

# Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations

Mickaël Lelimosin,<sup>†,‡</sup> Vittorio Limongelli,<sup>\*,§,||</sup> and Mark S. P. Sansom<sup>\*,†</sup>

<sup>†</sup>Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

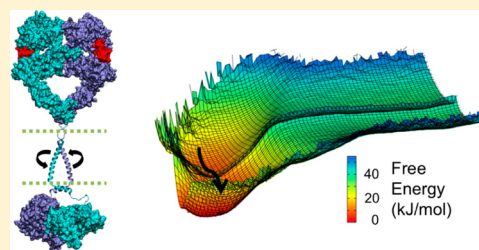
<sup>‡</sup>CERMAV, Université Grenoble Alpes and CNRS, BP 53, F-38041 Grenoble Cedex 9, France

<sup>§</sup>Università della Svizzera Italiana (USI), Faculty of Informatics, Institute of Computational Science – Center for Computational Medicine in Cardiology, via G. Buffi 13, CH-6900 Lugano, Switzerland

<sup>||</sup>Department of Pharmacy, University of Naples “Federico II”, via D. Montesano 49, I-80131 Naples, Italy

## Supporting Information

**ABSTRACT:** The epidermal growth factor receptor (EGFR) is a dimeric membrane protein that regulates key aspects of cellular function. Activation of the EGFR is linked to changes in the conformation of the transmembrane (TM) domain, brought about by changes in interactions of the TM helices of the membrane lipid bilayer. Using an advanced computational approach that combines Coarse-Grained molecular dynamics and well-tempered MetaDynamics (CG-MetaD), we characterize the large-scale motions of the TM helices, simulating multiple association and dissociation events between the helices in membrane, thus leading to a free energy landscape of the dimerization process. The lowest energy state of the TM domain is a right-handed dimer structure in which the TM helices interact through the N-terminal small-X<sub>3</sub>-small sequence motif. In addition to this state, which is thought to correspond to the active form of the receptor, we have identified further low-energy states that allow us to integrate with a high level of detail a range of previous experimental observations. These conformations may lead to the active state via two possible activation pathways, which involve pivoting and rotational motions of the helices, respectively. Molecular dynamics also reveals correlation between the conformational changes of the TM domains and of the intracellular juxtamembrane domains, paving the way for a comprehensive understanding of EGFR signaling at the cell membrane.



## INTRODUCTION

Membrane proteins play a key role in the structure and function of cells. Most membrane proteins have transmembrane (TM) domains formed by bundles of  $\alpha$ -helices that span the lipid bilayer. The dynamics of TM domains are responsible for conformational changes that underlie, e.g., receptor activation and solute transport by membrane proteins. Thus, understanding the underlying free energy landscapes of TM helix interactions provides the key to mechanisms of membrane protein function. A relatively simple and well characterized example of such interactions is provided by TM helix dimers such as those found in receptor tyrosine kinases (RTKs).<sup>1–3</sup>

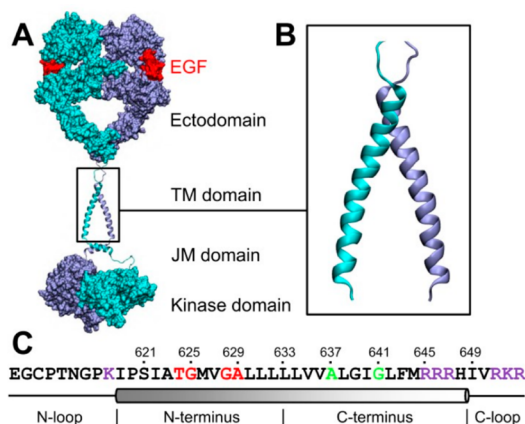
The receptor tyrosine kinases are a major family of membrane receptors controlling many cellular activities.<sup>4</sup> Among the RTKs, epidermal growth factor receptors (EGFR/ErbB1/Her1) regulate processes including cell proliferation, migration and differentiation. Aberrant ErbB activity is associated with a number of illnesses, including cancers and neurodegenerative diseases,<sup>5,6</sup> and therefore ErbB receptors are important drug targets.<sup>7,8</sup> Understanding the molecular mechanisms of signal transduction by the EGFR is of fundamental importance in providing a paradigm for other

RTKs<sup>9–11</sup> and will facilitate development of a range of therapeutic interventions.<sup>12–14</sup>

There has been extensive structural characterization of the extracellular ectodomain and the intracellular kinase domain of the EGFR.<sup>15</sup> These studies indicate that the EGFR functions as a dimer.<sup>16</sup> Binding of the ligand (EGF) to the ectodomain of the receptor induces a conformational change that is transmitted through an allosteric mechanism to the kinase domain.<sup>17</sup> Formation of an asymmetric dimer of the kinase domains enables tyrosine phosphorylation, which in turn activates the downstream signal transduction cascade in the cytoplasm.<sup>18</sup> Although cryo-electron microscopy has yielded images of the intact EGFR,<sup>19</sup> in the absence of a high resolution structure of the complete receptor, including the transmembrane (TM) domain, our understanding of its activation mechanism remains incomplete (Figure 1).<sup>1</sup> The TM helix dimer is believed to play a key role in the activation mechanism.<sup>20</sup> Mutagenesis experiments have also highlighted the role of the juxtamembrane domain (JM), immediately C-terminal to the TM domain, in regulating the allosteric transition between inactive

Received: June 8, 2016

Published: July 26, 2016



**Figure 1.** (A) Schematic diagram of the structure of a ligand-bound EGFR dimer. The EGFR monomers are in cyan and ice blue, the EGF ligands in red. The model was built using the structures of the ligand-bound ectodomain dimer (pdb id: 3NJP), of the TM helix dimer (pdb id: 2M20), and of the asymmetric kinase domain dimer (pdb id: 3GOP). (B) Structure of the TM domain used in the simulations. (C) Sequence of the TM domain and schematic representation of the secondary structure. Cationic residues are in purple; residues from N-terminal and C-terminal small-X<sub>3</sub>-small motifs are in red and green, respectively.

and active states of EGF-bound receptor dimers<sup>11</sup> and in their interactions with lipids.<sup>21</sup> Truncation of the ectodomain activates intracellular phosphorylation,<sup>22</sup> demonstrating that the ectodomain has an autoinhibitory function, and that the TM domain can independently activate the receptor.<sup>20</sup> It is suggested that the active state is characterized by dimerization contacts in the N-terminal region of the TM helices (Figure 1C, motif highlighted in red) along with an antiparallel interaction between JM segments.<sup>23</sup> The nature of the inactive state(s) remains somewhat less clear but is thought to involve a more C-terminal interaction of the TM helices (as recently discussed in the literature<sup>24</sup>).

In this context it is important to elucidate modes of interaction of the helices within the dimeric TM domain.<sup>25–27</sup> The conformational dynamics of these interactions may be explored through computer simulations, which have been used to provide detailed descriptions of different ensembles of conformations visited by TM helix dimer systems.<sup>28–30</sup> However, due to the limiting time scale of atomistic molecular dynamics (MD) simulations only partial description of EGFR dynamics have so far been achieved.<sup>31,32</sup> Consequently, models which employ a simplified (e.g., coarse-grained: CG) molecular representation are valuable,<sup>33</sup> as they may be used to investigate TM helix dimerization events beyond the accessible time scale of atomistic simulations.<sup>34–36</sup>

CG simulations have been used to obtain one-dimensional projections of the free energy landscape of TM helix dimerization for glycoprotein A,<sup>37,38</sup> for ErbB,<sup>39</sup> and for EphA1<sup>40</sup> receptors. However, in order to link the conformational motions of TM helices to possible activation mechanisms, an accurate and exhaustive (i.e., multidimensional) description of the underlying free energy landscape is essential. This remains beyond the accessible time scale of simple CG simulations, even with the availability of advanced supercomputing resources. A potential solution is provided by the use of enhanced sampling techniques.<sup>41–43</sup> These methods accelerate the simulation of the process of interest, allowing

capture of long time scale motions at affordable computational cost.

Here we present an innovative computational protocol that by combining CG simulations with well-tempered metadynamics<sup>44,45</sup> (CG-MetaD) greatly extends the time scales typically reached by simulations. In particular, CG models reduce the dimensionality of the system under investigation,<sup>36</sup> while metadynamics enhances the phase space exploration, thus allowing practical investigation of long time scale and large-scale motions of biomolecules.<sup>46–48</sup> Pioneering applications of metadynamics to CG models have indeed shown great potential.<sup>49,50</sup> In the present study we employ  $\sim 380 \mu\text{s}$  of CG-MetaD simulations to fully inspect both association and dissociation events of the EGFR TM helices. This provides an accurate and comprehensive description of the associated free energy landscape, which allows identifying the low free-energy basins that correspond to distinctive TM dimer conformations. The reversible pathways of transition between these dimeric states reveal key functional motions of the TM helices to reach the active dimer state, thus unveiling previously unknown features of the activation mechanism. This information is of paramount relevance to develop an exogenous control of the receptor activity. The conformational stability of the identified low energy dimer states has been confirmed by over  $3 \mu\text{s}$  of atomistic MD simulations, thus providing structural details that explain and complement previous studies. This study demonstrates that CG-MetaD provides a powerful method to investigate large-scale motions in the TM domains of membrane proteins, thus representing an important advance in simulation of protein conformational dynamics in a membrane environment.

## METHODS

**Coarse-Grained Models.** The starting structure of the TM domain dimer was a homology model based on the NMR structure of the ErbB1/ErbB2 heterodimer (pdb id: 2KS1),<sup>51</sup> generated using Modeler.<sup>52</sup> Coarse-graining used an in-house variant<sup>53,54</sup> of the Martini force field.<sup>55,56</sup> In addition to the standard bonded interactions, harmonic restraints are generally required to conserve secondary structure elements. Therefore, in the TM models, harmonic restraints were applied between backbone particles separated by three residues in the  $\alpha$ -helical region (from Ile<sup>619</sup> to His<sup>648</sup>) to mimic hydrogen-bonds, while the intrinsic flexibility of the N-loop and C-loop regions were maintained with no additional restraints. In the TM+JM models (based on the 2M20 NMR structure), an elastic network model acts between pairs of backbone particles separated by less than 0.7 nm.

**CG-MetaD Simulation.** The CG-MetaD simulation was performed with the well-tempered version<sup>45</sup> of metadynamics,<sup>44</sup> implemented in Plumed 1.3.<sup>57</sup> The estimate  $F(s,t)$  of the free-energy surface  $F(s)$  at time  $t$  was determined as

$$F(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) \quad (1)$$

where  $V(s,t)$  is the bias potential added to the system at time  $t$  as a function of a set of selected CVs,  $s$ .  $\Delta T$  is the difference between the fictitious temperatures ( $T + \Delta T$ ) of the selected CVs and the temperature  $T$  of the system. Therefore,  $\Delta T$  influences the exploration rate of the CVs space. The bias potential  $V(s,t)$  is made up by the sum of the Gaussians deposited so far.

In our case the main simulation was performed at 323 K (i.e., the temperature at which the MARTINI force field was parametrized) with a time step of 40 fs. The lateral separation ( $L$ ) between the center of mass of the helices was chosen as single active collective variable. Gaussians of height 0.05 kJ/mol and width 0.05 nm were initially used and deposited every 5000 steps with a bias factor  $(T + \Delta T)/\Delta T$  of 10. This means that the Gaussian height was gradually decreased along the

380  $\mu$ s of CG-MetaD simulation on the basis of the adaptive bias with a  $\Delta T$  of 2907 K (Supporting Information Figure S1).

Additional CG-MetaD simulations were performed using different parameters: the width of the Gaussians was reduced to 0.025 nm in one simulation or alternatively a soft wall potential was used to restrain the lateral helix separation  $L$  within 4.5 nm (Supporting Information Figure S2). The results obtained from the different simulations led to a coherent description of the free energy landscape (Supporting Information Figure S3), demonstrating the robustness of the approach. In contrast we observed both tempering and convergence issues with higher deposition rates (either with higher gaussians or higher deposition frequencies). Unless stated in the text, CG-MetaD results refer to the main simulation.

**CG-MD Simulations.** A series of standard CG simulations were performed using Gromacs 4.5.5<sup>58</sup> (i.e.,  $\sim 185 \mu$ s of CG-MD simulations, see Supporting Information Table S1). A leapfrog algorithm was used to integrate Newton's equations of motion, with a time step of 40 fs. Standard CG-MD and CG-MetaD simulations were performed in  $NpT$  and  $NVT$  ensembles, respectively. The temperatures of the individual monomers, the solvent and the lipid bilayer were separately maintained at 323 K using the Bussi thermostat with relaxation times of 1 ps.<sup>59</sup> In standard CG-MD simulations, semi-isotropic pressure coupling was used. The Berendsen barostat was employed to keep the average pressure at 1 bar,<sup>60</sup> with time constants of 10 ps and compressibilities of  $3 \times 10^{-5} \text{ bar}^{-1}$ . Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm, while electrostatic interactions were shifted to zero between 0 and 1.2 nm using a relative dielectric constant of 20. Neighbor lists were updated every 10 steps for the calculation of nonbonding interactions, using a 1.3 nm cutoff.

**AT-MD Simulations.** The populations of lowest energy identified by CG-MetaD were converted to atomistic level using a fragment-based approach.<sup>61</sup> A total of over 3  $\mu$ s of standard atomistic MD simulations (AT-MD) were performed using the AMBER force field ff14SB<sup>62</sup> and LIPID14<sup>63</sup> for protein and lipids, respectively. The TIP3 water model was used.<sup>64</sup> All the simulations were performed with the pmemd module of the AMBER MD code.<sup>65</sup> Short-range van der Waals interactions were switched to zero at a cutoff distance of 1 nm. The long-range electrostatic interactions were computed by means of the particle mesh Ewald (PME) method<sup>66</sup> using a real-space cutoff of 1 nm and a 0.1 nm grid spacing in periodic boundary conditions. The SHAKE algorithm was applied to constraint bonds involving hydrogen atoms, and thus an integration time step of 2 fs could be used. Harmonic constraints were applied to the protein and gradually released along the equilibration process. The temperatures of the individual monomers, the solvent and the lipid bilayer were maintained at 323 K using the Langevin thermostat with collision frequency of 3 ps<sup>-1</sup>. Production runs were performed in the  $NpT$  ensemble at 1 atm under anisotropic pressure scaling conditions using a Monte Carlo barostat and pressure relaxation time of 2 ps.

**Collective Variables (CVs).** Analysis of the simulations was carried out using Plumed 1.3,<sup>67</sup> Gromacs 4.5.5<sup>58</sup> and locally written scripts. The probability distribution function of various CVs was reweighted for free energy calculations.<sup>57</sup> The CVs used were:  $L$ , the lateral distance between the centers of mass of the TM helices;  $\Omega$ , the crossing angle between the TM helices;  $\Delta d$ , calculated as the difference in interhelix distance at the N-termini and the C-termini; and  $\Delta|\Theta|$ , which describes the rotation of the helices along their long axis relative to one another. More detailed definitions are provided in the paper when these CVs are used. However, the crossing angle CV ( $\Omega$ ) benefits here from a more complete description. The positions of the backbone particles of the TM helical region were used to determine, by singular value decomposition (SVD), the eigenvectors  $\mathbf{h}_1$  and  $\mathbf{h}_2$  associated with the helix axes. The absolute value of crossing angle was determined using the definition  $|\Omega| = \text{asin}(|\mathbf{h}_1 \times \mathbf{h}_2| / (|\mathbf{h}_1| |\mathbf{h}_2|))$ . One might also consider  $\text{atan2}(|\mathbf{h}_1 \times \mathbf{h}_2|, \mathbf{h}_1 \cdot \mathbf{h}_2)$  that leads to identical  $|\Omega|$  values. In order to compute the sign of  $\Omega$ ,  $\mathbf{h}_1$  was first aligned with the positive  $x$ -axis, with the same rotation applied to  $\mathbf{h}_2$ , which provided two new helix vectors  $\mathbf{h}_1'$  and  $\mathbf{h}_2'$ . Then, the cross product  $\mathbf{h}_1' \times \mathbf{h}_2'$  was aligned with the  $x$ -axis, and the same rotation

applied to the two helix vectors provided  $\mathbf{h}_1''$  and  $\mathbf{h}_2''$ . All rotations were done using the origin of the Cartesian coordinate system as center. The sign of  $\Omega$  was determined by considering two properties: (i) the sign of the  $x$ -component of the geometric center of the cross product  $\mathbf{h}_1'' \times \mathbf{h}_2''$ , (ii) the comparison of the  $x$ -component of the geometric centers of  $\mathbf{h}_1''$  and  $\mathbf{h}_2''$ . Negative and positive signs of  $\Omega$  characterized right- and left-crossings, respectively.

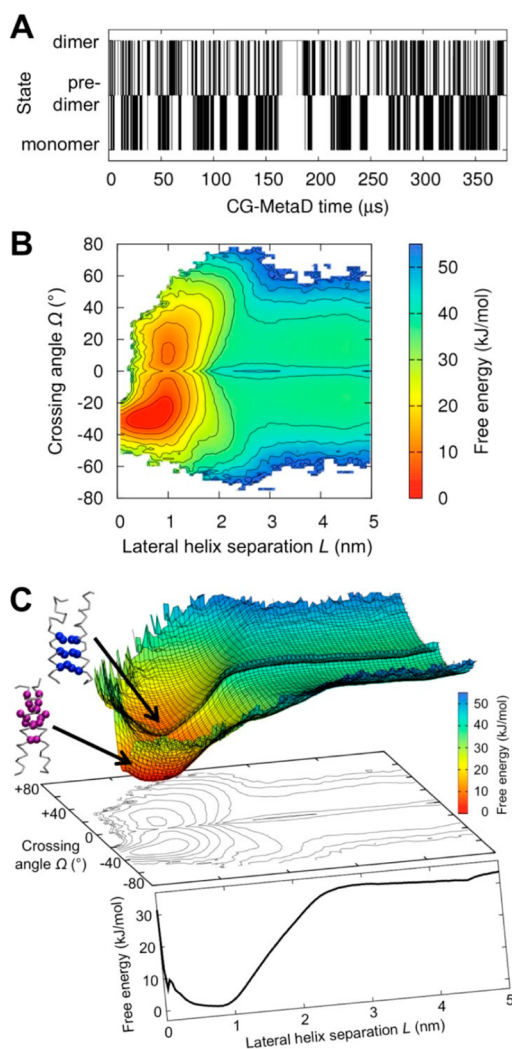
**Minimum Free Energy Paths (MFEPs).** All minimum free energy paths (MFEPs) presented in the paper were calculated in one of the two dimensions of the free energy surfaces obtained by CG-MetaD calculations. The paths were obtained by scanning of the lowest free energy points of the surface along the (CV) dimension of interest. We compared the MFEPs to free energy profiles obtained for each CV. For the determination of the free energy profiles, partition functions were calculated at each value of the CV of interest. Thus, contrary to MFEPs, all the free energy points obtained in the second dimension of the FES contributed to the free energy profiles. The comparison showed that both are really similar, although the MFEPs allowed a more accurate detection of the free energy basins of the FES. Therefore, we chose to present our results on the basis of the MFEPs. Importantly the relative free energies of the basins identified from the different MFEPs were similar for each dimer population (Supporting Information Figures S3 and S4). This guarantees both the validity of the free energy calculations and the suitability of the different CVs to describe the multidimensional conformational landscape of the EGFR TM domain (Supporting Information Table S2).

Structural figures were rendered using either VMD<sup>68</sup> or PyMOL (<http://www.pymol.org>).

## RESULTS AND DISCUSSION

**CG-MetaD Sampling.** A starting structure for the simulations of the EGFR TM helix dimer was based on the NMR structure of the ErbB1/ErbB2 heterodimer (pdb id: 2KS1). This structure includes flexible N-terminal (extracellular) and C-terminal (intracellular) loops, which we will refer to as the N-loop and C-loop, respectively (Figure 1). Therefore, in the CG simulations, harmonic restraints were applied (see Methods above) to maintain the secondary structure of the core  $\alpha$ -helical region (from Ile<sup>619</sup> to His<sup>648</sup>) while retaining the intrinsic flexibility of the N-loop and C-loop regions. The N-loop contains residue K618 that is thought to interact with the glycolipid GM3 in cell membranes.<sup>69,70</sup> The C-terminus of the TM helix plus the C-loop contains six basic residues (R645 to R653) that can interact with anionic lipids (e.g., PIP<sub>2</sub>) of the intracellular membrane leaflet.<sup>21,70,71</sup> The TM helix dimer was embedded in a phospholipid bilayer.<sup>34</sup> To describe the motion of one helix with respect to the other we used the lateral separation between the centers of mass of the helices ( $L$ ) as a collective variable (CV). The sampling of TM helix dimerization took a total of  $\sim 0.4$  ms of CG-MetaD simulation, during which 65 interconversion events between the dimeric and monomeric states of the TM helices were observed. Thus, the diffusive exploration of the CV space in our CG-MetaD simulation yielded a well-characterized and converged free energy surface (Figure 2 and Supporting Information Figure S1).

**Free Energy of Dimerization.** From the free energy surface (FES) one can accurately quantify free energies of interaction.<sup>72</sup> Thus, the free energy of TM helix dimerization was calculated from the FES as the difference between the dimeric and monomeric states, yielding an estimate of  $-38 \pm 3$  kJ/mol (error calculated as the standard deviation from the weighted average value of the TM helix dimerization free energy obtained from the last part of the simulation where the calculation is converged; see also Supporting Information

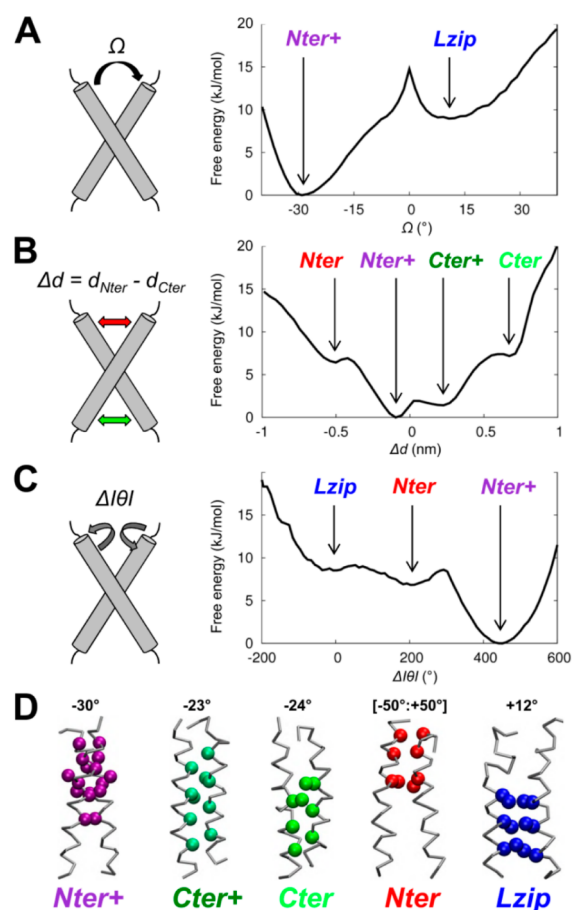


**Figure 2.** (A) Interconversion of the TM domain between three states: monomeric state (lateral helix separation,  $L > 3$  nm), predimeric ( $1.5 \text{ nm} < L < 3 \text{ nm}$ ), and dimeric ( $L < 1.5 \text{ nm}$ ). (B) Top-view of the free energy surface (FES) as a function of  $L$  and of the crossing angle  $\Omega$ . Isoenergy lines are plotted every 5 kJ/mol on the FES. (C) 3D-view of the FES (upper part). Conformations corresponding to the free energy basins are drawn with gray tubes for the helix backbone and colored spheres for residues engaged in interhelical contacts. Projection of the FES in the  $L$  and  $\Omega$  dimensions shows isoenergy lines every 5 kJ/mol (middle panel). One dimensional free energy profile as a function of the active CV  $L$  (lower panel).

Figure S1C). This is in reasonable agreement with an experimental estimate of the dimerization free energy of  $-31$  kJ/mol, derived from a  $K_d$  of  $\sim 2 \mu\text{M}$  in FRET experiments on EGFR TM helices in LDAO micelles.<sup>73</sup> We do note that these values are larger than some other computational and experimental estimates, of  $-26$  kJ/mol<sup>39</sup> and  $-11$  kJ/mol<sup>74</sup> respectively. However, considering the sensitivity of the dimerization process to experimental conditions (e.g., lipid vs detergent), the approximations implicit in the CG force fields used in simulations, and especially the different construct sequences of the EGFR TM domain used in the various studies (Supporting Information Figure S5), it is perhaps unsurprising that we observe differences. Indeed, we note that a wide range of values (from  $-30$  kJ/mol to  $-60$  kJ/mol) has been estimated for the dimerization free energy of the canonical

glycophorin A helix dimer on the basis of atomistic<sup>75,76</sup> and CG<sup>37,38</sup> simulations.

**2D Free Energy Landscape.** Using a reweighting protocol we computed the Boltzmann distribution of CVs different from those biased in the original calculation.<sup>57</sup> This allows for a more detailed inspection and characterization of the motions observed during the simulation. In this manner we reconstructed the FES as a function of both the original CV (the interhelix distance  $L$ ) and of a CV corresponding to the interhelix crossing angle ( $\Omega$ ). The latter is a frequently used structural descriptor of the packing interactions within TM helix dimers.<sup>34</sup> The resulting FES (Figure 2B) reveals two free-energy minima. The lower minimum corresponds to right-handed ( $\Omega = -30^\circ$ ) TM helix dimer conformations, while the second minimum is 9 kJ/mol higher and corresponds to a left-handed dimer ( $\Omega = \sim +12^\circ$ ; Figures 2BC and 3A). Within the right-handed energy basin of the TM dimer, we refer to the



**Figure 3.** Schematic definition and associated free energy profiles of the CVs: (A)  $\Omega$ , (B)  $\Delta d$ , and (C)  $\Delta|\theta|$ .  $\Omega$  is the crossing angle between helices;  $\Delta d$  CV is calculated as the difference in interhelix distance at the N-termini and the C-termini (defined as the centers of mass of  $\text{TG}^{625}\text{xxGA}^{629}$  and  $\text{A}^{637}\text{xxxG}^{641}$  motifs, respectively);  $\Delta|\theta|$  describes the rotation of the TM helices about their long axes relative to one another (see also Supporting Information Figure S4). Each free energy profile was calculated as the minimum free energy path (MFEP) located at low  $L$  values of the associated FES (see Methods). (D) Structure of the dimer populations identified in the basins of the free energy landscape of EGFR TM domain. Colored spheres represent residues mainly engaged in interhelical contacts, while gray tubes show helix backbones. Averages or ranges of crossing angles are provided for each dimer.

most populated ensemble of conformations as *Nter+* (Figure 3D) since it shows the two helices interacting through residues of the N-terminal motif along with additional contacts at the helix core ( $G^{625}\text{xxGA}^{629}\text{xxLL}^{633}\text{xxxA}^{637}$ ). We will refer to the left-handed dimer as the *Lzip* population (Figure 3D) since the helices form leucine zipper-like interactions ( $V^{635}\text{xxL}^{638}\text{xxxL}^{642}$ ). The *Nter+* state is  $11 \pm 2$  kJ/mol more stable than the *Lzip* state.

Although the helix crossing angle is a metric commonly used to characterize TM helix dimers,<sup>34</sup> it may not allow us to readily distinguish subtly different dimer conformations within the same macro-ensembles (i.e., within the right-handed and left-handed energy basins). Therefore, we undertook a series of local analyses of all dimer conformations.

**Multidimensional Free Energy Landscape.** The sequence of the EGFR TM domain contains a tandem small- $X_3$ -small sequence motif toward the N-terminus ( $TG^{625}\text{xxGA}^{629}$ ) and a single small- $X_3$ -small motif ( $A^{637}\text{xxxG}^{641}$ ) toward the C-terminus (Figure 1).<sup>51</sup> Such small- $X_3$ -small motifs play a key role in TM helix dimerization in a number of membrane proteins.<sup>77</sup> Interconversion between different TM helix packing modes has been implicated in activation of the EGFR and related receptors (see the literature<sup>24</sup> for a recent discussion). On this basis, we defined a new CV,  $\Delta d$ , which corresponds to the difference in interhelix distances between these two motifs, such that a negative value of  $\Delta d$  indicates an N-terminal helix/helix interaction and a positive value indicates a C-terminal interaction (Figure 3B). The minimum free energy path (MFEP) along the  $\Delta d$  CV, corresponding to the lowest energy points computed in the  $\Delta d$  range from  $-1$  to  $+1$  nm, reveals the presence of four energy minima (Figure 3B). The deepest energy minimum (at  $\Delta d = -0.1$  nm) corresponds to the previously described *Nter+* population. An additional low free-energy minimum, named *Cter+*, shows the two helices interacting through both the C-terminal small- $X_3$ -small motif and some of the core hydrophobic residues ( $A^{629}\text{xxxL}^{633}\text{xxxA}^{637}\text{xxxG}^{641}$ ). We note that the *Nter+* population is slightly more stable than *Cter+* ( $\sim 1.4$  kJ/mol), as suggested by previous experimental studies.<sup>78</sup> The other two free-energy minima, *Nter* and *Cter*, are 6 to 7.5 kJ/mol higher than *Nter+* and correspond to helix dimers in contact primarily through their N-terminal and C-terminal small- $X_3$ -small motifs, respectively. These interactions are similar to those found in the *Nter+* and *Cter+* conformations, with the lack of additional contacts at the core of the helices explaining the higher free energy values of these populations (Supporting Information Figure S4A–D).

From the analyses presented so far, the *Nter+* dimer always represents the lowest free-energy minimum. However, the left-handed *Lzip* dimer is absent from the FES computed as a function of  $\Delta d$  (Figure 3B). Accordingly, we defined an additional CV,  $\Delta|\theta|$ , which is able to further characterize specific conformational differences among the diverse dimer structures. This CV describes the rotation of the TM helices about their long axes relative to one another. The MFEP computed as a function of  $\Delta|\theta|$  reveals three out of the five dimer populations previously described (Figure 3C). In this representation the *Nter* dimer appears as an intermediate conformation between *Lzip* and *Nter+*.

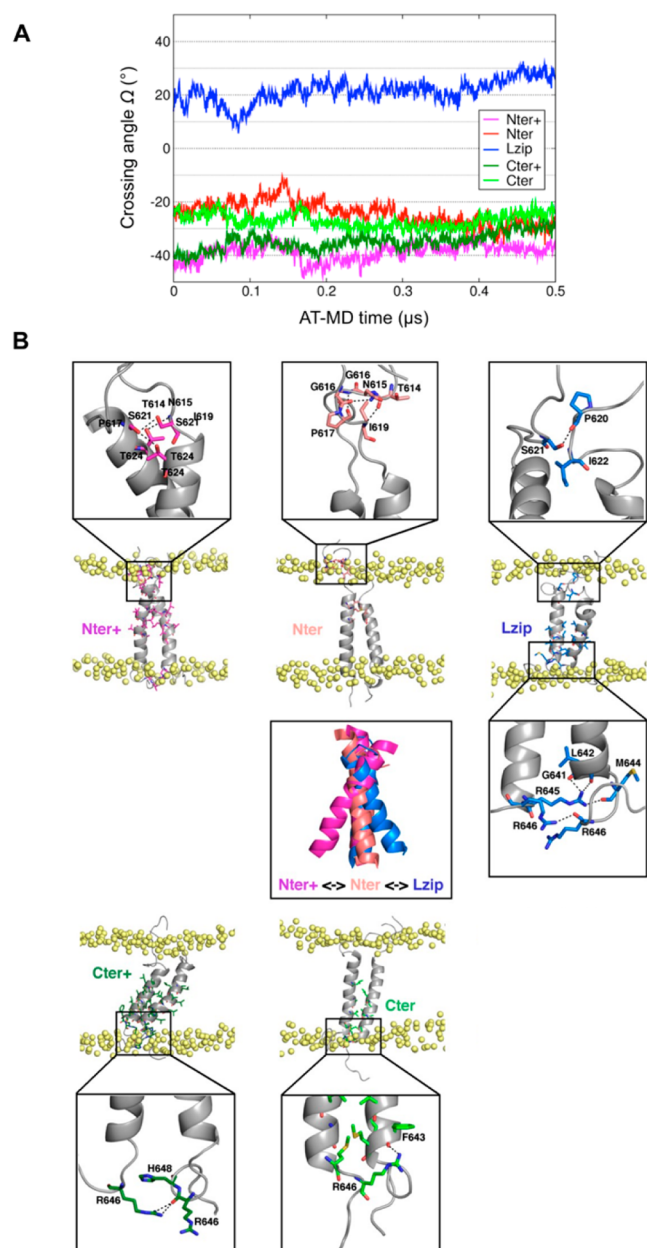
CG-MetaD simulations have allowed us to explore the large-scale motions of the TM domain, revealing the existence of five dimeric states (Figure 3D, Supporting Information Table S2). We have used each of these dimeric state models as the starting

point for extended atomistic simulations, in order to explore in more detail specific inter-residue interactions between the two helices. Thus, each dimeric model of the five lowest energy minima identified by CG-MetaD was used to initiate a  $0.5 \mu\text{s}$  of unrestrained atomistic (AT) MD simulations (yielding a total simulation time of  $2.5 \mu\text{s}$ ). We found that the dimer conformations were stable during the AT-MD simulations, with, e.g., no significant change in helix crossing angles over the course of the  $0.5 \mu\text{s}$  simulations (Figure 4A; also see Supporting Information Figures S6). Examination of the structures at the end of the atomistic simulations confirmed that the hydrophobic residues of the TM domain form interactions with the lipid aliphatic chains within the bilayer core, while interhelix contacts are formed by the interaction motifs identified through the CG-MetaD simulation. We found that a number of H-bonding interhelix interactions involving the N-ter and C-ter loop residues also contribute to stabilize the dimeric states, in addition to electrostatic interactions between cationic residues and lipid headgroups (Figure 4B; also see Supporting Information SI text).

Examination of the CG-MetaD derived free energy landscape reveals that in the unbound state (i.e., monomeric TM helices at  $L > 3$  nm) the surface is not flat along the  $\Omega$  dimension (Figure 2C). The shallow free-energy dependence on the crossing angle, with favorable free energies at  $\Omega \sim -18^\circ$  and  $\Omega \sim +18^\circ$ , corresponds to the (monomeric) helices adopting a tilted orientation, with a lowest free-energy tilt angle of  $\sim 20^\circ$  relative to the bilayer normal (Supporting Information Figure S7), close to values reported for the tilt of synthetic peptides in model lipid bilayers.<sup>79–81</sup>

**Dimerization Pathways.** The bias added to the system during the CG-MetaD calculation allows acceleration of sampling, and the reconstruction at the end of the simulation of the thermodynamics of the event under study (i.e., the free-energy landscape and free-energy minima identification). The characterization of the two-dimensional FES (Figure 2B) also therefore provides information on the possible pathways connecting the monomeric state to the different dimeric structures. We divided the FES into two regions separated by the free energy barrier observed at a crossing angle of  $0^\circ$  in order to calculate MFEPs for dimerization (i.e., from high to low  $L$  values). We observed similar features for the calculated pathways in the two regions (Figure 5A). At high  $L$  values, the MFEPs linearly follow the free energy basins at  $\Omega \sim -18^\circ$  and  $\Omega \sim +18^\circ$  discussed above, which represents the tilted orientations of TM helices in their monomeric state. In contrast at intermediate  $L$  values, ( $1.5 \text{ nm} < L < 3 \text{ nm}$ ) corresponding to predimeric states, the MFEPs show a nonlinear pattern with an increase of the magnitude of the crossing angle between the TM helices when they first contact each other. This is followed by a decrease in the magnitude of the crossing angle when the helices pack together more closely (Figure 5A). Subsequently, the MFEPs follow the curvature of the free energy wells observed at low  $L$  values.

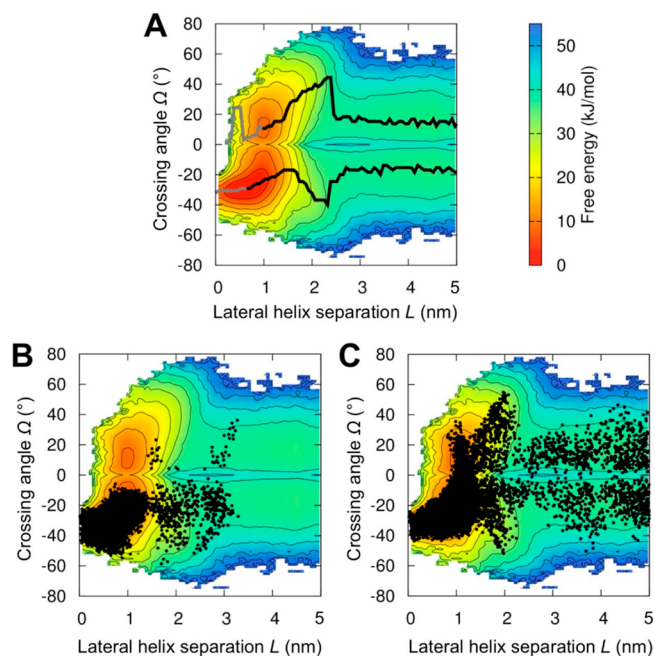
These findings prompted us to compare the dimerization pathways predicted by CG-MetaD to those observed through unbiased CG simulations. We therefore performed three standard CG-MD simulations, each of  $5 \mu\text{s}$  duration starting from monomeric states ( $L \sim 3$  nm) with different  $\Omega$  values. These simulations displayed dimerization of the TM helices with features similar to those predicted by the MFEPs. However, comparison of the CG-MD simulations showed a



**Figure 4.** Atomistic MD simulations initiated from each dimer conformation of the five lowest energy minima identified by CG-MetaD. (A) Crossing angles as a function of time for the five AT-MD simulations. (B) Atomistic models selected from the five EGFR dimer populations obtained from AT-MD simulations. Helix-helix interactions at loop regions are shown as insets. H-bonds are displayed as black dashed lines. In the central panel a superimposition of the *Nter+*, *Nter* and *Lzip* conformations is shown, highlighting the intermediate role played by *Nter* in the *Lzip*-to-*Nter+* transition. TM domains are represented as gray cartoons, zwitterionic headgroups of phospholipids as yellow spheres, while waters and lipid tails are omitted for clarity.

range of trajectories on the FES, highlighting the stochastic behavior of individual dimerization events.

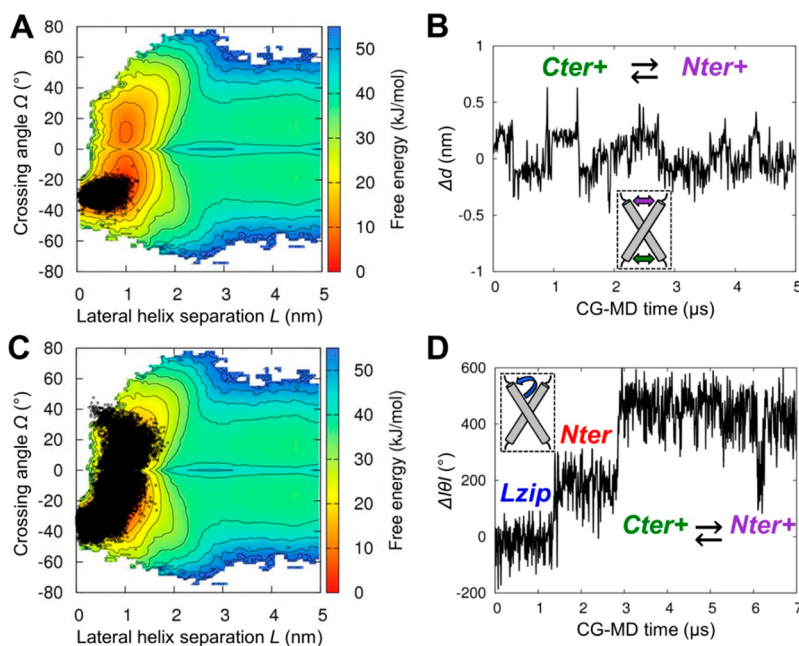
Dimerization of the TM helices occurred early after  $\sim 50$  ns in the first simulation, following the MFEP calculated in the region of negative  $\Omega$  values of the FES (Figure 5B). In the second simulation a wider exploration of the FES at high  $L$  values was seen (Figure 5C). The longer diffusion time of the helices in the monomer state resulted in a well-balanced representation of the two low energy populations, with both



**Figure 5.** (A) Minimum free energy pathways (MFEPs) between monomeric and dimeric states in the regions of the FES with  $\Omega$  above or below the zero crossing angle. MFEPs are drawn as black lines between monomeric states with high  $L$  ( $>5$  nm) and the free energy minima of the dimeric states, while MFEPs between these latter points and the dimeric states with lower  $L$  values are shown as gray lines. Free energy profiles of these pathways are in the Supporting Information (Figure S1F). (B,C) Regions of the FES explored during 5  $\mu$ s of standard CG-MD simulation starting from two randomly chosen initial monomeric states: (B)  $L = 3$  nm and  $\Omega = -28^\circ$ ; and (C)  $L = 3$  nm and  $\Omega = -7^\circ$ . Each frame of the standard simulations is represented by a black dot. Isoenergy lines are plotted every 5 kJ/mol on the FES.

right-handed and left-handed monomeric states sampled. In this simulation the two TM domains first interacted with each other through the extracellular N-loops. These predimeric states can readily convert from left- to right-handed forms in agreement with the quasi-linear isoenergy lines observed along the  $\Omega$  axis in this region of the FES ( $1.5 \text{ nm} < L < 2 \text{ nm}$ ; Figure 5A). Subsequently, more stable interactions are formed via the residues at the core of the helices. In these dimeric states ( $L \approx 1$  nm) the system still explores both right-handed and left-handed forms, through an equilibrium between the *Lzip* and the *Nter* populations. The final stage of the simulation shows the stabilization of the dimer in right-handed forms corresponding to the global minimum of the FES (Figure 5C). The third simulation (Supporting Information, Figure S8) showed a mixed profile with respect to the first two simulations, namely a fast association of the helices followed by the exploration of both right-handed and left-handed forms to reach the final dimeric state. Interestingly these results suggest that the equilibrium between the *Lzip* and the *Nter* populations may be a preliminary pathway to access the most stable *Nter+* state.

The agreement between CG-MetaD and standard CG-MD in finding the lowest energy pathways in the monomeric, predimeric, and dimeric states demonstrates that the metadynamics bias did not corrupt the thermodynamics of the system, and that the CVs  $L$  and  $\Omega$  properly describe the motion of the helices during the dimerization/dissociation processes.



**Figure 6.** (A) Region of the FES explored during 5  $\mu\text{s}$  of standard CG-MD simulation starting from a *Cter+* conformation, and (B) time-evolution of the  $\Delta d$  CV. (C) Region of the FES explored during 7  $\mu\text{s}$  of standard CG-MD simulation starting from a *Lzip* conformation, and (D) time-evolution of the  $\Delta|\theta|$  CV. In (B,D) conformations observed during the simulations are indicated by the colored labels, and the insets depict the observed pivot and rotational motions.

### Functionally Relevant Motions of the TM Domain.

Elucidating the structural transitions that follow the initial dimerization step is fundamental to our mechanistic understanding of EGFR activation. In particular, the equilibrium dynamics of the TM domain may reveal how the helix interactions change when the active and inactive dimeric states interconvert. Thus, we evaluated through standard (i.e., unbiased) CG-MD simulations whether the four higher energy dimeric states identified in the multidimensional landscape could convert into the *Nter+* conformation, which is thought to correspond to the activated state of the receptor.<sup>23</sup>

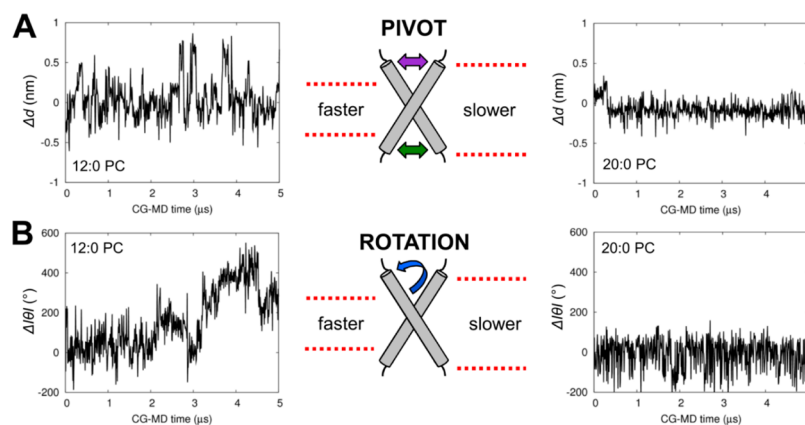
First, we performed a standard CG-MD simulation (of 5  $\mu\text{s}$  duration) starting from the *Cter+* conformation, which corresponds to a possible inactive state of the receptor.<sup>23,24,32</sup> This explored the various low energy conformations of the right-handed (negative  $\Omega$  value) region of the FES (Figure 6A). Analysis and structural examination revealed that there is equilibrium between the *Cter+* and the *Nter+* populations, in agreement with the free energy calculations that indicated a low energy barrier between the two dimers ( $\sim 2$  kJ/mol from *Nter+* to *Cter+*; see Figure 3B). In particular, we found that these two populations can interconvert through a “pivot” motion whereby the contacts formed by the N-terminal small- $X_3$ -small motif alternate with those formed by C-terminal motif (Figure 6B). A similar mechanism has been invoked to explain the correlation between the TM conformational transitions and the activation of the EGFR intracellular kinase domains.<sup>23,24,32</sup>

Comparable standard CG-MD simulations starting from the (left-handed) *Lzip* dimer showed a transition to the more stable right-handed *Nter+* ensemble (Figure 6C). This transition occurred through a two-step mechanism involving helix rotation as characterized by evaluation of  $\Delta|\theta|$  (Figure 6D). In particular, after  $\sim 1.5$   $\mu\text{s}$  one of the two helices rotated leading to the intermediate (asymmetric dimer) *Nter* population. Subsequently rotation of the second helix leads to

the *Cter+* conformation that rapidly evolves to the active *Nter+* ensemble (Figure 6D). Alongside some experimental studies,<sup>82</sup> our results suggest that the “rotation” motion of the TM helices of the EGFR may correspond to another possible activation pathway.

We also performed CG-MD simulations starting from *Cter* and from *Nter* (Supporting Information Figure S9). We found a rapid transition (after  $\sim 0.25$   $\mu\text{s}$ ) from the *Cter* population to *Nter+*, which then existed in equilibrium with *Cter+*. This is in line with the free energy landscape previously described (Figures 3B). Starting from the *Nter* dimer, an equilibrium with the *Lzip* population was reached and followed by transition to *Nter+*. These observations confirm our suggestion that *Nter* is an intermediate state between *Lzip* and *Nter+* (Figure 3C). Some structural asymmetry, observed in *Nter* conformations (Figure S9C), may play a role in the kinetics of conversion to the *Nter+* active state. Such asymmetry of TM helix dimers was also recently discussed for other RTKs.<sup>27,40</sup> Taken together these simulations support the picture suggested by, e.g., NMR studies<sup>24</sup> of a number of different structures of the TM dimer, interconversion between which may provide pathways for activation of the EGFR.

**Influence of Membrane Thickness.** For a number of TM helix dimers, the lipid environment has been suggested to modulate the stability of specific conformations.<sup>25,37,83</sup> Our CG-MetaD simulations have allowed us to identify *Nter+* as the most stable dimeric state. We therefore simulated the dimeric states in membranes differing in their bilayer thickness (Supporting Information Figure S10) to investigate the effect of bilayer thickness on the conversion pathways from the higher energy states (*Cter+*, *Nter*, *Cter*, *Lzip*) to this stable conformation. Overall, these simulations confirm the results obtained in the 16:0 PC bilayer used in the majority of the simulations. However, some interesting differences were found. In particular, in a thinner (12:0 PC) bilayer the amplitude and



**Figure 7.** Influence of the bilayer thickness (as shown schematically in the center panels) on the kinetics of (A) pivot, and (B) rotational mechanisms. In (A)  $\Delta d$  is monitored during CG-MD simulations starting from the *Cter+* conformation in either 12:0 PC or 20:0 PC bilayers. In (B)  $\Delta|\theta|$  is monitored during CG-MD simulations starting from the *Lzip* conformation in either 12:0 PC or 20:0 PC bilayers.

the frequency of the pivot motions were increased, whereas these appeared to be slowed in a thicker (20:0 PC) bilayer (Figure 7A). Similar behavior was observed regarding the rotational motions, even leading in some case to the absence of interconversion in the course of simulations (Figure 7B). These findings support the view that the bilayer thickness can modulate TM helix dimer stability,<sup>37,38,76</sup> and that the physical properties of plasma membrane (nano)domains may be able to modulate receptor activation.<sup>2,84,85</sup> Indeed, in future studies it would be of interest to explore the dependence of free energy landscape on more complex lipid bilayer compositions, including interactions with, e.g., cholesterol and phosphatidylinositol 4,5-bisphosphate such as have been observed in, e.g., simulations with GPCRs.<sup>86</sup>

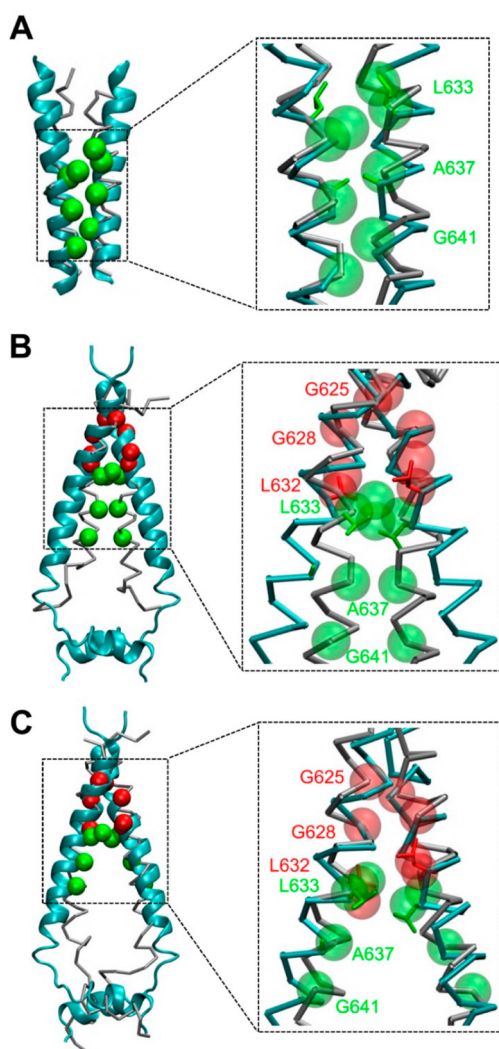
**Structural Correlations.** The structure of the TM domain of the EGFR has been determined by NMR in DPC micelles (pdb id: 2M0B).<sup>3</sup> It shows interhelix contacts at the C-terminal small- $X_3$ -small motif similar to those found in the *Cter+* population identified in our simulations (Figure 8A). Furthermore, we note that the structure of the ErbB1/B2 TM helix dimer (pdb id: 2KS1) obtained by NMR in DMPC/DHPC bicelles resembles more closely our *Nter+* ensemble.<sup>51</sup> Another NMR structure of the EGFR TM domain plus the adjacent intracellular juxtamembrane (JM) region (pdb id: 2M20) also revealed interhelix contacts at the N-terminal small- $X_3$ -small motif within the TM domain.<sup>23</sup> This dimer structure is very similar to the *Nter+* ensemble identified as the population of lowest free-energy in our CG-MetaD calculations (Figure 8B). This agreement prompted us to undertake a series of 5  $\mu$ s standard CG-MD simulations using the TM+JM NMR structure (pdb id: 2M20) as a starting conformation. In these simulations, we found that the experimental conformation is stable, displaying contacts between the TM helices at the N-terminal motifs with additional hydrophobic leucine-zipper interactions between the JM segments.<sup>23</sup> Encouragingly, the coarse-grained model reproduced the global structure and especially the interhelical contacts determined by NMR (Figure 8C).

**Correlated Motions of the TM and JM Domains.** We also simulated the behavior of the NMR TM+JM dimer using CG-MD in a phospholipid bilayer containing anionic lipids (POPC:POPG 85:15) in order to mimic the environment presented by a mammalian plasma membrane.<sup>32,87</sup> After  $\sim 2.7$   $\mu$ s a conformational transition was observed from the *Nter+* to

the *Cter+* population (Supporting Information Figure S11A). This change occurred through a pivot motion of the TM domain and was associated with the separation of the JM segments (Figure 9A). Interestingly a recent simulation study suggests that a similar motion is associated with an inhibitory effect of binding of the Herceptin monoclonal antibody to the Her2 receptor.<sup>33</sup> In another set of CG-MD simulations of the TM+JM dimer, this time performed in a zwitterionic (PC) bilayer and starting from the *Nter+* conformation, we observed a rotational motion of the helices that led to the *Lzip* conformation, before returning to the initial *Nter+* conformation (Supporting Information Figure S11B). Similarly to what was observed for the TM domain, the transition between the right-handed and the left-handed conformations is mediated by an asymmetric *Nter* conformation (Supporting Information Figure S11C). Structural characterization also shows that the JM leucine-zipper interactions are disrupted during this process and reform at the end of the simulation (Figure 9B). Extrapolating back to the full-length receptor, reformation of the antiparallel leucine-zipper interactions in the JM region would be expected to latch the kinase domains in the asymmetric active conformation.<sup>22</sup> Thus, our CG-MD simulations reveal reversible transitions between conformations, which may be identified as corresponding to the likely active and inactive states of the receptor, providing structural insights into the dynamics of EGFR activation.

**Mechanisms of EGFR Activation.** Overall, our simulations suggest that when the TM helices are in contact through the C-terminal small- $X_3$ -small or *Lzip* motifs, the JM segments are distant and cannot readily form antiparallel leucine-zipper interactions (Supporting Information Figure S11). This further suggests that the *Lzip*, *Cter* and *Cter+* conformations all correspond to inactive states of the receptor. We note that all these inactive dimer states exhibit an increased distance between the two K618 residues in the N-loop that have been previously identified as critical for regulation of EGFR activation by glycolipids such as GM3.<sup>69</sup> We also note that such variation of distance between charged residues in this specific upper region of the TM domain was found sufficient to regulate activation of the growth hormone receptor.<sup>30</sup> Thus, our findings provide a description of the possible role of the TM domain in relaying conformational changes between the extracellular and the intracellular regions of the receptor (Supporting Information Figure S11). We suggest that the



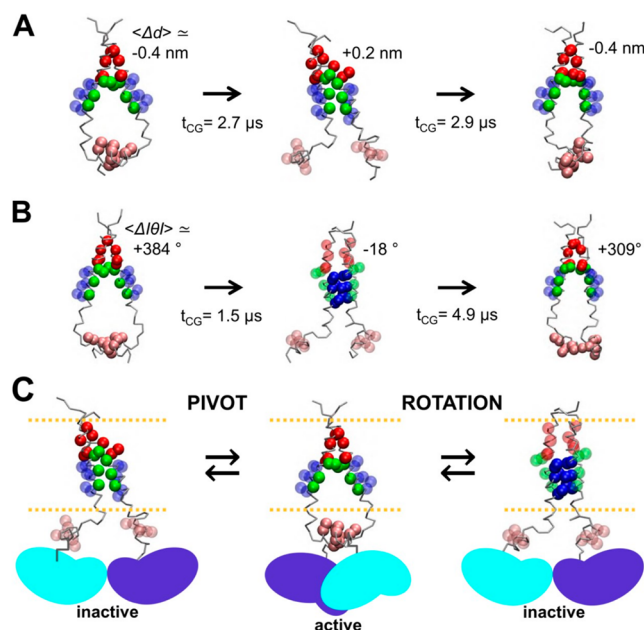


**Figure 8.** (A) Superposition of the NMR structure of the TM domain (pdb id: 2M0B) and a representative structure of the *Cter*+ ensemble. (B, C) Superposition of the NMR structure of the TM+JM dimer (pdb id: 2M20) with representative structures of (B) the *Nter*+ population; and (C) CG-MD simulation #1 in a 16:0 PC bilayer (see Supporting Information Figure S11A). NMR structures are in cyan, and the backbone of CG models are shown as gray traces. Red and green spheres respectively represent the *Nter* and *Cter* motifs of interhelical contact in the TM domain of CG models. In the insets the helix backbones of the NMR structures are shown as cyan traces.

EGFR can be activated either through a “pivot” motion of the TM helices and/or through an alternative “rotational” mechanism, both leading to the right-handed *Nter*+ active dimeric state (Figure 9C).

## CONCLUSIONS

Large-scale motions of TM domains play a key role in the activation mechanisms of membrane receptors such as RTKs. These functional motion of TM helices can be difficult to characterize by either experimental or theoretical approaches. Here, we have described an advanced computational protocol that combines two well-established techniques, namely Coarse-Grained molecular dynamics<sup>88</sup> and well-tempered MetaDynamics (CG-MetaD),<sup>45</sup> allowing us to overcome the time scale limits of current simulations. Using this approach, we have simulated helix/helix interactions of the EGFR TM domain to



**Figure 9.** (A) Structural summary of the dynamics observed during a CG-MD simulation starting from the NMR structure of the TM+JM dimer (pdb id: 2M20) in a bilayer containing anionic lipids (POPC:POPG 85:15), indicating pivot motion and average values of  $\Delta d$  for each population. (B) Structural summary of the dynamics observed during a simulation in a 16:0 PC lipid bilayer, indicating rotation events of individual monomer and average value of  $\Delta|\theta|$  for each population. Residues from motifs involved in direct interhelix contacts are drawn as opaque spheres, while motif residues not engaged in direct contacts are shown as transparent spheres. Red, green and blue spheres represent *Nter*, *Cter* and *Lzip* motifs of the TM domain, and leucine zipper residues in the JM domain are in pink. (C) Suggested mechanisms of EGFR activation via either pivot or rotational motions. Yellow broken lines represent the lipid head groups. Kinase domains of inactive and active states are shown schematically with symmetric and asymmetric states, respectively, in cyan and blue.

reconstruct the underlying free energy landscape, providing a detailed description of the conformational motions of this canonical RTK TM dimer at near-atomic resolution. In particular, the enhancement of the phase space exploration using CG-MetaD allows us to describe states and transitions between states that would be otherwise only partially described by more standard approaches. Running large ensembles of multiple standard CG-MD simulations with different initial configurations<sup>89</sup> would be an alternative approach but requires *a priori* structural knowledge on starting configurations that is achievable only by a thorough sampling such as is possible by CG-MetaD.

Our study reveals the various modes of interaction between EGFR TM helices. The agreement between the simulated packing modes and the available structural data supports our thermodynamic characterization. Our calculations predict the existence of five dimeric states of the EGFR TM domain, which are separated from one another by less than 10 kJ/mol (i.e.,  $\sim 4$  kT), and for which the relatively fast kinetics of interchange may be modulated by the thickness of phospholipid bilayers (Figures 6 and 7). These features underlie considerable conformational flexibility of the TM domain, supporting its likely role in the activation mechanism of the EGFR. We have described two mechanisms leading to the population of lowest

free-energy,  $N_{ter+}$ , corresponding to the activated state of the whole receptor. The first involves a “pivoting” motion of the helices that converts the  $C_{ter+}$  (inactive) population in the  $N_{ter+}$  (active) population, while the second is a “rotational” mechanism that leads from a left-handed  $L_{zip}$  (inactive) population to the right-handed  $N_{ter+}$  (active) population. Simulations in the presence of the juxtamembrane (JM) domain confirm this scenario and demonstrate the likely critical role of the TM domain in the allosteric regulation of the EGFR (Figure 9). Thus, we are able to fully describe functionally relevant motions of the TM domain “switch” which links the ecto and intracellular domains within the intact receptor dimer.

It is of course worthwhile to consider possible limitations of the CG model (i.e., MARTINI) employed in these studies.<sup>36</sup> The simplifications of the molecular systems include a mapping of four non-hydrogen atoms into one single CG particle, as well as the short-range nature of the interactions between particles. These features lead to smoothed free energy landscape, which intrinsically modifies the kinetics of the system.<sup>55</sup> However, the enthalpy-driven parametrization of the MARTINI force field potential aids prediction of energetically relevant states, including intermediate states.<sup>36</sup> It has been demonstrated that the CG force field reproduces physicochemical properties of solvated phospholipid bilayers<sup>55,90</sup> and describes protein–lipid<sup>84,91–94</sup> and protein–protein<sup>95</sup> interactions. To date, MARTINI has successfully complemented experimental investigations of various TM protein domains.<sup>30,83,96,97</sup> In particular, the in-house variant used in the present study has previously shown quantitative agreement with experimental measures on a number of TM systems.<sup>81,98,99</sup> Furthermore, comparisons of PMFs for helix/helix interactions of glycoporphin A derived from MARTINI v2.1 vs v2.2 suggest that the overall well depth changes by less than kT (unpublished observations). Although the standard formulations of MARTINI present limitations that may affect the dynamics of the system, such as the lack of solvent polarizability and entropic components,<sup>36,100</sup> the correlation between our results and available experimental data suggests that this CG force field is able to predict thermodynamically relevant structural properties (Figure 8).

Using the CG-MetaD approach we have addressed the challenging problem<sup>89</sup> of characterizing accurately the free energy landscape of protein/protein interactions within membranes. The association and dissociation of the TM helices of EGFR in a phospholipid bilayer environment are described by a large number of interconversion events between the dimeric and monomeric states (see Movie S1). Through the accelerated sampling of the slowest motions of the system, CG-MetaD simulations allow us to overcome the time scale limits of current simulations. Estimating effective rates of the observed events will be important for future developments of the CG-MetaD methodology. A protocol has been recently proposed to extract transition rates from atomistic metadynamics simulations.<sup>101,102</sup> However, in the present case the kinetic calculations are rather complex since the sampling acceleration observed with the CG model along the CV space needs to be rigorously evaluated.

By enabling us to go beyond the time scale and sampling of current simulations, our CG-MetaD studies alongside standard equilibrium CG and AT simulations reveal new structural information which complements those obtained previously. Our results demonstrate that the combination of CG and MetaD provides an exhaustive exploration of the phase space

leading to an accurate description of the free energy landscape underlying the association of the EGFR TM helices. This allows identification of new dimeric states and characterization of multiple, reversible interconversion events between these states of the TM domain dimer of the EGFR. This contributes to our understanding of the mechanisms of EGFR signaling by helping to bridge the gap between the extracellular and intracellular regions. The intrinsic functional motions of the TM domain will reciprocally modulate and be modulated by additional molecular interactions in the structurally intact receptor *in vivo*.<sup>103</sup> Future investigations will need to analyze the possible effects of more complex membrane environments<sup>86</sup> and to probe the similarities and/or differences among homologous RTKs.<sup>21</sup> This study endorses CG-MetaD as a powerful method that opens new avenues in free energy calculations of biomolecules<sup>104–106</sup> and in the investigation of large-scale conformational motions underlying the activation mechanism of RTKs and related membrane proteins.

## ■ ASSOCIATED CONTENT

### 📎 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05602.

Movie S1 (AVI)

Tables with details of simulation setups, additional analysis, text and figures (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*vittoriolomongelli@gmail.com

\*mark.sansom@bioch.ox.ac.uk

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by grants from the Wellcome Trust and the BBSRC (to MS), and from the Swiss National Supercomputing Center (CSCS under project ID s451), the Italian MIUR-PRIN 2010/2011 (E61J12000210001), the Swiss National Science Foundation (Project N. 200021\_163281) and the COST action CA15135 (Multitarget paradigm for innovative ligand identification in the drug discovery process MuTaLig) (to VL). The authors thank Prof. Michele Parrinello as well as many other colleagues for fruitful discussions. In particular we thank Dr. Khairul Bariyyah Abd Halim for the homology model of the intact EGFR.

## ■ REFERENCES

- (1) Bessman, N. J.; Freed, D. M.; Lemmon, M. A. *Curr. Opin. Struct. Biol.* **2014**, *29C*, 95–101.
- (2) Bocharov, E. V.; Mayzel, M. L.; Volynsky, P. E.; Mineev, K. S.; Tkach, E. N.; Ermolyuk, Y. S.; Schulga, A. A.; Efremov, R. G.; Arseniev, A. S. *Biophys. J.* **2010**, *98*, 881–889.
- (3) Bocharov, E. V.; Lesovoy, D. M.; Pavlov, K. V.; Pustovalova, Y. E.; Bocharova, O. V.; Arseniev, A. S. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 1254–1261.
- (4) Lemmon, M. A.; Schlessinger, J. *Cell* **2010**, *141*, 1117–1134.
- (5) Li, E.; Hristova, K. *Biochemistry* **2006**, *45*, 6241–6251.
- (6) Bublil, E. M.; Yarden, Y. *Curr. Opin. Cell Biol.* **2007**, *19*, 124–134.
- (7) Hynes, N. E.; Lane, H. A. *Nat. Rev. Cancer* **2005**, *5*, 341–354.
- (8) Tebbutt, N.; Pedersen, M. W.; Johns, T. G. *Nat. Rev. Cancer* **2013**, *13*, 663–73.
- (9) Finger, C.; Escher, C.; Schneider, D. *Sci. Signaling* **2009**, *2*, 1–8.

- (10) Groves, J. T.; Kuriyan, J. *Nat. Struct. Mol. Biol.* **2010**, *17*, 659–665.
- (11) He, L.; Hristova, K. *Sci. Rep.* **2012**, *2*, 854.
- (12) Bennisroune, A.; Fickova, M.; Gardin, A.; Dirrig-Grosch, S.; Aunis, D.; Cremel, G.; Hubert, P. *Mol. Biol. Cell* **2004**, *15*, 3464–74.
- (13) Yin, H.; Slusky, J. S.; Berger, B. W.; Walters, R. S.; Vilaire, G.; Litvinov, R. I.; Lear, J. D.; Caputo, G. A.; Bennett, J. S.; DeGrado, W. F. *Science* **2007**, *315*, 1817–1822.
- (14) Sinclair, J. K.; Denton, E. V.; Schepartz, A. *J. Am. Chem. Soc.* **2014**, *136*, 11232–5.
- (15) Ferguson, K. M. *Annu. Rev. Biophys.* **2008**, *37*, 353–373.
- (16) Landau, M.; Ben-Tal, N. *Biochim. Biophys. Acta, Rev. Cancer* **2008**, *1785*, 12–31.
- (17) Liu, P.; Cleveland, T.E.T.; Bouyain, S.; Byrne, P. O.; Longo, P. A.; Leahy, D. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10861–6.
- (18) Linggi, B.; Carpenter, G. *Trends Cell Biol.* **2006**, *16*, 649–56.
- (19) Mi, L. Z.; Lu, C.; Li, Z.; Nishida, N.; Walz, T.; Springer, T. A. *Nat. Struct. Mol. Biol.* **2011**, *18*, 984–9.
- (20) Cymer, F.; Schneider, D. *Cell Adh. Migr.* **2010**, *4*, 299–312.
- (21) Hedger, G.; Sansom, M. S. P.; Koldsø, H. *Sci. Rep.* **2015**, *5*, 9198.
- (22) Jura, N.; Endres, N. F.; Engel, K.; Deindl, S.; Das, R.; Lamers, M. H.; Wemmer, D. E.; Zhang, X. W.; Kuriyan, J. *Cell* **2009**, *137*, 1293–1307.
- (23) Endres, N. F.; Das, R.; Smith, A. W.; Arkhipov, A.; Kovacs, E.; Huang, Y. J.; Pelton, J. G.; Shan, Y. B.; Shaw, D. E.; Wemmer, D. E.; Groves, J. T.; Kuriyan, J. *Cell* **2013**, *152*, 543–556.
- (24) Bragin, P.; Mineev, K.; Bocharova, O.; Volynsky, P.; Bocharov, E.; Arseniev, A. J. *Mol. Biol.* **2016**, *428*, 52–61.
- (25) Matthews, E.; Zoonens, M.; Engelman, D. *Cell* **2006**, *127*, 447–450.
- (26) Cymer, F.; Veerappan, A.; Schneider, D. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 963–973.
- (27) Volynsky, P. E.; Polyansky, A. A.; Fakhruddinova, G. N.; Bocharov, E.; Efremov, R. G. *J. Am. Chem. Soc.* **2013**, *135*, 8105–8108.
- (28) Fleishman, S. J.; Schlessinger, J.; Ben-Tal, N. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15937–40.
- (29) Polyansky, A. A.; Volynsky, P. E.; Efremov, R. G. *J. Am. Chem. Soc.* **2012**, *134*, 14390–14400.
- (30) Brooks, A. J.; Dai, W.; O'Mara, M. L.; Abankwa, D.; Chhabra, Y.; Pelekanos, R. A.; Gardon, O.; Tunny, K. A.; Blucher, K. M.; Morton, C. J.; Parker, M. W.; Sierecki, E.; Gambin, Y.; Gomez, G. A.; Alexandrov, K.; Wilson, I. A.; Doxastakis, M.; Mark, A. E.; Waters, M. J. *Science* **2014**, *344*, 1249783.
- (31) Kastner, J.; Loeffler, H. H.; Roberts, S. K.; Martin-Fernandez, M. L.; Winn, M. D. *J. Struct. Biol.* **2009**, *167*, 117–28.
- (32) Arkhipov, A.; Shan, Y.; Das, R.; Endres, N. F.; Eastwood, M. P.; Wemmer, D. E.; Kuriyan, J.; Shaw, D. E. *Cell* **2013**, *152*, 557–569.
- (33) Franco-Gonzalez, J. F.; Ramos, J.; Cruz, V. L.; Martinez-Salazar, J. J. *Comput.-Aided Mol. Des.* **2014**, *28*, 1093–1107.
- (34) Psachoulia, E.; Marshall, D.; Sansom, M. S. P. *Acc. Chem. Res.* **2010**, *43*, 388–396.
- (35) Benjamini, A.; Smit, B. *Biophys. J.* **2012**, *103*, 1227–1235.
- (36) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.
- (37) Janosi, L.; Prakash, A.; Doxastakis, M. *Biophys. J.* **2010**, *99*, 284–292.
- (38) Sengupta, D.; Marrink, S. J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12987–12996.
- (39) Prakash, A.; Janosi, L.; Doxastakis, M. *Biophys. J.* **2010**, *99*, 3657–3665.
- (40) Chavent, M.; Chetwynd, A. P.; Stansfeld, P. J.; Sansom, M. S. *Biochemistry* **2014**, *53*, 6641–6652.
- (41) Chipot, C.; Pohorille, A. *Free Energy Calculations Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007.
- (42) Grant, B. J.; Gorfie, A. A.; McCammon, J. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142–7.
- (43) Abrams, C.; Bussi, G. *Entropy* **2014**, *16*, 163–199.
- (44) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
- (45) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (46) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. *J. Am. Chem. Soc.* **2012**, *134*, 453–63.
- (47) Limongelli, V.; Bonomi, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.
- (48) Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V. *Nucleic Acids Res.* **2014**, *42*, 5447–5455.
- (49) Johnston, J. M.; Aburi, M.; Provasi, D.; Bortolato, A.; Urizar, E.; Lambert, N. A.; Javitch, J. A.; Filizola, M. *Biochemistry* **2010**, *50*, 1682–1690.
- (50) Cossio, P.; Marinelli, F.; Laio, A.; Pietrucci, F. *J. Phys. Chem. B* **2010**, *114*, 3259–65.
- (51) Mineev, K. S.; Bocharov, E. V.; Pustovalova, Y. E.; Bocharova, O. V.; Chupin, V. V.; Arseniev, A. S. *J. Mol. Biol.* **2010**, *400*, 231–243.
- (52) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (53) Bond, P. J.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2006**, *128*, 2697–2704.
- (54) Bond, P. J.; Wee, C. L.; Sansom, M. S. P. *Biochemistry* **2008**, *47*, 11321–11331.
- (55) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (56) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (57) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (58) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.
- (59) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (60) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (61) Stansfeld, P. J.; Sansom, M. S. P. *J. Chem. Theory Comput.* **2011**, *7*, 1157–1166.
- (62) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (63) Dickson, C. J.; Madej, B. D.; Skjerve, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. *J. Chem. Theory Comput.* **2014**, *10*, 865–879.
- (64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (65) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, I. T.E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*, University of California: San Francisco, 2014.
- (66) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (67) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (68) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (69) Coskun, Ü.; Grzybek, M.; Drechsel, D.; Simons, K. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9044–9048.
- (70) Hedger, G.; Koldsø, H.; Sansom, M. S. P. *J. Phys. Chem. B* **2016**, DOI: 10.1021/acs.jpcc.6b01387.

- (71) Michailidis, I. E.; Rusinova, R.; Georgakopoulos, A.; Chen, Y.; Iyengar, R.; Robakis, N. K.; Logothetis, D. E.; Baki, L. *Pfluegers Arch.* **2011**, *461*, 387–397.
- (72) Troussicot, L.; Guillièrre, F.; Limongelli, V.; Walker, O.; Lancelin, J. M. *J. Am. Chem. Soc.* **2015**, *137*, 1273–1281.
- (73) Duneau, J. P.; Vegh, A. P.; Sturgis, J. N. *Biochemistry* **2007**, *46*, 2010–2009.
- (74) Chen, L. R.; Merzlyakov, M.; Cohen, T.; Shai, Y.; Hristova, K. *Biophys. J.* **2009**, *96*, 4622–4630.
- (75) Hénin, J.; Pohorille, A.; Chipot, C. *J. Am. Chem. Soc.* **2005**, *127*, 8478–8484.
- (76) Kuznetsov, A. S.; Polyansky, A. A.; Fleck, M.; Volynsky, P. E.; Efremov, R. G. *J. Chem. Theory Comput.* **2015**, *11*, 4415–4426.
- (77) Russ, W. P.; Engelman, D. M. *J. Mol. Biol.* **2000**, *296*, 911–919.
- (78) Escher, C.; Cymer, F.; Schneider, D. *J. Mol. Biol.* **2009**, *389*, 10–16.
- (79) Holt, A.; Koehorst, R. B.; Rutters-Meijneke, T.; Gelb, M. H.; Rijkers, D. T.; Hemminga, M. A.; Killian, J. A. *Biophys. J.* **2009**, *97*, 2258–2266.
- (80) Kim, T.; Im, W. *Biophys. J.* **2010**, *99*, 175–183.
- (81) Vostrikov, V. V.; Hall, B. A.; Greathouse, D. V.; Koeppe, R. E.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2010**, *132*, 5803–5811.
- (82) Moriki, T.; Maruyama, H.; Maruyama, I. N. *J. Mol. Biol.* **2001**, *311*, 1011–26.
- (83) Dominguez, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. *J. Am. Chem. Soc.* **2014**, *136*, 854–857.
- (84) Muhle-Goll, C.; Hoffmann, S.; Afonin, S.; Grage, S.; Polyansky, A.; Windisch, D.; Zeitler, M.; Bürck, J.; Ulrich, A. *J. Biol. Chem.* **2012**, *287*, 26178–26186.
- (85) Ruysschaert, J. M.; Loney, C. *Biochim. Biophys. Acta, Biomembr.* **2015**, *1848*, 1860–1867.
- (86) Koldso, H.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2015**, *137*, 14694–14704.
- (87) Abd Halim, K. B.; Koldso, H.; Sansom, M. S. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1017–25.
- (88) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press, 2008.
- (89) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. *J. Chem. Theory Comput.* **2015**, *11*, 2278–2291.
- (90) Vanni, S.; Hirose, H.; Barelli, H.; Antonny, B.; Gautier, R. *Nat. Commun.* **2014**, *5*, 4916.
- (91) van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willig, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmüller, H.; Diederichsen, U.; Jahn, R. *Nature* **2011**, *479*, 552–555.
- (92) Arnarez, C.; Marrink, S. J.; Periole, X. *Sci. Rep.* **2013**, *3*, 1263.
- (93) Arnarez, C.; Mazat, J.-P.; Elezgaray, J.; Marrink, S.-J.; Periole, X. *J. Am. Chem. Soc.* **2013**, *135*, 3112–3120.
- (94) Stansfeld, P. J.; Goose, J. E.; Caffrey, M.; Carpenter, E. P.; Parker, J. L.; Newstead, N.; Sansom, M. S. P. *Structure* **2015**, *23*, 1350–1361.
- (95) Rassam, P.; Copeland, N. A.; Birkholz, O.; Toth, C.; Chavent, M.; Duncan, A. L.; Cross, S. J.; Housden, N. G.; Kaminska, R.; Seger, U.; Quinn, D. M.; Garrod, T. J.; Sansom, M. S. P.; Piehler, J.; Baumann, C. G.; Kleanthous, C. *Nature* **2015**, *523*, 333–336.
- (96) Sharma, S.; Juffer, A. H. *J. Am. Chem. Soc.* **2013**, *135*, 2188–97.
- (97) Cybulski, L. E.; Ballering, J.; Moussatova, A.; Inda, M. E.; Vazquez, D. B.; Wassenaar, T. A.; de Mendoza, D.; Tieleman, D. P.; Killian, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6353–8.
- (98) Psachoulia, E.; Bond, P. J.; Fowler, P. W.; Sansom, M. S. P. *Biochemistry* **2008**, *47*, 10503–105012.
- (99) Kalli, A.; Hall, B. A.; Campbell, I. D.; Sansom, M. S. P. *Structure* **2011**, *19*, 1477–1484.
- (100) Wu, Z.; Cui, Q.; Yethiraj, A. *J. Phys. Chem. Lett.* **2011**, *2*, 1794–1798.
- (101) Tiwary, P.; Parrinello, M. *Phys. Rev. Lett.* **2013**, *111*, 230602.
- (102) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386–91.
- (103) Kaszuba, K.; Grzybek, M.; Orlowski, A.; Danne, R.; Rog, T.; Simons, K.; Coskun, U.; Vattulainen, I. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4334–4339.
- (104) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5411–5416.
- (105) Formoso, E.; Limongelli, V.; Parrinello, M. *Sci. Rep.* **2015**, *5*, 8425.
- (106) Comitani, F.; Limongelli, V.; Molteni, C. *J. Chem. Theory Comput.* **2016**, *12*, 3398–3406.