Optimization of Site-Directed Mutagenesis. 2. Application of Random-Centroid Optimization to One-Site Mutation of *Bacillus stearothermophilus* Neutral Protease To Improve Thermostability

Shuryo Nakai,* Soichiro Nakamura, and Christine H. Scaman

Department of Food Science, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

The random-centroid optimization program was modified for application to site-directed mutagenesis of proteins and renamed random-centroid optimization for genetics (RCG). RCG was used to select both the site in the protein sequence for mutation and the amino acid for substitution. The substituting amino acid was selected on the basis of a hydrophobicity scale. Other indices of amino acids, namely propensities of α-helix, β-strand, and β-turn as well as bulkiness, were also included in the RCG program. For validating the RCG approach, single-site mutations were conducted in the 16 amino acid residues of the active-site helix (G139−Y154) of *Bacillus stearothermophilus* neutral protease. After 13 mutations, it was found that the substitution V143E substantially improved the thermostability of the recombinant enzyme. The half-survival temperature of mutant V143E was higher by 6.5 °C than the 68.3 °C of the wild-type enzyme. Lower hydrophobicity and bulkiness at the N-terminal side of the active region of enzyme were important for maintaining high thermostability as determined from mapping, based on the different amino acid scales.

Keywords: *One-site-directed mutagenesis; B. stearothermophilus; neutral protease; thermostability; random-centroid optimization*

INTRODUCTION

Directed molecular evolution in a test tube is a mimic of natural evolution (Ellington and Matsumura, 1997). Initially, a weakly active protein is selected from a random-sequence population and then optimized by mutation and reselection. This cycle is repeated, thus mimicking natural evolution and providing a model for investigating the origin of life. Random mutagenesis of a protein can be useful when few comparative data or little structural information is available for a specific protein. It is also useful when we wish to investigate the effects of all possible amino acid substitutions at a predetermined position within a protein molecule. Of three methods of random mutagenesis, that is, chemical mutagenesis, PCR mutagenesis, and oligonucleotide synthesis (ONS), a truly random library can be generated by ONS, although this method is costly (Hedstrom et al., 1991). Breaker and Joyce (1994) suggested that generating ribozymes with novel catalytic function could be best carried out using in vitro random selection by mimicking Darwinian evolution of organisms in nature. A more systematic approach to the design cycle was carried out by the recursive ensemble mutagenesis (REM) of Delagrave et al. (1993) that involves the recursive use of combinatorial cassette mutagenesis (CCM). It is an iterative strategy, which seeks to continually improve the CCM library. Fuellen and Youvan (1994) applied the genetic algorithm to REM. Delagrave et al. (1993) stated that the generation of new and informative mutant proteins was necessary to elucidate the structure−function relationships of proteins.

Blundell (1994) has discussed rational design in protein engineering. A great deal of effort has been directed to modeling mutant proteins from their sequences. However, it is not possible to accurately predict molecular structure, or rationally design proteins, using computer-aided molecular modeling. Reliable three-dimensional structural details of the proteins can be obtained only by using X-ray crystallography or NMR. Furthermore, the fact that no reliable relationships between molecular structure and functions of mutant proteins are currently available makes the situation even more difficult. Due to a lack of appropriate working equations, that is, structure−function relationships in this case, computational optimization is hard to apply to protein engineering.

In the preceding paper, we reported that random-centroid optimization (RCO) suitable for application to experimental optimization could efficiently find the global optimum of a multimodal function (Nakai et al., 1998). As a logical consequence, we have modified the RCO program so that it can be applied to site-directed mutagenesis of proteins. First, it was essential to have reliable evidence showing that RCO could be applied to global optimizations, because biological phenomena are frequently nonlinear. In general, when multiple objective functions are optimized, the combined functions become multimodal because the best conditions for each function will vary. As well, an efficient optimization strategy is especially important in the case of mutagenesis, because of the complexity and the time as well as the expense involved in genetic research projects. To

* Address correspondence to this author at the Department of Food Science, University of British Columbia, 6650 N.W. Marine Dr., Vancouver, BC, Canada V6T 1Z4 [telephone (604) 822-4427; fax (604) 822-3959; e-mail nakai@unixg.ubc.ca].

S0021-8561(97)00610-9 CCC: $15.00 © 1998 American Chemical Society
Published on Web 03/04/1998

ensure that the RCO program was applicable to such a global optimization, the computer programs of multimodal functions were randomized (Nakai et al., 1998). This process avoided subjectivity in selecting search spaces on maps due to a priori knowledge of the location of optimum. Application of optimization algorithms to randomized model functions, therefore, improved the objectivity and reliability of the optimization efficiency assessment.

Unlike REM, our optimization program was designed to optimize the selection of the location of the mutation and the amino acid for replacement, simultaneously. Therefore, two factors are required to alter for each mutation experiment. Sander (1994) has suggested that it would be beneficial to use knowledge of a protein structure to select a few key residue positions, in which mutations may have a functionally beneficial effect. The residues lining the active site of an enzyme provide such an example. These positions can be randomized in a vector carrying the original sequence, using suitably synthesized oligonuclotides. It was suggested that the number of residues chosen for randomization be kept reasonably small to achieve good coverage of possible sequence combinations.

The objective of this study was to modify the RCO program so that it could be applied to site-directed mutagenesis. For verifying our approach, the modified RCO, renamed random-centroid optimization for genetics (RCG), was applied to a simple one-site-directed mutagenesis of a short peptide chain composed of 16 amino acid residues in the active-site helix of the *Bacillus stearothermophilus* neutral protease. This strategy seems to be appropriate according to the recommendation made by Sander (1994) as discussed above.

EXPERIMENTAL METHODS

Modifications to the RCO Program. Factors with odd factor numbers were used for entering the site numbers in the sequence to be mutated into the RCG program. For example, factors 1, 1 and 3, and 1, 3, and 5 are assigned for the sites to be substituted in simultaneous one-, two-, and three-site mutations, respectively. The site numbers at both ends of the region to be optimized in the sequence define the minimum and maximum of the search space. Factors with even factor numbers were assigned to the amino acids to be used for substitution. The range of indices on the amino acid scale to be used for selecting the substituting amino acids is entered into the RCG program. Therefore, factors 2, 2 and 4, and 2, 4, and 6 are assigned for one-, two-, and three-site mutations, respectively. Any scale of an amino acid index, stored in the RCG program, can be used for selecting the amino acid to substitute in optimization computation. It is feasible to use different scales for guiding the mutation at different sites during an optimization run. These are the only changes to the RCO program required to allow its use in optimizing the site-directed mutagenesis of a protein. After these changes have been incorporated, the use of the program is similar to that described in the preceding paper (Nakai et al., 1998).

In this study, the initial selection of an amino acid substitution was done using a hydrophobic scale. The use of such a scale in cycle 1 is preferable. It should be noted, however, that it is possible to change scales in the following cycles of optimization, if it is found during mapping that another mechanism appears to play an important role in the specific function of interest. It is also possible to use an alternative amino acid scale, such as helix propensity in cycle 1.

An empirical hydrophobicity scale of amino acids derived from elution data of peptides from reversed-phase (RP-) HPLC was employed, rather than a theoretically derived scale. Indices used were obtained from an extensive study conducted by Wilce et al. (1995). The hydrophobic indices used in the RCG program were −2.24 (His), −1.62 (Lys), −0.85 (Arg), −0.62 (Ser), −0.2 (Asp), −0.1 (Glu), 0.06 (Ala), 0.15 (Gly), 0.21 (Met), 0.25 (Asn), 0.31 (Gln), 0.49 (Cys), 0.65 (Thr), 0.71 (Pro), 1.59 (Val), 1.89 (Tyr), 2.29 (Trp), 3.0 (Ile), 3.5 (Leu), and 4.8 (Phe). For helix and strand propensities, the scales of Muñoz and Serrano (1994) were selected because of the extensive survey done in their work. For bulkiness, the scale for protein compressibility reported by Gromiha and Ponnuswamy (1993) was adopted. A β-turn scale was not included but left blank as no adequately reliable scale has been published. Regarding β-turn propensity, there are only a few papers published, including the paper by Wilmot and Thornton (1988).

Hypothetical Model Programs for Site-Directed Mutagenesis. The active-site helix of *B. stearothermophilus* neutral protease consists of 16 amino acid residues (GIDVVGHELTHAVTDY) including the HEXXH motif, in which the glutamic acid acts as a catalytic base (Jiang and Bond, 1992). A model of the active-site helix was constructed to test the reliability of the RCG approach. It was hypothesized, for computation trials, that the center portion of the peptide should be strongly hydrophobic relative to the terminal sections for a required function (model 1). The hypothetical response value, that is, hydrophobic ratio (HpT), to be optimized was defined as HpT = Hp2/(Hp1 + Hp3), where Hp1, Hp2, and Hp3 are the total hydrophobicities of sites 1−5, 6−11, and 12−16. To facilitate map reading, the response values were adjusted to be within a range of −1 to 10. Two functions, hydrophobicity and helix propensity, regulated model 2. To search for the optimum, cycles 1 and 2 were optimized using HpT and helix ratio (HxT), respectively. For computing a hypothetical response value for the helix model in cycle 2, HxT = Hx2/(Hx1 + Hx3) was used, where Hx1, Hx2, and Hx3 were total helix propensities of the sites defined in model 1. Response values were adjusted to the same range as model 1 to facilitate map reading.

Site-Directed Mutagenesis of *B. stearothermophilus* Neutral Protease. The active-site helix of the neutral protease was mutated by single-site substitutions to maximize the thermostability of the enzyme. The methods used for mutation and thermostability evaluation were those reported in our previous paper (Nakamura et al., 1997).

RESULTS AND DISCUSSION

Modification of the RCO Program for Site-Directed Mutagenesis. It is generally agreed that hydrophobic effects are the major stabilizing factor in the folded structure of globular proteins (Matthews, 1993). Therefore, it is reasonable to use a hydrophobicity index of amino acids for selecting amino acid substitutions in cycle 1.

Several amino acid hydrophobicity scales were examined in this study. The scale of Jones (1975) was tried first. Extension of the original scale of Nozaki and Tanford (1971) that was based on solubility was used to derive Jones's indices. Since solvation effects are the basis of polarity as well as hydrophobicity, this selection may be appropriate. However, the hydrophobicity scale used in this study was later changed to that of Wilson et al. (1981), which computed hydrophobicity indices of amino acids derived from the elution volume of 96 peptides (2−65 amino acids in length) subjected to RP column chromatography. As proteins cannot be applied to RP columns because of possible surface denaturation, the scale computed from chromatographic data of peptides may be the next best choice. Model computation using Wilson's scale yielded data in good agreement with the hypothesis preset for model formulation as

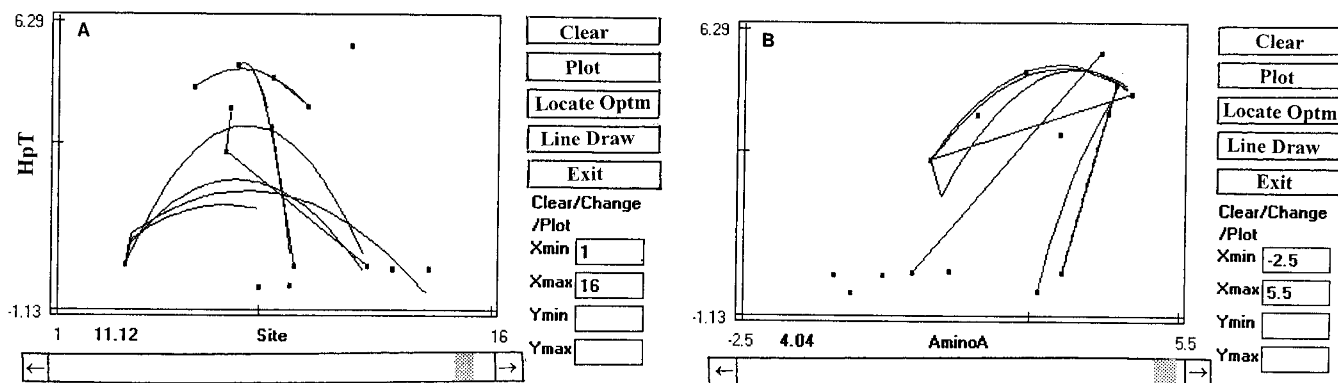

Figure 1. Maps generated from a model random-centroid optimization for one-site mutation of neutral protease (model 1, hypothetical hydrophobicity−hydrophobicity model): (A) site 1−16 of active-site helix; (B) hydrophobicity scale, index from −2.5 to 5.5. Each datapoint represents a one-site mutation for maximizing HpT. Cycle 1 included six mutations for the random search and two mutations for the centroid search. Using the narrowed search spaces determined from the maps in cycle 1, a random search was carried out, consisting of five mutations followed by centroid search with two mutations. Values 11.12 and 4.04 on the *X*-axes in (A) and (B), respectively, are the locations of the highest HpT.

described above, whereas the scale of Jones (1975) gave occasionally deviations from the set hypothesis.

Recently, an extensive study on the elution behavior of peptides on RP columns was made by Wilce et al. (1995). They used 1738 peptides ranging in size from 2 to 50 amino acid residues. The correlation coefficient (r) between scale 1 of Wilce et al. (1995) and the scale of Wilson et al. (1981) is 0.81, which is comparable with $r = 0.84$ between scale 1 and the scale reported by Meek and Rosetti (1981). It may be worth noting that all three scales have been computed on the basis of the elution volume from C_{18} HPLC columns. Meanwhile, the correlation coefficient values between Wilce's scale 1 and the scales based on solubility data of amino acids were only 0.68 and 0.66 for those of Zimmerman et al. (1968) and Jones (1975), respectively. It was ultimately decided to use the scale of Wilce et al. (1995) for genetic optimization because it was based on the most recent and thorough investigation reported in the literature.

The scales chosen for the α-helix and β-strand propensity were those of Muñoz and Serrano (1994) because these were the most detailed survey available. These propensity scales were calculated from $\phi-\psi$ matrices derived from a database of 279 3-D structures. Good correlations were obtained with other experimental scales reported in the literature according to Muñoz and Serrano (1994).

For bulkiness, the scale in the paper on protein compressibility by Gromiha and Ponnuswamy (1993) was employed for our study. However, no scale for turn propensity was included at this time. Two scales, that of Chou and Fasman (1978) for turns or loops and that of Wilmot and Thornton (1988) for two turn types based on 59 nonidentical proteins, have been reported. However, a cluster analysis conducted by Nakai et al. (1988) could not reliably classify helix and turn propensities into separate groups. Therefore, it was felt that it may be premature to use a turn scale in the optimization program. This subprogram was left blank but may be completed in the future when an appropriate scale becomes available.

Model Computation. Figure 1 shows the maps of 15 mutations in two cycles of optimization using the hydrophobicity scale in both cycles of model 1. $X_{\min}$ and $X_{\max}$ for maps were set as 1 and 16 and −2.5 and 5.5 for parts A and B of Figure 1, respectively. These values were the upper limits and lower limits of the region to

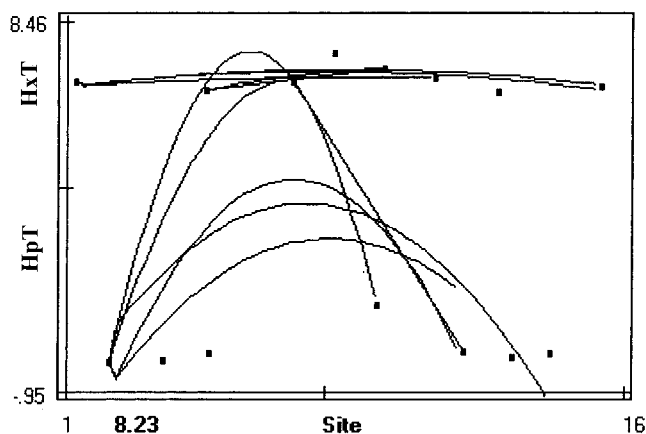

Figure 2. Map as a function of sites, generated from a model for one-site mutation of the neutral protease (model 2, hypothetical hydrophobicity−helix model). The optimization process carried out was similar to that described in Figure 1. In this case HpT was maximized in cycle 1, and HxT was maximized in cycle 2 by mutating the best sequence found in cycle 1. Value 8.23 on the *X*-axis is the location of the highest HxT.

be optimized in the sequence and the range of hydrophobic scale to be used in selecting amino acids to substitute, which were entered into the computer to start the random search in cycle 1. The lines on the map of Figure 1A, drawn by linking datapoints according to the rule described in the preceding paper (Nakai et al., 1998), show a distinct curvature in the response surface, indicating a trend toward the optimum in the middle of the map. This matches the working hypothesis for this model that sites around 9, ranging from site 6 to site 12 (Figure 1A), should be highly hydrophobic as shown in Figure 1B.

For model 2 in Figure 2, in which a combination of hydrophobicity and helix propensity was used as the basis of optimization, the following result was obtained: After eight mutations in cycle 1 based on hydrophobicity, experiment 8 that replaced site 7 with phenylalanine yielded the best result (Figure 2). Further mutations were carried out in cycle 2 with mutant 8 for optimization of helix propensity. The best result was obtained in experiment 10, by substituting site 8 with glycine. However, compared to cycle 1 based on hydrophobicity alone, cycle 2 experiments resulted in only a slight improvement in responses at the center

Table 1. Random-Centroid Optimization of Mutation of Neutral Protease

optimization steps		site	mutation	ΔT_{50}[a] (°C)	proteolytic activity[b] (%)
cycle 1	random	13	V151D	5.7	45.2
		12	A150W	4.4	27.2
		7	H145G	1.7	−87.7
		3	D141P	5.0	76.3
		6	G144F	−0.7	82.9
		10	T148I	2.1	3.6
	centroid	8	E146N	3.8	−96.3
		13	V151P	4.2	−12.8
cycle 2	random	11	H149W	3.5	−91.5
		7	H145K	6.0	−80.7
		5	V143E	6.5	32.1
		9	L147K	1.9	−1.2
		12	A150E	4.4	21.6

[a] ΔT_{50}, difference in T_{50} from that of wild-type enzyme. [b] Proteolytic activity, difference from that of wild-type enzyme (100%) as measured by the method of Nakamura et al. (1997).

(upper level trend lines in Figure 2). This is not unexpected, because the native 16 amino acid region was initially highly helical in nature. The map of cycle 2 based on the helix scale of amino acids showed a trend in agreement with the preset hypothesis that sites 6−11 are supposed to be strongly helix. At any rate, a reasonable optimization was achieved in seven mutations in cycle 2.

Considering the total possible number of experiments required to be carried out, if all 16 amino acids in the peptide were replaced with 19 residues, completion of the optimization of this model within 15 mutations can be considered extremely efficient. Furthermore, the RCG has an additional advantage as mapping can reveal which property of the amino acids is playing an important role in the function of interest by using the different amino acid scales as independent variables to plot maps. Mapping is an approximation of the response surface useful in searching for the optimum as discussed in the preceding paper (Nakai et al., 1998). If datapoints on the maps scatter widely without demonstrating a clear trend, the property of amino acids used in mapping computation is irrelevant to the function and this factor should be eliminated as soon as possible from the optimization.

One-Site-Directed Mutagenesis of Neutral Protease. The results of this RCG optimization are shown in Table 1 and Figure 3. The technique of intensified line drawing, which eliminates factors on purpose as described in the preceding paper (Nakai et al., 1998), could not be applied to this mutagenesis as it required more than two factors. This optimization of a single-site mutation uses only two factors, and thus, this powerful method could not be applied.

Table 1 shows the result of RCG optimization with six random mutations and two centroid mutations in cycle 1 followed by five random mutations in cycle 2. The third mutation in the random search for cycle 2 yielded the largest ΔT_{50} with an associated increase in proteolytic activity of 32.1%. Additional information obtained from other sources, if any, can be added to the RCG optimization to improve the accuracy of mapping process. In this case, data obtained from an independent study (Nakamura et al., 1997), in which proline was introduced to four locations in the active-site helix (without using any optimization algorithms), were included in the maps shown in Figure 3. Addition of these extra data enhances the optimization efficiency, as well as the reliability of interpreting the maps to explain the mechanism of protein function. Mutant V143E (site 5 in the 16 amino acid peptide chain) gave an increase of 6.5 °C for the T_{50} [half-survival temperature defined by Nakamura et al. (1997)] of 68.3 °C for the wild-type enzyme. However, ΔT_{50} of 7.5 °C was obtained for the mutant I140P (site 2 in 16 amino acid residues) for the same active-site helix (Nakamura et al., 1997). It is expected that continuation of the RCG optimization might exceed this value obtained by proline introduction, but we decided to discontinue further mutation experiments as it was felt that the objectives of the project had been achieved.

Figure 3A indicates that mutations near the N-terminal end are slightly more efficient than those at the C-terminal end of the active region for enhancing thermostability. Figure 3B shows the effect of bulkiness. The left-most datapoint for glycine appears distinct from the trend of the other amino acid residues. Within the bulkiness scale, excluding glycine, there is a direct correlation between a decrease in bulkiness and an increase in thermostability. This is demonstrated by four left-slanting trend curves with negative slopes, compared to only one with a positive slope. This trend is in agreement with the conclusion of the study involving proline introduction (Nakamura et al., 1997). In this previous work, it was found that high rigidity in the N-terminal end favors greater thermostability. This conclusion was derived from α-chymotrypsin susceptibility trials and a study of the dihedral angle of molecular models of the protease and its mutants. Figure 3C shows the trend favorable toward low hydrophobicity as demonstrated not only by lines with negative slopes but also by a similar overall trend in the datapoints. This may not mean that lower hydrophobicity will always be effective in improving the thermostability of proteins. The function of proteins is a net result of different forces and interactions (Matthews, 1993), which is especially true when two causes are contradictory such as the relation between hydrophobicity and bulkiness in this study. Parts D and E of Figure 3 show almost no relation of β-strand and α-helix with thermostability as the maps demonstrate no special trends. This is reasonable because the site 1−16 is originally a helix region and a drastic change was not expected from a one-site mutation.

This study was conducted to confirm the effectiveness of our optimization approach by applying the RCG program to site-directed mutagenesis. Therefore, the optimization was discontinued after 13 mutations. At this stage, it was felt that the study had validated the RCG approach. Using the RCG program for mutation of more than one site will be the next important application, since more substantial structural changes can be expected under these circumstances. As shown in the preceding paper, the application of the mapping method to multimodal functions with up to six factors was proven to be feasible, by using randomized model functions. The implication is that simultaneous mutations of up to three sites in molecules can be optimized, despite the fact that the usefulness and reliability of trend lines are generally reduced as the number of factors increases. The RCG program can accommodate the use of different amino acid scales for different mutation sites in the same protein for multisite mutagenesis.

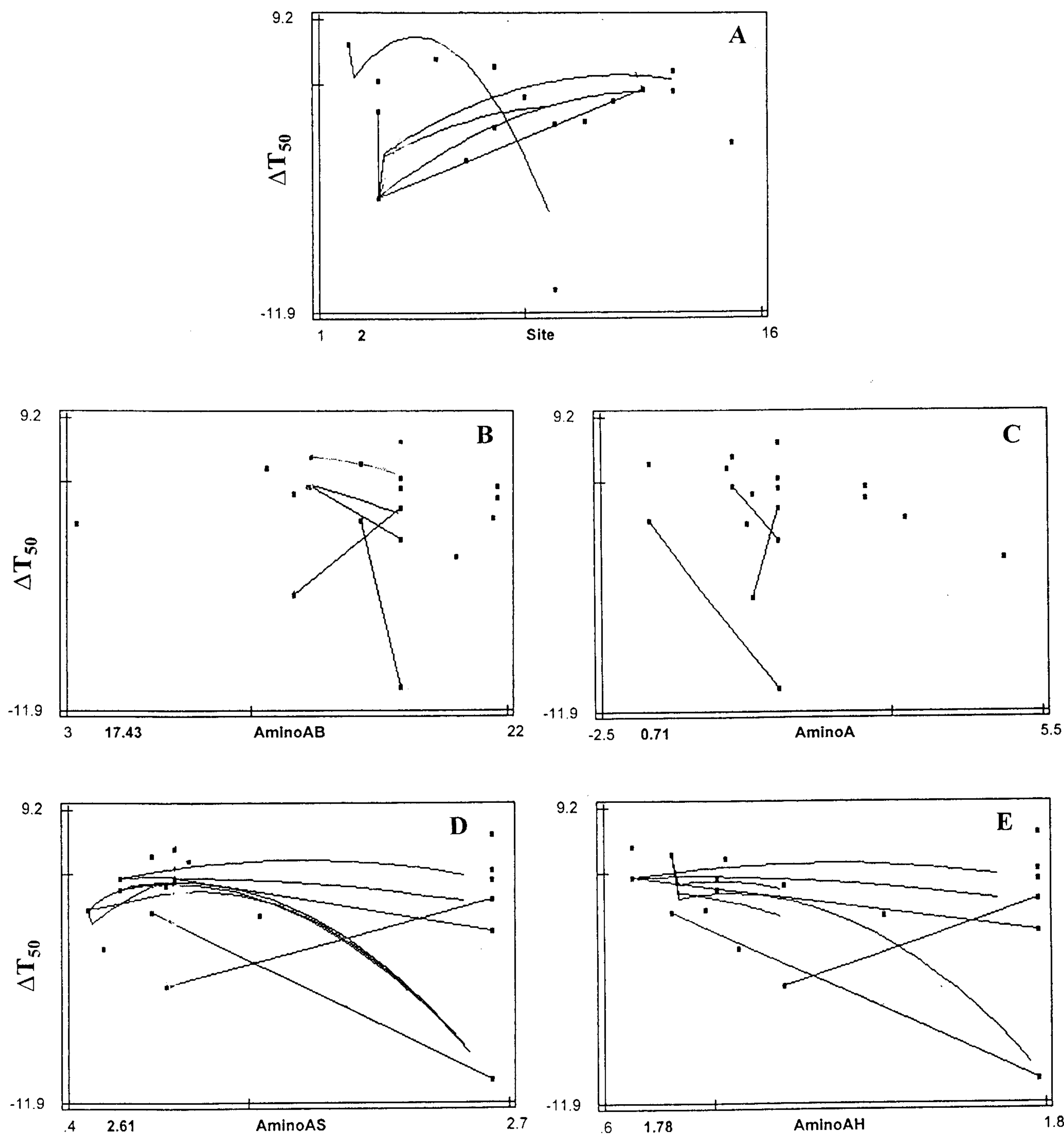

Figure 3. Maps from the random-centroid optimization directed mutation of neutral protease. AminoAB, amino acid scale for bulkiness; AminoA, amino acid scale for hydrophobicity; AminoAH, amino acid scale for α-helix propensity; AminoAS, amino acid scale for β-strand propensity; ΔT_{50}, difference in half-survival temperature (T_{50}) from that of the wild-type enzyme.

General Discussion. There are several good review papers that discuss the effects of genetic engineering on protein stability (Matthews, 1993; Vieille and Zeikus, 1996). Vieille and Zeikus (1996) stated that, owing to the complexity of protein structure, the immense number of mutational combinations for a single protein prevents the definition of any universal protein stabilization mechanisms. Both papers discussed the importance of noncovalent interactions and secondary structures in protein molecules. Among all of the amino acids, only the introduction of proline affords characteristic effects on the thermostability of proteins (Vieille and Zeikus, 1996). In our previous work (Nakamura et al., 1997), it was found that I140P increased the T_{50} of *B. stearothermophilus* protease by 7.5 °C, which surpasses the best value of 6.5 °C of V143E obtained by using the RCG optimization in this study. Interestingly, the conclusions of both studies were the same—that increased thermostability of the mutants was due to molecular rigidification at the N-terminal end of the active-site helix. It may be worthwhile to introduce proline to the mutant V143E protease, as an example of a two-site mutation.

Warren and Petsko (1995) discussed the importance of helical coil stability in thermostable proteins. They found that the temperature differences (Δs) in helix

propagation values (s) more accurately defined the direction of the probability shift in thermostable protein α-helices than "s" alone. This phenomenon is true not just for helix stability, since surface hydrophobicity, rather than average hydrophobicity, plays a more important role in protein function (Nakai and Modler, 1996). However, structure and function cannot be predicted from sequence data at the present time. Structural parameters, such as surface hydrophobicity and Δs, should be useful in structure–function relationship studies but not for mutagenesis. Therefore, it is logical to use the helix propensity and hydrophobicity of amino acids per se for protein engineering purposes since site-directed mutagenesis alters amino acid sequences, which in turn results in protein structural changes. It may be worth noting that the information needed for mutagenesis is that of amino acids constituting protein molecules, whereas those useful for elucidating functions are the properties relating to protein structure.

Tagaya et al. (1989) have reported that since variation of the dihedral angle (ϕ) of a prolyl residue is highly limited due to its pyrrolidine ring, this residue facilitates the turn structure of a loop and might contribute to the stability of a protein as a whole. Hardy et al. (1993) reported ΔT_{50} of 5.6 °C for A69P for the same *B. stearothermophilus* protease. The mutation was introduced in a solvent-exposed flexible region at sites 63–69. Therefore, it may be useful to apply computer-aided optimization to broader regions of a protein in the future, despite the complexity in interpreting the mechanism of functions in question. The propensity to form a β-turn or loop may become increasingly important in the optimization of site-directed mutagenesis.

Although it may be limited in use from the aspect of genetic optimization of protein modification, the RCG program may provide an efficient approach to site-directed mutagenesis for studying the structure–function relationships of protein molecules. It should be emphasized that the RCG approach may be important for creating new functions in protein molecules. Efforts should be continued to accumulate examples of the RCG optimization to establish reliable and versatile methodology. Advantages of the RCG program are that (1) the algorithm is simpler than that of any other global optimization techniques and (2), as a result, it is relatively easy for even nonmathematicians to apply RCG to a protein for optimization of site-directed mutagenesis, to improve a function or create a new function.

The RCO computer program is available to the public via the Internet for downloading to an individual's PC computer. The RCG program also will be more generally available when it is better established. The reader can scrutinize the progress of our RCG study through the UBC home page at http://www.interchange.ubc.ca/agsci/foodsci/nakai.htm.

CONCLUSION

Random-centroid optimization was successfully applied to hypothetical models of the active-site helix of a microbial neutral protease as well as directly to the helix to improve the thermostability of the neutral protease. The mechanism of thermostability could be explained using mapping and was in good agreement with that reported in our previous study on proline introduction to the same helix.

LITERATURE CITED

Blundell, T. L. Problems and solutions in protein engineering—towards rational design. *Trends Biotechnol.* **1994**, *12*, 145–148.

Breaker, R. R.; Joyce, G. F. Inventing and improving ribozyme function: Rational design versus iterative selection methods. *Trends Biotechnol.* **1994**, *12*, 268–274.

Chou, P. Y.; Fasman, G. D. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **1978**, *47*, 251–276.

Delagrave, S.; Goldman, E. R.; Youvan, D. C. Recursive ensemble mutagenesis. *Protein Eng.* **1993**, *6*, 327–331.

Ellington, A. D.; Matsumura, I. Engineering evolution—with molecules. *Yearbook Sci. Future*; Encyclopaedia Britanica: Chicago, 1997; pp 178–187.

Fuellen, G.; Youvan, D. Genetic algorithm and recursive ensemble mutagenesis in protein engineering. *Complexity Int.* **1994**, *1* (http://www.csu.edu.au/ci/voll/fuellen/REM.html).

Gromiha, M. M.; Ponnuswamy, P. K. Relationship between amino acid properties and protein compressibility. *J. Theor. Biol.* **1993**, *165*, 87–100.

Hardy, F.; Vriend, G.; Veltman, O. R.; van der Vinne, B.; Venema, G.; Eijsink, V. G. H. Stabilization of *Bacillus stearothermophilus* neutral protease by introduction of prolines. *FEBS Lett.* **1993**, *317*, 89–92.

Hedstrom, L.; Graf, L.; Stewart, C.-B.; Rutter, W.; Phillips, A. Modulation of enzyme specificity by site-directed mutagenesis. *Methods Enzymol.* **1991**, *202*, 671–687.

Jiang, W.; Bond, J. S. Families of metalloendopeptidases and their relationships. *FEBS Lett.* **1992**, *312*, 110–114.

Jones, D. D. Amino acid properties and side-chain orientation in proteins: A cross correlation approach. *J. Theor. Biol.* **1975**, *50*, 167–183.

Matthews, B. W. Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **1993**, *62*, 139–160.

Meek, J. L.; Rosetti, Z. L. Factors affecting retention and resolution of peptides in high- performance liquid chromatography. *J. Chromatogr.* **1981**, *211*, 15–28.

Muñoz, V.; Serrano, L. Intrinsic secondary structure propensities of the amino acids, using statistical ϕ-ψ matrices: Comparison with experimental scales. *Proteins* **1994**, *20*, 301–311.

Nakai, S.; Modler, H. W. *Food Proteins: Properties and Characterization*; VCH Publishers: Weiheim, Germany, 1996.

Nakai, K.; Kidera, A.; Kanehisa, M. Cluster analysis of amino acid indices for prediction of protein structure and function. *Protein Eng.* **1988**, *2*, 93–100.

Nakai, S.; Dou, J.; Lo, K. V.; Scaman, C. H. Optimization of site-directed mutagenesis. 1. A new random-centroid optimization program for Windows useful in research and development. *J. Agric. Food Chem.* **1998**, *46*, 1642–1654.

Nakamura, S.; Tanaka, T.; Yada, R. Y.; Nakai, S. Improving the thermostability of *Bacillus stearothermophilus* neutral protease by introducing proline into the active site helix. *Protein Eng.* **1997**, *10*, 1263–1269.

Nozaki, Y.; Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solution. *J. Biol. Chem.* **1971**, *246*, 2211–2217.

Sander, C. Design of protein structures: Helix bundles and beyond. *Trends Biotechnol.* **1994**, *12*, 163–167.

Tagaya, M.; Yagami, T.; Noumi, T.; Futai, M.; Kishi, F.; Nakazawa, A.; Fukui, T. Site-directed mutagenesis of Pro-17 located in the glycine-rich region of adenylate kinase. *J. Biol. Chem.* **1989**, *264*, 990–994.

Vieille, C.; Zeikus, J. G. Thermoenzymes: Identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol.* **1996**, *14*, 183–190.

Warren, G. L.; Petsko, G. A. Composition analysis of α-helices in thermophilic organisms. *Protein Eng.* **1995**, *8*, 905–913.

Wilce, M. C. J.; Aguilar, M.-I.; Heam, M. T. Physicochemical basis of amino acid hydrophobicity scales: Evaluation of four

new scales of amino acid hydrophobicity coefficients derived from RP-HPLC of peptides. *Anal. Chem.* **1995**, *67*, 1210–1219.
Wilmot, C. M.; Thornton, J. M. Analysis and prediction of the different types of β-turn in proteins. *J. Mol. Biol.* **1988**, *203*, 221–232.
Wilson, K. J.; Honegger, A.; Stotzel, R. P.; Hughes, G. J. The behavior of peptides on reverse-phase supports during high-pressure liquid chromatography. *Biochem. J.* **1981**, *199*, 31–41.
Zimmerman, J. M.; Eliezer, N.; Simha, R. The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.* **1968**, *21*, 170–201.

Received for review July 15, 1997. Revised manuscript received January 14, 1998. Accepted January 22, 1998. We are grateful to the Natural Sciences and Engineering Research Council of Canada for the research grant to support this study.

JF970610R