5:229-235

Introduction

The development of energy functions and force fields for studying the behaviour of molecular systems is a major goal in physical chemistry. Prediction of native structures of proteins from amino acid sequences, simulation of the folding process, and calculation of protein stabilities are among the most ambitious goals of contemporary research in biomolecular theory [1].

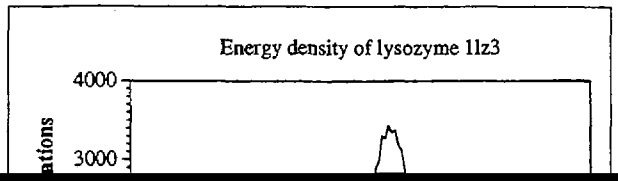
Research on these topics already has a respectable history, and the difficulties encountered over the past two decades seemed to indicate that they might be intractable because of our lack of a suitable theory of molecular interactions, and because of the computational complexities involved. We now, however, have computational tools at hand that enable the recognition of errors in experimentally determined and model structures. Furthermore, fold-recognition techniques are enabling molecular architectures of proteins to be correctly predicted, before the experimentally determined structures are

too928.35

Scheraga [11], and many others have reported subsequent attempts in the intervening period (see e.g. [12–16] and [17–24] for more recent developments).

A characteristic feature of molecular force fields

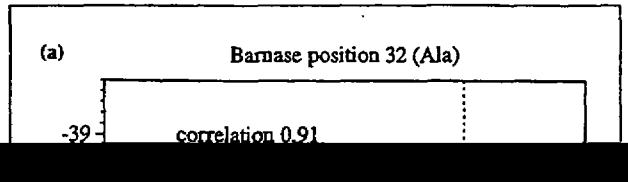
The detailed features of molecular energy functions that



Cβ pairwise and −6.2 for protein–solvent interactions), but when the two terms are combined the scores increase significantly to −9.66 (M Jaritz, MJ Sippl, unpublished data). In other words, the information contained in intramolecular pair interactions is quite different from protein–solvent interactions and both components are

actions in this molecule are unfavourable. The z-scores and energy graphs were calculated using the program PROSA II [26••] (PROtein Structure Analysis), which

There are three critical components of fold recognition techniques: first, energy functions or parameter sets providing a reasonable description of protein–solvent



Protein stabilities

A vital requirement for rational protein engineering and design is the ability to predict the effect of amino acid replacements on the stability of proteins. In some cases experimental results are well documented. In the case of

Acknowledgements

This work was supported by the Fonds Wissenschaftlicher Forschung Austria, grant 9661-MOB and by the Nationalbank of Austria, grant 5158.

References and recommended reading

17. Holm L, Sander C: **Evaluation of protein models by atomic solvation preference.** *J Mol Biol* 1992, **225**:93–105.

18. Maiorov VN, Crippen GM: **Contact potential that recognizes the correct folding of globular proteins.** *J Mol Biol* 1992, **227**:876–888.

19. Goldstein RA, Luthey-Schulten ZA, Wolynes PG: **Optimal protein folding codes from spin-glass theory.** *Proc Natl Acad Sci USA* 1992, **89**:4918–4922.

20. Avbelj F: **Use of a potential of mean force to analyze free energy contributions in protein folding.** *Biochemistry* 1992

When pair potentials are combined with solvent terms the predictive value of the energy function increases almost twofold, demonstrating the complementary nature of these energy terms.

32. Bränden C-I, Jones TA: **Between objectivity and subjectivity.** *Nature* 1990, **343**:687–689.

33. Janin J: **Errors in three dimensions.** *Biochimie* 1990, **72**:705–709.

34. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M: **The protein data bank: a computer-assisted archival file for macromolecular structures.** *J Mol Biol* 1977, **112**:535–542.

35. MacArthur MW, Laskowski RA, Thornton JM: **Knowledge based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy.** *Curr Opin Struct Biol* 1994, **4**:731–737.

36. Luethy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles.** *Nature* 1992, **356**:83–85.

37. Vriend G, Sander C: **Quality control of protein models: directional atomic contact analysis.** *J Appl Crystallogr* 1993, **26**:47–60.

38. Novotny J, Bruccoleri R, Karplus M: **An analysis of incorrectly folded protein models.** *J Mol Biol* 1984, **177**:787–818.

39. •• Holm L, Sander C: **Searching protein structure data bases has come of age.** *Proteins* 1994, **19**:165–173.
Structures in the Brookhaven data base are compared and a spanning tree is constructed displaying the structural hierarchy among the various folds. Detailed databases of structural relationships among proteins are extremely important for fold recognition tecchniques.

40. •• Orengo CA, Jones DT, Thornton JM: **Protein superfamilies and domain superfolds.** *Nature* 1994, **372**:631–634.
This paper describes work similar to that in [39••], but the relationships are obtained by different techniques. In cases of very similar structures both techniques yield the same results. For very distantly related cases, the methods produce complementary data.

41. Bowie JU, Luethy R, Eisenberg D: **A method to identify protein sequences that fold into a known three-dimensional structure.** *Science* 1991, **253**:164–170.

42. Sippl MJ, Weitckus S: **Detection of native like models for amino acid sequences of unknown three dimensional structure in a data base of known protein conformations.** *Proteins* 1992, **13**:258–271.

43. Jones DT, Taylor WR, Thornton JM: **A new approach to protein fold recognition.** *Nature* 1992, **358**:86–89.

44. Godzik A, Kolinski A, Skolnick J: **Topology fingerprint approach to the inverse folding problem.** *J Mol Biol* 1992, **227**:227–238.

45. Ouzounis C, Sander C, Scharf M, Schneider R: **Prediction of protein structure by evaluation of sequence-structure fitness. Aligning sequences to contact profiles derived from 3D structures.** *J Mol Biol* 1993, **232**:805–825.

46. Wilmans M, Eisenberg D: **Three-dimensional profiles from residue-pair preferences: identification of sequences with beta/alpha-barrel fold.** *Proc Natl Acad Sci USA* 1993, **90**:1379–1382.

47. Bryant SH, Lawrence CE: **An empirical energy function for threading protein sequence through folding motif.** *Proteins* 1993, **16**:92–112.

48. Johnson MS, Overington JP, Blundell TL: **Alignment and searching for common protein folds using a data bank of structural templates.** *J Mol Biol* 1993, **231**:735–752.

49. Sippl MJ, Weitckus S, Floeckner H: **Fold recognition.** In *Modelling of biomolecular structures and mechanisms.* Edited by Pullman A, Jortner J, Pullman B. Kluwer; 1995 in press.

50. •• Lathrop RH: **The protein threading problem with sequence amino acid interaction preferences is NP-complete.** *Protein Eng* 1994, **7**:1059–1068.
One goal in fold recognition is the calculation of optimal alignments between sequences and structures. In contrast to sequence/sequence comparison, the problem is found to be NP-complete. The consequence is that the problem is computationally intensive, and it is unlikely that fast algorithms can be found that at the same time guarantee an optimal solution.

51. •• Krogh A, Brown M, Mian IS, Sjolander K, Haussler D: **Hidden Markov models in computational biology. Applications to protein modeling.** *J Mol Biol* 1994, **235**:1501–1531.
Hidden Markov chains are applied to calculate local structures in proteins. As demonstrated recently by Tim Hubbard in a blind test, results obtained by Hidden Markov chains can be used to identify protein folds.

52. Fersht AR, Serrano L: **Principles of protein stability derived from protein engineering experiments.** *Curr Opin Struct Biol* 1993, **3**:75–83.

53. •• Jones DT: **De-novo protein design using pairwise potentials and a genetic algorithm.** *Protein Sci* 1994, **3**:567–574.
Mean force potentials are used to derive sequences compatible with a given fold. The sequence of a starting protein is forced to change while the conformation is kept fixed. Mutations are only accepted if the stability of the protein does not deteriorate.

54. Sippl MJ, Hendlich M, Lackner P: **Assembly of polypeptide and protein backbone conformations from low energy ensembles of short fragments. Development of strategies and construction of models for myoglobin, lysozyme and thymosin β-4.** *Protein Sci* 1992, **1**:625–640.

55. • Gunn JR, Monge A, Friesner RA, Marshall CH: **Hierarchical algorithm for computer modeling of protein tertiary structure — folding of myoglobin to 6.2 Å resolution.** *J Phys Chem* 1994, **98**:702–711.
Ab initio folding studies are attempted on the myoglobin sequence using knowledge-based potentials.

56. • Monge A, Friesner RA, Honig B: **An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure.** *Proc Natl Acad Sci USA* 1994, **91**:5027–5029.
The native topology of four helix bundle proteins can be found by minimizing knowledge-based potentials, when the known secondary structure is preformed.

MJ Sippl, Center for Applied Molecular Engineering, Institute for Chemistry and Biochemistry, University of Salzburg, Jakob Haringer Straße 1, A-5020 Salzburg, Austria.