CMLS, Cell. Mol. Life Sci. 54 (1998) 394–404
1420-682X/98/050394-11 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Screening of conformationally constrained random polypeptide libraries displayed on a protein scaffold

N. Doi[a,b] and H. Yanagawa[b,*]

[a]Graduate School of Environmental Earth Sciences, Hokkaido University, Sapporo, Hokkaido 060 (Japan)
[b]Mitsubishi Kasei Institute of Life Sciences, Machida, Tokyo 194 (Japan), Fax +81 427 24 6317

Received 9 October 1997; received after revision 6 January 1998; accepted 19 January 1998

Abstract. The selection of novel proteins or enzymes from random protein libraries has come to be a major objective in current biology, and these enzymes should prove useful in various biological and biomedical fields. New technologies such as in vitro selection of proteins in cell-free systems have high potential to realize evolutionary molecular engineering of proteins. This review highlights an application of insertional mutagenesis of proteins to evolutionary molecular engineering. Random sequence proteins are inserted into the surface of a host enzyme which serves as a scaffold to display random protein libraries. Constraints on random polypeptide conformations owing to the proximity of N- and C-termini on the scaffold would result in greater screening efficiency of libraries. The scaffold enzyme is also used as a probe for monitoring the hill climbing of random sequence proteins on a fitness landscape and navigating rapid protein folding in the sequence space.

Key words. *Escherichia coli* RNase HI; evolutionary molecular engineering; insertional mutagenesis; protein evolution; protein folding; random sequence protein.

Introduction

The advent of genetic engineering has enabled the design and construction of new proteins with virtually any sequences. One of the important goals of de novo protein design is to create useful enzymes that catalyse novel reactions not carried out by natural enzymes. The design of totally new protein domains is very difficult, since the relationship of amino acid sequence to protein structure or function has yet to be understood. There are presently two experimental approaches for clarifying the sequence–structure/function relationships of proteins. One is site-directed mutagenesis of natural enzymes for investigating the nearest area of natural proteins in the sequence space (protein engineering; white circles in fig. 1). The other is an evolutionary approach involving random

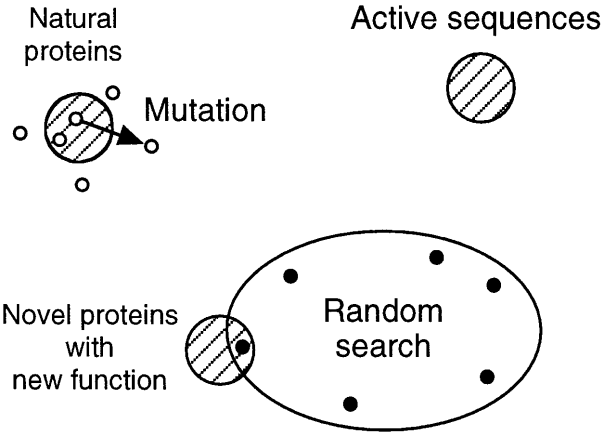

Figure 1. Schematic representation of amino acid sequence space. Clusters of active sequences with identical functions are distributed in the sequence space. White circles indicate sequences of natural proteins and mutant proteins. Black circles indicate random sequence proteins, distributed widely throughout the sequence space.

* Corresponding author.

search and optimization of artificial protein sequences in a wide region of the sequence space (evolutionary engineering; black circles in fig. 1).

In protein engineering, substitution mutagenesis of natural enzymes, by which effects of small perturbations on the structure and function of proteins are studied, is primarily used. Most changes in side chains have been shown to be tolerated surprisingly well, except when they occur near active-site residues. Studies of the structural and functional effects of relatively large perturbations in polypeptide backbones have also been performed by insertion and deletion mutagenesis [1], with results similar to those of substitution mutagenesis.

Evolutionary molecular engineering, first proposed by Eigen and Gardiner [2], applies the concept of molecular evolution to creating novel molecules blindly by a process of mutation and selection. Efficient screening of large libraries of randomly generated molecules for desired binding activity is now possible. For example, RNA molecules [3] and small peptides [4] capable of binding to a particular protein or dye have been selected by alternate rounds of affinity selection and polymerase chain reaction (PCR) amplification. It is still not possible to select novel proteins from libraries with completely random sequences of more than 100 amino acid residues.

In this review, we will discuss the possibilities of the evolutionary approach for selecting novel proteins or enzymes from completely random sequence libraries. The key concept is that of a 'protein scaffold'; so far scaffolds have been used to display relatively short peptide libraries [5, 6]; we seek to display random sequences of more than 100 amino acid residues on a surface loop region of a protein scaffold and to screen them. In the next section, the design of random sequence proteins with long open reading frames (ORFs) and merits of selecting functional random proteins on a protein scaffold are discussed. Next, we take up the choices of scaffold proteins and screening methods. Finally, we will discuss evolutionary aspects of proteins from random sequences.

Mimicking enzymes

For the past decade or more years, catalytic antibodies [7, 8] have been considered ideal for mimicking enzyme catalysis and have been used with limited success. Catalytic antibodies bind to transition state analogues with high affinity, but their catalytic efficiency generally falls far short of that of natural enzymes [9]. The rigid antibody framework is optimal for tight binding to substrates on protein surfaces, but enzyme catalysis may require substrate–enzyme interactions in an 'adventitious pocket' [10] to achieve high turnover and induced fit for substrates (fig. 2). Present-day enzymes may possibly have evolved from flexible or molten globule-like proteins, and this would have provided a plastic medium in which random mutations might give rise to enzymatic activity [11]. To imitate this process, we propose a model system, in which a long random polypeptide is displayed on a protein scaffold (fig. 2B). Establishing this system requires the design of random sequence proteins with long ORFs and the insertion of large polypeptides into the protein scaffold.

Design of random sequence proteins

The design of gene libraries encoding fully random sequence proteins with long ORFs is difficult. The transcription and translation of completely random DNA sequences produce only short peptides (about 20 residues in length on average), since stop codons appear at a probability of 3/64. To prevent this in DNA would give rise to biased amino acid composition and reduce the diversity of amino acid sequences. Long sequence length is generally incompatible with high sequence diversity [12].

A few attempts have been made to construct random protein libraries with limited diversity. Kamtekar et al. constructed a partially randomized library of a four-helix bundle protein in which the pattern of polar and nonpolar amino acids was constrained but the identity of these residues was not constrained, and thus it varied [13]. In this library, polar and nonpolar amino acid residues of protein sequences were encoded by the degenerate codons NAN and NTN, re-

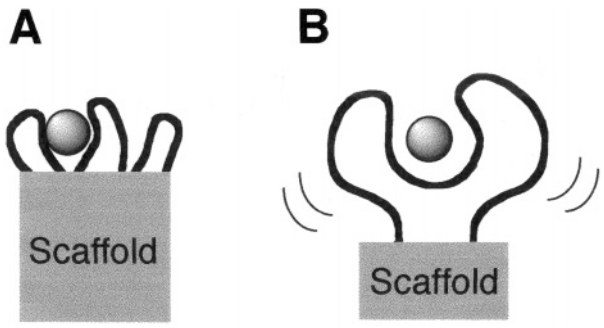

Figure 2. Schematic illustration of methods for mimicking enzymes. (*A*) Several short random loops are displayed together on a scaffold such as an antibody framework. Substrate binding to the surface of protein is strong. (*B*) A long random loop is displayed on a scaffold protein. Substrate enters the pocket of a long flexible polypeptide.

spectively (N represents an equimolar mixture of A, G, C and T). Davidson et al. constructed libraries of 80- to 100-residue random proteins consisting of glutamine (Q), leucine (L), and arginine (R) [14, 15]. QLR proteins were encoded by randomized CDR codons (D, an equimolar mixture of A, T and G; R, an equimolar mixture of A and G). In this library, the locations of glutamine, leucine and arginine in the protein sequence were completely random, with content controlled by that of A, T and G at the second position of each randomized codon. In these cases, however, the sequence patterns or the amino acid types were highly constrained.

In an early study, Mandecki constructed long ORFs encoding 200–800 amino acid polypeptides by random ligation of 75-bp-long DNA fragments [16]. The sequences contained randomized NNY (Y = C or T) or RNN codons, and thus there was no stop codon or glutamine or tryptophan codon. Prijambada et al. recently constructed a highly random library containing 140-residue random proteins comprising 20 naturally occurring amino acids [17]. In this library, the semirandom sequence was designed with technical care: (i) No stop codon was present in any of six frames even though one-base insertions or deletions occur during the synthesis and construction of DNA libraries. (ii) Mean GC content was not high. High GC content is suited to avoidance of stop codons such as TAA, TAG and TGA, but often interferes with polymerase chain reactions (PCR). These random protein libraries with high diversity and complexity may serve as sources of useful sequences to be selected.

Large ORFs can also be constructed by the polymerization of a minigene possessing no termination codons in the six frames [18]. Libraries of repeating polypeptides constructed by shuffling protein building blocks such as secondary structures or modules [19, 20] may provide useful sources of unevolved proteins that mimic exon shuffling [21].

Constrained polypeptide conformations

Why is it difficult to select novel enzymes from libraries with such highly random sequence proteins? This question is closely related to another question regarding what fractions of all random sequence proteins have desired functions. Lau and Dill estimated the fraction of 100-residue polymer sequences which can fold and active is in the range of 10^{-8} to 10^{-17} [22]. Assuming their result is correct, libraries containing more than 10^8 to 10^{17} variants must be screened to obtain a functional molecule. The present screening of polypeptides makes it possible to prepare as many as 10^{12} variants. Screening efficiency should thus be enhanced over several orders of magnitude to make possible the selection of enzymes with all desired functions.

A selection system in which random sequence proteins are inserted into the surface of a scaffold enzyme has the following merits. First, the random polypeptide is conformationally constrained due to the proximity of N- and C-termini on the scaffold (fig. 2B). When termini are constrained early in folding, conformational space to be searched is greatly reduced [23]. Constraint is n^{-2} for a polymer molecule consisting of n monomers, according to a simple lattice model [24], e.g. the number of possible conformations, Nc, is reduced to $N\mathrm{c} \times 10^{-4}$ for a 100-residue polymer. This reduction may directly reflect the probability of polypeptide folding, and this would enhance the screening efficiency by 10^4-fold. That is, the required number of variants can be reduced from 10^8–10^{17} to 10^4–10^{13}.

Second, the scaffold protein serves as a means for high-level expression and stabilization of foreign polypeptides in vivo. Foreign sequences introduced in host cells often lead to low-level expression and peptidase degradation. Insertion of random protein at the middle of the scaffold protein prevents it from undergoing exopeptidase digestion in vivo. Endopeptidase digestion would also occur to a lesser extent by conformational constraint.

Insertional mutagenesis

Insertion and deletion mutagenesis of proteins has been used mainly to screen for functionally important regions in proteins whose structures are little understood [25–27]. Proteins have been found generally to tolerate most mutations surprisingly well, based on their stability and catalytic activity. The sites at which insertions are tolerated without loss of activity are called 'permissive' sites [28].

Insertional mutagenesis is also used as a structural probe for proteins of known three-dimensional structure. Permissive sites of some amino acid insertions have been found restricted almost exclusively to loop regions, appearing only rarely in secondary structure elements such as α-helices and β-strands. This has also been shown for insertions and deletions in a data bank by multiple sequence alignment studies [29, 30]. Examination of 1400 examples of insertions and deletions in the Protein Data Bank indicated only 24 to be in α-helices and 20 in β-strands [29]. In addition, little or no effect of elongation of loop length on protein folding was quite recently reported [31, 32].

Insertions can be used not only to modify structural and functional properties of proteins but also to impart

entirely new properties to host proteins. To use a host protein as a scaffold for random polypeptides, the means for inserting large segments into permissive sites must be understood. Thus, we consider the permissive *length* of insertions in this review, with no thought to deletion mutants of proteins. Permissive sites and length are described for three types of proteins: membrane proteins, secretory soluble proteins and intracellular soluble proteins:

Membrane proteins

Membrane proteins of confirmed three-dimensional structure are rather few, and scanning insertion mutagenesis of membrane proteins thus still serves as an efficient means to screen for functionally important regions of membrane proteins and to confirm membrane-spanning regions and topology. Typical membrane proteins have one or several hydrophobic membrane-spanning helices connected to hydrophilic extramembrane loops. The transmembrane regions contain 20 to 30 residues of continuous hydrophobic amino acids, which can easily be predicted by hydropathy plots of the hydrophobicity of adjacent amino acid residues averaged over a moving window of suitable length [33].

Permissive sites are generally present in helix-connecting loops [34–37]. A recent study on 31-codon insertions in *Escherichia coli* lac permease at 21 different sites indicated a simple pattern of the lactose transport activity of mutant proteins: most proteins (10/12) with insertions in the loops are at least partially active, and all proteins (9/9) with insertions in membrane-spanning helices are inactive [37]. Insertions in transmembrane helices respond more strongly to activity than do extramembrane loops. A single amino acid (mainly alanine) was inserted into transmembrane helices containing active-site regions in integral membrane proteins, such as bacteriorhodopsin [38], the α subunit of F_1F_0-ATP synthase [39] and glycophorin A [40]. As expected, activity was altered considerably when insertions were made at or quite close to active-site regions. This is an effective means for examining membrane-spanning helices and mapping helix–helix interactions.

Foreign peptide insertions in outer-membrane proteins, such as *Pseudomonas* OprF [41], OprH [42], *E. coli* FhuA [43], OmpA [44], PhoE [45] and LamB [46] have been studied extensively. A typical outer-membrane protein structure is the eight-stranded β-barrel motif; the transmembrane regions are not α-helices but β-strands of about 10 amino acids in length. Permissive sites are generally distributed in extramembrane loop regions. An epitope sequence of 10 amino acids was inserted into OprH at 10 different sites, and most mutant proteins (3/4) with insertions in predicted cell-surface loop regions were expressed essentially as much as the wild type, whereas all proteins (6/6) with insertions in predicted transmembrane β-strands were expressed at undetectable levels [42]. Insertion length at permissive sites in cell-surface exposed areas has been extended for several outer-membrane proteins: a 21-amino acid (aa) insertion in OmpA [44], 30–50 aa in PhoE [45] and about 60 aa in LamB [46]. Such extension should considerably broaden the applications of live bacterial vaccine (cell-surface engineering; see below).

Secretory protein

A pioneering study of insertional mutagenesis of soluble proteins has been performed by Barany [34] for a secretory protein, TEM-1 β-lactamase. Two-codon insertions gave rise to mutants at eight different sites, six of which had residual enzyme activity. The three-dimensional structure of β-lactamase was determined, and all but one insertion present in loops or near the ends of α-helices or β-strands [47], as has been confirmed by recent results of studies on pentapeptides insertions in β-lactamase at 23 different sites [48].

Structural effects of a few amino acid insertions were precisely investigated for a large number of insertion mutants of staphylococcal nuclease [49–52] and T4 lysozyme [53–55]. These studies demonstrated that it is possible for proteins to accommodate the addition of residues in helices without a gross loss of function. The X-ray crystal structures of insertion mutants in helices demonstrated two distinct responses, termed 'bulge'; insertions caused a looping out within the helix [51], and 'register shift'; the inserted amino acids were incorporated into the helix, and wild-type amino acids were translocated from the helix into the preceding loop [53].

To date, permissive length of insertions at permissive sites of secretory proteins has been extended by only 20 residues. Two amino acids were inserted into adenylate cyclase at 14 different sites, and consequently 5 permissive sites tolerated the subsequent insertion of 16 amino acids [56]. *E. coli* alkaline phosphatase was generally tolerant of insertions in exterior loops, but was inactivated by insertions within α-helices or β-strands. These surface loops serve as vehicles for larger foreign peptide sequences, such as the 15-residue analogue of dynorphin [57] and 18-residue somatostatin [58].

Secretion of proteins is more sensitive to insertional alterations than is enzymatic activity. Small linkers (usually two codons) have been inserted into the cholesteryl ester transfer protein (CETP) gene at 18 different sites [59]. Although the secretion of each mutant protein was less than that of wild-type CETP, most of the mutants had normal cholesteryl ester transfer activity. Essentially the same has been noted for insertions into human lysozyme [60]. The low productivity of mutant lysozymes

containing long extra sequences in the secreted fraction may possibly be due to impairment of the secretory pathway of these proteins. Activity was much the same as that of the native form, indicating no effect of insertions on the active cleft of human lysozyme.

Intracellular protein

Many cytoplasmic proteins, particularly enzymes which recognize nucleic acids as substrates, have been examined by insertional mutagenesis [61–69]. Li and Chandrasegaran cause the cleavage distance of Fok I restriction endonuclease to change on being inserted within a linker region between recognition and cleavage domains [67]. Although amino acid insertions usually lessen the stability of mutants, Ishikawa et al. produced a stable mutant of *E. coli* RNase HI by single glycine insertion [69]. Relatively large insertions (more than 20 amino acids) have been made at the permissive sites of intracellular proteins; 23 aa insertion in Fok I [68], 25 aa in thioredoxin [70], 27 aa in β-galactosidase [71] and 31 aa in *E. coli* lac repressor [66], without gross loss of function.

Recently, we attempted more than 100 amino acid insertions in the active enzyme *E. coli* RNase HI [72]. To find permissive sites for insertions, two to five codons were preliminarily inserted at seven naturally occurring restriction sites in *E. coli* RNase HI. As shown in figure 3, the insertions at three sites within secondary structure elements abolished RNase H activity, whereas those at four sites within turns or at ends of secondary structures retained RNase H activity (N. Doi, M. Itaya and H. Yanagawa, unpublished data). The former sites were near active sites (open circles in fig. 3), and the latter on the surface of the enzyme. Random sequences of 120–130 amino acid residues were thus inserted at a permissive site (His 62) of *E. coli* RNase HI. The library was screened, and 10% of the clones were found to retain >1% of wild-type RNase H activity [72].

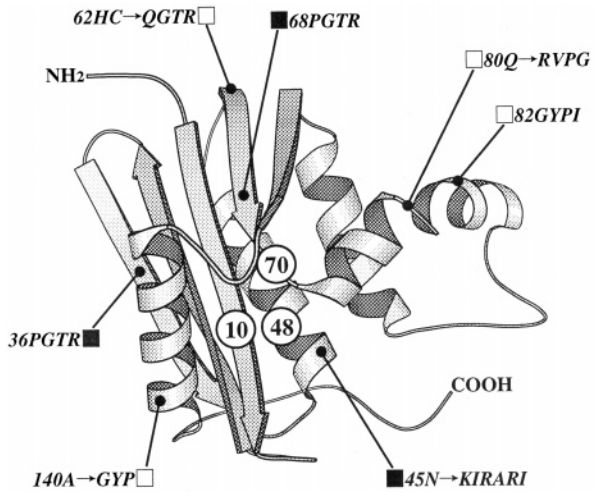

Figure 3. Sites of two to five codon insertions in the three-dimensional structure of *E. coli* RNase HI. Numbers corresponding to insertion positions and sequences of individual insertions are indicated. Open and filled boxes next to insertion numbers indicate insertions that maintained and caused the loss of RNase H activity, respectively. The three open circles indicate active-site residues. The crystal structure of *E. coli* RNase HI was determined by Katayanagi et al. [105].

Such large polypeptide insertions may also occur in nature. There are several examples of inserted multidomain proteins whose domains are connected by insertion of one domain (the inserted domain) into another (the parent domain) [73]. For example, the disulfide oxidoreductase DsbA would be constructed by inserting a large segment into the middle of ancestral thioredoxin. Nearly all parent domains observed so far are intracellular proteins such as thioredoxin, phosphoribosyl anthranilate isomerase, ribosomal protein L7/L12 and the nucleotide binding domain of glyceraldehyde 3-phosphate dehydrogenase [73].

In summary, examination of the effects of insertions in the three types of proteins indicated the following. Loop regions located far from active sites highly tolerate relatively large insertions, whereas helices rarely accept even a few amino acid insertions. The limits of permissive length of insertions for secretory, membrane and intracellular proteins were 20, 60 and 120 amino acids, respectively (fig. 4A). Insertional alterations tend to prevent the transfer of membrane proteins into membranes and translocation of secretory proteins through membranes. Intracellular proteins may thus possibly be best suited for large polypeptide insertions.

Evolutionary molecular engineering

Differing from the case of RNA ribozymes [3], application of efficient screening methods to proteins or peptides requires the establishment of the physical linkage of each polypeptide to its encoding nucleic acid. Several systems are available for constructing peptide libraries, including display of peptides on the surface of bacteria cells [74], filamentous phage [4] and ribosomes [75]. The direct bonding of messenger RNA (mRNA) to the encoded polypeptide through puromycin in a cell-free translation system has recently been conducted [76, 77]. To construct conformationally constrained polypeptide libraries, insertional mutagenesis of membrane proteins, secretory proteins and intracellular proteins can be used for cell-surface display, phage display and cell-free systems, respectively (fig. 4B).

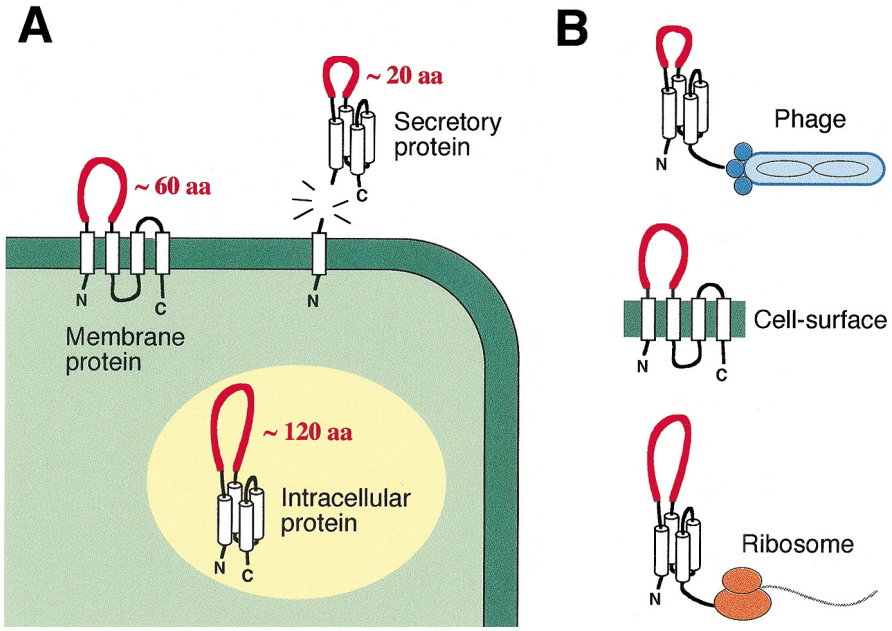

Figure 4. Application of insertional mutagenesis to evolutionary engineering. (*A*) Present upper limits of permissive length of insertions in secretory proteins, membrane proteins and intracellular proteins are indicated. (*B*) The physical linkage of each polypeptide to its encoding nucleic acid, which is necessary for recovery of desired molecules, can be conducted by displaying peptides on filamentous phage, bacterial cell-surface and ribosomes.

Cell-surface display

Expression systems for the display of heterologous proteins on cell surfaces of microorganisms have been developed [74, 78]. Glycosylphosphatidylinositol (GPI) anchored proteins [79, 80] and extracellular proteinaceous appendages such as pili and flagella [81–84] are used for peptide displays. Lu et al. inserted peptides into a thioredoxin domain as a protein scaffold which was introduced into a major structural component of the *E. coli* flagellum [84].

As described above, peptides have been inserted into the surface-exposed loops of outer-membrane proteins as protein scaffolds. The cell-surface engineering using outer-membrane proteins was used with considerable success in practical applications, such as live bacterial vaccines [74], display of antibody and peptide libraries [85, 86], and production of antipeptide antibodies [87]. The screening of bacterial libraries can be done by affinity chromatography or, mainly, by flow cytometry [74].

Phage display

Since development of the fusion phage method in 1985 [4], many peptides and proteins have been displayed on the surface of filamentous bacteriophage as a fusion to either pIII or pVIII coat protein, and the mutants have been selected for improved binding [88, 89]. Phage-displayed random peptide libraries have been used to isolate hormones, epitopes, enzyme inhibitors and receptor antagonists [90]. Phage can also be used to display folded antibody fragments [91], providing means for making antibody fragments in bacteria and as an alternative to hybridoma technology [92]. Phage display is also useful for modifying the substrate specificity of enzymes, such as *E. coli* alkaline phosphatase, trypsin, β-lactamase and hen egg-white lysozyme [93–96].

Most phage-displayed domains attempted so far have been secretory proteins that can be translocated across *E. coli* inner membranes into periplasm and may serve as protein scaffolds for displaying conformationally constrained peptides on phage. Several protein domains are presently in use as scaffolds, such as Tendamistat [97], minibody [98] and cytochrome b_{562} [99]. The randomized loops of scaffold proteins represent only the replacement of residues, such as variable loops of antibodies. O'Neil et al. probably first applied the insertional mutagenesis of scaffold protein to a phage display system, in which six residues of foreign random peptides were inserted into a turn of the immunoglobulin G (IgG)-binding domain Strp G displayed on a phage [100]. Constraint of peptide conformations that will enable their conformation to be determined by

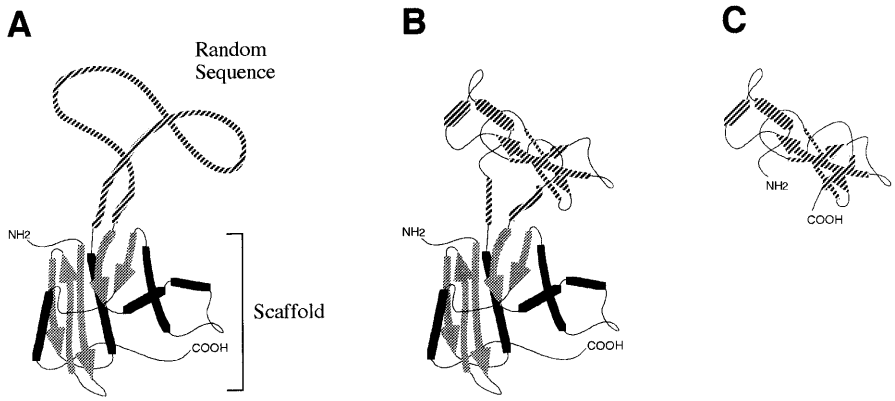

Figure 5. Schematic diagram of new folded-protein construction from random sequences. (*A*) Random sequences are inserted on the surface of a scaffold enzyme. (*B*) The structures of random sequences are optimized using subsequent mutagenesis of the random region by monitoring scaffold enzyme activity. If greater instability or fluctuation of a random protein causes the scaffold enzyme to incur more stress with subsequent reduction in enzyme activity, restoration of activity with mutagenesis of random proteins may result in greater stability and foldability. (*C*) A new folded domain is detached from the scaffold. This process naturally yields the N- and C-terminal proximity of evolved proteins, which has long been observed for many natural proteins [23, 106].

structural analysis should prove useful for drug design. The constraint of conformation of peptide ligands also contributes to increase the affinity, due to a decrease in conformational entropy in unbound states [101].

Cell-free systems

In cell-surface display and phage display systems, peptides and scaffold proteins are synthesized in the cytoplasm and must be translocated across membranes. Recent display systems remove the limitations of proteins that are to be secreted. One system uses peptides fused to the C-terminus of the *lac* repressor [102]. The repressor fusion proteins bind to *lac* operator sequences on plasmids, and are screened in the same manner as for a phage system. The repressor fusions are cytoplasmic, in contrast to the phage fusions, which are exported to the periplasm.

All the above display systems rely on in vivo gene expression, and the size and diversity of the library are limited by transformational efficiency and the biological constraints of the host cells. Hanes and Plückthun constructed a ribosome display system in which a folded protein and its encoding mRNA are linked on the ribosome in vitro [75]. Ribosome display or polysome display [103] systems are worked completely in vitro, and thus larger and a greater variety of molecules are available for selection.

Nemoto et al. quite recently constructed a molecule called 'in vitro virus', in which a translated protein bonds directly to its encoding mRNA [76]. In a cell-free translation system, optional puromycin from the 3′ end of mRNA enters the A site of a ribosome to covalently bind to the C-terminus of the translated polypeptide, and the in vitro virus is then released from the ribosome. The bonding efficiency has been shown to be 10%; thus, for a population of molecules, there are $\sim 10^{12}$ protein variants in 1 ml of solution [76].

Even if selection of novel enzymes from completely random sequence libraries becomes possible using cell-free systems, the screening efficiency of a library can still be increased on a protein scaffold. In ribosome display, a disulfide containing single-chain fragment of antibody served as the scaffold [75]. Antibody domains must form intramolecular disulfide bonds, and this makes the experimental procedures more complicated [75, 104]. Cytoplasmic globular proteins with no disulfide bonds may thus prove more suitable as scaffolds than antibodies in cell-free translation systems.

Directed evolution of random sequence proteins

How did present-day enzymes acquire their well-defined, compact structures? Directed evolution of artificial proteins from random sequences in a laboratory can provide the means for experimental approaches to answering this question. Active enzymes have yet to be isolated from libraries with completely random sequences, but random sequence proteins are useful for examining inherent properties of proteins as a whole. Clarification of the structural properties of unevolved random proteins would yield pertinent information on the origin of folded sequences. The complete or virtual absence of function of present random sequence

proteins, however, makes monitoring hill climbing on a 'fitness landscape' in the sequence space difficult. A library of random sequence proteins displayed on a scaffold enzyme may thus serve as a preliminary model enzyme system, the foldability of whose random sequence proteins can be followed by scaffold enzymatic activity.

Selection of folded proteins

Some attempts have been made to isolate folded proteins from synthetic gene libraries encoding random sequences. The random QLR proteins constructed by Davidson et al. possess highly stable α-helical structures, though this stability is most likely due to the highly helical propensity of glutamine, leucine and arginine [14]. A library of random sequence proteins comprising 20 naturally occurring amino acids (140 residues) designed by Yomo [17] yielded 10% soluble proteins, but no random protein having any notable secondary structure (T. Yomo, unpublished results). Such studies are particularly difficult owing to the absence of an efficient screening system for selecting folded proteins from expression libraries.

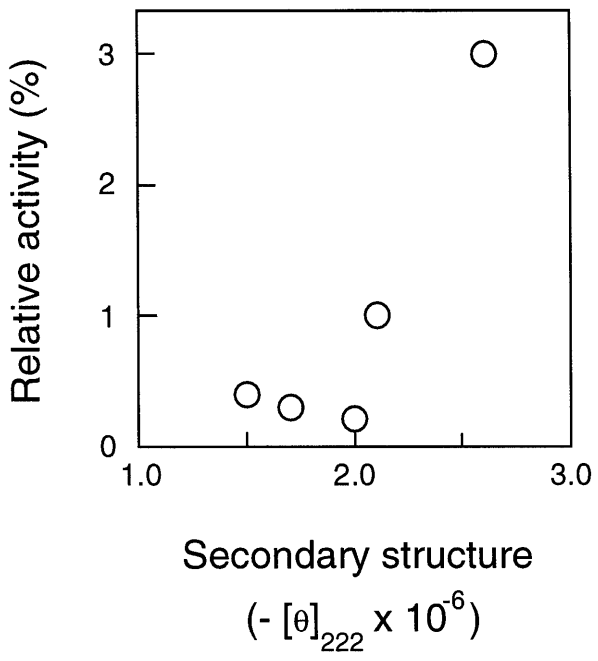

Figure 6. Correlation between the secondary structure and RNase H activity of five purified mutant proteins of *E. coli* RNase HI. The content of secondary structure (abscissa) was estimated from circular dichroism ellipticity at 222 nm, $[\theta]_{222}$ [72]. RNase H activity (ordinate) was determined from radioactivity of the acid-soluble digestion product from ^{32}P-poly(rA) · poly(dT) [72]. Specific activity of the wild-type (100%) was 150 U/mg, where one unit is defined as that generating 1 nmol of acid-soluble nucleotide per 15 min at 37 °C [107].

In the present scaffold enzyme system that uses *E. coli* RNase HI [72], the foldability of random sequence proteins can be monitored by scaffold RNase H activity (fig. 5). The selection of folded proteins from a random library displayed on the surface of an enzyme requires a correlation between scaffold enzyme activity and structural features of inserted random polypeptides. Relative activity of purified mutants of RNase HI remains low, up to 3% wild-type specific activity [72]. When greater instability or fluctuation of a random sequence results in more stress to a scaffold enzyme with subsequent reduction in its activity, insertion mutants with high scaffold activity may be expected to contain random sequence proteins with high foldability and/or stability. In fact, structural and functional analysis of insertion mutants of *E. coli* RNase HI [72] demonstrated the correlation between secondary structure and RNase H activity (fig. 6). The random region of a mutant protein possessing relatively high RNase H activity was detached from the RNase HI scaffold, and its characterization indicated the presence of a significant extent of secondary and tertiary structures (N. Doi, M. Itaya, T. Yomo, and H. Yanagawa, unpublished observations).

De novo proteins with completely native-like structures have yet to be fully designed. The essential elements of folding and design of proteins might be obtained by comparison of folded and nonfolded sequences, not by inspection of folded sequences alone. If the selection of folded proteins from random sequence libraries were made available along with a series of random proteins of gradually increasing foldability, examination of the structures of artificial evolutionary intermediates of random sequence proteins would provide some clarification of sequence–structure relationships of proteins.

Conclusions

The selection of novel enzymes from completely random sequence libraries is an important objective in current biology. The evolutionary engineering of proteins requires (i) diversity and complexity of random sequence libraries with long ORFs, and (ii) a screening method to select novel catalytic enzymes efficiently.

Constraints on polypeptide conformations due to N- and C-terminal proximity on a protein scaffold should result in greater screening efficiency of random libraries. Outer-membrane proteins and secretory proteins are presently being used as protein scaffolds in cell-surface display and phage display systems, respectively. Intracellular proteins with stable compact structures and no disulfide bonds may be considered to be suitable scaffolds for carrying long polypeptides linkable to cell-free systems.

Scaffold enzymes may also serve as structural probes for random sequence proteins. By monitoring scaffold enzymatic activity, optimized sequences with appropriate primary, secondary and tertiary structures can be selected. The feedback of information on optimized sequences to initial library design should make it possible to obtain desired sequences with long ORFs.

Acknowledgements. The authors thank S. Tokura, N. Nemoto and M. Ishizaka for valuable comments on the manuscript and/or figures. We are particularly grateful to M. Itaya for the series of 'magic' strains of *E. coli* and advice on how to assay RNase H activity, and to T. Yomo for important comments and for a plasmid mixture containing designed random sequences. Thanks are also due to K. Katayanagi for the illustration of the crystal structure of *E. coli* RNase HI.

1 Shortle D. and Sondek J. (1995) The emerging role of insertions and deletions in protein engineering. Curr. Opin. Biotechnol. **6:** 387–393
2 Eigen M. and Gardiner W. (1984) Evolutionary molecular engineering based on RNA replication. Pure Appl. Chem. **56:** 967–978
3 Bartel D. P. and Szostak J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. Science **261:** 1411–1418
4 Smith G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science **228:** 1315–1317
5 Nygren P.-Å. and Uhlén M. (1997) Scaffolds for engineering novel binding sites in proteins. Curr. Opin. Struct. Biol. **7:** 463–469
6 Vita C. (1997) Engineering novel proteins by transfer of active sites to natural scaffolds. Curr. Opin. Biotechnol. **8:** 429–434
7 Tramontano A., Janda K. D. and Lerner R. A. (1986) Catalytic antibodies. Science **234:** 1566–1570
8 Pollack S. J., Jacobs J. W. and Schultz P. G. (1986) Selective chemical catalysis by an antibody. Science **234:** 1570–1573
9 Hollfelder F., Kirby A. J. and Tawfik D. S. (1996) Off-the-shelf proteins that rival tailor-made antibodies as catalysts. Nature **383:** 60–63
10 Benkovic S. J. (1996) The key is in the pocket. Nature **383:** 23–24
11 Betz S. F., Raleigh D. P. and DeGrado W. F. (1993) De novo protein design: from molten globules to native-like states. Curr. Opin. Struct. Biol. **3:** 601–610
12 LaBean T. H. and Kauffman S. A. (1993) Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics. Protein Sci. **2:** 1249–1254
13 Kamtekar S., Schiffer J. M., Xiong H., Babik J. M. and Hecht M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science **262:** 1680–1685
14 Davidson A. R. and Sauer R. T. (1994) Folded proteins occur frequently in libraries of random amino acid sequences. Proc. Natl. Acad. Sci. USA **91:** 2146–2150
15 Davidson A. R., Lumb K. J. and Sauer R. T. (1995) Cooperatively folded proteins in random sequence libraries. Nature Struct. Biol. **2:** 856–864
16 Mandecki W. (1990) A method for construction of long randomized open reading frames and polypeptides. Protein Eng. **3:** 221–226
17 Prijambada I. D., Yomo T., Tanaka F., Kawama T., Yamamoto K., Hasegawa A. et al. (1996) Solubility of artificial proteins with random sequences. FEBS Lett. **382:** 21–25
18 Shiba K., Takahashi Y. and Noda T. (1997) Creation of libraries with long ORFs by polymerization of a microgene. Proc. Natl. Acad. Sci. USA **94:** 3805–3810
19 Yoshida K., Shibata T., Masai J., Sato K., Noguti T., Go M. et al. (1993) Protein anatomy: spontaneous formation of filamentous helical structures from the N-terminal module of barnase. Biochemistry **32:** 2162–2166
20 Yanagawa H., Yoshida K., Torigoe C., Park J.-S., Sato K., Shirai T. et al. (1993) Protein anatomy: functional roles of barnase module. J. Biol. Chem. 268: 5861–5865
21 Gilbert W. (1987) The exon theory of genes. Cold Spring Harbor Symp. Quant. Biol. **52:** 901–904
22 Lau K. F. and Dill K. A. (1990) Theory for protein mutability and biogenesis. Proc. Natl. Acad. Sci. USA **87:** 638–642
23 Christopher J. A. and Baldwin T. O. (1996) Implications of N and C-terminal proximity for protein folding. J. Mol. Biol. **257:** 175–187
24 de Gennes P.-G. (1979) In: Scaling Concepts in Polymer Physics, pp. 39–42 and 201–202, Cornell University Press, Ithaca, NY
25 Stone J. C., Atkinson T., Smith M. and Pawson T. (1984) Identification of functional regions in the transforming protein of Fujinami sarcoma virus by in-phase insertion mutagenesis. Cell **37:** 549–558
26 Lyman S. D. and Rohrschneider L. R. (1987) Analysis of functional domains of the v-*fms*-encoded protein of Susan McDonough strain feline sarcoma virus by linker insertion mutagenesis. Mol. Cell. Biol. **7:** 3287–3296
27 Goff S. P. and Prasad V. R. (1991) Linker insertion mutagenesis as probe of structure–function relationships. Methods Enzymol. **208:** 586–603
28 Hofnung M., Bedouelle H., Boulain J. C., Clément J. M., Charbit A., Duplay P. et al. (1988) Genetic approaches to the study and use of proteins: random point mutations and random linker insertions. Bull. Inst. Pasteur **86:** 95–101
29 Pascarella S. and Argos P. (1992) Analysis of insertions/deletions in protein structures. J. Mol. Biol. **224:** 461–471
30 Benner S. A., Cohen M. A. and Gonnet G. H. (1993) Empirical and structural models for insertions and deletions in the divergent evolution of proteins. J. Mol. Biol. **229:** 1065–1082
31 Viguera A.-R. and Serrano L. (1997) Loop length, intramolecular diffusion and protein folding. Nature Struct. Biol. **4:** 939–946
32 Ladurner A. G. and Fersht A. R. (1997) Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. J. Mol. Biol. **273:** 330–337
33 Kyte J. and Doolittle R. F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132
34 Barany F. (1985) Two-codon insertion mutagenesis of plasmid genes by using single-stranded hexameric oligonucleotides. Proc. Natl. Acad. Sci. USA **82:** 4202–4206
35 Borjigin J. and Nathans J. (1994) Insertional mutagenesis as a probe of rhodopsin's topography, stability and activity. J. Biol. Chem. **269:** 14715–14722
36 McKenna E., Hardy D. and Kaback H. R. (1992) Insertional mutagenesis of hydrophilic domains in the lactose permease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **89:** 11954–11958
37 Manoil C. and Bailey J. (1997) A simple screen for permissive sites in proteins: analysis of *Escherichia coli lac* permease. J. Mol. Biol. **267:** 250–263
38 Marti T., Otto H., Rösselet S. J., Heyn M. P. and Khorana H. G. (1992) Consequences of amino acid insertions and/or deletions in transmembrane helix C of bacteriorhodopsin. Proc. Natl. Acad. Sci. USA **89:** 1219–1223
39 Wang S. and Vik S. B. (1994) Single amino acid insertions probe the α subunit of the *Escherichia coli* F_1F_0-ATP synthase. J. Biol. Chem. **269:** 3095–3099
40 Mingarro I., Whitley P., Lemmon M. A. and von Heijne G. (1996) Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix: a rapid way to map helix–helix interactions in integral membrane proteins. Protein Sci. **5:** 1339–1341

41 Wong R. S. Y., Wirts R. A. and Hancock R. E. W. (1995) *Pseudomonas aeruginosa* outer membrane protein OprF as an expression vector for foreign epitopes: the effects of positioning and length on the antigenicity of the epitope. Gene **158:** 55–60
42 Rehm B. H. A. and Hancock R. E. W. (1996) Membrane topology of the outer membrane protein OprH from *Pseudomonas aeruginosa*: PCR-mediated site-directed insertion and deletion mutagenesis. J. Bacteriol. **178:** 3346–3349
43 Koebnik R. and Braun V. (1993) Insertion derivatives containing segments of up to 16 amino acids identify surface- and periplasm-exposed regions of the FhuA outer membrane receptor of *Escherichia coli* K-12. J. Bacteriol. **175:** 826–839
44 Freudl R. (1989) Insertion of peptides into cell-surface-exposed areas of the *Escherichia coli* OmpA protein does not interfere with export and membrane assembly. Gene **82:** 229–236
45 Agterberg M., Adriaanse H., van Bruggen A., Karperien M. and Tommassen J. (1990) Outer-membrane PhoE protein of *Escherichia coli* K-12 as an exposure vector: possibilities and limitations. Gene **88:** 37–45
46 Charbit A., Molla A., Saurin W. and Hofnung M. (1988) Versatility of a vector for expressing foreign polypeptides at the surface of Gram-negative bacteria. Gene **70:** 181–189
47 Zebala J. and Barany F. (1991) Mapping catalytically important regions of an enzyme using two-codon insertion mutagenesis: a case study correlating β-lactamase mutants with the three-dimensional structure. Gene **100:** 51–57
48 Hallet B., Sherratt D. J. and Hayes F. (1997) Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. Nucleic Acids Res. **25:** 1866–1867
49 Sondek J. and Shortle D. (1990) Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. Proteins Struct. Funct. Genet. **7:** 299–305
50 Sondek J. and Shortle D. (1992) Structural and energetic differences between insertions and substitutions in staphylococcal nuclease. Proteins Struct. Funct. Genet. **13:** 132–140
51 Keefe L. J., Sondek J., Shortle D. and Lattman E. E. (1993) The α aneurism: a structural motif revealed in an insertion mutant of staphylococcal nuclease. Proc. Natl. Acad. Sci. USA **90:** 3275–3279
52 Keefe L. J., Quirk S., Gittis A., Sondek J. and Lattman E. E. (1994) Accommodation of insertion mutations on the surface and in the interior of staphylococcal nuclease. Protein Sci. **3:** 391–401
53 Heinz D. W., Baase W. A., Dahlquist F. W. and Matthews B. W. (1993) How amino acid insertions are allowed in an α-helix of T4 lysozyme. Nature **361:** 561–564
54 Heinz D. W., Baase W. A., Zhang X.-J., Blaber M., Dahlquist F. W. and Matthews B. W. (1994) Accommodation of amino acid insertions in an α-helix of T4 lysozyme. Structural and thermodynamic analysis. J. Mol. Biol. **236:** 869–886
55 Vetter I. R., Baase W. A., Heinz D. W., Xiong J.-P., Snow S. and Matthews B. W. (1996) Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. Protein Sci. **5:** 2399–2415
56 Ladant D., Glaser P. and Ullmann A. (1992) Insertional mutagenesis of *Bordetella pertussis* adenylate cyclase. J. Biol. Chem. **267:** 2244–2250
57 Freimuth P. I., Taylor J. W. and Kaiser E. T. (1990) Introduction of guest peptides into *Escherichia coli* alkaline phosphatase. J. Biol. Chem. **265:** 896–901
58 Langen H. T. and Taylor J. W. (1992) Alkaline phosphatase–somatostatin hybrid proteins as probes for somatostatin-14 receptors. Proteins Struct. Funct. Genet. **14:** 1–9
59 Wang S., Deng L., Brown M. L., Agellon L. B. and Tall A. R. (1991) Structure–function studies of human cholesteryl ester transfer protein by linker insertion scanning mutagenesis. Biochemistry **30:** 3484–3490
60 Yamada T., Matsushima M., Inaka K., Ohkubo T., Uyeda A., Maeda T. et al. (1993) Structural and functional analyses of the Arg-Gly-Asp sequence introduced into human lysozyme. J. Biol. Chem. **268:** 10588–10592
61 Barany F. (1987) A genetic system for isolation and characterization of *Taq* I restriction endonuclease mutants. Gene **56:** 13–27
62 Barany F. and Zebala J. (1992) Correlation between insertion mutant activities and amino acid sequence identities of the *Taq* I and *Tth* HB8 restriction endonucleases. Gene **112:** 13–20
63 Starzyk R. M., Burbaum J. J. and Schimmel P. (1989) Insertion of new sequences into the catalytic domain of an enzyme. Biochemistry **28:** 8479–8484
64 Kimura M. and Ishihama A. (1995) Functional map of the alpha subunit of *Escherichia coli* RNA polymerase: insertion analysis of the amino-terminal assembly domain. J. Mol. Biol. **248:** 756–767
65 Dykxhoorn D. M., St. Pierre R., Van Ham O. and Linn T. (1997) An efficient protocol for linker scanning mutagenesis: analysis of the translational regulation of an *Escherichia coli* RNA polymerase subunit gene. Nucleic Acids Res. **25:** 4209–4218
66 Nelson B. D., Manoil C. and Traxler B. (1997) Insertion mutagenesis of the *lac* repressor and its implications for structure-function analysis. J. Bacteriol. **179:** 3721–3728
67 Li L. and Chandrasegaran S. (1993) Alteration of the cleavage distance of *Fok* I restriction endonuclease by insertion mutagenesis. Proc. Natl. Acad. Sci. USA **90:** 2764–2768
68 Kim Y.-G., Li L. and Chandrasegaran S. (1994) Insertion and deletion mutants of *Fok* I restriction endonuclease. J. Biol. Chem. **269:** 31978–31982
69 Ishikawa K., Nakamura H., Morikawa K., Kimura S. and Kanaya S. (1993) Cooperative stabilization of *Escherichia coli* ribonuclease HI by insertion of Gly-80b and Gly-77→Ala substitution. Biochemistry **32:** 7136–7142
70 LaVallie E. R., DiBlasio E. A., Kovacic S., Grant K. L., Schendel P. F. and McCoy J. M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Bio/Technology **11:** 187–193
71 Benito A. and Villaverde A. (1994) Insertion of a 27 amino acid viral peptide in different zones of *Escherichia coli* β-galactosidase: effects on the enzyme activity. FEMS Microbiol. Lett. **123:** 107–112
72 Doi N., Itaya M., Yomo T., Tokura S. and Yanagawa H. (1997) Insertion of foreign random sequences of 120 amino acid residues into an active enzyme. FEBS Lett. **402:** 177–180
73 Russell R. B. (1994) Domain insertion. Protein Eng. **7:** 1407–1410
74 Georgiou G., Stathopoulos C., Daugherty P. S., Nayak A. R., Iverson B. L. and Curtiss R. III (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nature Biotechnol. **15:** 29–34
75 Hanes J. and Plückthun A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA **94:** 4937–4942
76 Nemoto N., Miyamoto-Sato E., Husimi Y. and Yanagawa H. (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. FEBS Lett. **414:** 405–408
77 Roberts R. W. and Szostak J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA **94:** 12297–12302
78 Georgiou G., Poetschke H. L., Stathopoulos C. and Francisco J. A. (1993) Practical applications of engineering Gram-negative bacterial cell surfaces. Trends Biotechnol. **11:** 6–10
79 Medof M. E., Nagarajan S. and Tykocinski M. L. (1996) Cell-surface engineering with GPI-anchored proteins. FASEB J. **10:** 574–586
80 Murai T., Ueda M., Yamamura M., Atomi H., Shibasaki Y., Kamasawa N. et al. (1997) Construction of a starch-utilizing yeast by cell surface engineering. Appl. Environ. Microbiol. **63:** 1362–1366

81 Thiry G., Clippe A., Scarcez T. and Petre J. (1989) Cloning of DNA sequences encoding foreign peptides and their expression in the K88 pili. Appl. Environ. Microbiol. **55:** 984–993
82 Steidler L., Remaut E. and Fiers W. (1993) Pap pili as a vector system for surface exposition of an immunoglobulin G-binding domain of protein A of *Staphylococcus aureus* in *Escherichia coli*. J. Bacteriol. **175:** 7639–7643
83 Pallesen L., Poulsen L. K., Christiansen G. and Klemm P. (1995) Chimeric FimH adhesin of type 1 fimbriae: a bacterial surface display system for heterologous sequences. Microbiology **141:** 2839–2848
84 Lu Z., Murray K. S., Cleave V. V., LaVallie E. R., Stahl M. L. and McCoy J. M. (1995) Expression of thioredoxin random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: a system designed for exploring protein–protein interactions. Bio/Technology **13:** 366–372
85 Francisco J. A., Campbell R., Iverson B. L. and Georgiou G. (1993) Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc. Natl. Acad. Sci. USA **90:** 10444–10448
86 Brown S. (1992) Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage λ receptor. Proc. Natl. Acad. Sci. USA **89:** 8651–8655
87 Martineau P., Charbit A., Leclerc C., Werts C., O'Callaghan D. and Hofnung M. (1991) A genetic system to elicit and monitor anti-peptide antibodies without peptide synthesis. Bio/Technology **9:** 170–172
88 Roberts B. L., Markland W., Ley A. C., Kent R. B., White D. W., Guterman S. K. et al. (1992) Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. Proc. Natl. Acad. Sci. USA **89:** 2429–2433
89 Lowman H. B., Bass S. H., Simpson N. and Wells J. A. (1991) Selecting high-affinity binding proteins by monovalent phage display. Biochemistry **30:** 10832–10838
90 Burritt J. B., Bond C. W., Doss K. W. and Jesaitis A. J. (1996) Filamentous phage display of oligopeptide libraries. Anal. Biochem. **238:** 1–13
91 McCafferty J., Griffiths A. D., Winter G. and Chiswell D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. Nature **348:** 552–554
92 Harrison J. L., Williams S. C., Winter G. and Nissim A. (1996) Screening of phage antibody libraries. Methods Enzymol. **267:** 83–109
93 McCafferty J., Jackson R. H. and Chiswell D. J. (1991) Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. Protein Eng. **4:** 955–961
94 Corey D. R., Shiau A. K., Yang Q., Janowski B. A. and Craik C. S. (1993) Trypsin display on the surface of bacteriophage. Gene **128:** 129–134
95 Soumillion P., Jespers L., Bouchet M., Marchand-Brynaert J., Winter G. and Fastrez J. (1994) Selection of β-lactamase on filamentous bacteriophage by catalytic activity. J. Mol. Biol. **237:** 415–422
96 Maenaka K., Furuta M., Tsumoto K., Watanabe K., Ueda Y. and Kumagai I. (1996) A stable phage-display system using a phagemid vector: phage display of hen egg-white lysozyme (HEL), *Escherichia coli* alkaline phosphatase and anti-HEL monoclonal antibody, HyHEL10. Biochem. Biophys. Res. Commun. **218:** 682–687
97 McConnell S. J. and Hoess R. H. (1995) Tendamistat as a scaffold for conformationally constrained phage peptide libraries. J. Mol. Biol. **250:** 460–470
98 Venturini S., Martin F. and Sollazzo M. (1994) Phage display of the minibody: a β-scaffold for the selection of conformationally-constrained peptides. Protein Peptide Lett. **1:** 70–75
99 Ku J. and Schultz P. G. (1995) Alternate protein frameworks for molecular recognition. Proc. Natl. Acad. Sci. USA **92:** 6552–6556
100 O'Neil K. T., DeGrado W. F. and Hoess R. H. (1994) Phage display of random peptides on a protein scaffold. In: Techniques in Protein Chemistry V, pp. 517–524, Crabb J. (ed.), Academic Press, San Diego
101 Fairlie D. P., Abbenante G. and March D. R. (1995) Macrocyclic peptidomimetics – forcing peptides into bioactive conformations. Curr. Med. Chem. **2:** 654–686
102 Cull M. G., Miller J. F. and Schatz P. J. (1992) Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. Proc. Natl. Acad. Sci. USA **89:** 1865–1869
103 Mattheakis L. C., Bhatt R. R. and Dower W. J. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. Proc. Natl. Acad. Sci. USA **91:** 9022–9026
104 Ryabova L. A., Desplancq D., Spirin A. S. and Plückthun A. (1997) Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. Nature Biotechnol. **15:** 79–84
105 Katayanagi K., Miyagawa M., Matsushima M., Ishikawa M., Kanaya S., Ikehara M. et al. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. Nature **347:** 306–309
106 Thornton J. M. and Sibanda B. L. (1983) Amino and carboxy-terminal regions in globular proteins. J. Mol. Biol. **167:** 443–460
107 Itaya M. (1990) Isolation and characterization of a second RNase H (RNase HII) of *Escherichia coli* K-12 encoded by the *rnhB* gene. Proc. Natl. Acad. Sci. USA **87:** 8587–8591