Appl Microbiol Biotechnol (2007) 75:1225–1232
DOI 10.1007/s00253-007-0964-2

MINI-REVIEW

Protein engineering: opportunities and challenges

Matti Leisola · Ossi Turunen

Received: 28 February 2007 / Revised: 20 March 2007 / Accepted: 21 March 2007 / Published online: 3 April 2007

Abstract The extraordinary properties of natural proteins demonstrate that life-like protein engineering is both achievable and valuable. Rapid progress and impressive results have been made towards this goal using rational design and random techniques or a combination of both. However, we still do not have a general theory on how to specify a structure that is suited to a target function nor can we specify a sequence that folds to a target structure. There is also overreliance on the Darwinian blind search to obtain practical results. In the long run, random methods cannot replace insight in constructing life-like proteins. For the near future, however, in enzyme development, we need to rely on a combination of both.

Keywords Protein engineering · Directed evolution · Enzymes

Introduction

Proteins, when properly configured, have a remarkable capacity for carrying out complex molecular processes with extreme precision and efficiency. If humans could master protein engineering at this level, the implications would be staggering. Where do we stand in this learning process? The answer depends on our point of reference. Relative to the starting point, it is important to ask how far we have come and to keep track of significant accomplishments. To that end, there has been a steady succession of thorough reviews covering all aspects of protein design, selection, and modification (Bolon et al. 2002; Cherry and Fidantsef 2003; Khersonsky et al. 2006; Hibbert and Dalby 2005; Johannes and Zhao 2006; Rubin-Pitel and Zhao 2006; Wong et al. 2006; Butterfoss and Kuhlman 2006; Pleiss 2006; Bommarius et al. 2006). Equally important, though, is an assessment of progress from the other point of reference—the end point. From this perspective, the question becomes: How far have we come? This review will aim to answer this and to assess the most significant obstacles to progress.

Structural or mechanical functions of proteins depend both on the overall shape and on the material properties (e.g., rigidity, elasticity, and adhesion) in ways that enzymatic functions typically do not. Conversely, enzymatic functions are uniquely dependent on local geometry and chemical environment within their active sites. A number of other categorical distinctions present themselves, such as between the mechanical roles of extended fibrous structures like that of collagen and those of globular structures like flagellin or between such disparate nonmechanical roles as photon capture, electron conduction, ion transport, chemical catalysis, signal transduction, and chaperone-assisted protein folding. The physics of these processes is so varied that we should expect considerably different design rules to apply. Enzymes pose a particular challenge, in that all the action appears to be happening in a small part of the structure. It is much easier to understand the roles of a handful of active-site residues than it is to understand the functional requirements of the scaffold that gives shape to the active site. In this review, we shall concentrate mainly on enzymes.

Competing engineering principles: Diversity and specificity

Ultimately, the objective is to make proteins perform for us as well as they perform in life. The variety of methodolog-

M. Leisola (✉) · O. Turunen
Laboratory of Bioprocess Engineering,
Helsinki University of Technology,
P.O. Box 6100, 02015 HUT Espoo, Finland
e-mail: matti.leisola@tkk.fi

ical approaches aiming to achieve this may be categorized, roughly, along a spectrum defined by two extremes (Fig. 1). At one end is an approach commonly referred to as a rational design, which aims to understand the principles of protein structure and function well enough to apply them in designing new properties or even novel proteins using de novo design. The value of this approach in purely scientific terms is indisputable. However, because the difficulty is likewise indisputable, any approach that might succeed sooner is worth exploring. That realization has motivated work at the other end of the spectrum, where the emphasis is on *finding* what works rather than predicting what works. Darwinian evolution is the inspiration behind this. In the extreme form, this means avoiding protein design principles altogether and relying instead on huge sequence libraries and carefully designed selection methods.

Random search

An average-sized protein has about 300 amino acids and can be put together in an enormous number of different ways. The number of different possibilities is beyond comprehension (20^{300}). This enormous sea of possibilities forms the sequence space. The idea that life-like properties might be reasonably common among random polypeptides dates back to the 1950s, with the work of Sidney Fox on proteinoid microspheres (Fox 1980). Where does this idea stand now, 50 years later?

The most prominent recent study arguing this position is that of Keefe and Szostak (2001). Using a powerful in vitro selection method that links protein chains to their encoding mRNA (Roberts and Szostak 1997), an initial library of nearly 10^{13} protein sequences was cycled through eight rounds of selection on an ATP-affinity column. Then, after several rounds of mutagenesis and 17 more cycles of affinity selection, the authors ended up with a protein that shows significant ATP-binding activity despite having only marginal structural stability. A shortened version of that protein, called artificial nucleotide-binding protein (ANBP), was found to have sufficient structural integrity for crystallization, resulting in the first reported structure of a protein with a random-library pedigree (Lo Surdo et al. 2004). However, when compared to ATP-binding enzymes, the ANBP fold has very limited potential as an enzyme, even considering the possibility of amino-acid substitutions near the binding site.

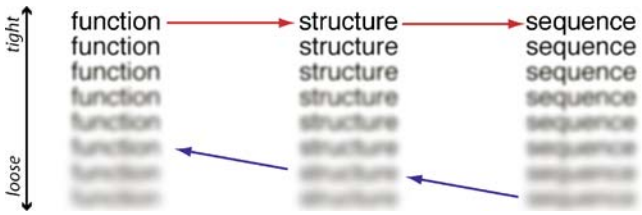

Fig. 1 Rational design goes from function to structure to sequence, while random approaches try to create a sequence that forms a structure producing the desired function

A good example of the great distances between functional islands in sequence space comes from the studies with β-lactamases. A natural variation of β-lactamase has occurred in response to different penicillin derivatives, cephamycins, and four generations of cephalosporins. Resistance to all the different types of antibiotics is the result of only 13 point mutations of 290 residues (Orencia et al. 2001). When structural and genetic variation among β-lactamases from different sources was analyzed, it was concluded that out of 10^{199} possible structures, 10^{122} are active (Axe 2004). Thus, only 1 out of 10^{77} structures is functional.

Functional native-like proteins seem thus to be very rare in random sequence libraries, and if there are proteins that show folding, solubility tends to be a problem (Doi et al. 2005). Even when randomly generated, sequences were fused 50/50 to an N-terminal half of the cold shock protein CspA, a very low amount (1 in 10^7) of folded proteins was obtained (Riechmann and Winter 2000). In conclusion, a fully random search is hardly a way to create novel proteins for biotechnological use. This does not, however, prevent us from utilizing a random search in focused areas.

Randomizing within a local sequence space

Protein engineers can point to solutions employed in life, but we have hardly begun to ask how it is that one protein fold lends itself to a certain class of applications but not to others. We have categories for fold shapes but neither like a theory of structural mechanics or dynamics for proteins nor a theory of active-site design for enzymes; although, we have information on the reaction chemistry of the active site of enzymes. Justifiably, the first priorities have been the structural and functional characterizations of biological proteins and protein complexes. At this point, though, data collection has progressed far beyond data assimilation, making the need of a theory increasingly important. This means that we do not have a theoretical framework for identifying the right protein fold to accomplish a new functional objective. This is the reason why most enzyme engineers prefer directed evolution approaches in improving enzyme activities.

Random or directed evolution methods involve a set of techniques to create random changes in a protein structure by genetic methods and selection of new protein variants. The gene of interest is diversified through mutations, and the created library of mutated genes is tested against a specific selection pressure, for example a particular prop-

erty of a protein. The best variants are iteratively subjected to a new round of mutations. The selection is directed towards a desired activity, hence the name-directed evolution. The diversity of the library can be produced with random point mutations, recombination of genes (homologous or nonhomologous), or a combination of both. Arnold (2007) describes the requirements for successful directed evolution as follows:

1. The desired function must be physically possible.
2. The function must also be biologically or evolutionarily feasible. In practice, this means that there exists a mutational pathway to get from here to there through ever-improving variants.
3. You must be able to make libraries of mutants complex enough to contain rare beneficial mutations.
4. You must have a rapid screen or selection that reflects the desired function.

Requirement no. 2 (*There must be a mutational pathway through ever-improving variants*) limits the search to a functional/structural island. Therefore, the starting point is always a functional protein. Bearing this in mind, some impressive practical achievements (Table 1) have been done using directed evolution methodologies.

Enzymes by design

A rational design of an enzyme catalyst from scratch is a demanding task because the catalyst must have a stable soluble scaffold, and it has to be able to bind and orient both substrates and transition states to catalytic groups. Recently, some remarkable successes have been reported. Kaplan and DeGrado (2004) succeeded in creating de novo-designed diiron proteins with phenol oxidase activity from helical peptides using computational methods and screening of a large number of variants. Kuhlman et al. (2003) used computational strategy that iterated between sequence and structure prediction and designed a stable 93-residue α/β protein fold not found in nature. Both of these studies demonstrate that it is becoming possible to create de novo protein designs, although the created structures still fall far behind the properties of life-like enzymes. Walter et al. (2005) constructed an active enzyme from only nine different amino acids, and de novo proteins have been found from a designed combinatorial library with a quite high frequency (e.g., Wei et al. 2003). Hecht et al. (2004) reviewed the progress in de novo protein design. As interesting as these developments are, practical applications most likely require much more work than the use of traditional methods starting from already functional enzymes.

Another way forward is the rational redesign approach, which means putting a new function on a natural scaffold. The TIM barrel, which was first discovered in triose phosphate isomerase, is a good example of this approach (Sterner and Höcker 2005). About 10% of all known protein structures contain at least one TIM barrel domain in which the active site is formed of or surrounded by connecting loops from eight beta strands and eight helices. Using a computer-based rational redesign, Dwyer et al. (2004) turned ribose-binding protein into a triose phosphate isomerase enzyme, and Bolon and Mayo (2001) redesigned a catalytically inert 108-residue thioredoxin scaffold to function as a *p*-nitrophenyl acetate hydrolyzing enzyme. Despite of some successes in altering enzyme activities, it is often very difficult to attain the results predicted by the rational design. Corey and Corey (1996) listed several failed attempts to produce de novo-designed biocatalysts.

Rational design has been mainly used in biotechnology to improve the properties (especially thermostability) of natural enzymes. The analysis of proteins for rational design involves sequence comparison of the protein family, which gives information about the structurally important amino acids and variability at each amino acid site. In engineering, e.g.,

Table 1 Examples of protein engineering

Target	Enzyme	Result	Method	Reference
Enantioselectivity	Epoxide hydrolase	13-fold improvement	epPCR; DNA shuffling	van Loo et al. (2004)
Increase in promiscuous activity	Carbonic anhydrase	40-fold increase with 2-naphthyl acetate	Mutagenesis+recombination	Gould and Tawfik (2005)
Catalytic efficiency	Glyphosate *N*-acetyltransferase	10,000-fold increase	Eleven rounds of DNA shuffling	Castle et al. (2004)
Thermostability	Xylanase	T_m increased 35°C	Site saturation mutagenesis™	Palackal et al. (2004)
	Phytase	T_m increased 33°C	Consensus method	Lehmann et al. (2002)
Stability in organic solvent	Subtilisin E	170-fold increase in 60% dimethylformamide	Error-prone PCR+screening	Chen and Arnold (1993)
Antibiotic resistance against cetotaxime	β-Lactamase	32,000-fold increase	DNA shuffling	Stemmer (1994)
Cofactor dependency	Lactate dehydrogenase	Specificity from NAD to NADP	Consensus approach	Flores and Ellington (2005)

thermostability, earlier mutational studies, comparison of amino acid sequences, total amino acid content, and crystal structures between mesophilic and thermophilic enzymes may give information on the key factors behind the elevated stability (Hakulinen et al. 2003; Eijsink et al. 2004). When the protein structure is known, computer simulations can be used to study the active site properties, substrate binding, thermostability, and unfolding of the enzymes. Simulations provide information that is useful in planning mutations. For example, molecular dynamics simulations have been used to identify flexible regions in proteins (Daggett and Levitt 1993; Pikkemaat et al. 2002), and subsequently, the protein stability has been increased considerably by introducing a disulfide bridge into such a region.

One of the simplest stabilization methods is actually the construction of a disulfide bridge into a protein structure. It has to be done by rational design because random mutations are not likely to form the required simultaneous double mutation. The thermostability problems of an industrial enzyme have been solved efficiently by this method (Fenel et al. 2004), resulting in over 20°C thermostability increase when two disulfide bridges and other mutations were combined (Xiong et al. 2004). One of the most impressive thermostability increases has been obtained by a single point mutation (Glu→Gln); the thermostability of triosephosphate isomerase increased by 26°C (Williams et al. 1999).

A new semirational method to improve enzyme stability is to make a consensus sequence to the protein family. This approach is based on the assumption that conserved amino acid properties have been selected in nature because of their impact on protein stability. The consensus method efficiently explores the local sequence space. About 30°C increase in thermostability has been achieved by this method (Lehmann and Wyss 2001). In limited areas, rational design methods can be very effective in improving the properties of industrial enzymes.

Engineering the protein fold

Basic structural scaffolds (on the order of few thousands) represent only a tiny fraction of the sequence space. Although alteration of enzyme function has become a routine operation, the reshaping of the active center by modifying the basic fold is a different story. Very few experiments have been carried out trying to change one fold or basic scaffold to another.

Blanco et al. (1999) studied experimentally the sequence space between two different small proteins having different folds. One was a 62-amino acid protein that folds as an eight-stranded orthogonal β-sheet sandwich and another was a 57-amino acid protein that has a central α-helix packed against a four-stranded β-sheet. The authors designed a gradual series of mutants in trying to understand whether there would be an evolutionary path from one fold to another. The conclusion of their study was that the sequence space between the two proteins is enormous. The results suggested that only a small fraction of this space would have adequate properties for folding into a unique structure. The sequence spaces of the two small proteins did not overlap and a change from one fold to another could not be reached within a valid evolutionary trajectory.

However, there are some studies that demonstrate a fold change as a result of a single or few mutations. Cordes et al. (2000) described a single point mutation that changed a part of a small 39-amino acid protein fold from β-sheet to α-helix. However, it is not evident what the selective advantage of a mutation causing structural heterogeneity in natural environment would be. Structural intermediate is not necessarily the same thing as evolutionary intermediate. Meier et al. (2007) studied a small 27-amino acid cysteine-rich polypeptide. The secondary structure of this small peptide is limited to a single helix turn, and the structure is held together by two disulfide bonds. Alternative disulfide patterns can form different folded structures in this type of atypical polypeptide; the formation is controlled by a key amino acid site. Such abrupt folding changes are more likely in small peptides than in large proteins, in which "misfolding" typically leads to a nonfunctional protein.

Alexander et al. (2005) studied two different bacterial IgG binding domains that were nonhomologous and had different folds but were similar in size and function. Their sequence identity is 14%. By directed evolution approach, the authors changed the sequence of one of the domains so that the sequence identity was finally 59%. Both domains, however, still retained their original fold, although the binary sequence space separating these two domains had decreased almost eight orders of magnitude. The authors concluded that the folding information must be in the nonidentical parts of the domains.

A change in folded structure is sometimes possible through a short walk in the sequence space. When occurring at the posttranslational level, this kind of event is usually called miss-folding. To get something meaningful from genetically guided "miss-folding," it must give selective advantage to the organism. At the moment, we can only expect that huge sequence libraries are needed to find novel functions through fold change events.

Obstacles in protein engineering

In view of the very substantial challenges remaining and the considerable effort expended thus far, we should pause to ask what things are most impeding our progress. Questions of this kind always reveal diversity of opinion, which is not

a bad thing. In offering our opinions here, we hope merely to stimulate an important discussion that might enhance our collective progress. With that in mind, we suggest the following as the most significant obstacles to be overcome:

1. Lack of a theory for structure design (i.e., specifying a structure that is suited to a target function)
2. Lack of a general approach for sequence design (i.e., specifying a sequence that folds to a target structure)
3. Overreliance on the Darwinian methodology

There is information about the rules on how the catalytic amino acid residues have to be located in three dimensions for the enzymatic reaction to happen. This knowledge can be used to design active sites (Dwyer et al. 2004). Similarly, we have some ideas of protein scaffolds into which an active site can be grafted. Despite of all the progress, we are in the very beginning in designing de novo active enzymes. Thus, we are still missing general theories that would help us to design novel enzymes without a need to use methods that are based on a random search in the local sequence space.

It is often said that random genetic methods to improve enzyme properties "rely on simple but powerful Darwinian principles of mutation and selection" (Johannes and Zhao 2006). We agree. It is also said that "every protein has become adapted by step-by-step improvement and refinement of its function over millions of years" (McLachlan 1987). The present theories, however, only partly explain the protein diversity, although a recent study (Poelwijk et al. 2007) shows that even the key-lock dilemma can be resolved by the Darwinian approach when the operation field for random search is within the same protein family, and the new key-lock pair closely resembles the original (ancestral).

As discussed above, the transition from one fold to another is very problematic. Even inside a fold family, most transitions are very challenging. For example, family 10 xylanase and xylose isomerase (XI) have a TIM barrel fold, but the active site of xylanase is open and large, whereas the active site of XI is buried inside the protein so that the loops between secondary structure elements cover the active site. In xylanase, a long filament has to fit into the active site, and in XI, the isomerization of sugars has to happen in an environment in which the amount and positioning of water are controlled. These two examples demonstrate the wide diversity potential of the TIM barrel fold. It is also evident that evolutionary transitions in this fold family can be very complex. Development of loop grafting libraries might produce interesting results in TIM barrel scaffold.

Gene duplication

Gene duplication and subsequent divergence as mechanisms to create natural variety and novel structures are now decade's old theories (Ohno 1970). Basically, directed evolution approach is an application of the gene duplication concept. Gene duplication is seen as a way to avoid random sequences in evolution, because random sequences most often are not functional. Mutations in the duplicated genes explore the local sequence space and expand the number of members in a gene family.

It appears, however, that the full potential of duplicated genes are not effectively explored in nature, as only a small number of duplicated genes do not experience a dead end (Lynch and Conery 2000). Positive selection may be needed to retain the duplicated genes in the population (Ohta 2002; Kondrashov and Kondrashov 2006; Shiu et al. 2006). Divergence after gene duplication, sometimes under lowered selection pressure, may create new activities via amino acid substitutions or small deletions and insertions. However, small amino acid changes are not likely to form a new protein fold (Bogarad and Deem 1999). This can be easily understood as even after, e.g., 70% divergence, the protein is still likely to retain its original fold (Chothia and Lesk 1986). Park et al. (2006) have demonstrated that the task is not easy even for a protein engineer; even minor changes in the scaffold need huge design efforts. Thus, gene duplication methods can be used to find novel solutions but probably mainly inside a gene family.

Exon shuffling

Exon shuffling (and shuffling of protein blocks) is often considered as a more powerful evolutionary method to create novel proteins than gene duplication. Exon shuffling basically applies a gene duplication concept but in smaller units. It seems that in practice, single units (exons or protein blocks) must be fused to a target gene sequentially and not several exons from different sources in a single event (Peisajovich et al. 2006). Too radical changes disrupt the protein function (and protection by selection) or cause damage in the genome. Because of these reasons, the exon shuffling process has to occur in reality so that the overall protein fold does not change.

DNA shuffling methods based on homologous recombination keep the resulting hybrid genes inside the original gene family. Nonhomologous recombination, like exon (or protein block) shuffling can, in principle, create proteins with novel folds, provided that the new proteins have such an activity that a screening/selection method can detect them. Nonhomologous random recombination has been studied in vitro, e.g., with chorismate mutase (CM; Bittker et al. 2004). Functional enzyme variants contained insertions, deletions, and rearrangements. The authors also randomly recombined CM with fumarase—another α-helical protein. The resulting active CM proteins contained

the original CM core and fumarase sequences only at the termini or one loop. This experiment demonstrates that the destruction of the stable core leads to the loss of activity supporting the view that protein scaffold can change mainly by incremental modifications of the fold. DNA shuffling methods based on nonhomologous recombination may find fully novel functions, but apparently, a very large number of variants have to be screened. This creates a practical limit that may not be easily broken with current methods.

Novelty through promiscuous activities

Enzymes are not always too accurate in their substrate specificity. Pastinen et al. (1999), for instance, showed that the industrially used glucose (xylose) isomerase has a number of side activities isomerizing a large number of both natural and rare hexoses and pentoses. Such weak side activities can be improved by design (Karimäki et al. 2004) or by random methods. It has been proposed that in nature, a new function is evolved from an initially promiscuous activity (Aharoni et al. 2005). Khersonsky et al. (2006) reviewed the present views of divergence of today's enzymes from ancestral proteins catalyzing a whole range of activities at low levels. A promiscuous activity can be a promising starting point for directed evolution methods to produce biotechnologically relevant enzymes (Bornscheuer and Kazlauskas 2004).

A way forward: hybrid approaches

Practical experience shows that directed evolution can produce remarkable changes that are, at present, not easily achieved by rational design. However, these methods have their limits as discussed above. Furthermore, selection and screening in directed evolution methods require that each step improves the enzyme. When the desired change involves a simultaneous change in several amino acids, it is not likely to be reached by the random approach (Behe and Snoke 2004). In many cases, a combination of design to create the needed structure or function and its improvement by random techniques is a better approach. As Taylor et al. (2001) put it: "The low frequency of protein catalysts in sequence space indicates that it will not be possible to isolate enzymes from unbiased random libraries in a single step." Recently, Park et al. (2006) changed a metallohydrolase through designed deletion and insertion of several structural loops in the active site to form a new enzyme with a different catalytic function and then applied random techniques to improve the designed activity.

To focus mutations near the active site when trying to change the catalytic properties of an enzyme appears to be more effective than distant mutations (Morley and Kazlauskas 2005). Yoshikuni et al. (2006) used saturation mutagenesis and systematic recombination approach inside or near the active site to expand the synthetic capacity of γ-humulene synthase to produce different sesquiterpenes. The engineered enzyme variants used different reaction pathways to synthesize a variety of reaction products while maintaining the specific activity of the original enzyme. The remarkable results in changing the active site properties by a blind search in a limited number of amino acids indicate that there exists a number of biotechnologically interesting enzyme variants beyond the sequence space that can be easily reached by the directed evolution method in which the whole gene is a target, and tens or hundreds of thousands of variants has to be screened. Computational methods may also be increasingly used in reducing the sequence space that is screened by high-throughput screening methods (Voigt et al. 2001; Hayes et al. 2002).

Bloom et al. (2006) recently demonstrated that the stability of the protein scaffold improves its evolvability. In other words, when the stability of a protein scaffold is increased, it may be possible to make mutations that create a truly novel property, which without increased stability would be too harmful for the protein. This approach can be very promising in finding novel enzymatic solutions that otherwise cannot be detected in the functional screening.

Conclusions

Twenty years of protein engineering has resulted in an impressive array of genetic engineering and computational tools as well as several concrete results in modifying and improving the properties of enzymes. We have learned to understand, on one hand, the rarity of functional proteins in the sequence space and on the other hand, the large variation potential of biotechnologically relevant protein functions. We only need methods to dig them out. We know superficially how a large number of life-like proteins look like, but we are still far from understanding how life-like proteins are designed and even further from being able to design life-like nonnatural proteins.

In spite of the progress, we still do not have a general theory on how a sequence produces a specific structure and how a structure determines a function. Therefore, a blind Darwinian search within a known protein scaffold is often used to modify proteins. Unfortunately, blind searches have hard resource limits whereas insight has not. Therefore, in the long run, blind searches are of limited value in compensating our present ignorance. We still have a long way to go before we are able to design a suitable protein scaffold, position binding, and catalytic groups correctly into this scaffold and optimize the designed protein for life-like efficiency.

Acknowledgment The authors thank Douglas Axe for his helpful criticism and for revising the text.

References

Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS (2005) The 'evolvability' of promiscuous protein functions. Nat Genet 37:73–76

Alexander PA, Rozak DA, Orban J, Bryan PN (2005) Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code. Biochemistry 44:14045–14054

Arnold (2007) Directed enzyme evolution http://www.che.caltech.edu/groups/fha/directed_evolution.html

Axe D (2004) Estimating the prevalence of protein sequences adopting functional enzyme folds. J Mol Biol 341:1295–1315

Behe MJ, Snoke DW (2004) Simulating evolution by gene duplication of protein features that require multiple amino acid residues. Protein Science 13:2651–2664

Bittker JA, Le BV, Liu JM, Liu DR (2004) Directed evolution of protein enzymes using nonhomologous random recombination. Proc Natl Acad Sci USA 101:7011–7016

Blanco FJ, Angrand I, Serrano L (1999) Exploring the conformational properties of the sequence space between two proteins with different folds: an experimental study. J Mol Biol 285:741–753

Bloom JD, Labthavikul ST, Otey CR, Arnold FA (2006) Protein stability promotes evolvability. Proc Natl Acad Sci USA 103:5869–5874

Bogarad LD, Deem MW (1999) A hierarchical approach to protein molecular evolution. Proc Natl Acad Sci USA 96:2591–2595

Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. Proc Natl Acad Sci USA 98:14274–14279

Bolon DN, Voigt CA, Mayo SL (2002) De novo design of biocatalysts. Curr Opin Chem Biol 6:125–129

Bommarius AS, Broering JM, Chaparro-Riggers JF, Polizzi KM (2006) High-throughput screening for enhanced protein stability. Curr Opin Biotechnol 17:606–610

Bornscheuer UT, Kazlauskas RJ (2004) Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. Angew Chem Int Ed 43:6032–6040

Butterfoss GL, Kuhlman B (2006) Computer-based design of novel protein structures. Ann Rev Biophys Biomol Struct 35:49–65

Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, Bertain S, Cho HJ, Duck N, Wong J, Liu D, Lassner MW (2004) Discovery and directed evolution of a glyphosate tolerance gene. Science 304:1151–1154

Chen K, Arnold FH (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of Subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA 90:5618–5622

Cherry JR, Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. Curr Opin Biotechnol 14:438–443

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826

Cordes MHJ, Burton RE, Walsh NP, McKnight CJ, Sauer RT (2000) An evolutionary bridge to a new protein fold. Nature Struct Biol 7(12):1129–1132

Corey MJ, Corey E (1996) On the failure of de novo-designed peptides as biocatalysts. Proc Natl Acad Sci USA 93:11428–11434

Daggett V, Levitt M (1993) Protein unfolding pathways explored through molecular dynamics simulations. J Mol Biol 232:600–619

Doi N, Kakukawa K, Oishi Y, Yanagawa H (2005) High solubility of random-sequence proteins consisting of five kinds of primitive amino acids. Prot Eng Des Sel 18:279–284

Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. Science 304:1967–1971

Eijsink VGH, Gåseidnes S, Synstad B, Bjørk A, Sirevåg R, Van den Burg B, Vriend G (2004) Rational engineering of enzyme stability. J Biotechnol 113:105–120

Fenel F, Leisola M, Jänis J, Turunen O (2004) A de novo designed N-terminal disulfide bridge stabilizes the *Trichoderma reesei endo*-1, 4-b-xylanase II. J Biotechnol 108:137–143

Flores H, Ellington AD (2005) A modified consensus approach to mutagenesis inverts the cofactor specificity of *Bacillus stearothermophilus* lactate dehydrogenase. Prot Eng Des Sel 18:369–377

Fox SW (1980) Metabolic microspheres. Origins and evolution. Naturwissenschaften 67:378–383

Gould SM, Tawfik DS (2005) Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. Biochem 44:5444–5452

Hakulinen N, Turunen O, Jänis J, Leisola M, Rouvinen J (2003) Three-dimensional structures of thermophilic β-1,4-xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa*. Eur J Biochem 270:1399–1412

Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, KUndu A, Dahiyat BI (2002) Combining computational and experimental screening for rapid optimization of protein properties. Proc Natl Acad Sci USA 99:15926–15931

Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. Protein Sci 13:1711–1723

Hibbert EG, Dalby PA (2005) Directed evolution strategies for improved enzymatic performance. Microbial Cell Fact 4:29

Johannes TW, Zhao H (2006) Directed evolution of enzymes and biosynthetic pathways. Curr Opin Microbiol 9:261–267

Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. Proc Natl Acad Sci USA 101:11566–11570

Karimäki J, Parkkinen T, Santa H, Pastinen O, Leisola M, Rouvinen J, Turunen O (2004) Crystallographic, molecular dynamics simulation and site-directed mutagenesis study of the reaction of D-xylose isomerase with L-arabinose. Prot Eng Des Select 17:861–869

Keefe AD, Szostak JW (2001) Functional proteins from a random-sequence library. Nature 410:715–718

Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Curr Opin Chem Biol 10:498–508

Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. J Theor Biol 239:141–151

Kuhlman B, Dantas G, Ireton, GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302:1364–1368

Lehmann M, Wyss M (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr Opin Biotechnol 12:371–375

Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon A, Wyss M (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. Prot Eng 15:403–411

Lo Surdo P, Walsh MA, Sollazo M (2004) A novel ADP- and zinc-binding fold from function-directed in vitro evolution. Nat Struct Mol Biol 11:382–383

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

McLachlan AD (1987) Gene duplication and the origin of repetitive protein structures. In: Cold Spring Harbor symposium on quantitative biology, vol. LII. Cold Spring Harbor laboratory, Cold Spring Harbor, p. 411–420

Meier S, Jensen PR, David CN, Chapman J, Holstein TW, Grzesiek S, Ozbek S (2007) Continuous molecular evolution of protein-domain structures by single amino acid changes. Curr Biol 17:173–178

Morley KL, Kazlauskas RJ (2005) Improving enzyme properties: when are closer mutations better? Trends Biotechnol 23:231–237

Ohno S (1970) Evolution by gene duplication. Springer, Berlin Heidelberg New York

Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. Proc Natl Acad Sci USA 99:16134–16137

Orencia MC, Yoon JS, Ness JE, Stemmer WPC, Stevens RC (2001) Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nature Struct Biol 8:238–242

Palackal N, Brennan Y, Callen WN, Dupree P, Frey G, Goubet F, Hazlewood GP, Healey S, Kang YE, Kretz KA, Lee E, Tan X, Tomlinson GL, Verruto J, Wong VW, Mathur EJ, Short JM, Robertson DE, Steer BA (2004) An evolutionary route to xylanase process fitness. Protein Sci 13:494–503

Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS (2006) Design and evolution of new catalytic activity with an existing protein scaffold. Science 311:535–538

Pastinen O, Visuri K, Schoemaker H, Leisola M (1999) Novel reactions of xylose isomerase from *Streptomyces rubiginosus*. Enzyme Microb Technol 25:695–700

Peisajovich SG, Rockah L, Tawfik DS (2006) Evolution of new protein topologies through multistep gene rearrangements. Nat Genet 38:168–174

Pikkemaat MG, Linssen ABM, Berendsen HJC, Janssen DB (2002) Molecular dynamics simulations as a tool for improving protein stability. Prot Eng 15:185–192

Pleiss J (2006) The promise of synthetic biology. Appl Microbiol Biotechnol 73:735–739

Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445:383–386

Riechmann L, Winter G (2000) Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. Proc Natl Acad Sci USA 97:10068–10073

Roberts RW, Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci USA 94:12297–12302

Rubin-Pitel SB, Zhao H (2006) Recent advances in biocatalysis by directed enzyme evolution. Comb Chem High Throughput Screen 9:247–257

Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH (2006) Role of positive selection in the retention of duplicate genes in mammalian genomes. Proc Natl Acad Sci USA 103:2232–2236

Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391

Sterner R, Höcker B (2005) Catalytic versatility, stability, and evolution of the (β/α)8-barrel enzyme fold. Chem Rev 105:4038–4055

Taylor SV, Walter KU, Kast P, Hilvert D (2001) Searching sequence space for protein catalysts. Proc Natl Acad Sci USA 98:10596–10601

van Loo B, Spelberg JH, Kingma J, Sonke T, Wubbolts MG, Janssen DB (2004) Directed evolution of epoxide hydrolase from A. radiobacter toward higher enantioselectivity by error-prone PCR and DNA shuffling. Chem Biol 11:981–990

Voigt CA, Mayo SL, Arnold FH, Wang Z-G (2001) Computational method to reduce the search space for directed protein evolution. Proc Natl Acad Sci USA 98:3778–3783

Walter KU, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet. J Biol Chem 280:37742–37746

Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Sci 12:92–102

Williams JC, Zeelen JP, Neubauer G, Vriend G, Backmann J, Michels PAM, Lambeir, A-M, Wierenga RK (1999) Structural and mutagenesis studies of *leishmania* triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a super-stable enzyme without losing catalytic power. Prot Eng 12:243–250

Wong TS, Zhurina D, Schwaneberg U (2006) The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen 9:271–288

Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440:1078–1082

Xiong H, Fenel F, Leisola M, Turunen O (2004) Engineering the thermostability of *Trichoderma reesei endo*-1,4-β-xylanase II by combination of disulfide bridges. Extremophiles 8:393–400