RSC Advances

PAPER

Cite this: *RSC Adv.*, 2020, **10**, 790

Calculating the absolute binding free energy of the insulin dimer in an explicit solvent

Qiankun Gong, Haomiao Zhang, Haozhe Zhang and Changjun Chen *

Received 11th October 2019
Accepted 23rd December 2019

DOI: 10.1039/c9ra08284k

rsc.li/rsc-advances

Insulin is a significant hormone in the regulation of glucose level in the blood. Its monomers bind to each other to form dimers or hexamers through a complex process. To study the binding of the insulin dimer, we first calculate its absolute binding free energy by the steered molecular dynamics method and the confinement method based on a fictitious thermodynamic cycle. After considering some special correction terms, the final calculated binding free energy at 298 K is -8.97 ± 1.41 kcal mol^{-1}, which is close to the experimental value of -7.2 ± 0.8 kcal mol^{-1}. Furthermore, we discuss the important residue–residue interactions between the insulin monomers, including hydrophobic interactions, $\pi-\pi$ interactions and hydrogen bond interactions. The analysis reveals five key residues, Vla^{B12}, Tyr^{B16}, Phe^{B24}, Phe^{B25}, and Tyr^{B26}, for the dimerization of the insulin. We also perform MM-PBSA calculations for the wild-type dimer and some mutants and study the roles of the key residues by the change of the binding energy of the insulin dimer.

Introduction

Insulin is a crucial hormone which regulates glucose level in blood. Usually it is stored in the pancreas as a hexamer, and it only shows the biological activity in a monomeric form.[1,2] The whole process follows three steps in the physiological conditions. First, an insulin hexamer dissociates into dimers and the zinc ions. And then each dimer further dissociates into the monomers. At last the monomer interacts with its receptor.[1] Nowadays, there remain many challenges in the study of insulin.[2,3] For example, how to figure out the process by which one insulin monomer binds to another and how to reduce the affinity between them. For the first problem, despite many studies having been done in this field,[4–8] the real binding mechanism of the insulin monomers is still unclear.[9] In order to solve the second problem, some fast-acting insulin analogues with low affinities have been exploited, such as lispro insulin (Lys^{B28}, Pro^{B29}),[10] and insulin glulisine (Lys^{B3}, Glu^{B29}).[11]

Determination of the binding affinity of a complex in an experiment is difficult because of time and money. Compared to the experiment, molecular simulation on the computer has an advantage of low cost. It becomes an important assistant tool at present. Based on the molecular simulations, some methods have been developed to predict the affinities of the complexes, such as free energy perturbation molecular dynamics (FEP/MD),[12] Poisson–Boltzmann surface area (MM-PBSA),[13] linear interaction energy (LIE),[14] adaptively biased molecular dynamics (ABMD),[15] steered molecule dynamics (SMD)[16] and so on. Among them, SMD may be a promising tool in drug design,[17] because it is able to produce lots of short and non-equilibrium trajectories.

Now there are two ways to compare the binding affinities of a complex by the SMD method. The first way uses the maximum force (rupture force) on the pulling collective variable to compare the affinities.[18–22] For example, Mai *et al.* study two ligands binding to influenza virus and find that there is a strong relationship between the maximum force and the experimental free energy. The correlation coefficient is about 0.97.[19] However, a recent study also indicates that there might be at least two different relationships (correlation coefficients) between the maximum force and the experimental value.[23] For a specific complex, the real correlation depends on the predominant interaction type between the receptor and the ligand, like the electrostatic interaction, hydrophobic interaction or H-bonds. It makes the comparison work much complicated.

The second way of comparison is calculating the free energy profile.[16] It is based on an equality proposed by Jarzynski that connects the external work and the free energy differences.[24] With the equality, one can obtain the free energy profile and the related binding free energy of a complex by lots of non-equilibrium trajectories. The whole calculation is quite suitable for parallelization. Up to now, there are numerous successful applications, like two ligands binding to FKBP protein[25] and numerous ligands binding to protein FKBP, trypsin and cyclin-dependent kinase 2.[26] These studies present that the structural flexibility and the steered speed have large impact on the calculations of the free energies. To obtain accurate free energy, the low speed should be used in SMD simulation. Moreover, to improve the calculation accuracy, a lot

Biomolecular Physics and Modeling Group, School of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China. E-mail: cjchen@hust.edu.cn

of practical methods have been proposed, including the fluctuation-dissipation theorem of Brownian dynamics (BD-FDT),[27] adaptive SMD,[28] hybrid SMD,[29,30] and the non-equilibrium friction correction method.[31]

A few years ago, Tyka and co-workers proposed a confinement method to calculate the free energy differences between two different stable states of a single molecule.[32] Suppose these two stable states are P and Q, respectively. The free energy difference between the two states is ΔF_{PQ}. Based on a thermodynamic cycle, the calculation of ΔF_{PQ} is divided into three independent stages. Two of them are the calculations of the free energy differences, ΔF_{PP^*} and ΔF_{QQ^*}. They correspond to the transitions from the flexible state, P and Q, to the restrained state, P* and Q*, respectively (method: thermodynamic integration (TI)[33]). The last stage is the calculation of the free energy difference between the two restrained states P* and Q* (method: normal mode analysis (NMA)[34]). The confinement method has been shown to be practical for an isolated molecule. Recently, Perthold and Oostenbrink also construct a thermodynamic cycle to calculate the binding free energies of some protein–protein complexes with the perturbed distance restraints.[35] Their free energy results are close to the experimental values.

In this work, we borrow the ideas of the above papers into the calculation of the absolute binding free energy of an insulin dimer system. Since TI and NMA are very time-consuming for large systems, we replace them by SMD. In our work, one SMD simulation is performed in a virtual binding process of the two insulin monomers. To remove the effect of the restraints from the results, we also perform two additional SMD simulations to produce the free energy differences between the restrained and unrestrained states. Just as the work in ref. 35, the three processes constitute a fictitious but complete thermodynamic cycle from the unbound state to the bound state for the insulin monomers. The final result in our work is -8.97 ± 1.41 kcal mol^{-1}, which is close to the experimental value -7.2 ± 0.8 kcal mol^{-1}.[36]

Besides the calculation of the absolute binding energy by SMD, we also investigate the binding mechanisms of the insulin dimer. A set of simulations are performed for the dimer at different distances. The results show the importance of the π–π interaction, hydrophobic interaction and the hydrogen bond interaction in the dimerization process. Finally, we perform the MM-PBSA calculations[13] to study how the key residues affect the binding energies of the dimer.

Materials and methods

Structural information

The conformation of the insulin dimer is shown in Fig. 1(b). Its monomeric structure consists of two chains (Fig. 1(a)). Chain A in red color has 21 residues and chain B in green color has 30 residues. They are associated with each other by two disulfide bonds (Cys^{A7}–Cys^{B7}, Cys^{A20}–Cys^{B19}). Two monomers bind together to form a dimer (Fig. 1(b)). On the interaction surface between the monomers, there is a hydrophobic core including two antiparallel β-sheets (B24–B26) and two α-helixes (B9–B18). And the β-sheets are stabilized by four hydrogen bonds (Fig. 1(b)). Many experiments have been done to investigate the structure and dynamics of insulin, such as atomic force microscopy (AFM) experiment,[37] alanine scanning experiment,[38] amide I two-dimensional infrared spectroscopy (2D IR) experiment,[39] laser induced temperature jump (T-jump) experiment,[40] time-resolved X-ray scattering experiment[41] and so on. Moreover, at 298 K, concentration difference spectroscopy experiment presents that the insulin dimer's binding free energy is -7.2 ± 0.8 kcal mol^{-1} (ref. 36) and the isothermal titration microcalorimetry (ITC) dilution experiment gives its dissociation free energy 6.88 ± 0.03 kcal mol^{-1}.[42] These studies suggest that both of the insulin dimerization and dissociation are enthalpy control. As to the simulations, the dynamic properties of the insulin monomer,[43,44] dimer[45–50] and hexamer[51,52] in water are also investigated by some groups. For instance, two distinct surfaces (the dimer forming surface and the hexamer forming surface) on an insulin monomer are characterized in ref. 44. The role of the confined water molecules to the stability of the insulin hexamer is studied in ref. 52.

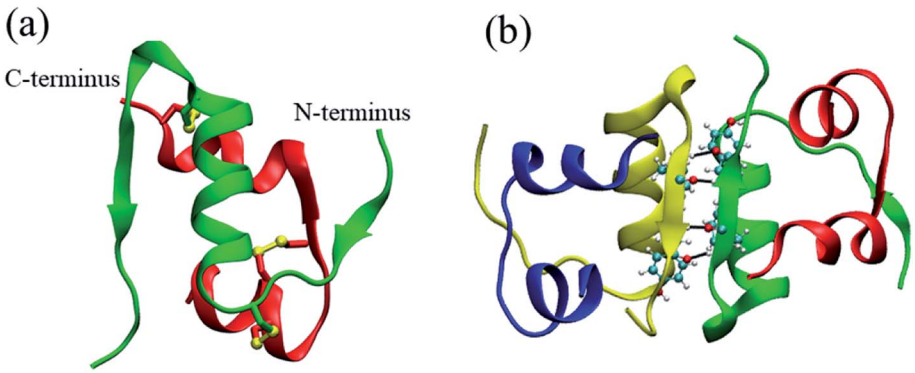

Fig. 1 (a) Monomeric structure from the pig insulin dimer in Protein Data Bank[53] (PDB ID: 4INS[54]). Chain A and chain B are rendered in red color and green color, respectively. Disulfide bonds, Cys^{A20}–Cys^{B19}, Cys^{A6}–Cys^{A11}, Cys^{A7}–Cys^{B7} from top to bottom, are plotted in the yellow balls and sticks model. (b) Structure of the pig insulin dimer from the Protein Data Bank[53] (PDB ID: 4INS[54]). Four hydrogen bonds between the two monomers are represented by the black solid lines. The figure is produced by VMD.[55]

Thermodynamic cycle

Since insulin tends to exist as dimers or hexamers in aqueous solvent, the monomeric structure of the insulin is still unknown in experiment. In an early NMR experimental study, the monomeric structure of an insulin variant des-[Phe^{B25}] (insulin without the Phe^{B25} residue) was determined.[56] It does not show a large structural difference to that in the dimer of the wild insulin. To verify the structural stability, we extract the monomer from the crystal structure of the insulin dimer and perform four independent 100 ns equilibrium simulations. The fluctuations of the RMSD of C_α atoms are about 2.0 Å (the first two and the last two residues are not included). It proves again that the monomeric structure from the dimer is relatively stable in aqueous solvent.

In 1985, concentration difference spectroscopy experiment obtained the thermodynamic quantities of porcine insulin, including the changes of enthalpy, entropy and free energy in the binding process.[36] The experimental values can be used as reference but the process is too complicated to be simulated on computer, so does the calculation of the absolute binding free energy. In current work, inspired by the confinement method[32] and the work by Perthold *et al.*,[35] we also use a thermodynamic cycle to circumvent the issue. The schematic diagram of the cycle is illustrated in Fig. 2.

In the figure, the whole thermodynamic cycle consists of four independent processes. Process D at bottom is corresponding to the real binding process from the unbound flexible monomers to the dimer. Process A, B and C are the virtual processes to close the thermodynamic cycle. Among them, process A at left is a virtual relaxing process from the restrained unbound state to the flexible unbound state. In the restrained unbound state, eleven geometric restraints are applied to the system (see the details below). And in the flexible unbound state, all the restraints are all removed from the system. Process C at right is similar to process A except that the SMD simulation is only performed for the bound state. As to process B at top, it is a virtual binding process from the unbound state to the bound state with a fixed force constant in all the restraint potentials. It must be noted that the simulation of the virtual binding process B is much easier than the real binding process D. This is due to the decrease of the degrees of freedom in the system.

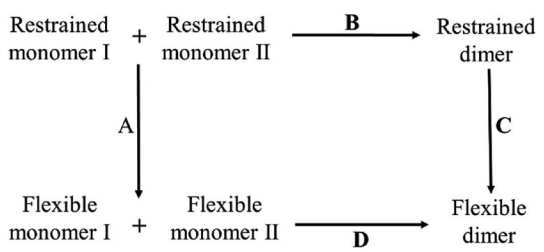

Fig. 2 Thermodynamic cycle in the calculation of the absolute binding free energy of insulin dimer. Process D at bottom is the real binding process of a dimer. All the molecules at the bottom are flexible in the process. Process A, B and C are the virtual processes for SMD simulation.

Based on the thermodynamic cycle, we divide the calculation of the absolute binding free energy of insulin dimer ΔF_D into three parts, ΔF_A, ΔF_B and ΔF_C. Each part represents a free energy difference in the corresponding process. All the simulations are performed by SMD.[16] Finally, the absolute binding free energy of insulin dimer is obtained by the following formula:

$$\Delta F_D = -\Delta F_A + \Delta F_B + \Delta F_C \quad (1)$$

Standard-state correction and rotational restriction correction

In this work, we also involve the standard-state correction ΔF_V and the rotation restriction correction ΔF_R. ΔF_V gives the free-energy difference when one molecule transfers from the available volume in the unbound state V_u to the standard-state volume V_0 (1661 Å^3).[35,57]

$$\Delta F_V = -RT \ln\left(\frac{V_u}{V_0}\right) = -RT \ln\left(\frac{\Delta r_{x,u}\Delta r_{y,u}\Delta r_{z,u}}{V_0}\right) \quad (2)$$

Here $\Delta r_{x,u}$, $\Delta r_{y,u}$ and $\Delta r_{z,u}$ are the differences between the minimum and maximum monomer–monomer distances in the x, y and z directions at the unbound state with restraints, respectively.

The rotational restriction correction term ΔF_R is introduced to account for the effect of the restraints in the bound state, which is given in the following formula:[35]

$$\Delta F_R = -RT \ln\left(\frac{4\pi {r_b}^2}{\Delta r_{y,b}\Delta r_{z,b}}\right) \quad (3)$$

Here $\Delta r_{y,b}$ and $\Delta r_{z,b}$ are the maximum deviations of the distances in the y and z directions at the bound state with restraints, and r_b is the average radial monomer–monomer distance in the bound state without restraints. As a consequence, the absolute binding free energy of the insulin dimer becomes

$$\Delta F = -\Delta F_A + \Delta F_B + \Delta F_C + \Delta F_V + \Delta F_R \quad (4)$$

We want to note that the reaction coordinate in the SMD simulation for the virtual binding process B is the distance between the monomers along a fixed X-axis. The free energy does not need the Jacobian correction.[58]

Free energy calculation

Here we give a brief introduction of the existing SMD method. Thermodynamics theory says that when a system changes infinitely slow along a path from its state X to state Y, the Helmholtz free energy difference between the states equals the external work W performed on this system:

$$W = \Delta F = F_Y - F_X \quad (5)$$

where F_X, F_Y is the free energy of state X and state Y, respectively. This equation only works for an extremely slow process. As to a transition process with a finite rate, Jarzynski proposes a new equality:[24]

$$\exp(-\beta\Delta F) = \langle \exp(-\beta W) \rangle \quad (6)$$

Here β is the inverse temperature $1/k_B T$, and the bracket indicates an ensemble average. This equality allows us to obtain the free energy difference from lots of non-equilibrium short trajectories.

To force the system to move along a fast transition path, some kind of restraint potential should be applied. The potential may have many different forms,[59] such as that with the constant forces or torques. But the most widely used one is the spring potential. It has the following form:

$$V(\mathbf{r};\ \lambda) = \frac{k}{2}[\xi(\mathbf{r}) - \lambda]^2 \quad (7)$$

where $\mathbf{r}$ is the positions of the atoms in a 3N dimensional space. $\xi(\mathbf{r})$ is the collective variable to be restrained on the path. In this work, it is set as the distance of center-of-mass of the insulin monomers at *X*-axis for process A, B and C. k is the force constant of the restraint potential. The center of the spring potential, λ, moves according to following formula:

$$\lambda = \lambda_0 + vt \quad (8)$$

λ_0 is the initial position of the system on the path, t is the simulation time and v is the constant moving velocity.

With such spring potential, the new Hamiltonian of the restrained system is

$$H(\mathbf{r},\mathbf{p};\lambda) = H_0(\mathbf{r},\mathbf{p}) + V(\mathbf{r};\lambda). \quad (9)$$

Here H_0 is the Hamiltonian of the original system and $\mathbf{p}$ is the momenta of the atoms. According to the Jarzynski's equality, the free energy on the path is given as follows:

$$\exp\{-\beta[F(\lambda_t) - F(\lambda_0)]\} = \langle \exp(-\beta W_{0\to t}) \rangle \quad (10)$$

$W_{0\to t}$ is the external work done on the system when it moves from λ_0 to λ_t on the path. The work can be calculated by an integral on time from 0 to t:

$$W_{0\to t} = \int_0^t \mathrm{d}t' \left(\frac{\partial H(\mathbf{r}_{t'},\ \mathbf{p}_{t'};\ \lambda_{t'})}{\partial \lambda} \right) \left(\frac{\partial \lambda}{\partial t} \right) \quad (11)$$

Substitute the restrained Hamiltonian into the formula, the external work becomes:

$$W_{0\to t} = -kv \int_0^t \mathrm{d}t' \left[\xi(\mathbf{r}_{t'}) - \lambda_0 - vt' \right] \quad (12)$$

Definition of the restraint potentials

Besides the distance of center-of-mass, there are ten more restraints on the two insulin monomers. These restraints are used to ensure the monomers to form a correct complex conformation in the SMD simulation. The first two restraints are the heavy atoms RMSD restraints. They restrain the structures of the individual monomers in the virtual binding process. The third and fourth restraints are the center-of-mass (C_α atoms) restraint for each monomer, which is used to ensure the monomers move along the *X*-axis direction (binding direction). The restraint potential has the form

$$V = \frac{k}{2}\left(\Delta y^2 + \Delta z^2\right) \quad (13)$$

Here Δy and Δz are the deviations of the center-of-mass of the monomers to the *X*-axis in the *y* or *z* direction, respectively. The form of the restraint potential does not change with time. And it does not have the steered coordinate in the SMD calculation (center-of-mass distance between the monomers on the *X*-axis).

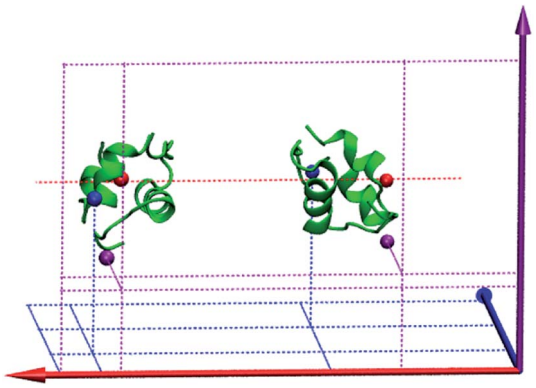

Fig. 3 Six single-atom restraints in the SMD simulation for the insulin dimer. The red, purple and blue solid arrows represent the *X*, *Y* and *Z* axis, respectively. Two red atoms are fixed at the horizontal red dotted line, two purple atoms are fixed on the vertical *X*–*Y* plane with a fixed *z* coordinate and two blue atoms are fixed on the horizontal *X*–*Z* plane with a fixed y coordinate. The figure is produced by VMD.[55]

The rest are six restraints on six different atoms of the monomers. They are used to avoid the rotation of the monomers. As shown in Fig. 3, two red atoms (OE1 of Gln^{A15} in monomer I and CD1 of Leu^{A16} in monomer II) are fixed on the red horizontal dotted line. Two purple atoms (ND1 of Hid^{B5} in monomer I and CG1 of Val^{B2} in monomer II) are fixed on the purple vertical *X*–*Y* plane with a fixed *z* coordinate. Two blue atoms (N of Lys^{B29} in monomer I and CA of Thr^{A8} in monomer II) are fixed on the blue horizontal *X*–*Z* plane with a fixed *y* coordinate. All these restraints are the harmonic potentials with a force constant of 50 kcal mol^{-1} $Å^{-2}$.

Simulation details

The initial structure of insulin dimer comes from the Protein Data Bank[53] (PDB ID: 4INS[54]). First, the complex is rotated to be aligned with *X*-axis (red solid arrow in Fig. 3). Then it is solvated in a periodic box of 88.6 Å × 55.5 Å × 55.3 Å with 6498 TIP3P water molecules.[60] Na^+ and Cl^- ion concentration is 12 mM, which is similar to experiments.[36] Second, the solvated system is minimized for 10 000 steps by the steepest descent method and 10 000 steps by the conjugate gradient method. Third, a 20 ns equilibrium simulation is performed to relax the system. In the simulation, the center-of-mass of each monomer is restrained to *X*-axis. Finally, the distance between the monomers is

gradually increased to 36.5 Å with the restraints mentioned above. The final unbound conformation of the system is shown in Fig. 3.

In this work, the force field is ff14SB AMBER force field[61] and the time step is 2.0 fs. SHAKE algorithm[62] is used to constrain the bonds involving the hydrogens. The non-bonded interaction is calculated with a cut-off of 10.0 Å. All the SMD simulations are performed by AMBER 18 (ref. 63) in the NVT ensemble at 298 K.

In the virtual binding process B in Fig. 2, the center-of-mass distance between the insulin monomers decreases from 36.5 Å to 17.5 Å with a low speed of 0.2 Å ns^{-1}. All the restraints are kept in the simulation. The simulation time of the process lasts for 95 ns. Process A and C are also simulated by SMD but with a little difference. In these two processes, the driven coordinate is the force constant k in the restraint potential that changes from a restrained state ($k = 50$ kcal mol^{-1} Å^{-2}) to a flexible state ($k = 0$ kcal mol^{-1} Å^{-2}). As a comparison, a standard SMD uses the distance as the driven coordinate. The simulation time of process B and C is 20 ns. To obtain the average free energy difference, we perform 40 independent simulations for every process.

To study the residue–residue interactions in the dimerization of the insulin (hydrophobic interaction, π–π interaction and hydrogen bond interaction), we run five simulations at a center-of-mass distance of 26 Å, 24 Å, 22 Å, 20 Å and 18 Å (native state). Each simulation consists of eight independent 100 ns trajectories, whose initial structure is extracted from process B and further equilibrated for 20 ns. To fix the distance between the monomers, an umbrella restraint potential U is applied to the system in the simulation.

$$\begin{aligned} U &= k(r_1 - r_2)R && (R \le r_1) \\ U &= \frac{1}{2}k(R - r_2)^2 && (r_1 < R \le r_2) \\ U &= 0 && (r_2 < R \le r_3) \\ U &= \frac{1}{2}k(R - r_3)^2 && (r_3 < R \le r_4) \\ U &= k(r_4 - r_3)R && (R > r_4) \end{aligned} \tag{14}$$

Here R is the center-of-mass distance between the monomers. k is 50 kcal mol^{-1} Å^{-2}. r_1, r_2, r_3 and r_4 determine the region. For instance, in the simulation at a distance of 26.0 Å, r_1, r_2, r_3 and r_4 are set to 25.5 Å, 25.6 Å, 26.4 Å and 26.5 Å, respectively.

We also perform MM-PBSA calculations to get the change of the binding energy between the wild-type dimer and the mutants. To get the mutants, we remove all the sidechain atoms of the selected residues except the C_β atom. And then change their names to 'ALA' in the pdb file. The missing hydrogen atoms are added automatically by the tleap tool in AMBER 18.[63] In all the MM-PBSA calculations, we use the three-trajectory mode and run 27 trajectories for each mutant. For each trajectory, the simulation time is 100 ns and 25 snapshots are extracted evenly from the last 50 ns. So, in total we have 225 snapshots in the MM-PBSA calculation. The calculation is performed by the MMPBSA.py script in AMBER 18.[63] The dielectric constants of the solute and the solvent are 1.0 and 80.0, respectively. The nonpolar part of the solvation energy is estimated by:

$$\mathbf{G}_{\mathrm{np,solv}} = 0.0072 \times \mathrm{SASA} \tag{15}$$

Results and discussions

As shown in the thermodynamics cycle in Fig. 2, the calculation of the binding free energy of the insulin dimer is completed by three independent processes. The main process is the virtual binding process from the restrained unbound state to the restrained bound state (process B). And the other two are the virtual relaxing processes at the unbound and bound state, respectively (process A and C). For each process, we perform

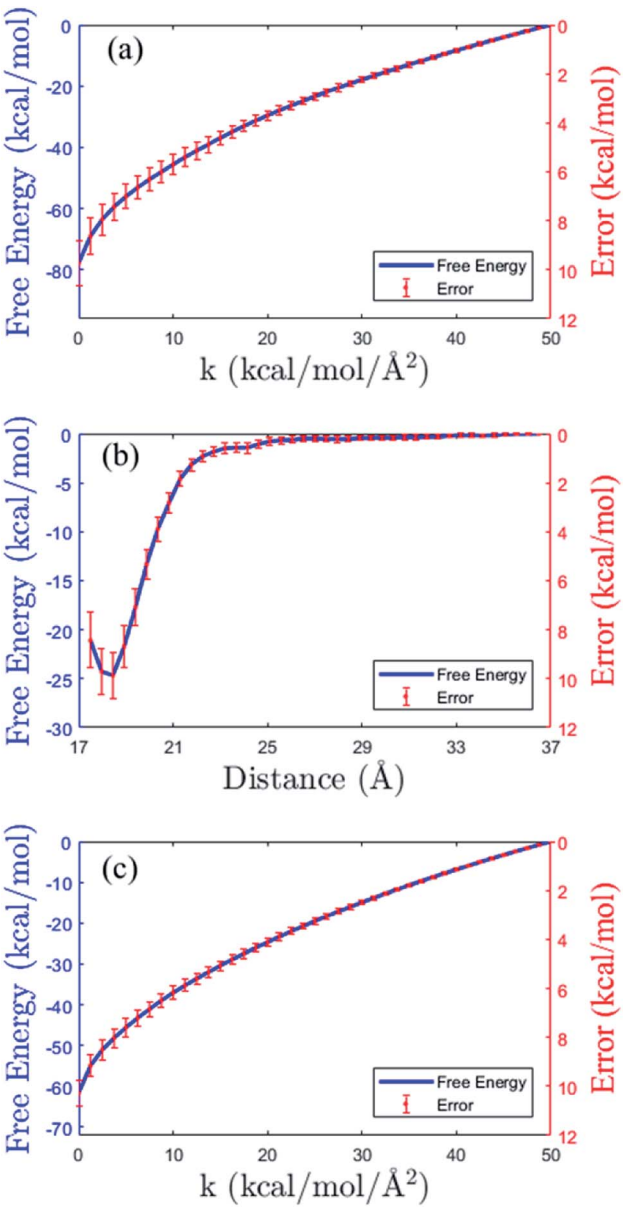

Fig. 4 Free energy profiles (blue solid lines) and the errors (in red color) in the three processes of the thermodynamic cycle in Fig. 2. (a) Virtual relaxing process A. (b) Virtual binding process B. (c) Virtual relaxing process C. See the descriptions of the processes in the main text.

a repeated SMD simulation[16] to generate 40 trajectories. These trajectories are used for the calculation of the free energies (eqn (10)) and the corresponding errors.

The free energy profiles of the three processes are shown in Fig. 4. Panel (b) is the data of the virtual binding process B. In the process, the center-of-mass distance between the monomers decreases from 36.5 Å to 17.5 Å. And the least free energy ΔF_B is -25.02 ± 0.95 kcal mol^{-1} at distance 18.2 Å. Panel (a) and (c) are the data of the two virtual relaxing processes A and C at the unbound state (distance 36.5 Å) and bound state (distance 18.2 Å), respectively. In these two processes, the force constants in all the restraint potentials decrease from 50.0 kcal mol^{-1} to 0 kcal mol^{-1} gradually. So, at the end of the processes, the molecule turns into a flexible structure completely. The end-to-end free energy differences of the two processes are -77.88 ± 0.89 kcal mol^{-1} and -61.66 ± 0.53 kcal mol^{-1}, respectively. As expecting, all the free energy errors increase when the molecule moves along the path.

In order to have the absolute binding free energy, the standard-state correction and the rotational restriction correction are both considered in our work. The standard-state correction (eqn (2)) is 5.45 kcal mol^{-1} and the rotational restriction correction (eqn (3)) is -5.62 kcal mol^{-1}. Combining all these data together, the final absolute binding free energy of the insulin dimer is -8.97 ± 1.41 kcal mol^{-1} (eqn (4)). In experimental aspect, the pig insulin dimer has been studied by the concentration difference spectroscopy. Strazza *et al.* measure the equilibrium constant K_a at five uniformly-spaced temperatures from 394 K to 311 K.[36] At 298 K, the corresponding binding free energy is -7.2 ± 0.8 kcal mol^{-1}.[36] It shows that our calculation result is close to the experimental data.

In a previous study, Zoete *et al.* calculate the absolute binding free energy of the insulin dimer by the MM-GBSA method.[46] In this method, the binding free energy is constituted of an enthalpy term and an entropy term. The former includes the difference of the gas phase energy and the solvation free energy between the unbound and the bound state. And the latter contains the difference of the transitional, rotational and vibrational entropies. The MM-GBSA result is -11.91 ± 6.7 kcal mol^{-1},[46] which is comparable to the experimental value.[36] But, its enthalpy term and the entropy term are -38.65 ± 5.8 kcal mol^{-1} and -26.74 ± 3.6 kcal mol^{-1}, respectively.[46] The experimental enthalpy and entropy ($T\Delta S$) is -10.0 kcal mol^{-1} and -3.3 kcal mol^{-1} respectively, which are obtained approximately by van't Hoff relation. These two large quantities are greatly different to that in experiment. This may be because the MM-GBSA method includes many parameters and approximations.[13,34,64,65] For the enthalpy term, the error might come from the implicit solvent model,[13,66] and for the entropy term, the error might come from the harmonic approximation.[34] And there are some improved methods aim to decrease the errors in the MM-GBSA or MM-PBSA calculation.[67–70]

Compared to MM-GBSA, SMD is quite simple. Its calculation is based on work W in eqn (6), instead of the separated enthalpy and entropy terms. But SMD do have its own parameters, such as the steered speed and the number of the trajectories.

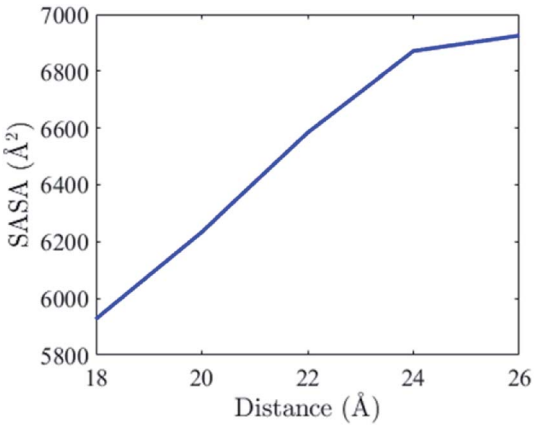

Fig. 5 The average SASA of the insulin dimer at different monomer–monomer distance.

Table 1 Eight residues that have the largest SASA differences from the isolated monomers (distance 26 Å) to the complex (distance 18 Å)

Residue	Phe^{B24}	Tyr^{B26}	Glu^{B13}	Val^{B12}	Tyr^{B16}	Ser^{B9}	Pro^{B28}	Phe^{B25}
SASA	132.3	129.0	106.7	105.5	98.0	93.2	76.0	69.6

Previous studies have shown that both slowing down the steered speed and increasing the number of the trajectories can improve the accuracies of the results.[25–27] However, preparation of a proper restraint is still difficult in SMD.

One more thing we want to note is that SMD completes the calculation in the explicit solvent. And MM-GBSA or MM-PBSA are not. In MM-GBSA or MM-PBSA, the snapshots are extracted from the explicit solvent, but the free energy calculation is done in implicit solvent. This may bring the underlying problem to the calculation of the absolute binding free energy. For example, determination of the reference solute dielectric constant value is not easy. It can be 1.0, 2.0 or 4.0 in different MM-PBSA calculations.[71–74]

Insulin is an essential protein in treating the disease. Understanding the binding mechanism of insulin dimer will be useful to design drug. Fig. 5 shows the change of the average solvent-accessible-surface area (SASA) in the binding process of the insulin monomers. The data come from five independent MD simulations at the center-of-mass (COM) distance of 26 Å, 24 Å, 22 Å, 20 Å and 18 Å (native state), respectively.

The change of the total SASA in the binding process (COM distance from 26 Å to 18 Å) mainly comes from 22 residues on the binding interface (B9–B30). We present eight residues that have the largest SASA differences in Table 1. Their SASA changes are plotted in Fig. 6. Among these residues, three are hydrophobic residues (Val^{B12}, Phe^{B24}, Phe^{B25}), four are polar residues (Ser^{B9}, Tyr^{B16}, Tyr^{B26}, Pro^{B28}) and one is the charged residues (Glu^{B13}). According to the definition of the hydrophobic interaction, the SASA reductions of the three hydrophobic residues (Val^{B12}, Phe^{B24}, Phe^{B25}) should favor the binding of the monomers. These residues are important to the insulin dimerization.

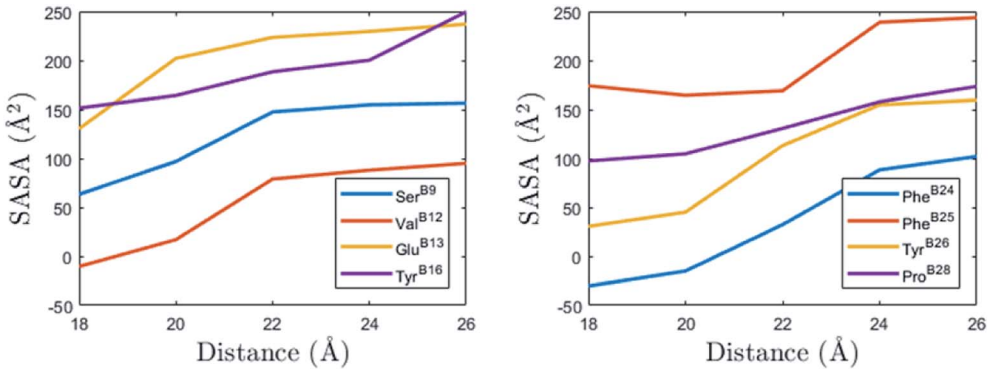

Fig. 6 The changes of the SASA of the eight residues in Table 1 from the isolated monomers (distance 26 Å) to the complex (distance 18 Å).

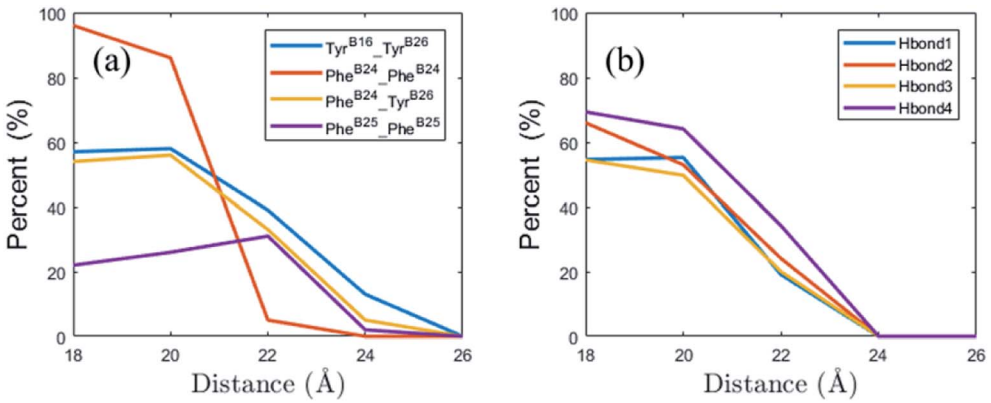

Fig. 7 (a) The percent of the four strongest π–π interaction between the aromatic amino acid pairs at different distances between the insulin monomers. (b) The percent of four inter-strand hydrogen bonds at different distances. (O–N distance smaller than 3.0 Å and donor–H–acceptor angle larger than 135°).

Moreover, Phe^{B24} and Phe^{B25} are on the flexible β-strand, their hydrophobic interactions also favor the formation of the β-sheet between the monomers.

π–π interaction between the aromatic amino acids is a typical non-covalent intermolecular force. Recently, Zhao *et al.* investigate lots of pairs of aromatic amino acids and find some statistical laws.[75] In their definition, π–π interaction satisfies some geometry conditions. First, the center-of-mass distance between two aromatic rings should be less than 7.2 Å. Second, the normal–normal acute angle should be larger than 50° (T-shaped conformation) or less than 30° (parallel conformation). In this work, we use the definition to study the π–π interactions between insulin monomers.

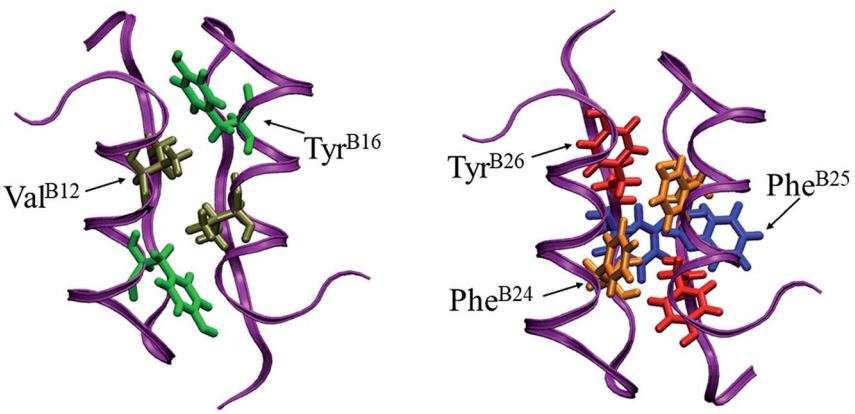

Fig. 8 The positions of five key residues on the binding interface of the insulin dimer. They are Val^{B12} (tan), Tyr^{B16} (green), Phe^{B24} (orange), Phe^{B25} (blue) and Tyr^{B26} (red). The figure is produced by VMD.[55]

Table 2 The Enthalpy and the structural RMSD of the wild-type insulin dimer and the analogues

Mutant	Enthalpy (kcal mol^{-1})	Dimer RMSD (Å)
Wile-type	−61.4	1.8
Phe^{B24}Ala	−62.0	2.2
Phe^{B25}Ala	−54.8	2.0
Asn^{A21}Ala	−48.7	1.9
Thr^{B27}Ala	−48.7	1.9
Val^{B12}Ala	−44.4	2.0
Tyr^{B26}Ala	−37.7	2.2
Tyr^{B16}Ala	−35.7	2.3

There are four aromatic amino acids (Tyr^{B16}, Phe^{B24}, Phe^{B25} and Tyr^{B26}) for each monomer on the binding interface. In the binding process, the percent of the π–π interactions between four strongest aromatic pairs are shown in Fig. 7(a) (Tyr^{B16}_Tyr^{B26}, Phe^{B24}_Phe^{B24}, Phe^{B24}_Tyr^{B26} and Phe^{B25}_Phe^{B25}). From the figure, we find that the first three π–π interactions increase quickly in the binding process. Besides the hydrophobic interaction, the inter-strand hydrogen bond interaction is also important to the stabilization of the complex. Fig. 7(b) gives the change of the percent of four inter-strand hydrogen bonds at different distances between the monomers. The locations of these bonds in the complex are shown in Fig. 1. It presents that all the four inter-strand hydrogen bonds become stable at 20 Å (percent larger than 50%). They make the two flexible β-strands form into one global β-sheet.

In a word, the dimerization of the insulin is favored by the hydrophobic interaction, π–π interaction and hydrogen bond interaction. Among the 22 residues on the binding interface, five key residues Val^{B12}, Tyr^{B16}, Phe^{B24}, Phe^{B25}, and Tyr^{B26} are picked out according their involvement in the interactions. The locations of the key residues in the dimer are shown in Fig. 8. These findings are close to previous alanine scanning experiment.[38] The experiment studies some mutants of the insulin

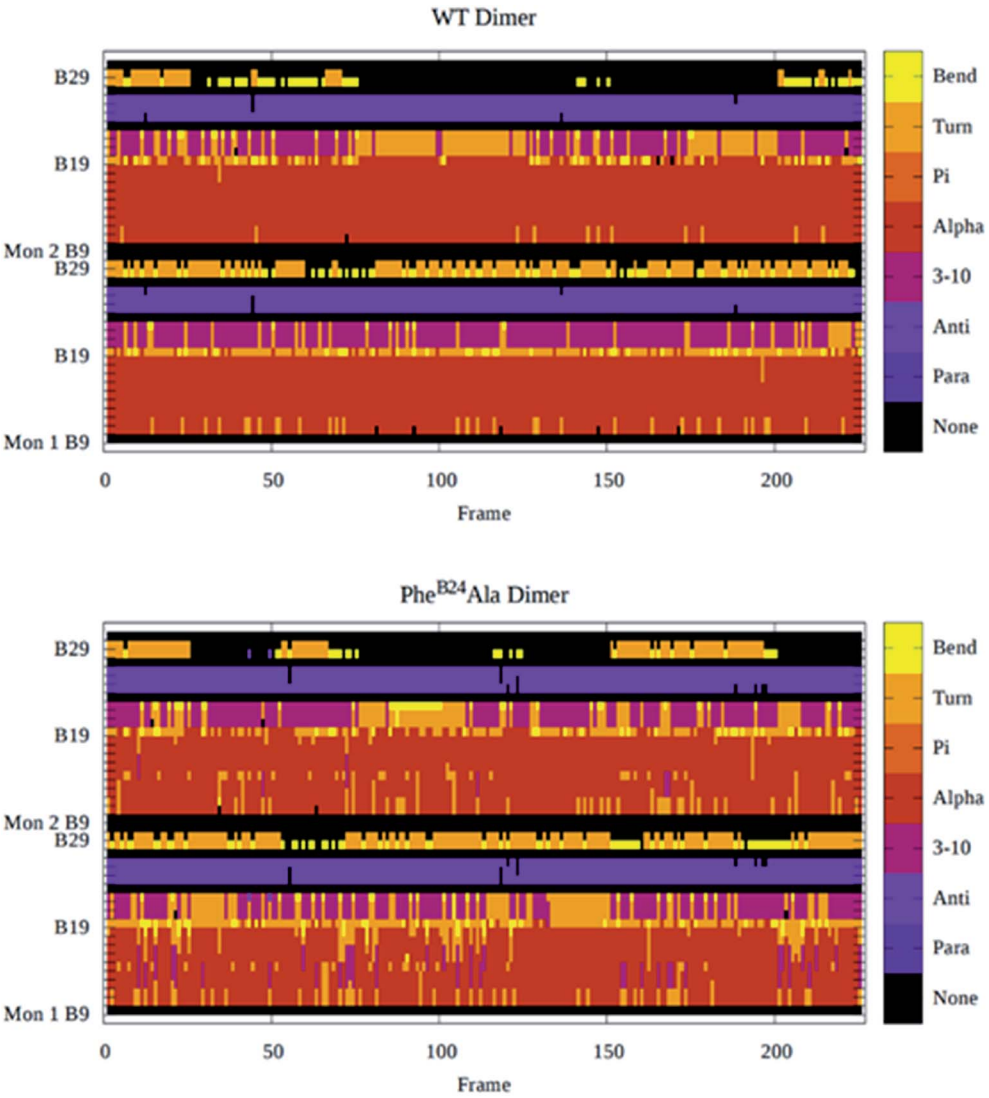

Fig. 9 The change of the secondary structures of the wild-type dimer (WT dimer) and the mutant dimer Phe^{B24}Ala in the simulation. Different secondary structures are plotted in different colors (defined in the color bar at right). The calculation is performed by DSSP.[78]

dimer ($Val^{B12}Ala$, $Tyr^{B16}Ala$, $Phe^{B24}Ala$, $Tyr^{B26}Ala$, $Asn^{A21}Ala$, $Phe^{B25}Ala$, $Thr^{B27}Ala$) and finds that the first four mutants exist as a monomeric insulin after the mutation. Why these residues (Val^{B12}, Tyr^{B16}, Phe^{B24}, Tyr^{B26}) are important to the dimerization of the insulin can be well explained by the analysis of the residue–residue interactions here. Of course, the hydrophobic residue Phe^{B25} is an exception. In the mutation experiment, Phe^{B25} does not present to be critical to the insulin dimerization apparently. In the simulation, its SASA reduction is smaller than the other hydrophobic residues (Table 1). And its π–π interaction is weaker than the other aromatic pairs (Fig. 7(a), magenta line). So, compared to the other key residues, Phe^{B25} may have less contribution to the binding of the monomers.

MM-GBSA or MM-PBSA are the popular approaches to study the relative binding affinity between similar complexes.[76,77] Because in this case, the entropy contribution can be safely ignored. For instance, Huiyong *et al.* study lots of protein–ligand systems and find that there is a good correlation between the enthalpy value and the binding affinity in experiment.[76] To verify the idea, we further investigate the variation of the enthalpy of the insulin dimer and its mutants by the MM-PBSA method. Seven residues Asn^{A21}, Val^{B12}, Tyr^{B16}, Phe^{B24}, Phe^{B25}, Tyr^{B26} and Thr^{B27} are mutated to Ala one by one. The entropy term is not included in the calculation.

Table 2 gives the enthalpy of the wild-type dimer and the mutants. The changes of the structural RMSD in the simulations are also presented. Compared to the wild-type dimer, the three mutants, Val^{B12}, $Tyr^{B16}Ala$, $Tyr^{B26}Ala$, have a much higher enthalpy. Mutations on these residues decrease the binding affinities of the dimer remarkably. The results are close to the findings in previous alanine scanning experiment[38] except the mutant $Phe^{B24}Ala$. It presents that MM-PBSA is successful in the prediction of the relative binding affinity of the complex. As to the average structural C_α RMSD, all the mutants show larger RMSD values than the wild-type dimer. The mutations destabilize the structure of the native state.

To describe the effect of the mutation, we also calculate the secondary structures of the residues on the binding interface (B9–B30) for the wild-type dimer and all the mutants by DSSP.[78] Among them, the mutant $Phe^{B24}Ala$ shows the most difference than the wild-type dimer. The changes of the secondary structures of the wild-type dimer and the mutant $Phe^{B24}Ala$ are shown in Fig. 9. In the figure, red color represents the α-helix and purple is the β-sheet. We find that both of the α-helix (B9–B18) and the β-sheet (B24–B26) are quite stable in the wild-type dimer. The dimerization is helpful to the stabilization of the secondary structures. But for the mutant $Phe^{B24}Ala$, the α-helix changes into the "turn" conformation frequently in the whole simulation time. The average percent of the α-helix (B9–B18) decrease from 83.0% (wild-type) to 72.8% (mutant). As to the individual residues in the helix, in monomer I, the percent of Ala^{B14} and Val^{B18} decrease 26.0% and 21.0%, respectively. The other residues, Val^{B12}, Glu^{B13}, Leu^{B15}, Tyr^{B16} and Leu^{B17}, decrease about 15%. And in monomer II, the residue Ala^{B14} decreases 19.0%, and Val^{B12}, Glu^{B13} and Val^{B18} decrease about 5%.

Conclusion

Protein–protein association is an important process in cell biology. Many typical systems have been successfully investigated in the simulations, such as the melittin aggregation[79,80] and the fibritin foldon domain assembly.[81] In this work, we study the dimerization of insulin. Combing the SMD method[16] with the confinement method,[32] we calculate its absolute binding free energy based on a fictitious thermodynamic cycle. The calculation result is close to the experimental value, which proves its feasibility to small protein–protein systems. Moreover, in order to understand the association mechanism of the insulin dimer, we analyze the variation of SASA, π–π interaction and hydrogen bond interaction during the binding process. And the residues Vla^{B12}, Tyr^{B16}, Phe^{B24}, Phe^{B25}, and Tyr^{B26} have the most favorable contributions to the dimerization. The current work may have implications for the development of the insulin analogues.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 31770773.

References

1 J. Brange and A. Vølund, *Adv. Drug Delivery Rev.*, 1999, **35**, 307–335.
2 A. N. Zaykov, J. P. Mayer and R. D. DiMarchi, *Nat. Rev. Drug Discovery*, 2016, **15**, 425–439.
3 M. Akbarian, Y. Ghasemi, V. N. Uversky and R. Yousefi, *Int. J. Pharm.*, 2018, **547**, 450–468.
4 S. H. Nakagawa and H. Tager, *J. Biol. Chem.*, 1986, **261**, 7332–7341.
5 R. Mirmira and H. Tager, *J. Biol. Chem.*, 1989, **264**, 6349–6354.
6 C. Kristensen, T. Kjeldsen, F. C. Wiberg, L. Schäffer, M. Hach, S. Havelund, J. Bass, D. F. Steiner and A. S. Andersen, *J. Biol. Chem.*, 1997, **272**, 12978–12983.
7 L. Zakova, E. Kletvikova, V. Veverka, M. Lepsik, C. J. Watson, J. P. Turkenburg, J. Jiracek and A. M. Brzozowski, *J. Biol. Chem.*, 2013, **288**, 10230–10240.
8 J. G. Menting, Y. Yang, S. J. Chan, N. B. Phillips, B. J. Smith, J. Whittaker, N. P. Wickramasinghe, L. J. Whittaker, V. Pandyarajan, Z. L. Wan, S. P. Yadav, J. M. Carroll, N. Strokes, C. T. Roberts Jr, F. Ismail-Beigi, W. Milewski, D. F. Steiner, V. S. Chauhan, C. W. Ward, M. A. Weiss and M. C. Lawrence, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E3395–E3404.
9 G. Scapin, V. P. Dandey, Z. Zhang, W. Prosise, A. Hruza, T. Kelly, T. Mayhood, C. Strickland, C. S. Potter and B. Carragher, *Nature*, 2018, **556**, 122–125.

10 D. C. Howey, R. R. Bowsher, R. L. Brunelle and J. R. Woodworth, *Diabetes*, 1994, **43**, 396–402.
11 T. Heise, L. Nosek, H. Spitzer, L. Heinemann, E. Niemoller, A. D. Frick and R. H. Becker, *Diabetes, Obes. Metab.*, 2007, **9**, 746–753.
12 J. Wang, Y. Deng and B. Roux, *Biophys. J.*, 2006, **91**, 2798–2814.
13 J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
14 J. Åqvist, V. B. Luzhkov and B. O. Brandsdal, *Acc. Chem. Res.*, 2002, **35**, 358–365.
15 V. Babin, C. Roland and C. Sagui, *J. Chem. Phys.*, 2008, **128**, 134101.
16 S. Park, F. Khalili-Araghi, E. Tajkhorshid and K. Schulten, *J. Chem. Phys.*, 2003, **119**, 3559–3566.
17 P. C. Do, E. H. Lee and L. Le, *J. Chem. Inf. Model.*, 2018, **58**, 1473–1482.
18 B. K. Mai, M. H. Viet and M. S. Li, *J. Chem. Inf. Model.*, 2010, **50**, 2236–2247.
19 B. K. Mai and M. S. Li, *Biochem. Biophys. Res. Commun.*, 2011, **410**, 688–691.
20 M. Suan Li and B. Khanh Mai, *Curr. Bioinf.*, 2012, **7**, 342–351.
21 H. Nguyen, N. Do, T. Phan and T. Pham, *Appl. Biochem. Biotechnol.*, 2018, **184**, 401–413.
22 H. Nguyen, T. Pham, H. L. Nguyen and T. Phan, *Appl. Biochem. Biotechnol.*, 2018, **186**, 834–846.
23 J. K. Marzinek, P. J. Bond, G. Lian, Y. Zhao, L. Han, M. G. Noro, E. N. Pistikopoulos and A. Mantalaris, *J. Chem. Inf. Model.*, 2014, **54**, 2093–2104.
24 C. Jarzynski, *Phys. Rev. Lett.*, 1997, **78**, 2690.
25 F. M. Ytreberg, *J. Chem. Phys.*, 2009, **130**, 164906.
26 N. Okimoto, A. Suenaga and M. Taiji, *J. Biomol. Struct. Dyn.*, 2017, **35**, 3221–3231.
27 L. Y. Chen, D. A. Bastien and H. E. Espejel, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6579–6582.
28 G. Ozer, E. F. Valeev, S. Quirk and R. Hernandez, *J. Chem. Theory Comput.*, 2010, **6**, 3026–3038.
29 L. Y. Chen, *J. Chem. Theory Comput.*, 2015, **11**, 1928–1938.
30 R. A. Rodriguez, L. Yu and L. Y. Chen, *J. Chem. Theory Comput.*, 2015, **11**, 4427–4438.
31 S. Wolf and G. Stock, *J. Chem. Theory Comput.*, 2018, **14**, 6175–6182.
32 M. D. Tyka, A. R. Clarke and R. B. Sessions, *J. Phys. Chem. B*, 2006, **110**, 17212–17220.
33 D. L. Beveridge and F. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, 1989, **18**, 431–492.
34 B. Tidor and M. Karplus, *J. Mol. Biol.*, 1994, **238**, 405–414.
35 J. W. Perthold and C. Oostenbrink, *J. Chem. Theory Comput.*, 2017, **13**, 5697–5708.
36 S. Strazza, R. Hunter, E. Walker and D. W. Darnall, *Arch. Biochem. Biophys.*, 1985, **238**, 30–42.
37 C. M. Yip, C. C. Yip and M. D. Ward, *Biochemistry*, 1998, **37**, 5439–5449.
38 H. Chen, M. Shi, Z.-Y. Guo, Y.-H. Tang, Z.-S. Qiao, Z.-H. Liang and Y.-M. Feng, *Protein Eng.*, 2000, **13**, 779–782.
39 Z. Ganim, K. C. Jones and A. Tokmakoff, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3579–3588.
40 X. X. Zhang, K. C. Jones, A. Fitzpatrick, C. S. Peng, C. J. Feng, C. R. Baiz and A. Tokmakoff, *J. Phys. Chem. B*, 2016, **120**, 5134–5145.
41 D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, I. Kosheleva and L. X. Chen, *J. Phys. Chem. Lett.*, 2017, **8**, 4413–4418.
42 E. Antolikova, L. Zakova, J. P. Turkenburg, C. J. Watson, I. Hanclova, M. Sanda, A. Cooper, T. Kraus, A. M. Brzozowski and J. Jiracek, *J. Biol. Chem.*, 2011, **286**, 36968–36977.
43 V. Zoete, M. Meuwly and M. Karplus, *J. Mol. Biol.*, 2004, **342**, 913–929.
44 K. Bagchi and S. Roy, *J. Phys. Chem. B*, 2014, **118**, 3805–3813.
45 M. Falconi, M. T. Cambria, A. Cambria and A. Desideri, *J. Biomol. Struct. Dyn.*, 2001, **18**, 761–772.
46 V. Zoete, M. Meuwly and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**, 79–93.
47 T. Kim, A. Rhee and C. M. Yip, *J. Am. Chem. Soc.*, 2006, **128**, 5330–5331.
48 P. Banerjee, S. Mondal and B. Bagchi, *J. Chem. Phys.*, 2018, **149**, 114902.
49 K. El Hage, P. Mondal and M. Meuwly, *Mol. Simul.*, 2018, **44**, 1044–1061.
50 S. Raghunathan, K. El Hage, J. L. Desmond, L. Zhang and M. Meuwly, *J. Phys. Chem. B*, 2018, **122**, 7038–7048.
51 H. Vashisth and C. F. Abrams, *Biophys. J.*, 2008, **95**, 4193–4204.
52 S. Mukherjee, S. Mondal, A. A. Deshmukh, B. Gopal and B. Bagchi, *J. Phys. Chem. B*, 2018, **122**, 1631–1637.
53 F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, 1977, **112**, 535–542.
54 E. N. Baker, T. L. Blundell, J. F. Cutfield, E. J. Dodson, G. G. Dodson, D. M. C. Hodgkin, R. E. Hubbard, N. W. Isaacs, C. D. Reynolds and K. Sakabe, *Philos. Trans. R. Soc., B*, 1988, **319**, 369–456.
55 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
56 A. M. M. Jørgensen, H. B. Olsen, P. Balschmidt and J. J. Led, *J. Mol. Biol.*, 1996, **257**, 684–699.
57 S. Doudou, N. A. Burton and R. H. Henchman, *J. Chem. Theory Comput.*, 2009, **5**, 909–918.
58 D. Trzesniak, A. P. Kunz and W. F. van Gunsteren, *ChemPhysChem*, 2007, **8**, 162–169.
59 S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers and K. Schulten, in *Computational molecular dynamics: challenges, methods, ideas*, Springer, 1999, pp. 39–65.
60 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
61 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
62 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
63 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden,

R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.

64 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.

65 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.

66 V. Tsui and D. A. Case, *Biopolymers*, 2000, **56**, 275–291.

67 J. Kongsted and U. Ryde, *J. Comput.-Aided Mol. Des.*, 2009, **23**, 63–71.

68 K. W. Harpole and K. A. Sharp, *J. Phys. Chem. B*, 2011, **115**, 9461–9472.

69 W. M. Menzer, C. Li, W. Sun, B. Xie and D. D. L. Minh, *J. Chem. Theory Comput.*, 2018, **14**, 6035–6049.

70 L. Duan, X. Liu and J. Z. H. Zhang, *J. Am. Chem. Soc.*, 2016, **138**, 5722–5728.

71 L. Li, C. Li, Z. Zhang and E. Alexov, *J. Chem. Theory Comput.*, 2013, **9**, 2126–2136.

72 E. Wang, G. Weng, H. Sun, H. Du, F. Zhu, F. Chen, Z. Wang and T. Hou, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18958–18969.

73 B. Wang, L. Li, T. D. Hurley and S. O. Meroueh, *J. Chem. Inf. Model.*, 2013, **53**, 2659–2670.

74 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. Zhang and T. Hou, *Chem. Rev.*, 2019, **119**, 9478–9508.

75 Y. Zhao, J. Li, H. Gu, D. Wei, Y.-c. Xu, W. Fu and Z. Yu, *Interdiscip. Sci.: Comput. Life Sci.*, 2015, **7**, 211–220.

76 H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.

77 J. Zhu, X. Li, S. Zhang, H. Ye, H. Zhao, H. Jin and W. Han, *J. Biomol. Struct. Dyn.*, 2017, **35**, 3140–3151.

78 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

79 C. Liao, M. Esai Selvan, J. Zhao, J. L. Slimovitch, S. T. Schneebeli, M. Shelley, J. C. Shelley and J. Li, *J. Phys. Chem. B*, 2015, **119**, 10390–10398.

80 M. Y. Shelley, M. E. Selvan, J. Zhao, V. Babin, C. Liao, J. Li and J. C. Shelley, *J. Chem. Theory Comput.*, 2017, **13**, 3881–3897.

81 A. Barducci, M. Bonomi, M. K. Prakash and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E4708–E4713.

