December 8, 2020

A structural mechanism for the generation of biased agonism in the epidermal growth factor receptor

Yongjian Huang[1,2,3], Jana Ognjenović[4], Deepti Karandur[1,2,3], Alan Merk[4], Sriram Subramaniam[5*], John Kuriyan[1,2,3,6,7*]

[1] Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA.

[2] California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720, USA.

[3] Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA 94720, USA.

[4] Frederick National Laboratory for Cancer Research, Frederick, MD 21701, USA.

[5] University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.

[6] Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA.

[7] Divisions of Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*Correspondence: sriram.subramaniam@ubc.ca, kuriyan@berkeley.edu

Abstract

The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase that couples the binding of extracellular ligands, such as EGF and transforming growth factor-a (TGF-α), to the initiation of intracellular signaling pathways[1–3]. EGFR binds to EGF and TGF-α with similar affinity but exhibits biased agonism, generating different signals from these ligands[4–6]. The mechanistic basis for this phenomenon remains unclear. We now present cryo-EM analyses of human EGFR bound to EGF and TGF-α. We find, unexpectedly, that the extracellular module adopts an ensemble of dimeric conformations when bound to either EGF or TGF-α. The two extreme states of this ensemble represent distinct ligand-bound quaternary structures in which the membrane-proximal ends of the extracellular module are either juxtaposed or separated. EGF and TGF-α differ in their ability to maintain the conformation with the membrane-proximal ends of the extracellular module separated, and this conformation is stabilized preferentially by an oncogenic EGFR mutation. Close proximity of the transmembrane helices at the junction with the extracellular module is known to increase EGFR activity. Our results show how EGFR can couple the binding of different ligands to differential modulation of this proximity, thereby providing a mechanism for the generation of biased agonism in this receptor family.

Main

Human EGFR can be activated by seven related ligands that generate biased agonism, with ligand-dependent differences in cellular outputs[4–6]. While the classical mechanism for receptor tyrosine kinase activation, ligand-induced dimerization, is essential for all these ligands to activate the receptor, it cannot explain how the extracellular module of EGFR is able to respond differently to different ligands once the receptor is dimerized. The extracellular module of EGFR consists of four domains, denoted Domains I, II, III, and IV (Fig. 1a). Domains I and III sandwich the ligands between them[7,8]. Domain II, which bridges the ligand-binding domains, contains the "dimerization arm", a loop that mediates the principal interaction with the other subunit in an activated dimer[9]. Domains I, II, and III form a compact unit that we refer to as the ligand-binding "head", and Domain IV forms a rigid and elongated "leg" that connects the ligand-binding head to the single transmembrane helix (Fig. 1a). Previous crystallographic analyses of fragments of the EGFR extracellular module have revealed different conformations of the ligand-binding head when bound to EGF, TGF-α, epiregulin, and epigen[7,8,10,11]. It appeared that these conformational differences might explain how different EGFR ligands generate different signals[4,12], but the significance of these structural difference is unclear because only fragments of the receptor were analyzed. For weakly-binding ligands, such as epiregulin and epigen, the lifetime of the dimeric receptor is shortened relative to that for the tighter-binding ligands, such as EGF, and the reduction in dimer lifetime affects signaling output[11]. However, EGF and TGF-α are both high-affinity ligands[13], and they also generate different outputs from EGFR. How this happens is still unclear.

Cell-based studies using fluorescent reporters of EGFR conformation have shown that the binding of EGF and TGF-α results in different conformations of the intracellular juxtamembrane segments of the receptor[12,14,15]. The balance between alternative conformations of the

juxtamembrane segments can be modulated by point mutations in the transmembrane helices [16] and oncogenic mutations in the kinase domains[15]. These studies demonstrated that the extracellular module, the transmembrane helix, the intracellular juxtamembrane segments, and the intracellular kinase domains are coupled allosterically, and that they respond differently to EGF and TGF-α. How the two ligands generate these differences remained unknown. To gain structural insight into this mechanism, we analyzed full-length human EGFR in complex with EGF and TGF-α by cryo-EM.

Linkages between the extracellular modules and the transmembrane helices are not rigidly defined

We purified full-length human EGFR expressed in human HEK293S GnTI$^-$ cell line[17], and reconstituted the complex of EGFR with EGF into four membrane-mimetic environments: detergent micelles, lipid nanodiscs[18], amphipols[19], and peptidiscs[20,21] (Extended Data Fig. 1). Cryo-EM reconstructions of the full-length EGFR:EGF complex in all four reconstitutions look very similar, and only the extracellular module is resolved in the cryo-EM densities (Fig. 1b, c, Extended Data Fig. 2). The lack of resolvable density for the transmembrane helices and the intracellular kinase domains suggests that the linkage between the extracellular module and the transmembrane helices is not stereochemically defined, consistent with 2D class averages of negative-stain EM images (Extended Data Fig. 1). Thus, mechanisms for biased agonism in EGFR that require strict stereochemical coupling between the extracellular ligand-binding module and the transmembrane helices appear to be ruled out. Recently published cryo-EM analyses of intact insulin receptor and intact type 1 insulin-like growth factor receptor also did not resolve the transmembrane and intracellular modules[22–24], suggesting that a lack of rigidity in the connection

between the extracellular module and the transmembrane helices is a feature common to receptor tyrosine kinases.

Cryo-EM single-particle analysis reveals a range of conformations for the EGF-bound extracellular module

After focusing the cryo-EM reconstructions on the extracellular module alone, the dataset for the full-length EGFR:EGF complex reconstituted in peptidiscs yielded the highest resolution reconstruction of the extracellular module, with an overall resolution of 2.9 Å (Extended Data Fig. 3). In this reconstruction, cryo-EM density for the Domain IV legs is less well resolved than for the heads. To better understand the origin of this structural heterogeneity, we used 3D classification to further partition the entire dataset of 765,883 particles into 10 subclasses. Cryo-EM reconstructions corresponding to individual subclasses resulted in overall resolutions ranging between 3.1 Å and 3.7 Å (Extended Data Fig. 3, 4).

The resulting structures for different subclasses differ from one another in the extent of a "scissor-like" rotation between the two ligand-binding heads of the dimeric extracellular module (Fig. 2a-c, Supplementary Videos 1-3). The rotation between the two ligand-binding heads is due to a rigid-body rotation of each head about an axis parallel to the dimerization arms, and is correlated with a slight twist in the dimerization arm of each EGFR subunit (Fig. 2a-c). The extent of this intersubunit rotation can be estimated by defining a dihedral angle corresponding to rotation about a virtual bond connecting the Cα atoms of Thr 249 in each EGFR subunit (in this study, the numbering of amino acids in EGFR starts after the 24-residue signal sequence). The Cα atoms of Ile 190 in each subunit are used to further specify the dihedral angle, and the value of this dihedral angle ranges from ~10° to ~25° in the 10 representative structures (Fig. 2a-c).

The scissor-like rotation between ligand-binding heads in the cryo-EM structures is reminiscent of differences noted previously between crystal structures of fragments of EGFR extracellular modules bound to different ligands[25]. At one extreme, the EM-derived structure with the largest intersubunit rotation (~25°) resembles closely the crystal structure of the ligand-binding head of EGFR bound to EGF (PDB code: 3NJP[7,10], pair-wise Cα RMSDs calculated over EGFR residues 1-501: 0.8 Å, Extended Data Fig. 5a). The EM-derived structure with the smallest intersubunit rotational angle (~10°) resembles the crystal structure of the EGFR ligand-binding head bound to TGF-α (PDB code: 1MOX[8], pair-wise Cα RMSDs calculated over EGFR residues 1-501: 1.5 Å, Extended Data Fig. 5b). Thus, the partitioning of the cryo-EM data into subclasses reflects an intrinsic conformational variability of the ligand-bound EGFR extracellular module. Importantly, these results demonstrate that the conformations ascribed previously to specific binding of EGF or TGF-α to the receptor are actually part of an ensemble of conformations that are readily accessible to EGFR when bound to either one of these ligands.

Different dimeric conformations of the ligand-bound extracellular module have different separations between the ends of the two legs

The two Domain IV legs in the dimer behave differently when the scissor-like rotational angle between the ligand-binding heads changes. The structure of the Domain IV leg from a previously determined crystal structure of the EGFR extracellular module (PDB code: 3NJP[10]) can be docked into our cryo-EM densities reasonably well (Extended Data Fig. 6a, b). This allowed us to determine the position and orientation of the Domain IV legs associated with each representative conformation from the ensemble revealed by cryo-EM. When the rotational angle between the two heads is at its largest value (~25°), the two legs adopt similar conformations, and the membrane-

proximal ends of the two legs are juxtaposed, with a separation of ~5 Å between the Cα atoms of the last residue in the legs (Thr 614) (the "juxtaposed" conformation, Fig. 2a, d, Extended Data Fig. 7). When the rotation between the two subunits is small, maintenance of the same conformation in both legs would lead to steric clashes between the two legs (Extended Data Fig. 6c). The clashes are relieved by one of the two Domain IV legs undergoing a hinge rotation around the connection between the head and the leg, and we refer to this leg as "flexible". This hinge rotation causes the flexible leg to swing upwards, closer to the ligand-binding head, increasing the separation between the membrane-proximal ends of the two Domain IV legs in the dimer. When the angle between the heads is at its minimal value (~10°), the flexible leg swings maximally towards the ligand-binding head, increasing the distance between the membrane-proximal ends of the extracellular module by nearly 15 Å (the "separated" conformation, Fig. 2c, d, Extended Data Fig. 7).

An oncogenic mutation in the kinase domain stabilizes the separated conformation of the extracellular module

We wondered whether the ~15 Å difference in the spacing between the membrane-proximal ends of the extracellular module seen in the juxtaposed and separated conformations could be associated with differences in intracellular kinase activity. We examined an oncogenic variant of EGFR seen frequently in cancer patients, in which a leucine residue in the activation loop of the kinase domain is substituted by arginine (L834R; this is residue 858 when the signal sequence is included). The activation loop is a key allosteric control element at the active site of the kinase domain, and the L834R mutation promotes the formation of the catalytically-active asymmetric dimer of the kinase domains, leading to hyper-activation of EGFR[26,27]. We carried out cryo-EM analysis of full-length

EGFR(L834R) bound to EGF, using the same procedures as for wild-type EGFR. The two structures representing the juxtaposed and the separated conformations of the EGFR(L834R):EGF complex were determined at resolution of 3.3 Å and 3.4Å, respectively (Extended Data Table 1).

The overall structures of the juxtaposed and separated conformations of the EGFR(L834R):EGF complex are very similar to that seen for the corresponding complex of the wild-type receptor. The effect of the activating mutation, L834R, becomes apparent when comparing the quality of the cryo-EM densities in the Domain IV legs for the wild-type and mutant EGFR. For the juxtaposed conformations, the cryo-EM densities are of comparable quality (Extended Data Fig. 8a, b). An important result emerges when comparing the separated conformations for wild-type EGFR and the oncogenic variant. Although the oncogenic L834R mutation involves a residue that is located in the intracellular kinase domain, this mutation markedly stabilizes the flexible leg in the separated conformation of the extracellular module, compared to wild-type EGFR. The cryo-EM density for the flexible leg is much better resolved in the EGFR(L834R):EGF complex than for wild-type EGFR, with secondary structural features clearly visible (Fig. 3a, b). Thus, activation of the intracellular kinase domains by oncogenic mutation preferentially stabilizes the flexible leg in the separated conformation of the extracellular module, providing direct evidence for conformational coupling between the extracellular and intracellular modules of EGFR.

EGF and TGF-α differ in their ability to stabilize the flexible Domain IV leg in the separated conformation

We examined if TGF-α differs from EGF in its ability to stabilize different conformations of the extracellular module. We carried out cryo-EM analysis of EGFR bound to TGF-α using the same

procedures as before. Compared to the EGFR:EGF complex, the flexible leg in the separated conformation of the EGFR:TGF-α complex is destabilized. The cryo-EM density for the flexible leg is poorly defined, and there is no interpretable density for this leg in the reconstruction of the separated conformation of the EGFR:TGF-α complex (Fig. 3c). For the juxtaposed conformation, the reconstructions for the EGFR:EGF and EGFR:TGF-α complexes are similar (Extended Data Fig. 8b, c), with near-symmetrical densities for the Domain IV legs in both complexes. The only noticeable difference is that the cryo-EM density at the very end of the Domain IV legs in the EGFR:TGFα complex is slightly weaker than the corresponding density in the EGFR:EGF complex (Extended Data Fig. 8c).

The structures we have determined provide a plausible explanation for why EGF and TGF-α differ in their ability to stabilize the flexible leg in the separated conformation. Several key interactions with Domain I of the receptor are different between EGF and TGF-α (Fig. 4a). As a result, the ligand-binding pocket is slightly compressed when TGF-α is bound. This compression between Domains I and III causes Domain II to bend in the TGF-α complex relative to its structure in the EGF complex (Fig. 4a), leading to a small difference in the orientation of the dimerization arm with respect to Domains III and IV. EGF and TGF-α both maintain the same interface between the dimerization arms of the two EGFR subunits in the dimer, but the additional bending of Domain II in the TGF-α complex requires that the two Domain IV legs move closer to one another. In the separated conformation of the TGF-α complex, this inward motion would cause a steric clash between the two legs if the legs maintained the conformation seen in the EGF complex (Fig. 4b). The increased disorder in the Domain IV leg when TGF-α is bound appears to be a result of this close contact being avoided.

These results demonstrate that TGF-α is less effective than EGF at stabilizing the legs-separated conformation of the extracellular module. Our analysis of the oncogenic L834R variant shows that stabilization of the flexible leg in the separated conformation is correlated with increased levels of activity in the mutant EGFR. We speculate that the observed differences in the ability of EGF and TGF-α to stabilize the Domain IV leg in the separated conformation underlies the ability of these two ligands to trigger different levels of kinase activity from EGFR.

The legs-separated conformation of the dimeric extracellular module can couple more readily to the N-terminal dimer of the transmembrane helices

The EGFR transmembrane helix contains two dimerization motifs, one located toward the N-terminal end of the transmembrane helix, near the extracellular face of membrane, and the other located toward the C-terminal end[28]. Dimerization of the transmembrane helices through the N-terminal dimerization motif is associated with increased kinase activity[29,10,16,30] and so our cryo-EM results suggest that the legs-separated conformation of the extracellular module is coupled to the N-terminal dimer of the transmembrane helices. To understand how this coupling occurs, we analyzed two structures of EGFR transmembrane helix dimers determined previously by NMR. One structure shows N-terminal association of the transmembrane helices (PDB code: 5LV6[31]) and the other shows C-terminal association (PDB code: 2M0B[32]) (Extended Data Fig. 9a).

The seven-residue linker connecting the extracellular module and the transmembrane helices (Asn 615-Ser 621) is visualized in both structures. The three C-terminal residues of the linker (Ile 619-Ser 621), located at the junction with the transmembrane helix, adopt well-defined conformations in both NMR structures. The hydrophobic sidechain of Ile 619 packs against the sidechain of Ala 623 in the first turn of the transmembrane helix. Ile 619 is followed by a 3_{10}-

helical turn spanning Pro 620 to Ala 623 (Extended Data Fig. 9b). These structural constraints cause the linker-transmembrane helix junction to orient parallel to the membrane surface in both structures (Extended Data Fig. 9a). As a consequence, in the N-terminal dimer of the transmembrane helices, the two linkers are oriented away from each other, with the two Ile 619 residues separated by ~20 Å. In contrast, the C-terminal dimer has the two linkers oriented towards each other, with the distance between the two Ile 619 residues at ~10 Å. Thr 614 is at the junction between the linker and the Domain IV leg of the extracellular module. In the ensemble of structures for the N-terminal dimer of the transmembrane helices[31], the two threonine residues (Thr 614) are separated by an average distance of ~35 Å (Extended Data Fig. 9a). The separation between the two threonine residues is ~5 Å in the cryo-EM reconstruction of the juxtaposed conformation, and ~20 Å in the separated conformation. This suggests that the N-terminal dimer of the transmembrane helices may couple more readily to the separated conformation of the extracellular module, rather than to the juxtaposed conformation.

We used molecular dynamics simulations to explore the constraints on the separation between the two threonine residues at position 614 in the linker. We generated a 175 ns trajectory for the transmembrane helices with the linkers present, embedded in a neutral POPC membrane, starting from one of the structures in the NMR-based ensemble for the N-terminal dimer conformation (PDB code: 5LV6[31]). We then initiated twenty-five independent molecular dynamics trajectories from the end of the first trajectory, each 100 ns long, for a total simulation time of 2.5 μs (Fig. 5a). The conformation of the junction between the linker and the transmembrane helices is very stable, with the sidechain of Ile 619 remaining associated with Ala 623 over the course of the simulations (Extended Data Fig. 9c). The linkers explore a range of conformations but are oriented predominantly such that they point away from one another, and the

distance between Thr 614 from each subunit ranges between 20 Å and 35 Å (Fig. 5b). These results confirm that the junction between the linker and the transmembrane helices imposes a constraint on the orientation of the linkers that favors its connection to the separated conformation of the extracellular module.

Concluding Remarks

Our cryo-EM analysis provides a molecular explanation for biased agonism in the response of EGFR to EGF and TGF-α (Fig. 5c, Extended Data Fig. 10). The differential activation of EGFR by these two ligands arises from structural differences in the receptor-ligand interactions. As a result, EGF and TGF-α differ in their abilities to stabilize the leg-separated conformation of the extracellular module that couples to the N-terminal dimer conformation of the transmembrane helices, which is further coupled to a specific conformation of the juxtamembrane segments[16,33–35], ultimately promoting the formation of kinase dimers with higher activity[36]. In this way, EGF and TGF-α can generate biased EGFR signaling by differentially modulating the structure of the junction point between the extracellular module and the transmembrane helices, altering the balance between alternative dimeric conformations of the transmembrane helices (Fig. 5c). A lack of rigidity in the linkage between the extracellular module and the transmembrane helices is emerging as a common feature in receptor tyrosine kinases[10,22–24], and the mechanism we have defined could potentially be utilized by other receptor tyrosine kinases to generate ligand-specific signaling outputs[37].

Members of the EGFR family are important targets for the development of cancer therapeutics, including antibodies that block receptor dimerization and activation[38,39]. Our structural analysis suggests new strategies for the modulation of EGFR signaling. For example,

antibodies or peptides that can bind to the Domain IV legs and stabilize the juxtaposed conformation would allow the receptor to dimerize but send weaker signals. Such a strategy may circumvent resistance mechanisms that arise in response to antibodies and inhibitors that shut down EGFR signaling completely[40,41,38].

Methods

Cell culture

Insect cells

Spodoptera frugiperda Sf9 cells were maintained as suspension cultures at 27 ^{o}C in ESF-921™ Insect Cell Culture Medium (Expression Systems). Isolation of bacmid DNA, transfection of Sf9 cells, and amplification of baculovirus were performed following the protocols detailed in the Bac-to-Bac Baculovirus Expression System (ThermoFisher Scientific).

Mammalian cells

Human HEK293S GnTI^{-} cells were cultured as suspension at 37 ^{o}C in FreeStyle293 expression medium (Invitrogen) supplemented with 2% (vol/vol) FBS, in the presence of 8% CO_2. The HEK293S GnTI^{-} suspension cells were maintained between 0.2 and 3 x 10^6 cells/ml.

Full-length EGFR protein expression and purification

Protein expression

A cDNA fragment encoding full-length human EGFR (UniProt accession no. P00533) was cloned into the pEG_BacMam vector[42]. To facilitate affinity purification, a FLAG-tag (with sequence DYKDDDDK) was inserted after the signal sequence of EGFR. To monitor protein

expression and to reduce non-specific degradation of the C-terminal tail of EGFR, EGFP was fused after the C-terminus of EGFR. An oncogenic EGFR variant (L834R) was generated by site-directed mutagenesis.

Plasmid DNA carrying the EGFR sequence was transformed into DH10Bac competent cells (ThermoFisher Scientific) to produce the bacmid DNA, following the protocol detailed in the Bac-to-Bac Baculovirus Expression System. To produce the recombinant baculovirus, purified bacmid DNA was transfected into Sf9 cells using Cellfectin II (ThermoFisher Scientific). After two rounds of amplification, the recombinant baculovirus was used to infect HEK293S GnTI$^-$ cells at a 1:20 (vol/vol) ratio. After 24 hours, the infected HEK293S GnTI$^-$ suspension culture was supplemented with 2 μM of Erlotinib, an EGFR inhibitor, and 8 mM sodium butyrate, for optimal protein expression and stability. Cells were harvested by centrifugation 72 hours after the infection.

Protein purification

Harvested cells were resuspended in lysis buffer (25 mM Tris pH 8.0, 150 mM NaCl, 2 μM Erlotinib and cOmplete Protease Inhibitor Cocktail (Roche)) and were lysed by sonication in pulse mode for 10 minutes, with 2s intervals between each 1s pulse. Large debris and unbroken cells were removed from the crude lysate by centrifugation for 12 mins at 19,000g. The membrane fraction was pelleted by ultra-centrifugation at 190,000g for 1 hr and resuspended in lysis buffer. To extract EGFR from the membrane, 1% DDM (Dodecyl-β-D-Maltoside, Anatrace) and 0.1% CHS (Cholesteryl Hemisuccinate, Anatrace) were added to the suspended membrane fraction. After mixing for 2 hrs at 4 ^{o}C, detergent solubilized proteins were collected by ultra-centrifugation at 190,000g for 30 mins. The supernatant was incubated with Anti-FLAG M2 Affinity agarose gel (Sigma-Aldrich) for 4 hrs at 4 ^{o}C. The gel was washed with lysis buffer supplemented with 0.05%

DDM and 0.005% CHS. EGFR was eluted from the gel by lysis buffer containing 0.16 mg/ml FLAG peptide, 0.05% DDM and 0.005% CHS.

Affinity-purified full-length EGFR was concentrated to 50 μM and treated with 100 μM human EGF (Novoprotein) or TGF-α (Sino Biological) for at least 5 mins on ice, before further fractionation by size-exclusion chromatography using a Superose 6 Increase 10/300 GL column (GE Healthcare), with buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, 2 μM Erlotinib, 0.02% DDM and 0.002% CHS). The peak fractions containing ligand-bound dimeric EGFR were collected and immediately used for reconstitution and the reconstituted receptor was applied to EM grids. The fractions were also flash-frozen in liquid nitrogen and used for reconstitution at later times. Freezing and thawing the fractions more than once leads to deterioration in cryo-EM sample quality.

Reconstitution of ligand-bound full-length EGFR complex

Reconstitution into lipid nanodiscs

Purified protein complexes of EGFR bound to EGF or TGF-α were reconstituted into lipid nanodiscs as previously described, with modifications[43]. Briefly, membrane scaffold proteins MSP1D1 and MSP1E3D1 were expressed and purified from *Escherichia coli*. 10 mg soybean PC (L-α-phosphatidylcholine (soy), Avanti) dissolved in chloroform was dried using nitrogen stream and residual chloroform was further removed by desiccation under vacuum overnight. Rehydration buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, 0.2 mM DTT, 1% DDM and 1% sodium cholate was added to the dried lipid and vortexed briefly, resulting in a semi-clear solution at 10 mM lipid concentration. After 3 cycles of freezing (in liquid nitrogen)-and-thawing (warm water bath), the lipid solution was bath sonicated for 1 hr. The resulting clear lipid stock (10 mM) was

immediately used for lipid nanodiscs assembly or stored at -80 ^{o}C in small aliquots. Purified complexes of EGFR bound to EGF or TGF-α (~10 μM, solubilized in 0.02% DDM and 0.002% CHS) was supplemented with 20 μM extra EGF or TGF-α to keep the ligand-binding site on the receptor saturated, before it was mixed with the soybean PC stock and membrane scaffold proteins at molar ratios of 1:6:240 (EGFR monomer:MSP1D1:soy PC) and 1:6:600 (EGFR monomer:MSP1E3D1:soy PC). The reaction mixture (~500 μl) was incubated on ice for 30 mins, before 20 mg Bio-beads SM2 absorbent (Bio-Rad) were added to initiate the removal of detergents from the mixture and to start the nanodiscs assembly process. The mixture was incubated at 4 ^{o}C by continuous rotation. After 1 hour, a second batch of Bio-beads (20 mg) was added to the mixture to replace the first batch, and the sample was incubated for overnight at 4 ^{o}C. A third and final batch of Bio-beads (20 mg) was added to replace the second batch, 2 hrs before the reconstitution mixture was cleared by removing the Bio-beads through centrifugation. After another centrifugation at 14,000 rpm for 10 mins, the reconstitution mixture was fractionated by size-exclusion chromatograph using Superose 6 Increase 10/300 GL column (GE Healthcare) in detergent-free buffer containing 25 mM Tris pH 8.0, 150 mM NaCl and 5 μM Erlotinib. The peak fraction corresponding to dimeric EGFR:ligands complexes reconstituted in lipid nanodisc was collected and analyzed by SDS-PAGE, before being applied to negative-stain and cryo-EM grids.

Reconstitution into amphipols

Samples containing purified complexes of EGFR bound to EGF or TGF-α (~10 μM, solubilized in 0.02% DDM and 0.002% CHS) were supplemented with 20 μM extra EGF or TGF-α, before the amphipol A8-35 (Anatrace) was added to reach a mass ratio of 1:3 (EGFR:A8-35). After incubation at 4 ^{o}C overnight, free amphipol molecules were removed by size-exclusion

chromatography using a Superose 6 Increase 10/300 GL column in detergent-free buffer containing 25 mM Tris pH 8.0, 150 mM NaCl and 5 μM Erlotinib. Peak fractions corresponding to the reconstituted EGFR:ligand complexes in amphipol were analyzed by SDS-PAGE, before EM studies.

Reconstitution into peptidiscs

As for the reconstitution into amphipols, samples containing purified complexes of EGFR bound to EGF or TGF-α (~10 μM, solubilized in 0.02% DDM and 0.002% CHS) were supplemented with 20 μM extra EGF or TGF-α, before the addition of NSP_r peptide[20] to the samples at a mass ratio of 1:1.5 (EGFR:NSP_r peptide). After overnight incubation at 4 ^{o}C, the sample was fractionated using a Superose 6 Increase 10/300 GL column (detergent-free buffer: 25 mM Tris pH 8.0, 150 mM NaCl and 5 μM Erlotinib) and free NSP_r peptides were removed during fractionation. Peak fractions corresponding to the reconstituted EGFR:ligand complexes in peptidiscs were analyzed by SDS-PAGE, before applied to EM grids.

Preparation of electron microscopy (EM) grids and data acquisition

For negative-stain EM, a 5 μL sample of reconstituted full-length human EGFR (~0.05 mg/mL protein in 25 mM Tris pH 8.0, 150 mM NaCl and 2 μM Erlotinib) was placed on the continuous-carbon side of a glow-discharged copper grid (Ted Pella, Redding, CA, USA), and the excess sample was removed by wicking with filter paper after 1 min incubation. The bound particles were stained by floating the grids on four consecutive 30 μL droplets of freshly prepped 2% uranyl formate solution and incubating the grids with each droplet for 15 s. The excess stain was removed by blotting with filter paper and the grids were air-dried for at least 5 min. Images

of negative-stained full-length human EGFR reconstituted in lipid nanodisc, amphipol, and peptidisc were recorded on a 4049 x 4096 pixels CMOS camera (TVIPS TemCam-F416) using the automated Leginon data collection software[44]. Samples were imaged using a Tecnai 12 transmission electron microscope (FEI, Hillsboro, OR, USA) at 120 keV with a nominal magnification of 49,000x (2.18 Å calibrated pixel size at the specimen level) using a defocus range of -0.8 to -1.5 μm. All data were acquired under low dose conditions, allowing a dose at around 35 $e^-/Å^2$.

For cryo-EM, UltrAuFoil R 1.2/1.3 300 mesh grids (Electron Microscopy Sciences, Q350AR13A) were rendered hydrophilic by pretreatment with a O_2/H_2 gas mixture using a Solarus 950 plasma cleaning system (Gatan, Inc., Pleasanton, CA) for 15s at 50W. After application of 3 μL sample of reconstituted full-length human EGFR (~0.5 mg/ml protein in 25 mM Tris pH 8.0, 150 mM NaCl and 2 uM Erlotinib), grids were plunge-frozen using a Leica EM GP (Leica Microsystems Inc., Buffalo Grove, IL) with blotting time of 8 seconds, chamber temperature and humidity of +4°C and 95%, and a liquid ethane temperature of -184°C. The grids were imaged using a Titan Krios transmission electron microscope (Thermo Fisher Scientific, Waltham, MA) operated at 300 kV. Images were recorded on a K3 camera (Gatan, Inc.) operated in super-resolution counting mode with a physical pixel size of 0.54 Å. The detector was placed at the end of a GIF Quantum energy filter (Gatan, Inc.), operated in zero-energy-loss mode with a slit width of 20 eV. Due to strongly preferred particle orientations, all movies were collected at 40° tilt, at a nominal magnification of 81,000x and a defocus range between -0.75 and -1.5 μm., using the automated data collection software Latitude (Gatan, Inc.). The nominal dose per frame was ~1 $e^-/Å^2$ and the total dose per movie is ~50 $e^-/Å^2$. For the full-length EGFR:EGF complex reconstituted in peptidiscs, a total of 4,331 movies were collected. For the full-length EGFR:TGF-α complex

reconstituted in peptidiscs, a total of 7,406 movies were collected. For the full-length EGFR(L834R):EGF complex reconstituted in peptidiscs, a total of 5,013 movies were collected.

EM image processing

For negative-stain EM images of full-length EGFR:EGF complexes reconstituted in lipidic nanodiscs (MSP1E3D1), the initial image processing and classification steps were performed using the Appion image processing environment[45]. Particles were first selected from micrographs using DoG Picker[46]. The contrast transfer functions (CTFs) of the micrographs were estimated using the CTFFIND3[47]. CTF correction of the micrographs was performed by Wiener filter using ACE2[48]. A total of 112,408 particles were extracted using a 192 x 192 box size and binned by a factor of 2. Each particle was normalized to remove pixels whose values were above or below 4.5 σ of the mean pixel value using the XMIPP normalization program[49]. In order to remove incorrectly selected protein aggregates or other artifacts, particles with extreme intensity values were removed. The remaining 84,260 particles were subjected to 2D iterative reference-free alignment and classification using IMAGIC multi-reference alignment (MRA) and a topology-representing network classification[50,51], to produce the 200 2D class averages.

For Cryo-EM images of full-length EGFR:EGF complex reconstituted in peptidiscs, Extended Data Fig. 3 shows a summary of the image processing workflow. In brief, movies were motion-corrected using the MotionCor2 wrapper[52] in Relion3[53]. Corrected micrographs were imported to cryoSPARC v2[54,55] for further processing. After local CTF determination using Patch-Based CTF Estimation, a total of 4,234,635 particles from 4,431 micrographs were automatically selected using template-free Blob_picker and were extracted with a box size of 320 x 320 pixels. Due to strongly preferred particle orientations, a relatively low threshold value was used during

the particle picking to include particles with less contrast than the preferred views. The majority of the initial 4,234,635 picks are from the background. After iterative 2D classification, most of the background particles were removed based on 2D class averages. The remaining 1,587,872 particles were further cleaned-up by ab-initio reconstruction into 4 classes. The resulting 3D maps showed that 3 of the ab-initio reconstruction classes were mostly "junk" particles without recognizable structural features, and the other class showed clear resemblance to the dimeric extracellular module of EGFR. This clean class of particles (814,089) were selected and subjected to a final round of 2D classification. After visual inspection of the final 2D class averages, 765,883 particles were retained and ab-initio reconstruction was performed to generate the initial model, which was used as the reference for a round of homogeneous refinement as well as a non-uniform refinement[56]. Based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion, the overall resolution of the reconstructions from homogeneous refinement and non-uniform refinement are determined as 3.0 Å and 2.9 Å, respectively. After visual inspection of both reconstructions, the map generated from non-uniform refinement with better quality was chosen for further analysis and used as the reference in the subsequent round of heterogeneous refinement (3D classification) using the 765,883 particles, resulting in 10 classes. A final round of homogeneous refinement and non-uniform refinement were performed for each of the resulting 10 classes from the heterogeneous refinement run. After visual inspection, the 10 maps resulting from non-uniform refinement were chosen for model building and refinement for each class. The overall resolution of these 10 maps ranges between 3.1 Å and 3.7 Å.

Cryo-EM data processing of full-length EGFR(L834R):EGF complex and full-length EGFR:TGF-α complex reconstituted in peptidiscs followed similar procedures described above. For full-length EGFR(L834R):EGF complex, after local CTF estimation by Patch-Based CTF

Estimation, a total of 3,250,237 particles were automatically picked by Blob_Picker. After iterative 2D classification and ab-initio reconstruction, background and spurious particles were removed. 987,367 cleaned particles were subjected to homogeneous as well as non-uniform refinement, using the initial model generated from ab-initio reconstruction as reference. After visual inspection, the map generated by non-uniform refinement (overall resolution: 3.0 Å) was chosen as the reference for the subsequent heterogeneous refinement, resulting in 10 classes. A final round of non-uniform refinement was performed for each of the 10 classes and the resulting 10 maps were used for model building and refinement. The resolutions of these 10 maps were determined to be in the range of 3.2-4.2 Å.

For the full-length EGFR:TGF-α complex, a total of 6,846,834 particles were initially picked by Blob_Picker. Iterative 2D classification and ab-initio reconstruction were performed to remove background and spurious particles, and 1,019,024 cleaned particles were retained. These cleaned particles were subjected to homogeneous refinement and non-uniform refinement. The map resulting from non-uniform refinement was at an overall resolution of 3.0 Å and it was used as the reference for the subsequent heterogeneous refinement, resulting in 10 classes. A final round of non-uniform refinement was performed for each of the 10 classes and the resulting 10 maps were determined at resolutions between 3.2 Å and 4.2 Å. These maps were used for model building and refinement.

3D variability analysis, local resolution estimation, and B-factor sharpening of all refined maps were performed using the cryoSPARC v2.

Model building and refinement

To generate models for different conformations of the EGFR(WT):EGF and EGFR(L834R):EGF complexes, a crystal structure of the isolated EGFR extracellular module bound to EGF (PDB code: 3NJP) was split into four fragments: two heads and two legs. The two heads contain EGFR residues 1-500 and EGF residues 5-51 from each side of the dimeric structure, respectively. The two legs contain EGFR residues 501-614 from each EGFR molecule, respectively. These four fragments were fit into the reconstructed cryo-EM maps of the EGFR(WT):EGF and EGFR(L834R):EGF complexes, using UCSF Chimera[57]. UCSF chimera-fitted models were subjected to one round of real-space refinement[58] in *Phenix*[59]. To prevent overfitting due to the lower local resolution in the map region corresponding to the flexible Domain IV legs, the coordinates of the flexible Domain IV leg was not refined against the cryo-EM maps. The models were manually adjusted in *Coot*[60], followed by iterative rounds of real-space refinement in *Phenix* and manual fitting in *Coot*. Model quality was assessed using the comprehensive model validation tools from *Phenix* (Extended Data Table 1). All maps and models were visualized using UCSF Chimera and PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Similar procedures were followed to generate models for the different conformations of the EGFR:TGF-α complex, except that a crystal structure of the truncated EGFR extracellular domains bound to TGF-α (PDB code: 1MOX) was used to generate the fragments of the heads for the initial model fitting in UCSF Chimera.

Molecular dynamics simulations

System preparation for molecular dynamics simulations

We built a system of the two transmembrane helices (residues 610-653) associating at the N-terminal ends starting from model 1 of the NMR structure (PDB code: 5LV6[31]). We used the Membrane Builder module on the CHARMM-GUI website[61,62] to insert the protein into a lipid bilayer comprising of neutral POPC lipids. The x- and y- dimensions of the bilayer was set to 100 Å each. This lipid bilayer comprised of 148 lipids in the upper leaflet, and 146 lipids in the lower leaflet. The N-termini of each protein chain were capped with acetyl groups and the C-termini were capped with methyl groups. The system was solvated such that the solvent layer extended to 20 Å beyond the lipid bilayer on either side, and ions were added such that the ionic strength was 0.1 M, using VMD[63]. The final edge dimensions of this system were 103 Å x 103 Å x 118 Å.

Minimization, equilibration and production protocols

First the energy was minimized while holding the protein, waters, ions, and lipid heads fixed, and allowing only the lipid tails to move, for 1000 steps, then while holding the protein and lipid heads fixed for 5000 steps, then while holding only the protein fixed for 1000 steps, and finally for a 1000 steps while allowing all the atoms to move. This system was then equilibrated using molecular dynamics for 5 ns while holding the protein fixed, to allow the lipids to relax about the protein. A longer trajectory of 175 ns then was generated from this equilibrated structure, during which all the atoms were allowed to move. During the course of this simulation, the N-terminal ends of the linkers (residues 610-613), which are disordered in this structure, interacted frequently with each other. In the full-length structure, these residues are a part of the ordered region in Domain IV of the extracellular module and would not be available to form such interactions. We therefore deleted these 4 residues, such that the protein chains started at Thr 614, and capped these N-terminal ends with acetyl groups. The energy of this system was minimized with the same multi-step protocol as before and used to initiate 25 independent molecular

dynamics trajectories. These simulations were run for 100 ns, for a total simulation time of 2.5 μs. All equilibration and production simulations were performed with the temperature and pressure held constant at 300 K and 1 atm, respectively.

The NAMD package was used to run the minimization and equilibration simulations[64] with the CHARMM36m force field[65]. The velocity Verlet algorithm was used to calculate the trajectories of the atoms. A time step of 2 fs was used. Particle Mesh Ewald was used to calculate long-range electrostatic interactions, with a maximum space of 1 Å between grid points[66]. Long-range electrostatics were updated at every time step. Van der Waals interactions were truncated at 12 Å. Hydrogen atoms bonded to the heavy atoms were constrained using the ShakeH algorithm[67]. The temperature was controlled with Langevin dynamics with a damping coefficient of 1/ps, applied only to non-hydrogen atoms. Pressure was controlled by the Nose-Hoover method with the Langevin piston, with an oscillation period of 200 fs and a damping time scale of 50 fs[68,69]. Distance variations over the course of the simulations were calculated with the AmberTools18 package[70].

References

1. Leahy, D. J. Structure and function of the epidermal growth factor (EGF/ErbB) family of receptors. *Adv Protein Chem* **68,** 1–27 (2004).

2. Lemmon, M. A., Schlessinger, J. & Ferguson, K. M. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* **6,** a020768 (2014).

3. Kovacs, E., Zorn, J. A., Huang, Y., Barros, T. & Kuriyan, J. A structural perspective on the regulation of the epidermal growth factor receptor. *Annu Rev Biochem* **84,** 739–764 (2015).

4. Wilson, K. J., Gilmore, J. L., Foley, J., Lemmon, M. A. & Riese, D. J. Functional selectivity of EGF family peptide growth factors: implications for cancer. *Pharmacol Ther* **122,** 1–8

(2009).

5. Singh, B., Carpenter, G. & Coffey, R. J. EGF receptor ligands: recent advances. [version 1; peer review: 3 approved]. *F1000Res* **5,** (2016).

6. Ronan, T. *et al.* Different epidermal growth factor receptor (EGFR) agonists produce unique signatures for the recruitment of downstream signaling proteins. *J Biol Chem* **291,** 5528–5540 (2016).

7. Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110,** 775–787 (2002).

8. Garrett, T. P. J. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* **110,** 763–773 (2002).

9. Dawson, J. P. *et al.* Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol Cell Biol* **25,** 7734–7742 (2005).

10. 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).

11. Freed, D. M. *et al.* EGFR ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell* **171,** 683–695.e18 (2017).

12. Scheck, R. A., Lowder, M. A., Appelbaum, J. S. & Schepartz, A. Bipartite tetracysteine display reveals allosteric control of ligand-specific EGFR activation. *ACS Chem Biol* **7,** 1367–1376 (2012).

13. Jones, J. T., Akita, R. W. & Sliwkowski, M. X. Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Lett* **447,** 227–231 (1999).

14. Doerner, A., Scheck, R. & Schepartz, A. Growth Factor Identity Is Encoded by Discrete Coiled-Coil Rotamers in the EGFR Juxtamembrane Region. *Chem Biol* **22,** 776–784 (2015).

15. Lowder, M. A., Doerner, A. E. & Schepartz, A. Structural Differences between Wild-Type and Double Mutant EGFR Modulated by Third-Generation Kinase Inhibitors. *J Am Chem Soc* **137,** 6456–6459 (2015).

16. Sinclair, J. K. L., Walker, A. S., Doerner, A. E. & Schepartz, A. Mechanism of Allosteric Coupling into and through the Plasma Membrane by EGFR. *Cell Chem. Biol.* **25,** 857–870.e7 (2018).

17. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative

HEK293S stable mammalian cell line. *Proc Natl Acad Sci U S A* **99,** 13419–13424 (2002).

18. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett* **584,** 1721–1727 (2010).

19. Gohon, Y. *et al.* Well-defined nanoparticles formed by hydrophobic assembly of a short and polydisperse random terpolymer, amphipol A8-35. *Langmuir* **22,** 1281–1290 (2006).

20. Carlson, M. L. *et al.* The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *elife* **7,** (2018).

21. Angiulli, G. *et al.* New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins. *elife* **9,** (2020).

22. Uchikawa, E., Choi, E., Shang, G., Yu, H. & Bai, X.-C. Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *elife* **8,** (2019).

23. Li, J., Choi, E., Yu, H. & Bai, X.-C. Structural basis of the activation of type 1 insulin-like growth factor receptor. *Nat Commun* **10,** 4567 (2019).

24. Zhang, X. *et al.* Visualization of Ligand-Bound Ectodomain Assembly in the Full-Length Human IGF-1 Receptor by Cryo-EM Single-Particle Analysis. *Structure* **28,** 555–561.e4 (2020).

25. Liu, P. *et al.* A single ligand is sufficient to activate EGFR dimers. *Proc Natl Acad Sci U S A* **109,** 10861–10866 (2012).

26. Shan, Y. *et al.* Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* **149,** 860–870 (2012).

27. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proc Natl Acad Sci U S A* **110,** E3595–604 (2013).

28. Mendrola, J. M., Berger, M. B., King, M. C. & Lemmon, M. A. The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J Biol Chem* **277,** 4704–4712 (2002).

29. Miloso, M., Mazzotti, M., Vass, W. C. & Beguinot, L. SHC and GRB-2 are constitutively by an epidermal growth factor receptor with a point mutation in the transmembrane domain. *J Biol Chem* **270,** 19557–19562 (1995).

30. Pahuja, K. B. *et al.* Actionable activating oncogenic ERBB2/HER2 transmembrane and juxtamembrane domain mutations. *Cancer Cell* **34,** 792–806.e5 (2018).

31. Bocharov, E. V. *et al.* The conformation of the epidermal growth factor receptor

transmembrane domain dimer dynamically adapts to the local membrane environment. *Biochemistry* **56,** 1697–1705 (2017).

32. Bocharov, E. V. *et al.* Alternative packing of EGFR transmembrane domain suggests that protein-lipid interactions underlie signal conduction across membrane. *Biochim Biophys Acta* **1858,** 1254–1261 (2016).

33. Arkhipov, A. *et al.* Architecture and membrane interactions of the EGF receptor. *Cell* **152,** 557–569 (2013).

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

35. Jura, N. *et al.* Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **137,** 1293–1307 (2009).

36. 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).

37. Ahmed, F., Paul, M. & Hristova, K. The biophysical basis of receptor tyrosine kinase ligand functional selectivity: Trk-B case study. *Biochem J* (2020). doi:10.1042/BCJ20200671

38. Cai, W.-Q. *et al.* The latest battles between EGFR monoclonal antibodies and resistant tumor cells. *Front Oncol* **10,** 1249 (2020).

39. Arteaga, C. L. & Engelman, J. A. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25,** 282–303 (2014).

40. Brand, T. M., Iida, M. & Wheeler, D. L. Molecular mechanisms of resistance to the EGFR monoclonal antibody cetuximab. *Cancer Biol Ther* **11,** 777–792 (2011).

41. Chong, C. R. & Jänne, P. A. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med* **19,** 1389–1400 (2013).

42. Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat Protoc* **9,** 2574–2585 (2014).

43. Ritchie, T. K. *et al.* Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Meth Enzymol* **464,** 211–231 (2009).

44. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151,** 41–60 (2005).

45. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* **166,** 95–102 (2009).

46. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166,** 205–213 (2009).

47. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142,** 334–347 (2003).

48. Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. ACE: automated CTF estimation. *Ultramicroscopy* **104,** 8–29 (2005).

49. Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc* **3,** 977–990 (2008).

50. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116,** 17–24 (1996).

51. Ogura, T., Iwasaki, K. & Sato, C. Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *J Struct Biol* **143,** 185–200 (2003).

52. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14,** 331–332 (2017).

53. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *elife* **7,** (2018).

54. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14,** 290–296 (2017).

55. Punjani, A. & Fleet, D. J. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *BioRxiv* (2020). doi:10.1101/2020.04.08.032466

56. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: Adaptive regularization improves single particle cryo-EM reconstruction. *BioRxiv* (2019). doi:10.1101/2019.12.15.877092

57. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

58. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74,** 531–544 (2018).

59. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and

electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75,** 861–877 (2019).

60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).

61. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29,** 1859–1865 (2008).

62. Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35,** 1997–2004 (2014).

63. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. of Mol. Graph.* **14,** 33–38 (1996).

64. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26,** 1781–1802 (2005).

65. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14,** 71–73 (2017).

66. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089 (1993).

67. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23,** 327–341 (1977).

68. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101,** 4177 (1994).

69. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **103,** 4613–4621 (1995).

70. Case, D. A. *et al.* AMBER 18. (2018).

Acknowledgments

We thank the Cell Culture Facility at the University of California, Berkeley, for insect cell and mammalian cell cultures. We thank the Eric Gouaux group at the Vollum Institute at the Oregon Health & Science University, for the pEG_BacMam plasmid. This research was in part supported

by the National Cryo-EM Facility of the National Cancer Institute at the Frederick National Laboratory for Cancer Research, under contract HSSN261200800001E. We also acknowledge the Berkeley Bay Area Cryo-EM Facility and the HHMI Janelia Cryo-EM Facility. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1548562. This work is in part supported by an award from the Canada Excellence Research Chair program (to S.S.) and by a philanthropic gift to the VGH and UBC Hospital Foundation (to S.S.).

Author Contributions

Y.H., S.S., and J.K. conceived the project; Y.H. designed and performed experiments; Y.H., J.O., and A.M. performed EM experiments; Y.H. analyzed data; D.K. performed molecular dynamics simulations; S.S. and J.K. supervised the project; Y.H. and J.K. drafted the manuscript; Y.H., S.S., and J.K. edited the manuscript. All authors commented on the manuscript.

Competing interests

J.K. is a cofounder of Nurix Therapeutics and is on the scientific advisory boards of Carmot and Revolution Medicine.

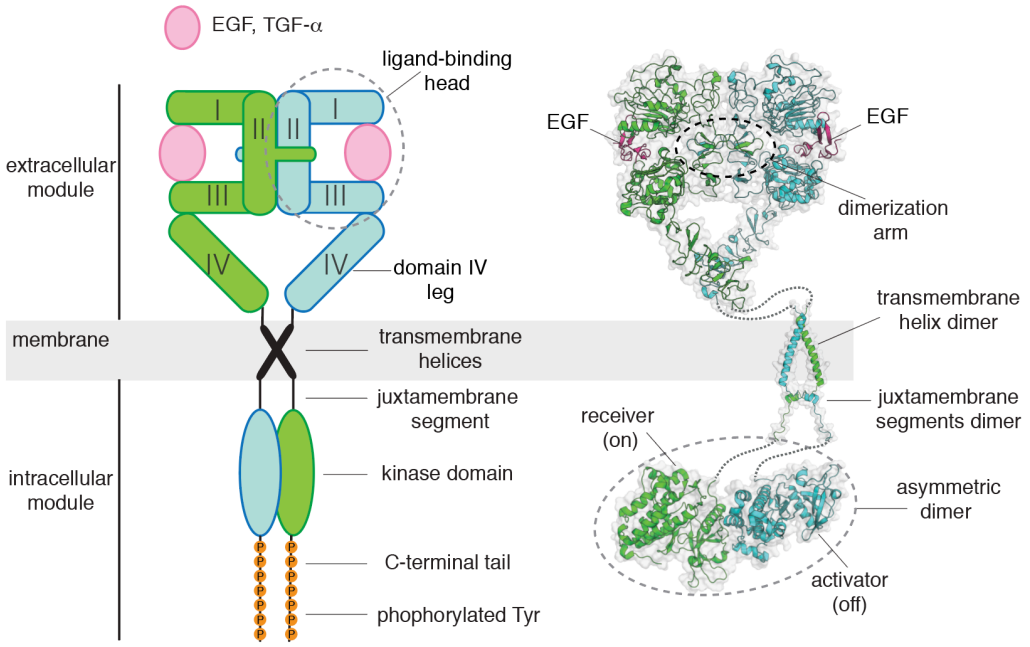
Supplementary information

Supplementary Videos 1-3

3D variability analysis of the EGFR:EGF complex performed in cryoSPARC v2[55], shown from the front (SI_video1), the side (SI_video2), and the top (SI_video3). Note the intersubunit rotation between the two heads and the corresponding movement of the Domain IV legs.

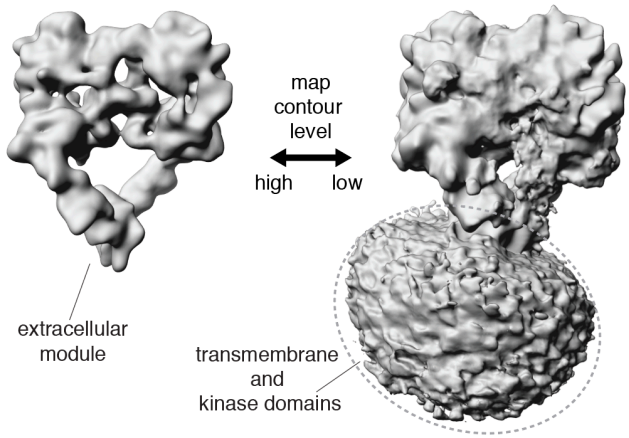
Fig. 1

a



b

Cryo-EM density map of EGFR:EGF complex in nanodiscs



c

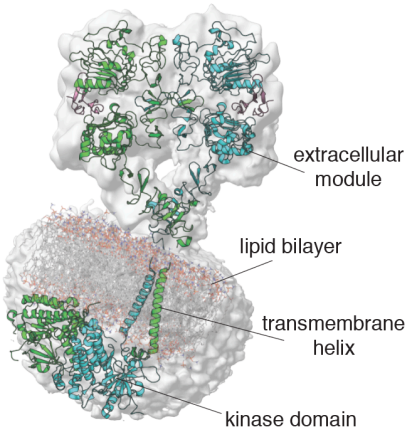


Fig.1 Linkage between the extracellular module and transmembrane helices in full-length human EGFR is not rigid

a, Left, a schematic diagram of ligand-bound dimeric EGFR. Right, structures of EGFR fragments, determined previously: a dimeric extracellular module bound to EGF (PDB code: 3NJP[10]), a transmembrane helix dimer (PDB code: 2M20[34]), and an asymmetric dimer of kinase domains (PDB code: 2GS6[36]).

b, Cryo-EM density map for a full-length EGFR:EGF complex reconstituted in nanodiscs, shown at two contour levels. Note the lack of EM density for the transmembrane and kinase domains of EGFR at higher contour level.

c, Superposition of the cryo-EM map shown in **b** with a model for full-length EGFR embedded in the lipid bilayer, described previously[33].

Fig. 2

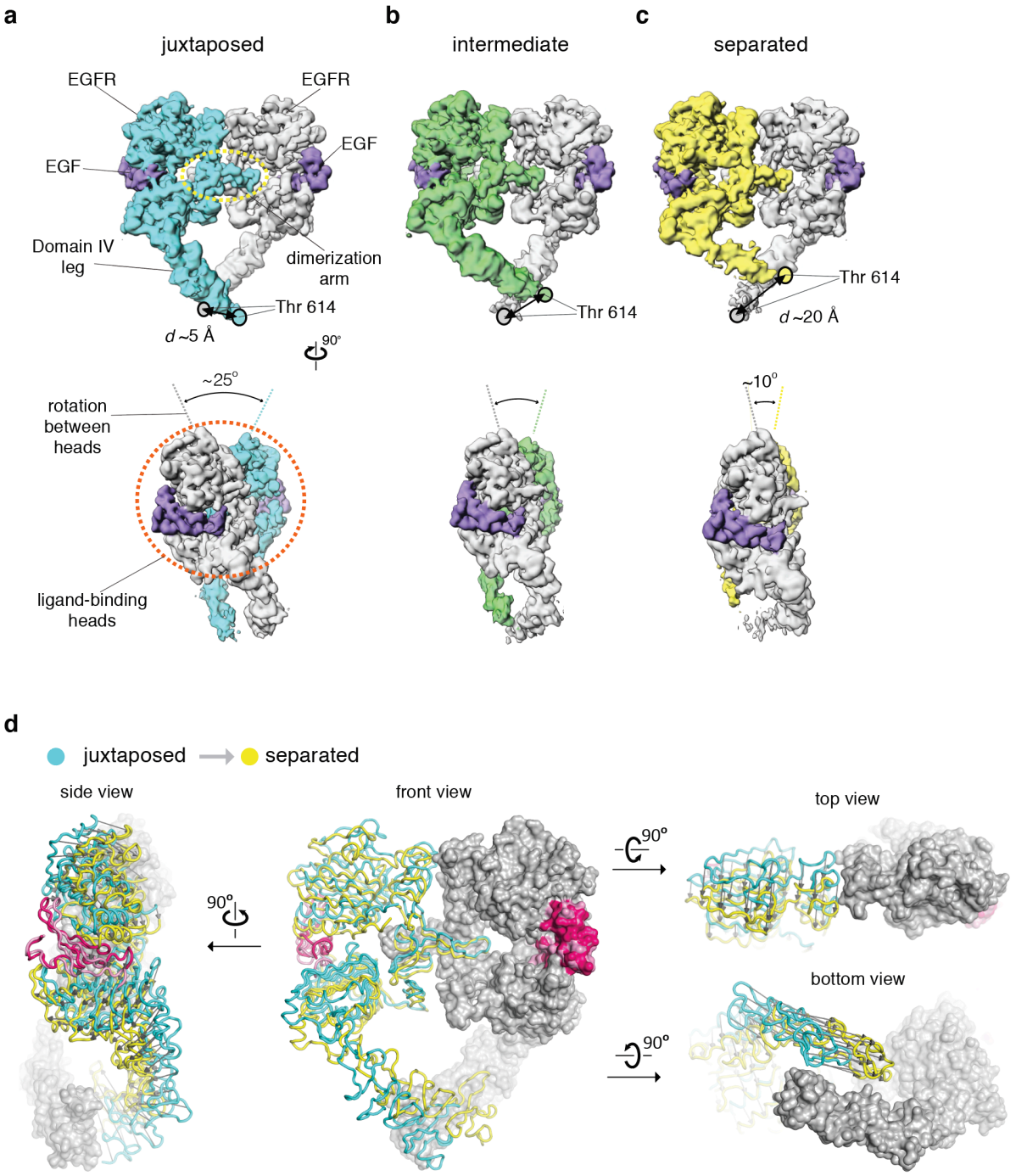

a
juxtaposed
EGFR
EGFR
EGF
EGF
Domain IV leg
dimerization arm
Thr 614
d ~5 Å
90°
~25°
rotation between heads
ligand-binding heads
b
intermediate
Thr 614
c
separated
Thr 614
d ~20 Å
~10°
d
juxtaposed
separated
side view
front view
top view
90°
90°
90°
bottom view

Fig. 2 Cryo-EM analysis reveals a range of conformations for the EGF-bound extracellular module

a-c, Two orthogonal views of the cryo-EM maps for three representative states of the conformational ensemble of the EGF-bound extracellular module: the juxtaposed conformation (**a**), an intermediate conformation (**b**), and the separated conformation (**c**). These conformations differ from one another in the rotations between the ligand-binding heads, as well as in the distances between the two Thr 614 residues located at the membrane-proximal ends of the Domain IV legs.

d, Comparison of the cryo-EM structures of the juxtaposed and the separated conformations of the extracellular module of the EGFR:EGF complex, from four different views. The structures representing the juxtaposed conformation (cyan) and the separated conformation (yellow) are superimposed on the right-hand EGFR subunits which have similar Domain IV leg conformations (light grey). A vector (dark grey) is used to indicate the displacement of Cα atoms between the two structures, pointing from the juxtaposed structure to the separated structure. See also Extended Data Fig.7.

Fig. 3

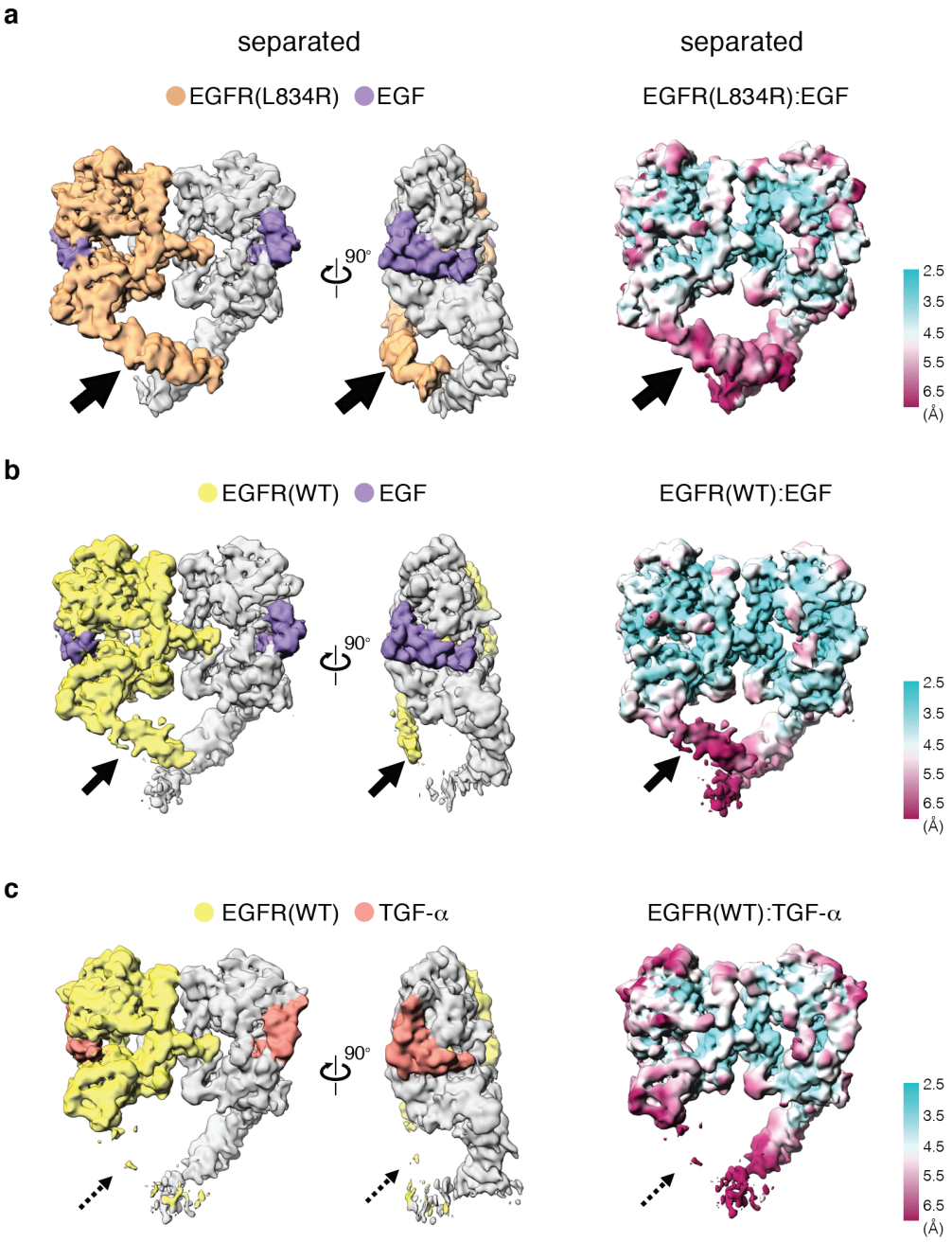
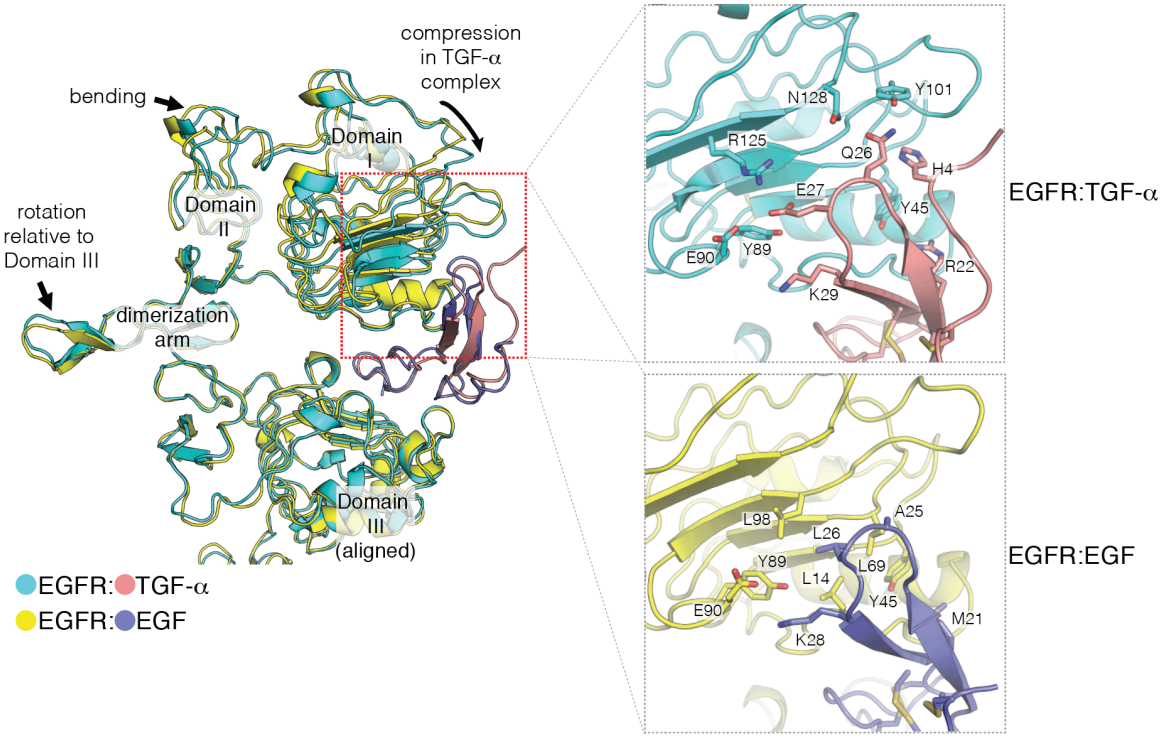

a
separated
separated
EGFR(L834R)
EGF
EGFR(L834R):EGF
90°
2.5
3.5
4.5
5.5
6.5
(Å)
b
EGFR(WT)
EGF
EGFR(WT):EGF
90°
2.5
3.5
4.5
5.5
6.5
(Å)
c
EGFR(WT)
TGF-α
EGFR(WT):TGF-α
90°
2.5
3.5
4.5
5.5
6.5
(Å)

Fig. 3 An oncogenic mutation in the intracellular kinase domain (L834R) stabilizes the separated conformation of the extracellular module

In each panel, the cryo-EM map of the ligand-bound extracellular module is shown on the left, in two orthogonal views. The diagram on the right corresponds to the first view, with the surface colored according to the local resolution of the cryo-EM map. The color scale of the local resolution estimation is shown on the lower right. In each panel, the map shown is that of the separated conformation. **a**, EGFR(L834R) bound to EGF. **b**, EGFR(WT) bound to EGF. **c**, EGFR(WT) bound to TGF-α. The black arrows highlight the different stabilities of the flexible Domain IV leg.

Fig. 4

a


compression
in TGF-α
complex
bending
Domain
I
Domain
II
rotation
relative to
Domain III
dimerization
arm
Domain
III
(aligned)
EGFR:TGF-α
EGFR:EGF
N128
Y101
R125
Q26
H4
E27
Y45
E90
Y89
R22
K29
EGFR:TGF-α
L98
A25
L26
Y89
L69
L14
E90
Y45
M21
K28
EGFR:EGF

b

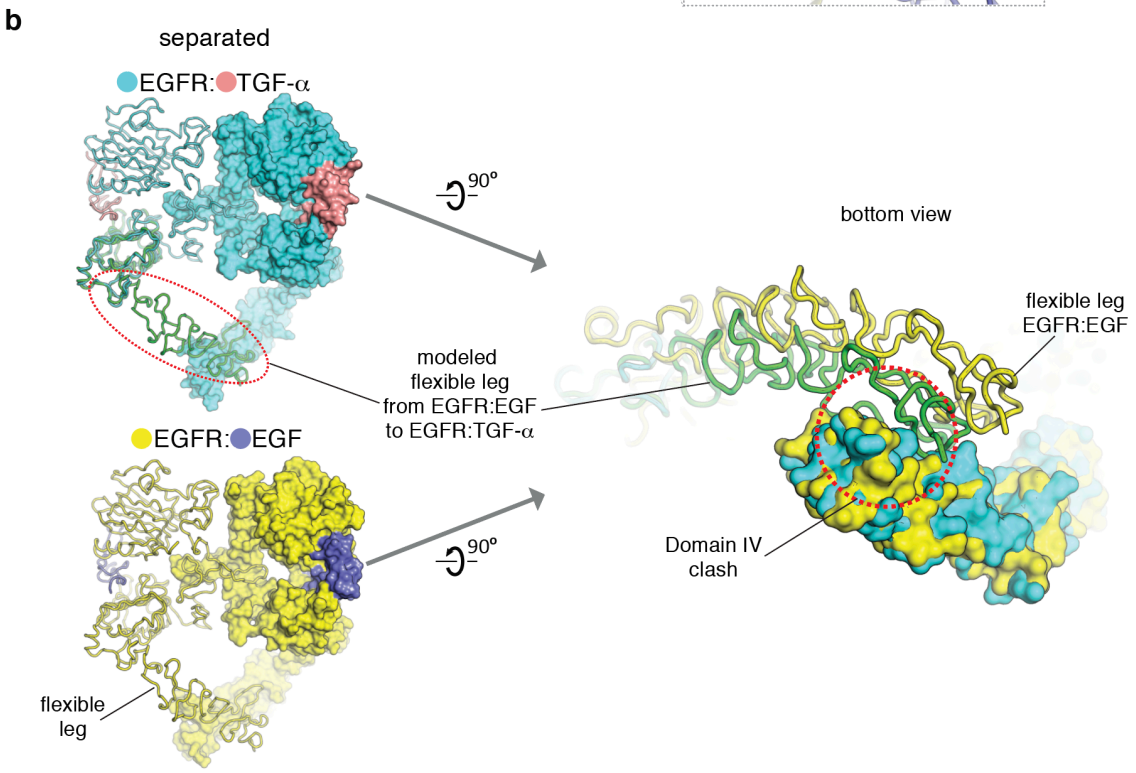

separated
EGFR:TGF-α
90°
bottom view
flexible leg
EGFR:EGF
modeled
flexible leg
from EGFR:EGF
to EGFR:TGF-α
EGFR:EGF
90°
Domain IV
clash
flexible
leg

Fig. 4 EGF and TGF-α differ in their ability to stabilize the separated conformation

a, The differences in structure between the separated conformations of the EGF and TGF-α complexes of EGFR are illustrated. Structures corresponding to the separated conformations of the EGFR:EGF complex (yellow:slate) and the EGFR:TGF-α complex (cyan:salmon) are aligned on Domain III of each structure. Enlarged views of the interface between Domain I of EGFR and the ligands are shown on the right. The interface between Domain I of EGFR (cyan) and TGF-α (salmon) features hydrophilic residues that are shown (upper right). The equivalent interface between Domain I of EGFR (yellow) and EGF (slate) is characterized by hydrophobic residues (lower right).

b, Destabilization of the Domain IV legs in the separated conformation of the EGFR:TGF-α complex. The EGFR:TGF-α complex (upper left) and the EGFR:EGF complex (lower left) are colored as in **a**. The flexible leg in the EGFR:TGF-α complex (green, red dotted line) is modeled based on the conformation of Domains III and IV in the left-hand subunit of the EGFR:EGF complex (lower left). When viewed from the bottom (right), the two Domain IV legs are accommodated in the EGFR:EGF complex (yellow), but not in the EGFR:TGF-α complex. There are stereochemical clashes between the legs (green and cyan), within the dotted red line.

Fig. 5

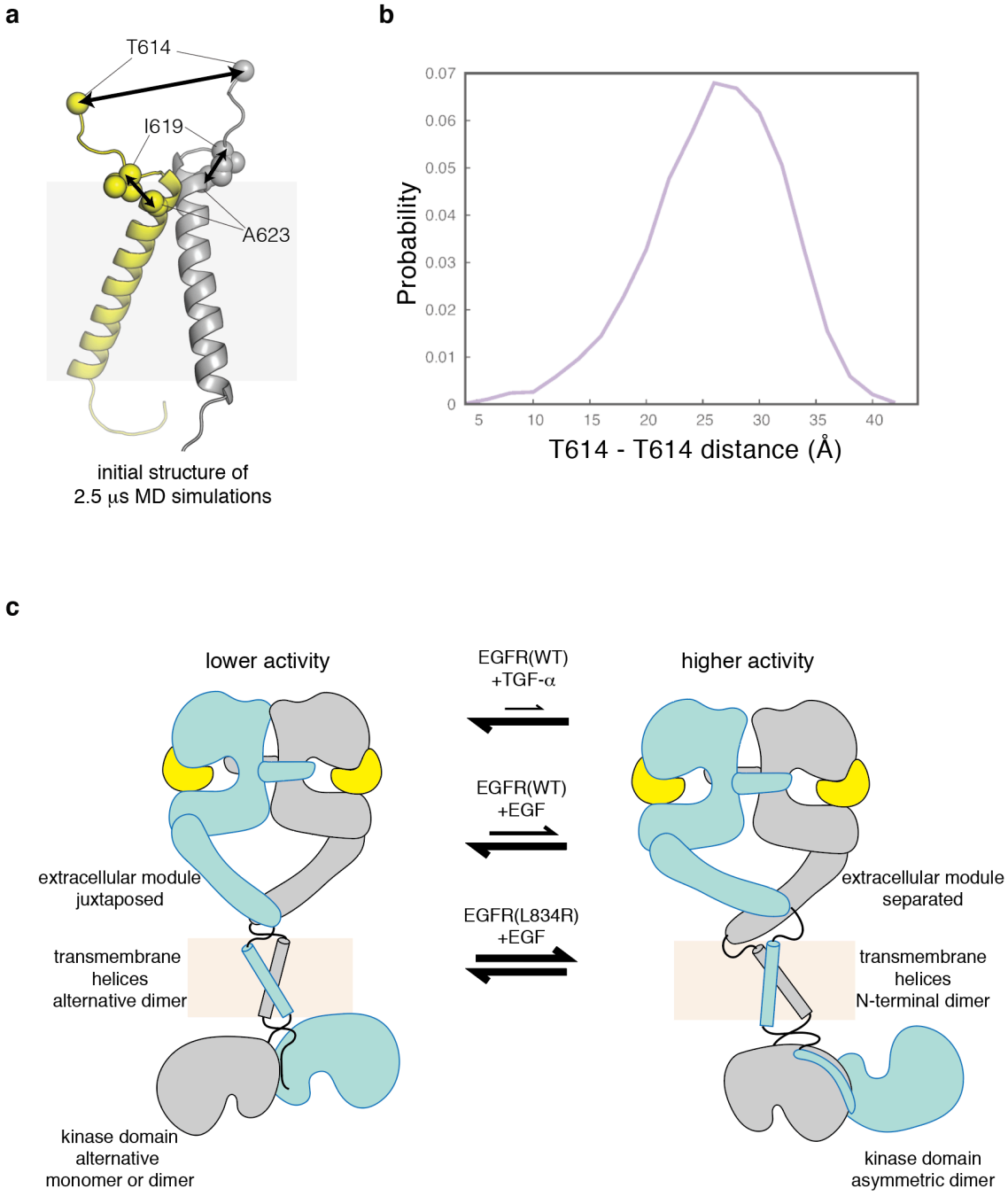

a
T614
I619
A623
initial structure of
2.5 μs MD simulations
b
Probability
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
5
10
15
20
25
30
35
40
T614 - T614 distance (Å)
c
lower activity
EGFR(WT)
+TGF-α
higher activity
EGFR(WT)
+EGF
EGFR(L834R)
+EGF
extracellular module
juxtaposed
transmembrane
helices
alternative dimer
kinase domain
alternative
monomer or dimer
extracellular module
separated
transmembrane
helices
N-terminal dimer
kinase domain
asymmetric dimer

Fig. 5 A mechanism for biased agonism in EGFR by EGF and TGF-α

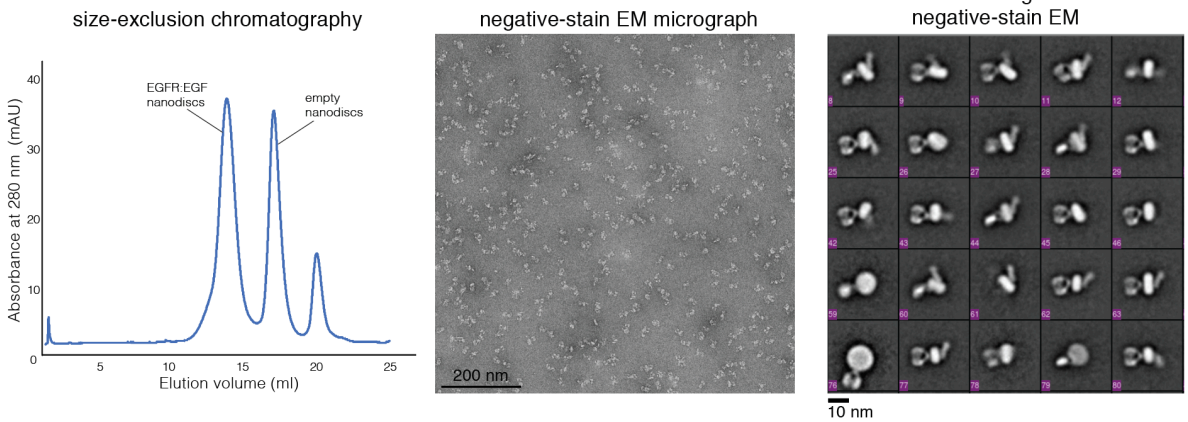
a, Initial structure of the N-terminal transmembrane helices dimer with the linker present, embedded in a neutral POPC membrane, for the 2.5 μs molecular dynamics (MD) simulations (chain A in yellow, chain B in grey). Residues Thr 614 located at the end of the extracellular module are shown as spheres. Also shown as spheres are residues Ile 619 and Ala 623, located at the junction between the linker and the transmembrane helix, and the interaction between these two residues is maintained over 2.5 μs MD simulations (see Extended Data Fig. 9c).

b, Probability distribution of the distances between the Cα atoms of the two Thr 614 residues, sampled every 100 picoseconds over 2.5 μs of MD simulations.

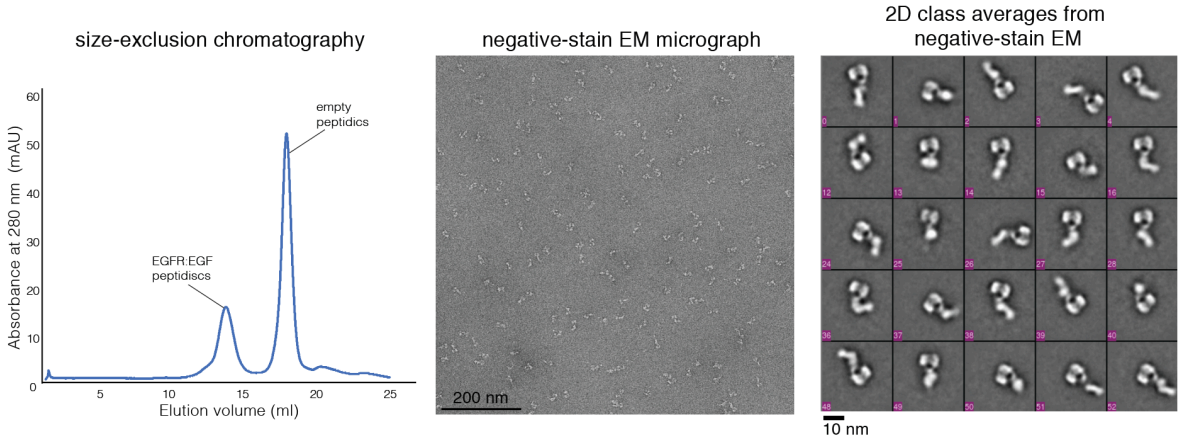
c, Schematic diagrams illustrate two alternative dimeric conformations of the ligands-bound full-length EGFR, representing a lower activity state (left) and a higher activity state (right). It shows how different EGFR ligands and oncogenic mutation could generate different EGFR activity by altering the balance between these dimeric conformations of the full-length EGFR.

Extended Data Fig. 1

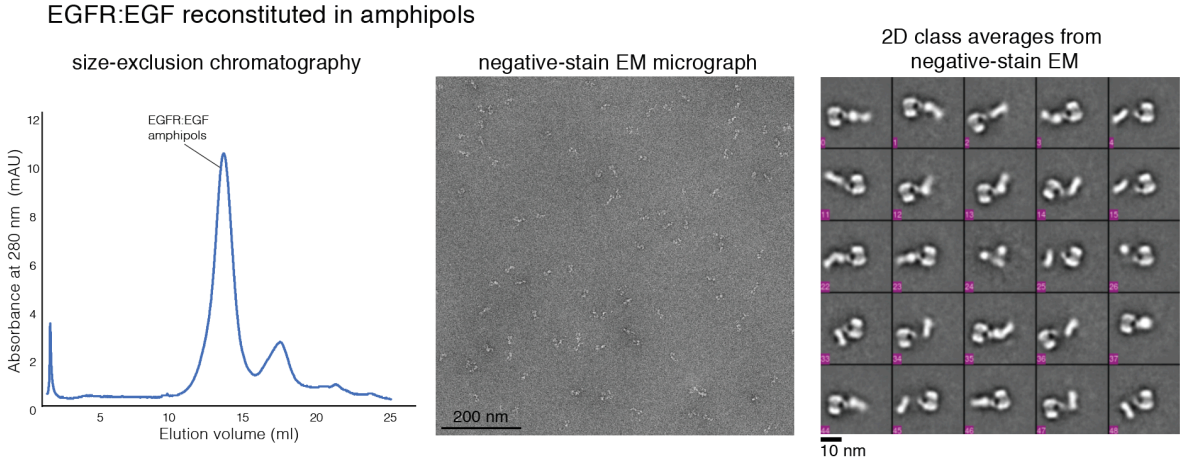
a EGFR:EGF reconstituted in nanodiscs



b EGFR:EGF reconstituted in peptidiscs



c EGFR:EGF reconstituted in amphipols



Extended Data Fig. 1 Reconstitution of full-length human EGFR:EGF complex.

a, Sample preparation of full-length EGFR:EGF reconstituted into nanodiscs.

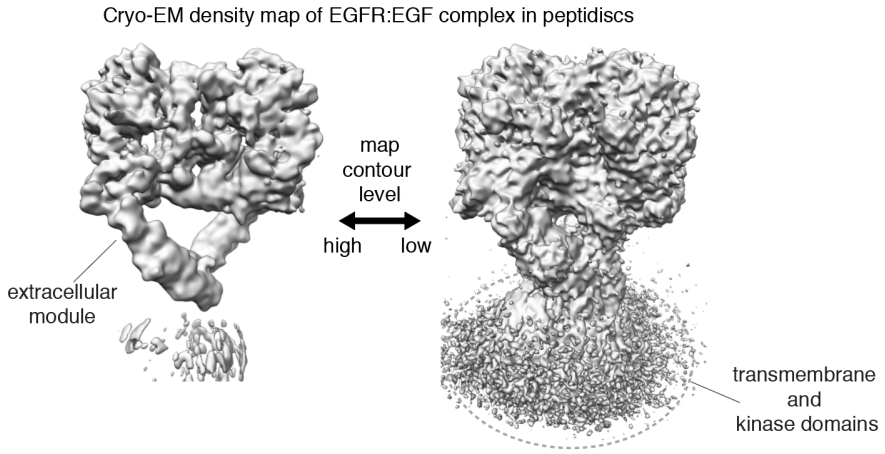
b, Sample preparation of full-length EGFR:EGF reconstituted into peptidiscs.

c, Sample preparation of full-length EGFR:EGF reconstituted into amphipols.

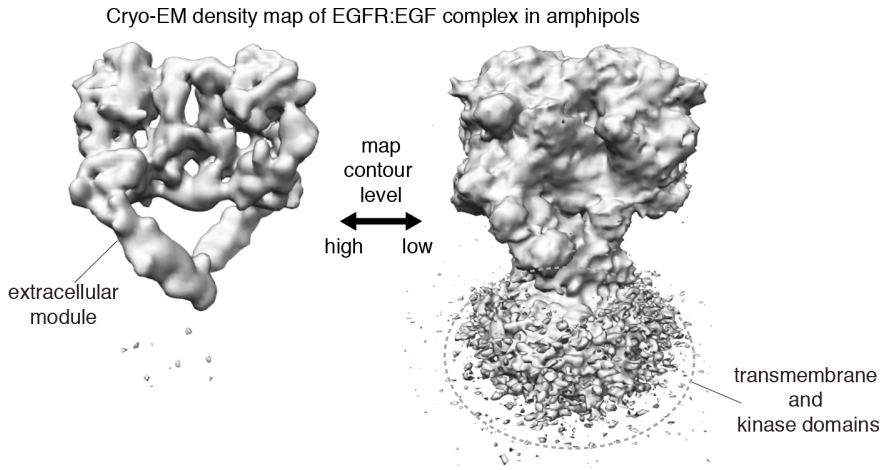
For each sample preparation, a representative profile of size-exclusion chromatography is shown on the left, a representative negative-stain micrograph is shown in the middle, and the representative 2D class averages are shown on the right.

Extended Data Fig. 2

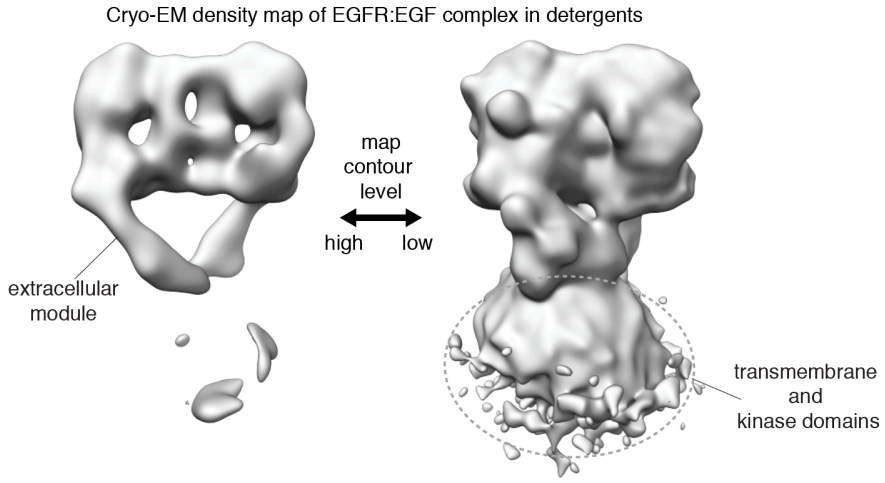
a



b



c



Extended Data Fig. 2 Cryo-EM density maps of reconstituted full-length EGFR:EGF complexes.

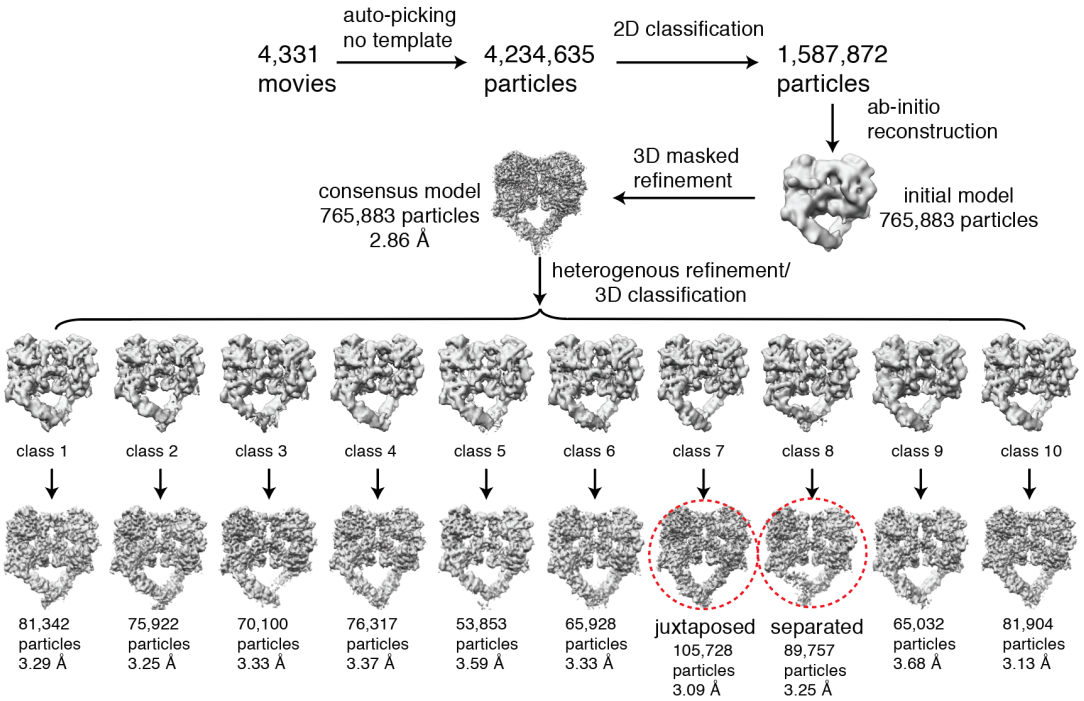
In each panel, the cryo-EM density map for a reconstituted full-length EGFR:EGF complex is shown at two contour levels. Note the lack of EM density for the transmembrane and kinase domains of EGFR at higher contour level (left) in all these complexes.

a, Full-length EGFR:EGF complex reconstituted in peptidiscs.

b, Full-length EGFR:EGF complex reconstituted in amphipols.

c, Full-length EGFR:EGF complex purified in detergents.

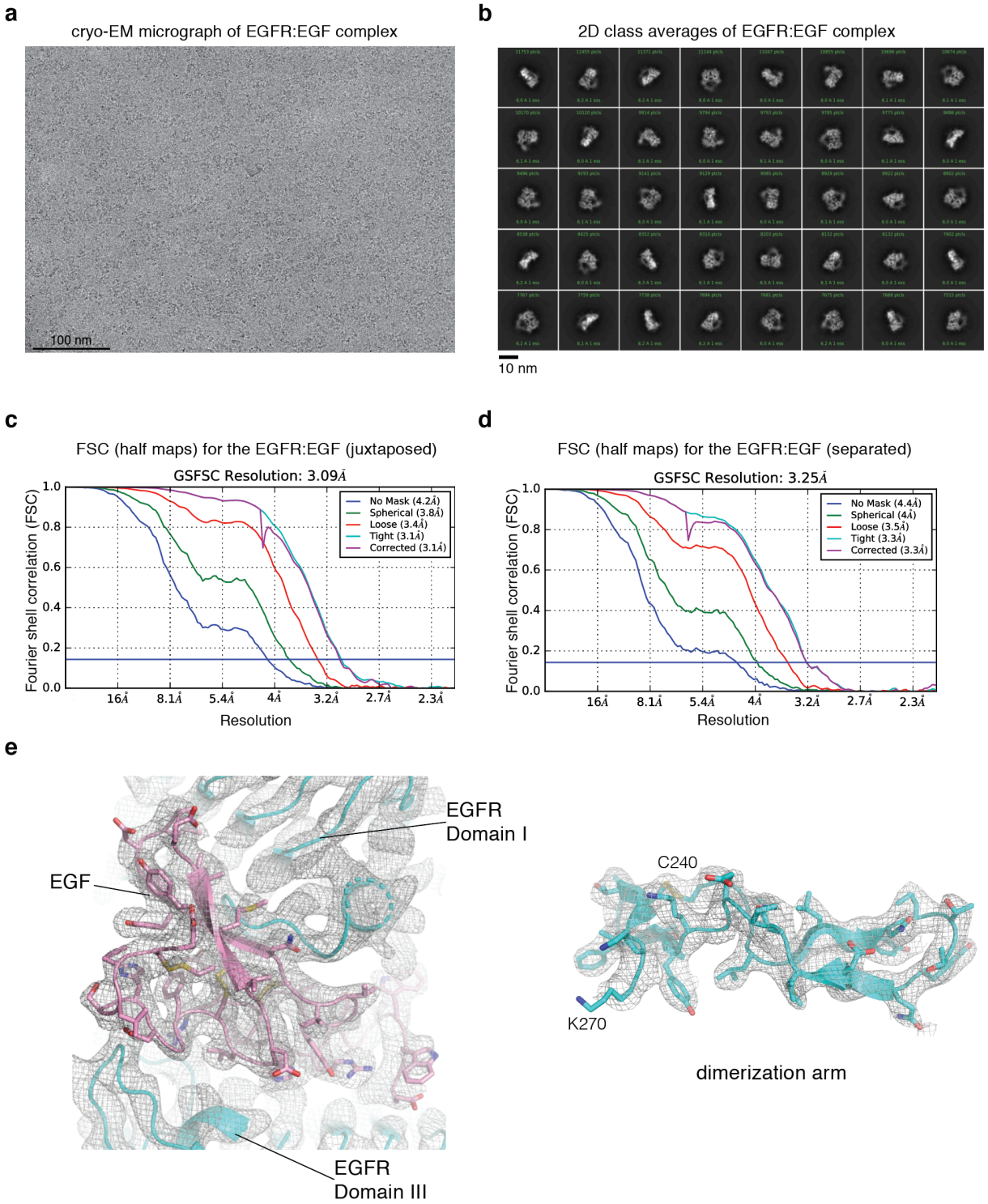
Extended Data Fig. 3



Extended Data Fig. 3 Cryo-EM data processing workflow.

Summary of the cryo-EM data processing workflow for the full-length EGFR:EGF complex reconstituted in peptidiscs. The two extreme conformations of the ensemble, the juxtaposed and the separated, are highlighted by red dotted circle. See Methods for more details.

Extended Data Fig. 4



Extended Data Fig. 4 Cryo-EM image processing for focused reconstruction on the extracellular module of full-length EGFR:EGF complex reconstituted in peptidiscs.

a, Representative cryo-EM micrograph after motion-correction.

b, Representative 2D class averages generated from reference-free 2D classification.

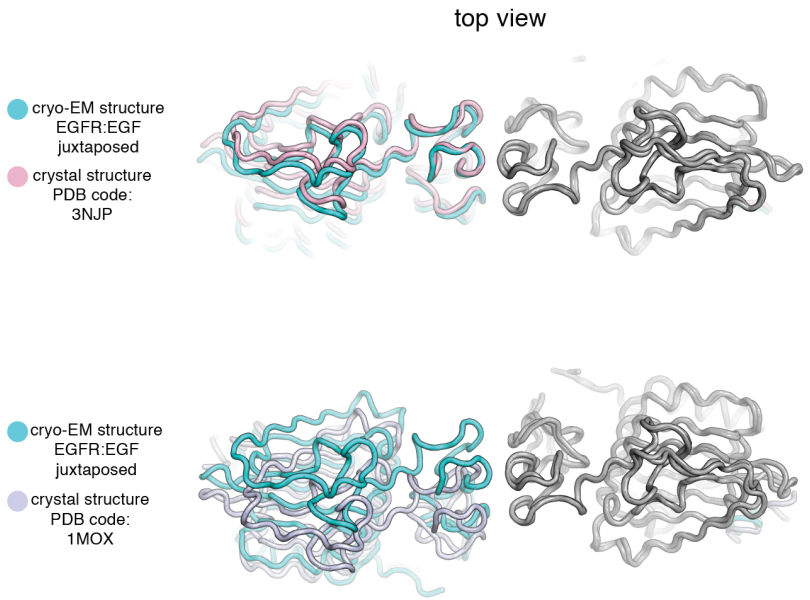
c, Fourier shell correlation (FSC) curves output from cryoSPARC v2 Non-uniform refinement job, for the cryo-EM reconstruction of the juxtaposed conformation of the EGFR:EGF complex reconstituted in peptidiscs. The resolution is estimated at FSC = 0.143.

d, Fourier shell correlation (FSC) curves output from cryoSPARC v2 Non-uniform refinement job, for the cryo-EM reconstruction of the separated conformation of the EGFR:EGF complex reconstituted in peptidiscs. The resolution is estimated at FSC = 0.143.

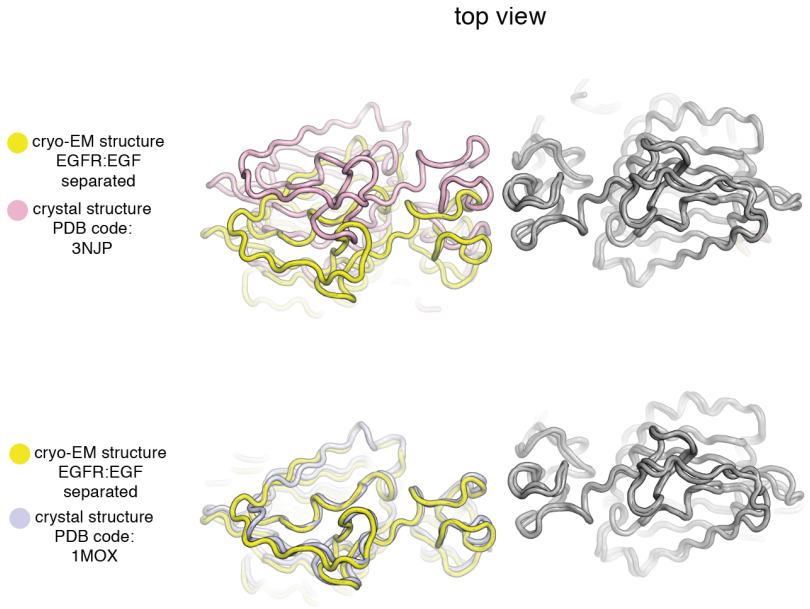
e, Representative cryo-EM densites for the juxtaposed conformation of the EGFR:EGF complex. The structural model and the cryo-EM density corresponding to one of the EGF binding site (left) and the dimerization arm (right, Cys 240-Lys 270) are shown.

Extended Data Fig. 5

a



b



Extended Data Fig. 5 Comparison of Cryo-EM structures of EGFR:EGF complex in two extreme conformations with previous crystal structures of the ligand-binding head of EGFR bound to EGF and TGF-α.

In each panel, a cryo-EM structure of the extracellular module of the full-length EGFR:EGF complex (juxtaposed or separated) is compared to a crystal structure of the ligand-binding head of EGFR bound to ligands (EGF or TGF-α), from top view. Structures are superimposed on the right-hand EGFR subunits (grey).

a, Top, the cryo-EM structure of juxtaposed conformation of EGFR:EGF (cyan) is compared to a crystal structure of the ligand-binding head of EGFR bound to EGF (pink, PDB code: 3NJP).

Bottom, the cryo-EM structure of juxtaposed conformation of EGFR:EGF (cyan) is compared to a crystal structure of the ligand-binding head of EGFR bound to TGF-α (light purple, PDB code: 1MOX).

Note that the cryo-EM structure of the juxtaposed conformation of the EGFR:EGF complex resembles closely the crystal structure of the ligand-binding head of EGFR bound to EGF (PDB code: 3NJP), but deviates from the crystal structure of the ligand-binding head of EGFR bound to TGF-α (PDB code: 1MOX).

b, Top, the cryo-EM structure of separated conformation of EGFR:EGF (yellow) is compared to a crystal structure of the ligand-binding head of EGFR bound to EGF (pink, PDB code: 3NJP).

Bottom, the cryo-EM structure of separated conformation of EGFR:EGF (yellow) is compared to a crystal structure of the ligand-binding head of EGFR bound to TGF-α (light purple, PDB code: 1MOX).

Note that the cryo-EM structure of the separated conformation of the EGFR:EGF complex resembles closely the crystal structure of the ligand-binding head of EGFR bound to TGF-α (PDB

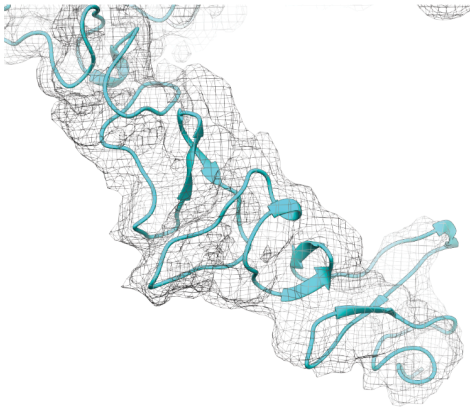
code: 1MOX), but deviates from the crystal structure of the ligand-binding head of EGFR bound to EGF (PDB code: 3NJP).

Extended Data Fig. 6

a

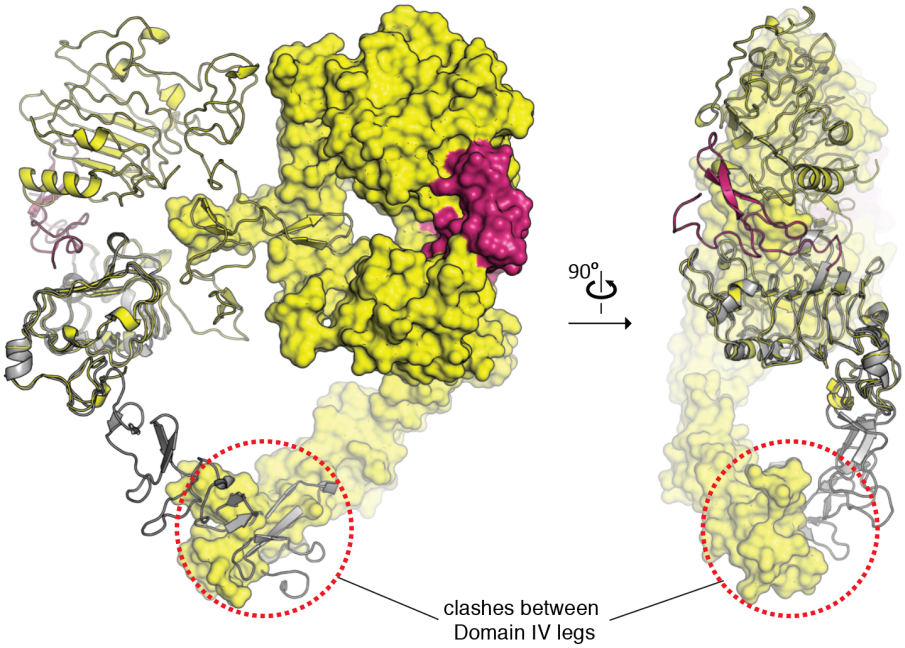
superposition of the Domain IV legs
from crystal structures of EGFR

b



Docking of the Domain IV leg
into cryo-EM density

c



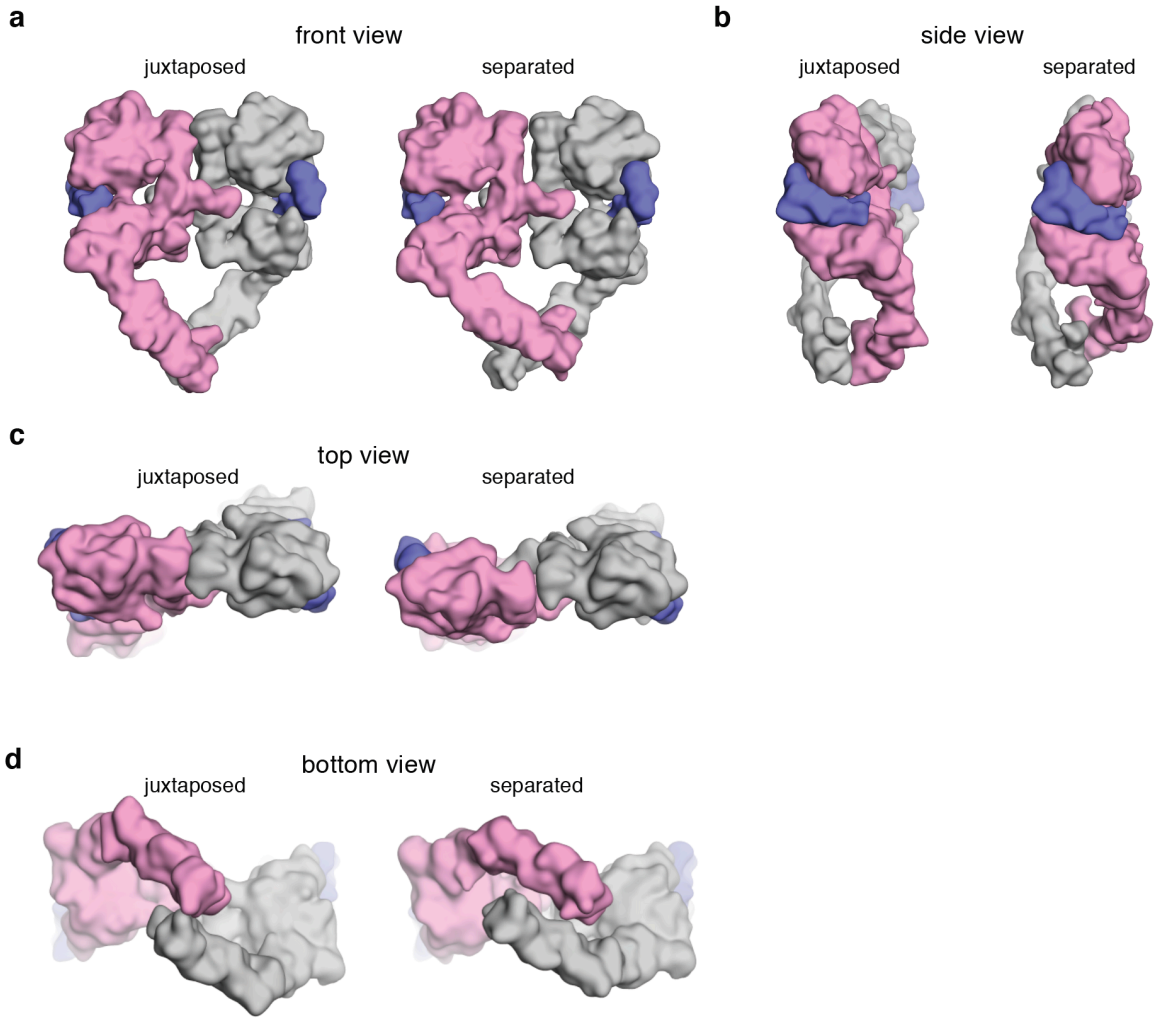
Extended Data Fig. 6 Conformation of the Domain IV leg.

a, Superposition of the Domain IV leg from the dimeric crystal structure of the isolated EGFR extracellular domains bound to EGF (cyan, PDB code: 3NJP), with the Domain IV legs from seven different crystal structures of monomeric extracellular domains of EGFR (grey, PDB codes: 4UV7, 5XWD, 1NQL, 3QWQ, 1YY9, 6ARU, 4UIP). This comparison shows that the Domain IV leg is a fairly rigid structural component.

b, The Domain IV leg from a crystal structure of the isolated EGFR extracellular domains bound to EGF (PDB code: 3NJP) is fitted into the cryo-EM density corresponding to the flexible leg in the cryo-EM map for the juxtaposed conformation of the EGFR:EGF complex.

c, The extracellular module of the EGFR:EGF complex in separated conformation is shown in two orthogonal views. The Domain IV leg shown on the left-hand subunit (grey) is modeled based on the conformation of Domains III and IV in the right-hand subunit. Maintaining the same conformation in both legs would lead to steric clashes between the two legs (circled by red dotted lines), when the extracellular module of EGFR adopts the separated conformation.

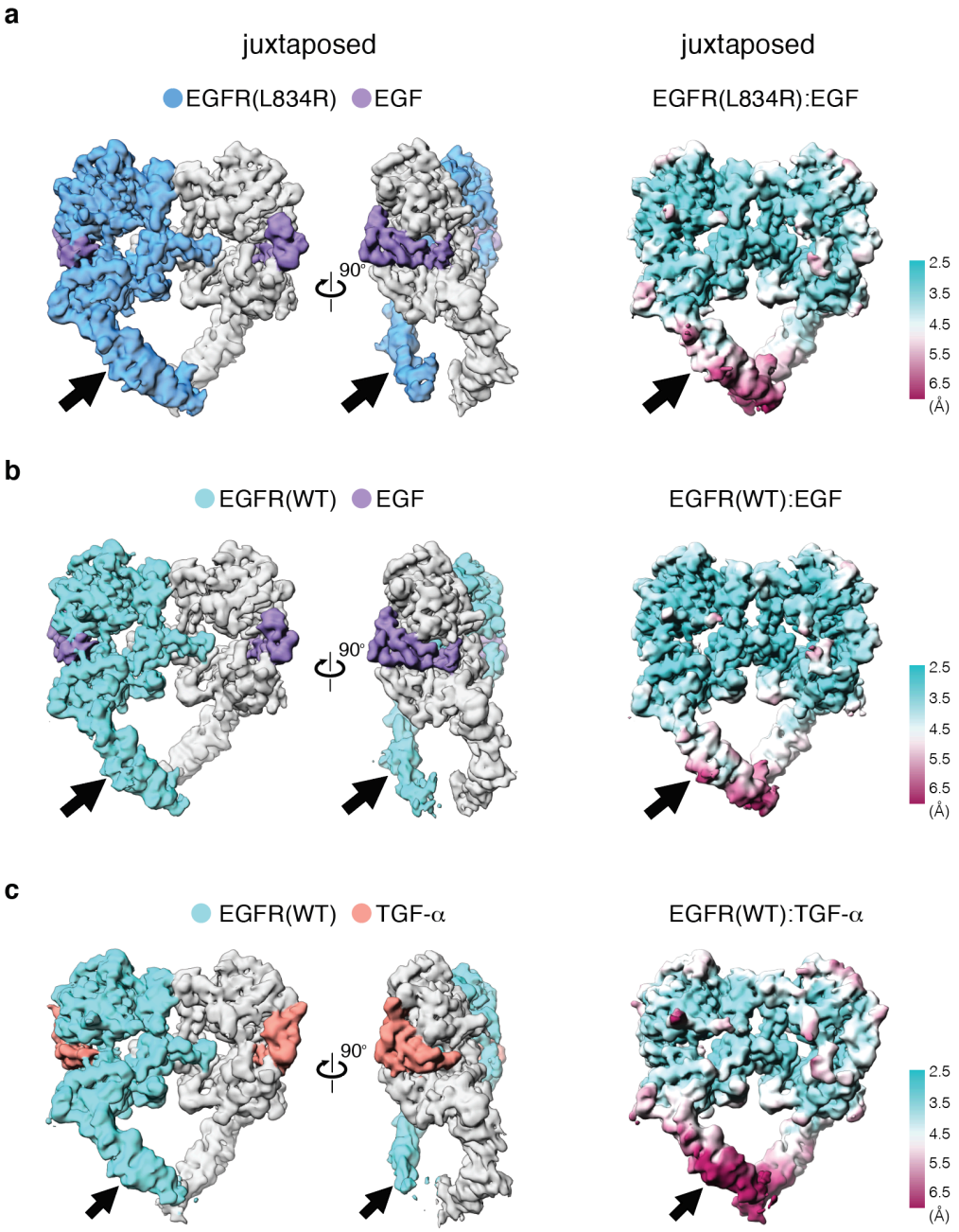
Extended Data Fig. 7



Extended Data Fig. 7 Comparison between the juxtaposed conformation and the separated conformation.

Side by side comparison between the juxtaposed conformation and the separated conformation of the extracellular module of the EGFR:EGF complex, see also Fig. 2d. Molecular surface of the cryo-EM models of the juxtaposed and separated conformations are shown from the front (**a**), the side (**b**), the top (**c**), and the bottom (**d**). EGF molecules are shown in slate blue. The same EGFR subunit used for superposition in Fig. 2d is colored grey and the other EGFR subunit is light pink.

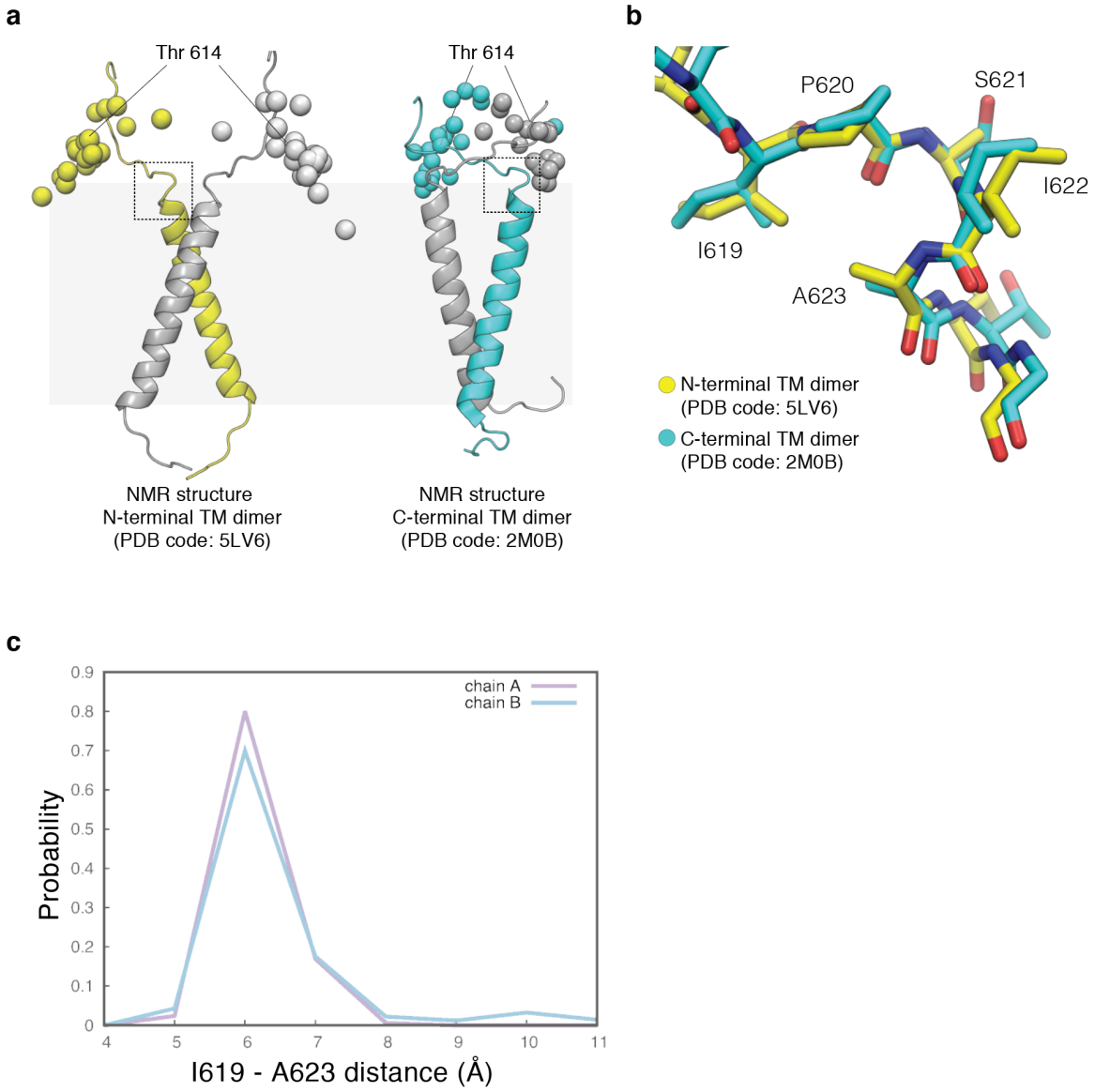
Extended Data Fig. 8



Extended Data Fig. 8 Comparison between Cryo-EM density maps of the juxtaposed conformation in different EGFR complexes.

In each panel, the cryo-EM map of the ligand-bound extracellular module is shown on the left, in two orthogonal views. The diagram on the right corresponds to the first view, with the surface colored according to the local resolution of the cryo-EM map. The color scale of the local resolution estimation is shown on the lower right. In each panel, the map shown is that of the juxtaposed conformation. **a**, EGFR(L834R) bound to EGF. **b**, EGFR(WT) bound to EGF. **c**, EGFR(WT) bound to TGF-α. The black arrows highlight the flexible Domain IV legs.

Extended Data Fig. 9



Extended Data Fig. 9 Defined conformation of the linker at the junction with the transmembrane helix.

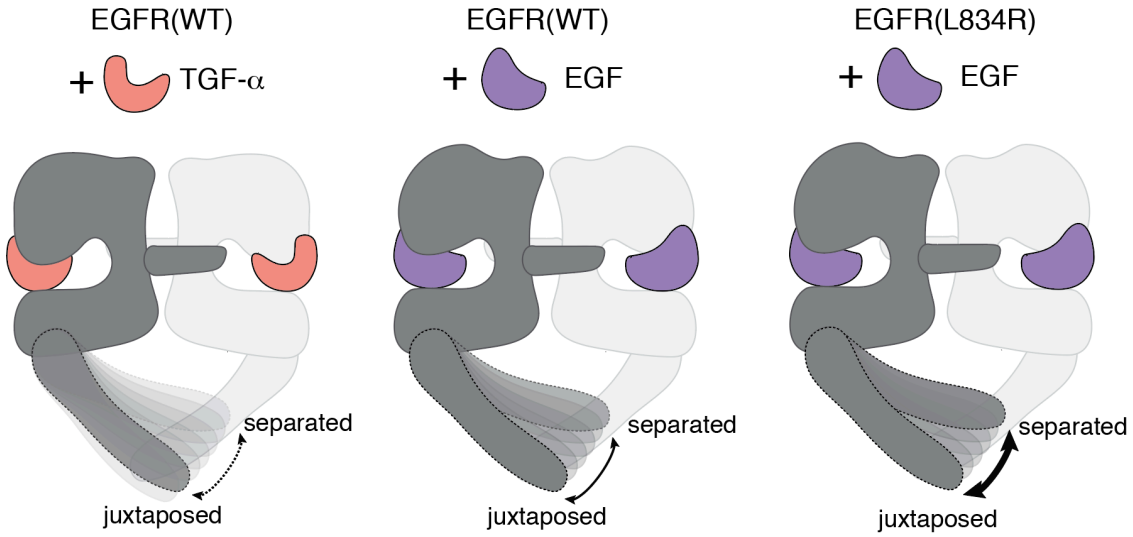
a, Models of the transmembrane helices in two conformations based on NMR structures. One of the structures in the NMR-based ensemble for the N-terminal dimer conformation (PDB code: 5LV6) is shown in yellow (chain A) and grey (chain B), and the structure for the C-terminal dimer conformation (PDB code: 2M0B) is shown in cyan (chain A) and grey (chain B). Shown in spheres are superimposed Cα atoms of the residue Thr 614, located at the end of the extracellular module, from the 20 structures in the NMR-based ensemble for the N-terminal dimer conformation (left) and for the C-terminal dimer conformation (right), respectively. The black dotted square indicates the linker region that is imposing a constraint on the orientation of the linkers. In the N-terminal dimer conformation of the transmembrane helices (left), the linkers from each subunit are oriented away from one another. In the C-terminal dimer conformation of the transmembrane helices (right), the linkers are oriented towards each other.

b, Enlarged view of the linker located at the junction with the transmembrane helices, zoomed in from the areas highlighted by dotted black square in **a**, showing well-defined conformations for this region in both NMR structures.

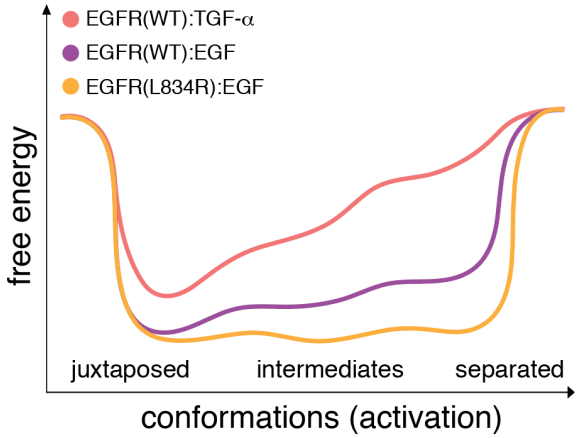
c, Probability distribution of the distances between the Cα atoms of Ile 619 and Ala 623, sampled every 100 picoseconds over 2.5 μs of MD simulations. See also Fig. 5a.

Extended Data Fig. 10

a



b



Extended Data Fig. 10 Differential modulation of the conformation of the ligand-bound extracellular module.

a, Schematic representations of the ligand-bound extracellular module of EGFR illustrate the differential modulation of the conformation of the Domain IV legs by TGF-α (left), EGF(middle) and oncogenic L834R mutation (right). Cryo-EM densities for the Domain IV legs in the juxtaposed conformation is similarly defined in all three complexes (EGFR(WT):TGF-α, EGFR(WT):EGF, and EGFR(L834R):EGF). In contrast, density for the Domain IV legs in the separated conformation is poorly defined in the EGFR(WT):TGF-α complex and best defined in the EGFR(L834R):EGF complex.

b, A schematic diagram showing a simplified energy landscape for the ligand-bound extracellular module of EGFR. Cryo-EM analysis revealed the ensemble of conformations and the effects of the indicated ligands and mutation on the energy landscape.

Extended Data Table 1
Statistics of cryo-EM data collection and structure determination

	EGFR(WT) :EGF juxtaposed	EGFR(WT) :EGF separated	EGFR(WT) :TGF-α juxtaposed	EGFR(WT) :TGF-α separated	EGFR(L834R) :EGF juxtaposed	EGFR(L834R) :EGF separated
Data collection						
Electron microscope	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios
Camera	Gatan K3	Gatan K3	Gatan K3	Gatan K3	Gatan K3	Gatan K3
Voltage (kV)	300	300	300	300	300	300
Magnification	81,000	81,000	81,000	81,000	81,000	81,000
Pixel sise (Å)	0.54	0.54	0.54	0.54	0.54	0.54
Nominal dose (e^-/Å^2)	50	50	50	50	50	50
Tilt angle (°)	40	40	40	40	40	40
Defocus range (μm)	-0.75 to -1.5	-0.75 to -1.5	-0.75 to -1.5	-0.75 to -1.5	-0.75 to -1.5	-0.75 to -1.5
Image processing						
Micrographs	4,331	4,331	7,406	7,406	5,013	5,013
Final particle number	105,728	89,757	126,471	102,206	123,670	100,615
Map resolution (Å)	3.1	3.3	3.4	3.6	3.3	3.4
Map sharpening B factor	-97.6	-94.7	-108.2	-108.2	-103.5	-92.8
Model refinement						
Rms deviation bonds (Å)	0.007	0.006	0.004	0.007	0.004	0.004
Rms deviation angles (°)	0.691	0.732	0.639	0.820	0.663	0.705
Map CC (mask)	0.82	0.76	0.79	0.74	0.81	0.72
Validation						
Molprobility score	1.96	2.20	2.05	2.45	1.91	2.19
Clash score	7.50	11.14	8.84	17.78	6.41	10.87
Rotamer outliers (%)	0.09	0.19	0.00	0.38	0.00	0.00
Cβ outliers (%)	0.00	0.00	0.00	0.00	0.00	0.00
Ramachandran plot						
Favored (%)	90.08	86.00	88.97	82.75	89.55	86.14
Allowed (%)	9.92	14.00	11.03	17.25	10.45	13.86
Outliers (%)	0.00	0.00	0.00	0.00	0.00	0.00