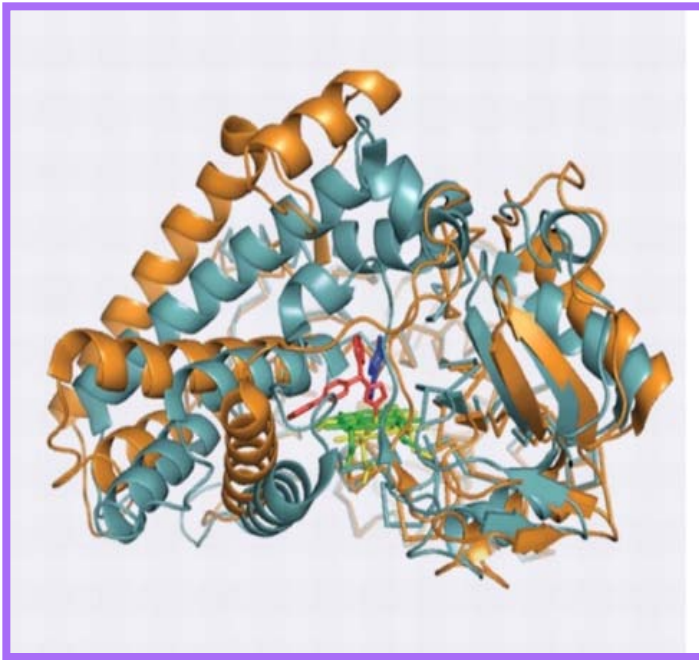


# Intrinsically accessible motions enable Optimal binding of substrate or drugs



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</i>					<i>i</i> +5	<i>i</i> +7	<i>i</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

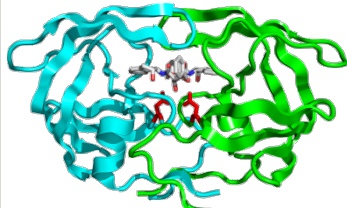
$$= [ I(i, x) I(j, x) ] / \langle I(i, j) \rangle$$

	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  $i = \sum_j I(i, j)$

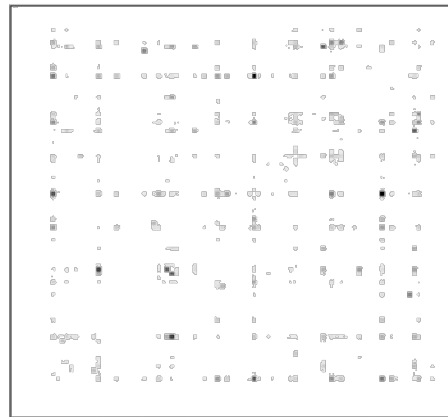
# HIV-1 protease correlated mutation analysis (CMA)

MSA of HIV-1 protease



```
FLKIIQLLDDYPKCF  
FLKIIQLLDDYPKCF  
FLKIIQLLNDYPKCF  
FIKVVELFDEFPKCF  
LEKATKLFTTYDKMI
```

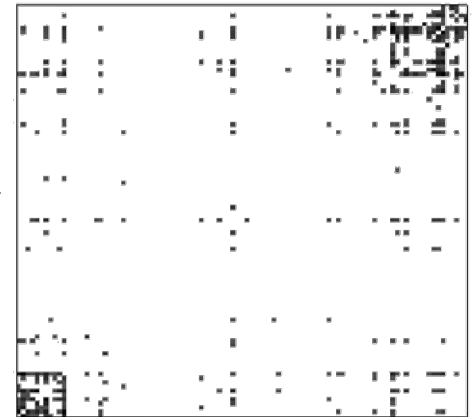
MI matrix  $\mathbf{I}_{ij} = I(i, j)$



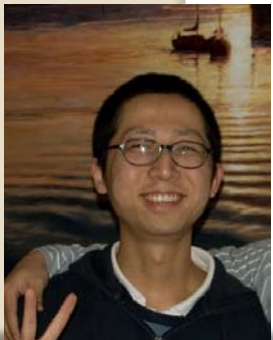
residue index

Shi and Malik (2000)

spectral clustering



reordered residue index



Dr. Ying Liu

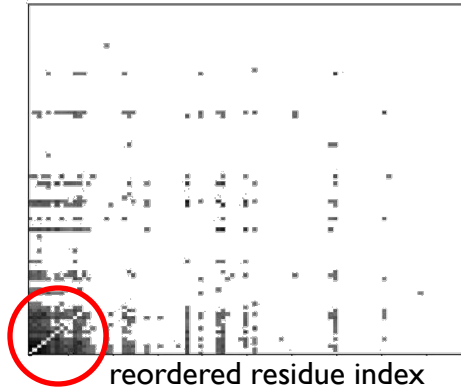
# MDR mutations distinguished by CMA

## MSA of HIV-1 protease

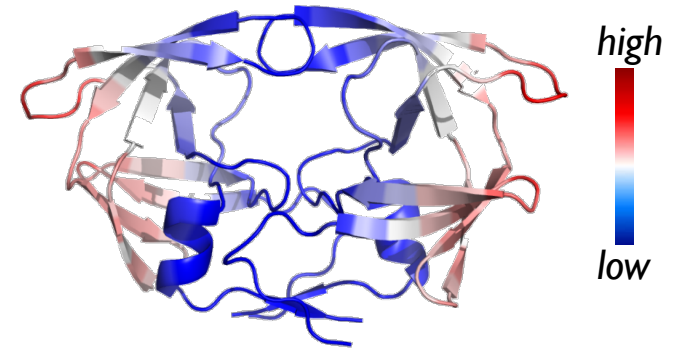
Stanford HIV Drug Resistance Database  
<http://hivdb.stanford.edu/>

```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR  
CDKFGIVVHELGHVVGFWHEHTRPDREDHVV  
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL  
NFTVGSLEIHEIGHAFGLIHEHQRPDRDDYVI  
CLTYGTPIHELMHALGFFHEQNRHERDSYVR  
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV  
CDKFGIVVHELGHVVGFWHEHTRPDREHVV  
CAYFGTIVHEIGHAIGFHEQSRPDRDDYIN  
CVYHGIIQHELSHALGFYHEHTRSRNKYVR  
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

untreated

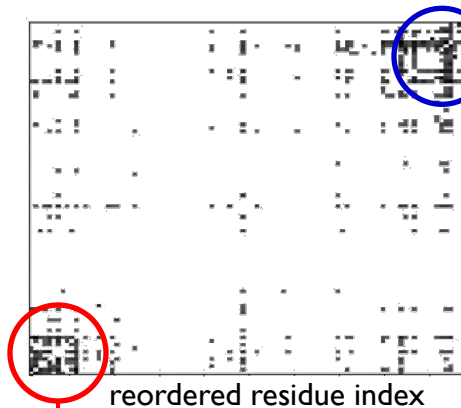


mobility profile



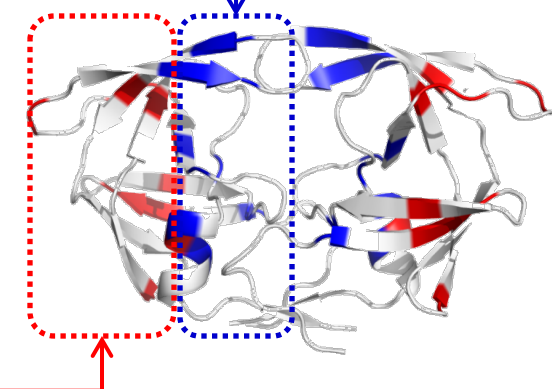
```
CTLVGTAIHEMMHALGFLHEQNREDRDDWVR  
CDKFGIVVHELGHVVGFWHEHTRPDREDHVV  
CFRFGTVIHEFIHALGFYHAQSAYTRDDYVL  
NFTVGSLEIHEIGHAFGLIHEHQRPDRDDYVI  
CLTYGTPIHELMHALGFFHEQNRHERDSYVR  
CDKFGIVVHELGHVVGFWHEHTRPDREKHVV  
CDKFGIVVHELGHVVGFWHEHTRPDREHVV  
CAYFGTIVHEIGHAIGFHEQSRPDRDDYIN  
CVYHGIIQHELSHALGFYHEHTRSRNKYVR  
CINSGTIIHEVLHALGVHHEQARADRDGYVT
```

treated by at least  
one drug



Drug-resistant cluster

Phylogenetic cluster



# Summary

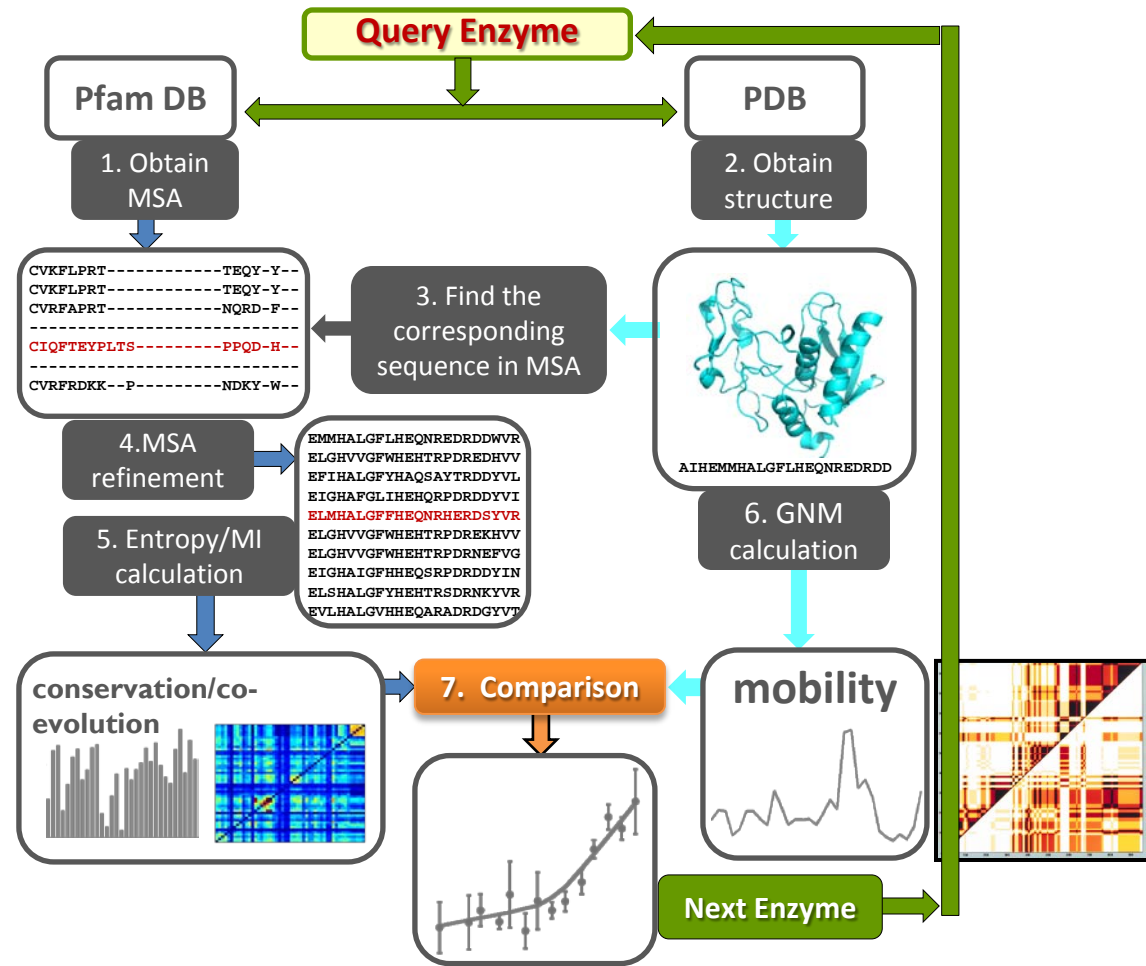
- two groups of correlated mutation sites

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

# Questions:

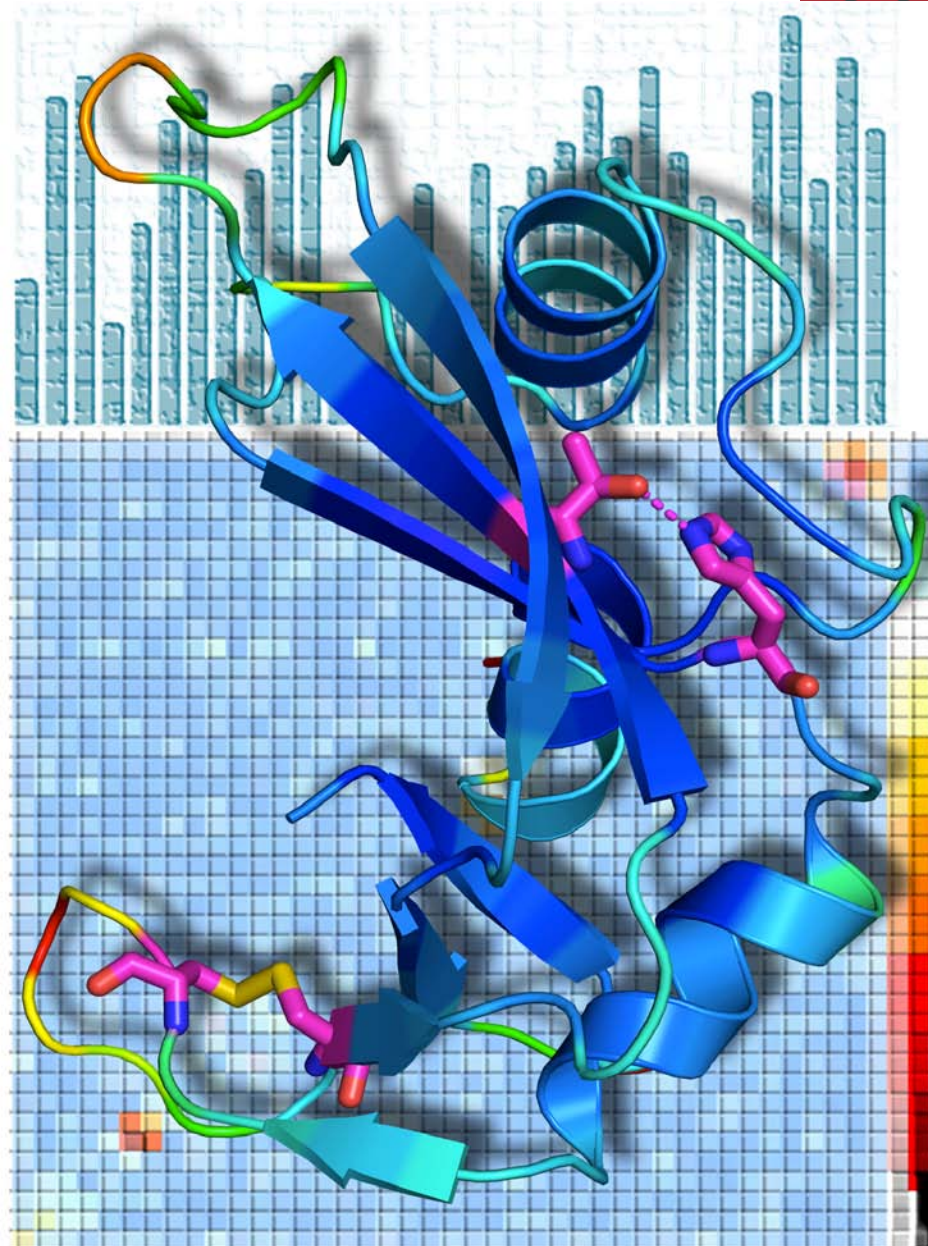
- Are key mechanical sites (e.g. hinges) conserved?
- Is there any correlation between sequence variability and structural dynamics?
- How does the structure ensure substrate specificity *and* conformational adaptability?

# A systematic study of a set of enzymes

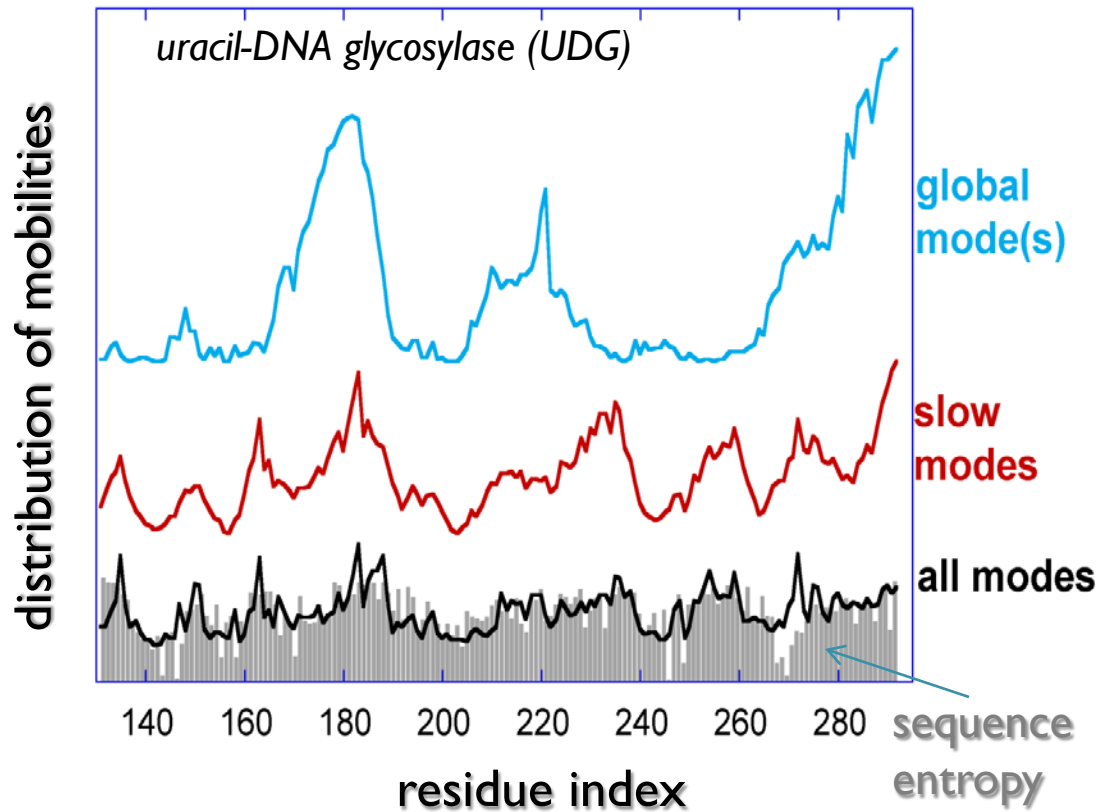




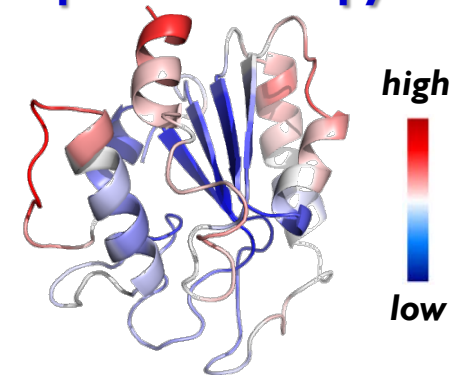
# Evol



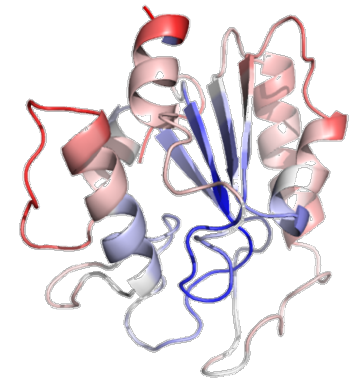
# Correlation between sequence entropy & conformational mobility



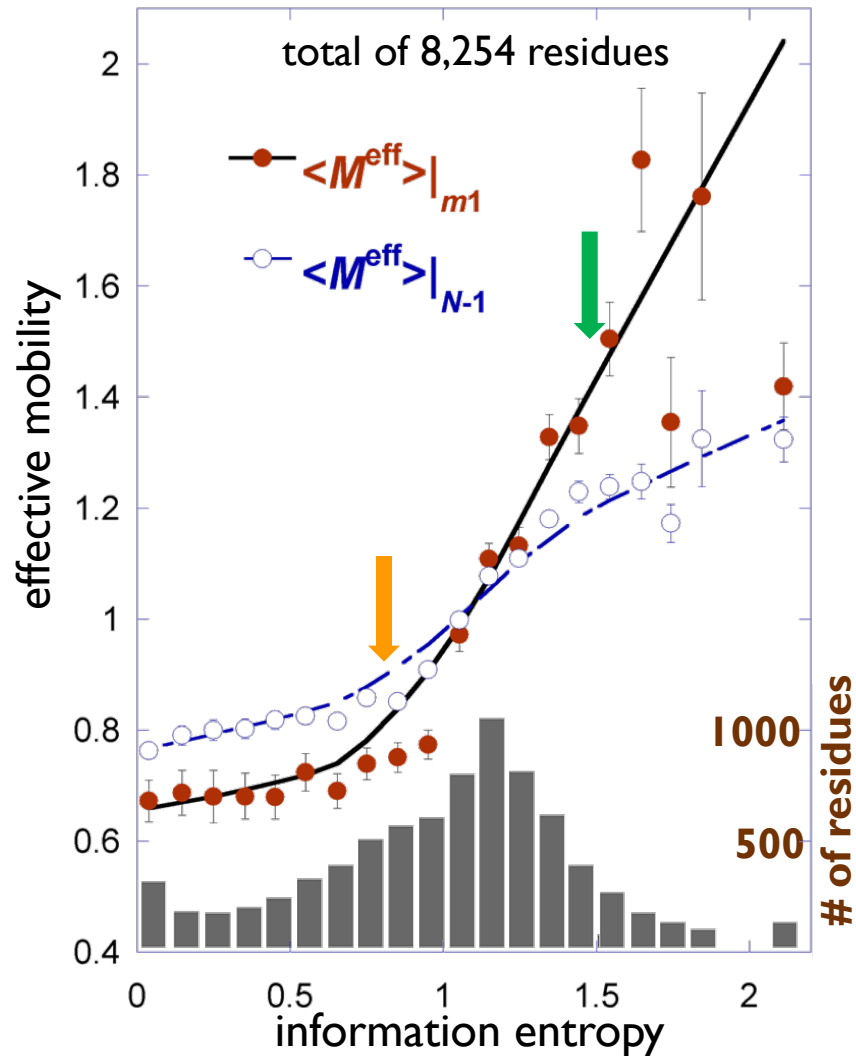
sequence entropy



structural dynamics

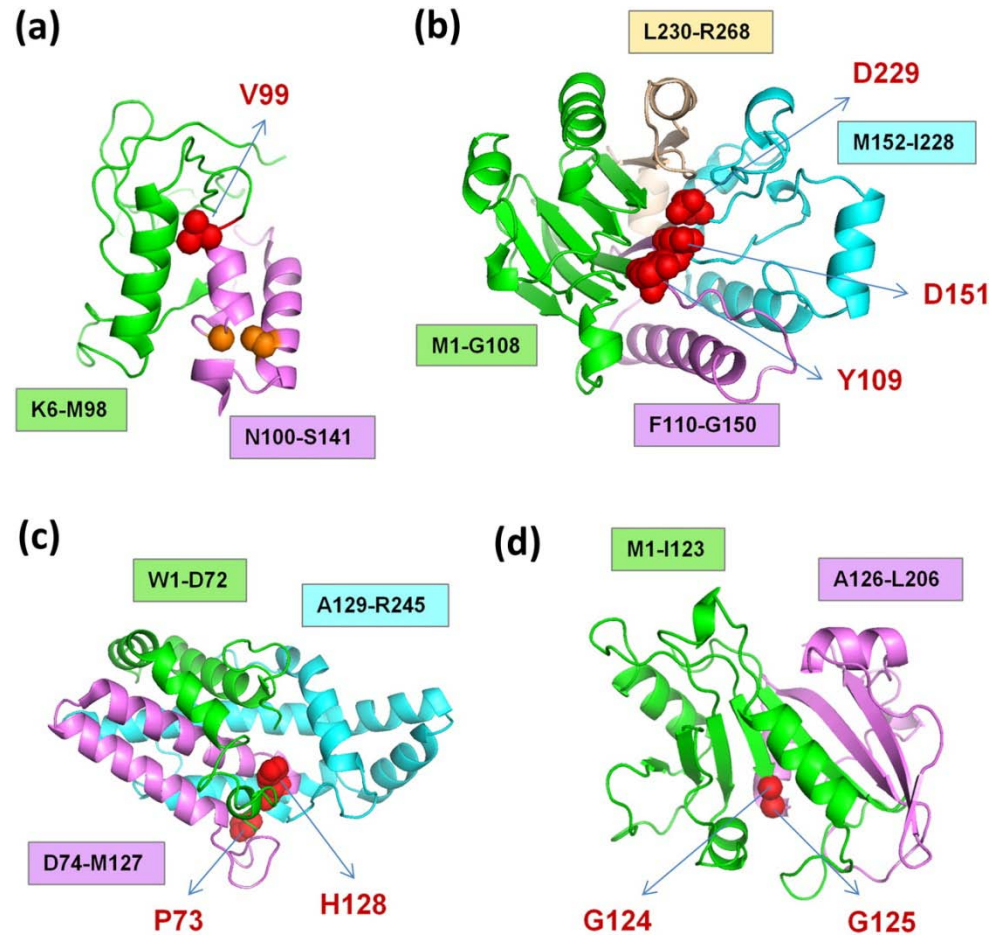


# Mobility increases with sequence entropy



# Hinge sites are evolutionarily conserved

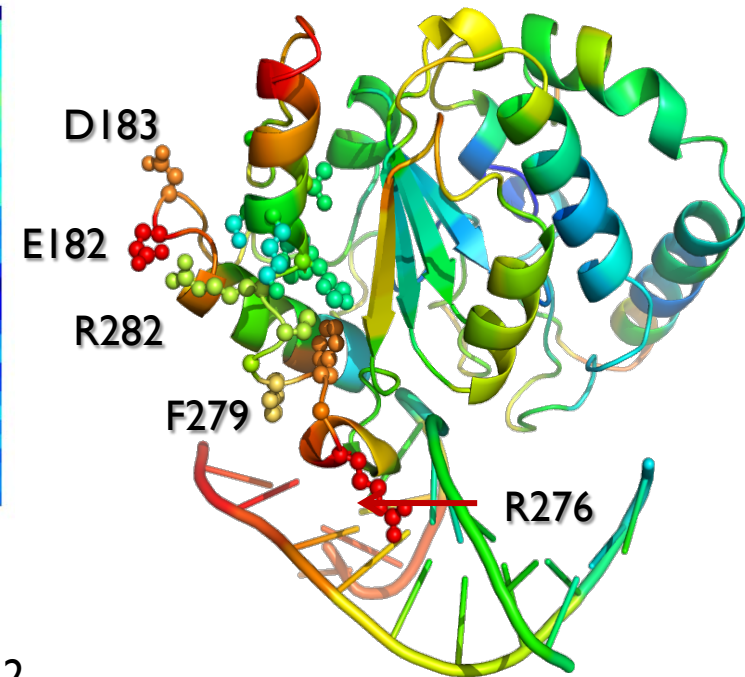
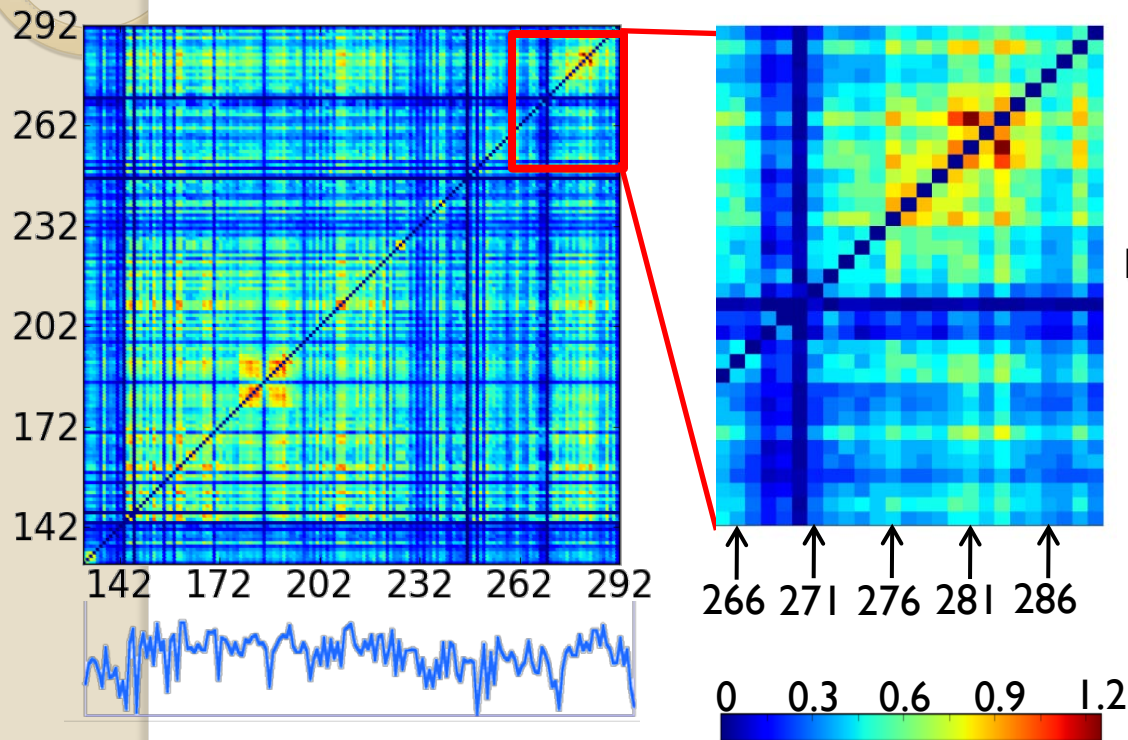
*despite their moderate-to-high exposure to environment*



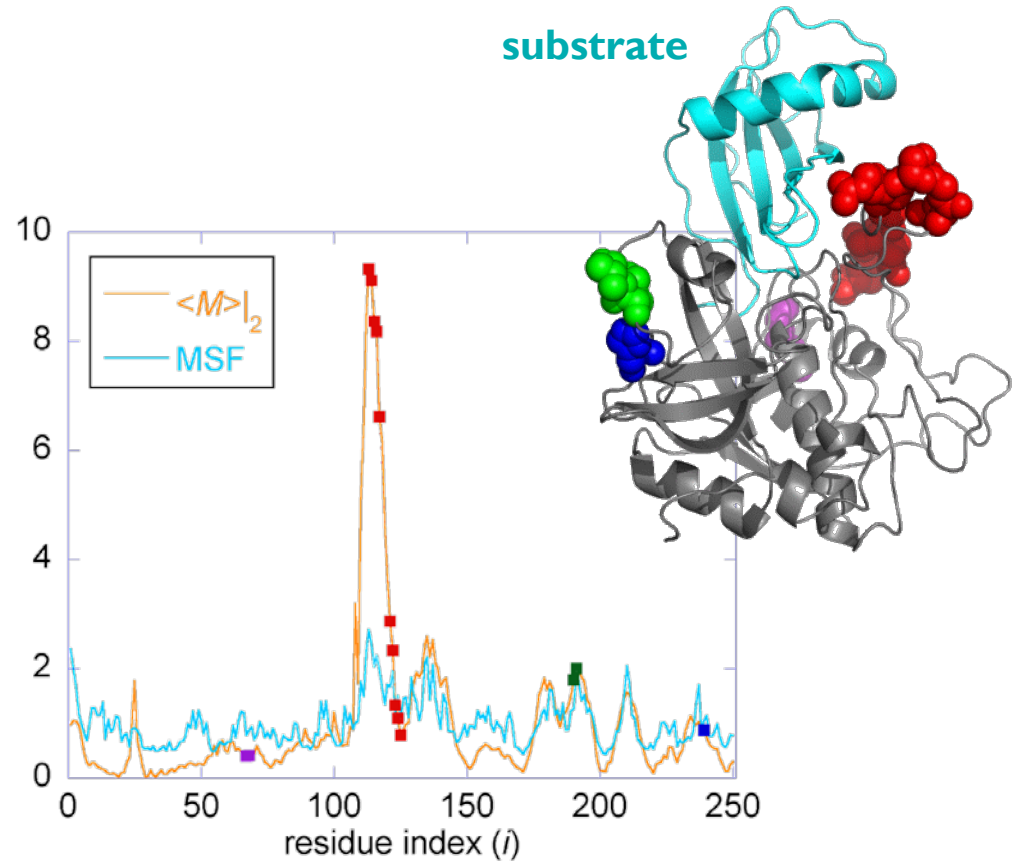
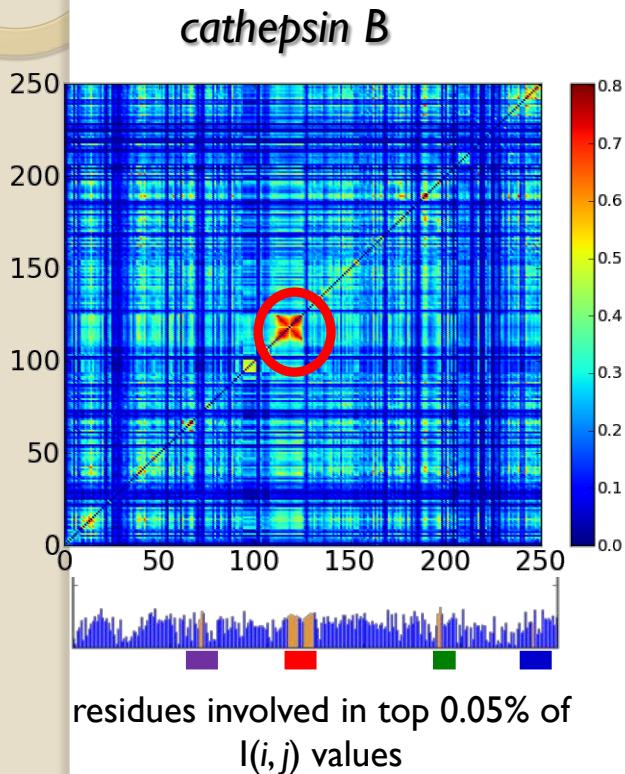


3

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**

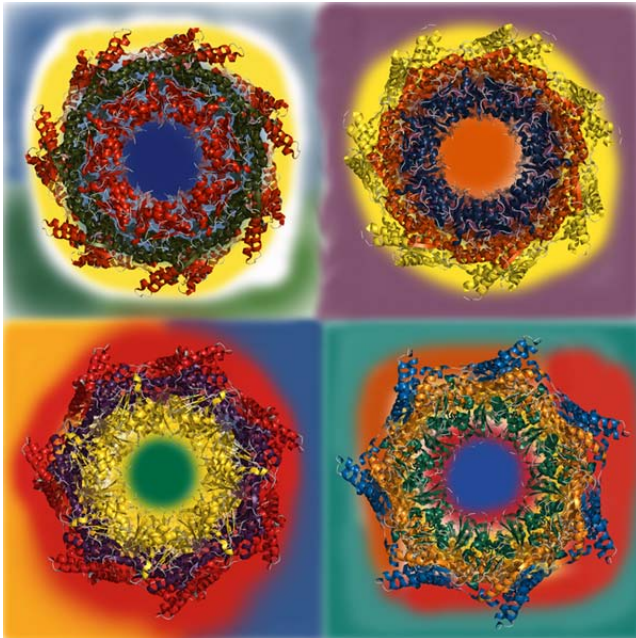


# Summary

## 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

# CONCLUSION



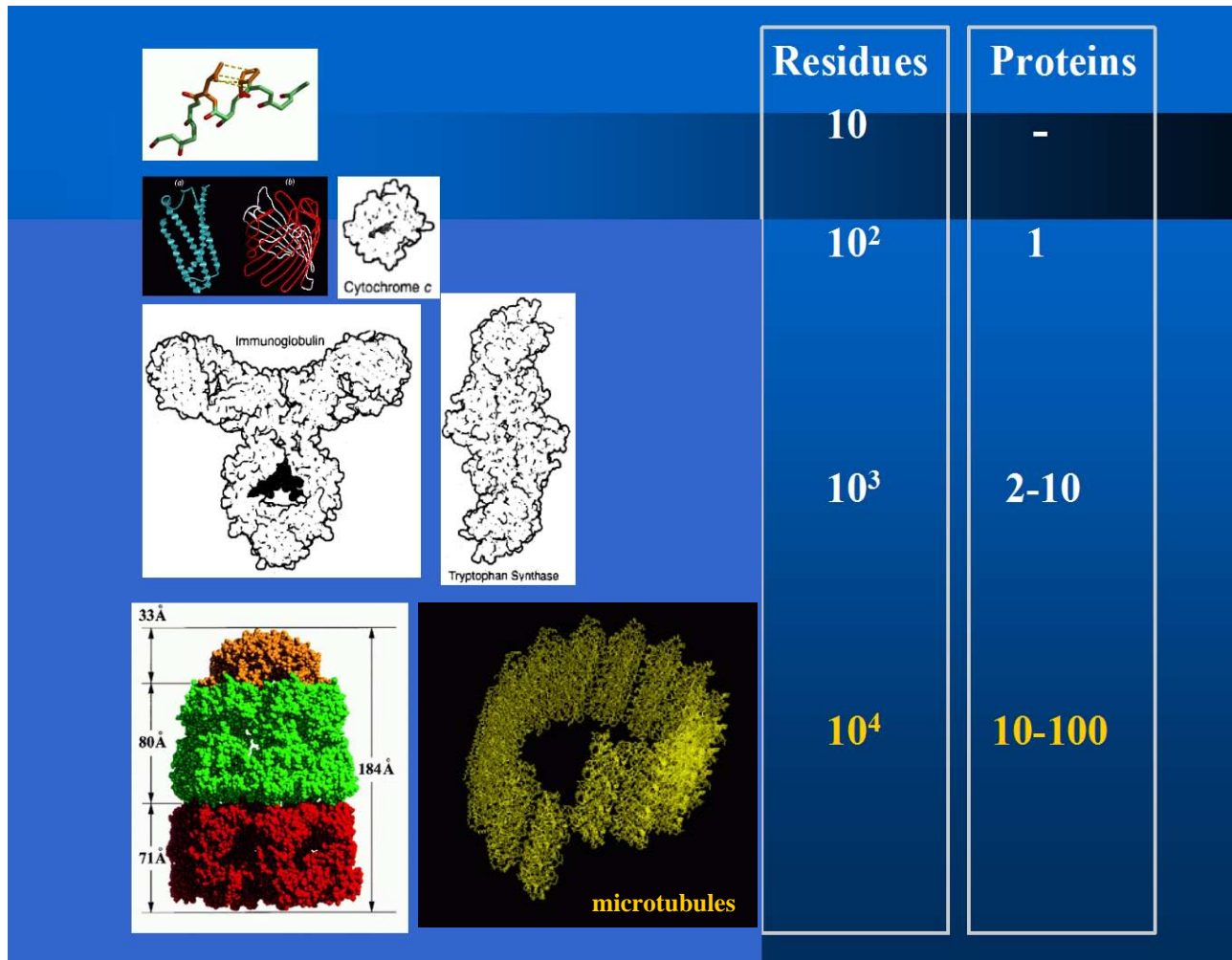
- Proteins are designed to favor functional changes in their structure. Pre-existing soft modes facilitate substrate binding.
- Collective mechanics/allosteric dynamics are mediated by conserved residues
- The intrinsic motions confer enhanced flexibility at substrate **recognition** sites
- Correlated 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



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



# ProDy

Protein Dynamics Analysis in Python



Wenzhi Mao



Dr. Indira Shrivastava  
Assist Prof, DCSB, Pitt



Markus Dittrich, PhD  
NRBSC Group Leader  
Pitt Supercomputing Center



Drs. Ahmet Bakan and Anindita Dutta



Dr. Chakra Chennubhotla  
Assist Prof, DCSB, Pitt



Dr. Timothy R Lezon  
Assistant Prof, DCSB, Pitt

Acknowledgment: NIH - 5 P41 GM10371202