MEMBRANE PROTEIN MODELS

EDITED

BY

J.B.C. FINDLAY

Membrane Protein Models

J.B.C. Findlay
Department of Biochemistry and Molecular Biology,
University of Leeds, Leeds, UK

First published 1996

A CIP catalogue record for this book is available from the British Library.

ISBN 1 85996 080 4

BIOS Scientific Publishers Ltd
9 Newtec Place, Magdalen Road, Oxford OX4 1RE, UK
Tel. +44 (0)1865 726286. Fax +44 (0)1865 246823

DISTRIBUTORS

Australia and New Zealand
DA Information Services
648 Whitehorse Road, Mitcham
Victoria 3132

India
Viva Books Private Limited
4325/3 Ansari Road, Daryaganj
New Delhi 110002

Singapore and South East Asia
Toppan Company (S) PTE Ltd
38 Liu Fang Road, Jurong
Singapore 2262

USA and Canada
Books International Inc.
PO Box 605, Herndon, V A 22070

The colour section in this book is kindly sponsored by Orion-Farmos

Typeset by Euroset, Alresford, Hampshire, UK.
Printed by Information Press Ltd, Oxford, UK.

Membrane Protein Models

Contributors

Argos, P. European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, D-69012 Heidelberg. Germany

Bhogal, N. Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Blaney, F.E. SmithKline Beecham Pharmaceuticals, Computational Chemistry Department, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW, UK

Bradshaw, C. Glaxo Institute for Molecular Biology, Plan-les-Quates, CH-1228 Geneva, Switzerland

Chollet, A. Glaxo Institute for Molecular Biology, Plan-les-Quates, CH-1228 Geneva, Switzerland

Cocchi, M. Dipartimento di Chimica, Universita' di Modena, Via Campi 183, I-41100 Modena, Italy

De Benedetti, P.G. Dipartimento di Chimica, Universita' di Modena, Via Campi 183, I-41100 Modena, Italy

Donnelly, D. Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Fanelli, F. Dipartimento di Chimica, Universita' di Modena, Via Campi 183, I-41100 Modena, Italy

Finbow, M.E. Beatson Institute, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK

Findlay, J.B.C. Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Harrison, M.A. Department of Biochemistry and Molecular Biology, University of Leeds LS2 9JT, UK

Hurrell, C.R. Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

IJzerman, A.P. Leiden-Amsterdam Center for Drug Research, Division of Medicinal Chemistry, Gorlaeus Laboratories, P.O. Box 9502, NL-2300 RA, The Netherlands

Jones, P.C. Department of Biochemistry and Molecular Biology, University of Leeds LS2 9JT, UK

Kim, Y.-I. Department of Biochemistry and Molecular Biology, University of Leeds LS2 9JT, UK

Kuipers, W. Department of Medicinal Chemistry, Solvay Duphar B.V., P.O. Box 900, NL-1380 DA Weesp, The Netherlands

Lybrand, T.P. University of Washington, Center for Bioengineering, P.O. Box 351750, Seattle, WA 98195-1750, USA

Menziani, M.C. Dipartimento di Chimica, Universita' di Modena, Via Campi 183, I-41100 Modena, Italy

Milpetz, F. European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, D-69012 Heidelberg. Germany

Nemeth, K. Glaxo Institute for Molecular Biology, Plan-les-Quates, CH-1228 Geneva, Switzerland

Oliveira, L. Department of Biophysics, Escola Paulista de Medicina, Sao Pualo 04043-971, Brasil

Paiva, A.C.M. Department of Biophysics, Escola Paulista de Medicina, Sao Paulo 04043-971, Brasil

Persson, B. Department of Medical Biochemistry and Biophysics, Karolinska Institute, S-17199 Stockholm, Sweden

Rippmann, F. E. Merck, Preclinical Pharmaceutical Research, D-64271 Darmstadt, Germany

Sander, C. European Molecular Biology Laboratory, BIOcomputing department, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

Sankararamakrishnan, R. Department of Physiology and Biophysics, Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Sansom, M.S.P. Laboratory of Molecular Biophysics, University of Oxford, The Rex Richards Building. South Parks Road, Oxford OX1 3QU, UK

Tennant, M. SmithKline Beecham Pharmaceuticals, Computational Chemistry Department, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW, UK

Thomas, P. Department of Biomolecular Structure, Glaxo Research and Development Limited, The Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

Turcatti, G. Glaxo Institute for Molecular Biology, Plan-les-Quates, CH-1228 Geneva, Switzerland

Vriend, G. European Molecular Biology Laboratory, BIOcomputing department, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

Zuurmond, H.M. Leiden-Amsterdam Center for Drug Research, Division of Medicinal Chemistry, Gorlaeus Laboratories, P.O. Box 9502, NL-2300 RA, The Netherlands.

Abbreviations

5-CT	5-carboxamidotryptamine
5-HT	5-hydroxytryptamine
ASP	atomic solvation parameter
B-M	benzophenone-4-maleimide
bR	bacteriorhodopsin
BSA	bovine serum albumin
CFP	channel-forming peptides
DCCD	dicyclohexylcarbodiimide
DHA	(-)-[^{3}H]dihydroalprenolol
DOI	1-(2,5-dimethoxy-4-Br-phenyl)-2-aminopropane
8-OH-DPAT	2-(di-*n*-propylamino)-8-hydroxytetralin
ET	endothelin
F-M	fluorescein-5-maleimide
FTIR	Fourier transform infrared
GA	gramicidin A
GPCR	G protein-coupled receptors
HLA	histocompatibility antigen molecules
IE	interaction energies
LSD	lysergic acid diethylamide
MD	molecular dynamics
nAChR	nicotinic acetylcholine receptor
NBD	7-nitrobenz-2-oxa-1,3-diazol-4-yl
NK	neurokinin
PBS	phosphate-buffered saline
QMD	quenched molecular dynamics
QSAR	quantitative structure–activity relationship
SA/MD	stimulated annealing via restrained molecular dynamics
SAR	structure–activity relationship
SDS–PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
TCA	trichloracetate
TM	transmembrane

Preface

Integral membrane proteins

The membrane has represented one of the most intractable elements of the biological cell. From the time when it was realized that there was a barrier within the plant cell wall and that this barrier represented the most important and fundamental regulator of the cellular environment, considerable effort has been expended to determine its nature and properties. The lipid fraction was the first to be described but although that was achieved some time ago and interest has waned somewhat since then, there are still fundamental mysteries to explain, not least the real link between function and composition. Knowledge of the protein fraction lagged some way behind, principally due to the intractability of these polypeptides to the methodologies then available. But with the advent/use of such simple molecules as detergents and organic solvents has come the dramatic advances in our understanding of membrane proteins from the primitive days of scaffold-like fragments to the tantalizing mechanisms suggested by the high-resolution 3-dimensional structures of the photosynthetic reaction complex and cytochrome oxidase.

Now that the molecular description of biology is in full flood, it is clearly important to be able to describe the structure and mechanisms of action of integral membrane proteins, both the types that are responsible for the translocation of material and those which mediate the transfer of information. But this urgent need has not so far been fulfilled because yet again, routine methodology is still not available. This has given rise to alternative strategies to gain impressions, albeit crude and unreliable ones, of the structures of integral membrane proteins. The usefulness of such representations in facilitating experimental design and in stimulating the development of new concepts should not be underestimated, however, for there is ample evidence of their predictive potential at least at a low resolution level. During the conference of which these are the proceedings, many of the strategies used to further our structural understanding of integral membrane proteins were described and discussed. Many models were presented and their strengths, weaknesses, contradictions and inconsistencies thoroughly explored.

G protein-coupled receptors

Most attention was devoted to the G protein-coupled receptors (GPCRs), surely now the most avidly pursued and widespread family of proteins in eukaryotic biology. Little did those few groups who wrestled with the mechanism of action, sequence and topography of the visual pigment rhodopsin appreciate the avalanche of structural and functional relatives that lie in the various eukaryotic genomes. The elucidation of the seven transmembrane segments of opsin has given rise to one of the most potent signatures of any protein family. There are now in the database over 700 (and rising) sequences which are recognized by this composite signature. Clearly, this is still only a small proportion of the total if the estimates of the number of olfactory receptors alone are in any way accurate. Already, however, interesting divisions are appearing which in some ways reflect evolutionary distance. The classical sequence-based motifs are incapable of recognizing sub-families such as those for yeast mating factor receptors, the dictyostelium cAMP receptors, the secretin subgroup and receptors for glutamate, Ca, gonadotrophin hormones etc. All appear to be involved in some way with G proteins but perhaps the mechanism of interaction and activation differs.

The assumption, so far unproven, is that the basic framework of these receptors remains more or less intact but considerable functionality is incorporated, often in association with new structural elements or additional domains. Thus, as the ligand gets larger, as with the peptide receptors, the hydrophilic surface regions including the extended N-terminus are recruited in to generate specific binding pockets. This development reaches full expression in the large N-terminal extensions seen in the follicle-stimulating hormone, thyroid-stimulating hormone and luteinizing hormone receptors which have a major role in interacting with the protein ligand. Even larger domains also occur in the glutamate and calcium receptors but here it is harder to give a rational explanation for these regions given the small size of the ligand. Whatever their precise role, these large domains are nevertheless both structurally and functionally intimately associated with the transmembrane segments through which the information flow must pass. It is to be expected that additional elements of structure will also be found at the intracellular face of the receptor. This has already been seen in the cephalopod visual receptors which possess a very unusual region at the C-terminus, one that intriguingly occurs in other, unrelated proteins.

On the basis of limited biophysical evidence, bacteriorhodopsin was the template used for the first tertiary structure representations of GPCRs. This assumption was justified when the low resolution structure of rhodopsin appeared. But it is also clear that the template was only an approximate one for there is variation in the relative position and pitch of some of the transmembrane helices. It is reasonable to assume that rhodopsin is now a more accurate template but, at the same time, it would not

be unexpected to find that further differences occur throughout the family and more likely still amongst the more distant and less conserved sub-families.

This conference was designed as a discussion forum to explain and examine the various approaches to modelling integral membrane proteins in general and GPCRs in particular. It revealed wide differences in interpretation particularly of the mutagenesis data. The protein models themselves, despite having different origins and being generated by different methods, apparently had a surprising degree of convergence. The really striking differences were seen in the area of ligand docking where quite different interpretations were put on much the same sets of data. The inevitable conclusion from this was that models whilst stimulating, provocative and reasonably predictive required much firmer structure-based data before one could reliably move from low resolution representations to higher resolution 'structures'.

Finally, one should not overlook the evolving but still unconfirmed concepts in GPCR structure/function relationships. The ternary complex model and the suggestions of a range of conformational intermediates between the activated R* (agonist binding) and inactive R (antagonist binding) modes present new structural challenges. So too do the increasingly frequent suggestions of different binding epitopes for agonists and antagonists, particularly when the former is large and the latter is relatively small. The even more detailed models of GPCRs will have to take on board these subtle concepts and observations. Part of this appreciation must involve a rational description of the activated state, particularly as it applies to the fascinating array of constitutively active mutants.

Transporters/channels

Whilst there is some structural basis for modelling GPCRs, the situation for transporters/channels is much more desperate, at almost all levels of resolution and types of data. The best information so far comes from the structure of bacterial outer membrane porins but it is not at all clear how representative these will be of the majority of channels, since the entire membrane domain consist of an emaculate β–barrel. The suggestions made for the acetylcholine (nicotinic) receptor are intriguing in that a bundle of helices appear to provide the transport route but these may be embedded in a continuous shell of β-sheet. Such data as exist for other transporters/channels emphasize a substantial α-helical content, sometimes incorporating a channel-lining segment of unknown secondary structure. Thus, it looks as though we may have a fascinating mixture of types with much scope for imagination. However, there are both modelling and experimental approaches which might provide useful insights into the structural and functional properties of proteins which mediate material transport. A few examples of these are presented here to illustrate their potential application to model construction.

The meeting was entitled: 'Membrane Protein Models: Experiment, Theory and Speculation' (Leeds, UK, March/April 1994, under the auspices of The Molecular Graphics Society). From the excellent presentations that are included in this text, the reader will quickly appreciate the lively and stimulating debates that took place on all three aspects of the structure and function of integral membrane proteins.

J.B.C. Findlay

Contents

Chapter 1

Prediction of transmembrane segments in proteins using multiple sequence alignments

Bengt Persson, Frank Milpetz and Patrick Argos

1. Introduction

Membrane proteins are important for several processes and functions in all biological systems. For example, they act as receptors for neurotransmitters or hormones (Savarese and Fraser, 1992; Stephenson, 1991), form a wide variety of ion channels (Barnard, 1992; Miller, 1991), or serve as the respiratory chain (Capaldi, 1991) and transport proteins for different molecules (Griffith *et al.*, 1992; Marger and Saier, 1993; Schloss *et al.*, 1992). Bacterial toxins which form membrane pores (Li, 1992) also belong to this large group of lipid-associated molecules. There are several surface molecules which are anchored to the membrane by one transmembrane segment, for example histocompatibility antigen molecules (HLA), neuraminidases and haemagglutinins (cf. Popot and de Vitry, 1990).

Despite their biological significance tertiary structures have been determined for only a few membrane proteins. Three-dimensional structures at medium to high resolution are available for bacteriorhodopsin (Henderson *et al.*, 1990), photosynthetic reaction centre (Deisenhofer *et al.*, 1985), light-harvesting complexes (Kühlbrandt *et al.*, 1994; McDermott *et al.*, 1995), prostaglandin H_2 synthase-1 (Picot *et al.*, 1994), porin (Weiss and Schulz, 1992) and cytochrome *C* oxidase (Iwata *et al.*, 1995). Other structures are not yet fully resolved, for example photosystem I (Krauss *et al.*, 1993) and nicotinic acetylcholine receptor (Unwin, 1993). Given the scarcity of tertiary structural information, many experimental methods have been applied to determine membrane topology (Jennings, 1989), including analyses of gene fusion proteins and studies of biochemically modified membrane proteins (cf. Traxler *et al.*, 1993).

Membrane proteins constitute a ubiquitous group of structures with members representing several different types of molecular architecture. However, since each traverses the lipid bilayer once or several times, they generally possess hydrophobic sequence segments. Various prediction methods use this characteristic to determine the location of these membrane-spanning regions, albeit with varying degrees of accuracy.

2. Present prediction methods

Theoretical prediction algorithms have been shown to be useful in detecting membrane-spanning segments from the primary structure alone, especially as an aid to designing experiments investigating protein topology. One of the most widely used is that of Kyte and Doolittle (1982), where mean residue hydrophobicity values are calculated for consecutive 19-residue sequence spans. Segments with hydrophobicity above a certain threshold are predicted to be membrane-spanning. A similar approach is adopted by Rao and Argos (1986), who also considered residues that break the transmembrane helices in order to improve reliability of prediction.

Different prediction methods were reviewed and evaluated in 1990 by Degli Esposti *et al.* They examined the correlations amongst the various amino acid hydrophobicity scales used and compared the accuracy of the various prediction approaches. They also calculated a new set of parameters derived from seven different scales (Degli Esposti *et al.*, 1990). A trapezoidal sliding window was used by von Heijne in his hydrophobic analysis of the sequence together with a consideration of positively charged residues interior to the membrane (von Heijne, 1986) particularly with application to the topology of bacterial inner membrane proteins (von Heijne, 1992). These rules were also applied to a number of eukaryotic membrane proteins (Sipos and von Heijne, 1993). Several studies have also been effected regarding helix–helix interactions in membrane proteins (e.g. Lemmon and Engelman, 1992).

Here we present a transmembrane helix prediction algorithm, based upon multiple sequence alignments of related proteins. This method takes advantage of the extended information not found in analysis of a single sequence as is characteristically used by other approaches. The algorithm is described in detail elsewhere (Persson and Argos, 1994). Present primary structural databases are large and expanding at such a rate that homologous sequences are often found. We show that this technique has higher accuracy in predicting transmembrane segments than previous methods based on individual sequences.

3. Residue distributions in membrane proteins

The tertiary structure of only a few membrane proteins are known. However, other types of data exist from various experiments and predictions to deduce the location