Drug Discovery Today: Technologies
Vol. xxx, No. xx 2004



Target validation

# Reverse pharmacology and the de-orphanization of 7TM receptors

Knut Kotarsky[1,*], Niclas E. Nilsson[2]

[1]Section of Immunology, Department of Cellular and Molecular Biology, Lund University, BMC 113, Tornavägen 10 221 84 Lund, Sweden
[2]Division of Molecular Neurobiology, Department of Physiological Sciences, Lund University, BMC A12, Tornavagen 10 221 84 Lund, Sweden

**Approximately 800 genes coding for seven-transmembrane, G-protein-coupled receptors have so far been recognized. In spite of this, many of these receptors are defined by their sequence only, and are therefore classified as orphan receptors. Without knowing what their endogenous ligands are, we lack the information needed to understand their physiological role and hence cannot make use of them as drug targets. In this communication, we discuss different strategies, as well as difficulties in the deorphanizing process.**

**Section Editors:**
Luis Menandez-Arias, Pierre Chatelain, Bernard Masareel

Approximately 800 genes coding for seven-transmembrane (7TM), G-protein-coupled receptors (GPCRs) have so far been recognized. In spite of this, many of these receptors are defined by their sequence only, and are therefore classified as orphan receptors. Without knowing what their endogenous ligands are, we lack the information needed to understand their physiological role and hence cannot understand the physiological role and possibly make use of them as drug targets. In this communication, Kotarsky and Nilsson discuss critically different strategies, as well as difficulties in the deorphanizing process. Reverse pharmacology and de-orphanization of 7TM receptors is a powerful approach beacuse 50–70 orphan GPCRs have been de-orphanized to date, leading to significant progress.

## Introduction

A large proportion of all drugs used today in medicine interact with members of the seven-transmembrane (7TM), G-protein-coupled receptor (GPCR) family. However, only a small number of 7TM receptors are actually targeted by drugs [1]. For many 7TM receptors no physiological ligand has yet been identified; therefore, they are termed ''orphan'' receptors. Their potential as drug targets is intimately connected with a thorough knowledge of the expression patterns and signaling pathways, as well as their physiological role and ligand(s). To identify new drug targets, the screening for novel 7TM receptor ligands has gained much attention over the past decade.

Considering the significant number of orphan receptors, the question has been asked; why do so many of them still lack a known ligand? Although specific answers are most likely dependent on each individual receptor, difficulties in the process of ligand identification might be grouped into three main categories (i) the receptor itself, (ii) the assay systems used for identification, and (iii) the nature of the ligand. Some of the pitfalls are shared by all strategies, other problems are more closely connected to one or another approach.

## Key technologies and strategies

### *Orphan receptors*

The genetic identification of unknown receptors in the past was accomplished by using molecular biological techniques (e.g. polymerase chain reaction with degenerated primers). Nowadays, after the completion of the human genome sequencing project, sequences with a certain degree of homology to known 7TM receptors have been continuously discovered using *in silico* approaches and are stored in public

*Corresponding author: (KnutKotarsky) knut.kotarsky@immuno.lu.se

 DOI: 10.1016/j.ddtec.2004.07.003

**Glossary**

**7TM or GPCR:** seven-transmembrane receptors traverse the membrane seven times. The N-terminus of the receptor is located outside the cell. The receptors relay their actions to the inside of the cell activating so-called heterotrimeric G-proteins, hence, they are G-protein-coupled.
**De-orphanization:** is the process, which results in the identification of a natural ligand acting on an orphan receptor.
**Odorant:** a chemical compound with a distinct smell. The odorant receptors form a large group of receptors within the rhodopsin family of 7TM receptors.
**Orphan receptor:** is a receptor without an assigned and accepted ligand, often purely defined by a novel sequence with a certain degree of similarity to other known receptors.
**Z-factor:** combines the difference between background and signal with the differences in the standard deviation of both values resulting in a number between 0 and 1, where 1 reflects the ideal assay.

databases. The estimated number of 7TM receptors (GPCRs) varies to some extent, but not much [2,3]. Recently, Fredriksson *et al*. recognized and phylogenetically analyzed more than 800 7TM receptors in humans. The authors subdivided these receptors into five branches, according to the sequence similarity of their transmembrane regions. Approximately 350 of these receptors belong to the group of sensory receptors (primary odorant receptors). Of the remaining 342 non-odorant receptors [3], about 100 receptors still lack a described and confirmed ligand (Table 1).

However, not all orphan 7TM receptors are suitable as drug targets. Of the 2507 TM receptors which are already identified, no more than 30–40 are actually used as drug targets (15–20%) [1]. A restricted expression pattern in the area of interest and a physiological role (involvement in regulatory circuits) are prerequisites for a ''drugable'' receptor.

### *Applied strategies to deorphanize 7TM receptors*

Classical physiology and pharmacology explain an observed phenomenon by the characterization of the active principle (compound). The opposite approach has been described as ''reverse pharmacology'' [4]. In this case, the identification of a receptor sequence is followed by the discovery of the corresponding ligand. Subsequently the pharmacological and physiological context is investigated for example, by gene knockout or over-expression of the receptor and its ligand.

Three different strategies have been applied in the process of reverse pharmacology (Fig. 1). The pharmaceutical industry and some larger academic laboratories have designed extensive substance libraries, which contain, in addition to all the known 7TM receptor ligands, predicted ligands for such receptors. These substance libraries were subsequently matched with ''orphan'' receptor libraries, containing as many orphan receptors as possible, to identify positive hits [5]. This approach will be referred to as the ''library-based approach'' (Fig. 1a).

Smaller laboratories, by contrast, work mainly on a few receptors at a time, starting with crude tissue extracts and isolating the active compound in a laborious procedure. The second approach is often referred to as the ''orphan receptor strategy'' [6] but will here be referred to as the ''tissue-extract-based approach'' (Fig. 1b).

A third strategy considers the reported effects of certain substances, which are suspected to act as a ligand for 7TM receptors. The available information regarding affected cells or tissues is used to identify a small number of suspected orphan receptors, which are then challenged with the compounds of interest [7,8]. This can be referred to as the information-based approach (Fig. 1c).

### *Common difficulties*

There are a couple of common problems, which have to be solved regardless of the strategy used (Fig. 2). The first problem is connected to the receptor itself, due to the fact that a correct open reading frame (ORF) has to be identified. The N-terminus of a 7TM receptor can be spliced and might result in the expression of different protein variants. Some of the splice variants miht even exhibit differences in their pharmacological properties [9] (Fig. 2a).

**Table 1. The number of expressed human 7TM receptors (GPCRs), excluding odorant receptors, based on three different sources [2,3] and database of the IUPHAR at: http://www.iuphar-db.org/iuphar-rd/index.html**

| 7TM receptor group | Vassilatis *et al*. "endo GPCR" | Fredriksson *et al*. "non-olfactory" | IU-Pharm |
|---|---|---|---|
| **Class A (rhodopsin type)** | 284 (98) | 241 (52) | 280 (95) |
| **Class B (secretin family)** | 52 (34) | 39 (22) | 55 (35) |
| **Class C (metabotropic family)** | 17 (6) | 15 (1) | 19 (9) |
| **Class F/S** | 11 (0) | 24 (2) | 11 (0) |
| **Others** | 5 (5) | 23 (17) | |
| **Total** | 369 (143) | 342 (94) | 365 (139) |

The number in parentheses indicates the proposed number of orphan receptors within each sub-family. Taken all together the total number of orphan receptors is approximately 100.

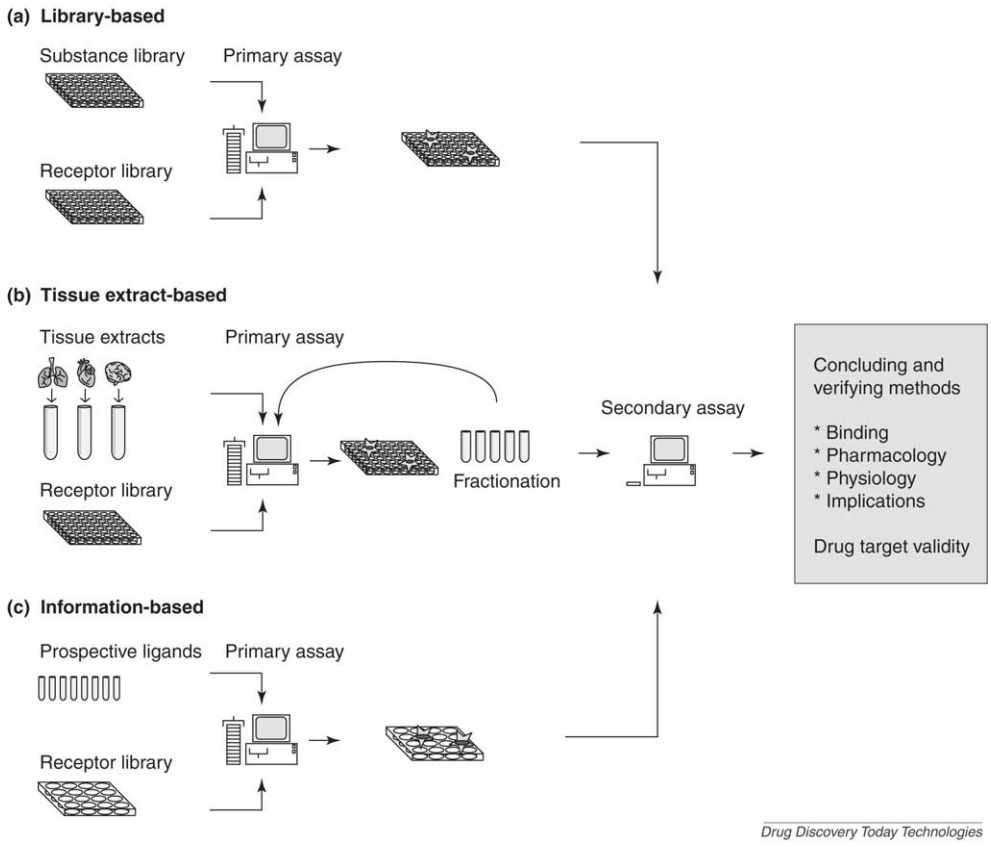


**Figure 1.** Three different strategies of reverse pharmacology used to identify ligands acting on 7TM, G-protein-coupled receptors. (a) A library-based approach, which matches an extensive substance library with a large receptor library. (b) A tissue-extract-based approach that identifies hits in fractionated tissue samples. (c) An information-based approach that identifies prospective ligands by database screening and testing them on a small number of selected orphan receptors.

Once the ''proper'' reading frame of a gene is identified and cloned, the receptor gene has to be expressed heterologously to obtain the receptor protein. However, recombinantly expressed receptor proteins are not always correctly targeted to the cell surface. To verify the cell surface expression, different techniques have been applied, but none of them with completely satisfying results. For example, the receptor can be tagged with either a short peptide (on the N-terminus) or EGFP (usually on the C-terminus). However, both of these techniques could potentially influence receptor protein targeting, disturb ligand binding, or alter the signaling behavior of the receptor. Hence, the cell surface expression of the receptor has to be controlled in functional assays.

Another approach to verify the desired cell surface expression of a receptor relies on the signaling behavior of the receptor. Intrinsic signaling of a receptor could either be achieved by over-expression or by a mutation, which makes the receptor constitutively active [10]. The over-expressed receptor might also attract G-proteins from other endogenously expressed receptors, thus decreasing signals obtained from these receptors. None of the techniques described is straightforward; all require a considerable amount of work, yet, still they do not guarantee success.

Other problems related to the receptor protein might arise from its possible interaction with other molecules in the cell membrane. The most prevalent examples are: the dimerization of two gene products to form a functional $GABA_B$ receptor (NP_005449 and NP_001461) [11] and the association of the calcitonin receptor-like receptor with receptor associated proteins (RAMPs). In the latter case, the interaction of RAMP2 (NP_005845) or 3 (NP_005847) with the receptor leads to an adrenomedullin receptor, whereas interaction with RAMP1 (NP_005846) results in the formation of a calcitonin-gene-related-peptide receptor [11] (Fig. 2b) (see also the ''Outstanding issue box'').

Another potential problem is the coupling of the receptor to effectors inside the cell. The difficulty related to receptor signaling and assay systems are shared by all three approaches. To identify ligands acting on 7TM receptors, the receptor protein is used as a ''fishing tool'', either in binding assays or in functional tests. Traditionally, assay techniques screen for changes in intracellular second messengers, hereby, assuming that the receptor couples to G-proteins (see [12] for review). There is a plethora of different assay systems described in the literature. However, because of the high demands placed on the assay used for the initial characterization,

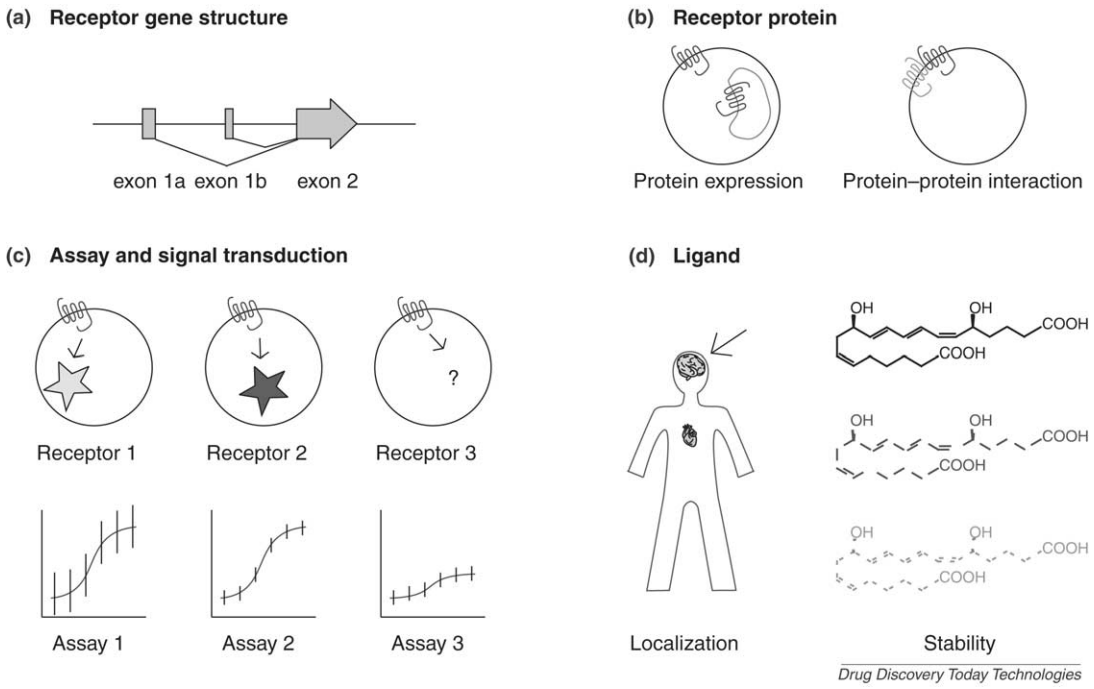


**Figure 2.** The functional identification of 7TM, G-protein-coupled receptors might face different difficulties and problems. (a) The correct gene structure (open reading frame) has to be determined. (b) The surface expression of the receptor and its potential dependence on interacting proteins in the membrane should be considered. (c) Receptors might activate different signaling pathways and therefore it has to be ensured that a receptor's activation is monitored by the assay. Additionally, assay techniques display different qualities (signal-to-noise ratio; sensitivity; statistical deviation), which can be compared using the Z-factor. (d) The localization of the ligand and its stability are crucial in tissue-based fractionation approaches.

techniques successfully used for identification purposes are not as numerous [13]. These assay techniques have to display high reliability and quality, a property that is defined by a high Z-factor [14] (Fig. 2c). The main techniques used today monitor changes in intracellular calcium concentrations, either with the help of calcium sensitive fluorescent dyes, such as Fura-2, or by using the calcium-sensitive luciferase, aequorin (P07164) [13].

The prevalent use of these techniques relies on the introduction of chimeric, and/or promiscuous, G-proteins which are supposed to redirect signaling of any 7TM receptor into the calcium signaling pathway [15]. Thus, a mixture of chimeric and promiscuous G-proteins is often co-expressed with the ''orphan'' receptor. However, there are several reports questioning the over-all applicability of this ''G-protein cocktail'' method (as reviewed in [16]).

Besides fluorescent indicators and the aequorin technique, half a dozen other assay systems have been successfully used. However, these techniques, are performing at a lower throughput, are hampered by long incubation times, or have other disadvantages [13,17].

Once an orphan receptor's ligand has been identified, the initial hit has to be verified by a secondary assay with lower assay quality (with a decreased Z-factor). Ideally, the hit would be confirmed by binding studies showing that the labeled ligand indeed interacts with the receptor protein.

### *Approach specific difficulties*

In the library-based approach earlier described, the success depends to a great extent on the substance library, which contains known and potential ligands for 7TM receptors. The sizes of the compound libraries vary greatly, ranging from several dozen to hundreds of thousands of substances. Depending on the assay, this strategy can be performed in a high-throughput format. Using this approach, a large number of receptors has already been identified [5]. One advantage of this method is that pure substances can be applied at relatively high concentrations. The major drawbacks of the compound library are twofold: (i) the library is relatively expensive to purchase and (ii) obtained hits depend entirely on the composition of the library. Completely unknown mediators will not be found using this method and even peptide ligands might be missed, exemplified by the description of the discovery of the ghrelin peptide [18].

The second, ''tissue-extract-based'' approach, sometimes termed the ''orphan receptor strategy'' involves applying a tissue extract on a recombinantly expressed receptor. In theory, the ligand should be present at locations where the

**Table 2. Summary of the main techniques used in the de-orphanization of 7TM receptors, their pros and cons, as well as groups, which have extensively applied these techniques**

| Name of the approach | Library-based | Tissue-extract-based | Information-based |
|---|---|---|---|
| **Pros** | Pure substances can be tested at high concentrations | Unknown ligands can be found irrespective structure | Known relevance of the ligand |
| **Cons** | Composition of the library | Demands high quality assay and tissue extracts | Work only for relative, well investigated ligands |
| **Prerequisites** | High-throughput screening laboratory | Tissue availability and compound fractionation | Thorough database and literature screening |
| **Groups and companies using it** | GlaxoSmithKline; Stevenage, UK | Civelli group at University of California, Irvine and Takeda Chemicals, Japan | Offermann group at University of Heidelberg, Germany |
| **References for examples** | [5] | [6,18,19] | [7,8] |

receptor is expressed. In addition, the extracts of those tissues can be prepared, fractionated, and screened in a functional assay.

Because the orphan receptor strategy commences with a tissue extract containing a complex mixture of different molecules, it requires a particularly high quality assay. Other signaling substances present in the extract might activate the cellular assay system used for receptor identification. False positives can result from non-specific activation that is independent of the orphan receptor expression. Sub-fractionation of tissue extracts, before testing, might reduce the background noise caused by other substances.

Another problem of this approach is the stability of the ligand, which could complicate its purification, particularly if its amount is limited in the starting material (Fig. 2d). Furthermore, some ligands might consist of several parts which are not covalently bound together and which would be eluted in different fractions during purification. Finally, the identification of a suitable source might be difficult, because the natural ligand might be present only during certain physiological conditions.

However, by means of the tissue-extract-based strategy, a large number of natural ligands, including unusual ligands with extraordinary features, have been discovered. For example, the peptide ghrelin, which acts on the GHS receptor, was identified using this strategy [18]. This unusual peptide has a covalent bound octanoyl modification, which is required for functional ligand–receptor interaction. Yet another example is the discovery of a bromidated peptide, which is the ligand for the GPR7 receptor. In this case, bromidation is not required for receptor–ligand interaction [19].

The starting point for the third, information-based approach is a substance, with known or proposed physiological effects, in a certain tissue or even cell type. Orphan receptors, which expression coincident with the effect pattern of the suspected ligand, are cloned, expressed and eventually tested. Hence, a large proportion of the work is performed *in silico* to identify potential targets. In the performed screening, a small number of related ligands are eventually matched with a few selected orphan receptors. This approach has resulted in some, physiologically interesting identifications [7,8].

## Strategy comparison

Different techniques can be utilized to deorphanize the remaining 100 orphan receptors. Three main strategies of reverse pharmacology have been used in the past (Table 2). The library-based approach is hampered mainly by the composition of the library tested. In this approach, pure test substances are applied at high concentrations, which facilitate the discovery of unexpected receptor–ligand pairs. The tissue-extract-based approach, by contrast, requires an extraordinarily sensitive assay procedure and high quality tissue extracts. It is, however, the only technique available to identify previously unknown mediators. The third, information-based approach commences with a proposed physiological relevant ligand. It will neither discover unexpected receptor–ligand pairs, nor reveal novel ligands. However, it reduces practical screening work to a minimum.

### Links

- Database of the IUPHAR Committee on Receptor Nomenclature and Drug Classification: http://www.iuphar-db.org/iuphar-rd/index.html.
- HUGO Gene Nomenclature committee: http://www.gene.ucl.ac.uk/nomenclature/.
- GPCRDB: Information system for G-protein-coupled receptors (GPCRs): http://www.gpcr.org/7tm/.
- Alliance for Cellular Signaling: http://www.signaling-gateway.org/.
- Olfactory Receptor DataBase (ORDB): http://www.senselab.med.yale.edu/senselab/ORDB/.

### Related articles

Wise, A. *et al.* (2004) The identification of ligands at orphan G-protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* 44, 43–66
Cacace, A. *et al.* (2003) An ultra-HTS process for the identification of small molecule modulators of orphan G-protein-coupled receptors. *Drug Discov. Today* 8 (17), 785–792
Im, D.S. (2004) Discovery of new G-protein-coupled receptors for lipid mediators. *J. Lipid Res.* 45 (3), 410–418
George, S.R. *et al.* (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat. Rev. Drug Discov.* 1 (10), 808–820
Milligan, G. (2002) Strategies to identify ligands for orphan G-protein-coupled receptors. *Biochem. Soc. Trans.* 30 (4), 789–793

### Outstanding issues

- How many receptors are activated by more then one ligand under physiological conditions and might we miss important ligands, because their receptors are already identified (with another ligand)?
- How important are oligomerization and receptor associated protein interactions for the overall receptor function?
- How relevant is ligand specific activation of different signal transduction pathways through the same receptor?
- Are there hitherto unrecognized protein-subfamilies, which are G-protein-coupled?

Receptors, which have been deorphanized in the past five years have for the most part, peptides or lipid mediators as ligands. This fact might reflect that the overwhelming majority of 7TM receptors are activated by such substances. However, it is possible that these ligand classes are the ones most easily isolated and characterized with the techniques available at present.

The approach used for a concrete deorphanization project is, as discussed above, dependent on the circumstances. All information regarding the receptor, as well as its suspected ligand(s), has to be carefully analyzed and eventually put into a physiological context.

## Conclusion

The principle of reverse pharmacology remains valid and as a result of its use, roughly 50–60 receptors have been deorphanized in the past five years. For the first time, information is available about all members of the 7TM GPCR family. There is now a reliable estimate of the number of remaining orphan 7TM receptors and their sequence identity. However, not all of the proposed orphan 7TM receptors will be relevant in the process of drug development. Today, the pharmacological significance of the receptors and their proposed ligands is used to identify tomorrow's drug targets.

The identification of receptor–ligand interaction is only the starting point in a demanding mission. To reveal the function of a receptor and to understand its physiological impact are the main goals. And last, but not least, the increasing amount of knowledge should to be used in the development of new and better drugs.

## Acknowledgements

We would like to thank Dr. W. Agace, Dr. H. Kotarsky and S. Ellis Nilsson for reading the manuscript and providing valuable advice.

## References

[1] Drews, J. (2004) Drug discovery: a historical perspective. *Science* 287 (5460), 1960–1964
[2] Vassilatis, D.K. *et al.* (2003) The G protein-coupled receptor repertoires of human and mouse. *Proc. Natl. Acad. Sci. USA* 100 (8), 4903–4908
[3] Fredriksson, R. *et al.* (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63 (6), 1256–1272
[4] Libert, F. *et al.* (1991) Current developments in G-protein-coupled receptors. *Curr. Opin. Cell. Biol.* 3 (2), 218–223
[5] Brown, A.J. *et al.* (2003) The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* 278 (13), 11312–11319
[6] Civelli, O. *et al.* (1999) Orphan receptors, novel neuropeptides and reverse pharmaceutical research. *Brain Res.* 848 (1/2), 63–65
[7] Tunaru, S. *et al.* (2003) PUMA-G and HM74 are receptors for nicotinic acid and mediate its anti-lipolytic effect. *Nat. Med.* 3 (3), 3
[8] Kotarsky, K. *et al.* (2003) A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem. Biophys. Res. Commun.* 301 (2), 406–410
[9] Wise, A. *et al.* (2003) Molecular identification of high and low affinity receptors for nicotinic acid. *J. Biol. Chem.* 9, 9
[10] Milligan, G. *et al.* (1997) Inverse agonism at adrenergic and opioid receptors: studies with wild type and constitutively active mutant receptors. *Receptors Channels* 5 (3/4), 209–213
[11] Pierce, K.L. *et al.* (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell. Biol.* 3 (9), 639–650
[12] Hall, R.A. and Lefkowitz, R.J. (2002) Regulation of G protein-coupled receptor signaling by scaffold proteins. *Circ. Res.* 91 (8), 672–680
[13] Szekeres, P.G. (2002) Functional assays for identifying ligands at orphan G protein-coupled receptors. *Receptors Channels* 8 (5/6), 297–308
[14] Zhang, J.H. *et al.* (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* 4 (2), 67–73
[15] Milligan, G. (2002) Strategies to identify ligands for orphan G-protein-coupled receptors. *Biochem. Soc. Trans.* 30 (4), 789–793
[16] Kostenis, E. (2001) Is Galpha16 the optimal tool for fishing ligands of orphan G-protein-coupled receptors?. *Trends Pharmacol. Sci.* 22 (11), 560–564
[17] Kotarsky, K. *et al.* (2003) Progress in methodology. Improved reporter gene assays used to identify ligands acting on orphan seven-transmembrane receptors. *Pharmacol. Toxicol.* 93 (6), 249–258
[18] Kojima, M. *et al.* (1999) Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* 402 (6762), 656–660
[19] Fujii, R. *et al.* (2002) Identification of a neuropeptide modified with bromine as an endogenous ligand for GPR7. *J. Biol. Chem.* 277 (37), 34010–34016