

Dmitrij Frishman
Editor

Structural Bioinformatics of Membrane Proteins



SpringerWienNewYork

CONTENTS

Evolutionary origins of membrane proteins (*Armen Y. Mulikidjanian, Michael Y. Galperin*) 1

- 1 Introduction 1
- 2 Comparative analysis of F/V-type ATPases: example of function cooption? 3
- 3 Emergence of integral membrane proteins 9
- 4 Emergence of lipid membranes 10
- 5 Scenario for the origin and evolution of membranes and membrane proteins 17

Molecular archeological studies of transmembrane transport systems (*Milton H. Saier Jr, Bin Wang, Eric I. Sun, Madeleine Matias, Ming Ren Yen*) 29

- 1 Introduction 29
- 2 Molecular transport 30
- 3 Techniques to establish homology or the lack of homology 30
- 4 Transport protein diversity 31
- 5 The ABC superfamily 32
- 6 Independent origins for ABC porters 33
- 7 The phosphoenolpyruvate-dependent sugar transporting phosphotransferase system (PTS) 35
- 8 Independent origins for PTS permeases 37
- 9 Reverse (retro)-evolution 38
- 10 Conclusions and perspectives 40

Resource for structure related information on transmembrane proteins (*Gábor E. Tusnády, István Simon*) 45

- 1 Introduction 45
- 2 3D structure resources 46
 - 2.1 Protein Data Bank 46
 - 2.2 Manually curated structure resources of TMPs 47
 - 2.3 TMDET algorithm 48
 - 2.4 PDBTM database 51

Contents

- 2.5 OPM database 52
- 2.6 Modeling protein–lipid assembly 52
- 3 2D structure resources 53
 - 3.1 TOPDB database 54
 - 3.2 TOPDOM database 55
 - 3.3 Prediction methods incorporating experimental results 56

Topology prediction of membrane proteins: how distantly related homologs come into play (*Rita Casadio, Pier Luigi Martelli, Lisa Bartoli, Piero Fariselli*) 61

- 1 Introduction 61
- 2 From membrane protein sequence to topologic models 62
 - 2.1 Datasets of membrane proteins 63
 - 2.2 Scoring the accuracy of different methods 64
 - 2.3 Propensity scales versus machine learning-based methods 65
 - 2.4 Methods for optimizing topologic models 66
 - 2.5 Single sequence versus multiple sequence profile 68
 - 2.6 Prediction of signal peptides and GPI-anchors 69
 - 2.7 More methods are better than one: CINTHIA 69
 - 2.8 A large-scale annotator of the human proteome: the PONGO system 71
- 3 From membrane protein sequence to function and structure 73
 - 3.1 Membrane proteins: how many with known functions and folds? 74
 - 3.1.1 All-alpha membrane proteins 74
 - 3.1.2 All-beta membrane proteins 75
 - 3.2 What do BAR clusters contain? 76
 - 3.2.1 The cluster of glyceroporins 76
 - 3.2.2 The cluster of multidrug transporter proteins (EmrE proteins) 78
 - 3.3.3 The cluster of P-glycoproteins 80

Transmembrane beta-barrel protein structure prediction (*Arlo Randall, Pierre Baldi*) 83

- 1 Introduction 83
 - 1.1 1D feature prediction 84
 - 1.2 β -Contact and tertiary structure prediction 84
- 2 Data 85
 - 2.1 Benchmark sets 85
 - 2.2 Cross-validation 87
 - 2.3 Template construction 87
- 3 Methods 87
 - 3.1 Secondary structure prediction 87

3.1.1	Neural network implementation	87
3.1.2	Two-class prediction (β , -)	88
3.1.3	Three-class prediction (M, C, -)	89
3.2	β -Contact prediction	90
3.3	Tertiary structure prediction	90
3.3.1	Search energy	90
3.3.2	Template usage	91
3.3.3	Move types	92
3.3.4	Conformational search	93
4	Results	93
4.1	Secondary structure prediction results	93
4.1.1	Secondary structure evaluation metrics	93
4.1.2	Results using <i>SetTransfold</i>	94
4.1.3	Results using <i>SetPRED-TMBB</i>	95
4.2	β -Contact prediction results	95
4.2.1	β -Contact evaluation metrics	95
4.2.2	Results using <i>SetTransfold</i>	96
4.2.3	Results using <i>SetPRED-TMBB</i>	96
4.3	Tertiary structure prediction results	96
4.3.1	Tertiary structure evaluation metrics	97
4.3.2	Prediction results	98
4.3.3	Self-consistency results	98
5	Discussion	99

Multiple alignment of transmembrane protein sequences (*Walter Pirovano, Sanne Abeln, K. Anton Feenstra, Jaap Heringa*) 103

1	Introduction	103
2	Factors influencing the alignment of transmembrane proteins	105
2.1	Transmembrane substitution rates	105
2.2	Transmembrane alignment gaps	107
3	Overview of TM MSA methods	107
3.1	TM-aware multiple sequence alignment by the Praline method	108
3.1.1	Profile pre-processing	108
3.1.2	Bipartite alignment scheme	109
3.1.3	Tree-based consistency iteration	110
3.2	Bipartite MSA compared to standard MSA	111
3.3	Comparing PRALINE-TM with non-TM MSA methods	112
4	Benchmarking transmembrane alignments	114
4.1	Defining TM regions	115
5	Applications for TM multiple alignments	116
5.1	Homology searches of TM proteins	117

Contents

- 6 Current bottlenecks 117
- 7 Avenues for improvement 118
- 8 Conclusions 119

Prediction of re-entrant regions and other structural features beyond traditional topology models (*Erik Granseth*) 123

- 1 Introduction 123
- 2 Background 125
 - 2.1 The Z-coordinate as a measure of distance to the membrane 125
- 3 Interface helices 125
 - 3.1 Prediction of interface helices 127
 - 3.2 Prediction of amphipathic membrane anchors 128
- 4 Helical kinks in transmembrane helices 128
 - 4.1 Prediction of helix kinks 129
- 5 Re-entrant regions 129
 - 5.1 Prediction of re-entrant regions 130
 - 5.1.1 TOP-MOD 130
 - 5.1.2 TMloop 131
 - 5.1.3 OCTOPUS 131
 - 5.1.4 MEMSAT-SVM 131
- 6 Prediction of the Z-coordinate 132
- 7 Free energy of membrane insertion ΔG 133
- 8 The frequency of re-entrant regions and interface helices 134
- 9 Summary 135

Dual-topology: one sequence, two topologies (*Erik Granseth*) 137

- 1 Introduction 137
- 2 Background 139
 - 2.1 A brief history of dual-topology research 139
 - 2.2 The difference between dual- and multiple-topology 139
 - 2.3 Topology mapping 139
 - 2.4 Arginines and lysines are important for the topology 140
 - 2.5 Internal structural repeats – evidence of former gene duplication events 140
- 3 Prediction of dual-topology 142
 - 3.1 The small multidrug resistance family: one family, different topologies 142
 - 3.2 The DUF606 family contains fused genes 143
- 4 Examples of membrane proteins with dual- or multiple-topology 144
 - 4.1 MRAP 144
 - 4.2 Ductin 144
 - 4.3 *Hepatitis B* virus L protein 145

- 4.4 Hepatitis C virus protein NS4B 146
 - 4.5 TatA 146
 - 4.6 PrP 147
- 5 Using topology inversion for function 147
 - 5.1 SecE 147
- 6 Using dual-topology as a targeting system 148
 - 6.1 Cytochrome p450-2E1 148
 - 6.2 Epoxide hydrolase 148

Predicting the burial/exposure status of transmembrane residues in helical membrane proteins (*Volkhard Helms, Sikander Hayat, Jennifer Metzger*) 151

- 1 Introduction 151
- 2 Hydrophobicity analysis 154
- 3 Amino acid propensity scales 155
- 4 Methods using sequence conservation 158
- 5 Applications of burial prediction 162

Helix–helix interaction patterns in membrane proteins (*Dieter Langosch, Jana R. Herrmann, Stephanie Unterreitmeier, Angelika Fuchs*) 165

- 1 Introduction 165
- 2 Technical approaches to identify transmembrane helix–helix interfaces 167
- 3 Structure of transmembrane helix–helix interfaces 170
 - 3.1 Amino acid side-chain packing 170
 - 3.2 GxxxG motifs 171
 - 3.3 Hydrogen bonding 173
 - 3.4 Charge–charge interactions 174
 - 3.5 Aromatic interactions 176
- 4 Dynamic TMD–TMD interactions 177

Predicting residue and helix contacts in membrane proteins (*Angelika Fuchs, Andreas Kirschner, Dmitrij Frishman*) 187

- 1 Introduction 187
- 2 Biological background 188
 - 2.1 Diversity of helix–helix contacts in membrane proteins 189
 - 2.2 Frequency of residue contacts in membrane and soluble proteins 190
- 3 Prediction of lipid accessibility 191
 - 3.1 Hydrophobicity-based predictions 191

Contents

- 3.2 Amino acid propensity scales derived from membrane protein sequences and structures 192
- 3.3 Sequence conservation of exposed and buried transmembrane residues 193
- 3.4 Best performing methods in the field of lipid accessibility 193
- 4 Prediction of helix–helix contacts 194
 - 4.1 Co-evolving residues in membrane proteins 194
 - 4.2 Prediction of helix–helix contacts with machine-learning techniques 195
- 5 Prediction of helix interactions 197
- 6 Modeling of membrane proteins with predicted contact information 199

Natural constraints, folding, motion, and structural stability in transmembrane helical proteins (*Susan E. Harrington, Nir Ben-Tal*) 205

- 1 Folding background 205
 - 1.1 Two-stage hypothesis 205
 - 1.2 Translocon-aided folding 206
- 2 Overview of non-interhelical stabilizing forces and natural constraints 206
 - 2.1 Membrane constraints and interactions 206
 - 2.1.1 Hydrophobic mismatch 207
 - 2.1.2 Specific flanking and anchoring interactions with polar headgroups 207
 - 2.1.3 Positive-inside rule 207
 - 2.2 Loop constraints 207
- 3 Interhelical interactions and constraints 208
 - 3.1 Helix–helix packing 208
 - 3.2 Motifs and stabilizing specific interactions 208
 - 3.2.1 Packing motifs 209
 - 3.2.2 Hydrogen bonds 209
 - 3.2.3 Aromatic interactions 209
 - 3.2.4 Salt bridges 209
 - 3.3 The five types of specific stabilizing interhelical interactions considered 209
 - 3.4 Structural hot spots 210
 - 3.5 Experimental data on residue contributions to stabilization 211
 - 3.6 Particularly stabilizing interactions as geometric constraints 212
 - 3.7 Helix pairs revisited 214
 - 3.8 Constraint perspective and underlying rigid-body geometry 214
 - 3.9 Iterative reassembly of full TM helix bundles using interactions of the five types 216
 - 3.10 The sets of the five types of particularly favorable interactions determine the packing of helices in the native structures of a diverse test set 217
 - 3.11 Distribution of particularly stabilizing residues, folding funnels, and the construction of low-energy minima 218

3.12	Cooperativity with packing	219
3.13	Static structures versus ensembles	219
4	Conservation and diversity of determining sets of stabilizing interactions	219
4.1	Conservation and diversity of the determining sets of interactions of bR	221
5	Determining sets, multiple states, and motion	221
5.1	Multiple states and motion in the ErbB family	222
6	Conclusion	225

Prediction of three-dimensional transmembrane helical protein structures (*Patrick Barth*) 231

1	Introduction	231
2	Goal of the chapter	232
3	Methods	232
3.1	De novo membrane protein structure prediction	232
3.1.1	MP topology predictions	234
3.1.2	The first MP structure prediction methods developed during the past decade	234
3.1.3	Solutions to the conformational search problem: folding with predicted constraints and contact predictors	237
3.1.3.1	Folding with predicted constraints	237
3.1.3.2	Contact predictors	239
3.1.4	MP-specific energy functions for decoy discrimination	240
3.2	Sequence-based modeling with experimental constraints	241
3.3	Comparative modeling of MP structures	244
4	Conclusions and future directions	245

GPCRs: past, present, and future (*Bas Vrolijk, Robert P. Bywater, Laerte Oliveira, Gert Vriend*) 251

1	Introduction	251
2	A short history	252
3	GPCR structures	259
3.1	Rhodopsin	259
3.2	Ligand-mediated GPCRs	261
4	From sequence to structure	266
4.1	The conserved cysteine bridge in the extracellular domain	266
4.2	Loop IV-V, cysteine bridges, and ligand binding	266
5	The future	270

List of Contributors	279
----------------------	-----