生物物理学报 2011, 27(1) 28-37 DOI: ISSN: 1000-6737 CN: 11-1992/Q

综述

基于模板的蛋白质结构预测

黄俊峰, 段鹏, 吴文言

中山大学生命科学学院有害生物防治与资源利用国家重点实验室，广州 510275

摘要：

基于模板的蛋白结构预测和不依赖模板的蛋白结构预测是计算预测蛋白质三维结构的两种方法，前者由于具有快速和较高准确性的优点，而得到了广泛的应用。基于模板的结构预测是通过寻找与目标蛋白序列相似并且有实验测定的结构作为模板，进而构建目标序列的结构模型的方法。文章详细综述了基于模板的结构预测方法的步骤、关键环节，并对影响结构预测精确性的因素进行了分析和讨论。

关键词： 基于模板的蛋白结构预测 序列比对 结构精修 结构模型筛选

Template-Based Protein Structure Prediction

HUANG Junfeng, DUAN Peng, WU Wenyan

State Key Laboratory of Biocontrol, School of Life Science, Sun Yat-Sen University, Guangzhou 510275, China

Abstract:

Methods for predicting three-dimensional structure of protein molecules can be classified into two categories: template-based modeling and template free modeling. Template-based prediction is widely used in application due to its speed and relatively high accuracy. This technique predicts the three-dimensional structure of a given protein sequence based primarily on its alignment to one or more proteins of known structure. In this paper, the progress in the methodology for template-based modeling is reviewed. The critical steps and factors that influence the prediction accuracy are analysed and discussed.

Keywords: Template-based modeling Sequence alignment Model refinement Model selection

收稿日期 2010-06-07 **修回日期** 2010-08-10 **网络版发布日期**

DOI:

基金项目:

国家自然科学基金(30772722)、广州市科技计划(2009JI-C541)资助项目

通讯作者: 吴文言，电话：(020)84115570，E-mail：lsswwy@mail.sysu.edu.cn

作者简介:

作者Email: lsswwy@mail.sysu.edu.cn

参考文献：

1. Klug A. The Human Genome Project. IUBMB Life, 2001, 51(1): 1~4
2. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol, 2008, 26(10): 1135~1145
3. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL, 2nd, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC and others. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods, 2009, 6(8): 606~612
4. Anfinsen. Principles that govern the folding of protein chains. Science, 1973, 181: 223~230
5. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA, 1987, 84(21): 7524~7528

6. Wolynes PG. Energy landscapes and solved protein-folding problems. Philos Transact A Math Phys Eng Sci, 2005, 363(1827): 453~464; discussion 464~467
7. Feng HQ, Zhou Z, Bai YW. A protein folding pathway with multiple folding intermediates at atomic resolution. Proc Natl Acad Sci USA, 2005, 102(14): 5026~5031
8. Xiang ZX. Advances in homology protein structure modeling. Curr Protein Pept Sc, 2006, 7(3): 217~227
9. Jones D, Thornton J. Protein fold recognition. J Comput Aided Mol Des, 1993, 7(4): 439~456
10. Osguthorpe DJ. Ab initio protein folding. Curr Opin Struct Biol, 2000, 10(2): 146~152
11. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins, 1995, 23(3): ii~v
12. Ginalski K, Grishin NV, Godzik A, Rychlewski L. Practical lessons from protein structure prediction. Nucleic Acids Res, 2005, 33(6): 1874~1891
13. Du PC, Andrec M, Levy RM. Have we seen all structures corresponding to short protein fragments in the Protein Data Bank? An update. Protein Engin, 2003, 16(6):407~414
14. Chothia C. Proteins. One thousand families for the molecular biologist. Nature, 1992, 357(6379): 543~544
15. Lee J, Liwo A, Ripoll DR, Pillardy J, Scheraga HA. Calculation of protein conformation by global optimization of a potential energy function. Proteins, 1999, Suppl 3:204~208
16. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA, 1988, 85(8): 2444~2448
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol, 1990, 215(3):403~410
18. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res, 1997, 25(17): 3389~3402
19. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics, 1998, 14(10): 846~856
20. Zhou H, Zhou Y. SPARKS 2 and SP3 servers in CASP6. Proteins, 2005, 61(Suppl 7): 152~156
21. Wallner B, Elofsson A. All are not equal: a benchmark of different homology modeling programs. Protein Sci, 2005, 14(5): 1315~1327
22. Boniecki M, Rotkiewicz P, Skolnick J, Kolinski A. Protein fragment reconstruction using various modeling techniques. J Comput Aid Mol Des, 2003, 17(11): 725~738
23. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res,2004, 32: 526~531
24. Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. Proteins, 2007, 69 (Suppl 8): 108-117
25. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics, 2006, Chapter 5: Unit 5~6
26. Bruccoleri RE, Karplus M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers, 1987, 26(1): 137~168
27. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. Proteins, 2004, 55(2): 351~367
28. Lee DS, Seok C, Lee J. Protein loop modeling using fragment assembly. J Korean Phys Soc, 2008, 52(4):1137~1142
29. Xiang ZX, Soto CS, Honig B. Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA, 2002,99(11): 7432~7437
30. Felts AK, Gallicchio E, Chekmarev D, Paris KA, Friesner RA, Levy RM. Prediction of protein loop conformations using the AGBNP implicit solvent model and torsion angle sampling. J Chem Theory Comput, 2008, 4(5): 855~868
31. Bowers KE, Fierke CA. Positively charged side chains in protein farnesyltransferase enhance catalysis by stabilizing the formation of the diphosphate leaving group. Biochemistry, 2004, 43(18): 5256~5265
32. Jackson RM. Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. Protein Sci, 1999, 8(3): 603~613
33. SmockRG, Gierasch LM. Sending signals dynamically. Science, 2009, 324(5924): 198~203
34. Schluter KD. PTH and PTHrP: Similar Structures but Different Functions. News Physiol Sci, 1999, 14: 243~249
35. Orengo CA, Pearl FM, Bray JE, Todd AE, Martin AC, Lo Conte L, Thornton JM. The CATH Database provides insights into protein structure/function relationships. Nucleic Acids Res, 1999, 27(1): 275~279
36. Bower MJ, Cohen FE, Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. J Mol Biol,1997, 267(5): 1268~1282
37. Canutescu AA, Shelenkov AA, Dunbrack RL Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci, 2003, 12(9): 2001~2014
38. Lazaridis T, Karplus M. Effective energy functions for protein structure prediction. Curr Opin Struct

Biol, 2000, 10(2): 139~145
39. Tsai J, Bonneau R, Morozov AV, Kuhlman B, Rohl CA, Baker D. An improved protein decoy set for testing energy functions for protein structure prediction. Proteins, 2003, 53(1): 76~87
40. Misura KM, Baker D. Progress and challenges in high-resolution refinement of protein structure models. Proteins, 2005, 59(1): 15~29
41. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR, 1996, 8(4): 477~486
42. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol, 1997, 277: 396~404
43. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci, 2006, 15(11):2507~2524
44. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using rosetta. Methods Enzymol, 2004,383: 66~93
45. Boas FE, Harbury PB. Potential energy functions for protein design. Curr Opin Struct Biol, 2007, 17 (2): 199~204
46. Mucherino A, Costantini S, di Serafino D, D'Apuzzo M, Facchiano A, Colonna G. Understanding the role of the topology in protein folding by computational inverse folding experiments. Comput Biol Chem, 2008, 32(4): 233~239
47. Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J Mol Biol, 1992, 228(1): 13~22
48. Lobanov MY, Bogatyreva NS, Galzitskaya OV. Radius of gyration as an indicator of protein structure compactness. Molecular Biology, 2008, 42(4): 623~628
49. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol, 1998, 277(4): 985~994
50. Pons JL, Labesse G. @TOME-2: A new pipeline for comparative modeling of protein-ligand complexes. Nucleic Acids Res, 2009; 37(Web Server issue): W485~91
51. Thielmann Y, Weiergraber OH, Ma P, Schwarten M, Mohrluder J, Willbold D. Comparative modeling of human NSF reveals a possible binding mode of GABARAP and GATE-16. Proteins, 2009, 77(3): 637~646
52. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D. High-resolution structure prediction and the crystallographic phase problem. Nature, 2007, 450(7167):259~264
53. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei JM, Kim D, Kellogg E, DiMaio F, Lange O and others. Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins-Struct Funct Bioinform, 2009, 77:89~99
54. Zhang Y. Protein structure prediction: When is it useful? Curr Opin Struc Biol, 2009, 19(2): 145~155
55. Ben-David M, Noivirt-Brik O, Paz A, Prilusky J, Sussman JL, Levy Y. Assessment of CASP8 structure predictions for template free targets. Proteins-Struct Funct Bioinform, 2009,77: 50~65
56. Zhang Y. I-TASSER: Fully automated protein structure prediction in CASP8. Proteins-Struct Funct Bioinform, 2009,77: 100~113
57. Alexander PA, He Y, Chen Y, Orban J, Bryan PN. The design and characterization of two proteins with 88% sequence identity but different structure and function. Proc Natl Acad Sci USA, 2007, 104(29): 11963~11968
58. He Y, Chen Y, Alexander P, Bryan PN, Orban J. NMR structures of two designed proteins with high sequence identity but different fold and function. Proc Natl Acad Sci USA, 2008, 105(38): 144112~144127
59. Roessler CG, Hall BM, Anderson WJ, Ingram WM, Roberts SA, Montfort WR, Cordes MH. Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity butdifferent folds. Proc Natl Acad Sci USA, 2008, 105(7): 2343~2348
60. Davidson AR. A folding space odyssey. Proc Natl Acad Sci USA, 2008, 105(8): 2759~2760
61. Takano K, Katagiri Y, Mukaiyama A, Chon H, Matsumura H, Koga Y, Kanaya S. Conformational contagion in a protein: Structural properties of a chameleon sequence. Proteins, 2007, 68(3): 617~625
62. Taylor WR, Chelliah V, Hollup SM, MacDonald JT, Jonassen I. Probing the "dark matter" of protein fold space. Structure, 2009, 17(9): 1244~1252
63. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit, 2005, 18(5): 343~384
64. Shimizu K, Toh H. Interaction between intrinsically disordered proteins frequently occurs in a human protein-protein interaction network. J Mol Biol, 2009, 392(5): 1253~1265
65. Noivirt-Brik O, Prilusky J, Sussman JL. Assessment of disorder predictions in CASP8. Proteins, 2009, 77 (Suppl 9): 210~216
66. Lin J, Yang XY, Deng RQ, Yu BG, Lai HJ, Sun WL, Wu WY. Soluble expression of a strong thrombolytic pro-urokinase mutant in Escherichia coli. Protein Expres Purif, 2006, 48(1): 69~73

67. Liu YL, Tao L. Discussion of the mothods of protein structure prediction. China J Bioinform, 2007, 5 (4): 185~187

本刊中的类似文章

文章评论

反馈人		邮箱地址	
反馈标题		验证码	8656