J. Biochem. 2010;147(3):371–379 doi:10.1093/jb/mvp179

JB THE JOURNAL OF BIOCHEMISTRY

Mimicking the evolution of a thermally stable monomeric four-helix bundle by fusion of four identical single-helix peptides

Received September 6, 2009; accepted October 16, 2009; published online November 4, 2009

Satoshi Akanuma, Taku Matsuba, Emi Ueno, Naoki Umeda and Akihiko Yamagishi*

Department of Molecular Biology, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan.

*Akihiko Yamagishi, Department of Molecular Biology, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan, Tel.: +81-426-76-7139; Fax: +81-426-76-7145, E-mail: yamagish@ls.toyaku.ac.jp

Internal symmetry is a common feature of the tertiary structures of proteins and protein domains. Probably, because the genes of homo-oligomeric proteins duplicated and fused, their evolutionary descendants are proteins with internal symmetry. To identify any advantages that cause monomeric proteins with internal symmetry to be selected evolutionarily, we characterized some of the physical properties of a recombinant protein with a sequence consisting of two tandemly fused copies of the *Escherichia coli* Lac repressor C-terminal α-helix. This polypeptide exists in solution mainly as dimer that likely maintains a four-helix bundle motif. Thermal unfolding experiments demonstrate that the protein is considerably more stable at elevated temperatures than is a homotetramer consisting of four non-covalently associated copies of a 21-residue polypeptide similar in sequence to that of the Lac repressor C-terminal α-helix. A tandem duplication of our helix–loop–helix polypeptide yields an even more thermally stable protein. Our results exemplify the concept that fusion of non-covalently assembled polypeptide chains leads to enhanced protein stability. Herein, we discuss how our work relates to the evolutionary selective-advantages realized when symmetrical homo-oligomers evolve into monomers. Moreover, our thermally stable single-chain four-helix bundle protein may provide a robust scaffold for development of new biomaterials.

Keywords: four-helix bundle/fusion/internal symmetry/tandem duplication/thermal stability.

Abbreviation: CD, circular dichroism; PCR, polymerase chain reaction.

Internal symmetry is often an attribute of the tertiary structures of natural proteins. Gene duplication and subsequent fusion of the duplicated genes is presumed to be a very common evolutionary mechanism for the appearance of proteins and protein domains with internal symmetry (*1*–*3*). The triosephosphate isomerase (TIM) barrel is a highly symmetrical protein motif, consisting of eight repeats of a β-strand/α-helix module (*4*). The crystal structures of two TIM-barrel-type enzymes that are members of the histidine biosynthetic pathway show internal structural 2-fold symmetry (*3*). Experimental data has been obtained that is consistent with the idea that both enzymes evolved from an ancestral half-barrel (*5*–*7*). The β-propeller is another protein motif with distinctive symmetry. Recently, a stable β-propeller structure was generated by fusion of six identical WD (tryptophan-aspartate) repeats (*8*). Functional β-propeller lectins have also been reconstructed through homo-oligomeric assembly of smaller fragments (*9*). These studies have substantiated the concept that internally symmetric protein structures emerge as a consequence of gene duplication and fusion. Because the risk associated with residue misincorporation should be smaller and because the coding economy should be better (*10*), it may be thought that biosynthesis of a polypeptide that subsequently associates to form a homo-oligomer would be favoured over the biosynthesis of a single-chain (monomeric) protein that arose via gene duplication and fusion. Nevertheless, because single-chain proteins with internal symmetry exist, certain putative ancestral homo-oligomers seem to have evolved into single polypeptide chains. Given that internal symmetry is a common feature among protein structures, there must be some selective advantages driving the appearance of such proteins.

The four-helix bundle is another protein motif with striking internal symmetry. Proteins with four-helix bundles are abundant in nature, though diverse in function (*11*). The bundle consists of four amphiphilic α-helices packed against each other so as to form a hydrophobic core. In nature, this fold often occurs as an isolated form. For instance, cytochrome b_{562}, which is involved in electron transfer, is a monomeric four-helix bundle protein (*12*, *13*); the RNA-binding protein, Rop is a four-helix bundle protein that is formed by the assembly of two identical helix-turn-helix monomers, *i.e.* it is a homodimer (*14*, *15*). The helix-bundle motif is also prevalent at protein subunit interfaces where it helps to stabilize subunit–subunit interactions (*16*, *17*). The fact that four-helix bundles exist naturally as monomers, homodimers and homotetramers gives credence to the idea that the monomers and homodimers arose via gene duplication and subsequent fusion events.

The Lac repressor is a DNA-binding protein that negatively regulates transcription of bacterial *lac* genes (*18*). The *Escherichia coli* Lac repressor is a

tetramer of four identical subunits. Historically, the results of genetic analyses first indicated that the C-terminal α-helices of the Lac repressor subunits associate to form a four-helical bundle with an antiparallel arrangement (*19*). Mutations and deletions in the α-helix abolished the protein's tetrameric structure (*20–22*). Thus, the four C-terminal helices were thought to form an antiparallel four-helix bundle, which then significantly stabilized the Lac repressor's other subunit interactions. This model was confirmed when the X-ray crystal structure of the Lac repressor was solved (Fig. 1A) (*23*). Within the structure, the four-helix bundle is clearly a separate domain, and it should therefore be stable even if it was separated from the rest of the protein. In fact, four copies of a synthetic oligopeptide Lac21—a 21-amino acid residue peptide containing a sequence nearly identical to that of the Lac repressor C-terminal α-helix—fold into a stable, antiparallel four-helix bundle in solution (*24*).

For this report, we reconstructed the evolution of a single-chain four-helix bundle from a single α-helical peptide with the goal of identifying the advantages acquired when a homo-oligomer evolves into a monomer. To this end, using the C-terminal α-helix sequence of the Lac repressor as the template, we synthesized a helix–loop–helix protein, named LARTH (Lac repressor two-helix protein) and its duplicated version, named LARFH (Lac repressor four-helix protein). We show that both LARTH and LARFH have similar secondary and tertiary structures, with the tertiary structures likely to be four-helix bundles. We also show that monomeric LARFH is significantly more thermally stable than is dimeric LARTH, which in turn is significantly more thermally stable than tetrameric Lac21. The inverse relationship between the number of discrete polypeptide chains in the four-helix bundles and the thermal stabilities of the bundles suggests that, in nature, duplication and fusion of a gene that encodes a subunit of a homo-oligomeric protein results in a more robust protein.

Materials and Methods

Construction of the genes for LARTH and LARFH

The gene encoding LARTH was synthesized using a polymerase chain reaction (PCR)-mediated method (Fig. 1B). The synthetic

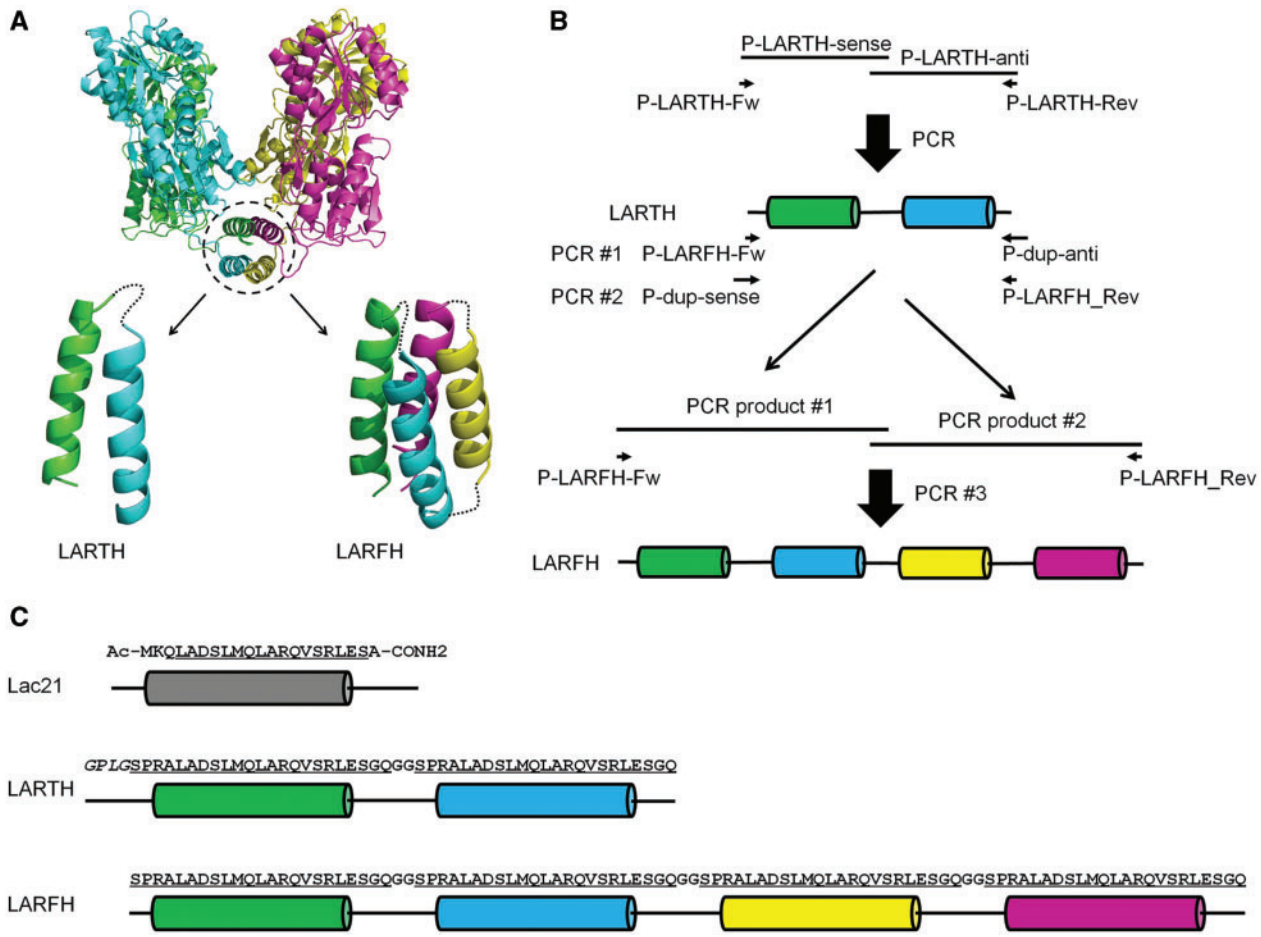

Fig. 1 (A) The tertiary and quaternary structures of an *E. coli* Lac repressor missing its N-terminal DNA binding domains (top); and LARTH and LARFH models (bottom). The structures were drawn using PyMOL (http://www.pymol.org). For the Lac repressor, the subunits are each in a different colour. The dotted lines in the LARTH and LARFH models represent the –S–G–Q–G–G– linkers. **(B)** The synthesis of the genes for LARTH and LARFH. Experimental details are described in 'MATERIALS AND METHODS'. **(C)** The amino acid sequences of Lac21 (*24*), LARTH and LARFH. The sequences originating from the *E. coli* Lac repressor are underlined. The amino acids originating from the vectors' sequences are italicized. Cylinders represent segments that are likely to form α-helices.

Table 1. Oligonucleotide primers used in this study.

Primer	Sequence
P-LARTH-sense	GCGCCGGAATTCCCATATGTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACA GGTTTCCCGACTGGAAAGCGGGCAGGGTGGTAGCCC
P-LARTH-antiz	GGGCCCAAGCTTTTATTATTGGCCAGACTCCAAGCGGCTCACTTGGCGAGCTAATTGCATCAG GCTGTCAGCCAGGGCACGCGGGCTACCACCCTGCCC
P-LARTH-Fw	GGGGGGGGGATCCCCCCGCGCGTTGGCC
P-LARTH-Rev	GGGGGGAATTCTTATTATTGGCCAGACTCCAAG
P- dup-sense	TCAGGGTGGTTCGCCCCGCGCGTTGGCC
P-dup-anti	GGGGCGAACCACCCTGACCAGACTCCAAGCGGCT
P-LARFH-Fw	TCTCCCCGCGCGTTGGCC
P-LARFH-Rev	GGGGGGAAGCTTATTATTGGCCAGACTCCAAG

The *Bam*HI (GGATCC), *Eco*RI (GAATTC) and *Hin*dIII (AAGCTT) restriction sites are underlined.

oligonucleotides used are listed in Table 1. The overlapping oligonucleotides, P-LARTH-sense and P-LARTH-anti, have complementary sequences at their 3′-ends. Another pair of primers, P-LARTH-Fw and P-LARTH-Rev, was included to ensure sufficient DNA amplification. P-LARTH-Fw and P-LARTH-Rev contain *Bam*HI and *Eco*RI restriction sites at their 5′ ends, respectively. For amplification, the PCR-reaction mixture contained 1× PCR buffer for KOD-plus polymerization (Toyobo, Osaka, Japan), 1 mM $MgSO_4$, 0.2 mM each of the dNTPs, 0.2 μM each of the synthetic oligonucleotides and 1.0 unit KOD-plus DNA polymerase. The time–temperature program was: step 1, 95°C, 3 min; step 2, 95°C, 30 s; step 3, 55°C, 30 s; step 4, 68°C, 1 min; steps 2–4 were repeated 16 times. After amplification, the PCR product was treated with *Bam*HI and *Eco*RI (New England Biolabs, Tokyo, Japan) and then purified by agarose gel electrophoresis. To construct the protein carrying a GST-tag at its N-terminus, the purified DNA was ligated into pGEX-6p-1 (GE Healthcare, Piscataway, NJ, USA). The gene encoding LARFH was constructed using the splicing by overlap extension (SOE) PCR-method (*25*) (Fig. 1B). SOE-PCR consists of two rounds of PCR using the outer primer pairs that are common to all reactions and the inner primer pairs that define the splice site. P-LARFH-Fw and P-dup-anti were used for amplification of the 5′ gene segment (PCR #1). P-dup-sense and P-LARFH-Rev were used for amplification of the 3′ gene segment (PCR #2). Both amplifications used pGEX-6p-1 that carried the gene encoding LARTH as the template. The two amplified fragments were mixed and PCR amplified again in the presence of P-LARFH-Fw and P-LARFH-Rev (PCR #3). The conditions and program for PCR were same as those used for amplification of the gene encoding LARTH. After amplification, the PCR product was digested with *Hin*dIII (New England Biolabs, Tokyo, Japan) and then cloned into the *Stu*I-*Hin*dIII site of pQE30Xa (Qiagen, Tokyo, Japan) so that LARFH would be N-terminally His-tagged.

Protein preparation

LARTH was expressed and purified as an N-terminally GST-tagged chain. For its production, *E. coli* BL21 (DE3) was transformed with the expression plasmid and then cultivated in LB medium supplemented with ampicillin (150 μg/ml). Expression was induced with Overnight Express Autoinduction System 1 (Merck, Tokyo, Japan). After overnight cultivation, cells were harvested by centrifugation and disrupted by sonication. Precipitate was removed by centrifugation at 60,000*g* for 20 min and soluble proteins were purified using a GSTrap HP affinity column (GE Healthcare, Piscataway, NJ, USA). The N-terminal GST-tag (excepting residues $G_{-4}P_{-3}L_{-2}G_{-1}$) was then removed by digestion with PreScission Protease (GE Healthcare, Piscataway, NJ, USA) (Fig. 1C). The released GST-tag and PreScission Protease were removed from the solution by passage through a GSTrap HP affinity column. For preparation of LARFH, *E. coli* M15 (pREP4) (Qiagen, Tokyo, Japan) harbouring the expression plasmid was cultivated in LB medium supplemented with ampicillin (150 μg/ml) and kanamycin (25 μg/ml). Expression was induced by adding isopropyl-thio-β-D-galactopyranoside to a final concentration of 0.5 mM when the OD_{600} of the culture medium was between 0.6 and 0.8. After cultivation for an additional 6 h at 37°C, cells were harvested by centrifugation and disrupted by sonication. N-terminally His_6-tagged LARFH was purified from the supernatant of the crude extract by HisTrap FF nickel affinity column chromatography. The purified fusion protein was treated with Factor Xa protease (Qiagen, Tokyo, Japan) to cleave the N-terminal His-tag. The cleaved tag and Factor Xa were then removed from the protein solution by passage through a HisTrap FF nickel affinity column and a HiTrap Q anion exchange column (GE Healthcare, Piscataway, NJ, USA). The purity of each protein was >95% as judged by SDS–polyacrylamide gel electrophoresis.

Analytical methods

Since LARTH and LARFH have no aromatic residues, their protein concentrations were determined using the reagents of a BCA protein assay kit (Pierce, Rockford, IL, USA) and LARFH-W, a LARFH variant containing a C-terminal tryptophan residue, as the standard. The concentration of the LARFH-W solution was determined using its absorbance at 280 nm as described by Pace *et al. (26)* who modified the procedure of Gill and von Hippel (*27*). For analytical gel filtration, proteins were eluted at a flow rate of 0.5 ml/min over Superdex 75 resin (1.0 × 30 cm column dimensions; GE Healthcare, Piscataway, NJ, USA) that was equilibrated with 20 mM potassium phosphate, pH 7.6, 150 mM KCl. Apparent molecular masses of the experimental proteins were determined using their elution volumes and a calibration curve produced using the molecular masses and elution volumes of protein standards. Circular dichroism (CD) measurements were done using a J-720 spectropolarimeter (Jasco, Hachioji, Japan). Far-UV CD spectra were recorded from 190 to 250 nm at 25°C, using a 0.1 cm path-length cell. Protein concentrations were adjusted to 0.1 mg/ml with 20 mM potassium phosphate, pH 7.6, 7.5 mM KCl. For thermal denaturation, the spectropolarimeter was equipped with a programmable temperature controller and a pressure-proof cell compartment that prevented the aqueous solution from bubbling and evaporating at high temperatures. The resistance of LARTH and LARFH to protease digestion was tested by incubating 0.20 mg/ml of each protein with 30-μg/ml proteinase K (Wako, Osaka, Japan) at 37°C in 20 mM potassium phosphate, pH 7.6. The reactions were stopped at various times by adding the same volumes of 2× SDS–PAGE sample buffer to the digests and heating them at 100°C for 90 s. The proteolysis time-courses were analysed after electrophoresis of the digests through 18% polyacrylamide gels in Tris–tricine, pH 8.3.

Results and Discussion

Design and synthesis of LARTH and LARFH

In nature, four-helix bundles are found as monomers, homodimers and homotetramers; therefore, the appearance over time of the monomeric and dimeric structures is likely the result of duplication and fusion of genes that encoded for ancestral helical polypeptides that associated non-covalently. For this study, we experimentally mimicked the outcomes of these putative evolutionary events. The Lac repressor of *E. coli* is a tetramer consisting of four identical subunits. The C-terminal helices of the subunits assemble to form an antiparallel four-helix bundle (Fig. 1A).

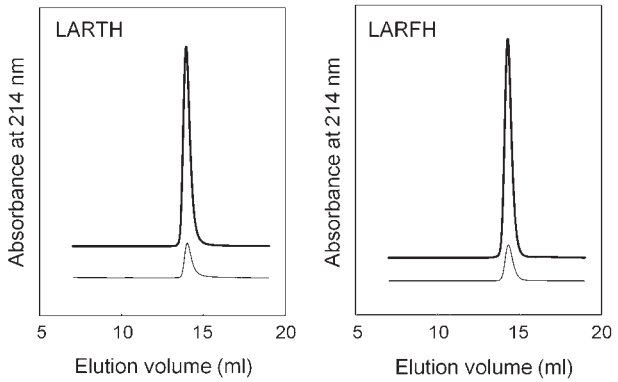

Fig. 2 Superdex 75 analytical gel filtration chromatograms of LARTH and LARFH. The elution rates were always 0.5 ml/min. The initial protein concentrations are 10 μM (thin line) or 50 μM (thick line).

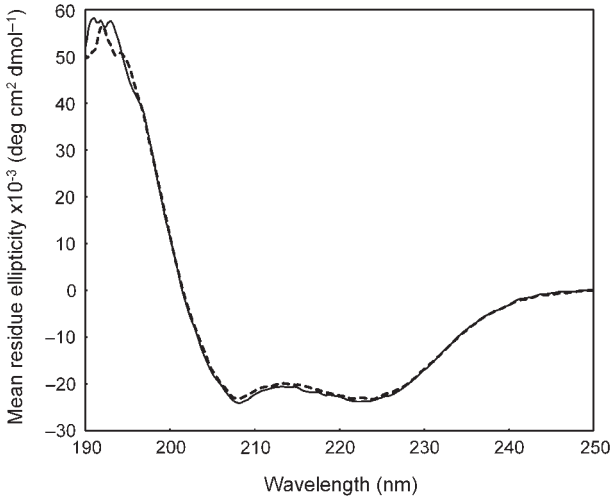

Fig. 3 Far-UV CD spectra of LARTH (solid line) and LARFH (dashed line). Each protein concentration was 0.10 mg/ml in 20 mM potassium phosphate, pH 7.6, 7.5 mM KCl. The cell path-length was 0.1 cm.

Using recombinant techniques, we synthesized LARTH that contains two copies of the Lac repressor C-terminal α-helix sequence tandemly linked via a modified version of the C-terminal sequence of the Lac repressor (–S–G–Q). The last three residues are disordered in the crystal structure of the Lac repressor (*23*), indicating that they have no preferred structural orientation. Two glycines were added to the terminal sequence (SGQGG; Fig. 1B and C). We expected that because glycine lacks a side chain, the presence of the –G–G– sequence, in conjunction with the –S–G–Q– triad would ensure sufficient flexibility, so that the helices flanking the linker could orient properly. This polypeptide was expressed in *E. coli* and then purified from the supernatant of the crude cell extract. This helix–loop–helix protein appears to be very soluble.

We next duplicated and fused the gene for LARTH to generate the gene for the monomeric four-helix bundle LARFH (Fig. 1B). The amino acid sequence of LARFH thus contains four repeats of the C-terminal α-helix sequence of the *E. coli* Lac repressor (Fig. 1A and C). When expressed in *E. coli*, most of it was found in the soluble fraction of the cell homogenate and could be purified from the supernatant of the crude extract.

Structural characterization of LARTH and LARFH

The oligomeric structures of LARTH and LARFH were investigated using analytical gel filtration chromatography (Fig. 2). For LARTH, initial protein concentrations of 10 and 50 μM yield very similar elution profiles, suggesting that protein concentration does not significantly affect the oligomeric state of LARTH. The apparent molecular mass of LARTH (13.7 kDa) is slightly greater than that calculated assuming it exists as a dimer (10.9 kDa). It is well-known that rod-shaped molecules migrate faster than expected, as does the Lac repressor tetramerization domain (*19*). The peaks for LARTH are somewhat asymmetric, suggesting that, at the experimental concentrations of LARTH used, it rapidly equilibrated between a monomer and a dimer on the chromatographic time scale and existed predominantly as a dimer during chromatography. The apparent molecular mass of LARFH, determined by analytical gel filtration is 12.1 kDa, slightly greater than that calculated assuming it is a monomer (10.3 kDa), suggesting that LARFH exists as a pure monomer.

The far-UV CD spectra of LARTH and LARFH are presented in Fig. 3. Far-UV CD spectra reflect the secondary structure contents of proteins. The CD measurements of both proteins show typical spectra for helical proteins, with characteristic minima at 208 and 222 nm. The shape and intensity of the far-UV CD spectrum of LARFH is almost the same as that of LARTH, indicating that the two proteins share similar proportions and types of secondary structure. The analytical gel filtration chromatograms and the CD spectra, taken together, indicate that LARTH and LARFH both form four-helix bundles containing two and one polypeptide chains, respectively.

Stabilities of LARTH and LARFH

Denaturations of LARTH and LARFH were followed by monitoring the changes in ellipticity at 222 nm that occurred when the temperature was increased. The temperature-induced unfolding of each protein is almost reversible, as demonstrated by the recovery of most of the original CD intensity upon cooling (Fig. 4). The effects of three different protein concentrations (0.02, 0.06 and 0.20 mg/ml) and low, physiologic and high salt concentrations (15, 150 and 500 mM KCl) were examined. The melting curves show that both proteins undergo cooperative secondary structure unfolding transitions under all conditions tested (Fig. 5). Similar melting curves were obtained when the temperature was increased at rates of 0.5°C min^{-1} and 1.0°C min^{-1}, indicating that equilibrium was reached at each data point (Fig. 5A). The temperature required for 50% denaturation (T_m) was used to compare the thermal stabilities of the proteins. For LARTH, smaller T_m values were obtained when the protein concentration was decreased; whereas, for

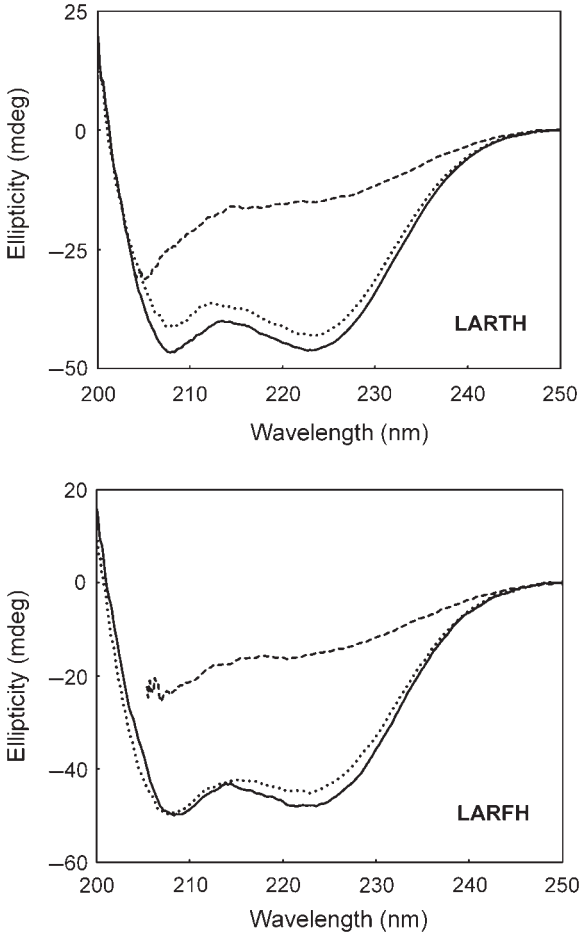

Fig. 4 Far-UV CD spectra of LARTH and LARFH in their native, unfolded and refolded states. The protein concentrations were 0.20 mg/ml and the buffer solution was 20 mM potassium phosphate, pH 7.6, 150 mM KCl. The CD spectra of the native states were recorded at 50°C (LARTH) or 55°C (LARFH). The temperature was increased at a rate of 0.5°C/min to 100°C (LARTH) or 120°C (LARFH) and then the spectra of the unfolded states recorded. The protein solutions were next cooled at a rate of 0.5°C/min and the spectra recorded again at the initial temperatures. Assuming a two-state transition for each protein, more than 90% of each protein refolded upon cooling as judged by the values of the ellipticities at 222 nm. Solid lines, native states; dashed lines, unfolded states; dotted lines, refolded states.

LARFH, protein concentration did not significantly affect the T_m value (Fig. 5B). These results are in agreement with those of the analytical gel filtration experiments and indicate that LARTH exists in a monomer–dimer equilibrium and LARFH as a monomer.

The T_m values for LARTH at a protein concentration of 0.2 mg/ml are 86°C, 89°C and 90°C, respectively, for 15, 150 and 500 mM KCl concentrations (Fig. 5C). Thus, LARTH shows unusually high thermal stability that is similar to those of thermophilic proteins, although its sequence is derived from that of a mesophilic protein. The duplicated helix–loop–helix protein LARFH is much more stable than is LARTH—the T_m values of LARFH are larger than those of LARTH by around 18°C under all salt conditions. The superior resistance to denaturation by LARFH in general suggests that, in nature, it is

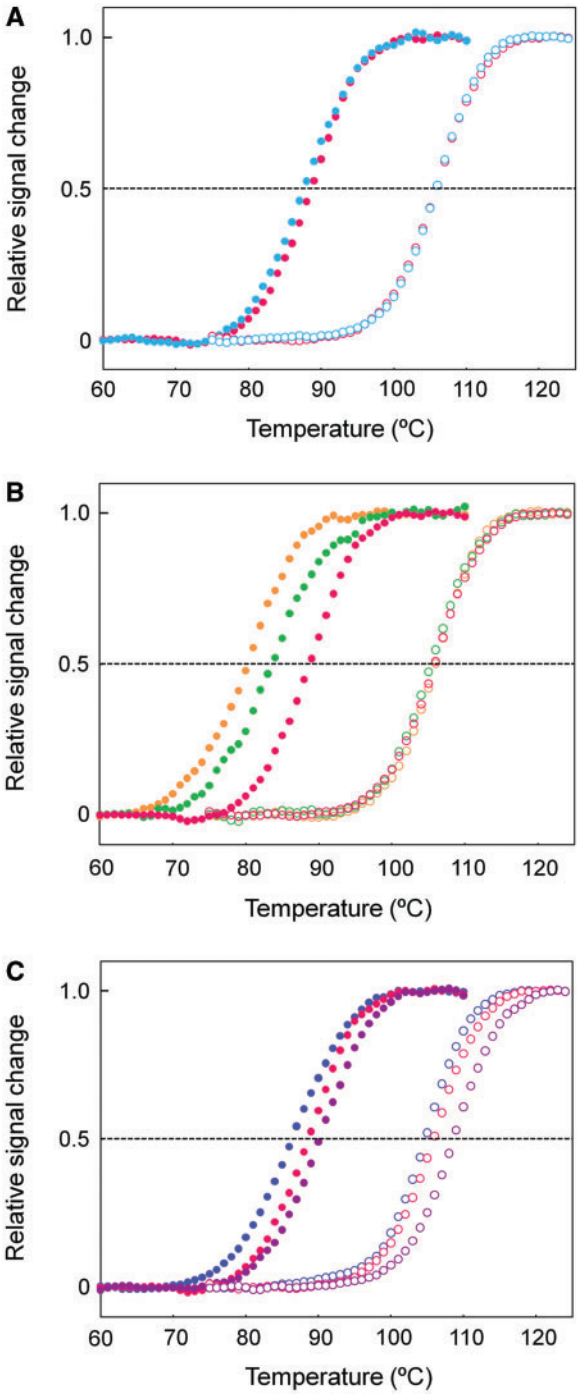

Fig. 5 Thermal melting profiles of LARTH (filled circles) and LARFH (open circles). Ellipticities were monitored at 222 nm. The melting plots were normalized assuming linear temperature dependencies for the baselines of the native and denatured states. (A) Protein solutions were 20 mM potassium phosphate, pH 7.6, 0.20 mg/ml protein, 150 mM KCl. The temperature was increased at 0.5°C/min (magenta) or 1.0°C/min (cyan). For both proteins, increasing the rate did not affect the unfolding curves significantly. Therefore, for each measurement, equilibrium must have been reached. (B) Each protein sample was dissolved in 20 mM potassium phosphate, pH 7.6, 150 mM KCl. The protein concentration was 0.02 mg/ml (orange), 0.06 mg/ml (green) or 0.20 mg/ml (magenta). The temperature was increased at a rate of 0.5°C/min. (C) Each protein was dissolved in 20 mM potassium phosphate, pH 7.6, containing 15 mM (blue), 150 mM (magenta) or 500 mM KCl (purple). The protein concentration was always 0.20 mg/ml. The temperature was increased at a rate of 0.5°C/min.

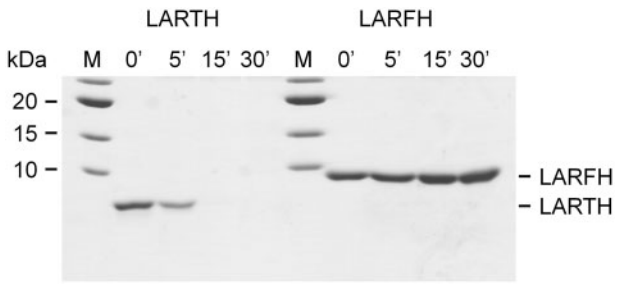

Fig. 6 Resistance of LARTH and LARFH to proteolytic digestion. LARTH and LARFH (0.20 mg/ml each) were incubated for the indicated time periods with proteinase K (30 μg/ml) in 20 mM potassium phosphate, pH 7.6, at 37°C. The proteins were separated through a Tris–tricine, pH 8.3, 18% PAGE gel and their amounts determined by densitometry.

advantageous for a small gene segment to duplicate and fuse in order to generate a larger monomeric protein.

In four-helix bundles, the association of the four helices is driven and stabilized by the hydrophobic core made up of the apolar faces of the amphiphilic helices (*11*). In the *E. coli* Lac repressor four-helix bundle, the interior hydrophobic core is composed mostly of leucine side chains. Accordingly, as the KCl concentration increases, the T_m value of both polypeptides increases owing to enhanced hydrophobic interactions involving the interior core residues.

Artificial proteins containing α-helices have sometimes been found to be unusually stable (*28*). Regan and DeGrado performed an experiment similar to ours using monomeric and oligomeric four-helix bundles constructed from a *de novo*-designed helical peptide (*29*). An inverse correlation was found for the number of discrete polypeptide chains in the four-helix bundles and the stabilities of the bundles. A subsequent NMR study revealed a relatively well-folded helical conformation for the monomeric four-helix bundle but also indicated that its hydrophobic core had molten globule characteristics (*30*). We tested the resistance of LARTH and LARFH to proteinase K digestion, which mainly cleaves at the C-termini of hydrophobic residues. The proteins were incubated with proteinase K for various time periods, after which they were subjected to SDS—-polyacrylamide gel electrophoresis. LARTH was completely digested after 15 min, whereas LARFH remained intact after 30 min (Fig. 6). Thus, LARFH is significantly more stable towards proteolysis than is LARTH. Unlike the monomeric four-helix bundle protein designed by Regan and DeGrado, LARFH mimics a naturally occurring four-helix bundle whose hydrophobic core has been optimized during the course of evolution. Accordingly, our monomeric four-helix bundle protein may have a well-structured hydrophobic core that confers unusually thermal stability upon LARFH, so that it is able to resist proteolysis.

Implications for the origins of monomeric proteins with internal symmetry

It has been argued that, in general, large protein structures are prerequisites for certain protein properties and functions (*10*). As the means to increase protein size, homo-oligomerization has several advantages in comparison with the construction of long, single-chain proteins (reviewed by Ref. (*10*) and references therein). For example, oligomerization increases the size of a protein without increasing its gene size. Therefore it protects against the occurrence of deleterious errors in replication and protein synthesis by allowing a large protein to be built without using additional genetic space. Nevertheless, some oligomers with subunits arrayed symmetrically seemed to have evolved into single polypeptide chains. Therefore, the question arises as to what is the 'evolutionary driving force' that prompts the conversion of an oligomeric protein into a monomer. Oligomeric proteins are usually unstable at very low concentrations, and therefore, secreted proteins are often monomeric (*10*). Similarly, the thermal stability of LARTH is dependent on its concentration; whereas, protein concentration does not affect the thermal stability of the monomeric LARFH (Fig. 5B). It has been postulated that subunit fusion of homo-oligomeric proteins generally improves conformational stability because the entropy of the denatured state is reduced (*31*, *32*). Liang *et al. (33)* engineered single chain variants of the homodimeric protein, gene V protein of bacteriophage f1, by fusing two copies of its subunit. The monomers are significantly more stable against guanidine hydrochloride denaturation than is the parent dimer. They also showed that the increased stabilities correlate with the calculated value for the reduction in entropy expected when the termini of two 'denatured' subunit chains are connected. The connection of two LARTH chains may also reduce the translational and rotational entropies of its denatured state, thus producing a single-chain protein with enhanced thermal stability.

Höcker *et al. (5)* reported that, in isolation, the last four βα units of the $(\beta/\alpha)_8$-barrel protein, imidazoleglycerolphosphate synthase from *Thermotoga maritima*, form primarily a homodimeric protein having significant amounts of native-like secondary and tertiary structures. Tandem duplication and fusion of the $(\beta/\alpha)_4$ half-barrel segment yields a monomeric protein that is more resistant to urea denaturation than is its parent dimeric protein (*6*). Similarly, we demonstrated that an isolated $(\beta/\alpha)_4$ half-barrel segment of the $(\beta/\alpha)_8$-barrel protein, phosphoribosylanthranilate isomerase from *E. coli*, folds into a well-ordered structure that equilibrates between single chain and dimeric states. Further, when two half-barrel segments are covalently connected in tandem, the resulting monomer has increased stability (*34*). Fairman *et al. (24)* reported that four copies of the 21-residue polypeptide, Lac21, containing a sequence nearly identical to that of the C-terminal α-helix of the Lac repressor, form a homotetramer with a (likely) four-helix bundle motif. They reported that its T_m value is 54°C when its protein concentration is 500 μM. Our helix–loop–helix protein LARTH is considerably more thermally stable than is Lac21 even when the solution concentration of LARTH is 0.02 mg/ml (3.6 μM). Furthermore, the tandem duplication of LARTH yields the even more stable monomeric protein LARFH at the given

protein concentrations. Thus, our results and those of others mentioned above illustrate how fusion of the subunits of homodimeric proteins improves protein stability.

The relative stability of a protein can be an important positive influence during the evolution of a protein's function, because an initially large stability can alleviate the potential damage caused by the decrease in stability that often accompanies the appearance of a new function. Such stability-function tradeoffs have been found for a number of functionally beneficial mutations that also are destabilizing (*35*–*38*). Arnold and her co-workers have theoretically and experimentally demonstrated that more stable proteins can more readily develop new functions (*39*, *40*). Similarly, Tawfik and his co-workers recently pointed out that acquisition of a new protein function depends on compensatory, stabilizing mutations in regions that are not involved in the protein's function (*41*). Therefore, the evolutionary process likely involves enhanced protein stability as a prerequisite for the appearance of a new function, and the appearance of monomeric proteins from oligomeric ancestors would be one means of enhancing protein stability especially when the protein concentration is low.

Another property driving the fusion of peptide chains may be their increased resistance to the damaging effects of mutations. Once a mutation occurs in a subunit of a homo-dimeric protein, its effect is compounded. The structures of protein active sites have been optimized for function (*42*, *43*). A single genetic mutation would result in the replacement of two residues in the active site of a homo-dimeric protein—perhaps causing more damage than would replacement of a single residue in the corresponding single-chain protein. This outcome must be a selective 'disadvantage' for dimeric proteins over monomers. Conversely, the potentially smaller mutational effect in a single-chain protein may cause only a subtle structural change, which might be important for fine-tuning of the protein structure during the course of evolution, and therefore could provide an evolutionary selective advantage.

Versatility of thermally stable four-helix bundles

The single-chain four-helix bundle protein engineered in this study may offer a robust molecular building block to fabricate novel biomaterials, such as artificial biocatalysts, biosensors and bioelectronics devices. Designed four-helix bundle polypeptides have provided scaffolds to engineer a variety of functions. A *de novo* designed synthetic helix–loop–helix polypeptide that forms a four-helix bundle upon dimerization has been utilized to control the assembly of gold nanoparticles, which can serve as scaffolds for biosensing materials (*44*). Efforts have been made to introduce specific Zn-binding, haeme-binding and even non-biological cofactor-binding sites into the interior cores of designed bundles (*45*–*52*). Other *de novo* design experiments showed the potential to engineer artificial biocatalysts (*53*–*56*). The successful design of these functional helical proteins has shown that designed four-helix bundles can be used as structural and functional elements in the fabrication of novel biomaterials.

Proteins with a high thermal stability have certain advantages over less stable proteins. Thermally stable proteins are necessary when high process-temperatures are required (*57*). In addition, thermally stable proteins are often resistant to other protein-inactivating factors, such as organic solvents, acidic and alkaline solutions and detergents (*58*). By combining the versatility of the four-helix bundle motif with the unusually high thermal stability of LARFH, it may be an appropriate scaffold that can be used during the design of new biomaterials. However, before such biomaterials can be designed, the structure of LARFH must be determined at the atomic level, and efforts to do so are underway.

Conflict of interest
None declared.

References

1. Weeds, A. G. and McLachlan, A.D. (1974) Structural homology of myosin alkali light chains, troponin C and carp calcium binding protein. *Nature* **252**, 646–649
2. Tang, J., James, M.N., Hsu, I.N., Jenkins, J.A., and Blundell, T.L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**, 618–621
3. Lang, D., Thoma, R., Henn Sax, M., Sterner, R., and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **289**, 1546–1550
4. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D., and Waley, S.G. (1975) Structure of chicken muscle triosephosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609–614
5. Höcker, B., Beismann Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nat. Struct. Biol.* **8**, 32–36
6. Höcker, B., Claren, J., and Sterner, R. (2004) Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. *Proc. Natl Acad. Sci. USA* **101**, 16448–16453
7. Seitz, T., Bocola, M., Claren, J., and Sterner, R. (2007) Stabilisation of a $(\beta\alpha)_8$-barrel protein designed from identical half barrels. *J. Mol. Biol.* **372**, 14–129
8. Nikkhah, M., Jawad-Alami, Z., Demydchuk, M., Ribbons, D., and Paoli, M. (2006) Engineering of beta-propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat. *Biomol. Eng.* **23**, 185–194
9. Yadid, I. and Tawfik, D.S. (2007) Reconstruction of functional β-propeller lectins via homo-oligomeric assembly of shorter fragments. *J. Mol. Biol.* **365**, 10–17
10. Goodsell, D.S. and Olson, A.J. (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153
11. Kamtekar, S. and Hecht, M.H. (1995) Protein Motifs. 7. The four-helix bundle: what determines a fold? *FASEB J.* **9**, 1013–1022
12. Lederer, F., Glatigny, A., Bethge, P.H., Bellamy, H.D., and Matthew, F.S. (1981) Improvement of the 2.5 Å

resolution model of cytochrome b562 by redetermining the primary structure and using molecular graphics. *J. Mol. Biol.* **148**, 427–448
13. Hamada, K., Bethge, P.H., and Mathews, F.S. (1995) Refined structure of cytochrome b562 from *Escherichia coli* at 1.4 Å resolution. *J. Mol. Biol.* **247**, 947–962
14. Banner, D.W., Kokkinidis, M., and Tsernoglou, D. (1987) Structure of the ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.* **196**, 657–675
15. Eberle, W., Pastore, A., Sander, C., and Rosch, P. (1991) The structure of ColE1 rop in solution. *J. Biomol. NMR* **1**, 71–82
16. Lin, S.L., Tsai, C.J., and Nussinov, R. (1995) A study of four-helix bundles: investigating protein folding via similar architectural motifs in protein cores and in subunit interfaces. *J. Mol. Biol.* **248**, 151–161
17. Kohn, W.D., Mant, C.T., and Hodges, R.S. (1997) Alpha-helical protein assembly motifs. *J. Biol. Chem.* **272**, 2583–2586
18. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356
19. Alberti, S., Oehler, S., von Wilcken-Bergmann, B., and Muller-Hill, B. (1993) Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. *EMBO J.* **12**, 3227–3236
20. Brenowitz, M., Mandal, N., Pickar, A., Jamison, E., and Adhya, S. (1991) DNA-binding properties of a *lac* repressor mutant incapable of forming tetramers. *J. Biol. Chem.* **266**, 1281–1288
21. Chen, J. and Matthews, K.S. (1992) Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation. *J. Biol. Chem.* **267**, 13843–13850
22. Chen, J. and Matthews, K.S. (1994) Subunit dissociation affects DNA binding in a dimeric lac repressor produced by C-terminal deletion. *Biochemistry* **33**, 8728–8735
23. Lewis, M., Chang, G., Horton, N. C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254
24. Fairman, R., Chao, H.G., Mueller, L., Lavoie, T.B., Shen, L., Novotny, J., and Matsueda, G.R. (1995) Characterization of a new four-chain coiled-coil: influence of chain length on stability. *Protein Sci.* **4**, 1457–1469
25. Horton, R.M., Ho, S.N., Pullen, J.K., Hunt, H.D., Cai, Z., and Pease, L.R. (1993) Gene splicing by overlap extension. *Methods Enzymol.* **217**, 270–279
26. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423
27. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326
28. 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
29. Regan, L. and DeGrado, W.F. (1988) Characterization of a helical protein designed from first principles. *Science* **241**, 976–978
30. Handel, T.M., Williams, S.A., and DeGrado, W.F. (1993) Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **261**, 879–885
31. Erickson, H.P. (1989) Co-operativity in protein–protein association. The structure and stability of the actin filament. *J. Mol. Biol.* **206**, 465–474
32. Terwilliger, T.C. (1995) Engineering the stability and function of gene V protein. *Adv. Protein Chem.* **46**, 177–215
33. Liang, H., Sandberg, W.S., and Terwilliger, T.C. (1993) Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc. Natl Acad. Sci. USA* **90**, 7010–7014
34. Akanuma, S. and Yamagishi, A. (2008) Experimental evidence for the existence of a stable half-barrel subdomain in the $(\beta/\alpha)_8$-barrel fold. *J. Mol. Biol.* **382**, 458–466
35. Schreiber, G., Buckle, A.M., and Fersht, A.R. (1994) Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* **2**, 945–951
36. Beadle, B.M. and Shoichet, B.K. (2002) Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol.* **321**, 285–296
37. Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B., and Sanchez-Ruiz, J.M. (2004) Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. *J. Mol. Biol.* **336**, 313–318
38. Nagatani, R.A., Gonzalez, A., Shoichet, B.K., Brinen, L.S., and Babbitt, P.C. (2007) Stability for function trade-offs in the enolase superfamily "catalytic module". *Biochemistry* **46**, 6688–6695
39. Bloom, J.D., Wilke, C.O., Arnold, F.H., and Adami, C. (2004) Stability and the evolvability of function in a model protein. *Biophys. J.* **86**, 2758–2764
40. Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006) Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874
41. Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002
42. Benkovic, S.J., Fierke, C.A., and Naylor, A.M. (1988) Insights into enzyme function from studies on mutants of dihydrofolate reductase. *Science* **239**, 1105–1110
43. Shoichet, B.K., Baase, W.A., Kuroki, R., and Matthews, B.W. (1995) A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA* **92**, 452–456
44. Aili, D., Enander, K., Baltzer, L., and Liedberg, B. (2007) Synthetic de novo designed polypeptides for control of nanoparticle assembly and biosensing. *Biochem. Soc. Trans.* **35**, 532–534
45. Handel, T. and DeGrado, W.F. (1990) De novo design of a Zn^{2+}-binding protein. *J. Am. Chem. Soc.* **112**, 6710–6711
46. Handel, T.M., Williams, S.A., and DeGrado, W.F. (1993) Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **261**, 879–885
47. Robertson, D.E., Farid, R.S., Moser, C.C., Urbauer, J.L., Mulholland, S.E., Pidikiti, R., Lear, J.D., Wand, A.J., DeGrado, W.F., and Dutton, P.L. (1994) Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432
48. Choma, C.T., Lear, J.D., Nelson, M.J., Dutton, P.L., Robertson, D.E., and DeGrado, W.F. (1994) Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116**, 856–865
49. Rojas, N.R., Kamtekar, S., Simons, C.T., McLean, J.E., Vogel, K.M., Spiro, T.G., Farid, R.S., and Hecht, M.H. (1997) De novo heme proteins from designed combinatorial libraries. *Protein Sci.* **6**, 2512–2524
50. Cochran, F.V., Wu, S.P., Wang, W., Nanda, V., Saven, J.G., Therien, M.J., and DeGrado, W.F. (2005) Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J. Am. Chem. Soc.* **127**, 1346–1347

51. Lehmann, A. and Saven, J.G. (2008) Computational design of four-helix bundle proteins that bind nonbiological cofactors. *Biotechnol. Prog*. **24**, 74–79
52. Koder, R.L., Anderson, J.L., Solomon, L.A., Reddy, K.S., Moser, C.C., and Dutton, P.L. (2009) Design and engineering of an O_2 transport protein. *Nature* **458**, 305–309
53. Baltzer, L., Broo, K.S., Nilsson, H., and Nilsson, J. (1999) Designed four-helix bundle catalysts—the engineering of reactive sites for hydrolysis and transesterification reactions of p-nitrophenyl esters. *Bioorg. Med. Chem*. **7**, 83–91
54. Allert, M. and Baltzer, L. (2002) Setting the stage for new catalytic functions in designed proteins—exploring the imine pathway in the efficient decarboxylation of oxaloacetate by an Arg–Lys site in a four-helix bundle protein scaffold. *Chemistry* **8**, 2549–2560
55. Rossi, P., Tecilla, P., Baltzer, L., and Scrimin, P. (2004) De novo metallonucleases based on helix–loop–helix motifs. *Chemistry* **10**, 4163–4170
56. Razkin, J., Nilsson, H., and Baltzer, L. (2007) Catalysis of the cleavage of uridine 3'-2,2,2-trichloroethylphosphate by a designed helix–loop–helix motif peptide. *J. Am. Chem. Soc*. **129**, 14752–14758
57. Sterner, R. and Liebl, W. (2001) Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol*. **36**, 39–06
58. Cowan, D.A. (1997) Thermophilic proteins: stability and function in aqueous and organic solvents. *Comp. Biochem. Physiol. A. Physiol*. **118**, 429–438