

# **High-Throughput Evaluation of Protein Folding Conditions and Expression Constructs for Structural Genomics**

Dissertation  
zur Erlangung des Doktorgrades  
der Naturwissenschaften (Dr. rer. nat.)  
an der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität Potsdam

vorgelegt von  
**Christoph Scheich**

**Berlin  
Mai 2004**

1. Gutachter: Prof. Dr. Robert Seckler
2. Gutachter: Prof. Dr. Hans Lehrach
3. Gutachter: PD Dr. Hauke Lilie

Die Disputation fand am 31.08.2004 statt

## Abstract

For recombinant production of proteins for structural and functional analyses, the *E. coli* expression system is the most widely used due to high yields and straightforward processing. However, particularly the expression of eukaryotic proteins in *E. coli* is often problematic, e.g. when the protein is not folded correctly and is deposited in insoluble inclusion bodies. In some cases it is favourable to analyse deletion constructs of a protein or an individual protein domain instead of the full-length protein. This implies the generation of a set of expression constructs that need to be characterised. In this work methods to optimise and evaluate *in vitro* folding of inclusion body proteins as well as high-throughput characterisation of expression constructs were developed.

Transferring inclusion body proteins to their native state involves two steps: (a) solubilisation with a chaotropic reagent or a strong ionic detergent and (b) folding of the protein by removal of the chaotrop accompanied by the transfer into an appropriate buffer. The yield of natively folded protein is often substantially reduced due to aggregation or misfolding; it may, however, be improved by certain additives to the folding buffer. These additives need to be identified empirically. In this thesis a screening procedure for folding conditions was developed. To reduce the number of possible combinations of screening additives, empirical observations documented in the literature as well as well known properties of certain screening additives were considered. To decrease the amount of protein and work invested, the screen was miniaturised and automated using a pipetting robot. Twenty rapid dilution conditions for the denatured protein are tested and two conditions for folding of proteins using the detergent/cyclodextrin protein folding system of Rozema et al. (1996). 100 µg protein is used per condition. In addition, eight conditions can be tested for folding of His-tagged proteins (approx. 200 µg) immobilised on metal chelate resins.

The screen was successfully applied to fold a human protein, the p22 subunit of dynactin that is expressed in inclusion bodies in *E. coli*. For p22 dynactin – as is the case for many proteins – there was no biological assay available to assess the success of the folding screen. Protein solubility can not be used as a stringent criterion because beside natively folded protein, soluble misfolded species and microaggregates may occur. This work evaluates methods to detect small amounts of natively folded protein after automated folding screening. Before folding screening with p22 dynactin, two model

enzymes, bovine carbonic anhydrase II (CAB) and pig heart mitochondrial malate dehydrogenase, were used for evaluation. Recovered activity after refolding was correlated to different biophysical methods. 8-anilino-1-naphtalenesulfonic acid binding-experiments gave no useful information when refolding CAB, due to low sensitivity and because misfolded protein could not be readily distinguished from native protein. Tryptophan fluorescence spectra of refolded CAB were used to assess the success of refolding. The shift of the intensity maximum to a shorter wavelength, compared to the denaturant unfolded protein, as well as the fluorescence intensity correlated to recovered enzymatic activity. For both model enzymes, analytical hydrophobic interaction chromatography (HIC) was useful to identify refolded samples that contain active enzyme. Compactly folded, active enzyme eluted in a distinct peak in a decreasing ammonium sulfate gradient. The detection limit of analytical HIC was approx. 5  $\mu$ g. In case of CAB, tryptophan fluorescence spectroscopy and analytical HIC showed that both methods in combination can be useful to rule out false positives or false negatives obtained with one method. These two methods were also useful to identify conditions for folding of p22 dynactin. However, tryptophan fluorescence spectroscopy can lead to false positives because in some cases spectra of soluble microaggregates are not well distinguishable from spectra of natively folded protein. In summary, a fast and reliable screening procedure was developed to make inclusion body proteins accessible to structural or functional analyses.

In a separate project, 88 different *E. coli* expression constructs for 17 human protein domains that had been identified by sequence analysis were analysed using high-throughput purification and folding analysis in order to obtain candidates suitable for structural analysis. After 96 deep-well microplate expression and automated protein purification, solubly expressed protein domains were directly analysed using 1D  $^1$ H-NMR spectroscopy. It was found that isolated methyl group signals below 0.5 ppm are particularly sensitive and reliable probes for folded protein. In addition – similar to the evaluation of a folding screen – analytical HIC proved to be an efficient tool for identifying constructs that yield compactly folded protein. Both methods, 1D  $^1$ H-NMR spectroscopy and analytical HIC, provided complementary results. Six constructs, representing two domains, could be quickly identified as targets that are well suitable for structural analysis. The structure of one of these domains was solved recently by co-workers, the other structure was published by another group during this project.

---

## Abbreviations

|                                     |   |
|-------------------------------------|---|
| [ $\Theta$ ]                        | molar ellipticity   |
| 1D $^1\text{H}$ NMR                 | one-dimensional $^1\text{H}$ nuclear magnetic resonance                               |
| 2D $^{15}\text{N}$ $^1\text{H}$ NMR | two-dimensional $^1\text{H}$ - $^{15}\text{N}$ heteronuclear single quantum coherence |
| Å                                   | angstrom  |
| a.u. or AU                          | arbitrary units   |
| ANSA                                | 8-anilino-1-naphthalenesulfonic acid  |
| borax                               | sodium tetraborate, anhydrous   |
| BSA                                 | bovine serum albumin  |
| CAB                                 | bovine carbonic anhydrase II (from bovine erythrocytes)                               |
| CD                                  | circular dichroism  |
| cDNA                                | complementary DNA   |
| CTAB                                | hexadecyltrimethylammonium bromide  |
| DLS                                 | dynamic light scattering  |
| DTT                                 | dithiothreitol  |
| <i>E. coli</i>                      | <i>Escherichia coli</i>   |
| EDTA                                | ethylenediaminetetraacetic acid   |
| FTIR                                | Fourier-transform infrared  |
| g                                   | acceleration due to gravity, $g = 9.81 \text{ ms}^{-2}$                               |
| GdnHCl                              | guanidine hydrochloride   |
| GSH                                 | reduced glutathione   |
| GSSG                                | oxidized glutathione  |
| HIC                                 | hydrophobic interaction chromatography  |
| HRP                                 | horse radish peroxidase   |
| HSQC                                | heteronuclear single quantum coherence  |
| IPTG                                | isopropyl $\beta$ -D-1-thiogalactopyranoside  |
| kDa                                 | kilodalton  |
| KRAB protein                        | kruppel associated box protein  |
| M                                   | mole/liter  |
| MDH                                 | malate dehydrogenase (from pig heart mitochondria)                                    |
| min                                 | minute  |
| mS                                  | millisiemens  |
| NADH                                | $\beta$ -nicotinamide adenine dinucleotide, reduced                                   |
| NiNTA                               | nickel-nitrilotriacetic acid  |
| NMR                                 | nuclear magnetic resonance  |
| PAGE                                | polyacrylamide gel electrophoresis  |
| PEG                                 | polyethyleneglycol  |
| PCR                                 | polymerase chain reaction   |
| PMSF                                | phenylmethylsulfonyl fluoride   |
| PSF                                 | Protein Structure Factory   |
| ppm                                 | parts per million   |
| rpm                                 | revolutions per minute  |
| SB                                  | super broth   |
| SDS                                 | sodium dodecyl sulfate  |
| SMART                               | simple modular architecture research tool   |
| sec                                 | second  |
| TEV                                 | tobacco etch virus  |
| Tris                                | tris(hydroxymethyl)aminomethan  |
| v/v                                 | volume per volume   |
| w/v                                 | weight per volume   |
| YT                                  | yeast tryptone  |

For amino acids the three letter or one letter standard abbreviations were used.

---

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction.....</b>  | <b>1</b>  |
| 1.1      | Protein structure and stability.....  | 2         |
| 1.2      | <i>In vitro</i> protein folding .....   | 3         |
| 1.3      | Folding screen .....  | 6         |
| 1.4      | Detection of natively folded protein upon folding screening.....  | 9         |
| 1.5      | Structural analysis of human protein domains.....   | 13        |
| 1.6      | Objectives of this work.....  | 14        |
| <b>2</b> | <b>Materials .....</b>  | <b>16</b> |
| 2.1      | Chemicals.....  | 16        |
| 2.2      | Media.....  | 17        |
| 2.3      | Miscellaneous materials .....   | 17        |
| 2.4      | Instruments.....  | 18        |
| <b>3</b> | <b>Methods.....</b>   | <b>19</b> |
| 3.1      | Bioinformatics tools.....   | 19        |
| 3.2      | Cloning .....   | 19        |
| 3.3      | Protein expression .....  | 19        |
| 3.3.1    | Expression of protein in <i>E.coli</i> for folding sceening.....  | 19        |
| 3.3.2    | Expression of human protein domains in 96 deep-well microplates .....   | 20        |
| 3.4      | Protein preparation and purification .....  | 21        |
| 3.4.1    | Perfusion chromatography.....   | 21        |
| 3.4.2    | Denaturation of model enzymes for refolding screening and unfolding transitions<br>of CAB and myoglobin ..... | 21        |
| 3.4.3    | Inclusion body preparation and denaturing His-tag purification .....  | 22        |
| 3.4.4    | Automated folding screening .....   | 22        |
| 3.4.5    | Large scale folding and purification of the p22 dynactin subunit.....   | 23        |
| 3.4.6    | Medium scale folding attempts .....   | 24        |
| 3.4.7    | Robot protein purification of human protein domains.....  | 25        |
| 3.4.8    | Preparation of <sup>15</sup> N labeled samples for 2D <sup>1</sup> H- <sup>15</sup> N-HSQC.....               | 28        |
| 3.5      | Protein analysis .....  | 28        |
| 3.5.1    | Determination of protein concentration.....   | 28        |
| 3.5.2    | SDS-polyacrylamidegelelectrophoresis (SDS-PAGE).....  | 29        |
| 3.5.3    | Tryptophan fluorescence spectroscopy.....   | 29        |
| 3.5.4    | Analytical hydrophobic interaction chromatography (HIC) .....   | 29        |

|          |  |           |
|----------|--|-----------|
| 3.5.5    | Determination of CAB and MDH activity.....   | 31        |
| 3.5.6    | ANSA fluorescence of refolded CAB.....   | 32        |
| 3.5.7    | Dynamic light scattering (DLS).....  | 32        |
| 3.5.8    | UV CD spectroscopy with p22 dynactin.....  | 33        |
| 3.5.9    | Fourier transform infrared spectroscopy with p22 dynactin.....   | 34        |
| 3.5.10   | Detection of folded protein using analytical 1D <sup>1</sup> H-NMR and 2D <sup>1</sup> H- <sup>15</sup> N-HSQC spectroscopy..... | 36        |
| <b>4</b> | <b>Result I: Protein Folding Screening .....</b>   | <b>38</b> |
| 4.1      | Screening design.....  | 38        |
| 4.2      | Evaluation of the automated folding screen and design of a detection system for folded proteins using model enzymes.....         | 41        |
| 4.2.1    | Refolding of MDH.....  | 41        |
| 4.2.2    | Tryptophan fluorescence spectroscopy of protein unfolding transitions.....   | 43        |
| 4.2.3    | Refolding of CAB.....  | 45        |
| 4.2.4    | Evaluation of ANSA fluorescence as a monitor for native protein after refolding screening of CAB.....                            | 49        |
| 4.3      | Folding screening of proteins expressed in <i>E. coli</i> inclusion bodies.....  | 50        |
| 4.3.1    | Screening for folding conditions for p22 dynactin.....   | 52        |
| 4.3.2    | Folding, purification and detailed biophysical analysis of p22 dynactin.....   | 55        |
| 4.3.3    | Unsuccessful folding screening.....  | 57        |
| <b>5</b> | <b>Results II: Analysis of human protein domains.....</b>  | <b>61</b> |
| 5.1      | Domain selection and cloning.....  | 61        |
| 5.2      | Solubility of different domain constructs.....   | 62        |
| 5.3      | Detection of natively folded protein by 1D <sup>1</sup> H-NMR and analytical HIC.....  | 63        |
| <b>6</b> | <b>Discussion .....</b>  | <b>71</b> |
| 6.1      | A new automated protein folding screen.....  | 71        |
| 6.1.1    | Design of a folding screen.....  | 71        |
| 6.1.2    | A combination of biophysical monitors for folding detection upon folding screening..   | 73        |
| 6.1.3    | Possible reasons for unsuccessful folding screens of proteins.....   | 77        |
| 6.1.4    | Alternative <i>in vivo</i> -approaches to obtain aggregation-prone protein in a soluble form                                     | 77        |
| 6.2      | Fast identification of expression constructs of human protein domains suitable for structural analysis.....                      | 78        |
| 6.3      | Concluding remark.....   | 81        |
| <b>7</b> | <b>References .....</b>  | <b>82</b> |
| <b>8</b> | <b>Supplement.....</b>   | <b>91</b> |
| <b>9</b> | <b>Publication list .....</b>  | <b>92</b> |

## 1 Introduction

Protein structure analysis has gained increasing interest in recent years. The elucidation of protein structures is important to understand protein functions and facilitate pharmaceutical drug development (Buchanan et al., 2002; Zhang & Kim, 2003). The structures are helpful to determine binding partners and functional sites of proteins (Jones & Thornton, 2004), in addition to protein-protein- or protein-ligand interaction studies. The human genome is estimated to have 35,000–45,000 genes (Pennisi, 2001). In the beginning of 2004, the ENSEMBL database (<http://www.ensembl.org>) lists three-dimensional structures for approx. 1,200 different human proteins. However, most structures only span part of the protein sequence. When structural homologues from other species are available – particularly mammalian structures – these can be used to build models for unknown human structures.

Structural information for a large number of proteins need to be determined to yield a comprehensive mechanistic understanding of cell physiology at the molecular level. In recent years a structural genomics initiative has been launched. Structural genomics projects develop and utilise techniques for high-throughput structural analysis (Heinemann et al., 2001). This thesis was carried out at the Protein Structure Factory (PSF, <http://www.proteinstrukturfabrik.de>) where large sets of human proteins are expressed in *E. coli* or yeast for structural analysis by X-ray crystallography and NMR spectroscopy. Using bioinformatics tools, target proteins are selected according to the following criteria: (i) no or little structural information available, (ii) no membrane proteins, (iii) low content of coiled-coil regions, (iv) intracellular location and (v) no low complexity regions that have extended sequence repeats with a limited repertoire of amino acids over long polypeptide segments.

A prerequisite for structure determination is the preparation of pure, natively folded protein. The expression of target proteins can be performed in different organisms, e.g. *E. coli*, yeast or insect cells. The *E. coli*-expression system is the most commonly used. Its advantages in comparison to other systems are high yields, cost effectiveness, the easy handling with respect to cloning, cultivation and protein preparation and – as a consequence – its readiness to be implemented in high-throughput processes. Its drawbacks are the failure to express a lot of recombinant proteins from higher organisms

in a native, soluble form and its inability to posttranslationally modify proteins from higher organisms. The first part of this work comprises the development of a semi-automated screening procedure to fold proteins that are expressed insolubly in *E. coli*, thereby making them available for structural and functional analyses.

At the PSF large numbers of expression constructs are prepared in different vectors and strains, either as full-length proteins or as single domains. The constructs have to be screened for soluble expression and suitable clones are selected for scale up of expression and purification. To achieve this, methods to express, purify and analyse these proteins in a parallel, time and cost effective manner had to be implemented. The second part of this work shows how such a screening procedure to analyse a large number of expression constructs for human protein domains can be carried out.

## 1.1 Protein structure and stability

Globular proteins in solution are compact structures with the interior having a density similar to that of a protein crystal. The smallest units of three-dimensional structure in a protein are referred to as secondary structure elements comprising  $\alpha$ -helices and  $\beta$ -sheets. These structures are formed through certain patterns of hydrogen bonding of the peptide backbone. A protein domain is a folding entity with a certain arrangement of secondary structures and most hydrophobic residues packed in the interior of the domain. Small proteins with only one domain exist but most proteins consist of more than one domain which are joined together to form tertiary structure. Proteins are heavily solvated (approx. 50%), and in case where hydrophilic residues are enclosed in the interior of the protein, these are usually hydrogen-bonded to water molecules, which is also called structural water. Most hydrophilic side chains protrude to the protein surface and are hydrogen bonded to water molecules that form a stabilising solvation shell around the protein. Protein structures are mainly stabilised by (i) van der Waals forces that occur between pairs of non-bonded atoms, brought close together in space as a result of attraction between transient magnetic dipoles, (ii) a hydrophobic effect that results from the gain in entropy upon transferring water-exposed hydrophobic side chains from the polar solvent into the hydrophobic interior of a protein, (iii) hydrogen bonds. Stronger forces like salt bridges between polar residues or disulfide bonds formed between cysteine residues only have a minor impact on protein stability simply because they occur with less frequency than the other stabilising forces (Jaenicke &

---

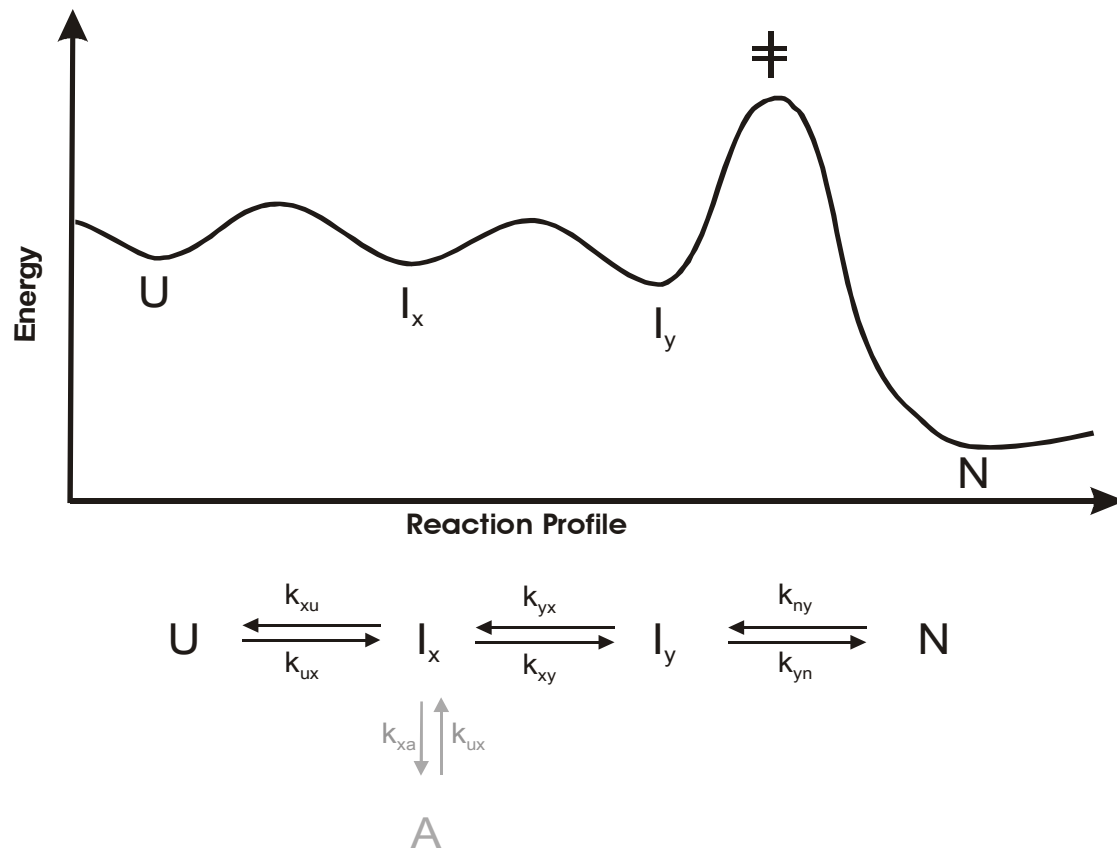
Seckler, 1997; Sheehan, 2000). Exceptions to this are small proteins that are secreted from the cell. These are commonly stabilised by disulfide bridges (Creighton et al., 1996).

On the other hand, proteins are destabilised by unfavourable interactions or by the chain entropy, which is highest in the completely unfolded molecule. In solutions with a salt composition and pH similar to physiological conditions, the folded (native) structure is thermodynamically favoured because stabilising forces are slightly stronger than the destabilising forces. Protein unfolding may be induced by exposure to extremes of temperature, pH or pressure, or by treatment with denaturants, e.g. urea or guanidine hydrochloride (GdnHCl), or detergents.

## 1.2 *In vitro* protein folding

Protein folding has been studied extensively *in vitro* in order to understand its mechanisms. Because the denatured state of a protein can be discriminated from the native state due to different spectroscopic properties, protein folding is usually followed by the change of a spectroscopic signal upon dilution or dialysis of a denaturant unfolded protein into a solution in which native protein is stable. In most cases, folding from the completely denatured state takes place in a short time interval, i.e. a few seconds or faster. From such a fast reaction it can be concluded that folding follows an ordered compulsory pathway rather than a random search for conformations (Levinthal, 1968). Small proteins may fold without the occurrence of any detectable intermediate states (Jackson & Fersht, 1991; Schindler et al., 1995). However, particularly in case of multidomain or oligomeric proteins, intermediates can be detected (Bodenreider et al., 2002; Martin & Schmid, 2003). Kinetically, a folding pathway from the unfolded state through intermediates to the folded protein may be represented as a sequence of consecutive, reversible first order reactions (see Figure 1.1). In intermediate states only parts of the interactions present in native protein have been formed. Hydrophobic collapse (compaction of hydrophobic side chains) and secondary structure formation to form the molten globule state are early events during folding and happen on the millisecond time range. Both processes were proposed to occur simultaneously (Uversky & Fink, 2002). Molten globules are intermediates that have been observed during folding of several proteins (Fink, 1995). They are almost as compact as native protein and have secondary structure but little or no tertiary structure. Because the docking of two folded

domains is a reaction that is sometimes considerably slower than hydrophobic collapse and secondary structure formation, undocked domains may be detected as intermediates (Martin & Schmid, 2003). In case of fast folding small proteins, intermediates are assumed to accelerate the folding reaction because they reduce the number of possible unfolded or partially folded conformations (Kim & Baldwin, 1990). On the other hand, intermediates may contain non-native interactions and may therefore represent a kinetic trap (Sosnick et al., 1994).



**Figure 1.1: Schematic energy diagram and kinetics of a hypothetical folding reaction and a competing aggregation reaction, adapted from Kiefhaber (1995)**

U represents the unfolded state, N is the native state,  $I_x$  and  $I_y$  are intermediates and A is an aggregate formed from intermediate  $I_x$ . The activated state of the rate limiting step during the folding reaction is indicated by ††. The aggregation reaction is not included in the energy diagram. k are rate constants for the respective reactions.

Intermediates are usually thermodynamically very unstable and have a very short half-life so that they are not readily detectable. Hence in most cases the overall (un)folding

reaction may be regarded as a co-operative transition between only two macroscopic states, the unfolded and the native state. This is reflected in Figure 1.1 showing a hypothetical folding reaction. The energy barrier of the activated transition state that needs to be surpassed in the final reaction to native protein is high compared to the energy barriers between U and the intermediates  $I_x$  and  $I_y$ . In most but not all cases folding reactions can be characterised by the depicted reaction profile. Intermediates with higher stability that are flanked by high-energy barriers may occur. It is also possible that intermediates are higher in energy than the unfolded state.

Early intermediates that not yet built up a hydrophobic core of the molten globule can be prone to aggregation due to unspecific intermolecular hydrophobic interactions. The aggregation reaction (a second or higher order reaction) competes with the folding reaction (in case of monomeric proteins a first order reaction) and is favoured at higher protein concentrations (Goldberg et al., 1991). Aggregated protein is trapped in a minimum of conformational energy distinct to the one present in native protein. Upon refolding of denaturant unfolded protein under conditions where the native state is stable, the equilibrium between intermediates and native protein or aggregates is far on the side of the latter species. Under moderate denaturing (destabilising) conditions (e.g. at low concentrations of denaturants) denatured protein is in equilibrium with native protein and/or aggregates; in some cases aggregates are more stable under these conditions and will predominate (Zettlmeissl et al., 1979). When folding these proteins, extended exposure to moderate denaturing conditions should be avoided by *fast* dilution from strongly denaturing conditions under vigorous stirring. In contrast, for other proteins aggregation is only observed under folding conditions where native protein is stable. In those cases gradual removal of the denaturant by dialysis may be superior to refolding by rapid dilution.

Only few proteins fold quantitatively to the native state under standard buffer conditions, in most cases aggregates and/or trapped intermediates or misfolded not-aggregated protein are present beside the native protein. Eventually no native protein can be obtained. The time required to complete the folding reaction can vary significantly, ranging from less than a second to several hours or even days. Certain reactions have been identified that may slow down protein folding: (i) *cis-trans* prolyl isomerisation, (ii) disulfide exchange (see 1.3, Figure 1.2) and (iii) subunit assembly in case of oligomeric proteins. *In vivo* protein folding is in many cases catalysed by chaperones. Chaperones protect the folding molecules from aggregation by shielding hydrophobic patches that

---

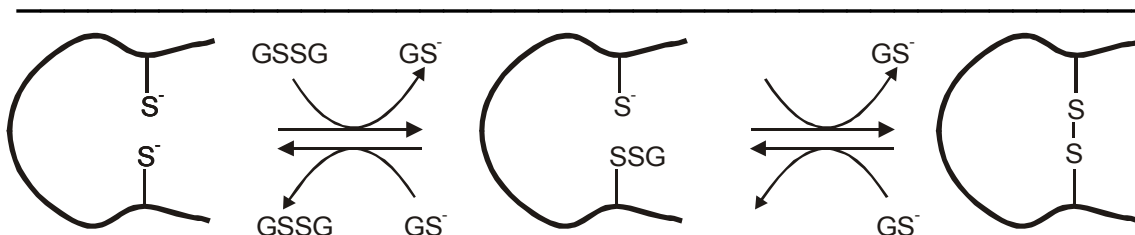
are prone to intermolecular interactions; they may also rescue misfolded proteins by unfolding them, hence giving them a new chance to fold (Hartl & Hayer-Hartl, 2002). *Cis-trans* prolyl isomerisation as well as disulfide bridging (see 1.3) is catalysed by peptidyl prolyl *cis-trans* isomerases and protein disulfide isomerases, respectively (Schiene & Fischer, 2000).

### 1.3 Folding screen

Upon heterologous over-expression in *E. coli*, many proteins accumulate in insoluble, intracellular aggregates. These inclusion bodies contain high amounts of the target protein in relatively high purity. The structures of inclusion body proteins differ from their respective native conformations, e.g. an increased level of nonnative  $\beta$ -sheets appears to be common among them (Fink, 1998). It has been proposed that inclusion body formation results from a specific aggregation reaction of a folding intermediate (featuring a specific aggregation mechanism like intermolecular  $\beta$ -sheet formation) (Przybycien et al., 1994; Speed et al., 1996). Proteins with native disulfide bonds may not be expressed solubly due to the reducing environment of the *E. coli* cytoplasm (see also below, Figure 1.2)

Inclusion body protein, solubilized by denaturants or detergents, can be folded to the native conformation by transfer into an appropriate buffer, using either dilution or dialysis. As outlined above (see 1.2), this process is in some cases prone to high losses due to misfolding and aggregation, although the addition of “folding helpers” (e.g. polyethylene glycol (PEG), glycerol, detergents, arginine) may lead to higher yields. Many other factors including pH, temperature, protein concentration or the redox milieu can influence the success of folding of a particular protein. Which additive/factor is beneficial for *in vitro* folding of a particular protein can not be reliably predicted, but needs to be determined in a screening procedure.

The beneficial effect of most folding additives as folding helpers has been found empirically. In some cases, the mechanisms by which protein folding is supported is not completely understood. Obviously, adding redox pairs like reduced glutathione (GSH) and oxidised glutathione (GSSG) to the folding buffer facilitates rapid reshuffling of incorrect disulfide bonds. Such a thiol-disulfide exchange (see Figure 1.2) is accelerated at high pH where more thiol-groups of cysteins (which have a pK value of approx. 9.3) are deprotonated.



**Figure 1.2: Thiol-disulfide exchange reaction catalysed by glutathione**

S<sup>-</sup> represents the thiolate ion of a free cysteine, GSSG denotes oxidised glutathione and GS<sup>-</sup> reduced, deprotonated glutathione.

Detailed studies about the PEG-enhanced refolding of bovine carbonic anhydrase II (CAB) indicated that PEG binds to an early intermediate and prevents self-association and subsequent aggregation (Cleland & Randolph, 1992; Cleland et al., 1992b). Arginine or urea interact directly with the protein and act by weakening nonpolar interactions and decelerating the folding reaction. In the Hofmeister series that ranks electrolytes (or osmolytes) due to their ability to salt out (precipitate) proteins, arginine and urea are rated as “salting in” (solubilising) agents. They are also referred to as chaotrops because they disturb hydrogen bonding between water molecules. The positive effect of arginine or low concentrations of urea during *in vitro* folding was proposed to be due to their weakly destabilising effect: in their presence it may be possible to shuffle aggregates back to the native state because native protein is in a lower energy minimum than aggregates (Jaenicke & Seckler, 1997). Conversely, additives like glycerol or sucrose are stabilisers of native protein. They are preferentially excluded from the water sheath of both the native and denatured states of a protein (Arakawa et al., 1990). However, glycerol interaction is indirect through strong interaction with water molecules surrounding nonpolar groups of the protein (Gekko & Timasheff, 1981). The stabilising effect of glycerol has been attributed to unfavourable interaction between the peptide backbone and the osmolyte (Bolen & Baskakov, 2001). Bolen and Baskakov (2001) proposed that this effect is stronger than favourable interactions between osmolyte and side chains. Hence the osmolyte will destabilise both the native and the unfolded state, but the unfolded state is destabilised more than the native state, because the peptide backbone is more solvent-exposed in the unfolded state. This causes the Gibbs energy of the unfolded state to be raised more than of the native state and leads to stabilisation

## Introduction

of the native protein (Bolen & Baskakov, 2001). However, this stabilisation effect can be overcome when the protein is at the same time destabilised by other factors, e.g. by denaturants or elevated temperature (Jacob et al., 1999; Jacob et al., 1997). During *in vitro* folding, glycerol has been shown to increase the rate constant of refolding whereas the rate constant of unfolding is decreased (Russo et al., 2003). Such an accelerated folding reaction may reduce the time during which aggregation-sensitive folding intermediates occur giving them less time for self-association. In contrast, under conditions where the stabilising effect of sucrose or ethylene glycol is overcome by addition of denaturants or by increased temperature, refolding rates are slowed down because increased viscosity slows down chain diffusion of the folding molecule, e.g. during hydrophobic collapse (Jacob et al., 1999; Jacob et al., 1997).

The exposure of hydrophobic residues is weakened at lower temperatures due to decreased molecular movement. It could therefore be assumed that low temperatures generally promote folding because aggregation of folding intermediates *via* hydrophobic patches is slowed down. In fact, this has been reported in some cases (Buchner & Rudolph, 1991; Mendoza et al., 1991; Viitanen et al., 1990). However, the converse has also been observed, namely that elevated temperatures between 20–40 °C resulted in higher refolding yields than those observed at lower temperature (Fischer et al., 1993; Peralta et al., 1994).

Beside dilution and dialysis, protein folding can also be achieved by binding the protein to a chromatographic resin in the unfolded state and subsequent washing with an appropriate buffer that contains no denaturant. Affinity resins like Ni-NTA sepharose (Rehm et al., 2001; Zahn et al., 1997) or heparin sepharose (Stempfer et al., 1996) have been used for both poly-His-tagged or poly-Arg-tagged proteins, respectively. Compared to folding by dialysis or rapid dilution, this method has two advantages: aggregation due to intermolecular interaction of partly folded species is prevented, and the folded protein is obtained in higher concentration. However, an interference of the chromatographic support with the folding protein molecule can be detrimental, causing precipitation of the protein on the matrix.

In a different approach, detergents like CTAB in combination with  $\beta$ -cyclodextrin were used to fold proteins (Hanson & Gellman, 1998; Rozema & Gellman, 1996). Folding with this system comprises two steps. First, the denatured protein is diluted into a buffer containing detergent like CTAB or SDS. The protein is captured in a complex with the detergent, thus preventing aggregation. In case of bovine carbonic anhydrase II, CD

spectroscopy of these complexes revealed an increased level of non-native secondary structure. This is attributed to  $\alpha$ -helix formation in the detergent-complexed state (McCoy & Wong, 1981). In the second step, refolding is initiated by adding  $\beta$ -cyclodextrin that removes the detergent from the protein.

Protein folding of denaturant- or detergent-solubilised proteins has been extensively reviewed (Clark, 2001; Guise et al., 1996; Lilie et al., 1998; Middelberg, 2002; Misawa & Kumagai, 1999; Rudolph & Lilie, 1996). A protein folding screen for identification of optimal folding conditions using rapid dilution or dialysis has been developed by Gouaux and colleagues (Armstrong et al., 1999; Chen & Gouaux, 1997). Related screens have been used by others (Heiring & Muller, 2001; Tobbell et al., 2002). This so-called fractional factorial screen is also commercially available (FoldIt Screen, Hampton Research) and consists of an initial screen of sixteen conditions, in which twelve parameters (various additives, pH, protein concentration) are altered. Each parameter appears in eight of the sixteen conditions. Therefore, a couple of parameters are varied and tested at the same time, instead of checking each parameter independently. Thus, this is an approach to screen a high number of parameters with a relatively small number of folding buffers.

### **1.4 Detection of natively folded protein upon folding screening**

When screening for the best conditions of protein folding, it is difficult to determine the yield of native protein, if no biological assay is available. This problem is particularly important for proteomics projects that deal with large numbers of diverse proteins, but may also be inherent in various other fields where a functional test to detect native protein is not available or very time- and/or product consuming. Centrifugation or filtration can remove precipitates, but solubility cannot be used as a stringent criterion, as soluble microaggregates and partially folded intermediates are not distinguished from the natively folded protein. To prevent aggregation, low protein concentrations are commonly used in folding experiments. Therefore, sensitive methods are required for detection. Some methods that can be commonly applied are discussed here:

1. Analytical gel filtration, which has been applied by Chen et al. (Chen & Gouaux, 1997), separates molecules according to size and shape, assuming that all non-aggregated protein is natively folded. In my own experience, the major drawback of

---

this method is that protein concentrations are often too low to detect natively folded protein. Misfolded non-aggregated protein or trapped partially folded intermediates may exist and are difficult to distinguish from native protein. In addition, the method is time-consuming and columns may clog easily because misfolded hydrophobic species may precipitate on the column resin.

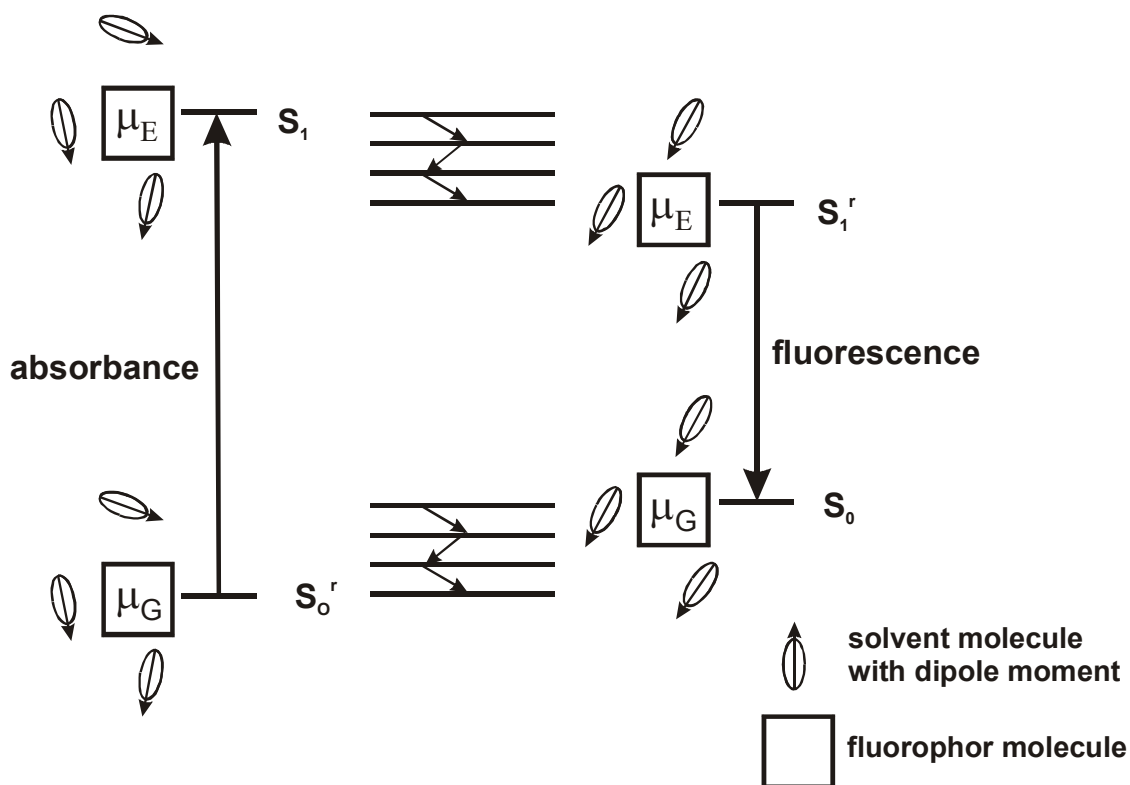
2. Limited proteolysis, which has been used by Heiring and Muller (Heiring & Muller, 2001). Partially folded intermediates are assumed to be more susceptible to subtilisin-induced proteolysis than native protein. Although this approach is very sensitive and does not require high sample purity, there are some disadvantages. Firstly, range-finding experiments for the protease concentration may be necessary. Additionally, false positives or false negatives can occur because soluble nonnative microaggregates might be equally or even less susceptible to proteolysis than native protein, thus leading to a misinterpretation of the results.
3. Circular dichroism (CD) spectroscopy to probe for secondary and tertiary structure (see also 3.5.8). A drawback of this method is that a  $\beta$ -sheet spectrum observed in far UV-CD spectra could result from soluble aggregates but not from native protein (Chiti et al., 1999). The acquisition of a near-UV CD spectrum to monitor close-packed structure around aromatic rings is limited by low sensitivity. Additionally, CD spectroscopy is difficult to perform when a large number of samples need to be analysed.
4. One dimensional  $^1\text{H}$  nuclear magnetic resonance (1D  $^1\text{H}$ -NMR) spectroscopy (see also 3.5.10) is an established and relatively sensitive method to check for protein folding (Rehm et al., 2002). Large chemical shift dispersions, especially in the methyl group region can be used as an indicator for protein folding. But this method is very expensive and sensitivity may in some cases not be sufficient. NMR-spectrometers are not standard laboratory equipment. However, the application of this method to detect small quantities of natively folded human protein domains will be described in the second part of this work.
5. Tryptophan fluorescence spectroscopy. Because this method will be important in this thesis it is introduced here in more detail. Intrinsic fluorescence of proteins mainly results from tryptophan, tyrosine and phenylalanine residues, with tryptophans being the dominating residue due to higher absorbance and higher quantum yield of emission. Depending on the distance to tryptophan residues, fluorescence resonance energy of tyrosine and phenlyalanine can be transferred to tryptophan

---

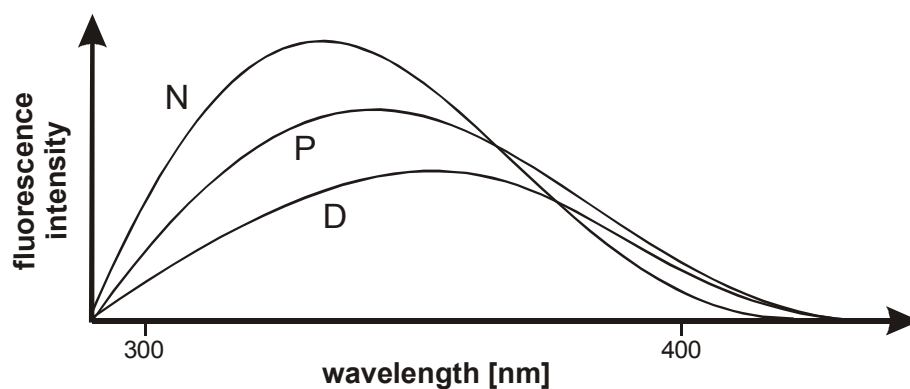
with the result that mainly tryptophan fluorescence is observed. Folding intermediates and natively folded proteins are distinguishable in their respective tryptophan spectra (Royer, 1995). Fluorescence occurs when a chromophore is raised to a higher energy level by absorption (within approx.  $10^{-15}$  sec) and subsequently (after  $10^{-7} - 10^{-9}$  sec) leaves its excited electronic state by emission of light (Figure 1.3 A). The excited state possesses a stronger dipole moment than the ground state. If a polar solvent surrounds the chromophore, these solvent molecules need to reorientate to the chromophore. Therefore there will be loss of vibrational energy in the excited electronic state as heat by collision with neighboring atoms. This will proceed until the molecule reaches the lowest vibrational level of the excited state ( $10^{-11} - 10^{-14}$  sec). After the emission of light in the excited state, the polar solvent molecules have to relax a second time to their original orientation, which again is accompanied by loss of energy and brings the molecule back to the ground state. Hence, the polarity of the environment of a tryptophan residue of a protein defines its fluorescence energy. Tryptophan residues are commonly more solvent-exposed in unfolded or partially folded proteins than in native proteins, where they are buried in a hydrophobic environment in the interior of the protein. Therefore, upon folding of denaturant-unfolded protein to the native state, the wavelength of the fluorescence intensity maximum is shifted to smaller values (i.e. higher energy, blue-shift). This well-known effect is illustrated in Figure 1.3 B. The occurrence of such a blue-shift can be used as a monitor for folded protein. In addition, solvent-exposure usually causes quenching of fluorescence (mainly by oxygen). Therefore it may be assumed that high fluorescence intensity is an indicator for folded protein and indeed this is observed in most cases. However, a few exceptions to this assumption are known (e.g.  $\gamma$ F-crystallin (Das & Liang, 1998), where the tryptophan fluorescence is strongly quenched in the native conformation by histidines, disulfide bonds or chromophores. Hence, fluorescence intensity should only be used in conjunction with other folding monitors. Tryptophan fluorescence is very sensitive and usually can be measured with relatively dilute samples (at protein concentrations in the low microgram per milliliter range)

With the exception of analytical gel filtration, samples that are analysed with the other methods need to be dialysed against a standard buffer prior to analysis.

A



B



**Figure 1.3: (A) Schematic depiction of the physical basis of fluorescence in solutions (adapted from Lankowicz, 1999) and (B) hypothetical protein fluorescence spectra (adapted from Royer, 1995)**

(A)  $S_0^r$ ,  $S_0$ ,  $S_1^r$  and  $S_1$  represent energy levels for the “relaxed” ground state, the ground state, the “relaxed” excited state and the excited state, respectively;  $\mu_G$  and  $\mu_E$  indicate the energy of the fluorophore dipole in the ground state and in the excited state, respectively.  $\mu_G < \mu_E$  (see text). (B) N, P and D indicate spectra of native, partially folded and unfolded protein, respectively.

---

### 1.5 Structural analysis of human protein domains

Structural genomics projects aim to produce large sets of recombinant proteins that are suitable for structural analysis. "Suitable for structural analysis" shall indicate that these proteins will yield well diffracting crystals or well analysable NMR-spectra. Unstructured regions in a protein have been shown to impair crystallisation (Pantazatos et al., 2004) as well as they may complicate NMR spectra. In a lot of cases it will be necessary to eliminate unstructured regions from a full-length protein to make them "suitable targets". This shall not imply that unstructured regions generally do not have biological relevance. Although there is no doubt that protein structure and function are intimately linked, it should be noted that numerous unstructured regions have indeed been identified to possess biological function (Wright & Dyson, 1999). It has been estimated that about 25% of residues in proteins in a range of different genomes are in regions unlikely to fold to globular structures (Lovell, 2003). Eventually, these unstructured regions become ordered when they interact with *in vivo* binding partners that provide stabilising atomic contacts (Wright & Dyson, 1999). In those cases it is advantageous to know the interaction partner so that they may be included in crystallisation trials and NMR experiments.

By producing only a domain of a protein that represents a compact folding entity one may reduce the possibility that the expressed construct contains long unfolded stretches. Because protein domains are of limited size (usually not bigger than 15 kDa) they represent targets well suitable for analysis with NMR-spectroscopy. A domain can occur repeatedly in one protein or it may be found in different proteins. Domains may be defined in different ways according to Jaenicke (1999). They are (i) stable globular substructures which may be disconnected by limited proteolysis, (ii) co-operative thermodynamic units, detectable by distinct folding/unfolding transitions (iii) genetic units, defined by sequence homologies (based on the consensus that one-dimensional homology or identity translates to three-dimensional relationships) (iv) substructures with specific functional properties. Because many exons encode for one domain, it was hypothesised that the occurrence of a domain in a variety of proteins was caused by a "shuffling of exons" during evolution. Protein domains of a multidomain protein might fold independently without being linked to the other domains, although they may be stabilised when being docked to their neighbouring domains (Martin & Schmid, 2003; Mayr et al., 1994).

---

In context of a project at the PSF, a set of human protein domains without known structures were analysed. Domain boundaries were predicted using the SMART database of Hidden Markov Models (<http://smart.embl-heidelberg.de/>) (Letunic et al., 2002; Schultz et al., 2000). Such predictions are based on the analysis of a statistical model that has been derived from a multi-sequence alignment of the protein domain family (Schultz et al., 1998).

To express an isolated domain successfully, the expression construct needs to represent a compact folding entity. This should not contain overhanging unstructured amino acid stretches that may hamper correct folding, lead to aggregation or may complicate analysis with NMR spectroscopy. The domain boundary prediction is putative for domains of unknown structure. Domain prediction by SMART does not always span the complete folding entity, but might miss weakly conserved residues at the ends. This could result in truncated domain constructs lacking stabilising residues at the termini. Therefore different constructs should be generated for each domain, with small variations at the predicted domain boundaries. The outcome is a large set of expression clones. Out of this set, constructs suitable for structure determination have to be identified, namely constructs that express a folded protein in reasonable quantities. For this purpose, methods need to be implemented for fast and reliable screening of the expressed proteins that can be used as early as possible in the overall process from cloning to structure determination.

### **1.6 Objectives of this work**

The production of recombinant proteins for structural or functional analysis using the *E. coli* expression system may be problematic when the protein is not folded correctly and is deposited in insoluble inclusion bodies. In some cases it is favourable to analyse deletion constructs of a protein or an individual protein domain instead of the full-length protein. This implies the generation of a set of expression constructs that need to be characterised. The motivation of this work was to develop methods that facilitate *in vitro* protein folding as well as to develop procedures for quick and reliable evaluation of *in vivo* protein folding. Both parts include one methodological problem to be solved, namely the detection of low microgram-amounts of natively folded protein.

---

Expression of proteins in inclusion bodies has some advantages: (i) the yield is usually high, (ii) the protein is protected from protease digestion, and (iii) proteins that are toxic to the expression host in their native conformations can be expressed. Despite these advantages, the necessity to refold the target protein to its native conformation remains a major obstacle. In a lot of cases *in vitro* protein folding is paralleled by misfolding and/or aggregation (see 1.2). The overall folding reaction is dependent on various parameters (see 1.3). Therefore screening strategies need to be developed. This involves mainly the choice of the refolding method (rapid dilution, dialysis etc.) and conditions (which buffers and additives are used, how should those be combined), but also preliminary work like inclusion body preparation and protein purification prior to screening. The presented work describes a screening strategy for folding of globular proteins and compares it to a strategy proposed previously (Armstrong et al., 1999). The amount of protein and work invested into the screening procedure increases with the number of conditions (parameters) tested. To overcome this problem this work deals with the automatisisation and miniaturisation of a refolding screen using a pipetting robot. This should make such a screen useful for high-throughput analysis carried out at the PSF or similar projects.

Beside the problem to fold denatured proteins correctly, there is the difficulty to detect small amounts of natively folded protein produced in particular conditions of the folding screen. As functional assays are usually not available, generally applicable methods need to be implemented. The applied methods should be sensitive, reliable and adaptable to high-throughput processing.

Furthermore, the implementation of high-throughput, small-scale procedures for soluble expression in *E. coli*, purification and folding analysis of different constructs of a set of human protein domains was attempted (see also 1.5). The aim is to establish a procedure for quick and reliable identification of suitable candidates for structural analysis. It was investigated whether folding analysis procedures evaluated for protein folding screening of inclusion body proteins can also be applied in this context.

---

## 2 Materials

### 2.1 Chemicals

All chemicals were obtained from Merck in high purity (p.a.), except:

|  |           |
|--|-----------|
| ammomiumpersulfat                              | ICN       |
| ampicillin                                     | Sigma     |
| 8-anilino-1-naphtalenesulfonic acid (ANS)      | Sigma     |
| bovine erythrocyte carbonic anhydrase II (CAB) | Sigma     |
| chloramphenicol                                | Sigma     |
| $\beta$ -cyclodextrin                          | Fluka     |
| guanidine hydrochloride (GdnHCl, ultra pure)   | ICN       |
| hexdecyltrimethylammonium bromide (CTAB)       | Fluka     |
| kanamycin                                      | Sigma     |
| $\beta$ -mercaptoethanol                       | Fluka     |
| myoglobin, from horse heart                    | Sigma     |
| $\beta$ -NADH                                  | Sigma     |
| Ni-NTA-agarose                                 | Qiagen    |
| Ni-NTA-superose                                | Qiagen    |
| oxalacetic acid                                | Fluka     |
| PMSF   | Roche     |
| p-nitrophenyl acetate                          | Sigma     |
| Rotiphoresegel <sup>®</sup> Gel 30             | Roth      |
| anhydrous sodium tetraborate (Borax)           | Fluka     |
| TEMED  | Aldrich   |
| Low molecular weight marker (2 – 17 kDa)       | Pharmacia |
| Precision Plus Protein Standard (10 – 250 kDa) | Bio Rad   |

## 2.2 Media

|      |   |
|------|---|
| 2xYT | 16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, pH 7.0, sterilised by autoclaving  |
| SB   | 12 g/l Bacto-tryptone, 24 g/l yeast extract, 0.4% (v/v) glycerol, 17 mM KH <sub>2</sub> PO <sub>4</sub> , 72 mM K <sub>2</sub> HPO <sub>4</sub> , pH 7.0, sterilised by autoclaving |

### Antibiotic stock solutions

|                 |                                |
|-----------------|--------------------------------|
| Ampicillin      | 100 mg/ml in 50% (v/v) ethanol |
| Chloramphenicol | 34 mg/ml in ethanol            |
| Kanamycin       | 30 mg/ml in aqua bidest        |

## 2.3 Miscellaneous materials

|   |                       |
|---|-----------------------|
| 22 – well dialysis unit   | self-made, AG Seckler |
| 96 deep well plates   | Brandt                |
| 96-well plate for fluorescence measurements   | Corning               |
| cuvettes (quartz) for UV or CD spectroscopy and DLS   | Hellma, Müllheim      |
| filterplate for automated protein purification (Protino <sup>®</sup> M96)   | Macherey & Nagel      |
| filterplate for (re)folded protein samples and purified human protein domains for analytical HIC  | MAGV N22 (Millipore)  |
| HiTrap desalting column   | Amersham Bioscience   |
| 10/10 XK column for Ni-NTA superose   | Amersham Bioscience   |
| PEEK columns (0.46 mm * 50 mm, 0.83 ml bedvolume for analytical HIC, 0.46 mm * 100 mm, 1.6 ml or 10 mm * 100 mm, 7.8 ml for large scale preparations) | Applied Biosystems    |
| POROS 20 HP2 (for HIC)  | Applied Biosystems    |
| POROS 20 HQ (for anion exchange chromatography)   | Applied Biosystems    |
| POROS 20 MC (for metal chelate chromatography)  | Applied Biosystems    |
| POROS 20 S (for cation exchange chromatography)   | Applied Biosystems    |
| Superose 12 16/50 column  | Amersham Bioscience   |
| Ultrafree – 15 or 4 (ultrafiltration)   | Millipore             |
| Zellu Trans 4,000 – 6,000 Da cut-off membrane   | Roth                  |

---

## 2.4 Instruments

|                            |  |
|----------------------------|--|
| Balances:                  | APX-200 (Denver Instruments)<br>BL 1500S (Sartorius) |
| Centrifuges:               | Anvanti™ J-20 XP (Beckman Coulter)<br>4K15 (SIGMA)   |
| Dynamic light scattering:  | Laser spectroscatter 201 (RiNA)                      |
| Electrophoresis:           | Mighty Small II (Hoefer)                             |
| Infrared spectroscopy      | Bruker iFS113  |
| Fluorescence spectroscopy: | Cary Eclipse (Varian instruments)                    |
| HPLC:                      | Vision Workstation (Applied Biosystems)              |
| pH meter:                  | pHM 82 (Radiometer Copenhagen)                       |
| Pipetting robot:           | Speedy (Zinsser Analytic)                            |
| Shaker:                    | Orbital Shaker (Forma Scientific)                    |
| Sonication:                | Sonoplus (Bandelin)                                  |
| Spektropolarimeter:        | J-715 (Jasco)  |
| UV-Vis spectroscopy:       | Cary 50 Scan (Varian Instruments)                    |
| NMR spectrometer           | Bruker, DRX 600                                      |

---

## 3 Methods

### 3.1 Bioinformatics tools

Secondary structure prediction of p22 dynactin was performed with PSIPRED (Department of Computer Science, University College London, United Kingdom).

Domain prediction was performed using SMART, EMBL, Heidelberg (<http://smart.embl-heidelberg.de>).

Prediction of protein pI was done using the EMBL WWW Gateway to Isoelectric Point Service (<http://www.embl.de/cgi/pi-wrapper.pl>).

### 3.2 Cloning

All target genes were obtained as cDNA-clones from the RZPD, German Resource Center for Genome Research GmbH (<http://www.rzpd.de>) and were PCR-amplified with 5'- and 3'-gene specific primers. PCR reactions have been performed using the Expand High Fidelity system (Roche Diagnostics) according to manufacturers' instructions. PCR products were restriction digested using restriction sites introduced into the primers, ligated into the vector pQTEV (GenBank Accession Number AY243506) and cloned into *E. coli* SCS-1 cells carrying the pRosetta (Novagen) helper plasmid using a chemotransformation method (Inoue et al., 1990). Proteins expressed from the pQTEV vector carried an N-terminal heptahistidine-tag (His-tag) followed by a TEV protease cleavage site. All cloning steps were carried out in a parallel 96-well format where appropriate and were performed by Volker Sievert (PSF).

### 3.3 Protein Expression

#### 3.3.1 Expression of protein in *E.coli* for folding scening

In a preliminary experiment, proteins for folding screening were expressed in small scale at both 37 °C and 22 °C to evaluate the effect of a lower expression temperature on protein solubility. 300 µl-overnight cultures were prepared in 2xYT medium, supplemented with 2% glucose, 100 µg/ml ampicillin and 30 µg/ml chloramphenicol. This culture was transferred into 2.7 ml of SB medium, supplemented with 20 µg/ml thiamine,

100 µg/ml ampicillin and 30 µg/ml chloramphenicol. After 3 h protein expression was induced for 3 h at 37 °C or for 5 h at 22 °C by adding IPTG to a final concentration of 1 mM. Cells were harvested by centrifugation and lysed for 30 min at room temperature by adding 200 µl lysozym buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.1 mM EDTA, 0.2 mg/ml lysozym, 1 mM PMSF). To reduce viscosity caused by nucleic acids, 35 µl benzonase buffer (50 mM Tris-HCl, pH 8.0, 1.2 mM mgCl<sub>2</sub>) containing 0.1 Unit/µl benzonase was added followed by a further 30 min incubation at room temperature. A sample is withdrawn (whole cellular protein) and the lysate was cleared by centrifugation (16000xg, 4 °C, 10 min). A second sample was collected from the supernatant (soluble cellular protein). Samples are subjected to SDS-PAGE (see 3.5.2).

For large-scale preparation, 100 ml overnight culture was prepared in 2xYT medium, supplemented with 2% glucose, 100 µg/ml ampicillin and 30 µg/ml chloramphenicol. This culture was transferred into 1,9 L of SB medium, supplemented with 20 µg/ml thiamine, 100 µg/ml ampicillin and 30 µg/ml chloramphenicol. At an optical density at 600 nm of 1.5, protein expression was induced for 4 h by adding IPTG to a final concentration of 1 mM. Cells were harvested by centrifugation (4000xg for 20 min at 4°C) and washed with 20 mM Tris-HCl, pH 7.7, 150 mM NaCl. They were frozen in liquid N<sub>2</sub> and stored at -80 °C until use.

### **3.3.2 Expression of human protein domains in 96 deep-well microplates**

A 96 deep-well microplate with 2 ml cavities was filled with 0.1 ml 2xYT-medium supplemented with 2% glucose, 100 µg/ml ampicillin and 15 µg/ml kanamycin. Pre-cultures were inoculated with 96 expression clones using a custom made steel replicator carrying 96 pins of 6 cm length. The plate was sealed tightly and bacteria were grown overnight at 37 °C. 0.9 ml pre-warmed SB-medium was added. The plate was covered with a lid allowing air exchange and incubated for 3 h at 37 °C. Protein expression was induced for 3 h at 37 °C, or alternatively at 28 °C, by adding IPTG to 1 mM final concentration. Cells were harvested by centrifugation at 4 °C at 2,000 g for 10 min and frozen at -80 °C.

---

### 3.4 Protein preparation and purification

#### 3.4.1 Perfusion chromatography

All column purification steps and analytical HIC (see 3.5.4) were performed using perfusion chromatography with POROS media (except a separation of TEV protease from purified p22 dynactin using Ni-NTA sepharose and size exclusion chromatography with p22 dynactin using superose 12). POROS media are based on spherical polystyrene divinylbenzene particles with high chemical and mechanical stability. In contrast to conventional chromatography material where beads with long, thin pores are used into which molecules need to diffuse, POROS media contain large (6,000–8,000 Å) perfusive channels (through pores) that allow for flow of the mobile phase through the particle. Using a network of smaller interconnecting diffusive pores between the through pores enhances the surface area of this material. Such media reduce stagnant mobile mass transfer encountered in conventional chromatography, an effect that leads to peak broadening and is caused by diffusion (Kaufmann, 1997). During perfusion chromatography, solutes enter the interior of the perfusive particles through a combination of convective and diffusive transport. The higher the flow rate - and concomitantly the pressure differential across each bead - the more dominant is convective flow compared to diffusive flow. Hence, higher flow rates may be applied while separation quality should increase (typically > 1000 cm/h).

#### 3.4.2 Denaturation of model enzymes for refolding screening and unfolding transitions of CAB and myoglobin

CAB and MDH were denatured for 5 h at room temperature. Conditions used were: (i) 6 mg/ml CAB in 20 mM Tris-HCl, pH 7.7, containing 6 M GdnHCl and 2 mM EDTA and (ii) 2.24 mg/ml MDH in 20 mM Tris-HCl, pH 7.7, containing 7 M GdnHCl and 10 mM DTT.

For GdnHCl-unfolding transitions of bovine carbonic anhydrase II (CAB) and myoglobin, CAB (50 µg/ml, A) and myoglobin (50 µg/ml, B) were incubated for 5 h in 20 mM Tris pH 7.5, enriched with different GdnHCl concentrations. Unfolding transitions were monitored by fluorescence spectroscopy (3.5.3) as well as enzymatic activity for CAB

(3.5.5) and soot absorbance at 409 nm for myoglobin. Data points were sigmoidally fitted using Origin<sup>®</sup>.

### 3.4.3 Inclusion body preparation and denaturing His-tag purification

Cells were suspended in ice cold 20 mM Tris-HCl, pH 7.7, 300 mM NaCl, 2 mM DTT, 1 mM EDTA, before disruption by sonification. Sonification of ice-cooled sample was performed using a Sonoplus with the TT13 titanium flat tip (Bandelin) in two 30 sec-steps (with 3 min pause in-between), in 50% cycles with a 75% amplitude. The crude extract was centrifuged at 25,000xg for 45 min, 10 °C, and the pellet was washed twice in 20 mM Tris-HCl, pH 7.7, 300 mM NaCl. The resulting pellet was dissolved in 20 mM Tris-HCl, pH 7.7, 6.6 M GdnHCl, 30 mM NaCl, 5 mM imidazole, 2 mM  $\beta$ -mercaptoethanol over night under stirring at room temperature. The GdnHCl-insoluble material was removed by centrifugation (25,000xg, 60 min) and the supernatant was applied to a 7.8 ml Ni-POROS20-column, previously equilibrated in urea buffer (20 mM Tris-HCl, pH 7.7, 8 M urea, 30 mM NaCl, 5 mM imidazole, 2 mM  $\beta$ -mercaptoethanol). After washing with fifteen column volumes of urea buffer, bound protein was eluted using urea buffer enriched with 300 mM imidazole. The protein concentration was adjusted to 5 mg/ml by dilution into urea buffer. EDTA and DTT was added to achieve final concentrations of 1 mM and 15 mM, respectively, followed by a 60 min incubation at room temperature to reduce cysteines (DTT was not added to prepared KREB protein, AAD20972). The obtained protein solution was used for rapid dilution screen conditions (Table 4.1). The purified proteins were analysed with SDS-PAGE (see 3.5.2), where in all cases a high purity could be observed (>90%). For screening conditions where proteins are immobilised on Ni-NTA agarose or Fractogel EMD Chelate during folding, EDTA, DTT and imidazole had to be removed by a buffer exchange against urea buffer using a HiTrap desalting column. A solution containing approx. 1.5 mg protein was incubated for 20 min whilst shaking with 350  $\mu$ l Ni-NTA agarose or 350  $\mu$ l Fractogel EMD Chelate, loaded with Ni<sup>2+</sup>, both equilibrated in urea buffer. After binding, the suspension with the affinity resins was aliquoted in eight portions, four for each resin. After the resins had settled at the bottom of the tubes, the supernatant with unbound protein was removed leaving 100  $\mu$ l in every tube.

---

### 3.4.4 Automated folding screening

The pipetting robot used has four steel pipetting needles and is equipped with a temperature-controlled plate that keeps all solutions at 16 – 18 °C. Protein folding was carried out in a rack containing 2 ml tubes for conditions 1 – 22 (see Table 4.1) and 1.5 ml tubes for folding of proteins that were immobilised on metal chelate resins (conditions 23 – 30, Table 4.1). All buffer components for rapid dilution screening were pipetted together from stock solutions by the robot. Before adding protein, the rack was transferred onto a stirrer to ensure effective mixing. Screening conditions 1–22 had a final volume of 1 ml. For conditions 1–20, 20 µl aliquots of denatured protein were added in four 5µl-steps, divided by 150min-incubation. In case of conditions 21 and 22 (CTAB/cyclodextrin-system), protein was added in one step to 680 µl CTAB-containing buffer and incubated for 150 min, followed by addition of 300 µl β-cyclodextrin and a further 10 h incubation. 350 µl aliquots were submitted to dialysis and subsequent fluorescence spectroscopy and 400 µl were directly applied to HIC (see 3.5.4).

For conditions 23–30, denatured protein bound to affinity beads (prepared as described in 3.4.3) was washed with 1.3 ml urea buffer (see 3.4.3). To achieve this, the pipetting needles detect the liquid surface and remove the liquid slowly while moving down with the liquid level. All conditions, except for 24 and 26, were washed three times in 1.3 ml folding buffers and incubated for 10 h to induce folding. For conditions 24 and 26, the urea concentration was stepwise decreased by dilution with folding buffer (see table 4.1, 15 min intervals) followed by two washes and a 9 h incubation. In summary, all samples were washed five times in total, including a final transfer into the same buffer (20 mM Tris-HCl, pH 7.7, 30 mM NaCl, 2 mM β-mercaptoethanol). Finally, protein was eluted with 400 µl of 20 mM Tris-HCl, pH 7.7 150 mM NaCl, 2 mM β-mercaptoethanol, 500 mM imidazole. The obtained samples were analysed by fluorescence spectroscopy before application to the analytical HIC column (3.5.4).

The programmed method is documented in the file “refolding method.csv” as well as further manual information is provided in the file “refolding screen info.pdf”. Both files can be accessed on <http://www.molgen.mpg.de/~psf/robot>.

### **3.4.5 Large scale folding and purification of the p22 dynactin subunit**

Inclusion bodies of p22 dynactin from 4 L cultures were prepared and purified as described above. 15 ml of pure denatured protein (5 mg/ml) in 20 mM Tris-HCl, pH 7.5, 8 M urea, 30 mM NaCl, 15 mM DTT, 1 mM EDTA, 300 mM imidazole was diluted into 510 ml of 20 mM Tris-HCl, 50 mM NaCl, 0.2 mM  $\text{CaCl}_2/\text{MgCl}_2$ , 0.57 mM CTAB, 2 mM  $\beta$ -mercaptoethanol under vigorous stirring. After 2 h incubation, 225 ml of 15.7 mM  $\beta$ -cyclodextrin was added under vigorous stirring to induce folding at 18 °C. After 1 h, 7.5 mg TEV protease was added, followed by an overnight incubation to cut off the His-tag and to ensure that the folding reaction is completed. After centrifugation to remove precipitates, the folded protein was subjected to anion exchange chromatography (POROS HQ, 1.8 ml bed volume) where p22 dynactin did not bind. Subsequently, the p22 dynactin/TEV protease mixture was applied to Ni-NTA-superose to remove the TEV protease. To the flow-through containing p22 dynactin, 3.8 M ammonium sulfate was added under vigorous stirring to achieve a final ammonium sulfate concentration of 2.2 M. The resulting 2.2 L were applied onto a POROS HP2 column (7.8 ml bed volume) and purified as described above. The protein eluate from the ammonium sulfate gradient was applied to a Superose 12 16/50 column (Pharmacia), equilibrated in 20 mM Tris-HCl, pH 7.7, 50 mM NaCl. Eluted protein fractions that correspond to monomeric p22 dynactin (determined using standard molecular weight markers) were collected. A yield of 25 mg of p22 dynactin was obtained in 25 ml, representing 33% of the urea-denatured protein subjected to the folding and purification procedure. An aliquot of the gel filtration eluate was directly analysed with CD spectroscopy (see 3.5.8). The protein was concentrated to a final concentration of 24.7 mg/ml by ultrafiltration (Biomax centrifuge tubes, Millipore) for crystallisation experiments (data not shown). It was examined for occurrence of aggregation by DLS (see. 3.5.7). The purity of the sample was assessed by SDS-PAGE (see 3.5.2).

### **3.4.6 Medium scale folding attempts**

Except for p22 dynactin, all proteins listed in Table 4.3, page 59, did not give a peak during analytical HIC (see 3.5.4) in a decreasing ammonium sulfate gradient during folding screening. However, for some conditions a blue-shift in tryptophan fluorescence spectroscopy (see 3.5.3) was observed. These conditions were further analysed by

diluting the denatured proteins (approx. 5 mg/ml in 8 M urea) into 100 ml of the respective buffer matching the tested conditions. Like in small scale, dilution was performed in four 1:200 steps with at least 2.5 hours incubation in-between. 50 ml were diluted 1:2 with 20 mM Tris, pH 7.7 (to reduce ionic strength) and subjected to ion exchange chromatography. The theoretical pI of the respective proteins was calculated and for predicted pI lower than eight anion exchange chromatography (on POROS HQ, 1.8 ml bedvolume) was tested first. The wash buffer was 20 mM Tris, pH 7.7, and elution was carried out in a NaCl-gradient from 0 M–0.75 M over 12 column volumes. Flow through and eluate fractions were collected and analysed with SDS-PAGE (see 3.5.2). If no binding occurred, cation exchange chromatography was attempted with the flow-through (same protocol as for anion exchange chromatography). The remaining 50 ml were concentrated 10–20 times using ultrafiltration and subjected to UV-Vis spectroscopy to determine protein concentration and yield after the concentration procedure. An absorption scan from 500–190 nm was performed to also check for stray light, an indicator for larger aggregates. Additionally, the concentrated samples were analysed using DLS (see 3.5.7).

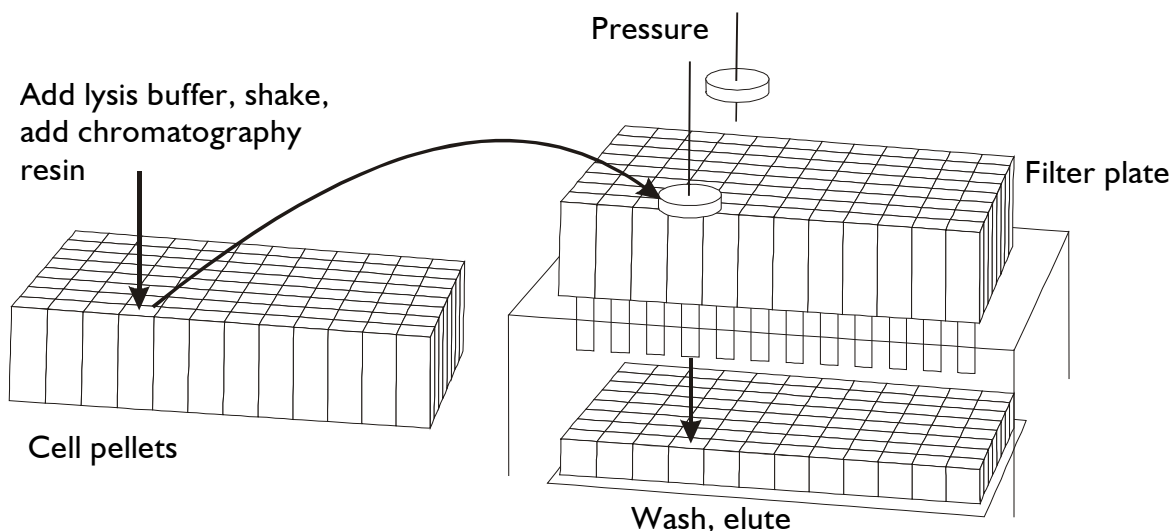
CCR4 was tested under condition 23 (Table 4.3) in medium scale. 2 ml of a 2.5 mg/ml denatured CCR4 solution (in 8 M urea; imidazole, DTT and EDTA had been removed by buffer exchange using a HiTrap column) were incubated with 1 ml Ni-NTA agarose, previously equilibrated in 8 M urea, 20 mM Tris, pH 7.7 and 2 mM  $\beta$ -mercaptoethanol. The suspension was shaken for 20 min at room temperature, unbound protein was removed and beads were washed in equilibration buffer. Protein was eluted using 1.5 ml equilibration buffer enriched with 400 mM imidazole. This sample was stored at 8 °C and analysed over the following days using DLS.

### **3.4.7 Robot protein purification of human protein domains (Scheich et al., 2003)**

The Zinsser Speedy pipetting robot has four steel pipetting needles and is equipped with a cooling plate that is temperature controlled by an external cooling device. Furthermore, the robot has a vacuum block that can be loaded with standard 96-well filter plate. Filtration can be achieved by either applying vacuum to the vacuum block or by applying pressure. The pressure is delivered through the four pipetting needles that can be filled with air and that can pick up plastic plugs (“shutter”) to close the filter plate’s wells (see

Figure 3.1). Additionally, the machine is equipped with a high-speed microplate shaker and a plate handling device.

Proteins were purified on the pipetting robot according to the following protocol. Shaking is performed at 1,500 rpm. Reactions were carried out at room temperature except otherwise noted. A microplate containing bacterial pellets stored at  $-80^{\circ}\text{C}$  was thawed and placed on the robot. Cell pellets were resuspended by adding  $150\ \mu\text{l}$  of lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.1 mM EDTA, 10 mM imidazole, 1 mM PMSF) and 5 min shaking.  $35\ \mu\text{l}$  lysozyme buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.2 mg/ml lysozyme, 0.5% Brij 58, 10 mM imidazole) were added to the resuspended cells, followed by 30 sec shaking (1500 rpm). The cultures were incubated for 30 min on the robot's cooling plate to reduce the temperature in the wells to approx.  $13\ ^{\circ}\text{C}$ .  $35\ \mu\text{l}$  benzonase buffer (50 mM Tris-HCl, pH 8.0, 1.2 mM  $\text{mgCl}_2$ ) containing 0.1 Unit/ $\mu\text{l}$  benzonase was added and plates are shaken for 30 sec followed by a 30 min incubation on the cooling plate and another 30 sec of shaking.  $20\ \mu\text{l}$  aliquots (whole cellular protein



**Figure 3.1 Automated purification procedure**

Crude lysates containing the target protein bound to affinity beads were transferred into a filter plate. Beads are washed and bound protein is eluted by applying pressure onto the filter plate wells

sample) were transferred to a PCR microplate containing 7  $\mu$ l of 4xSDS-PAGE sample buffer (0.2 M Tris-HCl pH 6.8, 8% SDS, 0.004% bromphenol blue, 40% glycerol). 50  $\mu$ l of a 50% suspension of Ni NTA agarose equilibrated in wash buffer (20 mM Tris, pH 8, 300 mM NaCl, 10 mM imidazole) are added. Immediately before this step, the 50% slurry is mixed with the pipetting needles. The plate is shaken for 30 min to bind the protein to the affinity beads. At this stage the volume in each well is 300  $\mu$ l considering that the cell pellet comprised 50  $\mu$ l. 150  $\mu$ l wash buffer is added to increase the volume to over 400  $\mu$ l. This volume increase improves the performance of the robot's liquid detection system which is required in the next step when the affinity beads/cell extract - suspension is transferred into the filter plate.

To reliably transfer the beads settled on the bottom of the well, the robot's pipetting needles detect the liquid surface and carefully resuspend the beads twice by slowly aspirating 100  $\mu$ l and dispensing with higher velocity. Then the bead suspension is transferred to a 96-well filter plate residing on the robots vacuum manifold. The filter plate is equipped with large-pore polyethylene frits (Macherey & Nagel, Düren, Germany, 96-well filter plate).

To ensure that no beads have been residing in the deep well plate, 400  $\mu$ l wash buffer is pipetted into the plate and a transfer to the filter plate is carried out as described above. The beads in the filter plate are washed by adding 400  $\mu$ l wash buffer and subsequently pressing out all liquid by applying positive pressure through the pipetting needles. The robot then places a collection microplate inside the vacuum block to receive protein eluates. 80  $\mu$ l elution buffer (20 mM Tris, pH 8, 300 mM NaCl, 400 mM imidazole) is added into the wells of the filter plate. Eluted protein is transferred into a collection plate by applying pressure. The obtained samples are subjected to SDS-PAGE (see 3.5.2). 36 samples that had relatively high yield of purified protein judged by the purified band in SDS-PAGE (denoted with “++” or “+” in Table 5.2) were selected and arrayed in duplicate fashion on a new plate. From this plate two 96 deep-well plates with cultures were grown (see 3.3.2), one for expression of proteins for the analysis with 1D  $^1$ H-NMR (3.5.10) and the other for analysis with analytical HIC (3.5.4). Proteins were purified again using the automated procedure and 10  $\mu$ l aliquots were used for SDS-PAGE analysis (see Figure 5.1). Due to the perturbing effect of Tris-HCl and imidazole resonance signals during analysis with 1D  $^1$ H-NMR, the automated purification protocol was slightly modified for analysis by 1D  $^1$ H-NMR. Before elution with 80  $\mu$ l 100 mM EDTA in 20 mM NaHPO<sub>4</sub>, pH 7.0, 50 mM NaCl, Ni-NTA-beads were washed using the same buffer

without EDTA to remove Tris-HCl. Eluates were analysed with SDS-PAGE (10  $\mu$ l, see 3.5.2) and 1D  $^1$ H-NMR (70  $\mu$ l). All constructs were run in duplicate which were pooled for NMR analysis. As a negative control (buffer spectrum), two 1 ml cultures from SCS1 cells containing no expression vector were subjected to the same purification protocol. The resulting 140  $\mu$ l were made up to 500  $\mu$ l with NMR-buffer stock solution to a final concentration of 20 mM  $K_2HPO_4/KH_2PO_4$ , pH 7.0, 50 mM NaCl and 10% (v/v)  $D_2O$ .

The programmed method is documented in the files “automated protein purification part 1.csv” and “automated protein purification part 2.csv”, as well as further manual information is provided in the file “automated protein purification info.pdf”. These files can be accessed on <http://www.molgen.mpg.de/~psf/robot>.

### 3.4.8 Preparation of $^{15}N$ labeled samples for 2D $^1H$ - $^{15}N$ -HSQC

This preparation was performed in Anne Diehls' laboratory (PSF). For expression of  $^{15}N$ -labeled proteins, *E. coli* was grown on M9 minimal medium with  $^{15}NH_4Cl$  (0.5 g/l) in 400 ml. Protein expression was induced for 4 h at 30 °C by adding 1 mM IPTG. Cells were collected and desintegrated using a french press. Purification of the soluble His-tagged domains was achieved *via* a 1.7 ml Ni-POROS column, TEV cleavage and a second metal chelating chromatography. The protein in the flow-through was concentrated to 0.5 ml by ultrafiltration and washed with 20 mM phosphate buffer, pH 7.0 and 50 mM NaCl.  $D_2O$  was added (10% final concentration) for NMR lock signal. 2D  $^1H$ - $^{15}N$ -HSQC is described in section 3.5.10.

## 3.5 Protein analysis

### 3.5.1 Determination of protein concentration

Unless otherwise indicated, protein concentrations were determined from the absorbance at 280 nm using the extinction coefficient calculated from the amino acid sequence (Mach et al., 1992). Absorbance was corrected for stray light according to the light scattering theory (Tyndall effect,  $I_{(s)} \sim \lambda^{-4}$ ) assuming that no absorption due to protein chromophores occur above 320 nm (Levine & Federici, 1982).

For protein concentration determination of conditions 1–18 of refolded MDH as well as of refolded CAB for ANS fluorescence studies, 250  $\mu$ l aliquots were centrifuged at 25000xg

for 30 min at 8 °C, and supernatants were analysed with the Bio-Rad Bradford Assay with native MDH or CAB as standards. This method is based on the formation of a complex of protein with Coomassie-Brilliant-Blue G250 which is followed by its absorption at 595 nm. Arginine and CTAB present in conditions 19–22 are not compatible with this assay, and could therefore not be determined. Protein concentrations of purified human domains were determined using the same assay with BSA as a standard.

### **3.5.2 SDS-polyacrylamidegelelectrophoresis (SDS-PAGE)**

The separation of proteins according to their molecular weight was performed using discontinuous SDS-PAGE (Laemmli, 1970) with an acrylamide concentration of 5% and 15% (v/v) in stacking and separating gels, respectively. Due to the small size of the human protein domains, a modified protocol for the separation of smaller proteins was applied (Okajima et al., 1993). In this procedure twice the concentration of Tris/glycin (0.05 M/0.5 M) in the running buffer is used as well as twice the concentration of Tris (0.75 M) in the separation gel, which contained 20% (v/v) acrylamide.

Gels were stained for 20 min with Coomassie-Brilliant-Blue G250 (in 45% ethanol, 10% acetic acid) and destained in 20% ethanol, 7.5% acetic acid. The sensitivity of Coomassie brilliant blue staining is in the range of 30–100 ng per band (Steinberg et al., 2001).

### **3.5.3 Tryptophan fluorescence spectroscopy (see also 1.3 and Schmid, 1997))**

350 µl of samples of conditions 1–22 of the folding screen were dialyzed for two days against 20 mM Tris-HCl, pH 7.7, 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT, using a membrane with a 4–6 kDa cut-off. The relatively long dialysis procedure was necessary to ensure effective removal of larger screening components like PEG 1000. Samples obtained from conditions 23–30 were not dialyzed. 320 µl of all samples were filtrated through a 96-well 0.22 µm PVDF membrane and collected in a black 96-well plate that has an UV-transparent bottom (Corning). Absorbance at 280 or 340 nm were measured before fluorescence emission spectra were taken between 285 and 420 nm upon

---

excitation at 280 nm in 0.5 nm intervals and with a scanning rate of 30 nm/min. The reference spectra were subtracted from the data.

### **3.5.4 Analytical hydrophobic interaction chromatography (HIC) (Wheelwright, 1991)**

#### Theory of hydrophobic interaction chromatography

The addition of kosmotropic salts like ammonium sulfate to a protein solution leads to distraction of water from the hydration shell of the protein to the salt ions because these have a higher charge density compared to the protein surface. In high ammonium sulfate concentrations, proteins can be bound to a hydrophobic resin, for example carrying phenyl ligands. Depending on the respective surface hydrophobicity, bound proteins can be eluted in a decreasing ammonium sulfate gradient, with more hydrophobic proteins eluting at lower salt concentrations. The exact mechanism of binding of proteins to hydrophobic surfaces is not known. The more kosmotropic a salt is (according to the Hofmeister series) the more it promotes binding. Strong chaotropic salts will favour desorption. Hence, there seems to be a correlation between the salting out ability of a salt and its ability to promote binding to hydrophobic resins. It has been proposed that the driving force for binding is the increase of entropy that results from displacement of water away from the hydrophobic patches on the protein surface.

#### Experimental procedure

Analytical HIC was performed using the Vision System (Applied Biosystems). For sample application, an autosampler is used that allows successive application of samples in reproducible runs. The program procedure is listed in the supplements.

**Screening *in vitro* (re)folded proteins:** to 400  $\mu$ l aliquots from the folding screen (see above) 3.8 M ammonium sulfate was added under vigorous stirring to achieve a final ammonium sulfate concentration of 2.0 M (or alternatively 1.5 M). Samples were centrifuged (30 min, 25000xg, 8 °C) and applied to a POROS HP2 column (0.83 ml bed volume, HP2 refers to high density phenyl), equilibrated in 20 mM Tris-HCl, pH 7.7, 2.0 M (1.5 M) ammonium sulfate. Chromatography was performed at 8 °C. The column was washed with equilibration buffer and a gradient to 0 M ammonium sulfate was performed over 18, 14 or 10 column volumes for CAB, MDH and for the inclusion body proteins,

respectively. After each run, the column was washed with 5 M GdnHCl to regenerate the column. A flow rate of 3.5 ml/min (1260 cm/h) was applied; each run took approx. 15 min. The relatively low absorption coefficient at 280 nm ( $\epsilon = 10,264 \text{ (M cm)}^{-1}$ ) of the 36.6 kDa monomeric MDH made it necessary to use the absorbance of peptide bonds at 220 nm to detect elution peaks from HIC for this protein. All shown chromatograms were baseline corrected.

**Screening of human protein domains:** the ammonium sulfate concentration of the eluates obtained from robot protein purification (70  $\mu$ l) was adjusted to 1 M under pipette mixing in 96-well format using a 2 M stock solution. Samples were filtrated through a 96-well 0.22  $\mu$ m PVDF membrane and chromatography was performed as described above. The column was washed with equilibration buffer (20 mM Tris-HCl, pH 7.7, 1 M ammonium sulfate) and a gradient to 0 M ammonium sulfate was performed over 10 column volumes. Finally, a gradient from 0 to 5 M GdnHCl over 5 column volumes was performed to elute tightly bound, presumably misfolded, protein. The flow rate was 3.5 ml/min; each run took approx. 15 min. Absorbance at 280 nm and 220 nm was monitored in parallel to detect eluted proteins. All shown chromatograms were baseline corrected.

When analysing the elution behavior of proteins of different sizes on the POROS HP2 column (see Figure 6.1), 100–200  $\mu$ g of cytochrome C, myoglobin, HRP, CAB and MDH (in 20 mM Tris, pH 7.5, 2 M ammonium sulfate) were first loaded separately on the column (equilibrated in the same buffer) and eluted in a decreasing ammonium sulfate gradient over 15 column volumes. After the different elution volumes (ammonium sulfate concentrations measured by conductivity, cytochrom C was in flow-through) were determined, all proteins were separated in one run, during which proteins eluted at the same volumes as was observed in the individual runs.

### 3.5.5 Determination of CAB and MDH activity

CAB activities were measured using the pNPac esterase assay (Rozema & Gellman, 1996). 50  $\mu$ l of dialyzed samples from the refolding screen were pipetted into 400  $\mu$ l 20 mM Tris-HCl, pH 7.7. After addition of 45  $\mu$ l of 52 mM pNPac in dry acetonitril and mixing, the formation of the hydrolysis product, *p*-nitrophenolate, was monitored by measuring the linear increase in absorbance at 400 nm between 30 and 60 seconds (background hydrolysis was subtracted). MDH activities were assayed as described

before (Hutchinson et al., 1994). To 980  $\mu\text{l}$  of 150 mM  $\text{K}_2\text{HPO}_4$ , pH 7.6, 2 mM  $\beta$ -mercaptoethanol, 0.5 mM oxalacetate and 0.2 mM NADH (30 °C), 20  $\mu\text{l}$  enzyme solution was added and mixed by pipetting and the decrease of absorbance at 360 nm was followed. The measured activities for both refolded CAB and MDH were related to the activity of a control sample with native enzymes. Absorbance changes were measured using a spectrophotometer.

### 3.5.6 ANS fluorescence of refolded CAB

ANS is a hydrophobic dye that is widely used to detect the molten globule state of a protein because it undergoes an increase in fluorescence on binding to hydrophobic surfaces (John et al., 2001). Neither fully folded nor fully unfolded proteins bind ANS.

Refolded samples of conditions 1–22 (see Table 4.1, final concentration was 100  $\mu\text{g/ml}$ , 3.5  $\mu\text{M}$ ) were dialysed as described in 3.5.3. Dialysis is necessary because additives of folding screen may alter ANS-fluorescence. ANS was added from a 2.1 mM stock solution to give a final concentration of 30  $\mu\text{M}$ . Fluorescence emission spectra were taken between 400 and 600 nm upon excitation at 350 nm in 0.5 nm intervals and a scanning rate of 30 nm/min. The blank spectrum (ANS without protein) as well as a spectrum of ANS in 50% ethanol were also recorded.

### 3.5.7 Dynamic light scattering (DLS) (Müller & Schuhmann, 1996)

DLS measures the time-dependent fluctuations of the intensity of straylight resulting from Brownian molecular motion. Smaller particles diffuse faster and thus have higher fluctuations of intensities. A correlator determines an auto correlation function, which is fitted to a theoretical correlation function:

$$g(\tau) = \exp(-2DK^2\tau) \quad \text{where } D = \text{diffusion constant}$$

$K = \text{absolute value of straylight vector}$   
 $\tau = \text{delay time}$

From this curve  $D$  can be determined ( $K$  and  $\tau$  are defined by the experimental settings). With increasing  $\tau$ ,  $g(\tau)$  decreases. This decrease occurs more rapidly for smaller

---

particles (i.e.  $D$  has a higher value). The particle radius  $r$  can now be determined from the Stokes-Einstein equation:

$$r = (k * T) / (6 * \pi * \eta * D)$$

where  $k$  = Boltzmann constant  
 $T$  = absolute temperature  
 $\eta$  = dynamic viscosity

In practice, a laser beam is directed on a sample and the stray light is detected by a photomultiplier at small time intervals. By comparing all obtained values, the correlation function is determined. This function reflects a continuous comparison of the position of a particle with temporally subsequent positions. For small particles, the correlation decreases in a smaller time frame than for bigger particles, which is reflected in the curvature of the correlation function. The outcome of such an analysis is a logarithmic plot of the samples' particle distribution. Because stray light intensity is proportional to the sixth power of the particle diameter, the sensitivity of this analysis is strongly dependent on particle size. If the hydrodynamic radius of a small monomeric protein shall be determined (usually in the range of 1.5–5 nm), a concentration of approx. 0.5 mg/ml or more is necessary for a reliable analysis. Aggregated protein can be detected at lower concentrations.

All samples were centrifuged at 16,000 $xg$  for 10 min, 4 °C, to remove large aggregates before analysis with DLS.

### **3.5.8 UV CD spectroscopy with p22 dynactin (Schmid, 1997; Sheehan, 2000)**

#### Theory of UV CD spectroscopy

When light passes through a protein solution it may be refracted (or delayed) or it may be absorbed. If plane polarised light is used, each asymmetric centre in a protein (mainly the  $\alpha$ -C atoms, which are L-enantiomers) interacts differently with the left and right polarised components of the light beam. Circular dichroism (CD) occurs, because extinction coefficients for left and right circularly polarised light are different. A CD-spectrum is obtained by plotting this difference against the scanned wavelengths. To measure the absorbance difference the protein sample needs to be selectively exposed to circularly polarised light for short time periods, which is achieved by a photoelastic

modulator. Such a modulator normally is a quartz piezoelectric crystal subjected to an oscillating electrical field. A photomultiplier detects the absorbance difference and converts it into ellipticity (millidegree-units). The selective absorption of one circularly polarised component lowers its amplitude and causes the resulting electric vector to trace an elliptical path around the axis of the transmitted beam (the electric vector is elliptically polarised). A photomultiplier detector converts incident light intensity into an electric current composed of a direct current (which is related to total light absorption) and an alternating current, which is a direct measure of CD.

CD signals of proteins are observed in the same spectral regions as absorbance occurs for their chromophores, i.e. at 170–250 nm for peptide bonds (far-UV CD) and at 250–300 nm for aromatic residues (near UV CD). Hence, far-UV CD reports the backbone (secondary) structure of a protein. CD bands in the near-UV region are observed when, in a folded protein, aromatic side chains are immobilised in an asymmetric environment.

### Experimental procedure

CD spectroscopy was carried out with the help of Monika Walter (AG Seckler), using a Jasco 715 CD spectropolarimeter, calibrated with camphor sulfonic acid. The concentration of p22 dynactin in 20 mM Tris/Cl, 50 mM NaCl, pH 7.7 was 1.03 mg/ml. Data were averaged at 10 °C over 25 scans accumulated at 0.5 nm intervals with 50 nm/min scan speed and 1 s response time. In the far UV, overlapping spectra were measured in 1 mm (250 – 207 nm) and 0.1±0.005 mm (250–186 nm) path length cells (Hellma, Müllheim, Germany). Spectra were baseline corrected and the overlapping part of the spectra was used to correct for variations in the path length of the 0.1 mm cell. The spectrum in the near UV region (350 – 250 nm) was measured in a 4 mm path length cell calibrated by absorption spectroscopy.

### **3.5.9 Fourier transform infrared spectroscopy with p22 dynactin (Jackson & Mantsch, 1995; Sheehan, 2000)**

#### Theory of Fourier transformed infrared spectroscopy

Infrared spectroscopy applied to proteins probes for the vibrational energy level of C=O, N–H and C–N bonds of the peptide backbone. The vibrational energy level is characterised by the bond length, the bond angle and electron density. In practice, the amid I and amid II absorptions of proteins are investigated to assess secondary

structure. Amid I absorption (in the spectral range of 1,600–1,700  $\text{cm}^{-1}$ ) contains mainly contributions for the C=O stretching vibration, whereas the amid II absorption (1,500–1,600  $\text{cm}^{-1}$ ) is less “pure” and contains N–H bending as well as C–N stretching vibrations. Amid I absorption is more commonly used and was applied in this work.

Different types of hydrogen bonding of the C = O group in proteins are distinguishable in their respective amid I absorption. In denatured, extended chains mainly intermolecular hydrogen bonding persist. These bonds are very strong, hence causing the C=O bond to be stretched. Accordingly, an intense amid I band at a relatively low wavenumber (the reciprocal of the wavelength) is observed (1,610–1,627  $\text{cm}^{-1}$ ). Intramolecular  $\beta$ -sheet formation exhibits hydrogen bonding of polypeptide chains still in an extended state, but hydrogen bonds are weakened compared to intermolecular ones due to steric constraints. This is reflected in an amid I maximum observed at 1,630–1,640  $\text{cm}^{-1}$ . In  $\alpha$ -helices, the C=O group of residue  $n$  forms a hydrogen bond with residue  $n + 4$  resulting in a slightly longer hydrogen bond. Therefore,  $\alpha$ -helical proteins possess amid I maxima at an even higher wavenumber (1,648–1,658  $\text{cm}^{-1}$ ).

In a Fourier transformed infrared (FTIR) spectrometer, an incident infrared beam is split by passage through a half transmissive beam splitter positioned at 45° to the incident beam. The reflected beam is again reflected by a stationary mirror, whereas the transmitted beam is reflected from a moving mirror. The moving of this mirror causes destructive (beams out of phase) or constructive (beams in phase) interference. Such a setup allows fast scanning in a relatively wide range of the infrared spectrum. A number of scans are performed of reference buffer and then repeated with sample present. Both interference patterns are analysed by Fourier transform, then reference was subtracted from the sample spectrum.

Water gives high background signals because of its high absorption within the amid region. Therefore, it needs to be replaced by  $\text{D}_2\text{O}$ .

### Experimental procedure

FTIR analysis was performed in collaboration with Frank Niesen (PSF).

Prior to the sample measurement, single beam spectra of  $\text{D}_2\text{O}$  buffer (20 mM Na/K phosphate, pH 7.0, 150 mM NaCl), injected into a temperature-controlled, freshly assembled 30 mm diameter transmission cell with two  $\text{BaF}_2$  windows and a 50  $\mu\text{m}$  polytetrafluorethylene gasket chamber through injection ducts, are taken as reference. Spectra are taken using a Bruker iFS113, equipped with a liquid  $\text{N}_2$ -cooled HgMnTe

detector (J15D series, EG&G Judson) and continuously flushed with dry air. p22 dynactin (eluate from gel filtration) was transferred into D<sub>2</sub>O buffer by ultra-filtration and injected into the cell after removal of reference buffer. The second derivatives of the difference spectra are calculated using the instrument software (Opus 4.0, Bruker Optik GmbH, Rheinstetten, Germany). The temperature was raised linearly (1 K/min), spectra were taken every 5 K.

### **3.5.10 Detection of folded protein using analytical 1D <sup>1</sup>H-NMR and 2D <sup>1</sup>H-<sup>15</sup>N-HSQC spectroscopy (Sheehan, 2000)**

#### Theory of NMR spectroscopy

Atom nuclei have magnetic spin moments that will orientate in a parallel or antiparallel fashion to an externally applied magnetic field. The two orientations are referred to as spin states. The magnetic spin rotates around the axis of the external field in a certain angle and with a particular frequency, the Larmor frequency. The spin states have certain potential energy levels that are dependent on the magnetic spin moment, the strength of the magnetic field and the rotational angle. Nuclear magnetic resonance (NMR) spectroscopy depends on absorption of electromagnetic radiation in the range of radiowaves causing the nuclei to undergo a transition from a low to a high energy spin state. To achieve this, the nucleus is exposed to a strong magnetic field as well as to a pulse of radiowaves. The weak oscillating magnetic field of the radiowaves needs to rotate around the spinning nucleus in a plane perpendicular to the strong magnetic field. Absorption occurs when the frequency of rotation is in resonance with the Larmor frequency of the spinning nucleus. This induces a short magnetisation perpendicular to the external field followed by the relaxation (decay of the induced magnetisation). The relaxation causes a tiny measurable current that is converted to a NMR signal.

The resonance frequency is dependent on the type of nucleus as well as its immediate chemical environment. Magnetic fields of nearby nuclei (in 3–8 Å distance) also alter the magnetic field experienced by the investigated nucleus (chemical shielding). For instrumental reasons the referencing of resonance signals is taken from a standard compound such as tetramethylsilane (TMS). More recently calibration has been performed with water signals. The frequency of the measured signals is related to the frequency of the standard compound and is expressed as a number called the chemical shift (in ppm, the higher the resonance frequency, the higher the chemical shift).

Electron withdrawing groups bonded (or close) to the group under investigation tend to increase the chemical shift. That is why protons of methyl groups in unfolded proteins that are exposed to solvent have a higher chemical shift than the methyl protons in folded proteins where methyl groups are commonly located in a tightly packed, hydrophobic environment.

Complex molecules like proteins show a high number of signals (chemical shifts) that overlap. To resolve these overlaps, two-dimensional NMR spectroscopy is applied. In this approach the samples are subjected to a radiowave pulse for certain periods of time after which a second pulse is applied. With application of the second pulse information of the magnetisation after the different incubation times resulting from the first pulse may be retrieved (indirect signal) and additionally another magnetisation is induced (direct signal). After Fourier transformation, the obtained signals can be plotted in two dimensions. Two-dimensional NMR may be homonuclear (one type of nuclei is investigated) or heteronuclear (two types of nuclei are successively transferred to resonance). For example, in two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC) spectroscopy, the frequencies of  $^{15}\text{N}$  and  $^1\text{H}$  are measured. Signals represent a proton bound to  $^{15}\text{N}$ , hence signals result mainly from the peptide backbone. Homogenous samples with folded proteins exhibit a large dispersion of distinct signals along both chemical shift axis (the  $^1\text{H}$ -axis and the  $^{15}\text{N}$ -axis) whereas in case of unfolded protein, signals tend to cluster.

### Analysis procedure

All NMR experiments were performed by Dietmar Leitner (FMP Berlin).

NMR measurements were carried out at 15 °C with a Bruker DRX600 spectrometer in standard configuration equipped with triple resonance probe head and a Bruker Sample changer. NMR-data were processed using the XWINNMR software (Bruker Biospin GmbH). For the 1D NMR experiments 3072 scans were accumulated with a relaxation delay of 1.2 s. 8 k complex data points were acquired with a sweep of 10,000 Hz. Water suppression was achieved by WATERGATE (Piotto et al., 1992).

2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (Kay et al., 1992) spectra were measured at 15 °C, recording 256 x 1024 complex points in the indirect ( $^{15}\text{N}$ ) and acquisition ( $^1\text{H}$ ) dimensions. Four scans were acquired per increment with a relaxation delay of 1.2 s.

### 4 Result I: Protein Folding Screening

#### 4.1 Screening design

The folding screen presented in this work includes folding by rapid dilution, folding with the CTAB/cyclodextrin system, and - in case of His-tagged proteins - folding on metal chelate resins (nickel-nitrilotriacetic acid (Ni-NTA) or EMD fractogel chelate, loaded with Ni<sup>2+</sup>). Dialysis into a screening set of different folding buffers was not used because it is more difficult to be implemented in a high-throughput approach using a pipetting robot system. For certain proteins, aggregation may occur during the folding reaction only under conditions where native protein is stable. For these proteins, dialysis should be the superior approach to rapid dilution (see 1.2, introduction). However, when folding those proteins by rapid dilution, they may be rescued by addition of weakly destabilising agents like arginine, hence shifting the equilibrium between aggregation reaction and the reaction of productive folding to productive folding (see also 1.3, introduction). Additionally, folding under conditions where the protein is immobilised may be helpful in those cases, because a self-aggregation reaction does not occur. The temperature during the folding reaction was kept as low as possible using the cooling device of the pipetting robot system (approx. 18 °C).

The folding screen consists of 20 conditions for rapid dilution with eight parameters (pH and additives, see Table 3.1). The used additives represent a selection of additives commonly found in literature to promote protein folding. The influence of those additives is briefly summarised in Table 4.2 (see also section 1.3). Rather than a fractional factorial screen (see 1.3, page 9), a screening-strategy was chosen in which only one or sometimes two parameter are changed at a time. Screening buffers are pipetted from stock solutions using a pipetting robot (see Figure 4.1). The procedure is flexible and the screen can easily be modified by reprogramming the robot or by exchanging stock solutions, e.g. for altering the pH-values. For example, a small modification of the screen for proteins containing no cysteines was carried out, e.g. for the KRAB protein. The redox pair GSH/GSSG, which catalyses disulfide bridging during folding, was omitted. When screening this protein, GSH/GSSG was replaced by borax, pH 10, to examine the effect of a higher pH on folding (see Table 4.1). For rapid dilution screening, several small protein aliquots (5-8 µl, containing approx. 25 µg protein) were added to each

## Result I: Protein Folding Screening

| condition | additives for rapid dilution and CTAB/cyclodextrin screening for CAB, MDH and p22 dynactin      |
|-----------|---|
| 1         | -   |
| 2         | 0.8 M urea  |
| 3         | 1.6 M urea  |
| 4         | 0.03 % PEG 1000   |
| 5         | 0.03 % PEG 1000 + 0.8 M urea  |
| 6         | 0.03 % PEG 1000 + 1.6 M urea  |
| 7         | 39 % glycerol   |
| 8         | 39 % glycerol + 0.8 M urea  |
| 9         | EDTA (*)  |
| 10        | EDTA + 0.03 % PEG 1000 + 0.8 M urea(*)  |
| 11        | EDTA + 39 % glycerol (*)  |
| 12        | 20 mM citrate, pH 5.3 (**)  |
| 13        | 20 mM citrate, pH 5.3 + 0.03 % PEG 1000 (**)  |
| 14        | 20 mM citrate, pH 5.3, + 39 % glycerol (**)   |
| 15        | 5 mM/0.5 mM GSH/GSSG  |
| 16        | 5 mM/0.5 mM GSH/GSSG + 0.03 % PEG 1000 + 0.8 M urea(***)  |
| 17        | 5 mM/0.5 mM GSH/GSSG + 39 % glycerol (***)  |
| 18        | 20 mM citrate, pH 5.3 + 0.8 M urea (**)   |
| 19        | 0.75 M arginine   |
| 20        | 0.75 M arginin, 5 mM/0.5 mM GSH/GSSG (***)  |
| 21        | 0.57 mM CTAB/ 4.71 mM $\beta$ -cyclodextrin   |
| 22        | 0.57 mM CTAB/ 4.71 mM $\beta$ -cyclodextrin, 5 mM/0.5 mM GSH/GSSG (***)                         |
| condition | folding of His-tagged p22 dynactin on metal chelate resins                                      |
| 23        | folding on Ni-NTA-agarose, direct removal of urea   |
| 24        | folding on Ni-NTA-agarose, stepwise dilution of urea (8M, 4M, 2M, 1M)                           |
| 25        | folding on Fractogel EMD Chelate, direct removal of urea  |
| 26        | folding on Fractogel EMD Chelate, stepwise dilution of urea (8M, 4M, 2M, 1M)                    |
| 27        | folding on Ni-NTA-agarose, direct removal of urea + 0.03 % PEG 1000                             |
| 28        | folding on Fractogel EMD Chelate, direct removal of urea + 0.03 % PEG 1000                      |
| 29        | folding on Ni-NTA-agarose, direct removal of urea + 0.1 mM $\text{CaCl}_2/\text{MgCl}_2$        |
| 30        | folding on Fractogel EMD Chelate, direct removal of urea + 0.1 mM $\text{CaCl}_2/\text{MgCl}_2$ |

**Table 4.1: Composition of the automated folding screen**

Conditions 1 – 22 contain 20 mM Tris HCl, pH 7.7, 50 mM NaCl, 1 mM DTT, 0.2 mM  $\text{CaCl}_2/\text{MgCl}_2/\text{ZnCl}_2$  (when screening His-tagged p22 dynactin no  $\text{ZnCl}_2$  was used), except (\*) = no  $\text{CaCl}_2/\text{MgCl}_2/\text{ZnCl}_2$ , (\*\*) = no Tris or (\*\*\*) = no DTT (in case of refolding of KRAB protein, GSH/GSSG and Tris were substituted by borax, pH 10). Folding buffer for condition 23 – 30 is 20 mM Tris, pH 7.7, 30 mM NaCl, 2 mM  $\beta$ -mercaptoethanol with or without additives as indicated. % values for PEG are w/v and for glycerol v/v.

## Result I: Protein Folding Screening

folding buffer in 2.5 h intervals to achieve low concentrations of aggregation-sensitive folding intermediates but relatively high final concentrations of folded protein (Rudolph, 1990), thus facilitating detection. Before each addition, the rack containing plastic tubes with 1 ml screening solutions and small stirrers in each tube was transferred onto a stirrer which had been put on the robot work platform (Figure 4.1). The solution with denatured proteins was “spat” into each screening solution. Reliable dispensing of the small volumes of protein was achieved by optimisation of liquid handling parameters.

| <b>additive</b>                      | <b>effect</b>   |
|--------------------------------------|---|
| urea / arginine                      | Destabiliser, may shuffle aggregation-prone folding intermediates back to the native state; decelerates folding reaction  |
| PEG                                  | Can bind to early intermediates and prevent their self-association and aggregation  |
| glycerol                             | Destabilises the denatured state more than the native state, thus forcing proteins to fold; accelerates folding reaction  |
| buffer, different pH                 | pH values lower or higher than neutral and different buffer-molecules may stabilise or destabilise proteins, the effect on folding may be complex, disulfide bond formation is favoured at basic pH |
| GSH/GSSG                             | Facilitate reshuffling of incorrect disulfide bonds   |
| CaCl <sub>2</sub> /MgCl <sub>2</sub> | Ca <sup>2+</sup> or Mg <sup>2+</sup> may be cofactors; divalent ions may oxidise cysteines to form incorrect inter- or intramolecular disulfide bonds   |
| EDTA                                 | Chelates divalent ions hence protects proteins from oxidation   |

**Table 4.2: Effects on protein folding of different screening components for rapid dilution (see also 1.3, introduction)**

Two conditions using the detergent/cyclodextrin folding system were tested (Hanson & Gellman, 1998; Rozema & Gellman, 1996) (see 1.3) and eight conditions for folding of His-tagged proteins on metal chelate resins were also included in the automated screening procedure (see Table 4.1). Folding on metal chelate resins was performed in plastic tubes. After the beads (approx. 100 µl) settled on the bottom of the tube the supernatant can be discarded by the robot pipetting system. To achieve this, the pipetting needles detect the liquid surface and remove the liquid slowly while moving down with the liquid level. Using this procedure, the metal chelate resins can be washed with different buffers.



**Figure 4.1: Automation of the refolding screen**

Left picture: robot working platform, right picture: “folding rack” placed on stirrer and containing tubes with folding buffers and small stirring fish inside, a pipetting needle “spits” denatured protein sample on top of the stirred solutions

### **4.2 Evaluation of the automated folding screen and design of a detection system for folded proteins using model enzymes**

To evaluate the automated folding screen and methods to detect natively folded proteins, two model enzymes were analysed. The obtained data can be correlated to enzymatic activity.

#### **4.2.1 Refolding of MDH**

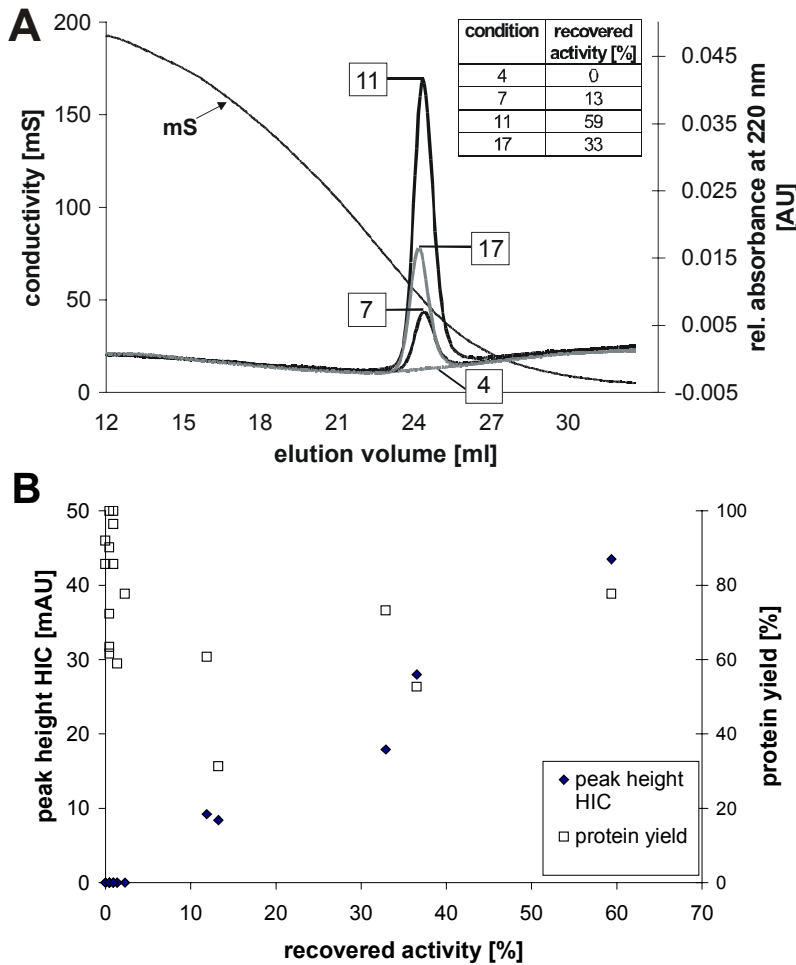
Malate dehydrogenase from pig heart (MDH) is a homodimeric enzyme of 36.5 kDa subunits. MDH was shown to be a substrate of the GroEL/GroES folding system during *in vitro* refolding (Weber et al., 1998). In this study, the chaperones enhanced MDH-refolding efficiency by more than a magnitude compared to refolding without GroEL/GroES (60–70% compared to less than 5%).

## Result I: Protein Folding Screening

MDH was refolded using the described automated screen after denaturation in 7 M GdnHCl (conditions 1–22, table 1). Conditions without glycerol resulted in no significant recovery of enzymatic activity ( $\leq 2\%$  yield of active MDH). Most beneficial for the folding in the presence of glycerol was EDTA (condition 11, 59% recovered activity). In addition a low pH (condition 14, 37%) and the presence of the redox pair GSH/GSSG (condition 17, 33%) favoured refolding of MDH.

For many proteins a biological assay may not be available as a folding monitor. In those cases, sensitive biophysical folding monitors to detect natively folded protein are needed which can be applied generally after folding screening. MDH is not usable for analysis with tryptophan fluorescence spectroscopy as it contains no such residues. In a new approach, analytical hydrophobic interaction chromatography (HIC) was tested as a folding monitor. During HIC, proteins are separated due to differences of surface hydrophobicity. Therefore, HIC should discriminate the compact native protein from partially folded or misfolded species that usually expose more hydrophobic patches.

Ammonium sulfate was added to all refolded samples before application onto a small column packed with a hydrophobic resin using the autosampler of an automated chromatography system. Bound protein was eluted in a decreasing ammonium sulfate gradient. Eluted protein was detected at 220 nm and not at 280 nm due to the relatively low extinction coefficient at 280 nm (see 3.5.4). All conditions that resulted in no enzymatic activity did not show an elution peak in HIC (e.g. condition four), whereas refolded active MDH eluted in a close range from 49 to 53 mS from the HIC column (see Figure 4.2 A). A linear correlation between recovered activity and peak height was observed, but protein yield (after removal of precipitates) was not correlated to activity (Figure 4.2 B). Condition 7 resulted in 13 % activity yield, which corresponds to 6  $\mu\text{g}$  native protein. This amount was clearly detected with the employed analytical HIC (Figure 4.2 A). In summary, it was shown that HIC can be used as a monitor for refolding MDH.



**Figure 4.2: Analysis of refolded MDH with HIC**

(A) Recovered enzymatic activity (inset) and elution profiles of selected refolding conditions (see table 4.1). (B) Correlation of recovered activity to peak heights (◆) observed during HIC and to protein yields (□, 100% yield corresponds to 45 μg/ml) for conditions 1–18; precipitates removed by centrifugation; conditions 19–22 excluded because arginine and CTAB are incompatible with the Bradford assay used to determine protein concentrations.

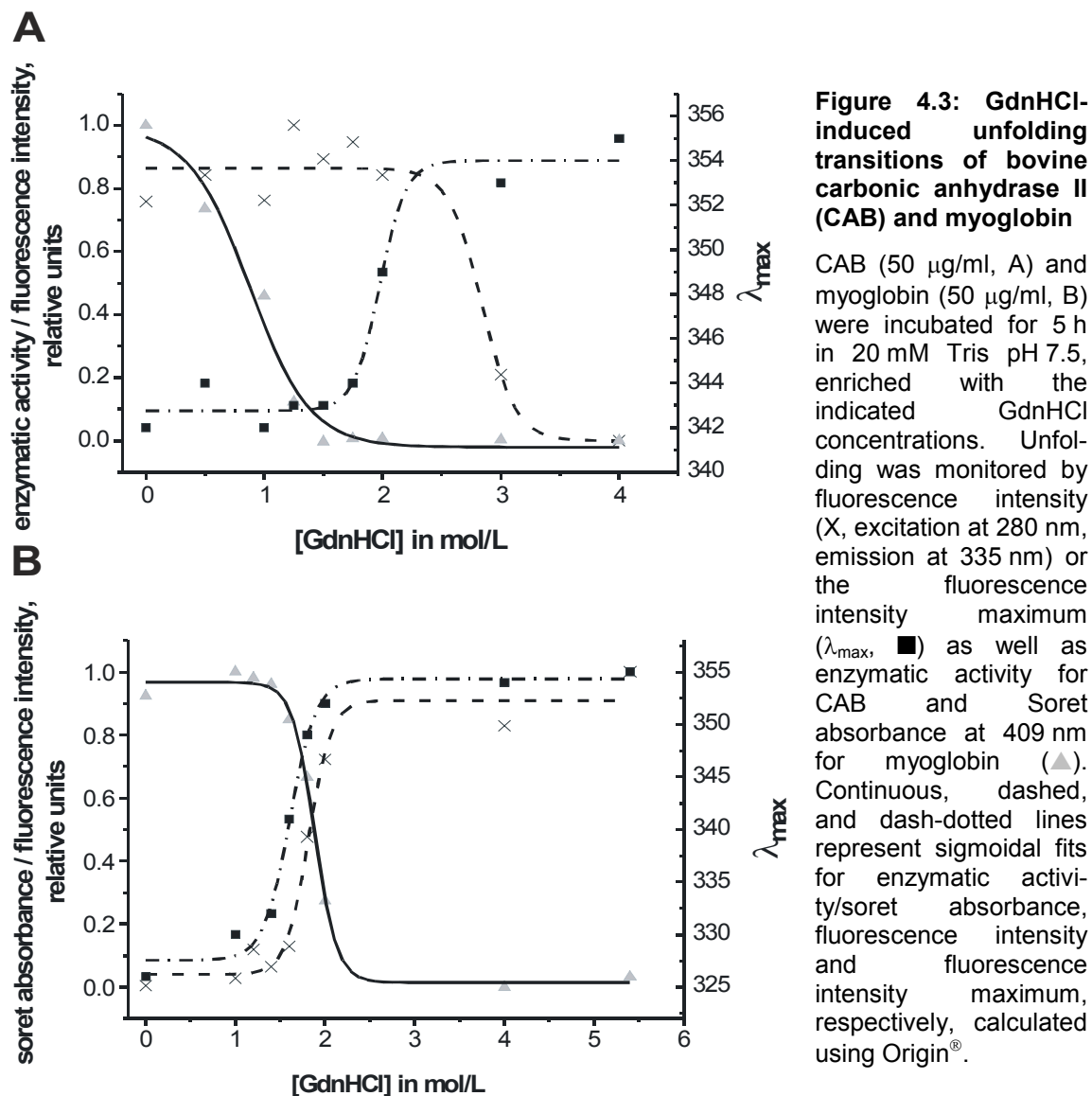
## 4.2.2 Tryptophan fluorescence spectroscopy of protein unfolding transitions

Because tryptophan fluorescence spectroscopy is a sensitive probe to distinguish unfolded, partially folded and folded protein (see 1.4), it was evaluated as a monitor for folded protein upon folding screening. In contrast to analytical HIC, samples for fluorescence spectroscopy need to be dialysed after refolding screening because the different screening components influence the fluorescence properties of tryptophans.

In a preliminary experiment, two proteins, CAB and myoglobin, were unfolded for 5 h at different concentrations of GdnHCl. Then fluorescence spectra were measured in a 96-well microplate. Additionally, enzymatic activity was determined for CAB and Soret absorbance for myoglobin at 409 nm. Soret absorbance occurs because the heme of

## Result I: Protein Folding Screening

native myoglobin is coordinated with the side chain of a histidine residue. Fluorescence signals were strong enough to allow comprehensive analysis although relatively low concentrations of the two proteins were used (50  $\mu\text{g/ml}$ ).



The fluorescence intensity maximum was shifted to a higher wavelength upon denaturation with GdnHCl for both proteins (see Figure 4.3). The fluorescence intensity decreased with higher concentrations of GdnHCl for CAB, probably due to solvent quenching. In case of myoglobin fluorescence intensity increased with higher GdnHCl concentrations, indicating that in native myoglobin tryptophan fluorescence is quenched

## Result I: Protein Folding Screening

---

by heme. With myoglobin, both processes, blue-shift of fluorescence maximum and loss of fluorescence intensity, occurred in parallel with loss of Soret absorbance. For CAB, they happened at higher concentrations than loss of enzymatic activity indicating the presence of a partially folded species having no activity but similar fluorescence properties as native CAB.

These results support the assumption that tryptophan fluorescence (assayed in 96-well format) may be useful to detect small amounts of folded protein after folding screening.

### 4.2.3 Refolding of CAB

CAB is a monomeric 28.5 kDa enzyme with seven tryptophan residues and a zinc ion in its catalytic centre. *In vitro* refolding of CAB is strongly hampered by aggregation. CAB has therefore been used as a model to study aggregation processes during refolding (Cleland & Wang, 1990). With this model enzyme the folding monitoring with analytical HIC that had been applied to MDH was repeated and additionally tryptophan fluorescence spectroscopy was tested. Figure 4.4 shows the elution profiles during HIC and the yield of enzymatic activity (inset) of four of the 22 conditions tested for the refolding of GdnHCl-unfolded CAB. All 22 chromatograms of HIC can be divided into three groups: (i) no peak observed (ii) peaks at approx. 125 mS and (iii) peaks at approx. 111 mS. Only conditions nine, ten and eleven belonged to group (iii). These three conditions were the only ones that contained EDTA instead of ZnCl<sub>2</sub> (Table 1). Hence, HIC could discriminate CAB molecules with a zinc ion bound from molecules with no zinc ion.

From the fluorescence spectra of refolded CAB (Figure 4.4) (i) the blue-shift of the fluorescence intensity maximum compared to the denaturant-unfolded protein and (ii) the fluorescence intensity at 340 nm were evaluated. The GdnHCl-unfolded CAB had a fluorescence intensity maximum of 355 nm. Spectra of refolded CAB that exhibited fluorescence had intensity maxima ranging from 355 nm down to 346 nm.

## Result I: Protein Folding Screening

**Figure 4.4: Analysis of refolded CAB with HIC and tryptophan fluorescence spectroscopy**

(A) Elution profiles and recovered enzymatic activity (inset) of selected refolding conditions (see table 1). (B) Fluorescence emission spectra (upon excitation at 280 nm) of the refolding conditions shown in (A) and of denatured CAB in 6 M GdmHCl (120  $\mu\text{g/ml}$ ), indicated with 'D'.

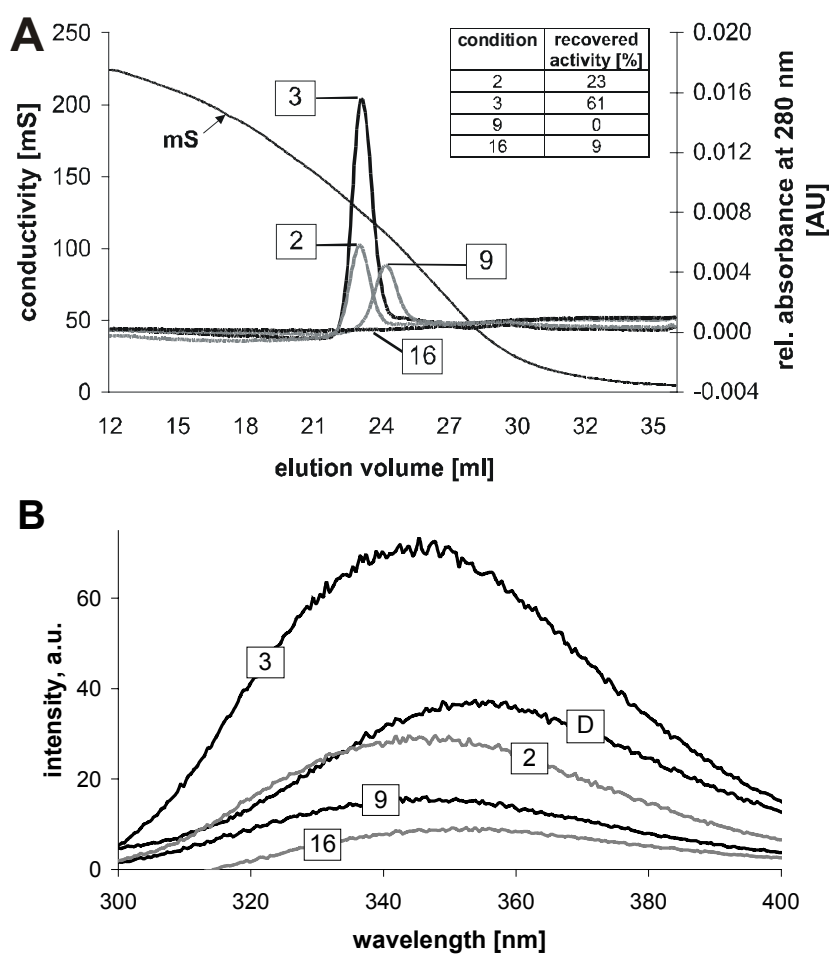
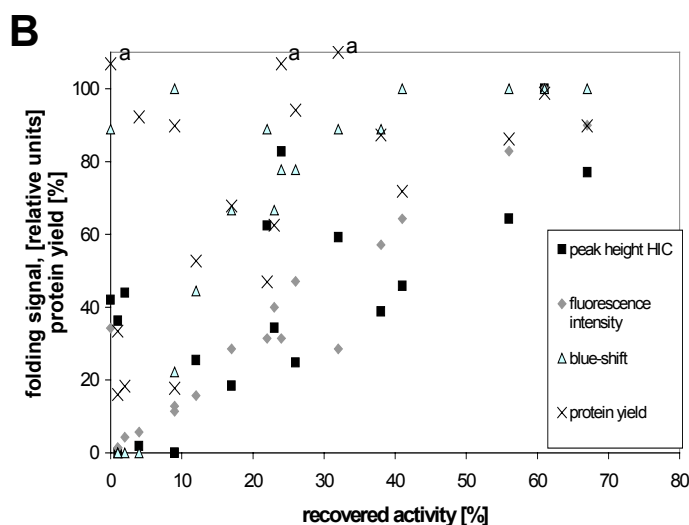
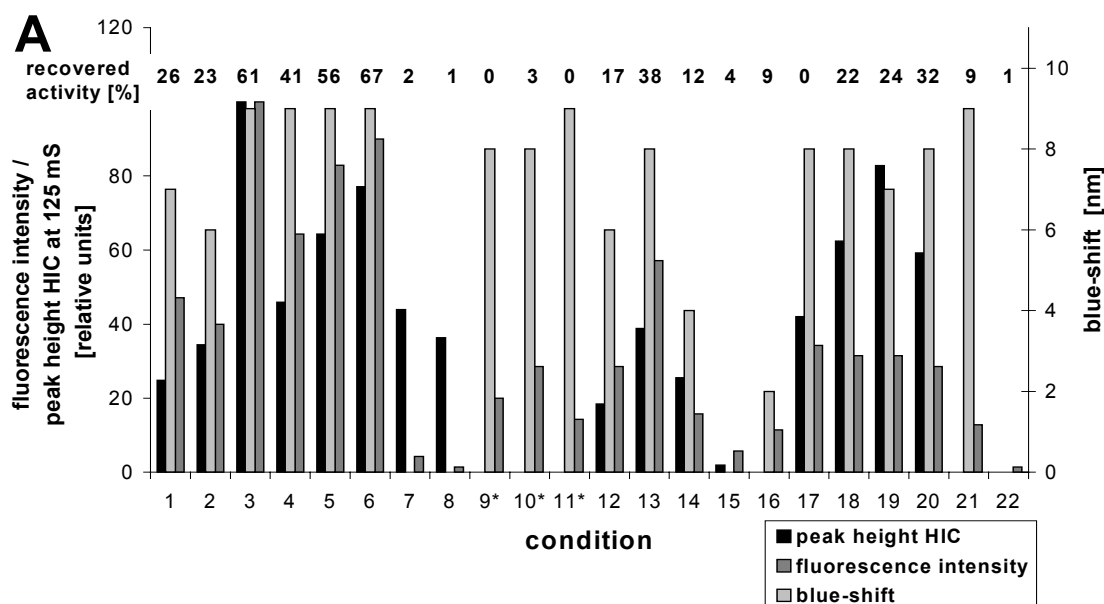


Figure 4.5 A summarizes all data obtained from HIC, tryptophan fluorescence spectroscopy and the recovery of enzymatic activity. Conditions three to six showed the highest fluorescence intensities of all tested conditions and, together with other samples (conditions 11 and 21), the strongest blue-shifts of the fluorescence intensity maxima

## Result I: Protein Folding Screening

(9 nm). The peak eluted from HIC at 125 mS for condition three, four, five and six was the highest, the seventh highest, the fourth highest and the third highest, respectively. These four conditions also showed the highest recovery of enzymatic activity (61%, 41%, 56% and 67%, respectively). It can be concluded that CAB can be refolded most efficiently in the presence of PEG 1000 and/or urea. This beneficial effect was reported previously (Cleland et al., 1992b).



**Figure 4.5: Correlation of recovered activities of refolded CAB to peak heights observed during HIC, fluorescence intensities, blue-shifts of maximum fluorescence intensities and protein yields**

(A) Schematic depiction of the different parameters (recovered enzymatic activity, peak heights observed during HIC at 125 mS, fluorescence intensities, blue-shifts of maximum fluorescence intensity and protein yields (after removal of precipitates with filtration)) measured for the tested conditions; \* = detected peaks during HIC eluted at 111 mS (see text and

figure 4.4). (B) Correlation of recovered activities to the folding signals obtained with peak heights of HIC at 125 mS (■), fluorescence intensities (◆), blue-shifts of maximum fluorescence intensities (△) and to protein yields (X, 100% yield corresponds to 120 µg/ml); conditions 9–11 are excluded due to a different behaviour in HIC (eluted peak at 111 mS). Protein yields calculated from absorbance at 280 nm, with correction for stray light using absorbance at 340 nm; note that despite this correction, for samples denoted with 'a' the strong stray light gave rise to yields higher than 100%. Fluorescence intensities were measured at 340 nm upon excitation at 280 nm. All values given as relative units with respect to the conditions that resulted in the highest value obtained with the respective folding monitor.

## Result I: Protein Folding Screening

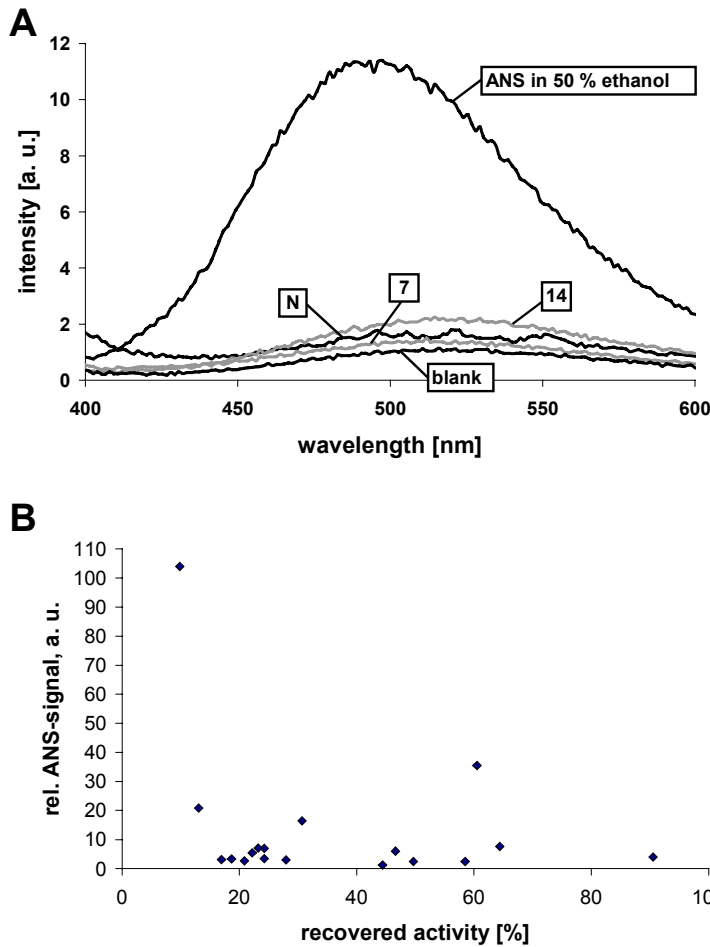
In Figure 4.5 B the recovery of enzyme activity is correlated to four parameters: (i) peak height at 125 mS during HIC, (ii) fluorescence intensity, (iii) blue-shift of the fluorescence intensity maximum, and (iv) protein yield. Conditions 9–11 of the folding screen were excluded from the graph due to their behaviour in HIC, as described above and shown in Figure 4.4. Fluorescence intensity showed a linear correlation to recovered activity, with the exception of condition 17 (a false positive) which had high intensity but no enzymatic activity. The values of the HIC peak heights could also be correlated linearly to activity. However, three false positives occurred (conditions 7, 8 and 17, compare with Figure 4.5 A). Additionally, conditions 16 and 21 could be considered as false negatives, because some enzymatic activity was detected (both 9%), but HIC showed no elution peak. For all samples with a recovered enzymatic activity higher than 15%, significant blue-shifts of fluorescence intensity maxima could be observed (Figure 4.5 B). Among the samples with less than 15% recovered activity shown in Figure 3B (seven data points), five samples showed small or no blue-shifts. The two remaining samples exhibited strong blue-shifts; one of them showed no recovered activity (condition 17) and the second (condition 21, Figure 4.5 A) had 9% recovered activity. Protein yield was the least reliable of the four parameters for analyzing the refolding screen of CAB (Figure 4.5 B). Figure 4.5 A demonstrates that the combination of three monitors was useful to rule out false positives or false negatives that were obtained with one monitor. Conditions seven and eight exhibited no fluorescence signal and correspondingly led to almost no activity yield (2% and 1%, respectively) even though peaks at 125 mS were observed during HIC. On the other hand, no recovered activities were observed for conditions nine and eleven, even though fluorescence was detected as well as a blue-shift of the fluorescence intensity maximum. However, these samples could be ruled out due to their deviating behavior in HIC. Condition 17 was the only condition showing no enzymatic activity, but positive signals with all three monitors. Considering all results, no false negatives occurred, because all conditions that led to enzymatic activity were either clearly detected by analytical HIC or showed a blue-shift in fluorescence spectroscopy. The beneficial effect of the CTAB/cyclodextrin-system on the refolding of CAB reported before (Rozema & Gellman, 1996) was not observed in our screen. This could be due to the higher protein concentration used in this study during refolding (120  $\mu\text{g/ml}$ , compared to 30  $\mu\text{g/ml}$ ).

## Result I: Protein Folding Screening

### **4.2.4 Evaluation of ANS fluorescence as a monitor for native protein after refolding screening of CAB**

8-anilino-1-naphthalenesulfonic acid (ANS) is a dye that has been widely used to detect the molten globule state of a protein because it undergoes an increase in fluorescence upon binding to hydrophobic surfaces (John et al., 2001). Less binding is observed for the completely unfolded and the completely folded protein since they do not provide large hydrophobic patches for ANS to bind. When denaturant-unfolded proteins are folded by transferring them to a buffer without denaturant, this solution will not contain completely unfolded protein any more. It will rather contain native protein and/or misfolded, partly folded or aggregated species. It may be assumed that the latter species exhibit pronounced ANS fluorescence compared to native protein and will therefore be distinguished from native protein. Hence it was tested with CAB whether low ANS fluorescence is an indicator for folded protein after refolding screening. CAB was refolded at a final protein concentration of 100  $\mu\text{g/ml}$  (3.5  $\mu\text{M}$ ). Prior to measurements, all samples were dialysed and filtrated to remove large aggregates. Protein yields and enzymatic activities were determined. ANS was added in a nine-fold molar excess to CAB. The detected signals from ANS fluorescence were very low, for a lot of samples even lower than of 100  $\mu\text{g/ml}$  native CAB (see Figure 4.6 A). Low ANS signals might not only be due to the fact that few molecules with large hydrophobic patches are present; low signals may also occur because the overall amount of protein present in the sample is low. Therefore, fluorescence signals needed to be related to protein yield (after removal of large aggregates by filtration), otherwise samples with low protein yield could appear as false positives. Some samples had significantly higher protein yield than yield of active enzyme. Obviously, these samples contained high amounts of incorrectly folded protein or soluble aggregates that could not be removed by filtration. However, in contrast to the assumption made above, relative ANS fluorescence was low in those cases. Figure 4.6 B shows that there was no correlation between relative ANS fluorescence and recovered activity.

## Result I: Protein Folding Screening



**Figure 4.6: ANS fluorescence of CAB after refolding screening**

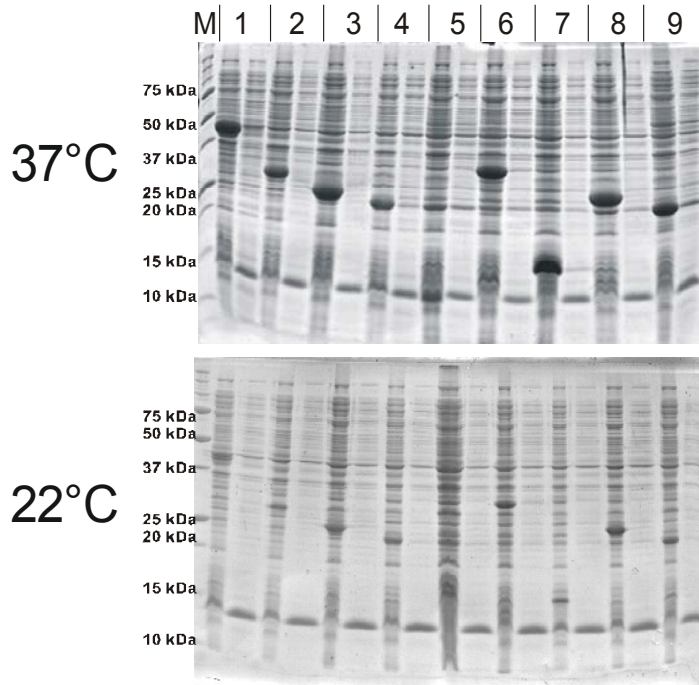
(A) ANS fluorescence emission spectra (30  $\mu\text{M}$  ANS, upon excitation at 350 nm) of native CAB (3.5  $\mu\text{M}$ , denoted with 'N') and two selected samples of the refolding screen. Denatured CAB (3.5  $\mu\text{M}$  in 7 M GdmHCl) exhibited an almost identical fluorescence spectrum to the native spectrum. Note that fluorescence intensities of all refolded samples were in a close range between 1 and 2.5 a.u. A spectrum of ANS in 50% ethanol is also shown. (B) Correlation of recovered activities of refolded CAB to relative ANS fluorescence. (ANS-fluorescence was related to protein yield, see text). Only samples with a protein yield  $\geq 10\%$  are depicted.

### 4.3 Folding screening of proteins expressed in *E. coli* inclusion bodies

The screening system established for the two model enzymes was applied to the folding of human proteins (PSF-target proteins, see page 1) that are expressed in inclusion bodies in *E. coli*. Lowering the temperature from 37 °C to 22°C during expression of these proteins did not result in soluble expression (see Figure 4.7). Sequence analysis of those expression clones revealed that for one protein, the hypothetical protein DKFZp434B156, a 17 amino acid deletion at the N-terminus was introduced which evolved during PCR-amplification of the source cDNA-clone. In another case, sequence comparison revealed that the expressed sequence did not span the full protein sequence but missed N-terminal parts (hypothetical protein DKFZp762K197). This may

## Result I: Protein Folding Screening

be due to a wrong annotation of cDNA start codon. In case of 60 kDa SS-A/Ro alternative protein 60e2 the template cDNA clone was truncated 20 bp upstream the 3' end. These three proteins were therefore not used for folding analysis. The remaining six expression clones contained no or single point mutations (see Table 4.3 page 59) and were subjected to folding screening.



**Figure 4.7: Analysis of expression clones for folding screening with SDS-PAGE**

Protein expression was induced for 3 h at 37 °C and for 6 h at 22 °C. The first lane per clone represents the total cellular protein and the second lane shows soluble cellular protein. M : molecular weight standard, 1: ubiquitin fusion degradation 1 like protein, 2: 60 kDa SS-A/Ro alternative protein 60e2, 3: hypothalamus protein HT002, 4: p22 dynactin, 5: KRAB, 6: CCR4-associated factor1, 7: hypothetical protein DKFZp564M173, 8: hypothetical protein DKFZp762K197, 9: hypothetical protein DKFZp434B156 (see also Table 4.3, page 59)

Proteins were solubilized with GdnHCl and purified under denaturing conditions using metal chelate chromatography. Bound protein was eluted using imidazole, EDTA was added to capture contaminating  $\text{Ni}^{2+}$  and the sample was reduced with 15 mM DTT for several hours at room temperature prior to folding screening with rapid dilution. Sample purity was higher than 90 % in all cases, judged from SDS-PAGE analysis (data not shown).

The N-terminal His-tag was not only used for purifying the protein prior to the folding screen, but also to test folding of proteins that are immobilised on metal chelate resins (Table 4.1, page 39, conditions 23–30). During this procedure, an interaction of the folding protein molecule with the chromatographic resin can be counterproductive for refolding. This interaction was shown to be dependent on the salt concentration when

## Result I: Protein Folding Screening

---

folding poly-Arg-tagged  $\alpha$ -glucosidase on heparin sepharose (Stempfer et al., 1996). In this study, NaCl concentrations in the range of 30 mM gave the best results for folding. Therefore, this salt concentration was chosen in the attempts to fold His-tagged proteins on Ni-NTA-agarose or Fractogel EMD Chelate (loaded with  $\text{Ni}^{2+}$ ). In contrast to Ni-NTA-agarose, Fractogel EMD Chelate is a chromatographic support modified according to the tentacle technology (Merck). Long chains of iminodiacetic acids act as functional groups that bind  $\text{Ni}^{2+}$ . This leads to protein immobilisation in greater distance to the chromatographic support and should reduce unfavourable interaction with the support. Before binding purified and reduced denatured proteins to the metal chelate resin, DTT and imidazole were removed by buffer exchange. All expression constructs contained a TEV protease cleavage site to remove the N-terminal His-tag.

### **4.3.1 Screening for folding conditions for p22 dynactin**

Dynactin is a macromolecular complex that is an activator for cytoplasmic dynein-mediated transport (Gill et al., 1991). The first report on the smallest subunit, p22 (21 kDa), was by Holzbaur and co-workers (Karki et al., 1998). In this study the authors noted a novel localization for cytoplasmic dynein and dynactin at the cleavage furrow and the midbody of dividing cells. So far, no structural information of the dynactin subunits is known. For the p22 subunit the closest sequence homologue is the  $\beta$ -domain of streptokinase, with an identity of 25% covering 53% of the sequence of p22 dynactin. Therefore it was attempted to obtain a preparation of pure and native p22 dynactin for protein crystallisation screening carried out at the PSF.

## Result I: Protein Folding Screening

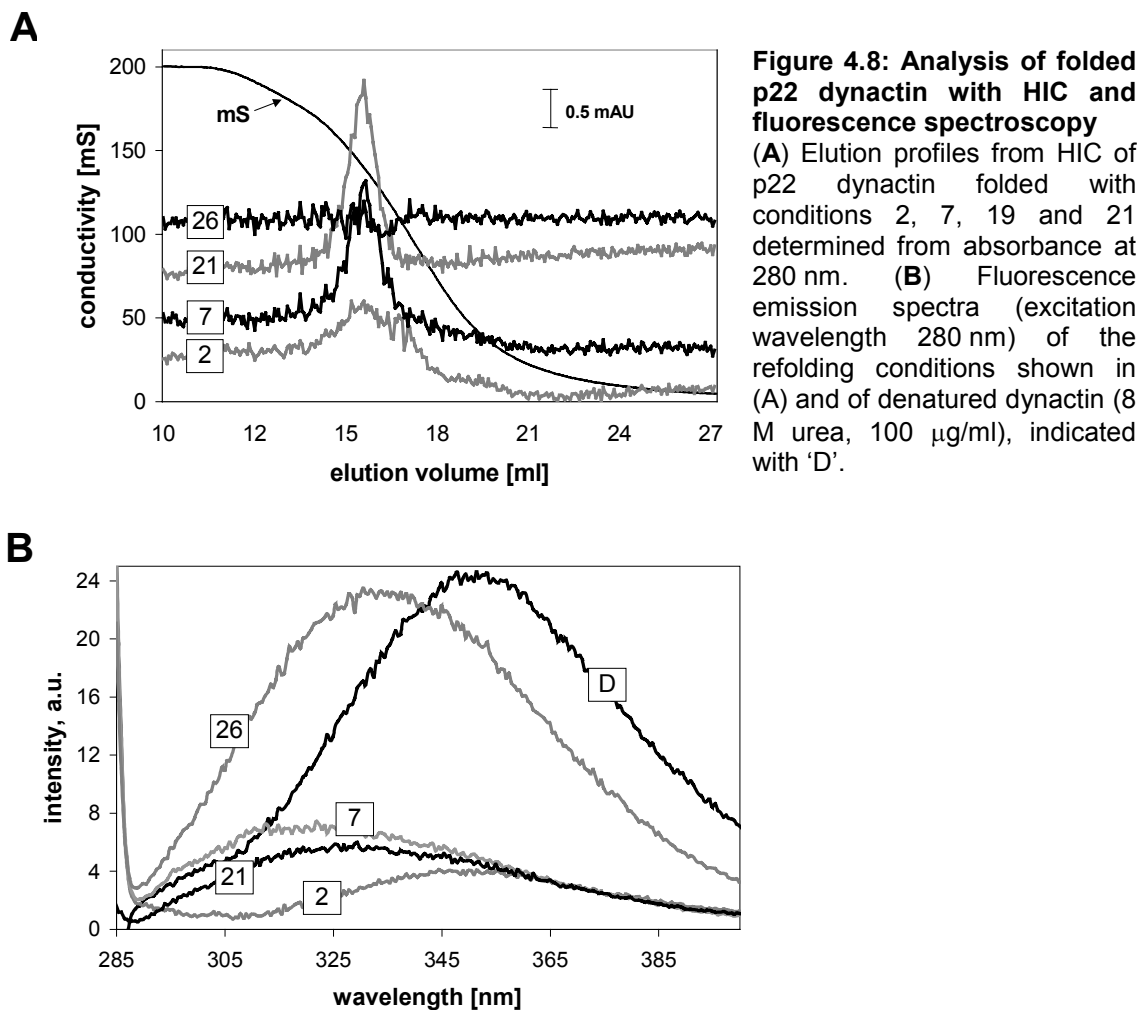
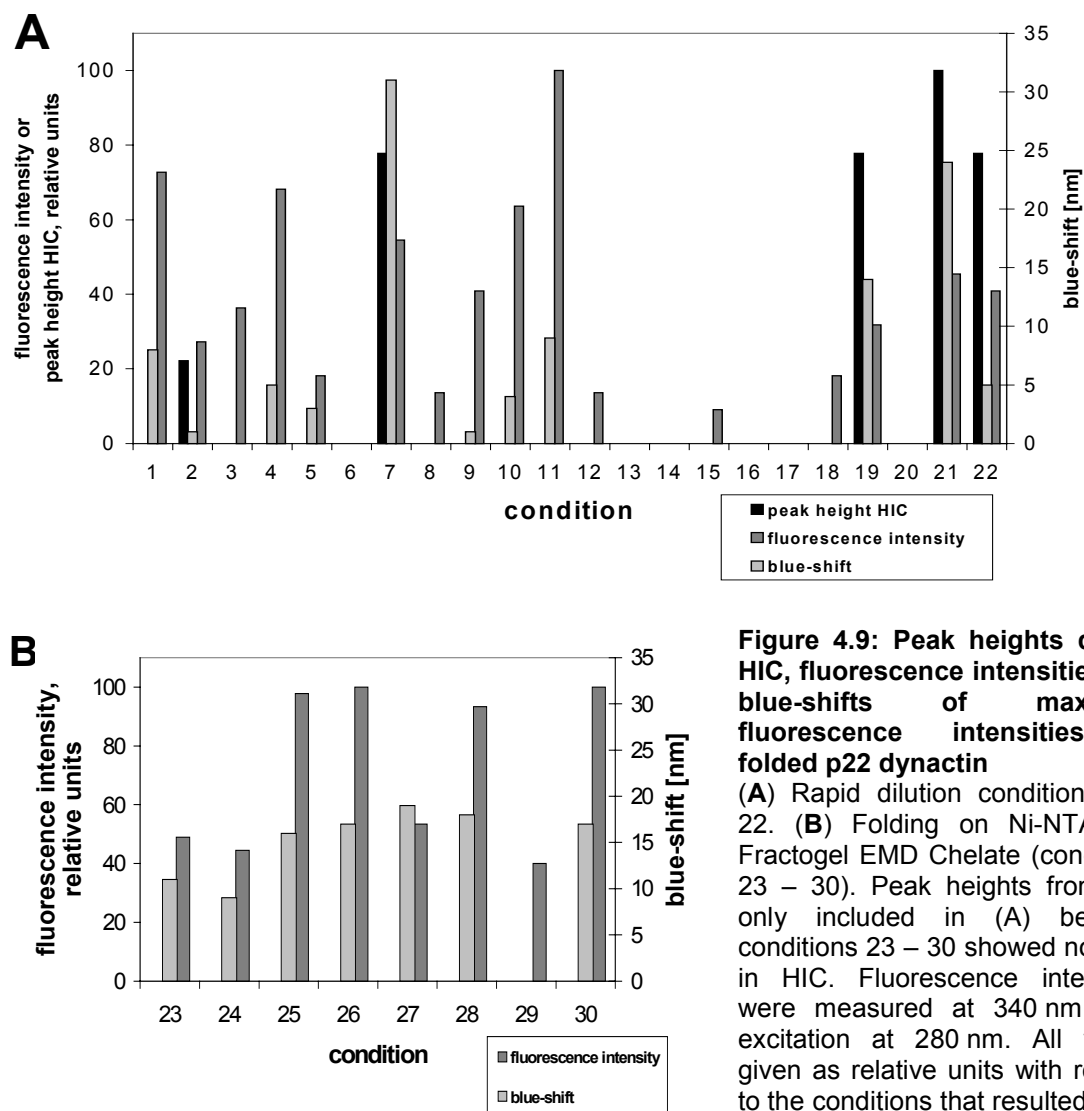


Figure 4.8 shows the results of three folding conditions of p22 dynactin that gave a peak during HIC, plus condition 26 with no peak. All peaks during HIC eluted at approx. 138 mS. The highest peaks were observed with the CTAB/cyclodextrin-conditions 21 and 22, as well as with condition 19 and condition 7 (Figure 4.9 A). Conditions 7, 19 and 21 also showed a strong blue-shift of the fluorescence maximum compared to denatured p22 dynactin. The sample obtained from condition 26 had no peak in HIC but exhibited both strong fluorescence and a significant blue-shift, as shown in Figure 4.8 B. It has to be noted that more denatured protein was used for the conditions 23–30 (folding of p22 immobilised on metal chelate resins) than for 1–22 (folding by rapid dilution). Therefore, fluorescence intensities between those conditions could not be compared directly.

## Result I: Protein Folding Screening

Intensities of rapid dilution samples were relatively low as p22 dynactin only contains two tryptophans.



**Figure 4.9: Peak heights during HIC, fluorescence intensities and blue-shifts of maximum fluorescence intensities of folded p22 dynactin**

(A) Rapid dilution conditions 1 – 22. (B) Folding on Ni-NTA and Fractogel EMD Chelate (conditions 23 – 30). Peak heights from HIC only included in (A) because conditions 23 – 30 showed no peak in HIC. Fluorescence intensities were measured at 340 nm upon excitation at 280 nm. All values given as relative units with respect to the conditions that resulted in the highest value obtained with the respective folding monitor. Rapid dilution conditions and the conditions for folding on metal chelate resins are shown separately because different amounts of denatured protein were used for those two groups of conditions.

dilution conditions and the conditions for folding on metal chelate resins are shown separately because different amounts of denatured protein were used for those two groups of conditions.

Figure 4.9 shows that conditions 2, 7, 19, 21 and 22 showed “positive” signals for all measured parameters, with condition 2 showing the lowest signals. For condition 22, a small blue-shift of the fluorescence maximum was observed in comparison to the other four conditions.

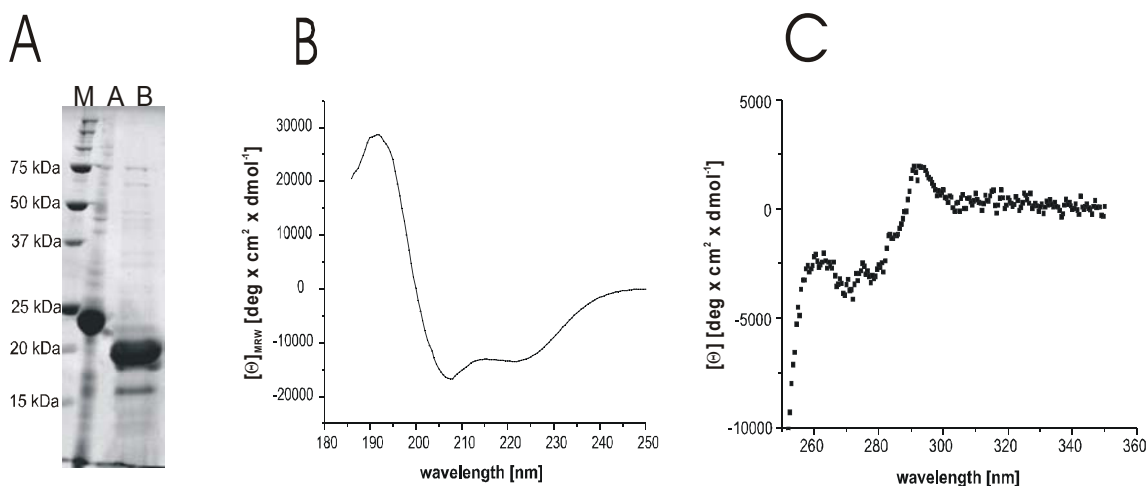
## Result I: Protein Folding Screening

From these results it can be assumed that p22 dynactin can be folded using glycerol, arginine or the CTAB/cyclodextrin system. The presence of 5 mM/0.5 mM GSH/GSSG seems to be counterproductive. None of the samples folded on metal chelate supports gave a peak in HIC, even though their fluorescence intensities were high and a significant blue-shift could be observed – especially for the samples that had been refolded on Fractogel EMD Chelate. Hence, as with MDH (see 4.2.1), HIC seems to be a reliable monitor for analysing the folding screen for p22 dynactin.

### 4.3.2 Folding, purification and detailed biophysical analysis of p22 dynactin

p22 dynactin was folded on a large scale using condition 21. Inclusion bodies resulting from a 4 L *E. coli*-culture were solubilised in 6.6 M GdnHCl and His-tagged p22 dynactin was purified using metal chelate chromatography under denaturing conditions. 75 mg denatured p22 dynactin were diluted into a CTAB-containing buffer. After addition of cyclodextrin to the mixture to induce folding, TEV protease was added to remove the His-tag and the sample was incubated overnight at 18 °C. Folded p22 was then purified in two further steps using the same HIC-resin used for analytical HIC and size exclusion chromatography. The yield was 25 mg protein, which corresponded to 33% of the denatured protein subjected to the folding and purification procedure. The molecular size detected by size exclusion chromatography corresponded to monomeric protein. Figure 4.10 shows a SDS-PAGE analysis of urea-denatured and purified p22 dynactin. The major impurity after folding and purification was a band at 17 kDa, which most likely was a degradation product of p22 dynactin that was formed during the incubation time of the folding reaction. It was not a product of TEV protease cleavage but seemed to result from a protease that was folded in small amounts together with p22 dynactin (data not shown). This protease was separated from p22 dynactin during purification. The truncated p22 dynactin could not be entirely removed during the purification procedure. The preparation was analysed using CD spectroscopy. A near-UV spectrum of p22 dynactin (Figure 4.10 C) exhibited pronounced fine structure, showing that the aromatic residues were immobilised in an asymmetric environment and that the protein

## Result I: Protein Folding Screening

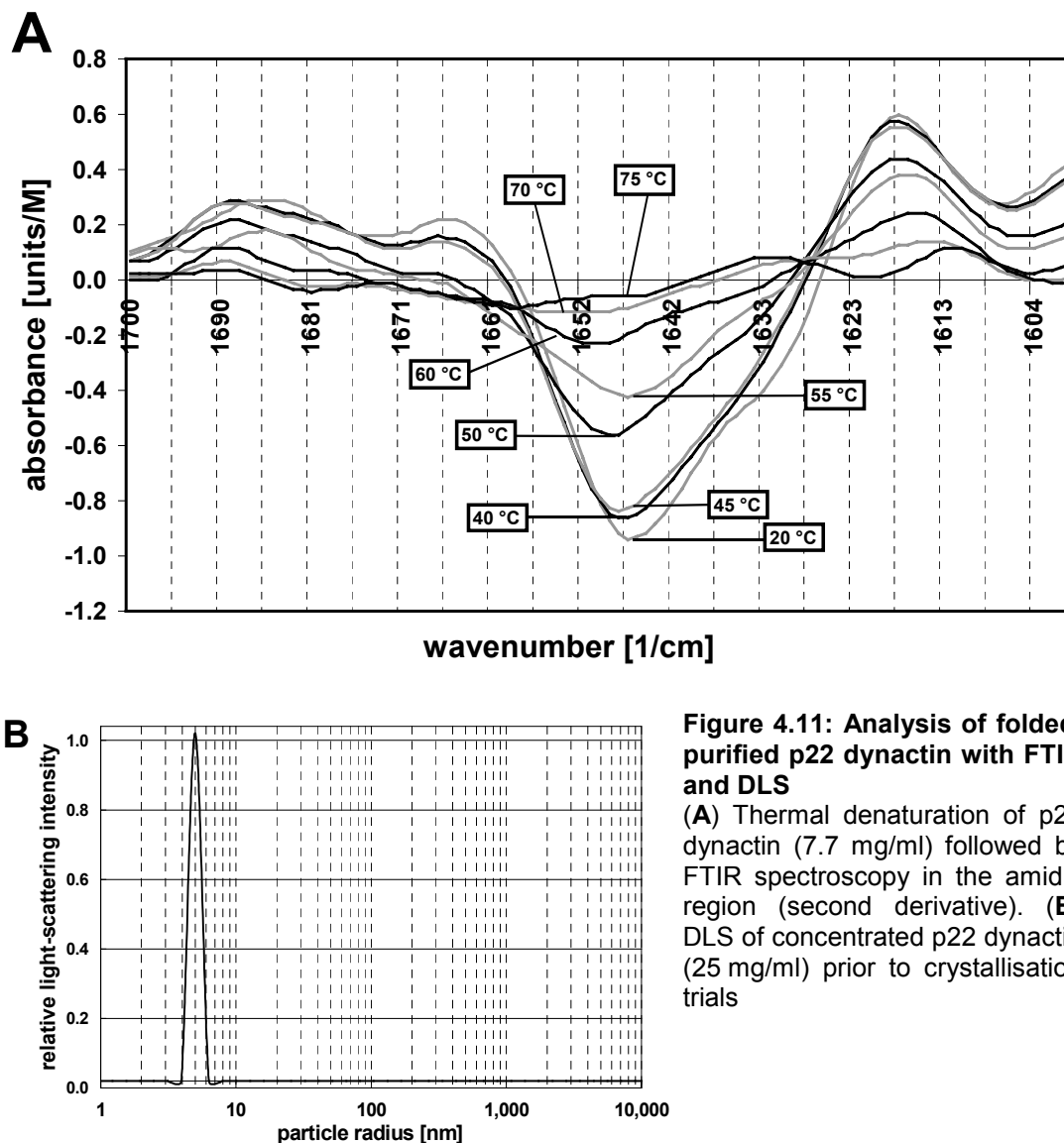


**Figure 4.10 – Analysis of folded, purified p22 dynactin with SDS-PAGE and CD spectroscopy**

(A) Lane A - denatured protein after purification with Ni-POROS20, lane B - natively folded, purified protein. Due to removal of the His-tag with TEV protease the native protein is approx. 3 kDa smaller than the denatured protein. M = molecular weight marker. (B) and (C) far- and near- UV CD spectra, respectively. Sample concentration: 1.03 mg/ml.

attained a well-defined, stable fold. The far-UV spectrum (Figure 4.10 B) indicated a high  $\alpha$ -helix content (minima at 222 nm and 208 nm), in accordance with secondary structure prediction for p22 dynactin (63%  $\alpha$ -helix, 9%  $\beta$ -sheet, calculated with PSIPRED, see supplement). FTIR spectra recorded at temperatures between 20 °C and 40 °C (Figure 4.11 A) showed a strong amide I band at 1646 – 1648  $\text{cm}^{-1}$ , also suggesting an  $\alpha$ -helical fold.

Thermal denaturation was monitored by FTIR spectroscopy. Loss of secondary structure started at 45 – 50 °C and was complete at 70 °C (Figure 4.11 A) as can be observed by the strong decrease of absorption of the  $\alpha$ -helix band concomitant with a slight band broadening. At 75 °C a weak absorbance band appeared at 1620  $\text{cm}^{-1}$  indicating intermolecular hydrogen bonding. Folded p22 dynactin could be concentrated to 25 mg/ml without any occurrence of aggregation, as was shown by dynamic light scattering (a hydrodynamic diameter of 4.9 nm was observed, Figure 4.11 B). p22 dynactin was subjected to crystallisation trials (Heinrich Delbrück, PSF) using 25 mg/ml but only small crystals unsuitable for X-ray experiments could be obtained so far (data not shown).



**Figure 4.11: Analysis of folded, purified p22 dynactin with FTIR and DLS**  
**(A)** Thermal denaturation of p22 dynactin (7.7 mg/ml) followed by FTIR spectroscopy in the amid I region (second derivative). **(B)** DLS of concentrated p22 dynactin (25 mg/ml) prior to crystallisation trials

### 4.3.3 Unsuccessful folding screening

Table 4.3 shows a list of proteins that were expressed as inclusion bodies at the PSF and that were subjected to *in vitro* folding screening. Although some proteins contained single point mutations that may influence folding, they were considered for screening. Mutations evolved during PCR-amplification of the source cDNA-clone.

Except for p22 dynactin, for all proteins listed in Table 4.3 a peak in a decreasing ammonium sulfate gradient during analytical HIC was not observed with any condition

## Result I: Protein Folding Screening

---

after folding screening. In some cases fluorescence spectroscopy gave promising conditions as high fluorescence intensity as well as a blue-shift of the fluorescence maximum was observed. Those conditions were further investigated by folding the respective proteins on a larger scale (usually 5 mg). Half of the volume of the resulting samples were subsequently concentrated by ultrafiltration and analysed with DLS, the other half was submitted to ion exchange chromatography. In some cases, the majority of the protein preparation was lost during concentration, presumably it bound irreversibly to the filter membrane due to high hydrophobicity. If this was not the case, the concentrated sample showed large particles in DLS, or - as shown in Fig 4.12 in case of CCR4 - an aggregation reaction was occurring. Figure 4.12 shows that the tryptophan fluorescence spectrum of an aggregated protein may exhibit a substantial blue-shift of intensity maximum compared to the denatured protein, therefore giving rise to false positives during analysis of a folding screen with fluorescence spectroscopy only.

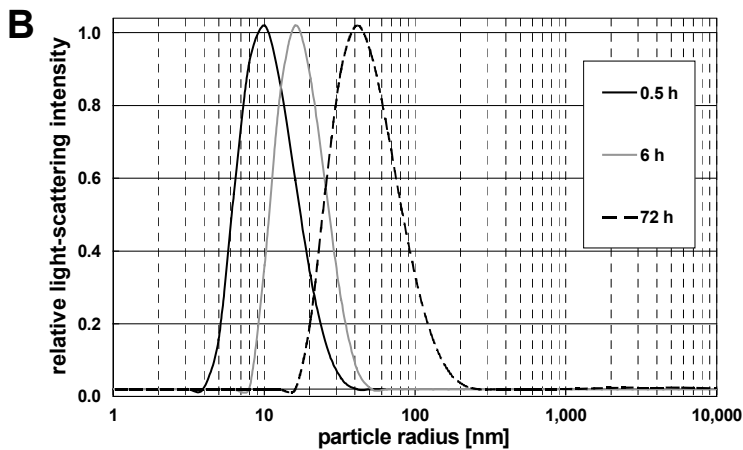
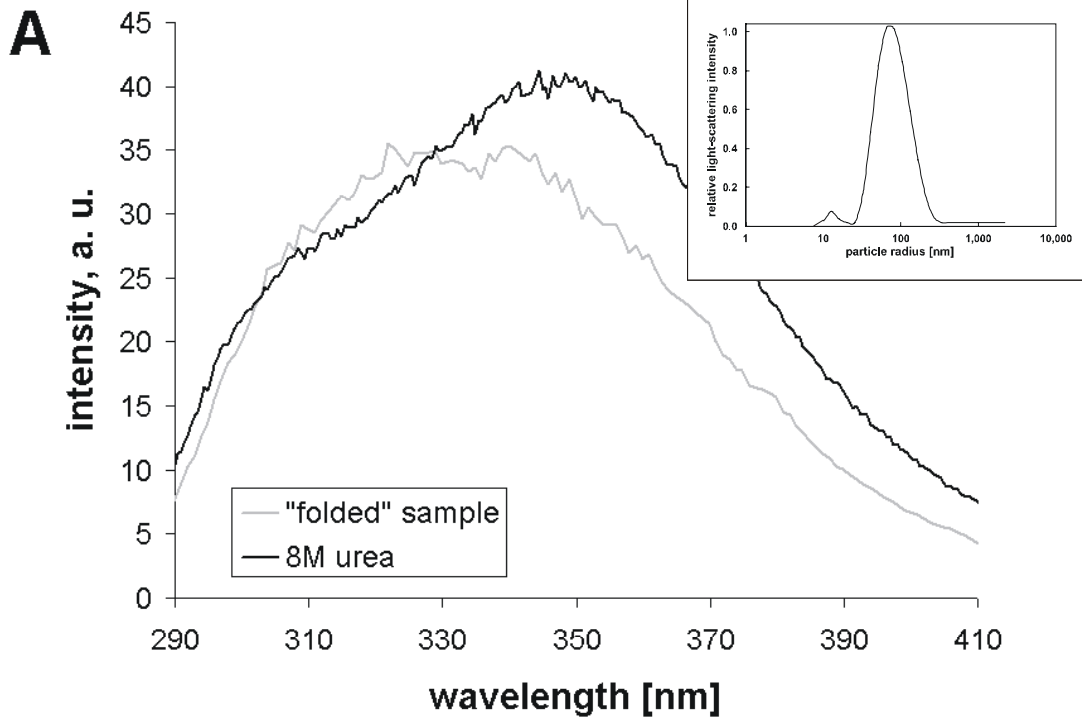
Ion exchange chromatography of protein samples folded under conditions leading to “promising” tryptophan fluorescence spectra was performed. In all cases the yield was very low and/or the protein eluted erratically in a NaCl-gradient (i.e. not in a distinct peak), probably because of sample inhomogeneity. Hence, protein preparation failed with those samples.

## Result I: Protein Folding Screening

| W/Y/C  | GenBank accession number | Name or NCBI-entry                          | Function/ Characteristics  | MW | Mutations   |
|--------|--------------------------|---|--|----|---|
| 2/7/2  | AAC61280                 | p22 dynactin subunit                        | see 4.3.1  | 21 | D134→G  |
| 1/4/0  | AAD20972                 | kruppel associated box (KRAB) protein       | transcriptional repressor (Collins et al., 2001)                   | 14 | T30→S   |
| 3/5/6  | AAH17366                 | CNOT8 protein                               | RNA processing and modification                                    | 27 | A5→T  |
| 3/13/6 | AAD28788                 | ubiquitin fusion-degradation 1 like protein | see name   | 42 | contains 36 amino acid deletion which obviously is a splice variant |
| 3/12/5 | AAD02685                 | CCR4-associated factor 1                    | part of the CCR4-NOT complex, a transcriptional regulatory complex | 36 | E5→G  |
| 2/12/3 | CAB66579                 | hypothetical protein DKFZp564M173           | contains an Ocnus-domain   | 17 | no mutation   |

**Table 4.3:** Proteins with insoluble expression (see Figure 4.6) that were submitted to the automated folding screen/biophysical analysis procedure (NCBI: National Center of Biotechnology Information, MW: molecular weight).

# Result I: Protein Folding Screening



**Figure 4.12: Analysis of ubiquitin fusion-degradation 1 like protein and CCR4 after folding screening**

(A) Fluorescence spectra of ubiquitin fusion-degradation 1 like protein in 8 M urea (100  $\mu\text{g/ml}$ ) and after folding using condition 2 from Table 4.1 ("folded" sample), the inset shows DLS analysis of the "folded" sample. (B) Aggregation over time of CCR4 upon folding using condition 23 from Table 4.1 followed by DLS

### 5 Results II: Analysis of human protein domains

#### 5.1 Domain selection and cloning

A human cDNA expression library from fetal brain (hEx1) (Büssow et al., 2000) was screened for clones that express their inserts as soluble proteins. Such expression clones were identified for 1,200 human proteins. These protein sequences were analysed using the SMART database of Hidden Markov Models (Letunic et al., 2002; Schultz et al., 2000). A total of 26 different protein domains (contained in 30 different proteins and represented by 41 clones) for which no three-dimensional structure was solved at that time. This analysis was performed by Konrad Büssow (PSF), Bernd Simon and Ivica Letunic (both EMBL).

For each domain, it was attempted to generate four constructs: one corresponding to the SMART-prediction and three further constructs with either five amino acid extensions at the N- or C-termini or at both. This is indicated in Table 5.1 by the columns 'start residue' and 'end residue'. However, some constructs could not be generated due to failure during PCR-amplification or transformation (e.g. two of four L27 domain constructs, Table 5.1). All constructs were cloned into an *E. coli* expression vector pQTEV that allows for expression of His-tag fusion proteins. Between the His-tag and the target protein a TEV protease recognition site was placed such that tags could be cleaved off during purification. The clones were provided by Volker Sievert (PSF).

G\_patch domains of four different source proteins (CAD38742.1, AAH07871.1, CAA68976.1, AAA99715.1) and JAB\_MPN of two source proteins (AAD03468.1, AAD03465.1) were cloned, whereas all other domains originate from one representative protein (Table 5.1). Doublecortin (AAC31696.1) contains two adjacent DCX-domains (from residue 48–139 and 175–263). Constructs of the single DCX-domains as well as others covering both domains were generated. The analysed SEP domain from protein p47 (AAH02801.1) was cloned individually (residues 176–270) as well as together with its neighbouring UBX domain (residues 289–368). Altogether, 88 expression constructs were generated, representing 17 different domains.

## Results II: Analysis of human protein domains

---

### 5.2 Solubility of different domain constructs

Expression of the cloned constructs was tested at 37 °C and 28 °C. Domains fused to N-terminal His-tags were purified with metal chelate chromatography using an automated purification procedure (Scheich et al., 2003) (see Materials and Methods). This procedure was developed to purify protein from 1 ml-cultures in 96-well format. In brief, cell pellets in 96 deep-well plates were resuspended in lysis buffer and subsequently lysed with lysozyme. A small sample representing the total cellular protein was withdrawn for SDS-PAGE analysis. Affinity beads were added and the plate was shaken to bind affinity-tagged proteins. For His-tagged proteins analysed in this study, Ni-NTA-agarose was used, but other affinity chromatography procedures (e.g. glutathione S-transferase-tag purification) may also be applied. After binding, the cell extract/affinity bead suspension was transferred onto a filter plate that retains only the affinity beads. These were washed and bound protein was eluted with imidazole in a small volume (70 µl) to obtain a relatively concentrated sample (see 3.4.7). This procedure had previously been shown to be robust and up to 90 µg of His-tagged protein could be purified when a high-expression clone was used (Scheich et al., 2003).

Samples (total cellular protein and purified protein) were analysed with SDS-PAGE. One domain could not be expressed (OSTEO), seven could only be expressed insolubly (CNH, RWD, JAB\_MPN, MA3, eIF5C, SPRY and TLDC) and the PRY domain only yielded a small fraction of solubly expressed protein (see Table 5.1). These constructs were therefore not considered for further detection of natively folded protein (see 5.3). The Zpr1 domain could only be expressed solubly at 28 °C. In summary, for 36 of the 88 expression constructs (representing 9 different domains) a reasonable amount of purified protein could be detected, judged from SDS-PAGE analysis (entries “+” and “++”, Table 5.1).

For some domains only particular constructs lead to soluble expression. For instance, the N-terminal zinc finger (Zpr1) domain of protein AAC33514.1 could only be expressed solubly when five amino acid residues were added on both N- and C-terminus (PTEIE and PHAPQ, respectively) to the *in silico* predicted Zpr1 sequence (Table 5.1). These extensions may represent non-conserved regions of the Zpr1 domain but are necessary for soluble expression (see 1.5). In contrast, the second Zpr1 domain of protein AAC33514.1 was only found in the soluble fraction when its shorter construct lacking the C-terminal extension (VYAPE) was expressed. Three of the four analysed G\_patch

## Results II: Analysis of human protein domains

domains behaved in a similar manner, i.e. particular constructs of a domain yielded more soluble protein than others (see Table 5.1).

| Domain   | Protein ID | start residue   | end residue    | total cellular protein | soluble purified protein |
|----------|------------|-----------------|----------------|------------------------|--------------------------|
| CNH      | AAB48435.1 | 506/511         | 825/830        | ++                     | -                        |
| RWD      | CAB52345.1 | 8/13            | 134/140        | +                      | -                        |
| SEP-UBX  | AAH02801.1 | 171/176         | 368            | ++                     | ++                       |
| SEP      | AAH02801.1 | 171/176         | 270/275        | ++                     | ++                       |
| L27      | AAG34117.1 | 13              | 68/73          | ++                     | ++                       |
| Zpr1     | AAC33514.1 | <b>252/257</b>  | <b>416/421</b> | ++                     | ++                       |
| Zpr1     | AAC33514.1 | <b>44/49</b>    | <b>207/212</b> | ++                     | ++                       |
| ZnF_C3H1 | AAF28981.1 | 126/131         | 157/162        | ++                     | ++                       |
| G_patch  | CAD38742.1 | 406/ <b>411</b> | <b>459/464</b> | +                      | +                        |
| G_patch  | AAH07871.1 | <b>228/233</b>  | <b>279/284</b> | +/-                    | ++                       |
| G_patch  | CAA68976.1 | 92/97           | 142/147        | +                      | ++                       |
| G_patch  | AAA99715.1 | <b>736/741</b>  | <b>787/792</b> | +/-                    | +                        |
| OSTEO    | AAA59974.1 | 14/19           | 314            | -                      | -                        |
| ZnF_RBZ  | CAC28312.1 | 185             | 214            | +                      | ++                       |
| PINT     | AAB58732.1 | 316/321         | 404/409        | +                      | +                        |
| JAB_MPN  | AAD03468.1 | 54              | 191            | +                      | -                        |
| JAB_MPN  | AAD03465.1 | 33/38           | 172/177        | ++                     | -                        |
| DCX      | AAC31696.1 | 85/90           | 181/186        | ++                     | ++                       |
| DCX      | AAC31696.1 | 85/90           | 305/310        | ++                     | ++                       |
| DCX      | AAC31696.1 | 212/217         | 305/310        | ++                     | ++                       |
| MA3      | AAC02903.2 | 1215/1220       | 1332/1327      | ++                     | -                        |
| eIF5C    | AAC02903.2 | 1487/1492       | 1579/1583      | ++                     | -                        |
| PRY      | AAA36564.1 | 310/315         | 367/372        | ++                     | +/-                      |
| SPRY     | AAA36564.1 | 363/368         | 493/498        | ++                     | -                        |
| TLDc     | BAB13435.1 | 250/255         | 424/429        | ++                     | -                        |

**Table 5.1: : Expression strength and solubility of the domain constructs.**

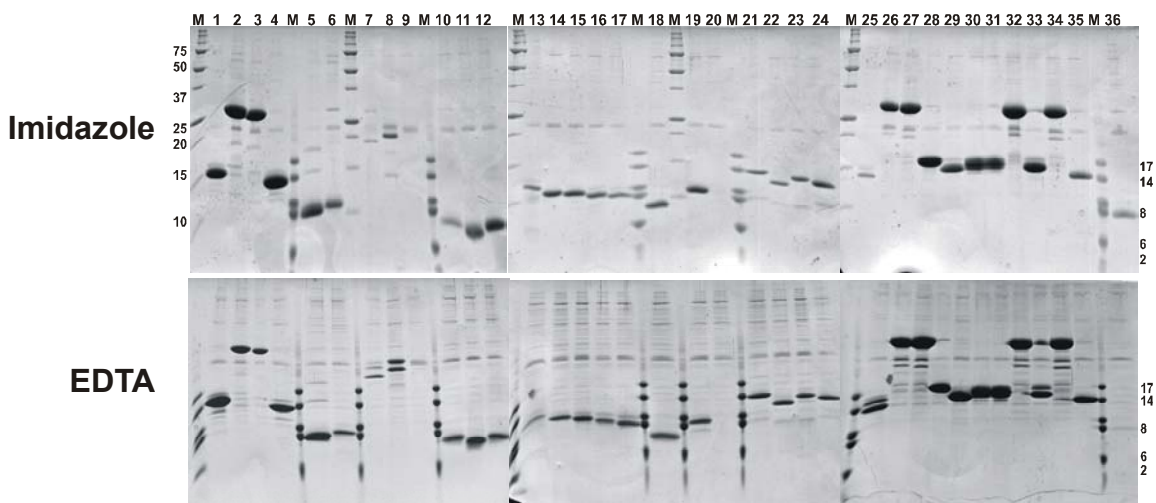
Relative amounts of purified protein were judged from SDS-PAGE analysis. The numbers in the columns start and end residues identify the combinations of constructs tested. Bold numbers indicate that these constructs yielded more soluble expression for the respective domain than the other constructs (see Results). '-' = no expression, '+/-' = low expression, '+' = intermediate expression, '++' = strong expression.

### 5.3 Detection of natively folded protein by 1D <sup>1</sup>H-NMR and analytical HIC

The 36 expression construct with reasonable yields of purified protein were arrayed in duplicate fashion on a new microplate. Of this plate, two 96-well culture plates for 1 ml cultures were grown in parallel and one was purified for analysis with 1D <sup>1</sup>H-NMR and

## Results II: Analysis of human protein domains

the other for analysis with analytical HIC. Prior to these analyses, a small aliquot of the 70  $\mu$ l-eluate was withdrawn (10  $\mu$ l) and analysed with SDS-PAGE. Due to the perturbing effect of imidazole on NMR signals, the elution of the purified His-tagged proteins for analysis with 1D  $^1$ H-NMR was performed with EDTA instead of imidazole. The purity of His-tagged proteins eluted with imidazole was higher than of those eluted with EDTA (Figure 5.1), particularly when yields of His-tagged proteins were low. This may be attributed to a higher proportion of unspecifically bound protein being eluted with EDTA. Expression clones 9 and 20 showed almost no yield of purified proteins in this experiment. Protein yields vary significantly within this batch of different expression constructs.



**Figure 5.1: Automatically purified human protein domains**

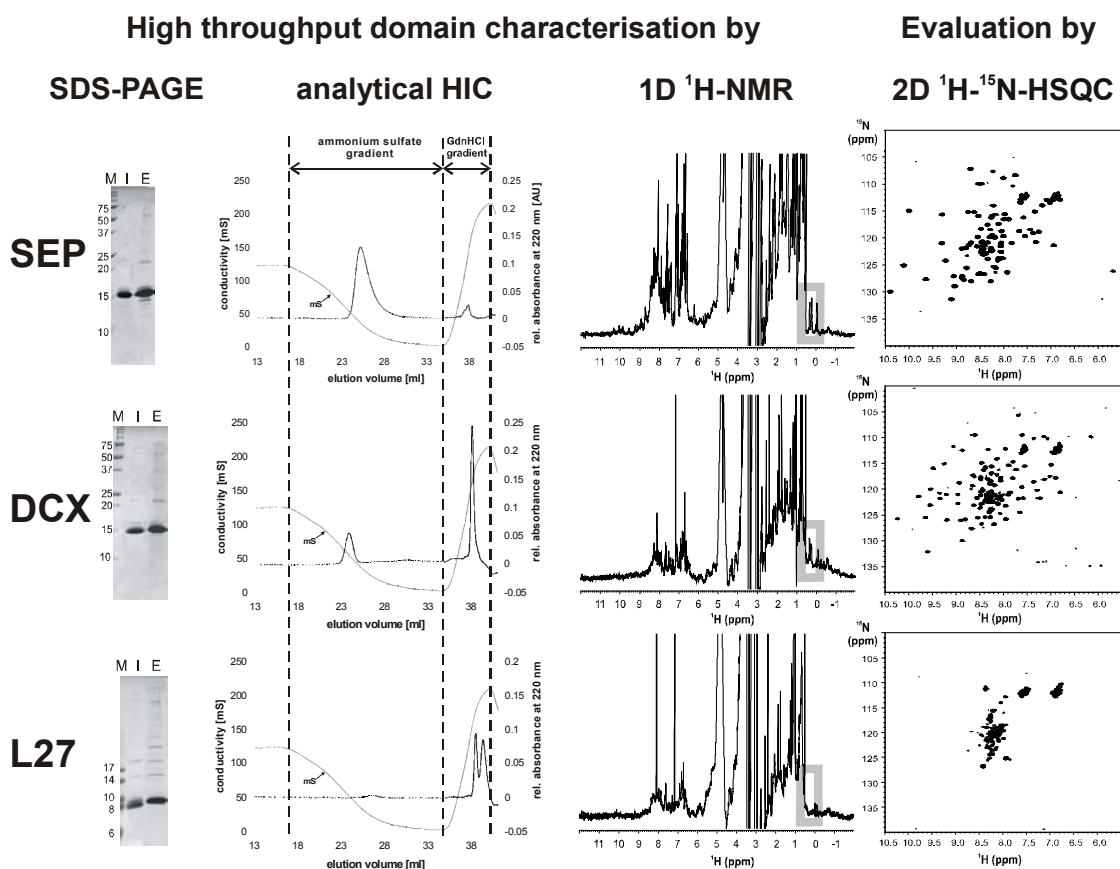
M: high molecular weight marker (indicated on the right side, kDa) or low molecular weight marker (indicated on the left side, kDa), numbers above correspond to domain expression clones listed in Table 5.2, page 67. In the upper panel His-tagged protein bound to Ni-NTA was eluted with imidazole (for analytical HIC), in the lower panel elution was performed with EDTA (for 1D  $^1$ H-NMR spectroscopy)

In collaboration with Anne Diehl and Dietmar Leitner (PSF) it was tested, whether the amount of purified protein resulting from 1 ml cultures is sufficient to be reliably analysed for natively folded protein using 1D  $^1$ H-NMR. NMR is a suitable method for evaluating the structural integrity of a protein either by hetero- or homonuclear techniques (Gronenborn & Clore, 1996; Rehm et al., 2002; Woestenenk et al., 2003). Especially in the context of structural genomics projects acquisition of 2D  $^1$ H- $^{15}$ N-HSQC spectra is widely used despite the disadvantage of having to employ  $^{15}$ N-minimal medium, which slows down the production process significantly. Alternatively, labelled, commercially

## Results II: Analysis of human protein domains

available rich media can be used for protein production, but these media are comparably expensive.

The fact that a 1D  $^1\text{H-NMR}$  spectrum contains enough information to evaluate if a protein is folded was exploited in the described analysis procedure. The advantage is that high-throughput protein expression and purification procedures based on unlabeled rich media can be used. Two 70  $\mu\text{l}$  EDTA eluates from two 1 ml cultures of the same construct that had been purified using automated protein purification were pooled (see Figure 5.1). From these samples, 1D  $^1\text{H-NMR}$  spectra were acquired.



**Figure 5.2: High-throughput domain characterisation and evaluation by 2D  $^1\text{H-}^{15}\text{N-HSQC}$  spectroscopy.**

Analytical results for the SEP-domain (residues 171 - 270 from protein AAH02801.1), the N-terminal DCX domain from protein AAC31696.1 (residues 90 - 186) and the L27-domain from protein AAG34117.1 (residues 13 - 68) are shown. Samples for SDS-PAGE, analytical HIC and 1D  $^1\text{H-NMR}$  spectra were obtained from two separate automated small-scale purifications of 1 ml cultures. 2D  $^1\text{H-}^{15}\text{N-HSQC}$  spectra resulted from the purification of a 400 ml culture. SDS-PAGE lanes: M: marker; I: Ni-NTA-bound protein eluted with imidazole for analytical HIC; E: Ni-NTA-bound protein eluted with EDTA for 1D  $^1\text{H-NMR}$  analysis.

## Results II: Analysis of human protein domains

In Figure 5.2, three selected examples from Table 5.1 are shown: (i) the SEP domain (residues 171-270 from protein AAH02801.1), (ii) the N-terminal DCX domain from protein AAC31696.1 (residues 90–186) and (iii) the L27-domain from protein AAG34117.1 (residues 13–68). The chemical shift dispersion of the proton resonance signals is indicative of protein folding. Large chemical shift dispersions, especially in the methyl group region between 1.0 and -1.0 ppm and the amide region downfield of 8.5 ppm (i.e. at higher ppm values) can be used as an indicator for protein folding. However, the resonance signals of the amide protons occurring in this region can be broadened due to chemical exchange especially at non acidic pH. This can be a drawback when amide signals at protein concentrations close to the detection limit are to be detected.

Isolated methyl group signals below 0.5 ppm are further sensitive and reliable probes for folded protein (see 3.5.10). They are not affected by chemical exchange broadening. In addition their peak intensities are higher because three equivalent protons contribute rather than one proton of the amide group. The SEP and DCX domains showed a number of resonance signals below 0.5 ppm (highlighted in the grey box in Figure 5.2) whereas L27 has no signals below 0.5 ppm except of two signals that resulted from the buffer (spectrum not shown). In case of SEP a number of amide resonance signals were observed above 8.5 ppm indicating folded protein, whereas for DCX and L27 no signals occurred. In order to confirm these results SEP, DCX and L27 were expressed with  $^{15}\text{N}$  labels in 400 ml scale in Anne Diehls' laboratory. The domains were purified and His-tags were cleaved using TEV protease. Subsequently, 1D  $^1\text{H}$ -NMR spectra and 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra were acquired. In the methyl-group region, 1D  $^1\text{H}$ -NMR spectra from the small-scale approach exhibited the same characteristics as those obtained from the  $^{15}\text{N}$ -labelled samples (data not shown). 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra (Figure 5.2, right side) clearly support the findings from the 1D spectra recorded with unlabelled tagged protein domains. The peaks of SEP and DCX show a large dispersion along both dimensions indicating folded protein. On the contrary, the resonance signals of L27 cluster around 8.3 ppm suggesting an unfolded or only partly folded protein. The absence of amide signals downfield of 8.5 ppm for the DCX domain might be due to the fact that the intensity of single amide signals fell below the detection limit under the applied conditions. An acquired 1D  $^1\text{H}$ -NMR spectrum from the more concentrated  $^{15}\text{N}$ -sample showed indeed amide resonances downfield of 8.5 ppm (data not shown).

## Results II: Analysis of human protein domains

| Corresponding lane number in Figure 5.1 | Domain   | start residue | end residue | <sup>1</sup> H 1D-NMR | analytical HIC |
|---|----------|---------------|-------------|-----------------------|----------------|
| 4                                       | SEP      | 171           | 270         | +                     | +              |
| 1                                       | SEP      | 176           | 270         | +                     | +              |
| 31                                      | DCX      | 90            | 181         | +                     | +/-            |
| 28                                      | DCX      | 90            | 186         | +                     | +              |
| 30                                      | DCX      | 85            | 181         | +                     | +              |
| 33                                      | DCX      | 85            | 186         | +                     | +              |
| 29                                      | DCX      | 212           | 310         | -                     | -              |
| 35                                      | DCX      | 217           | 310         | n. d.                 | -              |
| 25                                      | DCX      | 217           | 305         | n. d.                 | -              |
| 32                                      | DCX      | 85            | 305         | +/-                   | -              |
| 34                                      | DCX      | 90            | 305         | +/-                   | -              |
| 27                                      | DCX      | 90            | 310         | -                     | -              |
| 26                                      | DCX      | 85            | 310         | -                     | -              |
| 2                                       | SEP-UBX  | 171           | 368         | +                     | +              |
| 3                                       | SEP-UBX  | 176           | 368         | -                     | +/-            |
| 21                                      | PINT     | 316           | 409         | n. d.                 | -              |
| 22                                      | PINT     | 321           | 404         | n. d.                 | -              |
| 23                                      | PINT     | 316           | 404         | n. d.                 | -              |
| 24                                      | PINT     | 321           | 409         | n. d.                 | -              |
| 13                                      | G_patch  | 406           | 464         | n. d.                 | -              |
| 20*                                     | G_patch  | 228           | 284         | n. d.                 | -              |
| 19                                      | G_patch  | 233           | 284         | n. d.                 | +/-            |
| 16                                      | G_patch  | 91            | 141         | n. d.                 | -              |
| 14                                      | G_patch  | 91            | 146         | -                     | -              |
| 17                                      | G_patch  | 96            | 141         | -                     | -              |
| 15                                      | G_patch  | 96            | 146         | -                     | +/-            |
| 36                                      | G_patch  | 736           | 787         | n. d.                 | -              |
| 7                                       | Zpr1     | 252           | 416         | n. d.                 | -              |
| 8                                       | Zpr1     | 257           | 416         | n. d.                 | -              |
| 9                                       | Zpr1     | 44            | 212         | n. d.                 | -              |
| 10                                      | ZnF_C3H1 | 131           | 157         | n. d.                 | -              |
| 11                                      | ZnF_C3H1 | 126           | 162         | n. d.                 | -              |
| 12                                      | ZnF_C3H1 | 131           | 162         | -                     | -              |
| 18                                      | ZnF_RBZ  | 185           | 214         | n. d.                 | -              |
| 5                                       | L27      | 13            | 68          | -                     | -              |
| 6                                       | L27      | 13            | 73          | -                     | -              |

**Table 5.2: Analysis of soluble constructs with 1D <sup>1</sup>H-NMR and analytical HIC.**

For column 1D <sup>1</sup>H-NMR: '-' = unfolded, '+/-' = presumably folded, '+' = folded, 'n. d.' = not detectable (protein amount insufficient for detection). For column analytical HIC: '-' = no peak upon ammonium sulfate gradient, but protein elutes only during GdnHCl gradient, '+/-' = very small or broad peak observed upon ammonium sulfate gradient, '+' = distinct peak observed upon ammonium sulfate gradient. \* sample 20 showed no peak in HIC, which corresponds to the observation that there was no protein purified (see Figure 5.1)

In summary, when analysing purified samples from small-scale cultures, NMR-signals in the methyl-group region seem to be a reliable indicator for folded protein. In addition, the

## Results II: Analysis of human protein domains

---

amide proton signals can provide useful information but are not as strong as proton signals from methyl-groups. Because most protein samples analysed in this work were rather dilute (see below), resonance signals from amid protons might in most cases not be detectable. The occurrence of proton signals in one of the two chemical shift regions is sufficient to detect folded protein.

The results of all screened protein domains are summarised in Table 5.2. Out of 36 tested protein samples, only 19 contained sufficient protein to observe  $^1\text{H}$  signals within 1 h and 20 min. For nine samples, folded proteins were detected. A '+' for folded protein was assigned whenever NMR signals could be detected either in the amide and/or methyl-group region within the chemical shift ranges described above. Two domain constructs were annotated with '+/-' because  $^1\text{H}$  signals occurred in the respective regions but with very low intensities. For 10 samples protein  $^1\text{H}$  signals are present but neither amide signals downfield of 8.3 ppm nor methyl-group resonances upfield of 0.5 ppm could be observed (classified with '-'). It was found by Anne Diehl and Dietmar Leitner that the minimal protein concentration to get an evaluation by NMR in 1 h and 20 min on a 600 MHz spectrometer (equipped with a 5 mM triple resonance probe head) is approx. 3  $\mu\text{M}$  (e.g. 15  $\mu\text{g}$  of a 10 kDa protein in 500  $\mu\text{l}$ ). This value corresponds to an expression level of about 15 mg recombinant protein per litre of culture. This level was not achieved for almost half of the analysed samples (entries 'n.d.' in Table 5.2). Hence 1D  $^1\text{H}$ -NMR is in a lot of cases not sensitive enough to detect folded protein from 1 ml *E. coli*-cultures. This may be overcome by using 5 ml cultures in 24 well format.

Having shown the effectiveness of analytical HIC to judge the outcome of a folding screen (see 4.2), it was attempted to apply this method to the analysis of the investigated domain constructs. The data from analytical HIC and 1D  $^1\text{H}$ -NMR were compared. In a separate experiment ammonium sulfate was added to samples obtained from the automated folding screen to a final concentration of 1 M. Protein domains were bound to a HIC-resin and eluted in a decreasing ammonium sulfate gradient (see Figure 5.2). It was assumed that compact, natively folded proteins elute in a distinct peak. Some constructs of the SEP- and the DCX-domain behaved in this manner. These were the constructs that yielded natively folded protein shown by NMR analysis (see above and Figure 5.2). The detection limit for folded proteins was in the range of 5  $\mu\text{g}$  protein. To elute tightly bound protein and to regenerate the column, a gradient from 0 to 5 M GdnHCl was performed. Protein eluted during this step was presumed to be unfolded or only partially folded, particularly when it was only eluted with high concentrations of

## Results II: Analysis of human protein domains

GdnHCl. This presumption proved to be correct in case of L27 (Figure 5.2, bottom) or the C-terminal DCX domain (Table 5.2) which were shown to be only partly folded or unfolded by NMR (see above). In case of the N-terminal DCX domain shown in Figure 5.2, that showed a peak of folded protein during the ammonium sulfate gradient, an additional fraction of protein was eluted during the GdnHCl-gradient. This could be due to two species present upon expression: a compactly folded one and a misfolded one that binds tightly to the HIC resin and is only eluted with high concentrations of GdnHCl. However, an evaluation of the signals obtained from 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC did not support this assumption. If two species were present, such a heterogeneity may be detected in 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC because additional signals to the expected number of signals calculated from amino acid sequence should be observed. Another explanation for the GdnHCl-eluted protein would be that a fraction of the DCX domain was unfolded upon addition of ammonium sulfate or upon binding to the HIC resin.

Using the described high-throughput analysis procedure (1D  $^1\text{H}$ -NMR and/or analytical HIC) a priority list of constructs suitable for structural analysis could be generated (Table 5.2). The SEP domain and the N-terminal DCX domain of doublecortin were selected for structural analysis. The structure of the SEP-domain was solved within the PSF using NMR spectroscopy (PDB ID 1ss6). However, the NMR-structure of the N-terminal DCX domain was solved recently by Kim et al. (Kim et al., 2003). In this study the authors noted that the C-terminal DCX domain of doublecortin expressed in *E. coli* is thermodynamically unstable with a pronounced surface hydrophobicity as assayed by ANS (8-anilino-1-naphthalene-sulphonic acid) binding. This is also reflected in the observation with analytical HIC as this domain could only be eluted with high concentrations of GdnHCl. Correspondingly, 1D  $^1\text{H}$ -NMR suggested that the C-terminal DCX domain of doublecortin was not folded (Table 5.2). Kim et al. recorded a  $^1\text{H}$ - $^{15}\text{N}$ -HSQC and observed significant line broadening which presumably is due to conformational disorder on the millisecond timescale and hampered structure solution (Kim et al., 2003).

The SEP-UBX construct covering amino acids 171 – 368 was found to be folded according to both analytical methods, 1D  $^1\text{H}$ -NMR and analytical HIC. Accordingly, large scale preparation of this construct (Frank Götz and Dinh-Thrung Pham, PSF) and biophysical analysis including dynamic light scattering, differential scanning calorimetry and Fourier-transform infrared spectroscopy (Frank Niesen, PSF) indicated that this protein is compact and folded (data not shown).

## Results II: Analysis of human protein domains

---

In most cases, the data from 1D  $^1\text{H}$ -NMR and analytical HIC provided complementary results (Table 5.2). Only minor deviations could be observed. For example, the G\_patch-domains that were presumably folded according to analytical HIC (the observed peak during the ammonium sulfate gradient was very small). A 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum showed, however, that the chemical shift dispersion of the peaks is small suggesting an unfolded species present. Purifying this G\_patch domain with protein material from a 200 ml culture using HIC showed that the observed peak resulted from a non-protein molecule with higher absorbance at 260 nm than at 280 nm and gave rise to a false positive signal (data not shown). During small scale screening, absorbance at 220 nm and 280 nm were monitored. 220 nm was used because some of the analysed domains do not contain any tryptophans and/or only few tyrosines and therefore have relatively low extinction coefficients at 280 nm. The chromatography system used can only detect absorption at two wavelengths in parallel.

---

## 6 Discussion

### 6.1 A new automated protein folding screen

A large proportion of globular eukaryotic proteins can only be expressed in *E. coli* in inclusion bodies. The folding of proteins that are expressed insolubly in *E. coli* cells is a major obstacle in proteomics projects as well as in general biochemical research and in industrial processes that involve production of proteins as pharmaceuticals. Screening for suitable conditions to fold proteins can be very labor-intensive. To facilitate this process, an automated folding screen was implemented in this work and its efficiency was proved by refolding MDH and CAB. The procedure is flexible and the screen can easily be modified by reprogramming the robot or by exchanging stock solutions, e.g. for altering the pH-values. The screen facilitates folding of proteins by rapid dilution as well as folding of proteins that are immobilised on affinity resins. For the complete screen of 30 conditions only four milligram pure, denaturant-solubilised protein is needed. This is the first published folding screen using a robot system.

#### 6.1.1 Design of a folding screen

During *in vitro* protein folding, yields are affected by multiple factors. A screen where all these factors are checked in all possible combinations will result in a very high assay number. However, in a primary screen some factor combinations may be excluded. This exclusion can be done on the basis of known properties of particular factors, with respect to their effects on proteins. For example, disulfide bond formation is favoured at basic pH (Misawa & Kumagai, 1999)(see also 1.3). Hence, the combination of an acidic pH with a redox pair is unlikely to be beneficial. In contrast, other additives proved to be beneficial in conjunction, like PEG with urea or GdnHCl (Cleland et al., 1992a; Cleland et al., 1992b). Such assumptions and observations have been considered when composing the folding screen presented in this work (Table 4.1).

In this work, a screening procedure in which only one or two parameters at a time is changed was chosen. Most of the other screens that are described in literature consist of changes of many parameters at once. This approach is commonly known as fractional factorial screen (Armstrong et al., 1999). With such a screen a high number of factors

---

(additives, pH or protein concentration) can be assayed with a low number of screening buffers. For fractional factorial screening, a functional assay or a reliable method for the quantification of correctly folded protein (e.g. the peak heights resulting from gel filtration experiments) is necessary for the calculation of response values for the screened factors. The response values are used to determine whether a particular factor is beneficial for the folding of a certain protein. However, a functional assay or one stringent quantification method may often not be available. After a folding screen, protein concentrations are often too low for gel filtration experiments. In this work, it was demonstrated that a reliable evaluation of a folding screen can sometimes only be accomplished by monitoring more than one (biophysical) parameter. Such a multi-data analysis procedure seems to be difficult to apply to the evaluation strategy used for fractional factorial screening. Another disadvantage of a fractional factorial screen is that interferences between the different additives may distort the factor calculation. Armstrong et al. (1999) refolded CAB with a fractional factorial screen and found a negative effect of PEG 3350, even though the beneficial effect of PEGs in the molecular weight range of 600 to 8000 was shown before (Cleland et al., 1992b) and was confirmed in this work for PEG 1000. The authors attributed this observation to multi-factor interactions (Armstrong et al., 1999).

Two denaturant unfolded enzymes (MDH and CAB) as well as one denaturant solubilised inclusion body protein, p22 dynactin, were (re)folded in this work. The determined additives that promoted folding of those proteins using the developed screen (Table 4.1, page 39) are listed in Table 6.1. For each protein different additives were determined. This indicates that – within the screen used in this work - there may not be a “superior” additive, which can be applied to fold many proteins. An additive may be helpful for folding one protein but have a negative effect for another protein. For example, glycerol promoted folding of both MDH and p22 dynactin while it had a detrimental effect on folding of CAB: condition one (without glycerol) showed 26% recovered CAB-activity whereas the same buffer enriched with glycerol (condition 7) yielded 2% recovered activity (see Figure 4.5, Table 4.1)

| protein      | additives that promote (re)folding  |
|--------------|---|
| MDH          | glycerol is essential for refolding, together with glycerol an acidic pH, the presence of GSH/GSSG or EDTA enhanced refolding |
| CAB          | urea or PEG 1000, the combination of both has an additive effect  |
| p22 dynactin | glycerol, arginine, CTAB/cyclodextrin   |

**Table 6.1:** Determined (re)folding conditions for three proteins using the described automated folding screen (4.1). MDH and CAB were refolded after GdnHCl-unfolding and p22 was folded after GdnHCl-solubilisation of inclusion bodies.

### 6.1.2 A combination of biophysical monitors for folding detection upon folding screening

A major issue of this work is the problem of reliable and rapid detection of small amounts of natively folded protein from a folding screen when no biological assay is available. This problem is particularly important for proteomics projects that deal with large numbers of diverse proteins, but may also be an issue in various other fields where a functional test to detect native protein is not available or very time- and/or product consuming. Methods that can be generally applied need to be implemented. The efficiency of those methods with respect to sensitivity, adaptability to high-throughput processing and reliability have to be evaluated. Commonly used methods are analytical gel filtration, limited proteolysis and CD spectroscopy, but these methods have drawbacks (see 1.4) that make them unfavourable to be used in context of the evaluation of protein folding screening.

This work shows for two proteins analysed that can be successfully (re)folded, CAB and p22 dynactin, that tryptophan fluorescence spectroscopy can be helpful when evaluating the outcome of a folding screen. It can be performed in 96-well format making it suitable for the analysis of large numbers of samples. From the fluorescence spectra it was examined: (i) the blue-shift of the fluorescence intensity maximum compared to the denaturant-unfolded protein, and (ii) the fluorescence intensity. As explained in the introduction (section 1.4), fluorescence intensities should only be used in conjunction with other monitors. False negatives did not occur, as all folding conditions that contained folded protein, i.e. that showed CAB activity or, in the case of p22 dynactin, gave a peak in analytical HIC, were indeed detected. Detailed biophysical

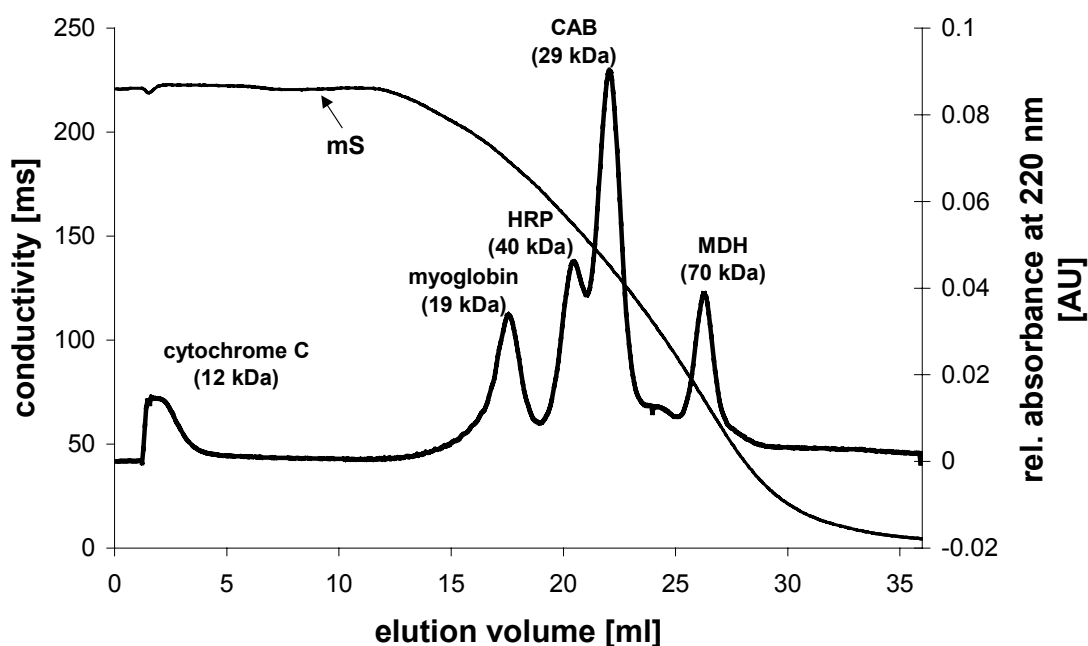
characterization showed that the p22 dynactin peak during HIC in a large-scale preparation represented folded protein, presumably in the native conformation.

However, when detecting folded p22 dynactin with fluorescence spectroscopy, some samples with high intensities as well as significant blue-shifts were observed, even though there was no peak in HIC. Blue-shifts were also observed for other proteins that were attempted to be folded, e.g. for ubiquitin fusion-degradation 1 like protein, although those samples mostly contained aggregated protein (see Figure 4.12). Misfolded microaggregates formed after a folding experiment may be structurally very compact thereby mimicking the fluorescence properties of native protein and giving rise to false positives. Another limitation of tryptophan fluorescence spectroscopy is that some proteins can not be analysed because they do not contain tryptophan residues or the tryptophan fluorescence has a low quantum yield. Therefore an alternative method to fluorescence spectroscopy needed to be employed.

In a new approach HIC was shown to be a powerful tool to judge the outcome of a folding screen. Adding an appropriate amount of ammonium sulfate and performing HIC can be used to separate misfolded from natively folded protein as shown for MDH, CAB and p22 dynactin. Using automated chromatography, HIC can easily be applied to large numbers of samples. In almost all cases, natively folded proteins showed a distinct peak in HIC due to their conformational uniformity. Misfolded or aggregated protein might (i) precipitate upon the addition of ammonium sulfate at lower concentrations, (ii) bind irreversibly to the column material or (iii) be eluted erratically from the column and not in a distinct peak due to conformational heterogeneity. Misfolded or partially folded protein exposes more hydrophobic patches than native protein. Aggregates were described to form *via* interaction of hydrophobic patches between folding intermediates (Fink, 1998). Hence large hydrophobic patches will be buried in the interior of the particle upon aggregation. Therefore, aggregated species may have an average surface hydrophobicity (surface hydrophobicity related to surface area) similar to that of native protein. This raises the question of how aggregates are separated from native protein during HIC. It was proposed that the driving force for binding to a HIC resin may be attributed to an increase in entropy that results from displacement of water from hydrophobic patches (Wheelwright, 1991). Due to its size, an aggregate will provide more hydrophobic binding sites than a small native molecule even when its average surface hydrophobicity is the same. Hence the entropic effect will be larger for a large particle. This may explain the separation of aggregates from native molecules and is

also supported by the observation that large globular proteins usually bind stronger to a HIC resin than smaller ones (see Figure 6.1, (Fausnaugh et al., 1984)).

Using analytical HIC, the yield of natively folded protein can be estimated quite reliably by simply comparing peak heights. Less than 10  $\mu\text{g}$  of natively folded protein could be readily detected for the analysed proteins. The analytical column may easily be regenerated by washing with a GdnHCl-containing buffer. Interestingly, HIC could detect the binding of the metal cofactor zinc to CAB. Although one would expect only a small effect on the structure of CAB caused by the binding of zinc, Zn-CAB elutes earlier from the HIC column and is presumably a less hydrophobic molecule.



**Figure 6.1: Separation of various test proteins of different size on the used analytical HIC column**

The type of HIC resin needs some consideration. For small globular proteins, i.e. up to 50 kDa, a resin with relatively high hydrophobicity should be used. In this work such a strongly hydrophobic resin was used, namely a POROS high-density phenyl resin (POROS HP2, Applied Biosystems). As outlined above, smaller proteins need in most cases a higher concentration of ammonium sulfate to bind to the column (see Figure 6.1, (Fausnaugh et al., 1984)). However, if the concentration is too high the target protein

---

may precipitate. To find an optimum concentration, it may be necessary to test two concentrations: first a low ammonium sulfate concentration (e.g. 1.5 M). If no peaks are detected with aliquots tested in these runs, more ammonium sulfate can be added and HIC is repeated. It should be noted that analytical HIC may not in all cases be the adequate method. Some natively folded proteins may precipitate even at ammonium sulfate concentrations of less than 1 M; additionally, the exposure to a strongly hydrophobic resin may cause unfolding of some proteins.

Ideally, a detection method for folded protein upon folding screening should lead to neither false negatives nor false positives. False negatives are highly detrimental because they lead to a complete loss of a positive sample and should therefore be avoided by all means. False positives can always be excluded later by a more detailed investigation. With the folding screen analysis of CAB and p22 dynactin it was demonstrated how two biophysical folding monitors, analytical HIC and tryptophan fluorescence spectroscopy, can be used in conjunction to exclude false positives or negatives obtained with one of the respective monitors. In case of CAB, one condition (# 17, see Figure 4.5) resulted in no enzymatic activity but a peak during HIC as well as a blue-shift of the fluorescence intensity maximum was observed. As small structural changes may result in complete loss of activity, it can be assumed that the refolded product had a slightly different structure to native CAB.

For the model enzyme CAB it was attempted to analyse a folding screen with ANS fluorescence. However, this approach appeared to be prone to high occurrence of false positives (see 4.2.4). At least for refolding screening with CAB, ANS fluorescence did not serve as a useful monitor to evaluate protein folding. The reason for this finding could be that (i) the overall fluorescence signal intensity in the protein concentration range where refolding is commonly performed is too low to allow for a comprehensive analysis, or (ii) ANS does not bind to misfolded soluble microaggregates.

In summary, the automation and miniaturisation of a protein folding screen coupled to a biophysical detection system consisting of tryptophan fluorescence spectroscopy and analytical HIC is described in this thesis. The outcome is an effective approach to quickly obtain natively folded protein from *E. coli* inclusion bodies when no functional assay is available as a folding monitor. For three proteins investigated, CAB, MDH and p22 dynactin productive conditions for (re-)folding were identified. The folding of p22 dynactin was successfully scaled up to yield a sufficient amount of protein for detailed analysis and crystallisation trials. The presented approach may help to overcome well-known

---

obstacles encountered during protein folding and facilitate the availability of inclusion body proteins for proteomics.

### 6.1.3 Possible reasons for unsuccessful folding screens of proteins

Except for p22 dynactin, for five proteins listed in Table 4.3 (page 59) that were submitted to the described folding screen, no folded protein could be detected using HIC (no peak in a decreasing ammonium sulfate gradient). In some cases tryptophan fluorescence spectra of proteins folded under particular conditions exhibited blue-shifts compared to the spectra of urea-denatured proteins. However, large-scale refolding under these conditions never resulted in preparations of natively folded protein. Proteins either (i) were lost during concentration or (ii) were aggregated or aggregated during concentration or (iii) could not be efficiently purified by ion exchange chromatography, probably due to inhomogeneity of the sample. Three of those five proteins contained single point mutations (see Table 4.3). Those mutations may hamper folding to a compact native species. In some cases the tested proteins might have been folded to the native form under certain conditions, but to such low quantities that the product could not be detected with the methods deployed. Different refolding conditions would need to be tested in those cases to accomplish a high yield that is sufficient for structural analysis. For some proteins, the presence of certain chaperones may be a prerequisite for folding. Some proteins may only be obtained (and maintained) in the native conformation *in vitro* at high concentrations when stabilising ligands are present that bind to the proteins *in vivo*. The only such “*in vivo*-cofactors” used in the presented screening procedure were magnesium and calcium ions. Furthermore, if posttranslational modifications, e.g. glycosylation, are essential for the stability of a particular eukaryotic protein, successful production will only be feasible by expression in higher organisms (e.g. yeast or insect cells). Proteins found in complexes with others *in vivo* may be significantly more stable within these complexes than without contact to their interaction partners. For structural analysis, these proteins would need to be co-expressed and prepared together with their respective binding partners.

---

#### **6.1.4 Alternative *in vivo*-approaches to obtain aggregation-prone protein in a soluble form**

All proteins analysed in the PSF are expressed at two temperatures, 37 °C and 28 °C. In some cases the lower temperature leads to an increase in soluble protein. This may be attributed to the lower mRNA translation rate which gives synthesised peptide chains more time to fold correctly. Such a beneficial effect has been reported previously (Han et al., 1999). Proteins that were submitted to the automated folding screen in this work could not be expressed solubly even at lower temperatures than 28 °C (see Figure 4.7). An alternative approach that may help to obtain proteins in a soluble form is the expression with N-terminal fusion partners such as maltose binding protein (Fox et al., 2003; Kapust & Waugh, 1999), NusA (Davis et al., 1999) or thioredoxin (Hammarstrom et al., 2002). These proteins seem to function as an “artificial” chaperone. However, it has been observed eventually that the solubly expressed target proteins fused to those proteins were inactive and not folded (Louis et al., 1999; Nomine et al., 2001; Saavedra-Alanis et al., 1994; Sachdev & Chirgwin, 1998; Sachdev & Chirgwin, 1999). Coexpression of chaperones as well as the addition of ethanol to the culture medium to upregulate chaperone expression has also been reported to enhance yields of soluble native expression in *E. coli* in certain cases (Nygaard & Harlow, 2001; Thomas & Baneyx, 1996; Thomas & Baneyx, 1997; Vonrhein et al., 1999). The overall efficiency of these procedures to produce large sets of target proteins for structural analysis needs to be evaluated. They may, however, be valuable alternative/additional approaches to express “problematic” proteins in *E. coli*.

#### **6.2 Fast identification of expression constructs of human protein domains suitable for structural analysis**

Because protein domains usually represent compact folding entities they are well suitable for structural analysis (see 1.5). The identification of domains may be performed using limited proteolysis coupled to mass spectrometry. However, this method is time-consuming, frequently requiring that multiple proteolytic reactions be refined for optimal cleavage. Peptide loops may be cut, leading to destabilisation and further proteolytic degradation of what originally was a structured region. In this work, domains were identified *in silico* using sequence homology models (Letunic et al., 2002; Schultz et al.,

2000). Because the prediction of domains within proteins only cover regions that are conserved within the respective domain family, expression constructs with variations at the domain boundaries were generated (see 5.1 and Table 5.1). Such a set of expression constructs is first screened for soluble expression with reasonable protein yield in a small scale. Conventionally, positively checked samples are then prepared on a larger scale to produce enough protein to identify targets suitable for structural analysis. Methods that have been used to identify suitable protein samples comprise CD spectroscopy or 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (Woestenenk et al., 2003). For analysis with 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC  $^{15}\text{N}$ -minimal media need to be employed for cultivation.

This work shows that the amount of His-tagged protein purified from 1 ml cultures of *E. coli* expression clones with reasonable expression levels can be used to quickly identify samples that are suitable for structural analysis. A prerequisite for such an analysis system in the context of a structural genomics project is the presented, robust automated purification procedure to handle large sets of expression clones. The method is performed in 96-well format and protein yields between approx. 5  $\mu\text{g}$  and 90  $\mu\text{g}$  purified protein can be obtained, depending on the expression level (Scheich et al., 2003). Expression clones with reasonable expression yielded more than 10  $\mu\text{g}$  purified protein. To identify samples with folded protein among the set of tested expression constructs, 1D  $^1\text{H}$ -NMR was used. Particularly isolated methyl group signals below 0.5 ppm are sensitive and reliable probes for folded protein when such small amounts of proteins are analysed (see Figure 5.2). However, almost half of the 36 expressed domain constructs gave resonance signals under the detection limit of 1D  $^1\text{H}$ -NMR (approx. 3  $\mu\text{M}$  or 15  $\mu\text{g}$  in case of a 10 kDa domain), even though eluates of purified protein from two 1 ml cultures of the same construct were pooled. For 1D  $^1\text{H}$ -NMR analysis of those expression constructs larger culture volumes should be used.

Analytical HIC was tested as an alternative approach to 1D  $^1\text{H}$ -NMR. It was shown throughout this work that analytical HIC can distinguish compact natively folded from misfolded or aggregated protein upon folding screening (see 4.2.1, 4.2.3 and 4.3). Accordingly, it was found that this technique can also distinguish compactly folded domain constructs (that eluted in the decreasing ammonium sulfate gradient) from those that are misfolded, aggregated or contain large unstructured regions. The latter molecules could only be eluted when high concentrations of GdnHCl were applied to the analytical column (see Figure 5.2 and Table 5.2). Analytical HIC has a minor drawback: the optimal ammonium sulfate concentration that is high enough to bind the analysed

protein to the HIC resin but not too high to precipitate the protein is not known or may not exist within a batch of different proteins. In case of the set of human protein domains studied here, 1 M ammonium sulfate seemed to be adequate for all constructs. Adjusting the concentration to 2 M ammonium sulfate was accompanied with precipitation in almost all samples. In a general screening procedure it seems to be necessary to test two concentrations (e.g. 1 M and 2 M), similar to how it was proposed for the detection of natively folded protein after folding screening (see 6.2).

A similar approach to detect folding of proteins was used before (Woestenenk et al., 2003). In this work a  $^{15}\text{N}$  labelled 50 ml set up was used, followed by purification and removal of the affinity tag. 60- $\mu\text{M}$  samples were evaluated using 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC, CD spectroscopy and ANS fluorescence. It was shown that CD spectroscopy analysis can lead to false positives or negatives, whereas ANS fluorescence provided auxiliary information. However, ANS fluorescence can not be used as an exclusive monitor because the fluorescence intensity of the natively folded protein is not known. For this thesis, an unlabelled expression for up to 47 clones in parallel in a 1 ml scale, a single purification step and 1D  $^1\text{H}$ -NMR spectroscopy/analytical HIC was used. This analysis was performed in a 96-well plate and 47 samples (plus buffer as control) can be analysed because all samples are run in duplicates. This is necessary because for 1D  $^1\text{H}$ -NMR spectroscopy the duplicates are pooled and for analytical HIC two ammonium sulfate concentrations may need to be tested (see above). Alternatively, it may be better to work with 5-ml cultures in 24-well format. The outcome is a classification into folded or unfolded proteins leading to a decision which construct is most promising for structure determination (see Table 5.2). Hence, identification of suitable constructs for structural analysis is facilitated without time-consuming and expensive large-scale preparations. The data obtained from both analysis approaches, 1D  $^1\text{H}$ -NMR and analytical HIC, provided complementary results. HIC is the more cost-effective and more sensitive approach (the detection limit is approx. 5  $\mu\text{g}$  compared to 15  $\mu\text{g}$  for 1D  $^1\text{H}$ -NMR), but may be more error-prone because some purified domain constructs might precipitate on the highly hydrophobic resin although they were natively folded.

Altogether, only two out of seventeen chosen target domains were found suitable for NMR-structure analysis. This low success rate is typical for eukaryotic proteins/domains expressed in *E. coli*. In some cases a single domain may not be stable without being linked to its neighbouring domains. For domains that may be expressed

---

stably as individual units it is important to generate a large set of constructs for a particular domain with different boundaries as well as from different source proteins. Applying high-throughput protein production coupled to biophysical methods (1D  $^1\text{H}$ -NMR spectroscopy and/or analytical HIC) can be used to distinguish suitable from non-suitable proteins for detailed structural investigations. Domains predicted by SMART that were expressed insolubly may not represent complete folding entities, but miss weakly conserved residues at the C- or N-terminus. For more reliable estimation of domain boundaries it may be helpful to consider the prediction of secondary structure in the boundary region to avoid disruption of secondary structure elements.

### 6.3 Concluding remark

The *E. coli* expression system is the most widely used for producing recombinant proteins, although a high proportion of proteins are not expressed in the natively folded form. The analysis procedures presented in this work comprise possible solutions for making such proteins accessible to structural and functional analyses. The procedures are designed to quickly determine individually whether it is feasible to use *E. coli* expression. If the outcome of the analysis is negative, alternative expression systems like yeast or baculovirus have to be applied, although these are more expensive and time-consuming.

---

## 7 References

- Arakawa, T., Bhat, R., and Timasheff, S.N. 1990. Why preferential hydration does not always stabilize the native structure of globular proteins. *Biochemistry* **29**: 1924-1931.
- Armstrong, N., de Lencastre, A., and Gouaux, E. 1999. A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Sci* **8**: 1475-1483.
- Bodenreider, C., Kellershohn, N., Goldberg, M.E., and Mejean, A. 2002. Kinetic analysis of R67 dihydrofolate reductase folding: From the unfolded monomer to the native tetramer. *Biochemistry* **41**: 14988-14999.
- Bolen, D.W., and Baskakov, I.V. 2001. The osmophobic effect: Natural selection of a thermodynamic force in protein folding [Review]. *Journal of Molecular Biology* **310**: 955-963.
- Buchanan, S.G., Sauder, J.M., and Harris, T. 2002. The promise of structural genomics in the discovery of new antimicrobial agents. *Current Pharmaceutical Design* **8**: 1173-1188.
- Buchner, J., and Rudolph, R. 1991. Renaturation, purification and characterization of recombinant Fab-fragments produced in *Escherichia coli*. *Biotechnology (N Y)* **9**: 157-162.
- Büssow, K., Nordhoff, E., Lubbert, C., Lehrach, H., and Walter, G. 2000. A human cDNA library for high-throughput protein expression screening. *Genomics* **65**: 1-8.
- Chen, G.Q., and Gouaux, E. 1997. Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: application of a novel protein folding screen. *Proc Natl Acad Sci U S A* **94**: 13431-13436.
- Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C.M. 1999. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 3590-3594.
- Clark, E.D. 2001. Protein refolding for industrial processes. *Curr Opin Biotechnol* **12**: 202-207.

## References

---

- Cleland, J., and Randolph, T. 1992. Mechanism of polyethylene glycol interaction with the molten globule folding intermediate of bovine carbonic anhydrase B. *J Biol Chem* **267**: 3147-3153.
- Cleland, J.L., Builder, S.E., Swartz, J.R., Winkler, M., Chang, J.Y., and Wang, D.I. 1992a. Polyethylene glycol enhanced protein refolding. *Biotechnology (N Y)* **10**: 1013-1019.
- Cleland, J.L., Hedgepeth, C., and Wang, D.I. 1992b. Polyethylene glycol enhanced refolding of bovine carbonic anhydrase B. Reaction stoichiometry and refolding model. *J Biol Chem* **267**: 13327-13334.
- Cleland, J.L., and Wang, J.L. 1990. Refolding and aggregation of bovine carbonic anhydrase B: quasi-elastic light scattering analysis. *Biochemistry* **29**: 11072-11078.
- Collins, T., Stone, J.R., and Williams, A.J. 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains [Review]. *Molecular & Cellular Biology* **21**: 3609-3615.
- Creighton, T.E., Darby, N.J., and Kemmink, J. 1996. The roles of partly folded intermediates in protein folding [Review]. *FASEB Journal* **10**: 110-118.
- Das, B.K., and Liang, J.J. 1998. Thermodynamic and kinetic characterization of calf lens gammaF- crystallin. *Int J Biol Macromol* **23**: 191-197.
- Davis, G.D., Elisee, C., Newham, D.M., and Harrison, R.G. 1999. New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnology & Bioengineering* **65**: 382-388.
- Fausnaugh, J.L., Kennedy, L.A., and Regnier, F.E. 1984. Comparison of hydrophobic-interaction and reversed-phase chromatography of proteins. *J Chromatogr.* **317**: 141-155.
- Fink, A.L. 1995. Molten Globules. In *Methods in Molecular Biology*, B. A. Shirley), pp 343-360. Humana Press, Totowa, New Jersey.
- Fink, A.L. 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**: R9-23.
- Fischer, B., Sumner, I., and Goodenough, P. 1993. Renaturation of lysozyme-- temperature dependence of renaturation rate, renaturation yield, and aggregation: identification of hydrophobic folding intermediates. *Arch Biochem Biophys.* **306**: 183-187.

- Fox, J.D., Routzahn, K.M., Bucher, M.H., and Waugh, D.S. 2003. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Letters* **537**: 53-57.
- Gekko, K., and Timasheff, S.N. 1981. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry*. Aug 4;20(16):4667-76 **20**: 4667-4676.
- Gill, S.R., Schroer, T.A., Szilak, I., Steuer, E.R., Sheetz, M.P., and Cleveland, D.W. 1991. Dynactin, a conserved, ubiquitously expressed component of an activator of vesicle motility mediated by cytoplasmic dynein. *J Cell Biol* **115**: 1639-1650.
- Goldberg, M.E., Rudolph, R., and Jaenicke, R. 1991. A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme. *Biochemistry* **30**: 2790.
- Gronenborn, A.M., and Clore, G.M. 1996. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Science* **5**: 174-177.
- Guise, A.D., West, S.M., and Chaudhuri, J.B. 1996. Protein folding in vivo and renaturation of recombinant proteins from inclusion bodies. *Mol Biotechnol* **6**: 53-64.
- Hammarstrom, M., Hellgren, N., Van den Berg, S., Berglund, H., and Hard, T. 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Science* **11**: 313-321.
- Han, K.G., Lee, S.S., and Kang, C.W. 1999. Soluble expression of cloned phage K11 RNA polymerase gene in *Escherichia coli* at a low temperature. *Protein Expression & Purification* **16**: 103-108.
- Hanson, P.E., and Gellman, S.H. 1998. Mechanistic comparison of artificial-chaperone-assisted and unassisted refolding of urea-denatured carbonic anhydrase B. *Fold Des* **3**: 457-468.
- Hartl, F.U., and Hayer-Hartl, M. 2002. Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein [Review]. *Science* **295**: 1852-1858.
- Heinemann, U., Illing, G., and Oschkinat, H. 2001. High-throughput three-dimensional protein structure determination [Review]. *Current Opinion in Biotechnology* **12**: 348-354.

- Heiring, C., and Muller, Y.A. 2001. Folding screening assayed by proteolysis: application to various cystine deletion mutants of vascular endothelial growth factor. *Protein Eng* **14**: 183-188.
- Hutchinson, J.P., el-Thaher, T.S., and Miller, A.D. 1994. Refolding and recognition of mitochondrial malate dehydrogenase by *Escherichia coli* chaperonins cpn 60 (groEL) and cpn10 (groES). *Biochem J* **302**: 405-410.
- Inoue, H., Nojima, H., and Okayama, H. 1990. Abstract High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.
- Jackson, M., and Mantsch, H.H. 1995. The use and misuse of FTIR spectroscopy in the determination of protein structure [Review]. *Critical Reviews in Biochemistry & Molecular Biology* **30**: 95-120.
- Jackson, S.E., and Fersht, A.R. 1991. Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* **30**: 10428 - 10435.
- Jacob, M., Geeves, M., Holtermann, G., and Schmid, F.X. 1999. Diffusional barrier crossing in a two-state protein folding reaction. *Nature Structural Biology* **6**: 923-926.
- Jacob, M., Schindler, T., Balbach, J., and Schmid, F.X. 1997. Diffusion Control in an Elementary Protein Folding Reaction. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 5622-5627.
- Jaenicke, R. 1999. Stability and folding of domain proteins [Review]. *Progress in Biophysics & Molecular Biology* **71**: 155-241.
- Jaenicke, R., and Seckler, R. 1997. Protein misassembly in vitro. *Adv Protein Chem.* **50**: 1-59.
- John, B., D'Silva, P.R., and Lala, A.K. 2001. Analysis of protein folding using polarity-sensitive fluorescent probes. *Current Science* **80**: 287-290.
- Jones, S., and Thornton, J.M. 2004. Searching for functional sites in protein structures. *Curr Opin Chem Biol.* **8**: 3-7.
- Kapust, R.B., and Waugh, D.S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**: 1668-1674.

## References

- Karki, S., LaMonte, B., and Holzbaur, E.L. 1998. Characterization of the p22 subunit of dynactin reveals the localization of cytoplasmic dynein and dynactin to the midbody of dividing cells. *J Cell Biol* **142**: 1023-1034.
- Kaufmann, M. 1997. Unstable proteins - how to subject them to chromatographic separations for purification procedures [Review]. *Journal of Chromatography B* **699**: 347-369.
- Kay, L.E., Keifer, P., and Saarinen, P. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**: 10663-10665.
- Kiefhaber, T. 1995. Protein Folding Kinetics. In *Methods in Molecular Biology*, B. A. Shirley), pp 313-341. Humana Press, Totowa, New Jersey.
- Kim, M.H., Cierpicki, T., Derewenda, U., Krowarsch, D., Feng, Y.Y., Devedjiev, Y., Dauter, Z., Walsh, C.A., Otlewski, J., Bushweller, J.H., and Derewenda, Z.S. 2003. The DCX-domain tandems of doublecortin and doublecortin-like kinase. *Nature Structural Biology* **10**: 324-333.
- Kim, P.S., and Baldwin. 1990. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem.* **59**: 631-660.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Lankowicz, J.R. 1999. *Principles of Fluorescence Spectroscopy*. Kluwer Academic/Plenum Publisher, New York.
- Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Research* **30**: 242-244.
- Levine, R.L., and Federici, M.M. 1982. Quantitation of aromatic residues in proteins: model compounds for second-derivative spectroscopy. *Biochemistry* **21**: 2600-2606.
- Levinthal, C. 1968. Are there pathways in folding. *J Chim Phys* **65**: 44-45.
- Lilie, H., Schwarz, E., and Rudolph, R. 1998. Advances in refolding of proteins produced in *E. coli*. *Curr Opin Biotechnol* **9**: 497-501.

## References

- Louis, J.M., Wondrak, E.M., Kimmel, A.R., Wingfield, P.T., and Nashed, N.T. 1999. Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. *Journal of Biological Chemistry* **274**: 23437-23442.
- Lovell, S.C. 2003. Are non-functional, unfolded proteins ('Junk proteins') common in the genome? *FEBS Letters* **554**: 237-239.
- Mach, H., Middaugh, C.R., and Lewis, R.V. 1992. Detection of proteins and phenol in DNA samples with second-derivative absorption spectroscopy. *Anal Biochem* **200**: 20-26.
- Martin, A., and Schmid, F.X. 2003. The folding mechanism of a two-domain protein: Folding kinetics and domain docking of the gene-3 protein of phage fd. *Journal of Molecular Biology* **329**: 599-610.
- Mayr, E.M., Jaenicke, R., and Glockshuber, R. 1994. Domain interactions and connecting peptides in lens crystallins. *Journal of Molecular Biology* **235**: 84-88.
- McCoy, L.F., and Wong, K.P. 1981. Renaturation of bovine erythrocyte carbonic anhydrase B denatured by acid, heat, and detergent. *Biochemistry* **20**: 3062-3067.
- Mendoza, J.A., Rogers, E., Lorimer, G.H., and Horowitz, P.M. 1991. Unassisted refolding of urea unfolded rhodanese. *J Biol Chem.* **266**: 13587-13591.
- Middelberg, A.P. 2002. Preparative protein refolding. *Trends Biotechnol* **20**: 437-443.
- Misawa, S., and Kumagai, I. 1999. Refolding of therapeutic proteins produced in *Escherichia coli* as inclusion bodies. *Biopolymers* **51**: 297-307.
- Müller, R.H., and Schuhmann, R. 1996. *Teilchengrößenmessung in der Laborpraxis*. Wissenschaftliche Verlagsgesellschaft mbH.
- Nomine, Y., Ristriani, T., Laurent, C., Lefevre, J.F., Weiss, E., and Trave, G. 2001. Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein. *Protein Expression & Purification* **23**: 22-32.
- Nygaard, F.B., and Harlow, K.W. 2001. Heterologous expression of soluble, active proteins in *Escherichia coli*: The human estrogen receptor hormone-binding domain as paradigm. *Protein Expression & Purification* **21**: 500-509.
- Okajima, T., Tanabe, T., and Yasuda, T. 1993. Nonurea sodium dodecyl sulfate-polyacrylamide gel electrophoresis with high-molarity buffers for the separation of proteins and peptides. *Anal Biochem.* **211**: 293-300.

## References

- Pantazatos, D., Kim, J., Klock, H.E., Stevens, R.C., Wilson, I.A., Lesley, S.A., and Woods, V.L.J. 2004. Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc Natl Acad Sci U S A* **101**: 751-756.
- Pennisi, E. 2001. The human genome. *Science* **291**: 1177-1180.
- Peralta, D., Hartman, D.J., Hoogenraad, N.J., and Hoj, P.B. 1994. Generation of a stable folding intermediate which can be rescued by the chaperonins GroEL and GroES. *FEBS Lett.* **339**: 45-49.
- Piotto, M., Saudek, V., and Sklenar, V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2**: 661-665.
- Przybycien, T.M., Dunn, J.P., Valax, P., and Georgiou, G. 1994. Secondary structure characterization of beta-lactamase inclusion bodies. *Protein Engineering* **7**: 131-136.
- Rehm, B.H., Qi, Q., Beermann, B.B., Hinz, H.J., and Steinbuchel, A. 2001. Matrix-assisted in vitro refolding of *Pseudomonas aeruginosa* class II polyhydroxyalkanoate synthase from inclusion bodies produced in recombinant *Escherichia coli*. *Biochem J* **358**: 263-268.
- Rehm, T., Huber, R., and Holak, T.A. 2002. Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure* **10**: 1613-1618.
- Royer, C.A. 1995. Fluorescence Spectroscopy. In *Methods in Molecular Biology*, B. A. Shirley), pp 65 - 90. Humana Press, Totowa, New Jersey.
- Rozema, D., and Gellman, S.H. 1996. Artificial chaperone-assisted refolding of carbonic anhydrase B. *J Biol Chem* **271**: 3478-3487.
- Rudolph, R. 1990. *Renaturation of recombinant, disulfide-bonded proteins from inclusion bodies*. Walter de Gruyter, New York.
- Rudolph, R., and Lilie, H. 1996. In vitro folding of inclusion body proteins. *Faseb J* **10**: 49-56.
- Russo, A.T., Rosgen, J., and Bolen, D.W. 2003. Osmolyte effects on kinetics of FKBP12C22A folding coupled with prolyl isomerization. *Journal of Molecular Biology* **330**: 851-866.

## References

- Saavedra-Alanis, V.M., Rysavy, P., Rosenberg, L.E., and Kalousek, F. 1994. Rat liver mitochondrial processing peptidase. Both alpha- and beta-subunits are required for activity. *J Biol Chem.* **269**: 9284-9288.
- Sachdev, D., and Chirgwin, J.M. 1998. Order of Fusions Between Bacterial and Mammalian Proteins Can Determine Solubility in Escherichia Coli. *Biochemical & Biophysical Research Communications* **244**: 933-937.
- Sachdev, D., and Chirgwin, J.M. 1999. Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. *Journal of Protein Chemistry* **18**: 127-136.
- Scheich, C., Sievert, V., and Büssow, K. 2003. Comparison of His-tag and GST-tag affinity chromatography by automated protein purification. *BMC Biotechnol* **3**: 12.
- Schiene, C., and Fischer, G. 2000. Enzymes that catalyse the restructuring of proteins [Review]. *Current Opinion in Structural Biology* **10**: 40-45.
- Schindler, T., Herrler, M., Marahiel, M.A., and Schmid, F.X. 1995. Extremely Rapid Protein Folding in the Absence of Intermediates. *Nature Structural Biology* **2**: 663-673.
- Schmid, F.X. 1997. *Optical spectroscopy to characterise protein conformation and conformational changes*. University Press, New York, Oxford.
- Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* **28**: 231-234.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. Smart, a Simple Modular Architecture Research Tool - Identification of Signaling Domains. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 5857-5864.
- Sheehan, D. 2000. *Physical Biochemistry: Principles and Applications*. Wiley.
- Sosnick, T.R., Mayne, L., Hiller, R., and Englander, S.W. 1994. The barriers in protein folding. *Nat Struct Biol.* **1**: 149 - 156.
- Speed, M.A., Wang, D.I.C., and King, J. 1996. Specific Aggregation of Partially Folded Polypeptide Chains - the Molecular Basis of Inclusion Body Composition. *Nature Biotechnology* **14**: 1283-1287.

## References

- 
- Steinberg, T.H., Top, K.P.O., Berggren, K.N., Kemper, C., Jones, L., Diwu, Z.J., Haugland, R.P., and Patton, W.F. 2001. Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics* **1**: 841-855.
- Stempfer, G., Holl-Neugebauer, B., and Rudolph, R. 1996. Improved refolding of an immobilized fusion protein. *Nat Biotechnol* **14**: 329-334.
- Thomas, J.G., and Baneyx, F. 1996. Protein Misfolding and Inclusion Body Formation in Recombinant Escherichia Coli Cells Overexpressing Heat-Shock Proteins. *Journal of Biological Chemistry* **271**: 11141-11147.
- Thomas, J.G., and Baneyx, F. 1997. Divergent Effects of Chaperone Overexpression and Ethanol Supplementation On Inclusion Body Formation in Recombinant Escherichia Coli. *Protein Expression & Purification* **11**: 289-296.
- Tobbell, D.A., Middleton, B.J., Raines, S., Needham, M.R., Taylor, I.W., Beveridge, J.Y., and Abbott, W.M. 2002. Identification of in vitro folding conditions for procathepsin S and cathepsin S using fractional factorial screens. *Protein Expr Purif* **24**: 242-254.
- Uversky, V.N., and Fink, A.L. 2002. The chicken-egg scenario of protein folding revisited. *FEBS Letters* **515**: 79-83.
- Viitanen, P.V., Lubben, T.H., Reed, J., Goloubinoff, P., O'Keefe, D.P., and Lorimer, G.H. 1990. Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K<sup>+</sup> dependent. *Biochemistry* **29**: 5665-5671.
- Vonrhein, C., Schmidt, U., Ziegler, G.A., Schweiger, S., Hanukoglu, I., and Schulz, G.E. 1999. Chaperone-assisted expression of authentic bovine adrenodoxin reductase in Escherichia coli. *FEBS Letters* **443**: 167-169.
- Weber, F., Keppel, F., Georgopoulos, C., Hayerhartl, M.K., and Hartl, F.U. 1998. The Oligomeric Structure of Groel/Groes Is Required For Biologically Significant Chaperonin Function in Protein Folding. *Nature Structural Biology* **5**: 977-985.
- Wheelwright, S.M. 1991. *Protein purification: design and scale up of downstream processes*. Hanser.

## References

- 
- Woestenenk, E.A., Hammarström, M., Härd, T., and Berglund, H. 2003. Screening methods to determine biophysical properties of proteins in structural genomics. *Anal Biochem* **318**: 71-79.
- Wright, P.E., and Dyson, H.J. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**: 321-331.
- Zahn, R., von Schroetter, C., and Wuthrich, K. 1997. Human prion proteins expressed in *Escherichia coli* and purified by high-affinity column refolding. *FEBS Lett* **417**: 400-404.
- Zettlmeissl, G., Rudolph, R., and Jaenicke, R. 1979. Reconstitution of lactic dehydrogenase. Noncovalent aggregation vs. reactivation. 1. Physical properties and kinetics of aggregation. *Biochemistry* **18**: 5567-5571.
- Zhang, C., and Kim, S.H. 2003. Overview of structural genomics: from structure to function. *Current Opinion in Chemical Biology* **7**: 28-32.



## 9 Publication list

\* Scheich, C., Sievert, V., Bussow, K.

An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography.  
BMC Biotechnol. 2003; **3**(1):12.

\* Scheich, C., Niesen, F.H., Seckler, R., Büssow, K.

An automated refolding screen applied to the human dynactin p22 subunit.  
Protein Sci. 2004; **13**(2):370-80

\* Scheich, C., Leitner, D., Sievert, V., Leidert, M., Schlegel, B., Simon, B., Letunic, I., Büssow, K., Diehl, A.

Fast identification of folded human protein domains expressed in E. coli suitable for structural analysis,  
BMC Struct Biol. 2004; **4**(1):4

\* Scheich, C., Niesen, F.H., Seckler, R., Büssow, K.

An automated *in vitro* folding method applied to a human dynactin subunit.  
Poster at the "Faltertage at the Leucorea Wittenberg", September 2003

Manjasetty, B.A., Delbrück, H., Pham, D.-T., Mueller, U., Fieber-Erdmann, M., Sievert, V., **Scheich, C.**, Büssow, K., Niesen, F.H., Weihofen, W., Loll, B., Saenger, W., Heinemann, U.:

Crystal Structure of homo sapiens translational inhibitor protein hp 14.5.  
Proteins. 2004; **55**(3):783

Niesen, F.H., **Scheich, C.**, Delbrück, H., Götz, F., Koch, A., Quedenau, C., Lenski, U., Büssow, K., Harttig, U., Hofmann, K.P., Heinemann, U.

Combined light scattering / precipitation analysis to assess protein crystallizability  
Submitted

\* Results from these publications have been integrated in this thesis

## Danksagung

Mein besonderer Dank geht an meinen Doktorvater, Prof. Dr. Robert Seckler, daß er meine Dissertation an der Universität Potsdam unterstützt hat, daß er mich im Herbst 2001 in sein Doktorandenseminar aufgenommen hat und insbesondere für die zahlreichen Anregungen und Ratschläge, die enorm wichtig für die Erstellung der Arbeit waren.

Die vorliegende Arbeit wurde im Verlauf meiner Tätigkeit im Labor von Dr. Konrad Büssow, bei dem ich seit März 2000 beschäftigt bin, angefertigt. Ohne ihn wäre diese Arbeit nicht zustande gekommen. Er hat mich großartig unterstützt und war immer hilfsbereit und ansprechbar. Dafür bin ich ihm zu großem Dank verpflichtet.

Weiterhin danke ich den Mitarbeitern aus Dr. Konrad Büssows Gruppe, insbesondere Dr. Volker Sievert für die Herstellung der zahlreichen Expressionsklone und Claudia Quedenau und Janett Tischer.

Auch möchte ich Prof. Dr. Hans Lehrach danken, in dessen Abteilung am Max-Planck-Institut für molekulare Genetik in Berlin diese Arbeit durchgeführt wurde.

Diese Arbeit wurde im Rahmen des BMBF-Leitprojektverbundes Proteinstrukturfabrik Berlin angefertigt. Allen Mitarbeitern der Proteinstrukturfabrik danke ich für die angenehme Arbeitsatmosphäre. Mein besonderer Dank geht an Dr. Frank Niesen für die vielen hilfreichen Diskussionen und für die wertvollen Anregungen zum Manuskript. Außerdem danke ich ihm und Anja Koch für die Hilfe bei den FTIR-Studien. Des weiteren danke ich Dr. Anne Diehl und Dr. Dietmar Leitner für die fruchtbare Zusammenarbeit bei der Untersuchung der Expressionsklone der humanen Proteindomänen.

Ich bedanke mich bei allen Teilnehmern des Doktorandenseminars von Prof. Dr. Robert Seckler für ihre wertvollen, fachlichen Anregungen, für die gemeinsamen Abendessen und Ausflüge, und insbesondere bei Dr. Monika Walter für die Hilfe beim Messen der CD-Spektren.

Außerdem danke ich Dr. Grant Langdon und meiner Schwester Monika Scheich für die wertvollen Anregungen zum Manuskript.

Schließlich geht ein herzlicher Dank an meine Eltern für ihre Unterstützung.