TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie

Lehrstuhl für Biotechnologie

Protein renaturation with the help of a genetic algorithm

Danae Dominique Helene Bärend

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Aymelt Itzen
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Johannes Buchner
	2.	Univ.-Prof. Dr. Dirk Weuster-Botz

Die Dissertation wurde am 02.05.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 12.07.2012 angenommen.

Für meinen Großvater

Content

1 Declaration

I, Danae Dominique Helene Bärend, hereby declare that this thesis was prepared by me independently and using only the references and resources stated here. The work has so far not been submitted to any audit commission. Parts of this work have been published in scientific journals.

Hiermit erkläre ich, Danae Dominique Helene Bärend, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit wurde bisher keiner Prüfungskommission vorgelegt. Teile dieser Arbeit wurde in wissenschaftlichen Journalen veröffentlicht.

_____

Danae Dominique Helene Bärend

München,

2 Summary

Protein folding is the series of events that lead to the assembly of a polypeptide chain into its unique three-dimensional conformation. The adoption of this so-called native state is mandatory for proteins to fulfill their function in cells. To produce active proteins in high amounts and preferably at low cost is a task of indispensable industrial and academic importance. The bacterium *Escherichia coli* is the expression system of choice. *E. coli*, however, produces foreign proteins often as inactive aggregates. Refolding and reactivation of those so-called inclusion bodies is possible, if the suitable physical and chemical parameters for the individual protein are identified. Very often, this is a time- and cost-consuming process, in many cases not even leading to a satisfying result.

Refolding of proteins can be mathematically described as a complex problem with multiple parameters contributing to the result and unclear correlations between them. Genetic algorithms are stochastic search methods that have been shown to be a powerful tool for finding solutions for this kind of problems. By mimicking the processes of evolution, they randomly sample a pool of variables as a starting point and further optimize these solutions using the Darwinian principles of selection and variation in an iterative procedure.

In the presented work, a genetic algorithm was applied to search for efficient solutions for the problem of protein reactivation. The aim was to screen for buffer conditions that provide an optimal environment for a specific protein to regain the active state. For this purpose, an experimental setup was developed to perform and evaluate refolding in a high throughput format, allowing the search for efficient refolding conditions. A set of different well characterized model proteins was subjected to the algorithm-based optimization procedure to test the performance of the established system. The selection included the jellyfish protein GFP (green fluorescent protein), bacterial glucokinase, hen egg white lysozyme, and a murine antibody fragment. Furthermore, two mammalian proteins for which no successful refolding was shown so far were examined: a human protease and the murine endoplasmic reticulum resident cochaperone ERdJ3, which is involved in the assembly of immunoglobulins.

For each protein, an individual functional assay was developed in the 96-well microplate format to evaluate the renaturation efficiency. For all assayed proteins, efficient reactivation could be achieved within 6 – 8 optimization cycles, with yields mostly up to 100 % efficiency. The algorithm-based modulation of the buffer environment was shown to potentially also improve the activities of native proteins as well. Additionally, the challenging task to initiate the correct formation of disulfide bridges was taken into account and could be further optimized by adjusting the setup to the specific requirements.

Taken together, it could be shown in this thesis that the non-empirical approach of combining the multidimensional biochemical problem of protein renaturation with a computational method can be employed successfully in the case of all investigated proteins. The established system provides a robust and adaptable setup for screening protein folding conditions with reasonable experimental effort and cost.

3 Zusammenfassung

Proteinfaltung bezeichnet die Vorgänge, während derer eine Polypeptidkette ihre individuelle dreidimensionale Konformation einnimmt. Diese, als nativer Zustand bezeichnete Struktur ist Voraussetzung für die biologische Funktion der Proteine. Die Herstellung aktiver Proteine in hohen Mengen und vorzugsweise zu geringen Kosten ist eine Aufgabe von großer Wichtigkeit in Industrie und Akademie. Dabei ist das Bakterium *Escherichia coli* das Expressionssystem der Wahl. *E. coli* produziert jedoch fremde Proteine oft in Form inaktiver Aggregate. Das Rückfalten und Reaktivieren dieser Aggregate ist möglich, wenn die richtigen physikalischen und chemischen Parameter für das entsprechende Protein identifiziert werden können. Dies ist sehr häufig ein Zeit- und Kosten-aufwendiges Unterfangen, welches in vielen Fällen kein zufriedenstellendes Ergebnis liefert.

Die Rückfaltung von Proteinen kann mathematisch als komplexes Problem mit vielen zum Ergebnis beitragenden Parametern beschrieben werden, wobei spezifische Zusammenhänge zwischen diesen bisher unklar sind. Genetische Algorithmen sind stochastische Suchmethoden, die sich als leistungsfähiges Mittel zur Lösungsfindung für diese Art von Problemen erwiesen haben. Unter Nachahmung des Evolutionsprozesses tasten diese ein Reservoir von Variablen zunächst zufallsbasiert ab und optimieren die ermittelten Lösungen weiter in einem iterativen Verfahren, basierend auf den Darwin'schen Prinzipien der Selektion und Variation.

In der hier vorgestellten Arbeit wurde ein genetischer Algorithmus für die Suche nach effizienten Lösungen für das Problem der Proteinreaktivierung angewandt. Ziel war es, Pufferbedingungen zu finden, die eine optimale Umgebung für ein spezifisches Protein bieten, um den aktiven Zustand zu erlangen. Zu diesem Zweck wurde ein experimenteller Ansatz zur Durchführung und Auswertung von Rückfaltungsexperimenten im Hochdurchsatz-Format entwickelt. Eine Reihe verschiedener und gut charakterisierter Modellproteine wurde dem Algorithmus-basierten Prozess unterzogen, um die Leistungsfähigkeit des entwickelten Systems zu testen. Die Auswahl beinhaltete das Quallen-Protein GFP (green fluorescent protein), bakterielle Glucokinase, Lysozym aus Hühnereiweiß, und ein murines Antikörperfragment. Weiterhin wurden zwei humane Proteine untersucht, für welche bisher

keine erfolgreiche Rückfaltung gezeigt wurde: eine humane Protease und das Cochaperon ERdJ3 aus dem endoplasmatischen Retikulum, welches bei der Biosynthese von Immunglobulinen eine wichtige Rolle spielt. Für jedes der Proteine wurde ein funktioneller Assay im 96-well-Format etabliert, um Rückschlüsse auf die Rückfaltungseffizienz ziehen zu können. Für alle untersuchten Proteine konnte eine erfolgreiche Reaktivierung innerhalb 6 – 8 Optimierungszyklen erzielt werden, oft mit Ausbeuten bis zu 100%. Durch die Algorithmus-basierte Anpassung der Pufferumgebung konnten ebenfalls die Aktivitäten von nativen Proteinen verbessert werden. Zusätzlich wurde die anspruchsvolle Aufgabe der korrekten Ausbildung von Disulfidbrücken gelöst.

Zusammengefasst konnte in dieser Arbeit gezeigt werden, dass der nicht-empirische Ansatz, der das mehrdimensionale biochemische Problem der Proteinrückfaltung mit einer Berechnungsmethode zur Unterstützung der Problemlösung kombiniert, in allen untersuchten Fällen zu erfolgreichen Ergebnissen führt. Das entwickelte System ist eine robuste und anpassungsfähige Methode, um Proteine mit vermindertem experimentellen Aufwand und Kosten erfolgreich zurückzufalten.

4 Introduction

4.1 Protein folding

To follow their individual functions in the cell, proteins have to arrange themselves into a specific three-dimensional conformation, which is commonly referred to as the native state. During the processes of protein folding, newly synthesized, denatured or disordered polypeptide chains adopt their native tertiary structure. There is also evidence for the presence of natively unfolded proteins, where folding occurs only upon ligand- or substrate-binding or during oligomerization (Bachmann, et al. 2011; Fink 2005; Orosz and Ovadi 2011; Tompa 2005)

As Anfinsen postulated, the thermodynamic point of view of the native state represents a global minimum in the accessible Gibbs energy of a given environment. Since folding is a spontaneous process (Anfinsen 1973; Anfinsen, et al. 1961), the gain of structure implies the transition from a less ordered to a more ordered state, concomitant with an energetically unfavorable loss of entropy, which has to be compensated. The hydrophobic factor and peptide hydrogen bonds are regarded as the major driving forces during protein folding (Baldwin 2007; Dill 1990), with further interactions, such as van der Waals forces, and salt bridges, conferring additional stability. Already one unsatisfied hydrogen bond contributing to backbone interactions can tip the balance of the entire free energy of folding for a globular protein (Rose, et al. 2006), illustrating the sensibility of the folding process. While it could be shown for small proteins, that folding occurs in two steps via a transition state as the highest energy point (Jackson and Fersht 1991), pathways leading through local energy minima, so called intermediate states, are commonly observed (Dobson 2003; Jennings and Wright 1993; Kim and Baldwin 1990; Ptitsyn 1995).

The distinct insights into the folding process gave rise to the folding funnel model (see Figure 1), which illustrates the process of folding being multidimensional for most proteins and providing multiple folding routes within a rugged energy landscape. For two-state folders the

folding funnel is smoothened (see inlet) (Dill and Chan 1997; Leopold, et al. 1992; Wolynes, et al. 1995).

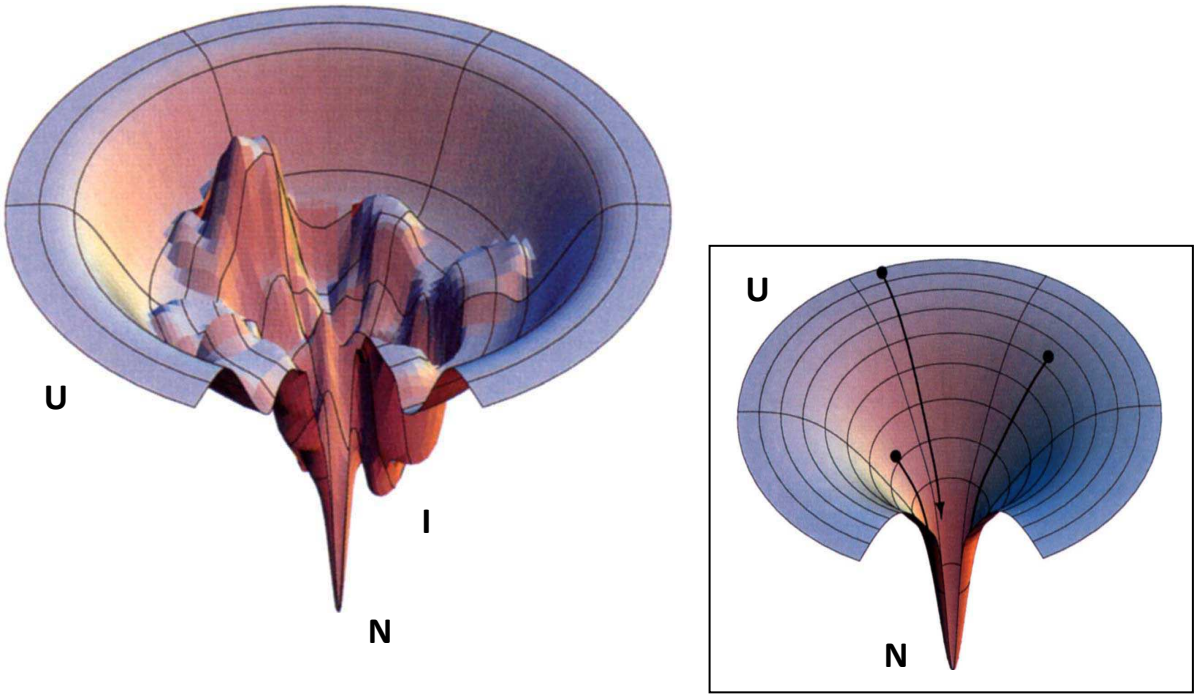


Figure 1 The rugged energy landscape of a folding protein in a three-dimensional visualization, with the free energy of each conformation as a function of the degrees of freedom, U represents the ensemble of unfolded states, I the folding intermediates and N refers to the native state as the global minimum. The inlet shows a folding funnel for a two-state folder.

Adopted and modified from (Dill and Chan 1997).

Additionally, formation of the stable, minimal free energy conformation occurs quickly on biological timescales, typically within milliseconds, despite the astronomical number of possible conformations – a fact that excludes random searching, but rather postulates folding to be an intrinsically guided procedure (Levinthal 1968; Zwanzig, et al. 1992). To date, it is not known in detail which principle governs those reactions, but it is generally assumed, that the information lies in the amino acid sequence (Anfinsen 1973; Fersht and Daggett 2002). Several, computational- and algorithm-based concepts exist to model or simulate the events that take place during folding of a single protein, using e.g. Monte Carlo simulations

(Hansmann and Okamoto 1999; Li and Scheraga 1987) or Markov state models (Bowman, et al. 2011; Pande, et al. 2010). Additionally, advances have been achieved in solving the folding paradoxon and possible native conformations by combining the creativity and rationality of humans with the power of computational approaches in the FOLDIT project (Cooper, et al. 2010). The aim is a growing capacity in the prediction of tertiary structures on the basis of the protein sequences.

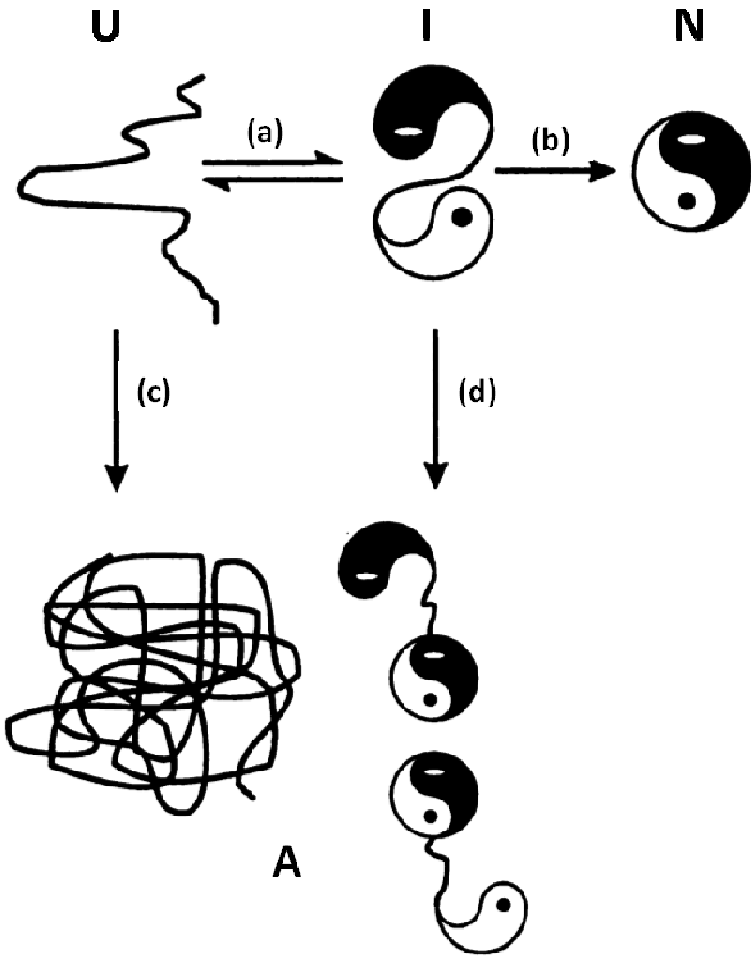


Figure 2 Scheme of reactions in the folding pathway.

The reactions leading to correct folding ((a), (b)) compete with misfolding and aggregation ((c), (d)). Starting from the unfolded polypeptide chain (U) and folding intermediates (I), aggregates can be formed (A) instead of the adoption of the native state (N).

Figure adopted and modified from (Rudolph and Lilie 1996).

During folding, many proteins are prone to misfolding and aggregation, especially those with a rugged folding landscape. This results in the formation of inactive folding products that are trapped in their energy level. A schematic overview of the protein folding pathway is given in Figure 2.

4.2 Recombinant protein production: soluble vs. insoluble

The development of recombinant DNA techniques and the understanding of the translational processes in organisms enabled the expression of foreign, often human proteins in transformed microorganisms at significant amounts for structural and clinical studies (Kamionka 2011). One of the most commercially viable host organisms is the bacterium *Escherichia coli* because it is easy to be genetically manipulated. An additional asset is the fast duplication rate especially in simple and inexpensive media. Besides the expression of proteins for research purposes, approximately 40 % of all biopharmaceuticals are meanwhile produced in *E. coli* cells (Huang, et al. 2012; Terpe 2006). The other available strategies e.g. expression in mammalian cells, insect cells, or fungi render the production more expensive, although posttranslational modifications such as glycosylation or correct disulfide bridge formation are not possible in bacterial cells in general. The production of heterologous proteins in *E. coli*, however, often results in the formation of inactive and insoluble aggregates, so called inclusion bodies (IBs) (Kane and Hartley 1988). These large intracellular proteinaceous granules are visible by light microscopy and most often occur at the poles of the cell (see Figure 3). As several studies of recombinant protein expression in *Escherichia coli* revealed, the formation of IBs is rather a common phenomenon than an exception (Baneyx 1999; Marston 1986). The rapid accumulation of high protein amounts together with the insufficient supply of chaperones guiding their correct folding lead to kinetic changes that favor the aggregation reaction (Georgiou and Valax 1996). In the case of human proteins, for example only 13% can be recombinantly expressed in their soluble form in *E.coli* (Braun, et al. 2002).

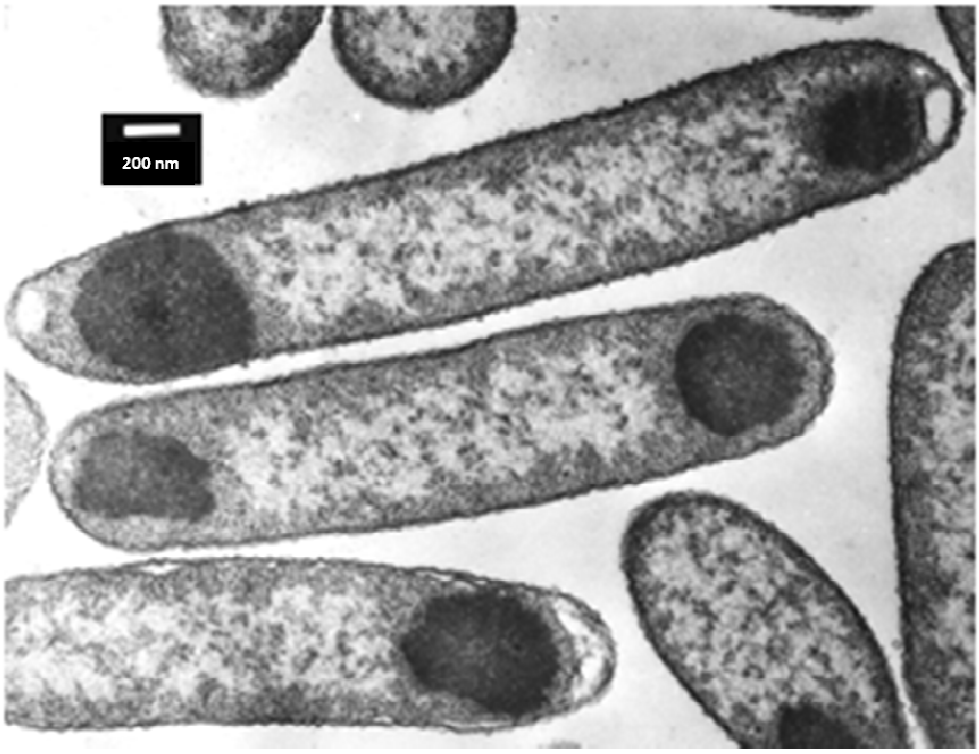


Figure 3 Electron micrograph of inclusion bodies in *E.coli*. Image adopted and modified from http://web.mit.edu/king-lab/www/research/Scott/Scott-Research.html.

Based on the fact that folding is a reversible process, several experimental studies in the past decades were able to show that it is possible to untangle IBs and initiate the folding reaction, thereby recovering the structure and function of the native protein (De Bernardez-Clark and Georgiou 1991; Guise, et al. 1996; Mayer and Buchner 2004). If the activity of the protein can be regained through refolding, *E.coli* can be the superior expression organism compared to eukaryotic systems allowing cheap and high amount production of recombinant proteins. Additionally, IB formation offers several advantages, such as the increased protection from proteolytic degradation and the high degree of purity of the target protein after IB isolation compared to the soluble counterpart making downstream processing an easier task.

4.3 Protein refolding

Despite the potentially less challenging purification steps, the bottleneck in the production of recombinant proteins from IBs is the folding step to the native state. Methods to solubilize and renaturate this paracrystalline protein form have been developed and were constantly improved (Buchner, et al. 1992; Eiberle and Jungbauer 2010; Rudolph and Lilie 1996; Singh and Panda 2005). Commonly, the aggregates are solubilized with chaotropic buffers containing e.g. guanidine hydrochloride or urea and providing a reducing environment if cysteines are present to allow their complete reduction in case non-native disulfide bridge formation has occurred in the IBs. Solubilization might be followed by purification steps before refolding which is mandatory to recover the native and active conformation of the protein.

Kinetically, folding and aggregation compete as first and second order reactions, respectively, upon refolding, with their rates determining the yield of reactivated protein (Kiefhaber, et al. 1991). A strong dependency of the success on the refolding conditions and physicochemical environment is therefore given by favoring either of the two reactions.

Different strategies exist to initiate refolding by transferring the protein to an appropriate buffer environment that favors the formation of native structure (Lilie, et al. 1998; Vallejo and Rinas 2004). They mainly comprise I) dilution, II) dialysis, or III) chromatography-based refolding steps of the denatured, solubilized protein to allow denaturant removal and transfer to the refolding buffer conditions. The simplest refolding technique is that of direct dilution into the respective buffer, demanding no special equipment and providing freedom in choice of buffer substances. Within this procedure additional concentration steps and high buffer volumes are drawbacks during scale-up. Final protein concentrations in the 1-50 µg/ml range are most frequently used to favor the folding instead of the aggregation pathway (Rudolph and Lilie 1996; Vallejo and Rinas 2004). This method can be improved by adding several aliquots of the denatured and solubilized protein to the refolding buffer in a consecutive manner, a procedure termed pulsed refolding (Buchner, et al. 1992). This reduces the risk of unwanted side reactions such as aggregation and allows an increase of the volume yield. The second strategy is the stepwise transfer of the denatured protein to

refolding buffer conditions with the use of diafiltration and dialysis. Though successful in some cases, it leads in most cases to more precipitation than the dilution method because of the accumulation of aggregation-prone folding intermediates (Clark 2001). This procedure also bears the risk of non-specific adsorption of the protein sample to the membrane material. Furthermore, refolding on column is a promising method, since it offers the advantage of reactivating proteins at high concentrations, needing comparably low buffer volumes due to the reduction of intermolecular interactions by spatial isolation (Dashivets, et al. 2009). However, the choice of substances in the refolding buffer is constrained due to the limitations of the respective column matrix material (e.g. pH value, salt concentration, redox agents, detergents).

4.4 Oxidative refolding

Refolding of proteins is a more demanding task for disulfide-bridged proteins such as antibody fragments or lysozyme, as the correct oxidation of the cysteine pair has to be achieved autonomously under the chosen refolding conditions. In contrast, the formation of non-native disulfide bridges will lead to non-functional proteins.

As Anfinsen has shown in his seminal experiments, the oxidation can be performed during *in vitro* refolding (Anfinsen and Haber 1961). Many studies reported the successful reoxidation of cysteines, resulting in correctly refolded and reactivated proteins (Buchner and Rudolph 1991; Hevehan and De Bernardez Clark 1997; Rudolph 1990). Distinct oxidizing agents have been investigated, ranging from the utilization of molecular oxygen up to agents such as the more recently described selenoglutathione (Beld, et al. 2007). The addition of the naturally occurring redox system of glutathione in its reduced and oxidized form (GSH and GSSG, respectively) is most widely used (Rudolph, et al. 1993). In terms of the buffer composition, the correct ratio of oxidizing and reducing agents has to be experimentally determined in combination with the other buffer components, since several thiol-species are present in equilibrium during disulfide bond formation (see Figure 4). In general, slightly basic pH values favor the formation of correct cysteine pairs as thiolate anions are necessary for the

obligatory formation of mixed disulfides in the oxidation reaction due to the pKa value of a typical thiol around 8.3 (Shaked, et al. 1980).

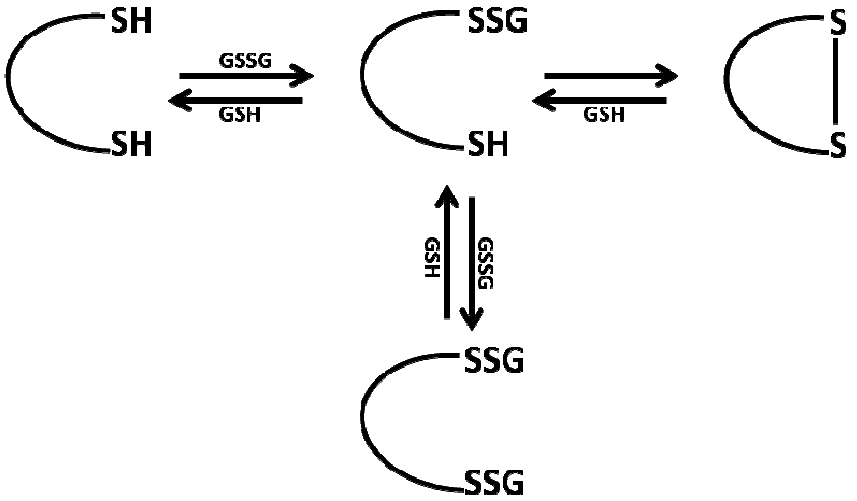


Figure 4 Scheme of redox equilibria during the formation of an intramolecular disulfide bond

4.5 State of the art for refolding

To date, most protocols developed for the efficient refolding of proteins are elaborated empirically on a case-by-case basis. Suitable strategies have to be tested and improved individually for the targeted protein, comprising extensive and often time consuming search episodes (Alibolandi and Mirzahoseini 2011; Eiberle and Jungbauer 2010; Singh and Panda 2005; Vallejo and Rinas 2004). Obtaining high amounts of active proteins at low experimental cost is the main objective in *in vitro* refolding but no general single strategy exists that satisfies all refolding projects. The constant growth of the REFOLD database gives an insight into the enormous number of renaturation protocols being established every year (Chow, et al. 2006a; Phan, et al. 2011). Though, the increasing pressure to produce and refold proteins from IBs for industrial and academic purposes also gave rise to rational

approaches to tackle the refolding problem in a more systematic attempt by applying high-throughput screening and performing statistical trend analysis of the obtained results (Basu, et al. 2011; Coutard, et al. 2012; Vincentelli, et al. 2004).

4.6 Genetic algorithm

In order to search for solutions for a complex multidimensional problem, where multiple factors are known to be involved without direct correlations, genetic algorithm-based methods provide a powerful tool. They are stochastic search and optimization strategies that computationally mimic the Darwinian process of evolution. The evolutionary principles of selection, inheritance, and recombination based on mutation and crossover are used to iteratively improve the qualities of an initially random assembled generation of solutions to the defined problem. In informatics, genetic algorithms are a popular tool since they were first introduced (Holland 1975; Rechenberg 1973) but they were also successfully employed for other experimental optimizations, e.g. in combinatorial chemistry (Gobin, et al. 2007; Weber 1998), in the bioengineering sector or in targeting fermentation media optimization (Havel, et al. 2006; Weuster-Botz 2000; Zuzek, et al. 1996).

In an experimental optimization based on a genetic algorithm, the process starts with a random population of solutions for the addressed problem. This set is generated through the combination of different parameters and variables, which were determined as being critical for solving the problem and are compiled in the parameter space. Each solution is experimentally tested and the experiments are evaluated with regard to one or more chosen quality features that represent the defined problem (i.e. objectives). This evaluation results in the assignment of a fitness value to each solution. According to their fitness value, satisfying solutions are selected and, by incorporation of the evolutionary principle of variation, recombined to a new population. This circle of events is repeated until the process is terminated by the user (see Figure 5).

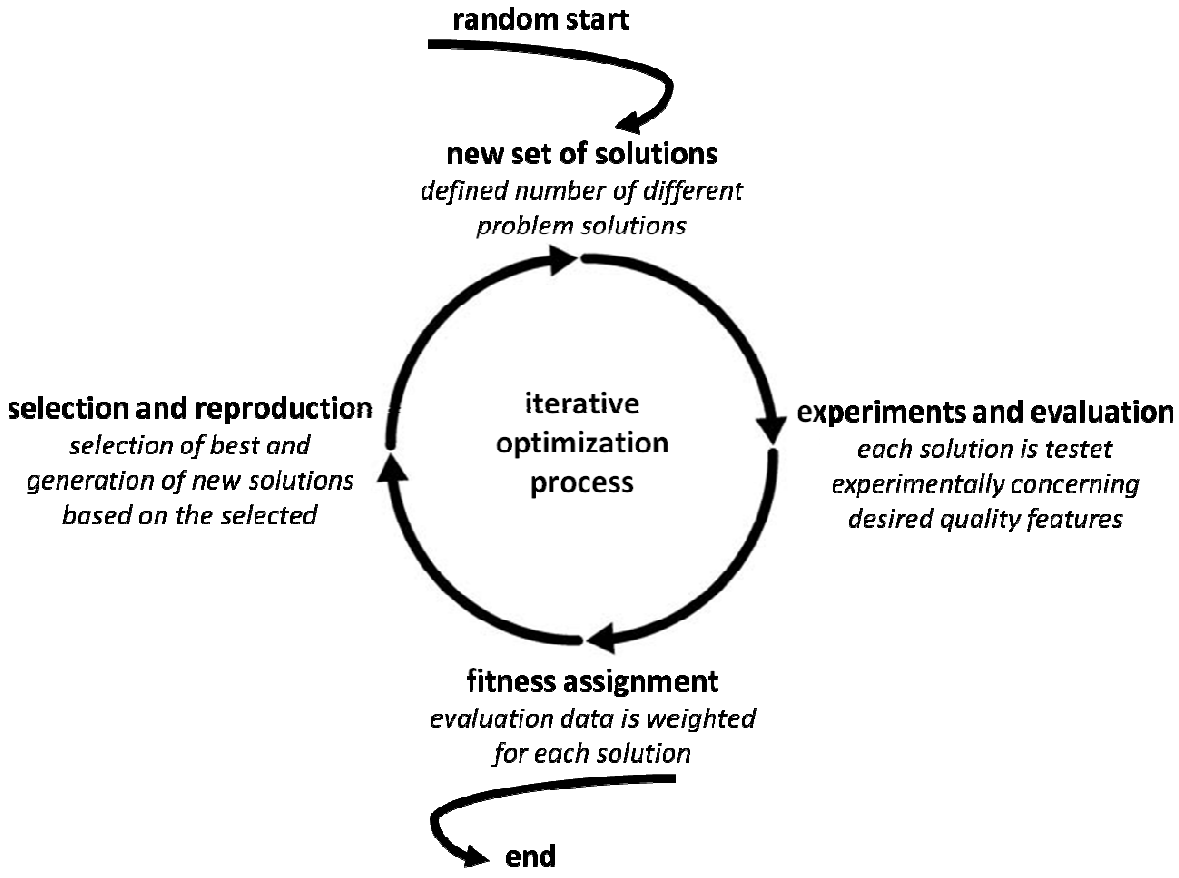


Figure 5 The iterative process of an experimental optimization, based on a genetic algorithm. Adopted and modified from (Anselment, et al. 2010).

5 Objective

In the process of recombinant protein production, which often leads to the formation of non-functional inclusion bodies, the step of refolding is a crucial bottleneck. Protein refolding is still a poorly understood biophysical event, which so far was addressed by random screens to search for suitable refolding conditions.

In this work, a multiobjective stochastic search method should be applied to find and optimize buffer conditions, which provide an efficient environment for recombinantly expressed insoluble proteins to (re-)gain their native structure and function. In contrast to empiric approaches displaying a time consuming, often not satisfying method, a computational format represents a promising tool to overcome this challenge. Therefore, a genetic algorithm based optimization system should be established and evaluated experimentally.

Protein refolding is dependent on individual yet often unknown buffer components, with many variables possibly contributing to the success. An important task for this approach therefore is the definition of an appropriate multi-parameter space to allow efficient and global screening.

Model proteins of different function, origin, and individual characteristics should be chosen as well as less well characterized proteins and subjected to the refolding optimization strategy, allowing statements on the performance and the feasibility of the designed system.

The optimization process should be multidimensional, including objectives such as refolding efficiency, and provide a general applicability with the experiments carried out in a high throughput format.

Thus, suitable refolding procedures for each protein should be assessed. Additionally, a respective read out to evaluate the refold success should be developed for the selected proteins and transferred to a microplate format.

During the folding process of eukaryotic endoplasmic reticulum – resident or extracellular proteins that contain disulfide bridges in their native conformation, the correct pairing of

cysteines is essential and often represents the limiting step. A further objective of this thesis therefore comprises a modulation of the genetic algorithm to allow refolding of such demanding and challenging proteins.

Having achieved a proof of concept for well characterized model proteins, the suitability of the approach for defining efficient refolding conditions of so far only poorly characterized proteins should be addressed. Furthermore, reliable enzymatic assays in high throughput formats for proteins without a native control should be established.

6 Results

Prior to the implementation of the genetic algorithm, the search space comprising all possible buffer variables and their concentration ranges were defined. Application and implementation of the algorithm was done by the project cooperation partner Bernd Anselment (TU München, Faculty of mechanical engineering). Having accomplished the design of experiments, the refolding progress of different proteins was examined in an iterative optimization of the buffer environment. With regard to the individual functions and properties of the proteins, assays in the multi-well format were purchased or developed to assign the respective refolding efficiencies.

6.1 The experimental setup

In order to work out a setup for the realization of generating refolding conditions to be tested and evaluated experimentally, two main issues were addressed first: which substances in which concentrations should be offered to the algorithm for selection (part 6.1.1), and which objectives should be optimized (part 6.1.2).

6.1.1 Design of the parameter space

In order to compose the pool of substances, from which the algorithm assembles the different buffer conditions, experimental information from the refolding literature was gathered and analyzed, together with the broad spectrum of ~1000 datasets on refolding experiments from the REFOLD database (Buckle, et al. 2005; Chow, et al. 2006b). Based on this data analysis, a set of refolding components was chosen and sorted into clustered subgroups. Further, the applied range for those substances was defined.

The clustering resulted in 8 parameter groups, and for the generation of a set of solutions, referred to as generation in the following, different rules were defined concerning the options for buffer assembly: the choice of a pH-value and buffer system is stringent, whereas for the other subgroups either two options are possible – none of the grouped components or one of them in a certain concentration, or more options, depending on the number of group members – none of them, one of them, or an arbitrary combination. The components enclosed in the parameter space and the respective adoptable concentrations or values are listed in Table 1.

The defined variables and their adoptable values were coded into bit-form and embedded into the search program, constituting the basis for the composition of the evolving protein renaturation environments, which were experimentally tested in this work.

A number of 22 solutions were set to constitute one generation.

Table 1: The components in the parameter space

The variables are clustered into parameter subgroups (left column); within some of the parameter groups, combinations of the compounds are possible to be chosen by the algorithm; the optional range is indicated in the central columns.

Parameter Group			
Component	**Minimum**	**Maximum**	**Unit**
pH - value			
pH - value	6.0	9.5	
Buffer system			
phosphate	20	100	mM
Hepes	20	100	
MOPS	50	100	
Tris/HCl	20	1250	
Salt Combination			
NaCl	0	350	mM
KCl		20	
Amino acids			
arginine	0	750	mM
glutamine		100	
glycine		150	
Additives Combination			
glycerol	0	15	% v/v
PEG-4000	0	0.2	% w/v
Cofactors			
Cu, Zn, Mg, Ca, Mn	0	5	mM
EDTA	0	2	mM
Detergents			
zwitterionic			
CHAPS	0	10.7	mM
Zwittergent 3-12		4	
NDSB 201		1500	
nonionic			
Tween 20	0	0.08	mM
Triton X-100		0.08	
Brij 35		0.12	
Ionic			
SDS	0	12	mM
Sodium desoxycholate		8	
Redox agents combination			
DTT	0	10	mM
TCEP		10	
GSH		5	
GSSG		5	

6.1.2 Definition of optimization objectives

The experiments introduced in the present work are based on a two-objective optimization strategy. This enables a search for solutions providing fitness in two quality features at the same time and optimizes them towards both objectives (see Figure 6). Solutions with higher efficiency values dominate those with lower ones and are selected to be recombined for the next generation, based on the principles of evolution.

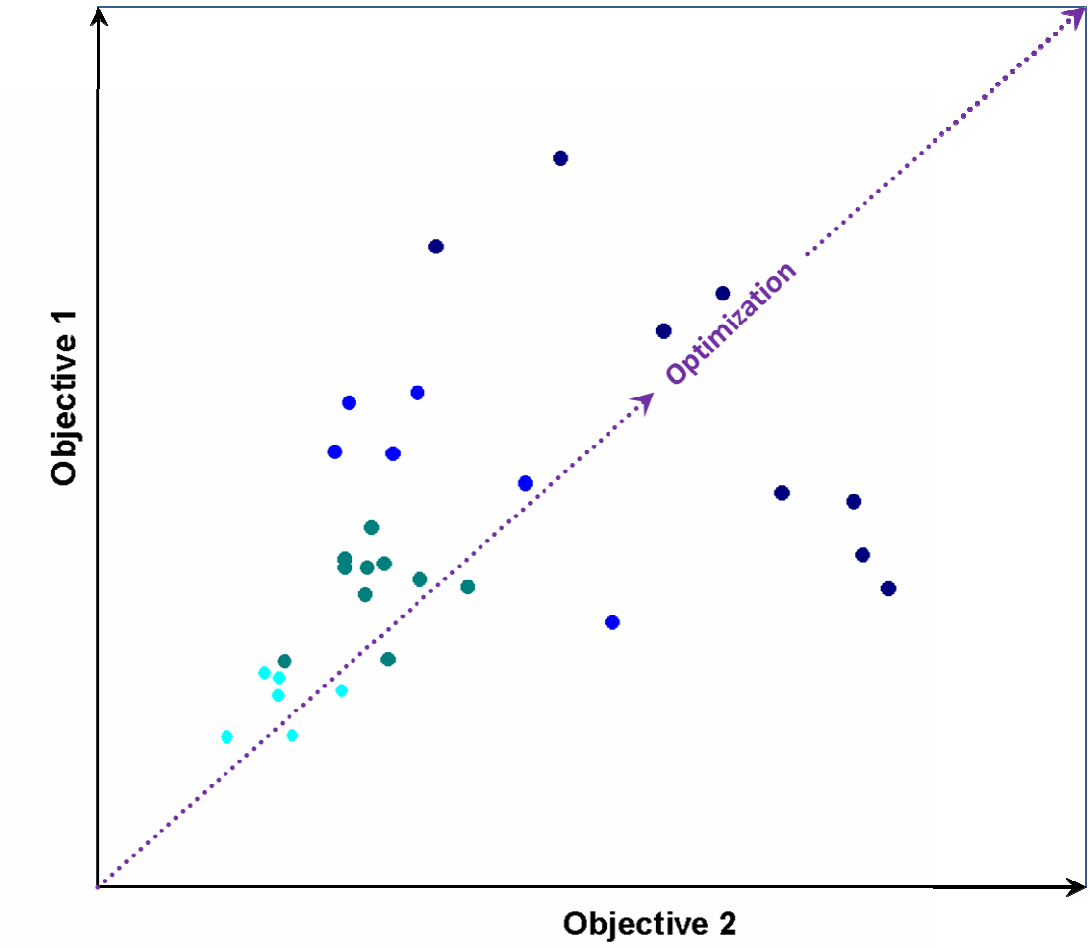


Figure 6 Graphical illustration of a dataset in the two-objective-optimization process

After evaluation, a fitness value concerning two objectives is assigned to each of a set of solutions subjected to the optimization. Graphically, the optimum of both objectives is located in the upper right corner of the diagram, and in a functional optimization, datapoints should move over the generations towards that region of ideal solutions.

In the presented context, the following objectives were chosen first:

I. The minimization of the experimental cost, calculated by addition of substance prizes (prizes adopted from sigma Aldrich, Germany, catalogue 2009) as objective 1

II. An increase in refolding yield, obtained by comparison of the functional performance of the refolded with the native protein, calculated in percent, as objective 2

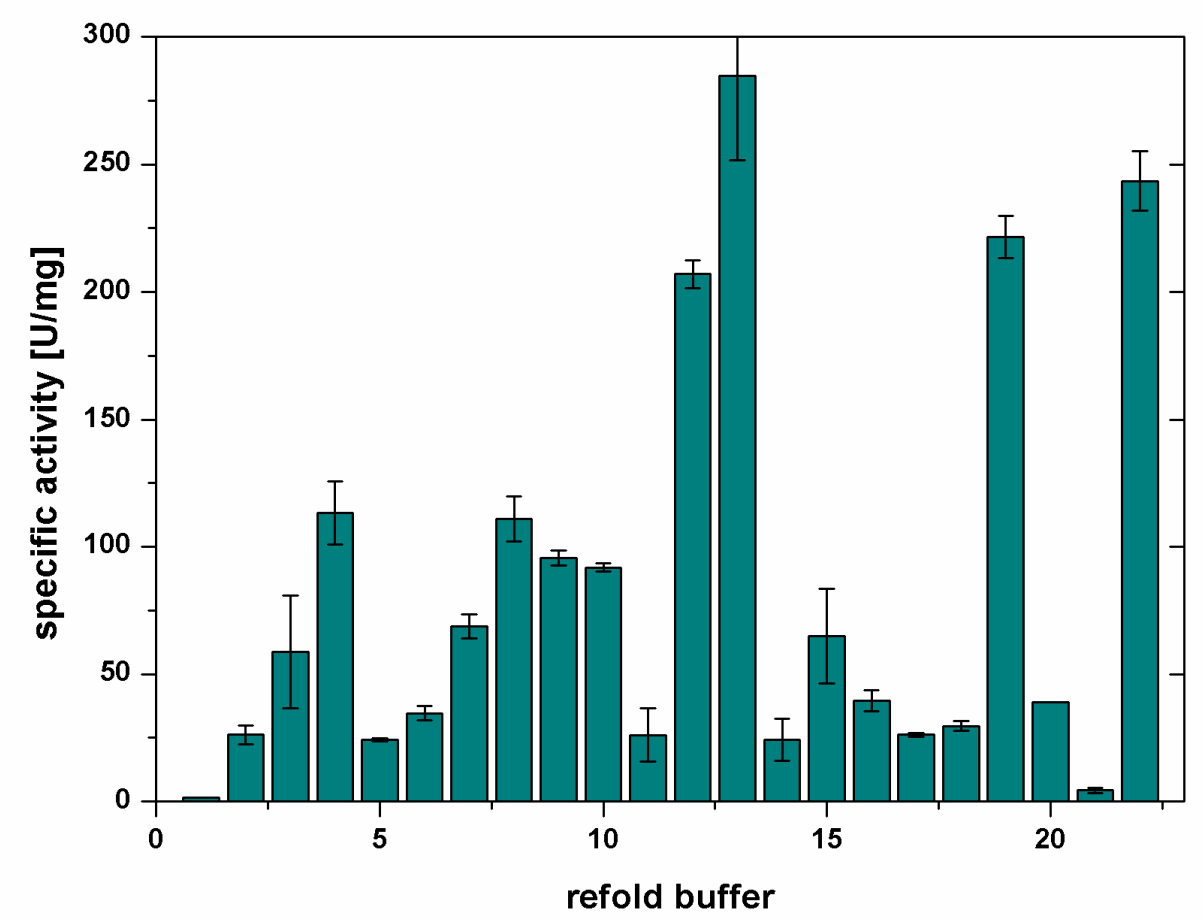


Figure 7: Influence of differently composed buffers on the native activity of glucokinase

Determination of the specific activities of the native protein (y-axis), incubated in 22 randomly generated refolding buffers (x-axis). For all incubation samples, the same protein batch was used to rule out any side effects deriving e.g. from storage time or purification factors. All experiments were done in triplicates.

During several test runs, it became clearly visible, that a strong influence of the buffer environment on the function of the native proteins exists. An example is given in Figure 7, showing the behavior of native glucokinase in the enzymatic assay (described in part 6.2.2 and 8.6.8.2) after incubation in 22 distinct buffers. In the depicted case, specific activities ranging from ~0.5 to ~280 U/mg are possible within one set of solutions. This effect in turn could during optimization of folding parameters in some cases suggest enormously high refolding yields despite of activity levels being actually low (Figure 8).

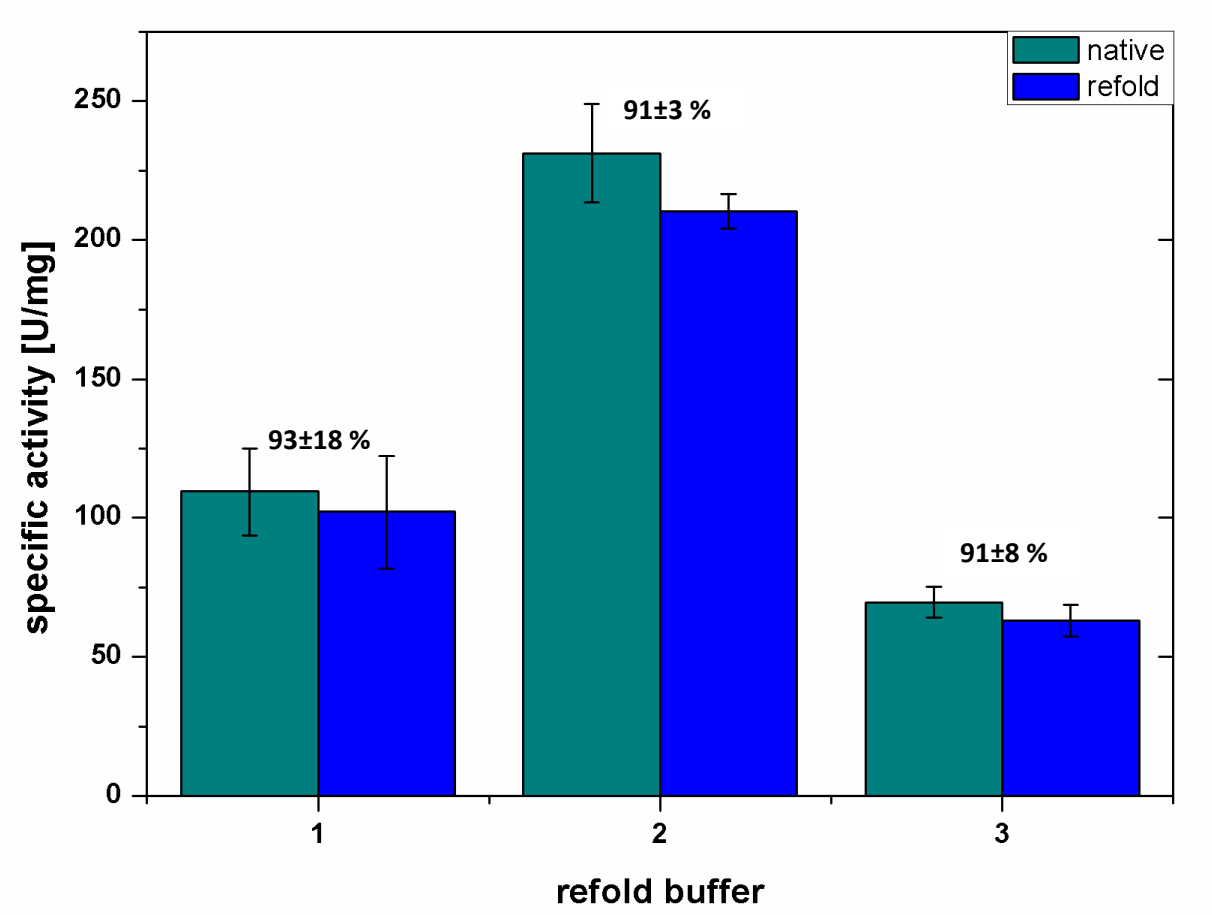


Figure 8 Influence of different buffer effects on yield calculation

Determination of specific kinase activities (y-axis) after refolding of denatured, and incubation of native protein in three different buffers (x-axis). Despite of distinct activity levels, all three samples provide 90-95% yield values due to buffer effects limiting the maximum possible performance. 1) 40 mM phosphate-buffer, pH 6, 0.25 M NaCl, 20 mM KCl, 0.15 mM PEG-4000, 0.5 M arginine, 50 mM glutamine, 25 mM glycine, 5 mM EDTA, 7.5 mM TCEP. 2) 0.2 M Tris/HCL, pH 8.5, 0.15 M NaCl, 0.05 mM PEG-4000, 0.25 M arginine, 0.1 M glutamine, 12.5 mM glycine, 2 mM EDTA, 1 mM Zwittergent3-12, 6.25 mM DTT. 3) 20 mM Hepes, pH 9.5, 0.35 M NaCl, 5 mM EDTA, 4 mM Zwittergent3-12, 2.5 mM DTT. The same batch of glucokinase was used for all samples, and all experiments were done in triplicates.

It was therefore decided to optimize proteins, which turn out to be affected in the native state by different buffers, towards

I. The activity enhancement of the refolded protein as objective 1
II. The activity enhancement of the native protein as objective 2

The general process of experimental optimization was thereby adapted to the thus defined task of protein refolding: 22 buffer solutions are to be generated for each population by the algorithm, tested in experiment and selected and recombined to new generations to iterative optimize the chosen objectives (Figure 9).

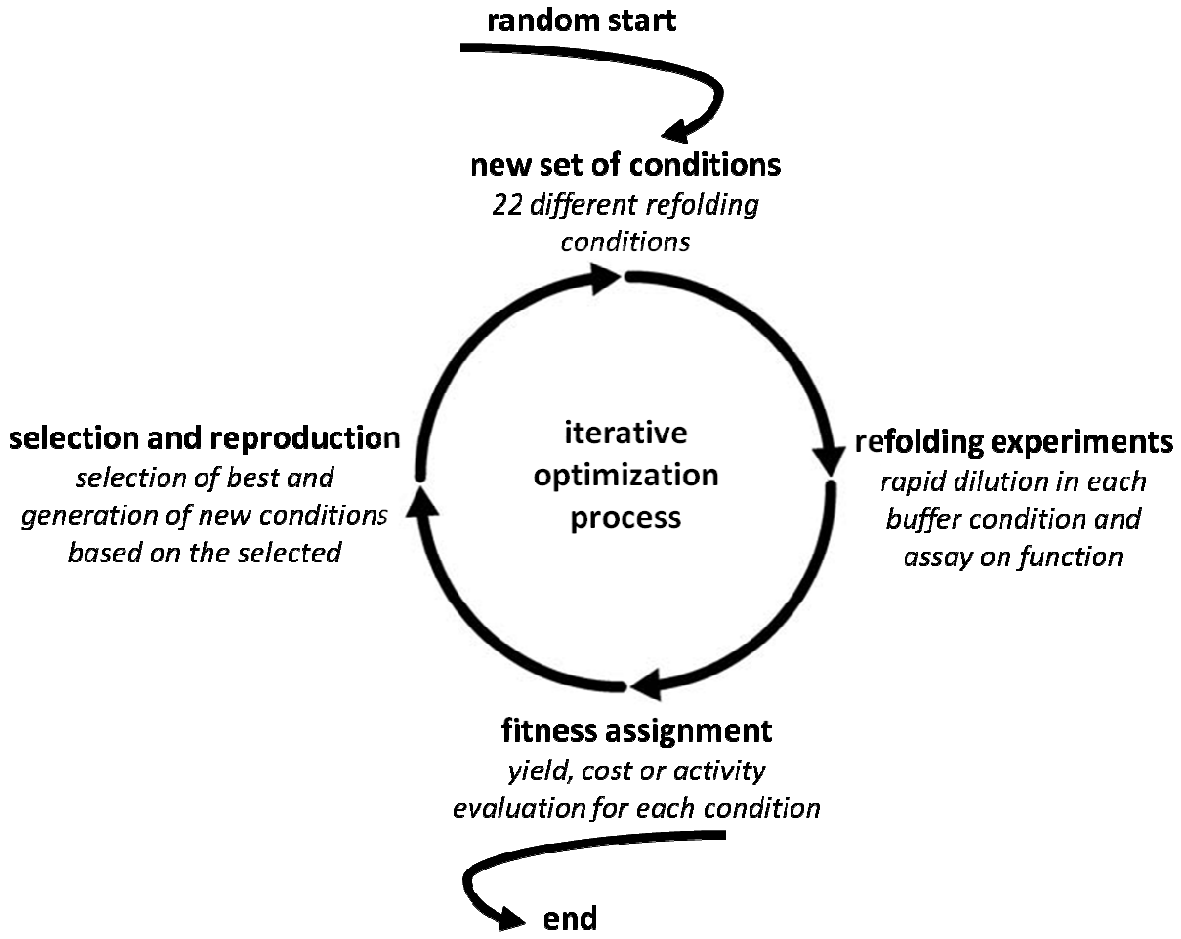


Figure 9 The genetic algorithm based optimization process, adapted to the problem of protein refolding.

6.2 Optimization processes of protein refolding experiments

Based on the established experimental setup, a set of well characterized model proteins was subjected to the optimization process. Recombinantly expressed green fluorescent protein, bacterial glucokinase, and lysozyme from chicken egg white were chosen for this purpose. To expand insights gained in the lysozyme refolding experiments, the modified setup of the parameter space for optimization of refolding proteins containing disulfide bridges was applied in a refolding screen for the MAK33 Fab fragment consisting of the light chain and the V_H and C_H1 domains of the heavy chain of IgG. The screen was performed using a native purified Fab fragment for denaturation, but the results were also verified by refolding purified inclusion bodies in selected buffer conditions. Further, a suitable assay was developed to monitor the refolding success of the J-protein ERdJ3 (mouse), which can be expressed recombinantly only in its insoluble form. In the case of an extracellular protease, the refolding experiments and efficiency read out were carried out without having a native reference within the screen, giving an example for the possibility of a 'bottom up'-optimization with the system.

The renaturation experiments as well as the different assay principles were executed in the 96-well format and the structural integrity deduced in most cases from monitoring the functional performance of the investigated proteins. All proteins were refolded by rapid dilution into 1 ml buffer with a final concentration of 1-10 µg/ml and incubated over night at 4°C with the exception of MAK33 Fab (72h at 10°C).

Native eGFP and glucokinase were purified by Tetyana Dahivets (TU München), native Lysozyme and the native MAK33 Fab fragment were from Invitrogen and Roche, respectively.

Computational analysis and recombination of the datasets was done by the cooperation partner Bernd Anselment (Faculty of Mechanical Engineering, Institute of Bioverfahrenstechnik).

6.2.1 GFP

Green fluorescent protein (GFP) originates from the jellyfish *Aequorea Victoria*. It has a calculated molecular mass of 28 kDa and contains a chromophore consisting of modified amino acid residues, fluorescing green when excited with ultraviolet blue light with excitation maxima at 395 and 475 nm and the emission maximum at 509 nm (Chalfie 1995; Shimomura, et al. 1962; Tsien 1998).

Formation of the chromophore and the ability to emit fluorescence require the native protein fold (Ormö, et al. 1996). In *E.coli*, the protein can fold and mature, without the need of exogenous cofactors. With those properties, the protein constituted an ideal tool to evaluate the capacity of the refolding strategy to be established.

In this work, a point mutant (enhanced GFP (EGFP)) that enables efficient folding at the *E. coli* growth temperature of 37°C (Thastrup, et al. 2001) was used.

Denaturation was carried out in the presence of 6 M GdmCl and refolding occured by dilution to a final concentration of 1 µg/ml. The amount of native structured refolded protein was concluded from monitoring the fluorescence signal of the refolded samples compared to the respective native samples (see part 8.6.8.1). For efficiency assignment, the following optimization objectives were set as desired quality features:

I. high yield
II. low experimental cost

To investigate the optimization process of GFP, a refolding screen over 6 generations was performed and terminated at that point due to a high number of efficient results. During the optimization, the amount of successful solutions increased over the generations and highly efficient buffer compositions were achieved, providing yields of up to 100 % with minimal experimental costs (figure 4). An overview over the best identified buffer conditions during the optimization screen is given in Table 2.

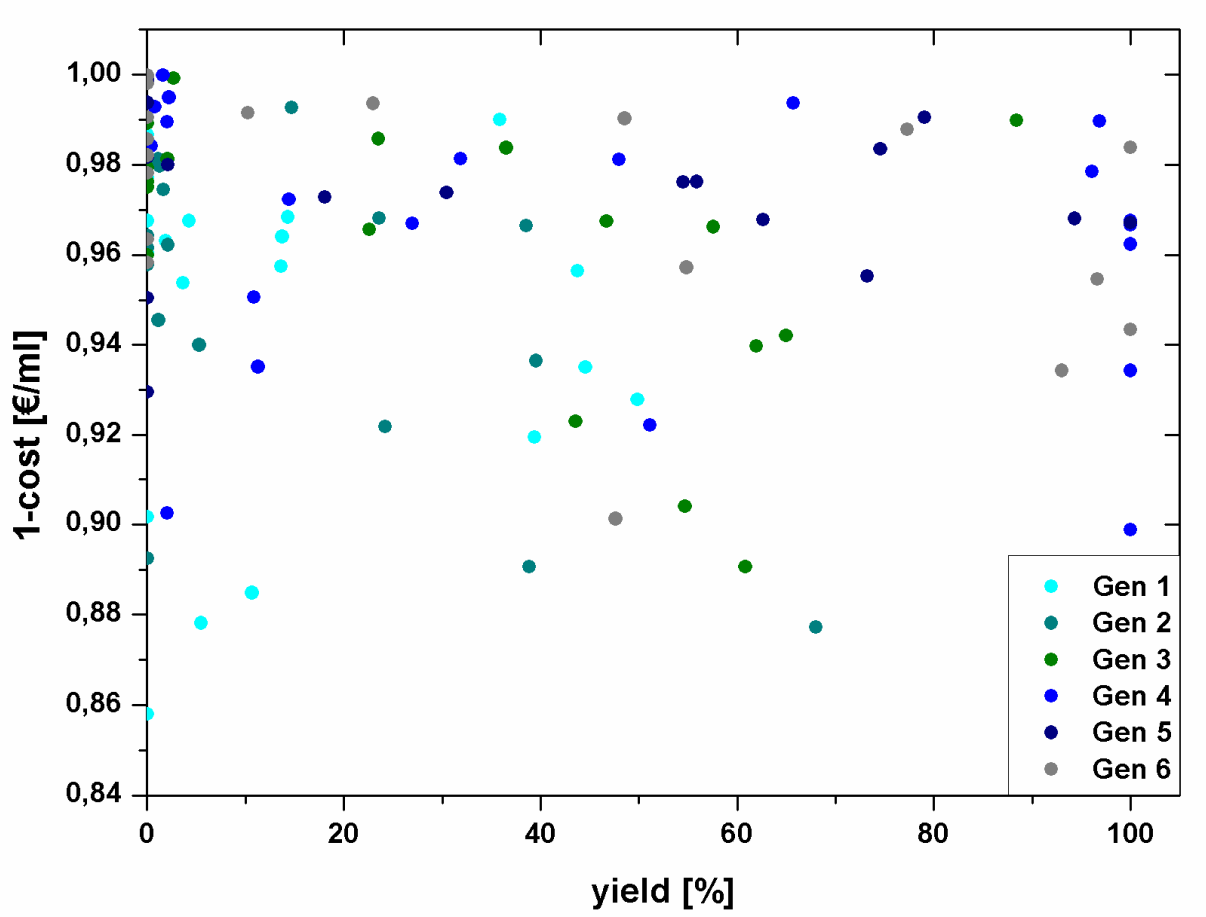


Figure 10 Optimization progress of GFP refolding

Efficiency evaluation of refolding solutions for GFP as function of yield (x-axis) and cost (y-axis). The refolding experiments over six generations are depicted as dots, with datasets of one generation sharing the same color, respectively, and being referred to as Gen 1-6 in the plot legend. All experiments were performed using identical incubation and fluorescence measurement conditions (see part 8.6.8.1). For the individual datasets see appendix (part 6.2.7)

Table 2 most efficient buffer compositions of the GFP refold optimization process

Buffer ID	Buffer composition	Efficiency
Gen 6 / 5	40 mM phosphate buffer, pH 7.0, 100 mM NaCl, 10 % v/v glycerol, 50 mM arginine, 50 mM glutamine, 5 mM EDTA, 7.5 mM DTT	100 ± 9.6 % yield 0.012 € ml^{-1}
Gen 4 / 16	50 mM Tris/HCl, pH 7.0, 250 mM NaCl, 15 % v/v glycerol, 100 mM arginine, 50 mM glutamine, 2.5 mM TCEP	100 ± 7.5 % yield 0.033 € ml^{-1}

6.2.2 Glucokinase

The *E.coli* glucokinase was chosen as a further model protein to test the efficiency of the refolding system. The 35 kDa protein catalyzes the first step of glycolysis, the phosphorylation of glucose to glucose-6-phosphate.

The activity range of the native enzyme is well described in the literature (Meyer, et al. 1997); it therefore represents an interesting target to evaluate the performance of the optimization system.

An assay in 96-well-format was established to monitor the refolding efficiency of glucokinase. In this assay, an ATP-regenerating system couples ATP-consumption to NADH-oxidation (Norby 1988); ATP-consuming phosphorylation of D-glucose by glucokinase can thus be monitored spectroscopically via a decrease of NADH absorption.

Prior to refolding, the protein was denatured in the presence of 6 M GdmCl, followed by refolding through rapid 1:50 dilution and incubation of the samples at 10°C. Glucose phosphorylation was assayed for 20 minutes to calculate specific activities (see part 8.6.8.2).

Due to strong influences of the different refolding buffers on the activities of the native protein (see part 6.1.2), we choose the enhancement of the following objectives of the optimization:

I. The activity of the refolded protein
II. The activity of the native protein in the respective buffer system

The optimization process (Figure 11) generated efficient solutions over the generations, improving in terms of the activities of the refolded kinase as well as those of the native protein, and with activity values that were in good match with those from the literature (Meyer, et al. 1997).

Showing no further progress in the 7^{th} generation, the optimization screen was terminated at that point. The best identified buffer conditions for glucokinase during the experimental optimization are listed in Table 3.

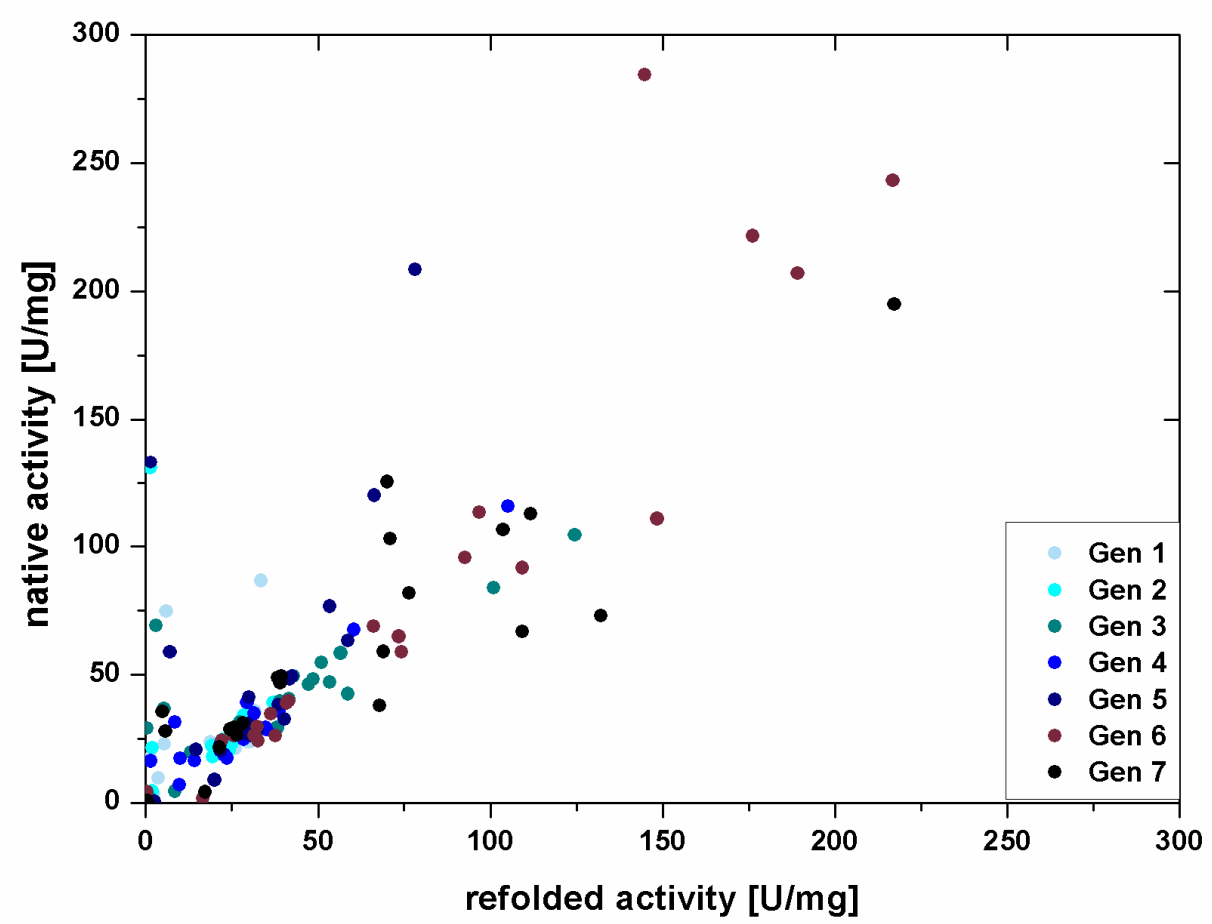


Figure 11 Optimization process of glucokinase refolding

Efficiency evaluation of refolding solutions for Glucokinase by determination of specific activities of the refolded (x-axis) and the native protein (y-axis). The refolding experiments over seven generations are depicted as dots, with datasets of one generation sharing the same color and being referred to as Gen 1-7 in the plot legend. All experiments were performed using identical incubation and activity measurement conditions (see part 8.6.8.2). For the individual datasets see appendix (part 6.2.7)

Table 3 most efficient buffer compositions of the Glucokinase refold optimization process

Buffer ID	Buffer composition	Efficiency
Gen 7 / 18	20 mM Hepes, pH 9.5, 350 mM NaCl, 0.05 mM PEG-4000, 5 mM EDTA, 5 mM DTT	195 ± 21 U mg^{-1} native 217 ± 7 U mg^{-1} refolded
Gen 6 / 22	20 mM Tris/HCl, pH 9.5, 50 mM NaCl, 0.15 mM PEG-4000, 50 mM arginine, 50 mM glutamine, 5 mM EDTA, 7.5 mM TCEP	243 ± 12 U mg^{-1} native 216 ± 7 U mg^{-1} refolded

6.2.3 Lysozyme

Lysozyme is a monomeric 14 kDa enzyme that catalyzes the hydrolyzation of ß-(1-4)-glycosidic linkages between N-acetylmuramic acid and N-acetyl-d-glucosamine in the cell walls of various microorganisms (Phillips 1967).

Its three-dimensional structure being well known (Blake, et al. 1965) and as its function is recordable in a commercially available activity assay, lysozyme represents a suitable model system for refolding studies. In its folding process, four disulfide bridges have to be formed.

The functionality of Lysozyme was assayed according to the EnzChek® Lysozyme Assay. Providing fluorescein-labeled *Micrococcus lysodeikticus* cell walls as substrate, it allows monitoring an increase in fluorescence upon lysozyme activity in a microplate reader.

We performed three approaches to screen for optimized refolding conditions, using the following objectives:

I. The activity of the refolded
II. The activity of the native protein

In the first approach, four generations were evaluated using our standard setup, leading to only 2 cases with refold success out of 88 solutions tested (Figure 12)

A common feature of the efficient buffer conditions was the presence of the oxidizing agent GSSG. A second optimization was therefore started, using a modified setup, which allows only solutions containing GSSG.

Despite a progress of the optimization process in that second approach, still a large amount (~80%) of the refold conditions were ineffective (Figure 13).

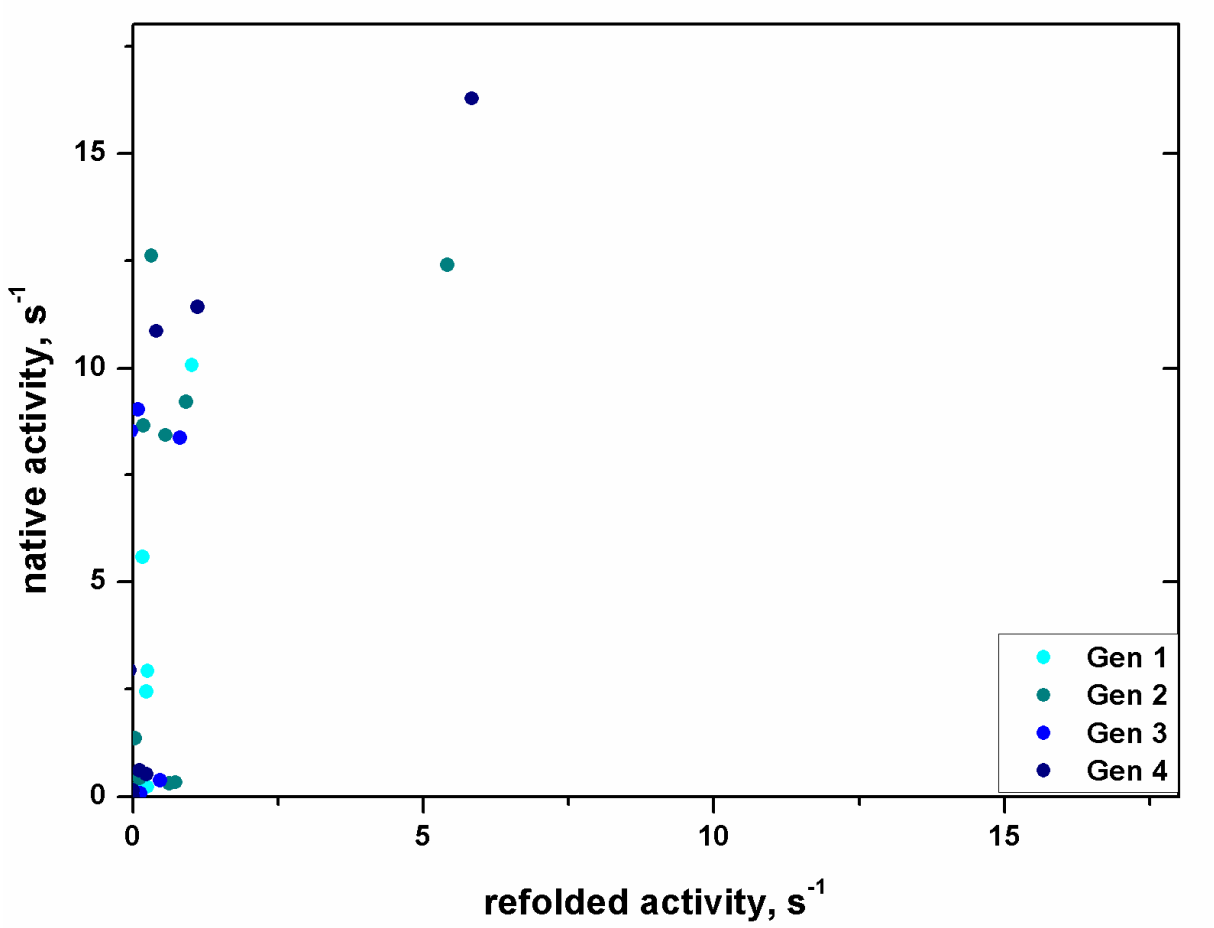


Figure 12 First approach of lysozyme refolding optimization

Efficiency evaluation of refolding solutions for lysozyme by activity determination for the refolded (x-axis) and the native protein (y-axis), given as change in relative fluorescent unit (ΔRFU s^{-1}). The refolding experiments over four generations are depicted as dots, with datasets of one individual generation sharing the same color and being referred to as Gen 1-4 in the plot legend. All experiments were performed using identical incubation and activity measurement conditions (see part 8.6.8.3). For the individual datasets see appendix (part 6.2.7)

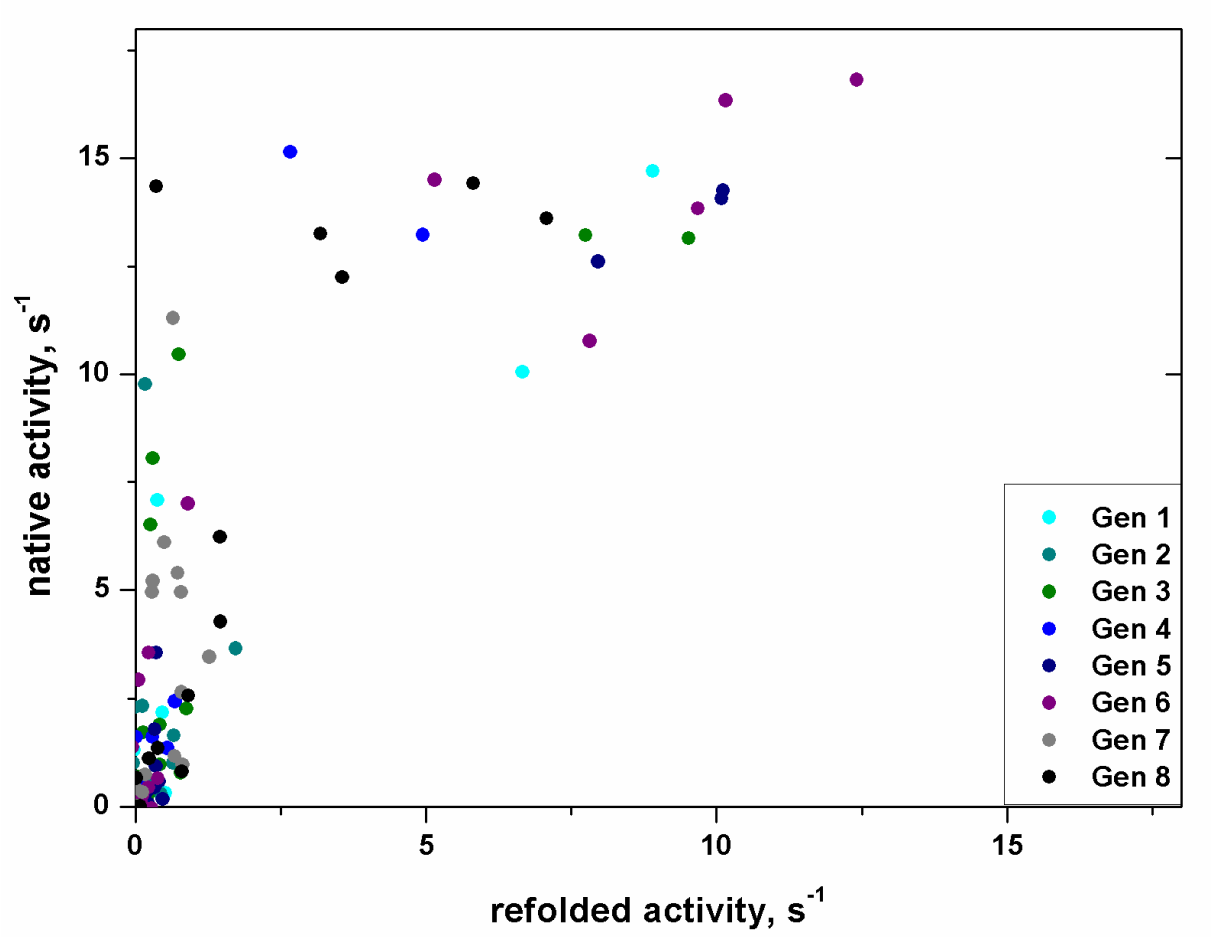


Figure 13 Second approach of lysozyme refolding optimization

Efficiency of refolding solutions for lysozyme by activity determination for the refolded (x-axis) and the native protein (y-axis), given as change in relative fluorescent unit (ΔRFU s^{-1}). The refolding experiments over eight generations are depicted as dots, with datasets of one individual generation sharing the same color and being referred to as Gen 1-8 in the plot legend. All experiments were performed using identical incubation and activity measurement conditions (see part 8.6.8.3). For the individual datasets see appendix (part 6.2.7)

A further analysis of the datasets revealed the success of lysozyme reactivation being strongly dependent on the ionic strength of the buffer environment. This observation is also consistent with previous reports, showing that the activity of lysozyme is strongly reduced in the presence of high ionic strength (Davies, et al. 1969). In the third approach, the parameter space was therefore modified concerning three limitations; only solutions providing the following features were allowed:

I. a redox system containing GSSG
II. the ionic strength being minimum 0.2 M and maximum 0.7 M
III. a pH-value higher than 7.0

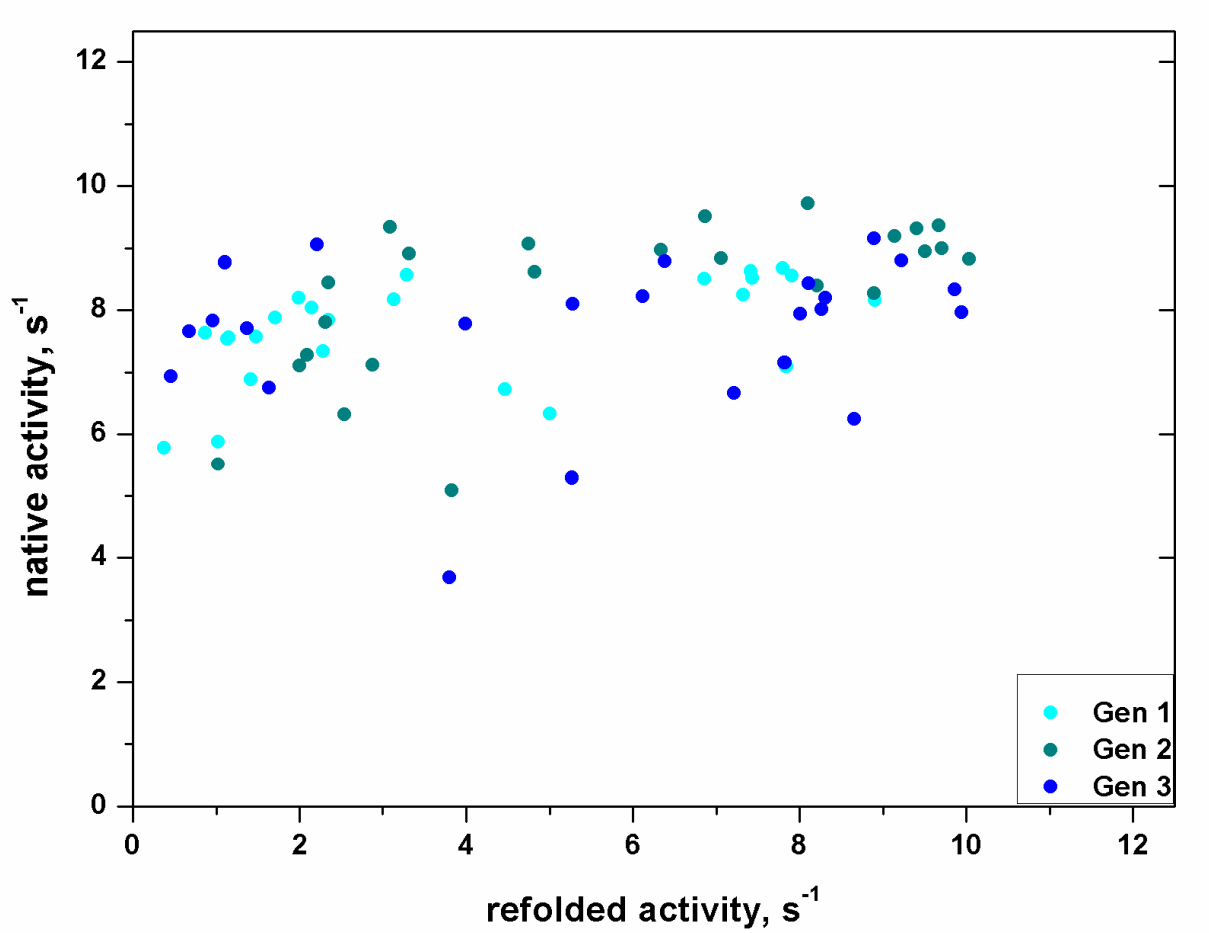


Figure 14 Third approach of lysozyme refolding optimization

Efficiency evaluation of refolding solutions for lysozyme by activity determination for the refolded (x-axis) and the native protein (y-axis), given as change in relative fluorescent unit (ΔRFU s^{-1}). The refolding experiments over the three additional generations are depicted as dots, with datasets of one individual generation sharing the same color and being referred to as Gen 1-3 in the plot legend. All experiments were performed using identical incubation and activity measurement conditions (see part 8.6.8.3). For the individual datasets see appendix (part 6.2.7)

Applying those changes, high refolding yields could be achieved already in three new generations (see Figure 14). With several buffers providing success in both quality features, the optimization was terminated at that point (see Table 4 for features of best buffer conditions).

Table 4 efficient buffer compositions of the Lysozyme refold optimization process

Buffer ID	Buffer composition	Efficiency
Gen 2 / 8	40 mM Tris/HCL, pH 7.5, 25 mM KCl, 0.2 mM PEG-4000, 25 mM glycine, 5 mM EDTA, 1 mM GSH, 0.5 mM GSSG	9.4 ± 0.24 ΔRFU s^{-1} native 9.7 ± 0.25 ΔRFU s^{-1} refolded
Gen 2 / 19	40 mM Hepes, pH 7.5, 25 mM NaCl, 5 % glycerol v/v, 2.5 mM GSH, 0.5 mM GSSG	9.0 ± 0.5 ΔRFU s^{-1} native 10.1 ± 1.5 U mg^{-1} refolded

6.2.4 MAK33 Fab-fragment

To test the applicability of the limited parameter space on a further disulfide-bridged protein, the MAK33 Fab-fragment was subjected to an optimzation screen with the above described adjustments.

MAK33 is a murine monoclonal antibody belonging to the κ/IgG1 subclass, directed against human muscle type creatine kinase (Buckel, et al. 1987). The Fab fragment is a product of papain cleavage of the full-length antibody and comprises the entire light chain together with the two N-terminal domains of the heavy chain, constituting the functional part of the antibody regarding antigen binding. During folding, internal disulfide bridges within the four domains have to be formed, accompanied by correct association of the heterodimers, which are linked covalently by cysteine pairing as well. Due to these features, the renaturation of the antibody Fab fragment, which cannot be expressed solubly in the cytosol of *E.coli* represents a challenging task.

Based on previous studies on the refolding kinetics of Fab (Buchner and Rudolph 1991), which were consistent with our findings, the renaturation samples were incubated for 72 h at 10 °C.

6 generations of optimization were carried out, searching for optimal solutions in terms of

I. High refolding yield (performance of refolded protein compared to native)
II. Low experimental cost

The Fab fragment was denatured using 6 M GdmCl and 5 mM DTT to reduce all cysteine residues. Refolding was carried out by rapid dilution into the respective buffer to a final concentration of 4 µg/ml.

To determine the refolding yields, an indirect ELISA (see part 8.6.8.4) was performed with the native and the refolded protein samples both incubated in the refolding buffers. Native Fab in 100 mM Tris/HCl pH 7.5; 2 mM EDTA buffer was used for a standard curve as an intrinsic control of the ELISA performance. Binding of Fab to biotinylated antigen occurs in streptavidin-coated microplates and can be quantified indirectly via the detection of a colorimetric reaction. 40 ng protein per well were applied to achieve optimal assay performance and refolding yields were calculated as percents of the native Fab incubated in the corresponding refolding buffers.

During the refolding optimization process, solutions providing up to 100 % yield and decreased experimental cost were found experimentally (Figure 15).

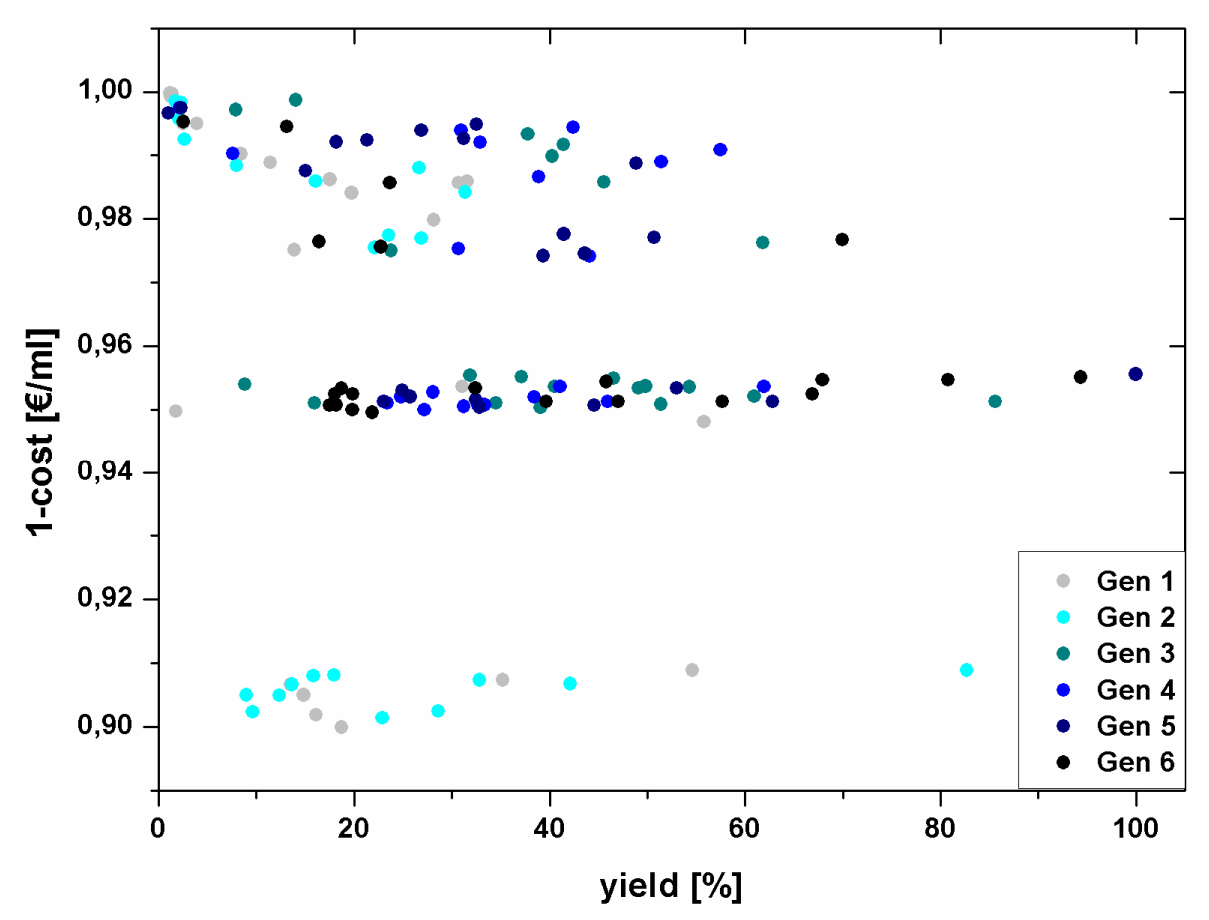


Figure 15 Optimization process of MAK33 Fab fragment refolding

Efficiency evaluation of refolding solutions for the MAK33 Fab fragment by determination of refolding yield (x-axis, given in % of the native protein) and the experimental cost (y-axis). The refolding experiments over six generations are depicted as dots, with datasets of one generation sharing the same color, respectively, and being referred to as Gen 1-6 in the plot legend. Fitness assignments were calculated according to ELISA performance. All experiments were performed using identical incubation and assay conditions (see part 8.6.8.4). For the individual datasets see appendix (part 6.2.7)

Several common features concerning the buffer conditions leading to the highest efficiencies (Table 5) could be observed; the combination of a pH value in the basic range, high glycerol concentrations, and the presence of the oxidative agent GSSG leads in all cases to high yields.

Table 5 Most efficient buffer compositions of the MAK33 Fab refold optimization process

Buffer ID	Buffer composition	Efficiency
Gen 3/ 3	40 mM Hepes, pH 9.0, 25 mM NaCl, 25 mM KCl, 15 % Glycerol, 50 mM glutamine, 5 mM GSH, 5 mM GSSG	86 ± 12 % yield 0.048 € ml^{-1}
Gen 5 / 12	40 mM Tris/HCL, pH 8.75, 100 mM NaCl, 25 mM KCl, 15% glycerol, 25 mM glycin, 7,5 mM EDTA, 5 mM GSSG	100 ± 10 % yield 0.044 € ml^{-1}
Gen 6 / 4	40 mM Hepes, pH 9.0, 25 mM NaCl, 25 mM KCl, 15 % glycerol, 25 mM L-Arg, 25 mM glycine, 0.5 mM GSH, 5 mM GSSG	81 ± 18 % yield 0.045 € ml^{-1}
Gen 6 / 12	40 mM Hepes, pH8.5, 75 mM NaCl, 50 mM KCl, 15 % glycerol, 0.5 mM GSH, 5 mM GSSG	94 ± 0.8 % yield 0.045 € ml^{-1}

To compare these results with the direct refolding of inclusion bodies, the MAK33 light chain and the Fd fragment were recombinantly expressed as insoluble aggregates in *E.coli*. The purification of the inclusion bodies is described in detail in part 8.6.2. Shortly, remaining cell components were removed by several wash- and centrifugation steps in the presence of a detergent, solubilized in the presence of 6 M GdmCl and 5 mM DTT, followed by a size-exclusion-chromatography run to separate undesired proteins. Refolding followed the same procedure as described above and was tested with four different buffers from generation 5 and 6.

In the entire chosen buffer conditions, no significant difference between refolding of purified inclusion bodies and denatured native Fab fragments could be observed (see Figure 16).

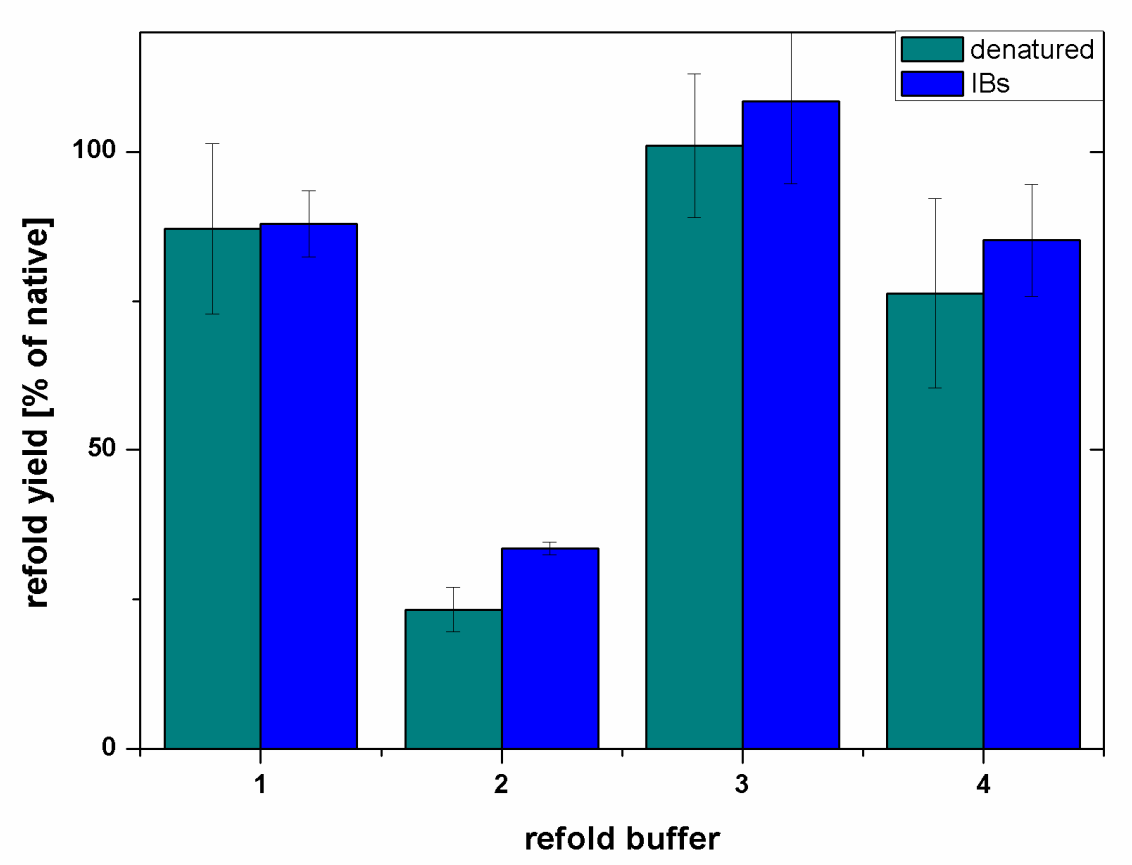


Figure 16 Refolding of inclusion bodies vs refolding of denatured native Fab

Yield comparison after refolding solubilized inclusion bodies (IB's) or denatured native protein in four different buffers: **1**) 40 mM Tris/HCl, pH 8.75, 0.1 M NaCl, 25 mM KCl, 15% glycerol, 25 mM glycine, 7.5 mM EDTA, 5 mM GSSG. **2**) 20 mM Tris/HCl, pH 9.0, 0.1 M KCl, 50 mM glycine, 5 mM GSH, 5 mM GSSG. **3**) 40 mM Hepes, pH 8.5, 75 mM NaCl, 50 mM KCl, 15% glycerol, 0.5 mM GSH, 5 mM GSSG. **4**) 40 mM Tris/HCl, pH 8.75, 0.1 M NaCl, 12.5% glycerol, 50 mM glutamine, 7.5 mM EDTA, 5 mM GSSG. All experiments were refolded and measured in triplets at 10°C and 30°C, respectively.

6.2.5 ERdJ3

We further wanted to investigate whether it is possible to find suitable refold conditions for the mammalian cochaperone ERdJ3 with the help of the algorithm-based search. This endoplasmic reticulum – resident J protein is organized in four domains and contains two intramolecular, non-sequential disulfide bridges. It regulates the functional cycle of BiP, the Hsp70 of the endoplasmic reticulum and is involved e.g. in immunoglobulin folding (Meunier, et al. 2002). Since it could not be recombinantly expressed in a soluble form in *E. coli*, nor be refolded efficiently, *in vitro* studies could only be performed with expensively in small amounts produced protein from eukaryotic cell culture so far.

ERdJ3 has been shown to accelerate the ATPase activity of the chaperone BiP (Shen and Hendershot 2005) *in vitro* up to 2.5 fold. The functional assay for refolding success was based on those findings, setting the BiP ATPase activity in the respective buffer environment as reference for the function of ERdJ3. BiP was purified for this purpose as described in 8.6.3, and applied in the assay with a ratio of 1 µM BiP to 2 µM ERdJ3, thereby observing the grade of acceleration mediated by endogenous ERdJ3 (Jin, et al. 2008).

Prior to refolding, ERdJ3 was recombinantly expressed as IBs in *E.coli*. These were isolated and further purified in the presence of 6 M GdmCl as described in detail in part 8.6.4. 2mM DTT was used for cysteine reduction. The refolding process was induced by 1:50 dilution into the respective buffers, which corresponded to a final concentration of 12 µM. Native ERdJ3, obtained from refolding in the previous standard buffer (250 mM Tris/HCl pH 8.0; 100 mM L-arginine; 10 mM EDTA; 1 mM GSSG; 0.5 mM GSH) served as a reference in the screen. After renaturation, BiP was added to the wells containing the assay components, and its ATPase activity was monitored upon addition of the respective refolded and native ERdJ3 samples.

For evaluation, the experimental costs as well as the acceleration levels conferred by ERdJ3 were determined (see Figure 17 and Figure 18, respectively).

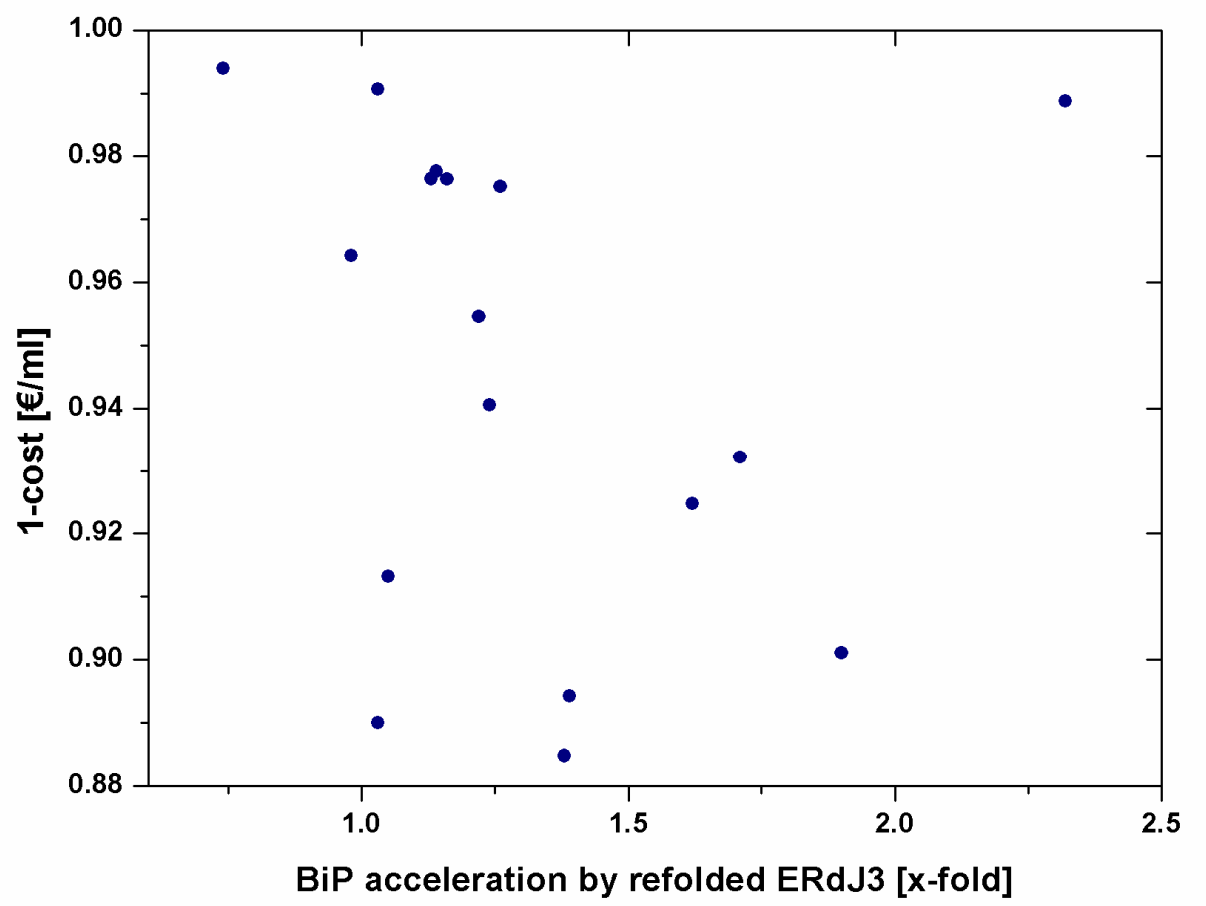


Figure 17 Refolding screen of mammalian ERdJ3 (1)

Evaluation of ERdJ3 refolding, depicted for experimental costs as a function of BiP's ATPase acceleration mediated by refolded ERdJ3. Each dot refers to one refolding experiment of the randomly generated set of solutions. All refolding experiments were performed as triplicates and were incubated and measured under identical conditions with the mean value depicted (see part 8.6.8.5).

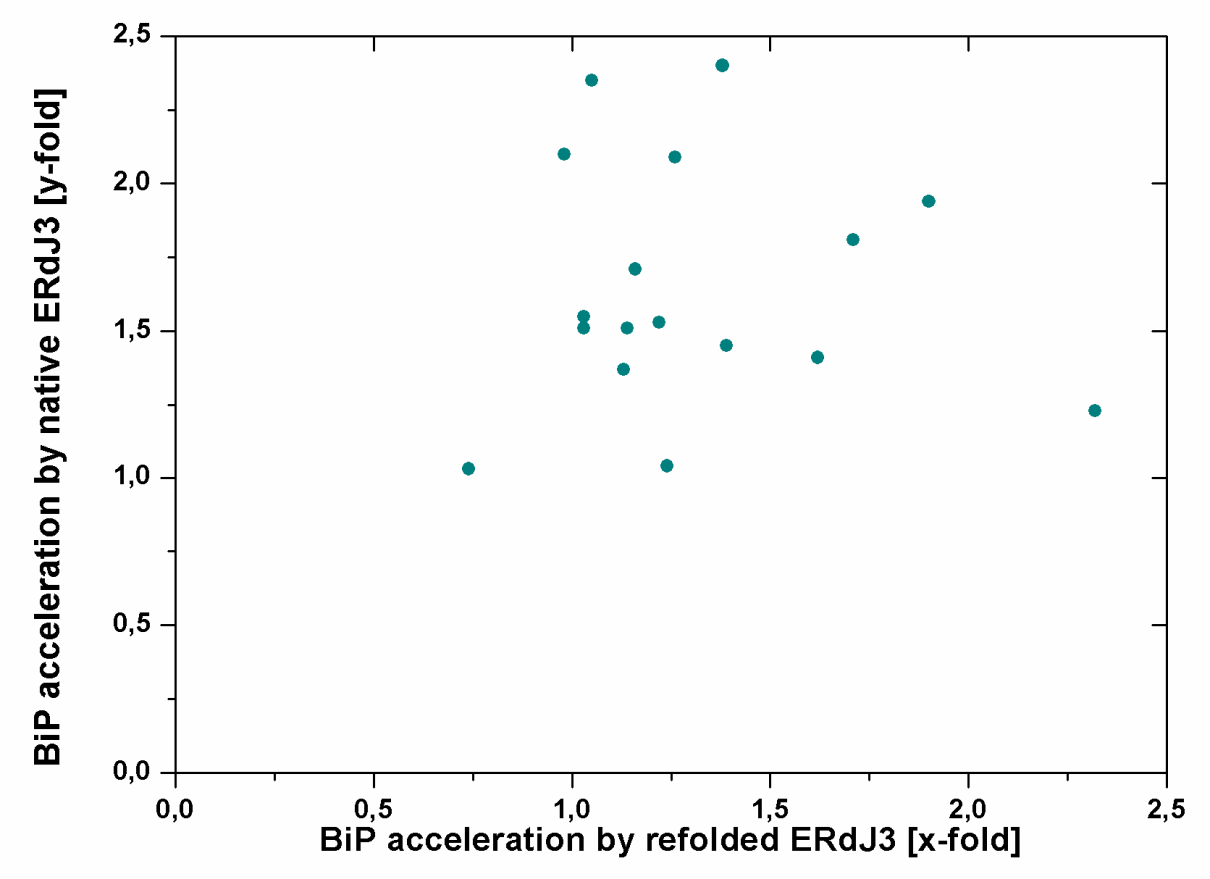


Figure 18 Refolding screen of mammalian ERdJ3 (2)

Evaluation of ERdJ3 refolding, illustrated as BiP's ATPase activity acceleration mediated by native ERdJ3 as function of refolded ERdJ3. Each dot refers to one refold experiment of the randomly generated set of solutions. All refolding experiments were performed as triplicates and were incubated and measured under identical conditions (see part 8.6.8.5).

To obtain realistic efficiency assignments, the performance of the native reference was factored into further evaluation, since a relevant influence of the different buffers in its activity was visible (Figure 18). For volume upscale experiments, solutions yielding lower than 2-fold BiP acceleration mediated by the native EdJ3 reference were therefore sorted out.

Refolding was upscaled to 250 ml, followed by further purification by size exclusion chromatography in 50 mM Hepes/KOH pH 7.5; 150 mM KCl; 10 mM $MgCl_2$ to remove aggregates and change the storage buffer. In this procedure, buffer 9 (40 mM Hepes/NaOH pH 9.0, 0.35 M NaCl, 20 mM KCl, 0.75 M L-arginine, 0.05 mM PEG-4000, 0.06 mM Tween20, 0.5 mM GSH, 5 mM GSSG) provided functional protein, which was utilized for further *in vitro*

studies on ERdJ3 (Marcinowski, et al. 2011). The final yield was ~40% purified refolded protein, which corresponded to the activity previously observed in the screen.

6.2.6 Protease

We additionally decided to examine the capacity of the system regarding a human protease with unknown parameters. The only available information was the molecular weight and the peptide cleavage recognition site.

The starting material were resuspended inclusion bodies, which were solubilized using buffer containing 5 M GdmCl. Renaturation was initialized by 1:100 dilution to 5 µg/ml in the respective refolding buffer.

For the determination of proteolytic activity, a synthetic peptide containing the recognition motif, flanked by a fluorophore and a quencher was used. Fluorescence increase upon peptide cleavage was monitored over 25 minutes at 37°C in a microplate-reader.

The optimization process was started with a limited setup of standard parameters (pH 7-9.5, no detergents and no cofactors), because we were interested if a parameter pool with the most widely used ingredients would be sufficient. It was set to proceed towards the following objectives:

I. low experimental cost
II. increasing enzymatic activity

After four generations with only slowly evolving improvement, the parameter space was expanded to its complete range. After this step, increasing activity slopes could be observed, indicating that the availability of the additional conditions was of importance for the reactivation of this protein (Figure 19).

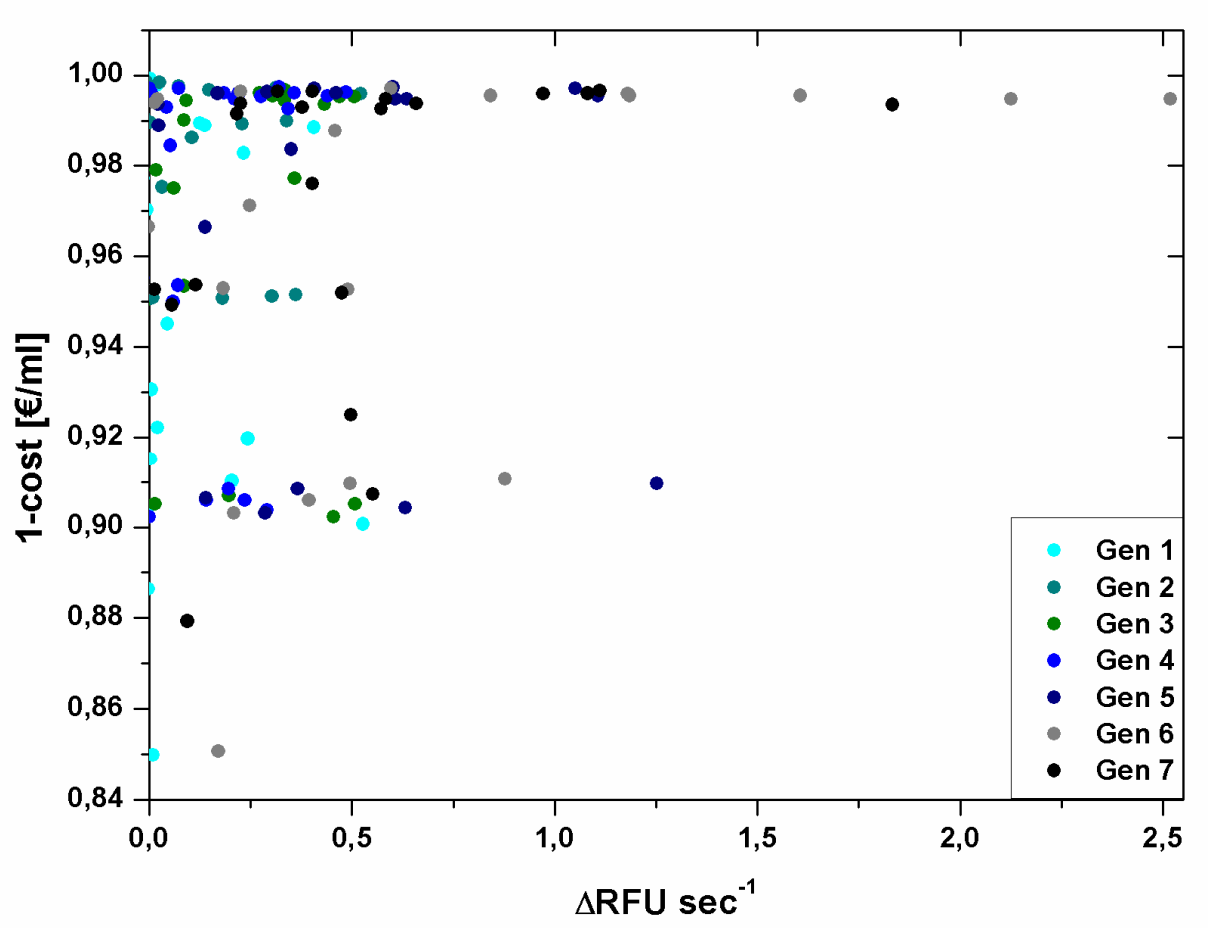


Figure 19 Optimization process of protease refolding

Efficiency evaluation of refolding solutions for the protease by fluorescence monitoring (x-axis) and cost calculation (y-axis). Changes in fluorescence were calculated as the change of relative fluorescent unit (RFU) per second. The refolding experiments over seven generations are depicted as dots, with datasets of one generation sharing the same color and being referred to as Gen 1-7 in the plot legend. All experiments were performed applying identical incubation and fluorescence measurement conditions (see 8.6.8.6). For the individual datasets see appendix (part 6.2.7)

Table 6 gives an overview over the buffer conditions that performed best within the seven generations we carried out. All of them share some common features, such as the presence of metal ions, non-ionic detergent within the CMC (critical micellar concentration), 1mM GSH, and amino acids.

Table 6 Most efficient buffer compositions of the protease refolding optimization process

Buffer ID	Buffer composition	Efficiency
Gen 6 / 15	40 mM Hepes pH 9.75, 0.1 M NaCl, 0.1 % PEG-4000, 0.025 M L-arginine, 0.05 M glutamate, 0,06 mM Cu/Zn/Mg/Mn, 0.8 mM Triton X-100, 1 mM GSH	1.6 ±0.01 ΔRFU sec^{-1} 0.045 € ml^{-1}
Gen 7 / 3	40 mM MOPS pH 8.25, 0.1 M NaCl, 0.05 M KCl, 0.1 M L-arginine, 0.05 M glycine, 0.04 mM Cu/Zn/Mg/Mn, 0.09 mM Brij, 1 mM GSH	1.8 ±0.18 ΔRFU sec^{-1} 0.065 € ml^{-1}
Gen 6 / 16	50 mM Tris/HCl pH 9.25, 5 % Glycerol, 0.1 M L-arginine, 0.05 M glycine, 0.06 mM Cu/Zn/Mg/Mn, 0.6 mM Triton X-100, 1 mM GSH	2.1 ±0.2 ΔRFU sec^{-1} 0.052 € ml^{-1}
Gen 6 / 8	50 mM Tris/HCL pH 8.25, 5 % glycerol, 0.1 M L-arginine, 0.05 M glycine, 0.04 mM Cu/Zn/Mg/Mn, 0.09 mM Brij, 1 mM GSH	2.5 ±0.28 ΔRFU sec^{-1} 0.051 € ml^{-1}

Volume upscale experiments were performed in buffer 6/8, all of them yielding slightly lower activities in the fluorescence assay, in the range of ~80 % compared to refolding in 1 ml (Figure 20).

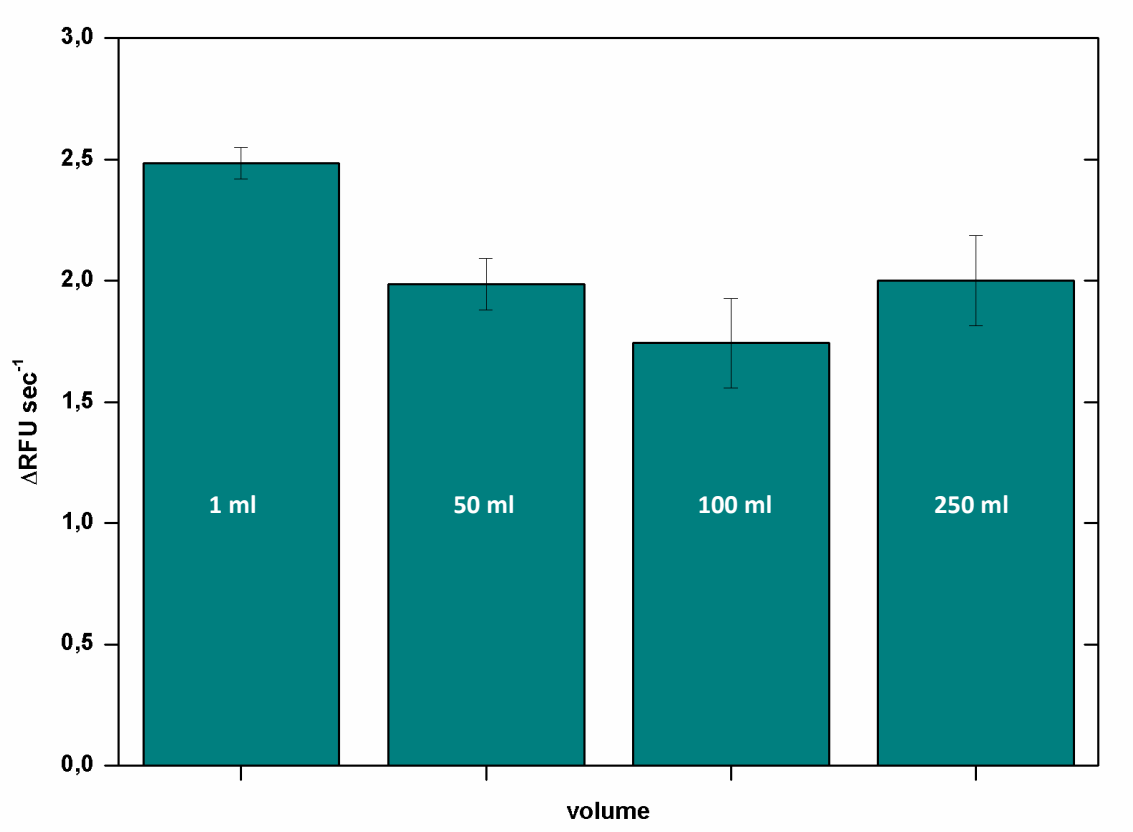


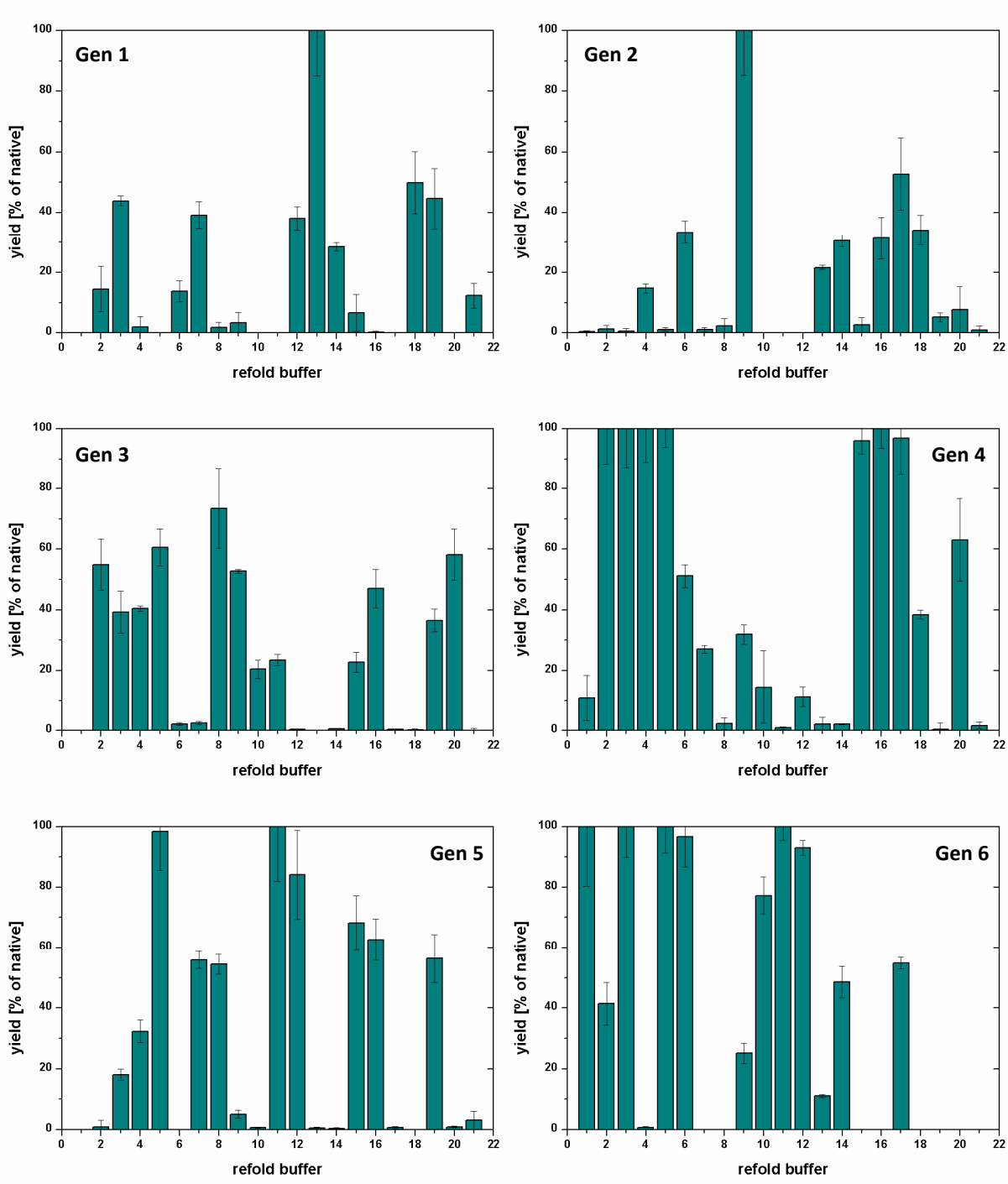
Figure 20 Refolding volume upscale experiments with protease

Protease refolding was performed in different renaturation sample volumes (1, 50, 100, and 250 ml, x-axis) and the samples subjected to the peptide cleavage assay under same conditions as in the optimization screen.

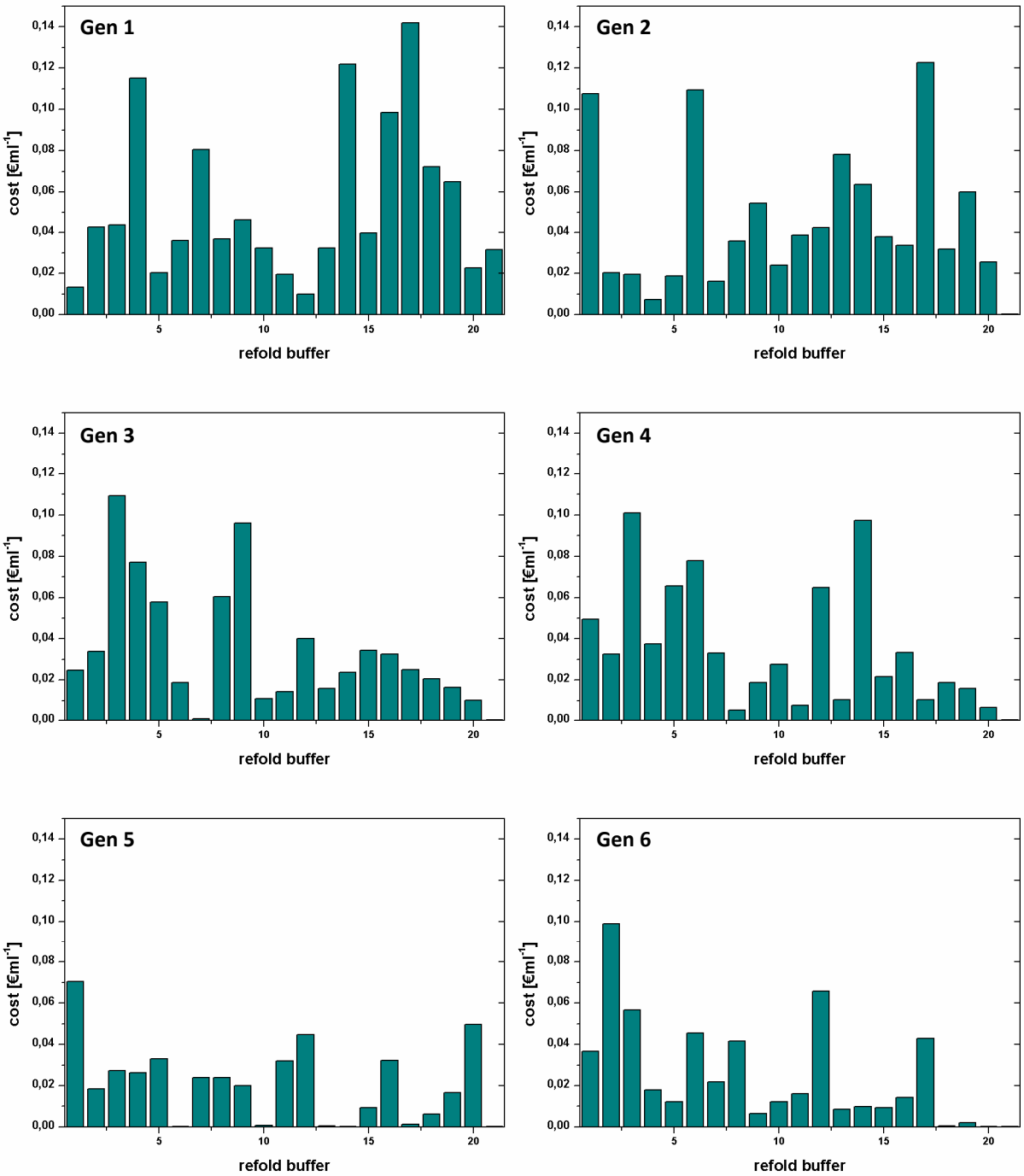
6.2.7 Appendix: Optimization progress evaluation for the individual objectives

6.2.7.1 GFP

- **Yield**

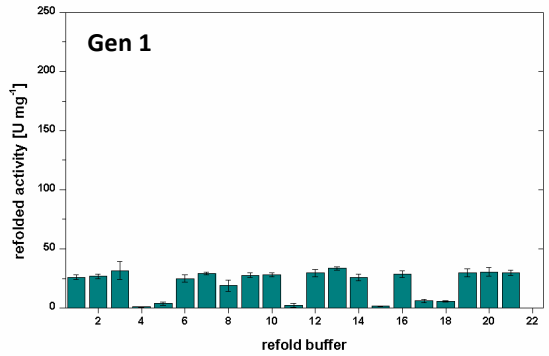


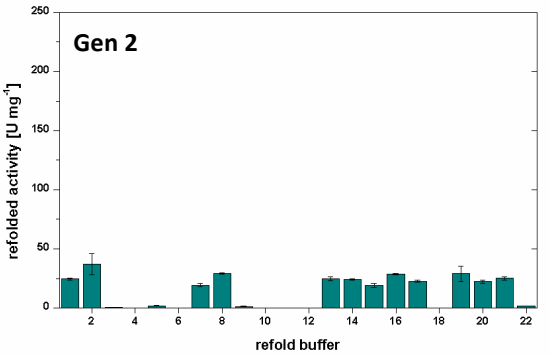
- **Cost**

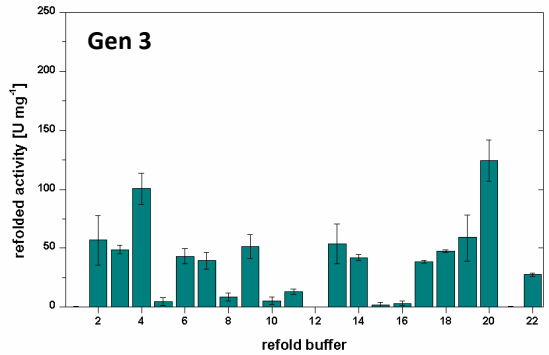

Gen 1
Gen 2
Gen 3
Gen 4
Gen 5
Gen 6
cost [€ml-1]
refold buffer
0,00
0,02
0,04
0,06
0,08
0,10
0,12
0,14
5
10
15
20

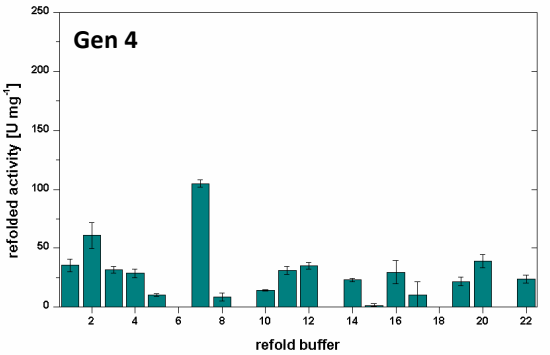
6.2.7.2 Glucokinase

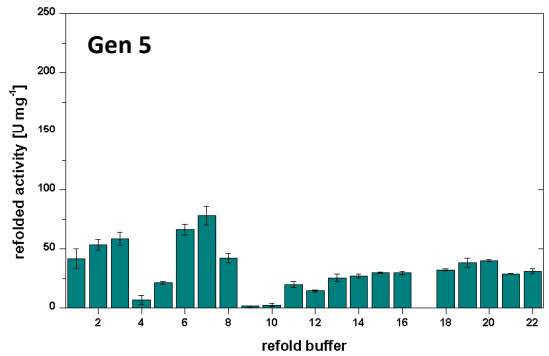
- **Refolded activity**

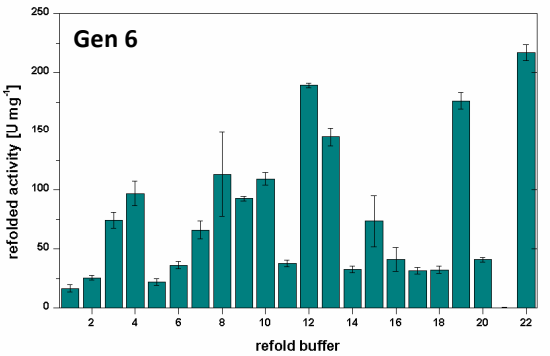


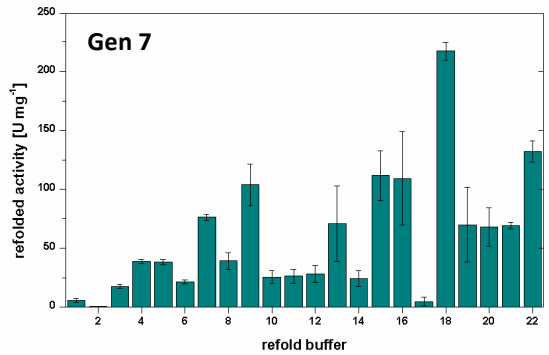




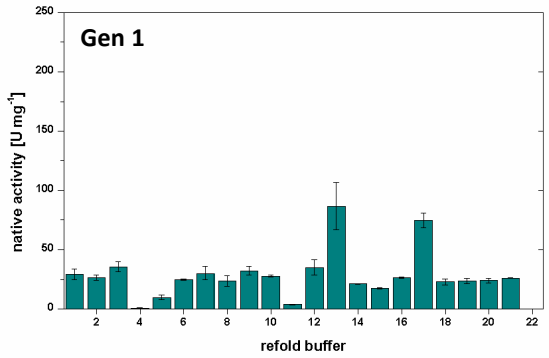


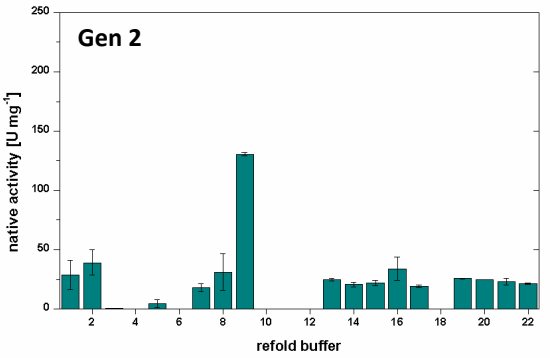


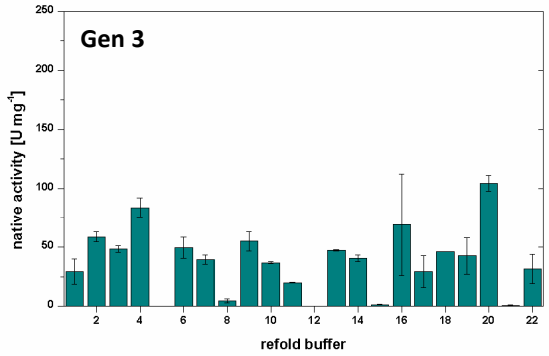


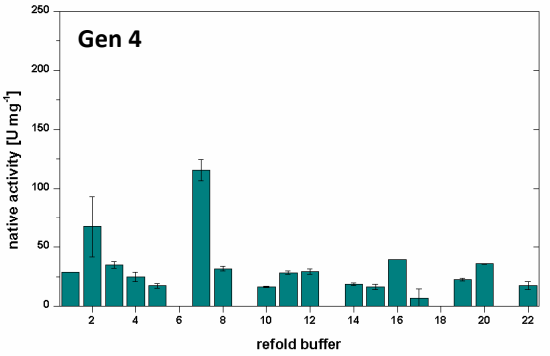


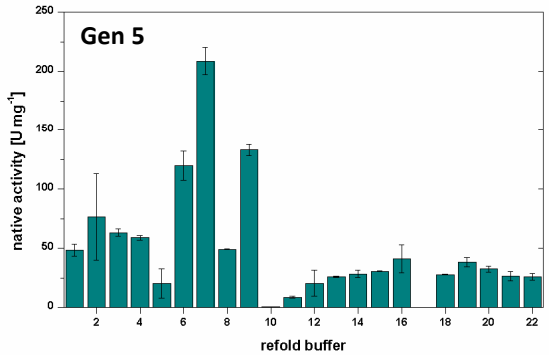
- **Native activity**

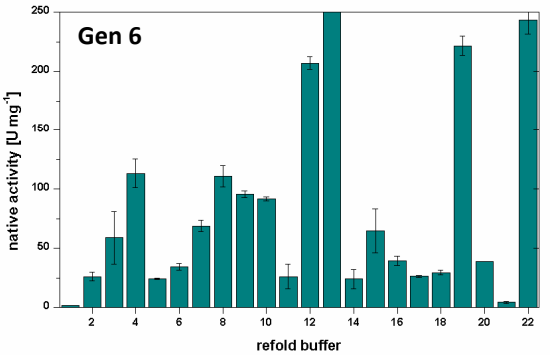


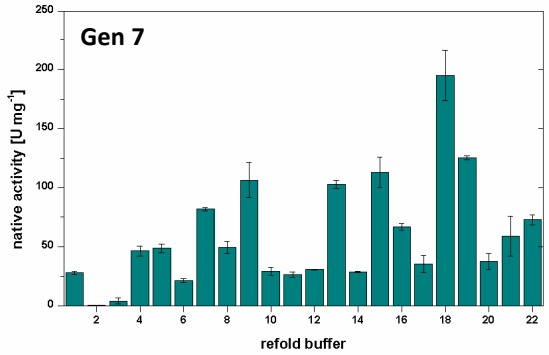






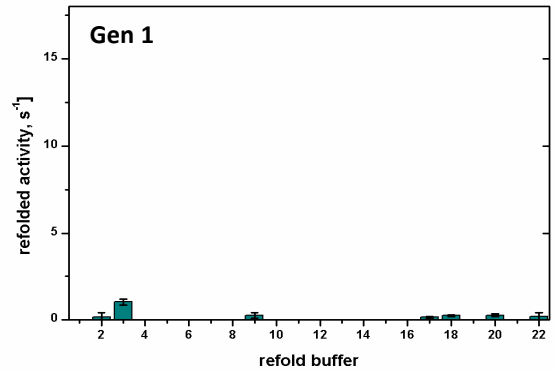


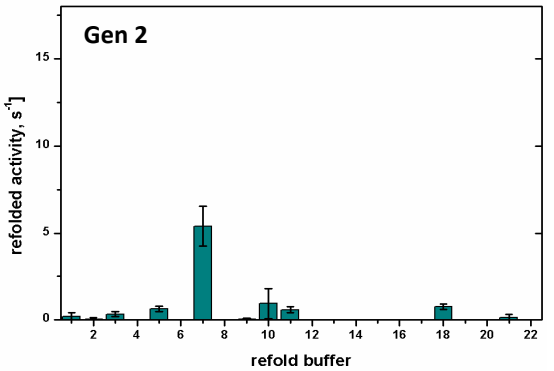


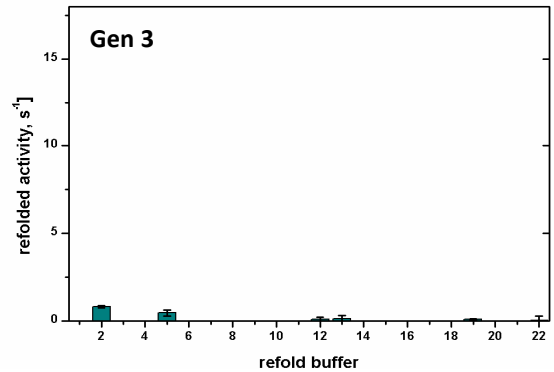


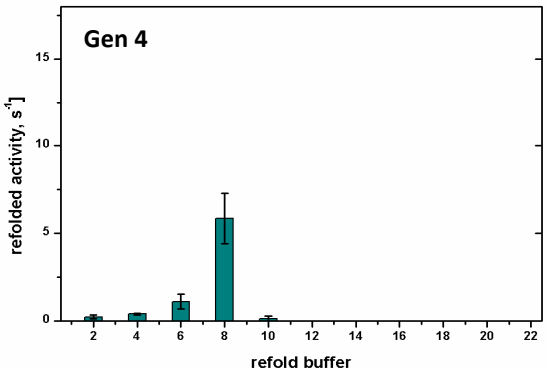
6.2.7.3 Lysozyme

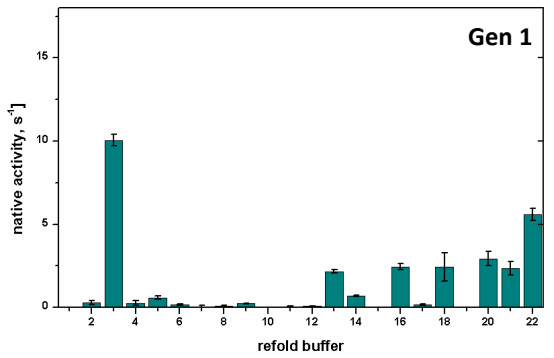
- **Approach 1, refolded and native activity**

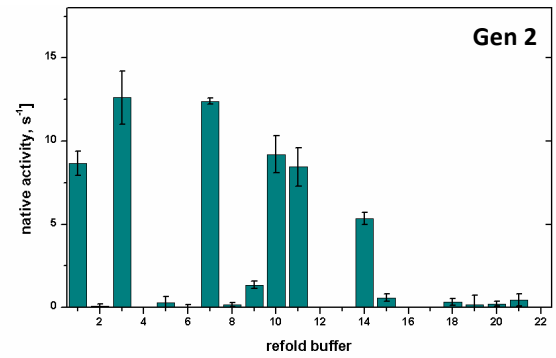


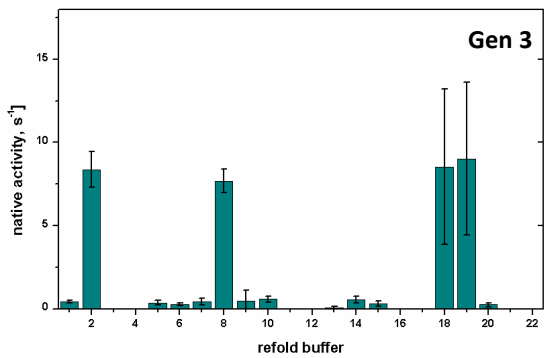


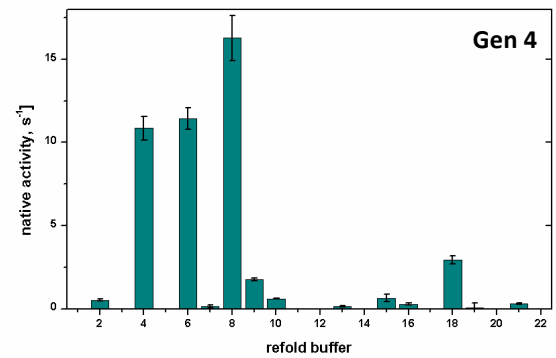




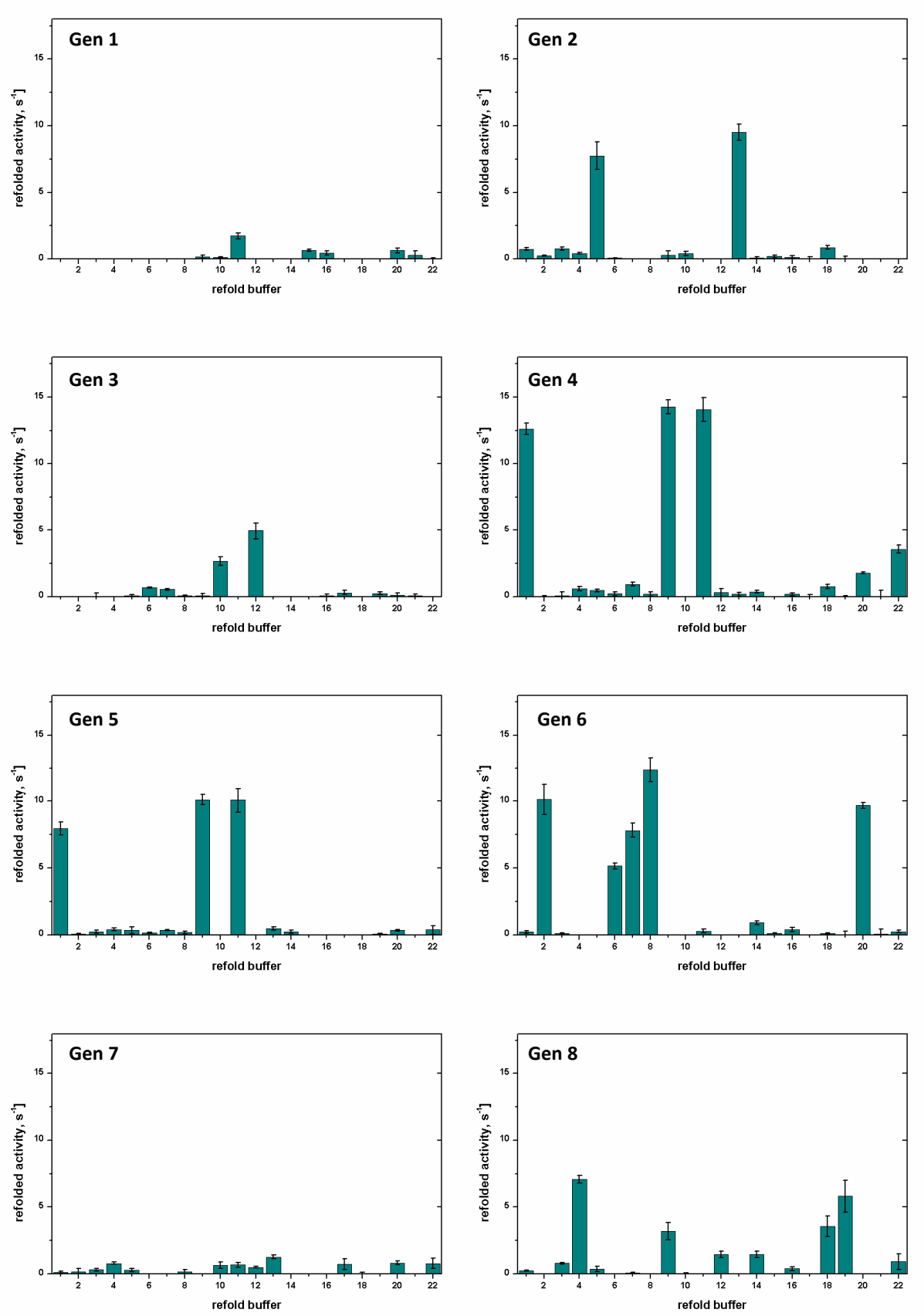




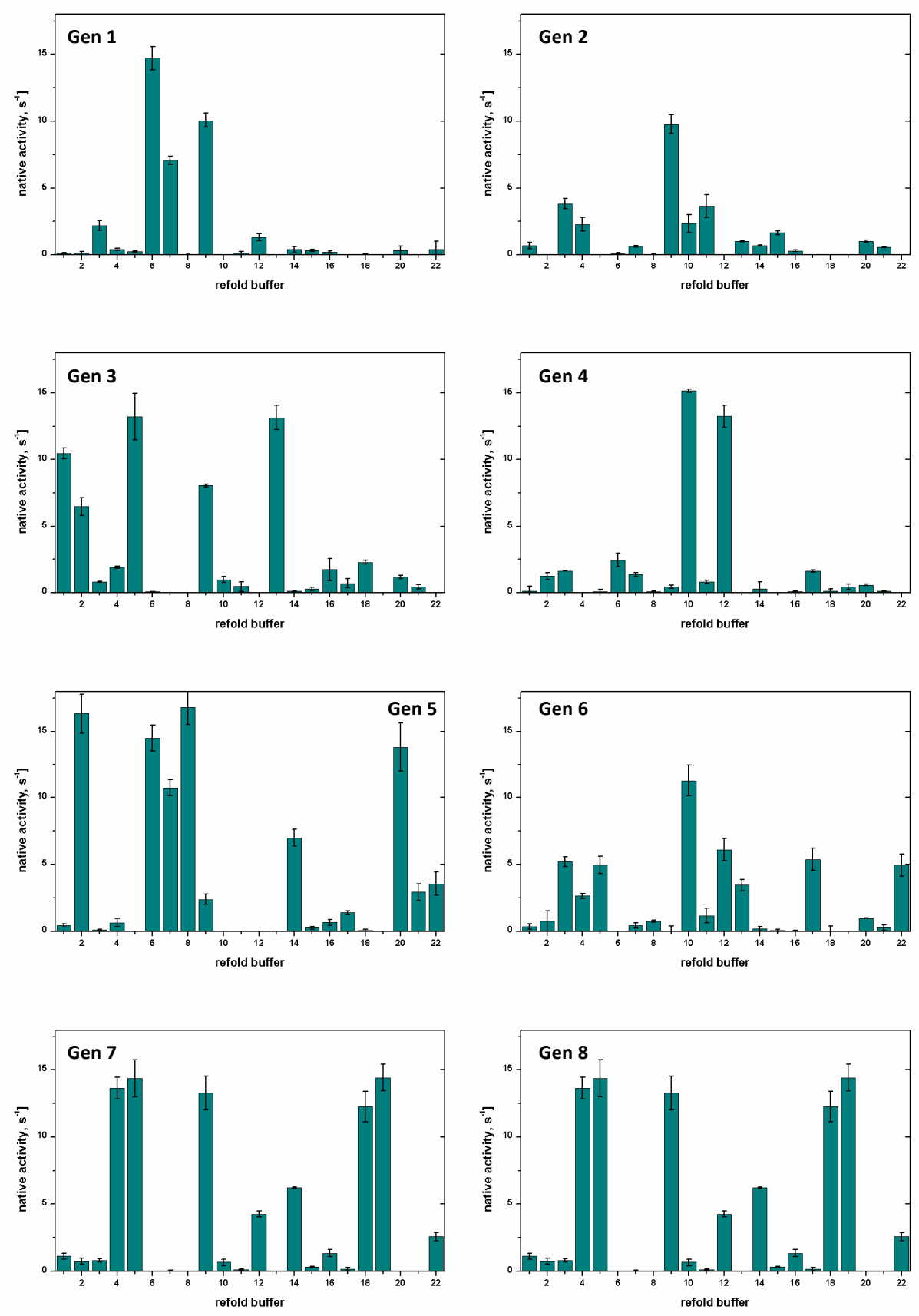




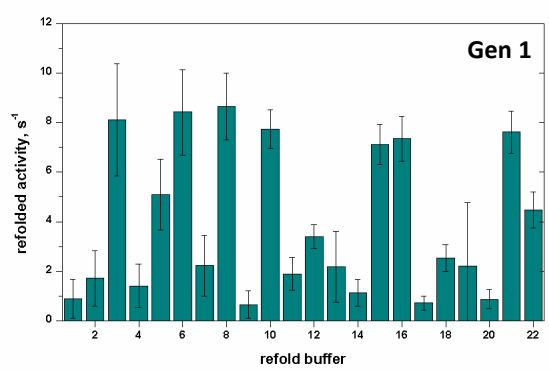
- **Approach 2, refolded activity**

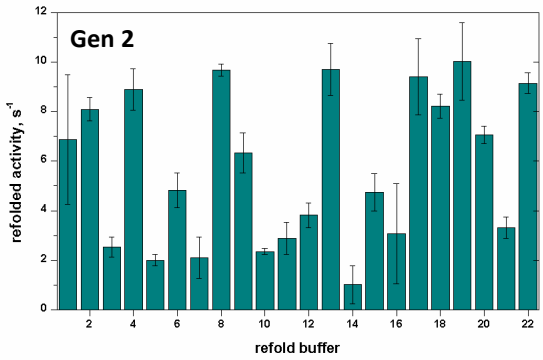


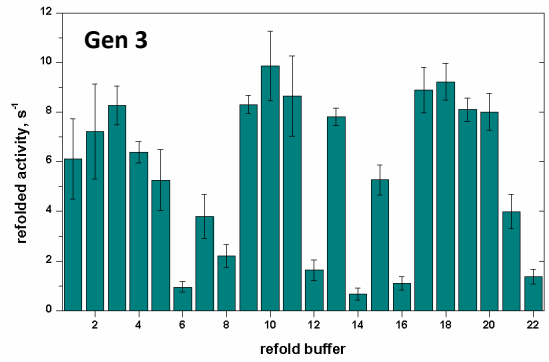
- **Approach 2, native activity**

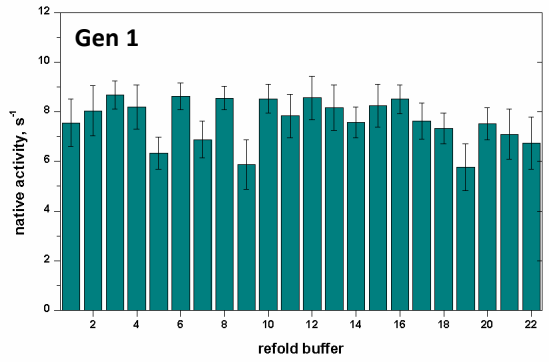


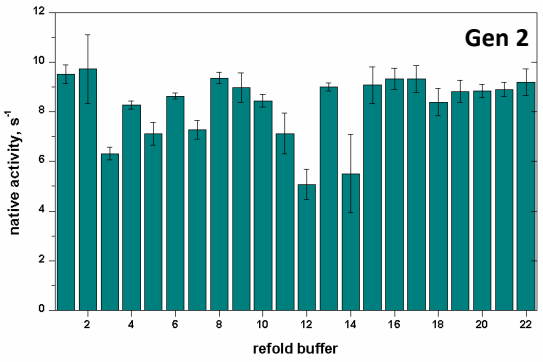
- **Approach 3, refolded and native activity**

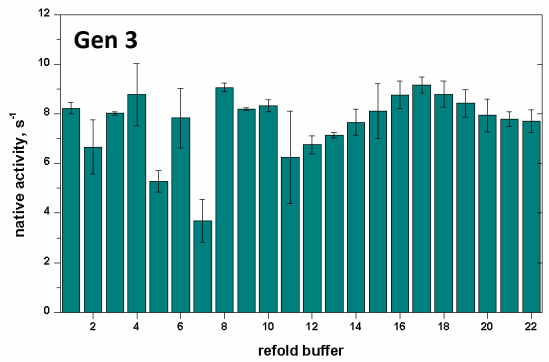






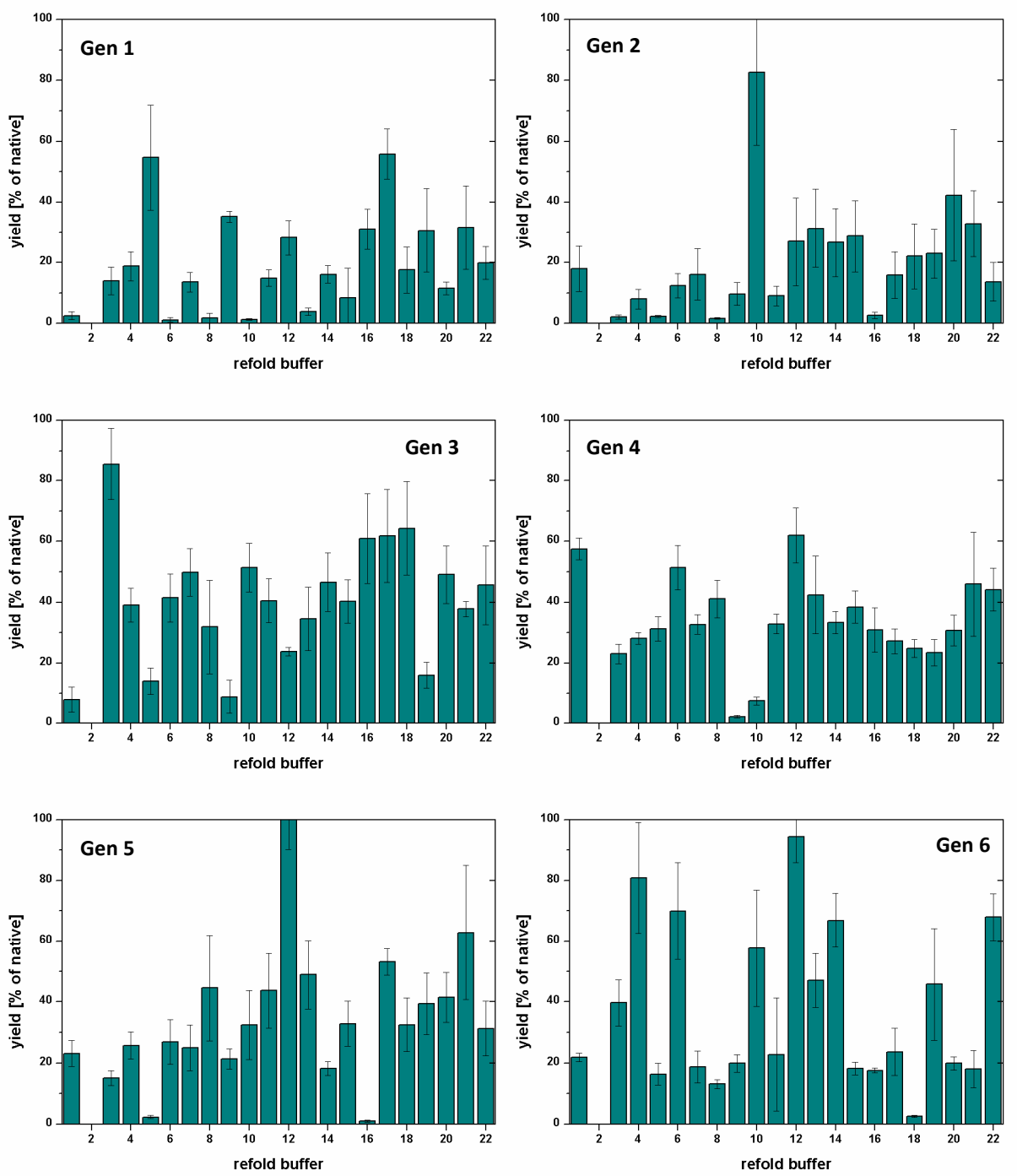




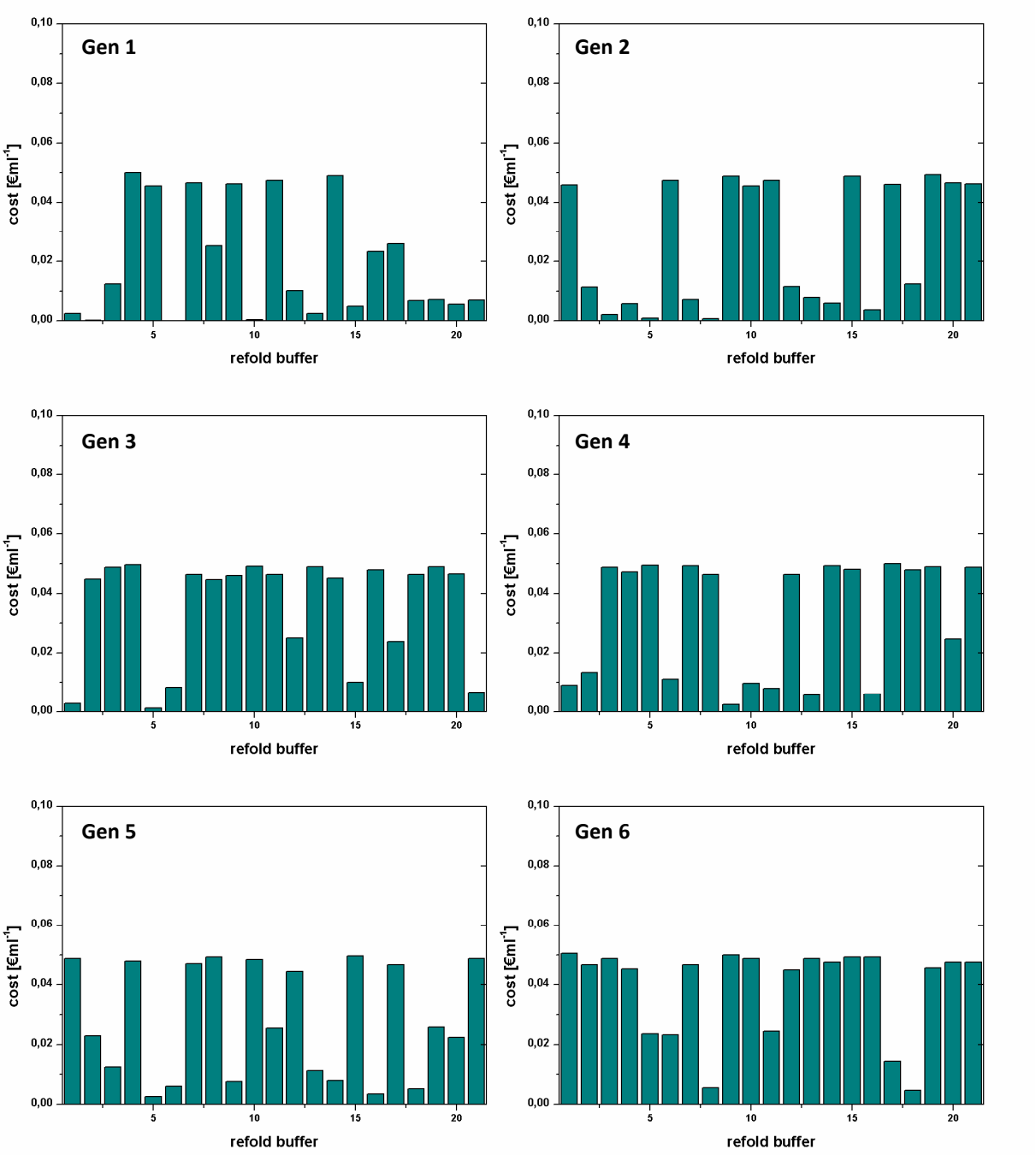


6.2.7.4 Fab fragment

- **Yield**

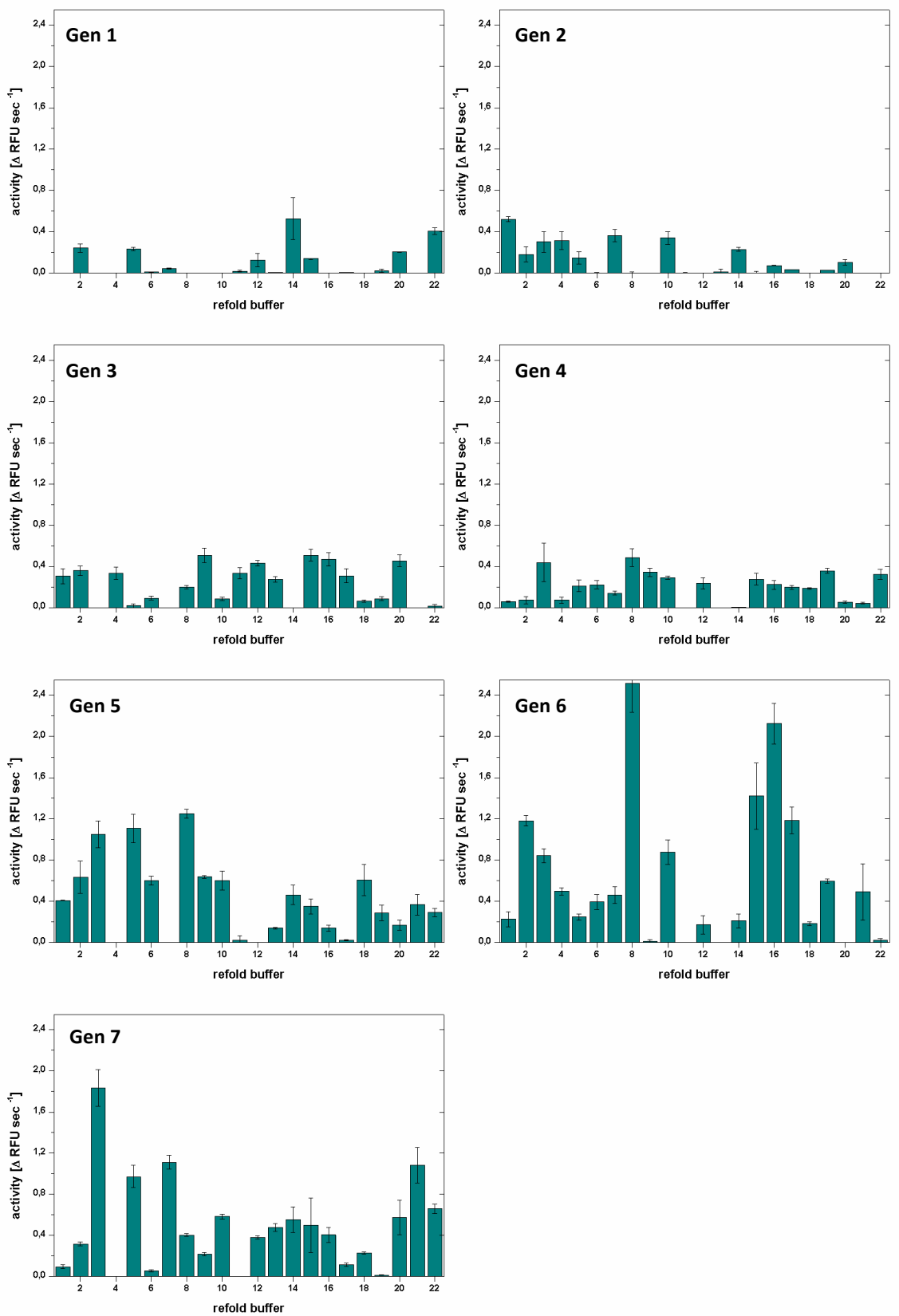


- **Cost**

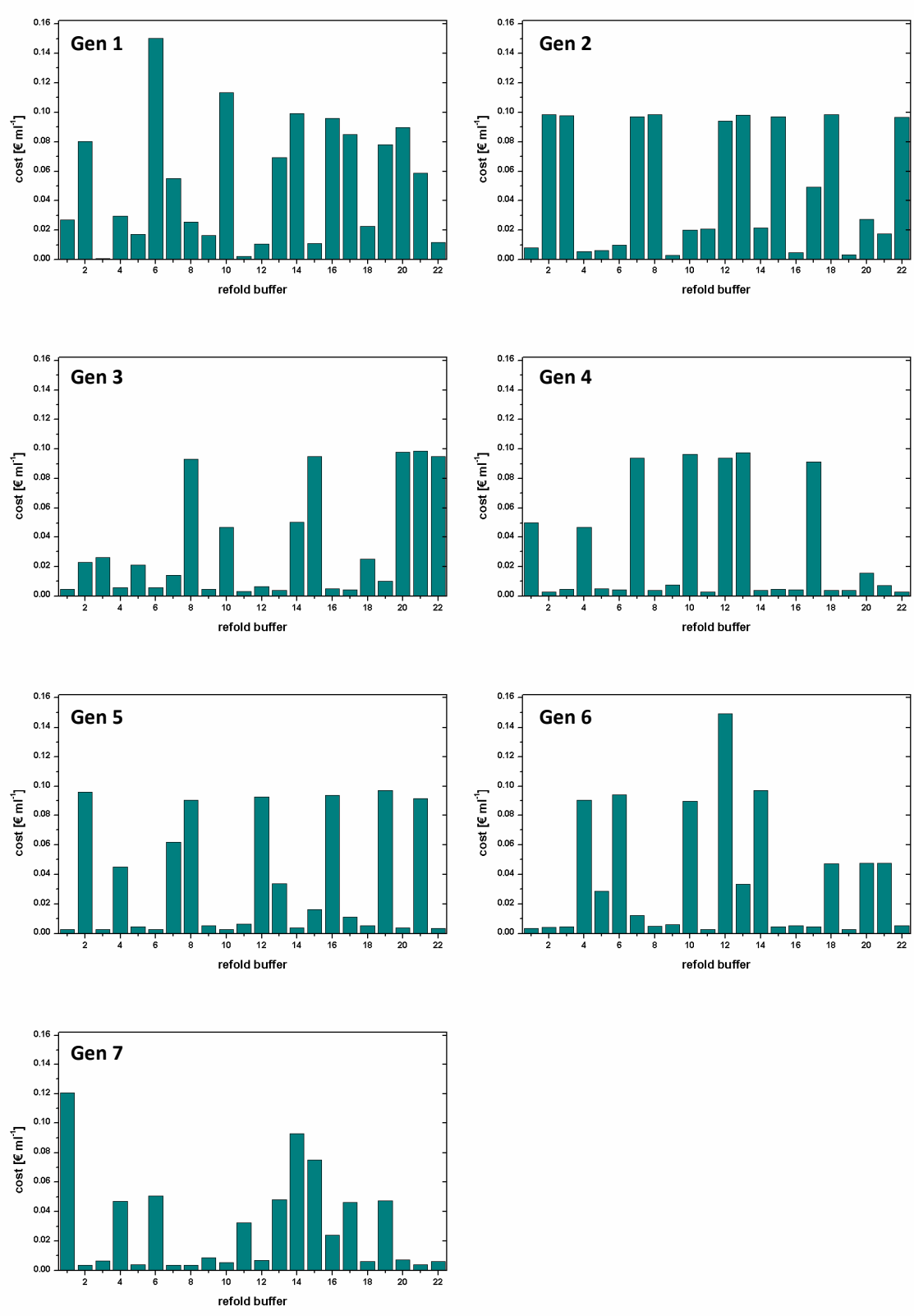

Gen 1
Gen 2
Gen 3
Gen 4
Gen 5
Gen 6
cost [€ml-1]
refold buffer
0,00
0,02
0,04
0,06
0,08
0,10
5
10
15
20

6.2.7.5 Protease

- **Refolded activity**



- **Cost**



7 Discussion

7.1 Feasibility of a genetic algorithm

This thesis presents a new strategy to refold and reactivate proteins from inclusion bodies. In a rational approach, a genetic algorithm was utilized to search for suitable refolding buffer conditions. This method allows not only the identification of solutions but also their further optimization. In general, the concept of the genetic algorithm is suitable for a wide range of refolding strategies. In this thesis, refolding experiments were carried out by rapid dilution allowing an easy to handle setup in a high throughput format in 96 well plates, yet having certain disadvantages in comparison to other setups used such as refolding on a column where e.g. higher protein concentrations are achievable due to separation effects on the matrix (Dashivets, et al. 2009). However, buffer limitations of the refolding on the column approach due to interactions with the matrix are overcome in solution and therefore this refolding technique was chosen.

Genetic algorithms are in general a powerful tool to solve and optimize problems with the following properties:

i. There is no detailed knowledge about the correlations between the contributing parameters and the attained result
ii. The search space is too large and complex to perform a linear search
iii. Several objectives are desired to be optimized in parallel

They therefore should be very well suited for the complex problem of protein renaturation, where numerous compounds contribute to the result. Though not providing a guarantee to reach the global optimum within acceptable time, they offer the chance to efficiently analyze the search space with low experimental effort.

In order to define the parameter space, extensive analysis of available data from the REFOLD database and the literature was performed (Chow, et al. 2006a; Chow, et al. 2006b; Phan, et al. 2011). In general, the most commonly used and most promising substances and their reasonable concentrations were integrated in the parameter space. Our setup allowed experiments to be carried out in the widely used 96-well format with a defined and limited number of experiments per generation. Therefore, the increments of the variables of the search space had to be adapted to the number of experiments per generation to guarantee the exploration of the complete parameter space. In case of a desired integration of additional e.g. more 'exotic' compounds and variables, the experimental setup could readily be adapted. We could also show that adjustments of the parameter space during an optimization run can efficiently accelerate an improvement of the results.

7.2 Investigated proteins

Several proteins were subjected to the established screening strategy. Their differences in size, structure and function can be extracted from Table 7.

In the first two approaches, GFP and Glucokinase were investigated and their renaturation was successfully optimized towards the chosen objectives. In the case of GFP this was the refolding yield and the experimental cost. Different refolding conditions with relative yields of 100 % were obtained. Additionally, the cost optimization proceeded towards a minimum during the performed 6 generations (part 6.2.1). A previously published refolding condition yielding 100 % still provides lower cost (Dashivets, et al. 2009), to further continue the optimization in this direction could therefore have been even more efficient regarding this point.

Table 7 Overview over the properties of the analyzed proteins

Protein	GFP	Glucokinase	Lysozyme	MAK33 Fab	ERdJ3	Protease
Function	fluorescent protein	phosphorylation of glucose	peptidoglycan cleavage	antibody fragment	Co-chaperone	peptide cleavage
M [kDa]	28	35	14	45	38	n.a.
Origin	*Aequorea Victoria*	*Escherichia coli*	*Gallus gallus*	*Mus musculus*	*Mus musculus*	*Homo sapiens*
Quaternary structure	Monomer	Dimer	Monomer	Heterodimer	Monomer / Dimer	n. a.
Secondary structure	β-barrel	α/β	α/β	β-sandwich	α/β	n. a.
Disulfides	-	-	4	5	2	n. a.
Theoretical pI	5.7	6.1	9.3	7.7	5.9	n.a.

During glucokinase refolding, we observed that the buffer environment can strongly affect the activities of the native protein. For the identical native protein batch, differently composed buffers led to changes in activity. This can lead to erroneous results if yields are calculated. To make use of this effect, we decided to optimize the activity of the refolded and the activity of the native protein instead of yield and cost. Interestingly, this even led to a pronounced improvement of the previously described activity level for this enzyme (Meyer, et al. 1997), thereby revealing an additional advantage of the algorithm-based system. The possibility to improve and optimize the buffer environment of native proteins for storage, formulation and activity can be a valuable tool for industrial as well as academic purposes. These demands were shown to be easy accessible by the method presented in this thesis. Additionally, modifications in the parameter space are possible to match with the needs and limitations of such buffer compositions.

Lysozyme activity also was observed to be influenced by buffer effects and was therefore optimized towards the new objectives as well. A reactivation of this disulfide-bridged protein turned out to be more complicated compared to previous cases. The adjustments that were applied in the second and third approach led to efficient results, although the 4 generations comprising first approach yielded very low success. All redox agents are clustered in the same subgroup, which contains more reducing than oxidizing compounds. The algorithm normally favors reducing conditions – which are of course undesirable when disulfide bridges have to be formed. But the clearly visible trend toward oxidizing conditions allowed us constraining the setup in this point. The further adaptations in response to the observed trends towards limiting the range of the ionic strength and the pH value finally led to satisfying results. Thus, even in cases with a high number of inappropriate starting buffer conditions, the algorithm is able to select appropriate conditions and enables the educated user to recognize a trend. The adaptable setup allows responding to the observations by conducting the respective adjustments. As for the glucokinase, both, the activity of refolded and native lysozyme were improved through the optimization. In the third approach, the overall activity level was lower compared to the previous screens, which was most probably because a different protein batch was used for those experiments. Nevertheless, the observed activities were in good agreement with the literature references (Hevehan and De Bernardez Clark 1997).

For the refolding of the MAK33 Fab fragment, the limited setup was applied as well to facilitate disulfide formation. Buffer effects on the native protein were not observed, we could therefore optimize yield and cost. The result was a successful optimization with high relative yields up to 100 % and minimized experimental cost. In earlier optimization approaches, a maximum of 40 % yield could be reached, in particular by introducing high L-arginine concentrations to the buffer composition (Buchner and Rudolph 1991). L-arginine has been discussed for several years to act as neutral crowding agent (Baynes, et al. 2005) and was recently shown to increase the equilibrium solubility during refolding (Tischer, et al. 2010). In our experiments, the majority of the most efficient renaturation buffers contained L-arginine, L-glutamine or glycine, though at lower concentrations (part 6.2.4), whereas in all of them a high glycerol content (15 %) was present. The compound has been widely shown to prevent protein aggregation as well (Vagenende, et al. 2009). For the optimization process, native MAK33 Fab was denatured and refolded. This allowed determining the

relative yields derived from the identical protein batch. However, the antibody fragment was also refolded from purified inclusion bodies in a selection of the generated buffers. It could be shown, that no significant differences in the obtained yields are present. Inclusion body refolding even resulted in overall slightly higher yields, which probably can be again traced back to the fact that two different protein batches are compared. Importantly, the interchain disulfide bridge was present in the cleaved Fab fragment as well as in the chains purified from IBs. This experiment with an identical yet differently generated and denatured protein demonstrates exemplarily that the results of the algorithm-based screen are transferable. Interestingly, a similar refolding approach with heavy and light chains produced in *E. coli* as IBs termed "inclonals" was recently reported for a complete IgG molecule using similar refolding conditions compared to the optimal buffer composition identified in this screen with the Fab fragment (Hakim and Benhar 2009).

In the case of ERdJ3 and protease screening, a new situation was faced, having only very limited amounts or no native protein at all to verify the results through a comparative assay, and dealing with proteins that are poorly characterized. Still, it was possible with the help of the optimization system to obtain positive results. For ERdJ3, the information from the available literature on its function was used to set up an assay performable in the multiwell-format. For this purpose, its interaction partner BiP was purified and integrated in the assay to monitor ERdj3 activity. This approach shows that also an indirect detection in a combinatory approach is generally possible as a read out for the refolding success of the presented screening method. Already one screening cycle enabled us to find suitable buffer conditions yielding ~40 % reactivation which allowed us to perform studies on a so far poorly characterized protein (Marcinowski, et al. 2011). Importantly, a scale-up from the 96 well-format to 250 ml showed similar refolding yields and allowed a large scale production of ERdJ3.

During the protease reactivation optimization, it again turned out to be valuable to interpret recognizable trends and react by adjusting the setup. The protein activity showed to be dependent on the presence of metal ions and non-ionic detergents and to supply the fully extended instead of a constrained parameter space was far more successful (see part 6.2.6). It was also shown that scale up to higher refold volumes is practicable, with overall yields of ~80 % compared to the 1 ml samples.

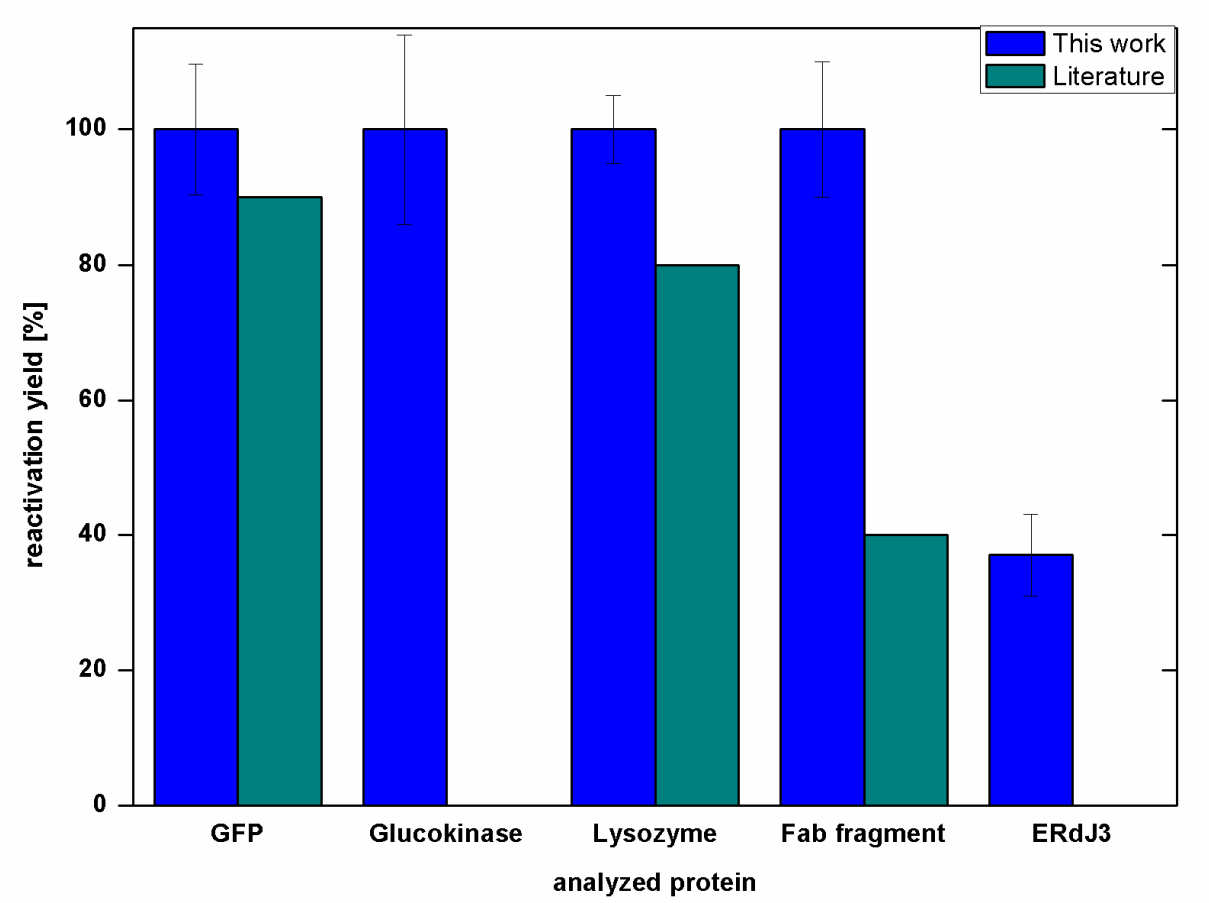


Figure 21 Overview over the reactivation yields obtained in this thesis in comparison with the literature, if available. Relative yields were calculated by activity comparison of the refolded to the native protein.

In general, the most popular database with information on refolding conditions REFOLD (Chow, et al. 2006a; Chow, et al. 2006b; Phan, et al. 2011) does not provide refolding yields, a problem which is overcome by our presented multiparameter optimization approach. In comparison to refolding conditions filed in this database we identified novel ones e.g. for lysozyme which is reported to refold in a buffer composed of 100 mM Tris/HCl pH 8.2, 3 mM GSH, 0.3 mM GSSG, 1 mM EDTA while we could obtain refolding rates of 100 % in a buffer composed of 40 mM Tris/HCl, pH 7.5, 25 mM KCl, 0.2 mM PEG-4000, 25 mM glycine, 5 mM EDTA, 1 mM GSH, 0.5 mM GSSG. Additionally, we could also identify refolding conditions for proteins such as ERdJ3 or glucokinase which were not analyzed before.

Summarizing, we could show for the example of six proteins with different properties and functions that the use of a genetic algorithm is a very well suited strategy in the field of protein renaturation (see Figure 21 for summary of obtained yields and Table 8 for activity data). The presented method allows for a quick and variable search for optimal conditions that provide a suitable environment for proteins to regain their native activity.

Table 8 Summary of analyzed proteins

Analyzed protein

	Experimental cost	
GFP	0.012 € ml^{-1}	
Fab fragment	0.044 € ml^{-1}	
Protease	0.045 € ml^{-1}	
	Activity	**Literature activity**
Glucokinase	Native 243 ± 12 U mg^{-1} Refolded 216 ± 7 U mg^{-1}	Native 158 U mg^{-1}
Lysozyme	Native 9.4 ± 0.24 ΔRFU s^{-1} Refolded 9.7 ± 0.25 ΔRFU s^{-1}	Native 11,2 ± 1,8 ΔRFU s^{-1} Refolded 9,2 ± 1,8 ΔRFU s^{-1}
ERdJ3	Native 2.1-2.4 fold Refolded 1.3-1.4 fold BiP acceleration (ATPase)	Native 2-2.5 fold BiP acceleration (ATPase)

8 Materials and Methods

All methods were adopted from standard protocols from the chair in biotechnology of the TU München if not stated otherwise.

8.1 *E. coli* strains

Strains	Geno- / Phenotype	Source / Reference
E. coli DH10B	*F-araD 139 Δ(ara leu) 7697 ΔlacX74 galU galK mcrA Δ(mrr- hsdRMS-mcrBC) rpsL decR 380 ΔlacZ ΔM15 endA1 nupG recA1*	Berthesda Research Laboratories, Berthesda, USA
E. coli BL21 (DE3)	F– *omp*T hsdS(rB– mB–) *dcm*+ *Tet*r *gal* l (DE3) *end*A *Hte* [*arg*U *ile*Y *leu*W *Cam*R]	Stratagene, La Jolla, USA

8.2 Bacterial vectors

Insert	Vector	Origin
enhanced GFP (eGFP)	pET28a	S. Töpell
MAK33 light chain	pET28a	Eva Seedig
MAK33 Fd	pET23a	Eva Seedig
ERdJ3 WT	pProEx-HT	Moritz Marcinowski

8.3 Chemicals

Acrylamide solution (38% with 2% bisacrylamide)	Roth, Karlsruhe, Germany
Adenosin-5´-diphosphate (ADP), disodium salt	Roche, Mannheim, Germany
Adenosin-5´-triphosphate (ATP), disodium salt	Roche, Mannheim, Germany
Agarose, ultra-pure	Roth, Karlsruhe, Germany
Albumin from bovine serum	Sigma, St. Louis, USA
Ammoniumperoxodisulfate (APS)	Roche, Mannheim, Germany
Ampicillin	Roth, Karlsruhe, Germany
L-Arginine	Sigma, St. Louis, USA
Bacto Peptone	Difco, Detroit, USA
Bacto Tryptone	Difco, Detroit, USA
Bromphenol blue S	Serva, Heidelberg, Germany
Coomassie Brilliant Blue G-250	Serva, Heidelberg, Germany
Dithiothreitol (DTT)	Roth, Karlsruhe, Germany
Ethanol, p.a.	Roth, Karlsruhe, Germany
Ethidiumbromide	Sigma, St. Louis, USA
Ethylen-diamine-tetraacetate (EDTA)	Serva, Heidelberg, Germany
L-Glutamine	Sigma, St. Louis, USA
L-Glutamate	Sigma, St. Louis, USA
Glutathione, oxidized (GSSG)	Sigma, St. Louis, USA
Glutathione, reduced (GSH)	Sigma, St. Louis, USA

Glycerine, 99%	ICN, Irvine, USA
Guanidinium hydrochloride, p.a.	ICN, Irvine, USA
Kanamycin	Roth, Karlsruhe, Germany
2-Mercaptoethanol, pure	Sigma, St. Louis, USA
N-(2-Hydroxyethyl)-piperazine-N'-2-ethansulfonic acid (Hepes)	ICN, Irvine, USA
N,N,N',N'-Tetramethylethylendiamin (TEMED)	Roth, Karlsruhe, Germany
Protease inhibitor mix HP	Serva, Heidelberg, Germany
Sodiumdodecylsulfate (SDS)	Roth, Karlsruhe, Germany
Tris-(hydroxymethyl)-aminomethan (Tris)	ICN, Irvine, USA
Tetracyclin	Roth, Karlsruhe, Germany
Isopropyl ß-D-1-thiogalactopyranoside (IPTG)	Merck, Darmstadt, Germany

All other chemicals were purchased at p.a. grade from Merck (Darmstadt, Germany). For the preparation of buffers double distillated water was used.

8.3.1 Standards and kits

1 kb DNA ladder molecular weight standard	Peqlab, Erlangen, Germany
Low-Range-molecular weight marker (LMW for SDS-PAGE)	BioRad, München, Germany
Wizard Plus Gel Extraction Kit	Promega, Madison, USA
Wizard Plus SV Minipreps DNA Purification System	Promega, Madison, USA

Reagents for ELISA

Blocking reagent	Roche Diagnostics, Mannheim, Germany
ABTS Tablets	Roche Diagnostics, Mannheim, Germany
Buffer for ABTS	Roche Diagnostics, Mannheim, Germany

8.3.2 Chromatography material

HisTrap FF 5 ml	GE Healthcare, Uppsala, Sweden
Superdex 75 Prep Grade	GE Healthcare, Uppsala, Sweden
Superdex 200 Prep Grade	GE Healthcare, Uppsala, Sweden

8.4 Equipment

Balances

Analysis balance BP 121 S	Sartorius, Göttingen, Germany
Halfmicro balance BL 310	Sartorius, Göttingen, Germany

Chromatography equipment

Äkta Explorer	Amersham, Uppsala, Sweden
FP-1520 fluorescence detector	Jasco, Groß-Umstadt, Germany
GradiFrac system	Amersham, Uppsala, Sweden
HighLoad system	Amersham, Uppsala, Sweden
LG-980-02S gradient unit	Jasco, Groß-Umstadt, Germany
PU-1580 HPLC Pump	Jasco, Groß-Umstadt, Germany
Super loop 150 ml	GE Healthcare, Uppsala, Sweden
Super loop 10 ml	GE Healthcare, Uppsala, Sweden
UV-1575 UV-VIS detector	Jasco, Groß-Umstadt, Germany

Centrifuges

Avanti J 25 with JA-10 and JA-25.50 rotors	Beckman, Wien, Austria
Eppendorf table-top centrifuge 5415 C	Eppendorf, Hamburg, Germany
Rotina 46 R centrifuge	Hettich, Tuttlingen, Germany

Cell cultivation and disruption

Air circulation incubator	New Brunswick Scientific, Nürtingen, Germany
Cell disruption machine Basic Z	Constant Systems, Warwick, England
Culture shaker Certomat S	Braun Biotech, Melsungen, Germany

Gelelectrophoresis devices

Hoefer Mighty Small II gelelectrophoresis unit	Amersham, Uppsala, Sweden
RHU10X	Roth, Karlsruhe, Germany

Multiwell-Plates and Cuvettes

Disposable cuvettes, 1,5 ml, halfmicro	Zefa, München, Germany
Silica glass Suprasil cuvettes	Hellma, Jena, Germany
96 Well Microplate, PS, µClear®, Chimney Well	Greiner bio-one, Germany
LUMITRAC™, 96 Well Microplate	Greiner bio-one, Germany
Streptawell, streptavidin-coated Microplate	Roche Diagnostics, Mannheim, Germany

Absorption Spectrophotometers

Biotech Ultrospec 3000 UV-VIS-spectrophotometer	Amersham, Uppsala, Sweden
Cary 100 Bio UV-VIS-spectrophotometer	Varian, Palo Alto, USA
Tecan Sunrise Platereader	Tecan, Crailsheim, Germany

Circular dichroism spectropolarimeter

J-715 spectropolarimeter with PTC 343 peltier unit	Jasco, Groß-Umstadt, tempering Germany

Fluorescence Spectrophotometers

Spex FluoroMaxII spectrofluorimeter	Jobin Yvon (München, Germany)
Tecan Genios Platereader	Tecan, Crailsheim, Germany

Voltage sources

LKB-GPS 200/400	Amersham, Uppsala, Sweden
EPS 3500, 301 and EPS 1001	GE Healthcare (Freiburg, Germany)

Further equipment

Centricon 10, Centricon 30	Amicon, Witten, Germany
Eppendorf thermomixer	Eppendorf, Hamburg, Germany
Digital thermometer	Keithley, Cleveland, USA
Icemachine	Ziegra, Isernhagen, Germany
Magnetic stirrer Heidolph MR 2000	Heidolph, Kelheim, Germany
Metal thermo block TB 1	Biometra, Göttingen, Germany
Millipore Ultra-15 device	Millipore, Bedford, USA
PE tubes (50/15 ml)	Greiner & Söhne, Nürtingen, Germany
pH-indicator strips	Roth, Karlsruhe, Germany
pH-meter	WTW, Weilheim, Germany
Sterile filters 0,2 µm	Zefa, München, Germany
Test tube roller	Heidolph, Kelheim, Germany
Varioklav steam autoclave EP-Z	H+P, Oberschleißheim, Germany
Water bath Haake F6-K	Haake, Karlsruhe, Germany

8.4.1 Computer software

Adobe Reader 5.0	Adobe Inc., San Jose, USA
Microsoft Office 2007	Microsoft, Unterschleißheim, Germany
Origin 8	OriginLab, Northhampton, USA
ProtParamTool	ExPasy (http://expasy.hcuge.ch/)
Reference Manager 12	ISI, Philadelphia, USA

8.5 Molecular biological methods

8.5.1 Cultivation and storage of *E.coli*

E. coli cells were cultivated in a thermoregulated incubator at 30°C or 37°C either on LB plates or in LB liquid media. Strains were selected by addition of appropriate antibiotics to the media, hence selecting cells containing the corresponding resistance genes either on the plasmid or in the genome. Liquid cultures were inoculated from fresh overnight cultures or by transferring single colonies from plates. Bacterial division was monitored at 600 nm (OD_{600nm} = 1 corresponding to approx. $8x10^8$ cells/ml). For long-term storage, 5 ml of a freshly inoculated culture were centrifuged at 5,000 x g and the sediment was resuspended in 1 ml medium. 300 µl 50 % glycerol were added to 700 µl of bacterial suspension resulting in a 15% glycerol culture stock. The culture was frozen using liquid nitrogen and stored at -80°C.

Applied concentrations of antibiotics:

Ampicillin: 100 µg/ml

Kanamycin: 40 µg/ml

Chloramphenicol: 35 µg/ml

8.5.2 Preparation of chemical competent E. coli cells

10 large colonies were transferred of a fresh overnight cultured plate into 250 mL SOB medium and incubated at 19°C to an OD_{600} of 0.5 (approx. 24 h). After a cold shock of 10 minutes (min) on ice, cells were spinned down at 4000 rpm for 10 min at 4°C. Cells were gently resuspended in 80 ml ice-cold TB medium and stored on ice for 10 min. After spinning the cells down again, the pellet was resuspended in 20 ml ice-cold TB containing 1.4 mL DMSO and the cells aliquoted to 100 µL, frozen in liquid nitrogen and stored at -80 °C.

8.5.3 Transformation of competent *E.coli* cells and amplification of plasmid DNA

100 µl of chemical competent *E.coli* cells were incubated with 50 ng of plasmid DNA on ice for 15min. After a heat step at 42°C for 60 sec, cells were left on ice for another 2 min. 700 µL LB0 were added and the cells incubated at 37°C for 45 min to let the resistance genes be transcribed and the antibiotics converting enzymes be synthesized. Cells were pelleted at 5000 rpm for 5 min and plated on selection media. Plates were incubated at 37°C ON.

For amplification of plasmid DNA, single colonies were transferred into 5 ml of LB containing the regarding antibiotics and grown in a shaker at 37°C ON. For plasmid preparation, cells were harvested and treated with the Wizard Plus SV Minipreps DNA Purification System.

8.5.4 Sequencing and DNA storage

Plasmid DNA was diluted to a concentration between 30 to 100 ng/µL and sent to GATC Biotech AG, Konstanz, Germany for sequencing.

DNA was stored in sterile H_2O at -20 °C.

8.6 Protein chemical methods

8.6.1 Protein expression

The proteins produced in this thesis were expressed in the BL21-CodonPlus(DE3)-RIL strain which carries an additional plasmid encoding extra copies of the *argU*, *ileY* and *leuW* tRNA genes and enables efficient high-level expression of heterologous protein of AT-rich genomes in *E. coli*. Over-expression of proteins can deplete the pool of rare tRNAs and stall translation. BL21-CodonPlus strains encode tRNAs that most frequently limit translation of heterologous proteins. The pET28b and pProex system were used as expression vector with the regarding genes. Expression was carried out on 6 l scale. After inoculation and growth at 37 °C to an OD of 0.8, expression was induced by the addition of 1 mM IPTG and allowed to proceed overnight for inclusion body production. Cells were harvested at 6000 rpm and 4°C for 15 min (JA 10 rotor).

8.6.2 Production of inclusion bodies

After harvest, cells were resuspended in an appropriate volume of inclusion body (IB) preparation buffer, supplement with DNase I and protease inhibitor mix HP. After cell disruption, 2.5 % Triton X-100 was added and the solution was stirred at 4°C for 30 minutes to solubilize membrane fragments. The inclusion bodies were sedimented by centrifugation

(20000 rpm; 20 minutes, 4°C; rotor: JA-25.50) and washed three times with IB preparation buffer. The IB pellet was stored at -20°C, or directly solubilized in GdmCl-containing buffer for further purification. In that case, the inclusion bodies were dissolved in ~ 100 ml of buffer supplemented with 10 mM β-mercaptoethanol and stirred at 4°C for 60 minutes. Following this, the solution was cleared by centrifugation (20000 rpm; 20 minutes; 4°C; rotor: JA25.50).

IB preparation buffer:	100 mM	Tris/HCl pH 7.5
	150 mM	NaCl
	10 mM	EDTA

8.6.3 Purification of BiP

Cell pellets of 6 l cultures were resuspended in 100 mL Ni-NTA buffer A containing 1 ml of Protease Inhibitor Mix HP and processed by a cell disruptor at 1.8 kbar. The lysate was centrifuged at 18,000 rpm and 4°C for 25 min in a JA 25.50 rotor.

The supernatant was loaded onto a in Ni-NTA buffer A pre-equilibrated 5 mL HisTrap FastFlow Ni-affinity chromatography column. The column was washed with 120 ml Ni-NTA buffer A and subsequently the bound fraction was eluted with 100 % Ni-NTA buffer B. Subsequently the elution fraction was applied on a Superdex 200 Prep Grade column equilibrated in HKM buffer. BiP containing fractions were analyzed by SDS-PAGE and homogeneity was confirmed by SDS-PAGE (for a detailed purification scheme see Dissertation Moritz Marcinowski, TU München).

Ni-NTA buffer A (BiP)	50 mM	Hepes/KOH pH 7.5
	300 mM	KCl
Ni-NTA buffer B (BiP)	50 mM	Hepes/KOH pH 7.5
	300 mM	KCl
	500 mM	imidazol
HKM buffer	50 mM	Hepes/KOH pH7.5
	150 mM	KCl
	10 mM	$MgCl_2$

8.6.4 Purification of ERdJ3 inclusion bodies

ERdJ3 is a glycosylated protein but recombinant production in *E. coli* only allows non-glycosylated protein expression. Hence, the protein used in this work is not glycosylated. ERdJ3 was purified from solubilised IBs by a denaturating Ni-NTA as the clone carries an N-terminal His-tag. Briefly, a 5 ml HisTrap FastFlow column was equilibrated in Ni-NTA buffer A. After a wash of 120 ml with Ni-NTA buffer A the bound fraction was eluted with 100 % Ni-NTA buffer B and the homogeneity of the pooled ERdJ3 fractions was determined by SDS-PAGE (for a detailed purification scheme see Dissertation Moritz Marcinowski, TU München).

Ni-NTA buffer A (ERdJ3)	50 mM	Hepes/KOH pH 7.5
	300 mM	KCl
	1 mM	β-mercaptoethanol
Ni-NTA buffer B (ERdJ3)	50 mM	Hepes/KOH pH 7.5
	300 mM	KCl
	500 mM	imidazol
	1 mM	β-mercaptoethanol

8.6.5 Purification of Fab from inclusion bodies

The light chain and the Fd fragment were produced, purified and stored independently. After solubilization of the IBs the fragments were purified on a Superdex Prep Grade column equilibrated in 50 mM Tris/HCl pH 8; 6 M GdmCl; 5 mM DTT. The homogeneity of the proteins was analyzed by SDS-PAGE and both chains were stored separately at -20°C.

8.6.6 SDS-PAGE

SDS-PAGE was performed in accordance to the protocol of Laemmli (Laemmli 1970) SDS-PAGE gels were poured and contained the following compounds:

Separation gel:

3.75 mL 40% Acrylamide (40% w/v, Acrylamide/Bisacrylamide 38:2)
2.5 mL 4 x SDS-Buffer (0.8% SDS, 1.5 M Tris/HCl, pH 8.8)
3.75 mL bidest. H_2O

Stacking gel:

0.625ml 40% Acrylamide (40% w/v, Acrylamide/Bisacrylamide 38:2)
2.5 mL 2 x SDS-Buffer (0.4% SDS, 0.25 M Tris/HCl, pH 6.8)
1.875 mL bidest. H_2O

Polymerization of the solution was induced by adding TEMED and APS. Electrophoresis was carried out at constant voltages (150 V) for 1 h. Gels were stained with Coomassie in accordance to the protocol of Fairbanks et al (Fairbanks, et al. 1971).

8.6.7 Refolding of proteins

Denaturation of native proteins and solubilization of IBs was performed by making use of the kinetically stable chaotropic agent guanidinium hydrochloride and the reducing agent DTT in case of cysteine presence. The applied buffer system was chosen according to the individual storage buffer. Denaturation samples were incubated at RT for 3 h. Refolding was initiated by rapid dilution into the respective refolding buffer in a microwell plate to a final volume of 1 ml. The refolding samples were incubated ON at 4 °C with the exception of the Fab fragment, which was incubated at 10°C for 72 h. Native and refolding controls were treated identically.

Protein	Denaturation buffer	Final concentration in denaturation buffer	Dilution factor for renaturation
GFP	40 mM phosphate buffer pH 7.5, 6 M GdmCl	0.5 mg/ml	1 : 100
Glucokinase	50 mM Tris/HCl pH 8.0, 5 mM DTT	0.8 mg/ml	1 : 100
Lysozyme	0.1 M phosphate buffer, 5 M GdmCl, 5 mM DTT	1 mg/ml	1 : 200
Fab fragment	0.1 M Tris/HCl pH 7.5, 6 M GdmCl, 2 mM EDTA, 5 mM DTT	0.2 mg/ml	1 : 50
ERdJ3 IBs	50 mM Hepes pH 7.5, 5 M GdmCl, 5 mM DTT	22 mg/ml	1 : 50
Protease IBs	50 mM Hepes pH 7.5, 5 M GdmCl	0.5 mg/ml	1 : 100

8.6.8 Functional protein assays

8.6.8.1 Assay on GFP folding and chromophore integrity

GFP refolding efficiency was deduced by monitoring the fluorescence intensities of the chromophore. The refolded and native samples were transferred to LUMITRAC^{TM} 96-well plates and the fluorescence measured at 30 °C in a Tecan genios plate reader equipped with the 405 nm excitation filter and the 512 nm emission filter. The following settings were applied: integration time: 80 µs, number of flashes: 10, gain: optimal, read mode: top, shake: 2 sec. Refolding yields were calculated by comparison with the native samples.

8.6.8.2 Assay on Glucokinase ATPse-activity

Glucokinase refolding efficiency was determined by monitoring the protein activities in a regenerative ATPase assay. This assay uses a coupled-enzymatic approach that rapidly converts produced ADP to ATP in the presence of phosphoenolpyruvate, pyruvate kinase, NADH and lactate dehydrogenase. The consumption of NADH to NAD^+ can be followed spectroscopically at a wavelength of λ = 340 nm. The following premix was prepared:

8500 µl	HKM buffer (50 mM Hepes/KOH pH 7.5, 150 mM KCl, 10 mM $MgCl_2$)
240 µl	100 mM phosphoenolpyruvate
35 µl	50 mM NADH
12 µl	Pyruvate kinase suspension (Roche, Mannheim, Germany)
44 µl	Lactate dehydrogenase suspension (Roche, Mannheim, Germany)

The assay premix containing additionally a final concentration of 600 µM D-glucose and 2.5 mM ATP was transferred to µClear® 96-well plates and the glucokinase samples were added to a final protein concentration of 66 ng/ml after baseline monitoring in a Tecan sunrise plate reader at 37°C. The following settings were applied for absorbance monitoring: absorption wavelength: 340 nm, read mode: normal, kinetic interval: 30 sec, number of cycles: 20 with shake duration: 2 sec. For evaluation, the specific activities (U/mg) of the refolded and the native samples were calculated from the slopes using the differential molar extinction coefficient of NADH and NAD^+ of 6200 $cm^{-1}M^{-1}$ at 340 nm.

$$v_{spez} = \frac{m}{d \cdot \left(-6200 \frac{1}{cm \cdot M}\right) \cdot c_{ATPase}}$$

m is the slope of the resulting lines, d is the thickness of the cuvette in cm and c_{ATPase} is the concentration of the respective ATPase in µM.

8.6.8.3 Assay on Lysozyme activity

Lysozyme activity was determinded according to the EnzCheck lysozyme assay from Invitrogen (Germany) according to the manufacturer's instructions. The kit provides fluorescein labeled *Micrococcus lysodeikticus* cell walls as substrate. In the intact cell walls, fluorescein is quenched, yet high emission levels are obtained upon peptidoglycan cleavage by lysozyme. The fluorescence assay was carried out at 37°C in a Tecan genios plate reader applying a final protein concentration of 2.5 µg/ml and the following settings: excitation wavelength: 485 nm, emission wavelength: 535 nm, gain: 76, number of flashes: 3, interval: 40 sec, number of cycles: 40.

8.6.8.4 Assay on MAK33 Fab antigen-binding

Fab antigen binding was determined by an indirect ELISA. The assay reagents and the streptavidin-coated microwell plates were purchased from Roche (Germany) and the assay carried out at RT according to the provided manufacturer's instructions. Shortly, the refolding and the native samples were transferred to the Strepta-well plates and incubated together with the biotinylated antigen creatine kinase. After washing, the horseradish-peroxidase-conjugated detection antibody (Millipore, California, 1:500 dilution) was added. The second washing-step was followed by the addition of the synthetic peroxidase substrate to initiate the colorimetric reaction, which was instantly monitored in a Tecan genios plate reader at 30°C. As final concentrations, 4 µg/ml Fab fragment and 40 µg/ml creatine kinase were applied in the assay. The reaction was monitored by absorbance measurement at 405 nm for 30 min.

8.6.8.5 Assay on BiP-stimulating activity of ERdJ3

ERdJ3 activity was determined by monitoring the acceleration of BiP's ATPase activity. For this purpose, the ATPase activity of BiP in assay premix (see part 8.6.8.2) was monitored in a Tecan sunrise plate reader at 37°C for 15 min, followed by the addition of the ERdJ3 samples and further activity monitoring for 40 min. As final concentrations, 1 µM BiP, 2 µM ERdJ3 and 5 mM ATP were applied in the assay.

8.6.8.6 Assay on protease activity

To determine the proteolytic activity of the protease, a synthetic peptide containing the recognition motif, a fluorophore and a quencher was used (purchased from Bachem, Switzerland). The fluorescence changes upon peptide cleavage were monitored over 25 minutes at 37°C in a Tecan genios plate-reader with the following settings: excitation

wavelength: 325 nm, emission wavelength: 392 nm, gain: 70, number of flashes: 3. The from the peptide manufacturer proposed buffer was applied as assay buffer (50 mM Tris/HCl pH 7.5, 10 mM $CaCl_2$, 0.15 M NaCl, 0.05 % Brij-35) and final assay concentrations of 20 µg/ml peptide and 1 µg/ml of the protease were used.

9 Acknowledgement

Meinem Doktorvater Prof. Dr. Johannes Buchner gilt mein besonderer Dank für seine Betreuung, wertvollen Anregungen und Hilfestellungen, sowie die Möglichkeit, diese Arbeit an seinem Lehrstuhl durchzuführen.

Dr. Martin Haslbeck danke ich herzlich für die Betreuung des Projekts und seine stete Unterstützung und Hilfsbereitschaft während meiner Doktorarbeit.

Großer Dank gebührt meinem Projektpartner Bernd Anselment für seinen hohen Beitrag während unserer Zusammenarbeit, ebenso Herrn Prof. Dr. Ing. Dirk Weuster-Botz für seine wichtige Kooperation und Unterstützung.

Für ihre bereichernde Mitwirkung im Team und den steten Austausch möchte ich meinen Team-Mitgliedern Tanya Dashivets und Moritz Marcinowski herzlich danken.

Weiterhin danke ich meinen Kollegen und Freunden am Lehrstuhl für die großartige gemeinsam verbrachte Zeit.

Nicht zuletzt möchte ich Frau Susanne Hilber für ihre außerordentlichen organisatorischen Tätigkeiten meine Anerkennung und meinen Dank aussprechen.

Meiner Familie und meinen Freunden gilt mein großer Dank für ihre liebevolle Unterstützung und Fürsorge während der letzten Jahre.

10 Literature

Alibolandi, M. and Mirzahoseini, H. 2011. "Chemical assistance in refolding of bacterial inclusion bodies." *Biochemistry research international* 2011.

Anfinsen, C. B. 1973. "Principles that Govern the Folding of Protein Chains." *Science* 181: 223-230.

Anfinsen, C. B. and Haber, E. 1961. "Studies on the reduction and re-formation of protein disulfide bonds." *J. biol. Chem* 236: 1361-1363.

Anfinsen, C. B., Haber, E., Sela, M. and White Jr, F. H. 1961. "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain." *Proceedings of the National Academy of Sciences of the United States of America* 47: 1309.

Anselment, B., Baerend, D., Mey, E., Haslbeck, M. and Buchner, J. 2010. "Experimental optimization of protein
refolding with a genetic algorithm." *Protein Science* 19: 2085—2095.

Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G. and Kiefhaber, T. 2011. "Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction." *Proceedings of the National Academy of Sciences* 108: 3952-3957.

Baldwin, R. L. 2007. "Energetics of protein folding." *Journal of molecular biology* 371: 283-301.

Baneyx, F. 1999. "Recombinant protein expression in Escherichia coli." *Current Opinion in Biotechnology* 10: 411-421.

Basu, A., Li, X. and Leong, S. S. J. 2011. "Refolding of proteins from inclusion bodies: rational design and recipes." *Applied microbiology and biotechnology*: 1-11.

Baynes, B. M., Wang, D. I. C. and Bernhardt, L. 2005. "Role of arginine in the stabilization of proteins against aggregation." *Biochemistry* 44: 4919-4925.

Beld, J., Woycechowsky, K. J. and Hilvert, D. 2007. "Selenoglutathione: Efficient oxidative protein folding by a diselenide." *Biochemistry* 46: 5382-5390.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. and Sarma, V. R. 1965. "Structure of hen egg-white lysozyme, a three dimensional fourier synthesis at 2~ Ã…ngstroms resolution." *Nature* 206: 757-761.

Bowman, G. R., Voelz, V. A. and Pande, V. S. 2011. "Taming the complexity of protein folding." *Current opinion in structural biology* 21: 4-11.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. 2002. "Proteome-scale purification of human proteins from bacteria." *Proceedings of the National Academy of Sciences* 99: 2654.

Buchner, J., Pastan, I. and Brinkmann, U. 1992. "A method for increasing the yield of properly folded recombinant fusion proteins: single-chain immunotoxins from renaturation of bacterial inclusion bodies." *Analytical biochemistry* 205: 263-270.

Buchner, J. and Rudolph, R. 1991. "Renaturation, purification and characterization of recombinant Fab-fragments produced in Escherichia coli." *Nature Biotechnology* 9: 157-162.

Buckel, P., Hübner-Parajsz, C., Mattes, R., Lenz, H., Haug, H. and Beaucamp, K. 1987. "Cloning and nucleotide sequence of heavy-and light-chain cDNAs from a creatine-kinase-specific monoclonal antibody." *Gene* 51: 13-19.

Buckle, A. M., Devlin, G. L., Jodun, R. A., Fulton, K. F., Faux, N., Whisstock, J. C. and Bottomley, S. P. 2005. "The matrix refolded." *Nature methods* 2: 3-3.

Chalfie, M. 1995. "Green fluorescent protein." *Photochemistry and photobiology* 62: 651-656.

Chow, M. K., Amin, A. A., Fulton, K. F., Whisstock, J. C., Buckle, A. M. and Bottomley, S. P. 2006a. "REFOLD: an analytical database of protein refolding methods." *Protein Expr Purif. 2006 Mar;46(1):166-71. Epub 2005 Aug 15.*

Chow, M. K. M., Amin, A. A., Fulton, K. F., Fernando, T., Kamau, L., Batty, C., Louca, M., Ho, S., Whisstock, J. C., Bottomley, S. P. and Buckle, A. M. 2006b. "The REFOLD database: a tool for the optimization of protein expression and refolding." *Nucleic acids research* 34: D207-D212.

Clark, E. D. B. 2001. "Protein refolding for industrial processes." *Current Opinion in Biotechnology* 12: 202-207.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D. and Popovic, Z. 2010. "Predicting protein structures with a multiplayer online game." *Nature* 466: 756-760.

Coutard, B., Danchin, E. G. J., Oubelaid, R., Canard, B. and Bignon, C. 2012. "Single pH buffer refolding screen for protein from inclusion bodies." *Protein Expression and Purification* 82: 352-359.

Dashivets, T., Wood, N., Hergersberg, C., Buchner, J. and Haslbeck, M. 2009. "Rapid Matrix-Assisted Refolding of Histidine-Tagged Proteins." *ChemBioChem* 10: 869-876.

Davies, R. C., Neuberger, A. and Wilson, B. M. 1969. "The dependence of lysozyme activity on pH and ionic strength." *Biochimica et Biophysica Acta (BBA)-Enzymology* 178: 294-305.

De Bernardez-Clark, E. and Georgiou, G. 1991. "Inclusion bodies and recovery of proteins from the aggregated state." *Protein refolding*: 1-20.

Dill, K. A. 1990. "Dominant forces in protein folding." *Biochemistry* 29: 7133-7155.

Dill, K. A. and Chan, H. S. 1997. "From Levinthal to pathways to funnels." *Nature structural biology* 4: 10-19.

Dobson, C. M. 2003. "Protein folding and misfolding." *Nature* 426: 884-890.

Eiberle, M. K. and Jungbauer, A. 2010. "Technical refolding of proteins: Do we have freedom to operate?" *Biotechnology journal* 5: 547-559.

Fairbanks, G., Steck, T. L. and Wallach, D. F. H. 1971. "Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane." *Biochemistry* 10: 2606-2617.

Fersht, A. R. and Daggett, V. 2002. "Protein Folding and Unfolding at Atomic Resolution." *Cell* 108: 573-582.

Fink, A. L. 2005. "Natively unfolded proteins." *Current opinion in structural biology* 15: 35-41.

Georgiou, G. and Valax, P. 1996. "Expression of correctly folded proteins in Escherichia coli." *Current Opinion in Biotechnology* 7: 190-197.

Gobin, O. C., Martinez Joaristi, A. and Schüth, F. 2007. "Multi-objective optimization in combinatorial chemistry applied to the selective catalytic reduction of NO with C3H6." *Journal of Catalysis* 252: 205-214.

Guise, A., West, S. and Chaudhuri, J. 1996. "Protein folding in vivo and renaturation of recombinant proteins from inclusion bodies." *Molecular Biotechnology* 6: 53-64.

Hakim, R. and Benhar, I. 2009. "Inclonals: IgGs and IgG-enzyme fusion proteins produced in an E. coli expression-refolding system." 281. Landes Bioscience.

Hansmann, U. H. E. and Okamoto, Y. 1999. "New Monte Carlo algorithms for protein folding." *Current opinion in structural biology* 9: 177-183.

Havel, J., Link, H., Hofinger, M., Franco-Lara, E. and Weuster-Botz, D. 2006. "Comparison of genetic algorithms for experimental multi-objective optimization on the example of medium design for cyanobacteria." *Biotechnology journal* 1: 549-555.

Hevehan, D. L. and De Bernardez Clark, E. 1997. "Oxidative renaturation of lysozyme at high concentrations." *Biotechnology and bioengineering* 54: 221-230.

Holland, J. H. 1975. *Adaptation in natural and artificial systems*. University of Michigan press.

Huang, C. J., Lin, H. and Yang, X. 2012. "Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements." *Journal of Industrial Microbiology & Biotechnology*: 1-17.

Jackson, S. E. and Fersht, A. R. 1991. "Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition." *Biochemistry* 30: 10428-10435.

Jennings, P. A. and Wright, P. E. 1993. "Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin." *Science* 262: 892-896.

Jin, Y., Awad, W., Petrova, K. and Hendershot, L. M. 2008. "Regulated release of ERdj3 from unfolded proteins by BiP." *EMBO J. 2008 Nov 5;27(21):2873-82. Epub 2008 Oct 16.*

Kamionka, M. 2011. "Engineering of therapeutic proteins production in Escherichia coli." *Current Pharmaceutical Biotechnology* 12: 268.

Kane, J. F. and Hartley, D. L. 1988. "Formation of recombinant protein inclusion bodies in Escherichia coli." *Trends in biotechnology* 6: 95-101.

Kiefhaber, T., Rudolph, R., Kohler, H. H. and Buchner, J. 1991. "Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation." *Nature Biotechnology* 9: 825-829.

Kim, P. S. and Baldwin, R. L. 1990. "Intermediates in the folding reactions of small proteins." *Annual review of biochemistry* 59: 631-660.

Laemmli, U. K. 1970. "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* 227: 680-685.

Leopold, P. E., Montal, M. and Onuchic, J. N. 1992. "Protein folding funnels: a kinetic approach to the sequence-structure relationship." *Proceedings of the National Academy of Sciences* 89: 8721.

Levinthal, C. 1968. "Are there pathways for protein folding?" *Journal of Medical Physics* 65: 44-45.

Li, Z. and Scheraga, H. A. 1987. "Monte Carlo-minimization approach to the multiple-minima problem in protein folding." *Proceedings of the National Academy of Sciences* 84: 6611.

Lilie, H., Schwarz, E. and Rudolph, R. 1998. "Advances in refolding of proteins produced in E. coli." *Current Opinion in Biotechnology* 9: 497-501.

Marcinowski, M., Höller, M., Feige, M. J., Baerend, D., Lamb, D. C. and Buchner, J. 2011. "Substrate discrimination of the chaperone BiP by autonomous and cochaperone-regulated conformational transitions." *Nat Struct Mol Biol* 18: 150-158.

Marston, F. A. 1986. "The purification of eukaryotic polypeptides synthesized in Escherichia coli." *Biochemical Journal* 240: 1.

Mayer, M. and Buchner, J. 2004. "Refolding of inclusion body proteins." *Methods in molecular medicine* 94: 239-254.

Meunier, L., Usherwood, Y. K., Chung, K. T. and Hendershot, L. M. 2002. "A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins." *Mol Biol Cell. 2002 Dec;13(12):4456-69.*

Meyer, D., Schneider-Fresenius, C., Horlacher, R., Peist, R. and Boos, W. 1997. "Molecular characterization of glucokinase from Escherichia coli K-12." *J Bacteriol. 1997 Feb;179(4):1298-306.*

Norby, J. G. 1988. "Coupled assay of Na+,K+ - ATPase activity." *Methods in enzymology* 156: 116-119.

Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. and Remington, S. J. 1996. "Crystal structure of the Aequorea victoria green fluorescent protein." *Science. 1996 Sep 6;273(5280):1392-5.*

Orosz, F. and Ovadi, J. 2011. "Proteins without 3D structure: definition, detection and beyond." *Bioinformatics* 27: 1449-1454.

Pande, V. S., Beauchamp, K. and Bowman, G. R. 2010. "Everything you wanted to know about Markov State Models but were afraid to ask." *Methods* 52: 99-105.

Phan, J., Yamout, N., Schmidberger, J., Bottomley, S. P. and Buckle, A. M. 2011. "Refolding your protein with a little help from REFOLD." *Methods Mol Biol. 2011;752:45-57.*

Phillips, D. C. 1967. "The hen egg-white lysozyme molecule." *Proceedings of the National Academy of Sciences of the United States of America* 57: 483-495.

Ptitsyn, O. B. 1995. "Molten globule and protein folding." *Advances in protein chemistry* 47: 83-229.

Rechenberg, I. 1973. "Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Verlag Frommann-Holzboog, Stuttgart-Bad Cannstatt, 1973. ISBN: 3-7728-0373-3." German.

Rose, G. D., Fleming, P. J., Banavar, J. R. and Maritan, A. 2006. "A backbone-based theory of protein folding." *Proceedings of the National Academy of Sciences* 103: 16623-16633.

Rudolph, D., Fischer, D. and Mattes, D. 1993. "Process for activating recombinant heterologous eucaryotic disulfide-bridged proteins after expression by procaryotes." EP Patent 0,219,874.

Rudolph, R. 1990. "Renaturation of recombinant, disulfide-bonded proteins from inclusion bodies." *Modern methods in protein and nucleic acid research*: 149-172.

Rudolph, R. and Lilie, H. 1996. "In vitro folding of inclusion body proteins." *The FASEB Journal* 10: 49-56.

Shaked, Z. e., Szajewski, R. P. and Whitesides, G. M. 1980. "Rates of thiol-disulfide interchange reactions involving proteins and kinetic measurements of thiol pKa values." *Biochemistry* 19: 4156-4166.

Shen, Y. and Hendershot, L. M. 2005. "ERdj3, a Stress-inducible Endoplasmic Reticulum DnaJ Homologue, Serves as a CoFactor for BiP's Interactions with Unfolded Substrates." *Molecular Biology of the Cell* 16: 40-50.

Shimomura, O., Johnson, F. H. and Saiga, Y. 1962. "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea." *Journal of Cellular and Comparative Physiology* 59: 223-239.

Singh, S. M. and Panda, A. K. 2005. "Solubilization and refolding of bacterial inclusion body proteins." *Journal of bioscience and bioengineering* 99: 303-310.

Terpe, K. 2006. "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Applied microbiology and biotechnology* 72: 211-222.

Thastrup, O., Tullin, S., Kongsbak Poulsen, L. and Bjørn, S. 2001. "Fluorescent Proteins." *US patent*.

Tischer, A., Lilie, H., Rudolph, R. and Lange, C. 2010. "L-Arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA." *Protein Science* 19: 1783-1795.

Tompa, P. 2005. "The interplay between structure and function in intrinsically unstructured proteins." *FEBS Letters* 579: 3346-3354.

Tsien, R. Y. 1998. "THE GREEN FLUORESCENT PROTEIN." *Annual review of biochemistry* 67: 509-544.

Vagenende, V., Yap, M. G. S. and Trout, B. L. 2009. "Mechanisms of protein stabilization and prevention of protein aggregation by glycerol." *Biochemistry* 48: 11084-11096.

Vallejo, L. and Rinas, U. 2004. "Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins." *Microbial Cell Factories* 3: 11.

Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. and Bignon, C. 2004. "High-throughput automated refolding screening of inclusion bodies." *Protein Science* 13: 2782-2792.

Weber, L. 1998. "Applications of genetic algorithms in molecular diversity." *Current Opinion in Chemical Biology* 2: 381-385.

Weuster-Botz, D. 2000. "Experimental design for fermentation media development: statistical design or global random search?" *Journal of bioscience and bioengineering* 90: 473-483.

Wolynes, P. G., Onuchic, J. N. and Thirumalai, D. 1995. "Navigating the folding routes." *Science* 267: 1619.

Zuzek, M., Friedrich, J., Cestnik, B., Karalic, A. and Cimerman, A. 1996. "Optimization of fermentation medium by a modified method of genetic algorithms." *Biotechnology techniques* 10: 991-996.

Zwanzig, R., Szabo, A. and Bagchi, B. 1992. "Levinthal's paradox." *Proceedings of the National Academy of Sciences* 89: 20.