Biochemical and Biophysical Research Communications 371 (2008) 122–126

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Engineering of protease variants exhibiting altered substrate specificity

Saravanan Sellamuthu, Bae Hyun Shin, Eui Seung Lee [1], Seong-Hwan Rho [2], Wangtaek Hwang, Yong Jae Lee [3], Hye-Eun Han, Jae Il Kim, Woo Jin Park *

Department of Life Science, Gwangju Institute of Science and Technology, 1 Oryong-dong, Buk-gu, Gwangju 500-712, Republic of Korea

ARTICLE INFO

Article history:
Received 1 April 2008
Available online 14 April 2008

Keywords:
HAV 3C protease
Protease engineering
Genetic assay for site-specific proteolysis (GASP)

ABSTRACT

By using an improved genetic screening system, variants of the HAV 3CP protease which exhibit altered P2 specificity were obtained. We randomly mutated the His145, Lys146, Lys147, and Leu155 residues that constitute the S2 pocket of 3CP and then isolated variants that preferred substrates with Gln over the original Thr at the P2 position using a yeast-based screening method. One of the isolated variants cleaved the Gln-containing peptide substrate more efficiently *in vitro*, proving the efficiency of our method in isolating engineered proteases with desired substrate selectivity.

Proteases play essential roles in a variety of cellular processes. A number of "rational" approaches have been taken to produce variants possessing improved and desired properties, such as increased protein stability [1,2] and better pharmacokinetic profiles [3]. However, it has proven difficult to generate protease variants exhibiting altered substrate specificity by such "rational" approaches. This is primarily due to a lack in comprehensive understanding of the relationship between protein structure and function. Therefore, a "directed evolution" approach in which target genes are partially or completely mutated remains the only viable option for generating such protease variants. This approach generally includes two steps: generation of genetic diversity within the target genes and selection for variants with the desired activity. While efficient methods for the generation of genetic diversity have already been implemented, a lack of efficient screening methods has greatly hampered the use of the "directed evolution" approach. A yeast-based assay suitable for detection of site-specific proteolysis was previously reported [4,5]. In this study, we showed that this method can be used as an efficient screening method for engineered protease variants.

Materials and methods

Vector constructions. The *GAL1* promotor and *STE2* membrane anchor sequences were introduced into the pRS303 vector between the Sac1 and BamH1 sites, following which the HAV 3CP coding sequence was inserted between the EcoR1 and Xho1 sites, resulting in pGAL-3CP. Construction of pADH-Ste-Lex was described previously [6]. Residues spanning the 2B/2C junction of the HAV polyprotein were chosen to represent the substrate. Two sets of oligonucleotides were synthesized for the substrate sequences: (i) for the TQ substrate, 5′-CAT GGA GAA TGA TGG AGC TGA GAA CTC AGA GCT TTT CTA ATA-3′ and 5′-GAT CTA TTA GAA AAG CTC TGA GTT CTC AGC TCC ATC ATT CTC-3′; and (ii) for the QQ substrate, 5′-CAT GGA GAA TGA TGG AGC TGA GAC AGC AGA GCT TTT CTA ATA-3′ and 5′-GAT CTA TTA GAA AAG CTC TGC TGT CTC AGC TCC ATC ATT CTC-3′. The oligonucleotides were annealed and cloned into the Nco1–BamH1 sites, resulting in pADH-Ste-TQ-Lex and pADH-Ste-QQ-Lex, respectively.

Yeast transformation and screening. Plasmids containing protease and substrate sequences were co-transformed into the yeast strain EGY48 (MATα, *ura*3, *his*3, *trp*1, $LexA_{op(X6)}$-*LEU2*) by a standard lithium acetate method [7]. The transformed yeast cells were plated onto non-selective plates lacking tryptophane, histidine, and uracil. After 3 days of incubation at 30 °C, the cells were replica-plated onto GAL-inducible selective plates containing 2% galactose and 1% raffinose and deprived of tryptophane, histidine, uracil, and leucine and onto GAL-repressible selective plates containing 2% glucose instead of galactose and raffinose. The faster growing colonies from the *GAL*-inducible selective plates were selected for further testing on X-gal plates.

β-Galactosidase assay. Three hundred microliters of yeast cells was mixed with 400 μl of Z-buffer (60 mM Na_2HPO_4, 40 mM $NaH_2PO_4{\cdot}H_2O$, 10 mM KCl, and 1 mM $MgSO_4{\cdot}7H_2O$, pH 7.0), 50 mM β-mercaptoethanol, 50 μl of chloroform, and 20 μl of 0.1% SDS, and then vortexed for 10 s. After adding 200 μl of ONPG (*o*-nitrophenyl-β-D-galactopyraniside) solution, the mixture was incubated at 37 °C until a yellow color appeared. Five hundred microliters of 1 M Na_2CO_3 was added to stop the reaction. The absorbance at A_{420} was measured and the specific activity was calculated as described elsewhere [8].

Construction of S2 pocket randomized 3CP protease library. The H145, K146, K147 residues of 3CP were randomized using the primer set: 5′-TCG AGA ATT CAT GTC AAC TTT GGA AAT AGC A-3′ and 5′-TGT TGT ACC ATC ATT MNN MNN MNN AAC

* Corresponding author. Fax: +82 62 970 2484.
E-mail address: wjpark@gist.ac.kr (W.J. Park).
[1] Present address: Department of Orthopedic Surgery, Chonnam National University Hwasun Hospital, 160 Ilsim-ri, Hwasun-eup, Hwasun-gun, Jeonnam 519-809, Korea.
[2] Present address: Physics Institute, University of Freiburg, Hermann-Herder-Strasse 3A, Freiburg 79102, Germany.
[3] Present address: Department of Pharmacology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA.

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.04.026

ATA AGT AGC TTT CTC TTC-3′, and the L155 residue was randomized using the primer set: 5′-ATG ATG GTA CAA CAG TTG ATN NKA CTG TGG ATC AGG CAT GGA G-3′ and 5′-TCG ACT CGA GTT ACT GAC TTT CAA TTT TCT T-3′. The resultant products were used as templates to make the full-length HKKL randomized 3CP. The final PCR product was digested with EcoR1–Xho1 and cloned into pGAL vector. The ligated plasmids were transformed into *E. coli* DH10B cells by electroporation. A total of 10^6 colonies were typically obtained.

Purification of 3CP and its variants. 3CP and its variants were expressed in the *E. coli* strain BL21 (DE3) and purified using the IMPACT-CN system (New England Biolabs) as described previously [9].

Determination of in vitro cleavage of peptide substrate. Cleavage reactions contained 40 mM Tris–HCl (pH 7.2), 1 mM dithiothreitol, 0.1% Nonidet P-40, various concentrations of substrate peptides, and purified 3CP or its variants in a total volume of 100 μl. Reactions were quenched by the addition of 100 μl of 0.1% trifluoroacetic acid and analyzed by reverse phase HPLC (Shimadzu) on a 10 cm Vydac C18 column using a 0–95% 30 min linear gradient of acetonitrile in 0.1% trifluoroacetic acid. The cleaved peptides were characterized by MALDI-TOF (Shimadzu).

Results and discussion

The rationale behind our method, referred to as GASP (genetic assay for site-specific proteolysis), is illustrated in Fig. 1A. We chose hepatitis A virus (HAV) 3C protease (3CP) [10,11] as a raw material for engineering. A DNA fragment encoding 3CP was inserted into pGAL, resulting in the protease plasmid, pGAL-3CP. Expression of 3CP was driven by the inducible *GAL1* promoter. Two fusion genes were created and inserted into the substrate plasmid pADH-Ste-Lex, resulting in pADH-Ste-TQ-Lex and pADH-Ste-QQ-Lex. These fusion genes were constitutively expressed under the control of *ADH* promoter. The encoded fusion proteins contained the transcription factor, LexA-b42, which is connected via linker sequence to the truncated cytoplasmic domain of the yeast integral membrane protein, STE2. While the pADH-Ste-TQ-Lex-encoded fusion protein contained the normal substrate sequence of 3CP (ELRTQ↓SFSN; designated "TQ substrate") of 3CP as the linker sequence between STE2 and LexA-b42, the pADH-Ste-QQ-Lex-encoded fusion protein contained a modified substrate sequence (ELRQQ↓SFSN; designated "QQ substrate") in which Thr was replaced by Gln in the P2 position. Due to the presence of the STE2 transmembrane domain, the fusion proteins were anchored in the plasma membrane.

When the fusion proteins were cleaved in the linker sequence by 3CP, LexA-b42 was released from the plasma membrane, entered the nucleus, and activated the reporter genes, *Leu2* and *LacZ*. A *Leu2* mutant reporter strain, EGY48 [12], was transformed with both pGAL-3CP and pADH-Ste-TQ-Lex, and plated on selective media lacking leucine and containing either glucose or galactose (galactose induces 3CP expression). The transformants displayed excellent growth on the selective media containing galactose but not glucose, and also stained strongly blue on X-gal plates only in the presence of galactose. This indicated that the TQ substrate was specifically cleaved by 3CP. In a similar experiment, the QQ substrate was not cleaved by 3CP (Fig. 1B). We sought to generate 3CP variants that could cleave the QQ substrate more efficiently than the TQ substrate.

A close examination of the 3CP crystal structure suggested that four amino acids (H145, K146, K147, L155) might constitute the S2

Table 1
Wt 3CP and the isolated 3CP variants

	145	146	147	155
3CP	His	Lys	Lys	Leu
Var1	Ser	Glu	Gln	Leu
Var2	Gly	Glu	Ala	Leu
Var3	Gly	Ile	Thr	Leu
Var4	Gly	Val	Ala	Leu
Var5	Cys	Glu	Gly	Val
Var6	Cys	Glu	Ser	Met
Var7	Cys	Trp	Gly	Trp
Var8	Gly	Asp	Asn	Leu
Var9	Gly	Glu	Thr	Ile
Var10	Gly	Leu	Gln	Leu
Var11	Gly	Val	Arg	Leu
Var12	Gly	Leu	Lys	Leu

The mutated amino acid residues in the isolated 3CP variants along with those of the Wt 3CP are shown.

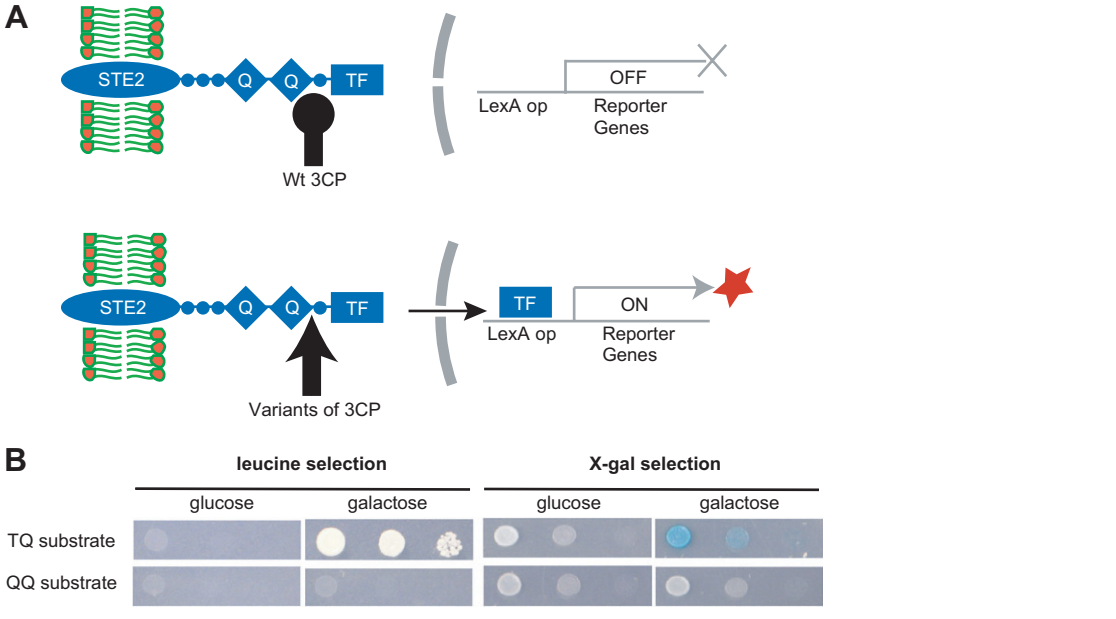

Fig. 1. (A) Principle of GASP. The triple fusion protein is anchored in the plasma membrane. Wild-type (Wt) 3CP cannot cleave the QQ substrate. When 3CP variants cleave the QQ substrate, the LexA-b42 moiety is released from the plasma membrane and activates the expression of reporter genes, *Leu2* and *LacZ*. (B) Yeast strain EGY48 was co-transformed with pGAL-3CP and either pADH-STE-TQ-Lex (TQ substrate) or pADH-STE-QQ-Lex (QQ substrate), and then plated on leucine selective and X-gal plates containing either glucose or galactose. The results indicated that Wt 3CP can only cleave the TQ substrate.

Substrate / Protease	ELRTQSFSN			ELRQQSFSN		
	WHUL⁻ glu	WHUL⁻ gal	β-galactosidase units	WHUL⁻ glu	WHUL⁻ gal	β-galactosidase units
None			0.8			7.6
Wt 3CP			418.1			3.9
Var 1			893.3			838.8
Var 2			14.2			655.6
Var 3			138.9			314.9
Var 4			14.3			152.5
Var 5			0.4			881.8
Var 6			16.7			839.2
Var 7			0.7			285.3
Var 8			7.5			131.3
Var 9			10.0			55.9
Var 10			21.1			33.4
Var 11			73.9			407.0
Var 12			59.4			169.6

Fig. 2. A yeast strain EGY48 was co-transformed with plasmids encoding Wt 3CP or isolated variants (Var1–12), and plasmids of fusion proteins containing either the TQ or QQ substrates. Proteolytic activity was determined by cell growth on selective media and β-galactosidase activity.

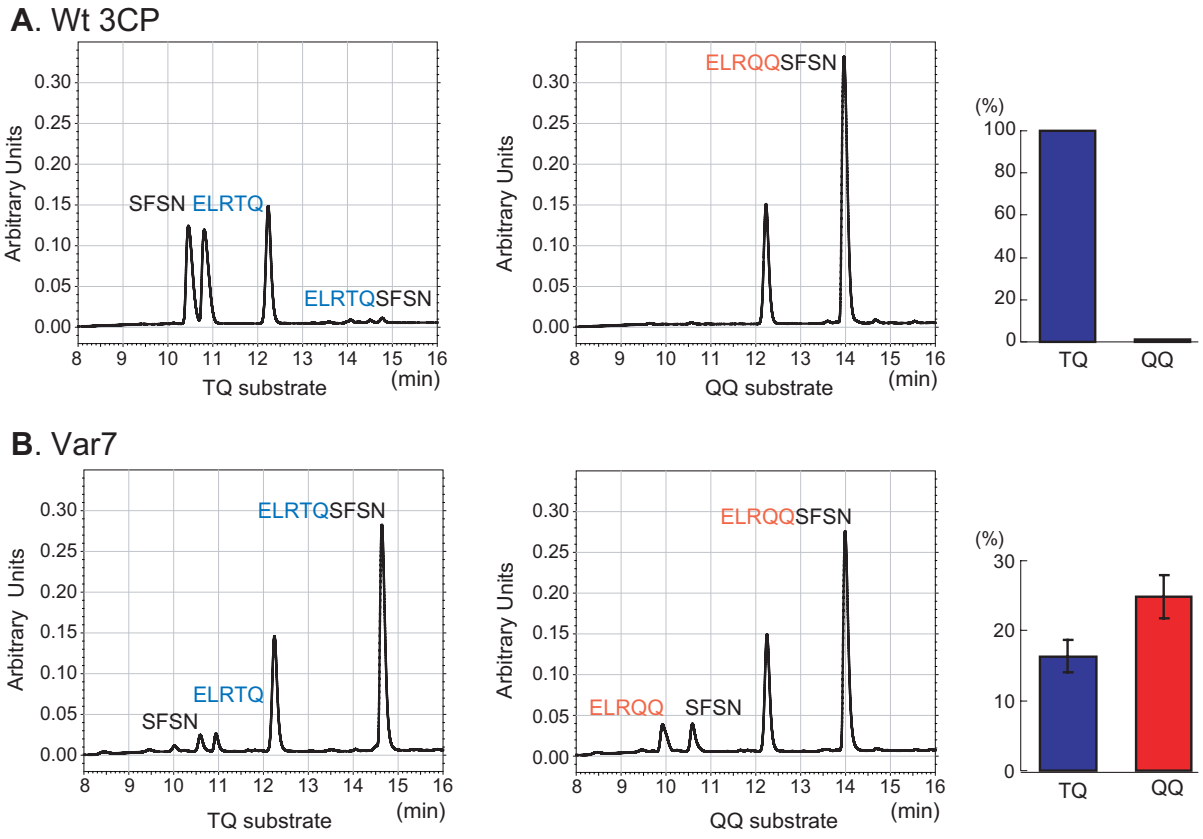

Fig. 3. *In vitro* cleavage of peptide substrates by purified enzymes. TQ and QQ peptide substrates (200 μM) were incubated with the purified (A) Wt 3CP (1 μM) and (B) Var7 (10 mM) for 6 h. The resulting fragments were separated by HPLC and their identities were further confirmed by MALDI-TOF.

pocket. Therefore, these residues were randomly mutated by successive PCR. The mutated 3CP genes were subcloned into pGAL, and a library of $\sim 1 \times 10^6$ transformants was obtained. Ten randomly picked colonies from this library contained completely randomized sequences at all four residues (data not shown). About 200,000 colonies co-transformed with the library and pADH-Ste-

QQ-Lex were plated on selective media lacking leucine and further tested for color development on X-gal plates. Only one colony was isolated from this initial screening, and contained mutations at all four positions (H145C, K146S, K147G, L155A). In addition, it contained a M29V mutation. Re-examination of the crystal structure suggested that M29 might be an essential element of the S2 pocket.

When the M29V mutation was corrected by site-directed mutagenesis, the isolated 3CP variant completely lost its ability to cleave the QQ substrate (data not shown) implying that the M29V mutation is critical. Therefore, we generated a library of mutations at the four residues starting with the 3CP gene in which M29 was mutated to Val. Again, a library of $\sim 1 \times 10^6$ transformants was generated and validated by sequencing. From the screening of about 250,000 colonies, we obtained 12 colonies which showed strong proteolytic activity toward the QQ substrate. The mutated sequences in these 3CP variants (Var1–12) are shown in Table 1. The plasmids pGAL-3CP-Var were isolated and then co-introduced into EGY48 with either pADH-Ste-TQ-Lex or pADH-Ste-QQ-Lex. The proteolytic activity of the variants for the TQ or QQ substrates was determined by the growth of the transformants on selective media and their β-galactosidase activity (Fig. 2). While Var1–4 and Var9–12 cleaved both substrates equally well, Var5–8 cleaved the QQ substrate much more efficiently than the TQ substrate.

Wild-type (Wt) 3CP and Var5–8 were then purified from *E. coli*, and two peptides for the TQ and the QQ substrates were synthesized. The cleavage of the peptide substrates by Wt 3CP and Var5–8 *in vitro* were analyzed by HPLC and MALDI-TOF. Wt 3CP efficiently cleaved the TQ but not the QQ substrate as expected (Fig. 3A), whereas Var5–8 cleaved both substrates. In particular, Var7 showed a relative preference for the QQ substrate, cleaving 25 ± 3% of the QQ substrate and 16 ± 2% of the TQ substrate in a same condition (Fig. 3B). Homology modeling of HAV3C (PDB ID: 2H9H) [13] bound with its substrate revealed that the S2 pocket of Var7 but not that of Wt 3CP may accommodate the bulky Gln residue (Fig. 4A and B). His145 lines the wall of the S2 pocket along with His44. Replacement of His145 with Cys in Var 7 appears to result in the expansion of S2 pocket, which enables the accommodation of Gln.

The results of this study indicated that protease variants possessing altered substrate specificity can be generated using GASP. Due to its simple approach for screening a large library of mutated protease variants, this method might be suitable

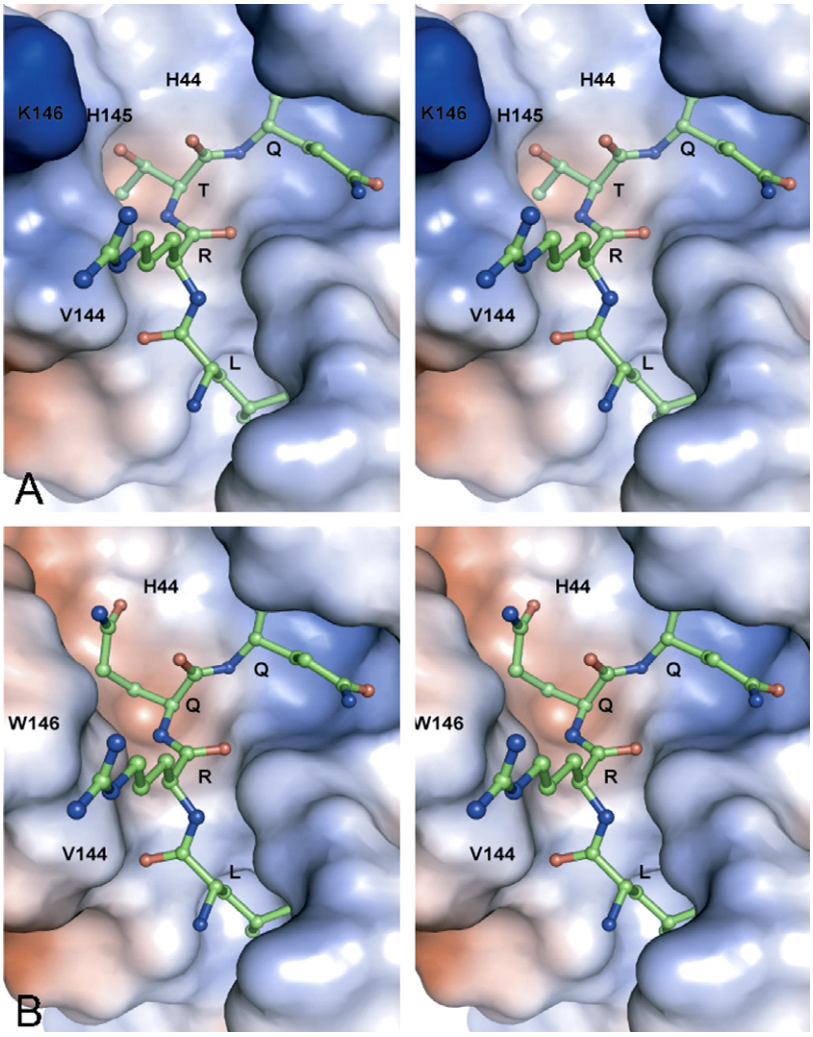

Fig. 4. Comparison of substrate binding pockets of (A) Wt 3CP and (B) Var7. The substrate residues are represented as ball-and-stick models and colored by green. The electrostatic potential ($-60 < kT < 60$) was mapped to the solvent accessible surfaces of the proteins. Figures were produced using PyMol [15]. Note that the bulky side chain of Gln at the P2 position can fit in Var7 but not Wt 3CP. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

for a "directed evolution" approach of generating protease variants. To our best knowledge, only one paper has previously reported a method for screening randomly-mutated protease variants [14]. The authors mutated an *E. coli* endopeptidase OmpT and isolated variants that referentially cleaved an Ala-Arg peptide bond rather than the Arg-Arg bond which is preferred by the wild-type OmpT. By exploiting a simultaneous selection and counter selection scheme for protease variant isolation, these researchers were able to isolate variants exhibiting both high catalytic activity and exquisite substrate selectivity. However, while the previous method may only be applicable to cell surface proteases, GASP can be used for proteases which function in the cytoplasm.

GASP is semi-quantitative, in the sense that the proteolytic activity can be roughly estimated by the growth rate of the transformants on selective plates. For example, the transformants harboring the Wt 3CP and the TQ substrate constructs appeared on selective plates after 3 days of incubation, while the transformants harboring Var7 and the QQ substrate constructs usually began to appear after 7 days of incubation. This was consistent with i*n vitro* kinetic analyses showing that the turnover rate of Var7 was >100 times lower than that of Wt 3CP (data not shown). Therefore, GASP can be used for improving the catalytic activity as well as altering substrate specificity of proteases. In conclusion, we present proof of principle that our genetic screening method, GASP, can be used to engineer protease variants possessing improved properties through directed evolution approaches.

Acknowledgments

This work was supported by a GRL grant (M6-0605-00-0001) and a NRL grant (M1-0104-00-0007) from the Ministry of Science and Technology, Korea.

References

[1] M.K. Ditursi, S.J. Kwon, P.J. Reeder, J.S. Dordick, Bioinformatics-driven, rational engineering of protein thermostability, Protein Eng. Des. Sel. 19 (2006) 517–524.
[2] L.D. Cabrita, D. Gilis, A.L. Robertson, Y. Dehouck, M. Rooman, S.P. Bottomley, Enhancing the stability and solubility of TEV protease using in silico design, Protein Sci. 16 (2007) 2360–2367.
[3] D.T. Berg, B. Gerlitz, J. Shang, T. Smith, P. Santa, M.A. Richardson, K.D. Kurz, B.W. Grinnell, K. Mace, B.E. Jones, Engineering the proteolytic specificity of activated protein C improves its pharmacological properties, Proc. Natl. Acad. Sci. USA 100 (2003) 4423–4428.
[4] C.J. Hawkins, S.L. Wang, B.A. Hay, A cloning method to identify caspases and their regulators in yeast: identification of *Drosophila* IAP1 as an inhibitor of the *Drosophila* caspase DCP-1, Proc. Natl. Acad. Sci. USA 96 (1999) 2885–2890.
[5] S.Y. Kim, K.W. Park, Y.J. Lee, S.H. Back, J.H. Goo, O.K. Park, S.K. Jang, W.J. Park, In vivo determination of substrate specificity of hepatitis C virus NS3 protease: genetic assay for site-specific proteolysis, Anal. Biochem. 284 (2000) 42–48.
[6] H. Kang, S.Y. Kim, W.J. Park, An improved strategy for a genetic assay for site-specific proteolysis, Mol. Cells 11 (2001) 263–266.
[7] D. Gietz, A. St Jean, R.A. Woods, R.H. Schiestl, Improved method for high efficiency transformation of intact yeast cells, Nucleic Acids Res. 20 (1992) 1425.
[8] J.H. Miller, Experiments in Molecular Genetics, first ed., Cold Spring Harbor Laboratory, New York, 1972.
[9] T.W. Muir, D. Sondhi, P.A. Cole, Expressed protein ligation: a general method for protein engineering, Proc. Natl. Acad. Sci. USA 95 (1998) 6705–6710.
[10] V. Gauss-Muller, D. Jurgensen, R. Deutzmann, Autoproteolytic cleavage of recombinant 3C proteinase of hepatitis A virus, Virology 182 (1991) 861–864.
[11] E.M. Bergmann, M.M. Cherney, J. McKendrick, S. Frormann, C. Luo, B.A. Malcolm, J.C. Vederas, M.N. James, Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV, Virology 265 (1999) 153–163.
[12] J. Estojak, R. Brent, E.A. Golemis, Correlation of two-hybrid affinity data with in vitro measurements, Mol. Cell Biol. 15 (1995) 5820–5829.
[13] J. Yin, M.M. Cherney, E.M. Bergmann, J. Zhang, C. Huitema, H. Pettersson, L.D. Eltis, J.C. Vederas, M.N. James, An episulfide cation (thiiranium ring) trapped in the active site of HAV 3C proteinase inactivated by peptide-based ketone inhibitors, J. Mol. Biol. 361 (2006) 673–686.
[14] N. Varadarajan, J. Gam, M.J. Olsen, G. Georgiou, B.L. Iverson, Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity, Proc. Natl. Acad. Sci. USA 102 (2005) 6855–6860.
[15] W.L. DeLano, The PyMOL molecular graphics system, 2002. Available from: <http://www.pymol.org>.