SUBSETS OF SLOW DYNAMIC MODES REVEAL GLOBAL INFORMATION SOURCES AS ALLOSTERIC SITES

by

Bengi Altıntel

B.S., Chemical Engineering, Boğaziçi University, 2020

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2022

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my thesis supervisor Prof. Türkan Haliloğlu for not only enlightening me with her immense knowledge from the very beginning to the end of this research but also helping me keep my faith in science always alive. It could not have been possible without her untiring support, loving patience, and enthusiasm throughout this process. It was a great pleasure and honor to work with her.

I would also like to give special thanks to my thesis committee members, Assoc. Prof. Nurcan Tunçbağ and Assist. Prof. Betül Uralcan for their time, attention, and invaluable comments.

I would like to give my sincere thanks to the members of the Polymer Research Center family, my dear friends; Dr. Burçin Acar, Özge Duman, Ayça Ersoy, Yiğit Kutlu, and Zeynep Erge Akbaş. They have taught me so much and created a so welcoming work environment that it would be very difficult to proceed without their support and positive energy.

I would like to give my deepest thanks to my beloved friend Yavuz Özbek for always being there for me whenever I needed him and keeping me motivated with his endless love. I would also like to express my sincere thanks to my dear housemate Kumsal Arat who always listened to me patiently, supported and cheered me up with her positive energy during all hard times.

I am grateful for my amazing friends whom I love and respect endlessly. I would like to express my infinite thanks to my dear sister Dr. Kübra Düz, who was with me throughout this journey, loving me the way I am for a decade and was always ready to give me coffee; to my beloved friend Dr. Pelin Pabuşçu, whom I know to be always there to help and support me even in the middle of the night; and lastly to dear Ayşen Kilitçioğlu for always making me feel her warmth even though there are thousands of kilometers between us.

Finally, I would like to extend my profound gratitude to my precious parents, Sevtap Altıntel and Bülent Altıntel for their unconditional love and unbelievable support. Thank you for always supporting my decisions without any hesitation. And thank you for always trusting me and making me feel special every single day. I feel so lucky to be your daughter.

It was a great opportunity for me that this study was financed by funds from The Scientific and Technological Research Council of Turkey (TUBITAK) and NATO Science for Peace Program with grant numbers 119F392 and G5685, respectively.

I would like to dedicate my thesis to my beloved grandmother, Gülten Güleryürek, whom I lost in the last semester of my master's degree, and whom I believe keep protecting and being proud of me from her lovely heaven with her amazing smile.

ABSTRACT

SUBSETS OF SLOW DYNAMIC MODES REVEAL GLOBAL INFORMATION SOURCES AS ALLOSTERIC SITES

Allostery is a crucial biological regulation mechanism, and dynamic information flow offers a framework to characterize allosteric interactions in causal links. Here, using a novel application of the Transfer Entropy (TE) calculations based on the Gaussian Network Model (GNM), it has been demonstrated how the dissection of dynamic information into subsets of slow dynamic modes reveals various layers of multi-directional allosteric pathways that are intrinsic in a particular protein structure. The degree of collectivity (Col) in the information transfer of residues with their TE values (TECol score) in these subsets of slow modes identifies particular residues as potent effectors and global information sources having a strong dynamic capacity to collectively disseminate information to other residues in the protein structure. These information source residues are linked to known active and allosteric sites, as demonstrated by aspartate transcarbamoylase (ATCase), Na^+/K^+-adenosine triphosphatase (Na^+/K^+-ATPase), and human transient receptor potential melastatin 2 (TRPM2), along with a dataset of 20 proteins. These specific residues provide feasible binding sites for structure-based rational drug design since they together affect/control others and direct pathways of allosteric communication.

ÖZET

YAVAŞ DİNAMİK MODLARIN ALT KÜMELERİ KOLLEKTİF BİLGİ KAYNAKLARINI ALLOSTERİK BÖLGELER OLARAK ORTAYA ÇIKARIR

Allosteri, biyolojik moleküller için önemli bir kontrol mekanizmasıdır ve dinamik bilgi akışı, nedensel allosterik etkileşimleri karakterize etmek için bizlere bir çerçeve sunar. Burada, Gauss Ağ Modeli (GNM) bazlı Transfer Entropi (TE) hesaplamalarının yeni bir uygulaması kullanılarak, dinamik bilginin yavaş dinamik modların alt kümelerine ayrımının, belirli bir protein yapısına özgü olan çok yönlü allosterik davranış katmanlarını nasıl ortaya çıkardığı gösterilmiştir. Yavaş modların bu alt kümelerinde rezidülerin sergilediği TE değerleri ile bu bilgi transferindeki kolektivite dereceleri (Col) bir arada kullanılarak, belirli rezidülerin güçlü efektörler olarak, yani protein yapısındaki diğer rezidülere toplu olarak bilgi yaymada güçlü dinamik kapasitelere sahip bilgi kaynakları olarak tanımlanması sağlanmıştır. Bu bilgi kaynağı rezidüler; aspartat transkarbamilaz (ATCaz), Na^+/K^+-adenozin trifosfataz (Na^+/K^+-ATPaz) ve insan geçici reseptör potansiyel melastatin 2 (TRPM2) proteinleri başta olmak üzere 23 proteinden oluşan bir veri seti üzerinde saptanarak, bu bilgi kaynağı rezidülerin bilinen aktif ve allosterik bölgelerle bağlantılı olduğu keşfedilmiştir. Bu spesifik rezidülerin protein yapısındaki birçok farklı rezidüyü etkileyerek yapı içindeki allosterik iletişim patikalarını doğrudan etkilediği/kontrol ettiği göz önüne alındığında, bu rezidülerin yapı bazlı rasyonel ilaç tasarımı için uygulanabilir bağlanma bölgelerine denk geldiği düşünülmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

i, j	Residue indeces
k	Mode component
k_B	Boltzmann constant
N	Residue number
N_s	The number of slow modes
n	The slow modes of 1 to 5
S	Entropy
T	Absolute temperature
t	Time
U	Matrix of eigenvectors
$\mathbf{R^0}$	Equilibrium position vector
R_{cut}	Cut-off radius
s	Subset of slow modes
V	Potential energy
X	Mean distance
Δ	Change in parameter
$\Delta T_{i\rightarrow j}$	The net information transfer from residue i to j
Å	Angstrom
α	Normalization factor
κ	The degree of collectivity
γ	Spring force constant
$\mathbf{\Gamma}$	Kirchhoff residue connectivity matrix
λ	Eigenvalue
$\mathbf{\Lambda}$	Diagonal matrix of eigenvalues
μ	Population mean
σ	Population standard deviation
τ	Time delay
τ_{opt}	Optimum time delay

LIST OF ACRONYMS/ABBREVIATIONS

0G6	D-Phenylalanyl-N-[(2S,3S)-6-{[Amino(Iminio)Methyl]Amino}-1-Chloro-2-Hydroxyhexan-3-yl]-L-Prolinamide
16G	2-Acetamido-2-Deoxy-6-O-Phosphono-Alpha-D-Glucopyranose
A	Alanine
2A2	4-[4-(3-Chlorophenyl)Piperazin-1-yl]-4-Oxobutane-1-Thiol
2D	Two Dimensional
3D	Three Dimensional
ABC	ATP-Binding Cassette
ACC	Accuracy
ACT	Acetate Ion
ADP	Adenosine Diphosphate
ADPR	ADP-Ribose
AGY	2-[(4-Tert-Butyl-3-Nitrophenyl)Carbonyl]-N-Naphthalen-1-Ylhydrazinecarboxamide
ALF	Tetrafluoroaluminate Ion
AMP	Adenosine Monophosphate
ANM-LD	Anisotropic Network Model Guided Langevin Dynamics
APS	Adenosine-5'-Phosphosulfate
ATCase	Aspartate Transcarbamoylase
ATP	Adenosine Triphosphate
ATPase	Adenosine Triphosphatase
BB3	3-(3,5-Dibromo-4-Hydroxy-Benzoyl)-2-Ethyl-Benzofuran-6-Sulfonic Acid Dimethylamide
BI4	3-(1h-indol-3-yl)-4-{1-[2-(1-Methylpyrrolidin-2-yl)Ethyl]-1H-indol-3-yl}-1Hpyrrole-2,5-Dione
BRCA1	Breast Cancer 1
BtuCD	Vitamin B12 Transporter
C	Cysteine
Col	Collectivity in Information Transfer

CTP	Cytidine-5'-Triphosphate
D	Aspartic Acid
DFY	5,6-Diphenyl-Furo[2,3-D]Pyrimidin-4-Ylamino)-Acetic
DH	Dehydrogenase
DNA	DeoxyriboNucleic Acid
DNA-PK	The DNA-dependent Protein Kinase
DNA-PKcs	Catalytic Subunit of DNA-PK
E	Glutamic Acid
F	Phenylalanine
F1G	1-Methyl-3-Trifluoromethyl-1h-Thieno[2,3c]Pyrazole-5-Carboxylic Acid (2-Mercapto-Ethyl)-Amide
F6P	Fructose-6-Phosphate
FBP	1,6-Di-O-Phosphono-Beta-D-Fructofuranose
Flim	Flagellar Motor Switch Protein
FN	False Negative
FP	False Positive
FRET	Fluorescence Resonance Energy Transfer
G	Glycine
GLC	Alpha-D-Glucopyranose
GLU	Glutamic Acid
GNM	Gaussian Network Model
GNP	Phosphoaminophosphonic Acid-Guanylate Ester
GTP	Guanosine-5'-Triphosphate
L	Leucine
MD	Molecular Dynamics
MRK	2-Amino-4-Fluoro-5-[(1-Methyl-1h-Imidazol-2-Yl)Sulfanyl]-N-(1,3-Thiazol-2-Yl)Benzamide
MSF	Mean Square Fluctuation
N	Asparagine
NAD	Nicotinamide-Adenine-Dinucleotide
NADH	1,4-Dihydronicotinamide Adenine Dinucleotide
NPF	2-Nitrophenyl Beta-D-Fucopyranoside
Opttau	Optimum Time Delay

PALA	N-Phosphonacetyl-L-Aspartate
PDB	Protein Data Bank
PGA	2-Phosphoglycolic Acid
PLP	Pyridoxal-5'-Phosphate
PMB	Paramercury-Benzenesulfonic Acid
PPS	3'-Phosphateadenosine-5'-Phosphate Sulfate
PRE	Precision
PTR	O-Phosphotyrosine
R	Arginine
R-state	Relaxed State
S	Serine
SER	Serine
SN	Sensitivity
SP	Specificity
TE	Transfer Entropy
TECol	Score of Collective Information Transfer
TN	True Negative
TP	True Positive
TRPM2	Human Transient Receptor Potential Melastatin 2
T-state	Tense State
U5P	Uridine-5'-Monophosphate
V	Valine
Y	Tyrosine
z-VAD-FMK	Carbobenzoxy-Valyl-Alanyl-Aspartyl-[O-Methyl]-Fluoromethylketone

1. INTRODUCTION

Proteins carry out a variety of tasks in cells while adapting to various environments and engaging numerous partners. This adaptability is a result of their dynamic nature, which enables dynamic allosteric interactions with the protein active sites rather than direct interactions to modify protein activity. Drug design efforts are especially interested in allostery since it is a crucial mechanism for regulating functions [1-3].

When Jacques Monod and François Jacob originally used the term allostery in 1961 [4], it was more of a structural phrase than a dynamical one. In years [5–14], the significance of dynamics in allostery has become increasingly clear. Dynamic fluctuations are strongly related to structural changes that result from perturbations. Allosteric modulation, however, is assumed by a shift in dynamic modes and their frequency and primarily entails an entropic effect, also known as dynamic allostery, in the absence of conformational modifications [10,14-18]. As has already been taken into account in various graph-based computations [19–23], allosteric pathways offer feasible paths for communication among functional regions within the network of dynamic interplay. Communities of residues on these interaction networks show functional significance and appear as potential sites for global transmission of allosteric information [24–26].

Correlated fluctuations between residue pairs can be explained in terms of the mutual information which measures how much the uncertainty in one residue's own fluctuations is decreased by the knowledge of fluctuations of the other. Transfer entropy (TE) [27] can therefore be used to assess the allosteric communication by seeing it as a straightforward information flow from one residue to another. In order to account for causal allosteric interactions, TE incorporates a time delay into the mutual information of residues using Shannon's entropy as a measure of information flow. Information flow from Molecular Dynamics (MD) trajectories could be retrieved using the transfer entropy approach [28–33], although the available MD simulation lengths are frequently in doubt.

Recently, a coarse-grained analytic approach based on Gaussian Network Model (GNM) combined with Schreiber's information transfer concept has been proposed [34] that it is utilized in the present TE calculations of this study.

GNM being straightforward, coarse-grained method for studying biological macromolecules represents these molecules as an elastic mass-and-spring model to analyse their mechanical characteristics of long-term dynamics. In an ensemble of substates, or in a dynamic mode spectrum, characterized by the 3D protein structure, GNM predicts complex structural fluctuations. The 3D protein structure specifies an ensemble of substates, i.e., conformers, or a spectrum of dynamic modes, in which GNM predicts complicated structural fluctuations (See Figure 1, Materials & Methods). These conformers such as ligand-bound, active/inactive, and open/closed states, are frequently inherently attainable, especially via a small group of slow dynamic modes which have the capacity to characterize specific functional movement patterns by moving several regions simultaneously and collectively just like choreographing out of basic dynamic aspects [10,17,35–38]. Using the vitamin B12 importer protein BtuCD as an example, the computational identification showed the impact of particular slow modes in the conformational change between the outward and inward conformers, which was consistent with the behavior seen in experimental FRET distances and point mutations [37,39]. Therefore, due to the dominant effect of slower modes or the slowest mode, certain functionally important slow modes of motion -thus related allosteric communication pathway- may not be evident if all slow mode residue fluctuations are contributed as an average.

Slow modes incorporate allosteric interactions, whose distinctive cooperativity allows coordinated action necessary for the generation of allosteric communication pathways within a protein structure. As a result, allosteric effects frequently result from activating or changing these global modes that already exist [10,17,37,38]. Global modes not only enable conformational changes between substates, but also render certain directional allosteric communication pathways [10,37,38,40,41]. For instance, the particular slow modes -which provide conformational transition between the outward and inward facing conformers of the two ATP-dependent ABC transporters BtuCD and MalFGK- are associated with opposite directional allosteric pathways pursuant to their distinctive functionalities [37]. Because of this, it is crucial to decompose dynamic behavior, in this case transfer entropy, into

individual or subsets of slow modes to be able to comprehend all possible causal relations and allosteric pathways that intrinsically exist within a particular topology.

The aim of this thesis is to show that subsets of slow dynamic modes reveal collective residues in the sense of information transfer, in an other saying global information source residues, as allosteric sites in a protein structure. An information-theoretic strategy to determine allosteric interactions between residues has been developed and used to identify the subsets of slow modes that maximize collective information transfer behaviour of residues. Thus, each residue is thought to have a specific amount of transferable information, namely degree of collectivity, which is based on the protein's vibrational modes of motion [34,37,38]. This collectivity measure is created based on the idea of transferable information/entropy capacity and the Bruschweiler's theory [42] and it assesses how globally a residue affects and controls the other residues in the protein structure. The term "global" is used for the residues demonstrating high collectivity in transferring this information. Here, these global information source residues are believed to be plausible binding sites for possible drugs as they collectively affect others and guide allosteric pathways.

Allosteric interactions between different sites produced by various subsets of slow modes may have causal links that are relevant to the functional mechanism involving a particular kind of motion. As a result, this may help to accurately comprehend the allosteric phenomena with functional modularity and make it possible to anticipate functionally important regions that are either active or allosteric sites within multi- or bi-directional allosteric communication pathways.

2. MATERIALS AND METHODS

2.1. Dataset

The proposed information-theoretic approach is validated using a dataset of the 20 known allosteric proteins from Amor *et al*. [21] and three additional allosteric examples of interest. The corresponding PDB codes of the active and allosteric state structures of the 23 test proteins with their known active and allosteric ligand binding sites are provided in Table 2.1. With consideration of multiple domain configurations, multimerization states, various ligands, and different protein sizes from 128 to 3432 residues, Amor's dataset covers a large region of the cytosolic protein structures [21]. Incorporation of three extra test proteins, two membrane proteins, human TRPM2 ion channel (PDB: 6MIX [43]) with 5348 residues and Na^+/K^+-ATPase transporter (PDB: 4HQJ [44]) with 1297 residues, and a large human DNA-PK holoenzyme (PDB: 5Y3R [45]) with 4670 amino acids, to this dataset offers further variety.

Table 2.1. Dataset of allosteric proteins.

Protein	**Res. Num.**	**PDB (Active)**	**Chain**	**Active Site Ligand**	**Binding Sites**	**PDB (Allos-teric)**	**Chain**	**Allos-teric Site Ligand**	**Binding Sites**
ATCase	2778	**1D09**	A, C, E, G, K, M (310)	PALA	51-55, 80, 84, 105, 134, 137, 167, 229, 231, 267.	**1RAC**	B, D, F H, L, N (153)	CTP	9, 11, 12, 17, 19, 60, 82, 84, 86, 89, 91, 94.
Gluco-kinase	447	**1V4S**	A (447)	GLC	151-153, 168, 169, 204, 205, 225, 229-231, 256, 287, 290.	**1V4S**	A (447)	MRK	61-66, 210, 211, 214, 215, 220, 235, 452, 455.
				ATP *(from 3ID8)*	78-83, 228, 295, 296, 332-336, 411-415.				

Table 2.1. Dataset of allosteric proteins (cont.).

Lac Repres-sor	658	**1EFA**	A, B (328)	DNA	5-7, 15-19, 21, 22, 24, 29, 47, 49, 50, 53, 54, 56, 57, 59.	**1EFA**	A, B (328)	NPF	73, 74, 75, 76, 79, 125, 148, 149, 161, 193, 197, 220, 246, 248, 274, 291, 293, 296.
H-Ras GTPase	166	**3K8Y**	A (166)	GNP	12-18, 28-32, 34, 35, 60, 116, 117, 119, 120, 145-147.	**3K8Y**	A (166)	ACT	97, 101, 106-109, 111.
Fructos e -1,6 Bisphos phatase	1324	**1EYI**	A, B, C, D (331)	F6P	121, 122, 212, 215, 243, 244, 246-248, 264, 274, 275, 280.	**1EYJ**	A, B, C, D (327)	AMP	17, 20, 21, 24, 26-31, 112, 113, 140, 160, 177.
Phosph oglycer ate DH	1592	**1YBA**	A, B, C, D (398)	NAD	84,106,108, 112, 158-162, 180-182, 185, 210-212, 216,217, 220, 238-240, 264, 265,292, 294,295.	**1PSD**	A, B, C, D (398)	SER	344-350 370.
Lactate DH	1252	**1LTH**	A, B, C, D (313)	NAD	13, 14, 16-18, 38-40, 44, 82-86, 103, 107, 123-125, 180, 230, 237, 240.	**1LTH**	A, B, C, D (313)	FBP	158, 170, 171, 173, 175.
Caspa-se-1	512	**2HBQ**	A, D (168)	z-VAD-FMK	179, 236-238, 283-285.	**2FQQ**	A, D (147)	F1G	258, 286.
			B, E (88)	z-VAD-FMK	339-343, 348, 383.		B, E (88)	F1G	331, 390, 391.
ATP Sulfu-rylase	3432	**1I2D**	A, B, C, D, E, F (572)	APS	196-200, 206, 209, 267,289, 291,292, 294, 295, 331-333.	**1M8P**	A, B, C, D, E, F (573)	PPS	405, 434, 437, 446, 451, 454, 455, 476-479, 515, 517, 527-530.

Table 2.1. Dataset of allosteric proteins (cont.).

Glutamate DH	3006	**6DHD**	A, B, C, D, E, F (501)	NADH-GLU	90-92, 94, 111, 114, 126, 134, 166-170, 211, 215, 250-255, 275-276, 295, 325-327, 330, 347-349, 374, 377, 378, 381.	**6DHD**	A, B, C, D, E, F (501)	NADH-GTP	85-87, 115, 116, 119-122, 195, 205, 206, 209, 210, 213, 217, 257, 258, 261, 262, 265, 292, 387, 388, 391-393, 446, 450, 488, 491.
glcN-6-P deaminase	1596	**1HOT**	A, B, C, D, E, F (266)	PO_4	42-44, 172, 208.	**1HOT**	A, B, C, D, E, F (266)	16G	1, 2, 151, 152, 158-161.
ADP-glucose phosphorylase	1727	**1YP3**	A, B, C, D (432)	ATP	26, 28,29, 43, 73,75, 118-120, 123, 143-145, 233, 253-256.	**1YP2**	A, B, C, D (428)	PMB	297, 298, 322, 324, 338-340, 354, 355,
Phosphofructokinase	1276	**4PFK**	A, B, C, D (319)	F6P	127, 169-171, 222, 249, 252.	**6PFK**	A, B, C, D (319)	PGA	21, 25, 55, 58, 59, 154, 185, 211, 213, 214.
UPRTase	846	**1XTT**	A, B, C, D (215)	U5P	81, 105, 140, 142-148, 201-203, 208-210.	**1XTU**	A, B, C, D (215)	CTP	29, 30, 33, 37, 87, 90, 91, 93, 94,96.
PTP1B	298	**1PTY**	A (298)	PTR	24, 27, 46, 48, 49, 182, 215-221, 254, 258, 259, 262.	**1T48**	A (292)	BB3	189, 192, 193, 196, 200, 276, 277, 280, 291, 292.
Glycogen Phosphorylase	3292	**7GPB**	A, B, C, D (823)	PLP	138, 491, 568, 648-650, 653, 675-677, 680.	**7GPB**	A, B, C, D (823)	AMP	42-45, 67, 68, 71, 72, 75, 76, 309, 310.
				SO_4	135, 569, 574.			SO_4	11, 14, 16, 69.
Thrombin	280	**1SFQ**	A (29)	-	-	**1SFQ**	A (29)	-	-
			B (251)	0G6	57, 60A, 60D, 97-99, 189-195, 214-216, 219, 220, 226.		B, E (251)	NA	221, 224.

Table 2.1. Dataset of allosteric proteins (cont.).

PDK1	277	**3ORZ**	A (277)	BI4	88-91, 109, 111, 143, 159-162, 209, 212, 222, 223.	**3ORZ**	A (277)	2A2	115, 118, 119, 124, 148, 149, 155.
CHK1	257	**2BRG**	A (257)	DFY	15, 16, 23, 36, 68, 84-87, 90, 91, 137, 147.	**3JVS**	A (256)	AGY	93-98, 133, 173, 200, 204-206.
CheY	128	**1F4V**	A (128)	BEF_3	57-59, 87, 88, 109, 129.	**1F4V**	A (128)	FliM Peptide	90-92, 95, 99, 103-108, 119, 122.
Human TRPM2 Ion Channel (Apo State)	5348	**6MIX**	A, B, C, D (1337)	Ca *(from 6PUU)*	842, 846, 869, 1073.	**6MIX**	A, B, C, D (1337)	ADPR *(from 6PUU)*	174, 175, 176, 177, 178, 295, 332, 333, 334, 335, 336.
Na+/K+ ATPase	1297	**4HQJ**	A (985)	ADP-ALF-Mg	369, 370, 371, 443, 445, 446, 475, 477, 482, 501, 502, 544, 546, 610-613, 685, 691, 710, 713.	**4HQJ**	A (985)	Na	322, 323, 325-327, 771, 774-776, 779, 804, 808, 854, 923, 926.
			B (285) G (27)	-	-		B (285) G (27)	-	-
Human DNA-PK Holo-enzyme	4670	**5Y3R**	C (3636)	Catalytic Sites	3921-3927.	**5Y3R**	A (493)	DNA	35, 80, 254-258, 275, 278, 282, 284, 285, 331, 363, 366, 403, 404.
							B (526)	DNA	265, 272, 275, 397-402, 404, 486.
							K (15)	DNA	12, 13.
							C (3636)	DNA	123-125, 128, 167-169, 259-261, 2231, 2311-2313, 2356, 2357.

2.2. Theoretical Background and Computational Methodology

2.2.1. The Gaussian Network Model (GNM)-based Transfer Entropy (TE)

GNM [46,47] predicts residue fluctuations and their correlations assuming a Gaussian probability distribution for instant caneous residue fluctuations ($\Delta\mathbf{R}$). It models the protein structure as an elastic network of α-carbon atoms of residues interacting with a harmonic potential function within a threshold radius (R_{cut}=10 Å). Using the γ force constant and $\Delta\mathbf{R}$ fluctuation vectors, potential energy V is defined as

$$V = (\gamma/2)\, \Delta\boldsymbol{R}^T \boldsymbol{\Gamma} \Delta\boldsymbol{R} \tag{2.1}$$

where $\boldsymbol{\Gamma}$ is the Kirchhoff residue connectivity matrix. With the pseudoinverse of $\boldsymbol{\Gamma}$ revealing N-1 intrinsic modes of motion, the correlation between fluctuations of residue pairs, i and j, is given as

$$\begin{aligned} < \Delta\boldsymbol{R}_i . \Delta\boldsymbol{R}_j > &= (3k_B T/\gamma)[\boldsymbol{\Gamma}^{-1}]_{ij} , \\ &= (3k_B T/\gamma)[\boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T]_{ij} , \\ &= (3k_B T/\gamma) \sum_k [{\lambda_k}^{-1} \boldsymbol{U}_k {\boldsymbol{U}^T}_k]_{ij} \end{aligned} \tag{2.2}$$

where $\boldsymbol{\Lambda}$ and $\mathbf{U}$ are respectively the eigenvalue and eigenvector matrices of Γ with λ_k and $\mathbf{U}_k$ representing the k^{th} mode component, k_B is the Boltzmann constant, and T is the absolute temperature in Kelvin. Modes are ranked in ascending order of eigenvalues with k=1 is the slowest mode and k=N-1 is the fastest mode.

Following, the correlation of zero-time fluctuations of i with future τ time fluctuations of j can be expressed as [47]

$$\begin{aligned} < \Delta\boldsymbol{R}_i(0) . \Delta\boldsymbol{R}_j(\tau) > &= (3k_B T/\gamma)[\boldsymbol{\Gamma}^{-1}]_{ij}\, e^{-\lambda_k \tau/\tau_0} , \\ &= (3k_B T/\gamma) \sum_k [{\lambda_k}^{-1} \boldsymbol{U}_k {\boldsymbol{U}^T}_k]_{ij}\, e^{-\lambda_k \tau/\tau_0} \end{aligned} \tag{2.3}$$

where τ_0 is a characteristic time of the vibrational dynamics of all folded proteins [47].

GNM based time-delayed correlations in conditional Shannon entropies, defining the amount of transfer entropy, i.e., information transfer, $T_{i\to j}(\tau)$ from residue i to j in time delay τ, using the expression by Schreiber [27] as

$$T_{i\to j}(\tau) = S(\Delta \boldsymbol{R}_j(t+\tau)|\Delta \boldsymbol{R}_j(\tau)) - S(\Delta \boldsymbol{R}_j(t+\tau)|\Delta \boldsymbol{R}_i(\tau), \Delta \boldsymbol{R}_j(\tau)). \quad (2.4)$$

The conditional Shannon entropies are calculated as described in [34] under the Equations 2.2 and 2.3. The net information transfer from residue i to j at time τ can be written as

$$\Delta T_{i\to j}(\tau) = T_{i\to j}(\tau) - T_{j\to i}(\tau) \quad (2.5)$$

where $\Delta T_{i\to j}(\tau)$ estimates the direction of information flow between residues i and j in a certain time delay τ, yielding the degree to which the present movement of residue i decreases the amount of uncertainty for the future movement of residue j. If, $\Delta T_{i\to j}(\tau) > 0$ then the dynamics of residue i affects the dynamics of residue j, meaning a causal directional relationship between fluctuations of residues i and j.

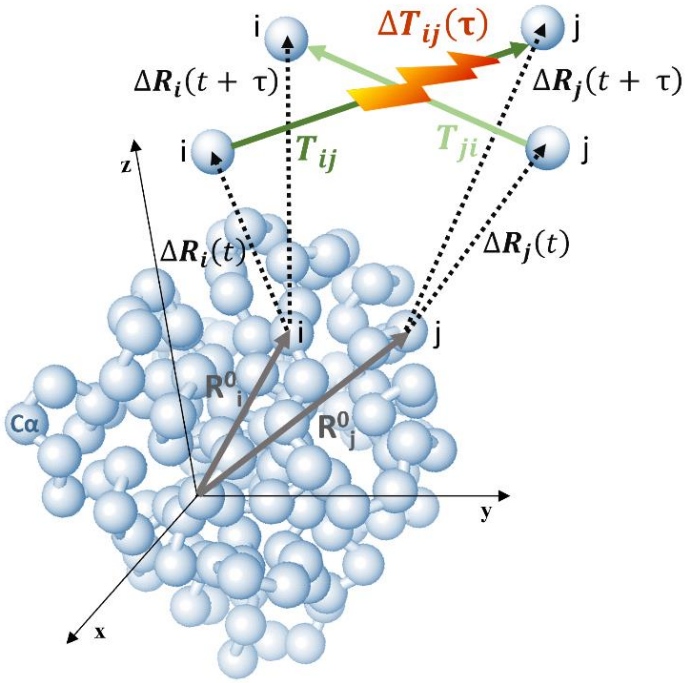

Figure 2.1. Schematic view of GNM-based information transfer.

In Figure 2.1, spheres represent C^{α} atoms of residues. $\mathbf{R^0}_i$ and $\mathbf{R^0}_j$ with gray arrows represent equilibrium position vectors of residues i and j. $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ with dashed black arrows are respectively the fluctuation vectors from the equilibrium positions of residues i and j at time t and t+τ. The information transfer between residue pairs of i and j with τ is described by green arrows. The difference between T_{ij} and T_{ji} described by red thunder reveals the net information transfer (TE) from residue i to j at time τ (ΔT_{ij} (τ)).

Additionally, the information transfer capability of a residue can be defined as how much it transfers information to the rest of the protein, as defined by cumulative TE formulation as

$$\Delta T_{i\to rest}(\tau) = \sum_{j=1}^{N} T_{i\to j}(\tau). \tag{2.6}$$

Maximum and minimum cumulative TE ($T_{i\to rest}(\tau)$) values are regarded as entropy/information sources and sinks respectively. Entropy sources send and entropy sinks receive information to/from the rest of the protein. Two parameters are effective in TE values: time delay τ and slow mode component k. Here instead of all modes, we develop a measure to identify distinct sets of slow modes to disclose plausible layers of allosteric communication pathways.

2.2.2. The Degree of Collectivity in the GNM-based Transfer Entropy

The residues with high net information transfer to many other residues are likely powerful effectors meaning that these residues affect most of the others, as dynamically key sites that lead and control rest of the protein. The motivation is thus to dissect dynamic information by the decomposition of the internal dynamics to disclose the subsets of slow modes maximizing the collectivities of residues in their information transfer: the net TE ($\Delta T_{i\to j}(\tau)$) values.

The degree of collectivity (Col) values can be calculated using the positive net TE values of residues with τ, to determine the most collective information source residues (effectors) benefiting from Bruschweiler's study [42] as

$$\kappa_{i,s} = \frac{1}{N}\, exp\left(-\sum_{j=1}^{N} \alpha(\Delta T_{ij,k}(\tau))^2\, log(\alpha(\Delta T_{ij,k}(\tau))^2)\right) \quad (2.7)$$

where $\kappa_{i,s}$ is the Col value of residue i in the information transfer through the subset of slow modes (s), N is the residue number, and $\Delta T_{ij,k}(\tau)$ is the positive net transfer entropy from residue i to j in slow mode k for time delay τ. In Equation 2.7, α is a normalization factor that is determined as

$$\sum_{i=1}^{N} \alpha(\Delta T_{ij,k}(\tau))^2 = 1. \quad (2.8)$$

One-dimensional plots comprised of the Col values of residues in the information transfer yield powerful effectors as global information source residues, which affect collectively the others. The multiplication of the Col values with cumulative positive net TE values of residues reveals powerful effectors collectively affecting the others with higher entropy signals, which we call as *TECol score*.

2.2.3. Degenerate and Nondegenerate GNM Modes

Degenerate and nondegenerate modes are determined for the test proteins which have structural symmetry such as homodimers, homotrimers, homotetramers, homohexamers, etc. Degeneracy analyses for these cases are performed by calculating the MSF (Mean Square Fluctuation) of residues in each individual GNM mode. The MSF values of the C^{α} atoms in the individual mode k, are obtained from the diagonal elements of $\mathbf{\Gamma}^{-1}$ using the definition of [46]

$$< \Delta \boldsymbol{R}_k{}^2 > = [\boldsymbol{\Gamma}^{-1}]_{kk} \,. \quad (2.9)$$

In these individual mode calculations; if the MSF shape of repating units are not the same with each other, i.e. not symmetrical, this individual mode is recorded as degenerate. Usually, two successive degenerate modes complement each other, revealing the symmetrical behavior of the protein together. On the other hand, nondegenerate modes by itself can fully reveal the structurally repetitive behavior of the protein, thus MSF shapes of the repetitive units are identical in an nondegenerate slow mode calculation.

The dissection of degenerate and nondegenerate modes of the dataset proteins with structural symmetry are summarized in Table 3.2 in the Results and Discussion Section, and the related MSF plots of these proteins are provided in Appendix A.

Information about active or allosteric sites is carried by both nondegenerate and degenerate slow modes. As recently demonstrated, degenerate slow modes may also be functionally significant in addition to nondegenerate slow modes that are necessary for symmetric functional patterns of complex proteins with different oligomerization states [48].

2.2.4. Identifying Subsets of Slow Modes Maximizing Information Transfer

To determine the collective capacity of residues in information transfer, and to develop a general protocol, collectivity expressed in Equation 2.7 is utilized as a dynamic measure. The procedure applied in this developed protocol is demonstrated in Figure 2.2. After the mode decomposition step, we have two -*simple and combinatorial*- search schemes for the subsets of slow modes of likely functional importance; the five basic subsets of slow modes (i.e., n-N_s, with n being the slow modes of 1 to 5) and the subsets of slow mode combinations (i.e., three-to-five slow mode combinations), where N_s represents the number of slow modes in each case having the major contribution to the overall dynamics, which is determined according to the degeneracy analysis. Furthermore, N_s ranging from 8 to 11 slow modes over 23 structures, takes the degeneracy for the cases with fold symmetries into account (See Table 3.2 in the Results and Discussion Section). Around ten slow modes on average are considered to represent the slow end of the dynamic spectrum [49,50].

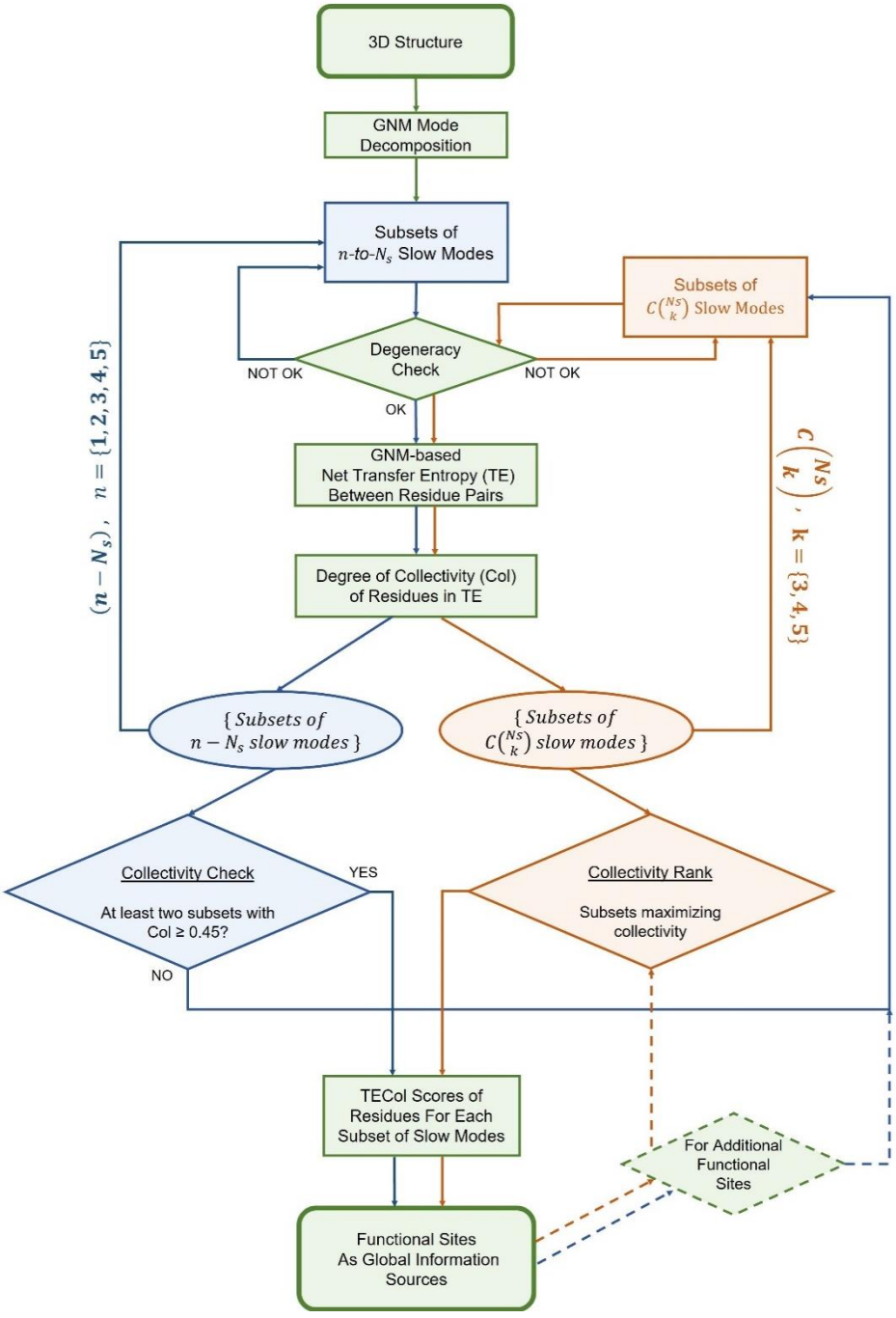

Figure 2.2. Method Flow Chart.

In the flow of this general protocol, it is started with the simple search scheme emphasizing the contribution of slowest/slower modes in these subsets as simply selected from the N_s slow modes. If this search scheme does not yield residues at least two subsets of slow modes with collectivities of residues higher than a threshold value (i.e., 0.45) or if additional subsets of slow modes with high collectivity in terms of transfer entropy are

desired, the combinatorial search scheme is followed. Then, the residues with maximum TECol score values are identified from these subsets of slow modes by both of these search schemes as global information source residues.

Using either the simple or combinatorial search scheme yields more than one subset of slow modes with which unique functional sites or the same functional sites in a different dynamic context could be disclosed. The proposed protocol is applied to a dataset of 23 proteins, and we have shown that active and allosteric sites are distinguishable based on their capacity to disseminate information to others collectively as global information sources.

2.2.5. Statistical Significance Analysis

The statistical significance of each functional site prediction is examined using random sampling. First, the average 3D distance between the TECol score peaks above the peaks' average and the known active/allosteric binding residues are calculated. Secondly, the average 3D distances are calculated for 10,000 randomly selected sample sets having the same number of residues as the number of those TECol score peaks. Then, Z score is calculated using the definition of

$$Z = {}^{(X-\mu)}/_{\sigma} \tag{2.10}$$

where X is the mean distance between the above-average TECol score peaks and the binding residues. μ and σ are the mean and the standard deviation of the population of random sampling, respectively. After calculating Z scores of each prediction, *p*-values are obtained by two-tailed hypothesis test using a significance level of 0.05.

The performance of the functional site predictions is also measured in terms of sensitivity (SN), specificity (SP), precision (PRE), and accuracy (ACC) with the formulations of

$$SN = TP/(TP + FN) \tag{2.11}$$

$$SP = TN/(FP + TN) \tag{2.12}$$

$$PRE = TP/(TP + FP) \tag{2.13}$$

$$ACC = (TP + TN)/(TP + FP + TN + FN) \tag{2.14}$$

where the number of true positives, the number of false positives, the number of true negatives, and the number of false negatives, are represented by TP, FP, TN, and FN, respectively. For each functional site prediction, TP corresponds to the number of above-average TECol score peaks -in that subset of slow mode- whose 3D distances to the known active/allosteric ligand binding residues are below or equal to the threshold distance, while FP corresponds to the number of the ones with 3D distances above the threshold. On the other hand, FN and TN correspond to the numbers of below-average TECol score peaks with 3D distances less and higher than the threshold, respectively. These numbers are determined using two different threshold distances, 7 Å and 10 Å.

3. RESULTS AND DISCUSSIONS

Aspartate transcarbamoylase (ATCase), Na^+/K^+-adenosine triphosphatase (Na^+/K^+-ATPase), and human transient receptor potential melastatin 2 (TRPM2) are used as three exemplary cases to illustrate the prediction method proposed in this thesis. Furthermore, the same strategy is applied for a test set composed of 20 allosteric proteins given in Table 2.1.

3.1. Three Exemplary Cases of the Prediction Method

3.1.1. Aspartate Transcarbamoylase (ATCase)

ATCase catalyzes the synthesis of N-carbamoyl-L-aspartate with inorganic phosphate from carbamoyl phosphate and L-aspartate in many prokaryotes, including Escherichia coli [51]. It is a system that has received extensive study and is a prime example of allosteric control and cooperativity [52-54]. It also holds relevance as a potential anticancer target for investigations on allosteric drug design because of the homology it has with human ATCase [55]. ATCase works as a heterododecamer composed of six catalytic subunits arranged in two trimers with three active sites at their interfaces [56] and six regulatory chains put together in three dimers [57]. Nucleotides (such as ATP, CTP, and UTP) serve an allosteric effect by binding to the regulatory subunits whereas carbamoyl phosphate or L-aspartate bind to the active site. It is strictly controlled to keep the purine/pyrimidine ratio in cell stable. This is maintained by the interaction of the tense (T) and relaxed (R) states, by which ATP induces activity by maintaining the R state and CTP and UTP inhibit function by giving priority to the T state. ATP and CTP bind to the same residues while UTP binds to a distinct but nearby position.

N-Phosphonacetyl-L-Aspartate (PALA), a well-known inhibitor of ATCase, functions by binding to the active site on the catalytic subunits and stabilizing the R-state [58]. The allosteric modulators ATP and CTP have a significant impact on the PALA's binding affinity [59]. Here, the PALA-bound R-state structure (PDB: 1D09 [51]) is used in the GNM-based TE computations.

Collectivities and TECol score values of residues in information transfer are calculated for five subsets of slow modes, and their distributions are illustrated in Figure 3.1.

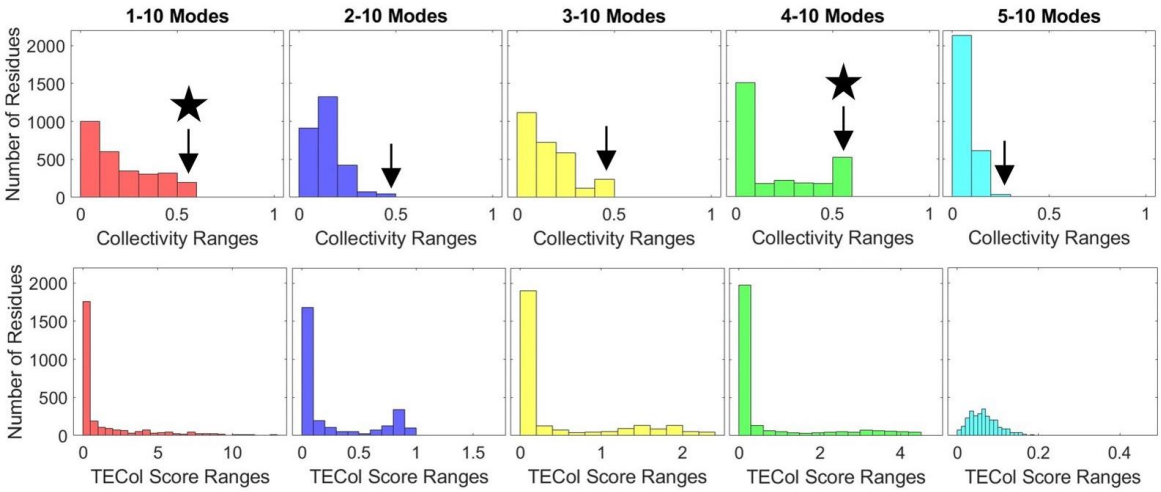

Figure 3.1. Collectivity and TECol score distributions for ATCase (PDB: 1D09).

The collectivity distributions of the TE values of residues in five subsets of slow modes show that the 1-10 and 4-10 slow modes maximize the collective information transfer of residues.

As shown in Figure 3.2 below, where all colored ribbon representations are from the highest (red) to lowest (blue) TECol score values, in the subsets of 1-10 and 4-10 slow modes, the residues with high TECol score values coincide with the active PALA (green) and allosteric ligand ATP/CTP (cyan) binding residues. That is to say, when all first ten slowest modes (1-10) are taken into account, the active ligand -PALA- binding residues seem to be the most collective/global information source residues in terms of information transfer. The allosteric ligand -CTP- binding residues, however, acquire the entire dynamic capacity for the collective information transfer in the 4-10 slow modes with the exclusion of the slowest to third slowest modes.

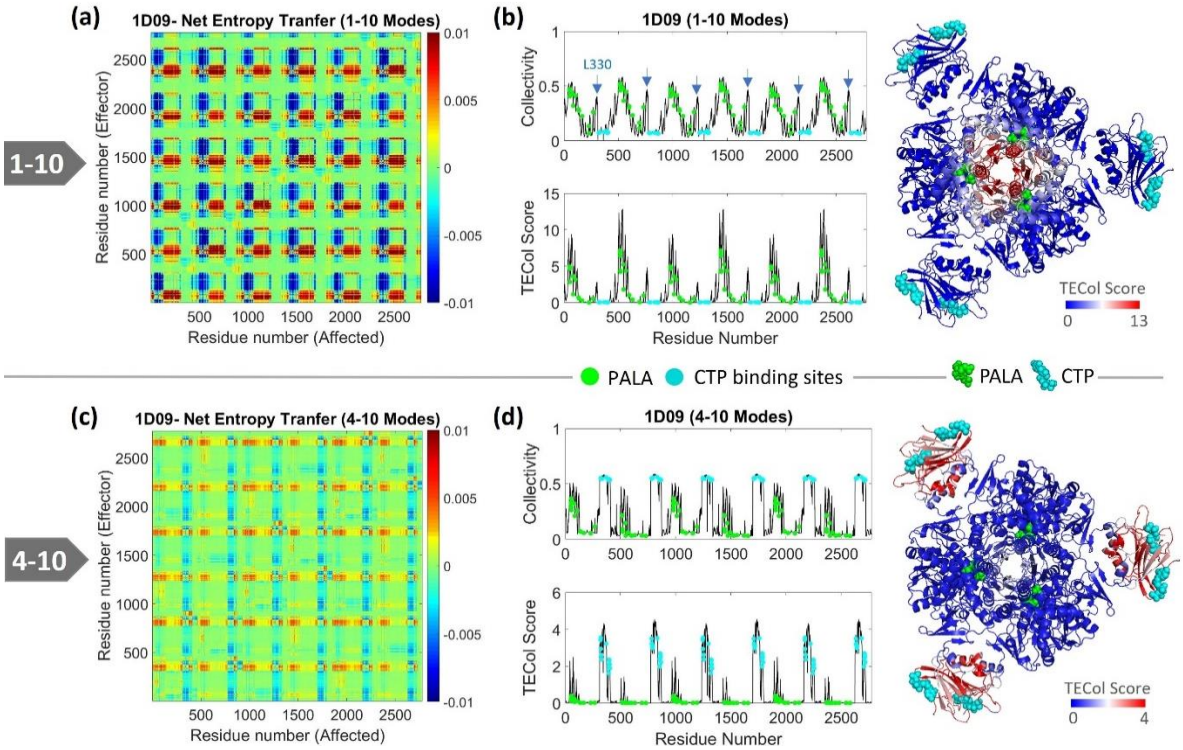

Figure 3.2. ATCase (PDB: 1D09). Net Transfer Entropy (TE) maps (a,c) and the corresponding collectivity and TECol score plots (b,d) in the subsets of 1-10 and 4-10 slow modes. The allosteric ligand CTP is taken from the CTP-bound T state structure (PDB ID: 1RAC).

The first three slowest modes in this case represent the prepotent function of active sites, whereas slow modes four through ten denote potential stimulability upon a perturbation (ligand binding, post-translational changes, etc.), leading particular slow modes to disappear, emerge, and/or rearrange. The stability of the R-state structure by the PALA binding is an example of how the active site residues may be anticipated to cause a stronger regulation in the dynamics within structure in this conformational state.

Different global information source residues with high collectivity but smaller TECol scores than active site residues in the subset of 1-10 slow modes are revealed with the C-terminal end of Helix 12 (A298-N305) in the catalytic subunit. According to previous research, the stability of the trimer and its interactions with the regulatory subunits are both affected by mutations at this region of the helix [60]. As a result, information source residues

identify also other parts with potential functional roles in a protein structure, such as the stability of quaternary structure, in addition to active/allosteric ligand binding sites.

3.1.2. Na^+/K^+-Adenosine Triphosphatase (Na^+/K^+-ATPase)

Na^+/K^+-ATPase transports three Na^+ ions in exchange with two K^+ ions through the cellular membrane by the ATP hydrolysis [61]. This needs an allosteric interplay between the intracellular ATP and transmembrane Na^+ and K^+ binding residues as well as the translocation tunnel. It is a key protein for several physiological activities to maintain concentrations of these ions across the cell membrane. It also regulates the energy-consuming action against the ion gradient.

Na^+/K^+-ATPase is composed of three different subunits which are the main catalytic α-subunit and two smaller subunits, β and γ. To carry out its function, it cycles between the two main conformations E1 and E2. E1 binds three Na^+ ions and ATP in the cytoplasmic part, releases these ions upon phosphorylation to the outside, and switched over to E2. Two K^+ ions are bound by E2 and dephosphorylated after binding them in the extracellular area. The E2 to E1 conformational transition that releases trapped K^+ ions into the cell depends on ATP hydrolysis [62]. During this conformational change, the β-subunit experiences considerable changes compared to the α and γ subunits [44].

The crystal structure of Na^+/K^+-ATPase in the E1 comformation with bound ADP and Na^+ (PDB: 4HQJ [44]) is used in the GNM-based TE computations. Collectivities and TECol score values of residues in information transfer are determined for five subsets of slow modes, and their distributions are illustrated in Figure 3.3.

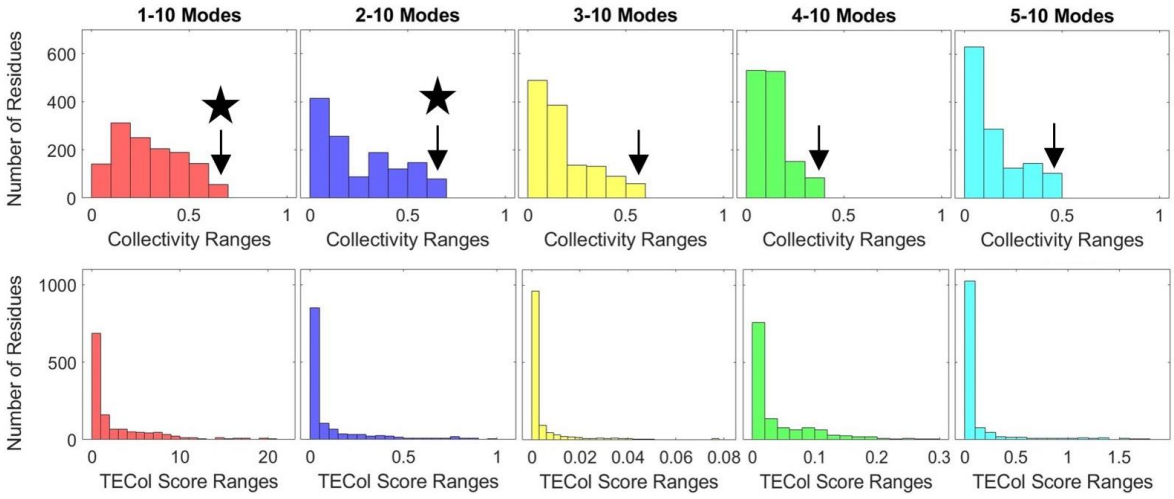

Figure 3.3. Collectivity and TECol score distributions for Na^+/K^+-ATPase (PDB: 4HQJ).

The collectivity distributions of the TE values of residues illustrate that the subsets of 1-10 and 2-10 slow modes maximize the collective information transfer of residues.

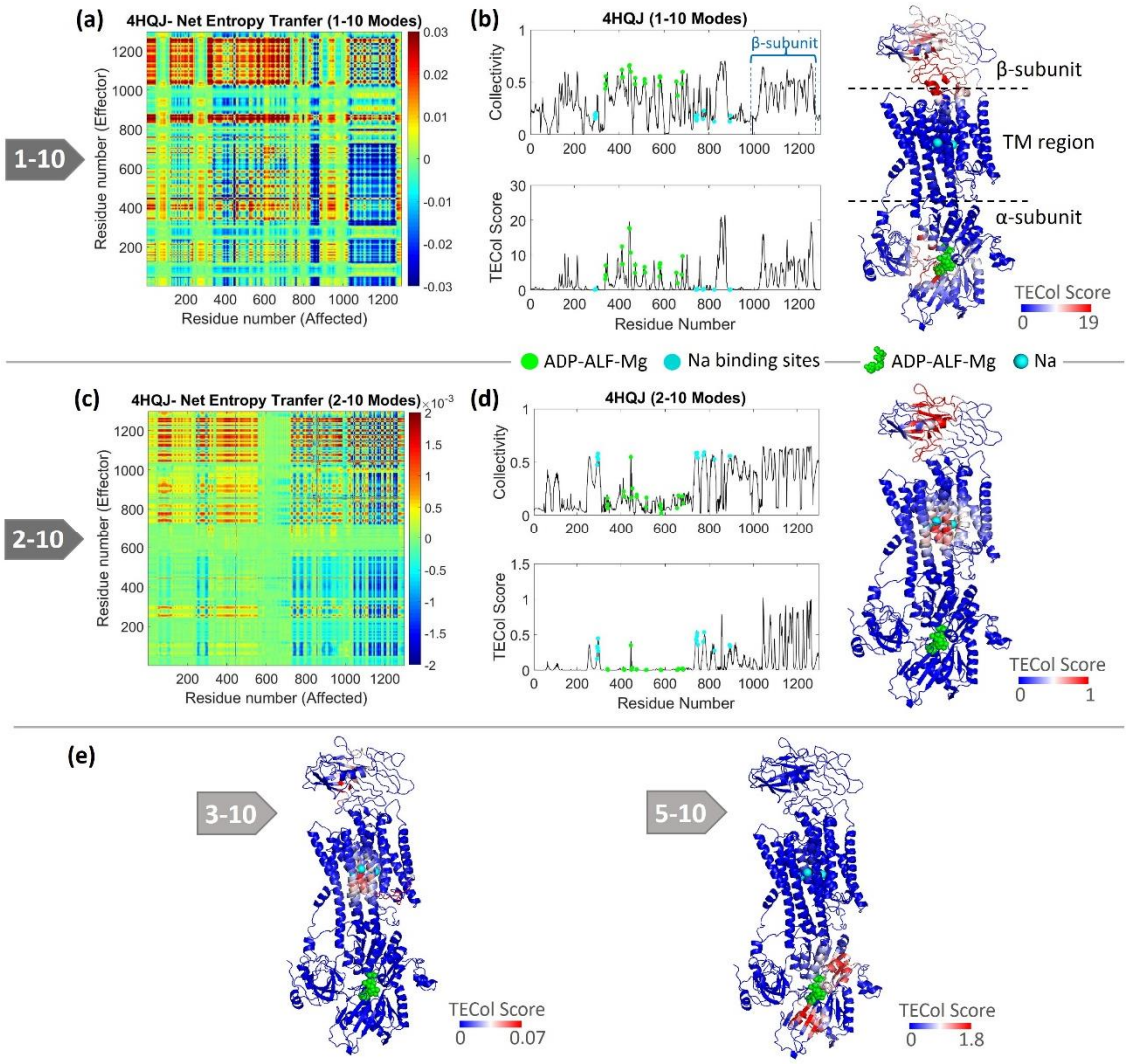

Figure 3.4. Na^+/K^+-ATPase (PDB: 4HQJ). Net Transfer Entropy (TE) maps (a,c) and the corresponding collectivity and TECol score plots (b,d) in the subsets of 1-10 and 2-10 slow modes. Coloured representations of TECol score values in the subsets of 3-10 and 5-10 (e).

As shown in Figure 3.4 -where all colored ribbon representations are from the highest (red) to lowest (blue) TECol score values as well as in Figure 3.2- in the subsets of 1-10 and 2-10 slow modes, the residues with high TECol score values appear at the ADP, tetrafluoroaluminate ion (ALF) and Mg^{2+} (green) and allosteric Na^+ (cyan) binding sites. The 3-10 and 5-10 slow modes likewise point to the allosteric and active sites, respectively, with comparable collectivities of the TE values. Thus, it can be said that the subsets of 1-10 and 2-10 slow modes, respectively, mask the behaviour of latter susbets of slow modes.

It is interesting to note that the β-subunit also exhibits highly collective residues in the TE values, especially more significantly in the 2-10 slow modes. This supports the idea that the β-subunit regulates ion binding on the α-subunit and stabilizes the Na^+-occluded E1-P state [63]. In particular, the β-subunit contains asparagine aminoacids that post-translationally bind oligosaccharides and are associated with cell-specific activity [64]. In the subset of 2-10 slow modes, two of these oligosaccharide binding residues -N158 and N193- appear as global information sources (Figure 3.4d). Also, the 2-10 slow modes reveal the cysteine residue pairs C159-C175 and C213-C276 forming disulphide bridges as collective information sources, while the latter is also revealed in the 1-10 slow modes. The assembly of the α-β subunits depends significantly on these disulphide bonds and oligosaccharide binding regions [65]. All of these are distant from the interface, yet possibly having an allosteric role in stabilization. Nevertheless, in the subset of 5-10 slow modes, the signalling effect of the β-subunit is replaced by the collective information transfer of the ADP-ALF-Mg^{2+} binding sites which appear as global information sources. This highlights the influence of the first four slowest modes on the allosteric communication led by the β-subunit.

Around E818 is another remarkable information source residue that has the highest collectivity in the information transfer in the 2-10 slow modes. A human mutation at this location has been demonstrated to impair proton transport and the E1>E2 conformational transition, resulting in severe alternating hemiplegia of childhood (AHC) disease [66]. The dynamic explanation provided by the GNM-based TE compoutations may help to identify the unidentified allosteric mechanism underlying this mutation.

3.1.3. Human Transient Receptor Potential Melastatin 2 (TRPM2)

TRPM2 is a nonselective Ca^{2+}-permeable cation channel on the cellular membrane activated by heat, redox signals and/or chemical binding of ADP-ribose (ADPR) and Ca^{2+}. It is a homotetramer with three tiers stacked on top of one another. ADPR is bound to the bottom pier and Ca^{2+} is bound to the top pier while a middle tier is in between. TRPM2 undergoes a significant conformational transition from closed to open state in order to permeate Ca^{2+} ions through the allosteric interaction between ADPR and Ca^{2+} [43]. The function of TRPM2 is crucial for various cells to manage oxidative stress and control

temperature [43]. However, the precise molecular process by which human TRPM2 is activated remains unknown.

The unbound structure of human TRPM2 in closed conformation (PDB: 6MIX [43]) is used in the GNM-based TE computations. Collectivities and TECol score values of residues in information transfer are determined for five subsets of slow modes, and their distributions are provided in Figure 3.5.

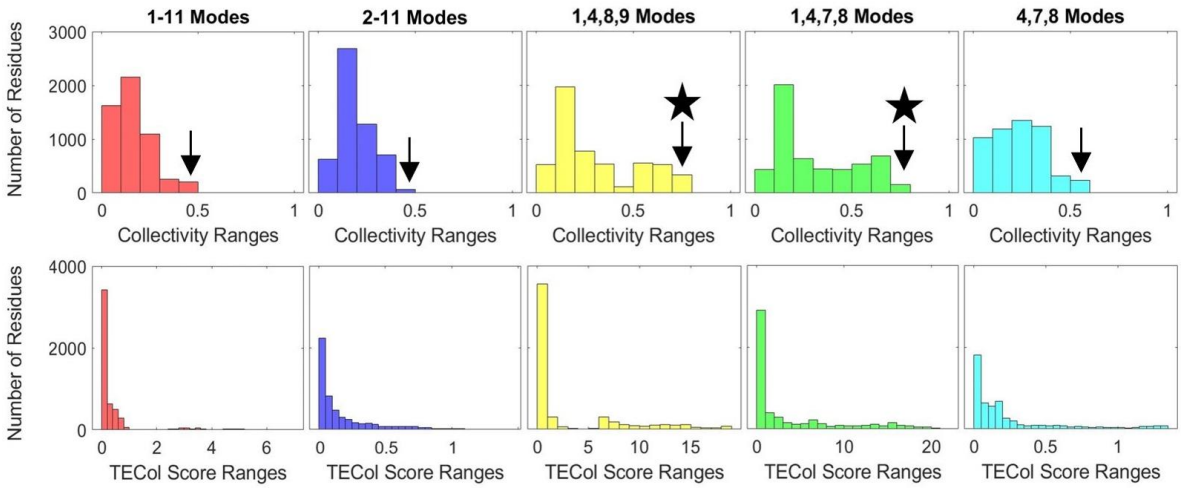

Figure 3.5. Collectivity and TECol score distributions for human transient receptor potential melastatin 2 (TRPM2; PDB: 6MIX).

The collectivity distributions of the TE values of residues show that the subsets of 1-11 and 2-11 slow modes maximize the collective information transfer (among other simple subsets).

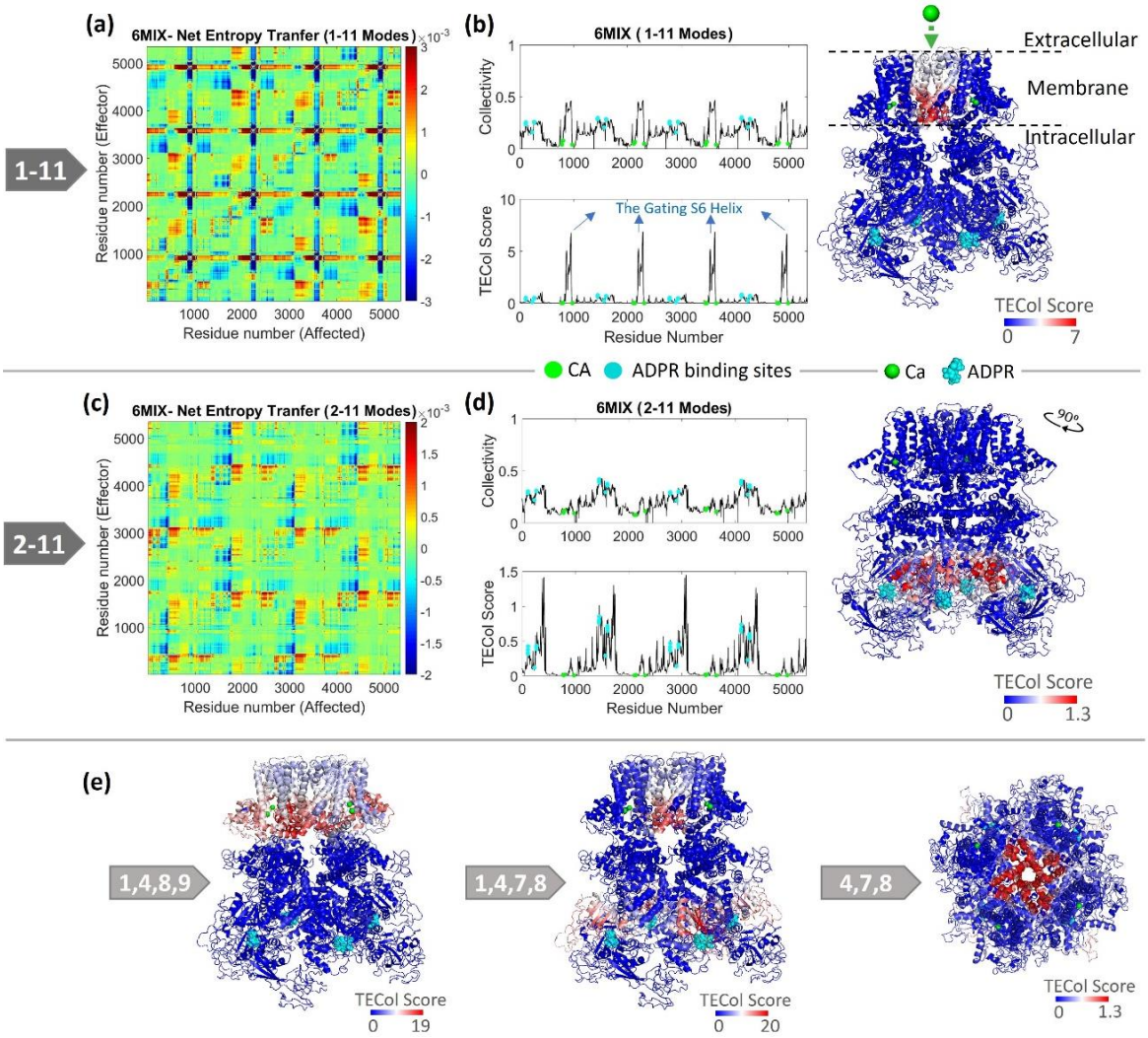

Figure 3.6. TRPM2 (PDB: 6MIX). Net Transfer Entropy (TE) maps (a,c) and the corresponding collectivity and TECol score plots (b,d) in the subsets of 1-11 and 2-11 slow modes. Coloured representations of TECol score values in the subsets of 1,4,8,9 , 1,4,7,8 and 4,7,8 (e).

As seen from Figure 3.6, the gating S6 helix -whose rotation is the key for the channel opening- is appeared as global information sources in the 1-11 slow modes, while the ADPR binding sites (cyan) -whose allosteric effect on the Ca^{2+} binding leads the gating S6 helix to rotate- are revealed to be collective information transfer residues in the subset of 2-11 slow modes. Despite the fact that both discovered sites are functionally important, the binding residues of Ca^{2+} (green) are absent in these five simple subsets of slow modes.

Additionally, there is a slight distortion of the symmetry in the TE patterns in such a symmetric structure although the degenerate modes of 10 and 11 having the same eigenvalues are taken together. Both of these need further analysis of the mode combinations to designate other possible slow mode subsets having higher collectivities in the information transfer than the five simple subsets of slow modes.

Indeed, it is determined by the combinatorial search scheme that, as shown in Figure 3.5, the 1,4,8,9 and 1,4,7,8 slow modes are the best slow mode combinations maximizing the collective information transfer of residues. Ca^{2+} binding sites (green) appear as collective information sources in the 1,4,8,9 slow modes while the ADPR binding sites (cyan) along with the cytoplasmic gate of the channel are disclosed in the 1,4,7,8 slow modes.

Moreover, the channel pore stretching from the extracellular to intracellular side of the top tier takes high TECol score values in an additional subset of 4,7,8 slow modes. Zn^{2+}, an allosteric modulator of TRPM2 inactivates the protein by binding to residues in the outer pore [67]. This exemplifies that already known different functionally important sites can also be identified with the proposed fine-tuned combinatorial search scheme strategy. Thus, with more subsets of slow modes with strong collectivity in information transfer, a new additional functional site may be discovered. For example, the cytoplasmic gate and ADPR binding sites, which appear separately in the 1-11 and 2-11 modes respectively, appear together in the 1,4,7,8 slow modes.

3.1.4. Directional Allosteric Pathways in ATCase, Na^+/K^+-ATPase and TRPM2

Global information source residues, which collectively affect many other residues in the protein structure, direct allosteric communication pathways within biological macromolecular machines. Thus, to be able to clarify allosteric interplay within the three exemplary cases (ATCase, Na^+/K^+-ATPase, and TRPM2), the information flow (net TE) from the strongest global information source residue to the others is determined and represented in Figures 3.7-9 that are color-coded from the highest (red) to the lowest (blue) Net TE values with yellow dash arrows displaing its direction.

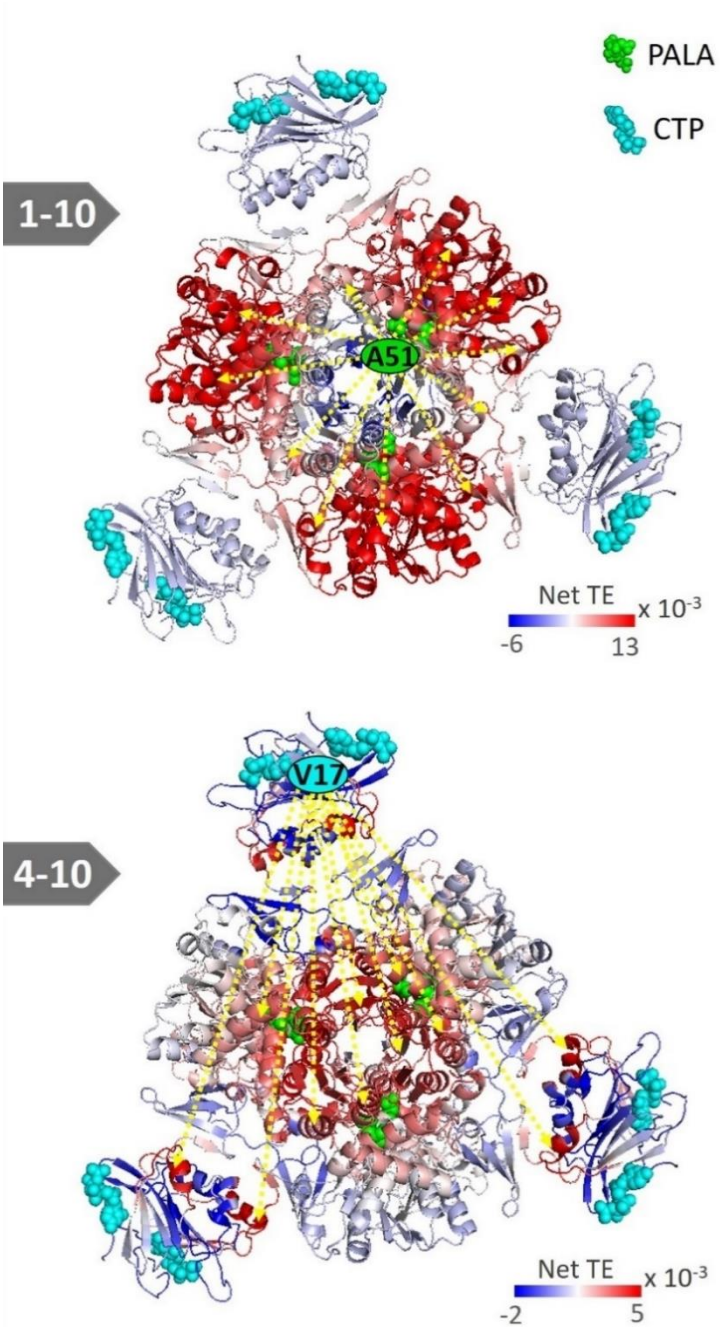

Figure 3.7. Directional allosteric pathways in ATCase. From the active ligand PALA binding residue A51 in the 1-10 slow modes and from the allosteric ligand CTP binding residue V17 in the 4-10 slow modes to the other residues, mainly to the catalytic sites of other subunits and to the active sites of all subunits, respectively.

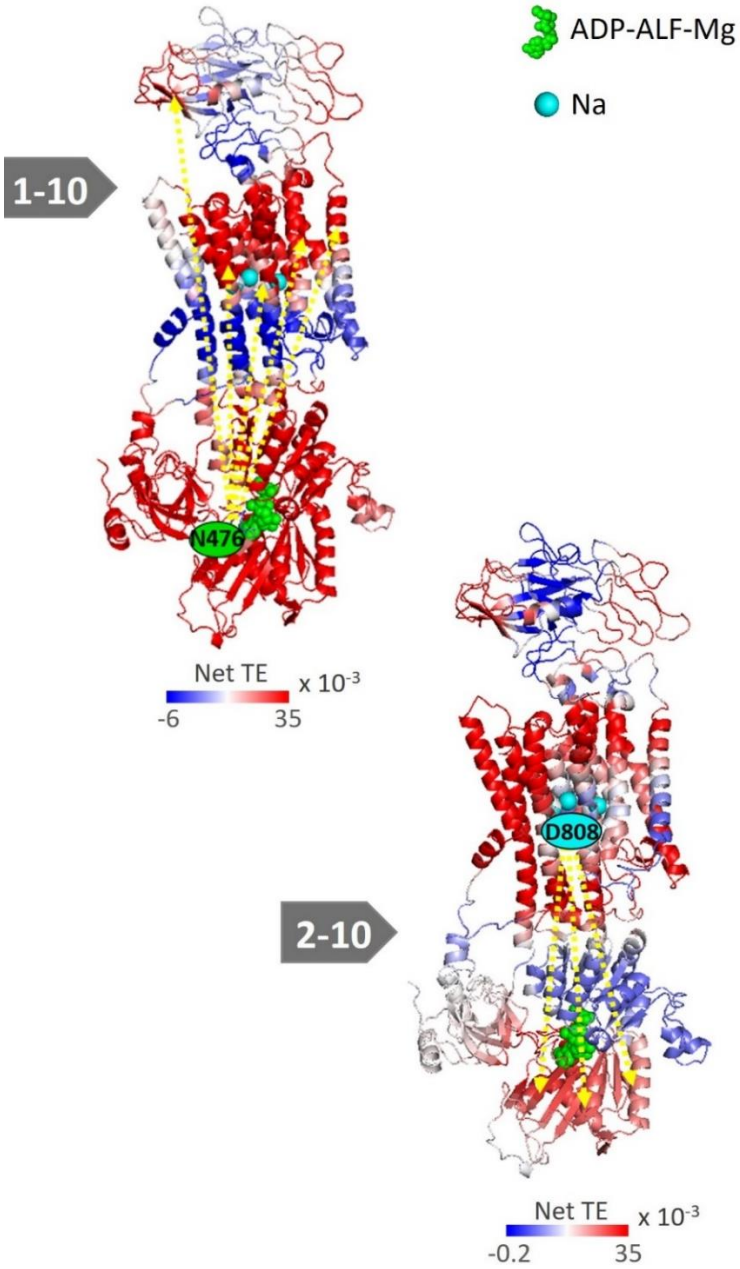

Figure 3.8. Directional allosteric pathways in Na^+/K^+-ATPase. From the active binding residue N476 in the 1-10 slow modes and from the allosteric ADP-ALF-Mg binding residue D808 in the 2-10 slow modes to the other residues, mainly to the allosteric and to the active sites, respectively.

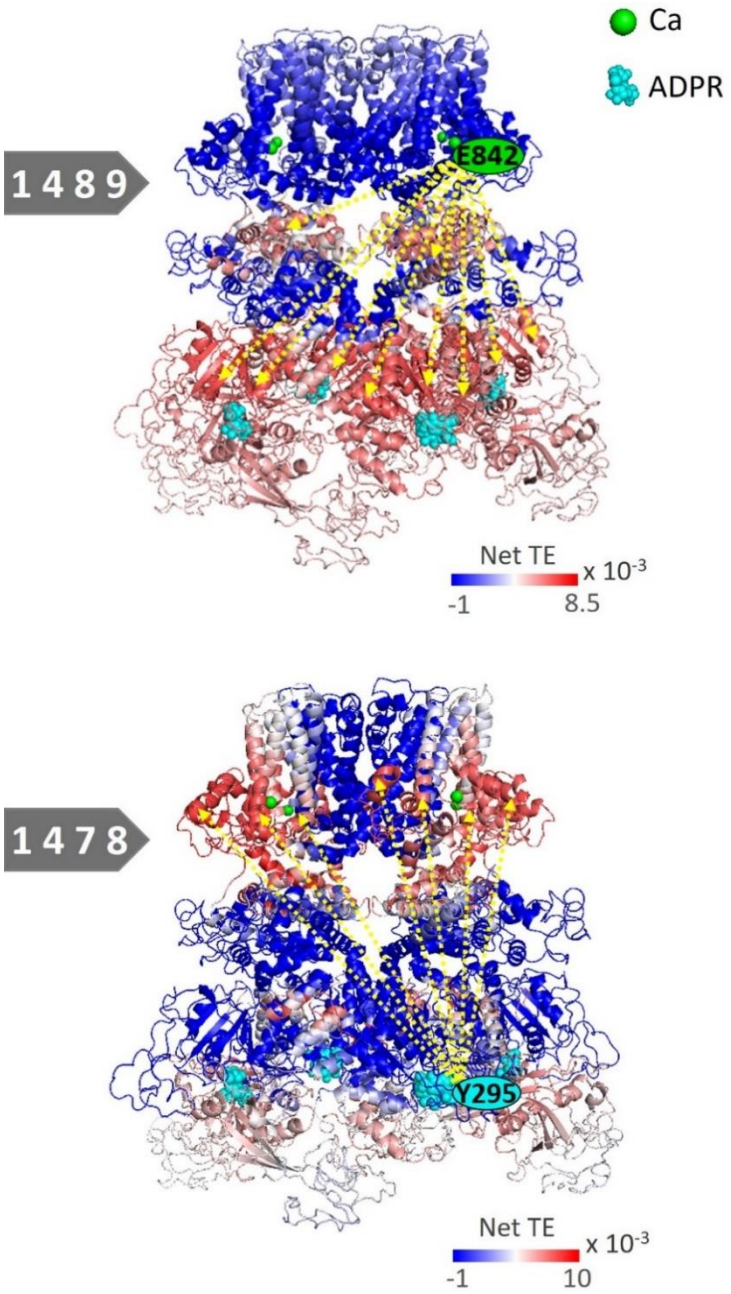

Figure 3.9. Directional allosteric pathways in TRPM2. From the Ca^{2+} ligand binding residue E842 in the 1,4,8,9 slow modes and from the ADPR binding residue Y295 in the 1,4,7,8 slow modes to the other residues, mainly to the ADPR and to the Ca^{2+} binding sites, respectively.

Figure 3.7 shows that, based on the PALA bound active structure of ATCase, residue A51 with highest collectivity in the TE values at the active site of one catalytic subunit collectively transfers information to the other catalytic subunits in the 1-10 slow modes. This implies cooperativity among catalytic sites/subunits for the ligand PALA binding in this inhibited R-state. On the other hand, residue V17 with high collectivity in the TE values at the allosteric site collectively transfers information to the active sites in the 4-10 slow modes even in the absence of allosteric ligands. This also provides a dynamic basis for the inhibitory effect of ATP/CTP on the binding affinity of PALA [59]. Notably, this behavior is latent in the 1-10 slow modes in this R state conformation of ATCase.

As illustrated in Figure 3.8 that in Na^+/K^+-ATPase with bound ADP and Na^+, N476 is one of the residues with high collectivity in its TE values of the 1-10 slow modes at the ADP binding site and directs allosteric communications pathways towards the allosteric region close to the NA binding sites and the β-subunit. On the other hand, D808 with high collectivity in the TE values of the 2-10 slow modes at the NA binding site collectively transfers information towards the ADP binding sites and to the β-subunit. That the ADP and NA binding sites exchange directed information shows bidirectionality in their allosteric communications. The altered directions are revealed in two different subsets of slow modes, indicating the importance of mode dissection for layering of allosteric communication pathways.

Figure 3.9 demonstrates that E842 is one of the residues with high collectivity in the TE values of the 1,4,8,9 slow modes at the Ca^{2+} binding region and directs allosteric communications pathways towards the ADPR binding sites in the apo structure of human TRPM2 in closed state. On the other hand, Y295 with high collectivity in the TE values of the 1,4,7,8 slow modes at the ADPR binding sites collectively transfers information mainly towards the Ca^{2+} binding sites. That the ADPR binding sites at the bottom tier and Ca^{2+} binding sites at the top tier exchange directed information shows bidirectionality in the allosteric crosstalk in the close to open transition of the channel to permeate the Ca^{2+} ions.

These three exemplary cases (Figures 3.7-9) show that the simple search scheme is already powerful to reveal the subsets of slow modes maximizing the collective information transfer and active/allosteric sites global information sources. Yet, as in the case of human TRPM2, the combinatorial search scheme may help identify new subsets of slow modes with functional importance and fine-tune the already notable subsets of slow modes.

3.1.5. Effect of Time Delay on Collectivity

The time delay τ is selected based on the maximization of the degree of collectivities of net TE values, which is 3 x τ_{opt}. τ_{opt} is the time window in which total TE of residues is maximized. Figure 3.10 illustrates the effect of different τ values on the collectivities in the TE values of residues for three main exemplary cases.

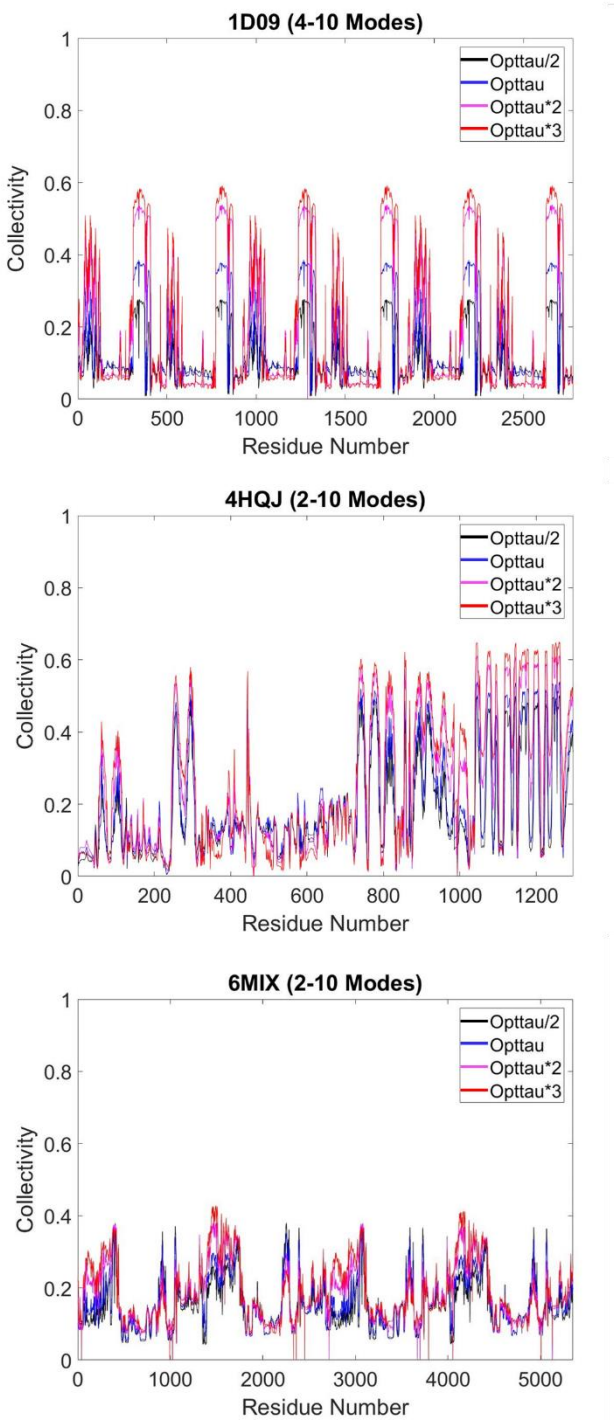

Figure 3.10. Collectivities at different time delays. The abbreviation *opttau* is τ_{opt}, which is the time window in which the total TE values are maximized in the structure. Here, the TE calculations are performed in the folds of τ_{opt} to identify τ that maximizes the collectivities in the residues' TE values.

3.2. Predictions on Allosteric Test Proteins

In addition to the three exemplary cases, the predictions are evaluated on 20 more allosteric proteins given the dataset table (Table 2.1). In the GNM based TE calculations, the active state structures are mostly used without explicitly including active and allosteric ligands or their binding sites information. As presented in Figures 3.11-3.31, the active and allosteric sites in these allosteric proteins are predicted as global information sources.

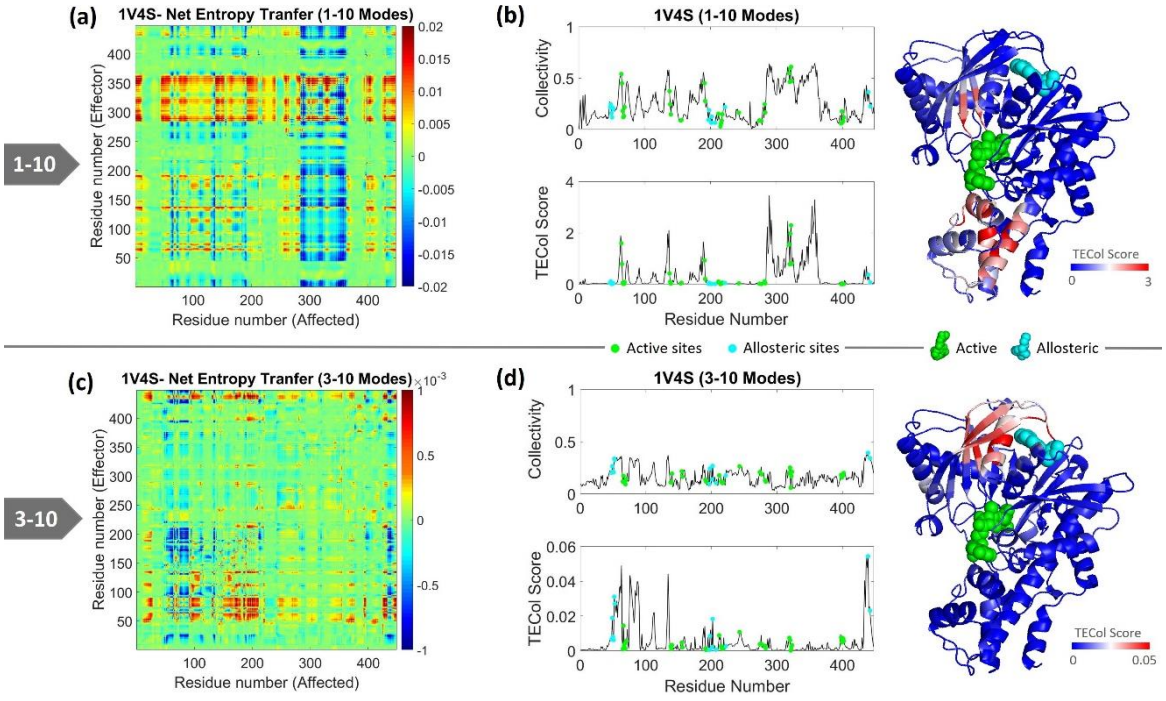

Figure 3.11. Human Glucokinase (PDB: 1V4S [68]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 3-10 slow modes, respectively, for the monomeric closed conformation with the active site ligands glucose-AMP-PNP(ATP)-Mg (green) and the allosteric ligand MRK (cyan). AMP-PNP is taken from PDB structure 3ID8 [69] aligned to 1V4S [68].

As shown in Figure 3.11 that the information source residues are at/around AMP-PNP and MRK binding residues in the subsets of 1-10 and 3-10 slow modes, respectively.

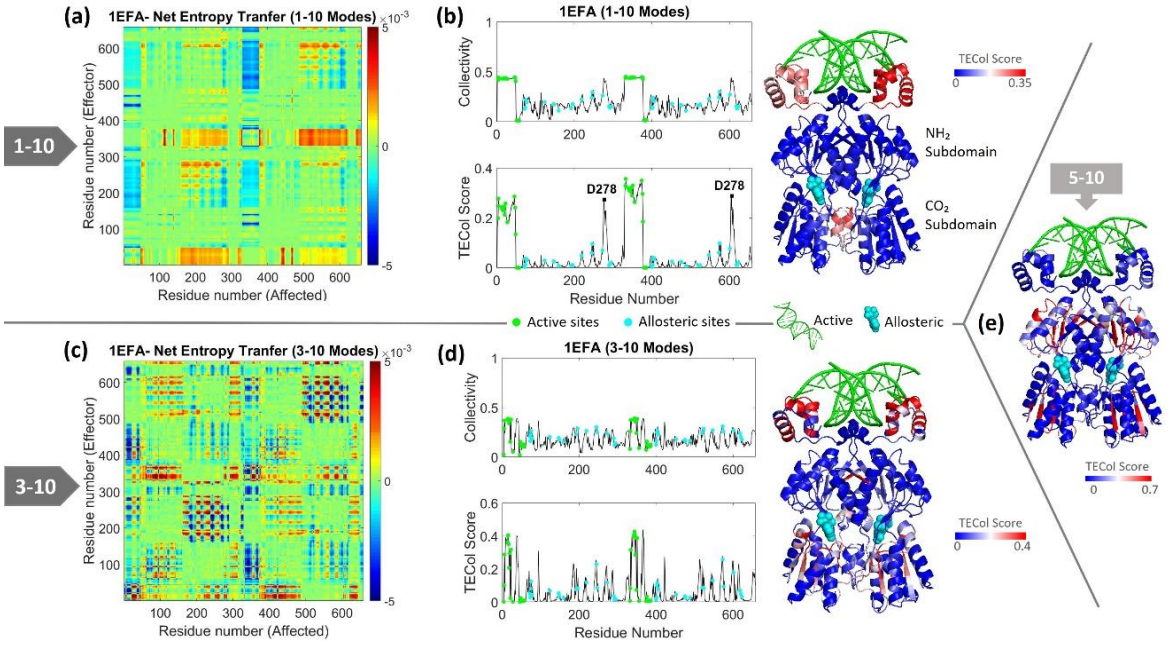

Figure 3.12. Lac Repressor (PDB: 1EFA [70]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 3-10 slow modes, respectively, for the DNA bound dimeric structure with the active site ligand DNA (green) and the allosteric ligand operator NPF (cyan).

The 1-10 slow modes reveal the DNA binding residues as information source residues (Figure 3.12b). Additionally, this subset of slow modes highlights the region including D278, which is stated as the key residue for the interaction between the NH_2-subdomain of one monomer and the CO_2-subdomain of the other in the inducer-bound T state conformation [70]. The DNA binding residues are also information source residues with the region extending to the NPF binding in the 3-10 slow modes (Figure 3.12d). Collective information transfer characteristics along the beta sheets in both subdomains are observed in the 5-10 slow modes (Figure 3.12e), also overlapping with some of the NPF binding residues.

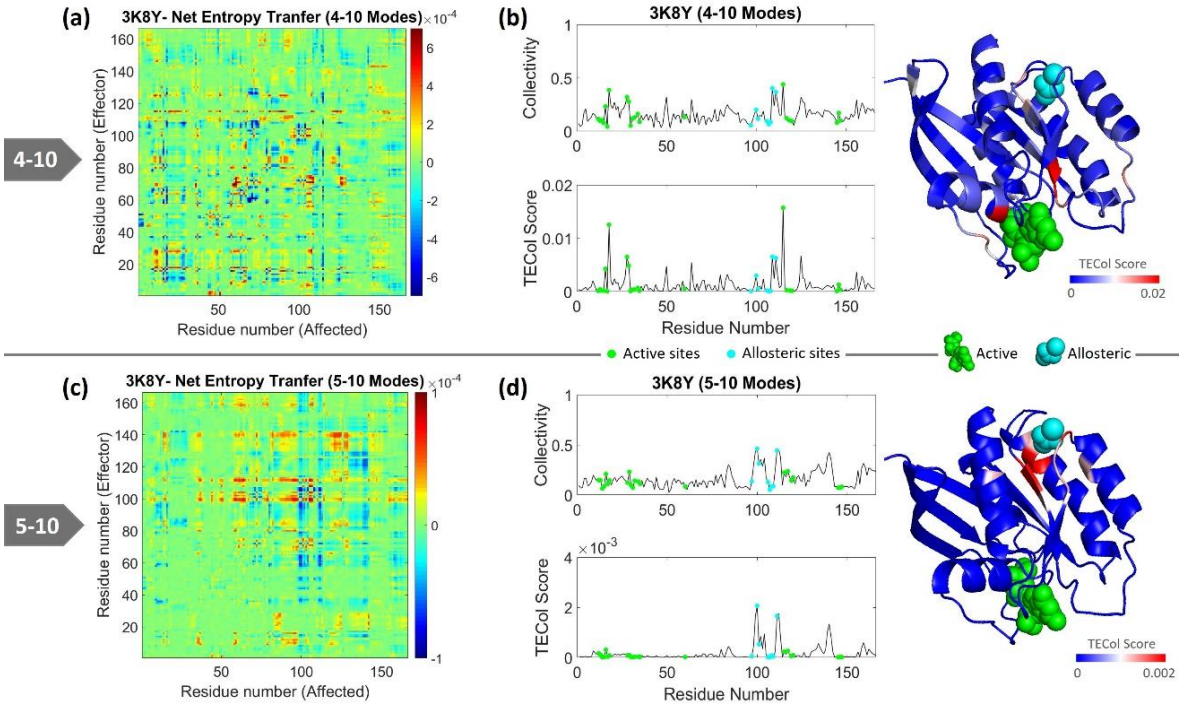

Figure 3.13. H-Ras GTPase (PDB: 3K8Y [71]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 4-10 and 5-10 slow modes, respectively, for the monomeric wild-type Ras with the active site ligand GNP (green) and the allosteric calcium acetate (cyan).

The subset of 4-10 slow modes detects most of the GNP binding residues as information source residues (Figure 3.13b). Besides, the calcium acetate binding residues, albeit slightly, coincide with lower local peaks here. The calcium acetate binding residues become more pronounced than the GNP binding residues when the fourth slowest mode is excluded (Figure 3.13d).

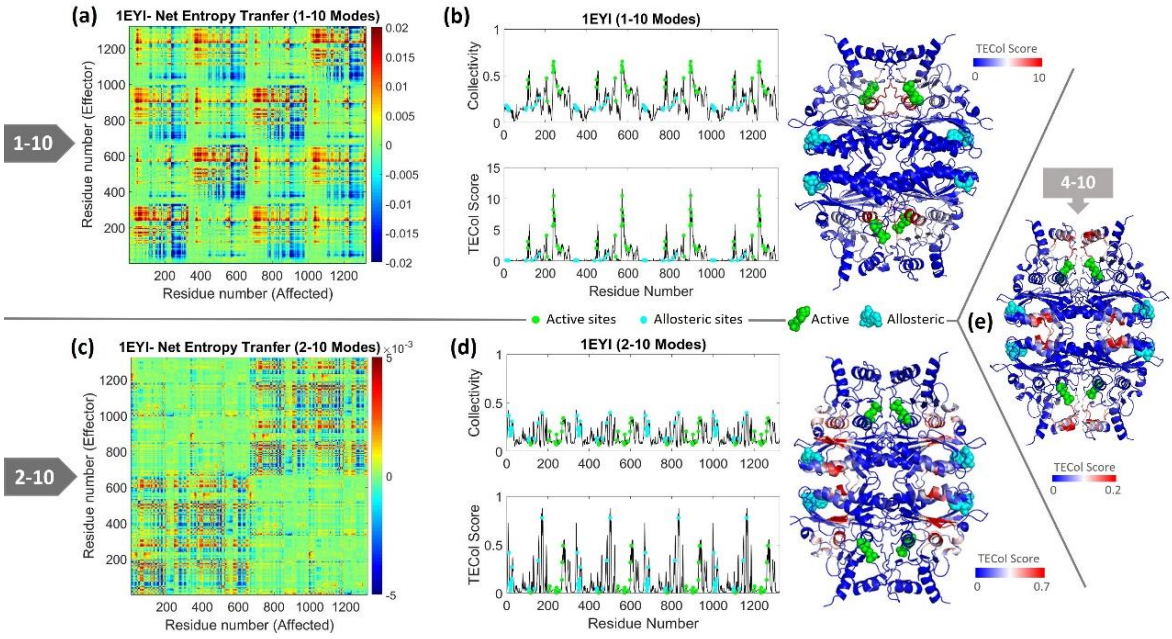

Figure 3.14. Fructose-1,6 bisphosphatase (PDB: 1EYI [72]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively, for the tetrameric R-state conformation with the active site ligand F6P (green) and the allosteric site ligand AMP (cyan) [72]. The allosteric AMP is taken from the T-state conformation (PDB: 1EYJ [72]).

All F6P binding residues and the adjacent loops and helices appear as information source residues in the 1-10 slow modes (Figure 3.14b). Mainly, the AMP binding residues appear as information source residues in the 2-10 slow modes yet some of the active sites also continue to emerge at lower peaks (Figure 3.14d). Interesting to note, although the collectivities remain at moderate values in the 2-10 slow modes, as seen on the 2D net TE maps, it shows quite high collectivities when evaluated based on the individual dimers (Figure 3.14a and c). This implies the collective behavior of the chains in the dimeric units of the tetramers in this topology. On the other hand, the 4-10 slow modes, although not as collective as in the 1-10 slow modes, leads to the emergence of the helix regions approaching both the F6P and AMP binding residues as more pronounced than they are in the 1-10 slow modes (Figure 3.14e).

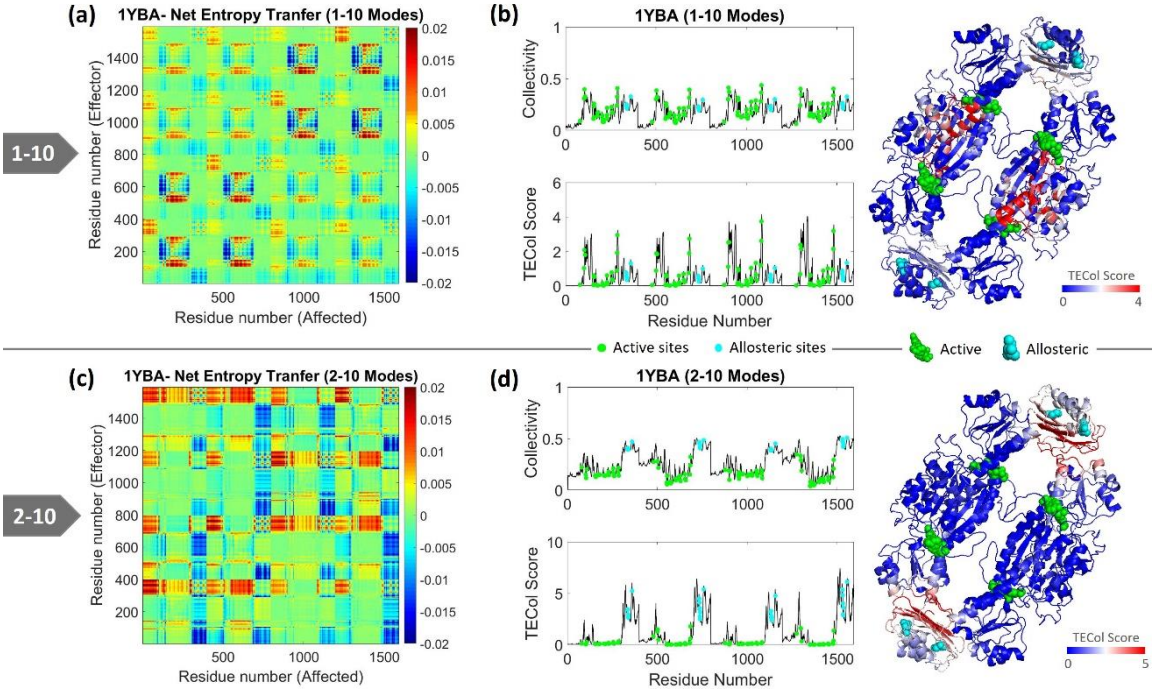

Figure 3.15. Phosphoglycerate Dehydrogenase (PDB: 1YBA [73]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively, for the quaternary active conformation whose subunits all bind the cofactor NAD (green) and the allosteric ligand serine (SER) (cyan). SER is taken from the effector-bound structure (PDB: 1PSD [74]).

In the 1-10 slow modes, the information source residues are most at the NAD binding residues and extending in-between active regions (Figure 3.15b). In the 2-10 slow modes, all SER binding residues are information source residues with significant high TECol score values than the other residues (Figure 3.15d).

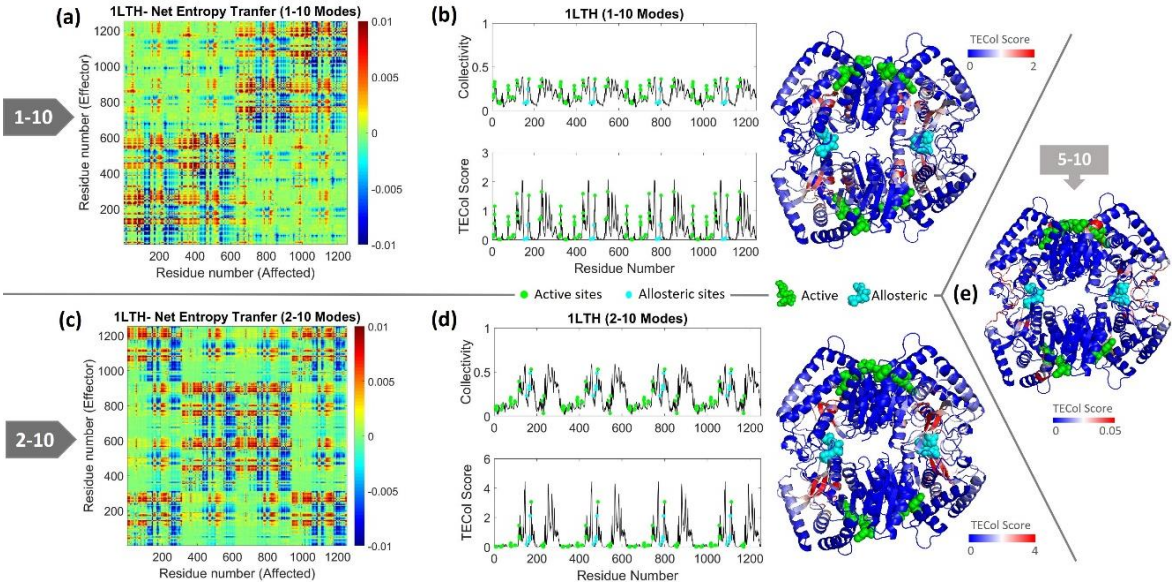

Figure 3.16. Lactate DH (PDB: 1LTH [75]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively, for the tetrameric structure of L-lactate dehydrogenase containing T- and R-state tetramers with the active site ligand NAD (green) and the allosteric ligand FBP (cyan).

In the 1-10 slow modes, the NAD binding residues are information source residues to a very large extent (Figure 3.16b). In the 2-10 slow modes, the FBP binding residues become more pronounced than they are in the 1-10 slow modes while most of the NAD binding residues as information source residues in the 1-10 slow modes reset to the 2-10 slow modes (Figure 3.16d). Interestingly also to note here, the dimeric collectivities of TE emerge in this conformational state of tetramer (Figure 3.16a and c). The ligand binding residues that are not visible in the first two subsets of slow modes are found in the 5-10 slow modes, albeit with relatively lower TECol score values (Figure 3.16e).

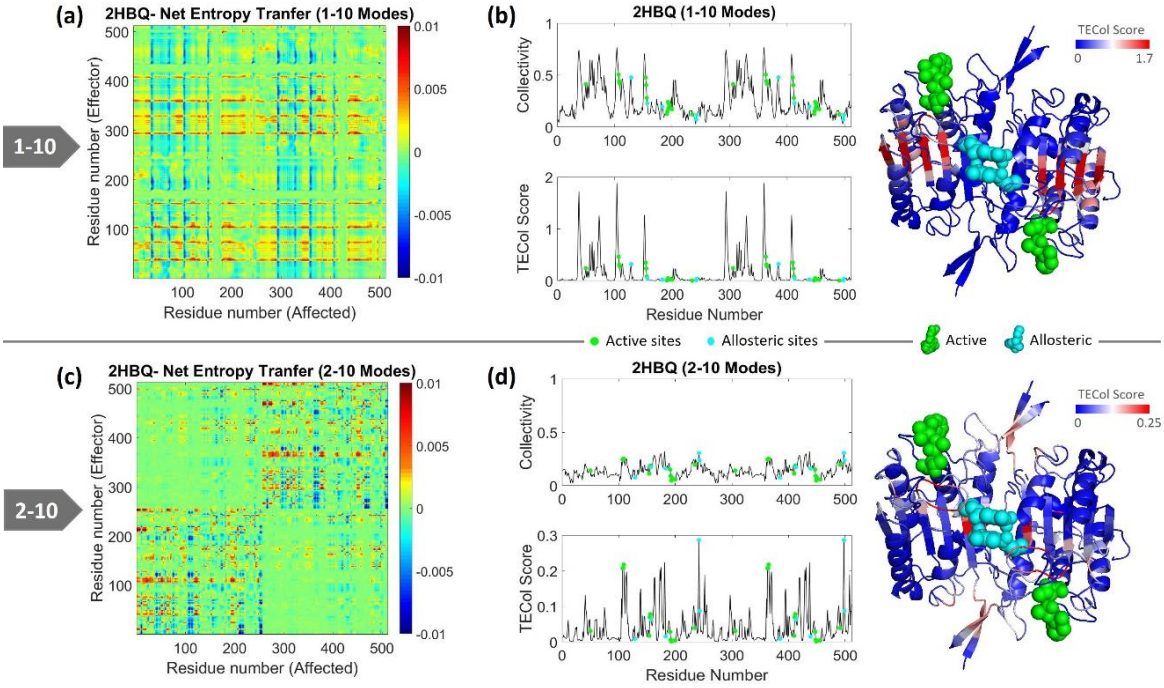

Figure 3.17. Caspase-1 (PDB: 2HBQ [76]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively, for the dimeric structure with the active site peptide z-VAD-FMK (green) and the allosteric ligand F1G (cyan) (PDB: 2FQQ [76]).

The peptide and F1G bindings appear right next to the information source residues, to the peaks of collectivity or TECol scores, which extends towards the ligand binding residues through beta sheets in both subunits in the 1-10 slow modes (Figure 3.17b). However, some of the peptide and FIG binding residues appear as information source residues in the 2-10 slow modes (Figure 3.17d). Interestingly to note here as well, the information source residues in these subsets of slow modes are effective mainly on the monomers as seen from the TE maps (Figure 3.17a and c).

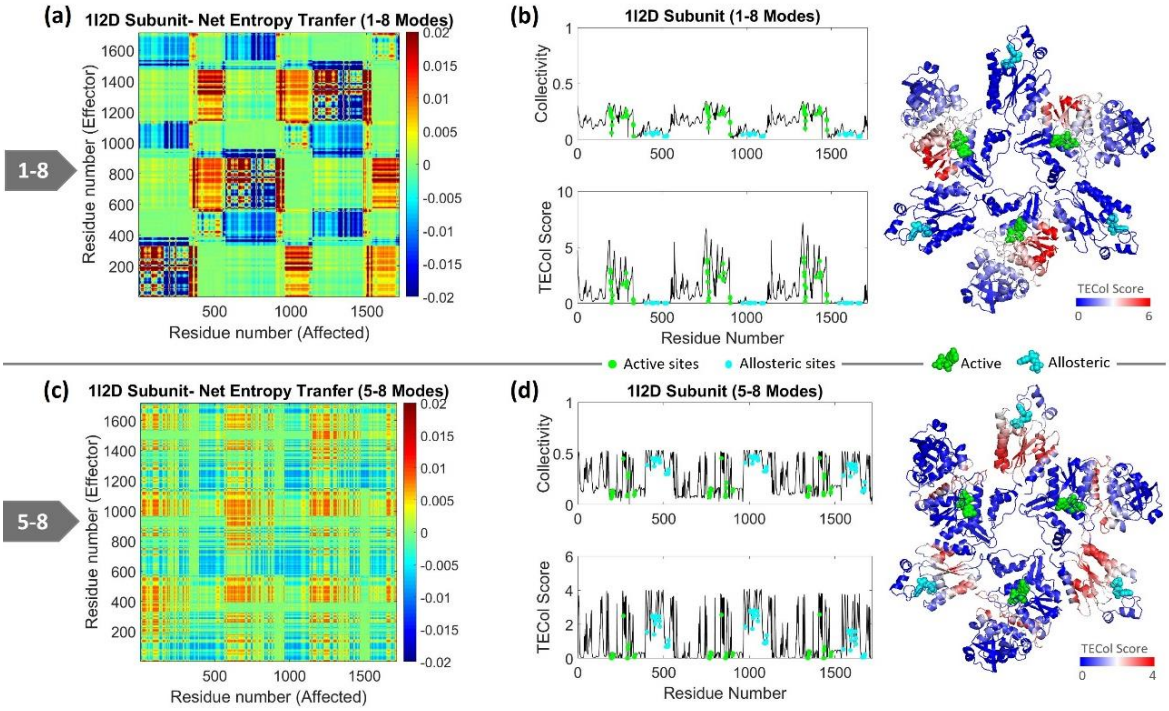

Figure 3.18. ATP Sulfurylase (PDB: 1I2D [77]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-8 and 5-8 slow modes, respectively, for an allosterically regulated hexameric enzyme with the R state conformation with the active site ligand APS (green) of the catalytic domain [77] and the allosteric inhibitor PPS (cyan) which is taken from the T-state conformation of the enzyme (PDB: 1M8P [77]).

The structure of ATP Sulfurylase is made up of two asymmetric units joined together [77]. Due to its asymmetric nature, the functional sites do not appear symmetrically in six monomers (see Figure 3.31b). Thus, here, subunit-based TE is evaluated to eliminate this asymmetry. APS binding residues appear as information source residues in the 1-8 slow modes (Figure 3.18b), and PPS binding residues appear as information source residues in the 5-8 slow modes (Figure 3.18d). Interesting to note here as well, although the collectivities are moderate in the 1-8 slow modes, as seen on the 2D TE map, it yields quite high collectivities when evaluated based on the three individual monomers (see Figure 3.31a).

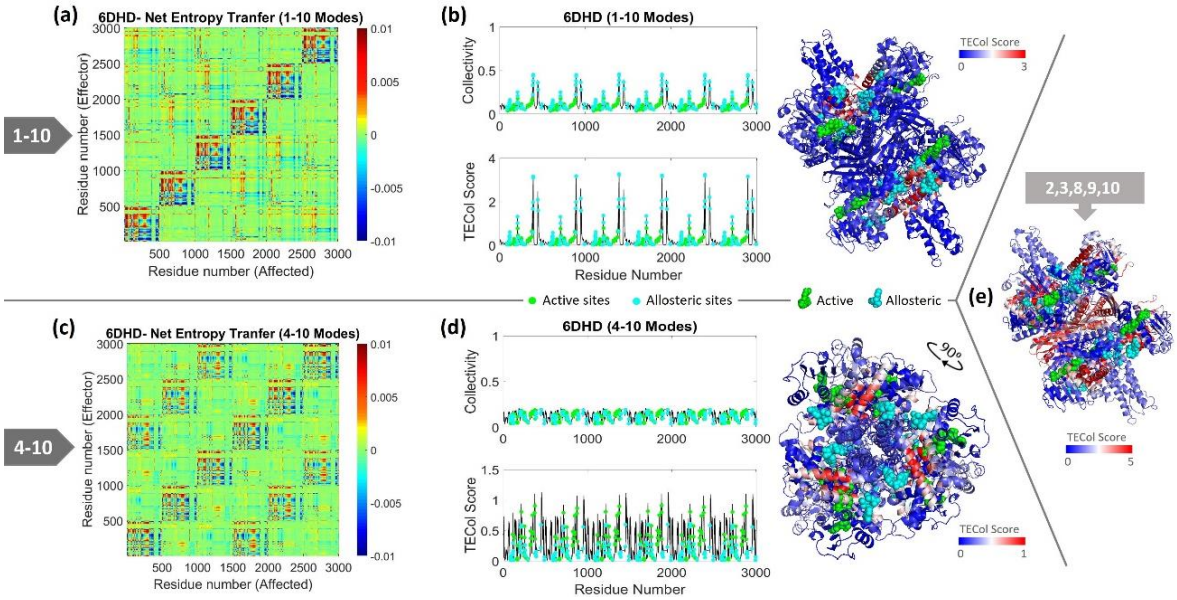

Figure 3.19. Glutamate DH (PDB: 6DHD [78]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 4-10 slow modes, respectivelyi for the hexameric structure of glutamate dehydrogenase with the active site ligand NADH-GLU (green) and in the inhibited state bound to the allosteric ligand NADH-GTP (cyan).

As viewed along the three-fold axis, most of the NADH-GTP binding residues are information source residues in the 1-10 slow modes (Figure 3.19b). Although the collectivities of TEs are relatively low in the 4-10 slow modes (Figure 3.19d), most of the NADH-GLU binding residues are information source residues, in fact, these two subsets of slow modes have the relatively highest TEs among the five basic slow mode subsets. However, interestingly to note as seen on the 2D maps, the collectivities of the TEs emphasize an extensive information flow within monomers in this conformation of the hexameric structure (Figure 3.19a and c). Indeed, the 2,3,8,9,10 slow modes reveal the NADH-GLU binding residues as information source residues with higher collectivities (Figure 3.19e) and most likely this could also emerge in the slower mode combinations of another conformation.

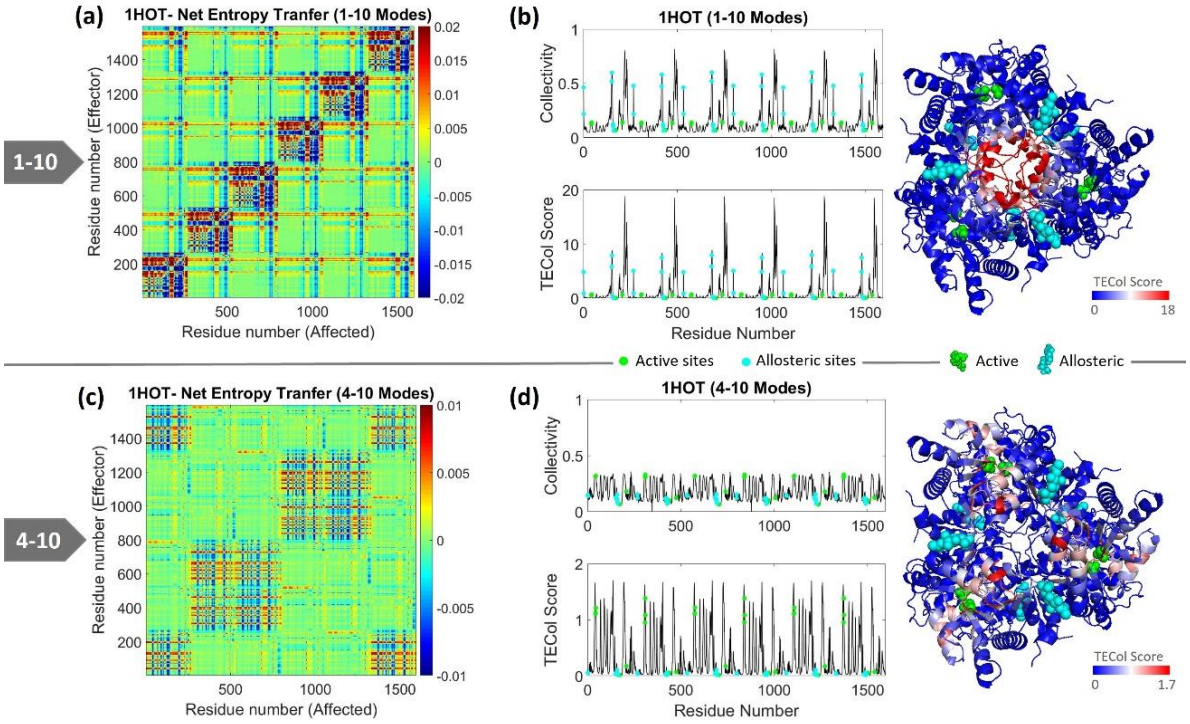

Figure 3.20. GlcN-6-P Deaminase (PDB: 1HOT [79]) (View along the threefold axis). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 4-10 slow modes, respectively, for the hexameric structure with the active site ligand PO_4 (green) and the allosteric activator 16G (cyan) [79].

Most of the 16G binding residues are information source residues circulating the allosteric region in the 1-10 slow modes (Figure 3.20b). Also, most of the PO_4 binding residues are information source residues in the 4-10 slow modes (Figure 3.20d). In fact, as seen on the 2D maps (Figure 3.20a and c), the information source residues are more effective in the dimeric and monomeric subunits and the collectivities of TE values could be much higher if evaluated over these subunits.

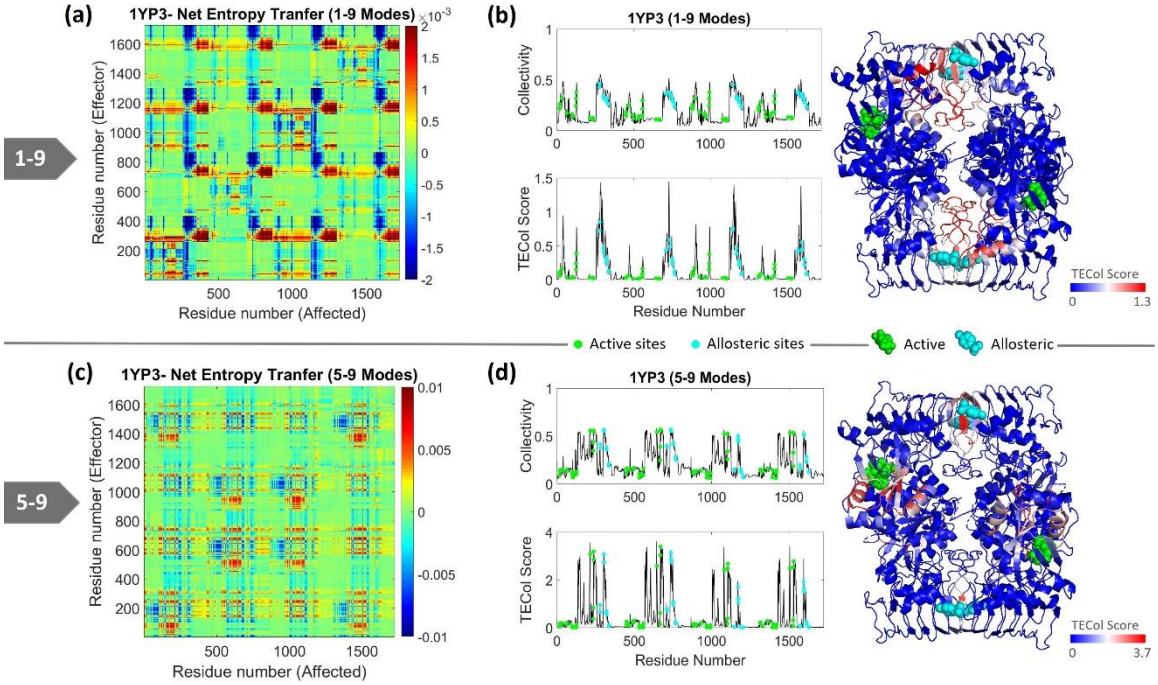

Figure 3.21. ADP-glucose Phosphorylase (PDB: 1YP3 [80]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-9 and 5-9 slow modes, respectively, for the tetrameric structure in complex with active site ligand ATP (green) and the allosteric ligand PMB (cyan) (PDB: 1YP2 [80]).

While the allosteric ligand PMB binding residues with the surrounding loops and sheets are major information source residues in the 1-9 slow modes (Figure 3.21b), the active ligand ATP binding residues are also information source residues in the 5-9 slow modes (Figure 3.21d).

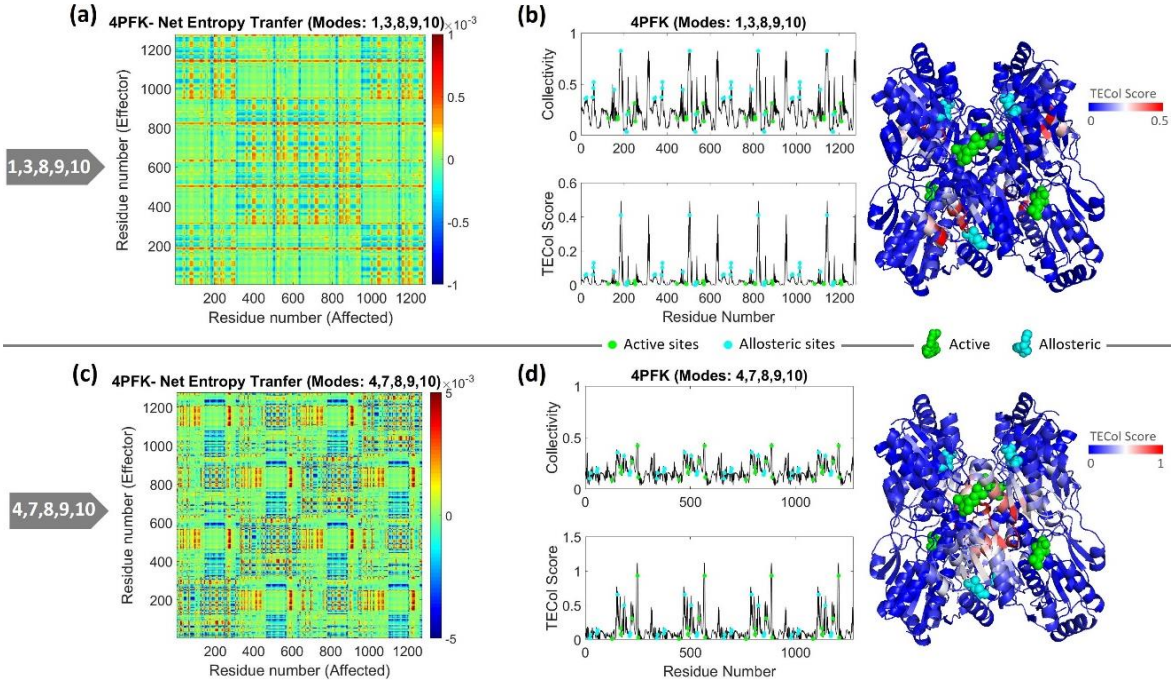

Figure 3.22. Phosphofructokinase (PDB: 4PFK [81]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1,3,8,9,10 and 4,7,8,9,10 slow modes, respectively, for the tetrameric R state conformation bound to the active site ligand F6P (green) [81] and the allosteric ligand PGA (cyan). PGA is taken from the inhibited T-state conformation (PDB: 6PFK [82]).

The 1-10 and 4-10 slow modes reveal the PGA and F6P binding residues, respectively. However, as they result in relatively lower collectivities in the behavior of information source residues on the tetrameric basis, as in the case of hexameric PDB structure 6DHD (see Figure 3.19), the results are presented here by the most collective mode combinations including the slow modes 1 and 4. This is to exemplify how the fine-tuned 1-10 and 4-10 slow modes have higher collectivities in the information source residues while revealing the same functional sites. The PGA binding residues are information source residues in the 1,3,8,9,10 slow modes (Figure 3.22b) and mainly the F6P bindings are information source residues in the 4,7,8,9,10 slow modes (Figure 3.22d).

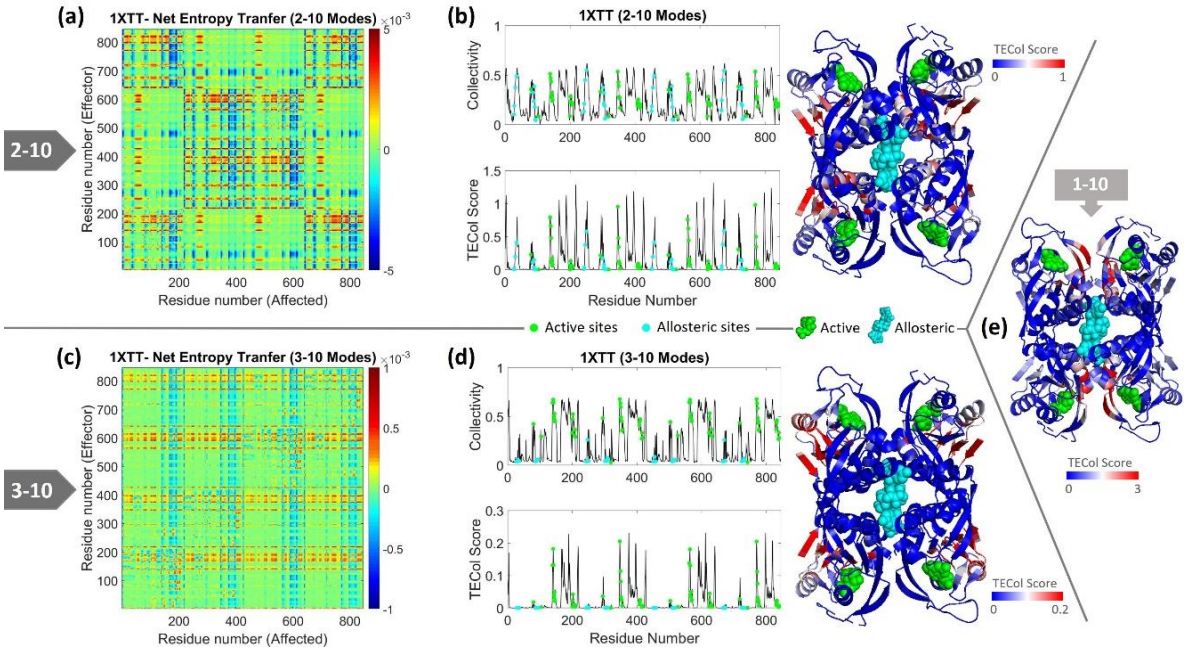

Figure 3.23. UPRTase (PDB: 1XTT [83]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 3-10 slow modes, respectively, for the tetrameric structure with the active site ligand U5P (green) [83] and the allosteric site ligand CTP (cyan). The CTP binding residues are taken from (PDB: 1XTU [83]).

The most collective slow mode subsets in the information transfer are the 2-10 and 3-10 slow modes (Figure 3.23a and c). Both U5P and CTP binding residues are information source residues in the 2-10 slow modes (Figure 3.23b). However, with the exclusion of the second slowest mode, the U5P binding residues continue to act as information source residues while the CTP binding residues disappear in the 3-10 slow modes (Figure 3.23d). This implies that without the second slowest mode, the active binding residues maintain the shear control over the structure. The subset of 1-10 slow modes, although the information transfer is not as collective as in the latter subsets of slow modes, reveals the different binding regions of U5P and CTP (Figure 3.23e). These regions do not contribute to the collective information transfer in the 2-10 and 3-10 slow modes.

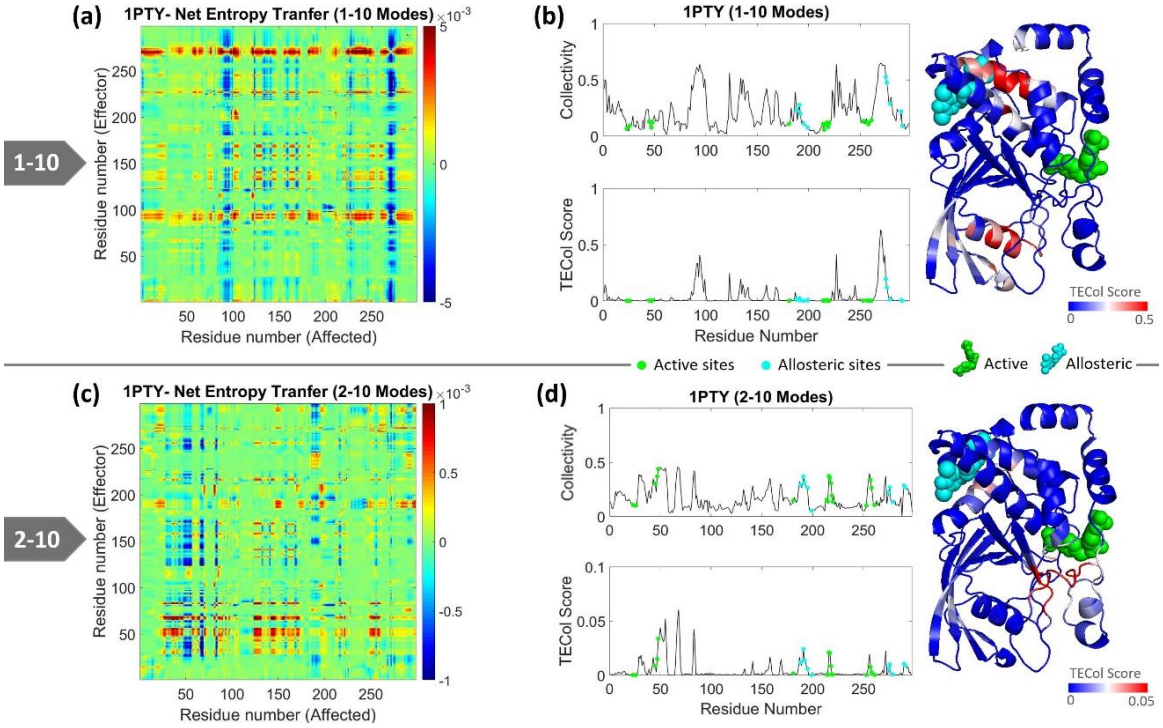

Figure 3.24. PTP1B (PDB: 1PTY [84]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively. The monomeric protein-tyrosine phosphatase 1B complexed with the active site ligand PTR (green) [84] and the allosteric ligand BB3 (cyan). BB3 is taken from the allosteric BB3-bound conformation (PDB: 1T48 [85]).

In the 1-10 slow modes, the BB3 binding residues only from one side come out as information source residues (Figure 3.24b). Here additional information source residues A228, F95, and G93 likely point to another allosteric site appearing away from the known ones. In the 2-10 slow modes, the PTR binding residues are predominant information source residues, and some of the BB3 binding residues not visible in the 1-10 slow modes appear in this subsets of slow modes (Figure 3.24d).

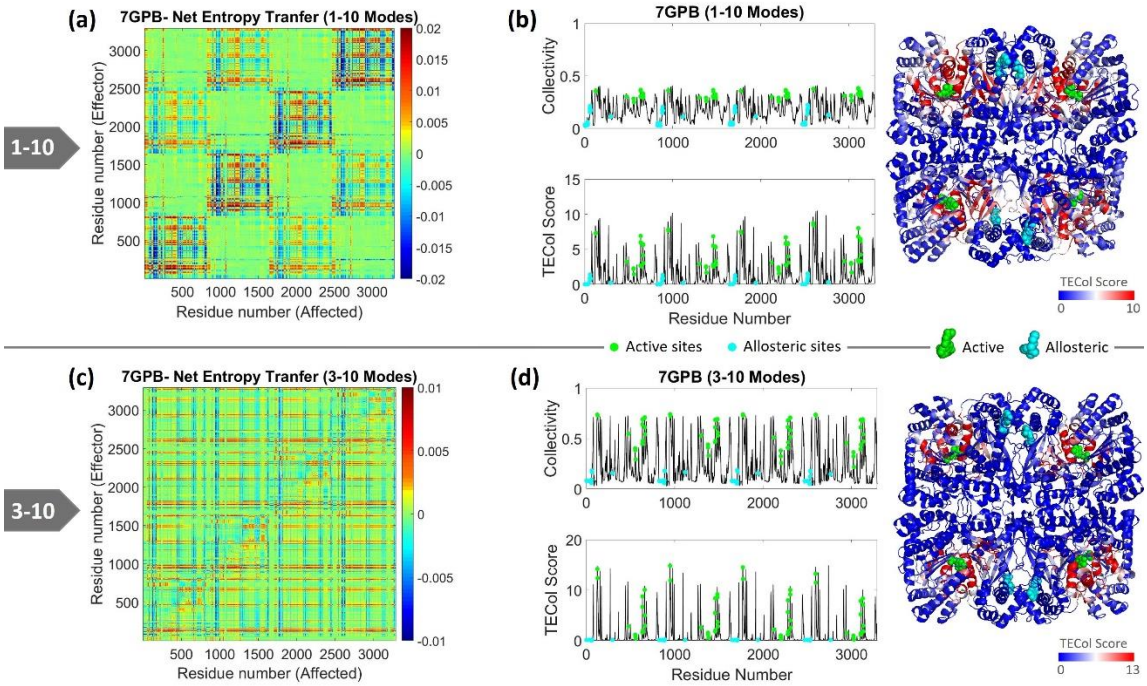

Figure 3.25. Glycogen Phosphorylase (PDB: 7GPB [86]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 3-10 slow modes, respectively, for the tetrameric structure with the active site ligand PLP-SO_4 (green) and the allosteric ligand AMP-SO_4 (cyan) [86].

The subsets of slow modes with relatively higher collectivities in the information transfer are the 1-10 and 3-10 slow modes. The PLP-SO_4 binding residues are mostly information source residues, which are also very close to and extending to the AMP-SO_4 binding residues, in the 1-10 slow modes (Figure 3.25b). The PLP-SO_4 binding residues are also information source residues in the 3-10 slow modes (Figure 3.25d). Worth to note, the information source residues are more collective in information transfer in the 1-10 slow modes within the individual subunits as seen from its TE map (Figure 3.25a).

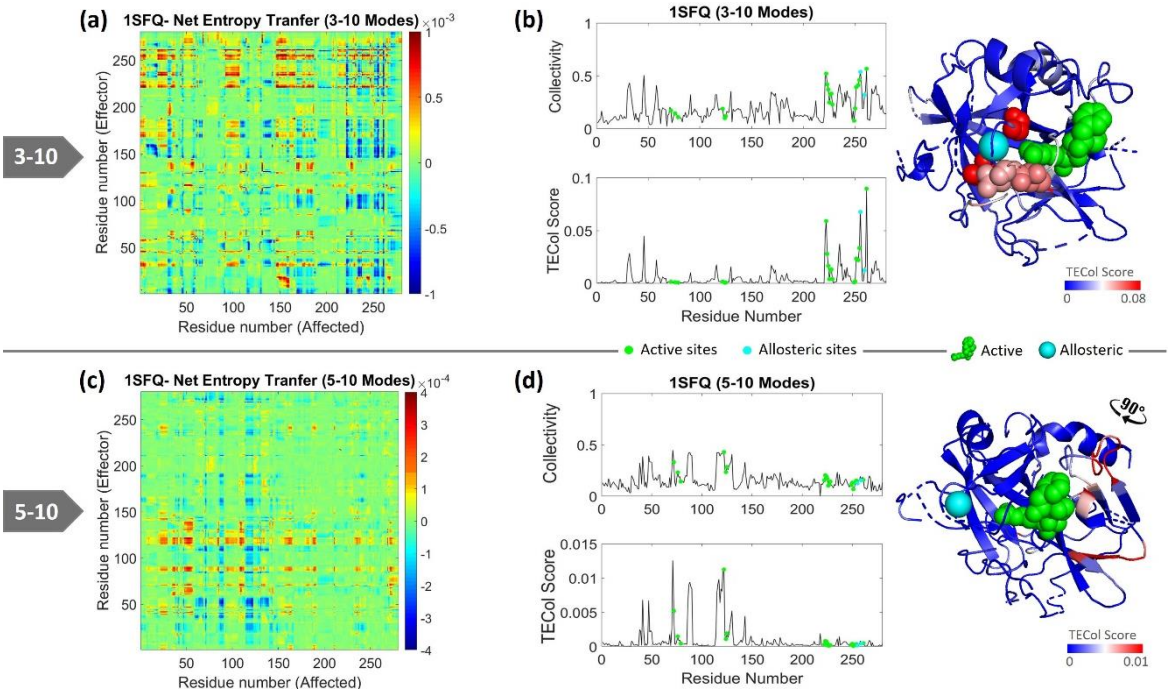

Figure 3.26. Thrombin (PDB: 1SFQ [87]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 3-10 and 5-10 slow modes, respectively, for the heterodimeric allosteric enzyme with the active site ligand 0G6 (green) and the allosteric ligand Na^+ (cyan) [87].

The subsets of slow modes maximizing the collectivities of residues in the information transfer are the 3-10 and 5-10 slow modes. The information source residues of the 3-10 slow modes reveal both the 0G6 and Na^+ binding residues (Figure 3.26b). However, some other 0G6 binding residues that are not visible in the 3-10 slow modes are captured as the information source residues of the 5-10 slow modes (Figure 3.26d).

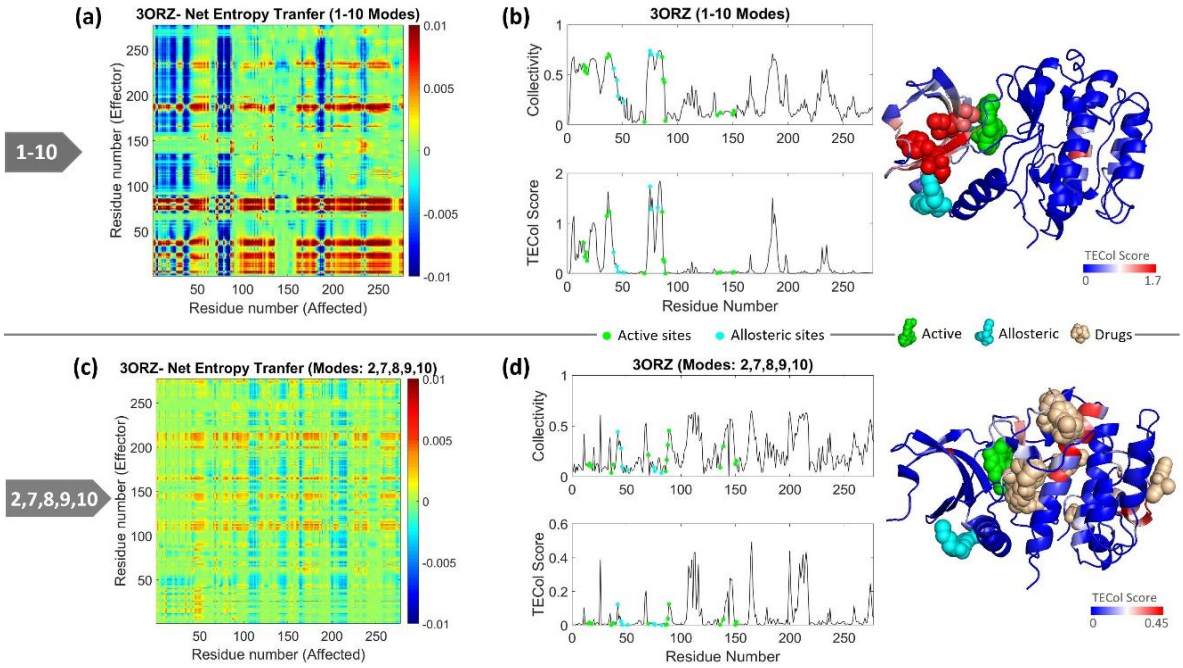

Figure 3.27. PDK1 (PDB: 3ORZ [88]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2,7,8,10 slow modes, respectively, for the monomeric protein kinase with the active site ligand BI4 (green) and the allosteric activator 2A2 (cyan) [88].

Information source residues of the 1-10 slow modes touching both ligands (represented in spheres) collectively transfer information to many others (Figure 3.27b). The five basic subsets of slow modes except the 1-10 slow modes yield very low collectivities, thus the most collective mode combination excluding the first mode (2,7,8,9,10) is evaluated (Figure 3.27c and d). This slow mode combination suggests plausible functional sites that overlay with the drug binding residues (shown as wheat spheres) obtained for the PDK1 kinase from the Kinase Atlas [89] (Figure 3.27d).

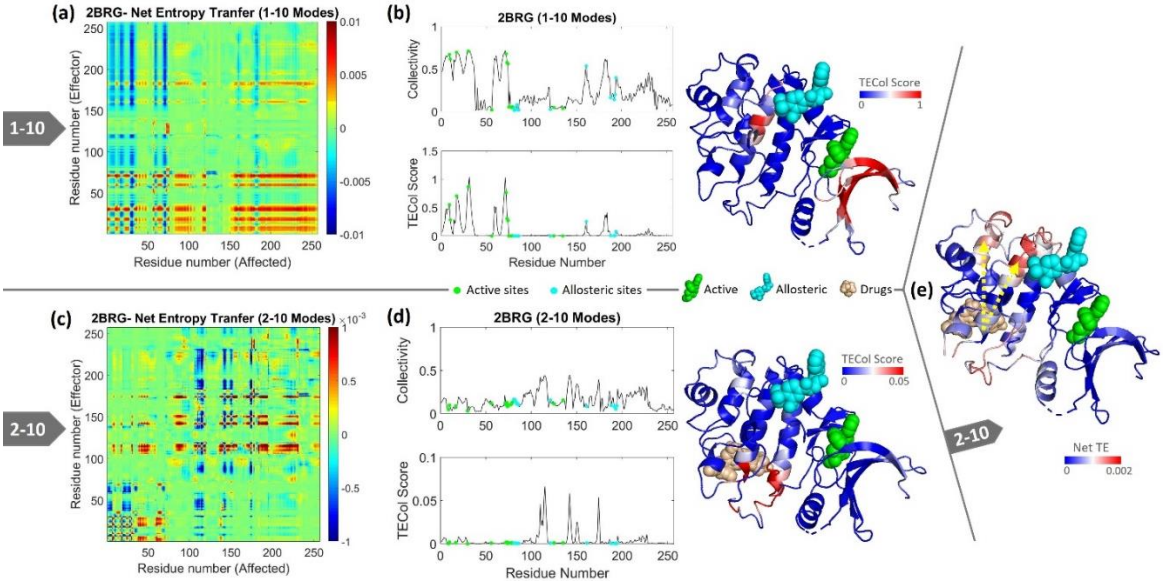

Figure 3.28. CHK1 (PDB: 2BRG [90]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively, for the monomeric kinase with the active site ligand DFY (green) [90] and the allosteric ligand AGY (cyan). AGY is taken from the allosterically inhibited structure (PDB: 3JVS [91]). Information flow from a global information source residue A186 which is also one of the drug binding sites (e).

The most prominent information source residues are at the DFY binding residues, yet one -at a lower peak- an information source residue is at the AGY binding residues in the 1-10 slow modes (Figure 3.28b). However, in the 2-10 slow modes, additional sites with information source residues appear that we observe to be associated with the drug (shown as wheat spheres) binding to the CHK1 kinase from Kinase Atlas [89] (Figure 3.28d).

Additionally, the information flow from a selected global information source residue A186 (one of the drug binding residues) to the others is shown from the highest (red) to the lowest (blue) net TE values (Figure 3.28e). Yellow dashed arrows show the direction of information flow to the allosteric ligand AGY binding region.

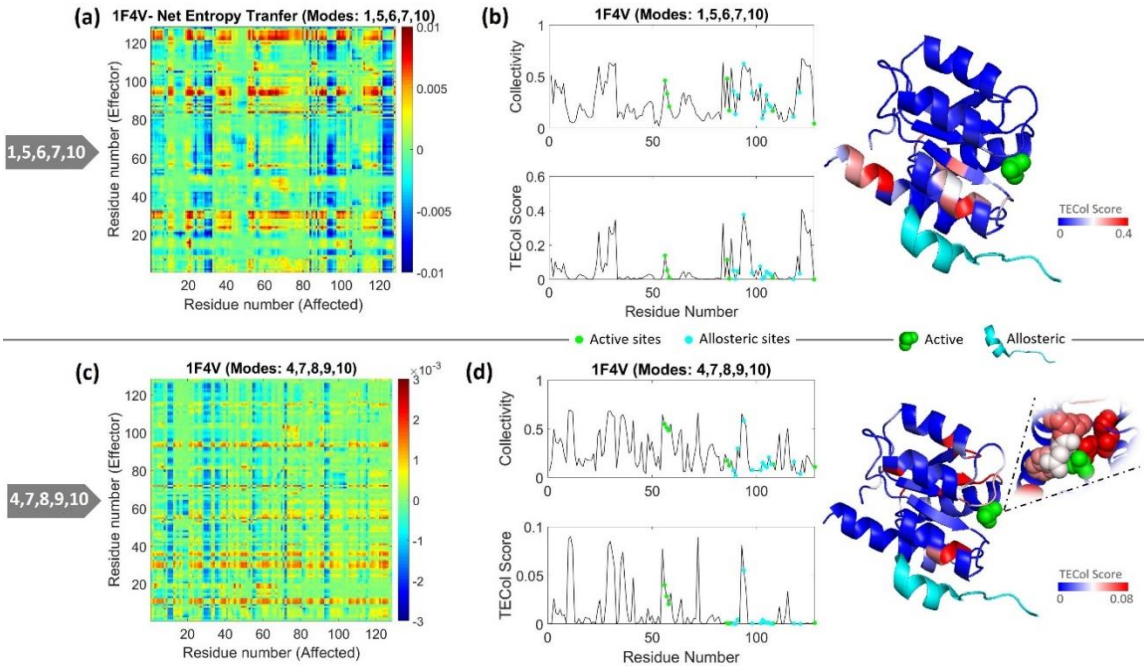

Figure 3.29. Che-Y (PDB: 1F4V [92]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1,5,6,7,10 and 4,7,8,9,10 slow modes, respectively, for the monomeric chemotaxis protein with the active ligand BeF_3 (green) and an N-terminal peptide Flim (cyan) [92].

Although the structure is monomer, the collectivities in information transfer are not high enough in any of the simplest basic five subsets of slow modes. Thus, the best results obtained by the subsets of slow modes by the combinatorial search scheme are presented in Figure 3.29. Both slow mode combinations, the 1,5,6,7,10 slow modes (Figure 3.29b) and the 4,7,8,9,10 slow modes (Figure 3.29d), give mainly the Flim binding residues as information source residues. However, the BeF_3 binding sites are also information source residues in the latter subset of slow modes. L25, F30 and V33 are proposed to be extra allosteric sites appearing as information source residues in both subsets of slow modes. Additionally, D12 and R73 are information source residues in the 4,7,8,9,10 slow modes and we see them to be functionally important residues, involved in salt bridge and an intermolecular hydrogen bond formation, respectively [21,92].

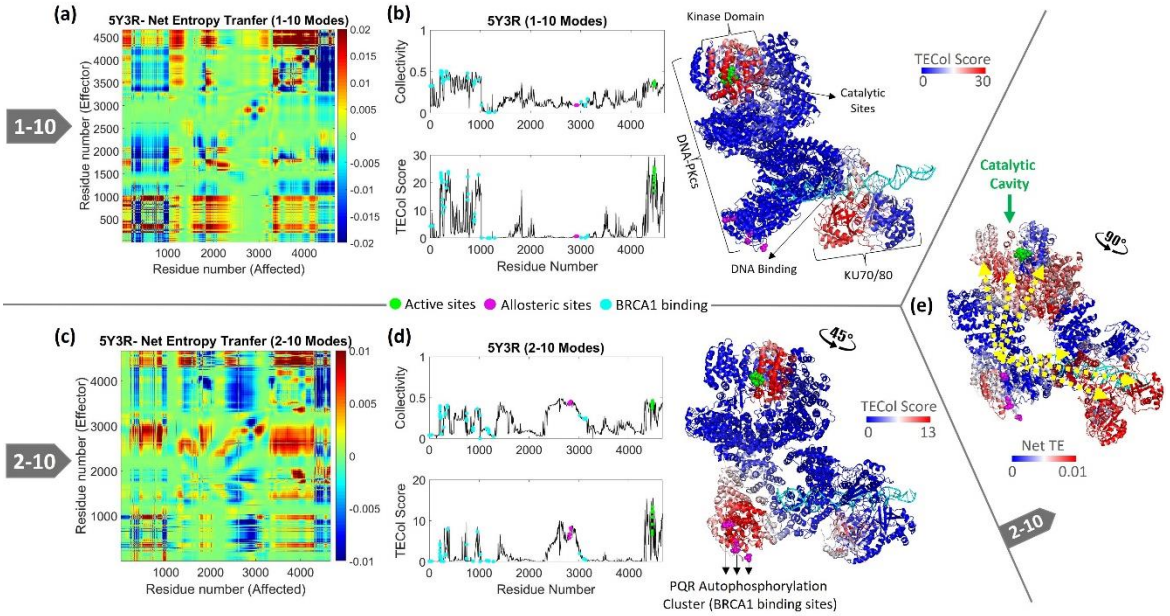

Figure 3.30. Human DNA-PK Holoenzyme (PDB: 5Y3R [45]). Net TE maps (a,c) and collectivity and TECol score plots (b,d) in the 1-10 and 2-10 slow modes, respectively. Information flow from a global information source residue in the 2-10 slow modes is represented by yellow dashed arrows (e).

The DNA-dependent protein kinase (DNA-PK) complex composed of a catalytic subunit (DNA-PKcs) and KU70/80 heterodimer bound to DNA (cyan), which allosterically stimulate the catalytic activity in kinase domain [45]. Most of the DNA binding residues of KU70/80 are information source residues in the 1-10 slow modes (Figure 3.30b). Also, catalytic sites (green) in the kinase domain of the DNA-PKcs appear as information source residues with the collective information transfer capability. Hidden information sources along with the catalytic sites appear at an extra allosteric region in the 2-10 modes (Figure 3.30d), which indeed overlap the breast cancer 1 (BRCA1) binding residues [93]. It is near to the PQR cluster, which is known to be regulating the autophosphorylation of the catalytic site of DNA-PK [94]. Additionally, the information flow from a selected global information source residue S2056 (one of the BRCA1 binding residues) to the others is shown from the highest (red) to the lowest (blue) net TE values (Figure 3.30e). Yellow dashed arrows show the direction of information flow to the regions including catalytic cavity and to the KU70/80 associated with DNA binding.

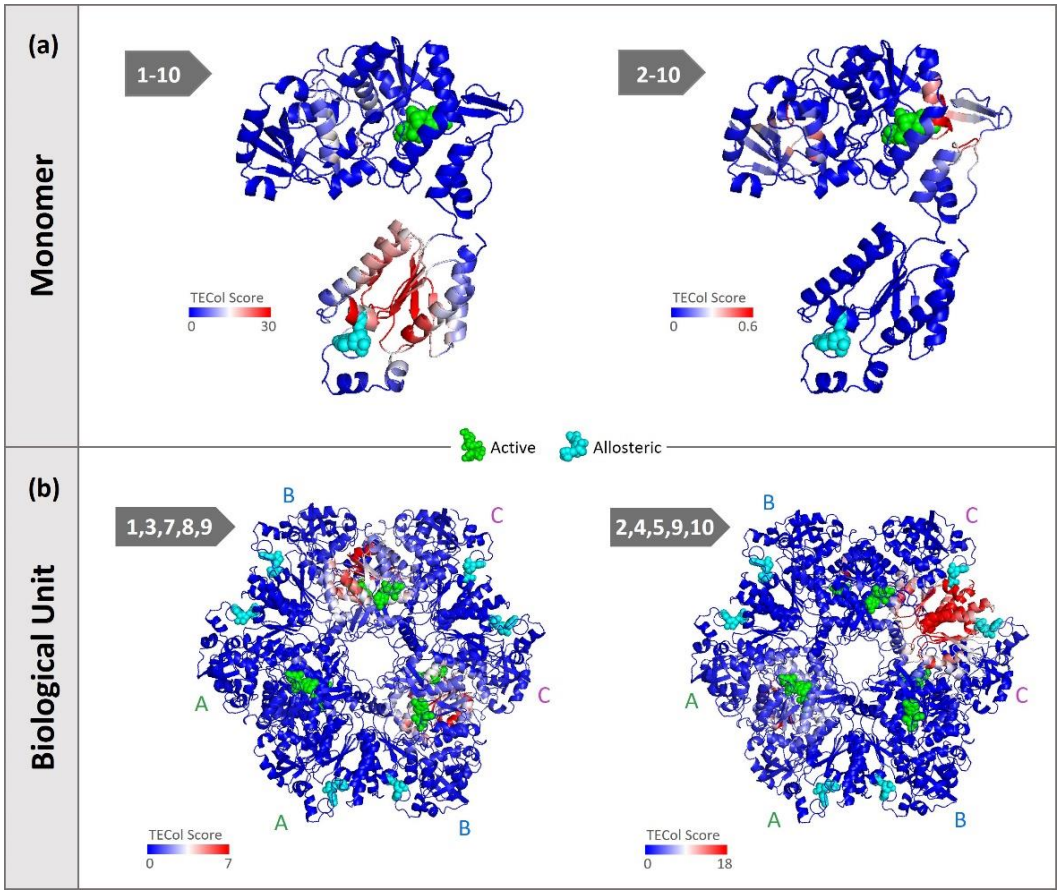

Figure 3.31. ATP Sulfurylase (PDB: 1I2D). TECol score values based on the individual/isolated monomer and the hexameric biological unit of the enzyme with the subsets of slow modes maximizing the collectivity of residues in the information transfer.

When the isolated monomeric subunit is considered, allosteric and active ligand binding residues appear as information source residues in the 1-10 and 2-10 slow modes, respectively (Figure 3.31a). However, due to the asymmetric nature when the biological unit is considered (see caption of Figure 3.18), the active sites appear as information source residues in the four chains (chains B and C) in the 1,3,7,8,9 slow modes, two of which are relatively more collective in information transfer than the others (Figure 3.31b). On the other hand, the PPS binding residues appear as information source residues in two chains of the hexamer (chains C) in the 2,4,5,9,10 slow modes.

3.2.1. Overview from Allosteric Proteins in the Dataset

The proposed method of collectivity in information transfer is applied on 20 more allosteric proteins in addition to the three illustrative scenarios for functional site predictions. The active state structures are mostly used in the GNM-based TE computations without giving any input about their active or allosteric ligand binding sites. As summarized in Figure 3.32 below, the active and allosteric sites in these allosteric proteins are revealed as global information sources.

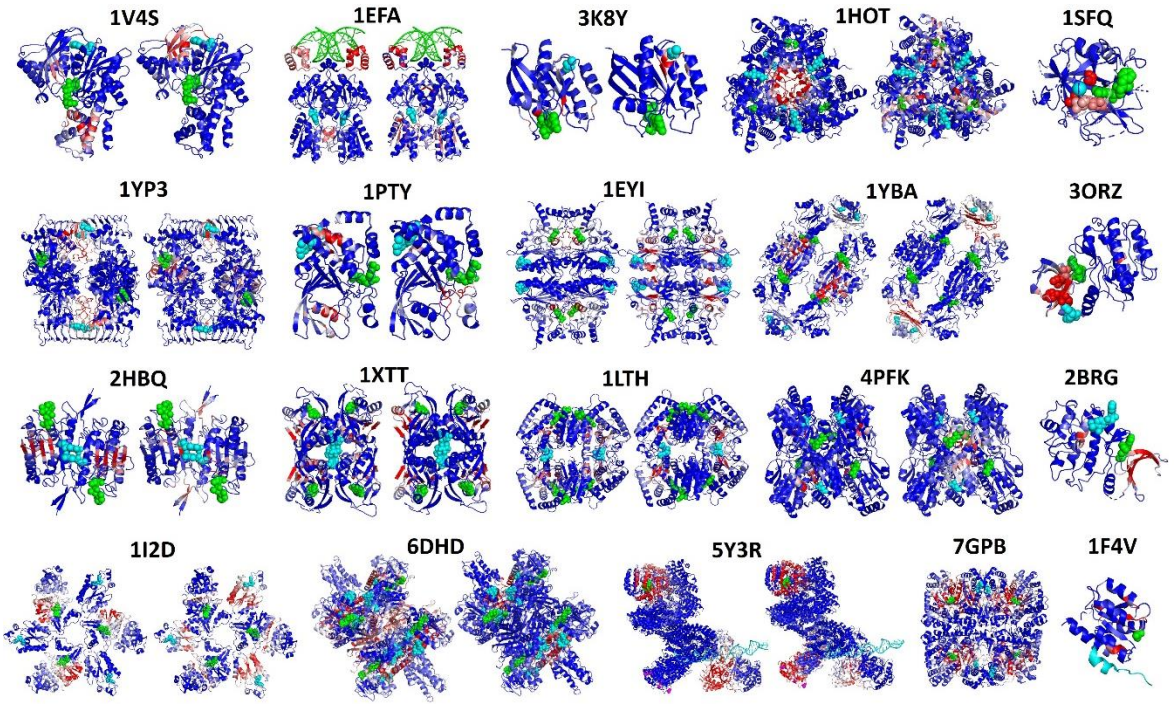

Figure 3.32. Active and allosteric sites as information sources for a test set of 20 proteins. Each ribbon structure is color-coded from the highest (red) to lowest (blue) TECol score values of residues in subsets of slow modes of maximal information transfer. Active (green) and allosteric (cyan) ligands are represented with spheres.

In the 18 cases, the simple search scheme of five subsets of slow modes that sequentially exclude the slowest mode disclose active and allosteric sites as global information sources. For the remaining two cases -phosphofructokinase (PDB: 4PFK) and chemotaxis protein Che-Y (PDB: 1F4V)- these subsets of slow modes partially detect

functional sites with relatively lower collectivities in their information transfer (Figures 3.22 and 3.29). However, the combinatorial search scheme of the three-to-five slow mode combinations based on the maximization of the collectivities of residues in the transfer of information also give active and allosteric sites as information sources in these structures.

For most of these cases, global information sources overlap with active site residues in the subsets of slow modes including the slowest mode. However, when the effects of the slowest mode and the next slower modes are sequentially removed, latent allosteric communications in which allosteric sites turn to the most pronounced information sources appear. There are the subsets of slow modes that we observe, both active and allosteric sites are detected as globular information sources. Appearance order of active and allosteric sites in the slowest mode may alter with the change in the conformational state, prioritizing a certain functional behavior or an information source over another.

The three-to-five slow mode combinations besides the simple five subsets modes are also evaluated for the 18 cases in the dataset. The slowest mode in a subset of slow modes mainly drives the general information transfer behavior. However, it is also possible to see more than one dynamic behavior in the subsets of different combinations starting with the same slowest mode. The averaging - the superimposition - of the information transfer in multiple modes of a crude dissection in such cases may preclude a clear view of information transfer patterns with eliminated “mode pollution” as well as lead to the identification of some unique functional sites that could not be observed otherwise.

3.2.2. Information Flow in Subunit Cooperativity

In multimeric proteins of identical subunits, the dynamic characteristics of functional -active and allosteric- sites as global information sources in different subsets of slow modes could be manifold as described below.

In tetrameric fructose-1,6 bisphosphatase (PDB: 1EYI) (Figure 3.14), the 1-10 slow modes result in active sites as the only information sources to integrate all subunits where the communication of diagonal monomer pairs are more emphasized. On the other hand, the

2-10 slow modes enable the allosteric sites of each monomer to transfer information to its neighboring monomer. In L-lactate dehydrogenase (PDB: 1LTH) (Figure 3.16), the 1-10 slow modes define information transfer within the dimers as a general pattern. The active site residues behave as exceptions to this general behavior and communicate with the other dimer at various positions. However, the 2-10 slow modes integrate all subunits with the information transfer from both active and allosteric sites. Similar behavior is observed in glycogen phosphorylase (PDB: 7GPB) (Figure 3.25) where the 1-10 slow modes enable information transfer from both active (relatively more collective) and allosteric sites within the dimers. On the other hand, the 3-10 slow modes having active sites with information transfer capacity integrate all subunits. The degeneracy shared by slow modes 1 and 2 is apparently at play for the allosteric communication on the level of dimers. ADP-glucose phosphorylase (PDB: 1YP3) (Figure 3.21) in the 1-9 slow modes shows that active and allosteric sites (relatively more collective) as global information sources integrate all subunits yet being more effective within dimers. Active and allosteric sites are also information sources with collective information transfer to all subunits in the 5-9 slow modes. A more repetitive communication pattern is apparent in glutamate dehydrogenase (PDB: 6DHD) (Figure 3.19), a hexameric protein. Here, allosteric sites maintain the information flow within each monomer in the 1-10 slow modes, whereas active sites display information flow within each dimer in the 4-10 slow modes. GlcN-6-P deaminase (PDB: 1HOT) (Figure 3.20) displays similar behavior to glutamate dehydrogenase. In the 1-10 slow modes, the information transfer is mainly within monomers in general, yet allosteric sites act as information sources and integrate all subunits. In the 4-10 slow modes, active site residues transfer information within dimers.

The appearance of active and allosteric sites alone or together and their control at different subunit levels revealed in various subsets of slow modes are based on the functional states represented by the structures on which the calculations are based. With conformational changes or in another functional state, it would be possible to see the slow modes/these subsets of slow modes to efface, appear, and/or shuffle, prioritizing certain functional motions according to the specific functional state.

3.2.3. Additional New Functional Sites

It is also possible to observe additional information source residues of likely functional importance, such as in PDK1 (PDB: 3ORZ) and CHK1 (PDB: 2BRG) kinases. In these two kinases, while the active and allosteric sites appear in the 1-10 slow modes, the information source residues also point to some other regions of the structure in the 2-10 slow modes that we see them overlapping with the drug bindings sites from Kinase Atlas [89] (Figures 3.27 and 3.28). Indeed, in CHK1, the information flow from a selected global information source residue A186 -which is one of the drug binding sites- is seen towards the allosteric sites (Figure 3.28e).

Along, in another kinase, human DNA-PK Holoenzyme (PDB: 5Y3R), the allosteric DNA binding region together with the catalytic sites in the kinase domain appear in the 1-10 slow modes (Figure 3.30b). However, other regions appear as additional allosteric sites in the 2-10 slow modes that we observe one of these coinciding with the breast cancer 1 (BRCA1) binding sites [93] (Figure 3.30d), which is known for its role in the autophosphorylation of the DNA-PK catalytic unit [94]. Indeed, the information flow from a selected global information source residue S2056 (one of the BRCA1 binding sites) is seen towards the catalytic cavity and the DNA binding sites associated with KU70/80 protein (Figure 3.30e). These examples indicate that information source residues are likely allosteric druggable sites to be targeted.

3.3. Statistical Analyses of Functional Site Predictions

The statistical significance of functional -active and allosteric- site predictions of each case has been measured by the *p*-value analysis as well as the performance metrics of sensitivity, specificity, precision, and accuracy. The results of the statistical analysis for the predictions are given in Table 3.1. where *p*-values that exceed the significance level of 0.05 are underlined.

Table 3.1. P-value analysis and performance metrics of functional site predictions.

Case	Subset of Slow Modes	Site Revealed	*p*-value ($\alpha = 0.05$)	Sensitivity		Specificity		Precision		Accuracy	
				7 Å	10 Å	7 Å	10 Å	7 Å	10 Å	7 Å	10 Å
1D09	1-10	Active	$p < .00001$	0.33	0.43	0.72	0.76	0.17	0.36	0.67	0.68
	4-10	Allosteric	$p < .00001$	1.00	0.97	0.81	0.84	0.35	0.51	0.83	0.86
1V4S	1-10	Active	$p = .04504$	0.31	0.37	0.72	0.77	0.24	0.50	0.63	0.61
	3-10	Allosteric	$p = .00004$	0.59	0.54	0.81	0.84	0.34	0.52	0.77	0.76
1EFA	1-10	Active	$p < .00001$	0.69	0.67	0.86	0.88	0.51	0.63	0.83	0.84
	3-10	Allosteric	$p = \underline{.29445}$	0.45	0.40	0.68	0.67	0.36	0.49	0.61	0.55
3K8Y	4-10	Active	$p = \underline{.45809}$	0.29	0.35	0.68	0.73	0.36	0.57	0.53	0.53
	5-10	Allosteric	$p = .00081$	0.70	0.67	0.87	0.94	0.58	0.83	0.84	0.86
1EYI	1-10	Active	$p < .00001$	0.75	0.67	0.74	0.79	0.50	0.67	0.74	0.74
	2-10	Allosteric	$p < .00001$	0.64	0.72	0.75	0.83	0.30	0.56	0.74	0.80
1YBA	1-10	Active	$p = .00411$	0.42	0.39	0.69	0.69	0.32	0.41	0.62	0.59
	2-10	Allosteric	$p < .00001$	1.00	1.00	0.78	0.82	0.32	0.50	0.80	0.85
1LTH	1-10	Active	$p < .00001$	0.54	0.54	0.73	0.79	0.48	0.69	0.67	0.68
	2-10	Allosteric	$p < .00001$	0.56	0.60	0.68	0.71	0.18	0.33	0.66	0.69
2HBQ	2-10	Active	$p = .02575$	0.54	0.50	0.70	0.71	0.36	0.46	0.66	0.64
		Allosteric	$p = \underline{.08727}$	0.57	0.54	0.67	0.69	0.21	0.31	0.66	0.66
1I2D	1-8	Active	$p < .00001$	0.67	0.77	0.65	0.72	0.16	0.42	0.65	0.73
	5-8	Allosteric	$p = .01336$	0.43	0.44	0.62	0.62	0.09	0.12	0.60	0.60
6DHD	1-10	Allosteric	$p < .00001$	0.52	0.49	0.92	1.00	0.64	1.00	0.83	0.81
	4-10	Active	$p = .00699$	0.48	0.40	0.59	0.56	0.26	0.35	0.56	0.50
1HOT	1-10	Allosteric	$p < .00001$	0.79	0.62	0.86	0.89	0.34	0.51	0.85	0.84
	4-10	Active	$p = .00611$	0.31	0.33	0.66	0.67	0.10	0.28	0.62	0.57
1YP3	1-9	Allosteric	$p < .00001$	0.54	0.62	0.77	0.81	0.14	0.34	0.75	0.78
	5-9	Both	$p < .00001$	0.32	0.32	0.79	0.83	0.36	0.61	0.66	0.60

Table 3.1. P-value analysis and performance metrics of functional site predictions (cont.).

4PFK	1 3 8 9 10	Allosteric	$p < .00001$	0.60	0.57	0.85	0.87	0.47	0.63	0.80	0.73
	4 7 8 9 10	Active	$p < .00001$	0.57	0.56	0.78	0.84	0.21	0.47	0.77	0.78
1XTT	2-10	Allosteric	$p = .00844$	0.44	0.50	0.74	0.77	0.32	0.46	0.67	0.70
	3-10	Active	$p < .00001$	0.39	0.35	0.83	0.86	0.53	0.70	0.68	0.60
1PTY	2-10	Active	$p = .24119$	0.31	0.35	0.72	0.74	0.21	0.42	0.64	0.61
		Allosteric	$p = .98963$	0.60	0.54	0.77	0.76	0.32	0.32	0.74	0.73
7GPB	1-10	Active	$p < .00001$	0.85	0.74	0.73	0.75	0.26	0.36	0.74	0.75
		Allosteric	$p < .00001$	0.04	0.08	0.65	0.65	0.01	0.03	0.61	0.60
1SFQ	3-10	Active	$p = .06681$	0.44	0.42	0.76	0.80	0.35	0.55	0.68	0.66
		Allosteric	$p = .08913$	1.00	0.71	0.74	0.76	0.10	0.25	0.74	0.76
3ORZ	1-10	Active	$p = .00856$	0.64	0.47	0.81	0.82	0.39	0.50	0.78	0.72
		Allosteric	$p = .00824$	0.50	0.54	0.77	0.80	0.22	0.39	0.74	0.75
2BRG	1-10	Active	$p = .00514$	0.43	0.43	0.86	0.90	0.46	0.69	0.76	0.75
1F4V	1 5 6 7 10	Active	$p = .33907$	0.57	0.46	0.74	0.76	0.36	0.54	0.71	0.65
		Allosteric	$p = .08876$	0.40	0.44	0.74	0.81	0.54	0.73	0.59	0.62
4HQJ	1-10	Active	$p = .02687$	0.90	0.85	0.72	0.76	0.23	0.37	0.74	0.77
	2-10	Allosteric	$p < .00001$	0.95	0.81	0.79	0.82	0.27	0.41	0.80	0.82
6MIX	1 4 8 9	Active	$p < .00001$	1.00	1.00	0.75	0.76	0.11	0.14	0.76	0.77
	2-11	Allosteric	$p < .00001$	0.90	0.91	0.70	0.72	0.10	0.17	0.71	0.73
	1 4 7 8	Allosteric	$p = .02464$	0.93	0.82	0.70	0.72	0.10	0.18	0.71	0.73
5Y3R	1-10	KD, KU70/80	$p < .00001$	0.58	0.59	0.77	0.79	0.08	0.14	0.76	0.77
	2-10	KD, KU70/80, BRCA1	$p < .00001$	0.44	0.45	0.73	0.73	0.07	0.13	0.71	0.71

37 out of 46 predictions are statistically significant according to their p-value analyses, corresponding to the overall success of 80.4% of our methodology. The predictions with p-values exceeding the significance level of 0.05 belong to the cases of small structures where the predictions may include both active and allosteric sites as they are in proximity on the structure. On the other hand, for the threshold distance of 7 Å, the values of four other performance metrics reached as high as 100%, 92%, 64%, and 85% respectively for sensitivity, specificity, precision, and accuracy.

Overall, sensitivity and precision values are relatively lower than specificity and accuracy values. Relatively lower precision values may be due to the fact that the predictions

in a given subset of slow modes may reveal both active and allosteric sites while here we analyze on the assumption of one functional group per each statistical set. For example, in ADP-glucose pyrophosphorylase (PDB: 1YP3), using both groups of allosteric and active sites improved the precision values. There could also be additional unknown binding sites and/or other plausible functional sites such as the ones important for subunit folding and assembly. Furthermore, neighboring information source residues of a functional site are not counted when the threshold distance is relatively small (<7 Å). Thus, precision increases up to 100% when a threshold distance of 10 Å is considered. Besides, relatively lower precision values are partially balanced by the higher specificity values for most of the cases, meaning nonfunctional residues are defined correctly.

On the other hand, lower sensitivity values reflect that some of the residues that make up a functional site do not appear as global information sources. This is expected that not all ligand or substrate binding residues need to have a dynamic capacity for collective information transfer, some may be important for additional roles such as the stabilization of the ligand-protein interactions.

3.4. Degenerate and Nondegenerate Modes

Degenerate and nondegenerate GNM modes of 14 test proteins with structural symmetry are determined by analyzing the MSF shapes of individual GNM modes, and are summarized in Table 3.2. The MSF analysis for degeneracy is exemplified by the case of ATCase in Figure 3.33. The MSF results for rest of the dataset proteins are provided in Appendix A.

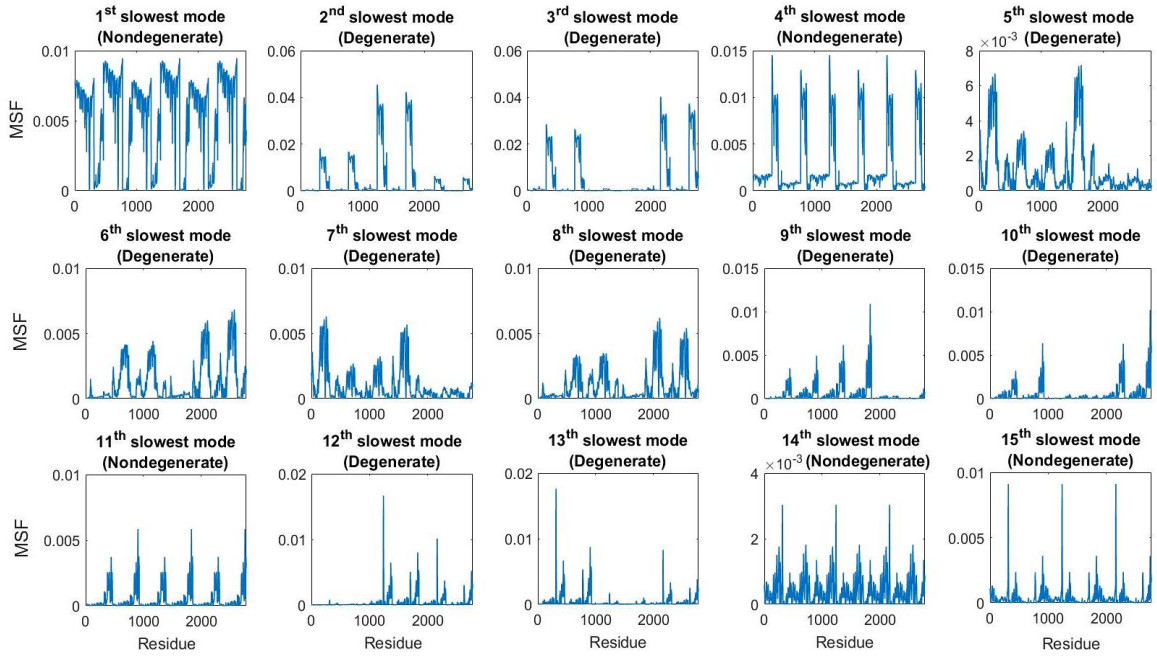

Figure 3.33. Individual GNM mode shapes for the ATCase protein (PDB: 1D09) as an example.

As seen from the individual mode shapes of ATCase, the slow modes of 1 and 4 are nondegenerate. N_s is determined as 10 for this case since the 9th and 10th slow modes are degenerate couples. Whether or not to include modes after 10th slowest mode is optional.

Table 3.2. Degeneracy analyses of 14 test proteins with structural symmetry.

PDB	# of Units	N_s	NONDEGENERATE			DEGENERATE		
			Slow Modes	Maximum Collectivity Gained	Sites Revealed with High TECol Scores	Slow Modes	Maximum Collectivity Gained	Sites Revealed with High TECol Scores
1D09	6	10	1, 4	0.63	Allosteric	2-3, 5-6, 7-8, 9-10	0.36	Allosteric
1EYI	4	10	All (1-10)	0.67	Active	None	-	-
1I2D	3	8	5, 6	0.44	Both	1-2, 3-4, 7-8	0.29	Active
1YP3	4	9	3, 6, 7	0.41	Active	1-2, 4-5, 8-9	0.44	Irrelevant

Table 3.2. Degeneracy analyses of 14 test proteins with structural symmetry (cont.).

2HBQ	2	10	All (1-10)	0.77	Active	None	-	-
4PFK	4	10	All (1-10)	0.30	Allosteric	None	-	-
7GPB	4	10	3, 4, 7, 10	0.78	Active	1-2, 5-6, 8-9	0.38	Active
1HOT	6	10	1, 8	0.44	Both	2-3, 4-5, 6-7, 9-10	0.22	Active
1LTH	4	10	All (1-10)	0.39	Active	None	-	-
6DHD	6	10	1, 6, 7, 8	0.77	Both	2-3, 4-5, 9-10	0.20	Both
6MIX	4	11	1, 4, 7, 8, 9	0.82	Active	2-3, 5-6, 10-11	0.42	Allosteric
1PSD*	4	10	All (1-10)	0.49	Active	None	-	-
1VST*	4	10	All (1-10)	0.51	Both	None	-	-
1JYF*	2	10	All (1-10)	0.41	Both	None	-	-

PDB structures of 1PSD, 1VST, and 1JYF are used for degeneracy analyses of 1YBA, 1XTT, and 1EFA, respectively, due to sligh asymmetry in their 3D structures.

Degenerate slow modes could be functionally relevant as well as nondegenerate slow modes that are essential for symmetric functional motions of the structure, as recently shown for toroidal proteins with different oligomerization states [48]. Although not revealing high collectivities, the functional significance of degenerate slow modes is evident in allosteric proteins with structural symmetries and both degenerate modes as well as nondegenerate ones carry information on active or allosteric sites (Table 3.2).

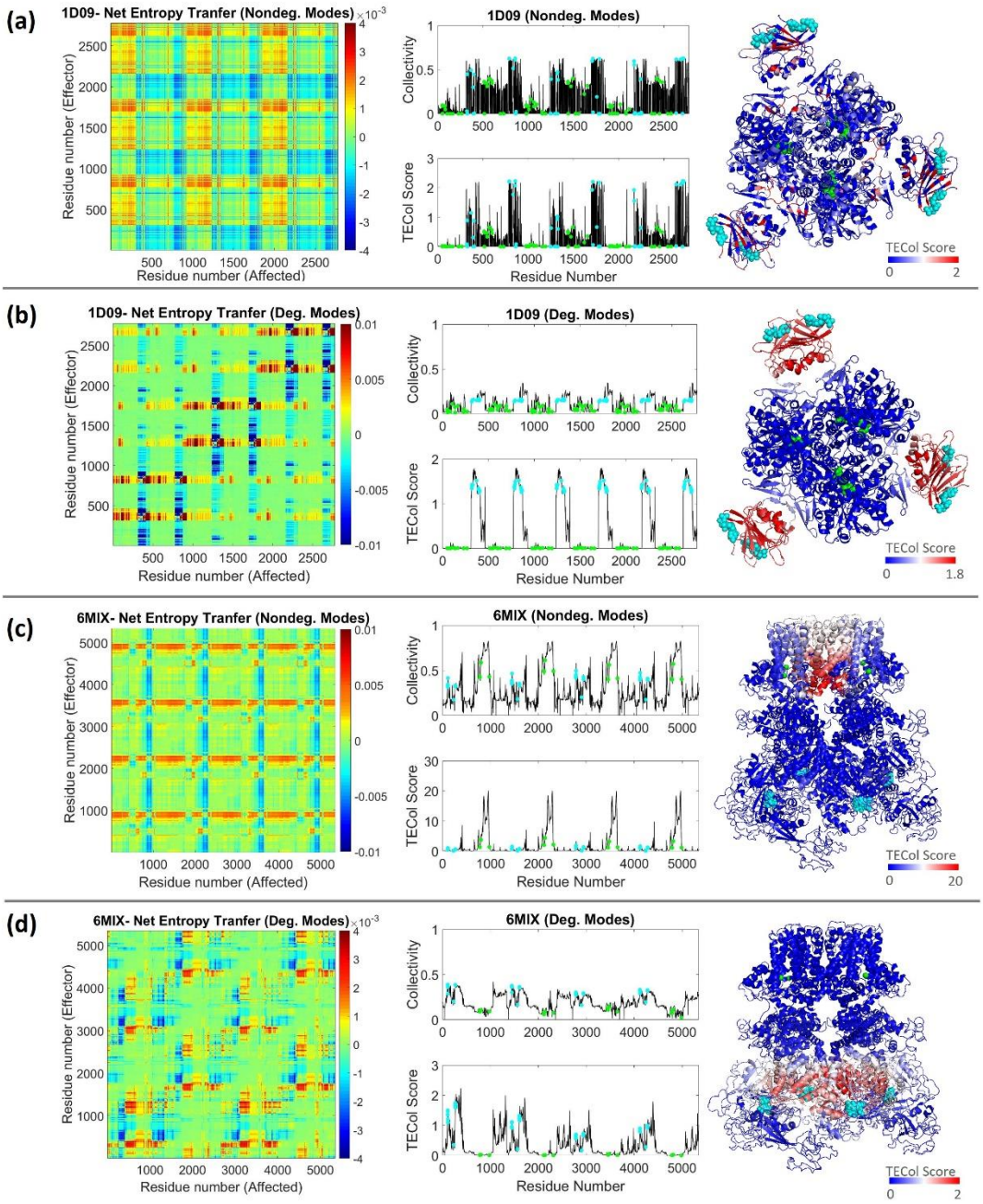

Figure 3.34. Subsets of degenerate and nondegenerate slow modes for ATCase and Human TRPM2 (see Table 3.2 for their nondegenerate and degenerate slow modes).

In Human TRPM2 (PDB: 6MIX), the subset of only degenerate slow modes mainly reveals the ADPR binding regions as seen in Figure 3.34d. Apparently, degenerate slow modes dominate the dynamics in the 2-11 slow modes that give the ADPR binding regions in this basic subset (Figure 3.6d). On the other hand, however, in some cases, it is not possible to identify either active or allosteric sites with the subsets of only nondegenerate or

only degenerate slow modes. That is for which their combinations can reveal functional sites with high collectivities such as in the case of aspartate transcarbamoylase (PDB: 1D09). For this case, two nondegenerate slow modes (1 and 4) together reveal allosteric sites (Figure 3.34a), while the slowest mode discloses active sites with contributions of some degenerate slow modes (the slow modes 2, 3, and 5-10) as seen in Figures 3.2b.

Further, not all nondegenerate or degenerate slow modes are of functional importance for a particular functional motion. Thus, the subsets of specific slow modes from nondegenerate or/and degenerate slow modes may yield higher collectivities in the information transfer and render functional sites more distinguishable as global information sources. The presence of the nondegenerate slowest mode is essential in the appearance of active sites in combinations with different slow modes, while its elimination discloses the allosteric sites in multiple subsets of slow modes being a major determinant of allosteric behavior (Figures 3.14-3.17), such as in fructose-1,6-bisphosphatase (PDB: 1EYI) (Figure 3.14) and human caspase-1 (PDB: 2HBQ) (Figure 3.17). For TRPM2 (PDB: 6MIX), this observation becomes more complex, since different subsets of slow modes yield distinct and different functional sites such as channel entrance along with ligand binding sites (Figures 3.6, 3.9 and 3.34c-d).

As a general observation, only nondegenerate slow modes mostly appoint high collectivities to both active and allosteric sites, while only degenerate slow modes are more likely to hint at either active or allosteric sites, yielding compartmentalization in the allosteric communication. Thus, these degenerate slow modes might be beneficial in dissecting the dynamics into layers of allosteric communication exclusive for entropy sources of either allosteric or active sites. Upon coupling with other degenerate/nondegenerate slow modes, this would help to complete the allosteric circuit. Furthermore, degenerate slow modes are likely those that respond to internal/external perturbation through splitting and their coupling to other slow modes are plausible, which may lead to some functional responses.

In combination with nondegenerate slow modes in the present analysis, it is paid attention not to set apart the degenerate modes of a given eigenvalue. On the other hand, it is still possible to observe the subsets of slow modes in split degeneracies with relatively maximum information transfer of residues. However, using them for functional site prediction is avoided.

3.5. Application of the GNM-based Prediction Method on DNA

As a novel application, GNM-based TE methodology has been applied to predict the functionally important sites on DNA. Unlike the model applied to proteins, now, the DNA structure is modelled as an elastic network of phosphorus atoms of nucleotides interacting with a harmonic potential function within a threshold radius (R_{cut}=19 Å). Then, nucleotides having high collectivities in information transfer are determined using the concepts explained in the Materials and Methods section.

The DNA structure bound to the human DNA-PK holoenzyme (PDB: 5Y3R) is chosen as the case study. The subset of 2-10 slow modes is designated as the subset with maximal collectivity in terms of TE and is presented in Figure 3.35.

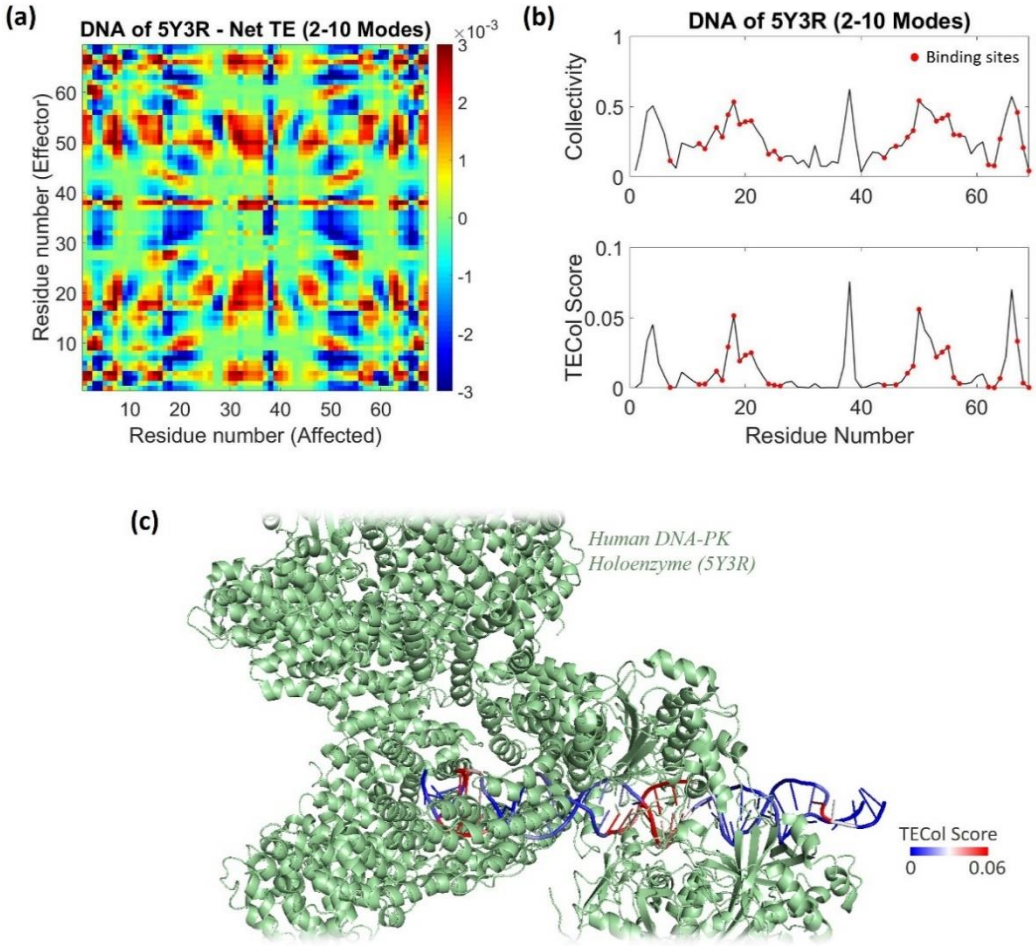

Figure 3.35. Net TE maps (a) and collectivity and TECol score plots (b) in the 2-10 slow modes. DNA structure (bound to human DNA-PK) is color-coded from the highest (red) to the lowest (blue) TECol score values of nucleotides in the subset of slow modes of maximal information transfer.

Binding nucleotides (red circles on the plots) are mostly at the rising peaks of the collectivity plot (Figure 3.35b). Thus, the global information sources appear mostly at the DNA-to-protein binding sites (Figure 3.35c). The fact that the prediction strategy also works on DNA shows the developability of the model.

4. CONCLUSION AND RECOMMENDATIONS

4.1. Conclusions

In this study, allosteric communication - which is fundamentally an information exchange between two distant protein sites - is defined by the GNM-based Transfer Entropy methodology derived from information theory [34]. It is demonstrated that known functionally important sites (active or allosteric) have the dynamic ability to collectively send information to the others - as global information sources - in different subsets of slow modes [95]. In other words, this is a unique dynamic behavior of functional residues that is only disclosed by particular slow modes, distinguishing them from one another by nature of their allosteric function in guiding other residues and thereby directing their motion.

Here, a dynamic measure is defined by adapting Bruschweiler's collectivity definition [42] to determine the subsets of slow modes from which it would be possible to reveal important residues with high collectivities in the information transfer. This allows to measure the collectivity of residues in information transfer and develop a procedure. In this developed protocol, there are two -*simple and combinatorial*- search schemes (Figure 2.2, Materials and Methods); the simple search scheme covers the basic dissection of slow modes into five subsets of slow modes by sequentially removing the slowest modes in each (i.e., 1-10, 2-10 slow modes) and the combinatorial search scheme takes the subsets of the three-to-five slow mode combinations into account. There may be more than one subset of slow modes produced by either the simple or combinatorial search schemes, allowing for the disclosure of different functional sites in various dynamic contexts. The proposed protocol is applied on a dataset of 23 proteins (Table 2.1), and it is demonstrated that the dissection of dynamic information into subsets of slow dynamic modes discloses different layers of multi-directional allosteric pathways inherent in a given protein structure.

The predictions are elaborated for three exemplary cases; aspartate transcarbamoylase (ATCase), Na^+/K^+–adenosine triphosphatase (Na^+/K^+-ATPase), and human transient receptor potential melastatin 2 (TRPM2) and are summarized for the rest of the allosteric

proteins in the dataset. As presented in Figure 3.32, functional sites can lead and direct feasible multi-directional allosteric communication pathways thanks to their collective information transfer abilities in distinct subsets of slow modes. The overlap between functional sites and highly collective information sources offers a way of optimizing the protein topology in the most cost-efficient approach of allosteric communication. The layering of the complex allosteric communication made possible by the suggested approach may allow for the chance to target only some layers of allosteric interactions while maintaining others. Especially, the allosteric interactions related with binding specificity might be a potential target to develop drugs with less side effects. With an ever-increasing number of protein structures available, the effectiveness and robustness of the proposed approach make it possible also to explore functional processes in addition to the functional site predictions.

4.2. Recommendations

In this thesis, subsets of slow modes maximizing collectivity in information transfer are identified and the results of these subsets are presented in terms of the predictions of functional sites. However, a different approach to the proposed methodology can also be brought by examining the TECol peaks produced by different mode combinations that passed the collectivity threshold. By such a mode selection algorithm, it can be observed which peaks appear in which specific subsets. Furthermore, by this approach, unique subsets in revealing distinct TECol peaks can be identified, adding a zoom-in approach to the method.

As seen in the case study on DNA, the proposed method also offers the possibility to make predictions on different types of biological molecules. It is recommended that the data set be developed in this sense. By adding new DNA cases, the results can be analyzed in more detail. Additionally, RNA molecules can be modelled using the idea of GNM adaptation for DNA.

Considering all conformational states would offer a comprehensive picture of allosteric control, i.e., a more complete dynamic allosteric landscape. A detailed conformational study based on this methodology can be performed to comprehend dynamic allosteric landscape of a single protein. Collective subsets of slow GNM modes –which disclose functional sites– can be used in ANM-LD simulations to be able to lead transitional pathways without providing target structure information. Thus, the importance of these selected modes in the conformational transition can be revealed.

REFERENCES

1. Zhang, J. and R. Nussinov (Eds.), *Protein Allostery in Drug Discovery*, Springer Singapore, 2019.

2. Verkhivker, G. M., S. Agajanian, G. Hu, and P. Tao, "Allosteric Regulation at the Crossroads of New Technologies: Multiscale Modeling, Networks, and Machine Learning", *Frontiers in Molecular Biosciences*, Vol. 7, p. 136, 2020.

3. Mannes, M., C. Martin, C. Menet, and S. Ballet, "Wandering beyond Small Molecules: Peptides as Allosteric Protein Modulators.", *Trends in Pharmacological Sciences*, Vol. 43, No. 5, pp. 406–423. 2022.

4. Monod, J. and F. Jacob, "Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation.", *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 26, pp. 389–401, 1961.

5. Cooper, A. and D.T.F Dryden, "Allostery without Conformational Change. A Plausible Model.", *European Biophysics Journal*, Vol. 11, No. 2, pp. 103–109.

6. Gunasekaran, K., B. Ma and R. Nussinov, "Is Allostery an Intrinsic Property of All Dynamic Proteins?", *Proteins*, Vol. 57, No. 3, pp. 433–443, 2004.

7. Brunori, M., "Allostery Turns 50: Is the Vintage Yet Attractive?", *Protein Science*, Vol. 20, No. 7, pp. 1097–1099, 2011.

8. Tzeng, S. R. and C. G. Kalodimos, "Protein Dynamics and Allostery: An NMR View", *Current Opinion in Structural Biology*, Vol. 21, No. 1, pp. 62–67, 2011.

9. Tsai, C. J. and R. Nussinov, “A Unified View of “How Allostery Works””, *PLoS Computational Biology*, Vol. 10, No. 2, p. e1003394, 2014.

10. Haliloglu, T. and I. Bahar, “Adaptability of Protein Structures to Enable Functional Interactions and Evolutionary Implications”, *Current Opinion in Structural Biology*, Vol. 35, pp. 17–23, 2015.

11. Guarnera, E. and I. N. Berezovsky, “Allosteric Sites: Remote Control in Regulation of Protein Activity”, *Current Opinion in Structural Biology*, Vol. 37, pp. 1–8, 2016.

12. Guo, J. and H. X. Zhou, “Protein Allostery and Conformational Dynamics”, *Chemical Reviews*, Vol. 116, No. 11, pp. 6503–6515, 2016.

13. Guarnera, E. and I. N. Berezovsky, “Toward Comprehensive Allosteric Control over Protein Activity”, *Structure*, Vol. 27, No. 5, pp. 866–878.e1, 2019.

14. Wodak, S. J., E. Paci, N. V. Dokholyan, I. N. Berezovsky, A. Horovitz, J. Li, V. J. Hilser, I. Bahar, J. Karanicolas, G. Stock, P. Hamm, R. H. Stote, J. Eberhardt, Y. Chebaro, A. Dejaegere, M. Cecchini, J. P. Changeux, P. G. Bolhuis, J. Vreede, P. Faccioli, S. Orioli, R. Ravasio, L. Yan, C. Brito, M. Wyart, P. Gkeka, I. Rivalta, G. Palermo, J. A. McCammon, J. Panecka-Hofman, R. C. Wade, A. Di Pizio, M. Y. Niv, R. Nussinov, C. J. Tsai, H. Jang, D. Padhorny, D. Kozakov and T. McLeish, “Allostery in Its Many Disguises: From Theory to Applications”, *Structure*, Vol. 27, No. 4, pp. 566–578, 2019.

15. Hawkins, R. J. and T. C. B. McLeish, “Coarse-grained Model of Entropic Allostery”, *Physical Review Letters*, Vol. 93, No. 9, p. 098104, 2004.

16. Popovych, N., S. Sun, R. H. Ebright and C. G. Kalodimos, “Dynamically Driven Protein Allostery”, *Nature Structural & Molecular Biology*, Vol. 13, No. 9, pp. 831–838, 2006.

17. Zhang, Y., P. Doruker, B. Kaynak, S. Zhang, J. Krieger, H. Li and I. Bahar, “Intrinsic Dynamics Is Evolutionarily Optimized to Enable Allosteric Behavior”, *Current Opinion in Structural Biology*, Vol. 62, pp. 14–21, 2020.

18. Kutlu, Y., N. Ben-Tal and T. Haliloglu, “Global Dynamics Renders Protein Sites with High Functional Response”, *The Journal of Physical Chemistry B*, Vol. 125, No. 18, pp. 4734–4745, 2021.

19. Kaya, C., A. Armutlulu, S. Ekesan and T. Haliloglu, “MCPath: Monte Carlo Path Generation Approach to Predict Likely Allosteric Pathways and Functional Residues”, *Nucleic Acids Research*, Vol. 41, Web Server issue, pp. W249–W255, 2013.

20. Van Den Bedem, H., G. Bhabha, K. Yang, P. E. Wright and J. S. Fraser, “Automated Identification of Functional Dynamic Networks from X-ray Crystallography”, *Nature Methods*, Vol. 10, No. 9, pp. 896–902, 2013.

21. Amor, B. R. C., M. T. Schaub, S. N. Yaliraki and M. Barahona, “Prediction of Allosteric Sites and Mediating Interactions through Bond-to-Bond Propensities”, *Nature Communications*, Vol. 7, pp. 1–13, 2016.

22. Wang, J., A. Jain, L. R. McDonald, C. Gambogi, A. L. Lee and N. V. Dokholyan, “Mapping Allosteric Communications within Individual Proteins”, *Nature Communications*, Vol. 11, No. 1, pp. 1–13, 2020.

23. Sheik Amamuddy, O., R. Afriyie Boateng, V. Barozi, D. Wavinya Nyamai and Ö. Tastan Bishop, “Novel Dynamic Residue Network Analysis Approaches to Study Allosteric Modulation: SARS-Cov-2 M(Pro) and Its Evolutionary Mutations as a Case Study”, *Computational and Structural Biotechnology Journal*, Vol. 19, pp. 6431–6455, 2021.

24. Barozi, V., T. M. Musyoka, O. Sheik Amamuddy and Ö. Tastan Bishop, "Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics", *ACS Omega,* Vol. 7, No. 15, pp. 13313–13332, 2022.

25. McClendon, C. L., A. P. Kornev, M. K. Gilson and S. S. Taylora, "Dynamic Architecture of a Protein Kinase", *Proceedings of the National Academy of Sciences,* Vol. 111, No. 43, pp. E4623–E4631, 2014.

26. Kornev, A. P. and S. S. Taylor, "Dynamics-Driven Allostery in Protein Kinases", *Trends in Biochemical Sciences,* Vol. 40, No. 11, pp. 628–647, 2015.

27. Schreiber, T., "Measuring Information Transfer", *Physical Review Letters,* Vol. 85, No. 2, pp. 461–464, 2000.

28. Kamberaj, H. and A. Van Der Vaart, "Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations", *Biophysical Journal,* Vol. 97, No. 6, pp. 1747–1755, 2009.

29. Barr, D., T. Oashi, K. Burkhard, S. Lucius, R. Samadani, J. Zhang, P. Shapiro, A. D. MacKerell and A. Van Der Vaart, "Importance of Domain Closure for the Autoactivation of ERK2", *Biochemistry,* Vol. 50, No. 37, pp. 8038–8048, 2011.

30. Corrada, D., G. Morra and G. Colombo, "Investigating Allostery in Molecular Recognition: Insights from a Computational Study of Multiple Antibody–Antigen Complexes", *The Journal of Physical Chemistry B,* Vol. 117, No. 2, pp. 535–552, 2013.

31. Zhang, L., T. Centa and M. Buck, "Structure and Dynamics Analysis on Plexin-B1 Rho GTPase Binding Domain as a Monomer and Dimer", *The Journal of Physical Chemistry B,* Vol. 118, No. 26, pp. 7302–7311, 2014.

32. Vatansever, S., Z.H. Gümüş and B. Erman, “Intrinsic K-Ras Dynamics: A Novel Molecular Dynamics Data Analysis Method Shows Causality Between Residue Pair Motions”, *Scientific Reports*, Vol. 6, No. 1, p. 37012, 2016.

33. Hacisuleyman, A. and B. Erman, “Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin”, *PLoS Computational Biology*, Vol. 13, No. 1, p. e1005319, 2017.

34. Hacisuleyman, A. and B. Erman, “Causality, Transfer Entropy, and Allosteric Communication Landscapes in Proteins with Harmonic Interactions.”, *Proteins*, Vol. 85, No. 6, pp. 1056–1064, 2017.

35. Tama, F. and Y. H. Sanejouand, “Conformational Change of Proteins Arising from Normal Mode Calculations”, *Protein Engineering*, Vol. 14, No. 1, pp. 1–6, 2001.

36. Ikeguchi, M., J. Ueno, M. Sato and A. Kidera, “Protein Structural Change upon Ligand Binding: Linear Response Theory”, *Physical Review Letters*, Vol. 94, No. 7, p. 078102, 2005.

37. Acar, B., J. Rose, B. Aykac Fas, N. Ben-Tal, O. Lewinson and T. Haliloglu, “Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters”, *Structure*, Vol. 28, No. 6, pp. 651–663.e5, 2020.

38. Ciftci, H., H. Tateishi, K. Koiwai, R. Koga, K. Anraku, K. Monde, Ç. Dağ, E. Destan, B. Yuksel, E. Ayan, G. Yildirim, M. Yigin, F. B. Ertem, A. Shafiei, O. Guven, S. O. Besler, R. G. Sierra, C. H. Yoon, Z. Su, M. Liang, B. Acar, T. Haliloglu, M. Otsuka, F. Yumoto, M. Fujita, T. Senda and H. DeMirci, “Structural Insight into Host Plasma Membrane Association and Assembly of HIV-1 Matrix Protein”, *Scientific Reports*, Vol. 11, No. 1, pp. 1–14, 2021.

39. Yang, M., N. Livnat Levanon, B. Acar, B. Aykac Fas, G. Masrati, J. Rose, N. Ben-Tal, T. Haliloglu, Y. Zhao and O. Lewinson, “Single-molecule Probing of the

Conformational Homogeneity of the ABC Transporter BtuCD", *Nature Chemical Biology*, Vol. 14, No. 7, pp. 715–722, 2018.

40. Dror, R. O., H. F. Green, C. Valant, D. W. Borhani, J. R. Valcourt, A. C. Pan, D. H. Arlow, M. Canals, J. R. Lane, R. Rahmani, J. B. Baell, P. M. Sexton, A. Christopoulos and D. E. Shaw, "Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs", *Nature*, Vol. 503, No. 7475, pp. 295–299, 2013.

41. Guarnera, E. and I. N. Berezovsky, "Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication", *PLoS Computational Biology*, Vol. 12, No. 3, p. e1004678, 2016.

42. Brüschweiler, R., "Collective Protein Dynamics and Nuclear Spin Relaxation", *The Journal of Chemical Physics*, Vol. 102, No. 8, pp. 3396–3403, 1995.

43. Wang, L., T. M. Fu, Y. Zhou, S. Xia, A. Greka and H. Wu, "Structures and Gating Mechanism of Human TRPM2", *Science*, Vol. 362, No. 6421, p. eaav4809, 2018.

44. Nyblom, M., H. Poulsen, P. Gourdon, L. Reinhard, M. Andersson, E. Lindahl, N. Fedosova and P. Nissen, "Crystal Structure of Na^+, K^+-ATPase in the Na^+-Bound State", *Science*. Vol. 342, No. 6154, pp. 123–127, 2013.

45. Yin, X., M. Liu, Y. Tian, J. Wang and Y. Xu, "Cryo-EM Structure of Human DNA-PK Holoenzyme", *Cell Research*, Vol. 27, No. 11, pp. 1341–1350, 2017.

46. Bahar, I., A. R. Atilgan and B. Erman, "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential", *Folding and Design*, Vol. 2, No. 3, pp. 173–181, 1997.

47. Haliloglu, T., I. Bahar and B. Erman, "Gaussian Dynamics of Folded Proteins", *Physical Review Letters*, Vol. 79, No. 16, pp. 3090–3093, 1997.

48. Li, H., P. Doruker, G. Hu and I. Bahar, “Modulation of Toroidal Proteins Dynamics in Favor of Functional Mechanisms upon Ligand Binding”, *Biophysical Journal*, Vol. 118, No. 7, pp. 1782–1794, 2020.

49. Kaynak, B. T. and P. Doruker, “Protein-Ligand Complexes as Constrained Dynamical Systems”, *Journal of Chemical Information and Modeling*, Vol. 59, No. 5, pp. 2352–2358, 2019.

50. Kaynak, B. T., I. Bahar and P. Doruker, “Essential Site Scanning Analysis: A New Approach for Detecting Sites That Modulate the Dispersion of Protein Global Motions”, *Computational and Structural Biotechnology Journal*, Vol. 18, pp. 1577–1586, 2020.

51. Jin, L., B. Stec, W. Lipscomb and E. Kantrowitz, “Insights into the Mechanisms of Catalysis and Heterotropic Regulation of Escherichia Coli Aspartate Transcarbamoylase based upon a Structure of the Enzyme Complexed with the Bisubstrate Analogue N-Phosphonacetyl-L-Aspartate at 2.1 Å”, *Proteins*, Vol. 37, No. 4, pp. 729–742, 1999.

52. Wang, J., K. A. Stieglitz, J. P. Cardia and E. R. Kantrowitz, “Structural Basis for Ordered Substrate Binding and Cooperativity in Aspartate Transcarbamoylase”, *Proceedings of the National Academy of Sciences*, Vol. 102, No. 25, pp. 8881–8886, 2005.

53. Cockrell, G. M., Y. Zheng, W. Guo, A. W. Peterson, J. K. Truong and E. R. Kantrowitz, “New Paradigm for Allosteric Regulation of Escherichia Coli Aspartate Transcarbamoylase”, *Biochemistry*. Vol. 52, No. 45, pp. 8036–8047, 2013.

54. Hodges, M., M. Barahona and S. N. Yaliraki, “Allostery and Cooperativity in Multimeric Proteins: Bond-to-Bond Propensities in ATCase”, *Scientific Reports*, Vol. 8, No. 1, p. 11079, 2018.

55. Lei, Z., B. Wang, Z. Lu, N. Wang, H. Tan, J. Zheng and Z. Jia, "New Regulatory Mechanism-Based Inhibitors of Aspartate Transcarbamoylase for Potential Anticancer Drug Development", *The FEBS Journal*, Vol. 287, No. 16, pp. 3579–3599, 2020.

56. Lipscomb, W. N. and E. R. Kantrowitz, "Structure and Mechanisms of Escherichia Coli", *Accounts of Chemical Research*, Vol. 45 No. 3, pp. 444–453, 2012.

57. Ke, H. M., R. B. Honzatko and W. N. Lipscomb, "Structure of Unligated Aspartate Carbamoyltransferase of Escherichia Coli at 2.6-Å Resolution", *Proceedings of the National Academy of Sciences*, Vol. 81, No. 13, pp. 4037–4040, 1984.

58. Moody, M. F., P. Vachette and A. M. Foote, "Changes in the X-ray Solution Scattering of Aspartate Transcarbamylase Following the Allosteric Transition", *Journal of Molecular Biology*, Vol. 133, No. 4, pp. 517–532, 1979.

59. Newell, J. O., D. W. Markby and H. K. Schachmann, "Cooperative Binding of the Bisubstrate Analog N-(Phosphonacetyl)-L-Aspartate to Aspartate Transcarbamoylase and the Heterotropic Effects of ATP and CTP", *The Journal of Biological Chemistry*, Vol. 264, No. 5, pp. 2476–2481, 1989.

60. Peterson, C. B. and H. K. Schachman, "Long Range Effects of Amino Acid Substitutions in the Catalytic Chain of Aspartate Transcarbamoylase. Localized Replacements in the Carboxyl-Terminal Alpha-Helix Cause Marked Alterations in Allosteric Properties and Intersubunit Interactions.", *The Journal of Biological Chemistry*, Vol. 267, No. 4, pp. 2443–2450, 1992.

61. Clausen, M. V., F. Hilbers and H. Poulsen, "The Structure and Function of the Na,K-ATPase Isoforms in Health and Disease", *Frontiers in Physiology*, Vol. 8, p. 371, 2017.

62. Glynn, I. M., “Annual Review Prize Lecture. ‘All Hands to the Sodium Pump’ ”, *The Journal of Physiology*, Vol. 462, No. 1, pp. 1–30, 1993.

63. Hilbers, F., W. Kopec, T. J. Isaksen, T. H. Holm, K. Lykke-Hartmann, P. Nissen, H. Khandelia and H. Poulsen, “Tuning of the Na,K-ATPase by the Beta Subunit”, *Scientific Reports*, Vol. 6, No.1, p. 20442, 2016.

64. Schmalzing, G. and S. Gloor, “Na^+/K^+-Pump Beta Subunits: Structure and Functions, *Cellular Physiology and Biochemistry*, Vol. 4, No. 3-4, pp. 96–114, 1994.

65. Beggah, A. T., P. Jaunin and K. Geering, “Role of Glycosylation and Disulfide Bond Formation in the Beta Subunit in the Folding and Functional Expression of Na,K-ATPase”, *The Journal of Biological Chemistry*, Vol. 272, No. 15, pp. 10318–10326, 1997.

66. Li, M., D. Jazayeri, B. Corry, K. M. McSweeney, E. L. Heinzen, D. B. Goldstein and S. Petrou, “A Functional Correlate of Severity in Alternating Hemiplegia of Childhood”, *Neurobiology of Disease*, Vol. 77, pp. 88–93, 2015.

67. Yang, W., P. T. Manna, J. Zou, J. Luo, D. J. Beech, A. Sivaprasadarao and L. H. Jiang, “Zinc Inactivates Melastatin Transient Receptor Potential 2 Channels via the Outer Pore”, *The Journal of Biological Chemistry*, Vol. 286, No. 27, pp. 23789–23798, 2011.

68. Kamata, K., M. Mitsuya, T. Nishimura, J. I. Eiki and Y. Nagata, “Structural Basis for Allosteric Regulation of the Monomeric Allosteric Enzyme Human Glucokinase”, *Structure*, Vol. 12, No. 3, pp. 429– 438, 2004.

69. Petit, P., M. Antoine, G. Ferry, J. A. Boutin, A. Lagarde, L. Gluais, R. Vincentelli and L. Vuillard, “The Active Conformation of Human Glucokinase Is Not Altered by Allosteric Activators”, *Acta Crystallographica. Section D, Biological Crystallography*, Vol. 67, No. Pt 11, pp. 929–935, 2011.

70. Bell, C. E. and M. Lewis, “A Closer View of the Conformation of the Lac Repressor Bound to Operator”, *Nature Structural Biology*, Vol. 7, No. 3, pp. 209–214, 2000.

71. Buhrman, G., G. Holzapfel, S. Fetics and C. Mattos, “Allosteric Modulation of Ras Positions Q61 for a Direct Role in Catalysis”, *Proceedings of the National Academy of Sciences*, Vol. 107, No. 11, pp. 4931–4936, 2010.

72. Choe, J. Y., H. J. Fromm and R. B. Honzatko, “Crystal Structures of Fructose 1,6-Bisphosphatase: Mechanism of Catalysis and Allosteric Inhibition Revealed in Product Complexes”, *Biochemistry*, Vol. 39, No. 29, pp. 8565–8574, 2000.

73. Thompson, J. R., J. K. Bell, J. Bratt, G. A. Grant and L. J. Banaszak, “Vmax Regulation through Domain and Subunit Changes. The Active Form of Phosphoglycerate Dehydrogenase”, *Biochemistry*, Vol. 44, No. 15, pp. 5763–5773, 2005.

74. Schuller, D. J., G. A. Grant and L. J. Banaszak, “The Allosteric Ligand Site in the Vmax-Type Cooperative Enzyme Phosphoglycerate Dehydrogenase”, *Nature Structural Biology*, Vol. 2, No. 1, pp. 69–76, 1995.

75. Iwata, S., K. Kamata, S. Yoshida, T. Minowa and T. Ohta, “T and R States in the Crystals of Bacterial L–Lactate Dehydrogenase Reveal the Mechanism for Allosteric Control”, *Nature Structural Biology*, Vol. 1, No. 3, pp. 176–185, 1994.

76. Scheer, J. M., M. J. Romanowski and J. A. Wells, “A Common Allosteric Site and Mechanism in Caspases”, *Proceedings of the National Academy of Sciences*, Vol. 103, No. 20, pp. 7595–7600, 2006.

77. MacRae, I. J., I. H. Segel and A. J. Fisher, “Crystal Structure of ATP Sulfurylase from Penicillium Chrysogenum: Insights into the Allosteric Regulation of Sulfate Assimilation”, *Biochemistry*, Vol. 40, No. 23, pp. 6795–6804, 2001.

78. Smith, T. J., P. E. Peterson, T. Schmidt, J. Fang and C. A. Stanley, “Structures of Bovine Glutamate Dehydrogenase Complexes Elucidate the Mechanism of Purine Regulation”, *Journal of Molecular Biology*, Vol. 307, No. 2, pp. 707–720, 2001.

79. Oliva, G., M. R. Fontes, R. C. Garratt, M. M. Altamirano, M. L. Calcagno and E. Horjales, “Structure and Catalytic Mechanism of Glucosamine 6-Phosphate Deaminase from Escherichia Coli at 2.1 Å Resolution”, *Structure*, Vol. 3, No. 12, pp. 1323–1332, 1995.

80. Jin, X., M. A. Ballicora, J. Preies and J. H. Geiger, “Crystal Structure of Potato Tuber ADP-Glucose Pyrophosphorylase”, *The EMBO Journal*, Vol. 24, No. 4, pp. 694–704, 2005.

81. Evans, P. R., G. W. Farrants and P. J. Hudson, “Phosphofructokinase: Structure and Control”, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 293, No. 1063, pp. 53–62, 1981.

82. Schirmer, T. and P. R. Evans, “Structural Basis of the Allosteric Behaviour of Phosphofructokinase”, *Nature*, Vol. 343, No. 6254, pp. 140–145, 1990.

83. Arent, S., P. Harris, K. F. Jensen and S. Larsen, “Allosteric Regulation and Communication between Subunits in Uracil Phosphoribosyltransferase from Sulfolobus Solfataricus”, *Biochemistry*, Vol. 44, No. 3, pp. 883–892, 2005.

84. Puius, Y. A., Y. Zhao, M. Sullivan, D. S. Lawrence, S. C. Almo and Z. Y. Zhang, “Identification of a Second Aryl Phosphate-Binding Site in Protein-Tyrosine Phosphatase 1B: A Paradigm for Inhibitor Design”, *Proceedings of the National Academy of Sciences*, Vol. 94, No. 25, pp. 13420–13425, 1997.

85. Wiesmann, C., K. J. Barr, J. Kung, J. Zhu, D. A. Erlanson, W. Shen, B. J. Fahr, M. Zhong, L. Taylor, M. Randal, R. S. McDowell and S. K. Hansen, “Allosteric

Inhibition of Protein Tyrosine Phosphatase 1B", *Nature Structural & Molecular Biology*, Vol. 11, No. 8, pp. 730–737, 2004.

86. Barford, D., S. H. Hu and L. N. Johnson, "Structural Mechanism for Glycogen Phosphorylase Control by Phosphorylation and AMP", *Journal of Molecular Biology*, Vol. 218, No. 1, pp. 233–260, 1991.

87. Pineda A. O., C. J. Carrell, L. A. Bush, S. Prasad, S. Caccia, Z. W. Chen, F. S. Mathews and E. Di Cera, "Molecular Dissection of Na^+ Binding to Thrombin", *The Journal of Biological Chemistry*, Vol. 279, No. 30, pp. 31842–31853, 2004.

88. Sadowsky J. D., M. A. Burlingame, D. W. Wolan, C. L. McClendon, M. P. Jacobson and J. A. Wells, "Turning a Protein Kinase On or Off from a Single Allosteric Site via Disulfide Trapping", *Proceedings of the National Academy of Sciences*, Vol. 108, No. 15, pp. 6056–6061, 2011.

89. Yueh, C., J. Rettenmaier, B. Xia, D. R. Hall, A. Alekseenko, K. A. Porter, K. Barkovich, G. Keseru, A. Whitty, J. A. Wells, S. Vajda and D. Kozakov, "Kinase Atlas: Druggability Analysis of Potential Allosteric Sites in Kinases", *Journal of Medicinal Chemistry*, Vol. 62, No. 14, pp. 6512–6524, 2019.

90. Foloppe, N., L. M. Fisher, R. Howes, P. Kierstan, A. Potter, A. G. S. Robertson and A. E. Surgenor, "Structure-Based Design of Novel Chk1 Inhibitors: Insights into Hydrogen Bonding and Protein-Ligand Affinity", *Journal of Medicinal Chemistry*, Vol. 48, No. 13, pp. 4332–4345, 2005.

91. Vanderpool, D., T. O. Johnson, C. Ping, S. Bergqvist, G. Alton, S. Phonephaly, E. Rui, C. Luo, Y. L. Deng, S. Grant, T. Quenzer, S. Margosiak, J. Register, E. Brown and J. Ermolieff, "Characterization of the CHK1 Allosteric Inhibitor Binding Site", *Biochemistry*, Vol. 48, No. 41, pp. 9823– 9830, 2009.

92. Lee, S. Y., H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. S. Huang, S. Kustu, E. A. Berry and D. E. Wemmer, "Crystal Structure of an Activated Response Regulator Bound to Its Target", *Nature Structural Biology*, Vol. 8, No. 1, pp. 52–56, 2001.

93. Sibanda, B. L., D. Y. Chirgadze, D. B. Ascher and T. L. Blundell, "DNA-PKcs Structure Suggests an Allosteric Mechanism Modulating DNA Double-Strand Break Repair", *Science*, Vol. 355, No. 6324, pp. 520– 524, 2017.

94. Davis, A. J., B. P. C. Chen and D. J. Chen, "DNA-PK: A Dynamic Enzyme in a Versatile DSB Repair Pathway", *DNA Repair*, Vol. 17, pp. 21–29, 2014.

95. Altintel, B., B. Acar, B. Erman and T. Haliloglu, "Subsets of Slow Dynamic Modes Reveal Global Information Sources as Allosteric Sites", *Journal of Molecular Biology*, Vol. 434, No. 17, p. 167644, 2022.

APPENDIX A: MSF SHAPES OF INDIVIDUAL GNM MODES (FOR SYMMETRIC STRUCTURES)

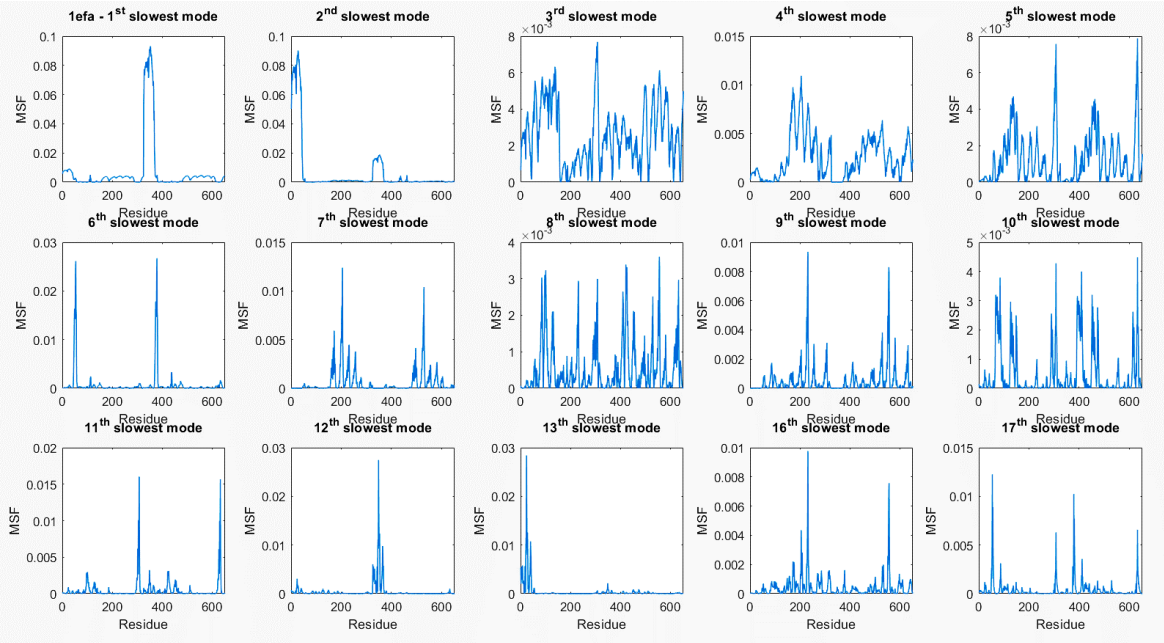

Figure A.1. Individual GNM mode shapes for Lac Repressor (PDB: 1EFA).

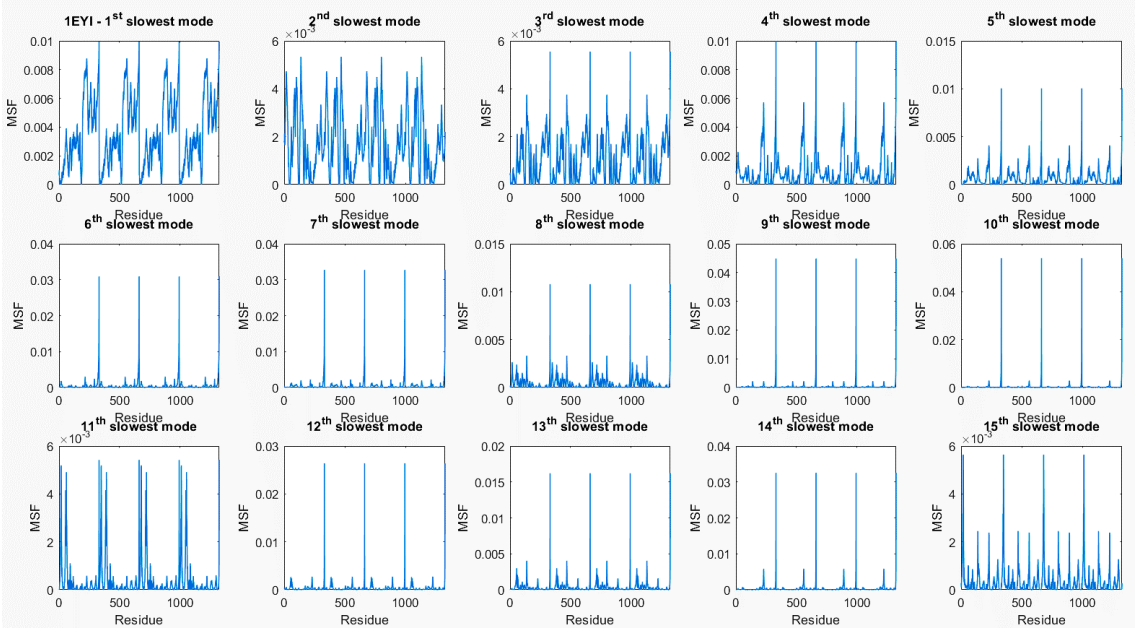

Figure A.2. Individual GNM mode shapes for Fructose-1,6 Bisphosphatase (PDB: 1EYI).

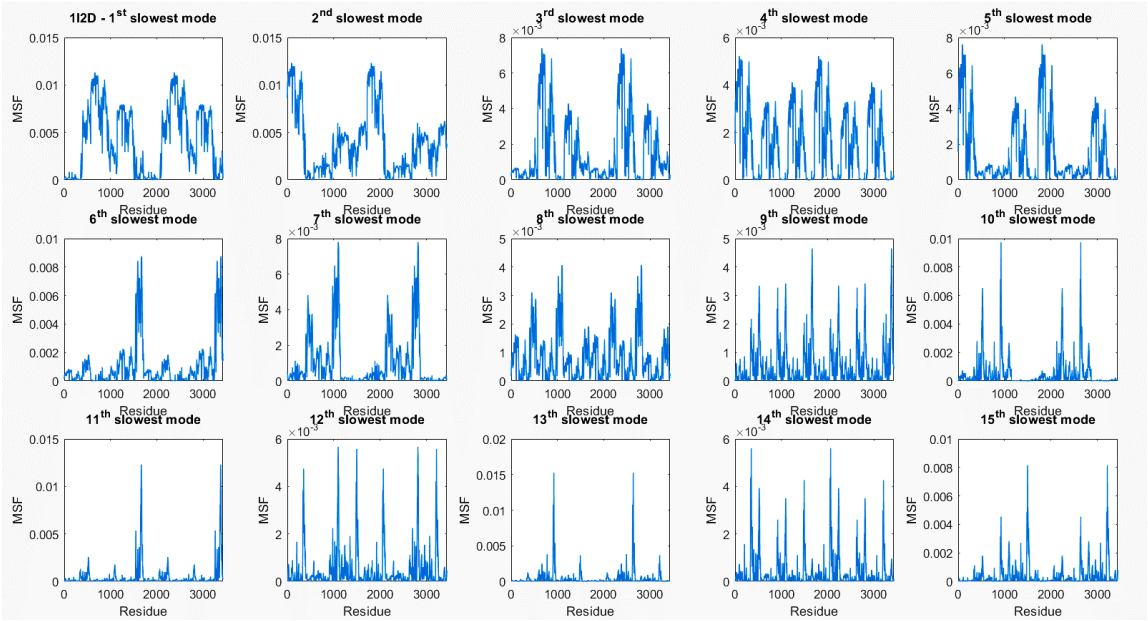

Figure A.3. Individual GNM mode shapes for ATP Sulfurylase (PDB: 1I2D).

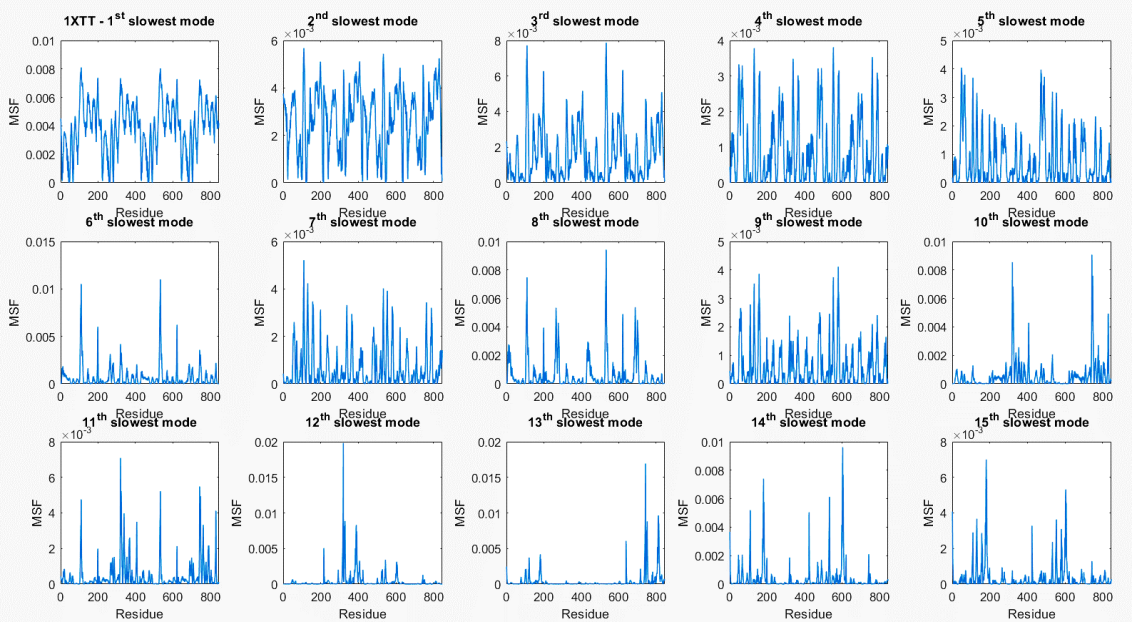

Figure A.4. Individual GNM mode shapes for UPRTase (PDB: 1XTT).

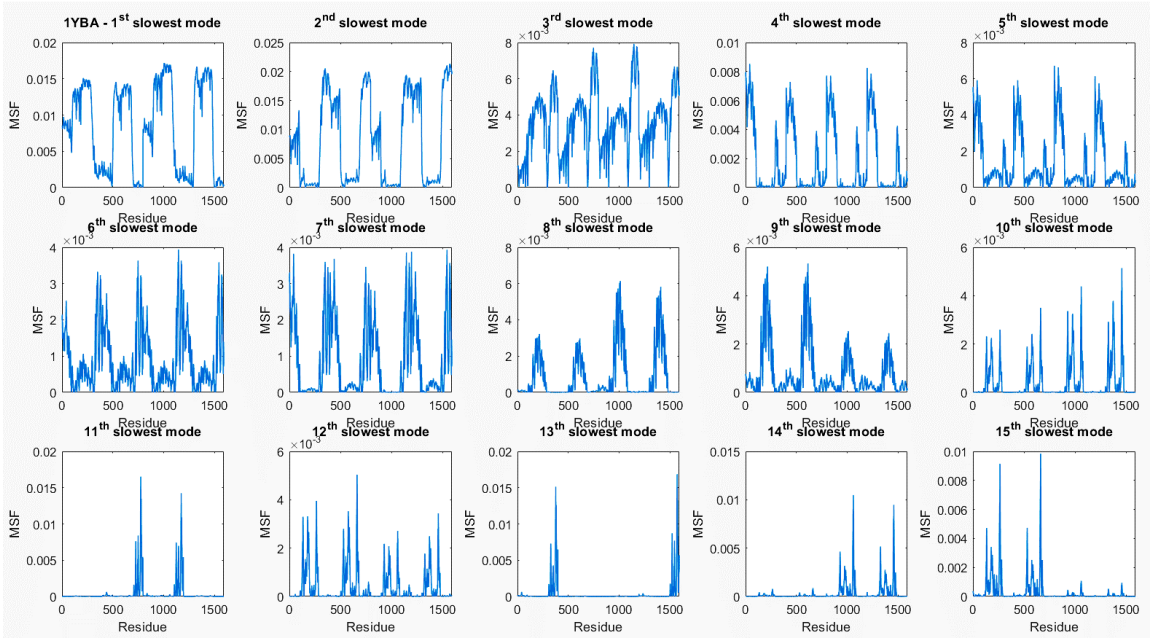

Figure A.5. Individual GNM mode shapes for Phosphoglycerate DH (PDB: 1YBA).

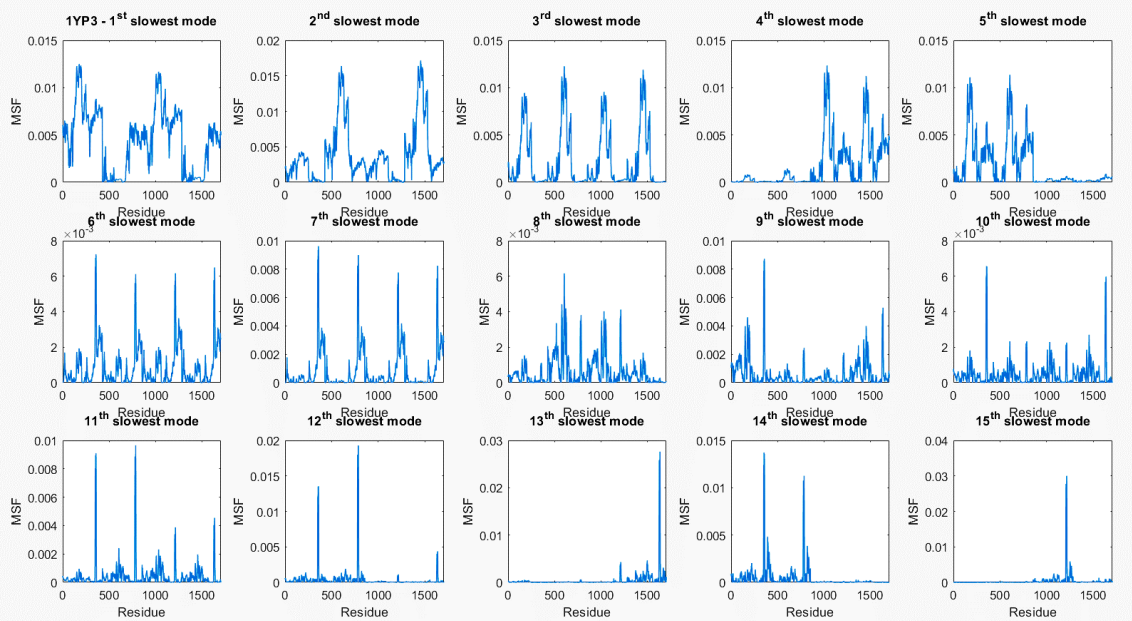

Figure A.6. Individual GNM mode shapes for ADP-glucose Phosphorylase (PDB: 1YP3).

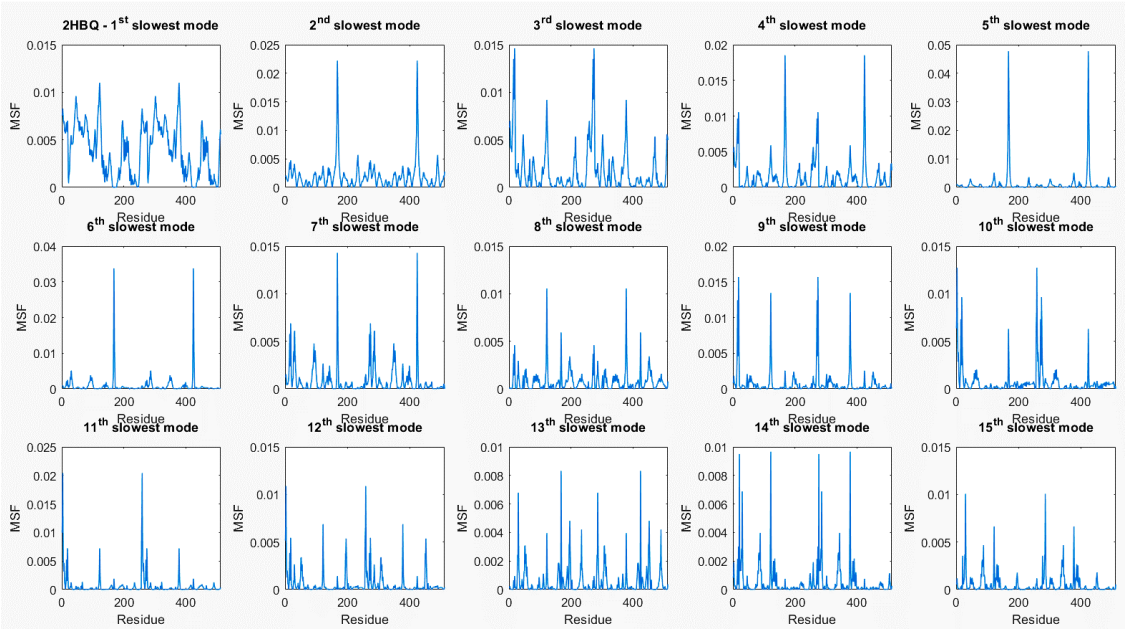

Figure A.7. Individual GNM mode shapes for Caspase-1 (PDB: 2HBQ).

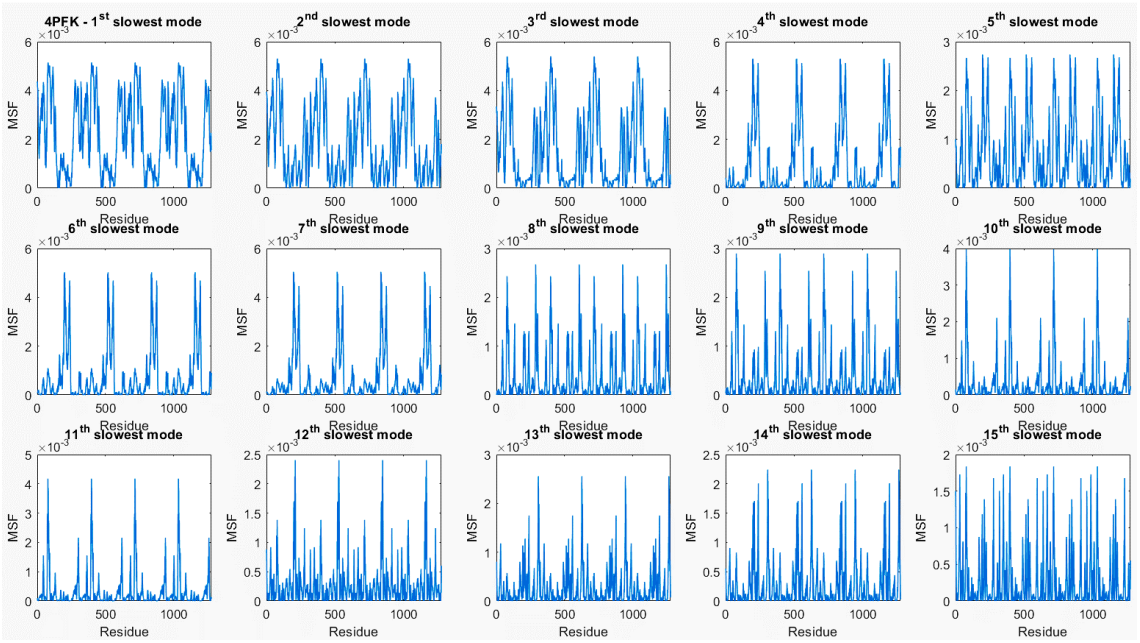

Figure A.8. Individual GNM mode shapes for Phosphofructokinase (PDB: 4PFK).

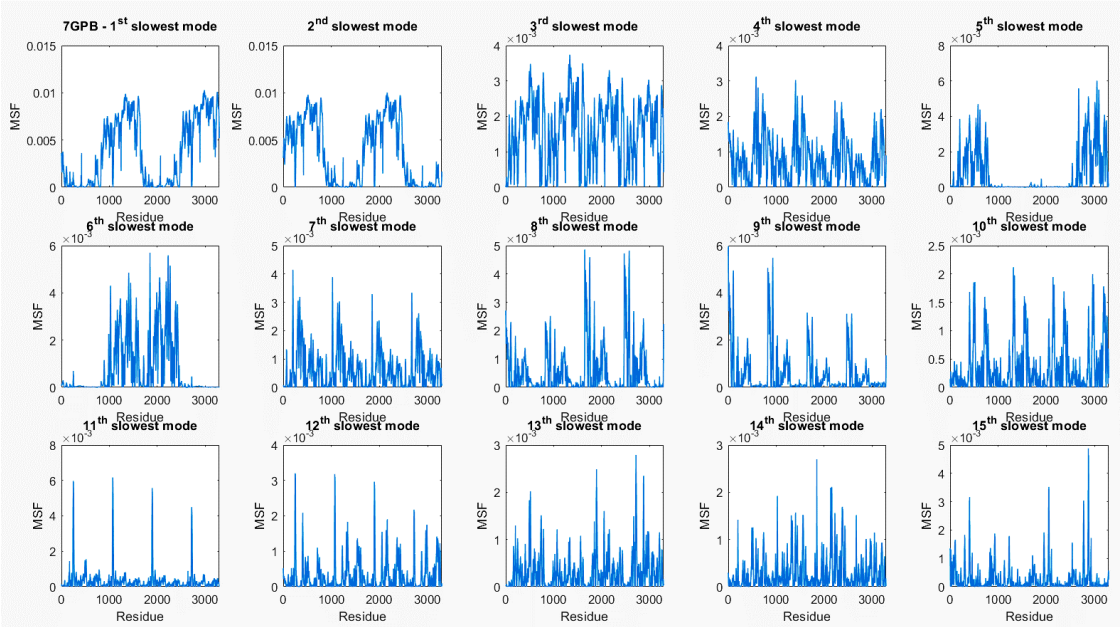

Figure A.9. Individual GNM mode shapes for Glycogen Phosphorylase (PDB: 7GPB).

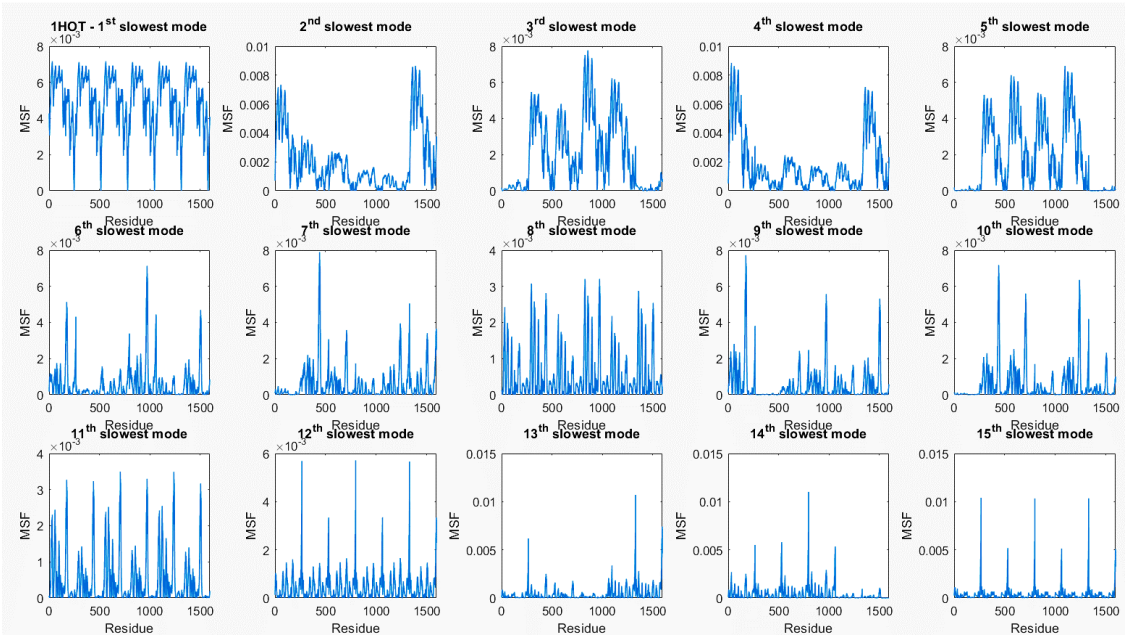

Figure A.10. Individual GNM mode shapes for glcN-6-P deaminase (PDB: 1HOT).

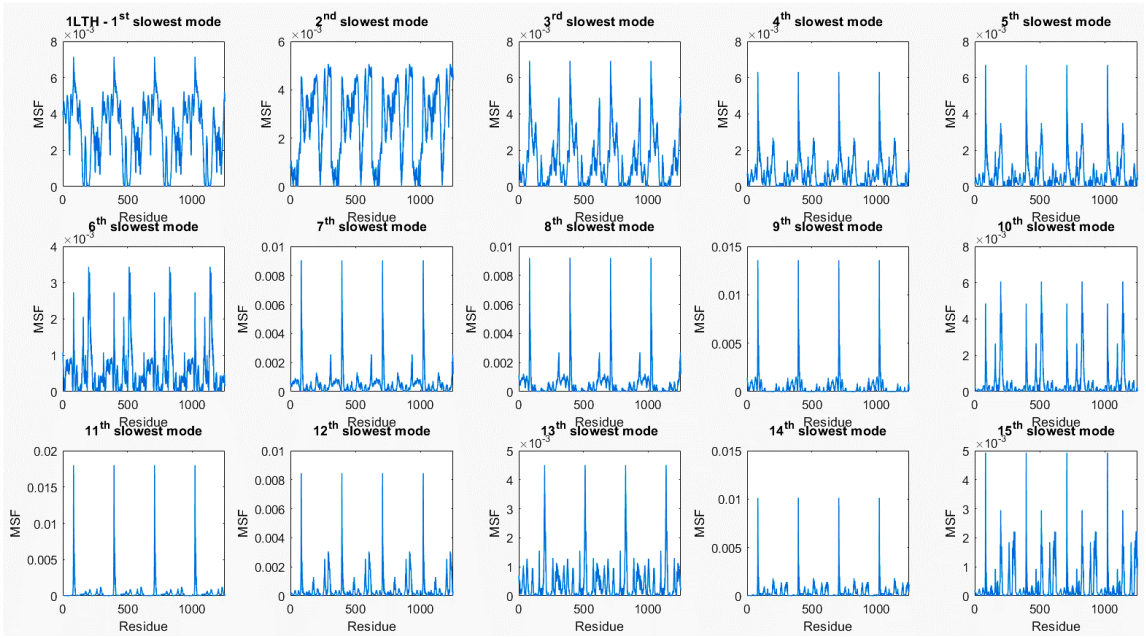

Figure A.11. Individual GNM mode shapes for Lactate DH (PDB: 1LTH).

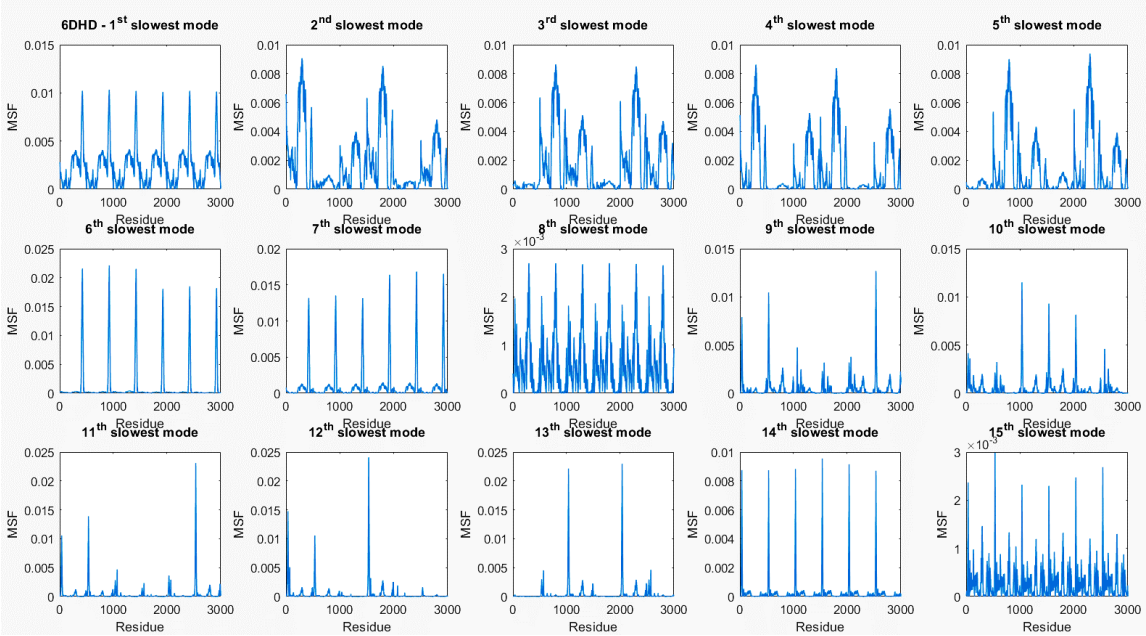

Figure A.12. Individual GNM mode shapes for Glutamate DH (PDB: 6DHD).

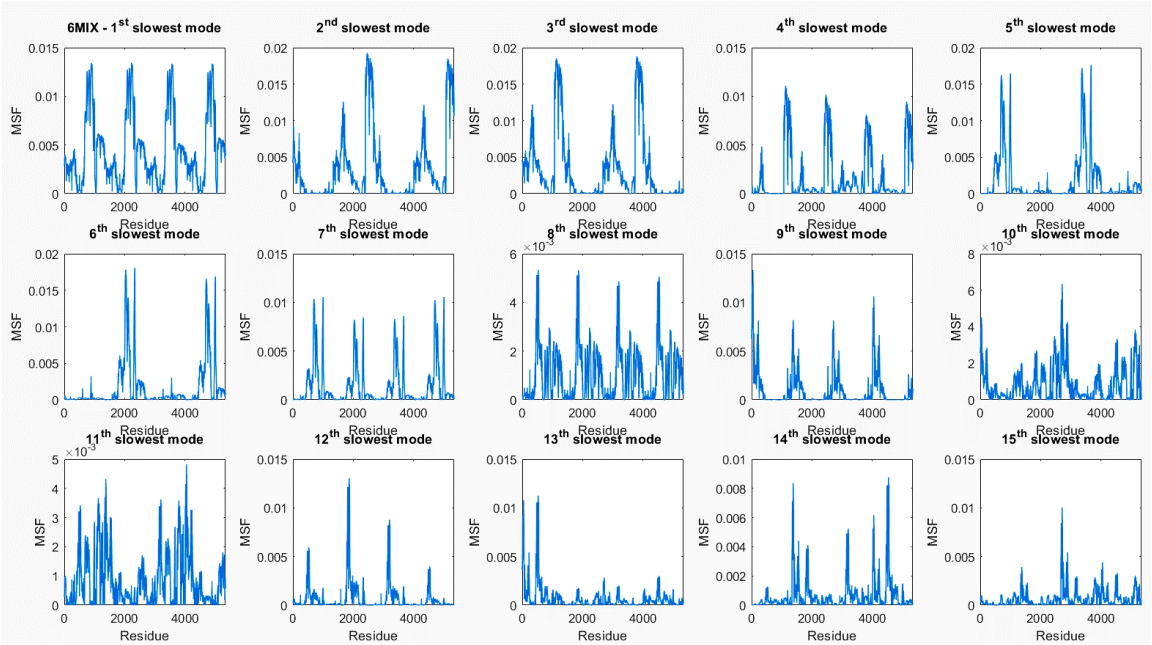

Figure A.13. Individual GNM mode shapes for human TRPM2 (PDB: 6MIX).

Images that have been created within the scope of this thesis work and whose copyrights have been transferred to the publisher, Elsevier, are reused in this thesis book according to the Elsevier's publishing policy for authors: "As the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source."