Bibliographie

[Aho *et al.*, 1974] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

[Aho *et al.*, 1983] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *Data Structures and Algorithms*. Addison-Wesley, 1983.

[Altschul and Erickson, 1986] S. F. Altschul and B. W. Erickson. Optimal sequence alignment using affine gap costs. *Bull. Math. Biol.*, 48:603–616, 1986.

[Altschul and Lipman, 1989] S. F. Altschul and D. J. Lipman. Tree, stars, and multiple biological sequence alignment. *SIAM J. Appl. Math.*, 49:197–209, 1989.

[Altschul and Lipman, 1990] S. F. Altschul and D. J. Lipman. Protein database searches for multiple alignments. *Proc. Natl. Acad. Sci. USA*, 87:5509–5513, 1990.

[Altschul *et al.*, 1990] 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.

[Altschul, 1989] S. F. Altschul. Gap costs for multiple sequence alignment. *J. Theor. Biol.*, 138:297–309, 1989.

[Altschul, 1991] S. F. Altschul. Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.*, 219:555–565, 1991.

[Altschul, 1993] S. F. Altschul. A protein alignment scoring system sensitive at all evolutionary distances. *J. Mol. Evol.*, 36:290–300, 1993.

[Apostolico and Preparata, 1983] A. Apostolico and F. P. Preparata. Optimal off-line detection of repetitions in a string. *Theoret. Comput. Sci.*, 22:297–315, 1983.

[Argos, 1987] P. Argos. Analysis of sequence-similar pentapeptides in unrelated protein tertiary structures: strategies for protein folding and a guide for site-directed mutagenesis. *J. Mol. Biol.*, 197:331–348, 1987.

[Bacon and Anderson, 1986] D. J. Bacon and W. F. Anderson. Multiple sequence alignment. *J. Mol. Biol.*, 191:153–161, 1986.

[Bacon and Anderson, 1990] D. J. Bacon and W. F. Anderson. Multiple sequence comparison. *Meth. Enzymol.*, 183:438–447, 1990.

[Baeza-Yates and Gonnet, 1992] R. Baeza-Yates and G. H. Gonnet. A new approach to text searching. *Commun. ACM*, 35:74–82, 1992.

[Bafna *et al.*, 1994] V. Bafna, E. L. Lawler, and P. A. Pevzner. Approximation algorithms for multiple sequence alignement. In M. Crochemore and D. Gusfield, editors, *Combinatorial Pattern Matching*, volume 807 of *Lecture Notes in Computer Science*, pages 43–53. Springer-Verlag, 1994.

[Bains, 1986] W. Bains. MULTAN: a program to align multiple DNA sequences. *Nucleic Acids Res.*, 14:159–177, 1986.

[Bains, 1989] W. Bains. MULTAN (2), a multiple string alignment program for nucleic acids and proteins. *Comput. Applic. Biosc.*, 5:51–52, 1989.

[Bairoch and Boeckmann, 1991] A. Bairoch and C. Boeckmann. The SWISS-PROT protein sequence data bank. *Nucleic Acids. Res.*, 19:2247–2249, 1991.

[Bairoch, 1992] A. Bairoch. PROSITE: A dictionary of protein sites and patterns. *Nucleic Acids Res.*, 20:2013–2018, 1992.

[Bajaj and Blundell, 1984] M. Bajaj and T. Blundell. Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng.*, 13:453–492, 1984.

[Barriel, 1994] V. Barriel. Phylogénies moléculaires et insertions-délétions de nucléotides. *C. R. Acad. Sci. Paris, Sciences de la vie*, 317:693–701, 1994.

[Barton and Sternberg, 1987a] G. J. Barton and M. J. E. Sternberg. Evaluation and improvements in the automatic alignment of protein sequences. *Protein Eng.*, 1:89–94, 1987.

[Barton and Sternberg, 1987b] G. J. Barton and M. J. E. Sternberg. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J. Mol. Biol.*, 198:327–337, 1987.

[Barton and Sternberg, 1990] G. J. Barton and M. J. E. Sternberg. Flexible protein sequence patterns: a sensitive method to detect weak structural similarities. *J. Mol. Biol.*, 212:389–402, 1990.

[Barton, 1990] G. J. Barton. Protein multiple sequence alignment and flexible pattern matching. In *Meth. Enzymol.*, volume 183, pages 403–428. Academic Press, 1990.

[Barton, 1993] G. J. Barton. An efficient algorithm to locate all locally optimal alignments between two sequences allowing for gaps. *Comput. Appl. Biosci.*, 9:729–734, 1993.

[Bashford *et al.*, 1987] D. Bashford, C. Chothia, and A. M. Lesk. Determinants of a protein fold: unique features of the globin amino acid sequence. *J. Mol. Biol.*, 212:389–402, 1987.

[Benner *et al.*, 1993] S. A. Benner, M. A. Cohen, and G. H. Gonnet. Empirical and structural models for insertions and deletions in the divergent evolution of proteins. *J. Mol. Biol.*, 229:1065–1082, 1993.

[Benner *et al.*, 1994] S. A. Benner, M. A. Cohen, and G. H. Gonnet. Amino acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Eng.*, 7:1323–1332, 1994.

[Berg and von Hippel, 1988] O. G. Berg and P. H. von Hippel. Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.*, 200:709–723, 1988.

[Bernstein *et al.*, 1977] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[Billoud *et al.*, 1996] B. Billoud, M. Kontic, and A. Viari. Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence databases. *Nucleic Acids Res.*, 24:1395–1403, 1996.

[Blaisdell *et al.*, 1993] B.E. Blaisdell, K.E. Rudd, A. Matin, and S. Karlin. Significant dispersed recurrent DNA sequences in the E*scherichia coli* genome. Several new groups. *J. Mol. Biol.*, 229:833–848, 1993.

[Blanken *et al.*, 1982] R. L. Blanken, L. C. Klotz, and A. G. Hinnebusch. Computer comparison of new and existing criteria for constructing evolutionary trees from sequence data. *J. Mol. Evol.*, 19:9–19, 1982.

[Blum, 1990] N. Blum. On locally optimal alignment in genetic sequences. manuscript, 1990.

[Blumer *et al.*, 1983] A. Blumer, J. Blumer, A. Ehrenfeucht, D. Haussler, and R. McConnell. Linear size finite automata for the set of all subwords of a text. *Bull. Europ. Assoc. Theoret. Comput. Sci.*, 21:12–20, 1983.

[Blumer *et al.*, 1985] A. Blumer, J. Blumer, A. Ehrenfeucht, D. Haussler, M. T. Chen, and J. Seiferas. The smallest automaton recognizing the subwords of a text. *Theoret. Comput. Sci.*, 40:31–55, 1985.

[Botsford and Harman, 1992] J. L. Botsford and J. G. Harman. Cyclic AMP in prokaryotes. *Microbiological Rev.*, 56:100–122, 1992.

[Brutlag *et al.*, 1990] D. L. Brutlag, J-P. Dautricourt, S. Maulik, and J. Relph. Improved sensitivity of biological sequence database searches. *Comput. Applic. Biosc.*, 6:237–245, 1990.

[Brutlag *et al.*, 1993] D. L. Brutlag, J.-P. Dautricourt, R. Diaz, , J. Fier, B. Moxon, and R. F. J. Stamm. BLAZE^{TM}: an implementation of the Smith-Waterman sequence comparison algorithm on a massively parallel computer. *Computers Chem.*, 17:203–207, 1993.

[Cailliez and Pages, 1976] F. Cailliez and J. P. Pages. *Introduction à l'Analyse de Données.* Société de Mathématiques Appliquées et de Sciences Humaines, 1976.

[Carafa *et al.*, 1990] Y. D'Aubenton Carafa, E. Brody, and C. Thermes. Prediction of Rho-independent E*scherichia coli* transciption terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, 216:835–858, 1990.

[Cardon and Crochemore, 1982] A. Cardon and M. Crochemore. Partioning a graph in $O(\mid A \mid \log_2 \mid V \mid)$. *Theoret. Comput. Sci.*, 19:85–98, 1982.

[Cardon and Stormo, 1992] L. R. Cardon and G. D. Stormo. Expectation Maximization algorithm for identifying protein-binding sites with variable lengths from unaligned DNA fragments. *J. Comp. Biol.*, 223:139–170, 1992.

[Carrillo and Lipman, 1988] H. Carrillo and D. J. Lipman. The multiple sequence alignment problem in biology. *SIAM J. Appl. Math.*, 48:1073–1082, 1988.

[Chen and Seiferas, 1985] M. T. Chen and J. Seiferas. Efficient and elegant subword tree construction. In A. Apostolico and Z. Galil, editors, *Combinatorial Algorithms on Words*, volume 12 of *F*, pages 97–107. NATO Advanced Science Institutes, 1985.

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

[Chothia, 1992] C. Chothia. One thousand families for the molecular biologist. *Nature*, 357:543–544, 1992.

[Clarke and Cooke, 1992] G. M. Clarke and D. Cooke. *A Basic Course in Statistics*. Edward Arnold, 1992.

[Clarke, 1986] A. R. Clarke. Molecular chaperones in protein folding and translocation. *Curr. Opin. Struct. Biol.*, 14:141–158, 1986.

[Clift *et al.*, 1986] B. Clift, D. Haussler, R. McConnell, T. D. Schneider, and G. D. Stormo. Sequence landscapes. *Nucleic Acids Res.*, 14:141–158, 1986.

[Cobbs, 1994] A.L. Cobbs. Fast identification of approximately matching substrings. In M. Crochemore and D. Gusfield, editors, *Combinatorial Pattern Matching*, volume 807 of *Lecture Notes in Computer Science*, pages 64–74. Springer Verlag, 1994.

[Collins and Coulson, 1990] J. F. Collins and A. F. W. Coulson. Significance of protein sequence similarities. In *Meth. Enzymol.*, volume 183, pages 474–487. 1990.

[Collins *et al.*, 1988] J. F. Collins, A. F. W. Coulson, and A. Lyall. The significance of protein sequence similarities. *Comput. Appl. Biosci.*, 4:67–71, 1988.

[Combrugghe *et al.*, 1984] B. Combrugghe, S. Busby, and H. Buc. Cyclic AMP receptor protein: role in transcription activation. *Science*, 224:831–838, 1984.

[Cornette *et al.*, 1987] J. L. Cornette, K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, and C. DeLisi. Hydrophobicity scales and computational techniques for detecting amphipatic structures in proteins. *J. Mol. Biol.*, 195:659–685, 1987.

[Corpet, 1988] F. Corpet. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, 16:10881–10890, 1988.

[Coulson *et al.*, 1987] A. F. W. Coulson, J. F. Collins, and A. Lyall. Protein and nucleic acid sequence database searching: a suitable case for parallel processing. *Comput. J.*, 30:420–424, 1987.

[Creighton, 1993] T. E. Creighton. *Proteins*. Freeman, 1993.

[Crochemore and Rytter, 1991] M. Crochemore and W. Rytter. Usefulness of the Karp-Miller-Rosenberg algorithm in parallel computations on strings and arrays. *Theoret. Comput. Sci.*, 88:59–82, 1991.

[Crochemore and Rytter, 1994] M. Crochemore and W. Rytter. *Text algorithms*. Oxford University Press, 1994.

[Crochemore, 1981] M. Crochemore. An optimal algorithm for computing the repetitions in a word. *Inf. Proc. Letters*, 12:244–250, 1981.

[Crochemore, 1985] M. Crochemore. Optimal factor transducers. In A. Apostolico and Z. Galil, editors, *Combinatorial Algorithms on Words*, volume 12 of *F*, pages 31–43. NATO Advanced Science Institutes, 1985.

[Crochemore, 1986] M. Crochemore. Transducers and repetitions. *Theoret. Comput. Sci.*, 45:63–86, 1986.

[Danchin, 1990] A. Danchin. *Une Aurore de Pierres. Aux Origines de la Vie.* Éditions Seuil, 1990.

[Danchin, 1995] A. Danchin. 1995. communication personnelle.

[Dandekar and Hentze, 1995] T. Dandekar and M. W. Hentze. Finding the hairpin in the haystack: searching for RNA motifs. *Trends Genet.*, 11:45–50, 1995.

[Dandekar *et al.*, 1991] T. Dandekar, R. Stripecke, N. K. Gray, B. Goossen, A. Constable, H. E. Johansson, and M. W. Hentze. Identification of a novel iron-responsive element in murine and human erythroid δ-amino-levulinic acid synthase mRNA. *EMBO J.*, 10:1903–1909, 1991.

[Daune, 1993] M. Daune. *Biophysique Moléculaire. Structures en Mouvement.* InterÉditions, 1993.

[Dayhoff and Eck, 1972] M. O. Dayhoff and R. V. Eck. A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, pages 89–99. Natl. Biomed. Res. Found., Washington, 1972.

[Dayhoff *et al.*, 1978] M.O. Dayhoff, R.M. Schwartz, and B.C. Orcutt. A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, volume 5 suppl.3, pages 345–352. Natl. Biomed. Res. Found., 1978.

[Dayhoff *et al.*, 1983] M. O. Dayhoff, W. C. Barker, and L. T. Hunt. Establishing homologies in protein sequences. In *Meth. Enzymol.*, volume 91, pages 524–544. Academic Press, 1983.

[Doolittle, 1981] R. F. Doolittle. Similar amino acid sequences: chance or common ancestry. *Science*, 214:149–159, 1981.

[Doolittle, 1986] R. F. Doolittle. *Of URFs and ORFs: a primer on how to analyze derived amino acid sequences.* University Science Books, 1986.

[Duret *et al.*, 1994] L. Duret, D. Mouchiroud, and M. Gouy. HOVERGEN: a database of homologous vertebrate genes. *Nucleic Acids Res.*, 22:2360–2365, 1994.

[Duret, 1995] L. Duret. *Évolution des séquences non-codantes chez les vertébrés : Étude des contraintes fonctionnelles et structurales par analyse comparative de gènes homologues.* PhD thesis, 1995. Thèse de doctorat - Université Claude Bernard - Lyon I.

[Ebright, 1993] R. H. Ebright. Transcription activation at Class I Cap-dependent promoters. *Molecular Microbiology*, 8:797–802, 1993.

[Escalier *et al.*, 1996] V. Escalier, J. Pothier, H. Soldano, and A. Viari. Pairwise and multiple identification of three dimensional common substructures in proteins. 1996. submitted to J. Comp. Biol.

[Fay, 1992] P. Fay. Oxygen relations of nitrogen fixation in Cyanobacteria. *Microbiol. Rev.*, 56:340–373, 1992.

[Feng and Doolittle, 1987] D.-F. Feng and R. F. Doolittle. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, 25:351–360, 1987.

[Feng and Doolittle, 1990] D.-F. Feng and R. F. Doolittle. Progressive alignment and phylogenetic tree construction of protein sequences. *Meth. Enzymol.*, 183:375–402, 1990.

[Feng *et al.*, 1985] D.-F. Feng, M. S. Johnson, and R. F. Doolittle. Aligning amino acids sequences: comparison of commonly used methods. *J. Mol. Evol.*, 21:112–125, 1985.

[Fickett, 1984] J. W. Fickett. Fast optimal alignment. *Nucleic Acids Res.*, 12:175–179, 1984.

[Finkelstein *et al.*, 1993] A. V. Finkelstein, A. M. Gutun, and A. Y. Badretdinov. Why are the same protein folds used to perform different functions? *FEBS*, 325:23–28, 1993.

[Fitch and Smith, 1983] W. M. Fitch and T. F. Smith. Optimal sequence alignments. *Proc. Natl. Acad. Sci. USA*, 80:1382–1386, 1983.

[Fitch, 1966] W. M. Fitch. An improved method of testing for evolutionary homology. *J. Mol. Biol.*, 16:9–16, 1966.

[Fitch, 1967] W. M. Fitch. Construction of phylogenetic trees. *Science*, 15:299–304, 1967.

[Fredman, 1984] M. L. Fredman. Algorithms for computing evolutionary similarity measures with length independent gap penalties. *Bull. Math. Biol.*, 46:553–566, 1984.

[Freifelder, 1990] D. Freifelder. *Biologie Moléculaire*. Masson, 1990.

[Galas *et al.*, 1985] D. J. Galas, M. Eggert, and M. S. Waterman. Rigorous pattern-recognition methods for DNA sequences. Analysis of promoter sequences from E*scherichia coli*. *J. Mol. Biol.*, 186:117–128, 1985.

[Galil and Giancarlo, 1989] Z. Galil and R. Giancarlo. Speeding up dynamic programing with application to molecular biology. *Theoret. Comput. Sci.*, 64:107–118, 1989.

[Garey and Johnson, 1979] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness*. Freeman, 1979.

[Gascuel and Danchin, 1986] O. Gascuel and A. Danchin. Protein export in prokaryotes and eukaryotes: indications of a difference in the mechanism of exportation. *J. Mol. Evol.*, 24:130–142, 1986.

[Gaston *et al.*, 1990] K. Gaston, A. Bell, A. Kolb, H. Buc, and S. Busby. Stringent spacing requirements for transcription activation by CRP. *Cell*, 62:733–743, 1990.

[Gautheret *et al.*, 1990] D. Gautheret, F. Major, and R. J. Cedergren. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput. Appl. Biosci.*, 6:325–331, 1990.

[Gerstein and Altman, 1995] M. Gerstein and R. B. Altman. Using a measure of structural variation to define a core for the globins. *Comput. Appl. Biosci.*, 11:633–644, 1995.

[Gething and Sambrook, 1992] M.-J. Gething and J. Sambrook. Protein folding in the cell. *Nature*, 355:33–45, 1992.

[Goldstein and Waterman, 1994] L. Goldstein and M. S. Waterman. Approximations to profile score distributions. *J. Comp. Biol.*, 1:93–104, 1994.

[Gombrich, 1987] E. H. Gombrich. *L'Art et l'Illusion. Psychologie de la Représentation Picturale.* Éditions Gallimard, 1987.

[Gonnet *et al.*, 1992] G. H. Gonnet, M. A. Cohen, and S. A. Benner. Exhaustive matching of the entire protein sequence database. *Science*, 256:1443–1445, 1992.

[Gonnet *et al.*, 1994] G. H. Gonnet, M. A. Cohen, and S. A. Benner. Analysis of amino acid substitution during divergent evolution: the 400 by 400 dipeptide substitution matrix. *Biochem. Biophys. Res. Commun.*, 199:489–496, 1994.

[Gotoh, 1982] O. Gotoh. An improved algorithm for matching biological sequences. *J. Mol. Biol.*, 162:705–708, 1982.

[Gotoh, 1986] O. Gotoh. Alignment of three biological sequences with an efficient traceback procedure. *J. Theor. Biol.*, 121:327–337, 1986.

[Gotoh, 1990] O. Gotoh. Optimal sequence alignment allowing for long gaps. *Bull. Math. Biol.*, 52:359–373, 1990.

[Grantham, 1974] R. Grantham. Amino acid difference formula to help explain protein evolution. *Science*, 185:862–864, 1974.

[Gribskov *et al.*, 1987] M. Gribskov, M. McLachlan, and D. Eisenberg. Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA*, 84:4355–4358, 1987.

[Gribskov *et al.*, 1988] M. Gribskov, M. Homyak, J. Edenfield, and D. Eisenberg. Profile scanning for three-dimensional structural patterns in protein sequences. *Comput. Appl. Biosci.*, 4:61–66, 1988.

[Gribskov *et al.*, 1990] M. Gribskov, R. Luthy, and D. Eisenberg. Profile analysis. *Meth. Enzymol.*, 183:146–159, 1990.

[Gribskov, 1992] M. Gribskov. The language metaphor in sequence analysis. *Comput. Chem.*, 16:85–88, 1992.

[Gupta *et al.*, 1995] S. K. Gupta, J. D. Kececioglu, and A. A. Schaffer. Making the shortest-paths approach to sum-of-pairs multiple sequence alignment more space efficient in practice. In Z. Galil and E. Ukkonen, editors, *Combinatorial Pattern Matching*, volume 937 of *Lecture Notes in Computer Science*, pages 128–143. Springer Verlag, 1995.

[Gusfield *et al.*, 1992] D. Gusfield, K. Balasubramanian, and D. Naor. Parametric optimization of sequence alignment. pages 432–439. Proc. of the Third Annual ACM-SIAM Symp. on Discrete Algorithms, 1992.

[Gusfield, 1993] D. Gusfield. Efficient methods for multiple sequence alignment with guaranteed error bounds. *Bull. Math. Biol.*, 55:141–154, 1993.

[Gusfield, 1995] D. Gusfield. Algorithms on strings: A dual view from computer science and computational molecular biology. manuscript - presented at 6th Annual Combinatorial Pattern Matching - Espoo, Finland, 1995.

[Hasemann *et al.*, 1995] C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, and J. Deisenhofer. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Curr. Opin. Struct. Biol.*, 2:41–62, 1995.

[Helmann, 1995] J. D. Helmann. Compilation and analysis of B*acillus subtilis* α-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.*, 23:2351–2360, 1995.

[Henikoff and Henikoff, 1991] S. Henikoff and J. G. Henikoff. Automatic generation of protein blocks for database searching. *Nucleic Acids Res.*, 19:6565–6572, 1991.

[Henikoff and Henikoff, 1992] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices for protein blocks. *Proc. Natl. Acad. Sci. USA*, 89:10915–10919, 1992.

[Henikoff, 1994] S. Henikoff. Comparative sequence analysis: finding genes. In D. W. Smith, editor, *Biocomputing. Informatics and Genome Projects*, pages 87–118. Academic Press, 1994.

[Heringa and Argos, 1993] J. Heringa and R. Argos. A method to recognize distant repeats in protein sequences. *Proteins: struct., funct., and genetics*, 17:391–411, 1993.

[Heringa, 1994] J. Heringa. The evolution and recognition of protein sequence repeats. *Comput. Chem.*, 18:233–244, 1994.

[Hertz *et al.*, 1990] G. Z. Hertz, G. W. Hartzell, and G. D. Stormo. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput. Appl. Biosci.*, 6:81–92, 1990.

[Higgins and Sharp, 1988] D. G. Higgins and P. M. Sharp. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73:237–244, 1988.

[Higgins *et al.*, 1992] D. G. Higgins, A. J. Bleasby, and R. Fuchs. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, 8:189–191, 1992.

[Hogeweg and Hesper, 1984] P. Hogeweg and B. Hesper. The alignment of sets of sequences and the construction of phylogenetic trees. An integrated method. *J. Mol. Evol.*, 20:175–186, 1984.

[Howard and Rees, 1994] J. B. Howard and D. C. Rees. Nitrogenase: A nucleotide-dependent molecular switch. *Ann. Rev. Biochem.*, 94:235–264, 1994.

[Hui, 1992] L. C. K. Hui. Color set size problem with applications to string matching. In A. Apostolico, M. Crochemore, Z. Galil, and U. Manber, editors, *Combinatorial Pattern Matching*, volume 644 of *Lecture Notes in Computer Science*, pages 230–243. Springer-Verlag, 1992.

[Jacob, 1970] F. Jacob. *La Logique du Vivant. Une Histoire de l'Hérédité*. Éditions Gallimard, 1970.

[Jacob, 1981] F. Jacob. *Le Jeu des Possibles*. Éditions Fayard, 1981.

[Jacob, 1987] F. Jacob. *La Statue Intérieure*. Éditions Odile Jacob, 1987.

[Jean *et al.*, 1996] P. Jean, J. Pothier, P. Dansette, D. Mansuy, and A. Viari. Automated multiple analysis of protein structures: application to homology modeling of cytochromes p450. 1996. submitted to *Proteins: struct., funct., and genetics*.

[Johnson and Doolittle, 1986] M. S. Johnson and R. F. Doolittle. A method for the simultaneous alignment of three or more amino acid sequences. *J. Mol. Evol.*, 23:267–278, 1986.

[Johnson and Overington, 1993] M. S. Johnson and J. P. Overington. A structural basis for sequence comparisons. An evaluation of scoring methodologies. *J. Mol. Biol.*, 233:716–738, 1993.

[Jones *et al.*, 1992] D. T. Jones, W. R. Taylor, and J. M. Thornton. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, 8:275–282, 1992.

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

[Karlin and Altschul, 1990] S. Karlin and S. F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA*, 87:2264–2268, 1990.

[Karlin and Altschul, 1993] S. Karlin and S. F. Altschul. Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. USA*, 90:5873–5877, 1993.

[Karlin and Ghandour, 1985a] S. Karlin and G. Ghandour. Multiple alphabet amino-acid sequence comparisons of the immunoglobulin κ-chain constant domain. *Proc. Natl. Acad. Sci. USA*, 82:8597–8601, 1985.

[Karlin and Ghandour, 1985b] S. Karlin and G. Ghandour. The use of multiple alphabets in κ-gene immunoglobulin DNA sequence comparisons. *The EMBO J.*, 4:1217–1223, 1985.

[Karlin *et al.*, 1988a] S. Karlin, M. Morris, G. Ghandour, and M.-Y. Leung. Algorithms for identifying local molecular sequence features. *Comput. Appl. Biosci.*, 4:41–51, 1988.

[Karlin *et al.*, 1988b] S. Karlin, M. Morris, G. Ghandour, and M.-Y. Leung. Efficient algorithms for molecular sequence analysis. *Proc. Natl. Acad. Sci. USA*, 85:841–845, 1988.

[Karp *et al.*, 1972] R. M. Karp, R. E. Miller, and A. L. Rosenberg. Rapid identification of repeated patterns in strings, trees and arrays. pages 125–136. Proc. 4th Annu. ACM Symp. Theory of Computing, 1972.

[Karplus and McCammon, 1983] M. Karplus and J. A. McCammon. Dynamics of proteins: elements and function. *Annu. Rev. Biochem.*, 53:263–300, 1983.

[Karplus *et al.*, 1987] M. Karplus, A. T. Brunger, R. Elber, and J. Kuriyan. Molecular dynamics: applications to proteins. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume LII, pages 381–390. NATO Advanced Science Institutes, 1987.

[Karplus, 1995] K. Karplus. Evaluating regularizers for estimating distributions of amino acids. pages 188–196, Cambridge, England, 1995. Third International Symposium on Intelligent Systems for Molecular Biology.

[Kaufmann, 1993] S. A. Kaufmann. *The Origins of Order: Self Organization and Selection in Evolution.* Oxford University Press, 1993.

[Kececioglu, 1993] J. Kececioglu. The maximum weight trace problem in multiple sequence alignement. In Z. Galil A. Apostolico, M. Crochemore and U. Manber, editors, *Combinatorial Pattern Matching*, volume 684 of *Lecture Notes in Computer Science*, pages 106–119. Springer-Verlag, 1993.

[Kimura, 1983] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1983.

[Klotz and Blanken, 1981] L. C. Klotz and R. L. Blanken. A practical method for calculating evolutionary trees from sequence data. *J. Theor. Biol.*, 91:261–272, 1981.

[Klotz *et al.*, 1979] L. C. Klotz, N. Komar, R. L. Blanken, and R. M. Mitchell. Calculation of evolutionary trees from sequence data. *Proc. Natl. Acad. Sci. USA*, 76:4516–4520, 1979.

[Knight and Myers, 1995] J. R. Knight and E. W. Myers. Super-pattern matching. *Algorithmica*, 13:211–243, 1995.

[Kosaraju, 1994] S. R. Kosaraju. Computation of squares in a string. In M. Crochemore and D. Gusfield, editors, *Combinatorial Pattern Matching*, volume 807 of *Lecture Notes in Computer Science*, pages 146–150. Springer-Verlag, 1994.

[Koshland *et al.*, 1966] D. E. Koshland, Nemethy, and Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5:365–385, 1966.

[Krogh *et al.*, 1994] A. Krogh, M. Brown, I. S. Mian, K. Sjoelander, and D. Haussler. Hidden Markov model in computational biology. Applications to protein modeling. *J. Mol. Biol.*, 235:1501–1531, 1994.

[Kruh, 1987] J. Kruh. *Biochimie. 1. Biologie Cellulaire et Moléculaire.* Hermann, 1987.

[Kruskal, 1956] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.

[Landraud *et al.*, 1989] A. M. Landraud, J. F. Avril, and P. Chretienne. An algorithm for finding a common structure shared by a family of strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):890–895, 1989.

[Lawrence and Reilly, 1990] C. E. Lawrence and A. A. Reilly. An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences. *Proteins: struct., funct., and genetics*, 7:41–51, 1990.

[Lawrence *et al.*, 1993] C. E. Lawrence, S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wooton. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, 262:208–214, 1993.

[Lefevre and Ikeda, 1993a] C. Lefevre and J.-E. Ikeda. Pattern recognition in DNA sequences and its application to consensus foot-printing. *Comput. Appl. Biosci.*, 9:349–354, 1993.

[Lefevre and Ikeda, 1993b] C. Lefevre and J.-E. Ikeda. The position end-set tree: A small automaton for word recognition in biological sequences. *Comput. Appl. Biosci.*, 9:343–348, 1993.

[Lefevre and Ikeda, 1994] C. Lefevre and J.-E. Ikeda. A fast word search algorithm for the representation of sequence similarity in genomic DNA. *Nucleic Acids Res.*, 22:404–411, 1994.

[Lesk and Chothia, 1980] 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.*, 136:225–270, 1980.

[Leung *et al.*, 1991] M.-Y. Leung, B. E. Blaisdell, C. Burge, and S. Karlin. An efficient algorithm for identifying matches with errors in multiple long molecular sequences. *J. Mol. Biol.*, 221:1367–1378, 1991.

[Levenshtein, 1966] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Cyber. Contr. Theory*, 10:707–710, 1966.

[Levin *et al.*, 1986] J. M. Levin, B. Robson, and J. Garnier. An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett.*, 205:303–308, 1986.

[Lewin, 1994] B. Lewin. *Genes V*. Oxford University Press, 1994.

[Li and Graur, 1991] W-H. Li and D. Graur. *Fundamentals of Molecular Evolution*. Sinauer Associates, 1991.

[Lipman and Pearson, 1985] D. J. Lipman and W. R. Pearson. Rapid and sensitive protein similarity searches. *Science*, 227:1435–1441, 1985.

[Lipman *et al.*, 1989] D. J. Lipman, S. F. Altschul, and J. D. Kececioglu. A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci. USA*, 86:4412–4415, 1989.

[Lorenz, 1975] K. Lorenz. *L'Envers du Miroir. Une Histoire Naturelle de la Connaissance.* Flammarion, 1975.

[Luthy *et al.*, 1991] R. Luthy, A. D. McLachlan, and D. Eisenberg. Secondary structure-based profiles: use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. *Proteins: struct., funct., and genetics*, 10:229–239, 1991.

[Martinez, 1983] H. M. Martinez. An efficient method for finding repeats in molecular sequences. *Nucleic Acids Res.*, 11:4629–4634, 1983.

[Martinez, 1988] H. M. Martinez. A flexible multiple sequence alignment program. *Nucleic Acids Res.*, 16:1683–1691, 1988.

[Masek and Paterson, 1983] W. J. Masek and M. S. Paterson. How to compute string-edit distances quickly. In D. Sankoff and J. B. Kruskal, editors, *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*, pages 337–349. Addison-Wesley, 1983.

[McCreight, 1976] E. M. McCreight. A space-economical suffix tree construction algorithm. *J. ACM*, 23:262–272, 1976.

[McLachlan, 1971] A. D. McLachlan. Test for comparing related amino acid sequences. Cytochrome c and cytochrome c_{551}. *J. Mol. Biol.*, 61:409–424, 1971.

[Médigue *et al.*, 1993] C. Médigue, A. Viari, A. Hénaut, and A. Danchin. Colibri: a functional data base for the E*scherichia coli* genome. *Microbiol. Rev.*, 57:623–654, 1993.

[Mehldau and Myers, 1993] G. Mehldau and E. W. Myers. A system for pattern matching applications on biosequences. *Comput. Applic. Biosc.*, 9:299–314, 1993.

[Miller and Myers, 1988] W. Miller and E. W. Myers. Sequence comparison with concave weighting functions. *Bull. Math. Biol.*, 50:97–120, 1988.

[Miyata *et al.*, 1979] T. Miyata, S. Miyazawa, and T. Yasunaga. Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.*, 12:219–236, 1979.

[Miyazawa and Jernigan, 1993] S. Miyazawa and R. L. Jernigan. A new substitution matrix for protein sequence searches based on contact frequencies in protein structures. *Protein Eng.*, 6:267–278, 1993.

[Monod *et al.*, 1963] J. Monod, J. P. Changeux, and F. Jacob. Allosteric proteins and cellular control mechanisms. *J. Mol. Biol.*, 6:306–329, 1963.

[Monod *et al.*, 1965] J. Monod, J. Wyman, and J.-P. Changeux. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, 12:88–118, 1965.

[Monod, 1970] J. Monod. *Le Hasard et la Nécessité*. Éditions du Seuil, 1970.

[Murata *et al.*, 1985] M. Murata, J. S. Richardson, and J. L. Sussman. Simultaneous comparison of three protein sequences. *Proc. Natl. Acad. Sci. USA*, 82:3073–3077, 1985.

[Murata, 1990] M. Murata. Three-way Needleman-Wunsch algorithm. *Meth. Enzymol.*, 183:365–375, 1990.

[Myers and Miller, 1988] E. W. Myers and W. Miller. Optimal alignment in linear space. *Comput. Applic. Biosci.*, 4:11–17, 1988.

[Myers, 1986] E. W. Myers. An O(ND) difference algorithm and its variations. *Algorithmica*, 1:251–266, 1986.

[Nakatsu *et al.*, 1982] N. Nakatsu, Y. Kambayashi, and S. Yajima. A longest common subsequence algorithm suitable for similar text strings. *Acta Informatica*, 18:171–179, 1982.

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

[Nei, 1987] M. Nei. *Molecular Evolutionary Genetics*. Columbia University Press, 1987.

[Nelson *et al.*, 1993] D. R. Nelson, T. Kamataki, D. J. Waxman, F. P. Guengerich, R. W. Estabrook, R. Feyereisen, F. J. Gonzalez, M. J. Coon, I. C. Gunsalus, O. Gotoh, K. Okuda, and D. W. Nebert. The P450 Superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and Cell Biology*, 12:1–50, 1993.

[Neuwald and Green, 1994] A. F. Neuwald and P. Green. Detecting patterns in protein sequences. *J. Mol. Biol.*, 239:698–712, 1994.

[Overington *et al.*, 1990] J. Overington, M. S. Johnson, A. Sali, and T. L. Blundell. Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. R. Soc. Lond. B*, 241:132–145, 1990.

[Overington *et al.*, 1992] J. Overington, D. Donnelly, M. S. Johnson, A. Sali, and T. L. Blundell. Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Science*, 1:216–226, 1992.

[Pascarella and Argos, 1992] S. Pascarella and P. Argos. Analysis of insertions/deletions in protein structures. *J. Mol. Biol.*, 225:461–471, 1992.

[Pau, 1989] R.N. Pau. Nitrogenases without molybdenum. *Trends Biochem. Sci.*, 14:183–186, 1989.

[Pearson and Lipman, 1988] W. R. Pearson and D. J. Lipman. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, 85:2444–2448, 1988.

[Pearson, 1990] W. R. Pearson. Rapid and sensitive sequence comparison with FASTP and FASTA. *Meth. Enzymol.*, 183:63–98, 1990.

[Pearson, 1991] W. R. Pearson. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, 11:635–650, 1991.

[Pevzner, 1992a] P. A. Pevzner. Multiple alignement with guaranteed error bounds and communication costs. In Z. Galil A. Apostolico, M. Crochemore and U. Manber, editors, *Combinatorial Pattern Matching*, volume 644 of *Lecture Notes in Computer Science*, pages 202–210. Springer-Verlag, 1992.

[Pevzner, 1992b] P. A. Pevzner. Multiple alignment, communication cost, and graph matching. *SIAM J. Appl. Math.*, 52:1763–1776, 1992.

[Posfai *et al.*, 1989] J. Posfai, A.S. Bhagwat, G. Posfai, and R.J. Roberts. Prediction motifs derived from cytosine methyltransferases. *Nucleic Acids Res.*, 17:2421–2435, 1989.

[Pothier, 1993] J. Pothier. 1993. communication personnelle.

[Prim, 1957] R. C. Prim. Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36:1389–1401, 1957.

[Ptitsyn and Volkenstein, 1986] O. B. Ptitsyn and M. V. Volkenstein. Protein structures and neutral theory of evolution. *J. Biomol. Struct. Dyn.*, 4(1):137–156, 1986.

[Ramachandran *et al.*, 1963] G. N. Ramachandran, C. Ramakrishnan, and V. Sesisakharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.

[Rao, 1987] J. K. M. Rao. New scoring matrix for amino acid residue exchanges based on residue characteristic physical parameters. *Int. J. Pept. Protein Res.*, 29:276–281, 1987.

[Risler *et al.*, 1988] J.-L. Risler, M.-O. Delorme, H. Delacroix, and A. Henaut. Amino acid substitutions in structurally related proteins. A pattern recognition approach. *J. Mol. Biol.*, 204:1019–1029, 1988.

[Robson and Postgate, 1980] R. L. Robson and J. R. Postgate. Oxygen and hydrogen in biological nitrogen fixation. *Ann. Rev. Microbiol.*, 34:183–207, 1980.

[Rooman and Wodak, 1988] M. J. Rooman and S. J. Wodak. Identification of predictive sequence motifs limited by protein structure database size. *Nature*, 335:45–49, 1988.

[Rooman *et al.*, 1990] M. J. Rooman, J. Rodriguez, and S. J. Wodak. Relations between protein sequence and structure and their significance. *J. Mol. Biol.*, 213:337–350, 1990.

[Rooman *et al.*, 1992] M. J. Rooman, A. K. J.-P., and S. J. Wodak. Extracting information on folding from the amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry*, 31:10226–10238, 1992.

[Russell, 1992] P. J. Russell. *Genetics*. Harper/Collins, 1992.

[Sagot and Viari, 1996] M.-F. Sagot and A. Viari. A double combinatorial approach to discovering patterns in biological sequences. In D. Hirschberg and G. Myers, editors, *Combinatorial Pattern Matching*, volume 1075 of *Lecture Notes in Computer Science*, pages 186–208. Springer-Verlag, 1996.

[Sagot *et al.*, 1995a] M.-F. Sagot, V. Escalier, A. Viari, and H. Soldano. Searching for repeated words in a text allowing for mismatches and gaps. pages 87–100, Viñas del Mar, Chili, 1995. Second South American Workshop on String Processing.

[Sagot *et al.*, 1995b] M.-F. Sagot, A. Viari, J. Pothier, and H. Soldano. Finding flexible patterns in a text - an application to 3D molecular matching. *Comput. Appl. Biosci.*, 11:59–70, 1995. presented at First International IEEE Workshop on Shape and Pattern Matching in Computational Biology, Seattle, Washington, USA.

[Sagot *et al.*, 1995c] M.-F. Sagot, A. Viari, and H. Soldano. A distance-based block searching algorithm. pages 322–331, Cambridge, England, 1995. Third International Symposium on Intelligent Systems for Molecular Biology.

[Sagot *et al.*, 1995d] M.-F. Sagot, A. Viari, and H. Soldano. Multiple comparison: a peptide matching approach. In Z. Galil and E. Ukkonen, editors, *Combinatorial Pattern Matching*, volume 937 of *Lecture Notes in Computer Science*, pages 366–385. Springer-Verlag, 1995. to appear in *Theoret. Comput. Sci.*

[Sagot *et al.*, 1996] M.-F. Sagot, A. Viari, J. Pothier, V. Escalier, and H. Soldano. Multiple comparison in biology: some mathematical formalizations of the problem and combinatorial approaches to solve it. unpublished, 1996.

[Saitou and Nei, 1987] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4:406–425, 1987.

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

[Sankoff and Cedergren, 1983] D. Sankoff and R. J. Cedergren. Simultaneous comparison of three or more sequences related by a tree. In D. Sankoff and J. B. Kruskal, editors, *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*, pages 253–263. Addison-Wesley, 1983.

[Sankoff and Kruskal, 1983] D. Sankoff and J. B. Kruskal. *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison.* Addison-Wesley, 1983.

[Sankoff *et al.*, 1973] D. Sankoff, C. Morel, and R. J. Cedergren. Evolution of 5S RNA and the nonrandomness of base replacement. *Nature New Biology*, 245:232–234, 1973.

[Sankoff *et al.*, 1976] D. Sankoff, R. J. Cedergren, and G. Lapalme. Frequency of insertion-deletion, transversion, and transition in the evolution of 5S ribosomal RNA. *J. Mol. Evol.*, 7:133–149, 1976.

[Sankoff, 1975] D. Sankoff. Minimal mutation trees of sequences. *SIAM J. Appl. Math.*, 28:35–42, 1975.

[Santibanez and Rohde, 1987] M. Santibanez and K. Rohde. A multiple alignment program for protein sequences. *Comput. Appl. Biosci.*, 3:111–114, 1987.

[Schneider *et al.*, 1986] T. D. Schneider, G. D. Stormo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, 188:415–431, 1986.

[Schuler *et al.*, 1991] G. D. Schuler, S. F. Altschul, and D. J. Lipman. A workbench for multiple alignment construction and analysis. *Proteins*, 9:180–190, 1991.

[Sellers, 1974] P. H. Sellers. On the theory and computation of evolutionary distances. *SIAM J. Appl. Math.*, 26:787–793, 1974.

[Sheridan and Venkataraghavan, 1992] R. P. Sheridan and R. Venkataraghavan. A systematic search for protein signature sequences. *Proteins: struct., funct., and genetics*, 14:16–28, 1992.

[Smith and Smith, 1990] R. F. Smith and T. S. Smith. Automatic generation of primary sequence patterns from sets of related protein sequences. *Proc. Natl. Acad. Sci. USA*, 87:118–122, 1990.

[Smith and Smith, 1992] R. F. Smith and T. F. Smith. Pattern-Induced Multi-sequence Alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for comparative protein modelling. *Protein Eng.*, 5:35–41, 1992.

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

[Smith *et al.*, 1990] H. O. Smith, T. M. Annau, and S. Chandrasegaran. Finding sequence motifs in groups of functionally releated proteins. *Proc. Natl. Acad. Sci. USA*, 87:826–830, 1990.

[Sneath and Sokal, 1973] H. A. Sneath and R. R. Sokal. *Numerical Taxonomy.* Freeman, 1973.

[Sobel and Martinez, 1986] E. Sobel and H. M. Martinez. A multiple sequence alignment program. *Nucleic Acids Res.*, 14:363–374, 1986.

[Soldano *et al.*, 1995] H. Soldano, A. Viari, and M. Champesme. Searching for flexible repeated patterns using a non transitive similarity relation. *Pattern Recognition Letters*, 16:233–246, 1995.

[Spouge, 1989] J. L. Spouge. Speeding up dynamic programming algorithms for finding optimal lattice graphs. *SIAM J. Appl. Math.*, 5:1552–1566, 1989.

[Spouge, 1991] J. L. Spouge. Fast optimal alignment. *Comput. Appl. Biosci.*, 7:1–7, 1991.

[Stormo, 1990] G. D. Stormo. Consensus patterns in DNA. *Meth. Enzymol.*, 183:211–221, 1990.

[Stryer, 1981] L. Stryer. *Biochemistry.* Freeman, 1981.

[Subbiah and Harrison, 1989] S. Subbiah and S. C. Harrison. A method for multiple sequence alignments with gaps. *J. Mol. Biol.*, 209:539–548, 1989.

[Tatusov and Koonin, 1994] R. L. Tatusov and E. V. Koonin. A simple tool to search for sequence motifs that are conserved in BLAST outputs. *Comput. Appl. Biosci.*, 10:0–0, 1994.

[Tatusov *et al.*, 1994] R. L. Tatusov, S. F. Altschul, and E. V. Koonin. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. USA*, 91:12091–12095, 1994.

[Taylor, 1986a] W. R. Taylor. The classification of amino acid conservation. *J. Theor. Biol.*, 119:205–218, 1986.

[Taylor, 1986b] W. R. Taylor. Identification of protein sequence homology by consensus template alignment. *J. Mol. Biol.*, 183:233–258, 1986.

[Taylor, 1986c] W. R. Taylor. Toward a practical grammar of protein structure: protein structure prediction by template fitting. In R. Fletterick and M. Zoller, editors, *Computer Graphics and Molecular Modeling*, pages 77–84. Cold Spring Harbor Laboratory, 1986.

[Taylor, 1987] W. R. Taylor. Multiple sequence alignment by a pairwise algorithm. *Comput. Appl. Biosci.*, 3:81–87, 1987.

[Taylor, 1989] W. R. Taylor. A template based method of pattern matching in protein sequences. *Prog. Biophys. Molec. Biol.*, 54:159–252, 1989.

[Taylor, 1990] W. R. Taylor. Hierarchical method to align large numbers of biological sequences. *Meth. Enzymol.*, 183:456–474, 1990.

[Thompson *et al.*, 1994a] J. D. Thompson, D. G. Higgins, and T. J. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignement through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673–4680, 1994.

[Thompson *et al.*, 1994b] J. D. Thompson, D. G. Higgins, and T. J. Gibson. Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Appl. Biosci.*, 10:19–29, 1994.

[Thornton *et al.*, 1991] J. M. Thornton, T. P. Flores, D. T. Jones, and M. B. Swindells. Prediction of progress at last. *Nature*, 354:105–106, 1991.

[Ukkonen and Wood, 1993] E. Ukkonen and D. Wood. Approximate string matching with suffix automata. *Algorithmica*, 10:353–364, 1993.

[Ukkonen, 1985a] E. Ukkonen. Algorithms for approximate string matching. *Inform. Contr.*, 64:100–118, 1985.

[Ukkonen, 1985b] E. Ukkonen. Finding approximate patterns in strings. *J. Algorithms*, 6:132–137, 1985.

[Ukkonen, 1992] E. Ukkonen. Constructing suffix trees on-line in linear time. pages 484–492. IFIP'92, 1992.

[Viari and Pothier, 1993] A. Viari and J. Pothier. *SmartMulti: a tool for the multiple alignment of protein sequences using flexible blocks.* Atelier de BioInformatique, 26, rue d'Ulm - 75005 Paris, 1993. in preparation.

[Viari, 1995] A. Viari. 1995. communication personnelle.

[Vingron and von Haesler, 1994] M. Vingron and A. von Haesler. Towards integration of multiple alignment and phylogenetic tree construction. presented at DIMACS Computational Molecular Biology Year - Sequence Alignment Workshop, November 1994.

[Vingron and Waterman, 1994] M. Vingron and M. S. Waterman. Sequence alignment and penalty choice. Reviews of concepts, case studies and implications. *J. Mol. Biol.*, 235:1–12, 1994.

[Vogt *et al.*, 1995] G. Vogt, T. Etzold, and P. Argos. An assessment of amino acid exchange matrices in aligning protein sequences: The twilight zone revisited. *J. Mol. Biol.*, 249:816–831, 1995.

[Wagner and Fischer, 1974] R. A. Wagner and M. J. Fischer. The string-to-string correction problem. *J. ACM*, 21:168–178, 1974.

[Watanabe, 1985] S. Watanabe. *Pattern Recognition: Human and Mechanical.* Wiley, 1985.

[Waterman and Byers, 1985] M. S. Waterman and T. H. Byers. A dynamic programming algorithm to find all solutions in a neighborhood of the optimum. *Math. Biosci.*, 77:179–188, 1985.

[Waterman and Eggert, 1987] M. S. Waterman and M. Eggert. A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons. *J. Mol. Biol.*, 197:723–728, 1987.

[Waterman and Perlwitz, 1984] M. S. Waterman and M. Perlwitz. Line geometries for sequence comparisons. *Bull. Math. Biol.*, 46:567–577, 1984.

[Waterman *et al.*, 1976] M. S. Waterman, T. F. Smith, and W. A. Beyer. Some biological sequence metrics. *Adv. Math.*, 20:367–387, 1976.

[Waterman *et al.*, 1984] M. S. Waterman, R. Arratia, and D. J. Galas. Pattern recognition in several sequences: consensus and alignment. *Bull. Math. Biol.*, 46:515–527, 1984.

[Waterman *et al.*, 1992] M. S. Waterman, Mark Eggert, and Eric Lander. Parametric sequence comparisons. *Proc. Natl. Acad. Sci. USA*, 89:6090–6093, 1992.

[Waterman, 1983] M. S. Waterman. Sequence alignments in the neighborhood of the optimum with general application to dynamic programming. *Proc. Natl. Acad. Sci. USA*, 80:3123–3124, 1983.

[Waterman, 1984a] M. S. Waterman. Efficient sequence alignment algorithms. *J. Theor. Biol.*, 108:333–337, 1984.

[Waterman, 1984b] M. S. Waterman. General methods of sequence comparison. *Bull. Math. Biol.*, 46:473–500, 1984.

[Waterman, 1986] M. S. Waterman. Multiple sequence alignments by consensus. *Nucleic Acids Res.*, 14:9095–9102, 1986.

[Waterman, 1989] M. S. Waterman. Consensus patterns in sequences. In M. S. Waterman, editor, *Mathematical Methods for DNA Sequences*, pages 93–116. CRC Press, 1989.

[Waterman, 1990] M. S. Waterman. Consensus methods for DNA and protein sequence alignment. In *Meth. Enzymol.*, volume 183, pages 221–237. Academic Press, 1990.

[Waterman, 1995] M. S. Waterman. *Introduction to Computational Biology. Maps, Sequences and Genomes.* Chapman and Hall, 1995.

[Watson and Crick, 1953] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[Watson *et al.*, 1987] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner. *Molecular Biology of the Gene.* Benjamin/Cummings, 1987.

[Weiner, 1973] P. Weiner. Linear pattern-matching algorithms. pages 1–11, Institute of Electrical Engineers, New York, USA, 1973. 14th IEEE Annual Symposium on Switching and Automata Theory.

[Wilbur and Lipman, 1983] W. J. Wilbur and D. J. Lipman. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. U.S.A.*, 80:726–730, 1983.

[Wilbur, 1985] W. J. Wilbur. On the PAM matrix model of protein evolution. *Mol. Biol. Evol.*, 2:434–447, 1985.

[Wong and Chandra, 1976] A. K. C. Wong and A. K. Chandra. Bounds for the string editing problem. *J. of the ACM*, 23:13–16, 1976.

[Wong *et al.*, 1993] A. K. C. Wong, S. C. Chan, and D. K. Y. Chiu. A multiple sequence comparison method. *Bull. Math. Biol.*, 55:465–486, 1993.

[Wu and Manber, 1992a] S. Wu and U. Manber. Agrep - a fast approximate pattern-matching tool. pages 153–162, San Francisco, CA, 1992. USENIX Technical Conference.

[Wu and Manber, 1992b] S. Wu and U. Manber. Fast text searching allowing errors. *Commun. ACM*, 35:83–91, 1992.

[Wu *et al.*, 1990] S. Wu, U. Manber, and E. W. Myers. An O(NP) sequence comparison algorithm. *Inf. Proc. Letters*, 35:317–323, 1990.

[Zuker, 1991] M. Zuker. Suboptimal sequence alignment in molecular biology. Alignment with error analysis. *J. Mol. Biol.*, 221:403–420, 1991.