On the Usage of Graph Spectra in Protein Structural Similarity*

Sheng-Lung Peng† Yu-Wei Tsay

Department of Computer Science and Information Engineering, National Dong Hwa University

Hualien 974, Taiwan

slpeng@mail.ndhu.edu.tw

Received 16 April 2012; Revised 21 May 2012; Accepted 4 June 2012

Abstract. The spectrum of a matrix *M* is the multiset that contains all the eigenvalues of *M*. If *M* is a matrix obtained from a graph *G*, then the spectrum of *M* is also called the graph spectrum of *G*. If two graphs has the same spectrum, then they are cospectral (or isospectral) graphs. In this paper, we compare four spectra of matrices to examine their accuracy in protein structural comparison. These four spectra are derived from *adjacency matrix*, *Laplacian matrix*, *signless Laplacian matrix*, and *Seidel adjacency matrix*. In the analysis of protein subunits, corroborating protein database CATH, SCOP, and RCSB PDB, results suggest that when Seidel adjacency matrix is applied, it performs a preferable outcome in protein complex structures. In addition, the proposed method outperforms than typical RMSD geometric metric distance. This graph-theoretic approach offers a practical direction for protein subunit comparison.

Keywords: Seidel adjacency matrix, graph spectra, protein structure comparison.

References

[1] A.M. Lesk, "Detection of Three-dimensional Patterns of Atoms in Chemical Structures," *ACM Commun.*, Vol. 22, pp. 219-224, 1979.

[2] P. Rogen and B. Fain, "Automatic Classification of Protein Structure by Using Gauss Integrals," in *Proceedings of Natl. Acad. Sci. USA*, Vol. 100, pp. 119-124, 2003.

[3] C.H. Hsu, S.L. Peng, Y.W. Tsay, "An Improved Algorithm for Protein Structural Comparison Based on Graph Theoretical Approach," *Chiang Mai J. Science*, Vol. 38, pp. 71-81, 2011.

[4] D. Cvetković, P. Rowlinson, S.K. Simić, "Signless Laplacians of Finite Graphs," *Linear Algebra and Its Application*, Vol. 423, pp. 155-171, 2007.

[5] M.R. Garey and D.S. Johnson, *Computers and Interactability: A Guide to the Theory of NP-Completeness,* W.H. Freeman & Co., 1990.

[6] N.L. Biggs, *Algebraic Graph Theory*, 2nd edition, Cambridge University Press, 1993.

[7] A.E. Brouwer and W.H. Haemers, "The Gewirtz Graph: An Exercise in the Theory of Graph Spectra," *Eur. J. Comb*., Vol. 14, pp. 397-407, 1993.

[8] J.J. Seidel, "A Survey of Two-graphs," *Colloquio Internazionale sulle Teorie Combinatorie*, Vol. 17, pp. 481-511, 1973.

[9] S.L. Peng and Y.W. Tsay, "Using Minimum Weighted Bipartite Matching in Graph Spectra for Protein Subunit Comparison," manuscript, 2012.

---

* The conference version appeared in the Proceedings of the 2011 International Computer Science and Engi-neering Conference, 209-214, September 7-9, 2011, Thailand.

† Corresponding author.

[10] H. Ehrig, G. Engels, H.J. Kreowski, *Handbook of Graph Grammars and Computing by Graph Transformation: Applications, Languages and Tools*, World Scientific Publishing Company, New Jersey, 1997.

[11] S. Vishveshwara, K. Brinda, N. Kannan, "Protein Structure: Insights from Graph Theory," *Journal of the Comp. Chem.*, Vol. 1, pp. 187-211, 2002.

[12] F.S. Roberts, *Graph Theory and Its Applications to Problems of Society*, Society for Industrial Mathematics, Philadelphia, 1987.

[13] J.S. Richardson, *The Anatomy and Taxonomy of Protein Structure*, Vol. 34, 1981.

[14] D.S. Goodsell, *The Machinery of Life*. 2nd edition, Springer, 2009.

[15] C. Janeway, *Immunobiology*, 5th edition, Garland Science, 2004

[16] E. Pamer and P. Cresswell, "Mechanisms of MHC Class I – restricted Antigen Processing," *Annual Review of Immunology*, Vol. 16, pp. 323-358, 1998.

[17] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton, "CATH – A Hierarchic Classification of Protein Domain Structures," *Structure*, Vol. 5, pp. 1093-1108, 1997.

[18] H. Hasegawa and L. Holm, "Advances and Pitfalls of Protein Structural Alignment," *Curr. Opin. Struct. Biol.*, Vol. 19, pp. 341-348, 2009.

[19] J. Felsenstein, "PHYLIP – Phylogeny Inference Packaage (Version 3.2)," *Cladistics*, Vol. 5, pp. 164-166, 1989.

[20] M. Nei and W.H. Li, "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases," in *Proceedings of Natl. Acad. Sci. USA*, Vol.76, pp. 5269-73, 1979.

[21] H.M. Berman, J. Westbrook, Z. Feng, etc, "The Protein Data Bank," *Nucl. Acids Res.*, Vol. 28, pp. 235-242, 2000.

[22] Jmol: An Open-source Java Viewer for Chemical Structures in 3D. http://www.jmol.org/