Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Structural genomics plucks high-hanging membrane proteins

Edda Kloppmann[1,2], Marco Punta[3] and Burkhard Rost[1,2,4]

Recent years have seen the establishment of structural genomics centers that explicitly target integral membrane proteins. Here, we review the advances in targeting these extremely high-hanging fruits of structural biology in high-throughput mode. We observe that the experimental determination of high-resolution structures of integral membrane proteins is increasingly successful both in terms of getting structures and of covering important protein families, for example, from Pfam. Structural genomics has begun to contribute significantly toward this progress. An important component of this contribution is the set up of robotic pipelines that generate a wealth of experimental data for membrane proteins. We argue that prediction methods for the identification of membrane regions and for the comparison of membrane proteins largely suffice to meet the challenges of target selection for structural genomics of membrane proteins. In contrast, we need better methods to prioritize the most promising members in a family of closely related proteins and to annotate protein function from sequence and structure in absence of homology.

Addresses
[1] Department of Bioinformatics and Computational Biology, Technical University Munich, Boltzmannstr. 3, 85748 Garching/Munich, Germany
[2] New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027, USA
[3] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK
[4] Institute of Advanced Study (TUM-IAS), Lichtenbergstr. 2a, 85748 Garching/Munich, Germany

Corresponding author: Kloppmann, Edda (kloppmann@rostlab.org)

Current Opinion in Structural Biology 2012, **22**:326–332

This review comes from a themed issue on
Sequences and topology
Edited by Christine Orengo and James Whisstock

Available online 21st May 2012

0959-440X/$ – see front matter

http://dx.doi.org/10.1016/j.sbi.2012.05.002

Structure determination for IMPs remains a major challenge

Integral membrane proteins (IMPs) link the cell to its environment and the constituencies of cell compartments to their functional pathways. They are importantly involved in ion transport, signaling and cell adhesion [1]. Owing to their central gate-keeping role, their functional and structural characterization is crucial to molecular and medical biology. Most common are alpha-helical IMPs (alpha-IMPs) that span the membrane by inserting long (15–40 residues) alpha helices [2]. These alpha-IMPs constitute about 25% of all proteins, a fraction that is surprisingly constant and varies more within the domains of life than between them [3,4].

In contrast to soluble proteins that interact with the aqueous solution of the cell, IMPs interact with at least three different environments: the hydrophobic core of the membrane, the charged head groups of lipids and the aqueous solution on both sides of the membrane. The fact that IMPs have been shaped by evolution to be stable under these particular conditions poses one of the enormous challenges for their experimental structure determination. Protein flexibility and low expression levels further add to the problems. This helps explaining why the class of proteins that is targeted by most drugs is structurally the most under-represented: alpha-IMPs constitute less than 2% of all proteins in the Protein Data Bank (PDB) [5–8]. Solving a single high-resolution structure for an alpha-IMP may take years of meticulous optimization of protein constructs and experimental conditions. In this review of the progress in structure determination for alpha-IMPs, we focus on polytopic alpha-IMPs, that is, IMPs with at least two transmembrane helices. These constitute about 70% of all alpha-IMPs [3,9].

Structural genomics for IMPs?

At the beginning of the millennium Structural Genomics (SG), the marriage between the world of sequences and the world of structures took off [10,11]. One important objective was to increase the impact of protein structure on biology, another to reduce the costs of structure determination through robotics (high-throughput cloning, expression and purification), and yet another to increase the structural coverage of sequence space by picking targets with a high impact upon comparative modeling [12,13].

Initially, SG initiatives would target so-called 'low-hanging fruits', that is, proteins whose structures were perceived to be easier to solve. Large SG centers have consistently determined several structures every day over the last five years. But the low-hanging fruits story ended with the focus of several new centers shifting to membrane proteins. Only proteins with long natively disordered regions seem to be even more elusive to structure determination [14].

The first SG center specializing in membrane proteins, the European Membrane Protein Consortium (E-MeP),

Table 1

SG initiatives that target IMPs. We report all SG centers explicitly targeting IMPs and, separately, SG centers which, although mainly focusing on soluble proteins, have contributed IMP structures to the PDB (ATCG3D, MCSG, RSGI and SGC).

Targeting	Center initial	Center name	IMPs in PDB
IMPs	CSMP[a]	Center for Structures of Membrane Proteins	15
	E-MeP	European Membrane Protein Consortium	11[b]
	GPCR network[a]		2
	MPID[a]	Center for Membrane Proteins of Infectious Disease	
	MPSbyNMR[a]	Membrane Protein Structures by Solution NMR	1
	MPSBC[a]	Membrane Protein Structural Biology Consortium	
	MPSi	Membrane Protein Structure Initiative	
	NYCOMPS[a]	New York Consortium on Membrane Protein Structure	26
	TEMIMPS[a]	Transcontinental EM Initiative for Membrane Protein Structure	
	TMPC[a]	Transmembrane Protein Center	
	TransportPDB[a]	Center for the X-ray Structure Determination of Human Transporters	
All proteins	ATCG3D[a]	Accelerated Technologies Center for Gene to 3D Structure	12
	MCSG[a]	Midwest Center for Structural Genomics	1
	RSGI	RIKEN Structural Genomics/Proteomics Initiative	5
	SGC	Structural Genomics Consortium	1

[a] Consortia part of PSI (Protein Structure Initiative) at NIH-NIGMS.
[b] Structures not classified as SG in PDB.

was launched in 2004. In 2005, two pilot projects from the Protein Structure Initiative (PSI) at the National Institutes of Health — National Institutes of General Medical Sciences (NIH-NIGMS) started operating: the Center for Structures of Membrane Proteins (CSMP) and the New York Consortium on Membrane Protein Structure (NYCOMPS). Today, 11 consortia exclusively target membrane protein structures (Table 1) with some of the centers focusing on proteins from specific families or related to particular processes, for example, the G-protein coupled receptor (GPCR) network and the Center for Membrane Proteins of Infectious Diseases (MPID).

Progress in structure determination for membrane proteins

Despite the experimental challenges, the number of new experimental IMP structures has been increasing remarkably [8] with 170 new polytopic alpha-IMP structures added to the PDB in 2011 as compared to 31 in 2001. It should be stressed however that many of these new entries were either similar or identical in sequence to IMPs previously in the PDB. Solving structures for new members of structurally covered families is often extremely useful, since homology models cannot always provide the level of resolution that may be required for downstream functional studies. The most obvious example is represented by targets of outstanding biomedical interest such as GPCRs for which two additional structures have recently been solved by the GPCR network [15,16••].

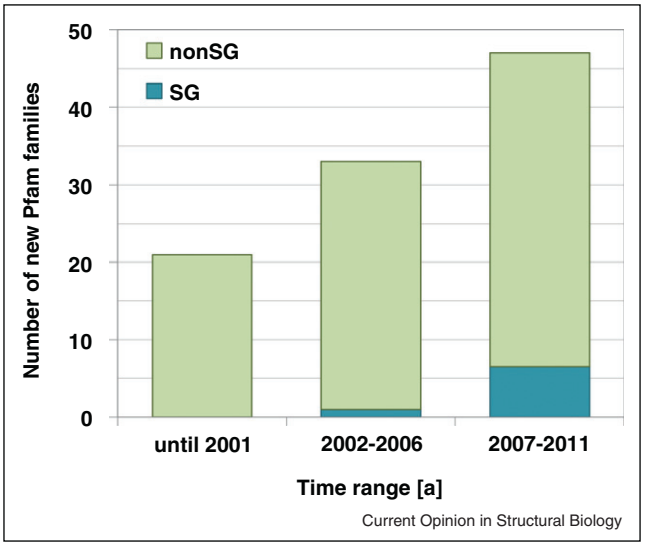
Our main objective in this opinion, however, is to assess progress in determining novel IMP structures. To this end, we use the family classification provided by the Pfam database (release 26.0 [17], see Figure 1 caption). In general, proteins found in different Pfam families are either evolutionary unrelated or only remotely related (a number of exceptions occur in so-called clans, where some families may be more closely related). In January 2012, all polytopic alpha-IMP structures in the PDB cover 101 Pfam families (Table 2), almost half of which were added over the last five years (Figure 1). These data show that, despite clear progress, determining IMP structures for new Pfam families is still extremely challenging.

After the first years, much of the success in structure determination of IMPs has been achieved in Bacteria (Figure 2). However, significant contributions also come from eukaryotic membrane protein structures. About 25% of the 101 Pfam families have structures for IMPs from more than one kingdom. In particular, we looked at structural coverage of polytopic alpha-IMP families in *Homo sapiens* (Figure 3). Note that 7% of human IMPs do not map to any Pfam family (Figure 3(a)). We see that 10% of Pfam human IMP families have a structural representative, with one fifth of those having a structure for a human protein (Figure 3(b)).

SG contributes already significantly to IMP structure determination

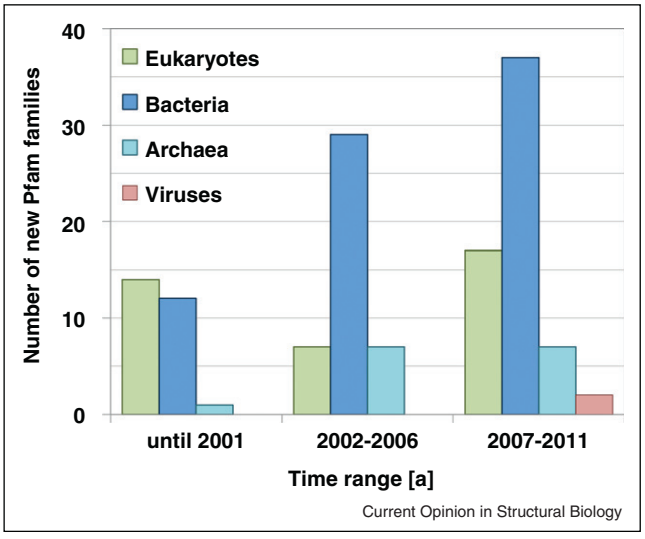
Large SG consortia contributed over 13% of all structures to the PDB [18,19]. Since IMP structures have not been on the list of SG for long, the impact on these is less impressive, albeit still significant (Table 2). In terms of novel leverage, SG has contributed the first structure for 7.5 Pfam families, or 7% of IMP families that are currently structurally covered (Figure 1, Table 2). Considering that SG work on IMPs started in earnest only in 2004–2005 and that the number of consortia involved has increased only very recently, this can be already considered an important contribution. Furthermore, Figure 1 shows that

Figure 1



Novel Pfam families covered by SG and non-SG alpha-IMP structures. We consider all Pfam families that have at least one structural representative in the PDB. Non-structural genomics (non-SG) and structural genomics (SG) contributions are shown in green and blue, respectively. We considered 1035 PDB IDs for polytopic alpha-IMP structures. These are all IMP structures found in the OPM (Orientations of Proteins in Membranes, a database of membrane protein structures [47]), or PDBTM (Protein Data Bank of Transmembrane Proteins [7]) databases, that have at least one representative in UniProt [48] and can be mapped to at least one Pfam family (Release 26.0 [17]). According to our definition, an IMP maps to a Pfam family when at least 12 transmembrane residues (annotation from PDBTM) align to the profile hidden Markov model of the family (using alignment coordinates). We found 107 such Pfam families. From this list, we excluded 6 Pfam families. These include two families that represent N-terminal or C-terminal soluble extensions of a transmembrane domain, one case of a dubious Pfam match, one case where the classification, but not the annotation in OPM is wrong, one case where the annotation in PDBTM is wrong and one case where we considered a bitopic IMP chain of an IMP structure with one polytopic and two bitopic chains. Therefore, here we consider 101 Pfam families that align to at least one structure of a polytopic IMP. *Note*: 159 IMPs of the initial set that could not be mapped to either a UniProt sequence or a Pfam family have been excluded from the analysis. Note also: the Pfam family 'formate/nitrite transporter' (PF01226) was covered by one SG and one non-SG structure in December 2009. Both structures were deposited in the PDB before the coordinates of the other were released, that is, became publicly known. We counted PF01226 as half a family for both SG and non-SG.

Figure 2



Taxonomic distribution of IMP structures covering novel Pfam families. We show the number of Pfam families covered by structures from Eukaryotes, Bacteria, Archaea and Viruses (in green, blue, light blue and red, respectively). The data are shown for three different time spans. For each combination of family and kingdom we consider the release date of the first structure solved for this family. For example, a family with several bacterial protein structures is counted in the time range during which the first structure was solved. On the other hand, Pfam families with protein structures from more than one kingdom are counted for each kingdom. For example, a Pfam family with a eukaryotic and bacterial protein structure is counted twice, that is, once for each kingdom. Seventeen of the 101 Pfam families are counted for 2 kingdoms and 8 families have eukaryotic, bacterial and archaeal protein structures. Mapping from PDB to Pfam as described in Figure 1.

the contribution of both SG and non-SG to novel leverage is increasing with time.

The Pfam families that have been covered first by SG are: (1) CorA-like Mg^{2+} transporter protein (PF01544) [20], (2) the formate/nitrite transporter (PF01226, covered at the same time by non-SG) [21], (3) the urea transporter (PF03253) [22], (4) the amino acid permease (PF13520) [23], (5) the C4-dicarboxylate transporter/malic acid transport protein (PF03595, including the anion channel TehA) [24••], (6) the phosphotransferase system, EIIC (PF02378) [25], (7) the cation transport protein, TrkH (PF02386) [26], and finally, (8) the etoposide-induced protein 2.4 (EI24) (PF07264, PDB ID 3tx3).

In terms of the estimated number of models that can be built by comparative modeling from each new structure [27], SG is lagging behind (5671 sequences/Pfam family for SG versus 13 591 sequences/Pfam family for non-SG). Interestingly, however, the median coverage of 3031 sequences for SG versus 3364 for non-SG suggests that once the bias due to the some large families is removed, the contribution to coverage is not very different.

SG finds the needle in the haystack

Expression, purification, stability, and crystallization outcomes can differ radically between very similar proteins even when subjected to the same experimental protocols executed by robots (Figure 4(a)). As a consequence, targeting several family members can increase the likelihood of solving a structure for that family. Structural knowledge acquired on one member of the family can then be at least partially transferred to other members with the help of comparative modeling [27,28,29•]. Thus,

Table 2

IMP structure statistics. 'All' is the sum of SG and non-SG depositions. In the first row 'PDB IDs', we report all PDB IDs of polytopic alpha-IMPs that we considered (see Figure 1). Row 'Pfam IDs', shows the number of Pfam families that overlap with the transmembrane region of the structures in row one. Row three 'Pfam IDs only in' reports unique Pfam family contributions (first structure deposited for that family) for SG and non-SG. Finally, row four 'Sequences in Pfam families' shows the total number of sequences belonging to the Pfam families in row four (full alignments).

	All	SG		Non-SG	
PDB IDs	1035	60	6%	975	94%
Pfam IDs	101	16	16%	96	95%
Pfam IDs only in		7.5[a]	7%	93.5[a]	93%
Sequences in Pfam families	$\sim 1.32 \times 10^6$	$\sim 45\,000$[a]	3%	1.28×10^{6}[a]	97%

[a] One Pfam family, 'formate/nitrite transporter' (PF01226), was covered by an SG and non-SG structure at the same time, both structures were deposited in the PDB before the first was released. This family is counted with a factor of 0.5 for SG and non-SG.

one experimental structure enables to study the basic sequence–structure–function relationships within the family. TehA, which is a homologue of the plant SLAC1 anion channel, is one remarkable recent example in which the structure of a *H. influenzae* protein (one of the highest resolution IMPs (1.2 Å) in the PDB, contributed by SG) explained important aspects about the mechanism of closing stomata in plant leaves [24••,30]. In the pictorial view of the channel crystal structure reported in Figure 4(b), we highlight the highly conserved Phe262, which is implicated in the gating mechanism.

More in general, recalcitrance of individual IMPs to experimental structure determination makes the pan-genomic, family-level approach of SG particularly relevant to their study. This is testified to by the fact that a similar approach has been adopted by a number of non-SG laboratories that target easier to handle bacterial IMPs instead of their functionally often better characterized eukaryotic relatives. A classic example is the structure of the bacterial KcsA potassium channel from *S. lividans*, the first high-resolution ion channel structure [31].

SG contributes more than 'just' structures

The ultimate and most important objective of structural genomics is the determination of protein structures. However, all consortia have generated much additional value over the last decade. For instance, the large consortia cloned and tried to express at least one representative for most structurally uncovered short bacterial proteins [18,32].

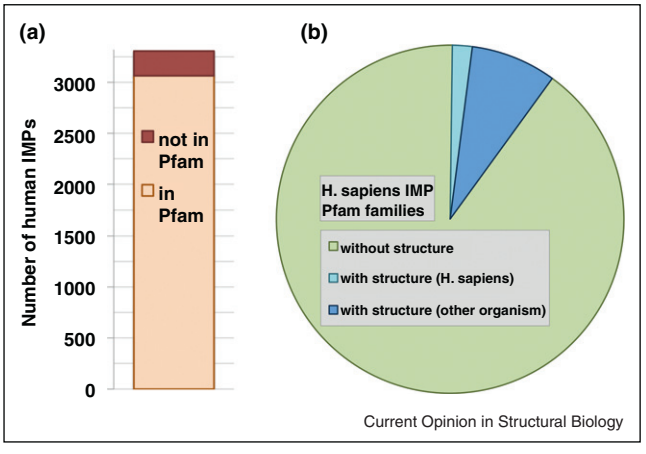
SG consortia targeting IMPs have created pipelines that hold important value [33•]. Over 15 000 IMPs have entered the SG pipelines (Figure 5). Notwithstanding high attrition rates at essentially each of the experimental stages, SG has managed to express and purify several thousands of IMPs. As all protocols, reagents and proteins are made available, there is a large added potential value to SG that reaches substantially beyond solved structures. Targets that failed in the late stages of an SG pipeline may still be successfully pursued by experts in the field, when time is taken to optimize experimental conditions.

Computational challenges for SG of IMPs

Methods from bioinformatics and computational biology are heavily used in SG, primarily for target selection and downstream analysis of solved structures, including reconstruction of phylogenies, homology modeling and function annotation. Starting from an initial pool of protein sequences, target selection for IMPs typically proceeds as follows:

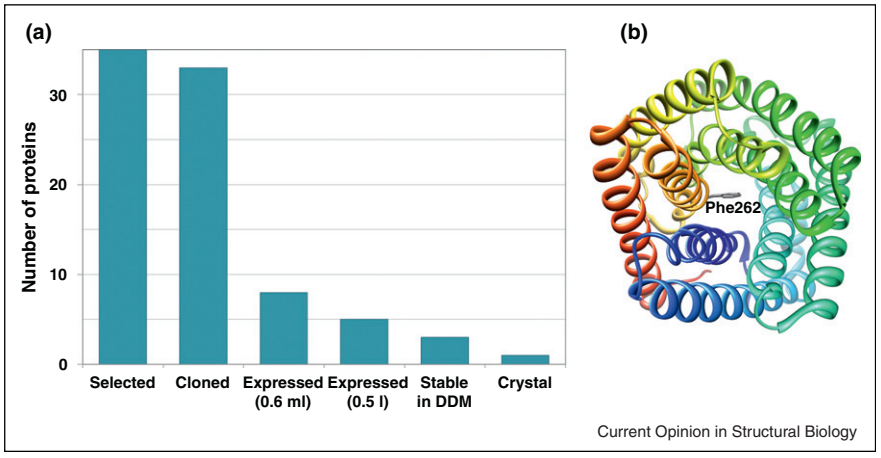
(1) Alpha-IMPs are separated from other proteins by methods that predict transmembrane helices

Figure 3



Human IMPs: Pfam families and PDB structures. **(a)** Mapping human IMPs to Pfam families. Three thousand three hundred and five polytopic alpha-IMPs were extracted from the 20 247 sequences part of the SwissProt *Homo sapiens* proteome (UniProt release February 22, 2012) using PolyPhobius [49]. Assignment of proteins to Pfam families was done as described in Figure 1 using the transmembrane assignment of PolyPhobius. Three thousand and sixty-three IMPs can be mapped to a Pfam family (orange); 242 IMPs fall outside of the current Pfam collection of families (red). **(b)** Human IMP Pfam families covered by structure. We show human IMP Pfam families with no structural representative (green) and with at least one structural representative (blue: representative is a human protein; light blue: representative is not a human protein).

Figure 4



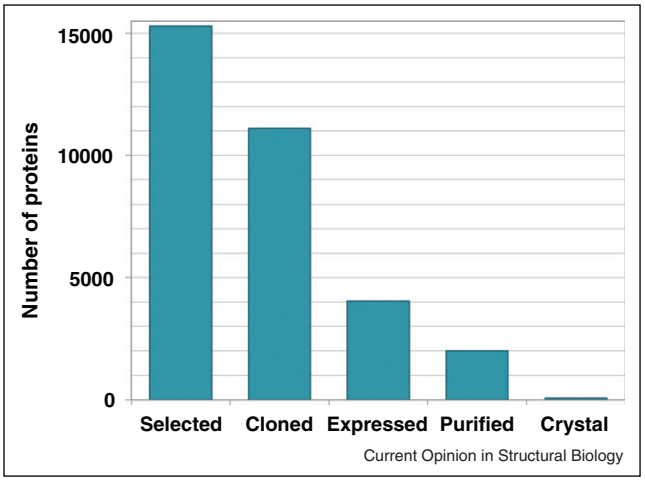
The TehA protein family. **(a)** Homologous proteins behave differentially. The NYCOMPS SG center cloned 35 homologous prokaryotic proteins belonging to the TehA family. Of the 35 homologous proteins experimentally pursued, 33 could be cloned, 8 and 5 expressed (small and large scale, respectively), and only one yielded a diffracting crystal and finally a high-resolution structure. Note in particular the dramatic attrition rate in the number of successfully cloned to successfully expressed proteins. Targets are cloned by ligation free cloning and C-terminal fusion expression vector. Expression and purification are assessed by Coomassie Blue stained SDS-PAGE gels and stability in the DMM detergent is determined by size exclusion chromatography [35]. Structures of the TehA family representative from *H. influenzae* have been solved and deposited in the PDB [24••]. **(b)** Structure of the SLAC1 homolog TehA. The anion channel structure is shown as seen from the periplasm (PDB ID: 3m71, ribbon representation). The highly conserved Phe262 occluding the ion permeation pathway is shown explicitly. The figure is created using Chimera [50].

(TMHs) [34]. Consortia may choose to target proteins with at least two [9,35] or more [36,37] transmembrane helices. Although their performance has often been overestimated, methods that predict TMHs have largely matured. Open challenges include the correct identification of re-entrant elements, and more generally the identification of non-helical regions inserted into the membrane bilayer [38,39].

(2) IMPs are then clustered into families such that one experimental structure in that family suffices to model structures for all family members using standard alignment methods. There is evidence that this is justified [40]. Recently, new tools have been developed that seem to improve alignments for membrane proteins [41–43]. This is most relevant for building better comparative models within a family covered by an experimental structure.

(3) Finally, targets promising a higher success rate are selected. Until now, this step largely filtered out proteins dominated by long regions that are natively disordered or used prior knowledge about success in, for example, protein expression. Some methods try to predict how feasible a globular protein is for some aspects of structure determination pipelines [44–46]. Lack of experimental data has so far hindered the development of such a method for alpha-IMPs. Given the over 15 000 IMPs in SG pipelines (Figure 5), it might be time to revisit this task.

Figure 5



Statistics for IMP structural genomics protein production pipelines. Depicted is the number of IMPs that were successful at different stages in the experimental pipelines. Data were extracted from TargetDB [51] in January 2012 for nine membrane protein structural genomics consortia: CSMP, GPCR network, MPID, MPSbyNMR, MPSCB, NYCOMPS, TEMIPS, TMPC and TransportPDB. For the NMR consortium (MPSbyNMR) we do not report data for the steps following purification.

Overall, we argue that for the purpose of SG, tasks (1) and (2) are sufficiently addressed by existing methods. The major challenge is task (3), that is, the prediction of which protein in a family is the most promising candidate for getting a structure from a pipeline. This task also remains essentially unsolved for globular proteins, but for membrane proteins advances are more urgently needed because the resources that still have to be

invested after the pipelines complete their work are considerable.

Another area in which further progress would be much welcome is prediction of functional attributes in proteins from sequence and structure. This information can be used in target selection to identify for example interaction partners (proteins or small molecules) that may be important for protein stability. In the same way, more accurate methods that predict protein function from sequence without using solely homology transfer would be extremely helpful. Finally, prediction from structure is important in cases in which a structure is solved for a functionally not annotated family.

Conclusion

Structural genomics has succeeded in plucking the high-hanging fruits of integral transmembrane proteins. Projecting from the recent progress, there is little doubt that structural genomics will play an important role toward tackling this important challenge in structural biology. Computational biology has contributed importantly to the recent experimental success, through methods that predict membrane regions and methods that align membrane proteins. Despite several challenges to those methods, the major challenge seems to be to develop better tools for prioritizing targets and for annotating function in the absence of homology information.

Acknowledgements

Thanks to Wayne Hendrickson (Columbia University), James Love (Albert Einstein College of Medicine), Brian Kloss and Renato Bruni (New York Structural Biology Center and New York Consortium on Membrane Protein Structure), and many others at the New York Consortium on Membrane Protein Structure for their important help and support. Thanks to Jaina Mistry (Wellcome Trust Sanger Institute) for providing the mapping between PDB entries and Pfam families. Thanks to Tim Karl and Laszlo Kajan (TUM) for invaluable help with hardware and software; to Marlena Drabik (TUM) and Ed Johnson (Columbia University) for their administrative support. This work was supported by a grant from the Alexander von Humboldt foundation through the German Federal Ministry for Education and Research (BMBF) and by the grant U54 GM095315 to the New York Consortium on Membrane Protein Structure (NYCOMPS) from the Protein Structure Initiative (PSI) of the National Institutes of Health (NIH). Last, not least, thanks to all those who deposit their experimental data in public databases, and to those who maintain these databases.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. von Heijne G: **The membrane protein universe: what's out there and why bother?** *J Intern Med* 2007, **261**:543-557.

2. von Heijne G: **Membrane-protein topology**. *Nat Rev Mol Cell Biol* 2006, **7**:909-918.

3. Liu J, Rost B: **Comparing function and structure between entire proteomes**. *Protein Sci* 2001, **10**:1970-1979.

4. Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L: **Prediction of the human membrane proteome**. *Proteomics* 2010, **10**:1141-1149.

5. Overington JP, Al-Lazikani B, Hopkins AL: **How many drug targets are there?** *Nat Rev Drug Discov* 2006, **5**:993-996.

6. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucleic Acids Res* 2000, **28**:235-242.

7. Tusnady GE, Dosztanyi Z, Simon I: **PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank**. *Nucleic Acids Res* 2005, **33**:D275-D278.

8. White SH: **Biophysical dissection of membrane proteins**. *Nature* 2009, **459**:344-346.

9. Punta M, Love J, Handelman S, Hunt JF, Shapiro L, Hendrickson WA, Rost B: **Structural genomics target selection for the New York consortium on membrane protein structure**. *J Struct Funct Genomics* 2009, **10**:255-268.

10. Rost B: **Marrying structure and genomics**. *Structure* 1998, **6**:259-263.

11. Montelione GT, Anderson S: **Structural genomics: keystone for a Human Proteome Project**. *Nat Struct Biol* 1999, **6**:11-12.

12. Levitt M: **Growth of novel protein structural data**. *Proc Natl Acad Sci U S A* 2007, **104**:3183-3188.

13. Liu J, Montelione GT, Rost B: **Novel leverage of structural genomics**. *Nat Biotechnol* 2007, **25**:849-851.

14. Uversky VN, Dunker AK: **Understanding protein non-folding**. *Biochim Biophys Acta* 2010, **1804**:1231-1264.

15. Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC: **Structure of an agonist-bound human A2A adenosine receptor**. *Science* 2011, **332**:322-327.

16. •• Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW *et al.*: **Structure of the human histamine H1 receptor complex with doxepin**. *Nature* 2011, **475**:65-70.
Yet another important GPCR structure, this time one related to allergies and inflammation processes. Although we have just started to crack the GPCR code, the GPCR structural genomics network promises to deliver the access keys for these most important biomedical targets.

17. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al.*: **The Pfam protein families database**. *Nucleic Acids Res* 2012, **40**:D290-D301.

18. Nair R, Liu J, Soong T-t, Acton TB, Everett J, Kouranov A, Fiser A, Godzik A, Jaroszewski L, Orengo C *et al.*: **Structural genomics is the largest contributor of novel structural leverage**. *J Struct Funct Genomics* 2009, **10**:181-191.

19. Burley SK, Joachimiak A, Montelione GT, Wilson IA: **Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI production centers**. *Structure* 2008, **16**:5-11.

20. Lunin VV, Dobrovetsky E, Khutoreskaya G, Zhang R, Joachimiak A, Doyle DA, Bochkarev A, Maguire ME, Edwards AM, Koth CM: **Crystal structure of the CorA Mg^{2+} transporter**. *Nature* 2006, **440**:833-837.

21. Waight AB, Love J, Wang DN: **Structure and mechanism of a pentameric formate channel**. *Nat Struct Mol Biol* 2010, **17**:31-37.

22. Levin EJ, Quick M, Zhou M: **Crystal structure of a bacterial homologue of the kidney urea transporter**. *Nature* 2009, **462**:757-761.

23. Shaffer PL, Goehring A, Shankaranarayanan A, Gouaux E: **Structure and mechanism of a Na^+-independent amino acid transporter**. *Science* 2009, **325**:1010-1014.

24. •• Chen YH, Hu L, Punta M, Bruni R, Hillerich B, Kloss B, Rost B, Love J, Siegelbaum SA, Hendrickson WA: **Homologue structure of the SLAC1 anion channel for closing stomata in leaves**. *Nature* 2010, **467**:1074-1080.
From the NYCOMPS structural genomics protein production pipeline into crystallization trials, the SLAC1 homolog from *H. influenzae* yielded a 1.20 Å resolution structure along with structures for several mutant variants. The first structural representative of the TehA family showed that not only structural genomics can deliver structures for high-hanging fruits, it can also solve them at a resolution that is competitive with the best one attainable for globular soluble proteins.

25. Cao Y, Jin X, Levin EJ, Huang H, Zong Y, Quick M, Weng J, Pan Y, Love J, Punta M *et al.*: **Crystal structure of a phosphorylation-coupled saccharide transporter**. *Nature* 2011, **473**:50-54.

26. Cao Y, Jin X, Huang H, Derebe MG, Levin EJ, Kabaleeswaran V, Pan Y, Punta M, Love J, Weng J *et al.*: **Crystal structure of a potassium ion transporter, TrkH**. *Nature* 2011, **471**:336-340.

27. Arnold K, Kiefer F, Kopp J, Battey JN, Podvinec M, Westbrook JD, Berman HM, Bordoli L, Schwede T: **The Protein Model Portal**. *J Struct Funct Genomics* 2009, **10**:1-8.

28. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC *et al.*: **ModBase, a database of annotated comparative protein structure models, and associated resources**. *Nucleic Acids Res* 2011, **39**:D465-D474.

29. • Faraldo-Gomez JD, Forrest LR: **Modeling and simulation of ion-coupled and ATP-driven membrane proteins**. *Curr Opin Struct Biol* 2011, **21**:173-179.

An interesting review that discusses how, starting from structure, good computational studies can advance the understanding of the biology of integral membrane proteins.

30. Thomine S, Barbier-Brygoo H: **Structural biology: a peep through anion channels**. *Nature* 2010, **467**:1058-1059.

31. Zhou P, Lugovskoy AA, McCarty JS, Li P, Wagner G: **Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45**. *Proc Natl Acad Sci U S A* 2001, **98**:6051-6055.

32. Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C: **PSI-2: structural genomics to cover protein domain family space**. *Structure* 2009, **17**:869-881.

33. • Mancia F, Love J: **High throughput platforms for structural genomics of integral membrane proteins**. *Curr Opin Struct Biol* 2011, **21**:517-522.

A perfect companion to the present review, explaining experimental techniques used in structural genomics of integral membrane proteins.

34. Punta M, Forrest LR, Bigelow H, Kernytsky A, Liu J, Rost B: **Membrane protein prediction methods**. *Methods* 2007, **41**:460-474.

35. Love J, Mancia F, Shapiro L, Punta M, Rost B, Girvin M, Wang DN, Zhou M, Hunt JF, Szyperski T *et al.*: **The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins**. *J Struct Funct Genomics* 2010, **11**:191-199.

36. Li M, Hays FA, Roe-Zurz Z, Vuong L, Kelly L, Ho CM, Robbins RM, Pieper U, O'Connell JD III, Miercke LJ *et al.*: **Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening**. *J Mol Biol* 2009, **385**:820-830.

37. Kelly L, Pieper U, Eswar N, Hays FA, Li M, Roe-Zurz Z, Kroetz DL, Giacomini KM, Stroud RM, Sali A: **A survey of integral alpha-helical membrane proteins**. *J Struct Funct Genomics* 2009, **10**:269-280.

38. Viklund H, Granseth E, Elofsson A: **Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes**. *J Mol Biol* 2006, **361**:591-603.

39. Kauko A, Illergard K, Elofsson A: **Coils in the membrane core are conserved and functionally important**. *J Mol Biol* 2008, **380**:170-180.

40. Forrest LR, Tang CL, Honig B: **On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins**. *Biophys J* 2006, **91**:508-517.

41. Park Y, Elsner M, Staritzbichler R, Helms V: **Novel scoring function for modeling structures of oligomers of transmembrane alpha-helices**. *Proteins* 2004, **57**:577-585.

42. Pirovano W, Feenstra KA, Heringa J: **PRALINETM: a strategy for improved multiple alignment of transmembrane proteins**. *Bioinformatics* 2008, **24**:492-497.

43. Hill JR, Kelm S, Shi J, Deane CM: **Environment specific substitution tables improve membrane protein alignment**. *Bioinformatics* 2011, **27**:i15-i23.

44. Price WN II, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J *et al.*: **Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data**. *Nat Biotechnol* 2009, **27**:51-57.

45. Kandaswamy KK, Pugalenthi G, Suganthan PN, Gangal R: **SVMCRYS: an SVM approach for the prediction of protein crystallization propensity from protein sequence**. *Protein Pept Lett* 2010, **17**:423-430.

46. Mizianty MJ, Kurgan L: **Sequence-based prediction of protein crystallization, purification and production propensity**. *Bioinformatics* 2011, **27**:i24-i33.

47. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI: **OPM: orientations of proteins in membranes database**. *Bioinformatics* 2006, **22**:623-625.

48. **Reorganizing the protein space at the Universal Protein Resource (UniProt)**. *Nucleic Acids Res* 2012, **40**:D71-D75.

49. Kall L, Krogh A, Sonnhammer EL: **An HMM posterior decoder for sequence feature prediction that includes homology information**. *Bioinformatics* 2005, **21(Suppl 1)**:i251-i257.

50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera – a visualization system for exploratory research and analysis**. *J Comput Chem* 2004, **25**:1605-1612.

51. Chen L, Oughtred R, Berman HM, Westbrook J: **TargetDB: a target registration database for structural genomics projects**. *Bioinformatics* 2004, **20**:2860-2862.