Expression, Purification and Characterization of Fungal and Viral Recombinant Proteins

Helena Vihinen

Department of Chemical Technology
Helsinki University of Technology
and
Institute of Biotechnology
University of Helsinki
Finland

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Chemical Technology for public examination and debate in Auditorium KE1 at Helsinki University of Technology (Espoo, Finland) on the 22nd of March, 2001, at 12 noon.

SUPERVISED BY:

Professor **Leevi Kääriäinen**
Institute of Biotechnology, University of Helsinki
and
Professor **Marja Makarow**
Institute of Biotechnology, University of Helsinki

REVIEWED BY:

Docent **Anu Jalanko**
Department of Human Molecular Genetics
National Public Heath Institute
Helsinki
and
Docent **Johan Peränen**
Institute of Biotechnology, University of Helsinki

OPPONENT:

Docent **Pertti Markkanen**
Biotechnology Program
Espoo-Vantaa Institute of Technology

ISSN 1239-9469
ISBN 951-45-9874-1
ISBN 951-45-9875-X (pdf)
Gummerus Kirjapaino Oy
Saarijärvi 2001

To my brother Harri

CONTENTS

Abbreviations

ACN	acetonitrile
AcMNPV	*Autographa californica* multiple nuclear polyhedrosis virus
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
BVS	baculovirus expression system
BHK	baby hamster kidney
BSA	bovine serum albumin
CD	circular dichroism
CHX	cycloheximide
CIAP	calf intestine alkaline phosphatase
CID	collision-induced dissociation
CPVI	cytopathic vacuoles type I
DOC	deoxycholate
DTT	dithiothreitol
ECL	enhanced chemiluminescence
ER	endoplasmic reticulum
ESI	electrospray ionization
FSC	fetal calf serum
Gnd-HCl	guanidinium hydrochloride
GPC	gel permeation chromatography
GST	glutathione-S-transferase
GT	guanylyltransferase
HPLC	high performance liquid chromotography
Hsp	heat shock protein
IEC	ion exchange chromatography
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-ß-D-thiogalacto-pyranoside
kDa	kilodalton
MALDI	matrix-assisted laser desorp-tion/ionization
MBP	maltose binding protein
m.o.i.	multiplicity of infection
MS	mass spectrometry
MT	methyltransferase
MVA	modified vaccinia virus Ankara
$NGFR_e$	ectodomein of rat nerve growth factor receptor
NMR	nuclear magnetic resonance
Nsp	nonstructural protein
NTA	nitrilotriacetic acid
ODV	occlusion derived virus
PBS	phosphate buffered saline
PDI	protein disulfide isomerase
PFU	plaque forming unit
p.i.	post infection
PMSF	phenylmethylsulfonyl fluoride
PPM	phosphopeptide mapping
PVDF	polyvinylidene difluoride
RPC	reversed phase chromatography
SC	synthetic complete
Sf9	*Spodoptera frugiperda* insect cell line
SFV	Semliki Forest virus
SIN	Sindbis virus
SRP	signal recognition particle
TCA	trichloroacetic acid
TGN	*trans* Golgi network
TLC	thin layer chromatography
TLE	thin layer electrophoresis
TM	tunicamycin
TOF	time-of-flight
Trx	thioredoxin
wt	wild type

Original publications

This thesis is based on the following original publications which are referred to in the text by their Roman numerals. Additional unpublished data will also be presented in the text.

I Jämsä, E., Holkeri, H., **Vihinen, H.**, Wikström, M., Simonen, M., Walse, B., Kalkkinen, N., Paakkola, J. and Makarow, M. (1995). Structural features of a polypeptide carrier promoting secretion of a ß-lactamase fusion protein in yeast. *Yeast* **11**, 1381-1391.

II Simonen, M., **Vihinen, H.**, Jämsä, E., Arumäe, U., Kalkkinen, N. and Makarow, M. (1996). The hsp150Δ-carrier confers secretion competence to the rat nerve growth factor receptor ectodomain in *Saccharomyces cerevisiae*. *Yeast* **12**, 457-466.

III Holkeri, H., Simonen, M., Pummi, T., **Vihinen, H.** and Makarow, M. (1996). Glycosylation of rat NGF receptor in the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* **383**, 255-258.

IV Ahola, T., Laakkonen, P., **Vihinen, H.** and Kääriäinen, L. (1997). Critical Residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *J. Virol.* **71**, 392-339.

V **Vihinen, H.** and Saarinen, J. (2000). Phosphorylation site analysis of Semliki Forest virus nonstructural protein 3. *J. Biol. Chem.* **275**, 27775-27783.

VI **Vihinen, H.**, Ahola, T., Tuittila, M., Merits, A. and Kääriäinen, L. (2001). Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. *J. Biol. Chem.* **276**, in press.

ABSTRACT

This work reports the production of recombinant yeast and viral proteins in a number of diverse *in vivo* model systems for enzymatic and structural studies. In the first part Hsp150Δ peptide, a derivative of the yeast (*Saccharomyces cerevisiae*) secretory heat-shock protein Hsp150, was investigated for its ability to act as a carrier in transporting the ectodomain of rat nerve growth factor ($NGFR_e$) out from the yeast cell. The Hsp150Δ-$NGFR_e$ fusion protein was efficiently secreted into the growth medium, where it constituted the majority of total secreted proteins. Inhibition experiments with purified Hsp150Δ-$NGFR_e$ showed that Hsp150Δ did not prevent $NGFR_e$ from folding into a ligand-binding conformation. Circular dichroism (CD) analysis revealed that the Hsp150Δ-carrier did not have any specific secondary structure, which was also suggested by NMR analysis of a synthetic polypeptide corresponding to the repetitive consensus sequence of subunit II of Hsp150. These findings suggest that Hsp150Δ can successfully act as a carrier for foreign proteins, such as $NGFR_e$, made and secreted by *S. cerevisiae*.

The second part of this study involved the expression and purification of an RNA animal virus, Semliki Forest virus (SFV), nonstructural proteins (Nsp1-4) using a number of *in vivo* protein expression systems. To ensure quantities large enough for structural and enzymatic studies of the Nsps, each of them was expressed either in bacteria (*Escherichia coli*) or in insect cells (Sf9). All the proteins were expressed in high quantities (10-100 mg/l), and purified by affinity and size exclusion chromatography under nondenaturing or denaturing conditions. Independent of the expression system used, all the partially purified Nsps aggregated and precipitated either upon concentration, dialysis, storing or thawing. No detergents were found that could alleviate the aggregation problem or assist in the purification process.

Despite the unsuccessful purification of Nsps for structural studies, the expression and partial purification of Nsp1 and Nsp3 permitted biochemical characterization of their enzyme activities and posttranslational modifications. Point mutational analysis of the Nsp1 methyltransferase domain revealed that residue His^{38} was essential for the guanylyltransferase activity of Nsp1. Furthermore, residues Asp^{64} and Asp^{90} were found to be important for the methyltransferase activity of Nsp1. Phosphorylation sites in Nsp3 were determinated by point mutational analysis, electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) mass spectrometry (MS) as well as by phosphopeptide mapping and Edman sequencing. A phosphorylated domain (aa 320-368) was located in the C-terminal, non-conserved region of Nsp3, where 12 serines and 4 threonines could be modified by phosphates. The phosphorylation of Nsp3 seemed not to affect the membrane association or the localization of Nsp3 in either transfected or infected cells. Furthermore, Nsp3 phosphorylation deficient mutant viruses were capable of replication in infected mammalian cells a similar manner to the wild type SFV, but their neuropathogenicity in adult mice was greatly reduced.

INTRODUCTION

Molecular biology offers technologies whereby proteins can be produced and purified easier and more efficiently than ever before. Using recombinant DNA techniques such as gene fusion it is possible to generate chimeric proteins, which are novel in structure and function. At the DNA level, amino acid coding regions are now routinely recombined to incorporate new functional domains, proteolytic cleavage sites, intra- and extracellular localization or targeting signals, stabilizing signals, and even amino acid sequences that facilitate protein purification. All of these directed mutations can be used to subtly or dramatically alter the properties (*e.g.*, solubility, electric charge, hydrophobicity, conformation, substrate binding and active sites) of the protein. Thus, protein engineering has become a powerful tool in molecular biology to investigate protein structure and function, in addition to production and purification of useful proteins (Sassenfeld, 1990). The main applications of recombinant proteins obtained by genetic engineering are in the medical therapeutic fields (*e.g.*, production of recombinant vaccines, and therapeutic proteins for human diseases), and medical diagnosis (*e.g.*, antigen engineering for poly- and monoclonal antibody production used in disease testing). Other areas where recombinant proteins are commonly utilized include enzymes for food and fiber production, testing food for microbial contamination and veterinary medicine (Nilsson *et al.*, 1992).

Besides producing valuable proteins for practical applications, the production of proteins using recombinant technology offers unprecedented possibilities in the basic research of protein function, structure, interaction, turnover, domain shuffling, and many other areas, which could ultimately lead to the development of novel protein applications. Most proteins are expressed in infinitesimal amounts in their native cells and tissues, and it is only by recombinant techniques that it is possible to produce amounts great enough for basic research or for practical uses. Therefore, the expression of engineered proteins in efficient heterologous protein expression system is integral to the production, and purification of many proteins of interest. Moreover, with the vast amounts of predicted protein sequences being generated by genomic research, the application of protein engineering will most certainly be used to better characterize many unknown protein functions, interactions, cellular locations and potential practical uses.

1. PRODUCTION SYSTEMS FOR RECOMBINANT PROTEINS

Although much of recombinant protein engineering is performed *in vitro*, the actual protein synthesis usually utilizes living prokaryotic or eukaryotic cells. Proteins produced *in vivo* can be expressed transiently, constitutively or by induction, with the last two requiring cell lines which stably maintain the genetic material coding for the protein. Another means consist of using live bioreactors, *i.e.*, live transgenic plants and animals, as recombinant protein production systems. The choice of production system depends on many factors including the original source of the protein and codon usage (*e.g.*, prokaryotic, mammalian, plant), posttranslational processing (*e.g.*, phosphorylation, proteolytic processing, glycosylation, acylation), its future use (*e.g.*, human therapy, plant pest control) and economic factors (Hodgson, 1993). However, production of any protein is a complex multistep cellular process involving gene transcription, translation, protein processing and localization. Thus, the expression of a recombinant protein in a heterologous host cell or organism is often problematic. For example, protein expression levels

can be low, or the purified protein may be insoluble or unstable. And although, certain precautions can be taken (*e.g.*, altering codon usage to match the host's cells), often trial and error is the only real method of operation.

1.1. Bacterial expression utilizing *E. coli*

Escherichia coli (*E. coli*) has long been the primary prokaryotic host for heterologous protein expression, therefore there is much information and technology regarding this system with many different proteins (Baneyx, 1999; LaVallie *et al.*, 1995). The advantages of *E. coli* include its relatively easy, rapid and inexpensive methods of growth, transformation and maintenance. Furthermore, most of the biochemical pathways of *E. coli* are understood in great detail, and its entire genome has been sequenced. Foreign proteins can be produced in *E. coli* in large amounts (5-50% of total protein). Nevertheless, prokaryotic cells such as *E. coli* are unable to perform some posttranslational modifications, which occur in eukaryotic cells (*e.g.*, glycosylation, phosphorylation and disulfidebond formation), and which may be crucial for a foreign protein to obtain an active form. The expression and accumulation of a foreign protein in *E. coli* may also cause aggregates of the protein to form (Lilie *et al.*, 1998; De Berbardez Clark, 1998). Sometimes the protein in these inclusion bodies can be refolded *in vitro* to produce functional protein, but this is not always possible and renaturation can be expensive and time consuming. Furthermore, proteolytic degradation of heterologous protein expression may be a potential obstacle (Murby *et al.*, 1996; Matsuo *et al.*, 1999). Even though *E. coli* may not be useful for all foreign protein production, it has been successfully utilized to produce many functional human proteins such as human growth hormone, proinsulin, interferon-gamma and antibody fragments (Patra *et al.*, 2000; Cowley and Mackin, 1997; Davis *et al.*, 1999; Plückthun, 1992).

1.1.1. Expression systems

Various *E. coli* expression systems are available commercially and in the public domain, which is shared within the scientific community. Each system offers different benefits for protein expression, detection, and purification, and should be considered according to the specific criteria and requirements each protein poses. The expression unit or cassette in which the foreign protein coding region will be cloned into should consist of a strong promoter (*e.g.*, T7, trc, lac, tac, PL, PR, phoA, ara, xapA, cad, recA), ribosomal binding site(s), and an efficient transcription terminator (LaVallie *et al.*, 1995). Tightly controlled transcription coupled with a strong and rapid transcription inducer (*e.g.*, IPTG, tryptophan, temperature, phosphatase starvation, arabinose, xanthosine, pH, nalidixic acid) are crucial for high protein expression. Other factors to be considered include convenient multiple cloning sites, appropriate codon usage, and plasmid copy number. Usually, high protein expression levels are desirable, but in cases where the protein may form insoluble inclusion bodies, or may be toxic to the bacterial host, lower and slower expression may be preferred. The ideal promoter for expression directs efficient transcription to allow high-level production, and is tightly regulated to minimize the metabolic burden and toxic effects of a foreign protein. Regulatable promoters are used to drive foreign protein expression to avoid the selection of non-expressing mutant cells during the cell growth phase. Promoters such as T3, SP6 and T7 are widely used and efficient promoters derived from *E. coli* bacteriophages.

The T7 RNA polymerase is very specific for its own promoter, and can transcribe RNA templates approximately five times faster than the *E. coli* RNA polymerase (Mertens *et al.*, 1995). The gene coding for T7 RNA polymerase is either lysogenic or it is transfected into the cell by the bacteriophage CE6. In DE3 *E. coli* cells the T7 RNA polymerase gene is

under the control of the *LacUV5* promoter, which is inducible by IPTG. The addition of IPTG to a culture of DE3 cells will induce the transcription of T7 RNA polymerase, which in turn transcribes any gene under the control of the T7 promoter. However, the basal level of T7 RNA polymerase activity will promote some transcription of the target gene in non-induced cells. More strict control mechanisms have been engineered to depress basal T7 RNA polymerase expression. One method is based on the addition of a regulatory *lacI* gene, which codes for a strong repressor of the *LacUV5* promoter (Dubendorff and Studier, 1991). Very low basal expression levels of target gene can be reduced even more by adding T7 lysozyme, yet high level expression can be achieved upon induction with IPTG. Moreover, the T7*lacO* expression vectors pBAT, pHAT and pRAT have been reported to be suitable for expression of toxic proteins in *E. coli* (Peränen *et al.*, 1996). In these vectors tight control is achieved by centering the *lac* operator 14 base pairs downstream from the major RNA transcription start point.

1.1.2. Solubility

The majority of foreign proteins produced in *E. coli* are cytosolic, but as mentioned above some proteins accumulate as aggregates called inclusion bodies. In these aggregates only a small fraction of the protein is folded correctly, and the rest is in a denatured and usually non-functional form (Geisow, 1991; Valax and Georgiou, 1993). The probability of any heterologous protein forming inclusion bodies correlates greatly with its charge average and turn forming residue fraction, whereas hydrophilicity and total number of residues do not appear to correlate with inclusion body formation (Wilkinson and Harrison, 1991). Inclusion bodies may consist of not only the foreign protein, but also of some host components such as T7 RNA polymerase, membrane components, 16S and 23S RNA and even plasmid DNA (Mitraki and King, 1989). In some cases the formation of inclusion bodies can be advantageous, since the expressed protein can be purified simply by washing and pelleting the aggregated bodies (Georgiou and Valax, 1999). However, often the aggregation of the protein is detrimental, therefore expression conditions must be optimized so that the possibility of producing soluble protein is increased. The maximized solubility of a protein can be achieved by optimizing several factors such as the expression temperature, *E. coli* strain used, pH of the growth medium, fusion partner the protein is fused with, rate and level of expression and co-expression with chaperones and foldases (Murby *et al.*, 1996; Zhang *et al.*, 1998; Weickert *et al.*, 1996). For example, fusions to ZZ (two domains derived from staphylococcal protein A) and thioredoxin have been reported to increase the solubility of insulin-like growth factor I and cytokines (Samuelsson *et al.*, 1994; LaVallie et al., 1993). Chaperone proteins mediate both the folding and rehabilitation of proteins (van Dyk *et al.*, 1989; Bukari and Zipser, 1973). The most extensively studied chaperons/foldases in *E. coli* are GroEL/GroES, which have been reported to assist in folding proteins into active forms (Goloubinoff *et al.*, 1989; Lee and Olins., 1992; Yasukawa *et al.*, 1995; Amrein *et al.*, 1995; Wall and Plückthun, 1995). Moreover, co-overproduction of thioredoxin has been reported to increase the solubility of insulin-like growth factor I as well as other vertebrate proteins expressed in *E. coli* (Yasukawa *et al.*, 1995). However, even if co-overexpression of chaperons/foldases (*e.g.*, GroES, GroEL, DnaK) is likely to be useful for some proteins, it is unlikely to provide a universal solution to the problem of inclusion formation (Hockney, 1994). Other factors affecting protein solubility include regulation of the expression vector and bacterial cell lysis methods. Protein inclusion bodies are usually solubilized using urea or guanidine hydrocloride (Gdn-HCl). After denaturation the protein must be renatured under conditions that favor their proper folding and native functions. Conditions that must be considered when

trying to refold a denatured protein involve optimizing the renaturation temperature (usually around 10°C), redox conditions, protein concentration, and addition of solubilizing additives such as polyethylene glycol. Soluble and active proteins have also been obtained with treatment with sarkosyl (N-dodecanoyl-sarcosinate, Frangioni and Neel, 1992).

1.1.3. Secretion

Secretion of foreign proteins to the medium can greatly facilitate their purification. However, secretion of a protein to the medium by Gram-negative bacteria, such as *E. coli*, requires translocation of the protein through the bacterial cell wall consisting of inner and outer membrane bilayers surrounding the periplasmic space. Only a limited number of *E. coli* strains are capable of secreting proteins from the cell (reviewed in Pugsley, 1993). Nevertheless, some genetically modified proteins with inserted secretion sequence signals, have been successfully secreted into the media (H⌀gset *et al.*, 1990; Wadensten *et al.*, 1991; Weiss *et al.*, 1994; Samuelsson *et al.*, 1994). More importantly, *E. coli* is capable of secreting genetically altered proteins into the periplasm, which may represent 20-40% of the total cellular volume. With this strategy the target protein can be separated from the cytoplasmic proteins and accumulate in a more oxidizing environment where disulfide-bond formation can occur (Missiakas *et al.*, 1993). Secretion into the periplasmic space has been successfully utilized in large scale production of biologically active antibody fragments (Plückthun, 1992).

1.2. Yeast (Saccharomyces cerevisiae)

Saccharomyces cerevisiae (bakers and brewers yeast) is generally considered as a safe organism for the production of foreign proteins, since it has been used in food production for thousands of years (Müller *et al.*, 1998). This unicellular eukaryote has been extensively studied at the biochemical, molecular and genomic levels, and its entire genome has been recently sequenced (Goffeau *et al.*, 1997). Like *E. coli*, yeast grows rapidly in relatively inexpensive and simple media, is easy to transform, and maintain (Buckholz and Gleeson, 1991). But unlike *E. coli*, yeast is an eukaryotic cell, and therefore expresses and processes proteins more similarly to higher eukaryotes. (Lin Cereghino and Cregg, 1999). The secretory pathway of yeast closely resembles that of mammalian cells, thus yeast is capable of many posttranslational modifications such as proteolytic processing, disulfide bond formation, and glycosylation (Eckart and Bussineau, 1996). Thus, yeast offers a eukaryotic protein production system that can retain the foreign protein intracellularly or the product can be secreted to the medium if fused to a carrier peptide. However, *S. cerevisiae* is unable to perform certain complex eukaryotic posttranslational modifications such as prolylhydroxylation and amidation. In addition, the glycosylation of yeast can differ from that of higher eukaryotes (Sudbery, 1996). Nevertheless, many foreign proteins have been produced on a large scale using *S. cerevisiae*, including human serum albumin, the hepatitis B viral surface antigen, insulin and hirudin (Goodly, 1993; Valenzuela *et al.*, 1982; Thim *et al.*, 1986; Ladisch and Kohlmann, 1992; Mendoza-Vega *et al.*, 1994).

1.2.1. Episomal vectors and chromosomal integration

Recombinant DNA which encodes foreign proteins can be expressed in *S. cerevisiae* as an episomal plasmid or integrated into its genome. Extra-chromosomal replicons are based either on plasmids containing autonomously replicating sequences, or on the native 2µ circle of *S. cerevisiae* (Romanos *et al.*, 1992). Most yeast expression vectors have been constructed from the multi-copy 2µ plasmid, which has also been engineered to replicate and be selectable in both *E. coli* and

yeast. Because the native 2µ circle is present in most *S. cerevisiae* strains at approximately 60-100 copies per haploid genome, and the plasmid is stable inherited, 2µ-based protein production can be very efficient (Futcher, 1988; Unternährer *et al.*, 1991). The most common selection markers are *LEU2*, *TRP1*, *URA3* and *HIS3* in corresponding amino acid (aa) uptake deficient mutant strains, which are auxotrophic for leucine, tryphophan, uracil and histidine, respectively. Selection can also be performed using *e.g.*, the antibiotic resistance element G418, a 2-deoxy-streptamine antibiotic (Jimenez and Davies, 1980). There are also yeast strains engineered to ensure plasmid maintenance irrespective of the culture conditions, for example, *ura3 fur1* -strains are nonviable on uracil minus media since they are blocked both in the *de novo* and salvage pathways of uridine 5'-monophosphate synthesis (Loison *et al.*, 1986; Napp and Da Silva, 1993). These strains can grow on uracil minus media only if they maintain a plasmid that contains the *URA3* gene, which complements the mutant allele.

Chromosomal integration offers a more stable alternative to episomal maintenance of plasmid DNA coding for foreign proteins. The most widely utilized technique to obtain integration is homologous recombination (reviewed in Romanos *et al.*, 1992). Integration vectors are similar to the above mentioned yeast plasmids with both *E. coli* and yeast selectable markers and bacterial replication sequences, but lack yeast replication sequences. Once the plasmid has been transformed into a yeast cell, usually in a linear form, it can not replicate, and is maintained only if it integrates into the yeast genome. Homologous recombination can occur by a single or double crossover, either as a multicopy integration into ribosomal DNA, or by transposable elements such as Ty or gamma (Fleer, 1992). Double cross-over vectors contain the foreign DNA and selection marker flanked by yeast DNA homologues to the 5' and 3' regions of the chromosomal DNA to be replaced. The frequency of transformation by this method is lower than episomal plasmid transformation, but it results in very stable, usually single copy transformants (Romanos *et al.*, 1992). Multicopy integration of heterologous genes usually employs integration into reiterated chromosomal DNA sequences that may be present from 20 to 140 repeats per haploid genome. Chromosomal integration can also be obtained by using Ty transposition vectors, which are analogous to the retroviral vectors of mammalian cells (Boeke *et al.*, 1988).

1.2.2. Promoter strength

Many different yeast transcriptional promoters are available for foreign protein expression in *S. cerevisiae* (Fleer 1992). Yeast promoters consist of at least three elements which regulate the efficiency and accuracy of the initiation of mRNA transcription. These elements are upstream activation sequences (UAS), the TATA-box and initiator elements (Romonos *et al.*, 1992). Numerous strong yeast promoters to be utilized are either inducible (*e.g.*, *GAL1*, *GAL7*, *GAL10*, *MET25*), or constitutive (*e.g.*, *CYC1*, *ADH1*, *TEF2*, *GPD* and *MF*α) (Mumberg *et al.*, 1994 and 1995). In the case of large scale protein production in yeast, it is crucial to be able to separate the cell growth phase from the protein production phase (Da Silva and Bailey, 1991). The promoter strength of three different promoters *i.e.*, *SUC2*, *PGK* and *GAL7* were studied when α-amylase was produced in 30 hour fed-batch yeast culture (Park *et al.*, 1993). According to the results the efficiency of the promoters appeared to be in the following order *PGK*> *GAL7*> *SUC2*, while plasmid stability and promoter strength appeared to be inversely correlated. Also expression vectors with a tetracycline and copper ion regulatable promoter systems have been developed (Gari *et al.*, 1997; Labbé and Thiele, 1999). Another approach to optimize gene expression in yeast utilizes transactivators such as mammalian steroid hormone receptors, which have been

reported to increase transcriptional activity (Fleer, 1992), as well as regulate the overproduction of the GAL4 protein (Schultz *et al.*, 1994).

1.2.3. Secretion of heterologous proteins

S. cerevisiae is capable of secreting heterologous proteins, but most of them have to be fused to an appropriate pre-pro-leader (Romanos *et al.*, 1992; Shuster, 1991). These carriers are relatively short amino acid sequences and include pre-pro-leaders from such proteins as α-factor, killer-toxin, bar protease, gp37, and Hsp150 (reviewed by Simonen, 1994). The carrier peptide may have a crucial role in the transport process from the ER to the Golgi, which has been suggested to be receptor- mediated (Balch *et al.*, 1994). The carrier may thus provide a positive secretion signal for the capture of fusion protein into vesicle budding sites at the ER membrane (Suntio *et al.*, 1999). The carrier may also assist the fusion partner to acquire a secretion-competent conformation (Simonen *et al.*, 1994; Simonen *et al.*, 1996; Hammond and Helenius, 1995).

Some proteins can be externalized in *S. cerevisiae* without a carrier. Human lipo-cortin-1 was first expressed as a fusion with the α-factor pre-pro-region but only 10% of the protein was secreted (Chung *et al.*, 1996). An expression system utilizing the inulinase signal peptide increased secreted lipocortin-1 5-fold, resulting up to 95% of the protein to be secreted into the medium (Chung *et al.*, 1996). Furthermore, fed-batch fermentation of the heterologous protein produced 2.1 g/l, representing more than 80% of the total extracellular protein (Chung *et al.*, 1999). Also, it has been observed that culture growth conditions can affect protein secretion (Rossini *et al.*, 1993). For example, ß-galactosidase, which is normally refractile to secretion, is secreted from yeast grown in a rich medium. ß-galactosidase secretion could be further increased by elevated growth temperatures.

1.2.3.1. Yeast secretory proteins as carriers

In *S. cerevisiae* the most commonly used carrier is the pre-pro-region of the α-factor (Fig. 1A and 1B). The pre-sequence (signal sequence) is cleaved upon translocation into the ER and the pro-region is cleaved by the Kex2 protease in the Golgi compartment (Julius *et al.*, 1984). However, the mechanisms which the α-factor pre-pro-region uses to confer

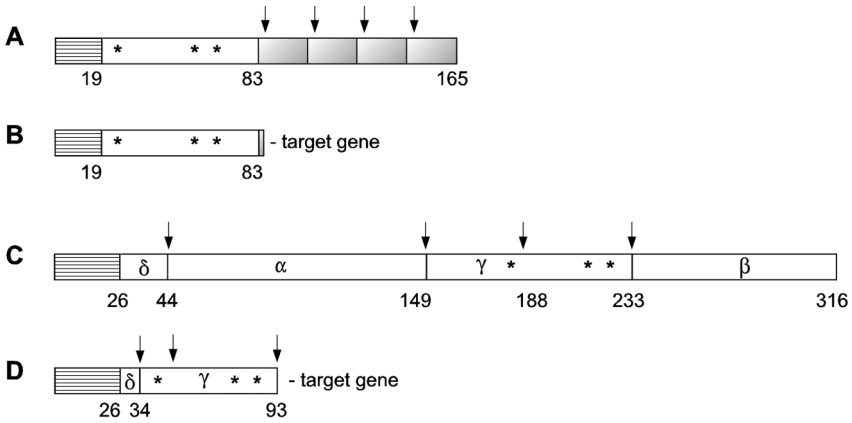

Figure 1. A Schematic presentation of primary translation products of S. cerevisiae mating α-factor (**A**), pre-pro-region of α-factor used as a carrier (**B**), killer toxin (**C**) and pre-pro-region of killer toxin used as a carrier (**D**). Pre-pro-killer toxin carrier comprises killer toxin residues 1-34, which are joined by alanine to the gamma-region residues 177-233. The Kex2 cleavage and N-glycosylation sites are indicated by arrows and asterisks, respectively. The signal sequence is marked by a horizontally striped area and the repetitive region of α-factor is colored gray.

secretion competence to the fusion partners, as well as its structure, are not known. Synthetic leaders have also been used as carriers (Kjeldsen *et al.*, 1997 and 1998b). Interestingly, it appears that three N-linked oligosaccharide chains are necessary for secretion competence of the α-factor pre-pro-leader, whereas synthetic pre-pro-leaders lacking the consensus N-glycosylation sites confers secretion competence of correctly folded insulin precursors (Kjeldsen *et al.*, 1998a and 1998b).

Another polypeptide used as a carrier in yeast is derived from the killer toxin (Fig. 1D). The K1 killer toxin is a heterodimer of two protein subunits (α and ß), maturated from the 316 residue pre-pro-toxin (Fig. 1C) (Bostian *et al.*, 1984). As in the pre-pro-region of α-factor, the signal sequence of the pre-pro-toxin is cleaved upon translocation into the ER and the pro-region is subsequently cleaved by the Kex2 protease in the Golgi compartment (Redding *et al.*, 1991). The ability of different fragments of the killer toxin pre-pro-region to confer secretion competence for heterologous protein was studied by fusing ß-lactamase to various fragments of pre-pro-toxin (Cartwright *et al.*, 1992). The most efficiently secreted construct had a carrier in which the genes of the ß-subunit and the control region of α were deleted (Fig. 1D). Secreted ß-lactamase reached level of several µg/ml which is equal to the level produced by the pre-pro-region of the α-factor (Cartwright *et al.*, 1994).

1.2.3.2. The secretion pathway of *S. cerevisiae*

The secretory pathway in eukaryotic cells is essential for posttranslational processing, targeting and transporting proteins to intra- and extracellular membranes and for proteins destined for export out of the cell. The secretory pathway of yeast is depicted in Figure 2. It consists of several membrane-bound compartments and vesicle-trafficing systems between these compartments. Protein secretion in yeast has been studied using temperature-sensitive secretion (*sec*) mutants, in which the transport of proteins can be reversibly blocked in distinct compartments (Kaiser and Schekman, 1990). Secretery proteins have a signal peptide, which is an endoplasmic reticulum (ER) targeting sequence present usually at their N-terminus, which allows the protein to enter the ER. Most posttranslational modifications, such as the initiation of glycosylation, disulfide bond formation and protein folding occur in the ER. Transport to the next compartment, the *cis* Golgi, occurs by vesicular transport, which is also the way proteins are transported through the Golgi (intra-Golgi) to the *trans* Golgi network (TGN) (Rexach and Sheckman, 1991). Transport, docking, and fusion of transport vesicles are regulated by proteins generally refered to as SNAREs (reviewed by Gerst, 1999). The anterograde and/or retrograde membrane traffic between the ER and Golgi occurs in vesicles coated with COPI and COPII proteins (Bednarek *et al.*, 1995; Schekman and Mellman, 1997; Orci *et al.*, 1997). In the TGN proteins are sorted into distinct vesicles, which

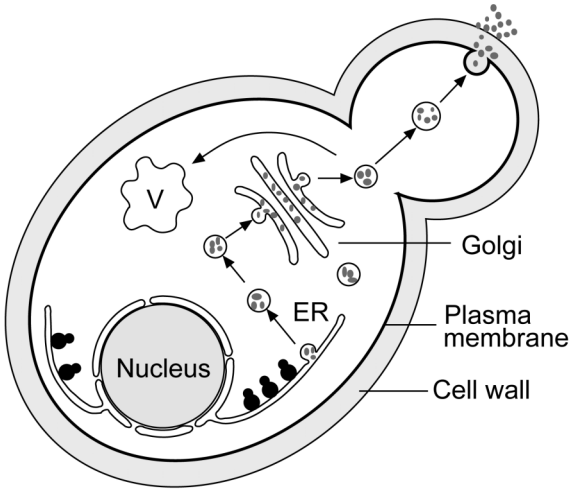

Figure 2. A schematic presentation of the secretory route of yeast *S. cerevisiae*. The nascent secretory proteins are translocated into the lumen of the endoplasmic reticulum (ER), from where they are transported by carrier vesicles *via* the Golgi to the vacuole, the plasma membrane, the cell wall or to the growth medium. Reproduction from Saris (1998) with permission.

are destined to the plasma membrane or to vacuoles (Griffiths and Simons, 1986; Conibear and Stevens, 2000). The sorting to vacuoles occurs *via* a late endosome-like prevacuolar compartment (Horazdovsky *et al.*, 1995; Gerrard *et al.*, 2000; Götte and Lazar, 1999).

The ER signal sequence consists of a short hydrophobic peptide with hydrophilic residues at both ends (von Heijne, 1985; Rapoport *et al.*, 1996). Polypeptides are translocated into the lumen of the ER usually cotranslationally, although some proteins translocate post-translationally. During co-translational translocation, the signal sequence of a nascent polypeptide is first recognized by the signal recognition particle (SRP) (Walter and Johnson, 1994). After binding to the SRP, translation is arrested until the complex is bound to the ER membrane *via* the SRP receptor. The translocation of proteins occurs through translocation channels containing hetero-trimeric complexes of the Sec61 protein (Hanein *et al.*,1996; Hamman *et al.*, 1997). The signal peptides of proteins which translocate posttranslationally are less hydrophobic and are thus independent of SRP (Ng *et al.*, 1996; Rapoport *et al.*, 1996). Posttranslational translocation requires a complex of Sec62, Sec63, Sec71 and Sec72 proteins in addition to the Sec61p translocon channel (Panzner *et al.*, 1995; Brodsky and Scheckman, 1993). The BiP chaperone has a critical role in translocation. It binds to the translocating polypeptide, acting as a molecular ratchet thus preventing the backsliding of the protein (Matlack *et al.*, 1999).

The ER is the main location where secretory proteins fold and it contains several folding enzymes and chaperone proteins essential for proper protein folding and transport (Hong, 1996). The lumen of the ER is an oxidizing environment and contains high levels of calcium, which provides proper conditions for chaperones and disulfide bond formation (Lodish *et al.*, 1992; Hwang *et al.*, 1992). Signal peptidase and glycosyltransferase complexes are functionally associated, and signal sequence cleavage and glycosylation, together with protein folding occur cotranslationally in the ER. In yeast cells, proteins are glycosylated with N-linked and O-linked glycans. Glycosylation is assumed to promote correct folding, protect against proteolysis and thermal denaturation, as well as to regulate intracellular trafficking (Lis and Sharon, 1993). In O-linked oligosaccharides, the first mannose residue is attached to a serine or threonine in the ER, which is unlike mammalian cells where mucin-type O-glycosylation is initiated in the Golgi by adding an N-acetyl-galactosamine to the fully folded protein (Tanner and Lehle, 1987; Lussier *et al.*, 1995; Van den Steen *et al.*, 1998). O-linked oligosaccharide chains in yeast can be elongated up to 5 mannoses in the Golgi (Fig. 3A) whereas O-glycans in mammalian cells are elongated by adding mannose, galactose, N-acetyl glucosamine, fucose, sialylate residues and/or polylactosamine-extensions (Varki, 1998, Van den Steen *et al.*, 1998). N-glycosylation of secreted glycoproteins in yeast is initiated in the ER (Helenius, 1994, Helenius *et al.*, 1992). In yeast, the core oligosaccharide consisting of two N-acetyl glucosamine, nine mannose residues and three glucose residues is similar to that of higher eukaryotes (Fig. 3B) (Kukuruzinska *et al.*, 1987). Unlike mammalian cells, the mannose residues are not replaced by other sugars (N-acetyl glucosamine, galactose, fucose and sialic acid) in the Golgi, but the outer chains are elongated by mannose residues only (Fig. 3C) (Byrd *et al.*, 1982; Tanner and Lehle, 1987).

1.2.4. Optimizing expression

The genetic background of both natural and recombinant yeast strains have been found to affect the quantity and structure of heterologous proteins (Eckart and Bussineau, 1996). Because different genetic backgrounds can influence transcription and translation efficiencies, the secretory pathway, protein quality, plasmid stability and plasmid copy number, it

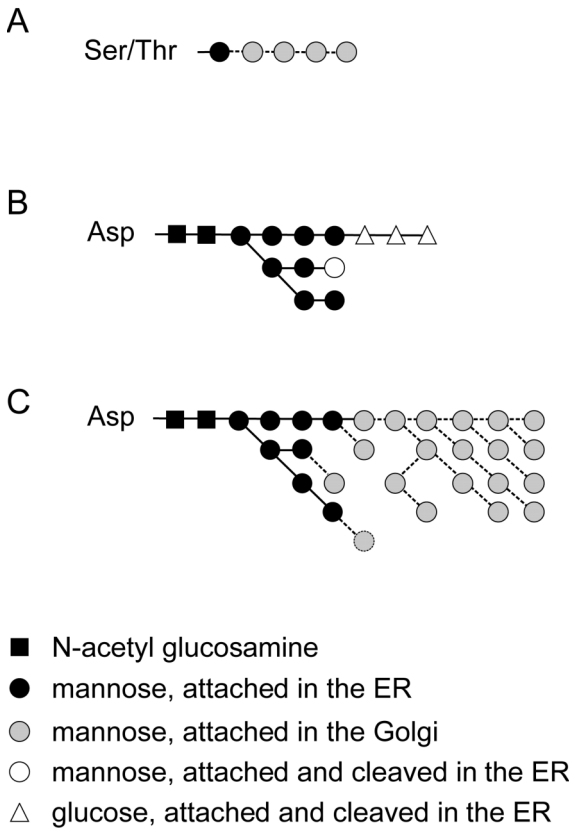

Figure 3. Carbohydrate structures of *S. cerevisiae* glycoproteins. (**A**) The first mannose of an O-linked oligosaccharide is attached to a serine (Ser) or threonine (Thr) residue in the ER and the chain is elongated up to pentamannosides in the Golgi compartment. (**B**) An N-acetyl glucosamine is attached to an asparagine (Asp) residue to from the core structure of N-linked oligosaccharide in the ER and (**C**) the outer chains are elongated in the Golgi. The length of the outer chain varies from 2 to 15 mannose residues.

may be necessary to screen a number of different host strains with varying genetic backgrounds when optimizing heterologous protein expression in yeast (Schultz *et al.*, 1994). Some genetic background considerations that may aid in foreign protein production in yeast include: the use of protease-deficient strains, which can decrease heterologous protein degradation thus increasing yields (Van den Hazel *et al.*, 1996), and the use of strains lacking hyperglycosylation of N-linked sites (Schultz *et al.*, 1994; Nakanishi-Shindo *et al.*, 1993). However, these strains usually grow slower than normal strains, but there are also yeast strains that require relatively few generations to obtain high production yields (Bussineau and Shuster, 1994). An N-terminus sequence of human interleukin 1ß could be utilized as an enhancer for heterologous protein expression in the same way as has been reported for two human growth hormones (Lee *et al.*, 1999). And, as in *E. coli*, the co-expression of chaperones have been reported to assist in the proper folding and secretion of foreign proteins in *S. cerevisiae* (Langer *et al.*, 1992; Chen *et al.*, 1994; Robinson *et al.*, 1995). Furthermore, the overexpression of disulfide isomerase (PDI) resulted in a 10-fold increase in secreted human platelet-derived growth factor, a four fold increase in acid phosphatase (Robinson *et al.*, 1994), a two to eight fold increase for five single chain antibody fragments (Shusta *et al.*, 1998), and a 15 to 24-fold increase in antistasin (Schultz *et al.*, 1994).

1.3. Insect cells (Baculovirus system)

The baculovirus expression system (BVS) is commonly used to express heterologous proteins in insect cells (Kost and Condreay, 1999). Since insect cells are eukaryotic, proteins expressed in them will be post-translationally modified in a manner similar to that of mammalian cells (Miller, 1988). Insect cells can be grown as suspension cultures, which permits the use of large-scale bioreactors for easier production scale-up. Furthermore, unlike mammalian cell lines insects cells do not require CO_2-incubators. The most commonly used baculovirus system utilized *Autographa californica* Multiple Nuclear Polyhedrosis virus (AcMNPV) (Jones and Morikawa, 1996) in a cell line derived from Lepidopteran *Spodoptera frugiperda* ovarian cells (Sf9). The AcMNPV's genome is a double-stranded, circular DNA, approximately 130 kilobases in length, and has been fully sequenced (Ayres *et al.*, 1994). The produc-

tion systems are commonly regulated by strong promoters like polyhedrin and p10 but also immediate early promoters (for example *ie1*) have been utilized (Jarvis *et al.*, 1990). Latter expression systems gave a continuous and stable expression of human glycoprotein in both infected and transformed Lepidopteran cells (Jarvis *et al.*, 1996).

As an alternative to baculovirus expression systems, an expression cassette for continuous protein expression by transformed insect cells have also been developed (Farrell *et al.*, 1998). The system utilizes the promoter of the silkmoth cytoplasmic actin gene, the *ie*1 transactivator gene and the HR3 enhancer region of *Bombyx mori* MNPV to stimulate gene expression. Levels of produced proteins in this system are comparable to the ones produced in BVS (Lu *et al.*, 1997; Keith *et al.*, 1999).

1.3.1. Systems utilizing polyhedrin and p10

In general, wild type baculoviruses exhibit lytic and occluded life cycles, producing extra-cellular and occlusion derived virus (ODV), respectively. ODV is embedded in proteinaceous viral occlusions called polyhedra (Rohrmann, 1986). Polyhedrin, a 29 kD protein, is the major component of polyhedra, which in turn are crystalline occlusion bodies visible in the nuclei of infected cells by light microscopy. The expression of viral genes is regulated in a successive cascade consisting of 4 distinguishable phases: early, delayed early, late and very late (Summers and Smith, 1988). During the very late phase of viral replication two very abundant mRNAs are produced, one of which codes for the polyhedrin protein and other codes for the p10 protein (10 kDa). The p10 polypeptide is involved in the formation of fibrillar structures found both in the nucleus and in the cytoplasm of baculovirus infected cells. The role of polyhedra is to protect occluded baculoviruses after the death of infected insects, and its presence is essential for the maintenance of the virus in nature. However, viral propagation in cell cultures is based on virions which bud from infected cells, not on ODV. Thus, baculovirus expression vectors have been engineered in which the *polyhedrin* and *p10* coding regions are deleted, but their strong promoters are retained to express foreign proteins in large amounts in the late and very late phases of infection (Smith *et al.*, 1983b; Kost and Condreay, 1999).

In the baculovirus expression system utilizing the *p10* promoter, the sequence encoding the p10 protein as well as the entire polyhedrin locus, including its promoter, have been deleted (Vlak *et al.*, 1990; Roelvink *et al.*, 1992; Bonning *et al.*, 1994; Naggie and Bentley, 1998). The remaining *p10* promoter is used to express recombinant protein sequences. In this way, expression of both the native polyhedrin and p10 proteins are abolished, which reduces background problems caused by the expression of large amounts of very late native proteins while increasing the expression of foreign recombinant proteins. Furthermore, the *p10* promoter is activated earlier in the infection cycle than the *polyhedrin* promoter (Roelvink *et al.*, 1992; Bonning *et al.*, 1994). The absence of the p10 protein results in delayed cell lysis, and thereby an extended period of recombinant protein synthesis. The yields of two reporter enzymes, juvenile hormone esterase and ß-galactosidase were higher under the *p10* promoter compared to *polyhedrin* controlled expression (Bonning *et al.*, 1994).

1.3.2. Recombination

In baculovirus expression systems the insertion of genes coding for foreign proteins can be accomplished by homologous recombination, site-specific transposition (Bac-to-Bac) or insertion directly into the viral genome *in vitro*. In homologous recombination, the gene of interest is cloned into a transfer vector containing a baculovirus promoter flanked by baculovirus DNA derived from the polyhedrin gene. After transfection into insect cells, the gene is inserted into the genome of the parent

virus by homologous recombination, at a rate of approximately 0.1 to 1% (Smith *et al.*, 1983a). Recombinants are identified by their altered plaque morphology, which appears as 'occlusion minus' plaques. A higher rate of recombination can be achieved when the parent virus genome is linearized at site(s) near the target site of foreign gene insertion (Kitts *et al.*, 1990). Kitts and Possee (1993) reported recombination rates approaching 100% when they used linearized viral DNA that is missing an essential portion of the baculovirus genome downstream from the polyhedrin locus. Nevertheless, by this method it can take more than a month to purify the plaques, and confirm the desired recombinants. A faster method to generate recombinant baculovirus can be achieved by utilizing baculovirus Ac-omega, which contains an unique restriction site downstream of the polyhedrin promoter (Ernst *et al.*, 1994). By this method recombinant viruses were obtained 8 days posttransfection and according to PCR analysis the non-recombinant background was 25 fold lower than that of the recombinant viral DNA. Another method utilizing site-specific transposition is also relatively fast (7 - 10 days) because plaque purification or virus amplification are not needed (Luckow *et al.*, 1993). In this method, the foreign gene is cloned into a baculovirus shuttle vector (bacmid) that can replicate in *E. coli*, but can also infect susceptible lepidopteran insect cells (Fig. 4). Bacmid is a recombinant virus that contains a mini-F replicon, a kanamycin resistance marker, and attTn7, the target site for the bacterial transposon Tn7 (Leusch *et al.*, 1995). Expression cassettes comprising a baculovirus promoter driving expression of a foreign gene that is flanked by the left and right ends of Tn7, can transpose to the target bacmid in *E. coli* when Tn7 transposition functions are provided by a helper plasmid.

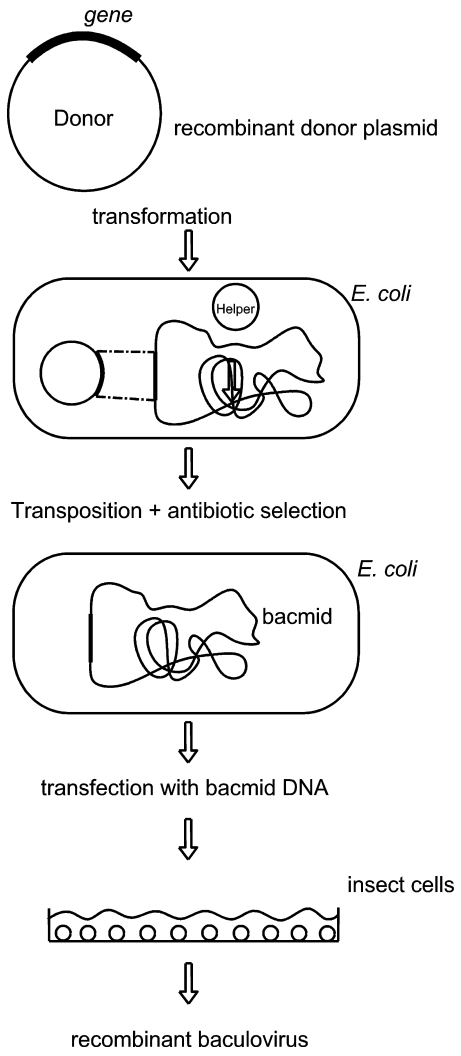

Figure 4. Schematic outline of the generation of recombinant baculovirus with a bac-to-bac expression system. A donor vector carrying the gene of interest flanked by a Tn7-element is transformed to *E. coli*. The Tn7 element can integrate into the attTn7 target site in the presence of transposition proteins provided by a helper plasmid. Recombinant bacmids are selected using their antibiotic resistance and recombinant bacmid DNA is used to transfect insect cells.

1.3.3. Optimizing expression

The density at which cells are infected, and the multiplicity of infection (m.o.i.) greatly affects the expression of heterologous proteins (Power *et al.*, 1994; Licari and Bailey, 1992; Wong *et al.*, 1995; Klaassen *et al.*, 1999). For instance, when the integral membrane protein bovine rhodpsin was expressed in large amounts, the highest volumetric yields were obtained with an m.o.i. of 0.01 during early to mid-exponential growth (Klaassen *et al.*,

1999). For easier purification, and cost saving reasons, it is possible to produce heterologous proteins in Sf9 cells in serum-free and protein-free media. The cells were gradually adapted to serum-free media in monolayer cultures and reached a doubling time of approximately 25 h, compared with 18 h in serum containing medium. Furthermore, the control of proteolysis in insect cells has a important role when optimizing the yield of recombinant protein (Naggie and Bentley, 1998). Like in *S. cerevisiae,* unwanted mannose residues can be added to secreted glycoproteins and they lack penultimate galactose as well as terminal sialic acid residues (Jarvis and Finn, 1996). The immediate early promoters (for example *ie1*), have been shown to be beneficial for protein production *e.g.*, when highly glycosylated or otherwise modified proteins are to be produced (Bonning *et al.*, 1994; Chazenbalk and Rapoport, 1995).

1.4. Mammalian expression systems utilizing vaccinia and alphavirus vectors

In some cases, mammalian cell systems are the only possible way to produce properly processed and active foreign proteins. Gene transfer into mammalian cells may be performed either by infection with a virus carrying the recombinant gene of interest (reviewed by Makrides, 1999) or by direct transfer of plasmid DNA (Geisse and Kocher, 1999).

1.4.1. Vaccinia virus vector

Recombinant vaccinia virus vectors have generally been acknowledged as versatile tools for the expression of foreign genes (reviewed by Moss, 1996). Since vaccinia is infectious to man, safety aspects must be taken into consideration, especially when working with large-scale preparations. To circumvent safety problems, an avian host-restricted vaccinia, modified vaccinia Ankara (MVA), can be used (Sutter *et al.*, 1994; Wyatt *et al.*, 1995). Furthermore, MVA does not produce the rapid cytophatic effect accompanied by destruction of the cell monolayer seen with the standard replication strains of vaccinia used. MVA has been shown to efficiently produce foreign proteins in several mammalian cell lines using either the homologous vaccinia promoter, *p11* (Sutter and Moss, 1992) or a hybrid vaccinia/bacteriophage T7 promoter system (Wyatt *et al.*, 1995). In the vaccinia/T7 polymerase hybrid system, the vaccinia virus contains the T7 RNA polymerase gene, which is under the control of an early vaccinia promoter. The foreign gene to be expressed is cloned into a separate expression vector under the control of the *T7* promoter. After the host cells are infected with the vaccinia/T7 virus, the cells are transfected with the expression plasmid. Problems with vaccinia systems include their cytopathic nature, and dependence on efficient transfection rates. Mammalian cell lines can be engineered so that the genes to be transfected are stably integrated into the cell´s genome, but then a tightly controlled-inducible promoter must be used. This method is especially advantageous when foreign protein production must be scaled-up. Furthermore, homologous vaccinia promoter systems can be more effective in producing secreted foreign proteins as compared to the heterologous T7 promoter system (Pfleiderer *et al.*, 1995). Another vaccinia virus based expression system utilizes defective vaccinia virus lacking the D4R open reading frame, and a complementary cell line providing the D4R gene product (Himly *et al.*, 1998). Experiments done with human secreted proteins (*i.e.*, factors VII and XI) showed that the defective system produces more secreted proteins than the wild type vaccinia. Surprisingly, recombinant human factor VII was more efficiently produced using the defective vaccinia recombinant under non-complementing conditions (Himly *et al.*, 1998). This suggests that the persistence of early phase vaccinia replication, combined with a delay in the shutoff of host protein synthesis, can be advantageous for foreign protein production.

1.4.2. Alphavirus vectors

Semliki Forest (SFV) and Sindbis (SIN) viruses together with Venezuelan equine encephalitis virus are used as vectors for the expression of heterologous proteins in many diverse cells types (Makrides, 1999; Garoff and Li, 1998; Agapov *et al.*, 1998). Alphavirus replicon vectors are self-replicating RNA molecules which include the genes for non-structural proteins (Nsps), while lacking the genes of the viral structural proteins. Instead of the structural genes, cloning sites have been engineered to accept foreign genes for expression. The replicons are introduced into host cells by transfection either as RNA, synthesized *in vitro*, or as DNA under the control of an eukaryotic promoter. Alternatively, the recombinant alphavirus RNA can be packaged into viral particles in cells that have been cotransfected with replication competent, but packaging incompetent helper RNA coding for the alphavirus structural proteins (Bredenbeek *et al.*, 1993). Although, recombination between vector and helper genomes can generate a fully infectious virus, this usually happens at very low frequencies (Berglund *et al.*, 1993). One way to prevent this recombination is to use conditional mutations to limit the infectivity of any such recombinants (Smerdou and Liljeström, 1999). In an improved packaging system, the helper virus has been modified so that the structural proteins are expressed in two separate plasmids.

As in vaccinia systems, alphavirus infections shuts off the host's own protein synthesis, which makes protein purification easier. Compared to vaccinia, SFV is less cytopathic to the host, and thereby potentially more efficient in foreign protein production (Liljeström and Garoff, 1991). Furthermore, non-pathogenic alphavirus vectors have been developed. For example, a Sindbis virus derivative, which has low host cell pathogenicity due to a single point mutation in Nsp2, demonstrates a persistent infection, while its replication rate is as high as the one of the wild type SIN (Dryga *et al.*, 1997). Although expression of foreign proteins using alphaviruses derived vectors can generate very high yields (up to 25% of total cellular proteins), cloning of large foreign genes into replicons can be problematic (Liljeström and Garoff, 1991). Foreign sequences of about 2 kb are usually stable in SIN vectors, whereas various inserts over 3 kb were rapidly deleted suggesting that the packaging capacity of the virion had been exceeded (Pugachev *et al.*, 1995). Besides tissue and suspension culture applications of alphavirus replicons, *in situ* infections and foreign protein expression have been accomplished in the neurons of rat hippocampal slices (Schlesinger and Dubensky, 1999). This system has been used to produce ß-galactosidase, tissue plasminogen, chloramphenicol acetyl transferase, hemagglutinin and Hepadnavirus proteins. Furthermore, derivatives of alphavirus vectors have been used to direct foreign protein expression in specific cell types in a tissue (Huang, 1996; Ohno *et al.*, 1997).

1.4.2.1. Semliki Forest virus

Semliki forest virus (SFV) is an enveloped RNA virus with a single strand genome of positive polarity. SFV belongs to the *Alphavirus* genus which in turn is a member of the family *Togaviridae*. The prototype virus of SFV, used in laboratories, is considered to be nonpathogenic to humans and has been long studied as a model virus for the *Alphavirus* genus. SFV enters the cell *via* receptor mediated endocytosis, resulting in nucleocapsid liberation of its RNA genome into the cytoplasm (Helenius *et al.*, 1980). Two-thirds of the genomic 42S RNA (total length ca. 11.5 kb) is translated into a polyprotein, which is then autoproteolytically cleaved into four non-structural proteins: Nsp1 - Nsp4 (Fig. 5) (reviewed in Kääriäinen *et al.*, 1987; Strauss and Strauss, 1994). In the initial stages of SFV infection, a RNA replication complex, is formed from the intermediate translation products Nsp123 and Nsp4 and is needed to recognize the 42S RNA plus strand for tran-

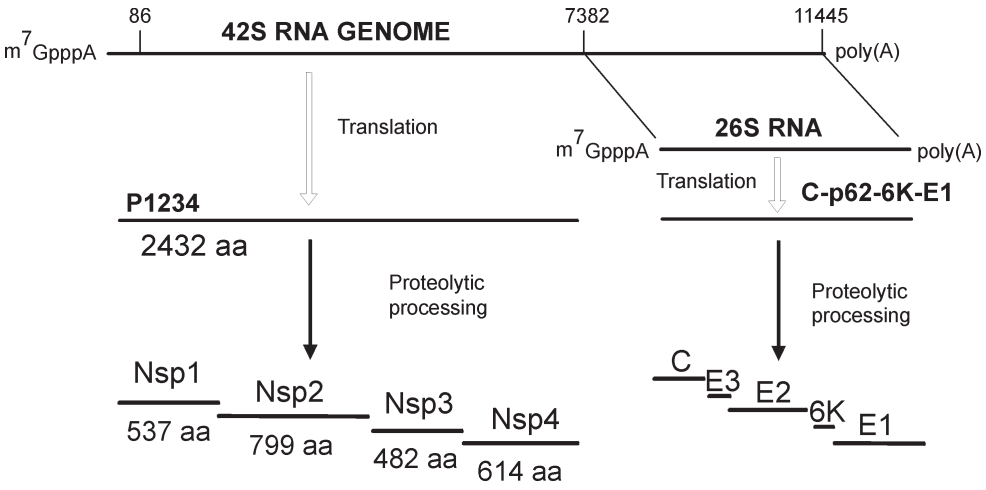

Figure 5. Proteolytic processing of the nonstructural and structural proteins of Semliki Forest virus.

scription of the negative strand 42S RNA (Lemm and Rice, 1993a, 1993b). Later in infection the intermediate Nsp123 is proteolytically processed into the final products, Nsp1, Nsp2 and Nsp3, after which the transcription of the minus strand 42S RNA will be terminated. According to the genetic criteria, all nonstructural proteins are needed to form the replication complex. Replication occurs in association with modified endosomes and lysosomes called cytopathic vacuoles type I (CPVI) (Froshauer *et al.*, 1998; Peränen and Kääriäinen, 1991). The structural proteins: capsid protein and envelope glycoproteins E1, E2 and E3 are transcribed from a subgenomic 26S RNA (reviewed in Strauss and Strauss, 1994). The functions of individual Nsps are briefly reviewed below.

1.4.2.1.1. Nsp1

Nsp1 (537 aa) is multifunctional enzyme catalyzing the capping of viral mRNA. Firstly, it is a methyltransferase (MT) (Mi *et al.*, 1989, Mi and Stollar, 1991; Laakkonen *et al.*, 1994). Nsp1, like the methyltransferase purified from the vaccinia virion can methylate GTP in the presence of S-adenosyl-L-methionine (AdoMet) (Laakkonen *et al.*, 1994, Martin and Moss, 1976). Secondly, Nsp1 has been shown to form a covalent complex with 7-methyl-GMP thus providing at least the first reaction catalyzed by guanylyltransferases (GT) (Ahola and Kääriäinen, 1995). The capping reaction of viral mRNA catalyzed by Nsp1 differs from the capping of eukaryotic mRNA and many other viral mRNA by requiring methylation of guanine before covalent complex formation between the guanylyltransferase and 7-methyl-guanylate (Fig. 6) (Ahola and Kääriäinen, 1995). The 7-methyl-guanylate is probably bound to Nsp1 through a phosphoamino bond either on a lysine or on a histidine. Furthermore, Nsp1 also assists in the initiation of minus strand synthesis (Hahn et al, 1989b).

When Nsp1 is expressed alone in HeLa cells by transfection, it associates with the plasma membrane as well as with endosomes and lysosomes (Peränen *et al.*, 1995). The tight membrane association of Nsp1 is partly due to palmitoylation on amino acids Cys^{418}-Cys^{420} (Peränen *et al.*, 1995; Laakkonen *et al.*, 1996). When the palmitoylation sites are mutated, the protein remains enzymatically active and bound to membranes (Laakkonen *et al.*, 1996), though less tightly as compared to the wild type (*i.e.*, the nonpalmitoylated mutant form could be released from the membranes with 1 M NaCl, whereas wt Nsp1 could be released only with 50 mM sodium carbonate, pH 12).

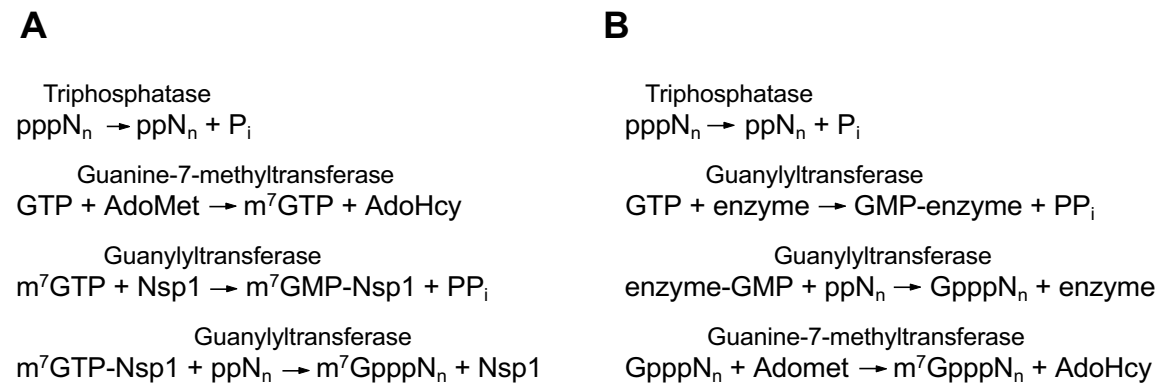

Figure 6. Reactions involved in capping of alphavirus mRNA (**A**) and eukaryotic mRNA (**B**).

In addition, the morphology of cells infected or transfected with nonpalmitoylated Nsp1 differs from wt by showing fewer filopodia like structures than the wt Nsp1.

1.4.2.1.2. Nsp2

According to what is presently known, Nsp2 (799 aa) is responsible for three different enzymatic activities in SFV replication. Firstly, it is an autoproteinase containing a papin-like thiol proteinase at its C-terminal region (Hardy and Strauss, 1989). Secondly, it is an RNA helicase (Gomez de Cedron *et al.*, 1999), which has single stranded RNA-stimulated ATPase and GTPase activities (Rikkonen *et al.*, 1994). Thirdly, Nsp2 is a triphosphatase, which catalyzes the first reaction in mRNA capping (Vasiljeva *et al.*, 2000). Moreover, Nsp2 has been reported to act on the regulation of negative strand synthesis, and is required for the synthesis of the subgenomic 26S RNA (Hahn *et al.*, 1989b). About half of the Nsp2 is transported into the nucleus during infection, and this transport is mediated by nuclear signal sequences (Peränen *et al.*, 1990; Rikkonen *et al.*, 1992). SFV in which the nuclear targeting signal of Nsp2 is removed, has been reported to be apathogenic in mice (Rikkonen *et al.*, 1996).

1.4.2.1.3. Nsp3

Nsp3 (482 aa) is an essential subunit for the replication complex although its specific functions are still not known. Studies of SIN Nsp3 have suggested that the polyprotein intermediates have distinct essential functions in the synthesis of negative strands during the early phases of RNA replication (Lemm *et al.*, 1994; Shirako and Strauss 1994). SIN Nsp3 has been reported to affect negative-strand RNA synthesis, and possibly also the synthesis of subgenomic mRNA (Wang *et al.*, 1994; LaStarza *et al.*, 1994b). When the localization of SFV Nsp3 was studied by indirect immunofluorescence it was found to be mainly in vacuole like structures both in infected as well as transfected cells (Peränen and Kääriäinen, 1991).

The N-terminal region of different Nsp3s are very conserved among the alphaviruses, and part of that sequence has homology to rubella virus, hepatitis E virus, and coronaviruses sequences (X-motif, Koonin and Dolja, 1993). Recently, through genome sequencing, it had become apparent that this domain can be found in bacteria, animals and plants and it shows homology to the nonhistone region of macro histone 2 (Pehrson and Fuji, 1998). The National Center for Biotechnology Information (NCBI) has assigned this domain as

DUF27 with an unknown function. Moreover, this domain has significant homologue to one yeast ORF which has been associated with adenosine-diphosphate-ribose-1-phosphate (Appr-1"-p) processing activity (Martzen *et al.*, 1999). Conservation of this region during evolution suggests an important function(s) for this domain.

The C-terminal part of alphavirus Nsp3 is not conserved among alphaviruses, and varies both in sequence and in length from 134 aa (Middel-burg virus) to 246 aa (O'nyong-nyong virus) (Strauss and Strauss, 1994). This domain is rich in acidic residues, as well as in serine and threonine and appears to be devoid of any predicted secondary structures. The region also contains duplicate amino acid (aa) sequences *i.e.*, ADVPEPA, PAPR and TFGDFD, suggesting that this region has evolved through duplication events. Nsp3 is the only alphavirus nonstructural protein modified by phosphorylation (Peränen *et al.*, 1988; Li *et al.*, 1990).

1.4.2.1.4. Nsp4

Even though direct evidence is still lacking, Nsp4 (614 aa) most probably is the catalytic subunit of the alphavirus RNA-replication complex (Barton *et al.*, 1988). Nsp4 has conserved a GDD motif found in other viral RNA polymerases (Argos, 1988), and a SIN temperature sensitive mutant, ts6, has been reported to fail in the synthesis of RNA due to a mutation in Nsp4 (Sawicki *et al.*, 1981a 1981b, Hahn *et al.*, 1989). Nsp4 has also been reported to be an autoproteinase (Takkinen *et al.*, 1990).

2. Purification strategies for recombinant proteins

Purification is an important step in the production of recombinant proteins. The characteristics of industrial-scale purification schemes, such as conventional chromatography, have a significant impact on the final cost of production. It is often more efficient to use one of the available tags to 'fish out' the target protein. The purpose for which the protein will be used determines required degree for its purity and authenticity. To purify intracellularly produced proteins the cells are harvested and lysed, which naturally contributes to the complexity of the protein mixture. The advantage of intracellular versus secreted protein is the volume to be handled, *i.e.*, the secreted proteins usually exist diluted in the culture medium. Isolation of a desired protein from the medium follows the general scheme: (1) concentration by precipitation, ultrafiltration, batch adsorption or partition in an aqueous phase system (2) enrichment by chromatography or partition (3) high resolution purification: by chromatography and/or immuno adsorption (4) final concentration (Menge *et al.*, 1987). If fetal calf serum (FCS) is used in cultivation, bovine serum albumin and globulins will be the most abundant proteins in the supernatant. The interaction of albumin especially with hydrophobic proteins represents a significant problem for the effective purification of minor constituents.

Even though there are many existing examples as well as heuristics and complex algorithms for suggesting the potential of purification processes, no universal scheme has been developed which could be applied to all proteins. There are a variety of different methods which can be used to purify proteins, but only chromatographic purification methods such as affinity, ion exchange, reversed phase and gel filtration chromatography will be briefly reviewed here.

2.1. Conventional chromatography

Chromatographic methods (size exclusion, ion exchange, reversed phase, hydrophobic interaction and affinity based) can be utilized either in traditional, low pressure or high-performance liquid chromatography instrumentation. In addition to liquid-solid chromatography there is liquid-liquid chromatogra-

phy which is analogous to gas-liquid chromatography (Cichna *et al.*, 1995). There are also many other types of chromatography methods *e.g.*, liquid adsorption chromatography, frontal chromatography, displacement chromatography covalent chromatography and membrane chromatography (Mao *et al.*, 1993; Heeter and Liapis, 1998; Freitag, 1999; Caldas *et al.*, 1998; Zeng and Ruckenstein, 1999).

2.1.1. Ion exchange chromatography

Ion exchange chromatography (IEC) has been the most widely used technique for the isolation and purification of biological macromolecules since the 1950s (Choudhary and Horváth, 1996). IEC is able to separate almost any type of charged molecules, from large proteins to small nucleotides and amino acids. The ion exchanger are insoluble solid matrices containing fixed ionogenic groups which bind reversibly to sample molecules (proteins, etc.). Desorption is then brought about by increasing the salt concentration or by altering the pH of the mobile phase. The two major classes of ion exhangers are cation exchangers and anion exchangers, having negatively and positively charged functional groups, respectively. Ion exchange containing diethyl aminoethyl (DEAE) or carboxymethyl (CM) groups are most frequently used.

Since the protein is covered by a hydrophilic layer, electrostatic interactions have a major role in the retention of proteins. In addition to size, geometric form as well as hydrophopic and van der Waals interactions affect the separation (Stålberg, 1999). Nevertheless, the major property which govern the adsorption to an ion exchanger is the net surface charge of the protein. Since surface charge is the result of weak acidic and basic groups on proteins, separation is highly pH-dependent. The optimum pH range for IEC for many proteins is within 1 pH units of the isoelectric point. Many retention models have been examined but the complexity of the adsorption process for a protein to solid surfaces makes it difficult to construct physical models for the interactions (Stålberg, 1999). Gradient elution with increasing salt concentration is most commonly used in the IEC. The higher the net charge of the protein, the higher the ionic strength needed to bring about desorption. Thus, to optimize selectivity in ion exchange chromatography, the pH of the running buffer is chosen so that sufficiently large net charge differences among the sample components are created.

2.1.2. Reversed phase chromatography

Reversed phase chromatography (RPC) can be utilized to separate compounds according to their hydrophobicity (Geng and Regnier, 1984). Unlike many other methods RPC is able to separate closely related and structurally disparate substances, even at picomolar levels (Aguilar and Hearn, 1996; Pearson *et al.*, 1982). In RPC, silica particles covered with chemically-bonded hydrocarbon chains represent the lipophilic phase (C2 to C18), while an aqueous mixture of an organic solvent surrounding the particle represents the hydrophilic phase (Henry, 1991; Zhou *et al.*, 1991). Depending on the extractive power of the eluent, a greater or lesser part of the sample component will be retained reversibly by the lipid layer of the particles. The partitioning of the sample components between the two phases will depend on their respective solubility characteristics. Less hydrophobic components will end up primarily in the hydrophilic phase while more hydrophobic ones will be found in the lipophilic phase. This can be affected by the addition of an organic solvent which is soluble in the hydrophilic phase. Some commonly used organic solvents, in order of increasing hydrophobicity are methanol, propanol, acetonitrile, and tetrahydrofuran. Separated components can be directly subjected to further analysis such as Edman sequencing or electrospray mass spectrometry.

The ability of a stationary phase (lipophilic phase) to discriminate between two components is reflected by the volume between the peak maxima of the corresponding zones after passing through the column (Aguilar and

Hearn, 1996). Along with partitioning, mechanism adsorption operates at the interface between the mobile and the stationary phases (Melander *et al.*, 1984). Thus, the retention of hydrophobic components will be greatly influenced by the thickness of the lipid layer. A C18 layer is able to accommodate more hydrophobic material than C8 or C2 ones. For hydrophilic components, changing from a C18 to a C2 layer influences the separation very little since only the surface area of the lipid layer is active. The mobile phase can be considered as an aqueous solution of an organic solvent, the type and concentration of which determines the extractive power. Moreover, according to experimental data, components interact with the chromatographic surface in an orientation-specific manner (Chicz and Regnier, 1990; Regnier, 1987). RPC has been utilized for purification of a variety of proteins and peptides *e.g.*, aprotinin, cytochrome C, bovine serum albumin, fibrinogen, insulin and lysosyme (Honda *et al.*, 1992; Nimura *et al.*, 1992). Even though some proteins have been reported to maintain their native structure during RPC (*e.g.*, insulin, thyroid-stimulating hormone, growth hormone), in most cases denaturing conditions are required which might limit the use of RPC (Welinder *et al.*, 1986; Welinder *et al.*, 1987; Forage, 1986; Chlenov *et al.*, 1993).

2.1.3. Gel permeation chromatography

The principle of gel permeation chromatography (GPC; size exclusion or gel filtration chromatography) is based on molecular volumes. Large molecules are excluded from the matrix, whereas intermediate size molecule can partly enter and only small molecules can freely enter the matrix. The porous three dimensional matrix acts as a steric barrier to solute molecules as they attempt to equilibrate with liquid inside and outside the bead. While purifying the protein GPC can also be used to estimate approximate molecular weights. Furthermore, gel filtration can be used for desalting or buffer changing when the fractionated size of the gel is small (*e.g.*, Bio-Gel 6, Sephadex G-25) allowing the protein to elute in the void volume.

The choice of the appropriate column type depends on the molecular size and physical properties of the proteins to be separated. A fraction of the internal volume which is accessible to a solute molecule can be described by a constant K_d, which can be calculated with experimentally determined elution volumes. A column should be chosen so that separations occur in the linear part of the K_d vs. the molecular weight plot (van Dijk and Smit, 2000). However, it is practically impossible to adjust the pH and ionic strength so that the proteins are in an equal state, since at the isoelectric point the shape of globular and polymer coil proteins relieve compact and soft sphere, respectively. Furthermore, the ionic strength of the buffer used may affect the size of the proteins. Experiments with neutral dextrans and charged proteins showed that the effective protein size increases with decreasing ionic strength due to the reduction in electrostatic shielding (Pujar and Zydney, 1998). Thus, although GPC is viewed as size-based separation process, there is considerable evidence for the importance of electrostatic interactions as well. GPC has been used in the renaturation procedure (Müller and Rinas, 1999; Batas and Chaudhuri, 1999). For example heterodimeric platelet-derived growth factor was purified from inclusion bodies after denaturation with Gdn-HCl (Müller and Rinas, 1999). Renaturation of this growth factor involved folding into an active heterodimeric form during GPC while circumventing aggregation observed when refolding was carried out by dilution.

2.2. Affinity chromatography

Affinity chromatography together with recombinant DNA-technology offers a simple and fast technique to purify proteins to high purity with a single purification step (Scouten, 1991). Fusion can be made on either side or

both sides of the target gene depending on specific application, but the majority of fusion proteins place the tag at the N-terminus of the protein (Nilsson *et al.*, 1997). Genetic manipulation of the protein can be used to form a cleavage site, which helps to remove the affinity tag after purification thereby resulting in an intact protein. Although affinity chromatography can be used for laboratory-scale purification, its utilization on a preparative scale can represent a major cost for the final protein product.

Successful separation by affinity chromatography requires that a biospecific ligand is available, and that it is covalently attached to a chromatographic bed material. It is important that the biospecific ligand (antibody, enzyme, or receptor protein) retains its specific binding affinity for the substance of interest (antigen, substrate, or hormone). These interactions typically have high affinities ($K_d < 10^{-6}$ M), yet are reversible when conditions are changed (Wilchek *et al.*, 1984). Due to the specificity of this recognition, it is often possible to obtain 100-, 1.000- or even 100.000-fold increases in purity of a protein sample (Clausen *et al.*, 1990). The packing material used, called the affinity matrix, must be inert and easily modified. Agarose is the most common substance used as a matrix, in spite of its relative high costs. The ligands, or "affinity tails", that are inserted into the matrix can be genetically engineered to possess a specific affinity. In a process similar to ion exchange chromatography, the desired molecules adsorb to the ligands on the matrix until desorption is carried out *e.g.*, with a high salt concentration, a competition reaction (*e.g.*, imidazole), strong chelating agents and/or low pH. Fouling of the matrix can occur when a large number of impurities are present, therefore, this type of chromatography is usually implemented late in the process. In addition to most common affinity chromatography utilizing HPLC, there are also other techniques which involve affinity, such as affinity precipitation, affinity partitioning of proteins using aqueous two-phase system, foam fractionation and dye ligand chromatography (Hoshino *et al.*,1998; Birkenmeier *et al.*, 1984; Lockwood *et al.*, 1997; Boyer and Hsu, 1993).

2.2.1. Affinity tags

To date, a large number of different fusion partners that range in size from one amino acid to whole proteins capable of selective interaction with ligand immobilized onto a chromatography matrix, have been described (Nilsson *et al.*, 1997). Although a multitude of systems have been introduced, no single affinity fusion partner is ideal for all expression or purification systems. Some of the most commonly used tags are reviewed below and listed in Table 1.

In 1975 Porath and co-workers introduced a method based on the interaction between the side chains of certain amino acids, especially histidines, on a protein surface and immobilized transition metal ions. Immobilized metal affinity chromatography (IMAC) systems have three basic components: an electron donor group, a solid support and a metal ion. The metal ion (usually Ni^{2+}, Co^{2+}, Cu^{2+} or Zn^{2+}) is restrained in a coordination complex where it still retains significant affinity towards macromolecules (Porath, 1992). The use of **polyhistidine tags** has been demonstrated in a wide range of host cells including *E. coli*, *S. cerevisiae*, insect cells as well as in mammalian cells (Table 1). IMAC is usually performed under nondenaturing conditions, but it is also compatible with non-ionic detergents allowing highly denaturing conditions with urea and guanidium-HCl (Smith and Roth, 1993). Also an organic solvent (isopropanol) can be used to increase the purification efficiency (Franken *et al.*, 2000). The elution is usually performed by competition with imidazole, lowering pH, or by adding strong chelating agents. IMAC is particularly suitable for preparative group fractionation of complex extracts and biofluids, but can also be used in the high-performance mode. Regardless of the relatively small size of the his-tag, is has been reported

to affect the activity of some finally purified recombinant proteins (Janssen *et al.*, 1995; Pekrun *et al.*, 1995; Büning *et al.*, 1996).

Maltose binding protein (MBP) from *E. coli* is frequently used as a fusion partner for proteins. MBP (40 kDa) is a periplasmic protein and thereby can be employed to induce the secretion into the periplasm. A mixture of cellulose and starch can be utilized as a ligand and elution can be achieved in the presence of maltose (Srinivasan and Bell, 1998). MBP has reported to increase the solubility of the fusion protein (Hennig and Schäfer, 1998; Ko and Pedersen 1995; Pérez-Martin *et al.*, 1997) and it has been successfully used as a C-terminal fusion partner (Hennig and Schäfer, 1998).

Glutathione S-Transferase (GST) -fusion vectors were first reported by Smith and Jonhson (1988). GSTs are a family of enzymes that can transfer sulfur from glutathione to substances such as nitro and halogenated compounds, leading to their detoxication. GST fusion proteins are bound to immobilized glutathione followed by competitive elution with reduced glutathione. Since GST is a relatively large tag (26 kDa), in many cases it is preferable to remove it by engineering a specific cleavage site (see section 2.2.2.). GST can also be utilized under denaturing conditions (Zhao and Siu, 1995). However, a possible complication associated with this system is the use of reduced glutathione for elution, since it can affect target proteins containing disulfides (Sassenfeld, 1990).

Staphylococcal protein A is a immunoglobulin-binding receptor on the surface of the gram-positive bacterium *Staphylococcus aureus*. Protein A (30 kDa) and its derivative Z shows a specific affinity for the Fc-region of immunoglobulin G. Protein A consists of five repetitive homologous domains (A-E), and each of these domains alone is capable of binding to IgG (Moks *et al.*, 1986). The structure of the B domain (from which Z has been derived) has been determined, revealing that the N- and C-termini of the three-helix bundle

Table 1. Examples of tags used for purification of fusion proteins

Tag	Ligand	host B/Y/I/M*	ref.$
Avidin	Biotin	B/I	1/2
in vivo biotionylated	Avidin/Streptavidin	B	3
FLAG	M1/M2	B/Y/I/M	4/5/6/7
Maltose	Amylose	B/Y	8/9
protein A or G	IgG	B/Y/I/M	10/11
calmodulin binding protein	calmodulin	M	12
Chitin-binding domain	Chitin	Y	13
Cellulose-binding domain	Cellulose	B	14
C-LYT	Choline	B	15
Strep	streptavidin	B	16
Poly-His	Ni^{2+}	B/I/M	17/18/19
GST	Glutathione	B/Y/M	20/21/20
Fc-Hinge	Protein A	M	22
Poly-Arg	Polyanionic groups	B	23
S-peptide	S-protein	B	24
Thioredoxin	Thiobond resin	B	25

*hosts: B bacteria, Y yeast, I insect cells, M mammalian cells

$References: **1.** Airenne and Kulomaa, 1995 **2.** Airenne *et al.*, 1999 **3.** Smith *et al.*, 1998 **4.** Hopp *et al.*, 1988 **5.** Clarke *et al.*, 1998 **6.** Airanne *et al.*, 1997 **7.** Chubert and Brizzard, 1996 **8.** Srinivasan and Bell, 1998 **9.** Henning and Schäfer, 1998 **10.** Zhou *et al.*, 2000 **11.** Hansson *et al.*, 1994 **12.** Neri *et al.*, 1995 **13.** Rigaut *et al.*, 1999 **14.** Ong *et al.*, 1989 **15.** Ortega *et al.*, 1992 **16.** Schmidt and Skerra, 1993 **17.** Pérez-Martin *et al.*, 1997 **18.** Schmidt *et al.*, 1998 **19.** Janknecht and Nordheim, 1992 **20.** Manoharan *et al.*, 1997 **21.** Mitchell *et al.*, 1993 **22.** Lo *et al.*, 1998 **23.** Stempfer *et al.*, 1996 **24.** Kim and Raines, 1994 **25.** La Vallie *et al.*, 1993.

structure are solvent exposed, which is supposed to favor an independent folding of the target protein (Gouda *et al.*, 1992; Jendeberg *et al.*, 1996). Analysis of different repeats of Z showed that there is no advantage in using more than two domains (Ljungquist *et al.*, 1989). Protein A is secretion competent and fusion proteins can be secreted to the *E. coli* periplasm and in some cases also to the culture medium (Abrahmsén *et al.*, 1986; Hansson *et al.*, 1994). Protein A has been reported to be highly soluble and capable of renaturation after being subjected to denaturants such as urea and guanidium HCl (Samuelsson *et al.*, 1994). The need for low pH for elution can be circumvented by the use of competitive elution strategies based on engineered competitor proteins (Nilsson *et al.*, 1997). Like protein A, **Staphylococcal protein G**, is a receptor present on the surface of certain strains of *Streptococci*. Protein G is a bifunctional protein having structurally separate regions for affinity to serum albumin and IgG. The serum albumin binding region has three helix-bundle domains, very similar to that of protein A (Kraulis *et al.*, 1996). Elution from a column is usually achieved with low pH or by alternative methods such as high pH, lithium diiodosalisylate and heat (Nilsson *et al.*, 1997). Protein G or different parts of the albumin binding domain have been used as affinity tags in protein expressed either intracellularly or as secreted proteins (Samuelsson *et al.*, 1994).

Thioredoxin (Trx, 12 kD), the product of the *E. coli trxA* gene, has been used to increase solubility as well as a fusion partner in *E. coli*. Proteins expressed as thioredoxin fusions are likely to be soluble, since thioredoxin remains soluble even at levels as high as 40% of the total cellular protein (LaVallie *et al.*, 1993; Yasukawa *et al.*, 1995). The other obvious benefit associated with the use of thioredoxin is that it provides partial purification *i.e.*, the recombinant protein accumulates in adhesion zones and it can be selectively released by rapid osmotic shock treatment. Thioredoxin also provides an affinity tag for rapid purification, since it specifically bonds to ThioBond^{TM}-resin. The resin consists of covalently bound phenylarsine oxide on an agarose-based support to which the thioredoxin's active site diothiol has a high affinity. The Trx-system has been utilized to produce short amino acid sequences (10-30 aa) which are usually extensively degraded or insoluble when expressed in *E. coli* (LaVallie *et al.*, 1993). Furthermore, a plasmid has been developed for *in vivo* biotinylation of Trx-fusion proteins in *E. coli* (Smith *et al.*, 1998).

FLAG peptide (about 1 kDa) is an eight amino acid long peptide (Hopp *et al.*, 1988). The sequence of the peptide (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys) is designed for maximum hydrophilicity and contains an enterokinase cleavage site. There are two common monoclonal antibodies raised against this peptide: M1 and M2. Antibody M1 binds to the peptide in a calcium-dependent manner, which allows gentle elution by the addition of a chelating agent. For antibody recognition the FLAG peptide must be located on the very N-terminal end of the target protein. The M1 antibody recognizes only the four first amino acids of the peptide (Asp-Tyr-Lys-Asp) allowing a shorter tag to be utilized (Knappik and Plückthun, 1994). Antibody M2 is capable of recognizing the FLAG peptide also at other locations in addition to the N-terminal site. Fusion proteins bond by antibody M2 can not be eluted with a chelating agent, but instead by low pH or competition with an excess of FLAG peptide (Brizzard *et al.*, 1994). There are also specific antibodies raised against the FLAG peptide sequence including a methionine at the N-terminus (Chubet and Brizzard, 1996).

The strong affinity between **biotin** and **avidin** or **streptavidin** have been used for various detection, immobilization and purification processes. Biotinylation of a protein can be achieved by covalent coupling of biotin to its lysine residues. The biotinylation method introduced by Cronan (1990) utilize *in vivo*

machinery for specific biotinylation which can not be achieved *in vitro*. A 75 residue fragment of transcarboxylase has been shown to be a sufficiently large substrate for biotinylation (Kohanski and Lane, 1990). Elution can be performed under mild conditions with excess biotin. Denaturing conditions may also be utilized since avidin, as a ligand, is fairly stable allowing purification in the presence of 1% SDS or 8 M urea (Cronan, 1990). The C-terminal 101 residues of the *E. coli* biotin carboxyl carrier protein of acetyl-CoA carboxylase, has been utilized as an affinity tag for Fab antibody fragment in *E. coli* (Weiss *et al.*, 1994). A subunit form of transcaboxylase has also been used as a biotinylated fusion partner in recombinant Semliki Forest virus-infected baby hamster kidney (BHK) and chinese hamster ovarian cells (Lundström *et al.*, 1995). PinPoint (13 kDa, Promega) binds to streptavidin/avidin or monomeric avidin and can be eluted with biotin. An alternative small affinity tag that binds to streptavidin is Strep (10 aa) which mimics biotin even though it is non-biotinylated (Schmidt and Skerra, 1993). Strep fusion proteins can be eluted under mild conditions with a biotin analog, named iminobiotin. An improved version of the strep tag is strep-tag II which retains its binding activity also when placed at positions other than the C-terminus (Voss and Skerra, 1997; Schmidt *et al.*, 1996).

Sometimes it is convenient to have more than one fusion partner (Scouten, 1991). The **Dual affinity fusion** concept, which in addition to providing the possibility for two affinity chromatographic purification steps, can also stabilize the target protein more than is possible with just one fusion partner (Nilsson *et al.*, 1997). Rigaut *et al.*(1999) successfully purified several proteins with a tandem affinity purification procedure. The tags used were protein A and calmodulin-binding peptide in combination with a TEV-cleavage site. The first purification step involved protein binding to an IgG matrix, followed by TEV-protease cleavage. The released protein could be separated from the protease by a second affinity chromatography step using calmodulin-coated beads. The His_6-tag together with the maltose-binding-protein have also been used for double-affinity purification system for several proteins in *E. coli* (Pryor and Leiting, 1997).

2.2.2. Cleavage

In many applications there is no need to remove the tag since it does not interfere with the activity of the protein of interest (Stempfer *et al.*, 1996). However, sometimes it is necessary to remove the tag, since it can cause unwanted immuno responses, inactivate the fused target protein or prevent structural determination. Several methods have been described for site-specific cleavage treatment of the fusion partner (Flaschel and Friehs, 1993). Careful upstream design of the fusion protein construct can facilitate the subsequent purification of the target protein. Furthermore, it may allow for integrated systems involving co-processing of the protein and efficient removal of the affinity fusion partner as well as the protease used for the cleavage. Chemical cleavage methods utilizing, for example, cyanogen bromide and hydroxylamine are relatively inexpensive. Nevertheless, the chemicals are relatively unspecific and the reaction conditions used can cause amino acid side chain modifications, denaturation or partial degradation of the target protein. Enzymatic methods are preferred due to their specificity and milder reaction conditions required. However, the proteases cleaving after a basic residue (enterokinase, thrombin and factor X_a) sometimes display nonspecific and unwanted cleavages (Nygren *et al.*, 1994). IgA-proteinase has been reported to be more specific since it requires a proline in the P2' position (see Table 2). After cleavage, the target protein can be purified from the tag by affinity chromatography. Protease can also be designed to contain a tag by which it can be separated from the target protein, and thus facilitate the purification (Nilsson *et al.*, 1997). For certain applications

Table 2. Enzymatic cleavage sites for fusion proteins. The cleavage window for protease is defined as P2-P1 ⇓ P1'-P2', where cleavage occurs at the bond between the P1 and P1' residues. X represent any aa.

Enzyme	Cleavage site [a]	Reference
Enterokinase	X-Asp-Asp-Asp-Asp-Lys ⇓ X	Hopp *et al.*, 1988
Thrombin	X-P4-P3-Pro-Arg ⇓ P1'-P2' where P4 and P3 hydrophobic and P1'-P2' nonacidic	Hakes and Dixon, 1992
Factor Xa	X- Ile-Glu (or Asp)-Gly-Arg ⇓ X (not Pro)	Pryor and Leiting, 1997
H64A subtilisin	X- Phe (Ala)-Ala-His-Tyr ⇓ X	Forsberg *et al.*, 1991
GST/His_6-protease 3C	X-Leu-Glu-Thr-Leu-Phe-Gln ⇓ Gly-X	Walker *et al.*, 1994
IgA protease	X-Pro-Pro ⇓ Thr-Pro-X	Pohlner *et al.*, 1992
Ubiquitin peptidase	Ubiquitin ⇓ relaxin α-chain	Miller *et al.*, 1989
Aminopeptidase	Poly-His ⇓	Pedersen *et al.*, 1999
TEV protease	X-Glu-Asn-Leu-Tyr-Phe-Gln ⇓ Gly	Rigaut *et al.*, 1999
ABP/His_6-protease 3C	X-Leu-Glu-Ala-Leu-Phe-Gln ⇓ Gly-Pro-X	Gräslund *et al.*, 1997

[a]For some of the enzymes, alternative sequences are also reported to be functional.

it is convenient to perform a site-specific cleavage when the fusion protein is still bound on the affinity column. An example of a protein purified by this method is the DNA binding domain of the glucocorticoid receptor fused to protein A (Dahlman *et al.*, 1989).

For some applications maintenance of an authentic N-terminus of the protein is required. For this purpose the initiation methionine can be removed by cyanogen bromide or by *in vitro* cleavage *e.g.*, within the sites of specific proteases such as aminopeptidase, enterokinase or factor X_a. The authentic N-terminus can also be obtained by using ubiquitin as a fusion partner (Miller *et al.*, 1989; Sabin *et al.*, 1989). The ubiquitin fused to human gamma-interferon and α1-proteinase was removed quantitatively *in vivo* by an endogenous ubiquitin-specific proteinase (Sabin *et al.*, 1989).

3. Analysis of posttranslational modifications

Many proteins are subsequently modified in various ways after or during the assembly of the polypeptide chain (Wold, 1981). The terminal methionine can be hydrolyzed, one or more N-terminal residues may be removed and/or polypeptides can be specifically cleaved. Disulfide bonds may be formed by oxidation of two cysteines and certain amino acids chains can be modified by *e.g.*, hydroxylation, sulfation, acylation, glycosylation and phosphorylation. Also prosthetic groups, such as lipoate, are covalently attached to some proteins. However, only the analysis techniques concerning three common posttranslational modification *i.e.*, glycosylation, acylation and phosphorylation are reviewed here.

3.1. Glycosylation

Many luminal, secreted and membrane proteins, as well as some cytoplasmic proteins, are modified by glycosylation. Together with glycolipids, membrane glycoproteins are responsible for many important functional properties of the eukaryotic cell surface, *e.g.*, cell-cell recognition and adhesion, blood group specific antigens, virus, bacteria and protozoan receptors, transplantation (histo-compatibility) antigens, ion channels and many others (Parekh, 1991). In general glycosylation is a complex process and varies in different cell types, and cannot be easily predicted by simple rules. Two types of carbohydrate-protein linkages are found: O-linked and N-linked. O-glycosidic bonds occur between the hydroxyl group of serine or threonine and an amino sugar, whereas N-glycosidic bonds are between the amide nitrogen of asparagine and

the C-1 of an amino sugar residue. N-glycosidic glycosylation can occur in the sequence motif of Asn-X-Ser, or Asn-X-Thr, where X can be any amino-acid except proline and asparagine. N-glycans can be subdivided into three distinct groups called 'high mannose type', 'hybrid type', and 'complex type' glucans.

The determination of oligosaccharide composition can be demanding since saccharides have very complex structure having branching and anomeric configurations (alpha and beta linkages). The complete characterization of the glycosylation of a glycoprotein entails the determination of three variables: Firstly, the occupancy of potential glycosylation sites; secondly, the type(s) of oligosaccharides present at these sites, and thirdly, the distribution of oligosaccharides structures at each site (Ashford, 1992). To study whether the protein is modified by glycosylation, metabolic labeling with radioactive saccharide and/or affinity chromatography (lectin, Con-A) can be utilized (Table 3). A radioactive sample can be obtained by labeling either metabolically or chemically, or a radioactive sugar nucleotide can be transferred from another saccharide using a glycosyltransferase. In many cases for the analysis of the saccharide composition and structure, it is necessary to release the saccharides from the glycoprotein. Methods used to release saccharides include the enzymatic approach (endo/exoglycosidases), chemical methods and digestion of the protein with proteases. Since the glycoside bond between the hydroxyl group and amino sugar can be disrupted by alkali treatment, O-linked saccharides are released by ß-elimination. However, this technique releases other O-linked residues *e.g.*, phosphate. Purified glycopeptides or saccharide moieties can be further analyzed by RP-HPLC, gas-liquid chromatography or thin layer chromatography (TLC). Mono- and disaccharides can also be determinated by paper chromatography. In addition, many inhibitors and comparative substrates can be utilized to alter the glycoconjugates, for example N-linked glycosylation can be prevented by tunicamycin and the elongation of the saccharide chain can be stopped at different stages by other specific inhibitors. Finally, there are many other approaches to analyze oligosaccharides such as the glycolipide anchor method, polysaccharide lyase, the phenolsulfuric acid assay, and mass spectrometry (MS).

3.2. Acylation

There are three common types of lipid modifications of cytoplasmic proteins: myristate (C14-saturated fatty acid), palmitate (C16-saturated fatty acid) and isoprenoids (farnesyl, geranyl) (Grand, 1989; Solski *et al.*, 1996).

Table 3. Different techniques for glycan analysis (modified from Merry, 1999).

Technique	Typical applications
Glycoprotein level	
Lectin analysis	Identification of type and amount of glucan present
Enzymatic deglycosylation (endoglycosidase F/H)	Degree and possibly information on the type of glycosylation
Capillary electrophoresis	Glycoform analysis
Mass spectrometry	Mass and site attachment of glucan
Glycan level	
Enzymatic release	N-glycans
Hydrazinolysis	N- and O-glycan release
HPLC	Quantitative profiling, sequencing, glucan analysis
Mass spectrometry	Mass profiling, glucan identification
NMR	Full structural analysis conformational studies

Some membrane proteins can also carry a stearate (C18-saturated fatty acid) and glycosylphosphatidylinositol anchor (Thomas *et al.*, 1990). Modification of the proteins by lipids appears often to be the first membrane binding step (McIlhinney, 1990). Lipid modifications with myristate or isoprenoids, are attached through chemically stable amine or thioether bonds, respectively, whereas palmitate is attached through a chemically fragile thioester linkage (Bizzozero, 1995). There is no recognition sequence for palmityolation, but often the palmitate is attached to cysteine residues (Dunphy and Linder, 1998). A conserved eight amino acid sequence is the minimum number of residues that provides an effective signal for myristoylation, but a large number of eukaryotic proteins are myristoylated at their N-terminal glycine-residue *via* amide linkage (Solski *et al.*, 1996). Prenyl groups are attached to C-terminal cysteine residues *via* a thioester linkage (Glomset *et al.*, 1990).

Preliminary indications that a protein is acylated can be provided by the incorporation of radiolabeled fatty acids or by mass spectrometry (MS; Neubert and Johnson, 1995). Acylation of a protein can be indicated when discrepancies between the observed molecular weight of a peptide and the calculated mass of the amino acids corresponds to the mass of an acyl sequence. The specific modified amino acid residues can be identified by tandem mass spectrometry (MS/MS) of proteolytic peptides (Hunt *et al.*, 1986). The incorporated fatty acid can also be identified by gas chromatography after hydrolysis and methyl esterification (Bizzozero, 1995). A sequential hydrolysis with acid, KOH/methanol and hydroxyamine might give information about the nature of the bonds between the acyl-group and the protein. Furthermore, the released fatty acids can then be identified by liquid chromatography or TLC. Nuclear magnetic resonance (NMR) has been used to study C-terminal alkylation with isoprenoid farnesyl or geranylgeranyl groups (Anderegg *et al.*, 1988).

3.3. Phosphorylation

Phosphorylation is generally considered to be the most prevalent posttranslational modification of proteins. It has been shown that many cellular processes such as signal transduction, enzymatic activities, protein translation and cell division are regulated by phosphorylation and dephosphorylation (Hunter, 1995). Proteins can be phosphorylated at serine, threonine and tyrosine to form O-phosphomonoesters. Besides the commonly occurring phosphorylated amino acid residues in proteins, phosphomono- and phosphodiesters have been found to be attached to sugar residues of some oligosaccharide side chains (Varki, 1998). One of the best known examples of the functional phosphate esters is the targeting of some hydrolytic enzymes to lysosome by mannose residues that are phosphorylated at the C-6 hydroxyl group, known as the mannose 6-P recognition signal. With radioactive labeling it is possible to determine if the protein is modified by phosphorylation. Other techniques to detect possible phosphorylation are dephosphorylation with phosphatases and immunodetection with anti-phosphoserine/threonine/tyrosine antibodies. To further study the function of phosphorylation it is necessary to identify the exact phosphorylation sites. This can be approached by site-directed mutagenesis, tryptic phosphopeptide mapping followed by Edman degradation or from pure peptide samples by mass spectrometry.

To analyze the type(s) of amino acids phosphorylated, the protein is subjected to acid hydrolysis, and phosphorylated amino acids are separated by chromatographic and/or electrophoretic systems. In addition to peptide bonds phosphodiester bonds are also unstable under the reaction conditions that lead to the release of free [^{32}P]phosphate. As a consequence, hydrolysis times are critical as usually no more than 20% of the radioactivity can be recovered as free phosphoamino acids. Sometimes one-dimensional migration is sufficient to

separate all three phosphoamino acids (Jelinek and Weber, 1993) but often two-dimensional (2D) analysis is required. 2D-systems can consist of two successive chromatography (Grimm and Nordeen, 1999), two successive electrophoresis or an electrophoresis followed by chromatography.

Peptide mapping or peptide fingerprinting is a powerful technique to analyze phosphoamino acid composition as well as other posttranslational modifications (*i.e.*, glycosylation, fatty acid acylation, methylation and acetylation). Phosphopeptide mapping (PPM) of a digested protein can give an overall view of how many sites are phosphorylated or it can be used to study the effects of different treatments and the site-directed mutagenesis on the relation of phosphorylated peptides. 2D-mapping is an extremely sensitive technique that requires only minor amounts of metabolically labeled product. Resolving the peptide mixture in two dimensions often yields subtle but important clues about peptide composition. Because cellulose is an inert material, the peptides can be recovered from the plate for further analysis, such as amino acid analysis or Edman degradation.

Two-dimensional electrophoresis is an effective method for the analysis of a broad range of complex protein samples, and posttranslational modifications like phosphorylation. The technique sorts proteins into two dimensions according to their isoelectric points (IEF) and sizes (SDS-PAGE). Drawbacks of this method are its extreme sensitivity to salts and other impurities as well as its unsuitability for membrane proteins. An improved variation of 2D electrophoresis utilizes immobilized pH gradients by which the inconsistencies among carrier ampholytes are circumvented (Görg *et al.*, 1988).

Mass Spectrometry (MS) is a rapid and extremely sensitive tool for the determination of posttranslational modifications, including phosphorylation (Costello, 1999). By utilizing techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) MS is able to analyze a sample below the picomolar level (Beavis and Chait, 1996; Banks and Whitehouse, 1996). MS with collision-induced dissociation (CID) after phosphopeptide precursor-ion scan of m/z of 79 can be utilized to sequence phosphopeptides (Carr *et al.*, 1996). However, identification can sometimes be ambiguous, especially when a protein/peptide exists in complex mixtures. Peptide characteristics can make the identification process difficult or occasionally impossible, *i.e.*, negatively charged peptides may be poorly ionized in the positive mode thus preventing sequence analysis by CID MS. Moreover, large peptides will not release an informative fragmentation pattern and small peptides (under 600 kDa) are more likely to be lost within the region of matrix signals. To avoid adventitious formation of acrylamide adducts during electrophoresis, alkylation with iodo-acetamide, 4-vinylpyridine and acrylamide can be performed prior to SDS-PAGE (Sechi and Chait, 1998). The peptides for MS can be purified by RP-HPLC either off- or on-line. Phosphopeptides can be also purified by binding them to an iron-chelating column (IMAC). MS can be used to identify phosphorylation sites alone or in combination with other methods such as phosphopeptide mapping, 2D-electrophoresis and Edman degradation. Phosphoprotein or phosphopeptides prepared by enzymatic (phosphatase treatment) or chemical degradation (ß-elimination) can be identified by their m/z of 79 $[PO^3]^-$ or 97 $[H_2PO^4]^-$ mass difference compared to the peptides expected from the sequence (Neubauer and Mann, 1999). An elegant way of determining phosphorylation sites utilizes on-line electrospray mass spectrometry, connected to an IMAC column and/or micro liquid chromatography column.

AIMS OF THE STUDY

This study focused on the expression and purification of truncated yeast secretory protein Hsp150, Hsp150Δ-$NGFR_e$ fusion protein and Semliki Forest virus nonstructural proteins (Nsp1-Nsp4). The purpose was to analyze the specific functions and structural features of these proteins. Specifically, the aims were:

- to express and purify truncated Hsp150 for structural studies (I)
- to evaluate the use of Hsp150Δ as a carrier for the ectodomain of rat nerve growth factor receptor ($NGFR_e$) (II)
- to characterize N- and O-glycosylation of Hsp150Δ-$NGFR_e$ (III)
- to express and purify SFV Nsp1-4 for enzymatic and structural studies (unpublished)
- to map the amino acids in SFV Nsp1 involved in methyltransferase and guanylyltransferease activities (IV)
- to identify and eliminate the phosphorylation sites in SFV Nsp3 (V)
- to study the effect of Nsp3 phosphorylation on membrane association, virus life cycle and neuropathogenicity (VI)

MATERIALS AND METHODS

1. Materials

1.1. Strains, cells and media

Plasmid construction was performed using the *E. coli* strain DH5α (Bethesda Research Laboratories or Life Technologies Inc.). *E. coli* strains BL21 (DE3, Novagen Promega Corp.) and JM109 (DE3, Novagen Promega Corp.) were used for protein expression of SFV Nsp1-3 and their mutations. The strains were grown at 37°C in L-broth medium supplemented with ampicillin (100 µg/ml).

The *S. cerevisiae* strains used are listed in Table 4 and described in detail in I-III. The yeast cells were grown at 24°C in YPD-medium or in synthetic complete (SC) medium (Sherman, 1991). Strains H487 and H640 were grown in SC medium lacking tryptophan.

Spodoptera frugiperda (Sf9) insect cells and *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) were kindly provided by Dr. Christian Oker-Blom (University of Jyväskylä, Finland). The Sf9-cells were grown at 28°C as monolayer cultures in TNM-FH medium (Norvacc Media) containing 10% fetal calf serum (PAA Laborund Forschungsges M.B.H.), 100 U/ml of penicillin, 100 µg/ml of streptomycin and 300 µg/ml of glutamine. The propagation of AcMNPV has been previously described (Summers and Smith, 1988).

HeLa cells were grown in Dulbecco's modified minimal essential medium, supplemented with 10% inactivated fetal calf serum (PAA Laboratories GmbH) and 100 U/ml of penicillin and streptomycin (Gibco BRL). The origin and culturing of baby hamster kidney cells (BHK21) has been previously described (Keränen and Kääriäinen, 1974).

The Semliki Forest virus (SFV) prototype strain was propagated in BHK cells as previously described (Keränen and Kääriäinen, 1974). Vaccinia virus vTF-3 and modified vaccinia Ankara virus MVA/T7 were kindly provided by Dr. B. Moss (National Institutes of Health, Bethesda, MD, USA) and propagated in HeLa and BHK cells, respectively, as described by Fuerst *et al.* (1986) and Wyatt *et al.* (1995).

Table 4. *S. cerevisiae* strains.

Strain	Genotype
H1	SEy2101a, *Mat***a** *ade2-101 ura3-52 leu2-3,112 suc2*Δ*9 gal2*
H3	SF821-8A, *Mat***a** *sec7-1 his4-580 ura3-52 leu2-3,112 trp1-289*
H4	mBy12-6D, *Mat*α *sec18-1 trp1-289 leu2-3,112 ura3-52 his*$^-$
H23	*Mat*α *his3-11,15 leu2-3,112 trp1-1 ade2-1 can1-100 HSP150::URA3*
H431	*Mat*α *his3-11,15 trp1-1 ade2-1 can1-100 HSP150::URA3 LEU2::*truncated *HSP150*
H433	*Mat*α *sec18-1 trp1-289 his*$^-$ *HSP150::URA3 LEU2::*truncated *HSP150*
H426	*Mat*α *his3-11,15 trp1-1 ade2-1 can1-100 HSP150::URA3 LEU2::HSP150*Δ*-NGFR*$_e$
H451	*Mat*α *his3-11,15 leu2-3,112 trp1-1 ade2-1 can1-100 ura3-1 HSP150::URA3 LEU2::HSP150*Δ*-fX*$_a$*-NGFR*$_e$
H487	*Mat*α *his3-11,15 leu2-3,112 trp1-1 ade2-1 can1-100 ura3-1 HSP150::URA3* [pKTH4616] (plasmid containing *SUC2- NGFR*$_e$)
H538	*Mat*α *sec18-1 trp1-289 HSP150::URA3 his*$^-$ *LEU2::HSP150*Δ*-NGFR*$_e$
H640	*Mat*α *sec18-1 trp1-289 leu2-1,112 his*$^-$ *HSP150::URA3* [pKTH4616]

1.2. Plasmid constructions

All DNA manipulations were done according to standard methods (Sambrook *et al.*, 1989). DNA fragments from the agarose gel were isolated using a commercial kit (Qiagen) and nucleotide sequences were determinated by an automated ALF DNA sequencer (Pharmacia LKB).

1.2.1. Yeast (I-III)

S. cerevisiae strains H431, H433 and H426 were constructed with a shuttle vector derived from YEp24 (Botstein *et al.*, 1979) by removing the *bla* gene and the 2µ replication origin (Simonen *et al.*, 1994). A detailed description of plasmid constructions is given in I-III.

1.2.2. *E. coli* (IV, unpublished)

Nsp1-Nsp3 were expressed using vector pHAT coding for six histidines in a row at the 5' end (Peränen *et al.*, 1996, Hyvönen, unpublished). The sequences encoding Nsp1, Nsp3 (*Nco*I-*Hind*III fragment) and Nsp2 (*Nco*I-*Bgl*II fragment) were cloned under the control of a T7 promoter, which could be induced by IPTG. Nsp1 and its mutants were expressed using the expression vector pBAT4 (Peränen *et al.*, 1996).

1.2.3. Baculovirus system (unpublished)

The gene fragment encoding for Nsp4 was amplified by PCR using linearized pPLH214 as a template (Peränen *et al.*, 1990). The PCR-fragment was cloned into a pVL1392 transfer plasmid (kindly provided by Dr. M. Summers, Texas A&M University, USA) to generate recombinant the plasmid pBNS4 (Laakkonen, unpublished). Sf9 cells were co-transfected with AcMNPV and pBNS4 by lipofectin according to the manufacturer's instructions. The recombinant virus AcMNPV-Nsp4 was isolated and subcloned as described by Summers and Smith (1988).

1.2.4. Vaccinia infection/DNA-transfection (IV- IV)

The vector used for DNA transfection (pTSF) was a derivative of pGEM3 (Promega), where the multiple cloning site has been changed into a site encoding the following restriction sites: *Nco*I-*BamH*I-*Bgl*II-*Hind*III (Takkinen *et al.*, 1990). The gene encoding Nsp3 (*Nco*I-*Hind*III) was cloned into pTSF under the control of T7 promoter (Peränen, 1991).

2. Methods

2.1. Deletion constructs and point mutations (IV-VI)

Deletion and point mutations of Nsp3 were made using pTSF3. Briefly, a linearized plasmid was used to amplify fragments of Nsp3 by the polymerase chain reaction, and primers which generated the mutations used are listed in V and VI. Point mutations made in Nsp3 and Nsp1 were made using the unique site elimination method according to the manufacturer's instructions (U.S.E. mutageneis kit, Pharmacia Biotech). All the mutations were verified by sequencing.

2.2. Immunoprecipitation, immunoblotting and SDS-PAGE

The immunoprecipitation methods have been previously described (Peränen *et al.*, 1988). Proteins were analyzed by SDS-PAGE in 8% or 10% gels according to Laemmli (1970). Western analysis of proteins were done on gels blotted onto nitrocellulose membranes (Hybond ECL, Amersham) and visualized by using the ECL-kit (Amersham). Anti-Nsp3 antibody was used in a 1 to 10,000 dilution. Horse radish peroxide-conjugated anti-rabbit antibody was used as a second antibody in a 1 to 3,000 dilution.

2.3. Protein expression and purification

2.3.1. Yeast (I-III)

Yeast cells were grown overnight and the supernatant was subjected to further purification as illustrated in Fig. 7. Briefly, the proteins were concentrated by filtration or ion exchange chromatography (IEC) and further purified by gel filtration and IEC. Between different purification steps the sample was concentrated with membrane filtration (PM10, Centricon, Amicon). In IEC and gel filtration, detection was at 280 nm, in RP-HPLC detection was at 216 nm.

Authentic Hsp150 was initially purified using IEC (DEAE cellulose). Hsp150 was collected from the flow-through, and subjected to dialysis against 10 mM Tris-HCl pH 8.0. After concentration, Hsp150 was purified by IEC (Q-Sepharose) equilibrated with 10 mM Tris-HCl pH 8.0. Elution was performed using an increasing gradient of NaCl, and fractions containing Hsp150 were pooled and desalted by gel filtration (Bio-Gel P100). The preparation was again subjected to IEC (Mono Q), and eluted under the same conditions as truncated Hsp150 and Hsp150Δ-NGFR$_e$ (Fig. 7).

2.3.2. *E.coli* (IV, unpublished)

BL21 (DE3) or JM109 (DE3) cells carrying the plasmid encoding target gene under T7 promoter, were grown at 37°C to an optical density of 0.7-1.0 (measured at 600 nm). The culture was transferred to 15°C or

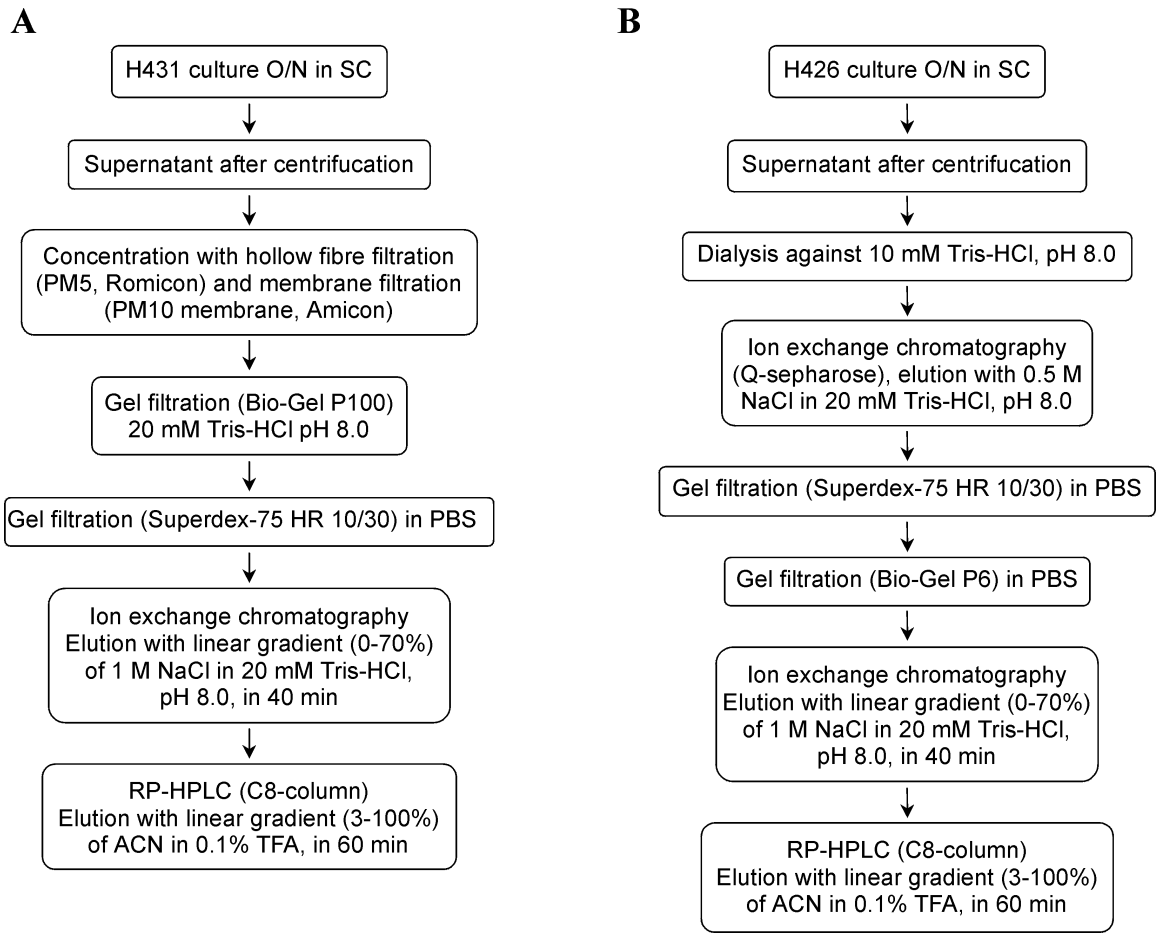

Figure 7. The purification schemes of truncated Hsp150 (**A**) and Hsp150Δ-NGFR$_e$ (**B**).

to 24°C, in the case of N-Nsp2, and production was induced by IPTG (see the results section for the concentrations used in different cases). After an 18 h incubation, 6 h in the case of N-Nsp2, the cells were harvested by centrifugation at 5000 rpm for 5 min (JA-14 rotor, Beckman) and lysed by French Press (550 Psi, three times; French Pressure Cell Press, SLM Instruments Inc.). The lysis buffer used contained: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 20% glycerol, 0.1% tween 20, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM dithiothreitol (DTT). The lysate was then centrifuged for 20 min at 15 000 x g at 4°C and the supernatant, designated as S15, was used for further studies or purification steps.

N-Nsp2 and Nsp3 were purified under nondenaturing conditions with immobilized metal affinity chromatography (IMAC) at 10°C followed by gel filtration. Generally, 0.5 M NaCl was added to the S15 fraction and after 30 min incubation on ice the sample was applied to a Ni^{2+}-NTA-column (Qiagen) equilibrated with buffer A (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 0.1% tween, 20% glycerol). The column was washed with 10 vol. of buffer A and stepwise eluted with 20-250 mM imidazole in buffer A. In the case of N-Nsp2, before elution the column was washed with 5 vol. of buffer A containing 5 mM ATP and 5 mM $MgCl_2$ followed by washing with 10 vol. of buffer A. Nsp1 was purified under denaturing conditions in 6 M guanidinium hydrochloride (Gnd-HCl), 10 mM Tris, 100 mM Na-phosphate, pH 8.0 or 7.5. The solubilized protein was applied to a Ni^{2+}-NTA column, and the buffer was changed to buffer B (8 M urea, 10 mM Tris, 20% glycerol, 100 mM Na-phosphate, pH 8.0 or 7.5). The amount of urea was reduced by a linear gradient (0-100%) of buffer C (20% glycerol, 50 mM Tris pH 8.0 or 7.5, 0.5 M NaCl, 0.4 M urea or no urea) in buffer A. Nsp1 was eluted by 167 mM or 250 mM imidazole in buffer C.

Gel filtration was performed in buffer containing 50 mM Tris-HCl, pH 7.5 or 8.0, 20% glycerol at a flow rate of 0.5 ml/min and detection was at 280 nm. The molecular weight of the purified proteins were estimated according to a molecular weight standard consisting of thyroglobulin, 670 kDa, gammaglobulin, 158 kDa, ovalalbumin, 44 kDa, myoglobin, 17 kDa and vitamin B-12, 13.5 kDa (Bio-Rad).

2.3.3. Sf9 cells (unpublished)

Sf9 cells grown in plastic bottles were infected with AcMNPV-Nsp4 at 50 PFU/cell with occasional shaking. After 2 hours p.i. 8 ml of medium was added and the cells were collected at 36-48 h p.i. in cold PBS and washed twice with RS-buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 1 mM PMSF). The cells were harvested by centrifugation at 5000 rpm for 5 min (Sorvall) at 4°C and resuspended in RS-buffer. After incubation on ice for 15 min, the cell were lysed using a Dounce homogenizator. The cell lysate was fractionated by centrifugation and Nsp4 was concentrated by flotation on a sucrose gradient, *i.e.*, the sample was suspended in 67% sucrose to a final sucrose concentration of 55% and layered on top of 60% sucrose and overlayered by 10% sucrose in 50 mM Tris-HCl pH 7.5, 25 mM KCl, 5 mM $MgCl_2$. Nsp4 was also subjected to solubilization by adding Triton X-100 to 0.5% prior to flotation. The discontinuous gradient was centrifuged at 35,000 rpm (rotor SW50.1, Beckman) for half of hour at 4°C. Several detergents *e.g.* Triton X-100, octylglycoside (17.1 mM), Thesit (8.6 mM), Chaps (8.1 mM) in the presence (1 M or 0.5 M) or absence of NaCl as well as sonication were used in order to solubilize the protein.

2.3.4. HeLa cells (IV-VI)

The cells were grown on plastic dishes until approximately 90% confluent. The cells were then infected with MVA or vTF7-3, 30-

50 PFU/cell at 37°C for 45 min with shaking. The cells were washed twice with Dulbecco-BSA and transfection with plasmid containing the desired gene under the T7 promoter was carried out in OPTIMEM1 using lipofectin (Gibco BRL). The DNA and lipofectin amounts used for 3 cm, 6 cm and 10 cm plastic plates were 1-2 µg and 5 µl, 7 µg and 12 µl, 10 µg and 50 µl, respectively. After 3 hours post transfection the transfection mixture was replaced with MEM medium and protein production continued for another 2-4 hours. Unless otherwise stated the cells were harvested in PBS (65°C), containing 1% SDS and 10 mM NaF (Merck).

For mass spectrometric analysis Nsp3 was immunoprecipitated and subjected to 10% SDS-PAGE. To obtain peptides suitable for MS, Nsp3 was subjected to CNBr degradation followed by trypsin digestion. The purification scheme is described in Fig 8.

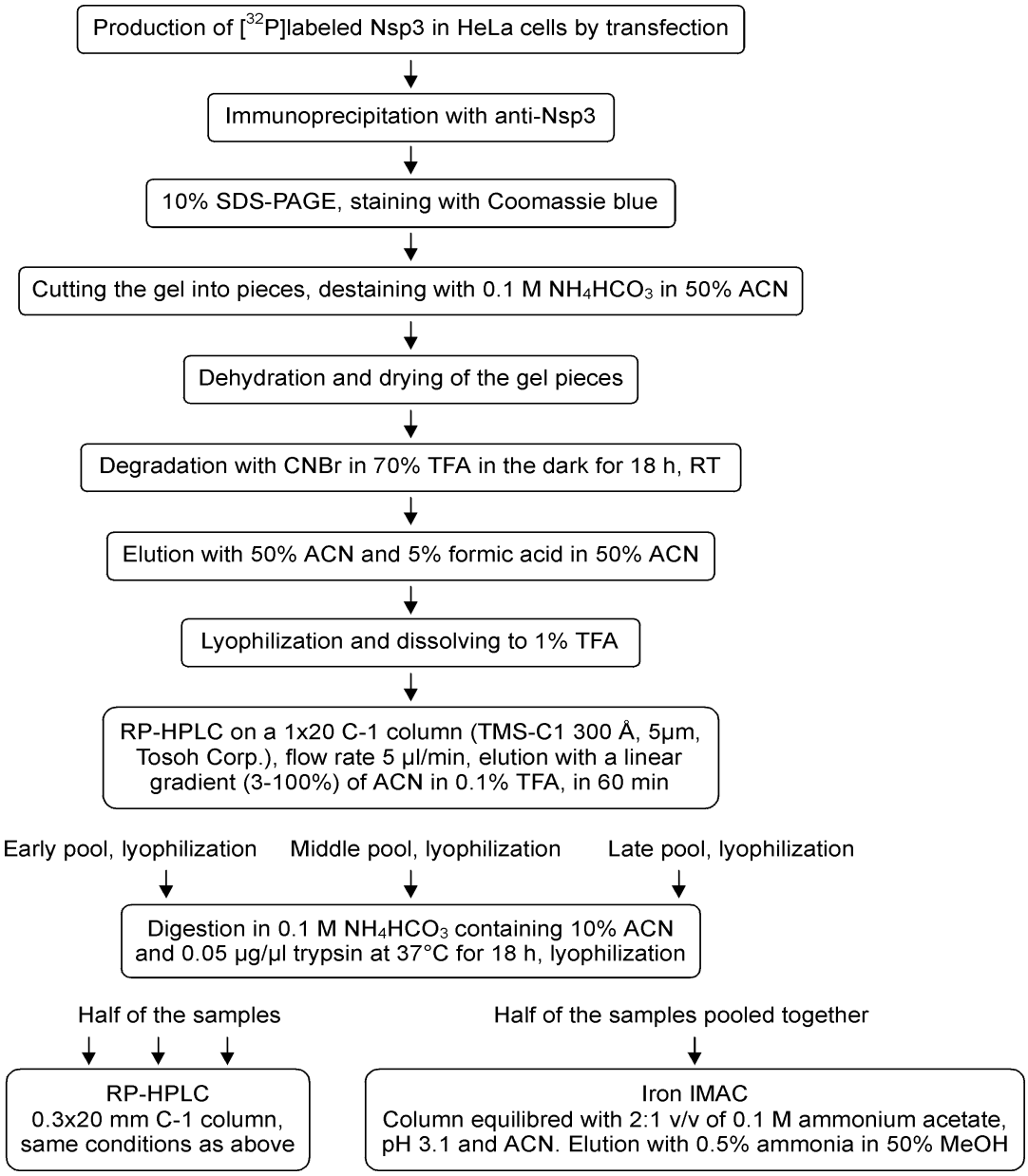

Figure 8. The purification scheme of Semliki Forest virus Nsp3 for mass spectrometric analysis.

2.4. Metabolic labeling (I-III,V,VI)

Yeast cells (10^8/200 µl) were labeled in synthetic medium lacking methione with 50 µCi/ml of [^{35}S]methionine/cysteine (>1000 Ci/mmol, Amersham). For glycan analysis the yeast cells were labeled with 300 µCi/ml of [^{3}H]mannose (11.5 Ci/mM, Amersham) in medium including 0.1% glucose.

HeLa cells were labeled with [^{35}S]methionine/cysteine (>1,000 Ci/mmol; Amersham) or [^{32}P]orthophosphate (10 mCi/ml; Amersham) as described in detail in V. The fluorographic detection was performed by Fuji Bas 1500 Bioimaging Analyzer instrument.

2.5. Immunofluorescence (VI)

For immunofluorescence HeLa cells were grown to low confluence on coverslips and transfected with pTSF -plasmid containing Nsp3 or its mutants as described in the section for protein expression and purification. After 3-4 hours post transfection the cells were fixed for 20 min with a solution containing 4% paraformaldehyde, 11.1% (w/v) sucrose 10 mM MES, 138 mM KCL, 3 mM $MgCl_2$ and 2 mM EGTA. After the fixation cells were permeabilized with 0.1% Triton X-100 for 1 min. Anti-Nsp3 antisera were raised in a rabbit or in a guinea pig against recombinant Nsp3. To avoid unspecific binding both antibodies were purified by adsorption to uninfected PFA-fixed and permeabilized HeLa cell monolayers. The antibodies were used in a 1 to 500 dilution in Dulbecco including 0.2% BSA. The secondary antibody was donkey rhodamine red-X conjugated anti-rabbit IgG (Jackson Research Laboratories Inc.) used in a 1 to 300 dilution. Microscopy was performed using a Bio-Rad MRC1024 confocal microscope (Cambridge, Mass.) connected to a Zeiss Axiovert M135 inverted microscope (Göttingen).

2.6. Flotation analysis (VI)

Flotation studies of Nsp3 and N-Nsp3, expressed in HeLa, were performed with protein metabolically labeled with [^{35}S]methionine/cysteine. The cells were harvested in RS-buffer and after a 15 min incubation on ice the cells were lysed with a Dounce homogenizator. A sucrose solution (67%) was added to the post nuclear supernatant (PNS, 10 min at 500 x g) to a final concentration of 60%. The sample (0.5 ml) was layered on top of a 0.5 ml 67% (w/w) sucrose cushion and overlaid by 3 ml of 50%, and 1 ml of 5% sucrose, modified from Caliquiri and Tamm (1970). The sucrose solutions were prepared in buffer containing 50 mM Tris-HCl, pH 7.5 and 100 mM or 500 mM NaCl. The discontinuous gradient was centrifuged at 35,000 rpm for 18 hr at 4°C (SW50.1 rotor, Beckman). Fractions of 0.5 ml were collected from the bottom of the tube and the presence of Nsp3 or N-Nsp3 was analyzed by immunoprecipitation followed by SDS-PAGE. The quantitation of proteins was performed utilizing digital imaging.

2.7. Phosphoamino acid analysis (V)

Phosphoamino acids were analyzed by 2-dimensional phosphoamino acid analysis. [^{32}P]-labeled proteins were Western blotted onto polyvinylidene difluoride (PVDF, Immobilon) membrane and the protein band were excised and rehydrated with methanol and water. Hydrolysis was carried out in 6 M HCl at 110°C for 120 min. The hydrolysate was lyophilized and phosphoamino acid standards (1 µg of phosphoserine, phosphothreonine and phosphotyrosine) were added. The separation of amino acids was performed using 2D-electrophoresis on a 20 cm x 20 cm cellulose plate (Merck). Electrophoresis in the first dimension was at 1.0 kV for 120 min in pH 1.9 buffer (88% formic acid:acetic acid:water, 25:78:897 v/v/v) and

in the second dimension at 1.3 kV for 45 min in pH 3.5 buffer (acetic acid:pyridine :water, 10:1:189 v/v/v), using the HTLE-7000 electrophoresis system (C.B.S. Scientific company, Del Mar, CA). The amino acid standards were detected by ninhydrin (0.2% ninhydrin (w/v) in ethanol:acetic acid, 19:1 v/v) staining, and [^{32}P]-labeled amino acid by phosphoimager.

2.8. Phosphopeptide mapping (V)

[^{32}P]-labeled Nsp3 or its mutants were immunoprecipitated with anti-Nsp3 as described above. Cysteines were alkylated with 20 mM iodoacetamide (10 mM DTT, 3 mM EDTA) for 15 min and resolved by 10% SDS-PAGE. Proteins were Western blotted onto nitrocellulose (Hybond C-extra, Amersham) or polyvinylidene difluoride membrane (PVDF, Immobilon). The PVDF membrane was dried prior to detection by phosphoimager, but the nitrocellulose membrane was prevented from drying by placing it into a plastic bag. The membrane pieces were excised and saturated with 0.5% polyvinylpyrrolidone 360 at 37°C for 1 h followed by several washes with 10% ACN. The protein was digested with trypsin (Promega) in 50 mM ammonium bicarbonate in 10% ACN -buffer overnight at 37°C and the bicarbonate buffer was evaporated by several rounds of lyophilization. The tryptic peptide mix was resuspended in ammonium carbonate buffer pH 8.9 and spotted onto a TLC plate (20 x 20 cm, cellulose, Merck). Electrophoresis in the first dimension was performed at 1.0 kV for 24 min in the same buffer, using a HTLE-7000 electrophoresis system (C.B.S. Scientific Company, Del Mar, CA). The plate was air-dried overnight before ascending chromatography in n-butanol : pyridine : acetic acid : water (15:10:3:12, v/v/v/v). After air-drying the plate, the phosphopeptides were detected by a phosphoimager.

2.9. Other specific methods

Other methods used in this study are listed in Table 5.

Table 5. Other methods used in publications I-VI.

Method		Publication
Amino-terminal sequence analysis		I
CD analysis		I
Cross linking assay	Nsp1	IV
	Hsp150Δ-NGFR$_e$	II
Edman sequencing		V
Factor X$_a$ digestion		I
Glycan analysis		I
Methyltransferase activity assay		IV
Mass spectrometric analysis		V
NMR analysis		III
Phosphatase treatment		V
Plaque assay		VI
RNA synthesis		VI
In vitro transcription		VI
Trichloroacetic acid precipitation		III

RESULTS

1. HSP150Δ AS A CARRIER FOR SECRETION OF RECOMBINANT PROTEINS IN *S. CEREVISIAE*

To obtain proper folding and secretion of foreign proteins in yeast, most proteins must be fused to a carrier polypeptide. A novel carrier polypeptide (Hsp150Δ) was developed by Simonen *et al.* (1994) from a natural yeast secretory glycoprotein Hsp150 (Russo *et al.*, 1992a). The primary translation product of the *HSP150* gene consists of a signal sequence (SS) and two subunits (I and II), which remain noncovalently associated after cleavage by the Kex2 protease in the Golgi (Fig. 9) (Russo *et al.*, 1992a, Jämsä *et al.*, 1994). Subunit II contains a repetitive region consisting of a 19 amino acid peptide repeated 11 times followed by a unique C-terminal region. Furthermore, subunit II is heavily O-glycosylated, since its apparent molecular weight determined by SDS-PAGE was reduced by hydrogen fluoride treatment from 150 kDa to 47 kDa, which corresponds to its predicted molecular weight (Russo *et al.*, 1992a). Hsp150 is secreted to the culture medium rapidly, with a half-time of 2 min (Jämsä *et al.*, 1994). The expression of Hsp150 is upregulated by heat and nitrogen starvation (Russo *et al.*, 1992; 1993). The Hsp150Δ-carrier consists of 321 N-terminal amino acids of Hsp150, which includes an N-terminal 18 aa signal sequence (SS), subunit I (54 aa), the repetitive region and 22 aa from the non-repetitive region of subunit II (Fig. 9).

1.1. CHARACTERIZATION OF HSP150Δ-CARRIER (I)

To understand better the properties of Hsp150Δ as a carrier, we expressed a truncated form of Hsp150 in yeast from which the *HSP150* gene was deleted. The truncated Hsp150 consists of 315 N-terminal aa of Hsp150 and two additional aa (Fig. 9). This

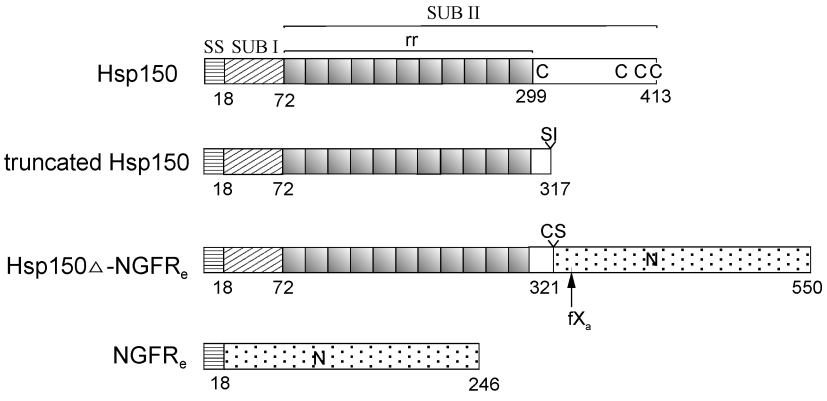

Figure 9. A schematic presentation of primary translation products of the coding for Hsp150 protein, truncated Hsp150, Hsp150Δ-$NGFR_e$ and $NGFR_e$. SS = signal sequence peptide (horizontal striped box), SUB I/II = subunit I/II (diagonally striped box / gray boxes together with white box), rr = repetitive region of subunit II (gray boxes), C = cysteine, S = serine, I = isoleusine. The potential N-glycosylation site in $NGFR_e$ (dotted box) is indicated by an N-letter. The factor X_a (fX_a) cleavage site (isoleusine-glutamate-glysine-arginine) in Hsp150Δ-fX_a-$NGFR_e$-construct is indicated by an arrow.

protein was purified along with the authentic Hsp150 and subjected to circular dichroism (CD) analysis. The ability of Hsp150Δ to act as a carrier was further studied by fusing it to the ectodomain of the rat nerve growth factor receptor $NGFR_e$.

1.1.1. Expression of truncated Hsp150 (I)

The gene encoding truncated Hsp150 was integrated into the *leu2* locus in strain H23 lacking the chromosomal *HSP150* gene (Russo *et al.*, 1992). It was placed under the control of the *HSP150* promoter. However, according to Northern blotting analysis, the truncated *hsp150* was not heat-regulated like the authentic gene, but was constitutively expressed at 24°C, at levels higher than authentic the *HSP150*. Because the truncated Hsp150 lacks cysteine residues, it was metabolically labeled with [^{3}H]mannose in low glucose (0.1%) medium to enhance labeling. To study the expression and secretion of truncated Hsp150, the protein was labeled and chased for different times with cycloheximide (CHX) (I, Fig. 1B). Samples from culture medium and cell lysates were immunoprecipitated using anti-Hsp150-antibody, and analyzed by SDS-PAGE (I, Fig. 1C). Truncated Hsp150 migrated in gels as a diffuse band with an apparent molecular weight over 200 kDa. Truncated Hsp150 was not as efficiently secreted as the wild type Hsp150, since less than half of the truncated Hsp150 was secreted to the medium compared to the authentic Hsp150.

1.1.2. Purification of truncated and authentic Hsp150 (I)

Strain H431, expressing truncated Hsp150 instead of Hsp150, was grown overnight in SC medium and truncated Hsp150 was purified from the medium under non-denaturing conditions for structural analysis (Materials and Methods, Fig. 7A). In brief, cell culture concentrate was then subjected to gel filtration and the fractions containing the protein were further concentrated and subjected to ion exchange chromatography (IEC). With hollow fibre filtration roughly half of the material was lost with the permeate, whereas with the membrane filter no flow through was detected. Elution in IEC was performed with a linear gradient of 0-70% of 1 M NaCl. Samples were analyzed by SDS-PAGE, followed by detection by Western blotting. Two layers of membranes (nitrocellulose) were used in immunoblotting due to the facile penetration of the protein. The majority of truncated Hsp150 eluted in the void volume (I, Fig. 2A, peak 2.5), whereas a small portion eluted later (I, Fig. 2A, peaks 13.7-17.7). The IEC fractions containing the protein were pooled, concentrated and gel filtrated. Four fractions were collected according to the absorbance at 280 nm. Three out of the four fractions contained truncated Hsp150 (data not shown), but due to the protein's heterogeneous migration on gel filtration, only the fraction eluting at 15.8-17.7 min was used for further analysis (I, Fig. 2B). Approximately 10% of this fraction was further purified by reversed phase high performance liquid chromatography (RP-HPLC) (I, Fig. 2C). Three main peaks (I, Fig. 2C) were collected and subjected again to RP-HPLC. N-terminal sequencing from all three peaks gave identical results, suggesting that the differences observed during the purification steps were due to heterogeneous glycosylation. In authentic Hsp150, subunit I is cleaved off in late Golgi, but remains non-covalently attached to subunit II (Russo *et al.*, 1992a). Since subunit I was not detected in any of the purification steps of truncated Hsp150, it is apparently attached to the C-terminal part of authentic Hsp150.

Authentic Hsp150 was also purified for structural characterization. When Hsp150 was expressed in *sec7* mutant cells, it was not retained in the Golgi complex under restrictive conditions like other secretory proteins, but was secreted into the medium (Russo *et al.*, 1992a). Thus, to facilitate its

purification, Hsp150 was expressed in SF821-8A cells (*sec7*) and purified from YPD and SC medium at 37°C. Hsp150 was purified from the supernatant with three different IECs (for details see Materials and Methods). In the last IEC (Mono Q) the proteins eluted with an increasing gradient of NaCl. (I, Fig. 3A). According to N-terminal sequence analysis, the major elution peak contained both Hsp150 subunits, which could be resolved by RP-HPLC into separate fractions (I, Fig. 3B).

1.1.3. Structural features (I)

The circular dichroism (CD) spectrum of purified, truncated subunit II was then determined. Half of the fraction identified in I, Fig. 2B, was analyzed by CD. The spectrum of truncated subunit II had its minimum at 200 nm, demonstrating that it lacked any regular secondary structure (I, Fig. 5A). When authentic Hsp150, containing subunits I and II non-covalently associated to each other, was analyzed, both random coils and some ß-sheets were detected (I, Fig. 5B). Thus, the ß-sheet elements evidently were in subunit I and/or the C-terminal region of subunit II.

To analyze the secondary structure of the consensus peptide of the repetitive region of Hsp150 subunit II (I, Fig. 6A and 6B), a consensus 19 aa peptide was synthesized and subjected to NMR spectroscopy. The peptide lacked any preferred secondary structure (I, Fig. 7A and 7B). Thus, the repetitive region of subunit II has a random coil structure. This is due to the amino acid sequence alone, and not to the O-glycans.

To characterize the O-glycans of Hsp150 subunit II, authentic Hsp150 was produced in SF821-8A cells, labeled with [^{3}H]mannose and subjected to SDS-PAGE. Labeled subunit II was eluted from the gel, and the buffer was changed by gel filtration (Bio-Gel P6). The O-glycans were detached from the polypeptide by ß-elimination and purified by gel filtration (Bio-Gel P6 I, Fig. 4A). According to thin layer chromatography (TLC), the O-glycans were di-, tri-, tetra- and pentamannoses presented in the ratio of 4:1:1:1, respectively (I, Fig. 4B and 4C).

1.2. Secretion and authenticity of Hsp150Δ-NGFR$_e$(II)

To further study Hsp150Δ as a carrier, it was fused to the N-terminus of the extracellular domain of the rat nerve growth factor receptor p75 (NGFR$_e$). The fusion protein was designated Hsp150Δ-NGFR$_e$ (Fig. 9). NGFR$_e$ is a plasma membrane protein, and binds all known neurotrophic factors with low affinity (Chao, 1992). NGFR$_e$ is a cysteine rich (24 cysteines) protein, and has a high sequence homology to the ectodomain of tumor necrosis factor receptor (TGFR$_e$) (Radeke *et al.*, 1987). Based on the crystal structure of TGFR$_e$ and molecular modeling, NGFR$_e$ is supposed to consists of four domains, each having three disulfide bonds arranged into an elongated structure (Holkeri *et al.*, 1998). NGFR$_e$ contains one potential N-glycosylation site at Asn^{32} (Baldwin *et al.*, 1992).

1.2.1. Expression of Hsp150Δ-NGFR$_e$ (II)

To study whether Hsp150Δ can act as a carrier for NGFR$_e$ and lead it out from the yeast ER to the culture medium, a DNA fragment encoding NGFR$_e$ (aa 1-223) was fused to the C-terminal end of *HSP150Δ*. The chimeric gene was integrated into the genome of strain H23 (Russo *et al.*, 1992a), which contains a disrupted copy of the *HSP150* gene, to create strain H426. Strain H426 was grown in SC medium to different cell densities, and samples from the medium were analyzed for the fusion protein by SDS-PAGE followed by Coomassie blue staining. The fusion protein appeared as a heavy band at the top of the gel indicating that it was efficiently expressed, stable, heterogenous and efficiently secreted into the medium. It was the major protein in the medium (several

milligrams per liter). Only small portions of the fusion protein remained cell-associated. Anti-Hsp150 antiserum and anti-$NGFR_e$-antiserum recognized two species of the fusion protein, *i.e.*, a band at approximately 180 kDa and a larger protein. The larger form was N-glycosylated, since the synthesis of this form could be prevented by tunicamycin (TM), an inhibitor of N-glycosylation. Inhibition of N-glycosylation did not prevent the fusion protein from being secreted.

To study secretion of $NGFR_e$ in the absence of Hsp150, it was fused to the signal peptide of yeast invertase (Fig. 9). When expressed in *S. cerevisiae*, $NGFR_e$ was translocated into the yeast ER, since it was N-glycosylated (II, Fig. 2B). However, no $NGFR_e$ was detected in the medium. Apparently it remained in the ER because the N-glycan was not elongated. Thus, $NGFR_e$ was able to acquire a secretion-competent conformation only when fused to the Hsp150Δ-carrier.

1.2.2. Purification of Hsp150Δ-$NGFR_e$ (II)

Hsp150Δ-$NGFR_e$ was purified from the culture medium (see Materials and Methods, Fig. 7B) of strain H426. In brief, the protein was purified with IEC (Q-sepharose) and gel filtration (Superdex-75). In gel filtration the fusion protein eluted in the void volume, in an estimated concentration of 0.4 µg/µl. This preparation is referred to as semipurified Hsp150Δ-$NGFR_e$, and it was used for ligand-binding assays (see below) as well as for polyclonal antibody production.

About one fifth of the semipurified Hsp150Δ-$NGFR_e$ preparation was further purified by IEC and RP-HPLC (see Materials and Methods for details). The fusion protein was eluted in the 18.5 - 27 min fractions as detected by immunoblotting (II, Fig. 4). These fractions were pooled, concentrated and further purified by RP-HPLC. The fusion protein was eluted in the 14-20 min fractions, which were pooled. N-terminal sequencing revealed two signals corresponding to the N-terminal sequences of subunit II, and subunit II missing the three N-terminal amino acids (II, Fig. 4B). The molar ratio between these two species was 3:1.

Folding of the $NGFR_e$ portion was studied by examining its ability to bind NGF. In the competition experiments, the semipurified fusion protein reduced the cross-linking of NGF to authentic NGF receptors of human A875 melanoma cells (II, Fig. 5). This indicates that $NGFR_e$ fused to Hsp150Δ could obtain its ligand-binding conformation in yeast cells, and that the yeast N-glycan did not prevent ligand-binding.

1.3. Glycosylation of $NGFR_e$ (III)

Hsp150Δ is heavily O-glycosylated at most of its potential sites, but has no N-glycans (I). In contrast, authentic $NGFR_e$ contains 41 potential sites for O-glycosylation, none of which are used, and one N-glycosylation site, which is glycosylated. To determinate whether the $NGFR_e$ portion was O-glycosylated in yeast, we made a construct where $NGFR_e$ could be detached from the carrier *in vitro*. The factor X_a-cleavage (fX_a) site was cloned between Hsp150Δ and $NGFR_e$, and the chimeric gene was integrated into the genome of strain H23 to create strain H451. The fusion protein was labeled with [^{35}S] methionine/cysteine chased with CHX, and immunoprecipitated from medium and cell lysate samples with anti-Hsp150 and anti-$NGFR_e$ antisera. According to these pulse-chase experiments, the fusion protein Hsp150Δ-fX_a-$NGFR_e$ was efficiently secreted to the medium, and had an apparent size of over 200 kDa (III, Fig. 2). The fusion protein was the major protein in the medium, and only a small percentage remained unsecreted inside the cells (III, Fig. 2). Furthermore, prevention of N-glycosylation by TM appeared to have only a minor effect on the secretion efficiency of the fusion protein.

The fusion protein was labeled with [^{35}S]methionine/cysteine in the presence of TM and subjected to *in vitro* digestion with fX_a. The digestion produced two fragments: one cleavage product migrated in SDS-PAGE like a 48 kDa and another like a 36 kDa protein (III, Fig. 3). Since Hsp150Δ has no methionines and only one cysteine, and $NGFR_e$ has 24 cysteine residues, the released Hsp150Δ-carrier was not visible in the fluorograph. The 48 kDa protein was most probably $NGFR_e$ since it was similar in size (Welcher *et al.*, 1991). In addition, authentic $NGFR_e$ comigrated with the 48 kDa protein (III, Fig. 4). Since the 36 kDa product contained similar amounts of [^{35}S]-label as Hsp150Δ-$NGFR_e$, the secondary cleavage site was apparently located at the cysteine-free C-terminal portion of $NGFR_e$. Cleavage at the fX_a-site probably exposed a secondary cleavage site for the fX_a preparation, since the fusion protein which lacked a fX_a site was not cleaved at all (not shown).

To analyze O-glycosylation of $NGFR_e$ the fusion protein was labeled with [^{3}H]mannose in the presence of TM. Hsp150Δ-$NGFR_e$ was subjected to factor X_a digestion, precipitated with trichloroacetic acid (TCA) and analyzed by SDS-PAGE. Since no label was detected in the 48 kDa and 36 kDa products, the $NGFR_e$ portion was not detectably O-glycosylated (III, Fig. 3). To study whether the Hsp150Δ-carrier inhibited O-glycosylation of $NGFR_e$, the gene encoding *$NGFR_e$* was cloned into strains H23 and H4 (*sec18-1*) to create strains H487 and H640, respectively. The protein was labeled with [^{35}S] methionine/cysteine, in the presence or absence of TM, immunoprecipitated using anti-Hsp150Δ-$NGFR_e$-antiserum and resolved with SDS-PAGE. In the absence of TM, $NGFR_e$ migrated as a 48 kDa protein and when N-glycosylation was prevented, like a 45 kDa protein in both strains (III, Fig. 4). When the oligosaccharides of $NGFR_e$ were labeled with [^{3}H]mannose in strain H487, a 48 kDa protein could be seen whereas in the presence of TM no incorporated [^{3}H]mannose could be seen, indicating that $NGFR_e$ is not O-glycosylated when expressed alone. As a control, *HSP150Δ-$NGFR_e$* was integrated into the genome of strain H4 (*sec18*) to create strain H538. When protein secretion was blocked in the ER, the Hsp150Δ-$NGFR_e$ fusion protein migrated like a 130 kDa protein and it could be labeled with [^{3}H]mannose (III, Fig. 4). Thus, the Hsp150Δ portion of Hsp150Δ-$NGFR_e$ is heavily O-glycosylated, whereas the $NGFR_e$ portion has one N-glycan but no O-glycans.

Next, the effect of folding on glycosylation was studied. The conformation of Hsp150Δ-$NGFR_e$ was distorted by preventing disulfide bond formation between cysteines using the reducing agent dithiothreitol (DTT). Hsp150Δ-$NGFR_e$ was labeled with [^{35}S] methionine/cysteine in the presence of DTT, followed by a chase in the absence of DTT but with CHX present. Immunoprecipitation from the medium was performed with anti-Hsp150Δ-$NGFR_e$ antiserum, followed by SDS-PAGE. The fusion protein synthesized in the presence of DTT migrated more slowly than authentic molecules, indicating more extensive N-glycosylation (III, Fig. 5, lanes 7 and 5). When N-glycosylation was prevented by TM, the DTT block did not affect the size of the fusion protein (III, Fig. 5, lanes 3 and 1). We suggest that under unusual conditions disulfide bond formation and folding compete with N-glycosylation. Once Asn^{34} emerges in the ER lumen, it is N-glycosylated only if folding and disulfide bond formation in its vicinity have not occurred. This leads to heterogenous occupation of the Asn^{34} site. When disulfide bond formation is prevented, the Asn^{34} of all the molecules are glycosylated.

2. Expression, purification and properties of SFV Nsps

Even though Semliki Forest virus (SFV) has been well studied as a model for alphaviruses, its viral RNA replication and host interactions during infection are not well understood. To overcome the problems of infinitesimal quantities of nonstructural proteins (Nsp) produced during infection and to characterize the properties of individual Nsps, these proteins were expressed individually in different expression systems.

2.1. Expression and purification of Nsps (unpublished)

The approach we took was to express the Semliki Forest virus nonstructural proteins (Nsp1-4) individually in *E. coli* and a baculovirus expression system (BVS) to produce and purify enough of each protein for enzymatic and structural analysis. Optimal growth conditions (*e.g.*, temperature, time) and IPTG-concentration used for induction, were determined experimentally and conditions producing the highest solubility of the protein (*i.e.*, highest protein concentration in S15) during expression were chosen. All the proteins produced were resistant to proteolytic degradation and relatively stable against thermal degradation even though some structural decomposition was observed with Nsp3 and N-Nsp2 upon freezing and thawing (data not shown).

2.1.1. Expression of Nsp1 in *E. coli* (unpublished)

There have been several attempts to purify Nsp1 from *E. coli* (Hyvönen, unpublished; Ahola, unpublished, Vihinen, unpublished; Karhi, unpublished). To facilitate purification the *nsp1* gene was cloned into pHAT, which encodes a tag of six histidines that is used in metal-chelating affinity chromatography (Peränen *et al.*, 1996; Hyvönen, unpublished). The cells were grown to an optical density of 0.7-1.0 at 600 nm, and expression was induced at 15°C with 500 µM IPTG. Initially, when Nsp1 was expressed from this plasmid after IPTG induction, it did not bind to an IMAC-column (Hyvönen, unpublished). However, Nsp1 with its his-tag could be purified under denaturing conditions when the S15 fraction from the cell lysate was treated with 6 M guanidinium-HCl (Fig. 10). Nsp1 was eluted in 167 (Fig. 10, lane 5) or 250 mM imidazole in 0.4 M urea. Purification was also performed in the presence of ß-mercaptoethanol in the washing and elution buffers, which also contained 0.1 M KCl and 2 mM $MgCl_2$. Unfortunately, all attempts to

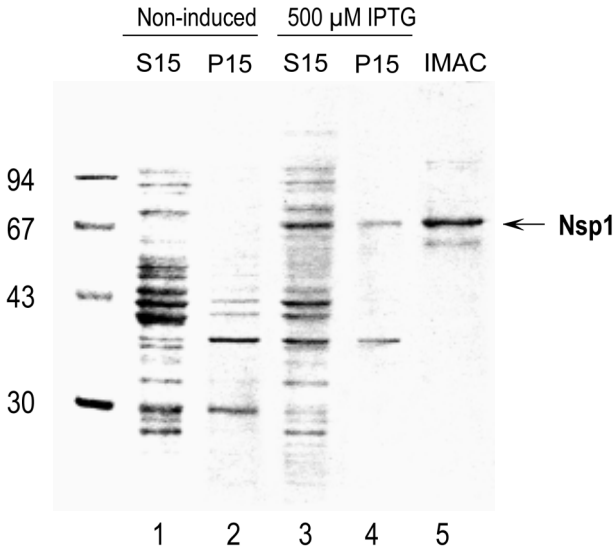

Figure 10. Expression of Nsp1 and its purification utilizing denaturing conditions. His_6-Nsp1 was expressed in *E. coli* by induction with 500 µM IPTG at 15°C. After overnight incubation the cells were harvested and lysed by French Press and lysate was fractionated to S15 and P15 by centrifugation (lanes 3 and 4). Lanes 1 and 2 represent uninduced control cells as indicated above. S15 fraction was denatured with 6 M guanidinium-HCl and subjected to Ni^{2+}-NTA column. The buffer was changed to urea buffer and the urea concentration was decreased by linear gradient to 0.4 mM. Nsp1 was eluted with 167 mM imidazole in 0.4 M urea (lane 5). The proteins were analyzed in 10% SDS-PAGE with Coomassie blue staining.

renature the protein by dialysis against 20 mM Tris, pH 8.0 or PBS failed. Furthermore, several detergents (for example octylglucoside, Triton X-100, thesit and chaps) have also been used to solubilize Nsp1, resulting in inactive enzyme and/or precipitation (Ahola, unpublished; Karhi, unpublished).

2.1.2. Expression of N-terminal fragment of Nsp2 in *E. coli* (unpublished)

Since the N-terminal part of Nsp2 is relatively small (N-Nsp2, aa 1-314, approx. 50 kDa) and has the predicted domains for RNA-helicase- and NTPase-activities, an effort was made to purify it for structural studies. N-Nsp2 was produced in JM109 (DE3) cells as described earlier (Rikkonen *et al.*, 1994), and was initially purified using IMAC-column (Fig. 11A, lane 5). All attempts to concentrate the sample using membrane filtration failed because of aggregation of the protein. However, the elute from IMAC was further purified by gel filtration chromatography (Fig. 11B). In addition to being found in the void volume, N-Nsp2 was also eluted at 22-24 min fractions corresponding to an apparent size of a monomer (40 kDa) in a concentration of approx. 0.1 µg/µl. The concentration of N-Nsp2 after gel filtration indicated that precipitation occurred within the column. Again, any attempts to concentrate the GPC purified product by membrane filtration or dialysis led to irreversible precipitation of N-Nsp2 (data not shown).

2.1.3. Expression of Nsp3 in *E.coli* (unpublished)

To facilitate the purification process the gene coding for Nsp3 was cloned into a pHAT-vector, which includes a gene for a histidine tag (Peränen *et al.*, 1996; Hyvönen, unpublished). The production of Nsp3 was induced by 100 µM IPTG at 15°C, which was determined to be the optimal expression conditions for soluble Nsp3. The S15 fraction of the cell lysate was subjected to IMAC (Ni^{2+}-NTA) and Nsp3 was eluted with 50 mM imidazole (Fig. 12A). The Nsp3 containing fractions were concentrated using membrane filtration (PM30) at 10°C or at RT. The concentrate was further purified by gel filtration.

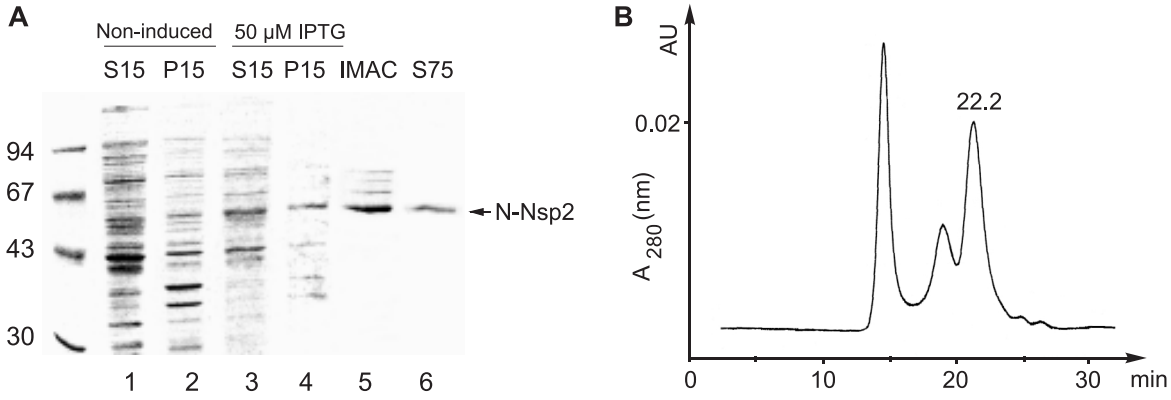

Figure 11. Expression and purification of his-tagged N-Nsp2 produced in *E. coli*. (**A**) The expression of N-Nsp2 was induced by 50 µM IPTG and the cells were harvested and lysed after 6 hours incubation at 24°C. P15 fraction (lane 4) was separated from S15 fraction (lane 3) which was applied to Ni^{2+}-NTA column. Lanes 1 and 2 represent S15 and P15 fractions from uninduced control cells. N-Nsp2 was eluted with 40 mM imidazole (lane 5) and further purified by gelfiltration (lane 6). Proteins were analyzed by 10% SDS-PAGE and Coomassie blue staining. (**B**) Gelfiltration of N-Nsp2 on a Superdex75-column. N-Nsp2 eluted in the void volume and in the peak 22.2 min.

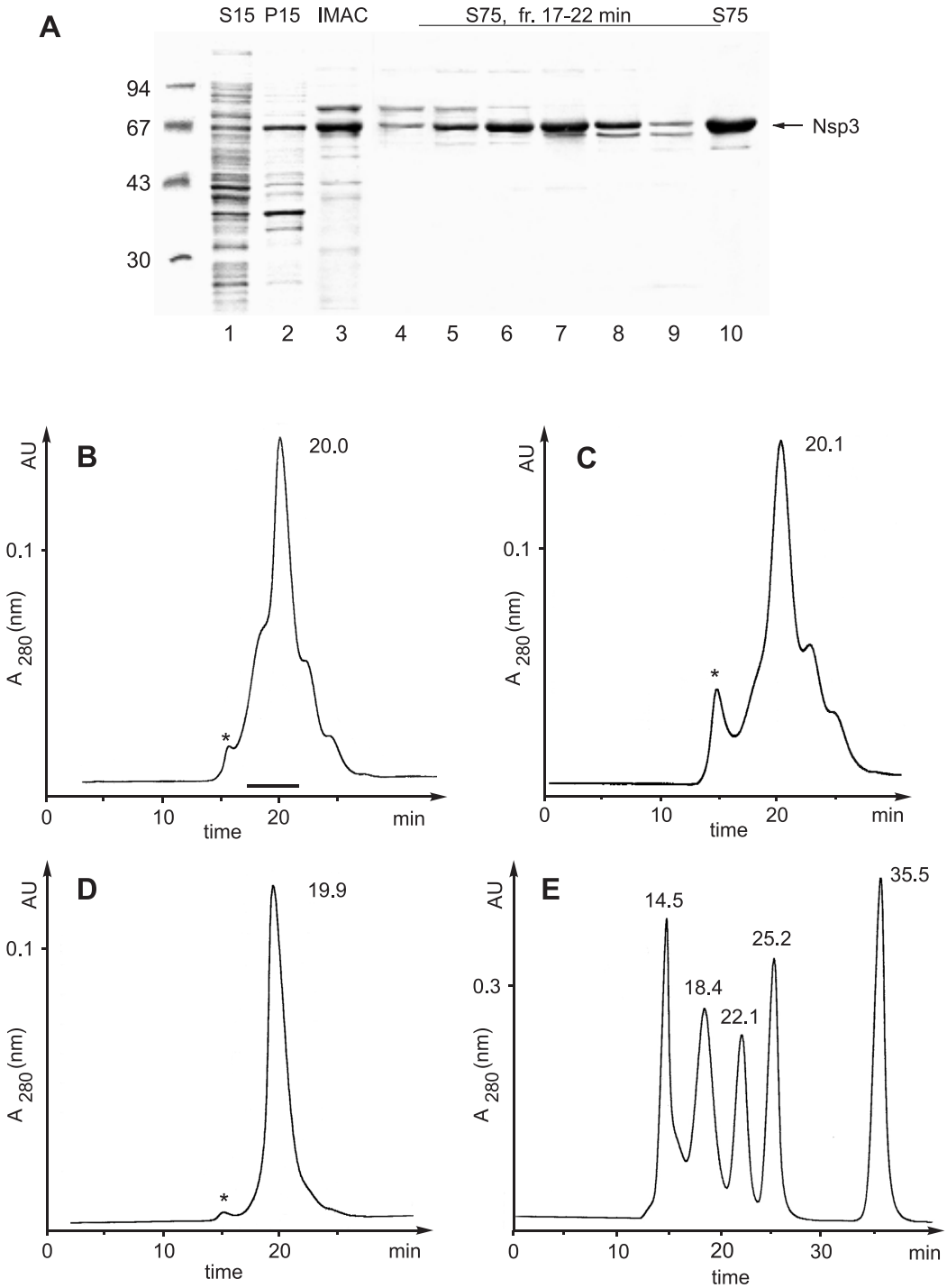

Figure 12. Expression and purification of Nsp3. (**A**) His-tagged Nsp3 was expressed in *E. coli* by induction of 100 µM IPTG at 15°C. After harvesting and lysing the cell, the lysate was fractionated into P15 and S15 (lanes 1 and 2). Nsp3 in S15 fraction was initially purified by Ni^{2+}-NTA-column from which Nsp3 was eluted with 50 mM imidazole (lane 3). Nsp3 was further purified by gel filtration (lanes 4-9). Gel filtration of Nsp3 was performed on a Superdex 75 column (panels **B**-**E**). Nsp3 partially eluted in the void volume (peak marked with *) but mostly at 19-21 min (peak 20.0 in panel **B**). Storage of the preparation overnight at 4°C increased the amount of Nsp3 in the void volume (Panel **C**, peak marked with an asterisk). The peaks at 20 min (fraction in lane 7 in panel **A**) were collected, concentrated and applied again to gel filtration column (**D**). The size of Nsp3 (90 kDa) was estimated according to molecular size standards (**E**).

Initially, a small fraction of the Nsp3 eluted in the void volume, though most of the protein eluted in the 19-21 min fraction (Fig. 12B). When the elute from the IMAC column was stored overnight on ice, at 4°C or frozen and thawed, the amount of Nsp3 in the void volume increased (Fig. 12C). Nevertheless, the peaks from several runs were collected (Fraction in Fig. 12, lane 7), concentrated (precipitated material was removed by centrifugation) and subjected again to the gel filtration column (Fig. 12D). The preparation of Nsp3 was fairly pure containing only minor contaminants and had a protein concentration of 1 µg/µl (Fig. 12A, lane 10). The estimated molecular size of Nsp3 in gel filtration was 90 kDa which indicated that Nsp3 was either in the mono or dimeric form (Fig. 12D and 12E). Nevertheless, storage, and freezing and thawing of this material caused precipitation of the Nsp3.

2.1.4. Expression of Nsp4 in Sf9 cells (unpublished)

To further study the putative polymerase activity of Nsp4, many efforts have been made to purify this protein from *E. coli* for structural analysis, with no success so far (Russo unpublished; Pakula, unpublished; Vihinen, unpublished; Ehsani, unpublished). Here only the results concerning Nsp4 expression in baculovirus system are briefly presented. Sf9 cells were infected with AcMNPV-Nsp4 at 50 PFU/cell and were harvested at 36, 40, 44 and 48 hours p.i. After cell harvesting, lysing and fractionation, nuclear, S15 and P15 fractions were analyzed by SDS-PAGE, followed by Coomassie blue staining (Fig. 13). At 36 hours p.i. the amount of Nsp4 in the cells was undetectable, while at 40 h p.i. Nsp4 was equally distributed in the nuclei and P15 fractions, though only small amounts were seen in the S15 fraction (Fig. 13, lanes 1-3). At 44 and 48 h p.i. most of the protein was found in the nuclear fraction (Fig. 13, lanes 4-6). Next, the effect of inoculated virus dose was studied as the cells were infected with 100, 50, 5, 1 and 0.5 PFU/cell. The analysis of cell fractions after 48 hours p.i. demonstrated that the expression level raised with increased PFU, but as soon as the protein reached a visible level (detected by Coomassie blue staining) it was distributed in the nucleus and P15 fractions, and only small amounts could be observed in the S15 fraction (Fig. 13, lanes 7-9). With higher PFU/cell

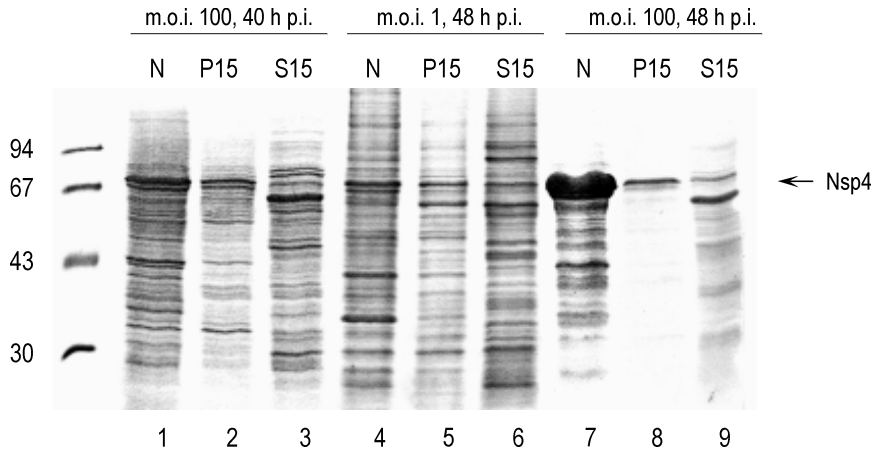

Figure 13. Expression of Nsp4 in the baculovirus system (BSV). Sf9 cells were infected with AcMNPV-Nsp4 with m.o.i. of 100 or 1 PFU/cell. The cells were harvested at 40 hours (lanes 1-3) and at 48 hours (lanes 4-9) p.i., lysed by Dounce homogenizator and fractionated to the nucleus (N), P15 and S15. The proteins were analyzed by 10% SDS-PAGE followed by Coomassie blue staining.

values, most of the protein appeared in the nuclear fraction (data not shown). NaCl as well as some detergents (thesit, octylclucoside, CHAPS) were used to solubilize Nsp4 in P15 fractions. Since no enzymatic activity assay for Nsp4 was available, only mild detergents at low concentrations were tested in order to maintain Nsp4 in its native conformation. None of the treatments solubilized Nsp4, instead it could be pelleted with a 15,000 x g centrifugation (not shown). Attempts were also made to purify Nsp4 from the nuclear fraction (48 hours p.i.) by a sucrose gradient in the presence and absence of 0.5% Triton X-100. Most of the Nsp4 in the absence of detergent floated to the border of 10 and 55% sucrose (not shown). This buoyancy indicates that Nsp4 is associated with membranes. This association could be disrupted by treatment with 0.5% triton X-100, though most of the protein precipitated during this flotation experiment. Also, sonication together with treatment with various detergents was tried to solubilize Nsp4 from the nuclear fraction. However, in all the preparations, Nsp4 remained insoluble according to the general definition of solubility (*i.e.*, remains in the supernatant after centrifugation at 100,000 x g, 1 h).

2.2. Characterization of Nsp1 and Nsp3 (IV-VI)

2.2.1. Critical residues for enzymatic activities of Nsp1 (IV)

To localize amino acid residues essential for the methyltransferase (MT) and guanylyltransferase (GT) activities, several deletions and point mutations were made in Nsp1 (IV, Fig. 1). BL21 cells were transformed with pBATns1 and protein expression was induced by IPTG. Expression levels similar to wt were obtained, except in the case of $Cys^{135}Ala$, $Cys^{142}Ala$ and $Asp^{180}Ala$, which all produced lower amounts compared to wt. The MT and GT activities of Nsp1 and its derivatives were analyzed using S15 fractions (IV, table 1). The mutations where the conserved aa were altered destroyed or greatly reduced both MT and GT activities, whereas mutations at less conserved residues appeared to have no effect (IV, table 1). Two constructs: $Cys^{81/83}Ala$ and $Lys^{169}Ala$ showed moderately reduced MT activity, while $Cys^{81/83}Ala$ also appeared to reduce GT activity. From the deletion constructs (aa 1-429, aa 1-478 and aa 1-515), only the truncated construct aa 1-515 was active in MT and GT analysis, whereas the larger deletions caused loss of both activities. However, mutation $His^{38}Ala$ showed a moderate increase in MT activity (160% compared to the wt Nsp1), though it was totally deficient in forming a covalent 7-methyl-Nsp1 complex (IV, table 1).

To study the effect of palmitoylation on MT and GT activities, some of the Nsp1 derivatives were expressed in HeLa cells using the vaccinia virus vTF7-3 expression system. Expression levels were similar to that of wt, except for mutants $Cys^{135}Ala$ and $Cys^{142}Ala$, which were expressed in lower amounts (data not shown). The cells expressing Nsp1 and its derivatives were labeled with [^{3}H] palmitate, immunoprecipitated with anti-Nsp1 antiserum and analyzed by SDS-PAGE and autoradiography. All the derivatives were palmitoylated to a level similar to that of wt Nsp1 except for the mutant $Leu^{19}Glu$ (IV, Fig. 3). According to quantification by digital imaging, the level of the palmitoylation in this mutant was about 10% that of wt. It appears that palmitoylation does not correlate with MT and GT activities, since this mutant was totally deficient in both MT and GT activities, whereas other MT and GT deficient mutant constructs were palmitoylated like wt.

To study the GT activity independently from MT activity [^{14}C]-labeled 7-methyl GTP was synthesized. Contrary to all expectations Nsp1 did not form a covalent complex with this GTP derivative neither in the presence nor in the absence of AdoMet, but required the presence of AdoHcy for complex forma-

tion (IV, Fig. 4). All the Nsp1 derivatives which had both MT and GT activities, formed complexes with 7-methyl GTP like wt. The derivatives lacking either of the activities (MT or GT) were not able to form the complex. Finally, the different point mutations were analyzed for their ability to bind AdoMet and GTP. The binding properties were analyzed by UV cross-linking in the presence of EDTA which prevents the GT-reaction. The derivatives which were deficient in MT activity were also deficient in binding AdoMet (IV, Fig. 6), though all the derivatives were able to bind GTP (IV, Fig. 5).

2.2.2. Phosphorylation of Nsp3 (V and VI)

In order to study the function of phosphorylation of Nsp3, the phosphorylation sites were determined by electrospray and matrix-assisted laser desorption/ionization mass spectrometry, point mutational analysis, two-dimensional (2D) phosphopeptide mapping, and Edman sequencing. Nsp3 was produced in HeLa cells by transfection, which supplied sufficient amounts of the protein for this study. According to phosphoamino acid analysis, Nsp3 is phosphorylated on serine and threonine residues in a ratio of approximately 2:1 in infected BHK cells (Peränen *et al.*, 1988). The relation of phosphorylated serines to threonines turned out to be different for Nsp3 produced in transfected HeLa cells compared to SFV infected BHK cells, *i.e.*, the relation between phosphoserines and phosphothreonines was approx. 6:1 (data not shown). However, according to phosphopeptide mapping the phosphorylation sites in Nsp3 were the same regardless of the expression type or cell line (V, Fig. 4).

2.2.2.1. Determination of phosphorylation sites (V)

The threonines at aa positions 344 and 345 were identified as phosphate acceptor sites by point mutational analysis, when all 11 threonines in the C-terminal region of Nsp3 were mutated to alanines (VI, Fig. 2). These two sites were further confirmed by mass spectrometric (MS) analysis. For MS, Nsp3 was purified and cleaved into peptides as described in the Materials and Methods section (Fig. 8). A precursor ion scan of m/z 79 in the negative mode was done with trypsin digested Nsp3, and revealed two major peaks which could be tentatively identified as peptides Val^{308}-Arg^{332} and Val^{308}-Lys^{323}, both carrying one phosphate (V, Fig. 1B). Further analysis of these major peaks by tandem MS identified the peptides as Val^{308}-Arg^{322} and Val^{308}-Lys^{323} with one phosphate moiety, and the assignment of phosphorylation at Ser^{320}. The phosphorylation of Ser^{320} was also observed by point mutational analysis followed by phosphopeptide mapping and Edman sequencing (V, Fig. 4 and 5A).

Fe^{3+}-IMAC was used to enrich the phosphopeptides of Nsp3 digested with trypsin and CNBr before MALDI mass analysis. Masses revealed from MS were assigned to peptides Lys^{324}-Arg^{337} and Tyr^{324}-Arg^{337} carrying from 1 up to 3 phosphates (V, Fig. 3A). The signals corresponding to the phosphopeptides disappeared after calf intestinal alkaline phosphatase (CIAP) treatment providing further evidence for three phosphates on this peptide (V, Fig. 3B). These peptides were also identified by phosphopeptide mapping followed by Edman degradation, which implicated the phosphorylation of serines 327 and 335 (V, Fig. 5B, 5D and 5E). Furthermore, according to phosphoamino acid analysis, these peptides included only phosphoserines supporting the result that threonines 328 and 329 are not phosphorylated (data not shown).

Peptide Ser^{356}-Met^{372} was putatively identified to be phosphorylated by MALDI-TOF MS in the negative ion mode, which released a series of peptide masses having a mass difference of 80 Da (V, Fig. 3C). This peptide

appeared to carry from 0 to 4 phosphates from which up to three could be removed by phosphatase treatment (V, Fig. 3C and 3D). The final identification of this peptide was achieved by tandem MS identifying the peptide being modified with an acrylamide adduct (V, Fig. 2C). This peptide has up to four possible phosphorylation sites, *i.e*, four serines, which can, thus all be potentially phosphorylated.

By ESI MS, 'bump-like' signals were revealed and tentatively identified as a charge series of peptide Gly^{338}-Lys^{415}. This peptide appeared to contain 7 to 12 phosphates distributed over its 13 potential phosphorylation sites. The same peptide was also most likely seen in phosphopeptide mapping as a large diffuse spot (V, Fig. 4B), but unfortunately, it turned out to be impossible to sequence it (V, Fig. 5C). This relatively long peptide (78 aa) has 11 aspartatic acid and 4 glutamic acid residues, which can covalently attach to membranes and result in an incorrect signal upon Edman sequencing. Furthermore, this peptide was the only one analyzed by phosphopeptide mapping to contain both phosphoserine and phosphothreonine (data not shown).

2.2.2.2. Phosphorylation of Nsp3 derivatives (VI)

To study the phosphorylation of individual amino acids, various point mutations in Nsp3 and a C-terminal deletion, N-Nsp3 (aa 1-328) were made. The mutated derivatives were expressed in HeLa cells by transfection, and the Nsp3 proteins were labeled with [^{32}P] orthophosphate. The total phosphorylation level of threonine double mutant (Thr344/345Ala) was approximately 40% less than that of wt Nsp3 as quantified with a phosphoimager (VI, Fig. 4). When either of the threonines Thr^{344} or Thr^{345} were individually mutated, a similar decrease in phosphorylation was observed (VI, Fig. 4). The mutation Ser^{320}Ala, which in mass spectrometric analysis seemed to be a major serine phosphorylation site, also decreased the total phosphorylation of Nsp3 by 40-50% (VI, Fig. 4). However, when mutations $Thr^{344,345}$Ala and Ser^{320}Ala were combined together, no further decrease in phosphorylation was observed (VI, Fig. 4, lane 5). The phosphorylation level of N-Nsp3 was about 1% as compared to wt (VI, Fig. 4). Mutation of the known phosphorylation sites located within N-Nsp3, serine 320 to alanine did not greatly affect its phosphorylation level (VI, Fig. 4). Furthermore, a small internal deletion of 26 amino acids, Nsp3Δ26 (Nsp3Δ343-368), reduced Nsp3 phosphorylation by 90% (VI, Fig. 4). These results suggest that the different phosphorylation sites may influence each other in a rather complex manner, and therefore phosphorylation of individual sites can not be simply quantified based on point mutational studies.

To study whether the full-length Nsp3 was needed either to provide appropriate conformation and/or kinase activity to phosphorylate the serines/threonines in the region 320-368, a N-terminal deletion construct (C-Nsp3) was made (VI, Fig. 1).This peptide could be labeled with [^{32}P]orthophosphate when expressed in HeLa cells by transfection (VI, Fig. 3). However, the phosphorylation level of the C-terminal peptide was about 30% of the full-length protein quantified by double labeling with [^{35}S]methionine/cysteine and [^{32}P]orthophosphate. This difference may be due to an altered conformation of the peptide or to the subcellular localization of C-Nsp3, which differed from the wt Nsp3 (VI, Fig. 5). The possibility of Nsp3 autophosphorylation was studied by co-transfecting wt Nsp3 with the C-terminal peptide, but no increase in the phosphorylation level of C-Nsp3 was detected (VI, Fig. 3). Since it is unlikely that the non-conserved C-terminal region could posess any kinase activity, these results suggest that the phosphorylation of C-Nsp3 and wt Nsp3 is catalyzed by cellular kinases.

2.2.2.3. Elimination of phosphorylation sites (V and VI)

Based on phosphopeptide mapping followed by Edman degradation as well as mass spectrometric and point mutational analysis of Nsp3, the phosphorylation sites of Nsp3 were determined as serines 320, 327, 332, 335 and 12 Ser/Thr in the peptide Gly^{338}-Lys^{415}. Two non-phosphorylated mutants were generated. The first had a deletion of 50 aa (Nsp3Δ319-368) that removed the whole potential region of phosphorylation (Nsp3Δ50). The second construct had a 26 aa deletion (Δ343-368) together with serines 320, 327, 332 and 335 being mutated to alanines, and designed as Nsp3Δ26+4S→4A. This later construct maintained highly conserved residues such as Tyr^{324}. Although these constructs were expressed in HeLa cells at amounts similar to wt Nsp3, no [^{32}P]-incorporation could be detected with these Nsp3 derivatives (V, Fig. 6 and unpublished). Assuming that the deletions did not affect the overall confirmation of Nsp3, it appears that Thr^{378} was not modified by phosphorylation. Both mutants were also cloned into the genome of SFV and the deficiency of phosphorylation was confirmed by double labeling (VI, Fig. 6 and unpublished).

2.2.2.4. Effect of phosphorylation on membrane association (VI)

To study the effects of phosphorylation of Nsp3 on its membrane association, a C-terminal deletion mutant (aa 1-328) was constructed (VI, Fig. 1). Two putative phosphorylation sites (Ser^{320} and Ser^{327}) remained in this truncated Nsp3, and it was phosphorylated at one percent of wt (VI, Fig. 3). However, according to flotation assays this derivative has an affinity to membranes similar to wt Nsp3, and like Nsp3 this attachment could be dissociated by 1 M NaCl (VI, Fig. 5). This result was further confirmed by indirect immunofluorescence studies when non-phosphorylated derivatives, expressed by transfection or in SFV infection, were found in similar vacuole like structures as wt Nsp3 (VI, Fig. 5 and unpublished).

2.2.2.5. Phosphorylation deficient Nsp3 in the context of SFV (VI)

To study the effect of phosphorylation on virus replication, Nsp3 phosphorylation deficient mutants were placed in the SFV genome (SFV Nsp3Δ50 and SFV Nsp3Δ26+4S→4A), and mutant viruses were grown in BHK, Vero and NIH cells. All the mutants were able to propagate in these cell lines and showed titers close to those of the wt SFV (VI, Fig. 7A and unpublished). This indicated that Nsp3 phosphorylation was not required for SFV replication in these cell lines.

The effects of these Nsp3 mutants on pathogenicity in mice were also investigated. At three weeks post-inoculation, 20% and 66% of the mice infected with doses of 10^6 PFU/mouse SFV Nsp3Δ26+4S→4A and wt SFV, respectively, were killed (unpublished). With higher doses (10^7 PFU/mouse) no mice survived the infection by wt SFV, whereas 66% of the mice infected with SFV Nsp3Δ26+4S→4A were still alive at three weeks. Viral doses up to 10^7 PFU/mouse of SFV Nsp3Δ50 were avirulent for adult mice (VI).

DISCUSSION

1. Hsp150Δ as a carrier for protein production

The ability of Hsp150Δ to confer secretion competence to heterologous proteins has been first studied with ß-lactamase, which is a periplasmic protein of *E. coli* (Simonen *et al.*, 1994). ß-lactamase was chosen as the reporter protein, since it is not secreted in yeast with its own signal peptide and its conformation is easy to monitor by an enzymatic assay (Roggenkamp *et al.*, 1985; Carwright *et al.*, 1994). When ß-lactamase is fused to C-terminus of Hsp150, the protein is secreted to the medium but is enzymatically inactive. Fusion of ß-lactamase to subunit I results in a fusion protein with poor activity and secretion competence (Simonen *et al.*, 1994; Holkeri and Makarow, 1998). The fusion protein that is enzymatically active and secreted efficiently to the growth medium, is the construct in which ß-lactamase is fused to the repetitive region of Hsp150 (Simonen *et al.*, 1994). Moreover, the first four repeats of subunit II have been shown to be sufficient to confer secretion competence although slower than the carrier with 11 repeats (Suntio *et al.*, 1999).

The ability of Hsp150Δ to confer secretion competence was further evaluated using the extracellular domain of a mammalian transmembrane protein, rat nerve growth factor receptor ($NGFR_e$) (II). Since $NGFR_e$ (223 aa) has 12 disulfide bridges, it was thought to offer a challenge for the folding machinery of the yeast ER. When $NGFR_e$ was fused to the signal peptide of yeast invertase, it was translocated into the ER, but unable to exit the ER. $NGFR_e$ was degraded in the ER lumen by a currently unknown mechanism (Holkeri *et al.*, 1998). However, when $NGFR_e$ was fused to Hsp150Δ, the fusion protein was secreted to the growth medium and acquired a conformation capable of binding its natural ligand, NGF. Hsp150Δ-$NGFR_e$ could also inhibit the cross-linking of NGF to the authentic NGFR. However, only approximately half of the binding could be inhibited by the fusion protein, which could possibly be explained by its linkage to the Hsp150 carrier. Hsp150Δ-$NGFR_e$ was secreted to the medium at a concentration of approximately 10 milligrams per liter (II). A similar level of expression (10-15 mg/l) was obtained when tick anticoagulant peptide was produced in *S. cerevisiae* utilizing the pre-pro-region of α-factor (Cook *et al.*, 1998).

Subunit II of Hsp150 was found to be O-glycosylated with di-, tri-, tetra- and pentamannosides in a ratio of 4:1:1:1 (I). Based on mass spectrometric analysis 65 of the potential 85 Ser/Thr-residues in subunit II were glycosylated, as well as most, if not all of the potential 21 Ser/Thr residues in subunit I (Suntio *et al.*, 1999). When fused to Hsp150Δ, $NGFR_e$ was N-glycosylated at its one potential site, but not glycosylated at any of its 41 potential O-glycosylation sites, whereas the carrier portion was heavily O-glycosylated (III). Moreover, the distortion of protein folding caused by preventing disulfide bridge formation did not affect the O-glycosylation of the fusion protein but enhanced N-glycosylation of $NGFR_e$ portion. Thus, it appears that the O-glycosylation machinery of yeast is highly selective.

Yeast BiP/Kar2p, a member of Hsp70 chaperones, is thought to assist the translocation of *de novo* synthesized proteins across the ER membrane as well as conformational maturation after translocation (Gething and Sambrook, 1992; Simons *et al.*, 1995; Jämsä *et al.*, 1994). It has been shown that, Hsp150Δ-ß-lactamase requires BiP/Kar2p for both translocation and maturation, even though it has only one disulfide bridge (Holkeri *et al.*, 1998). However, the folding

of $NGFR_e$ is assumed to occur domain by domain during the on going translocation (Netzer and Hartl, 1997). Like TGFR, the structure is more likely to be determined by hydrogen bond formation rather than disulfides (Banner *et al.*, 1993). Indeed, when the conformation of $NGFR_e$ is distorted by DTT, only mild conformation changes occur and the proteins are rapidly secreted after removal of DTT. Thus, even if BiP/Kar2p is required for the translocation of Hsp150Δ-$NGFR_e$, the conformational maturation occurs in the absence of the chaperone (Holkeri *et al.*, 1998). The numerous disulfide bridges of $NGFR_e$ seem to have only a minor role in the folding process making this molecule relatively easy to be secreted.

Hsp150, truncated Hsp150, and Hsp150Δ-$NGFR_e$ were purified from the growth medium by membrane filtration, ion exchange, and gel filtration chromatography, and in some cases using reserved phase chromatography. The purity was verified by N-terminal sequencing. All of the proteins were purified to homogeneity since no other signal was detectable (II). Probably due to the high level of O-glycosylation, these proteins were hydrophilic which made the purification process relatively easy. CD and NMR spectroscopy revealed that the repetitive region of the carrier protein lacked secondary structure, and ß-sheet occurred in the C-terminal part of subunit II and/or in subunit I. It is possible, that the lack of a secondary structure in this carrier protein promotes the fusion partner to acquire proper conformation *e.g.*, by attracting chaperones.

2. Expression and characterization of SFV Nsps

2.1. Expression and purification of Nsps

One of the main reasons for expressing and purifying the Nsps was to obtain enough protein for crystallization studies. For this purpose two essential requirements are that the protein is extremely pure, and it is dissolved in an aqueous medium at relatively high concentrations (5-20 mg/ml) (Durbin and Feher *et al.*, 1996). Even crystals can occasionally be grown from impure samples, the size and quality of the crystals improve as the purity of the protein increases (Ollis and White, 1990). Proteins with hydrophobic regions such as membrane proteins are problematic, since they easily form aggregates and oligomers. In SFV infected cells, all the Ns-proteins are associated with cellular membranes (Ahola *et al.*, 1999). For example, Nsp1 is attached to membranes by at least two different mechanisms. Firstly, it is palmitoylated at cysteine residues 418-420, which probably aids in membrane association (Laakkonen *et al.*, 1996). Secondly, it has a short segment (aa 245-264) which appear to be involved with membranes consisting of anionic phospholipids such as phosphatidylserines (Ahola *et al.*, 1999). When Nsp1 was expressed with the baculovirus system, it was mainly concentrated in the P15 fraction (Laakkonen *et al.*, 1994). After solubilization of the membranes by DOC, the protein remained enzymatically active, but the molecular size of Nsp1 aggregates were still over 300 kDa. The methyltransferase (MT) activity of Nsp1 produced in *E. coli* was mainly recovered in the S15 fraction, but the protein was precipitated in a sucrose gradient centrifuged at 100,000 x g (Laakkonen *et al.*, 1994). It has been suggested that both enzymatic activities (methyltransferase and guanylyltransferase) of Nsp1 are dependent on the membrane attachment (Ahola *et al.*, 1999).

When a protein is to be refolded from aggregates *i.e.*, inclusion bodies, it must be solubilized prior to the refolding into its native conformation can occur. The refolding step is complex and depends greatly on the renaturing conditions, such as, pH, rates of dialysis, protein concentration and redox conditions (Valax and Georgiou, 1993). In addition, the concentration of the protein should be fairly low (10-100 µg/ml) for proper folding (Maachupalli-Reddy *et al.*, 1997). Thus,

usually after the refolding step the protein must be concentrated. In the case of Nsp1, preliminary attempts to renature it after purification under denaturing conditions were not successful, and no enzymatically active protein could be obtained. To further investigate the potential of this approach, the following parameters should be varied: protein concentration, temperature, pH and dialysis buffers, so that the optimal conditions for Nsp1 renaturation could be found. In a related study, it has been reported that applying high hydrostatic pressure (1-2 kbar) to suspensions containing aggregates of recombinant human growth hormone, lysozyme or ß-lactamase resulted in properly folded proteins (St. John *et al.*, 1999).

In addition to Nsp1, fractions of Nsp2, Nsp3 and Nsp4 are tightly bound to the cytoplasmic membranes of SFV infected cells (Ahola *et al.*, 1999). The membrane association of Nsp3, when expressed alone by transfection, did not appear to be very tight since it could be released from the membranes with salt treatment (VI). A fraction of all Nsps can be found in vesicular structures in infected and transfected cells (Kujala *et al.*, 2001). All of these proteins have hydrophobic regions in their amino acid sequences. For example, Nsp3 has a potential segment for membrane attachment between aa 99-116, which in contact with membranes could form a amphiphilic helix (Fig. 14). This region is highly conserved among alphaviruses, and it can be found also in rubella and hepatitis E viruses. These hydrophobic segments may partly explain the characteristic tendency of Nsps to from aggregates. The removal of the hydrophobic C-terminal region (55 aa) from hepatitis C NS5B protein allowed its purification from *E. coli*, and structural characterization of its active RNA-dependent RNA polymerase unit by crystallization studies (Bressanelli *et al.*, 1999). The 55 aa deletion did not appear to affect its polymerase activity, and a further deletion of 21 aa in a hydrophobic region seemed to increase its RNA polymerase activity (Bressanelli *et al.*, 1999; Tomei *et al.*, 2000). Structural studies of the catalytic subunit of Hepatitis-C, poliovirus and two retrovirus replicase proteins suggest similar structures of the catalytic unit that appear to be shared in all of these viruses (Hansen *et al.*, 1997; Jäger and Pata, 1999; Bressanelli *et al.*, 1999). It remains to be seen whether the structural features of the catalytic subunit of Hepatitis-C, poliovirus and retrovirus are also conserved within alphaviruses.

The insolubility problem is difficult to overcome and varies from case to case. For example, some proteins such as immunoglobulins need high ionic strength to remain soluble (Harris *et al.*, 1995). In case of Nsps, salt concentration had only small effects on solubility. Also pH is crucial for the purification process, and the more the pH of crystallization differs from the isoelectric pH, the higher is the protein net charge and its solubility. However, calculated isoelectric pH values are only theoretical since some posttranslational modifications can greatly affect the final net charge. Detergents can play a crucial role in the crystallization of membrane proteins

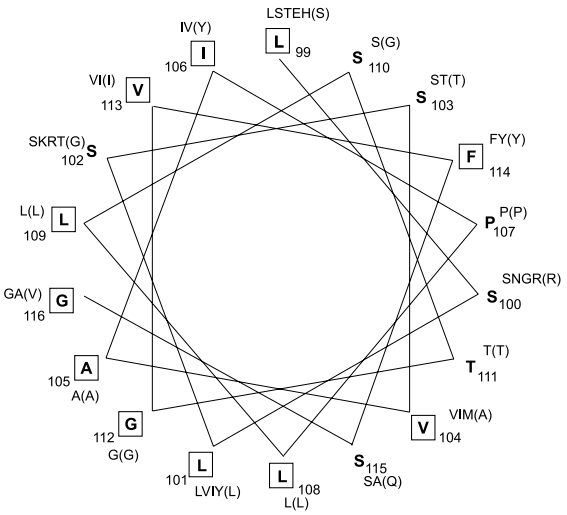

Figure 14. An amphiphilic helical wheel of Nsp3 aa 99-116. The amino acid residues of other alphavirus Nsp3s, as well as hepatitis E in brackets, are marked. □ = hydrophobic residue.

(Marston and Hartley, 1990; Thomas and McNamee, 1990; Hjelmeland, 1990). The detergents used to crystallize membrane proteins are generally simple nonionic or zwitterionic compounds. At detergent concentrations below critical micelle concentrations (CMC) most membrane proteins are not very soluble and may form two dimensional crystals (Thomas and McNamee, 1990). 3D crystals are often found close to the condition where phase separation occurs. The detergent to protein ratio may be a critical parameter for successful crystallization of membrane proteins. Many proteins appear to have well defined detergent-to-protein ratios for optimal solubilization. For most detergents with high CMC (>1%), protein solubilization occurs at or near the CMC. Even though some detergents were utilized to solubilize Nsp1, no conditions were found suitable for purification (Ahola, personal communication). Furthermore, disruption of essential protein-lipid or protein-protein interactions by the detergent might lead to inactivation or denaturation. Purification of proteins with unknown functions limit the choices of detergents and concentrations in which they can be used.

2.2. GT AND MT ACTIVITIES OF NSP1

Nsp1 has been found to catalyze two reactions in the capping of viral mRNA, *i.e.*, firstly, the transfer of a methyl group from AdoMet and secondly, the formation of a covalent complex, m^7GMP-Nsp1 (IV; Ahola and Kääriäinen, 1995). The structural region and amino acid residues involved in these activities were elucidated with point mutational analysis (IV). These studies revealed that a mutation at His^{38} destroyed GT activity, which made this residue a putative binding site for m^7GMP. SFV His^{38} is conserved among the alphavirus superfamily, and a mutation of a respective residue in SIN Nsp1 resulted in a noninfectious viral RNA (Rozanov *et al.*, 1992; Niesters and Strauss, 1990). Structural studies of five AdoMet-dependent MTs suggested similarities in AdoMet-binding sites, which are consistent with an antiparallel ß-sheet flanked by alpha-helixes (Schluckebier *et al.*, 1995). The secondary structure predictions of Nsp1 showed that it shares similarities with the structures of cellular MTs especially in its N-terminal domain (IV). Also the G-loop (UU-Asp-UoooXo; U=large hydrophobic residue, o=Gly/Ala/Ser, X=any residue) which is characteristic for MTs can also be found in the Nsp1 sequence (Koonin *et al.*, 1995; Ingrosso *et al.*, 1989). Mutational analysis indicated that residues Asp^{64} and Asp^{90} are required for AdoMet binding activity. Furthermore, Asp^{64} is conserved within the alphavirus-like superfamily and Asp^{90} may represent the conserved acidic residue involved in binding hydroxyl groups in AdoMet (Djordjevic and Stock, 1997). Cys^{142} was also required for the binding of AdoMet, whereas the residues Arg^{93}, Cys^{135} and Tyr^{249} were essential for binding of AdoMet and required for MT activity (IV). The conserved residues His^{39}, Asp^{91}, Arg^{94} and Tyr^{249} were also mutated in SIN Nsp1 and found to be critical for MT activity as well as for virus replication (Wang *et al.*, 1995).

2.3. PHOSPHORYLATION OF NSP3

Using different analytical methods, *i.e.*, mass spectrometry (MALDI TOF MS and ESI MS) point mutational analysis, phosphopeptide mapping and Edman sequencing, it was possible to identify major phosphorylation sites in Nsp3. The phosphorylation region was located within 48 aa, Ser^{320}-Ser^{368}, where 12 serines and 4 threonines could be phosphorylated. No direct data was obtained regarding the phosphorylation state of Thr^{378}, but assuming that a 26 aa internal deletion together with 4 point mutations in Nsp3Δ26+4S→4A did not affect the accessibility of Thr^{378} to kinase(s) by causing conformational changes, it appears that in Nsp3, Thr^{378} is not

modified by phosphorylation (V). The phosphorylation of Nsp3 was heterogeneous as suggested by the finding that some of the phosphorylation sites were also detected in unphosphorylated state by MS. Within this area, mutation of one or more serines/threonines affected the phosphorylation of the other residues, *i.e.*, the phosphorylation of some residues was dependent on the phosphorylation of other residues. Furthermore, the ratio of phosphorylated serines and threonines in Nsp3 varied according to the method and cell line, being 6:1 (V) and 2:1 (Peränen *et al.*, 1988), in transfected HeLa cells and SFV infected BHK cells, respectively. However, phosphopeptide mapping revealed similar patterns for Nsp3 from both of these expression types, which indicates that in spite of the ratio difference, the sites of phosphorylation were the same. Difference in the ratio of phosphorylated serines to threonines could be due to the different cell types used, differences in the cellular kinases involved, and/or the interacting of host factors with the viral proteins.

Even though Ser^{320} appears to be a major phosphorylation site in transfected HeLa cells and is in the conserved region of SFV Nsp3, it is not conserved among all alphaviruses. In general, the common features of the C-terminal domains of alphavirus Nps3s are negatively charged regions which are serine and threonine rich with short proline rich domains. The C-terminal region of SIN Nsp3 is longer (34 aa) and has 19 Ser/Thr residues more compared to SFV Nsp3. No systematic phosphorylation site analysis of SIN Nsp3 have been made, but a C-terminal deletion has been observed to dramatically reduce phosphorylation (LaStarza *et al.*, 1994a). Interestingly, the C-terminal region of SIN Nsp3 has two separate domains rich in serines and threonines, both of which could be phosphorylated resulting in different forms of SIN Nsp3.

2.3.1. Determination of phosphorylation sites

Nsp3 was a challenging molecule for phosphorylation analysis since it has 39 serine and 35 threonine residues, all potential phosphorylation sites. Also, the resistance of Nsp3 to trypsin digestion together with one large tryptic peptide (78 aa) make it difficult to analyze. This large peptide was identified as a phosphopeptide by precursor ion scanning, and it contained from 7 to 12 phosphates over its 13 potential sites. Unfortunately, no exact phosphorylation sites could be determined due to its large size and extensive phosphorylation. In addition, its large size and high number of acidic residues (15 Glu/Asp) prevented this peptide to be analyzed by Edman sequencing as well.

MS is a very sensitive method but extremely pure peptide samples are required for analysis. In addition, the phosphopeptide must have a molecular weight of more than 600 Da in order not to be masked by signals coming from the matrix material itself. Large peptides or peptides containing multiple phosphates are not likely to produce informative fragmentation patterns, which would allow for site specific assignment of phosphorylation over the sequence. Data from the MALDI/CIAP experiments, point mutational analyses or Edman sequencing alone would not have been sufficient for unequivocal identification of the phosphorylated residues of Nsp3 (V). Several independent methods were needed to localize phosphorylated residues, since any of these methods used alone would have resulted in only partial information about the phosphorylation sites.

2.3.2. Kinases

The C-terminal region of Nsp3 (aa 312-482) was phosphorylated when expressed alone, although at reduced levels as compared to the wt Nsp3 (VI). This reduced phospho-

rylation may reflect conformational changes in the polypeptide that lead to differential kinase recognition. Co-expression with the wt Nsp3 did not increase the phosphorylation of the C-terminal peptide, indicating that Nsp3 is unlikely be a kinase by itself. An alternative explanation for the diminished phosphorylation of C-Nsp3 is offered by the fact that it is localized in the cell cytoplasm, whereas the wt Nsp3 was associated with vesicular structures. Thus, C-Nsp3 might not be equally accessible to membrane-associated kinases, which may play a role in phosphorylating the wt Nsp3. The phosphorylation of Nsp3 was, however, not required for its peripheral membrane association. In addition, Nsp3 lacks any catalytic domains present in almost all eukaryotic kinases determined by sequence comparisons. These results indicate that the phosphorylation of Nsp3 is most likely to be mediated by kinase(s) of cellular origin.

SIN Nsp3 has been suggested to be phosphorylated by casein kinase II (CK II) (Li *et al.*, 1990). Most of the phosphorylation sites identified in this study are putative phosphorylation sites for CKII evincing that CKII could be a kinase involved in phosphorylation of SFV Nsp3. Serines 327 and 367 have a preferred recognition sequence, [Ser/Thr]-x-x-[Asp/Glu], for CKII and most of the other phosphorylation sites in Nsp3 are also putative sites for CKII assuming that the adjacent Ser/Thr-sites have been already phosphorylated by additional kinases (Kennelly and Krebs, 1991). However, Ser^{320} and Ser^{335} do not have a consensus sequence for CKII, but instead have a recognition sequence for protein kinase C, [Ser/Thr]-x-[Arg/Lys], as does Ser^{332} (Kennelly and Krebs, 1991). SFV Nsp3 Ser^{320} is also in a consensus site for $p34^{cdc2}$-kinase, which in turn is phosphorylated by CKII (Russo *et al.*, 1992b). NS5A, the phosphoprotein of hepatitis C and bovine viral diarrhea virus as well as NS5 of yellow fever virus (all members of family *Flaviviridae*) are suggested to be phosphorylated by a kinase or kinases from the CMGC group of serine/threonine kinases which includes *e.g.*, CKII and proline-directed kinases (Reed *et al.*, 1997; Reed *et al.*, 1998). Among negative strand RNA viruses, the phosphoprotein P of measles, rabies, vesicular stomatis and Sendai virus have been reported to be phosphorylated by two different kinases (Gupta *et al.*, 2000; Barik and Banerjee, 1992; Huntley *et al.*, 1997; Byrappa *et al.*, 1996).

2.3.3. Effects on the virus cycle and neurovirulence

The identified phosphorylation sites of SFV Nsp3 were eliminated by a deletion or by a deletion combined with point mutations resulting in Nsp3 phosphorylation defective mutant viruses. Surprisingly, these viruses showed only minor differences in replication compared to the wt SFV, in BHK, Vero and NIH cells. Thus, phosphorylation of SFV Nsp3 is not essential for virus replication in the cell lines tested. Similarly, a large deletion in the C-terminal region of SIN Nsp3 did not affect its plaque forming ability in mammalian cells (LaStarza *et al.*, 1994a). However, this derivative showed a poor plaque formation in mosquito cells. Thus, even though phosphorylation of SFV Nsp3 is not essential in BHK, Vero nor NIH cells, it could play an important role in replication in other cell lines such as insect cells. The ultimate function of phosphorylation remains to be resolved, as well as the function of Nsp3 in the SFV infection cycle. Among negative-strand RNA viruses, phosphoproteins of paramyxoviruses and rhabdoviruses act as chaperons for nucleoprotein, preventing its polymerization and nonspecific binding to cellular RNA (Tarbouriech *et al.*, 2000; Curran *et al.*, 1995). The phosphorylation state of P proteins have been reported to be essential for efficient transcription of viral polymerase among negative-strand RNA viruses (Mazumder and Barik, 1994; Barik and Banerjee, 1992).

The pathogenesis of Nsp3 phosphorylation defective viruses were studied in mice. Groups of five mice were infected intraperitoneally with 10^6 and 10^7 PFU/mouse of the wt and mutant viruses. The phosphorylation defective mutant SFV Nsp3Δ26+4S→4A showed a significant reduction in pathogenicity whereas SFV Nsp3Δ50 in the same experiment turned out to be apathogenic. Since both of these constructs were Nsp3 phosphorylation defective the apathogenic feature of latter mutant virus may be a result of the deletion of some conserved residues in region 319-342. Furthermore, the route of infection (*i.e.*, intraperitoneal or intranasal route) may also contribute to the resulting pathogenicity (Atkins *et al.*, 1999).

SFV strain A7(74) has been derived from an apathogenic SFV strain A7 by further selection for avirulence (Bradish *et al.*, 1971). In the neurons of adult mice A7(74) is unable to form virions and all the infected mice remained asymptomatic (Tuittila *et al.*, 2000). There are multiple mutations in A7(74) compared to the wt SFV4 strain, which in the region of Nsp3 include 9 single aa mutations, 7 aa deletion and an opal termination codon between Nsp3 and Nsp4. The 7 aa deletion (Gly-Ile-Ala-Asp-Leu-Ala-Ala) did not confer the attenuation of SFV4 virulence whereas opal and an other aa mutation in the region of Nsp3 reconstituted a virulent virus (Tuittila *et al.*, 2000). Nevertheless, none of the mutated aa were in the deleted region of Nsp3Δ50 suggesting that in addition to the opal codon, also the non-conserved C-terminal region, including the phosphorylation sites, is significant in the regulation of virus pathogenesis in mice. Thus, neither of these virus strains are capable of the rapid spread and replication in the central nervous system that is required for neurovirulence.

3. Conclusions and future aspects

The power of utilizing molecular engineering to produce recombinant proteins for various purposes has revolutionized many areas in biotechnology. In this study, modern molecular techniques were applied to produce and purify fungal and viral recombinant proteins in a variety of production systems.

Structural studies of Hsp150Δ revealed the first conformation information of a yeast carrier peptide. The extended structure of this carrier may allow the attached foreign proteins to be freely folded into their native conformations. We showed that the Hsp150Δ carrier is able to secrete $NGFR_e$ in a ligand binding form, but further work must be done to improve the recombinant protein production and secretion efficiencies. The possible introduction of the Hsp150Δ carrier into other yeast strains (*e.g.*, *Pichia pastoris*) might be useful in the attempts to increase of the production level.

All efforts to date to purify individually expressed Nsps amounts sufficient enough for structural studies have been unsuccessful. The conditions applied in this study for expression of Nsps in *E. coli* or by the baculovirus system, differ dramatically from a natural SFV infection. During SFV infection Nsps are produced in very low amounts, most of them are membrane associated, and they are folded from a large, autoproteolytically cleaved, polyprotein precursor. The enhanced recombinant protein expression in bacteria and insect cells led to undesired aggregation of the individually produced Nsps. However, it might still be possible to find new, sophisticated systems to facilitate the expression and purification of Nsps for their crystallization. Such system might utilize polyprotein expression in eukaryotic cell lines to produce Nsps, in their native conformations and in quantities high enough to enable their structural studies.

This study also reports the first results regarding the phosphorylation sites of an alphavirus Nsp. The techniques used to map the phosphorylated aminoacid residues in Nsp3 were a combination of several different methods, and the same analytical strate-

gies could be applied to other phosphoproteins. The article V as well as the phosphorylation sites of SFV Nsp3 has been entered in the internet database of Protein Research Foundation (http://www.prf.or.jp/en/dbi.html under accession number 2620611), which is collecting the information related to amino acids, peptides and proteins. The localization of the Nsp3 phosphorylation sites provided the opportunity to study the effects of elimination of Nsp3 phosphorylation, both biochemically and in the context of SFV infection. The Nsp3 phosphorylation defective mutant virus was able to replicate in cell culture to titers similar to that of wt SFV, but appeared to be avirulent for adult mice. This observation can be utilized in generating non-pathogenic SFV-vector for protein expression in a number of different higher eukaryotic systems. In addition, future studies could investigate the potential phosphoesterase activity at the N-terminal portion of Nsp3. A careful characterization might add knowledge to the function(s) of Nsp3 phosphorylation, as well as the function of Nsp3 in the SFV infection cycle.

ACKNOWLEDGEMENTS

This work was carried out at the Institute of Biotechnology during the years 1994-2001, and financially supported by the University of Helsinki and the Finnish Academy of Sciences. I'd like to thank Prof. Mart Saarma, the head of the Institute, for providing an excellent work environment as well as for helping me in various occasions during my work. Prof. Simo Laakso, the head of the Laboratory of Biotechnology and Microbiology in the Department of Chemical Technology, I would like to thank for his encouragement and positive attitude towards this thesis.

I wish to express my gratitude to my supervisors Professors Marja Makarow and Leevi Kääriäinen for patiently introducing me to the world of molecular biology, which for me, was a totally new field of knowledge. Both not only provided me with well-equipped work facilities, but also the needed encouragement to continue when everything was not proceeding so smoothly.

The reviewers Doc. Johan Peränen and Doc. Anu Jalanko are thanked for their careful reading of the manuscript as well as their constructive comments and suggestions to improve it. Dr. Patrick Russo is gratefully acknowledged for the revision of the text of this thesis.

I want to thank M.Sc. Juhani Saarinen, from the protein laboratory, for a pleasant collaboration. Besides being well versed in state of the art mass spectrometry (MS), he is also extremely co-operative and patient. I'm also grateful to Dr. Nisse Kalkkinen for introducing me to the techniques of protein purification and advising me on numerous occasions. The assistants from the Biochemistry Department, Dr. Leena Valmu and M.Sc. Tiina Hildén, are thanked for introducing me to phosphoamino acid and phosphopeptide analyses. I have never received such exact and detailed instructions from any other laboratory.

I would like to thank all the personnel at the Institute, especially those of the former and present Yeast and SFV groups, for creating dynamic and enjoyable working environments. Especially Airi Arffman, Anne Ikäheimonen, Jukka Hiltunen, Kimmo Karhi, Reetta Kettunen, Pekka Kujala, Julia Magden, Tiina Pummi, Patrick Russo, Nina Saris, Marjo Simonen, Airi Sinkko, Erja Tiihonen and Tarja Välimäki are thanked for their help and friendship. Anne, Pekka, Julia ja Tarja are also thanked for the most informative and important moments during the day "coffee time", as well as for the nice occasions outside of work. I also want to thank Prof. Per Saris and Dr. Nina Saris for adding to my knowledge of bacteria and yeast, as well as for other occasions outside work. Eija Jokitalo is acknowledged for guiding me though the very first step in my carrier as a 'protein purifier'. Tero Ahola is acknowledged for pieces of advise such as "be my own supervisor" and "do only those experiments, which can be published". I only wished I had followed more of his advise earlier.

I also want to acknowledged the "computer fellows" Jonne Viljanen, Christophe Roos and Iikka Salo. Their facilities and assistance concerning the net and programs have been excellent, and on many occasions I've been proud to be an Operoni-user. Special thanks are also due to all the users of the Confocal microscope, especially Prof. Marjatta Raudaskoski and M.Sc. Ulla Karvonen, who kept me sane by providing a distraction from my routine in many different ways.

I own my deepest gratitude to my friends: Pat, Jouko, Ilkka, Jonna, Med and Fynn for their support and encouragement, which have been crucial for me during the years I've been purifying SFV Nsps and calculating the number of CPVIs. I really feel lucky to have you as my friends to share life's joys and sorrows.

Finally, I want to express my gratitude to my parents for understanding and supporting me in all my different life phases.

Helsinki, 21st of February, 2001

Helena

REFERENCES

Abrahmsén, L., Moks, T., Nilsson, B. and Uhlén, M. (1986). Secretion of heterologous gene products to the culture medium of *Escherichia coli*. *Nucleic Acids Res.* **14**, 7487-7500.

Agapov, E.V., Frolov, I., Lindenbach, B.D., Pragai, B.M., Schlesinger, S. and Rice, C.M. (1998). Noncytophathic Sindbis virus RNA vectors for heterologous gene expression. *Proc. Natl. Acad. Sci. USA* **95**, 12989-12994.

Aguilar, M.-I. and Hearn, M.T.W. (1996). High-resolution reversed-phase high-performace liquid chromatography of peptides and proteins. *Methods Enzymol.* **270**, 3-26.

Ahola, T. and Kääriäinen, L. (1995). Reaction in alphavirus mRNA capping: Formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl. Acad. Sci. USA* **92**, 507-511.

Ahola, T., Lampio, A., Auvinen, P. and Kääriäinen, L. (1999). Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity. *EMBO J.* **11**, 3164-3172.

Airanne, K.J. and Kulomaa, M.S. (1995). Rapid purification of recombinant proteins fused to chicken avidin. *Gene* **167**, 63-68.

Airanne, K.J., Oker-Blom, C., Marjomäki, V.S., Bayer, E.A., Wilchek, M. and Kulomaa, M.S. (1997). Production of biologically active recombinant avidin in baculovirus -infected insect cells. *Protein Expr. Purif.* **9**, 100-108.

Airenne, K.J., Laitinen, O.H., Alenius, H., Mikkola, J., Kalkkinen, N., Arif, S.A.M., Yeang, H.Y., Palosuo,T. and Kulomaa, M.S. (1999). Avidin is a promising tag for fusion proteins produced in baculovirus-infected insect cells. *Protein Expr. Purif.* **17**, 139-145.

Aitken, R., Gilchrist, J. and Sinclair, M.C. (1994). Vectors to facilitate the creation of translational fusion to the maltose-binding protein of *Escherichia coli*. *Gene* **144**, 69-73.

Amrein, K.E., Takacs, B., Stieger, M., Molnos, J., Flint, N.A. and Burn, P. (1995). Purification and characterization of recombinant human p50csk protein-tyrosine kinase from *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. *Proc. Natl. Acad. Sci. USA* **92**, 1048-1052.

Anderegg, R.J., Betz, R., Carr, S.A., Crabb, J.W. and Duntze, W. (1988). Structure of *Saccharomyces cerevisiae* mating hormone α-factor. Identification of S-farnesyl cysteine as a structural component. *J. Biol. Chem.* **263**, 18236-18240.

Argos, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909-9916.

Ashford, D.A. (1992). Carbohydrate analysis. *Curr. Opin. Biotechnol.* **3**, 45-48.

Atkins, G.J., Sheahan, B.J. and Liljeström, P. (1999). The molecular pathogenesis of Semliki Forest virus: a model virus made useful? *J. Gen. Virol.* **80**, 2287-2297.

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M. and Possee R.D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**, 586-605.

Balch, W.E., McCaffery, J.M., Plutner, H., Farquhar, M.G. (1994). Vesicular stomatitis virus glycoprotein is sorted and concentrated during export from the endoplasmic reticulum. *Cell* **76**, 841-852.

Baldwin, A.N., Bitler, C.M., Welcher, A.A. and Shooter, E.M. (1992). Studies on the structure and binding properties of the cysteine-rich domain of rat low affinity nerve growth factor receptor ($p75^{NGFR}$). *J. Biol. Chem.* **267**, 8352-8359.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411-421.

Banks Jr., J.F. and Whitehouse, C.M. (1996). Electrospray ionization mass spectrometry. *Methods Enzymol.* **270**, 486-519.

Banner, D.W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.J., Broger, C., Loetscher, H. and Lesslauer, W. (1993). Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* **73**, 431-445.

Barik, S. and Banerjee, A.K. (1992). Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6570-6574.

Barton, D.J., Sawicki, S.G. and Sawicki, D.L. (1988). Demonstration *in vitro* of temperature-sensitive elongation of RNA in Sindbis virus mutant *ts6*. *J. Virol.* **62**, 3597-3602.

Batas, B. and Chaudhuri, J.B. (1999). Considerations of sample application and elution during size-exclusion chromatography-based protein refolding. *J. Chromatogr.* **864**, 229-236.

Beavis, R.C. and Chait, B.T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol.* **270**, 519-534.

Bednarek, S.Y., Racazzola, M., Hosobuchi, M., Amherdt, M. and Orci, L. (1995). COPI and COPII-coated vesicles bud directly from the endoplasmic reticulum. *Cell* **83**, 1183-1196.

Berglund, P., Sjöberg, M., Garoff, H., Atkins, G.J., Sheahan, B.J. and Liljeström, P. (1993). Semliki Forest virus expression system: production of conditionally infectious recombinant particles. *Bio/Technology* **11**, 916-920.

Birkenmeier, G., Usbeck, E. and Kopperschlager, G. (1984). Affinity partitioning of albumin and alpha-fetoprotein in an aqueous two-phase system using poly(ethylene glycol)-bound triazine dyes. *Anal. Biochem.* **136**, 264-271.

Bizzozero, O.A. (1995). Chemical analysis of acylation sites and species. *Methods Enzymol.* **250**, 361-379.

Boeke, J.D., Xu, H. and Fink, G.R. (1998). A general method for the chromosomal amplification of genes in yeast. *Science* **239**, 280-282.

Bonning, B.C., Roelvink, P.W., Vlak, J.M., Possee, R.D. and Hammock, B.D. (1994). Superior expression of juvenile hormone esterase and ß-galactosidase from the basic protein promoter of *Autographa californica* nuclear polyhedrosis virus compared to the p10 protein and polyherdin promoters. *J. Gen. Virol.* **75**, 1551-1556.

Bostian, K.A., Elliott, Q., Bussey, H., Burn, V., Smith, A. and Tipper, D.J. (1984). Sequence of the preprotoxin dsRNA *gene* of type I killer yeast: multiple processing events produce a two-component toxin. *Cell* **36**, 741-751.

Botstein, D., Falco, S.C., Stewart, S.E., Brennan, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979). Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. *Gene* **8**, 17-24.

Boyer, P.M. and Hsu, J.T. (1993). Protein purification by dye-ligand chromatography. *Adv. Biochem. Eng. Biotechnol.* **49**, 1-44.

Bradish, C.J., Allner, K. and Maber, H.B. (1971). The virulence of original and derived strains of Semliki forest virus for mice, guinea-pigs and rabbits. *J. Gen. Virol.* **12**, 141-160.

Bredenbeek, P.J., Frolov, I., Rice, M.C. and Schlesinger, S. (1993). Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. *J. Virol.* **67**, 6439-6446.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De francesco, R. and Rey, F.A. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. USA.* **96**, 13034-13039.

Brizzard, B.L., Chubet, R.G. and Vizard, D.L. (1994). Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *BioTechniques* **16**, 730-734.

Brodsky, J.L. and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteo-liposome. *J. Cell Biol.* **123**, 1355-1363.

Buckholz, R.G. and Gleeson, M.A.G. (1991). Yeast systems for the commercial production of heterologous proteins. *Bio/Technology* **9**, 1067-1072.

Bukari, A.L. and Zipser, D. (1973). Mutants of *Escherichia coli* with a defect in the degration of nonsense fragments. *Nature* **243**, 238-241.

Bussineau, C.M. and Shuster, J.A. (1994). Genetic stability for protein expression systems in yeast. *Dev. Biol. Stand.* **83**, 13-19.

Büning, H., Gärtner, U., von Schack, D., Baeuerle, P.A. and Zorbas, H. (1996). The histidine of recombinant DNA binding proteins may influence the quality of interaction with DNA. *Anal. Chem.* **234**, 227-230.

Byrappa, S., Pan, Y.-B. and Gupta, K.C. (1996). Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein. *Virology* **216**, 228-234.

Byrd, J.C., Tarentino, A.L., Maley, F., Atkinson, P.H. and Trimble, R.B. (1982). Glycoprotein synthesis in yeast. Identification of Man8GlcNAc2 as an essential intermediate in oligosaccharide processing. *J. Biol. Chem.* **257**, 14657-14666.

Caldas, T.D., El Yaagoubi, A., Kohiyama, M. and Richarme, G. (1998). Purification of elongation factors EF-Tu and EF-G from *Escherichia coli* by covalent chromatography on thiol-sepharose. *Protein Expr. Purif.* **14**, 65-70.

Caliguiri, L.A. and Tamm, I. (1970). The role of cytoplasmic membranes in poliovirus biosynthesis. *Virology* **42**, 100-111.

Carr, S.A., Huddleston, M.J. and Annan, R.S. (1996). Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* **239**, 180-192.

Cartwright, C.P., Li, Y., Zhu, Y.-S., Kang, Y.-S. and Tipper, D.J. (1994). Use of ß-lactamase as a secreted reporter of promoter function in yeast. *Yeast* **10**, 497-508.

Cartwright, C.P., Zhu, Y.-S. and Tipper, D.J. (1992). Efficient secretion in yeast based on fragments from K_1 killer preprotoxin. *Yeast* **8**, 261-272.

Chao, M.V. (1992). Neurotrophin receptors: a window into neuronal differentiation. *Neuron* **9**, 583-593.

Chazenbalk, G. and Rapoport, B. (1995). Expression of the extracellular domain of the thyrotropin receptor in the baculovirus system using a promoter active earlier than the polyhedrin promoter. Implications for the expression of functional highly glycosylated proteins. *J. Biol. Chem.* **270**, 1543-1549.

Chen, Y., Pioli, D. and Piper, P.W. (1994). Overexpression of the gene for polyubiquitin in yeast confers increased secretion of human leucocyte protease inhibitor. *Bio/Technology* **12**, 819-823.

Chicz, R.M. and Regnier, F.E. (1990). Microenvironmental contributions to the chromatographic behavior of subtilisin in hydrophobic-interaction and reversed-phase chromatography. *J. Chromotogr.* **500**, 503-518.

Chlenov, M.A., Kandyba, E.I., Nagornaya, L.V., Orlova, I.L. and Volgin, Y.V. (1993). High-performance liquid chromatography of human glycoprotein hormones. *J. Chromatogr.* **631**, 261-267.

Choudhary, G. and Horváth, C. (1996). Ion-exchange chromatography. *Methods Enzymol.* **270**, 47-82.

Chubet, R.G. and Brizzard, B.L. (1996). Vectors for expression and secretion of FLAG epitope-tagged proteins in mammalian cells. *BioTechniques* **20**, 136-141.

Chung, B.H., Nam, S.W., Kim, B.M. and Park, Y.H. (1996). Highly efficient secretion of heterologous proteins from *Saccharomyces cerevisiae* using inulinase signal peptides. *Biotechnol. Bioeng.* **49**, 473-479.

Chung, B.H., Seo, D.J. and Nam, S.W. (1999). High-level secretory production of recombinant human lipocortin-I by *Saccharomyces cerevisiae*. *Proc. Biochem.* **35**, 97-101.

Cichna, M., Markl, P. and Huber, J.F. (1995). Determination of true octanol-water partition coefficients by means of solvent generated liquid-liquid chromatography. *J. Pharm. Biomed. Anal.* **13**, 339-351.

Clarke, H.R.G., Wolfson, M.F., Rauch, C.T., Castner, B.J., Huang, C.-P., Gerhart, M.J., Johnson, R.S., Cerretti, D.P., Paxton, R.J., Price, V.L. and Black, R.A. (1998). Expression and purification of correctly processed active human TACE catalytic domain in *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **13**, 104-110.

Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L. and Hakomori, S. (1990). Isolation to homogeneity and partial characterization of a histo-blood group A defined Fuc alpha 1-2Gal alpha 1-3-N-acetyl-galactosaminyltransferase from human lung tissue. *J. Biol. Chem.* **265**, 1139-1145.

Conibear, E. and Stevens, T.H. (2000). Vps52p, Vps53p and Vps54p from a novel multisubunit complex required for protein sorting at the yeast late Golgi. *Mol. Biol. Cell* **11**, 305-323.

Cook, J.C., Schultz, L.D., Huang, J., George, H.A., Wayne, K., Herber, C. Ip., Joyce, J.G., Mao, S.-S., Markus, H.Z., Miller, W.J., Sardana, M.K. and Lehman, E.D. (1998). Expression and purification of recombinant tick anticoagulant peptide (Y1W/D10R) douple mutant secreted by *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **13**, 291-300.

Costello, C.E. (1999). Bioanalytic applications of mass spectrometry. *Curr. Opin. Biotechnol.* **10**, 22-28.

Cowley, D.J. and Mackin, R.B. (1997). Expression, purification and characterization of recombinant human proinsulin. *FEBS Lett.* **402**, 124-130.

Cronan, J.E. (1990). Biotination of proteins *in vivo*: A post-translational modification to label, purify, and study proteins. *J. Biol. Chem.* **265**, 10327-10333.

Curran, J., Marq, J.B. and Kolakofsky, D. (1995). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69**, 849-855.

Da Silva, N.A. and Bailey, J.E. (1991). Influence of plasmid origin and promoter strength in fermentation s of recombinant yeast. *Biotechnol. Bioeng.* **37**, 318-324.

Dahlman, K., Strömstedt, P.-E., Rae, C., Jörnvall, H., Flock, J.-I., Carlstedt-Duke, J. and Gustafsson, J.-Å. (1989). High level expression in *Escherichia coli* of DNA-binding domain of glucocortioid receptor in a functional form utilizing domain-specific cleavage of a fusion protein. *J. Biol. Chem.* **264**, 804-809.

Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999**).** New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* **65,** 382-388.

De Bernardez Clark, E. (1998). Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* **9**, 157-163.

Djordjevic, S. and Stock, A.M. (1997). Crystal structure of the chemotaxis receptor methyl-transferase CheR suggests a conserved structural motif for binding S-adenosylmethionine. *Structure* **5**, 545-558.

Dryga, S.A, Dryga, O.A and Schlesinger, S. (1997). Identification of mutation in a Sindbis virus variant to estabilish persistant infection in BHK cell: the importance of a mutation in the nsP2 gene. *Virology* **228**, 74-83.

Dubendorff, J.W. and Studier, F.W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**, 45-59.

Dunphy, J.T. and Linder, M.E. (1998). Signalling functions of protein palmitoylation. *Biochim. Biophys. Acta* **1436**, 245-261.

Durbin, S.D. and Feher, G. (1996). Protein crystallization. *Annu. Rev. Phys. Chem.* **47**, 171-204.

Eckart, M.R. and Bussineau, C.M. (1996). Quality and authenticity of heterologous proteins synthesized in yeast. *Curr. Opin. Biotechnol.* **7**, 525-530.

Ernst W., Grabherr R. and Katinger H. (1994). Direct Cloning into the Autographa californica Nuclear Polyhedrosis Virus Genome for the Generation of Recombinant Baculovirus. *Nucleic Acids Res.* **22**, 2855-2856.

Farrell, P.J., Lu, M., Prevost, J., Brown, C., Behie, L. and Iatrou, K. (1998). High-level expression of secreted glycoproteins in transformed lepidopteran insect cells using a novel expression vector. *Biotechnol. Bioeng.* **60**, 656-663.

Flaschel, E. and Friehs, K. (1993). Improvement of downstream processing of recombinant proteins by means of genetic engineering methods. *Biotech. Adv.* **11**, 31-78.

Fleer, R. (1992). Engineering yeast for high level expression. *Curr. Opin. Biotechnol.* **3**, 486-496.

Forage, R.G., Ring, J.M., Brown, R.W., McInerney, B.V., Cobon, G.S., Gregson, P., Robertson, D.M., Morgan, F.J., Hearn, M.T.W., Findlay, J.K., Wettenhall, R.E.H., Burger, H.G. and de Kretser, D.M. (1986). Cloning and sequence analysis of cDNA species coding for the two subunits of inhibin from bovine follicular fluid. *Proc. Natl. Acad. Sci. USA* **83**, 3091-3095.

Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K.M. and Hartmanis, M. (1991). Thrombin and H64A subtilisin cleavage of fusion proteins for preparation of human recombinant parathyroid hormone. *J. Protein Chem.* **10**, 517-526.

Frangioni, J.V. and Neel, B.G. (1992). Solubilization and purification of enzymatically active glutathione s-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**, 179-187.

Franken, K.L.M.C., Hiemstra, H.S., van Meijgaarden, K.E., Subronto, Y., den Hartigh, J., Ottenhoff, T.H.M. and Drijfhout, J.W. (2000). Purification of His-tagged proteins by immobilized chelated affinity chromatography: the benefits from the use of organic solvent. *Protein Expr. Purif.* **18**, 95-99.

Freitag, R. (1999). Displacement chromatography for biopolymer separation. *Nat. Biotechnol.* **17**, 300-302.

Froshauer, S., Kartenbeck, J. and Helenius, A. (1988). Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J. Cell Biol.* **107**, 2075-2086.

Fuerst, T.R., Niles, E.G., Studier, F.W. and Moss, B. (1986). Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **83**, 8122-8126.

Futcher, A.B. (1988). The 2µm circle plasmid of *Saccharomyces cerevisiae*. *Yeast* **4**, 27-40.

Gari, E., Piedrafita, L., Aldea, M. and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* **13**, 837-848.

Garoff, H. and Li, K.-J. (1998). Recent advances in gene expression using alphavirus vectors. *Curr. Opin. Biotechnol.* **9**, 464-469.

Geisow, M.J. (1991). Both bane and blessing-inclusion bodies. *Trends Biotechnol.* **9**, 368-369.

Geisse, S. and Kocher, H.P. (1999). Protein expression in mammalian and insect cell systems. *Methods Enzymol.* **306**, 19-41.

Geng, X. and Regnier, F.E. (1984). Retention model for proteins in reverse-phase liqued chromatography. *J. Chrom.* **296**, 15-30.

Georgiou, G. and Valax, P. (1999). Isolating inclusion bodies from bacteria. *Methods. Enzymol.* **306**, 48-58.

Gerrard, S.R., Bryant, N.J. and Stevens, T.H. (2000). *VPS21* controls entry of endocytosed and biosynthetic proteins into the yeast prevacuolar compartment. *Mol. Biol. Cell* **11**, 613-626.

Gerst, J.E. (1999). SNAREs and SNARE regulators in membrane fusion and exocytosis. *Cell Mol. Life Sci*. **55**, 707-734.

Gething, M.J. and Sambrook, J. (1992). Protein folding in the cell. *Nature* **355**, 33-45.

Glomset, J.A., Gelb, M.H. and Farnsworth, C.C. (1990). Prenyl proteins in eukaryotic cells: a new type of membrane anchor. *Trends Biochem. Sci.* **15**, 139-142.

Goffeau, A., Aert, R., Agostini-Carbone, M.L., Ahmed, A., Aigle, M., Alberghina, L., Albermann, K., Albers, M., Aldea, M., Alexandraki, D., Aljinovic, G., Allen, E., Altmann, R., Alt-Mörbe, J., André, B., Andrews, S., Ansorge, W., Antoine, G., Anwar, R., Aparicio A. *et al.* (1997). *Saccharomyces cerevisiae* Genome. *Nature* **387** (6632 Suppl), 1-105.

Goloubinoff, P., Gatenby, A.A. and Lorimer, G.H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44-47.

Gomez De Cedron, M., Ehsani, N., Mikkola, M.L., García, J.A. and Kääriäinen, L. (1999). RNA helicase activity of Semliki Forest virus replicase protein NSP2. *FEBS Lett.* **448**, 19-22.

Goodly, A.R. (1993). The production of heterologous plasma proteins. *Trends Biotechnol.* **11**, 430-433.

Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y. and Shimada, I. (1992). Three dimensional solution structure of the B domain of staphylococcal protein protein A: Comparison of the solution and crystal structure. *Biochemistry* **31**, 9665-9672.

Grand, R.J.A. (1989). Acylation of viral and eukaryotic proteins. *Biochem. J.* **258**, 625-638.

Griffiths, G. and Simons, K. (1986). The *trans* Golgi network: sorting at the exit site of the Golgi complex. *Science* **234**, 438-43.

Grimm, S.L. and Nordeen, S.K. (1999). Alternative procedure for two-dimensional separation of phosphoamino acids. *BioTechniques* **27**, 62-64.

Gräslund, T., Nilsson, J., Lindberg, M., Uhlén, M. and Nygren, P.-Å. (1997). Production of a thermostable DNA polymerase by site specific cleavage of a heat-eluted affinity fusion protein. *Protein Expr. Purif.* **9**, 125-132.

Gupta, A.K., Blondel, D., Choudhary, S. and Banerjee, A.K. (2000). The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J. Virol.* **74**, 91-98.

Görg, A., Postel, W. and Günther, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531-546.

Götte, M. and Lazar, T. (1999). The ins and outs of yeast vacuole trafficking. *Protoplasma* **209**, 9-18.

Hahn, Y.S., Grakoui, A., Rice, C.M., Strauss, E.G. and Strauss, J.H. (1989a). Mapping of RNA$^-$ temperature-sensitive mutants of Sindbis virus: Complementation group F mutants have lesions in nsP4. *J. Virol.* **63**, 1194-1202.

Hahn, Y.S., Strauss, E.G. and Strauss, J.H. (1989b). Mapping of RNA$^-$ temperature-sensitive mutants of Sindbis virus: Assignment of complementation groups A, B, and G to non-structural proteins. *J. Virol.* **63**, 3142-3150.

Hakes, D.J. and Dixon, J.E. (1992). New vectors for high level expression of recombinant proteins bacteria. *Anal. Biochem.* **202**, 293-298.

Hamman, B.D., Chen, J.C., Johnson, E.E. and Johnson, A.E. (1997). The aqueous pore through the translocon has a diameter of 40-60 Å during cotranslational protein translocation at the ER membrane. *Cell* **89**, 535-44.

Hammond, C. and Helenius, A. (1995). Quality control in the secretory pathway. *Curr. Opin. Cell Biol.* **7**, 523-529.

Hanein, D., Matlack, K.E., Jungnickel, B., Plath, K., Kalies, K.U., Miller, K.R., Rapoport, T.A. and Akey, C.W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87**, 721-32.

Hansen, J.L., Long, A.M. and Schultz, S. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **5**,1109-1122.

Hansson, M., Ståhl, S., Hjorth, R., Uhlén, M. and Moks, T. (1994). Single-step recovery of a secreted recombinant protein by expanded bed adsorption. *Bio/Technology* **12**, 285-288.

Hardy, W.R. and Strauss, J.H. (1989). Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. *J. Virol.* **63**, 4653-4664.

Harris, L.J., Skaletsky, E. and McPherson, A. (1995). Crystallization of intact monoclonal antibodies. *Proteins* **23**, 285-289.

Heeter, G.A. and Liapis, A.I. (1998). Frontal chromatography of proteins. Effect of axial dispersion on column performance. *J. Chromatogr.* **796**, 157-164.

Helenius, A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* **5**, 253-265.

Helenius, A., Kartenbeck, J., Simons, K. and Fries, E. (1980). On the entry of Semliki Forest virus in BHK-21 cells. *J. Cell Biol.* **84**, 404-420.

Helenius, A., Marquardt, T. and Braakman, I. (1992). The endoplasmic reticulum as a protein folding compartment. *Trends Cell Biol.* **2**, 227-231.

Henderson, L.E., Oroszlan, S. and Konigsberg, W. (1979). A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction. *Anal. Biochem.* **93**, 153-157.

Henning, L. and Schäfer, E. (1998). Protein purification with C-terminal fusion of maltose binding protein. *Protein Expr. Purif.* **14**, 367-370.

Henry, M.P. (1991). Design requirements of silica-based matrices for biopolymer chromatography. *J. Chromatogr.* **544**, 413-443.

Himly, M., Pfleiderer, M., Holzer, G., Fischer, U., Hannak, E., Falkner, F.G. and Dorner, F. (1998). Defective vaccinia virus as abiologically safe tool for the overproduction of recombinant human secretory proteins. *Protein Expr. Purif.* **14**, 317-326.

Hjelmeland, L.M. (1990). Removal of detergents from membrane proteins. *Methods Enzymol.* **182**, 277-282.

Hockney, R.C. (1994). Recent developments in heterologous protein production in *Escherichia coli*. *Bio/Technology* **12**, 456-463.

Hodgson, J. (1993). Expression systems: A user's guide. *Bio/Technology* **11**, 887-893.

Holkeri, H. and Makarow, M. (1998). Different degradation pathways for heterologous glycoproteins in yeast. *FEBS Lett.* **429**, 162-166.

Holkeri, H., Paunola, E., Jämsä, E. and Makarow, M. (1998). Dissection of the translocation and chaperoning functions of yeast BiP/Kar2p *in vivo*. *J. Cell Sci.* **111**, 749-757.

Honda, F., Honda, H. and Koishi, M. (1992). Application of non-porous silica ultramicrospheres to high-performance liquid chromatography column packings. *J. Chromatogr.* **609**, 49-59.

Hong, E., Davidson, A. and Kaiser, C. (1996). A pathway for targeting soluble misfolded proteins to the yeast vacuole. *J. Cell Biol.* **135**, 623-633.

Hopp, T.H., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Cerretti, D.P., Urdal, D.L. and Conlon, P.J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6**, 1204-1210.

Horazdovsky, B., DeWald, D. and Emr, S. (1995). Protein transport to the yeast vacuole. *Curr. Opin. Cell Biol.* **7**, 544-551.

Hoshino, K., Taniguchi, M., Kitao, T., Morohashi, S. and Sasakura, T. (1998). Preparation of a new thermo-responsive adsorbent with maltose as a ligand and its application to affinity precipitation. *Biotechnol. Bioeng.* **60**, 568-579.

Huang, H.V. (1996). Sindbis virus vector for expression in animal cells. *Curr. Opin. Biotechnol.* **7**, 531-535.

Hunt, D.F., Yates, III, J.R., Shabanowitz, J., Winston, S. and Hauer, C.R. (1986). Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA* **83**, 6233-6237.

Hunter, T. (1995). Protein kinase and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225-236.

Huntley, C.C., De, B.P. and Banerjee, A.K. (1997). Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase C zeta. *J. Biol. Chem.* **272**, 16578-16584.

Hwang, C., Sinskey, A.J. and Lodish, H.F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. *Science* **257**, 1496-502.

Høgset, A., Blingsmo, O.R., Saether, O., Gautvik, V.T., Holmgren, E., Hartmanis, M., Josephson, S., Gabrielsen, O.S., Gordeladze, J.O., Aleström, P. and Gautvik, K.M. (1990). Expression and characterization of recombinant human parathyroid hormone secreted by *Escherichia coli* employing the Staphylococcal protein A promoter and signal sequence. *J. Biol. Chem.* **265**, 7338-7344.

Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989). Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. *J. Biol. Chem.* **264**, 20131-20139.

Janknecht, R. and Nordheim, A. (1992). Affinity purification of histidine-tagged protein transiently produced in HeLa cells. *Gene* **121**, 321-324.

Janssen J.J.M., Bovee-Geurts, P.H.M, Merkx, M. and DeGrip, W.J. (1995). Histidine tagging both allows convenient single-step purification of bovine rhodopsin and exerts ionic strength-dependent effects on its photochemistry. *J. Biol. Chem.* **270**, 11222-11229.

Jarvis, D.L. and Finn, E.E. (1996). Modifying the insect cell N-glycosylation pathway with immediate early baculovirus expression vectors. *Nat. Biotechnol.* **14**, 1288-1292.

Jarvis, D.L., Fleming, J.-A.G.W., Kovacs, G.R., Summers, M.D. and Guarino, L.A. (1990). Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stable transformed Lepidopteran cells. *Bio/Technology* **8**, 950-955.

Jarvis, D.L., Weinkauf, C. and Guarino, L.A. (1996). Immediate-early baculovirus vectors for foreign gene expression in transformed or infected insect cells. *Protein Expr. Purif.* **8**, 191-203.

Jelinek, T. and Weber, M.J. (1993). Optimization of the resolution of phosphoamino acids by one-dimensional thin-layer electrophoresis. *BioTechniques* **15**, 628-630.

Jendeberg, L., Tashiro, M., Tejero, R., Lyons, B.A., Uhlén, M., Montelione, G.T. and Nilsson, B. (1996). The mechanism of binding staphylococcal protein A to immunoglobin G does involve helix unwinding. *Biochemistry* **35**, 22-31.

Jimenez, A. and Davies, J. (1980). Expression of a transposable antibiotic resistance element in *Saccharomyces*. *Nature* **287**, 869-871.

Jones, I. and Morikawa, Y. (1996). Baculovirus vectors for expression in insect cells. *Curr. Opin. Biotechnol.* **7**, 512-516.

Julius, D., Schekman, R. and Thorner, J. (1984). Glycosylation and processing of prepro-alpha-factor through the yeast secretory pathway. *Cell* **36**, 309-318.

Jäger, J. and Pata, J.D. (1999). Getting a grip: polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **9**, 21-28.

Jämsä, E., Simonen, M. and Makarow, M. (1994). Selective retention of secretory proteins in the yeast endoplasmic reticulum by treatment of cells with a reducing agent. *Yeast* **10**, 355-370.

Kaiser, C.A. and Schekman, R. (1990). Distinct sets of *SEC* genes govern transport vesicle formation and fusion early in the secretory pathway. *Cell* **61**, 723-733.

Keith, M.B., Farrell, P.J., Iatrou, K. and Behie, L.A. (1999). Screening of transformed insect cell lines for recombinant protein production. *Biotechnol. Prog.* **15**, 1046-1052.

Kennelly, P.J. and Krebs, E.G. (1991). Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266**, 15555-15558.

Keränen, S. and Kääriäinen, L. (1974). Isolation and basic characterization of temperature-sensitive mutants from Semliki Forest virus. *Acta Path. Microbiol. Scand. Sect. B* **82**, 810-820.

Kitts, P. and Possee, R. (1993). A method for producing recombinant baculovirus expression vectors at high frequency. *BioTechniques* **14**, 810-817.

Kim, J.S. and Raines, R.T. (1993). Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci.* **2**, 348-356.

Kim, J.S. and Raines, R.T. (1994). Peptide tags for a dual affinity fusion system. *Anal. Biochem.* **219**, 165-166.

Kitts, P., Ayres, M. and Possee, R. (1990). Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vector. *Nucleis Acids res.* **18**, 5667-5672.

Kjeldsen, T., Andersen, A.S., Hach, M., Diers, I., Nikolajsen, J. and Markussen, J. (1998a). α-factor pro-peptide N-linked oligosaccharides facilitate secretion of the insulin precursor in *Saccharomyces cerevisiae*. *Biotechnol. Appl. Biochem.* **27**, 109-115.

Kjeldsen, T., Hach, M., Balschmidt, P., Havelund, S., Petterson, A.F. and Markussen, J. (1998b). Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **14**, 309-316.

Kjeldsen, T., Pettersson, A.S., Hach, M., Diers, I., Havelund, S., Hansen, P.H. and Andersen, A.S. (1997). Synthetic leaders with potential BiP binding mediate high-yield secretion of correctly folded insulin precursors from *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **9**, 331-336.

Klaassen, C.H.W., Bovee-Geurts, P.H.M, DeCaluwé, G.L.J. and DeGrip, W.J. (1999). Large-scale production and purification of functional recombinant bovine rhdopsin with the use of the baculovirus expression system. *Biochem. J.* **342**, 293-300.

Knappik, A. and Plückthun, A. (1994). An improved affinity tag based on the FLAG peptide for detection and purification of recombinant antibody fragments. *BioTechniques* **17**, 754-761.

Ko, Y.H. and Pedersen, P.L. (1995). The first nucleotide binding fold of the cystic fibrosis transmembrane conductance regulator can function as an active ATPase. *J. Biol. Chem.* **270**, 22093-22096.

Kohanski, R.A. and Lane, M.D. (1990). Monovalent avidin affinity columns. *Methods Enzymol.* **184**, 195-200.

Koonin, E.V. and Dolja, V.V. (1993). Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28**, 375-430.

Koonin, E.V., Tatusov, R.L. and Rudd, K.E. (1995). Sequence similarity analysis of *Escherichia coli* proteins: functional and evolutionary implications. *Proc. Natl. Acad. Sci. USA* **92**, 11921-11925.

Kost, T.A. and Condreay, J.P. (1999). Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr. Opin. Biotechnol.* **10**, 428-433.

Kraulis, P.J., Jonasson, P., Nygren, P.-Å., Uhlén, M., Jendeberg, L., Nilsson, M. and Kördel, J. (1996). The serum albumin-binding domain of streptococcal protein G is a three-helical bundle: A heteronuclear NMR study. *FEBS Lett.* **378**, 190-194.

Kujala, P., Ikäheimonen, A., Ehsani, N., Vihinen, H., Auvinen, P. and Kääriäinen, L. (2001). The biogenesis of Semliki Forest virus replication complex utilizes endosomal apparatus. *J. Virol.* in press.

Kukuruzinska, M.A., Bergh, M.L.E. and Jackson, B.L. (1987). Protein glycosylation in yeast. *Ann. Rev. Biochem.* **56**, 915-944.

Kääriäinen, L., Takkinen, K., Keränen, S. and Söderlund, H. (1987). Replication of the genome of alphaviruses. *J. Cell Sci. Suppl.* **7**, 231-250.

Laakkonen, P., Ahola, T. and Kääriäinen, L. (1996). The effects of palmitoylation on membrane association of Semliki Forest virus RNA capping enzyme. *J. Biol. Chem.* **271**, 28567-28571.

Laakkonen, P., Auvinen, P., Kujala, P. and Kääriäinen, L. (1998). Alphavirus replicase protein Nsp1 induces filopodia and rearrangement of actin filaments. *J. Virol.* **72**, 10265-10269.

Laakkonen, P., Hyvönen, M., Peränen, J. and Kääriäinen, L. (1994). Expression of Semliki Forest virus nsP1-specific methyltransferase in insect cells and in *Escherichia coli*. *J. Virol.* **68**, 7418-7425.

Labbé, S. and Thiele, D.J. (1999). Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol.* **306**, 145-153.

Ladisch, M.R. and Kohlmann, K.L. (1992). Recombinant human insulin. *Biotechnol. Prog.* **8**, 469-478.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **259**, 680-685.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. and Hartl, F.U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**, 683-689.

LaStarza, M.W., Grakoui, A. and Rice, C.M. (1994a). Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: Effects on phoshorylation and on virus replication in vertebrate and invertebrate cells. *Virology* **202**, 224-232.

LaStarza, M.W., Lemm, J.A. and Rice, C.M. (1994b). Genetic analysis of the nsP3 region of Sindbis virus: Evidence for roles in minus-strand and subgenomic RNA synthesis. *J. Virol.* **68**, 5781-5791.

LaVallie, E.R. and McCoy, J.M. (1995). Gene fusion expression system in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 501-506.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in *E. coli* cytoplasm. *Bio/Technology* **11**, 187-193.

Lee, J., Choi, S.-I., Jang, J.S., Jang, K., Moon, J.W., Bae, C.S., Yang, D.S. and Seong, B.L. (1999). Novel secretion system of recombinant Saccharomyces cerevisiae using an N-terminus residue of human IL-1ß as secretion enhancer. *Biotechnol. Prog*. **15**, 884-890.

Lee, S.C. and Olins, P.O. (1992). Effect of overproduction of heat shock chaperones GroESL and DnaK on human procollagenase production in *Escherichia coli*. *J. Biol. Chem.* **267**, 2849-2852.

Lemm, J.A. and Rice, C.M. (1993a). Assembly of functional Sindbis virus RNA replication complexes: Requirement for coexpression of P123 and P34. *J. Virol.* **67**, 1905-1915.

Lemm, J.A. and Rice, C.M. (1993b). Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. *J. Virol.* **67**, 1916-1926.

Lemm, J.A., Rümenapf, T., Strauss, E.G., Strauss, J.H. and Rice, C.M. (1994). Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J.* **13**, 2925-2934.

Leusch, M.S., Lee, S.C. and Olins, P.O. (1995). A novel host-vector system for direct selection of recombinant baculoviruses (bacmids) in *Escherichia coli*. *Gene* **160**, 191-194.

Li, G., LaStarza, M.W., Hardy, W.R., Strauss, J.H. and Rice, C.M. (1990). Phosphorylation of Sindbis virus nsP3 *in vivo* and *in vitro*. *Virology* **179**, 416-427.

Licari, P. and Bailey, J.E. (1992). Modeling the population dynamics of baculovirus: infected insect cells: Optimizing infection strategies for enhanced recombinant protein yields. *Biotechnol. Bioeng.* **39**, 432-441.

Lilie, H., Schwarz, E. and Rudolph, R. (1998). Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol.* **9**, 497-501.

Liljeström, P. and Garoff, H. (1991). A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Bio/Technology* **9**, 1356-1361.

Lin Cereghino, G.P. and Cregg, J.M. (1999). Applications of yeast in biotechnology: protein production and genetic analysis. *Curr. Opin. Biotechnol.* **10**, 422-427.

Lis, H. and Sharon, N. (1993). Protein glycosylation: Structural and functional aspects. *Eur. J. Biochem.* **218**, 1-27.

Ljungquist, C., Jansson, B., Moks, T. and Uhlén, M. (1989). Thiol-directed immobilization of recombinant IgG-binding receptors. *Eur. J. Biochem.* **186**, 557-561.

Lo, K.-M., Sudo, Y., Chen, J., Li, Y., Lan, Y., Kong, S.M., Chen, L.L., An, Q. and Gillies, S.D. (1998). High level expression and secretion of Fc-X fusion proteins in mammalian cells *Protein Eng.* **11**, 495-500.

Lockwood, C.E., Bummer, P.M. and Jay, M. (1997). Purification of proteins using foam fractionation. *Pharm. Res.* **14**, 1511-1515.

Lodish, H.F., Kong, N. and Wikström, L. (1992). Calcium is required for folding of newly made subunits of the asialoglyco-protein receptor within the endoplasmic reticulum. *J. Biol. Chem*. **267**, 12753-12760.

Loison, G., Nguyen-Juilleret, M., Alouani, S. and Marquet, M. (1986). Plasmid-transformed *URA3 FUR1* douple mutants of *S. cerevisiae*: an autoselection system applicable to production of foreign proteins. *Bio/Technology* **4**, 433-437.

Lu, M., Farrell, P.J., Johnson, R. and Iatrou, K. (1997). A baculovirus (*Bombyx mori* nuclear polyhedrosis virus) repeat element functions as a powerful constitutive enhancer in transfected insect cells. *J. Biol. Chem.* **272**, 30724-30728.

Luckow, V., Lee, S., Barry, G. and Olins, P. (1993). Efficient generation of infectious recombinant baculoviruses by site- specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J. Virol.* **67**, 4566-4579.

Lundström, K., Vargas, A. and Allet, B. (1995). Functional activity of biotinylated human neurokinin 1 reseptor fusion expressed in the Semliki Forest virus system. *Biochem. Biophys. Res. Commun.* **208**, 260-266.

Lussier, M., Gentzsch, M., Sdicu, A.M., Bussey, H. and Tanner W. (1995). Protein O-glycosylation in yeast. The PMT2 gene specifies a second protein O-mannosyltransferase that functions in addition to the PMT1-encoded activity. *J. Biol. Chem.* **270**, 2770-2775.

Maachupalli-Reddy, J., Kelley, B. and De Bernadez-Clark, E. (1997). Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. *Biotechnol. Prog.* **13,** 144-150.

Makrides, S.C. (1999). Components of vector for gene transfer and expression in mammalian cells. *Protein Expr. Purif.* **17**, 183-202.

Manoharan, H.T., Gallo, J., Gulick, A.M. and Fahl, W.E. (1997). High-level production and purification of biologically active proteins from bacterial and mammalian cells using the tandem pGFLEX expression system. *Gene* **193**, 229-237.

Mao, Q.M., Stockmann, R., Prince, I.G. and Hearn, M.T. (1993). High-performance liquid chromatography of amino acids, peptides and proteins. CXXVI. Modelling of protein adsorption with non-porous and porous particles in a finite bath. *J. Chromatogr*. **646**, 67-80.

Marston, F.A.O and Hartley, D.L. (1990). Solubilization of protein aggregates. *Methods Enzymol*. **182**, 264-276.

Martin, S.A. and Moss, B. (1976). mRNA guanylyltransferase and mRNA (guanine-7-)-methyltransferase from Vaccinia virions. *J. Biol. Chem.* **251**, 7313-7321.

Martzen, M.R., McCraith, S.M., Spinelli, S.L., Torres, F.M., Fields, S., Grayhack, E. and Phizicky, E.M. (1999). A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153-1155.

Matlack, K.E.S., Misselwitz, B., Plath, K. and Rapoport, T.A. (1999). BiP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane. *Cell* **97**, 553-565.

Matsuo, E., Sampei, G., Mizobuchi, K. and Ito, K. (1999). The plasmid F Omp protease, a homologue of OmpT, as a potential obstacle to *E. coli*-based protein production. *FEBS Lett.* **461**, 6-8.

Mazumder, B. and Barik, S. (1994). Requirement of casein kinase II-mediated phosphorylation for the transcriptional activity of human respiratory syncytial viral phosphoprotein P: transdominant negative phenotype of phosphorylation-defective P mutants. *Virology* **205**, 104-111.

McIlhinney, R.A.J. (1990). The fats of life: the importance and function of protein acylation. *Trends Biochem. Sci.* **15**, 387-391.

Melander, W.R., Corradini, D. and Horvath, C. (1984). Salt-mediated retention of proteins in hydrophobic-interaction chromatography. Application of solvophobic theory. *J. Chromatogr.* **317**, 67-85.

Mendoza-Vega, O., Hebert, C. and Brown, S.W. (1994). Production of recombinant hirudin by high cell density fed-batch cultivations of *Saccharomyces cerevisiae* strain: physiological considerations during the bioprocess design. *J. Biotechnol.* **32**, 249-259.

Menge, U., Fraune, E., Lehmann, J. and Kula, M.R. (1987). Purification of proteins from cell culture supernatants. *Dev. Biol. Stand.* **66**, 391-401.

Merry, T. (1999). Current techniques in protein glycosylation analysis. A guide to their application. *Acta Biochim. Pol.* **46**, 303-314.

Mertens, N., Remaut, E. and Fiers, W. (1995). Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter - based expression plasmids. *Bio/Technology* **13**, 175-179.

Mi, S. and Stollar, V. (1991). Expression of Sindbis virus nsP1 and methyltransferase activity in *Escherichia coli*. *Virology* **184**, 423-427.

Mi, S., Durbin, R.K., Huang, H.V., Rice, C.M. and Stollar, V. (1989). Association of the Sindbis virus RNA methyltransferase activity with the nonstructural protein nsP1. *Virology* **170**, 385-391.

Miller, H.I., Henzel., W.J., Gidgway, J.B, Kuang, W.-J., Chisholm, V. and Liu, C.-C. (1989). Cloning and expression of a yeast ubiquitin-protein cleaving activity in *Escherichia coli*. *Bio/ Technology* **7**, 698-704.

Miller, L.K. (1988). Baculoviruses as gene expression vectors. *Ann. Rev. Microbiol.* **42**, 177-199.

Missiakas, D., Georgopoulos, C. and Raina, S. (1993). Identification and characterization of the *Escherichia coli* gene *dsbB*, whose product is involved in the formation of disulfide bonds *in vivo*. *Proc. Natl. Acad. Sci. USA* **90**, 7084-7088.

Mitchell, D.A., Marshall, T.K. and Deschenes, R.J. (1993). Vectors for the inducible over-expression of glutathione S-transferase fusion proteins in yeast. *Yeast* **9**, 715-723.

Mitraki, A. and King, J. (1989). Protein folding intermediates and inclusion body formation. *Bio/ Technology* **7**, 690-697.

Moks, T., Abrahmsén, L., Nilsson, B., Hellman, U., Sjöquist, J. and Uhlén, M. (1986). Staphylo-coccal protein A consists of five IgG-binding regions. *Eur. J. Biochem.* **156**, 637-643.

Moss, B. (1996). Genetically engineered pox-viruses for recombinant gene expression, vaccination, and safety. *Proc. Natl. Acad. Sci. USA* **93**, 11342-11348.

Mumberg, D., Muller, R. and Funk, M. (1994). Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767-5768.

Mumberg, D., Muller, R. and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic back-grounds. *Gene* **156**, 119-122.

Murby, M., Uhlén, M. and Ståhl, S. (1996). Upstream strategies to minimize proteolytic degratation upon recombinant pruduction in *Escherichia coli*. *Protein Expr. Purif.* **7**, 129-136.

Müller, C. and Rinas, U. (1999). Renaturation of heterodimeric platelet-derived growth factor from inclusion bodies of recombinant *Escherichia coli* using size-exclusion chromatography. *J. Chromatogr.* **855**, 203-213.

Müller, S.S., Kamp-Hansen, P. and Dalboge, H. (1998). Comparison of expression systems in the yeast *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizo-saccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promotors from *Y. lipolytica*. *Yeast* **14**, 1267-1283.

Naggie, S. and Bentley, W.E. (1998). Appearance of protease activities coincides with p10 and polyhedrin-driven protein production in the baculovirus expression system: Effects on yield. *Biotechnol. Prog.* **14**, 227-232.

Nakanishi-Shindo, Y., Nakayama, K., Tanaka, A., Toda, Y. and Jigami, Y. (1993). Structure of the N-linked oligosaccharides that show the complete loss of α-1,6-polymannose outer chain from *och1*, *och1 mnn1* and *och1 mnn1 alg3* mutants of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **268,** 26338-26345.

Napp, S.J. and Da Silva, N.A. (1993). Enhancement of cloned gene product synthesis *via* autoselection in recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **41**, 801-810.

Neri, D., De Lalla, C., Petrul, H., Neri, P. and Winter, G. (1995). Calmodulin as a versatile tag for antibody fragments. *Biotechnol.* **13**, 373-377.

Netzer, W.J. and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343-349.

Neubauer, G. and Mann, M. (1999). Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal. Chem.* **71**, 235-242.

Neubert, T.A. and Johnson, R.S. (1995). High-resolution determination of protein-linked acyl groups. *Methods Enzymol.* **250**, 487-494.

Ng, D., Brown, J. and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **143**, 269-278.

Niesters, H.G. and Strauss, J.H. (1990). Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. *J. Virol.* **64**, 1639-1647.

Nilsson, B., Forsberg, G., Moks, T., Hartmanis, M. and Uhlén, M. (1992). Fusion proteins in biotechnology and structural biology. *Curr. Opin. Struct. Biol.* **2**, 569-575.

Nilsson, B., Ståhl, S., Lundeberg, J., Uhlén, M. and Nygren, P.-Å. (1997). Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif.* **11**, 1-16.

Nimura, N., Itoh, H., Kinoshita, T., Nagae, N. and Nomura, M. (1992). Fast protein separation by reversed-phase high performance liquid chromatography on octadecylsilyl-bonded nonporous silica gel. *J. Chromatogr.* **585**, 207-211.

Nygren, P.-Å., Ståhl, S. and Uhlén, M. (1994). Engineering proteins to facilitate bioprocessing. *Trends Biotechnol.* **21**, 184-188.

Ohno, K., Sawai, K., Iiijima, Y., Levin, B. and Meruelo, D. (1997). Cell-spedific targeting of Sindbis virus vectors displaying IgG binding domains of proteins A. *Nat. Biotechnol.* **15**, 763-767.

Ollis, D. and White, S. (1990). Protein crystallization. *Methods Enzymol.* **182**, 646-659.

Ong, E., Greenwood, J.M., Gilkes, N.R., Kilburn, D.G., Miller Jr, R.C. and Warren, R.A.J. (1989). The cellulose-binding domains of cellulases: tools for biotechnology. *Trends Biotechnol.* **7**, 239-243.

Orci, L., Stamnes, M., Ravazzola, M., Amherdt, M., Perrelet, A., Söllner, T.H. and Rothman, J.E. (1997). Bidirectional transport by distinct populations of COPI-coated vesicles. *Cell* **90**, 335-349.

Ortega, S., García, J.L., Zazo, M., Varela, J., Muñoz-Willery, I., Cuevas, P. and Giménez-Gallego, G. (1992). Single-step purification on DEAE-sephacel of recombinant polypeptides produced in *Escherichia coli*. Bio/Technology **10**, 795-798.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S. and Rapoport, T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561-570.

Parekh, R.B. (1991). Effects of glycosylation on protein function. *Curr. Opin. Struct. Biol.* **1**, 750-754.

Park, Y.S., Shiba, S., Iijima S., Kobayashi, T. and Hishinuma, F. (1993). Comparison of 3 different promoter system for secretory alpha-amylase production in fed-batch cultures of recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **41**, 854-861.

Patra, A.K., Mukhopadhyay, R., Mukhija, R., Krishnan, A., Garg, L.C. and Panda, A.K. (2000). Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. *Protein Expr. Purif.* **18**, 182-192.

Pearson, J.D., Lin, N.T. and Regnier, F.E. (1982). The importance of silica type for reverse-phase protein separations. *Anal. Biochem.* **124**, 217-230.

Pedersen J., Lauritzen, C., Madsen, M.T. and Dahl, S.W. (1999). Removal of N-terminal poly-histidine tags from recombinant proteins using engineered aminopeptidases. *Protein Expr. Purif.* **15**, 389-400.

Pehrson, J.R. and Fuji, R.N. 1998. Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res.* **26**, 2837-2842.

Pekrun, K., Petry, H., Jentsch, K.D., Moosmayer, D., Hunsmann, G. and Luke, W. (1995). Expression and characterization of the reverse transcriptase enzyme from type 1 human immunodeficiency virus using different baculoviral vector systems. *Eur. J. Biochem.* **234**, 811-818.

Pérez-Martin, J., Cases, I. and de Lorenzo, V. (1997). Design of a solubilization pathway for recombinant polypeptides *in vivo* through processing of a bi-protein with a viral protease. *Protein Eng.* **10**, 725-730.

Peränen, J. (1991). Localization and phosphorylation of Semliki Forest virus non-structural protein nsP3 expressed in COS cells from a cloned cDNA. *J. Gen. Virol.* **72**, 195-199.

Peränen, J. and Kääriäinen, L. (1991). Biogenesis of type I cytopathic vacuoles in Semliki Forest virus-infected BHK cells. *J. Virol.* **65**, 1623-1627.

Peränen, J., Laakkonen, P., Hyvönen, M. and Kääriäinen, L. (1995). The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles. *Virology* **208**, 610-620.

Peränen, J., Rikkonen, M., Hyvönen, M. and Kääriäinen, L. (1996). T7 vectors with a modified T7*lac* promoter for expression of proteins in *Escherichia coli*. *Anal. Biochem.* **236**, 371-373.

Peränen, J., Rikkonen, M., Liljeström, P. and Kääriäinen, L. (1990). Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. *J. Virol.* **64** , 1888-1896.

Peränen, J., Takkinen, K., Kalkkinen, N. and Kääriäinen, L. (1988). Semliki Forest virus-specific non-structural protein nsP3 is a phosphoprotein. *J. Gen. Virol.* **69**, 2165-2178.

Pfleiderer, M., Falkner, F.G. and Dorner, F. (1995). Requirements for optimal expression of secreted and nonsecreted recombinant proteins in vaccinia virus systems. *Protein Expr. Purif* **6**, 559-569.

Plückthun, A. (1992). Mono- and bivalent antibody fragments produced in *Escherichia coli*: Engineering, folding and antigen binding. *Immunol. Rev.* **130**, 151-188.

Pohlner, J., Klause, T., Kuttler, E. and Halter, R. (1992). Sequence-specific cleavage of protein fusion using a recombinant *neisseria* type 2 IgA protease. *Bio/Technology* **10**, 799-804.

Porath, J. (1992). Immobilized metal ion affinity chromatography. *Protein Expr. Purif* **3**, 263-281.

Porath, J., Carlsson, J., Olsson, O. and Belfrage, G. (1975). Metal chelate affinity chromatography, a new approch to protein fractionation. *Nature* **258**, 598-599.

Power, J.F., Reid, S., Radford, K.M., Greenfield, P.F. and Nielsen, L.K. (1994). Modeling and optimization of the baculovirus expression vector system in batch suspension culture. *Biotechnol. Bioeng.* **44**, 710-719.

Pryor, K.D. and Leiting, B. (1997). High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* **10**, 309-319.

Pugachev, K.V., Mason, P.W., Shope, R.E. and Frey, T.K. (1995). Double-subgenomic Sindbis virus recombinants expressing immunogenic proteins of Japanese encephalitis virus induce significant protection in mice against lethal JEV infection. *Virology* **212**, 587-594.

Pugsley, A.P. (1993). The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* **57**, 50-108.

Pujar, N.S. and Zydney, A.L. (1998). Electrostatic effects on protein partitioning in size-exclusion chromatography and membrane ultrafiltration. *J. Chromatogr.* **796**, 229-238.

Radeke, M., Misko, T., Hsu, C., Herezenberg, L. and Shooter, E. (1987). Gene transfer and molecular cloning of the rat nerve growth factor receptor. *Nature* **325**, 593-597.

Rapoport, T.A., Jungnickel, B. and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271-303.

Redding, K., Holcomb, C. and Fuller, R.S. (1991). Immunolocalization of Kex2 protease identifies a putative late Golgi compartment in the yeast *Saccharomyces cerevisiae*. *J. Cell Biol.* **113**, 527-538.

Reed, K.E., Gorbalenya, A.E. and Rice, C.M. (1998). The NS5A/NS5 protein of viruses from three genera of the family *Flaviviridae* are phosphorylated by associated serine/thronine kinases. *J. Virol.* **72**, 6199-6206.

Reed, K.E., Xu, J. and Rice, C.M. (1997). Phosphorylation of the hepatitis C virus NS5A protein *in vitro* and *in vivo*: properties of the NS5A-accosiated kinase. *J. Virol.* **71**, 7187-7197.

Regnier, F.E. (1987). The role of protein structure in chromatographic behavior. *Science* **238**, 319-323.

Rexach, M.F. and Schekman, R.W. (1991). Distinct biochemical requirements for the budding, targeting, and fusion of ER-derived transport vesicles. *J. Cell Biol.* **114**, 219-229.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-1032.

Rikkonen, M. (1996). Functional significance of the nuclear-targeting and NTP-binding motifs of Semliki Forest virus nonstructural protein nsP2. *Virology* **218**, 352-361.

Rikkonen, M., Peränen, J. and Kääriäinen, L. (1992). Nuclear and nucleolar targeting signals of Semliki Forest virus nonstructural protein nsP2. *Virology* **189**, 462-473.

Rikkonen, M., Peränen, J. and Kääriäinen, L. (1994). ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. *J. Virol.* **68**, 5804-5810.

Robinson, A.S., Hines, V. and Wittrup, K.D. (1994). Protein disulphide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Bio/Technology* **12**, 381-384.

Robinson, A.S., Bocknaus, J.A., Voegler, A.C. and Wittrup, K.D. (1995). Reduction of BiP levels decreases heterologous protein secretion in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **271**, 10017-10022.

Roelvink, P.W., van Meer, M.M.M., de Kort, C.A.D., Possee, R.D., Hammock, B.D. and Vlak, J.M. (1992). Dissimilar expression of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus polyhedrin and p10 genes. *J. Gen. Virol.* **73**, 1481-1489.

Roggenkamp, R., Dargatz, H. and Hollenberg, C. (1985). Precursor of ß-lactamase is enzymatically inactive. Accumulation of the preprotein in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **260**, 1508-1512.

Roggenkamp, R., Dargatz, H. and Hollenberg, C.P. (1985). Precursor of beta-lactamase is enzymatically inactive. Accumulation of the preprotein in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **260**, 1508-1512.

Rohrmann, G. (1986). Polyhedrin structure. *J. Gen. Virol.* **67**, 1499-1513.

Romanos, M.A., Scorer, C.A. and Clare, J.J. (1992). Foreign gene expression in yeast: a review. *Yeast* **8**, 423-488.

Rossini, D., Porro, D., Brambilla, L., Venturini, M., Ranzi, B.M., Vanoni, M. and Albegrghina, L. (1993). In *Saccharomyces cerevisiae*, proteins secretion into the growth medium depends on environmental factors. *Yeast* **9**, 77-84.

Rozanov, M.N., Koonin, E.V. and Gorbalenya, A.E. (1992). Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J. Gen. Virol.* **73**, 2129-2134.

Russo, P., Kalkkinen, N., Sareneva, H., Paakkola, J. and Makarow, M. (1992a). A heat shock gene from *Saccharomyces cerevisiae* encoding a secretory glycoprotein. *Proc Natl Acad Sci USA* **89**, 3671-5. **Published erratum** correcting a sequencing error (1992). *Proc. Natl. Acad. Sci. USA* **89**, 8857.

Russo, G.L., Vandenberg, M.T., Yu, I.J., Bae, Y.-S., Franza, Jr.B.R. and Marshak, D.R. (1992b). Casein kinase II phosphorylates $p34^{cdc2}$ kinase in G1 phase of the HeLa cell division cycle. *J. Biol. Chem.* **267**, 20317-20325.

Russo, P., Simonen, M., Uimari, A., Teesalu, T. and Makarow, M. (1993). Dual regulation by heat and nutrient stress of the yeast HSP150 gene encoding a secretory glycoprotein. *Mol. Gen. Genet.* **239**, 273-280.

Sabin, E.A., Lee-Ng, C.T., Shuster, J.R. and Barr, P.J. (1989). High-level expression and *in vivo* processing of chimeric ubiquitin fusion proteins in *Saccharomyces cerevisiae*. *Bio/Technology* **7**, 705-709.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Samuelsson, E., Moks, T., Nilsson, B. and Uhlén, M. (1994). Enhanced *in vitro* refolding of insulin-like growth factor I using a solubilizing fusion partner. *Biochemistry* **33**, 4207-4211.

Saris, N. (1998). Conformational repair of heat-denaturated proteins in the endoplasmic reticulum of yeast cell. Thesis, University of Helsinki, Yliopistopaino.

Sassenfeld, H.M. (1990). Engineering proteins for purification. *Trends Biotechnol.* **8**, 88-93.

Sawicki, D.L., Sawicki, S.G., Keränen, S. and Kääriäinen, L. (1981a). Specific Sindbis virus-coded function for minus-strand RNA synthesis. *J. Virol.* **39**, 348-358.

Sawicki, S.G., Sawicki, D.L., Kääriäinen, L. and Keränen, S. (1981b). A Sindbis virus mutant temperature-sensitive in the regulation of minus-strand RNA synthesis. *Virology* **115**, 161-172.

Schekman, R. and Mellman, I. (1997). Does COPI go both ways? *Cell* **90**, 197-200.

Schlesinger, S. and Dubensky,T.W. (1999). Alphavirus vectors for gene expression and vaccines. *Curr. Opin. Biotechnol.* **10**, 434-439.

Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16-20.

Schmidt, T.G.M. and Skerra, A. (1993). The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Eng.* **6**, 109-122.

Schmidt, T.G.M., Koepke, J., Frank, R. and Skerra, A. (1996). Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J. Mol. Biol.* **255**, 753-766.

Schmidt, M., Tuominen, N. Johansson, T., Weiss, S.A., Keinänen, K. and Oker-Blom, C. (1998). Baculovirus-mediated large-scale expression and purification of a polyhistidine-tagged Rubella virus capsid protein. *Protein Expr. Purif.* **12**, 323-330.

Schultz, L.D., Markus, H.Z., Hofmann, K.J., Montgomery, D.L., Dunwiddie, C.T., Kniskern, P.J., Freedman, R.B., Ellis, R.W. and Tuite, M.F. (1994). Using molecular genetics to improve the production of recombinant proteins by the yeast *Saccharomycys cerevisiae*. *Ann. NY Acad. Sci.* **721**, 148-157.

Scouten, W.H. (1991). Affinity chromatography for protein isolation. *Curr. Opin. Biotechnol.* **2**, 37-43.

Sechi, S. and Chait, B.T. (1998). Modification of cysteine residues by alkylation. A tool on pepdide mapping and protein identification. *Anal. Chem.* **70**, 5150-5158.

Sherman, F. (1991). Getting started with yeast. *Methods Enzymol.* **194**, 3-21.

Shirako, Y. and Strauss, J.H. (1994). Regulation of Sindbis virus RNA replication: Uncleaved P123 and nsP4 function in minus- strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus- strand RNA synthesis. *J. Virol.* **68**, 1874-1885.

Shusta, E.V., Bockhaus, J.A., Voegler, A.C. and Wittrup, K.D. (1998). Increasing the secretory capasity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* **16**, 773-777.

Shuster, J.R. (1991). Gene expression in yeast: protein secretion. *Curr. Opin. Biotechnol.* **2**, 685-690.

Simonen, M. (1994). Secretion of heterologous proteins in *Bacillus sublitis* and *Saccharomyces cerevisiae*. Thesis, University of Helsinki, Laser-Paino Oy.

Simonen, M., Jämsä, E. and Makarow, M. (1994). The role of the carrier protein and disulfide formation in the folding of beta-lactamase fusion proteins in the endoplasmic reticulum of yeast. *J. Biol. Chem.* **269**, 13887-13892.

Simons, J.F., Ferro-Novick, S., Rose, M.D. and Helenius, A. (1995). BiP/Kar2p serves as a molecular chaperone during carboxypeptidase Y folding in yeast. *J. Cell Biol.* **130**, 41-49.

Smerdou, C. and Liljeström, P. (1999). Two-helper RNA system for production of recombinant Semliki Forst virus particles. *J. Virol.* **73**, 1092-1098.

Smith, D.B. and Jonhson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusion with glutathione S-transferase. *Gene* **67**, 31-40.

Smith, J.S. and Roth, M.J. (1993). Purification and characterization of an antive human immuno-deficiency virus type 1 Rnase H domain. *J. Virol.* **67**, 4037-4049.

Smith, G.E., Frasier, M.J. and Summers, M.D. (1983a). Molecular engineering of the *Autographa californica* nuclear polyhedrosis- virus genome: Deletion mutations within polyhedrin gene. *J. Virol.* **46**, 584-593.

Smith, G.E., Summers, M.D. and Fraser, M.J. (1983b). Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell Biol.* **3**, 2156-2165.

Smith , P.A., Tripp, B.C., DiBlasio-Smith, A., Lu, Z., LaVallie, E.R. and McCoy, J.M. (1998). A plasmid expression system for quantitative *in vivo* biotinylation of thioredoxin fusion proteins in *Escherichia coli*. *Nucleic Acids Res.* **26**, 1414-1420.

Solski, P.A., Quilliam, A., Coats, S.G., Der, C.J. and Buss, J.J.E. (1996). Targeting proteins to membranes using signal sequence for lipid modification. *Methods Enzymol.* **250**, 435-454.

Srinivasan, U. and Bell, J.A. (1998). A convenient method for affinity purification of maltose binding protein fusions. *J. Biotechnol* **62**, 163-167.

St John, R.J., Carpenter, J.F. and Randolph, T.W. (1999). High pressure fosters protein refolding from aggregates at high concentrations. *Proc. Natl. Acad. Sci. USA* **96**, 13029-13033.

Stempfer, G., Höll-Neugebauer, B. and Rudolph, R. (1996). Improved refolding of an immobilized fusion protein. *Nat. Biotechnol.* **14**, 329-334.

Strauss, J.H. and Strauss, E.G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491-562.

Stålberg, J. (1999). Retention models for ions in chromatography. *J. Crhomatogr. A* **855**, 3-55.

Sudbery, P.E. (1996). The expression of recombinant proteins in yeasts. *Curr. Opin. Biotechnol.* **7**, 517-524.

Summers, M.D. and Smith, G.E. (1988). A manual of methods for baculovirus vectors and insect cell culture procedures. Texas Agricultural Experiment Station, bulletin No. 1555, College Station, Texas, 57 pages.

Suntio, T., Shmelev, A., Lund, M. and Makarow, M. (1999). The sorting determinant guiding Hsp150 to the COPI-independent transport pathway in yeast. *J. Cell Sci.* **112**, 3889-3898.

Sutter, G. and Moss, B. (1992). Nonreplicating vaccinia vector efficiently expresses recombinant genes. *Proc. Natl. Acad. Sci. USA* **89**, 10847-10851.

Sutter, G., Wyatt, L.S., Foley, P.L., Bennink, J.R. and Moss, B. (1994). A recombinant vector derived from the host range-restricted and highly attenuated MVA strain of vaccinia virus stimulates protective immunity in mice to influenza virus. *Vaccine* **12**, 1032-1040.

Takkinen, K., Peränen, J., Keränen, S., Söderlund, H. and Kääriäinen, L. (1990). The Semliki-Forest-virus-specific nonstructural protein nsP4 is an autoproteinase. *Eur. J. Biochem.* **189**, 33-38.

Tanner, W. and Lehle, L. (1987). Protein glycosylation in yeast. *Biochim. Biophys. Acta* **906**, 81-99.

Tarbatt, C.J., Glasgow, G.M., Mooney, D.A., Sheahan, B.J. and Atkins, G.J. (1997). Sequence analysis of the avirulent, demyelinating A7 strain of Semliki Forest virus. *J. Gen. Virol.* **78**, 1551-1557.

Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R.W.H. and Burmeister, W.P. (2000). On the domain structure and the polymerization state of the Sendai virus P protein. *Virology* **266**, 99-109.

Thim, L., Hansen, M.T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G. and Fiil, N.P. (1986). Secretion and processing of insulin precursors in yeast. *Proc. Natl. Acad. Sci. USA* **18**, 6766-6770.

Thomas, J.R., Dwek, R.A. and Rademacher, T.W. (1990). Structure, biosynthesis, and function of glycosylphosphatidylinositols. *Biochemistry* **29**, 5413-5422.

Thomas, T.C. and McNamee, M.G. (1990). Purification of membrane proteins. *Methods Enzymol.* **182**, 499-520.

Tomei, L., Vitale, R.L., Incitti, I., Serafini, S., Altamura, S., Vitelli, A. and De Francesce, R. (2000). Biochemical characterization of hepatitis C virus RNA-dependent RNA polymerase mutant lacking the C-terminal hydrophobic sequence. *J. Gen. Virol.* **81**, 759-767.

Tuittila, M.T., Santagati, M.G., Röyttä, M., Määttä, J.A. and Hinkkanen, A.E. (2000). Replicase complex genes of Semliki Forest virus confer lethal neurovirulence. *J. Virol.* **74**, 4579-4589.

Unternährer, S., Pridmore, D. and Hinnen, A. (1991). A new system for amplifying 2µm plasmid copy number in *Saccharomyces cerevisiae*. *Mol. Mircobiol.* **5**, 1539-1548.

Valax, P. and Georgiou, G. (1993). Molecular characterization of beta-lactamase inclusion bodies produced in *Escherichia coli*. 1. Composition. *Biotechnol. Progr.* **9**, 539-547.

Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G. and Hall, B.D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**, 347-350.

Van den Hazel, H.B., Kielland-Brandt, M.C. and Winther, J.R. (1996). Review: biosynthesis and function of yeast vacuolar proteases. *Yeast* **12**, 1-16.

Van den Steen, P., Rudd, P.M., Dwek, R.A. and Opdenakker, G. (1998). Consepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151-208.

Van Dijk, J.A.P.P. and Smit, J.A.M. (2000). Size-exclusion chromatograhpy-multiangle laser light scattering analysis of ß-lactoglobulin and bovine serum albumin in aqueous solution with added salt. *J. Chromatogr.* **867**, 105-112.

Van Dyk, T.K., Gatenby, A.A. and LaRossa, R.A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* **342**, 451-453.

Varki, A. (1998). Factors controlling the glycosylation potential of the Golgi apparatus. *Trends Cell Biol.* **8**, 34-40.

Vasiljeva, L., Merits, A., Auvinen, P. and Kääriäinen, L. (2000). Identification of a novel function of the alphavirus capping apparatus. RNA 5'-TRIPHOSPHATASE ACTIVITY OF Nsp2. *J. Biol. Chem.* **275**, 17281-17287.

Vlak, J.M., Schouten A., Usmany, M., Belsham, G.J., Klinge-Roode, E.C., Maule, A.J., Van Lent, J.W.M. and Zuidema, D. (1990). Expression of cauliflower mosaic virus gene I using a baculovirus vector based on the p10 gene and a novel selection method. *Virology* **179**, 312-320.

Von Heijne, G. (1985). Signal sequences. The limits of variation. *J. Mol. Biol.* **184**, 99-105.

Voss, S. and Skerra, A. (1997). Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng.* **10**, 975-982.

Wadensten, H., Ekebacke, A., Hammarberg, B., Holmgren, E., Kalderén, C., Tally, M., Moks, T., Uhlén, M., Josephson, S. and Hartmanis, M. (1991). Purification and characterization of human insulin-like growth factor II (IGF-II) expressed as a secreted fusion protein in *Escherichia coli*. *Biotech. Appl. Biochem.* **13**, 412-421.

Walker, P.A., Leong, L.E.-C., Ng, P.W.P, Tan, H.S., Waller, S., Murphy, D. and Porter, A.G. (1994). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Bio/Technology* **6**, 601-605.

Wall, J.G. and Plückthun, A. (1995). Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 507-516.

Walter, P. and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87-119.

Wang, Y.-F., Sawicki, S.G. and Sawicki, D.L. (1994). Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *J. Virol.* **68**, 6466-6475.

Wang, H.L., O'Rear, J. and Stollar, V. (1996). Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. *Virology* **217**, 527-531.

Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. (1996). Optimization of heterologous protein production in *Escherichia coli*. *Curr. Opin. Biotechnol.* **7**, 494-499.

Weiss, E., Chatellier, J. and Orfanoudakis, G. (1994). *In vivo* biotinylated recombinant antibodies: Construction, characterization, and application of a bifunctional Fab-BCCP fusion proteins produced in *Escherichia coli*. *Protein Expr. Purif.* **5**, 509-517.

Welcher, A.A., Bitler, C.M., Radeke, M.J. and Shooter, E.M. (1991). Nerve growth factor binding domain of the nerve growth factor receptor. *Proc. Natl. Acad. Sci. U S A* **88**, 159-163.

Welinder, B.S., Sorensen, H.H. and Hansen, B. (1986). Reversed-phase high-performance liquid chromatography of insulin. Resolution and recovery in relation to column geometry and buffer components. *J. Chromatogr*. **361**, 357-367.

Welinder, B.S., Sorensen, H.H. and Hansen, B. (1987). Reversed-phase high-performance liquid chromatography of human growth hormone. *J. Chromatogr*. **398**, 309-316.

Wilchek, M., Miron, T. and Kohn, J. (1984). Affinity chromatography. *Methods Enzymol.* **104**, 3-55.

Wilkinson, D.L. and Harrison, R.G. (1991). Predicting the solubility of recombinant proteins in *Escherichia coli*. *Bio/Technology* **9**, 443-448.

Wold, F. (1981). *In vivo* chemical modification of proteins (posttranslational modification). *Ann. Rev. Biochem.* **50**, 783-814.

Wong, K.T.K., Peter, C.H., Greenfield, P.F., Reid, S. and Nielsen, L.K. (1995). Low multiplicity infection of insect cells with a recombinant baculovirus: The cell yield concept. *Biotechnol. Bioeng.* **49**, 659-666.

Wyatt, L., Moss, B. and Rozenblatt, S. (1995). Replicationdeficient vaccinia virus encoding bacteriophage T7 RNA polymerase for transient gene expression in mammalian cells. *Virology* **210**, 202-205.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T. Fujimoto, J., Yamamoto, T. and Ishii, S. (1995). Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* **270**, 25328-25331.

Zeng, X. and Ruckenstein, E. (1999). Membrane chromatography: preparation and application to protein separation. *Biotechnol. Prog.* **15**, 1003-1019.

Zhang, Y., Olsen, D.R., Nguyen, K.B., Olsen, P.S., Rhodes, E.T. and Mascarenhas, D. (1998). Expression of eukaryotic proteins in soluble form in *Escherichia coli*. *Protein Expr. Purif.* **12**, 159-165.

Zhao, X. and Siu, C.H. (1995). Colocalization of the homophilic binding site and the neuritogenic activity of the cell adhesion molecule L1 to its seconds Ig-like domain. *J. Biol. Chem.* **270**, 413-429.

Zhou, D.P. Malissard, M., Berger, E.G. and Hennet, T. (2000). Secretion and purification of recombinant beta 1-4 galactotransferase from insect cells using pFme1-protA, a novel transposition - based baculovirus transfer vector. *Arc. Biochem. Biophys.* **374**, 3-7.

Zhou, N.E., Mant, C.T., Kirkland, J.J. and Hodges, R.S. (1991). Comparison of silica-based cyanopropyl and octyl reversed-phase packings for the separation of peptides and proteins. *J. Chromatogr.* **548**, 179-193.