

Bibliographie

- [1] C. André, B. Vincens, J. Boisvieux, and S. Hazout. Mosaic: segmenting multiple aligned dna sequences. *Bioinformatics*, NA:-, 2001.
- [2] R. Aurora and G. Rose. Helix capping. *Protein Science*, 7:21–38, 1998.
- [3] D. Barlow and J. Thornton. Helix geometry in proteins. *J Mol Biol*, 201:601–619, 1988.
- [4] P. Bates and M. Sternberg. Model building by comparison at casp3: using expert knowledge and computer automation. *Proteins*, S3:47–54, 1999.
- [5] I. Berezovsky, A. Grosberg, and E. Trifonov. Closed loops of nearly standard size: common basic element of protein structure. *FEBS Letters*, 466:283–286, 2000.
- [6] F. Bernstein, T. Koetzle, G. Williams, E. Meyer, M. Brice, J. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112:535–540, 1977.
- [7] J. Bienkowska, R. j. Rogers, and T. Smith. Filtered neighbors threading. *Proteins*, 37:346–359, 1999.
- [8] J. Bienkowska, R. j. Rogers, and T. Smith. Protein fold recognition by total alignment probability. *Proteins*, 40:451–462, 2000.
- [9] T. Blundell, D. Carney, S. Gardner, F. Hayes, B. Howlin, and T. Hubbard. Knowledge-based protein modelling and design. *Eur. J. Biochem.*, 172:513–520, 1988.
- [10] T. Blundell, B. Sibanda, M. Sternberg, and J. Thornton. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, 326:347, 1987.
- [11] J. Boberg, T. Salakoski, and M. Vihinen. Selection of a representative set of structures from brookhaven protein databank. *Proteins*, 14:264–276, 1992.
- [12] T. bottom line for prediction of residue solvent accessibility. Richardson, cj and barlow, dj. *Protein Eng*, 12:1051–1054, 1999.
- [13] N. Boutonnet, A. Kajava, and M. Rooman. Structural classification of alphabeta and betabetaalpha supersecondary structure. *Proteins*, 30:193–212, 1998.

- [14] M. Bower, F. Cohen, and R. Dunbrack Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.*, 267:1268–1282, 1993.
- [15] J. Bowie, R. Lüthy, and D. Eisenberg. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253:164–169, 1991.
- [16] R. Britten. Precise sequence complementarity between yeast chromosome ends and two classes of just-subtelomeric sequences. *Proc Natl Acad Sci USA*, 26:5906–59120, 1998.
- [17] S. Bryant and C. Lawrence. An empirical energy function for threading protein sequence through the folding motif. *Proteins*, 16:92–112, 1993.
- [18] C. Bystroff and D. Baker. Blind predictions of local protein structure in casp2 targets using the i-sites library. *Proteins*, suppl.1:167–171, 1997.
- [19] C. Bystroff and D. Baker. Prediction of local structure in proteins using a library of sequence-structure motif. *J Mol Biol*, 281:565–577, 1998.
- [20] C. Bystroff, V. Thorsson, and D. Baker. Hmmstr: a hidden markov model for local sequence-structure correlations in proteins. *J Mol Biol*, 301:173–190, 2000.
- [21] A. Camproux, A. de Brevern, P. Tuffery, and S. Hazout. Exploring the use of a structural alphabet for a structural prediction of protein loops. *Theoretical Chemistry Accounts*, page en revision, 2001.
- [22] A. Camproux, F. Saunier, G. Chouvet, J. Thalabard, and G. Thomas. A hidden markov model approach to neuron firing patterns. *Biophysical Journal*, 71:2404–2412, 1996.
- [23] A. Camproux, P. Tuffery, L. Buffat, C. Andre, J. Boisvieux, and S. Hazout. Analyzing patterns between regular secondary structures using short structural blocks defined by a hidden markov model. *Theoretical Chemistry Accounts*, 101:33–40, 1999.
- [24] A. Camproux, P. Tuffery, J. Chevrolat, J. Boisvieux, and S. Hazout. Hidden markov model approach for identifying the modular framework of the protein backbone. *Protein Eng*, 12(12):1063–1073, 1999.
- [25] O. Carugo. Predicting residue solvent accessibility from protein sequence by considering the sequence environment. *Protein Eng*, 13:607–609, 2000.
- [26] H. Chan. Folding alphabet. *Nature Structural Biology*, 6:994–996, 1999.
- [27] J. Chandonia and M. Karplus. The importance of larger data sets for protein secondary structure prediction with neural networks. *Protein Sci.*, 5:768–774, 1996.
- [28] J. Chandonia and M. Karplus. New methods for accurate prediction of protein secondary structure. *Proteins*, 35:293–306, 1999.

- [29] P. Chou and G. Fasman. Prediction of protein conformation. *Biochemistry*, 13:222–245, 1974.
- [30] M. Claessens, E. Cutsen, I. Lasters, and S. Wodak. Modelling the polypeptide backbone with ‘spare parts’ from known protein structures. *Prot. Eng.*, 4:335, 1989.
- [31] N. Colloc’h, C. Etchebest, E. Thoreau, B. Henrissat, and J. Mornon. Comparaison of three algorithms for the assignement of secondary structure in proteins: the advantages of a concensus assignement. *Protein Eng.*, 6:377–382, 1993.
- [32] L. Conte and J. Chothia, C and Janin. The atomic structure of protein-protein recognition sites. *J Mol Biol*, 5:2177–2199, 1999.
- [33] J. Cuff and G. Barton. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*, 34:508–519, 1999.
- [34] J. Cuff, M. Clamp, A. Siddiqui, M. Finlay, and G. Barton. Jpred: a consensus secondary structure prediction server. *Proteins*, 14:892–893, 1998.
- [35] A. de Brevern, C. Etchebest, and S. Hazout. Bayesian probabilistic approach for prediction backbone structures in terms of protein blocks. *Proteins*, 41(3):271–287, 2000.
- [36] A. de Brevern and S. Hazout. Hybrid protein model (hpm): a method to compact protein 3d-structures information and physicochemical properties. *IEEE - Computer Society: Proceedings of the 7th Symposium on String Processing and Information Retrieval*, 1(1):49–57, 2000.
- [37] A. de Brevern and S. Hazout. Compacting local protein folds with a hybrid protein. *Theoretical Chemistry Accounts*, NA:–, 2001.
- [38] C. Deane and T. Blundell. A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins. *Proteins*, 40:135–144, 2000.
- [39] T. Defay and F. Cohen. Evaluation of current techniques for ab initio protein structure prediction. *Proteins*, 23:431–445, 1995.
- [40] G. Deléage, C. Blanchet, and C. Geourjon. Protein structure prediction. implications for the biologist. *Biochimie*, 79:681–686, 1997.
- [41] P. Derreumaux. Folding a 20 amino acid $\alpha\beta$ peptide with the diffusion process-controlled monte-carlo method. *J. Chem. Phys.*, 107:1941–1947, 1997.
- [42] V. Di Francesco, J. Garnier, and P. Munson. Improving protein secondary structure prediction with aligned homologous sequences. *Protein Science*, 5:106–113, 1996.
- [43] C. Dodge, R. Schneider, and C. Sander. The hssp database of protein structure-sequence alignments and family profiles. *Nucleic Acid Res*, 26:313–315, 1998.

- [44] L. Donate, S. Rufino, L. Canard, and T. Blundell. Conformational analysis and clustering of short and medium size loops connecting secondary structures: a database for modeling and prediction. *Protein Science*, 5:2600–2616, 1996.
- [45] R. Dunbrack Jr. and M. Karplus. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.*, 230:543–574, 1993.
- [46] A. Efimov. Super-secondary structures involving triple-strand beta-sheets. *FEBS Lett*, 334:253–256, 1993.
- [47] N. Eswar and C. Ramakrishnan. Secondary structures without backbone: an analysis of backbone mimicry by polar side chains in protein structures. *Protein Eng*, 12:447–455, 1999.
- [48] P. Fariselli and R. Casadio. A neural network based predictor of residue contacts in proteins. *Protein Eng*, 12:15–21, 1999.
- [49] P. Fariselli and R. Casadio. Prediction of the number of residue contacts in proteins. *Ismb*, 2000, 8:146–151, 2000.
- [50] P. Fariselli, P. Riccobelli, and R. Casadio. Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. *Proteins*, 36:340–346, 1999.
- [51] J. Fetrow. Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J*, 9:708–717, 1995.
- [52] J. Fetrow and G. Berg. Using information theory to discover side chain rotamer classes: analysis of the effects of local backbone structure. *Pacific Symposium*, 1:–, 1998.
- [53] J. Fetrow, M. Palumbo, and G. Berg. Patterns, structures, and amino acid frequencies in structural building blocks, a protein secondary structure classification scheme. *Proteins*, 27:249–271, 1997.
- [54] A. Finkelstein and B. Reva. A search for the most stable folds of protein chains. *Nature*, 351:497–499, 1991.
- [55] A. Finkelstein and B. Reva. Search for the most stable folds of protein chains: I. application of a self-consistent molecular field to a problem of protein three-dimensional structure prediction. *Protein Eng*, 9:387–397, 1996.
- [56] A. Fiser, N. Dosztányi, and S. I. The role of long-range interactions in defining the secondary structure of proteins is overestimated. *Comput. Appl. Biosci*, 13:297–301, 1997.
- [57] A. Fiser and I. Simon. Predicting the oxidation state of cysteines by multiple sequence alignment. *Bioinformatics*, 16:251–256, 2000.
- [58] D. Frishman and P. Argos. Knowledge-based protein secondary structure. *Proteins*, 23:566–579, 1995.

- [59] D. Frishman and P. Argos. The future of protein secondary structure accuracy. *Fold. Des.*, 2:159–162, 1997.
- [60] D. Frishman and P. Argos. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins*, 27:97–120, 1997.
- [61] J. Garnier, J. Gibrat, and B. Robson. Gor method for predicting protein secondary structure from amino acid sequence. *Methods. Enzymol*, 266:540–553, 1996.
- [62] J. Garnier, D. Osguthorpe, and B. Robson. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol*, 120:97–120, 1978.
- [63] V. Geetha and P. Munson. Linkers of secondary structures in proteins. *Protein Science*, 6:2538–2547, 1997.
- [64] C. Geourjon and G. Deléage. Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignements. *CABIOS*, 6:681–684, 1995.
- [65] J.-F. Gibrat, J. Garnier, and B. Robson. Further developments of protein secondary structure prediction using information theory. *J Mol Biol*, 198:425–443, 1987.
- [66] J.-F. Gibrat, B. Robson, and J. Garnier. Influence of the local amino acid sequence upon the zones of the torsional angles ϕ and ψ adopted by residues in proteins. *Biochemistry*, 30:1578–1586, 1991.
- [67] J. Gorodkin, O. Lund, C. Andersen, and S. Brunak. Using sequence motifs for enhanced neural network prediction of protein distance constraints. *Proceedings of the seventh international conference for molecular biology (ISMB'99)*, 1:95–105, 1999.
- [68] C. Gourceon and G. Deleage. Sopm: a self-optimized method for protein secondary structure prediction. *Prot. Eng.*, 7:157–164, 1994.
- [69] S. Govindarajan and R. Goldstein. Why are some proteins structures so common? *Proc. Natl. Acad. Sci. USA*, 93:3341–3345, 1996.
- [70] S. Govindarajan, R. Recabarren, and R. Goldstein. Estimating the total number of protein folds. *Proteins*, 35:408–414, 1999.
- [71] M. Gribskov, A. McLachlan, and D. Eisenberg. Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci U S A*, 84:4355–4358, 1987.
- [72] Y. Guermeur, C. Geourjon, P. Gallinari, and G. Deléage. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, 15:413–421, 1999.

- [73] C. Hadley and D. Jones. A systematic comparison of protein structure classifications scop, cath and fssp. *Structure*, 7:1099–1112, 1999.
- [74] K. Han and D. Baker. Recurring local sequence motifs in proteins. *J Mol Biol*, 241:176–187, 1995.
- [75] K. Han and D. Baker. Global properties of the mapping between local amino acid sequence and local structure in proteins. *Proc. Natl. Acad. Sci. USA*, 93:5814–5818, 1996.
- [76] K. Han, C. Bystroff, and D. Baker. Three dimensional structures and contexts associated with recurrent amino acid sequence patterns. *Protein Science*, 6:1587–1590, 1997.
- [77] J. Hanke and J. Reich. Kohonen map as a visualization tool for the analysis of protein sequences: multiple alignments, domains and segments of secondary structures. *Comput. Appl. Biosci.*, 12:447–454, 1996.
- [78] J. Hanke and J. Reich. Self-organizing hierarchic networks for pattern recognition in protein sequence. *Protein Science*, 5:72–82, 1996.
- [79] J. Hartigan and M. Wong. k-means. *Applied Statistics*, 28(1):100–115, 1979.
- [80] C. Haseman, R. Kurumbail, N. Boddupalli, J. Peterson, and N. Deisenhofer. Structure and function of cytochromes p450: a comparative analysis of three crystal structures. *Structure*, 2:41–62, 1995.
- [81] C. Hasemann, K. Ravichandran, J. Peterson, and J. Deisenhofer. Crystal structure and refinement of cytochrome p450terp at 2.3 a resolution. *J Mol Biol*, 4:1169–1185, 1994.
- [82] S. Hayward and J. Collins. Limits on α -helix prediction with neural networks models. *Proteins*, 14:372–381, 1992.
- [83] U. Hobohm and C. Sander. Enlarged representative set of protein structures. *Prot Sci*, 3:522–524, 1994.
- [84] U. Hobohm, F. Scharf, R. Schneider, and C. Sander. Selection of a representative set of structures from the brookhaven protein databank. *Prot Sci*, 1:409–417, 1992.
- [85] L. Holley and M. Karplus. Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci.*, 86:152–156, 1989.
- [86] L. Holm and C. Sander. The fssp database of structurally aligned protein fold families. *Nucl. Acid. Res.*, 22:3600–3609, 1994.
- [87] L. Holm and C. Sander. Dali/fssp classification of three-dimensional protein folds. *Nucl Acids Res*, 25:231–234, 1997.
- [88] E. Huang, R. Samudrala, and J. Ponder. Ab initio fold prediction of small helical proteins using distance geometry and knowledge-based scoring functions. *J Mol Biol*, 290:267–281, 1999.

- [89] T. Hubbard, A. Murzin, S. Brenner, and C. Chotia. Scop : a structural classification of proteins database. *Nucleic Acids Research*, 25:236–239, 1997.
- [90] E. Hutchinson and J. Thornton. A revised set of potentials for β -turn formation in protein. *Prot Sci*, 3:2207–2216, 1994.
- [91] J. Janin. Surface area of globular proteins. *J. Mol. Biol.*, 105:13, 1976.
- [92] J. Janin. Wet and dry interfaces: the role of solvent in protein-protein and protein-dna recognition. *Structure Fold Des*, 7:R277–279, 1999.
- [93] L. Jaroszewski, L. Rychlewski, B. Zhang, and A. Godzik. Fold prediction by a hierarchy of sequence, threading, and modeling methods. *Prot Sci*, 7:1431–40, 1998.
- [94] P. Jean, J. Pothier, P. Dansette, D. Mansuy, and A. Viari. Automated multiple analysis of protein structures: application to homology modeling of cytochromes p450. *Proteins*, 28:388–404, 1997.
- [95] D. Jones, W. Taylor, and J. Thornton. A new approach to protein fold recognition. *Nature*, 358:86–89, 1992.
- [96] D. Jones and J. Thornton. Protein fold recognition. *Journal of Computer-Aided Molecular Design*, 7:439–456, 1993.
- [97] D. Jones, M. Tress, K. Bryson, and C. Hadley. Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins*, S3:104–111, 1999.
- [98] S. Jones, M. Stewart, A. Michie, M. Swindells, C. Orengo, and J. Thornton. Domain assignment for protein structures using a consensus approach ; characterization and analysis. *Prot Sci*, 7:233–242, 1998.
- [99] T. Jones and S. Thirup. Using know substructures in protein model building and crystallography. *EMBO J*, 5:819–822, 1986.
- [100] W. Kabsh and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, 22:2577, 1983.
- [101] N. Kannan and S. Vishveshwara. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol*, 292:441–464, 1999.
- [102] M. Karpen, P. de Haseth, and K. Neet. Comparing short protein substructures by a method based on backbone torsion angles. *Proteins*, 6:155–167, 1989.
- [103] M. Karpen, P. de Haseth, and K. Neet. Differences in the amino acid distributions of 3_10 -helices and α -helices. *Protein Science*, 1:1333–1342, 1992.

- [104] T. Kawabata and J. Doi. Improvement of protein secondary structure prediction using binary word encoding. *Proteins*, 27:36–46, 1997.
- [105] R. King and M. Sternberg. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science*, 5:2298–2310, 1996.
- [106] I. Koch, T. Lengauer, and E. Wanke. An algorithm for finding maximal common subtopologies in a set of protein structures. *Journal of Computational Biology*, 3:289–306, 1996.
- [107] T. Kohonen. Learning vector quantization. *Neural Networks*, 1(suppl. 1):303, 1989.
- [108] T. Kohonen. Statistical pattern recognition revisited. *Advanced Neural Computers*, **R. Eckmiller (editor)**, Elsevier Science publisher (Holland), pages –, 1990.
- [109] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, Germany, 1997.
- [110] A. Kolinski, P. Rotkiewicz, B. Ilkowski, and J. Skolnick. A method for the improvement of threading-based protein models. *Proteins*, 37:592–610, 1999.
- [111] J. Kraulis. Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst*, 24:946–950, 1991.
- [112] S. Kullback and R. Leibler. On information and sufficiency. *Ann Math Stat*, 22:79–86, 1951.
- [113] S. Kumar and M. Bansal. Geometrical and sequence characteristics of α -helices in globular proteins. *Biophysical Journal*, 78:1935–1944, 1998.
- [114] J.-M. Kwasigroch, J. Chomilier, and J.-P. Mornon. A global taxonomy of loops in globular proteins. *J Mol Biol*, 259:855–872, 1996.
- [115] J. Kyte and R. Doolittle. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157(1):105–132, 1982.
- [116] G. Labesse, N. Colloc'h, J. Pothier, and J.-P. Mornon. P-sea: a new efficient assignment of secondary structure from α . *Comput. Appl. Biosci.*, 13:291–295, 1997.
- [117] R. Lathrop, R. Rogers Jr., T. Smith, and J. White. A bayes-optimal sequence-structure theory that unifies protein sequence-structure recognition and alignment,. *Bulletin of Mathematical Biology*, 60:1039–1071, 1998.
- [118] B. Lee and F. Richards. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, 55:379–392, 1971.
- [119] C. Levinthal. Molecular model-building by computer. *Sci Am*, 214:42–52, 1966.
- [120] M. Levitt. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.*, 226:507–533, 1992.
- [121] R. Lüthy, D. McLachlan, and D. Eisenberg. Secondary structure-based profiles: use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. *Proteins*, 10:229–239, 1991.

- [122] M. Liebman, C. Venanzi, and H. Weinstein. Structural analysis of carboxypeptidase a and its complexes with inhibitors as a basis for modeling enzyme recognition and specificity. *Biopolymers*, 24:1721–1758, 1985.
- [123] L. Lo Conte, B. Ailey, T. Hubbard, S. Brenner, A. Murzin, and C. Chotia. Scop : a structural classification of proteins database. *Nucleic Acid Research*, 28:257–259, 2000.
- [124] O. Lund, K. Frimand, J. Gorodkin, H. Bohr, J. Hansen, and S. Brunak. Protein distance constraints predicted by neural networks and probability density functions. *Protein Engineering*, 10:1241–1248, 1997.
- [125] M. MacArthur and J. Thornton. Deviations from planarity of peptide bond in peptides and proteins. *J Mol Biol*, 264:1180–1195, 1996.
- [126] T. Madej, J.-F. Gibrat, and S. Bryant. Threading a database of protein cores. *Proteins*, 23:356–369, 1995.
- [127] A. Michie, C. Orengo, and J. Thornton. Analysis of domain structural class using an automated class assignement protocol. *J Mol Biol*, 262:168–85, 1996.
- [128] Mucchielli-Giorgi. Thèse. *Paris 7*, 5:70–122, 1999.
- [129] M. Mucchielli-Giorgi, S. Hazout, and T. P. Predacc: prediction of solvent accessibility. *Bioinformatics*, 15:176–177, 1999.
- [130] M. Mucchielli-Giorgi, T. P, and S. Hazout. Prediction of solvent accessibility of amino acid residues: critical aspects. *Theoretical chemistry accounts*, 101:186–193, 1999.
- [131] A. Murzin. Structural classification of proteins: new superfamilies. *Current opinion in Structural Biology*, 6:386–394, 1996.
- [132] A. Murzin, S. Brenner, T. Hubbard, and C. Chotia. Scop : a structural classification of proteins database for the investigation of sequences and structures. *JMB*, 247:526–540, 1995.
- [133] S. Muskall and S. Kim. Predicting protein secondary structure content. a tandem neural network approach. *J Mol Biol*, 225:713–727, 1992.
- [134] K. Nadassy, S. Wodak, and J. Janin. Structural features of protein-nucleic acid recognition sites. *Biochemistry*, 38:1999–2017, 1999.
- [135] O. Olmea and A. Valencia. Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Fold Des*, 2:S25–32, 1997.
- [136] C. Orengo. Classification of protein folds. *Current opinion in Structural Biology*, 4:429–44, 1994.

- [137] C. Orengo, J. Bray, T. Hubbard, L. LoConte, and I. Sillitoe. Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. *Proteins*, suppl.3:149–170, 1999.
- [138] C. Orengo, S. Jones, and J. Thornton. Protein superfamilies and domain superfolds. *Nature*, 372:631–634, 1994.
- [139] C. Orengo, A. Michie, J. Jones, M. Swinells, and J. Thornton. Cath- a hierarchic classification of protein domain structures. *Structure*, 5:1093–1098, 1997.
- [140] C. Orengo, F. Peral, J. Bray, A. Todd, A. Martin, L. Lo Conte, and J. Thornton. The cath database provides insights into protein structure/function relationships. *Nucleic Acids Research*, 27:275–279, 1999.
- [141] D. Osguthorpe. ab initio protein folding. *Current Opinion in Structural Biology*, 10:146–152, 2000.
- [142] M. Ouali and K. RD. Cascaded multiple classifiers for secondary structure prediction. *Protein Science*, 9:1136–1176, 2000.
- [143] A. Panchenko, A. Marchler-Bauer, and S. Bryant. Threading with explicit models for evolutionary conservation of structure and sequence. *Proteins*, Suppl 3:133–140, 1999.
- [144] T. Petersen, C. Lundegaard, M. Nielsen, H. Bohr, J. Bohr, S. Brunak, G. Gippert, and O. Lund. Prediction of protein secondary structure at 80% accuracy. *Proteins*, 41:17–20, 2000.
- [145] L. Presta and G. Rose. Helix signals in proteins. *Science*, 240:1632–1641, 1988.
- [146] S. Prestelski, A. Williams Jr., and M. Liebman. Generation of a substructure library for the description and classification of protein secondary structure. i. overview of the methods and results. *Proteins*, 14:430–439, 1992.
- [147] S. Prestelski, A. Williams Jr., and M. Liebman. Generation of a substructure library for the description and classification of protein secondary structure. ii. application to spectra-structure correlations in fourier transform infrared spectroscopy. *Proteins*, 14:440–450, 1992.
- [148] N. Qian and T. Sejnowski. Prediction of secondary structure of globular proteins using a neural network. *J. Mol. Biol.*, 202:865–884, 1988.
- [149] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77:257–285, 1989.
- [150] K. Rajashankar and S. Ramakumar. Pi-turns in proteins and peptides: classification, conformation, occurrence, hydration and sequence. *Protein Sci*, 5:932–946, 1996.

- [151] K. Ravichandran, S. Boddupalli, C. Hasemann, J. Peterson, and J. Deisenhofer. Crystal structure of hemoprotein domain of p450bm-3, a prototype for microsomal p450's. *Science*, 6:731–736, 1993.
- [152] B. Reva, A. Finkelstein, and J. Skolnick. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Fold Des*, 3:141–147, 1998.
- [153] B. Reva, J. Skolnick, and A. Finkelstein. Averaging interaction energies over homologs improves protein recognition in gapless threading. *Proteins*, 35:353–359, 1999.
- [154] J. Richardson and D. Richardson. Amino acid preferences for specific locations at the end of α helices. *Science*, 240:1648–1652, 1988.
- [155] T. Richmond. Solvent accessible surface area and excluded volume in proteins. *J. Mol. Biol.*, 178:63–89, 1984.
- [156] F. Richards and C. Kundrot. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure,. *Proteins*, 3:71–84, 1988.
- [157] C. Rohl and A. Doig. Models for the 3.10-helix/coil, π -helix-coil and α -helix/3.10-helix/coil transitions in isolated peptides. *Protein Science*, 5:1689–1696, 1996.
- [158] J. Rooman, MJ, S. Wodak, and J. Thornton. Amino acid sequence templates derived from recurrent turn motifs in proteins: critical evaluation of their predictive power. *Protein Eng*, 3:23–27, 1989.
- [159] M. Rooman, J.-P. Kocher, and S. Wodak. prediction of protein backbone conformation based on seven structure assignments. influence of local interactions. *J Mol Biol*, 221:961–979, 1991.
- [160] M. Rooman, J.-P. Kocher, and S. Wodak. Extracting information on folding from amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry*, 27:10226–10238, 1992.
- [161] M. Rooman, J.-P. Kocher, and S. Wodak. Extracting information on folding from amino acid sequence: consensus regions with preferred conformation in homologous proteins. *Biochemistry*, 27:10239–10249, 1992.
- [162] M. Rooman, J. Rodriguez, and S. Wodak. Automatic definition of recurrent local structure motifs in proteins. *J Mol Biol*, 213:327–336, 1990.
- [163] M. Rooman, J. Rodriguez, and S. Wodak. Relations between protein sequence and structure and their significance. *J Mol Biol*, 213:337–350, 1990.
- [164] M. Rooman and S. Wodak. Identification of predictive sequence motifs limited by protein structure data base size. *Nature*, 335:45–49, 1988.

- [165] B. Rost. Phd : predicting one-dimensional protein structure by profile-based neural networks. *Methods. Enzymol*, 232:525–539, 1996.
- [166] B. Rost and C. Sander. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol*, 232:584–599, 1993.
- [167] B. Rost, C. Sander, and R. Schneider. Phd - an automatic mail server for protein secondary structure prediction. *CABIOS*, 10:53–60, 1994.
- [168] A. Salamov and V. Solovyev. Protein secondary structure prediction using local alignments. *J Mol Biol*, 268:31–36, 1997.
- [169] A. Sali and T. Blundell. Compartative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234:779–815, 1993.
- [170] R. Samudrala and J. Moult. A graph-theoric algorithm for comparative modeling of protein structure. *J Mol Biol*, 279:287–302, 1998.
- [171] R. Samudrala, Y. Xia, E. Huanh, and L. M. Ab initio protein structure prediction using a combined hierarchical approach. *Proteins*, 3 suppl.:194–198, 1999.
- [172] C. Sander and R. Schneider. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, 9:56–68, 1991.
- [173] F. Sasagawa and K. Tajima. Prediction of protein secondary structures by a neural network. *Comput. Appl. Biosci*, 9:147–152, 1993.
- [174] R. Schneider, A. de Daruvar, and C. Sander. The hssp database of protein structure-sequence alignments. *Nucl Acids Res*, 25:226–230, 1997.
- [175] J. Schuchhardt, G. Schneider, J. Reichelt, D. Schomburg, and P. Wrede. Local structural motifs of protein backbones are classified by self-organizing neural networks. *Protein Eng*, 9(10):833–842, 1996.
- [176] J. Segrest, H. De Loof, J. Dohlman, C. Brouillette, and G. Anantharamaiah. Amphipathic helix motif: classes and properties. *Proteins*, 8:103–117, 1990.
- [177] E. Shakhnovich. Modeling protein folding: the beauty of power and simplicity. *Fold. Des.*, 1:50–54, 1996.
- [178] K. Simons, R. Bonneau, I. Ruczinski, and D. Baker. Ab initio protein structure prediction of casp 3 targets using rosetta. *Proteins*, 34(suppl. 3):171–176, 1999.
- [179] K. Simons, C. Kooperberg, E. Huang, and B. D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J Mol Biol*, 268:209–225, 1997.

- [180] K. 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.
- [181] H. Sklenar, C. Etchebest, and R. Lavery. Describing protein structure: a general algorithm yielding complete helicoidal parameters and a unique overall axis. *Proteins*, 6:46–60, 1989.
- [182] V. Solovyev and A. Salamov. Predictiong α -helix and β -strand segments of globular proteins. *CABIOS*, 10:661–669, 1994.
- [183] R. Srinivasan and G. Rose. Linus: a hierarchic procedure to predict the fold of a protein. *Proteins*, 22:81–99, 1997.
- [184] R. Srinivasan and G. Rose. A physical basis for protein secondary structure. *PNAS*, 96:14258–14263, 1999.
- [185] M. Sternberg, P. Bates, L. Kelley, and M. MacCallum. Progress in protein structure prediction : assessment of casp 3. *Current opinion in Structural Biology*, 9:368–373, 1999.
- [186] M. Sternberg and S. Islam. Local protein sequence similarity does not imply a structural relationship. *Protein Eng*, 4:125–131, 1990.
- [187] P. Storloz, A. Lapedes, and Y. Xia. Predicting protein secondary structure using neural net and statistical methods. *J. Mol. Biol.*, 225:363–377, 1992.
- [188] J. Sun and A. Doig. Addition of side-chain interactions to 3_{10} -helix/coil and α -helix/ 3_{10} -helix/coil theory. *Protein Science*, 7:2374–2383, 1998.
- [189] X. Sun, X. Rao, L. Peng, and D. Xu. Prediction of protein supersecondary structures based on the artificial neural network method. *Protein Eng*, 10:763–769, 1997.
- [190] W. Taylor. The classification of amino acid conservation. *J Theor Biol*, 119:205–218, 1986.
- [191] W. Taylor. Protein structural domain identification. *Protein Eng*, 12:203–216, 1999.
- [192] M. Thompson and R. Goldstein. Predicting solvent accessibility: higher accuracy using bayesian statistics and optimized residue substitution classes. *Proteins*, 25:38–47, 1996.
- [193] M. Thompson and R. Goldstein. Predicting protein secondary structure with probabilistic schemata of evolutionarily derived information. *Protein Sci*, 6:1963–1975, 1997.
- [194] P. Tuffery. Xmmol: an x11 and motif program for macromolecular visualization and modeling. *J Mol Graphics*, 72:67–72, 1995.
- [195] P. Tufféry, C. Etchebest, and S. Hazout. Prediction of protein side chains conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. *Prot. Eng.*, 10:361–372, 1997.

- [196] P. Tufféry, C. Etchebest, S. Hazout, and R. Lavery. A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Struct. Dynam.*, 6:1267–1289, 1991.
- [197] P. Tufféry, C. Etchebest, S. Hazout, and R. Lavery. A critical comparison of search algorithms applied to the optimisation of protein side-chain conformations. *J. Comp. Chem.*, 14:790–798, 1993.
- [198] R. Unger, D. Harel, W. S, and S. JL. A 3d building blocks approach to analyzing and predicting structure of proteins. *Proteins*, 5:355–373, 1989.
- [199] R. Unger, D. Harel, W. S, and S. JL. Analysis of dihedral angles distribution: the doublets distribution determines polypeptides conformation. *Biopolymers*, 30:499–508, 1990.
- [200] R. Unger and S. JL. The importance of short structural motifs in protein structure analysis. *J Comput Aid Mol Des*, 7:457–472, 1993.
- [201] L. Wernish, M. Hunting, and S. Wodak. Identification of structural domains in proteins by a graph heuristic. *Proteins*, 36:338–352, 1999.
- [202] C. Wilmot and J. Thornton. Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol*, 203:221–232, 1988.
- [203] R. Wintjens, M. Rooman, and S. Wodak. Automatic classification and analysis of alpha alpha-turn motifs in proteins. *J. Mol. Biol*, 255:235–253, 1996.
- [204] J. Wodjick, J.-P. Mornon, and J. Chomilier. New efficient statistical sequence-dependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification. *J Mol Biol*, 289:1469–90, 1999.
- [205] Y. Xu, D. Xu, and E. Uberbacher. An efficient computational method for globally optimal threading. *J Comput Biol*, 5:597–614, 1998.
- [206] Q. Yi, C. Bystroff, P. Rajagopal, R. Klevit, and D. Baker. Prediction and structural characterization of an independant folding substructure in the src sh3 domain. *J Mol Biol*, 283:293–300, 1998.
- [207] A. Zamyatin. Protein volume in solution. *Prog Biophys Mol Biol*, 24:107–123, 1972.
- [208] X. Zhang, J. Fetrow, W. Rennie, D. Waltz, and G. Berg. Automatic derivation of substructures yoelds novel structure building blocks in globular proteins. *ISMB 93*, 1:438–446, 1993.
- [209] X. Zhang, J. Fetrow, W. Rennie, D. Waltz, and G. Berg. Design of an auto-associative neural network with hidden layer activation that were used to reclassify local protein structures. *Techniques in protein chemistery V*, 1:397–404, 1994.
- [210] K. Zimmermann. When awaiting 'bio' champollion; dynamic programming regularization of the protein secondary structure predictions. *Protein Eng*, 7:1197–1202, 1994.

- [211] K. Zimmermann and J. Gibrat. In unison : regularization of protein secondary structure predictions that makes use of multiple alignements. *Prot. Eng.*, 11:865–865, 1998.