Global Journal of Science Frontier Research: a Physics and Space Science
Volume 22 Issue 3 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Inspection of Remodeling Impacts of Domain Movements in Hydrosoluble Protein using Dual Artificial Intelligence Methods

By Katsuhiko Nishiyama

National Institute of Technology (KOSEN), Ishikawa College

Abstract- Remodeling impacts of domain movements in protease are of interest in many fields such as medical treatments, food processing, and bio-electronic devices. However, they are yet to be precisely explained. In this study, the remodeling effects in ficin were investigated via a deep neural network, genetic programming, and computer simulations. The replacement (Y113F) in ficin using domain movements exhibited a critical effect on the peptide compatibilities. Specifically, modification of amino acid allows the remodeling of the domain movements, and types of compatible peptides should be modulated by the remodeling. Moreover, the decision tree revealed important factors in peptides and ficin.

Keywords: protein, deep neural networks, genetic programming, molecular dynamics simulations.

GJSFR-A Classification: DDC Code: 006.3 LCC Code: Q335, DDC Code: 610.28 LCC Code: R853.C55

Inspection of Remodeling Impacts of Domain Movements in Hydrosoluble Protein using Dual Artificial Intelligence Methods

Katsuhiko Nishiyama

Abstract- Remodeling impacts of domain movements in protease are of interest in many fields such as medical treatments, food processing, and bio-electronic devices. However, they are yet to be precisely explained. In this study, the remodeling effects in ficin were investigated via a deep neural network, genetic programming, and computer simulations. The replacement (Y113F) in ficin using domain movements exhibited a critical effect on the peptide compatibilities. Specifically, modification of amino acid allows the remodeling of the domain movements, and types of compatible peptides should be modulated by the remodeling. Moreover, the decision tree revealed important factors in peptides and ficin.

Keywords: protein, deep neural networks, genetic programming, molecular dynamics simulations.

I. Introduction

Thermally produced movements of proteinous structure are perceived as the dominant factors involved in its biotic performance [1, 2, 3, 4], and their deliberate remodeling might lead to the discovery of an untapped resource. The majority of protein is organized via certain domain parts, and these domains are moving. Remodeling the domain movements would considerably influence molecular interactions between protein and ligand. Such interactions are of interest in many fields such as medical treatments, food processing, and bio-electronic devices [5, 6, 7, 8, 9], and their remodeling would provoke a breakthrough in these fields.

Hydrosoluble protein ficin separates certain peptides bound in sites near its active center, and its characteristics are beneficial to the improvement of transfusion safety, the development of novel antimicrobial therapeutics, and as a method for meat tenderization [10, 11, 12]. The structure of ficin is formed from two domains, and the binding areas of peptides are found in the two domains. Peptide binding is strongly affected by the domain movements, and therefore, remodeling the domain movements should affect the separation properties.

The two domains of ficin are joined by the β strand, which acts as a hinge. Modification of amino acid in the β strand would guide ficin to remodel the domain movements. It is desirable to be able to clearly explain the resulting effects of remodeling on peptide binding in more detail. Considering that ficin is a member of the protease family, reports on protease inhibitors play a large role in perceiving the effects [13, 14, 15, 16, 17]. However, remodeling effects of the domain movements in ficin to peptide bindings have not yet been clearly revealed.

Author: National Institute of Technology (KOSEN), Ishikawa College, Kitacyujo, Tsubata, Ishikawa 929-0392, Japan. e-mail: nisiyama@asagi.waseda.jp

The purpose of this research was to remodel the domain movements in ficin by modifying its local structure, and to clarify the effects of the remodeling to peptide bindings. Several possible combination patterns are possible for peptide, but the priority order of the search was unclear. The selection of combinations was conducted using deep neural networks (DNN) [18], and factors which enormously influence the peptide bindings were identified via genetic programming (GP) [19]. Characteristic features of various proteins were analyzed by molecular dynamics (MD) and docking simulations [20, 21, 13, 14, 22]. The moving structures of locally modified ficin were produced by MD simulation, while the binding states of peptides on the structures were assessed by docking simulations.

For the site near the center of the β strand between the two domains of ficin, the amino acid was replaced by another one (Y113F). For the first trial, a similar amino acid (F) was selected for this replacement. In the previous study, the binding properties between the tetrapeptides and ficin were investigated by DNN, GP, MD and docking simulations [23]. Tetrapeptide structures have multiple structures, and they often have weights comparable to the protease inhibitor E64. In this study, the behavior of tetrapeptides on the replaced structure of ficin (ficin_Y113F) was researched using the dual artificial intelligences (AIs) and the simulations.

II. Methods

The selection speed of candidate peptides was better for DNN than for GP, and thus DNN was utilized to select the candidates for this pursuit. The factors that influence the peptide bindings need to be clarified, but they cannot be derived using DNN. Therefore, the factors were extracted using GP.

The potential candidates were selected through the use of a convolutional neural network (CNN) [24]. CNN is a type of DNN, and its structure was constructed using the deep learning library TensorFlow [25]. CNN is actively used in fields of image processing, and is capable of treating a large amount of partially reliable data. Its structure is composed of seven layers, namely, input layer, first convolutional layer, first pooling layer, second convolutional layer, second pooling layer, fully connected layer, and output layer. A tetrapeptide comprises a combination of four amino acids, and the amino acids contain multiple atoms. For each amino acid, each atom was numbered. The numbers were set in the input layer. For more stable binding states between tetrapeptide and sites near active center of ficin_Y113F, the distance between atoms in the peptides and its active center was calculated. The binding characteristics of peptides were classified by the respective numbers of atoms nearer than 4.0 Åto its active center. The classification rule was as follows: NA1$\geqq$3 atoms, 3 atoms>NA2$\geqq$2 atoms, 2 atoms>NA3$\geqq$1 atoms, or 1 atoms>NA4. The successive numbers from 1 to 4 were assigned to NAα (α = 1–4), and were set in the output layer. Further, the first ten tetrapeptides were derived from the DNN for ficin [23], and their binding properties to ficin_Y113F were investigated by the simulations. Thereafter,

the first training data was created from the investigated results and training data of ficin [23]. Subsequently, the first DNN was constructed from the first training data, and consequently produced the second set of ten tetrapeptides. The properties of the second set of ten tetrapeptides were investigated in a similar manner, and second training data was made from the investigated results and the first training data. Consequently, the second DNN was constructed from the second training data, and subsequently produced the third ten tetrapeptides. Thus, this process was repeated, and various tetrapeptides were produced.

Highly Influencing factors to the peptide bindings were identified via GP. GP is an extended genetic algorithm (GA) technique, and it allows the exploration of the space of computer programs. The hyperparameters of GP were set with reference to a previous study [26], and evolutionary computing iteration was used to optimize the decision tree. The first time, entire areas of the decision tree were target of optimization. Thereafter, for the second and subsequent times, partial areas of the decision tree were randomly selected as optimization targets [27]. During the construction process of the decision tree, its size was expanded to satisfy the training data. Then, one of the successive numbers from 1 to 4 was set in the output of the decision tree, where the successive numbers corresponded to NAα (α = 1–4). For a tetrapeptide, four amino acids were labeled as amiβ, β = 1–4 from the N-terminus to the C-terminus. Each amino acid (amiβ) contained different types of atoms, and the number of each atom was represented by adding the atomic symbol at the end of amiβ. For instance, ami4S represented the number of sulfur atoms in amino acid of the C-terminus. The 20 attributes related to tetrapeptide were represented as amiβC, amiβN, amiβO, amiβS, and amiβAC (β = 1–4) [28]. 113 th amino acid of ficin was represented as ami113fic, and the number of each atom in ami113fic was represented by adding the atomic symbol at the end of ami113fic. The distance between domains of ficin was represented as domainDist, and it was assigned a numerical value of 0 for minimum distance, 1 for average distance, and 2 for maximum distance [23]. Further, 6 attributes related to ficin were represented as ami113ficC, ami113ficN, ami113ficO, ami113ficS, ami113ficAC, and domainDist. 19 operations (Add, Sub, Mul, Div, Fmod, Log, Log10, Sin, Cos, If, Equal, NotEqual, GT, GE, And, Or, Not, Fmod2, and Sqrt) and the 12 constants(a constant integer number CSTγ (γ) (γ = 0–9) and a boolean value TRUE (1), FALSE (0)) were inherited from our previous studies [23, 27, 29, 28]. Furthermore, the construction factors of the decision tree were chosen from the 26 attributes, 19 operations, and 12 constants.

The structures of ficin_Y113F with movements were produced by MD simulation through the use of the packaged software AMBER 12.0 [30]. The initial coordinate data of ficin_Y113F was produced by homology modeling through the use of the software MODELLER [31, 32]. For homology modeling, structural prediction was performed from sequence homology with known structures. The produced structure of ficin_Y113F was solvated with 11,843 TIP3P water molecules [33] inside a rectangular box. For the solvated system, its temperature was gradually increased from 5 to 300 K over

a period of 140 ps, and subsequently it was retained over a period of 100 ns until the precipitous structural changes vanished. Moreover, the ff03.r1 force field [34] was applied to manage MD simulation, and the hyperparameters were set by referencing the previous study [28]. In addition, the pressure and temperature were set using the Berendsen algorithm [35]. The long-distance electrostatic interactions were calculated using the particle mesh Ewald method [36]. The domain movements in ficin_Y113F were characterized as the distance variation between them for the period after the disappearance of precipitous structural changes. Thereafter, the average and snapshot structures were produced after the disappearance of the precipitous changes. The average structure was defined as $ficin\_Y113F\_ave$. Moreover, if the distance between the two domains of ficin_Y113F was at a maximum, the structure was defined as $ficin\_Y113F\_max$. Whereas, if the distance was at a minimum, the structure was defined as $ficin\_Y113F\_min$. The snapshot structures were made up of $ficin\_Y113F\_max$ and $ficin\_Y113F\_min$.

For $ficin\_Y113F\_ave$, $ficin\_Y113F\_max$ and $ficin\_Y113F\_min$, binding affinities of the peptides were evaluated through the use of AutoDock Vina [37]. This simulation software was executed by using the iterated local search global optimizer algorithm [38, 39] and the Broyden–Fletcher–Goldfarb–Shanno method [40], with its scoring function derived by machine learning. These methods contributed significantly to approximately two orders of size speed-up compared to the molecular docking software AutoDock 4 [41, 42, 43]. Further, the coordinate data of the tetrapeptides was built by using the LEaP module in AMBER 12.0 [30]. The exploration space over the surface of ficin_Y113F was restricted to the regions near the active center. The explorations were executed 10 times for each tetrapeptide, and the most stable state was chosen from them. For the most stable state, atoms in the tetrapeptide were counted only if they were found near the active center of ficin_Y113F. However, if there were certain members in the most stable docking state, the counted numbers were averaged.

III. Results

The temperature of the system was maintained at 300 K via MD simulations, and the precipitous structural changes of ficin_Y113F were surveyed at regular time intervals. For the structure of ficin_Y113F, the root-mean-square deviation (RMSD) between successive simulations was calculated to survey the precipitous changes(Figure S1). Consequently, the precipitous structural change was determined from the larger RMSD value, and they disappeared after 98,840 ps. The $ficin\_Y113F\_ave$ was estimated in the range of 99,901 to 100,000 ps, and the snapshot structures ($ficin\_Y113F\_max$ and $ficin\_Y113F\_min$) were derived from this range. Figure 1 shows the average structures of ficin and ficin_Y113F. The positional relation between the active center and the α–helix structure near the active center was changed by the replacement (Y113F). 10 tetrapeptides ((a) of Table S1) were selected as first candidates by using the data regarding ficin [23]. Thereafter, the candidates were docked to sites near active center of $ficin\_Y113F\_ave$, and the most stable styles were extracted.

Ficin is member of the family of cysteine protease, and the thiol (SH) group of its Cys25 can efficiently separate peptide. Consequently, the sulfur atom of Cys25 (S-Cys25) was considered as the active center. To discover the talented peptides for ficin_Y113F, the number of the atoms nearer than 4.0 Åto S-Cys25 in the most-stable docking state were estimated. In case of certain most-stable states, the estimated numbers were averaged. These results were represented by NAα ($\alpha = 1$–4), and are listed in Table S2. Within the first ten peptides, one member (Asn–Ser–Lys–Gln) was compatible with the sites near S-Cys25 of ficin_Y113F. Moreover, machine learning of the CNN was executed with the addition of these results, and the next ten candidates ((b) of Table S1) were guided. The characteristics of the ten peptides were estimated by docking simulations (Table S2), and one member (Arg–Ala–Val–Val) was compatible with the sites. This process was repeated eight times, and thereby, the characteristics of 80 peptides were obtained (Table S2), within which 17 members were compatible. The compatible 17 peptides were deposited in ave of Table 1. Simultaneously, 38 members were identified as especially incompatible peptides (Table S3). The docking simulations of the compatible 17 peptides (ave of Table 1) to the sites near S-Cys25 of two snapshot structures ($ficin\_Y113F\_max$ and $ficin\_Y113F\_min$) were executed, and the results are presented in Table S4. Within the 17 peptides, nine members (max of Table 1) were compatible to the sites near S-Cys25 of $ficin\_Y113F\_max$ and six members (min of Table 1) were compatible to those of $ficin\_Y113F\_min$. Furthermore, the five members (both of Table 1) were compatible to them of both $ficin\_Y113F\_max$ and $ficin\_Y113F\_min$.

To determine the important factors related to the peptide bindings, a decision tree was constructed from the results of 17 tetrapeptides to the three states of ficin_Y113F ($ficin\_Y113F\_min$, $ficin\_Y113F\_ave$, and $ficin\_Y113F\_max$)(Table 1)and those of 4 tetrapeptides to the three states of non-modified ficin ($ficin\_min$, $ficin\_ave$, and $ficin\_max$) [23]. The decision tree comprised 45 nodes, and its layout is indicated in Table 2 and Table 3. 11 attributes were included in the decision tree, and these were considered as the important factors. They were one ami1N, one ami1O, one ami1S, two ami2C, one ami3C, two ami3N, one ami113ficO, and two domainDist.
The acronyms were explained in Table S5.

IV. Discussion

The compatibilities of the 100 tetrapeptides to the sites near S-Cys25 of $ficin\_Y113F\_ave$ were revealed via a DNN, and MD and docking simulations. The 17 compatible peptides were found, and about 47 percent of them were not compatible for $ficin\_Y113F\_max$, and about 65 percent of them were not compatible for $ficin\_Y113F\_min$(Table 1). The compatibilities of the tetrapeptides to the sites near S-Cys25 of ficin_Y113F were grossly influenced by their domain movements. The five peptides were found to be compatible for the three states of ficin_Y113F(Table 1). The five peptides were not influenced by the domain movements of ficin_Y113F, and their structures would be an informative guide to drug development.

For average structure of non-modified ficin ($ficin\_ave$), only four peptides were compatible to the sites near S-Cys25 [23]. The replacement (Y113F) in ficin spread the compatible peptides within those of $ficin\_ave$. Within the four peptides, only one peptide was compatible for the structure of non-modified ficin when the distance between their domains was at a maximum ($ficin\_max$) [23]. In contrast, the replacement (Y113F) spread the compatible peptides within those of $ficin\_max$ as well. Further, all the four peptides were not compatible for the structure of non-modified ficin when the distance between their domains was at a minimum ($ficin\_min$) [23]. The replacement (Y113F) spread the compatible peptides within those of $ficin\_min$ as well as above two structures and had significant influence on the compatibilities between the peptides and ficin with domain movements. Moreover, modification of amino acid in the hinge structure of protease can enable us to remodel the domain movements, while the types of compatible peptides should be regulated by the remodeling. The collaboration of a DNN, and MD and docking simulations can be utilized as a tool in clarification of induced effects by modification of domain movements in protease.

The decision tree was constructed to satisfy the results for the three structures of ficin_Y113F ($ficin\_Y113 \quad F\_min$, $ficin\_Y113Fave$, and $ficin\_Y113Fmax$) (Table 1) and also for the three structures of non-modified ficin ($ficin\_min$, $ficin\_ave$, and $ficin\_max$) [23]. The tree was composed of 45 nodes, and the breakdown of its 45 nodes was 11 attributes, 25 operations, and 9 constants (Table 2 and Table 3). The 11 attributes included one ami1N, one ami1O, one ami1S, two ami2C, one ami3C, two ami3N, one ami113ficO, and two domainDist. Carbon atoms in second amino acid from the N-terminus of the peptide, nitrogen atoms in third amino acid from the N-terminus of it, and domain movements of the protein are the important factors governing the peptide compatibility. The three atoms (nitrogen, oxygen, and sulfur) in the N-terminus amino acid of the peptide, carbon atom in third amino acid from the N-terminus of it, and oxygen atom in 113 th amino acid in the protein must also be focused on. The 25 operations included four Add, three Sub, seven Mul, three Div, two Fmod, one Fmod2, three Sqrt, and two Sin. Many of the operations were the four basic arithmetic operations, and the peptide compatibility can be derived from comparatively elementary manner. The number of Sqrt was also relatively large, and numerical value could be be moderately changed. Furthermore, the logical operations were not included, and the derivation of the compatibility could not be expressed in binary form.

V. Conclusions

The remodeling effects of the domain movements in ficin were analyzed by using DNN, GP, MD, and docking simulations. The replacement (Y113F) in ficin had critical effect on the compatibilities between the peptides and ficin with domain movements. Further, modification of specific amino acid in protease can enable remodeling of the domain movements, and types of compatible peptides must be modulated by the remodeling. Moreover, the decision tree revealed that particular atom in particular amino acid in peptide and domain movements of the protease would be of central importance for the

peptide compatibility. Furthermore, the tree indicated the contribution of oxygen atom in 113 th amino acid in the protease to the compatibility. Thus, the analytical approach consisting of DNN, GP, MD and docking simulations can offer valuable lessons to understand the remodeling effects of the domain movements for some structures of protease, and the other structures should be analyzed for further research.

Acknowledgments

This study was supported by the National Institute of Technology (KOSEN).

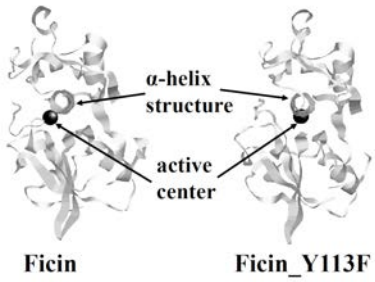

Figure 1: The average structures of ficin and ficin_Y113F.

Table 1: The compatible peptides.

ave	max	min	both
Arg–Ala–Val–Val			
Arg–Asp–Lys–Arg	◯	◯	◯
Arg–Gly–Arg–Gln	◯		
Arg–Gly–Asn–Ile	◯	◯	◯
Arg–Gly–Asn–Leu	◯	◯	◯
Arg–Gly–Asn–Lys	◯	◯	◯
Arg–Gly–Asn–Val	◯	◯	◯
Arg–Gly–Gln–Lys			
Arg–Gly–Pro–Ile	◯		
Arg–Gly–Pro–Leu			
Arg–Gly–Ser–Leu	◯		
Arg–Gly–Thr–Leu			
Arg–Gly–Val–Leu	◯		
Arg–Ser–Leu–Lys			
Asn–Ser–Lys–Gln		◯	
Lys–Gly–Leu–Gln			
Lys–Ser–Ile–Arg			

Table 2: Layout of first part of decision tree (α:Input, β:Output, γ:Node).

$\gamma 0$	Fmod2[$\alpha 0 = \beta$ of $\gamma 1$][β = Solution]
$\gamma 1$	Fmod[$\alpha 0 = \beta$ of $\gamma 2$][$\alpha 1 = \beta$ of $\gamma 33$][$\beta = \alpha 0$ of $\gamma 0$]
$\gamma 2$	Mul[$\alpha 0 = \beta$ of $\gamma 3$][$\alpha 1 = \beta$ of $\gamma 32$][$\beta = \alpha 0$ of $\gamma 1$]
$\gamma 3$	Fmod[$\alpha 0 = \beta$ of $\gamma 4$][$\alpha 1 = \beta$ of $\gamma 27$][$\beta = \alpha 0$ of $\gamma 2$]
$\gamma 4$	Mul[$\alpha 0 = \beta$ of $\gamma 5$][$\alpha 1 = \beta$ of $\gamma 6$][$\beta = \alpha 0$ of $\gamma 3$]
$\gamma 5$	ami3N[$\beta = \alpha 0$ of $\gamma 4$]
$\gamma 6$	Sub[$\alpha 0 = \beta$ of $\gamma 7$][$\alpha 1 = \beta$ of $\gamma 8$][$\beta = \alpha 1$ of $\gamma 4$]
$\gamma 7$	domainDist[$\beta = \alpha 0$ of $\gamma 6$]
$\gamma 8$	Mul[$\alpha 0 = \beta$ of $\gamma 9$][$\alpha 1 = \beta$ of $\gamma 20$][$\beta = \alpha 1$ of $\gamma 6$]
$\gamma 9$	Add[$\alpha 0 = \beta$ of $\gamma 10$][$\alpha 1 = \beta$ of $\gamma 18$][$\beta = \alpha 0$ of $\gamma 8$]
$\gamma 10$	Mul[$\alpha 0 = \beta$ of $\gamma 11$][$\alpha 1 = \beta$ of $\gamma 17$][$\beta = \alpha 0$ of $\gamma 9$]
$\gamma 11$	Mul[$\alpha 0 = \beta$ of $\gamma 12$][$\alpha 1 = \beta$ of $\gamma 16$][$\beta = \alpha 0$ of $\gamma 10$]
$\gamma 12$	Div[$\alpha 0 = \beta$ of $\gamma 13$][$\alpha 1 = \beta$ of $\gamma 15$][$\beta = \alpha 0$ of $\gamma 11$]
$\gamma 13$	Sin[$\alpha 0 = \beta$ of $\gamma 14$][$\beta = \alpha 0$ of $\gamma 12$]
$\gamma 14$	Cst2[$\beta = \alpha 0$ of $\gamma 13$]
$\gamma 15$	domainDist[$\beta = \alpha 1$ of $\gamma 12$]
$\gamma 16$	ami2C[$\beta = \alpha 1$ of $\gamma 11$]
$\gamma 17$	Cst6[$\beta = \alpha 1$ of $\gamma 10$]
$\gamma 18$	Sqrt[$\alpha 0 = \beta$ of $\gamma 19$][$\beta = \alpha 1$ of $\gamma 9$]
$\gamma 19$	Cst3[$\beta = \alpha 0$ of $\gamma 18$]
$\gamma 20$	Add[$\alpha 0 = \beta$ of $\gamma 21$][$\alpha 1 = \beta$ of $\gamma 25$][$\beta = \alpha 1$ of $\gamma 8$]
$\gamma 21$	Mul[$\alpha 0 = \beta$ of $\gamma 22$][$\alpha 1 = \beta$ of $\gamma 24$][$\beta = \alpha 0$ of $\gamma 20$]
$\gamma 22$	Sqrt[$\alpha 0 = \beta$ of $\gamma 23$][$\beta = \alpha 0$ of $\gamma 21$]

Table 3: Layout of second part of decision tree (α:Input, β:Output, γ:Node).

$\gamma 23$	Cst4[$\beta = \alpha 0$ of $\gamma 22$]
$\gamma 24$	Cst7[$\beta = \alpha 1$ of $\gamma 21$]
$\gamma 25$	Sqrt[$\alpha 0 = \beta$ of $\gamma 26$][$\beta = \alpha 1$ of $\gamma 20$]
$\gamma 26$	Cst5[$\beta = \alpha 0$ of $\gamma 25$]
$\gamma 27$	Sub[$\alpha 0 = \beta$ of $\gamma 28$][$\alpha 1 = \beta$ of $\gamma 31$][$\beta = \alpha 1$ of $\gamma 3$]
$\gamma 28$	Mul[$\alpha 0 = \beta$ of $\gamma 29$][$\alpha 1 = \beta$ of $\gamma 30$][$\beta = \alpha 0$ of $\gamma 27$]
$\gamma 29$	ami1O[$\beta = \alpha 0$ of $\gamma 28$]
$\gamma 30$	ami3C[$\beta = \alpha 1$ of $\gamma 28$]
$\gamma 31$	ami1S[$\beta = \alpha 1$ of $\gamma 27$]
$\gamma 32$	Cst2[$\beta = \alpha 1$ of $\gamma 2$]
$\gamma 33$	Add[$\alpha 0 = \beta$ of $\gamma 34$][$\alpha 1 = \beta$ of $\gamma 35$][$\beta = \alpha 1$ of $\gamma 1$]
$\gamma 34$	Cst6[$\beta = \alpha 0$ of $\gamma 33$]
$\gamma 35$	Div[$\alpha 0 = \beta$ of $\gamma 36$][$\alpha 1 = \beta$ of $\gamma 37$][$\beta = \alpha 1$ of $\gamma 33$]
$\gamma 36$	ami2C[$\beta = \alpha 0$ of $\gamma 35$]
$\gamma 37$	Div[$\alpha 0 = \beta$ of $\gamma 38$][$\alpha 1 = \beta$ of $\gamma 43$][$\beta = \alpha 1$ of $\gamma 35$]
$\gamma 38$	Add[$\alpha 0 = \beta$ of $\gamma 39$][$\alpha 1 = \beta$ of $\gamma 40$][$\beta = \alpha 0$ of $\gamma 37$]
$\gamma 39$	ami3N[$\beta = \alpha 0$ of $\gamma 38$]
$\gamma 40$	Sub[$\alpha 0 = \beta$ of $\gamma 41$][$\alpha 1 = \beta$ of $\gamma 42$][$\beta = \alpha 1$ of $\gamma 38$]
$\gamma 41$	ami113ficO[$\beta = \alpha 0$ of $\gamma 40$]
$\gamma 42$	Cst1[$\beta = \alpha 1$ of $\gamma 40$]
$\gamma 43$	Sin[$\alpha 0 = \beta$ of $\gamma 44$][$\beta = \alpha 1$ of $\gamma 37$]
$\gamma 44$	ami1N[$\beta = \alpha 0$ of $\gamma 43$]

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at

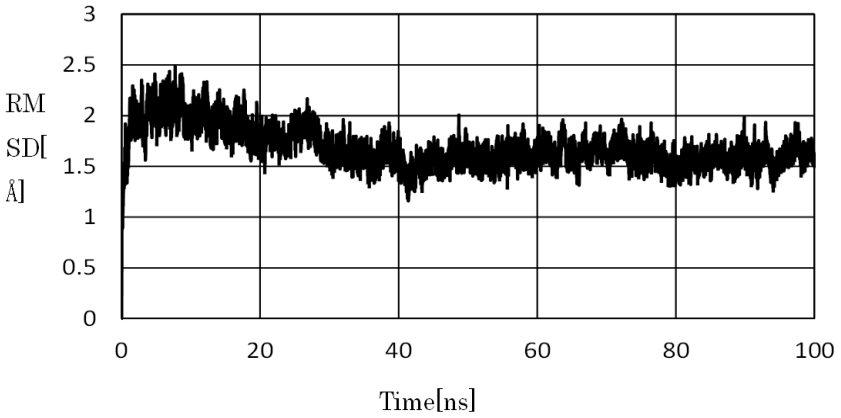

Figure S1.: RMSD between the successive simulations.

Table S1: The selected tetrapeptides.

(a)	(b)
Ser–Ser–Ile–Arg	Met–Gly–Lys–Gln
Arg–Cys–Lys–Arg	Lys–Cys–Arg–Gln
His–Cys–Arg–Ala	Gln–Ser–Arg–Asn
Arg–Asn–Arg–Asn	Arg–Ser–Lys–Lys
Thr–Asn–Arg–Arg	Gln–Ala–Arg–Asn
His–Gly–Leu–Gly	Arg–Ala–Val–Val
Arg–Gln–Arg–Arg	Lys–Gly–Lys–Val
Asn–Ser–Lys–Gln	His–Gly–Arg–Asn
Leu–Gly–Arg–Asn	Arg–Asp–Arg–Pro
Met–Asp–Ile–Arg	Lys–Ala–Lys–Lys

Table S2: Affinity of the 100 tetrapeptides in the active region of the average structure of locally modified ficin (Y113F).

Ser–Ser–Ile–Arg	NA3	Lys–Gly–Ile–Gln	NA4	Arg–Gly–Lys–Leu	NA2	Arg–Gly–Lys–Asn	NA4
Arg–Cys–Lys–Arg	NA2	Gln–Gly–Ile–Gln	NA4	Gln–Gly–Lys–Lys	NA4	Arg–Gly–Arg–Thr	NA4
His–Cys–Arg–Ala	NA3	Gln–Gly–Leu–Gln	NA4	Arg–Gly–Gln–Leu	NA2	Arg–Gly–Arg–Asn	NA2
Arg–Asn–Arg–Asn	NA2	Lys–Gly–Val–Gln	NA4	Arg–Gly–Asn–Leu	NA1	Val–Trp–Cys–Gly	NA4
Thr–Asn–Arg–Arg	NA3	Arg–Gly–Val–Leu	NA1	Arg–Gly–Ser–Lys	NA4	Met–Arg–Ile–Phe	NA4
His–Gly–Leu–Gly	NA4	Lys–Gly–Ile–Leu	NA4	Arg–Gly–Pro–Lys	NA3	Met–Gly–Glu–Arg	NA3
Arg–Gln–Arg–Arg	NA3	Arg–Gly–Lys–Met	NA3	Met–Gly–Val–Lys	NA4	Arg–Gly–Glu–Leu	NA4
Asn–Ser–Lys–Gln	NA1	Lys–Gly–Val–Leu	NA3	Arg–Gly–Ala–Leu	NA3	Lys–Gly–Gln–Arg	NA4
Leu–Gly–Arg–Asn	NA4	Arg–Gly–Val–Gln	NA2	Arg–Gly–Lys–Gln	NA3	Arg–Gly–Thr–Ile	NA3
Met–Asp–Ile–Arg	NA3	Arg–Gly–Arg–Gln	NA1	Arg–Gly–Gln–Gln	NA4	Ile–Gly–Asp–Arg	NA4
Met–Gly–Lys–Gln	NA3	Arg–Gly–Pro–Gln	NA4	Arg–Gly–Gln–Ile	NA4	Arg–Gly–Asp–Leu	NA4
Lys–Cys–Arg–Gln	NA4	Arg–Gly–Arg–Lys	NA3	Arg–Ser–Lys–Ile	NA3	Ile–Gly–Glu–Arg	NA4
Gln–Ser–Arg–Asn	NA4	Lys–Gly–Pro–Ile	NA2	Arg–Ser–Lys–Leu	NA4	Arg–Gly–Gln–Lys	NA1
Arg–Ser–Lys–Lys	NA2	Arg–Gly–Leu–Lys	NA4	Arg–Ala–Gln–Leu	NA2	Arg–Gly–Thr–Leu	NA1
Gln–Ala–Arg–Asn	NA3	Arg–Ala–Gln–Lys	NA4	Arg–Gly–Asn–Ile	NA1	Leu–Gly–Glu–Arg	NA3
Arg–Ala–Val–Val	NA1	Arg–Gly–Leu–Val	NA4	Gln–Asp–Lys–Arg	NA3	Arg–Gly–Ser–Ile	NA2
Lys–Gly–Lys–Val	NA4	Arg–Cys–Leu–Leu	NA4	Lys–Ser–Leu–Arg	NA3	Arg–Gly–Ser–Leu	NA1
His–Gly–Arg–Asn	NA4	Arg–Ser–Leu–Lys	NA1	Arg–Asp–Lys–Arg	NA1	Arg–Gly–Ile–Lys	NA3
Arg–Asp–Arg–Pro	NA4	Arg–Ser–Ile–Lys	NA4	Gln–Asp–Ile–Arg	NA4	Arg–Gly–Val–Arg	NA4
Lys–Ala–Lys–Lys	NA4	Arg–Gly–Leu–Ile	NA2	Lys–Ser–Ile–Arg	NA1	Arg–Ser–Gln–Arg	NA4
Arg–Gly–Val–Ile	NA3	Arg–Gly–Asn–Lys	NA1	Arg–Cys–Lys–Lys	NA2	Arg–Ser–Val–Arg	NA2
Arg–Gly–Pro–Leu	NA1	Arg–Gly–Lys–Pro	NA3	Arg–Gly–Thr–Lys	NA2	Arg–Ser–Ile–Arg	NA3
Lys–Gly–Pro–Gln	NA3	Arg–Ala–Leu–Lys	NA3	Arg–Gly–Asn–Val	NA1	Arg–Ser–Leu–Arg	NA3
Lys–Gly–Leu–Gln	NA1	Arg–Gly–Lys–Lys	NA3	Arg–Ala–Lys–Asn	NA4	Arg–Ser–Pro–Arg	NA2
Arg–Gly–Pro–Ile	NA1	Arg–Gly–Ile–Met	NA3	Arg–Cys–Arg–Asn	NA3	Arg–Gly–Lys–Arg	NA3

Table S3: The 38 especially incompatible tetrapeptides.

His–Gly–Leu–Gly	Gln–Gly–Lys–Lys
Leu–Gly–Arg–Asn	Arg–Gly–Ser–Lys
Lys–Cys–Arg–Gln	Met–Gly–Val–Lys
Gln–Ser–Arg–Asn	Arg–Gly–Gln–Gln
Lys–Gly–Lys–Val	Arg–Gly–Gln–Ile
His–Gly–Arg–Asn	Arg–Ser–Lys–Leu
Arg–Asp–Arg–Pro	Gln–Asp–Ile–Arg
Lys–Ala–Lys–Lys	Arg–Ala–Lys–Asn
Lys–Gly–Ile–Gln	Arg–Gly–Lys–Asn
Gln–Gly–Ile–Gln	Arg–Gly–Arg–Thr
Gln–Gly–Leu–Gln	Val–Trp–Cys–Gly
Lys–Gly–Val–Gln	Met–Arg–Ile–Phe
Lys–Gly–Ile–Leu	Arg–Gly–Glu–Leu
Arg–Gly–Pro–Gln	Lys–Gly–Gln–Arg
Arg–Gly–Leu–Lys	Ile–Gly–Asp–Arg
Arg–Ala–Gln–Lys	Arg–Gly–Asp–Leu
Arg–Gly–Leu–Val	Ile–Gly–Glu–Arg
Arg–Cys–Leu–Leu	Arg–Gly–Val–Arg
Arg–Ser–Ile–Lys	Arg–Ser–Gln–Arg

Table S4: Affinity of the 17 tetrapeptides in the active region of the snapshot structures of locally modified ficin (Y113F).

	max	min
Arg-Ala-Val-Val	NA3	NA3
Arg-Asp-Lys-Arg	NA1	NA1
Arg-Gly-Arg-Gln	NA1	NA3
Arg-Gly-Asn-Ile	NA1	NA1
Arg-Gly-Asn-Leu	NA1	NA1
Arg-Gly-Asn-Lys	NA1	NA1
Arg-Gly-Asn-Val	NA1	NA1
Arg-Gly-Gln-Lys	NA2	NA3
Arg-Gly-Pro-Ile	NA1	NA3
Arg-Gly-Pro-Leu	NA3	NA3
Arg-Gly-Ser-Leu	NA1	NA3
Arg-Gly-Thr-Leu	NA3	NA4
Arg-Gly-Val-Leu	NA1	NA3
Arg-Ser-Leu-Lys	NA2	NA3
Asn-Ser-Lys-Gln	NA3	NA1
Lys-Gly-Leu-Gln	NA3	NA3
Lys-Ser-Ile-Arg	NA2	NA3

Table S5: Explanation of the acronyms. (α: Input β: Output).

Add[α0][α1][β]	β = arithmetic addition of α0 and α1
Sub[α0][α1][β]	β = subtraction of α1 from α0
Mul[α0][α1][β]	β = arithmetic multiplication of α0 and α1
Div[α0][α1][β]	β = division of α0 by α1
Fmod[α0][α1][β]	β = remainder after division of α0 by α1
Log[α0][β]	β = natural logarithm of α0
Log10[α0][β]	β = common logarithm of α0
Sin[α0][α1][α2][β]	β = arithmetic multiplication of α0 and X, where X is sine of division of α1 by α2
Cos[α0][α1][α2][β]	β = arithmetic multiplication of α0 and X, where X is cosine of division of α1 by α2
If[α0][α1][α2][β]	β = if α0 = true then α1 else α2
Equal[α0][α1][β]	β = if α0 = α1 then true else false
NotEqual[α0][α1][β]	β = if α0 = α1 then false else true
GT[α0][α1][β]	β = if α0 > α1 then true else false
GE[α0][α1][β]	β = if α0 ≥ α1 then true else false
And[α0][α1][β]	β = logical multiplication of α0 and α1
Or[α0][α1][β]	β = logical addition of α0 and α1
Not[α0][β]	β = logical negation of α0
Fmod2[α0][β]	β = remainder after division of α0 by 4
Sqrt[α0][β]	β = square root of α0
ami γ	4 amino acids in a tetrapeptide, γ = 1-4 from the N-terminus to the C-terminus
ami γ C	the number of carbon atoms in ami γ
ami γ N	the number of nitrogen atoms in ami γ
ami γ O	the number of oxygen atoms in ami γ
ami γ S	the number of sulfur atoms in ami γ
ami γ AC	the number of aromatic carbon atoms in ami γ
ami113ficC	the number of carbon atoms in 113~th amino acid of ficin
ami113ficN	the number of nitrogen atoms in 113~th amino acid of ficin
ami113ficO	the number of oxygen atoms in 113~th amino acid of ficin
ami113ficS	the number of sulfur atoms in 113~th amino acid of ficin
ami113ficAC	the number of aromatic carbon atoms in 113~th amino acid of ficin
domainDist	the distance between domains of ficin (0:minimum, 1:average, 2:maximum)
CST δ	a constant integer number δ (δ = 0-9)
TRUE	a boolean value 1
FALSE	a boolean value 0
ficin_Y113F_ave	the average structure of ficin_Y113F
ficin_Y113F_max	the snapshot structure of ficin_Y113F when distance between its domains is at a maximum
ficin_Y113F_mini	the snapshot structure of ficin_Y113F when distance between its domains is at a minimum

References Références Referencias

1. D. W. Li, D. Meng, R. Bruschweiler, J. Am. Chem. Soc. 131 (2009) 14610.
2. A. Amadei, A. B. M. Linssen, H. J. C. Berendsen, PROTEINS 17 (1993) 412.
3. J. A. McCammon, B. R. Gelin, M. Karplus, P. G. Wolynes, Nature 262 (1976) 325.
4. J. Yguerabide, H. F. Epstein, L. Stryer, J. Mol. Biol. 51 (1970) 573.
5. H. Sun, H. Zhang, E. L. Ang, H. Zhao, Bioorg. Med. Chem 26 (2018)1275.
6. A. M. Bezborodov, N. A. Zagustina, Appl. biochem. biotechnol. 52 (2016) 237.
7. S. Raveendran, B. Parameswaran, S. B. Ummalyma, A. Abraham, A. K. Mathew, A. Madhavan, S. Rebello, A. Pandey, Food Technol. Biotechnol. 56 (2018) 16.
8. T. Tanii, T. Goto, T. Iida, K. M. Masahara, I. Ohdomari, Jpn. J. Appl. Phys. 40 (2001) L1135.
9. H. Itoh, A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida, K. Kinosita, Nature 427 (2004) 465.
10. D. R. Baidamshina, E. Y. Trizna, M. G. Holyavka, M. I. Bogachev, V. G. Artyukhov, F. S. Akhatova, E. V. Rozhina, R. F. Fakhrullin, A. R. Kayumov, Sci. Rep. 7 (2017) 46068.
11. B. C. Hill, C. A. Hanna, J. Adamski, H. P. Pham, M. B. Marques, L. A. W. III, Lab. Medicine 48 (2016) 24.
12. G. Walsh, Proteins: Biochemistry and Biotechnology., Chichester: John Wiley, 2002.
13. S. Matsuyama, A. Aydan, H. Ode, M. Hata, W. Sugiura, T. Hoshino, J. Phys. Chem. B 114 (2010) 521.
14. H. Ode, S. Neya, M. Hata, W. Sugiura, T. Hoshino, J. Am. Chem. Soc. 128 (2006) 7887.
15. P. Drabik, E. Politowska, C. Czaplewski, F. Kasprzykowski, L. Lankiewicz, J. Ciarkowski, Acta Biochim. Pol. 47 (2000) 1061.
16. X. Hu, S. Balaz, W. H. Shelver, J. Mol. Graphics Modell. 22 (2004) 293.
17. F. Osterberg, G. M. Morris, M. F. Sanner, A. J. Olson, D. S. Goodsell, PROTEINS 46 (2002) 34.
18. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Nature 529 (2016) 484.
19. J. R. Koza, Genetic Programming: on the Programming of Computers by Means of Natural Selection, MIT Press, 1992.
20. H. Yanagita, N. Yamamoto, H. Fuji, X. L. Liu, M. Ogata, M. Yokota, H. Takaku, H. Hasegawa, T. Odagiri, M. Tashiro, T. Hoshino, ACS Chem. Biol. 7 (2012) 552.
21. H. Yuki, T. Honma, M. Hata, T. Hoshino, Bioorg. Med. Chem. 20 (2012) 775.
22. K. Nishiyama, Chem. Phys. Lett. 647 (2016) 42.
23. K. Nishiyama, AIP Advances 11 (2021) 045325.
24. R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, Insights Imaging 9 (2018) 611.
25. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I.

Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Manje, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Vijegas, O. Vinyals, P.Warden, M.Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow. org (2015). URL https://www.tensorflow.org/

26. K. Nishiyama, AIP Advances 8 (2018) 125215.
27. K. Nishiyama, AIP Advances 10 (2020) 075102.
28. K. Nishiyama, AIP Advances 8 (2018) 055133.
29. K. Nishiyama, AIP Advances 9 (2019) 075001.
30. D. A. Case, T. A. Darden, T. E. CheathamIII, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Gotz, I. Kolossvjary, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, University of California, San Francisco, amber 12 Edition (2012).
31. A. Sali, T. L. Blundell, J. Mol. Biol. 234 (1993) 779.
32. M. A. Martjı-Renom, A. C. Stuart, A. Fiser, R. Sjanchez, F. Melo, A. Sali, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291.
33. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, J. Chem. Phys. 79 (1983) 926.
34. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, J. Comput. Chem. 24 (2003) 1999.
35. H. C. J. Berendsen, J. M. P. Postma, W. F. Gunsteren, A. DiNola, J. R. Haak, J. Chem. Phys. 81 (1984) 3684.
36. T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089.
37. O. Trott, A. J. Olson, J. Comput. Chem. 31 (2010) 455.
38. J. Baxter, Journal of the Operational Research Society 32 (1981) 815.
39. C. Blum, A. Roli, M. Sampels, Hybrid Metaheuristics: An Emerging Approach to Optimization., Springer, 2008.
40. J. Nocedal, S. J. Wright, Numerical optimization., Springer, 1999.
41. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, J. Comput. Chem. 30 (2009) 2785.
42. S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, A. J. Olson, Expert Opin. Drug Discovery 5 (2010) 597.
43. S. Forli, A. J. Olson, J. Med. Chem. 55 (2012) 623.