Bibliography

[1] R. A. Abagyan and S. Batalov. Do aligned sequences share the same fold? *J. Mol. Biol.*, 273:355–368, 1997.

[2] N. N. Alexandrov, R. Nussinov, and R. Zimmer. Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In L. Hunter and T. E. Klein, editors, *Pacific Symposium on Biocomputing*, pages 53–72, SingaporeSingapore, 1996. World Scientific Publishing Co.

[3] N. N. Alexandrov, K. Takahashi, and N. Go. Common spatial arrangements of backbone fragments in homologous and non-homologous proteins. *J Mol Biol*, 225(1):5–9, May 1992.

[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215:403–410, 1990.

[5] C. A. F. Andersen, A. G. Palmer, S. Brunak, and B. Rost. Continuum secondary structure captures protein flexibility. *Structure*, 10(2):175–184, Feb 2002.

[6] C. A. F. Andersen and B. Rost. Secondary structure assignment. *Methods Biochem Anal*, 44:341–363, 2003.

[7] A. Andreeva and A. G. Murzin. Evolution of protein fold in the presence of functional constraints. *Curr Opin Struct Biol*, 16(3):399–408, Jun 2006.

[8] A. Andreeva, A. Prlic, T. J. P. Hubbard, and A. G. Murzin. Sisyphus–structural alignments for proteins with non-trivial relationships. *Nucleic Acids Res*, 35(Database issue):D253–D259, Jan 2007.

[9] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(96):223–230, Jul 1973.

[10] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*, 47:1309–1314, Sep 1961.

[11] M. Arita. Graph modeling of metabolism. *J. Jap. Soc. Artific. Intell.*, 15:703–710, 2000.

[12] P. J. Artymiuk, H. M. Grindley, A. R. Poirette, D. W. Rice, E. C. Ujah, and P. Willett. Identification of beta sheet motifs, of psi loops, and of

patterns of amino acid residues in threedimensional protein structures using a subgraph isomorphism algorithm. *J. Chem. Inf. Comput. Sci.*, 229:707–721, 1994.

[13] P. J. Artymiuk, A. R. Poirrette, H. M. Grindley, D. W. Rice, and P. Willett. A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. *J Mol Biol*, 243(2):327–344, Oct 1994.

[14] S. Ash, A. Cline, R. W. Homer, T. Hurst, and G. B. Smith. SYBIL line notation (SLN): A versatile language for chemical structure representation. *J. Chem. Inf. Sci.*, 37:71–79, 1997.

[15] A. Aszodi and W. Taylor. Folding polypeptide α-carbon backbones by distance geometry methods. *Biopolymers*, 34:489–506, 1994.

[16] J. F. Atkins and R. F. Gesteland. The twenty-first amino acid. *Nature*, 407(6803):463, 465, Sep 2000.

[17] L. Babel and G. Tinhofer. A branch and bound algorithm for the maximum clique problem. *Methods Models Oper. Res.*, 34:207–217, 1990.

[18] R. Backofen, S. Will, and P. Clote. Algorithmic approach to quantifying the hydrophobic force contribution in protein folding. *Pac Symp Biocomput*, pages 95–106, 2000.

[19] E. N. Baker and R. E. Hubbard. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol*, 44(2):97–179, 1984.

[20] E. Balas and C. S. Yu. Finding a maximum clique in an arbitrary graph. *SIAM J. Comput.*, 15:1054–1068, 1986.

[21] U. Bastolla, J. Farwer, E.W. Knapp, and M. Vendruscolo. How to guarantee optimal stability for most representative structures in protein data bank. *Proteins*, 44:79–96, 2001.

[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shyndyalov, and P.E. Bourne. The protein data bank. *Nucl. Acids Res*, 28:235–242, 2000.

[23] Y. E. Bessonov. On the solution of a problem on the search for the best intersection of graphs on the basis of an analysis of the projections of the subgraphs of the modular product. *Vychisl. Sistemy*, 121:322 (in Russian), 1985.

[24] M. R. Betancourt and J. Skolnick. Universal similarity measure for comparing protein structures. *Biopolymers*, 59(5):305–309, Oct 2001.

[25] J. R Bienkowska, H. Hartman, and T. F. Smith. A search method for homologs of small proteins. ubiquitin-like proteins in prokaryotic cells? *Protein Eng*, 16(12):897–904, Dec 2003.

[26] F. Birzele, J. E Gewehr, G. Csaba, and R. Zimmer. Vorolign–fast structural alignment using voronoi contacts. *Bioinformatics*, 23(2):e205–e211, Jan 2007.

[27] D. N. Boobbyer, P. J. Goodford, P. M. McWhinnie, and R. C. Wade. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J Med Chem*, 32(5):1083–1094, May 1989.

[28] D. Bordo and P. Argos. The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. *J Mol Biol*, 243(3):504–519, Oct 1994.

[29] J. U. Bowie, R. Lthy, and D. Eisenberg. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016):164–170, Jul 1991.

[30] C. Brandon and J Tooze. *Introduction to Protein Structure*. London, New York: Garland, 1999.

[31] C. Bron and J. Kerbosch. Algorithm 457 - finding all cliques of an undirected graph. *Commun. ACM*, 16:575–577, 1973.

[32] R. D. Brown, G. Jones, and P. Willett. Matching two-dimensional chemical graphs using genetic algorithms. *J. Chem. Inf. Comput. Sci.*, 34:63–70, 1994.

[33] R. S. Cahn, C. K. Ingold, and V. Prelog. Spezifikation der molekularen chiralitt. *Angew. Chem.*, 78:413–447, 1966.

[34] O. Camoglu, T. Kahveci, and A. K. Singh. Index-based similarity search for protein structure databases. *J Bioinform Comput Biol*, 2(1):99–126, Mar 2004.

[35] I. D. Campbell and A. K. Downing. Building protein structure and function from modular units. *Trends Biotechnol*, 12(5):168–172, May 1994.

[36] A. Caprara, R. Carr, S. Istrail, G. Lancia, and G. Walenz. 1001 optimal PDB structure alignments: integer programming methods for finding the maximum contact map overlap. *J. Comp. Biol.*, 11(1):27–52, 2004.

[37] M. Carpentier, S. Brouillet, and J. Pothier. YAKUSA: A fast structural database scanning method. *Proteins*, 61:137–151, 2005.

[38] B. Carr, W. Hart, N. Krasnogor, J. Hirst, E. Burke, and J. Smith. Alignment of protein structures with a memetic evolutionary algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference - Gecco 2002*, pages 1027–1034. Morgan Kaufmann, 2002.

[39] P. Carter, C. A. F. Andersen, and B. Rost. Dsspcont: Continuous secondary structure assignments for proteins. *Nucleic Acids Res*, 31(13):3293–3295, Jul 2003.

[40] O. Carugo and S. Pongor. Protein fold similarity estimated by a probalistic approach based on cα-cα distance comparisons. *J. Mol. Biol.*, 315:887–898, 2002.

[41] J.M. Chandonia, N.S. Walker, L. Lo Conte, P. Koehl, M. Levitt, and S.E. Brenner. ASTRAL compendium enhancements. *Nucl. Acids Res.*, 30:264–267, 2002.

[42] S. Cheek, K. Ginalski, H. Zhang, and N. V. Grishin. A comprehensive update of the sequence and structure classification of kinases. *BMC Structural Biology*, 5:6, 2005.

[43] L. Chen and W. Robien. Application of the maximal common substructure algorithm to automatic interpretation of 13c-nmr spectra. *J. Chem. Inf. Comput.Sci.*, 34:934–941, 1994.

[44] Y. N. Chirgadze. Deduction and systematic classification of spatial motifs of the anti-parallel beta structure in globular proteins. *Acta Cryst.*, A43:405–417, 1987.

[45] C. Chothia. Principles that determine the structure of proteins. *Annu Rev Biochem*, 53:537–572, 1984.

[46] C. Chothia. Proteins. one thousand families for the molecular biologist. *Nature*, 357(6379):543–544, Jun 1992.

[47] C. Chothia and A. V. Finkelstein. The classification and origins of protein folding patterns. *Annu Rev Biochem*, 59:1007–1039, 1990.

[48] C. Chothia, M. Levitt, and D. Richardson. Structure of proteins: packing of alpha-helices and pleated sheets. *Proc Natl Acad Sci U S A*, 74(10):4130–4134, Oct 1977.

[49] M. Comin, C. Guerra, and G. Zanotti. Proust: a comparison method of three-dimensional structures of proteins using indexing techniques. *J Comput Biol*, 11(6):1061–1072, 2004.

[50] M. Cone, R. Venkataraghavan, and F. McLafferty. Molecular structure comparison program for the identification of maximal common substructures. *J. Am. Chem. Soc.*, 99:7668–7671, 1977.

[51] B. Contreras-Moreira, P. W. Fitzjohn, and P.A. Bates. In silico protein recombination: enhancing template and sequence alignment selection for comparative protein modelling. *J. Mol. Biol.*, 328(2):593–608, 2003.

[52] M. V. Cubellis, F. Cailliez, and S. C. Lovell. Secondary structure assignment that accurately reflects physical and evolutionary characteristics. *BMC Bioinformatics*, 6 Suppl 4:S8, Dec 2005.

[53] P. Dafas, D. Bolser, J. Gomoluch, J. Park, and M. Schroeder. Using convex hulls to extract interaction interfaces from known structures. *Bioinformatics*, 20(10):1486–1490, Jul 2004.

[54] T. Dandekar and P. Argos. Potential of genetic algorithms in protein folding and protein engineering simulations. *Protein Eng.*, 5:637–645, 1992.

[55] L. Davis. *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, 1991.

[56] S. de Bono, L. Riechmann, E. Girard, R. L. Williams, and G. Winter. A segment of cold shock protein directs the folding of a combinatorial protein. *Proc Natl Acad Sci U S A*, 102(5):1396–1401, Feb 2005.

[57] D. A. De Jong and W. M. Spears. Using genetic algorithms to solve NP-complete problems. In *Proceedings of the Third International Conference on Genetic Algorithms*, pages 124–132, 1989.

[58] W.L. DeLano. The pymol molecular graphics system. on World Wide Web `http://www.pymol.org`, 2002.

[59] U. Dengler, A. S. Siddiqui, and G. J. Barton. Protein structural domains: analysis of the 3dee domains database. *Proteins*, 42(3):332–344, Feb 2001.

[60] R. Diamond. On the comparison of conformations using linear and quadratic transformations. *Acta Crystallographica Section A*, 32(1):1–10, Jan 1976.

[61] R. F. Doolittle. The multiplicity of domains in proteins. *Annu Rev Biochem*, 64:287–314, 1995.

[62] O. Dror, H. Benyamini, R. Nussinov, and H. Wolfson. MASS: multiple structural alignment by secondary structures. *Bioinformatics*, 19:i95–i104, 2003.

[63] O. Dror, H. Benyamini, R. Nussinov, and H. Wolfson. Multiple structural alignments by secondary structure: algorithm and applications. *Protein Science*, 12:2492–2507, 2003.

[64] I. Eidhammer, I. Jonassen, and W. R. Taylor. Structure comparison and structure patterns. *J Comput Biol*, 7(5):685–716, 2000.

[65] I. Eidhammer, I. Jonasses, and W. R. Taylor. *Protein bioinformatics*. John Wiley & Sons Ltd., 2004.

[66] V. Escalier, J. Pothier, H. Soldano, and A. Viari. Pairwise and multiple identification of three-dimensional common substructures in proteins. *J Comput Biol*, 5(1):41–56, 1998.

[67] C. Ferrari and C. Guerra. Geometric methods for protein structure comparison. *LNCS*, 2666:5782, 2003.

[68] A. V. Finkelstein and O. B. Ptitsyn. Why do globular proteins fit the limited set of folding patterns? *Prog Biophys Mol Biol*, 50(3):171–190, 1987.

[69] D. Fischer, O. Bachar, R. Nussinov, and H. Wolfson. An efficient automated computer vision based technique for detection of three dimensional structural motifs in proteins. *J Biomol Struct Dyn*, 9(4):769–789, Feb 1992.

[70] D. Fischer and D. Eisenberg. Protein fold recognition using sequence-derived predictions. *Proteins*, 5:947–955, 1996.

[71] T. P. Flores, D. S. Moss, and J. M. Thornton. An algorithm for automatically generating protein topology cartoons. *Protein Eng*, 7(1):31–37, Jan 1994.

[72] D. R. Flower. Automating the identification and analysis of protein beta-barrels. *Protein Eng*, 7(11):1305–1310, Nov 1994.

[73] D. R. Flower. A topological nomenclature for protein structure. *Protein Engineering*, 11(9):723–727, 1998.

[74] D. Frishman and P. Argos. Knowledge-based secondary structure assignment. *Proteins*, 23:566–579, 1995.

[75] C. Frömmel, C. Gille, A. Goede, C. Gröpl, S. Hougardy, T. Nierhoff, R. Preissner, and M. Thimm. Accelerating screening of 3d protein data with a graph theoretical approach. *Bioinformatics*, 19(18):2442–2447, Dec 2003.

[76] M. R. Garey and D. S. Johnson. *Computers and Intractability: a guide to the theory of NP-completeness*. WH. Freeman and Co., 1979.

[77] F. Gavril. Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of chordal graphs. *SIAM J. Comput.*, 1:180–187, 1972.

[78] M. Gerstein and M. Levitt. Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci*, 7(2):445–456, Feb 1998.

[79] J.F. Gibrat, T. Madej, and S.H. Bryant. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.*, 6:377–385, 1996.

[80] E. Gifford, M. Johnson, D. Smith, and C. Tsai. Structure-reactivity maps as a tool for visualizing xenobiotic structure-reactivity relationships. *Network Science*, 2:1–33, 1996.

[81] D. Gilbert, D. Westhead, J. Viksna, and J. Thornton. A computer system to perform structure comparison using TOPS representations of protein structure. *Computers & Chemistry*, 26:20–23, 2001.

[82] C. Gille, A. Goede, R. Preissner, K. Rother, and C. Frömmel. Conservation of substructures in proteins: interfaces of secondary structural elements in proteasomal subunits. *J Mol Biol*, 299(4):1147–1154, Jun 2000.

[83] A. Godzik. The structural alignment between two proteins: is there a unique answer? *Protein Sci*, 5(7):1325–1338, Jul 1996.

[84] A. Godzik, A. Kolinski, and J. Skolnick. Topology fingerprint approach to the inverse protein folding problem. *J Mol Biol*, 227(1):227–238, Sep 1992.

[85] A. Godzik and J. Skolnick. Flexible algorithm for direct multiple alignment of protein structures and sequences. *CABIOS*, 10(6):587–596, 1994.

[86] A. Godzik, J. Skolnick, and A. Kolinski. Regularities in interaction patterns of globular proteins. *Protein Eng*, 6(8):801–810, Nov 1993.

[87] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addision-Wesley Publishing Co., 1989.

[88] D. Goldman, S. Istrail, and C. Papadimitriou. Algorithmic aspects of protein structure similarity. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 512–522, 1999.

[89] M. Golumbic. *Algorithmic graph theory and perfect graphs.* Academic Press New York, 1980.

[90] S. Govindarajan and R. A. Goldstein. Why are some proteins structures so common? *Proc Natl Acad Sci U S A*, 93(8):3341–3345, Apr 1996.

[91] I. V. Grigoriev, A. A. Mironov, and A. B. Rakhmaninova. Inter-helical contacts determing the architecture of alpha-helical globular proteins. *J. of Biomolecular Structure and Dynamics*, 12:559–572, 1994.

[92] H.M. Grindley, P.J. Artymiuk, D.W. Rice, and P. Willett. Identification of tertiary structure resemblance in proteins using a maximal common subgraph algorithm. *Journal of Molecular Biology*, 299:707–721, 1993.

[93] N. V. Grishin, Y. I. Wolf, and E. V. Koonin. From complete genomes to measures of substitution rate variability within and between proteins. *Genome Res*, 10(7):991–1000, Jul 2000.

[94] E. J. Gumbel. *Statistics of Extreme.* Columbia University Press, 1958.

[95] P. Güntert, C. Mumenthaler, and K. Wüthrich. Torsion angle dynamics for nmr structure calculation with the new program DYANA. *J Mol Biol*, 273(1):283–298, Oct 1997.

[96] C. Hadley and D. T. Jones. A systematic comparison of protein structure classifications: Scop, cath and fssp. *Structure*, 7(9):1099–1112, Sep 1999.

[97] M. Hahn, K. Piotukh, R. Borriss, and U. Heinemann. Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis. *Proc Natl Acad Sci U S A*, 91(22):10417–10421, Oct 1994.

[98] A. Harrison, F. Pearl, I. Sillitoe, T. Slidel, R. Mott, J. Thornton, and C. Orengo. Recognizing the fold of a protein structure. *Bioinformatics*, 19(14):1748–1759, 2003.

[99] A. R. Hawkins and H. K. Lamb. The molecular biology of multidomain proteins. selected examples. *Eur J Biochem*, 232(1):7–18, Aug 1995.

[100] A. Henschel, W. K. Kim, and M. Schroeder. Equivalent binding sites reveal convergently evolved interaction motifs. *Bioinformatics*, 22(5):550–555, Mar 2006.

[101] D. Hind and M. Levitt. How optimization of potential function affects protein folding. *J. Mol. Biol.*, 243:668–682, 1994.

[102] M. Hochstrasser. Evolution and function of ubiquitin-like protein-conjugation systems. *Nat Cell Biol*, 2(8):E153–E157, Aug 2000.

[103] L. Holm and C. Sander. Protein structure comparison by alignment of distance matrices. *J.Mol.Biol.*, 233:123–138, 1993.

[104] L. Holm and C. Sander. 3-D lookup: Fast protein structure searches at 90% reliability. In *Proc. Ann. Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, pages 179–187, 1995.

[105] L. Holm and C. Sander. Mapping the protein universe. *Science*, 273(5275):595–603, Aug 1996.

[106] L. Holm and C. Sander. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Research*, 23:231–234, 1997.

[107] W. L. Hsu, Y. Ikura, and G. L. Nemhauser. A polynomial algorithm for maximum weigthed vertex packings on graphs without long odd cycles. *Math. Programming*, 20:225–232, 1981.

[108] E. G. Hutchinson and J. M. Thornton. HERA - a program to draw schematic diagrams of protein secondary structure. *Proteins*, 8:202–212, 1990.

[109] E. G. Hutchinson and J. M. Thornton. Promotif - a program to identify and analyze structural motifs in proteins. *Protein Science*, 5(2):212–220, 1996.

[110] A. Hvidt and P. Westh. Different views on the stability of protein conformations and hydrophobic effects. *Journal of Solution Chemistry*, 27:395–402(8), 1998.

[111] V. A. Ilyin, A. Abyzov, and C. M. Leslin. Structural alignment of proteins by a novel topofit method, as a superimposition of common volumes at a topomax point. *Protein Sci*, 13(7):1865–1874, Jul 2004.

[112] G. A. Jeffrey and W. Saenger. *Hydrogen Bonding in Biological Structures.* Springer-Verlag, 1994.

[113] L. O. Johannissen and W. R. Taylor. Protein fold comparison by the alignment of topological strings. *Protein Eng*, 16(12):949–955, Dec 2003.

[114] D. T. Jones. Threader : Protein sequence threading by double dynamic programming. In S. Salzberg, D. Searls, and K. Kasif, editors, *Computational Methods in Molecular Biology*, pages 285–312. Elsevier Science, 1998.

[115] D. T. Jones, W. R. Taylor, and J. M. Thornton. A new approach to protein fold recognition. *Nature*, 358(6381):86–89, Jul 1992.

[116] J. Jung and B. Lee. Protein structure alignment using environmental profiles. *Protein Eng*, 13(8):535–543, Aug 2000.

[117] J. Jung and B. Lee. Circularly permuted proteins in the protein structure database. *Protein Sci*, 10(9):1881–1886, Sep 2001.

[118] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1978.

[119] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[120] F. Kaden, I. Koch, and J. Selbig. Knowledge-based prediction of protein structures. *J Theor Biol*, 147(1):85–100, Nov 1990.

[121] T. Kawabata and K. Nishikawa. Protein structure comparison using the markov transition model of evolution. *Proteins*, 41:108–122, 2000.

[122] S. K. Kearsley. On the orthogonal transformation used for structural comparisons. *Acta Cryst*, A45:208, 1989.

[123] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, Mar 1958.

[124] G. J. Kleywegt and T. A. Jones. Detecting folding motifs and similarities in protein structures. *Methods Enzymol.*, 277:525–545, 1997.

[125] I. Koch. *Ein graphentheoretischer Ansatz zum paarweisen und multiplen Vergleich von Proteinstrukturen.* Wissenschaft und Technik Verlag, 1998.

[126] I. Koch. Enumerating all connected maximal common subgraphs in two graphs. *Theoretical Computer Science*, 250:1–30, 2001.

[127] I. Koch, F. Kaden, and J. Selbig. Analysis of protein sheet topologies by graph theoretical methods. *Proteins*, 12:314–323, 1992.

[128] I. Koch and T. Lengauer. Detection of distant structural similarities in a set of proteins using a fast graph-based method. In T. Gaasterland, P. Karp, K. Karplus, C. Ouzounis, C. Sander, and A. Valencia, editors, *Proc. 5th Int. Conf. on Intelligent Systems for Molecular Biology*, pages 167–178. AAAI Press, Menlo Park, CA, 1997.

[129] I. Koch, T. Lengauer, and E. Wanke. An algorithm for finding maximal common sub-topologies in a set of protein structures. *J.Comp.Biol.*, 3(2):289–306, 1996.

[130] B. Kolbeck, P. May, T. Schmidt-Goenner, T. Steinke, and E. W. Knapp. Connectivity independent protein structure alignment: a hierarchical approach. *BMC Bioinformatics*, 7:510, 2006.

[131] R. Kolodny, P. Koehl, and M. Levitt. Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J Mol Biol*, 346(4):1173–1188, Mar 2005.

[132] A. S. Konagurthu, J. C. Whisstock, P. J. Stuckey, and A. M. Lesk. Mustang: a multiple structural alignment algorithm. *Proteins*, 64(3):559–574, Aug 2006.

[133] T. Krell, J. E. Coyle, M. J. Horsburgh, J. R. Coggins, and A. J. Lapthorn. Crystallization and preliminary x-ray crystallographic analysis of shikimate kinase from erwinia chrysanthemi. *Acta Crystallogr D Biol Crystallogr*, 53(Pt 5):612–614, Sep 1997.

[134] E. Krissinel and K. Henrick. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.*, D60:2256–2268, 2004.

[135] F. Kuhl and D. Crippen, G.and Friesen. A combinatorial algorithm for calculating ligand binding. *J. Comp. Chem.*, 5:24–34, 1984.

[136] G. Lancia, R. Carr, B. Walenz, and S. Istrail. 101 optimal PDB structure alignments: a branch-and-cut algorithm for the maximum contact map overlap problem. In *Proc. of the Fifth Annual International Conference on Computational Biology*, pages 193–202. ACM Press, 2001.

[137] G. Lancia and S. Istrail. Protein structure comparison: Algorithms and applications. *LNCS*, 2666:1–33, 2003.

[138] R. H. Lathrop. The protein threading problem with sequence amino acid interaction preferences is NP-complete. *Protein Eng.*, 7(9):1059–1068, 1994.

[139] N. Leibowitz, R. Nussinov, and H. J. Wolfson. Musta–a general, efficient, automated method for multiple structure alignment and detection of common motifs: application to proteins. *J Comput Biol*, 8(2):93–121, 2001.

[140] C. Lemmen and T. Lengauer. Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des*, 14(3):215–232, Mar 2000.

[141] A. M. Lesk. Extraction of geometrically similar substructures: least-squares and chebyshev fitting and the difference distance matrix. *Proteins*, 33(3):320–328, Nov 1998.

[142] A. M. Lesk. *Introduction to protein architecture.* Oxford University Press, 2001.

[143] A. M. Lesk and C. Chothia. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol*, 136(3):225–270, Jan 1980.

[144] G. Levi. A note on the derivation of maximal common subgraphs of two directed or undirected graphs. *Calcolo*, 9:341352, 1972.

[145] M. Levitt and C. Chothia. Structural patterns in globular proteins. *Nature*, 261(5561):552–558, Jun 1976.

[146] M. Levitt and M. Gerstein. A unified statistical framework for sequence comparison and structure comparison. *Proc Natl Acad Sci U S A*, 95(11):5913–5920, May 1998.

[147] W. Li, L. Jaroszewski, and A. Godzik. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*, 17:282–283, 2001.

[148] Y. Lindqvist and G. Schneider. Circular permutations of natural protein sequences: structural evidence. *Curr Opin Struct Biol*, 7(3):422–427, Jun 1997.

[149] G. Lu. TOP: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallogr.*, 33:176–183, 2000.

[150] H. Lu and J. Skolnick. A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins*, 44(3):223–232, Aug 2001.

[151] C. B. Lucasius, M. J. Blommers, L. M. C. Buydens, and G. Kateman. A Genetic Algorithm for Conformational Analysis of DNA. In L. Davis, editor, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.

[152] T. Madej, J. F. Gibrat, and S. H. Bryant. Threading a database of protein cores. *Proteins*, 23(3):356–369, Nov 1995.

[153] D. Madsen and G. J. Kleywegt. Interactive motif and fold recognition in protein structure. *J. Appl. Crystallogr.*, 85:137–139, 2002.

[154] V. N. Maiorov and G. M. Crippen. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol*, 235(2):625–634, Jan 1994.

[155] A. C. Martin. The ups and downs of protein topology; rapid comparison of protein structure. *Protein Eng*, 13(12):829–837, Dec 2000.

[156] A. C. W. May and M. S. Johnson. Improved genetic algorithm-based protein structure comparisons: pairwise and multiple superpositions. *Protein Eng.*, 8:873–882, 1995.

[157] P. May, S. Barthel, and I. Koch. PTGL–a web-based database application for protein topologies. *Bioinformatics*, 20(17):3277–3279, Nov 2004.

[158] P. May and T. Steinke. THESEUS - protein structure prediction at ZIB. *ZIB Report*, 06-24:1–34, 2006.

[159] J. McGregor and Willett P. Use of a maximal common subgraph algorithm in the automatic identification of the ostensible bond changes occurring in chemical reactions. *J. Chem. Inf. Comput.Sci.*, 21:137–140, 1981.

[160] A. D. McLachlan. Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.*, 128:49–79, 1979.

[161] L. Mirny and E. Domany. Protein fold recognition and dynamics in space of contact maps. *Proteins*, 26:391–410, 1996.

[162] E. M. Mitchell, P. J. Artymiuk, D. W. Rice, and P. Willett. Use of techniques derived from graph theory to compare secondary structure motifs in proteins. *J. Mol. Biol.*, 212:151–166, 1990.

[163] S. Miyazawa and R. L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol*, 256(3):623–644, Mar 1996.

[164] K. Mizuguchi and N. Go. Comparison of spatial arrangements of secondary structural elements in proteins. *Protein Eng*, 8(4):353–362, Apr 1995.

[165] K. Mizuguchi and N. Go. Seeking significance in three-dimensional protein structure comparisons. *Curr Opin Struct Biol*, 5(3):377–382, Jun 1995.

[166] P. Moscato. Memetic algorithms: A short introduction. In D. Corne and et al., editors, *New Ideas in Optimization*, pages 219–234. McGraw Hill, England, 1999.

[167] J. Moult. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.*, 15:285–289, 2005.

[168] A. G. Murzin. How far divergent evolution goes in proteins. *Curr Opin Struct Biol*, 8(3):380–387, Jun 1998.

[169] A.G. Murzin, S.E. Brenner, T. Hubbard, and C. Chothia. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247:536–540, 1995.

[170] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48:443–453, 1970.

[171] G. Neshich, R. C. Togawa, A. L. Mancini P. R. Kuser, M. E. B. Yamagishi, G. Pappas, W. V. Torres, T. Fonseca e Campos, L. L. Ferreira, F.M. Luna, A.G. Oliveira, R.T. Miura, M.K. Inoue, L.G. Horita, D.F. de Souza, F. Dominiquini, A. Alvaro, C.S. Lima, F. O. Ogawa, G.B. Gomes, J.F. Palandrani, G.F. dos Santos, E.M. de Freitas, A.R. Mattiuz, I.C. Costa, C.L. de Almeida, S. Souza, C. Baudet, and R.H. Higa. Sting millennium: A web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Res*, 31(13):3386–3392, Jul 2003.

[172] M. Novotny, D. Madsen, and G. J. Kleywegt. Evaluation of protein fold comparison servers. *Proteins*, 54:260–270, 2004.

[173] R. Nussinov and H. J. Wolfson. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. *Proc Natl Acad Sci U S A*, 88(23):10495–10499, Dec 1991.

[174] S. C. Nyburg and C. H. Faerman. A revision of van der waals atomic radii for molecular crystals: N, o, f, s, cl, se, br and i bonded to carbon. *Acta Cryst.*, B41:274–279, 1985.

[175] C. A. Orengo, D. T. Jones, and J. M. Thornton. Protein superfamilies and domain superfolds. *Nature*, 372(6507):631–634, Dec 1994.

[176] C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J.M. Thornton. CATH - a hierachic classification of protein domain structures. *Structure*, 28(1):1093–1108, 1997.

[177] C. A. Orengo, M. B. Swindells, A. D. Michie, M. J. Zvelebil, P. C. Driscoll, M. D. Waterfield, and J. M. Thornton. Structural similarity between the pleckstrin homology domain and verotoxin: the problem of measuring and evaluating structural similarity. *Protein Sci*, 4(10):1977–1983, Oct 1995.

[178] C. A. Orengo and W. R. Taylor. SSAP: sequential structure alignment program for protein structure comparison. *Methods Enzymol.*, 266:617–635, 1996.

[179] A. R. Ortiz, C. E. Strauss, and O. Olmea. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Science*, 11:2606–2621, 2002.

[180] L. Pauling and R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci U S A*, 37(5):251–256, May 1951.

[181] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A*, 37(4):205–211, Apr 1951.

[182] A. W. R. Payne and R. C. Glen. Molecular recognition using a binary genetic search algorithm. *J. Mol. Graphics*, 11:74–91, 1993.

[183] S. G. Peisajovich, L. Rockah, and D. S. Tawfik. Evolution of new protein topologies through multistep gene rearrangements. *Nat Genet*, 38(2):168–174, Feb 2006.

[184] M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21:11051120, 1999.

[185] G. A. Petsko and D. Ringe. *Protein structure and function.* New Science Press Ltd., 2004.

[186] C. P. Ponting and R. B. Russell. Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem Sci*, 20(5):179–180, May 1995.

[187] B. R. Preiss. *Data Structures and Algorithms with Object Oriented Design Pattern in C++.* Wiley & Sons, 1998.

[188] R. Preissner, A. Goede, and C. Frömmel. Dictionary of interfaces in proteins (dip). data bank of complementary molecular surface patches. *J Mol Biol*, 280(3):535–550, Jul 1998.

[189] T. Przytycka, R. Srinivasan, and G. D. Rose. Recursive domains in proteins. *Protein Sci*, 11(2):409–417, Feb 2002.

[190] O.B. Ptitsyn and A.V. Finkelstein. Similarities of protein topologies: evolutionary divergence, functional convergence or principles of folding. *Quarterly Reviews of Biophysics*, 13:339–386, 1980.

[191] G. N. Ramachandran and V. Sasisekharan. Conformation of polypeptides and proteins. *Adv Protein Chem*, 23:283–438, 1968.

[192] S. T. Rao and M. G. Rossmann. Comparison of super-secondary structures in proteins. *J Mol Biol*, 76(2):241–256, May 1973.

[193] John W Raymond and Peter Willett. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J Comput Aided Mol Des*, 16(7):521–533, Jul 2002.

[194] F. M. Richards and C. E. Kundrot. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins*, 3(2):71–84, 1988.

[195] J. S. Richardson. Beta-sheet topology and the relatedness of proteins. *Nature*, 268:495–500, 1977.

[196] J. S. Richardson. The anatomy and taxonomy of protein structure. *Adv Protein Chem*, 34:167–339, 1981.

[197] J. Roach, S. Sharma, M. Kapustina, and C. W. Carter. Structure alignment via delaunay tetrahedralization. *Proteins*, 60(1):66–81, Jul 2005.

[198] M. G. Rossmann and P. Argos. Exploring structural homology of proteins. *J Mol Biol*, 105(1):75–95, Jul 1976.

[199] M. G. Rossmann, D. Moras, and K. W. Olsen. Chemical and biological evolution of a nucleotide-binding protein. *Nature*, 250:194–199, 1974.

[200] B. Rost. Twilight zone of protein sequence alignments. *Protein Eng*, 12(2):85–94, Feb 1999.

[201] R. B. Russell. Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J Mol Biol*, 279(5):1211–1227, Jun 1998.

[202] R. B. Russell and G. J. Barton. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 14(2):309–323, Oct 1992.

[203] R. B. Russell and G. J. Barton. Structural features can be unconserved in proteins with similar folds. an analysis of side-chain to side-chain contacts secondary structure and accessibility. *J Mol Biol*, 244(3):332–350, Dec 1994.

[204] A. Sali and T. L. Blundell. Definition of general topological equivalence in protein structures. a procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J Mol Biol*, 212(2):403–428, Mar 1990.

[205] R. Sayle and E.J. Milner-White. Rasmol: Biomolecular graphics for all. *Trends in Biochemical Sciences*, 20:374–376, 1995.

[206] G. E. Schulz and R. H. Schirmer. *Principles of protein structures*. New York/Heidelberg/Berlin: Springer-Verlag, 1979.

[207] Maxim Shatsky, Ruth Nussinov, and Haim J Wolfson. A method for simultaneous alignment of multiple protein structures. *Proteins*, 56(1):143–156, Jul 2004.

[208] K. Shearer, H. Bunke, and S. Venkatesh. Video indexing and similarity retrieval by largest common subgraph detection using decision trees. *Pattern Recognition*, 34:1075–1091, 2001.

[209] Edward S C Shih and Ming-Jing Hwang. Alternative alignments from comparison of protein structures. *Proteins*, 56(3):519–527, Aug 2004.

[210] I. Shindyalov and P. Bourne. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, 11(9):739–747, 1998.

[211] A. P. Singh and D. L. Brutlag. Hierarchical protein structure superposition using both secondary structure and atomic representations. *Proc Int Conf Intell Syst Mol Biol*, 5:284–293, 1997.

[212] M. J. Sippl and S. Weitcus. Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins*, 13:258–271, 1992.

[213] H. Sklenar, C. Etchebest, and R. Lavery. Describing protein structure: a general algorithm yielding complete helicoidal parameters and a unique overall axis. *Proteins*, 6(1):46–60, 1989.

[214] T. S. Smith and M. S. Waterman. Identification of common molecular subsequences. *J.Mol.Biol.*, 147:195–197, 1981.

[215] J. Soding and A. N. Lupas. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays*, 25(9):837–846, Sep 2003.

[216] S. Subbiah, D. V. Laurents, and M. Levitt. Structural similarity of dna-binding domains of bacteriophage repressors and the globin core. *Curr Biol*, 3(3):141–148, Mar 1993.

[217] M. Suyama, Y. Matsuo, and K. Nishikawa. Comparison of protein structures using 3d profile alignment. *J Mol Evol*, 44 Suppl 1:S163–S173, 1997.

[218] J. D. Szustakowski and Z. Weng. Protein structure alignment using a genetic algorithm. *Proteins*, 38:428–440, 2000.

[219] J. D. Szustakowski and Z. Weng. Protein structure alignment using evolutionary computing. In G. Fogel and D. Corne, editors, *Evolutionary Computation in Bioinformatics*, pages 59–86. Morgan Kaufman, 2002.

[220] C. Tanford. The hydrophobic effect and the organization of living matter. *Science*, 200(4345):1012–1018, Jun 1978.

[221] R. E. Tarjan and A. E. Trojanowski. Finding maximum independent set. *SIAM J. Comput.*, 6:537–546, 1977.

[222] T. J. Taylor and I. I. Vaisman. Graph theoretic properties of networks formed by the Delaunay tessellation of protein structures. *Phys Rev E Stat Nonlin Soft Matter Phys*, 73(4 Pt 1):041925, Apr 2006.

[223] W. Taylor. Random structural models for double dynamic programming score evalution. *J. Mol. Evol.*, 44:174–180, 1997.

[224] W. R. Taylor. The classification of amino acid conservation. *J Theor Biol*, 119(2):205–218, Mar 1986.

[225] W. R. Taylor. Defining linear segments in protein structure. *J Mol Biol*, 310:1135–1150, 2001.

[226] W. R. Taylor, T. P. Flores, and C. A. Orengo. Multiple protein structure alignment. *Protein Sci*, 3(10):1858–1870, Oct 1994.

[227] W. R. Taylor and C. A. Orengo. Protein structure alignment. *J Mol Biol*, 208(1):1–22, Jul 1989.

[228] A. E. Todd, C. A. Orengo, and J. M. Thornton. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol*, 307(4):1113–1143, Apr 2001.

[229] M. Turcotte, S. H. Muggleton, and M. J. Sternberg. Automated discovery of structural signatures of protein fold and function. *J Mol Biol*, 306(3):591–605, Feb 2001.

[230] M. Turcotte, S. H. Muggleton, and M. J. Sternberg. Generating protein three-dimensional fold signatures using inductive logic programming. *Comput Chem*, 26(1):57–64, Dec 2001.

[231] M. Vendruscolo, E. Kussell, and E. Domany. Recovery of protein structure from contact maps. *Folding and Design*, 2:295–306, 1997.

[232] J. Vesterstroem and W. R. Taylor. Flexible secondary structure based protein structure comparison applied to the detection of circular permutation. *J Comput Biol*, 13(1):43–63, 2006.

[233] G. Voronoi. Nouvelles applications des parametres continus a la theorie des formes quadratiques. *J Reine Angew. Math.*, 134:198–207, 1908.

[234] G. Vriend and C. Sander. Detection of common three-dimensional substructures in proteins. *Proteins*, 11(1):52–58, 1991.

[235] R. C. Wade, K. J. Clark, and P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem*, 36(1):140–147, Jan 1993.

[236] C. J. Wallace. The curious case of protein splicing: mechanistic insights suggested by protein semisynthesis. *Protein Sci*, 2(5):697–705, May 1993.

[237] January Weiner, Geraint Thomas, and Erich Bornberg-Bauer. Rapid motif-based prediction of circular permutations in multi-domain proteins. *Bioinformatics*, 21(7):932–937, Apr 2005.

[238] D. Weininger. SMILES, a chemical language and information system. 1.introduction to methodologx and encoding rules. *J. Chem. Inf. Sci.*, 28:31–36, 1988.

[239] H. Whitney. Congruent graphs and the connectivity of graphs. *Am. J. Math.*, 54:150–168, 1932.

[240] R. W. Williams, A. Chang, D. Jureti, and S. Loughran. Secondary structure predictions and medium range interactions. *Biochim Biophys Acta*, 916(2):200–204, Nov 1987.

[241] C. M. Wilmot and J. M. Thornton. Turns and their distortions: a proposed new nomenclature. *Protein Eng*, 3:479–494, 1990.

[242] T. D. Wu, S. C. Schmidler, T. Hastie, and D. L. Brutlag. Regression analysis of multiple protein structures. *J Comput Biol*, 5(3):585–595, 1998.

[243] J. Xu, M. Li, D. Kim, and Y. Xu. RAPTOR: optimal protein threading by linear programming. *Journal of Bioinformatics and Computational Biology*, 1:95–117, 2003.

[244] A. S. Yang and B. Honig. An integrated approach to the analysis and modelling of protein sequences and structures. i. protein structure alignment and a quantitative measure for protein structural distance. *J. Mol. Biol.*, 301:665–678, 2000.

[245] X. Yuan and C. Bystroff. Non-sequential structure-based alignments reveal topology-independent core packing arrangements in proteins. *Bioinformatics*, 21(7):1010–1019, 2005.

[246] M. J. Zaki, V. Nadimpally, D. Bardhan, and C. Bystroff. Predicting protein folding pathways. *Bioinformatics*, 20 Suppl 1:I386–I393, Aug 2004.

[247] A. Zemla. LGA: A method for finding 3d similarities in protein structures. *Nucleic Acids Res.*, 31(13):3370–3374, 2003.

[248] S. Zhang, B. Haas, E. Eskin, and V. Bafna. Searching genomes for noncoding RNA using FastR. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 2(4):366–379, 2005.

[249] Y. Zhang and J. Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57(4):702–710, 2004.

[250] R. Zimmer, M. Wöhler, and R. Thiele. New scoring schemes for protein fold recognition based on voronoi contacts. *Bioinformatics*, 14(3):295–308, 1998.