Construction of Random Epitope Insertion cDNA libraries and Consecutive Selection for Folding Competent Proteins in Sz. pombe: a novel approach

Lyngsø, Christina; Mortensen, Uffe Hasbro

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

Construction of Random Epitope Insertion cDNA libraries and Consecutive Selection for Folding Competent Proteins in *Sz. pombe*: a novel approach

Danmarks Tekniske Universitet

PhD Thesis

Christina Lyngsø

ABSTRACT

Pharmexa's main research area focuses on therapeutic vaccines against diseases where a harmful self-protein is involved. In order to make a vaccine against a self-protein, the inherent tolerance of the immune system to self-antigens must be by-passed or broken. This is achieved by immunization with a modified form of the harmful protein.

The modification consists of an insertion of a highly immunogenic foreign Th-cell epitope in the target protein sequence. However, the Th-cell epitope is a 15-20 amino acid long peptide, which can destabilise the protein. Therefore, the modified protein may not be viable or its three-dimensional structure may be altered considerably. The aim of the present project has been to develop a systematic approach to find permissive insertion sites in protein vaccine candidates.

This strategy comprises: construction of a random insertional library of the protein target, representing insertions at all amino acid positions of the target using a transposon, and screening and selection for folded proteins from the insertion library in *Sz. pombe*. For screening, the cDNA library is fused upstream to the invertase cDNA and expressed in *Sz. pombe* as secreted fusion proteins with the invertase moiety. The chimeric proteins capable of folding and hence expression, are then selected based ability to genetically complement invertase activity in a host cell lacking an endogen invertase gene. In this study TNF-α was chosen as model protein.

The presented insertion library, which represents 17,500 distinct epitope insertions was transformed in *Sz. pombe* and screened for invertase activity. Out of 36,000 transformed colonies, 22 were found positive. In these permissive sites for insertion of the epitope were found mostly within the N-terminal region of the protein, as well as C-terminal and the in loops between beta-strands and at the ends of beta-strands. Thus, most insertion sites coincided with sites predictable by rational design and importantly, insertion sites, not predictable as permissive sites by rational design were also identified. Finally, 5 of the 22 TNF-α insertion variants, in which the invertase moiety had been removed by sub-cloning, were chosen for testing for expression in *E. coli*, *Sz. pombe,* and *Drosophila* cells. All tested proteins were expressed, whereas 2 negative controls (bearing insertions, but not displaying invertase activity) were not expressed, validation the approach.

RESUMÉ

Konstruktion af et bibliotek af tilfældigt indsatte Th-celle epitoper og efterfølgende screening og selektion af foldningskompetente proteiner I *Sz. pombe*: en ny tilgang.

Pharmexa A/S er en bioteknologisk virksomhed inden for aktiv immunterapi og vacciner til behandling af alvorlige kroniske sygdomme, hvor skadelige selv-proteiner er involveret. For at lave vacciner mod selv-proteiner må den naturlige tolerance mod disse omgås eller brydes. Dette opnås med immunisering med en ændret form af det skadelige protein. Ændringen består af indsættelsen af en immunogent fremmed T hjælper celle epitop i sekvensen af target proteinet, hvilket synliggør proteinet for immunesystemet. Epitopet består af 15-20 aminosyrer, og dette relativt lange peptid kan potentielt destabilisere proteinet, hvorved det mister foldningskompence eller dets tredimensionelle struktur er ændret betydeligt.

Målet med dette projekt har været, at udvikle en metodisk tilgang til, at finde indsættelsessteder af et epitop tolereret af selv-proteiner. Til dette konstrueres, ved hjælp af et transposon, et bibliotek med tilfældigt indsatte epitoper i target-proteinet. I biblioteket findes derfor varianter repræsenterende epitoper indsat på alle positioner mellem aminosyrerne. Efterfølgende selekteres foldningskompetente proteiner fra biblioteket, ved at cDNAet fra biblioteket fusioners opstrøms til et cDNA, der koder for invertase. Fusionsproteinet er desuden koblet til et signal peptid og kan derfor sekreteres til cellevæggen. Invertase er et enzym, der er essentielt for gærcellers vækst på almindeligt sukker, og biblioteket af chimeriske proteiner udtrykkes i *Sz. pombe* stamme deleteret for invertase genet. Selektionen foregår ved, at kun foldede proteiner vil kunne nå cellevæggen, hvor invertase skal være for at være aktiv pga. en kvalitets kontol, der foregår i sekresionsapperatet. Dvs. at fusionsproteiner med foldningskompetence vil da kunne vælges baseret på deres evne til at komplementere invertase aktivitet, og dermed opretholde vækst.

I dette studie er TNF-α valgt som modelprotein, hvorfra der er blevet lavet et insertionsbiblioteket, repræsenterende 17.500 distinkte epitop indsættelser. Ud af 36.000 transformerede kolonier, screenet blev 22 fundet positive og indsættelsesstedet analyseret. Epitopet hovedsageligt fundet i det N- terminale område af proteinet, dog var der også nogle C-terminalt og i loops mellem sekundære strukturer. Følgelig faldt de fleste positioner for epitopeindsættelser sammen med steder, som er forudsigelige ved hjælp af rationel design. Interessant nok var der et epitopeindsættelsessted, som ikke ville kunne forudsiges. Dette giver en indikation af anvendeligheden af denne empiriske metode. Til slut, valideredes systemet ved at udtrykke 5 af de 22 TNF-α varianter i *E. coli*, *Sz. pombe*, og *Drosophila* celler. Alle de testede proteiner kunne udtrykkes.

TABLE OF CONTEXT

PREFACE

This PhD project has been part of the industrial PhD initiative, jointly founded by Pharmexa A/S and the Danish Ministry of Science. The project has been carried out mainly at the division of Molecular Biology and Expression at Pharmexa and in part at the laboratory DNA repair at BioCentrum, DTU headed by Uffe Mortensen. Florence Dal Degan, research scientist at Pharmexa, and Uffe Mortensen have been supervisors.

Technician Helene Lykkegaard has carried out part of the practical work, Sven Muller has made the plasmid PADRE-pVax plasmid, and Søren Kjærulff the *Sz. pombe* screening host and the *Sz. pombe* vector. The *Sz. pombe* strain Eg660 Δinv1::ura4+ has kindly been provided by O. Nielsen, The University of Copenhagen.

The results presented here have formed the basis of patent applications 16535DK00 and 16535US00; Random Insertion of Peptides. These were filled at the May 31th, 2006. Furthermore, the results will be published in a peer reviewed international journal.

LIST OF ABBREVIATION

Adh	Alcohol Dehydrogenase
APC	Antigen presenting cell
ARS	Autonomous replication sequence
BiP	Immunoglobulin heavy chain binding protein
CAT	Chloramphenicol acetyl-transferase
CFA	Complete Freund's adjuvant
COPII	Cytoplasmic coat protein II
Cpy	Carboxy-peptidase Y
E. coli	*Escherichia coli*
E2	Ubiquitin conjugation enzymes
E3	Ubiquitin ligases
EMM	Edinburgh minimal medium
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
ERGIC-53	ER-Golgi intermediate compartment protein 53
ERp57	Endoplasmic reticulum protein 57
FACS	Fluorescence-activated cell sorting
GFP	Green florescent protein
GPR170	Glucose-regulated protein 170
GT	UDP-Glc:glycoprotein glucosyltranferase
HAC1	Homologous to ATF/CREB 1
HER-2	Human epidermal growth factor receptor-2
hCMV	Human cytomegalovirus
HIV	Human immunodeficiency virus
HLA	Human Leucocyte Antigen
Hsp70	Heat chock factor 70
HU	Histone like protein
IHF	Integration Host Factor
Inv1	Invertase gene in *Sz. pombe*
Ire1	Inositol requiring protein
ITR	Terminal inverted repeat
kDa	Kilo Dalton
LacZ-α	Subunit of the enzyme β-galactosidase
METI	Ministry of Economy, Trade and Industry
NMR	Nuclear magnetic resonance
nmt1	No message in thiamine promoter
OD	Optical density
OST	Oligosaccaryl-transferase complex
PADRE	Pan DR helper T cell epitopes
PDI	Protein disulfide isomerase
RANKL	Receptor activator of nuclear factor (NF)-κB ligand
Rpm	Rounds per minutes
S. cerevisiae	*Saccharomyces cerevisiae*
Sec61	Secretion complementation protein 61
SRP	Signal recognition particle
SV40	Simian virus 40
Sz. pombe	*Schizosaccharomyces pombe*
TACE	TNF-α converting enzyme
TCR	T-cell receptor
Th-cells	T-helper cells
TNF-α	Tumor necrosis factor alpha
TNFR1 and TNFR2	Tumor necrosis factor alpha receptor 1 and 2
TRAM	Translocating chain-association membrane
UPR	Unfolded protein response
YEA	Yeast extract + adenine media

AIM

A key feature of Pharmexa's AutoVac technology depends on the insertion of a peptide epitope into a give protein, to generate protein vaccines. Insertion of peptides into proteins may interfere with protein folding and structure. Presently, permissive insertion sites are predicted based on rational design and tested individually experimentally. The success rate of this approach is low as permissive sites for insertion have proven difficult to predict, and the successful generation of AutoVac vaccine candidates is therefore cumbersome and time consuming.

The aim of the presented project has been to develop a method that in principle allows to find all permissive insertion sites for foreign epitope into a given target protein. The method is based on the genetic screen in a eukaryotic host that selects for folding competent proteins from a library containing all possible AutoVac candidates. This method should not depend on any functional or structural features of the target molecule, and should therefore, be applicable to any protein of interest."

INTRODUCTION

Therapeutic vaccination by Autovac™ molecules

Pharmexa's main area of research is the development of therapeutic protein vaccines against diseases where harmful self-proteins are involved. The rationale behind this vaccine is to induce a strong, but controlled, antibody response against harmful self-proteins to reduce their levels, thereby creating a therapeutic effect (Dalum et al., 1996). Similarly to prophylactic vaccines, preventing infectious disease, therapeutic vaccines against self-proteins are based on a similarity to the putative cause of the disease. However, since these proteins are self-antigens, the immune system does not identify them as pathogenic, so the inherent tolerance of the immune system to self-antigens must be by-passed or broken.

To develop this vaccine, a genetic modification of the harmful protein is made by insertion of a foreign immuno-dominant T-helper cell epitope in its sequence, and the resulting modified antigen is expressed as a recombinant protein. The modified protein is able to evoke an immune response towards the wild-type protein.

Such recombinant proteins represent the proprietary technology of Pharmexa named AutoVac™ The vaccination can be used as preventive therapy to acquire disease resistance (as Receptor activator of nuclear factor (NF)-κB ligand (RANKL) in bone destruction (Juji et al., 2002)), to stop a disease state from developing further (as Tumor necrosis factor alpha (TNF-α)) in rheumatoid arthritis (Dalum et al., 1999)), or it can be used as therapeutic intervention against an already established disease state (as Human epidermal growth factor receptor-2 (HER-2) in breast cancer (Renard et al., 2003)).

Immunological background of the vaccine

The humoral immune response is comprised of B-cells and T-helper (Th) cells collaborating in starting the production of antibodies, in that B cell clonal selection and differentiation into antibody producing plasma cells are controlled in an antigen specific manner by Th-cells (Rajewsky et al., 1969) (Figure 1A). Antigen presenting cells (APC) randomly and continuously endocytose proteins in the bloodstream - self as well as foreign - without any prior selection (Kourilsky et al., 1987). The engulfed proteins are processed into small 15 amino acid fragments by the cells, and displayed on Human Leucocyte Antigen (HLA) class II receptors on the cell surface. The peptide fragments presented are specific parts of the proteins termed epitopes. Here, Th-cells continuously test if the displayed fragment can bind their T-cell receptors (TCR). If there is a match, the fragment is considered foreign, as Th-cells recognising self-peptides have been eliminated by clonal selection in the spleen and lymph

nodes. The Th-cell, matching the fragment presented on APC, are activated and stimulated to proliferate by the APC. The activated Th-cells can now provide help (via cytokines and other signals) for activation and proliferation of B-cells presenting a fragment identical to the one originally recognised by the Th-cell (Rock et al., 1984). Thus, in order to be activated, B-cells must be able to present the Th-epitope on their HLA II receptors. This is done by a small subset of the B-cells that, by using their cell surface bound antibody receptors, recognise, and bind three-dimensional structures on the foreign antigen. After engulfment by he B-cells, the protein is processed inside a proteolytic compartment of the cell (Mouritsen et al., 1992), and will be presented as several small fragments on the cell surface HLA II receptors (Lanzavecchia and Bove, 1985). If the B-cell displays the same epitope recognised by the Th-cells, it receives help and starts to secrete antibodies capable of cross-reacting with the antigen (Liu et al., 1992). These antibodies have the same specificity as the B-cell antibody receptors originally recognising the foreign protein.

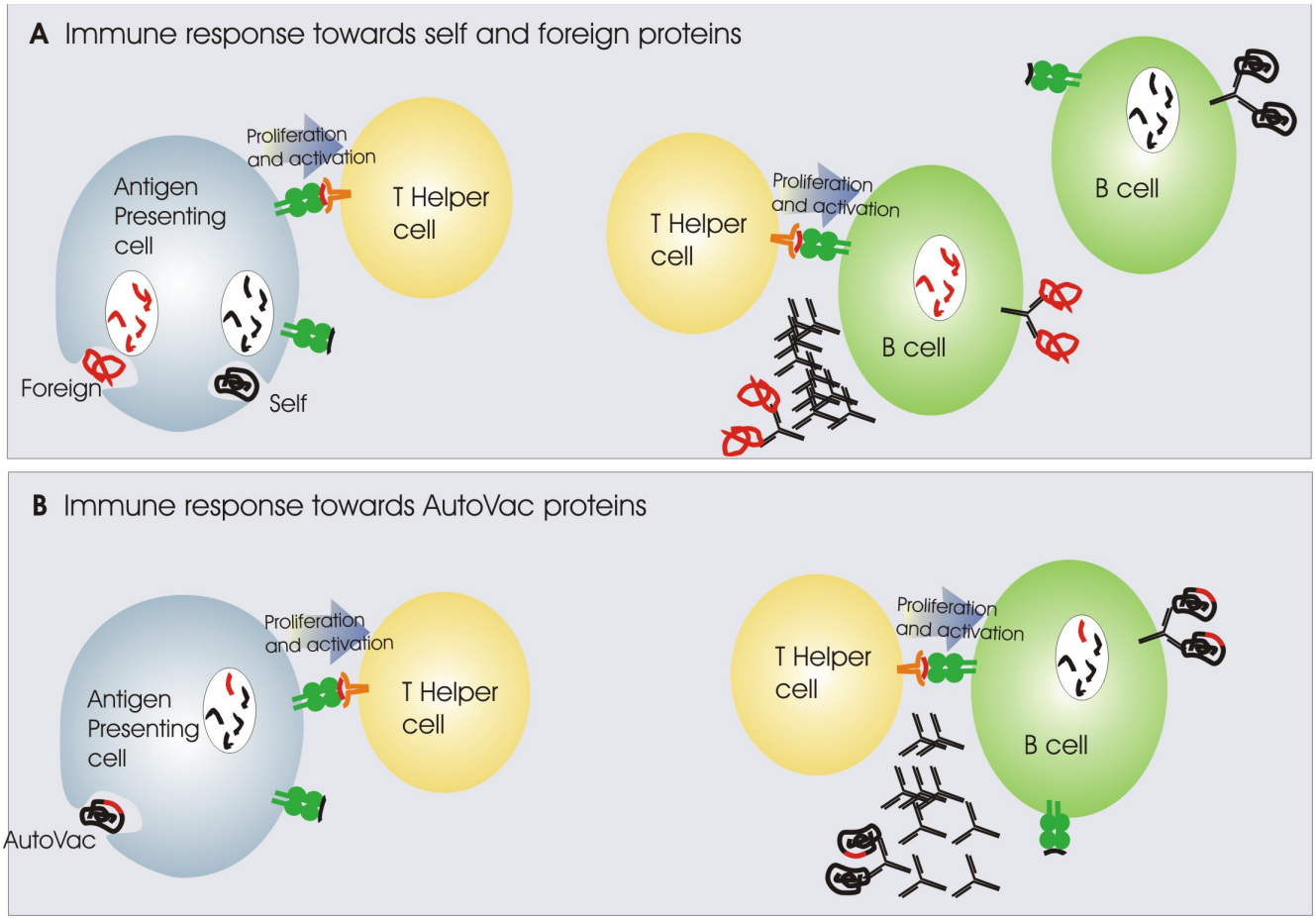


Figure 1. Outline of the Immune response against AutoVac protein

A Th-cell epitope is inserted into the primary sequence of the target. After immunisation the recombinant protein is endocytosed by APC and, after degradation, peptides are presented on the cell surface in complex with HLA class II. Th cells are not normally activated by the auto-peptides, as Th cells recognizing this complex have been down regulated, hence the tolerance against self proteins. Th cells recognizing the T cell epitope-HLA II are, on the contrary, activated and ready to activate B cells presenting the same peptide. Resting B cells recognize and endocytose antigens through membrane bound antibodies. B cell clones both recognizing the Th-cell epitope but also and more importantly, the target protein bind and endocytose vaccine molecule and will interact with epitope activated Th cells. The resting B cells will proliferate and differentiate into antibody producing plasma cells.

The vaccine strategy is based on the fact that the Th-cell epitope and the B-cell epitopes are not generally coinciding (Lanzavecchia, 1985) (Figure 1B). Vaccine molecules are comprised of two entities: a strong foreign Th-cell epitope and a therapeutic relevant protein. A small proportion of B-cells is self-reacting and has antibody receptors specific for the self-antigen corresponding to the vaccine molecule (Dalum et al., 1996; Hartley et al., 1993; Bretscher and Cohn, 1970), although the vast majority of B cells with receptors to self-proteins is removed by clonal deletion to induce tolerance to self and avoid autoimmunity (Hertz and Nemazee, 1998). These self-reacting B cells therefore internalise vaccine antigens, process and present the epitopes bound to HLA class II molecules. Among these is the foreign Th-epitope. Concurrently, APC also process the vaccine protein, and a subpopulation of Th-cells of the immune system respond by identifying the antigen as foreign, because of the inserted epitope. The Th-cells activated by this epitope provide help to the B cells presenting the same Th-epitope on their surface. In this way, B-cells producing antibodies react, not only against the injected modified antigen, but also against the corresponding non-modified pathogenic self-antigens which are activated and thereby proliferate (Dalum et al., 1996).

After the Th-cell-dependent activation, B-cells further increase the diversity and improve the affinity of their antigen receptors by rounds of mutation and selection in the germinal centre, resulting in high-affinity memory B-cells and antibody secreting plasma cells. After boosting with additional recombinant protein, this process is repeated, thereby enhancing the immune response (Han et al., 1996; Hikida et al., 1996).

Importantly, it has been demonstrated that the immunological process driving the cross-reactive antibody response against the self-antigen only occurs in the presence of the modified self-antigens, and frequent re-immunisations are needed to maintain a high antibody titer. When the vaccine injections are discontinued, the antibody concentration decreases, thereby eliminating the risk of inducing a permanent autoimmune response (Dalum et al., 1999).

PADRE – a highly immunogenetic and promiscuous Th-epitope

As explained above, helper T lymphocyte responses play an important role in the induction of both humoral and cellular immune responses. However, a lack of consistent help from T-cells can be a problem for generating general effective Th-cell responses, due to the extensive polymorphism of HLA molecules in the general human population, on which the peptide must bind. Therefore, the quality of the Th-cell epitopes is a crucial component of any vaccine. In Pharmexa's vaccine approach, the Th-cell epitope is genetically added to the target molecule which therefore has the advantage that the most effective and polymorph epitope can be chosen among both natural and synthetic epitopes.

HLA II molecules present epitopes to Th-cells; however, there are several HLA II alleles in the human population, displaying extensive polymorphism in the domains responsible for antigen binding and interactions with the variable regions of the T-cell receptors. This polymorphism is also effecting epitope selection and presentation (Corzo et al., 1995). Initially, the two universal P2 and P30 Th- cell epitopes of tetanus toxin were used, as epitope inserts. They bind several different HLA II molecules, and together they have a binding frequency of approximately 80% in the human population, individually P2 and P30 binds 56 % and 66%, respectively (Panina-Bordignon et al., 1989). The P2 epitope comprises 14 aa and the P30 epitope comprises 21 aa. Therefore, insertion both these epitopes is a rather large intervention in the target protein and one smaller universal epitope would be preferable.

There are many instances, as the P2 and P30, in which a given peptide has been shown to bind to more than one HLA specificity, giving rise to so-called promiscuous epitopes. However, no naturally occurring epitopes are completely universal in binding all HLA molecules. For this reason, Pan DR helper T cell epitopes (PADRE) was constructed to bind most common HLA II molecules with high affinity and act as a strong immunogen (van Bleek and Nathenson, 1990). By sequencing of pools of naturally processed peptides, as well as structural analysis of individual epitopes, there was found an indication of the presence of two or more anchor residues required for binding to the HLA II. The PADRE peptide was engineered by introducing anchor residues for different motifs of the HLA-DR alleles within a poly-alanine backbone, and subsequent testing of the resulting peptides for their binding capacity to HLA II molecules. Because of the small methyl group exposed to for T-cell recognition, these poly-alanine peptides were poor immunogens, but effective blockers of DR- restricted antigen presentation. By introducing bulky and charged residues at positions accessible for T-cell recognition a highly immunogenic epitope was obtained (Alexander et al., 1994).

The PADRE peptide consists of the amino acid sequence AKFVAAWTLKAA and is able to bind with high or intermediate affinity to 15 of the 16 most common HLA types (Alexander et al., 1994). It also binds most different haplotypes in mice, which is important when conducting pre-clinical investigations in this species. Further, when PADRE was evaluated for immunogenicity using human peripheral blood mononuclear cells in a proliferation assay, it was found to be approximately 100-fold more potent on a molar basis than a control tetanus toxoid-derived universal epitope. Because of its binding efficiency and promiscuity, PADRE should overcome the problems posed by the large polymorphism in the human population.

Active vaccination - an alternative to monoclonal antibody therapy

Therapeutic vaccination can be compared to treatment with monoclonal antibodies (mAb).

Like monoclonal antibodies, a vaccine technology is not limited to only one product or only one disease. In principle it can be used to target any self-protein associated with human chronic disease caused by or associated with over-production of certain proteins in the patients.

The protein vaccine is injected subcutaneously or intramuscularly in an appropriate adjuvant formulation. Pharmexa expects that the treatment regimen for humans eventually will involve a small number of initial injections followed by regular booster injections every 3-6 months. Since a vaccine activates the patient's own immune system to produce the antibodies, the dose of protein will be up to 50,000 times less per injection, compared to monoclonal antibody treatment. This results in less strain to the patient, as well as being considerably cheaper and probably a more effective treatment. Also, the principal adverse effect of monoclonal antibodies, namely, an immune reaction against the exogenous drug is not present in the vaccine strategy, since such a response is the basis for and desired effect of treatment. However, the injected protein can perform its natural function and therefore leads to an even more elevated level of the harmful effect; thus, it is of great importance to consider this issue, and if necessary, to inactivate the vaccine molecule through specific mutagenesis.

In vivo data from AutoVac vaccines

The first therapeutic vaccine developed by Pharmexa was an anti-TNF-α molecule. Although there are a number of cytokines in the proinflammatory immune pathways, inhibition of TNF-α has a more universal effect on inflammation than the blocking of other cytokines. TNF-α inhibition improves several inflammatory conditions, and can ameliorate the destruction associated with inflammatory arthritis and the granulomatous diseases as Crohn's disease (Dinarello, 2000).

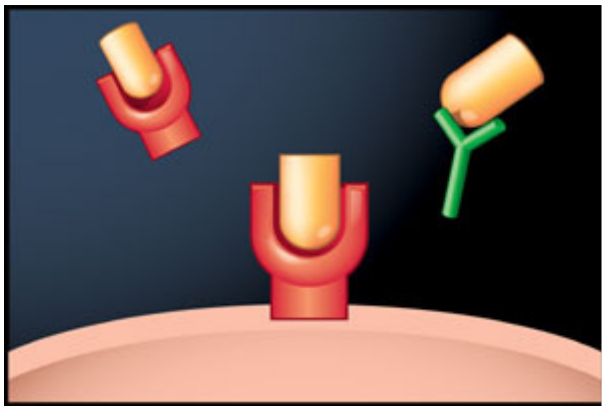


Figure 2. Action of Biological TNF-α drugs. Inflammation occurs when the immune system attacks the patient's own cells. Inhibiting the action of TNF-α can reduce this state. Two types of biological treatment are available; monoclonal antibodies and decoy TNF-α receptors. An anti-TNF-α vaccine will resemble the action of mABs, but will bind at several positions of the target molecule due to the polyclonal nature of the response. (*Illustration from the webpage of Gigt foreningen, www.gigtforenigen.dk*)

There are presently comparable drugs avalible (Figure 2). Two TNF-α mABs have been developed and a TNF-α decoy receptor: Infliximab (Lochs et al., 1999) and Adalimumab (Machold and Smolen, 2003). Infliximab binds to TNF-α and TNF-β and lyses TNF-producing cells to neutralise their activity. Adalimumab is a monoclonal fully human anti-TNF-α antibody binding with high affinity to TNF-α. By replacing murine with human elements, the production of antibodies that neutralise the Adalimumab injections is

reduced. Also, a soluble TNF-α receptor (Etanercept) is in use (Leonardi et al., 2003). It functions as a decoy receptor and thereby inhibits the signal transduction pathway of TNF-α.

The ability of Pharmexas TNF-α vaccine to induce neutralising antibodies was tested in a mouse model of rheumatoid arthritis (Figure 3) (Dalum et al., 1999). The data show that vaccination not only reduces signs and symptoms of arthritis, but also significally reduces the incidence of the disease. Furthermore, an anti-TNF-α immune response could be maintained for an extended period of time, but if immunitions are discontinued, this response declines.

Other AutoVac vaccines have been tested in animal models as RANKL against osteoporosis (Juji et al., 2002) and Her-2 against cancer with successful results (Renard et al., 2003). Her-2 is currently tested in clinical trails in breast cancer patients.

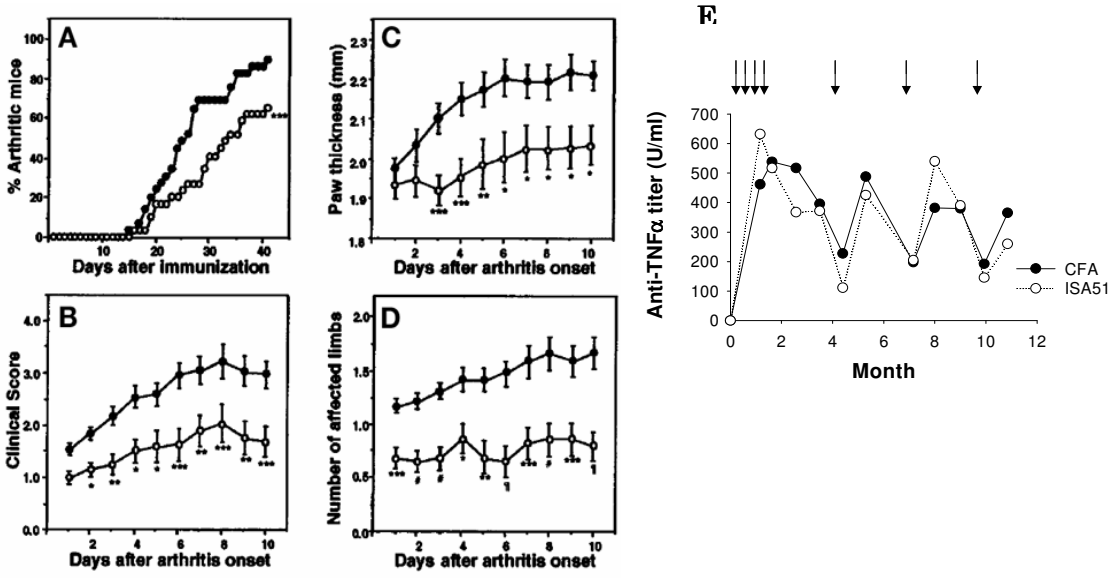


Figure 3. Immunisation with murine TNF-α vaccine molecules.

Groups of 29 mice immunised with murine TNF-α vaccine molecules (o) or placebo (•) and boosted 3 times at 2 weeks intervals. One week after last immunisation mice were immunised intradermally with bovine collagen to induce arthritis. A) Incidence of disease onset, B) clinical score C) swelling of the first affected limb, and D) the number of effected limbs E) level of TNF-α antibodies were measured after immunisation with CFA Complete Freunds adjuvant (•) or ISA51, an adjuvant approved for humans containing a mix of mineral oils and emulsifiers (o). Figures from Dalum *et al.* 1999

A novel approach for generation of vaccine candidate molecules

From an immunological point of view, it is of vital importance to retain the structural conformation of the target protein after introduction of the Th-cell epitope. Therefore, the most relevant requirement of the protein vaccine is that the insertion of the epitope does not disturb the three dimensional structure of the protein, because the immunogenicity and the conformation of the protein, and thereby its conservation of structural B-cell epitopes, are most likely are correlated. Thus, to acquire an effective neutralising antibody response, a high structural identity between vaccine molecule and native target protein is essential (Barlow et al., 1986).

However, the insertion of the epitope is not a trivial task. A 12-30 amino acids peptide has to be accommodated in the protein sequence of the target molecule without disturbing the overall three dimensional structure. This modification often leads to a misfolding of the protein, and many different insertion sites have to be tried before a folding competent molecule is found.

Today at Pharmexa the introduction of the epitopes is based on a search for areas or points in the target protein capable of accommodating the insertion based on rational design and serendipity. In brief, design of a vaccine candidate is largely based on the three-dimensional structure of the target protein when available. By studying the target protein and comparing it to homologous protein structures, suggestions can be made to where insertion of an epitope seems reasonable. Linkers between protein domains or loops connecting secondary structure elements as well as N- and C- terminals are the primary choices of insertion sites. Also, the computational folding prediction algorithms can be applied in the search. However, the calculation of such a large modification requires extensively large algorithms, and many of the variations in sequences and the stability of their conformations are distinguished only by relatively small energy differences. Furthermore, the effect of chaperones and other biological factors decisive for folding is difficult to include. Thus, it is difficult to make an accurate prediction of the structure and folding ability that is a consequence of insertion of the epitope. This leads to the construction of a significant number of variants by site-directed insertional mutagenesis and subsequent screening of these variants on the basis of expression capability. Therefore, a selection system for folded proteins among an extended library of variants with randomly inserted epitopes will be a valuable tool in the quest for finding potential vaccine molecules.

General overview over the presented system

The work presented here enables a search for permissive insertion sites in the target protein by a random approach, in other words; try out all - or many - possibilities. This not only

excludes the many unsuccessful insertions, but also identifies permissive sites that could not be predicted on the basis of current structural or functional knowledge of the protein. When using rational design the loops often are the target for epitope insertion. PADRE, the epitope used in the AutoVac molecules, is very susceptible to proteases (unpublished data from Pharmexa); therefore, the screen will be an advantage since it might enable insertion onto buried regions of the target molecule (Sheridan et al., 2002). Furthermore, it is an empirical system; no functional assay or beforehand knowledge of the target molecule structure or biological function is required. Therefore, the approach can be applied to non-characterised proteins for which no rational design approach is available.

The system for variant generation- and selection system is designed in the following way; (Figure 4); i) a transposon-based mutagenesis procedure is used to generate a unique restriction site in the cDNA at random into which a sequence coding for an epitope is situated, thereby creating a library of putative AutoVac candidates; *ii)* the library is fused to the N-terminal of the invertase protein and the fusion product is linked to the Cpy signal sequence directing the chimera to the secretory pathway; iii) the plasmids encoding the fusion protein are transformed into *Sz. pombe*; iv) the transformed cells are plated on selective media. Transformants expressing folding competent variants are selected, as only cells with invertase activity can utilise sucrose as carbon source and therefore grow. Hereby are foldable sequences are selected among millions of yeast transformants.

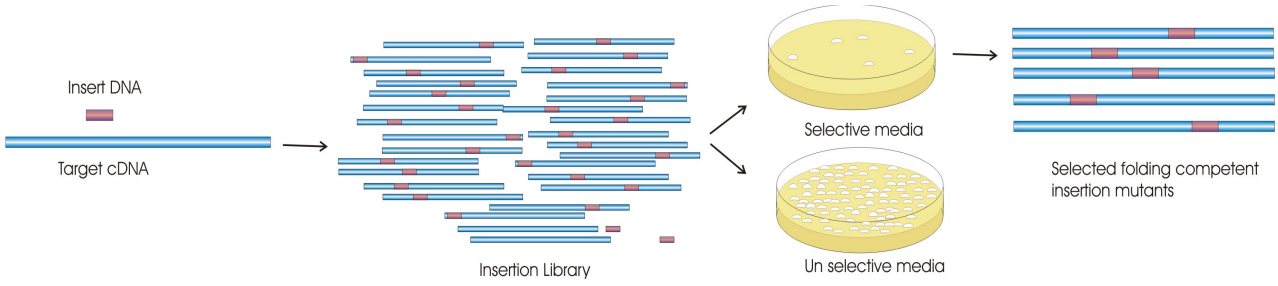


Figure 4. Principle of the complete Selection System for foldable AutoVac proteins
A Th-cell epitope (purple) is introduced into a target cDNA (blue) at random, making an insertional library. The library is subjected to selection in *Sz. pombe* where folding competent variants fused to the invertase marker supports growth on selective media. The sequence of the selected variants can then be analysed for epitope insertion sites.

Considerations before making an exhaustive library

To make an exhaustive library, compromising insertions at all positions in the amino acid sequence, the creation of double strand break at random in the encoding DNA is needed. Traditionally, this has been achieved by DNase1 digestion (Heffron et al., 1978) or chemical induced disruption (Murakami et al., 2002). However, these techniques have the disadvantage of being quite difficult to manage and require titration for each target. Furthermore, they

produce breaking point hot spots, and are therefore not completely random. DNase1 and chemical disruption may introduce arbitrary deletions at the site of the cut.

More recently, *in vitro* transposon has been used as a simple and rapid method to make random modifications in double-stranded DNA (Hayes and Hallet, 2000). In order to function in this context, it is important that the system used readily yield large numbers of independent, random, and simple insertions into any DNA target of choice. Another important attribute is its easy handling of the transposase and the transposon substrate. A comparison of the randomness of DNase1 cleavage, chemical cleavage, and transposon insertion can be viewed in figure 5. Briefly, figure 5A shows the distribution of double strand breaks in the GFP gene. A linker labelled with fluorecein has been attached to the 5´end of the sense DNA strand after the gene has been disrupted by either with Ce (IV)-EDTA treatment or DNAse 1 digestion. To map the distribution of double strand breaks formed by the two different agents the resulting DNA samples were analysed for size distribution on a high resolution PAGE (Murakami *et al,* 2002). As a comparison, shown in figure 5B, the insertion of transposon in the pBC plasmid has been presented. Here, the randomness has been investigated by a PCR based assay (Haapa et al., 1999). As seen, the chemical cleavage is more random than the DNase1 generated where a very high degree of hotspots is seen. However, the transposon-generated insertions are more evenly distributed, and even though there are differences in the intensities of the single bands, insertions are found between all phosphodiester bonds.

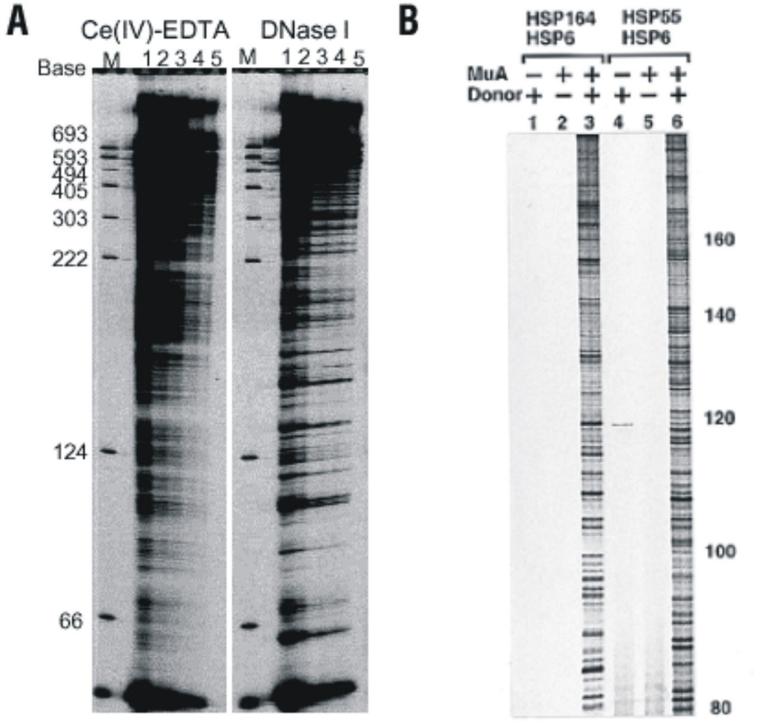


Figure 5. Distribution of double strand breaks: Comparison of the randomness of DNase1 cleavage, Chemical cleavage, and transposon insertion of fluorescence labelled DNA fragments
A. Ce (IV)-EDTA or DNase1 Cleavage. A rather large degree of hotspots is displayed. Figure from Murakami *et al,* 2002.
B. MuA transposon. All positions are represented, but some in a higher degree than others. Figure from Haapa *el at.* 2002.

Current systems for making insertional libraries

A method for generation of an insertion library has been reported recently (Murakami et al., 2002). This method, termed random insertion/deletion mutagenesis (RID), is based on Cerium (IV) as oxidative cleavage agent to generate the double strand breaks in the DNA fragment of interest. The method is a complex procedure requiring 9 reaction steps to complete construction of the library, including a PCR reaction that will generate unwanted errors in the insertional sequence. Ce (IV) is not completely random and multiple cuts in each fragment are difficult to avoid. Furthermore, the result showed that not all the insertions were as anticipated, in that additional bases had been deleted probably during treatment by the Klenow fragment.

The random insertion of a transposon and a subsequent excision of most of the transposon sequence resulting in a characteristic 5 amino acid insertion has been published by several research groups (Hayes and Hallet, 2000). The 5 amino acids are composed of residues left from the transposon and a target sequence duplicated during the transposition reaction. This pentapeptide scanning mutagenesis of cloned target protein has proved a powerful method for dissecting protein activity and function.

A transposon-based approach inserting a specific DNA sequence into a target protein at random has been described (Sheridan et al., 2002). This approach was used for finding insertion sites that can accommodate green florescent protein (GFP) without destroying the biological activity of a given protein. The glutamate receptor subunit, GluR1, and a G protein subunit were used as model proteins for this approach. Some of the proteins retained their signalling function after insertion, and the random nature of the transposition process revealed permissive sites for insertion that one would not have been predicted on the basis of structural or functional models of how that protein works.

Current *in vivo* screens for folding-capability

Different assays for detection of soluble recombinant proteins have been developed in *E. coli*. It is known that misfolding leads to redistribution of proteins to the insoluble fraction in the cell due to aggregation. Therefore, folding of a protein of interest can be detected by creating hybrid proteins comprising the protein of interest fused to the N-terminal of a reportable protein. If the N-terminally protein moiety is soluble, the fused reporter protein domain leads to the reportable phenotype. In contrast, aggregation and inclusion body formation of the N-terminal moiety result in the relocation of the reporter to the insoluble protein fraction in the cells, and the reportable phenotype is not displayed.

The first of these screens was made by expressing proteins fused to chloramphenicol acetyl-

transferase (CAT) fused to HIV integrase (Maxwell et al., 1999). An insoluble protein exhibited a decreased resistance to chloramphenicol, compared to fusions with soluble proteins. Further, a soluble mutant of an insoluble mutant of HIV integrase could be selected from a pool of un-soluble wild-type protein by plating on high levels of chloramphenicol. No actual selection experiments on a random mutagenised gene were preformed. This system contains the useful feature of having an amber codon between the gene of interest and the reporter gene. This gives the possibility to express protein without the CAT moiety, merely by changing expression strain, without additive cloning procedures after the selection scheme.

The basis of another assay is structural complementation of the α- and Ω- fragments of β-galactosidase (Wigley et al., 2001). Fusions of the α-fragment to the C-terminus of the Alzheimer's amyloid β peptide and a non-amyloidogenic mutant, revealed a clear correlation between β-gal activity and the solubility/folding of the target. Additionally, the assay was tested using different proteins widely varying in solubility levels fused to the α-fragment, displaying β-gal activity correlating to their solubility. This assay is solely a screening system - no selection is included.

Also, a screening system based on GFP fluorescence has been described (Waldo et al., 1999). Here GFP is used as indicator of protein folding, when fused to a panel of 20 different proteins. The fluorescence of *E. coli* cells expressing the GFP fusions is shown to be related to the productive folding of the N-terminal protein domains expressed alone. This system was used to evolve proteins, normally prone to aggregation during expression in *E. coli* into closely related proteins that fold robustly and are fully soluble and functional.

Two eucaryotic screening systems for protein folding have recently been developed. Both screening methods are based on the quality-control system of the yeast secretory pathway that is demonstrated to prevent the release of misfolded or incompletely folded proteins (Ellgaard and Helenius, 2003). In the first method the secreted proteins are displayed on *S. cerevisiae* cell surface receptors, labelled with antibodies, and quantified by flow cytometric analysis. This procedure was used for identification of protein with high stability and secretion properties among a library of random single site mutations (Boder and Wittrup, 2000; Boder and Wittrup, 1997). However, this system is based on fluorescence-activated sorting (FACS) and requires the availability of this expensive instrument and also antibodies specific for the target.

In the other method, a cellular screening is combined with a small, generic tag for antibody detection. This system was used to screen a library of random mutants of an unstable variant of bovine pancreatic trypsin inhibitor (BPTI). 3 mutants with improved folding properties were found. However, the screen uncovered many false positives, so subsequent use of a liquid-

culture secretion screen and a gel electrophoresis assay, were needed to eliminate false-positive mutants. Further, in this type of assay, based on antibody detection of positive colonies, there is no selection of positives included in the scheme (Hagihara and Kim, 2002).

Considerations prior to development of the selection system

The main difference between a screen and a selection procedure is that in a screen all the members of a population must be examined. Accordingly, screens are not as effective as selections because fewer individuals can be examined. However, screens can be useful because they allow the detection of a broad dynamic range of the investigated activity, whereas selection provides a positive/negative conclusion. In this study the interest is to select for folded sequences among a library of variants. Therefore, in principle the answer is simply if the molecule can fold or not. Consequently, a selection procedure is the most appropriate to the task.

It has been proven in the different vaccine projects at Pharmexa, as well as in other laboratories around the world, that human proteins can be difficult to express in *E. coli,* with aggregation of improper folded proteins as a consequence (Baneyx and Mujacic, 2004). This can be due to many factors, but especially the difference in transcription, folding (chaperones and other foldases), and secretions system in pro- and eucaryotes may be limiting. The probability of a successful expression of eucaryotic proteins is considered better if a eucaryotic host cell is used. Moreover, *E. coli* lacks the ability of post-translational modifications of the eukaryotic cells that can be important for recognition by the immune system (Makrides, 1996). Therefore, it was an initial perquisite of the presented project that the system developed was in a eucaryotic host. However, the existing eucaryotic systems were found to be inadequate for the task due to the drawback described above.

The screening and selection system is based on the knowledge that proteins folding up correctly in the ER is exported, whereas, insoluble protein variants are sent to the proteasome for degradation. The mechanism of ER dependent degradation of misfolded proteins is not completely elucidated. That these aberrant products, however, are not exported to the surface of the cell is well-established (Helenius et al., 1992). It is a fair assumption that if one domain of a protein, especially the N-terminal domain, is not folded correctly, the protein is not exported and does not reach the cell wall.

Sz. pombe was chosen as host for the screening /selection system since; it resembles higher eucaryotes in the structure and function of genes, proteins and processing of RNA and proteins. The primary components of the selection system are Carboxy-peptidase Y (Cpy) and invertase genes and their products, as this system is based on the redistribution of a hybrid

protein, as are the described *E. coli* selection systems. These proteins have been the primary substrates for the investigation of the secretion pathway in yeast; both *S. Cerevisiae* and *Sz. pombe*, so their behaviour is known in detail (Figure 6). The *Sz. pombe* Cpy signal sequence has been shown to support secretion in *Sz. pombe,* both of invertase (Tabuchi et al., 1997), and TNF-α (Kjaerulff and Jensen, 2005). The invertase protein has the enzymatic activity of hydrolysing sucrose to glucose and fructose, thus only yeast having invertase activity can be grown on sucrose media. Moreover, since protein stability/folding is expected to correlate with secretion of the invertase enzyme (Kowalski et al., 1998), which in turn correlates with cell growth on sucrose media, growth should likewise correlate with folding.

This technique should greatly accelerate the discovery of functional vaccine proteins. This approach does not require functional assays for the protein of interest and provides a simple route to find folding competent proteins. It is possible, in a screen based on folding competence, that positive candidates are folded in a manner not resembling the wild-type protein. However, since the majority of the primary sequence resembles the wild-type target molecule, the innate folding preference most likely also bears a resemblance to this.

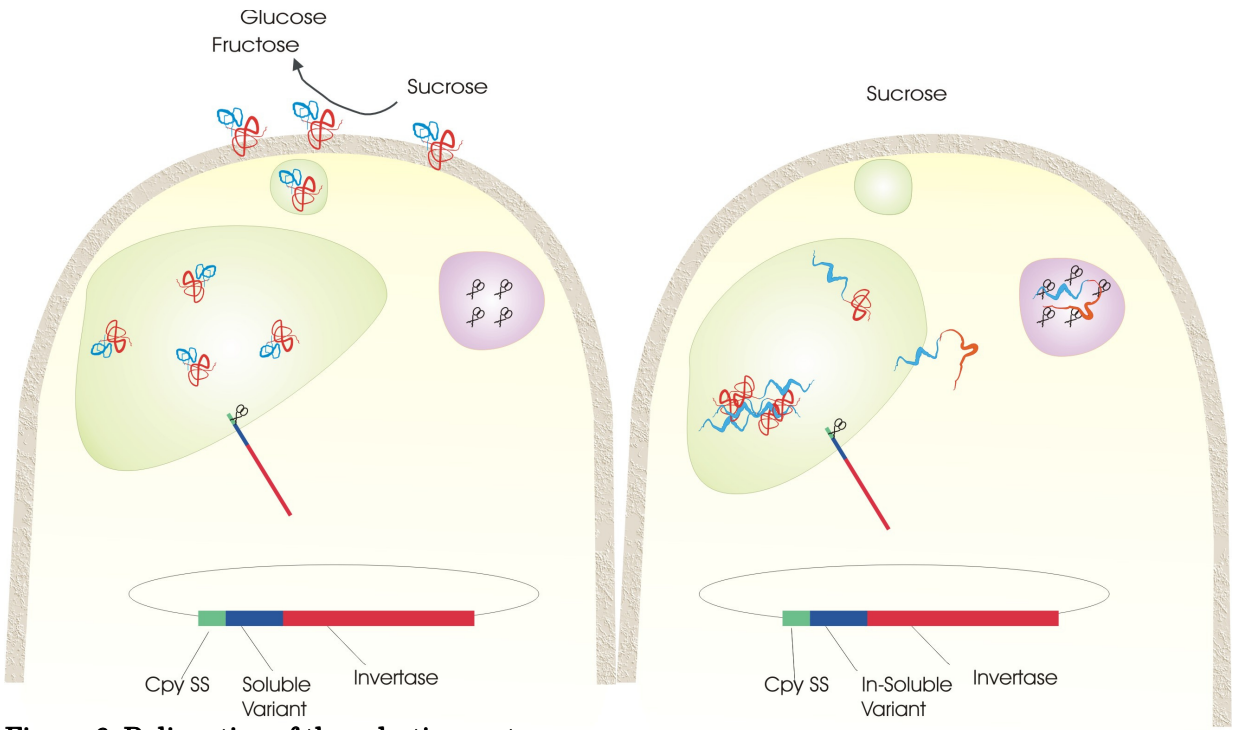


Figure 6. Delineation of the selection system
The gene of interest (Blue) is fused to a selectable genetic marker invertase (Red), and the Cpy signal sequence (Green). Thereby, proteins variants will be directed to the secretory pathway along with the detectable reporter. After translation, the variants go through the secretory pathway passing cell organelles, including endo-plasmic reticulum (Green) and Golgi apparatus. In these organelles the proteins are correctly folded and disulphide linkage and glycosylations are formed. After the posttranslational events, correctly folded proteins are secreted, whereas misfolded proteins are retained and degraded within the cells proteasome (Purple). This means that only colonies containing the genetic sequence for a folding competent protein variant will be able to survive on sucrose plates.

Random insertion of epitopes using genetically modified transposons.

In order to make use of the selection and screening scheme, a library of vaccine variants with randomly introduced epitope needs to be generated. The primary requirement for creating such a library is randomness of the epitope insertion. Thus, a straightforward and efficient technique for generating a random double-strand break in the cDNA sequence is needed. *In vitro* transposition using genetically modified bacteriophage Mu mini-transposons fulfilled this criterion. The rapid and random nature of the transposition process makes it possible to generate many different vaccine variants to be screened in the folding assay.

Biology of transposons

Transposons are discrete mobile DNA segments able to insert into non-homologous target sites. They are ubiquitous and many different transposons have been isolated on DNA level from all known organisms - more than 500 in bacteria alone (Mahillon and Chandler, 1998). They have been regarded as selfish DNA elements living within the genome of cellular organisms, usually conferring no selective advantage to the host and in this manner similar to viruses. This, however, is not a completely adequate description. For example, in bacteria the antibiotic resistance genes and colicin genes confer a synergistically beneficial effect for both host and element. Indeed, the host cell acquires an advantage to survive exposure to hostile environmental factors, and the transposable element achieves an enhanced copy multiplication and the opportunity to spread to other hosts as a consequence of beneficial effect. Also, recent data suggest that the transposons have some valuable effect on the eucaryotic host since transposons may have been adopted into the vertebrate immune system as a way of generating antibody diversity. This notion is implicated by the fact that the V (D)J recombination system operates by similar mechanisms as transposons (Agrawal et al., 1998; Zhou et al., 2004). Another, profound biological effect is shown by transposons being involved in many different processes in genome alteration including element insertion and deletion, homologous recombination between element copies, and horizontal gene transfer. Further, transposons are now known to be a major component of repetitive sequences of all characterized eukaryotic genomes and a major cause of genomic expansion in eucaryotes (Lander et al., 2001) – in mammals they account for nearly half of the genome. These events may not be beneficial to the individual host cell, but is driver of evolution contributing to a wider diversity of a population.

Among the phages of gram-negative bacteria, bacteriophage Mu is the most efficient transposon known. The genome of bacteriophage Mu is a transposon using transposition both

to integrate into the DNA of a new host cell and to replicate before phage induced lysis (Mizuuchi, 1992; Lavoie and Chaconas, 1996). When injected into the host cell, the phage DNA is integrated at a random site of the host genome by the phage-encoded transposition apparatus. Although the initial integration event does not involve replication of Mu, lytic development involves many replicative transposition events exploiting the host replication apparatus to form multiple integrated copies of Mu.

Functional features of DNA and protein elements of transposons

Transposition reactions occur in large catalytic protein-DNA complexes called transposomes. These contain the DNA substrate, oligomers of the transposase, and accessory proteins and involve multiple protein-protein and protein-DNA interactions. The assembly of the transposome is likely to be a key regulator part of the transposition, since it can assure that the appropriate substrates, being the terminal ends of the transposon, and sometimes the target is present as well, and hereby promotes coordination of the reaction steps.

The majority of the transposons is characterized by terminal inverted repeat (ITR) sequences of 10 –40 bp, flanking an internal transposon sequence encoding at least a transposase catalysing the transposition reaction, and frequently additional proteins (Mahillon and Chandler, 1998). In Mu, and most other transposons, two functional domains are found in the ITRs, both of them required for transposition. One region, counting the 3' terminal 5'-CA bp, is involved in the cleavages and strand transfer in the strand transfer reaction. The other region has multiple and asymmetrical sites implicated in binding the transposase, the enzyme catalysing a transfer. In Mu there are three recognition sites, distinct from the cleavage sites, sharing a 22-base pair consensus sequence (Craigie et al., 1984). The sites are referred to as R1, R2, and R3 for the right end and accordingly, L1, L2, and L3 for the left end (Kuo et al., 1991; Craigie et al., 1984). The internal part of the transposons are genetically compact, and usually encode no functions other than those involved in their mobility, though the antibiotic resistance and colicin genes are notable exceptions. The most important gene encodes the transposase, in Mu called MuA. In the DNA, flanking the transposon DNA element, there are always two direct repeats. These sequences are not an actual component of the transposon and are not included in the sequence of the moving element when jumping to a new target. The direct target repeat is always a duplication of a fixed length characteristic for a given element. In Mu the length is 5 bp, other elements have various lengths between 2 and 14 bp (Mahillon and Chandler, 1998).

The most important protein, which in some transposons is the only encoded protein, is the transposase. All transposases that have been investigated have a common significant three-

dimensional structure. The protein is comprised of two functionally distinct domains: N-terminally, a sequence specific DNA binding domain for recognition of the ITRs (Nakayama et al., 1987), and C-terminally, the catalytic core domain, containing the protein's active site with an invariant structural triad D,D35,E-motif (Asp, Asp, 35 aa spacer, Glu), responsible for the DNA cleavage (Rice and Mizuuchi, 1995). Both the recognition sequences and the 5'-CA cleavage sequences are required for transposition (Baker and Mizuuchi, 1992). This domain organisation is a general feature of most transposases and is also shared by the recombinases responsible for integration of the HIV virus (Yang and Steitz, 1995) and immunoglobulin reassembly (Agrawal et al., 1998).

Transposases are able to form multimeric complexes essential to their function (Polard and Chandler, 1995). MuA forms a tetramer (Figure 7), where three subunits tightly bind the R1, R2, and L1 sites, the fourth is weakly associated to the L2 site but, in contrast, have a strong interaction to other subunits in the complex. At least two of the four MuA subunits individually bridge the transposon ends; the subunit bound to the R1 connect to the left 3' AC cleavage site, and likewise the subunit bound to the L1 connect to the right 3' AC cleavage site (Rice and Mizuuchi, 1995; Savilahti and Mizuuchi, 1996; Aldaz et al., 1996). This means that the active site of the MuA subunit bound at R1 promotes reactions *in trans* to the left transposon-end and *vice versa.* This is probably in order to coordinate the reaction at the two ITRs ensuring that DNA transposition always occurs properly, avoiding damage to the host

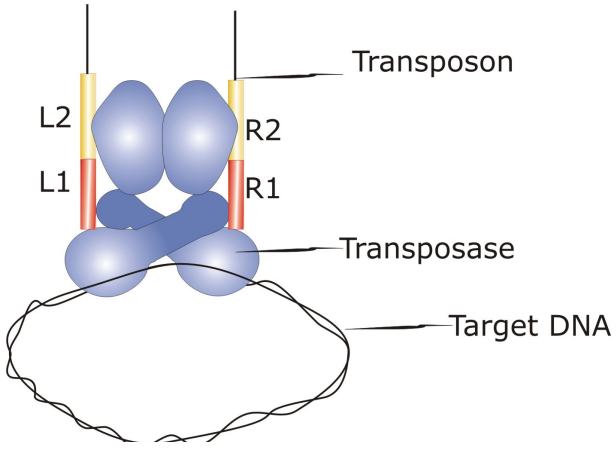


Figure 7. Model of transposome assembly
Individual MuA monomers bind to the recognition sites of the transposon and form a stable tetramer, synapsing the ends and engaging the cleavage site.

genome by partially completed reactions.

The number of other expressed transposon proteins varies considerably between the different types of transposons (Mahillon and Chandler, 1998). In the case of the Mu, the most critical ones are the phage encoded MuB, and the enhancer protein together with the host-encoded

protein HU and IHF. MuB is an ATP-dependent DNA-binding protein regulating the activity of MuA. It is involved in target selection and capture acting scaffold for transposome assembly, thus providing a preferential target for MuA transposase complexed with the ITR ends. MuB also functions as an allosteric activator of MuA, and displays an ATPase activity stimulated by DNA and MuA (Lavoie and Chaconas, 1996).

The enhancer protein participates in the assembly of the MuA tetramer on the DNA and further functions as positive regulator of expression of the phage encoded transposition proteins (Lavoie and Chaconas, 1996). Recruitment of the DNA bending enzymes HU and IHF to a DNA site induces a bending in the super-coiled DNA required for transposase assembly *in vivo*. The HU binding site is the Mu spacer between the L1 and L2 transposase binding sites (Kobryn et al., 1999). These auxiliary factors have been shown dispensable for *in vitro* transposition, where MuA can work separately, still integrating transposons in target DNA (Mizuuchi, 1992; Lavoie and Chaconas, 1996).

DNA transfer: The mechanism of reaction.

The DNA transposition follows a common mechanism with some variations depending on the element (Mizuuchi, 1992; Mahillon and Chandler, 1998). The transposase is expressed and identifies the ITR within the donor DNA, by highly sequence specific protein-DNA interactions. After binding and multimerisation, the transposase catalyse a blunt double-stranded break flanking the transposon. This is done by a nucleophile attack at the 5' phosphate group in the DNA by a water molecule, to expose a free 3' OH-group (Figure 8). The gapped donor DNA resulting from transposition is repaired by homologous recombination by the host cell repair apparatus (Hagemann and Craig, 1993). The transposase and transposon ends are now assembled into an enzyme complex, with an active 3' OH-group at each end of the transposon. One strand of the transposon DNA attacks, again by nucleophile attack, the target through trans-esterification (Goldhaber-Gordon et al., 2003; Mizuuchi and Adzuma, 1991). The target double-strand is now cleaved open by a staggered cut, where each end of the transposon DNA is joined to an overhanging end of the recipient DNA. The DNA polymerase completes insertion by repairing the short, overhanging sequences, resulting in a characteristic short direct repeat on each side of the transposon insertion in the recipient DNA (Bainton et al., 1991). Although the two reactions performed by the transposase are distinct, they are chemically related, both being one-step substitutions of the oxygen groups on a DNA phosphate (Engelman et al., 1991; Mizuuchi and Adzuma, 1991). The major differences between the two reactions are the identity of the nucleophile and the identity of the phosphate at the reaction centre.

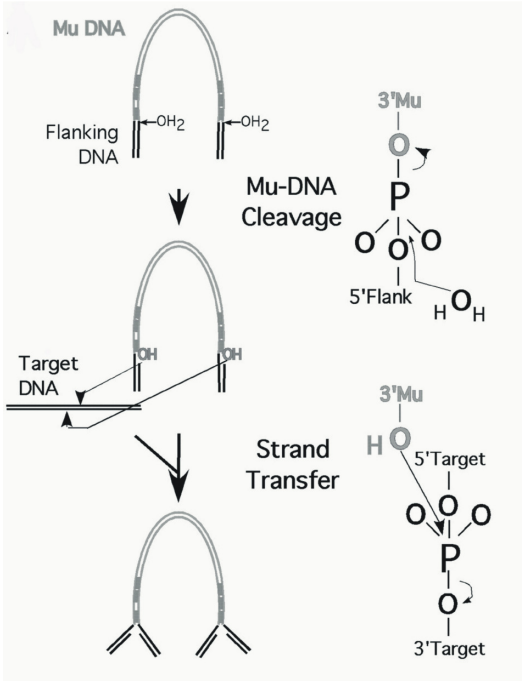


Figure 8. Schematic representation of the transposition mechanism.

Transposomes mediate two sequential chemical reactions, when transferring the transposon DNA from one site to another: (i) DNA cleavage, in which a nick is introduced precisely at the end of the transposon, on the 3' strand and (ii) DNA strand transfer, in which the nicked strand is joined to a separate target molecule. Figure from Mizuuchi *et al,* 1991.

The transfer process results in loss of the transposon from the donor DNA and insertion of the transposon into a recipient DNA. So, the transposon is not duplicated in the process. But, because transposition often occurs after replication of the donor DNA molecule, it is not possible to determine if a transposon moves by a non-replicative mechanism by merely looking for loss of the transposon at the original site.

Unique to the lytic part of the replication cycle of bacteriophage Mu and the Tn3 family of transposons, the DNA element remains attached to the donor DNA while inserting into the new target. In these cases a 3' nicking reaction occurs at each transposon ends (rather than double strand breaks), thus the 5' end of the element remains attached to the donor DNA. The exposed 3' end then joins the target DNA. This transposition reaction is followed by DNA replication. This generates two copies of the element in a structure called a cointegrate in which the donor and target backbone are connected by directly repeated copies of the transposon at each junction (Mizuuchi, 1992). This structure is subsequently resolved by recombination between the two copies of the element.

Target site selectivity and Target site immunity

Some transposons exhibit essentially random patterns of insertion, whereas others have considerable target site selectivity. The selectivity can result both from the quality of the DNA such as silenced DNA, or promoter regions, or can be a preferred consensus sequence of varying stringency. Mu has low target site specificity and shows a very random distribution and can transpose to essentially any DNA sequence, although it exhibits a relative preference for the degenerate sequence Y (G/C)R (Mizuuchi and Mizuuchi, 1993). It has been reported that Mu displays a dramatic preference for insertion into mismatched DNA sites (Yanagihara and Mizuuchi, 2002)

The bacterial transposons Mu, Tn3 and Tn7 are known to exhibit target immunity, a process that prevents the element from transposing into target DNAs already containing a copy of the

transposon (Robinson et al., 1977). This immunity can have two biological functions. One is to ensure a spreading of the transposons to as many different DNA units as possible, promoting the distribution and survival of the transposon. Another is to prevent the destruction of one transposon by the integration of another. The presence of a transposon end in a target DNA reduces the frequency of a second transposon event with 100-1000 fold *in vivo,* and it has been shown that the Tn7 element can inhibit insertion of a second element over a distance of 190 kb to (DeBoy and Craig, 1996). Transposition into other DNA molecules is not inhibited, so target immunity is a *cis*-acting effect, not a global *trans*-inhibition. Analysis using plasmid targets shows that the signal for the inhibition lies within the ITR sequences (Adzuma and Mizuuchi, 1989; Darzins et al., 1988; Hauer and Shapiro, 1984; Lee et al., 1983) and is transmitted by the MuB protein via the MuA. Interaction of MuB with MuA bound to the recognition sequences in the ITRs of the transposon provokes ATP hydrolysis by MuB following a release of MuB from the target DNA. This in turn reduces the ability for the DNA to function as a target, and the MuB is redistributed to those DNA molecules not containing a transposon element (Adzuma and Mizuuchi, 1991).

Insertion mutagenesis with a heterologous transposon

The development of *in vitro* transposition systems has been based on experiments showing that it is possible to construct artificial mini-transposons with the only requirement being that of containing some of the transposase binding sites of the ITRs and the catalytic 3' terminal base pairs (Savilahti et al., 1995). Furthermore, some systems exploit the fact that transposition *in vitro* is possible using only the transposase and in absence of other protein factors (Hayes and Hallet, 2000). In principle, any DNA fragment flanked by the trimmed terminal inverted repeats can be used as a mini-transposon and the core sequence of the transposon is totally open to genetic manipulation.

As described above most transposons function similarly, and a number of analogous systems has been developed based on different transposons as tools for biotechnological applications in insertional and truncation mutagenesis. These include Ty1 from yeast, MoMLV retroviral integrase (Singh et al., 1997), the mariner element *Himar1* (Lampe et al., 1999; Akerley et al., 1998) from the horn fly and Tn*7* (Biery et al., 2000a; Biery et al., 2000b) and Tn*5* (Goryshin and Reznikoff, 1998), Tn552 (Griffin et al., 1999), and Mu (Haapa et al., 1999) of bacterial origin. Several are commercially available; however, if necessary, the components are readily produced by standard laboratory techniques. Tn7 demonstrates high transposition efficiency, and target immunity. However, Tn*7* has long, complex ends impeding the construction of alternative transposon derivatives. Additionally, transposition requires three proteins and a two-step reaction. Normally, Tn*7* shows a very high degree of target specificity; however, the

version used for mutagenesis or sequencing exploits a mutant of the TnsC protein (a factor analogue to MuB) with a relatively little target specificity. The Ty1 insertion, although essentially random, shows quite low efficiency and needs a relatively large quantity of both transposon and target DNA. In the Tn*552, Tn5, and* Mu systems, all employing a linear modified version of the transposon, efficient strand transfer has been obtained with purified transposase, facilitating the generation of very large numbers of independent insertions in a single reaction. Furthermore, the transposon ends have a sufficient degree of plasticity to enable genetic engineering and artificially introduce restriction enzyme recognition sequences close to both ends in ITRs. This allows most of the transposon DNA sequence to be removed, subsequently leaving only a small sequence of the transposon, plus the duplication of the target sequence.

Since the Mu based systems appear equally simple, efficient, and random, the different artificial restriction enzyme recognition sites introduced at the ITR ends have been the decisive factor in the choice. In the Mu system, the *Not1* restriction enzyme fitted the PADRE insertion sequence so well that only two additional amino acids were present after insertion. Furthermore, the target selectivity for Mu *in vitro* transposons is the best studied, and has proven to be effectively random (Haapa et al., 1999). The method exploits the properties of a mini-Mu transposon and uses only the MuA transposase. The accessory protein and DNA cofactors from the host cell and from the original phage transposon involved in and controlling the Mu transposition are not included (Mizuuchi, 1992). The *in vitro* Mu reaction differs from the authentic *in vivo* reaction, as the two-step process of cleavage and strand transfer is reduced to only the latter step. This is possible by using pre-cut donor DNA, thereby miming a `cut and paste' type of transposition, and eliminating the need for the DNA bending factors and enhancer protein. The target preferences observed in vivo are most likely to be a result from the MuB protein, host factors, or features of gene expression. These factors are eliminated *in vitro*, resulting in a more uniform distribution of insertions (Haapa-Paananen et al., 2002). However, since target immunity is probably facilitated by the MuB protein, this cannot be expected in the *in vitro* reaction.

Protein folding, quality control and secretion

The screening assay established here is based on a sorting between folded and non-folded proteins in the secretion pathway and the resulting link between protein folding and protein secretion in eucaryotic cells. The current knowledge that forms the basis for this rationale is presented below.

The intracellular compartments involved in secretion

Many proteins, such as hormones and components of the extracellular matrix, must be directed to the cell surface and secreted. The secretion system is made up of separate organelles: The endoplasmic reticulum (ER), the Golgi complex, and the plasma membrane. Additionally, tubulovesicular transport elements mediate intracellular membrane transport between these organelles. Membrane interchange within this system flows along highly organized directional routes, balanced by retrieval pathways bringing membrane and selected proteins back to the compartment of origin (Cole and Lippincott-Schwartz, 1995) (Figure 9).

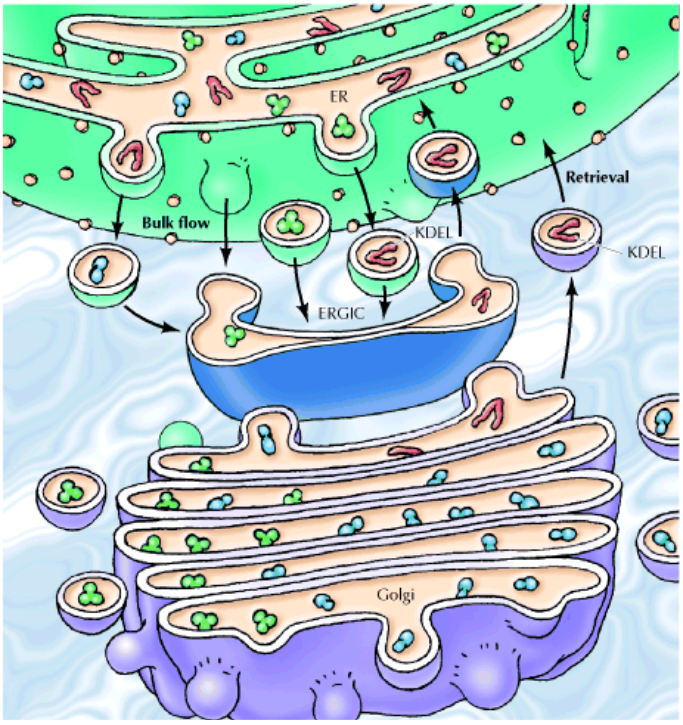


Figure 9. The intracellular compartments involved in secretion
Proteins from the ER are transported to the *cis* Golgi network by ER derived vesicles that fuse and form the ER-Golgi intermediate compartment (ERGIC). ER resident proteins that contain the KDEL retrieval signal are returned from the ERGIC and the *cis* Golgi network via the retrograde pathway. The *medial* and *trans* compartments of the Golgi stack are the sites of most protein modifications. Next, proteins are carried to the *trans* Golgi network, and sorted for transport to the plasma membrane, secretion, or lysosomes. *Figure from Cooper,G.M 2000. The Cell: A Molecular Approach.*

The ER is an elaborate network, extending from the nuclear membrane throughout the cytoplasm, of inter-connected membrane-enclosed tubules and layered flattened sacs (cisternae). It is the largest organelle of most eucaryotic cells representing about 10% of the total cell volume. There are two different types of ERs. One is the smooth ER that is not associated with ribosomes and is involved in metabolism of organelles, as lipid donation to other organelles, Ca+ homeostasis, and biogenesis of organelles take place here. Most notably in this context, is the rough ER covered by ribosomes on its outer surface. It is starting point for secretion, and functions in protein processing of secreted proteins. The protein concentration here is very high, more than 100 mg/ml (Marquardt et al., 1993). In this gel like protein matrix, chaperones and folding enzymes are very abundant, greatly

outnumbering the newly synthesised substrates. The ER is highly specialised for folding and secondary modifications of proteins. To carry out this function, the compartment has a unique high oxidising potential, allowing disulphide bond formation (Tu et al., 2000), a glycosylation apparatus, and a specialised subset of chaperones and other protein-modifying enzymes.

The secretory proteins continue their journey towards the exterior, in small protein coated transport vesicles budding off from specialised exit sites on the ER, which are smooth ribosome-free sections (Palade, 1975). The vesicles fuse to form vesicular-tubular clusters and from here secretory proteins move on to the *cis*-Golgi reticulum and move by cisternal migration to the *trans* face of the Golgi, and next into the *trans*-Golgi reticulum, another complex network of vesicles. Here they are sorted and packaged into post-Golgi secretory vesicles moving through the cytoplasm to fuse with the cell surface while unloading their cargo (Hauri and Schweizer, 1992; Saraste and Kuismanen, 1992).

Signal sequences directs membrane translocation

The synthesis of all proteins starts in the cytosol; however, proteins destined for secretion are repositioned into the ER as an early event of translation. This targeting of nascent polypeptide chains are initiated by a signal peptide sequence (Keenan et al., 2001), directing secretory proteins to the membrane of the ER. There is not one universal signal, but generally, the signals are N-terminal extensions with a short N-terminal region, a hydrophobic core region, and a polar C-terminal. However, the signals can also be located within a protein, or at its C-terminal end (Martoglio and Dobberstein, 1998). When comparing a large number of signal sequences a large variability is found in their relative size, ranging from 15-50 residues. The hydrophobic core consists of six to fifteen amino acids, and mutation analysis reveals that this is the most essential part required for ER-targeting and membrane translocation (von Heijne, 1985). Flanking the hydrophobic region C-terminally are polar residues often containing a helix-breaking proline, glycine residues, and small-uncharged residues in position –1 and –3. The latter determines the cleavage site of a peptidase, fitting into a shallow depression in the signal peptidase (von Heijne, 1990). Several polar residues are often found at the N-terminal side of the hydrophobic region giving net positive charge. This region is the main contributor to the variation in the overall length of the signal sequence. Notably, different signals can be interchanged between proteins, even from different organisms and, nevertheless, follow the original targeting pathway (Martoglio and Dobberstein, 1998). This feature is widely used for recombinant expression of heterologous proteins, where a few common signal sequences are used to target any gene of interest for secretion.

In the most common pathway in eukaryotes, co-translational translocation, a signal

recognition particle (SRP) identifies the hydrophobic segment of the signal peptide, just after emerging from the ribosome, and it binds the ribosome as well as the signal sequence (Matlack et al., 1998). Then the SRP arrests polypeptide chain elongation on the ribosome and the complex binds to the ER-bound SRP- receptor, attaching the ribosomes to the ER. SRP is released and the protein synthesis restarts into the protein translocation channel. This is a pore formed by the heterodimeric TRAM/Sec61 complex, through which the precursor protein is entering the ER lumen (Johnson and van Waes, 1999). After entry into the lumen the signal sequence is cleaved off the precursor protein by a membrane bound signal peptidase. Due to co-translational, the transport growing chain is never exposed to the cytosol and does not fold until it reaches the ER lumen. Thus, the protein crosses the membrane in an unfolded state and the co-translational translocation of growing polypeptide chains is driven directly by the process of protein synthesis.

Protein folding in the ER

Each newly synthesised chain of amino acids has to fold into a unique active conformation following its synthesis. For proteins crossing the ER membrane while translation is still in progress, the emerging protein immediately starts to fold when entering the lumen. The concurrent event of translocation across the ER membrane and folding permits a sequential folding, potentially improving the folding efficacy, particularly in multi-domain proteins (Netzer and Hartl, 1997b). As soon as enough amino acids extend beyond the ribosome, they will fold to the lowest available energy for that part of the chain. This will presumably involve the correct secondary structure, and in all probability an entire domain formation for larger multi-domain proteins, thereby minimising the error rate for complex proteins (Netzer and Hartl, 1997a). The correct tertiary sequence is acquired when the whole protein has been translocated. This is done by gradually combining and recombining secondary structures and single domains (Jaenicke, 1991). In this way protein folding is simplified by breaking it down into sequential steps. The N-terminal domains often seeds the folding of the subsequent parts of the proteins, hence directing and hastening the folding process (Netzer and Hartl, 1997a; Baldwin, 1989).

Even though the native conformation of a protein lies encoded in its primary structure, the ER enhances protein-folding efficiency immensely. As in the ER lumen, a complex chaperone system has evolved to assist the folding process, by facilitating rate limiting steps, preventing aggregation by shielding the folding protein from unwanted intra-and inter molecular interactions, and preventing premature transport of unfolded proteins along the secretory pathway (Feldman and Frydman, 2000). The best studied are the Hsp70 chaperones BiP (Kar2p in yeast) (Gething and Sambrook, 1992) and Glucose-regulated protein 170 (GPR170)

(Spee et al., 1999), which recognises hydrophobic patches on proteins, and the lectin-like chaperones calnexin and calreticulin, which bind monoglucosylated N-glycans (Ellgaard and Frickel, 2003). In addition, other specific folding enzymes lower the activation energy required for processes as disulphide bond formation, by interaction with substrate proteins; this can be protein disulfide isomerase (PDI) and oxidoreductase ERp57 (Ruddock et al., 2000). Also, the oligosaccharyl transferase complex (OST) can carry out *N*-glycosylation.

ER Quality control mechanisms

The synthesis of defective proteins is a common phenomenon – it is estimated that as many as 30% of newly synthesised proteins contain translational errors (Schubert et al., 2000). Therefore, a quality control system has evolved in the ER to function as a checkpoint for discriminating between properly folded proteins and terminally misfolded species, as well as unassembled protein subunits (Figure 10). At first, the misfolded polypeptides and orphan subunits are retained in the ER to undergo new attempts for folding. If this fails, they are subsequently subjected to elimination though ER-associated degradation (ERAD), a protective mechanism of vast importance for cell survival (Ellgaard and Helenius, 2003). The quality control seems to be based on folding ability alone and not on functionality. How the cell makes the conformational sorting is unclear, but most, if not all, chaperones and folding enzymes seem to contribute (Anken et al., 2005). Also, distinct systems seem to be employed for glycoproteins and non-glycoproteins.

When chaperones are able to bind exposed hydrophobic stretches of a polypeptide, it means that the protein is not fully folded or assembled to a matching subunit. This retains the protein in the ER, even though the chaperone and immature protein go through cycles of binding and release. Analogues, unpaired cysteines cause PDI mediated retention (Anelli et al., 2003; Sitia et al., 1990). It is unclear how BiP and PDI decide whether an unfolded protein is a folding intermediate, or a terminal folding failure ripened for proteasome degradation.

The majority of the secreted proteins is glycosylated. For glycoproteins in mammalian cells, the quality control system engages the action of calnexin and calreticulin (Ellgaard and Frickel, 2003), UDP Glc: glycoprotein glucosyltranferase (GT), and glucosidase II enzymes (Parodi, 2000). Here, GT serves as a folding sensor, since it glycosylates only misfolded or partly folded glycoproteins. Hydrophobic residues exposed near glycosylation sites in incompletely folded conformers mediate GT recognition. This structural feature is common to all non-native conformers (Caramelo et al., 2004). Also, an enhanced dynamic mobility of folding intermediates may play a part of the selectivity of GT (Ritter and Helenius, 2000). Proteins re-glycosylated by this factor have a prolonged association with calnexin and

calreticulin (Ritter et al., 2005). In yeast, there is no GT equivalent, but increasing evidence proposes that the Kar2p chaperone could play a similar role. Kar2p mutants are defective in ERAD and exhibit aggregation of misfolded proteins (Plemper et al., 1997).

ERAD activity is partly regulated by the unfolded protein response (UPR) (Casagrande et al., 2000). The UPR is a signal transduction pathway connecting the ER and nucleus to monitor and respond to the changing needs of the early secretory pathway. In *Saccharomyces cerevisiae*, Ire1p, a transmembrane kinase localized to the ER/nuclear envelope, senses the concentration of unfolded proteins in the ER by a binding competition to Kar2p between the Ire1p lumenal domain and unfolded proteins. As a consequence of increasing amounts of unfolded proteins there is an exhaustion of unbound Kar2p molecules by sequestration, leading to the dimerisation of Ire1p. The dimerisation induces a conformational change transmitting a signal across the membrane, by activating a cytoplasmic kinase activity, and causing a splicing of the *HAC1* mRNA, thereby allowing synthesis of the Hac1p transcription factor, which in turn, upregulates genes containing a UPR response element. This signal cascade leads to an increase in the level of proteins required for folding and quality control (Travers et al., 2000). The UPR and ER degradation are thereby closely connected. Overexpression or accumulation of unfolded proteins due to the absence of a component of the ER degradation machinery induces the UPR which, consecutively, upregulates the total level of components of the ER degradation machinery (Meusser et al., 2005).

Another fate for the misfolded protein can be aggregation, which may then be segregated into sub-regions of the ER. The size of the aggregates varies from small oligomers to complexes of several million daltons. Because aggregates are more stable than the intermediate conformation from which they are derived, the ultimate fate of the unfolded proteins is determined by the kinetic partitioning between proteasomal degradation and aggregation (Marquardt and Helenius, 1992; Sitia and Braakman, 2003).

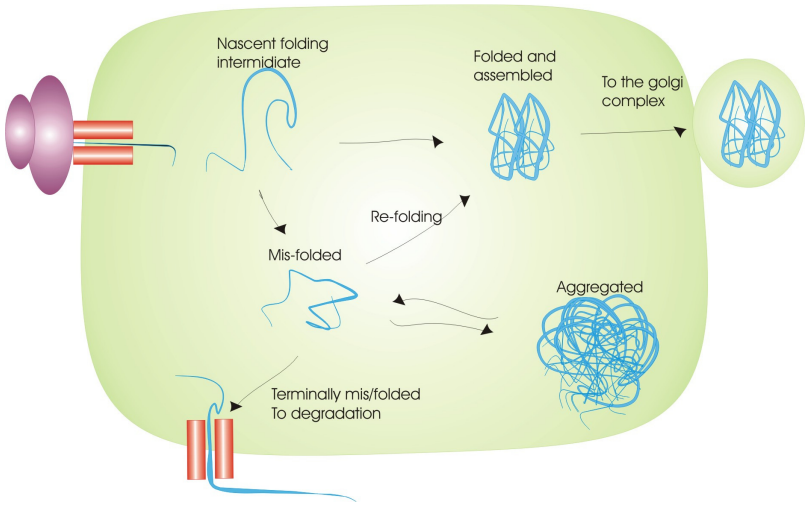


Figure 10. Possible fates of newly translocated proteins in the ER. A nascent protein is transported across the ER membrane trough the protein translocation channel (Red). The unfolded protein can either fold or be mis-folded. Folded proteins leave the ER through vesicle transport. Mis-folded proteins are tried refolded by chaperones and foldases. Terminally mis-folded proteins are either aggregated or degraded.

ER-associated degradation

Misfolded proteins are not degraded in the ER but require retrotranslocation across the ER membrane, probably via the Sec61 translocon (Matlack et al., 1998), into the cytoplasm and subsequent degradation by the 26S proteasome (Ellgaard and Helenius, 2003; Zapun et al., 1999; Hirsch et al., 2004; Sommer and Wolf, 1997). The retrotranslocation through the Sec61 requires unfolding of the misfolded protein. Consistent with this, it has been shown that unfolding and reduction of disulphide bonds are a perquisite for dislocation to the proteosome, and there is evidence that both BiP and PDI also are involved in this step (Tsai et al., 2001; Winkeler et al., 2003)

Following retrotranslocation from the ER, nearly all misfolded proteins are polyubiquitniated, this serves as signal for proteasomal degradation of cytosolic proteins, as well as proteins in the secretion pathway (Thrower et al., 2000). The ubiquitin conjugation enzymes (E2) and ubiquitin ligases (E3) mediate ubiquitination, where E3 is the key determinant of specificity in ubiquitination pathways, as it recognises the target proteins (Hershko and Ciechanover, 1998). These enzymes are found both in the cytosol as well as bound to the ER. However, the enzymatic domains of the ER associated E2 and E3 face the cytosol. Thus accordingly, only proteins completely or partially retro-translocated can be ubiquitinated. Therefore, the most important signal for degradation must involve lumenal factors, before the ubiquitination, in, or prior to the retrotranslocation process.

Protein Exit from the ER

After release from the chaperones, the folded and modified polypeptide chains move from the ER by vesicle-mediated transport to the Golgi complex. The transport vesicles are coated with the cytoplasmic coat protein II (COPII) made up of the subunits Sar1p, Sec23/Sec24p, and Sec13/Sec31p (Barlowe et al., 1994; Salama et al., 1993). Little is known about how fully folded and assembled proteins are recognized and packaged into these vesicles. Presumably, both receptor-mediated export, as well as diffusion into the transport vesicles (bulk-flow) combined with ER-protein retention or Golgi-to-ER retrieval, are in play (Belden and Barlowe, 2001). The specific protein recruitment is thought to occur as a result of direct interactions between sorting signals present on cytosolic domains of cargo molecules and specific receptors. The Sec24 of the coat machinery is considered to be the receptor for membrane proteins (Miller et al., 2003; Mossessova et al., 2003) and the cycling lectin ERGIC-53 has been suggested to act as the lumenal sorting receptor for glycoproteins (Appenzeller et al., 1999). However, these receptors seem dispensable for the secretion of soluble proteins, like invertase (Springer et al., 2000), and the identification of an export signal on these proteins remains elusive. But the

Erv29 has recently been suggested to function as a transmembrane receptor for the export of certain soluble proteins (Otte and Barlowe, 2004; Belden and Barlowe, 2001).

Evidence for the bulk-flow model is apparent, since cytosolic proteins are secreted when they are translocated into the ER lumen via fusion to signal peptides. Indeed, the rate of secretion is in some cases comparable to that of natural secretory proteins, but the mechanisms of bulk-flow, are poorly understood. Retention of some ER residents, through competition or steric hindrance, could account for the enrichment of cargo molecules in transport vesicles without a requirement for active sorting. The fact that chaperones and other folding proteins are part of large complexes could also minimise diffusion into transport vesicles, and since unfolded or partially folded proteins bind to the chaperons, they are retained as well (Sitia and Braakman, 2003). Another model is recycling of ER chaperones and folding helpers that often contain signals recognized in the Golgi complex and promote retrieval into the ER by retrograde vesicles. Misfolded proteins can thus be recovered from the Golgi complex by virtue of their association with a recycling chaperone (Phillipson et al., 2001). The retrieval and retention mechanisms are not mutually exclusive, but can very well coexist.

Sz. pombe - a uni-cellular host resembling higher eucaryotes

Schizosaccharomyces pombe (*Sz. pombe*) – or fission yeast – is a unicellular eukaryote in the class of Ascomycetes yeasts and taxonomically and evolutionary distant from budding yeast (Russell and Nurse, 1986). In the 1890's it was isolated from millet beer from East Africa ("pombe" means beer in Swahili). *Sz. pombe* proliferates by dividing like animal and plant cells, while other common yeast as *Saccharomyces cerevisiae* proliferate by budding. "Schizo" means division or fission, and accordingly the name *Schizosaccharomyces pombe* derives from the fact that it originates from beer and its mode of proliferation.

Sequencing the *Sz. pombe* genome was completed in 2002, thereby making it the sixth eukaryote genome to be completely sequenced (Wood et al., 2002). The genome composed of three chromosomes is 13.8 Mb, and the total number of genes is estimated to 4824, about 1000 less than in *S. cerevisiae*. Notably, about 50 disease-related genes in humans are also present in *Sz. pombe*. Due to the apparent similarity to higher organisms in terms of the mode of proliferation, *Sz. pombe* has long been used widely as a model for cell division in research in molecular genetics and cell biology. In 2001 Paul Nurse received the Nobel Prize in Medicine for his research on the *cdc2* gene, which is the key regulator of the cell cycle in *Sz. pombe*.

Although, *Sz. pombe* is classified as yeast, it has several properties similar to higher eucaryotic organisms, but different from that of other yeast. Apart from the cell cycle control these properties include: i) promoter organization – a number of mammalian promoters including SV40 and hCMV promoters are active in *Sz. pombe*. This is not surprising since, the transcription mechanism is similar to that of higher animals and the transcription initiation site is located 25 to 30 bp downstream of TATA box (Dieci et al., 2000; Jones et al., 1988; Brys et al., 1998); ii) gene splicing – introns can be found in 43% of *Sz. pombe* genes (Kaufer et al., 1985); iii) organelle morphology (Russell and Nurse, 1986); iv) post-translational modification of proteins, like glycosylation (Chappell and Warren, 1989; Tanaka et al., 2005), v) chromosome structure and function (Yanagida, 1990), vi), and codon usage (Giga-Hama and Kumagai, 1999).

Heterologous protein expression in *Sz. pombe*

Due to similarities to higher eukaryotic cells, together with the features of its facile use as a unicellular microbial host, *Sz. pombe* is widely adopted as a model organism for cell biology studies (Giga-Hama and Kumagai, 1999). Also, since eucaryotic proteins expressed in *Sz. pombe* are more likely to be properly folded, recombinant expression in *Sz. pombe* can reduce the protein insolubility problems found in *E. coli*.

Wild-type *Sz. pombe* is prototrophic and therefore capable of growing on minimal media i.e.,

media that have the minimal number of nutritional components to support growth. Auxotrophic strains have been developed to provide the basis for selection of successful transformation of complementing plasmids on minimal media. The most commonly used selectable markers used in yeasts are requirements for leucine (leu2), uracil (ura3), histidine (his3), and tryptophan (trp1) (Siam et al., 2004).

Classical studies in *S. cerevisiae* and *Sz. pombe* genetics have generated a wide array of potential cloning vectors (Siam et al., 2004), and, in the process, defined which plasmid and host genomic sequences are important in expression technology. Autonomous replication sequences (ARS) in the chromosomes of *Sz. pombe* have been isolated, which can promote replication and maintenance of plasmids (Okuno et al., 1997; Dubey et al., 1994; Maundrell et al., 1988). The region required for replication is several hundred base pairs long, much longer than that of *S. cerevisiae*. No single copy plasmids have been described for *Sz. pombe,* because the centromere is too large to be placed in a typical bacterial shuttle vector (Siam et al., 2004). Because of this, fission yeast has a high loss of plasmid even under selection, so 10-50% of cells in a culture have lost the plasmid (Hayles and Nurse, 1992). There have, however, been developed integrating vectors to circumvent this problem.

Effective protein secretion into the media can be achieved in *Sz. pombe* employing different signal peptides, among these are the pGKL killer toxin signal peptide used for expression of Mouse α-amylase (Tokunaga et al., 1993), Plus (p) pheromone signal peptide for expressing streptokinase (Kumar and Singh, 2004) and interleukin-6 secreted employing its own signal peptide (Giga-Hama and Kumagai, 1999). Secretion is beneficial when expressing recombinant proteins, since expression of the protein into the culture media makes purification much easier. Furthermore, while going through intracellular organelles as ER and the Golgi apparatus, secreted proteins undergo correct folding, disulfide bonding, and glycosylation. As in eukaryotic cells the Golgi apparatus in *Sz. pombe* is well organised and morphologically distinct (Smith and Svoboda, 1972), while *S. cerevisiae* has poor Golgi morphology (Chappell et al., 1994; Chappell and Warren, 1989), thus affecting the nature of the post-translational modifications carried out. However, as in *S. cerevisiae,* there is a considerable protease activity in *Sz. pombe* both intracellular and extracellular (Kjaerulff and Jensen, 2005; Kumar and Singh, 2004; Tokunaga et al., 1993). The proteolytic degradation can be impaired by adding peptone to the media (Kjaerulff and Jensen, 2005). The peptides from the peptone may function as substrate for the proteases, thereby protecting the protein of interest from proteolysis.

The OD_{600} at the starting point of the expression influences the yield from secreted recombinant protein, probably due in part to the presence of proteases. An optimal starting

point is observed at OD_{600}=0.3 for the protein streptokinase. Furthermore, higher expression efficiency is seen when the pre-culture is grown in YEA media opposed to minimal media containing thiamine. Expression of the recombinant protein causes a decrease in growth rate of approximately 50% (Kumar and Singh, 2004).

The post-translational modifications of proteins carried out in *Sz. pombe* reflect the elaborate secretion apparatus. The oligosaccharide processing is more complex than in *S. cerevisiae*, although different from mammalian cells, and it has been shown that *Sz. pombe* has a membrane localised galactosyltransferase in the Golgi, homologous to the one found in mammalian cells (Chappell et al., 1994; Chappell and Warren, 1989). Also N-acetylation, isopenylation (Giannakouros et al., 1992), phosphorylation, and myrisylation (Jakubowicz et al., 1993; Koegl et al., 1995) are reported to occur in *Sz. pombe.*

One of the most commonly used promoters for recombinant expression in *Sz. pombe* is the promoter *nmt1* (no message in thiamine) (Siam et al., 2004). This promoter is regulated by the concentration of thiamine in the media – it is active in the absence of thiamine, and turned off by the presence of more than 15 μM thiamine. The induction ratio of the promoter is about 300-fold. In *Sz. pombe* the *nmt1* promoter is more than 80% stronger than the commonly used *S. cerevisiae* derived *adh* promoter. A shortcoming of the *nmt1* promoter is that it is very slowly induced following exclusion of thiamine from the growth medium. The promoter requires approximately 16 hours for full induction, probably as an effect of exhaustion of thiamine in the cells. Importantly, in rich media containing yeast extract, enough thiamine is present to repress the promoter; therefore, expression requires synthetic media without thiamine (Siam et al., 2004).

Lastly – as an important feature for an expression host for therapeutic proteins – *Sz. pombe* meets the basic requirements for safety described in the Recombinant DNA Technology Industrialization Guideline (Ministry of Economy, Trade and Industry Notice No. 259 in 1998). In addition, Microorganisms Safety Information Data of Recombinant DNA Technology Industrialization Information System of METI shows that *Sz. pombe* is found in fruits (apples and grapes) and food (grape juice, palm wine, sugar cane and syrup) as reference data for long-term use of microorganisms.

Invertase – as selectable marker in yeast

Invertase has been used extensively as a selectable genetic marker in yeast and is therefore attractive as a reporter for screening and selection. The gene encoding invertase in *Sz. pombe*, Inv1, has been cloned and is the sole gene responsible for invertase activity in this yeast. The molecular mass of the protein part of invertase is 60 kDa per subunit and in the glycosylated

form the molecular weight is 120 kDa (Moreno et al., 1990). The enzyme (also called beta-D-fructofuranoside fructohydrolase) is almost exclusively located in the cell wall outside the plasma membrane, attached to the cell wall, where it conducts the enzymatic reaction of hydrolysing sucrose to glucose and fructose (Figure 11).

Fructose and glucose, but not sucrose, can be imported into the cell and utilised as carbon source in the metabolic pathway. Therefore, the secretion of invertase is essential for cell survival when growing in an environment containing only sucrose (Carlson et al., 1983). Several assays have been developed to measure the presence of invertase both quantitatively and qualitatively (Darsow et al., 2000).

Because invertase secretion is essential only for growth on sucrose, the cell viability can be maintained under growth on other carbon sources. Therefore, the enzyme is a convenient marker in genetic selection. It has been widely used to measure the effects of mutations within signal sequences (Perlman et al., 1986; Ngsee et al., 1989; Kaiser et al., 1987), to identify amino acid sequences mediating localization to yeast vacuoles (Saalbach et al., 1991; Tague et al., 1990; Klionsky et al., 1988), and to monitor the amounts of human proteins expressed for commercial purposes (Hitzeman et al., 1990). Furthermore, the choise of invertase marker is supported by the fact that the invertase protein has been shown to be tolerant to many different amino terminal extensions (Jacobs et al., 1997; Kaiser et al., 1987), and invertase as reporter protein has been utilised in the signal sequence trap developed by Klein *et al.* (Klein et al., 1996). This *S. cerevisiae* based system identifies signal sequences by fusing a library N-terminally to the mature form of invertase; thus, signal sequences can be detected on the basis of the ability to grow on sucrose as carbon source.

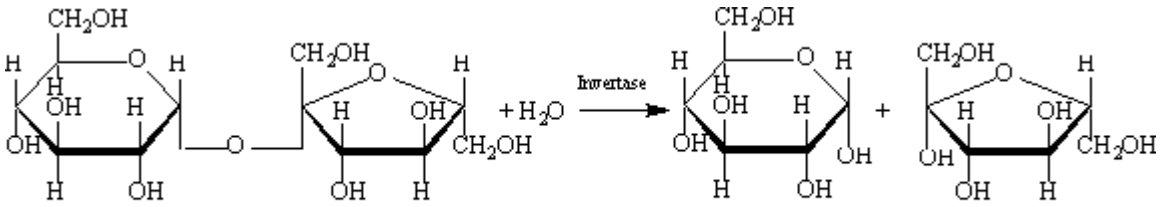


Figure 11. Sucrose is hydrolysed to fructose and glucose by invertase

Sucrose, is a disaccharide composed of an alpha-D-glucose molecule and a beta-D-fructose molecule linked by an alpha-1,4-glycosidic bond. Invertase is a b-fructofuranosidase that hydrolyses sucrose as well as other b-fructans such as raffinose.

Tumor necrosis factor α - a relevant and recognised target

TNF-α has often been used as model target to test new therapeutic approaches for protein down regulation, as it is a significant contributor to the pathology of a number of diseases associated with chronic inflammation such as Rheumatoid Arthritis and Crohn's disease. Inhibition of TNF-α activity significantly benefits these patients (Dinarello, 2000; Feldman et al., 1998). It has been recognised as target for therapeutic vaccination and has been a target for an AutoVac vaccine at Pharmexa (Dalum et al., 1999). The considerable in-house knowledge of TNF-α as well as the existence of TNFα specific tools made it an obvious model protein for the work presented here. Recombinant TNF-α has been produced in *E. coli* as soluble protein both at Pharmexa and other laboratories (Van, X et al., 1994). The task to introduce foreign epitopes in TNF-α has not been trivial. A prerequisite for soluble recombinant expression in *E. coli* is the formation of trimers - monomeric TNF-α is not soluble. Also, it has been reported that many single-point mutations results in insoluble protein expression (Nielsen et al., 2004).

Biology of Tumor necrosis factor α

TNF-α, named after its activity to cause tumor necrosis *in vivo* when injected into tumor-bearing mice, is a multifunctional proinflammatory cytokine belonging to the TNF super-family (Bodmer et al., 2002). Many names have been previously given to the protein originating from its biological activity as cachectin, necrosin, macrophage cytotoxic factor, macrophage cytotoxin, cytotoxin, haemorrhagic factor, and differentiation-inducing factor. These factors have in time been recognised as the same and are now known collectively as TNF-α (Old, 1988; Das, 1991). TNF-α exerts an extreme spectrum of biological processes and is capable of acting independently and in conjunction with a variety of other factors to affect the phenotype and metabolism of cells in every tissue of the body. The actions of the TNF-α can therefore be quite complex. TNF-α shows a notable functional duality, being involved in both tissue development and destruction (Sugarman et al., 1985). Tissue type, cellular context and TNF-α receptor composition, timing, and duration of TNF-α action are important parameters influential on the overall effect of TNF-α action.

TNF-α is mainly produced by macrophage (Ruddle, 1992), but also by a broad variety of different tissues and cells. However, it is generally thought that TNF-α is not produced constitutively by normal cells, but instead, potently induced by invasive challenges caused by neoplastic or infectious disease perpetuating inflammatory responses (Roach et al., 2002).

A trimeric structure is characteristic for all members of the TNF family, and the ligand-induced trimerisation of their receptors seems to be the critical event in initiating signalling

(Bazan, 1993). Both the membrane-bound and the soluble secreted forms of TNF-α show biological activity (Kriegler et al., 1988; Luettig et al., 1989).

The most profound TNF-α activity is that of proinflammatory mediator, but the molecule has optional capacity to induce apoptosis. In concert with other cytokines, TNF-α is considered to be a key player in the development of septic shock (Mannel and Echtenacher, 2000). Whereas high concentrations of TNF-α induce shock-like symptoms, the prolonged exposure to low concentrations of TNF can result in a wasting syndrome, that is, cachexia. This can be found, for example, in tumor patients. Also, many other effects are known, including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. TNFα is cytotoxic for many cells, especially in combination with γ-interferon.

The multiple biological effects of TNF-α are mediated by two distinct cell surface TNF-α receptors (TNFR1 and TNFR2). TNFR1 and TNFR2 activate different intra-cellular signal-transduction pathways and modulate expression of a broad number of genes (Brockhaus et al., 1990). Not only is the ratio of TNFR1 to TNFR2 important to decide which effect TNF-α has on a specific cell type, but also the levels of the several different intracellular mediators are critical for the physiological effect (Locksley et al., 2001). TNFR1 is constitutively and ubiquitous expressed in most tissues, whereas expression of TNFR2 is highly regulated and is typically found in cells of the immune system. So, in most cells, TNFR1 appears to be the key mediator of TNF-α signalling, whereas in the lymphoid system TNFR2 seems to play a major role. TNFR1 binds membrane-integrated as well as soluble TNF-α. However, only membrane bound TNF-α, and not soluble TNF-α, can activate this receptor (Grell et al., 1995).

One TNF-α trimer binds three TNF-α receptor molecules (Vandenabeele et al., 1995) and this ligand-dependent trimerisation of the receptors was long considered as the key event for signal initiation. However, the receptor activation appears to be more complicated (Chan et al., 2000). Ligand binding to the preformed TNFR complex either induces an activating conformational change of an already signal competent receptor complex or it allows the formation of higher-order receptor complexes, whereby it acquires signal competence.

Clinical indications of TNF-α up-regulation

Neutralisation of TNF-α by soluble receptors or monoclonal anti-TNF-α antibodies (Feldman et al., 1998; Dalum et al., 1999; Feldmann and Steinman, 2005; Taylor et al., 2004; Elliott et al., 1993) point to attractive therapies for TNF-α-mediated disorders. There are several clinical indications, where TNF-α plays a central pathogenic role generally characterised by chronic inflammation. These indications include, rheumatoid arthritis, Crohns disease, psoriasis (Feldman et al., 1998), and cancer cachexia (Natanson et al., 1989; Tracey et al.,

1988). Given the far-reaching effects of TNF-α on both the innate and adaptive part of the immune system, disruption of TNF-α signalling pathways leads to increased susceptibility to a variety of pathogens, in particular, those known to survive intracellularly (Marino et al., 1997). TNF-α or TNFR1-deficient mice show enhanced sensitivity when challenged with Mycobacterium tuberculosis, Lysteria monocytogenes, Toxoplasma gondii, or Leishmannia major (Flynn et al., 1995; Rothe et al., 1994; Vieira et al., 1996; Deckert-Schluter et al., 1998). Furthermore, data from patients who have been treated with an antagonising TNF-α activity, show susceptibility to infection and reactivation of latent tuberculosis early after commencement of anti-TNF-α therapy (Wolfe et al., 2004; Wagner et al., 2002). However, no serious adverse effects have been shown, and anti-TNF-α therapy is considered safe.

Structure of TNF-α

The TNF-α molecule is expressed as a non-glycosylated type II transmembrane protein arranged in stable homotrimers (Kriegler et al., 1988; Tang et al., 1996). This membrane-integrated form is processed by posttranslational proteolysis by the metalloprotease TNF-α converting enzyme (TACE) releasing a soluble homotrimeric molecule (Black et al., 1997). The TNF-α monomers self associates into a soluble non-covalently bound symmetrical trimer of 51 kDa (Jones et al., 1989).

The crystal structure of the human 17 kDa soluble TNF-α has been resolved to a 1.8 Å resolution (Eck and Sprang, 1989; Cha et al., 1998; Jones et al., 1989) and the murine

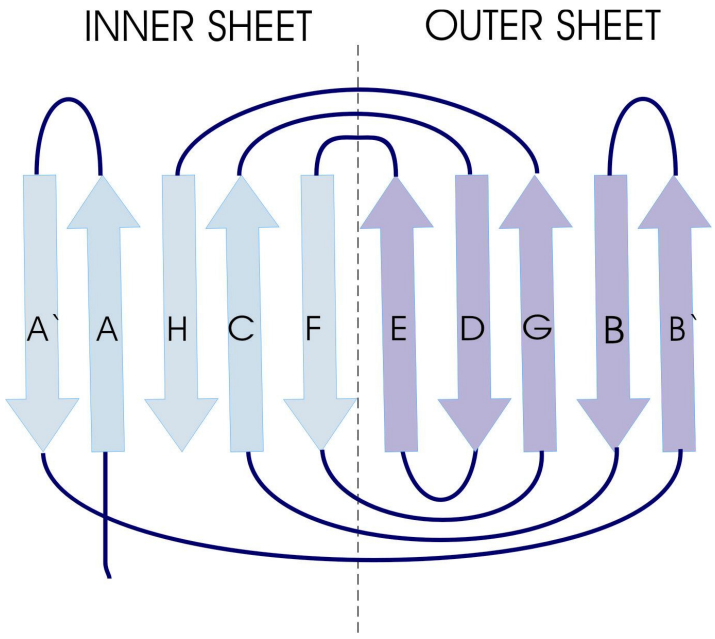


Figure 12. Diagram depicting the β-strands of TNF-α. The TNF monomer is a β-sheet sandwich composed of an inner and outer sheet β sheets each made up of 5 β-strands. The inner sheet faces the trimer interface (Eck& Spangs, 1989). The strands are named starting from the N-terminal

structure to a resolution of 1.4Å (Baeyens et al., 1997). The main chain topology of the TNF-α monomer is a β-sandwich formed by two anti-parallel β-sheets, both comprising five β-strands, forming a 'jelly roll' -structure (Narhi et al., 1996), typical for the TNF ligand family and some viral capsid proteins (Bazan, 1993) (Figure 12). The β-strands A, H, C, and F make up the inner sheet of the jelly roll and are involved in trimer contacts. The outer sheet is made up of strands B, G, D, and E, which form the outer surface of the trimer. Strand b in this outer sheet is interrupted, at its amino terminus (residue 26) with a 20 amino acid chain forming strands A' and B' A disulfide bridge is found between the loop connecting strands C and D with the loop connecting strands E.

The measure of relative flexibility yielded from the crystallographic data (Jones et al., 1990) shows that the β-strands forms a quite rigid scaffold, particularly the inner sheet situated at the core of the trimer (Figure 13). This type of structure provides a very stable core arrangement from which connecting loops can provide structural variations. The loops linking the strands of the outer sheet are more flexible than the ones of the inner. The C-terminus is embedded in the base of the inner β-sheet, in close proximity to the N-terminus. The N-terminus is highly flexible and, as far as residue 10, fairly independent of the rest of the molecule, and the first six amino acids can be deleted without loss of activity (Beutler and Cerami, 1988). Lysine-11 forms an iron pair with the C-terminal leucine-157, stabilising the tertiary and quaternary structure.

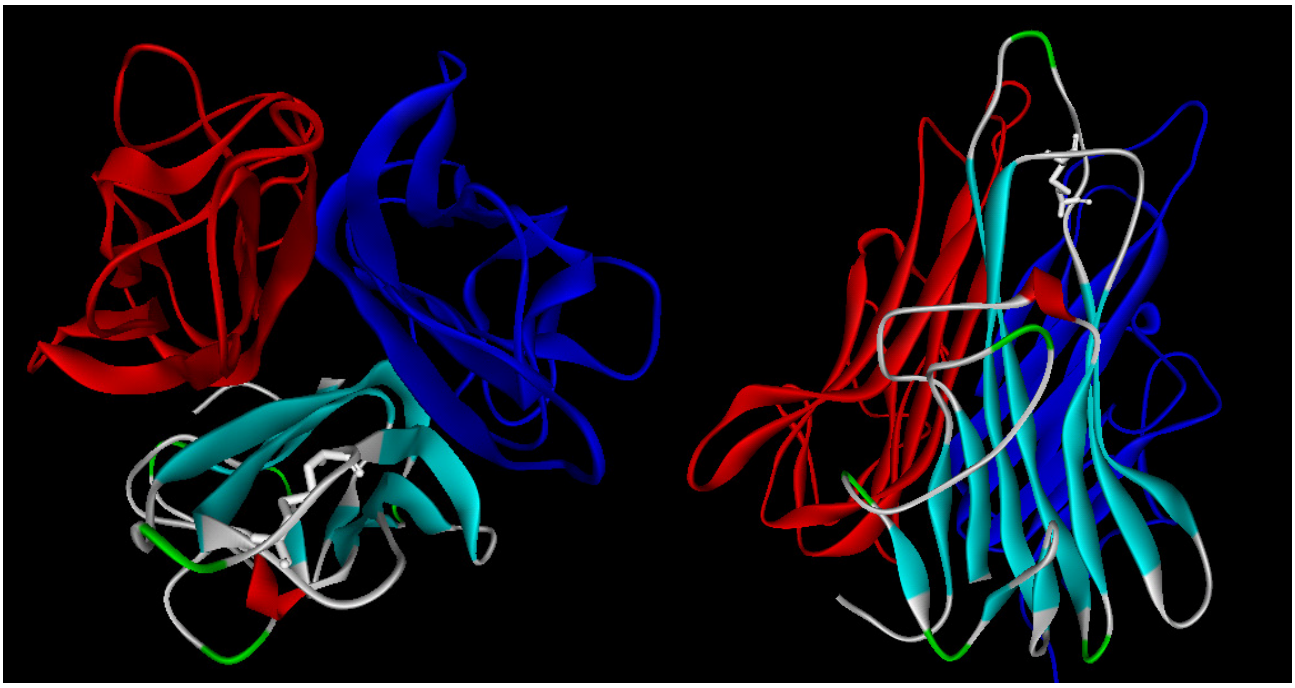


Figure 13 Three-dimensional structure of TNF-α. The molecule is presented as timer shown top-view or side-view. *Structure coordinates Protein Data Bank code tnf1A*

The inner β-sheet is relatively flat, but the surface of the outer is quite curved. The spatial distribution of residue side chains is typical of β-sandwich proteins; the interior is closely packed, largely with hydrophobic amino acids. The TNF-α trimer is a compact molecule and the three subunits tightly interact to form a solid conical shape with loops protruding out from the core. The majority of the interactions between the subunits implicate hydrophobic residues. The receptor-binding domain is found in the bottom of the bell shape of the trimer (Eck and Sprang, 1989), where also the N- and C-terminus as well as lysine-11 are situated.

Rational design of TNF-α vaccine candidates

In order to make rational designed vaccine molecules, suitable sites for T cell epitope insertions have to be identified (Nielsen et al., 2004). By examining the crystal structure of TNF-α, a flexible loop was identified as being the only region in the molecule which was not to proposed to have intra- or inter-molecular interactions, and accordingly it was chosen for epitope insertion. The insertion causes extension of the loop potentially destabilising the TNF-α molecule; therefore an extra cysteine bridge was introduced in some vaccine candidate molecules. Another approach was to connect the three TNF-α monomers that then could form a trimer-like molecule. The "monomers" were interconnected by two Th-cell epitopes that consequently need not to be included in the actual TNF-α structure, and in this way accomplished minimal disturbens to the structure. Both approaches resulted in a soluble protein expression in *E. coli* of some of the variants. Subsequent testing revealed that these TNF-α variants were as stable as wt-TNF-α and their biological activity comparable to wt-TNF-α.

In the TNF-α vaccine project at Pharmexa 12 different insertion sites have been predicted by rational design (Nielsen *et al.* 2004, unpublished in-house data from Pharmexa). Further, due to partial substitutions and other modifications 30 different TNF-α AutoVac proteins were suggested, designed and tested for folding competence. These rational predictions were, among other things, based on the crystal structure of human TNF-α, superimposing human and murine crystal structuresm, as well as by evaluating sequence alignments of TNF-α from species in which orthologue homologues have been identified and from other members of the TNF-α superfamily (Nielsen et al., 2004). The AutoVac candidates were made by site-directed mutagenesis and tested for expression capability in *E. coli*, before a candidate for animal testing was selected. Of the 12 different insertion sites 4 were able to accommodate the epitope insertion and only two showed a satisfactory level of expression. Hence, finding AutoVac variants requires considerably time and resources. Therefore, a complementary approach will be much appreciated.

RESULTS

The result section is divided into 5 main parts: 1) Generation of the TNF-α library, 2) analysis of the library, 3) generation of the *Sz. pombe* screening host, 4) screening/selection of folding competent variants, and 5) analysis of selected variants.

Generation of the TNF-α library

Rationale of the library plasmid construction.

The first step in project was the generation of a library TNF-α AutoVac variants with PADRE distributed randomly along the whole TNFα sequence. A library generation plasmid was made from the pUC19 plasmid (Yanisch-Perron et al., 1985) (Figure 14), which, besides being small in size, has the notable feature of being free of *Not1* restriction sites used in subsequent steps. The promoter sequence was removed from the plasmid to delete any expression from the plasmid that could introduce bias in the library. Further, the restriction sites for *Nco1* and *Sac1* were introduced at the 5'-end and 3'-end of TNF-α, respectively. These sites facilitate that the library readily can be moved from this plasmid into a plasmid suitable for screening in *Sz. pombe*. The resulting plasmid is only 2910 base pairs long, including the TNF-α sequence. The small size ensures that the transposon integrates with high frequency in the TNF-α cDNA sequence.

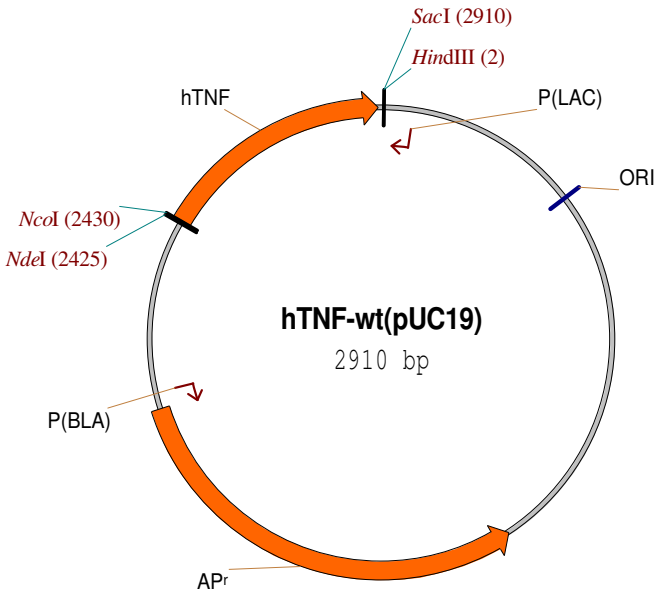


Figure 14. Schematic representation of the library generation plasmid

TNFα cloned into the pUC19 plasmid. Flanking the TNF-α coding cDNA is *Nco1* and *Sac1* restriction sites, facilitating the transfer of the library to the screening plasmid. The plasmid contains no *Not1* sites, which is flanking the transposon and will be used in the subsequent steps. The promoter sequence of the plasmid has been removed to avoid any expression of the insert.

Overview of the library generation steps

In order to generate a PADRE insertion library of TNF-α, a *Not1* restriction site is randomly inserted in the TNF-α cDNA sequence by transposition. After the insertion the transposon is removed by the Not1 restriction enzyme and replaced by the PADRE epitope. At the same time all wt-TNF-α encoding cDNA must be eliminated from the pool of variants to avoid background of wild-type protein later in the screen. Thus, a 5-step process was applied to the library plasmid (Figure 15); i) *in vitro* transposition, ii) isolation of mutagenised TNF-α cDNA iii) reintroducing mutagenised TNF-α cDNA into vector, iv) interchanging transposon with the PADRE epitope, v) transferring the library to the screening vector.

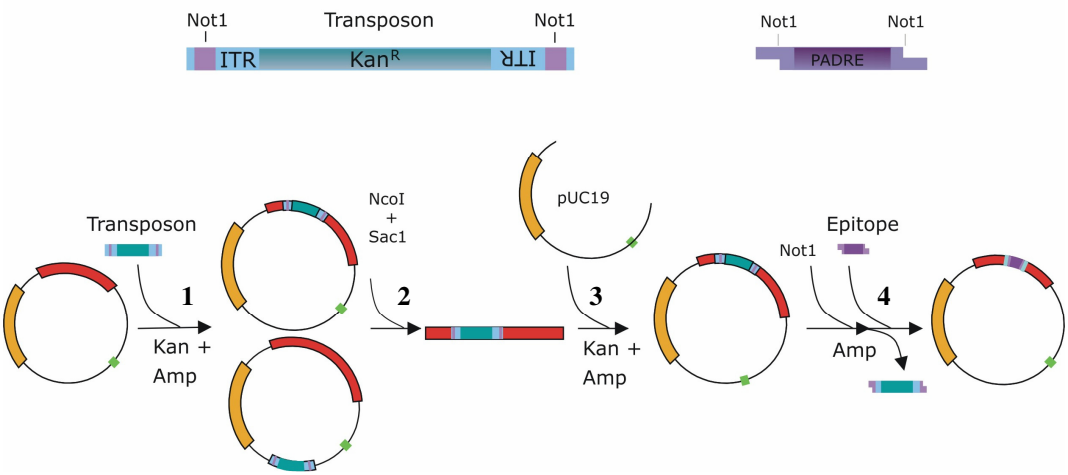


Figure 15. Construction of the TNF-α insertional Library.

1) The Entranceposon and transposase are reacting with the acceptor plasmid integrating transposon randomly herein.The plasmids are transformed into *E. coli* and amplified on LB solid media containing both ampicillin and kanamycin, ensuring selection only of plasmids with integrated transposon. 2) The transposon can be integrated both in the vector as well as in the TNF gene. So, to remove plasmids with vector integrated transposons, the TNF-α + Entranceposons insert is reintroduced into the original library vector by use of the restriction enzymes NcoI and SacI. 3) Transformants are again selected on LB solid media containing ampicillin and kanamycin. This is important to eliminate any wt TNF from the pool of variants that later could show up as false positives in the screen. 4) Now the Entranceposon is ready to be replaced with the PADRE epitope. To do so the Entranceposon is removed by use of the restriction enzyme *Not1*. The open plasmid is ligated together with a DNA fragment encoding the epitope, having convertible ends. The resulting plasmid is amplified on LB solid media containing ampicillin. The TNF-α library is ready for introduction into the screening vector.

In vitro transposition

The transposon was inserted into the plasmid using the MuA based transposon MGS Kit from Finnzymes (Figure 16). This kit contains a MuA transposases and a linearly DNA transposon (Entranceposon). The transposon has been genetically engineered by introducing *Not1* sites near each rim of the transposon, whereby the bulk of the transposon can be removed. After transposon removal, the residual sequence comprises 15 base pairs made up by 10 bases holding the *Not1* restriction site and a 5 base pair duplication of the target site. In order to

ensure as many hits as possible, the reaction was performed on a PCR thermoscycler where the temperature can be precisely regulated and remains constant to achieve optimal conditions for transposition. Furthermore, to attain as many insertions as possible, the reaction time was extended from 1 (suggested in the suppliers manual) to 3 hours (in my reserch).

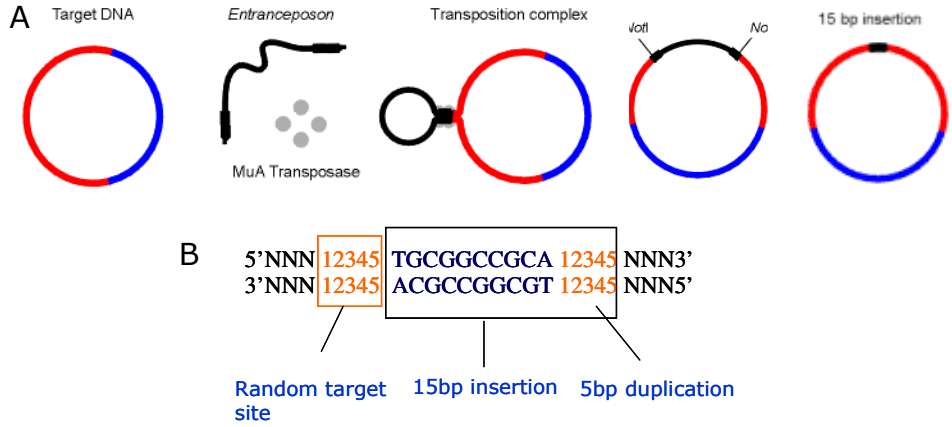


Figure 1.3 Reaction steps in the transposon MGS Kit from Finnzymes
A) The MGS system introduces 15 bp insertions at random positions in target plasmid DNA by use of Entranceposons and the MuA Transposase. The Entranceposon holds the kanamycin resistance gene.
B) A 15 bp insertion is left after transposon removal, including the *Not1* recognition sequence and a 5 bp duplication

After the transposon reaction the reaction was divided into 20 aliquots, and 7 of these were transformed into *E. coli* cells plated onto 7 plates of solid LB-media containing kanamycin (antibiotic marker for the Entranceposon) and carbencillin (antibiotic marker for the plasmid). The double selection ensures that only plasmids containing the transposon are able to amplify. In all, 35,000 colonies appeared on the seven plates representing equal numbers plasmid with independent transposon insertions. The 7 reactions were kept independently as parallel reactions throughout the library generation.

Isolating the mutagenesised TNF-α cDNA

Plasmids where the transposon has integrated in the vector, and hence left the TNF-α sequence intact, are not acceptable in the pool of plasmids, since they would lead to a background during the screen and must therefore be removed. Therefore, the plasmids are digested with *Nco1* and *Sac1* flanking the TNF-α cDNA, to select inserts holding the transposon and the fragment comprising TNF-α plus the transposon (Figure 17).

On the basis of the intensities of the bands appearing on the agarose gel, the transposon integration was estimated to have occurred in the cDNA of TNF-α in 50% of the plasmid molecules. Assuming that the modified Mu transposon behaviour is random, the probability of insertion in the TNF-α coding sequence during transposition should be the ratio of the coding

sequence to the size of the total plasmid (approximately 20%). However, transpositions that disrupt critical elements of the plasmid (the plasmid origin or the *Amp*^{r} gene, constituting about half of the vector) will not be recovered after transformation, thereby increasing the relative number of hits in the TNF-α cDNA compared to the overall number of hits in the plasmid. Consequently, as half of the transposon integrations are in the TNF-α cDNA, the other half is in the vector now removed, the number of individual transposon insertions in the TNF- cDNA in the library was estimated to 17,500.

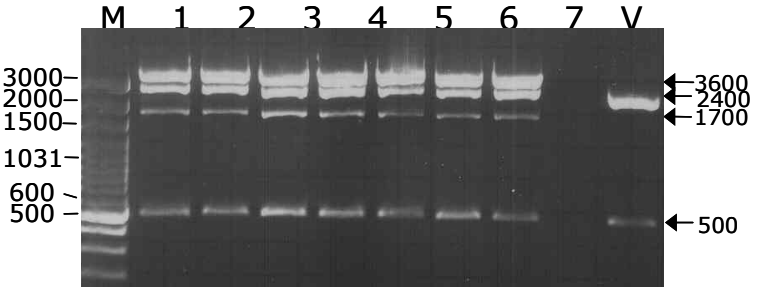


Figure 17. Inserts containing transposon are selected. The plasmids are digested with *Nco1* and *Sac1*.
Fragments from the 7 independent transposon reactions are run at an argarose gel. Fragment of 3600bp (transposon + vector), 2400bp (vector), 1700bp (TNF-α cDNA + transposon) and 500bp (TNF-α cDNA) are seen. V is the wt-TNF-α pUC19 vector prior transposition.

Reintroducing the cDNA into the library vector

After isolation the fragment with the TNF-α and transposon was ligated into a pristine library vector. In order to achieve as many transformants as possible and thereby maintain a high diversity of the library despite of the many steps in the library protocol, several measures were taken. Firstly, *E. coli* preparation of high transformation efficiency ($1x10^{10}$ transformants /μg) was used. Secondly, the ligase was heat inactivated before transformation, a step that has shown to increase the transformation efficiency 10-260 times (Ymer, 1991; Michelsen, 1995). And thirdly, the ligation mix was desalted by dialysis against milliQ water through a Millipore 0.05 um VMTP membrane in order remove residual salt from the reaction (Jacobs et al., 1990).

The transformed *E. coli* cells were again subjected to double antibiotic resistance selection. The *E. coli* transformations were very efficient. The colonies were confluently covering the plates so their number could not be estimated accurately. Ideally the transformants should have been plated out on additional plates. However, based on practical experience, colony numbers up to around 6000 per plate can be estimated by counting. Therefore, I estimate that the confluent layer represents, at least 3 fold more, which is 126,000 in this experiment comprising 7 plates. Thus, as intended, the high number of transformants ensured that the size of the library was not compromised. Hence, the number of single mutagenic events in the TNF-α library was kept at 17,500.

Interchanging the transposon with the PADRE epitope

The next step was to remove the transposon and replace it with the PADRE epitope. The library plasmids were digested with *Not1*, as two *Not1* sites flanks the transposon, and subsequently treated with calf intestine phosphatase to remove the phosphate units at the DNA fragments (Figure 18). Digesting a PADRE containing plasmid (PADRE pVax) with *Not1* retrieved a PADRE insertion fragment. In this way the PADRE sequence is compatible with the *Not1* restriction sites left after transposon. The reason for using a plasmid-derived PADRE fragment is that it now has phosphate moieties on the 5' ends, and therefore can be ligated together with the un-phosphorylated vector fragment. This is necessary in order to avoid self-ligation of the plasmid. If the plasmid closes on itself, it generates TNF-α variants holding the 15 base pairs insert deriving from the transposon event and not the epitope insertion.

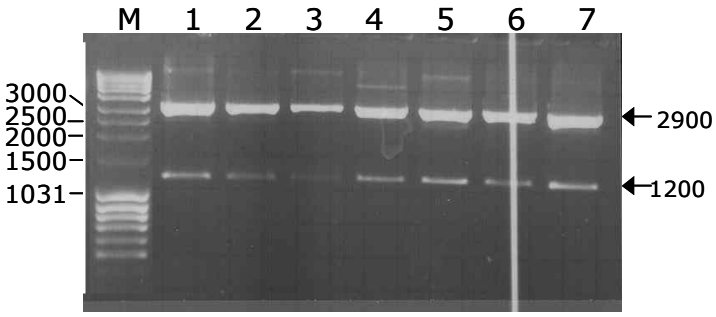


Figure 18. Removal of the Entranceposon from the TNF-α cDNA fragment.
After Digestion with *Not1*, fragments were run on an agarose gel. 2 bands appeared at 2900 base pairs (vector + TNF-α) and 1200 base pairs (transposon).

The resulting plasmids were amplified in *E. coli* cells on carbencillin selection. Again the transformation efficiency was very high, leading to confluent growt, and prohibiting the exact number of transformants to be known. It was presumed, however, that the number of variants was above 126,000 and therefore that the size of the library was not compromised.

Transferring the library to the screening vector

The TNF-α variant library was now ready, and introduced into the vector used for screening, described in detail below. In order to do so, the plasmid was digested with *Sac1* and *Nco1* and

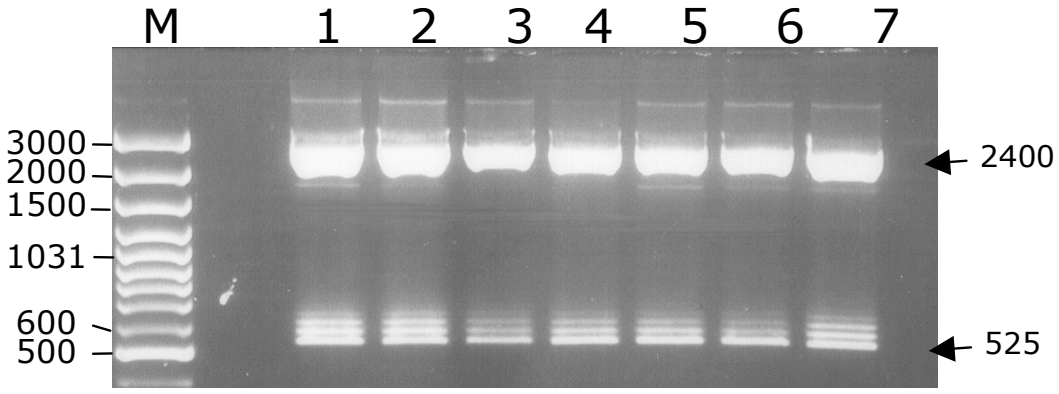


Figure 19. Concatenation of TNF-α insert fragment. In order to move the TNF-α variants to the screening vector, the plasmid is digested with *Sac1* and *Nco1* and 2 bands are expected on the agarose gel at 525 base pairs (TNF-α + insert) and 2400 base pairs (pUC19 library vector).

2 bands are expected on the agarose gel at 525 base pairs (TNF-α + PADRE) and 2400 base pairs (vector). Surprisingly, a ladder of inset fragments was observed starting by 545 base pairs as the smallest fragment (Figure 19). This was due to an extensive concatenation of the epitope DNA fragment during ligation, effected by the phosphorylated insert fragment.

The solution was to extract the lower DNA fragment from the gel and use it as insert in the screening vector. The result from this procedure showed that the approach was successful, since no repetition of the epitope was later found in any of the variant sequences. The 545 base pair fragment was cloned into the screening vector and about 84,000 colonies appeared after carbencillin selection. This number of transformants was estimated as sufficient to maintain the estimate of 17,500 single transposon hits.

Analysis of the Library

The TNF-α sequence comprises 474 base pairs between which the insert can be situated and the insert can be found in 2 different directions, giving 948 possible different variants of TNF-α in the library. The library represented 17,500 single transposon hits, so each possible sequence would in theory be represented approximately 20 times if the library were completely random. This means that the library was likely to be exhaustive.

The library was now ready for selection, but prior to this, the quality of the library was analysed. This was done by randomly picking and sequencing the insert of 44 clones (18 from the pUC19 library and 26 from the screening *Sz. pombe* vector library) (Figure 20). Each clone contained one single insert. Insertion points are evenly distributed along the 474 base pair long TNF-α sequence, and all, except 2, were unique.

```
ATGGTCAGATCATCTTCTCG AACCCCGAGTGACAAGCCTG TAGCCCATGTTGTAGCAAAC CCTCAAGCTGAGGGGCAGCT
CCAGTGGCTGAACCGCCGGG CCAATGCCCTCCTGGCCAAT GGCGTGGAGCTGAGAGATAA CCAGCTGGTGGTGCCATCAG
AGGGCCTGTACCTCATCTAC TCCCAGGTCCTCTTCAAGGG CCAAGGCTGCCCCTCCACCC ATGTGCTCCTCACCCACACC
ATCAGCCGCATCGCCGTCTC CTACCAGACCAAGGTCAACC TCCTCTCTGCCATCAAGAGC CCCTGCCAGAGGGAGACCCC
AGAGGGGGCTGAGGCCAAGC CCTGGTATGAGCCCATCTAT CTGGGAGGGGTCTTCCAGCT GGAGAAGGGTGACCGACTCA
GCGCTGAGATCAATCGGCCC GACTATCTCGACTTTGCCGA GTCTGGGCAGGTCTACTTTG GGATCATTGCCCTCGGAGCT
                                                                  SacI
```

Figure 20. Sequence analysis of 44 randomly picked clones.
44 *E. coli* colonies containing the screening plasmid were picked and the insert was sequenced in order to establish the insertion point of the insert. The insertion site is displayed by highlighting the last base before the insertion in red, insertions form the screening *Sz. pombe* vector are underlined.

The distribution of the insertion sites was tested at the last two steps of the construction of the library (see figure 15), to ensure that any biases introduced both during transposition event, but also importantly, in the following sub-cloning steps, are displayed in the analysis. The library was inserted into the pUC19 library vector, and secondly, in the S. pombe

screening vector. The results from sequencing of 18 randomly picked clones from the pUC19 vector and 26 clones from the S. pombe vector are presented in figure 20 in red/bold and red/bold/underlined, respectively. In both cases, the results show that the insertion sites are distributed randomly. Furthermore, this shows that no biases occurred during the transposition events, nor during the following sub-cloning step.

The DNA fragment can translate into 6 different amino acid sequences. This derives from the fact that the insert can be found in 2 orientations (direct and reverse) and 3 reading frames (RF1, RF2, RF3) (Figure 21). Therefore, the sequences were analysed for the distribution of the different amino acid sequences, so as to reveal if any biases were present in the library.

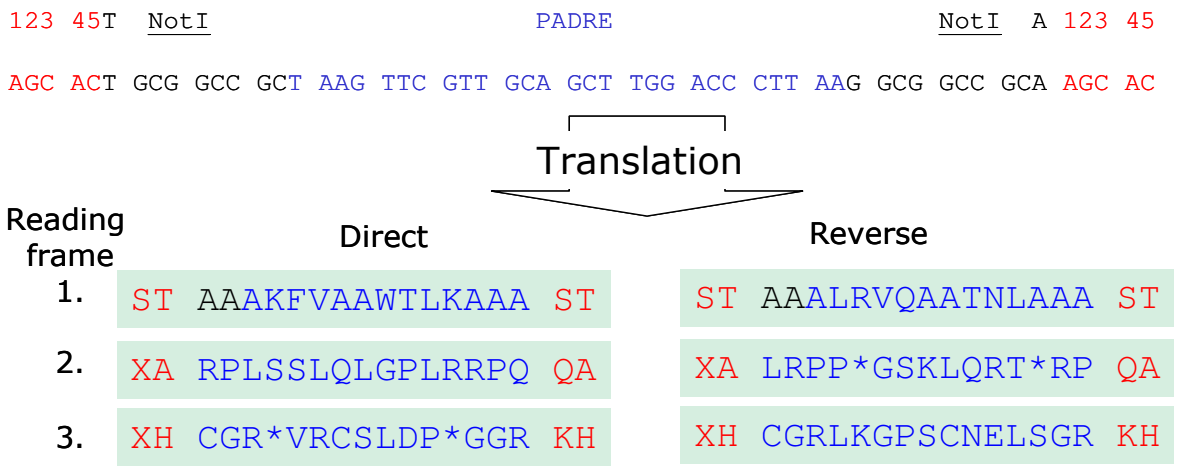


Figure 21. Possible amino acid sequences of the inserted fragment.
The red bases represent the target DNA (these bases were chosen arbitrarily here but they differ depending on the target sequence), the black bases the *Not1* restriction sites, and the blue are representing the insert. The insert can be found in 2 orientations (direct and reverse) and 3 reading frames (RF1, RF2, RF3) translating into 6 different insert sequences, 2 of which contains stop codons.

The distribution of insertion sites and sequences is shown in Table 1. Theoretically, if the transposition were truly random, the inserts should be found in 33% in each of the 3 reading frames. The data show that the distribution varies between 30-36%, indicating that the transposition has indeed been random, and analysing the two libraries individually, the same conclusion can be made. Of the 44 clones analysed, 18 have insertion sequences in the direct orientation (41 %) and 27 (59%) in reverse orientation, correlating well to the expected 50%.

If both the orientation and the reading frame be taken into account, the expected distribution would be 16% for each of the 6 possibilities. A distribution between 9% and 22% was found, corresponding well with the value of 16%. The sequences translate into 6 different amino acid sequences, 2 of which contain stop codons. Of the 4 viable sequences, the distribution was

between 14% and 37%, correlating with the expected distribution of 25%. The PADRE sequence was found in 6 of the 44 clones, representing 13% of the cDNA library and 22% of the translated library. So, no apparent biases were present in the analysed subset of clones.

In conclusion, the library exceeds the number of the possible events 20 times and the insertions appear random, both in terms of insertion site and of fragment orientation. Thus, the library is presumably exhaustive, so all possible sequences can be subjected to the subsequent screen of folding competence.

Frequency of different insert sequences	Reading frame of insertion point (33%)			Resulting insert DNA sequence (16%)	Resulting amino acid insert sequence	Frequency prior screening (25%)
	pUC19 library	*Sz. pombe* library	Total			
RF1-d (PADRE)	6 (33 %)	10 (38 %)	16 (36 %)	6 (13 %)	AAAKFVAAWTLKAAA	6 (22%)
RF1-r				10 (22%)	AAALRVQAATNLAAA	10 (37%)
RF2-d	5 (28 %)	8 (31 %)	13 (30 %)	4 (9%)	RPLSSLQLGPLRRPQ	4 (14%)
RF2-r				9 (20%)	LRPP*GSKLQRT*RP	-
RF3-d	7 (38 %)	8 (31 %)	15 (34 %)	8 (20%)	CGR*VRCSLDP*GGR	-
RF3-r				7 (16%)	CGRLKGPSCNELSGRT	7 (25%)
Total	18	26	44	44		27

Table 1. Frequency of each possible insert sequence found in 44 randomly picked clones.
The insert can be found in 3 reading frames (RF1, RF2, RF3) and 2 orientations direct or reverse (d, r) translating into 6 different insert amino acid sequences, 2 of which contains stop codons. The direct orientation was found 18 times (41 %), whereas the reverse orientation was found 26 times (59%). The expected distribution is specified in the parentheses in the top row.

Generation of the screening host and screening plasmid

The *Sz. pombe* screening host

In order to use invertase as selectable marker to screen the TNF-α insertion library, an invertase-knockout strain of *Sz. pombe* was constructed. The invertase knockout was generated from the *Sz. pombe* strain Eg660 (mat1-P Dmat2,3::LEU2-leu1 ura4-D18). A one-step gene replacement technique (Rothstein, 1983) was used to disrupt the endogenous invertase gene of the cell whereby the entire open reading frame was replaced with the *ura4* gene. The resulting strain (PY-17) showed no invertase activity and was unable to grow on sucrose media.

Rational of the screening vector constructions

The screening vector was generated using the pREP3X1 (Moreno et al., 2000) vector as background, because this is a shuttle vector (Figure 22). The invertase gene was fused to the Cpy signal sequence (Tabuchi et al., 1997; Kjaerulff and Jensen, 2005), to allow secretion of the library-invertase hybrid protein. The Cpy-signal sequence has previously been shown able to accommodate secretion of invertase, TNF-α, and GFP in *Sz. pombe* (Kjaerulff and Jensen, 2005). Between the Cpy signal sequence and the invertase gene, the *Nco1* and *Sac1* restriction sites were situated flanking a STOP codon. The restriction sites allow introduction of the library, and the STOP codon prevents background from self-ligated vector to contaminate the screen by preventing expression of the invertase gene. A similar strategy to eliminate background by introducing a stop-codon has previously been described (Olesen & Kielland, 1993).

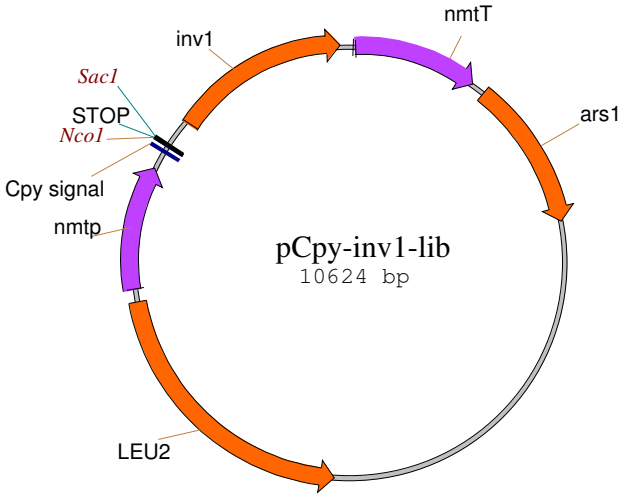


Figure 22 Schematic representation of the *Sz. pombe* screening vector.
The *Sz. pombe* screening vector plasmid consists off the following elements: 1) The Nmt1 promoter (nmtp) and terminator (nmtT) for expression in *Sz. pombe*, 2) the signal sequence from the Cpy gene for supporting secretion, 3) the mature invertase gene (minus the signal peptide) (Inv1) as folding marker, 4) NcoI and SacI restriction sites, in which the TNF-α variants can be interchanged with the STOP codon and be fused in frame to the invertase gene, 5) the leucine gene as selectable plasmid marker for transformants in Sz. *pombe*, 6) a origin for replication in *E. coli* (ori), 6) and the ampicillin gene (Amp) as selectable marker for transformants in *E. coli*.

Prediction of Signal peptide cleavage

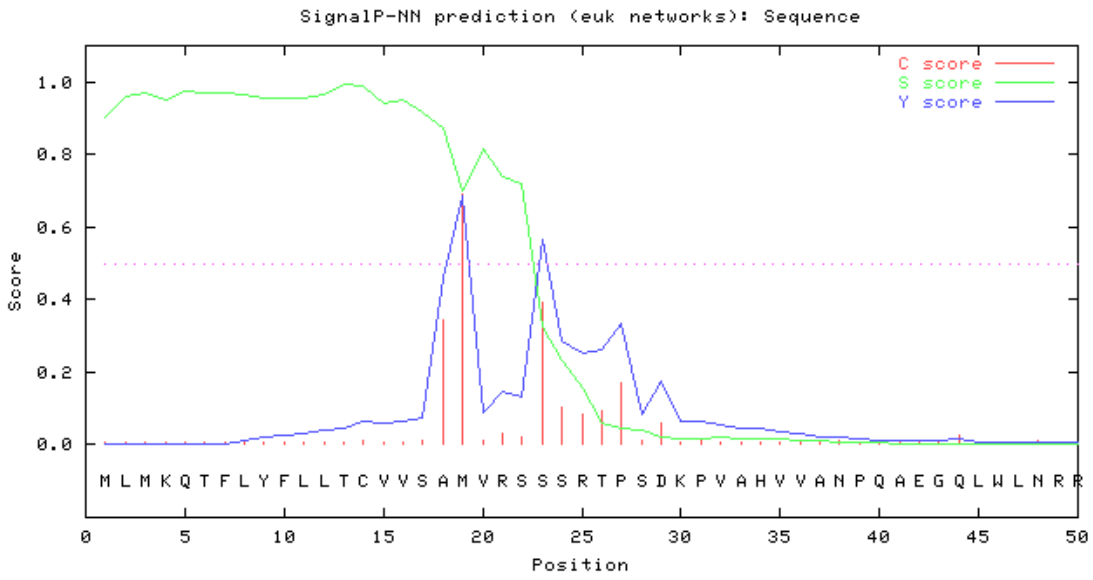
The SignalP 3.0 prediction server at Centre for Biological Sequence Analysis at DTU was used to evaluate the N-terminus of the Cpy-TNF-α fusion (Bendtsen et al., 2004). This server predicts the signal sequence of classically secreted proteins and the cleavage hereof through the analysis of the N-termini protein sequence. This predictor uses both Neural Networks and Hidden Markov Model algorithms, and is among the most precise methods for protein signal sequence prediction which reach accuracy above 80% (Klee and Ellis, 2005). The N-terminus of the Cpy-TNF-α fusion protein was submitted to the SignalP prediction server (Figure 23).

The first output is from the neural network (Figure 23A). This algorithm gives 3 outputs in a graphical form; The C-score assigns the ``cleavage site'', which is significantly high only at the cleavage site. The S-score is the predicted signal sequence. Y-max is a derivative of the C-score combined with the S-score ensuing in a better cleavage site prediction than the unprocessed C-score, since multiple high-peaking C-scores can be found in one sequence only having one factual cleavage site. The cleavage site is designated by the Y-score at the point where the slope of the S-score is steep and a significant C-score is found. The prediction from the neural network of the Signal 3.0 server gives a most likely cleavage site between residue 18 and 19. These residues constitute the border between the Cpy-signal sequence and the N-terminal of the TNF-α molecule, indicating that the fusion has not destroyed the features of termination sequences and that cleavage should occur as expected. There is, moreover, also a rather high Y-value between the residues 23 ands 24.

The Hidden Markov Model is a statistical model describing the probability of distribution between different states, which can be used for pattern recognition applications. The signal peptide model contains sub-models describing each of the 3 distinct sequential regions of a signal sequence: the N-terminal region, the hydrophobic-region, and the C-terminal-region. Furthermore, a likely cleavage site is assigned a probability score. This prediction algorithm describe the most likely cleavage site is between residue 23 and 24, but also here a cleavage site between 18 and 19 can be predicted. Thus, both the site between residue 18 and 19 and the one between 23 and 24 are possible cleavage sites according to the Signal 3.0 server. This gives a possibility that the signal sequence be elongated by 5 amino acids, when fusing Cpy and TNF-α giving a shorter hybrid protein. Alternatively, a subset of the protein is cleaved at one of the positions and another subset at the other.

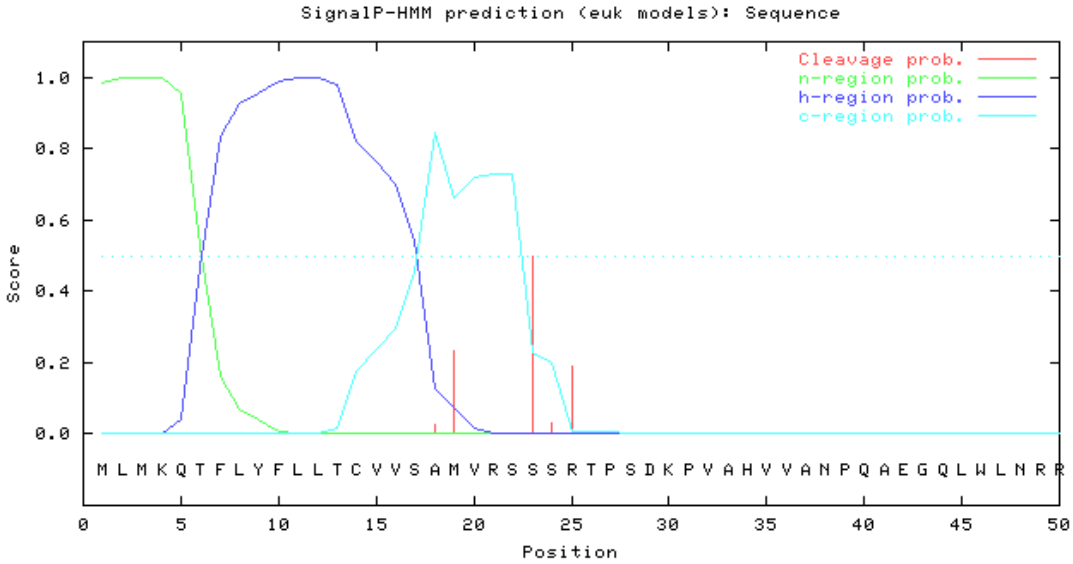
The difference between the outputs of the two models is due to the fact, that two different algorithms are employed using different input parameters. The prediction is also more tricky, since it is a fusion protein that is been analysed. This means that information from both the

Cpy signal sequence and the TNF-α combined in the output scores. What cleavage site that is used *in vivo* can be determined empirically by N-terminal peptide sequencing. But importantly, there is a strong prediction of a signal sequence and a cleavage site to direct the protein to the secretory pathway. This is confirmed by the experimental data showing that the Cpy-TNF-α protein is expressed and secreted (Kjaerulff and Jensen, 2005).




```
# Most likely cleavage site between pos. 18 and 19: VSA-MV
```

B




```
Prediction: Signal peptide
Signal peptide probability: 0.986
Signal anchor probability: 0.012
Max cleavage site probability: 0.499 between pos. 22 and 23
```

Figure 23 Output data from the SignalP 3.0 signal sequence prediction server
A. The N-terminal of the CPY-TNF-α fusion protein was submitted for signal sequence analysis.
The output from the neural network of SignalP contains 3 different scores, in a graphical form: the C (cleavage sites), S (signal peptide), and Y (combined) score. The scores are reported for each single amino acid position in the submitted sequence.
B. The Hidden Markov Model calculates the probability of whether the submitted sequence contains a signal peptide or not. The signal peptide model contains sub-models that describe each of the 3 distinct sequential regions of a signal sequence: the N-terminal region, the hydrophobic-region, and the C-terminal-region. A likely cleavage site is assigned a probability score.

Library screening

In order to obtain selected folding competent proteins 3 individual screens were done in the Py-17 *Sz. pombe* strain, using the library described above inserted. In the 2 first screens different obstacles in the set-up caused abortion of the ongoing screen. The experiences obtained were implemented to the following selection round.

First round of selection

Newly transformed *Sz. pombe* cells are in general known not to survive on sucrose media. Therefore, transformed cells were plated out on glucose plates and two different methods for screening for invertase activity known from the literature were tested. One is replica-plating from glucose to sucrose (or sucrose containing antimycin, an inhibitor of the respiratory-chain (Jacobs et al., 1999), to enhance the selection pressure), and the other is an invertase overlay assay using red/white colour screening (Paravicini et al., 1992; Johnson et al., 1987). However, much background was seen in these approaches, so the positive colonies, having a plasmid encoding a foldable sequence, could not be distinguished from negative colonies encoding non-foldable sequences.

In the growth assay at least 1/3 of the colonies should not have grown solely on the basis of containing stop codons in the insert sequence, and hence the invertase moiety will not be expressed. Therefore, it is obvious that a degree of background growth is present. This

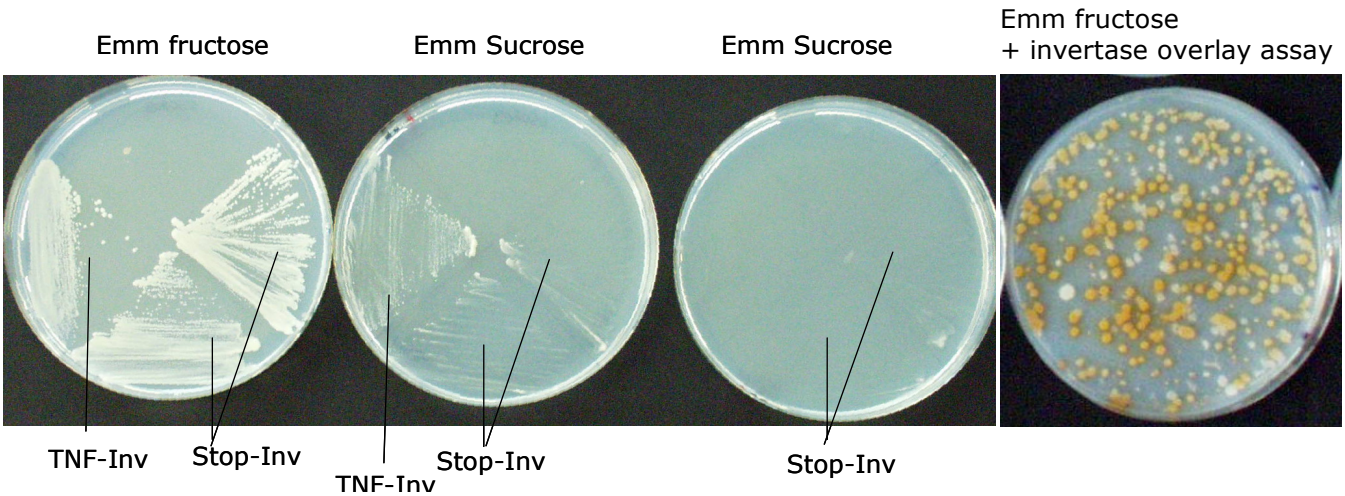


Figure 24. The cross-feeding of the growth assay and the reaction in the invertase assay are shown.

Growth assay plate 1-3: TNF-inv: The PY-17 strain transformed with Cpy-TNF-α-inv plasmid possesses invertase activity. Stop-inv: The PY-17 strain transformed with Cpy-STOP-invertase encoding plasmid does not possess invertase activity. On the first plate is Emm-fructose media giving no selection, both invertase positive and invertase negative strains are expected to grow. The second plate is Emm-sucrose selection media. Therefore, only positive cells should sustain growth. But as seen, the invertase negative strain also grew. The closer to the positive cells the more profound the growth. If however, no invertase positive cells have been streaked onto the plate, no growth is seen in the in the invertase negative cells.

Invertase assay plate 4: PY-17 cells transformed with the library containing screening plasmid and subjected to invertase overlay assay. As seen, about 2/3 of the colonies turn red. There is some differentiation between how fast the colour develops in the individual clones, but the development of colour is completed only in seconds in all cases.

background can derive from cross-feeding of glucose and fructose diffusing from positive cells, making all colonies invertase positive, as well as negative, multiply on the selective sucrose plates. To test the hypothesis of cross feeding, the growth among negative cells growing in proximity to positive cells was displayed (Figure 24, plate 1-3), using standard plasmids. From this test it was apparent that positive cells did sustain growth of negative cells.

In the invertase assay 2/3 the colonies developed colour corresponding to the amount of colonies without stop codon in the insertion sequence (Figure 24, plate 4). Thus, the sensitivity of the invertase assay seems too high for the assay to be useful in this context.

Second round of selection

To circumvent the problems from the previous screen, a new approach using growth on limiting glucose was developed. The next round of selection was done by growing transformants on solid EMM sucrose media without glucose or with a limiting concentration of glucose. The rationale behind the addition of glucose to the sucrose solid media is that the limited amount of glucose allows newly transformed colonies to survive on the sucrose media. The glucose in the media is soon exhausted, and after this point, only clones containing a folded protein-variant, and therefore invertase activity, are capable of continued growth. The concentrations tested were 2% sucrose containing 0%, 0.1%, and 0.01% glucose (Figure 25). On 0.1% of glucose all colonies grew just as well as the ones on glucose plates. But on the 0.01% glucose 20 out of 600 colonies grew (3 %). Surprisingly, 2 colonies appeared on the sucrose plates.

Colony PCR was preformed on the positive clones in order to determine the sequence of the TNF-α variant present in the colony. However, when attempting to sequence the insert most clones yielded no readable sequence, even though getting a DNA fragment of the right size from the colony PCR was possible. To investigate this, the DNA fragments from the separate selected colonies were cloned into the TOPO vector and transformed into *E. coli* cells. For each transformation a single *E. coli* colony represented a different plasmid in the yeast cell. From each set of the transformation 4 *E. coli* clones were picked and sequenced. The number of different variants per yeast colony varied from 4 out of 4 colonies to a single variant out of 4. The multiple plasmids in each cell/colony are likely to derive from the transformation where a transformation competent cell can uptake several plasmids. And since the plasmid used is of the multi-copy type, it is possible that plasmids encoding variants not able to sustain invertase activity are co-transformed and segregate along with those who are.

Plasmids are easily lost from the *Sz. pombe* cells, even under selection pressure, and it is estimated that up to 50% of the cells in a culture in liquid media are not holding any plasmids

(Hayles and Nurse, 1992). Therefore, the answer for reducing the multiple plasmids to one was to re-streak the selected colonies twice on sucrose media. Hereafter, only one variant was found in each colony displayed by sequencing.

After re-streaking and performing liquid invertase assay, 4 of the 22 of the colonies turned out to be false positives. Therefore, this selection round was aborted, but the use of limiting glucose in sucrose media was further investigated, and a double re-streaking was implemented in a new round of selection.

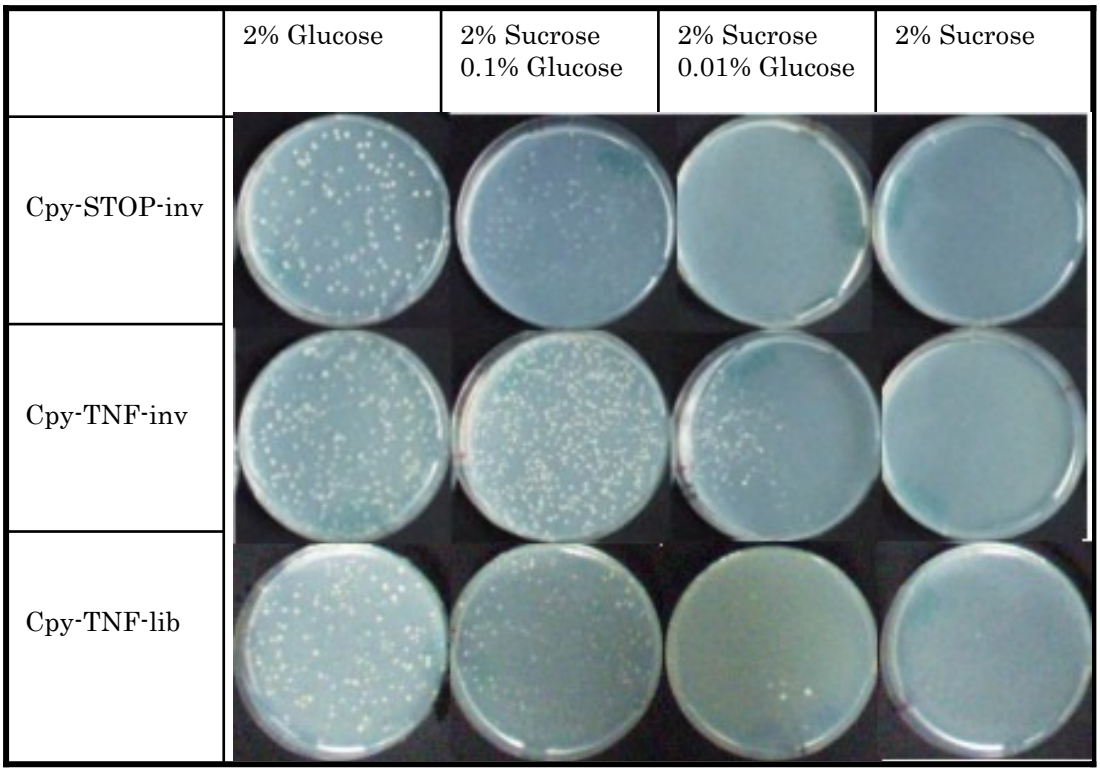


Figure 25. Optimisation of the glucose concentration in the screening plates.
The Py-17 *Sz. pombe* strain transformed with the screening plasmids either containing a STOP codon, TNF-α, or library fused to the invertase gene was plated out on Solid Emm-media with different concentrations of glucose.

Optimisation of the glucose concentrations using control proteins

First, more glucose concentrations in the sucrose plates were tried out (0.01%, 0.02%, or 0.03%). The test was done using TNF-α protein fused to invertase, but also the RANKL fused to invertase was included in the test (Figure 26). RANKL is used in an AutoVac project at Pharmexa, belongs to the TNF-α superfamily, and is demonstrated folded/expressed in *E. coli*, and was used inorder to investigate whether different proteins display different growth patterns on selective media.

The *Sz. pombe* strains containing the TNF-α or RANKL encoding plasmids grew on the selected concentrations, also on sucrose media. No differences were noted between the

proteins. Notably, the cells with empty vector only grew on glucose media and none on the selective media. At the glucose concentrations 0.01%, 0.02%, and pure sucrose, a reduced number of colonies was seen and at the concentration 0.03% as many colonies appeared as on glucose plates. The fact that all colonies grew at the concentration of 0.03% does not mean that this concentration is too high for selection, as all the transformed cells contain plasmids giving a positive phenotype. More important is that there was no growth at this concentration on the plates with an empty vector. In all, the results indicate that the right range of glucose concentrations for screening had been found.

It is apperent that an there is an uneven distribution of colonies on glucose concentrations at 0.01%, and 0.02% and sucrose. It is not quite clear what causes this, but a hypothesis is that some colonies have a higher invertase activity than others (maybe because of a higher plasmid copy number in these colonies) and, therefore, grow relatively faster than the majority of the colonies. These cells then drive the neighbour colonies to grow faster due to cross-feeding with glucose. In this context it should be noted that all the clones on these plates contain identical plasmids, all with folding competence and hence putative invertase activity. Therefore, this tendency may not be a problem when screening a library. Alternatively, subtle differences in the environment on the plates, e.g. temperatures or moistness result in the differences in growth rate.

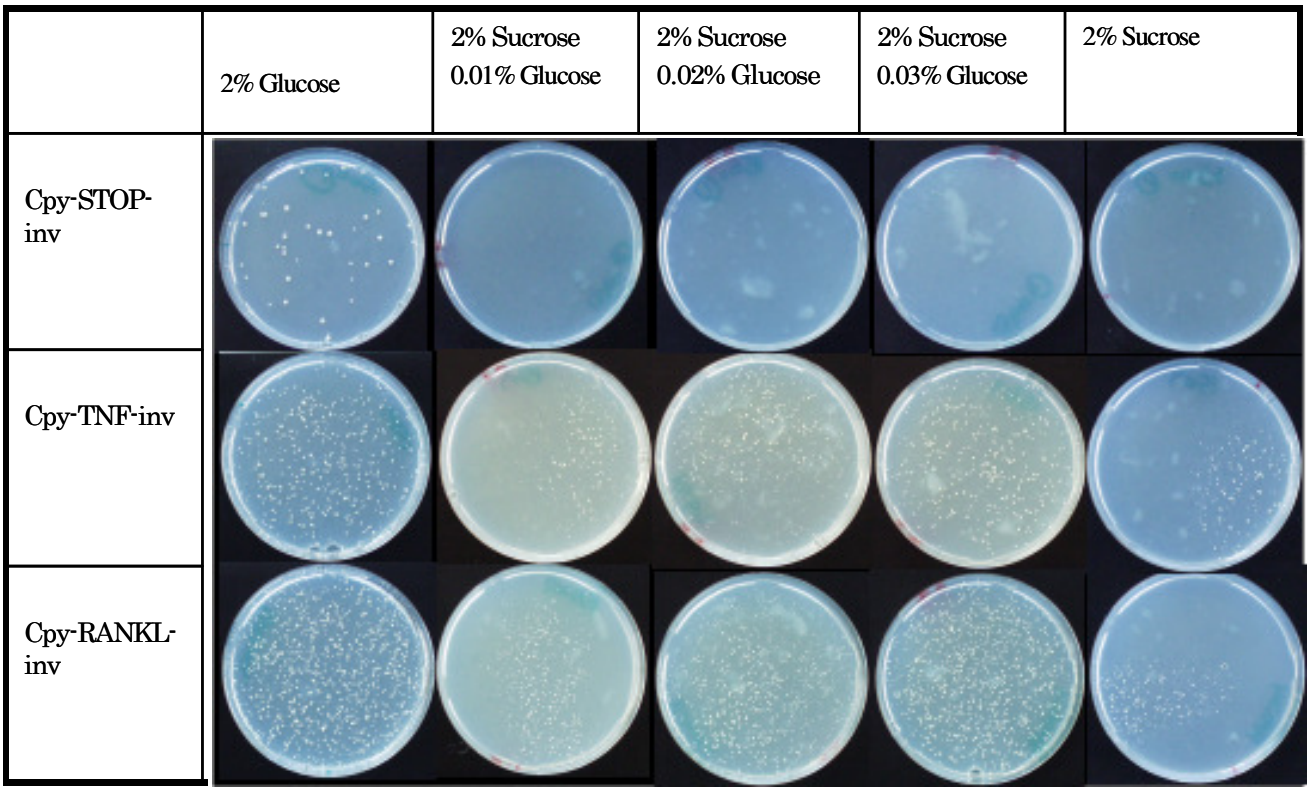


Figure 26. Optimisation of the glucose concentration in the screening plates.
The Py-17 *Sz. pombe* strain transformed with the screening plasmids either containing a STOP codon, TNF-α or RANKL fused to the invertase gene was plated out on Solid Emm-media with different concentrations of glucose

Third round of selection

The experiences of the 2 preceding screens and the glucose optimisation were applied to the last and successful screen. Consequently, *Sz. pombe* cells transformed with the library were spread out on 2% sucrose plates with the addition of 0%, 0.01%, 0.02%, or 0.03% glucose (Table 2). 3 plates of each glucose concentration were used. Furthermore, one plate of 2% glucose was used in order to estimate the number of transformed cells.

4000 colonies appeared on the glucose plate after 4 days. The selection plates were examined daily. No colonies appeared on the sucrose plates. In total 18 and 7 colonies appeared, at the concentrations 0.01% and 0.02% glucose, respectively and were picked at day 7 and 8. Growing colonies were picked from the selection plates as they appeared, in order to reduce growth of satellite colonies due to cross feeding. At the concentration of 0.03 % glucose 1068 colonies (10%) grew at day 7, and it was suspected that this concentration of sucrose was not selective enough. Therefore, these colonies were replica plated onto pure sucrose media and the fastest growing colonies were picked the following day. In total, 10 colonies were picked.

After the colonies were picked, they were re-streaked twice on pure sucrose media to further refine the selection and to avoid having multiple plasmids in each colony, as observed in the second screen. After this a total of 25 colonies were able to grow on sucrose media.

Glucose concentration	0.01%	0.02%	0.03%	Total
Number of screened colonies	12000	12000	12000	36000
Positive colonies before re-streak	18	7	(1068) 10	35
Positive colonies after re-streak	12	5	8	25
Positive colonies after rescue	11	5	7	22

Table 2

The table shows the total number of colonies screened, the number of invertase positive of colonies after the primary selection, and number of positive colonies after double re-streaking and plasmid-rescue. The number in the parentheses is the number of colonies before replica plating at the glucose concentration 0.03%.

Plasmids were rescued from the selected colonies and re-transformed into the *Sz. pombe* screening host PY-17. This was done in order to conclude whether one particular plasmid in selected TNF-α colony, whose sequence is later deduced, is in fact that which cause a selectable phenotype. After the initial growth on glucose media, a liquid invertase assay was performed on the re-transformed colonies in order to make a quantitative measurement of sucrose hydrolysis (Figure 27). Wt-TNF-α was used as positive control and variant 59 and 60 as negative control. These negative controls derive from the non-selected randomly picked clones from the library. The negative controls do not grow on sucrose plates and importantly

do not have a stop codon in the insertion sequence. In other words, these variants are bearing insertions at non-permissive sites (for sequence see figure 30).

Wt-TNF-α displayed activity in the liquid invertase assay, variant 60 did not display any activity, and variant 59 displayed a low level. All the selected plasmids were able to complement invertase activity in Py-17, except for three (clone 1, 20, and 22). These results were complemented by the sucrose growth assay (data not shown). Presumably, this means that the three colonies initially isolated, might have contained more than one plasmid after screening and re-streaking, and that the plasmid rescued is not the one responsible for the invertase positive phenotype.

From the data it is apparent that the clones exhibit different levels of invertase activity. This assay has been reported to be quantitative (Moreno et al., 1985); however, previous experience has proven that the levels vary from experiment to experiment (data not shown), and therefore no interpretation is placed on the level of activity. However, the three negatives have turned up as negative respectively.

The reason for the poor reproducibility of the liquid invertase assay is illusive but can be a sign that more optimisation is needed to achieve consensus among different analogues experiments. Though, it can also represent that differences in plasmid copy-number or growth conditions influence the outcome of this assay.

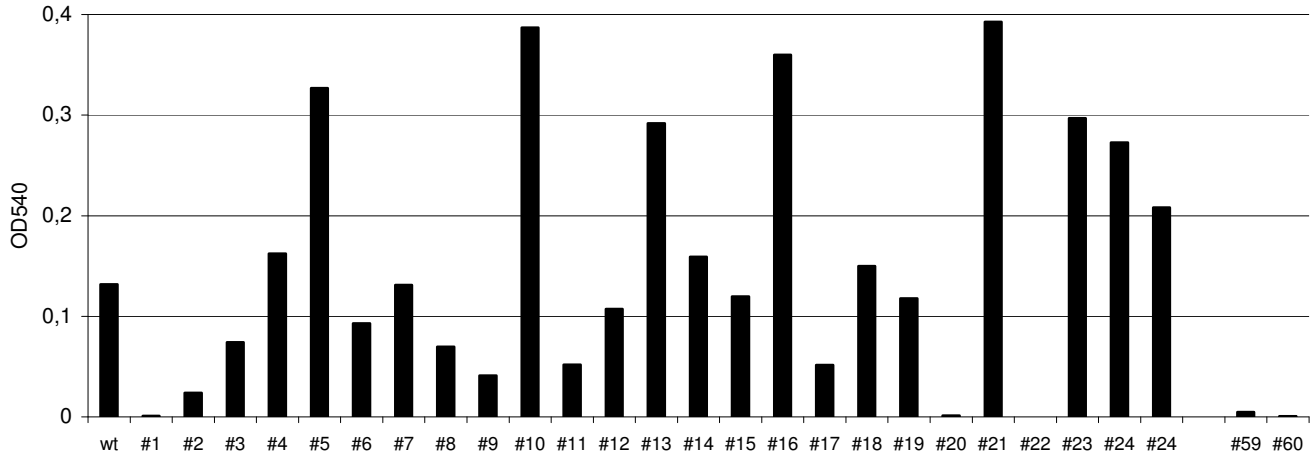


Figure 27. Liquid invertase assay on re-transformed colonies.
The retransformed colonies were grown in Emm-fructose media ON and the normalised to OD_{600} =1.5. The wt-TNF-α is included as positive control and two negative controls, which is variants from the randomly picked clones before selection (variant 59, and 60).

Analysis of insertion point and sequence of the selected variants

After having completed the screening, the plasmids from the positive colonies were sequenced to identify the insertion point and the sequence of the inserted peptide (Figure 30).

The linear positioning of the insertion sites

The distribution of the site of insertion prior to and after selection is examined and several differences are apparent (Figure 28). Firstly, the even distribution along the TNF-α sequence seen before the selection has shifted to a distribution largely focused in the N-terminal of the target protein - especially before the first β-strand, where 10 of the insertion points are situated. 14 of the 22 insertions are found between residue 1 and 35. In the middle part of the protein between aa 59-119, no insertions are found. In the C-terminal 5 insertion sites are found, 2 of which are appended at the C-terminal.

Secondly, before selection, no discrimination of insertion into secondary structure elements is seen. However, the insertion sites after selection are found primarily outside of β-strands, with the exception of variant 11 where the insertion is situated in the middle of β-strand number 8. Thus, after selection, the insertions are mostly situated in areas of the protein, which convincingly can, from a structural point of view, accept insertions.

And thirdly, before selection a maximum of two insertions was found per site. After selection, up to 5 insertions are found, at a single position between proline-13 and valine-14.

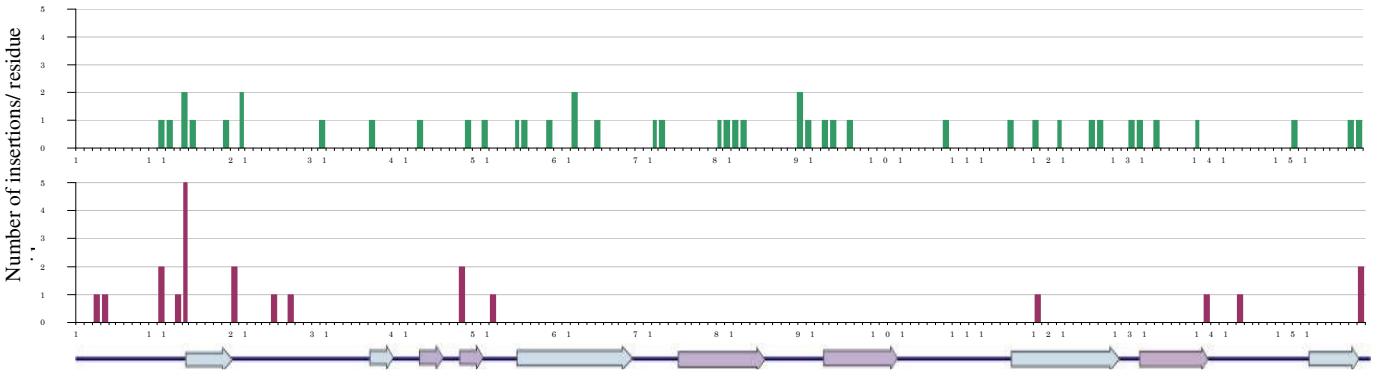


Figure 28. Diagram depicting the distribution of insertion sites prior to and after screening
The distribution of insertion sites in the amino acid sequence prior and after screening is shown. The barr represent an insertion site and the height the number of insertions found. The representation does not discriminate between the different reading frame and orientations, so each entry can represent any of the 6 different insertions sequences. The colouring of the illustration of the secondary structures below the diagram corresponds to that of figure 12.

The spatial positioning of the insertion sites

When looking at the spatial positioning of the insertions (Figure 29), it is apparent that the insertion sites fall into three categories; in the receptor-binding domain in the bottom of the bell-shape of the trimer, in three up-wards facing loops on the exterior of the lower half of the

trimer, and in the interface between the molecules in the trimer (for description of the constitution of TNF-α. See introduction page 45).

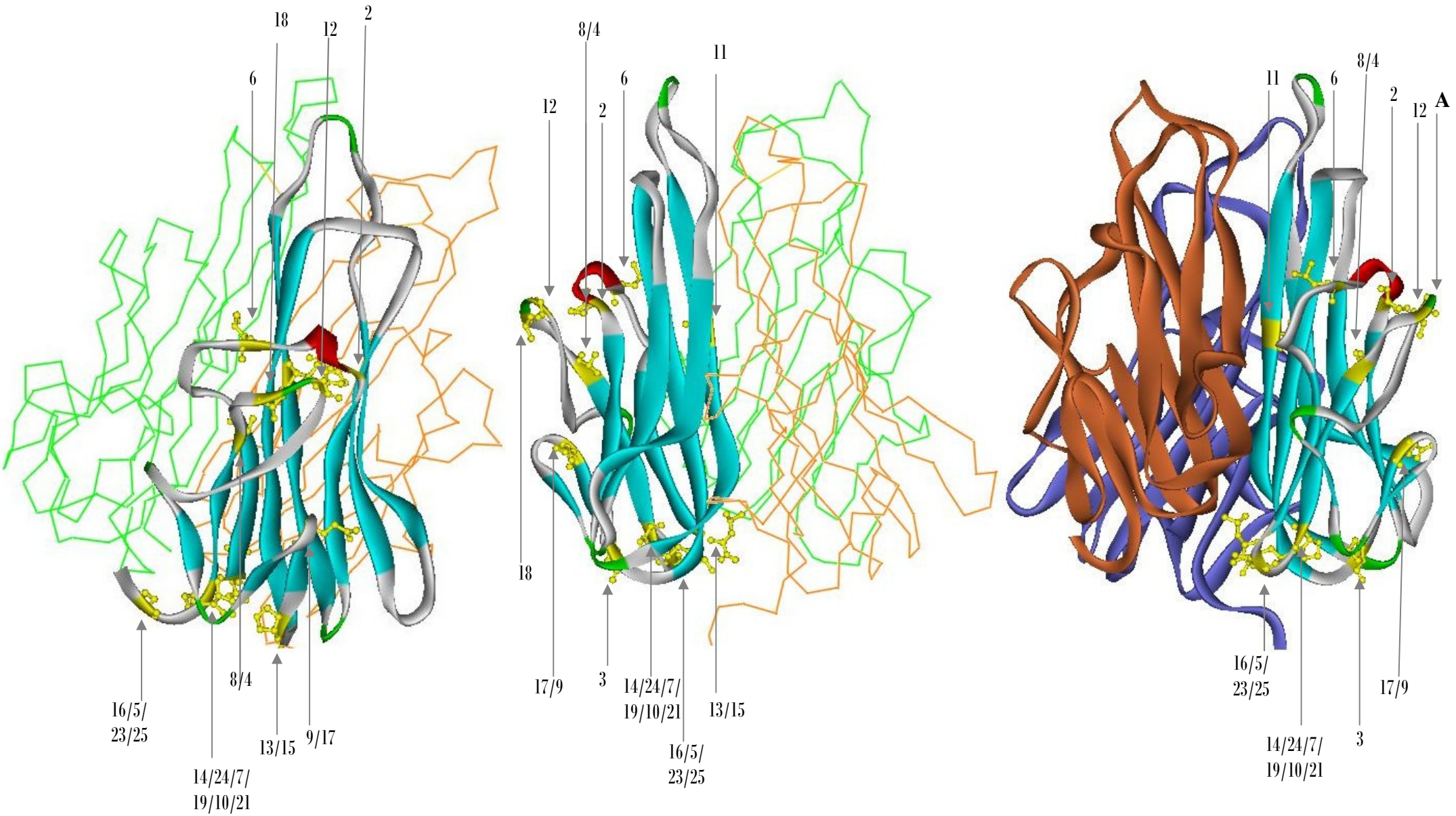
The receptor-binding domain comprises the N-terminus and the C-terminus together with the loop between β-strands B and C, and the loop between β-sheet F and G (see figure 12). So, the insertions at the N-terminal (variant 16, 5, 23, 25, 14, 24, 7, 19, 10, and 21), the loop between β-strands B and C (variant 3) and the 2 insertions in the very C-terminal of the molecule (variant 13 and 15) are accommodated here.

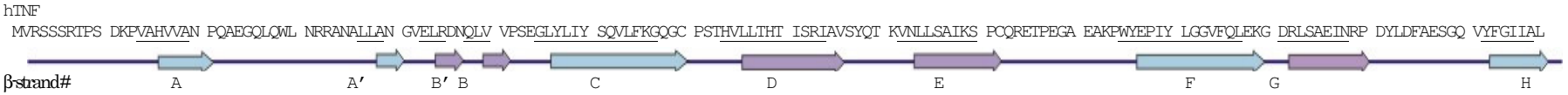
The three upward-facing surface orientated loops in which there are found insertions are the lower loop between B' and B (variant 9 and 17), the second lowest loop between A and A' (variant 4, 18, and 12), and the third lowest loop between B' and B (variant 2 and 6). Notably, no insertions were found in the loop between β-sheet E and F, which is the long loop situated at the top of the molecule where the insertion of Pharmexa's current AutoVac candidate molecule is found.

One of the insertions is exceptional (variant 11) in the respect that it is found both in a β-sheet and at the interface between the subunits in the trimeric molecule and thereby buried regions of TNF-α pointing towards the interfaces of the subunits in the trimer. This is controversial when using a rational view on protein folding.

Figure 29 Graphic 3D representation of insertion points in the TNF-α trimer (opposite page).

The image illustrates the insertion points of the peptide insertions after selection. The insertion points in the selected variants are indicated as yellow areas and sticks and balls in the accentuate subunit. The number of the variant is indicated by an arrow. If there is ore than one insertion at a given point (both identical or in different reading frames), all the variants numbers are indicated in the illustration, separated by slashes.
The images are identical, but rotated in relation to each other. One subunit is displayed in grey with indications of secondary structure in colour (red: helixes, green: β-strands), the two other sub-units of the trimer is shown as wire frame or uni-colour ribbons.


18
12
2
6
8/4
16/5/
23/25
13/15
9/17
14/24/7/
19/10/21
8/4
11
6
12
2
18
17/9
3
14/24/7/
19/10/21
13/15
16/5/
23/25
11
8/4
6
2
12
A
16/5/
23/25
17/9
14/24/7/
19/10/21
3



Possible insertions

AAAKFVAAWTLKAAA, AAALRVQAATNLAAA, RPLSSLQLGPLRRP, CGRLKGPSCNELSGRT, CGR.VRCSLDP.GGRT, RPP.GSKLQRT.RPH

Sequence of Selected clones

16.MVAAALRVQAATNLAAAMVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
5.MVRAAALRVQAATNLAAAVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
23.MVRSSSRTPSAAALRVQAATNLAAAPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
25.MVRSSSRTPSAAAKFVAAWTLKAAAPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
14.MVRSSSRTPS DKAAALRVQAATNLAAADKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
24.MVRSSSRTPS DKPVAAAKFVAAWTLKAAAPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
7.MVRSSSRTPS DKPVAAAAKFVAAWTLKAAAVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
19.MVRSSSRTPS DKPVCGRLKGPSCNELSGRTVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
10.MVRSSSRTPS DKPVAAAALRVQAATNLAAAVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
21.MVRSSSRTPS DKPALRPLSSLQLGPLRRPQVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
8.MVRSSSRTPS DKPVAHVVACGRLKGPSCNELSGRIAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
4.MVRSSSRTPS DKPVAHVVAAAKFVAAWTLKAAAVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
18.MVRSSSRTPS DKPVAHVVAN PQADAAAKFVAAWTLKAAAAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
12.MVRSSSRTPS DKPVAHVVAN PQAEGAAAKFVAAWTLKAAAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
17.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQCGRLKGPSCNELSGRNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
9.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQCGRLKGPSCNELSGRNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
3.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPLRPLSSLQLGPLRRPQPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
11.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIAAALRVQAATNLAAAPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
2.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRPVRPLSSLQLGPLRRPQP DYLDFAESGQ VYFGIIAL
6.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDCGRLKGPSCNELSGRIDFAESGQ VYFGIIAL
13.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL GAAALRVQAATNLAAAL
15 MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL CGRLKGPSCNELSGRTL

Sequence of Clones Negative after rescue and retransformation

1. MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRCGR.VRCSLDP.GGRSRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
22.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KCGRLKGPSCNELSGRT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
20.MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALCGRLKGPSCNELSGRTLLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL

Randomly Picked Library variants

59 MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIY LGAAAKFVAAWTLKAAALGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL
60 MVRSSSRTPS DKPVAHVVAN PQAEGQLQWL NRRANALLAN GVELRDNQLV VPSEGLYLIY SQVLFKGQGC PSTHVLLTHT ISRIAVSYQT KVNLLSAIKS PCQRETPEGA EAKPWYEPIAAAKFVAAWTLKAAAPIY LGGVFQLEKG DRLSAEINRP DYLDFAESGQ VYFGIIAL

Figure 30. The DNA sequence of all the selected variants. The top panel illustrates the secondary structure of TNF-α, where the arrows below show the β-sheets. The colours are identical to the ones used in figure 12. The amino acid sequence of the selected variants, the 3 variants failing rescue and 2 non-selected variants, used as negative controls are shown. The unselected variants are chosen as they contain PADRE, but fails to induce invertase activity. The underlines amino acids represent β-sheets and the coloured fragment inserts. PADRE inserts are green and others are yellow.

Reading frame and orientation

When analysing the insertion sequence frequencies in terms of reading frame and orientation, a change in distribution is noted relative to prior selection (Table 3). More sequences are found in RF3-r than before selection, whereas the number in RF1-r is reduced. The number of selected variants is too small to establish whether this is a significant trend. Logically, however, some of the sequences could be more acceptable insertions in the protein than others, as they differ in e.g. hydrophobicity/phility, charge, and volume as a consequence of the differences in the amino acid composition. Thus, it is not unexpected to observe a change in distribution before and after selection.

Frequency of different insert sequences	Resulting amino acid insert sequence	Frequency prior screening	Frequency after screening
RF1-d (PADRE)	AAAKFVAAWTLKAAA	6 (22%)	6 (25%)
RF1-r	AAALRVQAATNLAAA	10 (37%)	7 (29%)
RF2-d	RPLSSLQLGPLRRPQ	4 (14%)	3 (13%)
RF2-r	LRPP*GSKLQRT*RP	-	-
RF3-d	CGR*VRCSLDP*GGR	-	-
RF3-r	CGRLKGPSCNELSGRT	7 (25%)	8 (33%)
Total		27	24

Tabel 3. The frequencies of the different translated insert sequences found in the selected variants.

The position of the different insertion sequences

Not only is the position of the insertions interesting, but also the nature of the different translated sequences. Firstly, as expected, no stop codons are present after the screen, whereas, prior to screening, they constituted approximately 40% of the analysed clones and theoretically 33% of un-screened clones. By examining the position of the different insertion sequences in the target protein (Figure 31), it becomes apparent that between the amino acids proline-13 and valine-14, right before the first β-strand, many different insertion sequences can be accommodated without leading to instability of the TNF-α protein. All the 4 possible sequences are present here.

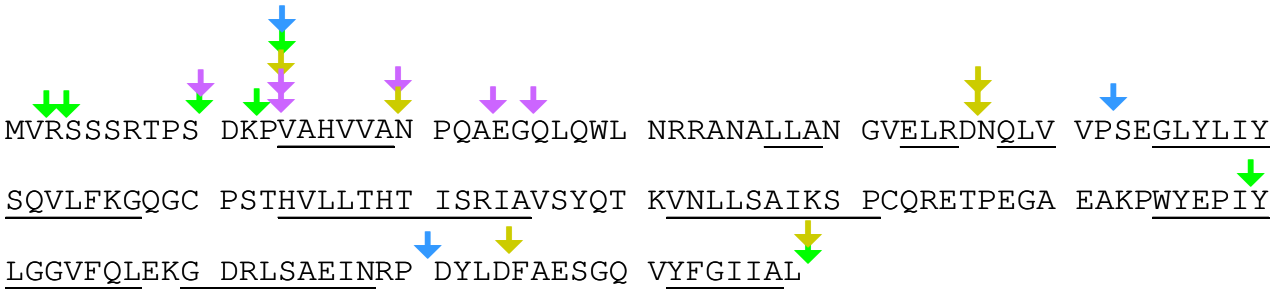


Figure 31. Insertions points in the selected variants
Insertions points are shown as arrows in the primary structure of TNF-α. The 4 different colors represent the 4 possible amino acid insertion sequences.
AAAKFVAAWTLKAAA, AAALRVQAATNLAAA, RPLSSLQLGPLRRPQ, CGRLKGPSCNELSGRT

In contrast, at the position between the amino acids glutamine-47 and leusine-48 (found between the β-strand B' and B) insertions with the sequence resulting from translation of RF3-r were found twice. Likewise, at the first 4 amino acid positions at the N-terminus, the sequence resulting from translation of RF1-r was found twice, and in the loop between β-strand 1 and 2 the sequence translating from RF1-d was found twice, indicating that at these positions only a particular sequence might be acceptable insertions.

Variants containing the insertion sequence encoding the PADRE epitope, and, therefore, likely vaccine candidates, have been found 6 times at 5 different insertion sites, all situated in the N-terminal region of TNF-α.

Recombinant Expression of selected TNF-α variants in 3 systems

Recombinant Expression of selected TNF-α variants in *Sz. pombe*

The 5 variants containing the PADRE epitope as insert (variant 4, 7 12, 18, and 25) are possible AutoVac candidates. They were chosen therefore to verify that the selected proteins can fold and be expressed in *Sz. pombe,* without being fused to the invertase moiety. If the variants can be expressed, it will demonstrate that invertase does not drive or influence folding. Thus, the cDNA sequence of these TNF-α variants was amplified by PCR and reintroduced into the screening vector deleted for the invertase gene. As negative controls, variants (59, and 60) were used. These variants do not have the capacity for invertase activity as fusion protein, i.e. bearing insertions at non-permissive sites. The wt-TNF-α was included as positive control and for investigating the relative expression level between wild type and variants.

Upon expression the TNF-α variants will be secreted into the medium due to the presence of the Cpy-signal sequence. Accordingly, the cell culture supernatants were tested using a specific TNF-α sandwich ELISA assay for quantification of the protein level of each variants. The expression level was normalised relative to the OD_{600} value of the single cultures at the point of harvest (Figure 32).

The induction time after thiamine repression of the promoter is long, and as the data below indicate, longer than the 16 hr described in the literature (Siam et al., 2004), therefore, the expression without thiamine was also tested to see if an increased expression could be

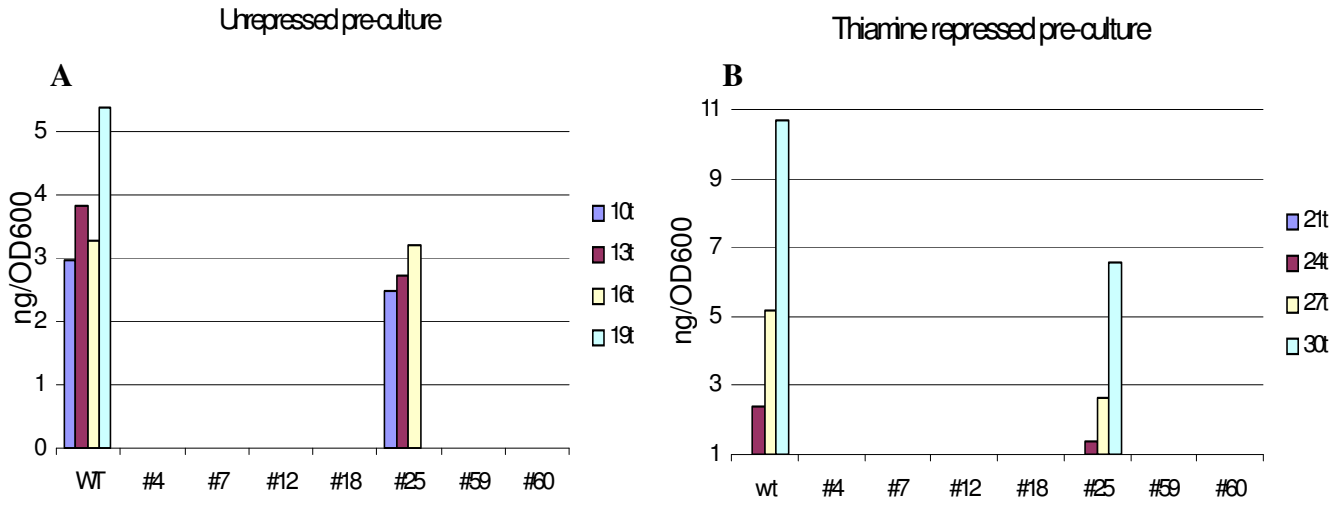


Figure 32. Expression of wt TNF-α and AutoVac variants in *Sz. pombe*.
Two rounds of expression are shown at several time points. A) pre-culture grown in standard EMM-glucose medium and B) pre-culture grown in EMM-glucose medium with the addition of 15 ug/ml thiamine. There are no data available for variant 25 at 19 hr in un-repressed media

obtained by constitutive expression.

The results show that only wt-TNF-α and variant 25 displayed any protein detectable protein levels in the expression medium. wt-TNF-α had a protein level at 5.5 and 10.5 ng/OD_{600} at its highest in unrepressed and thiamine-repressed culture, respectively. The Variant 25 showed a comparable expression level at 3 and 6.5 ng/OD_{600}. As seen, the promoter preformed most efficiently in the presence of thiamine.

The lack of detectable protein the other of the selected variants can be due to a lack of folding, and hence, expression, but also that the expressed proteins were below the detection level of the assay. It is known that *Sz. pombe* displays large protease activity in the growth media (Kumar and Singh, 2004; Tokunaga et al., 1993; Kjaerulff and Jensen, 2005), making the proteins liable to degradation, and detection made even more difficult. Therefore, the expression was also tested in *E. coli* and *Drosophila* cells.

Recombinant Expression of selected TNF-α variants in *E. coli*

Since not all of the selected variants could be expressed in *Sz. pombe,* the 5 PADRE variants, wt-TNF-α, and the negative controls, variant 59 and 60 were tested for soluble expression in *E. coli* (Figure 33). The proteins were expressed from the pET28 vector in BL21-star cells and tested in the ELISA assay. A pET28 vector coding for RANKL is used as control for cross reactivity of *E. coli* protein in the ELISA. This control displayed no colour reaction in the ELISA. The wt-TNF-α showed an expression level at 1000 ng/ml and variant 25 a level at 1500 ng/ml. Also, variant 7, 12, and 18 displayed expression levels at 11 ng/ml, 9 ng/ml, and 5 ng/ml, respectively. Variant 4 displayed a considerable lower expression level at 0.5 ng/ml. The controls displayed no expression. Thus, all of the selected variants showed protein expression. However, expression of variant 4 was very low, 2000 times lower than that of wt-TNF-α. Also, for the variants 7, 12, and 18 the expression was 100 times lower than that of the wt-TNF-α.

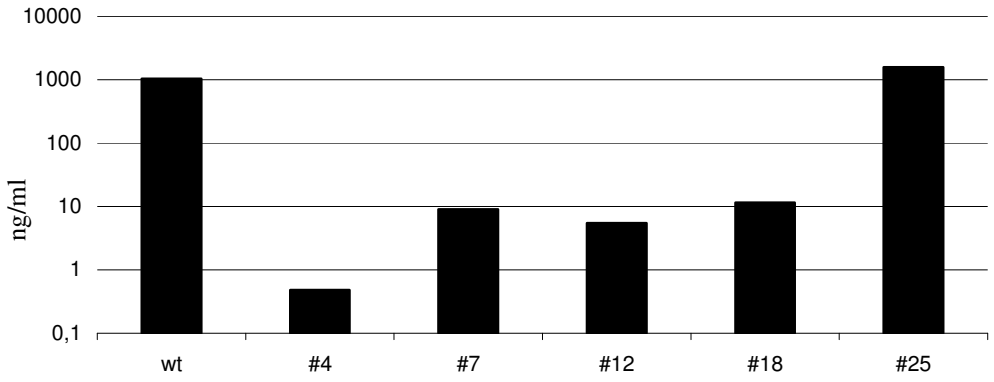


Figure 33. Expression of wt TNF-α and 5 AutoVac variants in *E. coli* Bl21-star. The values are displayed on a semi-logarithmic scale due to the large differentces between the values

Recombinant Expression of selected TNF-α variants in *Drosophila*

The last expression system in which the selected proteins were tested was *Drosophila* cells. This expression system is an important in-house system at Pharmexa, proven very successful for expression of human proteins. The variants were expressed transiently in S2 *Drosophila* cells from the p2Zop2F plasmid and secretion was supported by the BiP signal peptide. The media were tested with the TNF-α ELISA (Figure 34)

All selected proteins were tested for detectable levels of protein in the media in ELISA assay. The wt-TNF-α showed a protein level at 2.1 ug/ml, and variant 25 an even higher level at 2.4 ug/ml. The variants 7, and 12, expressed 1.4 ug/ml and the variant 18 1.2 ug/ml. Also, in this expression system, variant 4 showed a low, but detectable level of protein at 0.1 ug/ml. None of the negative controls showed any TNF-α expression.

Further, the expression levels of the variants were tested by western blotting. 25ul sample was loaded in each lane, which correlates to approximately 50 ng of wt-TNF-α according to the ELISA assay. To validate the results from the ELISA, 50 and 100ng wt-TNF-α from a standard preparation of protein were also loaded on the gel. As seen, the protein level at wt-TNF-α and variant 25 corresponds to the expected 50ng, the levels from the variants 12 and 18 are less than this, and the levels seen for variant 4 are barely detectable. Theses results validate the data obtained by ELISA assay. Surprisingly, the protein level of variant 7 is at the same level as wt-TNF-α and variant 25, eventhoug the ELISA results indicate a lower expression level of this variant. As expected, none of the negative controls displayed any TNF-α protein.The results from the western blot confirm the protein levels measured by the ELISA for all variants, but variant 7. Thus, the ELISA might generally underestimate the protein level of variant 7. Furthermore, no degradation of the expressed proteins is detected.

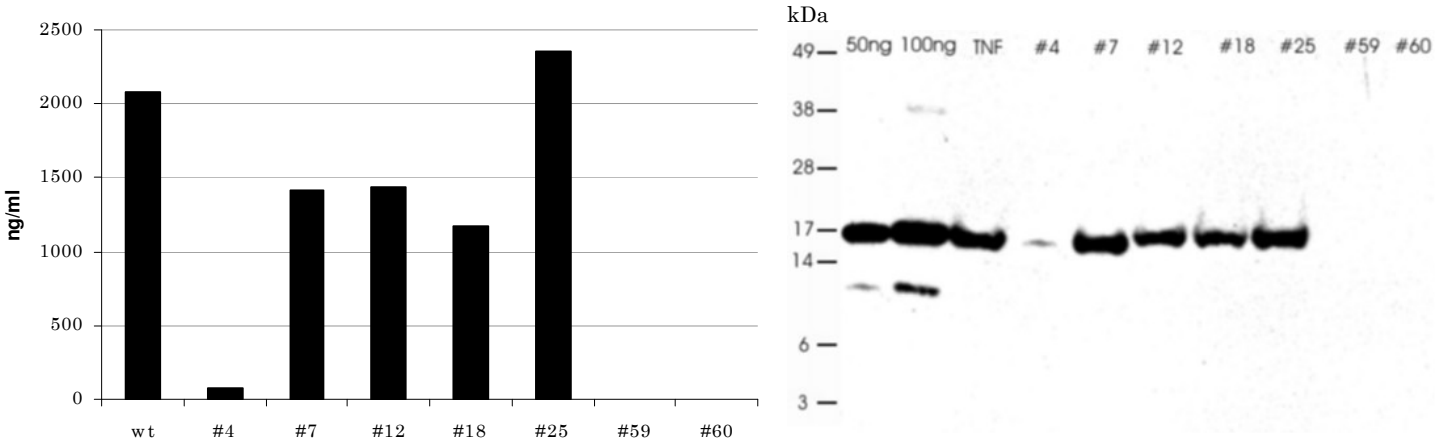


Figure 34. **Expression of wt TNF-α and the selected AutoVac variants in *Drosophila* cells**. Proteins are expressed transiently for 4 days in S2 *Drosophila* cells. The expression levels are displayed by ELISA and westen blot.

DISCUSSION

The need for correctly folded proteins in the AutoVac technology is apparent, as the immunological response generated upon vaccination is highly dependent on the recognition of spatial B-cell epitopes. However, inserting 15 amino acids of foreign peptide into the target protein is a difficult task. Until now the AutoVac proteins have been constructed on the basis of predictions from rational design, but a simpler and a less labour-intensive method is of obvious value. Furthermore, when using rational design, loops are often the target for insertion (Sheridan et al., 2002), but since PADRE is highly susceptible to proteases (unpublished data from Pharmexa) a screen might enable identification permissive insertion sites within buried regions of the target molecule, resulting in more protease resistant AutoVac proteins. In my option, these two approaches can complement each other in the quest for the most efficient and stable vaccine molecules.

Here, I present a complete system for a novel way of generating AutoVac proteins; a folding detection assay applied to a random epitope insertion library of TNF-α. 5 different putative AutoVac variants of TNF-α were found, 4 of which were expressed at a satisfactory level. The system is now ready to generate variants of other targets and/or insertions and to evaluate their folding competence. A number of adjustments to the procedure can be made, but a highly functional system has been provided. I have focused on the development of the complete system, from start to end, in order to prove the functionality, and remove the major obstacles of the library and selection system.

In order to establish this method, I have created a number of novel approaches not previously described, including: i) a genetic complementation system for detection of foldable sequences in a eukaryotic host, ii) a selection approach to detect the presence of invertase activity using a limiting amount of glucose in the selective sucrose media, enabling a selection scheme for positive clones.

The epitope insertion library

Size and randomness of the library

For generating the library of insertion sequences, a strategy including *in vitro* transposition was used, and this method proved successful. Although the procedure includes several reaction steps, it is straightforward and produces a library of good quality. No background of wt-TNF-α or transposon hits without insert sequence was found. All the insertions were

simple i.e. containing the expected base pairs without any arbitrary insertions or deletions.

Randomness is another important parameter of the library. On the basis of the results the library emerged as random. Only few insertions appeared twice and the insertions were evenly distributed in respect to insertion point and direction of the inserted sequence. This randomness, combined with the relative size of the library and the size of the TNF-α cDNA suggests that the library is exhaustive. This is in agreement with other reports of MuA mini transposon integration showing that all phosphodiester bands between the bases in a DNA fragment served as targets even though some are hit more frequently than others (Haapa et al., 1999). An investigation of the relative frequencies of insertion at the single phosphodiester bands could give an estimation of a minimal library size needed and the minimal number of variants to screen. Such a randomness analysis could be done by capillary electrophoresis using UV or fluorescent detection where the resolution level is at single base pairs. Hereby, the relative randomness of different commercially available transposon systems could also be evaluated.

The Insertion sequence

Insertion libraries made by introducing a DNA fragment of interest into random double-strand breaks have the inherent feature of translating into 6 different inserts, depending on the reading frame and orientation. In the library used in this project, 4 out of the 6 different insertion sequences do not contain a stop codon and thus can appear after the screen. Consequently, only 5 of the 22 positive clones contained PADRE.

The insertion points of all the positive variants hold much information as to where in the TNF-α molecule there is enough plasticity to accommodate modifications of the protein. Furthermore, the fact that there were 4 possible insertion sequences, reveals for each of the permissive sites found, whether only one specific sequence is allowed, or if various sequences have equal opportunity to be accommodated at this point, and thus stipulate some insertion sites are more tolerant than others.

The insertions translated into other sequences than the PADRE epitope expand information attained from the selection procedure. Nevertheless, they are not functional as AutoVac molecules. Therefore, generating a library which only translates into TNF-α-PADRE variants and introducing stop-codons in the non-PADRE insertions, could be attractive. Making different modifications in the insertion sequence could facilitate this. The simplest thing to do is to make mutations introducing stop codons in the redundant sequences that are silent in the PADRE epitope. This can be done for the sequences of RF2-d and RF3-r (Table 4). This

reduces the number of open reading frames to from 4 to 2, one of which codes for the PADRE sequence.

Current insert TGCGGCCGCTAAGTTCGTTGCAGCTTGGACCCTTAAGGCGGCCGCA
Insert with silent mutations TGCGGCCGCTAAGTTCGT**A**GCAGCTTGGACCCT**A**AAGGCGGCCGCA

	Translation of Current insert	Translation of insert with silent mutations
RF1-d	AAAKFVAAWTLKAAA	AAAKFVAAWTLKAAA
RF1-r	AAALRVQAATNLAAA	AAALRVQAATNLAAA
RF2-d	RPLSSLQLGPLRRP	RPLSS*QLGP*RRP
RF2-r	RPP*GSKLQRT*RP	RPPLGSKLLRT*RP
RF3-d	CGR*VRCSLDP*GGR	CGR*VRSSLDP*GGR
RF3-r	CGRLKGPSCNELSGR	CGRL*GPSCYELSGR

Table 4. The consequence of introducing stop codons in RF2-d and RF3-r is shown. The cDNA sequence is depicted showing the insert before and after mutation, and the table below illustrates the effect in the 6 different translated sequences this change.

Because of the way RF1-r relates to RF1-d (PADRE), the introduction of a stop codon in RF1-r will affect the amino acid sequence in RF1-d. Non-silent, but conservative, changes in the epitope could be acceptable. Such changes are not likely to influence the outcome of the selection and could be changed back afterwards. However, bases which are reverse complementary to stop codons in the RF1-r sequence result in histidine, tyrosine, glutamine, or aspartic acid residues, or in a stop codon in the PADRE sequence. The introduction of any of these residues will lead to non-conservative changes in the PADRE and alter the composition of the epitope dramatically; they might obstruct binding capacity to the MHC II molecules or the immunogenicity. Therefore, keeping the non-PADRE RF-1r sequence in the library is likely giving a much better outcome of the screen.

The PADRE sequence was identified as a promiscuous and efficient T cell helper epitope based on *in vitro* and *in vivo* MHC binding tests at Epimmune. Since Epimmune is now a subsidiary company of Pharmexa, it may be feasible to evaluate PADRE mutants containing a stop codon in their reverse sequence in the MHC binding assays developed at Epimmune to observe which amino acid change leading to a stop codon in the reverse sequence is acceptable in PADRE; this in respect to immunogenicity or binding capacity to the MHC II molecules.

Concatenation of the insertion fragments

A high degree of concatenation of the insertion fragments is seen and is based in the fact that the insert DNA was phosphorylated in order to avoid self-ligation of the vector. Another approach is to allow discrimination between plasmids containing insert fragment and self-

ligated plasmids. Destroying the Not1 sites of the variants containing insert fragment, and at the same time maintaining the *Not1* site, is in the unwanted variants only containing transposon derived residues can achieve this. In order to do so, the part of the *Not1* recognition sequences sites present in the insertion fragment is changed, but the sequence of the overhang sequence is preserved. Thereby, the capability of annealing to a fragment cut with *Not1* is maintained, but the recognition site of the restriction enzyme is removed. So, after insertion of the fragment and prior to transformation the pool of variants can be subjected to *Not1* and plasmids without insert being digested. In this way, the insert fragment can be made from large un-phosphorylated primers, simplifying the approach for obtaining the fragment.

Including substitutions and deletions in the library

The *Not1* site at the Entranceposon ends results in an extension of the epitope by two alanines, and the duplication derived from the transposition includes two extra amino acids in the insertion. Reducing the size of the insertions, or making substitutions, may reduce how much the insert compromises the stability of the variant protein and can be a means to diversify the complexity of the library. This could be accomplished by introducing type 2 restriction enzymes in the termini of the transposon sequence as previously described for other purposes (Jones, 2005). The redundant residues can be removed by including a Mly1 restriction site flanking one of the terminal Not 1 sites in the transposon. Digesting with Mly1 and then *Not1* after transposition removes the terminal alanines made by the *Not1* site and 4 bases of the 5 bp duplication. Alanine has an interchangeable third codon position; therefore, the residual base pair can be integrated in the PADRE sequence.

Mly1 makes a blunt end cut in the DNA, whereby overhanging bases are only found in one of the vector ends and making them incompatible for self-ligation. Therefore, cloning becomes easier, and the formation of the concatamers is inhibited. Other type 2 restrictions enzymes could make other variations of the mutagenic transposons. E.g. Mme1 removes 20 bp and produces a 2 base pair overhang. This end can be filled out by the Klenow fragment or alternatively, the insert fragment can contain tails with random overhangs as described by Murakami *et. al,* 2002. The number of bases removed by this procedure can be adjusted by the location of the Mme1 restriction sites relative to the Transposon end.

The Selection system for foldable protein variants

The system presented here is carried out on agar solid media in petri dishes, using standard laboratory techniques and the availability of an incubator with a shaking device at 30°C as sole prerequisite. It is a selection procedure that in contrast to a screen gives the possibility of examining a larger number of variants more readily. Furthermore, the selection was shown to be highly stringent. The stringency may derive form the use of *Sz. pombe* which has a highly structured and more elaborate secretion system than the one found in *S. cerevisiae*

The selection procedure

The screening procedure consists of three steps: i) selection on sucrose plates with limiting glucose, ii) re-streaking of selected clones on sucrose, iii) rescue of the plasmids of the variants and re-testing in an invertase assay. 36 colonies were selected on the plates, 25 clones showed growth on the sucrose plates, and 22 were positive in the invertase assay.

For practical reasons, the search for putative AutoVac candidates could progress from the re-streaking step via yeast colony PCR - where plasmids containing non-PADRE variants can be disregarded - directly to testing expression level, and still achieve adequate selection. Only 3 variants failed the last step of selection i.e. they were not positive in the invertase assay after re-streaking - one of which being an obvious false positive after DNA sequencing, as a stop codon is present in the sequence. Plasmid rescue and retransformation are important in order to deduce the quality of the selection system when it is being developed to know how many true positives there are at each step, but this is labour intensive. Therefore, it will be simpler to test for expression level and discard the false positives at this step, due to the limited number of colonies. This means that the information, as to whether lack of expression is caused by a false positive, or by a variant that cannot be expressed, since it is no longer fused to the invertase moiety is disregarded.

The multiple plasmids in the screening host are a drawback of the system and in *Sz. pombe* no low copy plasmids are available, but re-streaking the plasmids twice has largely resolved this obstacle. So when using colony PCR to discover the sequence of the variant caution should be placed on the possibility of finding more than one plasmid in each *Sz. pombe* colony.

Limiting glucose and other parameters for selection

In this system, the approach using limiting glucose in sucrose media enables a successful selection of invertase positive colonies, shown to be not otherwise obtainable. Because the

transformed cells can be plated directly on selective media, the necessity of replica plating or invertase plate assays is eliminated or reduced thereby reducing workload and occurrence of false positives.

Although secretion of invertase is crucial for growth on sucrose, as little as 0.6% of the wild-type invertase activity is sufficient to sustain growth on selective medium, and 4.7% results in growth rates indistinguishable of that of wild-type cells (Kaiser et al., 1987). Further, as data presented here, only 0.1% of glucose in the growth media is sufficient to support growth of invertase knockout strains, and these strains are very susceptible for glucose from positive neighbour colonies. This is probably the reason that a proportion of the clones 11 out of 36 picked on the selection plate turned up to be false positives; however, these were easily eliminated by the double re-streak procedure on sucrose media. This inconvenience is also known from the replica plating procedure (Jacobs et al., 1997), but in this approach false positives seem to occur much less frequently than when replica-plating from non-selective glucose plates. Altogether, the use of limiting glucose for direct selection of the transformed cells is an important improvement of the selection assays using invertase described so far. The technique represents, therefore, a new selection assay that could most likely be exploited in other systems based on invertase activity as selection marker like the signal sequence trap (Jacobs et al., 1997)

Further experience with the sucrose/glucose concentration ratios in the screen will become available as the limiting glucose procedure is used more extensively, e.g. other concentrations may be optimal for other targets. Also, the addition of antimycin to the selection plates should be tested. This may reduce the number of false positive and/or further enhance the stringency of the selection.

Many proteins fold better at lower temperatures (Baneyx and Mujacic, 2004). In addition, less stable proteins not viable at 30°C may be stable at lower temperatures. Therefore, temperature could be included as an additional parameter in the selection procedure and/or to adjust the sensitivity of the selection.

The dimension of the screen

In total, 22 colonies were selected from the 36,000 screened colonies (0.1%). This is equivalent to only 2 of the 948 possible variants having folding competence if the screen was exhaustive (36000 colonies /948 possible variants ≈ 40 colonies of each variant). The sequences of the selected variants show that 20 different insertions were found; therefore, the screen was not exhaustive. If the 20 different sequences represented the comprehensive selection of foldable

sequences 800 (40 colonies of each variant *20 different variants) colonies could have displayed invertase activity among the 36,000 colonies (≈2%).

The fact that only two of the picked colonies contained the same variant indicates that more different sequences are found as folding competent. This could imply that, the true result of positive colonies among screened colonies, due to foldable sequences encoded from the plasmids, is the 9% of growing colonies seen on the 0.03% glucose plates, before replica plating. This taking into account, that $^1/_3$ of the selected clones turned out to be false positives. If the number from the 0.03% plates is considered the true number of positives in the library, there are approximately 56 folding competent variants in the library (9% of 948 = 85, $^1/_3$ of 85 = 56).

However, this screen is not indented as a search for all foldable sequences in the library, but simply a proof of concept for the screening procedure. Thus, in order to demonstrate the competence of the selection procedure, a limited, and so more manageable, number of colonies is picked for analysis. If the screening had been allowed to proceed for more days, more positive clones could have been picked. This, in turn, might have resulted in relatively more false positives in the different steps of the screen, but also more true positives in total. Some of the same variants would probably have been found repetitively, but also new ones.

The current library has a maximum of only 948 different variants, but the complexity of the library can be increased as suggested in paragraph 1.4. Therefore, it is an apparent advantage of the invertase-based system that it is a selection procedure in contrast to a screening procedure giving the system the dimension to test a large number of colonies. Utilising invertase assay may include the qualitative aspect of a screen in the procedure. However, the invertase activity, and thus output of the invertase assay, is not only dependent on folding competence, but possibly on also folding velocity, as well as, on pre-translational factors differing in the variant sequences. Also, the results from the invertase assay are highly fluctuant between different rounds of experiments at the present, so some optimisation of must be done before full utilisation of this assay.

Analogous selection and screening systems for foldable sequences

A comparable protein solubility selection assay has been described in *E. coli* using the CAT resistance gene as selection marker (Maxwell et al., 1999). The selection capacity of this system was demonstrated by selecting soluble proteins from a larger pool of insoluble protein by mixing the insoluble HIV integrase and a soluble variant in the ratio 10,000:1, and select among 400,000 *E. coli* transformants. After screening 42 colonies grew and out of 18 analysed

clones 16 contained the soluble mutant. The selection was prominent, but not absolute and the number of non-soluble false positives is comparable to what was found in the invertase selection system after re-streaking. Notably, no actual selection was made demonstrating the stringency of the system on a complex library.

Nature has optimised ERAD as detector for folding competence (Ellegaard et al., 2003), and therefore, a selection system based on this feature is likely to be highly specific. The exploitation of the ERAD systems requires the use of eukaryotic cells with a functional ER and Golgi apparatus. Earlier work in *S. cerevisiae* shows that this feature is in fact functional in screens to assess stability of proteins. However, these screens have important drawbacks. Firstly, the yeast surface display system (Boder and Wittrup, 2000; Boder and Wittrup, 1997) seems specific and finds molecules with the desired qualities. But the system requires both the opportunity to perform FACS assay and the availability of highly specific florescent-labelled antibodies. Hence, it cannot be carried out by standard laboratory techniques. Secondly, the requirement of antibodies makes the system inapt for newly discovered proteins or proteins, which beforehand are difficult to express and therefore may have no, or a limited number of, available antibodies. Thirdly, if a modification is present in the area for antibody recognition, a folded variant will not be detected. The second screen, based on immunodetection of a generic tag (Hagihara & Kim, 2002), is applicable to any sequence of interest. Additionally, the screen is preformed on colonies growing on solid media in petri dishes. Thereby, the method solved the shortcomings of the yeast display system, but the screening procedure is not very stringent; many mutations, which passed the genetic screen, did not confer protein stability, necessitating additional selection procedures.

The system presented here improves the possibilities for searching for folding competent protein among libraries of variants, compared to other based yeast-systems described in the literature. It is stringent, simple, employing common laboratory techniques, relatively labor sparing, and can be applied to any targets without prior knowledge or specific detection aids for it.

Analysis of selected variants

Permissive insertions must allow energetically favourable and stable protein folding. The addition of residues in a protein can be accommodated either by the formation of local projections at the exterior of the protein causing small deformations, or by lateral insertion induced displacements of the polypeptide toward the protein surface producing loop enlargements. If such adjustments in the structure are not possible, the overall structure will be distorted. The variants where the insertions are situated in loops protruding at the surface of the protein are probably tolerated by extending these loops, whereas the insertions before the first and after the last β-strand are probably tolerated by lateral displacement of the terminal part of the protein.

The insertion sites of the selected variants – secondary structure

After selection 20 different insertions at 16 different insertion sites were found. All, but one of the insertions, leave the β-sheets intact. Thus, these insertion points correlate well with the secondary structure of the TNF-α molecule. The one insertion site interfering with the secondary structure is that of variant 11. This insertion is situated in the middle of a β-strand, and thus interferes with the secondary structure.

The fact that at some positions many different sequences are accommodated demonstrates that the selection assay can not only identify permissive sites, but also pinpoint to sites of highly plasticity. A site like this could be the obvious target for further construction of variants, including substitutions and deletions.

The insertion sites of the selected variants – tertiary and quaternary structure

The bulk of insertions is situated at the N-terminal of the TNF-α molecule, which is not surprising from the crystallographic data and deletion analysis, revealing that part of the area is highly flexible and can be deleted without effect on the structure or biological activity (Jones et al., 1989). Moreover, the stabilising properties of lysine-11, around which many of the insertions cluster, may help in accommodating the insertions by providing stability to the structure in this region, despite the considerable change the insertion of 15 heterologous derived amino acids presents to it.

The N- and C-termini are spatially in proximity, and are part of the receptor-binding domain. This domain accommodates 13 of the 22 insertions found, indicating an overall inherent flexibility of this region. This agrees with the finding that this receptor-binding domain is very diverse among the members of the TNF-α superfamily. This region facilitates contact between

ligands and receptors, and it contributes to the specific interaction between the different ligand and receptor combinations (Bodmer et al., 2002). Therefore, it is not surprising that this domain also has the plasticity to accept insertions.

The three lower up-wards facing loops on the solvent-accessible surface of the molecule are shown to be of a flexible nature (Jones et al., 1989). Also, there are reduced structural constraints from the rest of the molecule to extend the loops, relative to loops in buried regions. Therefore, it makes good sense that insertions are found here. It is puzzling as to why no insertions are found in the two upper loops of the molecule. However, this can be due to the fact that the screen was not exhaustive, which may account for the lack of inserts.

Variant 11 is situated toward the interface of the trimer TNF-α molecule, and is, furthermore, situated in a β-strand; this variant is, therefore, somewhat contentious in respect to the structural foundation of the molecule. A lateral displacement of β-strand toward the region of the N-terminus could be suggested as the means to accommodate this insertion, thereby providing space for the insertion. Alternatively, this TNF-α variant may be expressed as a monomeric protein, leaving space for the insertion to bulge out in a space otherwise occupied by the other subunits in the trimer. Although most TNF-α molecules are found as trimers, monomeric TNF-α molecules are to be found in the plasma of humans, mice and rats (Aderka et al., 1992; Corti et al., 1992). The oligomerisation state of this variant could be an interesting question to address, This can be measured by gel filtration and light scattering after expression.

The ultimate consequences of the insertions on the overall protein structure are simply speculative suggestions, inviting to experimental exploration. The exact outcome in the protein structure caused by the insertions ultimately must be resolved by protein crystallography. Nevertheless, most importantly the selection procedure pinpoints a specific region of TNF-α as insertion tolerant.

TNF-α variants made by rational design compared to selected variants

It is interesting to compare the insertions sites found by the selection method and the ones found by rational design (Figure 35). Variant 4 is found to be foldable both in the screen and by rational design. Again the sequence of the two versions is not exactly the same since the library-derived insertion sequence has two extra alanines at the end of the insertion. Both the library derived and the designed version displayed a very low level of expression. Importantly, the fact that this variant is found by both approaches is an indication of the validity of the screening approach.

The most N-terminal insertion site (variant 16) is also found in the rational design, but the sequences differ. From the screen only the RF1-r derived sequence (AAALRVQAATNLAAA) was found at the most N-terminal part, suggesting that this sequence is most acceptable here. But as shown by the designed variant, the PADRE sequence can also be inserted here.

Two variants with insertions between TNF-α and invertase in the fusion were found after screening the library. None of these has the PADRE sequence as insert. A complementary variant with PADRE at the C-terminus has been made by rational design, but this variant did not show any expression when tested. Although the insertion site is similar, there are several dissimilarities between the variants; most notably that the sequences in the insert differ. But also, the variants derived from the library are expressed as invertase fusion proteins, and this may allow the insertion of a sequence in the C-terminus of the TNF-α protein.

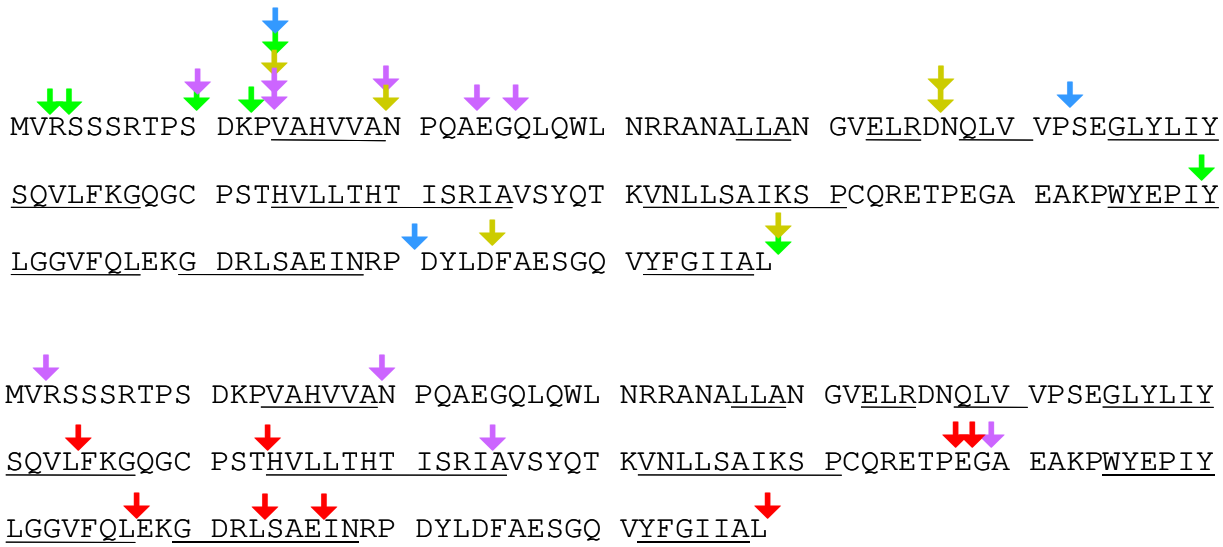


Figure 35. Insertions points are shown as arrows in the primary structure of TNF-α.
Top: The insertions found by selection. The 4 different colors represent the 4 possible amino acid insertion sequences. ↓AAAKFVAAWTLKAAA (PADRE), ↓AAALRVQAATNLAAA, ↓RPLSSLQLGPLRRPQ, ↓CGRLKGPSCNELSGRT
Bottom: The insertions predicted by rational design. The insertion sites of PADRE are shown as purple arrows or red arrows. Purple arrows indicate an expressible variant, whereas red indicate no expression of the variant.
↓ Expression, ↓ No expression

As seen in figure 35 some regions of the protein can accommodate more than one insertion sequence, whereas in others, only the same insert sequence has been found. Although the screen is not exhaustive, this gives an indication of the tolerance of insertion of the area, and can be utilised if the results from the screen should make the basis for further modifications by design. If a variant were to be designed after the results of the screen, it would be obvious to choose the N-terminal region, especially the area just before the first β-strand as target for insertions.

The designed variant resulting in the highest expression level is the TNF-α AutoVac molecule

with PADRE inserted in the large flexible loop between β-sheet E and F. Here 6 different insertions were tried out; of which 3 were expressible (Figure 36). One of these is a simple insertion into the flexible loop (TNF4). Although it could be anticipated that a variant corresponding to TNF4 should have appeared after selection from the library presented here, such a variant was not found. This could simply be because the screen was not exhaustive, and screening a larger number of clones would have revealed this variant. Another reason may be that the variant from the library corresponding to TNF4 does not have folding competence. Variants generated in the library have at least 2 additional residues on each side of the inserted peptide, compared to variants based on rational design (see Figure 36). As seen, even small differences in the insertion may be sufficient to compromise folding and cause the cessation of expression; this may explain that the selection did not result in a TNF4-like variant.

	103	112 expr.	
TNFα		RETPEG----------AEAK	+
TNF4		RETPEG--**AKFVAAWTLKAAA**--AEAK	+
TNF5		RETP----**AKFVAAWTLKAAA**EGAEAK	–
TNF6		RETPE---**AKFVAAWTLKAAA**-GAEAK	–
TNF7		RETPEG--**AKFVAAWTLKAAA**---EAK	+
TNF8		RETPEG--**AKFVAAWTLKAAA**-----K	+
TNF9		RETP----**AKFVAAWTLKAAA**-----K	–
TNFlib		RETPEGAA**AKFVAAWTLKAAA**EGAEAK	?

Figure 36. The TNF-α AutoVac made by rational design
The figure depicts the TNF-α variants in the flexible loop 3 (arginine-103 to lysine-112) made from suggestions based on rational design. The latter variant depicts how the corresponding library variant looks. The figure is modified from Nielsen *et al.*

Expression of variants in different expression systems

In order to verify that invertase does not drive or affect folding and the variants could thus be expressed after removal of the invertase moiety, the 5 selected variants holding the PADRE insertion were tested for expression in *Sz. pombe*, *E. coli* and *Drosophila*, as these are potential AutoVac protein candidates. All five variants were expressed as soluble proteins in *E. coli* and were secreted in *Drosophila*. It proved difficult to express the TNF-α protein in the *Sz. pombe* system and in the variants even more so. Other reports of *Sz. pombe* expression have shown this to be a common phenomenon, due to extended protease activity in the growth media (Kumar and Singh, 2004; Tokunaga et al., 1993; Kjaerulff and Jensen, 2005).

All tested proteins could be expressed after removal of the invertase moiety, indicating that not only the tested, but in fact all the selected proteins have folding and expression competence without the invertase moiety. Hence, the selection procedure appears to be very stringent. Furthermore, even if *Sz. pombe* will not be used as expression host due to the low expression level and protease secretion, the system is able to detect proteins with folding

competence that exceeds the host system, making apparent its value as a screening host. The low expression level of the *Sz. pombe* in fact may be an advantage to the screen, as it might be a crucial factor of the high stringency of the system.

Differences expression level between the selected variants

Variant 4 only showed a detectable level of protein in the *Drosophila* and *E. coli* expression systems, but in both, it was the variant with the lowest yield. A variant similar to that of variant 4 had been made by rational design, and also this designed variant did show expression in *E. coli.* However, both the selected and the designed variant displayed very low yields of soluble protein. Variant 4 is the sole of the tested variants where the insertion violates a β-strand. This β-strand disturbance could lead to instability of the protein structure; it may be at the rim of mis-folding, and therefore giving the poor expression.

The variant with expression in all systems and the highest expression level both in *Sz. pombe* and *E. coli* is variant 25. The level even exceeds that of the wt-TNF-α. Since the stability and secretions levels of proteins have been reported to correlate (Kowalski et al., 1998), the high expression level of this variant may indicate a protein stability comparable to the wt-TNF-α protein. Further, since the low levels of detectable protein in *Sz. pombe* are generally owing to protease activity in the expression media, this system gives an indication of the variants resistance to proteases and following the quality of the tightness of globular protein structure, as protruding peptide segments are more prone to be substrate for proteases. Therefore, the viability of the variants in this environment can indicate the degree interruption of the insertion on the protein structure. Variant 25 is the only protein variant detectable in *Sz. pombe*, and this point towards a very stable protein structure in which the disturbance of the insert in the TNF-α structure is negligible. The fact that the expression level of variant 25 in *E. coli* is comparable to that of wt-TNF-α would in principle make it a candidate to move further on in the process of vaccine development, which comprises large-scale fermentation, protein purification, and immunological testing. In particular, it would be highly interesting to analyse the protein chemical and immunological properties of these new TNF-α variants, including structural comparison with wt-TNF-α, stability and immunogenicity. Structural comparison could be done e.g. by CD-spectroscopy and light scattering. Stability could be investigated by in thermal denaturation studies using CD-spectroscopy and/or differential calorimetry. Immunogenicity studies could comprise vaccination of mice and characterisation of the resulting antisera as to TNF-α antibody titer, and ability to cross-react with and neutralise wt-TNF-α. Ultimately, the new TNF-α variants should be evaluated in a relevant disease model such as rheumatoid arthritis murine models for therapeutic effect.

In conclusion, the screening approach is able to useful in order to find foldable protein capable of being expressed in a soluble form. There was found one AutoVac candidate that could be moved on as an *E. coli* expression project. In *Drosophila* 4, out of the 5 variants showed an expression level beyond 1:g/ml, which is the level considered satisfactory for AutoVac candidates in a *Drosophila* expression project at Pharmexa. This means that the task set out for my PhD project has been fulfilled.

Other applications of the library-selection system

The expression of heterologous proteins is a general requirement in many areas of protein science and molecular biology, as well as in biotechnology. A solubility assay can be useful when improving protein functions or solubility by optimising the primary sequence using single amino acid substitutions (Petrounia and Arnold, 2000; Jones, 2005). Furthermore, more extreme changes, as by the AutoVac technology, calls for the use of a strategy for selection of soluble proteins.

Application: Protein dissection

The vaccine targets, to which the AutoVac technology can be applied, may be large multi domain proteins. In general, large proteins are often difficult to express recombinant and often result in heterogeneous products. For such proteins, it will be advantageous only to use a smaller part of the protein in a vaccine. This part with an epitope insertion will be sufficient to induce an immune response. The described system would be very applicable for identifying small domains in a larger protein having the potential to achieve a foldable structure. The transposons can generate both N-, and C-terminal mutants at random, and this, combined with the selection system identification of properly folded truncation mutants, is possible.
This system could also be used outside Pharmexa for identification of important regions of protein structure and function in general (Stemmer, 1994). Moreover, structural genomics are often limited by the ability to produce or acquire soluble protein variants for crystallography and high solubility strongly correlates with success in structural studies. The need for a mean to obtain soluble proteins is enhanced after the sequencing project for several organisms has been completed. It may not be possible to produce proteins their full length e.g. due to membrane spanning domains, and the method described here could provide a tool for definition of properly structured domains of proteins of interest.

Application: Multi component proteins

Similarly, the transposon system can also facilitate the generation of a library of stable hybrid molecules (Naumann et al., 2002). Using transposition, two proteins can be linked together via transposon insertion in the N-terminus of one protein of interest and C-terminally of another. The two proteins are assembled using the restriction sites from the transposons. In the context of AutoVac the epitope could be situated between the protein moieties. Alternatively, a variant containing an epitope could be found prior to protein fusion. The selection procedure can now be applied to find hybrid variants that are both in-frame with each other and

foldable. The described system could also be of value for other applications, e.g. making multi-component enzymes (Wells et al., 1985).

Conclusion

I have developed a general method to find expressible vaccine proteins which contain a foreign Th-cell epitope. With this project I have provided a means to generate an exhaustive insertion library with precise epitope insertions, and a stringent eukaryotic selection system for folding competent proteins, and thus a valuable tool in the generation of AutoVac candidates for Pharmexa. This method can advantageously complement the "trial and error" and rational design approaches used so far; it enables the generation of a larger panel of identified candidates, it potentially provides more candidates for immunological evaluation, and it reduces current bottlenecks in the research phase.

During the course of this work 4 novel putative TNF-α vaccine candidates were found. These candidates revealed interesting epitope insertion sites, not earlier investigated in Pharmexa's TNF-α vaccine development program. These novel TNF-α AutoVac variants represent therefore promising candidates to move further in the process of vaccine development via large scale fermentation, protein purification and immunological testing.

Conclusively, based on the presented results, I would strongly advocate for Pharmexa to implement this method in the future quest for AutoVac candidates.

MATERIALS & METHODS

Cloning procedures

Standard cloning procedures were carried out according to Samsbrook and Russell, 2001. All restriction enzymes were pursed from New England Biolab and all primers from DNA technology. PCR reactions were preformed using Expand PCR system (Roche) according to manufactures instructions. The *E. coli* strain DH10b (Invitrogen) is used for cloning purposes, and grown in liquid or agar solid LB-media (Invitrogen) containing 50 μg/ml Carbencillin (Sigma). Vectors and inserts were ligated together using T4 ligase. All plasmids are sequenced in a region covering modifications to check the fidelity of the constructs, using the BigDye kit (Applied Biosystem) and run at the ABI PRISM 310 Sequence analyser (Applied Biosystem).

Library plasmid

Insertion mutagenesis by transposon was carried out in a modified pUC19 vector. The pUC19 plasmid (Invitrogen) was linearised by the restriction enzymes *Nde1* and *Hind III*. The wt-TNF-α gene was amplified using PCR and the primers 5'GCACCATATGCCATGGTGCGCTCAAGCTCGCGCACGCC GA3' and 5'TTACGCCAAGCTTGAGCTCCCAATGCGATAATGCCGAAGTAGACC3'. The resulting DNA fragment is cut by the restriction endonucleases *Nde1* and *Hind III*.

PADRE-pVax

The pVAX plasmid (Invitrogen) was lineased by restriction with *Not1*. The insert was made by annealing the two primers 5'GCGGCCGCTAAGTTCGTTGCAGCTTGGACCCTTAAGGCGG3' and 5'CCGCCTTAA GGGTCCAAGCTGCAACGAACTTAGCGGCCGC3'. (*The PADRE –pVAX plasmid was made by Sven Muller)*

Transposition

The MuA transposon reaction kit was acquired from Finnzymes and carried according directions. The transposon was removed by the restriction enzyme Not 1 succeeded by CIP dephosphorylation. The transposon was interchanged with insert fragment from plasmid PADRE-pVAX also cut with *Not1*.

Sz. pombe Screening host, Screening Plasmid, and transformation

The *Sz. pombe* invertase knockout strain was generated from Eg660 Δinv1::ura4+ obtained from O. Nielsen (University of Copenhagen), using a one-step gene replacement technique described by Rothstein, 1983. The primers used for amplification of the invertase gene and flanking sequences were 5'ATCTAGACAAAATACATCTTATGTATTGAACATTAG3' and 5'AGAATTCGCAAATGCCACTAGCTAA AATATATTTC3'. Correct integration was confirmed by PCR.

The invertase gene was cloned from the Eg660 strain using the primers 5'ACCATGGCTCCCCGTCACT TATATGTAA3' and 5'AGCGGCCGCTTAGCAATTCCAGATAGTCTTTAATG'3 was introduced in the pREB3X1 vector by the restriction enzymes *Nco1* and *Not1*. A *Not1* and *Sac1* restriction site was destroyed by restriction of the relevant enzyme, end-filling by T4 ligase and re-ligation. A *Sac1*

restriction site and a stop codon was introduced in frame between the Cpy-signal sequence and the Inv gene by the primers 5'TATCCATGGCATGAGCTCCCCGTCACTTATATGTAAAACG'3 and 5'CAGTGG TATACGCAATAGCCTG'3. The fragment and vector were digested with the restriction enzymes *Nco1* and BspEI. The library of variants was transferred from the pUC19 -based plasmid to the screening plasmid using the restriction enzymes *Nco1* and *Sac1*. (*The screening vector and host was made by Søren Kjærulff*)

Sz. pombe transformation

Sz. pombe transformations were performed using a PEG [poly (ethylene glycol)]/lithium acetate procedure (Moreno et al., 1991). The expected transformation frequency is 10^4 to 10^6 per ug of DNA. This can, however, vary significanltly. This may by a concequence of the exact growth status of the cells at harvest time. Grow 150 ml culture in EMM medium at 30°C to a density of 1 - 2 x107 cells/ml. Harvest the cells at 3000 rpm for 5 min at room temperature. Wash cells in 40 ml of H2O and spin them down as before. Resuspend the cells at 1 x 109 cells / ml in 0.1 M lithium acetate (adjusted to pH 4.9 with acetic acid) and dispense 100 µl aliquots into Eppendorf tubes. Incubate at 30°C for 60 - 120 min. Cells will sediment at this stage. Add 1 µg of plasmid DNA in 15 µl TE (pH 7.5)/H20 to each tube and mix by gentle vortexing, completely resuspending cells sedimented during the incubation. Do not allow the tubes to cool down at this stage. Add 290 µl of 50 % (w/v) PEG 4000 prewarmed at 30°C. Mix by gentle vortexing and incubate at 30°C for 60 min. Heat shock at 43°C for 15 min. Cool the tubes to room temperature for 10 min. Centrifuge at 2000 rpm for 2 min in an Eppendorf centrifuge. Carefully remove the supernatant by aspiration. Resuspend the cells in 1 ml of 1/2 X YEA by pipetting up and down with a P1000. Transfer the suspension to a 50 ml flask and dilute with 9 ml of 1/2 YEA Incubate with shaking at 30°C for 60 min or longer. Cells are washed twice in Emm-sucrose media before plating on selection plates with freshly made ultra pure sucrose (filter-sterilised).

Edinburgh minimal medium (EMM)

EMM with the different carbon sources were prepared as described (Moreno et al., 1991). The medium contains Potassium phthalate (3 g/L), Na2SO4 (0.04 g/L), ZnSO4 (0.4 mg/L), Na2HPO4 (2.2 g/L), Pantothenic acid (1 mg/L), FeCl2 (0.2 mg/L), NH4Cl (5 g/L), Nicotinic acid (10 mg/L), Molybdic acid (40 µg/L), Myo-inositol (10 mg/L), Potassium Iodide (0.1 mg/L), MgCl2 (1.05 g/L), Biotin (1 mg/L), CuSO4 (40 µg/L), CaCl2 (14.7 mg/L), Boric acid (0.5 mg/L), Citric acid (1 mg/L), KCl (1 g/L), MnSO4 (0.4 mg/L).

Plasmid Rescue

The plasmid rescue were made by harvesting 5 ml of *Sz. pombe* culture OD_{600}=1. The cells are harvested and washed once in 5 ml milliQ H_2O. 500µl extraction buffer (100mM Tris (pH 7.6), 10 mM EDTA, 1% SDS, 0.2 M NaCl) and glass beads until 2 mm below the liquid surface was added. 500 ul Phenol/ Chloroform/isoamylalcohol (25:24:1) is added and the sample is submitted to 15 sec of shaking in the Fast-Prep. The Phenol phase is removed after centrifugation for 5 min at 20000 rpm. One more phenol extraction is done and the DNA is extracted by making EtOH precipitation twice. The plasmid DNA is now ready for transformation into *E. coli*. The plasmid yield is not low, so highly competent cells must be used.

Yeast Colony PCR

Due to an extended DNase activity in the cells, a successful colony PCR is highly dependent of the time passing between cell lysis and the onset of the PCR reaction. Hence, the PCR mix are prepared prior to cell lysis and kept on ice. The cell lysis is achieved by harvesting cells from 1 ml *Sz. pombe* culture of in exponential growth, adding 500 ml of extraction buffer (100mM Tris (pH 7.6), 10 mM EDTA, 1% SDS, 0.2 M NaCl) and 250 ml glass beads. The mixture is submitted to 15 sec in the Fastprep and cell debris is spun down for 1 min at 13000 rpm. 1 ul of the supernatant are transferred to the PCR reaction and standard PCR (prepared according to manufactures instructions) are preformed.

Invertase overlay solution:

50ml (10 plates), 30 ml H2O, 2.15 g Sucrose, (Ultrapure invitrogen), 5.0 ml 1.0M Sodium acetate buffer (pH5.5), 1.0 ml 20mM N-ethylamide (NEM) , 0,5 mg Horseradish peroxidase (HRP), 100 ul Glucose oxidase (1000 u/ml) , 30 mg Dianisidine, 10 ml Agar solutin 3% in H2O (50°C). (Darsow et al., 2000)

Liquid Invertase assay:

Grow Cell culture ~ 5x 10^6 cells/ml, Normalize the different cell cultures to the same OD600, Spin 1000 µl culture, Resuspend cells in 1 ml media, wash in 1 ml Media, Resuspend cells in 1 ml media.
10 µl cell culture+ 80 µl Media + 10 µl Sodium acetate buffer, 1.0 M (pH 4.9) (made from 17,5 M acetic acid pH adjusted with 6M NaOH). Add 25 µl Sucrose 30%, incubate 30 min at 30°C. Add 19,5 µl 0,2M K_2HPO_4 (pH 10.0) incubate 3 min at 100°C. The change to a basic pH takes the enzyme out of its active range and also renders the enzyme more sensitive to heat treatment. Spin down. Transfer supernatant to cuvettes. Add 1 ml invertase assay solution (39 ml K_2HPO_4 Buffer 0.1 M pH 7.0, 200 µl N-ethylamide (NEM), 20 mM, 100 µl Horseradish peroxidase (HRP), ML/ml, 80 µl Glucose oxidase, 1000 units/ml, 600 µl Dianisidine 10 mg/ml. Stocks of NEM, HRP, and dianisine can be made in advance and stored at 4°C. Dianisdine should be stored in a light-proof bottle). Incubate 30 min at 30°C. Add 1 ml of 6 N HCl. Measure OD_{540} (Darsow et al., 2000).

Sz. pombe plasmid and recombinant expression

Plasmids of the variants *Sz. pombe* plasmids containing Wt and variant protein were made by PCR amplification of the insert cDNA from the screening vector by the primers 5'CGTTTTACATATAA GTGACGGGGAGCTCAGAGGGC3' and 5'CCGGGTCATCATTACTATTACTAACGAGCT3', and cut with the restriction enzymes *Nco1* and *Sac1*. The pET28 vector was linearised with the restriction enzymes *Nco1* and *Sac1*, and insert and vector ligated together.
A single *Sz. pombe* colony was grown ON in a 3 ml pre-culture Emm-glucose media containing 2% Broad bean peptone and 15 uM thiamine or no thiamine. Expression was commenced at OD_{600} =0.2 inoculated from the ON cultures. The expression media was 10 ml EMM-glucose media containing 5% Broad bean peptone (Fluka) and CS-LEU amino acid supplement (Q-BIO gene). Media and cells were separated by centrifugation at 3000 rpm for 5 min at room temperature. Media supernatants were tested in the TNF-α ELISA.

E. coli expression plasmids and recombinant expression

5'CGTTTTACATATAAGTGACGGGGAGCTCAGAGGGC3' and 5'CCGGGTCATCATTACTATTACTAAC GAGCT3' and cut with the restriction enzymes *Nco1* and HindIII. The pET28 vector was linearised with the restriction enzymes *Nco1* and HindIII, and insert and vector ligated together.

Bl21-star cells transformed with the relevant plasmids and grown in a 5 ml ON culture. Cells were inoculated to an OD_{600} = 0.04 in 50 ml LB media with 60 ug/ml Kanamycin in shake flasks at 37°C. At OD_{600} = 0.7 IPTG was added to a concentration of 1 mM to initiate expression. After 5 hr of expression 1 ml cells were harvested and 1 ml sonication buffer were added (20mM Bis-Tris pH 6.0,40 mM NaCl, 10 mM EDTA) together with Complete protease inhibitor, according to manufactures instructions. Cells were sonicated on ice 4 times 15 sec at 11 Amp, with 20 sec interruption between. Cells were centrifuged at 20000 rpm in 4°C for 20 minutes and supernatants were tested in the TNF-α ELISA.

Drosophila plasmid and recombinant expression

The TNF-α and variants cDNA were amplified from the screening vector by the primers 5'TCAGCTG AATTCCCGATCTCAATATGAAGTTATGCATATTACTGGCCGTCGTGGCCTTTGTTGGCCTCTCGCTG GATGGTCAGATCATCTTCTCGAACCCCG3' and 5'CGGTCTAGAGGATCCATCACTATTAGAGGGCAA TGATCCCAAAGTAGACCTGCCC3' and introduced into p2Zop2F vector linearised with the restriction enzymes HindIII and BamHI.

Drosophila S2 cells (ATCC CRL-1963) were used as expression host. Plasmid transfection was done using the Saint-18 transfection reagent (Cytotech). 3 ug of DNA is mixed with 300ul serum and antibiotic free media. 25 ul of Saint 18 transfection agent is mixed with 300 ul serum and antibiotic free media the two solutions are mixed and incubated at room temp 10-15 min. Media is remove from cells grown to a density of 10^6 cells/ml, and they are rinsed with 5 ml serum-free media. 2.5 ml serum free media is mixed with transfection solution and added to the cells, and they are incubated for 18 hrs at 37°C. Hereafter, the media is replaced with 4 ml Excell-420 media (JRH Biosciences) with 10% FBS in 25 cm^2 T-flasks at 37°C and grown for 4 days. Media and cells were separated by centrifugation at 3000 rpm for 5 min at room temperature. Media supernatants were tested in the TNF-α ELISA and in western blot.

ELISA quantification of protein

To 96 well Maxisorp micro titter plates were added 100 ul coat antibody Z0109 (Dako), diluted 1:1000 in coating buffer (3.18 g/l Na_2CO_3, 5.85 g/l $NaHCO_3$, pH 9.0); 150 ul blocking buffer (1 % PBS in 10mM phosphate/150 mM NaCl buffer, pH 7.2); 100 ul Capture antibody mAB210 (R&D Biosystems) diluted 1:1000 in Dilution buffer (0.05% Tween 20 (Sigma), and 1 % PBS (Sigma) in 10mM phosphate/150 mM NaCl buffer, pH 7.2) ; 100 ul Detection antibody BAF210 (R&D Biosystems) diluted 1:500 in Dilution buffer; 100 ul Streptavidine-HRP RPN1231V (Amersham) in Dilution buffer; 100 ul of OPD (Sigma) 10 mg in 25 ml colour buffer (7.3 g/l Citric acid*H_2O,11.86 g/l Na_2HPO_4*H_2O) with 10 ul H_2O_2 (Riedel-deHaën). Between each step the plates were washed 3 times with wash buffer (10 ml/l TritonX100, 29.3 g/l NaCl, 0.2 g/l KCl, 0.2 g/l KH_2PO_4, 1.15 g/l $NaHPO_4$*$2H_2O$, pH 7.2). The colour reaction was stopped

adding 100ul of 1M H_2SO_4 and the plates were read at 490/630 nm.

Western blot

25ul of sample were loaded on NuPAGE 4-12% SDS gels (Invitrogen) together with 50ng or 100ng of *E. coli* derived standard wt-TNF made at Pharmexa. Samples run and blotted according to manufacturers instructions. Antibody dilution, blocking, and wash was preformed in ECL buffer (250mM Tris, 750 mM NaCl, 25mM EDTA, 0,5% NP-40, pH7.4) The blot was blocked for 2 hr and then incubated with primary antibody (polyclonal anti-TNF-α, generated in-house at Pharmexa) diluted 1:1000 for 1 hr, and secondary antibody goat-anti-rabbit (DAKO p448) diluted 1:2000 for 1 hr. Between each step the blot was washed for 3 x 20 min in ECL buffer.

LIST OF REFERENCES

Aderka,D., Engelmann,H., Maor,Y., Brakebusch,C., and Wallach,D. (1992). Stabilization of the bioactivity of tumor necrosis factor by its soluble receptors. J. Exp. Med. *175*, 323-329.

Adzuma,K. and Mizuuchi,K. (1989). Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. Cell *57*, 41-47.

Adzuma,K. and Mizuuchi,K. (1991). Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage mu. Involvement of protein oligomerization in the ATPase cycle. J. Biol. Chem. *266*, 6159-6167.

Agrawal,A., Eastman,Q.M., and Schatz,D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744-751.

Akerley,B.J., Rubin,E.J., Camilli,A., Lampe,D.J., Robertson,H.M., and Mekalanos,J.J. (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. Proc. Natl. Acad. Sci. U. S. A *95*, 8927-8932.

Aldaz,H., Schuster,E., and Baker,T.A. (1996). The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. Cell *85*, 257-269.

Alexander,J., Sidney,J., Southwood,S., Ruppert,J., Oseroff,C., Maewal,A., Snoke,K., Serra,H.M., Kubo,R.T., Sette,A., and . (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. Immunity. *1*, 751-761.

Anelli,T., Alessio,M., Bachi,A., Bergamelli,L., Bertoli,G., Camerini,S., Mezghrani,A., Ruffato,E., Simmen,T., and Sitia,R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. EMBO J. *22*, 5015-5022.

Anken,E., Braakman,I., and Craig,E. (2005). Versatility of the endoplasmic reticulum protein folding factory. Crit Rev. Biochem. Mol. Biol. *40*, 191-228.

Appenzeller,C., Andersson,H., Kappeler,F., and Hauri,H.P. (1999). The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. Nat. Cell Biol. *1*, 330-334.

Baeyens,K.J., De Bondt,H.L., Raeymaekers,A., De Ranter,C.J., and Fiers,W. (1997). Crystallization and preliminary X-ray studies of mouse tumor necrosis factor. Acta Crystallogr. D. Biol. Crystallogr. *53*, 329-330.

Bainton,R., Gamas,P., and Craig,N.L. (1991). Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell *65*, 805-816.

Baker,T.A. and Mizuuchi,K. (1992). DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. *6*, 2221-2232.

Baldwin,R.L. (1989). How does protein folding get started? Trends Biochem. Sci. *14*, 291-294.

Baneyx,F. and Mujacic,M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nat. Biotechnol. *22*, 1399-1408.

Barlow,D.J., Edwards,M.S., and Thornton,J.M. (1986). Continuous and discontinuous protein antigenic determinants. Nature *322*, 747-748.

Barlowe,C., Orci,L., Yeung,T., Hosobuchi,M., Hamamoto,S., Salama,N., Rexach,M.F., Ravazzola,M., Amherdt,M., and Schekman,R. (1994). COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell *77*, 895-907.

Bazan,J.F. (1993). Emerging families of cytokines and receptors. Curr. Biol. *3*, 603-606.

Belden,W.J. and Barlowe,C. (2001). Role of Erv29p in collecting soluble secretory proteins into ER-derived transport vesicles. Science *294*, 1528-1531.

Bendtsen,J.D., Nielsen,H., von Heijne,G., and Brunak,S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783-795.

Beutler,B. and Cerami,A. (1988). Tumor necrosis, cachexia, shock, and inflammation: a common mediator. Annu. Rev. Biochem. *57*, 505-518.

Biery,M.C., Lopata,M., and Craig,N.L. (2000a). A minimal system for Tn7 transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. J. Mol. Biol. *297*, 25-37.

Biery,M.C., Stewart,F.J., Stellwagen,A.E., Raleigh,E.A., and Craig,N.L. (2000b). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic Acids Res. *28*, 1067-1077.

Black,R.A., Rauch,C.T., Kozlosky,C.J., Peschon,J.J., Slack,J.L., Wolfson,M.F., Castner,B.J., Stocking,K.L., Reddy,P., Srinivasan,S., Nelson,N., Boiani,N., Schooley,K.A., Gerhart,M., Davis,R., Fitzner,J.N., Johnson,R.S., Paxton,R.J., March,C.J., and Cerretti,D.P. (1997). A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. Nature *385*, 729-733.

Boder,E.T. and Wittrup,K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. *15*, 553-557.

Boder,E.T. and Wittrup,K.D. (2000). Yeast surface display for directed evolution of protein expression, affinity, and stability. Methods Enzymol. *328*, 430-444.

Bodmer,J.L., Schneider,P., and Tschopp,J. (2002). The molecular architecture of the TNF superfamily. Trends Biochem. Sci. *27*, 19-26.

Bretscher,P. and Cohn,M. (1970). A theory of self-nonself discrimination. Science *169*, 1042-1049.

Brockhaus,M., Schoenfeld,H.J., Schlaeger,E.J., Hunziker,W., Lesslauer,W., and Loetscher,H. (1990). Identification of two types of tumor necrosis factor receptors on human cell lines by monoclonal antibodies. Proc. Natl. Acad. Sci. U. S. A *87*, 3127-3131.

Brys,R., Nelles,L., van der,S.E., Silvestre,N., Huylebroeck,D., and Remacle,J.E. (1998). Identical cis-acting elements and related trans-acting factors control activity of nonviral promoter in Schizosaccharomyces pombe and mammalian cells. DNA Cell Biol. *17*, 349-358.

Caramelo,J.J., Castro,O.A., Prat-Gay,G., and Parodi,A.J. (2004). The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. J. Biol. Chem. *279*, 46280-46285.

Carlson,M., Taussig,R., Kustu,S., and Botstein,D. (1983). The secreted form of invertase in Saccharomyces cerevisiae is synthesized from mRNA encoding a signal sequence. Mol. Cell Biol. *3*, 439-447.

Casagrande,R., Stern,P., Diehn,M., Shamu,C., Osario,M., Zuniga,M., Brown,P.O., and Ploegh,H. (2000). Degradation of proteins from the ER of S. cerevisiae requires an intact unfolded protein response pathway. Mol. Cell *5*, 729-735.

Cha,S.S., Kim,J.S., Cho,H.S., Shin,N.K., Jeong,W., Shin,H.C., Kim,Y.J., Hahn,J.H., and Oh,B.H. (1998).

High resolution crystal structure of a human tumor necrosis factor-alpha mutant with low systemic toxicity. J. Biol. Chem. *273*, 2153-2160.

Chan,F.K., Chun,H.J., Zheng,L., Siegel,R.M., Bui,K.L., and Lenardo,M.J. (2000). A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. Science *288*, 2351-2354.

Chappell,T.G., Hajibagheri,M.A., Ayscough,K., Pierce,M., and Warren,G. (1994). Localization of an alpha 1,2 galactosyltransferase activity to the Golgi apparatus of Schizosaccharomyces pombe. Mol. Biol. Cell *5*, 519-528.

Chappell,T.G. and Warren,G. (1989). A galactosyltransferase from the fission yeast Schizosaccharomyces pombe. J. Cell Biol. *109*, 2693-2702.

Cole,N.B. and Lippincott-Schwartz,J. (1995). Organization of organelles and membrane traffic by microtubules. Curr. Opin. Cell Biol. *7*, 55-64.

Corti,A., Fassina,G., Marcucci,F., Barbanti,E., and Cassani,G. (1992). Oligomeric tumour necrosis factor alpha slowly converts into inactive forms at bioactive levels. Biochem. J. *284 (Pt 3)*, 905-910.

Corzo,D., Salazar,M., Granja,C.B., and Yunis,E.J. (1995). Advances in HLA genetics. Exp. Clin. Immunogenet. *12*, 156-170.

Craigie,R., Mizuuchi,M., and Mizuuchi,K. (1984). Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell *39*, 387-394.

Dalum,I., Butler,D.M., Jensen,M.R., Hindersson,P., Steinaa,L., Waterston,A.M., Grell,S.N., Feldmann,M., Elsner,H.I., and Mouritsen,S. (1999). Therapeutic antibodies elicited by immunization against TNF-alpha. Nat. Biotechnol. *17*, 666-669.

Dalum,I., Jensen,M.R., Hindersson,P., Elsner,H.I., and Mouritsen,S. (1996). Breaking of B cell tolerance toward a highly conserved self protein. J. Immunol. *157*, 4796-4804.

Darsow,T., Odorizzi,G., and Emr,S.D. (2000). Invertase fusion proteins for analysis of protein trafficking in yeast. Methods Enzymol. *327*, 95-106.

Darzins,A., Kent,N.E., Buckwalter,M.S., and Casadaban,M.J. (1988). Bacteriophage Mu sites required for transposition immunity. Proc. Natl. Acad. Sci. U. S. A *85*, 6826-6830.

Das,U.N. (1991). Tumour necrosis factor/cachectin: biology and relevance to disease. J. Assoc. Physicians India *39*, 201-204.

DeBoy,R.T. and Craig,N.L. (1996). Tn7 transposition as a probe of cis interactions between widely separated (190 kilobases apart) DNA sites in the Escherichia coli chromosome. J. Bacteriol. *178*, 6184-6191.

Deckert-Schluter,M., Bluethmann,H., Rang,A., Hof,H., and Schluter,D. (1998). Crucial role of TNF receptor type 1 (p55), but not of TNF receptor type 2 (p75), in murine toxoplasmosis. J. Immunol. *160*, 3427-3436.

Dieci,G., Percudani,R., Giuliodori,S., Bottarelli,L., and Ottonello,S. (2000). TFIIIC-independent in vitro transcription of yeast tRNA genes. J. Mol. Biol. *299*, 601-613.

Dinarello,C.A. (2000). Proinflammatory cytokines. Chest *118*, 503-508.

Dubey,D.D., Zhu,J., Carlson,D.L., Sharma,K., and Huberman,J.A. (1994). Three ARS elements contribute to the ura4 replication origin region in the fission yeast, Schizosaccharomyces pombe. EMBO J. *13*, 3638-3647.

Eck,M.J. and Sprang,S.R. (1989). The structure of tumor necrosis factor-alpha at 2.6 A resolution. Implications for receptor binding. J. Biol. Chem. *264*, 17595-17605.

Ellgaard,L. and Frickel,E.M. (2003). Calnexin, calreticulin, and ERp57: teammates in glycoprotein folding. Cell Biochem. Biophys. *39*, 223-247.

Ellgaard,L. and Helenius,A. (2003). Quality control in the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *4*, 181-191.

Elliott,M.J., Maini,R.N., Feldmann,M., Long-Fox,A., Charles,P., Katsikis,P., Brennan,F.M., Walker,J., Bijl,H., Ghrayeb,J., and . (1993). Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis Rheum. *36*, 1681-1690.

Engelman,A., Mizuuchi,K., and Craigie,R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211-1221.

Feldman,D.E. and Frydman,J. (2000). Protein folding in vivo: the importance of molecular chaperones. Curr. Opin. Struct. Biol. *10*, 26-33.

Feldman,M., Taylor,P., Paleolog,E., Brennan,F.M., and Maini,R.N. (1998). Anti-TNF alpha therapy is useful in rheumatoid arthritis and Crohn's disease: analysis of the mechanism of action predicts utility in other diseases. Transplant. Proc. *30*, 4126-4127.

Feldmann,M. and Steinman,L. (2005). Design of effective immunotherapy for human autoimmunity. Nature *435*, 612-619.

Flynn,J.L., Goldstein,M.M., Chan,J., Triebold,K.J., Pfeffer,K., Lowenstein,C.J., Schreiber,R., Mak,T.W., and Bloom,B.R. (1995). Tumor necrosis factor-alpha is required in the protective immune response against Mycobacterium tuberculosis in mice. Immunity. *2*, 561-572.

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

Giannakouros,T., Armstrong,J., and Magee,A.I. (1992). Protein prenylation in Schizosaccharomyces pombe. FEBS Lett. *297*, 103-106.

Giga-Hama,Y. and Kumagai,H. (1999). Expression system for foreign genes using the fission yeast Schizosaccharomyces pombe. Biotechnol. Appl. Biochem. *30 (Pt 3)*, 235-244.

Goldhaber-Gordon,I., Early,M.H., and Baker,T.A. (2003). The terminal nucleotide of the Mu genome controls catalysis of DNA strand transfer. Proc. Natl. Acad. Sci. U. S. A *100*, 7509-7514.

Goryshin,I.Y. and Reznikoff,W.S. (1998). Tn5 in vitro transposition. J. Biol. Chem. *273*, 7367-7374.

Grell,M., Douni,E., Wajant,H., Lohden,M., Clauss,M., Maxeiner,B., Georgopoulos,S., Lesslauer,W., Kollias,G., Pfizenmaier,K., and Scheurich,P. (1995). The transmembrane form of tumor necrosis factor is the prime activating ligand of the 80 kDa tumor necrosis factor receptor. Cell *83*, 793-802.

Griffin,T.J., Parsons,L., Leschziner,A.E., DeVost,J., Derbyshire,K.M., and Grindley,N.D. (1999). In vitro transposition of Tn552: a tool for DNA sequencing and mutagenesis. Nucleic Acids Res. *27*, 3859-3865.

Haapa,S., Taira,S., Heikkinen,E., and Savilahti,H. (1999). An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res. *27*, 2777-2784.

Haapa-Paananen,S., Rita,H., and Savilahti,H. (2002). DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. J. Biol. Chem. *277*, 2843-2851.

Hagemann,A.T. and Craig,N.L. (1993). Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. Genetics *133*, 9-16.

Hagihara,Y. and Kim,P.S. (2002). Toward development of a screen to identify randomly encoded, foldable sequences. Proc. Natl. Acad. Sci. U. S. A *99*, 6619-6624.

Han,S., Zheng,B., Schatz,D.G., Spanopoulou,E., and Kelsoe,G. (1996). Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells. Science *274*, 2094-2097.

Hartley,S.B., Cooke,M.P., Fulcher,D.A., Harris,A.W., Cory,S., Basten,A., and Goodnow,C.C. (1993). Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death. Cell *72*, 325-335.

Hauer,B. and Shapiro,J.A. (1984). Control of Tn7 transposition. Mol. Gen. Genet. *194*, 149-158.

Hauri,H.P. and Schweizer,A. (1992). The endoplasmic reticulum-Golgi intermediate compartment. Curr. Opin. Cell Biol. *4*, 600-608.

Hayes,F. and Hallet,B. (2000). Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. Trends Microbiol. *8*, 571-577.

Hayles,J. and Nurse,P. (1992). Genetics of the fission yeast Schizosaccharomyces pombe. Annu. Rev. Genet. *26*, 373-402.

Heffron,F., So,M., and McCarthy,B.J. (1978). In vitro mutagenesis of a circular DNA molecule by using synthetic restriction sites. Proc. Natl. Acad. Sci. U. S. A *75*, 6012-6016.

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

Hershko,A. and Ciechanover,A. (1998). The ubiquitin system. Annu. Rev. Biochem. *67*, 425-479.

Hertz,M. and Nemazee,D. (1998). Receptor editing and commitment in B lymphocytes. Curr. Opin. Immunol. *10*, 208-213.

Hikida,M., Mori,M., Takai,T., Tomochika,K., Hamatani,K., and Ohmori,H. (1996). Reexpression of RAG-1 and RAG-2 genes in activated mature mouse B cells. Science *274*, 2092-2094.

Hirsch,C., Jarosch,E., Sommer,T., and Wolf,D.H. (2004). Endoplasmic reticulum-associated protein degradation--one model fits all? Biochim. Biophys. Acta *1695*, 215-223.

Hitzeman,R.A., Chen,C.Y., Dowbenko,D.J., Renz,M.E., Liu,C., Pai,R., Simpson,N.J., Kohr,W.J., Singh,A., Chisholm,V., and . (1990). Use of heterologous and homologous signal sequences for secretion of heterologous proteins from yeast. Methods Enzymol. *185*, 421-440.

Jacobs,K.A., Collins-Racie,L.A., Colbert,M., Duckett,M., Evans,C., Golden-Fleet,M., Kelleher,K., Kriz,R., La Vallie,E.R., Merberg,D., Spaulding,V., Stover,J., Williamson,M.J., and McCoy,J.M. (1999). A genetic selection for isolating cDNA clones that encode signal peptides. Methods Enzymol. *303*, 468-479.

Jacobs,K.A., Collins-Racie,L.A., Colbert,M., Duckett,M., Golden-Fleet,M., Kelleher,K., Kriz,R., LaVallie,E.R., Merberg,D., Spaulding,V., Stover,J., Williamson,M.J., and McCoy,J.M. (1997). A genetic selection for isolating cDNAs encoding secreted proteins. Gene *198*, 289-296.

Jacobs,M., Wnendt,S., and Stahl,U. (1990). High-efficiency electro-transformation of Escherichia coli with DNA from ligation mixtures. Nucleic Acids Res. *18*, 1653.

Jaenicke,R. (1991). Protein folding: local structures, domains, subunits, and assemblies. Biochemistry *30*, 3147-3161.

Jakubowicz,T., Cytrynska,M., Kowalczyk,W., and Gasior,E. (1993). Phosphorylation of acidic ribosomal proteins by ribosome-associated protein kinases of Saccharomyces cerevisiae and Schizosaccharomyces pombe. Acta Biochim. Pol. *40*, 497-505.

Johnson,A.E. and van Waes,M.A. (1999). The translocon: a dynamic gateway at the ER membrane. Annu. Rev. Cell Dev. Biol. *15*, 799-842.

Johnson,L.M., Bankaitis,V.A., and Emr,S.D. (1987). Distinct sequence determinants direct intracellular sorting and modification of a yeast vacuolar protease. Cell *48*, 875-885.

Jones,D.D. (2005). Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion. Nucleic Acids Res. *33*, e80.

Jones,E.Y., Stuart,D.I., and Walker,N.P. (1989). Structure of tumour necrosis factor. Nature *338*, 225-228.

Jones,E.Y., Stuart,D.I., and Walker,N.P. (1990). The structure of tumour necrosis factor--implications for biological function. J. Cell Sci. Suppl *13*, 11-18.

Jones,R.H., Moreno,S., Nurse,P., and Jones,N.C. (1988). Expression of the SV40 promoter in fission yeast: identification and characterization of an AP-1-like factor. Cell *53*, 659-667.

Juji,T., Hertz,M., Aoki,K., Horie,D., Ohya,K., Gautam,A., Mouritsen,S., Oda,H., Nakamura,K., and Tanaka,S. (2002). A novel therapeutic vaccine approach, targeting RANKL, prevents bone destruction in bone-related disorders. J. Bone Miner. Metab *20*, 266-268.

Kaiser,C.A., Preuss,D., Grisafi,P., and Botstein,D. (1987). Many random sequences functionally replace the secretion signal sequence of yeast invertase. Science *235*, 312-317.

Kaufer,N.F., Simanis,V., and Nurse,P. (1985). Fission yeast Schizosaccharomyces pombe correctly excises a mammalian RNA transcript intervening sequence. Nature *318*, 78-80.

Keenan,R.J., Freymann,D.M., Stroud,R.M., and Walter,P. (2001). The signal recognition particle. Annu. Rev. Biochem. *70*, 755-775.

Kjaerulff,S. and Jensen,M.R. (2005). Comparison of different signal peptides for secretion of heterologous proteins in fission yeast. Biochem. Biophys. Res. Commun. *336*, 974-982.

Klee,E.W. and Ellis,L.B. (2005). Evaluating eukaryotic secreted protein prediction. BMC. Bioinformatics. *6*, 256.

Klein,R.D., Gu,Q., Goddard,A., and Rosenthal,A. (1996). Selection for genes encoding secreted proteins and receptors. Proc. Natl. Acad. Sci. U. S. A *93*, 7108-7113.

Klionsky,D.J., Banta,L.M., and Emr,S.D. (1988). Intracellular sorting and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targeting information. Mol. Cell Biol. *8*, 2105-2116.

Kobryn,K., Lavoie,B.D., and Chaconas,G. (1999). Supercoiling-dependent site-specific binding of HU to naked Mu DNA. J. Mol. Biol. *289*, 777-784.

Koegl,M., Courtneidge,S.A., and Superti-Furga,G. (1995). Structural requirements for the efficient regulation of the Src protein tyrosine kinase by Csk. Oncogene *11*, 2317-2329.

Kourilsky,P., Chaouat,G., Rabourdin-Combe,C., and Claverie,J.M. (1987). Working principles in the immune system implied by the "peptidic self" model. Proc. Natl. Acad. Sci. U. S. A *84*, 3400-3404.

Kowalski,J.M., Parekh,R.N., Mao,J., and Wittrup,K.D. (1998). Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. J. Biol. Chem. *273*, 19453-19458.

Kriegler,M., Perez,C., DeFay,K., Albert,I., and Lu,S.D. (1988). A novel form of TNF/cachectin is a cell surface cytotoxic transmembrane protein: ramifications for the complex physiology of TNF. Cell *53*, 45-53.

Kumar,R. and Singh,J. (2004). Expression and secretion of a prokaryotic protein streptokinase without glycosylation and degradation in Schizosaccharomyces pombe. Yeast *21*, 1343-1358.

Kuo,C.F., Zou,A.H., Jayaram,M., Getzoff,E., and Harshey,R. (1991). DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. EMBO J. *10*, 1585-1591.

Lampe,D.J., Akerley,B.J., Rubin,E.J., Mekalanos,J.J., and Robertson,H.M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. Proc. Natl. Acad. Sci. U. S. A *96*, 11428-11433.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczky,J., LeVine,R., McEwan,P., McKernan,K., Meldrim,J., Mesirov,J.P., Miranda,C., Morris,W., Naylor,J., Raymond,C., Rosetti,M., Santos,R., Sheridan,A., Sougnez,C., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Wyman,D., Rogers,J., Sulston,J., Ainscough,R., Beck,S., Bentley,D., Burton,J., Clee,C., Carter,N., Coulson,A., Deadman,R., Deloukas,P., Dunham,A., Dunham,I., Durbin,R., French,L., Grafham,D., Gregory,S., Hubbard,T., Humphray,S., Hunt,A., Jones,M., Lloyd,C., McMurray,A., Matthews,L., Mercer,S., Milne,S., Mullikin,J.C., Mungall,A., Plumb,R., Ross,M., Shownkeen,R., Sims,S., Waterston,R.H., Wilson,R.K., Hillier,L.W., McPherson,J.D., Marra,M.A., Mardis,E.R., Fulton,L.A., Chinwalla,A.T., Pepin,K.H., Gish,W.R., Chissoe,S.L., Wendl,M.C., Delehaunty,K.D., Miner,T.L., Delehaunty,A., Kramer,J.B., Cook,L.L., Fulton,R.S., Johnson,D.L., Minx,P.J., Clifton,S.W., Hawkins,T., Branscomb,E., Predki,P., Richardson,P., Wenning,S., Slezak,T., Doggett,N., Cheng,J.F., Olsen,A., Lucas,S., Elkin,C., Uberbacher,E., Frazier,M., Gibbs,R.A., Muzny,D.M., Scherer,S.E., Bouck,J.B., Sodergren,E.J., Worley,K.C., Rives,C.M., Gorrell,J.H., Metzker,M.L., Naylor,S.L., Kucherlapati,R.S., Nelson,D.L., Weinstock,G.M., Sakaki,Y., Fujiyama,A., Hattori,M., Yada,T., Saurin,W., Artiguenave,F., Brottier,P., Bruls,T., Pelletier,E., Robert,C., Wincker,P., Smith,D.R., Doucette-Stamm,L., Rubenfield,M., Weinstock,K., Lee,H.M., Dubois,J., Rosenthal,A., Platzer,M., Nyakatura,G., Taudien,S., Rump,A., Yang,H., Yu,J., Wang,J., Huang,G., Gu,J., Hood,L., Rowen,L., Madan,A., Qin,S., Davis,R.W., Federspiel,N.A., Abola,A.P., Proctor,M.J., Myers,R.M., Schmutz,J., Dickson,M., Grimwood,J., Cox,D.R., Olson,M.V., Kaul,R., Raymond,C., Shimizu,N., Kawasaki,K., Minoshima,S., Evans,G.A., Athanasiou,M., Schultz,R., Roe,B.A., Chen,F., Pan,H., Ramser,J., Lehrach,H., Reinhardt,R., McCombie,W.R., de la,B.M., Dedhia,N., Blocker,H., Hornischer,K., Nordsiek,G., Agarwala,R., Aravind,L., Bailey,J.A., Bateman,A., Batzoglou,S., Birney,E., Bork,P., Brown,D.G., Burge,C.B., Cerutti,L., Chen,H.C., Church,D., Clamp,M., Copley,R.R., Doerks,T., Eddy,S.R., Eichler,E.E., Furey,T.S., Galagan,J., Gilbert,J.G., Harmon,C., Hayashizaki,Y., Haussler,D., Hermjakob,H., Hokamp,K., Jang,W., Johnson,L.S., Jones,T.A., Kasif,S., Kaspryzk,A., Kennedy,S., Kent,W.J., Kitts,P., Koonin,E.V., Korf,I., Kulp,D., Lancet,D., Lowe,T.M., McLysaght,A., Mikkelsen,T., Moran,J.V., Mulder,N., Pollara,V.J., Ponting,C.P., Schuler,G., Schultz,J., Slater,G., Smit,A.F., Stupka,E., Szustakowski,J., Thierry-Mieg,D., Thierry-Mieg,J., Wagner,L., Wallis,J., Wheeler,R., Williams,A., Wolf,Y.I., Wolfe,K.H., Yang,S.P., Yeh,R.F., Collins,F., Guyer,M.S., Peterson,J., Felsenfeld,A., Wetterstrand,K.A., Patrinos,A., Morgan,M.J., de Jong,P., Catanese,J.J., Osoegawa,K., Shizuya,H., Choi,S., and Chen,Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lanzavecchia,A. (1985). Antigen-specific interaction between T and B cells. Nature *314*, 537-539.

Lanzavecchia,A. and Bove,S. (1985). Specific B lymphocytes efficiently pick up, process and present antigen to T cells. Behring Inst. Mitt. 82-87.

Lavoie,B.D. and Chaconas,G. (1996). Transposition of phage Mu DNA. Curr. Top. Microbiol. Immunol. *204*, 83-102.

Lee,C.H., Bhagwat,A., and Heffron,F. (1983). Identification of a transposon Tn3 sequence required for transposition immunity. Proc. Natl. Acad. Sci. U. S. A *80*, 6765-6769.

Leonardi,C.L., Powers,J.L., Matheson,R.T., Goffe,B.S., Zitnik,R., Wang,A., and Gottlieb,A.B. (2003). Etanercept as monotherapy in patients with psoriasis. N. Engl. J. Med. *349*, 2014-2022.

Liu,Y.J., Johnson,G.D., Gordon,J., and MacLennan,I.C. (1992). Germinal centres in T-cell-dependent antibody responses. Immunol. Today *13*, 17-21.

Lochs,H., Adler,G., Beglinger,C., Duchmann,R., Emmrich,J., Ewe,K., Gangl,A., Gasche,C., Hahn,E., Hoffmann,P., Kaskas,B., Malchow,H., Pohl,C., Raedler,A., Renner,E., Scholmerich,J., Schreiber,S., Stange,E., Tilg,H., Vogelsang,H., Weigert,N., and Zeitz,M. (1999). Anti-TNF antibody in Crohn's

disease--status of information, comments and recommendations of an international working group. Z. Gastroenterol. *37*, 509-512.

Locksley,R.M., Killeen,N., and Lenardo,M.J. (2001). The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell *104*, 487-501.

Luettig,B., Decker,T., and Lohmann-Matthes,M.L. (1989). Evidence for the existence of two forms of membrane tumor necrosis factor: an integral protein and a molecule attached to its receptor. J. Immunol. *143*, 4034-4038.

Machold,K.P. and Smolen,J.S. (2003). Adalimumab - a new TNF-alpha antibody for treatment of inflammatory joint disease. Expert. Opin. Biol. Ther. *3*, 351-360.

Mahillon,J. and Chandler,M. (1998). Insertion sequences. Microbiol. Mol. Biol. Rev. *62*, 725-774.

Makrides,S.C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol. Rev. *60*, 512-538.

Mannel,D.N and Echtenacher,B (2000). TNF in the inflammatory response. Chem. Immunol. *74*, 141-161.

Marino,M.W., Dunn,A., Grail,D., Inglese,M., Noguchi,Y., Richards,E., Jungbluth,A., Wada,H., Moore,M., Williamson,B., Basu,S., and Old,L.J. (1997). Characterization of tumor necrosis factor-deficient mice. Proc. Natl. Acad. Sci. U. S. A *94*, 8093-8098.

Marquardt,T., Hebert,D.N., and Helenius,A. (1993). Post-translational folding of influenza hemagglutinin in isolated endoplasmic reticulum-derived microsomes. J. Biol. Chem. *268*, 19618-19625.

Marquardt,T. and Helenius,A. (1992). Misfolding and aggregation of newly synthesized proteins in the endoplasmic reticulum. J. Cell Biol. *117*, 505-513.

Martoglio,B. and Dobberstein,B. (1998). Signal sequences: more than just greasy peptides. Trends Cell Biol. *8*, 410-415.

Matlack,K.E, Mothes,W, and Rapoport,T.A (1998). Protein translocation: tunnel vision. Cell *92*, 381-390.

Maundrell,K., Hutchison,A., and Shall,S. (1988). Sequence analysis of ARS elements in fission yeast. EMBO J. *7*, 2203-2209.

Maxwell,K.L., Mittermaier,A.K., Forman-Kay,J.D., and Davidson,A.R. (1999). A simple in vivo assay for increased protein solubility. Protein Sci. *8*, 1908-1911.

Meusser,B., Hirsch,C., Jarosch,E., and Sommer,T. (2005). ERAD: the long road to destruction. Nat. Cell Biol. *7*, 766-772.

Michelsen,B.K. (1995). Transformation of Escherichia coli increases 260-fold upon inactivation of T4 DNA ligase. Anal. Biochem. *225*, 172-174.

Miller,E.A., Beilharz,T.H., Malkus,P.N., Lee,M.C., Hamamoto,S., Orci,L., and Schekman,R. (2003). Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. Cell *114*, 497-509.

Mizuuchi,K. (1992). Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu. Rev. Biochem. *61*, 1011-1051.

Mizuuchi,K. and Adzuma,K. (1991). Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. Cell *66*, 129-140.

Mizuuchi,M. and Mizuuchi,K. (1993). Target site selection in transposition of phage Mu. Cold Spring

Harb. Symp. Quant. Biol. *58*, 515-523.

Moreno,M.B., Duran,A., and Ribas,J.C. (2000). A family of multifunctional thiamine-repressible expression vectors for fission yeast. Yeast *16*, 861-872.

Moreno,S., Klar,A., and Nurse,P. (1991). Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. Methods Enzymol. *194*, 795-823.

Moreno,S., Ruiz,T., Sanchez,Y., Villanueva,J.R., and Rodriguez,L. (1985). Subcellular localization and glycoprotein nature of the invertase from the fission yeast Schizosaccharomyces pombe. Arch. Microbiol. *142*, 370-374.

Moreno,S., Sanchez,Y., and Rodriguez,L. (1990). Purification and characterization of the invertase from Schizosaccharomyces pombe. A comparative analysis with the invertase from Saccharomyces cerevisiae. Biochem. J. *267*, 697-702.

Mossessova,E., Bickford,L.C., and Goldberg,J. (2003). SNARE selectivity of the COPII coat. Cell *114*, 483-495.

Mouritsen,S., Hansen,A.S., Petersen,B.L., and Buus,S. (1992). pH dependence of the interaction between immunogenic peptides and MHC class II molecules. Evidence for an acidic intracellular compartment being the organelle of interaction. J. Immunol. *148*, 1438-1444.

Murakami,H., Hohsaka,T., and Sisido,M. (2002). Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat. Biotechnol. *20*, 76-81.

Nakayama,C., Teplow,D.B., and Harshey,R.M. (1987). Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. Proc. Natl. Acad. Sci. U. S. A *84*, 1809-1813.

Narhi,L.O., Philo,J.S., Li,T., Zhang,M., Samal,B., and Arakawa,T. (1996). Induction of alpha-helix in the beta-sheet protein tumor necrosis factor-alpha: acid-induced denaturation. Biochemistry *35*, 11454-11460.

Natanson,C., Eichenholz,P.W., Danner,R.L., Eichacker,P.Q., Hoffman,W.D., Kuo,G.C., Banks,S.M., MacVittie,T.J., and Parrillo,J.E. (1989). Endotoxin and tumor necrosis factor challenges in dogs simulate the cardiovascular profile of human septic shock. J. Exp. Med. *169*, 823-832.

Naumann,T.A., Goryshin,I.Y., and Reznikoff,W.S. (2002). Production of combinatorial libraries of fused genes by sequential transposition reactions. Nucleic Acids Res. *30*, e119.

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

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

Ngsee,J.K., Hansen,W., Walter,P., and Smith,M. (1989). Cassette mutagenic analysis of the yeast invertase signal peptide: effects on protein translocation. Mol. Cell Biol. *9*, 3400-3410.

Nielsen,F.S., Sauer,J., Backlund,J., Voldborg,B., Gregorius,K., Mouritsen,S., and Bratt,T. (2004). Insertion of foreign T cell epitopes in human tumor necrosis factor alpha with minimal effect on protein structure and biological activity. J. Biol. Chem. *279*, 33593-33600.

Okuno,Y., Okazaki,T., and Masukata,H. (1997). Identification of a predominant replication origin in fission yeast. Nucleic Acids Res. *25*, 530-537.

Old,L.J. (1988). Tumor necrosis factor. Sci. Am. *258*, 59-75.

Olesen,K. and Kielland-Brandt,M.C. (1993). Altering substrate preference of carboxypeptidase Y by a novel strategy of mutagenesis eliminating wild type background. Protein Eng *6*, 409-415.

Otte,S. and Barlowe,C. (2004). Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. Nat. Cell Biol. *6*, 1189-1194.

Palade,G. (1975). Intracellular aspects of the process of protein synthesis. Science *189*, 347-358.

Panina-Bordignon,P., Tan,A., Termijtelen,A., Demotz,S., Corradin,G., and Lanzavecchia,A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. Eur. J. Immunol. *19*, 2237-2242.

Paravicini,G., Horazdovsky,B.F., and Emr,S.D. (1992). Alternative pathways for the sorting of soluble vacuolar proteins in yeast: a vps35 null mutant missorts and secretes only a subset of vacuolar hydrolases. Mol. Biol. Cell *3*, 415-427.

Parodi,A.J. (2000). Protein glucosylation and its role in protein folding. Annu. Rev. Biochem. *69*, 69-93.

Perlman,D., Raney,P., and Halvorson,H.O. (1986). Mutations affecting the signal sequence alter synthesis and secretion of yeast invertase. Proc. Natl. Acad. Sci. U. S. A *83*, 5033-5037.

Petrounia,I.P. and Arnold,F.H. (2000). Designed evolution of enzymatic properties. Curr. Opin. Biotechnol. *11*, 325-330.

Phillipson,B.A., Pimpl,P., daSilva,L.L., Crofts,A.J., Taylor,J.P., Movafeghi,A., Robinson,D.G., and Denecke,J. (2001). Secretory bulk flow of soluble proteins is efficient and COPII dependent. Plant Cell *13*, 2005-2020.

Plemper,R.K., Bohmler,S., Bordallo,J., Sommer,T., and Wolf,D.H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. Nature *388*, 891-895.

Polard,P. and Chandler,M. (1995). Bacterial transposases and retroviral integrases. Mol. Microbiol. *15*, 13-23.

Rajewsky,K., Schirrmacher,V., Nase,S., and Jerne,N.K. (1969). The requirement of more than one antigenic determinant for immunogenicity. J. Exp. Med. *129*, 1131-1143.

Renard,V., Sonderbye,L., Ebbehoj,K., Rasmussen,P.B., Gregorius,K., Gottschalk,T., Mouritsen,S., Gautam,A., and Leach,D.R. (2003). HER-2 DNA and protein vaccines containing potent Th cell epitopes induce distinct protective and therapeutic antitumor responses in HER-2 transgenic mice. J. Immunol. *171*, 1588-1595.

Rice,P. and Mizuuchi,K. (1995). Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell *82*, 209-220.

Ritter,C. and Helenius,A. (2000). Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. Nat. Struct. Biol. *7*, 278-280.

Ritter,C., Quirin,K., Kowarik,M., and Helenius,A. (2005). Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. EMBO J. *24*, 1730-1738.

Roach,D.R., Bean,A.G., Demangel,C., France,M.P., Briscoe,H., and Britton,W.J. (2002). TNF regulates chemokine induction essential for cell recruitment, granuloma formation, and clearance of mycobacterial infection. J. Immunol. *168*, 4620-4627.

Robinson,M.K., Bennett,P.M., and Richmond,M.H. (1977). Inhibition of TnA translocation by TnA. J. Bacteriol. *129*, 407-414.

Rock,K.L., Benacerraf,B., and Abbas,A.K. (1984). Antigen presentation by hapten-specific B lymphocytes. I. Role of surface immunoglobulin receptors. J. Exp. Med. *160*, 1102-1113.

Rothe,J., Mackay,F., Bluethmann,H., Zinkernagel,R., and Lesslauer,W. (1994). Phenotypic analysis of TNFR1-deficient mice and characterization of TNFR1-deficient fibroblasts in vitro. Circ. Shock *44*, 51-

56.

Rothstein,R.J. (1983). One-step gene disruption in yeast. Methods Enzymol. *101*, 202-211.

Ruddle,N.H. (1992). Tumor necrosis factor (TNF-alpha) and lymphotoxin (TNF-beta). Curr. Opin. Immunol. *4*, 327-332.

Ruddock,L.W., Freedman,R.B., and Klappa,P. (2000). Specificity in substrate binding by protein folding catalysts: tyrosine and tryptophan residues are the recognition motifs for the binding of peptides to the pancreas-specific protein disulfide isomerase PDIp. Protein Sci. *9*, 758-764.

Russell,P. and Nurse,P. (1986). Schizosaccharomyces pombe and Saccharomyces cerevisiae: a look at yeasts divided. Cell *45*, 781-782.

Saalbach,G., Jung,R., Kunze,G., Saalbach,I., Adler,K., and Muntz,K. (1991). Different legumin protein domains act as vacuolar targeting signals. Plant Cell *3*, 695-708.

Salama,N.R., Yeung,T., and Schekman,R.W. (1993). The Sec13p complex and reconstitution of vesicle budding from the ER with purified cytosolic proteins. EMBO J. *12*, 4073-4082.

Saraste,J. and Kuismanen,E. (1992). Pathways of protein sorting and membrane traffic between the rough endoplasmic reticulum and the Golgi complex. Semin. Cell Biol. *3*, 343-355.

Savilahti,H. and Mizuuchi,K. (1996). Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. Cell *85*, 271-280.

Savilahti,H., Rice,P.A., and Mizuuchi,K. (1995). The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J. *14*, 4893-4903.

Schubert,U., Anton,L.C., Gibbs,J., Norbury,C.C., Yewdell,J.W., and Bennink,J.R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature *404*, 770-774.

Sheridan,D.L., Berlot,C.H., Robert,A., Inglis,F.M., Jakobsdottir,K.B., Howe,J.R., and Hughes,T.E. (2002). A new way to rapidly create functional, fluorescent fusion proteins: random insertion of GFP with an in vitro transposition reaction. BMC. Neurosci. *3*, 7.

Siam,R., Dolan,W.P., and Forsburg,S.L. (2004). Choosing and using Schizosaccharomyces pombe plasmids. Methods *33*, 189-198.

Singh,I.R., Crowley,R.A., and Brown,P.O. (1997). High-resolution functional mapping of a cloned gene by genetic footprinting. Proc. Natl. Acad. Sci. U. S. A *94*, 1304-1309.

Sitia,R. and Braakman,I. (2003). Quality control in the endoplasmic reticulum protein factory. Nature *426*, 891-894.

Sitia,R., Neuberger,M., Alberini,C., Bet,P., Fra,A., Valetti,C., Williams,G., and Milstein,C. (1990). Developmental regulation of IgM secretion: the role of the carboxy-terminal cysteine. Cell *60*, 781-790.

Smith,D.G. and Svoboda,A. (1972). Golgi apparatus in normal cells and protoplasts of Schizosaccharomyces pombe. Microbios *5*, 177-182.

Sommer,T. and Wolf,D.H. (1997). Endoplasmic reticulum degradation: reverse protein flow of no return. FASEB J. *11*, 1227-1233.

Spee,P., Subjeck,J., and Neefjes,J. (1999). Identification of novel peptide binding proteins in the endoplasmic reticulum: ERp72, calnexin, and grp170. Biochemistry *38*, 10559-10566.

Springer,S., Chen,E., Duden,R., Marzioch,M., Rowley,A., Hamamoto,S., Merchant,S., and Schekman,R. (2000). The p24 proteins are not essential for vesicular transport in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A *97*, 4034-4039.

Stemmer,W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389-391.

Sugarman,B.J., Aggarwal,B.B., Hass,P.E., Figari,I.S., Palladino,M.A., Jr., and Shepard,H.M. (1985). Recombinant human tumor necrosis factor-alpha: effects on proliferation of normal and transformed cells in vitro. Science *230*, 943-945.

Tabuchi,M., Iwaihara,O., Ohtani,Y., Ohuchi,N., Sakurai,J., Morita,T., Iwahara,S., and Takegawa,K. (1997). Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport, and processing of carboxypeptidase Y from Schizosaccharomyces pombe. J. Bacteriol. *179*, 4179-4189.

Tague,B.W., Dickinson,C.D., and Chrispeels,M.J. (1990). A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. Plant Cell *2*, 533-546.

Tanaka,N., Fujita,Y., Suzuki,S., Morishita,M., Giga-Hama,Y., Shimoda,C., and Takegawa,K. (2005). Characterization of O-mannosyltransferase family in Schizosaccharomyces pombe. Biochem. Biophys. Res. Commun. *330*, 813-820.

Tang,P., Hung,M.-C., and Klostergaard,J. (1996). Human pro-tumor necrosis factor is a homotrimer. Biochemistry *35*, 8216-8225.

Taylor,P.C., Williams,R.O., and Feldmann,M. (2004). Tumour necrosis factor alpha as a therapeutic target for immune-mediated inflammatory diseases. Curr. Opin. Biotechnol. *15*, 557-563.

Thrower,J.S., Hoffman,L., Rechsteiner,M., and Pickart,C.M. (2000). Recognition of the polyubiquitin proteolytic signal. EMBO J. *19*, 94-102.

Tokunaga,M., Kawamura,A., Yonekyu,S., Kishida,M., and Hishinuma,F. (1993). Secretion of mouse alpha-amylase from fission yeast Schizosaccharomyces pombe: presence of chymostatin-sensitive protease activity in the culture medium. Yeast *9*, 379-387.

Tracey,K.J., Wei,H., Manogue,K.R., Fong,Y., Hesse,D.G., Nguyen,H.T., Kuo,G.C., Beutler,B., Cotran,R.S., Cerami,A., and . (1988). Cachectin/tumor necrosis factor induces cachexia, anemia, and inflammation. J. Exp. Med. *167*, 1211-1227.

Travers,K.J., Patil,C.K., Wodicka,L., Lockhart,D.J., Weissman,J.S., and Walter,P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249-258.

Tsai,B., Rodighiero,C., Lencer,W.I., and Rapoport,T.A. (2001). Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. Cell *104*, 937-948.

Tu,B.P., Ho-Schleyer,S.C., Travers,K.J., and Weissman,J.S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. Science *290*, 1571-1574.

van Bleek,G.M. and Nathenson,S.G. (1990). Isolation of an endogenously processed immunodominant viral peptide from the class I H-2Kb molecule. Nature *348*, 213-216.

Van,O., X, Tavernier,J., and Fiers,W. (1994). Structure-activity studies of human tumour necrosis factors. Protein Eng *7*, 5-22.

Vandenabeele,P., Declercq,W., Beyaert,R., and Fiers,W. (1995). Two tumour necrosis factor receptors: structure and function. Trends Cell Biol. *5*, 392-399.

Vieira,L.Q., Goldschmidt,M., Nashleanas,M., Pfeffer,K., Mak,T., and Scott,P. (1996). Mice lacking the TNF receptor p55 fail to resolve lesions caused by infection with Leishmania major, but control parasite replication. J. Immunol. *157*, 827-835.

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

von Heijne,G. (1990). The signal peptide. J. Membr. Biol. *115*, 195-201.

Wagner,T.E., Huseby,E.S., and Huseby,J.S. (2002). Exacerbation of Mycobacterium tuberculosis enteritis masquerading as Crohn's disease after treatment with a tumor necrosis factor-alpha inhibitor. Am. J. Med. *112*, 67-69.

Waldo,G.S., Standish,B.M., Berendzen,J., and Terwilliger,T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. *17*, 691-695.

Wells,J.A., Vasser,M., and Powers,D.B. (1985). Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. Gene *34*, 315-323.

Wigley,W.C., Stidham,R.D., Smith,N.M., Hunt,J.F., and Thomas,P.J. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. Nat. Biotechnol. *19*, 131-136.

Winkeler,A., Godderz,D., Herzog,V., and Schmitz,A. (2003). BiP-dependent export of cholera toxin from endoplasmic reticulum-derived microsomes. FEBS Lett. *554*, 439-442.

Wolfe,F., Michaud,K., Anderson,J., and Urbansky,K. (2004). Tuberculosis infection in patients with rheumatoid arthritis and the effect of infliximab therapy. Arthritis Rheum. *50*, 372-379.

Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S., Basham,D., Bowman,S., Brooks,K., Brown,D., Brown,S., Chillingworth,T., Churcher,C., Collins,M., Connor,R., Cronin,A., Davis,P., Feltwell,T., Fraser,A., Gentles,S., Goble,A., Hamlin,N., Harris,D., Hidalgo,J., Hodgson,G., Holroyd,S., Hornsby,T., Howarth,S., Huckle,E.J., Hunt,S., Jagels,K., James,K., Jones,L., Jones,M., Leather,S., McDonald,S., McLean,J., Mooney,P., Moule,S., Mungall,K., Murphy,L., Niblett,D., Odell,C., Oliver,K., O'Neil,S., Pearson,D., Quail,M.A., Rabbinowitsch,E., Rutherford,K., Rutter,S., Saunders,D., Seeger,K., Sharp,S., Skelton,J., Simmonds,M., Squares,R., Squares,S., Stevens,K., Taylor,K., Taylor,R.G., Tivey,A., Walsh,S., Warren,T., Whitehead,S., Woodward,J., Volckaert,G., Aert,R., Robben,J., Grymonprez,B., Weltjens,I., Vanstreels,E., Rieger,M., Schafer,M., Muller-Auer,S., Gabel,C., Fuchs,M., Dusterhoft,A., Fritzc,C., Holzer,E., Moestl,D., Hilbert,H., Borzym,K., Langer,I., Beck,A., Lehrach,H., Reinhardt,R., Pohl,T.M., Eger,P., Zimmermann,W., Wedler,H., Wambutt,R., Purnelle,B., Goffeau,A., Cadieu,E., Dreano,S., Gloux,S., Lelaure,V., Mottier,S., Galibert,F., Aves,S.J., Xiang,Z., Hunt,C., Moore,K., Hurst,S.M., Lucas,M., Rochet,M., Gaillardin,C., Tallada,V.A., Garzon,A., Thode,G., Daga,R.R., Cruzado,L., Jimenez,J., Sanchez,M., del Rey,F., Benito,J., Dominguez,A., Revuelta,J.L., Moreno,S., Armstrong,J., Forsburg,S.L., Cerutti,L., Lowe,T., McCombie,W.R., Paulsen,I., Potashkin,J., Shpakovski,G.V., Ussery,D., Barrell,B.G., and Nurse,P. (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Yanagida,M. (1990). Higher-order chromosome structure in yeast. J. Cell Sci. *96 (Pt 1)*, 1-3.

Yanagihara,K. and Mizuuchi,K. (2002). Mismatch-targeted transposition of Mu: a new strategy to map genetic polymorphism. Proc. Natl. Acad. Sci. U. S. A *99*, 11317-11321.

Yang,W. and Steitz,T.A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. Structure. *3*, 131-134.

Yanisch-Perron,C., Vieira,J., and Messing,J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*, 103-119.

Ymer,S. (1991). Heat inactivation of DNA ligase prior to electroporation increases transformation efficiency. Nucleic Acids Res. *19*, 6960.

Zapun,A., Jakob,C.A., Thomas,D.Y., and Bergeron,J.J. (1999). Protein folding in a specialized compartment: the endoplasmic reticulum. Structure. Fold. Des *7*, R173-R182.

Zhou,L., Mitra,R., Atkinson,P.W., Hickman,A.B., Dyda,F., and Craig,N.L. (2004). Transposition of hAT elements links transposable elements and V (D)J recombination. Nature *432*, 995-1001.