Ligand-Based Peptide Design and Combinatorial Peptide Libraries to Target G Protein-Coupled Receptors

Christian W. Gruber¹,², Markus Muttenthaler³ and Michael Freissmuth¹,*

¹Institute of Pharmacology, Center of Biomolecular Medicine & Pharmacology, Medical University of Vienna, Waehringer Str. 13a, A-1090 Vienna, Austria, ²Department of Pharmacognosy, University of Vienna, Althanstr. 14, A-1090 Vienna, Austria, ³Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia

Abstract: G protein-coupled receptors (GPCRs) are considered to represent the most promising drug targets; it has been repeatedly said that a large fraction of the currently marketed drugs elicit their actions by binding to GPCRs (with cited numbers varying from 30-50%). Closer scrutiny, however, shows that only a modest fraction of (~60) GPCRs are, in fact, exploited as drug targets, only ~20 of which are peptide-binding receptors. The vast majority of receptors in the human genome have not yet been explored as sites of action for drugs. Given the druggability of this receptor class, it appears that opportunities for drug discovery abound. In addition, GPCRs provide for binding sites other than the ligand binding sites (referred as to the “orthosteric site”). These additional sites include (i) binding sites for ligands (referred to as “allosteric ligands”) that modulate the affinity and efficacy of orthosteric ligands, (ii) the interaction surface that recruits G proteins and arrestins, (iii) the interaction sites of additional proteins (GIPs, GPCR interacting proteins that regulate G protein signaling or give rise to G protein-independent signals). These sites can also be targeted by peptides. Combinatorial and natural peptide libraries are therefore likely to play a major role in identifying new GPCR ligands at each of these sites. In particular the diverse natural peptide libraries such as the venom peptides from marine cone-snails and plant cyclotides have been established as a rich source of drug leads. High-throughput screening and combinatorial chemistry approaches allow for progressing from these starting points to potential drug candidates. This will be illustrated by focusing on the ligand-based drug design of oxytocin (OT) and vasopressin (AVP) receptor ligands using natural peptide leads as starting points.

Keywords: GPCR, drug, ligand-based design, peptide, cyclotide, conotoxin, oxytocin, vasopressin.

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute probably the largest known protein families of cell surface receptors that regulate various cellular processes via signal transduction. The receptors consist of seven transmembrane-spanning α-helices, with an extracellular N-terminus, an intracellular C-terminus and three inter-helical loops on each side of the membrane [1-3]. They recognize extracellular signaling molecules (ligands) of various nature (e.g., neurotransmitters, hormones, growth factors, odorant molecules and light) and size (from small molecules to peptides to large proteins) [4] that elicit a response inside the cell. The signal transmission leads to the activation or inactivation of a particular signaling pathway and hence a specific cellular response.

The human genome contains 791 - 865 genes coding for GPCRs [5-7]. This estimate is subject to continuous revision because of uncertainties and ambiguities in the nature of some genes: in the human genome, for instance, most pheromone receptors of the vomeronasal organ (VR1-class) are pseudogenes. There may also be pseudogenes in the abundant family of odorant receptors. More importantly from the perspective of drug discovery, at least ~150 are still “orphan receptors”: their endogenous ligand is not known; accordingly their physiological role is obscure [8-10]. The family of G protein-coupled receptors has been classified in various attempts according to the IUPHAR scheme [11] or the GPCRDB (database) [12] into four classes of GPCRs in humans; Class A: rhodopsin-like, with over 80% of all GPCRs in humans; Class B: secretin-like; Class C: metabotropic glutamate receptors; and the much smaller Class F: frizzled/smoothened family. Two additional classes (Class D: pheromone receptors; and Class E: cAMP receptors) are only found in non-mammalian species. These classes can be further divided, based on their function and their ligands.

In contrast to plants, animals are highly dependent on GPCRs to afford communication between cells. GPCRs have adapted to bind to a large variety of ligands; the range of endogenous agonists comprises lipids (e.g., prostanoids, lysophosphatidic acid, retinal), sugars (recognized by taste receptors), a bewildering array of volatile organic compounds (recognized by odorant receptors), amino acids (e.g., glutamate), organic acids (e.g., hydroxybutyrate, succinic acid), inorganic ions (e.g., Ca²⁺), nucleosides and nucleotides (e.g., adenosine, ATP, ADP, UDP, UDP), biogenic amines (e.g., dopamine, (nor-)epinephrine), peptides (e.g., secretin, gastrin, vasopressin, oxytocin), large proteins (e.g., TSH/hydrotopin, the gonadotropins). The mode of binding is highly divergent; binding may occur within the hydrophobic core of the receptor (the presumed binding mode for small molecular ligands other than GABA, glutamate and Ca²⁺), on the extracellular face of the receptor (for peptide ligands) or to an extended N-terminal domain (for receptors binding proteohormones, GABA, glutamate and Ca²⁺). In addition, even within topologically related binding sites, the structure of GPCRs accommodates distinct binding modes: X-ray crystal structures are available for two related group A (i.e., rhodopsin-like) receptors in the antagonist-bound state: within the binding pockets of the β₂ and the β₁ adrenergic receptors, the orientation of the antagonists cyanopindolol [13] and carazolol [14], respectively, is parallel to the membrane plane (and perpendicular to the axis of the transmembrane helices. In contrast, in the A₂a-adenosine receptor the binding pocket is displaced towards helices VI and VII (relative to that of the β-adrenergic receptors) and the antagonist (ZM241385, a methylxanthine derivative) is bound in an extended conformation perpendicular to the plane of the membrane [15].

Typically, the “drugability” of GPCRs is highlighted by stressing that 30-50% of all currently registered drugs act on GPCRs [16-18], but these numbers obscure the untapped opportunities: regardless of whether these estimates are based on market share or on the percentage of the total currently approved chemical entities, the
number of receptors that are targeted by these drugs is modest (~60 receptors, see Table 1). It is also evident from Table 1 that only a fraction (i.e., ~20) of those receptors are peptide-binding GPCRs [18]. In fact, “drugability” has not yet been explored for the vast majority of GPCRs. Last but not least, it is worth pointing out that GPCRs have binding sites other than the orthosteric binding site (i.e., the binding site of physiological agonists and competitive antagonists); these include an allosteric site for compounds that modulate agonist binding to the orthosteric site (for review see [19-20]). In addition, the interface between receptor and G protein may also be targeted [21, 22].

Structure- and ligand-based design and the molecular diversity created by combinatorial peptide libraries are not mutually exclusive, but offer in combination the ability for the development of new specifically targeted peptide drugs [23]. Here, we focus on the use of ligand-based combinatorial peptide design for drug discovery to provide an overview on “peptides as ligands for GPCRs”. First, the reader is guided through the basic principles of GPCR signaling - explaining the G protein cycle and the role of GPCR accessory proteins. We will further present an overview of peptide-binding GPCRs and point out how peptides can interfere with GPCR signaling. We devote a section to discuss the role of combinatorial peptide libraries (including naturally-occurring peptide libraries) and ligand-based combinatorial peptide design for the purpose of drug discovery and the development of novel therapeutics. As an applied example we will review some recent work, which was carried out on the design of selective peptide agonists/antagonists for oxytocin and vasopressin receptors.

GPCR SIGNALING AND ACCESSORY PROTEINS - NEW OPPORTUNITIES?

The eponymous term “G protein-coupled receptor” reflects the canonical signal transduction mechanism, namely the flow of information from an extracellular agonist, which activates a given receptor to engage its cognate G protein(s); this generates active subunits that regulate cellular effector systems and thus give rise to a biological response.

For the past 20 years, the G protein cycle has been understood in considerable detail [24]: the G protein cycles between an inactive GDP-bound conformation and an active GTP-bound conformation. Heterotrimeric G proteins consist of three subunits: α, β and γ. The G protein heterotrimer acts as a molecular switch in signal transduction pathways mediated by the GPCR [2] (Fig. 1). Upon receptor activation the GDP is released from the α-subunit; three models have been proposed to explain this GDP-exit mechanism, they are referred to as the (i) lever-arm, (ii) lever-shift and (iii) sequential fit models. The mechanistic details of these models explaining GDP-release have been recently reviewed [25] and the pertinent arguments will not be further considered. Suffice it to say that the GDP-release is brought about by a conformational change of the receptor, which leads to a structural movement in the Gα subunit resulting in a change of the microenvironment of the GDP-binding pocket and the release of GDP. The three models differ in the molecular mechanism of these structural movements that lead to the release of GDP. Once the binding pocket of the nucleotide is empty, the G protein heterotrimer is in its activated conformation and forms a stable, high-affinity complex with the activated receptor. Binding of GTP to the Gα subunit destabilizes the complex, which leads to dissociation of Gα and Gαγ subunits from the receptor and their interaction with downstream effector proteins (E1 and E2). The signal is terminated by the intrinsic GTPase activity of the Gα subunit, which hydrolyses the GTP to GDP. This GTP-GDP “exchange” leads to re-formation of the heterotrimer Gαγ and results in the inactive or basal state of the G protein. In summary, the G protein cycle is defined by a conformational switch of the receptor and hence G protein heterotrimer upon stimulation. The signal is initiated by GDP-GTP exchange and terminated by a conversion of GTP to GDP and the mutual inactivation of Gα and Gβγ. Thus the GDP/GTP-cycle drives the conformational cycle between inactive and active conformation of Gα, which results in a superimposed cycle of subunit association and dissociation.

Several proteins impinge on the G protein cycle by interacting directly with individual G proteins [26]: (i) the intrinsic GTPase of the G protein α-subunit functions as the timed turn-off switch; the rate of GTP-cleavage (kcat) of the intrinsic GTPase is in the range of 2 - 15 min⁻¹ and thus too slow to allow for rapid signaling events (following receptor activation, the turn-off reaction becomes rate-limiting in the cycle; the turn-off reaction, however, also determines the speed at which the agonist-induced response reaches its plateau). Thus, the turn-off reaction is - in most instances - assisted by the eponymous action of GAPs (GTPase activating proteins) (see Fig. 1F). For historical reasons (the proteins were originally identified in functional screens in Saccharomyces cerevisiae and Caenorhabditis elegans), these GAPs are referred to as RGS proteins (regulators of G proteins signaling) [27, 28]. Additional proteins are (ii) non-receptor guanine nucleotide exchange factors (GEFs) and (iii) guanine nucleotide dissociation inhibitors (GDIs) [26, 29], which modulate the G protein cycle in opposite ways (Fig. 1F).

There are many different isoforms or closely related subtypes of G proteins. In mammalian organisms, for example, there are 20 Gα, 5 Gβ and 12 Gγ isoforms assembled in various combinations to generate a diverse population of G protein heterotrimers [26, 30]. In spite of this combinatorial diversity, G proteins are classified according to their Gα-subunit because this highlights the principal signaling mechanism (e.g., the isoforms of Gαs and Gαi stimulate and inhibit, respectively, adenyl cyclases and the production of cAMP, those of Gα12/13 stimulate phospholipase C and the breakdown of PIP2; Gαq/11 Rho-mediated cytoskeletal rearrangements). It is evident that this approach is an oversimplification, because it ignores the contribution of Gβγ.

However, GPCRs do not only trigger cellular response via the canonical signaling pathway, i.e., via the G-protein activation/inactivation cycle. The agonist-activated GPCR is subject to phosphorylation by regulatory kinases (GRK1-6, G protein-coupled receptor kinases). Phosphorylation triggers recruitment of arrestins, which associate with the phosphorylated receptor. This interaction precludes the recruitment of G protein and thus leads to desensitization of G protein-dependent signaling [31]. During the past decade, it was appreciated that upon internalization, the complex of GPCR and arrestin triggers a second round of signals that involves non-receptor tyrosine kinases of the SRC-family, MAP kinase family members (ERK1/2, jun-N-terminal kinase, p38 MAP kinase etc.) and regulators of small G proteins [32]. In this context, it is interesting to note that (partial) agonists can be identified that bias the receptor conformation; i.e., they promote recruitment of arrestins although they do not stimulate the canonical (i.e., G protein-dependent signaling) pathway [33]. These observations were generated by using the β2-adrenergic receptor, which has a rich pharmacology: thus a long list of compounds is available that can systematically be tested for their ability to raise cAMP and to support stimulation of ERK1/2 phosphorylation. However, it is very likely that similar bias for agonists will be discovered with many other receptors. In fact, desensitization of μ-opioid receptor is an instructive example: peptide agonists act as full agonists that efficaciously trigger signaling, desensitization and internalization; morphine and several semi-synthetic alkaloids are agonists that also give rise to desensitization, but they are poorly effective in triggering internalization. While these observations were originally made in transfected cells, they have been recapitulated ex vivo (i.e., in slice preparation) in neurons indicating that these differences also exist in the native environment [34].

In addition, GPCRs can interact with numerous proteins other than their cognate G protein(s), GRKs and arrestins [35-37]. These
### Table 1. Currently Targeted G Protein-Coupled Receptors

<table>
<thead>
<tr>
<th>G Protein-Coupled Receptor(s)</th>
<th>Representative Targeting Drug(s) (Generic Name)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>receptors for amines, lipid mediators and nucleosides/nucleotides:</strong></td>
<td></td>
</tr>
<tr>
<td>(A_1), (A_{2A})-adenosine receptor</td>
<td>caffeine, theophylline</td>
</tr>
<tr>
<td>(\alpha_{1A,B,C})-adrenergic receptors</td>
<td>prazosin, doxazosin, terazosin, tamsulosin</td>
</tr>
<tr>
<td>(\alpha_{2A,B,C})-adrenergic receptors</td>
<td>clonidine, moxonidine</td>
</tr>
<tr>
<td>(\beta_1)-adrenergic receptor</td>
<td>bisoprolol, metoprolol, atenolol, esmolol</td>
</tr>
<tr>
<td>(\beta_2)-adrenergic receptor</td>
<td>salbutamol, salmeterol, fenoterol, formoterol</td>
</tr>
<tr>
<td>CB1-cannabinoid receptor</td>
<td>rimonabant</td>
</tr>
<tr>
<td>(D_1, D_3, D_5)-dopamine receptors</td>
<td>L-DOPA</td>
</tr>
<tr>
<td>(D_2)-receptor</td>
<td>pramipexol, ropinirol, haloperidol, fluphenazine, sulpiride</td>
</tr>
<tr>
<td>(D_3)-receptor</td>
<td>clozapine</td>
</tr>
<tr>
<td>GPR109a= HM74 = PUMA-G</td>
<td>nicotinic acid, acipimox</td>
</tr>
<tr>
<td>(H_1)-histamine receptor</td>
<td>diphenhydramine, dimetindene, desloratadine</td>
</tr>
<tr>
<td>(H_2)-histamine receptor</td>
<td>cimetidine, ranitidine</td>
</tr>
<tr>
<td>(5HT_{1a})-serotonin receptor</td>
<td>buspirone</td>
</tr>
<tr>
<td>(5HT_{1b,d})-receptors</td>
<td>sumatriptan, zolmitriptan</td>
</tr>
<tr>
<td>(5HT_{2A,C})-receptors</td>
<td>risperidone, quetiapine, olanzapine</td>
</tr>
<tr>
<td>(5HT_2)-receptor</td>
<td>cispamide, metoclopramide</td>
</tr>
<tr>
<td>(M_1)-muscarinic receptor</td>
<td>pirenzepine</td>
</tr>
<tr>
<td>(M_{2,3,4})-muscarinic receptors</td>
<td>atropine, ipratropamide</td>
</tr>
<tr>
<td>(M_4)-muscarinic receptor</td>
<td>tiotropium</td>
</tr>
<tr>
<td>(MT_1,2)-melatonin-receptor</td>
<td>agomelatine</td>
</tr>
<tr>
<td>CysLT1-cysteinyl-leukotriene-receptor</td>
<td>montelukast, zafirlukast</td>
</tr>
<tr>
<td>(EP_2)-prostanoid receptor</td>
<td>misoprostol</td>
</tr>
<tr>
<td>(EP_4)-prostanoid receptor</td>
<td>alprostadil (PGE1)</td>
</tr>
<tr>
<td>(FP_{1,2,3})-prostanoid receptor</td>
<td>latanoprost, PGF2(\alpha)</td>
</tr>
<tr>
<td>IP prostaglandin (I_2)-receptor</td>
<td>treprostinil, epoprostenol</td>
</tr>
<tr>
<td>DP(1)-prostaglandin (D_2) receptor</td>
<td>laroniprant</td>
</tr>
<tr>
<td>P2(y_1)-purinergic receptor</td>
<td>clopidogrel</td>
</tr>
<tr>
<td>rhodopsin</td>
<td>retinal/vitamin A (for night blindness)</td>
</tr>
<tr>
<td>S1P(_{1,3,4,5})-sphingosin-1-phosphate receptors</td>
<td>fingolimod</td>
</tr>
<tr>
<td><strong>“flytrap receptors”:</strong></td>
<td></td>
</tr>
<tr>
<td>(Ca^{2+})-sensing receptor</td>
<td>cinacalcet</td>
</tr>
<tr>
<td>GABA(_{A})-receptor (1/2)</td>
<td>baclofen</td>
</tr>
<tr>
<td><strong>receptors for peptides/proteohormones:</strong></td>
<td></td>
</tr>
<tr>
<td>angiotensin-II-receptor type 1</td>
<td>losartan, candesartan, telmisartan</td>
</tr>
<tr>
<td>(B_1)-bradykinin receptor</td>
<td>Icatibant</td>
</tr>
<tr>
<td>calcitonin-receptor</td>
<td>calcitonin</td>
</tr>
</tbody>
</table>
so-called GPCR interacting proteins (GIPs) can elicit specific cellular responses, independently of the G protein and/or the arrestin-mediated cellular signaling. A case in point is the A2A-adenosine receptor, where stimulation of mitogen-activated protein kinase/ERK1/2 is independent of the cognate G protein Gs [38] and of any other heterotrimeric G protein [39], but requires ARNO (ARF-nucleotide binding site opener, the guanine nucleotide exchange factor of the monomeric G protein ARF6) and ARF6 [40, 41]; ARNO binds directly to the C-terminus of the receptor. However, in many instances, the significance of the interaction between a given receptor and the GIPs is not fully understood. It is specifically not clear to what extent these interactions contribute to the signal transduction elicited by the receptor and what the relative role of each interaction is. There have only been a few attempts to characterize the interactome of GPCRs, in particular at the receptors C-terminus. Two strategies have been reported: (i) using the isolated C-terminus as bait in pull-down experiments coupled to LC-MS/MS identification; e.g., for the interactomes of the 5-hydroxytryptamine receptor [42] and the melanotonin receptors [43] or (ii) tandem affinity purification of tagged receptor from recombinant cells coupled to MS identification [44]. These surveys enabled a first glimpse of possible GIPs for two GPCRs. The “interactome” of other GPCRs as well as the mechanism of these interactions, their regulation and their exact physiological role is not yet understood. It is therefore of interest to characterize the interactome of GPCRs, in particular at the receptors C-terminus. Two strategies have been reported: (i) using the isolated C-terminus as bait in pull-down experiments coupled to LC-MS/MS identification; e.g., for the interactomes of the 5-hydroxytryptamine receptor [42] and the melanotonin receptors [43] or (ii) tandem affinity purification of tagged receptor from recombinant cells coupled to MS identification [44]. These surveys enabled a first glimpse of possible GIPs for two GPCRs. The “interactome” of other GPCRs as well as the mechanism of these interactions, their regulation and their exact physiological role is not yet understood. It is therefore of interest to characterize the (i) list of accessory proteins of GPCRs, in particular in vivo, (ii) how the interactome of the receptor changes in the presence of agonists/antagonists and within various tissues and (iii) identify the signaling pathways that are regulated by these additional interactors.

In principle, G proteins – like any other protein interacting with multiple partners – can act (i) as signal integrators (two or more competing signals impinge on the receptor and the output is determined by the relative strength of each input), (ii) coincidence detectors (two or more signals must be present simultaneously to elicit a response) and (iii) molecular switches (in the presence of interactor A, signal 1 is generated; in the presence of interactor B, a different effector is activated). Charting the interactome of individual GPCRs and its functional network may translate into opportunities for chemical biology and drug discovery. It is obvious that blockage of the protein-protein interaction surfaces will elicit biological effects that are distinct from those that can be triggered by agonists or antagonists acting at the receptor ligand binding site.

PEPTIDES AS LIGANDS FOR GPCRS AND MODULATORS OF SIGNALING

Mechanisms by which Peptides can Interfere with GPCR Signaling

Generally there are at least five mechanisms by which peptides can interfere with GPCR signaling. These are:
(I). as orthosteric ligands of GPCRs,
(II). as allosteric ligands of GPCRs that enhance or block the action of the orthosteric ligand (for review see [19 20]),
(III). at the receptor-G protein interface,
(IV). at the interface of the receptor with other accessory proteins and
(V). by direct interaction with G proteins.

As mentioned above, there are numerous examples of peptides that act as natural ligands for GPCRs (see Tables 1 and 2). These peptide ligands can either stimulate the receptor and thus act as agonists or inhibit receptor-induced signaling and thus act as an-
tagonists. Besides the direct interaction of the peptides with the receptor as ligands on the extracellular side, peptides can interfere with the receptor on the intracellular side. By this interaction they may block the recruitment of the G protein and hence inhibit signaling. This has been shown, for instance, for synthetic peptides mimicking the C-terminal tail of G₁₅ subunits and hence blocking an efficient binding of the G proteins to the receptor (summarized in [21, 22, 29]). In similar fashion peptides could block interactions of the receptor with other accessory proteins. GPCR interacting proteins (GIPs) and their respective receptors are dynamic signaling complexes that form differentially, depending on the cellular signal status, thus allowing therapeutic intervention. This approach has for example been used to inhibit protein interactions of the metabotropic glutamate receptor with the aim of preventing neurodegenerative diseases [46, 47]. Last but not least there are many possibilities how peptides/ligands can directly interfere with the G protein side of GPCR signaling (reviewed in [21, 22, 29]), for instance by binding to either G protein subunit (G₁₅, β and γ) and inhibiting the formation of the heterotrimer or by acting as GEFs, GAPs or GDIs. Obviously, the above-mentioned mechanisms of signal interception are only viable for in vivo pharmacological applications if the peptides can cross cellular membranes. In instances where this is not feasible, the use of small organic molecules may be the better alternative [48, 49]. It is likely that a collection of new ligands will emerge because high-throughput assays have been developed to screen for peptides that bind selectively to different conformations of G protein subunits [50] or target the interface with a specific subset of RGS proteins [51].

GPCRs that Bind Peptides and Proteins

Of the ~800 distinct GPCRs, the list of known peptide- or protein-targeted GPCRs is readily comprehensible with “only” a few hundred [18]. Nevertheless peptide- and protein-GPCRs are
Table 2. G Protein-Coupled Receptors for Peptides and Proteins

<table>
<thead>
<tr>
<th>Receptor Family</th>
<th>Receptor Subtypes</th>
<th>Cellular/Biological Function</th>
<th>Natural Ligand</th>
<th>Nature of Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensin receptors</td>
<td>AT1, AT2</td>
<td>Blood pressure, water and electrolyte homeostasis; tissue development and repair</td>
<td>Angiotensin</td>
<td>Peptides (8-22 amino acids)</td>
</tr>
<tr>
<td>Apelin R</td>
<td>APJ</td>
<td>Cardiovascular, fluid homeostasis</td>
<td>Apelin</td>
<td>Peptides (13/36 aa)</td>
</tr>
<tr>
<td>Bombesin R</td>
<td>BB1, BB2, BB3</td>
<td>Wide range of physiological and pathological functions</td>
<td>neuromedin B, gastrin-releasing peptide</td>
<td>Peptides</td>
</tr>
<tr>
<td>Bradykinin R</td>
<td>B1, B2</td>
<td>Blood vessel vasodilatation</td>
<td>Bradykinin</td>
<td>Small peptide (9 aa)</td>
</tr>
<tr>
<td>Calcitonin R</td>
<td>AMY1-3, AM1.2, CT, CALCRL, CGRP</td>
<td>Calcium and phosphorus metabolism</td>
<td>amylin, adrenomedullin, CGRP</td>
<td>Peptide (32 aa)</td>
</tr>
<tr>
<td>Chemokine R</td>
<td>CXCR1-5, CCR1-10, CX3CR1, XCR1, CXCR6</td>
<td>Immunity</td>
<td>CXCL1-3, 5-13,16, CCL …, CX3CL1, XCL1,2, macrophage derived lectin</td>
<td>Proteins</td>
</tr>
<tr>
<td>Cholecystokinin R</td>
<td>CCK1.2</td>
<td>Gastrointestinal tract hormone, digestion</td>
<td>choleystokinin, gastrin</td>
<td>Peptides (17/33 aa)</td>
</tr>
<tr>
<td>Corticotropin releasing factor R</td>
<td>CRF1.2</td>
<td>Stress response hormone, CNS</td>
<td></td>
<td>Peptide (41 aa, α-helical)</td>
</tr>
<tr>
<td>Endothelin R</td>
<td>ETA, B</td>
<td>Vasoconstriction of smooth muscle cells</td>
<td>endothelin 1-3</td>
<td>Peptides (21 aa)</td>
</tr>
<tr>
<td>Formylpeptide R</td>
<td>FPR, FPRL1.2</td>
<td>Chemotaxis</td>
<td>N-formyl peptides</td>
<td>Amino acids and peptides with modifications (N-formyl)</td>
</tr>
<tr>
<td>Frizzled R</td>
<td>-</td>
<td>Development, cell proliferation</td>
<td>Wnt-family lipoglycoproteins</td>
<td>Proteins</td>
</tr>
<tr>
<td>Galanin R</td>
<td>GAL1-3</td>
<td>Neuropeptide, regulation of neurotransmitter release</td>
<td>galanin</td>
<td>Peptides (30 aa)</td>
</tr>
<tr>
<td>Ghrelin R</td>
<td>GRLN</td>
<td>Peptide hormone, involved in digestion and obesity (see ghrelin)</td>
<td>ghrelin</td>
<td>Peptides (modified)</td>
</tr>
<tr>
<td>Glucagon R</td>
<td>GHRH, GIP, GLP1.2, Glucagon, Secretin</td>
<td>Peptide hormone, blood-sugar regulation</td>
<td>glucagon, secretin</td>
<td>Peptides</td>
</tr>
<tr>
<td>Glycoprotein hormone R</td>
<td>FSH, LH, TSH</td>
<td>Reproduction, Development</td>
<td>follicle-stimulating hormone (FSH), luteinizing hormone (LH), chorionic gonadotropin, thyroid-stimulating hormone</td>
<td>Proteins</td>
</tr>
<tr>
<td>Gonadotropin releasing hormone R</td>
<td>GNRH, GNRH2</td>
<td>Neuropeptide hormone, release of FSH and LH</td>
<td>gonadotropin releasing hormone</td>
<td>Peptide/protein</td>
</tr>
<tr>
<td>KiSS1 derived peptide R</td>
<td>KiSS1</td>
<td>Regulation of endocrine function</td>
<td>kissepeptin</td>
<td>Peptide</td>
</tr>
<tr>
<td>Melanocortin R</td>
<td>MC1-5</td>
<td>Peptide hormone; appetite and sexual arousal</td>
<td>α, β, γ-melanocyte stimulating hormone</td>
<td>Peptide (39 aa)</td>
</tr>
<tr>
<td>Motilin R</td>
<td>MTLN</td>
<td>Peptide hormone, digestion (see ghrelin)</td>
<td>motilin</td>
<td>Peptide (22 aa)</td>
</tr>
<tr>
<td>Neurotensin R</td>
<td>NTS1.2</td>
<td>Peptide hormone, digestion, CNS</td>
<td>neurotensin</td>
<td>Peptide (13 aa)</td>
</tr>
<tr>
<td>Neuromedin U R</td>
<td>NMU1.2</td>
<td>Bombesin-related function</td>
<td>neuromedin</td>
<td>Peptide (8/25 aa)</td>
</tr>
<tr>
<td>Receptor Family</td>
<td>Receptor Subtypes</td>
<td>Cellular/Biological Function</td>
<td>Natural Ligand</td>
<td>Nature of Ligands</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Neuropeptide S R</td>
<td>NPS</td>
<td>Behavioral functions, linked to asthma susceptibility</td>
<td>neuropeptide S</td>
<td>Peptide</td>
</tr>
<tr>
<td>Neuropeptide Y R</td>
<td>NPY1,2,4,5</td>
<td>neurotransmitter</td>
<td>neuropeptide Y, pancreatic polypeptide</td>
<td>Peptide (36 aa)</td>
</tr>
<tr>
<td>Neuropeptide B/W R</td>
<td>NPBW1,2</td>
<td>Regulation of feeding, obesity</td>
<td>neuropeptide W, B</td>
<td>Peptides</td>
</tr>
<tr>
<td>Neuropeptide FF/AF</td>
<td>NPFF1,2</td>
<td>Several physiological functions</td>
<td>neuropeptide FF, AF</td>
<td>Peptides</td>
</tr>
<tr>
<td>Opioid R</td>
<td>μ-, κ-, δ-OP, NOP</td>
<td>Analgesic</td>
<td>β-endorphin, dynorphin A, nociceptin/orphanin FQ</td>
<td>Opioid peptides</td>
</tr>
<tr>
<td>Orexin R</td>
<td>OX1,2</td>
<td>Neuropeptide hormone, food intake, energy metabolism</td>
<td>orexin A, B</td>
<td>Peptides (28/33 aa)</td>
</tr>
<tr>
<td>Parathyroid hormone R</td>
<td>PTH1,2</td>
<td>Calcium metabolism</td>
<td>parathyroid hormone, TIP-39</td>
<td>Peptide/protein (84 aa)</td>
</tr>
<tr>
<td>Peptide P518 R</td>
<td>QRFP</td>
<td>Orphan receptor (neuroendocrine system)</td>
<td>RF-amide P518 gene product</td>
<td>Amidated peptides</td>
</tr>
<tr>
<td>Prokineticin R</td>
<td>PK1,2</td>
<td>Contraction of gastro-intestinal smooth muscle</td>
<td>prokineticin</td>
<td>Cysteine-rich proteins (81/86 aa)</td>
</tr>
<tr>
<td>Prolactin releasing peptide R (GPR10)</td>
<td>PRRP</td>
<td>processing of nociceptive signals in the brain</td>
<td>prolactin releasing peptide</td>
<td>Peptides (20/31 aa)</td>
</tr>
<tr>
<td>Relaxin family peptide R</td>
<td>RXFP1-4</td>
<td>Reproduction</td>
<td>relaxin</td>
<td>Peptides</td>
</tr>
<tr>
<td>Somatostatin R</td>
<td>sst1-5</td>
<td>inhibition of growth hormone release and modulation of neuronal activity</td>
<td>somatostatin, cortistatin</td>
<td>Cyclic peptides 14/28 aa (somatostatin); 17 aa (cortistatin)</td>
</tr>
<tr>
<td>Tachykinin R</td>
<td>NK1-3</td>
<td>behavioral responses; potent vaso dilators; contraction (directly or indirectly) of many smooth muscles</td>
<td>substance P, neurokinin A, neurokinin B</td>
<td>Peptides (10-12 aa)</td>
</tr>
<tr>
<td>Thyrotropin releasing hormone R</td>
<td>TRH1</td>
<td>release of thyroid-stimulating hormone and prolactin by the anterior pituitary</td>
<td>thyrotropin releasing hormone</td>
<td>Tripeptide</td>
</tr>
<tr>
<td>Urotensin R</td>
<td>UT</td>
<td>Neurosecretory system</td>
<td>urotensin II</td>
<td>Cyclic dodecapeptide</td>
</tr>
<tr>
<td>Vasopressin and oxytocin R</td>
<td>V1a, V1b, V2, OT</td>
<td>See text</td>
<td>vasopressin, oxytocin</td>
<td>Small peptides (cyclic, disulfide-bonds)</td>
</tr>
<tr>
<td>VIP and PACAP R</td>
<td>PAC1, VPAC1,2</td>
<td>circadian rhythms, pancreatic insulin secretion, control of immunity and inflammation</td>
<td>vasoactive intestinal peptide (VIP) and pituitary adenylate cyclase-activating polypeptide (PACAP)</td>
<td>Peptides (VIP 28 aa, PACAP 27/38 aa)</td>
</tr>
</tbody>
</table>

Information extracted from [11]; without consideration of orphan GPCRs

involved in the regulation of basic and indispensable physiological and cellular processes. Table 2 summarizes the currently known GPCRs, their cognate physiological agonist, which are peptides and proteins (as extracted from the IUPHAR database [11]). For the sake of simplicity, the 19 identified chemokine receptors and their 44 endogenous ligands were not included. It is emphasized that peptide-binding GPCRs are spread over all GPCR subfamilies, but the majorities are members of the rhodopsin and secretin classes.

A recent review has examined ~120 GPCRs and their respective peptide ligands with respect to structural properties of ligand binding [18]. This comparison has highlighted common ligand binding motifs allowing receptor recognition: these GPCRs apparently recognize (broadly defined) peptide turn conformations in their diverse peptide ligands. This information may help to better understand peptide-mediated binding and activation of GPCRs and will be discussed in more detail below.

**COMBINATORIAL PEPTIDE LIBRARIES AND DESIGN OF PEPTIDES AS LIGANDS FOR GPCRS**

**Common Structural Motifs of GPCR-Binding Peptides**

The majority of registered pharmaceuticals that act on GPCRs are derived from ligand-based drug design strategies that were originally based on structure-activity relationships (SARs) of the endogenous ligand and derivatives thereof that interacted with the
receptor. Future drugs are still likely to rely on ligand-based drug design due to the limited availability of structural data on GPCRs. It is unlikely that structural information will be available for the majority of GPCRs, regardless if the number is limited to the non-sensory GPCRs with known ligands (180 out of 210) that have not been targeted yet or whether the > 150 “orphan” receptors are included. For these latter receptors different strategies have been devised, which generally involve the screening of large combinatorial libraries and in silico molecular modeling.

Many of the known endogenous ligands have been studied extensively and common structural binding motifs have been identified [18, 52, 53]. At the primary structure level similar pattern of amino acid sequences are found in secretin, glucagons, growth hormone-releasing hormone, glucose-dependent insulinotropic polypeptide, glucagon-like-peptide 1 and 2 [54]. An even richer source of recognition motifs can be found at the secondary structure level, particularly considering that information content in proteins/peptides is evolutionary more conserved through three-dimensional structures rather than through linear amino acid sequences [55]. The main structural motif identified is the turn [18]. A turn may be defined by 3 residues (γ-turn), 4 residues (β-turn) and 5 residues (α-turn) (see Fig. 2). These can form 7-, 10- and 13-membered hydrogen bonded rings, respectively. Many of these turn structures are found to be stabilized by cyclic ring and loop moieties, in particular in the case of smaller and more flexible peptides that require conformational stabilization to maintain a rigid three-dimensional structure. Examples of such cyclic peptides targeting GPCRs are the calcitonins, chemokines, endothelins, melanin-concentrating hormone, oxytocin, relaxins, somatostatin, vasopressin and urotensin II. Recognition of turn motifs generally only involves interactions of the spatially-orientated side chain residues of the ligand with the receptor and they can therefore be considered as scaffolds, which could theoretically be substituted by alternative rigid non-peptidic scaffolds that maintain the functional side chains in the right conformation. This field of peptidomimetics has been thoroughly reviewed and the reader is directed to some key articles [56-59]. Certainly, these common recognition motifs (Fig. 2) can be used as well-defined starting points for ligand-based drug design that can lead with the help of combinatorial chemistry to novel bioactive peptides as well as non-peptidic entities.

Similar to the peptide recognition motifs, so called privileged structures have been identified within small molecule drugs. A privileged structure is considered to be “a single molecular framework able to provide ligands for diverse receptors” [60]. The concept originated from the study of 1,4-benzodiazepin-2-ones that bind to cholecystokinin, gastrin and central benzdiazepine receptors. In addition, the benzodiazepine substructure is found in neuropeptide Y antagonists, κ-secretase inhibitors, farnesyl: protein trans-

![Fig. (2). Common structural recognition motifs of peptides targeting GPCRs.](image-url)

(A) α-Helix of the human parathyroid hormone [199]. (B) Type II β-turn of deamino-oxytocin [180]. (C) Stromal cell-derived factor-1α (SDF-1α), a member of the chemokine superfamily, exhibiting β-sheets and helices as structural motifs [200].
ferase inhibitor and ion channel ligands. Privileged structures are generally identified by virtual screening of the structures of approved drugs employing various filters and algorithms. Once a privileged structure has been identified, it can be utilized as a scaffold for the design of drug-like libraries that can be screened against other GPCR targets. In this way many small organic privileged substructures have been identified and their use in library design has been extensively reviewed [53, 61-64]. Interestingly, some of them, including benzodiazepines have been reported to be β-turn mimetics [65, 66] and it is speculated that the multiple receptor binding properties are mainly due to the specific orientation of the attached side chains similar to above discussed peptide turn motifs [53]. It can therefore be argued that the term privileged structure can be extended to common structural recognition motifs, such as the turn structures [67, 68].

Ligand-Based Peptide Design and Combinatorial Approaches

Over the last two decades the production of efficient and high quality combinatorial libraries has been one of the most rapidly developing fields in the pharmaceutical industry driven by the unmet compound supply for high throughput screening and the vast number of novel drug targets derived from genome projects across the world [69]. The promise of a seemingly inexhaustible supply of compound libraries with drug-like properties resulted in global down sizing of the more tedious (time intensive, expensive, structurally and synthetically challenging, difficult to scale-up, unclear intellectual property, etc.) natural product discovery divisions in most pharmaceutical companies [69]. It is now considered to have been a premature decision: the vast, random compound libraries did not produce the results initially anticipated; in fact, there is only one FDA-approved drug that emerged from this approach (Sunitinib™ for renal carcinoma in 2005 [69-71]). Nonetheless, the techniques developed in the field of combinatorial chemistry certainly revolutionized drug development. Highly automated solid-phase synthesis, high yield reactions, convergent scaffold approaches in combination with split and pool or parallel techniques increased the efficiency and high yield reactions, convergent scaffold approaches in combination with split and pool or parallel techniques increased the efficiency of SAR studies and optimization of lead compounds drastically. It has been estimated that 10^6 small drug-like molecule compounds or 20^50 (10^6) 50-residue peptides (not including different folds and conformers) could in principle be synthesized [72-74]. Realistically, this chemical space can neither be covered nor characterized systematically, with the largest libraries at present only containing up to 10^6 compounds. Conversely, biological relevant space is restricted to biosynthesis and natural amino acids, and functional domains are found to be conserved in their three-dimensional structure rather than through their underlying sequences [55, 75]. Consequently, biologically relevant space explored by nature during evolution is highly conserved and strongly limited in size compared to the possible chemical space [74, 76, 77]. It is therefore not surprising that random libraries based on commercially available reagents do not often overlap with structurally complex biological-relevant space. This realization resulted in a paradigm shift in library design to more focused, rational designed and smaller libraries that have an enhanced likelihood of obtaining biologically active compounds.

Two main synthetic approaches emerged: biology-orientated synthesis (BIOS) and diversity-orientated synthesis (DOS) [77]. BIOS libraries primarily employ combinatorial techniques to build compounds around privileged structures or known pharmacophores. In contrast, DOS libraries rely on methods that generate structurally complex and natural product-like compounds: the aim is to probe chemical space by structural and functional diversity. Both approaches are not mutually exclusive. DOS libraries can be based on privileged structures and BIOS libraries can be generated around scaffolds derived from DOS. The compounds of both library designs can be screened against a wide range of biological targets, considering that privileged structures interact with multiple proteins differently. These two interactive drug discovery platforms in combination with the rapid advances in computational modeling and in silico screening are powerful tools in the quest for novel GPCR modulators or in general new bioactive entities.

Implementing recognition motifs in library designs has already yielded novel leads. For example, cyclic heterohedral penta- and hexapeptides that adopt a βII/γ-turn motif were used as conformational scaffolds for probing receptor recognition, where a recognition motif (such as Arg-Gly-Asp) was systematically shifted around the cyclic peptide backbone structure to spatially sample various conformations, leading to the identification of a vittneoin binding inhibitor to the α₁β3-inteumin [78, 79]. Micromolar non-peptidic heterocyclic agonists associated with the melanocortin-1 receptor were discovered by screening a library of 951 compounds based upon a β-turn motif [80, 81].

Another interesting approach takes advantage of “turn-scans”: an initial library of 2 x 10^7 peptides with a β-turn constrained mimetic block in their centre was screened against the human opioid receptors µ, δ, and κ, and the opioid receptor like ORL1 [82]. The initial hits underwent optimization towards affinity and selectivity by exchanging the original β-turn mimetic moiety with other turn mimetics (“turn-scan”), which resulted in identification of a novel class of opioid ligands that act as inverse agonists as well as peptide antagonist on the k-receptor and possess different selectivity profiles. Considering biodurability shortcoming of peptides as drugs, the advantages of this approach towards drug development are evident: incorporation of peptidomimetics design within the library cuts out the challenging process of transferring the biological relevant peptide/protein topology into a small molecule structure.

The value of any drug discovery strategy can be evaluated by the quality and quantity of drug leads produced, with “quality” being associated to the drug-like properties of the hits (such as pharmacokinetics, toxicity, ADMET, i.e., Absorption-Distribution-Metabolism-Excretion), their specificity, potency, synthetic availability, intellectual property and the target novelty, while “quantity” depends on the hit success rate and how rapidly a hit can be moved into the clinic. No doubt that combinatorial chemistry has changed the drug discovery landscape and time will tell whether these new approaches can use combinatorial chemistry in combination with rational library design to their advantage to overcome the early setbacks of random libraries and to provide the new generation of pharmaceuticals.

Peptide Libraries

Attempts to indentify privileged structures of GPCR ligands - as pointed out above - can lead to some success in the development of new and optimized peptide drugs. The lack of three-dimensional receptor structures was a major limitation to structure-based drug design. Recently several X-ray crystal structures of GPCRs with their bound ligands have been reported [13, 15, 83-87]. Although these reports only include small-molecule but not peptide ligands, the structural information can be used in bioinformatics and molecular modeling approaches to advance the field of structure- and ligand-based peptide design as targets for GPCRs [88].

Alternatively, peptide libraries offer a unique way to screen the natural diversity of peptide-protein interactions for the modification of GPCR signaling response and efficiency [23, 89] and drug lead discovery. Natural product libraries have been applied to oligonucleotides, synthetic oligomers, oligosaccharides, small molecules proteins and peptides (summarized in [90]). Naturally occurring peptides and synthetic peptides derived from portions of receptors, G proteins and effector proteins have been used in the past to study protein-peptide and receptor-peptide interactions and how these peptides can interfere with GPCR signaling [29, 91]. Several approaches are known to generate molecular diversity through synthetic or biological (encoded) peptide libraries [16, 23, 29, 89, 92, 93]. A combinatorial library approach generally involves three main steps: (i) preparation of the library, (ii) screening of the components
(peptides) and (iii) identification and characterization of the active components [90]. The following section gives a brief general introduction to combinatorial peptide library approaches.

**Encoded and Synthetic Approaches**

Biological (also known as encoded or DNA-based libraries) include mainly mRNA display, phage display and ribosome display techniques (summarized in [89, 94]). A typical experimental library setup includes: (i) construction of a DNA library, (ii) expression of the library using the above mentioned methods, (iii) affinity selection against an immobilized target and (iv) amplification of the recovered nucleotide sequence to identify the sequence corresponding to the functional peptide. In vitro methods that do not require an in vivo transformation step, such as ribosome display, exhibit the greatest library sizes (>10^{12} unique peptides). Typical DNA-based (selection) libraries such as phage, bacteria, yeast, insect and mammalian cells have smaller library sizes (10^{8} – 10^{9}), but exhibit higher biosynthetic capabilities [29, 94]. Furthermore, mRNA display has several advantages, such as higher complexity of the library and the possibility to identify high affinity sequences and is therefore the preferred biological library method [29, 95].

Chemical (or synthetic) combinatorial peptide library approaches include positional scanning libraries, mixture based oriented peptide libraries and one bead-one peptide (or split pool) solid phase libraries [23, 89, 90, 96]. Synthetic libraries have been successful to identify bioactive peptides, including antimicrobial peptides, ligands for cell surface receptors, protein kinase inhibitors, peptide binding to MHC molecules, peptide mimetics of receptor binding sites (summarized in [23]). Furthermore there are some examples to use this approach for the identification of GPCR ligands, in particular α-MSH antagonists [90, 97], opioid receptor antagonists [98] and ligands for the D_{2} dopamine receptor [90, 99].

Both approaches, biological and chemical libraries, are equally powerful tools in drug discovery, but they differ specifically in the diversity (number of different molecules, library size), incorporation of modified and non-proteinogenic amino acids (only possible in chemical libraries), codon degeneracy and hence bias towards amino acids (in biological libraries), higher synthesis capabilities and cost efficiency (for biological libraries) [94]. Overall, encoded and synthetic peptide library approaches play an important role in the identification of drug leads and drug development for GPCRs and will remain a powerful tool for basic research and molecular recognition studies [23, 90].

**Natural Peptide Libraries**

The diversity in nature has long been and still is one of the biggest resources of pharmacological lead compounds. Many natural products often exhibit biological activity against unrelated biological targets, thus providing researchers with starting points for drug development [69, 70, 100, 101]. Natural peptides of great number and diversity occur in all organisms from microbes to man (see Table 3). One such rich and yet largely untapped library of bioactive compounds can be found in venom peptides. Rich sources of venom peptides can be found in spiders [102, 103], scorpions [104], snakes [105-107] and marine animals, in particular cone-snails [108-110], to name only a few. Evolutionary pressures have afforded a pre-optimized, structurally sophisticated collection of disulfide-rich peptide toxins that have been produced in a combinatorial fashion and fine-tuned over million of years. Hence it does not come as a surprise that within the venom of these animals structural scaffolds are found that give rise to a very large number of agonists and antagonists, which act on functionally diverse targets, e.g., ion channels, transporters and GPCRs (see Fig. 3). In particular the well-studied “three-finger” scaffold found within the snake neurotoxins (Fig. 3A) highlights the functional versatility of this defined fold: it is able to act on unrelated receptor families with high potency. Another group of extensively studied peptides can be found in the venom of the predatory marine cone snail (genus Conus).

These conotoxins have become a rich source for highly potent and subtype selective molecular probes for neuroscience [108, 110, 111]. Conotoxins are between 6 and 50 amino acids in length, they possess a highly conserved cysteine framework. The latter gives rise to well-defined three-dimensional structures containing many of the structural motifs present in proteins, including α-helices, β-sheets and β-turns. In total only approximately 100 conotoxins have been characterized so far - a mere 0.2 % of the estimated 50,000 biologically active peptides. Nevertheless, this rather small sample has already afforded a drug of proven clinical utility (ziconotide, a derivative of α-conopeptide MVIIA marketed under the name Prialt®), several pre-clinical leads for CNS disorders and various valuable tools to probe receptor subtypes [112]. Although they are thought to mainly target ligand-gated ion channels, some of these peptides have also been shown to modulate GPCRs [113-117]. The case of the α-conotoxins Vc1.1 and RglA (Fig. 3B) is in particular interesting: these bicuspid peptides block the neuronal α9/10 subtype of the nicotinic acetylcholine receptor with nanomolar affinity and are also able to bind to the GABA_{B} receptor. In order to extract information for the design of peptidomimetics and strategic scaffolds for libraries, it certainly would be of interest to address questions: (i) Is the helix in the highly conserved first loop the motif responsible for high affinity interaction with both receptors? (ii) What is the role of the second loop?

Another intriguing example are the conopressins [114-117]. These cyclic nonapeptides act on the closely related vasotocin/vasopressin/oxytocin GPCRs. These receptors are present in many different species, including mollusc, non-mammalian vertebrates, anemids, fish, mammals and humans [118] and they are characterized by a high degree of interspecies conservation among orthologues. It is not clear, which evolutionary advantage is conferred by the presence of these peptides in the venom of the cone snail. However, the discovery and characterization of conopressin-T in the venom of the marine invertebrate Conus kefferi led to the identification of an agonist/antagonist switch, which is currently investigated regarding novel antagonist design for the human receptors [117], which will be discussed in more detail below.

Another abundant group of naturally-occurring peptides, yet less explored in particular with respect to possible receptor targets, are bioactive plant peptides, especially circular peptides, so-called cyclotides. They have recently attracted much attention due to their impressive stability, their intrinsic bioactivity and their natural abundance and sequence diversity [119]. Cyclotides are disulfide-rich peptides discovered in plants of the Violaceae, Rubiaceae and Apocynaceae families [120-122]. They are about 30 amino acids in size and have the unique structural features of a head-to-tail cyclized backbone and a knotted arrangement of three-disulfide bonds, referred to as a cyclic cystine knot (CCK) motif [120] (see Fig. 4A). The compact CCK motif makes cyclotides exceptionally resistant to thermal, chemical or enzymatic degradation [121, 123].

Recent studies on the evolution, distribution and abundance of cyclotides in plants has led to the conclusion that there are at least 9,000 novel cyclotides to be discovered in the violet family (Violaceae) [124] and an even greater number (up to 50,000) in the coffee family (Rubiaceae) [122]. The discovery of cyclotides and cyclotide-like proteins in other monocotyleon [125] and dicotyleon plant families suggests that cyclotides are much more numerous than originally anticipated and it is likely that cyclotides will ultimately comprise one of the largest protein families within the plant kingdom [122] and hence constitute an immense “library of natural peptides”. The family of cyclotides impresses not only by their sheer number, but also by their diversity. Typically one species can express 15–60 different, unique cyclotides [122, 124, 126, 127] and a recent report suggests this number may be increased by varying growth conditions [128]. Their diversity is a result of the conserved cysteine framework and the backbone chains (loops)

---

**Note:** The full text is a summary of the content, providing key points and insights related to the development and use of peptide libraries in drug discovery. The text highlights the diversity in nature and the importance of natural peptides in the field of pharmaceutical research. It discusses encoded and synthetic approaches, as well as natural peptide libraries, including venom and plant peptides. The text emphasizes the role of these peptides in understanding protein-ligand interactions and their potential as drug candidates.
between those cysteines, which are tolerant to various amino acid combinations (see Fig. 4B). The plasticity of the cyclotide framework and its tolerance to substitution [129] have been recently used as a tool for the development of novel anti-cancer [130] and anti-

infectious agents [131]. These examples underpin the application of cyclotides as combinatorial peptide scaffolds in pharmaceutical drug design [132-134] and it makes the cyclotides a large naturally-occurring combinatorial peptide library.

### Table 3. Bioactive Peptides from Natural Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Peptide Class</th>
<th>Estimated Number of Peptides</th>
<th>Activity/Targets</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria, plants, animals (invertebrates, vertebrates) humans</td>
<td>Anti-microbial peptides (AMPs), defensins</td>
<td><em>-</em></td>
<td>Immunity, anti-microbial, insecticidal, antifungal, enzyme inhibitors</td>
<td>[190-193]</td>
</tr>
<tr>
<td>Animals/venom toxins</td>
<td>Bees, wasps and ants</td>
<td>Hundreds to thousands</td>
<td>Hemolytic, anti-microbial, cytolytic, receptor/G-protein interface/receptomimetic (e.g., mastoparan)</td>
<td>[107 194-196]</td>
</tr>
<tr>
<td>Cone-snails, Conotoxins</td>
<td>50,000</td>
<td>Anti-pain (ion-channels, acetylcholine receptor), neurotensin R, entodethlin R, adreno R, 5-HT R, OT/AVP R, and many more</td>
<td>[109 197]</td>
<td></td>
</tr>
<tr>
<td>Scorpions</td>
<td>100,000</td>
<td>Ion-channels, cytolytic</td>
<td></td>
<td>[104 197]</td>
</tr>
<tr>
<td>Snakes</td>
<td>Many thousands</td>
<td>Neurotoxins; muscarinic-, α-adrenergic- and neurotensin R</td>
<td>[106, 107]</td>
<td></td>
</tr>
<tr>
<td>Spiders</td>
<td>1.5 – 16 million</td>
<td>Ion channels, cytolytic</td>
<td></td>
<td>[103 197]</td>
</tr>
<tr>
<td>Plants</td>
<td>Cyclotides</td>
<td>&gt;19,000-59,000</td>
<td>Insecticidal, anti-HIV, cytotoxic, uterotonic, neurotensin R</td>
<td>[122, 124, 125 198]</td>
</tr>
</tbody>
</table>

*Total number has not been reliably been determined, but due to the wide distribution and diversity, these peptides make up large natural peptide libraries

*Only few peptides have been reported from individual species, although some venom components have been studies in great detail (e.g., melittin and mastoparan); generally arthropod venom peptides are thought to have great pharmacological potential

*Including peptides and larger proteins
Apart from their unique structural framework, abundance and sequence diversity, cyclotides make interesting starting points for pharmacological analysis. They exhibit a range of biological activities, including uterotonic [135], hemolytic [136], anti-neurotensin [137], anti-HIV [138], cytotoxic [139], anti-bacterial [140], anti-fouling [141] and insecticidal properties [142-144]. Only little is known about the mechanism of how they elicit these activities, but this probably involves a combination of “non-specific” membrane disruption and directed, specific receptor interaction [145].

The proposed uterotonic and anti-neurotensin activities of cyclotides are of particular interest in the present context as this would be the first example of cyclotides acting as ligands on GPCRs. Gran et al. studied the pharmacological properties of kalata B1 and found that the peptides induce oxytocin-like uterine muscle contractions in vitro (isolated rat and rabbit uteri) at concentrations of 20 μg/mL and similar effects in vivo (rabbit) at concentrations of 100 μg/kg. However, doses of > 1 mg/kg, were found to be toxic (increased blood pressure and ventricular tachycardia) and eventually lethal (ventricular fibrillation) [135, 146, 147]. Nevertheless, it is intriguing to further explore the uterotonic effect of cyclotides in molecular detail, in particular because kalata B1 [148], and other cyclotides (for instance tricyclons A and B from the violet plant Viola tricolor [149]), contain sequence- and structural-motifs (β-turns) that are similar to the presumed activity-bearing motifs in the oxytocin peptide [18] (see Fig. 4C). Another interesting cyclotide is cyclopsychotride A, which reportedly inhibits the activity (antagonistic) of the neuro-peptide neurotensin, which is the natural ligand of the neurotensin receptor and of interest as a target in neurodegenerative disease [150]. The cyclotide was originally isolated from the tropical tree Psychotria vellosiana (formerly known as Psychotria longipes) and inhibits the binding of neurotensin to its receptor with an IC₅₀ of 3 μM and also was found to increase the intracellular Ca²⁺ concentration in a dose-dependent manner [137]. The neurotensin receptor is also targeted by the conotoxin contulakin-G, which is currently undergoing clinical trials for the treatment of neuropathic pain [151].

These are just a few of many examples that show how nature’s vast libraries can be a source of diverse, biologically pre-validated and evolutionary fine-tuned scaffolds that can be used as biological relevant starting points in the quest of GPCR modulators. One could argue that natural products optimized to defend against and prey on a large selection of species would have an enhanced likelihood to preserve common structural motifs that are active on more than one target. It can therefore be hoped that, in times of high-throughput screening, such libraries will be screened also against unrelated target receptors to identify such interesting privileged motifs systematically rather than by serendipity.

**Design of Selective Peptide Agonists and Antagonists for Oxytocin and Vasopressin Receptors**

Oxytocin (OT) and vasopressin (AVP, arginine-vasopressin) are closely related, highly conserved, multifunctional nonapeptides...
with an N-terminal cyclic 6-residue ring structure stabilized by an intramolecular disulfide bond, and a flexible C-terminal 3-residue tail. They are structurally very similar: they differ only by two amino acids in position 3 and 8 (see Fig. 5), however their physiological functions are quite distinct. In the periphery, OT is involved in uterine smooth muscle contraction during parturition, ejaculation, and milk ejection during lactation [152]. In the central nervous system, OT functions as a neurotransmitter involved in complex social behaviors [153-157], maternal care [158, 159], bonding [160], stress and anxiety [96, 161]. AVP regulates peripheral fluid balance and blood pressure [162, 163] and is centrally implicated in memory and learning [164, 165], stress related disorders [166] and aggressive behavior [167, 168]. Both peptides act on GPCRs, OT via one oxytocin receptor (OTR) and AVP via three vasopressin receptors (AVPRs, vasopressor V1aR, pituitary V1bR, renal V2R) [169]. Their distinct physiological effects can be accounted for by the low sequence homology of the intracellular domains of the OTR and the AVPRs. In contrast, the transmembrane portion and extracellular domains have high sequence homology. The related docking sites and the structural similarity of OT and AVP create a problem of specificity and result in significant cross reactivity [170, 171]. In fact, receptor specificity is not only controlled by (modest) ligand selectivity, but rather by a complex interplay comprising the enzymes oxytocinase and vasopressinase and by cell-specific up- and down-regulation of individual receptor expression [172]. 

Species variability further complicates the quest for selective agonist/antagonists: many compounds selective in rat turned out to be unspecific in humans [173]. These factors constitute a major challenge and although thousands of OT and AVP analogues have been synthesized, there remains still a shortage of receptor selective agonists and antagonists, in particular for the human receptors [174-176]. Clinically, OT is administered intravenously to induce labor and treat post-partum hemorrhage, and intra-nasally to elicit milk letdown; the antagonist atosiban is employed during premature labor. Vasopressin is administrated orally in the form of deamino-o-Arg8-vasopressin (desmopressin) for the treatment of diabetes insipidus and bedwetting; it is also used in coagulation disorders such as von Willebrand disease and mild forms of factor VIII-deficiency [174, 177, 178]. However, all of these ligands have been shown to be non-selective at the human OT and AVP receptors [174, 175]. Other potential applications for selective OT/AVP agonists and antagonists include treatment of congestive heart failure, blood pressure, anxiety, stress, anger, depression, hypotension, hypo-osmolality, hyponatremia and bleeding disorders [174, 179]. Understanding the mechanism of ligand-receptor interaction and the structural motifs involved in binding and GPCR activation is essential for the library design of such a delicate and exigent system. There are significant differences in the design of peptide versus small molecule modulators and the focus will be on peptide library design. However, selective peptide agonist/antagonists represent excellent starting point for further small molecule drug designs.

The crystal structure of deamino-OT and NMR structures of OT and AVP showed secondary structure motifs in form of a Type-II β-turn between the backbone of –NH of Tyr 2 and the carbonyl of Asn5 and a Type-III β-turn between the Cys 6 and the Gly 9 imino group [180-183]. Structure-activity relationship and NMR studies further revealed that the disulfide bond is crucial for agonistic activity because it acts as a conformational constraint to limit the flexibility and to align tyrosine in position 2 and the carboxylic tripeptide tail [181, 184]. These observations indicate that binding might occur via a receptor-induced fit, which is further reinforced by the comparison of the deamino-OT crystal structure, (two β-turns) [180] with the crystal structure of OT bound to its carrier-protein neurophysin, where no β-turns were observed [185]. To obtain receptor-specific ligand selectivity it is hence essential for the agonist library design to preserve the recognition motifs and the conformational flexibility of the ring structure. The emphasis is on agonist design as subtle differences exist between the mechanism of agonist and antagonist. To act as an agonist, the ligand needs to undergo receptor recognition followed by binding and conformational changes in both, the ligand and the receptor, which results in a functional response. In contrast, receptor inhibition can result from simple blockage of the agonist binding pocket or by binding to an allosteric site that changes the conformation of the receptor suppressing the functional response. Hence SAR studies relevant for agonists do not necessarily yield results that are predictive for the SAR of antagonists. This has also been realized during the antagonist development for oxytocin and vasopressin receptors [186-188]. Nevertheless the overall methodology of finding selective peptide agonists as well as antagonists remains the same (see Fig. 6).

Modifications include changes in hydrophobicity, charge states, conformational constraints, polarity and aromaticity to extract information on their effect on potency and selectivity to guide the direction of subsequent efforts. For selectivity optimization it is especially helpful to screen analogues from four receptor species. The information obtained from the single residue mutations series and from the literature is the basis for the next series that incorporates additive changes in more than one position. This cycle of analysis and design is repeated until the desired parameters have been optimized. Use of synthetic/unnatural building blocks is of benefit not only because it significantly extends the chemical diversity around the structural framework important for selectivity, but also because it often results in analogues with improved bioavailability.

High-throughput synthesis and biological screening is required to drive this type of approach. A strategy in combinatorial chemistry was originally developed for the drug discovery pipeline of venom peptides that gives rapid synthetic access to a large quantity of disulfide-rich peptides [189]. This has been adapted for the oxytocin/vasopressin library design. It utilizes a safety-catch acid labile linker (SCAL) that allows complete side-chain deprotection of the assembled peptides without cleaving the peptide from the solid support. The “naked” peptides are folded on-resin and a chemo-selective reaction activates the SCAL linker, which releases the folded peptides from the solid support. Depending on the quality of assembly and analogue separation, the folded peptides are either screened directly or purified prior as mixtures by reverse phase HPLC. When combined with highly automated solid-phase peptide synthesis, this approach allows for rapid production of a large number of analogues. Biological screening occurs via radioligand binding assays or functional assays that analyze the intracellular secondary messengers (accumulation of IP3 or cAMP; Ca2+-ion transients) signal. In this way, hundreds of peptide ligands can be synthesized and characterized efficiently. Selectivity and other pharmacological properties should improve significantly within a few iterative cycles. Once the structural features for binding and selectivity have been identified, it might be possible to further modify the structure to enhance stability against proteolytic degradation.

![Fig. (5)](image-url) Amino acid sequences of oxytocin and vasopressin.

The amino acid of mature oxytocin and vasopressin peptide hormones is presented. The two differences in each of the peptides are highlighted in blue. Disulfide bonds are indicated.
increase the half-life of the peptide in the circulation, improve membrane permeability and/or create a prodrug. There are various well-established approaches such as implementation of peptidomimetic design, truncations, cyclizations and PEGylation that can be applied to improve the bioavailability of the peptide and improve its drug properties.

SUMMARY AND FUTURE PERSPECTIVE

There is a school of thoughts that assumes that all therapeutically useful, readily druggable GPCRs have already been discovered. Future investigations in these areas are then argued to be futile because they are subject to the law of diminishing returns. However, it is evident from Table 1 that several compounds (agomelatine, aprepitant, cinacalcet, exenatide, fangolimo, icatibant, laropram, maraviroc, rimonabant, sitaxentane) have been approved for clinical pharmacotherapy in the past 5-6 years. In most instances, the cognate receptors of these drugs had previously not been exploited as drug targets. Thus > 10% of the extant drugable GPCRs were dis-

REFERENCES

References 201-203 are related articles recently published.

Ligand-Based Peptide Design and Combinatorial Peptide Libraries


Escobas P, Sollod B, King GF. Venom landscapes: Mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon 2006; 47: 650-63.


Livett BG, Sandall DW, Keays D, et al. Therapeutic applications of conotoxins that target the neuronal nicotinic acetylcholine receptor. Toxicon 2006; 48: 810-29.


Ligand-Based Peptide Design and Combinatorial Peptide Libraries


