

Day 2 – Lecture 1

Bridging Structure and Function, Experiments and Computations

Pemra Doruker

Department of Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260

Summary

1. Theory

- a. Gaussian Network Model (GNM)
- b. Anisotropic Network Model (ANM)
- c. Resources/Servers/Databases (ProDy, DynOmics)

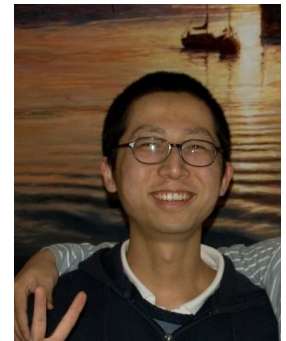
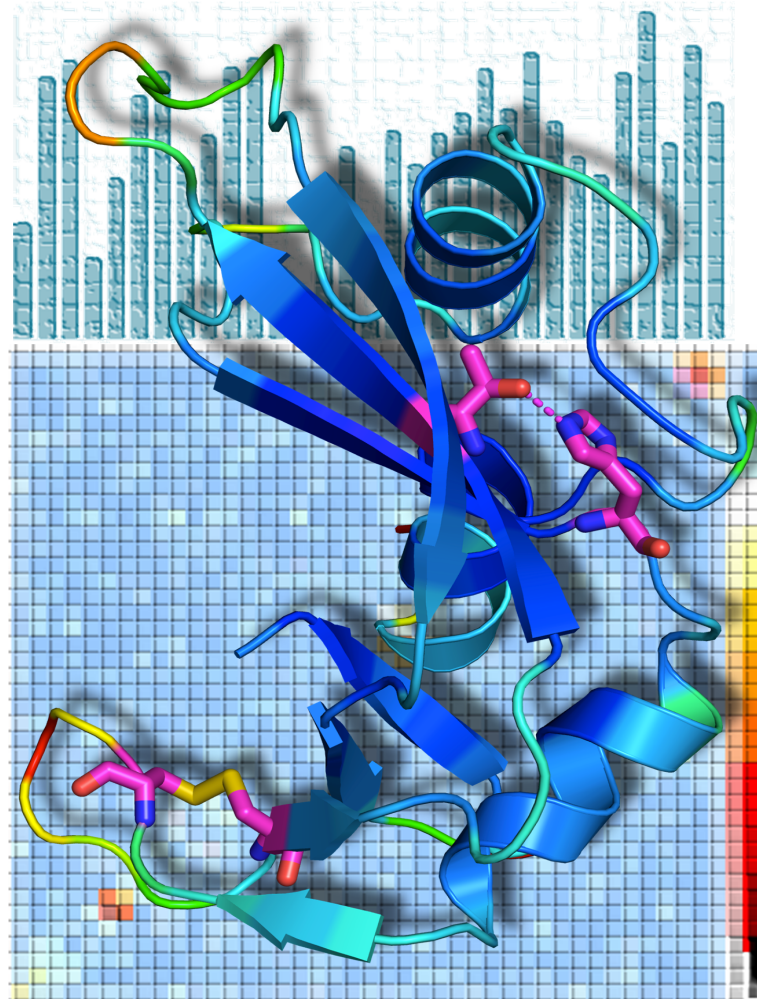
2. Bridging Sequence, Structure and Function

- a. Ensemble analysis and functional modes of motion
- b. Combining sequence and structure analyses – signature dynamics
- c. Modeling membrane proteins and lipid environment with ANM

3. Allostery and druggability

- a. Essential site scanning and allosteric pocket prediction
- b. Druggability simulations

Evol



Dr. Ying Liu

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?

Two recent reviews:

- Haliloglu, T, Bahar I (2015) Adaptability of protein structures to enable functional interactions and evolutionary implications *Curr Opin in Struct Biol* 35: 17-23
- Zhang Y, Doruker P, Kaynak B, Zhang S, Krieger JM, Li H, Bahar I. (2020) Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr Opin Struct Biol* 62:14-21.

Sequence evolution

an information-theoretic approach

F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F
F	L	K	I	I	Q	L	L	D	D	Y	P	K	C	F
F	L	K	I	I	Q	L	L	N	D	Y	P	K	C	F
F	I	K	V	V	E	L	F	D	E	F	P	K	C	F
L	E	K	A	T	K	L	F	T	T	Y	D	K	M	I

Residue index (up to N)

	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)

K, R: + charge; E, D: - charge → salt bridge

Mutual Information

without the influence of phylogeny

MI_p - to eliminate random noise and phylogenetic components

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

Average product correction

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

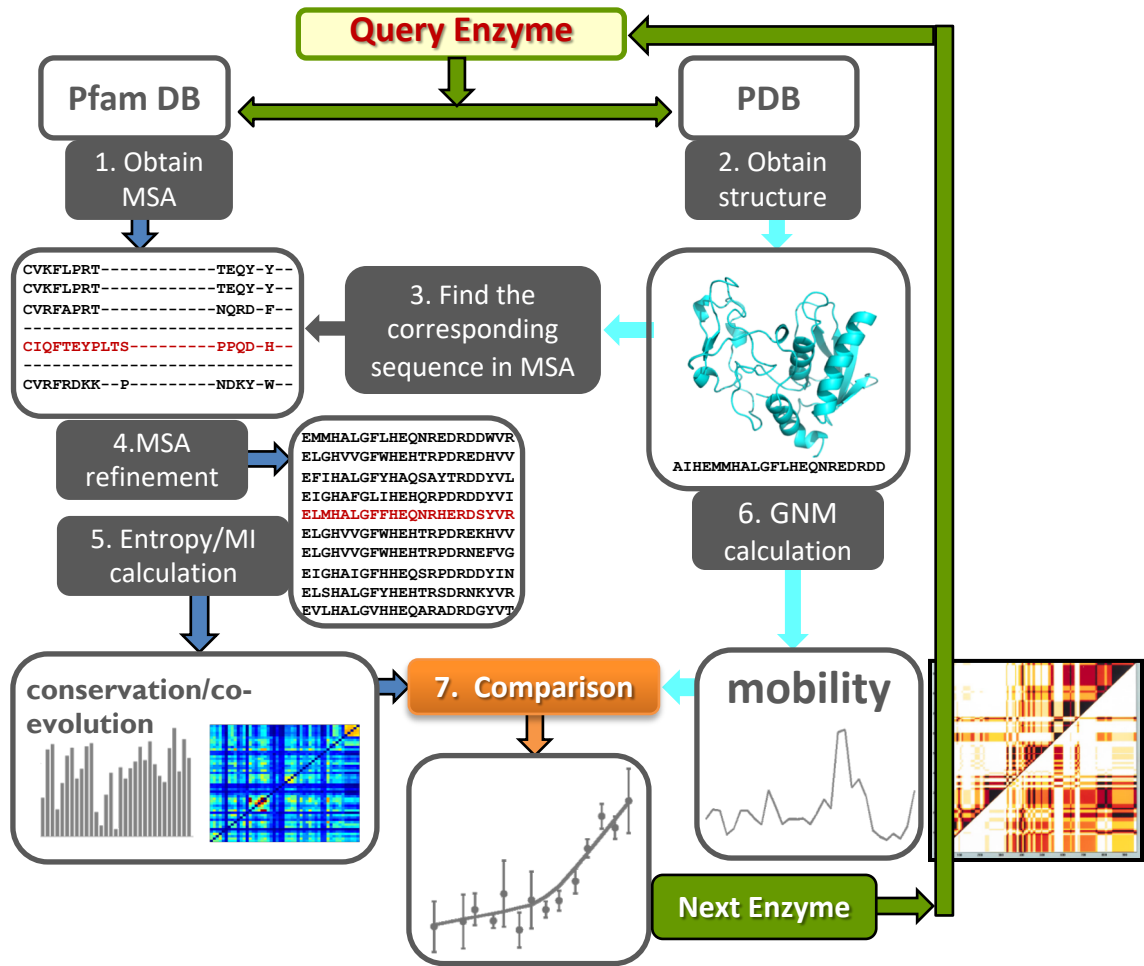
$\langle I(i) \rangle$: the mean mutual information of column i

$\langle I(i, j) \rangle$: average over all MI values

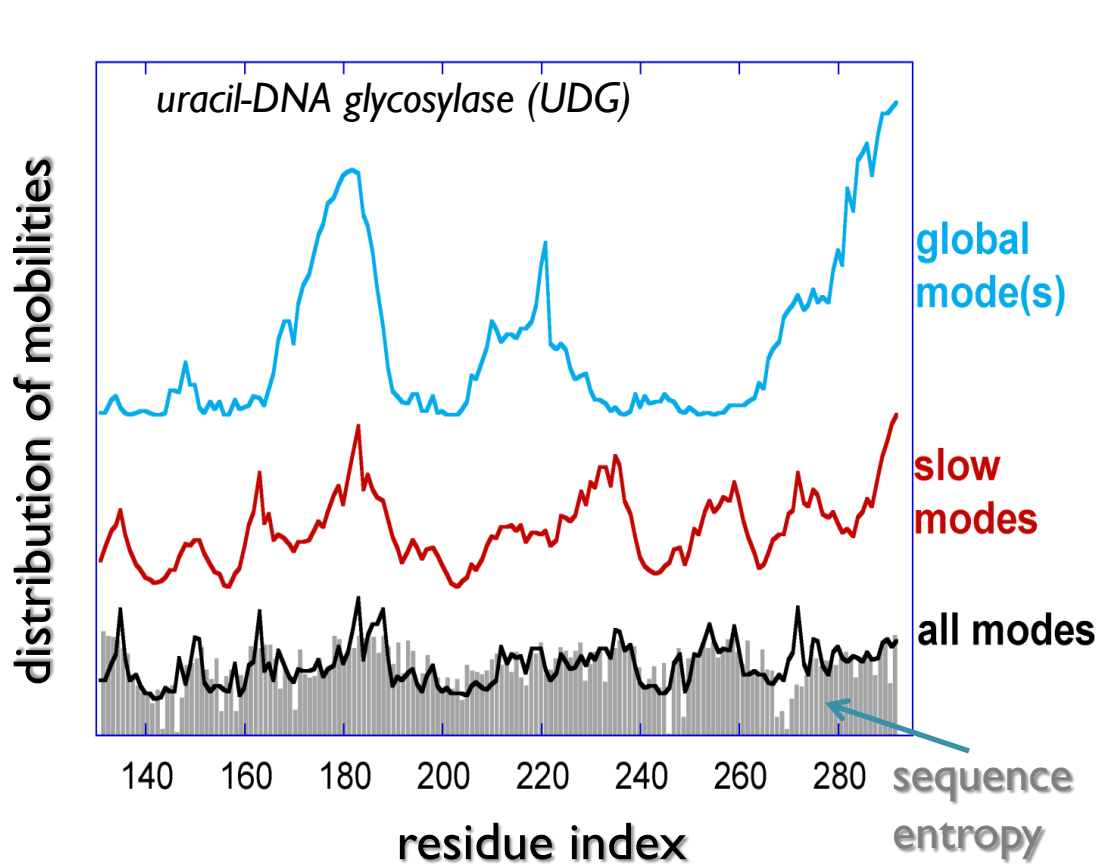
$$\langle I(i) \rangle = \sum_{j=1, j \neq i}^N I(i, j) / N$$

	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

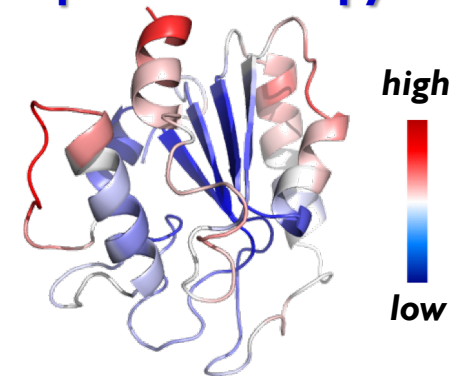
A systematic study of a set of enzymes



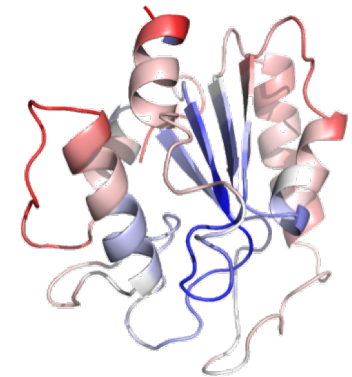
Correlation between sequence entropy & conformational mobility



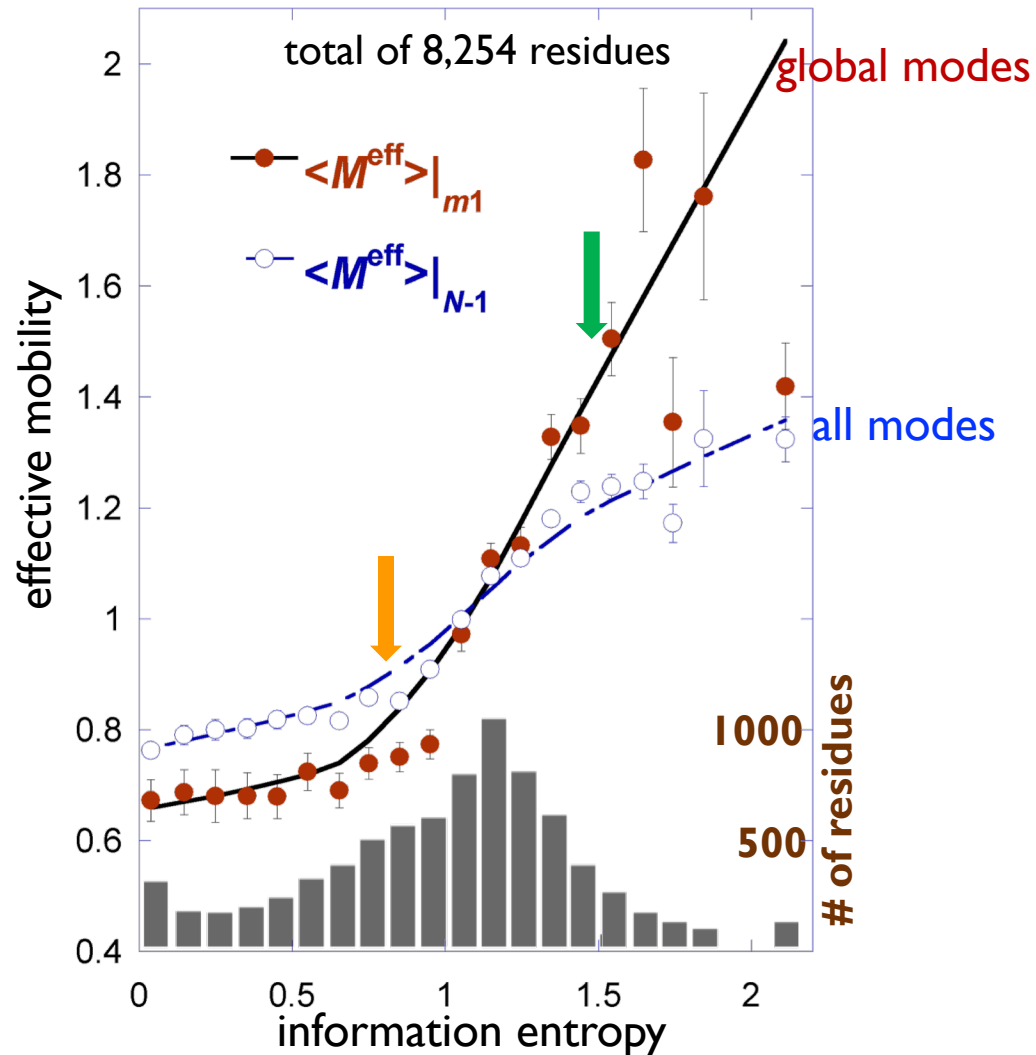
sequence entropy



structural dynamics

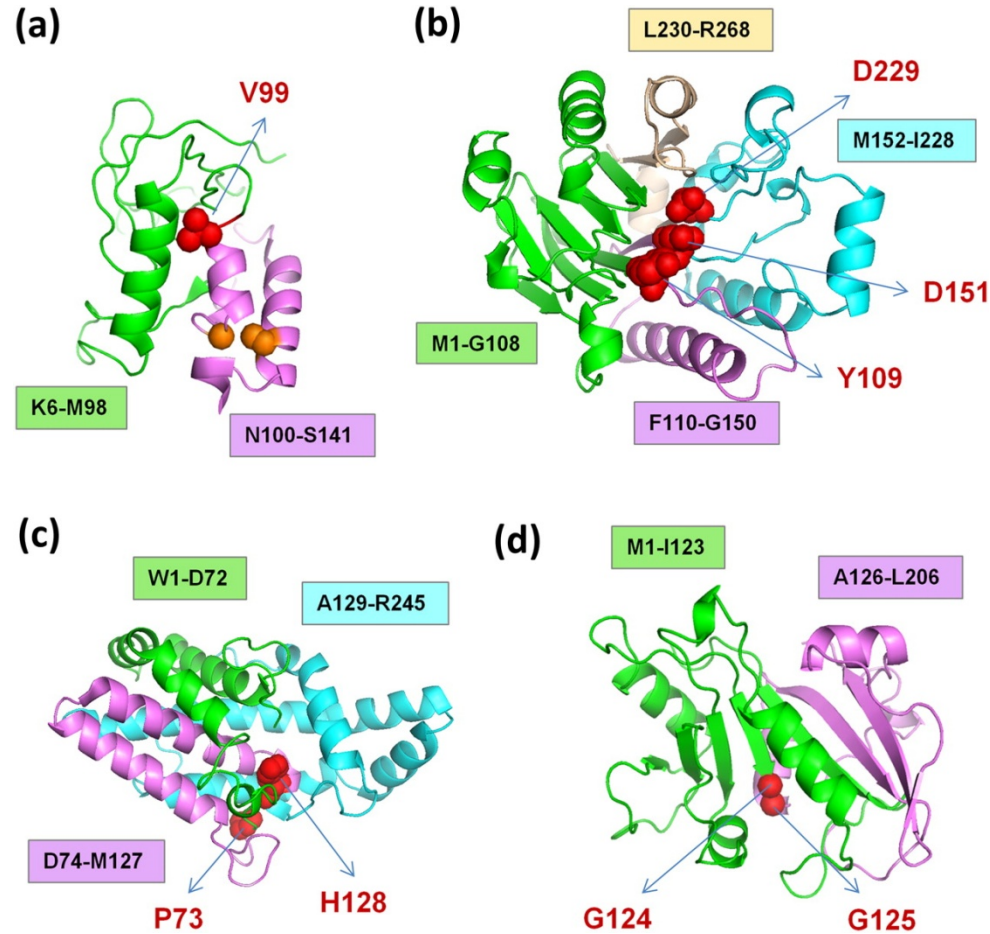


Mobility increases with sequence entropy

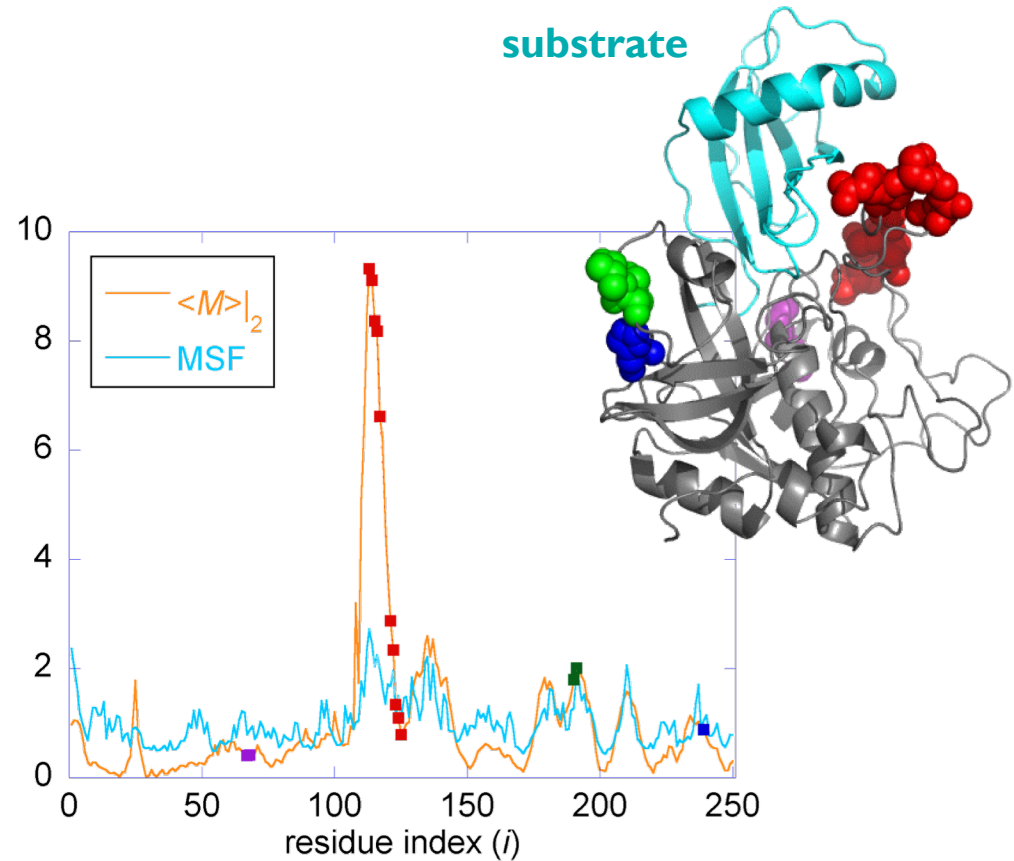
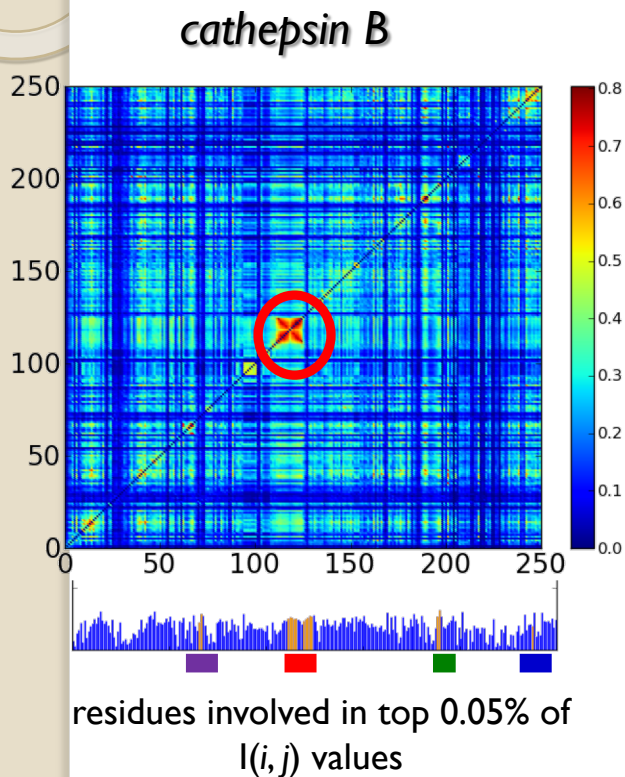


Hinge sites are evolutionarily conserved

despite their moderate-to-high exposure to environment



Amino acids involved in intermolecular recognition exhibit **high global mobility and co-evolution**



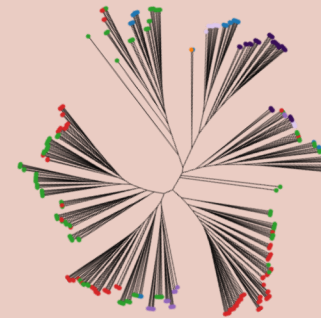
Summary

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

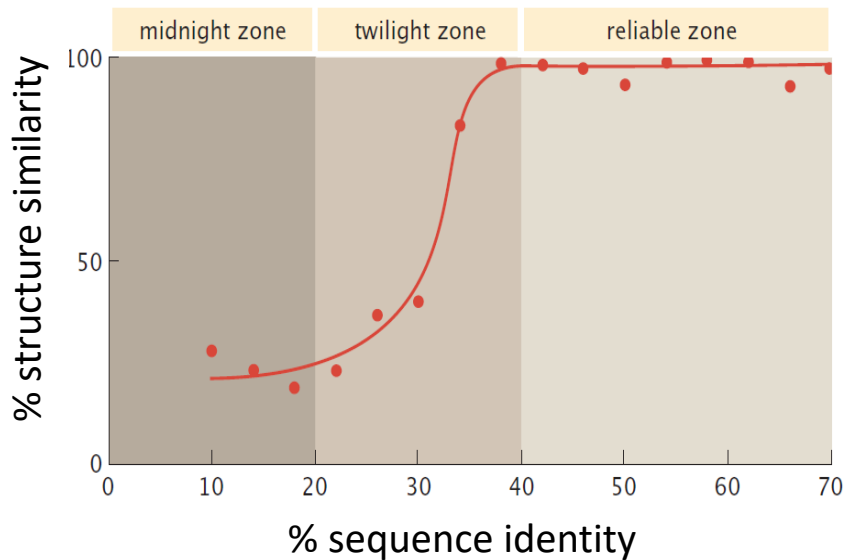
SignDy: Signature dynamics of families

- How does functional differentiation take place while maintaining the fold?
- What are the shared/differentiated dynamics of family members?
- Can we categorize family members based on dynamics?

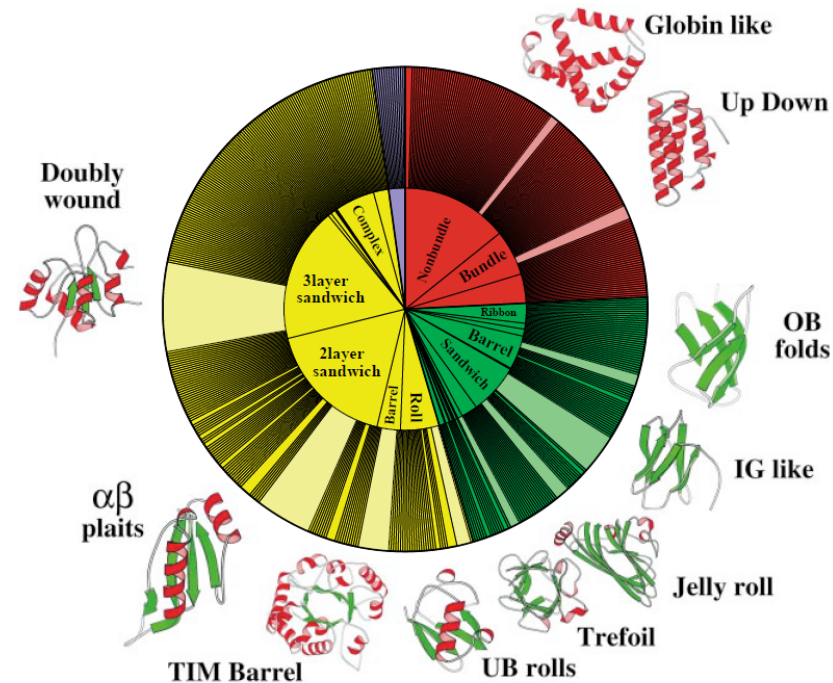


Protein Family Dynamics

Sequence evolves faster than structure.
 Proteins with different sequences may share similar structures (functions).

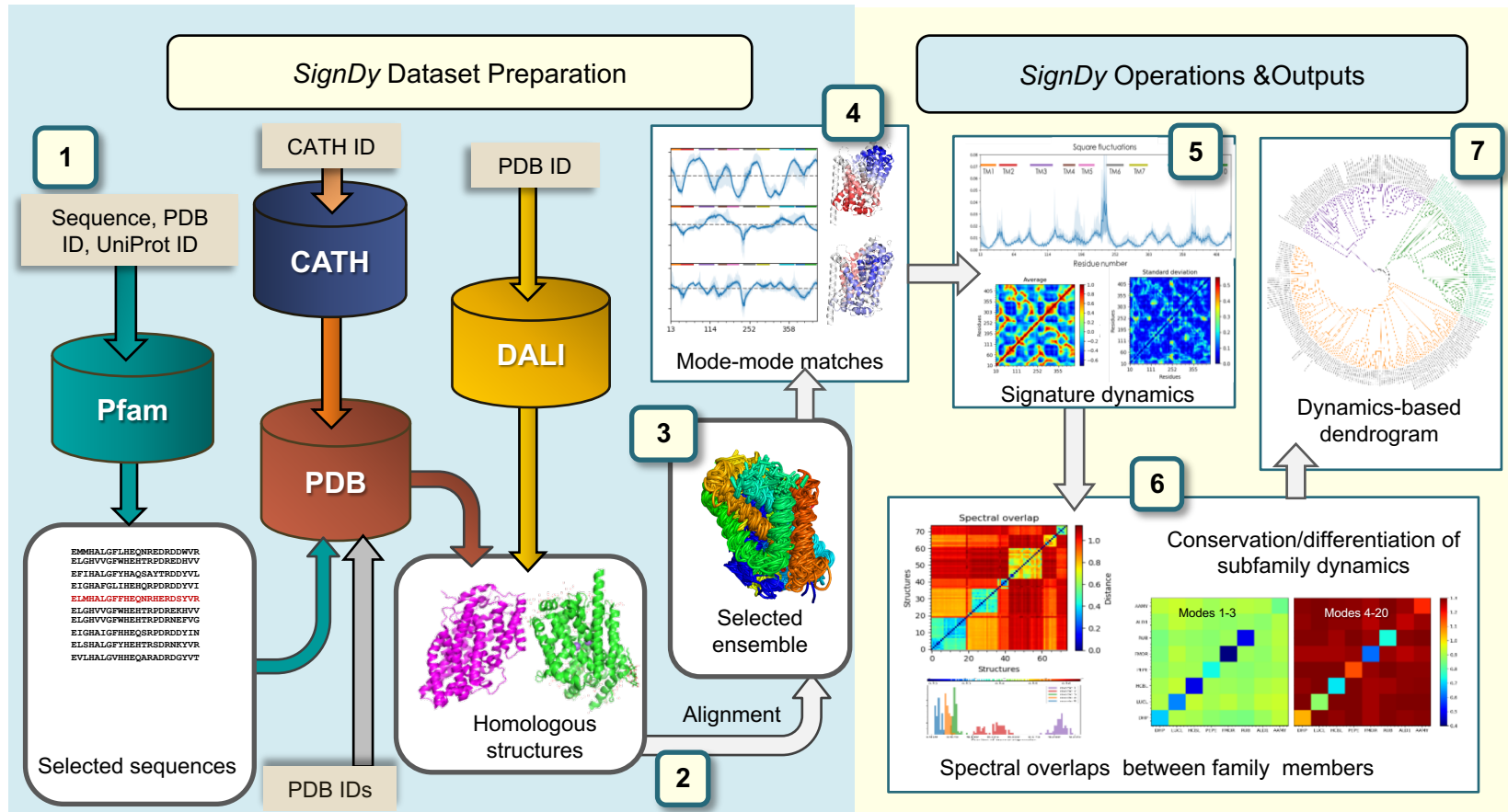


What is special about these folds that lend themselves to different functionalities?

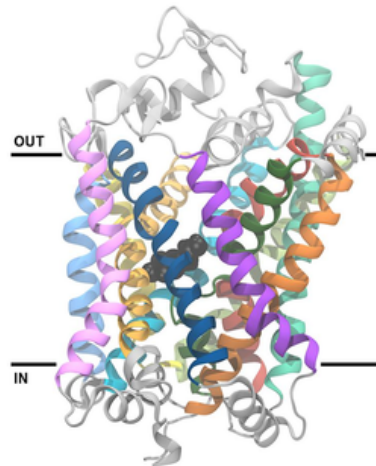
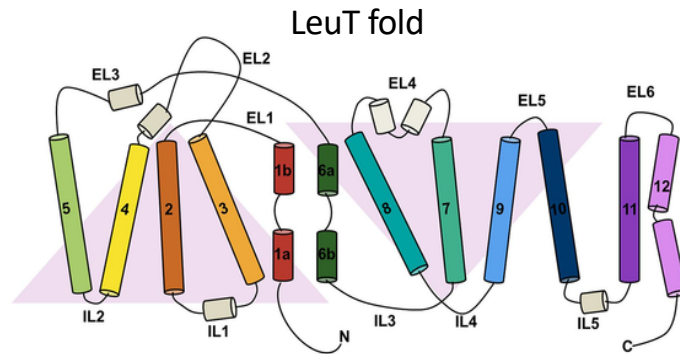


Bahar, Ivet, Robert L. Jernigan, and Ken A. Dill. *Protein Actions: Principles and modeling*. Garland Sci, 2017.
 Orengo, C.A., Pearl, F.M.G., et al., 1999. *Nucleic acids research*, 27(1), pp.275-279.

SignDy pipeline for evolution of dynamics

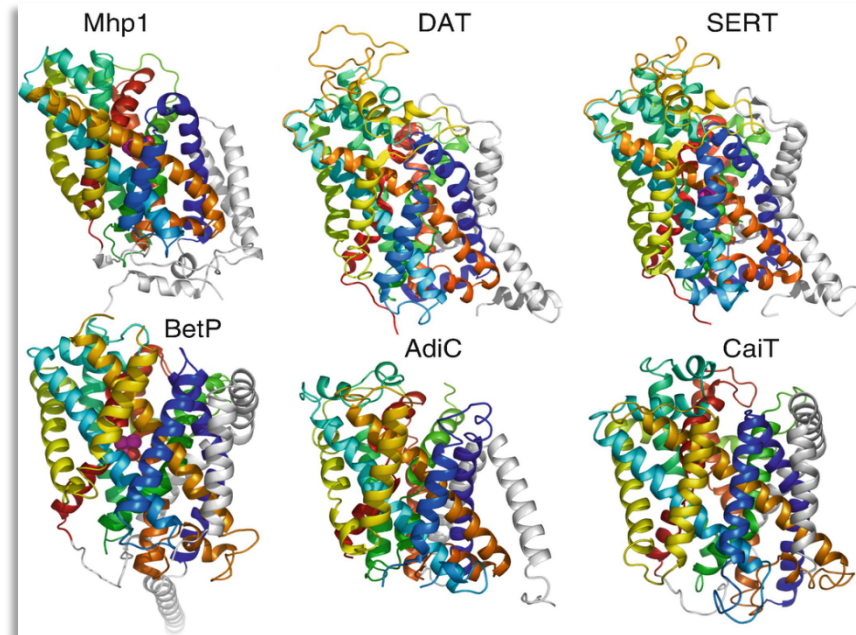


Case Study: LeuT Fold Superfamily



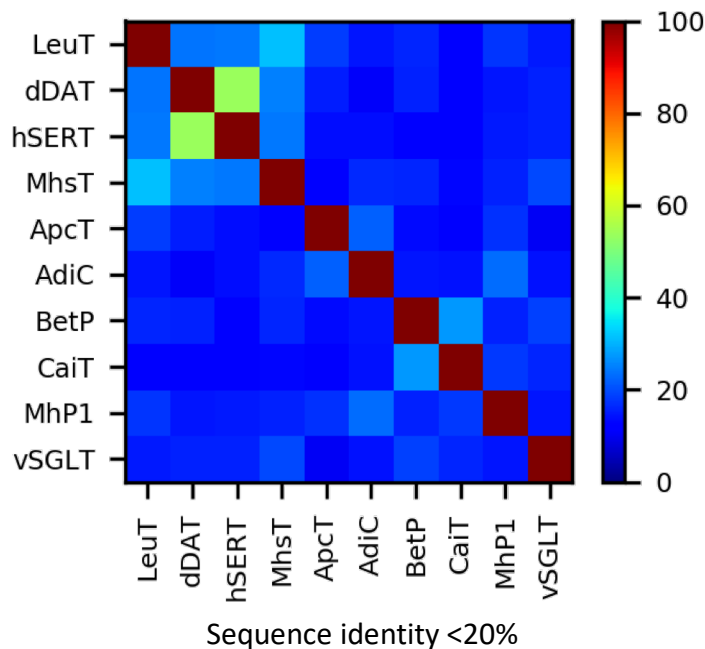
Bacterial Leucine Transporter
(LeuT)

Neurotransmitter transporters

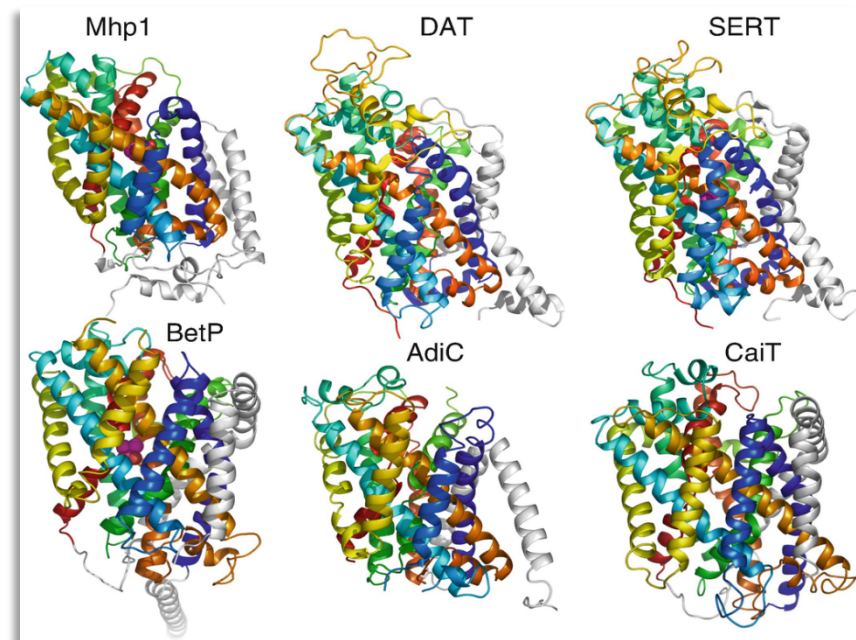


Case Study: LeuT Fold Superfamily

Minimal sequence identity



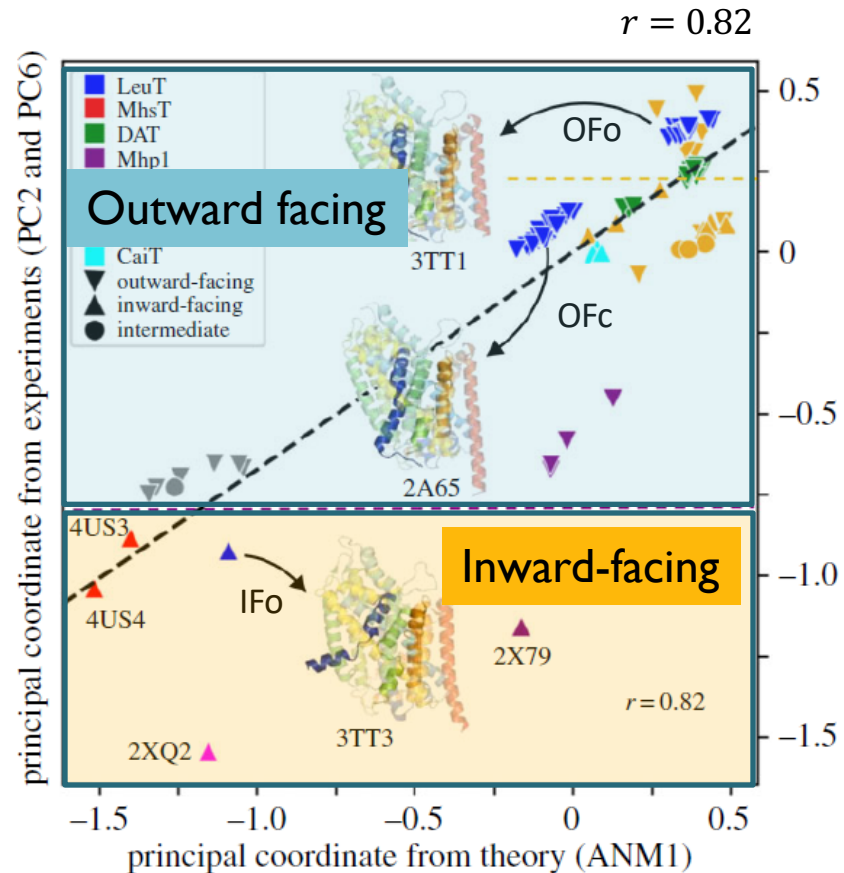
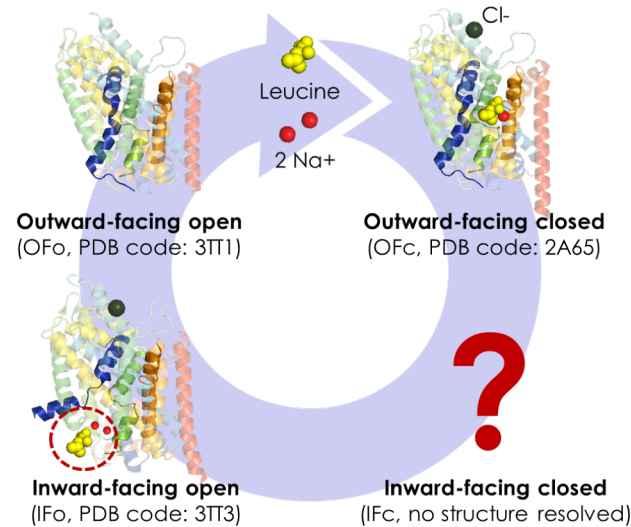
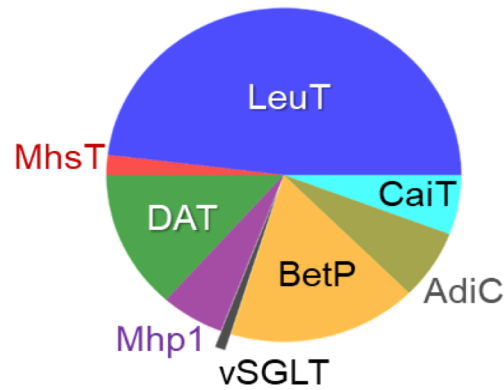
Same fold



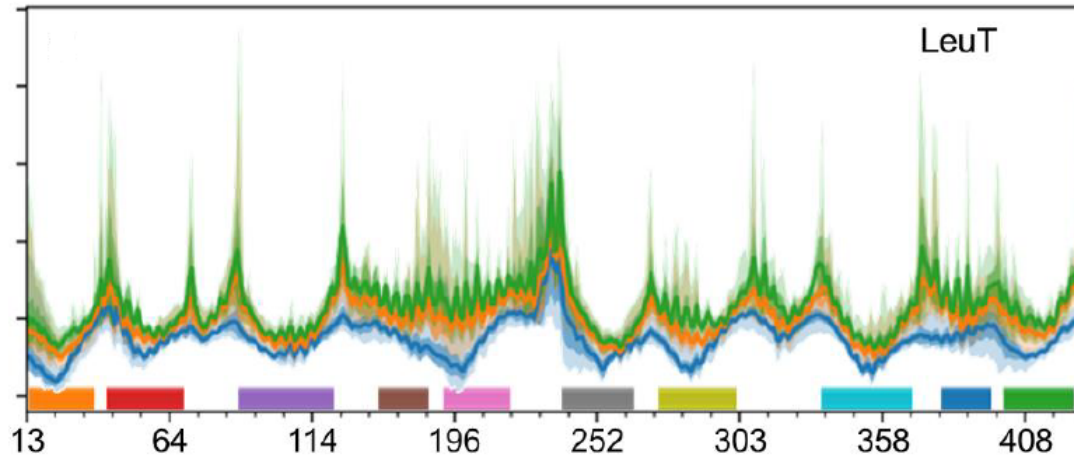
Zhang S, Li H, Krieger JM, Bahar I. (2019) Shared signature dynamics tempered by local fluctuations enables fold adaptability and specificity. *Mol Biol Evol* 36: 2053-2068.

Intrinsic Dynamics predicted by the ANM

104 conformers used as input

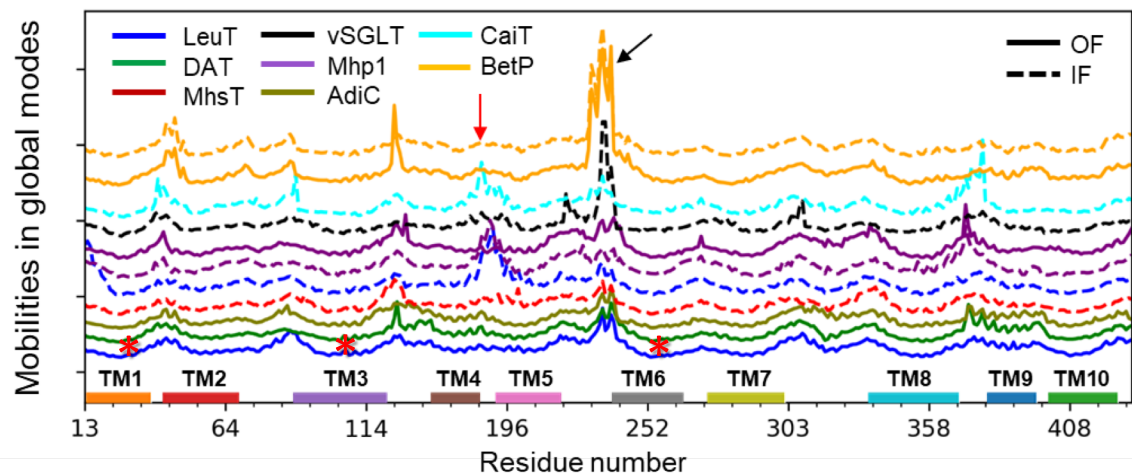
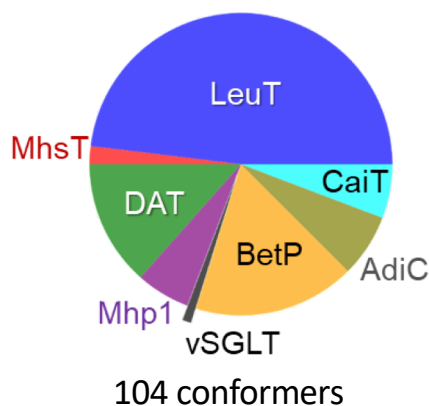


SignDy results for LeuT family

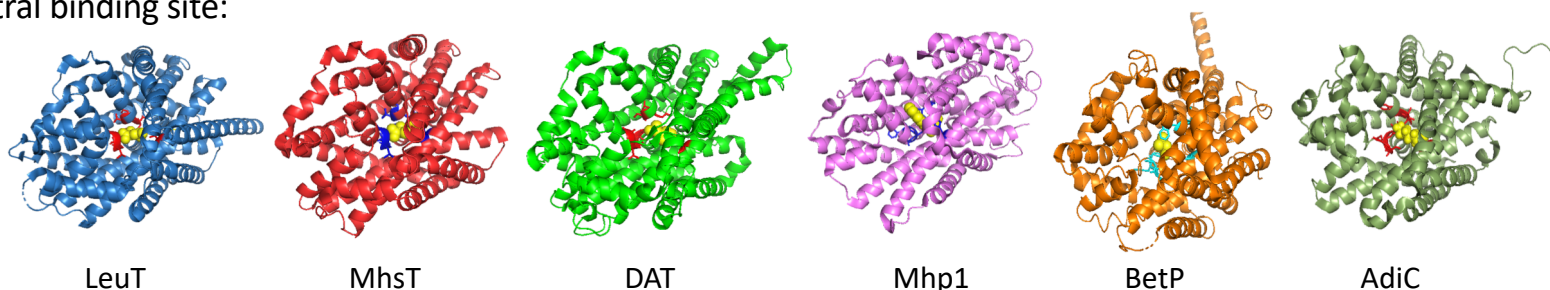


Signature-dynamics of each family is robustly defined by the global motions that are unique to the fold

Signature Dynamics: Shared fold-specific mobility profile

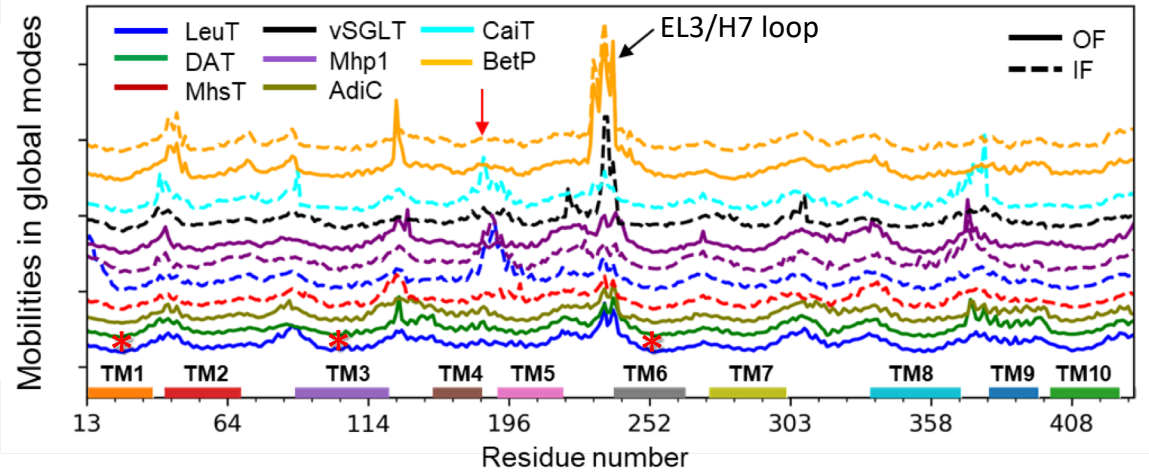
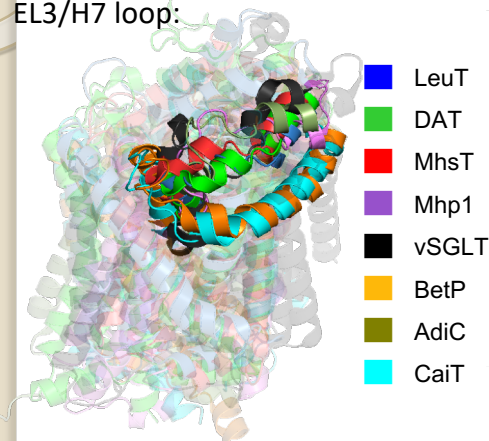


Central binding site:

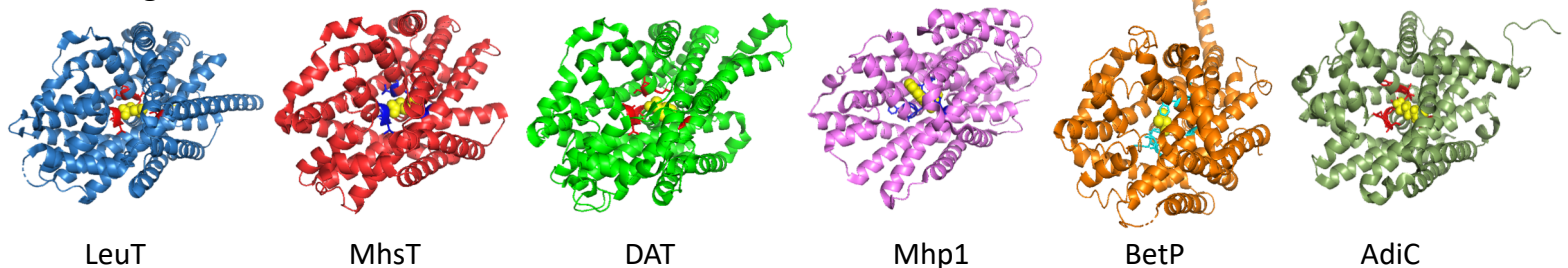


Signature Dynamics: Shared fold-specific mobility profile

EL3/H7 loop:

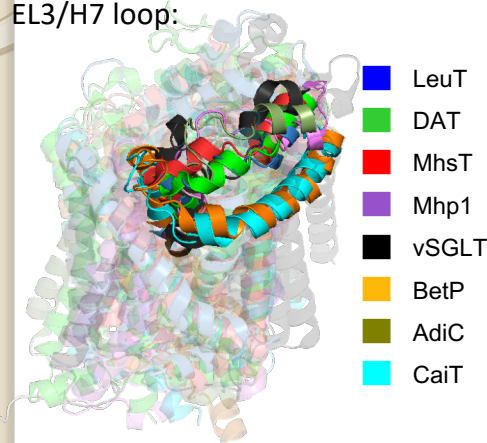


Central binding site:



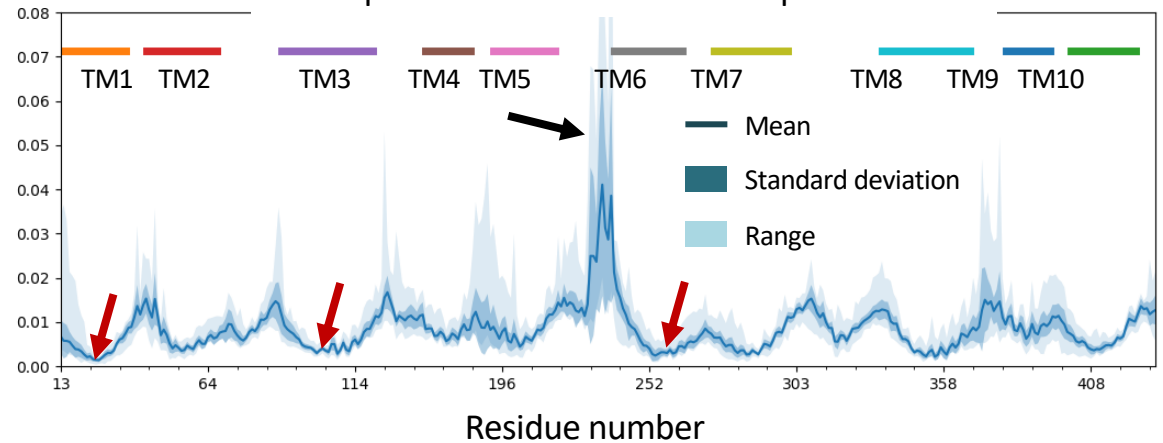
Signature Dynamics: Shared fold-specific mobility profile

EL3/H7 loop:

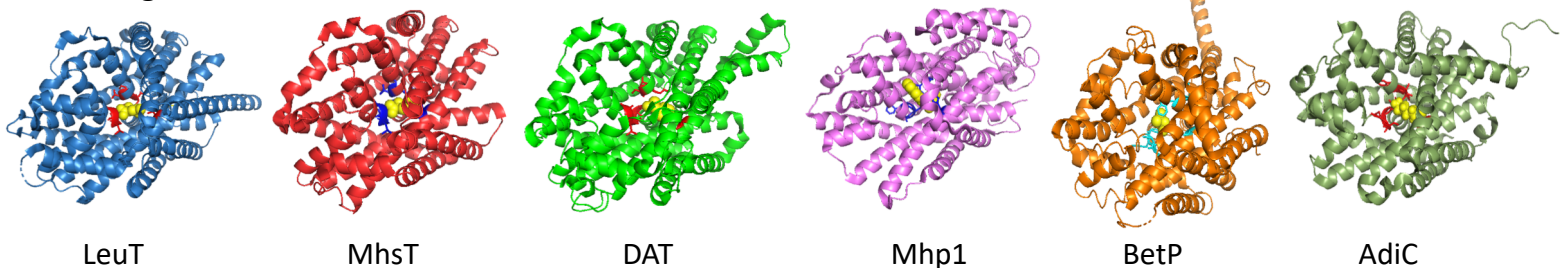


- LeuT
- DAT
- MhsT
- Mhp1
- vSGLT
- BetP
- AdiC
- CaiT

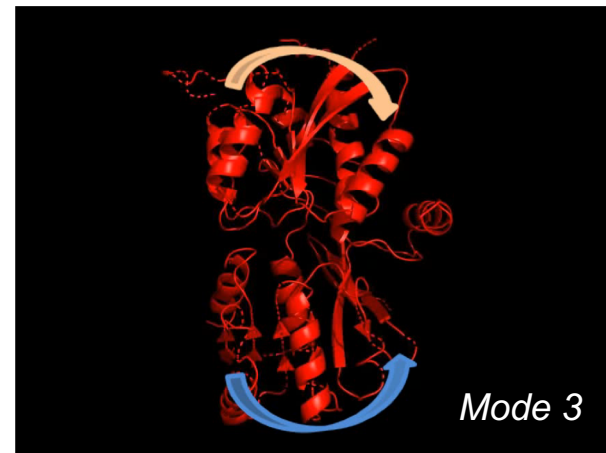
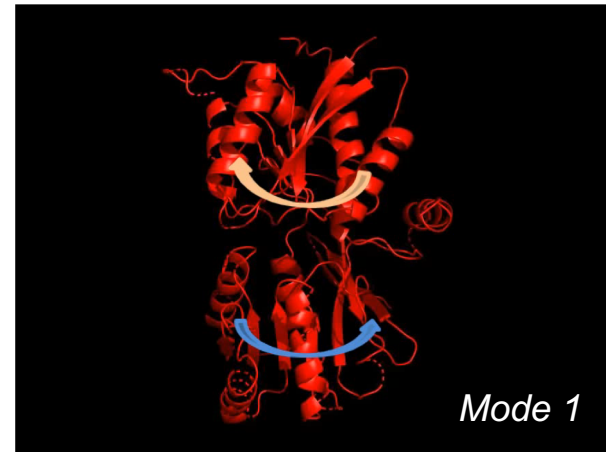
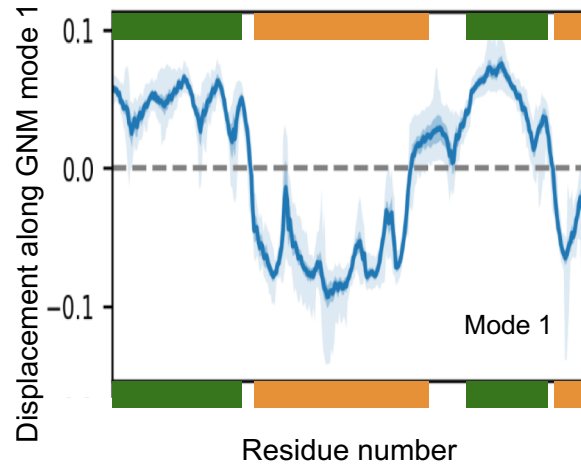
Square fluctuations in residue positions



Central binding site:

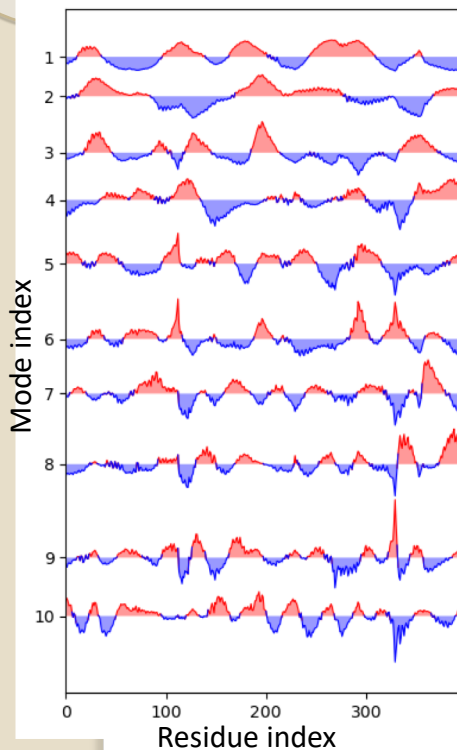


Signature modes match functions



How conserved are the global modes?

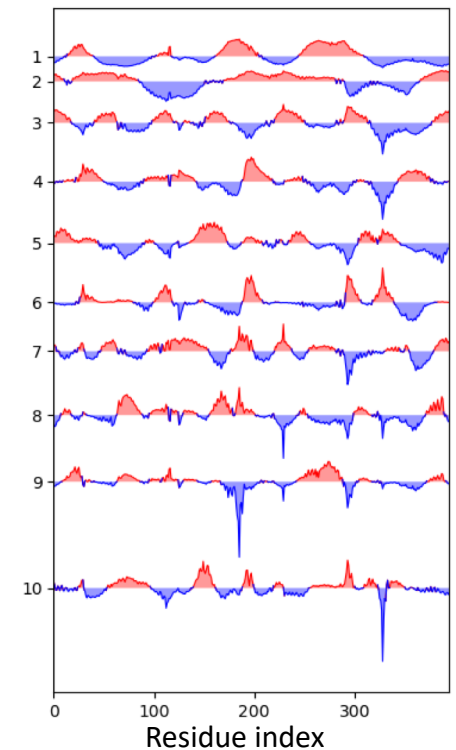
LeuT



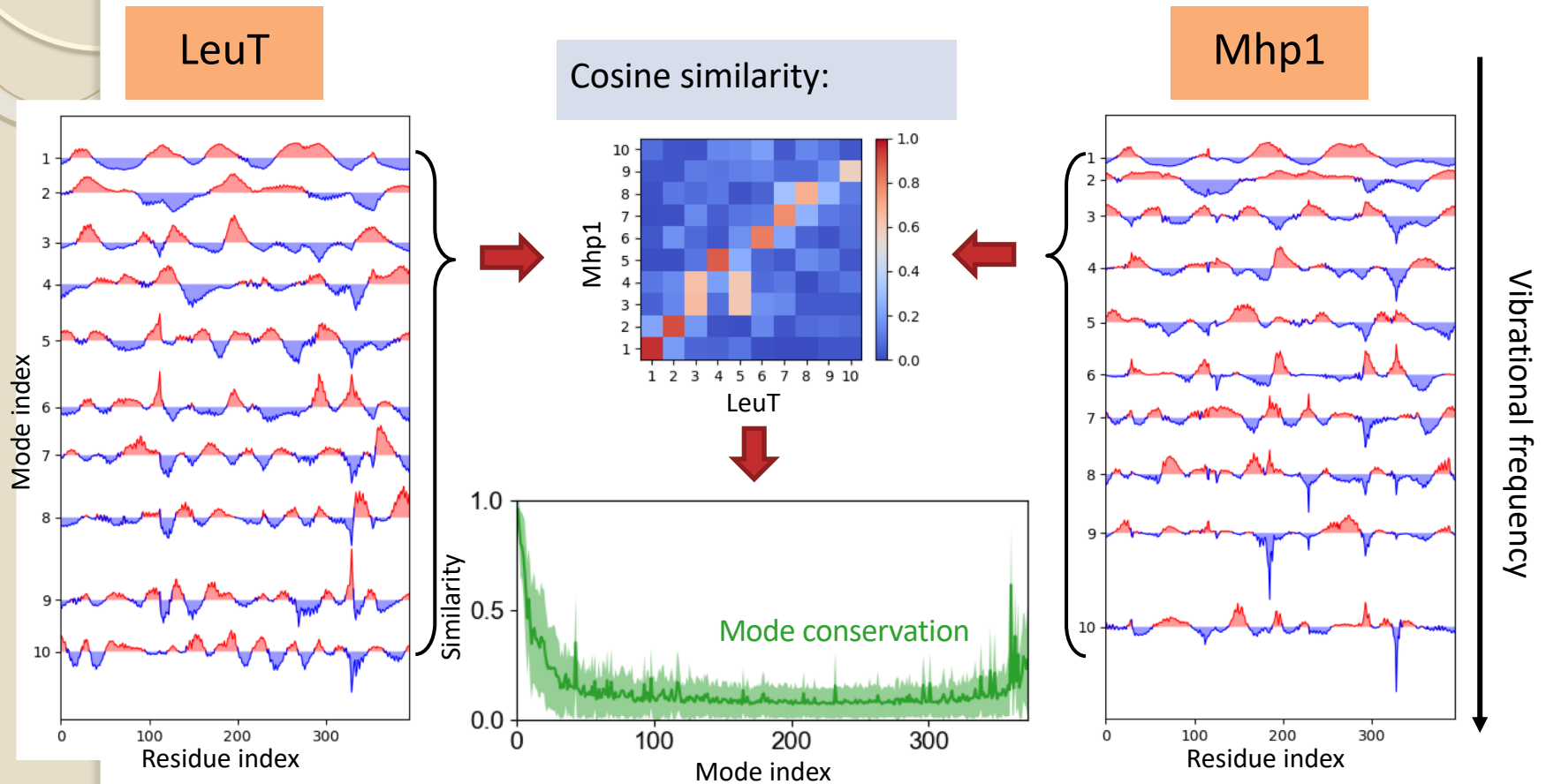
Cosine similarity:

$$cc = \mathbf{v}_i^{(LeuT)} \mathbf{v}_j^{(Mhp1)}$$

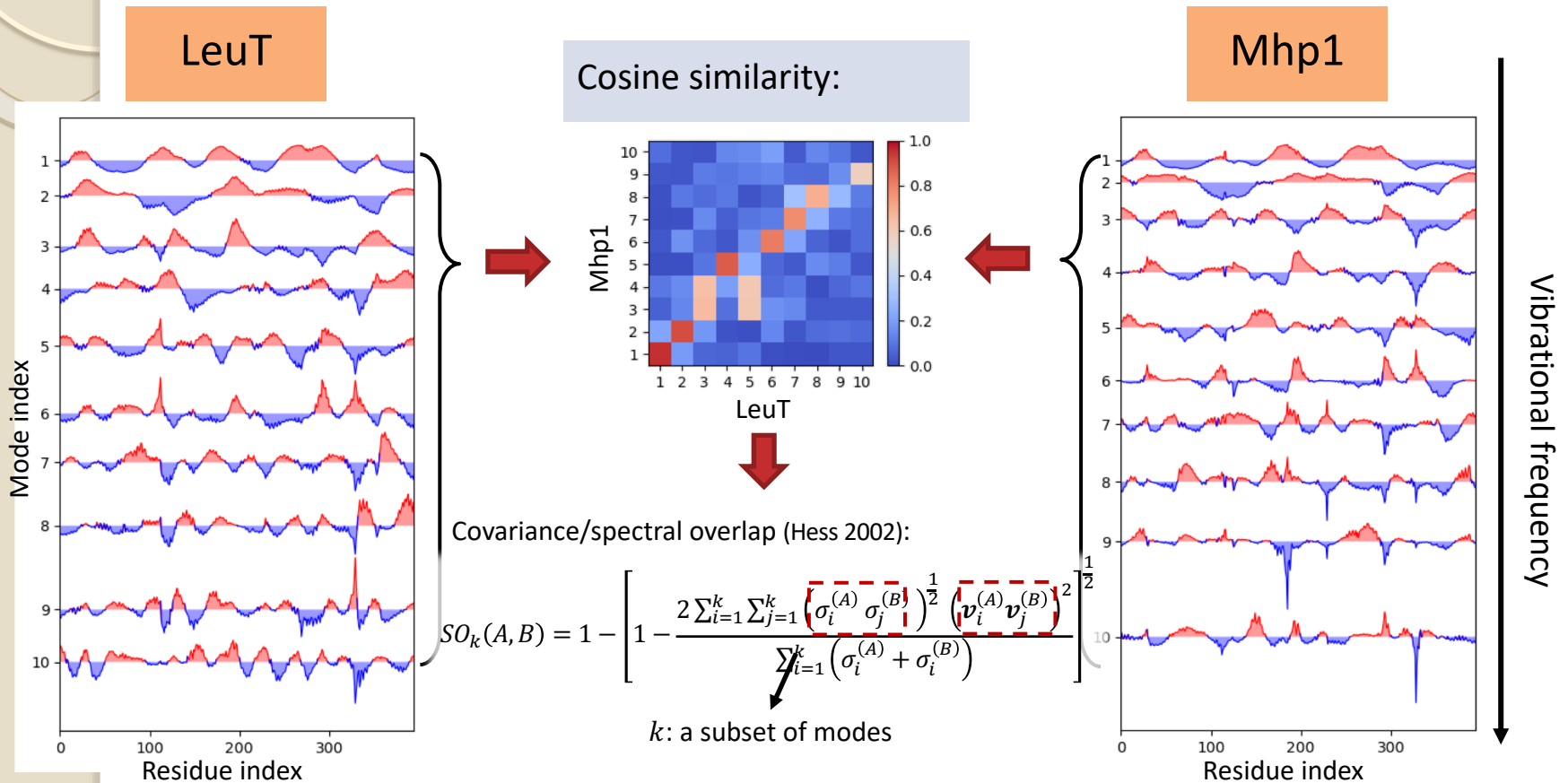
Mhp1



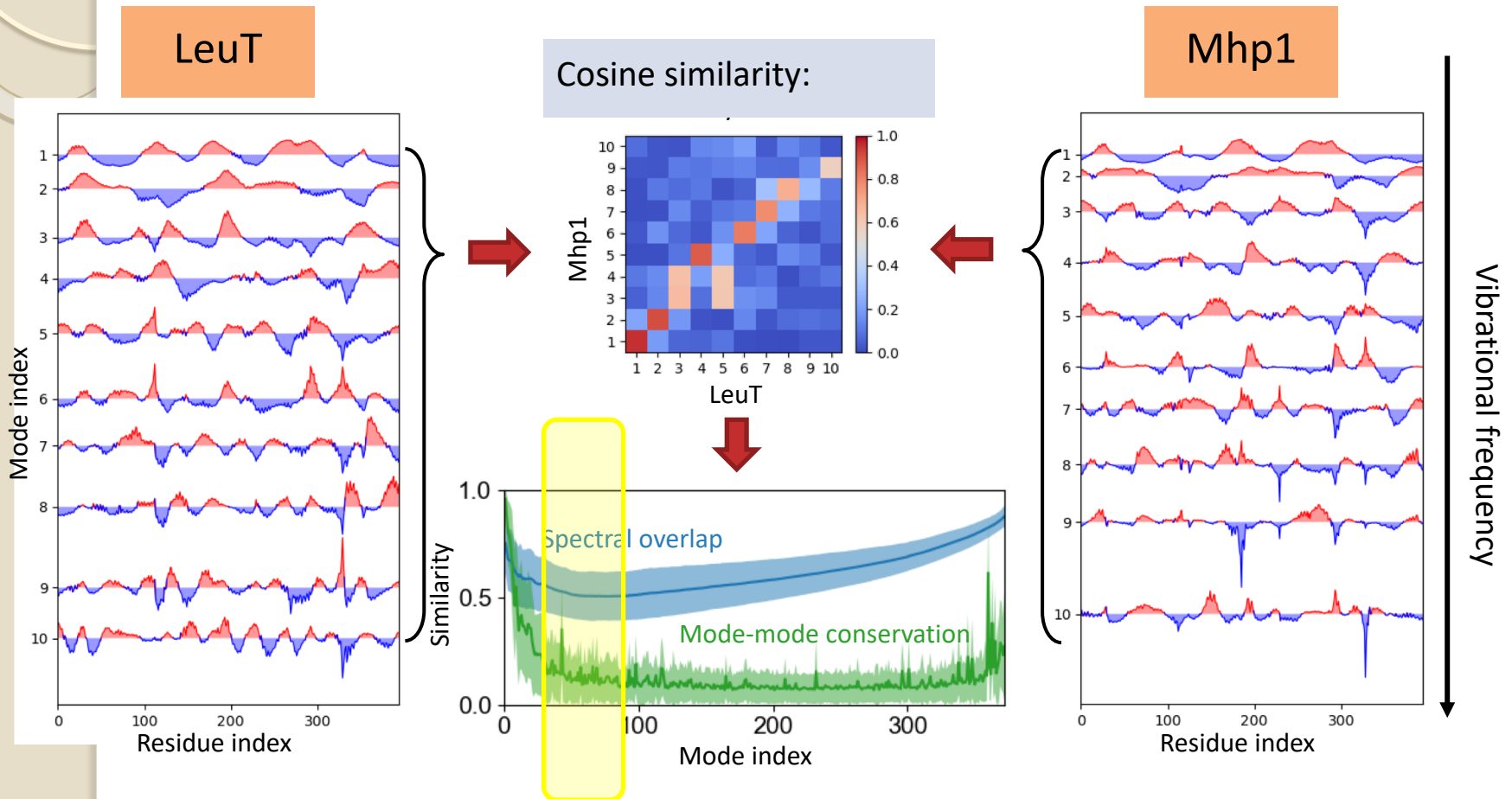
How conserved are the global modes?



How conserved are the global modes?



How conserved are the global modes?

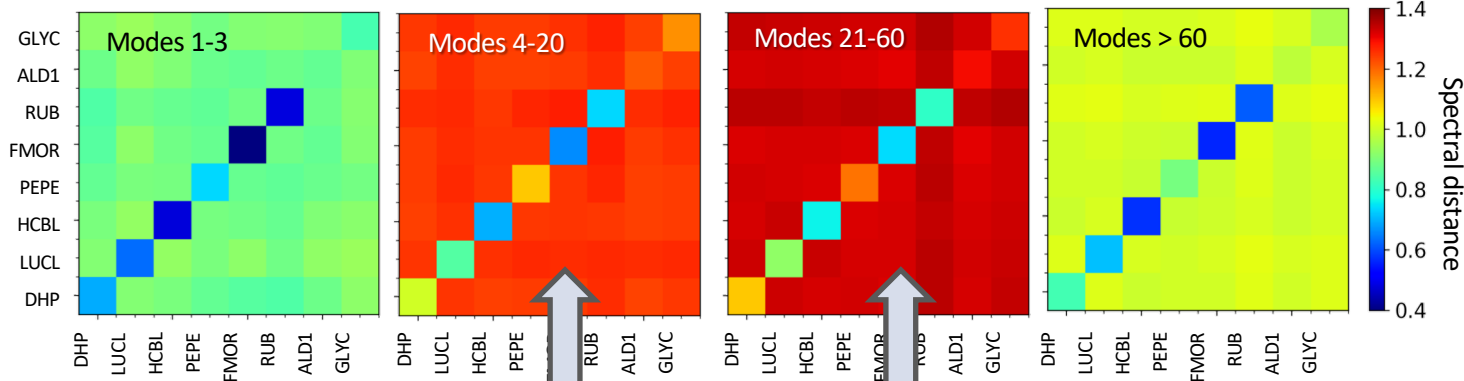
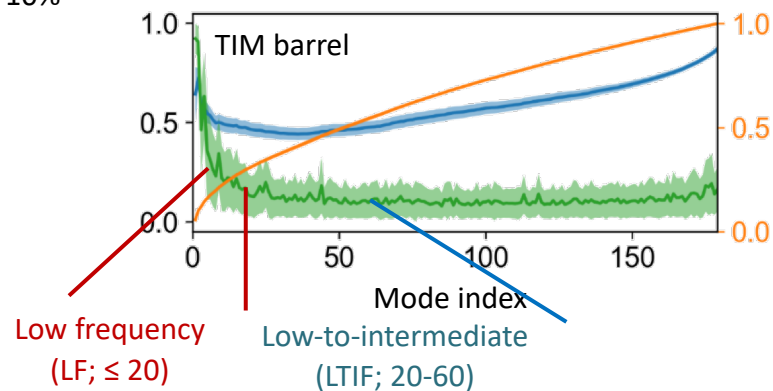
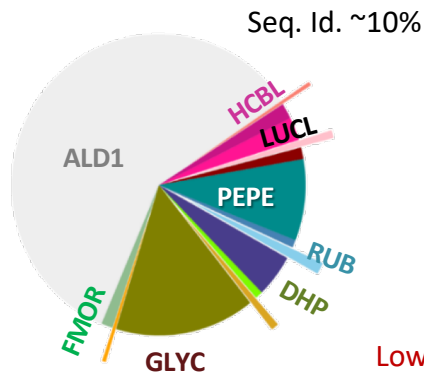
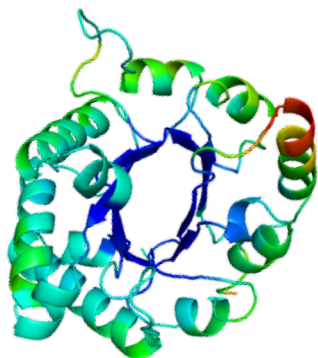


A region where overlap is minimal → these modes enable functional differentiation

Functional differentiation of subfamilies is enabled by LTIF modes

LTIF: low-to-intermediate frequency

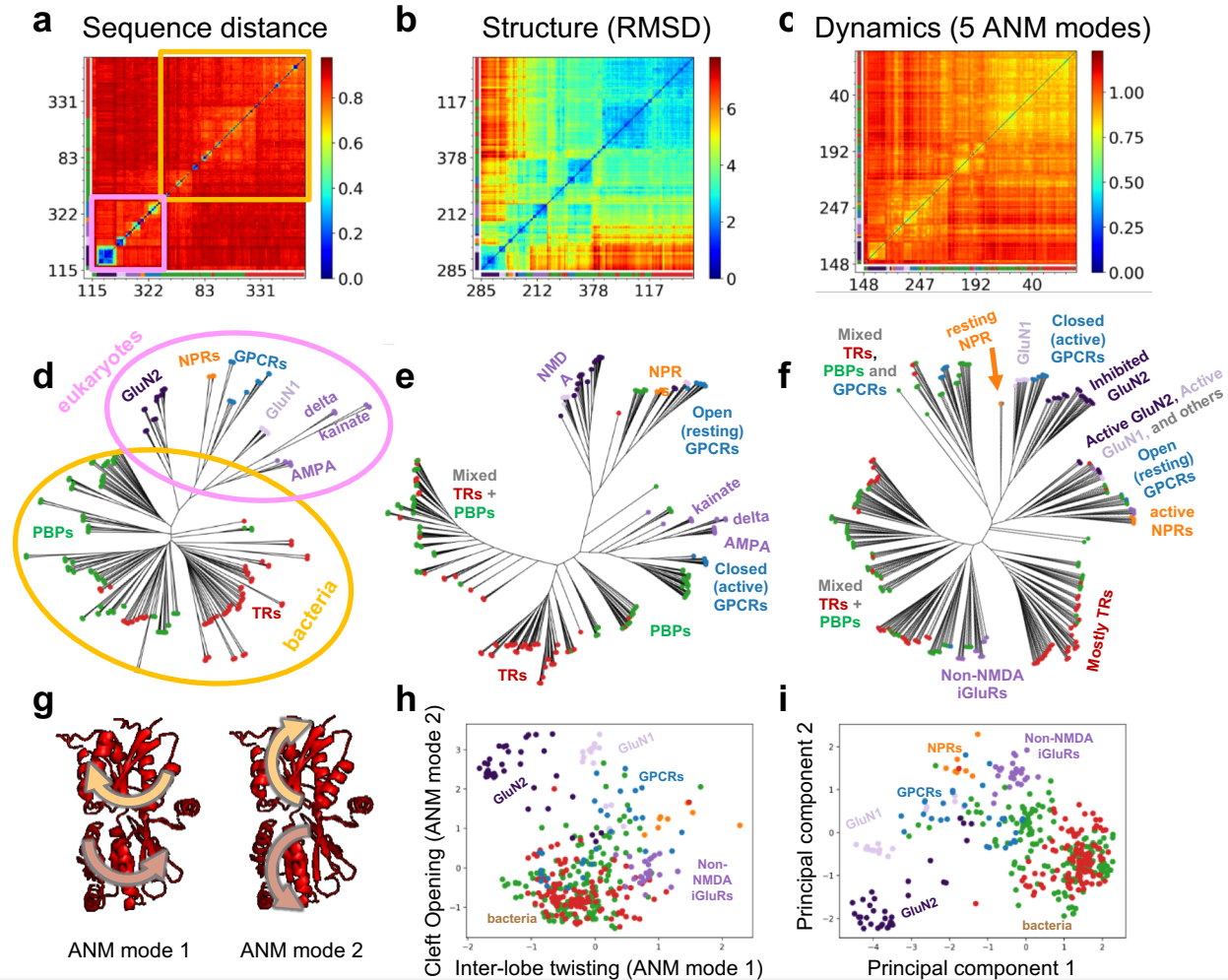
Application to TIM barrel proteins



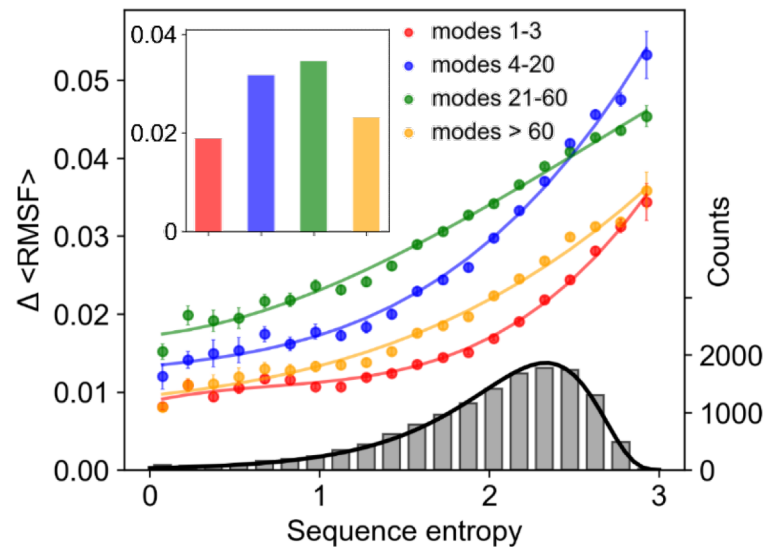
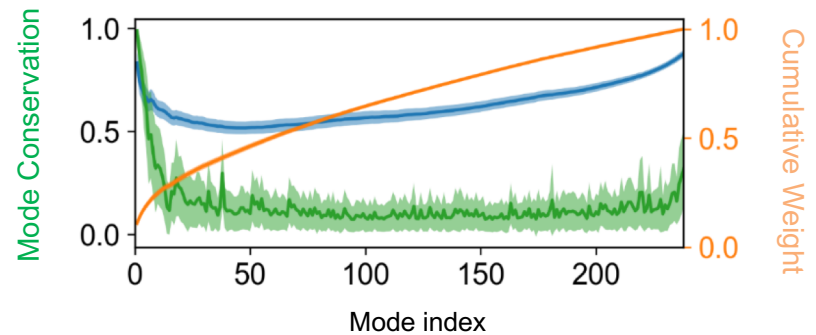
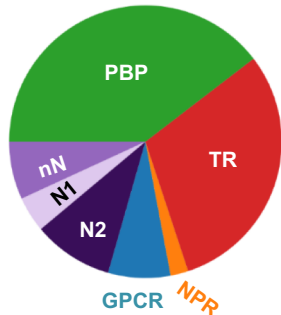
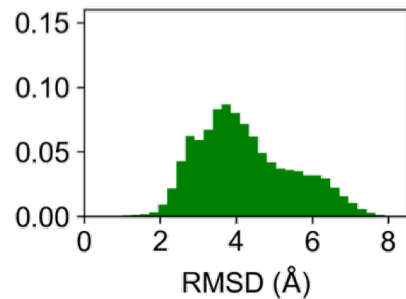
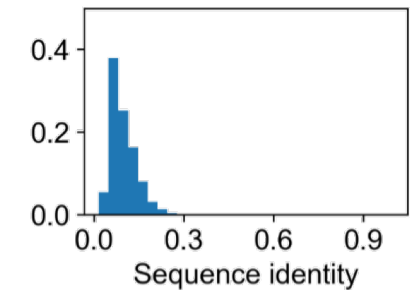
Largest spectral distance observed in modes 4-60.

Dynamics allows for classification

(complementing sequence- and structure- based classification)

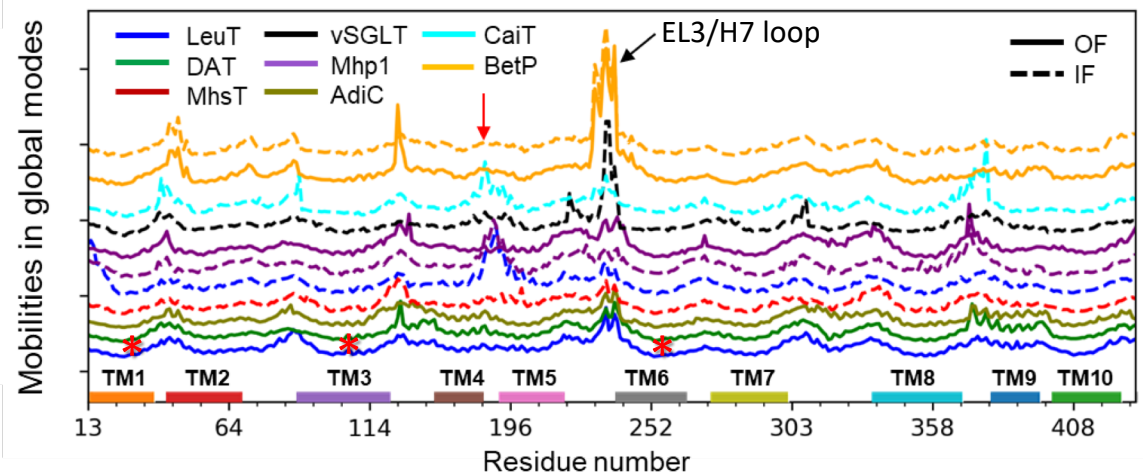
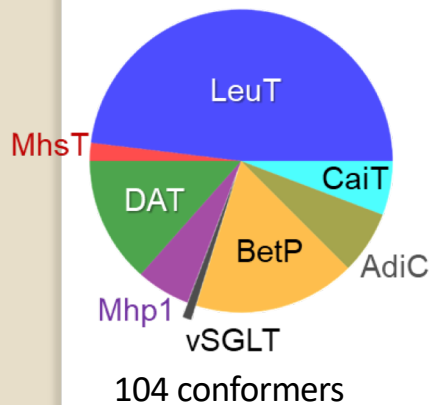


SignDy reveals shared and divergent motions of domains/folds



Summary: SignDy analysis

- Characterization of the shared structural and dynamic features, or **signature dynamics**
- Identification of sites important for the specific function of subfamilies,
- Insights into **functional differentiation** based on the differentiation of dynamics
- Learning design principles: focus on LTIF regime for altering function
- Data generated for the most common 77 CATH superfamilies (containing 15,000 proteins)
- Generation of dendrograms based on evolution of structural dynamics and thereby function



Summary

1. Theory

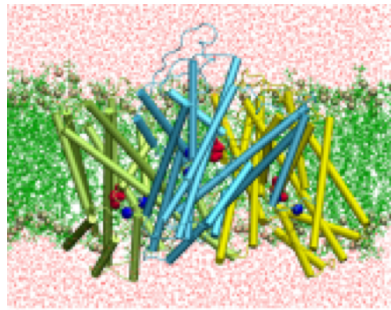
- a. Gaussian Network Model (GNM)
- b. Anisotropic Network Model (ANM)
- c. Resources/Servers/Databases (ProDy, DynOmics)

2. Bridging Sequence, Structure and Function

- a. Ensemble analysis and functional modes of motion
- b. Combining sequence and structure analyses – signature dynamics
- c. Modeling membrane proteins and lipid environment with ANM

3. Allostery and druggability

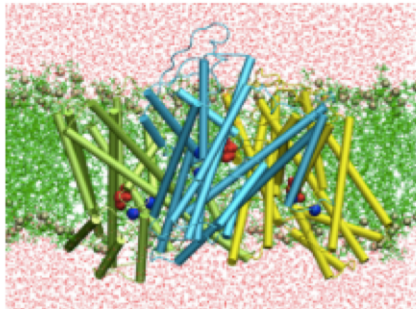
- a. Essential site scanning and allosteric pocket prediction
- b. Druggability simulations



membrANM

Membrane Anisotropic Network Model

ANM for membrane proteins: membrANM



Membrane ANM (membrANM)

Learn how to include the effect of lipid bilayer in ENM study of membrane proteins dynamics.

[Go to Tutorial](#)

- Evaluating membrane proteins' dynamics in the presence of lipid bilayer
- Comparing global motions in the presence and absence of membrane
- Understanding mechanisms of protein-membrane remodeling

Implemented in ProDy and DynOmics

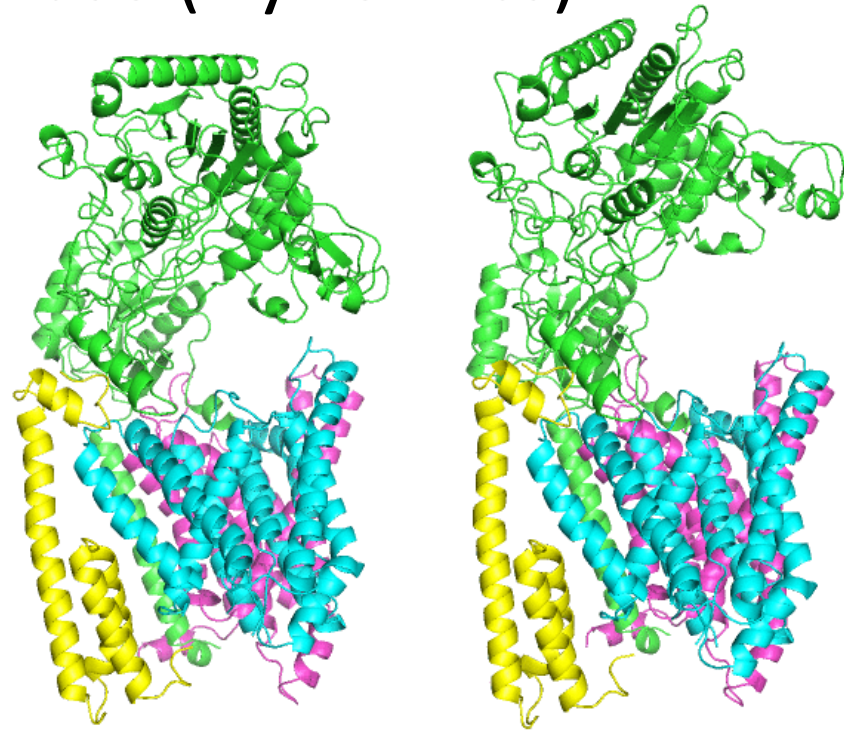
MembrANM for γ -secretase (DynOmics)

The effect of the membrane environment can be incorporated in the ANM analysis by adopting

- envANM-membrANM or
- substructure-membrANM.

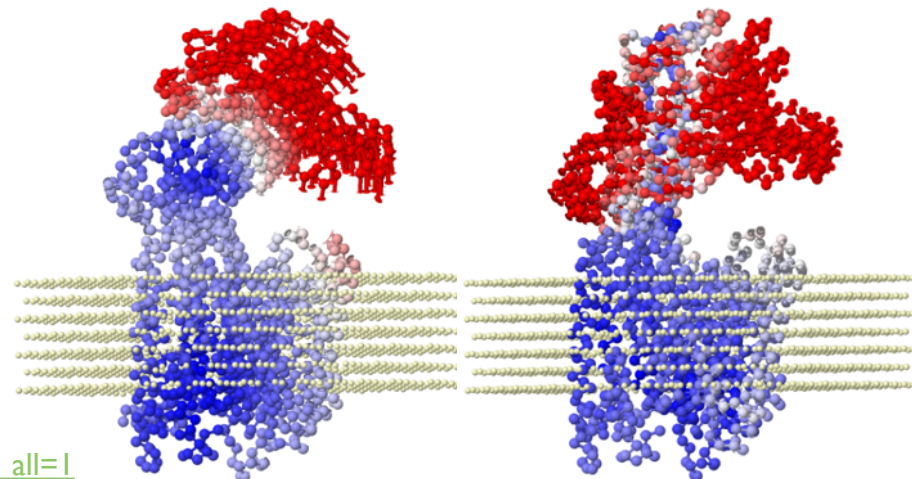
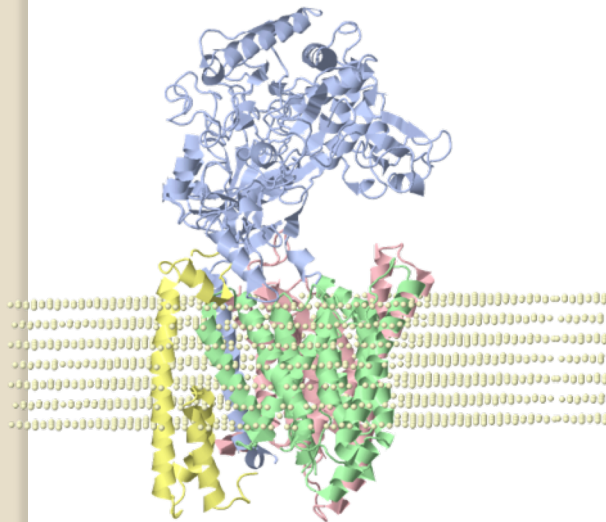
MembrANM results for the γ -secretase (PDB ID: 5FN2) can be viewed on the ENM server:

PyMol movies can be downloaded from DynOmics ENM server by generating "Full Atomic Structures for ANM-Driven Conformers".



Mode 1 - bending

Mode 2 - twisting



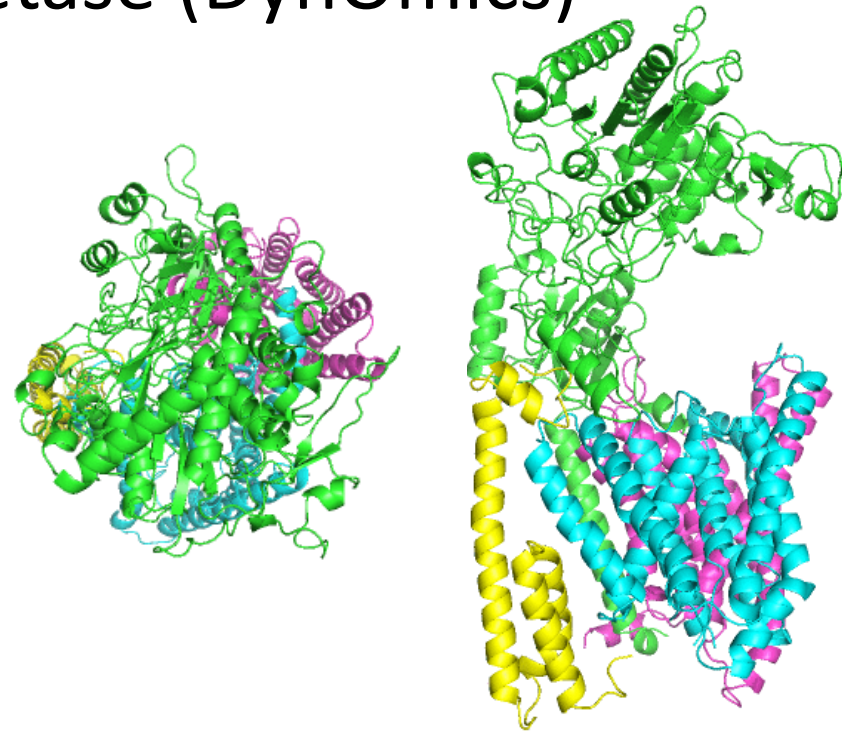
MembrANM for γ -secretase (DynOmics)

The effect of the membrane environment can be incorporated in the ANM analysis by adopting

- envANM-membrANM or
- substructure-membrANM.

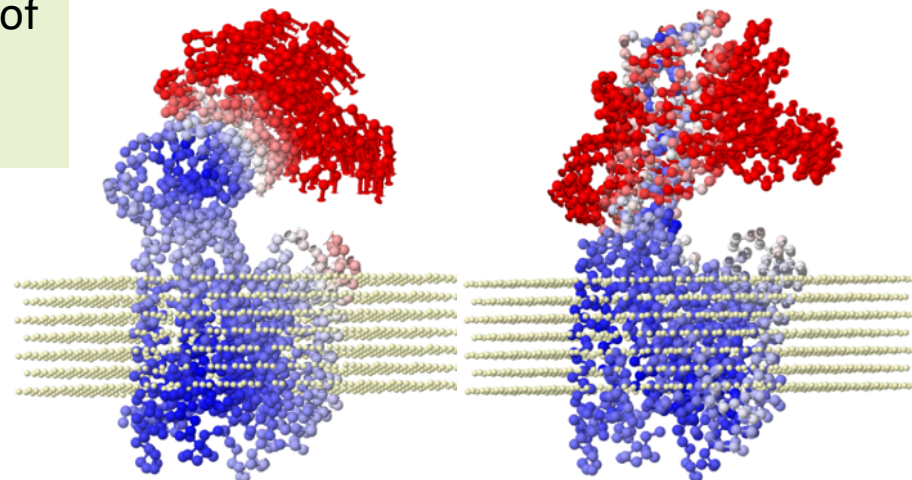
MembrANM results for the γ -secretase (PDB ID: 5FN2) can be viewed on the ENM server:

PyMol movies can be downloaded from DynOmics ENM server by generating “Full Atomic Structures for ANM-Driven Conformers”.



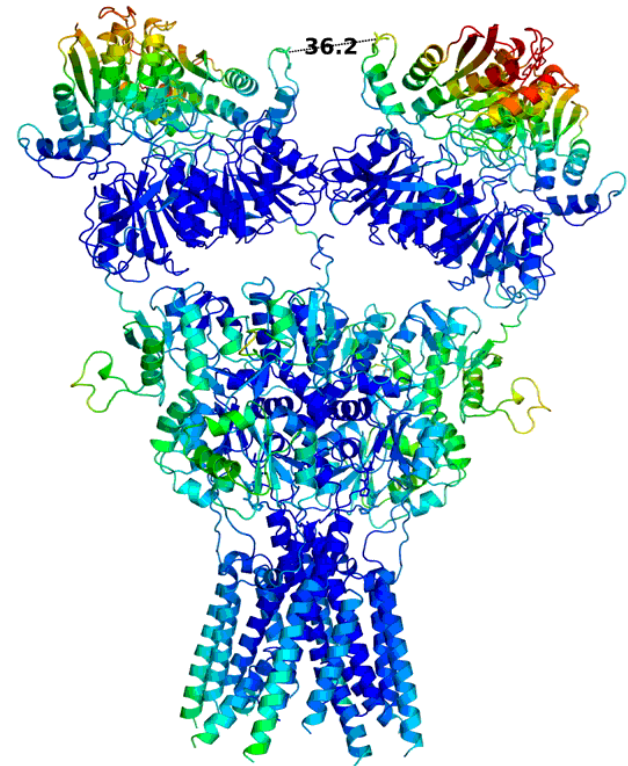
Mode 2 - twisting

Lee et al (2017) Allosteric Modulation of Intact γ -Secretase Structural Dynamics. *Biophys J* 113:2634-2649

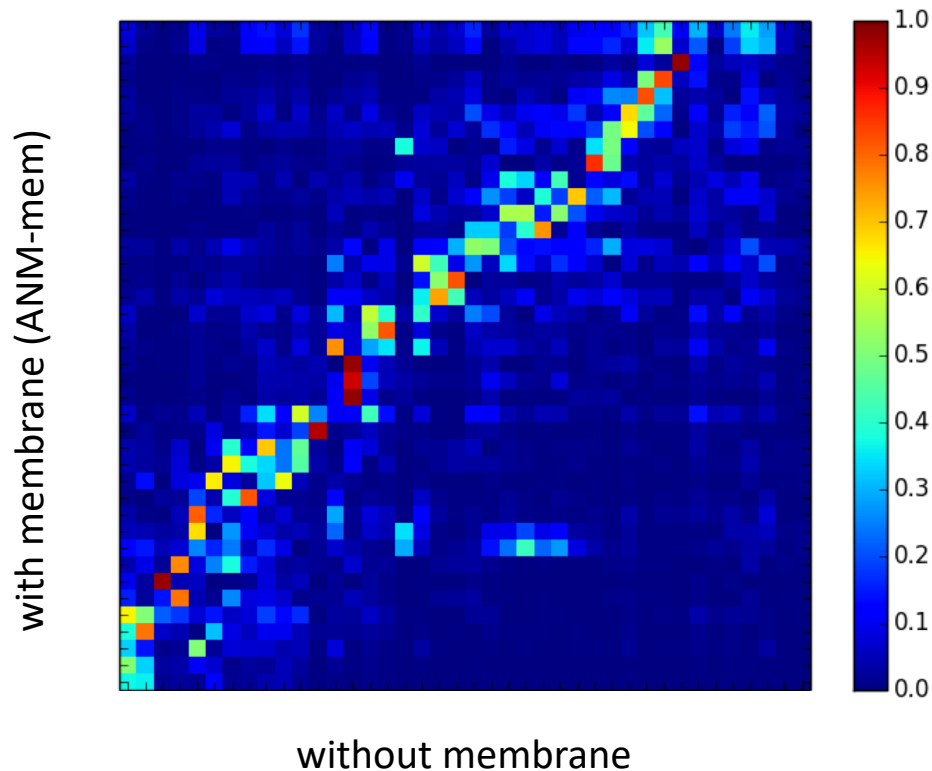


Collective dynamics of AMPA receptors

Experimentally verified by cross-linking experiments. Substitution of cysteines in the presence of an oxidizing agent promotes cross-linking.



Mode Comparison (DynOmics)

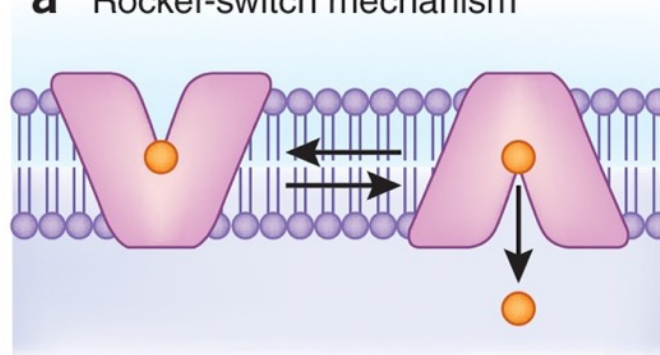


In the presence of membrane:

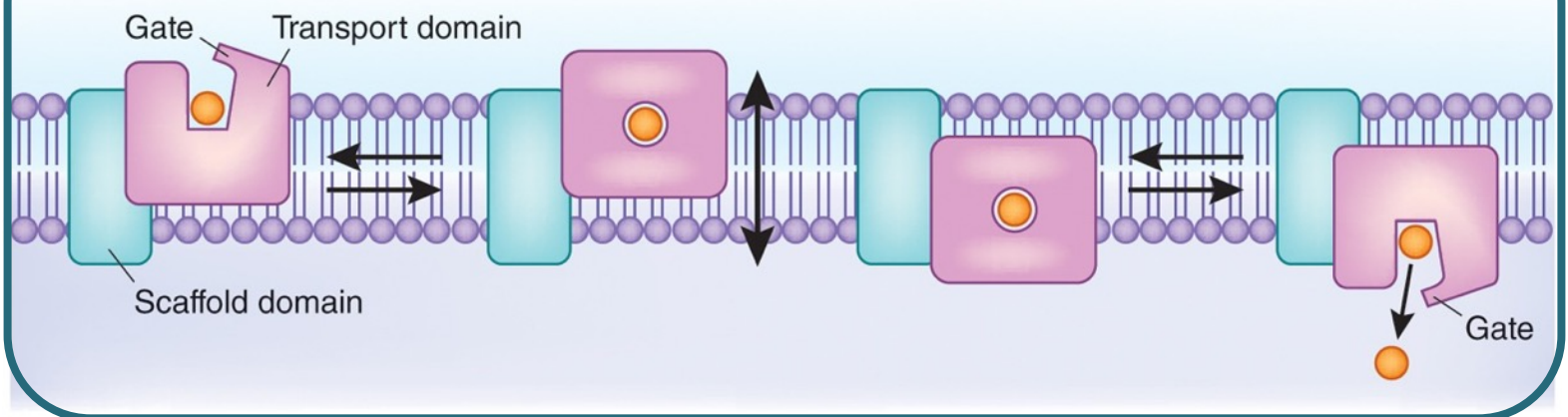
- Some intrinsic motions are preserved (*red*)
- Others are altered (*yellow*)
- New motions (mainly dominated by membrane fluctuations) emerge (blue rows)
- Diagonal shift due to new modes in the membrane

Alternating-access model. Transitions between outward-facing (OF) and inward facing (IF) states of membrane transporters

a Rocker-switch mechanism



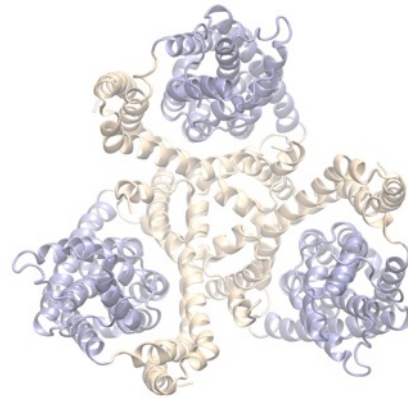
b Elevator mechanism



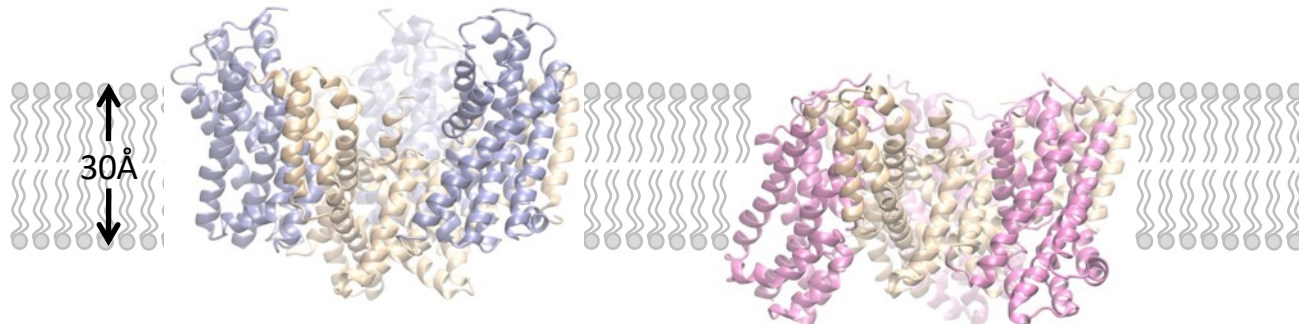
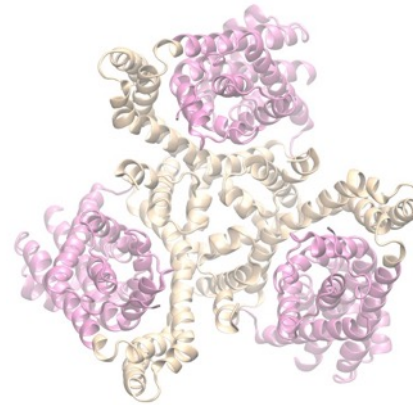
Global transitions of Glutamate Transporters

Trimeric membrane proteins, each subunit composed of:
scaffolding + transport domains

Outward Facing (OF)

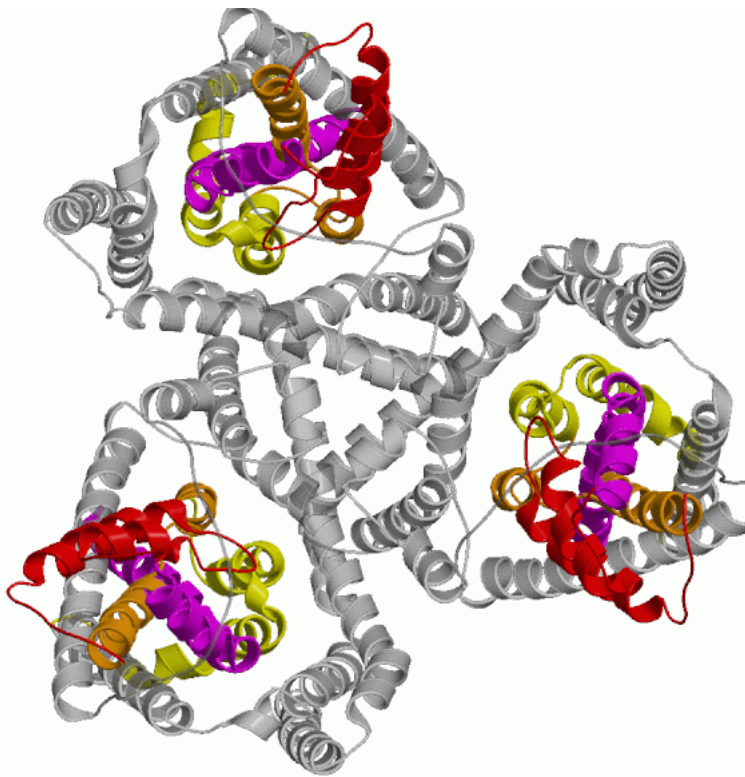


Inward Facing (IF)

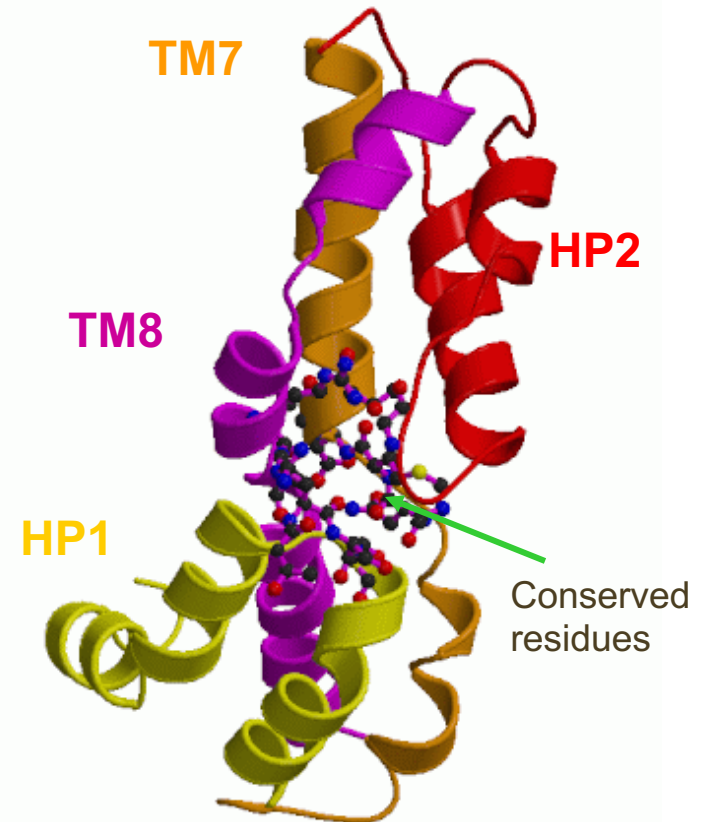


Domains of Aspartate Transporter

(also called EAAT – for excitatory amino acid transporter)



N-terminal (trimerization) domains in gray (TM1-TM6): **scaffolding domain**



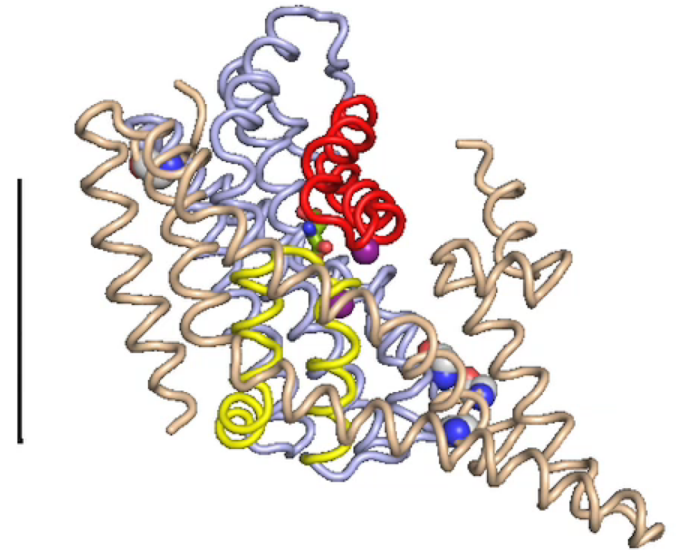
C-terminal Core (HP1-TM7-HP2-TM8): **transport domain**

Global transitions

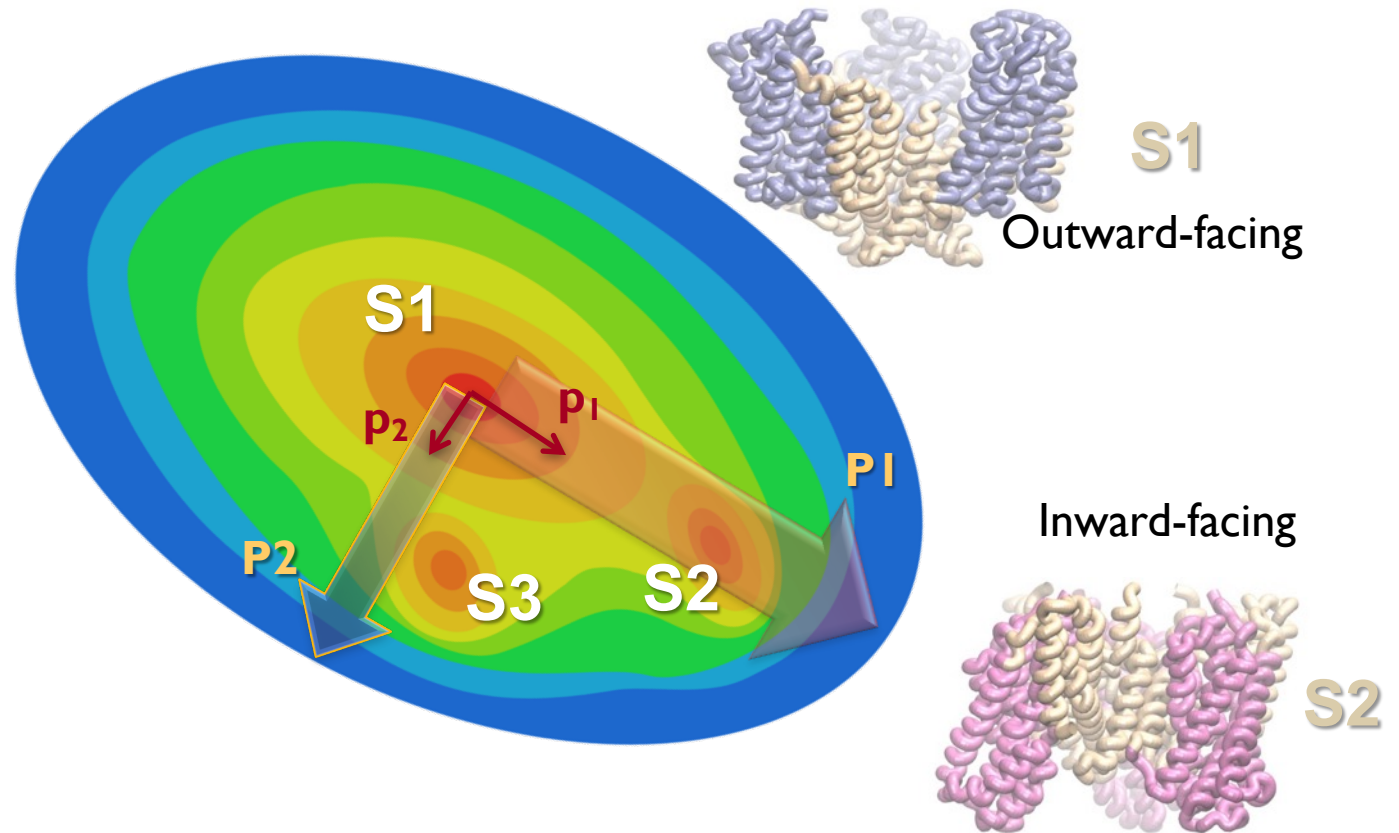
Transport domain undergoes elevator-like motions, while the trimerization scaffolding domain is rigidly affixed to the membrane

- Single subunit showing the transport domain moving across the membrane

- Translation of about 15 Å



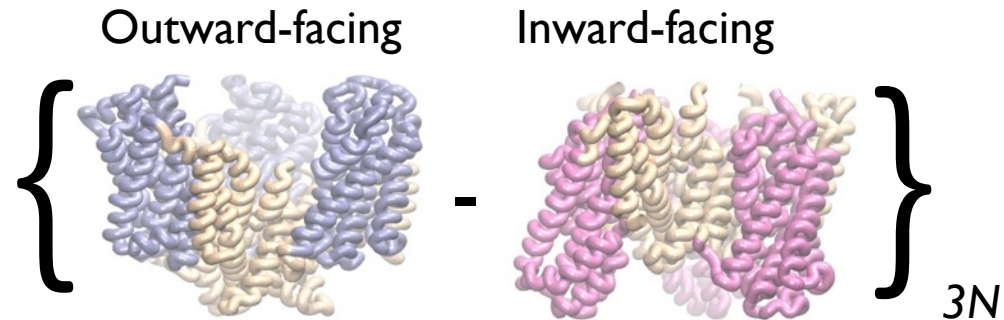
Substates are sampled along soft modes



Is this transition along a soft mode?

Correlation between soft modes & observed structural change

1. Evaluate the deformation \mathbf{d} ($3N$ unit vector)



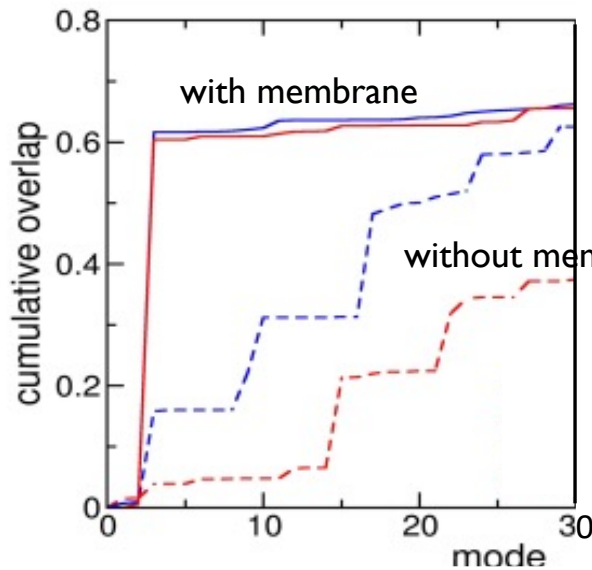
2. Calculate the **correlation cosine** $\cos(\mathbf{v}_i, \mathbf{d})$ between each mode vector, \mathbf{v}_i , and \mathbf{d}

3. **Cumulative overlap** = $[\sum_i \cos^2(\mathbf{v}_i, \mathbf{d})]^{1/2}$

Reference:

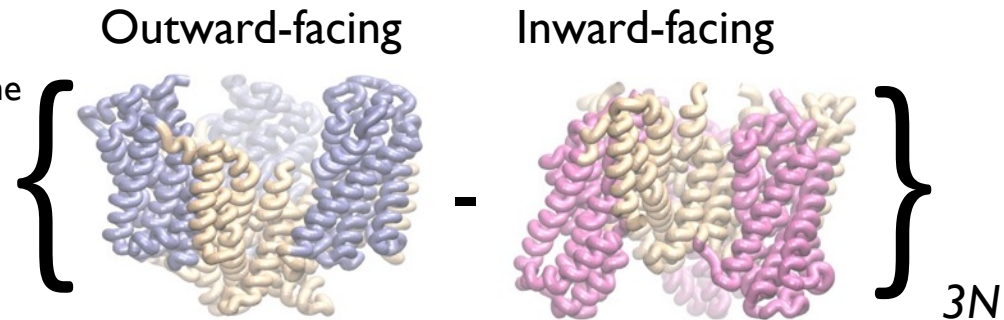
Lezon TR, Bahar I. (2012) Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter Glt_{ph} . *Biophys J.* **102**:1331-40.

Correlation between soft modes & observed structural change



Dr. Timothy R. Lezon

1. Evaluate the deformation \mathbf{d} ($3N$ unit vector)



2. Calculate the **correlation cosine** $\cos(\mathbf{v}_i, \mathbf{d})$ between each mode vector, \mathbf{v}_i , and \mathbf{d}

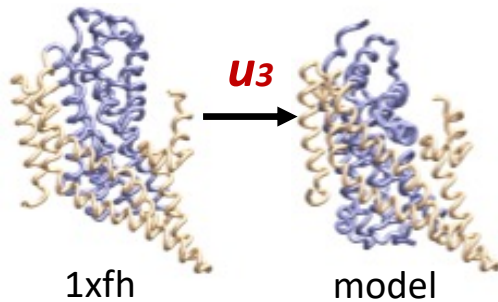
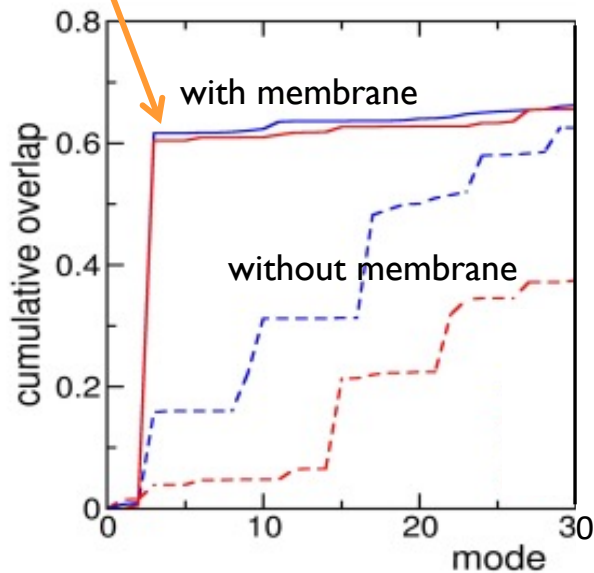
3. **Cumulative overlap** = $[\sum_i \cos^2(\mathbf{v}_i, \mathbf{d})]^{1/2}$

Reference:

Lezon TR, Bahar I. (2012) Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter Glt_{ph} . *Biophys J.* **102**:1331-40.

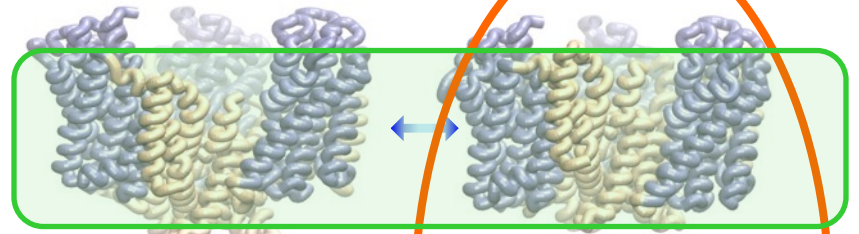
Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of $> 3,000$ modes

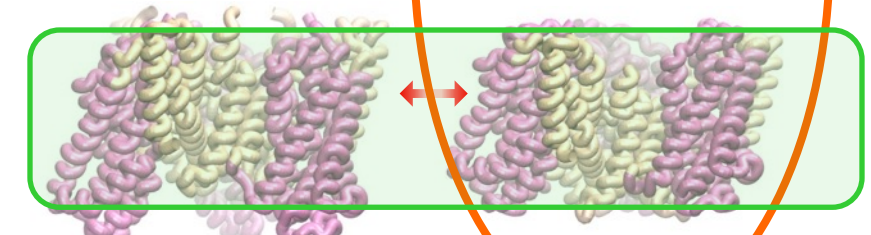


ENM global motions

Outward-facing



Inward-facing



in the presence of lipid constraints

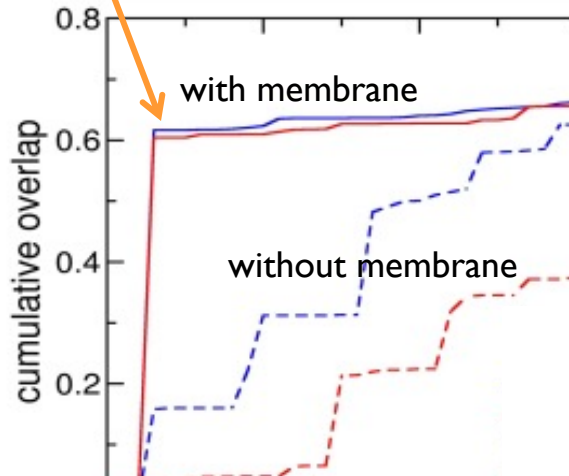
Reference:

Lezon TR, Bahar I. (2012) Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter Glt_{ph} . *Biophys J.* **102**:1331-40.

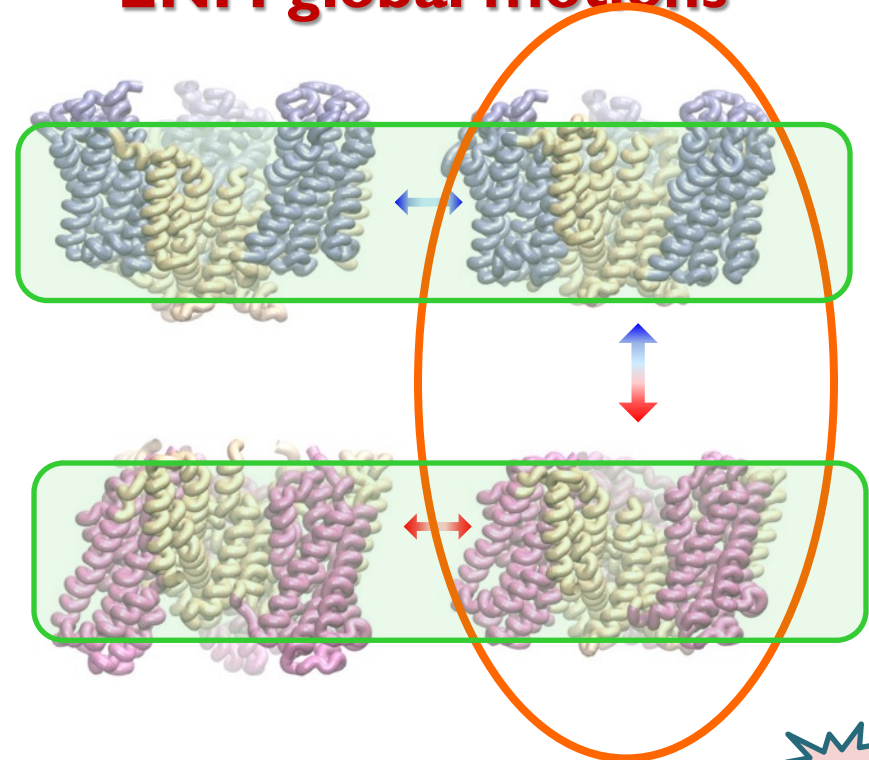
Membrane facilitates alternating access

Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of $> 3,000$ modes



ENM global motions



If the predicted modes were 'random', each mode would contribute by $1/3N$ to the cumulative overlap, i.e.

$$\cos(\mathbf{v}_k \cdot \mathbf{d}_{exp}) = (1/3N)^{1/2} = 0.0167$$

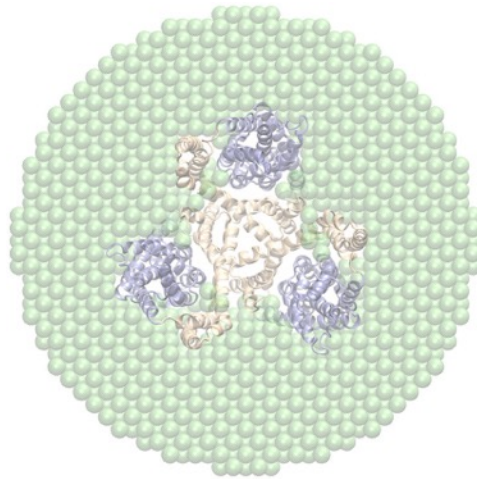
in the presence of lipid constraints

Reference:

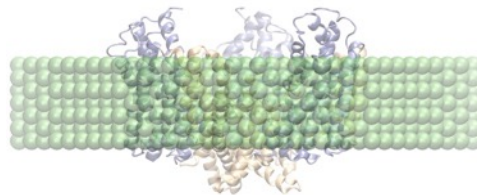
Lezon TR, Bahar I. (2012) Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter Glt_{ph} . *Biophys J.* **102**:1331-40.

Two approaches for including the lipid bilayer:

- Explicit membrane (a network model for the membrane)
- Implicit membrane (change in Hessian force constants)

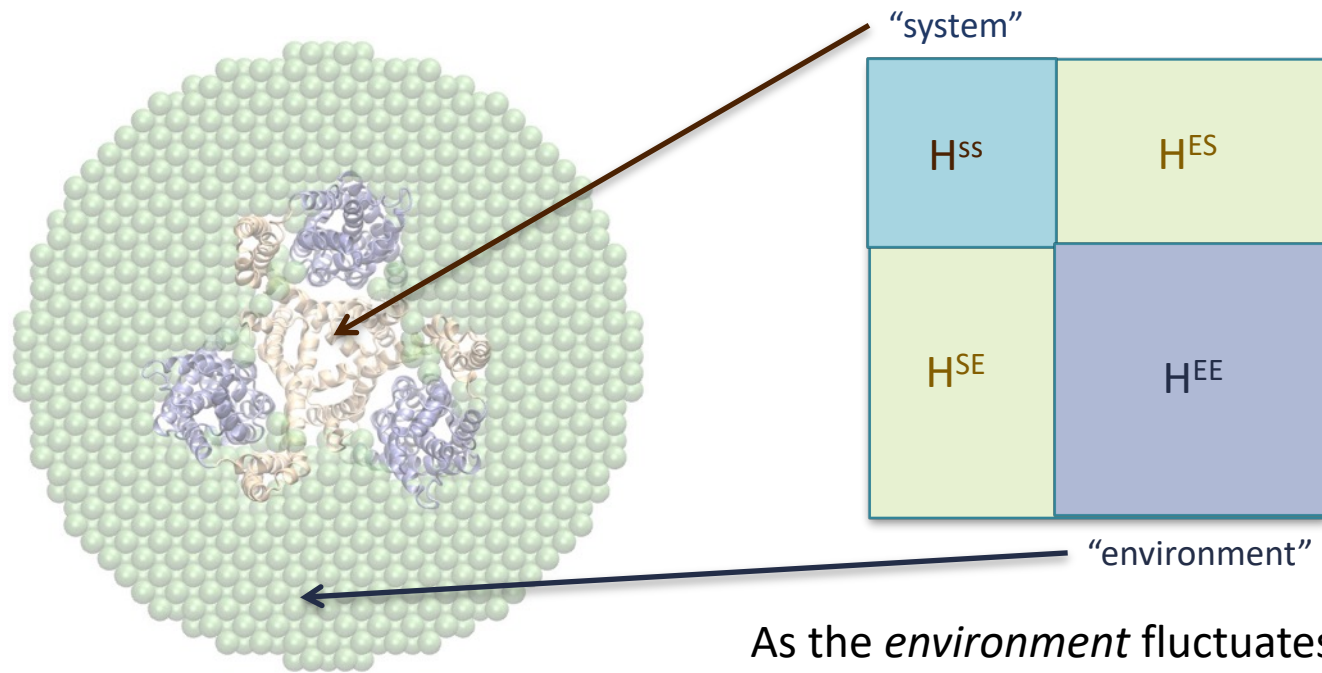


Explicit membrane top view



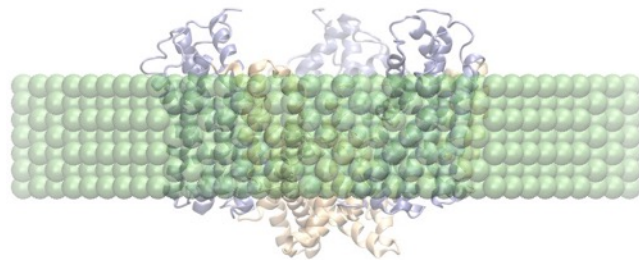
Explicit membrane side view

System/environment approximation



As the *environment* fluctuates randomly, the effective motion of the *system* is given by

$$V_{eff}(\mathbf{s}) = \frac{1}{2} \Delta \mathbf{s}^T \left(\mathbf{H}^{ss'} \right) \Delta \mathbf{s}$$
$$\mathbf{H}^{ss'} = \mathbf{H}^{SS} - \mathbf{H}^{SE} \left(\mathbf{H}^{EE} \right)^{-1} \mathbf{H}^{ES}$$



Implicit model for membrane effect

$$\mathbf{H}_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\ x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix}$$

Constraining effect of the lipids
in the radial direction (xy plane)

Altered radial force constants:

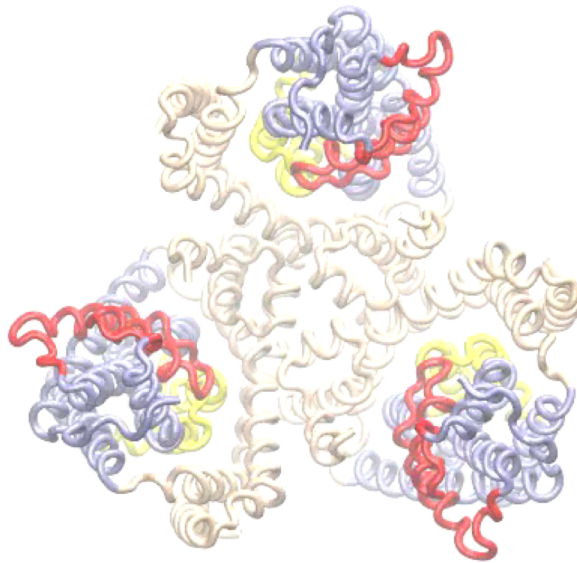
$$\gamma_x = \gamma_y = c^2 \gamma_z \quad c > 1$$

$$\mathbf{H}_{ij} = -(R_{ij}^0)^{-2} \begin{bmatrix} (x_{ij}^0 \sqrt{\gamma_x})^2 & x_{ij}^0 y_{ij}^0 \sqrt{\gamma_x \gamma_y} & x_{ij}^0 z_{ij}^0 \sqrt{\gamma_x \gamma_z} \\ x_{ij}^0 y_{ij}^0 \sqrt{\gamma_x \gamma_y} & (y_{ij}^0 \sqrt{\gamma_y})^2 & y_{ij}^0 z_{ij}^0 \sqrt{\gamma_y \gamma_z} \\ x_{ij}^0 z_{ij}^0 \sqrt{\gamma_x \gamma_z} & y_{ij}^0 z_{ij}^0 \sqrt{\gamma_y \gamma_z} & (z_{ij}^0 \sqrt{\gamma_z})^2 \end{bmatrix}$$

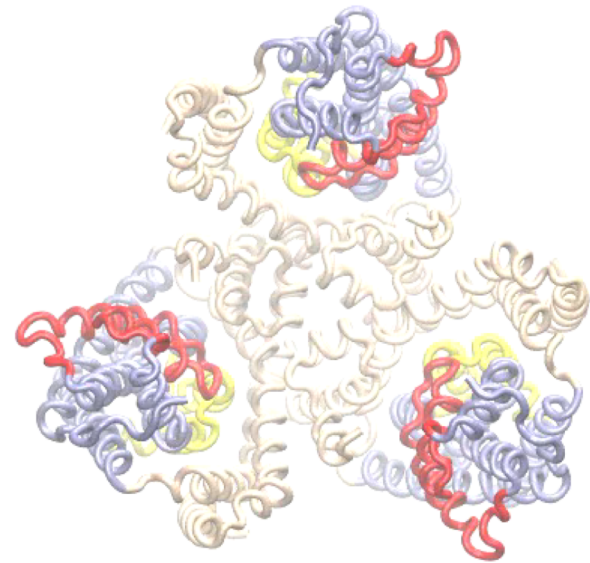
$$\mathbf{H}_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & c x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & c y_{ij}^0 z_{ij}^0 \\ c x_{ij}^0 z_{ij}^0 & c y_{ij}^0 z_{ij}^0 & (c z_{ij}^0)^2 \end{bmatrix}$$

Rotations and Translation of Blocks (RTB)
Tama F, Gadea FJ, Marques O, Sanejouand YH.
Proteins **2000** 41 1-7

Lipid bilayer favors elevator-like motions



ANM in the absence of membrane



ANM in the presence of membrane



Thank you!

