Journal of Molecular Modeling (2019) 25: 337
https://doi.org/10.1007/s00894-019-4225-y

ORIGINAL PAPER

Designing a less immunogenic nattokinase from *Bacillus subtilis* subsp. *natto*: a computational mutagenesis

Yoanes Maria Vianney[1] · Stanley Evander Emeltan Tjoa[1] · Reza Aditama[2] · Sulisyto Emantoko Dwi Putra[1]

Received: 19 December 2018 / Accepted: 9 October 2019 / Published online: 9 November 2019

Abstract

Nattokinase is an enzyme produced by *Bacillus subtilis* subsp. natto that contains strong fibrinolytic activity. It has potential to treat cardiovascular diseases. In silico analysis revealed that nattokinase is considered as an antigen, thus hindering its application for injectable therapeutic protein. Various web servers were used to predict B-cell epitopes of nattokinase both continuously and discontinuously to determine which amino acid residues had been responsible for the immunogenicity. With the exclusion of the predicted conserved amino acids, four amino acids such as S18, Q19, T242, and Q245 were allowed for mutation. Substitution mutation was done to lower the immunogenicity of native nattokinase. Through the stability of the mutated protein with the help of Gibbs free energy difference, the proposed mutein was S18D, Q19I, T242Y, and Q245W. The 3D model of the mutated nattokinase was modeled and validated with various tools. Physicochemical properties and stability analysis of the protein indicated that the mutation brought higher stability without causing any changes in the catalytic site of nattokinase. Molecular dynamics simulation implied that the mutation indicated similar stability, conformation, and behavior compared to the native nattokinase. These results are highly likely to contribute to the wet lab experiment to develop safer nattokinase.

Keywords B-cell epitopes · *Bacillus subtilis* subsp. *natto* · Bioinformatics · Immunogenicity · In silico mutagenesis

Abbreviation

3D	Three-dimensional
EFSA	European Food Safety Authority
GRAVY	Grand average of hydropathy
MSMS	Michel Sanner's molecular surface
MD simulation	Molecular dynamics simulation
SASA	Solvent-accesible surface area
R_g	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SVM	Support vector machine

Yoanes Maria Vianney and Stanley Evander Emeltan Tjoa contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00894-019-4225-y) contains supplementary material, which is available to authorized users.

✉ Sulisyto Emantoko Dwi Putra
emantoko@staff.ubaya.ac.id

[1] Faculty of Biotechnology, University of Surabaya, Raya Kalirungkut Street, Surabaya, East Java 60292, Indonesia

[2] Biochemistry Research Group, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Bandung Institute of Technology, Jalan Ganesha 10, Bandung, West Java 40132, Indonesia

Introduction

Nattokinase was first found by Sumi, and then it was purified and characterized by Fujita from the vegetable cheese natto, a soybean fermented food in Japan [1, 2]. Nattokinase is produced by *Bacillus subtilis* subsp. *natto* under the gene *apr*N, a serine protease composed of 275 amino acids, and well crystalized by Yanagisawa [3, 4]. This protein exhibits an antioxidant activity and fibrinolytic activity to degrade fibrin. Nattokinase also inhibits the angiotensin I converting enzyme, thus reducing blood pressure activity, with various pathways, highlighting the importance of this enzyme to treat cardiovascular diseases [5–8]. Other agents to treat thrombosis, such as urokinase, streptokinase, staphylokinase, and plasminogen activator, only functioning by converting plasminogen to plasmin, have a short half-life, and display uncontrollable adverse effects, such as accelerated fibrinolysis and hemorrhaging [8,

9]. The adverse effects of nattokinase oral treatment are not found in both animals and humans (2000 FU/capsule) [10, 11]. Thus, this protein is considered as a potential and natural supplement to prevent heart and cardiovascular diseases [12].

Oral consumption of nattokinase has some drawbacks, such as the absence of clear data on the status of the pharmacokinetics of nattokinase, which is the absorption or the intactness of the protein in the blood stream, and the dose is also considered high [8]. The formulation for injection provides one of the solutions to these drawbacks, especially to treat acute cardiovascular diseases. Unfortunately, nattokinase also displays immunogenicity effect [13]. Moreover, NSK-SD® product is also considered inflicting similar allergy risks compared to the other soy-derived products [14].

One solution is to conduct mutagenesis experiment to develop an entirely new type of nattokinase that displays low antigenicity. Computational analysis used in a preliminary study determines which amino acid residues are immunogenic and allow mutation to occur. Some strategies have been well demonstrated in arginine deaminase by Zarei et al. [15] and in peroxidase by Fattahian et al. [16]. Specifically, the B-cell epitopes of nattokinase, will be determined both continuously (linearly) and discontinuously (conformationally) by various tools [17–22]. Besides, the protein engineering of nattokinase has been conducted to increase the activity and to maintain the stability of the proteins using computational analysis, further emphasizing the usefulness of in silico analysis prior to experiment [23]. The nattokinase with less immunogenicity also suggests a new innovation of nattokinase type, thus broadens the current nattokinase market.

In the current study, both continuous and discontinuous B-cell epitopes of nattokinase were examined. Those selected amino acids were subjected to mutagenesis using rational design by considering the conservation of the residues and its 3D structure and stability. In addition, the optimization of codon for the production of mutated nattokinase in *Escherichia coli* as host was developed.

Materials and methods

Materials

The protein sequence of nattokinase was retrieved from UniProtKB (accession no. Q93L66). It comprised of 275 amino acids and was saved in FASTA format. The 3D structure of nattokinase was obtained by Yanagisawa et al. [4] by an X-ray crystallography method and retrieved from RCSB PDB under the code 4dww. HLA-DM 3D structure was retrieved from RCSB PDB (code: 2bc4) as a receptor for the molecular docking with nattokinase as ligand.

Antigenicity preliminary prediction of nattokinase

Nattokinase probability as an antigen was predicted with VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [24]. This web server measured the probability using Auto Cross Covariance method; thus, it is independent from the alignment method. The cut-off to consider the protein as antigen from bacteria is 0.4 by default.

Determination of B-cell epitopes

Various tools were used to predict both continuous and discontinuous B-cell epitopes. For the continuous B-cell epitopes, Bepipred 2.0 (http://www.cbs.dtu.dk/services/BepiPred/) was employed with a threshold of 0.5, specificity of 0.57158, and sensitivity of 0.58564. Bepipred 2.0 was based on random forest algorithm trained using various epitopes and non-epitopes of amino acids interpreted from antibody-antigen crystal structures [17]. ABCpred (http://crdd.osdd.net/raghava/abcpred/) was used to determine B-cell epitopes with a threshold of 0.8 as long as 16 mers [21]. SVMTriP (http://sysbio.unl.edu/SVMTriP/) was also used to predict the continuous 20-mer-long B-cell epitopes, which, in this case, achieved sensitivity of 80.1% and precision of 55.2% [25]. Flagged sequences were selected as epitopes. Another tool to predict continuous B-cell epitopes was Kolaskar and Tongaonkar antigenicity scale [22], accessible at Immune Epitope Database (IEDB) (http://tools.iedb.org/bcell/). This tool stored data of epitopes obtained from various experiments.

CBTOPE (conformational B-cell epitope prediction) (http://crdd.osdd.net/raghava/cbtope/) server was used to predict the discontinuous B-cell epitopes with an accuracy of more than 85%. By default, the threshold was − 0.3 and the residues were above 5 regarded as epitopes. Amino acid composition as an input for SVM algorithm was analyzed using CBTOPE [26]. BCEPred was also employed (http://crdd.osdd.net/raghava/bcepred/) to predict the discontinuous B-cell epitopes by evaluating the combination of various physicochemical properties of protein. The accuracy was 58.7% at the threshold of 2.38 by default [27]. Emini surface accessibility, accessed at IEDB Analysis Resource [28] was used to analyze the exposed surface of protein accessible to antibody. The threshold for the probability score was 1.

Determination of the mutated residues

Conservation of each residue of nattokinase had to be evaluated first before doing mutational analysis. Amino acid conservation was analyzed with Swiss-Model ExPASy using entropy method [29–31]. Entropy score below 2 was regarded as conserved residue. Thus, this residue should be excluded in the mutational analysis as not to disturb the structure and the stability of the protein. Residues were not located on the

protein surface, and predicted Emini surface accessibility was excluded. Residues allowed for mutation must conform to three of the predicted continuous epitopes and one of the predicted discontinuous epitopes.

In silico mutagenesis of nattokinase

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) was used to perform the mutagenesis using substitution mutation. 3D structure protein (4dww) was subjected to the mutagenesis analysis. This tool could predict the change of protein stability from the result of single point mutation with the help of Gibbs' free energy difference [32]. The formula applied was ΔG (mutein) – ΔG (wild-type protein), and positive free energy difference indicating that the mutation was permissible. The condition for mutation was set to correspond with the human physiological condition, with the pH of 7.4 and the temperature of 37 °C. For every acceptable single mutation, the antigenicity score was reevaluated with the help of VaxiJen web server independently. Then, the suggestible mutations from various residues were re-analyzed with VaxiJen in combined techniques to decide whether multiple mutations demonstrate a lower antigenicity probability score compared to a single mutation In addition, dDFIRE was used to evaluate mutant stability compare to native structure [33].

3D modeling and validation of the mutein

The determined mutein was then 3D modeled by PHYRE2 protein fold recognition server (http://www.sbg.bio.ic.ac.uk/phyre2) [34] and I-TASSER for comparison purpose. Both best 3D mutein models were compared with the native nattokinase and validated with molprobity (http://molprobity.biochem.duke.edu) generating Ramachandran plot that analyzed the favored and permissible regions for the dihedral angle of the protein backbone of amino acid residues [35, 36]. ProQ3/ProQ3D (http://proq3.bioinfo.se/pred/) was used to assess the quality of the 3D protein structure by single model method. ProQ3 was based on the combination of ProQ2, ProQRosCen (centroid model), and ProQCenFA (full atom model). ProQ3D is the deep-learning version of the ProQ3 that generates better correlation regarding the estimation of protein quality [37]. Moreover, the protein model quality was also analyzed with VERIFY3D [38, 39] and ERRAT [40] from SAVES 5.0 web servers (http://servicesn.mbi.ucla.edu/SAVES/).

The native and mutated nattokinase physicochemical properties and stability, such as Pi, molecular weight, instability index, aliphatic index, hydrophobicity by GRAVY, and half-life estimation at mammalian reticulocytes were analyzed with ProtParam. The secondary structure of protein was confirmed in SOPMA secondary structure prediction (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). Catalytic site and binding pocket predictions of native and mutated nattokinase were compared using tools in the PHYRE2 based on catalytic site atlas [41] and fpocket2 program [42].

Molecular dynamics simulation

MD simulation was used to predict the fluctuation and stability of amino acid residues in the 3D structure protein. Native nattokinase and the 3D models of mutein generated by PHYRE2 were subjected to MD simulation. Protonation state at pH 7.4 from native and mutated nattokinase was done with PROPKA 3.1 [43]. Both structures were solvated with TIP3P water [44] and both system were neutralized with counter ion; hence, the dimension of the both system were 7.39 mm × 7.392 mm × 7.39 mm. Gromacs 2018 with OPLS force field [45, 46] was employed. Minimization was done for 50 ps, continued by NVT and NPT equilibration for 100 ps. Simulation temperature was kept at 310 K using V-rescale thermostat [47]. The MD simulation time was 50 ns with the 2 fs of time step and 10 Å of cutoff. VMD program was used to analyze the trajectory simulation [48]. Several parameters were computed, such as RMSD, RMSF of all residues, Rg, and SASA.

Codon optimization for in silico cloning

Reverse translation web server tool (http://www.bioinformatics.org/sms2/rev_trans.html) was employed to predict and to optimize the DNA sequence of the mutated nattokinase gene using *E. coli* as expression host. Evaluation of the appropriate codon for *E. coli* expression was done in GenScript rare codon analysis (https://www.genscript.com/tools/rare-codon-analysis) using codon adaptation index (CAI) with an ideal value of 0.8-1, GC content (ideal at 30–70%), codon frequency distribution (CFD) (ideal if below 30%), and detection of negative regulatory elements plus repeat sequences. The restriction site prediction of the optimized linear DNA sequence was employed (http://nc2.neb.com/NEBcutter2) to suggest the restriction enzymes which should not be used if this gene was cloned in expression vector, or to consider changing the DNA sequence with Wooble hypothesis concept to remove the recognition sites from the gene.

Results and discussion

Nattokinase antigenicity and B-cell epitopes determination

Determination of nattokinase antigenicity by VaxiJen web servers acted as a preliminary step to decide whether or not the mutagenesis should be done to reduce its immunogenicity. Indeed, the VaxiJen antigen probability score of nattokinase sequence (Q93L66) was 0.7391, far above the 0.4 threshold. Thus, it raised an importance to develop a less-immunogenic

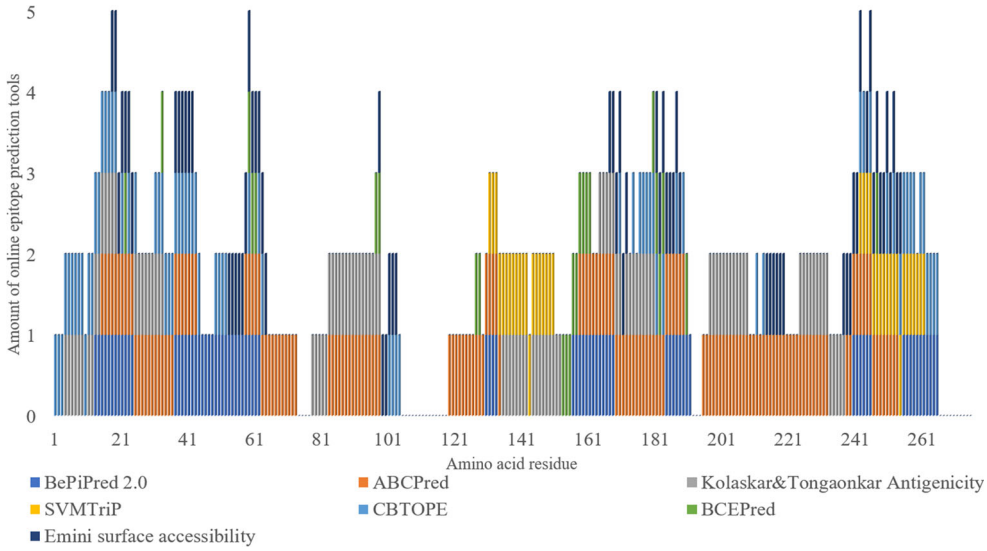

Fig. 1 Illustration of the determination of the continuous and discontinuous B-cell epitopes together with the surface accessibility analysis to reveal which residues should be mutated

nattokinase by mutagenesis. In silico analysis aided various researchers to determine both continuous and discontinuous B-cell epitopes [15, 16, 49–53]. In general, four tools were used to predict continuous B-cell epitopes, two web servers to predict the discontinuous B-cell epitopes, and Emini surface accessibility assessment. The B-cell epitopes were overlapped by each of the conformational B-cell epitopes. The epitopes were determined based on the requirements that the residues must comply with three out of four continuous B-cell epitopes prediction, and one out of two discontinuous B-cell epitopes

Table 1 In silico mutagenesis and stability analysis of nattokinase by single substitution mutation technique. The mutation condition was set at 37 °C for the temperature and at the pH of 7.4

	S18		Q19		T242		Q245	
Alternative residue	ΔΔG (kcal mol^{-1})	VaxiJen score	ΔΔG (kcal ·mol^{-}1)	VaxiJen score	ΔΔG (kcal mol^{-}1)	VaxiJen score	ΔΔG (kcal mol^{-}1)	VaxiJen score
V	0.78	0.7498	0.68	0.7195	− 0.67		0.21	0.7234
L	0.97	0.7459	2.14	0.7182	− 0.98		1.46	0.7167
I	0.88	0.7484	*0.99*	*0.7163*	− 0.45		0.73	0.7162
M	0.79	0.7408	0.27	0.7258	− 0.29		− 0.12	
F	0.33	0.7368	0.32	0.7232	− 0.19		0.5	0.7111
W	0.74	0.7282	0.7	0.7284	− 0.28		*0.28*	*0.7098*
Y	0.56	0.7303	0.75	0.7342	*0.05*	*0.7266*	0.24	0.7238
G	0.19	0.7516	0.22	0.7344	− 1.6		− 0.81	
A	0	0.7422	− 0.09		− 0.44		0.11	0.7316
P	0.09	0.7317	− 0.8		− 1.78		− 1.01	
S	0	0.7391	0.12	0.7391	0.06	0.7354	0.24	0.7379
T	0.12	0.7447	− 0.04		0	0.7391	0.25	0.7364
C	-0.1		1.03	0.745	− 0.32		0.76	0.7292
H	0.35	0.7266	0.07	0.7471	− 1.65		− 0.33	
R	0.72	0.7303	0.32	0.7388	0.12	0.7419	0.2	0.7428
K	0.97	0.7336	0.25		− 1.01		− 0.19	
Q	0.74	0.7342	0	0.7391	− 1.04		0	0.7391
E	0.81	0.7322	0.62	0.7377	0.25	0.7364	0.18	0.7414
N	0.71	0.7271	− 0.01		− 0.89		− 0.32	
D	*0.98*	*0.7255*	0.58	0.7528	0.22	0.7304	− 0.38	

The italicized numbers denote the amino acid that will be used to substitute native residue of nattokinase, which is combination of positive ΔΔG and the lowest VaxiJen score for all predicted amino acids

Table 2 Determination of the lowest antigenicity score by multiple mutagenesis of nattokinase. Italicized entry indicates that the mutein gave lowest VaxiJen score and was subjected for further analyses

	VaxiJen score
Native nattokinase	0.7391
Mutated protein	
S18D; Q19I; T242Y; Q245W	*0.6560*
S18D; Q19I; T242Y	0.6839
S18D; Q19I; Q245W	0.6777
Q19I; T242Y; Q245W	0.6653
S18D; Q19I	0.7070
S18D; T242Y	0.7133
S18D; Q245W	0.6962
Q19I; T242Y	0.7038
Q19I; Q245W	0.687
T242Y; Q245W	0.6987

prediction. Nonetheless, the surface accessibility assessment was compulsory [15]. It was the exposed residues that usually acted as functional epitopes, which was well demonstrated by bauA from *Acinetobacter baumanii* [54]. The selected epitopes can be seen schematically in Fig. 1. It can be seen clearly that there were 5 residues (residue no. 18, 19, 59, 242, 245) that comply with five out of the seven web server tools. Yet, residue number 59 was excluded since this residue only conformed to the two out of the four continuous B-cell prediction tools. Linear B-cell epitopes prediction tools generated better accuracy than the discontinuous B-cell epitopes prediction. Several studies even only used linear B-cell tools for the prediction for the epitopes [46, 55, 56].

Entropy method (ExPaSy) was one way to determine the conserved residues. Less entropy indicated more conservancy. Besides, the conserved residues usually contributed to stability of the protein [15]. These residues must be excluded as not to reduce the stability of the protein. Nevertheless, the four putative B-cell epitopes were all not suggested as conserved (Supplementary file 1). Based on all these requirements, S18, Q19, T242, and Q245 were determined and continued for the mutagenesis analysis.

In silico mutagenesis of nattokinase

In silico mutagenesis was done with the help of I-Mutant 2.0. Substitution mutation for point mutation to the other 19 amino acids could be performed using this tool to predict their stability through the differences of Gibbs energy [32]. The choice of substituted amino acids were based on two rationales, the increase of the difference of Gibbs energy and the largest decrease of VaxiJen score, which indicated reduction in antigenicity. An increase in the Gibbs energy differences indicated a rise in stability. Thus, only the residues showing positive values from the difference of Gibbs energy after substitution mutation were used. Combined with VaxiJen, the antigenicity of the protein was reevaluated after substitution mutation with each residue. In Table 1, the mutation result was described in terms of differences of Gibbs energy and VaxiJen antigenicity score. Accordingly, the decision of which residues that should be mutated is the result of combination of positive values of difference of Gibbs energy and the largest decrease of VaxiJen score, which gave preliminary prediction of retainment or rise in stability and a decrease of immunogenicity in the mutated protein. With a single substitution mutation of amino acid, selected mutations were S18D, Q19I, T242Y, and Q245W. All the changed four amino acids were polar. Surface antigenic residues are likely to be hydrophilic amino acid or aromatic [15]. With the exception of S18D, all the residues were changed into non-polar amino acid. These changes reduced both protein surface accessibility and the immunogenicity as shown by Ramya and Pulicherla [57] in the in silico mutation of asparaginase for deimmunization.

Although Zarei et al. [15] showed that mutation of single amino acid residues generated lower immunogenicity scores than multiple mutations, some studies indicated that the combinations of multiple mutations on epitopes dramatically reduced immunogenicity and IgE-binding activity [16, 58]. Thus, a series of combination mutation with its antigenicity score can be seen in Table 2. The combinations of all four substitution (S18D, Q19I, T242Y, and Q245W) mutations in one protein generated the lowest antigenicity. Protein conformational free energy calculation confirms that the mutation of nattokinase does not cause structural destabilization. The energy obtained from dDFIRE gives score − 588.11 and − 584.99 for mutated and native nattokinase, respectively.

In addition to the antigenicity score of the mutated protein, the antigenicity scores for each surface exposed area determined by Emini surface accessibility assessment were also evaluated. The exposed areas of 237–242 and 244–254 residues were unified because the areas did not meet the requirement of an epitope. This mutation dramatically reduced the antigenicity score of the exposed residues (Table 3).

Table 3 Antigenicity score of the epitopes that was surface exposed, before and after mutation

Before mutation	VaxiJen score	After mutation	VaxiJen score
18–24 (SQGYTGS)	1.3814	18–24 (DIGYTGS)	− 0.1772
237-254 (KHPTWTNAQVRDRLESTA)	0.9024	237–254 (KHPTWYNAWVRDRLESTA)	0.175

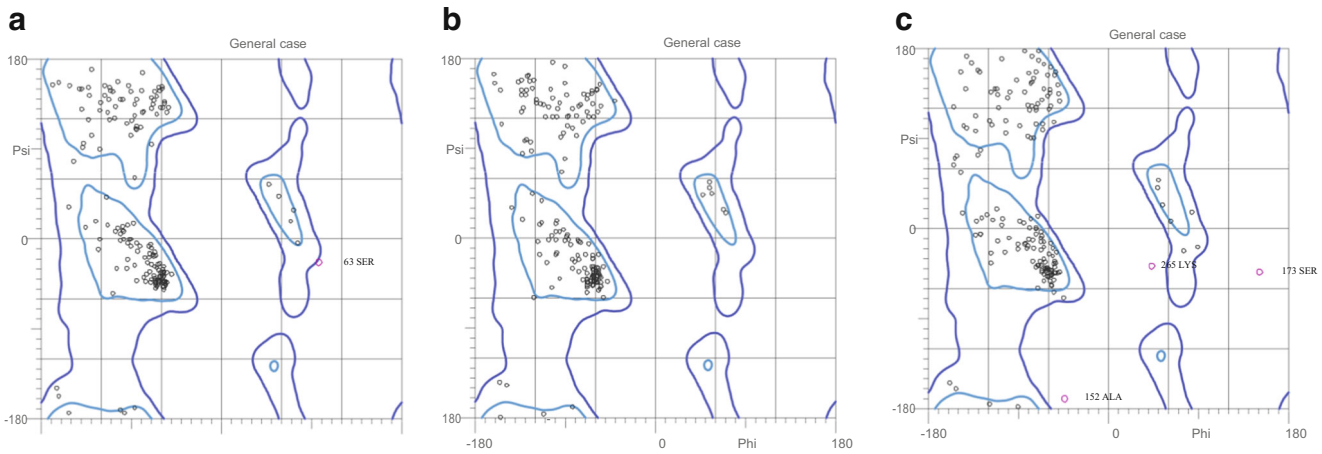

Fig. 2 Ramachandran plot for general amino acid residues of **a** nattokinase, **b** mutated nattokinase generated by PHYRE2, and **c** mutated nattokinase generated by I-TASSER

Mutein validation, physicochemical, and stability analysis

The recommended mutein was then 3D modeled with the two web-server tools, PHYRE2, and I-TASSER. Validation analyses were employed to compare which web servers generated best geometric quality of the 3D-modeled protein. Figure 2 showed the Ramachandran plot for the general amino acid residues for the native nattokinase and mutein generated by both tools. The complete Ramachandran plot can be seen in supplementary file 2. The 3D structure of the mutein generated by PHYRE2 generated a better geometry related to the torsional angle between two adjacent residues than the mutein generated by I-TASSER and native nattokinase. There were no outlier residues in the mutein generated by PHYRE2, compared to 5 outlier residues given by the 3D mutein modeled by I-TASSER (Table 4). Moreover, the ProQ3D score of the mutein generated by PHYRE2 was higher than the one generated by I-TASSER. Compared to the native structure, higher ProQ score indicated better overall quality of the model [37].

VERIFY3D can be used to analyze the 3D model using its constituent component sequences (1D) and its environment [59]. More than 90% of all 3 protein model residues met the standard of the 1D-3D score which was ≥ 0.2 (Table 4). The overall quality factor for non-bonded interactions within three types of atoms: carbon (C), nitrogen (N), and oxygen (O) was analyzed using ERRAT. The error function versus the position of every nine-residue sliding windows was plotted [40]. Both ERRAT scores from the mutein were lower than the native

Table 4 Quality assessment and validation of the mutated nattokinase 3D model

	Native protein	Mutein (generated by PHYRE2)	Mutein (generated by I-TASSER)
Ramachandran analysis			
Favored region	265	264	240
Allowed region	6	9	28
Outlier region	1	0	5
ProQ3D score	0.822	0.817	0.804
Average 1D-3D score[a]	91.61 % (pass)	99.64 % (pass)	99.64 % (pass)
ERRAT overall quality factor A	95.6938	90.6367	89.5131

[a] The requirement to pass the standard average 1D-3D score is 80%

Table 5 Physicochemical properties, stability, and hydrophilicity of the native and mutated 3D model nattokinase

	Native protein	Mutein (generated by PHYRE2)
Isoelectric point	6.65	6.3
Molecular weight	27612.60	27720.8
Instability index[a]	25.62	25.98
Aliphatic index[b]	83.05	84.47
GRAVY[c]	0.039	0.088
Estimated half-life	4.4 h	4.4 h
Secondary structure prediction		
Alpha-helix	23.64%	22.18%
Beta-turn	12.36%	10.91%
Extended strand	22.55 %	24 %
Random coil	41.45 %	42.91 %

[a] Value below 40 predicts that the protein is stable

[b] Increase in value of the aliphatic index is positively linear with the increase of thermostability

[c] Protein is considered more hydrophobic if the value is bigger positive number

[d] The half-life of the protein was predicted with N-terminal rule using mammalian reticulocytes as model in vivo

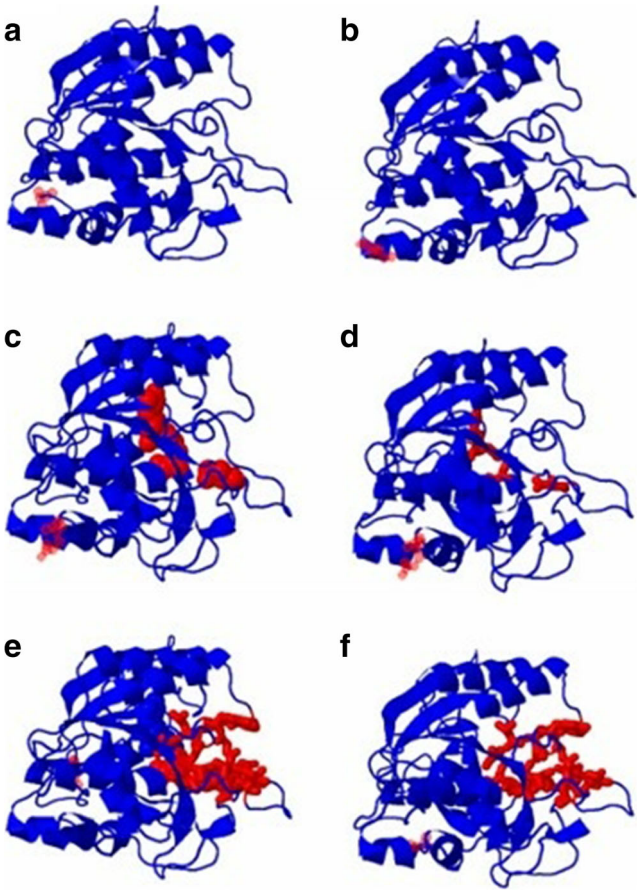

Fig. 3 3D modeling of native nattokinase (**a**) and mutated nattokinase (**b**); catalytic site prediction of native nattokinase (**c**) and mutated nattokinase (**d**); and binding pocket of native nattokinase (**e**) and mutated nattokinase (**f**)

protein related to the resolution of the model [60]. However, both mutein ERRAT scores were 50 above the threshold in general [61]. Still, the ERRAT score of the mutein model generated by PHYRE2 was above 90%, and this score was higher than the ERRAT score of the mutein model generated by I-TASSER. The ERRAT score above 90% showed the tendency of the protein model with the maximum resolution of 3 Å [40]. Overall, these validation results indicated that the 3D model of the mutein built by PHYRE2 was more proper. Muteins built by PHYRE2 were subjected to the various stability and the analysis of the physicochemical properties.

Regarding protein stability by instability index, the mutein was considered stable although the index was slightly higher than the native. The instability index algorithm was generated from the empirical data of the stability of protein and dipeptide composition [62]. Higher aliphatic index signified higher thermostability due to the hydrophobic interaction within nonpolar groups from the aliphatic side chain of amino acids [63], in which mutein aliphatic index was higher than the native one. With hydropathy analysis by GRAVY, the mutein was more hydrophobic than native nattokinase. Lower GRAVY value indicated more residues were to be exposed on the surface of the protein [53]. Hydrophilicity is one method that correlates with the antigenicity of a protein. Regarding the secondary structure, there was a reduction in the beta-turn conformation in the mutein, compared to the native protein. Beta-turn conformation was well-correlated with antigenicity because the residues that composed the beta-turn were usually hydrophilic, flexible, and accessible [64]. Reduction in beta-

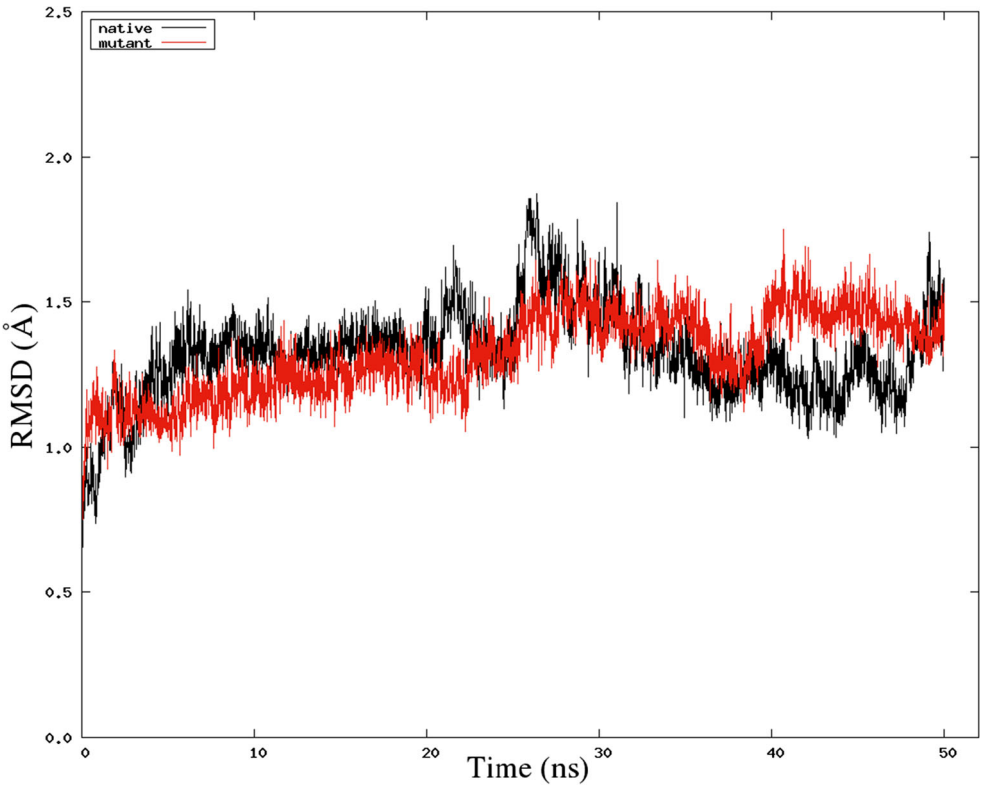

Fig. 4 Time evolution of the RMSD for native and mutated nattokinases

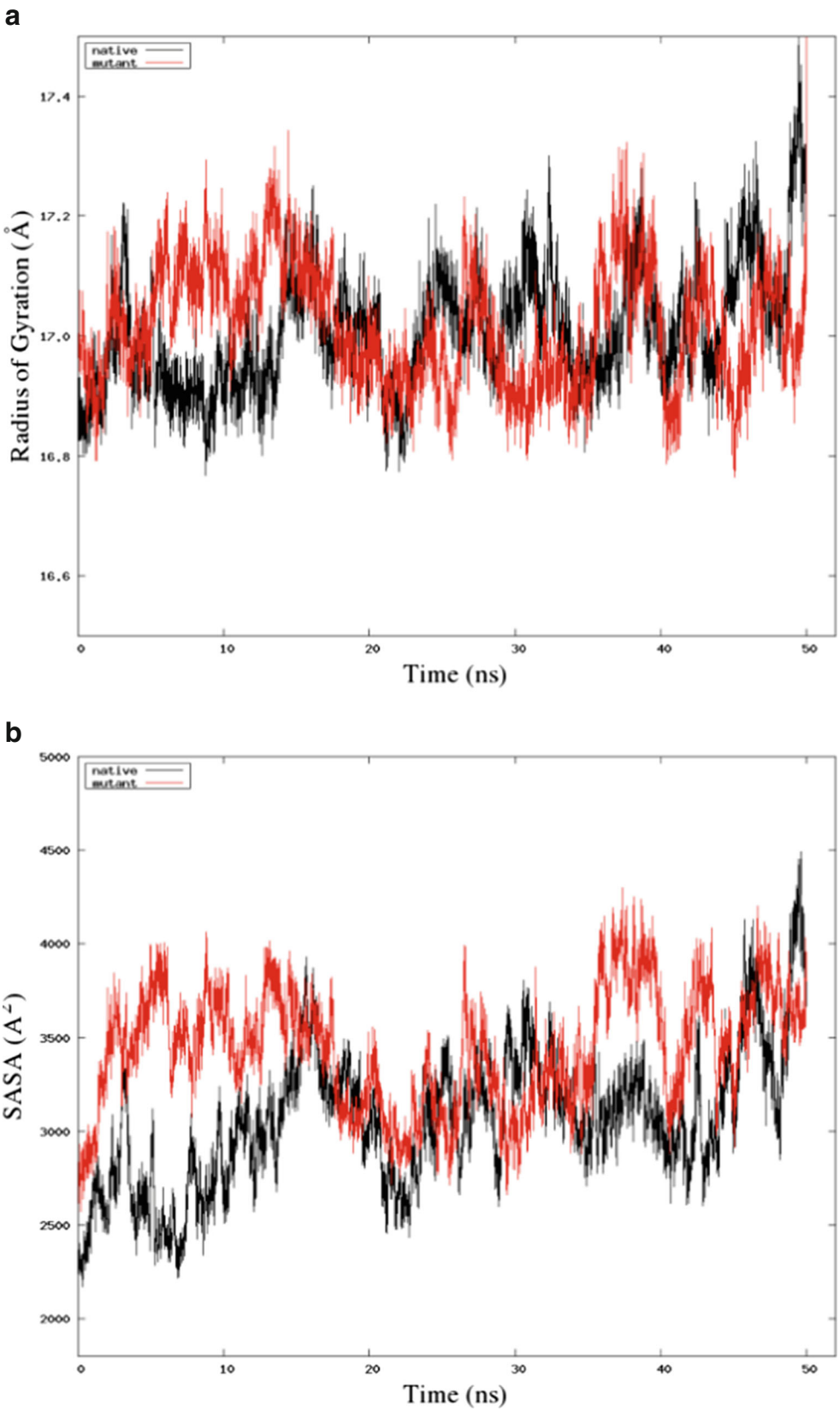

Fig. 5 R_g plot (**a**) and SASA (**b**) as a function of time for native and mutated nattokinase

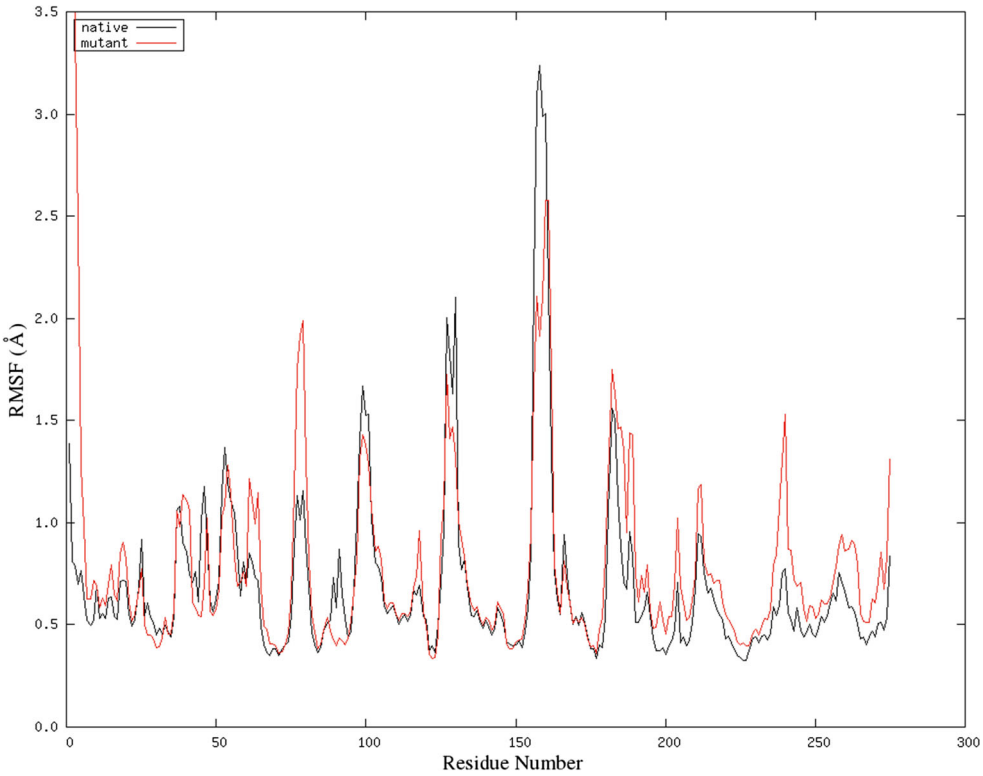

Fig. 6 RMSF analysis of the native nattokinase and mutated nattokinase

turn conformation in the mutein also indicated reduction in immunogenicity (Table 5).

Functionality of an enzyme was also evaluated by comparing the catalytic site and binding pocket of both mutated and native nattokinase. As indicated in Fig. 3, the catalytic site (D32, H64, N155, and S221) and the binding pocket (S125-P130; A152-E156; S163-Y167, A169, S191, T220, and S221) of the mutated proteins were completely similar with those of the native nattokinase. This was further supported with structure imposition after 50 ns of molecular dynamics simulation which showed there is no major difference after simulation (Supp file 3). Thus, the mutein was predicted to have similar mode of action compared to the native nattokinase.

Molecular dynamics simulation

Molecular dynamic simulations were employed to investigate and to compare the protein and mutein behaviors, conformational stability, and flexibility. Some dynamic properties, such as RMSD, RMSF of all residues, Rg, SASA, and secondary structure projection along the trajectory were studied. RMSD of all the backbone atoms from the initial structure indicated the convergence of the protein [65]. Both native and mutated nattokinases generated similar fluctuations along with the simulation, whereas all RMSD values are below 2.5 Å. After 40 ns of the simulation, both protein and mutein only gained a small increase of RMSD. This small magnitude of fluctuations indicated that the simulation reached stable trajectories (Fig. 4). The superimposed structure resulted from the simulation also revealed a 1.45 Å RMSD difference between native and mutated nattokinase which indicated there was no significant difference in structure (Supp file 3).

Rg and SASA were employed to analyze the geometrical behaviors of the native and mutated nattokinases. Both native and mutated nattokinase does not exhibit significant difference of Rg as depicted in Fig. 4a. The Rg relatively remained constant from the beginning until the end of the simulation. The implication of similar compactness between native and mutated nattokinase was further supported by SASA analysis, in which both SASA values along the simulation time did not differ significantly (Fig. 5). RMSF was employed to measure the flexibility of each residue in the protein. RMSF profiles between native and mutated nattokinase were similar (Fig. 6). However, there is an increase of flexibility in the amino acid residues of 150–160 due to the mutation. The gain of flexibility did not disturb the functional behavior of protein [66].

Table 6 Codon optimization parameter of the mutated nattokinase

Parameter	Range	Score
CAI	0.8–1	1
%GC	30–70%	62.79%
CFD	< 30%	0%
Negative cis element and negative repeat elements	-	0

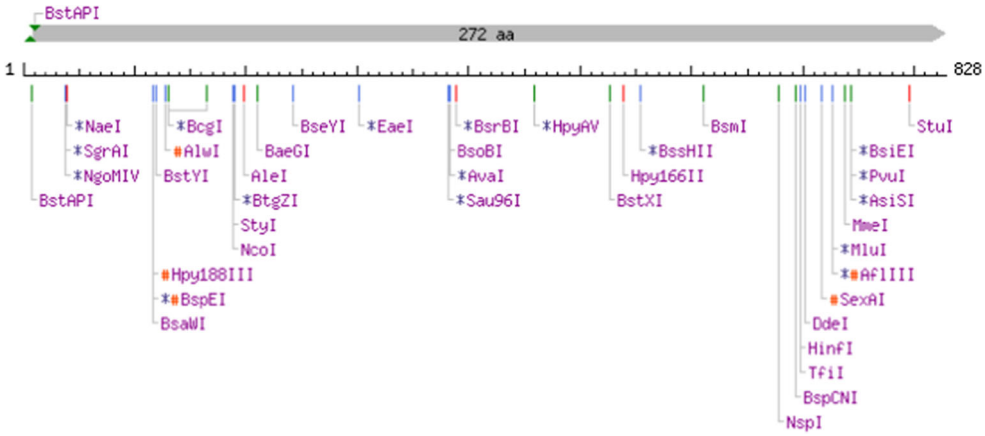

Fig. 7 Restriction sites prediction of the mutated nattokinase in silico synthetic gene

In reference to the RMSD, structure superimposition Rg, SASA, and RMSF, it can be interpreted that the mutation did not disrupt protein conformation and behavior. Thus, the mutation option for the less immunogenic nattokinase was allowed.

Codon optimization and in silico transcription

Reverse translation was employed to determine the proper codon and DNA sequences of the mutated nattokinase. It was optimized by using *E. coli* as the host for the protein production (Supp file 4). The optimized DNA sequence was analyzed on its codon adaptation index related to the codon usage bias [67], %GC content, and the codon frequency distribution (CFD) was related to the rare codon hindering the translation process. All parameters indicated that the generated gene sequence was within the range of the threshold or even showed a maximum value for the CAI and CFD (Table 6). The gene could be well-expressed in the *E. coli*. However, there was no negative cis element and repeated elements which could negatively regulate the expression of the gene or cancel out the transcription process. The restriction sites of this in silico synthetic gene were also described in Fig. 7. These sites acted as prediction to add the adaptors in the cis and the trans-region of the gene. These adaptors should contain restriction sites based on the selected plasmid but should not be present in the gene itself. Various restriction sites of common restriction endonuclease, such as EcoRI, BamHI, NotI, HindIII, and SacI, were not present in the optimized codon for the mutated nattokinase synthetic gene. Thus, there was no need to exploit the Wooble hypothesis to remove any restriction sites from the suggested codons. This could ease the cloning process of the mutated nattokinase gene.

Conclusions

In this study, the strategy to determine nattokinase B-cell epitopes and to generate mutated nattokinase with less immunogenicity but similar stability and functionality as proposed. Four mutations of S18D, Q19I, T242Y, and Q245W, altogether in one protein, generated the lowest antigenicity score. The similar stability as well as conformation of the mutated nattokinase was implied through stability analysis, RMSD, RMSF, Rg, SASA, and structure superimposition. This suggested nattokinase mutation be more applicable in the formulation of the injection due to the better safety and the same efficacy to the native one. This could act as suggestion to reduce or minimize cost and time in the wet-lab mutagenesis experiment.

Acknowledgments Thanks to Helen Hendaria Kamandhari, Ph.D. for her proofreading and comments.

Funding information This study is funded by the Faculty of Biotechnology, University of Surabaya.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Sumi H, Hamada H, Tsushima H, Mihara H, Muraki H (1987) A novel fibrinolytic enzyme (nattokinase) in the vegetable cheese Natto; a typical and popular soybean food in the Japanese diet. Experientia 43(10):1110–1111. https://doi.org/10.1007/BF01956052
2. Fujita M, Nomura K, Hong K, Ito Y, Asada A, Nishimuro S (1993) Purification and characterization of a strong fibrinolytic enzyme (nattokinase) in the vegetable cheese natto, a popular soybean fermented food in Japan. Biochem Biophys Res Commun 197(3): 1340–1347. https://doi.org/10.1006/bbrc.1993.2624
3. Nakamura T, Yamagata Y, Ichishima E (1992) Nucleotide sequence of the subtilisin NAT gene, *aprN*, of *Bacillus subtilis* (natto). Biosci Biotechnol Biochem 56(11):1869–1871. https://doi.org/10.1271/bbb.56.1869
4. Yanagisawa Y, Chatake T, Chiba-Kamoshida K, Naito S, Ohsugi T, Sumi H, Yasuda I, Morimoto Y (2010) Purification, crystallization and preliminary X-ray diffraction experiment of nattokinase from *Bacillus subtilis* natto. Acta Crystallogr Sect F Struct Biol Cryst

Commun 66(12):1670–1673. https://doi.org/10.1107/S1744309110043137
5. Murakami K, Yamanaka N, Ohnishi K, Fukayama M, Yoshino M (2012) Inhibition of angiotensin I converting enzyme by subtilisin NAT (nattokinase) in natto, a Japanese traditional fermented food. Food Funct 3(6):674–678. https://doi.org/10.1039/C2FO10245E
6. Dabbagh F, Negahdaripour M, Berenjian A, Behfar A, Mohammadi F, Zamani M, Irajie C, Ghasemi Y (2014) Nattokinase: production and application. Appl Microbiol Biotechnol 98(22):9199–9206. https://doi.org/10.1007/s00253-014-6135-3
7. Lee BH, Lai YS, Wu SC (2015) Antioxidation, angiotensin converting enzyme inhibition activity, nattokinase, and antihypertension of *Bacillus subtilis* (natto)-fermented pigeon pea. J Food Drug Anal 23(4):750–757. https://doi.org/10.1016/j.jfda.2015.06.008
8. Chen H, McGowan EM, Ren N, Lal S, Nassif N, Shad-Kaneez F, Qu X, Lin Y (2018) Nattokinase: a promising alternative in prevention and treatment of cardiovascular diseases. Biomark Insights 13:1177271918785130. https://doi.org/10.1777/1177271918785130
9. Feng R, Li J, Chen J, Duan L, Liu X, Di D, Deng Y, Song Y (2018) Preparation and toxicity evaluation of a novel nattokinase-tauroursodeoxycholate complex. Asian J Pharm Sci 13(2):173–182. https://doi.org/10.1016/j.ajps.2017.11.001
10. Kurosawa Y, Nirengi S, Homma T, Esaki K, Ohta M, Clark JF, Hamaoka T (2015) A single-dose of oral nattokinase potentiates thrombolysis and anti-coagulation profiles. Sci Rep 5:11601. https://doi.org/10.1038/srep11601
11. Lampe BJ, English JC (2016) Toxicological assessment of nattokinase derived from *Bacillus subtilis* var. natto. Food Chem Toxicol 88:87–99. https://doi.org/10.1016/j.fct.2015.12.025
12. Weng Y, Yao J, Sparks S, Wang KY (2017) Nattokinase: an oral antithrombotic agent for the prevention of cardiovascular disease. Int J Mol Sci 18(3):523. https://doi.org/10.3390/ijms18030523
13. Chitte RR, Deshmukh SV, Kanekar PP (2011) Production, purification, and biochemical characterization of a fibrinolytic enzyme from thermophilic *Streptomyces* sp. MCMB-379. Appl Biochem Biotechnol 165(5-6):1406–1413. https://doi.org/10.1007/s12010-011-9356-2
14. EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) (2016) Safety of fermented soybean extract NSK-SD® as a novel food pursuant to Regulation (EC) No 258/97. EFSA J 14(7):e04541. https://doi.org/10.2903/j.efsa.2016.4541
15. Zarei M, Nezafat N, Rahbar MR, Negahdaripour M, Sabetian S, Morowvat MH, Ghasemi Y (2018) Decreasing the immunogenicity of arginine deiminase enzyme via structure-based computational analysis. J Biomol Struct Dyn 37(2):523–536. https://doi.org/10.1080/07391102.2018.1431151
16. Fattahian Y, Riahi-Madvar A, Mirzaee R, Asadikaram G, Rahbar MR (2017) In silico locating the immune-reactive segments of *Lepidium draba* peroxidase and designing a less immune-reactive enzyme derivative. Comp Biol Chem 70:21–30. https://doi.org/10.1016/j.compbiolchem.2017.07.003
17. Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45(W1):W24–W29. https://doi.org/10.1093/nar/gkx346
18. Potocnakova L, Bhide M, Pulzova LB (2016) An introduction to B-cell epitope mapping and in silico epitope prediction. J Immunol Res 2016(6760830):1–11. https://doi.org/10.1155/2016/6760830
19. Singh H, Ansari HR, Raghava GP (2013) Improved method for linear B-cell epitope prediction using antigen's primary sequence. PloS One 8(5):e62216. https://doi.org/10.1371/journal.pone.0062216
20. Kringelum JV, Lundegaard C, Lund O, Nielsen M (2012) Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol 8(12):e1002829. https://doi.org/10.1371/journal.pcbi.1002829
21. Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65(1):40–48. https://doi.org/10.1002/prot.21078
22. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276(1-2):172–174. https://doi.org/10.1016/0014-5793(90)80535-Q
23. Weng M, Deng X, Bao W, Zhu L, Wu J, Cai Y, Jia Y, Zheng Z, Zou G (2015) Improving the activity of the subtilisin nattokinase by site-directed mutagenesis and molecular dynamics simulation. Biochem Biophys Res Commun 465(3):580–586. https://doi.org/10.1016/j.bbrc.2015.08063
24. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8(4):1–7. https://doi.org/10.1186/1471-2105-8-4
25. Yao B, Zhang L, Liang S, Zhang C (2012) SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PloS One. 7(9):e45152. https://doi.org/10.1371/journal.pone.0045152
26. Ansari HR, Raghava GP (2010) Identification of conformational B-cell epitopes in an antigen from its primary sequence. Immunome Res 6(6):1–9. https://doi.org/10.1186/1745-7580-6-6
27. Saha S, Raghava GPS (2004) BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. International Conference on Artificial Immune Systems. Springer, Berlin, pp 197–204. https://doi.org/10.1007/978-3-540-30220-9_16
28. Emini EA, Hughes JV, Perlow D, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55(3):836–839
29. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 22(2):195–201. https://doi.org/10.1093/bioinformatics/bti770
30. Benkert P, Biasini M, Schwede T (2010) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27(3):343–350. https://doi.org/10.1093/bioinformatics/btq662
31. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(W1):W252–W258. https://doi.org/10.1093/nar/gku340
32. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33(suppl_2):W306-W310. https://doi.org/10.1093/nar/gki375
33. Yang Y, Zhou Y (2008) Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. Protein Sci 17(7):1212–1219. https://doi.org/10.1110/ps.033480.107
34. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858. https://doi.org/10.1038/nprot.2015.053
35. Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66(1):12–21. https://doi.org/10.1107/S0907444909042073
36. Lovell SC, Davis IW, Arendall III WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 50(3):437–450. https://doi.org/10.1002/prot.10286

37. Uziela K, Menéndez Hurtado D, Shu N, Wallner B, Elofsson A (2017) ProQ3D: improved model quality assessments using deep learning. Bioinformatics. 33(10):1578–1580. https://doi.org/10.1093/bioinformatics/btw819
38. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364):83. https://doi.org/10.1038/356083a0
39. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253(5016):164–170. https://doi.org/10.1126/science.1853201
40. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519. https://doi.org/10.1002/pro.5560020916
41. Porter CT, Bartlett GJ, Thornton JM (2004) The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 32(suppl_1): D129–D133. https://doi.org/10.1093/nar/gkh028
42. Schmidtke P, Le Guilloux V, Maupetit J, Tuffery P (2010) Fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Res 38(suppl_2):W582-W589. https://doi.org/10.1093/nar/gkq383
43. Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7(2):525–537. https://doi.org/10.1021/ct100578z
44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935. https://doi.org/10.1063/1.445869
45. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible and free. J Comp Chem 26:1701–1718. https://doi.org/10.1002/jcc.20291
46. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105(28):6474–6487. https://doi.org/10.1021/jp003919d
47. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126(1):014101. https://doi.org/10.1063/1.2408420
48. Humphrey W, Dalke A, Schulten K (1996) VMD - visual molecular dynamics. J Mol Graph 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5
49. Bazmara H, Rasooli I, Jahangiri A, Sefid F, Astaneh SDA, Payandeh Z (2019) Antigenic properties of iron regulated proteins in *Acinetobacter baumannii*: an in silico approach. Int J Pept Res Ther 25(1):205–213. https://doi.org/10.1007/s10989-017-9665-6
50. Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R (2016) From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. Immunology 149(4):386–399. https://doi.org/10.1111/imm.12656
51. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y (2016) Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction-based approaches. Comp Biol Chem 62:82–95. https://doi.org/10.1016/j.compbiolchem.2016.04.006
52. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, Nabi AN (2016) In silico proposition to predict cluster of B-and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of *C. jejuni*. In Silico Pharmacol 4(5):1–10. https://doi.org/10.1186/s40203-016-0020-y
53. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, Li J, Li Y, Hu N, Wang H, Hu Y (2015) Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. PloS One 10(12):e0144475. https://doi.org/10.1371/journal.pone.0144475
54. Sefid F, Rasooli I, Jahangiri A, Bazmara H (2015) Functional exposed amino acids of BauA as potential immunogen against *Acinetobacter baumannii*. Acta Biotheor 63(2):129–149. https://doi.org/10.1007/s10441-015-9251-2
55. Pahil S, Taneja N, Ansari HR, Raghava GPS (2017) In silico analysis to identify vaccine candidates common to multiple serotypes of Shigella and evaluation of their immunogenicity. PloS One 12(8): e0180505. https://doi.org/10.1371/journal.pone.0180505
56. Talukdar S, Bayan U, Saikia KK (2017) In silico identification of vaccine candidates against *Klebsiella oxytoca*. Comp Biol Chem 69:48–54. https://doi.org/10.1016/j.compbiolchem.2017.05.003
57. Ramya LN, Pulicherla KK (2015) Studies on deimmunization of antileukaemic L-asparaginase to have reduced clinical immunogenicity-an in silico approach. Pathol Oncol Res 21(4): 909–920. https://doi.org/10.1007/s12253-015-9921-0
58. Beezhold DH, Hickey VL, Sussman GL (2001) Mutational analysis of the IgE epitopes in the latex allergen Hev b 5. J. Allergy Clin Immunol 107(6):1069–1076. https://doi.org/10.1067/mai.2001.115482
59. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404. https://doi.org/10.1016/S0076-6879(97)77022-8
60. Khor BY, Tye GJ, Lim TS, Noordin R, Choong YS (2014) The structure and dynamics of BmR1 protein from *Brugia malayi*: In silico approaches. Int J Mol Sci 15(6):11082–11099. https://doi.org/10.3390/ijms150611082
61. Messaoudi A, Belguith H, Hamida JB (2013) Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. Theor Biol Med Model 10(1):22. https://doi.org/10.1186/1742-4682-10-22
62. Guruprasad K, Reddy BB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng Des Sel 4(2):155–161. https://doi.org/10.1093/protein/4.2.155
63. Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88(6):1895–1898. https://doi.org/10.1093/oxfordjournals.jbchem.a133168
64. Pellequer JL, Westhof E, Van Regenmortel MH (1993) Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunol Lett 36(1):83–99. https://doi.org/10.1016/0165-2478(93)90072-A
65. Rajendran V, Purohit R, Sethumadhavan R (2012) In silico investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein. Amino acids 43(2):603–615. https://doi.org/10.1007/s00726-011-1108-7
66. Kamaraj B, Purohit R (2013) In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. Biomed Res Int 2013: 697051. https://doi.org/10.1155/2013/697051
67. Sharp PM, Li WH (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3):1281–1295. https://doi.org/10.1093/nar/15.3.1281