

UNIVERSITAT DE BARCELONA

FACULTAT
FARMÀCIA

DEPARTAMENT
BIOQUÍMICA I BIOLOGIA MOLECULAR

LES PROPIETATS FÍSiques DE L'ADN EN ESCALA GENÒMICA

Josep Ramon Goñi Macià 2008

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT
BIOQUÍMICA I BIOLOGIA MOLECULAR

8 REFERÈNCIES BIBLIOGRÀFIQUES

- Abeel, T., Saeys, Y., Bonnet, E., Rouzé, P., & Van de Peer, Y. (2008). Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Research*, 18(2), 310-23. doi: gr.6991408.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The Genome Sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195. doi: 10.1126/science.287.5461.2185.
- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., et al. (1999). *Biología Molecular de la Célula* (3rd ed.). Barcelona: Ediciones Omega.
- Alexandersson, M., Cawley, S., & Pachter, L. (2003). SLAM: cross-species gene finding and alignment with a generalized pair hidden Markov model. *Genome Research*, 13(3), 496-502. doi: 12618381.
- Antony, S., Arimondo, P. B., Sun, J., & Pommier, Y. (2004). Position- and orientation-specific enhancement of topoisomerase I cleavage complexes by triplex DNA structures. *Nucl. Acids Res.*, 32(17), 5163-5173. doi: 10.1093/nar/gkh847.
- Aviñó, A., Cubero, E., González, C., Eritja, R., & Orozco, M. (2003). Antiparallel triple helices. Structural characteristics and stabilization by 8-amino derivatives. *Journal of the American Chemical Society*, 125(51), 16127-38. doi: 14678005.
- Ayoubi, T., & Van De Ven, W. (1996). Regulation of gene expression by alternative promoters. *FASEB J.*, 10(4), 453-460.
- Bafna, V., & Huson, D. H. (2000). The conserved exon method for gene finding. *ISMB. International Conference on Intelligent Systems for Molecular Biology*, 8, 3-12. doi: 10977061.
- Bagasra, O. (2006). A unified concept of HIV latency. *Expert Opinion on Biological Therapy*, 6(11), 1135-49. doi: 10.1517/14712598.6.11.1135.
- Bagasra, O., Stir, A. E., Pirisi-Creek, L., Creek, K. E., Bagasra, A. U., Glenn, N., et al. (2006). Role of micro-RNAs in regulation of lentiviral latency and persistence. *Applied Immunohistochemistry & Molecular Morphology: AIMM / Official Publication of the Society for Applied Immunohistochemistry*, 14(3), 276-90. doi: 00129039-200609000-00005.
- Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34(Web Server issue), W369–W373. doi: 10.1093/nar/gkl198.

- Bajic, V. B., & Seah, S. H. (2003). Dragon Gene Start Finder identifies approximate locations of the 5' ends of genes. *Nucleic Acids Res*, 31, 3560-3563. doi: 10.1093/nar/gkg570.
- Bajic, V. B., Brent, M. R., Brown, R. H., Frankish, A., Harrow, J., Ohler, U., et al. (2006). Performance assessment of promoter prediction on ENCODE regions in the EGASP experiment. *Genome Biol*, 7(Suppl I), S3-S13. doi: 10.1186/gb-2006-7-s1-s3.
- Bajic, V. B., Seah, S. H., Chong, A., Krishnan, S. P. T., Koh, J. L. Y., & Brusic, V. (2003). Computer model for recognition of functional transcription start sites in RNA polymerase II promoters of vertebrates. *J Mol Gaph Mod*, 21, 323-332. doi: 10.1016/S1093-3263(02)00179-1.
- Barre, F. X., Ait-Si-Ali, S., Giovannangeli, C., Luis, R., Robin, P., Pritchard, L. L., et al. (2000). Unambiguous demonstration of triple-helix-directed gene modification. *Proceedings of the National Academy of Sciences of the United States of America*, 97(7), 3084-8.
- Bates, A. A. D., & Maxwell, A. (2005). *DNA Topology*. Oxford University Press.
- Berg, J. M., Tymoczko, J. L., & Stryer, L. (2006). *Biochemistry* (Biochemistry (6th ed., p. 1120). W. H. Freeman.
- Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818), 363-6. doi: 11201747.
- Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A., Cheatham, T. E., Dixit, S. B., et al. (2004). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophysical Journal*, 87(6), 3799-813. doi: 15326025.
- BioMoby. Retrieved from <http://www.biomoby.org/>.
- Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., et al. (2006). 3[prime] UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Meth*, 3(3), 199-204. doi: 10.1038/nmeth854.
- Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., et al. (2004). An Overview of Ensembl. *Genome Res.*, 14(5), 925-928. doi: 10.1101/gr.1860604.
- Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome*

Res., 14(5), 988-995. doi: 10.1101/gr.1865504.

Blackburn, G. M., & Gait, M. J. (1996). *Nucleic Acids in Chemistry and Biology* (2nd ed., p. 546). Oxford University Press, USA.

Blake, R. D., & Delcourt, S. G. (1998). Thermal stability of DNA. *Nucleic Acids Research*, 26(14), 3323-32. doi: 9649614.

Blake, R., Bizzaro, J., Blake, J., Day, G., Delcourt, S., Knowles, J., et al. (1999). Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics*, 15(5), 370-375. doi: 10.1093/bioinformatics/15.5.370.

Blanco, E., Farré, D., Albà, M. M., Messeguer, X., & Guigó, R. (2006). ABS: a database of Annotated regulatory Binding Sites from orthologous promoters. *Nucleic Acids Research*, 34(Database issue), D63–D67. doi: 10.1093/nar/gkj116.

Boeckmann, B., Blatter, M., Famiglietti, L., Hinz, U., Lane, L., Roechert, B., et al. (2005). Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *Comptes Rendus Biologies*, 328(10-11), 882-99. doi: S1631-0691(05)00110-1.

Bolshoy, A., McNamara, P., Harrington, R. E., & Trifonov, E. N. (1991). Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), 2312-6. doi: 2006170.

Bonen, L., & Vogel, J. (2001). The ins and outs of group II introns. *Trends in Genetics*, 17(6), 322-331. doi: 10.1016/S0168-9525(01)02324-1.

Brent, M. R., & Guigó, R. (2004). Recent advances in gene structure prediction. *Current Opinion in Structural Biology*, 14(3), 264-272. doi: 10.1016/j.sbi.2004.05.007.

Breslauer, K. J., Frank, R., Blöcker, H., & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 83(11), 3746-50. doi: 3459152.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187-217. doi: 10.1002/jcc.540040211.

Brukner, I., Sánchez, R., Suck, D., & Pongor, S. (1995). Trinucleotide models for DNA bending propensity: comparison of models based on DNaseI digestion and nucleosome packaging data. *Journal of Biomolecular Structure & Dynamics*,

13(2), 309-17. doi: 8579790.

- Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol*, 212, 563-578. doi: 10.1016/0022-2836(90)90223-9.
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1), 78-94. doi: 9149143.
- Burset, M., & Guigó, R. (1996). Evaluation of gene structure prediction programs. *Genomics*, 34(3), 353-67. doi: 8786136.
- Butler, J. E. F., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development*, 16(20), 2583-92. doi: 12381658.
- Carbone, G. M., Napoli, S., Valentini, A., Cavalli, F., Watson, D. K., & Catapano, C. V. (2004). Triplex DNA-mediated downregulation of Ets2 expression results in growth inhibition and apoptosis in human prostate cancer cells. *Nucl. Acids Res.*, 32(14), 4358-4367. doi: 10.1093/nar/gkh744.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). The transcriptional landscape of the mammalian genome. *Science*, 309, 1559-1563. doi: 10.1126/science.1112014.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 38, 626-635. doi: 10.1038/ng1789.
- Chastain, M., & Tinoco, I. (1991). Structural elements in RNA. *Progress in Nucleic Acid Research and Molecular Biology*, 41, 131-77. doi: 1715587.
- Chi, K. R. (2008). The year of sequencing. *Nat Meth*, 5(1), 11-14. doi: 10.1038/nmeth1154.
- Churchill, F. B. (1974). William Johannsen and the genotype concept. *Journal of the History of Biology*, 7(1), 5-30. doi: 10.1007/BF00179291.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Third Edition (Third., p. 736). Lawrence Erlbaum.
- Coma, S., Noé, V., Eritja, R., & Ciudad, C. J. (2005). Strand displacement of double-stranded DNA by triplex-forming antiparallel purine-hairpins. *Oligonucleotides*, 15(4), 269-83. doi: 10.1089/oli.2005.15.269.

- Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, (32 Database), D258-61. doi: 10.1093/nar/gkh036.
- Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921. doi: 10.1038/35057062.
- Consortium, T. E. P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 306, 636-640. doi: 10.1126/science.1105136.
- Crick, F. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138-63.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561-3.
- Dauxois, T., Peyrard, M., & Bishop, A. R. (1993). Entropy-driven DNA denaturation. *Physical Review E*, 47(1), R44. doi: 10.1103/PhysRevE.47.R44.
- Davuluri, R. V., Grosse, I., & Zhang, M. Q. (2001). Computational identification of promoters and first exons in the human genome. *Nat Genet*, 29, 412-417. doi: 10.1038/ng780.
- Dennis, J. H., Fan, H., Reynolds, S. M., Yuan, G., Meldrim, J. C., Richter, D. J., et al. (2007). Independent and complementary methods for large-scale structural analysis of mammalian chromatin. *Genome Research*, 17(6), 928-39. doi: 17/6/928.
- Dickerson, R. (1989). Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Research*, 17(5).
- Down, T. A., & Hubbard, T. J. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res*, 12, 458-461. doi: 10.1101/gr.216102.
- Duca, M., Vekhoff, P., Oussedik, K., Halby, L., & Arimondo, P. B. (2008). The triple helix: 50 years later, the outcome. *Nucleic Acids Research*. doi: gkn493.
- Dutta, S., Burkhardt, K., Swaminathan, G. J., Kosada, T., Henrick, K., Nakamura, H., et al. (2008). Data deposition and annotation at the worldwide protein data bank. *Methods in Molecular Biology* (Clifton, N.J.), 426, 81-101. doi: 10.1007/978-1-60327-058-8_5.
- el Hassan, M. A., & Calladine, C. R. (1996). Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *Journal of Molecular Biology*, 259(1), 95-103. doi: 8648652.

- Fagard, M., Boutet, S., Morel, J. B., Bellini, C., & Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21), 11650-4. doi: 11016954.
- Ferrer-Costa, C., Gelpí, J. L., Zamakola, L., Parraga, I., de la Cruz, X., & Orozco, M. (2005). PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics (Oxford, England)*, 21(14), 3176-8. doi: bti486.
- Ferrer-Costa, C., Orozco, M., & de la Cruz, X. (2002). Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *Journal of Molecular Biology*, 315(4), 771-86. doi: 11812146.
- Ferrer-Costa, C., Orozco, M., & de la Cruz, X. (2005). Use of bioinformatics tools for the annotation of disease-associated mutations in animal models. *Proteins*, 61(4), 878-87. doi: 10.1002/prot.20664.
- Florquin, K., Saeys, Y., Degroeve, S., Rouze, P., & Van de Peer, Y. (2005). Large-scale structural analysis of the core promoter in mammalian and plant genomes. *Nucleic Acid Res*, 33, 4235-4264. doi: 10.1093/nar/gki742.
- Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol*, 196, 261-282. doi: 10.1016/0022-2836(87)90689-9.
- Gelfand, M. S., Mironov, A. A., & Pevzner, P. A. (1996). Gene recognition via spliced sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*, 93(17), 9061-6. doi: 8799154.
- Gelpí, J. L., Kalko, S. G., Barril, X., Cirera, J., de La Cruz, X., Luque, F. J., et al. (2001). Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins*, 45(4), 428-37. doi: 11746690.
- Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbelt, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome research*, 17(6), 669-81.
- Gidden, J., Baker, E. S., Ferzoco, A., & Bowers, M. T. (2005). Structural motifs of DNA complexes in the gas phase. *International Journal of Mass Spectrometry*, 240(3), 183-193. doi: 10.1016/j.ijms.2004.09.011.
- Goñi, J. R., Fenollosa, C., Pérez, A., Torrents, D., & Orozco, M. (2008). DNALive: A tool for the physical analysis of DNA at the genomic scale. *Bioinformatics (Oxford, England)*. doi: btn259.

- Goñi, J. R., Pérez, A., Torrents, D., & Orozco, M. (2007). Determining promoter location based on DNA structure first-principles calculations. *Genome biology*, 8(12), R263.
- Goñi, J. R., Vaquerizas, J. M., Dopazo, J., & Orozco, M. (2006). Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC genomics*, 7, 63.
- Goodsell, D. S., & Dickerson, R. E. (1994). Bending and curvature calculations in B-DNA. *Nucleic Acids Research*, 22(24), 5497–5503.
- Gorin, A. A., Zhurkin, V. B., & Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. *Journal of Molecular Biology*, 247(1), 34-48. doi: 7897660.
- Gotoh, O., & Tagashira, Y. (1981). Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers*, 20(5), 1033-1042. doi: 10.1002/bip.1981.360200513.
- Gowers, D., & Fox, K. (1999). Towards mixed sequence recognition by triple helix formation. *Nucl. Acids Res.*, 27(7), 1569-1577. doi: 10.1093/nar/27.7.1569.
- Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, 47(15), 965-978. doi: 10.1016/j.infsof.2005.09.005.
- Gross, S. S., & Brent, M. R. (2006). Using multiple alignments to improve gene prediction. *J Comput Biol*, 13, 379-393. doi: 10.1089/cmb.2006.13.379.
- Guigó, R. (1998). Assembling genes from predicted exons in linear time with dynamic programming. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 5(4), 681-702. doi: 10072084.
- Gupta, S., Dennis, J., Thurman, R. E., Kingston, R., Stamatoyannopoulos, J. A., & Noble, W. S. (2008). Predicting human nucleosome occupancy from primary sequence. *PLoS Computational Biology*, 4(8), e1000134. doi: 10.1371/journal.pcbi.1000134.
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucl. Acids Res.*, 33(suppl_1), D514-517. doi: 10.1093/nar/gki033.
- Hewett, P. W., Daft, E. L., Laughton, C. A., Ahmad, S., Ahmed, A., & Murray, J. C. (2006). Selective Inhibition of the Human tie-1 Promoter with Triplex-Forming Oligonucleotides Targeted to Ets Binding Sites. *Molecular Medicine*, 12(1-3),

8–16. doi: 10.2119/2005-00046.Hewett.

Hide, W. (2007). Faculty of 1000 Biology: evaluations for Goñi JR et al BMC Genomics 2006 7 :63 . Faculty of 1000. Retrieved September 6, 2008, from <http://www.f1000biology.com/article/id/1068779/evaluation>.

Ho, P. S., Zhou, G. W., & Clark, L. B. (1990). Polarized electronic spectra of Z-DNA single crystals. *Biopolymers*, 30(1-2), 151-63. doi: 2224047.

Hobza, P., & Sponer, J. (1999). Structure, energetics, and dynamics of the nucleic Acid base pairs: nonempirical ab initio calculations. *Chemical Reviews*, 99(11), 3247-76. doi: 11749516.

Human Genome Sequencing ConsortiumInternational. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945. doi: 10.1038/nature03001.

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, 33(9), 2908-16. doi: 33/9/2908.

Ioshikhes, I., & Trifonov, E. N. (1993). Nucleosomal DNA sequence database. *Nucleic Acids Research*, 21(21), 4857-9. doi: 8177730.

Ito, T. (2007). Role of Histone Modification in Chromatin Dynamics. *J Biochem*, 141(5), 609-614. doi: 10.1093/jb/mvm091.

Ivanov, V. I., & Minchenkova, L. E. [The A-form of DNA: in search of the biological role]. *Molekuliarnaia Biologiia*, 28(6), 1258-71. doi: 7885327.

Jacques, P., Rodrigue, S., Gaudreau, L., Goulet, J., & Brzezinski, R. (2006). Detection of prokaryotic promoters from the genomic distribution of hexanucleotide pairs. *BMC Bioinformatics*, 7(1), 423. doi: 10.1186/1471-2105-7-423.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, 33, 245-254.

Jaumot, J., Aviña, A., Eritja, R., Tauler, R., & Gargallo, R. (2003). Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution. *Journal of biomolecular structure & dynamics*, 21(2), 267-78.

Kanhere, A., & Bansal, M. (2005). A novel method for prokaryotic promoter prediction based on DNA stability. *BMC Bioinformatics*, 6, 1. doi: 1471-2105-6-1.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler,

- D., et al. (2004). The UCSC Table Browser data retrieval tool. *Nucl. Acids Res.*, 32(suppl_1), D493-496. doi: 10.1093/nar/gkh103.
- Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., et al. (2004). EnsMart: A Generic System for Fast and Flexible Access to Biological Data. *Genome Res.*, 14(1), 160-169. doi: 10.1101/gr.1645104.
- Kent, W. J. (2002). BLAT---The BLAST-Like Alignment Tool. *Genome Res.*, 12(4), 656-664. doi: 10.1101/gr.229202.
- Kim, T. H., Barrera, L. O., Qu, C., Van Calcar, S., Trinklein, N. D., Cooper, S. J., et al. (2005). Direct isolation and identification of promoters in the human genome. *Genome Res*, 15, 830-839. doi: 10.1101/gr.3430605.
- Knauert, M. P., & Glazer, P. M. (2001). Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Human MolecularGenetics*, 10, 2243.
- Knudsen, S. (1999). Promoter2.0: for the recognition of PolIII promoter sequences. *Bioinformatics*, 15, 356-361. doi: 10.1093/bioinformatics/15.5.356.
- Korf, I., Flicek, P., Duan, D., & Brent, M. R. (2001). Integrating genomic homology into gene structure prediction. *Bioinformatics*, 17, S140-148.
- Lander, E. S. (1999). Array of hope. *Nature Genetics*, 21(1 Suppl), 3-4. doi: 9915492.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.
- Lankas, F., Sponer, J., Langowski, J., & Cheatham, T. E. (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophysical Journal*, 85(5), 2872-83. doi: 14581192.
- Larsen, B. S., & McEwen, C. N. (1998). *Mass Spectrometry of Biological Materials*, Second Edition (2nd ed., p. 488). CRC.
- Larsen, F., Gundersen, G., Lopez, R., & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, 13, 1095-1107. doi: 10.1016/0888-7543(92)90024-M.
- Latchman, D. S. (1997). Transcription factors: an overview. *The international journal of biochemistry & cell biology*, 29(12), 1305-12.
- Laughlan, G., Murchie, A., Norman, D., Moore, M., Moody, P., Lilley, D., et al. (1994). The high-resolution crystal structure of a parallel-stranded guanine

- tetraplex. *Science*, 265(5171), 520-524. doi: 10.1126/science.8036494.
- Levitsky, V. G., Podkolodnaya, O. A., Kolchanov, N. A., & Podkolodny, N. L. (2001). Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis. *Bioinformatics (Oxford, England)*, 17(11), 998-1010. doi: 11724728.
- Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. P., et al. (2003). *Molecular Cell Biology* (5th ed., p. 973). W. H. Freeman.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8[thinsp]Å resolution. *Nature*, 389(6648), 251-260. doi: 10.1038/38444.
- Mahalanobis, P. (1936). On the generalised distance in statistics. . *Proceedings of the National Institute of Science of India* , 12 , 49–55.
- Majumdar, A., Khorlin, A., Dyatkina, N., Lin, F. L., Powell, J., Liu, J., et al. (1998). Targeted gene knockout mediated by triple helix forming oligonucleotides. *Nat Genet*, 20, 212. doi: 10.1038/2530.
- McCarthy, B. J., & Holland, J. J. (1965). Denatured DNA as a direct template for in vitro protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 54(3), 880-6.
- Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., & Albà, M. M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics (Oxford, England)*, 18(2), 333-4. doi: 11847087.
- Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), 335-341.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-1092.
- Miele, V., Vaillant, C., d'Aubenton-Carafa, Y., Thermes, C., & Grange, T. (2008). DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucl. Acids Res.*, 36(11), 3746-3756. doi: 10.1093/nar/gkn262.
- Mito, Y., Henikoff, J. G., & Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genetics*, 37(10), 1090-7. doi: ng1637.
- Mito, Y., Henikoff, J. G., & Henikoff, S. (2007). Histone Replacement Marks the Boundaries of cis-Regulatory Domains. *Science*, 315(5817), 1408-1411. doi:

10.1126/science.1134004.

Mowserver. <http://www.inab.org/MOWServ/modules/help/Main/index.html>.

Nikolov, D. B., & Burley, S. K. (1997). RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences of the United States of America*, 94(1), 15-22.

Noy, A. (2008). *La flexibilitat en els àcids nucleics: un estudi de dinàmica molecular*. Universitat de Barcelona. Retrieved July 29, 2008, from <http://www.ncbi.nlm.nih.gov/sites/entrez/>.

O'Geen, H., Squazzo, S. L., Iyengar, S., Blahnik, K., Rinn, J. L., Chang, H. Y., et al. (2007). Genome-Wide Analysis of KAP1 Binding Suggests Autoregulation of KRAB-ZNFs. *PLoS Genetics*, 3(6), e89. doi: 10.1371/journal.pgen.0030089.

Ohler, U., Liao, G. C., Niemann, H., & Rubin, G. M. (2002). Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol*, 3, RESEARCH0087. doi: 10.1186/gb-2002-3-12-research0087.

Ohler, U., Nierman, H., Liao, G. C., & Rubin, G. M. (2001). Joint modeling of DNA sequence and physical properties to improve eukaryotic promoter recognition. *Bioinformatics*, 17, S199-206.

Ohno, M., Fukagawa, T., Lee, J. S., & Ikemura, T. (2002). Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma*, 111(3), 201-13. doi: 12355210.

Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(19), 11163-8. doi: 9736707.

Ornstein, R. L., Rein, R., Breen, D., & Macleroy, R. (1978). An optimized potential function for the calculation of nucleic acid interaction energies. I. Base stacking. *Biopolymers*, 17, 2341-2360.

Orozco, M., Rueda, M., Blas, J., Cubero, E., Luque, F., & Laughton, C. (2004). Molecular dynamics simulations of nucleic acids. In *Encyclopedia of Computational Chemistry*. Retrieved July 29, 2008, from <http://mmb.pcb.ub.es/~manu/papers.html>.

Ozsolak, F., Song, J. S., Liu, X. S., & Fisher, D. E. (2007). High-throughput mapping of the chromatin structure of human promoters. *Nature Biotechnology*, 25(2),

244-8. doi: nbt1279.

- Parra, G., Blanco, E., & Guigo, R. (2000). GeneID in *Drosophila*. *Genome Res.*, 10(4), 511-515. doi: 10.1101/gr.10.4.511.
- Pearson, H. (2006). Genetics: what is a gene? *Nature*, 441(7092), 398-401.
- Peckham, H. E., Thurman, R. E., Fu, Y., Stamatoyannopoulos, J. A., Noble, W. S., Struhl, K., et al. (2007). Nucleosome positioning signals in genomic DNA. *Genome Research*, 17(8), 1170-7. doi: gr.6101007.
- Pedersen, A. G., Baldi, P., Chauvin, Y., & Brunak, S. (1998). DNA structure in human RNA polymerase II promoters. *J Mol Biol*, 281, 663-673. doi: 10.1006/jmbi.1998.1972.
- Pedersen, A. G., Baldi, P., Chauvin, Y., & Brunak, S. (1999). The biology of eukaryotic promoter prediction - a review. *Comput Chem*, 23, 191-207. doi: 10.1016/S0097-8485(99)00015-7.
- Pedersen, A. G., Jensen, L. J., Brunak, S., Staefeldt, H. H., & Ussery, D. W. (2000). A DNA structural atlas for *Escherichia coli*. *J Mol Biol*, 299, 907-930. doi: 10.1006/jmbi.2000.3787.
- Pérez, A., Luque, F. J., & Orozco, M. (2007). Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society*, 129(47), 14739-45. doi: 10.1021/ja0753546.
- Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., et al. (2007). Refinement of the AMBER force field for nucleic acids. Improving the description of alpha/gamma conformers. *Biophys J*, 92, 3817-3829. doi: 10.1529/biophysj.106.097782.
- Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., et al. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophysical Journal*, 92(11), 3817-29. doi: biophysj.106.097782.
- Plum, G. E., & Breslauer, K. J. (1995). Thermodynamics of an intramolecular DNA triple helix: a calorimetric and spectroscopic study of the pH and salt dependence of thermally induced structural transitions. *Journal of Molecular Biology*, 248(3), 679-95. doi: 7752233.
- Ponger, L., & Mouchiroud, D. (2002). CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics*, 18, 631-633. doi: 10.1093/bioinformatics/18.4.631.

- Ponomarenko, J. V., Ponomarenko, M. P., Frolov, A. S., Vorobyev, D. G., Overton, G. C., & Kolchanov, N. A. (1999). Conformational and physicochemical DNA features specific for transcription factor binding sites. *Bioinformatics*, 15, 654-668. doi: 10.1093/bioinformatics/15.7.654.
- Prescott, L. M., Harley, J. P., & Klein, D. A. (2004). *Microbiology* (6th ed., p. 992). McGraw-Hill Science/Engineering/Math.
- Prody, G., Bakos, J., Buzayan, J. M., Schneider, I., & Bruening, G. (1986). Autolytic Processing of Dimeric Plant Virus Satellite RNA. *Science*, 231(4745), 1577-1580. doi: 10.1126/science.231.4745.1577.
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33(Database issue), D501-4. doi: 10.1093/nar/gki097.
- Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem*, 26, 51-56. doi: 10.1016/S0097-8485(01)00099-7.
- Reese, M. G., & Guigo, R. (2006). EGASP: Introduction. *Genome Biol*, 7(Suppl I), S1-3. doi: 10.1186/gb-2006-7-s1-s1.
- Reese, M. G., Kulp, D., Tammana, H., & Haussler, D. (2000). Genie---Gene Finding in *Drosophila melanogaster*. *Genome Res.*, 10(4), 529-538. doi: 10.1101/gr.10.4.529.
- Roberts, R. W., & Crothers, D. M. (1996). Prediction of the stability of DNA triplexes. *Proceedings of the National Academy of Sciences of the United States of America*, 93(9), 4320-5. doi: 10.1073/pnas.93.9.4320.
- Rueda, M. (2005). Estudio teórico sobre la influencia del solvente en la estructura y dinámica del ADN. Univeritat de Barcelona.
- Rueda, M., Ferrer-Costa, C., Meyer, T., Pérez, A., Camps, J., Hospital, A., et al. (2007). A consensus view of protein dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), 796-801. doi: 10.1073/pnas.0605534104.
- Rueda, M., Kalko, S. G., Luque, F. J., & Orozco, M. (2003). The structure and dynamics of DNA in the gas phase. *Journal of the American Chemical Society*, 125(26), 8007-14. doi: 10.1021/ja03023a023.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag.

- Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research*, 32(Database issue), D91-4. doi: 14681366.
- Santalucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA*, 95, 1460. doi: 10.1073/pnas.95.4.1460.
- Satchwell, S. C., Drew, H. R., & Travers, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. *Journal of Molecular Biology*, 191(4), 659-75. doi: 3806678.
- Schones, D. E., Cui, K., Cuddapah, S., Roh, T., Barski, A., Wang, Z., et al. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, Vol 132, 887-898.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., et al. (2006). A genomic code for nucleosome positioning. *Nature*, 442(7104), 772-778. doi: 10.1038/nature04979.
- Sequencing Consortium, M. G. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), 520-562. doi: 10.1038/nature01262.
- Shchyolkina, A., Borisova, O., Livshits, M., Pozmogova, G., Chernov, B., Klement, R., et al. (2000). Parallel-Stranded DNA with Mixed AT/GC Composition: Role of trans G•C Base Pairs in Sequence Dependent Helical Stability. *Biochemistry*, 39(33), 10034-10044.
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29(1), 308-11.
- Shields, G. A., Laughton, C. A., & Orozco, M. (1997). Molecular dynamics simulations of the d(T * A * T) triple helix. *J Am Chem Soc*, 119, 7463. doi: 10.1021/ja970601z.
- Shpigelman, E. S., Trifonov, E. N., & Bolshoy, A. (1993). CURVATURE: software for the analysis of curved DNA. *Comput Appl Biosci*, 9, 435.
- Singhal, P., Jayaram, B., Dixit, S. B., & Beveridge, D. L. (2008). Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. *Biophysical Journal*, 94(11), 4173-83. doi: biophysj.107.116392.
- Sivolob, A. V., & Khrapunov, S. N. (1995a). Translational positioning of nucleosomes on DNA: the role of sequence-dependent isotropic DNA bending stiffness.

Journal of Molecular Biology, 247(5), 918-31. doi: 7723041.

- Smale, S. T., & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual review of biochemistry*, 72, 449-79.
- Smith, G. R. (1981). DNA supercoiling: another level for regulating gene expression. *Cell*, 24(3), 599-600. doi: 6265097.
- Soliva, R., Laughton, C., Luque, F., & Orozco, M. (1998). Molecular Dynamics Simulations in Aqueous Solution of Triple Helices Containing d(G•C•C) Trios. *Journal of the American Chemical Society*, 120(44), 11226-11233.
- Solovyev, V., & Salamov, A. (1997). The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proc Int Conf Intell Syst Mol Biol*, 5, 294-302.
- Solovyev, V., & Shahmuradov, I. (2003). PromH: promoters identification using orthologous genomic sequences. *Nucl. Acids Res.*, 31(13), 3540-3545. doi: 10.1093/nar/gkg525.
- Spacková, N., Cubero, E., Sponer, J., & Orozco, M. (2004). Theoretical study of the guanine --> 6-thioguanine substitution in duplexes, triplexes, and tetraplexes. *Journal of the American Chemical Society*, 126(44), 14642-50. doi: 10.1021/ja0468628.
- Spanish National Institute of Bioinformatics. <http://www.inab.org/>.
- Sponer, J., Gabb, H. A., Leszczynski, J., & Hobza, P. (1997). Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys J*, 73, 76-87.
- Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. *Annual Review of Biochemistry*, 58, 1051-77. doi: 2673007.
- Sugimoto, N., Nakano, S., Yoneyama, M., & Honda, K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Research*, 24(22), 4501-5. doi: 8948641.
- Sun, J. S., Garestier, T., & Hélène, C. (1996). Oligonucleotide directed triple helix formation. *Current Opinion in Structural Biology*, 6(3), 327-33. doi: 8804836.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., et al. (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res*, 11, 677-684. doi: 10.1101/gr.1640R.

- Svinarchuk, F., Nagibneva, I., Cherny, D., Ait-Si-Ali, S., Pritchard, L. L., Robin, P., et al. (1997). Recruitment of transcription factors to the target site by triplex-forming oligonucleotides. *Nucleic Acids Research*, 25(17), 3459-64. doi: 9254704.
- Talavera, D., Vogel, C., Orozco, M., Teichmann, S. A., & de la Cruz, X. (2007). The (in)dependence of alternative splicing and gene duplication. *PLoS Computational Biology*, 3(3), e33. doi: 06-PLCB-RA-0259R3.
- The C. elegans Sequencing Consortium. (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396), 2012-2018. doi: 10.1126/science.282.5396.2012.
- The Gene Ontology Consortium. (2008). The Gene Ontology project in 2008. *Nucl. Acids Res.*, 36(suppl_1), D440-444. doi: 10.1093/nar/gkm883.
- The Genome Sequence of Drosophila melanogaster . (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814), 796-815.
- The HAVANA Team. <http://www.sanger.ac.uk/HGP/havana/>.
- van Erp, T. S., Cuesta-Lopez, S., Hagmann, J., & Peyrard, M. (2005). Can one predict DNA transcription start sites by studying bubbles? *Physical Review Letters*, 95(21), 218104. doi: 16384189.
- Vasquez, K. M., & Wilson, J. H. (1998). Triplex-directed modification of genes and gene activity. *Trends in biochemical sciences*, 23(1), 4-9. doi: 9478127.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), 1304-51.
- Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S., & Nakai, K. (2008). DBTSS: database of transcription start sites, progress report 2008. *Nucl. Acids Res.*, 36(suppl_1), D97-101. doi: 10.1093/nar/gkm901.
- Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losick, R. (2003). *Molecular Biology of the Gene, Fifth Edition (5th ed., p. 830)*. Benjamin Cummings.
- Watson, J., & Crick, F. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171, 737-738.
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1985). An all atom force

field for simulations of proteins and nucleic acids. , 7(2), 230-252.

- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, 7(2), 230-252. doi: 10.1002/jcc.540070216.
- Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvermin, V., et al. (2008). Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.*, 36(suppl_1), D13-21. doi: 10.1093/nar/gkm1000.
- Wiehe, T., Gebauer-Jung, S., Mitchell-Olds, T., & Guigó, R. (2001). SGP-1: prediction and validation of homologous genes based on sequence alignments. *Genome Research*, 11(9), 1574-83. doi: 11544202.
- Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59(5), 871-80. doi: 2590943.
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., et al. (2001). The TRANSFAC system on gene expression regulation. *Nucleic Acids Research*, 29(1), 281-3. doi: 11125113.
- Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, 34(Web Server issue), W6–W9. doi: 10.1093/nar/gkl164.
- Yuan, G., Liu, Y., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., et al. (2005). Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae*. *Science*, 309(5734), 626-630. doi: 10.1126/science.1112178.
- Zamecnik, P. C., & Stephenson, M. L. (1978). Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences of the United States of America*, 75(1).
- Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. *Human molecular genetics*, 7(5), 919-32.
- Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nat Rev Genet*, 3, 698-709. doi: 10.1038/nrg890.