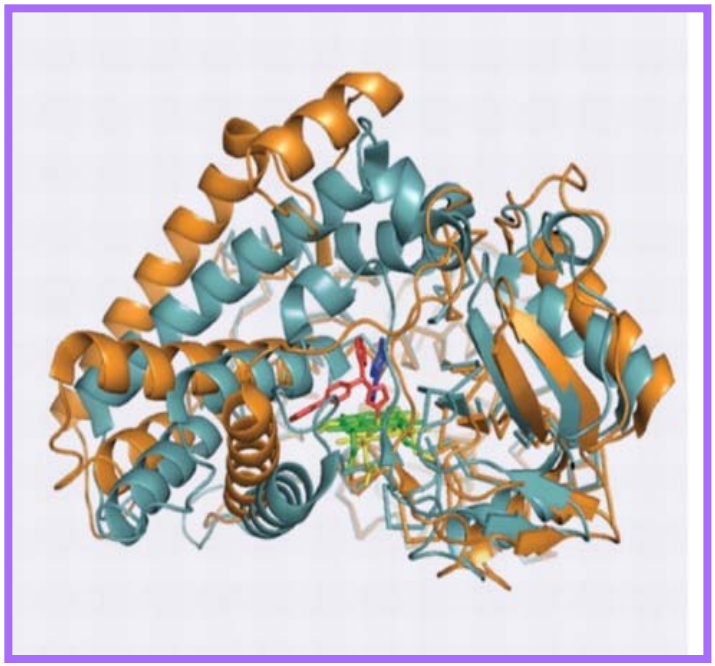
Intrinsically accessible motions enable
Optimal binding of substrate or drugs



Conformational flexibility + sequence variability mediates substrate selectivity

- Two conformations of P450-CYP2B4:

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

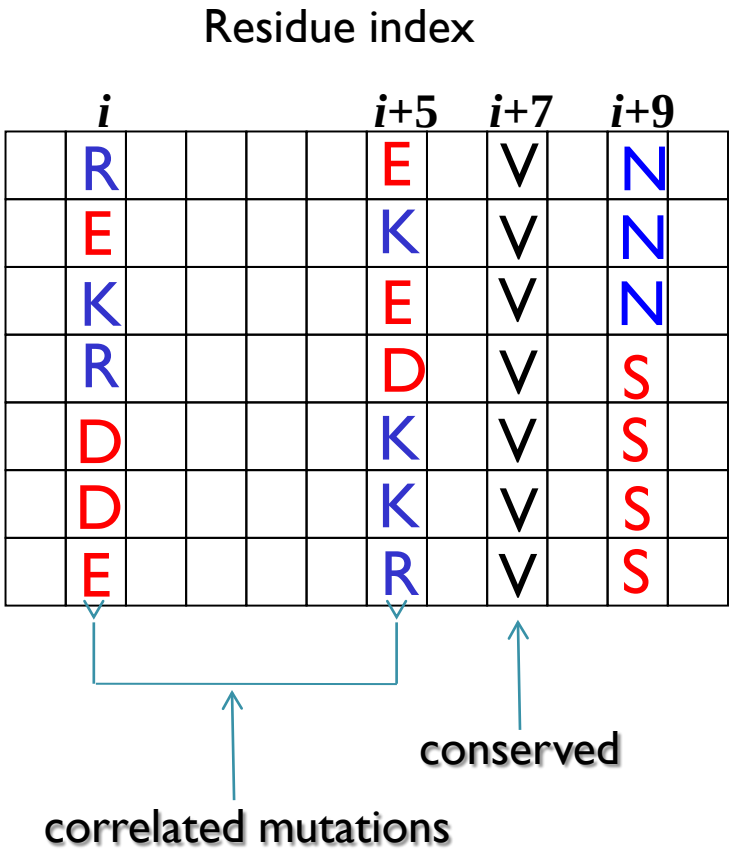
  closed (light blue) with the smaller substrate 4-(4-chlorophenyl) imidazole (blue)

See...

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

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

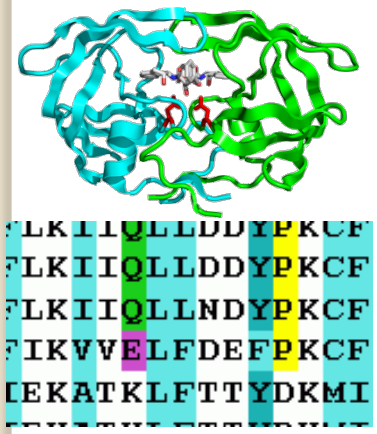
$$= [\, I(i, x)\, I(j, x) \,] \;/\; \langle I(i, j) \rangle$$

	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	

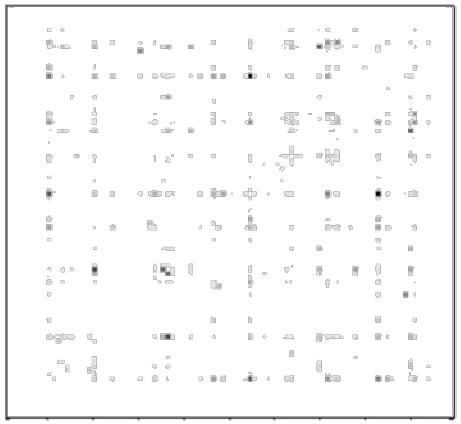
where $I(i, x)$ is the mean mutual information of column $i = \sum_j I(i, j)$

HIV-1 protease correlated mutation analysis (CMA)

MSA of HIV-1 protease



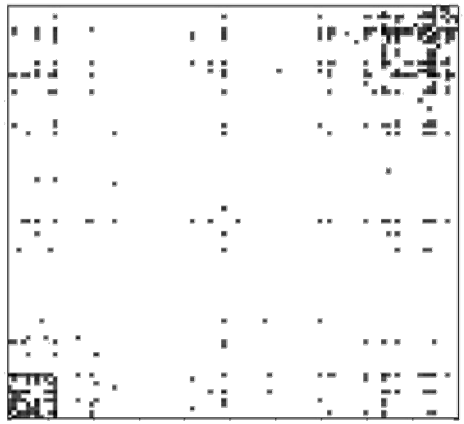
MI matrix $\mathbf{I}_{ij} = \mathrm{I}\,(i, j)$



residue index

Shi and Malik (2000)

spectral clustering



reordered residue index

Dr. Ying Liu

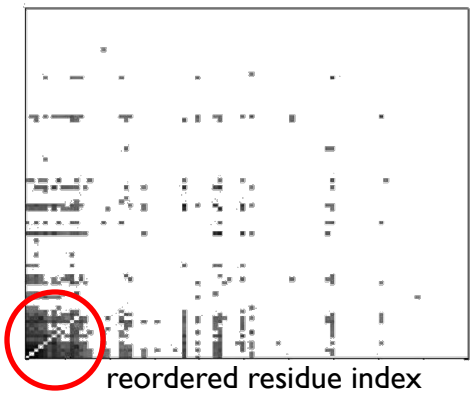
Liu, Eyal & Bahar (2008) *Bioinformatics*

MDR mutations distinguished by CMA

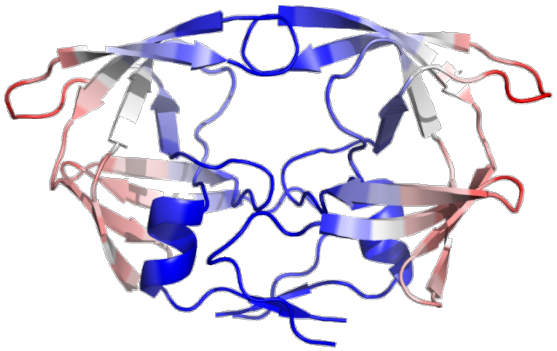
MSA of HIV-1 protease
Stanford HIV Drug Resistance Database
http://hivdb.stanford.edu/

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

untreated

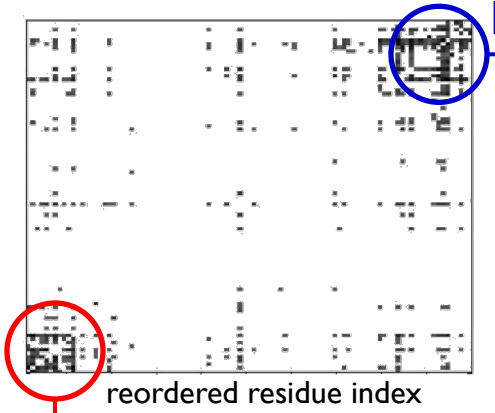


mobility profile

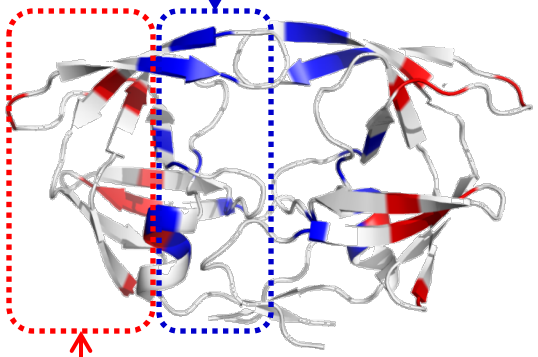



```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR
CDKFGIVVHELGHVVGFWHEHTRPDREDHVV
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL
NFTVGSLIHEIGHAFGLIHEHQRPDRDDYVI
CLTYGTPIHELMHALGFFHEQNRHERDSYVR
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV
CDKFGVVVHELGHVVGFWHEHTRPDRNEFVG
CAYFGTIVHEIGHAIGFHHEQSRPDRDDYIN
CVYHGIIQHELSHALGFYHEHTRSDRNKYVR
CINSGTIIHEVLHALGVHHEQARADRDGYVT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

treated by at least one drug



Drug-resistant cluster



Phylogenetic cluster

4

Summary

- two groups of correlated mutation sites

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

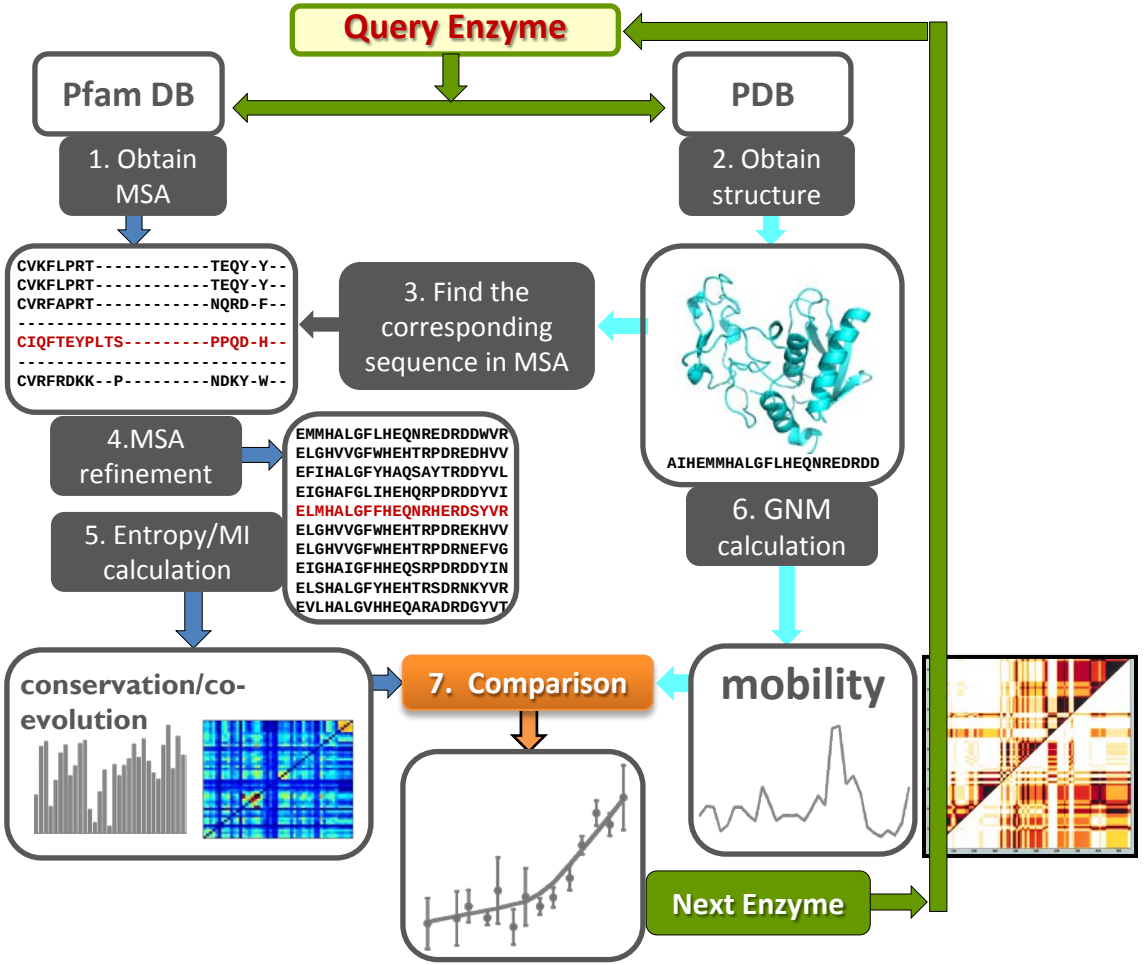
Liu, Eyal & Bahar (2008) *Bioinformatics* **15**, 1243.

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?

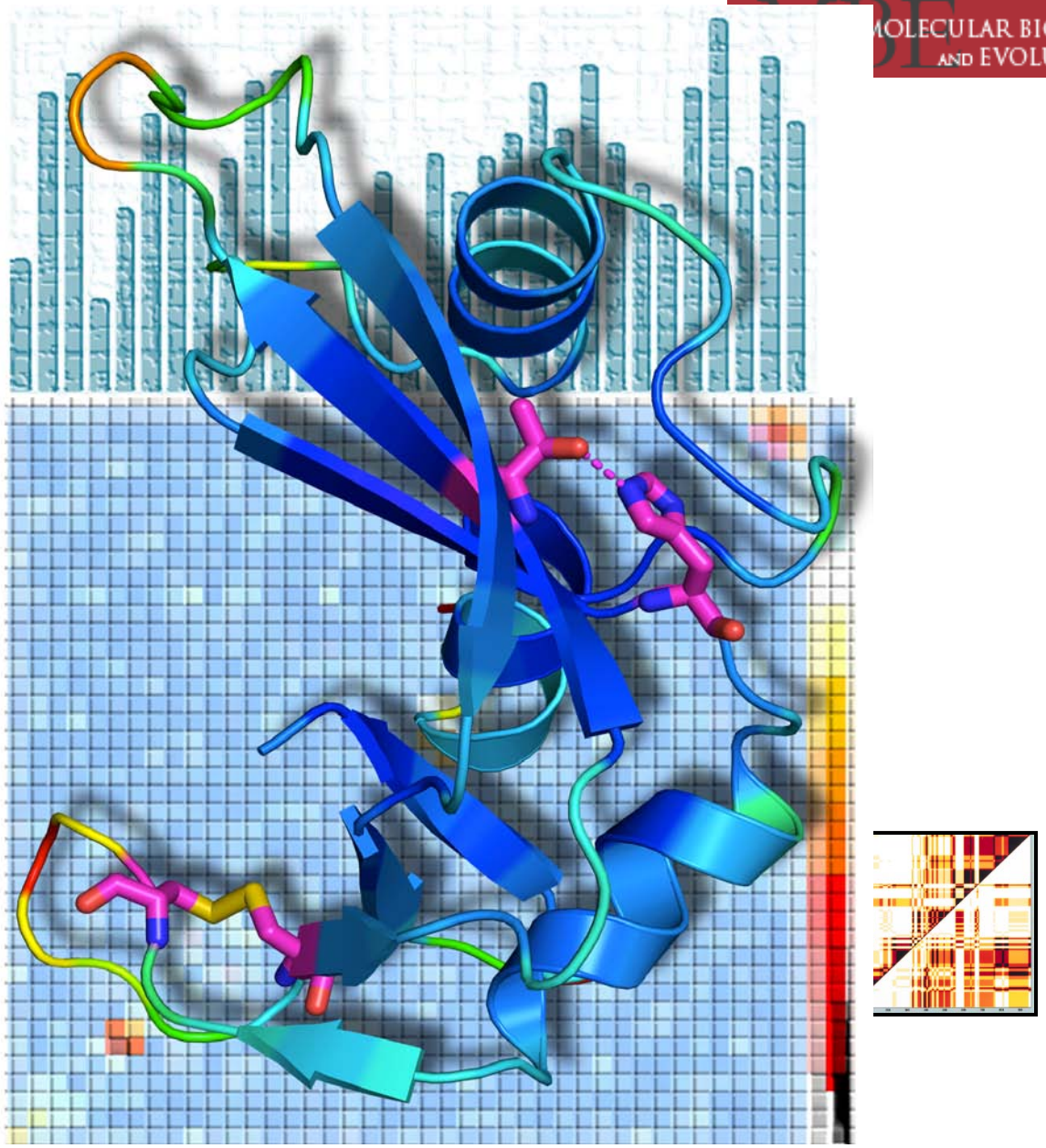
MBE MOLECULAR BIOLOGY AND EVOLUTION

A systematic study of a set of enzymes



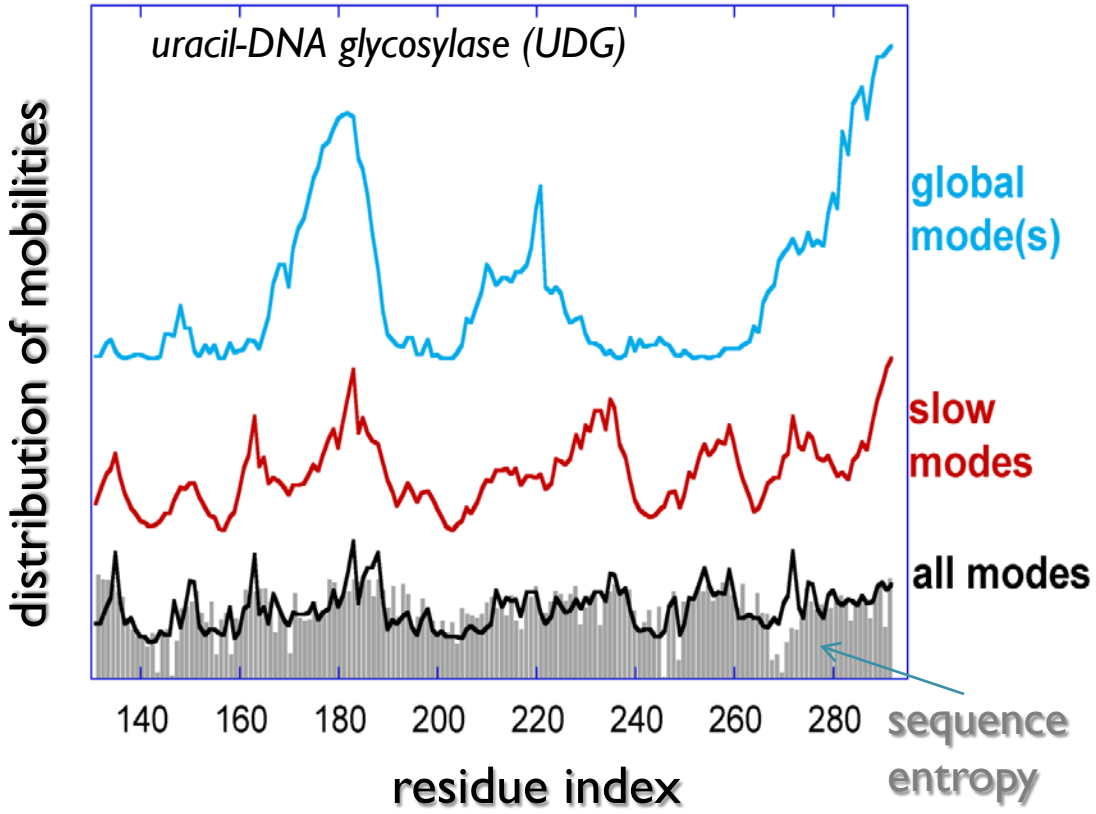
Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

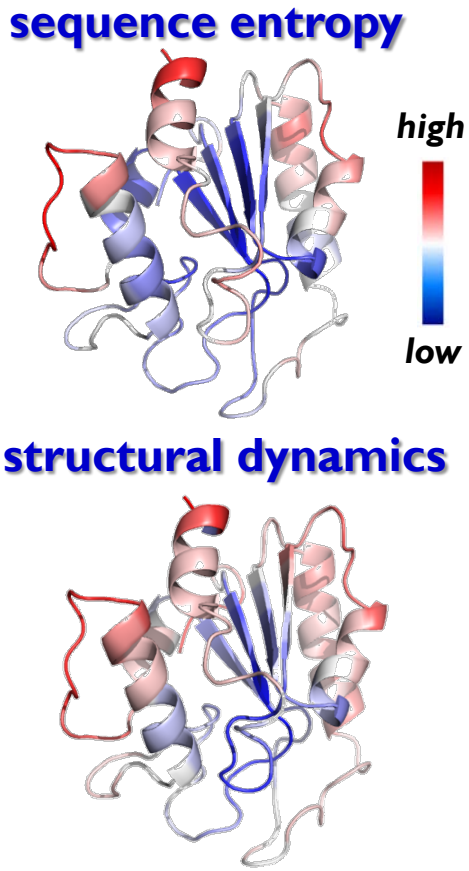
Evol



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

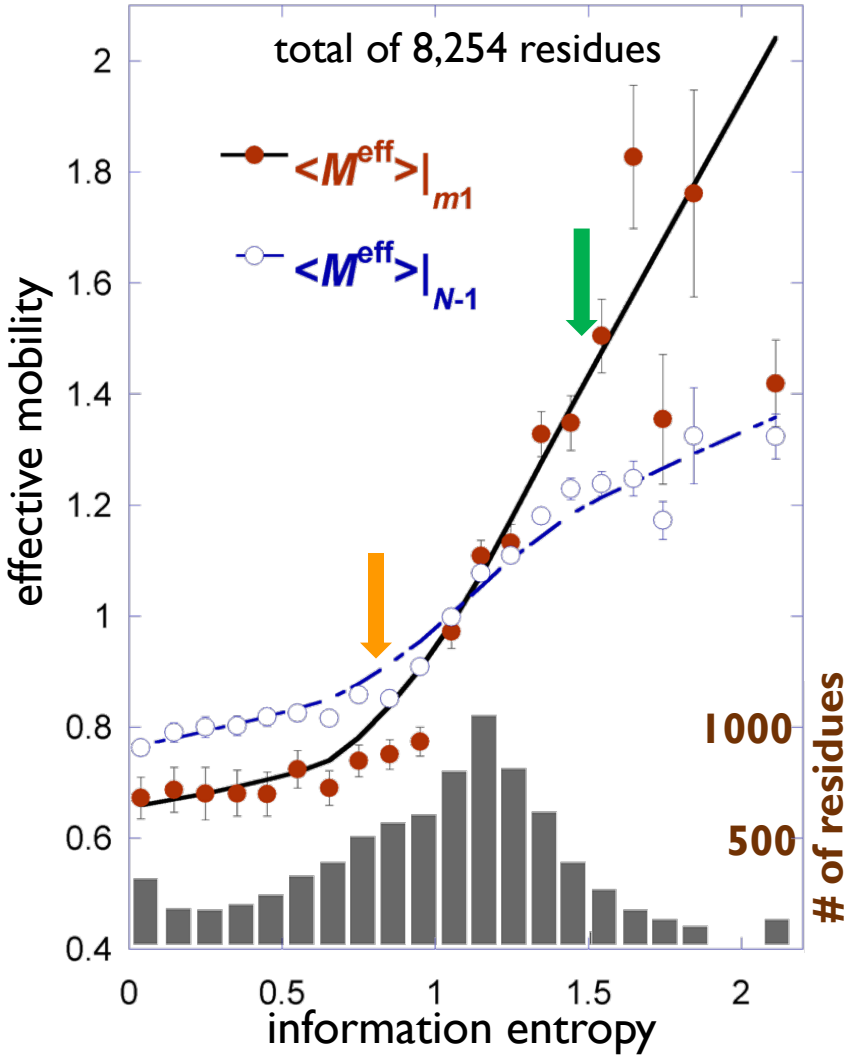
Correlation between sequence entropy & conformational mobility





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

Mobility increases with sequence entropy

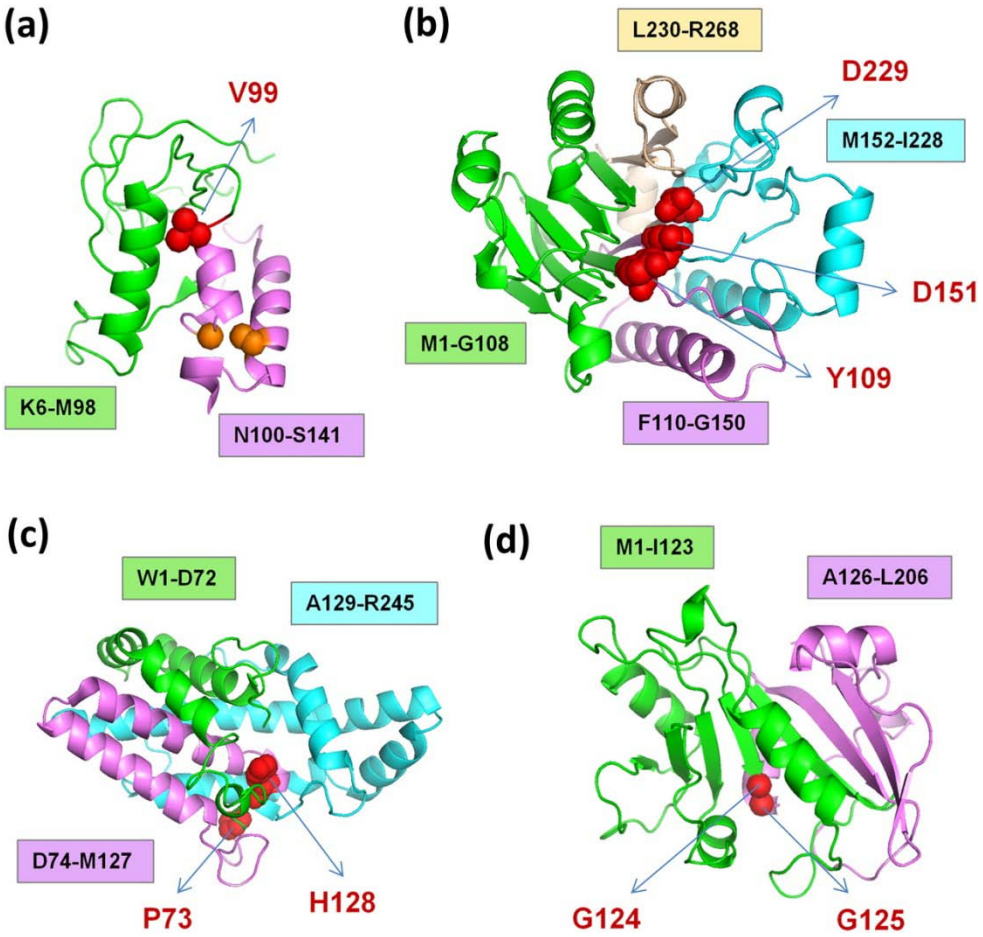


Liu & Bahar *Mol Biol Evol* (2012)

MBE MOLECULAR BIOLOGY AND EVOLUTION

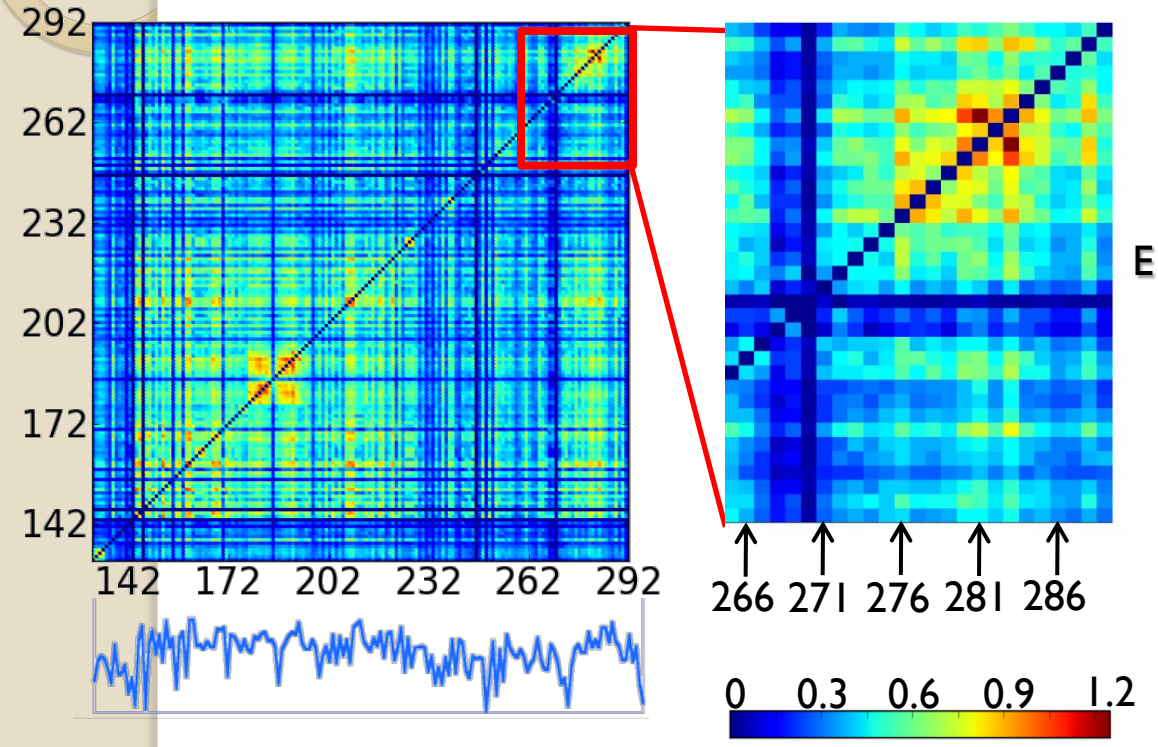
Hinge sites are evolutionarily conserved

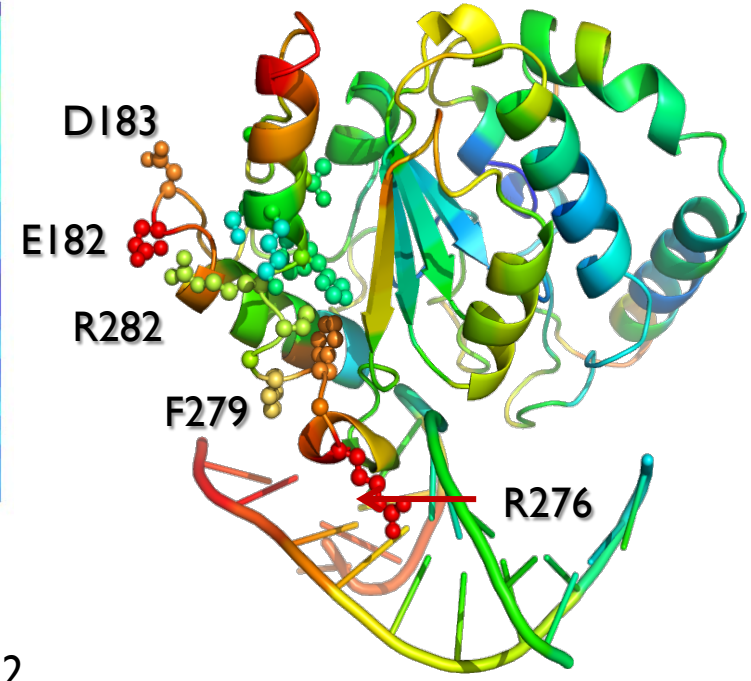
despite their moderate-to-high exposure to environment



3

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



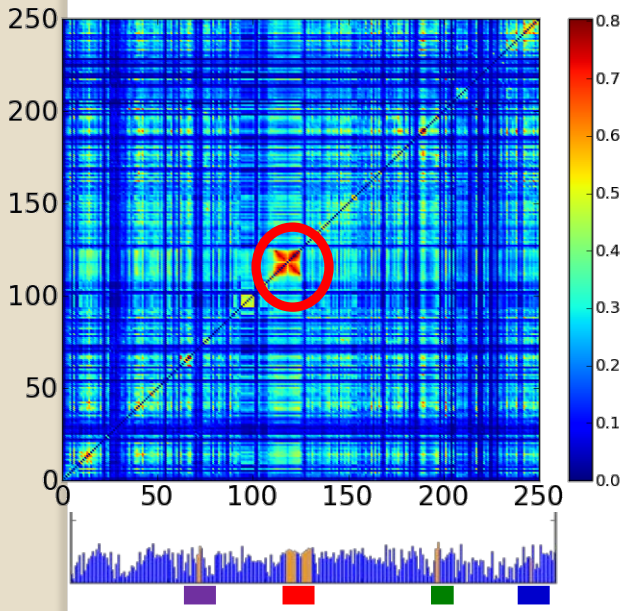


Liu Y, Bahar I (2012) "Sequence Evolution Correlates with Structural Dynamics" *Mol Biol Evol 9, 2253-63*

3

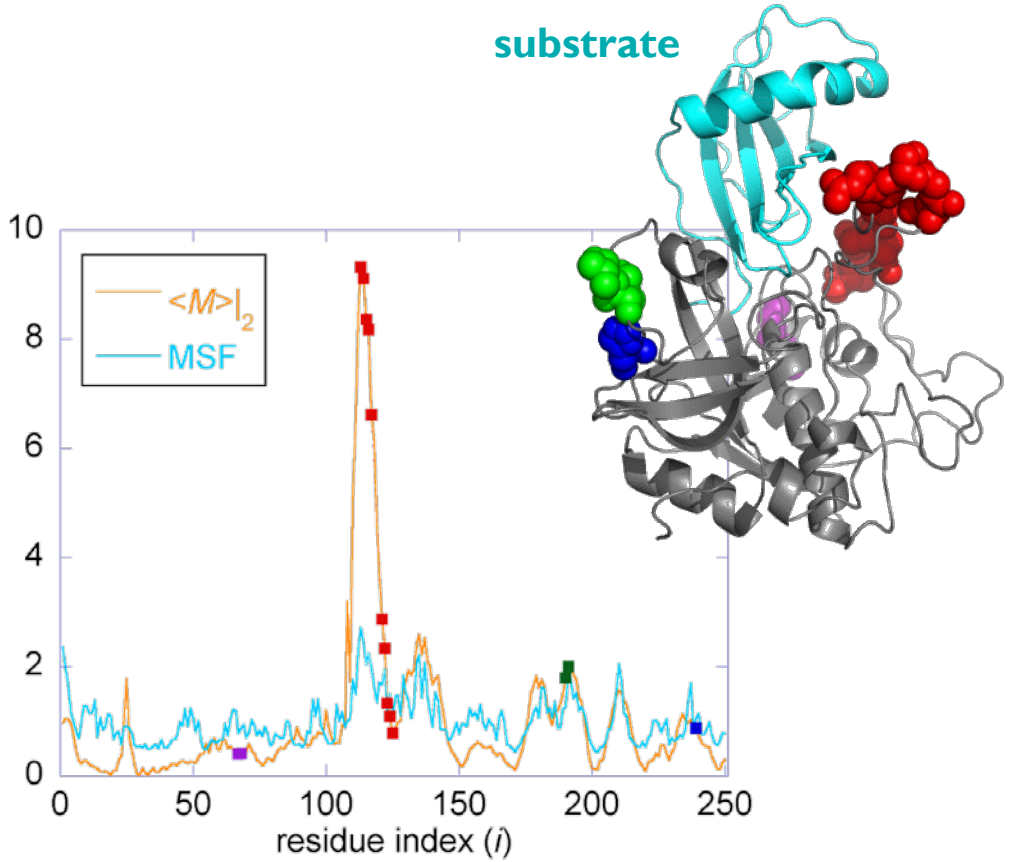
Amino acids involved in intermolecular recognition are distinguished by **their high global mobility**

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

Allosteric communication mechanisms explored by network models

- Diffusion of signal obeys a Markov process
- The structure is modeled as a network
- Network connectivity given by Γ

References

Laplacian based manifold methods (Belkin & Niyogi)

Chennubhotla & Bahar Mol Systems Biology (2006); PLoS Comp Bio (2007)

Markov Model of Network Communication

$\Gamma = D - A$ where A = connectivity/affinity matrix and D = diagonal matrix of degrees

A discrete-time, discrete-state Markov process is defined by setting the conditional probability of signal transduction from residue *j* to *i* as

$$m_{ij} = a_{ij} / d_j$$

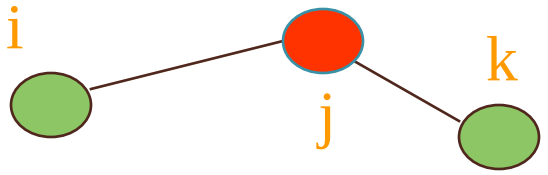
The conditional probability matrix $\mathbf{M} = \{m_{ij}\}$, also called the Markov transition matrix, is

$$\mathbf{M} = \mathbf{A}\,\mathbf{D}^{-1}$$

M completely defines the stochastics of information transfer over the network of residues.

Hitting time: a measure of communication efficiency between two endpoints

Based on all possible pathways



path	# of steps	Path Probability
$j \to i$	1	0.5
$j \to k \to j \to i$	3	0.5^2
$j \to k \to j \to k \to j \to i$	5	0.5^3

$$\boldsymbol{H}(j,i) = 1 \times 0.5 + 3 \times 0.5^2 + \cdots = \sum_{j=1}^{\infty} (2j-1) \times 0.5^j, = 3.$$

path	# of steps	Path Probability
$i \to j \to k$	2	0.5
$i \to j \to i \to j \to k$	4	0.5^2
$i \to j \to i \to j \to i \to j \to k$	6	0.5^3

$$\boldsymbol{H}(k,i) = 2 \times 0.5 + 4 \times 0.5^2 + \cdots = 2\sum_{j=1}^{\infty} j \times 0.5^j = 4.$$

P(t) = M P(0), where $M = AD^{-1}$ is the conditional prob matrix for signal diffusion

Fluctuations as determinant of communication

$$H(n,i) = 1 + \sum_{k=1}^{n-1} H(n,k) m_{ki}$$

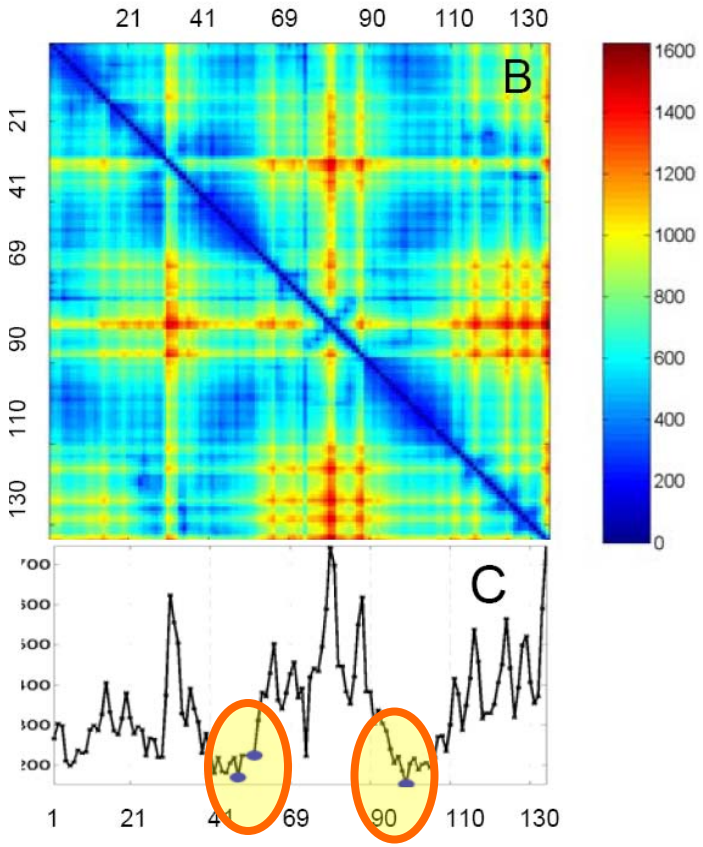
$$H(j,i) = \sum_{k=1}^{n} \left[\Gamma_{ki}^{-1} - \Gamma_{ji}^{-1} - \Gamma_{kj}^{-1} + \Gamma_{jj}^{-1} \right] d_k$$

$$C(i,j) = \left[\Gamma_{ii}^{-1} + \Gamma_{jj}^{-1} - 2\Gamma_{ij}^{-1} \right] \sum_{k=1}^{n} d_k.$$

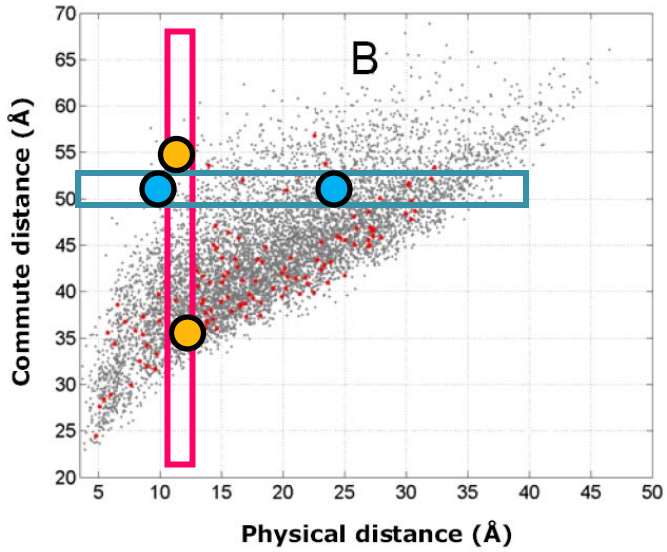
Commute distance ~ $\langle (\Delta R_{ij})^2 \rangle$

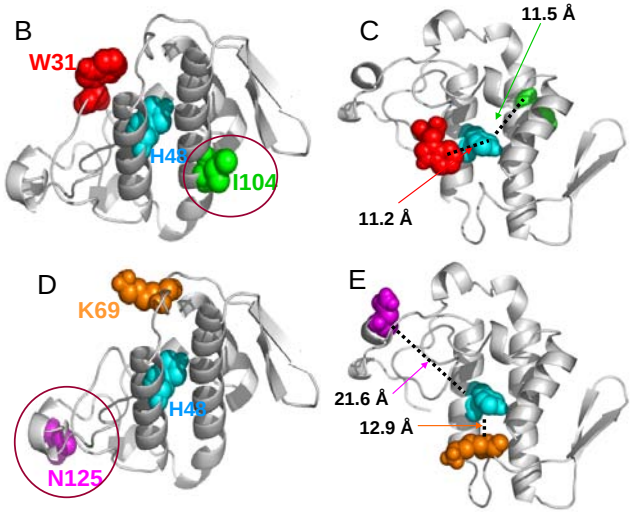
Chennubhotla & Bahar (2007) *PLoS Comp Bio*

Communication times



Distribution of Commute Times for Phospholipase A2 (1bk9)



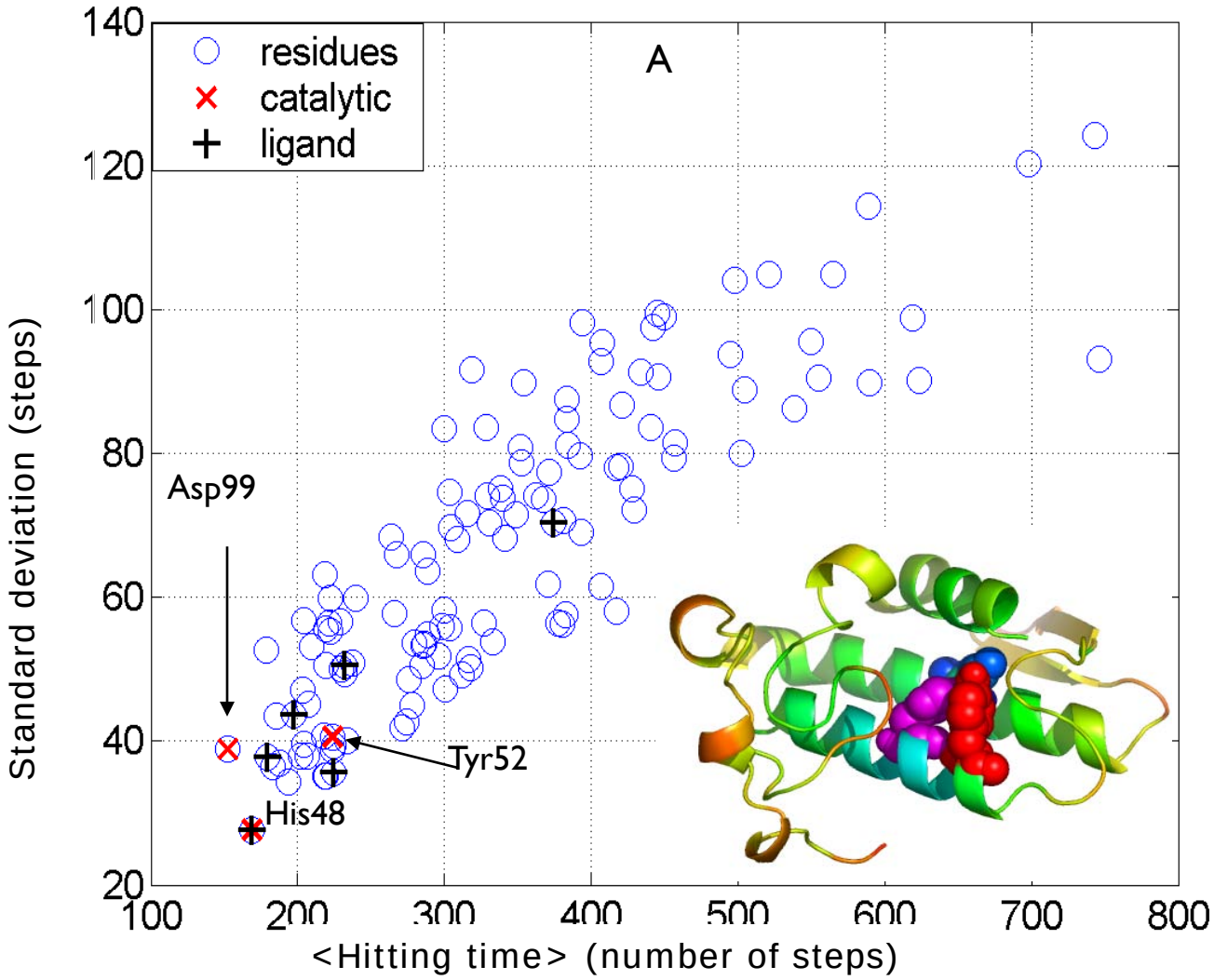


See also

His48 ,Tyr52, Asp99 – catalytic residues

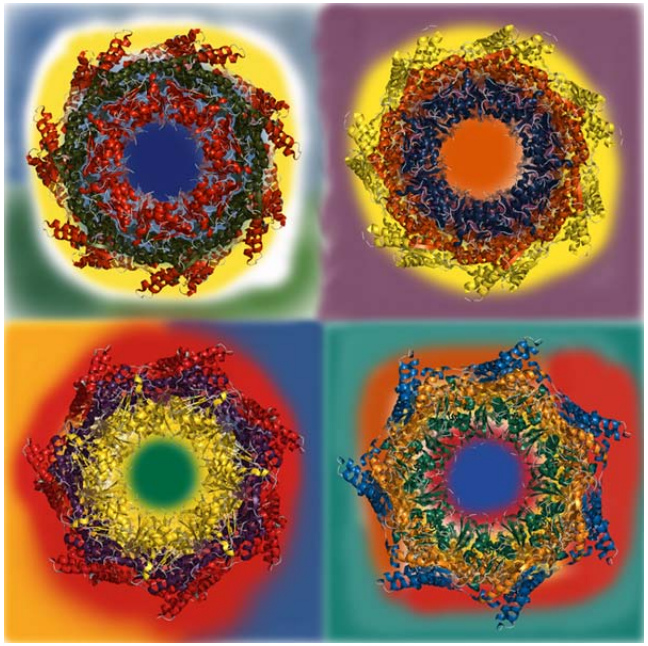
Nadler, Lafon, Kevrekidis & Coifman (2005) Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker-Planck Operators, NIPS 18; Coifman et al (2005) PNAS 102, 7426.

Active sites are distinguished by effective communication properties



Chennubhotla & Bahar (2007) *PLoS Comp Bio*

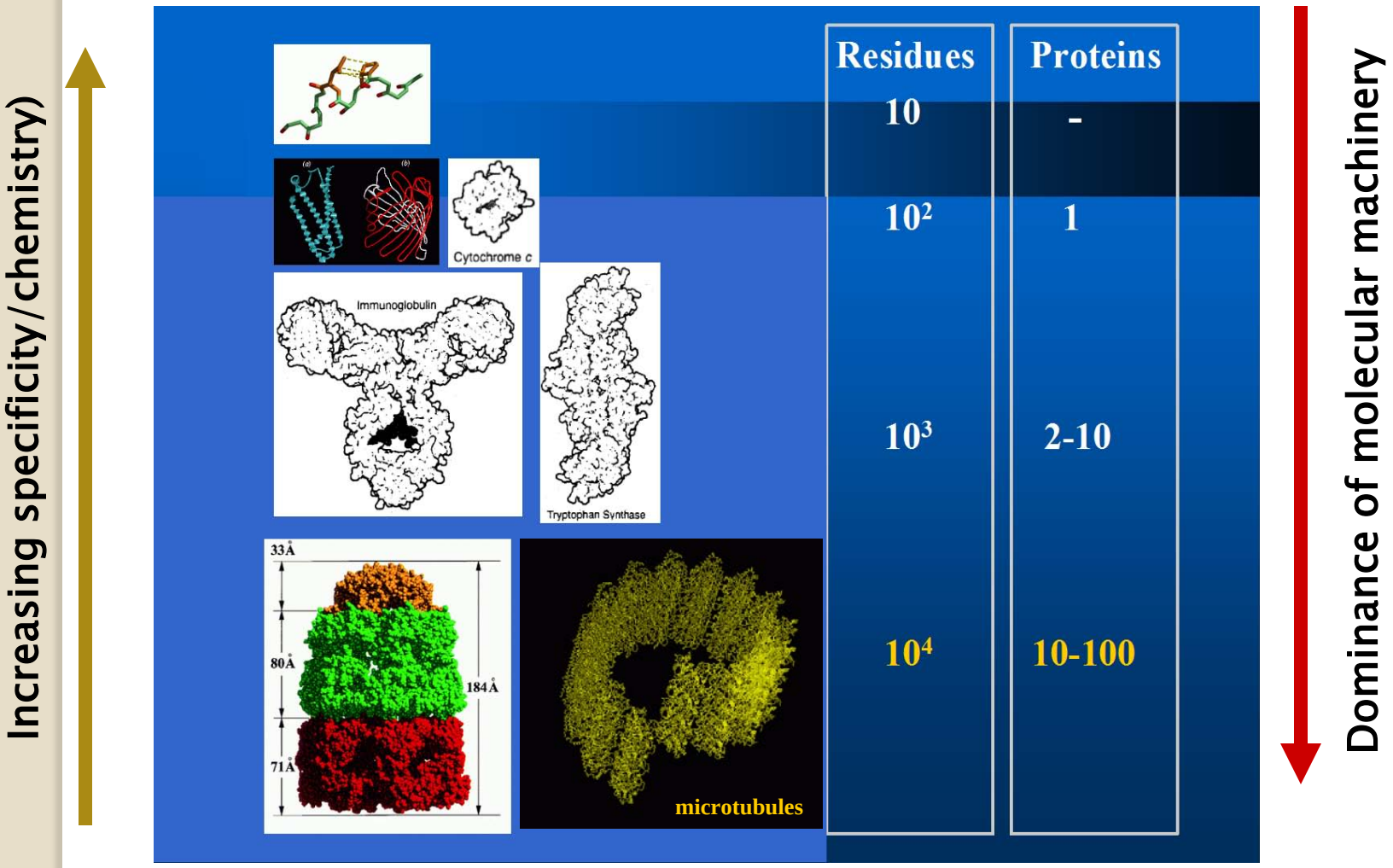
CONCLUSION



- 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?



DISCUSSION

- Different tools for different time/length windows: MD cannot explore long-time processes for multimeric systems; ANM does not incorporate detailed atomic forces

- Not all evolutionarily correlated sites refer to structural or dynamic correlations

- Accurate modeling of protein dynamics is essential to computer-aided drug discovery, but not sufficient for quantitative evaluation of binding affinity

- Druggability simulations identify druggable sites, but not the type of drugs that optimally bind those sites

Acknowledgment

Support from NIGMS, NLM, NIDDK & NIAID

Group members
- Elia Zomot
- Anindita Dutta
- Ahmet Bakan
- Ignacio General
- **Murat Can Cobanoglu**
- Tim Lezon
- Mary Cheng
- Filippo Pulara
- **Indira Shrivastava**
- Mert Gur
- Kaitlyn Hu

Former members
- Lee-Wei Yang
- **Eran Eyal**
- **Dror Tobi**
- Basak Isin
- AJ Rader
- Chakra Chennubhotla
- Enrique Marcus
- Zheng Yang
- Enrique Markus
- **Ying Liu**
- **Lin Liu**
- Lidio Meireles
- Divesh Bhatt

Collaborators
- Angela Gronenborn
- Lila Gierasch
- Benoit Roux
- Michael Tsang
- John Lazo
- Andreas Vogt
- Mike Widom
- Andrej Sali
- Klaus Schulten
- Susan Amara
- Pemra Doruker

Dr. Ahmet Bakan
Comp & Systems Biology
U of Pittsburgh

Dr. Tim Lezon
Comp & Systems Biol,
U of Pittsburgh

Anindita Dutta
CMU/Pitt PhD Program

Dr. Eran Eyal
Cancer Research Institute
Sheba Medical Center, Israel

Burak Erman
Koc University, Istanbul

Ali Rana Atilgan
Sabanci University, Istanbul

Turkan Haliloglu
Bogazici University, Istanbul