OPEN

Correspondence and
requests for materials
should be addressed to
H.K. (heidi.koldsoe@
gmail.com; heidi.
koldsoe@bioch.ox.ac.
uk)

The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids

George Hedger, Mark S. P. Sansom & Heidi Koldsø

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom.

Receptor tyrosine kinases (RTKs) play a critical role in diverse cellular processes and their activity is regulated by lipids in the surrounding membrane, including PIP_2 (phosphatidylinositol-4,5-bisphosphate) in the inner leaflet, and GM3 (monosialodihexosylganglioside) in the outer leaflet. However, the precise details of the interactions at the molecular level remain to be fully characterised. Using a multiscale molecular dynamics simulation approach, we comprehensively characterise anionic lipid interactions with all 58 known human RTKs. Our results demonstrate that the juxtamembrane (JM) regions of RTKs are critical for inducing clustering of anionic lipids, including PIP_2, both in simple asymmetric bilayers, and in more complex mixed membranes. Clustering is predominantly driven by interactions between a conserved cluster of basic residues within the first five positions of the JM region, and negatively charged lipid headgroups. This highlights a conserved interaction pattern shared across the human RTK family. In particular predominantly the N-terminal residues of the JM region are involved in the interactions with PIP_2, whilst residues within the distal JM region exhibit comparatively less lipid specificity. Our results suggest that JM–lipid interactions play a key role in RTK structure and function, and more generally in the nanoscale organisation of receptor-containing cell membranes.

The structure and function of membrane proteins are influenced by their lipid environment. Receptor tyrosine kinases (RTKs) are ubiquitous receptors in mammalian cell membranes, which transduce information about cellular environment and interactions across the plasma membrane to complex intracellular signalling networks[1]. Human RTKs comprise a family of 58 members divided into 20 subfamilies[2,3]. The diverse cellular processes regulated by these proteins include cell growth and division, differentiation, metabolism, migration and apoptosis[4,5]. The importance of RTK function is illustrated by the array of diseases, including cancer, diabetes, bone disorders, atherosclerosis and inflammatory conditions, which have been linked to pathogenic mutations in RTKs and the signalling systems they control[6,7]. Several RTK inhibitors are currently in clinical trials. Therefore, understanding the structure and function of these receptors is of considerable biomedical importance[8].

Almost all members of the human RTKs share a common molecular architecture consisting of an extracellular ligand binding ectodomain, a single transmembrane (TM) helix, and an intracellular region composed of a flexible juxtamembrane (JM) region followed by a protein kinase domain and a C-terminal region (see Fig. 1a).

Most RTKs can exist as monomers in an inactive state. However the insulin receptor (INSR) and insulin-like growth factor 1 receptor (IGF1R) exist as constitutive disulfide linked dimers[4,5,9]. Increasingly it is recognised that certain receptors, including the epidermal growth factor receptor (EGFR/ErbB1), can exist as inactive predimers and also in higher oligomeric states, while larger scale clustering within the membrane has also been observed[10–12]. It is well established that ligand binding to the extracellular ectodomains promotes receptor dimerization, and induces conformational reorganisation in existing dimers, leading to activation of protein kinase activity within the intracellular domains. The activated kinase domain of each monomer trans-autophosphorylates tyrosine residues in the C-terminal domain, JM region and/or activation loop of the opposing monomer leading to further downstream activation. RTK activity is regulated by multiple factors including interactions with membrane lipids such as GM3[13], phosphorylation by downstream *trans*-factors in feedback loops, clathrin mediated receptor
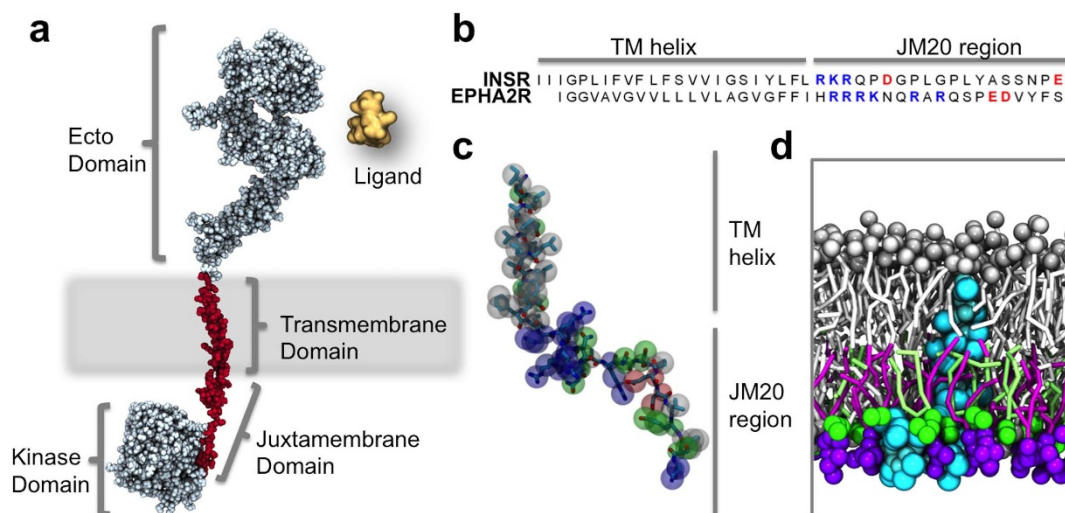
Figure 1 | RTK architecture includes a JM region enriched in basic residues. (a) Schematic representation of a generic RTK model (generated using MODELLER[69] based on PDB id: 2GS6[70], 3GOP[26], 2JIU[71] 2M20[72] and 3NJP[73]) with an extracellular ligand binding ectodomain, a single transmembrane spanning helix, a basic JM region, and a kinase domain. The TM and JM regions have been highlighted in red. (b) Sequence of INSR and the EphA2 receptor with charged residues within the JM region highlighted with basic residues in blue and anionic residues in red. (c) Protein model of the TM-JM region of the EphA2 receptor with CG beads overlaid onto an atomistic model: basic residues are shown in blue, acidic in red, polar in green, and non-polar in white. (d) Cross section of the same EphA2 receptor TM-JM model embedded in a lipid bilayer containing 20% PS and 20% PIP_2 within the inner leaflet.

endocytosis *via* NPxY motifs[14], ubiquitination, and by *cis*-elements including the kinase domain activation loop and the JM region[1].

The JM region is a flexible region normally consisting of 40 or more residues. It links the C-terminus of the transmembrane helix to the intracellular protein kinase domain. Previous efforts to understand RTK structure and function have primarily been focused on the ligand binding and protein kinase domains. The JM region is generally absent from crystal structures as it is often unstructured and intrinsically flexible. Due to difficulties in obtaining high resolution structures of full length RTKs, a modular interdisciplinary approach to RTK structure determination has been employed, with the X-ray crystallographic and NMR structures of the individual domains of many RTKs now known. These approaches have been successful, and in the case of ErbB1 structural data on isolated parts has been assembled into models of the full length receptor *in silico*[15–17]. The full length models are illustrative of a more recent shift in focus towards understanding whole receptors including the JM region, rather than isolated components; a goal which is also being pursued utilizing CryoEM[18].

The JM region is not simply an inert linker between the TM section and the kinase domain. It serves important regulatory roles and is essential for ErbB1 dimerization and activation of protein kinase activity[19,20]. Additionally the JM region often contains trafficking signal motifs[1], and binding sites for calmodulin[21]. Autoinhibition of protein kinase activity by the JM regions has been reported for numerous RTKs, including Fms-like tyrosine kinase (FLT3)[22], KIT[23], Muscle-specific kinase (MuSK)[24] and Ephrin (Eph) family RTKs[25]. It has been shown in most cases that the JM regions form inhibitory contacts with the kinase domain, which must be disrupted for receptor activation[9] to occur. In contrast, however, the JM region of EGFR acts as an activation domain, with alanine scanning mutagenesis defining the C-terminal JM region as being essential for receptor activation. A crystal structure of the asymmetric dimer reveals that the JM regions form multiple stabilizing contacts with each other[26]. Whether autoinhibitory or acting as an activation domain, in all cases studied to date, the JM region is essential for normal receptor function.

To date, the interactions between JM regions and anionic lipids have been studied for ErbB1 using experimental[27,28] and computa-tional approaches[15,29]. PIP_2 is an essential signalling lipid that is present in the inner leaflet of the plasma membrane. *In vitro* FRET labelling studies with ErbB1 JM peptides and phospholipid vesicles showed the JM construct was able to induce PIP_2 clustering[30]. Surface plasmon resonance (SPR) studies demonstrated strong binding between PIP_2 and the JM region of ErbB1[27]. PIP_2-JM interactions have also been observed in TM-JM constructs of rat epidermal growth factor receptor 2 (ErbB2)[31]. This interaction is likely to be mediated *via* electrostatic interactions between a cluster of basic residues in the JM region and the multiple phosphoryl groups of the headgroup of PIP_2, with interaction abolished in R/K > N neutralisation mutants[27].

Clustering of the anionic lipid phosphatidylserine (PS) around the JM region has been observed in atomistic MD simulations of full-length ErbB1[15], and specific interactions with PIP_2 were recently observed in multiscale simulations[29]. Interactions of the JM region with PIP_2 were favoured over PS interactions and a neutralising Asn mutant exhibited fewer interactions[29], consistent with an 'electrostatic engine model' for ErbB1 JM involvement in receptor activation[30]. In addition to these studies demonstrating PIP_2/PS-JM interactions of ErbB1/2, manipulation of PIP_2 levels by pharmacological knockdown of PIP_2-metabolizing enzymes demonstrated that ErbB1 receptor activity is PIP_2 dependent[27], thus adding ErbB1 to the list of membrane proteins modulated by PIP_2[32,33].

The presence of basic residues within the cytoplasmic JM region of several signalling receptors has been suggested to play a role in protein-lipid interactions[27,34]. Indeed the presence of basic residues within the JM region has been shown to be an inherent feature of all single transmembrane spanning proteins within eukaryotes. This significant enrichment in basic residues is present in proteins both within the plasma membrane and the golgi, and to a lesser extend the ER[35]. Despite the importance of lipid interactions revealed by these studies, the interactions of receptors with the membrane lipids have in the past been somewhat neglected within the RTK field[1] (as noted by e.g. Coskun *et al.*[13]). However, methodological advances and a growing body of literature demonstrating their critical nature has led to an appreciation of protein-lipid interactions as a key aspect of the biology of RTKs[36].

In contrast to previous studies, which have been limited to a single species of receptor, here we explore the specific interactions of the JM regions of the entire family of human RTKs with the anionic lipids PS and PIP_2. We initially sought to test the hypothesis that multiscale MD simulations can reveal TM-JM interactions of RTKs with anionic lipids by focussing on two systems: the INSR and EphA2 receptors. Based on these two test systems, we then created a high throughput approach to allow us to perform simulations of all 58 human RTKs, and test whether the interactions between the JM region and anionic lipids were conserved over this entire family of diverse receptors. This allows us to, for the first time, obtain insights into the specific nature of lipid interactions and their degree of conservation across the entire family of human RTKs.

Results

Models of the INSR and EphA2 receptors TM-JM regions consisting of the TM helix and first 20 residues of the JM region (Fig. 1b) were converted into coarse-grained (CG) representation (Fig. 1c) and inserted into phosphatidylcholine (PC) bilayers using self-assembly simulations[37]. Reduced representation methods such as CG[38–40] and hybrid models[41] have previously proven to be valuable approaches in the exploration of protein-lipid interactions. Asymmetric bilayers were constructed utilizing the method previously described in Koldsø *et al.*[42]. The asymmetric bilayers had a composition of $80:10:10$ of $PC:PS:PIP_2$ which closely mimics the *in vivo* anionic lipid composition of the inner (i.e cytoplasmic) leaflet (Fig. 1d). The simulations were run for 1 μs, with triplicate repeats performed for each system.

JM lipid interactions. The JM region moved towards the bilayer during the simulation and oriented itself to lie along the membrane surface, thereby allowing the basic residues to form favourable interactions with the anionic lipid headgroups (Fig. 1d and 2b). In particular, we observed favorable interactions between the headgroups of PIP_2 and the JM region of both the INSR and the EphA2 receptor as evident from the mapping of the mean number of interactions during the simulation onto the sequences of the proteins (Fig. 2a). Not surprisingly, PIP_2 exhibited the highest frequency of contacts with Arg and Lys residues, an intermediate level of contact with polar residues, and a low level contact with acidic and hydrophobic residues (Fig. 2).

Lipid clustering. PIP_2 was seen to form the overall highest degree of contact with JM residues, followed by PS, and then PC; as observed from the density maps for each lipid (Fig. 3 and Supplementary Fig. 1). Density maps of the inner membrane showing the spatial occupancy of each lipid headgroup over the course of the simulation were used to visualise lipid distributions within the INSR system (Fig. 3a). The EphA2 receptor system showed the same overall pattern (Supplementary Fig. 1). This analysis revealed clustering of lipids into ordered ring-like patterns around the protein. The mean number of PIP_2 lipid headgroups within 6 Å of the protein (corresponding to the first interaction shell) was between 4 and 6 for both systems. No such clustering was observed in a control system with only the transmembrane helix of INSR and no JM region (Supplementary Fig. 2).

Spherical radial distribution functions (RDFs) were computed for each lipid species to explore lipid distributions with respect to the protein (Fig. 3b). For both systems, lipids were seen to be distributed into shells around the protein with RDF peaks at 5 Å, 10 Å and 15 Å, consistent with results for the gp130 JM region which was observed to induce similar patterns of PS clustering[34]. Within the INSR system, the magnitude of these peaks for PIP_2 and PS decreased 5 Å > 10 Å > 15 Å, indicating decreasing probability of finding these lipids upon moving away from the protein (Fig. 3b). The same pattern was observed for the EphA2 receptor system (Supplementary Fig. 1b). The normalised RDF at 5 Å decreased when going from PIP_2 to PS and further for PC, indicating PIP_2 has the strongest preference for interaction with the JM region, followed by PS, and then PC. No such preference was observed in control simulations of the INSR TM helix without a JM region (Supplementary Fig. 2). In simulations with only $PS:PC$ and with only $PIP_2:PC$, the RDFs for PIP_2 and PS were very similar at 5 Å (Supplementary Fig. 3), indicating similar degrees of clustering within the first interaction shell. This dramatic drop in RDF for PS in the presence of PIP_2 suggests the more anionic PIP_2 headgroup outcompetes PS for interaction with the JM region. Both of these observations are consistent with clustering of anionic lipids *via* electrostatic interactions with basic JM residues.

To enable analysis of the dynamics of protein-lipid contacts, we also calculated the number of each lipid within 6 Å of each protein residue for each frame (Fig. 4). The results showed the number of contacts for each residue oscillates over the course of the simulations. This demonstrates that lipids do not bind irreversibly, but rather
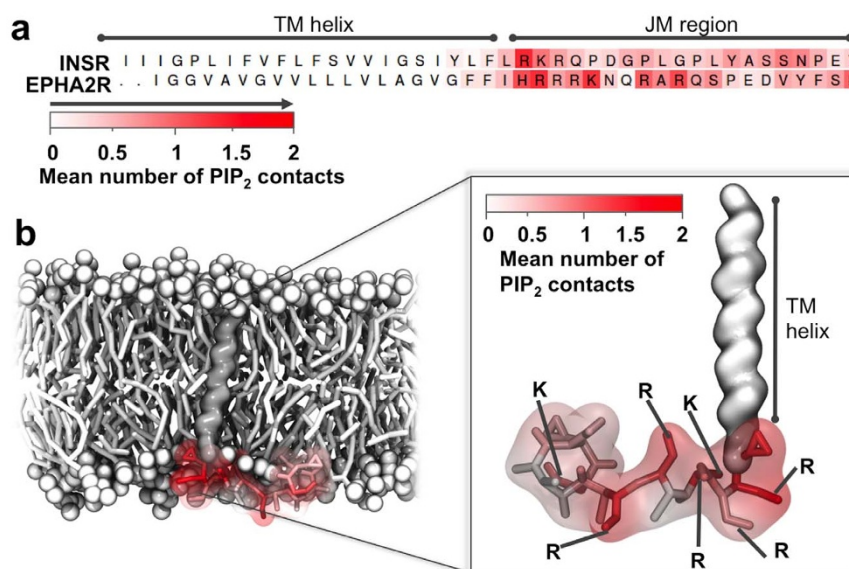
Figure 2 | Interactions between PIP_2 and the JM region. (a) TM helix-JM sequence of INSR and the EphA2 receptor coloured by the extent of PIP_2 headgroup contacts from white (low) to red (high). (b) The CG structure of the EphA2 receptor TM-JM model with residues also coloured according to the extent of contacts with the PIP_2 headgroup. Contacts were calculated over 3×1 μs repeat simulations using a 6 Å cutoff.
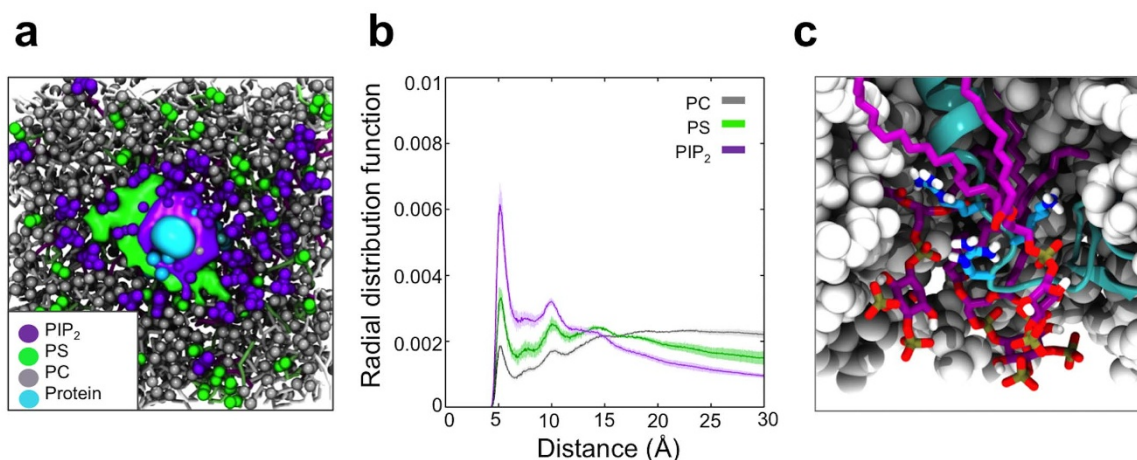
Figure 3 | Lipids clustering around the JM region within the INSR system. (a) Occupancy density maps of the inner leaflet surface showing spatial occupancy for the phosphoryl group attached to the glycerol moiety of the PIP_2 headgroup (purple), the PS headgroup (green), and the Arg and Lys residues of the JM region (cyan). Occupancy maps were computed from concatenated 3×1 μs simulations of each system, using Volmap in VMD. (b) Radial distribution functions (RDFs) for lipid headgroups of PIP_2, PS, and PC with respect to INSR. The RDFs shown are the mean values from 3×1 μs simulation repeats, with the standard deviation displayed as error bars. The area under each RDF curve is normalized to unity to aid comparison between the different lipids species. (c) A cluster of PIP_2 lipids around the INSR receptor JM region in an AT simulation. The protein has been shown in cyan with the basic juxtamembrane regions in sticks coloured according to atom type with carbon atoms in cyan. PIP_2 lipids within 6 Å of the protein have been shown as sticks coloured according to atom type, with carbon atoms in purple. The PC and PS lipids are shown as gray spheres.

their association is highly dynamic such that they associate and dissociate with the JM region multiple times over the course of the simulation (Fig. 4a). Basic residues within the JM region were seen to form the most stable long-lived interactions with PIP_2 and PS, whereas polar residues formed significantly more transient interactions with greater dynamic exchange with the surrounding lipid population. The percentage of time at least one of each lipid species was in contact with the JM region was also explored (Fig. 4b), which further supported the favourable interactions between PIP_2 and basic residues revealed by our RDF and time-averaged contact data (Fig. 3). Based on the occupancy time and number of association events we additionally calculated a mean residence time between each lipid species and each residue of the JM region (Fig. 4c). This calculation clearly indicates that the interactions between basic residues in INSR (in particular Arg24) are long lived and stable compared to more transient interactions between polar residues and lipids. Again it was observed that the interactions with PIP_2 over the other lipid species were very selective. A similar pattern was seen for the EphA2 receptor where all basic residues exhibit interactions with PIP_2 for most of the simulation time; and the interactions between PIP_2 and Lys26 were observed to be particular favourable (Supplementary Fig. 4). Additionally, we performed *in silico* mutagenesis in the EphA2 receptor in which all basic JM residues were mutated to leucine (Supplementary Fig. 5). This was chosen as leucine has a substantive hydrophobic side chain and thus differs from the basic residues mainly in terms of charge rather than size. In these simulations we observed loss of stable contacts at mutated positions and an overall decrease in the level of PIP_2 and PS clustering. Some level of contact between lipids and the JM region was maintained due to interactions with polar residues and random lateral diffusion.

AT simulations. To refine our CG observations, the final snapshots from 1 μs CG simulations of the INSR and EphA2 receptor systems were converted to AT detail and simulated for 50 ns of AT-MD. Root mean square fluctuations (RMSFs) were calculated for the Cα particle for both the CG and AT simulations, and used to compare the dynamic behaviour of each protein system (Supplementary Fig. 6). The RMSF values and profiles vs. residue number were very similar for the CG and AT simulations (Supplementary Fig. 6), indicating comparable degrees of JM flexibility at the two levels of

resolution. Over the course of each 50 ns AT simulation the same patterns of lipid clustering were observed as in the CG simulations, with PIP_2 having the greatest propensity to cluster and interact with the JM region (Fig. 3c and Supplementary Fig. 1c).

Analysis of lipid-protein contacts over time showed that the stable contacts of the PIP_2 lipid headgroup with basic residues were maintained in the atomistic systems (Supplementary Fig. 7). Visual inspection of the trajectory also revealed that this contact occurred *via* both electrostatic interactions between the phosphoryl groups of the PIP_2 headgroup and positively charged Lys and Arg side chains, as well as *via* some level of interaction of these residues with hydroxyl groups of the inositol moiety of PIP_2. PS was seen to interact with basic side chains *via* both its phosphoryl and carboxylate moieties. Interestingly, a high level of contact was also seen at the C-terminus of the INSR JM region, which contained no basic residues. This contact may be a function of the multiple proline kinks found in the INSR JM region which appear to cause it to adopt a conformation in which the JM region is slightly bent upwards to sit comparatively deep in the membrane (Supplementary Fig. 8 and Fig. 3c). This observation demonstrates slight differences in how the JM regions of RTKs interact with the membrane surface, despite overall broad similarities.

The high similarity between the behaviour observed within the CG and the AT simulations suggest that the CG representation is sufficient to fully capture the overall pattern and dynamics of RTK TM-JM region interactions with anionic lipids.

More complex membranes. Simple model bilayers provide useful test systems for exploring specific interactions. However the lipid composition of a mammalian plasma membrane is significantly more complex[43,44]. To better understand how this increased complexity may influence the interaction between anionic lipids and JM residues, we embedded both the INSR and the EphA2 receptors in bilayers of composition approximating that found in mammalian cell membranes[42], mainly composed of PC and phosphatidylethanolamine (PE) but also including glycolipids (GM3) within the outer leaflet and anionic lipids within the inner leaflet (PS and PIP_2), and performed 1 μs of CG simulation on these systems. The final composition within the outer leaflet was PC : PE : GM3 : Cholesterol (50 : 15 : 10 : 25), while the inner leaflet
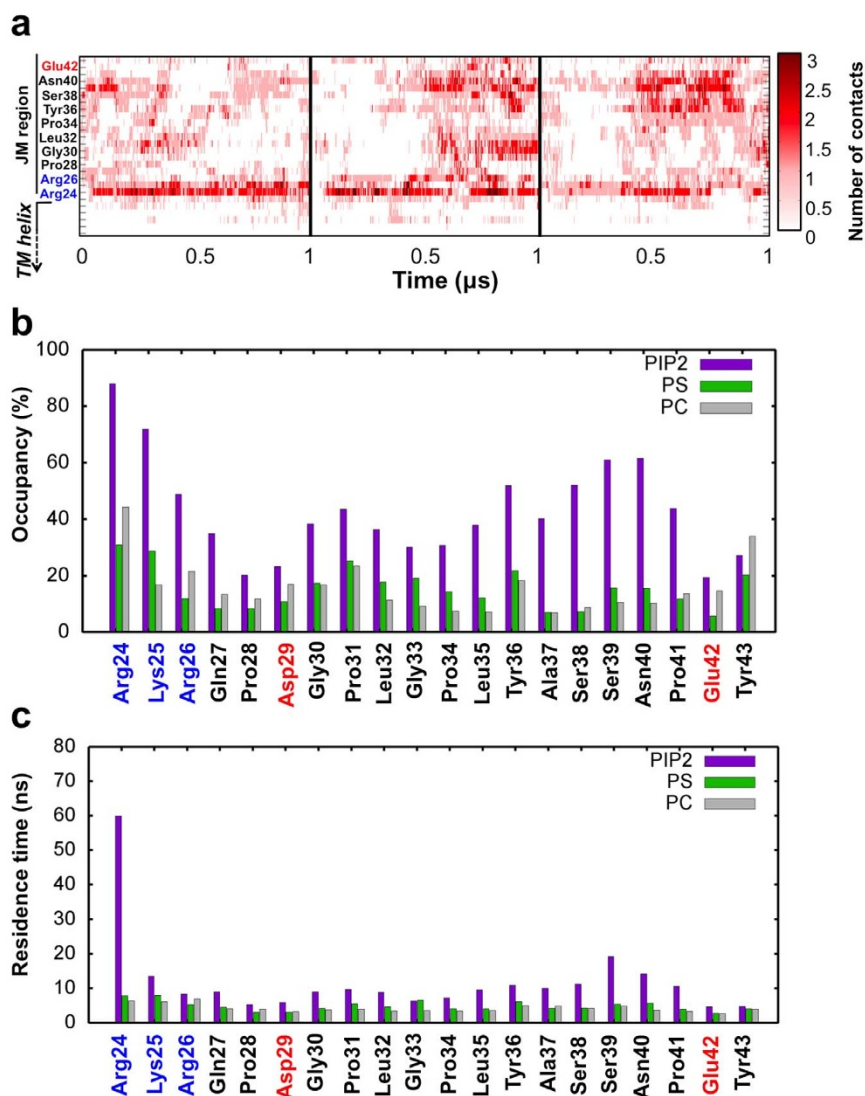
Figure 4 | Dynamics of PIP₂ interaction with the JM region. (a) Matrix showing the number of contacts per frame between each residue of the INSR JM region and the PIP$_2$ lipid headgroup as a function of time. The time course shown is from concatenated 3 × 1 µs CG simulation repeats. A cutoff of 6 Å was used to define contact. (b) Percentage of simulation time at least one lipid is in contact with each residue within the JM region. PIP$_2$, PS and PC are shown in purple, green and gray respectively. (c) Mean lipid residence time (occupancy time divided by number of association events) for concatenated 3 × 1 µs runs plotted for each JM residue.

had the following composition; PC : PE : PS : PIP$_2$: Cholesterol (10 : 40 : 15 : 10 : 25), see Supplementary Fig. 9.

The RDFs for PIP$_2$ and PS around the JM region indicated no significant deviation in their relative distribution from those observed in simple bilayers (Supplementary Fig. 10a + b). However the normalised probability decreased slightly. This qualitative decrease is likely to be a function of the reduced PIP$_2$ and PS concentrations within the plasma membrane systems. The diffusion for all lipid types were also decreased, which is a general trend as the complexity and crowding of membranes increases[42,45]. Protein-lipid interaction plots (Supplementary Fig. 10c + d) showed good agreement with those for lipids in simple bilayers, with PIP$_2$ forming the most stable interactions with basic JM residues and exhibiting the overall highest degree of contact. Our simulations of RTKs within more complex mixed lipid membranes further support our CG and AT findings from simple asymmetric membrane systems.

All 58 human RTKs. We extended our CG simulations to the TM-JM regions of all 58 human RTKs in order to evaluate the extent to which lipid interactions and clustering are conserved across the entire family. TM-JM models of each RTK were built and embedded in asymmetric PC : PS : PIP$_2$ (80 : 10 : 10) bilayers using the protocol already described for the INSR and EphA2 receptors. Three runs of 1 µs simulations were performed for each RTK system (Fig. 5).

Contacts between each residue of the protein and different lipid headgroups were mapped onto the protein structure (Fig. 5) and onto the amino acid sequences (Fig. 6). The overall trend in contacts between the JM region and lipids was in agreement with the findings from our more detailed analysis of the INSR and the EphA2 receptors. Arginine and lysine residues showed a strikingly high level of contact with the PIP$_2$ lipid headgroup, with polar residues also showing an overall intermediate level of contact. Similar patterns of contact were observed for PS (Supplementary Fig. 11), but not for PC (Supplementary Fig. 12). PIP$_2$ was seen to form the highest overall levels of contact with the JM residues, followed by PS, and then PC. This agrees with RDF and density map data from the INSR and EphA2 receptor systems, showing PIP$_2$ has the highest propensity to be distributed around the JM region, followed by PS.
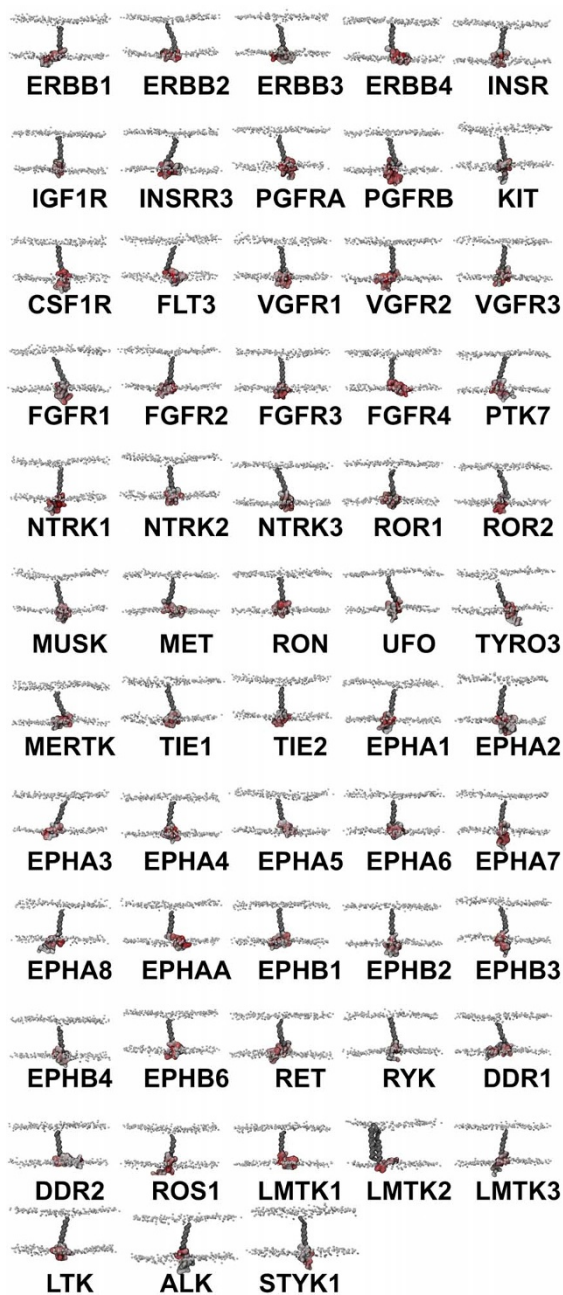
Figure 5 | Simulations of all 58 human RTKs. All proteins were embedded into PC:PS:PIP$_2$ (80:10:10) bilayers and simulated for 3 × 1 μs. The lipid headgroups are shown by white spheres and all the TM-JM models are coloured according to the extent of contacts with the PIP$_2$ headgroup. Contacts were calculated over 3 × 1 μs repeat simulations using a 6 Å cutoff.

Correlation between all systems. In all systems we found the same patterns of lipid distribution around the JM region as we saw for the INSR and the EphA2 receptors, with PIP$_2$ exhibiting the greatest propensity to cluster around the JM region, followed by PS, and then PC. For those RTKs whose JM region exhibits a net positive charge, a degree of correlation between the net charge and the degree of lipid clustering was observed (Fig. 7). However the degree of clustering seems to saturate at around a net charge of +5. Multiple other factors including charge distribution, JM dynamics, and relative numbers of polar and hydrophobic residues will also influence the degree of lipid clustering and it is therefore not unexpected that a perfect correlation is not observed. Charge

distribution is likely to be particularly important; a JM region with all the basic residues grouped at one end will physically be able to accommodate fewer lipids within 6 Å of the charged residues than one in which the same number of charged residues are evenly distributed along the JM region.

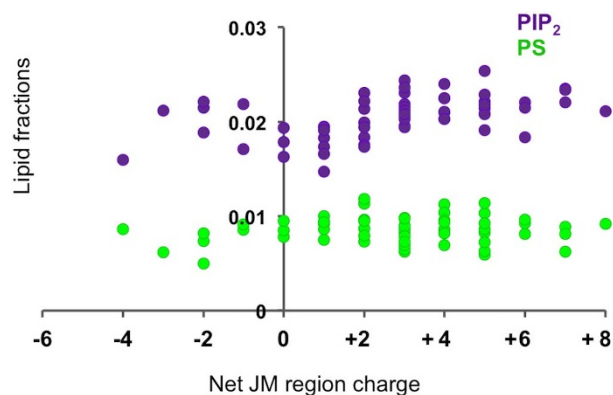**Global analysis of basic residue position and lipid interactions.** We analysed all 58 human RTK simulations more globally. The mean number of lipid contacts formed at each JM residue position over all RTKs was mapped onto a single 'average' JM structure (Fig. 8). Our results reveal that the overall pattern of PIP$_2$ lipid interaction is conserved across the entire RTK family. The highest level of contact with the PIP$_2$ lipid headgroup tends to occur at the beginning of the JM region, adjacent to the TM helix (Fig. 8a). Comparison of all 58 RTKs also revealed the highest probability of arginine and lysine occurrence is found within the first five residues of the JM region (Fig. 8b), corresponding to the positions with the highest levels of PIP$_2$ lipid headgroup contact. We did not observe a significant enrichment in the mean number of PS lipids within any particular region (Fig. 8c). Our data clearly illustrate the selectivity and specificity of JM-PIP$_2$ interactions, and since the same pattern is not observed for PS, the observed PIP$_2$ interaction sites are unlikely to be just an effect of the increased membrane surface proximity of this early JM region. The high occurrence of basic residues within the first five residues of the JM region is additionally the overall most conserved sequence feature when considering all 20 residues within all 58 human RTKs simulated here (Fig. 8d).

Discussion

The present study illustrates the nanoscale lateral organisation of lipids in bilayers around the JM region of the entire human RTK family, both in relatively simple and more complex lipid bilayers. Previous investigations of RTK-lipid interactions have been limited to single species of RTK, and in particular EGFR. The present study extends over the entire family of 58 human RTKs and thus provides novel and more general insights into the molecular details of these interactions and their conserved nature across an *entire family* of receptor proteins. Our simulations reveal that models of the TM helix and a short 20 residue JM regions of all 58 known human RTKs are able to induce bilayer reorganizations in the form of clustering of PIP$_2$ and PS lipid molecules into shells around the JM region. Selectivity for certain lipid species was evident, with PIP$_2$ found to have the greatest propensity to cluster around the JM regions of all receptors. The observed patterns of lipid distribution agreed with studies of the gp130 cytokine receptor[34] and dimers of ErbB1[29], as well as being corroborated by atomistic simulations of full-length ErbB1 which showed PS clustering around the full-length JM regions[15], and by experimental studies of ErbB1 peptide binding to PIP$_2$ monitored by FRET and SPR[27,30]. In bilayers with only PS and PC within the inner leaflet, we found that the degree of PS clustering was comparable to that of PIP$_2$ in systems containing all three lipid species, suggesting that when present PIP$_2$ is able to compete with PS for interaction with the protein. This highlights the importance of lipid diversity in bilayers both *in vivo* and *in silico*, and has implications for the physiological relevance of studies in oversimplified model bilayers. We tested the stability of interactions within more complex lipid membranes, and the same qualitative patterns of PIP$_2$ clustering and lipid-protein interaction were observed.

Analysis of contacts formed between different lipid molecules and protein residues over the course of the simulations demonstrated conserved patterns of lipid interaction over all 58 human RTKs, revealing that that these specific interactions were mainly located within the N-terminal part of the JM region. Anionic lipids, and in particular PIP$_2$, were seen to form the highest degree of contact with basic residues, an intermediate level of contact with polar residues, and a low level of contact with hydrophobic and acidic residues. This

Figure 6 | PIP₂ interactions in all 58 human RTKs. Sequence of the TM-JM region of all 58 RTKs, with each protein residue coloured by the mean number of contacts it formed the phosphoryl group attached to the glycerol moiety of the PIP₂ headgroup per frame. Contacts were calculated over $3 \times 1\ \mu s$ CG simulation repeats for each RTK. A 6 Å cutoff was used to define contact between the two groups.

suggests electrostatic interactions between the anionic headgroup moieties of PIP₂ and PS and clusters of basic residues within the early JM region are the predominant drivers of the observed clustering patterns, with the greater propensity for PIP₂ to interact with the JM region being a function of its significantly more negative head-

group. In agreement with this observation, *in silico* mutation of basic residues within the JM region of EphA2 receptor to leucines caused a decrease in PIP₂ and PS clustering and loss of stable contacts at mutated positions. These qualitative conclusions agree with *in vitro* mutagenesis[27] and computational[29] studies of ErbB1 in which R/K >

Figure 7 | Correlation between JM charge density and PIP$_2$ interactions in all 58 human RTKs. Relationship between the relative propensity of PS and PIP$_2$ lipid species to cluster around the JM, and net JM region charge for all 58 RTKs. The clustering propensity is measured by computing the mean fraction of each lipid type within 6 Å of the protein per frame, and normalised by dividing by the percentage of that lipid in the inner leaflet.

N neutralisation mutants were seen to abolish PIP$_2$ binding, as well other simulation studies of ErbB1 and the gp130 cytokine receptor which also found high levels of contact between basic JM residues and anionic lipids[15,34].

The results we present here reveal the ability of the short membrane-proximal JM regions of all 58 human RTKs to induce local bilayer reorganizations and clustering of anionic lipids in the inner leaflet around the embedded transmembrane receptors. This behaviour is primarily a result of electrostatic interactions that are localised to a cluster of conserved basic residues found within the first five positions of the JM region, and anionic lipid head groups. The insights that these simulations provide are of interest both for understanding the overall nanoscale organization of proteins and lipids within cell membranes, as well as in terms of RTK function. PIP$_2$ is a major modulator of membrane protein activity. For example, it is known to modulate ErbB1 activity[27], and also to regulate the behaviour of psychostimulant drugs *via* interactions with the dopamine transporter[46]. Thus, the apparent ability of all RTKs to induce clustering of PIP$_2$ lipid molecules around their JM regions *via* interactions with conserved residues may therefore have important implications for their regulation. Additionally, PIP$_2$ is a central downstream signalling component in a number of cascades controlled by receptor tyrosine kinases, including the PI3K-PKB and PLC-PKC cascades. It has indeed been shown experimentally and computationally that PIP$_2$ can form microdomains as a result of protein-lipid interactions[47] and experimentally that the signalling pathways of RTKs such as EphA2 receptors can be regulated by the surrounding environment[48,49]. We can speculate that localization of PIP$_2$ around the JM regions of RTKs may therefore be beneficial to the enhancement of signal transduction efficiency within these signalling systems, especially when it is considered that a number of RTKs form higher order arrays at the cell surface[11,12], which could potentially result in an substantial local pool of PIP$_2$. Furthermore, the conservation of the physiochemical properties of residues of the N-terminal section of the JM across the entire family of human RTKs and of the resulting lipid interaction hotspots are strongly suggestive of their importance in receptor function, not only on a single protein level but from a broader evolutionary perspective within cellular function.

Methods

Model proteins consisting of the predicted transmembrane helix-JM region of each receptor tyrosine kinase were built using PyMOL (PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). The protein sequence was obtained from UniProt[50] with the predicted transmembrane regions further supported using TMpred[51]. We tested the possible influence of the N-terminal region prior to the TM helix by inclusion of +4 residues at the N-terminus of the INSR TM-JM model, but found no significant impact the predicted on JM-lipid interactions (Supplementary Fig. 13). In the interest of simplicity the N-terminal region was therefore omitted from the protein models used in our simulations. PSIPRED[52], YASPIN[53], and JPRED[54]
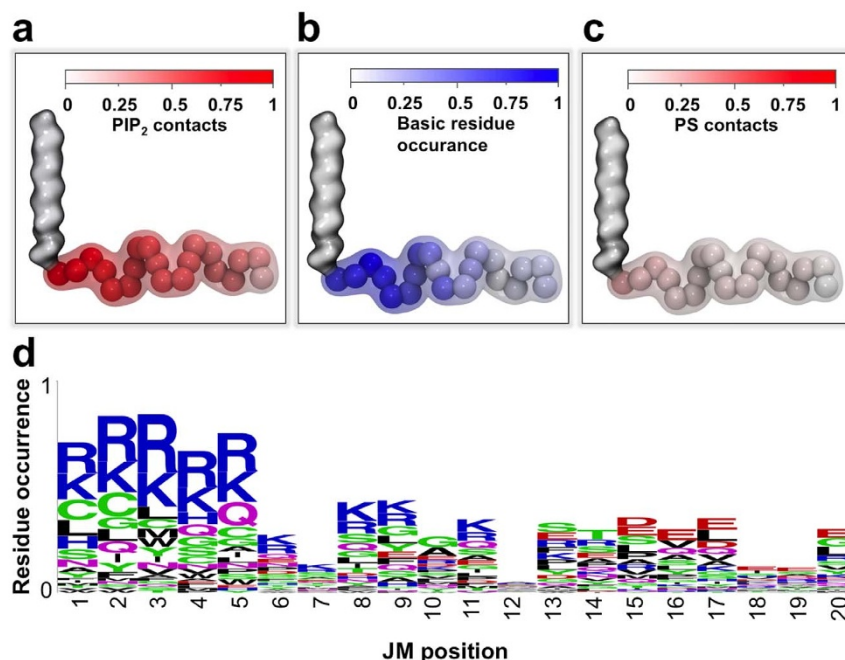
Figure 8 | Interactions with PIP$_2$ are mainly found in N-terminal end of the JM region. (a) Mean number of PIP$_2$ contacts for all 58 human RTKs. TM-JM models coloured by the degree of lipid contact, using contact data from all 58 RTKs. The number of contacts per frame at each JM residue position was summed over all receptors and divided by 58 to give the mean number of contacts per frame at each JM residue position. (b) A hypothetical TM-JM model coloured by the probability of basic residue occurrence at each position of the JM based on all 58 human RTKs. (c) Mean number of PS contacts for all RTKs was calculated as described for (a). (d) Residue occurrence shown as a function of position within the JM region.

were used to predict secondary structure within the JM region. Secondary structure was then modelled when agreement was found between two or more prediction programs. The JM regions of each RTK were modelled onto the transmembrane region in an extended conformation to prevent bias and encourage successful self-assembly of the lipid bilayer around the transmembrane helix. This modelling of JM tails was achieved using the sculpting tool in PyMOL and using the command line to set backbone (φ, ψ) angles.

Coarse-grain molecular dynamics (CG-MD) simulations. All simulations were performed using GROMACS 4.6[55]. Most simulations were performed in membranes containing phosphatidylcholine (PC), phosphatidylserine (PS), and phosphatidylinositol-4,5-bisphosphate (PIP$_2$). Additionally, some simulations were performed in more complex membranes containing PC, PS, PIP$_2$, GM3, phosphatidylethanolamine (PE), and cholesterol. Atomistic (AT) structures were energy minimized and converted to coarse-grained (CG) representations using the martinize.py workflow, and the MARTINI 2.2 force field was applied with periodic boundary conditions[56,57]. A timestep of 20 fs was used for all simulations with Coulomb and van der Waals interactions modelled with a cutoff of 10 Å. Temperature was controlled at 323 K for self-assembly simulations and at 310 K for production run simulations using a Berendsen thermostat[58] with a coupling constant of 1 ps. Pressure was controlled at 1 bar using a Berendsen barostat with a 1 ps coupling constant. The LINCS algorithm was used to constrain covalent bonds to their equilibrium values[59]. The proteins were inserted into a POPC bilayer using self-assembly simulations[37] for 25 ns with a time step of 20 fs as previously described for the gp130 receptor[34]. The simulations were run with the standard MARTINI water[57] and then neutralised and ionized using 0.15 M NaCl. The membrane composition was subsequently changed into the desired composition utilizing an exchange lipids approach previously published[42]. The script randomly exchanges POPC with specified lipids in each leaflet before re-solvating, re-ionizing and energy minimizing the system. The production runs of the TM-JM models embedded in asymmetric bilayers were simulated for 1 μs of CG MD simulation at 310 K. Each 1 μs production run was repeated three times with different initial velocity seeds and with different random position of the lipids.

Atomistic molecular dynamics (AT-MD) simulations. For INSR and the EphA2 receptor TM-JM region systems, a CG to AT fragment based protocol[60] was used to convert the final frame of each 1 μs CG repeat to AT detail. Simulations were performed using the united-atom GROMOS 53a6[61] force field in GROMACS 4.6[55]. Electrostatics were modelled using the Particle Mesh Ewald (PME) model[62] with a cutoff of 10 Å, while a cutoff of 10 Å was also used for van der Waals interactions. Temperature was controlled at 310 K by a V-scale thermostat[63] and pressure controlled at 1 bar utilizing the Parinello-Rahman barostat[64]. The LINCS algorithm was used to constrain covalent bonds to their equilibrium values[59].

After conversion to AT detail each system was re-solvated, re-ionized and energy minimized before being equilibrated for 1 ns with position restraints applied to the Cα atoms of the protein with a force constant of 10 kJ/mol/Å^2. Each AT simulation was then run for 50 ns with a 2 fs time step.

Analysis. VMD plugins were used to compute radial distribution functions and generate VolMap density surfaces[65,66]. The RDF data was normalised such that the area under the curve for each lipid species was equal to one to enable ready comparison between different lipid species. Contact data was mapped onto sequence using ALINE[67], whilst mapping of contact data to structure was achieved utilizing the β-factor columns of .pdb structure files, and visualised using VMD[66]. Further analysis was carried out through VMD using locally developed scripts. Figure 8d was generated utilizing the WebLogo server[68].

1. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
2. Robinson, D. R., Wu, Y. M. & Lin, S. F. The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548–5557 (2000).
3. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
4. Ullrich, A. & Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203–212 (1990).
5. Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **103**, 211–225 (2000).
6. Gschwind, A., Fischer, O. M. & Ullrich, A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat. Rev. Cancer* **4**, 361–370 (2004).
7. Mendelsohn, J. & Baselga, J. The EGF receptor family as targets for cancer therapy. *Oncogene* **19**, 6550–6565 (2000).
8. Zwick, E., Bange, J. & Ullrich, A. Receptor tyrosine kinases as targets for anticancer drugs. *Trends Mol. Med.* **8**, 17–23 (2002).
9. Hubbard, S. R. Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nat. Rev. Mol. Cell Biol.* **5**, 464–471 (2004).
10. Nagy, P., Claus, J., Jovin, T. M. & Arndt-Jovin, D. J. Distribution of resting and ligand-bound ErbB1 and ErbB2 receptor tyrosine kinases in living cells using number and brightness analysis. *Proc. Natl. Acad. Sci. U S A* **107**, 16524–16529 (2010).
11. Clayton, A. H. A., Tavarnesi, M. L. & Johns, T. G. Unligated epidermal growth factor receptor forms higher order oligomers within microclusters on A431 cells

12. Saffarian, S., Li, Y., Elson, E. L. & Pike, L. J. Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys. J.* **93**, 1021–1031 (2007).
13. Coskun, Ü., Grzybek, M., Drechsel, D. & Simons, K. Regulation of human EGF receptor by lipids. *Proc. Natl. Acad. Sci. U S A* **108**, 9044–9048 (2011).
14. Goh, L. K. & Sorkin, A. Endocytosis of receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **5** (2013) doi:10.1101/cshperspect.a017459.
15. Arkhipov, A. *et al.* Architecture and membrane interactions of the EGF receptor. *Cell* **152**, 557–569 (2013).
16. Arkhipov, A., Shan, Y., Kim, E. T. & Shaw, D. E. Membrane interaction of bound ligands contributes to the negative binding cooperativity of the EGF Receptor. *PLoS Comput. Biol.* **10**, e1003742 (2014) doi:10.1371/journal.pcbi.1003742.
17. Koland, J. G. Coarse-grained molecular simulation of epidermal growth factor receptor protein tyrosine kinase multi-site self-phosphorylation. *PLoS Comput. Biol.* **10**, e1003435 (2014) doi:10.1371/journal.pcbi.1003435.
18. Mi, L. *et al.* Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. *Nat. Struct. Mol. Biol.* **18**, 984–989 (2011).
19. Thiel, K. W. & Carpenter, G. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proc. Natl. Acad. Sci. U S A* **104**, 19238–19243 (2007).
20. Aifa, S. *et al.* A basic peptide within the juxtamembrane region is required for EGF receptor dimerization. *Exp. Cell Res.* **302**, 108–114 (2005).
21. Martín-Nieto, J. & Villalobo, A. The human epidermal growth factor receptor contains a juxtamembrane calmodulin-binding site. *Biochemistry* **37**, 227–236 (1998).
22. Griffith, J. *et al.* The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol. Cell* **13**, 169–178 (2004).
23. Mol, C. D. *et al.* Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **279**, 31655–31663 (2004).
24. Till, J. H. *et al.* Crystal structure of the MuSK tyrosine kinase: Insights into receptor autoregulation. *Structure* **10**, 1187–1196 (2002).
25. Wybenga-Groot, L. E. *et al.* Structural basis for autoinhibition of the EphB2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell* **106**, 745–757 (2001).
26. Brewer, M. R. *et al.* The juxtamembrane region of the EGF receptor functions as an activation domain. *Mol. Cell* **34**, 641–651 (2009).
27. Michailidis, I. E. *et al.* Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation. *Pflugers Arch.* **461**, 387–397 (2011).
28. McLaughlin, S., Smith, S. O., Hayman, M. J. & Murray, D. An electrostatic engine model for autoinhibition and activation of the epidermal growth factor receptor (EGFR/ErbB) family. *J. Gen. Physiol.* **126**, 41–53 (2005).
29. Abd Halim, K. B., Koldsø, H. & Sansom, M. S. P. Interactions of the EGFR juxtamembrane domain with PIP$_2$-containing lipid bilayers: Insights from multiscale molecular dynamics simulations. *Biochim. Biophys. Acta.* (2014) In Press, doi:10.1016/j.bbagen.2014.09.006.
30. McLaughlin, S. & Murray, D. Plasma membrane phosphoinositide organization by protein electrostatics. *Nature* **438**, 605–611 (2005).
31. Matsushita, C. *et al.* Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides. *Proc. Natl. Acad. Sci. U S A* **110**, 1646–1651 (2013).
32. Hilgemann, D. W., Feng, S. & Nasuhoglu, C. The complex and intriguing lives of PIP$_2$ with ion channels and transporters. *Sci. STKE.* **2001**, re19 (2001).
33. McLaughlin, S., Wang, J., Gambhir, A. & Murray, D. PIP$_2$ and proteins: Interactions, organization, and information flow. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 151–175 (2002).
34. Koldsø, H. & Sansom, M. S. P. Local lipid reorganization by a transmembrane protein domain. *J. Phys. Chem. Lett.* **3**, 3498–3502 (2012).
35. Sharpe, H. J., Stevens, T. J. & Munro, S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169 (2010).
36. Bessman, N. J., Freed, D. M. & Lemmon, M. A. Putting together structures of epidermal growth factor receptors. *Curr. Opin. Struct. Biol.* **29**, 95–101 (2014).
37. Scott, K. A. *et al.* Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16**, 621–630 (2008).
38. Stansfeld, P. J., Hopkinson, R., Ashcroft, F. M. & Sansom, M. S. P. PIP$_2$-binding site in Kir channels: Definition by multiscale biomolecular simulations. *Biochemistry* **48**, 10926–10933 (2009).
39. Arnarez, C., Marrink, S. J. & Periole, X. Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Sci. Rep.* **3**, 1263 (2013).
40. Arnarez, C., Mazat, J. P., Elezgaray, J., Marrink, S. J. & Periole, X. Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. *J. Am. Chem. Soc.* **135**, 3112–3120 (2013).
41. Ohkubo, Y., Pogorelov, T., Arcario, M., Christensen, G. & Tajkhorshid, E. Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. *Biophys. J.* **102**, 2130–2139.
42. Koldsø, H., Shorthouse, D., Hélie, J. & Sansom, M. S. P. Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid

bilayers. *PLoS Comput. Biol.* **10**, e1003911 (2014) doi:10.1371/journal.pcbi.1003911.

43. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124 (2008).

44. van Meer, G. & de Kroon, A. I. P. M. Lipid map of the mammalian cell. *J. Cell Sci.* **124**, 5–8 (2011).

45. Goose, J. E. & Sansom, M. S. P. Reduced lateral mobility of lipids and proteins in crowded membranes. *PLoS Comput. Biol.* **9**, e1003033 (2013) doi:10.1371/journal.pcbi.1003033.

46. Hamilton, P. J. *et al.* PIP$_2$ regulates psychostimulant behaviors through its interaction with a membrane protein. *Nat. Chem. Biol.* **10**, 582–589 (2014).

47. van den Bogaart. *et al.* Membrane protein sequestering by ionic protein-lipid interactions. *Nature* **479**, 552–555 (2011).

48. Groves, J. T. & Kuriyan, J. Molecular mechanisms in signal transduction at the membrane. *Nat. Struct. Mol. Biol.* **17**, 659–665 (2010).

49. Salaita, K. *et al.* Restriction of receptor movement alters cellular response: Physical force sensing by EphA2. *Science* **327**, 1380–1385 (2010).

50. Magrane, M. & Unirprot Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011** (2011).

51. Hofmann, K. & Stoffel, W. TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993).

52. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

53. Lin, K., Simossis, V. A., Taylor, W. R. & Heringa, J. A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* **21**, 152–159 (2005).

54. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucl. Acids Res.* **36**, W197–W201 (2008).

55. Hess, B., Kutzner, C., van, d. S. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).

56. Monticelli, L. *et al.* The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

57. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

58. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

59. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).

60. Stansfeld, P. J. & Sansom, M. S. P. From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* **7**, 1157–1166 (2011).

61. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).

62. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

63. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126** (2007).

64. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

65. Levine, B. G., Stone, J. E. & Kohlmeyer, A. Fast analysis of molecular dynamics trajectories with graphics processing units—Radial distribution function histogramming. *J. Comput. Phys.* **230**, 3556–3569 (2011).

66. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

67. Bond, C. S. & Schüttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Cryst.* **65**, 510–512 (2009).

68. Crooks, G. E., Hon, G., Chandonia, J. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

69. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

70. Zhang, X., Gureasko, J., Shen, K., Cole, P. A. & Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125**, 1137–1149 (2006).

71. Yun, C. *et al.* The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U S A* **105**, 2070–2075 (2008).

72. Endres, N. *et al.* Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **152**, 543–556 (2013).

73. Lu, C. *et al.* Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. *Mol. Cell Biol.* **30**, 5432–5443 (2010).

Acknowledgments

Author contributions

G.H. and H.K. performed all the experiments. G.H. and H.K. performed the data analyses. M.S.P.S. and H.K. designed the experiments. G.H., M.S.P.S. and H.K. wrote the article.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hedger, G., Sansom, M.S.P. & Koldsø, H. The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Sci. Rep.* **5**, 9198; DOI:10.1038/srep09198 (2015).

10