Journal of Microbiology and Biotechnology
Vol. 7. No. 2, 132-137, 1997

Design and Expression of High Nutritional Peptide (HEAAE) in *E. coli*

KIM, JAE-HO*, CHANG-KOOK LEE[1], AND BUM-SHIK HONG

Graduate School of Biotechnology, Korea University, Seoul 136-701, Korea
[1]Utilization Research Lab, National Fisheries Research and Development Agency, Pusan 626-900, Korea

A novel protein (HEAAE, High Essential Amino Acid Encoding Protein), rich in essential amino acids (75% of total), was designed and constructed in our laboratory. The designed peptides were analyzed by SYBLE and stable secondary and tertiary structures were predicted. The monomeric form (HEAAE-1) of the protein consists of 20 amino acid residues with four additional amino acids comprising a potential β-turn (HEAAE-4). Size exclusion analysis demonstrated that the monomer is self-aggregates in aqueous solution to form higher ordered multimeric structures, which are very reminiscent of natural plant storage proteins. The DNA encoding this amino acid sequence was synthesized, and from this monomeric gene fragment (*heaae*-1), the stable tetrameric form of the gene (*heaae*-4) was generated by subcloning into the *E. coli* expression vector pKK223-3. A clear 6 kDa polypeptide band corresponding to the molecular weight of the dimeric form (HEAAE-2) was detected. The smeared band which appeared around the molecular weight corresponding to HEAAE-4 of 11 kDa suggested that the tetramer form of this protein might be processed into smaller size products.

Recently, a new field in protein research, *de novo* design of proteins, has made remarkable progress due to better understanding of the rules which govern protein folding and topology. Protein design has two components; the design of activity and the design of structure. The usual approach for the design of helical protein bundles consists of linking sequences that have propensity for forming an α-helix via short loop sequences (1, 5, 8, 9, 13, 16). These chains can be folded into predetermined 'globular type' tertiary structures in aqueous solution.

Short range interactions account for different amino acids having different conformational preferences. Both statistical and experimental methods (6) show that residues such as Glu, Ala, and Met tend to stabilize helices, whereas residues such as Gly and Pro are destabilizing. However, these intrinsic preferences are not sufficient to determine the stability of helices in globular proteins. Analysis of the free-energy requirements for helix initiation and propagation indicates that peptides of 10 to 20 residues should show little helix formation in water (2, 27). Nevertheless, the 13 amino acid C-peptide obtained from RNase A does show measurable helicity (~25%) at low temperature (2). The stability of this peptide is 1000-fold greater than the value calculated by the Zimm-Bragg equation. Specific side-chain interactions, factors that are not considered in the Zimm-Bragg model, are responsible at least in part for the fact that the C-peptide is much more helical than predicted (27, 28).

Medium-range interactions are responsible for the additional stabilization of secondary structures (8). Interaction between the side-chains are regarded as important medium range interactions (15, 28). These include electrostatic interactions, hydrogen bonding, and the perpendicular stacking of aromatic residues. An α-helix possesses a dipole moment as a result of the alignment of its peptide bonds. The positive and negative ends of the amide group dipole point toward the helix amino-terminus and carboxyl-terminus, respectively, giving rise to a significant macrodipole. Appropriately charged residues near the ends of the helix can favorably interact with the helical dipole and stabilize helix formation (23, 24, 27, 28).

Protein structures contain several long-range stabilizing interactions which include hydrophobic and packing interactions, and hydrogen bonds. Among these, the hydrophobic effect is a prime contributor to the folding and stabilizing of protein structures. The driving force for helix formation in RNase A arises from long-range interactions between C-peptide and S-protein, a large frag-

*Corresponding author
Phone: 82-415-60-1435. Fax: 82-415-865-0220.
E-mail: jhkim@tiger.korea.ac.kr.
Key words: high-essential amino acids encoding protein, peptide design, four helix bundle protein

ment of the protein from which C-peptide was excised. The role of hydrophobic interactions in determining secondary structures was studied for a series of peptides containing only Glu and Lys in their sequence (7).

Hydrophobic residues often repeat every three to four residues in an α-helix and form an amphiphilic structure (8, 11). Amphiphilicity is important for the stabilization of the secondary structures of peptides and proteins which bind in aqueous solution to extrinsic apolar surfaces, including phospholipid membranes, air, and the hydrophobic binding sites of regulatory proteins (7). This amphiphilic secondary structure can be stabilized relative to other conformations by self-association. Therefore, short peptides often form the α-helix in water only because the helix is amphiphilic and is stabilized by peptide aggregation along the hydrophobic surface. Native globular proteins are folded by a similar mechanism, involving hydrophobic interaction between neighboring segments of the secondary structure (22).

The free energy associated with dimerization or tetramerization of the designed peptides could be experimentally determined from the concentration dependence of the CD spectra for the peptides (7, 8). At low concentrations, the peptides were found to be monomeric and have low helical contents, whereas at high concentration they could self-associate and stabilize the secondary structure. Therefore, possible hairpin loops between helices can affect the stability of the secondary structure by enhancing the self-association between the helical monomers. A strong helix breaker (6) was included as the first and last residue to set the stage for adding a hairpin loop between the helices.

The next level of complexity in analysis of helical proteins is the factor determining the packing of helices within a protein. Structural stability of proteins is directly related to *in vivo* proteolysis (20, 22). Proteolysis depends on the accessibility of the scissile peptide bonds to the attacking protease. The sites of proteolytic processing are generally in relatively flexible interdomain segments or on the surface of the loops, in contrast to the less accessible interdomain peptide bonds (17).

Synthetic oligonucleotides have been employed in the construction of genes coding for entirely artificial, nutritionally-significant proteins to improve the nutritional quality of plant proteins (1, 3, 4, 12). There are two fundamental difficulties in achieving efficient expression of novel synthetic storage proteins. First, it is not yet known what stabilizes a protein against proteolytic breakdown. Second, the mechanisms for the folding of an amino acid sequence into a biologically-stable tertiary structure have not yet been fully delineated. Thus, for the design of HEAAE, we focused on constructing a physiologically-stable as well as highly nutritious, storage-protein-like, artificial protein.

MATERIALS AND METHODS

Bacterial Strain and Vector

E. coli strains DH5α and CSR603 were used as host cells for the transformation of *heaae*-4 and other DNA constructs for the *in vivo* expression of HEAAE-4.

Synthesis and Analysis of Peptides

HEAAE monomer (HEAAE-1), 24 amino acids in length, was synthesized using an solid phase peptide synthesizer from Milligen/Biosearch (Model 9050 Pep-Synthesizer). After TFA cleavage and extraction, the crude peptide was washed twice with ether and dried under vacuum. Organic contaminants were removed by gel filtration through a Sephadex G 25-100 column in 25% acetic acid. Fractions were collected and lyophilized. Analysis of the crude peptide was done with a Waters Delta Prep HPLC, using Bondapak C-18 analytical and semi-preparative columns. The fragments from chromatography were quantitatively separated and lyophilized again. The molecular weight of the purified peptide were analyzed using a 252Cf Plasma Desorption Mass Spectrometer (PDMS) BIO ION 20 (Bio Ion Nordic AB, Uppsala, Sweden). The samples were applied to a nitrocellulose-coated target in a 50 : 50 water : ethanol solution and allowed to absorb for 10 minutes before loading into the instrument.

Size exclusion chromatography of the peptide was carried out by using a 1.6 × 90 cm column of Sephadex G50F (Pharmacia). The eluent was monitored by measuring absorbance at 280 nm. The protein standards and peptides (0.1 to 0.5 mg, Sigma) were applied to the column in 0.5 ml of 0.05 M MOPS buffer, pH 7.0, 5% glycerol, 0.1% sodium azide and eluted with the same buffer without glycerol at a flow rate of 0.3 ml/min.

Synthesis of Oligonucleotide and Gene Construction

Both strands of oligonucleotides for the construction of heaae were synthesized using an Applied Biosystems, Inc. (ABI) Model 380A DNA Synthesizer. Synthesized *heaae*-4 gene in pUC19 was sequenced by the procedure described by the manufacturer (USB) were followed.

In vivo Expression

A modified Sancar's method (25) was used for *in vivo* expression of HEAAE. The *heaae*-4 construct was tranformed and grown in K medium+uridine (11 g $Na_2HPO_4 \cdot 7H_2O$, 3 g KH_2PO_4, 5 g NaCl, 10 mg thiamine, 0.2 mM $MgSO_4$, 0.04 mM $CaCl_2$ per liter and 1% glucose, 1% casamino acid and 0.1% uridine pH 7.0) overnight at 37°C. The cells were irradiated with UV (450 ergs/cm^2/sec) in a petri-dish for 2 minutes with stirring and transferred to 20 ml K medium+uridine. Cells were harvested and washed twice with 2 × sulfate-free Hershey medium (NaCl 5.4 g, KCl 3.0 g, $CaCl_2$ $2H_2O$ 15 mg, $MgCl_26H_2O$ 203 mg, $FeCl_3 \cdot 6H_2O$ 0.2 mg, KH_2PO_4 87 mg and Tris 12.1 g per liter, 2 ml of 20% glucose, 0.5 ml of 2%

threonine, 1 ml of 1% leucine, 1 ml of 2% proline, 1 ml of 2% arginine, and 1 ml of 0.1% thiamine). The culture was labeled with [^{35}S]methionine (5 mCi/ml) and harvested. Sample was loaded on a 15% polyacrylamide gel and bands were detected by exposing on X-ray film.

RESULTS AND DISCUSSION

Design of Artificial Storage Protein (HEAAE)

HEAAE (High Essential Amino Acid Encoding gene) contains high percentages of the five most limiting essential amino acids (MLEAA) in plants, which are isoleucine, lysine, methionine, threonine and tryptophan. It has 1.8 times more of the essential amino acids compared to zein or phaseolin. The difference in MLEAA levels is much higher, containing 3 times more than phaseolin and 6.5 times more than zein. In addition to the nutritional quality, HEAAE has been designed to have a stable storage protein-like structure in plants.

The stability of the folded structure of a protein has a close relation to the *in vivo* proteolytic degradation rate (18-21). It was found that the sites of proteolytic processing are generally in relatively flexible interdomain segments or on the surface of the loops, in contrast to the less accessible interdomain peptide bonds (17). Therefore, stable folded native proteins are relatively resistant to cleavage by proteolytic enzymes, whereas denatured proteins are much more sensitive (18).

HEAAE is comprised of 4 helical repeating units, each 20 amino acids long (Fig. 1a, b). Three potential β-turn sequences (Gly-Pro-Gly-Arg) were inserted between four monomers (HEAAE-1) for the HEAAE-4 construction. The helical region of HEAAE-1 has been designed amphiphilically and stabilized by several Glu-Lys salt bridges (Fig. 1c). Glu and Lys residues were chosen as charged residues for the solvent-accessible exterior of the protein to help stabilize helix formation by electrostatic interaction (15, 23, 24, 28).

Hydrophobic residues repeat every three to four residues in an α-helix region of the HEAAE and form an amphiphilic structure. Amphiphilicity is important for the stabilization of the secondary structures of peptides and proteins which bind in aqueous solution. Therefore, short peptides often form the α-helix in water only because the helix is amphiphilic and is stabilized by peptide aggregation along the hydrophobic surface (22). Therefore, the hydrophobic effect is a prime contributor to the folding and stabilizing of HEAAE structures.

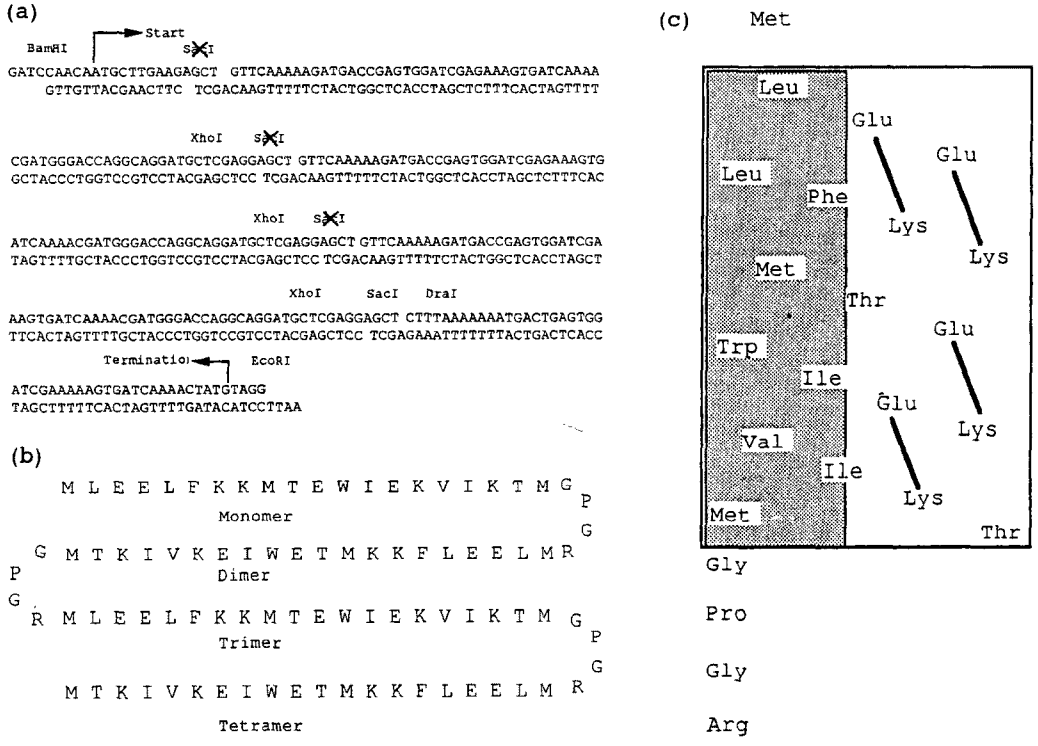

Fig. 1. (a) Nucleic and peptide sequence of HEAAE-4. The 276 bp long heaae-4 produces the 11.5 kDa HEAAE-4 protein. This protein consists of four helical and three turn regions. (b) Structural model of HEAAE-1. HEAAE-1 has an amphiphilic structure and possible salt bridge formed between glutamic acid and lysine residues. In the figure, shadowed and white region indicates hydrophobic and hydrophilic residues, respectively (c). The possible salt bridges between Lys and Glu are indicated by lines.

The β-turn could play an important role for the structural stability of the HEAAE-4 when it is expressed *in vivo*. It can also help stabilize tertiary structure formation. The proximate effect might be critical for folding at the low concentrations of HEAAE-4 when it is expressed *in vivo*. This increased gene copy number through the β-turn can increase the protein yields and the molecular mass of the encoded protein. Such an increase in size can significantly stabilize an otherwise unstable product (8, 26). In addition, this β-turn sequence has a tryptic digestion site (Gly-Arg) which could increase the digestibility of this protein when it is consumed by animals.

The secondary structures of the HEAAE-1 and -4 were predicted by PREDICT-SECONDARY in SYBYL. The percentage of α helix content predicted by information-theory showed 100% helical content for the monomer and 74% for the tetramer. A perfect amphiphilic α-helical conformation was predicted for the HEAAE-1 after minimization (Fig. 2a). The tertiary structure of the HEAAE-4 after minimization showed the antiparallel conformation as was designed (Fig. 2b). These minimization results suggested the high probability of stable secondary and tertiary structure formation of the HEAAE-1 and -4.

Fig. 2. (a) HEAAE-1: a computer predicted and then energy minimized structure. The structure has been predicted by information theory in SYBLE and then minimized using the Kollman force field. The blue color indicates hydrophobic amino acid residues while the red color indicates those which are hydrophobic. (b) HEAAE-4: a computer predicted and energy minimized structure. The structure was predicted by information theory.

Structural Analysis of HEAAE

The HPLC purified fraction of synthesized peptides was analyzed by mass spectrometry, and the molecular weight peak corresponding to the HEAAE-1 (2896.5 dalton) was used for structural analysis (Fig. 3). The α-helical secondary structure of the HEAAE-1 and its tertiary structure were analyzed by size exclusion analysis. The self-association capability of the HEAAE-1 was investigated by using size exclusion chromatography. The hydrodynamic behavior of this peptide showed that it was aggregated into a hexamer form with an apparent molecular weight of about 18 kDa (Fig. 4). This hexameric aggregate could be maintained in either the aqueous or high ionic concentration (0.5 M NaCl).

The stability of the α-helical secondary structure can be induced by the intermolecular interaction between the helical chains (8). Therefore, strong interactions among the monomeric HEAAE molecules could stabilize the secondary structure as well the tertiary structure of the HEAAE multimer. The solubility of proteins could also affect their proteolytic resistance. Some proteins aggregate to form inclusion bodies that escape proteolytic attack (14). Therefore, stable aggregation between monomers, through possible hydrophobic interactions, could protect against proteolytic attack as well as stabilize the secondary and tertiary structure of the HEAAE multimer. This result suggest the stable globular type tertiary structure formation of

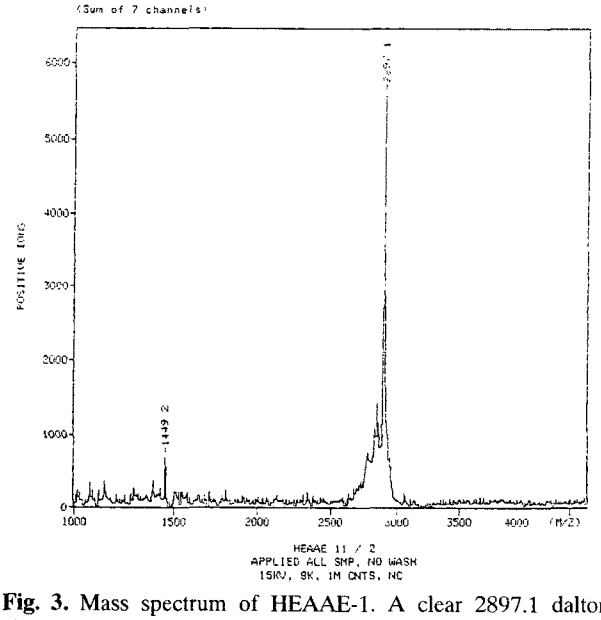

Fig. 3. Mass spectrum of HEAAE-1. A clear 2897.1 dalton signal was detected which is preciesly the molecular weight expected.

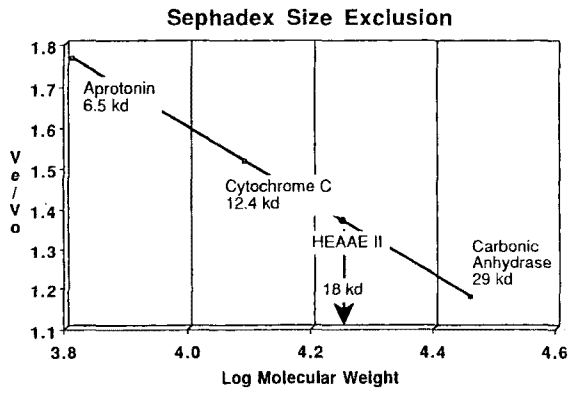

Fig. 4. Calibration curve for 1.6×90 cm G50F Sephadex column used in the size exclusion chromatography of aggregation of HEAAE-1.

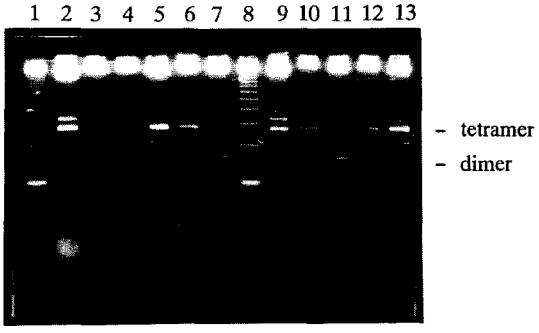

Fig. 5. Restriction enzyme digestion (*Eco*RI) of heaae.
HEAAE mutimeric genes bigger than tetramer were not stable and were processed into a smaller size in *E. coli* (lanes 2, 3, 4, 9, 11, and 12). The results indicated that *heaae*-4 and *heaae*-2 genes are the most stable construct (lanes 5, 6, 7, 10 and 13).

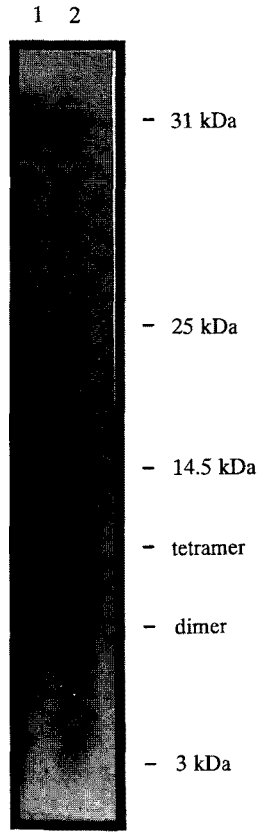

Fig. 6. Expression of *heaae*-4 cloned in vector pKK223-3 in *E. coli*.
The plasmid expressed protein were labeled with ^{35}S-Met. A distinct HEAAE-2 (5.4 kDa) and diffused HEAAE-4 (11.5 kDa) bands were seen in *E. coli* CSR 603 maxicell expression system. Lane 1, *E. coli* CSR 603 transformed with vector only. Lane 2, *E. coli* CSR 603 transformed with pKK *heaae*-4.

tetrameric HEAAE.

Gene Construction and *E. coli* Expression

The stability of the HEAAE mutimeric gene (*heaae*-4) and protein (HEAAE-4) were investigated by transforming maxi-cells with multimeric HEAAE gene construct. It was observed that HEAAE mutimeric genes bigger than tetramer were not stable and were processed into a smaller size in *E. coli* (Fig. 5. lanes 2, 3, 4, 9, 11, and 12). These results show that tetrameric and dimeric direct repeated forms are the most stable construct (Fig. 5. lanes 5, 6, 7, 10 and 13). With directly repeated genes, exact excision of one or more copies was occasionally observed by homologous recombination after transformation of *rec*$^+$, *recA*, and *recBC* host cells (10).

A clear 6 kDa polypeptide band corresponding to the molecular weight of the dimeric form (Fig. 6) was detected. The level of expression was not high in *E. coli* because the codon usuage of the HEAAE gene were selected for optimal expression in the plant system. The smeared band which appeared around the molecular weight corresponding to tetrameric 11 kDa suggested that the tetramer form of this protein might be processed into smaller size products. The dimeric form might be stable because it could associate into the most stable hexmaric form. The second β-turn of tetrameric-form of HEAAE might be exposed to proteolytic attack because of its unstable association. A similar observation was reported from an experiment with designed hexameric repeated peptides fused with β-galactosidase (3). The repeated cassette region of the synthetic peptide has been processed to monomeric form (4). This result suggests that HEAAE-2 might be the most stable product, at least in an *E. coli* system.

REFERENCES

1. Beauregard, M., C. Dupont, R. M. Teather, and M. A. Hefford. 1995. Design, expression and identical characterization of MB 1, a *De novo* protein enriched in essential

amino acids. *BioTechnol.* **13**: 974-981.
2. Bierzynski, A., P. S. Kim, and R. L. Baldwin. 1982. A salt bridge stabilizes the helix formed by isolated C-peptide of RNAseA. *Proc. Natl. Acad. Sci. USA.* **79**: 2470-2478.
3. Biernat, J., H. Hasselmann, B. Hofer, N. Kennedy, and H. Koster. 1987. The construction and cloning of synthtic genes coding for artificial proteins and expression studies to obtain fusion proteins. *Protein Eng.* **1**: 345-352.
4. Biernat, J. and H. Koster. 1987. Expression of synthtic genes coding for completely new, nutrtionally rich, artificial proteins. *Protein Eng.* **1**: 353-359.
5. Clifford, R. R. and G. S. Stephan. 1993. Electrostatic stabilization in four-helix bundle proteins. *Protein Sci.* **2**: 826-837.
6. Chou, P. Y. and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**: 45-148.
7. Degrado, W. F. and J. D. Lear. 1985. Induction of peptide conformation at apolar/water interface a study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **107**: 7684-7689.
8. Degrado, W. F., Z. R. Wasserman, and J. D. Lear. 1989. Protein design, a minimalist approach. *Science* **241**: 622-628.
9. Ho, S. P. and W. F. Degrado. 1987. Design of a 4-Helix bundle protein; synthesis of peptides which self-associate into helical protein. *J. Am. Chem. Soc.* **109**: 6751-6758.
10. Hofer, B. 1987. Construction and stability of a sixfold repeated artificial gene. *Eur. J. Biochem.* **167**: 307-313.
11. Iwata, T., S. Lee, O. Oshi, H. Aoyagi, M. Ohno, K. Anzai, Y. Kirino, and G. Sugihara. 1994. Design and synthesis of amphipathic helical peptide and their interaction with phospholipid bilayer and ion channel formation. *J. Biol. Chem.* **269**: 4928-4933.
12. Jaynes, J. M., M. S. Yang, N. Espinoza, and J. H. Dodds. 1986. Plant protein improvement by genetic engineering: use of synthetic genes. *Tibtech.* **4**: 314-320.
13. Kamtekar, S., J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht. 1993. Protein design by binary patterening of polar and nonpolar amino acids. *Science* **262**: 1680-1685.
14. Kane, J. F. and D. L. Hartley. 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli. Tibtech.* **6**: 95-101.
15. Marqusee, S. and R. Baldwin. 1987. Helix stabilization by Glu-Lys salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. USA.* **84**: 8898-8902.
16. Mutter, M. 1988. Nature's rules and chemist's tool: a way for creating novel peptide. *TIBS.* **13**: 260-265.
17. Neurath, H. 1989. Proteolytic processing and physiological regulation. *TIBS.* **14**: 268-271.
18. Pace, C. N. and A. J. Barret. 1984. Kinetics of tryptic hydrolysis of the arginine-valin bond in folded and unfolded ribonuclease T1. *Biochem. J.* **219**: 411-417.
19. Pakula, A. A. and R. T. Sauer. 1986. Bacteriophage 1 Cro mutation: effect on activity and intracellular degradation. *Proc. Natl. Acad. Sci. USA.* **82**: 8829-8833.
20. Pakula, A. A. and R. T. Sauer. 1989. Amino acid substitution that increase the thermal stability of the Cro protein. *Proteins* **5**: 202-210.
21. Parasell, D. A. and R. T. Sauer. 1989. The structural stability of protein is an important determinant of its proteolytic susceptability in *Escherichia coli. J. Biol. Chem.* **264**: 7590-7594.
22. Presnell, S. R. and F. E. Cohen. 1989. Topological distribution of a four α-helix bundle. *Proc. Natl. Acad. Sci. USA.* **86**: 6592-6596.
23. Presta, L. G. and G. D. Rose. 1988. Helix signals in proteins. *Science* **240**: 1632-1641.
24. Richardson, J. S. and D. C. Richardson. 1988. Amino acid preferances for specific location at the end of α-helices. *Science* **240**: 1648-1652.
25. Sancar, A., R. P. Wharton, S. Seltzer, B. M. Kacinski, N. D. Clarke, and W. P. Rupp. 1981. Identification of the *urvA* gene product. *J. Mol. Biol.* **148**: 45-62.
26. Shen, S-H. 1984. Multiple joined genes prevent product degradation in *E. coli. Proc. Natl Acad. Sci. USA.* **81**: 4627-4631.
27. Shoemaker, K. R., P. S. Kim, D. N. Bream, S. Marqusee, E. J. York, I. M. Chaiken, and R. Baldwin. 1985. Nature of the charged-group effect on the stability of the C-peptide helix. *Proc. Natl. Acad. Sci. USA.* **82**: 2349-2353.
28. Shoemaker, K. R., P. S. Kim, E. York, J. M. Stewart, and R. L. Baldwin. 1987. Tests of the helix dipole model for stabilization of a α-helices. *Nature* **326**: 563-567.

(Received December 13, 1996)