Rational design of a novel propeptide for improving active production of Streptomyces griseus trypsin in Pichia pastoris

Zhenmin Ling, Yi Liu, Shaolei Teng, Zhen Kang, Jingjing Zhang, Jian Chen, and Guocheng Du

The Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China
School of Biotechnology, Jiangnan University, Wuxi 214122, China
Department of Genetics & Biochemistry, Clemson University, Clemson, South Carolina, U. S. A.
National Engineering Laboratory for Cereal Fermentation Technology, Jiangnan University, Wuxi 214122, China
The Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China

*Corresponding authors: Zhen Kang, Address: School of Biotechnology, Jiangnan University, 1800 Lihu Avenue, Wuxi 214122, China, Tel: +86-510-85918307, Fax: +86-510-85918309, E-mail: zkang@jiangnan.edu.cn.
Guocheng Du, Address: School of Biotechnology, Jiangnan University, 1800 Lihu Avenue, Wuxi 214122, China, Tel: +86-510-85918309, Fax: +86-510-85918309, E-mail: gcdu@jiangnan.edu.cn.

Running title: Rational design of a propeptide for trypsin production
ABSTRACT

Applying *in silico* simulations and *in vitro* experiments, the amino acid proline was proved to have a profound influence on *Streptomyces griseus* trypsinogen, and the hydrogen bond between H^{57} and D^{102} was crucial for trypsin activity. By introducing an artificial propeptide IVEF, the titer of trypsin was increased by 6.71-fold.

KEYWORDS: *Streptomyces griseus*; trypsinogen; simulation; propeptide
Trypsin (E.C. 3.4.21.4), an archetypal serine protease, has been discovered in many organisms ranging from bacteria to mammals (1). Bovine trypsinogen, as a representative, has been well studied (2, 3) and function through the cleavage of an N-terminal hexapeptide (VDDDDK) by auto-activation (4) or hydrolyzation with enterokinase (5). In addition, previous studies have shown that four unusual aspartyl residues (DDDD) in trypsinogen are strictly conserved in mammals, including humans, and may have arisen by selective pressure during the course of evolution. In comparison, although activation of the bacterial Streptomyces griseus trypsinogen also involved removal of the propeptide (APNP), the enzymatic mechanism was considered to be different to mammalian trypsinogen (3, 6). Since 1970s, the amino acid sequence, open reading frame and crystal structure of S. griseus trypsin have been solved (7-10). However, the activation mechanism of S. griseus trypsinogen remains poorly understood. In 1990s, Kim et al deduced that S. griseus trypsin consists of three parts according to its protein amino acid sequence and activation of S. griseus trypsinogen was non-autocatalytic (8). Recently, we have successfully overexpressed S. griseus trypsin in Pichia pastoris (11) by including four residues (YVEF) at N-terminus (Fig. 1A). The four residues were inserted by digestion with restriction enzymes SnaBI and EcoRI, and cleavage of the α-factor signal peptide by peptidases KEX2 and STE13 (12-14). However, no activity was detected when substituting the native propeptide with mammalian trypsinogen or its mutant VDDDDD (11). Recently, many short synthetic amphiphilic peptides were discovered, designed and used for improving enzyme properties (15-17). Therefore, artificial modification of the N-terminal propeptide represents a suitable alternative for enzyme engineering.

In the past decade, computational methods for protein simulation and analysis have been developed and exploited, which in turn has promoted protein research by combining such data with in vitro experimental verification (18-20). In particular, when the crystal structure of the target protein is not available, a computer modeling method can be indispensable for determining the protein conformation. Currently, the S. griseus trypsinogen crystal structure has not been resolved. Therefore, computer-aided modeling should offer key information that further directs rational engineering.

Applying S. griseus ATCC 10137 genome as the template, the fragments encoding trypsinogen or trypsin (Fig. 1) were amplified with designed oligonucleotides (Table S2) and subcloned into pPIC9k (Table S1), which is controlled by the AOX1 promoter and the α-factor signal peptide from Saccharomyces cerevisiae (Fig. 1A). All the recombinants constructed (Table S1) were further confirmed by PCR amplification (Fig. S1) and trypsin activity was measured as described previously (11). As shown in Fig. 2, the recombinants Exmt, ExYmt and mt accumulated trypsin to 0.6 U ml$^{-1}$, 0.32 U ml$^{-1}$ and 0.21 U ml$^{-1}$, respectively. However, the recombinants that contained a proline residue showed no activity, despite the
observation of expression and secretion (Fig. 5A). The difference between recombinants ExPmt and ExYmt as well as Pmt and mt suggested that the proline was vital to the inhibition of trypsin. Moreover, compared with mt, the recombinants Exmt and ExYmt accumulated higher titers of trypsin (Fig. 2).

To study the inhibition mechanism of the native propeptide towards the _S. griseus_ trypsinogen, especially the role of the residue proline, we carried out molecular dynamics. Molecular dynamics (21) of the recombinants were simulated by the NAMD software with the charmm force field (http://www.ks.uiuc.edu/Research/namd) and comparatively analyzed with the active _S. griseus_ trypsin (PDB ID: 1SGT). Total electrostatic energy in a Particle Mesh Ewald periodic box was calculated by Ewald summation method and the whole system was minimized using the descent method plus the conjugate gradient method. As shown in Fig. 3A, the major features of native trypsin were characterized. First, three disulfide bonds between residues C^{168}-C^{182}, C^{191}-C^{220}, and C^{42}-C^{58} held the substrate binding pocket rigid and the correct fold was observed. Second, three Hydrogen (H) bonds among the catalytic triad (H^{57}, D^{102} and S^{195}) maintained the accurate conformation of the catalytic center (4). Third, one H-bond had formed between V^{16} and D^{194} and this interaction stabilized the structure of trypsin. In contrast, although ExPmt, APNPmt or Pmt were found to be expressed (Fig. 5A), their structures did not resemble the crystal structure (Fig. 3 C, E and F). Consequently, it could be concluded that the propeptide, especially proline, affected the formation of disulfide bonds and H-bonds, which eventually resulted in a native inactive trypsinogen. In contrast, both Exmt (YVEF) and ExYmt (YVEFY) were actively overexpressed. Furthermore, compared with the native trypsin, the recombinants Exmt and ExYmt appeared much more flexible presumably because of the missing disulfide bond (Fig. 3 A, B, D), suggesting the disulfide bond between residues C^{42} and C^{58} could be ignored, yet trypsin activity is retained. In the native trypsin and the recombinant Exmt, there are three H-bonds exist in the catalytic triad. Nevertheless, only one H-bond formed between H^{57} and D^{102} in ExYmt, suggesting the H-bond between H^{57} and D^{102} may be essential for active trypsin. Furthermore, in Exmt and the native trypsin, one H-bond formed between V^{16} and D^{194}, whereas in ExYmt, the H bond formed between V^{16} and D^{189}. As a result, it could be deduced that the H-bond between V^{16} and the loop (D^{189}-D^{194}) is important for active trypsin. In comparison with the native trypsin, the recombinants contained fewer H-bonds and ion pairs, which indicated the profound influence of the N-terminal propeptide (Table S3).

After exploring the inhibitory mechanism of the propeptide, we evaluated the significance of the three factors identified above. First, we designed and constructed four different recombinants YVEImt, RVEFmt, RIRImt and YVERmt (Fig. 1C) and simulated their structures. Interestingly, introduction of the peptide YVEI or RVEF resulted in successful formation of disulfide bonds (C^{168}-C^{182}, C^{191}-C^{220}) and H-bond (V^{16}-D^{194}) (Fig. 4A, observation of expression and secretion (Fig. 5A). The difference between recombinants ExPmt and ExYmt as well as Pmt and mt suggested that the proline was vital to the inhibition of trypsin. Moreover, compared with mt, the recombinants Exmt and ExYmt accumulated higher titers of trypsin (Fig. 2).

To study the inhibition mechanism of the native propeptide towards the _S. griseus_ trypsinogen, especially the role of the residue proline, we carried out molecular dynamics. Molecular dynamics (21) of the recombinants were simulated by the NAMD software with the charmm force field (http://www.ks.uiuc.edu/Research/namd) and comparatively analyzed with the active _S. griseus_ trypsin (PDB ID: 1SGT). Total electrostatic energy in a Particle Mesh Ewald periodic box was calculated by Ewald summation method and the whole system was minimized using the descent method plus the conjugate gradient method. As shown in Fig. 3A, the major features of native trypsin were characterized. First, three disulfide bonds between residues C^{168}-C^{182}, C^{191}-C^{220}, and C^{42}-C^{58} held the substrate binding pocket rigid and the correct fold was observed. Second, three Hydrogen (H) bonds among the catalytic triad (H^{57}, D^{102} and S^{195}) maintained the accurate conformation of the catalytic center (4). Third, one H-bond had formed between V^{16} and D^{194} and this interaction stabilized the structure of trypsin. In contrast, although ExPmt, APNPmt or Pmt were found to be expressed (Fig. 5A), their structures did not resemble the crystal structure (Fig. 3 C, E and F). Consequently, it could be concluded that the propeptide, especially proline, affected the formation of disulfide bonds and H-bonds, which eventually resulted in a native inactive trypsinogen. In contrast, both Exmt (YVEF) and ExYmt (YVEFY) were actively overexpressed. Furthermore, compared with the native trypsin, the recombinants Exmt and ExYmt appeared much more flexible presumably because of the missing disulfide bond (Fig. 3 A, B, D), suggesting the disulfide bond between residues C^{42} and C^{58} could be ignored, yet trypsin activity is retained. In the native trypsin and the recombinant Exmt, there are three H-bonds exist in the catalytic triad. Nevertheless, only one H-bond formed between H^{57} and D^{102} in ExYmt, suggesting the H-bond between H^{57} and D^{102} may be essential for active trypsin. Furthermore, in Exmt and the native trypsin, one H-bond formed between V^{16} and D^{194}, whereas in ExYmt, the H bond formed between V^{16} and D^{189}. As a result, it could be deduced that the H-bond between V^{16} and the loop (D^{189}-D^{194}) is important for active trypsin. In comparison with the native trypsin, the recombinants contained fewer H-bonds and ion pairs, which indicated the profound influence of the N-terminal propeptide (Table S3).

After exploring the inhibitory mechanism of the propeptide, we evaluated the significance of the three factors identified above. First, we designed and constructed four different recombinants YVEImt, RVEFmt, RIRImt and YVERmt (Fig. 1C) and simulated their structures. Interestingly, introduction of the peptide YVEI or RVEF resulted in successful formation of disulfide bonds (C^{168}-C^{182}, C^{191}-C^{220}) and H-bond (V^{16}-D^{194}) (Fig. 4A,
4B). However, no activity of YVEImt and RVEFmt was detected (Fig. 4F), although secretory expression was observed (Fig. 5B). In contrast, the recombinants RIRImt and YVERmt exhibited clear activity (0.29 U ml\(^{-1}\) and 0.22 U ml\(^{-1}\), respectively) (Fig. 4F). Consequently, by comparing the results of \textit{in silico} simulations and \textit{in vitro} experiments, it could be concluded the formation of the H-bond between residues H\(^{57}\) and D\(^{102}\) is essential for trypsin activity. Subsequently, we further designed a novel peptide IVEF and fused it at the N-terminus of the native trypsin (Fig. 1C). By \textit{in silico} simulation, the recombinant IVEFmt formed two disulfide bonds (C\(^{168}\)-C\(^{182}\), C\(^{191}\)-C\(^{220}\)) and two H-bonds (H\(^{57}\)-D\(^{102}\), V\(^{16}\)-D\(^{189}\)) (Fig. 4E), which indicated the expressed protein would be active. As predicted, the recombinant IVEFmt was overexpressed (Fig. 5B) and a remarkable activity of 1.45 U ml\(^{-1}\) was obtained (Fig. 4F). Compared with parameters between the native trypsin and the recombinant IVEFmt, we found that although H-bonds and ion pairs decreased in the recombinant IVEFmt, \(\pi\)-interactions and the proportion of the random coil increased, which may give the recombinant IVEFmt more flexibility (Table S3).

In conclusion, by applying the \textit{in silico} simulations and \textit{in vitro} experimental, we explored the inhibitory mechanism of the propeptide to the \textit{S. griseus} trypsinogen and interpreted the crucial role of the proline. The disulfide bonds and H-bonds were found to be the major factors involved in trypsin regulation. In particular, the H-bond between residues H\(^{57}\) and D\(^{102}\) was crucial for trypsin activity. More importantly, by rationally designing an artificial propeptide IVEF, the titer of trypsin was increased by 6.71-fold. Furthermore, our work demonstrated that the combination of \textit{in silico} simulations and \textit{in vitro} experiments would be favorable to protein engineering.

ACKNOWLEDGMENTS

We thank Professor Byong Lee at Jiangnan University for his discussion and revision. This work was financially supported by the National High Technology Research and Development Program of China (863 Program, 2011AA100905), Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1135), the National Natural Science Foundation of China (31200020, 31000031), the Independent Innovation Program of Jiangnan University (JUSRP111A23), the Doctor Candidate Foundation of Jiangnan University (JUDCF10014), the 111 Project and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

REFERENCES

FIGURE CAPTIONS

**FIG 1** Schematic presentation of the construction of the recombinants with different propeptides. (A) Map of the integrated vector pPIC9k and the cleavage process of the α-factor signal peptide in *Pichia pastoris*. (B) Structure of the recombinants with different propeptides. YVEF was residual sequence at the N-terminus because of the digestion by endonucleases SnaBI and EcoRI, and cleavage of the α-factor signal peptide by peptidases KEX2 and STE13. (C) Structure of recombinant trypsin by fusing different designed propeptides.

**FIG 2** Comparison of the recombinants with different propeptides. Cultivation was performed in baffled shake flasks with 30 ml medium at 30 °C for 72h. 5 g l⁻¹ methanol was used as inducer was added to trigger the *AOX1* promoter. The results are the average of three individual experiments.

**FIG 3** Simulated structure of the *S. griseus* trypsinogen recombinants by NAMD molecular calculation. The crystal structure of *S. griseus* trypsin was used as the template. (A) Crystal structure of *S. griseus* trypsin; (B), (C), (D), (E) and (F) represent the simulated structures of the recombinants Exmt, ExPmt, ExYmt, APNPmt and Pmt, respectively.

**FIG 4** Comparison of the recombinants with different designed propeptide. (A), (B), (C), (D) and (E) represent the simulated structures of the recombinants YVEImt, RVEFmt, RIRImt, YVERmt and IVEFmt, respectively. (F) Secretory trypsin accumulated in cultured supernatants. Cultivation was performed in baffled shake flasks with 30 ml medium at 30 °C for 72h. 5 g l⁻¹ methanol was used as the inducer to trigger the *AOX1* promoter. The results are the average of three individual experiments.

**FIG 5** SDS-PAGE analysis of the purified recombinants with different propeptides. (A) The recombinants with native propeptide and its mutants by stepwise deletion. Lanes 1 to 9 represent the purified recombinants Exmt, ExAPNPmt, ExNPmt, ExPmt, ExYmt, APNPmt, Pmt and mt, respectively. (B) Mutants with different desinged propeptides, Lanes 1 to 5 represent the recombinants RIRImt, YVERmt, YVERmt, RVEFmt and IVEFmt, respectively. M, molecular markers (Thermo, USA). All the purified enzymes were concentrated by ultrafiltration.