



OPEN

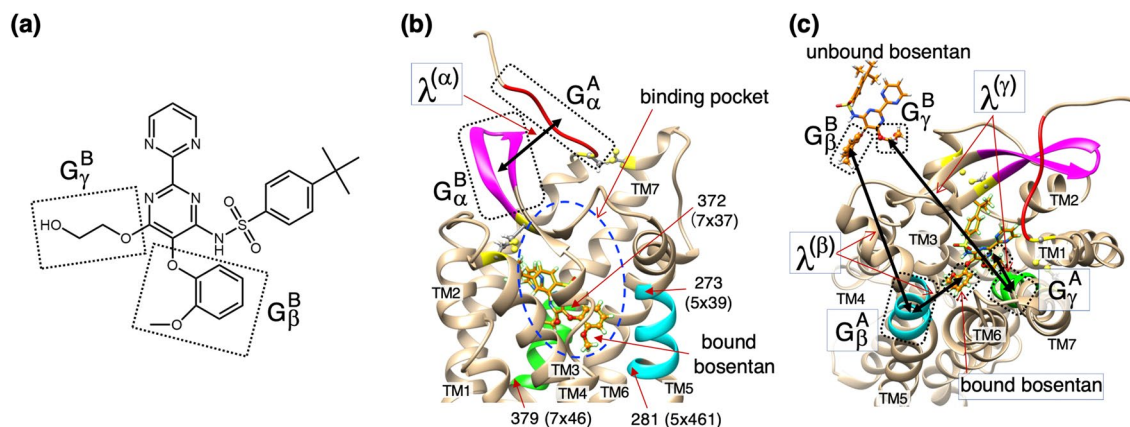
## Fly casting with ligand sliding and orientational selection supporting complex formation of a GPCR and a middle sized flexible molecule

Junichi Higo<sup>1,2✉</sup>, Kota Kasahara<sup>3</sup>, Gert-Jan Bekker<sup>4</sup>, Benson Ma<sup>4</sup>, Shun Sakuraba<sup>5</sup>, Shinji Iida<sup>6,8</sup>, Narutoshi Kamiya<sup>1</sup>, Ikuo Fukuda<sup>1</sup>, Hidetoshi Kono<sup>5</sup>, Yoshifumi Fukunishi<sup>7</sup> & Haruki Nakamura<sup>4</sup>

A GA-guided multidimensional virtual-system coupled molecular dynamics (GA-mD-VcMD) simulation was conducted to elucidate binding mechanisms of a middle-sized flexible molecule, bosentan, to a GPCR protein, human endothelin receptor type B (hETB). GA-mD-VcMD is a generalized ensemble method that produces a free-energy landscape of the ligand-receptor binding by searching large-scale motions accompanied with stable maintenance of the fragile cell-membrane structure. All molecular components (bosentan, hETB, membrane, and solvent) were represented with an all-atom model. Then sampling was conducted from conformations where bosentan was distant from the binding site in the hETB binding pocket. The deepest basin in the resultant free-energy landscape was assigned to native-like complex conformation. The following binding mechanism was inferred. First, bosentan fluctuating randomly in solution is captured using a tip region of the flexible N-terminal tail of hETB via nonspecific attractive interactions (fly casting). Bosentan then slides occasionally from the tip to the root of the N-terminal tail (ligand-sliding). During this sliding, bosentan passes the gate of the binding pocket from outside to inside of the pocket with an accompanying rapid reduction of the molecular orientational variety of bosentan (orientational selection). Last, in the pocket, ligand-receptor attractive native contacts are formed. Eventually, the native-like complex is completed. The bosentan-captured conformations by the tip-region and root-region of the N-terminal tail correspond to two basins in the free-energy landscape. The ligand-sliding corresponds to overcoming of a free-energy barrier between the basins.

So-called G protein-coupled receptors (GPCRs) are membrane proteins constructing a large evolutionarily related protein family with various molecular functions<sup>1</sup>. Generally speaking, GPCRs have seven transmembrane helices (TM1–TM7) embedded in the membrane. GPCRs activate cellular responses via detection of molecules outside the cell. Because GPCRs are related to many diseases, they have been important targets for drugs<sup>2</sup>. Endothelin-1 (ET1) is a peptide (21 residues long) with a strong vasoconstrictor action discovered in humans<sup>3</sup>. To exert its activity, ET1 transmits signals by interacting with two GPCRs: endothelin type A (ETA)<sup>4</sup> and endothelin type

<sup>1</sup>Graduate School of Information Science, University of Hyogo, 7-1-28 Minatojima Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. <sup>2</sup>Research Organization of Science and Technology, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan. <sup>3</sup>College of Life Sciences, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan. <sup>4</sup>Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan. <sup>5</sup>Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, 8-1-7 Umemi-dai, Kizu, Kyoto 619-0215, Japan. <sup>6</sup>Technology Research Association for Next-Generation Natural Products Chemistry, 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan. <sup>7</sup>Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan. <sup>8</sup>Present address: Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7, Aomi, Koto-ku, Tokyo 135-0064, Japan. ✉email: higo@protein.osaka-u.ac.jp



**Figure 1.** (a) Chemical structure of bosentan. Two atom groups  $G_{\beta}^B$  and  $G_{\gamma}^B$  are surrounded by dotted-line rectangles, which are used to set reaction coordinates (RCs)  $\lambda^{(\beta)}$  and  $\lambda^{(\gamma)}$ . (b) The first RC,  $\lambda^{(\alpha)}$ , defined by distance between centroids of atom groups  $G_{\alpha}^A$  (red segment in the N-terminal tail of hETB surrounded by dotted line; residues 85–89) and  $G_{\alpha}^B$  (magenta-colored segment in  $\beta$ -hairpin of hETB surrounded by another dotted line; residues 243–254). Binding pocket of hETB is indicated by blue broken-line circle. Bound bosentan is shown at the bottom of the pocket. Green and cyan regions are defined in panel (c). (c) The second and third RCs are, respectively,  $\lambda^{(\beta)}$  and  $\lambda^{(\gamma)}$ . Also,  $\lambda^{(\beta)}$  is defined by the distance between centroids of  $G_{\beta}^A$  (a cyan-colored part of the fifth transmembrane helix TM5; residues 273–281 ( $5 \times 39$ – $5 \times 461$ ), in which the first and last residues are indicated by “273( $5 \times 39$ )” and “281( $5 \times 461$ )”, respectively, in the panel) and  $G_{\beta}^B$  (a part of bosentan surrounded by dotted line).  $\lambda^{(\gamma)}$  is defined by  $G_{\gamma}^A$  (a green part of the seventh transmembrane helix TM7; residues 372–379 ( $7 \times 37$ – $7 \times 46$ ), in which the first and last residues are indicated by “372( $7 \times 37$ )” and “379( $7 \times 46$ )”, respectively, in the panel) and  $G_{\gamma}^B$  (other part of bosentan surrounded by the other dotted line). Bound and unbound bosentans are displayed. Yellow residues are cysteines that form disulfide bonds at the roots of the N-terminal tail,  $G_{\alpha}^A$ , and the  $\beta$ -hairpin,  $G_{\alpha}^B$ . Transmembrane helices (TM1–7) are labeled in panel (b) and (c).

B (ETB)<sup>5</sup> receptors. The complex structure of human endothelin type-B receptor (hETB) and ET1 was resolved using X-ray crystallography<sup>6</sup>. In this complex, ET1 is bound to a binding pocket of hETB on the cytoplasmic side.

Bosentan competes with ET1 when binding to human ETA (hETA) and human hETB (hETB). It inhibits ET1’s vasoconstriction effect as an antagonist<sup>7–9</sup>. The tertiary structure of the bosentan-hETB complex was solved using X-ray crystallography<sup>10</sup>. In this structure, bosentan binds to the binding pocket of hETB more deeply than ET1. However, interaction patterns between the C-terminal segment of ET1 and amino-acid residues in the binding pocket of hETB are similar to those in the bosentan-hETB complex.

Bosentan is a medium-sized (551.6 Da) drug molecule that is actually large compared to other commercial drug molecules. This molecule is flexible because it has long flexible sidechains (Fig. 1a), whereas the central ring, to which the sidechains are connected, is stiff. Consequently, bosentan will adopt temporal complex conformations (encounter complexes) before reaching the deep position of the hETB pocket (i.e., the native-complex position). A special molecular dynamics (MD) simulation that can sample various conformations in the complex-formation process is suitable for investigating those temporal conformations.

Recently, we introduced a generalized ensemble method, a multi-dimensional virtual-system coupled molecular dynamics (mD-VcMD) simulation<sup>11</sup>, which enhances conformational sampling with repetition of simulations (iteration) in a multi-dimensional (mD) reaction-coordinate (RC) space. By repeating the iterations, the sampled volume in the mD-RC space is expanded. The sampled conformations (snapshots) are then collected from the iterations. Importantly, mD-VcMD assigns a statistical weight (thermodynamic weight equilibrated at a simulation temperature) to each snapshot. Therefore, the ensemble of snapshots is equilibrated. Any equilibrated physicochemical quantity is computable from the ensemble if the quantity is expressed by the system’s coordinates. We designate this simulation method as the “original mD-VcMD method” or simply the “original method” in this paper. The original method was applied to a large and complicated system consisting of ET1 and hETB, where hETB was embedded in an explicit membrane and the membrane was immersed in an explicit solvent<sup>12</sup>. A benefit of mD-VcMD against other generalized ensemble methods is that RCs can be set arbitrarily to distances among atom groups. Consequently, this method can treat the molecular binding/unbinding process straightforwardly when RCs are set to distances between the two molecules (or between portions of the molecules). That is to say, enhancement of the motions of those RCs is related directly to molecular binding and unbinding.

Although the original mD-VcMD was applicable to the large and complicated system, we encountered an important difficulty: Once a poorly sampled region emerges in the mD-RC space in an iterative simulation, the conformation might be trapped in this region in subsequent iterations. Consequently, a more robust method is necessary to proceed with the sampling. To overcome this difficulty, we introduced a *genetic algorithm*-guided multi-dimensional virtual-system coupled molecular dynamics method (GA-mD-VcMD)<sup>13</sup>, which is an extension of the original mD-VcMD. We applied this method to some biological systems that consist of a protein and ligand in an explicit solvent<sup>14–16</sup>.

From the original mD-VcMD applied to binding of ET1 to hETB, two binding mechanisms were identified: Fly casting<sup>17–19</sup> and conformational selection<sup>20</sup>. The receptor hETB has an intrinsically disordered N-terminal tail. The ET1 floating in solution was captured by the fluctuating N-terminal tail (fly casting mechanism) because the interactions between ET1 and the hETB's N-terminal were attractive<sup>12</sup>. In fact, the root of the N-terminal tail is located at the gate of the binding pocket of hETB, which indicates that once ET1 is captured by the N-terminal tail, the ligand is restricted in a narrow volume around the gate of the binding pocket, resulting in an increased possibility that ET1 contacts the gate of the binding pocket. Next, only a fraction of various orientations of ET1 were allowed to enter the hETB pocket (conformational selection). This fractional entrance suggests that the pocket size is not sufficiently large for ET1 to rearrange the molecular orientation in the pocket, and that only ET1 with molecular orientations that are advantageous for the native complex formation is accepted.

For this study, we used GA-mD-VcMD simulation to sample interactions between the medium-sized molecule bosentan and hETB. Aside from the difference of ligand and simulation method, two differences exist between earlier studies and the present study. (1) We prolonged the intrinsically disordered N-terminal tail of hETB of the X-ray structure<sup>6,10</sup>, which were five and ten residues long for the earlier and the current studies, respectively<sup>12</sup>. This tail extension varies the features of the fly casting mechanism, as shown later. Because GPCRs have a long N-terminal tail in general<sup>21</sup>, this extension of the tail might support experimentation studies elucidating the role of the N-terminal tail of GPCRs<sup>22,23</sup>. Second, the initial simulation conformations for the present study were those in which bosentan was distant from hETB, whereas those for the previous study were the native-complex conformations of ET1 and hETB. In general, the difference of initial conformations influences the resultant conformational ensemble considerably. The reliability of findings obtained from the current study depends on the capacity of our method to sample native conformations. The native complex should be *discovered* as the lowest free-energy conformation out of many possible conformations. By contrast, in earlier studies, the native complex is always involved in the resultant ensemble, which means that the native complex might not necessarily correspond to the lowest free-energy conformation because a lower free-energy conformation might be overlooked in the sampling.

Our GA-mD-VcMD simulation produced a free-energy landscape of the bosentan–hETB system. Probability distribution functions of some quantities were calculated from the resultant conformational ensemble. Because the free-energy landscape covered both unbound conformations and the native-complex structures, and because the lowest free-energy conformation corresponded to the native-complex structure, we were able to analyze the complex formation comprehensively. We discuss the generality of the binding mechanism found in bosentan–hETB to ligand–GPCR binding. Because bosentan is a drug molecule, the present study helps drug development research not only by proposing the lowest free-energy conformations but also by investigating the binding process.

## Methods

The aim of this work is investigation of the complex-formation process of bosentan and hETB using GA-mD-VcMD method. First, we briefly explain this method, which provides a conformational ensemble of the system and a thermodynamic weight at the simulation temperature  $T$  (300 K in this study), which is assigned to conformations stored in the ensemble. Details of GA-mD-VcMD have been explained elsewhere<sup>13–16</sup>. Next, we explain the molecular system which consists of hETB, bosentan, membrane (cholesterol and POPC lipid molecules), and solvent (water molecules and ions). Then, we describe multiple reaction coordinates that are introduced, along with the initial conformations used for simulation. Finally, we account for the method used to calculate the spatial density of bosentan around hETB.

We frequently use the term “native-like complex”. In general, an MD simulation is influenced by some force-field errors and sampling errors. Therefore, when a simulation is initiated from a completely unbound conformation, the simulation does not perfectly reproduce the experimentally observed conformation. For this reason, we designate the computed lowest free-energy conformation as “native-like complex conformation”, even when the conformation closely resembles the experimentally observed one.

However, we use the word “native complex conformation” for a computed conformation when a short simulation is done starting from the experimentally observed conformation because this simulation can be regarded as refinement of the experimentally observed conformation.

**GA-mD-VcMD and a conformational ensemble.** To sample both the unbound and bound conformations of the system extensively, GA-mD-VcMD controls the system motions in a multi-dimensional (mD) space constructed by multiple reaction coordinates (RCs). The definition of an RC is presented in Section 1 of Supplementary Information (SI) and in Figure S1. We introduce three RCs denoted as  $\lambda^{(h)}$  ( $h = \alpha, \beta$ , or  $\gamma$ ), whereas the dimensionality of the RC space is arbitrary in theory<sup>13</sup>. Therefore, the word “multi-dimensional (mD)” designates three-dimensional (3D) for the present study. The actual definition for the three RCs is given later.

Conformational sampling of the system by GA-mD-VcMD at 300 K produces a distribution (density) of the system in the mD-RC space (not in the real space). Because the conformational space of a biological molecular system is wide, the GA-mD-VcMD simulation is executed iteratively, by which the sampled volume of 3D-RC space is expanded. Here, the density at a 3D-RC position  $[\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)}]$  is denoted as  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$ , with the density computed from the  $M$ -th iteration expressed as  $Q_{cano}^{(M)}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  when necessary.

The simulation is terminated when a convergence criterion is satisfied for  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$ . The convergence is discussed in the “Results and Discussion” section. Details of the convergence are discussed in Section 5 of SI. Snapshots of the system are saved from all iterations. After convergence, a thermodynamic weight

at simulation temperature  $T$  is assigned to each saved snapshot based on  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$ . Therefore, the snapshots construct a thermally equilibrated conformational ensemble (canonical ensemble) at  $T$  (300 K).

**Molecular system studied.** We generated the molecular system consisting of bosentan, hETB, cholesterol surrounding hETB, the POPC bilayer where hETB is embedded, and an explicit solvent (water molecules,  $\text{Na}^+$  and  $\text{Cl}^-$  ions). All of them were packed in a periodic boundary box. Details of system preparation are presented in Section 2 of SI. Figure S2 is a diagram outlining generation of the molecular system.

Two reference crystal structures are available for generating the molecular system for the simulation: The bosentan–hETB (PDB ID 5xpr; 3.6 Å resolution) and K8794–hETB (PDB ID 5x93; 2.2 Å) complex structures. Importantly, the two crystal structures are mutually similar to a considerable degree (Figure S3a) because K8794 is an analog of bosentan (Figure S3b). Because K8794–hETB complex has better resolution than bosentan–hETB complex, we used the K8794–hETB complex to prepare the molecular system.

First, we replaced K8794 with bosentan. This complex conformation is designated as “Complex structure of bosentan–hETB” in Figure S2. For crystallography, some mutations were applied to the amino-acid sequence of hETB to increase the diffraction data quality<sup>10</sup>. Then, we back-mutated these amino acids to those in the wild-type sequence of hETB. Finally, the computed hETB consists of residues 80–403. The N-terminal and C-termini of hETB were capped respectively by the acetyl and N-methyl groups denoted as Ace and Nme. Section 2 of SI presents details of system preparation.

Then the complex was embedded in the POPC bilayer. After four cholesterol molecules were introduced into the hETB–membrane interface (Section 2 of SI), the system was immersed in a periodic boundary box filled with water molecules. The x-, y-, and z-coordinate axes for specifying the atomic positions were set parallel to the sides of the periodic box (the x–y plane was parallel to the membrane surface). The box dimensions were 71.33 Å (x-axis), 71.33 Å (y-axis), and 132.17 Å (z-axis). Finally, sodium and chlorine ions were introduced by replacing water molecules randomly with ions to neutralize the system at a physiological ion concentration of 130 mM. This conformation is shown as the “Complex structure in periodic box” in Figure S2. The total number of atoms of the system was 69,062, with the numbers of protein and bosentan atoms as 5289 and 68, respectively, and with four cholesterol molecules (296 atoms), 127 POPC molecules (17,018 atoms), 15,440 water molecules (46,320 atoms), and 28 sodium and 43 chlorine ions.

After energy minimization of the above system, with subsequent NVT and NPT simulations, we obtained the complex structure shown in Figure S4, for which the resultant periodic-box size was  $69.37 \text{ \AA} \times 69.37 \text{ \AA} \times 140.29 \text{ \AA}$ . This structure is represented as “Complex structure in Figure S4 (native-complex structure)” in Figure S2. Because this complex structure is a relaxed conformation of the X-ray complex structure, bosentan was located at a position close to that in the X-ray structures (PDB IDs: 5xpr and 5x93). Therefore, we designate this conformation as the “native complex structure” in the present study. It is noteworthy that this native complex structure is *not* the initial conformation of the current GA-mD-VcMD simulations. As explained later, we generated conformations in which bosentan was separated completely from the binding pocket of hETB. They were the conformations that we used as the initial simulation conformations of GA-mD-VcMD.

The residue ordinal numbers used for this paper are based on two sources: the original PDB numbering (PDB ID 5xpr) and the GPCRdb generic numbering<sup>24,25</sup>. The GPCRdb generic residue numbers are presented in parentheses. When a residue is not numbered in the GPCRdb generic method, only the original PDB residue number is shown. For instance, residues of the N-terminal tail of hETB are not numbered in the GPCRdb generic method.

**Set of three RCs introduced to control system motions.** Before describing generation of the initial conformations of GA-mD-VcMD, the set of three RCs ( $\lambda^{(h)}$ ,  $h \in \{\alpha, \beta, \gamma\}$ ) introduced for the current study must be explained concretely. The setup of RCs is fundamentally important to increase the sampling efficiency, although RCs can be set arbitrarily in theory. We imposed three conditions on the RCs: the RCs can control (1) the gate opening and closing of the hETB binding pocket, (2) the ligand approaching to and departing from the receptor, and (3) the ligand orientation relative to the receptor.

Although various RCs can satisfy the three conditions presented above, we set them in a straightforward manner as shown in Fig. 1b,c. The exact specifications for the atom groups are given in the figure caption. We briefly explain the roles of the RCs. For condition (1), two atom groups  $G_\alpha^A$  and  $G_\alpha^B$  (Fig. 1b) are introduced to define the first RC  $\lambda^{(\alpha)}$ , where the increase and decrease of  $\lambda^{(\alpha)}$  correspond to the opening and closing of the gate of the hETB's binding pocket, respectively. We note that disulfide bonds are formed at the roots of the N-terminal tail,  $G_\alpha^A$ , and the  $\beta$ -hairpin,  $G_\alpha^B$ , as shown in Fig. 1b,c. This means that the roots of the N-terminal tail and  $\beta$ -hairpin are locked by the disulfide bonds and that other regions of hETB are only weakly influenced by the gate motions. Figure 1c shows atom groups  $G_\beta^A$  and  $G_\beta^B$  to define  $\lambda^{(\beta)}$ , as well as  $G_\gamma^A$  and  $G_\gamma^B$  to define  $\lambda^{(\gamma)}$ . Groups  $G_\beta^A$  and  $G_\beta^B$  are mutually close in the native complex;  $G_\gamma^A$  and  $G_\gamma^B$  are as well. Likewise, those atom groups are distant from each other in an unbound conformation. The increase and decrease of  $\lambda^{(\beta)}$  and  $\lambda^{(\gamma)}$  result in approaching and departing of bosentan to the hETB's binding pocket: Condition (2) is established. Finally, we note that bosentan rotates with respect to hETB when  $\lambda^{(\beta)}$  decreases concomitantly with increasing  $\lambda^{(\gamma)}$ : Condition (3) is established. Figure 1a portrays a closeup image of bosentan to clarify  $G_\beta^A$  and  $G_\gamma^B$ .

These atom groups are set as independent of the chemical properties of the ligand and receptor. The atom groups are set conveniently to establish the three conditions (1), (2), and (3) above. As described later, this set of RCs produced various gate-widths of the hETB's binding pocket as well as various bosentan conformations relative to hETB.

One might infer that enhancement of the conformational motions in the 3D RC space affect the resultant conformational distribution and snapshots because a bias is introduced for the enhancement<sup>13</sup>. This possible effect of bias is true for conventional MD simulation. In GA-mD-VcMD, however, the bias effect is removed entirely from the resultant ensemble using a reweighting technique<sup>13</sup>.

**Initial conformations for GA-guided mD-VcMD.** We explain generation of the initial conformations of GA-mD-VcMD. Our aim is to explore the conformational space from unbound to bound states. Consequently, the initial conformations should not be the native complex structure but should be a completely unbound conformation. Because we had only the native complex structure (Figure S4) at this stage, it was necessary to generate the unbound conformations from the complex structure. For this purpose, we performed ten simulations (called “separation simulations”) starting from the native complex structure with imposition of three requirements to RCs to the system’s conformation:  $24.0 \text{ \AA} \leq \lambda^{(\alpha)} \leq 25.0 \text{ \AA}$ ,  $60.0 \text{ \AA} \leq \lambda^{(\beta)} \leq 65.0 \text{ \AA}$ , and  $60.0 \text{ \AA} \leq \lambda^{(\gamma)} \leq 65.0 \text{ \AA}$ . Ten separation simulations were run at 300 K for  $5 \times 10^6$  steps, with a time step of 2 fs. The three RCs for the native complex structure are  $\lambda^{(\alpha)} = 8.22 \text{ \AA}$ ,  $\lambda^{(\beta)} = 8.35 \text{ \AA}$ , and  $\lambda^{(\gamma)} = 6.21 \text{ \AA}$ . During the separation simulations, bosentan moved from the binding pocket to unbound conformations. The resultant conformations are presented in Figure S5, where bosentans are separated completely from hETB, which are presented as “Unbound conformations in figure S5” in Figure S2.

Next, for additional relaxation of those unbound conformations, we performed 2000 runs: 200 ( $1 \times 10^6$  steps with a time step of 2 fs at 300 K) runs starting from each of those ten unbound conformations with different random seeds to the initial atomic velocities. For the 2000 runs, we applied the following three restrictions to RCs:  $\lambda^{(\alpha)} \leq 25.0 \text{ \AA}$ ,  $\lambda^{(\beta)} \leq 65.0 \text{ \AA}$ , and  $\lambda^{(\gamma)} \leq 65.0 \text{ \AA}$ . The upper limits were introduced to prevent bosentan from flying further away from hETB in the periodic box. We refer to those 2000 runs as “diffusion simulations”. The last snapshots from the 2000 runs were then used as the initial conformations of the GA-mD-VcMD simulations. Figure S6 shows the positions of the last snapshots of the simulations in the 3D-RC space and the native complex structure in the 3D-RC space. Those conformations are represented as “Distributed conformations in figure S6” in figure S2.

It is noteworthy that Figure S6 demonstrates that the conformations sampled in the diffusion simulations did not reach the native complex structure. We emphasize that none of the current results was influenced by the structural features of the native complex because the 2000 initial conformations are those diffused freely from the ten completely unbound conformations (Figure S5).

**GA-mD-VcMD simulations.** As described in the “Introduction” section, GA-mD-VcMD enhances conformational sampling in the 3D-RC space. The thermodynamic weights of sampled conformations at the simulation temperature (300 K) are computed using the converged distribution  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$ .

In the original mD-VcMD<sup>12</sup> and GA-mD-VcMD<sup>13</sup>, each RC axis  $\lambda^{(h)}$  ( $h \in \{\alpha, \beta, \gamma\}$ ) is divided into small 1D zones (originally designated as “virtual zones”<sup>13</sup>). Accordingly, the 3D-RC space is divided into small 3D zones. Table S2 lists the actual division of each 1D axis into zones. During a short time-interval  $\Delta\tau_{trns}$  of simulation, the system’s conformation freely moves only in a single 3D zone with a restoring force outside the zone. Then, an inter-zone transition from the zone to one of neighboring zones is attempted at every time interval. For the present study, we set  $\Delta\tau_{trns} = 20$  ps, which corresponds to  $1 \times 10^4$  simulation steps, with a time step of 2 fs. The inter-zone transition probabilities are determined using  $Q_{cano}$ <sup>12,13</sup>. In this way, the conformation moves in a wide 3D-RC space by inter-zone transitions.

We performed  $N_{run}$  ( $=2000$ ) runs at 300 K in parallel starting from different conformations in each iteration to increase the sampling efficiency further. A simple integration of the  $N_{run}$  trajectories can be regarded as a single, long trajectory by connecting the trajectories in arbitrary order<sup>26,27</sup>. The  $N_{run}$  initial conformations for the first iteration of GA-mD-VcMD are the last conformations from the diffusion simulations (Figure S6). We obtained snapshots from the first to  $M$ -th iterations when the  $M$ -th iteration had finished. The  $N_{run}$  initial conformations for the  $(M + 1)$ -th iteration were selected from those snapshots so that they distribute as evenly as possible in the 3D-RC space. However, when a certain RC region was poorly sampled, we selected more snapshots from or near the poorly sampled RC region than from well-sampled RC regions<sup>13</sup>.

We set the simulation length of each run to  $1 \times 10^6$  steps ( $1 \times 10^6 \times 2 \text{ fs} = 2.0 \text{ ns}$ ). Consequently, the total length for an iteration was  $4.0 \mu\text{s}$  ( $=2.0 \text{ ns} \times 2000$ ). We performed 55 iterations for the present study, which corresponds to  $220.0 \mu\text{s}$  ( $4.0 \mu\text{s} \times 55$ ) in all. A snapshot was stored every  $1 \times 10^5$  steps (0.2 ns) in each run. Therefore,  $1.1 \times 10^6$  ( $220.0 \mu\text{s}/0.2 \text{ ns}$ ) snapshots were stored in all. We calculated the distribution functions of various quantities at the simulation temperature (300 K) from the resultant ensemble of snapshots.

The GA-mD-VcMD algorithm<sup>13</sup>, which was initially available in a MD simulation program *omegagene/myPresto*<sup>28</sup>, has been implemented for the current study in GROMACS<sup>29</sup> using the PLUMED plug-in<sup>30</sup>. The simulation conditions were the following: LINCS<sup>31</sup> to constrain the covalent-bond lengths related to hydrogen atoms, the Nosé–Hoover thermostat<sup>32,33</sup> to control the simulation temperature, the zero-dipole summation method<sup>34–36</sup> to compute the long-range electrostatic interactions accurately and quickly, a time-step of 2 fs ( $\Delta t = 2 \text{ fs}$ ), and simulation temperature of 300 K. All simulations were performed on the TSUBAME3.0 supercomputer at the Tokyo Institute of Technology using GP-GPU.

The force fields used are AMBER99SB-ILDN<sup>37</sup> for hETB and AMBER LIPID14<sup>38</sup> for the cholesterol and POPC lipid molecules, with TIP3P<sup>39</sup> for water molecules and Joung–Cheatham model<sup>40</sup> for ions (chloride and sodium), respectively. The bosentan force field was prepared as follows: First, the bosentan atomic partial charges were calculated quantum-chemically using Gaussian09<sup>41</sup> at the HF/6-31G\* level, followed by RESP fitting<sup>42</sup>. Then, the obtained partial charges were incorporated into a general amber force field (GAFF) force-field file<sup>43</sup>, which

was designed to be compatible with conventional AMBER force-fields. The chemical structure of bosentan is shown in Fig. 1A.

**Spatial density of bosentan around hETB.** GA-mD-VcMD simulation produces a canonical ensemble of conformations (snapshots) and assigns a thermodynamic weight (300 K) to each snapshot. Therefore, using the ensemble, one can calculate a spatial density of the bosentan centroid around hETB at 300 K using the ensemble. Details of methods used to calculate the spatial density are presented in Subsection 3.1 of SI.

We briefly explain the notation used to express the spatial density. The 3D real space (not the 3D-RC space) is divided into cubes ( $3 \text{ \AA} \times 3 \text{ \AA} \times 3 \text{ \AA}$ ), with centers denoted as  $\mathbf{r}_{cube}$ . The spatial density  $\rho_{CMB}(\mathbf{r}_{cube})$  is the probability of detecting the bosentan centroid  $\mathbf{r}_{CMB}$  in the cube centered at  $\mathbf{r}_{cube}$ .

## Results and discussion

We repeated 55 iterations of GA-mD-VcMD while saving  $1.1 \times 10^6$  snapshots from the first to 55th iterations in all, as described in Methods. First, we show that the GA-mD-VcMD produced a distribution of the system that covered both the unbound and native-like complex conformations. We next show that the highest-density spot (the lowest free-energy spot) corresponds to the native-like complex structure. We then analyze the binding process from the unbound to intermediates, and from the intermediates to the native-like complex. For this purpose, the binding process is divided into three stages: the early stage of binding, where bosentan is outside the binding pocket; the second stage, where bosentan is passing the gate of the binding pocket; and the last stage, where bosentan is moving in the binding pocket.

**Conformational distribution in 3D-RC space.** GA-mD-VcMD produced the distribution  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  equilibrated at 300 K in the 3D-RC space (Figure S8). Importantly, the distribution covered both the unbound and bound conformations. The native-complex structure was located at the periphery of the lowest free-energy basin. Section 4 of SI explains the features of this distribution.

The convergence check of the distribution is argued in Section 5 of SI with a function,  $E_{local}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$ . This function was introduced originally in an earlier report of a study<sup>13</sup>. The value of this function is calculated at zones  $(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  in each iteration. Simply, when an inequality  $E_{local}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)}) < 0.25$  is satisfied at a zone, it is judged that  $Q_{cano}$  is determined accurately at the zone. The validity of this threshold value has been checked repeatedly<sup>14–16</sup>.

Figure S9 shows that iterations 1–10 (Figures S9a–c) sampled a large volume of the 3D-RC space involving the unbound and bound conformations. Therefore, the simulation could be halted at the tenth iteration. However, to obtain snapshots sufficient for analysis, we continued the simulation up to the 55th iteration. A snapshot of every 0.2 ns (100,000 steps) of the simulation was stored.

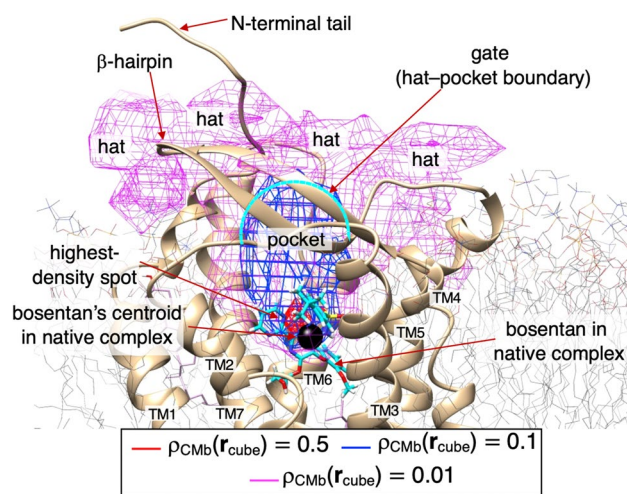
Next, we compared the distribution  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  with that obtained from an earlier study of the ET1–hETB system<sup>12</sup>. Section 6 of SI presents details of the comparative method. Figure S10 shows that the free-energy slope along the  $\lambda^{(\alpha)}$  axis (the motion along the gate opening and closing of the hETB's binding pocket) is gentler than those along the  $\lambda^{(\beta)}$  or  $\lambda^{(\gamma)}$  axis (bosentan's approaching and departing motions to hETB). This result agrees with that obtained from the earlier study. However, the slope in the current bosentan–hETB landscape was about one-tenth of that in the previous ET1–hETB landscape.

From the earlier study<sup>12</sup>, we inferred that the steep slope is probably caused by a shortcoming of the update method of the probability distribution  $Q_{cano}$  in the 3D RC space (i.e.,  $Q_{cano}^{(M)} \rightarrow Q_{cano}^{(M+1)}$ ) and that the steep slope would be weakened if two pieces of  $Q_{cano}^{(M)}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  and  $Q_{cano}^{(M)}(\lambda^{(\alpha)'}, \lambda^{(\beta)'}, \lambda^{(\gamma)'})$  were connected adequately, where  $(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  and  $(\lambda^{(\alpha)'}, \lambda^{(\beta)'}, \lambda^{(\gamma)'})$  are mutually neighboring zones in the RC space. GA-mD-VcMD provides such a method to connect  $Q_{cano}^{(M)}$  smoothly<sup>13</sup>. In fact, the free-energy slope decreased in the present study significantly.

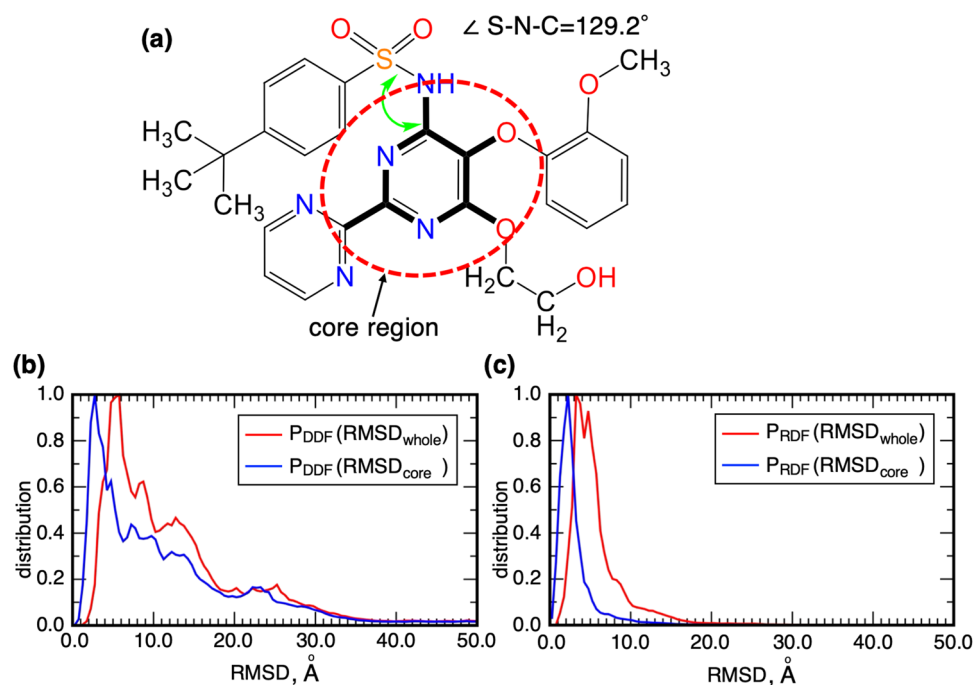
**Three-dimensional (3D) spatial distribution of bosentan around hETB.** Whereas GA-mD-VcMD samples the mD-RC space, the conformational distribution in the mD-RC space does not always provide a concrete or intuitive image for the molecular conformation of the system. Therefore, we converted the distribution  $Q_{cano}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$  (Figure S8) to the conformational distribution  $\rho_{CMB}(\mathbf{r}_{cube})$  defined in the 3D real space using a thermodynamic weight assigned to all the sampled snapshots using Eq. 31 of that earlier study<sup>13</sup>. Subsection 3.1 of SI presents the method to calculate  $\rho_{CMB}(\mathbf{r}_{cube})$ . Figure 2 portrays the spatial patterns of  $\rho_{CMB}(\mathbf{r}_{cube})$  of the bosentan centroid  $\mathbf{r}_{CMB}$  around hETB. Importantly, the highest-density spot (the red contours;  $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.5$ ) well agreed with the position of bosentan centroid in the native complex structure. This result is important because the reliability of the simulation data decreases unless the highest-density spot is assigned to the native bosentan position.

The binding pocket corresponds to the region of  $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.1$  (blue contours), which involved the highest-density spot. Outside of the binding pocket was characterized by a lower density of  $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.01$  (magenta-colored ones), which was named a “density hat” or simply “hat”. We denote the hat–pocket boundary (cyan-colored broken line in Fig. 2) as the “gate” of the binding pocket. We emphasize that the gate is not defined by residues of hETB but rather assigned to the boundary between the low- and high-density regions of  $\rho_{CMB}(\mathbf{r}_{cube})$ . In other words, the gate is a quantity that is dependent on the density field.

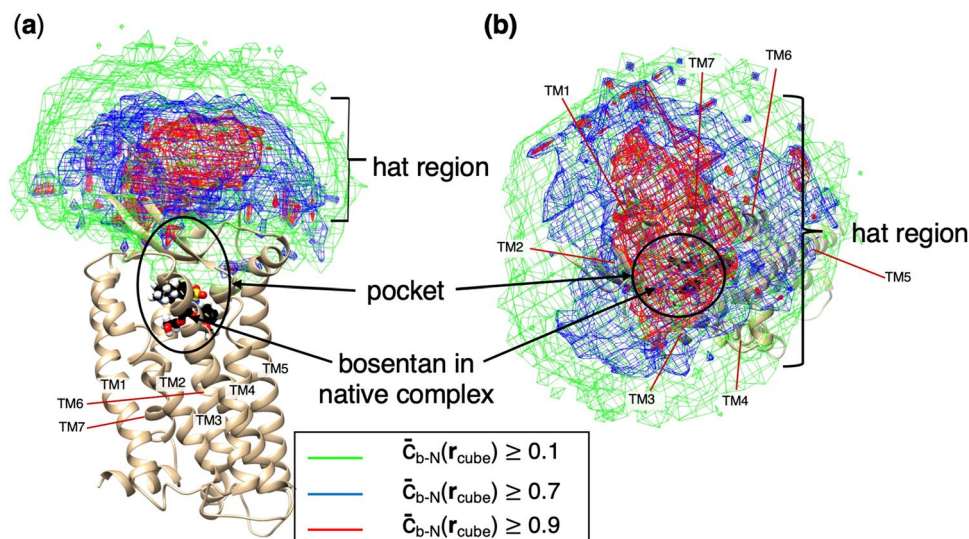
Figure S11 depicts the spatial density for the whole sampled region. Bosentan was distributed even in regions that are distant from hETB, as evidenced by the low density:  $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.0001$ . The bosentan density



**Figure 2.** The spatial density  $\rho_{CMb}(r_{cube})$  of the bosentan centroid around hETB. The iso-density map is presented at three contour levels with different colors, as shown in the inset. The molecular structure is the native complex (Figure S4) omitting solvent. Bosentan is shown as a cyan-colored stick model, with its centroid shown as a small black sphere. N-terminal tail and  $\beta$ -hairpin of hETB are indicated with labels. Red-colored contours are named “highest-density spots”. Blue-colored ones ( $\rho_{CMb}(r_{cube}) \geq 0.1$ ) well specify the binding pocket of hETB. Magenta-colored regions above the pocket are designated as “hat”. Cyan-colored broken-lines represent the boundary, the hat–pocket boundary, which is called “gate” of the binding pocket. Figure S11, presents  $\rho_{CMb}(r_{cube})$  with addition of two low-density contours of  $\rho_{CMb}(r_{cube}) \geq 0.001$  and  $0.0001$ . Transmembrane helices (TM1–7) are labeled.



**Figure 3.** (a) Bosentan’s “core region”, which is surrounded by a red broken line. The core region located at the center of bosentan can be regarded approximately as a rigid body. Therefore, the core region is a representative part to express the bosentan position and orientation. (b) Distance distribution function  $P_{DDF}(RMSD)$  and (c) radial distribution function  $P_{RDF}(RMSD)$  are presented for both  $RMSD_{whole}$  and  $RMSD_{core}$ . Definitions of  $RMSD$ ,  $P_{DDF}$ , and  $P_{RDF}$  are given in Subsection 3.2 of SI.



**Figure 4.** (a) Spatial patterns of cube-based bosentan–tail contact ratio  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  at three contour levels presented in the inset. The “hat” region, “pocket”, and “bosentan in native complex” are also shown. (b) Spatial patterns viewed from above the pocket. The structure which is presented is the native complex. Transmembrane helices (TM1–7) are labeled.

increased rapidly with approach to the binding pocket from the far region. This figure demonstrates that GAmD-VcMD is powerful for sampling the wide conformational space efficiently.

**RMSD of bosentan between the snapshot and the native complex.** As presented above, we showed visual agreement of the computed highest-density spot and the bosentan native-complex position. Here, to analyze the agreement quantitatively, we calculated the root-mean-square-deviations  $RMSD$  of bosentan of two types between a snapshot and the native complex:  $RMSD = RMSD_{whole}$  or  $RMSD = RMSD_{core}$ .  $RMSD_{whole}$  is calculated using all the heavy atoms in bosentan. Also,  $RMSD_{core}$  is calculated using the heavy atoms in the bosentan core region. The definition of the core region is presented in Fig. 3a.

The ten heavy atoms comprising the core region are located on a plane and are connected firmly by covalent bonds. Consequently, this region can be regarded as a rigid body approximately. Furthermore, this region is located at the bosentan center surrounded by four sidechains. Therefore, the core region is a representative part for expressing the bosentan position and orientation.

We calculated two probability distribution functions to analyze  $RMSD$ : the distance distribution function  $P_{DDF}(RMSD)$  and the radial distribution function  $P_{RDF}(RMSD)$ . Subsection 3.2 of SI presents details of  $P_{DDF}$  and  $P_{RDF}$ . The resultant distributions are presented in Fig. 3. For the whole bosentan, the highest-peak positions of  $P_{RDF}$  and  $P_{DDF}$  are, respectively, at  $RMSD_{whole} \approx 3.25 \text{ \AA}$  and  $RMSD_{whole} \approx 5.75 \text{ \AA}$ . Similarly, for the core region, those are respectively at  $RMSD_{core} \approx 2.25 \text{ \AA}$  and  $RMSD_{core} \approx 2.75 \text{ \AA}$ . The peak positions for the core region were smaller than those for the whole bosentan because of the disorder of the longest sidechain of bosentan, as explained later.

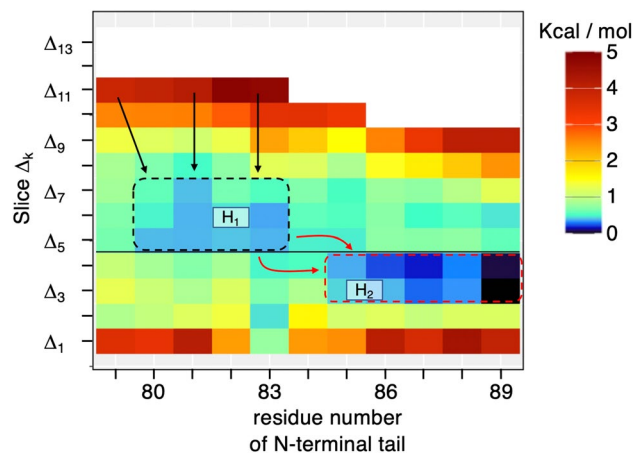
**Bosentan–tail contact ratio.** The structure of the N-terminal tail is not determined in the X-ray structures (PDB IDs: 5x93 and 5xpr) because the tail is intrinsically disordered. We presume that the tail might affect the binding process because the root of the tail is located at the entrance of the hETB binding pocket. Results of some experimentation studies have indicated that an extracellular loop of GPCRs plays a role as a lid to block the gate of the binding pocket<sup>44,45</sup>.

Figure S7b highlights the gate of the hETB binding pocket. As this figure shows, half of the gate, which is covered by the white solid-line tail, is closed by the contact of the N-terminal tail and the  $\beta$ -hairpin. The other half of the gate, which is covered by the white broken-line tail, is open. The definition of the N-terminal tail and the  $\beta$ -hairpin is given in the caption of Figure S7b. One can readily imagine from this figure that the gate situation can vary during tail fluctuations. Below, we investigate the tail fluctuations and the bosentan–tail contact.

First, to visualize fluctuations of the N-terminal tail, we calculated a spatial density function  $\rho_{CMNt}(\mathbf{r}_{cube})$  for the N-terminal tail tip. We also calculated the spatial density function  $\rho_{CM\beta h}(\mathbf{r}_{cube})$  for the tip of the  $\beta$ -hairpin because the  $\beta$ -hairpin is the contact partner of the N-terminal tail to open and close the gate of the binding pocket. Figure S12a exhibits that the N-terminal tail fluctuates widely with varying its orientation, which is a natural property of the intrinsically disordered segment. In contrast, the  $\beta$ -hairpin fluctuation is highly limited in a narrow volume, as shown in Figure S12b. Therefore, the gate opening and closing is induced by the large motions of the N-terminal tail (not by the fluctuation of  $\beta$ -hairpin).

Next, to investigate the bosentan–tail contact, we introduced a “cube-based bosentan–tail contact ratio” denoted as  $\bar{c}_{b-N}(\mathbf{r}_{cube})$ . It is noteworthy the bosentan centroid  $\mathbf{r}_{CMB}$  in a snapshot was assigned to a cube at  $\mathbf{r}_{cube}$ , which involved  $\mathbf{r}_{CMB}$ . Then  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  is the ratio by which bosentan in  $\mathbf{r}_{cube}$  contacts to the N-terminal tail.





**Figure 5.** Residue-based bosentan–tail contact ratio,  $\rho_{cnt}^{\Delta_k}(j)$ , converted to PMF:  $F^{\Delta_k}(j) = -RT \ln[\rho_{cnt}^{\Delta_k}(j)]$  (see Section 9 of SI), where  $j$  is the residue ordinal number (x-axis) of hETB’s N-terminal tail, and where  $\Delta_k$  is the  $k$ -th  $r_{bb}$  slice (y-axis), where the respective lower and upper boundaries are  $5(k-1)$  and  $5k$ . Unit is Å. PMF value is presented by color (see color bar). White is assigned to unsampled regions or regions for which the PMF is larger than 5 kcal/mol. The sequence of the N-terminal tail is Ace, Ser80, Pro81, Pro82, Arg83, Thr84, Ile85, Ser86, Pro87, Pro88, Pro89, where Ace is an acetyl group introduced to cap the N-terminal tail and where the residue number of 79 is assigned to Ace in the X-axis of the figure. Also,  $\Delta_k$  is the range of the  $r_{bb}$  slice given in Table S3. Region  $H_1$  ranges from Ser80 to Arg83 in x-axis and from  $\Delta_7$  to  $\Delta_5$  in y-axis.  $H_2$  does from Ile85 to Pro89 and from  $\Delta_4$  to  $\Delta_3$ . The meanings of arrows are described in the text. The gate of the binding pocket is defined by boundary (thin line) between  $\Delta_4$  and  $\Delta_5$  and the binding pocket is the region below the gate ( $\Delta_4$ ,  $\Delta_3$ ,  $\Delta_2$ , and  $\Delta_1$ ).

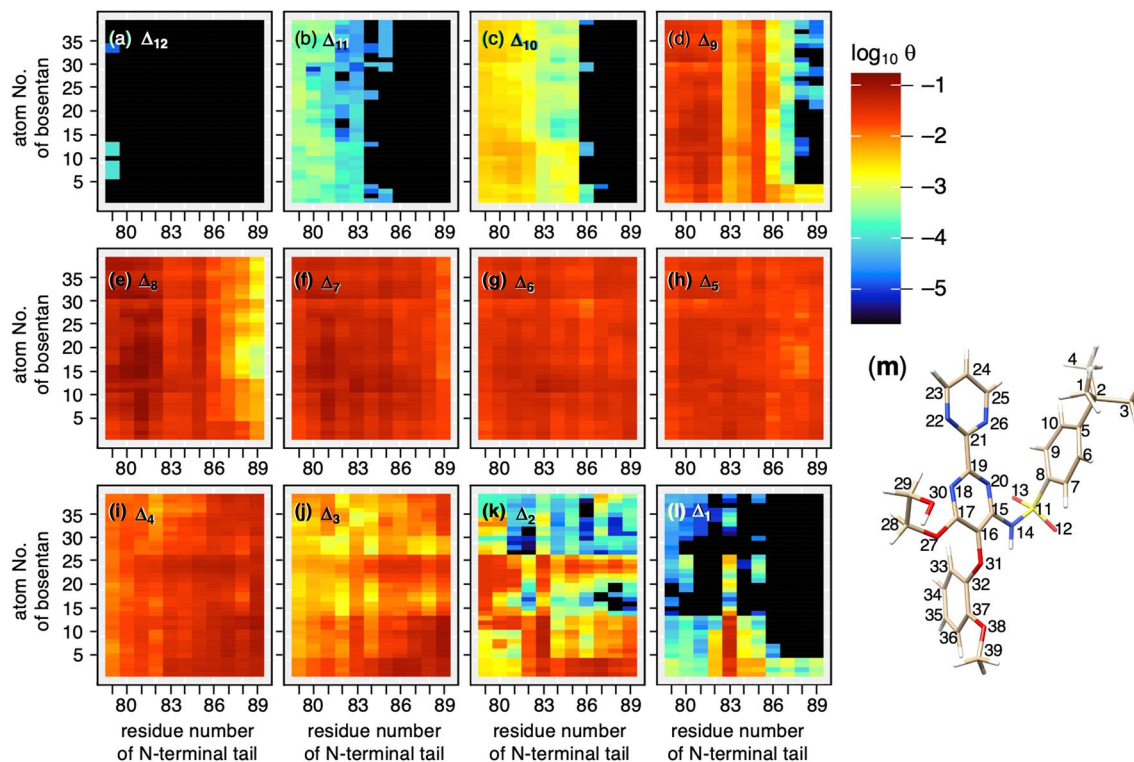
The maximum and minimum of  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  are, respectively, 1.0 and 0.0. Detailed procedures used to calculate  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  are presented in Subsection 3.3 of SI.

Figure 4a presents the spatial patterns of  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  at three contour levels:  $\bar{c}_{b-N}(\mathbf{r}_{cube}) \geq 0.1, 0.7$ , and  $0.9$ . The low-contact contours,  $\bar{c}_{b-N} \geq 0.1$ , extended to regions distant from the gate of the binding pocket, mean that the N-terminal tail contacted bosentan even when bosentan was floating in solution. With bosentan approaching the pocket, the contact ratio increased quickly and reached  $\bar{c}_{b-N} \geq 0.7$ . Because the region of  $\bar{c}_{b-N} \geq 0.7$  involved the density-hat region ( $\rho_{CMB} \geq 0.01$ ; magenta-colored contour regions of Fig. 2), the contact might act as an attractive interaction. The high-contact region of  $\bar{c}_{b-N} \geq 0.9$  (magenta-colored contours in Fig. 4a) was located immediately above the pocket, not in the hETB binding pocket. It is particularly interesting that  $\bar{c}_{b-N}(\mathbf{r}_{cube})$  was considerably small in the pocket. The high contact between the tail and bosentan right outside the pocket, followed by loss of contact once bosentan is in the pocket, indicates that the role of the N-terminal tail is the recruiting of bosentan into the pocket.

It is worth recalling that the gate-closing occurs because of the contact of the N-terminal tail and the  $\beta$ -hairpin (Figure S7b). In other words, gate-opening is induced by the N-terminal tail fluctuating in solution away from the gate. Figure 4 shows that the regions of  $\bar{c}_{b-N} \geq 0.7$  are widely extended far from the binding pocket. The gate-opening and closing, the large fluctuations of the N-terminal tail, and the bosentan–tail contact formation are mutually related.

**Bosentan slides along the N-terminal tail upon binding.** As shown in Fig. 4, the cube-based bosentan–tail contact ratio depends on the relative position of bosentan to hETB in the 3D real space. Here, we quantify the bosentan position relative to hETB by quantity  $r_{bb}$ , which is the distance between the bosentan centroid,  $\mathbf{r}_{CMB,i}$  in snapshot  $i$  and that in the native complex  $\mathbf{r}_{CMB}^{ref}$ :  $r_{bb} = |\mathbf{r}_{CMB,i} - \mathbf{r}_{CMB}^{ref}|$ . Figure S7c presents  $r_{bb}$  in a snapshot. Smaller  $r_{bb}$  is associated with closer bosentan to its natively binding position. Then, we divided the  $r_{bb}$  axis into 13 slices with thickness of 5 Å, which are designated respectively as  $\Delta_k$  ( $k = 1, \dots, 13$ ): The lower and upper boundaries of  $\Delta_k$  are given respectively as  $5(k-1)$  and  $5k$  (unit is Å). Table S3 presents the slice range exactly. As explained in Section 7 of SI and in Figure S13, the boundary between slices  $\Delta_4$  and  $\Delta_5$  coincides well with the gate of the binding pocket, which was introduced as the boundary between the blue contour region ( $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.1$ ) and the magenta-colored contour region ( $\rho_{CMB}(\mathbf{r}_{cube}) \geq 0.01$ ) in Fig. 2. By this arrangement, four slices  $\Delta_1, \dots, \Delta_4$  are in the binding pocket, and slices  $\Delta_5, \dots, \Delta_{13}$  are outside the pocket.

Figures S14–S17 present conformations selected from some slices. Section 8 of SI explains some features of those selected conformations. These figures visually imply that the contact site on the N-terminal tail to bosentan tends to move with  $r_{bb}$  motion. To verify this tendency quantitatively, we introduced a “residue-based bosentan–tail contact ratio”  $\rho_{cnt}^{\Delta_k}(j)$ , which is the contact ratio of residue  $j$  of the N-terminal tail to bosentan in a slice  $\Delta_k$ . Then, we converted  $\rho_{cnt}^{\Delta_k}(j)$  to a potential of mean force (PMF)  $F^{\Delta_k}(j)$  (equation S14 of Section 9 of SI). One



**Figure 6.** Atom–residue contact ratio  $\theta_{cnt}^{\Delta_k}(i, j)$ , which is the contact ratio of heavy atom  $i$  of bosentan to residue  $j$  of N-terminal tail in each slice  $\Delta_k$ . Its exact definition is given in equation S15 of SI. Panels (a–l) respectively show  $\theta_{cnt}^{\Delta_{12}}(i, j), \dots, \theta_{cnt}^{\Delta_1}(i, j)$ . (m) Atom ordinal numbers assigned to bosentan’s heavy atoms.

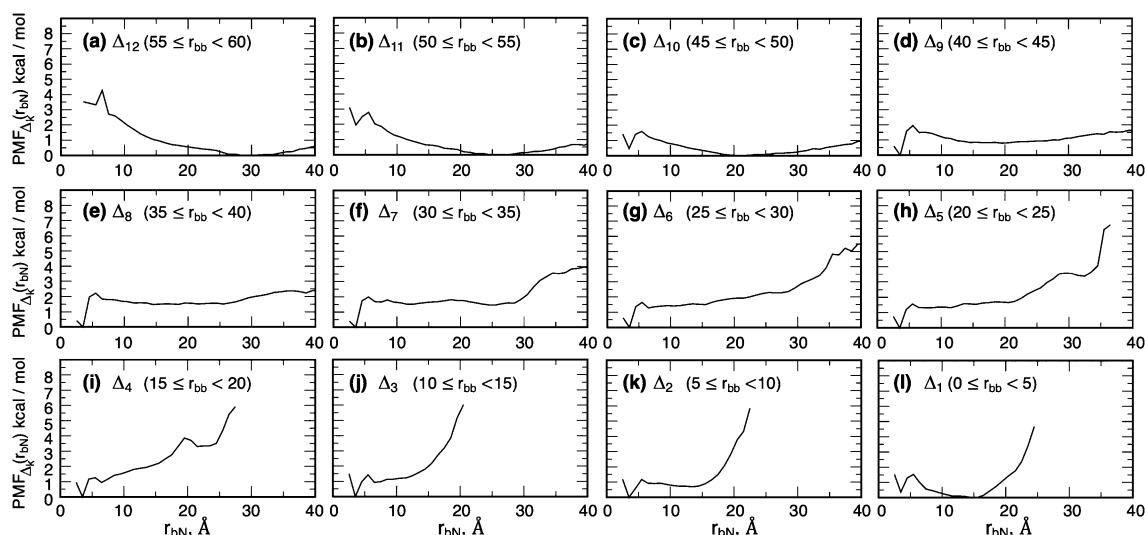
can imagine a 2D plane constructed by the residue ordinal number  $j$  and the slice  $\Delta_k$ . The lower the  $F^{\Delta_k}(j)$  at a site  $[j, \Delta_k]$  in the 2D plane, the stabler the contact becomes.

Figure 5 shows  $F^{\Delta_k}(j)$  in the 2D plane as the free-energy landscape. This figure indicates the existence of two free-energy basins, which are denoted as  $H_1$  and  $H_2$  in the figure. We discuss the binding process assuming that bosentan is approaching hETB from a far solvent region to the binding pocket. In  $\Delta_{13}$  ( $60 \text{ \AA} \leq r_{bb} < 65 \text{ \AA}$ ), the N-terminal tail does not touch bosentan:  $\rho_{cnt}^{\Delta_{13}}(j) = 0 (\forall j)$ . In  $\Delta_{12}$  ( $55 \text{ \AA} \leq r_{bb} < 60 \text{ \AA}$ ), bosentan first touches Ace (N-terminal capping group) of the N-terminal tail. Because the system moves generally toward a low free-energy region, possible transition paths are indicated by black arrows in Fig. 5, which lead the conformation to basin  $H_1$ . Those motions decrease  $r_{bb}$  and the bosentan–tail contact firm, thereby decreasing the free energy.

With further decrease of  $\Delta_k$ , the stable region moved from the basin  $H_1$  to  $H_2$ , i.e., the system moves from the tip region to the root region of the N-terminal tail. This move is natural because basin  $H_2$  is lower than basin  $H_1$ . One can recall that the boundary between  $\Delta_5$  and  $\Delta_4$  corresponds to the gate of binding pocket. Conformational changes from  $H_1$  to  $H_2$  can occur via multiple pathways indicated by the red arrows in Fig. 5. Because the sites to which the red arrows are assigned had a larger free energy than either  $H_1$  or  $H_2$ , the  $H_1$ – $H_2$  basin switch requires that the free-energy barrier be overcome.

In  $\Delta_2$  and  $\Delta_1$ , which are located at the bottom of the binding pocket, the bosentan–tail contact is disrupted. This disruption does not mean that the complex itself is dissociated but that the complex is stabilized without the bosentan–tail contact because the native-like complex conformations are stabler than the other state, as shown in Figs. 2 and 3. We also observed a specific contact at Arg83 in  $\Delta_2$  and  $\Delta_1$ . Nevertheless, this specific contact represents a weak interaction because the depth of free energy at Arg83 is shallow (Fig. 5).

The importance of the long N-terminal tail is demonstrated by comparing the current simulation with the simulations from GPCRmd<sup>46</sup>, which is a database of various MD simulations of GPCRs (<https://submission.gpcrmd.org/home/>). In the unbound-form MD simulation of hETB in GPCRmd (<https://submission.gpcrmd.org/view/886>), the N-terminal tail of hETB fluctuates greatly in solvent without forming a specific conformation. This result agrees with those obtained from our study. By contrast, in the bound-form simulation of hETB from GPCRmd (<https://submission.gpcrmd.org/view/887>), the N-terminal tail contacted frequently to a ligand, which bound in the deep binding pocket of hETB. Figure 5 shows that the bosentan–tail contact disappeared when bosentan moved to the bottom of the binding pocket in our simulation. This discrepancy of the tail conformation in the bound form between our simulation and the simulation from GPCRmd is caused simply by a difference in the length of the tail used for the two simulations. The N-terminal tail in the current simulation is six residues longer than that of the GPCRmd simulation. The longer N-terminal tail cannot be packed into the binding pocket of GPCR. We presume that our simulation is more realistic than the bound-form simulation of



**Figure 7.** Potential of mean force  $PMF_{\Delta_k}(r_{b-N})$  for the bosentan–tail distance  $r_{b-N}$  in slice  $\Delta_k$  ( $k = 1, \dots, 12$ ).  $PMF_{\Delta_k}(r_{b-N})$  is defined in the main text. Panels (a–l) respectively show  $PMF_{\Delta_{12}}(r_{b-N}), \dots, PMF_{\Delta_1}(r_{b-N})$ . Also,  $PMF_{\Delta_{13}}(r_{b-N})$  is omitted because  $PMF_{\Delta_{13}}(r_{b-N})$  is similar with  $PMF_{\Delta_{12}}(r_{b-N})$ .

GPCRmd because GPCRs have a long N-terminal tail in general. Therefore, this comparison of the two simulations demonstrates the importance of the length of the N-terminal tail.

**Bosentan–tail non-specific contacts.** Figure 5 does not show whether the bosentan–tail contact is specific or non-specific. To investigate the bosentan atom specificity, we calculated an “atom–residue contact ratio”  $\theta_{cnt}^{\Delta_k}(i, j)$  which is the contact ratio of a heavy atom  $i$  of bosentan contacting residue  $j$  of the N-terminal tail in each slice  $\Delta_k$ . The exact definition of  $\theta_{cnt}^{\Delta_k}(i, j)$  is presented in Section 9 of SI. Figure 6 shows  $\theta_{cnt}^{\Delta_k}(i, j)$  in each slice. Normalization is done for each panel of Fig. 6 to highlight the atom-specificity (color gradation) in each slice. Consequently, the existence of basins  $H_1$  or  $H_2$  is not recognized in Fig. 6.

Figure 6a shows that a few atoms of bosentan rarely contact Ace of the N-terminal tail in  $\Delta_{12}$ . Figures 6b–d depict that all the heavy atoms of bosentan contact non-specifically to the tip of the N-terminal tail in  $\Delta_{11} - \Delta_9$ . Figures 6e–h again show that bosentan contact is non-specific in  $\Delta_8 - \Delta_5$ , where all residues of the N-terminal tail contact to bosentan. Up to  $\Delta_{12} - \Delta_5$ , bosentan is outside of the binding pocket. The non-specific contact is still found in  $\Delta_4$  (Fig. 6i) and  $\Delta_3$  (Fig. 6j), although  $\Delta_4$  and  $\Delta_3$  are in the binding pocket.

In  $\Delta_2$  and  $\Delta_1$  (Fig. 6k,l, respectively), by contrast, Arg83 of the N-terminal tail contacts specifically to atoms 1–13 of bosentan, as shown in Fig. 6m. This specificity is consistent with the depiction in Fig. 5. Atoms 1–10 and 11–13 are associated respectively with the hydrophobic and hydrophilic groups. Figure S18 presents some conformations that involve the specific bosentan–tail contacts. Generally, the hydrophobic sidechain stem of Arg83 of the tail and the hydrophobic group (atoms 1–10) of bosentan share some contact, as do the nitrogen atoms of the sidechain tip of Arg83 and oxygen atoms of sulfonamide of bosentan.

As shown in Figs. 4 and 5, the bosentan–tail contact tends to be dissociated when bosentan proceeds to the bottom of the binding pocket ( $\Delta_1$  and  $\Delta_2$ ). Consequently, in many snapshots depicting  $\Delta_1$  and  $\Delta_2$ , the N-terminal tail does not contact bosentan. If the bosentan–tail contact has a role to capture bosentan in solution, then we consider that the contact in the binding pocket has no role for molecular binding.

**Fly casting with ligand sliding.** Here we check whether the bosentan–tail contact acts as an attractive interaction. First, we calculated a probability distribution function  $P_{\Delta_k}(r_{b-N})$  regarding the minimum heavy atomic distance  $r_{b-N}$  from bosentan to the N-terminal tail for snapshots, which are involved in a slice  $\Delta_k$ . Then,  $PMF$  regarding  $r_{b-N}$  is defined as  $PMF_{\Delta_k}(r_{b-N}) = -RT \ln[P_{\Delta_k}(r_{b-N})]$ . In fact, the effective interaction between two molecules is contributed by many elemental interactions. To estimate this interaction comprehensively,  $PMF$  is the best quantity. This quantity involves not only direct intermolecular interactions but also indirect interactions (ligand–solvent and solvent–receptor interactions). This quantity also involves effects of molecular motions (i.e., conformational variety, or entropy).

Figure 7 presents  $PMF$  for each slice. Apparently, the bottom of the lowest free-energy basin is assigned to a noncontact state in slices  $\Delta_{12} - \Delta_{10}$  ( $r_{bb} \geq 45$  Å) (Fig. 7a–c). Actually,  $PMF$  for  $\Delta_{13}$  is not shown because a bosentan–tail contact was not detected in  $\Delta_{13}$ . In contrast, the lowest free-energy basin is assigned to a contact state ( $r_{b-N} \approx 3.5$  Å) in  $\Delta_9 - \Delta_2$  (Fig. 7d–k), which indicates that the bosentan–tail contact acts effectively as an attractive interaction in the range of  $45$  Å  $>$   $r_{bb} \geq 5$  Å.

The analysis described above showed that a bosentan–tail attractive interaction exerted from  $\Delta_9$  ( $45$  Å =  $r_{bb}$ ) between the N-terminal tail and bosentan. Together with the residue sliding from the tip side to the root side with decreasing  $r_{bb}$ , we designate this binding mechanism “Fly Casting with ligand sliding”.

**Bosentan–membrane contact.** Bosentan occasionally contacts the membrane (Figures S14e, S15d and S15e). To investigate the bosentan–membrane contact, we introduced a quantity, a “bosentan–membrane contact ratio”  $\bar{c}_{b-m}(\mathbf{r}_{cube})$ , that is calculated with a similar procedure with  $\bar{c}_{b-N}(\mathbf{r}_{cube})$ . Subsection 3.3 of SI provides details of the calculation procedure. The bosentan–membrane contact was formed commonly on the membrane surface (Figure S19). Some GPCR ligands are known to diffuse from the lipid bilayer to reach the binding pocket of GPCRs rather than approaching the pocket from the solvent side<sup>47,48</sup>. From the present simulation, however, we were unable to find such a ligand-diffusive motion of bosentan in the membrane, whereas we observed that bosentan sank in the membrane (Figure S14e). One might consider that the absence of bosentan diffusive motions in membrane might be due to insufficient setting of the upper boundaries of  $\lambda^{(\beta)}$  and  $\lambda^{(\gamma)}$  (68.47 Å and 67.47 Å, respectively; Table S2). However, Figure S11 of SI shows that the upper boundaries were sufficiently large for bosentan to move in the membrane if no interaction is applied between bosentan and the membrane. Therefore, we conclude that bosentan mainly approaches the pocket from solvent. Because the membrane prevents bosentan from moving to the cytoplasmic side, the membrane has a role in restricting the phase volume for bosentan to move. This volume restriction might facilitate binding of bosentan to the receptor’s binding pocket<sup>49</sup>.

**Oriental selection when bosentan enters the binding pocket.** We investigated the molecular orientation of bosentan around hETB. First, we defined the bosentan molecular orientation by two vectors (orientation vectors)  $\mathbf{v}_{\downarrow}$  and  $\mathbf{v}_{\leftarrow}$ , which are fixed in the core region of bosentan (Figure S20a), which are approximately perpendicular to one another. Figure 3a gives the exact definition of the core region. Then, we normalized the orientation vectors as  $|\mathbf{v}_{\alpha}| = 1$  ( $\alpha = \leftarrow$  or  $\downarrow$ ). Because the bosentan core region is structurally stiff, the two orientation vectors are convenient quantities for defining the bosentan molecular orientation. In fact, in general, two nonparallel vectors can specify the orientation of a plane (the central ring of the core region) in 3D real space. Figure S20b presents an example of the vectors in the native complex and in a snapshot. Next, we assigned the orientation vectors to a cube  $\mathbf{r}_{cube}$ , in which the bosentan centroid was detected. Then,  $\Omega$  in equation S7 of Subsection 3.4 of SI is replaced by  $\mathbf{v}_{\alpha}$ . The resultant quantity  $\bar{\Omega}(\mathbf{r}_{cube})$  is designated as  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$  in the equation. We designated  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$  as an “averaged orientation vector”. Because  $\mathbf{v}_{\alpha}$  in a snapshot was normalized, the maximum of  $|\bar{\mathbf{v}}_{\alpha}|$  is unity when all the vectors detected in the cube  $\mathbf{r}_{cube}$  are perfectly ordered. The minimum is 0.0 when the vectors are randomized completely.

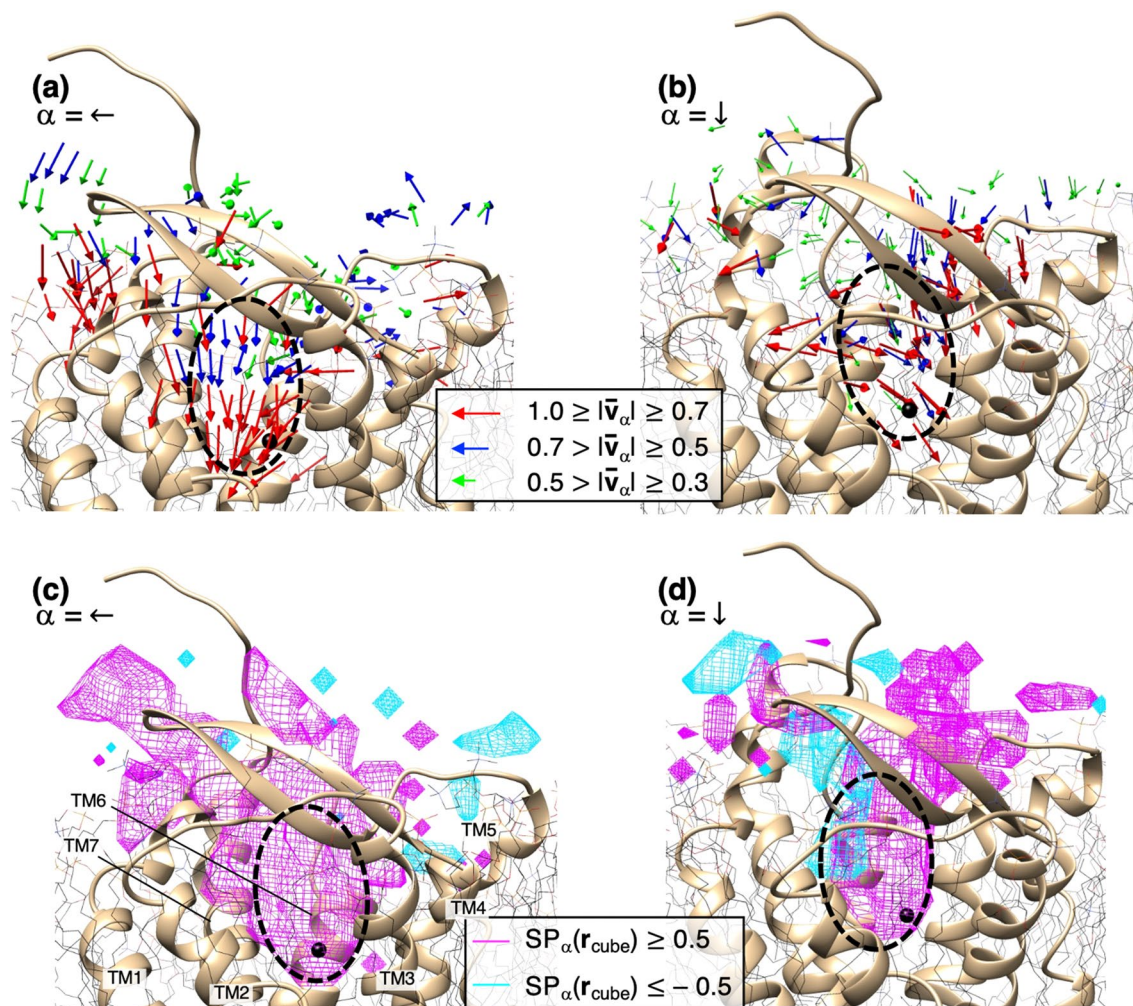
Figures 8a,b portray spatial patterns of  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$ , in which orientation vectors with large norms ( $|\bar{\mathbf{v}}_{\alpha}| > 0.3$ ) are shown. Importantly, the large-norm vectors were found mainly in the binding pocket. Therefore, the bosentan orientation is ordered in the pocket, although it is disordered outside the pocket. This finding suggests that configurational entropy of bosentan decreases quickly when bosentan passes the gate of the binding pocket. It is worth recalling that the free-energy basin of  $F^{\Delta k}(j)$  also switched quickly from  $H_1$  to  $H_2$  at the gate (Fig. 5). In fact, the decrease of entropy results in increased free energy. This entropic decrease should be compensated by another thermodynamic factor. Otherwise, bosentan in the binding pocket is destabilized, and the density  $\rho_{CMB}(\mathbf{r}_{cube})$  in the binding pocket decreases. We show in the next section that intermolecular native contacts, which act as attractive interactions between bosentan and hETB, are formed in the binding pocket. Consequently, those intermolecular attractive interactions (enthalpy decrease) compensate the entropy loss.

However, Fig. 8a,b do not necessarily mean that the bosentan orientations resemble those in the native-complex position. Then, we check the orientation similarity as follows: First, we calculated  $\mathbf{v}_{\leftarrow}$  and  $\mathbf{v}_{\downarrow}$  for the native complex, which are designated respectively as  $\mathbf{v}_{\leftarrow}^{(ref)}$  or  $\mathbf{v}_{\downarrow}^{(ref)}$ . Next, we calculated a scalar product as  $SP_{\alpha}(\mathbf{r}_{cube}) = \mathbf{e}_{\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})} \cdot \mathbf{v}_{\alpha}^{(ref)}$ , where  $\mathbf{e}_{\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})}$  is a unit vector that is parallel to  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$ . If  $\mathbf{e}_{\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})}$  is exactly parallel to  $\mathbf{v}_{\alpha}^{(ref)}$ , then  $SP_{\alpha}(\mathbf{r}_{cube}) = 1$ ; if  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$  is exactly anti-parallel to  $\mathbf{v}_{\alpha}^{(ref)}$ , then  $SP_{\alpha}(\mathbf{r}_{cube}) = -1$ . If  $\mathbf{e}_{\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})}$  and  $\mathbf{v}_{\alpha}^{(ref)}$  are mutually perpendicular, then  $SP_{\alpha}(\mathbf{r}_{cube}) = 0$ . Therefore, the spatial pattern of  $SP_{\alpha}(\mathbf{r}_{cube})$  is a useful measure to assess similarity between  $\bar{\mathbf{v}}_{\alpha}(\mathbf{r}_{cube})$  and  $\mathbf{v}_{\alpha}^{(ref)}$ .

Figure 8c demonstrates that the major volume of the binding pocket was occupied by  $SP_{\leftarrow}(\mathbf{r}_{cube}) \geq 0.5$ , which indicates similar orientations of  $\bar{\mathbf{v}}_{\leftarrow}$  and  $\mathbf{v}_{\leftarrow}^{(ref)}$ . For  $\bar{\mathbf{v}}_{\downarrow}$ , a large volume of the pocket was also characterized by  $SP_{\downarrow}(\mathbf{r}_{cube}) \geq 0.5$ , whereas a minor part (cyan-colored contours) of the pocket was characterized by  $SP_{\downarrow}(\mathbf{r}_{cube}) \leq -0.5$  (Fig. 8d). This minor part does not cover the bosentan native position at all. Therefore, we can infer that bosentan enters the pocket with limited orientations that are beneficial for reaching the native-complex position because the pocket is too narrow for bulky bosentan to flip. We designate this possible mechanism as “orientational selection”.

**Bosentan–hETB contacts and ring-C packing at the bottom of binding pocket.** In the above section, we suggested that the entropy loss (decrease in the orientational variety of bosentan) should be compensated by enthalpy decrease for the stabilization of bosentan in the binding pocket. To investigate this stabilization effect, we analyze attractive interactions (atomic contacts) of bosentan with amino-acid residues of hETB in the binding pocket.

First, we observe common intermolecular interactions in the three X-ray structures: the bosentan–hETB<sup>10</sup>, ET1–hETB<sup>6</sup>, and the K8794–hETB<sup>10</sup> complexes. In fact, the same binding scheme can be observed in the three complex structures (Figure S21). By viewing the three complexes closely, we identified four intermolecular native atomic contacts (interatomic distance less than 5 Å in the X-ray structures), which exist commonly in the three complex structures as attractive interactions, as shown in Section 10 of SI, Figures S22a, S22b, and Table S4: contact (1) between oxygen atoms in the sulfonyl group (SOO) of the bosentan’s sulfonamide and the nitrogen ( $N_{\zeta}$ ) of Lys 182 (3 × 33); (2) between the oxygens in bosentan’s SOO and the nitrogen of Lys 273 (5 × 39); (3) between



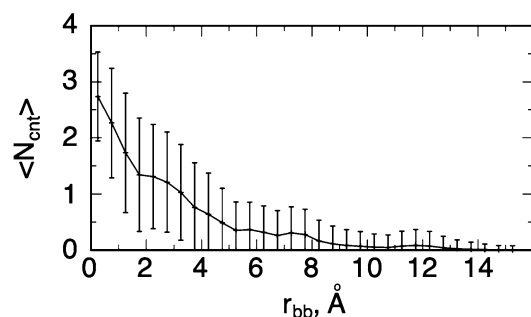
**Figure 8.** Spatial patterns of averaged orientation vectors (a)  $\bar{v}_{\leftarrow}(\mathbf{r}_{cube})$  and (b)  $\bar{v}_{\downarrow}(\mathbf{r}_{cube})$ . A vector is colored differently by its norm  $|\bar{v}_\alpha(\mathbf{r}_{cube})|$  ( $\alpha = \leftarrow$  or  $\alpha = \downarrow$ ) (inset). Vectors with  $|\bar{v}_\alpha(\mathbf{r}_{cube})| < 0.3$  are omitted. (c) Spatial patterns of scalar product  $SP_{\leftarrow}(\mathbf{r}_{cube}) (= \mathbf{e}_{\bar{v}_{\leftarrow}}(\mathbf{r}_{cube}) \cdot \mathbf{v}_{\leftarrow}^{(ref)})$  and (d) those of  $SP_{\downarrow}(\mathbf{r}_{cube}) (= \mathbf{e}_{\bar{v}_{\downarrow}}(\mathbf{r}_{cube}) \cdot \mathbf{v}_{\downarrow}^{(ref)})$ , where  $\mathbf{e}_{\bar{v}_\alpha}(\mathbf{r}_{cube})$  and  $\mathbf{v}_\alpha^{(ref)}$  are defined in main text. Regions of  $SP_\alpha(\mathbf{r}_{cube}) \geq 0.5$  and  $SP_\alpha(\mathbf{r}_{cube}) \leq -0.5$  are colored differently by their values (inset). For all panels, the hETB structure is from the native complex. Small black spheres are bosentan centroid in the native complex. The binding pocket is shown as a broken-line circle in all panels. Transmembrane helices (TM1–7) are labeled for panel c. Positions of the transmembrane helices for the other panels are presumable from panel c because hETB is shown from similar directions for all panels.

the oxygens in bosentan's SOO and nitrogen atoms of Arg 343 ( $6 \times 55$ ); and (4) between the nitrogen atom (N14) of the bosentan's sulfonamide and the nitrogen ( $N_\zeta$ ) of Lys 182 ( $3 \times 33$ ). The actual interatomic distances for the four contacts are listed in Table S4 for the three X-ray structures: Most of contacts are around 3 Å, with the sole exception of 4.56 Å for (2) in the bosentan–hETB complex. Figure S22c enumerates all intermolecular interactions in the bosentan–hETB complex.

Contacts (1)–(3) are associated with salt bridges. Contact (4) can be associated either with a salt bridge or with a hydrogen bond because we cannot infer whether a hydrogen atom is bound to N14 of the bosentan's sulfonamide in the X-ray crystallography, or not. Nevertheless, it is noteworthy that both the salt bridge and the hydrogen bond are attractive interactions. For the present study, we used the neutral form for bosentan, in which a hydrogen is attached to N14 (Fig. 1a). This ambiguity of protonation is discussed at the end of the “Results and Discussion” section.

The native atomic contacts introduced above are appropriate quantities to analyze the stabilization effect (enthalpic effect) that compensates the entropic disadvantage (decrease of molecular orientation). As more native contacts form in the complex, the attractive interactions become stronger while bosentan's structural entropy decreases.

To analyze the intermolecular distances of snapshots, we introduce the notation of  $r_{\alpha,\beta}$ , which is a distance between atoms  $\alpha$  and  $\beta$  belonging respectively to the ligand (bosentan, K8794, or ET1) and hETB. We use the original PDB residue numbering to specify atom  $\beta$ . The four distances related to the four native contacts are designated as  $r_{N14;Lys182N_\zeta}$ ,  $r_{SOO;Lys182N_\zeta}$ ,  $r_{SOO;Lys273N_\zeta}$ , and  $r_{SOO;Arg343N_\eta}$ . When multiple distances were



**Figure 9.** Dependence of  $N_{cnt}(r_{bb})$  on  $r_{bb}$  and its standard deviation (error bars). Lower limits of some error bars are smaller than 0 on the x-axis because the standard deviation was calculated using the usual form:  $[N_{cnt}(r_{bb})^2 - N_{cnt}(r_{bb})]^2$ .

possible in  $r_{\alpha;\beta}$ , the minimum of the distances was selected for  $r_{\alpha;\beta}$ . For instance, four distances are possible for  $r_{\text{SOO};\text{Arg}343\text{N}_\eta}$  because two oxygen atoms and two nitrogen atoms are involved respectively in the sulfonyl group and Arg 343 ( $6 \times 55$ ). Also, Table S4 presents distances  $r_{\alpha;\beta}$  in the X-ray structures, in which three distances are approximately 3 Å, except for  $r_{\text{SOO};\text{Lys}182\text{N}\zeta}$  ( $= 4.56$  Å) from the bosentan-hETB complex. We presume that  $r_{\text{SOO};\text{Lys}182\text{N}\zeta}$ ,  $r_{\text{SOO};\text{Lys}273\text{N}\zeta}$ , and  $r_{\text{SOO};\text{Arg}343\text{N}_\eta}$  are related to salt-bridges, and that  $r_{\text{N}14;\text{Lys}182\text{N}\zeta}$  might be done to either a salt-bridge or a hydrogen bond.

We judged that the heavy atoms  $\alpha$  and  $\beta$  are contacting if  $r_{\alpha;\beta} < 5.0$  Å is satisfied in a snapshot. The value of 5.0 Å is about 0.5 Å larger than the largest 4.56 Å of the experimentally obtained values (Table S4). Considering thermal fluctuations of the complex structure and the likelihood of error in experiment measurements (3.6 Å resolution for the bosentan-hETB complex), we added a tolerance figure of 0.5 Å to the inequality. Furthermore, not only does a heavy atom have a radius of approximately 2.0 Å, but the remaining distance between  $\alpha$  and  $\beta$  is 1.0 Å ( $= 5.0$  Å  $- 2 \times 2.0$  Å), which does not allow a water molecule to penetrate because the water molecule diameter is about 3.0 Å. Therefore, the 5.0 Å threshold is rational to check the atomic contacts.

Based on the criteria above, we calculated the number of contacts  $N_{cnt}$  ( $0 \leq N_{cnt} \leq 4$ ) regarding the four distances for all snapshots. Then, picking snapshots falling in an  $r_{bb}$ -range of  $R(r_{bb}) = [r_{bb} - \Delta r_{bb}; r_{bb} + \Delta r_{bb}]$ , we calculated the average of  $N_{cnt}$  for the selected snapshots and denoted it as  $N_{cnt}(r_{bb})$ . Section 11 of SI presents the exact method to calculate  $N_{cnt}(r_{bb})$  and its standard deviation  $SD_{cnt}(r_{bb})$ . Ideally,  $\Delta r_{bb}$  should be sufficiently small ( $\Delta r_{bb} \rightarrow 0$ ) to calculate  $N_{cnt}(r_{bb})$  exactly at  $r_{bb}$ . However, the number of snapshots in  $R(r_{bb})$  decreases concomitantly with decreasing  $\Delta r_{bb}$  because the simulation length is finite. To maintain statistical significance in  $N_{cnt}(r_{bb})$ , we set  $\Delta r_{bb} = 0.25$  Å for the present simulation. The value of  $N_{cnt}(r_{bb})$  is four if all snapshots in  $R(r_{bb})$  always have four contacts and zero if no snapshot has any contact.

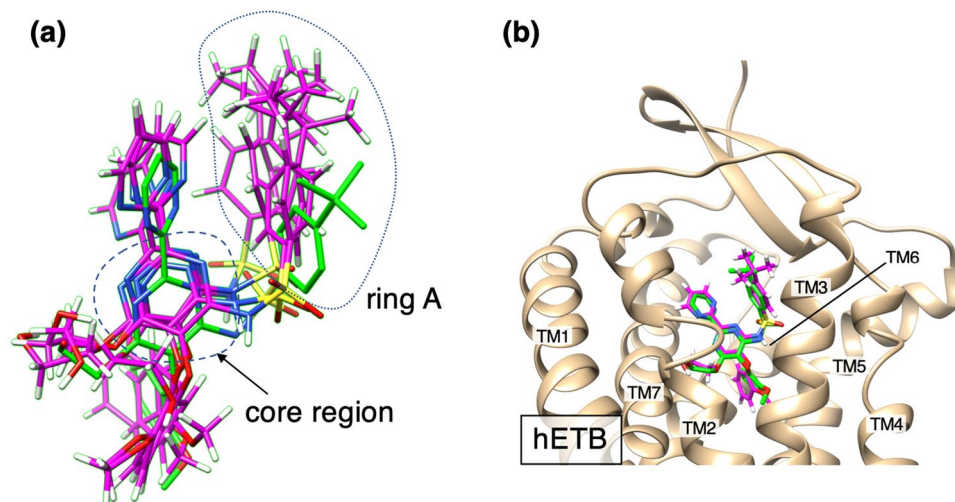
Figure 9 depicts the  $r_{bb}$  dependence of  $N_{cnt}(r_{bb})$ , where  $r_{bb}$  starts from 0.25 Å and where  $N_{cnt}(0.25$  Å) is calculated for the range of  $R(0.25$  Å) = [0.0 Å; 0.5 Å]. Note that  $R(r_{bb} < 0.25$  Å) is impossible to be calculated. We set the maximum of  $r_{bb}$  at 15.25 Å, where the used range is  $R(15.25$  Å) = [15.0 Å; 15.5 Å]. Also,  $R(15.25$  Å) is involved in the slice  $\Delta_4$ , immediately below the gate of the hETB's binding pocket (Figure S13). Therefore, Fig. 9 shows that almost no atom contact is formed at the gate:  $N_{cnt}(15.25$  Å) = 0.005. Also,  $N_{cnt}(r_{bb})$  increased concomitantly with decreasing  $r_{bb}$  (with bosentan approaching the native-complex position): The number of contacts exceeded the value of 1 at  $r_{bb} = 3.25$  Å:  $N_{cnt}(3.25$  Å) =  $1.03 \pm 0.85$ . Subsequently, the growth rate of  $N_{cnt}(r_{bb})$  rose, and  $N_{cnt}(r_{bb})$  reached the maximum at  $r_{bb} = 0.25$  Å:  $N_{cnt}(0.25$  Å) =  $2.74 \pm 0.79$ . We further investigated formation of each contact, as described in Section 12 of SI. Figure S23 shows that each native contact is formed with decreasing  $r_{bb}$ .

Because the four intermolecular native contacts act as attractive interactions (Section 10 of SI), Fig. 9 shows that the complex structure is stabilized enthalpically by the native contacts when bosentan goes to the bottom of the binding pocket. We presume that the instability induced by the decrease of the entropy when bosentan passing the gate of the binding pocket is compensated by formation of the native contacts.

We checked the reproducibility of the packing of ring C to the bottom of the binding pocket (see Figure S21c for ring C of bosentan). For this purpose, we calculated the density map  $\rho_{CMC}(\mathbf{r}_{cube})$  of the centroid of ring C using equation S3 of Subsection 3.1 of SI with replacement of the centroid of portion X by the centroid of ring C. Figure S24 presents the high-density spot ( $\rho_{CMC}(\mathbf{r}_{cube}) \geq 0.8$ ) for ring C. Importantly, the high-density spot is localized in a region that overlaps with the ring-C position in the native complex. However, the contours of  $\rho_{CMC}(\mathbf{r}_{cube}) \geq 0.4$  show deviation from the native-complex position. We discuss this discrepancy between the computation and the native complex later.

**Bosentan conformations in the highest-density spot.** As shown in Fig. 2, the bosentan centroid in the native complex structure was located on the fringe of the computed highest-density spot (red contours in Fig. 2). This section presents an explanation of bosentan conformations in the highest-density spot.

As stated earlier, the highest-peak position of  $P_{RDF}(RMSD_{whole}^{heavy})$  or  $P_{DDF}(RMSD_{whole}^{heavy})$  of Fig. 3 is larger than that of  $P_{RDF}(RMSD_{core}^{heavy})$  or  $P_{DDF}(RMSD_{core}^{heavy})$ , which suggests that the sidechains of bosentan were somewhat



**Figure 10.** Snapshots sampled in the highest density spot ( $\rho_{CMB}(r_{cube}) = 0.5$  in Fig. 3). In both panels, bosentan conformations from simulation are shown as magenta, and that in the native complex is shown as green. **(a)** Four conformations selected randomly from the range of  $3.5 \text{ \AA} \leq RMSD_{whole}^{heavy} < 4.0 \text{ \AA}$ . The bosentan core region and ring A are shown in the figure. **(b)** The conformation with  $RMSD_{whole}^{heavy} = 1.11 \text{ \AA}$ . hETB in the native complex is also shown. Although superimposition was not applied to bosentan, it was applied to transmembrane helices of hETB (see Figure S7a). Transmembrane helices (TM1–7) are labeled for panel (b).

disordered. To examine this disorder in greater detail, we collected snapshots from  $3.5 \text{ \AA} \leq RMSD_{whole}^{heavy} \leq 4.0 \text{ \AA}$ , which involve the highest peak of  $P_{RDF}(RMSD_{whole}^{heavy})$  (Fig. 3b).

Figure 10a displays four randomly selected conformations from the snapshots collected above. This figure demonstrates that the conformations from the highest density spot have some similarity with the native-complex structure (the X-ray structure). Therefore, we referred to the complex conformations in highest density spot as “native-like complex”. However, Fig. 10a shows that the longest sidechain of bosentan, ring A (Figure S21c), is deviated from the native-complex position with a conformational diversity, whereas the core region was converged to a position close to the native-complex position. For this reason, the highest-peak position of  $P_{RDF}(RMSD_{whole}^{heavy})$  or  $P_{DDF}(RMSD_{whole}^{heavy})$  was larger than that of  $P_{RDF}(RMSD_{core}^{heavy})$  or  $P_{DDF}(RMSD_{core}^{heavy})$  in Fig. 3, respectively. We discuss this imperfection in the next section.

Figure 10b depicts a snapshot with a small  $RMSD_{whole}^{heavy}$  (1.11 Å). Such small  $RMSD_{whole}^{heavy}$  conformations were the minority in the ensemble of snapshots because there was no peak at  $RMSD_{whole}^{heavy} \approx 1 \text{ \AA}$  in Fig. 3. This small  $RMSD_{core}^{heavy}$  was a result obtained because the ring A of this conformation overlapped well on that of the native-complex structure.

**Comparison of bosentan at the highest-density spot with that in the crystal complex structure.** We note several discrepancies between bosentan in the crystal complex structure and that at the computed highest-density spot. Actually, GA-guided mD-VcMD is sufficiently powerful to explore the ligand–receptor conformational space starting from completely unbound conformations (Figure S5). However, we remark that reproduction of the native-like complex was not perfect. As discussed, the bosentan’s longest sidechain (ring A; Figure S21c) exhibited structural diversity in native-like complex conformations, although the core region agreed well with the native-complex position.

We infer two possibilities for the imperfection: Accuracy of classical-mechanical force-fields and/or treatment of the bosentan’s sulfonamide. Despite multiple attempts to improve force fields, current force fields are well known to be imperfect. In future studies, novel force-field parameters might resolve details of conformational inconsistency with experimentation.

Regarding the second possibility, we used the bosentan’s sulfonamide in a normal (neutral) form in that a hydrogen atom is attached to the nitrogen atom (N14) of the sulfonamide. In the neutral form, the electronic state of N14 is the  $sp^2$  hybridization with planar three-chemical bonds (S–NH–C angle of approximately 130 degrees). In the deprotonated form, the electronic state is also the  $sp^2$  hybridization with planar two-chemical bonds and one electron-lone pair (S =  $\dot{N}$  – C angle of approximately 131°). Therefore, we cannot distinguish the electronic state of sulfonamide merely by viewing the X-ray structure of bosentan.

We might treat the N14 atom deprotonated because the  $pK_a$  value of bosentan is 5.1<sup>50</sup> or 5.79<sup>51</sup> in solution: 0.5% and 99.5% of bosentan molecules are in the neutral and ionic form, respectively, in human blood (pH = 7.4). However, the form of sulfonamide in the hETB binding pocket is unknown. Reportedly<sup>52</sup>, the  $pK_a$  of sulfonamide in various compounds, which are immersed in a mixture of organic solvent and water, increases greatly with increasing fraction of the organic solvent. In addition, as reported<sup>53</sup>, the electronic state of amino

acids in the protein interior might be different from that in solution. Furthermore, the cell membrane is negatively charged. The bosentan molecules in the negatively charged form (the major fraction in blood) should only slightly access the cell surface by the repulsive electrostatic force between the cell membrane and the negatively charged bosentan. In contrast, the neutral form (the minor fraction in blood) can reach the cell surface and the receptor. Finally, we decided to use the neutral form of bosentan with protonated N14.

As described earlier, an attractive interaction can act for either the neutral or charged form of bosentan to the nitrogen ( $N_{\zeta}$ ) of Lys 182 ( $3 \times 33$ ). In fact, we demonstrated that the distribution function  $\rho_{bs}(r_{N14;Lys182N_{\zeta}}; r_{bb} = 0.25 \text{ \AA})$  exhibited a remarkable peak at  $r_{N14;Lys182N_{\zeta}} = 3 \text{ \AA}$  (Figure S23f), which reproduces the experimentally obtained observations (Table S4). This result constitutes evidence that the currently used form of sulfonamide is plausible. To analyze the structure diversity of the ring A further, hugely numerous computations must be made in future studies, with varying of the force-fields and the charge state of sulfonamide.

**Binding mechanism from the current and other simulations.** Recent advances in MD simulations have enabled investigation of ligand–GPCR interactions<sup>54</sup>. Various inter-molecular interactions and semi-stable complexes (poses) were investigated when the ligands reach the GPCR's binding pocket<sup>55</sup>. Energetically favored binding pathways and stable states (poses) in a conformational space were investigated, and information transfer from the ligand binding pocket to the intracellular region of the receptor was analyzed from the obtained conformational ensemble<sup>56–58</sup>. A previous MD simulation study of ligands and the  $\beta_2$ -adrenergic receptor ( $\beta_2$  AR) showed an energy barrier at the extracellular entrance of binding pocket<sup>59</sup>. The study suggested that the barrier is induced by dehydration of the ligand to enter the binding pocket. It is likely that this extracellular entrance may correspond to the “gate” of the binding pocket in our study because our simulation also showed the existence of a free-energy barrier, though this one is related to the ligand sliding on the hETB's N-terminal tail and to the decrease of the ligand orientational variety (i.e., entropy loss) at the binding pocket gate. This difference of the origin for the free-energy barrier between these two studies implies the existence of a variety of GPCR binding pathways.

Most studies have focused on the ligand-binding pathways after the ligands have passed the gate of the GPCR's binding pocket<sup>56–59</sup>. On the other hand, there has been little investigation where ligands being far from the binding pocket gate are concerned. One main outcome from the current study is that the disordered N-terminal tail acts as an attracter to capture the ligand. Therefore, these studies and ours complement each other with respect to understanding different aspects of the ligand–GPCR binding process.

The mechanism through which N-terminal tail of a GPCR assists with ligand binding can also vary depending on the tail's structure. For example, the N-terminal segment of a GPCR, sphingosine-1-phosphate receptor 1 (S1P<sub>1</sub>R), is not disordered but adapts an  $\alpha$ -helix in the crystal structure (PDB ID: 3v2y). This segment covers the gate of the binding pocket of the GPCR. A previous MD study<sup>60</sup> reported that the ligand ML056 diffuses within the membrane to reach the entrance of the S1P<sub>1</sub>R's binding pocket. As the ligand passes the entrance, the  $\alpha$ -helix unfolds, increasing the width of the entrance, which allows the ligand to diffuse into the pocket more readily. This simulation study and ours demonstrate that there is variety in the ligand–GPCR binding process depending on the structure of the system being studied.

## Conclusions

Prediction of the ligand–receptor complex structure is crucially important for structure-based drug design. This strategy is reasonable to develop a new drug (especially antagonist) because the molecular system finally reaches the most thermodynamically stable structure irrespective of the binding pathways. However, the binding pathways are also important to elucidate the complex. As mentioned above, MD simulations showed ligand binding pathways of various GPCRs and semi-stable complex structures (poses) when the ligands enter into the binding pocket and reach the bottom of the binding pocket<sup>5,7,54–56,58–60</sup>. The current simulation study showed that an attractive interaction acts effectively between bosentan and the N-terminal tail of hETB via the fly casting mechanism<sup>17–19</sup>, even when bosentan is located distant from the binding pocket of hETB. Some experiments reported in the related literature have indicated that the N-terminal tails of GPCRs have important physiological functions<sup>61</sup>, whereas the N-terminal tails of GPCRs have been studied insufficiently in contrast to the binding pocket. Because the tail is conformationally disordered, a simulation like the one performed for this study is helpful for elucidating the functions or roles of the N-terminal tail in the binding process.

In general, GPCRs have a long and disordered N-terminal tail (40–200 amino-acid residues long)<sup>21</sup>. Results of an MD study revealed that an amino-acid substitution to the N-terminal tail of a GPCR (human  $\beta_2$ -adrenergic receptor) varies the tail dynamics, and that accessibility of the ligand to the binding pocket also varies by substitution<sup>62</sup>. This study, in addition to our present work, suggests that the ligand–receptor interaction should be studied even when the ligand is outside the binding pocket of GPCR.

The bosentan–hETB complex formation obtained from the present study is summarized as follows: First the tip-side of the N-terminal tail of hETB captures bosentan when bosentan is fluctuating in solution via nonspecific intermolecular contacts. The bosentan–capturing conformation is identified as the free-energy basin  $H_1$  in the free-energy landscape (Fig. 5). Then ligand–sliding occurs occasionally from the tip-side to the root-side of the N-terminal tail when bosentan passes the gate of the binding pocket, which corresponds to basin switching from  $H_1$  to the deeper basin  $H_2$  with overcoming of the inter-basin free-energy barrier. We designated this mechanism as “fly-casting with ligand–sliding”. Importantly, the molecular orientational variety of bosentan decreases quickly when bosentan passes the gate (Fig. 8). This decrement of the orientational variety has a role of screening out bosentan conformations for which the molecular orientations are improper to reach the binding site (orientational selection). In fact, this orientational selection is a variant of conformational selection<sup>20</sup>. Because the decrement of the bosentan orientational variety leads to a loss of configurational entropy of bosentan,



another thermodynamic quantity should compensate the free-energetic shortcoming by the decrement of the orientational variety. Otherwise, the density of the bosentan centroid,  $\rho_{CMb}(\mathbf{r}_{cube})$ , should decrease in the binding pocket. Our results indicate that the intermolecular native contacts (electronically attractive interactions) between bosentan and hETB are formed in the binding pocket (Fig. 9). It is likely that this native contact formation is the compensation factor, and that, eventually, the native-complex conformation is completed.

Results obtained from the present study suggest that a longer N-terminal tail adds a new property to the binding mechanism. Moreover, it presents the possibility that the binding speed might be enhanced by designing the length and sequence of the N-terminal tail. In fact, earlier reports describe that long N-terminal tails exert various physiological functions of GPCRs such as ligand recognition, receptor activation, and signaling<sup>22,23</sup>. Therefore, a simulation with a longer N-terminal tail is the next challenge for ligand-GPCR simulation.

We uploaded important snapshots obtained from the current simulation to the Biological Structure Model Archive (BSM-Arc)<sup>63</sup>. The entry for this study is <https://bsma.pdbj.org/entry/35>.

## Data availability

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request. Several important snapshots during the current simulation were available in the BSM-Arc (<https://bsma.pdbj.org/entry/35>).

Received: 3 March 2022; Accepted: 2 August 2022

Published online: 13 August 2022

## References

- Syrovatkina, V., Alegre, K. O., Dey, R. & Huang, X. Y. Regulation, signaling, and physiological functions of G proteins. *J. Mol. Biol.* **428**, 3850–3868 (2016).
- Filmore, D. It's a GPCR world. *Mod. Drug Discov.* **7**, 24–27 (2004).
- Yanagisawa, M. *et al.* A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature* **332**, 411–415 (1988).
- Arai, H., Hori, S., Aramori, I., Ohkubo, H. & Nakanishi, S. Cloning and expression of a cDNA encoding an endothelin receptor. *Nature* **348**, 730–732 (1990).
- Sakurai, T. *et al.* Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. *Nature* **348**, 732–735 (1990).
- Shihoya, W. *et al.* Activation mechanism of endothelin ET<sub>B</sub> receptor by endothelin-1. *Nature* **537**, 363–368 (2016).
- Clozel, M. *et al.* Pathophysiological role of endothelin revealed by the first orally active endothelin receptor antagonist. *Nature* **365**, 759–761 (1993).
- Clozel, M. *et al.* Pharmacological characterization of bosentan, a new potent orally active nonpeptide endothelin receptor antagonist. *J. Pharmacol. Exp. Ther.* **270**, 228–235 (1994).
- Neidhart, W. *et al.* The discovery of nonpeptide endothelin receptor antagonists Progression towards bosentan. *Chimia* **50**, 519–524 (1996).
- Shihoya, W. *et al.* X-ray structures of endothelin ET<sub>B</sub> receptor bound to clinical antagonist bosentan and its analog. *Nat. Struct. Mol. Biol.* **24**, 758–764 (2017).
- Hayami, T., Higo, J., Nakamura, H. & Kasahara, K. Multidimensional virtual-system coupled canonical molecular dynamics to compute free-energy landscapes of peptide multimer assembly. *J. Comput. Chem.* **40**, 2453–2463 (2019).
- Higo, J. *et al.* Free-energy landscape of molecular interactions between endothelin 1 and human endothelin type B receptor: Fly casting mechanism. *Protein Eng. Des. Sel.* **32**, 297–308 (2019).
- Higo, J. *et al.* GA-guided mD-VcMD: A genetic-algorithm-guided method for multi-dimensional virtual-system coupled molecular dynamics. *Biophys. Physicobiol.* **17**, 161–176 (2020).
- Higo, J. *et al.* Molecular interaction mechanism of a 14-3-3 protein with a phosphorylated peptide elucidated by enhanced conformational sampling. *J. Chem. Inf. Model.* **60**, 4867–4880 (2020).
- Higo, J., Takashima, H., Fukunishi, Y. & Yoshimori, A. Generalized-ensemble method study: A helix-mimetic compound inhibits protein-protein interaction by long-range and short-range intermolecular interactions. *J. Comput. Chem.* **42**, 956–969 (2021).
- Hayami, T. *et al.* Difference of binding modes among three ligands to a receptor mSin3B corresponding to their inhibitory activities. *Sci. Rep.* **11**, 6178 (2021).
- Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc. Natl. Acad. Sci. USA* **97**, 8868–8873 (2000).
- Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
- Arai, M. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophys. Rev.* **10**, 163–181 (2018).
- Spolar, R. S. & Record, M. T. Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777–784 (1994).
- Wallin, E. & von Heijne, G. Properties of N-terminal tails in G-protein coupled receptors: A statistical study. *Prot. Eng. Des. Sel.* **8**, 693–698 (1995).
- Ersoy, B. A. *et al.* Mechanism of N-terminal modulation of activity at the melanocortin-4 receptor GPCR. *Nat. Chem. Biol.* **8**, 725–730 (2012).
- Uddin, M. S., Hauser, M., Naider, F. & Becker, J. M. The N-terminus of the yeast G protein-coupled receptor Ste2p plays critical roles in surface expression, signaling, and negative regulation. *Biochem. Biophys. Acta.* **1858**, 715–724 (2016).
- Isberg, V. *et al.* Generic GPCR residue numbers—Aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
- Pandy-Szekeres, G. *et al.* The G protein database, GproteinDb. *Nucleic Acids Res.* **50**, D518–D525 (2022).
- Higo, J., Kamiya, N., Sugihara, T., Yonezawa, Y. & Nakamura, H. Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem. Phys. Lett.* **473**, 326–329 (2009).
- Ikebe, J. *et al.* Theory for trivial trajectory parallelization of multicanonical molecular dynamics and application to a polypeptide in water. *J. Comput. Chem.* **32**, 1286–1297 (2011).
- Kasahara, K. *et al.* myPresto/Omegagene: A GPU-accelerated molecular dynamics simulator tailored for enhanced conformational sampling methods with a non-ewald electrostatic scheme. *Biophys. Physicobiol.* **13**, 209–216 (2016).
- Kutzner, C. *et al.* More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40**, 2418–2431 (2019).

30. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961–1972 (2009).
31. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
32. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
33. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
34. Kamiya, N., Fukuda, I. & Nakamura, H. Application of zero-dipole summation method to molecular dynamics simulations of a membrane protein system. *Chem. Phys. Lett.* **568–569**, 26–32 (2013).
35. Fukuda, I., Kamiya, N., Yonezawa, Y. & Nakamura, H. Simple and accurate scheme to compute electrostatic interaction: Zero-dipole summation technique for molecular system and application to bulk water. *J. Chem. Phys.* **137**, 054314 (2012).
36. Fukuda, I., Yonezawa, Y. & Nakamura, H. Molecular dynamics scheme for precise estimation of electrostatic interaction via zero-dipole summation principle. *J. Chem. Phys.* **134**, 164107 (2011).
37. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct. Genet.* **78**, 1950–1958 (2010).
38. Dickson, C. J. *et al.* Lipid14: The Amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
39. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–936 (1983).
40. Joung, I. S. & Cheatham, T. E. III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
41. Frisch M. J. *et al.* Gaussian 09, Revision D.01.
42. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
43. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
44. Wacker, D. *et al.* Crystal structure of an LSD-bound human serotonin receptor. *Cell* **168**, 377–389 (2017).
45. Che, T. *et al.* Nanobody-enabled monitoring of kappa opioid receptor states. *Nat. Commun.* **11**, 1145–1145 (2020).
46. Rodriguez-Espigares, I. *et al.* GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* **17**, 777–787 (2020).
47. Szlenk, C. T., Gc, J. B. & Natesan, S. Does the lipid bilayer orchestrate access and binding of ligands to transmembrane orthosteric/allosteric sites of G protein-coupled receptors?. *Mol. Pharmacol.* **96**, 527–541 (2019).
48. Toyoda, Y. *et al.* Ligand binding to human prostaglandin E receptor EP<sub>4</sub> at the lipid-bilayer interface. *Nat. Chem. Biol.* **15**, 18–26 (2019).
49. McCloskey, M. A. & Poo, M. M. Rates of membrane-associated reactions: Reduction of dimensionality revisited. *J. Cell Biol.* **102**, 88–96 (1986).
50. Gatfield, J., Grandjean, C. M., Bur, D., Bolli, M. H. & Nayler, O. Distinct ETA receptor binding mode of macitentan as determined by site directed mutagenesis. *PLoS ONE* **9**, e107809 (2014).
51. Krupa, A., Majda, D., Mozgawa, W., Szlęk, J. & Jachowicz, R. Physicochemical properties of bosentan and selected PDE-5 inhibitors in the design of drugs for rare diseases. *AAPS PharmSciTech* **18**, 1318–1331 (2017).
52. Şanlı, N., Şanlı, S., Özkan, G. & Denizlic, A. Determination of pKa values of some sulfonamides by LC and LC-PDA methods in acetonitrile-water binary mixtures. *J. Braz. Chem. Soc.* **21**, 1952–1960 (2010).
53. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno, E. B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. USA* **108**, 5260–5265 (2011).
54. Marino, K. A. & Filizola, M. Investigating small-molecule ligand binding to G protein-coupled receptors with biased or unbiased molecular dynamics simulations. *Methods Mol. Biol.* **1705**, 351–364 (2018).
55. Dror, R. O. *et al.* Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **503**, 295–299 (2013).
56. Schneider, S., Provasi, D. & Filizola, M. How olliceridine (TRV-130) binds and stabilizes a  $\mu$ -opioid receptor conformational state that selectively triggers G protein signaling pathways. *Biochemistry* **55**, 6456–6466 (2016).
57. Shang, Y. *et al.* Proposed mode of binding and action of positive allosteric modulators at opioid receptors. *ACS Chem. Biol.* **11**, 1220–1229 (2016).
58. Kruse, A. C. *et al.* Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552–556 (2012).
59. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. USA* **108**, 13118–13123 (2011).
60. Stanley, N., Pardo, L. & Fabritiis, G. D. The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* **6**, 22639 (2016).
61. Coleman, J. L. J., Ngo, T. & Smith, N. J. The G protein-coupled receptor N-terminus and receptor signalling: N-tering a new era. *Cell Signal* **33**, 1–9 (2017).
62. Shahane, G., Parsania, C., Sengupta, D. & Joshi, M. Molecular insights into the dynamics of pharmacogenetically important N-terminal variants of the human  $\beta$ 2-adrenergic receptor. *PLoS Comput. Biol.* **10**, e1004006 (2014).
63. Bekker, G., Kawabata, T. & Kurisu, G. The biological structure model archive (BSM-Arc): An archive for *in silico* models and simulations. *Biophys. Rev.* **12**, 371–375 (2020).

## Acknowledgements

This work was supported by JSPS KAKENHI Grants No. 21K06052 (J.H.), 20H03229 (N.K.), and JP18H05534 (H.K.), and was performed in part under the Cooperative Research Program of the Institute for Protein Research, Osaka University, CR-21-05. J.H., K.K. and K.N. acknowledge support by the HPCI System Research Project (Project IDs: hp200025, hp200063, hp200090, hp210002, hp210005, and hp210008). J.H., S.I., N.H., I.F., N.K., and Y.F. are supported by the Development of Innovative Drug Discovery Technologies for Middle-Sized Molecules project from the Japan Agency for Medical Research and Development (AMED). J.H., N.K., I.F., and Y.F. are supported by Project Focused on Developing Key Technology for Discovering and Manufacturing Drugs for Next-Generation Treatment and Diagnosis (2018–2021 and 2021–) from AMED. J.H., K.K., S.S. and H.K. were partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP21am0101106. The authors thank Dr. Isao Yasumatsu from Daiichi Sankyo RD Novare Co., Ltd. (Japan) for helpful discussion. GA-mD-VcMD was performed on TSUBAME3.0 supercomputers at the Tokyo Institute of Technology. We used UCSF Chimera ver. 1.15 for drawing molecular structures and gnuplot for graphs. Other analyses were done using homemade programs.

### Author contributions

H.N. proposed the research plan; J.H., N.K., I.F., Y.F., and H.N. designed the actual research plan; K.K. and J.H. developed the GA-mD-VcMD algorithm; K.K., S.S. and H.K. implemented the GA-mD-VcMD algorithm to gromacs using the PLUMED plug-in; G.J.B., N.K. and B.M. generated the molecular system for simulation; B.M., G.J.B. and N.K. generated the gromacs input files and performed preparatory simulations; J.H. performed the main part of the simulation and analyses; S.I., I.F. and Y.F. performed some important parts of analyses; J.H. wrote the paper with acceptance of critical comments raised by coauthors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17920-7>.

**Correspondence** and requests for materials should be addressed to J.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022