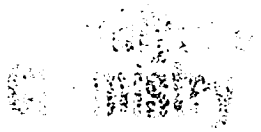
Fresenius J Anal Chem (1990) 337:1–3

Fresenius' Journal of



Symposium 1: Structure and engineering of proteins: New developments

Fresenius J Anal Chem (1990) 337:1 – © Springer-Verlag 1990

New developments in X-ray structure analysis: Facing up membranes

Hartmut Michel

Max-Planck-Institut für Biophysik,
Abt. Mol. Membranbiologie, Heinrich-Hoffmann-Strasse 7,
D-6000 Frankfurt/M. 71, Federal Republic of Germany

Manuscript not received

Fresenius J Anal Chem (1990) 337:1–2 – © Springer-Verlag 1990

From comparative structure analysis to protein engineering: knowledge-based protein modelling and design

Mark S. Johnson, John Overington, Andrej Šali, Zangyang Zhu, Dan Donnelly, Pam Thomas, Alasdair McCleod, Richard Goold, Chris Topham, and Tom L. Blundell

ICRF Unit of Structural Molecular Biology and Laboratory of Molecular Biology, Department of Crystallography, Birbeck College, London University, Malet Street, London WC1E7HX, England

Much can be learned from natural evolution that is useful for protein engineering. In particular the divergent evolution of protein families provides an important part of the knowledge base that is essential for the design of novel proteins (see Blundell et al. [1] for a review).

Knowledge-based modelling can be envisaged as a number of steps concerned with the establishment and use of rules to generate a model of a protein. One of the most powerful procedures in learning rules is the comparison of related structures either through alignment of sequences to identify conserved residues or through superposition of three-dimensional structures to identify conserved conformations or motifs. Thus the first step in a knowledge-based modelling is the systematic comparison of families of topologically similar structures. This step leads to the establishment of "equivalences" between the structures compared and to their clustering based on measures of similarity. From analysis of the comparisons rules are developed in the second stage that are useful for modelling proteins. The third step involves the projection of the results of the comparisons of three-dimensional structures down onto the level of sequence. This step uses rules relating structure to sequence. They are expressed as consensus sequences or templates for topologically equivalenced residues, or as key residues in canonical structures, which are then used to align the sequence of the protein of unknown tertiary structure. The final step uses the rules established in the second step to generate a three-dimensional model.

The classical form of knowledge-based modelling is modelling by homology or comparative modelling. This procedure depends on the knowledge that homologous structures have similar tertiary structures involving a conserved "framework" of helices and strands connected by structurally variable regions that accommodate much of the sequence variation and almost all the insertions or deletions. The method was first used over twenty years ago but has recently been developed into a systematic approach [COMPOSER] in which several homologous structures can be used in modelling the unknown (Sutcliffe et al. [22, 23]; Blundell et al. [2]). Rules are used to establish the precise relative positions of the framework (Sutcliffe et al. [22], to select appropriate fragments for variable regions not only from homologous proteins (Greer [6]; Chothia et al. [4, 5], but also from other protein structures (Blundell et al. [2]; Sibanda et al. [20]) and for the replacement of sidechains (Sutcliffe et al. [23]; Summers et al. [21]). This can be used successfully to model proteins of >40% identity (Overington et al. [15]). In a parallel development Jones and Thirup [11] have shown that modelling into electron density during protein crystallography can also be aided by selection of conformational fragments from a series of other proteins of known three-dimensional structures.

Approaches such as COMPOSER that depend on superposition of three-dimensional structures are restricted to closely related motifs or homologous structures. Chothia and Lesk [4, 5] showed that for increasingly divergent structures the number of topologically equivalenced residues obtained by

superposition decreases and the root mean square difference increases. This is mainly due to small relative translations and rotations of the secondary structural elements. This also affects the core residues (Hubbard and Blundell [7]) and results in an insufficient framework for modelling. Clearly a more flexible approach for definition of topological equivalence is required. The problem of defining topological equivalence was addressed more than a decade ago by Rossmann, Matthews and their colleagues (Matthews and Rossmann [12], for a review) who compared local mainchain direction and conformation to establish topological equivalence. An alternative approach is to simplify the structure to a series of vectors representing the axes of the helices and strands which are then compared (Murthy [13]; Richards and Kundrot [17]). In our approach we compare local properties and relations at each level in the hierarchy of protein structure and derive weight matrices from which the optimal alignment can be deduced using the dynamic programming approach of Needleman and Wunsch [14]. This approach works well for similar structures that have little or no significant sequence identity (Šali and Blundell [18]). A similar approach has been developed by Taylor and Orengo [24]). The aligned three-dimensional structures can also be clustered in a similar way to the formation of phylogenetic trees for sequences (Johnson et al. [8, 9]). This is helpful in the selection of proteins for modelling.

The next step is to derive rules that reflect the constraints of the three-dimensional structure on the sequence for a particular fold. We are approaching this problem by making detailed analyses of the substitution patterns at topologically equivalent positions in families of homologous proteins as a function of the local conformation, sidechain accessibility and hydrogen-bonding patterns (Overington et al. [16]). Templates are then derived on the basis of one or several equivalenced structures and used to align the sequence of the unknown (Johnson et al. [10]). New approaches are also being developed for the final step in the procedure whereby a model is generated (Šali et al. [19]). The fact that local properties and relations have been equivalenced indicates that internal coordinates should be used. In many ways the problem is closely related to that of reconstruction a model from upper and lower bounds on distances obtained from 2-D NMR experiments (see for example Braun and Go [3]).

We shall discuss the various approaches available in each step of the modelling procedure as they are being developed in our laboratory and then briefly discuss applications to protein engineering and design.

References

1. Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM (1987) Nature 326:347–352
2. Blundell TL, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, Overington J, Singh DA, Sibanda BL, Sutcliffe M (1988) Eur J Biochem 172:513–520
3. Braun W, Go N (1985) J Mol Biol 186:611–626
4. Chothia C, Lesk AM (1986) EMBO J 5:823–826
5. Chothia C, Lesk AM, Levitt M, Amit AG, Mariuzza RA, Phillips SEV, Poljak RJ (1986) Science 233:755–758
6. Greer J (1981) J Mol Biol 153:1027–1042
7. Hubbard TJP, Blundell TL (1987) Prot Eng 1:159–171
8. Johnson MS, Sutcliffe MJ, Blundell TL (1990) J Mol Evol 30:43–59
9. Johnson MS, Šali A, Blundell TL (1990) Methods Enzymol (in press)
10. Johnson MS, Overington J, Blundell TL (1990) (manuscript in preparation)
11. Jones TH, Thirup S (1986) EMBO J 5:819–822
12. Matthews BW, Rossmann MG (1985) Methods Enzymol 115:397–420
13. Murthy MRN (1984) FEBS Lett 168:97–102
14. Needleman SB, Wunsch CD (1970) J Mol Biol 48:443–453
15. Overington JP, Sutcliffe MJ, Watson F, James K, Campbell S, Blundell TL (1989) In: Durand G, Bobichon L, Florent J (eds) Proc Inter Biotechn Symp, pp 279–304
16. Overington J, Johnson M, Šali A, Blundell TL (1990) (manuscript submitted)
17. Richards FM, Kundrot CE (1988) Proteins 3:71–84
18. Šali A, Blundell TL (1990) J Mol Biol 212:403–428
19. Šali A, Donnelly D, Blundell TL (1990) (manuscript in preparation)
20. Sibanda BL, Blundell TL, Thornton JM (1989) J Mol Biol 207
21. Summers NL, Carson WD, Karplus M (1987) J Mol Biol 96:175–198
22. Sutcliffe MJ, Haneef I, Carney D, Blundell TL (1987) Prot Eng 1:377–384
23. Sutcliffe MJ, Hayes FRF, Blundell TL (1987) Prot Eng 1:385–392
24. Taylor WR, Orengo CA (1989) J Mol Biol 208:1–22

Fresenius J Anal Chem (1990) 337:2 – © Springer-Verlag 1990

The structural analysis of protein stability

Tom Alber

School of Medicine, Department of Biochemistry,
University of Utah, Salt Lake City, Utah 84132, USA

Manuscript not received