

Cell Engineering

Cell Engineering

Volume 5

Editor-in-Chief

Professor Mohamed Al-Rubeai
*University College Dublin,
Belfield, Dublin, Ireland*

Editorial Board

Dr Hansjorg Hauser
*GBF,
Braunschweig, Germany*

Professor Michael Betenbaugh
*Johns Hopkins University,
Baltimore, U.S.A.*

Professor Martin Fussenegger
*Swiss Federal Institute of Technology,
Zurich, Switzerland*

Dr Nigel Jenkins
*Serono Biotech Centre,
Fenil-sur-Corsier, Switzerland*

Dr Otto-Wilhelm Merten
*A.F.M.-Genethon 11,
Gene Therapy Program,
Evry, France*

The titles published in this series are listed at the end of this volume

CELL ENGINEERING

Vol. 5: Systems Biology

Edited by

Mohamed Al-Rubeai

University College Dublin, Ireland

and

Martin Fussenegger

*Swiss Federal Institute of Technology,
ETH Hoenggerberg HCI Institute of Chemical and Bioengineering ICB,
Zurich, Switzerland*

 Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-5251-0 (HB)
ISBN-13 978-1-4020-5251-4 (HB)
ISBN-10 1-4020-5252-9 (e-book)
ISBN-13 978-1-4020-5252-1 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© 2007 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

CONTENTS

List of Contributors	vii
1. Structural Genomics: A Special Emphasis on Membrane Proteins <i>Kenneth Lundstrom</i>	1
2. RNA Interference in Haematopoietic and Leukaemic Cells <i>Maria Thomas, Natalia Martínez Soria and Olaf Heidenreich</i>	29
3. Genomics and Proteomics of Chinese Hamster Ovary (CHO) Cells <i>Peter Morin Nissom and Robin Philp</i>	49
4. The Unfolded Protein Response <i>Natalie Strudwick and Martin Schröder</i>	69
5. Engineering of Cell Proliferation via myc Modulation <i>Vasiliki Ifandi and Mohamed Al-Rubeai</i>	157
6. The Molecular Response(s) during Cellular Adaptation to, and Recovery from, Sub-Physiological Temperatures <i>Sarah J Scott, Rosalyn J Marchant, Mohamed B Al-Fageeh, Michèle F Underhill, and C Mark Smales</i>	185
7. Molecular Response to Osmotic Shock <i>Susan T. Sharfstein, Duan Shen, Thomas R. Kiehl and Rui Zhou</i>	213
8. Metabolomics <i>Soo Hean Gary Khoo and Mohamed Al-Rubeai</i>	237
9. Metabolic Flux Analysis of Mammalian Cells <i>D.E. Martens</i>	275
10. Metabolic Engineering <i>Effendi Leonard, Zachary L. Fowler and Mattheos Koffas</i>	301

- | | |
|---|-----|
| 11. Chemical Organisation Theory | 361 |
| <i>Peter Dittrich and Pietro Speroni di Fenizio</i> | |
| 12. Prokaryotic Systems Biology | 395 |
| <i>Amy K. Schmid and Nitin S. Baliga</i> | |

LIST OF CONTRIBUTORS

- | | |
|--|--|
| Mohamed B. Al-Fageeh
University of Kent
Canterbury, UK | Mattheos Koffas
State University of New York at Buffalo
Buffalo, NY, USA |
| Mohamed Al-Rubeai
University College Dublin
Dublin, Ireland | Effendi Leonard
State University of New York at Buffalo
Buffalo, NY, USA |
| Nitin S. Baliga
Institute for Systems Biology
Seattle, WA, USA | Kenneth Lundstrom
Flamel Technologies
Vénisieux, France |
| Peter Dittrich
Friedrich Schiller University Jena,
Jena, Germany | Natalie Strudwick
University of Durham
Durham, UK |
| Zachary L. Fowler
State University of New York at Buffalo
Buffalo, NY, USA | Maria Thomas
Eberhard Karls University Tuebingen
Tuebingen, Germany |
| Olaf Heidenreich
Eberhard Karls University Tuebingen
Tuebingen, Germany | Michèle F. Underhill
University of Kent
Canterbury, UK |
| Vasiliki Ifandi
Warwick Medical School and University
of Warwick
Coventry, UK | Rui Zhou
Rensselaer Polytechnic Institute
Troy, NY, USA |
| Soo Hean Gary Khoo
University of Birmingham
Birmingham, UK | Natalia Martínez Soria
Eberhard Karls University Tuebingen,
Tuebingen, Germany |
| Thomas R. Kiehl
Rensselaer Polytechnic Institute
Troy, NY, USA | Peter Morin Nissom
Bioprocessing Technology Institute,
Singapore |

- | | |
|---|---|
| Robin Philp
Bioprocessing Technology Institute,
Singapore | Susan T. Sharfstein
Rensselaer Polytechnic Institute
Troy, NY, USA |
| Sarah J. Scott
University of Kent, Kent, UK | D.E. Martens
Wageningen University, Wageningen,
The Netherlands |
| Martin Schröder
University of Durham, Durham, UK | Amy K. Schmid
Institute for Systems Biology
Seattle, WA, USA |
| Rosalyn J. Marchant
University of Kent, Kent, UK | Pietro Speroni di Fenizio
Friedrich Schiller University Jena,
Jena, Germany |
| C. Mark Smales
University of Kent, Kent, UK | |

CHAPTER 1

STRUCTURAL GENOMICS

A Special Emphasis on Membrane Proteins

KENNETH LUNDSTROM

Flamel Technologies, 33, Avenue du Georges Lévy, 69693 Vénisieux, France

Abstract: Drug discovery based on structural knowledge has proven useful as several structure-based medicines are already on the market. Structural genomics aims at studying a large number of gene products including whole genomes, topologically similar proteins, protein families and protein subtypes in parallel. Particularly, therapeutically relevant targets have been selected for structural genomics initiatives. In this context, integral membrane proteins, which represent 60–70% of the current drug targets, have been of major interest. Paradoxically, membrane proteins present the last frontier to conquer in structural biology as some 100 high resolution structures among the 30,000 entries in public structural databases are available. The modest success rate on membrane proteins relates to the difficulties in their expression, purification and crystallography. To facilitate technology development large networks providing expertise in molecular biology, protein biochemistry and structural biology have been established. The privately funded MePNet program has studied 100 G protein-coupled receptors, which resulted in high level expression of a large number of receptors at structural biology compatible levels. Currently, selected GPCRs have been purified and subjected to crystallization attempts

Keywords: recombinant expression, purification, crystallization, structure-based drug design, structural genomics

1. INTRODUCTION

Classically the drug discovery process has relied on methods which involve selection of appropriate compounds for biological evaluation to define lead compounds¹ followed by evaluation of structure-activity relationships (SARs). Later, the development of methods for high throughput *in vitro* screening and the synthesis of combinatorial compound libraries have set the stage for drug design based on molecular targets and protein structures. As the requirements for drug safety and

efficacy have become more demanding and complex the costs for drug development have soared and the number of drug candidates even reaching the market has become very rare. The translation of *in vitro* efficacy to a desirable effect in animal models has been of major concern and even finally to obtain a drug which demonstrates the same profile in patients in clinical trials. There have been lengthy discussions related to the problems for pharmaceutical companies to fill their pipelines with novel innovative drug candidates. Cynically, it has been postulated that most major discoveries have already been made in drug development. On a more positive note, the sequencing of the human genome has revealed potential new drug targets. The advent of functional and structural genomics and their combination promises to become important tools in the discovery of novel cellular functions and pathways, which might uncover novel drug functions and open avenues for new medicines.

Structural genomics can be defined as the parallel structural characterization of a large number of gene products. This approach allows one to elaborate on the structure of protein subtypes and families, proteins with similar topology, i.e. membrane proteins, even whole genomes. Structural genomics therefore presents a great opportunity to improve the success in the discovery process of novel drugs. A drawback of studying a large number of protein targets is the need of large resources. For this reason, networks with expertise in various areas such as protein expression, purification and structure determination have been established. Several such consortia are described in this chapter. The MePNet project, which uniquely targets membrane proteins, is described in more detail. Moreover, a brief background in structure based drug design and examples of successful applications are presented. Today, one of the cornerstones for structural biology on therapeutically relevant proteins is recombinant protein production. The essential and most frequently used methods for expression are highlighted. Likewise, the procedures involved in protein purification and the various approaches for structure determination such as X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy-based techniques are described. As 60–70% of the current drugs are targeted to membrane proteins a special emphasis is dedicated to this type of proteins.

2. STRUCTURE-BASED DRUG DESIGN

Structure-based approaches in drug discovery have been frequently applied for lead optimization.² In this context, the drug potency and selectivity can be improved based on structural knowledge of proteins and their ligands.³ The significant impact structural biology has had on lead discovery can be measured by faster definition of drug binding properties and the easier identification of “hit” compounds through screening programs.⁴ Recent automated procedures such as AutoSolve® has enabled rapid structure resolution of protein-ligand complexes.⁵ For instance, in cocktails of as many as 100 molecules their different shapes can be distinguished by the variation in electron density.⁶ Another approach has been to use cocktails of smaller numbers of fragments at very high concentrations.⁷ Candidate fragments are ranked automatically, several datasets can be collected at a synchrotron hourly

and up to 1000 compounds screened in 2–3 days. In this approach a focused set of fragments from successful drug-like molecules is applied taken into account molecular weight (< 200 Da), the presence of hydrogen bond donors and acceptors and solubility.⁸ Also virtual screening can be applied, where large libraries of candidate fragments are systematically docked in the predefined binding site of the target protein in 3D computer models.⁵ Additionally, it was demonstrated that NMR-based screening can be used to design small molecule drug candidates to inhibit the aberrant overexpression of c-myc in a variety of tumors by inhibition of the far upstream element (FUSE) binding protein (FBP) expression.⁹

An interesting example of structure-based drug design comes from the program of synthesis of novel inhibitors of phosphodiesterase (PDE) to treat hypertension and other cardiovascular indications.³ Based on the X-ray structure, a rational drug design approach was initiated to develop a series of heterocyclic replacements for the parent ring system in zaprinast to obtain novel inhibitors of PDE. After intermediate steps, the synthesized compound UK 92480 (sildenafil) showed 100-fold increase in PDE5 inhibitory activity compared to zaprinast as well as unprecedented selectivity over other PDE enzymes.¹⁰ Although the pharmacological profile of sildenafil was as expected the clinical performance in patients with coronary heart disease was disappointing. However, further studies demonstrated that sildenafil blocked the PDE5 action which potentiated the natural activity of nitrose oxide and could be used for the treatment of erectile dysfunction (impotence).¹¹ Although originally developed for a completely different indication, the globally known blockbuster drug Viagra can therefore be considered as an example of how structural biology at least indirectly could facilitate the drug discovery process. Several other examples exist where structural knowledge has directly influenced drug design as briefly summarized below.

2.1 Examples of Successful Drug Design

Although structure-based drug design is a relatively new approach it has already established itself as a prominent alternative to classical drug screening. The major advantages of applying structural information are the shorter time required and the possibility to develop medicines with improved potency and selectivity. In addition to application in lead discovery many available protein structures have been used directly for the development of drug candidates. Today, 42 structure-based drug compounds have successfully entered clinical trials and at least 7 drugs have been approved on the market.¹² For instance, the AIDS drugs Agenerase® and Viracept® were designed based on the high resolution structure of the human immunodeficiency virus (HIV) proteinase.¹³ Similarly, the structure of the influenza virus neuraminidase was the basis for the flu drug Relenza®.¹⁴ The 3D structure of the kinase domain of c-Abl proved also helpful in designing and particularly avoiding resistance to the protein kinase drug Gleevec®.¹⁵

The number of structure-based drugs will certainly increase with time as more structures will be available. The study on kinase inhibitors already applies structure-based approaches and for example the crystal structure of the AKT kinase has

presented an attractive target for development of small molecule inhibitors as tumor therapeutics.¹⁶ However, the major boost in structure-based drug discovery is expected when finally structures of the largest topological group of drug targets, membrane proteins, will be available.

3. RECOMBINANT PROTEIN EXPRESSION

Very few therapeutically interesting drug target proteins are present in native tissues at such high levels that it is possible to isolate them in quantities that allow purification and further structural studies. Among membrane proteins, only the nicotinic acetylcholine receptor isolated from the electric organ of the *Torpedo marmorata* ray¹⁷ and the bovine rhodopsin from cow retina¹⁸ could be purified in quantities that allowed successful crystallization and structure determination. As a good example of the minute quantities of protein present in native tissue, the equivalent of 1000 pig brains was needed to purify 190 μg of the neuropeptide Y2 receptor for functional studies.¹⁹ Despite efforts of miniaturization, structural studies still require tens to hundreds of milligrams of purified protein. Naturally, in case of human proteins, ethical considerations will prevent any large scale purification efforts from native tissue. It has therefore been essential to establish robust recombinant protein expression systems to provide material for structural biology. In this context, all potential host cells including bacteria, yeast, insect and mammalian cells have been evaluated and expression vectors with a variety of properties engineered. Moreover, cell-free translation systems have been specifically developed for structural biology applications. The major expression systems are briefly presented below.

3.1 Cell-Free Translation

During the last few years the methods for cell-free protein expression have rapidly improved and made this approach a sincere alternative to cell-based recombinant protein expression methods.²⁰ Cell-free translation systems are currently commercially available based on both *E. coli* and wheat germ lysates. The advantages of cell-free expression are the possibility to directly use PCR fragments omitting any cloning procedures and the expression can be controlled in defined minimal media. Perhaps the most applied approach for cell-free translation is the simple amino-selective or uniform stable isotope labeling, which presents the opportunity of direct sample analysis by NMR.²¹

The strongest limitation of cell-free translation systems was for a long time the low expression levels obtained for membrane proteins. However, a cell-free translation system based on a modified *E. coli* S30 extract has provided significant improvement in expression levels also of membrane proteins.²² For instance, two bacterial membrane proteins, multidrug transporter TehA and YfiK, were expressed at levels of 2.7 mg/mL. Likewise, such GPCRs as the human β_2 adrenergic receptor (β_2 -AR), the human muscarinic acetylcholine M2 receptor (M2R) and the rat

neurotensin receptor (NTR) were evaluated for cell-free expression.²³ To obtain functional binding of the GPCRs the β 2-AR had to be fused to $G\alpha_s$,²⁴ the M2R to $G\alpha_i$ ²⁵ and the NTR to maltose binding protein (MBP).²⁶ The M2R had additionally the glycosylation sites and the third intracellular loop deleted.

3.2 Prokaryotic Expression

Prokaryotic expression has been the standard procedure for the expression of recombinant proteins. *E. coli*-based expression systems are definitely the most frequently used approach.²⁷ Alternative bacterial hosts have been verified as summarized below. The obvious advantage of prokaryotic expression is the ease of application and low production costs even at large scale level. The success rate has been very high for bacterial proteins, but also a large number of simple, especially soluble eukaryotic proteins, have been successfully expressed in prokaryotic systems. However, quite a few eukaryotic proteins require special post-translational modifications, which cannot be achieved in bacterial hosts. Whether these functions are necessary depends to a large extent on the further use of the produced material.

3.2.1 *E. coli*

The popularity of *E. coli* as an expression host stems from the ease to use and scale up of the system, the cheap production and high safety standards. Traditionally, *E. coli*-based expression has been reliable for soluble cytoplasmic proteins especially when applying fusion constructs.²⁸ Novel high throughput methods for *E. coli* expression including rapid cloning systems such as Gateway vectors²⁹ are reviewed elsewhere.³⁰ Recombinant protein secretion has also become feasible from Gram-negative bacteria, where the heterologously expressed protein is targeted to the periplasmic space or secreted into the culture medium.³¹

Expression of membrane proteins in *E. coli* has, however, been more difficult. Two approaches have been taken where the recombinant protein is either targeted to the bacterial membrane or accumulated in inclusion bodies as aggregates. In the former case, the success has been relatively good for bacterial membrane transport proteins and receptors.³² In contrast, expression of eukaryotic membrane proteins has been more difficult and the yields have been disappointingly low due to the toxicity to the host bacteria imposed by the foreign protein inserted in the membrane. However, major engineering involving evaluation of different deletions, fusion partners and tags has resulted in significantly improved expression of GPCRs. In this context, the fusion of the maltose binding protein (MBP) to the N-terminally deleted rat neurotensin receptor (NTR) resulted in milligram quantities of receptor protein.²⁶ Similarly, when the C-terminally truncated human adenosine A2a receptor was expressed as an MBP fusion protein 10–20 nmol receptor/L was produced in *E. coli* inner membranes.³³ Alternatively, the expression in bacterial inclusion bodies can generate high yields, but as the recombinant protein is present in aggregates refolding is mandatory to restore functional activity. The refolding process

has generally been difficult and inefficient for membrane proteins.³⁴ Recently, technology improvement has provided some promising results for the glucagon-like peptide 1 receptor (GLP-1),³⁵ the human leukotriene B4 receptor BLT1³⁶ and the serotonin 5-HT4 receptor.³⁷

3.2.2 *Lactococcus lactis*

In search of alternative expression hosts of prokaryotic origin, the Gram-positive bacterium *Lactococcus lactis* has been applied for the expression of both prokaryotic and eukaryotic recombinant proteins.³⁸ *L. lactis* grows rapidly to high densities and does not require aeration. Isotope labeling is feasible as most strains are auxotrophs for multiple amino acids. Transformation methods have been developed for *L. lactis* and for instance the commonly used nisin NisA promoter with the NisR and NisK regulatory trans-acting factors represents a tightly regulated system.³⁹ The *L. lactis* system has demonstrated high expression levels of such prokaryotic membrane proteins as ABC transporters and MSF (major facilitator subfamily) efflux pumps.³⁸ For the yeast mitochondrial carrier proteins CTP1 and AAC3 with a 6 TM topology yields up to 5% of total protein were obtained. However, attempts to express the human KDEL (Lys-Asp-Glu-Leu) receptor resulted only in low levels (< 0.1% of total membrane protein). In another study, 11 yeast mitochondrial transporter proteins were expressed in *L. lactis* in a functional form.⁴⁰ The expression levels were 10-fold higher when a lactococcal signal peptide was used or the N-terminus of the transporters was deleted. The expression levels are now compatible with amounts needed for structural biology. Preliminary data for GPCR expression in *L. lactis* also looks promising (Kunji, personal communication).

3.2.3 *Other prokaryotes*

There is a relatively large number of bacterial species (Table 1), which have been evaluated for homologous and heterologous gene expression. For instance, *Bacillus subtilis* was already in the 1980s considered as a potentially efficient host for recombinant protein production such as interferon⁴¹ and truncated viral membrane proteins.⁴² The drawback with this approach was that the recombinant proteins were heavily degraded by bacterial proteases. However, the expression of bacterial proteins such as α -amylase from *Bacillus amyloliquefaciens* resulted in secretion of large quantities of active enzyme.⁴³ Recently, novel plasmid vectors inducible by IPTG have been engineered for high level intra- and extracellular expression in *B. subtilis*.⁴⁴

The characteristic feature of the Archaeobacterium *Halobacterium salinarum* is its purple color, which relates to the accumulation of bacterio-opsin protein (Bop) complex formed with the chromophore retinol.⁴⁵ Attempts have been made to employ *H. salinarum* for heterologous gene expression. Fusion constructs of C-terminally tagged Bop to *E. coli* aspartate transcarbamylase (AT), human muscarinic M1 receptor, human serotonin 5-HT2 receptor and yeast α mating factor receptor Ste2 have been engineered.⁴⁶ The significance of the Bop sequence was evident as introduction of tags in this area reduced both mRNA and protein levels substantially. The Bop-AT fusion protein yields were 7 mg/L. In contrast,

Table 1. Structural Genomics Networks

Network	Expression	Targets
Berkeley Structural Genomics Center (BSGC) www.strgen.org	<i>E. coli</i>	<i>Mycoplasma genitalium</i> <i>Mycoplasma pneumoniae</i>
Center for Eukaryotic Structural Genomics (CESG) www.uwstructuralgenomics.org	<i>E. coli</i>	<i>Arabidopsis thaliana</i>
European Membrane Proteins (E-MeP) www.e-mep.org	<i>E. coli</i> <i>L. lactis</i> <i>S. cerevisiae</i> <i>P. pastoris</i> Baculo, SFV	prokaryotic MPs 100 GPCRs ion channels transporters other eukaryotic MPs
Joint Center for Structural Genomics (JCSG) www.jcsg.org	Cell-free, <i>E. coli</i> Baculo Adenovirus	<i>Thermotoga maritima</i> mouse genome human GPCRs
Membrane Protein Network (MePNet) www.mepnet.org	<i>E. coli</i> <i>P. pastoris</i> SFV	100 GPCRs
Membrane Protein Platform (MPP) www.swegene.org	<i>E. coli</i> <i>S. cerevisiae</i> <i>P. pastoris</i>	bacterial and yeast MPs human GPCRs
Midwest Center for Structural Genomics (MCSG) www.mcsg.anl.gov	<i>E. coli</i>	all 3 kingdoms of life
Northeast Structural Genomics Consortium (NSGC) www.nigms.nih.gov/Initiatives/PSI/Centers/NECSG.htm	<i>E. coli</i> , yeast insect cells	small proteins from yeast, <i>C. elegans</i> and <i>D. melanogaster</i>
New York Structural Genomics Research Consortium (NYSGRRC) www.nysgrc.org	<i>E. coli</i> yeast	bacterial, yeast and <i>C. elegans</i>
Paris-Sud Yeast Structural Genomics (YSG) www.genomics.eu.org	<i>E. coli</i>	250 non-membrane yeast proteins
Protein Structure Factory (PSF) www.proteinsturkturfabrik.de	<i>E. coli</i> <i>S. cerevisiae</i> <i>P. pastoris</i>	medically and biotechno-logically relevant proteins
Protein Wide Analysis of Membrane Proteins (ProAmp) www.pst-ag.com	<i>E. coli</i>	<i>S. typhimurium</i> and <i>H. pylori</i> MPs
RIKEN Structural Genomics Initiative (RSGI) www.rsgi.riken.go.jp/rsgi_e/index.html	Cell-free <i>E. coli</i>	<i>A. thaliana</i> , <i>T. thermophilus</i> and mouse proteins

(continued)

Table 1. (Continued)

Network	Expression	Targets
Southeast Collaboratory for Structural Genomics (SECSG) www.secsg.org	<i>E. coli</i> Baculo Lentivirus	<i>Pyrococcus furiosus</i> , <i>C. elegans</i> , human MPs
Structural Proteomics in Europe (SPINE) www.spineurope.org	<i>E. coli</i> Baculo	proteins/protein complexes of relevance to human health and disease
Structural Genomics Consortium (SGC) www.sgc.ox.ac.uk	<i>E. coli</i> Baculo	targets related to cancer, diabetes and malaria
Structure 2 Function Project (S2FP) www.s2fp.carb.nist.gov	<i>E. coli</i>	Haemophilus influenzae proteins
Swiss National Center for Competence in Research (NCCR) www.structuralbiology.ethz.ch	<i>E. coli</i> Baculo	bacterial MPs, transporters and human GPCRs
TB Structural Genomics Consortium (TBSGC) www.mbi-doe.ucla.edu/TB	<i>E. coli</i>	<i>Mycobacterium tuberculosis</i> proteins

MPs, membrane proteins

the fusions to the human GPCR constructs demonstrated no expression evaluated by immunoblotting. However, the Ste2 receptor showed a positive signal in Western blots, but the expression levels were much lower than for Bop-AT. In another study, the human muscarinic M1 and adrenergic α 2B receptors were expressed from Bop fusion vectors.⁴⁷ Membranes isolated from cells expressing the Bop-M1 fusion showed no specific signal, whereas membranes from Bop- α 2B cells were specifically recognized by a BR polyclonal antibody. Not unexpectedly, no binding activity was detected for Bop-M1. In contrast, specific binding was observed for Bop- α 2B, albeit 10 times weaker than obtained in yeast or mammalian cells.

3.3 Expression in Yeast

Yeast represents a good model organism of eukaryotic origin and has frequently been used as host for recombinant protein production. The advantages of yeast are the simple culture and scale-up methods, easy genetic manipulation, the cheap production and yeast cells possessing the machinery for eukaryotic post-translational modifications. Different types of yeast vectors have been engineered, both for episomal expression and chromosomal integration. Yeast vectors have been

frequently used for topologically different proteins such as cytoplasmic, membrane and secreted proteins and production of FDA-approved insulin⁴⁸ and hepatitis B surface antigen (HBsAg)⁴⁹ as pharmaceuticals. Typically, yeast and especially *Saccharomyces cerevisiae* has a strong tendency for hyperglycosylation of heterologous proteins, which could induce immunogenic or allergenic reactions if the product is aimed at therapeutic use. The fission yeast *Schizosaccharomyces pombe*⁵⁰ and the methylotrophic yeast *Pichia pastoris*⁵¹ have also been developed into efficient expression systems as highlighted below. Among other yeast strains the methylotrophic *Hansenula polymorpha* and the dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica* have not yet been much explored, but might present appropriate alternatives for future recombinant protein expression.⁵²

3.3.1 *Saccharomyces cerevisiae*

Saccharomyces cerevisiae is well characterized and its whole genome has been sequenced⁵³ so it is no surprise that the Baker's yeast has been used frequently for heterologous gene expression.⁵⁴ Among the many different proteins expressed from *S. cerevisiae* can be mentioned hepatitis B surface antigen (HBsAg),⁵⁵ α 1-Antitrypsin⁵⁶, human interferon- α ⁵⁷ and β -Endorphin.⁵⁸ A variety of yeast promoters have been used,⁵⁴ among which two inducible systems apply GAL1 and CUP1.⁵⁹

S. cerevisiae has also commonly been applied for the expression of membrane proteins. In this context, C-terminal FLAG- and His-tags were engineered to the yeast α -factor Ste2p receptor and expressed at milligram yields.⁶⁰ Ste2p receptor activity could be obtained after metal affinity column purification and reconstitution in artificial phospholipid vesicles, but only after addition of solubilized yeast membranes. Also the human dopamine D1A receptor was expressed at high levels in *S. cerevisiae*.⁶¹ Applying metal affinity and immunoaffinity chromatography, the D1A receptor was purified to near homogeneity and showed after reconstitution [³H] SCH23390 antagonist binding. Additionally, large-scale fermentation cultures at the 15 L level resulted in approximately 40 pmol of β 2 adrenergic receptor per milligram, which corresponds to yields of 20–30 mg of functionally active receptor.⁶² Recently, the biotin-tagged rabbit sarcoplasmic-endoplasmic reticulum Ca²⁺-ATPase isoform 1a (SERCA1a) was expressed in *S. cerevisiae*.⁶³ Purification of SERCA1a was performed by avidin agarose affinity chromatography followed by HPLC filtration. The purified protein was successfully crystallized and data collected at a 3.3 Å resolution.

3.3.2 *Schizosaccharomyces pombe*

Due to certain disadvantages of *S. cerevisiae* such as plasmid instability and hyperglycosylation recombinant protein expression has also been carried out in the fission yeast *Schizosaccharomyces pombe*. For instance, vectors for chromosomal integration to obtain stable expression have been engineered for *S. pombe*.⁶⁴ The

glycosylation pattern in *S. pombe* is different from *S. cerevisiae*, which could be of advantage for recombinant protein expression.⁶⁵ The human blood coagulation factor XVIIIa was expressed in *S. pombe* from an alcohol dehydrogenase (ADH) promoter.⁶⁶

Expression vectors based on *S. pombe* have been applied on membrane proteins. In this context, the human dopamine D2 receptor was expressed from a thiamine-repressible *nmt1* promoter, which resulted in 14.6 pmol receptor/mg and localization at the plasma membrane.⁶⁷ The expression levels were 5 fold higher compared to *S. cerevisiae*. In contrast, the rat dopamine D2 receptor⁶⁸ and the human neurokinin-2 receptor⁶⁹ were only expressed at 1 pmol/mg levels.

3.3.3 *Pichia pastoris*

The methylotrophic yeast strain *Pichia pastoris* has demonstrated high efficacy as an expression host.⁵¹ The expression systems developed for *P. pastoris* are based on integration of the heterologous gene construct in the yeast genome and utilization of strong yeast promoters such as the alcohol oxidase (AOX) promoter.⁷⁰ A clear advantage of culturing *P. pastoris* is the possibility to establish extremely high growth density with OD₆₀₀ values up to 500 U/ml. More than 200 recombinant proteins have been expressed in *P. pastoris*, including bacterial, fungal, plant, insect, mammalian and viral proteins. For instance, an endoglucanase from *Streptomyces viridosporus* resulted in yields up to 2.5 g/L.⁷¹ Likewise, human fibrinogen was produced at 100 mg/L.⁷²

A relatively large number of membrane proteins have been expressed from *P. pastoris* vectors. The mouse serotonin 5HT5 receptor was the first GPCR to be expressed in *P. pastoris*.⁷³ Engineering of an α -factor signal sequence from *S. cerevisiae* resulted in significantly higher yields (22 pmol/mg). Similarly, when the human dopamine D2 receptor was expressed in *P. pastoris* the receptor density was as low as 1200 receptors/cell, which could be enhanced by 20-fold by the introduction of the α -factor signal sequence.⁷⁴ Improvement of expression levels by several folds was observed for the mouse 5-HT5A, the dopamine D2S and the β 2 adrenergic receptors by introduction of the biotinylation (biotin-tag) of the transcarboxylase from *Propionibacterium shermanii* at the C-terminal.⁷⁵ It is thought that the biotin-tag stabilized the receptor directly against degradation or through prevention of the unfolding protein response. The pharmacological profile of GPCRs expressed in *P. pastoris* was similar to native tissue.⁷⁴ However, due to differences in lipid composition the ligand affinities were generally lower in yeast than in mammalian cells. The overexpression of GPCRs resulted in localization of receptors to the endoplasmic reticulum and the Golgi apparatus in contrast to the translocation to the plasma membrane in native mammalian cells. High-level overexpression in mammalian cells has, however, also resulted in prominent retention of receptors in the endoplasmic reticulum.⁷⁶ Receptor levels could be significantly enhanced by addition of specific agonists or antagonists or other supplements to the growth medium during methanol induction.⁷⁷

Among other membrane proteins, the water channel aquaporins were expressed in *P. pastoris* with C-terminal His- and myc-tags.⁷⁸ Recently, the structure of the spinach aquaporin SoPIP2 was solved after purification of the recombinant channel expressed in *P. pastoris*.⁷⁹ A rat neuronal voltage-sensitive K⁺-channel was overexpressed in *P. pastoris*, which allowed purification and single particle imaging and reconstitution of 2D crystals by cryo-EM.⁸⁰ The high resolution structure of another mammalian voltage-dependent K⁺ channel from the *Shaker* family was resolved after heterologous expression in *P. pastoris*.⁸¹

3.4 Expression in Insect Cells

Insect cell-based recombinant expression has been applied for mammalian proteins to a large extent because of the similarity of insect and mammalian cells. Insect cells possess many of the post-translational modification mechanisms also characteristic for mammalian proteins. However, differences exist and for instance the N-glycosylation pattern in insect cells is simpler and of high mannose type.⁸² The engineering of baculovirus expression systems presented the real breakthrough in application of insect cells for recombinant protein expression.⁸³ In this context, recombinant proteins have been produced for pharmacological characterization and drug screening purposes, but especially baculovirus has been the second most used production system in structural biology after *E. coli*. Alternatively to baculovirus, stable inducible expression systems in *Drosophila* Schneider-2 cells have also been established.⁸⁴

3.4.1 Baculovirus

The most frequently used system for expression in insect cells is based on baculovirus (*Autographa californica*).⁸³ A number of modifications have been introduced to facilitate the cloning and virus production procedures. The robust expression obtained from baculovirus vectors have made them attractive for production of various types of human recombinant proteins.⁸⁵ Although baculovirus vectors were originally developed for transduction of insect cells, modified vectors with mammalian-specific promoters has allowed efficient expression in different mammalian cells.⁸⁶ This approach has, however, mainly served functional studies and drug screening programs as relatively high virus concentrations are required to established efficient transduction rates in mammalian cells, which make large-scale applications unfeasible. In contrast, the scale-up procedure in insect cells is straight forward and presents a sensible alternative to bacterial and yeast expression for production of large quantities of recombinant protein for structural studies. Scale-up has been conducted in several insect cell lines, preferentially in *Spodoptera frugiperda* (Sf9 and Sf21 cells), *Mamestra brassica* and *Trichoplusia ni* (High Five) cells.⁸⁷

Baculovirus vectors have been applied for the expression of several membrane proteins and especially GPCRs. For instance, rhodopsin was expressed at yields of 4-6 mg/L of which 80% represented functional receptor.⁸⁸ Optimal expression has

been obtained under the following conditions:⁸⁷ bioreactor cultures with control of stirring speed, oxygen supply, removal of CO₂ and ambient temperature; dividing cells are infected at a low MOI; cells are harvested before they start to disintegrate. In an attempt to improve folding and transport of membrane proteins to the plasma membrane the signal sequence from the influenza hemagglutinin gene was placed in front of the human β 2 adrenergic receptor, which resulted in approximately two-fold increase in receptor expression.⁸⁹ In another study, 16 human GPCRs were expressed in three baculovirus-infected cell lines and monitored by radioligand binding assays.⁹⁰ The expression levels varied considerably from 1 to 250 pmol/mg protein. Recently, the C-terminally His-tagged human histamine H1 receptor was expressed in Sf9 cells at levels of 30–40 pmol per 10⁶ cells.⁹¹

Non-GPCR membrane proteins such as ion channels and transporter proteins have also been successfully expressed in baculovirus-insect cell systems. For instance, the voltage-gated AKT1 K⁺ channel⁹² and the NaPi-2 and NaSi-1 cotransporters⁹³ demonstrated functional activity when expressed in insect cells. Furthermore, a membrane spanning domain of the cystic fibrosis transporter (CFTR)⁹⁴ and the Na⁺-K⁺-ATPase⁹⁵ revealed functional activity.

3.4.2 Other insect cell systems

In addition to baculovirus other insect cell-based expression systems have been developed. Typically, *Drosophila* Schneider cells have been applied for heterologous expression of cytosolic, membrane and secreted proteins.⁸⁴ The generation of stable Schneider SL-3 cell lines is relatively easy and cheap. To verify the capacity of post-translational modifications and protein transport in insect cells the VIP36 (vesicular integral membrane protein) was expressed in SL-3 cells.⁸⁴ In parallel, a truncated form of VIP36 served as the model for a secreted protein and annexin XIIIb was studied to evaluate myristoylation in SL-3 cells. The secreted truncated VIP36 was N-glycosylated and the N-glycan of the Golgi-localized full-length VIP36 was endo-H resistant. Moreover, annexin XIIIb was myristoylated suggesting that SL-3 cells can be considered as suitable hosts for mammalian proteins.

Drosophila cells were used for the large scale production of human interleukin 5 (IL5) and soluble forms of its receptor alpha subunit.⁹⁶ The deglycosylated form of IL5 was active and a 2.6 Å resolution crystal structure determined from the purified recombinant protein. Also GPCRs have been subjected to expression studies in *Drosophila* cells.⁹⁷ In this context, the human mu opioid receptor (hMOR) was expressed in Schneider 2 cells with an N-terminal EGFP-tag. The recombinant EGFP-hMOR showed a similar pharmacological profile as detected in mammalian cells and functionality was demonstrated by cAMP stimulation and [³⁵S] GTP γ S binding. Comparison of binding data and quantitative EGFP fluorescence intensity analysis indicated that a relatively large number of the receptors did not present high-affinity binding, which might be at least partly due to retention of the receptors in intracellular structures.

3.5 Expression in Mammalian Cells

The most native environment for expression of mammalian proteins is obviously mammalian host cells. Two strategies have been applied by performing either transient transfection or establishing stable cell lines.⁹⁸ Both approaches have their advantages and disadvantages. Transient transfection provides faster expression and generally higher yields, whereas although time-consuming and relatively labor-intensive stable expression presents the means of obtaining clones for long-term use. One of the bottlenecks for using mammalian expression has been the expensive cell culture components and the more demanding scale-up procedure compared to bacterial and yeast systems. The choice of promoter and cell line plays also a significant role in obtaining high levels of expression and it was demonstrated that the full-length CMV promoter was superior.⁹⁹

3.5.1 *Transient expression*

A large number of mammalian expression vectors have been engineered for transient expression in mammalian cell lines such as BHK-21, CHO-K1, COS-7 and HEK293.¹⁰⁰ Methods have been developed for large scale transient expression.¹⁰¹ Recently, CHO-K1-S suspension cells cultured in serum-free medium generated mg/L quantities of bioactive antibody.¹⁰²

A number of GPCRs have been transiently expressed in mammalian cells. For instance, the cholecystokinin (CCK) A receptor was expressed transiently in both COS cells¹⁰³ and HEK293 cells.¹⁰⁴ Other membrane proteins such as the Na⁺- and Cl⁻-coupled GABA transporter GAT-1 was expressed transiently in Ltk-cells,¹⁰⁵ the muI Na⁺-channel in HEK293 cells¹⁰⁶ and the GLUT1 and GLUT4 transporters in COS-7 cells.¹⁰⁷

3.5.2 *Stable cell lines*

Although popular, instability of established mammalian cell lines has been of major concern and has led to the engineering and use of inducible expression vectors.¹⁰⁸ Generally, the expression levels obtained from stable cell lines have been lower than in transient expression.⁹⁹ The choice of promoter and host cell line are important factors for optimal expression. In a study where four secreted proteins were stably expressed from different promoters in various cell lines, the highest expression levels were obtained from the myeloproliferative sarcoma virus (MPSV) LTR promoter in CHO-K1 cells.⁹⁹

Expression levels for membrane proteins have generally been relatively low in stable cell lines. Typically, the kappa opioid receptor was expressed at 266 fmol/mg in CHO cells.¹⁰⁹ More recently, improved levels in the range of 1 to 20 pmol have been obtained for certain GPCRs, although resulting in only moderate yields of 0.1 mg/L in large scale production. Exceptionally high levels of rhodopsin were produced in an inducible mutant HEK293 suspension cell line.¹¹⁰ In this case, up to 6 mg receptor was obtained per liter culture. In addition to GPCRs the serotonin transporter (SERT) protein has been expressed in stable cell lines.¹¹¹ For instance,

the cold-inducible pCytTS system based on the Sindbis virus replicase¹¹² generated 250,000 SERT receptors per cell, whereas using the tetracycline-inducible T-Rex system produced 400,000 copies per cell.¹¹¹

3.5.3 Viral vectors

Viral mammalian vectors have been frequently used for recombinant expression. Generally viruses possess strong promoters from which high levels of expression can be obtained. Although the host range varies, many viruses such as adenoviruses, alphaviruses, lentiviruses and vaccinia viruses can infect many different types of mammalian cells. As previously mentioned, also baculovirus vectors have been successfully used for transduction of mammalian cells.⁸⁶ The drawback of viral vectors, especially those naturally infecting mammalian cells is the obvious safety concerns related to their use. However, these issues have been thoroughly addressed by the engineering of deletion mutant and replication-deficient vectors.

Adenovirus vectors are most commonly used for gene expression *in vitro* and *in vivo*.¹¹³ Various replication-deficient and –competent adenovirus vectors have been engineered to express a large number of recombinant proteins. For example, when the nonstructural glycoprotein NS1 from tickborne encephalitis virus (TBEV) was expressed from a CMV major early-immediate promoter in a replication-deficient adenovirus vector up to 25% of the total protein was represented by NS1.¹¹⁴ Several GPCRs have been expressed from adenovirus vectors. For instance, the mu (MOR) and kappa opioid receptors (KOR) were expressed in CHO cells from replication-deficient adenovirus vectors as fusion proteins with GFP or as such.¹¹⁵ The pharmacological properties of the recombinant receptors were identical to observations in native tissue. In comparison to expression in stable cell lines, 3 fold higher expression levels (B_{\max} values of 3 pmol/mg) were obtained. Likewise, the β_2 adrenergic receptor was expressed in rabbit myocytes at a level of 4 pmol/mg receptor.¹¹⁶

Vaccinia virus expression systems have been engineered as hybrid vectors with bacteriophage RNA polymerases, especially applying T7 phages¹¹⁷ and as replication-deficient vaccinia virus vectors.¹¹⁸ Hundreds of foreign genes have been expressed from vaccinia virus vectors.¹¹⁹ Several GPCRs have been expressed in mammalian cells as recombinant proteins using vaccinia virus vectors. In this context, it was demonstrated that the neuropeptide Y (NPY) receptor was localized in the plasma membrane and saturation binding experiments suggested that 5–10 million binding sites existed per cell.¹²⁰ Also the human dopamine D2 and D4 receptors were expressed from vaccinia virus vectors in rat-1 cells in a functional form.¹²¹ The GABA transporter was expressed in HeLa cells infected with vaccinia virus demonstrating similar expression levels to transient expression.¹⁰⁵

Replication-deficient **lentivirus** vectors generally based on the human immunodeficiency virus 1 (HIV-1) have been widely applied to gene expression in cell lines and *in vivo* especially in cases where long-term expression is advantageous.¹²² The expression from lentivirus vectors is generally driven by an internal cassette as lentiviral promoters are inefficient in human cells and biosafety concerns require

control mechanisms.¹²³ Efficient reporter gene expression (GFP and lacZ) was obtained both in cell lines and *in vivo* from the constitutive CMV-1E promoter.^{122,124} Lentivirus vectors have also been used for the expression of membrane proteins. The human retinal pigment epithelium (RPE) retinal GPCR was expressed in COS-7 and ARPE-19 (retinal pigment epithelium) cells from replication-deficient vectors.¹²⁵ The expression was verified by immunodetection and [³H] all-*trans*-retinal binding showing 100 fold higher expression levels in ARPE-19 cells than in COS-7 cells. The long-term expression was demonstrated by stable expression up to 6 months. Tranzyme Pharma has recently developed the lentivirus-based TranzExpression Technology (Text™), which has been successfully evaluated for a number of GPCRs (www.tranzyme.com).

Semliki Forest virus (SFV), a single strand RNA virus with an envelope structure¹²⁶ has been commonly used for *in vitro* and *in vivo* heterologous gene expression.¹²⁷ SFV is particularly well suitable for structural genomics approaches due to its fast high-titer virus production and broad host range. A variety of topologically different proteins have been expressed from SFV vectors including cytoplasmic, membrane and secreted proteins.¹²⁸ SFV vectors have frequently been used for overexpression of GPCRs and ion channels and large-scale production has yielded up to 10 mg/L receptor for structural studies.¹²⁹ As described below 101 GPCRs were overexpressed with a high success rate from SFV vectors in the MePNet consortium.¹³⁰

4. PROTEIN PURIFICATION

A basic requirement for any structural characterization of recombinantly expressed proteins is purification to high homogeneity. For soluble proteins this procedure is straight forward whereas for membrane proteins more complicated steps including solubilization and separation of protein and lipid components are necessary.¹³¹ To facilitate the solubilization process detergents are applied. There are a number of different detergents and it seems that each target protein has to be screened for appropriate detergents.¹³² The use of detergents is reviewed elsewhere.¹³³ Various biochemical purification methods have been developed, which are briefly summarized below.

4.1 Affinity Tag Purification

The most commonly used approach for protein purification is based on immobilized metal affinity chromatography (IMAC). To facilitate the purification of recombinant proteins from the mixture of other cellular proteins different purification tags have been engineered into the constructs. Histidine (His) tags, which bind to chromatographic media charged with Ni²⁺ have been extensively applied¹³⁴ and has allowed a one-step purification procedure. Successful positioning of the His-tag at the N- or C-terminus or even within the recombinant protein sequence has been demonstrated by numerous examples.

In addition to His-tags an eight amino acid streptavidin binding sequence (Strep) has been applied and has proven excellent for large scale purification.¹³⁵ Also biotin-, FLAG- and hemagglutinin tags are possible alternatives.

4.2 Other Means

In addition to using tags for recombinant protein purification classical biochemical methods including ammonium sulfate precipitation and sucrose gradients are possible. However, these methods require large quantities of available material, which is not the case for membrane proteins. Ion exchange chromatography can also be applied for protein purification. Gel filtration or size exclusion chromatography, which separates the proteins based on their molecular weights, has been commonly used for purification purposes. Additionally, hydrophobic interaction and reverse-flow chromatography can be applied.

5. STRUCTURE DETERMINATION

Different approaches are possible for collecting structural information on proteins. The highest structure resolution can generally be obtained by x-ray crystallography, but the drawback of this approach is that high quality crystals are required. Although this approach is more or less a routine procedure for soluble proteins today it has only been successful for a limited number of membrane proteins. Other methods such as nuclear magnetic resonance (NMR) have therefore proven useful although limitations have been seen in relation to structure resolution and only recently improved methodology has opened NMR for larger proteins. Finally, structure determination can be exercised by atomic force microscopy and electron microscopy techniques, which can be applied on crude samples but on the other hand results in lower resolution.

5.1 X-Ray Crystallography

One of the cornerstones in structure determination is obviously X-ray crystallography, which has allowed obtaining structure resolution below 2.0 Å. The process includes crystallization of highly purified protein samples, measurement of crystal diffraction, solving phase determination problems, phase and electron density calculations and model building and refinement. The crystallization procedure has during the last few years experienced major development with the introduction of automation and miniaturization.¹³⁶ The crystallization process has been strongly facilitated by the reduced volumes (nanoliter). Also the development of high density crystallization microplates in 96 or higher format has contributed to the high throughput nature.¹³⁷ It has therefore become possible to optimize solution variables such as pH, ionic strength, temperature, and concentrations of salts and detergents. This approach has allowed the establishment of up to 100,000 crystallization trials per day. Important issues are the harvesting and storage of crystals to be analyzed at

synchrotron radiation facilities, data collection at beam lines and characterization of obtained crystals. It is anticipated that the miniaturization of the crystal screening process will result in production of smaller crystals and will require the use of micro-diffractometer technologies.

5.2 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) technologies have generally been seen as a complementary technology to X-ray crystallography.¹³⁸ Although the resolution has been poorer and the molecular weight range has set limitations, NMR has routinely been used for the identification and evaluation of chemical leads.¹³⁹ Furthermore, NMR requires extensive isotope labeling of the protein, which is expensive and time consuming. Recent development in probe technology, software and NMR methodology itself has made it possible to obtain high resolution structures and increase the size of the studied proteins. For instance, generation of iterative protein-ligand complexes has become feasible with NMR.¹⁴⁰ Novel solid state and solution NMR methods have further provided means for utilization of NMR in structural biology on membrane proteins.¹⁴¹

5.3 Electron Microscopy and AFM

Although the use of electron microscopy (EM) and atomic force microscopy (AFM) have not seen the same rapid progress as X-ray crystallography recently, these methods can be applied to extract atomic resolution information on proteins.¹⁴² Cryoelectron microscopy of reconstituted membrane proteins in 2D crystals has been conducted for bacteriorhodopsin¹⁴³ and aquaporin AQP1.¹⁴⁴ Although the resolution was at 3.5 Å it made it possible to define the atomic structure, subsequently confirmed by X-ray crystallography. Additionally, AFM can be applied on native and reconstituted membranes in aqueous solutions, which has allowed the monitoring of polypeptide loops. For instance, AFM has been used to study disc membranes of vertebrate photoreceptor rod outer segments, where rhodopsin harbors 50% of the surface space. It could be demonstrated that rhodopsin was present in dimers and higher oligomeric forms.¹⁴⁵

6. STRUCTURAL GENOMICS NETWORKS

A large number of both national and international networks have been established to facilitate the interaction between experts in different areas such as protein expression, purification and crystallography, required for structural biology. Working in large consortia allows parallel studies on many targets, which aids significantly in understanding issues related to protein expression and the feasibility of structural approaches. The demands on network coordination are, however, high as the information flow between the different scientists and institutes is essential for achievement of success. Table 1 presents a comprehensive list of existing networks.

6.1 Networks on Soluble Proteins

Many of the established networks have focused their activities entirely on the structural biology of soluble proteins. This strategy is fully understandable as the success rate of structure resolution is very high and a large number diffracting high quality crystals can be obtained and many structures can be solved within a short time period. Soluble proteins are also well expressed in *E. coli* and the purification procedure is relatively straight forward. Quite a few networks have also focused on thermostable proteins, which are often stable at conditions where other proteins are easily degraded facilitating the purification procedure significantly.

6.2 Networks on Membrane Proteins

The importance of membrane proteins as drug targets is reflected by the number of networks that have included membrane proteins in their target list. In many instances whole genomes are studied, which is the case for structural genomics programs on for example *Mycobacterium tuberculosis*¹⁴⁶ and *Thermus thermophilus*.¹⁴⁷ Alternatively, the targets may represent certain types of proteins or protein families. Among networks studying membrane proteins, the most popular targets are GPCRs and ion channels due to their importance in drug discovery. Networks such as the RIKEN Structural Genomics Initiative, The Joint Center for Structural Genomics, Swegene, and the Swiss National Center for Competence in Research all have GPCRs in their programs.

The EU-funded E-MeP (European Membrane Proteins) uniquely focuses on membrane proteins. Among the 100 prokaryotic targets are ABC transporters and other bacterial membrane proteins. E-MeP also studies 200 eukaryotic proteins, of which 100 are GPCRs and 100 ion channels, transporters and other integral membrane proteins. An initial expression evaluation of eukaryotic membrane proteins is carried out in *E. coli*, *L. lactis*, *S. cerevisiae*, *P. pastoris*, baculovirus and SFV.

6.2.1 MePNet – Structural genomics on GPCRs

As an example of a program uniquely dedicated to GPCRs, a brief summary of the MePNet program is presented.¹⁴⁸ MePNet was established in 2001 through private funding from more than 30 pharmaceutical and biotech companies interested in structural biology on GPCRs. The aim of the program was to subject 100 GPCRs to structural biology by developing technologies for expression, purification and crystallization in a high throughput format. The target GPCRs were selected based on ligand availability, representation of GPCR families and subtypes and relevance to human disease. Initially, expression evaluation was carried out in three expression systems based on *E. coli*, *P. pastoris* and SFV vectors. The expression in bacteria was targeted to inclusion bodies, whereas the GPCRs in yeast and mammalian cells were aimed at membranes. Expression in *E. coli* resulted in a success rate of 46% measured by immunodetection, whereas in yeast and mammalian cells

94% and 96%, respectively, of the GPCRs were successfully expressed. Overall, more than 60 different GPCRs were expressed at structural biology compatible levels, i.e. at least 1 mg/L at least in one of the three systems. Further expression optimization improved the binding activity for *P. pastoris* and SFV, leading to the highest binding values of 180 and 287 pmol/mg,¹³⁰ respectively. The yields for well-expressed GPCRs in *E. coli* were further improved in fermentor cultures resulting in up to 350 mg/L. A limited number of GPCRs have been purified and subjected to refolding attempts. Similarly, selected GPCRs from yeast and mammalian cells have been solubilized and purified. Crystallization attempts on the first purified GPCRs are in progress. MePNet launched in 2005 the second phase of its program (MePNet2), which will have a strong focus on crystallography. The MePNet teams are confident that several high resolution structures on GPCRs will be available in the near future.

7. CONCLUSIONS

In summary, structural genomics approaches have presented fruitful ways to provide quickly novel structural information. Studies on whole genomes, topologically defined types of proteins and protein families in parallel have provided invaluable information. Technology development has played a key role in this process leading to improved expression vectors and systems, as well as automation and miniaturization of purification and crystallization methods. However, systematic approaches have also required large resources and broad expertise and in this context it has been appropriate to establish structural genomics networks. As technology improvement has been achieved it has also become feasible to study membrane proteins. Not only do they represent some 30% of the genomes for various organisms, but they are also targets for 70% of current drugs and potentially interesting novel targets for new medicines.

8. FUTURE ASPECTS

Structure determination has been highly successful for a large number of soluble proteins in studies on individual proteins and in parallel studies on a large number of targets within structural genomics networks. From a historical aspect there is a strong similarity today for membrane proteins what was experienced for soluble proteins in the 1970s. At the beginning of that decade the first high resolution structures had been obtained and even in 1979 less than 100 structures were available. The dramatic technology improvement seen during the last two decades of the 20th century resulted in an exponential growth in the number of solved structures, reaching a number of over 30,000 today. If we manage to achieve a similar development of technologies for membrane proteins we might finally be able to soon conquer the last frontier in structural biology.

9. REFERENCES

1. Lombardino, J.G. & Lowe III, J.A. The role of the medicinal chemistry in drug discovery – then and now. *Nat. Rev. Drug Discov.* **3**, 853–62 (2004).
2. Blundell, T.L. Structure-based drug design. *Nature* **384S**, 23–6 (1996).
3. Campbell, S.F. Science, art and drug discovery: A personal perspective. *Clin. Sci.* **99**, 255–60 (2000).
4. Blundell, T.L. & Patel, S. High-throughput X-ray crystallography for drug discovery. *Curr. Opin. Pharmacol.* **4**, 490–6 (2004).
5. Blundell, T.L., Jhoti, H. & Abell, C. High throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **1**, 45–54 (2002).
6. Nienaber, V.L., Richardson, P.L., Klighofer, V., Bouska, J.J., Giranda, V.L. & Greer, J. Discovering novel ligands for macromolecules using X-ray crystallography screening. *Nat. Biotechnol.* **18**, 1105–8(2000).
7. Blundell, T.L., Abell, C., Cleasby, H., Hartshorn, M.J., Tickle, I.J., Parasini, E. & Jhoti, H. High throughput X-ray crystallography for drug discovery, in *Drug Design, Cutting Edge Approaches*, D. Flower, ed., Royal Society Chemistry, London, pp 53–9 (2002).
8. Scharff, A. & Jhoti, H. High-throughput crystallography to enhance drug discovery. *Curr. Opin. Chem. Biol.* **7**: 340–5 (2003).
9. Huth, J.R., Yu, L., Collins, I., Mack, J., Mendoza, R., Isaac, B., Braddock, D.T., Muchmore, S.W., Comess, K.M., Fesik, S.W., Clore, G.M., Levens, D. & Hajduk, P.J. NMR-driven discovery of benzoylanthranilic acid inhibitors of far upstream element binding protein binding to the human oncogene c-myc promoter. *J. Med. Chem.* **47**, 4851–7 (2004).
10. Wallis, R.M., Corbin, J.D., Francis, S.H. & 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.* **83**, 3C–12C (1999).
11. Goldstein, I., Lue, T.F., Padma-Nathan, H., Rosen, R.C., Steers, W.D. & Wicker, P.A. Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. *N. Engl. J. Med.* **338**, 1397–404 (1998).
12. Hardy, L.W. & Malikayil, A. The impact of structure-guided drug design on clinical agents. *Curr. Drug. Discov.* **3**, 15–20 (2003).
13. Stoll, V., Qin, W., Stewart, K.D., Jakob, C., Park, C., Walter, K., Simmer, R.L., Helfrich, R., Bussiere, D., Kao, J., Kempf, D., Sham, H.L. & Norbeck, D.W. X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg. Med. Chem.* **10**, 2803–6 (2002).
14. Varghese, J.N. Development of neuroaminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* **46**, 176–96 (1999).
15. Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B. & Kuriyan, J. 2000., Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938–42 (2000).
16. Kumar, C.C. & Madison, V. AKT crystal structure and AKT-specific inhibitors. *Oncogene* **24**, 7493–501 (2005).
17. Unwin, N. Structure and action of the nicotinicacetylcholine receptor explored by electron microscopy. *FEBS Lett.* **555**, 91–5 (2003).
18. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, F., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., & Miyano, M. 2000., Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**: 739–45 (2000).
19. Wimalawansa, S.J. Purification and biochemical characterization of the neuropeptide Y2 receptor. *J. Biol. Chem.* **270**, 18523–30 (1995).
20. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23**, 150–6 (2005).
21. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39–43 (2003).

22. Klammt, C., Löhr, F., Schäfer, B., Haase, W., Dötsch, V., Rüterjahns, H., Glaubitz, C. & Bernhard, F. High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271**, 568–80 (2004).
23. 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.* **41**, 27–37 (2005).
24. Guo, Z.D., Suga, H., Okamura, M., Takeda, S. & Haga, T. Receptor-G fusion proteins as a tool for ligand screening. *Life Sci.* **68**, 2319–27 (2001).
25. Furukawa, H. & Haga, T. Expression of functional M2 muscarinic acetylcholine receptor in *Escherichia coli*. *J. Biochem.* **127**, 151–61 (2000).
26. Tucker, J. & Grisshammer, R. Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem. J.* **317**, 891–9 (1996).
27. Miroux, B. & Walker, J.E. 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.* **260**, 289–98 (1996).
28. La Vallie, E.R. & McCoy, J.M. Gene fusion expression systems in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 501–6 (1995).
29. Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S. & Vidal, M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–92 (2000).
30. Hunt, I. From gene to protein: A review of new and enabling technologies for multi-parallel protein expression. *Prot. Expr. Purif.* **40**, 1–22 (2005).
31. Mergulhao, F.J., Summers, D.K. & Monteiro, G.A. Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* **23**, 177–202 (2005).
32. Saidijam, M., Bettaney, K.E., Szakonyi, G., Psakis, G., Shibayama, K., Suzuki, S., Clough, J.L., Blessie, V., Abu-Bakr, A., Baumberg, S., Mueller, J., Hoyle, C.K., Palmer, S.L., Butaye, P., Walravens, K., Patching, S.G., O'reilly, J., Rutherford, N.G., Bill, R.M., Roper, D.I., Phillips-Jones, M.K. & Henderson, P.J. Active membrane transport and receptor proteins from bacteria. *Biochem. Soc. Trans.* **33**, 867–72 (2005).
33. Weiss, H.M. & Grisshammer, R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *Eur. J. Biochem.* **269**, 82–92 (2002).
34. Kiefer, H. In vitro folding of alpha-helical membrane proteins. *Biochim. Biophys. Acta* **1610**, 57–62 (2003).
35. 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 exendin peptides with much higher affinity than GLP-1. *J. Biol. Chem.* **278**, 10195–200 (2003).
36. Baneres, J.L., Martin, A., Hullot, P., Girard, J.P., Rossi, J.C. & 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.* **329**, 801–14 (2003).
37. Baneres, J.L., 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.* **280**, 20253–60 (2005).
38. Kunji, E.R., Slotboom, D.J. & Poolman, B. *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta* **1610**, 97–108 (2003).
39. de Ruyter, P.G., Kuipers, O.P. & de Vos, W.M. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62**, 3662–7 (1996).
40. Monne, M., Chan, K.W., Slotboom, D.J. & Kunji, E.R. Functional expression of eukaryotic membrane proteins in *Lactococcus lactis*. *Protein Sci.* **14**, 3048–56 (2005).
41. Palva, I., Lehtovaara, P., Kaariainen, L., Sibakov, M., Cantell, K., Schein, C.H., Kashiwagi, K. & Weissmann, C. Secretion of interferon by *Bacillus subtilis*. *Gene* **22**, 229–35 (1983).
42. Lundstrom, K., Palva, I., Kaariainen, L., Garoff, H., Sarvas, M. & Pettersson, R.F. Secretion of Semliki Forest virus membrane glycoprotein E1 from *Bacillus subtilis*. *Virus Res.* **2**, 69–83 (1985).
43. Palva, I. Molecular cloning of alpha-amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*. *Gene* **19**, 81–7 (1982).

44. Phuong Phan, T.T., Nguyen, H.D., and Schumann, W. Novel plasmid-based expression vectors for intra- and extracellular production of recombinant proteins in *Bacillus subtilis*. *Prot. Expr. Purif.*, *in press*.
45. Luecke, H., Schobert, B., Cartailler, J.P., Richter, H.T., Rosengarth, A., Needleman, R. & Lanyi, J.K. Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin. *J. Mol. Biol.* **300**, 1237–55 (2000).
46. Turner, G.J., Reusch, R., Winter-Vann, A.M., Martinez, L. & Betlach, M.C. Heterologous gene expression in a membrane-protein-specific system. *Prot. Expr. Purif.* **17**, 312–23 (1999).
47. Bartus, C.L., Jaakola, V.P., Reusch, R., Valentine, H.H., Heikinheimo, P., Levay, A., Potter, L.T., Heimo, H., Goldman, A. & Turner, G.J. Downstream coding region determinants of bacterio-opsin, muscarinic acetylcholine receptor and adrenergic receptor expression in *Halobacterium salinarum*. *Biochim. Biophys. Acta.* **1610**, 109–23 (2003).
48. Melmer, G. Biopharmaceuticals and the industrial environment, discovery, in *Production of Recombinant Proteins – Novel Microbial and Eukaryotic Expression Systems*, G. Gellissen, ed., Wiley-VCH, Weinheim, pp 361–383 (2005).
49. Harford, N., Cabezon, T., Colau, B., Delisse, A.M., Rutgers, T. & De Wilde, M. Construction and characterization of a *Saccharomyces cerevisiae* strain (RIT4376) expressing hepatitis B surface antigen. *Postgrad. Med. J.* **63 Suppl 2**, 65–70 (1987).
50. Giga-Hama, Y. & Kumagai, H. Expression system for foreign genes using the fission yeast *Schizosaccharomyces pombe*. *Biotechnol. Appl. Biochem.* **30**, 235–44 (1999).
51. Macauley-Patrick, S., Fazenda, M.L., McNeil, B. & Harvey, L.M. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* **22**, 249–70 (2005).
52. Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J.M., Berardi, E., Veenhuis, M. & van der Klei, I. New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adenivorans* and *Yarrowia lipolytica* – A comparison. *FEMS Yeast Res.* **5**, 1079–96 (2005).
53. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S.G. Life with 6000 genes. *Science* **274**, 563–7 (1996).
54. Gellissen, G. & Hollenberg, C.P. Application of yeasts in gene expression studies: a comparison of *Saccharomyces cerevisiae*, *Hansenula polymorpha* and *Kluyveromyces lactis* – A review. *Gene* **190**, 87–97 (1997).
55. Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G. & Hall, B.D. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**, 347–50 (1982).
56. Cabezon, T., De Wilde, M., Herion, P., Loriau, R. & Bollen, A. 1984., Expression of human alpha 1-antitrypsin cDNA in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U S A* **81**: 6594–8 (1984).
57. Hinnen, A., Meyhack, B. & Tsapis, R. High expression and secretion of foreign proteins in yeast, in *Gene Expression in Yeast: Foundation for Biotechnical and Industrial Fermentation Research*, M. Korhola, E. Väisänen, eds., Kauppakirjapaino, Helsinki, pp 157–66 (1983).
58. Sakai, A., Shimizu, Y. & Hishinuma, F. Isolation and characterization of mutants which show an oversecretion phenotype in *Saccharomyces cerevisiae*. *Genetics* **119**, 499–506 (1988).
59. Wang, Z. Controlled expression of recombinant genes and preparation of cell-free extracts in yeast. *Methods Mol. Biol.* **313**, 317–32 (2005).
60. David, N.E., Gee, M., Andersen, B., Naider, F., Thorner, J. & Stevens, R.C. Expression and purification of the *Saccharomyces cerevisiae* alpha-factor receptor (Ste2p), a 7-transmembrane-segment G protein-coupled receptor. *J. Biol. Chem.* **272**, 15553–61 (1997).
61. Andersen, B. & Stevens, R.C. The human D1A dopamine receptor: heterologous expression in *Saccharomyces cerevisiae* and purification of the functional receptor. *Prot. Expr. Purif.* **13**, 111–9 (1998).
62. Sizmman, D., Kuusinen, H., Keranen, S., Lomasney, J., Caron, M.G., Lefkowitz, R.J. & Keinanen, K. Production of adrenergic receptors in yeast. *Receptors Channels* **4**, 197–203 (1996).