Ligand Design
for G Protein-coupled Receptors

Edited by
Didier Rognan
Ligand Design
for G Protein-coupled Receptors

Edited by
Didier Rognan
Ligand Design
for G Protein-coupled Receptors

Edited by
Didier Rognan
Contents

Preface  

1 G Protein-coupled Receptors in the Human Genome  

Robert Fredriksson and Helgi B. Schiöth

1.1 Introduction  
1.2 The Adhesion Family  
1.3 The Secretin Family  
1.4 The Frizzled/Taste 2 Family  
1.4.1 The Frizzled Receptor Cluster  
1.4.2 The Taste 2 Receptor Cluster  
1.5 The Glutamate Family  
1.6 The Rhodopsin Family  
1.6.1 The Rhodopsin α-Group  
1.6.1.1 The Prostaglandin Receptor Cluster  
1.6.1.2 The Amine Receptor Cluster  
1.6.1.3 The Opsin Receptor Cluster  
1.6.1.4 The Melatonin Receptor Cluster  
1.6.1.5 The MECA Receptor Cluster  
1.6.1.6 Other Rhodopsin α-Receptors  
1.6.2 Rhodopsin β-Group  
1.6.3 Rhodopsin γ-Group  
1.6.3.1 The SOG Receptor Cluster  
1.6.3.2 The Melanocyte Concentrating Hormone Receptor Cluster  
1.6.3.3 The Chemokine Receptor Cluster  
1.6.3.4 Other Rhodopsin γ-Receptors  
1.6.4 The Rhodopsin δ-Group  
1.6.4.1 The MAS-related Receptor Cluster  
1.6.4.2 The Glycoprotein Receptor Cluster  
1.6.4.3 The Coagulation Factor Receptor Cluster  
1.6.4.4 The Purinergic Receptor Cluster  
1.6.4.5 The Olfactory Receptor Cluster  
1.6.4.6 Other Rhodopsin α-Receptors
Contents

1.7 Other GPCRs 21
1.8 Future Perspective 21

References 23

2 Why G Protein-coupled Receptors Databases are Needed 27
Jacques Haiech, Jean-Luc Galzi, Marie-Claude Kilhofer, Marcel Hibert, and Didier Rognan

2.1 Introduction 27
2.2 A Non-exhaustive List of the GPCR Data Models 27
2.3 Using the Central Dogma of Biology 28
2.4 Using the Tree of Life 30
2.5 Using a Chemogenomic Approach 35
2.6 Conclusion 38

References 38

3 A Novel Drug Screening Assay for G Protein-coupled Receptors 51
Brian F. O’Dowd, Xiaodong Ji, Mohammad Alijaniaram, Tuan Nguyen, and Susan R. George

3.1 Introduction 51
3.1.1 History 51
3.1.2 Nuclear Translocation of Endogenous GPCRs 52
3.1.3 The MOCA Method 52
3.2 The MOCA Strategy Demonstrated with the D1 Dopamine Receptor 53
3.2.1 Development of the Assay 53
3.2.2 Concentration-dependent Antagonist Blockade of Nuclear Transport 55
3.2.3 Measurement of Receptor Cell Surface Expression: Antagonist Binding of Receptors at Cell Surface 55
3.3 Development of Quantitative Methodology Suitable for High Throughput Analysis 56
3.3.1 Nuclear Translocation of Orphan GPCRs 58
3.4 Discussion of the MOCA Method 58
3.5 Conclusion 59

References 60

4 Importance of GPCR Dimerization for Function:
The Case of the Class C GPCRs 61
Laurent Prézeau, Cyril Goudet, Philippe Rondard, and Jean-Philippe Pin

4.1 Introduction 61
4.2 Class C GPCRs are Multidomain Proteins 62
4.2.1 The VFT 63
4.2.2 The CRD 64
4.2.3 The HD 65
4.2.4 CTail 66
4.3 Class C GPCRs are Constitutive Dimers 66
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Chemogenomics Approaches to Ligand Design</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td><em>Thomas Klabunde</em></td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction to Chemogenomics: Similar Receptors Bind Similar Ligands</td>
<td>115</td>
</tr>
<tr>
<td>7.2</td>
<td>Focused Libraries and Screening Collections Directed Against GPCRs</td>
<td>117</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Physicochemical Property-based Selection of GPCR Screening Sets</td>
<td>118</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Pharmacophore and Molecular Descriptors for GPCR Directed Libraries</td>
<td>118</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Privileged-fragment-based GPCR-directed Libraries</td>
<td>120</td>
</tr>
<tr>
<td>7.2.4</td>
<td>GPCR Collection and Subfamily-directed Library Design</td>
<td>121</td>
</tr>
<tr>
<td>7.3</td>
<td>Understanding Molecular Recognition: Impact on GPCR Ligand Design</td>
<td>124</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Sites for Ligand Recognition within Biogenic-amine-binding and Other GPCRs</td>
<td>125</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Design of GPCR-directed Libraries Using “Motifs” and “Themes”</td>
<td>127</td>
</tr>
<tr>
<td>7.3.3</td>
<td>“Chemoprints” for Recognition of GPCR-privileged Fragments</td>
<td>127</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Molecular Interaction Models by Proteochemometrics</td>
<td>131</td>
</tr>
<tr>
<td>7.4</td>
<td>Outlook</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>133</td>
</tr>
<tr>
<td>8</td>
<td>Strategies for the Design of pGPCR-targeted Libraries</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td><em>Nikolay P. Savchuk, Sergey E. Tkachenko, and Konstantin V. Balakin</em></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>137</td>
</tr>
<tr>
<td>8.1.1</td>
<td>Peptidergic GPCRs: Brief Overview</td>
<td>137</td>
</tr>
<tr>
<td>8.1.2</td>
<td>Endogenous Ligands for pGPCRs</td>
<td>140</td>
</tr>
<tr>
<td>8.1.3</td>
<td>Potential Therapeutic Targets of pGPCRs</td>
<td>140</td>
</tr>
<tr>
<td>8.2</td>
<td>Approaches to the Design of pGPCR-targeted Libraries</td>
<td>141</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Problems in Drug Discovery Directed Towards pGPCRs</td>
<td>143</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Docking and Pharmacophore-based Design</td>
<td>146</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Knowledge-based Data Mining Approaches</td>
<td>148</td>
</tr>
<tr>
<td>8.2.4</td>
<td>Chemogenomics Approaches</td>
<td>149</td>
</tr>
<tr>
<td>8.2.5</td>
<td>Incorporation of Specific Biomolecular Recognition Motifs</td>
<td>149</td>
</tr>
<tr>
<td>8.2.5.1</td>
<td>Privileged Structures</td>
<td>150</td>
</tr>
<tr>
<td>8.2.5.2</td>
<td>Mimetics of the Peptide Secondary Structure Elements</td>
<td>154</td>
</tr>
<tr>
<td>8.3</td>
<td>Synthesis of pGPCR-focused Libraries: Example of a Practical Methodology</td>
<td>156</td>
</tr>
<tr>
<td>8.4</td>
<td>Conclusions</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>160</td>
</tr>
</tbody>
</table>
## 9 Ligand-based Rational Design: Virtual Screening 165
*David E. Clark and Christopher Higgs*

### 9.1 Introduction 165

### 9.2 Why Use Ligand-based Virtual Screening? 166

#### 9.2.1 Speed 166

#### 9.2.2 Applicability 166

#### 9.2.3 Complementarity 166

### 9.3 Overview of Ligand-based Virtual Screening 167

#### 9.3.1 Starting Points 167

#### 9.3.2 Chemical Structure Databases 167

#### 9.3.3 Database Search Techniques 168

##### 9.3.3.1 2-D Substructure Searching 168

##### 9.3.3.2 2-D Similarity Searching 169

##### 9.3.3.3 3-D Substructure Searching 170

##### 9.3.3.4 3-D Similarity Searching 170

##### 9.3.3.5 Pharmacophore Searching 170

### 9.4 Successful Applications of Ligand-based Virtual Screening for GPCRs 170

#### 9.4.1 Somatostatin Agonists 171

#### 9.4.2 Muscarinic M3 Receptor Antagonists 172

#### 9.4.3 Urotensin II Antagonists 174

#### 9.4.4 Melanin-concentrating Hormone-1 Receptor Antagonists 176

#### 9.4.5 Growth Hormone Secretagogue Receptor Agonists 178

### 9.5 Conclusions 179

## 10 3-D Structure of G Protein-coupled Receptors 183
*Leonardo Pardo, Xavier Deupi, Cedric Govaerts, and Mercedes Campillo*

### 10.1 Introduction 183

### 10.2 Classification of G Protein-coupled Receptors 185

### 10.3 The Extracellular N-terminal Domain of G Protein-coupled Receptors 185

### 10.4 Sequence Analyses of the 7TM Segments of the Rhodopsin Family of G Protein-coupled Receptors 185

### 10.5 The Conformation of Pro-kinked Transmembrane α-Helices 186

### 10.6 Helix Deformation in the Rhodopsin Family of G Protein-coupled Receptors 187

#### 10.6.1 Transmembrane Helix 1 187

#### 10.6.2 Transmembrane Helix 2 188

#### 10.6.3 Transmembrane Helix 3 190

#### 10.6.4 Transmembrane Helix 4 190

#### 10.6.5 Transmembrane Helix 5 190

#### 10.6.6 Transmembrane Helix 6 192

#### 10.6.7 Transmembrane Helix 7 193

### 10.7 Structural and Functional Role of Internal Water Molecules 193

References 180
10.7.1 A Conserved Hydrogen Bond Network Linking D2.50 and W6.48  194
10.7.2 The Environment of the NPxxY Motif in TM7  195
10.8 Molecular Processes of Receptor Activation  195
10.8.1 Molecular Processes Initiated by the Recognition of the Extracellular Ligand by the Receptor  196
10.8.2 Molecular Processes that Propagate the Signal from the Ligand Binding Site to the Intracellular Amino Acids of the Transmembrane Bundle  196
10.9 Conclusions  198
References  199

11 7TM Models in Structure-based Drug Design  205
Frank E. Blaney, Anna-Maria Capelli, and Giovanna Tedesco
11.1 Introduction  205
11.2 Early Models of 7TM Receptors  207
11.3 Third Generation 7TM Models  208
11.3.1 Docking Ligands into Receptor Models  209
11.3.2 Designing 5-HT2C Selective Antagonists  210
11.3.3 Even Wrong Models Can be Useful: The Importance of SDM Studies  212
11.4 Fourth Generation Models  214
11.4.1 Revisiting the 5-HT2C Antagonist Binding Site  214
11.5 The Inclusion of Extracellular Loops in 7TM Models  215
11.6 Switching Selectivity in Neurokinin Antagonists  218
11.7 Homology (Fifth Generation) Models of 7TM Receptors  223
11.8 “Ligand-based” Design of 7TM Receptor Compounds  223
11.8.1 Pharmacophores Often do NOT Work  226
11.9 Refinement of 7TM Pharmacophores Using Current Receptor Models  228
11.10 Optimizing Properties of the CCR2 Antagonists  230
11.11 Some General Ligand Considerations When Docking  233
11.12 What the Future Holds  236
11.13 Abbreviations and Nomenclature  236
References  237

12 Receptor-based Rational Design: Virtual Screening  241
Didier Rognan
12.1 Introduction  241
12.2 Structure-based Screening Workflow  242
12.2.1 Setting Up a Ligand Library  242
12.2.2 Docking and Scoring  244
12.2.3 Data Post-processing  246
12.3 Retrospective Screening  247
12.3.1 Give it a Try  247
12.3.2 Several Alternative Screening Strategies  249
12.4  Prospective Screening  250  
   12.4.1  Screening Rhodopsin-based Ligand-biased Homology Models  250  
   12.4.2  Screening Ab Initio Models  253  
   12.4.3  A Few Difficult Screening Scenarios  253  
12.5  Conclusions  255  

References  256  

Subject Index  261
Preface

G protein-coupled receptors (GPCR) represent to the best of our knowledge more at least 60% of all receptors. This vast majority keeps them still alive as the most interesting group of targets in drug finding and development. Some 18,000 reviews are listed in Pubmed, many of them dealing with structural features and peculiarities of G protein-coupled receptors. Especially their functional categorization, association with other membrane-integral proteins and dimerization/oligomerization behaviour is still a hot topic in research.

Nevertheless, the existing body of knowledge at atomic resolution, enables us to propose interaction mechanism and activation models for this type of receptor. Here it is the merit of Didier Rognan, himself being one of the leading figures in the field of molecular modelling of GPCRs, that he started to collect a number of reputed researches sharing a history in the topic of GPCRs and edited a 12 chapter volume on the state-of-the-art in ligand design for those targets.

The volume starts with a genomic overview on GPCRs, which is followed by an appropriate review of the available data and their appearance and utilisation in databases. In more specialized chapters the question is raised how to de-orphanize receptors. Strategies in these fields are urgently needed since by HTS strategies, array technologies, etc., the number of orphan receptors has grown exponentially.

Ligand interaction does not mean at all that a drug will emerge from this knowledge. So, druggability analysis, which has overcome its infant years of rule-based estimates, has become a sophisticated methodology on its own. One chapter is devoted to druggability of human GPCRs. It’s the molecular mechanism which is illuminated in depth within the subsequent three chapters. Oligomerization or just dimerization, activation/inactivation processes and allosteric regulation are still complex puzzles to solve, last but not least because of the difficulties of understanding the entropy contribution.

Further chapters are dedicated to computational procedures. Chemical genomics approaches are going to be presented, the development detection of targeted libraries and privileged structures for GPCR interaction and leadhopping and virtual screening approaches to ligand design.

The final three chapters deal with the 3D-structures of GPCRs and the usefulness as a basis for rational design of ligands. Both, modelling approaches as well as virtual screening will be discussed in extenso.
Thus, we expect this new volume in the series to be of fundamental interest to a large community of scientists and researches devoted to GPCRs. The editors are deeply convinced that the contents of this book will help to fathom the potential of GPCRs and will generate new ideas and visions for their role in drug discovery.

The editors are indebted to Renate Doetzer and Frank Weinreich from Wiley-VCH for their invaluable support in this project which is hereby gratefully acknowledged.

_Raimund Mannhold, Düsseldorf_

_Hugo Kubinyi, Weisenheim am Sand_

_Gerd Folkers, Zürich_
A Personal Foreword

Describing in a single book all existing approaches to design ligands targeting G protein-coupled receptors (GPCRs) is an impossible challenge. However, giving some clues to assist drug designers in their daily work is feasible. This is precisely the aim of the current book whose contributors have been selected to reflect the current knowledge on an extraordinary diverse family of protein targets. We have chosen to address fundamental and methodological issues which are the most likely to promote rational design of GPCR ligands. Ligand selectivity cannot be answered without considering the entire protein family at a genomic level. Dissecting the fine molecular mechanisms underlying GPCR function is also necessary to design ligands with the desired pharmacological profile. Last, a precise knowledge of current biostuctural data is necessary to decide whether a ligand-based and/or a receptor-based design strategy is the most adequate. Many design strategies are indeed possible. Their potential is however very dependent on the current knowledge about a particular target and their ligands. It is therefore of outmost importance to be aware of all available information and design methods while beginning a drug discovery program. We do hope that this book will provide the reader with the necessary material to start with.

I would like to thank all contributors for their nice collaboration and their effort to respect a few deadlines in submitting their chapters. The series editors are also warmly acknowledged for their valuable comments in the early definition of the Table of Contents. I am grateful to Frank Weinreich and Renate Doetzer from Wiley-VCH for the nice collaboration over several months. Last, I extend my thanks to my family for their continuous support.

Illkirch, January 2006

Didier Rognan
List of Contributors

Mohammad Alijaniaram
Centre for Addiction and Mental Health
250 College St.
Toronto, Ontario M5T 1R8
Canada

Konstantin V. Balakin
ChemDiv, Inc.
11558 Sorrento Valley Road Ste. 5
San Diego, CA 92121
USA

Frank Blaney
Department of Computational, Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research NFSP (North), Third Avenue
Harlow, Essex CM19 5AW
United Kingdom

Robert P. Bywater
Adelard Institute
9 Limes Avenue
London NW7 3NY
and
14 TM Ltd.
100 Low Road
Burwell, Cambridge CB5 OEG
United Kingdom

Mercedes Campillo
Laboratori de Medicina Computacional Unitat de Bioestadistica and Institut de Neurociencies
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain

Anna-Maria Capelli
Department of Computational, Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research
Via A. Fleming 4
37135 Verona
Italy

David Clark
Argenta Discovery Ltd.
8/9 Spire Green Centre
Flex Meadow
Harlow, Essex CM19 5TR
United Kingdom

Paul Denny-Gouldson
14 TM Ltd
100 Low Road
Burwell, Cambridge CB5 OEG
United Kingdom
Xavier Deupi  
Laboratori de Medicina Computacional  
Unitat de Bioestadística and Institut de Neurociències  
Universitat Autònoma de Barcelona  
08193 Bellaterra  
Spain

Robert Fredriksson  
Department of Neuroscience  
Uppsala University  
Husargatan 3, Box 593  
75124 Uppsala  
Sweden

Jean-Luc Galzi  
Faculté de Pharmacie  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
74, route du Rhin  
67401 Illkirch  
France

Susan R. George  
Department of Pharmacology and Medicine  
University of Toronto  
Medical Science Building  
1 King’s College Circle  
Toronto, Ontario M5S 1A8  
Canada

Cyril Goudet  
Département de Pharmacologie  
CNRS UMR 5203 – INSERM U661  
Université de Montpellier  
141 rue de la Cardonille  
34094 Montpellier cedex 5  
France

Cedric Govaerts  
Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire  
Campus Erasme  
Université Libre de Bruxelles  
1070 Brussels  
Belgium

Muriel Hachet-Haas  
Département Récepteurs et Protéines Membranaires  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
Boulevard Sébastien Brant  
BP 10413  
67412 Illkirch  
France

Jacques Haiech  
Département Récepteurs et Protéines Membranaires  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
Boulevard Sébastien Brant  
BP 10413  
67412 Illkirch  
France

Marcel Hibert  
Faculté de Pharmacie  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
74, route du Rhin  
67401 Illkirch  
France

Christopher Higgs  
ArgentaDiscovery Ltd.  
8/9 Spire Green Centre  
Flex Meadow  
Harlow, Essex CM19 5TR  
United Kingdom
Brigitte Ilien  
Département Récepteurs et Protéines Membranaires  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
Boulevard Sébastien Brant  
BP 10413  
67412 Illkirch  
France

Emeline Maillet  
Département Récepteurs et Protéines Membranaires  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
Boulevard Sébastien Brant  
BP 10413  
67412 Illkirch  
France

Xiaodong Ji  
Centre for Addiction and Mental Health  
250 College St.  
Toronto, Ontario M5T 1R8  
Canada

Tuan Nguyen  
Centre for Addiction and Mental Health  
250 College St.  
Toronto, Ontario M5T 1R8  
Canada

Marie-Claude Kilhoffer  
Faculté de Pharmacie  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
74, route du Rhin  
67401 Illkirch  
France

Brian F. O’Dowd  
Department of Pharmacology and Medicine  
University of Toronto  
Medical Science Building  
1 King’s College Circle  
Toronto, Ontario M5S 1A8  
Canada

Thomas Klabunde  
Sanofi-Aventis Deutschland GmbH  
Science & Medical Affairs – Drug Design  
Industriepark Höchst G878  
65926 Frankfurt  
Germany

Leonardo Pardo  
Laboratori de Medicina Computacional Unitat de Bioestadística and Institut de Neurosciences Universitat Autònoma de Barcelona 08193 Bellaterra  
Spain

Sandra Lecat  
Département Récepteurs et Protéines Membranaires  
CNRS UMR 7175-LC1  
Institut Gilbert Laustriat  
Boulevard Sébastien Brant  
BP 10413  
67412 Illkirch  
France

Jean-Philippe Pin  
Département de Pharmacologie  
CNRS UMR 5203 – INSERM U661  
Université de Montpellier  
141 rue de la Cardonille  
34094 Montpellier cedex 5  
France
Laurent Prézeau
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Didier Rognan
Laboratoire de Pharmacochimie
de la Communication Cellulaire
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
74, route du Rhin
67401 Illkirch
France

Philippe Rondard
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Nikolay Savchuk
ChemDiv, Inc.
11558 Sorrento Valley
Road Ste. 5
San Diego, CA 92121
USA

Helgi B. Schiöth
Department of Neuroscience
Uppsala University
Husargatan 3, Box 593
75124 Uppsala
Sweden

Giovanna Tedesco
Department of Computational,
Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research
Via A. Fleming 4
37135 Verona
Italy

Sergey E. Tkachenko
ChemDiv, Inc.
11558 Sorrento Valley
Road Ste. 5
San Diego, CA 92121
USA
1
G Protein-coupled Receptors in the Human Genome

Robert Fredriksson and Helgi B. Schiöth

1.1 Introduction

The superfamily of G protein-coupled receptors (GPCRs) is one of the largest families of proteins in the human genome [1, 2] and probably also in most other vertebrate species [3]. GPCRs participate in a diversity of important physiological functions and are targets for many modern drugs. Their ligands are particularly diverse, and include ions, organic odorants, amines, peptides, proteins, lipids, nucleotides and photons, which are all able to activate GPCRs. The main structural characteristic of GPCRs is seven stretches of about 25–35 consecutive amino acid residues that show a relatively high degree of hydrophobicity and represent \( \alpha \)-helixes that span the plasma membrane in an anti-clockwise manner. These sequences stretch from the common area or a recognition and connection unit of all GPCRs, enabling an extracellular ligand to exert a specific effect on the cell. This area of the receptors is generally relatively well conserved and is used to identify and classify novel GPCRs as other areas of the receptors are frequently much more diverse. The name GPCRs indicates that these receptors interact with G-proteins. This has however not yet been demonstrated for most of the proteins classified as GPCRs. Moreover, GPCRs are known to have many alternative signaling pathways, interacting directly with a number of other proteins such as arrestins and kinases. Hence, it would perhaps be more technically correct to term this superfamily “seven transmembrane (TM) receptors”, but the GPCR terminology has become more established.

Both physiological and structural features have been used to classify GPCRs. The first classification system was introduced in 1994 by Attwood and Findley [4]. They used the term “clans” to designate the different GPCR families. The classified dataset at this time contained over 240 rhodopsin-like GPCRs from different species. Many of these receptors were olfactory and light-recognizing receptors of the opsin type. Independently, but around the same time, Kolakowski presented the well known “Family A–F classification system” [5]. This system included receptors shown to bind G-proteins while the other 7TM receptors were classified as O (other). In conjunction with this classification system the database GPCRdb was
developed and included at that time 777 unique GPCRs from various species. Family A contained receptors similar to rhodopsin and biogenic amine receptors. Family B contained receptors similar to the secretin and calcitonin receptors while Family C contained the metabotropic glutamate receptors. Family D and E contained only receptors that were not, and still have not been, identified in mammals, namely the fungal pheromone receptors and the cAMP binding receptors, respectively. Finally, Family F contained archebacterial opsins. The Kolakowski classification system was later extended independently, and differently, by Josefsson and Flower in 1999 [6, 7]. Moreover, another classification system was suggested in 1999 that contained in total five families based on the position of the ligand-binding pocket and the sequence length of the receptors. This system excluded the receptors that are not present in vertebrates [8]. This system used both structural and physiological features to classify the receptors. Recently, we have undertaken large-scale systematic phylogenetic analyses including the majority of the GPCRs in the human genome [9]. This provides us with the GRAFS system showing five main families named Glutamate (G; previous family C/3), Rhodopsin (R; previous family A/1), Adhesion (A; previously part of family B/2), Frizzled/Taste2 (F; previously O/5 and not included) and Secretin (S; previously part of family B/2). Moreover, we subdivided the large Rhodopsin family into 13 subgroups. The grouping was carried out using strict phylogenetic criteria and only a few human receptors did not group into these clusters and these receptors were thus placed into what we called Other 7TM receptors. There are several GPCRs that have been discovered since we published this classification [10–13] and here we present an updated version of the human repertoire. In this overview we describe each of the families and groups within the GRAFS classification system and include phylogenetic trees which were derived by Maximum Likelihood and show branch lengths.

1.2 The Adhesion Family

The Adhesion family is the second largest GPCR-family in humans with 33 members. The group is called Adhesion GPCRs according to a recent GPCR classification [9] and this nomenclature seems to prevail. This family has however been assigned various names through the years. These include EGF-TM7 to reflect the presence of epidermal growth factor (EGF) domains in the N-termini [14, 15] and LN-TM7 receptors where LN stands for long N-termini and B2/LNB-7TM to reflect their vague similarity to secretin receptors [16]. The Adhesion family members have several structural features that clearly separate them from all other groups of GPCRs. In a recent article we showed the entire repertoire in human and mouse where the diversity of their N-termini is highlighted [12]. Their long N-termini contain a high percentage of Ser and Thr residues that can create O- and N-glycosylation sites. These N-termini or stalk-like regions are thus thought to be highly glycosylated and act as a mucin-like domain with a rigid erect structure protrud-
ing from the cell surface. The long N-termini are believed to bind various proteins that promote cell-to-cell and cell-to-matrix interactions. All of the Adhesion GPCRs except GPR 123 contain a GPCR proteolytic domain (GPS). Additionally their N-termini can contain a number of different domains that are also found in various other proteins, such as cadherin, lectin, laminin, olfactomedin, immunoglobulin or trombospondin. It is likely that the repertoire of these domains plays an important role in the functional specificity of the receptors.

Phylogenetically, as can be seen in Fig. 1.1, this family forms three main subfamilies with the largest containing lectomedin, EGF-like module, cadherin EGF, EGF lathrophilin, CD97 and GPR 127 receptors. It is interesting to note that all receptors in this group, with the exception of lectomedin receptors, contain EGF domains and that no receptors outside this cluster contain this type of domain [12]. Continuing clockwise in Fig. 1.1 the second group contains 10 receptors termed GPR 110, GPR 111, GPR 113, GPR 115, GPR 116, GPR 123, GPR 124, GPR 125, GPR 133 and GPR 144. The receptors in this cluster have in general very few recognizable domains in their long N-termini, with GPR 123 having no known domains and GPR 110, GPR 111 and GPR 115 having only a GPS domain. The other receptors contain immunoglobulin domains (GPR 124, GPR 125 and GPR 116), hormone-binding domains (GPR 113), leucine-rich repeats (GPR 124 and GPR 125), a pentraxin domain (GPR 144) and a sea urchin sperm domain (GPR 116) [12]. The third family contains brain angiogenesis receptors, human epithelial receptors, the very large GPR 1 (over 6300 amino acids long) and GPR 56, GPR 97, GPR 112, GPR 114, GPR 126 and GPR 128. Also, several receptors in this group are rather sparse in known domains with GPR 56, GPR 97, GPR 114, GPR 126, GPR 128 and human epidymal receptor containing only GPS domains. The three Brain Angiogenesis Inhibitor GPCRs (BAI) contain only hormone-binding and trombospodin domains while GPR 112 contains a pentraxin domain. The very large GPR 1 contains several copies of the sodium–calcium exchange/integrin beta domains [12]. Although several of the more recently discovered Adhesion GPCRs have surprisingly few recognizable functional domains in their N-termini, it is likely that these receptors contain novel domains that are not recognizable using current bioinformatics tools. The majority of the Adhesion GPCRs are orphans and for the few that have been characterized with regard to ligand binding, none has been shown to bind their ligand within the TM regions. CD97 is one of the most studied receptors in this family and is found in several types of blood cell. CD97 interacts with the 312-amino acid membrane protein CD55 (or decay accelerating factor; DAF) which is expressed on most leukocytes [17]. Recently, it was also shown that a glycosaminoglycan (chondroitin sulfate) acts as a cellular ligand specific to the EGF-like domains of the EMR2 [15]. Receptors of the Adhesion family are expressed in various parts of the human body and many of them have prominent expression in the immune system, central nervous system, and in the reproductive organs, suggesting that they might take part in a large variety of physiological functions.
Fig. 1.1 Phylogenetic trees for each family and group of GPCRs. The topology is a consensus tree from 100 bootstrap replicas calculated using ordinary parsimony and the branch lengths are optimized using the Maximum Likelihood method with no assumptions of the presence of a molecular clock and hence the branch lengths correspond to protein distances, not evolutionary time. The calculations are performed on the amino acid sequences from transmembrane region 1 to the end of transmembrane region 7, meaning that no N-termini are used for the calculations as they are very divergent for the different groups. R is the abbreviation for receptors in the annotation on the figures.
1.3 The Secretin Family

The secretin family consists of 15 receptors and occurs widely in all animal species [16, 18]. The N-terminal regions of these receptors share some primary sequence similarity with the Adhesion family of GPCRs and in the A-F classification system both are considered to belong to the B family. Although this sequence similarity is clearly recognizable, it is evident that the Secretin and the Adhesion families are evolutionarily old and that they split into individual groups long ago. Both families are present as multimember families in both insects such as D. melanogaster and A. gambiae as well as in C. elegans [3] as are the other main families, i.e. the Adhesion, Rhodopsin, Frizzled and Glutamate families. The phylogenetic tree of this family has four main subgroups, the largest one consisting of the secretin, Growth Hormone Releasing Hormone, Vasoactive Intestinal Peptide and Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) receptors. The other groups contain, in clockwise order in Fig. 1.2, the Corticotropin Releasing Hormone/Calcitonin Gene Related Peptide Receptors, Glucagon/Glucagon-like Peptide/Gastric inhibitory peptide receptors and the Parathyroid Hormone Receptors. The receptors in the Secretin family bind rather large peptides and most often act in a paracrine manner. The Secretin family name is related to the fact that the secretin receptor was the first of this family to be cloned and the term secretin-like receptors has also frequently been used in the literature with reference to receptors in this cluster. The N-terminal, between about 60 and 80 amino acids long, contains conserved Cys bridges and is particularly important for the binding of the ligand to these receptors. For example, the N-terminal alone of the VIPR and PACAP receptor constitutes a functional binding site for the ligand. The receptors have a recognizable “hormone binding domain” in the N-termini and these receptors bind rather large peptides that most often act in a paracrine manner.

1.4 The Frizzled/Taste2 Family

Our phylogenetic studies on the human repertoire have indicated that two very different groups of receptors cluster together. There are few elements in the consensus sequence and the HMM models, such as the consensus sequence of IFL in TM2, SFLL in TM5, and SxKTL in TM7 which are motifs that do not seem to be present in the consensus sequences of the other four families, that could explain why these two groups of receptors cluster together. Further studies are needed to investigate whether these two groups have a common evolutionary history. Below we look at these groups separately.
1.4.1

The Frizzled Receptor Cluster

The Frizzled group consists of 10 frizzled receptors named Frizzled receptor 1–10 and the single Smoothened receptor. The topology of the tree in Fig. 1.3 shows four main clusters: the cluster containing Frizzled 1, 2 and 7 which have approximately 75% identity to each other; the Frizzled 8 and 5 that have 70% identity, the
Frizzled 10, 9 and 4 that have around 65% identity; and finally Frizzled 6 and 3 that have 50% amino acid identity. The identities shared by receptors from different clusters are between 20 and 40%, indicating that four parental genes from the Frizzled family were initially formed and subsequently the four clusters originated out of these. Smoothened is, as evident in Fig. 1.3, clearly the most divergent of the receptors from the Frizzled family, sharing only 24% amino acid identity to FZD2 and less to the others. The large evolutionary distance between the Smoothened receptor and the other Frizzled receptors also reflects a large evolutionary time of divergence as Smoothened are found as a distinct receptor back in *C. elegans*, which diverged from the lineage leading to mammals more than 600 million years ago [19]. Despite this large sequence divergence between Smoothened and the other Frizzled receptors, all these receptors clearly belong to the same family, which has been shown from phylogenetic analysis of the entire GPCR family [9]. The frizzled receptors control cell fate, proliferation, and polarity during metazoan development by mediating signals from secreted glycoproteins termed Wnt. The

**Fig. 1.3** Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.