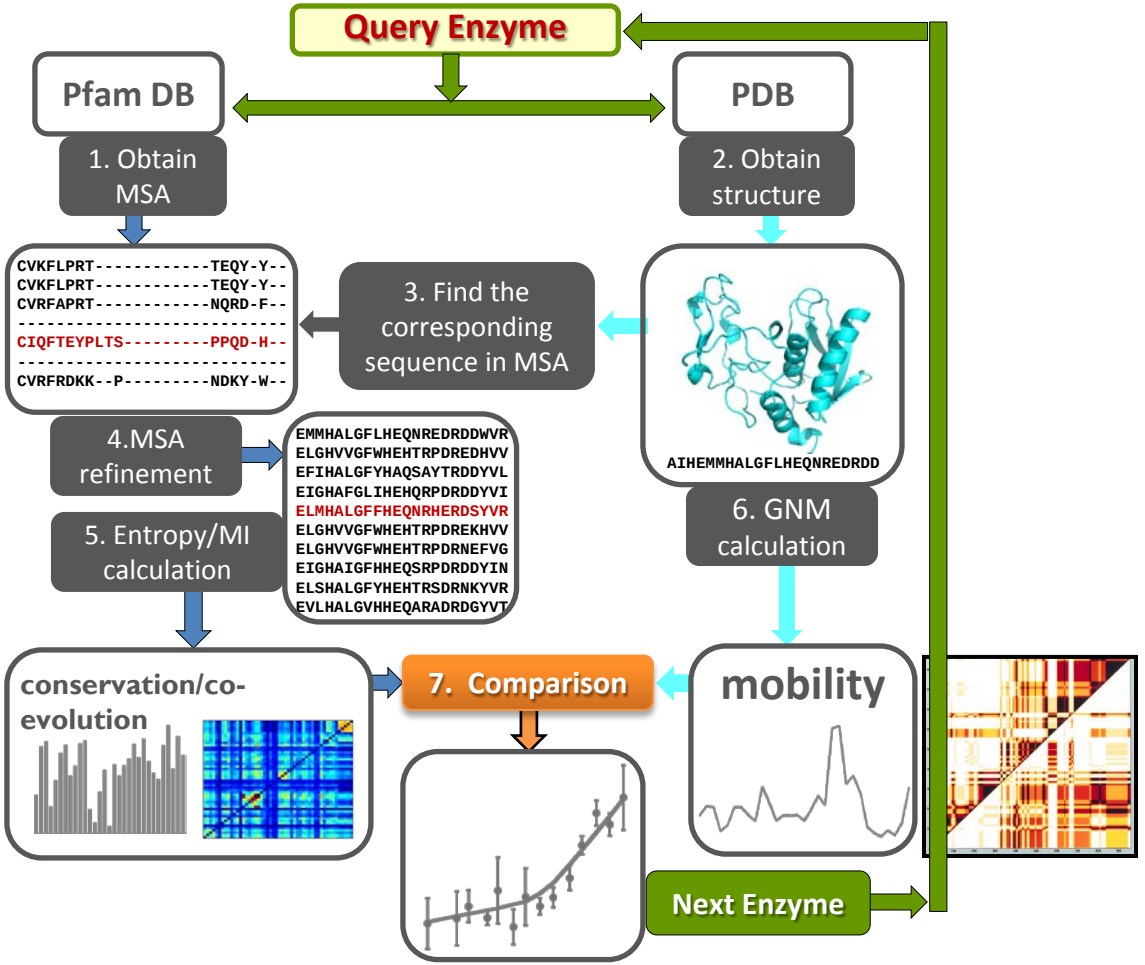
functional aspects	Structural location	structural dynamics
phylogenetic	exposed	mobile
multi-drug resistant	dimerization interface	restrained

Questions:

- Are key mechanical sites (e.g. hinges) conserved?
- Is there any correlation between sequence variability and structural dynamics?
- How does the structure ensure substrate specificity *and* conformational adaptability?

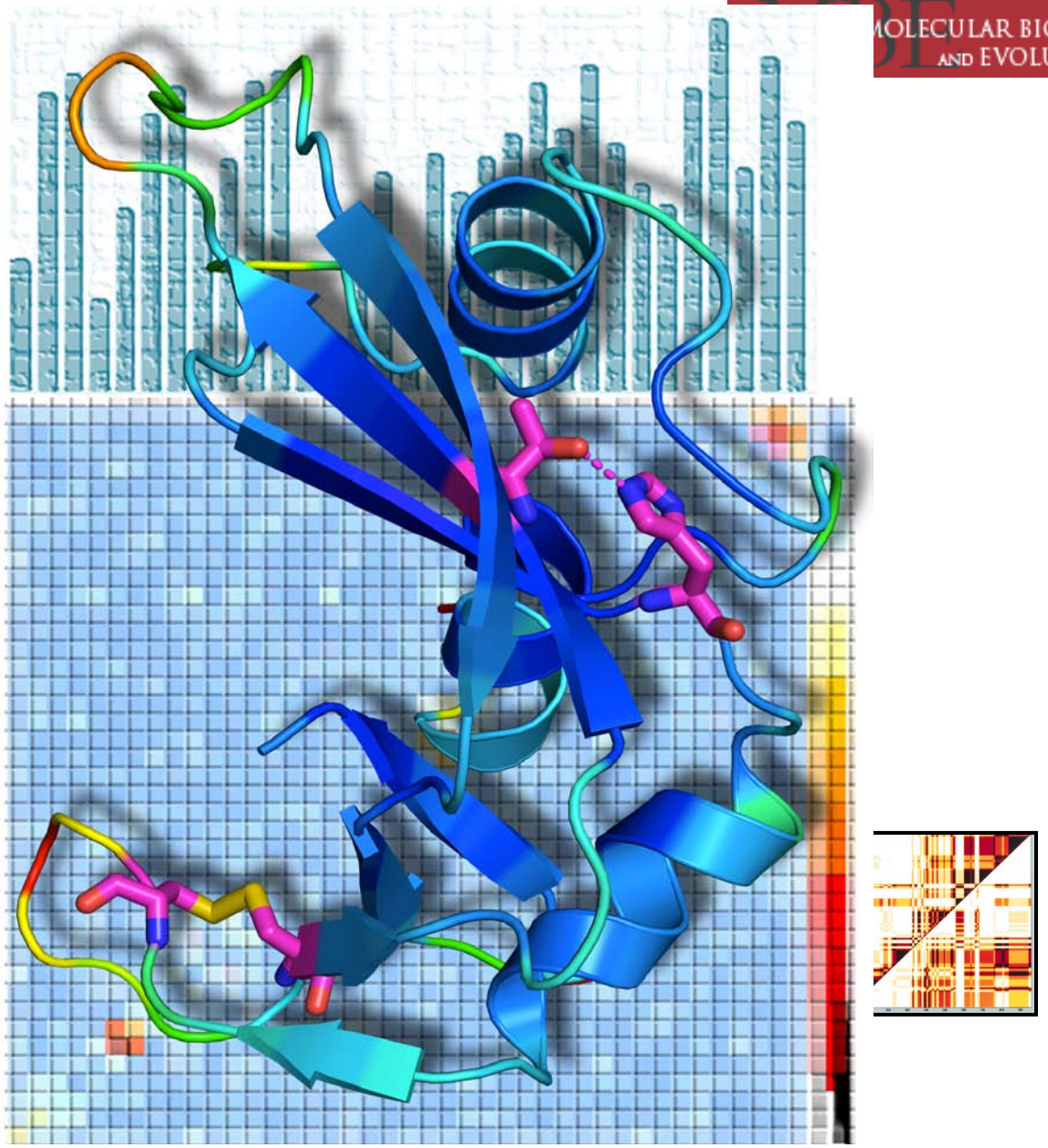
MBE MOLECULAR BIOLOGY AND EVOLUTION

A systematic study of a set of enzymes



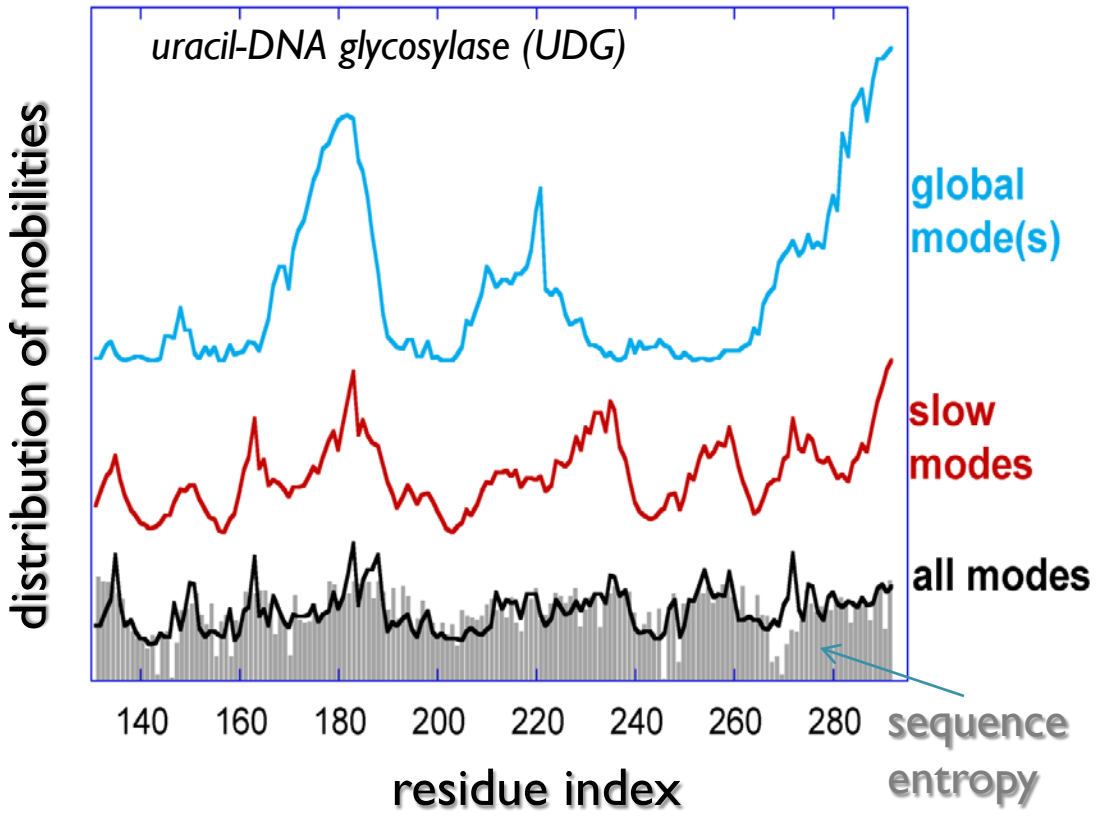
Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

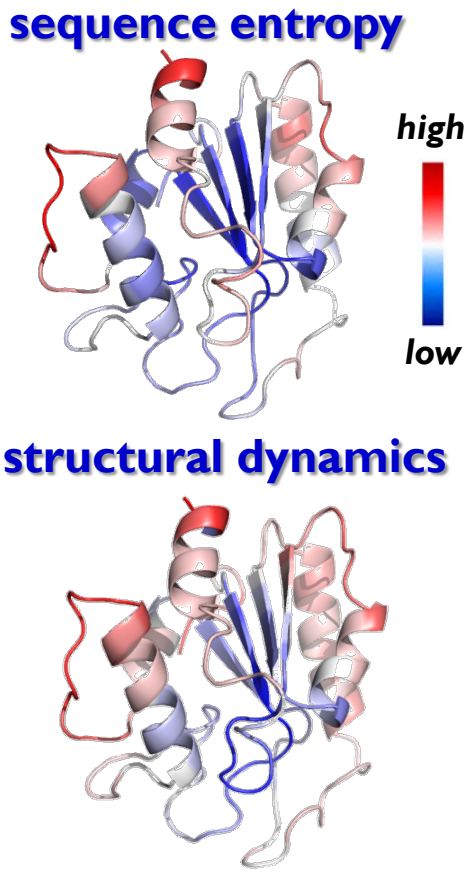
Evol



http://www.csb.pitt.edu/prody/tutorials/evol_tutorial/index.html

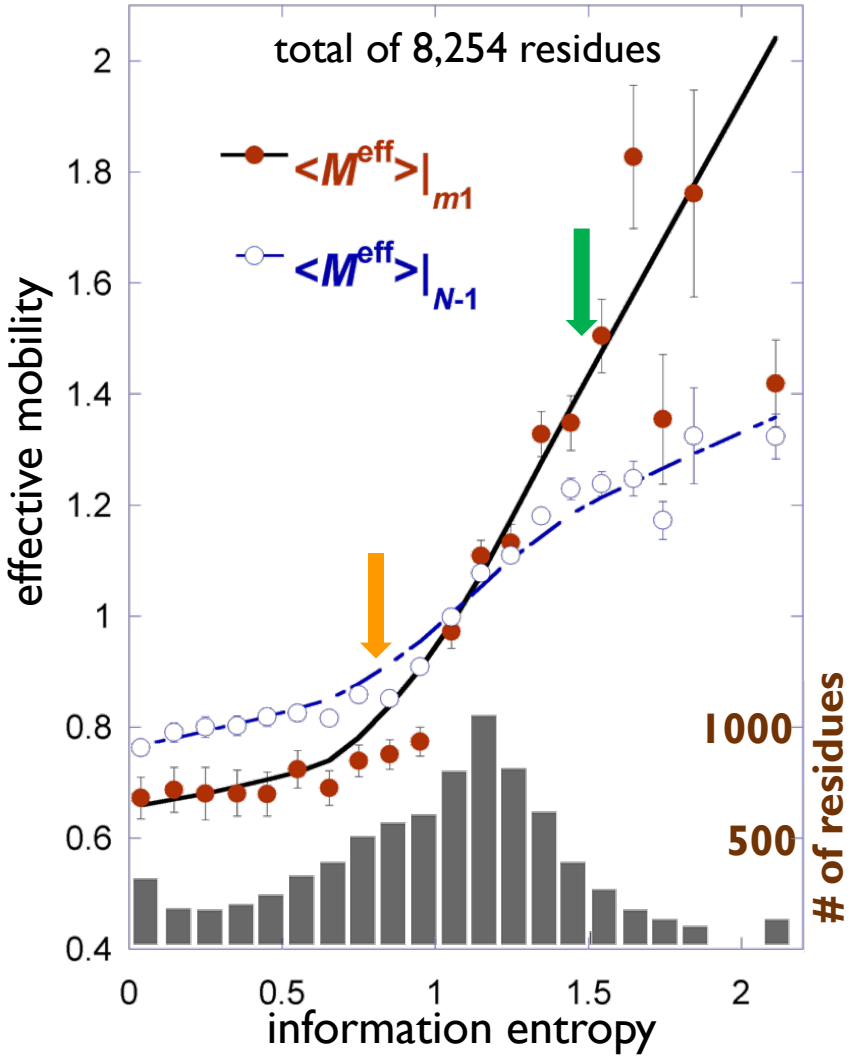
Correlation between sequence entropy & conformational mobility





Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

Mobility increases with sequence entropy

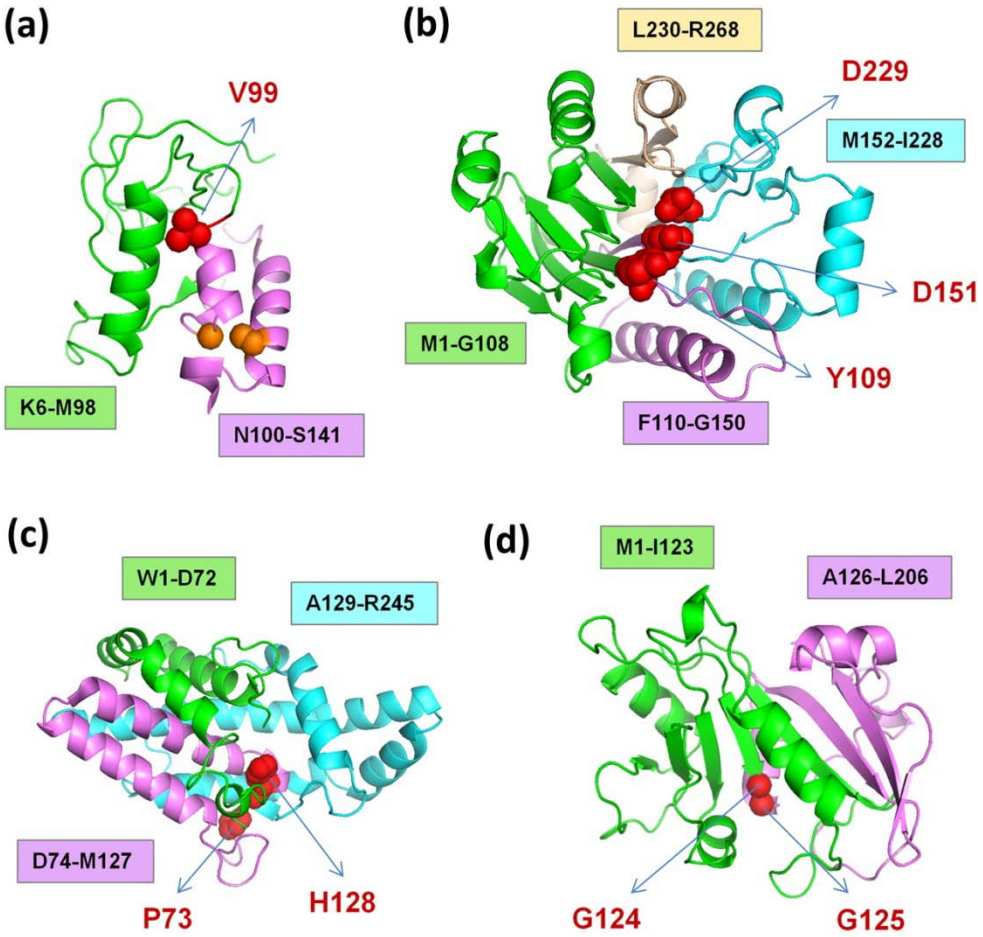


Liu & Bahar *Mol Biol Evol* (2012)

MBE MOLECULAR BIOLOGY AND EVOLUTION

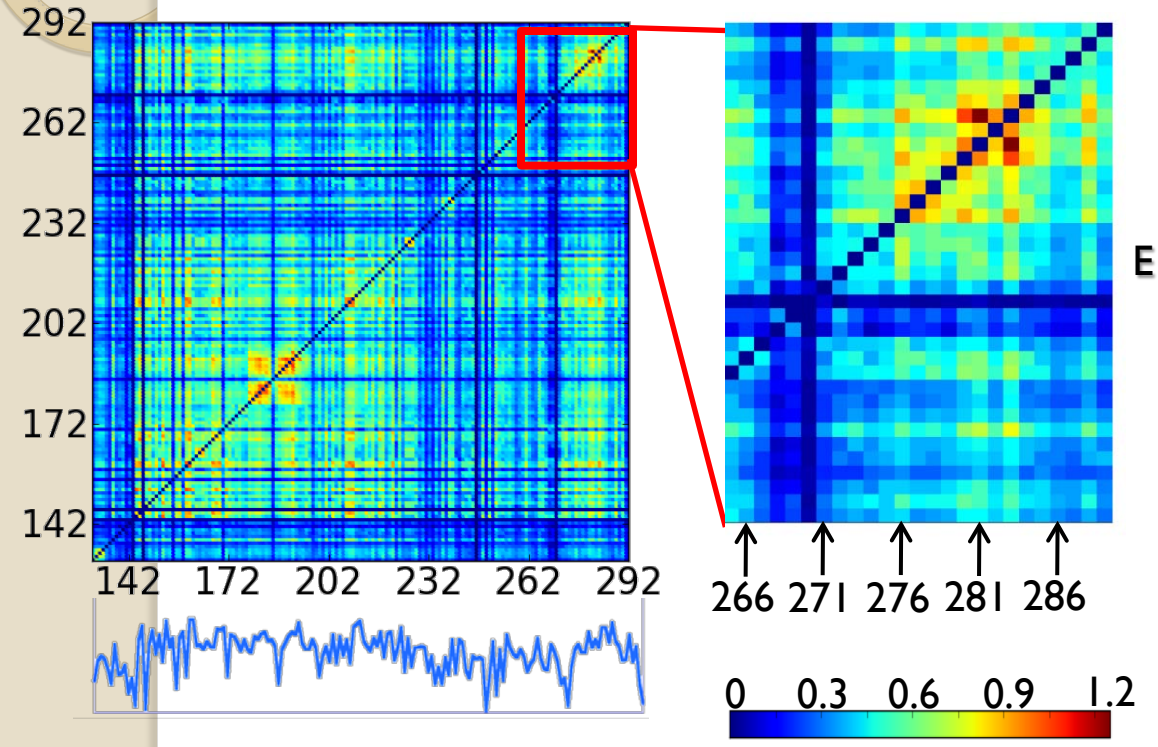
Hinge sites are evolutionarily conserved

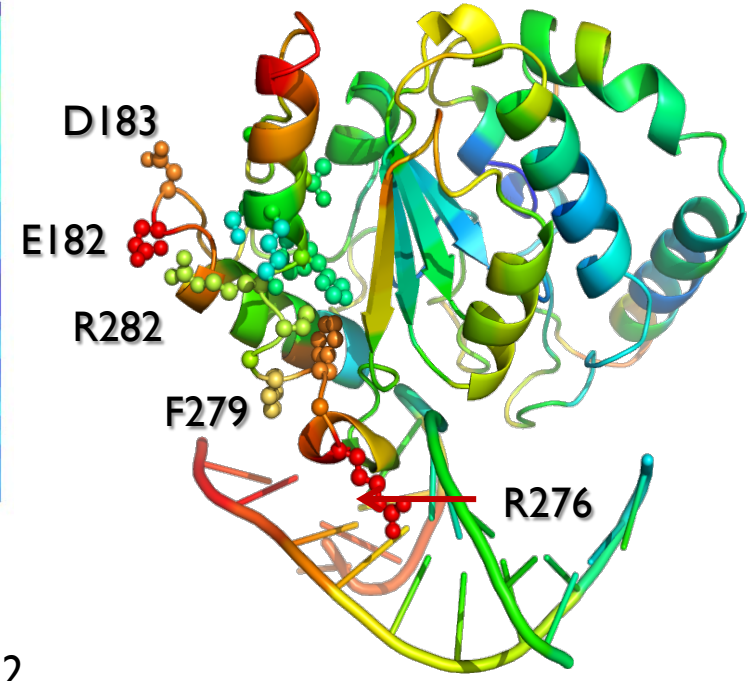
despite their moderate-to-high exposure to environment



3

Amino acids involved in intermolecular recognition are distinguished by **their co-evolution propensities**



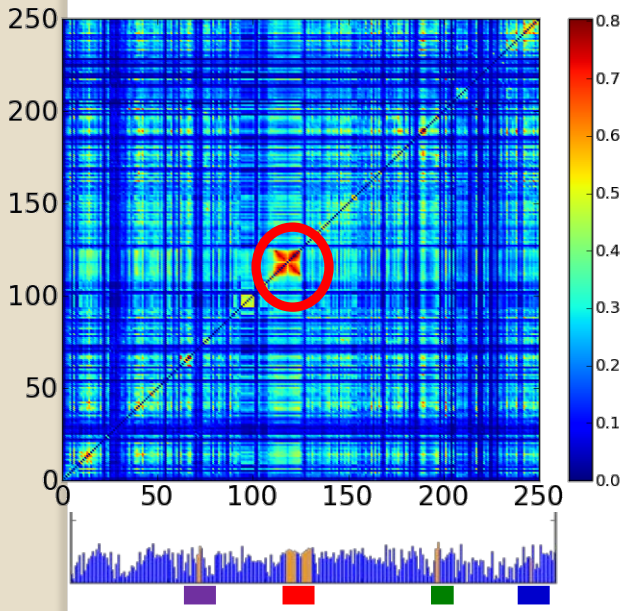


Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

3

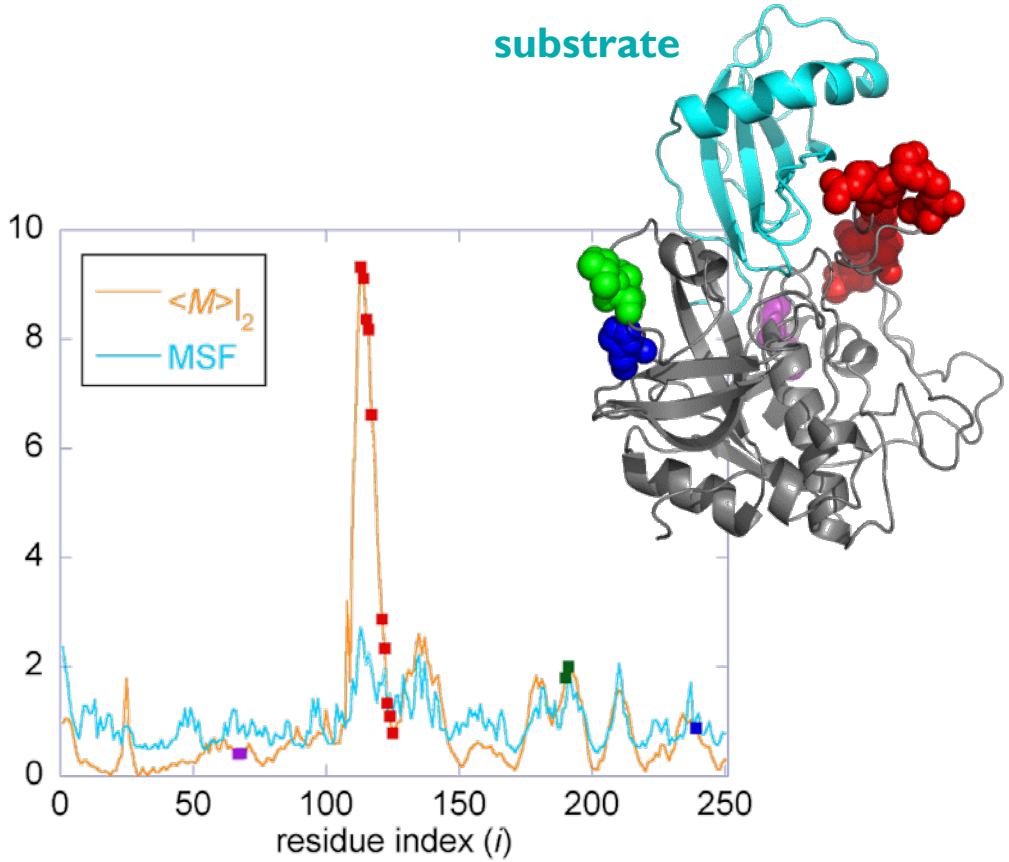
Amino acids involved in intermolecular recognition are distinguished by **their high global mobility**

cathepsin B



residues involved in top 0.05% of $I(i, j)$ values

substrate



Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

Summary

Four types of functional sites

Functional site	Mobility in global modes	Sequence evolution	Dominant Feature
Chemical (catalytic, ligand binding)	Minimal	Conserved	high fidelity, precision
Core	Minimal	Conserved	high stability
Hinge sites	Minimal	Conserved	rotational flexibility
Substrate recog-nition (specific)	High	High co-evolution propensity	adaptability

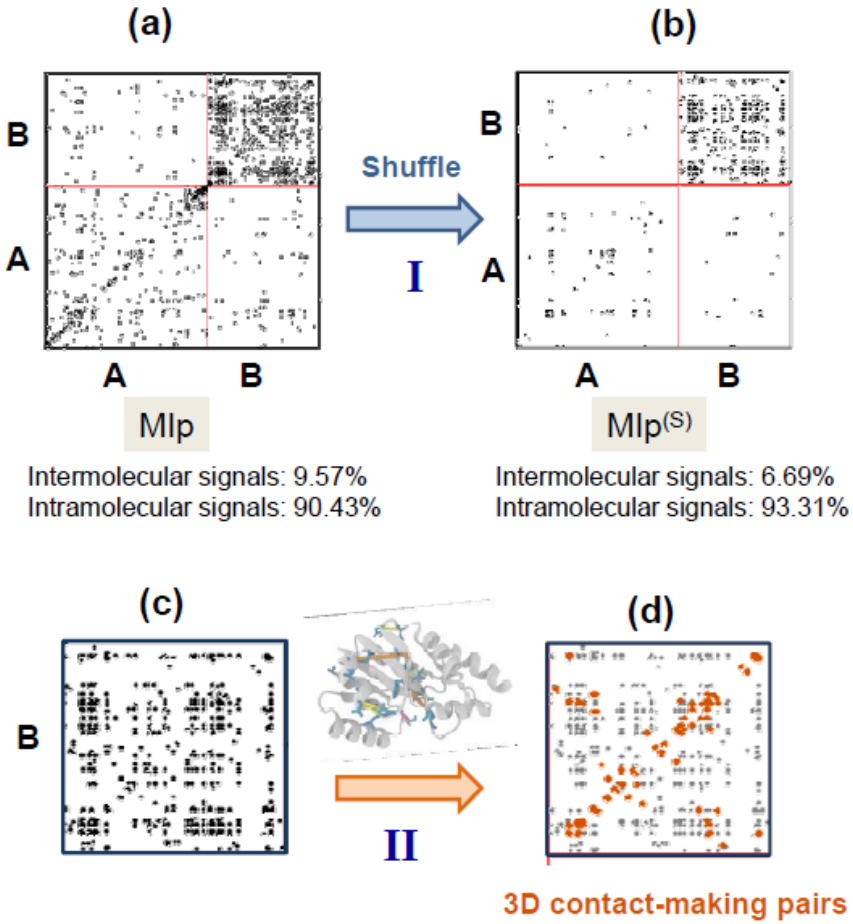
There are several methods for evaluating sequence co-evolution

Mao W, Kaya C, Dutta A, Horovitz A, Bahar I (2015) Comparative Study of the Effectiveness and Limitations of Current Methods for Detecting Sequence Coevolution *Bioinformatics* **pii**: btv103PMID: 25697822

Four possible outcomes:

- True positive (TP) – correctly predicted to be a hit
- False positive (FP); predicted but it is a miss
- True negative (TN) – correctly predicted to be a miss
- False negative (FN) – predicted as a miss, but is a hit

Two criteria for assessing the performance of different methods



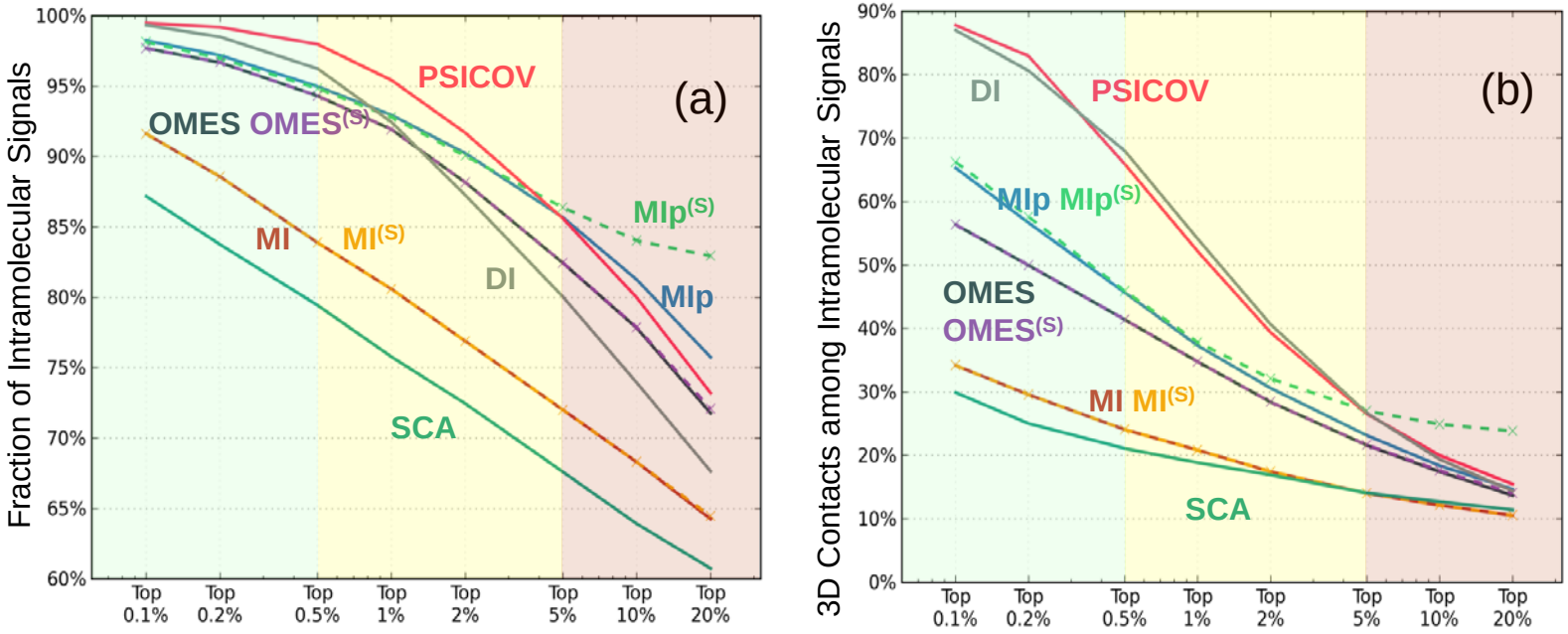
- Minimizing false positives (signals between non interacting proteins)
- Maximizing true positives (signals between contact making residues

Screening of large databases

For testing 9 methods, including

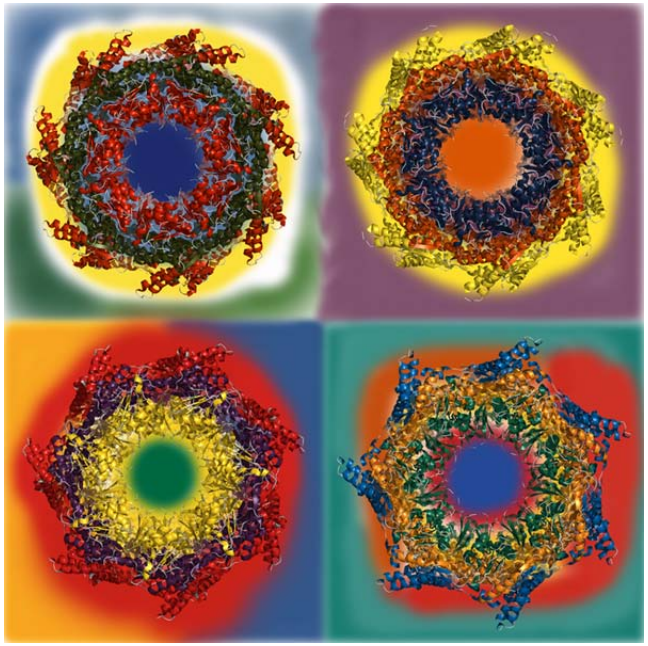
- observed-minus-expected-squared ((OMES) (Kass and Horovitz, 2002)
- statistical coupling analysis (SCA) (Halabi et al., 2009; Lockless and Ranganathan, 1999).
- Direct Coupling Analysis (DCA or DI for Direct Information) (Morcos et al., 2011; Weigt et al., 2009),
- Protein Sparse Inverse COVariance (PSICOV) (Jones et al., 2012),

PSICOV and DI are the best



Average performance of the nine methods based on two criteria, absence of intermolecular FPs (a), and fraction of 3D contact making pairs (b) among different subsets of top-ranking signals. The signals are classified to 3 groups: strong coevolution signals (0.1-0.5%), intermediate signals (0.5-5%) and relatively weak signals (5-20%), which also refer to relatively small, intermediate, and high coverage of coevolving pairs. PSICOV and DI outperform other methods in the strong coevolution region. For the intermediate signal, OMES and MIp exhibit performances similar to PSICOV and DI in panel **a**. MIp$^{(S)}$ shows the best performance in the weak signal regime. SCA and MI (and its shuffled version) have lower performance compared to all others for both criteria over the whole range.

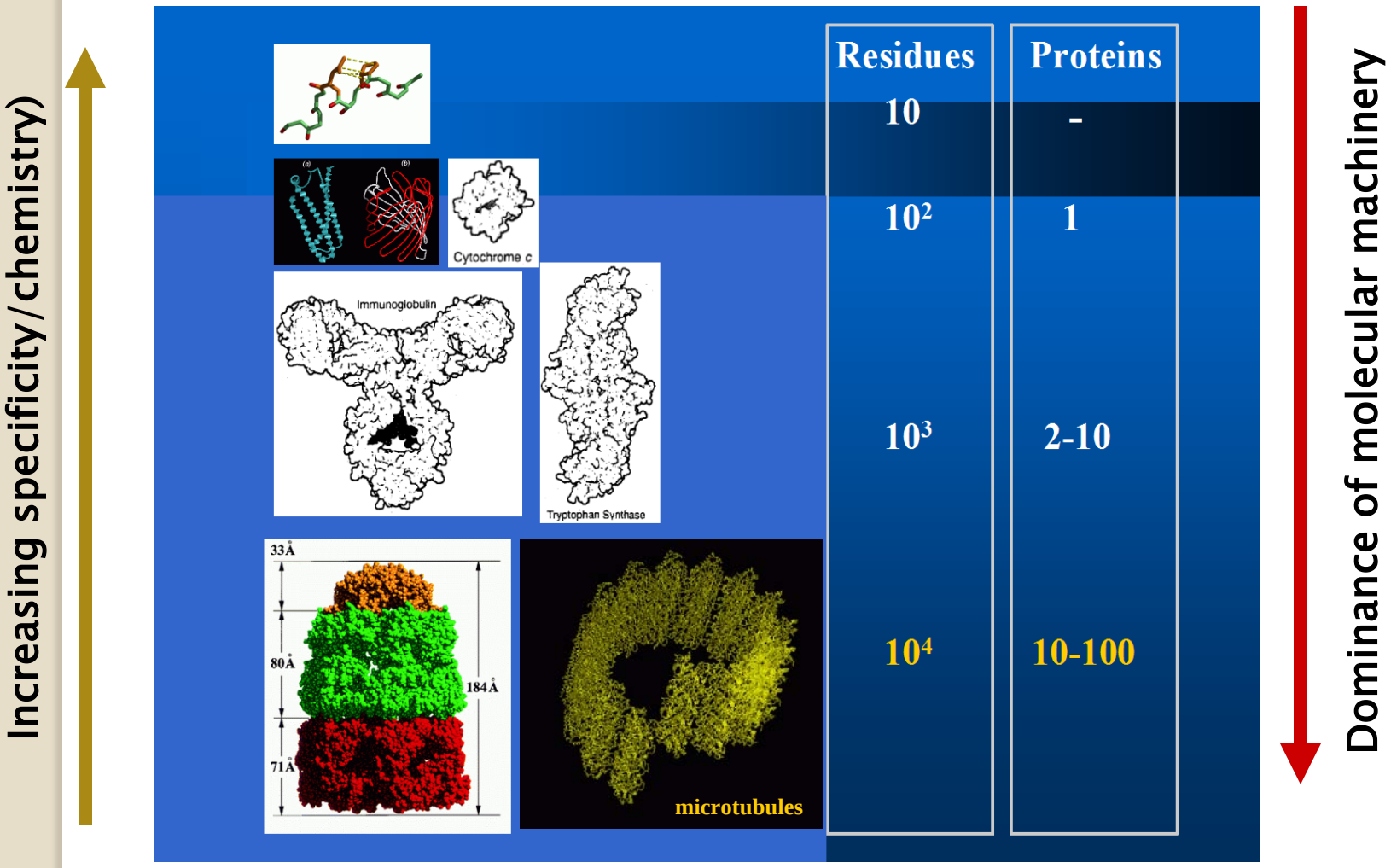
CONCLUSION



- Proteins are designed to favor functional changes in their structure. Pre-existing soft modes facilitate substrate binding.
- Collective mechanics/allosteric dynamics are mediated by conserved residues
- The intrinsic motions confer enhanced flexibility at substrate recognition sites
- Correlated mutations at recognition sites enable substrate specificity while conferring conformational adaptability
- Accurate modeling of protein dynamics is essential to assessing target druggability

Mechanics vs chemistry?

How does complexity scale with size of the system?



ProDy

Protein Dynamics Analysis in Python

NMWiz

Markus Dittrich, PhD
NRBSC Group Leader
Pittt Supercomputing Center

Dr. Timothy R Lezon
Assistant Prof, DCSB, Pitt

Drs. Ahmet Bakan and Anindita Dutta

Dr. Chakra Chennubhotla
Assist Prof, DCSB, Pitt

Acknowledgment: NIH - 5 P41 GM10371202