

# 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...

# Sequence evolution

## an information-theoretic approach

Residue index

$i$					$i+5$	$i+7$	$i+9$	
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	

correlated mutations

conserved

Information entropy (Shannon, 1951 )

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

Mutual information (MI)

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

for correlated mutations analysis (CMA)

# Mutual Information without the influence of phylogeny

MI<sub>p</sub> - to eliminate random noise and phylogenetic components

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

APC = Average product correction

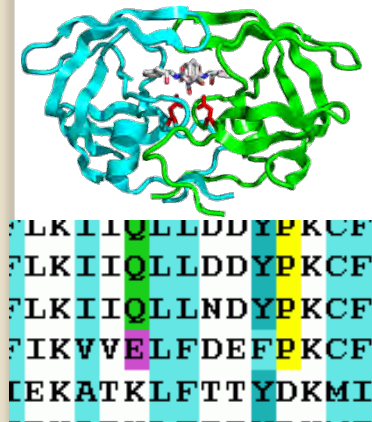
$$= [ I(i, x) I(j, x) ] / \langle I(i, j) \rangle$$

	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	

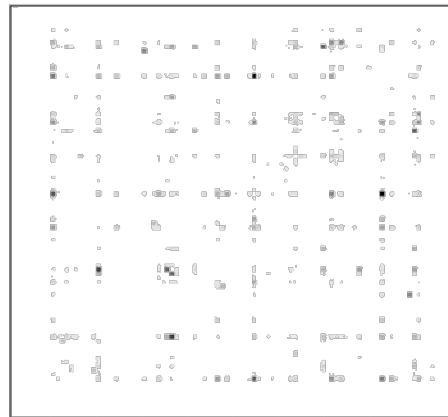
where  $I(i, x)$  is the mean mutual information of column  $i = \sum_j I(i, j)$

# HIV-I protease correlated mutation analysis (CMA)

MSA of HIV-I  
protease



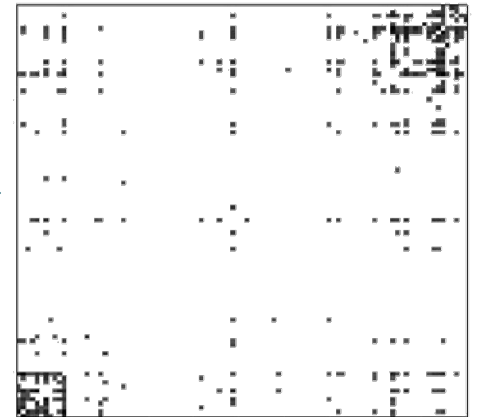
MI matrix  $\mathbf{I}_{ij} = \mathbf{I}(i, j)$



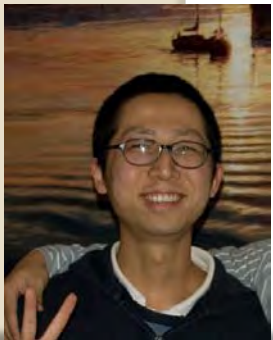
residue index

Shi and Malik (2000)

spectral clustering



reordered residue index



Dr. Ying Liu

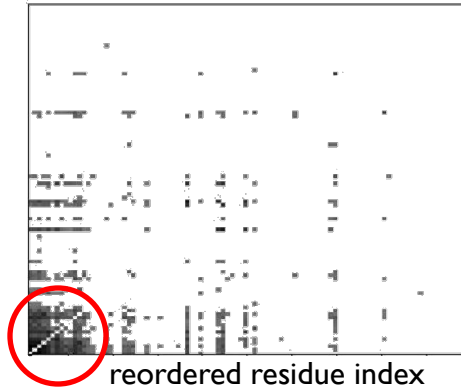
# MDR mutations distinguished by CMA

## MSA of HIV-1 protease

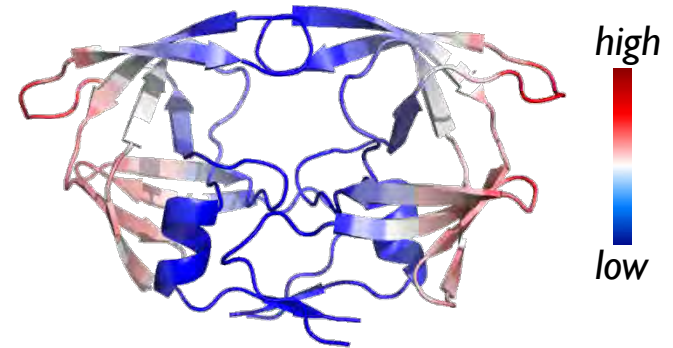
Stanford HIV Drug Resistance Database  
<http://hivdb.stanford.edu/>

```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR  
CDKFGIVVHELGHVVGFWHEHTRPDREDHV  
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL  
NFTVGSLLIHEIGHAFGLIHEHQRPRDDYVI  
CLTYGTPIELMHALGFFPHEQNRHERDSYVR  
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV  
CDKFGVVVHELGHVVGFWHEHTRPDNEFVG  
CAYFGTIVHEIGHAIGFHHQSRPDRDDYIN  
CVYHGIIQHELSHALGFYHEHTRSDRNKYVR  
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

untreated

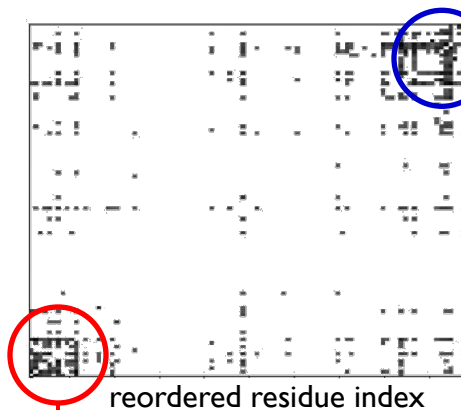


mobility profile



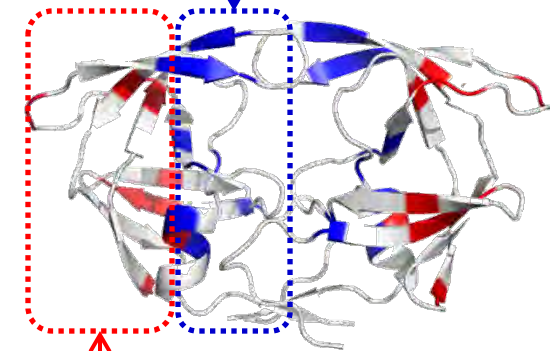
```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR  
CDKFGIVVHELGHVVGFWHEHTRPDREDHV  
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL  
NFTVGSLLIHEIGHAFGLIHEHQRPRDDYVI  
CLTYGTPIELMHALGFFPHEQNRHERDSYVR  
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV  
CDKFGVVVHELGHVVGFWHEHTRPDNEFVG  
CAYFGTIVHEIGHAIGFHHQSRPDRDDYIN  
CVYHGIIQHELSHALGFYHEHTRSDRNKYVR  
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

treated by at least  
one drug



Drug-resistant cluster

Phylogenetic cluster



# Summary

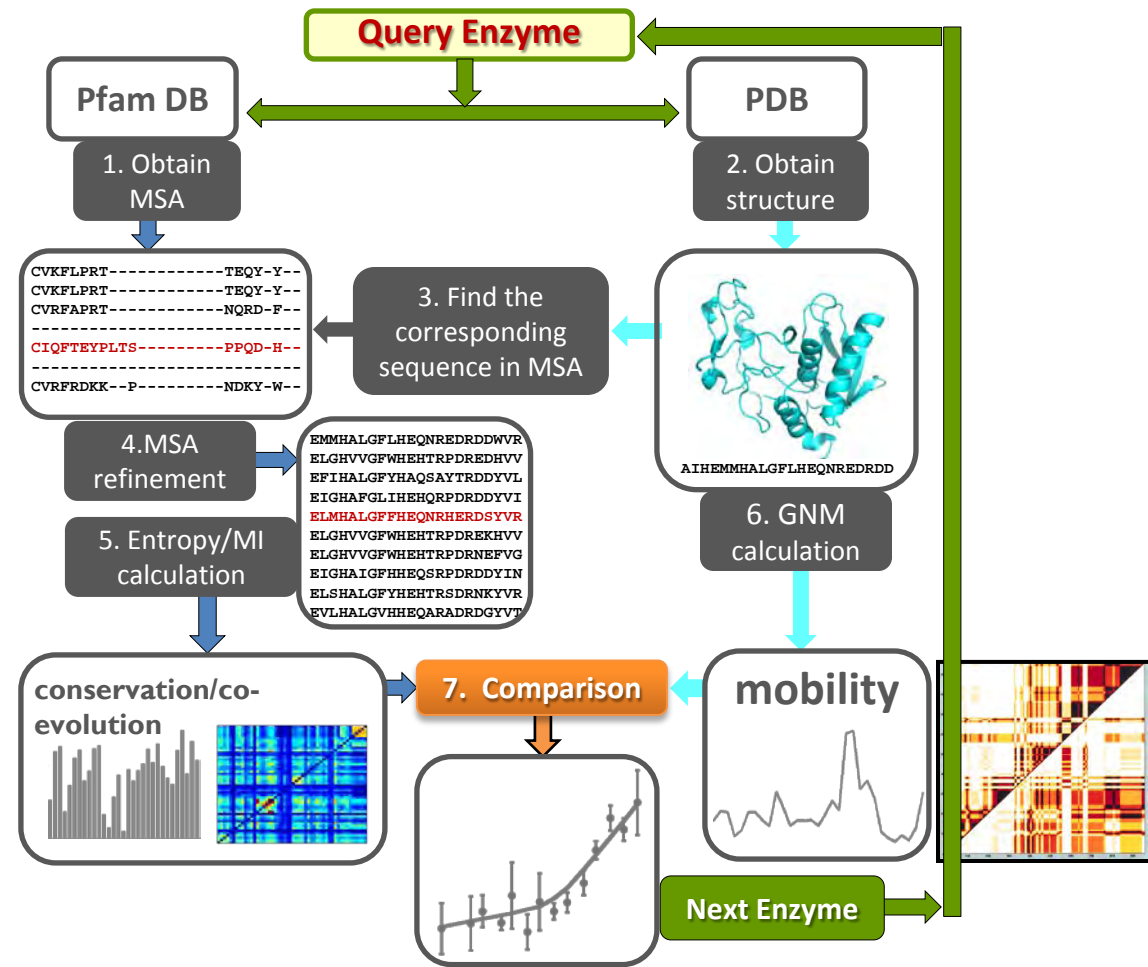
- two groups of correlated mutation sites

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

# Questions:

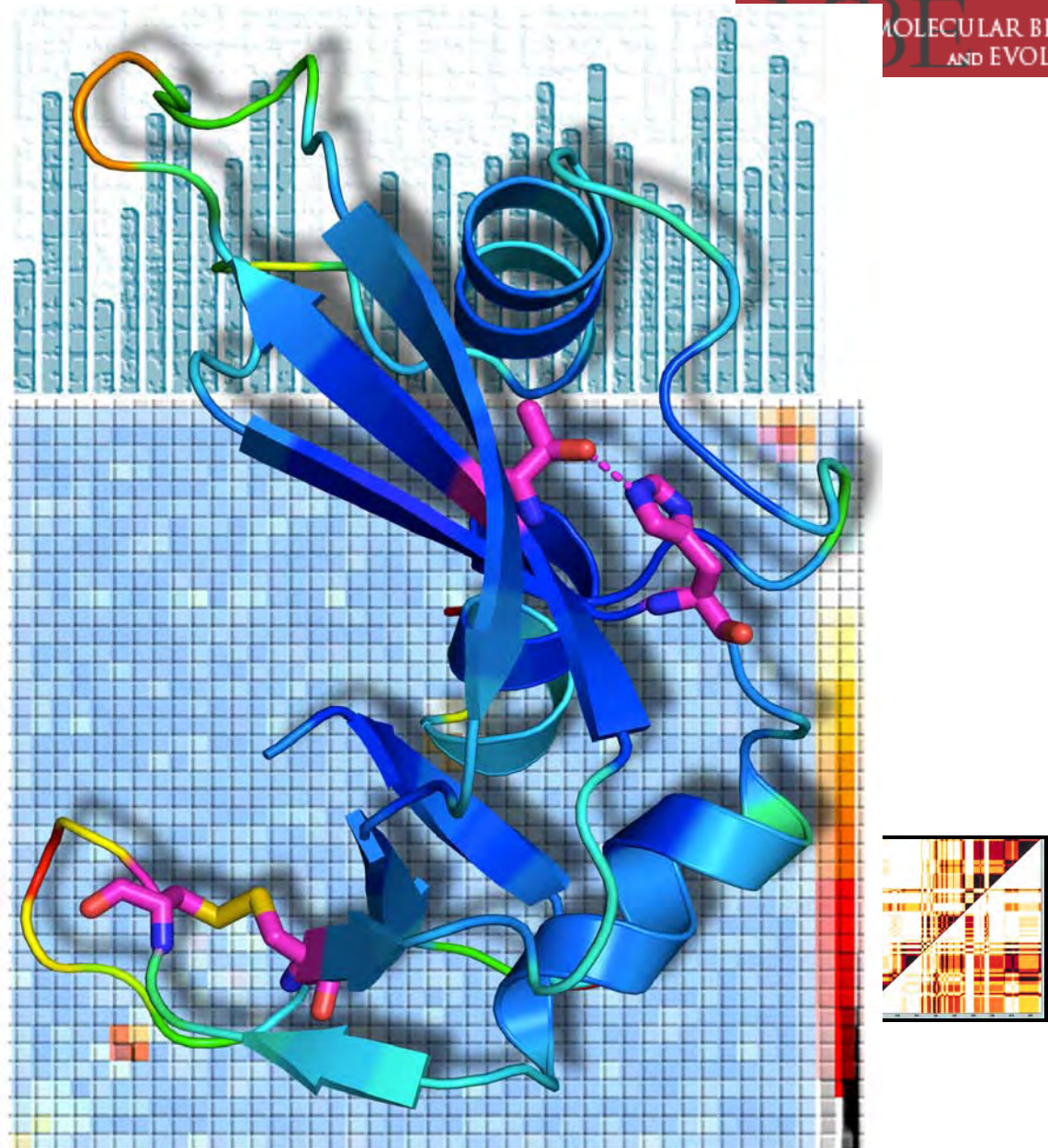
- Are key mechanical sites (e.g. hinges) conserved?
- Is there any correlation between sequence variability and structural dynamics?
- How does the structure ensure substrate specificity *and* conformational adaptability?

# A systematic study of a set of enzymes

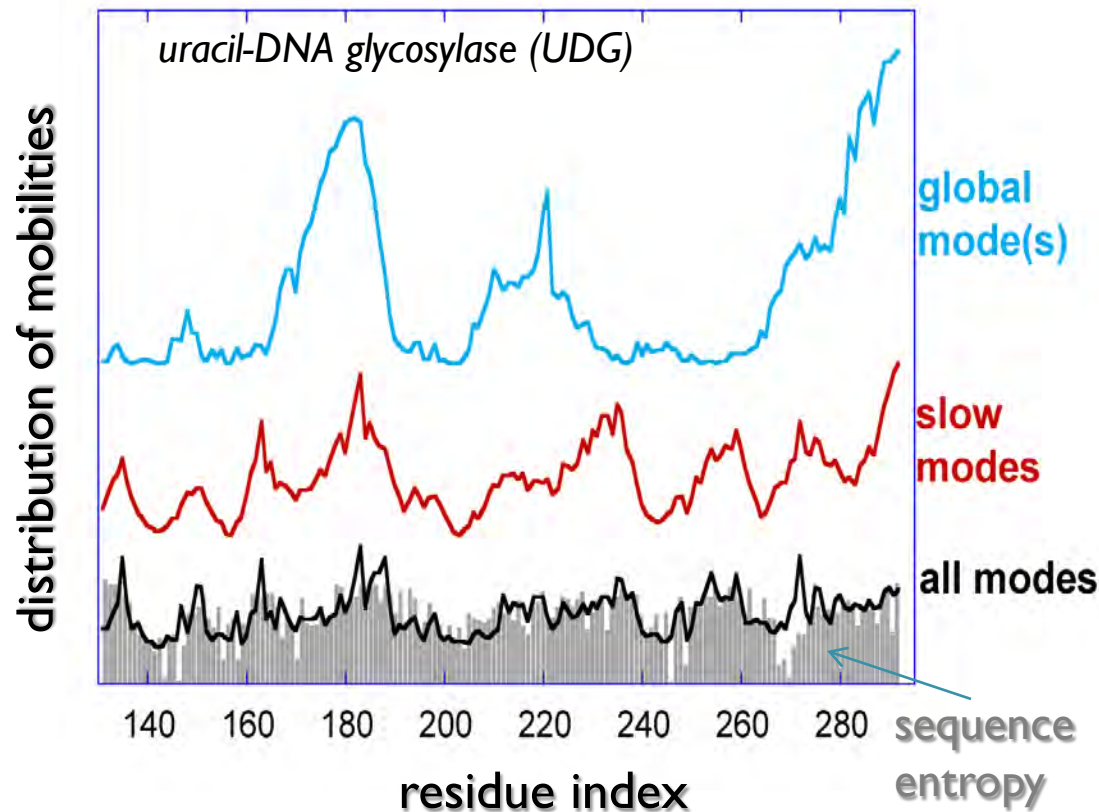




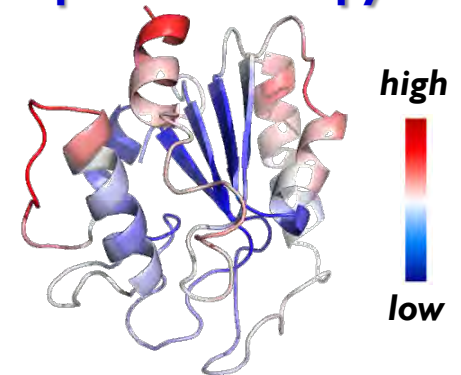
# Evol



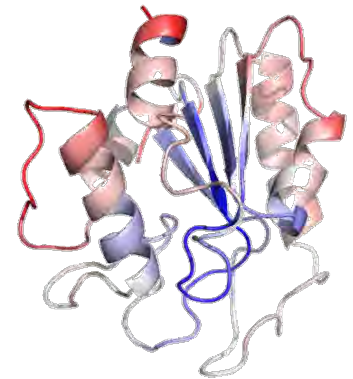
# Correlation between sequence entropy & conformational mobility



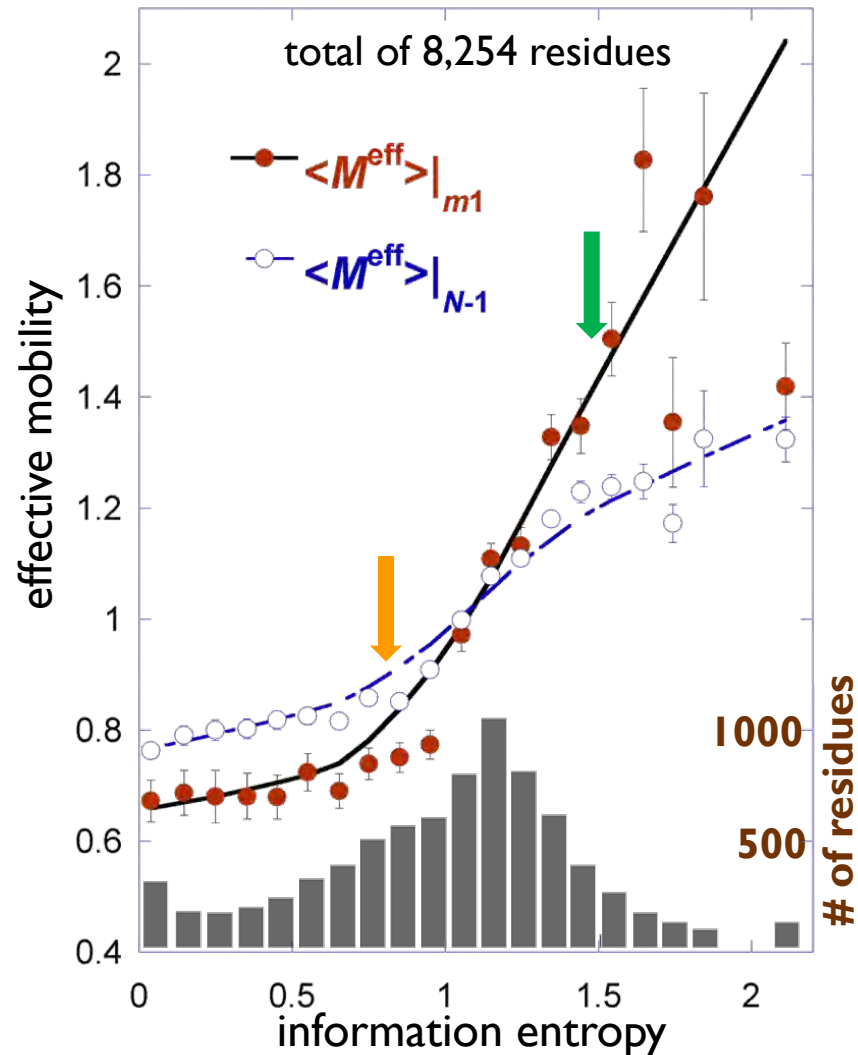
sequence entropy



structural dynamics



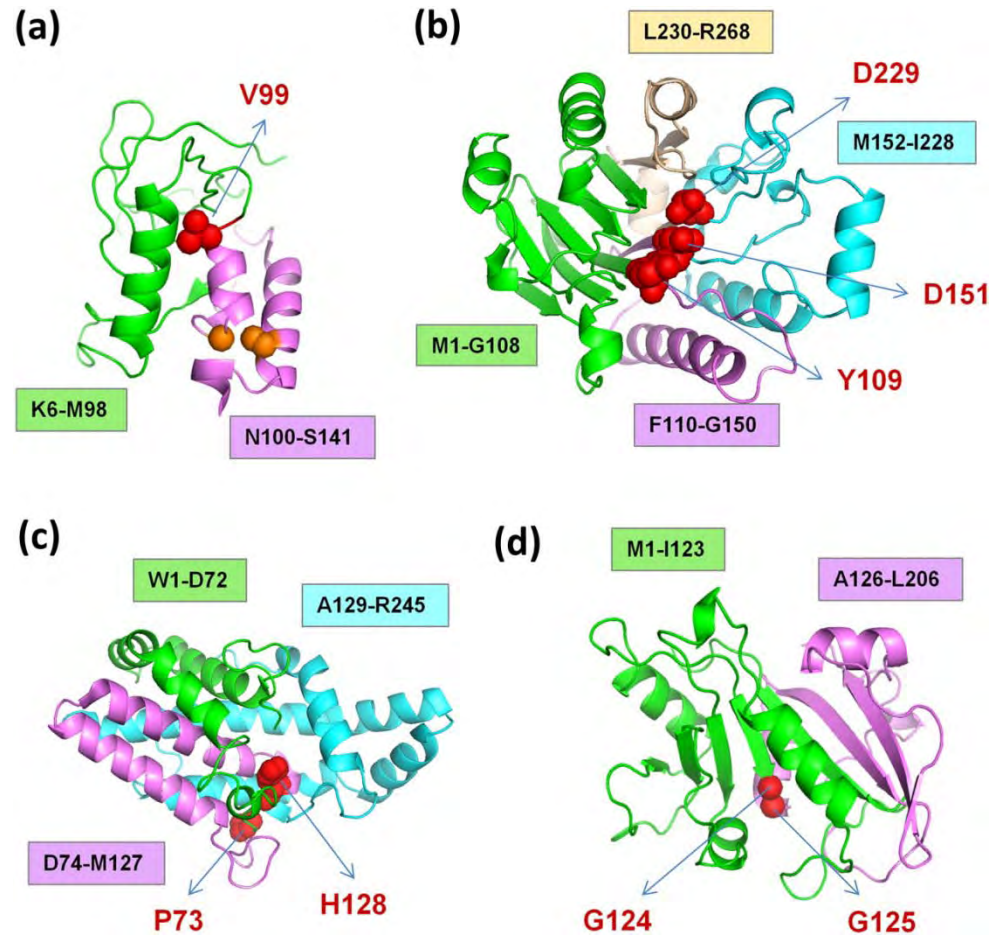
# Mobility increases with sequence entropy



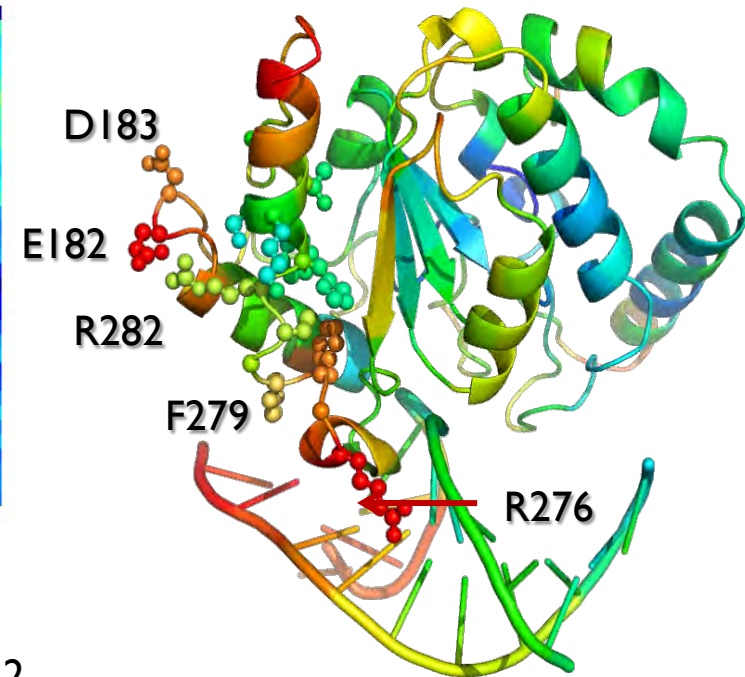
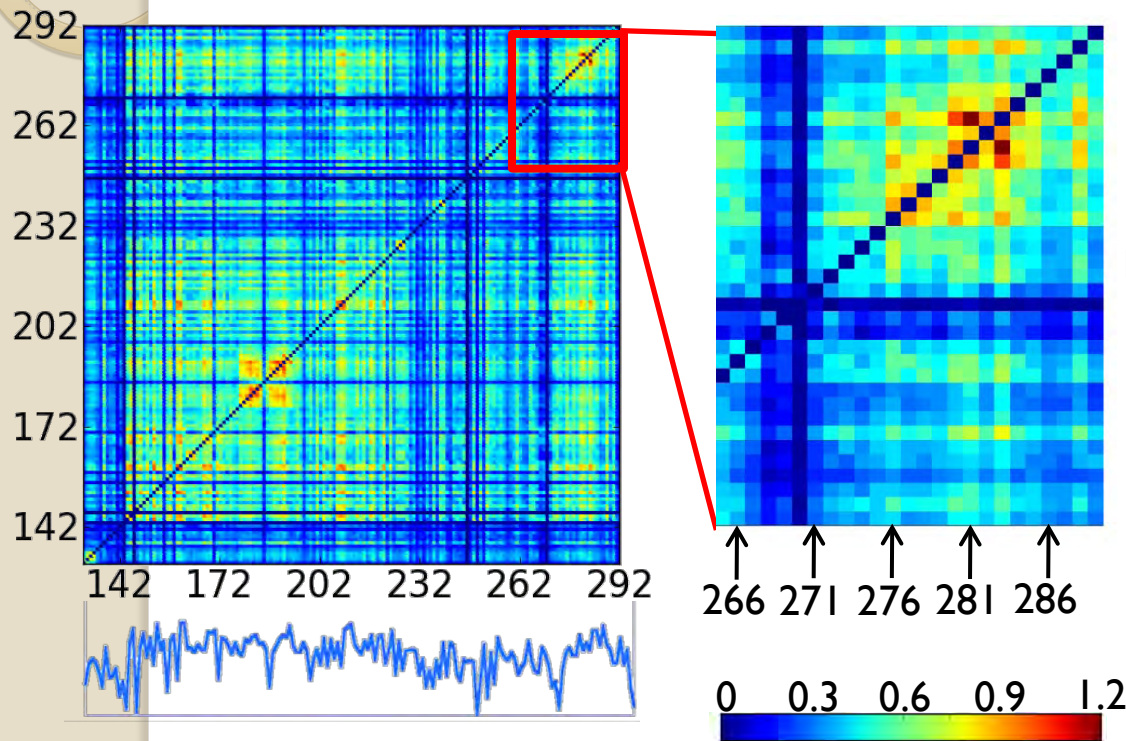


# Hinge sites are evolutionarily conserved

*despite their moderate-to-high exposure to environment*

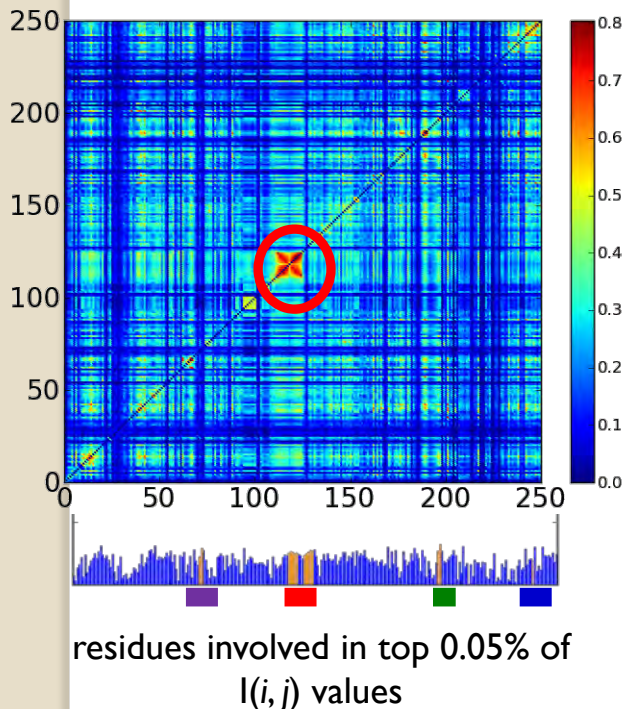


Amino acids involved in intermolecular recognition are distinguished by **their co-evolution propensities**

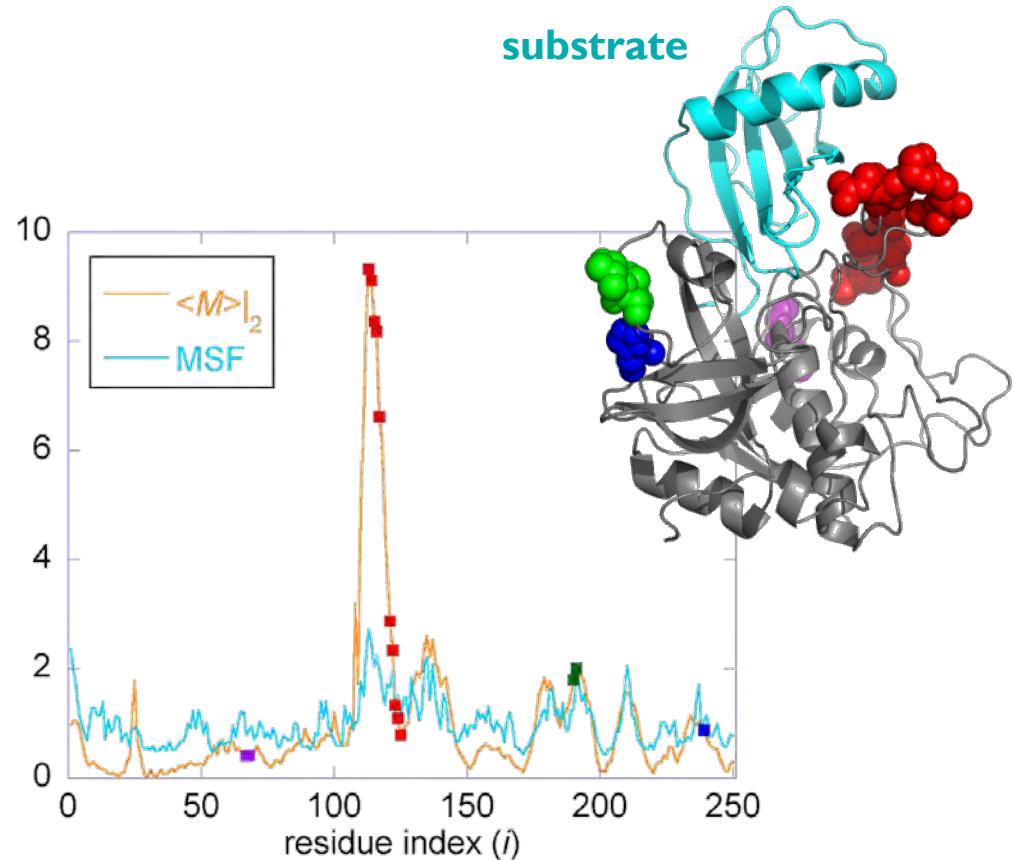


Amino acids involved in intermolecular recognition are distinguished by **their high global mobility**

*cathepsin B*



substrate



# 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 recognition (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  $\Gamma$

## 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

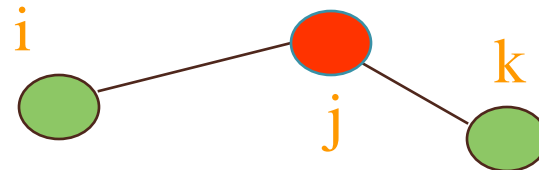
$$\mathbf{M} = \mathbf{A} \mathbf{D}^{-1}$$

$\mathbf{M}$  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
$j \rightarrow i$	1	0.5
$j \rightarrow k \rightarrow j \rightarrow i$	3	$0.5^2$
$j \rightarrow k \rightarrow j \rightarrow k \rightarrow j \rightarrow i$	5	$0.5^3$



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

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

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

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

# Fluctuations as determinant of communication

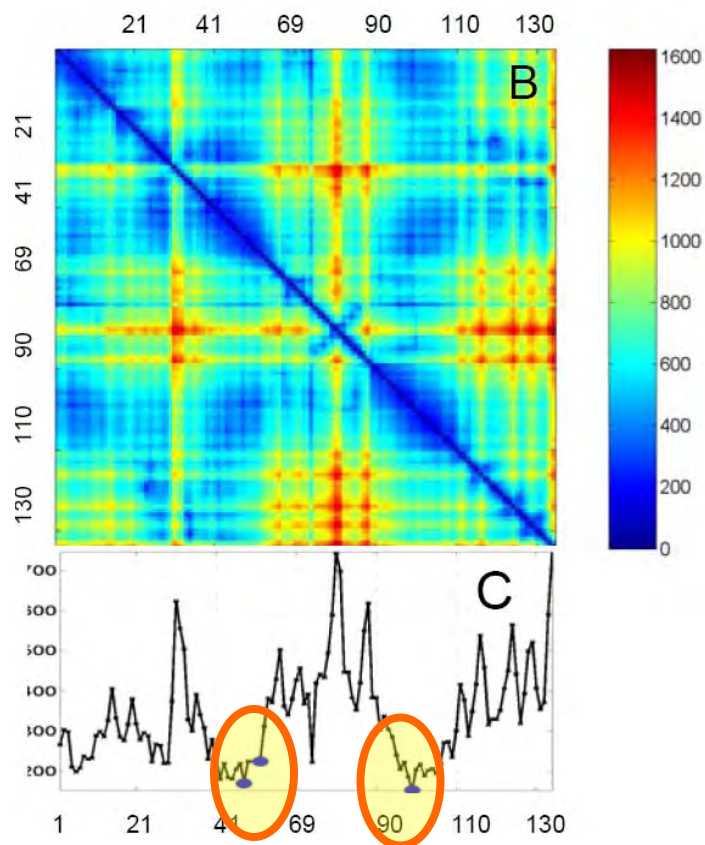
$$H(n, i) = 1 + \sum_{k=1}^{n-1} H(n, k) m_{ki}$$

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

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

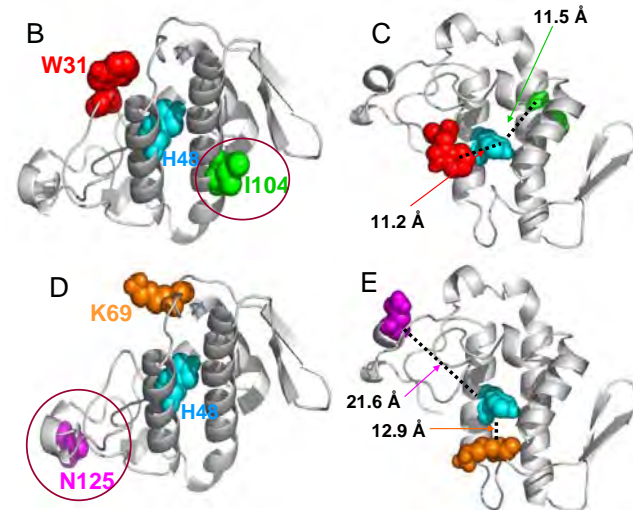
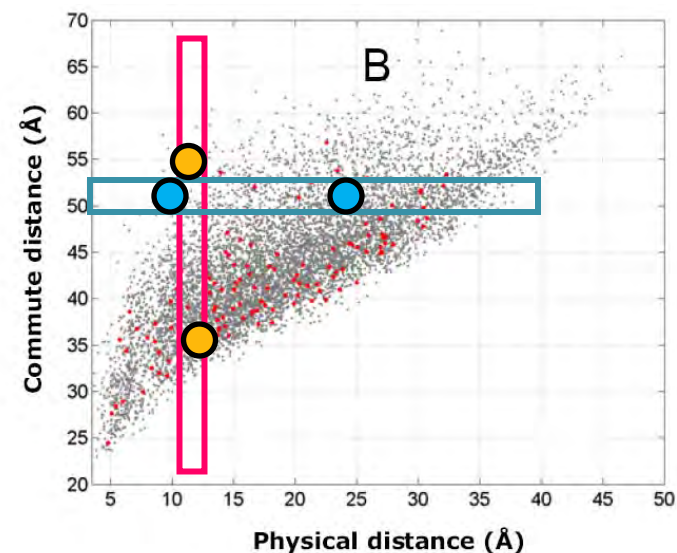
$$C(i, j) = [\Gamma_{ii}^{-1} + \Gamma_{jj}^{-1} - 2\Gamma_{ij}^{-1}] \sum_{k=1}^n d_k.$$

# Communication times



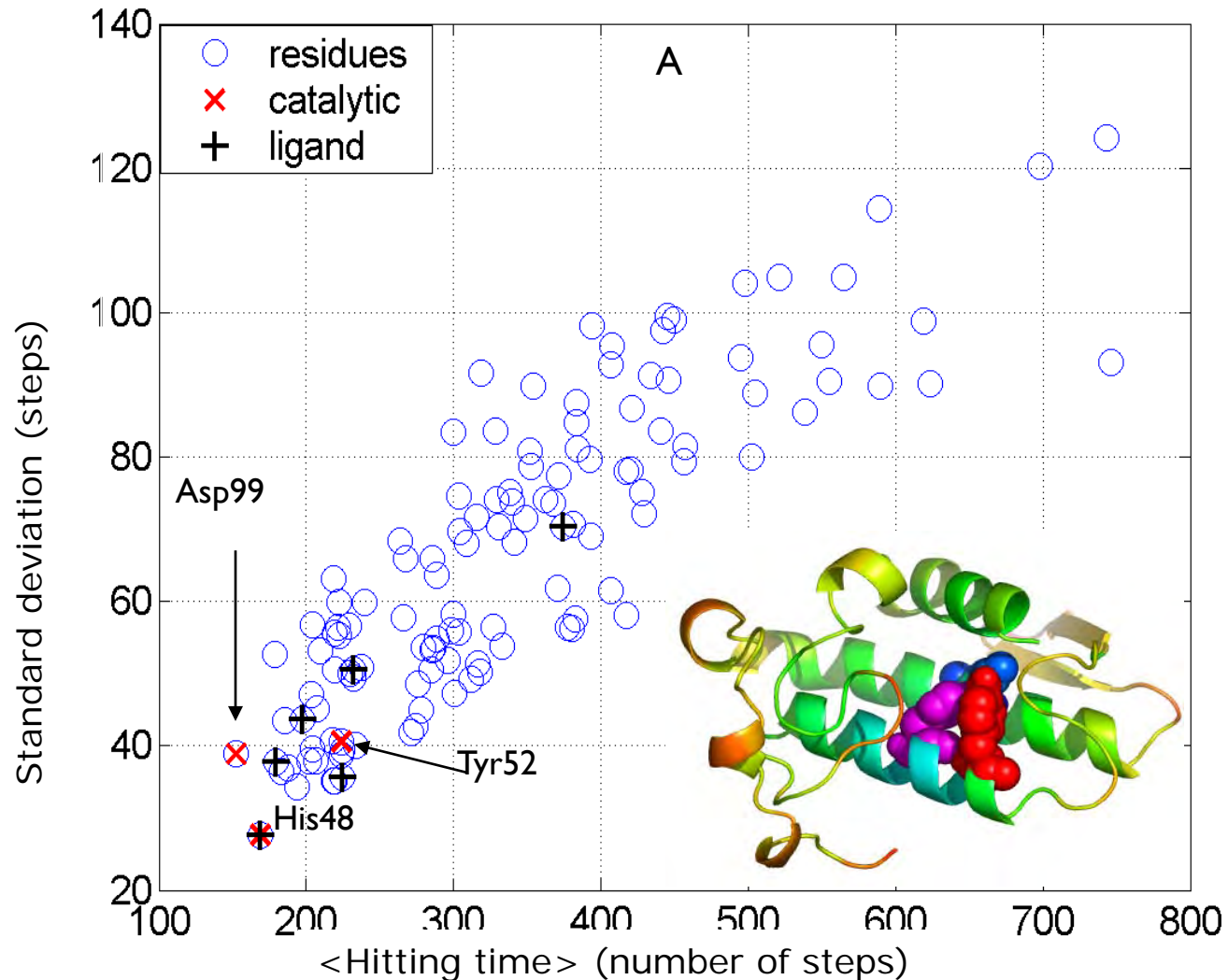
Distribution of Commute Times for Phospholipase A2 (1bk9)

His48 ,Tyr52, Asp99 – catalytic residues

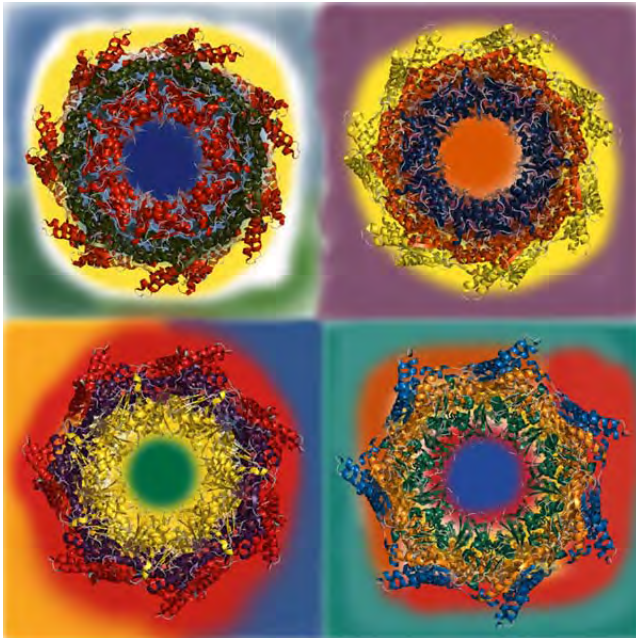


See also

# Active sites are distinguished by effective communication properties



# 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

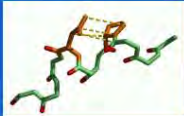
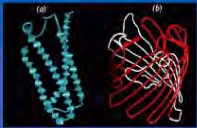

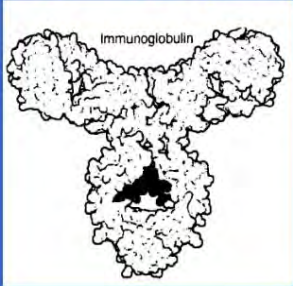
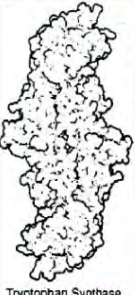
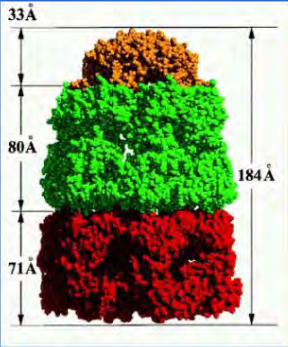



# Mechanics vs chemistry?

How does complexity scale with size of the system?

Increasing specificity/chemistry



		Residues	Proteins
		10	-
  Cytochrome c		$10^2$	1
 Immunoglobulin  Tryptophan Synthase		$10^3$	2-10
		$10^4$	10-100
 microtubules			



Dominance of molecular machinery

# 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



## 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 Salic
- Klaus Schulten
- Susan Amara
- Pemra Doruker

Support from NIGMS, NLM, NIDDK & NIAID



# Acknowledgment



**Dr. Ahmet Bakan**  
Comp & Systems Biology  
U of Pittsburgh



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



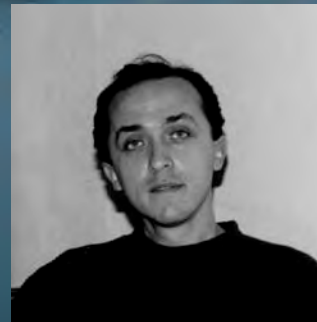
**Anindita Dutta**  
CMU/Pitt PhD Program



**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