G&I Genomics & Informatics

ORIGINAL ARTICLE

Structural Analysis of Recombinant Human Preproinsulins by Structure Prediction, Molecular Dynamics, and Protein-Protein Docking

Sung Hun Jung[1,2], Chang-Kyu Kim[3], Gunhee Lee[4], Jonghwan Yoon[4], Minho Lee[5]*

[1]Department of Biological Science, Sangji University, Wonju 26339, Korea, [2]Theragen Etex Bio Institute, Suwon 16229, Korea,
[3]WeGreen, Inc., Wonju 26493, Korea, [4]Department of Biomedicine & Health Sciences, Graduate School, The Catholic University
of Korea, Seoul 06591, Korea, [5]Catholic Precision Medicine Research Center, College of Medicine,
The Catholic University of Korea, Seoul 06591, Korea

More effective production of human insulin is important, because insulin is the main medication that is used to treat multiple types of diabetes and because many people are suffering from diabetes. The current system of insulin production is based on recombinant DNA technology, and the expression vector is composed of a preproinsulin sequence that is a fused form of an artificial leader peptide and the native proinsulin. It has been reported that the sequence of the leader peptide affects the production of insulin. To analyze how the leader peptide affects the maturation of insulin structurally, we adapted several *in silico* simulations using 13 artificial proinsulin sequences. Three-dimensional structures of models were predicted and compared. Although their sequences had few differences, the predicted structures were somewhat different. The structures were refined by molecular dynamics simulation, and the energy of each model was estimated. Then, protein-protein docking between the models and trypsin was carried out to compare how efficiently the protease could access the cleavage sites of the proinsulin models. The results showed some concordance with experimental results that have been reported; so, we expect our analysis will be used to predict the optimized sequence of artificial proinsulin for more effective production.

Keywords: leader peptide, molecular dynamics, preproinsulin, protein docking, structure prediction

Introduction

Human insulin, produced by beta-cells of the pancreatic islets, plays a critical role in the regulation of the metabolism of glucose [1]. Dysfunction of the synthesis or release of insulin may lead to diabetes mellitus [2]. Millions of people suffer from diabetes mellitus worldwide [3], and the most common medication to treat diabetes is insulin; so, a large number of studies on insulin have been done [4]. Insulin is first produced as an inactive protein, called preproinsulin. Preproinsulin, including a signal peptide, is a single, long protein. The chain evolves into proinsulin by cutting out the signal peptide. Then, proinsulin needs to be cleaved into insulin (an A chain and B chain) by removing the C-peptide linking the two chains [5].

Nowadays, recombinant DNA techniques enable us to produce insulin through biochemical processes using *Escherichia coli* [6]. However, the production of proinsulin by *E. coli* strains has to several drawbacks, such as low expression, difficulty in solubilizing the inclusion body, short half-life in the host cell, high proteolysis, and inefficient translation of the underlying coding sequences [7]. A new fusion protein system, which fuses an artificial leader peptide to the N-terminus of proinsulin, has been invented as a solution to these problems [8]. The specificity of cleavage and refolding rates are known to be dependent on the sequences of leader peptides, and different kinds of sequences have been proposed [6, 7, 9-11]. Thus, how the sequence of the artificial leader peptide affects the production of proinsulin and its modification into mature insulin needs to be investigated.

In this work, we used several kinds of computational approaches to find a better leader peptide. First, we

predicted the three-dimensional structures of fused proinsulins while changing the leader peptides. Although the 3D structure of proinsulin was determined recently [12] and although that of active insulin was determined long ago [13], the structure of fused proinsulin should be analyzed by prediction. Afterwards, the stabilities of the predicted structures were calculated by molecular dynamics (MD) simulation. MD is a method that simulates the movements of atoms and calculates their potential energies. Finally, the interaction of proinsulins with protease was evaluated by protein-protein docking. By comparing these structural analyses and experiments, we demonstrate that these structural analyses may contribute to determining whether a leader peptide results in efficient production of insulin.

Methods

Selection of artificial leader peptides

In previous work (patent WO 2004/044206 A1), a formula for constructing a leader peptide was proposed as follows: Met-Thr-Met-Ile-Thr···Lys(Arg). We selected 13 models, as shown in Table 1. Out of these 13 models, 5 have been used to generate experimental results in a previous work [7], and 7 are being introduced in this work.

Protein structure prediction

The full sequence (leader peptide of the model plus the native sequence of proinsulin) of each fused protein was used as an input for the structure prediction. Among various methods of 3D protein structure prediction, we used I-TASSER (Iterative Threading ASSEmbly Refinement) [14], which ranked as first in performance in recent community-based competitions [15, 16].
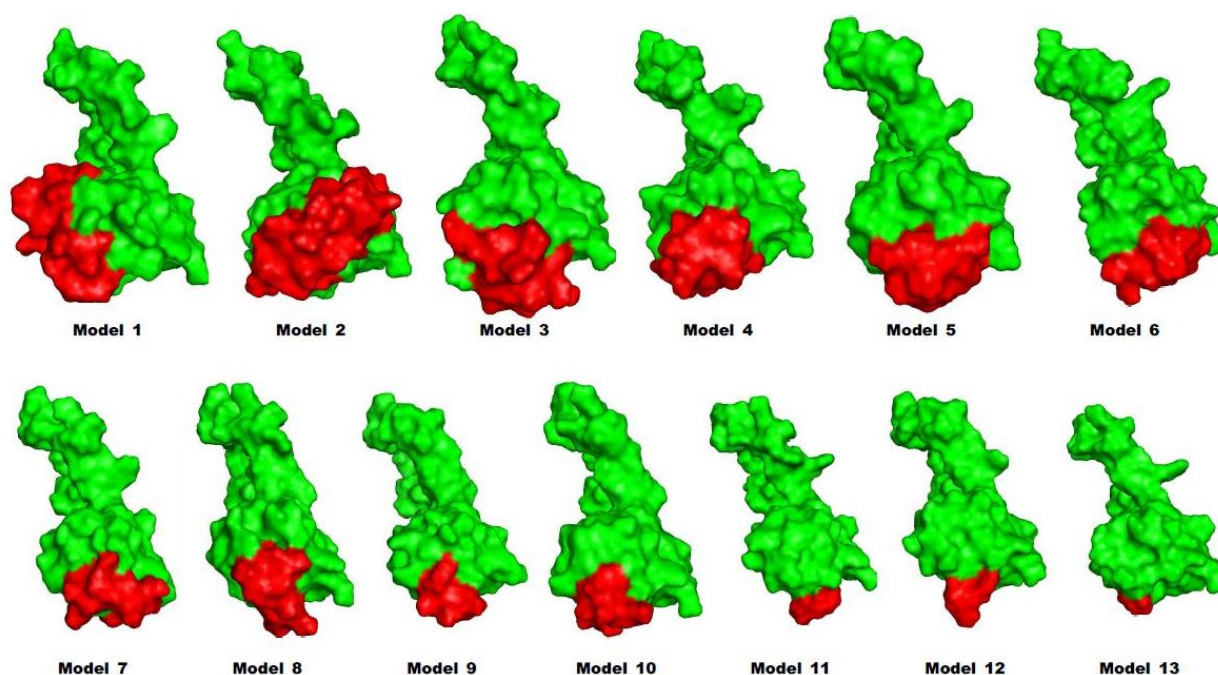
Table 1. Sequences of 13 artificial leader peptides used in the structural simulation and refolding yields

Model	Sequence of leader peptide	Refolding yield (%)
Model 1	MTMITNSPEISHHHHHHHHHHHQ LISEAR	68.7
Model 2	MTMITNSPEISHHHHHHHHHHHQ LISEAK	NA
Model 3	MTMITDSLAVVLQGSLQR	NA
Model 4	MTMITDSLAVVLQGSLQK	NA
Model 5	MTMITDSLAVVLQR	62.7
Model 6	MTMITDSLAVVLQK	NA
Model 7	MTMITDSLAR	57.9
Model 8	MTMITDSLAK	NA
Model 9	MTMITK	54.0
Model 10	MTMITR	42.2
Model 11	MK	NA
Model 12	MR	NA
Model 13	M	NA

Refolding yields were retrieved from our previous data [7]. The values of 5 models (models 1, 5, 7, 9, and 11) are reported. NA, not available.

Fig. 1. Predicted structures of proinsulin models. Structures of leader peptide regions and proinsulins are red and green, respectively.

MD simulation

The predicted structures were located in a cubic box of water; then, 3-ns MD simulations were carried out using GROMACS, ver. 5.1.2 [17]. Amber99sb was chosen as the force field [18]. Energies of the models were minimized using the steepest descent algorithm for 3 ns. The step size was 0.002 ps. The properties and stabilities of the models were evaluated based on the potential energy, total energy, and root-mean-square deviation of the atomic positions.

Protein-protein docking

Protein-protein docking between trypsin (PDB ID: 2PTN)

and recombinant proinsulins was performed using the InterEvDock server [19]. The InterEvDock server provides three kinds of scores: InterEvDock, FRODOCK [20], and SOAP_PP [21]. We chose the SOAP_PP score to evaluate the binding affinities, and docked poses were used to check whether two proteins were bound at the right position.

Results and Discussion

We predicted the 3D structures of 13 artificial models of proinsulin using I-TASSER (Fig. 1). As shown in Fig. 1, different leader peptides affected the overall structure of proinsulin. Especially, although the sequences of models 1
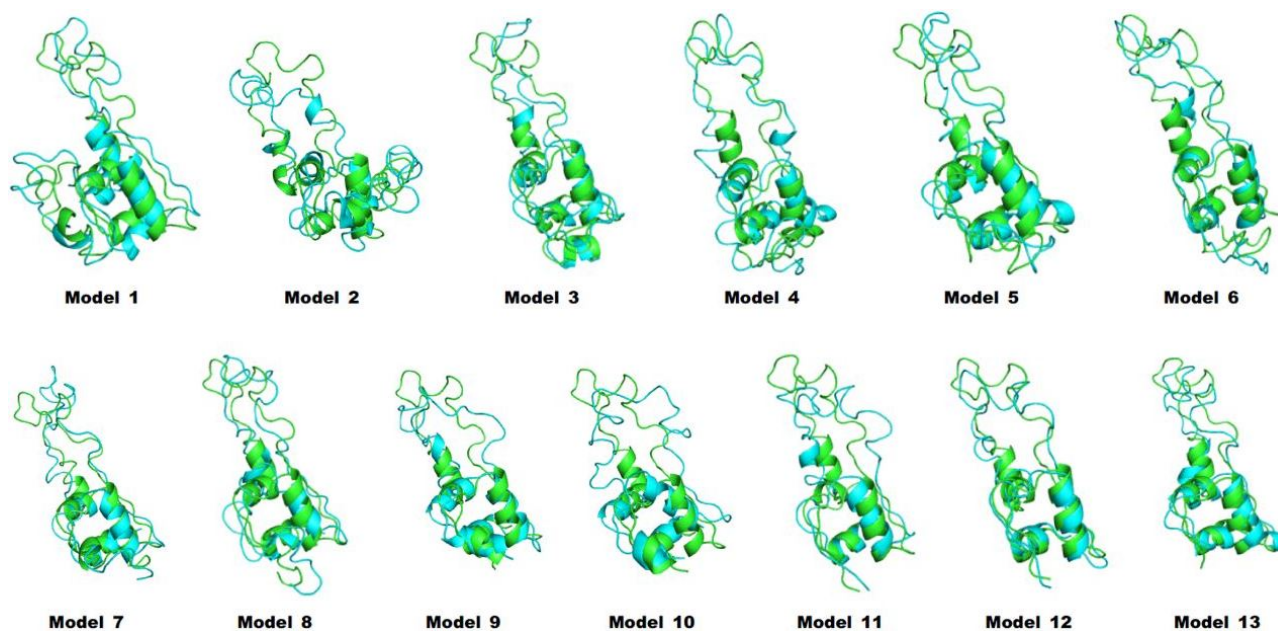

Fig. 2. Structural comparison of before (cyan) and after (green) the molecular dynamics simulation.
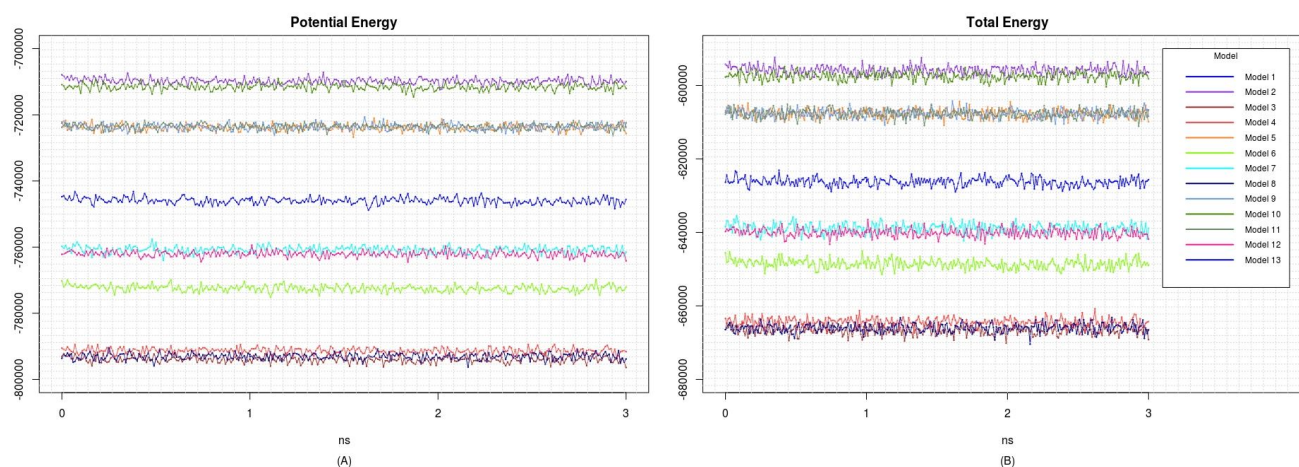

Fig. 3. Energy changes during molecular dynamics simulation of proinsulin models. (A) Potential energy. (B) Total energy.
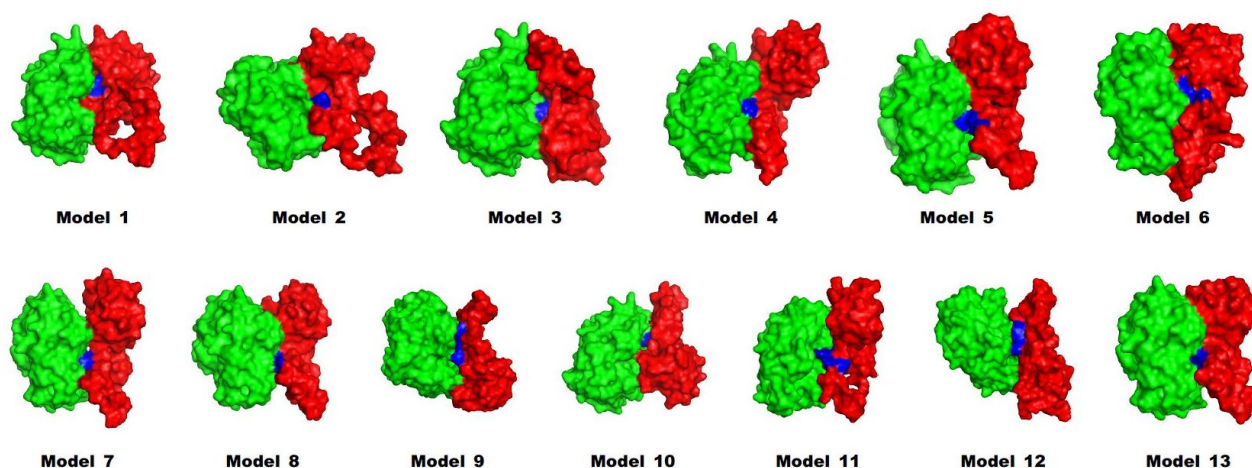
Fig. 4. Structures by docking between proinsulin models (red) and trypsin (green). Cleavage sites are in blue.

Table 2. Protein-protein docking energies

Model number	SOAP_PP score (kJ/mol)	
	Round 1	Round 2
Model 1	−12,021.42	−10,875.04
Model 2	−12,239.53	−11,204.59
Model 3	−11,580.68	−10,881.19
Model 4	−11,394.95	−10,484.51
Model 5	−11,280.89	−10,628.80
Model 6	−11,348.51	−10,292.90
Model 7	−11,596.85	−10,383.14
Model 8	−11,194.78	−10,107.16
Model 9	−10,933.44	−10,304.09
Model 10	−11,037.88	−10,266.31
Model 11	−10,690.47	−9,966.84
Model 12	−10,872.28	−9,782.31
Model 13	−10,614.70	−9,815.33

and 2 had only a 1-residue difference, the predicted positions of the leader peptides were quite different.

To refine the predicted structures, we observed the changes in structures by MD simulation at 3 ns to minimize energy. The structures after the MD simulation were superimposed onto the original structures by I-TASSER (Fig. 2). All potential energies of models were kept relatively stable (Fig. 3). The predicted structures were little refined; thus, they were naturally favorable. The total energy and potential energy of models 3 and 8 were the lowest. Among the models tested before (Table 1), no correlation between refolding yields and energies was observed.

Preproinsulin should be refolded and cleaved to become mature insulin. A previous work showed that trypsin could be used as an efficient protease in the maturation of insulin [7]. To investigate how the leader peptide affects the binding mode between proinsulin and trypsin, *in silico* docking was

carried out by InterEVDock [19].

Interestingly, in the docked structure (Fig. 4), the active sites (histidine 57, aspartate 102, and serine 195) of trypsin were located near lysine 64 and arginine 65 of proinsulin, which is the exact position where the first cleavage of proinsulin occurs to release chain A. Binding affinities were also predicted by InterEVDock (Table 2). Among all docked structures, models 1 and 2 were predicted to have the strongest binding affinities.

Because multiple steps of cleavage are needed for maturation, we performed additional protein docking of trypsin to model structures whose C-terminal chain after residue 64 (A chain) was cleaved out. As a result of this step, the preferable position for the active sites of trypsin were located near arginine 31 and arginine 32, which are known to form a cleavage site between the B and C chains. The binding affinities of the models were predicted to be strong (in descending order): model 2, 3, and 1. The two docking steps for artificial proinsulin models into trypsin revealed leader peptides that did not affect the order of cleavage sites, and models 1 and 2 are likely to be best accessed by trypsin in producing mature insulin. This coincides with previously tested refolding yields [7].

In summary, using 13 artificial models of leader peptides of proinsulin, we predicted the structures of fused proinsulins, the structures were refined by MD simulation, and protein-protein docking revealed binding modes between the artificial models and trypsin. The energies of the predicted structures of the models were not related to refolding yields, but the docking energies between the protease and the models showed some relation. We expect these analyses to provide basic information in a structural context for more effective production of insulin.

ORCID: Sung Hun Jung: http://orcid.org/0000-0001-666 6-7199; Chang-Kyu Kim: http://orcid.org/0000-0003-4753 -5320; Gunhee Lee: http://orcid.org/0000-0001-7142-907 2; Jonghwan Yoon: http://orcid.org/0000-0002-3399-096 6; Minho Lee: http://orcid.org/0000-0002-0168-9546

Authors' contribution

Conceptualization: CKK, ML
Data curation: SHJ, ML
Formal analysis: SHJ
Funding acquisition: ML
Methodology: SHJ, ML
Writing – original draft: SHJ, GL, JY, ML
Writing – review and editing: SHJ, CKK, ML

Acknowledgments

References

1. Wright JR Jr, Yang H, Hyrtsenko O, Xu BY, Yu W, Pohajdak B. A review of piscine islet xenotransplantation using wild-type tilapia donors and the production of transgenic tilapia expressing a "humanized" tilapia insulin. *Xenotransplantation* 2014;21:485-495.

2. Koolman J, Roehm KH. *Color Atlas of Biochemistry*. 2nd ed. Stuttgart: Thieme, 2005.

3. Shi Y, Hu FB. The global implications of diabetes and cancer. *Lancet* 2014;383:1947-1948.

4. Mohsin F, Craig ME, Cusumano J, Chan AK, Hing S, Lee JW, *et al*. Discordant trends in microvascular complications in adolescents with type 1 diabetes from 1990 to 2002. *Diabetes Care* 2005;28:1974-1980.

5. Chistiakov DA, Tyurina I. Current strategies and perspectives in insulin gene therapy for diabetes. *Expert Rev Endocrinol Metab* 2007;2:27-34.

6. Castellanos-Serra LR, Hardy E, Ubieta R, Vispo NS, Fernandez C, Besada V, *et al*. Expression and folding of an interleukin-2-proinsulin fusion protein and its conversion into insulin by a single step enzymatic removal of the C-peptide and the N-terminal fused sequence. *FEBS Lett* 1996;378:171-176.

7. Min CK, Son YJ, Kim CK, Park SJ, Lee JW. Increased expression, folding and enzyme reaction rate of recombinant human insulin by selecting appropriate leader peptide. *J Biotechnol* 2011;151:350-356.

8. Malik A, Jenzsch M, Lubbert A, Rudolph R, Söhling B. Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin. *Protein Expr Purif* 2007;55:100-111.

9. Chen JQ, Zhang HT, Hu MH, Tang JG. Production of human insulin in an *E. coli* system with Met-Lys-human proinsulin as the expressed precursor. *Appl Biochem Biotechnol* 1995;55:5-15.

10. Nilsson J, Jonasson P, Samuelsson E, Ståhl S, Uhlén M. Integrated production of human insulin and its C-peptide. *J Biotechnol* 1996;48:241-250.

11. Winter J, Lilie H, Rudolph R. Renaturation of human proinsulin: a study on refolding and conversion to insulin. *Anal Biochem* 2002;310:148-155.

12. Yang Y, Hua QX, Liu J, Shimizu EH, Choquette MH, Mackin RB, *et al*. Solution structure of proinsulin: connecting domain flexibility and prohormone processing. *J Biol Chem* 2010;285: 7847-7851.

13. Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, *et al*. The structure of 2Zn pig insulin crystals at 1.5 A resolution. *Philos Trans R Soc Lond B Biol Sci* 1988;319: 369-456.

14. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;5:725-738.

15. Zhang C, Mortuza SM, He B, Wang Y, Zhang Y. Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins* 2017 Oct 30 [Epub]. https://doi.org/10.1002/prot.25414.

16. Yang J, Zhang W, He B, Walker SE, Zhang H, Govindarajoo B, *et al*. Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. *Proteins* 2016;84 Suppl 1:233-246.

17. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, *et al*. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1:19-25.

18. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712-725.

19. Yu J, Vavrusa M, Andreani J, Rey J, Tufféry P, Guerois R. InterEvDock: a docking server to predict the structure of protein-protein interactions using evolutionary information. *Nucleic Acids Res* 2016;44:W542-W549.

20. Garzon JI, Lopéz-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernandez-Recio J, *et al*. FRODOCK: a new approach for fast rotational protein-protein docking. *Bioinformatics* 2009;25: 2544-2551.

21. Dong GQ, Fan H, Schneidman-Duhovny D, Webb B, Sali A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics* 2013;29:3158-3166.