Research Article

Insights Into The Effects of Amino Acid Substitutions on The Stability of Reteplase Structure: A Molecular Dynamics Simulation Study

Kaveh Haji-Allahverdipoor[1], Mokhtar Jalali Javaran[2], Sajad Rashidi Monfared[2], Mohamad Bagher Khadem-Erfan[1], Bahram Nikkhoo[1], Zhila Bahrami Rad[1], Habib Eslami[3]**, Sherko Nasseri[1]*

[1]Cellular and Molecular Research Center, Research Institute for Health Development, Kurdistan University of Medical Sciences, Sanandaj, Iran
[2]Department of Biotechnology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran
[3]Department of Pharmacology and Toxicology, School of Pharmacy, Hormozgan University of Medicinal sciences, Bandar Abbas, Iran

Corresponding author: Sherko Nasseri, Cellular and Molecular Research Center, Research Institute for Health Development, Kurdistan University of Medical Sciences, Sanandaj, Iran. Tel: +98-8733627228, Fax: +98-8733627228, E-mail: sherko.nasseri@muk.ac.ir

****Co-Corresponding author*: Habib Eslami, Department of Pharmacology and Toxicology Faculty of Pharmacy, Molecular Medicine Research Center, Hormozgan Health Institute, Hormozgan University of Medicinal sciences, Bandar Abbas, Iran. Tel: +98-7633710406, Fax: +98-7633710406, E-mail: habib.Eslamidoolabi@hums.ac.ir

Background: Reteplase (recombinant plasminogen activator, r-PA) is a recombinant protein designed to imitate the endogenous tissue plasminogen activator and catalyze the plasmin production. It is known that the application of reteplase is limited by the complex production processes and protein's stability challenges. Computational redesign of proteins has gained momentum in recent years, particularly as a powerful tool for improving protein stability and consequently its production efficiency. Hence, in the current study, we implemented computational approaches to improve r-PA conformational stability, which fairly correlates with protein's resistance to proteolysis.

Objectives: The current study was developed in order to evaluate the effect of amino acid substitutions on the stability of reteplase structure using molecular dynamic simulations and computational predictions.

Materials and Methods: Several web servers designed for mutation analysis were utilized to select appropriate mutations. Additionally, the experimentally reported mutation, R103S, converting wild type r-PA into non-cleavable form, was also employed. Firstly, mutant collection, consisting of 15 structures, was constructed based on the combinations of four designated mutations. Then, 3D structures were generated using MODELLER. Finally, 17 independent 20-ns molecular dynamics (MD) simulations were conducted and different analysis were performed like root-mean-square deviation (RMSD), root-mean-square fluctuations (RMSF), secondary structure analysis, number of hydrogen bonds, principal components analysis (PCA), eigenvector projection, and density analysis.

 Results: Predicted mutations successfully compensated the more flexible conformation caused by R103S substitution, so, improved conformational stability was analyzed from MD simulations. In particular, R103S/A286I/G322I indicated the best results and remarkably enhanced the protein stability.

Conclusion: The conformational stability conferred by these mutations will probably lead to more protection of r-PA in protease-rich environments in various recombinant systems and potentially enhance its production and expression level.

Keywords: Amino Acid Substitution, Homology Modeling, Molecular Dynamics Simulation, Protein Stability, Reteplase, Recombinant plasminogen activator

1. Background

Reteplase (recombinant plasminogen activator, r-PA) is a serine protease, which mimics endogenous tissue plasminogen activator (t-PA) by catalyzing the production of plasmin from plasminogen, leading to fibrinolysis (1). Being a third-generation thrombolytic agent, r-PA works in the presence of fibrin. Since r-PA is a non-glycosylated protein, it can be produced in prokaryotic as well as eukaryotic expression system, and for its activation requires *in vitro* folding. The *in vitro* enzymatic activity of r-PA is comparable to the native t-PA, and in clinical setting, it is a potent thrombolytic drug with prolonged half-life in humans. These properties result from the fact that r-PA diffuses more easily into clots instead of binding only on the surface because of its decreased affinity for fibrin. Additionally, the long half-life of r-PA is due to the removal of its binding domain for hepatocytes involved in the hepatic clearance of t-PA (2). Notwithstanding all merits, there are still some challenges associated with the production and application of reteplase need to be addressed, such as aggregation tendency, low solubility and instability. Therefore, researchers all around the world are recruiting different approaches to design novel forms of reteplase with improved features (2-4).

r-PA is constructed from only two domains of t-PA (kringle-2 and catalytic serine protease domains), and like t-PA is primarily synthesized as a single-chain polypeptide. The subsequent proteolytic cleavage generates a two-chain form of the enzyme with increased activity. However, unlike other serine proteases, both agents are also active in single-chain form, which is the preferred form for therapeutic administration in acute myocardial infraction (5, 6).

Therapeutic proteins such as r-PA can be expressed and produced in a wide range of recombinant systems, however, their susceptibility to proteolytic degradation represents a major problem related to the production process (7). For instance, cleavage and degradation of t-PA has been detected in plants, one of the valuable expression systems for therapeutic and industrial proteins. Moreover, production of desmoteplase in tobacco is inhibited by proteolysis (8). Other proteins such as human growth hormone, equistatin and different antibodies are also cleaved during production, as confirmed by immunoblot analysis of transgenic plant extracts (9, 10). It has been demonstrated that the resistance to proteolysis is strongly associated with the relative inherent stability of polypeptide chains expressed in heterologous environment. This factor influences recombinant protein quality and determines the overall yield of foreign proteins in plant cells as well. Besides, protein stability is also an essential issue in the storage of harvested plant biomass (11,12). Conversely, instability and more flexible conformation can potentially make proteins liable to proteolysis. This relies on the fact that proteolysis can happen only if the polypeptide chain can bind and easily adapt to the specific spatial conformation of the protease's active site (13). The other contributing factor that renders proteins vulnerable to cleavage and degradation is the existence of protease-susceptible amino acids sites, which interact favorably with endogenous proteases in the transgenic host system. In the case of t-PA, conversion of the two-chain to its one chain form is the main proteolytic susceptible reaction that has been observed during purification in the pH range of 5-9. In natural pH range, however, the stability-limiting reaction is conversion of one chain to two chain form, which occurs after proteolytic cleavage at Arg276 on a t-PA molecule (14, 15).

It has been shown that R276S mutation generates a non-cleavable single-chain form of t-PA, characterized with both lack of neurotoxicity and reduced amidolytic activity, while conserving fibrinolytic activity (16). Additionally, among various fibrinolytic drugs, desmoteplase is exclusively non-cleavable single chain protein secreted by the *D.rotundus* salivary glands and numerous studies have emphasized the absence of neurotoxicity for this protease. Indeed, the deleterious effects of r-PA in the brain parenchyma is mediated through its interaction with the N-methyl-D-aspartate receptor (NMDAR), with a subsequent NMDA-induced calcium influx and ultimately neuronal death. Studies with mutant enzymes that lack either the cleavage site of t-PA (Arg276) or the kringle-2 lysine-binding site have demonstrated the attenuation of this side effect (17-19). Therefore, in the current study, mutation at the cleavage site was employed in the r-PA structure to produce non-cleavable single chain enzyme, with the aim of increasing overall stability of protein, as well as enhancing its production and the expression level. Generally, expression and production of r-PA is limited by recombinant systems such as eukaryotic expression systems, owing to the complicated and expensive processes (20). The low expression level of

r-PA in transgenic tobacco plants has been also reported previously (21).

2. Objective

In the current study, in addition to the experimentally reported mutation at the cleavage site of r-PA (Arg103), we utilized stabilization strategy and stability prediction upon mutation servers to select appropriate mutations in order to attenuate the structural flexibility and solvent accessibility of this protein. The most conformationally stable mutant was attempted to be identified by performing computational procedures, including homology modeling and molecular dynamics (MD) simulations.

3. Materials and Methods

3.1. Protein Structure Prediction
Primary amino acid sequence of human r-PA was submitted to NCBI Protein BLAST (blastp) (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and searched against the latest version of PDB data bank for identifying suitable homologous template. The structural models of wild type r-PA and its non-cleavable form (R103S) were generated with MODELLER 9V15 (22) using 1TPK (kringle-2 domain of human t-PA) and 1BDA (catalytic domain of human single chain t-PA) X-ray crystal structures as templates. The input files used to build the models were template coordinate files in PDB format, alignment file of r-PA and Python commands file. After completion of modeling, a model with lowest discrete optimized energy (DOPE score) was chosen among 10000 models calculated for wild type and non-cleavable form of protein. Then, the model generated for r-PA was successfully submitted in protein model database (PMDB) available at http://srv00.recas.ba.infn.it/PMDB/ with the accession number of PM0080444.

3.2. Stabilizing Strategy and Mutant Designation
Hydrophobic interactions are primarily responsible for the stabilization of globular conformation of proteins. According to $\Delta\Delta G$ values calculated for hydrophobic mutants in experimental studies, burying -CH$_2$- group contributes largely to protein stability through two terms. The first one is the constant term, which depends on the hydrophobicity difference between mutant and wild-type side chains. The second is a variable term that depends on the difference in Van der Waals interaction of the side chains (23, 24). All possible mutations that increase hydrophobicity in the protein core were submitted to stability prediction upon mutation servers, including SDM (Site Directed Mutator), mCSM, and DUET to calculate the difference between ΔG of wild type and mutant ($\Delta\Delta G$). Stabilizing mutations were designated based on $\Delta\Delta G$ calculated for all possible mutations. SDM is energy which calculated based statistical potential function that recruits environment-specific amino-acid substitution frequencies, especially within homologous protein families to calculate a stability score. This score is analogous to the free energy difference between the wild-type and mutant protein (25). mCSM, which relies on graph based signatures, was successfully applied and evaluated in different predictive tasks and was shown to outperform earlier methods (26). DUET is a web server for an integrated computational approach to study missense mutations in proteins. DUET consolidates two complementary approaches (mCSM and SDM) in a consensus prediction, obtained by combining the results of the separate methods in an optimized predictor using Support Vector Machines (SVM) (27). Finally, the most stabilizing mutations (four mutations) among all possible mutations, were selected for the next steps.

3.3. In Silico Mutant Collection
All combinations of the four designated mutations were calculated and 3D structures of all models were built using MODELLER 9V15 through homology modeling. For each protein, a model with the lowest DOPE score was chosen from 10000 generated models. In silico mutant collection is comprised of four non-cleavable point mutants, six non-cleavable double mutants, four non-cleavable triple mutants and one non-cleavable quadruple mutant (**Fig. 1A**).

3.4. Structure Minimization and Validation
Gromacs 5.0.2 was used for the structure minimization of the models. Computations were done with parameters of Amber ff99SB-LIDN forcefield. The minimization protocol included 500 steps of steepest descent algorithm (28, 29). After optimization procedure, the generated models were validated by using ProSA-web (Protein Structure Analysis), RAMPAGE and Verify3D servers. ProSA-web server is widely used to assess the quality of three dimensional models.
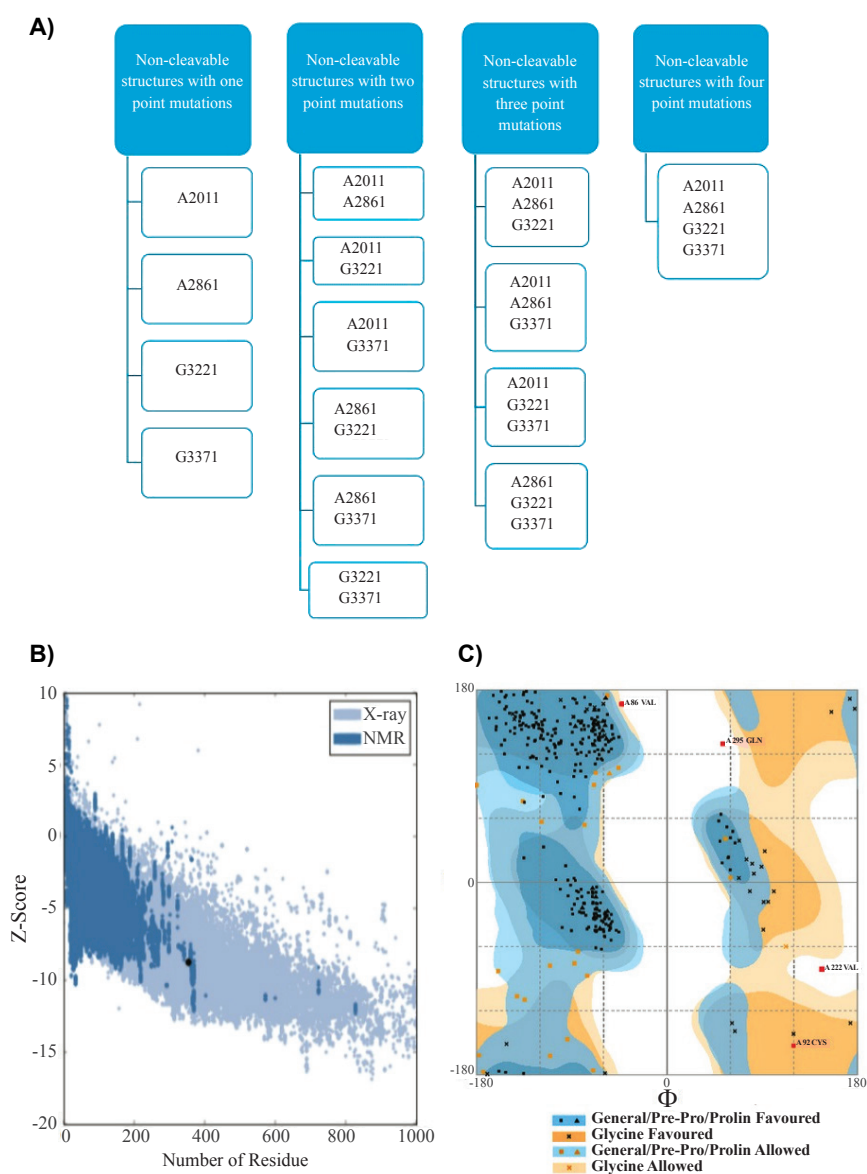
Figure 1. Combination of mutations, Z-score and Ramachandran plots. A) All possible combination of four designated mutations (A201I, A286I, G322I, and G337I). **B)** Z- score of reteplase structure, generated using x-ray structures of its two domains, is -8.76 marked as block dot and has similar quality as native structures. **C)** Ramachandran plot shows that the reteplase structure has 98.8% amino acids in allowed region, hence confirming the reliability of the structure.

The Z-score calculated by ProSA-web server is representative of model quality and indicates whether the modeled structure is within the range of scores obtained from native proteins of similar size. Backbone conformation was assessed by the Psi/Phi Ramachandran plot obtained from RAMPAGE server. Verify3D assesses compatibility of each residue against three dimensional structure which is based on 3D-1D profile derived from statistical potentials (22, 30).

3.5. Molecular Dynamics Simulation
Molecular dynamics simulation of wild type, non-cleavable form and mutants were performed with Gromacs 5.0.2 software under Amber ff99SB-LIDN forcefield. Each initial structure was centered in a cubic box with 1.0 nm distance from the box edge. Periodic

82

Iran. J. Biotechnol. January 2023;21(1): e3175

boundary conditions were applied to avoid artificial effects due to finite size of simulation box. The box was then filled with appropriate amount of TIP3P water molecules. System was made electrically neutral by replacing water molecules with required number of counter-ions using the genion tool of Gromacs package (31). Energy minimization was done using steepest descent algorithm. The energy-minimized system was equilibrated by the position restrained simulation under canonical ensemble (NVT) followed by isothermal isobaric ensemble (NPT) in 500 and 1000 ps respectively. The initial velocities were derived from Maxwell-Boltzman distribution at 300 k. Productive unrestrained MD simulations were carried out in the NPT ensemble for 20 ns and bond lengths constrained with LINCS algorithm. The leapfrog algorithm with 2 fs time step was applied to integrate the equation of motion. Conditions of constant pressure 1 bar and temperature 300 K were accomplished by application of the Parinello-Rahman barostat and Nosé-Hoover thermostat. The particle mesh Ewald (PME) method was used for calculating the long-range electrostatics interactions. The utilities of Gromacs 5.0.2 such as gmx rms, gmx rmsf, gmx gyrate, gmx sasa, gmx density and gmx hbond were used to analyze MD simulation trajectories (32, 33). The program DSSP was used to study the secondary structure fluctuations (34). Principal components analysis (PCA) reduces the dimensionality of the data obtained from MD simulations. PCA is described in detailed by Giuliani (35). Here, it was performed to examine the protein global motions. The Gromacs utility "gmx covar" was used to calculate and diagonalize the covariance matrix

of Cα atomic positions from the 20 ns trajectories of the wild type r-PA and its mutants. Then, the eigenvalues and corresponding eigenvectors of the matrix were produced. The gmx anaeig was used for analysis and plotting the eigenvectors (35, 36). All graphs were plotted using Xmgrace (37). The gmx sham was used to calculate the free energy (G) by the first two eigenvectors extracted after PCA. 2D and 3D representation of the free energy landscape (FEL) were plotted using GNUplot 5.2 (www.gnuplot.info). The dynamic cross-correlation matrix (DCCM) that shows the fluctuations of the fluctuations between any two pair of Cα atoms were calculated by the Bio3D package.

4. Results

4.1. Protein Modeling and Mutant Library Construction
Since there is no experimentally solved 3D structure for r-PA, the structure of wild type r-PA was modeled based on the available x-ray structures of its two domains through homology modeling and using the program MODELLER. The structure quality was confirmed with online servers, including ProSA-web, Rampage and Verify3D (**Fig. 1B , 1C**). This structure includes 355 amino acids, which comprise amino acids 1-3 and 176-527 of human t-PA. Thus, it contains only the kringle 2 and the protease domains of human t-PA. However, similar to t-PA, wild type r-PA undergoes proteolytic cleavage at Arg103 to generate a two-chain form of the enzyme, which is the stability-limiting reaction. This cleavage site is equal to Arg276 on wild type t-PA. In this study all structures, other than wild type r-PA, had R103S mutation and therefore were non-cleavable

Table 1. List of in-silico mutations were chosen based on biophysical principals of mutation and prediction servers.

Mutation	Secondary structure	Wild Propensity (Pβ) [*]	Mutant Propensity (Pβ)	mCSM (ΔΔG) [†] Kcal.moL [-1]	SDM (ΔΔG) Kcal.moL [-1]	DUET (ΔΔG) Kcal.moL [-1]
A201I	β-sheet	0.83	1.6	0.744	1.00	1.082
A286I	β-sheet	0.83	1.6	0.724	1.00	1.063
G322I	β-sheet	0.75	1.6	1.505	2.81	1.820
G337I	β-sheet	0.75	1.6	1.354	4.25	1.718

[*] Based on Chou-Fasman algorithm

[†] ΔΔG = ΔG wild – ΔG mutant upon single point mutation, The difference in folding free energy between wild type and mutant structure.

Table 2. Assessment of selected models quality and reliability

Protein	ProSA-web (Z-score) *	Ramachandran plot			Verify- 3D ‡
		Favored region †	Allowed region †	Disallowed region †	
Wild (Reteplase)	-8.76	92.0 %	6.8 %	1.2%	99.15 %
R103S	-8.75	91.2 %	6.8 %	2.0 %	97.7 %
R103S/A201I	-8.78	92 %	6.2 %	1.7 %	98.87 %
R103S/A286I	-8.53	92 %	6.2 %	1.7 %	96.90 %
R103S/G322I	-8.67	93.2 %	5.7 %	1.1 %	96.62 %
R103S/G337I	-8.77	92 %	6.0 %	2.0 %	96.06 %
R103S/A201I/A286I	-8.75	93.2 %	4.8 %	2.0 %	95.21 %
R103S/A201I/G322I	-8.84	94 %	4.5 %	1.4 %	95.49 %
R103S/A201I/G337I	-9.00	92 %	5.7 %	2.3 %	97.18 %
R103S/A286I/G322I	-8.76	92 %	6.2 %	1.7 %	97.75 %
R103S/A286I/G337I	-8.84	93.8 %	4.8 %	5 %	96.62 %
R103S/G322I/G337I	-8.81	92.6 %	5.1 %	2.0 %	96.34 %
R103S/A201I/A286I/G322I	-8.74	93.5 %	4.8 %	1.7 %	93.80 %
R103S/A201I/A286I/G337I	-9.15	93.2 %	5.7 %	1.1 %	97.18 %
R103S/A201I/G322I/G337I	-8.85	91.8 %	6.5 %	1.7 %	96.90 %
R103S/A286I/G322I/G337I	-8.77	92.9 %	5.4 %	1.7 %	98.59%
R103S/A201I/A286I/G322I/G337I	-8.85	93.2 %	5.1 %	1.7 %	99.72 %

* Indicative of overall homology model quality using ProSA-web server

† Based on Ramachandran plot of predicted model by RAMPAGE server

‡ Based on 3D-1D profile of verify3D server and indicative of the percentage of residues having scores greater than 0.2 in each model.

single chain form of the protein (16). The rationale behind this approach is that this form is resistant to the main stability-limiting reaction and also causes less neurotoxicity than the wild type.

Since hydrophobic interaction in the core of protein has the major influence on the stability of protein, all possible mutations that enhance this interaction were investigated by using stability prediction servers, including SDM, mCSM and DUET. Then, the more stabilizing mutations were selected based on the folding free energy differences in wild type and mutant structures (**Table 1**). In addition to stabilizing strategy and prediction servers, biophysical principals of mutation for the improvement of the secondary structure stability were taken into consideration as well. For example, isoleucine is a strong beta-sheet former and has a higher propensity in beta-sheet rather than alanine and glycine (38). Finally, among

all possible mutations, four more stabilizing mutations were selected according to stabilizing strategy, $\Delta\Delta G$ of SDM, mCSM and DUET as well as amino acid propensity in corresponding secondary structure. All combination of designated mutations, consist of four non-cleavable point mutants, six non-cleavable double mutants, four non-cleavable triple mutants and one non-cleavable quadruple mutant are represented in **Figure 1A** . The structure of mutants were modeled and the selected model from 10000 calculated ones (based on DOPE score), was submitted to energy minimization with Amber ff99SB-Lidn parameters. Energy minimized models were assessed by ProSA-web, Rampage and Verify3D servers. All Z-scores of ProSA-web were inside the range of characteristics for native proteins with similar size (**Table 2**). Based on the Ramachandran plot, all structures had noticeable number of residues in the favored and allowed regions. Verify3D scores near

the zero indicates erroneous in model and the great percentage of residues had an average scores greater than 0.2 in all models constructed in this study (**Table 2**).

4.2. Molecular Dynamics Simulation

4.2.1. Root Mean Square Deviation (RMSD)
To evaluate the stability of the mutant structures, RMSDs of the Cα atoms after 20 ns simulations with respect to the modeled structure was calculated. The RMSD of all Cα atoms was evaluated to study the convergence of protein structure through the simulations. RMSD analysis indicated that wild type and mutant structures are relatively stable. The average and standard deviation of RMSD values for all proteins have been presented in Table 3. From the RMSD values and deviations we could find out that non-cleavable double mutant, R103S/A286/G322I, is the most stable mutant. Therefore, the RMSD plot of this mutant has been shown in **Figure 2A** along with the plots of wild type and non-cleavable form. Moreover, RMSD analysis indicated non-cleavable forms of one point mutant (R103S/G337I), one double mutant (R103S/G322I/G337I), two triple mutants (R103S/A201I/G322I/G337I, R103S/A201I/A286I/G337I) and the quadruple mutant have stable structures in comparison with wild type, non-cleavable form and other mutants (**Table 3**).

4.2.2. Root Mean Square Fluctuation (RMSF)
To investigate the effect of mutations on dynamic behavior of flexible regions of protein, the RMSF values were calculated for all systems (**Table 3**). In addition to RMSF profile, average and standard deviation of RMSF values indicated that, among all mutations, non-cleavable double mutant (R103S/A286I/G322I) has significant decrease in the RMSF values globally. Hence, the RMSF plot of this mutant along with the plots of wild type and non-cleavable form has been illustrated in **Figure 2B**. The N-terminal of r-PA was quite flexible throughout MD simulations as shown in the RMSF plots (**Fig. 2B**). The large fluctuation of N-terminal of the protein would trigger instability in the structure of protein. According to RMSF plots of mutant proteins, one point mutant (R103S/G337I), two double mutants (R103S/A286/G322I and R103S/G322I/G337I), two triple mutants (R103S/A201I/G322I/G337I and R103S/A201I/A286I/G337I) and quadruple

mutant are affected structurally by mutations, as evidenced by less flexibility especially at N-terminus (**Table 3**). Analysis of fluctuation of each residue in mutant structures indicated that the highest flexibility reduction occurs in 1-80 residues as compared to wild type and non-cleavable form of r-PA. Therefore, the kringle2 domain of r-PA shows significant decrease in RMSF value in R103S/A286I/G322I mutant than in the wild type and R103S protein structure (**Fig. 2B**).

4.2.3. Solvent Accessible Surface Area (SASA)
The compactness of hydrophobic core of protein was analyzed by the solvent accessible surface area (SASA). Decreased values in mutant structures indicated that these structures have more compressed core compared with the wild type and non-cleavable structures (**Fig. 2C, Fig. 3B**) (**Table 3**). The average values of SASA indicated that R103S/A201I (183.333 ± 3.889 nm^2), R103S/A286I (183.753 ± 3.422 nm^2), R103S/G337I (184.528 ± 2.641 nm^2) and R103S/A286I/G322I (184.712 ± 2.938 nm^2) have more compacted structures than wild type and non-cleavable form of r-PA. The profile of SASA of these mutant structures in **Figure 3A** illustrates that R103S/A286I/G322I mutant has a relatively steady SASA profile in comparison with R103S/A201I, R103S/A286I and R103S/G337I mutants during simulation.

4.2.4. Radius of Gyration (Rg)
The radius of gyration (Rg) is the mass weighted root mean square distance of group of atoms from their common center of mass. The radius of gyration provided an observation into dimension of the protein through the simulation. Radius of gyration graph for Cα atoms of proteins with time at 300 K are tabulated in **Table 3**. Also, Rg plots of R103S/A286/G322I mutant, wild type, and non-cleavable form are depicted in **Figure 3B**. The wild type and mutant structures have a similar Rg profile and average values of Rg have limited range of variations (**Fig. 3B, Table 3**). Maximum and minimum Rg values are 2.252 ± 0.022 and 2.113 ± 0.020 nm for R103S/A201I/A286I and R103S/G337I mutant, respectively. Values and trend of variations of Rg showed that dimensions of proteins had no noticeable differences among wild type and mutant structures through mutations.

Iran. J. Biotechnol. January 2023;21(1): e3175

85
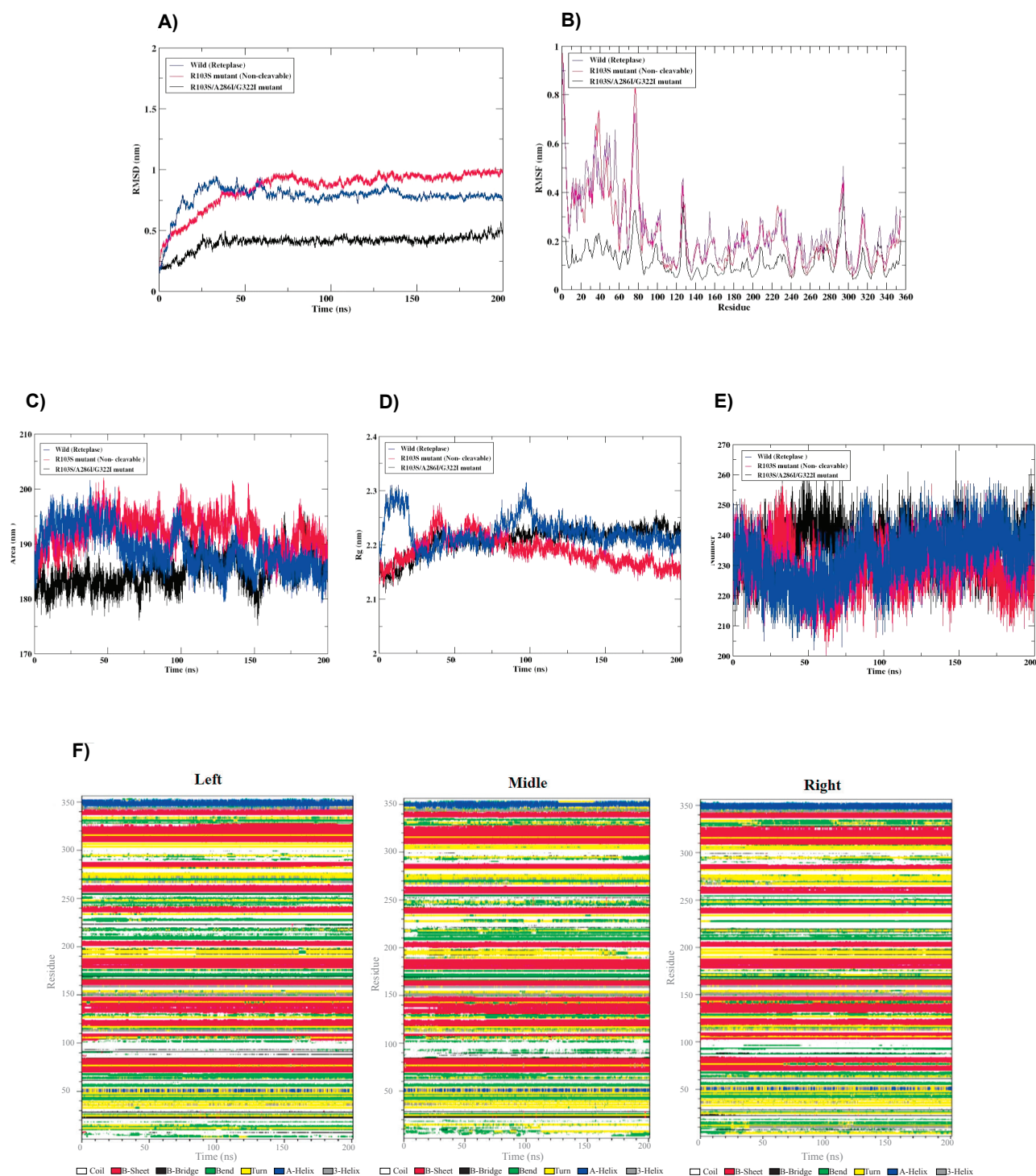
Figure 2. RMSD, RMSF, SASA, Rg and DSSP plots. A) The plots of Cα RMSD, and B. RMSF in wild-type reteplase (blue), R103S mutant as a non-cleavable form(red) and R103S/A286I/G322I mutant (black) structures at 300 K. **B)** This plots indicated that designated mutant has more stable and less flexible structure related to aim of the stabilization of Reteplase. **C)** the Plots of SASA, and **D)** The Plots of Rg of Cα atoms, and **E)** the number of intra-molecular hydrogen bonds of wild-type reteplase (blue), R103S mutant as a non-cleavable form (red) and R103S/A286I/G322I mutant (Black) structures at 300 K. **F)** Time evolution of the secondary structure fluctuations by DSSP analysis of wild-type reteplase A, R103S mutant as a non-cleavable form B and R103S/A286I/G322I C mutant structures at 300K.

86

Iran. J. Biotechnol. January 2023;21(1): e3175

4.2.5. Number of Hydrogen Bonds (NH bonds)

Hydrogen bonds play a major role on keeping the stable conformation of proteins. A variable profile of hydrogen bonds were observed among wild and mutant structures during simulation (**Table 3**). Mutant structures showed relatively greater number of intra-molecular hydrogen bonds formation than wild type and non-cleavable form. More intra-molecular hydrogen bonds might help protein to preserve rigidity. Plots of NH bonds for the R103S/A286/G322I mutant, wild type, and non-cleavable form of r-PA are shown in **Figure 3C**.

4.3. Analysis of Secondary Structure

Further information on the structural stability of protein was found by the analysis of time-dependent secondary structures. The DSSP algorithm was applied to study changes in secondary structure throughout the simulations (34). As evidenced from RMSF plots, the trend of variations in flexibility behavior of protein from its N terminus to the end of its C terminus was obviously related to the secondary structure content and variation of both ends of protein. Based on this conclusion, we easily distinguished 1-70 regions from the rest of the structure according to RMSF and secondary structure analysis (**Fig. 2B, Fig. 2F**). To understand how the mutations affect secondary structure and leads to more stable structures, variation in secondary structure of these mutants were carefully monitored. The severe secondary structure changes in the 100-110 region of wild and non-cleavable form of r-PA were observed. Tendency to loss of these variations and β-sheet formation were observed in all mutants that successfully control N-terminus fluctuations and have better RMSD values (**Fig. 2A**). Due to the massive data and for easier interpretation, secondary structure profiles of just R103S/A286/G322I mutant, wild type, and non-cleavable form are represented in **Figure 2F**.

Table 3. Summarized parameters of 20 ns molecular dynamics simulations at 300 K.

| Protein	RMSD [*] (nm)	RMSF [†] (nm)	Rg [‡] (nm)	SASA [§] (nm^2)	NH bonds [Ɩ]
Wild (Reteplase)	0.777 ± 0.102	0.248 ± 0.143	2.223 ± 0.027	189.025 ± 3.732	232.162 ± 8.890
R103S	0.849 ± 0.155	0.226 ± 0.151	2.188 ± 0.025	192.053 ± 3.119	229.767 ± 7.929
R103S/A201I	0.824 ± 0.130	0.188 ± 0.128	2.153 ± 0.035	183.333 ± 3.889	239.775 ± 9.314
R103S/A286I	0.694 ± 0.092	0.172 ± 0.116	2.138 ± 0.020	183.753 ± 3.422	243.040 ± 9.364
R103S/G322I	0.682 ± 0.124	0.195 ± 0.148	2.162 ± 0.020	188.482 ± 2.486	236.203 ± 8.000
R103S/G337I	0.573 ± 0.070	0.138 ± 0.090	2.113 ± 0.020	184.528 ± 2.641	241.430 ± 7.090
R103S/A201I/A286I	0.575 ± 0.064	0.163 ± 0.087	2.252 ± 0.022	189.821 ± 3.177	234.522 ± 7.370
R103S/A201I/G322I	0.651 ± 0.079	0.154 ± 0.098	2.161 ± 0.013	186.443 ± 2.424	232.412 ± 7.538
R103S/A201I/G337I	0.794 ± 0.137	0.192 ± 0.129	2.182 ± 0.018	188.188 ± 2.938	235.264 ± 7.443
R103S/A286I/G322I	0.406 ± 0.062	0.120 ± 0.060	2.209 ± 0.023	184.712 ± 2.938	236.645 ± 7.262
R103S/A286I/G337I	0.677 ± 0.101	0.199 ± 0.132	2.223 ± 0.028	188.177 ± 3.269	235.740 ± 7.298
R103S/G322I/G337I	0.517 ± 0.059	0.138 ± 0.088	2.167 ± 0.016	185.654 ± 2.873	237.176 ± 7.274
R103S/A201I/A286I/G322I	0.666 ± 0.156	0.205 ± 0.175	2.148 ± 0.016	187.638 ± 3.069	235.409 ± 7.573
R103S/A201I/A286I/G337I	0.530 ± 0.071	0.135 ± 0.089	2.184 ± 0.018	185.954 ± 3.198	239.987 ± 8.063
R103S/A201I/G322I/G337I	0.592 ± 0.062	0.137 ± 0.070	2.198 ± 0.014	186.262 ± 2.488	232.796 ± 7.158
R103S /A286/G322I/G337I	0.564 ± 0.090	0.151 ± 0.068	2.168 ± 0.015	189.570 ± 3.296	227.314 ± 6.991
R103S/A201I/A286I/G322I/G337I	0.534 ± 0.057	0.146 ± 0.087	2.130 ± 0.012	185.295 ± 2.743	239.619 ± 8.030

[*] Root mean square deviation
[†] Root mean square fluctuation
[‡] Radius of gyration
[§] Solvent accessible surface area
[Ɩ] Number of hydrogen bond
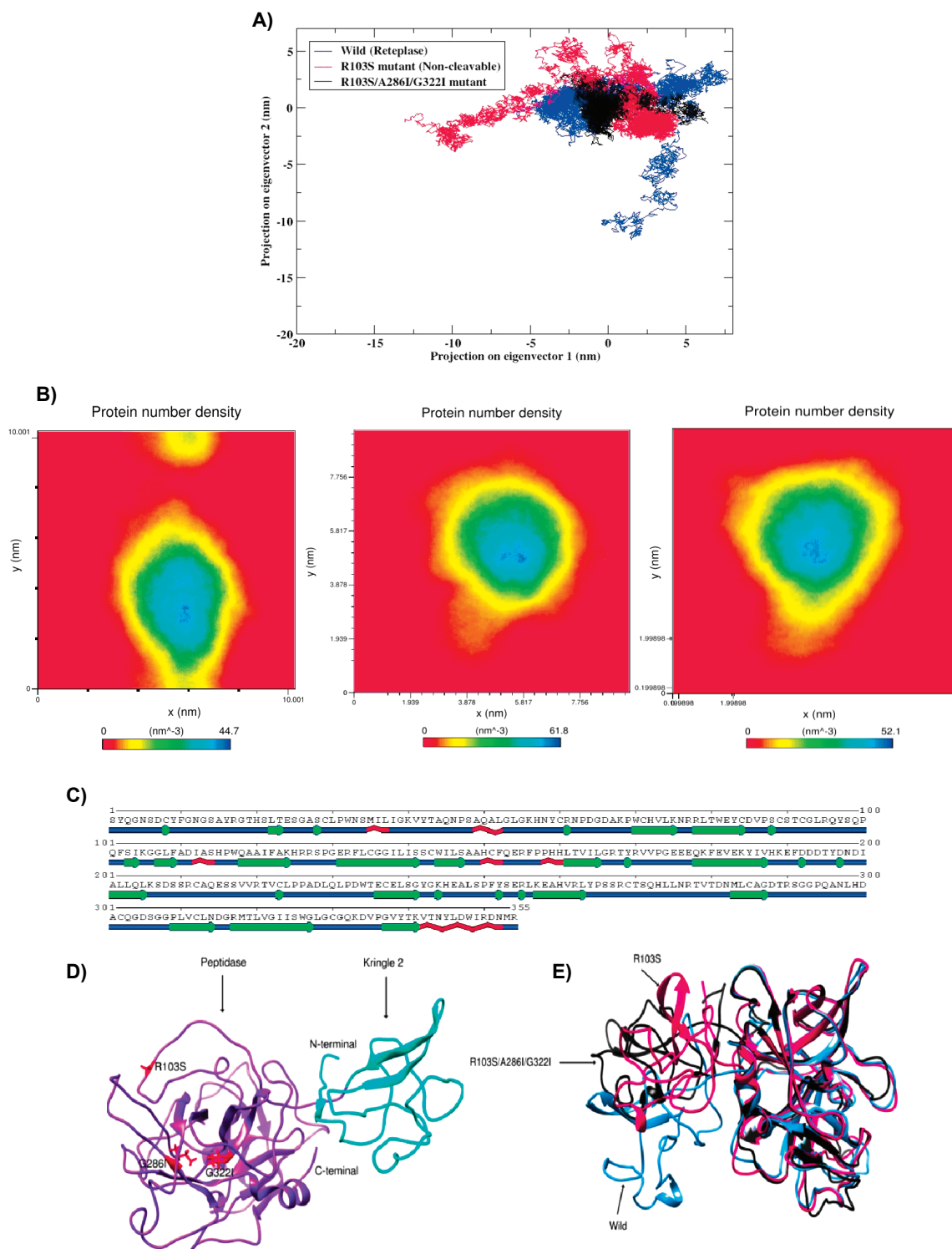
Figure 3. The projection of motion, density map, secondary structure, 3D structure and superimposition. A) The projection of motion of wild-type reteplase (blue), R103S mutant as a non-cleavable form (red), and R103S/A286I/ G322I mutant (black) structures in phase space along the first two principal eigenvectors at 300 K. **B)** The protein density plot of wild-type (left), non-cleavable form (middle), and R103S/A286I/G322I mutant (right) proteins. **C)** A simple secondary structure of r-PA depicted the Amino acid sequence witch indicated helixes, sheets, and loops. **D)** 3D structure of r-PA indicating its two domains, namely kringle 2 and the protease as well stabilizing mutations. **E)** Superimposition of structures illustrated the highest flexibility reduction in 1-80 residues of R103S/A286I/G322I mutant (black) as compared to wild-type (blue) and non-cleavable form (red) of r-PA.

88

Iran. J. Biotechnol. January 2023;21(1): e3175

4.4. PCA and Density Maps

PCA was performed to obtain a broader view of change in overall motion of protein, and also to support our MD simulation results (35, 39). The projection of the first two eigenvectors, containing the maximum motions, in conformational space for wild-type, non-cleavable form and R103S/A286I/G322I mutant, which was identified as the most stable structure, was done (**Fig. 3A**). The results demonstrated that the cluster of R103S/A286I/G322I mutant occupied less area in space and showed narrow variations in the conformational motion along the first two principle components, demonstrating its rigidity and compactness. Whereas wild type and non-cleavable form of r-PA were expanded in the conformational space due to flexibility. Furthermore, we examined the density of system as a function of a specified box vector. The average atomic density of R103S/A286I/G322I structure was 52.1 nm, while this value for the wild type and non-cleavable form of r-PA were 44.7 nm and 61.8 nm, respectively (**Fig. 3B left, middle & right**).the secondary structure of r-PA shown in **Figure 3**C and the tertiary structure of r-PA which indicating two cringle 2 and protease domains (**Fig. 3D**). in addition, Super imposition of structures of R103S/A286I/G322I mutant (black) as compared to wild-type (blue) and non-cleavable form (red) of r-PA illustrated the highest flexibility reduction in 1-80 residues in R103S/A286I/G322I mutant (**Fig. 3E**). Altogether, as a result of R103S/A286I/G322I mutations, the structure becomes more rigid than the wild type r-PA (**Fig. 3E**), which is evident from the PCA and density analysis.

4.5. DCCM and FEL

To gain further insight into the conformational changes, correlations between the dynamic motions of the intra- and inter-domains of reteplase were quantified through the dynamic cross-correlation matrix (DCCM). In comparison with the wild type and non-cleavable form of r-PA (**Fig. 4A left & middle**), negative correlative motions were found to weaken in the mutant structure (**Fig. 4A right**). This result suggests that mutation decreases the residue motions and the protein acquires rigid conformation. Specifically, more positive correlative motions were noticed in the kringle2 domain (residues 1–80) of the wild type structure, which is well supported by RMSF plot. To examine the sub conformational structural transitions of reteplase, we studied the Gibbs free energy landscape (FEL) against the first two principal components, namely PC1 and PC2, which describes how protein's free energy correlates with the three dimensional arrangement of the molecule. **Figure 4B** represents 2D and 3D views of the FEL analysis and ΔG value 0 to 15.3 15.0 and 14.0 kJ.moL^{-1} for wild type, non-cleavable form and mutant structure, respectively. The global energy minima conformations were designated by blue color. The size and shape of the minimal energy area indicates the stability of structure. Since the thermodynamically most stable structure resides in a minimum on the free-energy (ΔG) surface, more concentrated blue areas suggest more stability of the corresponding structures during MD simulation (40). The mutant showed more concentrated minima (in 2D) and less dispersed basins (in 3D) (**Fig. 4B right**), whereas the wild type and non-cleavable form of r-PA displayed more dispersed blue regions (**Fig. 4B left & middle**).

5. Discussion

Computational re-design of proteins is a promising strategy for improving the desired features of these biopharmaceuticals such as stability, affinity, solubility, and even production yield. In particular, there are numerous investigations for development of more stable therapeutic proteins via in silicon screening methods (41-43). In the same manner, we recruited computational approaches to design novel forms of r-PA with the improved conformational stability, which fairly correlates with protein's resistance to proteolysis. An appropriate 3D model is a prerequisite for any computational design of novel proteins. Therefore, we employed homology modeling approach for generating proper models of the wild type and mutant r-PAs. Protein structures were predicted using MODELLER which is a reliable software and widely utilized for comparative modeling (44). Modeling, energy minimization, and refinement processes were carried out meticulously to obtain high quality models which were confirmed by validation servers (**Table 2**). It has been well established that the r-PA proteolytic cleavage at Arg103 is a stability-limiting reaction. Accordingly, this cleavage site has been mutated to serine in all mutant forms studied in this work, because this substitution (R103S) confers resistance to the main stability-limiting reaction and also is less neurotoxic than the wild type (16, 17).

Iran. J. Biotechnol. January 2023;21(1): e3175

89

A)

B)
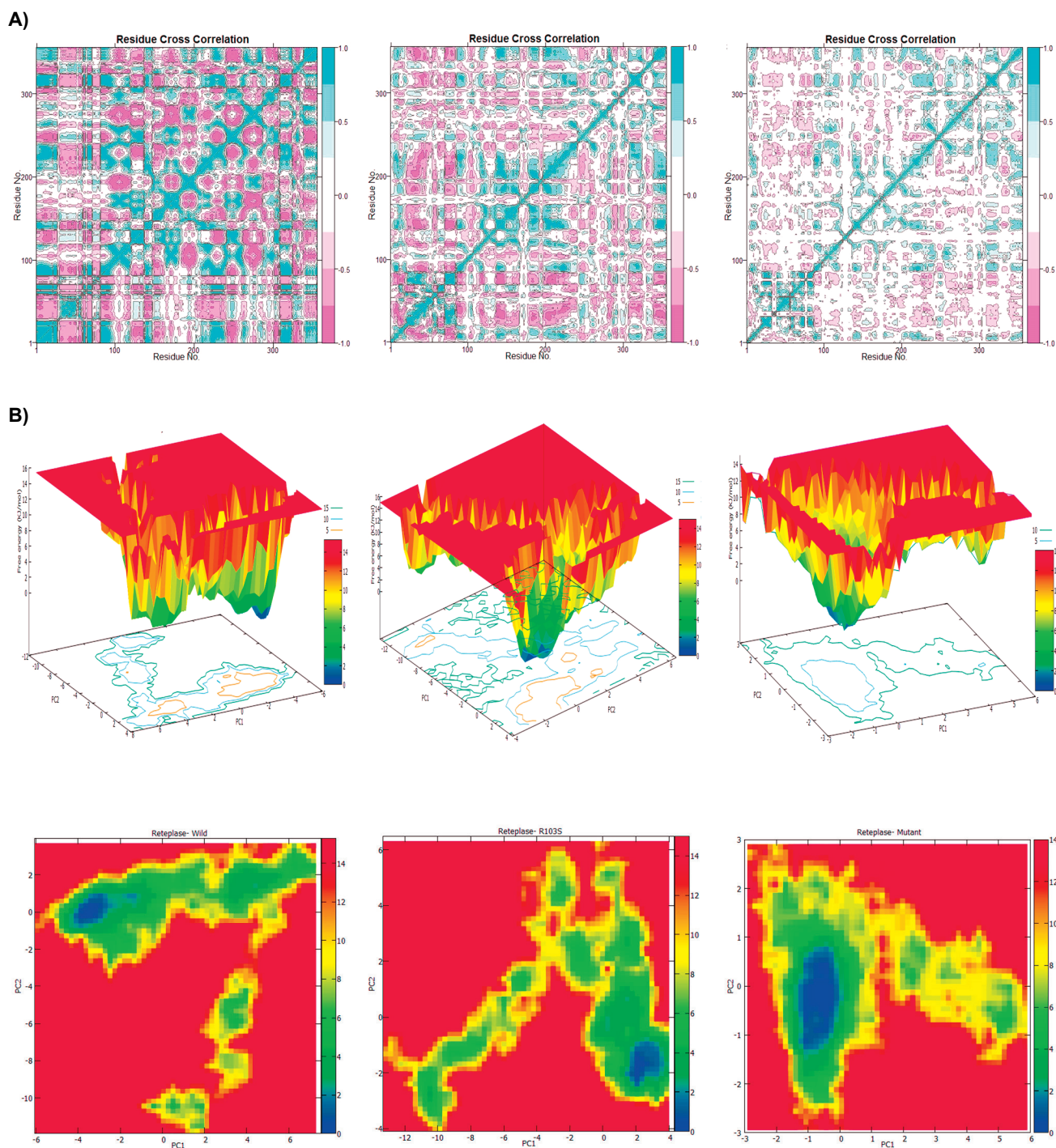
Figure 4. Cross-correlation matrix and free energy landscape. A) Cross-correlation matrix of the fluctuations of coordinates for Cα atoms, wild-type reteplase (left), non-cleavable form (middle) and R103S/A286I/G322I mutant (right) around their mean positions during the MD simulation: Cyan represents positive correlations, whereas magenta represents negative correlations. **B)** Gibbs free energy landscape (FEL) of wild-type reteplase (left), non-cleavable form (middle) and R103S/A286I/G322I mutant (right) along the two principal eigenvectors. The blue region depicts conformations with lower energy, the yellow color signifies the meta-stable states and the red corresponds to conformations having high energy.

Similarly, desmoteplase, an exclusively single chain form of t-PA, has been shown to have no interaction with NMDAR and consequently is devoid of neurotoxic effects compared to t-PA (19). Previous studies also have shown that mutations substituting serine for Arg (less hydrophobic properties) in the protein core could increase the compactness and enhance stability (24). For instance, circular dichorism (CD) study indicated that R157S lipase had increased thermostability with higher Tm value (71.6 °C) than its wild type (63.9 °C) showing a better compactness as revealed by spectrofluorescence study. Furthermore, MD results of R157S lipase showed lower values of RMSD, RMSF, SASA and Rg than native lipase, which corroborates with experimental data of CD (45).

In addition to the R103S substitution, other mutations were also introduced to r-PA in order to generate more stable forms of the enzyme. These mutations were selected based on the fact that hydrophobic interactions in the proteins core have a great impact on the stability of the protein (46), so, various mutations with this effect were created and studied using stability prediction servers, including SDM, mCSM and DUET. Then, considering ΔΔG and biophysical principals of substitutions for the improvement of the secondary structure stability, four more stabilizing mutations were selected including A201I, A286I, G322I, and G337I. Subsequently, all possible combinations of these four non-cleavable point mutants were generated and listed in **Figure 1A**. Overall, our mutant collection in this work is comprised of 16 non-cleavable members: the R103S mutant, four point mutants, six double mutants, four triple mutants and one quadruple mutant.

To understand the structural behavior of the wild type and mutant proteins, the molecular dynamics simulations were performed. Following mean factors, including RMSD, RMSF, solvent accessible SASA, number of intera-molecular hydrogen bonds (NH bonds), Rg and the density of system, which correspond to conformational stability were calculated. The advance analysis of the trajectories from MD simulation was performed by essential dynamics (ED) or PCA, which identifies the large concerted motions of protein (36). The compactness of proteins was analyzed by SASA and Rg values. Although Rg values were almost similar in the wild type and mutant proteins, decreased SASA values in mutants unveiled that these proteins were more compressed in comparison with the wild type and

non-cleavable forms (**Table 3**). Similar trend of Rg in wild type and mutants, despite decrease in SASA, was observed in the study of in silico stabilization of *Pseudomonas mendocina* Lipase (47) which is in agreement with the data obtained in the present study. Our findings of MD simulations revealed that R103S/A201I, R103S/A286I and R103S/A286I/G322I mutants have lesser SASA values than wild type and other mutants, which is attributed to the compactness in hydrophobic core of protein and less structural flexibility, giving rise to conformational stability. RMSD analysis is a well-known tool for providing a reliable estimation of the protein stability (48). According to the RMSD values tabulated in **Table 3**, the non-cleavable double mutant, R103S/A286/G322I, showed the lowest RMSD and deviation among all proteins (0.406 ± 0.062 nm) which indicates its noticeable stability during MD simulation. Also, based on the RMSF profiles, this mutant displayed the lowest average and standard deviation of RMSF among all mutants (0.120 ± 0.060 nm). Detailed analysis of the fluctuation of each residue showed that the highest decrease in the flexibility was occurred in the kringle2 domain.

On the basis of RMSD and RMSF observations as well as SASA analysis and the formation of less number of hydrogen bonds, it was confirmed that R103S mutation led to a more flexible conformation but removes the main susceptible site of proteolysis. The fewer propensities for proteolysis is advantageous to the production and extraction process of r-PA in protease-rich environments in various recombinant systems. In this study, the more flexible conformation and the consequent instability caused by R103S mutation in the protein structure was transparently compensated by introducing several mutations in the protein structure that improved parameters of conformational stability.

In fact, there is plenty of evidence in the literature indicating that structural properties such as conformational stability and flexibility can influence resistance to proteolysis. For instance, structural characteristics of nonspecific lipid transfer proteins explain their resistance to proteolysis (49). In the case of Apo A-I protein, flexibility is important factor on susceptibility to proteolysis (50). Moreover, investigations of limited proteolysis and MD simulation on Cu, Zn superoxide dismutase demonstrated that the high level of accessibility of the peptide bond is a necessary requirement but it is not sufficient to permit proteolysis. Thus, accessibility

must be coupled to flexibility to permit an efficient proteolytic cleavage (51). These evidences point to the fact that the introduction of mutations in the protein structure in such a way that enhance conformational stability will probably help to rationally design proteins that are less vulnerable to proteolysis during production and extraction in a wide range of recombinant systems. In this study, after MD simulations, the non-cleavable double mutant (R103S/A286I/G322I) was identified as the most stable structure based on the results of RMSD, RMSF and SASA values. The steady profile of Rg in R103S/A286I/G322I mutant showed that the protein was not undergoing an important structural transition. This finding was also supported by the similar trend of SASA profile. However, in contrast to the Rg profile in R103S/A286I/G322I mutant, we observed the major fluctuation in Rg in the wild type and the non-cleavable form of r-PA (**Fig. 2D**). In order to gain a comprehensive perspective of change in overall motion of proteins, their detailed conformational alterations, and also to corroborate our results, PCA and DCCM analysis were performed. PCA and density analysis revealed that the structure of R103S/A286I/G322I mutant was more rigid than the wild type r-PA (**Fig. 3C**). Moreover, this deduction was supported with the observation that negative correlative motions were found to weaken in the R103S/A286I/G322I compared to the wild type and R103S structures (**Fig. 4A**).

Reteplase, the recombinant clot-dissolving drug, is widely prescribed for the treatment of myocardial infarction. The main issue regarding the expression and production of this therapeutic protein by recombinant systems is its susceptibility to proteolytic degradation. Since conformational stability correlates remarkably to protein's resistance to proteolysis, several mutations including R103S, A201I, A286I, G322I, and G337I were employed in the reteplase structure to enhance its stability. The 3D structure of mutants were subjected to computational analysis to investigate the changes in parameters corresponding to structural flexibility and solvent accessibility. The introduction of R103S mutation was conducive to a more flexible conformation but removes the main susceptible site of proteolysis and produces non-cleavable single chain form of wild type r-PA. The other mutations, predicted by online servers, compensated the more flexible conformation caused by R103S mutation in the protein structure, and thus improved parameters of conformational

stability calculated by molecular dynamics simulation. Collectively, among all in silicon mutants, R103S/A286I/G322I had better structural stability (RMSD), noticeable reduction in flexibility (RMSF), more compactness (SASA), steadier profile of Rg and more rigid structure (NH bonds) than wild type r-PA. PCA and density analysis also indicated its stability. However, experimental studies and *in vitro* assessments are required for further structural and clinical validation of this in silicon mutant.

6. Conclusion
Findings of this study showed that selected mutations have fundamentally changed the stability of protein structure. These mutations will probably lead to more protection of r-PA in protease-rich environments in various recombinant systems and potentially enhance its production and expression level.

Author Contribution
K.H.Allahverdipoor, H.Eslami, S.Nasseri and M.J. Vavaran designed and performed Methods as well as drafted the manuscript. K.H.Alahverdipour, H.Eslami and S.Nasseri validate the methodology and tests. K.H.Alahverdipour, M.B.Khadem-Erfan, preparation of the manuscript. K.H.Alahverdipour, S.Nasseri, H.Eslami and M.J.Vavaran Scientific consultant. B.Nikkhoo and zh.Bahrami rad content validation and writing editor. K.H.Alahverdipour, H.Eslami, and S.Nasseri supervised the entire study. All authors revised and approved the final manuscript.

Conflict of interest
No conflict of interest to be declared by any of the authors.

References
1. Harter K, Levine M, Henderson SO. Anticoagulation drug therapy: a review. *West J Emerg Med*. 2015;**16**(1):11. doi: 10.5811%2Fwestjem.2014.12.22933
2. Cai Y, Bao J, Lao X, Zheng H, Chen J, Yu R. Computational

design, functional analysis and antigenic epitope estimation of a novel hybrid of 12 peptides of hirudin and reteplase. *J Mol Model*. 2015;**21**(9):1-9. doi: 10.1007/s00894-015-2774-2

3. Mohammadi E, Mahnam K, Jahanian-Najafabadi A, Sadeghi HMM. Design and production of new chimeric reteplase with enhanced fibrin affinity: a theoretical and experimental study. *J Biomol Str Dynamic*. 2021;**39**(4):1321-1333. doi: 10.1080/07391102.2020.1729865

4. Ghaheh HS, Ganjalikhany MR, Yaghmaei P, Pourfarzam M, Mir Mohammad Sadeghi H. Improving the solubility, activity, and stability of reteplase using in silico design of new variants. *Res Pharm Sci*. 2019;**14**(4):359-368. doi: 10.4103%2F1735-5362.263560

5. Gong L, Liu M, Zeng T, Shi X, Yuan C, Andreasen PA, *et al*. Crystal structure of the Michaelis complex between tissue-type plasminogen activator and plasminogen activators inhibitor-1. *J Biol Chem*. 2015;**290**(43):25795-25804. doi: 10.1074/jbc.M115.677567

6. Gao L, Zhang C, Li L, Liang L, Deng X, Wu W, *et al*. Construction, expression and refolding of a bifunctional fusion protein consisting of C-terminal 12-residue of hirudin-PA and reteplase. *Protein J*. 2012;**31**(4):328-336. doi: 10.1007/s10930-012-9407-8

7. Gurman P, Miranda O, Nathan A, Washington C, Rosen Y, Elman N. Recombinant tissue plasminogen activators (rtPA): a review. *Clin Pharm Therap*. 2015;**97**(3):274-285. doi: 10.1002/cpt.33

8. Schiermeyer A, Schinkel H, Apel S, Fischer R, Schillberg S. Production of Desmodus rotundus salivary plasminogen activator alpha1 (DSPAalpha1) in tobacco is hampered by proteolysis. *Biotechnol Bioeng*. 2005;**89**(7):848-858. doi: 10.1002/bit.20410

9. Kotb E. The biotechnological potential of fibrinolytic enzymes in the dissolution of endogenous blood thrombi. *Biotechnol Prog*. 2014;**30**(3):656-672

10. Adivitiya, Khasa YP. The evolution of recombinant thrombolytics: Current status and future directions. *Bioengineered*. 2017;**8**(4):331-358. doi: 10.1002/btpr.1918

11. Mousavi SB, Fazeli A, Shojaosadati SA, Fazeli MR, Hashemi-Najafabadi S. Purification and efficient refolding process for recombinant tissue-type plasminogen activator derivative (reteplase) using glycerol and Tranexamic acid. *Process Biochem*. 2017;**53**:135-144. doi: 10.1016%2Fj.procbio.2016.11.020

12. Magliery TJ. Protein stability: computation, sequence statistics, and new experimental methods. *Curr Opin Struct Biol*. 2015;**33**:161-168. doi: 10.1016/j.sbi.2015.09.002

13. Broom A, Jacobi Z, Trainor K, Meiering EM. Computational tools help improve protein stability but with a solubility tradeoff. *J Biol Chem*. 2017;**292**(35):14349-14361. doi: 10.1074/jbc.m117.784165

14. Nasab MA, Javaran MJ, Cusido RM, Palazon J. Purification of recombinant tissue plasminogen activator (rtPA) protein from transplastomic tobacco plants. *Plant Physiol Biochem*. 2016;**108**:139-144. doi: 10.1016/j.plaphy.2016.06.029

15. Zhuo X-F, Zhang Y-Y, Guan Y-X, Yao S-J. Co-expression of disulfide oxidoreductases DsbA/DsbC markedly enhanced soluble and functional expression of reteplase in Escherichia coli. *J Biotechnol*. 2014;**192**:197-203. doi: 10.1016/j.jbiotec.2014.10.028

16. Parcq J, Bertrand T, Baron A, Hommet Y, Anglès-Cano E,

Vivien D. Molecular requirements for safer generation of thrombolytics by bioengineering the tissue-type plasminogen activator A chain. *J Thromb Haemost*. 2013;**11**(3):539-346. doi: 10.1111/jth.12128

17. Goulay R, Naveau M, Gaberel T, Vivien D, Parcq J. Optimized tPA: A non-neurotoxic fibrinolytic agent for the drainage of intracerebral hemorrhages. *J Cereb Blood Flow Metab*. 2018;**38**(7):1180-1189. doi: 10.1177/0271678x17719180

18. Fredriksson L, Lawrence DA, Medcalf RL. tPA Modulation of the Blood-Brain Barrier: A Unifying Explanation for the Pleiotropic Effects of tPA in the CNS. *Semin Thromb Hemost*. 2017;**43**(2):154-168. doi: 10.1055%2Fs-0036-1586229

19. Lopez-Atalaya JP, Roussel BD, Levrat D, Parcq J, Nicole O, Hommet Y, *et al*. Toward safer thrombolytic agents in stroke: molecular requirements for NMDA receptor-mediated neurotoxicity. *J Cereb Blood Flow Metab*. 2008;**28**(6):1212-1221. doi: 10.1038/jcbfm.2008.14

20. Mohammadi E, Seyedhosseini-Ghaheh H, Mahnam K, Jahanian-Najafabadi A, Mir Mohammad Sadeghi H. Reteplase: Structure, Function, and Production. *Adv Biomed Res*. 2019;**8**:19-. doi: 10.4103%2Fabr.abr_169_18

21. Hidalgo D, Abdoli-Nasab M, Jalali-Javaran M, Bru-Martínez R, Cusidó RM, Corchete P, *et al*. Biotechnological production of recombinant tissue plasminogen activator protein (reteplase) from transplastomic tobacco cell cultures. *Plant Physiol Biochem*. 2017;**118**:130-137. doi: 10.1016/j.plaphy.2017.06.013

22. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2016;**54**(1):5.6.1-5.6. 37. doi: 10.1002/cpbi.3

23. Pace CN, Scholtz JM, Grimsley GR. Forces stabilizing proteins. *FEBS letters*. 2014;**588**(14):2177-2184. doi: 10.1016/j.febslet.2014.05.006

24. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, *et al*. Contribution of hydrophobic interactions to protein stability. *J Mol Biol*. 2011;**408**(3):514-528. doi: 10.1016/j.jmb.2011.02.053

25. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction.*Nucleic Acids Res*.2011;**39**(suppl_2):W215-W222. doi: 10.1093/nar/gkr363

26. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*. 2014;**30**(3):335-342. doi: 10.1093/bioinformatics/btt691

27. De P. ascher DB, Blundell tl. DUet: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res*. 2014;**42**:W314-W319. doi: 10.1093/nar/gku411

28. Childers MC, Daggett V. Insights from molecular dynamics simulations for computational protein design. *Mol Syst Des Eng*. 2017;**2**(1):9-33. doi: 10.1039/c6me00083e

29. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, *et al*. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct Funct Genet*. 2010;**78**(8):1950-1958. doi: 10.1002/prot.22711

30. Monzon AM, Zea DJ, Marino-Buslje C, Parisi G. Homology modeling in a dynamical world. *Protein Sci*. 2017;**26**(11):2195-2206.doi: 10.1002/pro.3274

31. Berendsen H, Hess B, Lindahl E, Van Der Spoel D, Mark A, Groenhof G. GROMACS: fast, flexible, and free. *J Comput*

Chem. 2005;**26**(16):1701-1718. doi: 10.1002/jcc.20291

32. 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. doi: 10.1016/j.softx.2015.06.001

33. Salsbury Jr FR. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr Opin Pharmacol.* 2010;**10**(6):738-744. doi: 10.1016/j.coph.2010.09.016

34. Touw WG, Baakman C, Black J, Te Beek TA, Krieger E, Joosten RP, *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 2015;**43**(D1):D364-D368. doi: 10.1093/nar/gku1028

35. Giuliani A. The application of principal component analysis to drug discovery and biomedical data. *Drug Discovery Today.* 2017;**22**(7):1069-1076. doi: 10.1016/j.drudis.2017.01.005

36. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Protein dynamics: Springer;* 2014. p. 193-226. doi: 10.1007/978-1-62703-658-0_11

37. Turner P. XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005

38. Steinbrecher T, Zhu C, Wang L, Abel R, Negron C, Pearlman D, *et al.* Predicting the effect of amino acid single-point mutations on protein stability—large-scale validation of MD-based relative free energy calculations. *J Mol Biol.* 2017;**429**(7):948-963. doi: 10.1016/j.jmb.2016.12.007

39. Kulshreshtha S, Chaudhary V, Goswami GK, Mathur N. Computational approaches for predicting mutant protein stability. *J Comput. Aided Mol Des.* 2016;**30**(5):401-412. doi: 10.1007/s10822-016-9914-3

40. Pontiggia F, Pachov D, Clarkson M, Villali J, Hagan M, Pande V, *et al.* Free energy landscape of activation in a signalling protein at atomic resolution. *Nat Commun.* 2015;**6**(1):1-14. doi: 10.1038/ncomms8284

41. Goldenzweig A, Fleishman SJ. Principles of Protein Stability and Their Application in Computational Design. *Annu Rev Biochem.* 2018;**87**:105-129. doi: 10.1146/annurev-biochem -062917-012102

42. Poorebrahim M, Sadeghi S, Ghorbani R, Asghari M, Abazari MF, Kalhor H, *et al.* In silico enhancement of the stability and activity of keratinocyte growth factor. *J Theor Biol.*

43. Kalhor H, Sadeghi S, Marashiyan M, Enssi M, Kalhor R, Ganji M, *et al.* In silico mutagenesis in recombinant human keratinocyte growth factor: Improvement of stability and activity in addition to decrement immunogenicity. *J Mol Graph Model.* 2020;**97**:107551. doi: 10.1016/j.jmgm.2020.107551

44. Ghavimi R, Mohammadi E, Akbari V, Shafiee F, Jahanian-Najafabadi A. In silico design of two novel fusion proteins, p28-IL-24 and p28-M4, targeted to breast cancer cells. *Res Pharm Sci.* 2020;**15**(2):200-208. doi: 10.4103/1735-5362.283820

45. Salleh AB, Rahim A, Rahman R, Leow TC, Basri M. The role of Arg157Ser in improving the compactness and stability of ARM lipase. *J Comput Sci Syst Biol.* 2012;**5**(2):39-46.

46. Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl Acad Sci.* 2006;**103**(35):13057-13061. doi: 10.1073/pnas.0605504103

47. Saravanan P, Dubey VK, Patra S. Emulating structural stability of Pseudomonas mendocina lipase: in silico mutagenesis and molecular dynamics studies. *J Mol Model.* 2014;**20**(11):1-12. doi: 10.1007/s00894-014-2501-4

48. Sadeghi S, Poorebrahim M, Rahimi H, Karimipoor M, Azadmanesh K, Khorramizadeh MR, *et al.* In silico studying of the whole protein structure and dynamics of Dickkopf family members showed that N-terminal domain of Dickkopf 2 in contrary to other Dickkopfs facilitates its interaction with low density lipoprotein receptor related protein 5/6. *J Biomol Struct Dyn.* 2019;**37**(10):2564-2580. doi: 10.1080/07391102.2018.1491891

49. Wijesinha-Bettoni R, Alexeev Y, Johnson P, Marsh J, Sancho AI, U. Abdullah S, *et al.* The structural characteristics of nonspecific lipid transfer proteins explain their resistance to gastroduodenal proteolysis. *Biochemistry.* 2010;**49**(10):2130-2139. doi: 10.1021/bi901939z

50. Sheridan DC, Cheng W, Carbonneau L, Ahern CA, Coronado R. Involvement of a heptad repeat in the carboxyl terminus of the dihydropyridine receptor β1a subunit in the mechanism of excitation-contraction coupling in skeletal muscle. *Biophys J.* 2004;**87**(2):929-942. doi: 10.1529/biophysj.104.043810

51. Falconi M, Parrilli L, Battistoni A, Desideri A. Flexibility in monomeric Cu, Zn superoxide dismutase detected by limited proteolysis and molecular dynamics simulation. *Proteins: Struct Funct Genet.* 2002;**47**(4):513-520. doi: 10.1002/prot.10094.