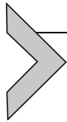


CHAPTER FOUR

Directed Evolution of G-Protein-Coupled Receptors for High Functional Expression and Detergent Stability

Karola M. Schlinkmann, Andreas Plückthun[1]

Department of Biochemistry, University of Zurich, Zürich, Switzerland

[1]Corresponding author: e-mail address: plueckthun@bioc.uzh.ch

Contents

1. Introduction 68
   1.1 G-protein-coupled receptors 68
   1.2 GPCR structures 69
   1.3 Engineering of GPCRs 70
2. Methods 74
   2.1 Generation of genetic diversity 74
   2.2 Expression and selection 86
   2.3 Characterization of selected GPCR variants 91

Acknowledgments 94

References 95

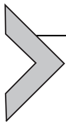
Abstract

G-protein-coupled receptors (GPCRs) are cell-surface receptors exhibiting a key role in cellular signal transduction processes, thus making them pharmacologically highly relevant target proteins. However, the molecular mechanisms driving receptor activation by ligand binding and signal transduction are poorly understood, since as integral membrane proteins, most GPCRs are very challenging for functional and structural studies. The biophysical properties of natural GPCRs, usually required by the cell in only low amounts, support their functionality in the lipid bilayer but are insufficient for high-level recombinant overexpression and stability in detergent solution. Current structural information about GPCRs is thus limited to a subset of GPCRs with either intrinsically favorable or properly improved biophysical behavior. Recently, directed protein evolution techniques for functional expression and detergent stability have been developed to increase the accessibility of GPCRs for functional and structural studies. Directed evolution does not rely on any preconceived notion of what might be limiting biophysical properties. By random mutagenesis combined with a high-throughput screening and selection system, directed protein evolution has the power to efficiently isolate rare

Methods in Enzymology, Volume 520
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-391861-1.00004-6

© 2013 Elsevier Inc.
All rights reserved.

phenotypes and thus contribute to the elucidation of the stability-determining factors, in addition to solving the practical problem of creating stable GPCRs. In the current chapter, protocols for generation of genetic diversity within GPCRs and selection are provided and discussed.



1. INTRODUCTION

1.1. G-protein-coupled receptors

Natural evolution has designed G-protein-coupled receptors (GPCRs) for functionality in a cellular context: As integral membrane proteins they are adapted to the lipid bilayer, and most of them are needed only in very small amounts by the cell. Their mode of action requires a considerable conformational change, transmitting the information of a ligand, which binds on the extracellular side, to the inside of the cell, where a heterotrimeric G-protein binds to the activated receptor (Rosenbaum, Rasmussen, & Kobilka, 2009). The GPCR acts as a guanine-nucleotide exchange factor. After activation, the GPCR is often also phosphorylated, internalized, and degraded to avoid continuous signaling (Ferguson, 2001). The biophysical protein properties of GPCRs have thus evolved to fulfill their key role in signal transduction processes, allowing for efficient and sensitive activation of signaling pathways upon changes in the extracellular environment. Nature has evolved GPCRs to optimally fulfill these requirements *in vivo*, which are almost antithetic to what is desired for *in vitro* characterization purposes.

About 1% of open reading frames in vertebrates code for GPCRs (Bjarnadottir et al., 2006) and the number of class A GPCRs is estimated to cover about 800 different receptors in humans, of which 50% are olfactory receptors (Bjarnadottir et al., 2006; Foord et al., 2005). The remaining 400 GPCRs play key roles in many signal transduction processes making non-olfactory GPCRs as important pharmaceutical targets, and it is estimated that about 30% of all currently marketed drugs target GPCRs (Lagerström & Schiöth, 2008; Overington, Al-Lazikani, & Hopkins, 2006). Almost certainly, among the 400 GPCRs of potential interest, there are many valid targets that have not even been explored.

Although GPCRs are physiologically expressed at low levels, high ligand affinities and strong amplification of downstream signals guarantee specific and efficient signal transduction. Moreover, the low expression levels further account for minimal background signaling activity. Similar to expression levels, the biophysical protein properties have evolved for functionality,

and the limited stability of GPCRs might even be a desired feature in the cellular context to facilitate fast turnover and degradation.

In vitro, however, the protein properties of GPCRs desired and evolved *in vivo* turn out to be a biochemist's nightmare. Many GPCRs cannot be expressed in *Escherichia coli*, otherwise representing the easiest, fastest, and cheapest system to produce high quantities of recombinant protein. Some of them are poorly expressed even in mammalian cells.

Receptor solubilization from membranes to obtain functional and stable receptor in detergent solution is a further crucial challenge along the purification process. Moreover, the intrinsic conformational flexibility of GPCRs, enabling activation of the receptor by conformational changes upon ligand binding, hinders structure determination.

Consequently, our understanding of GPCR architecture and mechanism has remained limited, and the design features of agonists and antagonists for the diverse set of receptors have remained mostly enigmatic. The evident discrepancy between the high pharmacological relevance and the poor status of GPCR characterization accounts for the enormous scientific effort in GPCR research. The effort of decades started to pay off only recently with the crystal structure determination of a handful of GPCR receptors.

1.2. GPCR structures

The first crystal structure of a GPCR, bovine rhodopsin, was solved in the year 2000 (Palczewski et al., 2000) and remained unchallenged for several years until 2007. The pioneer position of rhodopsin in GPCR structural research is a result of its extraordinary and unique features, namely its high natural abundance and high conformational stability and homogeneity as a result of its covalently bound ligand 11-*cis*-retinal, acting as a potent antagonist. However, its uniqueness limits implications for other GPCRs.

Recent GPCR structures, including the inactive states of the human adenosine receptor A_{2A} (Jaakola et al., 2008), human β2-adrenergic receptor (Cherezov et al., 2007; Rosenbaum et al., 2007), turkey β1-adrenergic receptor (Warne et al., 2008), CXCR4, and the human dopamine D3 receptor (Chien et al., 2010), pointed out differences between rhodopsin and the remaining class A GPCRs, which, different from rhodopsin, are liganded by diffusible molecules (reviewed, e.g., in Katritch, Cherezov, & Stevens, 2012). The recent determination of the β2-adrenergic receptor structure in complex with a heterotrimeric G-protein (Katritch et al., 2012; Rasmussen, DeVree, et al., 2011) depicts a landmark in the understanding of the signaling process itself.

Except for rhodopsin, most GPCR structure determination required changes within the protein sequence, including loop deletions, engineered domain insertions, and/or trial-and-error optimization of the protein sequence.

Domain insertions (e.g., T4-lysozyme, Rosenbaum et al., 2007) or deletions of flexible loops (e.g., Cherezov et al., 2007) and the use of binding proteins such as a specific camelid V_HH antibody domain (Rasmussen, Choi, et al., 2011) to stabilize the heterotrimeric G-protein or a conventional antibody F_{ab} fragment (Rasmussen et al., 2007) to facilitate crystal contact formation were successfully used. The binding proteins, however, cannot overcome intrinsic limitations in biophysical properties, and the domain insertions and loop deletions mask functional information about receptor activation mechanisms in the resulting structures. Trial-and-error stabilization in detergents using alanine scanning has been used as well (Serrano-Vega, Magnani, Shibata, & Tate, 2008; Warne et al., 2008).

Despite these efforts, the limited number of receptor structures and the redundancy of the datasets do not reflect the functional diversity of GPCRs, and thus general conclusions about their activation mechanism remain limited as well. Most importantly, fundamental rules for agonist and antagonist design have not yet emerged. So far, structure determination is limited to a subset of GPCRs that can be well expressed and purified and requires already detergent-stable GPCRs. Most GPCRs are not amenable to functional and structural studies, as the bottleneck lies in earlier steps, namely expression and purification, which involves detergent stability as well.

To increase the spectrum of GPCRs accessible for functional and structural analysis and to gain detailed understanding of the GPCR activation and inactivation mechanisms, two main parameters have to be optimized: the recombinant functional expression of a target GPCR and its biophysical protein properties. Improved biophysical properties, mainly stability in detergents, increase the chances of crystal formation, particularly when conformational homogeneity is achieved.

1.3. Engineering of GPCRs

1.3.1 Process engineering

Two different and orthogonal engineering strategies can be distinguished: First and conventionally, alterations are made to the "external" conditions of GPCR expression and purification. Here, we focus on heterologous expression of GPCRs in *E. coli*. This can be subsumed under process engineering, as opposed to protein engineering.

Empiric optimization of expression host, temperature and medium of expression, and plasmid copy number can substantially influence functional expression levels (see, e.g., Tucker & Grisshammer, 1996). For example, GPCRs are typically expressed from low copy plasmids to reduce toxicity of GPCR expression in *E. coli*. We showed in several of our studies that low copy number plasmids are essential for nonoptimized receptors (Sarkar et al., 2008), while GPCRs with improved biophysical properties can be expressed from higher copy number plasmids without toxic effects (Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012).

GPCR overexpression in yeast cells, mammalian cells, or the baculovirus expression system using *Spodoptera frugiperda* (*Sf9*) cells is an alternative. We observed, in the analysis of numerous mutants, a strong correlation between the relative expression levels in all these hosts (Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012). Thus, changes in the protein sequence that have been found to improve expression in *E. coli* have also been found to improve expression levels in eukaryotic hosts, including mammalian cells where the receptors came from.

As a production host, these alternative systems are more time consuming, laborious, and expensive, but need to be used of course when the posttranslational modifications are studied. Yet, for most GPCRs, posttranslational modifications, mainly glycosylation, are not imperative, and the respective site can be mutated or the flexible N-terminus can be deleted for expression in *E. coli*, even though a small subset of GPCRs might potentially remain nonexpressible.

Optimization of GPCR overexpression in eukaryotic hosts is further hampered by the fact that high GPCR levels can lead to increased basal signaling activity, which often interferes with cellular signaling pathways, also leading to high toxicity after overexpression.

Similar to the expression conditions, detergent solubilization from membranes has to be optimized for a given target protein. Many detergents are available, differing in solubilization efficiency and capability to retain the membrane protein in a functional state (see, e.g., Duquesne & Sturgis, 2010; le Maire, Champeil, & Moller, 2000; Seddon, Curnow, & Booth, 2004).

While probably every GPCR will require some kind of process optimization, it currently appears that most members of the family cannot be studied with focusing on the process alone. Furthermore, most processes are not transferable, as the above laborious empiric optimization has to be specifically optimized for each given target receptor. Moreover, in many cases, no feasible conditions at all will be found. Thus, with this conventional

approach alone, most GPCRs would remain inaccessible, as the biophysical protein properties themselves are limiting.

1.3.2 Protein engineering

The second strategy thus focuses on identifying a related protein sequence with improved biophysical properties. A commonly used strategy is to screen the "homology space" for target homologs with similar function, but better protein behavior with respect to expression and biophysical properties. It is commonly found that homologous proteins of bacterial origin, notably from thermophilic bacteria, have more favorable properties than proteins from eukaryotes, and this has been observed for membrane proteins as well (Granseth, Seppala, Rapp, Daley, & Von Heijne, 2007).

Unfortunately, the above strategy cannot be applied to GPCRs, as it appears that prokaryotes do not contain such proteins. Despite the fact that this strategy is popular in current structural biology, the homologs used often share only very low sequence identity, thus potentially limiting the relevance of homolog characterization.

In this situation, protein engineering techniques provide a valuable alternative to the classical (but limited) homology search to identify near-target-like GPCR variants with a desired phenotype, which are first of all functional, stable, and show high expression. Many protein engineering strategies are routinely applied to soluble proteins but cannot easily be transferred to integral membrane proteins. Rational design of a protein sequence relies on sufficient structural and functional information about the target protein in order to design a favorable phenotype—the very reason of writing this chapter is that this information is not available yet. For GPCRs, rational design is thus not an alternative, at least not yet, as the limited structural information and the difficulties of protein expression and purification constitute major roadblocks in the application of this strategy. Furthermore, using classical trial-and-error approaches, many variants would have to be individually designed and tested for a given target protein.

Directed evolution and selection for the desired phenotype provide a more attractive methodology, and the availability of a screening and selection technique would render sampling of highly diverse libraries possible in order to identify a rare mutant with the envisaged phenotype. A main focus in GPCR research is to improve receptor expression levels and stability, in order to increase the diversity of receptors that are accessible to functional and structural studies. It has been shown by Bowie and coworkers (summarized in Bowie, 2001) that stabilizing mutations are not rare in membrane proteins,

emphasizing the great potential of membrane proteins as a target for directed protein evolution.

In our laboratory, we had previously developed an *E. coli*-based selection system to evolve and engineer GPCRs for high functional expression and stability in detergent (Dodevski & Plückthun, 2011; Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012; Schlinkmann, Honegger, et al., 2012). The lack of GPCR-homologs in *E. coli* turns out as an advantage here, as these receptors cannot interfere with any cellular signaling pathway, as opposed to when using eukaryotic expression systems. By multiple and iterative rounds of gene randomization, followed by selection for high functional expression using a fluorescence-activated cell sorting (FACS)-based approach, functional and highly expressed receptor variants can be efficiently identified within less than a month time.

The main prerequisite for this approach is the availability of a functional ligand: The FACS selection is based on positive selection using a fluorescence-labeled ligand, thereby ensuring functionality of the receptor. No further functional or structural information other than the primary protein sequence is needed, as unfavorable variants are efficiently selected against. Furthermore, library sizes of $>10^7$ can easily be transformed into *E. coli* and efficiently be screened by the FACS-based selection system (screening of approximately 10^7 single cells per hour in yield mode). High-efficiency transformation of yeast cells or mammalian cells is not as straightforward and thus less suitable for highly diverse libraries.

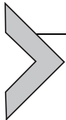
The robustness of our method allows application of stringent selection conditions on receptor libraries with high diversity, as only the coverage of a large mutational space increases the chances to identify rare receptor variants with the desired phenotype. We have successfully applied this methodology to substantially improve the expression levels of several GPCRs from hardly detectable (<500 receptors per cell) to well expressed (6000–25,000 receptors per cell) (Dodevski & Plückthun, 2011; Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012; Schlinkmann, Honegger, et al., 2012).

In the above studies, we further observed an inherent coevolution of detergent stability with functional expression levels, the latter one never being under direct selection pressure. During selection, pressure is applied on functional receptor expression, which is a result of the efficiency of correct protein folding, insertion into the lipid bilayer and stability within the lipid bilayer (Jungnickel, Rapoport, & Hartmann, 1994). It is thus an indirect selection for biophysical properties, emphasized by the fact that mutants selected in

E. coli also express better in eukaryotic cells (Dodevski & Plückthun, 2011; Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012). Nonetheless, functional expression and detergent stability are not directly linked as we find residue substitutions influencing one property but not the other (Dodevski & Plückthun, 2011; Sarkar et al., 2008; Schlinkmann, Hillenbrand, et al., 2012; Schlinkmann, Honegger, et al., 2012).

Different from our selection technique, *in vitro* alanine scanning solely for detergent stability leads to uncoupling of receptor expression and stability, and coevolution is thus unlikely to be detected (Shibata et al., 2009).

The selected receptor variants display very high sequence identity to their original target (usually above 95%). Only a few amino acid changes are necessary to significantly improve the expression and biophysical behavior of the target protein, clearly illustrating that GPCRs maintain a delicate balance between stability in the membrane, flexibility required for signaling, and the subsequent steps of receptor inactivation and degradation or recycling (Deupi & Kobilka, 2007). These constraints limit stability and at least partly explain the paucity of structural information from this large family, despite Herculean efforts.



2. METHODS

2.1. Generation of genetic diversity

The described diversification methodologies are applicable to all GPCRs for which the coding sequence is known and which can be successfully cloned into an *E. coli*-expression vector. If the below-described FACS-based selection method is applied, a respective receptor ligand has to be available (see above).

2.1.1 Error-prone PCR to randomize GPCR sequences

By default, random mutagenesis is a stochastic process, without any external influence on the distribution of the introduced mutations. This assumption-free process is an easy and fast method to create genetically diverse receptor libraries on DNA level. Random alterations are introduced to the receptor coding sequence by error-prone PCR, and the mutational load per receptor sequence can be controlled by adaptation of the reaction conditions. Error-prone PCR amplification should be limited to the receptor coding sequence to keep the fusion protein tags intact and functional. Evidently, diversification can be further restricted to a specific receptor region, if desired.

The obtained mutational load is a product of the number of base misincorporations per amplification round and the number of amplification rounds. Both factors need to be adjusted to ensure optimal results since the optimal mutational load highly depends on the target protein and the desired phenotype. From our experience, in the case of GPCRs under functional selection of ligand binding, a low mutational load of 1–5 amino acid substitutions per randomization round is suitable. The argument for choosing a low error rate and rather more selection cycles are as follows: The 7 transmembrane helices of the GPCR are long stretches where the introduction of even one charged residue would render the GPCR nonfunctional. Such a substitution would therefore mask the beneficial effect of other mutations, and the clone would be lost. Thus, we have to tread carefully and add the mutations slowly, "purifying" the population by selection rounds after mutagenesis. Error-prone PCR mainly introduces single-base changes, and a subset of substitutions will hence remain silent on the amino acid level, but they can become nonsilent with further diversifications rounds. Even some mutations on the amino acid level will only show their beneficial effect in the presence of existing mutations. These beneficial "neutral drift" phenomena are an inherent property of the evolutionary process.

Different alternatives exist for gene diversification: First, a low-fidelity DNA polymerase such as *Taq* DNA polymerase can be employed. To support base misincorporations, Mg^{2+} concentrations, Mn^{2+}, or the amount of polymerase are elevated to increase the likelihood of continuous strand synthesis after mismatches, and unbalanced nucleotide mixtures or nucleotide analogues can be used to favor mismatches (Cadwell & Joyce, 1994; Spee, de Vos, & Kuipers, 1993). However, the obtained diversification is highly nonrandom, since *Taq* DNA polymerase favors AT to GC substitutions over others (Wilson & Keefe, 2001). A different mutational bias (GC to AT) can be obtained by using Mutazyme® DNA polymerase (Stratagene GeneMorph Kit). The need for random diversification led to the development of a new and optimized enzyme blend, Mutazyme II® (Stratagene GeneMorph Kit II), a DNA polymerase with reduced mutational bias by combination of the Mutazyme® with a novel *Taq* DNA polymerase (e.g., Vanhercke, Ampe, Tirry, & Denolf, 2005). A routinely used method in our laboratory employs the Mutazyme II® DNA polymerase.

The GPCR coding sequence of interest is used for error-prone PCR, and 10 ng of template DNA is used as input. Flanking PCR primers should include suitable restriction sites for subcloning into the expression vector. Low template DNA input decreases the final fraction of wild-type sequence

in the diversified sequence pool and increases the likelihood that amplification and diversification in further error-prone PCR rounds starts from a previously diversified template. Thirty cycles of PCR amplification are routinely used for error-prone PCR, and the diversified PCR product is treated with DpnI to digest template sequences, as it is specific for methylated DNA as produced in *E. coli* in the form of the starting plasmid. A subsequent PCR amplification with a high-fidelity polymerase, for example Phusion® Polymerase (Finnzyme), is used to obtain sufficient quantities for the subsequent cloning steps. The diversified PCR product is then purified, if necessary from a preparative agarose gel to avoid carryover of PCR side products, and is subsequently digested with the flanking restriction enzymes. A further purification step is used to obtain the final PCR product for subcloning, typically, 2–3 μg of DNA.

Despite optimization of error-prone PCR conditions to minimize mutagenic bias, it has to be considered that error-prone PCR will still favor certain amino acid substitutions over others, as some substitutions would require two consecutive base changes, which are statistically unlikely to happen.

Random mutagenesis using error-prone PCR was successfully applied to create genetic diversity within the rNTR1 (Sarkar et al., 2008), the adrenergic receptors α_{1a} and α_{1b}, and the tachykinin receptor NK_1 (Dodevski & Plückthun, 2011) for selection for high functional expression and stability. Notably, the tachykinin receptor NK_1 did not only evolve to higher expression levels, but could also be evolved for functional extraction from the lipid bilayer by detergent treatment, which is not possible for the wild type (Dodevski & Plückthun, 2011). For all receptors, increase in expression levels was associated with improvement of biophysical protein properties.

2.1.2 Single amino acid scanning mutagenesis approaches

Alanine-scanning mutagenesis is a common approach to identify positions of a target protein that are crucial for a desired phenotype (Clackson & Wells, 1995; Wells, 1991). Every amino acid position of a given target protein is sequentially and separately replaced by alanine and every mutant is analyzed for the desired phenotype.

With respect to GPCRs, the high helical propensities, the small side chain size, and the relative inertness of alanine make it a most likely tolerated substitution in most receptor positions, and might in principle improve the biophysical properties of the receptor by improving helix propensities and packing. Yet in a comprehensive all-versus-all screen (Schlinkmann, Honegger, et al., 2012), almost no alanines were the most preferred amino

acid type. By alanine scanning alone, mutational space is only minimally covered, and the relevance of a specific position for detergent stability may not be identified if the favorable effect would not be conveyed by Ala, but only by a different amino acid substitution.

Nonetheless, alanine-scanning mutagenesis has been successfully applied to two GPCRs, rNTR1 (Shibata et al., 2009) and the human β1-adrenergic receptor (Serrano-Vega et al., 2008). Recently, leucine-scanning mutagenesis of the human β1-adrenergic receptor was performed and revealed further beneficial substitutions (Miller & Tate, 2011).

2.1.3 Technical aspects of mutant construction

The production of a collection of single mutants is a straightforward point mutagenesis, albeit becoming laborious with the number of mutants to be generated. For each mutant to be constructed, a mutagenic and complementary primer pair covering the codon of interest and introducing the mutation is required. Two cloning strategies can be applied: First, the mutagenic primer pair can be used for PCR amplification of the entire expression vector, for which only the mutagenic primer pair is needed. The PCR is performed for 15–18 cycles using a high-fidelity DNA polymerase such as *Pfu* DNA polymerase (Promega). For elongation, 2–2.5 min/1000 base pairs (bp) are recommended for optimal performance of the reaction. Digestion of the input template DNA is achieved by DpnI treatment, followed by transformation of *E. coli* by the amplified expression vector. With complete digestion of the input DNA, only mutant clones should be obtained. Plasmid DNA from a single colony is then isolated, the sequence is recloned into a fresh vector (to eliminate any spurious introduction of backbone mutations in the vector which could mask the true phenotype), and the sequence of the final construct is verified by sequencing.

Second, a two-step assembly PCR strategy (Fig. 4.1A) can be alternatively used. In a first PCR amplification round, the two following PCR fragments are generated: A flanking forward primer, introducing the 5′ restriction site, and the reverse mutagenic primer are used to obtain a PCR fragment covering the 5′-half of the receptor sequence, including the desired mutation. Similarly, the 3′-half fragment including the desired mutation is obtained from a PCR using the forward mutagenic primer and the reverse flanking primer, introducing the 3′- restriction site. The quality of the two PCR products is analyzed on an analytical agarose gel, and PCR products are purified from a preparative agarose gel in case additional side products are detectable. If large numbers of mutants have to be generated, it is worth to optimize PCR conditions such that a pure and single PCR product of correct length is obtained

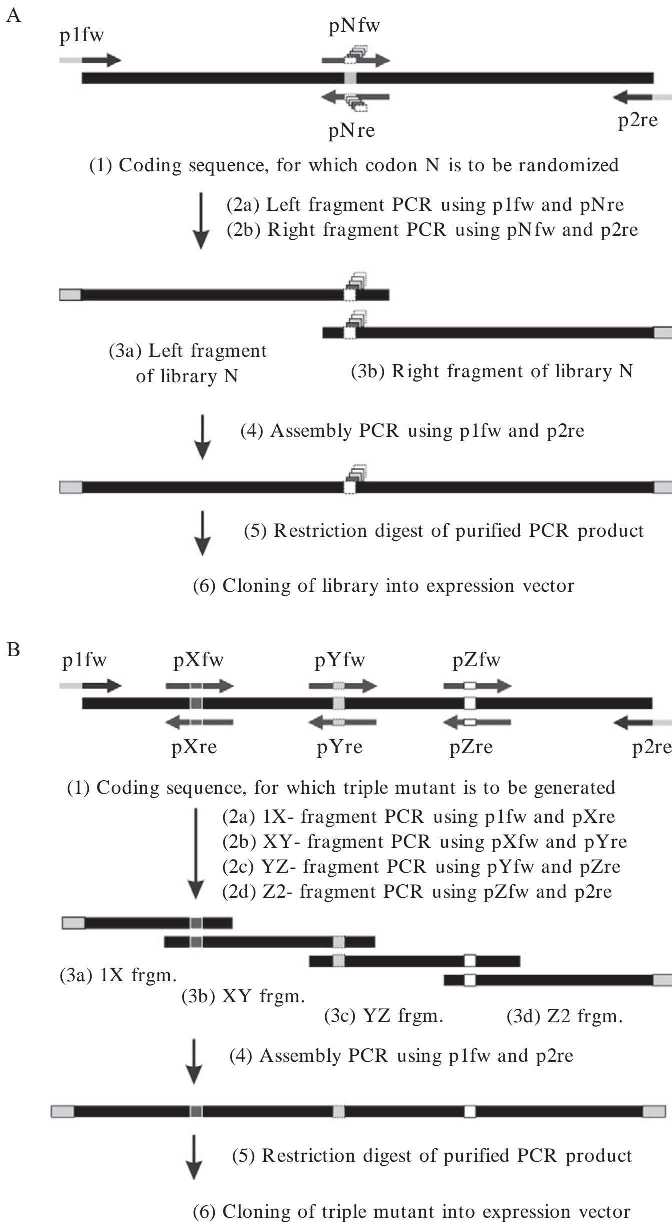


Figure 4.1 Generation of position-specific randomized GPCR libraries (A) and multiple GPCR mutants (B). (A) Fully randomized position-specific libraries are generated by a two-step PCR assembly strategy. First, two separate PCRs are performed with the GPCR coding sequence (1) as template: with primers p1fw and pNre (2a), the 5′-end of the library N is generated. Primers pNre and pNfw are NNN-randomized in the codon of library

every time. In this case, the output of the first PCR amplification round can be directly used as input for the second PCR assembly step. Twenty to one hundred nanogram of each purified PCR fragment or 1–2 μl of each unpurified PCR is used for extension of the two fragments and assembly of the full-length mutagenized receptor sequence. Both PCR products overlap in the mutagenized region, and the mutagenic primer pair should be designed such that the overlap is 25–30 bp. The flanking PCR primers are used for PCR amplification of the full-length mutagenic sequence.

For fragments of similar length, a standard amplification PCR protocol is working well. However, particularly for fragments with larger differences in length, the extension of the two fragments from the overlap region before assembling them is beneficial for obtaining the specific PCR product. For this purpose, 5–10 PCR cycles are performed in the absence of the flanking primers, and 25–30 amplification rounds are subsequently performed after addition of the flanking primers. If the position to be mutagenized is close to one end of the receptor sequence, the corresponding flanking primer can be elongated to include the target position and introduce the desired mutation. In this case, only one PCR amplification step is necessary. Primers of approximately 100 nucleotides have been successfully used by the authors to introduce mutations.

The assembly strategy can be easily employed for fast generation of multiple defined mutants by generation of multiple mutagenic and overlapping fragments that are assembled to the full-length construct in the second PCR step. A triple mutant can be successfully obtained within one round of assembly (Fig. 4.1B). For this purpose, mutagenic primer pairs covering the target positions, denoted X, Y, and Z here in sequential order, are designed to introduce the desired mutations. In the first PCR amplification, four fragments are generated: The first fragment covers the region from the 5′ end to the most upstream mutation X (obtained with the 5′-flanking primer and the reverse mutagenic primer X), the second fragment reaches from mutation X

position N, thus introducing the desired randomization. For a specific point mutation, primers pNfw and pNre have a defined sequence at the target position. Primers p1fw and p2re introduce restriction sites for subcloning into the expression vector (overhanging gray ends). With primers pNfw and p2re (2b), the 3′-end of library N is generated. The resulting PCR products (3a, 3b) are isolated and purified and used as template for the subsequent assembly PCR (4). Primers p1fw and p2re are used to generate and amplify the full-length library PCR product from the two fragments. The full-length library is purified and subsequently cloned into the expression vector. (B) A triple mutant is assembled from four separate fragment PCRs similar to the position-specific library in (A).

to mutation Y, and corresponding fragments are amplified from Y to Z and Z to the 3′ end. Individual fragment lengths should be 200 bp at minimum for optimal results. All fragments are then combined in a second PCR step for assembly of the full-length sequence by amplification with the flanking primers as described above. If the assembly of four fragments is inefficient or results in PCR side products, it is recommend to assemble two overlapping fragments in two separate reactions and to assemble the full-length sequence from these intermediate fragments.

The full-length mutagenic fragment is then purified, if necessary from a preparative agarose gel, and the flanking restriction sites are digested by the respective restriction enzymes. The purified mutagenic fragment is then ligated into the expression vector and *E. coli* is transformed with it. The DNA of a single colony is isolated and sequence-verified.

The second strategy seems to be more work-intensive at first sight. However, it contains only one cloning step, as every sequence is cloned directly into a fresh expression vector. In contrast, amplification of the entire expression vector by PCR can easily accumulate mutations in the vector backbone, affecting origin of replication or the promotor, for example. In order to avoid any spurious mutation, the mutagenized receptor sequence has to be cut and ligated into a fresh vector backbone by restriction digest, hence making this strategy actually more time consuming than the second strategy.

2.1.4 Comprehensive mutagenesis

A more integrative method is to explore the entire mutational space by full randomization of a specific receptor position with selection for the amino acid variant conveying the desired phenotype. With our FACS-based selection system, a powerful and efficient selection method exists to select for high functional expression and stability (Sarkar et al., 2008).

Because of the incomplete coverage of mutational space as well as target sequence space by random mutagenesis, we recently performed a saturating and exhaustive mutagenesis on rNTR1-D03 to determine for every position the residue types that are not permitted, permitted, and preferred (Schlinkmann, Honegger, et al., 2012). Importantly, the already improved mutant rNTR1-D03 was used as a framework, since rNTR1-wt expression levels were too low for these experiments.

A full coverage of mutational space should ideally include the full codon diversity, to ensure phenotype selection and exclude any bias from variable tRNA levels, mRNA secondary structure, or other undesired influences.

The occurrence of several or all codons of a preferred amino acid after selection internally verifies the selection outcome.

Generally, the strategy to fully randomize a specific amino acid position, that is to create a position-specific library, is based on the generation of a single point mutagenesis (Fig. 4.1A). The mutagenic primers contain an NNN sequence at the target codon, where N denotes an equimolar mixture of all four nucleotides. An NNN mixture also includes stop codons, and depending on the screening and selection technique that is employed, the primer design may have to be adapted to exclude stop codons. However, with FACS-based selection for functional expression using fluorescent ligands, stop-codon mutants are counter-selected, as most of these mutants do not contain a functional ligand binding site (unless the stop codon occurs after TM7). Thus, the counter-selection of stop codons is a useful internal quality control of the selection success.

Primers should be designed such that the hybridization temperature with the template is between 55 and 65 °C so that differences in specific primer sequences are negligible. Two overlapping half fragments of the target sequence are then generated similar to the approach described above, extended and amplified. Very importantly, throughout all steps of library generation, care must be taken to preserve library diversity. Thus, DNA amounts representing a number of molecules equivalent to 10- to 20-fold of the library diversity should be used as input in every PCR step. The expression vector for subcloning of the position-specific library should not contain a wild-type receptor sequence in order to avoid wild-type overrepresentation due to incomplete vector digest (for details, see Schlinkmann, Hillenbrand, et al., 2012).

2.1.5 In vitro *DNA shuffling of GPCR sequences to generate highly diverse chimeric receptor libraries*

In vitro DNA shuffling is used to generate chimeras from two receptor sequences. We have recently adapted and optimized the staggered extension process (StEP; Fig. 4.2A) for generation of chimeric libraries starting from the rNTR1-D03 and a mutagenized artificial receptor sequence, rNTR1-M30 or rNTR1-M303 (Schlinkmann, Hillenbrand, et al., 2012). StEP is a PCR-based approach (Aguinaldo & Arnold, 2002; Zhao & Zha, 2006), in which two or more different sequences are used as input templates. By using a very short combined annealing and extension step at a suboptimal DNA-polymerase elongation temperature, the amplification primers are only extended by a few nucleotides per StEP-PCR cycle. After the subsequent denaturation step, the primers are further extended, until eventually a full-length receptor sequence is obtained. Most importantly,

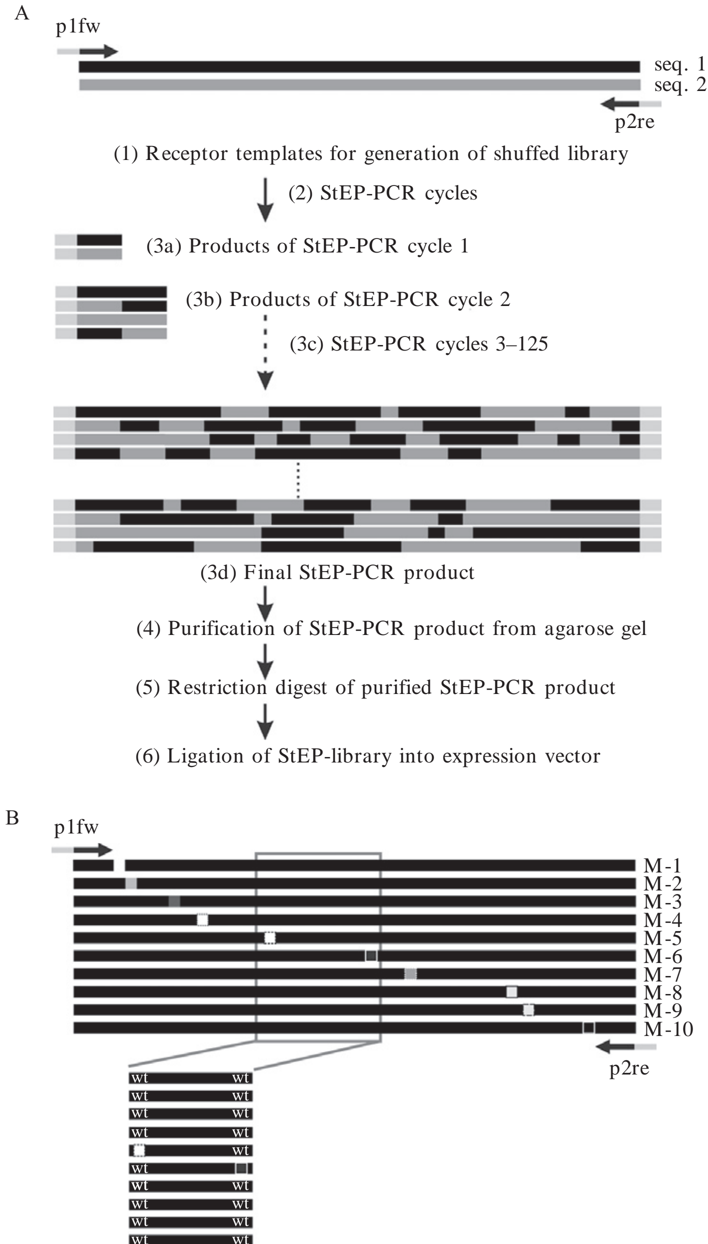


Figure 4.2 Generation of shuffled GPCR libraries using StEP-PCR. (A) Two (or more) different GPCR templates are used for *in vitro* DNA shuffling, denoted as seq. 1 and seq. 2 here (1). By using high numbers of very short StEP-PCR cycles, the flanking primers are only extended by a few nucleotides (3a–3c) until eventually, a full-length and

by using two or more different input templates, the growing primer fragment can switch templates between two PCR cycles, and thus accumulates mutations from two different templates in one chimeric receptor sequence.

For shuffling of two GPCR variants of approximately 1200 bp, 10 ng of each plasmid template DNA is mixed per 50 μl PCR using 2 units $Vent_R^{®}$ DNA Polymerase (NEB) and 30 pmol each of the two flanking primers, introducing a restriction site. The choice of DNA polymerase greatly affects shuffling efficiency, and the following aspects should be considered: High fidelity is required to avoid undesired mutations as a result of high number of StEP-PCR cycles. To yield short recombination distances, a slower DNA polymerase is preferred, for example $Vent_R^{®}$ (NEB, ~1000 bp/min) over Phusion® DNA polymerase (Finnzymes, 1000 bp/15 s).

The amplification yield in a StEP reaction is comparably low, and further PCR amplification of a single StEP reaction should be avoided, as it does not increase diversity. With a theoretical diversity of about 10^7, 12 reactions are setup in parallel to increase the product amount while simultaneously generating high diversity. StEP shuffling is performed for 125 PCR cycles on a Biometra T3 thermocycler (heating rate of 2 °C/s) with 30 s denaturation at 94 °C and 6 s annealing/elongation at 50 °C per cycle and 2 min of initial denaturation. A final and extended elongation step should be omitted, as it could lead to amplification of the template sequences without shuffling.

StEP is a delicate PCR-based process, reacting strongly to small changes in reaction conditions. The most important parameters for optimization and troubleshooting are duration and temperature of annealing and elongation, which are key determinants of shuffling efficiency. Under these conditions, recombination events within 30-bp distance are obtained. Importantly, the differences in heating and cooling rates between thermocyclers strongly influence the recombination efficiency and yield of the StEP process, and should be controlled and adjusted together with the elongation conditions. With slow heating and cooling rates (2 °C/s), the actual window of

---

chimeric GPCR sequence is generated (3d). By template switches between the StEP-PCR cycles, mutations from two templates are combined in one StEP-PCR product (3c). The StEP-PCR product is then purified from an agarose gel (4) and the flanking restriction sites are digested (5) for ligation into the expression vector (6). (B) Recombination efficiency of a StEP reaction on 10 individual point mutants will suffer from “dilution” of a given mutation with wild type (wt) from other mutant sequences, and the observed net recombination efficiency is much lower than the actual efficiency, as many recombination events will not result in sequence changes.

DNA-polymerase activity is longer than defined by the annealing and elongation cycle (here 6 s per cycle), compared to thermocyclers with fast heating rates (up to 6 °C/s). Under these conditions, elongation times might have to be extended to allow sufficient product formation.

$Vent_R$® and Phusion DNA polymerases both exhibit 3′- to 5′-proofreading exonuclease activity, which, for the case that an incorporated mutation locates at the 3′-end of the growing fragment, could lead to correction by the polymerase proofreading activity after template switching. Thus, a DNA polymerase lacking 3′- to 5′-proofreading exonuclease activity, for example Deep $Vent_R$™ (exo^-) (NEB), might at first seem attractive with respect to recombination efficiency, but it resulted in high amounts of PCR side products in our experiments and was thus not used.

Furthermore, despite the presence of 2 m*M* $MgSO_4$ in the PCR buffer, we observed that a further increase to a final of 4 m*M* $MgSO_4$ positively affected the reaction yield, probably by stabilizing DNA-polymerase complexes after mismatches.

The shuffled StEP product is digested with DpnI to minimize the carryover of input templates. The StEP product should be purified from a preparative agarose gel, as PCR side products are common to StEP reactions. The purified product is digested with the corresponding restriction enzymes and purified. At least 3 μg of final StEP product should be obtained for subcloning of a product with a theoretical diversity of 10^7.

Shuffling by StEP is an easy and fast technique to generate a chimeric library from two or more target sequences. However, mutations close in sequence (3–30 bp) are inefficiently separated by StEP, and sequences with coupled mutations are overrepresented, compared to the recombined sequences. If more than two templates are shuffled, the apparent recombination efficiency can suffer from a "dilution effect" (Fig. 4.2B): If 10 individual point mutants of a given receptor are used as input templates for StEP shuffling, one sequence will contain a particular mutation, while nine templates contain the wild-type codon in the respective position. Statistically, eight of nine recombination events will shuffle wild type against wild type and the accumulation of mutations in one shuffled sequence is consequently low.

Alternatively, an artificial receptor sequence combining all mutations of interest can be synthesized and shuffled against the wild-type sequence for an mutant to wild-type ratio of 1:1 (e.g., Schlinkmann, Hillenbrand, et al., 2012). Evidently, the above effect can be easily exploited to direct and influence recombination by adjustment of template ratios and template design.

The selection output from a diverse StEP-library can be readily subjected to a further StEP shuffling by plasmid DNA isolation from the selected cell pool.

2.1.6 Cloning and transformation of GPCR libraries with high diversity

Cloning and transformation of single point mutants, that is, single plasmids, is straightforward and explained elsewhere (Sambrook & Russel, 2001).

Library cloning and transformation is technically more demanding, and care must be taken to ensure full library diversity throughout all cloning steps. In case of random mutagenesis, library diversity can be estimated by multiplying the number of substitutions per sequence with the number of molecules in the reaction. For a StEP shuffling of two sequences, the theoretical diversity can be calculated from the mutational load of the input templates. In a StEP reaction from two templates, which differ in 30 amino acid positions, the theoretical diversity is given as 2^{30} ($\sim 10^9$).

For a theoretical diversity of 10^7 and a 1200-bp StEP product to be subcloned, 3 μg of purified product DNA is ligated into the expression vector with a threefold molar excess of insert DNA over vector DNA and 10 units of T4 DNA ligase per μg of DNA in the ligation mix in a total volume of 500 μl. Ligation is performed for 12–16 h at 16 °C. Ligation products are then purified using, for example, Qiagen MinElute columns. Column purifications are quick; the final concentration of the ligated product can be controlled by the elution volume and the product is quantitatively recovered. Other methods such as DNA precipitation or purification from agarose gels can be alternatively used, but are more time consuming and less quantitative.

Usually, the limiting step in keeping library diversity is transformation of *E. coli*. Electrocompetent *E. coli* cells are superior to chemocompetent cells with regard to transformation rates and should be routinely used for library transformations (for protocols, see, e.g., Chuang, Chen, & Chao, 1995; Dower, Miller, & Ragsdale, 1988). The amount of DNA per electroporation reaction has to be optimized with respect to the cell density of electrocompetent cells. Routinely, a maximum of 1 μg ligation product in a 5-μl volume, preferably water or a 5-m*M* Tris-buffered solution, is transformed per 100 μl of electrocompetent cells. DNA and cells are premixed on ice, and transferred to a prechilled electroporation cuvette (2 mm, for example Eurogentec electroporation cuvettes). A Gene Pulser® II electroporator (Biorad) is used to electroporate the DNA–cell mixture at 2500 V with a capacitance of 25 μF and 200 Ω resistance. Time constants should be above 4 ms, ideally 4.5–5 ms to ensure high efficiency of transformation. The given protocol is found to give optimal

electroporation efficiency, which is particularly affected by changes in the electroporation volume, and larger volumes per electroporation cuvette will dramatically decrease electroporation efficiency.

Electroporated cells are directly recovered in 1 ml SOC medium for 1 h at 37 °C in a shaking incubator. Directly after the 1 h recovery of the transformed cells, dilution series are plated on agar dishes to determine electroporation efficiency (10 μl, 10^1- to 10^7-fold dilutions). Typically, 5×10^7 to 3×10^8 colonies can be obtained using the described procedure. The recovered cells are subsequently diluted into 500 ml 2YT medium, supplemented with 1% glucose and antibiotic selection marker and grown for 12–16 h at 28 °C in a shaking incubator. Growth temperature should be low to minimize possible growth differences between mutants. The final cell density should be approximately 10^9 per ml of culture volume (OD_{600} of 1). For long-term storage, aliquots of $>10^9$ cells are supplemented with 20% glycerol, snap-frozen in liquid N_2 and stored at −80 °C until further use.

Optimally, single colonies of the naïve library should be analyzed by sequencing of the receptor sequence to ensure library quality and analyze randomization or shuffling efficiency. In case of high genetic diversity, 48–96 colonies should be assayed at minimum.

2.2. Expression and selection

2.2.1 Design of expression vector and GPCR fusion construct

As discussed above, the expression vector and the receptor construct have to be empirically optimized for a given target receptor, for which the following general considerations apply.

First, depending on the expression host, GPCRs are often expressed as fusion proteins to allow efficient targeting of the receptor to the lipid bilayer. For expression of the wild-type rat neurotensin receptor in *E. coli*, expression levels were highest when expressed with an N-terminal maltose binding protein (MBP) and a C-terminal thioredoxin fusion (Tucker & Grisshammer, 1996).

For receptors that do not contain any N-terminal domain that is directly involved in ligand binding and that does not contain critical modifications, the flexible N-terminus can often be truncated. For some receptors, however, the N-terminal domain is large, and involved in ligand binding and GPCR activation (see, e.g., Pin et al., 2004). The rNTR1 receptor, for example, is expressed with deletion of the first 42 N-terminal amino acids (Grisshammer, Duckworth, & Henderson, 1993). All crystallographically

solved nonrhodopsin GPCR structures have truncated or modified N- and C-termini (Katritch et al., 2012).

MBP is connected to the GPCR target by a flexible linker including a protease cleavage site, for example, tobacco etch virus protease. MBP is efficiently targeted to the periplasmic space by its signal sequence, thereby directing the GPCR to the inner membrane. Even though the GPCR can be incorporated into the membrane without this fusion and without a native signal sequence, the use of this MBP fusion system may better guide the receptor to the Sec translocon in *E. coli*.

C-terminally, thioredoxin is fused to the GPCR, again via a flexible linker and a protease cleavage site, and is followed by a His_{10}-tag for purification purposes. Whether thioredoxin, a small well-folding and soluble protein, really serves as a "folding chaperone" (Tucker & Grisshammer, 1996) will require further investigations, and it might suffice that it provides a defined soluble and folded domain that helps correct positioning of the C-terminal end on the cytoplasmic side. In any event, thioredoxin considerably affects the functional expression level of a GPCR in *E. coli*.

Different fusion protein tags and purification tags might be tested for optimal expression of a specific target GPCR. We have replaced the C-terminal His_{10}-tag by an AviTag sequence (GLNDIFEAQKIEWHE, biotinylation on K) for enzymatic *in vivo* biotinylation of the receptor fusion construct (Dodevski & Plückthun, 2011) by the *E. coli* biotin protein ligase (BirA). *In vivo* biotinylation using an AviTag sequence is simple, stoichiometric, specific, and quantitative, and hence superior to chemical biotinylation of the purified receptor. While quantitative biotinylation of a highly expressed recombinant protein requires coexpression of BirA ligase and addition of free biotin, these measures are not necessary for *in vivo* biotinylation of GPCRs expressed at only about 500–6000 receptors per *E. coli* cell.

Second, promotor strength and plasmid copy number are critical determinants of GPCR expression, as has already been mentioned above, and have to be adapted to avoid toxicity of GPCR expression in *E. coli*. A low copy plasmid should be used for difficult-to-express target receptors, combined with a tunable and tight promotor. For expression of wild-type rNTR1, a pBR322-derived origin of replication and the *lac*-promotor is used (Tucker & Grisshammer, 1996). As discussed above, with improvement of the biophysical receptor properties (e.g., rNTR1-D03, see Sarkar et al., 2008), higher plasmid copy numbers are tolerated, and plasmid copy number can even limit expression levels of superior receptor variants (Schlinkmann, Hillenbrand, et al., 2012) where thus still higher copy numbers should be used.

2.2.2 *Expression of highly diverse GPCR libraries in* E. coli

GPCRs are generally expressed at low temperatures in *E. coli*, leading to decreased protein synthesis rates, probably positively affecting targeting efficiency to and insertion efficiency into the lipid bilayer. The rNTR1 is expressed in the *E. coli* strain DH5α. A given volume of 2YT medium, supplemented with 0.2% glucose and selection marker, is inoculated to a cell density of $OD_{600}=0.05$, and grown at 37 °C in a shaking incubator to $OD_{600}=0.5$. Expression is then induced by addition of 250 μ*M* IPTG (isopropyl-β-D-thiogalactopyranoside), and expression is continued for 18–24 h at 20 °C. Receptor variants with improved biophysical properties tolerate higher expression temperatures up to 30 °C (Schlinkmann, Hillenbrand, et al., 2012).

For high-diversity libraries, inoculation density has to be controlled to sustain library diversity at this step. Expression temperatures should be kept low, that is, 20 °C, for the naïve library and optionally for the first rounds of selection to minimize possible differences in growth behavior of individual receptor variants.

If the cell density of the glycerol stock is known, the cell number needed for an inoculation density of $OD_{600}=0.05$ can be calculated on the assumption that 10^9 cells/ml in a cuvette with 1 cm pathlength equal 1 OD_{600}. The cell number used for inoculation should again cover 10- to 20-fold the library diversity. It is recommended to increase expression culture volumes, and not inoculation density, if the above recommendation does not hold for a given expression volume. A library with a diversity of 10^7–10^8 variants is expressed in 60-ml culture volume in a 300-ml Erlenmeyer flask. To reach an inoculation density of $OD_{600}=0.05$ in a volume of 60 ml, 3 OD units of cells, equaling 3×10^9 cells, are needed, thus oversampling library diversity 30- to 300-fold.

2.2.3 *Fluorescence-labeling of GPCR-expressing* E. coli *cells*

FACS selection is applicable to any receptor for which a known ligand with reasonable affinity exists that can be fluorescence-labeled. Peptide ligands and many small molecules are known to work well, whereas ligands that are too hydrophobic tend to bind nonspecifically to the cells, thus making their use more difficult with respect to receptors with low basal expression levels, as the ratio specific:nonspecific signal is low. The size of the labeled ligand affects permeability and labeling efficiency, and small ligands of approximately 1 kDa diffuse well through the permeabilized outer membrane, while also larger ligands of up to 10 kDa in size were shown to penetrate the outer membrane (Chen et al., 2001) after suitable permeabilization. However, for selection

using FACS, cell viability after recovery from FACS selection is a relevant parameter that might suffer from harsh permeabilization conditions and should be tested experimentally for a specific buffer.

The fluorescence label has to be compatible with excitation wavelengths and emission filter wavelengths on the respective FACS machine, which are most commonly equipped with a 488-nm laser and a 633-nm laser (for example, BD FACS Aria Series). Depending on the machine configuration, a 355- or 405-nm UV laser, or a 561-nm laser might be available for excitation.

For selection of well-expressed receptor variants by FACS, cells are labeled relative to their expression levels by use of a fluorescence-labeled receptor ligand. In case of rNTR1, a BODIPY-labeled neurotensin peptide is used (BODIPY-neurotensin (8–12), BP-NT) (Sarkar et al., 2008). The outer membrane of *E. coli* is gently permeabilized to allow diffusion of the BP-NT to the inner membrane, where the receptor is located. A Tris-salt buffer (50 m*M* Tris–HCl pH 7.4, 150 m*M* KCl, abbreviated TKCl buffer) is found optimal for permeabilization of *E. coli* DH5α cells expressing rNTR1 (Sarkar et al., 2008), but might differ for a specific target receptor and receptor ligand (Dodevski & Plückthun, 2011). Sodium salts should be avoided when working with rNTR1, as the receptor is sodium-sensitive (Martin, Botto, Vincent, & Mazella, 1999). Depending on the library diversity, an aliquot of cells covering 10-fold library diversity should be used for selection. The volume for permeabilization and labeling should be adjusted accordingly to ensure efficient labeling, and can be concentrated during later washing steps if necessary. Cell densities of 5×10^7 to 2×10^8 are suitable under these conditions. Cells are collected by centrifugation for 3 min at $6000 \times g$, washed once in TKCl buffer and resuspended in the appropriate volume of TKCl buffer. BP-NT (20 n*M*) is added, and permeabilization and labeling is performed for 1–2 h on ice in the dark. The optimal ligand concentration should be at least 10-fold above K_D, to ensure quantitative binding to receptors. The apparent K_D needs to be tested in a saturation binding experiment before selections, as the ligand diffusion across the permeabilized outer membrane might cause the system to not be at full equilibrium. Furthermore, if receptor expression levels per cell are expected to be high, the ligand concentration might have to be increased to prevent ligand depletion. Nonspecific binding is assayed in the presence of 10 μ*M* unlabeled neurotensin peptide (AnaSpec). Cells are then washed twice in 1 ml TKCl buffer, and resuspended in 1–2 ml of TKCl buffer and directly subjected to screening and selection for high functional expression by FACS.

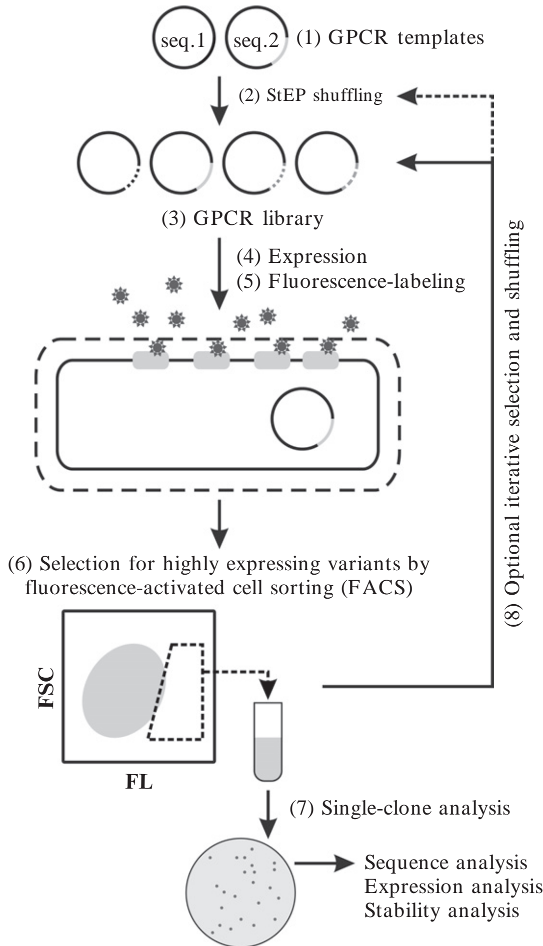


Figure 4.3 Selection for high functional expression. Here, a GPCR library is generated from two GPCR templates, denoted as seq. 1 and seq. 2, (1) using StEP shuffling (2). The resulting library (3) is electroporated and expressed in *E. coli* DH5α cells (4). The fluorescence-labeled agonist BP-NT (BODIPY-neurotensin (8–13)) is used to label the cells relative to their functional expression levels (5). Highly expressing cells are identified and isolated by fluorescence-activated cell sorting (FACS) (6). Individual selected clones are either grown for single-clone analysis (7) or the selected cell pool is subjected to iterative selection rounds with optional reshuffling by StEP (8).

2.2.4 Selection for high functional GPCR expression using FACS

The labeled and washed receptor-expressing cells are subjected to selection using FACS (Fig. 4.3). We routinely work with a BD FACS Aria I. Unlabeled cells are used to gate for the viable cell population. For selection for high functional expression, the fluorescence signal of the viable cells is used to gate for highest cell fluorescence, hence functional expression.

If applicable, multiple fluorescence parameters can be analyzed and gated in parallel or consecutively. The 1% highest fluorescent cells are gated and selected. Selection stringency can be adjusted by the gate size and the selection mode. Gate sizes of 0.5–2% are recommended, with higher stringency in subsequent rounds of selection. The purity of selection is adjusted by the sorting mode, where yield mode should be used for naïve libraries to recover any positive cell within the gate, while purity mode is routinely used for subsequent rounds to avoid carryover of nongated cells (refer to the manufacturer's FACS manual for further detail). On a BD FACS Aria I, flow rates of 5000–20,000 cells are working well. The most fluorescent cells, usually 10^6 cells for the naïve library and 10^5 cells for subsequent rounds, are then directly sorted into 2 ml recovery medium, 2YT medium supplemented with 1–2% glucose and selection marker. If cell viability is low, the selection marker should be decreased in the recovery medium or even omitted.

To screen and select very large libraries, a "fluorescence-threshold selection" can be performed. For that purpose, the labeled cells are highly concentrated to allow flow rates of about 500,000 cells/s. Note that not every FACS machine technically supports this application.

Next, a threshold is set to the fluorescence signal, until only the most highest fluorescent cells are recorded by the machine and the apparent flow rate is reduced to 20,000–30,000 cells/s. Finally, the most fluorescent cells of this population are gated and selected. By this approach, very high cell numbers can be screened, but the sort is of low purity, as the true flow rate, that is cell density, is much higher. Under these conditions, many nonrelevant cells, which are close to a cell falling within the sorting gate, are co-sorted, because they fall below the fluorescence and are not detected by the FACS machine, so that the sorting mask (purity or yield) does not apply. The output can be directly resorted, if a sufficient number of cells are isolated or subjected to purity selection in the subsequent round.

Cells are recovered for 1 h at 28–37 °C, then diluted into 5–20 ml of 2YT medium supplemented with 1–2% glucose and selection marker and grown at 25–28 °C to $OD_{600} < 1$. Aliquots of $> 10^9$ cells are supplemented with 20% glycerol, snap-frozen, and stored at −80 °C until further use. Expression for subsequent selection rounds can directly be inoculated from a glycerol stock.

2.3. Characterization of selected GPCR variants

2.3.1 Sequence analysis of selected GPCR variants

First of all, the sequence diversity, mutational load and sequence distribution of mutations in the selected cell pool is analyzed by sequencing. Either the plasmid DNA of single colonies is isolated and the receptor sequence is

sequenced, or the receptor sequence is directly amplified from a single colony by colony PCR (cPCR) for subsequent sequencing of the PCR product. cPCR is much faster than plasmid DNA isolation and suitable to analyze large numbers of different variants, as it can be easily adapted to a 96-well format. For this purpose, a 20-μl PCR is setup containing 0.1 μ*M* of each flanking primer, 0.8 m*M* dNTP mix in PCR buffer with 2 m*M* $MgCl_2$ and 1–2 units of DNA polymerase. *Taq* DNA polymerase is sufficient for sequences up to 1000 bp, while a DNA polymerase with higher fidelity should be used to amplify longer sequences, for example $Vent_R^{TM}$ (NEB) or Phusion® (Finnzymes) DNA polymerase. A 10-μl pipette tip is used to gently pick a single colony from an agar plate and transfer it into one well of a 96-well PCR plate. During transfer of cells, any carryover of agar should be avoided, as it inhibits the cPCR. For very large colonies, a small sample from the colony boundary should be used. Cells are resuspended by repeated pipetting, and the pipette tip is then transferred to the corresponding well on a 96-deep-well plate containing 1 ml 2YT medium supplemented with 1% glucose (to suppress expression) and selection marker. The mini-cultures are then grown at 37 °C for 6–12 h in a shaking incubator and either stored at 4 °C for a few days, or as glycerol stocks at −80 °C. Receptor variants of interest can thus be easily regrown from a stock culture for further analysis or storage.

The cPCR includes a 10-min initial denaturation step which ensures cell disruption, while PCR amplification conditions have to be adjusted to primer sequence, product length, and DNA polymerase.

The PCR products are then purified using, for example, MultiScreen $PCR_{\mu 96}$ Filter Plate (Millipore) and subsequently sequenced.

2.3.2 Expression levels of individual selected GPCR variants

The final selected pool of cells is plated on agar plates to obtain single colonies, which can then be individually analyzed. Single colonies are grown and expressed in 24-well plates in 3–5 ml of 2YT medium with 0.2% glucose and selection marker each, closed with gas-permeable seals, and expression is performed as for the library expression.

Analytical flow cytometry can then be used to analyze the relative expression level, using the fluorescence-labeled receptor ligand. The assay is performed as described for the FACS, except that smaller expression volumes and number of cells for analysis can be used (10^6–10^7 cells for flow cytometry analysis).

For quantification of functional receptor expression, a radioligand-binding assay (RLBA) can be used, with the assumption that 1 OD_{600} in a cuvette of 1 cm pathlength equals 10^9 cells/ml. RLBAs are high-throughput compatible and hence suitable for the screening of large variant numbers. For rNTR1, a [^{3}H]-labeled neurotensin peptide is used (PerkinElmer). All steps are performed in 96-well plates. For one measurement, 2×10^7 cells are collected by centrifugation, washed once in ligand binding buffer (LBB, 50 m*M* Tris–HCl pH 7.4, 1 m*M* EDTA, 0.1% BSA and 40 μg/ml bacitracin), and resuspended in 100 μl LBB buffer. Hundred microliters of LBB containing 20 n*M* [^{3}H]-neurotensin are added to a final concentration of 10 n*M* and incubated for 2–3 h at 4 °C to allow for ligand saturation. Nonspecific binding is determined in the presence of 5 μ*M* unlabeled neurotensin peptide (Anaspec). Unbound and free [^{3}H]-neurotensin is separated from the cell-bound ligand by vacuum filtration using 96-well glass fiber filter plates (Millipore MultiScreen-FB plates MAFBN0B50, pretreated with 100 μl of 0.01% polyethylenimine (PEI)), on a 96-well vacuum filtration device (e.g., Millipore MultiScreen Vacuum Manifold). The sample volume is applied to a well of the filter plate, and filtrated by application of vacuum, and the filters are washed four to five times with 200 μl of LBB buffer. Filters are dried for 30–60 min at 60 °C, and the filter- and cell-bound radioactivity is then quantified by liquid scintillation. For this, filters are transferred to scintillation plates (IsoPlate 96; PerkinElmer) containing 200 μl of OptiPhase SuperMix scintillation cocktail (Perkin Elmer). Filters are allowed to dissolve for 3–12 h and quantified for 2 min in a Wallac 1450 Microbeta plus liquid scintillation counter.

2.3.3 Detergent stability of selected variants in the presence and absence of receptor ligand

To assess the detergent stability of the selected variants, we have previously reported a fast and efficient method to screen large numbers of variants, which is explained in detail elsewhere (Dodevski & Plückthun, 2011). Briefly, receptor variants are *in vivo* biotinylated using the AviTag sequence, detergent-solubilized, and immobilized on magnetic, streptavidin-coated beads (MyOne Streptavidin T1 beads, Invitrogen). Detergents can be efficiently exchanged after immobilization by repetitive washing in the detergent of choice (Fig. 4.4). According to our experiments, stability measurements in a particular detergent are not affected by the choice of detergent used for solubilization or the rebuffering process (Fig. 4.4).

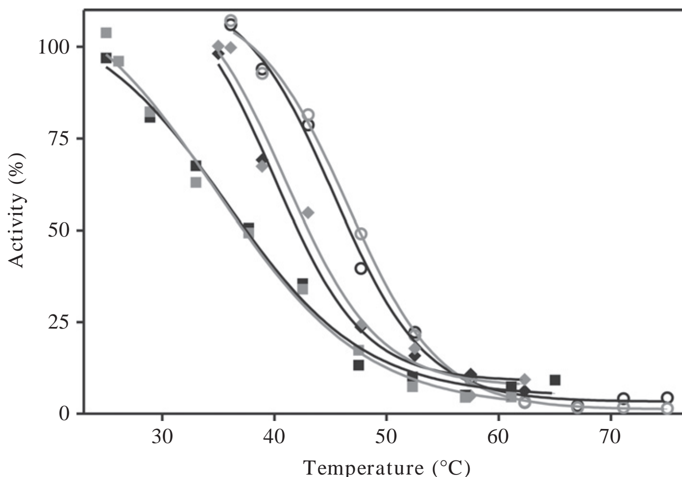


Figure 4.4 Influence of detergent exchange on thermostability measurements. GPCR variants are solubilized from *E. coli* membranes by DDM (*n*-dodecyl-β-D-maltopyranoside, black) or DM (*n*-decyl-β-D-maltopyranoside, gray) and immobilized on streptavidin-coated magnetic beads. Detergents are exchanged by repeated washing and pull-down of the beads in the final detergent buffer. Final detergents are DDM (open circles), UM (*n*-undecyl-β-D-maltopyranoside, diamonds), or DM (squares). Aliquots of solubilized and rebuffered GPCR are thermally challenged, and the remaining ligand binding affinity is analyzed by RLBA.

The solubilized and immobilized receptor is then thermally challenged, and the remaining receptor activity is quantified by RLBA. In this experimental setup, the apparent detergent stability in the absence of ligand is determined. The assay is easily adapted to study apparent detergent stability in the presence of ligand: For this purpose, the immobilized receptor is first saturated with [^{3}H]-neurotensin for 2 h, free ligand is washed away, and the receptor is then thermally challenged. Depending on the sample volume, the concentration of [^{3}H]-neurotensin has to be adjusted to allow ligand saturation under these conditions. LBB buffer containing 3–5 n*M* [^{3}H]-neurotensin is then added and incubated for 1 h before remaining receptor activity is quantified by liquid scintillation counting.

ACKNOWLEDGMENTS

We thank Drs. Christoph Klenk and Erik Sedlak for critical reading of the manuscript. Work in the author's laboratory was supported by the Swiss National Science Foundation through the NCCR Structural Biology.

REFERENCES

Aguinaldo, A. M., & Arnold, F. (2002). Staggered extension process (StEP) in vitro recombination. *Methods in Molecular Biology*, *192*, 235–239.

Bjarnadottir, T. K., Gloriam, D. E., Hellstrand, S. H., Kristiansson, H., Fredriksson, R., & Schiöth, H. B. (2006). Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics*, *88*, 263–273.

Bowie, J. U. (2001). Stabilizing membrane proteins. *Current Opinion in Structural Biology*, *11*, 397–402.

Cadwell, R. C., & Joyce, G. F. (1994). Mutagenic PCR. *PCR Methods and Applications*, *3*, S136–S140.

Chen, G., Hayhurst, A., Thomas, J. G., Harvey, B. R., Iverson, B. L., & Georgiou, G. (2001). Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nature Biotechnology*, *19*, 537–542.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., et al. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*, *318*, 1258–1265.

Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., et al. (2010). Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*, *330*, 1091–1095.

Chuang, S. E., Chen, A. L., & Chao, C. C. (1995). Growth of *E. coli* at low temperature dramatically increases the transformation frequency by electroporation. *Nucleic Acids Research*, *23*, 1641.

Clackson, T., & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, *267*, 383–386.

Deupi, X., & Kobilka, B. (2007). Activation of G protein-coupled receptors. *Advances in Protein Chemistry*, *74*, 137–166.

Dodevski, I., & Plückthun, A. (2011). Evolution of three human GPCRs for higher expression and stability. *Journal of Molecular Biology*, *408*, 599–615.

Dower, W. J., Miller, J. F., & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research*, *16*, 6127–6145.

Duquesne, K., & Sturgis, J. N. (2010). Membrane protein solubilization. *Methods in Molecular Biology*, *601*, 205–217.

Ferguson, S. S. (2001). Evolving concepts in G protein-coupled receptor endocytosis: The role in receptor desensitization and signaling. *Pharmacological Reviews*, *53*, 1–24.

Foord, S. M., Bonner, T. I., Neubig, R. R., Rosser, E. M., Pin, J. P., Davenport, A. P., et al. (2005). International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacological Reviews*, *57*, 279–288.

Granseth, E., Seppala, S., Rapp, M., Daley, D. O., & Von Heijne, G. (2007). Membrane protein structural biology—How far can the bugs take us? *Molecular Membrane Biology*, *24*, 329–332.

Grisshammer, R., Duckworth, R., & Henderson, R. (1993). Expression of a rat neurotensin receptor in *Escherichia coli*. *The Biochemical Journal*, *295*, 571–576.

Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., et al. (2008). The 2.6 Ångström crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science*, *322*, 1211–1217.

Jungnickel, B., Rapoport, T. A., & Hartmann, E. (1994). Protein translocation: Common themes from bacteria to man. *FEBS Letters*, *346*, 73–77.

Katritch, V., Cherezov, V., & Stevens, R. C. (2012). Diversity and modularity of G protein-coupled receptor structures. *Trends in Pharmacological Sciences*, *33*, 17–27.

Lagerström, M. C., & Schiöth, H. B. (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Reviews. Drug Discovery*, 7, 339–357.

le Maire, M., Champeil, P., & Møller, J. V. (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta*, *1508*, 86–111.
Martin, S., Botto, J. M., Vincent, J. P., & Mazella, J. (1999). Pivotal role of an aspartate residue in sodium sensitivity and coupling to G proteins of neurotensin receptors. *Molecular Pharmacology*, *55*, 210–215.
Miller, J. L., & Tate, C. G. (2011). Engineering an ultra-thermostable beta(1)-adrenoceptor. *Journal of Molecular Biology*, *413*, 628–638.
Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews. Drug Discovery*, *5*, 993–996.
Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., et al. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, *289*, 739–745.
Pin, J. P., Kniazeff, J., Goudet, C., Bessis, A. S., Liu, J., Galvez, T., et al. (2004). The activation mechanism of class-C G-protein coupled receptors. *Biology of the Cell*, *96*, 335–342.
Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., et al. (2011). Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature*, *469*, 175–180.
Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., et al. (2007). Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature*, *450*, 383–387.
Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., et al. (2011). Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*, *477*, 549–555.
Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., et al. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science*, *318*, 1266–1273.
Rosenbaum, D. M., Rasmussen, S. G., & Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature*, *459*, 356–363.
Sambrook, J. F., & Russel, D. W. (2001). (3rd ed). *Molecular cloning: A laboratory manual* (3 vols.). New York: Cold Spring Harbor Laboratory Press.
Sarkar, C. A., Dodevski, I., Kenig, M., Dudli, S., Mohr, A., Hermans, E., et al. (2008). Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 14808–14813.
Schlinkmann, K. M., Hillenbrand, M., Rittner, A., Künz, M., Strohner, R., & Plückthun, A. (2012). Maximizing detergent stability and functional expression of a GPCR by exhaustive recombination and evolution. *Journal of Molecular Biology*, *422*, 414–428.
Schlinkmann, K. M., Honegger, A., Türeci, E., Robison, K. E., Lipovsek, D., & Plückthun, A. (2012). Critical features for biosynthesis and functionality of a GPCR uncovered by all-versus-all mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 9810–9815.
Seddon, A. M., Curnow, P., & Booth, P. J. (2004). Membrane proteins, lipids and detergents: Not just a soap opera. *Biochimica et Biophysica Acta*, *1666*, 105–117.
Serrano-Vega, M. J., Magnani, F., Shibata, Y., & Tate, C. G. (2008). Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 877–882.
Shibata, Y., White, J. F., Serrano-Vega, M. J., Magnani, F., Aloia, A. L., Grisshammer, R., et al. (2009). Thermostabilization of the neurotensin receptor NTS1. *Journal of Molecular Biology*, *390*, 262–277.
Spee, J. H., de Vos, W. M., & Kuipers, O. P. (1993). Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. *Nucleic Acids Research*, *21*, 777–778.

Tucker, J., & Grisshammer, R. (1996). Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *The Biochemical Journal*, *317*, 891–899.

Vanhercke, T., Ampe, C., Tirry, L., & Denolf, P. (2005). Reducing mutational bias in random protein libraries. *Analytical Biochemistry*, *339*, 9–14.

Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., et al. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*, *454*, 486–491.

Wells, J. A. (1991). Systematic mutational analyses of protein-protein interfaces. *Methods in Enzymology*, *202*, 390–411.

Wilson, D. S., & Keefe, A. D. (2001). Random mutagenesis by PCR. *Current Protocols in Molecular Biology*, Chapter 8, Unit8.3.

Zhao, H., & Zha, W. (2006). In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature Protocols*, *1*, 1865–1871.