Mutated form (G52E) of inactive diphtheria toxin CRM197: molecular simulations clearly display effect of the mutation to NAD binding

Ramin Ekhteiari Salmas\textsuperscript{a,b}, Mert Mestanoglu\textsuperscript{c}, Ayhan Unlu\textsuperscript{d}, Mine Yurtsever\textsuperscript{a} and Serdar Durdagi\textsuperscript{b*}

\textsuperscript{a}Department of Chemistry, Istanbul Technical University, Istanbul, Turkey; \textsuperscript{b}Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey; \textsuperscript{c}School of Medicine, Bahcesehir University, Istanbul, Turkey; \textsuperscript{d}Faculty of Medicine, Department of Biophysics, Trakya University, Edirne, Turkey

Communicated by Ramaswamy H. Sarma

(Received 16 March 2015; accepted 8 November 2015)

Mutated form (G52E) of diphtheria toxin (DT) CRM197 is an inactive and nontoxic enzyme. Here, we provided a molecular insight using comparative molecular dynamics (MD) simulations to clarify the influence of a single point mutation on overall protein and active-site loop. Post-processing MD analysis (i.e. stability, principal component analysis, hydrogen-bond occupancy, etc.) is carried out on both wild and mutated targets to investigate and to better understand the mechanistic differences of structural and dynamical properties on an atomic scale especially at nicotinamide adenine dinucleotide (NAD) binding site when a single mutation (G52E) happens at the DT. In addition, a docking simulation is performed for wild and mutated forms. The docking scoring analysis and docking poses results revealed that mutant form is not able to properly accommodate the NAD molecule.

Keywords: CRM197; MD simulations; diphtheria toxin; binding interactions analysis

Introduction

Diphtheria was a main cause of death in humans over all countries until its vaccination was discovered (Malito et al., 2012). Even though diphtheria is no longer a common disease in developed countries, it remains as a large problem in other countries (Hadfield, McEvoy, Polotsky, Tzinerling, & Yakovlev, 2000; Golaz et al., 2000). Corynebacterium diphtheriae is indicated as a facultative agent of diphtheria by Vitek (2006). Roux and Yersin showed that an extracellular toxin (i.e. diphereria toxin (DT)) secreted by C. diphtheriae is responsible for toxicity (Malito et al., 2012). Generally, proteins employ the enzymatic transfer of the ADP ribose subfamily of nicotinamide adenine dinucleotide (NAD) to determine amino acids within a certain protein substrate (ADP-ribosylation) in order to adjust enzyme activity within the cell (Bell & Eisenberg, 1996). A group of this huge family of ADP-ribosyltransferase proteins includes DT, cholera toxin (CT), and pertussis toxin (PT) (Bell & Eisenberg, 1996). Detection of mutated states of DT was one of important steps in treatment of this disease (Malito et al., 2012). A positive antibody reaction is identified as immunological cross-reacting material (CRM). The main CRM determined was CRM197, an inactive enzyme and nontoxic form of DT that includes a single point mutation, G52E (Giannini, Rappuoli, & Ratti, 1984). Our laboratory focuses on investigation of molecular mechanisms of protein/drug and protein/protein interactions and optimization protocols for rational drug design (Durdagi, Papadopoulos, Papahatjis, & Mavromoustakos, 2007; Durdagi, Zhao, Cuervo, & Noskov, 2011; Fotakis et al., 2010; Leonis et al., 2014; Mavromoustakos et al., 2011; Salmas, Mestanoglu, Yurtsever, Noskov, & Durdagi, 2015; Salmas, Yurtsever, Stein, & Durdagi, 2015). In this study, we tried to investigate the influence of a single point G52E mutation on structural and dynamical profiles of DT and its effect on NAD binding on an atomic scale.

Methods

System setup

Crystal structure of nucleotide-free DT was retrieved from Protein Data Bank (PDB ID: 1SGK (Bell & Eisenberg, 1997)). The C domain of the structure was subjected to preparation and repairing processes using protein preparation module (Madhavi Sastry, Adzhigirey, Day, Annabhimoju, & Sherman, 2013) implemented in the Maestro molecular modeling package. The missing hydrogen atoms were completed, disulfide bond and the bond orders were checked. The pH was adjusted to physiological value of 7.4 using PROPKA module (Li, Robertson, & Jensen, 2005). OPLS-AA force field

*Corresponding author. Email: serdar.durdagi@med bahcesehir.edu.tr

© 2016 Informa UK Limited, trading as Taylor & Francis Group
(Jorgensen & Tirado-Rives, 1988) was employed for the initial total energy minimization of the protein structure. The constructed system was considered as a wild type of CRM197 and in addition to this form, a mutant form was manually assembled by modifying Gly to Glu at position 52 forming the active site of the protein (residues 38–52). The same protein preparation protocol was also used for mutant system.

**Molecular dynamic (MD) simulations**

The prepared proteins in the wild and mutant forms were hydrated using explicit TIP3P water model and neutralized by adding 0.15 M KCl using visual molecular dynamics (VMD) software (Humphrey, Dalke, & Schulten, 1996). In order to treat long-range electrostatic interactions during MD simulations, the periodic boundary conditions (PBC) and particle mesh Ewald (PME) method were used. Atomistic MD simulations in NPT ensemble at 310 K and by imposing periodic boundary conditions were performed using NAMD 2.9 (Phillips et al., 2005) software package and the CHARMM36 (MacKerell et al., 2001) force field for all-atom interactions. Prior to MD simulations, 10,000 steps for energy minimization and 2 ns restrained equilibration of the systems at fixed coordinates were performed. The long-range electrostatic interactions were treated using particle mesh Ewald (PME) method. The SHAKE algorithm was applied to constrain the hydrogen-heavy atom bonds, and the constraints at the atomic positions of protein atoms were then reduced gradually. Finally, 2 × 100 ns MD simulations for each system (wild and mutant) using different random seeds (because of chaotic behavior of initial velocity of simulations) were carried out as production run, under constant number of particles, pressure, and temperature (NPT) conditions at 310 K. 1 and 2 fs time steps were used in equilibration and production MD runs, respectively.

The derived MD trajectories were analyzed utilizing VMD, PyMOL (The PyMOL molecular graphics system), Bio3D (Grant, Rodrigues, ElSawy, McCammon, & Caves, 2006), and R programming language. In addition, a principal component analysis (PCA) was done on protein structures during MD simulations.

**Docking simulations**

Docking protocol of Glide (Friesner et al., 2004) implemented in Maestro molecular modeling package of Schrodinger was used to study the prediction of lowest energy docking poses and corresponding interaction energies of NAD molecule at the binding pockets of wild and mutant CRM197. We considered the derived clustered MD trajectory frames as targets of NAD in docking simulations, and then, the ensemble average of predicted binding energies and representative forms of docking positions were highlighted. It should be noted that before the docking of NAD, ligand preparation (i.e. protonation states, etc.) at physiological pH of 7.4 and geometry optimization of the chemical structure were performed.

**Results and discussion**

**Cognate study**

To determine whether the docking protocol implemented in the current study is to properly accommodate the ligand into the active site, we set up a cognate study. For this purpose, the crystal structure of DT dimer complexed with NAD was retrieved from PDB (PDB ID: ITOX (Bell & Eisenberg, 1996)). NAD molecule was removed from the active site and, ligand-free protein was targeted in docking simulation without using any constrain. Co-crystallized NAD molecule retrieved from the PDB file was afforded to dock into the protein, but prior docking, a conformational search protocol was set up for NAD molecule to initiate the docking using a low-energy conformer of the ligand. The docking simulation was carried out and the derived top-docking pose of NAD was superimposed with the one in the crystal structure. A 2.2 Å root mean square deviation (RMSD), (heavy atoms were considered) between docked pose and conformer of the ligand at the crystal structure is derived which is reasonable value to validate docking protocol used (Supplementary Figure S1).

**MD and molecular docking simulations**

Different distribution of initial velocities are used to initiate every two simulations for wild and mutant systems and the convergence of MD simulations is monitored using energy versus time and RMSD versus time graphs (Supplementary Figures S2, S3 and S4). In order to apply a comparative analysis on both wild and G52E targets, two representative snapshots (i.e. the frame that has the smallest RMSD to the average structure) picked from trajectory frames (10,000 frames for each) for the wild and mutant forms from derived MD trajectories are superimposed as shown in Figure 1. Active site is highlighted in MD snapshots. The importance of this binding site is indicated for access of NAD to the binding cleft and for elongation factor-2 (EF-2) recognition (Bell & Eisenberg, 1996). G52E mutation caused a high degree of deviation compared to the wild form in active site as shown in Figure 1. Their stereo views were also illustrated in Figure S5 at the supplementary materials. The loop of mutant form has more extended conformation compared to wild type. This extension contributed to an intra-loop H-bonds loss, as it will be discussed in next sections. A short helical part at the starting of the loop is
deformed during MD simulations in the mutant structure. The major factors of deformation of this critical site in the mutant CRM197 form are also investigated. In addition, model and crystal structures of the mutant and wild forms were superimposed based on their backbone atoms to illustrate their conformational deviations (Supplementary Figure S6). RMSDs between model and crystal structures are found as 1.16 and 1.51 Å for wild and mutant targets, respectively. In order to better understand the stability of the whole protein and especially the active-site loop, RMSD analysis on MD trajectory frames is carried out. RMSD plots for both wild and G52E targets during MD simulations are represented in Figure 2. It is clear that the mutation of Gly to Glu at position 52 leads to different structural orientations of Cα atoms of overall structure as well as at the active site. This structural change especially at the active site may influence the binding of NAD to CRM197. In order to better understand these different structural behaviors, RMSD histograms are also calculated and plotted in Supplementary Figure S7. These histograms clearly show that while a higher population was observed in greater RMSDs (i.e. RMSD, ~3.2 Å) for the mutant form, smaller RMSDs are observed for the wild form (RMSD, ~2.2 Å). In addition, root mean square fluctuations (RMSF) analysis was applied. As expected, active site has greater RMSF values in G52E CRM197 compared to its wild type throughout MD simulations. (Figure 3) Since this deformation and disstability at the binding site may lead to the loss of internal H-bonding, H-bond analysis on residues at this critical loop was also considered. A population analysis of H-bonds between residues which participate in H-bonding interactions at the active site was performed through 100 ns MD simulations (Figure 4). H-bond occupancy analysis showed a significant decrease in internal H-bond population in the mutant form between amino acids located in NAD binding site. While H-bonds between Lys51-Asn45, Gln32-Val28, and Ile31-Tyr27 were completely lost upon mutation, these H-bonds exist in the wild type during the simulations with different occupancy ratios (i.e. between 10 and 25%). Interestingly, a new H-bond between Gln36 and Gly34 amino acid residues was formed which may induce conformational changes in NAD binding site.
H-bond between Lys51 and Asp48 was in existence for both wild and mutated types during simulations. In total, as H-bond analysis described, G52E mutation has been influenced by the strength of intra-loop interactions and it leads to a less stable conformer. Thus, post-processing MD analyses helped us to better understand the structural differences of the wild and mutant forms that may lead to change in conformational dynamics and thus effecting the binding of the ligand. In order to understand the effect of mutation on correlation values between amino acid residues, dynamic cross-correlation (DCC) map analysis was carried out using MD trajectory frames (Supplementary Figure S8). It will be easily observed when the active site is more focused that while positive correlation motions are formed more at the wild type, higher number of anti-correlation motions is observed in the G52E mutated form. These results show the dis-stability at the mutant state in particular at the active site. Since anti-correlation becomes significant during the mutation, this may lead to inactivation of the system toward NAD binding.

Also, PCA was performed to check the linear motions on the mutant and wild forms of CRM197. 100 eigenvectors were used to calculate the relevant PCs for each motion. Supplementary Figure S9 displays PCA analysis results for both systems. Each dot in the figure indicates a single conformer which has retrieved from MD trajectory frames. Both systems were clustered and then profiled their deviations with respect to the first three PCs. First PCs involve 35.2 and 27.1% of whole harmonic motions for the wild and mutant systems, respectively. Second PCs indicate 10 and 13% of overall motion respect to wild and mutant systems. While there is a large jump between proportion of variances when passing from first to second or from second to third PCs.
in the mutant forms (i.e. differences of proportion of variances between PCs), corresponding values are much smaller in the wild case. Conformational clustering analysis shows distribution of different conformers for both systems based on PCA factor which have established during MD simulations. A comparative analysis between two systems indicates different established pathways for the wild and mutant forms. Mutant type underwent many different conformers rather than wild type as suggesting the higher mobility of the mutant form.

Figure 6. 2D ligand interaction diagrams for docked NAD into the wild and mutant targets.
Finally, we considered the derived clustered MD trajectory frames (10,000 frames from each run) as targets of NAD in docking simulations. The ensemble average of predicted binding energies and representative forms of docking positions were highlighted. Two docked poses were superimposed to compare derived docking poses. Distinct deviations were observed between docked NAD poses into wild and mutant proteins. (Figure 5) While binding pose of NAD in the mutant form was observed at the outside of the binding pocket, its docking poses in wild target were located at the active site. 2D diagrams of docking positions of NAD into the ligand binding sites of both systems are profiled in Figure 6. The main interactions such as hydrogen bonds, π–π stacking, and hydrophobic interactions are tracked between NAD and target structure. Upon mutation, binding pocket of NAD is affected; thus, it interacts with different amino acids which are indeed far from the real binding site of the ligand.

In order to estimate the interaction energy of NAD into the active site of wild and mutant systems, derived MD trajectories were clustered according to their RMSD values to the representative conformer and then among them, ensemble of conformers were formed and implemented in docking simulations. Glide docking protocol was used to estimate the interaction energies between NAD and mutant and wild targets. NAD revealed average binding energies of $-9.74 \pm 0.91$ and $-6.60 \pm 0.81$ kcal/mole against wild and mutant types, respectively. These results are in good agreement with experiments, as in principle, the mutant form could not accommodate the NAD molecule in the binding pocket, while the wild type is enzymatically so active in the presence of NAD molecule. Gly22 residue has made two strong hydrogen bonds with NAD from backbone in the wild form, while in the mutant type, it has not participated in any main interactions. Approximately, the same scenario is happened for His21, Gln36, Thy27, and Thr65 residues; π–π stacking and hydrogen bond interactions were established by participation of these critical amino acids to accommodate the NAD molecule into the wild form. Thus, this single mutation profoundly decreases the potentiality of the active site for binding NAD molecule.

**Conclusion**

In summary, we used long comparative MD simulations (two replicas are used for each wild and mutant forms) to better understand and clarify the structural and dynamical effects with G52E mutation on CRM197. In the light of our study, the influence of mutation on the whole and at the active site of protein is investigated and mechanistic differences between the wild and mutant forms are highlighted using post-processing MD analyses.
Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society, 110*, 1657–1666.

Leonis, G., Avramidopoulos, A., Salmas, R. E., Durdagi, S., Yurtsever, M., & Papadopoulos, M. G. (2014). Elucidation of conformational states, dynamics, and mechanism of binding in human κ-opioid receptor complexes. *Journal of Chemical Information and Modeling, 54*, 2294–2308.

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics, 61*, 704–721.

MacKerell, A. D., Banavali, N., & Foloppe, N. (2001). Development and current status of the CHARMM force field for nucleic acids. *Biopolymers, 56*, 257–265.

Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design, 27*, 221–234.

Malito, E., Bursulaya, B., Chen, C., Surdo, P. L., Picchianti, M., Balducci, E., … Spraggon, G. (2012). Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. *Proceedings of the National Academy of Sciences, 109*, 5229–5234.

Mavromoustakos, T., Durdagi, S., Koukoulitsa, C., Simcic, M., G. Papadopoulos, M. G., Hodoscek, M., & Grdadolnik, S. G. (2011). Strategies in the rational drug design. *Current Medicinal Chemistry, 18*, 2517–2530.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*, 1781–1802. doi:10.1002/(ISSN)1096-987X

Salmas, R. E., Mestanoglu, M., Yurtsever, M., Noskov, S. Y., & Durdagi, S. (2015). Molecular simulations of solved co-crystallized X-ray structures identify action mechanisms of PDE6 inhibitors. *Biophysical Journal, 109*, 1163–1168.

Salmas, R. E., Yurtsever, M., & Durdagi, S. (2015, August). Investigation of inhibition mechanism of chemokine receptor CCR5 by micro-second molecular dynamics simulations. *Scientific Reports, 5*, 13180. doi:10.1038/srep13180

Salmas, R. E., Yurtsever, M., Stein, M., & Durdagi, S. (2015). Modeling and protein engineering studies of active and inactive states of human dopamine D2 receptor (D2R) and investigation of drug/receptor interactions. *Molecular Diversity, 19*, 321–332.

The PyMOL molecular graphics system (Version 1.7.4). Schrödinger, LLC.

Vitek, C. R. (2006). Diphtheria. *Current Topics in Microbiology and Immunology, 304*, 71–94.