

# G PROTEIN-COUPLED TIME TRAVEL

## *Evolutionary Aspects of GPCR Research*

**T**he common seven-transmembrane-domain (TMD) architecture of G protein-coupled receptors (GPCRs) has been preserved over a vast period of time, and highly conserved amino acid motifs and residues have evolved to establish ligand and signal transduction specificities. The mining of evolutionary data from sequenced genomes and targeted retrieved orthologs has proven helpful for understanding the physiological relevance of individual GPCRs and for interpreting the clinical significance of GPCR mutations in structural terms. Sequence analysis of GPCR pseudogenes, which are considered as genomic traces of past functions, as well as recent success in sequence analysis of GPCR genes from extinct species, provide further information. This review discusses recent advances and approaches aimed at developing a better understanding of GPCR biology based on evolutionary data.

■ Holger Römpler<sup>1,3</sup>, Claudia Stäubert<sup>1</sup>, Doreen Thor<sup>1</sup>, Angela Schulz<sup>1</sup>, Michael Hofreiter<sup>2</sup>, and Torsten Schöneberg<sup>1</sup>

<sup>1</sup> Institute of Biochemistry, Molecular Biochemistry, Medical Faculty, University of Leipzig, Johannisallee 30, 04103 Leipzig, Germany. <sup>2</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>3</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

## INTRODUCTION

G protein-coupled receptors (GPCRs) constitute the most prominent family of validated pharmacological targets in biomedicine. Approximately sixty percent of approved drugs elicit their therapeutic effects by selectively targeting members of this receptor family (1). Despite the remarkable structural diversity of natural GPCR agonists and low sequence homology among GPCRs, hydrophathy analysis and biochemical data suggest that all GPCRs share a common molecular architecture consisting of seven transmembrane domains (TMDs) connected by three intracellular and three extracellular loops. Currently, a high-resolution structure is available only for bovine rhodopsin (2, 3), and so broader consideration of GPCRs must rely on other sources of structural information. The availability of in-depth genome sequence data for human (4), mouse (5), and many other vertebrates, as well as of numerous invertebrates, has provided the unique opportunity to compare and analyze genome-wide GPCR sequences. Such comparative genomic approaches can provide information about the origin and evolutionary history of GPCRs, illustrated below in the first two paragraphs by the example of rhodopsin-like GPCR. Comparative genomic data provide information about the origin and evolutionary history of GPCRs, and analyses of genomes have revealed hundreds of obviously non-functional gene-like sequences, so-called pseudogenes, many of which reflect GPCR-like sequences suggestive of past functions. Finally, comparative genomic approaches can help to identify conserved sequence motifs that are responsible for certain aspects of GPCR functionality, which, through mutation, may cause disease.

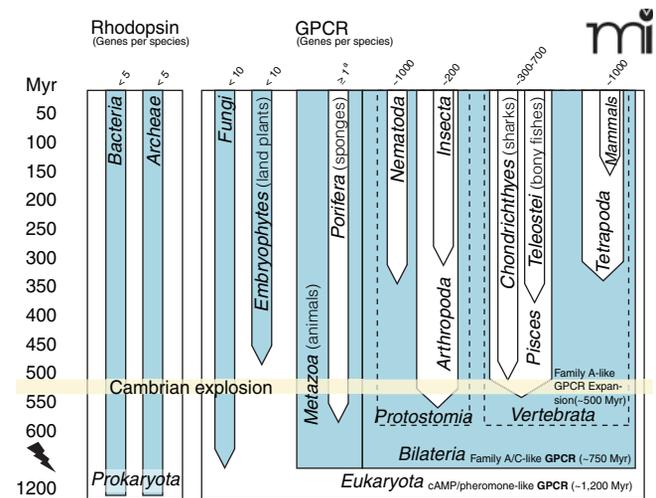
## THE ORIGIN OF MODERN GPCRS

The evolutionary success of the GPCR superfamily is reflected both by its presence in almost every eukaryotic organism and by its abundance in mammals. Based on the genome sequences of humans and other mammalian species, approximately three to four percent of mammalian genes code for GPCRs. The GPCR superfamily comprises at least five structurally distinct families: the Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, and Secretin (GRAFS) families (6, 7). The rhodopsin-like receptors (also called family A or 1) form the largest family in vertebrates. Because there is very little sequence homology among the rhodopsin-like, secretin-like (family B or 2), and glutamate-like (family C or 3) receptor families, the evolutionary origin of GPCRs and their ancestry remain a matter of debate.

Proteins that display a seven-TMD architecture are present in prokaryotes that utilize light-sensitive proteo-, halo- and bacteriorhodopsins to harvest energy and fix carbon via a non-chlorophyll-based pathway. In addition to the ion-translocating rhodopsins, sensory rhodopsins exist in Halobacteria that promote phototaxis by collaborating with transducer membrane-embedded proteins with no relation to G proteins [i.e., halobacterial transducer of sensory rhodopsin-(Htr)-I and -II]. Like the rhodopsins of bilateral animals, prokaryotic rhodopsins contain retinal covalently bound to

TMD7. Similarly, eukaryotes contain transmembrane proteins that are structurally similar to prokaryotic sensory rhodopsins (8, 9). The structural and functional features shared by prokaryotic and eukaryotic rhodopsins suggest a common ancestry; however, despite these similarities, sequence comparisons provide no convincing evidence of an evolutionary linkage between prokaryotic rhodopsins and eukaryotic GPCRs (10).

Signal transduction through G proteins is definitive of GPCRs. The rhodopsins of insects and vertebrates are coupled to the G proteins  $G_q$  and transducin, respectively. Several groups of worms (e.g., nematodes and trematodes) are the evolutionarily oldest extant lineages that exhibit evidence of rhodopsin- and G protein-mediated photoreception (11). These appeared, however, long after GPCRs and G protein signaling evolved. Structural and functional data clearly show that G protein signaling via GPCRs is present in yeast/fungi (12) and plants (13) (e.g., fungal pheromone receptors, slime mold cAMP receptor-like receptors, and glutamate receptor-like receptors), as well as in unicellular eukaryotes, such as the slime mold *Dictyostelium discoideum* (14). GPCRs with some structural relations to cAMP receptor-like receptors and adhesion receptors are also found in plant genomes. Glutamate receptor-like receptors are present in *D. discoideum* and the sponge *Geodia cydonium* (15), which diverged more than 600 million years ago. Sponges (Porifera) belong to the earliest occurring metazoans (animals). In contrast, the structural signatures of rhodopsin-like GPCRs do not appear to extend phylogenetically back before the appearance (approximately



**Figure 1. Representation of GPCRs in major evolutionary lineages.** Heptahelical transmembrane-spanning proteins of the proteo-, halo-, and bacteriorhodopsin families are present in prokaryotes. Except for their transmembrane architecture, there is no evidence of an evolutionary relation between the genuine GPCRs and bacteriorhodopsins, which are not coupled to G proteins. GPCRs and their signal transduction probably evolved ~1.2 billion years ago. Genomes of extant plants and fungi usually contain less than 10 GPCR genes. The first rhodopsin-like GPCRs, which compose the main GPCR family in vertebrates, appeared ~580-800 million years (Myr) ago. Expansion of rhodopsin-like GPCRs started ~500 Myr ago, giving rise to over 1000 members in mammalian genomes. The relationships of the major lineages, which remain controversial, are not indicated.

<sup>a</sup> Complete genome data are not yet available.

580–800 million years ago) of Bilateria such as insects, mollusks, nematodes, and trematodes (16–18). Given that the coupling of G proteins to seven-TMD proteins evolved before the plant/fungi/animal split (i.e., about 1.2 billion years ago), and that the first rhodopsin-like receptors appeared early in metazoan evolution (Figure 1), the retinal-based photosensory system likely represents a “re-invention” (i.e., a product of convergent evolution). Among the rhodopsin-like GPCRs, serotonin receptors appear to be among the oldest, as suggested by their presence in planarians (19) and in many nematodes (20).

### THE EVOLUTIONARY EXPANSION OF RHODOPSIN-LIKE GPCRS

Vertebrates appeared in evolution about 500 million years ago, during the Cambrian period. With the exception of odorant receptors (ORs), rhodopsin-like receptors occur in vertebrate genomes twice as frequently as they do in invertebrates (Figure 1). Although they are the predominant family of GPCRs in vertebrates as a whole, rhodopsin-like receptors occur significantly less frequently in more basal vertebrate lineages than in higher species, which primarily reflects the continuous expansion of the OR gene family (7). The evolutionary rise in the number of GPCR genes is generally explicable in terms of two mechanisms of duplication, described below. Subsequent to gene duplication, new functions can evolve as one or even both members of a duplicated gene pair mutate and acquire novel functionality; the fitness of the organism is presumptively less at risk to mutation in the context of multiple gene copies. Disadvantageous mutations in GPCRs are removed from a population through purifying selection, and many evolutionarily old GPCR genes, including the rhodopsins, display strong purifying selection (21–23).

### INTRACHROMOSOMAL GENE DUPLICATIONS

Clusters of GPCR genes occur as the result of contiguous duplication of genes. The most impressive example is found in ORs; human chromosome 11 encodes nearly half of the classical OR repertoire, including a single cluster of more than 100 OR genes (24). Multiple copies of related GPCR genes, such as the human protease-activated receptors (human 5q13), P2Y<sub>12</sub>-like receptors (human 3q24), trace amine associated receptors (TAAR; human 6q23.2) and CC-chemokine receptors (human 3p21.3), are arranged in a tandem-like fashion.

### ENTIRE CHROMOSOMAL AND WHOLE GENOME DUPLICATIONS

The dopaminergic receptor paralogs DRD5 and DRD1, and the adrenergic receptor paralogs ADRA2C and ADRA1 on chromosomes 4 and 5, respectively, seem to have originated from a chromosomal or even a whole genome duplication event (25–28). Polyploidy is common in certain vertebrates, such as fish and amphibian species; ploidy levels of up to 8n have been found in sturgeon (29). GPCR

duplications due to polyploidy have been proposed for the expansion of TAAR and lyso-phosphatidylserine receptors (e.g., GPR34) in the ancestor of the teleostei, the largest group of bony fishes (21, 30). A whole genome duplication likely occurred in the teleost lineage following its split from the tetrapod lineage, such that only a subset of the duplicates are retained in modern teleost genomes (31).

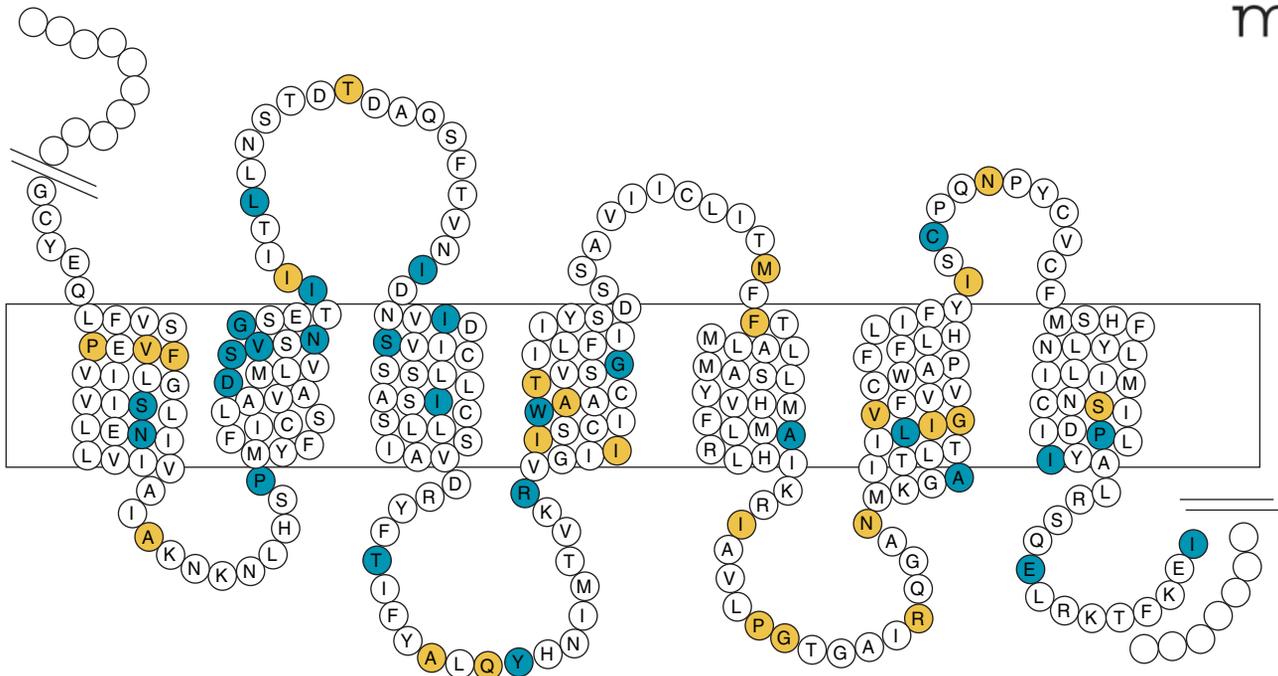
### ANCIENT RECORDS OF GPCR EVOLUTION

Whereas more than ninety-nine percent of all species that ever lived on earth are extinct, most of the genetic information about receptor repertoires and their evolutionary history must be garnered from the fossil record. Recent advances in DNA extraction and amplification from fossil remains (32) have made it possible to retrieve substantial amounts of ancient [e.g., Pleistocene (33) and Bronze age (34)] DNA sequences. Genome sequences of extant species, however, can also harbor valuable information about past functions and gene history. Pseudogenes in particular can generally be considered as genomic fossils, and they increasingly attract attention in GPCR research.

Pseudogenes are by definition non-functional genomic sequences that are homologous to a known gene (35). As a result of their non-functionality, most pseudogenes are released from selective pressure, and they therefore display an increased ratio of asynonymous to synonymous substitution rates ( $K_a/K_s$ ) and accumulate frameshifting and nonsense mutations. These features are frequently used to identify pseudogenes in a genome (36). The majority of mammalian pseudogenes can be classified as duplicated pseudogenes or retrotransposed pseudogenes (also called “processed” pseudogenes). The latter are created by the reverse transcription of mRNAs followed by genomic integration. Duplicated pseudogenes arise from local duplication or unequal crossing-over; thus, they often retain, at least to some extent, the intron structures of the parental genes. Nevertheless, duplication or retrotransposition events are not always directly involved in the emergence of pseudogenes. The mutational inactivation of functional genes can be pinpointed in the genesis of specific pseudogenes, including mutational inactivation of GPCR-encoding genes (37–39).

Over time, signatures of the original sequence gradually disappear, which complicates the identification of pseudogenes (40); the exact number of pseudogenes within a complex genome can only be estimated. Rough estimates suggest that signatures of genes can be detected more than eighty million years after “pseudogenization” and loss of constraint. For example, pseudogenes for the neuropeptide Y receptor type 6 (Y<sub>6</sub>R) exist in all primates so far investigated; an inactivating deletion mutation is therefore believed to have occurred in a common ancestor of primates. Large deletions, however, can remove informative sequences in a significantly shorter period; for example, the Y<sub>6</sub>R gene disappeared in rats after the mouse/rat split fourteen to sixteen million years ago.

Pseudogenes are important because they provide molecular records of ancient genes that existed millions of years ago. Pseudogenes are particularly plentiful in the case of ORs. Whereas



**Figure 2. Distribution of missense mutations in the melanocortin type 4 receptor (MC4R).** Like all GPCRs, the MC4R is composed of seven transmembrane domains connected by three intra- and three extracellular loops. Inactivating mutations in the MC4R are indicated in inherited forms of obesity. Amino acid substitutions at sixty positions have been found in obese humans (colored circles), but only half of missense mutations reveal a partial or complete loss of receptor function (blue circles).

the human genome contains signatures of at least thirty non-olfactory rhodopsin-like GPCR pseudogenes, out of approximately 900 ORs, at least sixty-three percent appear to be pseudogenes (24). Furthermore, some apparently intact human OR genes lack motifs that are very highly conserved in their mouse orthologs, suggesting that not all human OR genes with complete open reading frames express functional OR proteins. A similar accumulation of pseudogenes has been observed in human bitter taste receptors. By contrast, in the mouse genome only about twenty percent of OR sequences are pseudogenes, giving mice over three times as many intact OR genes as humans (41). It has been speculated that the evolution of trichromatic color vision in hominoids and other Old World monkeys has relaxed the functional constraints for many taste, odorant, and pheromone receptors, thereby facilitating the development of pseudogenes (42–44).

Although the original definition of a pseudogene implied transcriptional silence, a number of reports over the years have reported pseudogene transcripts (37, 45–48). Transcripts of processed pseudogenes can contain regions with significant antisense homology, which may suggest a regulatory role for transcribed pseudogenes through an RNAi-like mechanism (49). Transcripts of several GPCR pseudogenes have been detected, but evidence for their functional relevance has yet to be substantiated.

## STRUCTURAL INFORMATION FROM EVOLUTIONARY APPROACHES

Currently, the only available crystal structure of a GPCR is that of rhodopsin, which provides the basis to generate structural models for all other members of the superfamily (50). Efforts to mine evolutionary trends for additional structural information have thus been crucial in modeling and mutational studies for predicting the arrangements of the seven transmembrane helices (51). To identify distinct structural determinants, such as those that participate in ligand recognition or signal transduction for a particular GPCR, sequence analysis has focused on the comparison of receptor orthologs and paralogs. The basic concept of this approach is that the structural diversity between orthologs is the result of a long evolutionary process characterized by a continuous accumulation of non-deleterious mutations. Studies based on large numbers of orthologs have analyzed the conservation and relative orientation of TMDs (21) to address the functional relevance of distinct residues in GPCRs (38, 52).

GPCR pseudogenes and their evolutionary fate can also be monitored by sequence comparisons of orthologs. An illustrative example is the chemokine receptor GPR33. Having appeared in the mammalian genome about 125–190 million years ago, GPR33 underwent parallel pseudogenization in humans, other hominoids, and some rodent species—all within just the past one million years (37). This relatively sudden occurrence across so many disparate species likely reflects a selective pressure in favor of GPR33 inacti-

**Table 1. °Characterization of naturally occurring missense mutations in human MC4R**

<sup>b</sup> Degree that given amino acid position is conserved.	<sup>c</sup> Number of amino acid substitutions that retain wild-type functionality.	<sup>c</sup> Number of amino acid substitutions that change wild-type functionality.
Not conserved	18	3
Loosely conserved	8	9
Stringently conserved	5	24

<sup>a</sup> More than seventy missense mutations (among sixty positions) have been identified in the human MC4R.

<sup>b</sup> Stringency of conservation at the given position is determined by sequence comparison of over seventy vertebrate MC4R orthologs: Stringently conserved, no variation; loosely conserved, either of two amino acids occurs among orthologs; not conserved, more than two amino acids can occur at given position among orthologs.

<sup>c</sup> About ninety percent of all mutations have been functionally tested. The complete list of missense mutations included in the analysis and the references can be provided on request.

variation. A selective advantage has been established in the inactivation of the chemokine receptors CCR5 and DUFFY, which act as co-receptors for the cell entry of pathogens. CCR5 is involved in the internalization of human immunodeficiency virus-(HIV)-1, and DUFFY mediates the entry of *Plasmodium vivax*. Mutational inactivation of CCR5 and DUFFY leads to resistance to infection by HIV-1 (53) and *P. vivax* (54), respectively. Interestingly, the inactive CCR5-Δ32 allele exists in European populations at comparatively high rates (around ten percent), although the appearance of HIV-1 is too recent to explain such an allelic preponderance. Other possible selective advantages proposed for CCR5-Δ32 (55), along with non-selective scenarios (56), are still under debate.

In the case of GPR33, a cause of inactivation could be the possible interplay between receptor and a rodent-hominoid-specific pathogen; indeed, fixation of the GPR33 pseudogene is observed in rats and gerbils but not in dozens of other mammalian species (37, 38). Although the selective action of microbes upon the GPR33 pseudogene is still hypothetical, rats and gerbils are common hosts of zoonotic pathogens like hanta viruses and *Yersinia pestis*, and in fact both rodent species share their habitats with humans. Although evolutionary data cannot supplant functional studies of receptors, the relevance of evolutionary trends upon GPCR research is intriguing and greatly enriches our understanding of receptor repertoires.

## CLINICAL IMPLICATIONS OF EVOLUTIONARY DATA

GPCRs play a central role in almost all physiological functions, and their involvement in pathophysiological processes is becoming clear within multiple clinical contexts. Mutations in GPCR genes are responsible for more than thirty human diseases, including cancer (57). Classical examples for GPCR-related diseases are nephrogenic diabetes insipidus (NDI) and an inherited form of obesity caused by the concomitant mutations of the vasopressin type 2 receptor (V2R) and in the melanocortin type 4 receptor (MC4R). About one third of

all inactivating mutations found in GPCRs are nonsense and frame-shifting mutations, which are readily recognized upon sequence analysis. In contrast, interpreting the clinical relevance of missense mutations usually requires heterologous expression and functional characterization of the mutant receptor by ligand binding studies and second messenger assays, because missense mutations in GPCRs do not cluster in specific receptor regions. Figure 2 depicts positions of missense mutations of MC4R that decrease receptor activity and are associated with inherited obesity. It is of interest to note that almost ninety percent of all functionally relevant missense mutations in MC4R (C Stäubert, T Schöneberg, H Römler; unpublished results) and V2R (57) affect receptor sequences that have been conserved (negative selection) over 450 million years of vertebrate evolution.

Several examples demonstrate that in-depth functional characterization in combination with mining evolution as an additional source of structural information is an extremely powerful approach in the investigation of the molecular consequences of distinct disease-causing mutations in GPCRs. The likelihood that a residue is important for proper receptor function positively correlates to the degree to which the residue is conserved among vertebrates over the course of evolution. For example, most GPCRs of the rhodopsin family contain a highly conserved disulfide bond connecting extracellular loops 1 and 2. Interestingly, disease-causing missense mutations located in extracellular loops of the V2R often introduce cysteinyl residues in the expressed protein (e.g., the Y<sup>205</sup>C mutant receptor). An attractive explanation for the dysfunction is that the additional extracellular cysteine residue introduced by mutation disrupts the conserved disulfide bond or engages in other disulfide bridges. As attractive as this mechanistic explanation is, the replacement of certain wild-type residues by cysteine (e.g., the Y<sup>205</sup>C variant of V2R) can be mimicked by the introduction of residues that do not participate directly in disulfide linkages (e.g., the Y<sup>205</sup>H variant of V2R) (58), such that abolishment per se of an evolutionarily conserved and essential wild-type residue must be acknowledged.

For well-defined disease states such as NDI, the association of a phenotype with a missense mutation found in a respective GPCR can be straightforward. Obesity, in contrast, can have multiple causes; a mutation in MC4R is only one rare cause. Therefore, it is not surprising that only half of the missense variants of the MC4R found in obese humans have a partial or complete loss of function when studied *in vitro* (Table 1). Evolutionary approaches that exploit large sets of ortholog sequence data can supplement mutational analysis of obesity. For example, the Ala<sup>175</sup>Thr substitution has been associated with obesity and positively tested for dysfunction (59–62); however, Thr at this position is naturally found in wild-type amphibian, fish, and Western European hedgehog MC4R (63) (Figure 3). Similarly, the Val<sup>95</sup>Ile mutation has been identified in an obese patient and tested to be functionless, whereas Ile<sup>95</sup> is naturally found in polar bear and armadillo MC4R, findings that prompted the re-evaluation of the human MC4R; in our hands, the Val<sup>95</sup>Ile variant of the human MC4R in manifests wild-type activity (C Stäubert, T Schöneberg, H

	Ala <sup>175</sup> ↓
<i>H. sapiens</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>P. troglodytes</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>P. pygmaeus</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>M. mulatto</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>M. fascicularis</i>	VR <sup>I</sup> IISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>C. jacchus</i>	VGI <sup>T</sup> IISCIWA <sup>A</sup> ACTVSGILFIIYSE
<i>C. aethiops</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>S. sciureus</i>	VGI <sup>T</sup> IISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>O. garnettii</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>L. catta</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>T. belangeri</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>S. tridecemlineatus</i>	VGI <sup>T</sup> IISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>D. novemcinctus</i>	VGIIISGI <sup>W</sup> WA <sup>A</sup> ACTVSGVLFIIYSE
<i>F. catus</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>C. familiaris</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>V. vulpes</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>P. leo</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>L. lutra</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>P. sibirica</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>Z. californianus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>U. arctos</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>U. maritimus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>U. americanus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>U. thibetanus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>N. nasua</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>H. malayanus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>T. ornatus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>A. melanoleuca</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>B. taurus</i>	V <sup>A</sup> I <sup>I</sup> I <sup>S</sup> A <sup>I</sup> WA <sup>A</sup> ACTVSGVLFIIYSD
<i>S. scrofa</i>	VGIIISCIWA <sup>V</sup> CTVSGVLFIIYSD
<i>O. aries</i>	V <sup>A</sup> I <sup>T</sup> I <sup>S</sup> A <sup>I</sup> WA <sup>A</sup> ACTVSGVLFIIYSD
<i>C. liberiensis</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSE
<i>L. africana</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>E. maximus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>B. acutorostrata</i>	V <sup>A</sup> I <sup>I</sup> I <sup>S</sup> CIWA <sup>A</sup> ACTVSGILFIIYSD
<i>T. manatus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>O. anatinus</i>	VG <sup>T</sup> I <sup>I</sup> I <sup>G</sup> CIWA <sup>T</sup> CTVSGALFIIYSD
<i>E. telfairi</i>	VGII <sup>I</sup> I <sup>L</sup> SCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>E. europaeus</i>	VG <sup>V</sup> I <sup>I</sup> I <sup>S</sup> CIWA <sup>T</sup> ACTVSGILFIIYSD
<i>S. araneus</i>	VGIIISCIWA <sup>A</sup> ACS <sup>V</sup> SGILFIIYSD
<i>M. albescens</i>	VGII <sup>I</sup> I <sup>L</sup> SCIWA <sup>A</sup> ACTVAGILFIIYSE
<i>M. lucifugus</i>	VGII <sup>I</sup> I <sup>L</sup> SCIWA <sup>A</sup> ACTVAGILFIIYSE
<i>R. io</i>	VGII <sup>I</sup> I <sup>L</sup> SCIWA <sup>V</sup> CTVAGILFIIYSD
<i>M. spretus</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>M. musculus</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>R. norvegicus</i>	VGIIISCIWA <sup>A</sup> ACTVSGVLFIIYSD
<i>P. duprasi</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>P. obesus</i>	VGIIISCIWA <sup>A</sup> ACTVSGI <sup>V</sup> FIIYSD
<i>C. porcellus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>M. auratus</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>M. domestica</i>	VGII <sup>I</sup> I <sup>T</sup> CIWA <sup>A</sup> ACTVSGILFIIYSD
<i>C. simum</i>	VGIIISCIWA <sup>A</sup> ACTVSGILFIIYSD
<i>A. anser</i>	VG <sup>V</sup> I <sup>I</sup> I <sup>T</sup> CIWA <sup>A</sup> ACTVSGILFIIYSD
<i>G. gallus</i>	VG <sup>V</sup> I <sup>I</sup> I <sup>T</sup> CIWA <sup>A</sup> ACTVSGILFIIYSD
<i>G. aculeatus</i>	ATLV <sup>I</sup> IGSIW <sup>T</sup> CCTVSGILFIIYSE
<i>C. carpio</i>	AGT <sup>I</sup> I <sup>I</sup> TCIWT <sup>F</sup> FCM <sup>V</sup> SGVLFIIYSE
<i>T. nigroviridis</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>T. chinensis</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>T. radiatus</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>O. latipes</i>	AALV <sup>I</sup> ISSIWT <sup>C</sup> CTVSGILFIIYSE
<i>T. rubripes</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>T. poecilonotus</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>X. tropicalis</i>	AVV <sup>I</sup> IISCIWTAC <sup>S</sup> ISGVLFIIYSD
<i>C. auratus</i>	AGT <sup>I</sup> I <sup>I</sup> TCIWT <sup>L</sup> CTVSGVLFIIYSE
<i>D. rerio</i>	AGT <sup>I</sup> I <sup>I</sup> TCIWT <sup>F</sup> CTVSGVLFIIYSE
<i>T. porphyreus</i>	ASLV <sup>I</sup> ISSIWT <sup>C</sup> CCTVSGVLFIIYSE
<i>O. mykiss</i>	AMAV <sup>I</sup> ACIWS <sup>C</sup> CVASGVLFIIYSE
<i>S. acanthias</i>	ALM <sup>I</sup> IAAIWA <sup>A</sup> ACT <sup>G</sup> SGILFIIYSE



Römpler; unpublished results). These examples nicely illustrate the power of evolutionary data to inform functional analysis and structural evaluation of disease-causing mutations in GPCRs.

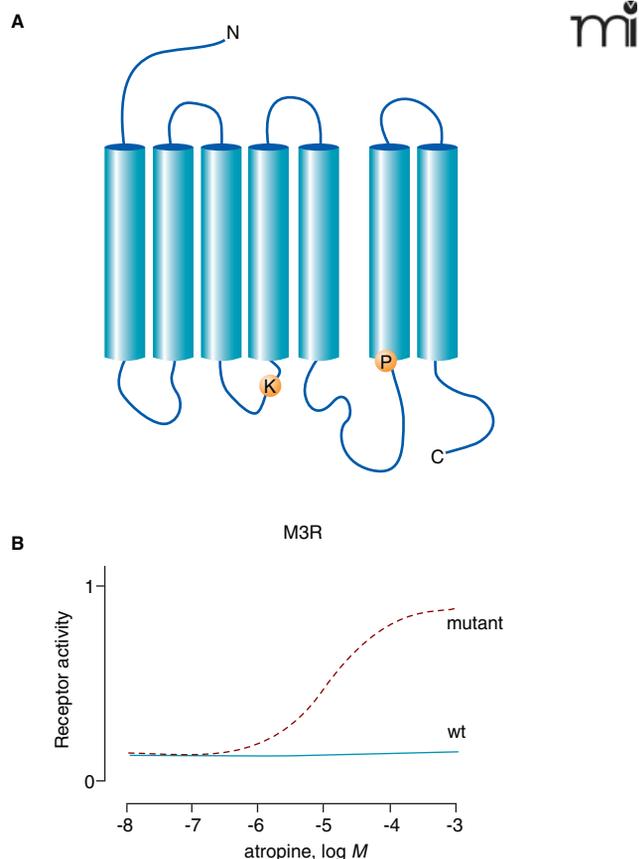
## FUTURE PERSPECTIVES

The growing number of sequenced genomes will further support our efforts to understand phylogenetic relationships and to identify signatures of past and ongoing evolutionary selection within GPCRs. The significance and reliability of interpretations depend on both the quality and quantity of sequence data obtained from different species and populations. Extant species and populations will be the main sources for genetic information, but sequence analysis of ancient nuclear DNA will be useful for dating molecular genetic events and analyzing adaptive processes.

In GPCR research, evolution technology depends on the very robust nature of G protein–receptor crosstalk, such that signaling elements of even distantly related species (e.g., yeast and mammals) can be brought together in vitro. To enable mammalian GPCRs to activate the yeast mating pathway, for instance, only one minor change in the C terminus of the yeast G protein Gpa1p is necessary (64). Through the tailoring of the yeast G protein, activation of a mammalian GPCR can control yeast cell growth via an engineered pheromone response pathway. This coupling of receptor activity to yeast growth, moreover, allows the rapid and economical screening of randomly modified receptor libraries. For example, atropine usually acts as a blocker on M3 muscarinic receptors (M3R). However, M3Rs from libraries of millions of randomly mutated receptors expressed in yeast have been selected for growth in the presence of atropine (Figure 4). Such approaches are suitable for ligand identification, mutational maturation, saturable structure-function analysis, genetic selection of constitutively active receptors, and generation of receptors with novel ligand recognition properties (65, 66).

Admittedly, the extraction of past evolutionary information cannot necessarily identify the selective forces that have shaped the structural evolution of GPCRs, and these past selective forces are obviously not amenable to experimental modulation and verification. However, for many biomolecules, such as RNA, enzymes, and antibodies, in vitro evolution techniques have been developed that guide structure and maturation under researcher-defined selection conditions. Over more than one billion years of GPCR evolution, the unique seven-TMD architecture has manifested discriminatory power and a remarkable degree of structural freedom to recruit almost any chemical molecule as a ligand. In keeping with the idea of a universal “receptor backbone,” artificial programming of the binding site by in vitro evolution approaches has indeed led to

**Figure 3. Evaluating the functional relevance of mutant MC4R by ortholog sequence comparison and functional testing.** Ala<sup>175</sup>Thr within transmembrane domain 4 has been associated with obesity and positively tested for dysfunction (59–62). In amphibian, fish, and Western European hedgehog, MC4R Ala<sup>175</sup> is naturally substituted by Thr; accordingly, the human Ala<sup>175</sup>Thr receptor protein has been examined and recently found to possess wild-type functionality (C Stäubert, H Römpler, and T Schöneberg; unpublished results).



**Figure 4. Identification of an atropine-agonized mutant M3 muscarinic acetylcholine receptor (M3R).** To generate an M<sub>3</sub>R that can be activated by atropine, we applied error-prone random mutagenesis and screened approximately  $1 \times 10^7$  mutant M<sub>3</sub>R clones for atropine sensitivity in a yeast expression system. In one clone we identified the mutations R<sup>179</sup>K and A<sup>489</sup>P (A) that enable the M<sub>3</sub>R to be activated by atropine (B). Both mutated positions are fully conserved within G<sub>q</sub>-coupled muscarinic acetylcholine receptors.

the generation of GPCRs with new ligand binding properties (67). Custom-designed biosensors are just one application of such evolutionary approaches.

### Acknowledgments

We would like to thank Christine Green and Adrian Briggs for many suggestions and critical reading of the manuscript. The work of the authors is supported by the Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, the Studienstiftung des deutschen Volkes, and the Max Planck Gesellschaft. ♥  
doi:10.1124/mi.7.1.5

### References

- Muller, G. Towards 3D structures of G protein-coupled receptors: A multi-disciplinary approach. *Curr. Med. Chem.* **7**, 861–888 (2000).
- Palczewski, K., Kumasaka, T., Hori, T. et al. Crystal structure of rho-

- dopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000). **This paper describes the first high resolution crystal structure of a GPCR. This structure inspired and formed the basis of hundreds of mutagenesis and modelling studies of this receptor family.**
- Edwards, P.C., Li, J., Burghammer, M., McDowell, J.H., Villa, C., Hargrave, P.A., and Schertler, G.F. Crystals of native and modified bovine rhodopsins and their heavy atom derivatives. *J. Mol. Biol.* **343**, 1439–1450 (2004).
  - International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
  - Waterston, R.H., Lindblad-Toh, K., Birney, E. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
  - Fredriksson, R., Lagerstrom, M.C., Lundin, L.G., and Schiöth, H.B. The G protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogous groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003). **Currently the most comprehensive study of sequence relationships among GPCRs. The analysis forms the basis of the GRAFS system of GPCR families.**
  - Fredriksson, R. and Schiöth, H.B. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.* **67**, 1414–1425 (2005).
  - Sineshchekov, O.A., Jung, K.H., and Spudich, J.L. Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* **99**, 8689–8694 (2002).
  - Waschuk, S.A., Bezerra, A.G., Jr., Shi, L., and Brown, L.S. Leptosphaeria rhodopsin: Bacteriorhodopsin-like proton pump from a eukaryote. *Proc. Natl. Acad. Sci. USA* **102**, 6879–6883 (2005).
  - Soppa, J. Two hypotheses—one answer. Sequence comparison does not support an evolutionary link between halobacterial retinal proteins including bacteriorhodopsin and eukaryotic G protein-coupled receptors. *FEBS Lett.* **342**, 7–11 (1994).
  - Hoffmann, K.F., Davis, E.M., Fischer, E.R., and Wynn, T.A. The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **112**, 113–123 (2001).
  - Elion, E.A. Pheromone response, mating and cell biology. *Curr. Opin. Microbiol.* **3**, 573–581 (2000).
  - Fujisawa, Y., Kato, H., and Iwasaki, Y. Structure and function of heterotrimeric G proteins in plants. *Plant Cell Physiol.* **42**, 789–794 (2001).
  - Kim, J.Y., Haastert, P.V., and Devreotes, P.N. Social senses: G protein-coupled receptor signaling pathways in *Dictyostelium discoideum*. *Chem. Biol.* **3**, 239–243 (1996).
  - Pin, J.P., Galvez, T., and Prezeau, L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* **98**, 325–354 (2003).
  - Benton, M.J. and Ayala, F.J. Dating the tree of life. *Science* **300**, 1698–1700 (2003).
  - Feng, D.F., Cho, G., and Doolittle, R.F. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94**, 13028–13033 (1997).
  - Peterson, K.J. and Butterfield, N.J. Origin of the Eumetazoa: Testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc. Natl. Acad. Sci. USA* **102**, 9547–9552 (2005).
  - Saitoh, O., Yuruzume, E., Watanabe, K., and Nakata, H. Molecular identification of a G protein-coupled receptor family which is expressed in planarians. *Gene* **195**, 55–61 (1997).
  - Tierney, A.J. Structure and function of invertebrate 5-HT receptors: A review. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **128**, 791–804 (2001).
  - Schulz, A. and Schöneberg, T. The structural evolution of a P2Y-like G protein-coupled receptor. *J. Biol. Chem.* **278**, 35531–35541 (2003).
  - Nielsen, R., Bustamante, C., Clark, A.G. et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
  - Go, Y., Satta, Y., Takenaka, O., and Takahata, N. Lineage-specific loss

- of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* **170**, 313–326 (2005).
24. Mombaerts, P. The human repertoire of odorant receptor genes and pseudogenes. *Annu. Rev. Genomics Hum. Genet.* **2**, 493–510 (2001).
  25. Perez, D.M. The evolutionarily triumphant G protein–coupled receptor. *Mol. Pharmacol.* **63**, 1202–1205 (2003).
  26. Lundin, L.G., Larhammar, D., and Hallbook, F. Numerous groups of chromosomal regional paralogs strongly indicate two genome doublings at the root of the vertebrates. *J. Struct. Funct. Genomics* **3**, 53–63 (2003).
  27. Larhammar, D., Lundin, L.G., and Hallbook, F. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.* **12**, 1910–1920 (2002).
  28. Wraith, A., Tornsten, A., Chardon, P., Harbitz, I., Chowdhary, B.P., Andersson, L., Lundin, L.G., and Larhammar, D. Evolution of the neuropeptide Y receptor family: Gene and chromosome duplications deduced from the cloning and mapping of the five receptor subtype genes in pig. *Genome Res.* **10**, 302–310 (2000).
  29. Ludwig, A., Belfiore, N.M., Pitra, C., Svinsky, V., and Jenneckens, I. Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus). *Genetics* **158**, 1203–1215 (2001).
  30. Gloriam, D.E., Bjarnadottir, T.K., Yan, Y.L., Postlethwait, J.H., Schiøth, H.B., and Fredriksson, R. The repertoire of trace amine G protein–coupled receptors: Large expansion in zebrafish. *Mol. Phylogenet. Evol.* **35**, 470–782 (2005).
  31. Woods, I.G., Wilson, C., Friedlander, B. et al. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* **15**, 1307–1314 (2005).
  32. Römpler, H., Dear, P.H., Krause, J. et al. Multiplex amplification of ancient DNA. *Nat. Protoc.* **1**, 720–728 (2006).
  33. Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schoneberg, T., and Hofreiter, M. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**, 62 (2006). **This study describes the cloning of the first nuclear gene (a GPCR gene) from an extinct species.**
  34. Hummel, S., Schmidt, D., Kremeyer, B., Herrmann, B., and Oppermann, M. Detection of the CCR5-Delta32 HIV resistance gene in Bronze Age skeletons. *Genes Immun.* **6**, 371–374 (2005). **By analysing CCR5 sequence from Bronze Age individuals this study challenges the hypotheses of recent selection of an inactive CCR5 allele by Middle Age epidemic plaque or viral infections.**
  35. Balakirev, E.S. and Ayala, F.J. Pseudogenes: Are they “junk” or functional DNA? *Annu. Rev. Genet.* **37**, 123–151 (2003).
  36. Zhang, Z. and Gerstein, M. Large-scale analysis of pseudogenes in the human genome. *Curr. Opin. Genet. Dev.* **14**, 328–335 (2004).
  37. Römpler, H., Schulz, A., Pitra, C., Coop, G., Przeworski, M., Paabo, S., and Schoneberg, T. The rise and fall of the chemoattractant receptor GPR33. *J. Biol. Chem.* **280**, 31068–31075 (2005). **This report established the concomitant and independent pseudogenization of the GPR33-encoding gene in multiple species, presumably driven by selection.**
  38. Römpler, H., Yu, H.T., Arnold, A., Orth, A., and Schoneberg, T. Functional consequences of naturally occurring DRY motif variants in the mammalian chemoattractant receptor GPR33. *Genomics* **87**, 724–732 (2006).
  39. Gaillard, I., Rouquier, S., Chavanieu, A., Mollard, P., and Giorgi, D. Amino acid changes acquired during evolution by olfactory receptor 912-93 modify the specificity of odorant recognition. *Hum. Mol. Genet.* **13**, 771–780 (2004).
  40. Torrents, D., Suyama, M., Zdobnov, E., and Bork, P. A genome-wide survey of human pseudogenes. *Genome Res.* **13**, 2559–2567 (2003).
  41. Young, J.M. and Trask, B.J. The sense of smell: Genomics of vertebrate odorant receptors. *Hum. Mol. Genet.* **11**, 1153–1160 (2002).
  42. Liman, E.R. and Innan, H. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc. Natl. Acad. Sci. USA* **100**, 3328–3332 (2003).
  43. Zhang, J. and Webb, D.M. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl. Acad. Sci. USA* **100**, 8337–8341 (2003).
  44. Wang, X., Thomas, S.D., and Zhang, J. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum. Mol. Genet.* **13**, 2671–2678 (2004).
  45. Bard, J.A., Nawoschik, S.P., O’Dowd, B.F., George, S.R., Branchek, T.A., and Weinschank, R.L. The human serotonin 5-hydroxytryptamine1D receptor pseudogene is transcribed. *Gene* **153**, 295–296 (1995).
  46. Nguyen, T., Sunahara, R., Marchese, A., van Tol, H.H., Seeman, P., and O’Dowd, B.F. Transcription of a human dopamine D5 pseudogene. *Biochem. Biophys. Res. Commun.* **181**, 16–21 (1991).
  47. Olsen, M.A. and Schechter, L.E. Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. *Gene* **227**, 63–69 (1999).
  48. Weinschank, R.L., Adham, N., Macchi, M., Olsen, M.A., Branchek, T.A., and Hartig, P.R. Molecular cloning and characterization of a high affinity dopamine receptor (D1 beta) and its pseudogene. *J. Biol. Chem.* **266**, 22427–22435 (1991).
  49. Korneev, S.A., Park, J.H., and O’Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* **19**, 7711–7720 (1999).
  50. Zhang, Y., Devries, M.E., and Skolnick, J. Structure modeling of all identified G protein-coupled receptors in the human genome. *PLoS Comput. Biol.* **2**, e13 (2006).
  51. Baldwin, J.M. The probable arrangement of the helices in G protein–coupled receptors. *EMBO J.* **12**, 1693–1703 (1993).
  52. Tarnow, P., Schöneberg, T., Krude, H., Gruters, A., and Biebermann, H. Mutationally induced disulfide bond formation within the third extracellular loop causes melanocortin 4 receptor inactivation in patients with obesity. *J. Biol. Chem.* **278**, 48666–48673 (2003).
  53. Samson, M., Libert, F., Doranz, B.J. et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
  54. Hadley, T.J. and Peiper, S.C. From malaria to chemokine receptor: The emerging physiologic role of the Duffy blood group antigen. *Blood* **89**, 3077–3091 (1997).
  55. Galvani, A.P. and Slatkin, M. Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. *Proc. Natl. Acad. Sci. USA* **100**, 15276–15279 (2003).
  56. Hedrick, P.W. and Verrelli, B.C. “Ground truth” for selection on CCR5-Delta32. *Trends Genet.* **22**, 293–296 (2006).
  57. Schoneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Rompler, H., and Sangkuhl, K. Mutant G protein-coupled receptors as a cause of human diseases. *Pharmacol. Ther.* **104**, 173–206 (2004).
  58. Sangkuhl, K., Rompler, H., Busch, W., Karges, B., and Schoneberg, T. Nephrogenic diabetes insipidus caused by mutation of Tyr205: A key residue of V2 vasopressin receptor function. *Hum. Mutat.* **25**, 505 (2005).
  59. Farooqi, I.S., Keogh, J.M., Yeo, G.S., Lank, E.J., Cheetham, T., and O’Rahilly, S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.* **348**, 1085–1095 (2003).
  60. Larsen, L.H., Echwald, S.M., Sorensen, T.I., Andersen, T., Wulff, B.S., and Pedersen, O. Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. *J. Clin. Endocrinol. Metab.* **90**, 219–224 (2005).
  61. Yeo, G.S., Lank, E.J., Farooqi, I.S., Keogh, J., Challis, B.G., and O’Rahilly, S. Mutations in the human melanocortin-4 receptor gene associated with severe familial obesity disrupts receptor function through multiple molecular mechanisms. *Hum. Mol. Genet.* **12**, 561–574 (2003).
  62. Buono, P., Pisanisi, F., Nardelli, C. et al. Six novel mutations in the proopiomelanocortin and melanocortin receptor 4 genes in severely obese adults living in southern Italy. *Clin. Chem.* **51**, 1358–64 (2005).
  63. Xiang, Z., Litherland, S.A., Sorensen, N.B., Proneth, B., Wood, M.S., Shaw, A.M., Millard, W.J., and Haskell-Luevano, C. Pharmacological characterization of 40 human melanocortin-4 receptor polymorphisms

with the endogenous proopiomelanocortin-derived agonists and the agouti-related protein (AGRP) antagonist. *Biochemistry* **45**, 7277–7288 (2006).

64. Pausch, M.H. G protein-coupled receptors in *Saccharomyces cerevisiae*: High-throughput screening assays for drug discovery. *Trends Biotechnol.* **15**, 487–494 (1997). **Comprehensive review on yeast-based GPCR expression systems used for high-throughput ligand screening and mutagenesis studies. The author's own experimental contributions pioneered this methodical field.**
65. Beukers, M.W. and Ijzerman, A.P. Techniques: How to boost GPCR mutagenesis studies using yeast. *Trends Pharmacol. Sci.* **26**, 533–539 (2005).
66. Ladds, G., Goddard, A., and Davey, J. Functional analysis of heterologous GPCR signalling pathways in yeast. *Trends Biotechnol.* **23**, 367–373 (2005).
67. Ault, A.D. and Broach, J.R. Creation of GPCR-based chemical sensors by directed evolution in yeast. *Protein Eng. Des. Sel.* **19**, 1–8 (2006).



[From left to right] **Michael Hofreiter, PhD**, is an independent Max Planck junior research group leader at the Max Planck Institute for Evolutionary Anthropology, Leipzig. **Holger Römpker, PhD**, and **Angela Schulz, PhD**, are postdoctoral fellows in the Institute of Biochemistry, Leipzig. **Claudia Stäubert** and **Doreen Thor** are graduate students in the laboratory of Torsten Schöneberg. **Torsten Schöneberg, MD**, is Professor of Molecular Biochemistry at the Institute of Biochemistry, Medical Faculty, University of Leipzig.