

Transmembrane protein structures without X-rays

Sarel J. Fleishman¹, Vinzenz M. Unger² and Nir Ben-Tal¹

¹Department of Biochemistry, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Ramat Aviv 69978, Israel

²Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208024, 333 Cedar Street, New Haven, CT 06520-8024, USA

Transmembrane (TM) proteins constitute 15–30% of the genome, but <1% of the structures in the Protein Data Bank. This discrepancy is disturbing, and emphasizes that structure determination of TM proteins remains challenging. The challenge is greatest for proteins from eukaryotes, the structures of which remain intractable despite tremendous advances that have been made towards structure determination of bacterial TM proteins. Notably, >50% of the membrane protein families in eukaryotes lack bacterial homologs. Therefore, it is conceivable that many more years will elapse before high-resolution structures of eukaryotic TM proteins emerge. Until then, integrated approaches that combine biochemical and computational analyses with low-resolution structures are likely to have increasingly important roles in providing frameworks for the mechanistic understanding of membrane-protein structure and function.

Introduction

It is estimated that transmembrane (TM) proteins constitute ~15–30% of eukaryotic genomes [1–4]. Owing to their strategic localization at the interfaces between the interior and exterior of the cell and between cellular compartments, membrane proteins have pivotal roles in many cellular processes, including cell-to-cell signaling events, solute transport and cellular organization. For this reason, membrane proteins are by far the most attractive targets for drug discovery. Despite their importance, however, only a few distinct folds of TM proteins have been solved to date by high-resolution methods such as X-ray crystallography [5] and nuclear magnetic resonance (NMR) [6]; therefore, TM protein structures constitute <1% of the entries in the Protein Data Bank (PDB). Disturbingly, only two of the current entries represent a membrane protein from human origin [7,8], whereas the majority of entries are of bacterial membrane proteins (Figure 1).

Part of the reason why progress has been faster for bacterial membrane proteins stems from the fact that they can more easily be expressed in large quantities in bacterial hosts, and that they lack many of the post-translational modifications that potentially complicate

crystallization. Moreover, the fast pace at which bacterial genomes are sequenced provides an almost unlimited repertoire of target proteins including homologs from thermophilic bacteria that are often more stable during detergent solubilization, purification and crystallization. By contrast, eukaryotic membrane proteins are more difficult to express than their bacterial homologs, are subject to post-translational modifications and, often, only few candidate genes are available for screens to identify the ideal target protein. It thus comes as no surprise that, over the past few years, efforts have been focused on identifying bacterial homologs of eukaryotic membrane proteins, and pursuing their structure determination by ‘brute-force’ approaches, sometimes using thousands of combinations of homologs of the protein and different crystallization conditions [9]. This strategy has begun to bear fruit (Figure 1) and, indeed, the recent growth in novel TM-protein structures was estimated to be exponential, as it is for soluble proteins, suggesting that, over the next few years, many new structures will emerge [5]. However, this growth has not been steady over the years and, more importantly, has been restricted mostly to TM proteins from bacteria; the pace of discovery of novel TM proteins from eukaryotes, however, has remained low (Figure 1). Notably, the use of bacterial homologs for eukaryotic TM proteins does not represent the ultimate solution because many eukaryotic membrane proteins do not have bacterial homologs. In fact, a search in the Pfam-A database of protein families [10] shows that only 47% of the eukaryotic TM protein families have bacterial or archaeal homologs.

In an attempt to overcome the problem of there being such a large proportion of eukaryotic proteins for which direct structure determination is likely to have to wait many years, data-based modeling approaches were developed that rely on inferences derived from biochemical, computational, evolutionary and intermediate-resolution structural methods. Here, we focus on the methods that have been used to model helical membrane proteins before their experimental structure determination at high resolution. Notably, helical proteins are the dominant class of TM proteins in eukaryotes and in bacterial inner membranes. We also delineate potentially productive venues for future research. We will not deal with comparative or homology modeling applied to TM proteins (but see recent reviews [11,12]).

Corresponding author: Ben-Tal, N. (nirb@tauex.tau.ac.il).

Available online 10 January 2006

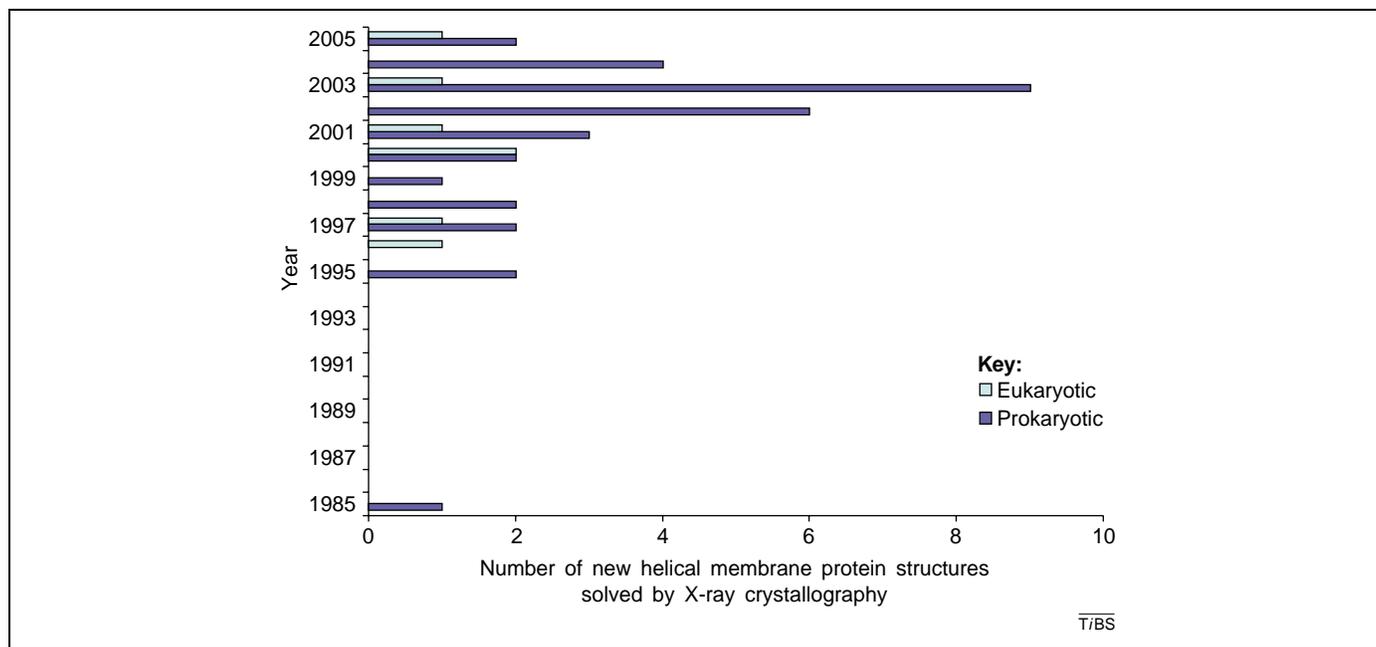


Figure 1. Number of new helical membrane-protein folds solved in recent years. Tremendous progress has been made over the past few years in crystallization of TM proteins from bacteria, although the growth in the discovery of novel structures has not been steady. Moreover, crystallization of eukaryotic TM proteins still lags far behind, and only a handful of structures have been obtained. The entry for 2005 includes structures up to and including November 2005.

Architecture of helical TM proteins

A simple rule that has guided many of the approaches to modeling helical TM proteins is the two-stage model of folding [13]. According to this model, hydrophobic segments are first inserted into the plasma membrane in the form of helices, which engage the polar carbonyl and amide groups on the backbone of the peptide chain through hydrogen bonds, and shield them from the hydrophobic lipid bilayer. Next, these helices associate with one another to shape the tertiary structure of the protein. One of the implications of the two-stage model for computational modeling is that each of the hydrophobic segments comprising the TM domain can be approximated as an energetically stable canonical α helix, the polar backbone and N and C termini of which are shielded from the membrane environment. Hence, TM-protein-structure prediction can concentrate on the relative configurations of preformed α helices. This constraint considerably reduces the number of degrees of freedom that must be explored computationally.

This quite simple picture of TM-protein architecture was supported by the first few membrane proteins to be solved [14–17] (e.g. that shown in Figure 2). Moreover, the extramembrane loops are short in these proteins, dictating that consecutive domains in the sequence are proximal in the 3D structure [18]. However, this simplistic picture collapsed when the first ion-channel structures revealed that helices need not span the entire width of the bilayer [19], and can be extremely long and highly tilted with respect to the membrane normal [20] (Figure 3a,b). Recent transporter structures have also shown marked deviations from α helicity; it has been suggested that these deviations have a role in the conformational changes underlying transporter functions by destabilizing the structures [21] (Figure 3c). All of these structural features are still beyond what can be reliably predicted by

computational methods, raising the question of how many membrane domains might have gone unnoticed by contemporary methods for the detection of TM spans [22]. More importantly, however, the observation that not all consecutive hydrophobic domains form physical contacts [19,20] heralded the end of naïve modeling of TM proteins, and underscored the importance of a joint experimental–computational approach to structure prediction. Over the past several years, two sources of experimental data have proven valuable in aiding most modeling exercises of membrane proteins: low-resolution structures obtained by

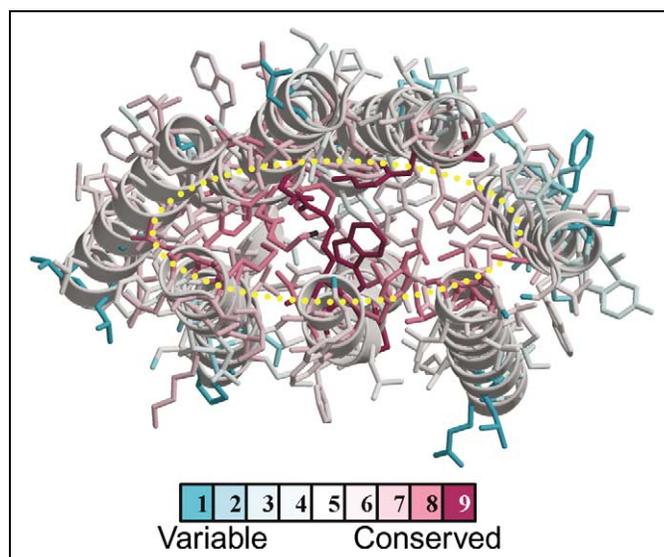


Figure 2. Evolutionary conservation can aid the orientation of TM helices. Evolutionary conservation is projected on the bacteriorhodopsin structure viewed from the direction vertical to the membrane plane, showing that the core of the protein (within the yellow ellipse) is more conserved than its periphery. Conservation was computed using the ConSurf webserver (<http://consurf.tau.ac.il/>) [66].

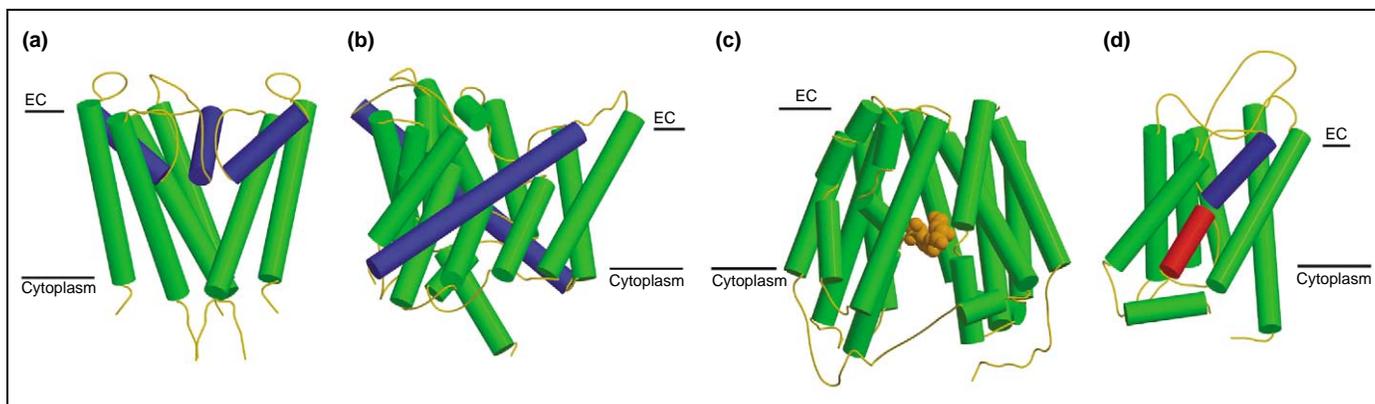


Figure 3. Recent structures reveal many discrepancies from the view that TM helices are canonical and span the entire lipid bilayer. **(a)** For clarity, only three of the four monomers comprising the K^+ ion channel are shown [19]. Blue cylinders represent the pore helix, which spans only half of the membrane width. **(b)** A monomer of the Cl^- channel [20]. The blue cylinders represent the locations of helices B and J, which are highly tilted with respect to the membrane normal and comprise ~ 35 amino acids each. **(c)** Structure of the transporter lac permease [21]. Some of the helices are kinked. Orange spheres represent a lactose analog. **(d)** Structure of the aquaporin 1 water channel [7]. Blue and red cylinders represent two half helices that meet at the mid-point of the membrane. (EC, extracellular.)

cryo-electron microscopy (cryo-EM) and mutational analyses of structure–function relationships.

Cryo-EM of 2D crystals of TM proteins

In contrast to the difficulties usually experienced in obtaining 3D crystals of TM proteins, in some cases, membrane proteins readily form 2D arrays in the membrane (e.g. bacteriorhodopsin [23], photosystem II [24], the gap junction [25], the bacterial translocon complex secYEG [26] and the bacterial multidrug-resistance transporter EmrE [27]). Added advantages of 2D crystals are that they mimic the native environment of the protein more closely than 3D crystals do, including interactions with the surrounding lipid molecules, which sometimes have important roles in determining the physiological structure [28]. For instance, substantial differences were observed between the cryo-EM map of EmrE [27] and a structure of the protein derived from X-ray analysis of 3D crystals [29]. Another demonstration of the importance of maintaining a membrane-like environment is provided by the differences between two recent X-ray structures of the voltage-gated K^+ channel [30,31], one of which was crystallized in the presence of lipids. In addition, it is sometimes possible to induce crystal formation in 2D, even when the proteins are dispersed in the membrane [32], and small and poorly ordered crystals can be used to derive data in the 5–10-Å resolution range thanks to digital-image-processing protocols that enable crystals to be corrected for translational disorder [14,33,34].

However, cryo-EM of 2D crystals usually produces structures at limited resolutions (typically, > 4.5 Å in the plane of the membrane) so that individual amino-acid sidechains are not visible and, often, flexible loops and extramembranous domains are unresolved owing to lack of crystallographic order. Moreover, the resolution in the direction vertical to the lipid bilayer is worse than the in-plane resolution. This reduced resolution entails an uncertainty regarding the actual length of each helical segment, and might obscure the helical register. The lower vertical resolution might also limit the detection of helices that do not span the entire bilayer. In the case of

the aquaporin-1 water channel for instance, an initial map at 6-Å in-plane resolution [35] did not reveal the surprising architecture of the channel, whereby two half-helices meet midway through the membrane (Figure 3d): misleadingly, these half-helices seemed to be one. A subsequent cryo-EM map at 4.5-Å resolution uncovered the two half-helices [36], and enabled a combination of sequence-based methods to be used to predict a model structure [37,38], which was found to be in agreement with the subsequently solved high-resolution structure [7]. The initially incorrect interpretation underscores the importance of improving resolution even marginally within the intermediate-resolution range to ascertain the general architecture of the protein.

Despite these shortcomings of intermediate-resolution maps, the fact that they provide an overall description of the protein architecture and the approximate packing of TM helices tremendously reduces the degrees of freedom for conformational search and the extent of uncertainty in constructing model structures. In fact, by assuming that ideal α helices occupy the locations observed in the map, the conformational search for the backbone positions can be limited to identifying the native-state orientation of each helix around its principal axis [39].

Building on this realization, and using further constraints obtained from multiple-sequence alignments and biochemical data (Box 1), Baldwin *et al.* [40] pioneered a structure-based modeling approach to derive the first model of the G-protein-coupled receptor (GPCR) rhodopsin based on a structure at 7-Å in-plane resolution [41]. Although rough, this model served as a template for modeling other GPCRs, which then provided a framework for interpreting the effects of mutations in the context of the receptor structure (see, for example, Refs [42,43]). Three years later, the first high-resolution structure of rhodopsin was solved by X-ray crystallography of 3D crystals [44], and showed that the previous model approximated the native-state structure to within 3.2-Å root-mean-square deviation. The orientations of all of the helices were predicted quite accurately by Baldwin *et al.* [40], and the main structural differences were due to

Box 1. Combinations of methods used in TM-protein-structure prediction

Many of the modeling applications for TM-protein structures have used at least some of the data sources and analyses shown in Figure 1 [37,40,47,67]. For many TM proteins, sufficient biochemical and biophysical data are available, specifying, for example, which sequence segments form helices [54] and make contact with other helices [48]. These data can be used to predict or verify the model. By contrast, cryo-EM maps at resolutions that enable the helix-packing arrangement to be discerned (typically better than 10 Å) have so far been obtained for only a few TM proteins, but more are expected to follow. The last two stages of modeling, in which modeling is refined

by direct experimentation, have not yet been implemented in structure prediction of TM proteins. Generating atomic-resolution models (final step in Figure 1) is complicated by the fact that even minor differences from the native-state structure often result in energetically unfavorable steric clashes and the abrogation of favorable polar bonds [68]. Because the positions of the backbone atoms inferred in the former steps of the flowchart are, at best, approximations of the native-state structure [39], conformational searching must also explore backbone degrees of freedom. Nevertheless, atomic resolution could considerably increase the quality and utility of TM-protein-structure prediction.

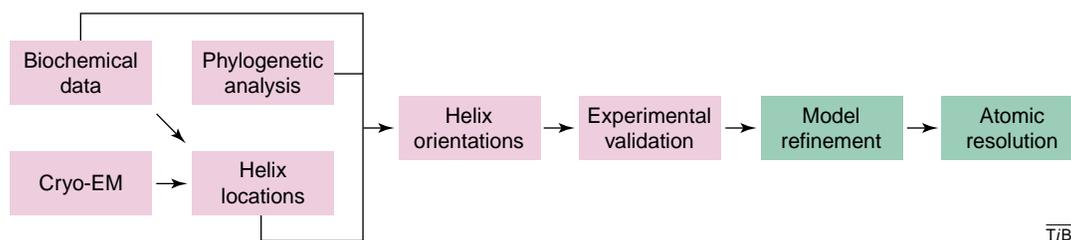


Figure 1. A flowchart for the modeling of TM-protein structures. Various sources of experimental and computational information are often integrated to model TM proteins. The last two stages (green) have, so far, not been implemented in TM-protein structural modeling.

deviations in the positioning of the kinked helices (Figure 4).

The successful combination of cryo-EM and computational methods for the modeling of rhodopsin served as a basis for developing automatic tools for modeling based on phylogenetic analysis (Box 2) and intermediate-resolution structures [39,45]. These methods were then used to predict the structure of the TM domain of the gap junction, which is a channel that connects neighboring cells in a tissue, and lacks bacterial homologs [46]. A map of the gap junction was solved initially at a resolution of 7.5 Å in the membrane plane [25], and was subsequently improved to 5.7 Å [47]. The intermediate-resolution structure revealed a large pore (~15-Å diameter at the point of constriction), and clearly distinguished the four helices (M1–M4) that comprise each of the six gap-junction forming connexin monomers [25]. Because the intermediate-resolution map did not reveal the connectivities between the TM helices, the four hydrophobic segments in connexin sequences (M1–M4) were assigned to the four helices seen in the structure based on a combination of experimental and computational data. Subsequently, the four helices were oriented using evolutionary conservation and evolutionarily correlated mutations (Box 2; Figure 5).

Using this combination of approaches and data sources, a 5.7-Å resolution map (in-plane resolution), evolutionary conservation and correlated mutations, the final model structure predicted previously undetected interactions between pairs of polar residues in the structure. The model also suggested a molecular cause for almost 30 disease-related mutations. Although not taken into account during modeling, most of these mutations were revealed as mapping to structurally packed regions of the helix bundle, whereas two physico-chemically radical polymorphisms localized to the more spacious regions of the structure facing the lipid or the pore lumen [47]

(Figure 5). Although it is clearly a model, it seems worthwhile to point out that the gap junction is an example of a eukaryotic membrane protein, the intermediate-resolution structure of which has not been superseded by a high-resolution crystal structure even six years after its original publication. Given the difficulties in obtaining well-ordered 3D crystals of eukaryotic membrane proteins, it seems likely that more cases are to follow, emphasizing why structure-based modeling is important and how it can help to generate a

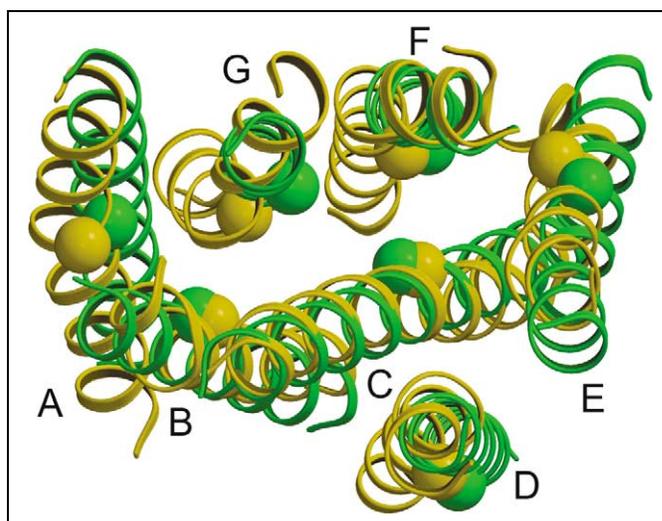


Figure 4. Comparison of the hypothetical and high-resolution structures of rhodopsin. The crystal and the hypothetical structures of rhodopsin are superimposed (yellow and green, respectively). The hypothetical structure was modeled on the basis of an electron-density map at 7-Å in-plane resolution [40]. The two structures deviate by 3.2-Å root-mean-square. Spheres are included to aid identification of identical positions in the hypothetical and crystal structures. The orientations of all of the helices are similar, and the main differences are in the locations of the helices within the plane of the membrane, particularly in the kinked helices F and G.

Box 2. Phylogenetic analysis used in TM-protein-structure prediction

Phylogenetic inference, and particularly conservation analysis, has found many applications in TM-protein-structure prediction [37,40,47]. Based on a multiple-sequence alignment of homologs of the target protein, individual amino-acid positions that show a low degree of sequence variation are considered important for protein structure or function [64], and are placed at strategic locations in the model structure, for example, at the interfaces between helices. Conversely, variable positions are considered to be unimportant, and are placed in lipid-facing positions. This type of sequence-based analysis is analogous to a large-scale mutagenesis scan conducted by evolution. To further refine the role of individual sidechain contributions, determination of evolutionarily co-varying sites can provide clues for contacts between positions. That is, if two positions form contact in 3D space, then a substitution in one site could be compensated by a substitution in the other. In this sense, determination of co-varying sites by any of several methods (see, for example, Refs [45,69]) can be regarded as an *in-silico* second-site suppression screen.

framework for planning and interpreting biochemical studies.

Biochemical and biophysical assays provide restraints for modeling

Mutagenesis and cross-linking assays have long been used to probe structure–function relationships in TM proteins,

where high-resolution structures were not available (for reviews see Refs [46,48,49]) (Box 1). One aspect in which these techniques can aid modeling is validation because models make specific predictions regarding physical contacts between pairs of residues. Mutation analyses can also be used in the earlier stages of modeling. For instance, they have been used to identify the packing interfaces between helices [50] and the positions of pore-lining residues in channels [51], and cross-linking data have been used to constrain distances between pairs of positions [52]. Biochemical and biophysical analyses can also be used to assign the hydrophobic domains in the protein sequence to the helices seen in low-resolution structures [47,53], and to identify the secondary structure and tilt angles of the helices with respect to the membrane normal [54]. However, a major pitfall – which has obscured structural interpretation of some of these data – is the fact that mutagenesis assays cannot be used to discriminate between direct and indirect effects on helix association.

Glycophorin A (GpA), which is a small and extensively characterized bitopic protein that forms homodimers in the plasma membrane [55], is a good example to illustrate this problem. Much work has been conducted by Engelman and co-workers to explore the determinants of stability in GpA dimerization, and to gain insights into the process of helix association in the membrane.

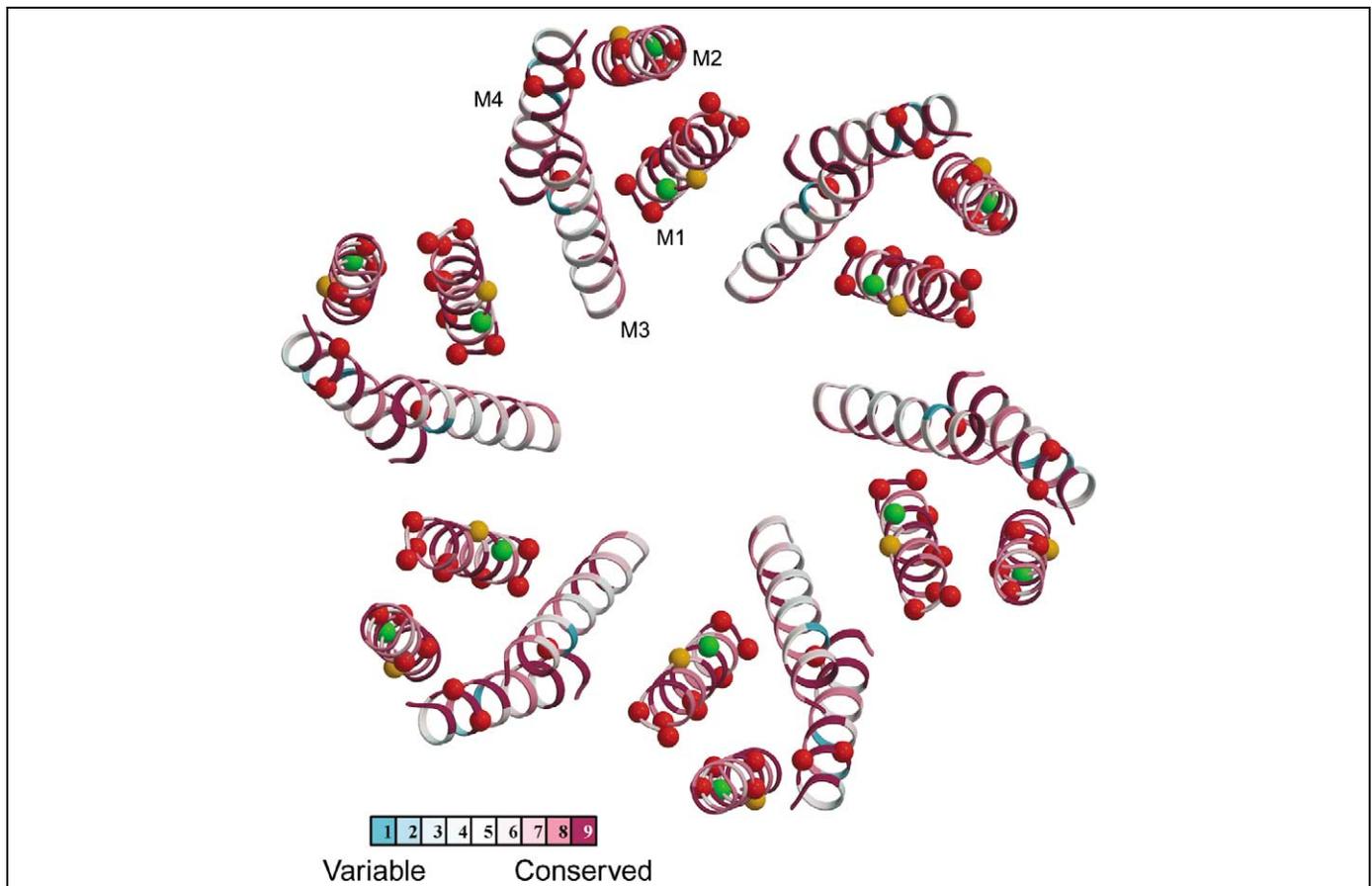


Figure 5. The model structure of the gap junction TM domain. The model structure is viewed from the cytoplasm of one cell looking in the direction vertical to the membrane [47]. Evolutionary conservation is color-coded on the structure according to the key. The positions of physico-chemically mild mutations that cause diseases and are packed within the bundle core are shown as red spheres; physico-chemically mild mutations that are not packed within the bundle core are shown as orange spheres; and physico-chemically radical polymorphisms are shown as green spheres. Almost all of the mild disease-causing mutations pack inside the bundle core, whereas the radical polymorphisms face the lipid or the pore lumen [47].

Systematic mutagenesis work by Lemmon *et al.* [50] identified a short sequence motif consisting of two glycine residues (Gly79 and Gly83), in which even physico-chemically mild substitutions abrogated dimerization. It was suggested that two glycine residues, which are small and polar, separated by three residues in the amino-acid sequence (Gly-Xaa-Xaa-Xaa-Gly) facilitate a closer approach of the two interacting α helices. It was later found that the Gly-Xaa-Xaa-Xaa-Gly motif can drive the dimerization of hydrophobic segments [56], and that it is statistically over-represented in TM sequences [57]. This and other sequence motifs were shown to have structural and functional roles in various TM proteins [58].

Although the role of the two glycine residues in the dimerization of GpA was deduced correctly from the mutagenesis assays, the same assays initially led to wrong conclusions with regard to Thr87 [50]. This position was also shown to be crucial for dimerization, and two different structural models based on molecular-dynamics-simulated annealing were suggested that supported the important roles of this triad of residues in dimerization [59,60]. One model consisted of an asymmetric right-handed supercoil [59]; the other model, suggested four years later, showed symmetric right-handed packing of the two helices [60]. The models agreed that the two glycine residues mediate much of the inter-helix contact. However, whereas Thr87 made a direct contribution to helix association in the earlier model by forming an inter-helical hydrogen bond, the residue stabilized the interface indirectly in the later model by forming an intra-helical hydrogen bond. The structure of GpA solved subsequently by NMR [8] supported the latter model (Figure 6). It is interesting to note that, because mutation analyses alone cannot discriminate between these two types of contributions to helix interaction, the interpretation of the experimental results led initially to the acceptance of an incorrect model [59].

A major problem with molecular dynamics (which was used in the prediction of GpA [59,60]) is that it is computationally demanding, essentially restricting its application to small homo-oligomers. A different approach for predicting the structures of pairs of TM helices was

recently suggested that is based, in essence, on an integration of the experimental data on the stability of TM oligomers [61]. Studies of model TM proteins such as GpA highlighted the important role of small and polar sidechains in mediating inter-helix contacts [56–58]. Thus, a simple scoring function was suggested that favored contact formation between such residues, and penalized contacts mediated by large residues. The scoring function can discriminate between decoys and the conformations of several native-state pairs of tightly packed helices with known structures, including GpA. Using this function, it has been found that the TM domains of the receptor tyrosine kinase ErbB2 could exist in two stable alternative conformations [62], which is in agreement with *in vitro* studies [63]. These results were used to suggest a model of activation for this receptor that is coupled to a switch between the two conformations of the TM domain. However, a major drawback of this method [61] is that it assumes that the pairs of helices under study are closely packed ($<9\text{-\AA}$ separation between the principal axes of the helices), thus, in effect, precluding its applicability to most polytopic proteins [61].

Recently, a different modeling strategy, based on a combination of biochemical and biophysical data, was applied to the lac-permease [53]. This 12-membrane-spanning bacterial protein catalyzes the stoichiometric transport of galactosides with a proton across the membrane. The transporter was extensively investigated using a combination of single-site mutagenesis, double-cysteine mutants, second-site suppressors and biophysical methods [48]. In modeling, the data on the membrane-spanning segments were interpreted as constraints that approximate helical structures, and other experimental results were employed to provide 99 long-range constraints (between residues that are not sequence neighbors). Several other constraints were derived from data about the residues that participate in the binding of ligand. These distance constraints were then used in modeling the protein structure based on algorithms that are employed in NMR studies [53]. Once the structure was solved at atomic resolution by X-ray crystallography of 3D crystals [21] (Figure 3c), it was possible to compare it to the model. The comparison revealed many global discrepancies but also confirmed many local interactions (e.g. residues that interact directly with sugar and positions of residues involved in proton translocation). A closer look at the constraints derived from the cross-linking experiments [52] showed that the distances implied by these data agreed with the crystal structure on the compact side of the protein that faces the periplasm (Figure 3c). However, many of the constraints in the cytoplasmic-facing part of the protein consistently underestimated the distances seen in the crystal structure by $\sim 10\text{ \AA}$ [64].

Two main reasons were suggested for the discrepancies between the model and the crystal structure [21]: (i) the transporter is a dynamic structure with alternating cytoplasmic- and periplasmic-facing conformations, and, consequently, results from the mutation and cross-linking analyses reflect a superposition of several conformational sub-states; (ii) using disulfide-bond formation as an indication for inter-residue proximity tends to

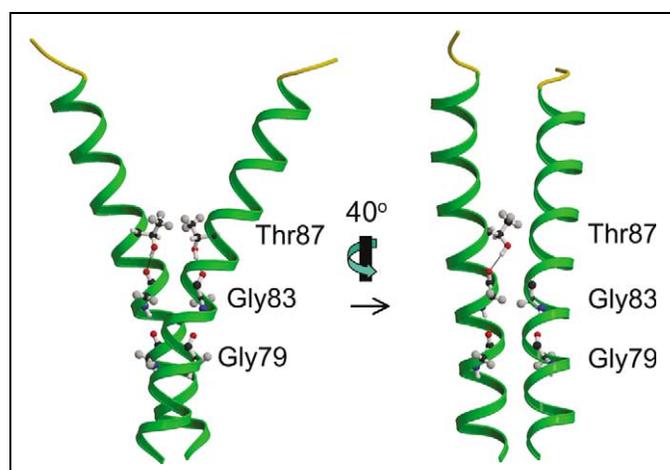


Figure 6. NMR Structure of GpA [8]. The two glycine amino acids enable the helices to pack tightly. In the view rotated by 40° (right), the intra-helical hydrogen bond between Thr87 and Gly83 is marked with a solid line.

underestimate the distances in the native state because the conformational changes of the protein bring into proximity residue pairs that might nevertheless be distal in the native structure. It is interesting to note that the cross-linking results indicated a model structure for the alternative periplasmic-facing conformation based on rotation of part of the structure with respect to the other [64]. In this alternative conformation, many of the experimental constraints in the periplasmic domain were consistent with the structure.

Concluding remarks

Structure determination of eukaryotic membrane proteins remains too slow to sustain hypothesis-driven experimentation aimed at understanding structure–function relationships in integral membrane proteins. Here, we have given examples for how mutational and computational techniques can be used to overcome this bottleneck by exploiting the information that is contained in intermediate-resolution structures obtained by cryo-EM. Although none of the techniques in isolation can provide anything more than clues, the sum of the different approaches yields insights into the structures of the targeted proteins at the level of individual amino-acid residues. Undoubtedly, as more intermediate-resolution structures emerge in the future, modeling techniques will be further refined and might be extended to include modeling of sidechains and non-canonical structures (Box 1) such as bulges and kinks [65]. Such refinements and extensions are likely to become crucial in the field of TM-protein structural studies, and will present researchers with a treasure trove of testable hypotheses to gain mechanistic insights into the function of integral membrane proteins.

Acknowledgements

We thank I.T. Arkin, D.M. Engelman, S. Harrington, H.R. Kaback and P.L. Sorgen for many useful comments. We regret that, owing to the focus on methods that have been used to predict novel structures, several new approaches are not reviewed. This study was supported by grant 222/04 from the Israel Science Foundation (ISF) to N.B.-T., and NIH grants GM66145 and GM071590 to V.M.U. S.J.F. was supported by a doctoral fellowship from the Clore Israel Foundation.

References

- 1 Liu, J. and Rost, B. (2001) Comparing function and structure between entire proteomes. *Protein Sci.* 10, 1970–1979
- 2 Rost, B. *et al.* (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* 5, 1704–1718
- 3 Mitaku, S. *et al.* (1999) Proportion of membrane proteins in proteomes of 15 single-cell organisms analyzed by the SOSUI prediction system. *Biochem. Biophys. Res. Commun.* 256, 165–171
- 4 Krogh, A. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580
- 5 White, S.H. (2004) The progress of membrane protein structure determination. *Protein Sci.* 13, 1948–1949
- 6 Opella, S.J. and Marassi, F.M. (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chem. Rev.* 104, 3587–3606
- 7 Murata, K. *et al.* (2000) Structural determinants of water permeation through aquaporin-1. *Nature* 407, 599–605
- 8 MacKenzie, K.R. *et al.* (1997) A transmembrane helix dimer: structure and implications. *Science* 276, 131–133

- 9 Chang, G. *et al.* (1998) Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* 282, 2220–2226
- 10 Bateman, A. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.* 32 (Database issue), D138–D141
- 11 Fanelli, F. and De Benedetti, P.G. (2005) Computational modeling approaches to structure-function analysis of G protein-coupled receptors. *Chem. Rev.* 105, 3297–3351
- 12 Oliveira, L. *et al.* (2004) Heavier-than-air flying machines are impossible. *FEBS Lett.* 564, 269–273
- 13 Engelman, D.M. *et al.* (2003) Membrane protein folding: beyond the two stage model. *FEBS Lett.* 555, 122–125
- 14 Henderson, R. *et al.* (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* 213, 899–929
- 15 Baldwin, J.M. (1993) The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* 12, 1693–1703
- 16 Kuhlbrandt, W. and Wang, D.N. (1991) Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature* 350, 130–134
- 17 Deisenhofer, J. *et al.* (1995) Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodospseudomonas viridis*. *J. Mol. Biol.* 246, 429–457
- 18 Bowie, J.U. (1997) Helix packing in membrane proteins. *J. Mol. Biol.* 272, 780–789
- 19 Doyle, D.A. *et al.* (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280, 69–77
- 20 Dutzler, R. *et al.* (2002) X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415, 287–294
- 21 Abramson, J. *et al.* (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301, 610–615
- 22 Chen, C.P. *et al.* (2002) Transmembrane helix predictions revisited. *Protein Sci.* 11, 2774–2791
- 23 Unwin, P.N. and Henderson, R. (1975) Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* 94, 425–440
- 24 Rhee, K.H. *et al.* (1998) Three-dimensional structure of the plant photosystem II reaction centre at 8 Å resolution. *Nature* 396, 283–286
- 25 Unger, V.M. *et al.* (1999) Three-dimensional structure of a recombinant gap junction membrane channel. *Science* 283, 1176–1180
- 26 Breyton, C. *et al.* (2002) Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* 418, 662–665
- 27 Ubarretxena-Belandia, I. *et al.* (2003) Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* 22, 6175–6181
- 28 Fujiyoshi, Y. (1998) The structural study of membrane proteins by electron crystallography. *Adv. Biophys.* 35, 25–80
- 29 Ma, C. and Chang, G. (2004) Structure of the multidrug resistance efflux transporter EmrE from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2852–2857
- 30 Jiang, Y. *et al.* (2003) X-ray structure of a voltage-dependent K⁺ channel. *Nature* 423, 33–41
- 31 Long, S.B. *et al.* (2005) Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309, 897–903
- 32 Hasler, L. *et al.* (1998) 2D crystallization of membrane proteins: rationales and examples. *J. Struct. Biol.* 121, 162–171
- 33 Henderson, R. *et al.* (1986) Structure of purple membrane from *Halobacterium halobium*: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* 19, 147–178
- 34 Amos, L.A. *et al.* (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Prog. Biophys. Mol. Biol.* 39, 183–231
- 35 Walz, T. *et al.* (1997) The three-dimensional structure of aquaporin-1. *Nature* 387, 624–627
- 36 Mitsuoka, K. *et al.* (1999) The structure of aquaporin-1 at 4.5-Å resolution reveals short α -helices in the center of the monomer. *J. Struct. Biol.* 128, 34–43
- 37 Heymann, J.B. and Engel, A. (2000) Structural clues in the sequences of the aquaporins. *J. Mol. Biol.* 295, 1039–1053
- 38 de Groot, B.L. *et al.* (2000) The fold of human aquaporin 1. *J. Mol. Biol.* 300, 987–994

- 39 Fleishman, S.J. *et al.* (2004) An automatic method for predicting the structures of transmembrane proteins using cryo-EM and evolutionary data. *Biophys. J.* 87, 3448–3459
- 40 Baldwin, J.M. *et al.* (1997) An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* 272, 144–164
- 41 Unger, V.M. *et al.* (1997) Arrangement of rhodopsin transmembrane α -helices. *Nature* 389, 203–206
- 42 Latronico, A.C. *et al.* (1998) A unique constitutively activating mutation in third transmembrane helix of luteinizing hormone receptor causes sporadic male gonadotropin-independent precocious puberty. *J. Clin. Endocrinol. Metab.* 83, 2435–2440
- 43 Scheer, A. *et al.* (2000) Mutational analysis of the highly conserved arginine within the Glu/Asp-Arg-Tyr motif of the α_{1b} -adrenergic receptor: effects on receptor isomerization and activation. *Mol. Pharmacol.* 57, 219–231
- 44 Palczewski, K. *et al.* (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289, 739–745
- 45 Fleishman, S.J. *et al.* (2004) An evolutionarily conserved network of amino acids mediates gating in voltage-dependent potassium channels. *J. Mol. Biol.* 340, 307–318
- 46 Harris, A.L. (2001) Emerging issues of connexin channels: biophysics fills the gap. *Q. Rev. Biophys.* 34, 325–472
- 47 Fleishman, S.J. *et al.* (2004) A C- α model for the transmembrane α -helices of gap-junction intercellular channels. *Mol. Cell* 15, 879–888
- 48 Kaback, H.R. *et al.* (2001) The kamikaze approach to membrane transport. *Nat. Rev. Mol. Cell Biol.* 2, 610–620
- 49 Karlin, A. (1993) Structure of nicotinic acetylcholine receptors. *Curr. Opin. Neurobiol.* 3, 299–309
- 50 Lemmon, M.A. *et al.* (1992) Sequence specificity in the dimerization of transmembrane α -helices. *Biochemistry* 31, 12719–12725
- 51 Karlin, A. and Akabas, M.H. (1998) Substituted-cysteine accessibility method. *Methods Enzymol.* 293, 123–145
- 52 Kwaw, I. *et al.* (2000) Thiol cross-linking of cytoplasmic loops in the lactose permease of *Escherichia coli*. *Biochemistry* 39, 3134–3140
- 53 Sorgen, P.L. *et al.* (2002) An approach to membrane protein structure without crystals. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14037–14040
- 54 Torres, J. *et al.* (2001) Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic $^{13}\text{C}=\text{^{18}}\text{O}$ group as a novel infrared probe. *Biopolymers* 59, 396–401
- 55 Furthmayr, H. and Marchesi, V.T. (1976) Subunit structure of human erythrocyte glycophorin A. *Biochemistry* 15, 1137–1144
- 56 Russ, W.P. and Engelman, D.M. (2000) The GxxxG motif: a framework for transmembrane helix–helix association. *J. Mol. Biol.* 296, 911–919
- 57 Senes, A. *et al.* (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β -branched residues at neighboring positions. *J. Mol. Biol.* 296, 921–936
- 58 Sternberg, M.J. and Gullick, W.J. (1990) A sequence motif in the transmembrane region of growth factor receptors with tyrosine kinase activity mediates dimerization. *Protein Eng.* 3, 245–248
- 59 Treutlein, H.R. *et al.* (1992) The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry* 31, 12726–12732
- 60 Adams, P.D. *et al.* (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* 26, 257–261
- 61 Fleishman, S.J. and Ben-Tal, N. (2002) A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane α -helices. *J. Mol. Biol.* 321, 363–378
- 62 Fleishman, S.J. *et al.* (2002) A putative activation switch in the transmembrane domain of erbB2. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15937–15940
- 63 Mendrola, J.M. *et al.* (2002) The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J. Biol. Chem.* 277, 4704–4712
- 64 Abramson, J. *et al.* (2003) The lactose permease of *Escherichia coli*: overall structure, the sugar-binding site and the alternating access model for transport. *FEBS Lett.* 555, 96–101
- 65 Yohannan, S. *et al.* (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* 101, 959–963
- 66 Glaser, F. *et al.* (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19, 163–164
- 67 Beuming, T. and Weinstein, H. (2005) Modeling membrane proteins based on low-resolution electron microscopy maps: a template for the TM domains of the oxalate transporter OxIT. *Protein Eng. Des. Sel.* 18, 119–125
- 68 Schueler-Furman, O. *et al.* (2005) Progress in modeling of protein structures and interactions. *Science* 310, 638–642
- 69 Gobel, U. *et al.* (1994) Correlated mutations and residue contacts in proteins. *Proteins* 18, 309–317

Elsevier.com – Dynamic New Site Links Scientists to New Research & Thinking

Elsevier.com has had a makeover, inside and out. Designed for scientists' information needs, the new site, launched in January, is powered by the latest technology with customer-focused navigation and an intuitive architecture for an improved user experience and greater productivity.

Elsevier.com's easy-to-use navigational tools and structure connect scientists with vital information – all from one entry point. Users can perform rapid and precise searches with our advanced search functionality, using the FAST technology of Scirus.com, the free science search engine. For example, users can define their searches by any number of criteria to pinpoint information and resources. Search by a specific author or editor, book publication date, subject area – life sciences, health sciences, physical sciences and social sciences – or by product type. Elsevier's portfolio includes more than 1800 Elsevier journals, 2200 new books per year, and a range of innovative electronic products. In addition, tailored content for authors, editors and librarians provides up-to-the-minute news, updates on functionality and new products, e-alerts and services, as well as relevant events.

Elsevier is proud to be a partner with the scientific and medical community. Find out more about who we are in the About section: our mission and values and how we support the STM community worldwide through partnerships with libraries and other publishers, and grant awards from The Elsevier Foundation.

As a world-leading publisher of scientific, technical and health information, Elsevier is dedicated to linking researchers and professionals to the best thinking in their fields. We offer the widest and deepest coverage in a range of media types to enhance cross-pollination of information, breakthroughs in research and discovery, and the sharing and preservation of knowledge. Visit us at Elsevier.com.

Elsevier. Building Insights. Breaking Boundaries.