Exploring the Intrinsic Dynamics of Proteins from Metabolic Pathways using Coarse-Grained Normal Mode Analysis

Sarah Bretl, Ryan McMunn, Students of Biophysical Chemistry (Fall 2014), and Sanchita Hati

University of Wisconsin-Eau Claire, Eau Claire, WI, 54702

Acknowledgements:
UWEC-Office of Research and Sponsored Programs

ABSTRACT

Protein dynamics play an important role in molecular recognition and catalysis. Each protein structure has unique dynamics encoded by its primary sequence. The collective intrinsic dynamics of a protein, which is defined by its overall architecture, is important in promoting a pre-organized active site conformation that is conducive to molecular recognition and effective catalysis. In an attempt to classify proteins based on their intrinsic dynamics, we are investigating the intrinsic dynamics of enzymes involved in primary metabolic pathways. We have employed Normal Mode Analysis to investigate the intrinsic dynamics of twelve enzymes, six species for each of them. Preliminary studies showed that each enzyme family exhibits unique patterns of motions that are conserved across multiple species. Completion of this study is expected to produce a catalog of characteristic dynamics that could be used for functional classification of proteins, which is expected to be more efficient than traditional sequence/structure-based classification.

BACKGROUND

- Proteins are intrinsically dynamic in nature.
- Each protein structure has unique dynamics encoded by its primary sequence.
- Several studies suggested that protein dynamics play key roles in molecular recognition and catalysis (1).
- It has been proposed that the connection between protein structure and function actually lies in its dynamics (2).
- To better understand the biological function of a protein, it is essential to have an understanding of its three-dimensional structure, as well as its intrinsic dynamics.

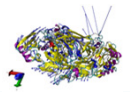
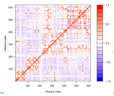
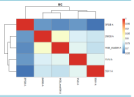
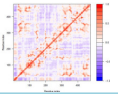
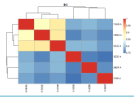
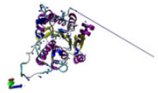
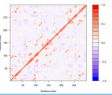
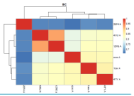
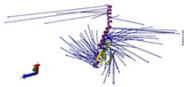
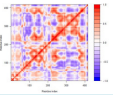
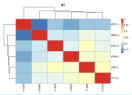
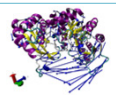
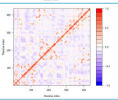
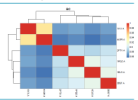
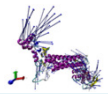
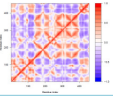
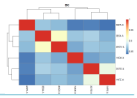
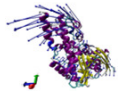
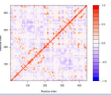
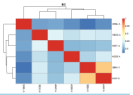
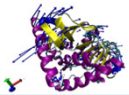
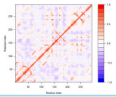
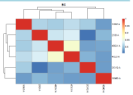
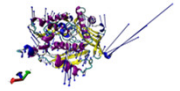
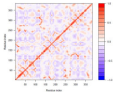
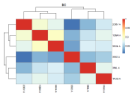
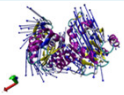
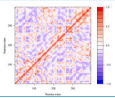
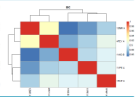
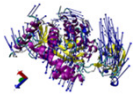
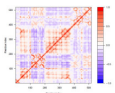
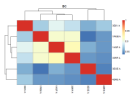
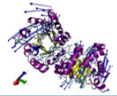
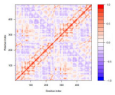
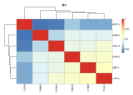
OBJECTIVES

Investigate the intrinsic dynamics of enzymes involved in metabolic pathways and analyze the correlation between structure, dynamics, and function. With the compiled data, develop a framework for classification and functional identification of proteins based on their intrinsic dynamics.

METHODS

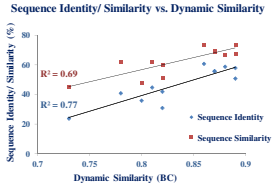
- Selection of proteins involved in common metabolic pathways and their atomic coordinates from Protein Data Bank (3).
- Identification of homologous proteins and determination of % sequence similarity and % sequence identity using BLAST (4).
- Multiple-sequence alignment using Clustal Omega (5).
- Determination of structural similarity using DALI (6).
- Normal mode analysis using WEBnm@ (7). The low frequency normal modes are calculated using the coarse-grained elastic network model.
- Normal mode **single analysis** using WEBnm@ to obtain dynamic cross-correlation matrices (DCCMs).
- Comparison of protein dynamics within families via normal mode **comparative analysis** to obtain Bhattacharyya Coefficients (8).
- Visualization of protein structure and the direction of motions using VMD (9).

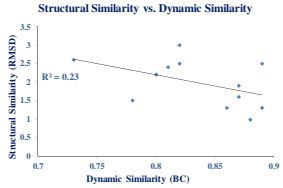
RESULTS

Protein	Structure	DCCM	Heat Map
Amine oxidase 1IVU, 1KS1, 2WO0, 3PGB, 3SX1			
Alkaline phosphatase 1EW2, 1SHN, 2X98, 3BDF, 3E2D, 4KJG			
Arginase 1CEV, 1ZPE, 2EF4, 4ITY, 4IXU, XXXX			
Citrate synthase 2H12, 2IBP, 3MSU, 3O8J, 4E6Y, 4G6B			
Enolase 1ELS, 1PDZ, 1IYX, 3B97, 4A3R, 2PTX			
Fumarase 1YFE, 3E04, 4APA, 3GTD, 4HGV, 1VDK			
Hexokinase 1BDG, 3B8A, 3W0L, 4JAX, 4QS8, 4IWV			
Malate dehydrogenase 2X0I, 1O6Z, 2CVQ, 4BGV, 4CL3, 1EMD			
Methionine adenosyl-transferase 2PO2, 1FUG, 1QM4, 3IML, 3S82, 3SO4			
Phosphoglycerate kinase 1PHP, 1V6S, 1VPE, 1ZMR, 4FEY			
Pyruvate kinase 1AQF, 1PKM, 3E0V, 3EOE, 3SRF, 4DRS			
Triose phosphate isomerase 1WYI, 1YPI, 2I9E, 3TH6, 4IOT, 8TIM			

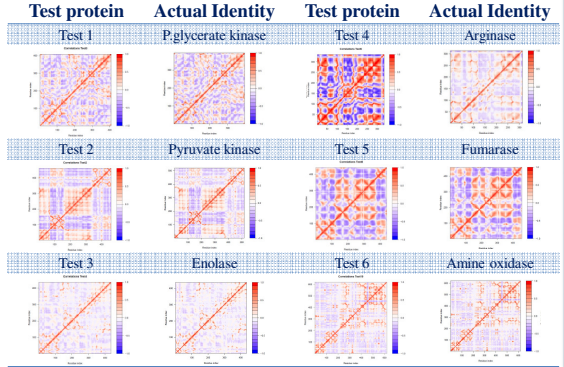
Proteins	Species (N)	Number of Amino Acids	Sequence Identity	Sequence Similarity	Structural Similarity	Dynamic Similarity
Alkaline Phosphatase	6	478 ± 30	36 ± 12	48 ± 13	2.2 ± 0.8	0.80 ± 0.05
Amine Oxidase	5	705 ± 60	24 ± 4	45 ± 3	2.6 ± 0.6	0.73 ± 0.05
Arginase	6 (1)*	306 ± 15	41 ± 8	62 ± 9	1.5 ± 0.4	0.78 ± 0.08
Citrate Synthase	6	422 ± 13	45 ± 14	62 ± 12	2.4 ± 0.4	0.81 ± 0.04
Enolase	6	434 ± 2	58 ± 7	74 ± 5	1.3 ± 0.3	0.89 ± 0.03
Fumarase	6	479 ± 12	56 ± 5	69 ± 6	1.6 ± 0.4	0.87 ± 0.02
Hexokinase	6	468 ± 15	42 ± 14	60 ± 12	3.0 ± 1.0	0.82 ± 0.04
Malate Dehydrogenase	6	311 ± 12	31 ± 6	51 ± 8	2.5 ± 0.7	0.82 ± 0.03
Methionine Adenosyltransferase	6	399 ± 11	61 ± 11	74 ± 8	1.3 ± 0.3	0.86 ± 0.02
Phosphoglycerate Kinase	5	393 ± 4	51 ± 8	68 ± 6	2.5 ± 1.2	0.89 ± 0.02
Pyruvate Kinase	6	531 ± 15	56 ± 21	70 ± 15	1.9 ± 0.7	0.87 ± 0.03
Triose Phosphate Isomerase	6	251 ± 5	59 ± 13	67 ± 7	0.98 ± 0.24	0.88 ± 0.03

* Number of models indeed in parenthesis.





Test protein	Actual Identity	Test protein	Actual Identity
Test 1	P.glycerate kinase	Test 4	Arginase
Test 2	Pyruvate kinase	Test 5	Fumarase
Test 3	Enolase	Test 6	Amine oxidase



CONCLUSIONS

- Intrinsic dynamics of proteins are maintained across species.
- Strong correlation between sequence identity and dynamic similarity suggested that a protein's primary sequence dictates its dynamics.
- Accurate identification of test proteins was possible using the developed catalog of DCCMs.

REFERENCES

1. Kamerlin, S. C. L. and Warshel, A. (2010) *Proteins: Structure, Function, and Bioinformatics*, 78: 1339-1375.
2. Howton, U., Meyer, T., Haas, J., Rex, R., Vriend, G., and Grahnmiller, H. (2012) *PLoS One 7, e33931.*
3. Berman, H.M., et. al. (2000) The Protein Data Bank. *Nucleic Acids Research*, 28: 235-242.
4. Altschul, S.F., et. al. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
5. Sievers, F., et. al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7: 539.
6. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucl. Acids Res.* 38: W545-549.
7. Hollup, SM, Salikmanidis, G, and Hunter, N. (2005) WEBnm@: a web application for normal mode analysis of proteins. *BMC Bioinformatics*. 2005 Mar 11;6(1):52.
8. Bhattacharyya, A. (1943) On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calc. Math. Soc. 35: 99-109.
9. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graphics*. 14: 33-38.