Computational and Structural Biotechnology Journal 17 (2019) 812–820

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Predicting the Most Stable Aptamer/Target Molecule Complex Configuration Using a Stochastic-Tunnelling Basin-Hopping Discrete Molecular Dynamics Method: A Novel Global Minimum Search Method for a Biomolecule Complex

Hung-Wei Yang [a], Shin-Pon Ju [b,c,*], Yu-Sheng Lin [b]

[a] Institute of Medical Science and Technology, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
[b] Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan.
[c] Department of Medicinal and Applied Chemistry, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

ARTICLE INFO

Article history:
Received 1 March 2019
Received in revised form 10 June 2019
Accepted 18 June 2019
Available online 20 June 2019

ABSTRACT

This study proposed a novel global minimum search method for predicting the most stable biomolecule complex, which combines the strengths of three global minimum search methods (stochastic tunnelling, basin hopping, and discrete molecular dynamics) to efficiently improve the spatial domain search ability of the stochastic tunnelling–basin hopping (STUN–BH) method from our previous study. The epithelial cell adhesion molecule (EpCAM, PDB code: 4MZV) was used as a benchmark target molecule for the EpCAM aptamer EpA (Apt_{EpA}). For the most stable Apt_{EpA}/EpCAM complex predicted by our new method, the Apt_{EpA} was attached to the entangling loop fragments of the two EpCAM molecules with the most Apt_{EpA} residues. After the Apt_{EpA}/EpCAM complex had equilibrated with the water environment through a molecular dynamics simulation at 300 K for 10 ns, stable hydrogen bonds formed between the bases of Apt_{EpA} and EpCAM residues of the secondary structures, which included the alpha helix and beta sheet becoming less stable in the water environment. Those hydrogen bonds formed between the bases of Apt_{EpA} and EpCAM loop fragment residues remained stable in the water environment.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

For predicting the most stable binding orientation of a drug molecule to a larger molecule with a known three-dimensional structure, a molecular docking simulation is a key process in computer-assisted drug design based on structural molecular biology [1–3]. For example, Chen used a docking simulation to explore the interaction mechanism of isorenieratene with human serum albumin (HSA) [4]. Available experimental results have shown two possible major binding sites (sites I and II) in HSA. However, through competitive ligand experiments, Chen indicated that isorenieratene could bind to both sites I and II, but that site II might be the main binding site. Therefore, Chen used a docking simulation to determine a possible binding configuration at site II that was in agreement with the experimental results. Grayzna [5] indicated that more butyrylcholinesterase (BChE) was observed in patients with Alzheimer's disease than in the general population. Isothermal titration calorimetry (ITC) has been used to confirm that ferulic acid and dihydrocaffeic acid could be used as inhibitors of BChE. However, experimental approaches still cannot clearly determine why these two compounds have a strong affinity with BChE. Grayzna's molecular docking simulation results indicated that both ferulic and dihydrocaffeic acids have a strong hydrophobic interaction with BChE and can be stabilised through hydrogen bond and pi–pi interactions with BChE.

For drug design, the structural prediction (docking simulation) of small drug molecules with large molecules having known three-dimensional structures is critical. Gareth employed the GOLD (Genetic Optimisation for Ligand Docking) program to verify the binding sites by using 100 benchmark protein-ligand complexes from Brookhaven Protein DataBank. GOLD is an automated ligand docking program that uses the genetic algorithm to explore the full range of ligand conformational flexibility and partial flexibility of protein. The results showed the GOLD program can accurately predict the ligand binding sites for 71 cases among 100 benchmark protein-ligand complexes [6]. In Alessandro's study [7], the metadynamics (MTD) algorithm was used to efficiently calculate the free energy and explore the reaction pathways during the system evolution. The MTD algorithm enables the system

* Corresponding author at: Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan.
E-mail address: jushin-pon@mail.nsysu.edu.tw (S.-P. Ju).

https://doi.org/10.1016/j.csbj.2019.06.021
2001-0370/© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

to escape from the local minimum easily during the search process, which can explore the spatial domain in a fast and rough way, but the results can still obtain the accurate free energy. Cerqueira indicated the flexibility of the receptor is generally not considered in the docking simulation, because the degrees of freedom are too large to completely go through the search domain. They developed an automated script, named MADAMM, to include the flexibility of receptors and ligands during docking simulation. The results show that the active site on the protein will be selected in the initial stage of the MADAMM method. Although this will affect the docking result of donor and receptor, but it also improves the limitation of rigid receptor in docking simulation [8]. Garrett used AutoDock4 docking simulation package to predict the binding free energies of small molecules to macromolecule targets. The force field is based on a comprehensive thermodynamic model that considers the incorporation of intramolecular energies for the predicted binding free energy. However, if the conformational space is wider, the free energy evaluation could lead to a great error. Therefore, the more accurate force filed is needed for docking simulation on the complex structures [9]. In Leonardo's study [10], they indicated most molecular docking programs can accurately predict the binding configuration of small-molecule ligand on the receptor binding sites. However, the docking score function still shows some inaccuracy problems when the solvent effects, entropic effects, and receptor flexibility are considered. In Paweł's study [11], they indicated the docking simulation can be used to predict the complex configuration of small molecule and protein, but the solvent effects on the prediction result is still a challenge for the docking simulation. For example, it is difficult for the docking simulation to predict what molecules are specific to the solvent molecules located at the binding sites, or which water molecules at the binding site can be replaced by ligands. Therefore, the results are more accurate if the binding affinity could be predicted directly through the MD in the explicit solvents. The docking simulation is mainly applied to small drug molecules docking with known larger biomolecules. If the docking simulation method used for small drug molecules is applied to a larger molecule (such as an aptamer), a more efficient method to consider the conformational change of molecular structure should be developed in the docking process because the spatial domain for a larger molecule is much wider than a small drug molecule.

Take aptamer molecules as an example; for the purpose of molecular recognition, aptamers are generally much larger than the drug molecules used to monitor the existence of a specific protein [12]. This is because the well-designed aptamer possesses high binding affinity to a specific protein; for example, Wang combined a single-stranded (ssDNA) aptamer with a multifunctional nanoparticle containing specific drug molecules [13]. Once the ssDNA aptamer recognised the particular tumour cell, the drug molecules within the nanoparticle could be precisely delivered into the tumour cell. In addition, an aptamer molecule can bind to its complementary strand containing a fluorescent signal to work as an aptamer sensor. Saberi [14] used a ssDNA aptamer attached by its complementary strand (hybrid sequence) with carbon dots to detect the existence of adenosine triphosphate (ATP). Once the ATP molecule was attached by the aptamer, the complementary strand with a weaker interaction strength compared with ATP immediately separated from the aptamer, leading to enhanced fluorescence intensity for the identification of ATP. In a study by Alshaer [15], the optimal sequence of ssDNA aptamer for epithelial cell adhesion molecule (EpCAM) identification was determined by the system evolution of ligands using exponential enrichment (SELEX) technology. This ssDNA aptamer successfully recognises EpCAMs such as ESA and CD326 [16,17]. Furthermore, it displays specificity for the identification of EpCAM-positive (KATO III) and -negative (NIH/3T3) cells. The sizes of the aforementioned aptamer molecules are much larger than those used in previous docking simulations.

For predicting the most stable binding orientation of a drug molecule or an aptamer molecule with a target molecule, both molecular docking and molecular recognition simulations must be used to conduct a global minimum binding energy search. Drug molecules are generally smaller than aptamer molecules, and thus a systematic search can be performed by randomly placing a drug molecule in a molecular docking system, and structural overlap between the drug and docked molecules during the molecular docking search can be easily avoided. However, for aptamer molecules, randomly placing one to change its orientation to the target molecule as well as to change its conformation simultaneously is unfeasible, because the overlap between molecules is not easily avoided. Consequently, most relevant studies have not utilised a systematic method for determining the most stable binding configuration of aptamers to the target protein. Only limited different initial binding configurations have been used to determine binding configurations with relatively lower energy instead of the most stable one. For example, Raheleh [18] proved retinol-binding protein 4 (RBP4) to be one of the biomarkers for type 2 diabetes (T2D); however, RBP4 had no structural information when it was bound to T2D. Therefore, the RBP4/T2D complexes with several different RBP4 conformations on T2D were simulated using a molecular dynamics (MD) simulation, during which the root mean square deviation (RMSD), RMS fluctuation (RMSF), and hydrogen bond interaction with RBP4 and T2D were used to monitor the local conformation change and final equilibrium configurations of the RBP4/T2D complexes. The results showed that the final equilibrium configurations of RBP4/T2D complexes depend on the initial RBP4 structures, implying that their MD results still could not predict the most stable structure of RBP4/T2D complexes as used in related experiments. Consequently, to prevent the MD simulation using an infinite number of different initial aptamer conformations with a target biomolecule, a systematic global minimum search method is first required to obtain the most stable configuration of the aptamer/target molecule before the MD simulation is conducted.

In this study, we improved the global search efficiency of the stochastic tunnelling–basin hopping (STUN–BH) method used in our previous study [19] by mutually using MD and discrete MD (DMD) [20] in the STUN process [21–23] to expand the spatial domain search. An EpCAM was selected as the target protein molecule for an available ssDNA aptamer, EpA [16]. A systematic search process using this new global minimum search method (named STUN–BH–DMD method) with the AMBER99sb force field [24,25] was used to determine the most stable adsorption configuration of EpA on EpCAM.

2. Simulation Model and the STUN–BH–DMD Method

2.1. Molecular Model

The EpCAM structure was constructed according to the PDB code 4MZV obtained from the Protein Data Bank (PDB) [17]. Fig. 1(a) presents a cartoon plot of EpCAM dimer containing two EpCAM molecules with a total atom number of 7610, in which the EpCAM dimer secondary structures (including an alpha helix, beta sheet, and loop) can clearly be observed. Furthermore, in the right panel of Fig. 1(a), the EpCAM dimer can be clearly seen to mainly form through the entanglement of loop fragments of two EpCAM molecules. SYL3C, the EpCAM aptamer discussed in Yanling's experimental study [16], can be truncated into three sequences; moreover, a fragment (named EpA) of SYL3C can maintain the same identification ability as SYL3C, and thus EpA was used as the aptamer molecule for EpCAM and was designated as Apt_{EpA}. Fig. 1(b) displays the molecular structure of Apt_{EpA} as well as its sequence (5'-ACA GAG GTT GCG TCT GT-3′). Apt_{EpA} is composed of 17 bases and possesses 543 atoms [26]. All molecular simulations were conducted using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package, and the AMBER99SB force field was used for describing the atomic interactions for the current Apt_{EpA}/EpCAM

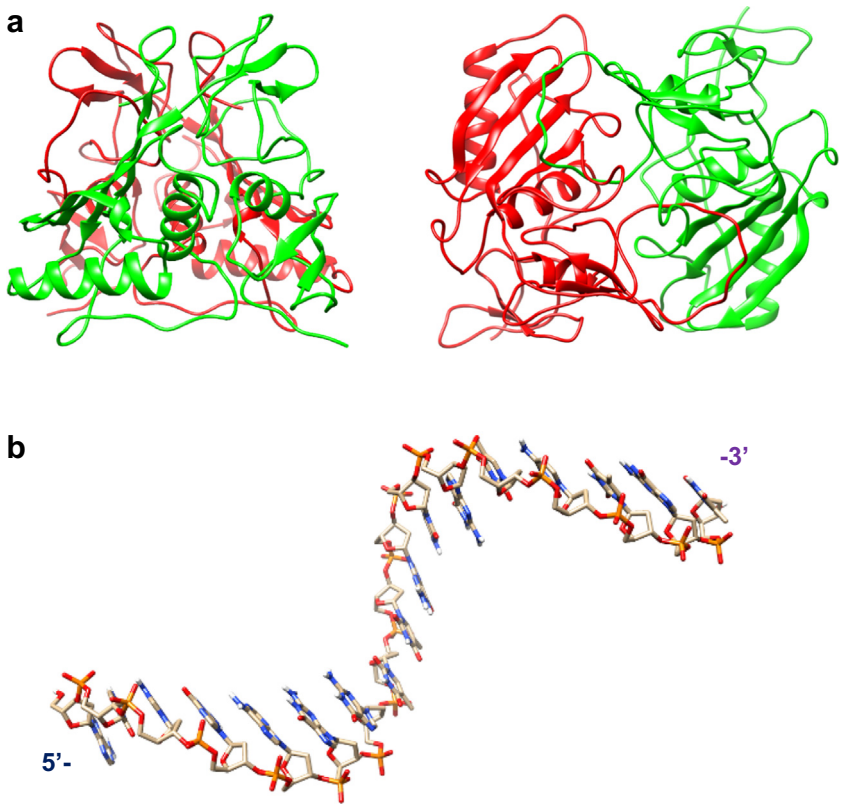

Fig. 1. (a) Illustration showing the dimer with the alpha helixes and beta sheets of EpCAM (PDB code: 4MZV). The left panel shows the entanglement of loop fragments of each EpCAM molecule to form the EpCAM dimer. (b) The molecular structure of EpA with the sequence 5'-ACA GAG GTT GCG TCT GT-3′.

system. The Chimera [23] packages were used to show the molecular structures as well as perform post-processing.

2.2. The STUN–BH–DMD Method

To determine the low-lying binding configuration of Apt_{EpA} on the EpCAM, the advantages of three global minimum search methods (STUN, BH, and DMD) were combined. Details of the differences between STUN-BH and STUN-BH-DMD methods are given in the supplementary file (Table S1 and Fig. S1). These three methods generally proceed through the two basic steps of calculating the energy and changing the atomic coordinates, as well as evaluating the system evolution using the Boltzmann factor. In the STUN method, MD at a higher temperature and with a larger integration time step were used to change the atomic coordinates. However, the closest local minimum structure after the MD step cannot be obtained using the original STUN method. Consequently, molecular statics using conjugate gradient (CG) optimisation in the BH method [27] were used to obtain the closest local minimum structure. In the original BH method for finding the most stable LJ nanocluster, the atom coordinates were randomly moved for the next search step. However, for the current Apt_{EpA}/EpCAM system, the random movements of atomic coordinates could cause the Apt_{EpA} and EpCAM to overlap; thus, the MD part of the STUN method was used to replace the random atom movement in the original BH method. During the MD process, Apt_{EpA} fragments that interacted more strongly with the EpCAM could undergo a very small orientation change compared with fragments with higher interaction energy. To extend the local spatial domain search, the DMD was also used with MD to change the atomic coordinates during the search process. DMD uses a step potential function to maintain repulsive interactions when the distance between two atoms is shorter than a threshold value as well as to significantly reduce the interaction strength between two atoms when the distance between them exceeds a threshold value. Although the advantage of DMD is to expand the spatial search domain, it could also destroy the stable local configuration built into the previous MD steps, and moreover, it cannot find a more stable configuration. Consequently, MD and DMD were iteratively used during the STUN–BH–DMD search process through which MD was conducted in a STUN–BH–DMD iteration, and DMD was performed in a subsequent STUN–BH–DMD iteration. By combining the advantages of MD and DMD, the STUN–BH–DMD method possesses the ability for a wide spatial domain search for the Apt_{EpA} orientation to the EpCAM as well as for keeping the local Apt_{EpA} fragment with relatively stronger energy with the EpCAM.

The effective potential $f_{STUN}(x)$ has the following formula:

$$E_{STUN}(x) = \ln\left((f(x)-f_0) + \sqrt{(f(x)-f_0)^2+1}\right) \quad (1)$$

where x stands for the coordinates of atoms, and f_0 is the lowest binding energy formed by the Apt_{EpA} bases and EpCAM amino acids obtained thus far. The value of f_0 in Eq. (1) is replaced when a lower binding energy (i.e., when the $E_{STUN}(x)$ value is negative) than the current f_0 value is found during the following STUN–BH–DMD search. Furthermore, $f(x)$ is the interaction energy between the Apt_{EpA} bases and EpCAM amino acids at the atom coordinates x. This effective potential preserved all minima locations lower than f_0, and the entire energy space from f_0 to the potential maximum was mapped onto the interval greater than 0 [28]. The binding energy G_{BE} was calculated according to the following Eq. (2):

$$G_{BE} = G_{EpA/EpCAM} - G_{EpA} - G_{EpCAM} \quad (2)$$

where $G_{EpA/EpCAM}$, G_{EpA}, and G_{EpCAM} are the potential energies of EpA/EpCAM complex, isolated EpA, and isolated EpCAM, respectively. The

effective potential energy of STUN shown in Eq. (1) converts the original potential energy surface (PES) into a smoother potential energy surface that allows the configuration to tunnel the forbidden regions for a wider spatial search. The effective potential energy surface still keeps the same local minimum structure as those within the original potential energy surface. Consequently, the BH part of STUN-BH-DMD method uses the conjugate gradient (CG) method to conduct the geometrical optimization of Apt_{EpA}/EpCAM complex. Then the binding energy of the optimized structure by AMBER99sb was used to obtain the effective energy by Eq. (1) for the Boltzmann factor evaluation.

The difference between DMD and traditional MD lies in the potential energy function of their interactions. In the original DMD, the step potential function was used for the hard sphere interaction, which makes atoms repulsive when they are close to each other; otherwise, no interaction was considered. Consequently, the atoms could have their relative coordinates changed more easily by DMD than by traditional MD. In the STUN–BH–DMD method, we did not use a step function for the DMD part, and the interaction strengths of nonbonding interactions (including the Lennard–Jones and Coulombic interactions) were reduced to 1/100 times smaller than their original values. Thus, the interaction between atoms works in a highly similar manner to those using a step potential function.

A total of 1700 independent initial conformations with different Apt_{EpA} orientations to the EpCAM were considered for enhancing the global spatial domain search. In order to generate 1700 different Apt_{EpA} orientation to EpCAM in the vacuum system, the EpCAM was randomly rotated at its center of mass for 1700 times, while the Apt_{EpA} keeps the same coordinates for each EpCAM random rotation. 300 STUN–BH–DMD iterations were performed for each initial orientation configuration. A total of 300,000 STUN–BH–DMD iterations are performed in the process to facilitate the search for the most stable adsorption configuration between Apt_{EpA} and EpCAM. For each STUN–BH–DMD iteration, CG optimisation (as used in the BH method [27]) was conducted by following the NVT MD or DMD simulation at 600 K for 300 steps (0.6 ps) to extend the local spatial search domain. The energy of the optimized structure using the CG method was used as $f(x)$ for Eq. (1), and thus the subsequent interaction between Apt_{EpA} and EpCAM can be performed through STUN to find a lower energy configuration. If the value of $f_{STUN}(x)$ is negative, it means that $f(x)$ is smaller than f_0 and the energy value of $f(x)$ is stored as f_0. The structure is preserved and the STUN–BH–DMD process evolves in the direction of the lower energy. If the current $E_{STUN}(x)$ is greater than the value of $STUN_{last}$ (the last $E_{STUN}(x)$ stored after Boltzmann factor evaluation) recorded by the system, it will be determined by generating a random number from 0 to 1 as the acceptance ratio and calculating the Boltzmann factor. If the value of the Boltzmann factor is greater than the receiving ratio, it is accepted; otherwise, the current structure is skipped, and the preferred configuration searched for in the previous one is advanced. The Boltzmann factor has the following formula:

$$\text{Boltzfactor} = \exp\left[\frac{STUN_{last} - E_{STUN}(x)}{kT}\right] \quad (3)$$

1. If $E_{STUN}(x)$ is lower than or equal to $STUN_{last}$, this STUN–BH–DMD step is accepted. The $STUN_{last}$ is renewed by the current $E_{STUN}(x)$ and xlast (the atom coordinates at $STUN_{last}$) is also replaced by the current atom coordinates.
2. If $E_{STUN}(x)$ is greater than $STUN_{last}$, a random number between 0 and 1 is generated and the Boltzmann factor is determined according to Eq. (3). If the Boltzmann factor is greater than the random number, this STUN–BH–DMD step is accepted. The $STUN_{last}$ is renewed by the current $E_{STUN}(x)$ and xlast is also replaced by the current atom coordinates.
3. If $E_{STUN}(x)$ is greater than $STUN_{last}$ and the corresponding Boltzmann factor is smaller than the random number generated between 0 and 1, this STUN–BH–DMD step is rejected. The system atom coordinates are replaced by xlast for the next STUN–BH step.

The value of kT shown in Eq. (3) was dynamically adjusted every 20 STUN–BH–DMD steps to maintain the acceptance percentage close to 50%. The STUN–BH–DMD method can significantly guide the system to evolve in the direction of lower energy.

Fig. 2 summarises the global minimum search process using the STUN–BH–DMD method. The binding energy formed between the Apt_{EpA} bases and EpCAM was first converted to the effective energy according to the function in the STUN method. The initial configuration

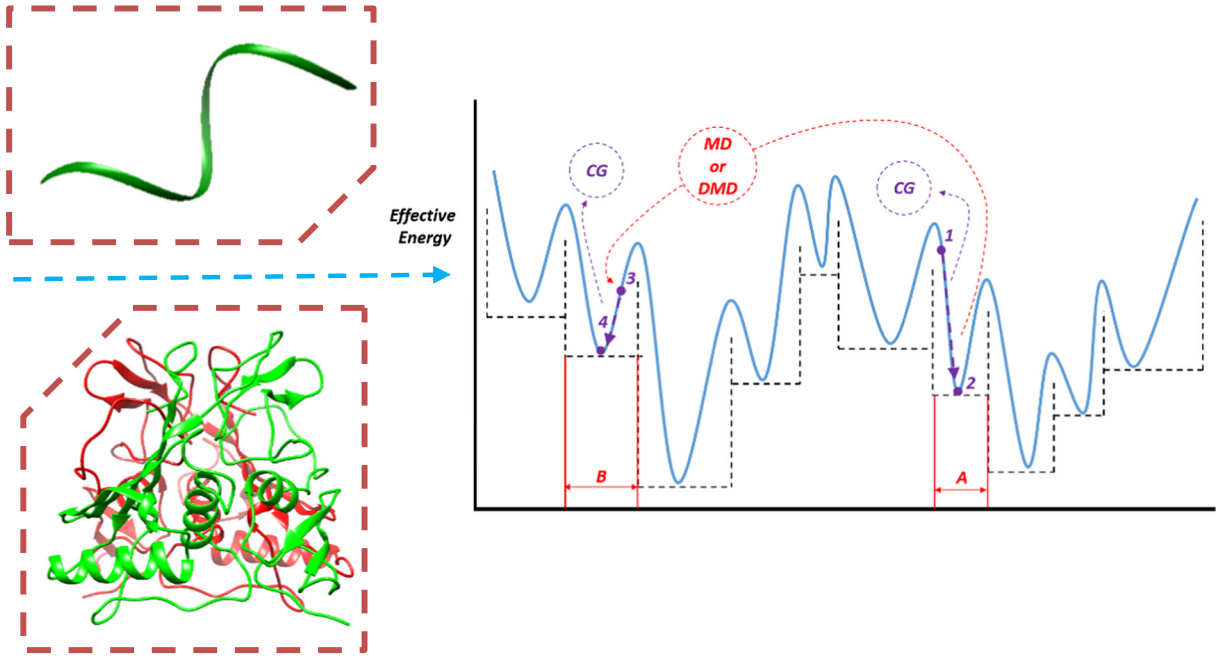

Fig. 2. Schematic of the STUN–BH–DMD method. A and B represent two energy basins (Basin A and Basin B) of the effective potential energy surface; Points 1 and 3 stand for the structures with the lowest effective potential energies at Basin A and Basin B obtained by CG method, respectively. Points 2 and 4 are the structures located at Basin A and Basin B, respectively.

was assumed to correspond to the effect energy at Point 1 within Basin A, and then molecular statics using the CG method (as adopted in the BH method) were used to quickly find the nearest local minimum structure at Point 2 within Basin A, leading to the system evolving from Point 1 to Point 2, seen in Fig. 2. Using the CG method, all structures within Basin A corresponded to the same local minimum structure at Point 2 of Basin A. This indicated that using the CG method can convert the effective energy of each basin into the flat PES, as shown by the dashed horizontal line.

Next, MD or DMD were performed for 300 steps for the Apt_{EpA} configuration at Point 2 to generate more Apt_{EpA} orientations for the target EpCAM to jump to another basin, as shown in Fig. 2, where the configuration at Point 2 in Basin A changed to that at Point 3 in Basin B. Subsequently, the local minimum structure at Point 4 in Basin B could be quickly found using the CG method. Many local minimum structures in different basins can be quickly obtained by repeating the above mentioned steps and compared with the previously calculated values of the stable structure using the Boltzmann factor. The configuration search is guided towards the direction of lower energy, and finally, the most stable structure (global minimum structure) can be found.

The most stable Apt_{EpA} configuration on the EpCAM obtained using the STUN–BH–DMD method was then used to observe the dynamic behaviour of the Apt_{EpA}/EpCAM complex in the water environment by MD simulation. The Apt_{EpA}/EpCAM complex was first placed into the simulation box, within which the complex is about 30 Å from its PBC image for each dimension. Then the water molecules are randomly inserted into the system until the system density is close to 1 g/cm^3. Approximately 80,000 water molecules were randomly inserted into the simulation box with a closest distance of 5 Å allowed between any atoms of Apt_{EpA} and EpCAM. Fourteen Na^+ ions were also randomly inserted into the system for neutralising the simulation system. After the CG minimisation, the system underwent the canonical ensemble (NVT) in the Nose–Hoover method at 300 K for 10 ns. A cut-off distance of 10 Å was used to calculate the van der Waals interactions and the particle-particle particle-mesh (PPPM) method was used for the long-ranged interactions.

3. Result and Discussion

To use the STUN-BH-DMD method for the water environment directly, a large amount of water molecules has to use in the simulation box. The required computational capacity becomes several orders higher than that for the model in a vacuum. In several previous docking studies [1,4,18,29–31], the molecular docking method was used to find the best docking configuration formed by the donor and acceptor using the evaluation of the scoring function. After the docking process, the complexes were then used in the water environment for a long-time MD simulation to investigate the stability of the complexes. Consequently, a similar process was used for making the STUN-BH-DMD method feasible. Fig. 3 presents the configurations of the Apt_{EpA}/EpCAM complex with the lowest binding energy in a vacuum after the STUN–BH–DMD search. In Fig. 3(a), most Apt_{EpA} are attached to surface loop fragments by which two EpCAM molecules are linked to a EpCAM dimer. Almost no Apt_{EpA} are attached to the alpha helical or beta sheet fragments of the EpCAM because the hydrogen bonding network between the residues of the same alpha helical or beta sheet fragments makes their local structures more stable than other parts of the EpCAM, resulting in a much lower probability of hydrogen bonds forming with the Apt_{EpA}. In Fig. 3(b), from the inverse viewpoint of Fig. 3(a), the Apt_{EpA} prefers to bind on the side of the EpCAM with more loop fragments, which are because the most secondary structures of EpCAM (including alpha helixes and beta sheets) can be clearly seen. In Fig. 3(c), the middle segment of Apt_{EpA} does not seem to have direct contact with EpCAM, mainly because of the three-dimensional barrier of EpCAM.

Fig. 4 shows the equilibrated configuration of the Apt_{EpA}/EpCAM complex after an MD simulation at 300 K for 10 ns. Fig. 4(a)–(c) shows that the Apt_{EpA} attached to highly similar surface areas of EpCAM, as shown in Fig. 3(a)–(c), indicating that the Apt_{EpA}/EpCAM complex obtained using the STUN–BH–DMD method under a vacumm condition remains stable in the water environment. The corresponding binding energies between Apt_{EpA} and EpCAM in vacuum and water conditions are listed in Table 1. In the case of the vacuum, the binding energy was sampled using an MD simulation at 10 K for 100 ps, whereas the binding energy in the case of water was sampled for the last 1 ns after the MD simulation was conducted at 300 K for 9 ns. Table 1 shows that the Apt_{EpA}/EpCAM binding energies in vacuum and water were approximately −568.77 kcal/mol and − 433.44 kcal/mol, respectively. The average binding energy of Apt_{EpA}/EpCAM at 300 K in vacuum was also obtained by the MD simulation for comparison with that in the water environment at 300 K. For the vacuum case, the averaged binding energy at 300 K is about −553.99 kcal/mol, which is 14.78 higher than that listed in Table 1. In the water environment, the average binding energy listed in Table 1 becomes further higher, indicating the decreasing binding energy of Apt_{EpA}/EpCAM in water environment was due to

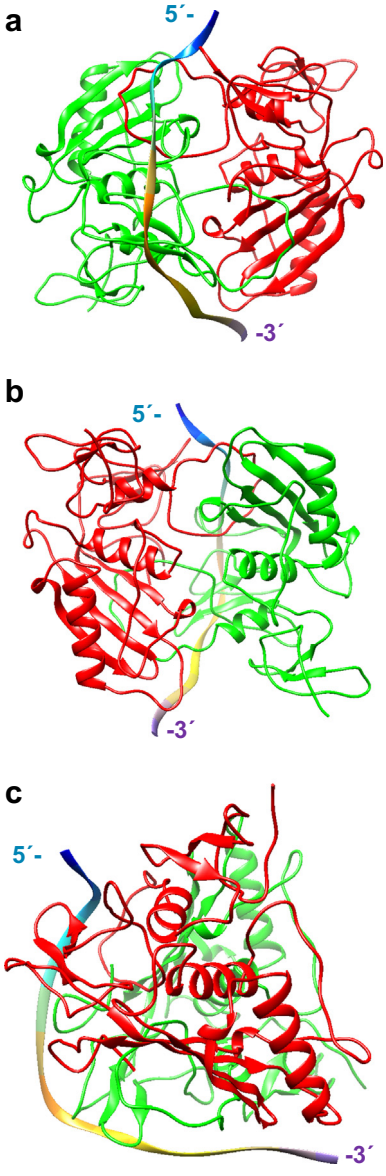

Fig. 3. Morphology of the most stable Apt_{EpA}/EpCAM complex in the vacuum from three different viewpoints.

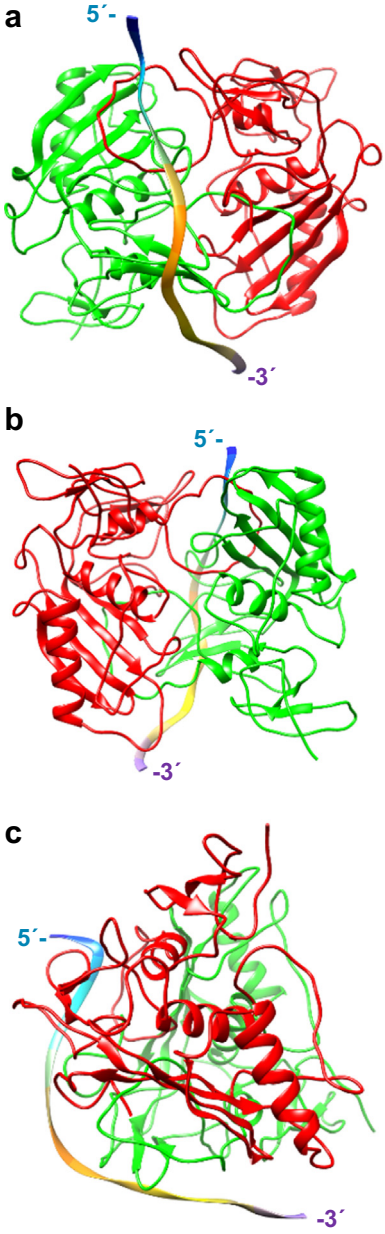

Fig. 4. Morphology of the equilibrated Apt_{EpA}/EpCAM complex in the water environment from three different viewpoints.

interactions with water molecules. Consequently, the Apt_{EpA}/EpCAM binding energy in the water environment became weaker than that in the vacuum by 23.8%.

For investigating the interaction strength between each EpCAM residue and each Apt_{EpA} residue, the upper panels of Fig. 5(a) and (b) present interaction energy contour maps for each EpCAM residue to each Apt_{EpA} residue in a vacuum and in water. For the interaction strengths of EpCAM to individual Apt_{EpA} residues, the lower panels of Fig. 5 (a) and (b) present histogram plots for the energies of EpCAM to individual residue of Apt_{EpA}. In the contour maps with the energy scale bar in kcal/mol, the vertical axis stands for the sequence of the EpCAM amino acid index, whereas the horizontal axis presents the 17 bases of Apt_{EpA} in sequence. According to the contour maps diagram in Fig. 5 (a), the Apt_{EpA} can be seen to stabilise on the EpCAM surface by simultaneously interacting with multiple amino acids, as indicated by dashed rectangles. The histogram plot in Fig. 5(a) shows that two Apt_{EpA} fragments (2–7 and 11–17) interacted relatively more strongly with the EpCAM. By contrast, the fragment 8–10 interacted very weakly with the EpCAM because this fragment at the middle of Apt_{EpA} almost makes no contact with the EpCAM, as seen in Fig. 3(c). In the water environment, the weaker averaged binding energy in Table 1 can be explained by the energy contour map in Fig. 5(b). The areas with stronger interaction energy can be seen to become narrower than those in Fig. 5(a). The histogram plot in Fig. 5(a) also shows that the Apt_{EpA} nucleotide interacts relatively more weakly with the EpCAM than those in Fig. 5(a). The energies shown in Fig. 5 are the interaction strength between the Apt_{EpA} nucleotide and EpCAM. Because the sequence of ssDNA aptamers has a critical influence on the recognition ability for a target molecule, the interaction between the bases of Apt_{EpA} and the EpCAM could also play a crucial role. Consequently, the histogram plot of Fig. 5(a) marks the five bases in orange (DG7, DC11, DG12, DC14, and DG16), which possess the five strongest energies with the EpCAM. In the vacuum, the binding energy Apt_{EpA} near the 5′ terminal can be seen to be mainly formed by the Apt_{EpA} backbone atoms with the EpCAM, whereas the interaction between the Apt_{EpA} bases and EpCAM is dominant near the 3′ terminal. In the water environment, the histogram plot of Fig. 5(b) indicates that the strongest five bases are still the same as those in the vacuum system, indicating that the STUN–BH–DMD method can determine the same base adsorption positions in the water environment as those in the vacuum.

The strongest five base interaction energies of DG7, DC11, DG12, DC14, and DG16 in the vacuum and in water are listed in Table 2 in descending order of interaction strength. Table 2 shows that the base interaction strength order in the vacuum was different to that in water. The interaction of DC14 in water was stronger than that in the vacuum by 15.4%, whereas the other four bases in water interacted much more weakly with the EpCAM than the corresponding ones in the vacuum, which reflects the average binding energies shown in Table 1. The STUN-BH-DMD simulation proposed the EpCAM protein may strongly interact with the Apt_{EpA} fragment near the 3′ terminal not the 5′ terminal. To prove the concept of STUN-BH-DMD simulation, we modified the thiol group at 5′ terminal (SH-$(CH_2)_3$–5′-ACA GAG GTT GCG TCT GT-3′; named 5′-Apt_{EpA}) and 3′ terminal (5′-ACA GAG GTT GCG TCT GT-3′-$(CH_2)_3$-SH; named 3′- Apt_{EpA}) then the 5′-Apt_{EpA} and 3′-Apt_{EpA} were immobilized on the well of 96-well plate through SH-group for EpCAM capture. After that, EpCAM were added into each well modified with 5′-Apt_{EpA} or 3′-Apt_{EpA} on the bottom and incubated for 1 h at 25 °C. The remaining EpCAM in supernatant was analyzed using 10% SDS-PAGE, showing more EpCAM can be capture on the well modified with 5′-Apt_{EpA} than that on the well modified with 3′-Apt_{EpA}, suggesting that functional groups should be better to modified at the 5′ end terminal for immobilization of Apt_{EpA} on the well (Fig. S2). The results indicated that the STUN-BH-DMD simulation can help forecast the best binding region of Apt_{EpA} with EpCAM protein to construct an ideal biosensor with high sensitivity and wide detection range.

To understand why the base energies in a vacuum significantly change in the water environment, Fig. 6(a) and (b) show the hydrogen bonding networks of DC11 and DC14 with their closest EpCAM residues. The base energy of DC11 decreases the most significantly (from −86.283 to −38.341 kcal/mol) and the base energy of DC 14 slightly increases in the water environment (from −82.530 to −97.509 kcal/mol). The hydrogen bonds formed between the DC11 (DC14) and EpCAM are indicated by solid lines. In Fig. 6(a), several hydrogen bonds form between DC11 and residual LYS84. However, LYS84 is one compositional residue of beta sheets that mainly forms through the hydrogen network of compositional residues and is more conservative in terms of changing

Table 1
Average binding energies between Apt_{EpA} and EpCAM in vacuum and water environments.

	Energy (kcal/mol)
Vacuum	−568.77
Water	−433.44

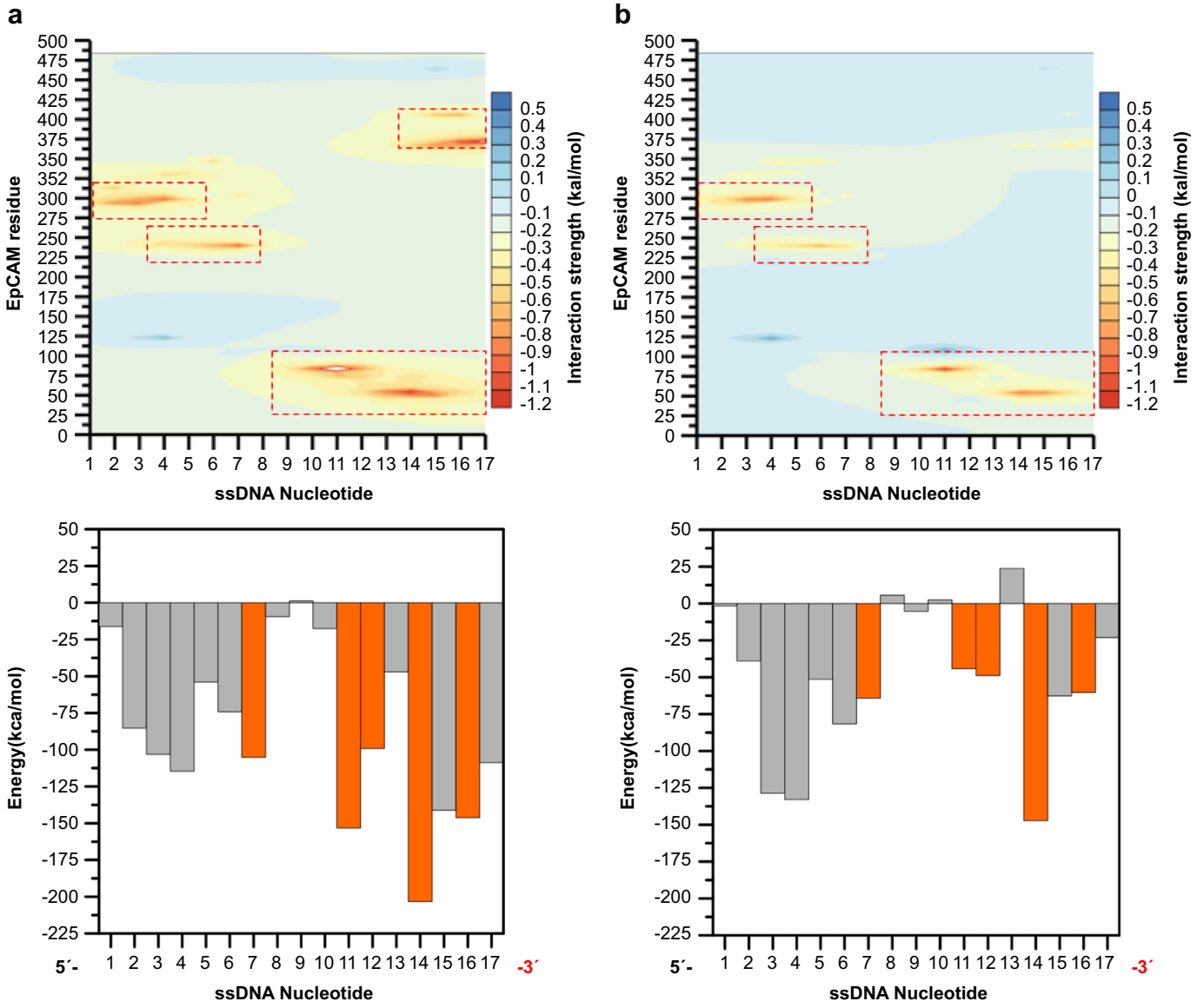

Fig. 5. Upper panel: Contour map showing the interaction energy distribution of each Apt_{EpA} nucleotide to each EpCAM residue. Lower panel: Histogram plot showing the interaction energy between the EpCAM and an Apt_{EpA} nucleotide in (a) the vacuum environment and (b) the water environment. The five nucleotides marked in orange indicate that the interaction energies of their bases are the five strongest ones among the bases of all nucleotides. The nucleotides marked in gray indicate their bases interact relatively weaker with the EpCAM.

its local structure. Consequently, the hydrogen bonds formed between the DC11 base and LYS84 cannot be maintained in the water system, indicating that the amino acids of the alpha helix and beta sheet do not easily form stable hydrogen bonds with bases of Apt_{EpA}. In Fig. 6(b), several hydrogen bonds form between DC14 and EpCAM residues, including SER52, LYS53, GLN50, and ASP72 in vacuum, and all these EpCAM residues belong to the loop fragment. In the water environment, DC14 still forms stable hydrogen bonds with SER52 and LYS53. Two loop residues, MET49 and LEU54, also form hydrogen bonds with DC14 in the water environment, resulting in the increasing base binding energy shown in Table 2.

Table 2
Five strongest interaction energies between the Apt_{EpA} bases and EpCAM in vacuum and water environments.

Vacuum	DC11	DC14	DG16	DG7	DG12
Base binding energy (kcal/mol)	−86.28	−82.53	−82.31	−79.55	−77.42
Water	DC14	DG7	DG16	DC11	DG12
Base binding energy (kcal/mol)	−97.50	−54.05	−47.61	−38.34	−35.58

To determine the stability of EpCAM for the EpCAM/Apt_{EpA} complex at room temperature, an MD simulation at 300 K was conducted for 10 ns. During the simulation, variations in RMSD [23–25] and the radius of gyration Rg [26,27] of EpCAM were used to monitor whether the structure of EpCAM had become stable. The definition of Rg can be expressed using Eq. (4):

$$R_g(t) = \left[\sum_{i=1}^{N} m_i (r_i(t) - r_{COM}(t))^2 \right]^{\frac{1}{2}} \quad (4)$$

where m_i is the mass of the i-th atom; $r_{COM}(t)$ stands for the centre of mass of the EpCAM at timestep t; and $r_i(t)$ indicates the coordinates of atom i at timestep t. For a molecule, a larger value of R_g indicates a more expansive structure, whereas a smaller value of R_g indicates the opposite. In the current study, the Rg was only used to monitor whether the structure of the EpCAM had become stable instead of evaluating

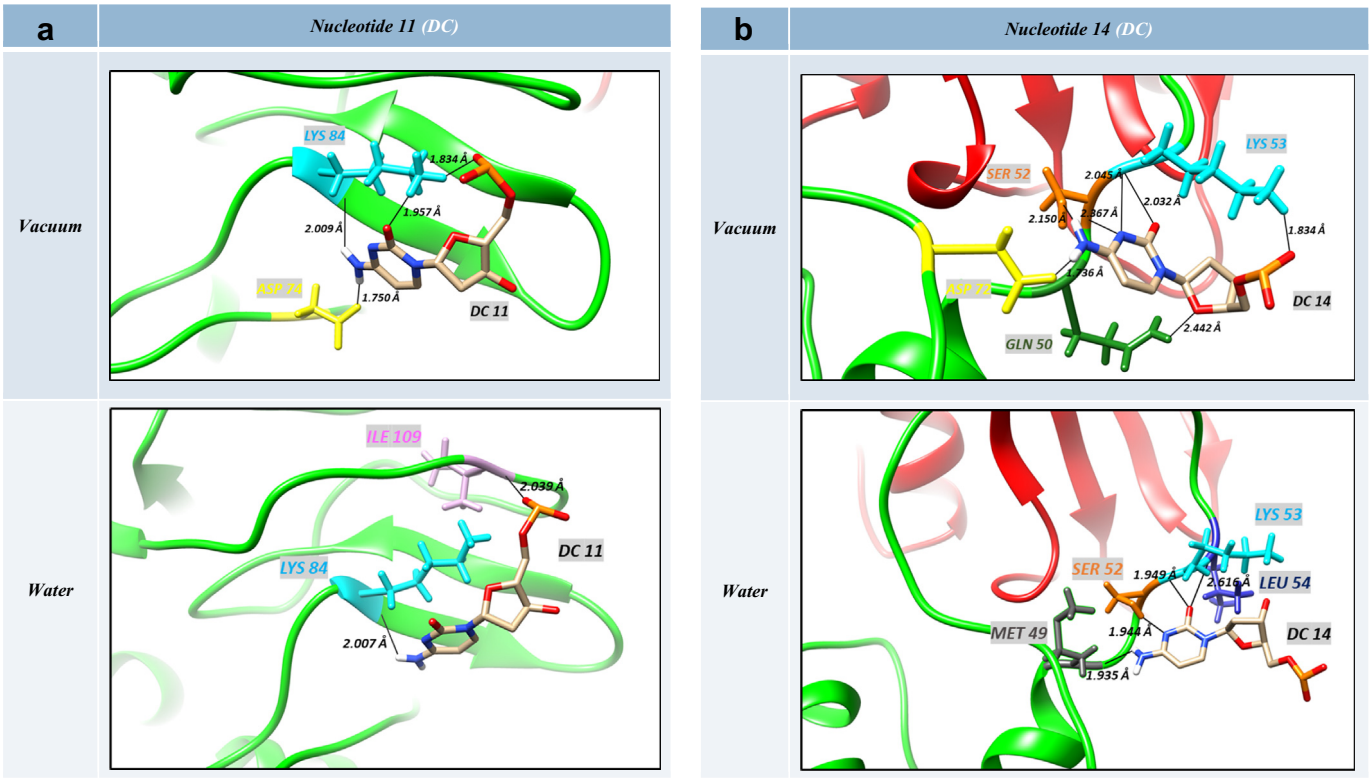

Fig. 6. Schematics of hydrogen bonds formed by Apt_{EpA} nucleotides and EpCAM residues for (a) DC11 and (b) DC14 in the vacuum and water environments.

whether the EpCAM had become more expansive or compact in the water environment. Fig. 7 shows the RMSD and Rg profile during the MD simulation for 10 ns. Both values underwent significant changes for the first 5 ns and almost fluctuated at constants after 5 ns. This indicates that the EpCAM configuration in the water environment became stable.

To observe the stability of the EpCAM/Apt_{EpA} complex at room temperature, Fig. 8 shows the variation in hydrogen bond numbers between EpA and EpCAM during the MD simulation at 300 K for 10 ns. One can see the hydrogen bond number continually decreases from 32 to 24 during the first 5 ns, indicating the EpCAM/Apt_{EpA} complex adjusts its local configuration to equilibrate with the water molecules. Furthermore, the considerable RMSD and Rg changes in Fig. 7 for the first 5 ns reflect the local configuration change of the EpCAM/Apt_{EpA} complex. From 5 to 8.5 ns, the hydrogen bond number fluctuated at approximately 24, and then slightly decreases to approximately 21 after 8.5 ns. Because the EpCAM configuration became stable after 5 ns according to the RMSD and Rg profiles in Fig. 7, the hydrogen number change after 5 ns can be attributed to the local configuration adjustment of Apt_{EpA} to form a stable hydrogen bond network with the EpCAM in the water environment.

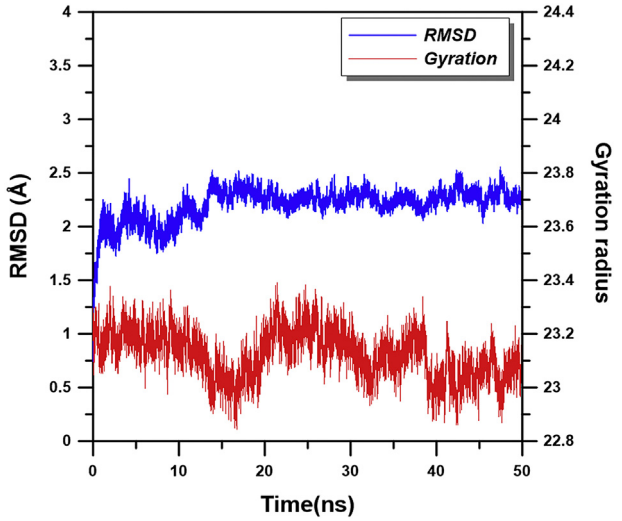

Fig. 7. RMSD and gyration radius profiles of the EpCAM in the water environment.

4. Conclusions

Based on our previous STUN–BH method, we proposed a novel global minimal search method, STUN–BH–DMD, for a larger aptamer molecule, to determine the most stable adsorption configuration on a target biomolecule. Combining the DMD method with the STUN–BH

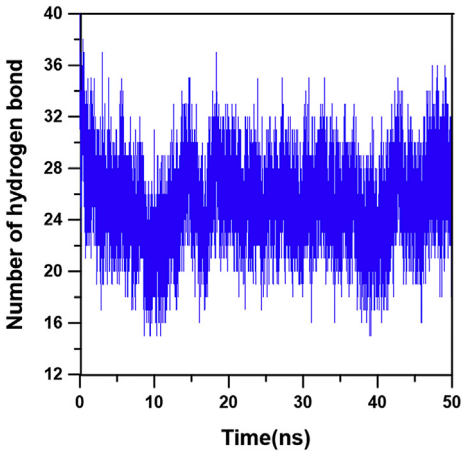

Fig. 8. Variations in hydrogen bond numbers between Apt_{EpA} and EpCAM during the MD simulation at 300 K for 50 ns in the water environment.

method could effectively expand the spatial domain search. A benchmark system for the Apt_{EpA}/EpCAM complex was used and the most stable Apt_{EpA} configuration on the EpCAM was obtained after 300,000 STUN–BH–DMD steps in the vacuum system. For the most stable Apt_{EpA}/EpCAM complex, the Apt_{EpA} was attached to the entangling loop fragments of the two EpCAM molecules with most Apt_{EpA}.

By monitoring the RMSD, Rg, and hydrogen bond number variations during the MD simulation at 300 K for 10 ns in the water environment, we observed that the local structures of both Apt_{EpA} and EpCAM in the vacuum changed to become equilibrated with the water molecules. After the long MD simulation, stable hydrogen bonds formed between the bases of Apt_{EpA} and EpCAM residues of the secondary structures, including the alpha helix and beta sheet becoming less stable in the water environment. Regarding the hydrogen bonds that formed between the bases of Apt_{EpA} and EpCAM residues of loop fragments, they remained stable in the water environment. This study has proposed a new numerical process to find the most stable complex configuration. The more reliable results can be obtained if a more recent force field is used for the STUN-BH-DMD method.

Declarations of Competing Interests

No potential conflict of interest was reported by the authors.

Acknowledgments

The authors would like to thank the (1) Ministry of Science and Technology of Taiwan, under Grant No. MOST105-2221-E-110-030-MY3, MOST106-2628-E-110-001-MY3, and MOST106-2628-B-110-001-MY4; (2) NSYSU-KMU JOINT RESEARCH PROJECT, (#NSYSUKMU 107-P032) from Kaohsiung Medical University for their support on this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2019.06.021.

References

[1] Heiat M, Najafi A, Ranjbar R, Latifi AM, Rasaee MJ. Computational approach to analyze isolated ssDNA aptamers against angiotensin II. J Biotechnol 2016;230:34–9.
[2] Mohan M, Pandya V, Anindya R. Escherichia coli AlkB and single-stranded DNA binding protein SSB interaction explored by molecular dynamics simulation. J Mol Graph Model 2018;84:29–35.
[3] Biniuri Y, Albada B, Willner I. Probing ATP/ATP-Aptamer or ATP-Aptamer mutant complexes by microscale Thermophoresis and molecular dynamics simulations: discovery of an ATP-Aptamer sequence of superior binding properties. J Phys Chem B 2018;122:9102–9.
[4] Chen Y, Zhou Y, Chen M, Xie B, et al. Isorenieratene interaction with human serum albumin: multi-spectroscopic analyses and docking simulation. Food Chem 2018;258:393–9.
[5] Budryn G, Grzelczyk J, Jaśkiewicz A, Żyżelewicz D, et al. Evaluation of butyrylcholinesterase inhibitory activity by chlorogenic acids and coffee extracts assed in ITC and docking simulation models. Food Res Int 2018;109:268–77.
[6] Jones Gareth, et al. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267(3):727–48.
[7] Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys 2008;71:126601.
[8] Cerqueira N, Bras N, Fernandes P, Ramos M. MADAMM: a multistaged docking with an automated molecular modeling protocol. Proteins Struct Funct Bioinforma 2009;74:192–206.
[9] Morris GM, Huey R, Lindstrom W, Sanner MF, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91.
[10] Ferreira L, dos Santos R, Oliva G, Andricopulo A. Molecular docking and structure-based drug design strategies. Molecules 2015;20:13384–421.
[11] Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. Curr Opin Struct Biol 2018;48:93–102.
[12] Kinghorn A, Fraser L, Liang S, Shiu S, Tanner J. Aptamer bioinformatics. Int J Mol Sci 2017;18:2516.
[13] Wang D, Liu B, Ma Y, Wu C, et al. A molecular recognition approach to synthesize nucleoside analogue based multifunctional nanoparticles for targeted cancer therapy. J Am Chem Soc 2017;139:14021–4.
[14] Saberi Z, Rezaei B, Khayamian T. A fluorescent aptasensor for analysis of adenosine triphosphate based on aptamer–magnetic nanoparticles and its single-stranded complementary DNA labeled carbon dots. Luminescence 2018;33:640–6.
[15] Alshaer W, Ababneh N, Hatmal M, Izmirli H, et al. Selection and targeting of EpCAM protein by ssDNA aptamer. PLoS One 2017;12:e0189558.
[16] Song Y, Zhu Z, An Y, Zhang W, et al. Selection of DNA aptamers against epithelial cell adhesion molecule for cancer cell imaging and circulating tumor cell capture. Anal Chem 2013;85:4141–9.
[17] Pavšič M, Gunčar G, Djinović-Carugo K, Lenarčič B. Crystal structure and its bearing towards an understanding of key biological functions of EpCAM. Nat Commun 2014;5:4764.
[18] Torabi R, Bagherzadeh K, Ghourchian H, Amanlou M. An investigation on the interaction modes of a single-strand DNA aptamer and RBP4 protein: a molecular dynamic simulations approach. Org Biomol Chem 2016;14:8141–53.
[19] Yang H-W, Ju S-P, Cheng C-H, Chen Y-T, et al. Aptasensor designed via the stochastic tunneling-basin hopping method for biosensing of vascular endothelial growth factor. Biosens Bioelectron 2018;119:25–33.
[20] Proctor EA, Ding F, Dokholyan NV. Discrete molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci 2011;1:80–92.
[21] Hamacher K, Wenzel W. Scaling behavior of stochastic minimization algorithms in a perfect funnel landscape. Phys Rev E 1999;59:938.
[22] Wenzel W, Hamacher K. Stochastic tunneling approach for global minimization of complex potential energy landscapes. Phys Rev Lett 1999;82:3003.
[23] Pettersen EF, Goddard TD, Huang CC, Couch GS, et al. UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.
[24] Genheden S, Akke M, Ryde U. Conformational entropies and order parameters: convergence, reproducibility, and transferability. J Chem Theory Comput 2013;10:432–8.
[25] Hornak V, Abel R, Okur A, Strockbine B, et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins Struct Funct Bioinforma 2006;65:712–25.
[26] Macdonald J, Henri J, Goodman L, Xiang D, et al. Development of a bifunctional aptamer targeting the transferrin receptor and epithelial cell adhesion molecule (EpCAM) for the treatment of brain cancer metastases. ACS Chem Nerosci 2017;8:777–84.
[27] Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 1993;6:525–33.
[28] Schug A, Herges T, Wenzel W. Reproducible protein folding with the stochastic tunneling method. Phys Rev Lett 2003;91:158102.
[29] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61.
[30] Wang Z, Sun H, Yao X, Li D, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys 2016;18:12964–75.
[31] Vu CQ, Rotkrua P, Soontornworajit B, Tantirungrotechai Y. Effect of PDGF-B aptamer on PDGFRβ/PDGF-B interaction: molecular dynamics study. J Mol Graph Model 2018;82:145–56.