Supplementary Information for:

CONFORMATIONAL CHANGES IN THE EPIDERMAL GROWTH FACTOR RECEPTOR: ROLE OF THE TRANSMEMBRANE DOMAIN INVESTIGATED BY COARSE-GRAINED METADYNAMICS FREE ENERGY CALCULATIONS

Mickael Lelimousin, Vittorio Limongelli, & Mark S. P. Sansom

Table S1. Summary of the simulations described in the manuscript.

Starting conformation	Lipid bilayer	Simulation type	Duration
Homology model	16:0 PC	CG-MetaD	3 x 380 μs
TM helix a	16:0 PC	CG-MD	3 x 0.1 μs
TM helix b	16:0 PC	CG-MD	3 x 0.1 μs
Monomeric states	16:0 PC	CG-MD	3 x 5 μs
Cter+ conformation	16:0 PC	CG-MD	2 x 5 μs
Cter conformation	16:0 PC	CG-MD	2 x 5 μs
Nter conformation	16:0 PC	CG-MD	2 x 9 μs
Lzip conformation	16:0 PC	CG-MD	2 x 7 μs
Cter+ conformation	12:0 PC	CG-MD	2 x 5 μs
Cter conformation	12:0 PC	CG-MD	2 x 5 μs
Nter conformation	12:0 PC	CG-MD	2 x 5 μs
Lzip conformation	12:0 PC	CG-MD	2 x 5 μs
Cter+ conformation	20:0 PC	CG-MD	2 x 5 μs
Cter conformation	20:0 PC	CG-MD	2 x 5 μs
Nter conformation	20:0 PC	CG-MD	2 x 5 μs
Lzip conformation	20:0 PC	CG-MD	2 x 5 μs
2M20 structure	12:0 PC	CG-MD	2 x 5 μs
2M20 structure	16:0 PC	CG-MD	2 x 5 μs
2M20 structure	20:0 PC	CG-MD	2 x 5 μs
2M20 structure	POPC/POPG	CG-MD	2 x 5 μs
Nter+ conformation	16:0 PC	AT-MD	1 x 0.5 μs
Cter+ conformation	16:0 PC	AT-MD	1 x 0.5 μs
Cter conformation	16:0 PC	AT-MD	1 x 0.5 μs
Nter conformation	16:0 PC	AT-MD	1 x 0.5 μs
Lzip conformation	16:0 PC	AT-MD	1 x 0.5 μs
TM helix (monomer)	16:0 PC	AT-MD	1 x 0.5 μs

Table S2. Relative free energy and CV coordinates characterizing the population basins identified on the multidimensional free energy landscape of EGFR TM domain. Average CV values are reported when low standard deviations were observed, while intervals of values reflect large fluctuations between extrema.

	Dimer populations				
	Nter+	*Cter+*	*Cter*	*Nter*	*Lzip*
F (kJ.mol^{-1})	0	1.4	6.4	7.2	11.5
L (nm)	0.64	0.84	0.90	0.91	0.97
Ω (°)	-30	-23	-24	[-50 : +50]	+12
Δd (nm)	-0.10	+0.22	+0.67	-0.50	[-1.00 : +0.69]
$\Delta\lvert\theta\rvert$ (°)	+443	+448	+485	+205	-5

SI text - Atomistic structure of the dimeric states. The structural stability of each of the five most stable dimer populations (Fig. 3D) identified by CG-MetaD calculations was investigated through over 0.5 μs MD simulations. We found that all the dimer populations are stable during the simulations (Fig. S6) as confirmed by the low average rmsd values (below 0.1 nm) computed for the Cα atoms of the TM alpha helices hereafter labeled "a" and "b". Each dimer structure is described at atomistic resolution in the following paragraphs.

***Nter*+ dimer.** This is the lowest free-energy population showing the two helices in the right-handed dimer conformation. This state is stabilized throughout the simulation by a series of hydrophobic and polar interhelical contacts at N-terminus. In particular, Ser^{621a} establishes a H-bond network with Ser^{621b} and Thr^{624b}, while the side chain of Thr^{624a} H-bonds with the backbone amide group of Ser^{621b}. In addition to these polar interactions, the two helices engage a number of hydrophobic contacts from the *Nter* motif up to the helix core (G^{625}xxGA^{629}xxLL^{633}xxxA^{637}), as also observed in the CG-MetaD simulations. All these interactions favor a close contact between the helices stabilizing the right-handed orientation of the dimer (Fig. 4). At variance with N-ter, the C-terminal residues remain relatively distant during the whole simulation. It is also interesting to note that the lysine residues of N-loop (Lys^{618}) interact with the phospholipid polar heads. Similar interactions are also engaged by arginine residues of C-loop. These contacts, together with those previously described, are important in the orientation of the two helices in the membrane bilayer and might play a role in the signal transduction at intracellular level (*1-3*).

***Lzip* dimer.** The left-handed *Lzip* dimer shows interactions different from those observed in the *Nter*+ conformation. In particular, the N-loop is rather flexible although Pro^{620a} can transiently engage H-bonds with Ser^{621b}. Furthermore, Ile^{622a} at N-loop points towards the bilayer core, forming hydrophobic interactions with the aliphatic chains of the phospholipids (Fig. 4). These interactions might play a stabilizing role of the left-handed orientation of this dimer state. At the helix core this population shows helix-helix contacts (V^{635}xxL^{638}xxxL^{642}) very similar to those observed in the CG-MetaD simulations. On the other hand, at intracellular level Arg^{645a} forms a stable H-bond network with residues of monomer "b" such as Gly^{641b}, Leu^{642b} and Met^{644b}. An additional H-bond is present between Arg^{646a} and the carbonyl group of Arg^{646b}. These interactions stabilize a conformation at C-terminus where the C-loop tails remain distant each other during the whole simulation, thus suggesting the inactivity of this dimer state. A similar structural organization at C-ter was also observed in the coarse-grained simulations on the TM+JM structures (Fig. 8).

***Cter*+ dimer.** In this dimer population, both the N-loops and C-loops are rather flexible and no long lasting interaction between the helices is found. On the contrary, at C-terminus stable hydrophobic contacts are present. In particular, Ala^{629a}, Leu^{633a}, Val^{636a}, Ile^{640a} and Met^{644a} interact with Leu^{630b}, Leu^{633b}, Ala^{637b} and Gly^{641b} (Fig. 4). We note that similar inter-helical interactions ($A^{629}xxxL^{633}xxxA^{637}xxxG^{641}$) are found in the CG-MetaD simulation. The arginine residues at C-loop might play an important role in stabilizing this state. In fact, despite the flexibility of the C-loop, Arg^{646a} is able to form H-bonds during the simulation with the carbonyl group of either Arg^{647b} or His^{648b}, thus favoring close contacts between the C-terminus of the two helices.

***Nter* dimer.** In this population the monomer "b" assumes a slightly more extended alpha helix structure with respect to the monomer "a". At variance with *Nter*+, fewer interhelical interactions are found in this state and most of these interactions, such as the hydrophobic contacts formed by Thr^{624a} with Gly^{625b} and by Val^{627a} and Gly^{628a} with Met^{626b} and Ala^{629b}, respectively, are present at N-terminus. Furthermore, a H-bond network is formed at N-loop by the triad Thr^{614a}-Asn^{615a}-Gly^{616a} of monomer "a" with Ile^{619b}, Gly^{616b} and Pro^{617b} of monomer "b", respectively. These interactions contribute to the stability of this dimer conformation (Fig. 4). At C-terminus and C-loops no inter-helical contact is formed. It is also relevant to note that in the *Nter* dimer conformation the orientation of one helix with respect to other is intermediate to those observed in *Nter*+ and *Lzip* (Fig. 4). This finding supports the CG-MetaD results that propose *Nter* as an intermediate state between the inactive left-handed *Lzip* and the active right-handed *Nter*+ state (Fig. 3).

***Cter* dimer.** In this population, at N-loops and N-terminus no interhelical contact is present. On the other hand, interactions between the helices are present at C-terminus ($L^{633}xxxA^{637}xxxG^{641}$) very similar to what found in the CG-MetaD simulation. In particular, Leu^{633a}, Leu^{634a}, Ala^{637a}, Leu^{638a}, Gly^{641a} and Met^{644a} form hydrophobic contacts with Leu^{633b}, Ala^{637b}, Ile^{640b}, Gly^{641b} and Met^{644b}. These interactions are similar to those observed in the *Cter*+ ensemble, however the additional contacts at the core of the helices are here not present. Similarly to what observed in the *Cter*+ state, the arginine residues at C-loops might play a stabilizing role for the dimer state. In fact, Arg^{646a} at C-loop forms a stable H-bond with the carbonyl oxygen of Phe^{643b} (Fig. 4), assisting the interaction between the C-terminus of the two helices.

Fig. S1. Evolution of (A) the lateral separation *L* between helices of EGFR TM domain and of (B) the height of the Gaussians deposited during the well-tempered CG-MetaD simulation. (C) Convergence of the binding free energy calculated as a function of the CG-MetaD simulation time: the solid line reports the free energy difference between the bottom of the well of the dimeric states and the plateau of the monomeric states; the dashed line shows the progressive average of the value from an equilibration time of 110 μs. (D) One dimensional free energy profiles as a function of the lateral helix separation *L*, obtained by integration of the bias, or (E) calculated with an umbrella sampling estimator. The comparison of the profiles D and E shows that the free energy calculations are converged. (F) Free energy profiles of the MFEPs calculated between monomeric and dimeric states in two parts of the Ω-FES above and below the zero vaue of crossing angle Ω, as a function of the lateral helix separation *L*. For both regions MFEPs are drawn in black lines between monomeric states with high *L* values (5 nm) and the free energy minima of the dimeric states, whilst MFEPs between these later points and the dimeric states with lower *L* values are shown in grey lines. (G) Evolution of the MFEP calculated on the Δd-FES, reported every 20 μs from 280 μs (dashed line at the top) to 380 μs (solid line at the bottom).

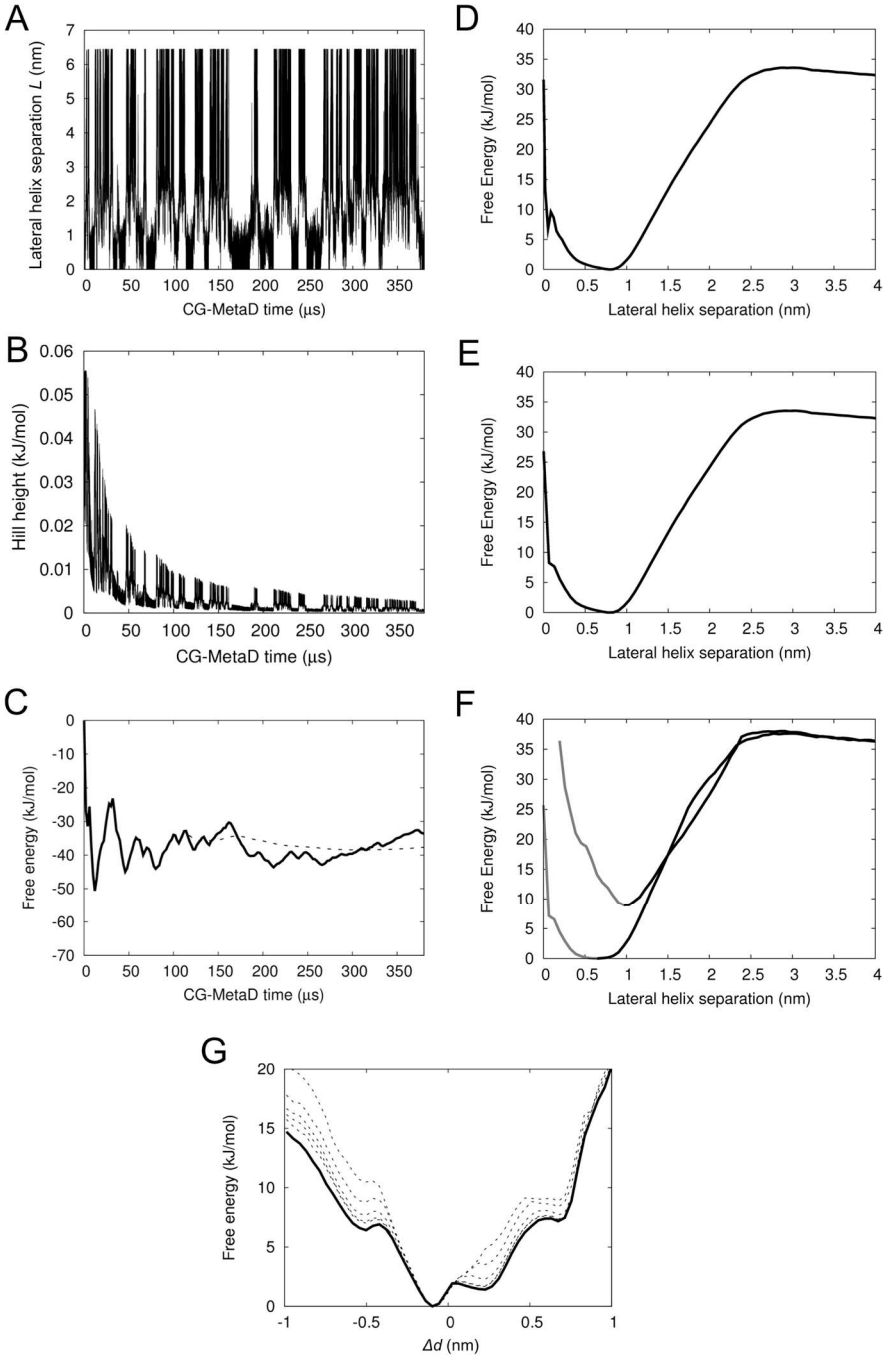


Fig. S2. Evolution of (A) the lateral helix separation L and of (B) the height of the Gaussians deposited during a well-tempered CG-MetaD simulation with identical parameters to those of the main simulation except for the width of the Gaussians fixed here to $\sigma = 0.025$ nm. (C) Convergence of the binding free energy calculated as a function of the CG-MetaD simulation time: the solid line reports the free energy difference between the bottom of the well of the dimeric states and the plateau of the monomeric states; the dashed line shows the progressive average of the value from an equilibration time of 110 μs. (D) One dimensional free energy profile as a function of the lateral helix separation L. (E), (F), (G), (H) show the results obtained from a CG-MetaD simulation with identical parameters to those of the main simulation, but with an additional soft wall potential restraining L below 4.5 nm.

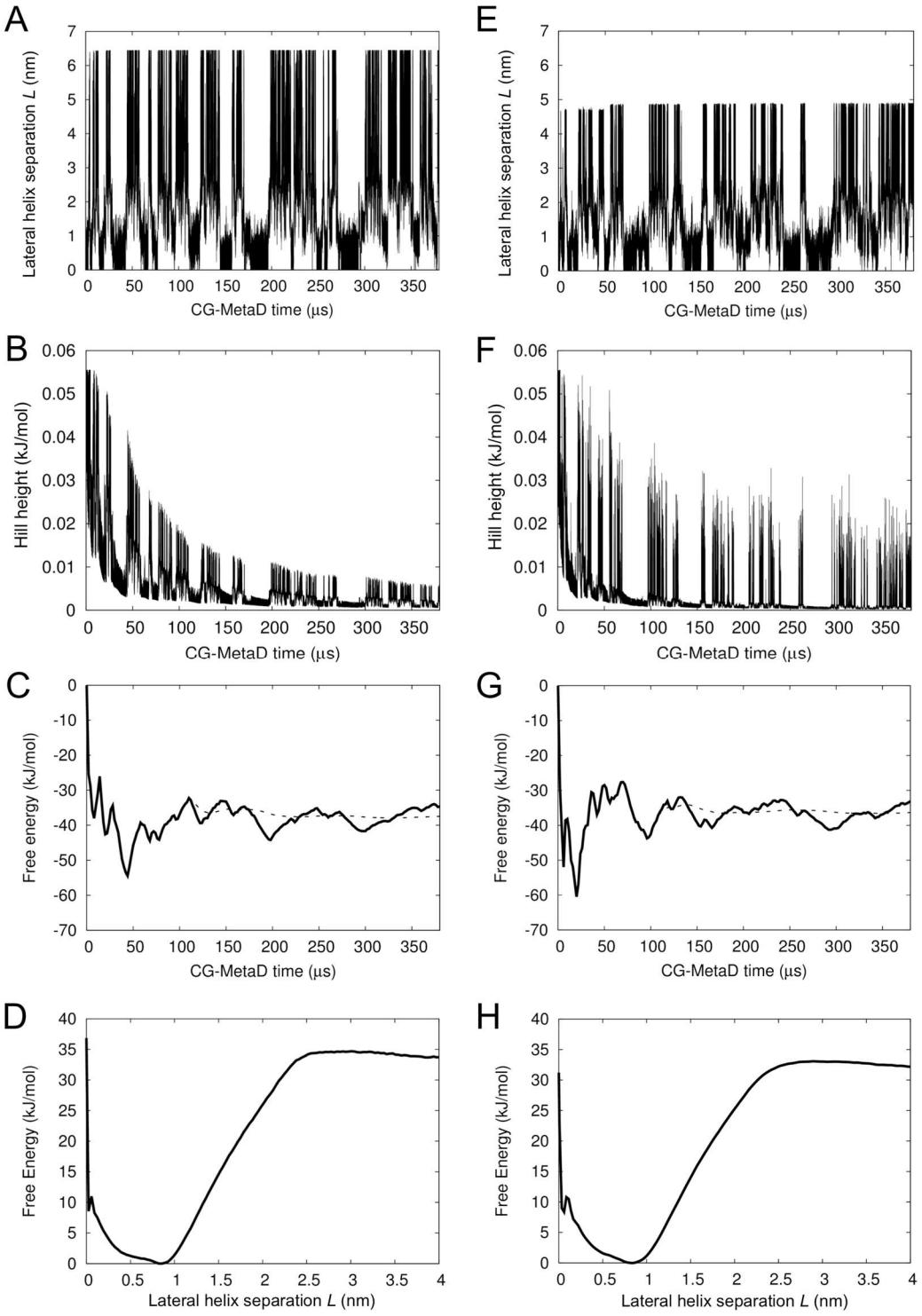


Fig. S3. (A) MFEPs calculated between monomeric and dimeric states on the Ω-FES obtained from a CG-MetaD simulation with identical parameters to those of the main simulation except for the width of the Gaussians fixed here to $\sigma = 0.025$ nm. In the two parts of the FES (above and below the zero value of crossing angle Ω) MFEPs are drawn in black bold lines between monomeric states with high L values (5 nm) and the free energy minima of the dimeric states, whilst MFEPs between these later points and the dimeric states with lower L values are shown in grey bold lines. One dimensional free energy profile as a function of (B) Ω, (C) Δd, (D) $\Delta|\theta|$. Each free energy profile was calculated as the MFEP located at low L values of the associated FES. (E), (F), (G), (H) show the results obtained from a CG-MetaD simulation with identical parameters to those of the main simulation, but with an additional soft wall potential restraining L below 4.5 nm.

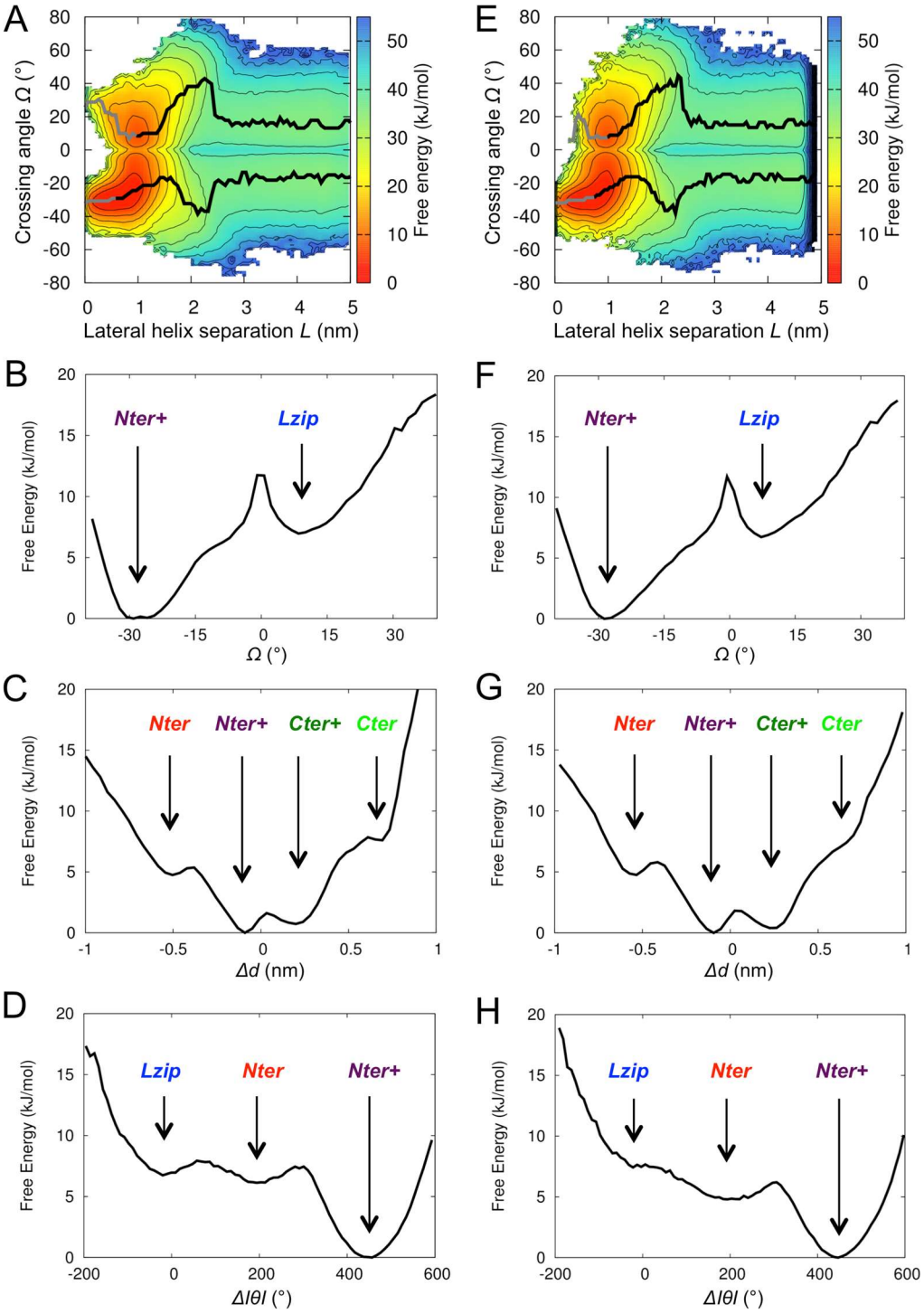


Fig. S4. Free energy profile as a function of the distances between the center of mass of the (A) N-terminal, (B) C-terminal, and (C) *Lzip* association residues defined as the $TG^{625}xxGA^{629}$, $A^{637}xxxG^{641}$, $V^{635}xxL^{638}xxxL^{642}$ motifs, respectively. Each free energy profile was calculated as the MFEP located at low L values of the associated FES. (D) Contact matrices characterizing the population identified in the basins of the multidimensional free energy landscape. Distances between pair of backbone particles (from residues Lys^{618} to Ile^{649}) in the helix dimers were averaged over ensembles of structures taken from the bottom of the free energy basins. Motifs of main residues in contact are provided with the color code associated to populations. (E) Crossing angle Ω as a function of the index of the structures observed in each helix dimer population. (F) Number of interhelical contacts observed in each population. The ensembles of structures were taken from the bottom of the wells identified on the free energy landscape. The numbers of contacts were calculated as a number of residues in interaction within pair of *Nter*, *Cter* and *Lzip* motifs in the dimer. Only symmetric contacts between same residues of identical motifs of the two helices were considered for summation. The number of contacts was normalized to 1 for any pair of residues (*e.g.* contacts between a leucine residue and its homolog in the second helix were divided by 4 to account for the 2 beads of each residue). The properties related to pair of *Nter*, *Cter* and *Lzip* motifs are plotted in red, green and blue lines, respectively. (G) Definition of the $\Delta|\theta|$ orientation CVs. In both side and top views helix backbones are shown in grey tubes, while transparent spheres of red, green and blue colours represent Leu^{632}, Leu^{633} and Val^{635}, respectively. Each dihedral angle is calculated from the sidechain particle of the residue, the backbone particles of the residue and of the residue $i+4$ along the chain, to the backbone particle of the residue $i+4$ in the other helix.

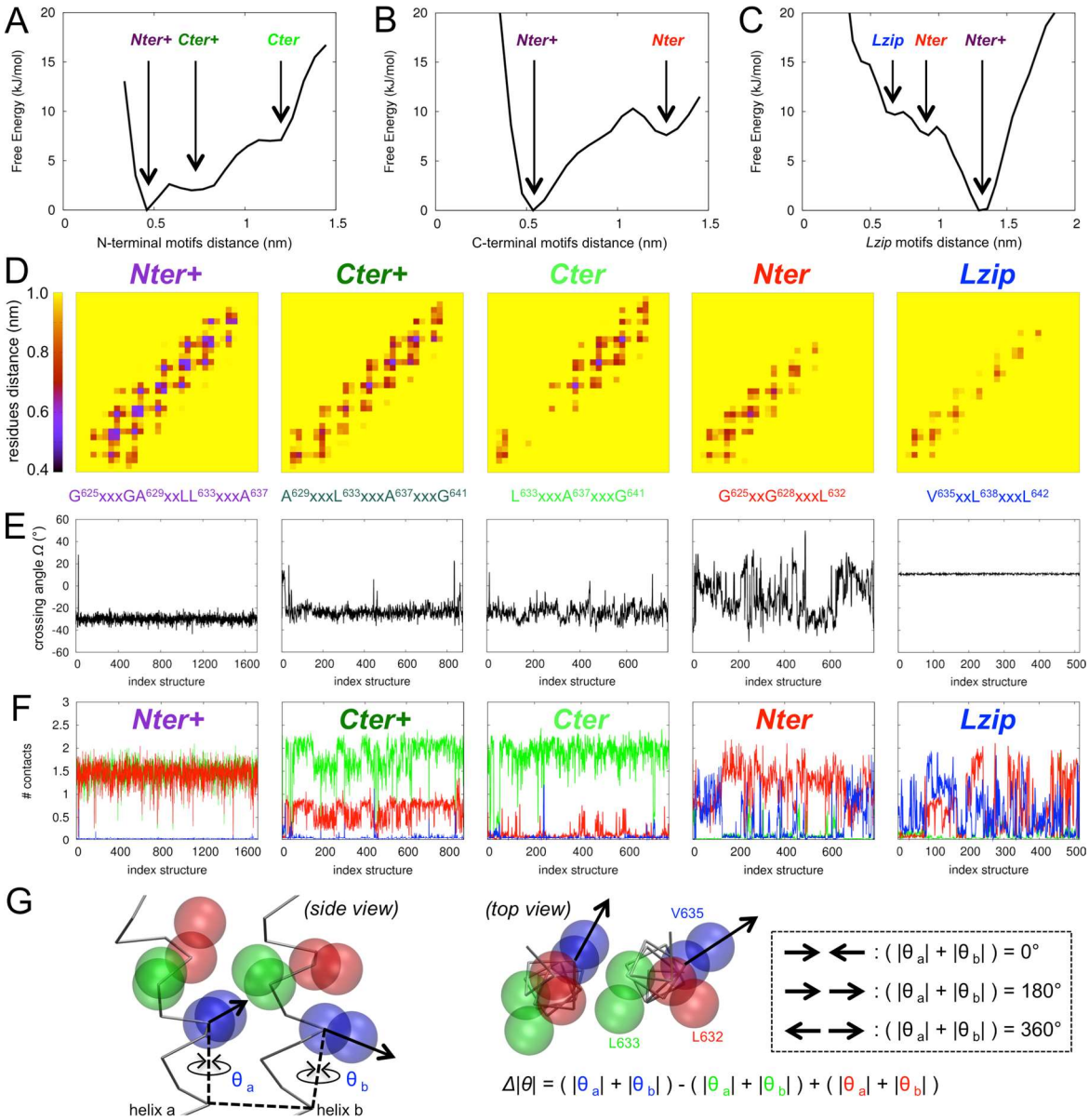


Fig. S5. Comparison of the sequences used in this study and the previous investigations of EGFR TM domain. Residues in bold display differences in length and sequence of the constructs used in the studies, whereas cyan highlights show variations from the native receptor sequence.

This study:	**EGCPTNGPKIP**SIATGMVGALLLLLVVALGIGL**F**M**RRRHIVRKR**
Duneau et al.	**KIP**SIATGMVGALLLLLVVALGIGL**Y**M**RRRH**
Chen et al.	**KKK**SIATGVVGALLLLLVVALGIGL**F**M**KKK**
Prakash et al.	SIATGVVGALLLLLVVALGIGL**F**M**RRR**

Fig. S6. Root mean square deviation (RMSD) observed through AT-MD (atomistic) simulations of the five EGFR dimer populations. The RMSD values were calculated over all the C-alpha of the helices backbone, in respect to the positions observed in the average structure of the simulation.

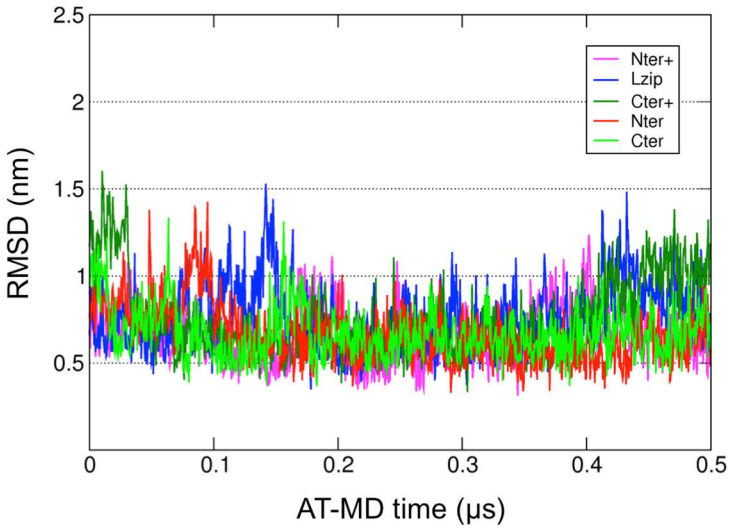


Fig. S7. (A) Structural representation of the tilt of an individual TM helix in the lipid bilayer, as observed in CG simulations. The backbone of the EGFR TM domain is shown in grey, whilst side chains of basic residues and phosphate headgroups of the phospholipids are represented in purple and yellow spheres, respectively. (B) Relative free energy of the monomeric helices, as a function of the tilt angle of the dissociated helices relative to the normal of the bilayer. The red and blue lines show the free energy profiles for the individual tilt of helix a and of helix b, respectively. The solid line and the error bars report the mean values and the standard deviations respectively observed on the tilt angle FES calculated by CG-MetaD at high L values (3 nm < L < 4.5 nm). The red and blue dashed lines show free energy profiles based on the Boltzmann inversion method applied to an ensemble of standard CG-MD simulations of helix a and of helix b, respectively. (C) Tilt angle values of the monomeric helix observed during the AT-MD simulation. (D) Relative free energy of the monomeric states of the TM domain, as a function of the crossing angle Ω measured between the dissociated helices from the CG-MetaD simulation. The solid line and the error bars report the mean values and the standard deviations respectively observed on the Ω-FES calculated by CG-MetaD, for lateral helix separations L comprised between 3 nm and 4.5 nm. (E) *On the left:* Model used to numerically calculate a theoretical probability distribution function of Ω in monomeric states. Two helices show a distribution of tilt and rotation angles relative to local vertical axes. The second helix rotates around the central one. *On the right:* Free energy profile of the monomeric states as a function of Ω, based on the theoretical model. The probability density functions ρ of the tilt angle applied to the two helices in the model were taken as discretized functions of the density functions of the tilt angle α calculated by CG-MetaD (see panel B).

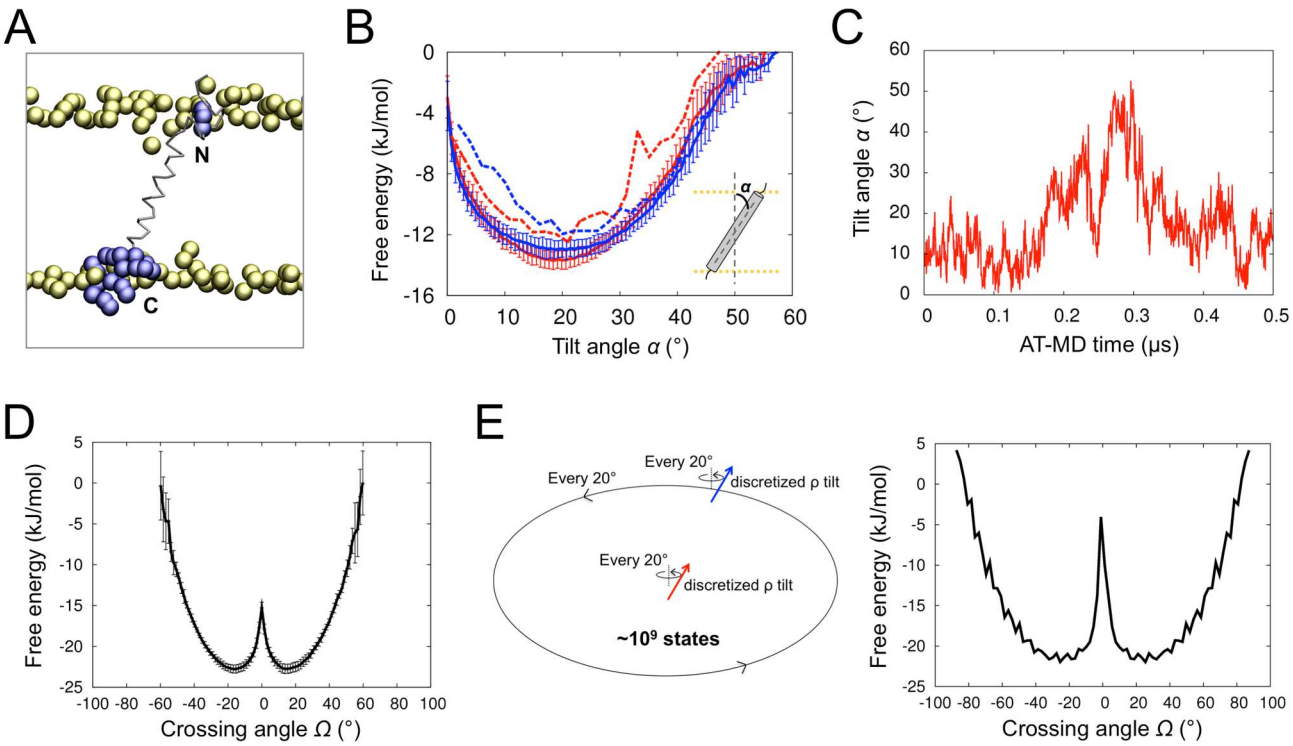


Fig. S8. (A) Region of the Ω-FES explored during 5 µs of standard CG-MD simulation starting from the third randomly chosen monomeric state of coordinates L = 3 nm and Ω = +29 °. Each frame of the simulation is represented by a black dot. (B) Crossing angle CV value as a function of the CG-MD simulation. The dashed vertical lines are coloured in respect to the appearance of specific dimeric states, namely red for *Nter*, blue for *Lzip* and purple for *Nter+*. (C) Structural representation of the changes of dimeric states during the CG-MD simulation. The backbone of TM helices is shown in grey tubes, while N-loop and C-loop are omitted for clarity. The residues from motifs involved in direct interhelical contacts are drawn in opaque spheres, while residues from other motifs are shown in transparent spheres. Colour code is used as previously for representation of the different motifs.

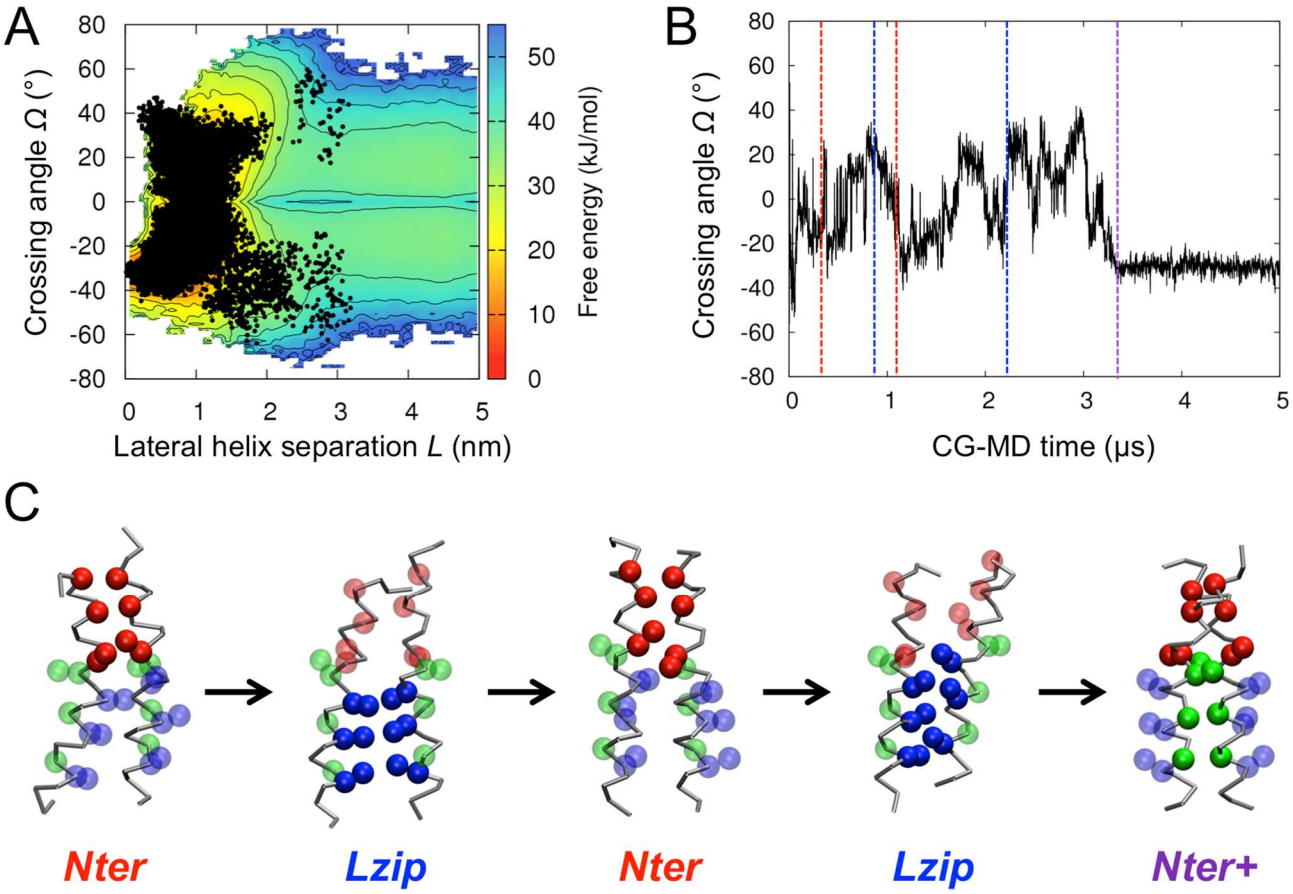


Fig. S9. (A) Number of interhelical contacts observed in the *Nter* (red), *Cter* (green) and *Lzip* (blue) motifs, during two plain CG-MD simulations starting from either the *Cter+* conformation or the *Cter* conformation in a 16:0 PC lipid bilayer. In the bottom panels, the Δd CV is plotted as a function of the simulation time. (B) Number of interhelical contacts observed during two plain CG-MD simulations starting from either the *Nter* conformation or the *Lzip* conformation in a 16:0 PC lipid bilayer. In the bottom panels, the $\Delta|\theta|$ CV is plotted as a function of the simulation time. Second repeats of the four simulations carried out with different initial particle velocities showed similar evolutions. (C) Structural summary of the dynamics observed during the CG-MD simulation starting from the *Lzip* conformation, indicating rotation events of individual helix and average value of $\Delta|\theta|$ for each population. The backbone of TM helices is shown in grey tubes, while N-loop and C-loop are omitted for clarity. The residues from motifs involved in direct interhelical contacts are drawn in opaque spheres, while residues from other motifs are shown in transparent spheres. Colour code is used as previously for representation of the different motifs.

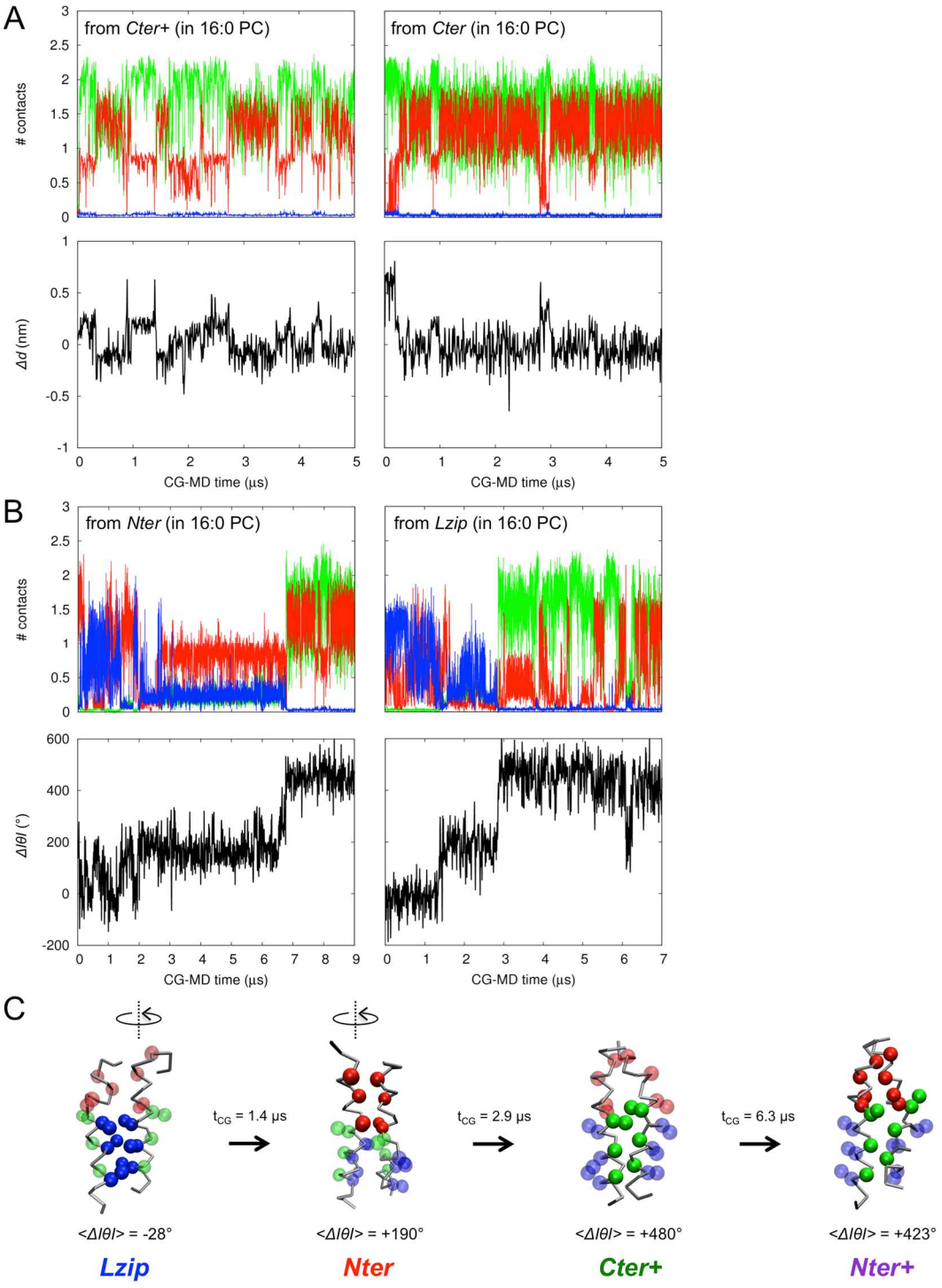


Fig. S10. Influence of the bilayer thickness on the kinetics of (A) pivot, and (B) rotation mechanisms. For a series of CG-MD simulations performed with either 12:0 PC or 20:0 PC bilayers, the Δd and $\Delta|\theta|$ CVs are monitored as a function of the simulation time in panels A and B, respectively.

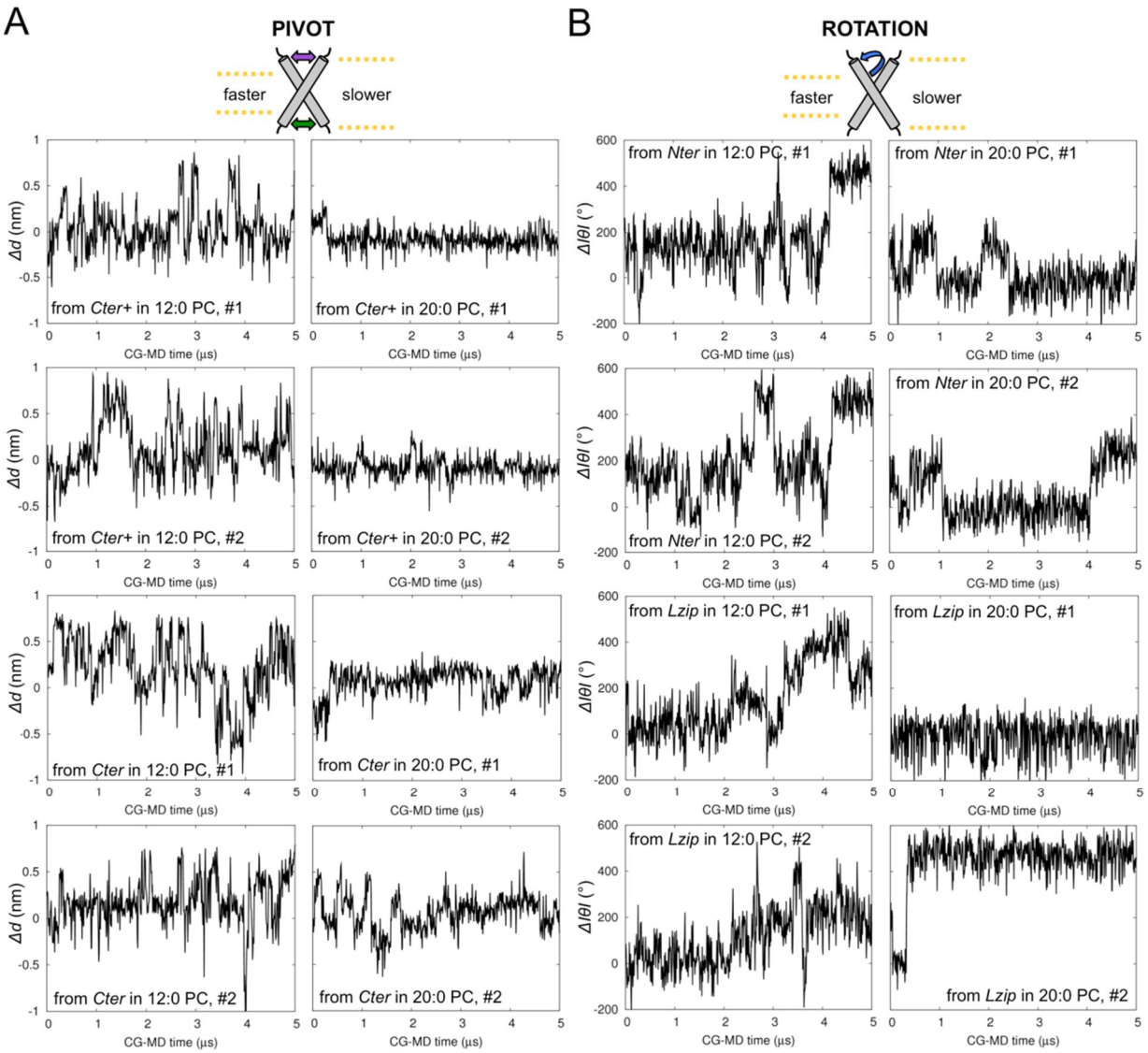
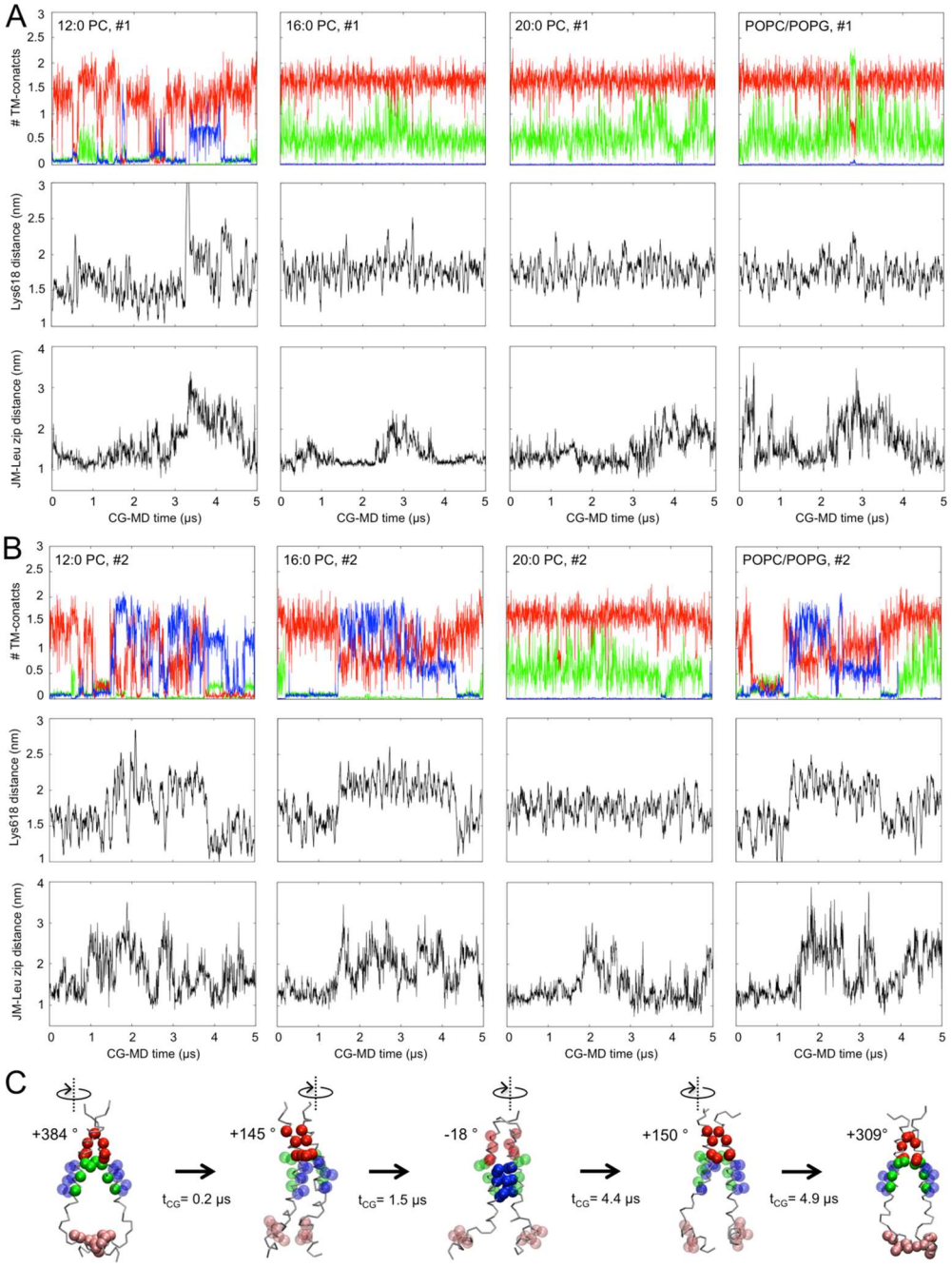


Fig. S11. (A) Number of interhelical contacts observed in the *Nter* (red), *Cter* (green) and *Lzip* (blue) motifs, during four plain CG-MD simulations starting from the NMR structure of Endres et *al.* (pdb id: 2M20) embedded either in pure bilayers containing 12:0 PC, or 16:0 PC, or 20:0 PC lipids, or mixed bilayer containing 85% of 16:0-18:1 PC and 15% of 16:0-18:1 PG lipids. In the middle panels, the distance between the center of mass of the residue Lys^{618} of each monomer is plotted as a function of the simulation time. In the bottom panels, the distance between the center of mass of the leucine zipper residues ($L^{655}xxL^{658}Leu^{659}$) in the helical part of the JM domain of each monomer is plotted as a function of the simulation time. (B) Number of interhelical contacts together with inter-residue distances observed during second repeats of the plain CG-MD simulations, for which a different seed was used to initiate particle velocities. (C) Structural summary of the dynamics observed during the second simulation repeat starting from the NMR structure in the 16:0 PC bilayer, indicating rotation events of individual monomer and average value of $\Delta|\theta|$ for each population. We note that a similar evolution was observed during the second simulation repeat starting from the NMR structure in the mixed POPC/POPG bilayer. The backbone of helices is shown in grey tubes, while the unstructured loops of the JM-B domain are omitted for clarity. The residues from motifs involved in direct interhelical contacts are drawn in opaque spheres, while residues from other motifs are shown in transparent spheres. Colour code is used as previously for representation of the different motifs in the TM domain, and leucine zipper residues in the helical part of the JM domain are shown in pink spheres.



Movie S1. Three consecutive crossing events of association/dissociation of the TM helices, observed during 3 μs of the CG-MetaD simulation. The backbone of TM helices is shown in grey tubes, while the phosphate groups of lipids are represented in transparent yellow spheres. The residues from motifs involved in interhelical contacts are drawn in opaque spheres, with the colour code used in the manuscript.

REFERENCES

(1) Coskun, U.; Grzybek, M.; Drechsel, D.; & Simons, K. *Proc Natl Acad Sci U S A* **2011** *108*, 9044-8.

(2) Michailidis, I.E.; Rusinova, R.; Georgakopoulos, A.; Chen, Y.; Iyengar, R.; Robakis, N.K.; Logothetis, D.E.; & Baki, L. *Pflugers Arch* **2011** *461*, 387-97.

(3) Halim, K.B.A.; Koldsø, H.; & Sansom, M.S.P. *Biochim. Biophys. Acta* **2015** *1850*, 1017-1025.