ISSN 0006-2979, Biochemistry (Moscow), 2010, Vol. 75, No. 7, pp. 881-891. © Pleiades Publishing, Ltd., 2010.
Original Russian Text © L. E. Petrovskaya, A. A. Shulga, O. V. Bocharova, Ya. S. Ermolyuk, E. A. Kryukova, V. V. Chupin, M. J. J. Blommers, A. S. Arseniev, M. P. Kirpichnikov, 2010, published in Biokhimiya, 2010, Vol. 75, No. 7, pp. 1001-1013.

Expression of G-Protein Coupled Receptors in *Escherichia coli* for Structural Studies

L. E. Petrovskaya[1]*, A. A. Shulga[1], O. V. Bocharova[1], Ya. S. Ermolyuk[1], E. A. Kryukova[1], V. V. Chupin[1], M. J. J. Blommers[2], A. S. Arseniev[1], and M. P. Kirpichnikov[1,3]

[1]*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (495) 330-6983; E-mail: lpetr65@yahoo.com*
[2]*Novartis Institutes for BioMedical Research, Basel, Switzerland*
[3]*Faculty of Biology, Lomonosov Moscow State University, 119991 Moscow, Russia*

Received December 12, 2009
Revision received February 22, 2010

Abstract—To elaborate a high-performance system for expression of genes of G-protein coupled receptors (GPCR), methods of direct and hybrid expression of 17 GPCR genes in *Escherichia coli* and selection of strains and bacteria cultivation conditions were investigated. It was established that expression of most of the target GPCR fused with the N-terminal fragment of OmpF or Mistic using media for autoinduction provides high output (up to 50 mg/liter).

DOI: 10.1134/S0006297910070102

Key words: autoinduction, expression in *E. coli*, GPCR, OmpF, Mistic

The G-protein coupled receptors (GPCR) form a major family of membrane proteins in human cells including about 1000 representatives. As biological sensors, they provide signal transduction and cell response to various extracellular stimuli, from inorganic ions to proteins [1, 2]. The common structural features of these proteins include seven hydrophobic transmembrane segments, extracellular N-terminus, and intracellular C-terminus. The binding sites of small organic molecules are often located within the transmembrane segments, while peptide hormones and proteins interact with the N-terminus of the GPCR and extracellular loops [3].

Proteins of the GPCR family are associated with various human diseases and are targets for many drugs. Development of new generation drugs possessing higher specificity and lacking side effects requires detailed study of mechanisms of ligand binding and signal transduction by these proteins, which can be ensured only by exact information about their spatial structure. Previously, molecular modeling of receptors of the GPCR family has been based on the spatial structure of bovine rhodopsin obtained by X-ray crystallographic analysis [4]. There were no data about other GPCR structures due to difficulties in obtaining protein preparations such as low concentration of the receptors in tissues, instability, and heterogeneity. There has been significant improvement in this field during the last two years. Construction of systems for heterologous expression and stabilization of GPCR molecules by protein engineering methods has enabled establishment of three-dimensional structure of three more family representatives: β1- and β2-adrenergic and adenosine receptors [5-8]. However, for crystallization of each of these receptors, specific methods including individual selection of the expression system and modification of the protein molecule were used. Obviously, deciphering of the spatial structure of most GPCR molecules by X-ray methods or NMR spectroscopy requires development of a more universal approach applicable for, with small variations, most of these receptors.

The most important part of this approach is selection of the optimal system for expression of the GPCR gene. Mammalian and insect cells provide the most suitable medium for synthesis and posttranslational processing of eukaryotic receptors, but their cultivation is quite expensive [9]. In yeast cells, hyperglycosylation of recombinant proteins is often observed, and folding of the GPCR is often incorrect [10]. Expression of foreign genes in bacteria, first of all, in *Escherichia coli*, is not followed by any posttranslational modifications of the recombinant pro-

Abbreviations: GPCR, G-protein coupled receptors; GST, glutathione-S-transferase; IPTG, isopropyl-β-D-thiogalactoside.

* To whom correspondence should be addressed.

teins and can provide enough material for study of the structure by NMR spectroscopy and X-ray analysis [11]. Cultivation of bacteria does not require costly media and the equipment and is easily scaled. Thus, *E. coli* cells provide a promising system for heterologous expression of GPCR genes.

The goal of this work was to elaborate a high-performance system for expression of human GPCR genes in *E. coli* cells. Various methods of direct and hybrid expression of 17 GPCR genes and strains and conditions of cultivation of bacteria were comparatively analyzed. Based on the data obtained, an effective approach for expression of GPCR genes in *E. coli* cells is suggested for studies of GPCR structure.

MATERIALS AND METHODS

Reagents from BioRad (USA), Sigma (USA), Merck (USA), and Panreac (Spain) and bacteria cultivation media components from Difco (USA) were used in this study. Solutions were prepared using MilliQ deionized water.

Construction of GPCR expression plasmids. DNA manipulations were performed using standard methods [12] in *E. coli* cells XL-1 Blue (Stratagene, USA) using Fermentas enzymes (Lithuania). The oligonucleotides were synthesized by Evrogen (Russia). The plasmids and fragments for cloning were purified using Qiagen kits (Germany).

cDNA from the Novartis collection (Switzerland) were amplified by PCR using gene-specific primers containing *Nde*I site at the 5′-end and *Xho*I site at the 3′-end (Table 1) and *Pfu* DNA polymerase. The following PCR conditions were used: 3 min template denaturation at 95°C; 25 cycles (denaturation at 95°C for 45 sec, primer annealing at 52-55°C for 45 sec, elongation at 72°C for 1 min); final elongation at 72°C for 5 min. The resulting fragments were treated with *Nde*I and *Xho*I, purified by electrophoresis in 1% agarose gel, and cloned into the pET32a vector obtained by treatment with the same restrictases.

The insert sequence of the resulting plasmids was confirmed by sequencing in both directions (Center of Multiple Access "Genome", Russia) using T7prom and T7term primers as well as gene-specific primers (Table 1). Then the plasmid DNA was treated with *Nde*I and *Xho*I and the gene-containing fragment was purified and cloned into the expression vectors treated with *Nde*I and *Xho*I.

To construct the pET28F plasmid, the DNA fragment encoding signal sequence and the first 12 amino acid residues (a.a.) of mature OmpF was amplified on a genomic DNA template of *E. coli* using PCR with primers OmpF-Nco (5′-ACATACCATGGTGAAGCGCAATATTC) and OmpF-Nde (5′-ACATACATATGATCTACTTTGTTGCCATC), treated with *Nco*I and *Nde*I restrictases, and cloned at the corresponding sites of the pET28a vector (Novagen, USA).

The pET28DA plasmid encoding DsbA signal sequence was constructed by cloning following the same scheme of synthetic DNA fragment obtained as a result of annealing of oligonucleotides DA1 (5′-CATGAAAAAGATTTGGCTGGCGCTGGCT), DA2 (5′-GGTTTAGTGTTAGCGTTTAGCGCTTCGGCGGCGCA), DA3 (5′-CTAACACTAAACCAGCCAGCGCCAGCCAAATCTTTTT), DA4 (5′-TATGCGCCGCCGAAGCGCTAAACG).

The pMT32 plasmid was obtained by cloning of the *Mistic* gene of *B. subtilis* amplified on a template of the plasmid DNA of pET-Mistic (provided by O. V. Nekrasova) using primers T7prom and MisRev (5′-GGAATTCCATATGAGATCCACGCGGAACCAGACCAGAAC) and treated with *Xba*I and *Nde*I restrictases at corresponding sites of the pET32 plasmid. The pMT32H10 plasmid was obtained by ligation of the *Xho*I-*Ade*I fragment of the pETH10 plasmid (provided by L. N. Shingarova) at corresponding sites of the pMT32 plasmid.

The pOmpF and pGST plasmids were constructed by replacement of the *Mistic* gene in pMT32 by the *OmpF* and *GST* genes, respectively. With this goal, the genes were amplified using *E. coli* genomic DNA and the pGEX2T plasmid (GE Healthcare, USA), respectively, as templates for PCR using primers OmpF-Nco and OmpF-rev (5′-ACATACATATGGAACTGGTAAACGATAC) for obtaining pOmpF and GST-Nco (5′-ACATACCATGGGTTCCCCTATACTAGGTTATTG) and GST-Nde (5′-ACATACATATGATCCGATTTTGGAGGAT) for obtaining pGST, treated with *Nco*I and *Nde*I restrictases, and cloned at the corresponding sites of pMT32. The structures of all of the resulting plasmids were confirmed by sequencing within the insert.

Screening of expression by dot-blot analysis. To examine the expression, *E. coli* strains Rosetta2(DE3)pLysS (Novagen) and C41(DE3)R2 obtained by transformation of the C41(DE3) strain (Avidis SA, USA) using the pRARE2 plasmid (Novagen) were used. Samples (0.2 ml) of the autoinduction medium [13] containing 100 μg/ml ampicillin (or 50 μg/ml kanamycin in case of pET28a-based plasmids) and 34 μg/ml chloramphenicol in 2 ml test tubes were inoculated with separate colonies obtained from the surface of the dish containing L-agar. The test tubes were incubated for 24 h at 37°C and 72 h at 18°C in a thermomixer (Thermomixer Comfort; Eppendorf, Germany). The culture was precipitated by centrifugation in an Eppendorf 5415R table-top centrifuge for 5 min at 7000 rpm, and the precipitate was resuspended in 50 μl of CelLyticB reagent (Sigma) with addition of 5 μg/ml DNase I solution (Sigma). After incubation at room temperature for 15 min, 50 μl of 2XTES buffer containing 20 mM Tris-HCl, pH 6.8, 2 mM EDTA, 2% SDS was

Table 1. GPCR genes used in this study

No.	Short name	Full protein name	Number of amino acid residues	Sequences of the primers used for cloning (5′-3′)
1	*GPR4*	G-protein-coupled receptor, 4	362	ACATACCATATGGGTAACCACACGTGG ACATCTCGAGTTGTGCTGGCGGCAGC
2	*CCR4*	chemokine receptor 4 (C-C motive)	361	ACATACCATATGAACCCCACGGATATAGC ACATCTCGAGCGGGGGCCCATTTGCC
3	*CNR2*	cannabinoid receptor 2 (from macrophages)	360	ACATACCATATGGAGGAATGCTGGGTG ACATCTCGAGGCAATCAGAGAGGTCTAG
4	*CCR1*	chemokine receptor 1 (C-C motive)	356	ACATACCATATGGAAACTCCAAACACC ACATCTCGAGGAACCCAGCAGAGAGTTC
5	*CX3CR1*	chemokine receptor 1 (C-X3-C motive)	355	ACATACCATATGGATCAGTTCCCTG ACATCTCGAGGAGAAGGAGCAATGC
6	*β2AR*	β2-adrenergic receptor	413	ACATACATATGGGTCAACCAGGGAACGG ACATCTCGAGCAGCAGTGAGTCATTTGTAC
7	*DRD3*	dopamine receptor 3	401	ACATACATATGGCATCTCTGAGCCAG ACATCTCGAGGCAAGACAGGATCTTGAGG
8	*EDG1*	sphingolipid receptor 1	381	AATTCCATATGGGTCCAACCAGCGTCCCGC CCGCTCGAGGGAAGAAGAGTTGACGTTTCC
9	*EDG3*	sphingolipid receptor 3	378	ACATACCATATGGCAACTGCCCTCCC ACATCTCGAGGTTGCAGAAGATCCCATTC
10	*CCR2b*	chemokine receptor 2b (C-C motive)	375	ACATACATATGCTGTCCACATCTCGTTC ACATCTCGAGTAAACCAGCCGAGACTTCC
11	*GPR68*	G-protein coupled receptor, 68	365	ACATACATATGCGTAGTGTGGCCCCTTC ACATCTCGAGGGCCAACCTGCCCGTGG
12	*SSTR5*	somatostatin receptor 5	364	ACATACCATATGGAGCCACTGTTCCC ACATCTCGAGCAGCTTGCTGGTCTGC
13	*GALR1*	galanin receptor 1	349	ACATACCATATGGAGCTGGCGGTCG ACATCTCGAGCACATGAGTACAATTGGTTG
14	*GPR174*	G-protein coupled receptor, 174	333	ACATACATATGCCTGCTAATTACACGTG ACATCTCGAGGCATAATTCAGGTGTCATGG
15	*GPR1*	G-protein coupled receptor, 1	333	AATTCCATATGACGCCAAACAGCACTGG CCGCTCGAGGTTCAAGTCCAGGTCGACAC
16	*MC2R*	melanocortin receptor 2	297	AATTCCATATGAAGCACATTATCAACTCG CCGCTCGAGCCAGTACCTGCTGCAGAAG
17	*CXCR4*	chemokine receptor 4 (C-X-C motive)	352	ACATACCATATGGAGGGGATCAGTATATAC ACATCTCGAGGTTTCTTTAGTTTTTGTG

Note: Cloning sites in primer sequences are underlined. GPCR gene sequences are in bold type.

added to the suspension, then incubation was continued for 10 min at 37°C and the mixture was centrifuged for 10 min at 13,000 rpm. A 2-μl sample of the supernatant was applied to a nitrocellulose membrane (BioRad) together with serial dilutions of purified GPCR with known concentrations as a standard. The protein was visualized using monoclonal antibodies to the His-tag (Novagen) and secondary antibodies conjugated with alkaline phosphatase (Sigma) according to the manufacturer's recommendations. For fractionation of the cell proteins, the suspension preliminarily incubated with CelLyticB was centrifuged and the supernatant (soluble fraction) was taken. A 50-μl volume of 2XTES buffer was added to the precipitate, and then the sample was treated

as described above (the fraction of inclusion bodies). Aliquots of both fractions (2 μl) were separately applied to the dot-blot.

Cultivation of producer strains. A 200-ml sample of 2ZYM5052 medium containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) in a 1-liter flask was inoculated with about 100 colonies of the strain transformed by one of the expression plasmid from the surface of a fresh dish with L-agar. To obtain the inclusion bodies, the culture was incubated at 37°C and 260 rpm for 24 h or, to obtain the membrane fraction, at 25°C for 72 h. The cells were precipitated by centrifugation at 8000 rpm for 15 min.

Isolation of membrane fraction. The cell precipitate obtained as described earlier was resuspended in sonication buffer (50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 20% sucrose), lysozyme was added up to 0.2 mg/ml, and the sample was incubated at room temperature for 20 min. Then the suspension was 5-fold diluted with cooled deionized water containing 1 mM phenylmethylsulfonyl fluoride (PMSF), kept on ice for 5 min, and sonicated (Branson Sonifier 450; Branson, USA) for 3 min. The suspension was centrifuged for 30 min at 6000 rpm, and the supernatant was repeatedly centrifuged for 1 h at 45,000 rpm using a 70Ti rotor (Beckman, USA). The resulting membrane fraction precipitate was resuspended in 50 mM Tris-HCl, pH 8.0. Total concentration of the protein was estimated using Protein Assay dye (BioRad).

For Western-blot visualization, the membrane fraction proteins separated by gel electrophoresis according to Laemmli [14] (50 μg per lane) or purified GPCR (15 μg per lane) were applied onto nitrocellulose membrane (BioRad) by the method described earlier [15]. The membrane was stained as described above for dot-blot.

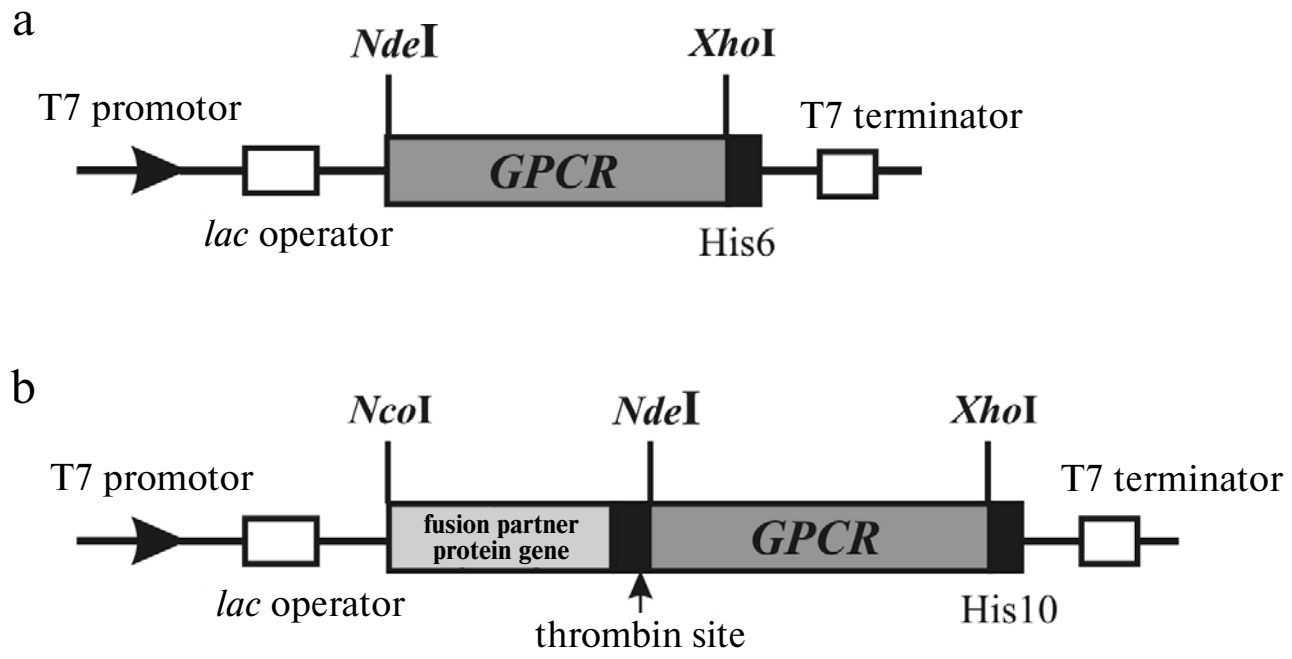
Purification of recombinant GPCR. The cell precipitate obtained as described above was resuspended in 20 mM Tris-HCl buffer, pH 8.0, 1 mM EDTA, 200 mM NaCl containing lysozyme (0.2 mg/ml) and then ultrasonicated. After centrifugation, the precipitate was washed twice with buffer containing 20 mM Tris-HCl, pH 8.0, 2 M urea, 1 mM EDTA, 10 mM β-mercaptoethanol. Then it was washed twice with buffer containing 0.5% Triton X-100 instead of urea and diluted in 20 ml of buffer A (20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 1.5% lauroyl sarcosine, 10 mM β-mercaptoethanol, and 10 mM imidazole) and incubated for 16 h at room temperature. The solution was centrifuged, the supernatant was applied to a 4-ml Ni^{2+}-charged column with Chelating-Sepharose CL-6B (GE Healthcare) equilibrated in the same buffer, washed with 40 ml of buffer B (20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1% lauroyl sarcosine, 10 mM β-mercaptoethanol, 50 mM imidazole) and eluted with buffer B containing 0.5 M NaCl and 300 mM imidazole. The fractions were analyzed by protein electrophoresis according to Laemmli in 12% polyacrylamide gel.

The concentration of the purified protein was estimated using individual absorption coefficients based on the results of 280-nm absorption measurement. In the case of the Mistic-GPCR fusion protein, the solution was two-fold diluted with buffer containing 20 mM Tris-HCl, pH 8.0, and thrombin solution (1 unit/mg protein; Technology-Standard, Russia) was added. The mixture was incubated for 4 h at room temperature, and the completeness of cleavage was analyzed by protein electrophoresis. One millimolar PMSF was added to the solution, and imidazole and salt concentrations were brought to 0.5 M and 30 mM, respectively. Then the protein was applied to a 2-ml Ni-Sepharose Fast Flow column (GE Healthcare) and washed with 10 ml of buffer B and 10 ml of the same buffer containing 10 mM NaCl. The protein was eluted with buffer B containing 0.5 M NaCl and 300 mM imidazole. Fractions containing protein as visualized by electrophoresis were combined, and then concentration was estimated as described above.

RESULTS

The literature contains data about the expression of GPCR genes in *E. coli* cells using various systems including direct [16] and hybrid expression, where glutathione-S-transferase [17], ketosteroid isomerase [18], and other proteins play the role of partner in a fusion protein. In a number of studies GPCR has been expressed as a protein fused with maltose-binding protein on the N-terminus and with thioredoxin on the C-terminus [19]. In this case, the fusion proteins are mainly found in the membrane fraction of the bacterial cells, and the yield is not more than 1 mg per liter of the culture [20]. Significantly more efficient synthesis of recombinant GPCR (dozens of milligrams per liter of culture) necessary for structural studies can be provided by formation of inclusion bodies in bacterial cells. The disadvantage of this approach for expression is the necessity to dissolve insoluble aggregates using denaturing agents and subsequently refolding the proteins obtained. Nevertheless, in a number of studies examples of successful use of this approach and restoration of native activity of the GPCR after renaturation have been described [21, 22]. At the first stage of the work, it was proposed to compare the potentials of both of the approaches to provide high level synthesis of the target proteins as well as efficiency of various partner proteins for hybrid expression.

To solve the problem, a collection of 17 cDNAs of GPCR was used (Table 1). To obtain plasmids for direct expression, the genes were amplified by PCR using gene-specific primers and cloned into the pET32a vector by *Nde*I-*Xho*I sites (Scheme, part (a)). Full-size GST, OmpF, Mistic, and DsbA and OmpF signal sequences were selected as fusion partners for hybrid expression. To provide an opportunity to simultaneously obtain various



Structural scheme of GPCR genes with adjacent regulatory elements within plasmids for direct (a) and hybrid (b) expression

expression constructs, the fusion partner protein genes were cloned into *Nco*I-*Nde*I cleavage sites so that they formed a single reading frame containing GPCR genes cloned by *Nde*I-*Xho*I sites (Scheme, part (b)). The pET28a plasmid was used as a vector for expression of hybrid proteins fused with DsbA and N-terminal fragment of OmpF or, for expression of fusion proteins Mistic-GPCR, GST-GPCR, and OmpF-GPCR, a plasmid based on pET32a containing *Nco*I site was used. Thus, all of the resulting genes were placed under the control of tightly regulated promotor T7*lac* and contained the sequence encoding six or ten histidine residues (His-tag) at the 3′-end for the subsequent purification of the fusion proteins by metal-affinity chromatography providing the standard conditions of expression and purification of the recombinant proteins in the ranges of the elaborated approach.

To study the expression, two *E. coli* strains were selected: Rosetta2(DE3)pLysS and C41(DE3)R2. The cultures were incubated in LB medium at 37 and 18°C. It was supposed that cultivation at the lower temperature would enable insertion of the recombinant receptors into the membrane of the bacterial cells [23]. At the first stage, the expression of the genes in the presence of 0.05-1 mM isopropyl-β-D-thiogalactoside (IPTG) was screened. It should be noted that a large number of membrane proteins, especially GPCR, display abnormal mobility during gel electrophoresis [24]. In addition, the corresponding bands on the gel are diffuse [17]. Therefore, to improve the screening efficiency the level of synthesis of the recombinant GPCR in culture was estimated by intensity of staining of dots on dot-blots using antibodies to the His-tag.

To simultaneously analyze the maximal number of variants, the cultures were incubated in 200 μl in 2 ml test tubes. It was first established that, on culturing in a small volume of the medium, the level of GPCR synthesis correlates with the amount of protein that can be obtained on scaling up the experiment (data not presented). The results of the dot-blot showed that IPTG induction enables direct expression of only two of 17 GPCR genes in both strains (GPR4 and GPR68, Table 2). Western-blot analysis of lysates of the cells expressing GPR68 gene with antibodies to the His-tag confirmed that the target recombinant protein is synthesized (data not presented).

At the next stage of the study, cultures were incubated in 2ZYM5052 autoinduction medium elaborated by Studier [13] at 37 and 18°C. Use of this medium enabled to reduce the number of analyzed parameters, especially the culture absorption at the time of induction and the concentration of the inducer, and, respectively, to improve the efficiency of screening. At the same time, the number of genes, the direct expression of which was detected by dot-blot, increased from two to six (Fig. 1a and Table 2). During study of hybrid expression of the GPCR genes with various partner proteins in autoinduction medium, expression of 13 more genes were detected (Table 2). Thus, during expression of 17 genes from the collection, negative result was obtained only in the case of DRD3. This suggests that the elaborated approach for expression of GPCR genes in bacterial cells is universal.

Table 2. Expression level of GPCR genes

No.	Direct expression	Hybrid expression with OmpF*	Hybrid expression with Mistic*	Yield of the purified protein, mg/liter**
1	+++	++	+++	50/4
2	++*	n.d.	+	10/10
3	–	+	+	n.d.
4	–	+	+++	n.d.
5	–	+	+++	n.d.
6	–	++	+++	20/4
7	–	–	–	n.d.
8	–	+	++	n.d.
9	–	+	++	n.d.
10	–	++	+++	20/4
11	+++	++	+++	50/10
12	+*	++	+++	50/10
13	–	–	+	n.d.
14	–	n.d.	++	n.d.
15	+*	+	++	n.d.
16	+*	++	+++	50/10
17	–	+	+	n.d.

Note: (–), <5 mg/liter; (+), 5-10 mg/liter; (++), 10-20 mg/liter; (+++), 20-50 mg/liter; n.d., not determined.

* Expression was observed only after culturing on autoinduction medium. Screening data confirmed by preparative purification of the recombinant proteins are designated by gray shading.

** Yield obtained using the most effective expression system and culturing the strain on rich/minimal autoinduction medium is presented. In case of similar yield using direct and hybrid expression, direct expression was used.

Notably, the proteins and the protein fragments used as fusion partners for hybrid expression of the GPCR genes have shown different efficiency. The use of full-sized OmpF as well as DsbA signal sequence was fully inefficient. Expression of GST hybrids promoted an increase in synthesis of seven GPCR, but high expression level was observed only in the case of GPR4. Expression of hybrid proteins containing the signal sequence and the N-terminal fragment of OmpF (12 a.a.) was more successful. We detected the expression of 13 of 15 hybrid OmpF-GPCR genes, and the expression level of six of them was more than 10 mg per liter of the culture (Table 2). However, full-sized Mistic was the most effective partner protein. Expression of 16 of 17 of the studied genes was detected by dot-blot analysis, and the expression level of 11 genes cultured in rich autoinduction medium was ≥10 mg protein per liter of culture (Fig. 1b and Table 2). Higher synthesis level (>20 mg protein per liter of culture) was detected in samples obtained by cultivation at 37°C. In most cases, use of the Rosetta2(DE3)pLysS and C41(DE3)R2 strains was accompanied by nearly the same level of synthesis of recombinant GPCR, but the yield of the hybrid proteins containing CCR1, CX3CR1, and some other target proteins was higher in the C41(DE3)R2 strain.

Fractionation of the proteins in culture samples expressing fusion proteins Mistic-GPCR at 37°C revealed that the expression products are mainly located in inclusion bodies (Fig. 2a). However, in a number of samples cultivated at 18-25°C, dot-blot analysis revealed that expression of the target proteins in soluble fraction was quite high (Fig. 2a). To specify the location of the Mistic-GPCR fusion proteins, the membrane fraction of the cells expressing β2AR fused with Mistic was isolated. Using Western-blot analysis with antibodies to the His-tag, the presence of the fusion protein was confirmed in

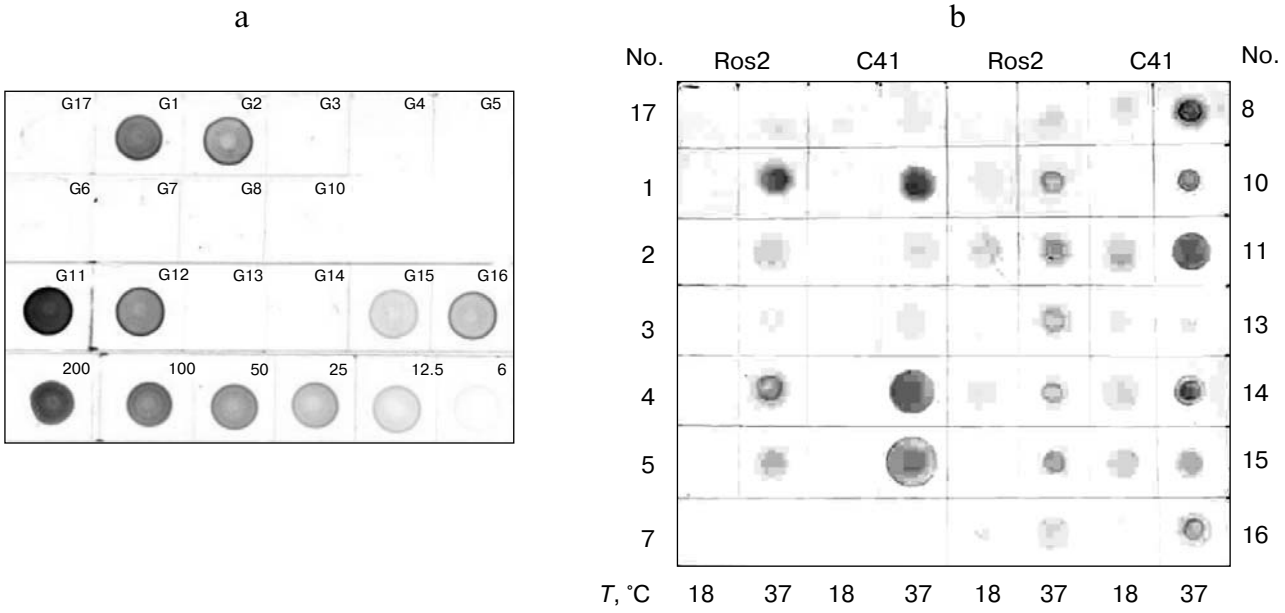


Fig. 1. Expression of GPCR in *E. coli* cells. Screening of direct (a) and hybrid (with Mistic as partner (b)) expression of GPCR genes by the dot-blot method using antibodies to the His-tag. Cells of Rosetta2(DE3)pLysS (a, b) and C41(DE3)R2 (b) strains containing corresponding plasmids were incubated at 18°C (a, b) and 37°C (b). Amount of protein (ng) added as a standard is designated by numbers.

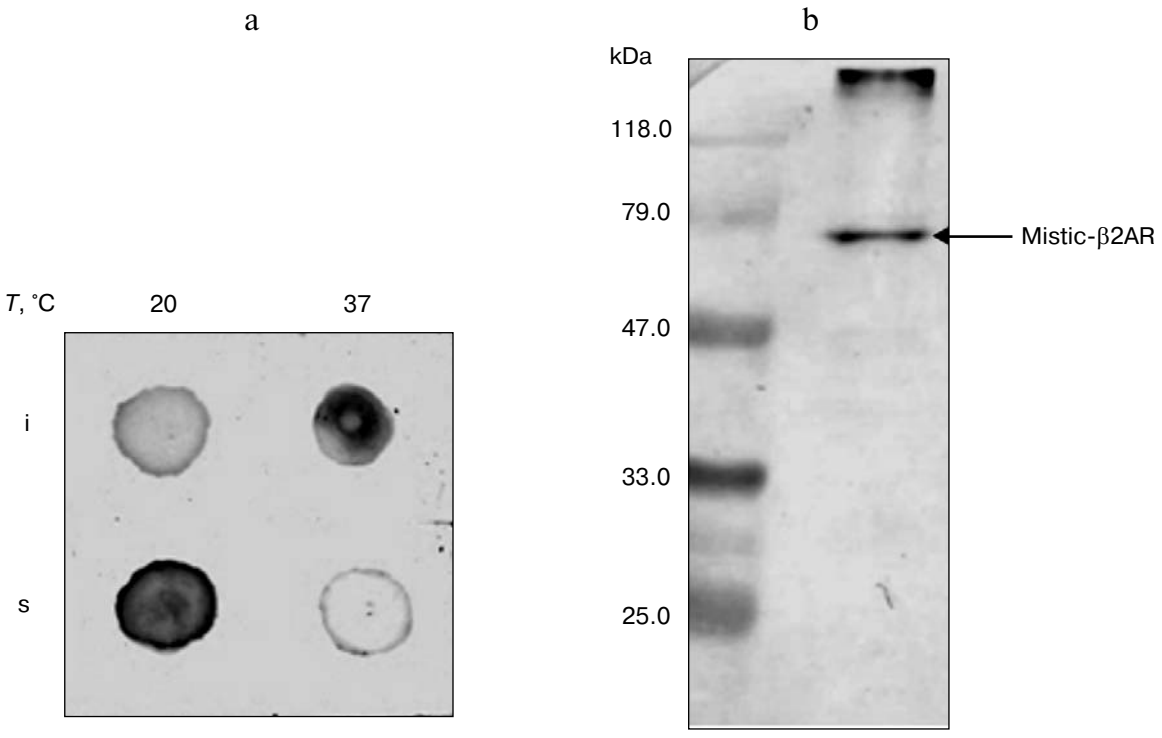


Fig. 2. Analysis of recombinant b2AR localization in *E. coli* cells. a) Dot-blot analysis of soluble (s) and insoluble (i) fractions of cell proteins of the Rosetta2(DE3)pLysS strain containing the pMT6H10 plasmid cultured in 2ZYM5052 medium at 20 and 37°C. b) Western-blot analysis of membrane fraction of cells expressing b2AR with antibodies to the His-tag. The lane on the left is molecular mass markers (Fermentas).

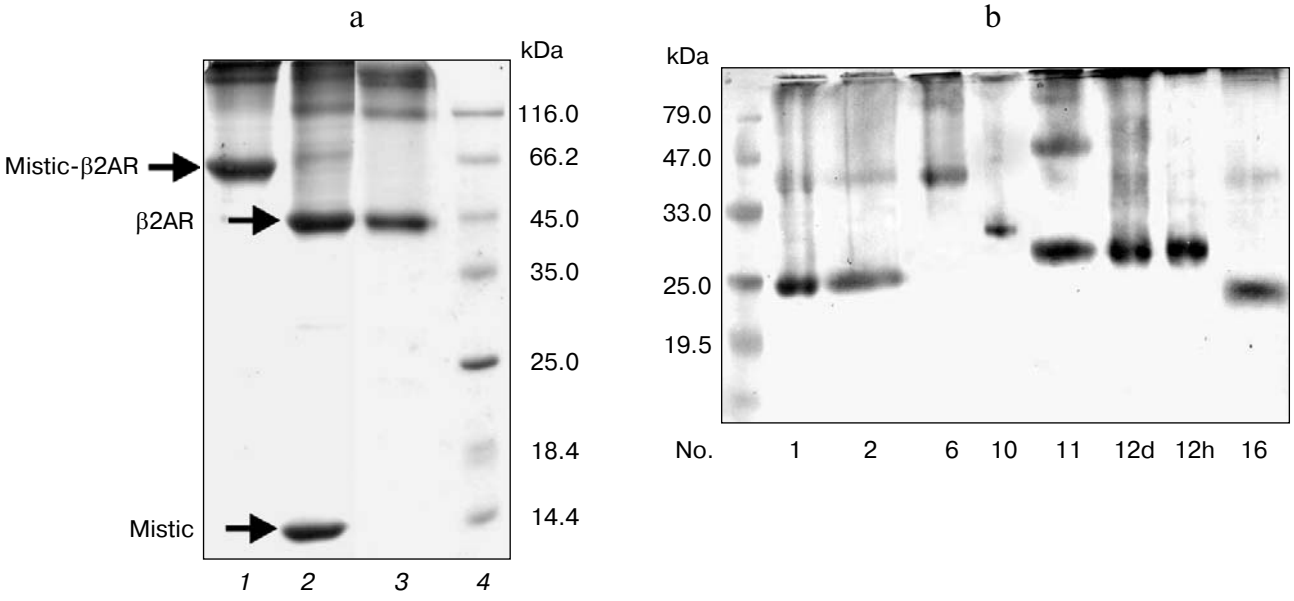


Fig. 3. Purification of recombinant GPCR. a) SDS-PAGE in 12% polyacrylamide gel of purified Mistic-b2AR fusion protein (*1*), the same protein after incubation with thrombin for 4 h (*2*), and purified b2AR (*3*). Positions of the Mistic-b2AR, b2AR, and Mistic monomers are designated by arrows. *4*) Molecular mass markers (Fermentas). b) Western-blot analysis of purified GPCR with antibodies to the His-tag. Lane on the left, molecular mass markers (Fermentas). Numeration of the GPCR is presented according to the Table 1. 12d and 12h, SSTR5 obtained by direct and hybrid (with Mistic) expression systems, correspondingly.

the inner membrane of the bacteria (Fig. 2b). However, along with the monomer, there was a significant amount of aggregated protein in the sample. Similar results were obtained during the study of localization of the Mistic-SSTR5 fusion proteins in cells (data not presented).

To confirm the screening results, eight recombinant GPCR from the fractions of insoluble cellular proteins were purified using metal-affinity chromatography in buffer containing lauroyl sarcosine and mercaptoethanol (Table 2). Due to thorough washing of the inclusion bodies, after one stage of chromatographic separation the preparation of the target protein is obtained with purity >90% (Fig. 3a, lane *1*). After analysis of the chromatography results, it was found that binding of the fusion proteins containing hexahistidine tag to the nickel-affinity resin is not strong enough. Elution of the target protein bound to the sorbent was observed already after washing with buffer containing 30 mM imidazole. To increase the sorbent-binding strength, the length of the histidine tag was increased to 10 residues, with increase in the concentration of imidazole in the washing buffer to 70 mM.

In the next stage, the purified Mistic-GPCR fusion proteins were treated with thrombin for separation of Mistic and the target protein. Cleavage of the fusion proteins appeared to be complete after 4 h of incubation at room temperature (Fig. 3a, lane *2*). Then the target proteins were separated from the partner proteins by repeated Ni-Sepharose chromatography (Fig. 3a, lane *3*). Using the standard purification method, we obtained >20 mg of recombinant GPR68, b2AR, CCR2b, SSTR5, MC2R, and other GPCR from 1 liter of rich medium (Fig. 3b). Analysis of the N-terminal sequence of three purified proteins (b2AR, SSTR5, and MC2R) confirmed correctness of cleavage of Mistic by thrombin (data not presented).

To study the spatial structure of proteins by NMR spectroscopy, it is necessary to obtain proteins containing isotope labels. With this goal, bacterial strains must be cultivated on minimal media containing the corresponding additives. Efficiency of expression of several hybrid Mistic-GPCR genes (6, 10, 12, 16) was studied by incubating the cultures on minimal autoinduction medium (M5052 [13]) at 37°C. It was established that, under the given conditions, higher level of synthesis of the recombinant GPCR is observed in the Rosetta2(DE3)pLysS strain. Therefore, it was selected as the basic strain for the elaborated expression system. The strains were cultivated on the minimal medium for 3-5 days, and the protein yields were several times lower than that when using the rich medium (Table 2). Nevertheless, even when cultivating the strains on the minimal medium, the suggested approach provided recombinant GPCR yield, which represents sufficient amount of material for structural studies.

DISCUSSION

Construction of systems for expression of human membrane receptors in bacterial cells is an extremely

important problem for fundamental research and molecular medicine. To solve the problem, accurate analysis of effects of a great number of parameters to find a compromise between level of synthesis of the recombinant protein and cell viability is necessary, since heterologous expression of membrane proteins is often toxic for bacteria [25]. The present work was oriented to obtain material for study of spatial structure of the proteins by NMR spectroscopy. This fact determined the presence of additional requirements on the constructed expression systems, such as high protein yield and possibility to obtain isotope-labeled preparations. Another feature was the attempt to develop a universal approach for efficient expression of large number of GPCR. The cDNA library used in the present work included 17 genes of various GPCR, and the data about the possibility of expression in bacteria were available for only two of them (β2AR and CNR2 [26, 27]).

The approach is based on the standard scheme of cloning of encoding sequences of the GPCR and partner proteins for hybrid expression at the same restriction sites into the plasmids based on the same vector. This fact enabled rapid construction of plasmids for expression of each of the GPCR genes as fusion proteins and comparative analysis of their efficiency. The expression level was estimated by dot-blot assay, which enabled simultaneously study of a great number of the samples using a standard reagent kit for all of the studied proteins.

One of the common expression problems is the presence of rare codons in the target gene. This can be a reason for low expression level of the gene in bacteria [28]. In a number of works, this problem has been solved by construction of synthetic GPCR genes taking into account usage of the codons in bacterial cells [21, 29]. However, use of this approach for expression of a great number of genes is quite expensive. Therefore, we have employed a more universal strategy – expression of the target genes in strains containing plasmid pRARE2, which provides synthesis of an extra amount of rare tRNA. It is known that *E. coli* strains C41 and C43 provide the increase in the synthesis level of recombinant membrane proteins and decrease their toxic effect on cells [30]. In several cases, we did observed an increase in expression of the GPCR genes using strain C41(DE3)R2. However, expression level of majority of genes in Rosetta2(DE3)pLysS and C41(DE3)R2 cells appeared to be comparable. The final choice of the strain was made by comparing yields of the target proteins after incubating the cultures on the minimal medium. Under these conditions, the expression level of recombinant GPCR was significantly higher in Rosetta2(DE3)pLysS cells.

Use of autoinduction medium played a significant role in the development of the universal approach for expression of GPCR genes in bacteria. Rational combination of three carbon sources (glucose, glycerin, and lactose) as well as additives promoting more intensive growth of the culture enables tighter control after the induction process and production of a great amount of biomass [13]. At the same time, it is not necessary to select individual parameters of cultivation (stages of culture growth, amount of inducer), which complicates standardization of traditional expression experiments by IPTG induction. Comparison of the results obtained by IPTG induction and using the autoinduction medium proved higher efficiency and ease of the second method for solution of the problem. Notably, using homemade autoinduction medium and OvernightExpress medium (Novagen), there was no significant difference in the amount of biomass obtained and the yield of the target proteins (data not presented).

The easiest method for heterologous expression of GPCR genes in bacteria is direct expression of the target protein gene under control of regulatory elements provided by the plasmid vector. Such systems have been obtained for a number of membrane proteins including GPCR [16]. The possibility of direct expression of a certain gene depends on various factors, especially on the secondary structure of the 5′-end region of the mRNA, which determines its translation efficiency [31]. It should be noted that the 5′-end regions of most genes from our collection had high content of G and C nucleotides that could promote stabilization of the hairpin structures of the corresponding mRNA and decrease protein synthesis level in direct expression. On screening the cultures incubated in the presence of IPTG, direct expression of only two GPCR of 17 (GPR4 and GPR68) was detected and, using autoinduction medium, the synthesis of CCR4, SSTR5, GPR1, and MC2R was additionally detected.

In a number of cases the level of synthesis of the recombinant proteins was increased by use of hybrid expression systems in which translation is initiated in the 5′-end region of the mRNA encoding a partner protein with high expression level. In addition, expression of the proteins fused with several partners promotes increase in solubility and protection of the target protein from degradation by bacterial proteases, which can be a reason for decreased yield. Thus, a hybrid expression system based on GST is widely used for increase in synthesis level and solubility of recombinant proteins in bacterial cells [32]. There are data about excretion of fusion proteins containing full-sized OmpF into the medium of *E. coli* culture [33]. Conjunction of encoding sequences of the signal peptides and N-terminal fragments of excreted bacterial proteins to the genes of the target proteins can enable their incorporation into inner membrane of the cells [34, 35]. Finally, it was established recently that expression of genes of a number of membrane proteins (including GPCR) in *E. coli* is facilitated by construction of proteins fused with the Mistic integral protein of *B. subtilis* membrane [36].

Our research has confirmed the efficiency of use of hybrid expression systems for obtaining recombinant

GPCR in *E. coli* cells. The total number of the GPCR genes with expression level more than 5 mg/liter of medium increased from 6, when expressed directly, to 16. Use of the N-terminal fragment (signal sequence and first 12 a.a.) of OmpF and full-sized Mistic appeared to be the most effective. According to our data, use of Mistic as a partner protein promotes higher synthesis level in *E. coli* of most of the GPCR. This must be due to the presence of different mechanisms of binding of the proteins with the membrane. OmpF is transported through the plasma membrane by a complex of bacterial proteins (translocon), and this process is limited by its "capacity" [37]. Mistic does not possess a typical signal sequence and integrates into the membrane without using the excretion apparatus of bacteria. So, its superproduction as well as expression of fusion proteins with Mistic does not cause toxicity for cells [36].

Most of the Mistic-GPCR fusion proteins are localized in a fraction of insoluble cell proteins (inclusion bodies) and in the membrane fraction. However, it was found that Mistic-β2AR and Mistic-SSTR5 fusion proteins alongside with the monomer found in membrane fraction of the bacteria contain the same amount of aggregates as the protein localized in inclusion bodies. It is known that orientation of bacterial proteins in membrane is determined by distribution of charged amino acid residues in their loop regions [38]. This distribution can differ dramatically in the molecules of eukaryotic membrane proteins. In addition, for correct folding of most of the GPCR, the presence of certain lipids absent in bacterial membrane is required [39]. Thus, the possibility of Mistic-GPCR fusion protein to incorporate into the membrane is likely to be determined by the properties of the target protein, and violation of this process leads to formation of aggregates.

To confirm the screening results, four target proteins obtained by direct expression (GPR4, GPR68, CCR4, and SSTR5) and four proteins expressed as proteins fused with Mistic (β2AR, CCR2b, SSTR5, MC2R) were isolated. The standard purification method using metal affinity chromatography in denaturing conditions yielded high-purity recombinant GPCR in all cases. Notably, in all preparations of the target GPCR obtained, high molecular weight aggregates as well as monomeric forms of the protein are observed (Fig. 3b). This phenomenon can be explained by formation of strong intermolecular interactions provided by hydrophobic transmembrane domains of the proteins during formation of inclusion bodies. Consequently, treatment with lauroyl sarcosine is not sufficient to break these interactions. In addition, all of the studied GPCR contain a large number of cysteine residues localized predominantly in hydrophobic zones, which limits the possibility of their reduction using water-soluble agents (β-mercaptoethanol or dithiothreitol). Thus, study of the structure of recombinant GPCR obtained by expression in bacterial cells in the form of inclusion bodies will require further elaboration of methods for purification and renaturation of the target proteins.

Comparison of the results of primary screening of expression of GPCR genes with the yield of the purified proteins has convincingly demonstrated efficiency of the suggested approach. In all cases when high expression level was observed during screening, it correlated with high yield of the purified target protein. When the same protein could be obtained using different expression systems, the screening results enabled us to use the more effective system. For example, according to the dot-blot results, expression of SSTR5 fused with Mistic was more efficient than direct expression, which was confirmed by comparison of the yields of the purified proteins.

Thus, as a result of our study, a high-efficiency approach for expression of genes of recombinant GPCR in *E. coli* cells including optimized plasmid vector, selection of partner protein for hybrid expression, producer strain, and cultivation strategy was elaborated. Efficiency of the approach has been demonstrated on 17 human GPCR. Our data can be used for construction of expression systems and production of preparative amounts of recombinant GPCR and other membrane proteins for study of their structures as well as for obtaining mutants of these proteins with the goal of determining the molecular mechanisms of their functioning.

We thank Dr. O. V. Nekrasova for providing the pET-Mistic plasmid and Dr. L. N. Shingarova for providing pETH10 DNA.

This work was supported by Novartis (Switzerland), the Russian Foundation for Basic Research (grant 09-04-01585a), and the program "Molecular and Cellular Biology" of the Russian Academy of Sciences.

REFERENCES

1. Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009) *Nature*, **459**, 356-363.
2. Blois, T. M., and Bowie, J. U. (2009) *Protein Sci.*, **18**, 1335-1342.
3. Kobilka, B., and Schertler, G. F. (2008) *Trends Pharmacol. Sci.*, **29**, 79-83.
4. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) *Science*, **289**, 739-745.
5. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) *Science*, **318**, 1258-1265.
6. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobilka, B. K. (2007) *Nature*, **450**, 383-387.

7. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G., and Schertler, G. F. (2008) *Nature*, **454**, 486-491.
8. Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P., and Stevens, R. C. (2008) *Science*, **322**, 1211-1217.
9. Mancia, F., and Hendrickson, W. A. (2007) *Mol. Biosyst.*, **3**, 723-734.
10. Sarramegna, V., Muller, I., Mousseau, G., Froment, C., Monsarrat, B., Milon, A., and Talmont, F. (2005) *Protein Expr. Purif.*, **43**, 85-93.
11. Lundstrom, K., Wagner, R., Reinhart, C., Desmyter, A., Cherouati, N., Magnin, T., Zeder-Lutz, G., Courtot, M., Prual, C., Andre, N., Hassaine, G., Michel, H., Cambillau, C., and Pattus, F. (2006) *J. Struct. Funct. Genomics*, **7**, 77-91.
12. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, N. Y.
13. Studier, F. W. (2005) *Protein Expr. Purif.*, **41**, 207-234.
14. Laemmli, U. K. (1970) *Nature*, **227**, 680-685.
15. Towbin, H., Strahelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350-4354.
16. Tian, C., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A., and Sanders, R. (2005) *J. Am. Chem. Soc.*, **127**, 8010-8011.
17. Kiefer, H., Krieger, J., Olszewski, J. D., von Heijne, G., Prestwich, G. D., and Breer, H. (1996) *Biochemistry*, **35**, 16077-16084.
18. Baneres, J. L., Mesnier, D., Martin, A., Joubert, L., Dumuis, A., and Bockaert, J. (2005) *J. Biol. Chem.*, **280**, 20253-20260.
19. Tucker, J., and Grisshammer, R. (1996) *Biochem. J.*, **317**, 891-899.
20. Attrill, H., Harding, P. J., Smith, E., Ross, S., and Watts, A. (2009) *Protein Expr. Purif.*, **64**, 32-38.
21. Baneres, J. L., Martin, A., Hullot, P., Girard, J. P., Rossi, J. C., and Parello, J. (2003) *J. Mol. Biol.*, **329**, 801-814.
22. Park, S. H., Prytulla, S., de Angelis, A. A., Brown, J. M., Kiefer, H., and Opella, S. J. (2006) *J. Am. Chem. Soc.*, **128**, 7402-7403.
23. Kirpichnikov, M. P., Goncharuk, M. V., Yermolyuk, Ya. S., Goncharuk, S. A., Shulga, A. A., Maslennikov, I. V., and Arsenyev, A. S. (2005) *Tekhnol. Zhivykh Sistem*, **2**, 20-27.
24. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009) *Proc. Natl. Acad. Sci. USA*, **106**, 1760-1765.
25. Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A., Wagner, C. S., Nord, O., Nygren, P. A., van Wijk, K. J., and de Gier, J. W. (2007) *Mol. Cell. Proteomics*, **6**, 1527-1550.
26. Freissmuth, M., Selzer, E., Marullo, S., Schutz, W., and Strosberg, A. D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8548-8552.
27. Yeliseev, A. A., Wong, K. K., Soubias, O., and Gawrisch, K. (2005) *Protein Sci.*, **14**, 2638-2653.
28. Zhang, S., Zubay, G., and Goldman, E. (1991) *Gene*, **105**, 61-72.
29. Ren, H., Yu, D., Ge, B., Cook, B., Xu, Z., and Zhang, S. (2009) *PLoS One*, **4**, e4509.
30. Miroux, B., and Walker, J. E. (1996) *J. Mol. Biol.*, **260**, 289-298.
31. Seo, S. W., Yang, J., and Jung, G. Y. (2009) *Biotechnol. Bioeng.*, **104**, 611-616.
32. Smith, D. B., and Johnson, K. S. (1988) *Gene*, **67**, 31-40.
33. Jeong, K. J., and Lee, S. Y. (2002) *Appl. Environ. Microbiol.*, **68**, 4979-4985.
34. Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D., and Beckwith, J. (2003) *J. Bacteriol.*, **185**, 5706-5713.
35. Luo, J., Choulet, J., and Samuelson, J. C. (2009) *Protein Sci.*, **18**, 1735-1744.
36. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) *Science*, **307**, 1317-1321.
37. Driessen, A. J., and Nouwen, N. (2008) *Annu Rev. Biochem.*, **77**, 643-667.
38. Von Heijne, G., and Gavel, Y. (1988) *Eur. J. Biochem.*, **174**, 671-678.
39. Opekarova, M., and Tanner, W. (2003) *Biochim. Biophys. Acta*, **1610**, 11-22.