Understanding the conformation transition in the activation pathway of \( \beta_2 \) adrenergic receptor via a targeted molecular dynamics simulation

Xiuchan Xiao, Xiaojun Zeng, Yuan Yuan, Nan Gao, Yanzhi Guo, Xuemei Pu and Menglong Li

G protein coupled receptors (GPCRs) play a crucial role in regulating signal recognition and transduction through their activation. The conformation transition in the activation pathway is of particular importance for their function. However, it has been poorly elucidated due to experimental difficulties in determining the conformations and the time limitation of conventional molecular dynamics (CMD) simulation. Thus, in this work, we employed a targeted molecular dynamic (TMD) simulation to study the activation process from an inactive structure to a fully active one for \( \beta_2 \) adrenergic receptor (\( \beta_2 \)AR). As a reference, 110 ns CMD simulations on wild \( \beta_2 \)AR and its D130N mutant were also carried out. TMD results show that there is at least an intermediate conformation cluster in the activation process, evidenced by the principal component analysis and the structural and dynamic differences of some important motifs. It is noteworthy that the activation of the ligand binding site lags the G-protein binding site, displaying uncoupled correlation. Comparisons between the CMD and TMD results show that the D130N mutation significantly speeds up ICL2 and key ionic lock to enter into the intermediate state, which to some extent facilitates the activation involved in the NPxxY, DRY region and the separation between TM3 and TM6. However, the contribution from the D130N mutation to the activation of the ligand binding site could not be observed within the scale of 110 ns time. These observations could provide novel insights into previous studies for better understanding of the activation mechanism for \( \beta_2 \)AR.

Introduction

G protein coupled receptors (GPCRs), the largest group of cell surface proteins, play a crucial role in regulating signal recognition and transduction. Moreover, GPCRs are one of the largest druggable families of genes in the human genome. Approximately 30% of currently marketed drugs target GPCRs.

As revealed, in response to stimulation by an external signal (e.g. ligand binding or mutation), GPCRs can be activated from their inactive states to active ones, which is accompanied by a series of conformational rearrangements. The activated GPCRs, in turn, trigger activation of G protein and arrestin, leading to specific biological responses. Thus, the conformation changes in the activation pathway are of particular importance for understanding the structure and function of GPCRs and related drug design. Considerable efforts from experiments have been devoted to studying the issues mainly through various spectroscopic techniques. Until now, the structural basis for understanding the activation mainly stem from a number of crystal structures with bound agonists or inverse agonists, which presumably represent the active and the inactive states, respectively. These crystal structures have already provided valuable information on the structural differences between the active state and the inactive one. However, the differences cannot reveal the dynamic mechanism by which GPCRs transit between the two states. Furthermore, the lack of high resolution structures resulting from difficulties in experimentally resolving the structure of the membrane protein, in particular for the active states, also significantly limits the advances in understanding the activation mechanism. In addition, the specific motion associated with certain functions can be observed by means of an extensive body of pharmacological evidence, which suggests that GPCRs can adopt multiple, distinct signalling states in the activation pathway and various ligands can bind differentially to induce or stabilize those states. However, the number of functionally distinct states and their conformations remain...
unknown due to indirectly and often uncertainly determination in these experimental techniques as well as the molecular nature of the issue. Molecular dynamics simulations could provide a molecular level view of the dynamic behavior of proteins. Thus, this method has been successfully used to study the structure and function of various proteins and has become an indispensable tool in the field.10–12 Previous MD studies on the GPCRs have focused largely on the effects of ligands,13–18 including ligand-induced changes in the structure of the receptor, interactions between ligands, receptors and G-proteins, allosteric communication pipelines, and internal water pathway. Although some MD studies19–21 have been concerned with the conformation changes of GPCRs upon activation, the information is still very limited to explore the activation mechanism due to the time scale of MD simulations. It is known that the fastest atomic vibration limits the size of the numerical time steps to 1–2 fs, which puts a practical restriction on the simulation length to hundreds to thousands of nanoseconds for large protein systems. Thus, global domain motions that occur on the millisecond to second time scale,22 which was considered to be activation time for GPCRs in experiments, were not obtained from the simulation length. In other words, the power of computer has not yet been able to produce the fully activated conformations of GPCRs starting from an inactive one. Thus, Dror20 studied the deactivation process of β2AR from the active state to the inactive one instead of the activation process through microsecond timescale molecular dynamics simulations, since the deactivation was considered to be much faster than the activation; thus, the microsecond time of MD could be accessible. However, further confirmation is still needed to see if the deactivation process is equal to the activation one. In addition, some coarse/biased MD methods with varying efficacies were used to bridge the gap between the computational and experimental timescales in order to investigate the active state and the activation process of GPCRs.23–25

The abovementioned observations provided very valuable information for understanding the activation process and related conformational change. However, the information obtained is still far from capable of elucidating the activation mechanism for GPCRs, and thus more and further studies are urgently required.

TMD simulations, which can accelerate the transition process between two existing states with the help of a constraint, are used increasingly for understanding the large-scale conformational transition of some proteins on time scales accessible to MD simulations.26–31 Although the time scales of the TMD trajectories do not equal to the actual time of transition processes, previous studies have demonstrated that this method can give qualitatively correct pathways of conformational changes. For example, TMD simulations have been successfully used to predict the reaction pathway for the allosteric processes of human glucokinase.26 The molecular switch in signal transduction and the pathways for the conformational transition in ras p21 have also been successfully studied using TMD method.27 TMD was also used to study the transition pathways of ATP-binding cassette transporter BtuCD,28 the conformational transition of response regulator RR46829 and how inhibitors open a connection “Gate” between allosteric and catalytic sites for HIV-1 reverse transcriptase.30

On the basis of these abovementioned considerations, we have used TMD to study the activation process of β2AR from the inactive state to the active one. β2AR, as an important receptor of Class A GPCRs, shares a common structural fold of seven α membrane-spanning helices connected by three intracellular and three extracellular loops. It is also an important target for cardiac and asthma drugs.32,33 Recently, a crystal structure of β2AR in complex with an agonist and the G protein was solved,14 which was first reported to be an active state fully activated. Previously reported β2AR structures with various bound-agonists were almost identical to the inactive β2ARs with inverse-agonists bound due to a lack of an intracellular binding partner. The real active β2AR significantly advances our understanding of the active GPCR structures34 and also provides a reasonable target structure for our TMD simulation. To the best of our knowledge, the activation pathway of β2AR from the inactive state to the active one should be studied first, based on the new active crystal structure and TMD method. Thus, our work could provide new insight to previous studies for elucidating the activation mechanism of β2AR.

In addition, as a reference, we also used a 110 ns conventional molecular dynamics (CMD) simulation to observe the conformational changes of the inactive β2AR and its D130N (Asp130–Asn) mutant that have been shown experimentally to increase ligand-independent activity.35 We hope that a comparison of the results between the CMD and TMD simulations can provide useful information for better understanding the conformational transition in the activation pathway and improve strategies of computational studies on GPCRs.

Materials and methods

System preparation

Initial coordinates for the inactive and active states of β2AR were taken from the X-ray crystal structures (PDB ID 2RH1 and 3SN6).34,36 In order to match the two initial structures, all the non-receptor molecules were removed.37 The two protein structures obtained were inserted into a well-prepared phospholipids bilayer POPC,38 in which the TM4 and half of the TM6 were perpendicular to the membrane plane.39 The lipids whose P atoms fell within 0.5 Å of the receptor were removed. Then, the whole system was solvated in a rectangular water box (TIP3P water model) and the water molecules within 5 Å of the lipid and the receptor eliminated. As a result, each of the systems contain about 30 000 water molecules.

To remove bad contacts in the initial geometries of the systems,39 the systems were 20 000-steps minimized by a steepest descent method combined with the conjugate gradient algorithm. Then, they were heated gradually from 0 to 310 K within 120 ps using the Berendsen temperature coupling.30 The coordinates and trajectories were stored for the following TMD and CMD simulations.
Targeted molecular dynamics simulation

Targeted molecular dynamics was then carried out in the NPT ensemble as implemented in the Sander module of the Amber 12 package by applying the following time-dependent energy function:

\[ U_{\text{TMD}} = \frac{1}{2}NK(\text{RMSD} - \rho(t))^2 \]  

(1)

where \( N \) gives the number of atoms used as the template, \( K \) is the harmonic force constant, and RMSD is the root-mean square deviation between a configuration at time point \( t \) and the target configuration. The reference RMSD value at time \( t \) is given by \( \rho(t) \). When \( \rho(t) \) is monotonically decreased, the moving structure is gradually driven toward the target structure.

Here, the TMD simulations were performed using the inactive and active MD coordinates obtained above as the initial structure and the target one, respectively. Moreover, the NPT ensemble was utilized at a constant temperature of 310 K and a constant pressure of 1 bar. The SHAKE algorithm was applied to constrain all the bonds involving a hydrogen atom with a tolerance of \( 1.0 \times 10^{-5} \) Å.41 Non-bonding interactions were handled with a 10 Å atom-based cut-off. The particle mesh Ewald (PME)42 method was applied to treat electrostatic interactions with a 10 Å non-bonded cut-off. Difference force constants (0.5, 1, 2, 5, 10 kcal mol\(^{-1}\) Å\(^{-2}\)) and simulation times (0.5 ns/1 ns/2 ns/5 ns) were tested to determine the optimal TMD conditions. The integration step for the TMD simulations was set to 2 fs. For analysis, the trajectories were saved at an interval of 1 ps.

Conventional molecular dynamics simulation

110 ns CMD simulations were carried out in the NPT ensemble for the wild β2AR and the D130N mutant using a similar procedure to that described above. The trajectories were saved at an interval of 5 ps for analysis.

Principal component analysis

Principal component analysis was performed using the Ptraj module of Amber 12. To obtain the proper trajectory matrix, the overall translation or rotation motion was removed by fitting the coordinate data to the average structure. The obtained trajectory data was utilized to generate a covariance matrix between the Cα atoms of \( i \) and \( j \), which is defined in eqn (2):

\[ C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle / N \quad (i, j = 1, 2, 3, \ldots, 3N) \]  

(2)

where \( x_i \) and \( x_j \) are Cartesian coordinates of the \( i \)th and \( j \)th Cα atoms, \( N \) is the number of the Cα atoms considered, and \( \langle x_i \rangle \) and \( \langle x_j \rangle \) represent the time average over all the configurations obtained in the molecular dynamics simulation.43

Clustering analysis

The conformations from the TMD trajectories were clustered by a K-means clustering method44 embedded in the Ptraj module of Amber 12, taking the RMSD of the Cα atoms as the similarity measure.

All MD simulations were performed using the Amber 12 package.45 Amber 03 and GAFF force fields were utilized for the receptor and POPC molecules, respectively. All MD trajectories were analyzed using the Ptraj module of AMBER 12.0 and VMD46 as well as some other developed specific trajectory analysis software.

Results and discussion

Targeted molecular dynamics simulation conditions

Simulations were started from the inactive crystal structure to the active crystal structure (see Materials and methods part for details). Herein, we tested difference force constants and simulation times in order to find a reasonable TMD condition. Five TMD simulations (labelled as A1–A5) were carried out with force constants of 0.5, 1, 2, 5 and 10 kcal mol\(^{-1}\) Å\(^{-2}\), respectively, in which the simulation times were 1 ns. In addition, the other four TMD simulations (labelled as B1–B4) were performed with four different simulation times (i.e., 0.5 ns, 1.5 ns, 2 ns and 5 ns) and the same force constant of 1 kcal mol\(^{-1}\) Å\(^{-2}\). We monitored changes in the root-mean-square deviation (RMSD) values (see Fig. 1 and 2) in order to observe conformation fluctuations in the overall structure of the receptor. The RMSD values were calculated with respect to the inactive crystal structure (PDB entry 2RH1). As shown in Fig. 1 and 2, the RMSDs along
the different TMD trajectories display similar change trends, indicating that the different simulation times and force constants give similar conformation transition pathways. It also confirms that the TMD method used in the study was reliable. Thus, we will use one trajectory with 1 ns simulation time and 1 kcal mol\(^{-1}\) Å\(^{-2}\) force constant (viz., A2 model) to discuss the activation process from the inactive state to the active state in detail.

### Cluster analysis

As reflected from Fig. 1 and 2, there are significant conformational changes in the activation pathway. Observations from fluorescence spectroscopy and ligand-binding affinity measurements suggested that the activation by full agonists involve at least one intermediate conformation.\(^{47,48}\) Thus, in order to identify the most significant intermediate state, the conformations in the trajectory were grouped into three clusters by a \(K\)-means method, taking the RMSD of the \(C_\alpha\) atoms as the similarity measure. The three clusters were projected on the RMSD trajectory and differentiated by three colors (see Fig. 3). It is not surprising that the three clusters obtained are orderly distributed along the simulation time. The representative conformations from the three clusters are selected and superimposed on the inactive and active crystal structures, as depicted in Fig. 3. A careful inspection of the RMSD values and the three superimposed structures suggests that the representative conformation from the first cluster most closely approaches the inactive crystal structure relative to the two other representative conformations. The conformation from the third cluster exhibits an opposite trend, and resembles the active crystal structure. The second one seems to be between the inactive and active structures, which is probably close to the intermediate conformation proposed by experiments. Thus, it is reasonable to assume that the first cluster should represent the inactive conformational group, denoting the first stage in the transition path. The second and third clusters should correspond to the intermediate/second and the active/third stages, respectively. In order to identify the significant structural features discriminating the three stages, we applied principal component analysis (PCA) to examine the \(C_\alpha\) atom motion along the principal components (PCs). The resulting eigenvalue contribution of PCA shows that 83% of these motions are captured in the first three components, as shown in Fig. 4a. Moreover, we have mapped all conformations of the TMD trajectory onto the space defined by the first three PCs, in which three significant conformational groups can be identified and are very similar to the three clusters obtained by the \(K\)-means method described above, as observed from Fig. 4b. Accordingly, the crystal structures of the inactive (blue star) (PDB: 2RH1) and the active \(\beta2\)Rs (red star) (PDB: 3SN6) were also projected into the space as a reference. As shown in Fig. 4, the inactive crystal structure fell into the inactive conformational cluster while the active crystal structure was located in the active conformational cluster, providing support for the reliability of the three stages obtained. These observations indicate that the first three PCs indeed represent the main geometrical features differentiating the three stages in the activation pathway. Thus, we further analyzed the contribution of each residue to the first three PCAs in order to gain insight into the natural motion characterized by the first three principal components. The results are shown in Fig. 5, where the height of each bar denotes the maximum atomic displacement of each residue for a given PC. As can be seen from Fig. 5, the subdomains of the structural...
variation captured by PC1 are almost distributed global regions, displaying contributions from the seven helices, intracellular loops (ICLs) and extracellular loops (ECLs) to the activation. However, it is noteworthy that the intracellular loop 2 (ICL2) exhibits a more remarkable contribution to the first PC compared to the other domains, indicating its important role in the activation. The subdomains with high contribution to PC2 include ECL2, ICL2, and the terminal parts of TM1, TM3, TM4 and TM6. PC3 characterizes ECL2 and the concerted maximum displacement of TM4, TM5 and TM6, which are located in the middle region of the GPCR structure. The important roles of ECL2, ICL2, TM3, TM5, TM6 and TM7 in the activation were also reported by previous experimental studies, in line with our PCA analysis. It is known that β2AR structurally contains extracellular terminal, intracellular terminal and connected regions. Functionally, β2AR is composed of a ligand-binding site (close to the extracellular loops of the receptor) and G-protein binding site (located on the intracellular loops of the receptor). The signal transmitting inspired by the activation is from the extracellular region to the intracellular G protein. Thus, we will give out a more detailed analysis on the conformation transition of these important regions revealed by the abovementioned observations, organizing the discussion in terms of the order from the extracellular region to the intracellular one.

ECL2: important regions for ligand entrance

The ECL2, which was revealed from the principal component analysis (PCA) above, significantly contributes to the first three PCs and has been proposed to act as a ‘gatekeeper’ for agonist binding and associated with the specificity of ligand binding. Thus, we are concerned with the conformation transition of ECL2 between the three conformational clusters. Fig. 6 shows the superimposition of the representative structures from the three clusters, which resemble the inactive, intermediate, active groups, respectively. The ECL2 in the active cluster displays most inward migration, moving outwards away from TM7 towards TM4 and facilitating the ‘opening’ movement of the binding crevice for the ligand entrance (see Fig. 6). Interestingly, the inward migration of ECL2 in the second/intermediate cluster was much more obvious than that found in the third one, indicating that ECL2 moves back in the third stage and exhibits a ‘closed’ state for the binding crevice. The observation suggests that ECL2 may start with an ‘open’ conformation in order to allow the entrance of the ligand, and ends with a ‘closed’ state after the ligand moved into the binding site, which is consistent with experimental observations from other Class A receptors of GPCRs and MD studies on the interactions between GPCRs and ligands.

Ligand binding site

As revealed, the ligand-binding site is located in the extracellular region and the primary change in the ligand binding site upon activation is the movement of the N-terminal of helices. Table 1 lists the residues involved in the ligand binding site. It is reported that there are no large conformation differences in the ligand binding site between the active crystal structure and the inactive one, which is consistent with the RMSD values calculated between the active and inactive crystal structures (see Table 1). However, our TMD calculation shows that the binding residues exhibit remarkable fluctuations during the activation process, as shown by the root means standard fluctuation (RMSF) values calculated relative to the average conformation derived from the TMD trajectory (see Table 1). The observation suggests that a simple comparison between the active and inactive crystal structures does not efficiently reflect the dynamics behavior in the activation process due to the high flexibility of these helices. As can be seen from Table 1, Tyr316 displays the highest flexibility and Ser207 is second to Tyr316. Furthermore, the active crystal structure of β2AR indicated that the greatest difference between the inactive and active structures in the ligand-binding site is an inward bulge of TM5 centered on Ser207. Thus, we take the distance between the Cz atoms of Ser207 and Tyr316 as an indicator to determine the
dynamics changes of the ligand-binding site in the activation pathway (see Fig. 7a). It can be seen from Fig. 7a that the distance almost fluctuates near the value of the inactive state (\( \sim 16.1 \) Å) in the first two stages. Only in the third stage, it starts to drop and gradually approaches the value of the active crystal structure (\( \sim 13.9 \) Å). This observation indicates that the binding site is still in its inactive state, although the overall structure of \( \beta 2 \)AR already exhibits the features of the intermediate state.

As a reference, we also performed 110 ns CMD simulations on the wild \( \beta 2 \)AR and its D130N mutant (see the Methods part for details). Fig. 7b shows the changes in distance between Ser207\(^{5.46}\) and Tyr316\(^{7.43}\) with respect to the 110 ns simulation time. It is clear that the distances in both the wild receptor and the mutated one are significantly larger than the values in the wild receptor at the 110 ns simulation time. It is also observed that the distance almost fluctuates near the value of the inactive state (i.e., the mutant system may experience a longer time to achieve the active state than the other subdomains. Thus, the CMD method will be difficult in observing the activation of the binding site.

### Important regions of the G-protein binding site

The G-protein binding site located in the cytoplasmic end is adjacent to the intracellular loops of \( \beta 2 \)AR. The highly conserved NPxxY in TM7 has been considered to be closely related to the G protein binding and plays an important role in maintaining the structure of the inactive state. Furthermore, the analysis described above on the contribution from each residue also shows that TM7 plays a significant role in discriminating the three stages, as reflected by the PC1 in Fig. 5. Thus, we have analyzed the dynamic changes in the NPxxY region of TM7 (Asn322\(^{7.49}\) – Tyr326\(^{7.53}\)) during the activation process through monitoring its RMSD change (see Fig. 8a). In addition, 3D models of the NPxxY region of the representative conformations from the three stages are also displayed in Fig. 8a. It can be seen from the RMSD values in Fig. 8a that there are no significant changes observed in the structures of the NPxxY region in the first stage and the beginning of the second stage, which are similar to the inactive state. From the middle stage of the intermediate state to the initial stage in the third state, the RMSD values of the NPxxY region gradually increase to be close to the active value and are maintained in the remaining time of the third stage. The 3D models in Fig. 8a further revealed that the intracellular end of TM7 was moved inward to TM3 upon activation. Such shift was reported from the microsecond MD simulation of the deactivation process for \( \beta 2 \)AR. The observations described above suggest that the significant change in the structure of the NPxxY region occurs in the second/intermediate stage, presenting partially activated features. As observed above, the ligand binding site was still in the inactive conformation in the second/intermediate stage. Thus, it is reasonable to assume that the NPxxY region near the G-protein binding site seems to be activated earlier than the ligand binding site although the NPxxY region is far away from the ligand binding region and the signal is transmitted from the ligand binding site to the G-protein bind region. Fig. 8b shows the RMSD changes along with the 110 ns CMD simulation time for the NPxxY region of the wild \( \beta 2 \)AR and the D130N mutant. It is observed from Fig. 8b

### Table 1: The RMSD and RMSF values for the ligand binding site residues.

<table>
<thead>
<tr>
<th>Resid ID</th>
<th>Hist93</th>
<th>Trp109</th>
<th>Thr110</th>
<th>Asp113</th>
<th>Val114</th>
<th>Val117</th>
<th>Thr118</th>
<th>Thr195</th>
<th>Tyr199</th>
<th>Ala200</th>
<th>Ser203</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>TM2</td>
<td>TM3</td>
<td>TM3</td>
<td>TM3</td>
<td>TM3</td>
<td>TM3</td>
<td>ECL2</td>
<td>TM5</td>
<td>TM5</td>
<td>TM5</td>
<td>TM5</td>
</tr>
<tr>
<td>RMSD/Å</td>
<td>0.309</td>
<td>0.401</td>
<td>0.728</td>
<td>0.760</td>
<td>1.186</td>
<td>0.904</td>
<td>1.346</td>
<td>0.950</td>
<td>1.402</td>
<td>1.348</td>
<td>1.040</td>
</tr>
<tr>
<td>RMSF/Å</td>
<td>5.953</td>
<td>1.742</td>
<td>1.844</td>
<td>2.556</td>
<td>2.511</td>
<td>2.778</td>
<td>2.284</td>
<td>2.549</td>
<td>4.905</td>
<td>4.050</td>
<td>5.459</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resid ID</th>
<th>Ser204</th>
<th>Ser207</th>
<th>Trp286</th>
<th>Phe289</th>
<th>Phe290</th>
<th>Lys305</th>
<th>Tyr308</th>
<th>He309</th>
<th>Asn312</th>
<th>Tyr316</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>TM5</td>
<td>TM5</td>
<td>TM6</td>
<td>TM6</td>
<td>TM6</td>
<td>TM6</td>
<td>TM7</td>
<td>TM7</td>
<td>TM7</td>
<td>TM7</td>
</tr>
<tr>
<td>RMSD/Å</td>
<td>1.150</td>
<td>1.900</td>
<td>0.592</td>
<td>0.617</td>
<td>0.593</td>
<td>0.640</td>
<td>0.581</td>
<td>0.824</td>
<td>0.411</td>
<td>0.535</td>
</tr>
</tbody>
</table>

**Fig. 7** The time-dependent changes in the distance of the two key residues (Ser207\(^{5.46}\) and Tyr316\(^{7.43}\)) in the ligand binding site for 1 ns TMD (a) and 110 ns CMD simulations on the wild \( \beta 2 \)AR (WT) and its D130N mutant (D130N) (b). The blue, yellow and cyan scales on the horizontal axis represent the inactive, intermediate and active stages or clusters identified by clustering analysis. The green and red horizontal lines represent the inactive, intermediate and active stages or clusters identified by clustering analysis.
that the D130N mutation causes more significant fluctuations in the NPxxY region than that found in the wild β2AR. The RMSD values of the NPxxY region are significantly larger in the mutant type, as confirmed by Student’s t-test (P < 0.05). However, the values of the mutated receptor in the last 30 ns are still close to that in the first stage/cluster identified by the TMD simulation. The observation suggests that the 110 ns CMD simulation was inaccessible to the real active state for the NPxxY region in either the mutated system or the wild β2AR, although the mutation to some extent speeds up the transition of the NPxxY motif from the inactive state to the active one.

ICL2

As revealed by the PCA described above, ICL2 exhibits a remarkable contribution to PC1 and PC2 and is an important feature in differentiating the three clusters or stages. The active crystal structure of β2AR-Gs revealed that there are interactions between ICL2 and the G-protein, leading to a significant difference in the ICL2 conformation between the inactive state and the active one. Our observations from the PCA described above further confirm the experimental finding. Thus, we examined in detail the dynamic changes in the ICL2 region during this part. Fig. 9a displays the RMSD change trend for ICL2 along with the activation pathway. As reflected by Fig. 9a, the RMSD values in the second/intermediate stage become larger than those in the first stage, but it is still much smaller than those in the third stage.

In the beginning of the third/active stage, the RMSDs of ICL2 continuously increase up to the value in the active crystal structure. The change trend is similar to that found in the NPxxY region; i.e., the ICL2 region also presents partially activated features in the intermediate state, but, not fully activated. Fig. 9b displays the change trend in the RMSD values of ICL2 with respect to the 110 ns CMD simulation time. An inspection of Fig. 9b and Student’s t-test (P < 0.05) confirms that the RMSD values of ICL2 are significantly higher in the mutant system than those in the wild one over the 110 ns simulation time, approaching the value in the second/intermediate state identified by the TMD simulation described above. After the 110 ns CMD simulation, the RMSD value for ICL2 in the wild receptor was still close to the first/inactive state. These observations suggest that the D130N mutation could advance the ICL2 motif to reach the intermediate state more quickly within the 110 ns CMD simulation time. Since it was reported that the stability of ICL2 is closely associated with the DRY motif, we also analyzed the changes in the DRY motif upon activation. It is known that the hydrophilic DRY motif is located in the intracellular ends of the transmembrane segments and consists of the residues Asp130ICL2, Arg131ICL2, and Tyr141ICL2 for β2AR. It is a highly conserved sequence in the GPCR subfamilies and has been considered to play an important role in stabilizing the inactive state and the GPCRs activation. Both the active and inactive crystal structures exhibit a hydrogen bond network in the DRY region, as shown in Fig. 10b. In the inactive crystal
structure, two H-bonds were observed between Asp1303.49 and Arg1313.50. However, in the active crystal structure, the H-bonds are broken, accompanied by the formation of two new H-bonds involving OD2@Asp1303.49...OG1@Thr682.39 and OD1@Asp1303.49...OH@Tyr141ICL2. This observation shows that there is difference in the type of residues involved in the H-bond network between the inactive and active states, implying significant conformation changes in the DRY motif upon activation.

To estimate the dynamic changes in the H-bond network during the activation process, we have analyzed the change in the distances between the H-bond donor atoms and the H-bond acceptor atoms for these H-bond residues, as shown in Fig. 10a. It is clear that these two H-bonds between Asp1303.49 and Arg1313.50 have been stable during the first two stages because the distances are always smaller than 3.0 Å. In the inactive stage, Asp1303.49 is far away from Thr682.39, as observed in Fig. 10a. However, the activation-induced TM3 deviation from TM6 makes Asp1313.50 tilt to Thr682.39 to form a new H-bond in the beginning of the intermediate stage and the H-bond is kept stable during the remaining time of the second stage, as reflected by Fig. 10. The H-bond between Asp1303.49 and Arg1313.50 in the first two stages could to some extent increase the conformational rigidity of the C-terminal domain of TM3, facilitating the departure movement of TM3 from TM6, which was proposed to the most significant feature for the activation.15 In the active stage, the H-bond between Asp1303.49 and Thr682.39 is kept stable. However, the distance between Asp1303.49 and Arg1313.50 was sharply increased to be larger than ~4.5 Å in the beginning of the third stage, indicating full disruption of the H-bonds involved in the two residues. In addition, the remarkable ICL2 conformation changes upon activation significantly shift Tyr141ICL2 to approach Asp1303.49, leading to the formation of a new H-bond between Asp1303.49 and Tyr141ICL2 in the active state, as shown in Fig. 10. The disruption of the H-bond between Asp1303.49 and Arg1313.50 also to some extent facilitates the H-bond formation of OD1@Asp1303.49...OH@Tyr141ICL2 in the third/active stage. Table 2 summarizes population of the H-bonds in the three stages, displaying the significant differences. The H-bond between Asp1303.49 and Tyr141ICL2 in the active state should favor the interfacial domain to maintain its structural integrity, facilitating the DRY hydrophilic cavity to interact with the G protein. This observation further confirmed that the H-bonds in the region play an important role in the conformational transition in the activation process. The CMD simulation for wild β2AR shows that the two H-bonds between Asp1303.49 and Arg1313.50 keep stable over the 110 ns simulation time, as confirmed by the short H-bond distances (<3 Å) calculated, while the H-bonds of Asp1303.49...Thr682.39 and Asp1303.49...Tyr141ICL2 do not exist since their distances are much larger than the H-bond distances, as reflected by Fig. 11. In the D130N mutant system, the H-bonds between Asp1303.49...Arg1313.50 are also broken due to the mutation of Asp1303.49 to Asn1303.49, as shown by their larger bond distances than the H-bond distance of 3.5 Å (see Fig. 11). Similarly, there are no H-bonds between Asn1303.49...Thr682.39 and Asn1303.49...Tyr141ICL2 in the mutant since their distances are higher than 5 Å. However, the distance between Asn1303.49 and Tyr141ICL2 in the mutant is much smaller than that in the wild system, approaching the

![Fig. 10](image-url)

**Table 2** The H-bond percentage occupation (%) in the three clusters over the TMD trajectory*

<table>
<thead>
<tr>
<th>H-bond</th>
<th>Cluster1 (%)</th>
<th>Cluster2 (%)</th>
<th>Cluster3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OD2@Asp1303.49...NH2@Arg1313.50</td>
<td>100</td>
<td>95.00</td>
<td>3.24</td>
</tr>
<tr>
<td>OD2@Asp1303.49...NE@Arg1313.50</td>
<td>99.20</td>
<td>96.57</td>
<td>14.47</td>
</tr>
<tr>
<td>OD2@Asp1303.49...OG1@Thr682.39</td>
<td>0</td>
<td>96.29</td>
<td>99.84</td>
</tr>
<tr>
<td>OD1@Asp1303.49...OH@Tyr141ICL2</td>
<td>0</td>
<td>0</td>
<td>93.55</td>
</tr>
</tbody>
</table>

*a OD2@Asp1303.49 depicts OD2 atom of Asp1303.49. NH2@Arg1313.50 depicts NH2 atom of Arg1313.50. NE@Arg1313.50 depicts NE atom of Asp1313.50. OG1@Thr682.39 depicts OG1 atom of Thr682.39. OH@Tyr141ICL2 depicts the OH atom of Tyr141ICL2.
value exhibited in the middle stage of the intermediate state. This observation suggests that the mutation could speed up the shift of Tyr141ICL2 to Asn1303.49, favoring the activation associated with H-bond formation between Asn1303.49 and Tyr141ICL2.

Ionic lock

It was reported that the ionic lock between Arg1313.50 and Glu2686.30 plays an important role in maintaining the inactive states of some GPCRs like rhodopsin, alpha (1b) adrenergic receptors, delta opioid receptors of 6 and 5. The breakage of the ionic lock has been suggested to regulate the activation of different family A GPCRs by in vitro and computational experiments. In order to observe the dynamic behavior of the ionic lock during the activation pathway, we monitored the salt bridge distance between Arg1313.50 and Glu2686.30 (see Fig. 12a). As shown in Fig. 12a, the distance was about 11 Å in the inactive stage, very close to the distance of 11.2 Å found in the inactive crystal structure. Then, it displays a sharp increase in the beginning of the second/intermediate stage and approaches ~17 Å (close to 17.3 Å in the active crystal structure) in the middle of the second stage and is maintained in the remaining stages (including the third/active stage). The observation reveals that the ionic lock was fully disrupted in the intermediate and active states. In addition, we also examined the distance changes between the two helices (viz., TM3 and TM6) associated with the ionic lock since their conformational changes have been considered as the most significant feature in the activation of GPCRs. As can be seen in Fig. 12b, the distance change between TM3 and TM6 is almost similar to that of the ionic lock, providing support for the experimental observations that the separation of TM3 and TM6 can be characterized by the disruption of the special ionic lock. However, it was further observed from the TMD results that when the ionic lock between the two helices already presents the full disruption conformation in the middle of the second stage, the distance between TM3 and TM6 is still close to that in the inactive state, suggesting that the full separation of TM3 and TM6 lags the disruption of the ionic lock.

Fig. 12c and d show the changes of the ionic lock and the distance between TM3 and TM6 along with the 110 ns CMD simulation time for the wild β2AR (WT) and its D130N mutant (D130N).
as confirmed by a Student’s t-test ($P < 0.05$). However, the distances between TM3 and TM6 in the wild and mutated β2ARs are close to those found in the inactive stage, far away from the active state, as shown in Fig. 12b and d. These observations indicate that the D130N mutation significantly favors the disruption of the ionic lock and the separation of TM3 and TM6 compared to the wild type. But, the 110 ns CMD simulation time is not enough to be accessible to the intermediate state for the wild and mutated β2ARs, and at most makes the ionic lock approach to the intermediate state while the distance between TM3 and TM6 is still in the inactive state.

**Conclusions**

In this work, we have mainly used targeted molecular dynamics (TMD) simulations to overcome the time limitation of the conventional molecular dynamic (CMD) simulations to explore the conformation transition of β2AR in the activation process from the inactive state to the active one, based on the real active structure recently solved and the inactive crystal structure. In addition, the 110 ns CMD simulations on wild β2AR and its D130N mutant were also carried out to gain more insight into the conformation changes and the effect of the mutation on the activation process.

Clustering analysis of the targeted molecular dynamic trajectory presents the three conformational clusters, which resembles the inactive, intermediate and active states/stages, respectively. We have further analyzed the dynamic changes in some important regions identified by principal component analysis (PCA), including regions involved in the ligand binding (ECL2, residues in the ligand binding site) and ones associated with G-protein binding (ICL2, NpXxY, the key ionic lock and the distance between TM3 and TM6). The observations from the dynamic transitions of these regions further confirm the existence of an intermediate state different from the inactive and active clusters. It is interesting to observe that the dynamic changes in some important regions of G-protein binding site, like ICL2, NpXxY, the ionic lock, the separation of terminal part of TM3 and TM6 and the H-bonding in the DRY region, are coupled. In other words, they synchronously experience the process from the inactive, intermediate to active states. However, it is surprising to observe that the activation of the ligand binding site lags the G-protein binding site, displaying uncoupled correlation, which is consistent with the findings from the microsecond CMD study on a deactivation process of β2AR from an active structure to an inactive one. Thus, the observations from our work provides further support for the assumption that the activation process probably starts with the G-protein-binding despite of the fact that the signal transits from the extracellular loop to the intracellular loop. In addition, the observations from the residues involved in the ligand binding site reveal that the ligand-binding residues exhibit high flexibility in the activation process despite the small differences observed in the ligand binding site between the active structure and the inactive one.

On the other hand, a comparison of some results from the TMD simulations with those from the two 110 ns CMD simulations revealed that the D130N mutation could significantly speed up the ICL2 and the key ionic lock to enter into the intermediate state and to some extent advance the activation of NPxxY, the H-bonding in the DRY region and the separation between TM3 and TM6. However, the D130N mutation does not exhibit any contribution to activation of the ligand binding site within the 110 ns timescale. In a whole, the findings from this work can provide valuable information for understanding the molecular mechanism concerning the conformational transition in the GPCRs activation process.

**Acknowledgements**

This project is supported by the National Science Foundation of China (Grant No. 21273154 and U1230121).

**References**
