

## Structural genomics and drug discovery

K. Lundstrom \*

Flamel Technologies, Vénissieux, France

Received: January 16, 2007; Accepted: February 15, 2007

- Background to drug discovery
- Structure determination and drug discovery
- Expression systems
  - Cell-free translation
  - Bacterial expression
    - Expression in *E. coli*
    - Other prokaryotic systems
  - Expression in yeast cells
  - Expression in insect cells
    - Baculovirus
    - Other insect vectors
- Expression in mammalian cells
  - Transient and stable expression
  - Viral vectors
- Protein purification
- Structure determination
  - X-ray crystallography
  - Nuclear magnetic resonance
  - Electron microscopy
  - Structural genomics programs
- Conclusions and future prospects

### Abstract

Structure determination has already proven useful for lead optimization and direct drug design. The number of high-resolution structures available in public databases today exceeds 30,000 and will definitely aid in structure-based drug design. Structural genomics approaches covering whole genomes, topologically similar proteins or gene families are great assets for further progress in the development of new drugs. However, membrane proteins representing 70% of current drug targets are poorly characterized structurally. The problems have been related to difficulties in obtaining large amount of recombinant membrane proteins as well as their purification and structure determination. Structural genomics has proven successful in developing new methods in areas from expression to structure determination by studying a large number of target proteins in parallel.

**Keywords:** structural genomics • drug discovery • structure-based drug design • recombinant protein expression • protein purification • structure determination • X-ray crystallography • nuclear magnetic resonance • electron microscopy

### Background to drug discovery

Drug discovery has relied to a large extent on medicinal chemistry, which obviously has evolved significantly during the past two decades [1]. Although screening of large numbers of compounds and

gigantic combinatorial libraries have generated a number of efficient drugs, increasing interest has been dedicated to structure-based drug design. The advantage of this approach is that drug molecules

---

\*Correspondence to: Kenneth LUNDSTROM, PhD  
Flamel Technologies, 33 Avenue du Dr. Georges Lévy  
69693 Vénissieux, France.

Tel.: +33 472 78 34 41  
Fax: +33 472 78 34 24  
E-mail: lundstrom@flamel.com

can be 'tailor-made' to interact with the drug target. This property should improve the drug efficacy and also specificity. Improved efficacy might allow the administration of lower drug doses and/or increasing the distribution frequency. Drugs presenting enhanced target specificity should limit the interaction with other non-specific target molecules and hence reduce the side effects of the drug substantially. Furthermore, the drug development process might be shortened.

To apply structure-based drug design, it is necessary to have access to structural information of the drug target. More than 30,000 high-resolution structures have been deposited in public databases, the majority of these however are on soluble proteins. In contrast, some 70% of current drugs are targeted against membrane proteins, for which only more than 100 structures are available [2]. This discrepancy relates mainly to the topological composition of membrane proteins. A large number of them such as G protein-coupled receptors (GPCRs) and ion channels have a topology, including multispansing transmembrane domains [3]. For this reason, it has been much more difficult to recombinantly express membrane proteins in quantitatively and qualitatively sufficient amounts compared with those of their soluble counterparts. Moreover, as the membrane proteins are embedded in membranes detergents are required for purification, which affects both yields and stability of the protein [4]. Furthermore, structure determination by X-ray crystallography has been negatively affected by low yields and presence of detergents [5]. Membrane proteins also possess characteristically flexible regions and short hydrophilic loops, which reduces significantly the potential crystal contacts and thereby crystallization success. The application of nuclear magnetic resonance (NMR) for structure determination is also more complicated for membrane proteins. Typically, only a single high-resolution structure has been solved for the family of GPCRs consisting of some 800 members [6]. Moreover, the structure for bovine rhodopsin was obtained on material isolated from native tissue, where the receptor occurs at high abundance.

The structural genomics approach allows the study of a large number of gene products in parallel. By targeting not only whole genomes but also specifically certain types or families of proteins, the understanding of the requirements for obtaining high levels of expression and the conditions for purification and

structure determination can significantly improve the success rate for membrane proteins.

In this review, although a general overview on structural genomics is presented, the focus is on membrane proteins. The status of applying structure-based drug discovery is briefly described. Much attention is given to the methodological development for structural biology. An overview of currently applied expression systems is given. Also various purification and structure determination procedures are described. Finally, examples of various structural genomics consortia are discussed.

## **Structure determination and drug discovery**

Structural biology has facilitated drug discovery already for some time. The application of structure-based drug design has become relatively common in lead optimization [7]. Structural knowledge of proteins and their ligands has aided in improving drug potency and selectivity [8]. This approach has resulted in faster definition of drug-binding properties and has made it easier to identify 'hit' compounds through screening programs [9]. Both the use of X-ray crystallography and NMR have allowed high throughput approaches for structure-based lead discovery [10]. Rapid structure resolution of protein-ligand complexes has been obtained applying such automated procedures as AutoSolve<sup>®</sup> [11]. The technology revealed that from cocktails of 100 molecules, electron density variations could be used to distinguish their different shapes [12]. Applying cocktails of smaller fragments at very high concentrations candidate fragment ranking can be performed automatically and up to 1000 compounds can be screened within 2–3 days [13]. In this context, fragments of successful drug-like molecules can be characterized in high throughput format taking into account molecular weight, the presence of hydrogen bond donors and acceptors and solubility. The technology has been expanded to virtual screening approaches by systematically docking large libraries of candidate fragments in pre-defined binding sites of target proteins applying three-dimensional (3D) computer models [11]. Furthermore, in structure-based drug design an NMR-based screening approach was

taken to design small molecule drug candidates to inhibit the aberrant over-expression of c-myc in different tumours by targeting the far upstream element (FUSE) binding protein (FBP) [14].

There are a number of drugs for which direct applications of structure-based design has been achieved. So far, more than 40 compounds based on structural data have entered clinical trials and at least 7 drugs have reached the market [15]. In this context, the high-resolution structure of the human immunodeficiency virus (HIV) proteinase was the basis for the AIDS drugs Agenerase<sup>®</sup> and Viracept<sup>®</sup> [16]. Another structure-based drug is the flu drug Relenza<sup>®</sup>, which is based on the structure of the influenza virus neuraminidase [17]. Furthermore, the design of the protein kinase drug Gleevec<sup>®</sup> profited from the 3D structure of the kinase domain of c-Abl especially in addressing resistance issues [18]. New structure-based drugs will certainly appear as novel high-resolution structures of relevant drug targets will become available. For instance, the crystal structure of AKT kinase has triggered structure-based drug discovery on kinase inhibitors [19].

Perhaps the most peculiar example of structure-based drug discovery comes from the program on novel phosphodiesterase (PDE) inhibitors to treat hypertension and other cardiovascular indications [8]. According to the structural information, novel PDE inhibitors were designed of which the compound UK 92480 (sildenafil) showed a 100-fold increase in PDE5 inhibition compared with zaprinast [20]. Despite the expected pharmacological profile of sildenafil its clinical performance in treating coronary heart disease was disappointing. However, as sildenafil inhibited efficiently PDE5 and potentiated nitrose oxide activity, the drug could be used for the treatment of erectile dysfunction. Although originally aimed for another indication structure-based drug design assisted in generating the globally well-known blockbuster drug Viagra<sup>®</sup>.

## Expression systems

As the minority of proteins is available in high abundance allowing direct isolation and purification from native tissues and furthermore ethical issues to large extent prevent any such action, structural biology strongly relies on heterologous expression systems for the production of recombinant proteins. During

the years since the advent of genetic engineering technologies, a number of various expression systems have been evaluated. Among these are heterologous expression in bacterial, yeast, insect and mammalian cells applying different types of expression vectors. More recently, cell-free translation systems have become available for general use. Today, the two most frequently used systems are based on expression in *Escherichia coli* and Baculovirus-infected insect cells. Generally, either of these two systems is applicable for expression of almost any soluble protein in quantities acceptable for structural studies. However, concerning membrane proteins the situation is quite different. The yields are substantially lower and the stability of the produced proteins low, which has forced major development activities for expression vectors and systems (Table 1). The main expression systems are briefly described later. Expression levels are commonly described for GPCRs based on binding activity ( $B_{\max}$  pmol receptor per milligram protein) or receptor yields (milligrams per liter). For an easier comparison, a  $B_{\max}$  value of 10 pmol/mg corresponds to 0.5 mg/l, 50 pmol/mg to 2 mg/l and 150 pmol/mg to 10 mg/l.

## Cell-free translation

The recent development of cell-free translation systems has made them attractive and competitive [21]. Currently, commercial systems based on *E. coli* and wheat germ extracts exist. Cell-free translation systems are applicable to rapid high throughput expression evaluation especially since the systems can be directly used for PCR fragments, which omits cloning procedures. The advantages of cell-free translation is the controlled expression in defined minimal medium and simple amino-acid selective and uniform stable isotope labelling for direct sample analysis by NMR [22].

Cell-free translation has supported structural biology well for soluble proteins. Recently, a modified *E. coli* S30 extract has improved the yields for membrane proteins and the bacterial multi-drug transporters TehA and YfiK were expressed at levels of 2.7 mg/ml [23]. Expression of three GPCRs (human  $\beta_2$  adrenergic, human muscarinic acetylcholine M2 and rat neurotensin receptor) resulted in functional binding activity albeit only after applying fusion protein partners [24]. Recently, the first demonstration of

**Table 1** Expression vectors and systems and their applications

System	Targets	Expression levels*	Applications/comments
Cell-free systems <i>E. coli</i> extracts Wheat germ extracts	S, MP, GPCR	relatively high mg/ml (S)	parallel expression, also PCR products isotope labelling for NMR analysis scale-up costly, fusion proteins helpful
<b>Bacteria</b> <i>E. coli</i>	S, MP, GPCR	relatively high in PM high in IB	structural biology: solubilization (PM) refolding from IBs required
<i>H. salinarum</i> <i>B. subtilis</i> <i>L. lactis</i>	MP, GPCR secreted transporters	high for Bop high relatively high	fusion proteins often helpful high protease activity good for mitochondrial transporters
<b>Yeast</b> <i>S. cerevisiae</i> <i>S. pombe</i> <i>Pichia pastoris</i>	S, MP, GPCR S, MP, GPCR S, MP, GPCR S, MP, GPCR	relatively high 1--10 pmol/mg 1--20 pmol/mg >100 pmol/mg; 5 mg/l	structural biology structure of SERCA 1a improved expression from <i>S. cerevisiae</i> structure SoP1P2, Shaker K+ channel
<b>Insect cells</b> <i>Baculovirus</i> <i>Drosophila</i>	S, MP, GPCR	high-very high 200 pmol/mg; 5 mg/l 10 pmol/mg	structure biology, large-scale structures on soluble proteins stable expression
<b>Mammalian cells</b> Transient	S, MP, GPCR	medium 20 pmol/mg; <1 mg/l	scale-up more expensive, complicated
Stable		low to high 200 pmol/mg	varies, inducible systems improvement codon-optimized, mutant HEK293 cells
<b>Viral</b> Adenovirus Lentivirus SFV Vaccinia virus	S, MP, GPCR	medium, 10 pmol/mg high 287 pmol/mg; 10 mg/l high-very high	broad host range commercial system easy scale-up T7 based vectors

Bop = bacteriopsin protein; IB = inclusion bodies; MP = membrane protein; PM = plasma membrane; S = soluble protein

\*Estimations of expression levels of GPCRs are presented as binding activity (Bmax) in pmol receptor per milligram protein or as receptor yields in milligrams per liter. Approximately, 10–20 pmol/mg is equivalent to 0.5–1 mg/l, 40–70 pmol/mg corresponds to 2–5 mg/l and 100–200 pmol/mg represents 7–15 mg/l.

cell-free translation-based high-resolution structure determination came from solving the X-ray structure of the EmrE multidrug transporter complex [25].

## Bacterial expression

### Expression in *E. coli*

The by far most commonly used expression system is based on *E. coli* [26]. The fast and simple expression profile, cheap and safe large-scale production and ver-

satility for different types of proteins are factors that have made *E. coli* the system of choice. However, although a number of high-resolution structures have been solved for soluble proteins recombinantly expressed in bacteria, membrane proteins have been less successful. Because of their transmembrane topology, membrane proteins have generally proven highly toxic to the bacterial host cells and therefore restricted their growth and thereby reducing the recombinant protein yields. Despite these shortcomings, a number of bacterial membrane proteins have

successfully been expressed in *E. coli* and subjected to purification and structural analysis [27]. Successful expression of eukaryotic membrane proteins has required modifications (deletions, mutagenesis) of the gene sequence as well as the application of appropriate fusion partners and purification tags. In this context, the fusion of the maltose-binding protein (MBP) to the *N*-terminal of the rat neurotensin receptor generated milligram quantities of purified functionally active receptor [28]. In a similar way, the adenosine A2a receptor with a truncated *C*-terminal resulted in 10–20 nM/l of receptor when expressed as a fusion protein with MBP [29]. Over-expression of GPCRs in bacterial membranes has recently been reviewed [30].

A completely different approach has been to over-express recombinant proteins in bacterial inclusion bodies as aggregates. The yields have been substantially higher by this procedure, but re-folding is necessary to obtain functionally active recombinant protein. Unfortunately, the refolding process has been difficult and inefficient [31]. However, recent technology development has allowed refolding of glucagon-like peptide 1 (GLP-1) [32], leukotriene B4 [33] and serotonin 5-HT4 [34] receptors.

### Other prokaryotic systems

In addition to *E. coli*, other bacterial expression systems have been evaluated for production of recombinant proteins. For instance, the Gram-positive bacterium *Lactococcus lactis* has been verified for the expression of both prokaryotic and eukaryotic proteins [35]. Specific vectors using the nisin NisA promoter and NisR and NisK regulatory sequences as well as transformation methods have been established for *L. lactis* [36]. A number of membrane proteins have been verified for expression. The prokaryotic ABC transporters and Major Facilitator Subfamily (MFS) efflux pumps [35] as well as yeast mitochondrial carrier proteins CTP1 and AAC3 were well expressed [36]. However, expression of the human KDEL receptor resulted in yields less than 0.1% of the total membrane protein [36]. In a recent study, 11 yeast mitochondrial transporter proteins were expressed at structural biology compatible levels in *L. lactis* [37]. Expression levels could be enhanced 10-fold by replacing the *N*-terminus of the transporter with a signal sequence from *L. lactis*. *Bacillus subtilis*, another Gram-positive organism, was already in

the 1980s considered as a good host for recombinant protein expression generating large quantities of recombinant interferon [38]. However, serious high endogenous protease activity and mainly restriction to secreted proteins has limited the use of the system, although recently recombinant human cystatins have been efficiently produced in *B. subtilis* [39].

*Halinobacterium salinarum* became an interesting alternative as a prokaryotic expression host because of the accumulation of bacterio-opsin protein (Bop), which with the chromophore retinol gives the bacteria their characteristic purple colour [40]. Various membrane proteins such as the *E. coli* aspartate transcarbamylase (AT), the yeast  $\alpha$  mating factor receptor and two human GPCRs (muscarinic M1 and serotonin 5-HT2 receptors) have been expressed from *H. salinarum* vectors [41]. It was essential to use Bop as fusion partner. The Bop-AT yielded 7 mg/l protein, whereas the GPCRs were expressed at extremely low levels. A fusion construct of Bop and the human adrenergic  $\alpha$ 2B receptor resulted in functional binding activity, although with 10 times lower binding values than obtained in yeast or mammalian cells [42].

### Expression in yeast cells

Yeast expression systems can be characterized by their easy use and large-scale production and their eukaryotic post-translation machinery. Various types of recombinant proteins have successfully been expressed in yeast cells [43]. Baker's yeast *Saccharomyces cerevisiae* has been used for heterologous expression of hepatitis B surface antigen (HBsAg) [44],  $\alpha$ 1-Antitrypsin [45], human  $\alpha$ -interferon [46] and,  $\beta$ -endorphin [47]. *S. cerevisiae* has also been frequently used for expression of membrane proteins. For instance, the yeast  $\alpha$ -factor Ste2p receptor [48] and human dopamine D1A receptor [49] generated high levels of expression. Moreover, large-scale production of the human  $\beta$ 2 adrenergic receptor generated yields of 20–30 mg of functional receptor [50]. Recently, the rabbit sarcoplasmic-endoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase isoform 1a (SERCA 1a) was purified from yeast cells by metal affinity chromatography and HPLC filtration and the structure solved at 3.3 Å resolution [51]. In addition to *S. cerevisiae*, the fission yeast *Schizosaccharomyces pombe* has also been applied to heterologous gene



expression [52]. Comparison of GPCR expression demonstrated that the human dopamine D2 receptor was expressed at fivefold higher levels in *S. pombe* than in *S. cerevisiae* [53]. In contrast, the rat dopamine D2 and human NK1 receptors were expressed at lower levels in *S. pombe* [54].

Currently, the most frequently used yeast strain for recombinant protein expression is *Pichia pastoris* [55]. This methylotrophic yeast expression system is based on chromosomal integration of the gene of interest and utilization of strong inducible promoters such as the alcohol oxidase (AOX) promoter [56]. One of the advantages of *P. pastoris* is the high biomasses obtained in large-scale cultures allowing production of g/L quantities of recombinant proteins [57]. Today, more than 200 recombinant proteins have been expressed in *P. pastoris*. A number of membrane proteins have been expressed in *P. pastoris*, particularly GPCRs [58, 59]. In a structural genomics-type study, 100 GPCRs were expressed in *P. pastoris*, for which a large number structural biology compatible levels (1–10 mg/L) were obtained [60]. The highest levels of expression of 180 pmol/mg receptor were measured for the human adenosine A2a receptor. Applying *P. pastoris* vectors, expression and purification allowed to obtain a high-resolution structure of the spinach aquaporin SOPIP2 channel [61]. Likewise, single particle imaging and cryo-EM two-dimensional (2D) crystals could be obtained for the rat neuronal voltage-sensitive K<sup>+</sup> channel over-expressed in *P. pastoris* [62]. Finally, the mammalian Shaker voltage-dependent K<sup>+</sup> channel was expressed in *P. pastoris*, which allowed efficient purification and high-resolution structure determination [63].

## Expression in insect cells

### Baculovirus

The second most commonly used expression system after *E. coli* is based on Baculovirus vectors for infection of insect cell lines [64]. High expression levels of topologically different recombinant proteins have been obtained in different insect cell lines from *Spodoptera frugiperda* (Sf9 and Sf21), *Mamestra brassica* and *Trichoplusia ni* (high five). Eukaryotic and especially mammalian membrane proteins have been relatively favourably expressed in insect cells

because of the similar post-translational processing mechanisms present in insect cells. In this context, expression of rhodopsin resulted in 80% functional receptor and yields up to 6 mg/l [65]. GPCRs have been popular targets for baculovirus expression, and in many cases expression levels of 40–60 pmol/mg have been achieved [66]. Expression levels between 1 pmol/mg and 250 pmol/mg were obtained in a study of 16 GPCRs in three insect cell lines [67].

### Other insect vectors

In addition to Baculovirus-based systems, expression of recombinant proteins has been conducted mainly in stable insect cell lines, typically *Drosophila* Schneider cells [68]. A soluble deglycosylated form of the human interleukin 5 (IL-5) alpha subunit was expressed in *Drosophila* cells in active form, purified and a 2.6 Å resolution crystal structure was solved [69]. Schneider cells have also been used as hosts for GPCR expression. For instance, the human mu opioid receptor (hMOR) showed a similar pharmacological profile as in mammalian cells and the functional coupling to G proteins was demonstrated by cAMP stimulation and GTP $\gamma$ S binding assays [70]. Engineering of an N-terminal EGFP tag to the hMOR for localization studies suggested that a large number of receptors were retained in intracellular compartments and not present on the plasma membrane.

## Expression in mammalian cells

### Transient and stable expression

Mammalian cell lines provide the most native environment for expression of recombinant mammalian proteins. However, immortalized cell lines are severely compromised and have significantly different features to primary cells. Furthermore, specific proteins may require accessory proteins for transport, folding and proper function. Recently, it was demonstrated that certain transmembrane proteins (RTP1, RTP2 and REEP1) [71], and the guanine nucleotide exchange factor Ric-8B [72] can facilitate the transport and enhance the expression of olfactory receptors. Transient expression has been conducted in various cell lines (BHK-21, CHO-K1, COS-7 and HEK293) [73]. Alternatively, stable cell lines have been generated [74]. Generally, the expression levels have been higher in transient expression, and

recently the codon-optimized hamster  $\beta 2$  adrenergic receptor ( $\beta 2$ -AR) was expressed in COS-1 cells at 18 pmol/mg [75]. Typically, GPCRs are expressed in the pmol range in stable cell lines [76]. However, the development of a tetracycline-inducible system resulted in up to 6 mg/l of rhodopsin production [77]. When a stable inducible HEK293- $\beta 2$ -AR cell line was established, 220 pmol/mg receptor was obtained [75]. This expression level resulted in up to 50  $\mu$ g of  $\beta 2$ -AR per 15-cm-cell culture plate.

### Viral vectors

Because of their broad host range and generally high expression levels, viral vectors have presented attractive alternatives for heterologous gene expression. However, the high transduction rate has also raised some concerns related to biosafety, which has required engineering of mutant replication-deficient vectors. A number of viral vectors such as adenoviruses, alphaviruses, lentiviruses and vaccinia viruses have been applied for recombinant protein expression [78]. In this context, adenovirus-based expression of the non-structural NS1 glycoprotein of tick borne encephalitis virus (TBEV) resulted in yields up to 25% of total protein [79]. A number of GPCRs have also been expressed from adenovirus vectors resulting in relatively high expression levels [80]. Owing to the broad host range of adenoviruses, expression studies could be conducted in various cell lines, and for instance the  $\beta 2$ -AR was expressed in rabbit myocytes [81]. Vaccinia virus vectors have also been frequently used for heterologous gene expression especially applying replication-deficient vectors [82]. More than 100 proteins have been expressed from vaccinia virus vectors [83]. Among GPCRs neuropeptide Y [84] and dopamine D2 and D4 receptors [85] have been expressed at densities of 5–10 million receptors per cell. Lentivirus vectors, characterized for their long-term expression pattern, have found more applications for recombinant protein expression with the commercialization of complete lentivirus expression systems [86]. Concerning membrane proteins, the human retinal pigment epithelium (RPE) retinal GPCR was expressed in COS-7 cells and in the retinal pigment epithelial cell line ARPE-19 [87]. Interestingly, the expression levels were 100 times higher in the ARPE-19 cells and long-term expression was detected up to 6 months.

The viral system probably most frequently used for recombinant protein expression and particularly membrane proteins is based on the Semliki Forest virus (SFV), a single-stranded enveloped RNA virus [88]. The easy and rapid production of recombinant SFV stocks has made it feasible to express a large number of proteins in parallel in various mammalian cell lines [89]. In this context, 103 GPCRs were evaluated for expression levels in three mammalian cell lines in a structural genomics program described in more detail below [90]. As large-scale production in mammalian suspension cultures has been established for SFV, ligand-gated ion channels [91] and GPCRs [92] have been expressed at levels of 5–10 mg/L, purified and subjected to structural biology.

## Protein purification

Structural characterization of protein generally requires access to highly homogenous and pure protein preparations although the solution NMR approaches has allowed analysis of labelled samples in the presence of a relatively high background on non-labelled proteins. Genetic engineering has strongly facilitated the purification procedure as various affinity tags have been introduced at the *N*- or *C*-terminal or even within the coding sequence of the gene of interest. The most commonly used tag is multi-histidine (either hexa or deca histidine), which binds to  $\text{Ni}^{2+}$  and therefore allows purification based on immobilized metal affinity chromatography (IMAC) [93]. Other common purification tags are streptavidin (Strep), biotin, FLAG and hemagglutinin tags. Obviously, when available, antigen-based affinity chromatography can be applied. Other means for purification include ammonium sulphate precipitation and sucrose gradients, although these methods require large quantities of material and might therefore not be suitable for recombinant proteins expressed at low levels or membrane proteins with low recovery yields. In addition, gel filtration, size exclusion chromatography, hydrophobic interaction and reverse-flow chromatography are methods to be considered for protein purification.

Naturally, membrane proteins require special conditions for purification. Because of their transmembrane topology, separation of proteins and lipids is

necessary by the addition of detergents [94]. Solubilization by detergents is a complicated process, and a vast number of detergents have been tested. In general, detergents are highly target-specific, which means that each target has to be screened for appropriate detergents [95]. Commonly used detergents are CHAPS (3-[(3-cholamidopropyl)dimethylamino]-1-propane-sulfonate, Triton X-100, n-Octylglucoside, n-Nonylglucoside, n-Dodecylmaltoside and FOS-Choline and cocktails thereof [95].

## Structure determination

### X-ray crystallography

The highest structure resolution, below 2 Å, can be achieved by X-ray crystallography only. Although a large number of X-ray structures are available today, crystallography still faces some serious challenges. This is mainly due to the success of purification and the stability of the purified protein. Needless to say, membrane proteins present additional obstacles as the purification is often inefficient and the presence of detergents may interfere with the crystallization process. However, major development has taken place with the introduction of automation and miniaturization [96]. The reduction of volumes to nanoliter scale has significantly reduced the material quantities required, and together with high throughput crystallization in 96 micro plate and higher format has permitted screening of numerous crystallization parameters and conditions in parallel [97]. In this context variables such as pH, ionic strength, temperature and concentration of salts and detergents can be screened and up to 100,000 crystallization trials conducted per day. The increasing number of parallel experiments also requires improved data collection and handling capacity. A drawback of the miniaturization process for the crystal screening might be an increase in production of smaller crystals, which can be addressed by improved micro-diffractometer technologies.

### Nuclear magnetic resonance

Complementary to X-ray crystallography, NMR can serve structure determination [98]. Especially, cell-

free translation systems for isotope labelling have been heavily tuned towards NMR applications. NMR has been routinely used for the identification and evaluation of chemical leads [99] and recent technology development for probes, software and NMR itself has made it possible to obtain high-resolution structures and subject NMR technology to larger-sized proteins. Improved technology has allowed to apply NMR also to iterative ligand-protein complexes [100]. Recent development of solid state and solution NMR technologies has further expanded the application range in structural biology especially for membrane proteins [101].

### Electron microscopy

In addition to X-ray crystallography and NMR, electron microscopy (EM) and atomic force microscopy (AFM) can also be used to obtain information at atomic resolution levels for protein structures [102]. Cryoelectron microscopy has successfully been applied for reconstituted membrane proteins in 2D crystals. In this context, a 3.5 Å resolution was achieved for bacteriorhodopsin [103] and aquaporin AQP1 [104]. Despite this rather low resolution it was possible to define the atomic structure, which was subsequently confirmed by X-ray crystallography. Structural characterization by AFM of polypeptide loops on native and reconstituted membranes in aqueous solutions demonstrated that rhodopsin in the disc membranes of vertebrate photoreceptor rod outer segments occurred as dimers or higher oligomeric forms [105].

### Structural genomics programs

Continuous technology development in the areas of molecular biology, protein expression and purification, structure determination requires increased expertise and resources in various areas. To facilitate substantial efforts in applying several expression systems in parallel for numerous targets it has been advantageous to form large national and international networks (Table 2). This development strongly encouraged structure biology in large scale and formed the basis for structural genomics. Many of the



**Table 2** Overview of selected structural genomics networks applications

Network	Activities
Berkeley Structural Genomics Center (BSGC) www.strgen.org	studies on <i>Mycoplasma genitalium</i> and <i>Mycoplasma pneumoniae</i> genomes; expression in <i>E. coli</i>
Center for Eukaryotic Structural Genomics (CESG) www.uwstructuralgenomics.org	studies on <i>Arabidopsis thaliana</i> transcriptome; expression in <i>E. coli</i>
European Membrane Proteins (E-MeP) www.e-mep.org	studies on 100 prokaryotic, 200 eukaryotic MPs; expression in <i>E. coli</i> , <i>Lactococcus lactis</i> , <i>P. pastoris</i> , <i>Saccharomyces cerevisiae</i> , baculovirus, SFV
Joint Center for Structural Genomics (JCSG) www.jcsg.org	studies on <i>Thermotoga maritima</i> proteome; human GPCRs expression in <i>E. coli</i> , baculovirus, adenovirus, SFV
Membrane Protein Network (MePNet) www.mepnet.org	studies on >100 GPCRs; expression in <i>E. coli</i> , <i>P. pastoris</i> and SFV/mammalian cells
Membrane Protein Platform (MPP) www.swegene.org	bacterial and yeast membrane proteins; human GPCRs; expression in <i>E. coli</i> , <i>S. cerevisiae</i> , <i>P. pastoris</i> ; lipid cubic phase crystallography
Midwest Center for Structural Genomics (MCSG) www.mcsg.anl.gov	targets from all three kingdoms of life; expression in <i>E. coli</i>
Northeast Structural Genomics Consortium (NESG) www.nigms.nih.gov/Initiatives/PSI/Centers/NECSG.htm	small proteins: <i>S. cerevisiae</i> , <i>C. elegans</i> and <i>D. melanogaster</i> ; expression in <i>E. coli</i> , yeast and insect cells
New York Structural Genomics Research Consortium (NYSGXRC) www.nysgrc.org	bacterial, yeast and <i>C. elegans</i> proteins; expression in <i>E. coli</i> and yeast
Paris-Sud Yeast Structural Genomics (YSG) www.genomics.eu.org	250 non-membrane yeast proteins; expression in <i>E. coli</i>
Protein Structure Factory (PSF) www.proteinstrukturfabrik.de	medically and biotechnologically valid proteins; expression in <i>E. coli</i> , <i>S. cerevisiae</i> and <i>P. pastoris</i>
Protein Wide Analysis of Membrane Proteins (ProAMP) www.pst-ag.com	<i>Salmonella typhimurium</i> and <i>Helicobacter pylori</i> MPs; expression in <i>E. coli</i>
RIKEN Structural Genomics Initiative (RSGI) www.rsgi.riken.go.jp/rsgi_e/index.html	mouse, <i>Arabidopsis thaliana</i> and <i>Thermus thermophilus</i> proteins; expression in <i>E. coli</i> and cell-free systems

Network	Activities
Southeast Collaboratory for Structural Genomics (SECSG) www.secsg.org	<i>Pyrococcus furiosus</i> , <i>C. elegans</i> and human proteins; expression in <i>E. coli</i> , baculovirus and lentivirus
Structural Proteomics in Europe (SPINE) www.spineurope.org	proteins and protein complexes with direct relevance to human health and diseases; expression in <i>E. coli</i> , baculovirus, transient mammalian cells
Structural Genomics Consortium (SGC) www.sgc.ox.ac.uk	targets related to human health: diabetes, cancer, infectious diseases (malaria); expression in <i>E. coli</i> , baculovirus
Structure 2 Function Project (S2FP) www.s2f.carb.nist.gov	structural genomics initiative on <i>Haemophilus influenzae</i> proteins; expression in <i>E. coli</i>
Swiss National Center of Competence in Research (NCCR) www.structuralbiology.ethz.ch	bacterial membrane proteins, transporters and GPCRs; expression in <i>E. coli</i> and baculovirus
TB Structural Genomics Consortium (TBSGC) www.mbi-doe.ucla.edu/TB	structural genomics initiative on <i>Mycobacterium tuberculosis</i> proteins; expression in <i>E. coli</i>

Bop = bacteriopsin protein; IB = inclusion bodies; MP = membrane protein; PM = plasma membrane; S = soluble protein  
 \*Estimations of expression levels of GPCRs are presented as binding activity (Bmax) in pmol receptor per milligram protein or as receptor yields in milligrams per liter. Approximately, 10–20 pmol/mg is equivalent to 0.5–1 mg/l, 40–70 pmol/mg corresponds to 2–5 mg/l and 100–200 pmol/mg represents 7–15 mg/l.

networks have selected their targets from a specific organism (whole genomes), topologically similar types of proteins or protein families. Understandably, quite a few of the networks have focused on the so-called low hanging fruits, which are the soluble proteins, relatively easy to express, purify and crystallize. In target selection a strong emphasis has been put to disease- and drug-related proteins. For instance, structural genomics on *Mycobacterium tuberculosis* [106] and *Helicobacter pylori* [107] aims at developing improved structure-based drugs against these microbes. Another approach has been to study genomes from thermophilic organisms such as *Thermotoga maritima* [108] and *Thermus thermophilus* [109], which present the advantage of possessing highly stable and temperature resistant proteins with good crystallization properties. Moreover, a structural genomics network has been initiated on *Caenorhabditis elegans*, a worm that has served as a model organism in neurobiology and developmental biology [110]. The EU-funded SPINE (Structural Proteomics in Europe) consortium consists of 20 European partners and has set the goal to determine 500 structures of soluble proteins [111].

A number of networks also work on membrane proteins. American and Japanese networks have included membrane proteins, especially GPCRs, in their programs (Table 2). However, at least two networks are completely dedicated to structural genomics on membrane proteins. The EU funded network E-MeP studies 100 prokaryotic and 200 eukaryotic membrane proteins in a consortium consisting of 18 European research groups. Among the eukaryotic targets 100 are GPCRs and the rest non-GPCR proteins such as ion channels, transporters, efflux pumps and other integral membrane proteins. Within E-MeP, initial expression studies are carried out in *E. coli*, *L. lactis*, *S. cerevisiae*, *P. pastoris*, baculovirus-infected insect cells and SFV-infected mammalian cells. A limited number of targets are also expressed in a cell-free translation system. The privately funded Membrane Protein Network (MePNet) uniquely concentrates on GPCRs. Expression systems based on *E. coli*, *P. pastoris* and SFV have been utilized to over-express more than 100 GPCRs [60, 90, 112]. More than 60 GPCRs were expressed at structural biology compatible levels (1–10 mg/L) in one or several expression systems. Selected GPCRs refolded from *E. coli* inclusion bodies

have been subjected to crystallization attempts. Likewise, GPCRs purified from membranes of yeast and mammalian cells have been introduced into screens to find optimal conditions for crystallization.

## Conclusions and future prospects

The drop in success rate in drug development programs has generated plenty of concern worldwide. It is not only of economical interest but, also naturally of medicinal reasons that the pharmaceutical industry should be able to deliver more efficient and safer drugs. Structure-based drug discovery and design presents an interesting approach to optimize the drug efficacy and selective and thereby reduce serious side effects. This approach has already proven its feasibility as demonstrated for flu [17] and HIV [16] drugs. Structure determination has also become a routine tool in lead discovery and optimization [7]. Current initiatives in structural genomics will certainly have a major impact on the number new drug target structures that will become available for modelling and drug design purposes. The established networks in the field, which generally have a strong orientation towards technology development should also be able to improve the success rate for structure determination of membrane proteins. As membrane proteins represent more than 70% of current drug targets, it is very likely that structural biology therefore will strongly influence the future of drug discovery.

## References

1. **Lombardino JG, Love III JA.** The role of the medicinal chemistry in drug discovery – then and now. *Nat Rev Drug Discov.* 2004; 3: 853–62.
2. **Michel H.** <http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html>.
3. **Lundstrom K.** The future of G protein-coupled receptors as targets in drug discovery. *Idrugs.* 2005; 8: 909–13.
4. **Helenius A, Simons K.** Solubilization of membranes by detergents. *Biochim Biophys Acta.* 1975; 415: 29–79.
5. **Loll PJ.** Membrane protein structural biology: the high throughput challenge. *J Struct Biol.* 2003; 142: 144–53.
6. **Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima F, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M.** Crystal structure of rhodopsin: a G protein-coupled receptor. *Science.* 2000; 289: 739–45.
7. **Blundell TL.** Structure-based drug design. *Nature.* 1996; 384S: 23–6.
8. **Campbell SF.** Science, art and drug discovery: a personal perspective. *Clin Sci.* 2000; 99: 255–60.
9. **Blundell TL, Patel S.** High-throughput X-ray crystallography for drug discovery. *Curr Opin Pharmacol.* 2004; 4: 490–6.
10. **Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Chelliah V, Worth CL, Harmer NJ, Davies O, Burke D.** Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos Trans R Soc Lond B Biol Sci.* 2006; 361: 413–23.
11. **Blundell TL, Jhoti H, Abell C.** High throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov.* 2002; 1: 45–54.
12. **Nienaber VL, Richardson PL, Klighofer V, Bouska JJ, Giranda VL, Greer J.** Discovering novel ligands for macromolecules using X-ray crystallography screening. *Nat Biotechnol.* 2000; 18: 1105–8.
13. **Blundell TL, Abell C, Cleasby H, Hartshorn MJ, Tickle IJ, Parasini E, Jhoti H.** High throughput X-ray crystallography for drug discovery. In: D Flower, editor. *Drug design, cutting edge approaches.* London: Royal Society Chemistry; 2002. p. 53–9.
14. **Huth JR, Yu L, Collins I, Mack J, Mendoza R, Isaac B, Braddock DT, Muchmore SW, Comess KM, Fesik SW, Clore GM, Levens D, Hajduk PJ.** NMR-driven discovery of benzoylanthranilic acid inhibitors of far upstream element binding protein binding to the human oncogene c-myc promoter. *J Med Chem.* 2004; 47: 4851–7.
15. **Hardy LW, Malikayil A.** The impact of structure-guided drug design on clinical agents. *Curr Drug Discov.* 2003; 3: 15–20.
16. **Stoll V, Qin W, Stewart KD, Jakob C, Park C, Walter K, Simmer RL, Helfrich R, Bussiere D, Kao J, Kempf D, Sham HL, Norbeck DW.** X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg Med Chem.* 2002; 10: 2803–6.
17. **Varghese JN.** Development of neuroaminidase inhibitors as anti-influenza virus drugs. *Drug Dev Res.* 1999; 46: 176–96.
18. **Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J.** Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science.* 2000; 289: 1938–42.
19. **Kumar CC, Madison V.** AKT crystal structure and AKT-specific inhibitors. *Oncogene.* 2005; 24: 7493–501.
20. **Wallis RM, Corbin JD, Francis SH, Ellis P.** Tissue distribution of phosphodiesterase families and the effects of sildenafil on tissue cyclic nucleotides,

- platelet function, and the contractile responses of trabeculae carneae and aortic rings *in vitro*. *Am J Cardiol*. 1999; 83: 3C–12C.
21. **Katzen F, Chang G, Kudlicki W.** The past, present and future of cell-free protein synthesis. *Trends Biotechnol*. 2005; 23: 150–6.
  22. **Yokoyama S.** Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol*. 2003; 7: 39–43.
  23. **Klammt C, Löhr F, Schäfer B, Haase W, Dötsch V, Rüterjahn H, Glaubitz C, Bernhard F.** High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem*. 2004; 271: 568–80.
  24. **Ishihara G, Goto M, Saeki M, Ito K, Hori T, Kigawa T, Shirouzu M, Yokoyama S.** Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Prot Expr Purif*. 2005; 41: 27–37.
  25. **Pornillos O, Chen YJ, Chen AP, Chang G.** X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science*. 2005; 310: 1950–3.
  26. **Miroux B, Walker JE.** Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*. 1996; 260: 289–98.
  27. **Clough J, Saidijam M, Bettaney K, Szakonyi G, Patching S, Meuller J, Suzuki S, Shibayama K, Bacon M, Barksby E, Groves M, Herbert R, Phillip-Jones M, Ward A, Gunn-More F, O'Reilly J, Rutherford N, Bill R, Henderson P.** Prokaryotic membrane transport proteins: amplified expression and purification. In: K. Lundstrom, editor. Structural genomics on membrane proteins. Boca Raton: CRC Press; 2006. p. 21–42.
  28. **Tucker J, Grisshammer R.** Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem J*. 1996; 317: 891–9.
  29. **Weiss HM, Grisshammer R.** Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *Eur J Biochem*. 2002; 269: 82–92.
  30. **Grisshammer R.** Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol*. 2006; 17: 337–40.
  31. **Kiefer H.** *In vitro* folding of alpha-helical membrane proteins. *Biochim Biophys Acta*. 2003; 1610: 57–62.
  32. **Lopez de Maturana R, Willshaw A, Kuntzsch A, Rudolph R, Donnelly D.** The isolated N-terminal domain of the glucagon-like peptide-1 (GLP-1) receptor binds expending peptides with much higher affinity than GLP-1. *J Biol Chem*. 2003; 278: 10195–200.
  33. **Baneres JL, Martin A, Hullot P, Girard JP, Rossi JC, Parello J.** Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1. *J Mol Biol*. 2003; 329: 801–14.
  34. **Baneres JL, Mesnier D, Martin A, Joubert L, Dumuis A, Bockaert J.** Molecular characterization of a purified 5-HT4 receptor: a structural basis for drug efficacy. *J Biol Chem*. 2005; 280: 20253–60.
  35. **Kunji ER, Slotboom DJ, Poolman B.** *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochim Biophys Acta*. 2003; 1610: 97–108.
  36. **de Ruyter PG, Kuipers OP, de Vos WM.** Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl Environ Microbiol*. 1996; 62: 3662–7.
  37. **Monne M, Chan KW, Slotboom DJ, Kunji ER.** Functional expression of eukaryotic membrane proteins in *Lactococcus lactis*. *Protein Sci*. 2005; 14: 3048–56.
  38. **Palva I, Lehtovaara P, Kaariainen L, Sibakov M, Cantell K, Schein CH, Kashiwagi K, Weissmann C.** Secretion of interferon by *Bacillus subtilis*. *Gene*. 1983; 22: 229–35.
  39. **Akiba S, Hayashi Y, Hakamada Y, Endo K, Ara K, Kawai S, Saitoh E.** Extracellular production of human cystatin S and cystatin SA by *Bacillus subtilis*. *Protein Expr Purif*. 2006; 49: 203–10.
  40. **Luecke H, Schobert B, Cartailler JP, Richter HT, Rosengarth A, Needleman R, Lanyi JK.** Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin. *J Mol Biol*. 2000; 300: 1237–55.
  41. **Turner GJ, Reusch R, Winter-Vann AM, Martinez L, Betlach MC.** Heterologous gene expression in a membrane-protein-specific system. *Prot Expr Purif*. 1999; 17: 312–23.
  42. **Bartus CL, Jaakola VP, Reusch R, Valentine HH, Heikinheimo P, Levay A, Potter LT, Heimo H, Goldman A, Turner GJ.** Downstream coding region determinants of bacterio-opsin, muscarinic acetylcholine receptor and adrenergic receptor expression in *Halobacterium salinarum*. *Biochim Biophys Acta*. 2003; 1610: 109–23.
  43. **Melmer G.** Biopharmaceuticals and the industrial environment, discovery. In: G. Gellissen, editor. Production of recombinant proteins – novel microbial and eukaryotic expression systems. Weinheim: Wiley-VCH; 2005. p. 361–83.
  44. **Valenzuela P, Medina A, Rutter WJ, Ammerer G, Hall BD.** Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature*. 1982; 298: 347–50.
  45. **Cabezon T, De Wilde M, Herion P, Loriau R, Bollen A.** Expression of human alpha 1-antitrypsin cDNA in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*. 1984; 81: 6594–8.



46. **Hinnen A, Meyhack B, Tsapis R.** High expression and secretion of foreign proteins in yeast. In: M. Korhola, E. Väisänen, editors. *Gene expression in yeast: foundation for biotechnical and industrial fermentation research*. Helsinki: Kauppakirjapaino; 1983. p. 157–66.
47. **Sakai A, Shimizu Y, Hishinuma F.** Isolation and characterization of mutants which show an oversecretion phenotype in *Saccharomyces cerevisiae*. *Genetics*. 1988; 119: 499–506.
48. **David NE, Gee M, Andersen B, Naider F, Thorner J, Stevens RC.** Expression and purification of the *Saccharomyces cerevisiae* alpha-factor receptor (Ste2p), a 7-transmembrane-segment G protein-coupled receptor. *J Biol Chem*. 1997; 272: 15553–61.
49. **Andersen B, Stevens RC.** The human D1A dopamine receptor: heterologous expression in *Saccharomyces cerevisiae* and purification of the functional receptor. *Prot Expr Purif*. 1998; 13: 111–9.
50. **Sizmann D, Kuusinen H, Keranen S, Lomasney J, Caron MG, Lefkowitz RJ, Keinanen K.** Production of adrenergic receptors in yeast. *Recept Channels*. 1996; 4: 197–203.
51. **Jidenko M, Nielsen RC, Sorensen TL, Moller JV, le Maire M, Nissen P, Jaxel C.** Crystallization of a mammalian membrane protein overexpressed in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*. 2005; 102: 11687–91.
52. **Broker M, Bauml O.** New expression vectors for the fission yeast *Schizosaccharomyces pombe*. *FEBS Lett*. 1989; 248: 105–10.
53. **Sander P, Grunewald S, Maul G, Reilander H, Michel H.** Constitutive expression of the human D2S-dopamine receptor in the unicellular yeast *Saccharomyces cerevisiae*. *Biochim Biophys Acta*. 1994; 1193: 255–62.
54. **Presland J, Strange PG.** Pharmacological characterization of the D2 dopamine receptor expressed in the yeast *Schizosaccharomyces pombe*. *Biochem Pharmacol*. 1998; 56: 577–82.
55. **Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM.** Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*. 2005; 22: 249–70.
56. **Cereghino JL, Cregg JM.** Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev*. 2000; 24: 45–66.
57. **Thomas L, Crawford, DL.** Cloning of clustered *Streptomyces viridosporus* T7A lignocellulose catabolism genes encoding peroxidase and endoglucanase and their extracellular expression in *Pichia pastoris*. *Can J Microbiol*. 1998; 44: 364–72.
58. **Weiss HM, Haase W, Michel H, Reilander H.** Expression of functional mouse 5-HT5A serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization. *FEBS Lett*. 1995; 377: 451–6.
59. **Reinhart C, Kettler C.** Expression of membrane proteins in yeast. In: K. Lundstrom, editor. *Structural genomics on membrane proteins*. Boca Raton: CRC Press; 2006. p. 115–52.
60. **Lundstrom K, Wagner R, Reinhart C, Desmyter A, Cherouati N, Magnin T, Zeder-Lutz G, Courtot M, Prual C, André N, Hassaine G, Michel H, Cambillau C, Pattus F.** Structural genomics on membrane proteins - comparison of more than 100 GPCRs in 3 expression systems. *J Struct Funct Genomics*. 2006; 7: 77–91.
61. **Tornroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P.** Structural mechanism of plant aquaporin gating. *Nature*. 2006; 439: 688–94.
62. **Parcej DN, Eckhardt-Strelau L.** Structural characterization of neuronal voltage-sensitive K<sup>+</sup> channels heterologously expressed in *Pichia pastoris*. *J Mol Biol*. 2003; 333: 103–16.
63. **Long SB, Campbell EB, MacKinnon R.** Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science*. 2005; 309: 897–903.
64. **Luque T, O'Reilly DR.** Generation of baculovirus-expression vectors. *Mol Biotechnol*. 1999; 13: 153–63.
65. **Klaassen CHW, De Grip WJ.** Baculovirus expression system for expression and characterization of functional recombinant visual pigments. *Meth Enzymol*. 2000; 315: 12–29.
66. **Bosman GJ, De Grip WJ.** Expression of functional membrane proteins in the baculovirus-insect cell system: challenges and developments. In: K. Lundstrom, editor. *Structural genomics on membrane proteins*. Boca Raton: CRC Press; 2006. p. 153–67.
67. **Akermoun M, Koglin M, Zvalova-looss D, Folschweiller N, Dowell SJ, Gearing KL.** Characterization of 16 human G protein-coupled receptors expressed in baculovirus-infected insect cells. *Prot Expr Purif*. 2005; 44: 65–74.
68. **Benting J, Lecat S, Zacchetti D, Simons K.** Protein expression in *Drosophila* Schneider cells. *Anal Biochem*. 2000; 278: 59–68.
69. **Johanson K, Appelbaum E, Doyle M, Hensley P, Zhao B, Abdel-Meguid SS, Young P, Cook R, Carr S, Matico R, Donna Cusimano, Edward Dul, Monica Angelichio, Ian Brooks, Evon Winborne, Peter McDonnell, Thomas Morton, Donald Bennett, Theodore Sokoloski, Dean McNulty, Martin Rosenberg, Irwin Chaiken.** Binding interactions of human interleukin 5 with its receptor alpha subunit. Large scale production, structural, and functional studies of *Drosophila*-expressed recombinant proteins. *J Biol Chem*. 1995; 270: 9459–71.



70. **Perret BG, Wagner R, Lecat S, Brillet K, Rabut G, Pattus F.** Expression of EGFP-amino-tagged human mu opioid receptor in *Drosophila* Schneider 2 cells: a potential expression system for large-scale production of G-protein coupled receptors. *Prot Expr Purif.* 2003; 31: 123–32.
71. **Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H.** RTP family members induce functional expression of mammalian odorant receptors. *Cell.* 2004; 119: 679–91.
72. **Von Dannecker LE, Mercadante AF, Malnic B.** Ric-8B promotes functional expression of odorant receptors. *Proc Natl Acad Sci USA.* 2006; 103: 9310–4.
73. **Makrides SC.** Vectors for gene expression in mammalian cells. In: S.C. Makrides, editor. Gene transfer and expression in mammalian cells. Amsterdam: Elsevier Science B.V. 2003.p. 9–26.
74. **Walter CA, Humphrey RM, Adair GM, Nairn RS.** Characterization of Chinese hamster ovary cells stably transformed by a plasmid with an inducible APRT gene. *Plasmid.* 1991; 25: 208–16.
75. **Chelikani P, Reeves PJ, Rajbhandary UL, Khorana HG.** The synthesis and high-level expression of a beta2-adrenergic receptor gene in a tetracycline-inducible stable mammalian cell line. *Protein Sci.* 2006; 15: 1433–40.
76. **Prather PL, McGinn TM, Claude PA, Liu-Chen LY, Loh HH, Law PY.** Properties of a kappa-opioid receptor expressed in CHO cells: interaction with multiple G-proteins is not specific for any individual G alpha subunit and is similar to that of other opioid receptors. *Brain Res Mol Brain Res.* 1995; 29: 336–46.
77. **Reeves PJ, Kim JM, Khorana HG.** Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc Natl Acad Sci USA.* 2002; 99: 13413–8.
78. **Lundstrom K.** Expression of membrane proteins in mammalian cells. In: K. Lundstrom, editor. Structural genomics on membrane proteins. Boca Raton: CRC Press; 2006. p. 169–78.
79. **Jacobs SC, Stephenson JR, Wilkinson GW.** High-level expression of the tick-borne encephalitis virus NS1 protein by using an adenovirus-based vector: protection elicited in a murine model. *J Virol.* 1992; 66: 2086–95.
80. **Zhen Z, Bradel-Tretheway BG, Drewhurst S, Bidlack JM.** Transient overexpression of kappa and mu opioid receptors using recombinant adenovirus vectors. *J Neurosci Methods.* 2004; 136: 133–9.
81. **Drazner MH, Peppel KC, Dyer S, Grant AO, Koch WJ, Lefkowitz RJ.** Potentiation of beta-adrenergic signaling by adenoviral-mediated gene transfer in adult rabbit ventricular myocytes. *J Clin Invest.* 1997; 99: 288–96.
82. **Niles EG, Studier FW, Moss B.** Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci USA.* 1986; 83: 8122–6.
83. **Carroll MW, Wilkinson GWG, Lundstrom K.** Mammalian expression systems and vaccination. In: Ring C.J.A., Blair E.D., editors. Genetically engineered viruses. Oxford: BIOS Scientific Publishers Ltd; 2001. p. 107–57.
84. **Walker P, Munoz M, Combe MC, Grouzmann E, Herzog H, Selbie L, Shine J, Brunner HR, Waeber B, Wittek R.** High level expression of human neuropeptide Y receptors in mammalian cells infected with a recombinant vaccinia virus. *Mol Cell Endocrinol.* 1993; 91: 107–12.
85. **Bouvier M, Chidiac P, Hebert TE, Loisel TP, Moffett S, Mouillac B.** Dynamic palmitoylation of G-protein-coupled receptors in eukaryotic cells. *Methods Enzymol.* 1995; 250: 300–14.
86. **Gasmi M, Wong-Staal F.** Virus-based vectors for gene expression in mammalian cells: lentiviruses. In: Makrides S.C., editor. Gene transfer in mammalian cells. Amsterdam: Elsevier Science B.V.; 2003. p. 251–64.
87. **Yang M, Wang XG, Stout JT, Chen P, Hjelmeland LM, Appukuttan B, Fong HK.** Expression of a recombinant human RGR opsin in Lentivirus-transduced cultured cells. *Mol Vis.* 2000; 6: 237–42.
88. **Liljestrom P, Garoff H.** A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Biotechnology (N Y).* 1991; 9: 1356–61.
89. **Lundstrom K.** Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim Biophys Acta.* 2003; 1610: 90–6.
90. **Hassaine G, Wagner R, Kempf J, Cherouati N, Hassaine N, Prual C, André N, Reinhart C, Pattus F, Lundstrom K.** Semliki Forest Virus vectors for overexpression of 101 G Protein-coupled receptors in mammalian host cells. *Prot Purif Expr.* 2006; 45: 343–51.
91. **Lundstrom K, Michel A, Blasey H, Bernard AR, Hovius R, Vogel H, Surprenant A.** Expression of ligand-gated ion channels with the Semliki Forest virus expression system. *J Recept Signal Transduct Res.* 1997; 17: 115–26.
92. **Lundstrom K.** Semliki Forest virus vectors: versatile tools for efficient large-scale expression of membrane receptors. In: Kühne S.R., de Groot H.J.M., editors. Perspectives on solid state NMR in biology. Netherlands: Kluwer Academic Publishers; 2001. p. 131–9.
93. **Lee AG.** How lipids interact with an intrinsic membrane protein: the case of the calcium pump. *Biochim Biophys Acta.* 1998; 1376: 381–90.
94. **Keyes MH, Gray DN, Kreh KE, Sanders CR.** Solubilizing detergents for membrane proteins. In: S. Iwata, editor. Methods and results in crystallisation of membrane proteins. San Diego: IUL; 2003. p. 17–38.

95. **Byrne B, Jormakka M.** Solubilization and purification of membrane proteins, In: Lundstrom K., editor. Structural genomics on membrane proteins. Boca Raton: CRC Press; 2006. p. 179–98.
96. **Abola E, Kuhn P, Earnest T, Stevens RC.** Automation of X-ray crystallography. *Nat Struct Biol.* 2000; 7: 973–7.
97. **Tickle I, Shariff A, Vinkovic M, Yon J, Jhoti H.** High-throughput protein crystallography and drug discovery. *Chem Soc Rev.* 2004; 33: 558–65.
98. **Shuker SB, Hajduk PJ, Meadows RP, Fesik SW.** Discovering high-affinity ligands for proteins: SAR by NMR. *Science.* 1996; 274: 1531–4.
99. **Moore JM.** NMR techniques for characterization of ligand binding: utility for lead generation and optimization in drug discovery. *Biopolymers.* 1999; 51: 221–43.
100. **Powers R.** Applications of NMR to structure-based drug design in structural genomics. *J Struct Funct Genomics.* 2002; 2: 113–23.
101. **Xie X-Q.** Membrane protein NMR. In: Lundstrom K., editor. Structural genomics on membrane proteins. Boca Raton: CRC Press; 2006. p. 211–59.
102. **Engel A.** Electron microscopy and atomic force microscopy of reconstituted membrane proteins. In: Lundstrom K., editor. Structural genomics on membrane proteins. Boca Raton: CRC Press; 2006. p. 300–20.
103. **Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH.** Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol.* 1990; 213: 899–929.
104. **Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y.** Structural determinants of water permeation through aquaporin-1. *Nature.* 2000; 407: 599–605.
105. **Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K.** Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature.* 2003; 421: 127–8.
106. **Goulding CW, Perry LJ, Anderson D, Sawaya MR, Cascio D, Apostol MI, Chan S, Parseghian A, Wang SS, Wu Y, Cassano V, Gill HS, Eisenberg D.** Structural genomics of *Mycobacterium tuberculosis*: a preliminary report of progress at UCLA. *Biophys Chem.* 2003; 105: 361–70.
107. **Essen LO.** Structural genomics of “non-standard” proteins: a chance for membrane proteins? *Gene Funct Dis.* 2002; 3: 39–48.
108. **Canaves JM, Page R, Wilson IA, Stevens RC.** Protein biophysical properties that correlate with crystallization success in *Thermotoga maritima*: maximum clustering strategy for structural genomics. *J Mol Biol.* 2004; 344: 977–91.
109. **Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S.** Structural genomics projects in Japan. *Nat Struct Biol.* 2000; 7: Suppl. 943–5.
110. **Wang BC, Adams MW, Dailey H, Delucas L, Luo M, Rose J, Bunzel R, Dailey T, Habel J, Horanyi P, Jenney FE Jr, Kataeva I, Lee HS, Li S, Li T, Lin D, Liu ZJ, Luan CH, Mayer M, Nagy L, Newton MG, Ng J, Poole FL 2<sup>nd</sup>, Shah A, Shah C, Sugar FJ, Xu H.** Protein production and crystallization at SECSG—an overview. *J Struct Funct Genomics.* 2005; 6: 233–43.
111. [www.spineurope.org](http://www.spineurope.org)
112. **Andre N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, Pattus F, Michel H, Wagner R, Reinhart C.** Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci.* 2006; 15: 1115–26.