REFERENCES

1. L. Adamian, R. Jackups, T. A. Binkowski, and J. Liang. Higher-order interhelical spatial interactions in membrane proteins. *J Mol Biol.*, 327:251–272, 2003.
2. L. Adamian and J. Liang. Helix-helix packing and interfacial pairwise interactions of r esidues in membrane proteins. *J. Mol. Biol.*, 311:891–907, 2001.
3. L. Adamian and J. Liang. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins.*, 47:209–218, 2002.
4. L. M. Amzel. Calculation of entropy changes in biological processes: folding, binding, and oligomerization. *Methods Enzymol*, 323:167–77, 2000.
5. C. Anfinsen, E. Haber, M. Sela, and F. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci.*, 47:1309–1314, 1961.
6. C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.
7. Park B. and Levitt M. Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. *J. Mol. Biol.*, 258:367–392, 1996.
8. U. Bastolla, J. Farwer, E. W. Knapp, and M. Vendruscolo. How to guarantee optimal stability for most representative structurs in the protein data bank. *Proteins*, 44:79–96, 2001.
9. U. Bastolla, M. Vendruscolo, and E. W. Knapp. A statistical mechanical method to optimize energy functions for protein folding. *Proc Natl Acad Sci USA*, 97:3977–3981, 2000.
10. A Ben-Naim. Statistical potentials extracted from protein structures: Are these meaningful potentials? *J. Chem. Phys.*, 107:3698–3706, 1997.

Models and Algorithms for Biomolecules and Molecular Networks, first edition.
By
Bhaskar DasGupta and Jie Liang

11. A. Ben-Naim. Statistical potentials extracted from protein structures: Are these meaningful potentials? *J. Chem. Phys.*, 107:3698–3706, 1997.
12. M Berkelaar. LP_Solve package. 2004.
13. M. R. Betancourt and D. Thirumalai. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.*, 8:361–369, 1999.
14. J. R. Bienkowska, R. G. Rogers, and T. F. Smith. Filtered neighbors threading. *Proteins*, 37:346–359, 1999.
15. A. J. Bordner and R. A. Abagyan. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. *Proteins*, 57(2):400–13, 2004.
16. B.R. Brooks, C.L. 3rd Brooks, A.D. Jr Mackerell, L Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus. CHARMM: The biomolecular simulation program. *J. Computational Chem.*, 30:1545–1614, 2009.
17. N. V. Buchete, J. E. Straub, and D. Thirumalai. Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures. *The Journal of Chemical Physics*, 118:7658–7671, 2003.
18. N. V. Buchete, J. E. Straub, and D. Thirumalai. Orientational potentials extracted from protein structures improve native fold recognition. *Protein Sci.*, 13:862–74, 2004.
19. C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Knowledge Discovery and Data Mining*, 2(2), 1998.
20. C. W. Carter Jr., B. C. LeFebvre, S. A. Cammer, A. Tropsha, and M. H. Edgell. Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. *J. Mol. Biol.*, 311(4):625–638, 2001.
21. H. S. Chan and K. A. Dill. Origins of structure in globular proteins. *Proc. Natl. Acad. Sci.*, 87(16):6388–6392, 1990.
22. T. L. Chiu and R. A. Goldstein. Optimizing energy potentials for success in protein tertiary structure prediction. *Folding Des.*, 3:223–228, 1998.
23. C. Czaplewski, S. Rodziewicz-Motowidlo, A. Liwo, D. R. Ripoll, R. J. Wawak, and H. A. Scheraga. Molecular simulation study of cooperativity in hydrophobic association. *Protein Sci*, 9:1235–1245, 2000.
24. J. Czyzyk, S. Mehrotra, M. Wagner, and S. Wright. PCx package. 2004.
25. B. I. Dahiyat and S. L. Mayo. *De Novo* protein design: Fully automated sequence selection. *Science*, 278:82–87, 1997.
26. J. M. Deutsch and T. Kurosky. New algorithm for protein design. *Phys. Rev. Lett.*, 76(2):323–326, 1996.
27. R. S. DeWitte and E. I Shakhnovich. SMoG: de novo design method based on simple, fast and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.*, 118:11733–44, 1996.
28. J. R. Dima, R. I. Banavar and A. Maritan. Scoring functions in protein folding and design. *Protein Sci.*, 9:812–819, 2000.
29. H. Dobbs, E. Orlandini, R. Bonaccini, and F. Seno. Optimal potentials for predicting inter-helical packing in transmembrane proteins. *Proteins*, 49(3):342–349, 2002.

30. Y. Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282(5389):740–744, 1998.
31. M. P. Eastwood and P. G. Wolynes. Role of explicitly cooperative interactions in protein folding funnels: A simulation study. *J Chem Phys*, 114(10):4702–4716, 2001.
32. H. Edelsbrunner. *Algorithms in combinatorial geometry.* Springer-Verlag, Berlin, 1987.
33. B. Fain, Y. Xia, and M. Levitt. Design of an optimal Chebyshev-expanded discrimination function for globular proteins. *Protein Sci.*, 11:2010–2021, 2002.
34. A. V. Finkelstein, A. Ya. Badretdinov, and A. M. Gutin. Why do protein architectures have boltzmann-like statistics? *Proteins.*, 23(2):142–50, 1995.
35. M. S. Friedrichs and P. G. Wolynes. Toward protein tertiary structure recognition by means of associative memory hamiltonians. *Science*, 246:371–373, 1989.
36. H. H. Gan, A. Tropsha, and T. Schlick. Lattice protein folding with two and four-body statistical potentials. *Proteins*, 43(2):161–174, 2001.
37. D. Gessmann, F. Mager, H. Naveed, T. Arnold, S. Weirich, D. Linke, J. Liang, and S. Nussberger. Improving the resistance of a eukaryotic beta-barrel protein to thermal and chemical perturbations. *J Mol Biol*, 413(1):150–161, Oct 2011.
38. S Geula, H Naveeds, J Liangs, and V Shoshan-Barmatz. Structure-based analysis of VDAC1 protein: Defining oligomer contact sites. *Journal of Biological Chemistry*, 287(3):2179–2190, 2012.
39. D. Gilis and M. Rooman. Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. *J Mol Biol*, 257(5):1112–26, 1996.
40. D. Gilis and M. Rooman. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J Mol Biol*, 272(2):276–90, 1997.
41. A. Godzik, A. Kolinski, and J. Skolnick. Topology fingerprint approach to the inverse protein folding problem. *J Mol Biol*, 227(1):227–238, 1992.
42. A. Godzik and J. Skolnick. Sequence-structure matching in globular proteins: application to supersecondary and tertiary structure determination. *Proc Natl Acad Sci*, 89(24):12098–102, 1992.
43. R. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes. Protein tertiary structure recognition using optimized hamiltonians with local interactions. *Proc. Natl. Acad. Sci. USA*, 89:9029–9033, 1992.
44. R. Guerois, J. E. Nielsen, and L. Serrano. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*, 320(2):369–87, 2002.
45. M. H. Hao and H. A. Scheraga. How optimization of potential functions affects protein folding. *Proc. Natl. Acad. Sci.*, 93(10):4984–89, 1996.
46. M-H. Hao and H. A. Scheraga. Designing potential energy functions for protein folding. *Curr. Opin. Struct. Biol.*, 9:184–188, 1999.
47. M-H. Hao and HA. Scheraga. Designing potential energy functions for protein folding. *Curr Opinion Structural Biology*, 9:184–188, 1999.
48. R. B. Hill, D. P. Raleigh, A. Lombardi, and W. F. DeGrado. *De novo* design of helical bundles as models for understanding protein folding and function. *Acc Chem Res.*, 33:745–754, 2000.

49. C. Hoppe and D. Schomburg. Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential. *Protein Sci*, 14:2682–92, 2005.

50. C. Hu, X. Li, and J. Liang. Developing optimal non-linear scoring function for protein design. *Bioinformatics*, 20(17):3080–98, 2004.

51. R. Jackups Jr and J. Liang. Interstrand pairing patterns in beta-barrel membrane proteins: the positive-outside rule, aromatic rescue, and strand registration prediction. *J Mol Biol*, 354(4):979–93, 2005.

52. R. Janicke. Folding and association of proteins. *Prog. Biophys. Mol. Biol.*, 49:117–237, 1987.

53. J. Janin, K. Henrick, J. Moult, L. T. Eyck, M. J. Sternberg, S. Vajda, I. Vakser, and S. J. Wodak. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins*, 52(1):2–9, 2003.

54. R. L. Jernigan and I. Bahar. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.*, 6:195–209, 1996.

55. Ronald Jackups Jr. and Jie Liang. Combinatorial analysis for sequence and spatial motif discovery in short sequence fragments. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7(3), 2010.

56. N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

57. M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, pages 631–639, 1990.

58. J. Khatun, S. D. Khare, and N. V. Dokholyan. Can contact potentials reliably predict stability of proteins? *J. Mol. Biol.*, 336:1223–1238, 2004.

59. J. A. Kocher, M. J. Rooman, and S. J. Wodak. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.*, 235:1598–1613, 1994.

60. P. Koehl and M. Levitt. *De Novo* protein design. I. In search of stability and specificity. *J. Mol. Biol.*, 293:1161–1181, 1999.

61. P. Koehl and M. Levitt. *De Novo* protein design. II. Plasticity of protein sequence. *J. Mol. Biol.*, 293:1183–1193, 1999.

62. K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes. Self-consistently optimized statistical mechanical energy functions for sequence structure alignment. *Protein Sci*, 5:1043–1059, 1996.

63. K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes. Self-consistently optimized energy functions for protein structure prediction by molecular dynamics. *Proc. Natl. Acad. Sci.*, 95(6):2932–7, 1998.

64. T. Kortemme and D. Baker. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci*, 99:14116–21, 2002.

65. T. Kortemme, D. E. Kim, and D. Baker. Computational alanine scanning of protein-protein interfaces. *Sci STKE*, 2004:pl2, 2004.

66. T. Kortemme, A. V. Morozov, and D. Baker. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol*, 326:1239–59, 2003.

67. B. Krishnamoorthy and A. Tropsha. Development of a four-body statistical pseudo-potential to discriminate native from non-native protein conformations. *Bioinformatics*, 19(12):1540–8, 2003.

68. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, pages 1364–8, 2003.
69. T. Lazaridis and M. Karplus. Effective energy functions for protein structure prediction. *Curr Opin Struct Biol*, 10:139–145, 2000.
70. K. H. Lee, D. Xie, E. Freire, and L. M. Amzel. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins*, 20:68–84, 1994.
71. C. M. R. Lemer, M. J. Rooman, and S. J. Wodak. Protein-structure prediction by threading methods - evaluation of current techniques. *Proteins*, 23:337–355, 1995.
72. M. Levitt and A. Warshel. Computer simulation of protein folding. *Nature.*, 253:694–8, 1975.
73. H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273:666–669, 1996.
74. H. Li, C. Tang, and N. S. Wingreen. Nature of driving force for protein folding: A result from analyzing the statistical potential. *Phs. Rev. Lett.*, 79:765–768, 1997.
75. X. Li, C. Hu, and J. Liang. Simplicial edge representation of protein structures and alpha contact potential with confidence measure. *Proteins*, 53:792–805, 2003.
76. X. Li and J. Liang. Computational design of combinatorial peptide library for modulating protein-protein interactions. *Pacific Symposium of Biocomputing*, pages 28–39, 2005.
77. X. Li and J. Liang. Geometric cooperativity and anti-cooperativity of three-body interactions in native proteins. *Proteins*, 60:46–65, 2005.
78. J. Liang and K. A. Dill. Are proteins well-packed? *Biophys. J.*, 81:751–766, 2001.
79. J. Liang, H. Naveed, D. Jimenez-Morales, L. Adamian, and M. Lin. Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochimica et Biophysica Acta - Biomembranes*, 1818(4):927–941, 2012.
80. S. Liu, C. Zhang, H. Zhou, and Y. Zhou. A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, 56:93–101, 2004.
81. L. L. Looger, M. A. Dwyer, J. J. Smith, and H. W. Hellinga. Computational design of receptor and sensor proteins with novel functions. *Nature*, 423:185–190, 2003.
82. H. Lu and J. Skolnick. A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins*, 44:223–232, 2001.
83. V. N. Maiorov and G. M. Crippen. Contact potential that pecognizes the correct folding of globular proteins. *J. Mol. Biol.*, 227:876–888, 1992.
84. B. J. McConkey, V. Sobolev, and M. Edelman. Discrimination of native protein structures using atom-atom contact scoring. *Proc. Natl. Acad. Sci. U.S.A.*, 100(6):3215–20, 2003.
85. C. S. Mészáros. Fast Cholesky factorization for interior point methods of linear programming. *Comp. Math. Appl.*, 31:49 – 51, 1996.
86. C. Micheletti, F. Seno, J. R. Banavar, and A. Maritan. Learning effective amino acid interactions through iterative stochastic techniques. *Proteins*, 42(3):422–431, 2001.
87. L. A. Mirny and E. I. Shakhnovich. How to derive a protein folding potential? a new approach to an old problem. *J. Mol. Biol.*, 264:1164–1179, 1996.

88. B. O. Mitchell, R. A. Laskowski, A. Alex, and J. M. Thornton. BLEEP: potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comp. Chem.*, 20:1177–85, 1999.

89. S. Miyazawa and R. L. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, 18:534–552, 1985.

90. S. Miyazawa and R. L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term. *J. Mol. Biol.*, 256:623–644, 1996.

91. S. Miyazawa and R. L. Jernigan. How effective for fold recognition is a potential of mean force that includes relative orientations between contacting residues in proteins? *The Journal of Chemical Physics*, 122:024901, 2005.

92. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.*, 79(22):2361–2381, 1975.

93. I. Muegge and Y. C. Martin. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem*, 42:791–804, 1999.

94. P. J. Munson and R. K. Singh. Statistical significane of hierarchical multi-body potential based on delaunay tessellation and their application in sequence-structure alignment. *Protein Sci*, 6:1467–1481, 1997.

95. H Naveed, R Jackups, Jr, and J Liang. Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):12735–12740, 2009.

96. Hammad Naveed, Yun Xu, Ronald Jackups, and Jie Liang. Predicting three-dimensional structures of transmembrane domains of Îš-barrel membrane proteins. *Journal of the American Chemical Society*, 134(3):1775–1781, January 2012.

97. K. Nishikawa and Y. Matsuo. Development of pseudoenergy potentials for assessing protein 3-D-1-D compatibility and detecting weak homologies. *Protein Eng.*, 6:811–820, 1993.

98. C. H. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity.* Dover, 1998.

99. Y. Park, M. Elsner, R. Staritzbichler, and V. Helms. Novel scoring function for modeling structures of oligomers of transmembrane alpha-helices. *Proteins*, 57(3):577–85, 2004.

100. J. A. Rank and D. Baker. A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci*, 6(2):347–354, 1997.

101. C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker. Protein structure prediction using rosetta. *Methods Enzymol*, 383:66–93, 2004.

102. Carol A. Rohl, Charlie E. M. Strauss, Kira M. S. Misura, and David Baker. *Protein Structure Prediction Using Rosetta*, volume 383 of *Methods in Enzymology*, pages 66–93. Elsevier, Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA., 2004.

103. A. Rossi, C. Micheletti, F. Seno, and A. Maritan. A self-consistent knowledge-based approach to protein design. *Biophys J*, 80(1):480–490, 2001.

104. W. P. Russ and R. Ranganathan. Knowledge-based potential functions in protein design. *Curr. Opin. Struct. Biol.*, 12:447–452, 2002.

105. K. Sale, J.L. Faulon, G.A. Gray, J.S. Schoeniger, and M.M. Young. Optimal bundling of transmembrane helices using sparse distance constraints. *Protein Sci*, 13(10):2613–27, 2004.

106. R. Samudrala and J. Moult. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.*, 275:895–916, 1998.

107. B. Schölkopf and A. J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. The MIT Press, Cambridge, MA, 2002.

108. E. I. Shakhnovich. Proteins with selected sequences fold into unique native conformation. *Phys. Rev. Lett.*, 72:3907–3910, 1994.

109. E. I. Shakhnovich and A. M. Gutin. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA.*, 90:7195–7199, 1993.

110. S. Shimizu and H. S. Chan. Anti-cooperativity in hydrophobic interactions: A simulation study of spatial dependence of three-body effects and beyond. *J Chem Phys*, 115(3):1414–1421, 2001.

111. S. Shimizu and H. S. Chan. Anti-cooperativity and cooperativity in hydrophobic interactions: Three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Proteins*, 48:15–30, 2002.

112. K. T. Simons, I. Ruczinski, C. Kooperberg, B. Fox, C. Bystroff, and D. Baker. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*, 34:82–95, 1999.

113. K. T. Simons, I. Ruczinski, C. Kooperberg, B. Fox, C. Bystroff, and D. Baker. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*, 34:82–95, 1999.

114. R. K. Singh, A. Tropsha, and I. I. Vaisman. Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues. *J Comput Biol*, 3(2):213–221, 1996.

115. R. K. Singh, A. Tropsha, and I. I. Vaisman. Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino-acid residues. *J. Comp. Bio.*, 3:213–221, 1996.

116. M. J. Sippl. calculation of conformational ensembles from potentials of the main force. *J. Mol. Biol.*, 213:167–180, 1990.

117. M. J. Sippl. Boltzmann's principle, knowledge-based mean fields and protein folding. an approach to the computational determination of protein structures. *J Comput Aided Mol Des.*, 7(4):473–501, 1993.

118. M. J. Sippl. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.*, 5(2):229–235, 1995.

119. S. Tanaka and H. A. Scheraga. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, 9:945–950, 1976.

120. S. Tanaka and H. A. Scheraga. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, 9:945–950, 1976.

121. P. D. Thomas and K. A. Dill. An iterative method for extracting energy-like quantities from protein structures. *PNAS*, 93(21):11628–33, 1996.

122. P. D. Thomas and K. A. Dill. Statistical potentials extracted from protein structures: How accurate are they? *J. Mol. Biol.*, 257:457–469, 1996.

123. D. Tobi, G. Shafran, N. Linial, and R. Elber. On the design and analysis of protein folding potentials. *Proteins*, 40:71–85, 2000.

124. D. Tobi, G. Shafran, N. Linial, and R. Elber. On the design and analysis of protein folding potentials. *Proteins*, 40:71–85, 2000.

125. R. J. Vanderbei. *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, 1996.

126. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y., 1995.

127. V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, 25, 1964.

128. V. Vapnik and A. Chervonenkis. *Theory of Pattern Recognition [in Russian]*. Nauka, Moscow, 1974. (German Translation: W. Wapnik & A. Tscherwonenkis, *Theorie der Zeichenerkennung*, Akademie–Verlag, Berlin, 1979).

129. E. Venclovas, A. Zemla, K. Fidelis, and J. Moult. Comparison of performance in successive CASP experiments. *Proteins*, 45:163–170, 2003.

130. M. Vendruscolo and E. Domanyi. Pairwise contact potentials are unsuitable for protein folding. *J. Chem. Phys.*, 109:11101–8, 1998.

131. M. Vendruscolo, R. Najmanovich, and E. Domany. Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading? *Proteins*, 38:134–148, 2000.

132. M. Vendruscolo, R. Najmanovich, and E. Domany. Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading? *Proteins: Structure, Function, and Genetics*, 38:134–148, 2000.

133. S. J. Wodak and M. J. Rooman. Generating and testing protein folds. *Curr. Opin. Struct. Biol.*, 3:247–259, 1993.

134. P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–20, 1995.

135. Y. Xia and M. Levitt. Extracting knowledge-based energy functions from protein structures by error rate minimization: Comparison of methods using lattice model. *J Chem. Phys.*, 113:9318–9330, 2000.

136. D. Xu, S. L. Lin, and R. Nussinov. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J Mol Biol*, 2651:68–84, 1997.

137. Y. Xu, C. Hu, Y. Dai, and J. Liang. On simplified global nonlinear function for fitness landscape: A case study of inverse protein folding. *Plos One*, submitted.

138. C. Zhang and S. H. Kim. Environment-dependent residue contact energies for proteins. *PNAS*, 97(6):2550–2555, 2000.

139. C. Zhang, S. Liu, H. Zhou, and Y. Zhou. An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state. *Protein Sci*, 13:400–411, 2004.

140. C. Zhang, S. Liu, H. Zhou, and Y. Zhou. The dependence of all-atom statistical potentials on structural training database. *Biophys J*, 86(6):3349–58, 2004.

141. C. Zhang, S. Liu, and Y. Zhou. The dependence of all-atom statistical potentials on training structural database. *Biophys. J.*, 86:3349–3358, 2004.

142. C. Zhang, S. Liu, Q. Zhu, and Y. Zhou. A knowledge-based energy function for protein-ligand, protein-protein, and protein-dna complexes. *J Med Chem*, 48:2325–35, 2005.

143. C. Zhang, G. Vasmatzis1, J. L. Cornette, and C. DeLisi. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, 267, 1997.

144. W. Zheng, S. J. Cho, I. I. Vaisman, and A. Tropsha. A new approach to protein fold recognition based on delaunay tessellation of protein structure. *Pac Symp Biocomput*, pages 486–497, 1997.

145. W. Zheng, S. J. Cho, I. I. Vaisman, and A. Tropsha. A new approach to protein fold recognition based on Delaunay tessellation of protein structure. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing'97*, pages 486–497, Singapore, 1997. World Scientific.

146. H. Y. Zhou and Y. Q. Zhou. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.*, 11:2714–26, 2002.

EXERCISES

2.1 To capture higher order interactions in proteins, one can construct the three-body propensity function. The propensity $P(i, j, k)$ for residues of type i, j, k to interact can be modeled as the odds ratio of the observed probability $q(i, j, k)$ of a three-body (triple) atomic contacts involving residue i, j, and k, and the expected probability $p(i, j, k)$ $P(i, j, k) \equiv \frac{q(i,j,k)}{p(i,j,k)}$. To compute the observed probability $q(i, j, k)$, we can use: $q(i, j, k) = a(i, j, k)/\sum_{i',j',k'} a(i', j', k')$, where $a(i, j, k)$ is the number count of atomic contacts among residue types i, j and k, and $\sum_{i',j',k'} a(i', j', k')$ is the total number of all atomic three-body contacts. For the random probability $p(i, j, k)$, let us assume it is the probability that three atoms are picked from a residue of type i, a residue of type j, and a residue of type k, when chosen randomly and independently from the pooled database of protein structures. Denote the number of interacting residues of type i as N_i, the number of atoms residue of type i has as n_i, and the total number of interacting atoms as n.

a) Assume all three interacting residues are of different types, *e.g.*, $i \neq j \neq k$, what is the probability that we first pick up an atom from a residue of type i, then an atom from a residue of type j, and with the third atom picked up to be from a residue of type k?

b) Now consider all other possible sequences of picking up an atom each from an i, j, and k residue type. Write down the formula for $p(i, j, k)$.

c) When two of the three interacting residues are of the same type, *i.e.*, $i = j \neq k$, what is the formula for $p(i, j, k)$?

d) When all three residues are of the same type, *i.e.*, $i = j = k$, what is the formula for $p(i, j, k)$?.

2.2 β-barrel membrane proteins are found in a large number of pathogeneic gram-negative bacteria. Their transmembrane (TM) segments are β-strands. We can obtain the empirical propensity $P(X, Y)$ for interacting pairs of residue types X and Y on neighboring β-strands as $P(X, Y) = f_{\text{obs}}(X, Y)/\mathbb{E}[f(X, Y)]$, where $f_{\text{obs}}(X, Y)$ is the observed count of X-Y contacts in the strand pair, and $\mathbb{E}[f(X, Y)]$ is the expected count of X-Y contacts in a null model.