



ELSEVIER

# Directed evolution of proteins for heterologous expression and stability

Cintia Roodveldt, Amir Aharoni and Dan S Tawfik

Recent developments have been made in the application of directed evolution to achieve the efficient heterologous expression of proteins in *Escherichia coli* and yeast by increasing the stability and solubility of the protein in the host environment. One interesting conclusion that emerges is that the evolutionary process often improves the stability and solubility of an intermediate (apoprotein, proprotein or folding intermediate) that otherwise constitutes a bottleneck to functional expression, rather than altering the protein's final state.

## Addresses

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding author: Tawfik, Dan S (tawfik@weizmann.ac.il)

**Current Opinion in Structural Biology** 2005, 15:50–56

This review comes from a themed issue on  
Folding and binding  
Edited by Gideon Schreiber and Luis Serrano

Available online 13th January 2005

0959-440X/\$ – see front matter

© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2005.01.001

## Introduction

Expressing high levels of stable and functional proteins remains the bottleneck of many scientific and biotechnological endeavors, including the determination of structures in a high-throughput fashion, the design, engineering and directed evolution of tailor-made proteins, and the metabolic engineering of bacteria [1]. In fact, data from a sample of more than 200 non-membrane genes from one bacterium species indicate that 50% of the genes will require further optimization to obtain soluble or stable proteins for crystallization [1,2].

Conventional approaches to the production of soluble and active proteins in heterologous expression systems include low-temperature expression, promoters with different strengths, modified growth media and a variety of solubility-enhancing fusion tags (reviewed in [3,4,5,6]). More recently, a series of vectors and fusion partners that can be screened for high-level functional expression of a target protein have been developed [1]. These approaches are often successful, as are some engineering approaches (see also Update) but, at the end of the day,

they cannot modify the intrinsic properties of the protein. Thus, even when a soluble fraction of the protein is obtained, the protein may still be inactive or aggregate during purification and storage.

An alternative to expression optimization is protein engineering by rational design (reviewed in [7]). There are now several examples of proteins that have been stabilized by the introduction of mutations with small yet cumulative stabilizing effects [8–11]. It is largely unknown, however, how the stability of a protein is encoded in its sequence and how individual amino acid changes contribute to stability [12], not to mention that most of the interesting targets of structural and mechanistic studies are obviously of unknown structure and are therefore not amenable to rational design.

## Directed evolution of proteins

In the past several years, directed evolution has emerged as an alternative approach to rational design, enabling the improvement of structural and functional properties, such as stability and performance under different conditions (e.g. at extreme temperatures and pH, and in organic co-solvents), or changes in their reaction and substrate specificity [13]. Rather than designing a limited number of site-directed mutants, directed evolution implements an iterative Darwinian optimization process, whereby the fittest variants are selected from an ensemble of random mutations. Improved variants are identified by screening or selection for the properties of interest and their encoding genes are then used as parent genes for the following round of evolution. This approach has proven particularly advantageous in cases in which prior knowledge of the protein's structure or mechanism was not available [14]. A radical example of the power of the 'blind watchmaker' is the selection of a novel ADP- and zinc-binding fold from a library of completely random polypeptide sequences [15,16]. Even in cases in which the structure of the protein was known, directed evolution has provided combinations of mutations that would not have been deduced from phylogenetic comparisons or from structural analyses (e.g. [12]).

This review addresses the literature of the past two years (for an earlier review, see [17]), and focuses on the application of directed evolution to achieve the efficient heterologous expression of proteins in *Escherichia coli* and yeast by increasing the stability and solubility of the protein in the host environment (Table 1) (for reviews on related issues, see [17,18]).

Table 1

Recent examples of directed evolution for heterologous expression, solubility or stability<sup>a</sup>.

Target protein	Library origin	Screening system	Results <sup>b</sup>		Refs
			Property	Improvement	
1. Anti-hag scFv	EP/DS	Ribosome display with translation and binding steps under reducing conditions	Functional cytoplasmic expression and chemical stability	200- to 400-fold higher expression; increase in urea denaturation midpoint from 4.1 M to 5.0 M.	[45]
2. Cold shock protein from <i>Bacillus subtilis</i>	Six positions diversified by SM	Stability to proteolytic digestion and GdmCl-induced unfolding, fd phage library.	Thermal stability	22 °C higher T <sub>M</sub> (from mesophile to thermophile)	[12]
3. Galactose oxidase from fungi	EP/StEP	Liquid bacterial cultures, assayed for native enzymatic activity	Functional expression, long-term stability (residual activity in presence of CuSO <sub>4</sub> and catalase).	18-fold, 60% higher long-term stability and twofold higher k <sub>cat</sub> /K <sub>M</sub>	[46]
4. Human cytochrome P450 (1A2) and bacterial 450 (BM3)	SHIPREC (16% homology)	CAT preselection and screening of colonies with 1A2's substrate	Functional expression	From <1.5% to 14% soluble protein in the cytosolic fraction.	[26]
5. N-carbamyl-D-amino acid amidohydrolase from <i>Agrobacterium tumefaciens</i>	DS	Filter screening of colonies for native activity after incubation at high temperature or with oxidizing reagent	Thermal and oxidative stability (residual activity after heat or oxidizer).	8-fold (incubation at 70 °C) and 16-fold (incubation with H <sub>2</sub> O <sub>2</sub> )	[47]
6. Methyl transferase, NDP-K and tartrate dehydratase β from <i>P. aerophilum</i>	DS	Fluorescence screening of <i>E. coli</i> colonies displaying correctly folded protein (GFP reporter)	Functional expression	From insoluble to 50–95% soluble expression. Crystal structure determination for NDP-K.	[24*]
7. Barnase from <i>Bacillus amyloliquefaciens</i>	17 positions diversified by oligonucleotide assembly.	Proteolytic selection by phage display, affinity selection for binding to ligand.	Thermodynamic stability (free energy of folding)	Similar or lower than wild-type barnase	[48]
8. Laccase from <i>Myceliophthora thermophila</i> (MtL)	EP/ <i>in vivo</i> shuffling in yeast/ <i>in vivo</i> gap repair/StEP.	Screening yeast cultures for native enzymatic activity	Functional expression	8–9-fold higher expression and 15–30-fold higher k <sub>cat</sub> .	[19*]
9. GlcDH from <i>Bacillus</i> species	FS of three wild-type genes and one improved mutant	Filter-based screening and selection for enzymatic activity after incubation at high temperature.	Thermal stability in the absence of NaCl	415-fold higher t <sub>1/2</sub> at high temperature (66 °C)	[38]
10. Mammalian serum PON1	FS of four mammalian PON1 genes/DS	Agar colony screens for a promiscuous esterase activity	Functional expression	>100-fold in soluble active protein. Crystal structure determination of the first PON family member.	[14*,39*]
11. Lipase B from <i>Candida antarctica</i>	<i>In vitro</i> and <i>in vivo</i> FS with two thermostable bacterial lipases B.	Screening yeast cultures for native enzymatic activity at ambient temperature.	Thermal stability and activity	12-fold higher t <sub>1/2</sub> at 45 °C, 11% higher T <sub>m</sub> , 16-fold higher k <sub>cat</sub> /K <sub>M</sub> .	[49]
12. Phosphotriesterase from <i>Pseudomonas diminuta</i> (PTE)	EP or BA/DS	Agar colony screens for a promiscuous esterase activity	Functional expression	~20-fold higher content in soluble fraction, owing to increased stability of the apoenzyme	(c)

<sup>a</sup>The examples shown (in chronological order) use established methods of sequence diversification in either isolation or combination: error-prone PCR (EP); base-analogue-induced mutagenesis (BA); saturation mutagenesis (SM); DNA shuffling (DS); family shuffling (FS); staggered extension process (StEP); sequence-homology-independent protein recombination (SHIPREC). <sup>b</sup>Note that the varying property was not part of the screen or selection in some cases. The improvement factor is based on the best clone(s) reported in terms of functional expression, stability or solubility. <sup>c</sup>C Roodveldt, DS Tawfik, unpublished.

### Directed evolution for heterologous expression – the methodology

Heterologous expression is performed primarily in *E. coli*, although yeast and insect cells are becoming increasingly popular, particularly for eukaryotic proteins that require post-translational processing. Considering transformation efficiency, stability of plasmid DNA and growth rate, *E. coli* and *Saccharomyces cerevisiae* are also best suited for directed evolution experiments [19\*].

Directed evolution in the laboratory, like natural evolution, involves two key steps: generation of genetic diversity and selection for function. In the laboratory, diversity in the gene of interest is typically created by random mutagenesis using a variety of methods based on error-prone DNA replication and chemical mutagenesis, or by homologous recombination of closely related genes using a process called family shuffling (for recent reviews, see [20–22]). Whereas methods of creating genetic diversity are generally applicable, the selection system needs to be tailored, or modified, for each target protein and aim — a selection for higher expression may not be suitable for altering substrate selectivity and vice versa. Broadly speaking, screening or selecting for heterologous expression can be done in two ways: screening/selecting for the activity of a reporter protein, or screening/selecting for the protein's own function. Particular examples of both approaches are discussed below.

A variety of generic 'C-fusion' approaches have been developed that rely on expression of the target protein as an N-terminal fusion to a reporter protein with a selectable function. An insoluble target protein leads to the aggregation of the reporter protein and loss of its function [17]. As an example, colonies overexpressing green fluorescent protein (GFP) fused to a soluble protein can be easily distinguished owing to GFP folding and fluorophore formation; colonies expressing an insoluble protein do not fluoresce [23,24\*] (Table 1, entry 6).

Another generally applicable method relies on chloramphenicol acetyltransferase (CAT) fusion, whereby selection for solubility of the target protein is performed by antibiotic resistance [25]. Using this methodology, Arnold and co-workers [26] converted membrane-associated insoluble human P450 cytochrome (1A2) into a fairly soluble protein by recombination of 1A2 with a distant bacterial homologue (Table 1, entry 4). The resulting library of hybrid proteins was first selected by fusion to CAT, to isolate all in-frame and correctly folded variants. A subsequent screen for active enzyme variants was performed with a fluorogenic P450 substrate.

In the  $\beta$ -galactosidase complementation assay, a small, approximately 50 amino acid  $\alpha$ -fragment is fused to the C terminus of the target protein, restoring — in those cases in which the target protein remains soluble — the  $\beta$ -

galactosidase activity of a truncated lacZ form (lacZ- $\Omega$ ) by complementation in *trans* [27]. Other potentially generic methodologies are based on screening by cellular stress responses to misfolded proteins [28,29], 'proteolytic selection' by phage display [30] or 'protein stability increased by directed evolution' (Proside) [31]; the last two technologies are based on the principle that infectivity of the phage is coupled to protease resistance of the protein variants. This assumption in turn relies on the observation that proteolysis resistance can be used as a marker of foldedness [32]. Because the C-fusion techniques do not depend on the function of the target protein, they are generally applicable and particularly advantageous when attempting to evolve new folds or when the target protein has no assigned function. However, a screen for soluble expression can turn into a major drawback. The blind fashion in which evolution acts ('you get what you select for' is the first rule of directed evolution [33]) implies that selection for soluble expression may be accompanied by significant changes in activity that may even render the soluble protein non-functional [24]. To drive the selection of soluble and stable variants without compromising the protein's functional properties, the selection system must be based on the protein's own activity. A large variety of assays are available for screening enzyme libraries [34–37]. Indeed, functional assays have been used in several cases. As an example, Baik *et al.* [38] have evolved glucose dehydrogenase (GlcDH), obtaining a 400-fold increase in thermal stability by screening a library generated by family shuffling for the enzyme's activity (Table 1, entry 9). Several other examples are listed in Table 1 (entries 1, 3–5, 7–12) (see also Update).

### Directed evolution for heterologous expression – recent examples

Recent examples of directed evolution for heterologous expression illustrate the power and versatility of this approach in facilitating structural and mechanistic studies. Directed evolution has thus far been the key to crystallizing two structures of general interest: a hyperthermophilic nucleotide diphosphate kinase (NDP-K) [24\*] and the first PON family member [39\*]. Using a GFP C-fusion screen, Waldo and co-workers [24\*] evolved the essentially insoluble NDP-K from *Pyrobaculum aerophilum* into a 90% soluble and active variant that enabled the determination of its crystal structure (Table 1, entry 6).

Serum paraoxonases (PONs) represent a particularly interesting target for directed evolution. PONs comprise a family of enzymes that play a key role in organophosphate detoxification and in preventing atherosclerosis. PONs are widespread in mammals and other vertebrates, as well as in invertebrates. Three subfamilies of PONs are known (PON1–3) and show approximately 60% sequence homology; however, the structure of not a single family member was known before the directed evolution of PON recombinant variants. Attempts to crystallize

serum-purified PON1 resulted in the crystallization of a contaminant protein, instead of PON1 [40]. Moreover, although PONs exhibit a range of hydrolytic activities towards esters, phosphotriesters and lactones, their native or physiological substrate is unknown.

Family shuffling and screening for esterase activity led to the first PON1 and PON3 variants that express in a soluble and active form in *E. coli* (Table 1, entry 10). However, in the absence of knowledge of the native substrate, the screen was based not on PON's native function, but on its promiscuous esterase and phosphotriesterase activities. Interestingly, whereas evolution of PON1 led to a variant with wild-type-like enzymatic properties and high *E. coli* expression, evolution of PON3 by the very same method led to a mutant with increased esterase activity relative to wild-type PON3, in addition to higher expression.

The promiscuous esterase activity was also used for evolving bacterial phosphotriesterase (PTE) (C Roodveldt, DS Tawfik, unpublished) (Table 1, entry 12). In this case, libraries created by random mutagenesis of the PTE gene were screened for higher esterase activity. However, attempts to evolve this promiscuous activity led to a variant with a 20-fold increase in functional expression and enzymatic properties that are essentially identical to those of wild-type PTE.

From a methodological point of view, these works highlight the advantages and drawbacks of using screens based on a promiscuous activity, particularly when the native substrate is unknown (e.g. as in the case of PON), when a facile screen for the native substrate is not available or when the detection of native activity falls out of the dynamic range of the screen (e.g. as in the case of PTE). The snag is that this approach can obviously lead to variants with increased promiscuous activity as well as, or instead of, increased expression.

Some proteins, especially eukaryotic ones, are notoriously difficult to express in heterologous systems. Arnold and co-workers [19<sup>•</sup>] have succeeded in obtaining high functional expression levels for *Myceliophthora thermophila* (MtL) laccase (a fungal glycoprotein), along with higher activity, by directed evolution [19<sup>•</sup>] (Table 1, entry 8). For this purpose, they used a yeast host (*S. cerevisiae*) and a high-throughput screen based on the enzyme's native activity. Both *in vitro* and *in vivo* shuffling were used to achieve random recombination of the isolated variants between rounds. In addition, PCR and gap repair tools [41,42] were used to ensure the recombination of neighboring mutations in a site-directed manner.

### What have we learnt?

Directed evolution can provide some general insights into the factors that dictate protein solubility, stability and

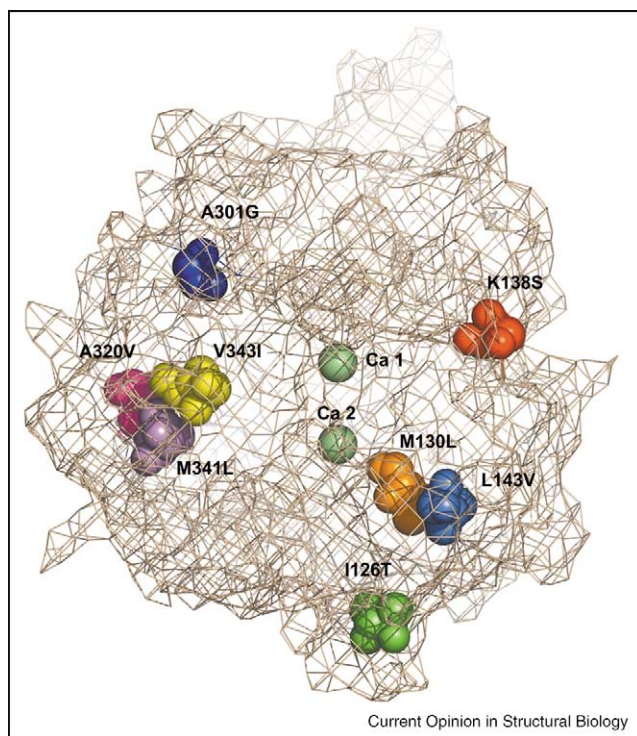
assembly in non-native environments. The directly evolved NDP-K variant contained six amino acid substitutions, which act in a synergistic manner [24<sup>•</sup>]. To explore the possible determinants of increased functional expression, the structures of wild-type and evolved proteins were modeled. This analysis suggested that two out of the six mutations (Glu40Lys and Gly33Asp) yielded more favorable charge distribution and hydrogen bonding at the dimeric interface of the evolved variant. The increased GFP fluorescence of the evolved NDP-K variants may also indicate that the mutations reduced the interference of fused NDP-K with GFP folding, possibly by preventing the formation of non-specific off-pathway aggregates [24<sup>•</sup>].

The highly expressing MtL laccase T2, evolved in yeast, contains 13 mutations [19<sup>•</sup>]. Three of them are located at the three processing sites of the protein, namely the signal sequence, the prosequence and the C-terminal sequence. Presumably, these mutations adjust the MtL sequence to the different specificities of the *S. cerevisiae* proteases compared to those of the natural fungal organism. Whereas N-terminal processing and glycosylation are probably required for secretion, C-terminal processing may play a role in activation of the enzyme [19<sup>•</sup>]. Indeed, the authors have shown that, whereas the His(c2)Arg substitution in the C-terminal tail resulted in an almost fivefold improvement in  $k_{cat}$ , a dramatic loss in activity was observed for the C-truncated mutant, suggesting that the C terminus is essential for early post-translational processing steps [19<sup>•</sup>]. That higher functional expression may require alteration of an intermediate, rather than the protein in its final form, is also apparent from several other examples.

A particularly clear example of how correct folding can be promoted by disfavoring the formation of an undesirable folding intermediate has been recently described. A proline residue within the sequence corresponding to the transmembrane segment of a chloride channel associated with cystic fibrosis has been shown to encode native ( $\alpha$ -helical) structure by disfavoring alternative misfolded conformations (namely  $\beta$ -sheet aggregates). This prevention of the formation of off-pathway structures may occur either by destabilizing the non-native conformation or by raising the kinetic barrier to its formation [43].

In the directed evolution of PON1 for soluble functional expression in *E. coli*, eight conserved mutations were identified. Initially (before we solved the structure), the nature of these mutations was quite puzzling. They were located in two regions of the sequence and most of them involved the substitution of one hydrophobic amino acid for another. The overall hydrophobicity of the variant appeared higher than that of wild-type PON1. At the time, we speculated that "the mutated regions may be involved in the oligomeric packing of PON1" and that

Figure 1



Mesh representation of the three-dimensional structure of a recombinant PON1 (rePON1), with the mutations that confer functional expression in *E. coli*. The sidechains of variant amino acids between rePON1 and wild-type rabbit PON1 are noted as colored spheres. The two calcium ions are represented by lime spheres: Ca1 is the catalytic ion, whereas Ca2 has a structural role. The active site is at the front-top part of the molecule. Eight conserved amino acid mutations were found in all selected variants (see supplementary information to Figure 8 in [14]). All eight mutations are distant from the active site. Six of the eight mutations are located within the hydrophobic core of the protein, five of which are grouped in two clusters (M130L-L143V and A320V-M341L-V343I). The two other substituted positions are partially exposed to the solvent, but only one (I126T) may increase surface polarity.

“subtle mutations in these exposed hydrophobic surfaces may have prevented misfolding and formation of high-order aggregates, and thereby facilitated soluble expression” [14]. A completely different picture emerged once the three-dimensional structure was solved. It turned out that six of the eight mutated residues are located within the hydrophobic core of the protein in two tightly packed clusters (Figure 1) [39]. These mutations may increase the stability of the core, although, in effect, PON1 is a very stable protein. However, PON1 contains two calcium ions, one of which is absolutely essential for its structural integrity. Even in the presence of calcium in the growth media, cytosolic calcium levels in *E. coli* are tightly controlled at the low micromolar range [44]. When overexpressed in *E. coli*, PON1 accumulates to millimolar concentrations and the calcium-free apoenzyme must therefore accumulate to a large degree. It is therefore

likely that the mutations stabilize the apo-form and not the final calcium-assembled holoenzyme form.

A similar scenario seems to have taken place in our directly evolved bacterial PTE variants (C Roodveldt, DS Tawfik, unpublished). Wild-type PTE, when expressed in *E. coli*, is mostly (98%) in an aggregated inactive form. Selection of PTE gene libraries led to a highly soluble variant (~35% soluble expression) containing three point mutations (Lys185Arg, Asp208Gly, Arg319Ser) located on the surface of the protein, away from the dimer interface and the active site. The evolved variant appears to be more sensitive to thermal denaturation and metal chelation than wild-type PTE. However, it appears that the low soluble expression of wild-type PTE is the result of limiting intracellular zinc concentrations and the low stability of the apoenzyme, not of the zinc-assembled holoenzyme form. The functional expression of the newly evolved variant, therefore, seems to reflect the higher stability of the metal-free apoenzyme. Indeed, we could show that most of the evolved enzyme (>90%) is maintained in its zinc-free apo-form until the cell is lysed and exposure to the zinc-supplemented buffer enables formation of the active holoenzyme. By contrast, almost all the wild-type apoenzyme (~98%) is aggregated and only an insignificant amount of the soluble apoenzyme is recovered upon cell lysis (C Roodveldt, DS Tawfik, unpublished).

### Concluding remarks

It is widely assumed that protein expression levels are determined by the ‘solubility’ or ‘stability’ of the expressed protein. This rationale is reflected in most strategies that have been designed to improve overexpression yields. However, recent works highlight a key, and often overlooked, factor. The main determinant of (active) overexpression for many proteins is defined by the thermodynamic and kinetic properties of an intermediate, be it an apoenzyme, proprotein, monomeric state or folding (on- or off-pathway) intermediate, rather than by the stability and solubility of the final state. Because these expression bottlenecks are largely unidentified, the design of stabilizing mutations may prove a daunting task. Evolution, which is directed only by the phenotypic outcome of the process and requires no knowledge of the protein’s structure and folding pathway, seems like an effective solution to this problem. Care must be taken, however, that selection is performed in a controlled manner that improves functional expression but does not alter the protein’s function and structure.

### Update

Stable variants of *Enterobacter cloacae*  $\beta$ -lactamase (BLA) have been generated by directed evolution [50]. A stable chimeric enzyme has been engineered by transplanting the function-determining domains of an unstable and difficult to handle F<sub>1</sub>-ATPase variant into the scaffold

of a stable homolog; this has been used to study the unique structural and functional properties of this enzyme within a stable framework [51].

A recent example of a directed evolution system based on the properties of a preprotein intermediate, applied to the search for an optimized sequence at the secretory signal-mature protein junction for the efficient export of a recombinant protein to the periplasm of *E. coli*, has been reported [52\*].

## Acknowledgements

We gratefully acknowledge financial support from the Israel Science Foundation and the Minerva Foundation. DST is the incumbent of the Elaine Blond Career Development Chair.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Berthold DA, Stenmark P, Nordlund P: **Screening for functional expression and overexpression of a family of diiron-containing interfacial membrane proteins using the univector recombination system.** *Protein Sci* 2003, **12**:124-134.
  2. Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I *et al.*: **Structural proteomics of an archaeon.** *Nat Struct Biol* 2000, **7**:903-909.
  3. Makrides SC: **Strategies for achieving high-level expression of genes in *Escherichia coli*.** *Microbiol Rev* 1996, **60**:512-538.
  4. Braun P, LaBaer J: **High throughput protein production for functional proteomics.** *Trends Biotechnol* 2003, **21**:383-388.
  5. Marsischky G, LaBaer J: **Many paths to many clones: a comparative look at high-throughput cloning methods.** *Genome Res* 2004, **14**:2020-2028.
- A comprehensive and detailed update on the available methods for recombinational cloning and their requirements for high-throughput purposes.
6. Pearlberg J, LaBaer J: **Protein expression clone repositories for functional proteomics.** *Curr Opin Chem Biol* 2004, **8**:98-102.
  7. Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G: **Rational engineering of enzyme stability.** *J Biotechnol* 2004, **113**:105-120.
  8. Wintrode PL, Arnold FH: **Temperature adaptation of enzymes: lessons from laboratory evolution.** *Adv Protein Chem* 2000, **55**:161-225.
  9. Wintrode PL, Miyazaki K, Arnold FH: **Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution.** *J Biol Chem* 2000, **275**:31635-31640.
  10. Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon AP, Wyss M: **The consensus concept for thermostability engineering of proteins: further proof of concept.** *Protein Eng* 2002, **15**:403-411.
  11. D'Amico S, Gerday C, Feller G: **Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase.** *J Mol Biol* 2003, **332**:981-988.
  12. Martin A, Sieber V, Schmid FX: **In-vitro selection of highly stabilized protein variants with optimized surface.** *J Mol Biol* 2001, **309**:717-726.
  13. Tao H, Cornish VW: **Milestones in directed enzyme evolution.** *Curr Opin Chem Biol* 2002, **6**:858-864.
  14. Aharoni A, Gaidukov L, Yagur S, Tokar L, Silman I, Tawfik DS:
    - **Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization.** *Proc Natl Acad Sci USA* 2004, **101**:482-487.
 Two members of an important and widespread family of detoxifying and anti-atherosclerotic enzymes were successfully evolved for bacterial functional expression using a promiscuous activity as the screen. Whereas PON3 increased its catalytic properties, PON1 evolution led to a highly expressed variant that enabled its crystallization.
  15. Keefe AD, Szostak JW: **Functional proteins from a random-sequence library.** *Nature* 2001, **410**:715-718.
  16. Lo Surdo P, Walsh MA, Sollazzo M: **A novel ADP- and zinc-binding fold from function-directed in vitro evolution.** *Nat Struct Mol Biol* 2004, **11**:382-383.
- The first example of the structure of a protein derived from *in vitro* evolution, selected from a completely random sequence. The ADP- and zinc-binding enzyme adopts a new protein fold.
17. Waldo GS: **Genetic screens and directed evolution for protein solubility.** *Curr Opin Chem Biol* 2003, **7**:33-38.
  18. Magliery TJ, Regan L: **Combinatorial approaches to protein stability and structure.** *Eur J Biochem* 2004, **271**:1595-1608.
  19. Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C,
    - Arnold FH: **Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution.** *Appl Environ Microbiol* 2003, **69**:987-995.
 A significant increase in heterologous functional expression of fungal laccase was achieved by directed evolution of the glycoenzyme, using yeast as a host system. Interestingly, the final expression and activity levels of the selected variants seem to rely on their stabilized intermediates during the intracellular post-translational processing stages.
  20. Neylon C: **Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution.** *Nucleic Acids Res* 2004, **32**:1448-1459.
  21. Arnold FH, Georgiou G (Eds): *Directed Enzyme Evolution: Screening and Selection Methods*, vol 230, edn 1. Totowa, New Jersey: Humana Press; 2003.
  22. Farinas ET, Bulter T, Arnold FH: **Directed enzyme evolution.** *Curr Opin Biotechnol* 2001, **12**:545-551.
  23. Waldo GS, Standish BM, Berendzen J, Terwilliger TC: **Rapid protein-folding assay using green fluorescent protein.** *Nat Biotechnol* 1999, **17**:691-695.
  24. Pedelacq JD, Piltch E, Liong EC, Berendzen J, Kim CY, Rho BS,
    - Park MS, Terwilliger TC, Waldo GS: **Engineering soluble proteins for structural genomics.** *Nat Biotechnol* 2002, **20**:927-932.
 The first example of the crystal structure of a protein generated by directed evolution, whereby correctly folded, soluble variants were selected using the GFP reporter system.
  25. Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR: **A simple in vivo assay for increased protein solubility.** *Protein Sci* 1999, **8**:1908-1911.
  26. Sieber V, Martinez CA, Arnold FH: **Libraries of hybrid proteins from distantly related sequences.** *Nat Biotechnol* 2001, **19**:456-460.
  27. Wigley WC, Stidham RD, Smith NM, Hunt JF, Thomas PJ: **Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein.** *Nat Biotechnol* 2001, **19**:131-136.
  28. Lesley SA, Graziano J, Cho CY, Knuth MW, Klock HE: **Gene expression response to misfolded protein as a screen for soluble recombinant protein.** *Protein Eng* 2002, **15**:153-160.
  29. Vostiar I, Tkac J, Mandenius CF: **Off-line monitoring of bacterial stress response during recombinant protein production using an optical biosensor.** *J Biotechnol* 2004, **111**:191-201.
  30. Kristensen P, Winter G: **Proteolytic selection for protein folding using filamentous bacteriophages.** *Fold Des* 1998, **3**:321-328.
  31. Sieber V, Pluckthun A, Schmid FX: **Selecting proteins with improved stability by a phage-based method.** *Nat Biotechnol* 1998, **16**:955-960.

32. Parsell DA, Sauer RT: **The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli***. *J Biol Chem* 1989, **264**:7590-7595.
33. Schmidt-Dannert C, Arnold FH: **Directed evolution of industrial enzymes**. *Trends Biotechnol* 1999, **17**:135-136.
34. Aharoni A, Griffiths AD, Tawfik DS: **High-throughput screens and selections of enzyme encoding genes**. *Curr Opin Chem Biol* 2005, in press.
35. Goddard JP, Reymond JL: **Enzyme assays for high-throughput screening**. *Curr Opin Biotechnol* 2004, **15**:314-322.
36. Goddard JP, Reymond JL: **Recent advances in enzyme assays**. *Trends Biotechnol* 2004, **22**:363-370.
37. Bornscheuer UT, Pohl M: **Improved biocatalysts by directed evolution and rational protein design**. *Curr Opin Chem Biol* 2001, **5**:137-143.
38. Baik SH, Ide T, Yoshida H, Kagami O, Harayama S: **Significantly enhanced stability of glucose dehydrogenase by directed evolution**. *Appl Microbiol Biotechnol* 2003, **61**:329-335.
39. Harel M, Aharoni A, Gaidukov L, Brumshtein B, Khersonsky O, Meged R, Dvir H, Ravelli RB, McCarthy A, Toker L *et al*: **Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes**. *Nat Struct Mol Biol* 2004, **11**:412-419.
- Directed evolution was applied to reveal the structural fold of the PON family and its enzymatic mechanism. The crystal structure of a directly evolved PON1 variant was solved, providing the first structure of a PON family member.
40. Fokine A, Morales R, Contreras-Martel C, Carpentier P, Renault F, Rochu D, Chabriere E: **Direct phasing at low resolution of a protein copurified with human paraoxonase (PON1)**. *Acta Crystallogr D Biol Crystallogr* 2003, **59**:2083-2087.
41. Muhlrad D, Hunter R, Parker R: **A rapid method for localized mutagenesis of yeast genes**. *Yeast* 1992, **8**:79-82.
42. Ma H, Kunes S, Schatz PJ, Botstein D: **Plasmid construction by homologous recombination in yeast**. *Gene* 1987, **58**:201-216.
43. Wigley WC, Corboy MJ, Cutler TD, Thibodeau PH, Oldan J, Lee MG, Rizo J, Hunt JF, Thomas PJ: **A protein sequence that can encode native structure by disfavoring alternate conformations**. *Nat Struct Biol* 2002, **9**:381-388.
44. Jones HE, Holland IB, Campbell AK: **Direct measurement of free Ca(2+) shows different regulation of Ca(2+) between the periplasm and the cytosol of *Escherichia coli***. *Cell Calcium* 2002, **32**:183-192.
45. Jermutus L, Honegger A, Schwesinger F, Hanes J, Pluckthun A: **Tailoring *in vitro* evolution for protein affinity or stability**. *Proc Natl Acad Sci USA* 2001, **98**:75-80.
46. Sun L, Petrounia IP, Yagasaki M, Bandara G, Arnold FH: **Expression and stabilization of galactose oxidase in *Escherichia coli* by directed evolution**. *Protein Eng* 2001, **14**:699-704.
47. Oh KH, Nam SH, Kim HS: **Directed evolution of N-carbamyl-D-amino acid amidohydrolase for simultaneous improvement of oxidative and thermal stability**. *Biotechnol Prog* 2002, **18**:413-417.
48. Pedersen JS, Otzen DE, Kristensen P: **Directed evolution of barnase stability using proteolytic selection**. *J Mol Biol* 2002, **323**:115-123.
49. Suen WC, Zhang N, Xiao L, Madison V, Zaks A: **Improved activity and thermostability of *Candida antarctica* lipase B by DNA family shuffling**. *Protein Eng Des Sel* 2004, **17**:133-140.
50. Amin N, Liu AD, Ramer S, Aehle W, Meijer D, Metin M, Wong S, Gualfetti P, Schellenberger V: **Construction of stabilized proteins by combinatorial consensus mutagenesis**. *Protein Eng Des Sel* 2004, in press.
- Genetic diversity of the  $\beta$ -lactamase gene was obtained by replacing non-consensus residues in a combinatorial manner according to a sequence homology alignment. Screening for protein stability yielded stable variants with small but cumulative effects from the acquired consensus mutations.
51. Lill H, Hisabori T, Groth G, Bald D: **A thermostable enzyme as an experimental platform to study properties of less stable homologues**. *Protein Eng Des Sel* 2004, **17**:553-555.
52. Kaderbhai MA, Davey HM, Kaderbhai NN: **A directed evolution strategy for optimized export of recombinant proteins reveals critical determinants for pre-protein discharge**. *Protein Sci* 2004, **13**:2458-2469.
- Directed evolution was applied to optimize a protein intermediate in order to obtain efficient export of secretory cytochrome  $b_5$  to the periplasm of *E coli*, by introducing genetic diversity in the N terminus of the preprotein.