

70. van Horck, F. P., Ahmadian, M. R., Haeusler, L. C., Moolenaar, W. H. & Kranenburg, O. Characterization of p190RhoGEF, a RhoA-specific guanine nucleotide exchange factor that interacts with microtubules. *J. Biol. Chem.* **276**, 4948–4956 (2001).
71. Ren, X. D., Kiosses, W. B. & Schwartz, M. A. Regulation of the small GTP-binding protein Rho by cell adhesion and the cytoskeleton. *EMBO J.* **18**, 578–585 (1999).
72. Fukata, Y., Amano, M. & Kaibuchi, K. Rho–Rho-kinase pathway in smooth muscle contraction and cytoskeletal reorganization of non-muscle cells. *Trends Pharmacol. Sci.* **22**, 32–39 (2001).
73. Bershadsky, A. D., Vaisberg, E. A. & Vasiliev, J. M. Pseudopodial activity at the active edge of migrating fibroblast is decreased after drug-induced microtubule depolymerization. *Cell Motil. Cytoskeleton* **19**, 152–158 (1991).
74. Dunn, G. A., Zicha, D. & Fraylich, P. E. Rapid, microtubule-dependent fluctuations of the cell margin. *J. Cell Sci.* **110**, 3091–3098 (1997).
75. Waterman-Storer, C. M., Worthylake, R. A., Liu, B. P., Burridge, K. & Salmon, E. D. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nature Cell Biol.* **1**, 45–50 (1999).
76. Bretscher, M. S. & Aguado-Velasco, C. Membrane traffic during cell locomotion. *Curr. Opin. Cell Biol.* **10**, 537–541 (1998).
77. Toomre, D., Keller, P., White, J., Olivo, J. C. & Simons, K. Dual-color visualization of trans-Golgi network to plasma membrane traffic along microtubules in living cells. *J. Cell Sci.* **112**, 21–33 (1999).
78. Bershadsky, A. D. & Futerman, A. H. Disruption of the Golgi apparatus by brefeldin A blocks cell polarization and inhibits directed cell migration. *Proc. Natl Acad. Sci. USA* **91**, 5686–5689 (1994).
79. Rodionov, V. I. *et al.* Microtubule-dependent control of cell shape and pseudopodial activity is inhibited by the antibody to kinesin motor domain. *J. Cell Biol.* **123**, 1811–1820 (1993).

## Acknowledgements

We thank O. Krylyskina and G. Resch for their invaluable help with the videos and animation. A.B. and B.G. are grateful to D. Riveline and J. Kirchner for providing the experimental data for figures 2 and 3. A. Huttenlocher is acknowledged for providing the DsRed zyxin construct that was used in figure 5. This work was supported in part by a grant from the Austrian Science Research Council to J.V.S. and I.K. B.G. is the incumbent of the E. Neter Chair in Cell and Tumor Biology; A.B. holds the J. Moss Chair of Biomedical Research.

## Online links

## DATABASES

The following terms in this article are linked online to: Swiss-Prot: <http://www.expasy.ch/> Cdc42 | Dia1 | Rho kinase

Access to this interactive links box is free online.

## TIMELINE

## Protein engineering 20 years on

James A. Brannigan and Anthony J. Wilkinson

It is 20 years since site-directed mutagenesis was first used to modify the active site of an enzyme of known structure and mechanism. Since then, this method has contributed far-reaching insights into catalysis, specificity, stability and folding of proteins. Engineered proteins are now being used in industry and for the improved treatment of human disease.

At the beginning of the 1980s, a major stumbling block to progress in biochemistry was our inability to direct chemistry specifically at macromolecular surfaces in a way that allowed the relationship between structure and activity to be examined in detail. Nowhere was this limitation more acutely felt than in the field of enzymology. The principles that govern enzyme catalysis were understood<sup>1,2</sup> and the number of enzyme structures solved by X-ray crystallography was beginning to grow, albeit slowly<sup>3</sup>. These structures, with their stereochemical clarity, provided a framework for formulating mechanisms of action in which precise roles were attributed to functional groups that were pinpointed in the active sites. The natural way to test emerging hypotheses was to

modify these functional groups specifically and to explore the effects on activity. This could only be achieved by painstaking chemical modification with its attendant problems of limited range and poor specificity. In this climate, the arrival of site-directed mutagenesis, a technique that allowed amino-acid sequences in proteins to be altered at will, was the answer to an enzymologist's prayer. The first uses of this technique, to mutate genes that encode enzymes of known mechanism and produce proteins with defined amino-acid residue substitutions, were reported for tyrosyl-transfer RNA synthetase (TyrRS) and  $\beta$ -lactamase at the end of 1982 (REFS 4–6).

These precise changes of only one or two amino-acid residues were later followed by changes of entire loops<sup>7</sup> and even domains<sup>8</sup>. This construction of modified proteins and the analysis of their properties coalesced to form a new field — that of protein engineering. In this article, we have traced the field of protein engineering over the past 20 years, principally following the thread of enzyme engineering that was pioneered in those early papers (TIMELINE). It is not possible in an article of this length to provide comprehensive

coverage of this subject. The examples chosen therefore reflect our own experiences and perspectives, and many significant topics and achievements have necessarily been omitted or abbreviated.

## Kicking off with TyrRS

TyrRS proved to be a fruitful system for the dissection of enzyme catalysis by site-directed mutagenesis. It is a central enzyme in molecular biology and is responsible for ligating the amino-acid tyrosine (Tyr) to its cognate tRNA<sup>Tyr</sup> in an ATP-dependent reaction that produces tyrosyl-tRNA<sup>Tyr</sup>. As with all aminoacyl-tRNA synthetases, the accurate selection of the cognate amino acid is important for the faithful translation of the genetic code. In particular, in the living cell, TyrRS discriminates against the most closely related amino acid to tyrosine — phenylalanine (Phe) — which lacks only the phenolic hydroxyl group of tyrosine.

The *Bacillus stearothermophilus* TyrRS was ripe for engineering. Greg Winter's laboratory had determined the sequence of the enzyme by using a combination of classical protein sequencing and DNA sequencing of the cloned gene, and this gene had also been expressed in *Escherichia coli*. David Blow's laboratory had crystallized the enzyme and was in the final stages of solving its structure. Mechanistic studies of the enzyme in Alan Fersht's laboratory had shown the existence of a remarkably stable aminoacyl-adenylate intermediate. This meant that the enzyme

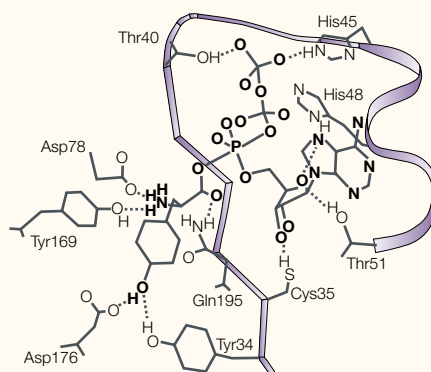
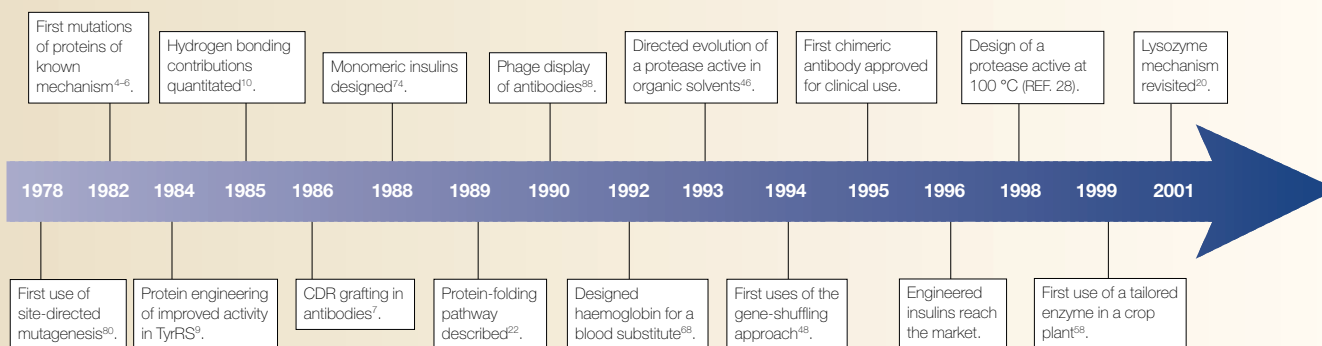


Figure 1 | **The active site of tyrosyl-transfer RNA synthetase.** The modelled structure of the transition state in tyrosyl-adenylate formation is shown. The transition state was extrapolated from the known structure of the enzyme-bound tyrosyl-adenylate. Hydrogen-bonding interactions between the enzyme and transition state species are shown as dashed lines. The roles of threonine (Thr) 40 and histidine (His) 45 are discussed in the main text. Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Tyr, tyrosine. Modified with permission from REF. 10 © (2002) National Academy of Sciences, USA.

Timeline | **Twenty years of protein engineering**

CDR, complementarity-determining regions; TyrRS, tyrosyl-tRNA synthetase.

could be assayed by active-site titration, and therefore accurate and reproducible steady-state kinetic measurements of reaction rates could be made. It also allowed the crystallographers to grow crystals and determine a structure of the enzyme in its complex with tyrosyl-adenylate, which showed that there were a plethora of contacts between the enzyme and the intermediate, including a dozen or so hydrogen bonds (FIG. 1). This detailed understanding of the structure and mechanism of the enzyme meant that site-directed mutagenesis could be used to address specific well-formulated questions. Furthermore, as it was the first enzyme of known structure to be probed by mutagenesis, fundamental questions about the contributions of specific types of interaction to enzyme–substrate binding, and catalysis in general, could be addressed.

The gene for TyrRS was re-cloned into the single-stranded bacteriophage vector M13 to facilitate oligonucleotide-directed mutagenesis. To get first-hand experience of this technique, Greg Winter visited the laboratory of Mike Smith in Vancouver where the technique was being pioneered. The first target for site-directed mutagenesis was cysteine 35 (Cys35), which was chosen because, in the crystal structure, the thiol group of this residue was seen to form a putative hydrogen bond with the ribose portion of the ATP (FIG. 1), and because this residue is conserved in other aminoacyl-tRNA synthetases.

Kinetic measurements were made first on the Cys35Ser mutant (where Ser is serine) of TyrRS in M13-infected *E. coli* cell extracts<sup>4</sup>, and then on purified TyrRS(Ser35) and TyrRS(Gly35) (where Gly is glycine) proteins<sup>9</sup>. The mutant enzymes were indeed less active than the wild-type enzyme, which confirmed the role of Cys35 in the catalytic mechanism. From the kinetic measurements,

it was also possible to show that the extra binding energy ( $\sim 1$  kcal mol<sup>-1</sup>) provided by the Cys35 side chain in the wild-type enzyme is used to stabilize the bound substrate in its transition state rather than its ground state.

This use of binding energy selectively in transition-state stabilization was shown in a dramatic way when modelling and mutagenesis were used to probe the contributions to catalysis of threonine 40 (Thr40) and histidine 45 (His45)<sup>10</sup>. These residues are far away from the seat of the reaction (FIG. 1), but together they contribute 300,000-fold to the catalytic rate, probably by binding the  $\gamma$ -phosphate of the ATP, which swings into the binding site as the transition state is formed.

**“The often-disastrous effects of ‘rationally introduced’ mutations on the stability and activity of proteins prompted Charles Craik to muse that ‘protein terrorism’ was a more suitable descriptor than protein engineering.”**

The contributions of other hydrogen-bonding side chains in TyrRS (FIG. 1) were evaluated systematically using the mutagenesis approach. Substitutions at one position, Thr51, even led to increases in the catalytic activity of TyrRS. The fact that this had, in part, been anticipated, gave us sufficient confidence to use the term ‘protein engineering’ when presenting these observations<sup>11</sup>.

The cumulative data on the hydrogen-

bonding contributions in TyrRS<sup>12</sup> indicated that deletions that leave an unpaired, uncharged hydrogen-bond donor or acceptor weaken binding by  $\sim 1$  kcal mol<sup>-1</sup>, but deletions that leave an unpaired, charged group weaken binding by  $\sim 4$  kcal mol<sup>-1</sup>. The avoidance of unsolvated charge is therefore an important component of biological specificity, which accounts for the presence of aspartate 176 (Asp176) at the base of the amino-acid side-chain binding pocket in TyrRS (FIG. 1). A Tyr ligand with its phenolic hydroxyl group can form a charge-dipole hydrogen bond to the negatively charged carboxylate of Asp176, but a Phe ligand that lacks this group cannot. The binding of Phe to TyrRS would therefore cause Asp176 to be buried in an apolar environment. This is highly unfavourable and largely explains the 10<sup>5</sup>-fold discrimination by TyrRS against Phe.

#### Protein engineering pervades

In the ten years that followed these studies, protein engineering enjoyed something of a golden era. As both gene cloning and the synthesis of oligonucleotides became standard laboratory activities, site-directed mutagenesis became part of the established repertoire of techniques that were used in protein chemistry and enzymology. New strategies that embrace *E. coli* host strains with deficient mismatch repair systems, and methods to select against parental (non-mutant) strands, greatly increased the efficiency of mutagenesis<sup>13</sup>.

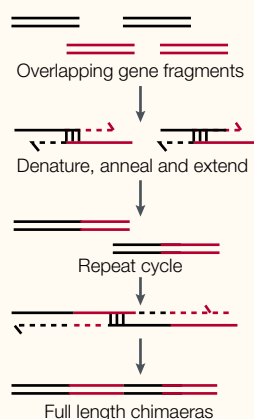
The roles of the ‘most famous’ residues in molecular biology, which are described in detail in numerous biochemistry textbooks, were scrutinized by mutagenesis and their contributions to function were quantitated. These studies variously confirmed prejudices, caused changes in emphasis, or forced com-

## Box 1 | Combinatorial mutagenesis and evolutionary methods

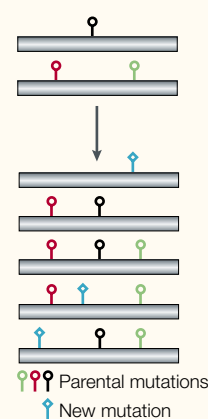
In combinatorial approaches to mutagenesis, small gene portions (see figure part a) are combined in a 'mix-and-match' fashion, the myriad combination of which generates a greater diversity of mutation. These fragments can be from a collection of mutants in a single gene (see figure part b), which leads to new permutations. If further diversity is desirable, a random factor (for example, error-prone polymerase chain reaction (PCR) under conditions that favour poor fidelity in replication) can be introduced to make more mutants, just as somatic hypermutation increases the repertoire of different antibodies from shuffled immunoglobulin genes. When the starting material is a family of genes that encode the same protein from several sources (see figure part c), the process can be considered similar to consensus mutagenesis, where the most typical, naturally occurring substitution at any given position is engineered to result in a protein with an averaged sequence<sup>59</sup>. However, whereas a consensus human  $\alpha$ -interferon has more than tenfold enhanced anti-viral activity, a variant that is derived from the shuffling together of different combinations of naturally occurring substitutions in five parental clones has up to 285,000-fold higher specific activity when compared with the best parental interferon<sup>81</sup>.

The 'sexual' PCR technique has been extended to distantly or completely unrelated gene families, which require methods that do not rely on homologous recombination because of the degree of sequence divergence, and their acronyms are as ingenious as their protocols. 'Sequence homology-independent protein recombination' (SHIPREC)<sup>82</sup> and 'incremental truncation for the creation of hybrid enzymes' (ITCHY)<sup>83</sup> lead to the formation of chimaeras — single-point fusions of unrelated sequences (see figure part d). These might then themselves be rearranged by shuffling (SCRATCHY) to yield functional hybrids, for example, between human and *Escherichia coli* glycinamide-ribonucleotide formyltransferase proteins as proof-of-principle<sup>84</sup>. The main advantage of these methods is that detailed knowledge of protein structure is not required, although such information can be used in the design of crossover points<sup>85</sup>. Typically, highly complex libraries of variants are screened for the desired activity, and a recurring finding is that the mutation combinations found to increase fitness are difficult to predict by rational means.

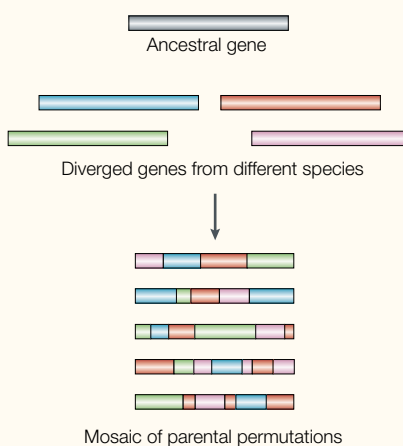
## a Recursive PCR and gene assembly



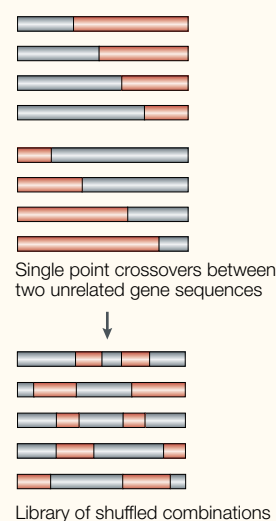
## b Single-gene shuffling



## c Family shuffling



## d The ITCHY and SCRATCHY show



plete revision of accepted mechanisms. For example, the ascribed roles of the residues of the catalytic triad and the oxyanion hole in serine proteases were substantiated<sup>14–16</sup>. In myoglobin, the crucial role of the distal histidine in discriminating in favour of binding oxygen ( $O_2$ ) and against binding carbon monoxide (CO) was confirmed. However, contrary to expectations, the stabilizing influence of hydrogen bonding on  $O_2$  binding was shown to outweigh the destabilizing influence of steric hindrance on bound CO (REF. 17). A famous 'dog that didn't bark', however, was the Tyr248 residue in carboxypeptidase. Its mutation to Phe, had no effect on enzyme activity, which discounted the possibility of acid catalysis by this residue<sup>18</sup>. Another textbook favourite that had its mechanism revised recently was hen egg-white lysozyme

— the first enzyme to have its structure determined by X-ray crystallography<sup>19</sup>. A glycosyl-enzyme intermediate was shown in a Glu35Gln mutant (where Glu is glutamate and Gln is glutamine), which dispelled the 30-year notion of a long-lived oxocarbenium-ion intermediate<sup>20</sup>.

Elsewhere, model systems were chosen carefully to study allostery, stability, folding and specificity, which provided fundamental insights into these phenomena<sup>21–24</sup>. The use of fluorescence and NMR spectroscopy techniques to monitor the refolding pathways in mutated proteins led to the identification of nucleation cores through which proteins attain their functional folded forms<sup>22</sup>, as well as to complete descriptions of protein-folding pathways<sup>23</sup>. The energetic contributions of electrostatic and apolar/hydrophobic

interactions were quantified, which provided a further platform for understanding both protein folding and stability<sup>24,25</sup>. The ground rules that were laid out in this way led to the design of enzymes with greater activity in non-aqueous solvents<sup>26</sup>, increased thermostability<sup>27,28</sup> and altered pH optima<sup>29</sup>. The ability to manipulate these characteristics is important for the engineering of industrial enzymes that are often required to function under conditions that are remote from those in the living cells for which they are adapted. Where the individual contributions of alterations could be measured, the frequent finding that the effects of non-interacting mutations were mostly additive indicated that progress was being made in understanding the links between protein sequence, form, folding and function<sup>30,31</sup>.

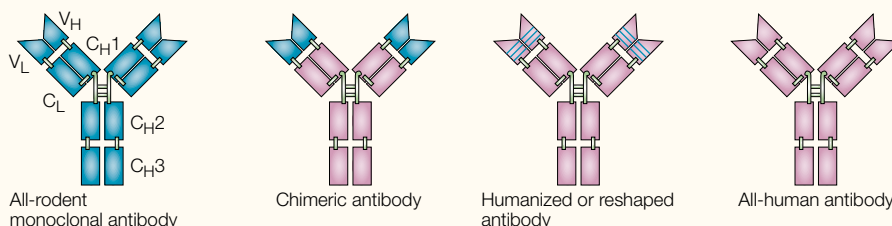
Alterations in substrate specificity were also achieved notably in lactate dehydrogenase, which was analysed extensively. A single mutation shifted the selectivity from pyruvate to oxaloacetate, which generated a malate dehydrogenase<sup>32</sup>; five mutations led to an enzyme that had a broader specificity and accommodated branched hydroxy-acids<sup>33</sup>; and the replacement of a protein loop that normally encloses the active site increased selectivity for the more bulky molecule phenylpyruvate. These experiments showed that specificity could be re-engineered, which led, in the latter case, to an enzyme activity with a possible application in monitoring phenylketonuria<sup>34</sup>.

It was not all plain sailing, however, as it became clear that although the local interactions in an enzyme were a dominant factor in shaping substrate specificity, they were not the sole determinant. This was exemplified by the difficulties that were encountered in converting the substrate specificity of trypsin to that of chymotrypsin. These two enzymes have almost superimposable substrate-binding sites and catalytic geometry, but they cleave peptides on the carboxylate side of positively charged and hydrophobic residues, respectively. Any notion that a single substitution of the negatively charged Asp189 in the proximal side-chain pocket of trypsin would effect this alteration in selectivity was soon dispelled<sup>35</sup>. Instead, 15 further mutations, many of residues that are not in contact with the substrate, were required to adapt the substrate specificity of trypsin to that of chymotrypsin<sup>36,37</sup>. This specificity-transplantation experiment was not significant in terms of potential applications, but it did serve to gauge the progress being made towards understanding the fundamental principles of protein design. It was clear that even conservative mutations could have unexpected results that arise because of the plasticity of proteins.

### New approaches

**Casting a wider net.** The often-disastrous effects of 'rationally introduced' mutations on the stability and activity of proteins prompted Charles Craik to muse that 'protein terrorism' was a more suitable descriptor than protein engineering. To increase the chances of being able to intelligently assess the role of individual residues, the late 1980s saw a gradual drift towards methods that were able to systematically trawl a larger number of mutations, both at single sites, and scattered along the protein. In one approach, specific residues were changed by saturation mutagenesis (to every other amino acid), either by using degenerate oligonucleotides that incorporate NNC/G

### Box 2 | Antibody engineering



As natural agents in host defence against infection and disease, the potential of engineered antibodies in the clinic is self-evident. Although monoclonal antibodies that bind a vast array of antigens with high affinity and selectivity can be raised more or less at will, their rodent origin can render them ineffective in patients. This is because the antibodies themselves are often immunogenic, which is a problem if chronic therapy is required. The modular structure and function of the immunoglobulin molecule has allowed the antigen recognition domains and surfaces (complementarity-determining regions, CDRs) of rodent immunoglobulins to be exchanged and grafted, respectively, onto the framework of a human antibody (see figure). These 'humanized' antibodies acquire the antigen-binding characteristics of the rodent antibody without the attendant problems of immunogenicity. Moreover, they are more efficient in recruiting human immune effector functions. These antibodies are now being used clinically to treat cancers and to prevent transplant rejection. In other strategies, antibodies are being engineered for new effector functions through conjugation to toxin/radionuclides and for improved retention in the serum<sup>63</sup>. The figure is modified with permission from REF. 63 © (2002) Kluwer Academic Publishers.

codons or through cassette mutagenesis. Targeting a methionine residue that is near the active site of the alkaline protease subtilisin in this way led to the isolation of an enzyme with enhanced resistance to oxidation, which has a potential application in biological washing powders<sup>38</sup>.

To sample the effects of a single mutation at many different sites, 'surface scanning' of 40 basic residues in TyrRS was implemented to delineate the sites for binding of tRNA<sup>39</sup>. Similarly, 'alanine-scanning' mutagenesis<sup>40</sup> was used to identify the binding site of the complement subcomponent C1q with complement<sup>41</sup>. Elsewhere, short segments of enzymes were mutated randomly<sup>42</sup> and mixed oligonucleotides were used to introduce missense mutations across the entire coding sequence of human immunodeficiency virus 1 (HIV-1) protease in a single reaction<sup>43</sup>.

It became clear that, rather than looking at single amino-acid alterations, methods of introducing multiple mutations in various combinations were required<sup>44</sup>. The brute-force inherent in these 'irrational' combinatorial approaches might evoke the basic engineering tenet of "when all else fails, use ruddy great nails". However, a rational element is maintained in these experiments by the use of sophisticated high-throughput methods of screening for the desired activity. In their simplest form, the target gene is mutagenized at random. In the tradition of *The Blind Watchmaker*<sup>45</sup>, the rare events that lead to

increased fitness are selected for and so they persist. By mirroring natural selection, these 'directed evolution' experiments have dominated recent attempts to isolate proteins with specific attributes — a good example is the generation of a subtilisin variant that is active in 60% dimethyl formamide<sup>46</sup>. Hydrolases selected in this way for their activity in organic solvents are used widely to favour synthetic rather than hydrolytic reactions.

### Sex and evolution: designing protein function.

Although site-directed mutagenesis proved to be highly effective for analysing the role of individual residues, it did not provide either a general method for making proteins with designed properties, or a simple route to new enzymes with novel activities. This had to wait for the advent of repertoire screening and selection technologies and methods for generating a rich diversity of mutation combinations, which were pioneered for antibodies by Richard Lerner and Greg Winter (for example, see REF. 47). For enzymes, the Californian groups of Willem Stemmer at Maxygen<sup>48,49</sup> and Frances Arnold at Caltech<sup>50</sup> enhanced directed evolution by shuffling mutation combinations, in the same way that recombination during meiosis generates diversity from parental traits. This 'sexual PCR' is akin to molecular breeding, and it even allows backcrossing to fix enhanced properties and eliminate deleterious ones, by recombining isolates of increased fitness with

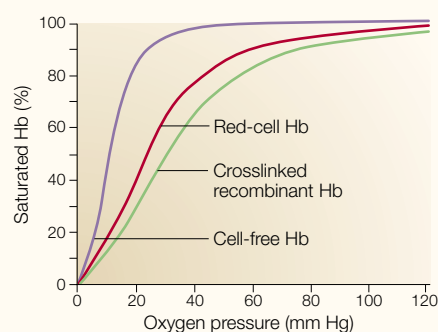


**Figure 2 | Oxygen saturation curves for haemoglobin.** The curves are shown for red-blood-cell haemoglobin (Hb), cell-free Hb and a crosslinked recombinant form of Hb that contains the  $\beta$ -chain Asn108Lys mutation (where Asn is asparagine and Lys is lysine). This figure was provided courtesy of K. Nagai (MRC Laboratory of Molecular Biology, Cambridge, UK), and was prepared using data from REFS 68,86.

parental sequences<sup>51</sup>. Successive generations can be made rapidly, and natural species barriers ignored. The method is limited only by the number of isolates that can be analysed, and how the selected property is quantified. Such *in vitro* evolution approaches might prove to be the most effective way to generate new enzymes with novel activities that are not represented in nature (BOX 1).

Several techniques have been used to increase the possible sequences that can be explored by combinatorial methods. As point mutations lead to rather conservative changes and limited sampling of the available sequence space, it is desirable to also explore those sequences that are accessible by two or three base substitutions in a single codon<sup>52</sup>. A codon-level combinatorial method that enables all possible replacements to be made at any given position has been used to mutate six contiguous residues of  $\beta$ -lactamase, and this led to the selection of a ceftazidime-resistant triple mutant<sup>53</sup>. Random insertion/deletion mutagenesis (RID)<sup>54</sup> mimics other evolutionary processes that are difficult to simulate by PCR methods.

The impressive developments that have been made in the techniques that are used to generate genetic diversity have focused attention on the bigger problem of selection methods. In one approach that exploited phage display, a DNA polymerase was evolved into an RNA polymerase by using an activity based selection system, in which the polymerase and its substrate were each attached to adjacent coat proteins on the phage<sup>55</sup>. Elsewhere, imaginative approaches to enzyme evolution that are based on gene selection in compartmentalized *in vitro* systems are

being pioneered<sup>56,57</sup>. In these systems, the genotype–phenotype connection is often maintained by covalently linking the target substrate to a library of gene fragments. These genes are transcribed and translated in isolated compartments that are formed in water-in-oil emulsions and selection is carried out for genes that are linked to the reaction product.

### Protein engineering leaves the lab

Despite much promise in agriculture — for example, in the engineering of a high proportion of essential amino acids into staple food proteins — the first example of a tailored enzyme being introduced into crop plants (to produce stearate and increase the proportion of unsaturated fatty acids in seed oil) has only recently taken place<sup>58</sup>. Enzymes are used increasingly as supplements to animal feeds to aid the digestion of celluloses and the release of phosphorus from phytic acids. The increasing prominence of engineered phytases<sup>59</sup> is a response to concerns over bovine spongiform encephalopathy (BSE) and the shunning of traditional bone-meal sources of phosphorus.

Proteins that are designed for household use remain rare, with the notable exceptions of engineered proteases and lipases that work well in the low-temperature cycles used in modern washing machines. In industry, engineered enzymes are used in various processes, in particular for the synthesis of fine chemicals and antibiotics. Combinatorial methods are rapidly replacing the bio-prospecting and strain-improvement philosophy for making new biocatalysts that are appropriate for an industrial setting<sup>60</sup>. The efficacy of proteins that have been engineered to degrade environmental pollutants, such as polychlorinated benzenes, by bioremediation<sup>61</sup> has been shown. Increasingly sophisticated applications, such as biosensors that can detect accurately a range of compounds, seem to be just around the corner<sup>62</sup>, but it is in the design of therapeutic proteins that protein engineering has made the most impact on society.

**Engineering of antibodies.** Protein engineering of antibodies has long been regarded as a rich seam, and this early promise is being realized with a variety of engineered antibodies now being used in the clinic<sup>7,63,64</sup> (BOX 2). Antibody engineering has been a principal driving force for many of the pioneering developments in the techniques that are described both in the previous section and in BOX 1. The basic strategies and some of the impressive achievements in this area have

been reviewed extensively elsewhere and are briefly summarized here in BOX 2.

**Haemoglobin-based blood substitutes.** After severe blood loss, blood volume and oxygen transport need to be restored quickly<sup>65</sup>. Cell-free oxygen carriers or ‘blood substitutes’ for emergency use have advantages over blood transfusion, because they do not need to be serotyped or screened for diseases. Purified recombinant human haemoglobin (Hb) is the obvious oxygen carrier to use, but several obstacles must first be overcome.

In the absence of the red-blood-cell allosteric effector diphosphoglycerate, the oxygen affinity of Hb must be lowered to allow sufficient oxygen dissociation at the tissues (FIG. 2). Among several hundred clinically reported Hb variants, many have been shown to have lower oxygen affinity. In one of these, which is known as Hb Presbyterian, asparagine 108 in the  $\beta$ -chain is replaced by lysine (Lys). This mutation was therefore engineered into recombinant human Hb with the desired effect (FIG. 2).

A second problem is caused by the tendency of  $\alpha_2\beta_2$ -Hb tetramers to dissociate to  $\alpha\beta$ -dimers that are small enough to pass through the glomerular membrane of the kidney, which causes severe damage to this organ. This problem was overcome by using genetic-engineering methods to fuse the sequences that encode the two  $\alpha$ -subunits together in head-to-tail fashion; the crystal structure of Hb had shown that this fusion could be accommodated in the Hb tetramer with the introduction of just a single intervening glycine residue<sup>66,67</sup>. The resulting Hb tetramer is stable and resists excretion through the kidneys<sup>68</sup>. Other problems, including the loss of activity through auto-oxidation of the ferrous iron and the hypertensive effects that are caused by nitric-oxide scavenging, remain, but we anticipate that, as these phenomena become better understood, they can be overcome<sup>69,70</sup>.

**Improved insulins for diabetics.** Successful treatment of diabetes mellitus relies on the continuous control of blood glucose levels. This is a major challenge, because the pharmacokinetics of traditional insulin preparations that are injected subcutaneously do not match the physiological profiles of insulin secreted by the pancreas. In particular, there is a need to accelerate the rate of arrival of insulin into the blood from the site of injection, and to prolong its lifetime once it is there<sup>71,72</sup>. The observation of a lag phase in the absorption of subcutaneously injected insulin and the fact that, in concentrated

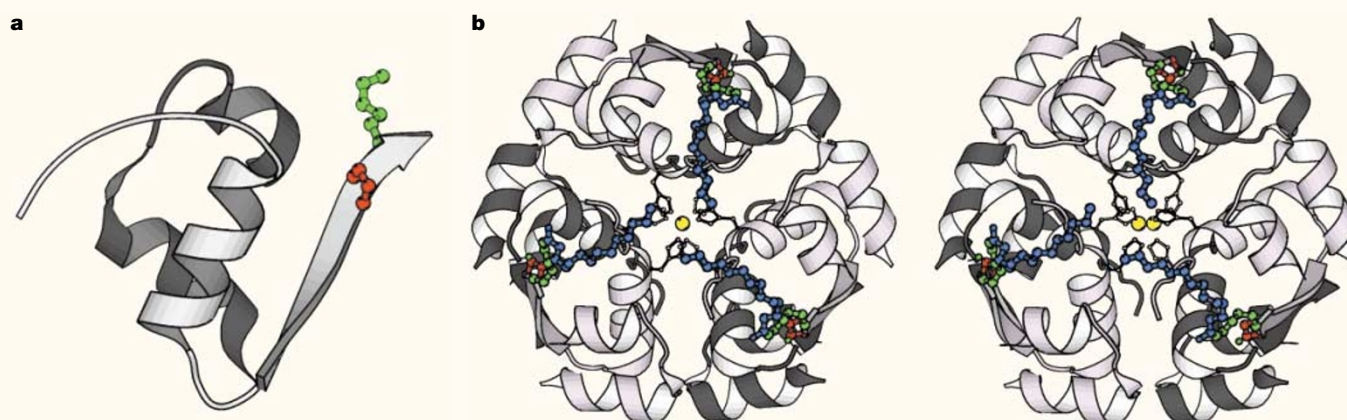


**Figure 3 | Ribbon diagrams of insulin.** **a** | An insulin monomer and **b** | a stereoview of the LysB29-tetradecanoyl des-(B30) insulin hexamer (where Lys is lysine). Protracted-acting insulins for therapy have been designed in which the  $\epsilon$ -amino group of LysB29 in the B chain (LysB29) is acylated with long-chain saturated fatty acids. In both panels, the side chains of proline 28 in the B chain (red) and LysB29 (green) are highlighted. The fatty acyl moiety is shown in blue. The zinc ions (yellow) and the B-chain histidine 10 ligands that coordinate them are shown at the centre of the hexamer<sup>87</sup>.

preparations, insulin self-associates to form dimers and hexamers, led to the idea that insulins with a reduced tendency to aggregate might be absorbed more rapidly. Self-assembly is not necessary for biological activity because insulin binds to its receptor as a monomer. The surfaces that are used in this self-assembly were readily identified from the crystal structure of the insulin hexamer<sup>73,74</sup> (FIG. 3). Protein engineering was used successfully to disrupt these interfaces through mutations that either juxtapose negative charges, such as when proline 28 in the B chain (Pro B28) is replaced by an Asp residue, or by destabilizing the intermolecular  $\beta$ -sheet interactions that are formed by the B-chain carboxyl termini, such as when ProB28 and LysB29 are exchanged. These fast-acting monomeric insulins retain biological activity and are now being used for therapy.

To match the basal levels of insulin in the circulation, protracted-acting insulins for therapy have been designed in which the  $\epsilon$ -amino group of LysB29 is acylated with long-chain saturated fatty acids (FIG. 3). These modifications cause insulin to associate with the fatty-acid-binding protein albumin that is present in the subcutaneous tissue fluids, which results in slower absorption into the blood and prolonged presence in the plasma<sup>71,72</sup>. These long-acting insulins are in advanced stages of clinical trials.

### Conclusion and prospects

In the field of enzymology, site-directed mutagenesis remains a potent technique for the dissection of the individual contributions of amino-acid residues and to test hypotheses. There is still an informed trial-and-error

element to the subject, even though there have been significant achievements. The challenge for protein scientists in the next 20 years will be to develop insights into folding and dynamics that give the protein engineer the confidence that a civil engineer enjoys<sup>75</sup>. A prevailing wisdom is that billions of years of evolution have shaped proteins to carry out their *in vivo* functions perfectly<sup>1</sup>. Protein engineering is therefore concerned with adapting proteins to function under different regimes. In this regard, most protein engineering so far has really been re-design, with nature having shown us that different catalytic and binding events can occur on a common scaffold. Now, however, shuffling technology can search for actions on compounds that are unlikely to have been encountered in nature. Fifty rounds of shuffling and directed pressure have been used to evolve ampicillin resistance of unknown mechanism from a completely unrelated gene<sup>76</sup>.

In the past, semi-synthetically misacylated tRNAs have been used successfully to direct the incorporation of unnatural amino acids into proteins<sup>77</sup>. Unnatural amino acids provide a new set of building blocks and a potential route to new chemistries that extend beyond those achieved through the 20 common amino acids that constrain natural selection. Until recently, the approach was limited to *in vitro* translation systems, but advances have made possible the site-specific incorporation of unnatural amino acids *in vivo* by using engineered aminoacyl-tRNA synthetases to 'misacylate' tRNAs that recognize termination codons<sup>78,79</sup>. This is a *tour de force* of engineering, as both RNA and protein components had to be adapted. This work echoes the first site-directed mutagenesis

experiment, in which an amber stop codon was engineered into a coding sequence<sup>80</sup>, and exploits the knowledge that was gleaned from the first protein engineering experiments on TyrRS to fool the translational machinery of the cell. How much difference emerging techniques such as this will make in the coming years is hard to gauge. They will certainly do well if they come close to matching the impact that site-directed mutagenesis — a technique so simple in its conception — has had on protein science over the past 20 years.

James A. Brannigan and Anthony J. Wilkinson are at the Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, UK.  
e-mails: [jab@ysbl.york.ac.uk](mailto:jab@ysbl.york.ac.uk);  
[ajw@ysbl.york.ac.uk](mailto:ajw@ysbl.york.ac.uk)

doi:10.1038/nrm975

- Knowles, J. R. & Alber, W. J. Perfection in enzyme catalysis: the energetics of triose phosphate isomerase. *Acc. Chem. Res.* **10**, 105–111 (1977).
- Fersht, A. R. *Enzyme Structure and Mechanism* (W. H. Freeman & Co., San Francisco, 1977).
- Campbell, I. D. The march of structural biology. *Nature Rev. Mol. Cell Biol.* **3**, 377–381 (2002).
- Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **299**, 756–758 (1982).
- Dalbadie-MacFarland, G. *et al.* Oligonucleotide-directed mutagenesis as a powerful method for studies of protein function. *Proc. Natl Acad. Sci. USA* **79**, 6409–6413 (1982).
- Sigal, I. S., Harwood, B. G. & Arentzen, R. Thiol  $\beta$ -lactamase: replacement of the active site serine of RTM  $\beta$ -lactamase by a cysteine residue. *Proc. Natl Acad. Sci. USA* **79**, 7157–7160 (1982).
- Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522–525 (1986).
- Neuberger, M. S., Williams, G. T. & Fox, R. O. Recombinant antibodies possessing novel effector functions. *Nature* **312**, 604–608 (1984).
- Wilkinson, A. J., Fersht, A. R., Blow, D. M. & Winter, G. Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. *Biochemistry* **22**, 3581–3586 (1983).

10. Leatherbarrow, R. J., Fersht, A. R. & Winter, G. Transition-state stabilisation in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc. Natl Acad. Sci. USA* **82**, 7840–7844 (1985).
11. Wilkinson, A. J., Fersht, A. R., Blow, D. M., Carter, P. & Winter, G. A large increase in enzyme-substrate affinity by protein engineering. *Nature* **307**, 187–188 (1984).
12. Fersht, A. R. *et al.* Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235–238 (1985).
13. Smith, M. *In vitro* mutagenesis. *Annu. Rev. Genet.* **19**, 423–462 (1985).
14. Carter, P. & Wells, J. A. Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564–568 (1988).
15. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin. *Phil. Trans. R. Soc. Lond. A* **317**, 415–423 (1986).
16. Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H.-Y. & Poulos, T. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc. Natl Acad. Sci. USA* **83**, 3743–3745 (1986).
17. Springer, B. A. *et al.* Discrimination between oxygen and carbon monoxide and inhibition of autooxidation by myoglobin. *J. Biol. Chem.* **264**, 3057–3060 (1989).
18. Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S. & Rutter, W. J. Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. *Nature* **317**, 551–555 (1985).
19. Blake, C. C. F. *et al.* Structure of hen egg white lysozyme. A three-dimensional fourier synthesis at 2Å resolution. *Nature* **206**, 757–763 (1965).
20. Vocadlo, D. J., Davies, G. J., Laine, R. & Withers, S. G. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835–838 (2001).
21. Kantrowitz, E. R. & Lipscomb, W. N. *Escherichia coli* aspartate transcarbamylase: the relation between structure and function. *Science* **241**, 669–674 (1988).
22. Matouschek, A., Kellis J. T. Jr, Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
23. Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002).
24. Matthews, B. W. Genetic and structural analysis of the protein stability problem. *Biochemistry* **26**, 6855–6888 (1987).
25. Shortle, D. Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem.* **264**, 5315–5318 (1989).
26. Arnold, F. H. Engineering enzymes for non-aqueous solvents. *Trends Biotechnol.* **8**, 244–249 (1990).
27. Perry, L. J. & Wetzel, R. Disulfide bond engineering into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* **226**, 555–557 (1984).
28. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G. & Eijssink, V. G. H. Engineering an enzyme to resist boiling. *Proc. Natl Acad. Sci. USA* **95**, 2056–2060 (1998).
29. Russell, A. J. & Fersht, A. R. Rational modification of enzyme catalysis by engineering surface charge. *Nature* **328**, 496–500 (1987).
30. Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. C. Structure–activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
31. Wells, J. A. Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509–8517 (1990).
32. Wilks, H. M. *et al.* A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* **242**, 1541–1544 (1988).
33. Wilks, H. M. *et al.* Designs for a broad substrate specificity keto acid dehydrogenase. *Biochemistry* **29**, 8587–8591 (1990).
34. Wilks, H. M. *et al.* Design of a specific phenyllactate dehydrogenase by peptide loop exchange on the *Bacillus stearothermophilus* lactate dehydrogenase framework. *Biochemistry* **31**, 7802–7806 (1992).
35. Graf, L. *et al.* Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry* **26**, 2616–2623 (1987).
36. Perona, J. J. *et al.* Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* **34**, 1489–1499 (1995).
37. Venekei, I., Szilagyi, L., Graf, L. & Rutter, W. J. Attempts to convert chymotrypsin to trypsin. *FEBS Lett.* **379**, 143–147 (1996).
38. Bedouelle, H. & Winter, G. A model of synthetase/transfer RNA interaction as deduced by protein engineering. *Nature* **320**, 371–373 (1986).
39. Cunningham, B. C. & Wells, J. A. High-resolution epitope mapping of HGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081–1085 (1989).
40. Wells, J. A. & Estell, D. A. Subtilisin — an enzyme designed to be engineered. *Trends Biochem. Sci.* **13**, 291–297 (1988).
41. Duncan, A. R. & Winter, G. The binding site for C1q on IgG. *Nature* **332**, 738–740 (1988).
42. Oliphant, A. R. & Struhl, K. An efficient method for generating proteins with altered enzymatic properties: application to  $\beta$ -lactamase. *Proc. Natl Acad. Sci. USA* **86**, 9094–9098 (1989).
43. Loeb, D. D. *et al.* Complete mutagenesis of the HIV-1 protease. *Nature* **340**, 397–400 (1989).
44. Merino, E., Osuna, J., Bolivar, F. & Soberon, X. A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC/M13 vectors. *Biotechniques* **12**, 508–510 (1992).
45. Dawkins, R. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design* (Longman, London, 1986).
46. Chen, K. & Arnold, F. H. 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 (1993).
47. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349**, 293–299 (1991).
48. Stemmer, W. P. R. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
49. Cramer, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
50. Arnold, F. H. Design by directed evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).
51. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA* **91**, 10747–10751 (1994).
52. Miyazaki, K. & Arnold, F. H. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J. Mol. Evol.* **49**, 716–720 (1999).
53. Gaytan, P., Osuna, J. & Soberon, X. Novel ceftazidime-resistance  $\beta$ -lactamases generated by a codon-based mutagenesis method and selection. *Nucleic Acids Res.* **30**, e84 (2002).
54. Murakami, H., Hohsaka, T. & Sisido, M. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nature Biotechnol.* **20**, 76–81 (2002).
55. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G. & Romesberg, F. E. Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl Acad. Sci. USA* **99**, 6597–6602 (2002).
56. Griffiths, A. D. & Tawfik, D. S. Man-made enzymes – from design to *in vitro* compartmentalisation. *Curr. Opin. Biotechnol.* **11**, 338–353 (2000).
57. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalised self-replication. *Proc. Natl Acad. Sci. USA* **98**, 4552–4557 (2001).
58. Facciotti, M. T., Bertain, P. B. & Yuan, L. Improved stearate phenotype in transgenic canola expressing a modified acyl–acyl carrier protein thioesterase. *Nature Biotechnol.* **17**, 593–597 (1999).
59. Lehmann, M. *et al.* The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **15**, 403–411 (2002).
60. Burton, S. G. The search for the ideal biocatalyst. *Nature Biotechnol.* **20**, 37–45 (2002).
61. Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T. & Furukawa, K. Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl oxygenase. *Nature Biotechnol.* **16**, 663–666 (1998).
62. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413**, 226–230 (2001).
63. Holliger, P. & Bohlen, H. Engineering antibodies for the clinic. *Cancer Metastasis Rev.* **18**, 411–419 (1999).
64. McCafferty, J. & Glover, D. R. Engineering therapeutic proteins. *Curr. Opin. Struct. Biol.* **10**, 417–420 (2000).
65. Winslow, R. M. Blood substitutes: refocusing an elusive goal. *Br. J. Haematol.* **111**, 387–396 (2000).
66. Perutz, M. F. Stereochemistry of the cooperative effects in haemoglobin. *Nature* **228**, 726–734 (1970).
67. Dickerson, R. E. & Geis, I. *Haemoglobin: Structure, Function, Evolution and Pathology* (Benjamin/Cummings Publishing Co., Inc., Menlo Park, California, 1983).
68. Looker, D. *et al.* A human recombinant haemoglobin designed for use as a blood substitute. *Nature* **356**, 258–260 (1992).
69. Olson, J. S., Eich, R., Smith, L. P., Warren, J. J. & Knowles, B. C. Protein engineering strategies for designing more stable hemoglobin-based blood substitutes. *Artif. Cells Blood Subst. Immobil. Biotechnol.* **25**, 227–241 (1997).
70. Doherty, D. H. *et al.* Rate of reaction with nitric oxide determines the hypertensive effect of cell-free hemoglobin. *Nature Biotechnol.* **16**, 672–676 (1998).
71. Brange, J. & Volund, A. Insulin analogs with improved pharmacokinetic profiles. *Adv. Drug Deliv. Rev.* **35**, 307–335 (1999).
72. Vajo, Z. & Duckworth, W. C. Genetically engineered insulin analogues: diabetes in the new millennium. *Pharmacol. Rev.* **52**, 1–10 (2000).
73. Blundell, T., Dodson, G., Hodgkin, D. & Mercola, D. Insulin: the structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem.* **26**, 279–402 (1972).
74. Brange, J. *et al.* Monomeric insulins obtained by protein engineering and their medical implications. *Nature* **333**, 679–682 (1988).
75. Knowles, J. R. Tinkering with enzymes: What are we learning? *Science* **236**, 1252–1259 (1987).
76. Yano, T. & Kagamiyama, H. Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: a laboratory model of molecular evolution. *Proc. Natl Acad. Sci. USA* **98**, 903–907 (2001).
77. Mendel, D., Cornish, V. W. & Schultz, P. G. Site-directed mutagenesis with an expanded code. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 435–462 (1995).
78. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–504 (2001).
79. Kiga, D. *et al.* An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc. Natl Acad. Sci. USA* **99**, 9715–9720 (2002).
80. Hutchison, C. A. *et al.* Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560 (1978).
81. Chang, C. C. *et al.* Evolution of a cytokine using DNA family shuffling. *Nature Biotechnol.* **17**, 793–797 (1999).
82. Sieber, V., Martinez, C. A. & Arnold, F. H. Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnol.* **19**, 456–460 (2001).
83. Lutz, S., Ostermeier, M. & Benkovic, S. J. Rapid generation of incremental truncation libraries for protein engineering using  $\alpha$ -phosphothioate nucleotides. *Nucleic Acids Res.* **29**, 16e (2001).
84. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D. & Benkovic, S. J. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl Acad. Sci. USA* **98**, 11248–11253 (2001).
85. O'Maille, P. E., Bakhtina, M. & Tsai, M.-D. Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.* **321**, 677–691 (2002).
86. Kilmartin, J. V. & Rossi-Bernardi, L. Interaction of haemoglobin with hydrogen ions, carbon dioxide and organic phosphates. *Physiol. Rev.* **53**, 836–890 (1973).
87. Whittingham, J. L., Havelund, S. & Jonassen, I. Crystal structure of a prolonged-acting insulin with albumin-binding properties. *Biochemistry* **36**, 2826–2831 (1997).
88. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).

## Acknowledgements

We thank G. Winter for his critical comments and helpful advice, and the Wellcome Trust and the Biotechnology and Biological Sciences Research Council for supporting our research.

 Online links

## DATABASES

The following terms in this article are linked online to: Swiss-Prot: <http://www.expasy.ch/>  $\beta$ -lactamase | insulin | lysozyme | myoglobin | TyrRS

## FURTHER INFORMATION

## York Structural Biology Laboratory:

<http://www.ysbl.york.ac.uk>

Access to this interactive links box is free online.