

Ligand binding and dynamics of the monomeric epidermal growth factor receptor ectodomain

Hannes H. Loeffler* and Martyn D. Winn

Scientific Computing Department, STFC Daresbury, Warrington WA4 4AD, United Kingdom

ABSTRACT

The ectodomain of the human epidermal growth factor receptor (hEGFR) controls input to several cell signalling networks via binding with extracellular growth factors. To gain insight into the dynamics and ligand binding of the ectodomain, the hEGFR monomer was subjected to molecular dynamics simulation. The monomer was found to be substantially more flexible than the ectodomain dimer studied previously. Simulations where the endogenous ligand EGF binds to either Subdomain I or Subdomain III, or where hEGFR is unbound, show significant differences in dynamics. The molecular mechanics Poisson–Boltzmann surface area method has been used to derive relative free energies of ligand binding, and we find that the ligand is capable of binding either subdomain with a slight preference for III. Alanine-scanning calculations for the effect of selected ligand mutants on binding reproduce the trends of affinity measurements. Taken together, these results emphasize the possible role of the ectodomain monomer in the initial step of ligand binding, and add details to the static picture obtained from crystal structures.

Proteins 2013; 81:1931–1943.

© 2013 The Authors. Proteins published by Wiley Periodicals, Inc

Key words: molecular dynamics; free energy; molecular mechanics Poisson–Boltzmann surface area; extracellular domain; receptor–ligand; Her1; ErbB1.

INTRODUCTION

The human epidermal growth factor receptor (hEGFR), also referred to as ErbB1 or Her1, is a receptor tyrosine kinase (RTK) and part of a family of four closely related receptors (ErbB1–ErbB4). These RTKs provide input to a complicated signalling network controlling such vital cell functions as proliferation, migration, differentiation, and apoptosis. Their activity is controlled by a set of 13 extracellular ligands which bind to the ectodomain.¹ Some ligands bind to more than one family member but none are known to bind to ErbB2. These RTKs are expressed in many kinds of tissues including skin, internal organs, breast, and prostate, and dysfunction in the network may lead to serious illnesses, in particular various types of cancers. Consequently, the ErbB network is one of the most studied systems of signal transduction.

The hEGFR monomer comprises an ectodomain with four subdomains (I–IV), a single transmembrane helix, the juxtamembrane region, a tyrosine kinase domain, and the C-terminal regulatory region. The active receptor is usually considered to be the dimer, with both homo and hetero dimers occurring within the ErbB family.

Structural information comes from crystallography of individual domains,^{2–6} Nuclear magnetic resonance (NMR) as well as infra-red and circular dichroism (CD) spectroscopy of the transmembrane helices and the juxtamembrane domain,^{7–9} small angle X-ray scattering (SAXS) of soluble ectodomains,¹⁰ and negative-stain electron microscopy.^{11–13} From crystallography, the ectodomain of the hEGFR monomer is known to exist either in an autoinhibited (so-called tethered) conformation^{2–4} or in an open conformation.^{5,6} The latter is capable of dimerising and thus helping to form the active receptor. The existence of these two conformations in solution has been confirmed by SAXS.¹⁰ Negative-stain

Additional Supporting Information may be found in the online version of this article.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Grant sponsor: UK Biotechnology and Biological Sciences Research Council; Grant number: BB/G006911/1.

*Correspondence to: Hannes H. Loeffler; Scientific Computing Department, STFC Daresbury, Warrington WA4 4AD, United Kingdom.

E-mail: Hannes.Loeffler@stfc.ac.uk

Received 29 January 2013; Revised 26 April 2013; Accepted 19 May 2013

Published online 12 June 2013 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/prot.24339

electron microscopy has also suggested distinct densities consistent with different ectodomain conformations.¹³

The exact mechanism of ligand binding is unknown but in principal two models are possible.¹⁴ One model assumes that a small fraction of the unbound monomer exists in the open conformation and ligand binding to that conformation drives the thermodynamic equilibrium to the side of the dimer.² The other model requires the ligand to bind first to the tethered conformation to initiate opening and subsequent dimerization. One crystal structure of the tethered hEGFR monomer does in fact include an epidermal growth factor (EGF) ligand bound to subdomain I, but this is not considered to be biologically relevant.² Scatchard analysis of ligand binding indicates two populations of binding sites, and this has been explained in terms of negative cooperativity of ligand binding to the receptor dimer.¹⁵ A possible structural basis for this negative cooperativity has been proposed based on crystal structures of *Drosophila* EGFR,¹⁶ and from simulations of human EGFR ectodomains on the plasma membrane.¹⁷ The parameterized model of MacDonald and Pike¹⁵ implies significant dimerization of unliganded ectodomain dimers, which would require ligand-independent conformational changes of the monomer.

Molecular dynamics (MD) simulations provide a means to supplement experimental structural data with atomic-level information about protein dynamics. Such information is vital to understanding a flexible multiconformation protein such as hEGFR. In recent years, several MD studies of the ectodomain dimer^{17–22} and of the tyrosine kinase domain^{21,23–28} have emerged. The hEGFR tethered monomer has not yet been the subject of MD simulations, although there is a recent study of the closely related ErbB4 monomer²⁹ whose dynamics and ligand-binding properties are less well-understood. An understanding of the tethered monomer would help clarify the earliest steps of ligand-induced activation of hEGFR.

In addition to exploring the protein dynamics, simulation can also help to understand the energetics of ligand binding. Numerous methods have been devised to estimate free energies of binding ΔG_{bind} of biologically relevant ligands to proteins.³⁰ Some approaches attempt to obtain ΔG_{bind} through direct simulation of the binding event, but such methods are typically computationally expensive as many intermediate states need to be explicitly enumerated and/or multiple runs are required to achieve sufficient numerical convergence. The ectodomain consist of the first 614 residues of hEGFR, whereas the EGF ligand has 53 residues, and so even the monomeric receptor-ligand system is computationally challenging. Some end-point methods on the other hand may conveniently be carried out after the MD trajectory has been computed. Necessarily, these approaches are less reliable in terms of accuracy and absolute binding

energies may not be quantitatively possible, but relative energies and ranking may be within reach.^{31,32} One such method is the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) approach.^{33–35} With MM–PBSA, coordinates from the trajectory are post-processed to evaluate gas-phase force field energies (the MM part) and to estimate the free energy of solvation through solution of the Poisson–Boltzmann equation (the PB part) and an expression relating to the surface area (the SA part). ΔG_{bind} can then be estimated from combining these contributions computed for complex, receptor, and ligand.

A large number of site-directed mutagenesis experiments on ligands have been carried out to gain a deeper understanding of the ligand’s interactions with the receptor and the functional role of individual residues. Relevant to the current study are mutations of the EGF ligand at Leu47,^{36,37} Arg41,^{38,39} Lys28,⁴⁰ and Ile23.^{37,40–42} To relate these experiments to structural models of ligand binding, a computational technique called “alanine scanning” can be used. *In silico* mutations of residues to alanine are analyzed with the MM–PBSA method to obtain relative free energies of binding for such ligand variants.

In this study, we investigate the dynamics and ligand binding of the monomeric hEGFR ectodomain via MD simulation. Experimental insights into the structural transitions of the monomer have come from crystal structures,^{2–4} from SAXS measurements,¹⁰ and from tryptophan fluorescence.¹⁴ Here, we use MD simulations and the MM–PBSA method to investigate a structural model of the monomeric ectodomain, and relate our results to those experiments. A picture emerges of a highly flexible protein, which complicates simple models for ligand binding.

The work here lays the foundation for understanding ligand binding in other members of the ErbB family, and in the yet larger dimeric system. It may help to rationalize the effects of dimer asymmetry on ligand binding, seen in *Drosophila* EGFR¹⁶ and in simulations of hEGFR on the plasma membrane.¹⁷

MATERIALS AND METHODS

Simulation setup

Three independent simulations of the hEGFR ectodomain in the tethered conformation were carried out with: (1) the EGF ligand bound to domain I (bI), (2) the EGF ligand bound to domain III (bIII), and (3) the ligand removed from the receptor (ub). The starting structure for bI was taken from PDBid 1NQL,² which is a crystal structure of the monomeric tethered hEGFR ectodomain with EGF bound to domain I. To construct the configuration where EGF is attached to domain III, we fitted domain III of PDBid 1IVO⁵ to 1NQL and

transferred the ligand coordinates. The unbound simulation was prepared by removing EGF from 1NQL.

Protonation states of titratable residues at pH 7 were determined with PROPKA^{43,44} using the PDB2PQR web interface.⁴⁵ The histidines in the domain I binding site, His23, and the domain III binding site, His346, His359, and His409, were all δ -protonated in each of the three simulations. These structures were subjected to 500 steps of initial minimization with heavy atoms restrained. The proteins were immersed in a cubic water box and counter ions were added to create an approximately 0.1M solution of NaCl. Thus, 177 Cl⁻ and 183 Na⁺ ions were added for the bI, 177 Cl⁻ and 184 Na⁺ ions were added for the bIII, and 180 Cl⁻ and 182 Na⁺ ions were added for the ub simulations, respectively. The total number of atoms was about 395,000 for each simulation system leaving a water buffer of at least 30 Å between the protein and the closest box face.

All heavy atoms of the receptor and the ligand were subjected to positional restraints of 5 kcal/mol/Å. Each system was energy minimized for 1000 steps to relax atom positions and remove bad contacts from the setup procedure. Next, the simulation boxes were heated incrementally from 25 K to the final temperature of 300 K in steps of 1000 over a period of 10,000 steps and kept at the final temperature for an additional 10,000 steps. Over a further 25,000 steps a constant pressure of 1 bar was applied to adjust the box size and the density. The temperature was controlled with a Langevin thermostat and a coupling coefficient of 1 ps⁻¹. The pressure was controlled by a Nosé–Hoover Langevin barostat^{46,47} with a decay time of 1000 fs and an oscillation period of 2000 fs. Finally, restraints were switched off step-wise over a period of 10,000 steps in increments of 1000 steps for initial preparation of the production run.

The total unrestrained simulation times were 150 ns for each of the three systems. All MD simulations were carried out at a constant temperature of 300 K and a constant pressure of 1 bar. The Langevin barostat decay time was set to 100 fs and the oscillation period to 200 fs. The simulations were run with NAMD⁴⁸ in versions 2.6 and 2.7. The force fields used were CHARMM 22 for proteins⁴⁹ with torsional backbone corrections (CMAP⁵⁰) and ions,⁵¹ and TIP3P for water.^{49,52} All bonds to hydrogens were constrained with the SHAKE algorithm which allowed for a time step of 2 fs. This simulation protocol is essentially the same as in our previous studies.^{17,22}

Binding free energies

Binding free energies were estimated by means of the MM–PBSA method.^{33,34,53} MM–PBSA is an endpoint method that allows convenient post analysis of the trajectory or trajectories. The binding free energy ΔG is

calculated from the individual free energies of the three species as

$$\Delta G = \langle G(\text{complex}) \rangle - \langle G(\text{protein}) \rangle - \langle G(\text{peptide}) \rangle \quad (1)$$

where complex is the hEGFR monomer plus the EGF ligand (bound to either Domain I or Domain III), protein is hEGFR and peptide is EGF. Each free energy G is computed as an average over the trajectory in the following way:

$$\langle G \rangle = \langle E_{\text{MM}} \rangle + E_{\text{rot/tr}} + \langle G_{\text{PBSA}} \rangle - T \langle S_{\text{MM}} \rangle \quad (2)$$

where E_{MM} is the molecular mechanical energy for the chosen force field including internal and nonbonding degrees of freedom, G_{PBSA} is the solvation free energy with the polar contributions computed through the Poisson–Boltzmann equation and the nonpolar part estimated from the scaled and shifted molecular surface, T is the simulation temperature, and S_{MM} is the conformational entropy of the solute. $E_{\text{rot/tr}}$ is the energy due to the six rotational and translational degrees of freedom and in the classical limit is $6 \times \frac{1}{2} RT$ ($= 1.79$ kcal/mol at 300 K).³⁵ In the computation of a relative free binding energy $\Delta \Delta G$, this term cancels, and is often neglected.

In principle, three independent simulations must be carried out to calculate each term in Eq. (1). In some situations, it may be possible to obtain a reasonable estimate for ΔG by sampling all three species from a single simulation of the complex. The assumption is that the conformations sampled for protein and peptide in the complex are representative for the unbound state. In the next section, we will discuss how far this is true for the systems studied here.

The MM–PBSA calculations have been carried out with CHARMM^{54,55} 35 using a script developed by us.⁵⁶ We validated our implementation by computing ΔG for the Ras-Raf complex (a standard test system), and comparing to the Amber results (see Supporting Information). The force field energies E_{MM} were calculated without nonbonded cutoffs. The nonpolar contributions to G_{PBSA} were evaluated with CHARMM's surface module by computing them as $G_{\text{nonpolar}} = \gamma A + \beta$ ^{55,57} where γ and β were 0.00542 kcal/Å² and 0.92 kcal/mol, respectively, and A is the molecular surface in Å² computed with the atomic radii by Nina *et al.*⁵⁸ The polar contributions were calculated with APBS⁵⁹ using the CHARMM 22 partial charges and a grid resolution of 0.3 Å (see Supporting Information) which required up to 13 GB of memory. Snapshots were taken every 100 ps (1400 data points) from the bIII simulation and every 200 ps (700 data points) from the bI and ub simulations. The trajectory averages were then calculated omitting the first 10 ns of simulation. For the three-trajectory approach, a 75 ns reference simulation of the EGF ligand was used, taking snapshots every 100ps



Figure 1

The bIII simulation structure (red, ligand in orange bound to domain III) taken from a snapshot at 106.27 ns, compared to the crystal structures 1NQL (grey, ligand removed) and 1YY9 (blue, antibody removed) aligned on residues 6–156 (domain I). Domain I is on the left, Domains II and IV in front, and Domain III on the right. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

(650 data points). Standard errors were estimated using the approach detailed in Supporting Information.

RESULTS

Dynamics of the bound and unbound receptor ectodomain

Flexibility

The root mean square deviations (RMSD) of C_{α} atoms were calculated relative to the structures from the initial minimization step for the four subdomains of the receptor ectodomain. For each simulation, the largest deviations are found for Subdomain II (see Supporting Information, Fig. S1) associated with a flexing along the length of this subdomain. This is in contrast to the soluble ectodomain dimer in which Subdomain II is stabilized by the dimer interface, and Subdomain IV is more flexible.²²

Relative motions of the subdomains were analyzed by computing RMSDs with respect to a fixed subdomain I (see Supporting Information, Fig. S2). Although Subdomain II maintains its position relative to Subdomain I, Subdomains III and IV fluctuate strongly, and usually in a correlated manner. The largest fluctuation is seen for the bIII simulation with a relative displacement of 70 Å reached after 106 ns of the simulation. This snapshot is compared in Figure 1 to the crystal structure from which

the starting configuration was taken (1NQL²), and a second crystal structure (1YY9³).

Analysis of RMSDs suggests a highly flexible hEGFR monomer, with Subdomains III and IV moving relative to Subdomain I, at least partly driven by flexing of Subdomain II. To obtain a clearer picture, we have computed angle distributions between two vectors V_1 and V_2 defining the orientations of Subdomains I and III (see Fig. 2). As reference points for these vectors, we chose the C_{α} atoms of residues Val36 and Glu118 in Subdomain I and the “equivalent” residues Ser340 and Glu431 in Subdomain III. As these residues are located at the ends of a relatively stable parallel β sheet and both domains are quite rigid (see Supporting Information, Fig. S1), effects of intramolecular motion are kept to a minimum.

As Figure 3 shows, the monomer simulations exhibit a broad distribution of angles between Subdomains I and III highlighting the large displacements that take place. Although there is considerable overlap in the distributions, there are clear differences between the three simulations. For comparison, the three available crystal structures of the tethered ectodomain show angles of 98.5° (1NQL), 113.0° (3QWQ⁴), and 130.1° (1YY9). Simulation bIII, with the EGF ligand bound to Domain III of the receptor, tends to larger angles than the other simulations, and larger than seen in any crystal structure.

We have also calculated the maximum curvature of a quadratic function fitted through the center of masses of the seven disulfide bond modules of Subdomain II following the procedure of Du *et al.*²⁹ This serves as a measure for the flexing of the “spine-like” structure⁶⁰ of the domain (see illustration in Fig. 2).

Figure 4 depicts the time series and the resulting distribution functions of the maximum curvatures from

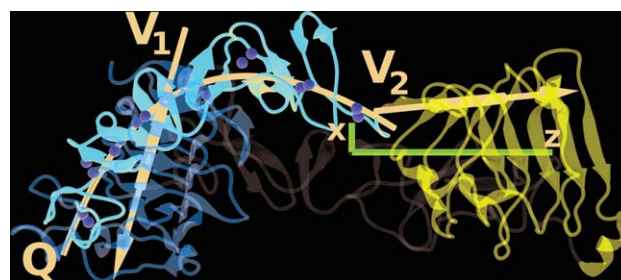
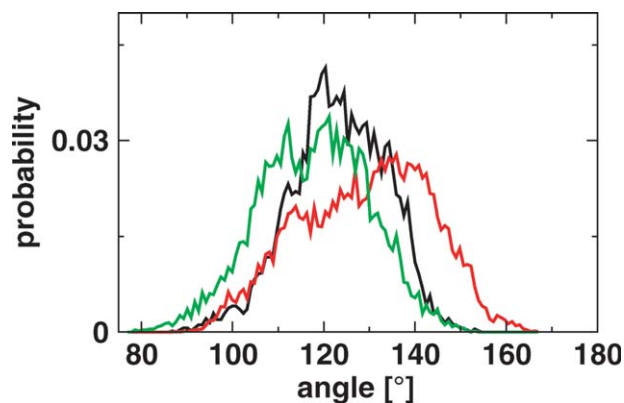


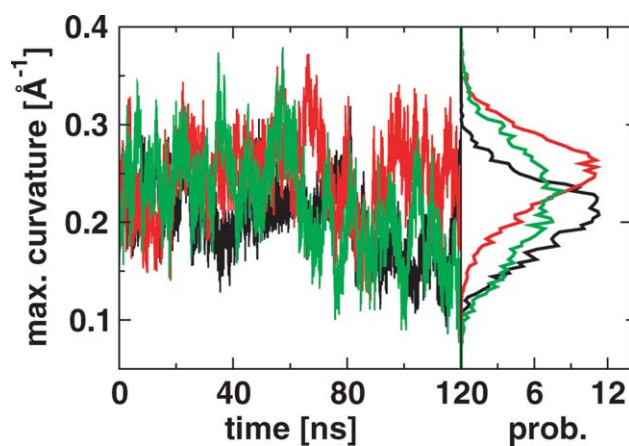
Figure 2

Definition of vectors V_1 and V_2 used to define the orientations of Subdomains I and III. Q denotes the quadratic fit curve through the center of mass of the disulfide modules of subdomain II. Sulfur atoms of disulfide bridges are shown as yellow spheres. The magenta coordinate system shows the eigenvectors (scaled by a factor of 1/20) of the R_g tensor with z the longest axis and x the shortest. Domain I (orange, top) is closest to the viewer and on top of domain II (red). Domain IV (grey) is farthest away. Domain III is shown in blue. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

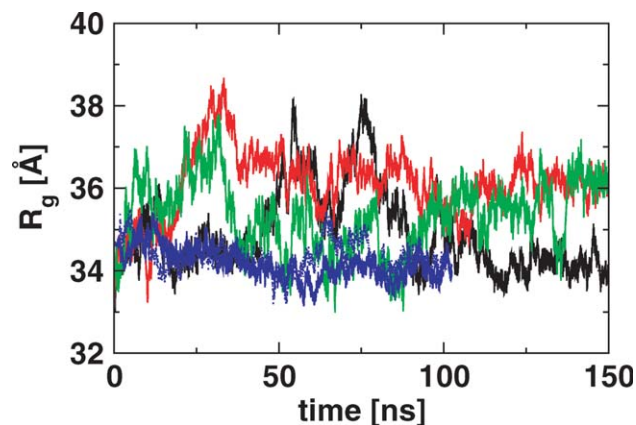
**Figure 3**

Angle distribution between the vectors V_1 and V_2 (for definitions see Fig. 2) for the monomeric simulations bI (black), bIII (red), and ub (green). All snapshots for each trajectory have been included in the calculation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

the three monomer simulations. The curvature of Subdomain II is consistently higher in the bIII simulation than in the bI simulation, and the distribution functions show clearly separated peaks (maxima at 0.21 \AA^{-1} for bI and at 0.25 \AA^{-1} for bIII). The unbound monomer shows a wider fluctuation of the curvature, and the distribution covers the range of the two other peaks. From the simulation of the soluble ectodomain dimer,²² the subdomain II curvatures of the constituent monomers peak at 0.28 \AA^{-1} and 0.34 \AA^{-1} . For comparison, the three available crystal structures of the tethered ectodomain show maximum curvatures of 0.36 \AA^{-1} (1NQL), 0.36 \AA^{-1} (3QWQ), and 0.19 \AA^{-1} (1YY9).

**Figure 4**

Maximum curvature of the quadratic fit curve Q (see Fig. 2) for the monomer simulations bI (black), bIII (red), and ub (green). The time series are shown on the left, and the resulting distribution on the right. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

**Figure 5**

Radius of gyration R_g as a function of time for the monomeric simulations bI (black), bIII (red), ub (green), and for both monomers of the dimeric hEGFR simulation (solid and dotted blue). The ligand has not been included in the calculations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Changes in shape

For each simulation, the gyration tensor has been calculated for the receptor (i.e., excluding the ligand atoms) as a function of simulation time. These tensors are highly anisotropic, reflecting the nonglobular shape of the ectodomain monomer; the average ratios of the eigenvalues are $833/311/76 \text{ \AA}^2$ for bI, $898/318/71 \text{ \AA}^2$ for bIII, and $875/302/71 \text{ \AA}^2$ for ub. The largest eigenvalue lies along the main axis of the molecule, and an example is shown in Figure 2. To compare with SAXS experiments, it is useful to compute the equivalent scalar radius of gyration R_g . Time courses of R_g are presented in Figure 5, and average values over the last 120 ns are summarized in Table I. The graphs show considerable fluctuations, with variations of up to 4 \AA in the case of the bI simulation.

The average R_g compares well with the experimental value of $35.4 \pm 0.11 \text{ \AA}$ obtained from SAXS.¹⁰ However, that value was obtained from a fully glycosylated protein,

Table I

Radius of gyration R_g and maximum dimension D_{\max} in \AA as obtained from MD simulation and experiment.¹⁰

	Simulation ^a		Experiment ^b	
	R_g	D_{\max}	R_g	D_{\max}
bI	35.0	126.3		
bIII	35.9	126.9		
Ub	35.2	125.4	35.4 ± 0.11	105 ± 5
Her1–Her1, monomer 1	34.1	139.0		
Her1–Her1, monomer 2	34.3	142.4		
Her1–Her1, dimer	42.3	142.4	46.1 ± 0.89	145 ± 5

^aNo carbohydrates.

^bFully glycosylated proteins. The ligand has been excluded.

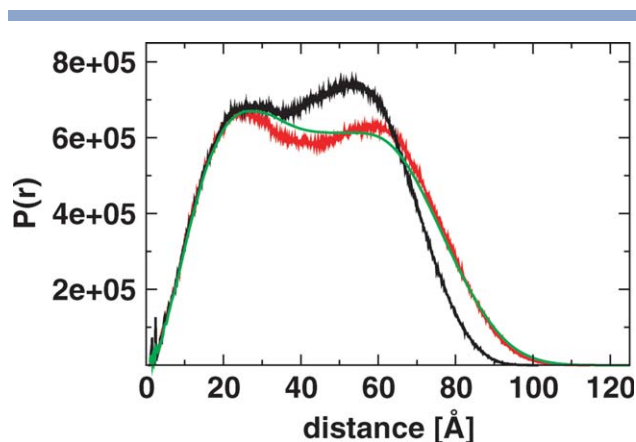


Figure 6

Pair distribution function $P(r)$ for the monomeric simulation ub (green) compared with data obtained from the receptor component of the X-ray structures 1NQL (black) and 1YY9 (red). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

whereas the model used in the current study is not glycosylated. Calculations on the crystal structure 1NQL have shown that model sugars ($\text{Man}_9\text{GlcNAc}_2$) attached to known and presumed glycosylation sites can add about 3 Å to the R_g when positioned approximately perpendicular on the molecular surface.¹⁰ Adding model sugars to the current MD simulations would likely lead to a smaller increase due to relaxation of the oligosaccharides. Furthermore, the oligosaccharides would be expected to interact with each other, with the protein surface, and with neighbouring receptor domains. These interactions may have a significant effect on the overall shape.

The R_g values computed for the individual monomers in the dimer simulation²² (with ligands excluded) are smaller than the values for the tethered monomers, with an average of around 34.2 Å. The average ratios of the eigenvalues of the gyration tensor are 840/244/85 Å² and 859/234/87 Å². Thus, the monomers in the dimer complex are more compact on average, despite being in the “extended” conformation. The latter is, however, reflected in the larger value for the maximum dimension D_{max} .

Figure 6 compares the average pair distribution function $P(r)$ obtained from the simulation of the unbound monomer (last 120 ns) with curves calculated directly from the receptor component of the crystal structures 1NQL² and 1YY9.³ 1NQL was used to construct the starting structures for the simulations, whereas 1YY9 (which is in complex with a Fab fragment) has a significantly different conformation (see Fig. 1). The simulation curve encompasses both the short-distance peak of 1NQL and the large-distance tail of the less compact 1YY9. Thus, although the simulation was started from a

particular crystal structure, it has explored different conformations, including ones similar to that captured in a different crystal structure. The differences between the three simulations (Supporting Information, Fig. S4) are relatively smaller, confirming that each system shows this flexibility.

Principal component analysis

Principal component analysis (PCA) has been carried out to gain further insights into the dynamics of the three hEGFR simulations. As large proteins are notoriously slow to converge;⁶¹ however, we applied this method here only to reveal differences between the dynamics of the three simulations. The mutual scalar products of the first five eigenvectors (Supporting Information, Tables SI and SII) show that with further progress of the simulations the individual modes become less similar, meaning that the three simulations diverge and lead to different global motions.

Projecting the trajectory against the first two PCA modes (Supporting Information, Fig. S7) for each simulation further confirms different dynamics in the three systems. For example, the bI simulation displays possible energy barriers in the 2D trajectory, while phase space is populated more densely in the other simulations. However, the probabilities obtained from the projections of the trajectory onto the first five eigenvectors (Supporting Information, Fig. S6) indicate strong non-Gaussian distributions in the lowest modes. The largest amplitude motions cannot be expected to be harmonic but the rather broad distribution suggests sampling issues. It would therefore be misleading to attach too much significance to the motions implied by the PCA modes.

Binding free energies of the ligand

The binding free energy ΔG of the ligand EGF to its monomeric receptor hEGFR has been estimated with the MM-PBSA method.^{33,34,53} We are particularly interested in the $\Delta\Delta G$ between EGF bound to Domain I and EGF bound to Domain III. This may suggest where (if at all) the ligand preferentially binds to the monomeric receptor, as well as identify the relative contributions of the two interfaces in the ligated dimer. Absolute ΔG s are difficult to obtain^{35,62,63} especially in the case of the large complexes discussed here.

The crystal structure 1NQL was found to bind the ligand weakly at domain I at pH 5.² The interactions of the EGF–Domain I interface primarily involve backbone hydrogen bonds which are not expected to be affected by changes in pH. In contrast, the EGF–Domain III interface may be disrupted at lower pH because of protonation of three histidine side chains in Domain III (see Simulation Setup). At neutral pH, a Domain III fragment was found to bind EGF strongly with a K_d of about 450 nM.⁶⁴

Table II

Free energies of binding of EGF to Domain I or Domain III of the receptor obtained with the MM-PBSA method.

	Res ^a	Domain I		Domain III	
		Single	Three	Single	Three
$\Delta G_{\text{noS}}^{\text{b}}$	0.75	-62.4	-52.4	-52.4	-78.9
ΔG_{noS}	0.5	-76.8	-70.4	-61.2	-96.7
ΔE_{ele}	0.3	-222.8 ± 45.2	-400.0 ± 80.9	-105.4 ± 29.9	-272.5 ± 75.8
ΔE_{vdW}		-69.1 ± 5.5	-106.7 ± 7.6	-61.2 ± 6.6	-68.1 ± 8.0
ΔE_{int}		0.0	28.3 ± 7.5	0.0	3.7 ± 7.0
ΔE_{MM}		-291.9 ± 49.0	-478.5 ± 83.2	-166.6 ± 34.2	-336.8 ± 78.6
ΔG_{PB}		222.4 ± 38.9	417.2 ± 74.6	107.8 ± 31.8	258.3 ± 72.6
ΔG_{SA}		-12.1 ± 0.6	-14.8 ± 0.8	-10.6 ± 0.3	-8.7 ± 0.7
ΔG_{noS}		-81.6 ± 16.5	-76.1 ± 13.2	-69.3 ± 9.4	-87.3 ± 9.4
$-T\Delta S_{\text{rt}}^{\text{c}}$			30.9		30.7
$-T\Delta S_{\text{v}}^{\text{d}}$			8.5		11.6
$-T\Delta S_{\text{tot}}^{\text{e}}$			39.3		42.3
$\Delta G_{\text{tot}}^{\text{f}}$			-36.8		-45.0

^aGrid resolution in PB calculation in Å.^b $\Delta G_{\text{noS}} = E_{\text{MM}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$.^cRotational plus translational entropy.^dVibrational entropy from normal mode analysis.^e $\Delta S_{\text{tot}} = \Delta S_{\text{rt}} + \Delta S_{\text{v}}$.^f $\Delta G_{\text{tot}} = \Delta G_{\text{noS}} - T\Delta S_{\text{tot}}$ based on 0.3 Å resolution.

Results from both the single trajectory and three trajectories approach are presented. Energies are in kcal/mol.

Binding to the full-length receptor is similar in strength, but falls off at low pH as Domain III binding is disrupted.² The residual Domain I binding is therefore presumed to be weak, and the binding seen in 1NQL to be an artifact of the protein concentrations used for crystallization.

In Table II, we summarize computed binding free energies obtained with both the single trajectory and three trajectories approaches (see Methods). The table also shows a test of the grid resolution used in the Poisson–Boltzmann calculation for the polar part of the solvation free energy. ΔG is sensitive to the grid resolution used, with a finer grid leading to a more negative ΔG in most cases. Although there is no obvious convergence with respect to the grid resolution, finer grids lead to smaller fluctuations over the trajectory (see Supporting Information, Fig. S8) and are assumed to be more accurate.

Comparing the single with the three trajectories approach, we find that they give very different ΔG values. Moreover, the two approaches suggest opposite binding preferences, with the ligand preferring Domain I with the single-trajectory approach and Domain III with the three trajectories approach. The high flexibility of the ectodomain and the differences seen between the bound and unbound simulations indicate that representative receptor or ligand conformations cannot realistically be taken from the trajectory of the complex. This can be quantified with the reorganization free energy^{65,66} $\Delta G_{\text{reorg,noS}}$ which is the free energy difference of a particular component in the complex and in an unbound state. We find (ignoring the conformational entropy) 5.9 and 12.3 kcal/mol for the ligand, and -0.4 and -30.2 kcal/mol for the receptor in the bI and bIII simulations, respectively. We thus conclude that the single-trajectory approach is not valid for the

system considered here. Finally, ΔG s were calculated only for the three trajectories approach.

The vibrational entropy ΔS_{v} has been estimated through normal mode analysis. However, obtaining entropies from the very large hEGFR–ligand complex ($\approx 10,000$ atoms) is not only very time and memory consuming but it is also difficult to calculate reliable minimized structures. Therefore, we followed a previous suggestion⁶⁷ to work on a reduced system. Normal modes are calculated for a flexible region consisting of the ligand and all receptor residues within 8 Å of the ligand. A buffer region with fixed atoms prevents distortions of the flexible region during energy minimization, and avoids problems of uncapped residues. Further details on the definition of the regions and the entropy calculation are given in Supporting Information.

As shown in Table II, ΔS_{v} is similar in magnitude for both Domain I and Domain III binding. As the ligand is also in close contact to Domains II and III in the Domain I binding case, we included also parts of Domain II and III (residues 261–375) in the entropy calculation and computed a new ΔS_{v} which was almost the same (8.8 kcal/mol) as in the smaller system (8.5 kcal/mol). We conclude that the change in vibrational entropy has only a minor effect on the final ΔG_{tot} . In many MM–PBSA studies, it is in fact neglected from the outset (see e.g., Refs. 27 and 68).

The final result for ΔG_{tot} predicts that domain III binding is more likely than Domain I binding by about 8 kcal/mol. The statistical uncertainty in ΔG_{noS} , however, is slightly larger than the final $\Delta\Delta G$ of binding, and so the preference is not clear cut.

Table II also summarizes component energies contributing to the free energy of binding. In general, the

Table III

Free energies of binding for selected mutations computed via alanine-scanning.

	L47A		R41A		K28A		I23A		A25A	
	dl	dIII	dl	dIII	dl	dIII	dl	dIII	dl	dIII
ΔE_{cle}	-400.0	-272.5	-382.6	-180.0	-376.7	-246.8	-400.1	-272.5	-400.0	-272.5
ΔE_{vdW}	-107.0	-63.5	-105.8	-67.7	-106.6	-68.0	-104.2	-68.0	-106.7	-68.1
ΔE_{int}	28.2	3.7	27.8	4.3	27.1	3.9	28.2	3.8	28.2	3.6
ΔE_{MM}	-478.8	-332.3	-460.6	-243.2	-456.2	-310.9	-476.2	-336.7	-478.5	-337.0
ΔG_{PB}	417.2	257.7	400.0	173.0	395.1	233.7	417.0	258.1	417.2	258.3
ΔG_{SA}	-14.8	-8.4	-14.7	-8.5	-14.6	-8.7	-14.8	-8.7	-14.8	-8.7
ΔG_{noS}	-76.4	-83.0	-75.3	-78.7	-75.7	-85.9	-74.0	-87.3	-76.1	-87.4
$\Delta\Delta G_{\text{noS}}$	-0.3	4.3	0.9	8.6	0.4	1.4	2.1	0.0	0.0	-0.1
Expt.		2.5 ³⁶ 2 ³⁷		0.05 ^{38,39} 180–188 ^b		79–95 ^a	3 ⁴¹	—		

^aRelative binding affinity for mutation K28R.⁴⁰^bRelative binding affinity for mutation K28L.⁴⁰The contribution of the conformational entropy is not included. All results are for the three trajectories approach. $\Delta\Delta G_{\text{noS}}$ gives the change with respect to the wild-type free energy of binding. Energies are in kcal/mol. Experimental relative binding affinities = $IC_{50}(\text{wild type})/IC_{50}(\text{mutant})$ are in %.

magnitudes of the components are larger in the three trajectories approach than they are in the single-trajectory approach, suggesting that the assumption behind the single-trajectory approach is suppressing some energetic changes. For both approaches, the gas-phase binding energy ΔE_{MM} favors complex formation with a preference for Domain I binding over Domain III binding. This is driven by more favorable changes in the electrostatic energy ΔE_{cle} , and to a lesser extent in the van der Waals energy ΔE_{vdW} . In contrast, solvation contributions to the free energy change, and in particular the Poisson–Boltzmann contributions ΔG_{PB} , oppose complex formation but are less unfavorable for Domain III binding. The overall free energy change for ligand binding is the result of a competition between gas-phase energetics and solvation effects. In the three trajectories approach, the latter effect is stronger, and ΔG_{tot} favors binding of EGF to Domain III.

Ligand mutations

In Table III, we summarize binding free energies for four single-point mutations of the EGF ligand, plus a null mutation included as a control. Leu47 and Arg41 are involved in Domain III binding, and are conserved across five hEGFR ligands: EGF, TGF α , betacellulin, epiregulin, and heparin-binding EGF-like growth factor. Lys28 and Ile23 are involved in Domain I binding; Lys28 is conserved in TGF α only, whereas Ile23 is replaced by similar side chains (Leu and Val) in other hEGFR ligands. Results were obtained through the alanine-scanning method in which a single residue is mutated into an alanine, but protein conformations are taken from the original simulations.³³ The alanine-scanning method applied here has the limitation that it assumes conformations of the mutant are identical to those of the wild-type, that is, no

new simulation is run. In general, NMR and binding replacement data find only small localized changes to the mutant–receptor interactions with no significant changes to the fold of the ligand,^{36,38} but this might not always be the case.⁶⁹

The L47A and R41A mutations have little effect on Domain I binding, but are unfavorable for Domain III binding as expected. Comparing the component energies with those of the wild-type (Table II), the difference for L47A comes largely from a change in ΔE_{vdW} , whereas the difference for R41A comes from large (and partially compensating) changes in ΔE_{cle} and ΔG_{PB} . The I23A mutation is close to the wild type in Domain III binding, but ΔG_{noS} is slightly less favorable for Domain I binding. As for L47A, this difference comes largely from a change in ΔE_{vdW} . The K28A mutation has only a small effect on both Domain I and Domain III binding. As for R41A, the electrostatic contributions vary strongly from the wild type, but in this case largely cancel out.

It must be noted that the $\Delta\Delta G_{\text{noS}}$ values calculated for these mutants are smaller than the standard errors of the individual ΔG_{noS} . On the other hand, we are reassured that the changes in the individual components make chemical sense, and that the control calculation for A25A gives essentially zero (not necessarily precisely zero because the alanine-scanning protocol rebuilds side chains in an idealized position and conformation).

DISCUSSION

Molecular simulation of the tethered hEGFR monomer provides a detailed picture of the dynamics of both the unbound state (ub) and bound states where the EGF ligand binds to either domain I (bI) or domain III (bIII). In all three cases, the monomer remains in the tethered

conformation for the duration of the simulation. Thus, any major structural rearrangement is likely to occur on a longer timescale than the 150 ns probed by our atomistic simulation. This is in agreement with a recent simulation study of the ErbB4 ectodomain monomer, with the ligand NRG1 β bound to Domain I, which showed a transition to a partially extended state at around 400 ns after breaking of the Domain II/Domain IV tether.²⁹ Recent experimental studies of tether mutants¹⁰ and a truncated form of ErbB4 missing Domain IV⁷⁰ show that the tether is not required to maintain a tether-like conformation, suggesting that there are several stabilizing interactions to be overcome.¹⁰

Nevertheless, our simulations do reveal that the tethered monomer is extremely flexible as demonstrated by, for example, angle (Fig. 3) and maximum curvature (Fig. 4) distributions. Correlated motions of Domain III and IV with respect to Domains I and II are clearly visible in RMSDs (see Supporting Information, Fig. S2). The largest deviations are seen in bIII, where the structure may move away from the starting structure by as much as 70 Å and fluctuations regularly span 30 Å. These motions are in the same overall direction as revealed by the difference between the two crystal structures 1NQL² and 1YY9³ (see Fig. 1). Li *et al.*³ suggested that these two structures “provide two snapshots of a flexible domain II, which have been trapped by different crystal packing environments,” and our simulation supports this interpretation. A recent crystal structure of the tethered monomer in complex with an adnectin⁴ shows a third relative position of Domain III, intermediate between 1NQL and 1YY9.

Dimerization has a strong effect on both intradomain and interdomain motions generally leading to a globally less dynamic system. In previous work,^{17,22} we have discussed the flexibility of the ectodomain dimer, in particular when placed in a membrane environment. Although such flexibility is significant, and may be relevant for ligand affinity in the dimer, the flexibility seen in the current study for the monomer is much larger.

From a study of tryptophan fluorescence, Kozar *et al.*¹⁴ found evidence for multiple conformations of the unliganded monomer of a truncated hEGFR ectodomain, as well as a rotational correlation time that was shorter than that for a rigid monomer. In contrast, in the presence of ligand there was only a single conformation with a rotational correlation time appropriate to a rigid complex. They interpreted this in terms of an equilibrium between tethered and open conformations in the absence of ligand, with ligand-binding stabilizing the open conformation (here, we use “tethered” as a convenient label, although their truncated form of hEGFR precluded the formation of the Domain II/Domain IV tether). Our simulation results suggest an alternative interpretation. Even in the tethered conformation, the unbound monomer shows considerable flexibility, with changes in the relative orientation between Domains I and III (see Fig. 3), and

this may be sufficient to explain the experimental results. The ligand bound dimer is certainly less flexible, and the same is probably true of the ligand-bound extended monomer.

A central question for growth factor binding is whether these ligands bind to the tethered conformation, and if so whether there is a preference for the Domain I or the Domain III binding site. In their simulation of the tethered ErbB4 ectodomain, Du *et al.*²⁹ conclude that the ligand neuregulin-1 β preferentially binds to Domain I, based on a trial simulation where the ligand moves to Domain I from a starting position between Domains I and III. We note that such a simulation is likely to be highly dependent on the starting position and orientation of the ligand. Furthermore, it is known that the Domain III ligand binding site is rotated 130° away from the Domain I binding site in going from the extended to the tethered conformation (see Ref. 2 for hEGFR or compare 2AHX⁶⁰ and 3U7U⁷¹ for ErbB4). Thus, the ligand cannot be orientated so as to potentially satisfy both binding surfaces.

We therefore consider both Domain I and Domain III binding to remain potentially relevant. Ligand binding in both systems (bI and bIII) appears stable over the 150 ns of simulation. Nevertheless, we also see clear differences between these simulations, suggesting that the location of ligand binding has an effect on the dynamics of the monomer. Figure 3 shows a difference in the interdomain angle distribution of bIII, compared to the bI and ub simulations. Although the change in the relative orientation of domain III is not in the direction required to form the extended conformation, it does suggest that ligand binding to Domain III induces a change in conformation which may be relevant, whereas (according to this measure) no such change occurs upon binding to Domain I. Correlated motions of Domains III and IV are also in general larger when the ligand is bound to Domain III (see Supporting Information, Fig. S2).

Conversely, binding of the ligand to Domain I leads to larger fluctuations in Domain II than when the ligand binds to Domain III (see Supporting Information, Fig. S1). Du *et al.*²⁹ highlight Domain II bending during the first stage of their simulation (in which the neuregulin-1 β ligand is bound to Domain I), suggesting that a large fluctuation eventually leads to the breaking of the tether. On the other hand, the distribution of the maximum curvature of Domain II (see Fig. 4) shows larger curvatures when the ligand is bound to Domain III, and so the mechanism proposed by Du *et al.* may be applicable in this case as well.

Application of the MM-PBSA method gives some insight into the free energy of ligand binding. The highly dynamical hEGFR monomers rule out the use of the single-trajectory approach in which it is assumed that representative conformations for all three of complex, receptor, and ligand can be obtained from a single

simulation of the complex. The large conformational changes seen over tens of ns also support the use of relatively long simulation times (i.e., 150 ns). An analysis of the correlation between snapshots used in the MM-PBSA calculation shows correlation times up to 10 ns for some components. For comparison, Luo *et al.*⁷² performed similar calculations for ErbB3 and ErbB4 complexes taking snapshots from 1 ns simulations, whereas Fuentes *et al.*⁶⁸ applied MM-GBSA to complexes of ErbB2 with antibodies taking snapshots from 20 ns simulations. Although the latter studies were applied to the more rigid extended conformation, it is clear that longer simulation times are necessary to get adequate statistics for the tethered conformation.

Both Domain I and Domain III of hEGFR contribute to ligand binding in the extended conformation.^{5,6} In contrast, if the ligand binds to the tethered conformation, then it can only form an interface with one of the domains. There may be an energetic preference, and the binding free energies estimated here from the three trajectories approach suggest that the EGF ligand binds more favorably to Domain III than to Domain I. Loosely, Domain I binding is driven by gas-phase electrostatic interactions, whereas Domain III binding is driven by solvation terms, and overall the latter is predicted to be the larger effect (Table II). However, the difference in binding free energy for bI and bIII is comparable to the associated standard errors and is perhaps not significant. Because Domain III binding is driven by solvation effects, its relative strength will be affected by changes in the solvent composition and this may be another factor behind the observed preference for Domain I in the crystal structure.²

In a recent comparison of different EGFR ligands, Sanders *et al.*²⁰ find a ΔG_{bind} for binding of EGF to dimeric EGFR of -102 kcal/mol. Entropy terms are not included, so this is to be compared to the values of -76 and -87 kcal/mol taken from Table II for binding to Domain I and Domain III, respectively. A comparison hints that binding to the extended conformation is more favorable, but that the contributions of Domains I and III are not simply additive.

The changes in binding free energy we have found for selected ligand mutants (Table III) are mostly in qualitative agreement with experiment.^{36–42} Experimentally, the largest reductions in binding affinity have been found for mutants of Arg41, with R41A having 0.05% of the affinity of the wild type (determined as the ratio of IC_{50} values).^{38,39} The affinity of L47A is reduced to 2³⁷ and 2.5%³⁶ of the wild-type, whereas the affinity of I23A is reduced to 5.9%.⁴¹ The relative order of these three mutants is the same as in our calculations (see Table III). A K28A mutant has not been reported in the literature, but the affinity of a K28L mutant was 79% of the wild-type, whereas the K28R variant was 188%.⁴⁰ Our calculation showing the K28A mutant to have only a minor

effect on binding is consistent with these results, and with the lack of conservation of Lys28 among hEGFR ligands.

Figure 7 shows zoomed snapshots of residues Leu47, Arg41, Lys28, and Ile23, together with their key interactions. In the simulation, Arg41 of EGF forms a strong salt bridge to Asp355, as is observed in the crystal structure.⁵ In the alanine variant, this interaction is removed and the aspartic acid is exposed to a mostly hydrophobic tightly packed environment (Val350, Phe357 and Thr358, and Leu15 from the ligand). This is consistent with the guanidinium group being an absolute requirement for high affinity.³⁹ In our simulations, as in the crystal structure,⁵ Leu47 is embedded in a hydrophobic pocket (Leu382, Phe412, Val417, Ser440, Ile438, Ile467). The mutation to alanine causes a disfavorable change in the van der Waals interactions as the smaller residue packs less well into the pocket. Despite breaking the strong salt bridge to Glu90 on Domain I of the receptor in the K28A variant, the predicted effect on domain I binding affinity is marginal. Measuring the heavy atom distance (CD–NZ) for that salt bridge we find that this distance is smaller than 4.0 Å in 91% of the total simulation time, which might suggest that the salt bridge does contribute to the binding. In contrast, if we look at the same salt bridge in our previous simulations of the liganded-dimer,^{17,22} then we find that it is often broken in the symmetric dimer in solution, and almost always broken in the asymmetric dimer on the membrane. We conclude that this salt bridge is in fact weak. Finally, Ile23 is in close contact to Leu14, Leu69, and Tyr45 in Domain I of the receptor, the latter making a hydrogen bond to the backbone of Leu14. Mutation to alanine will reduce the van der Waals interactions, but the calculated effect on Domain I binding is quite small.

Our results for key mutations thus make some sense in terms of the known structural interactions. They confirm the experimental affinity measurements and suggest that specific ligand interactions with Domain III are vital for ligand binding. Although ligand binding to Domain I is predicted to have a similar strength overall, partly due to backbone hydrogen bonding, specific side chain interactions appear to be less important.

In general, the current results may be limited by the simplicity of the structural model. First, we have not modeled the membrane environment, in contrast to our earlier study²² where it played a crucial role. Second, we have not modeled the oligosaccharide chains which can decorate the protein at up to 12 N-linked glycosylation sites.^{73,74} It has been argued that disruption of the steric restraints due to the presence of oligosaccharides may be one factor required to break the tether.¹⁰

In summary, our calculations suggest that the EGF ligand can bind to the tethered monomer, and that binding has an effect on the dynamics of this flexible protein. There are clear differences between Domain I and Domain III binding of the ligand, with some hints that

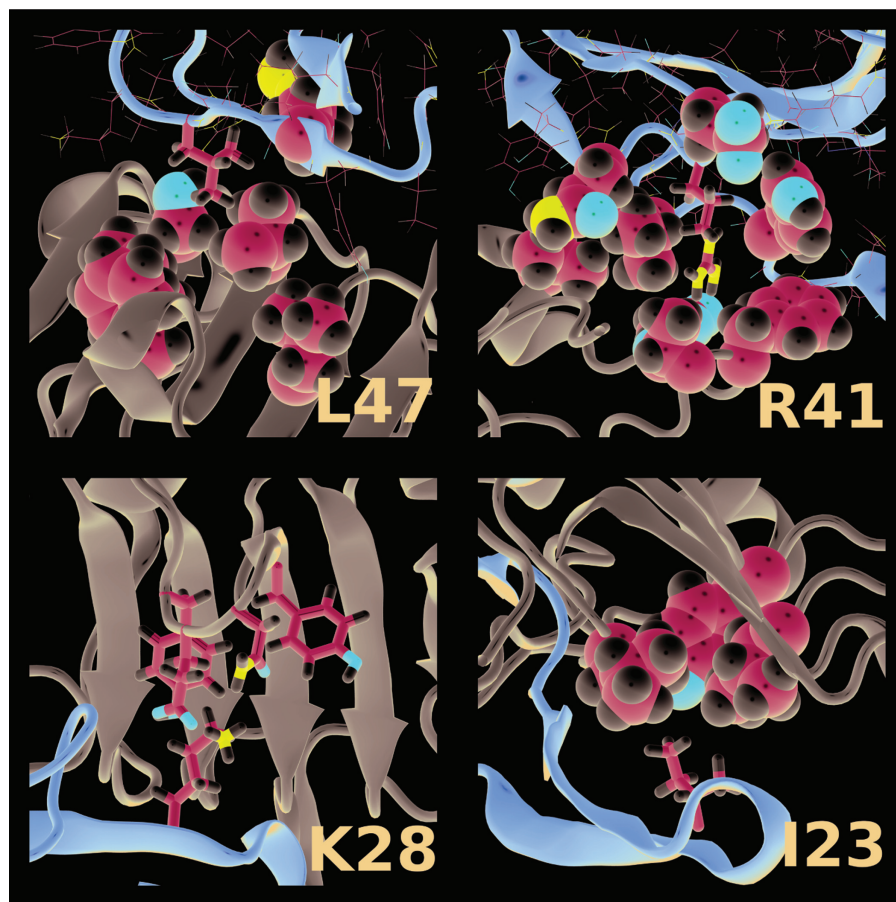


Figure 7

The interactions of the residues Leu47, Arg41, Lys28, and Ile23 with their immediate environment. Key residues are highlighted. Leu47: Val417, I438, S440 in direct contact with Leu, Phe412, and Leu382 flanking this region, Arg45 from the ligand; Arg41: salt bridge to Asp355, near neighbors Val350, Phe357, Thr358, and Tyr13, Leu15, Glu40, Gln43 from the ligand; Lys28: Tyr89, Glu90, Asn91, Tyr93; Ile23: Leu14, Leu69, Tyr45. hEGFR backbone in grey and EGF backbone in brown.

Domain III binding produces more significant changes in the average conformation. Free energy estimates also show a slight preference for ligand binding to Domain III. Although we find that ligand binding to Domain I is also possible, the evidence of a role for specific ligand residues is not so strong. These two binding interfaces are of course also relevant to ligand binding in the extended dimer, where both contribute to the overall binding. If asymmetric dimers^{16,17} occur, then the relative contributions of the two binding interfaces could vary.

ACKNOWLEDGMENT

Most simulations were run on an IBM Blue Gene/P system, provided by the Science and Technology Facilities Council. Additional simulations were run on HECToR, the UK's national high-performance computing service. The authors thank Prof. Mark Lemmon for providing us with the experimental data plotted in Figure S5.

REFERENCES

1. Yarden Y, Sliwkowski MX, Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2001;2:127–137.
2. Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA, EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell* 2003;11:507–517.
3. Li S, Schmitz KR, Jeffrey PD, Wiltzius JJW, Kussie P, Ferguson KM, Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer cell* 2005;7:301–311.
4. Ramamurthy V, Krystek SR, Jr, Bush A, Wei A, Emanuel SL, Das Gupta R, Janjua A, Cheng L, Murdock M, Abramczyk B, Cohen D, Lin Z, Morin P, Davis JH, Dabritz M, McLaughlin DC, Russo KA, Chao G, Wright MC, Jenny VA, Engle LJ, Furfine E, Sheriff S, Structures of adnectin/protein complexes reveal an expanded binding footprint. *Structure* 2012;20:259–269.
5. Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S, Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 2002;110:775–787.
6. Garrett TPJ, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Zhu HJ, Walker F, Frenkel MJ, Hoyne PA, Jorissen RN, Nice

- EC, Burgess AW, Ward CW, Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α . *Cell* 2002;110:763–773.
7. Choowongkamon K, Carlin CR, Sönnichsen FD, A Structural model for the membrane-bound form of the juxtamembrane domain of the epidermal growth factor receptor. *J Biol Chem* 2005;280:24043–24052.
 8. Mineev KS, Bocharov EV, Pustovalova YE, Bocharova OV, Chupin VV, Arseniev AS, Spatial structure of the transmembrane domain heterodimer of ErbB1 and ErbB2 receptor tyrosine kinases. *J Mol Biol* 2010;400:231–243.
 9. Sato T, Pallavi P, Golebiewska U, McLaughlin S, Smith SO, Structure of the membrane reconstituted transmembrane–juxtamembrane peptide EGFR(622–660) and its interaction with Ca^{2+} /calmodulin. *Biochemistry* 2006;45:12704–12714.
 10. Dawson JP, Bu Z, Lemmon MA, Ligand-induced structural transitions in ERBB receptor extracellular domains. *Structure* 2007;15:942–954.
 11. Mi LZ, Grey MJ, Nishida N, Walz T, Lu C, Springer TA, Functional and structural stability of the epidermal growth factor receptor in detergent micelles and phospholipid nanodiscs. *Biochemistry* 2008;47:10314–10323.
 12. Mi LZ, Lu C, Li Z, Nishida N, Walz T, Springer TA, Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. *Nat Struct Mol Biol* 2011;18:984–9.
 13. Lu C, Mi LZ, Schürpf T, Walz T, Springer TA, Mechanisms for kinase-mediated dimerization of the EGF receptor. *J Biol Chem* 2012;287:38244–38253.
 14. Kozer N, Rothacker J, Burgess AW, Nice EC, Clayton AHA, Conformational dynamics in a truncated epidermal growth factor receptor ectodomain. *Biochemistry* 2011;50:5130–5139.
 15. Macdonald JL, Pike LJ, Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proc Natl Acad Sci* 2008;105:112–117.
 16. Alvarado D, Klein DE, Lemmon MA, Structural basis for negative cooperativity in growth factor binding to an EGF receptor. *Cell* 2010;142:568–579.
 17. Tynan CJ, Roberts SK, Rolfe DJ, Clarke DT, Loeffler HH, Kastner J, Winn MD, Parker PJ, Martin-Fernandez ML, Human epidermal growth factor receptor (EGFR) aligned on the plasma membrane adopts key features of Drosophila EGFR asymmetry. *Mol Cell Biol* 2011;31:2241–2252.
 18. Zhang Z, Wriggers W, Polymorphism of the epidermal growth factor receptor extracellular ligand binding domain: the dimer interface depends on domain stabilization. *Biochemistry* 2011;50:2144–2156.
 19. Perilla JR, Leahy DL, Woolf TB, Molecular dynamics simulations of transitions for ECD epidermal growth factor receptors show key differences between human and drosophila forms of the receptors. *Proteins* 2013;81:1113–1126.
 20. Sanders JM, Wampole ME, Thakur ML, Wickstrom E, Determinants of epidermal growth factor binding: a molecular dynamics study. *PLoS ONE* 2013;8:e54136.
 21. Arkhipov A, Shan Y, Das R, Endres N, Eastwood M, Wemmer D, Kuriyan J, Shaw D, Architecture and membrane interactions of the EGF receptor. *Cell* 2013;152:557–569.
 22. Kästner J, Loeffler HH, Roberts SK, Martin-Fernandez ML, Winn MD, Ectodomain orientation, conformational plasticity and oligomerization of ErbB1 receptors investigated by molecular dynamics. *J Struct Biol* 2009;167:117–128.
 23. Wan S, Coveney PV, Molecular dynamics simulation reveals structural and thermodynamic features of kinase activation by cancer mutations within the epidermal growth factor receptor. *J Comput Chem* 2011;32:2843–2852.
 24. Shih AJ, Telesco SE, Choi SH, Lemmon MA, Radhakrishnan R, Molecular dynamics analysis of conserved hydrophobic and hydrophilic bond-interaction networks in ErbB family kinases. *Biochem J* 2011;436:241–249.
 25. Mustafa M, Mirza A, Kannan N, Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study. *Proteins* 2011;79:99–114.
 26. Liao QH, Gao QZ, Wei J, Chou KC, Docking and molecular dynamics study on the inhibitory activity of novel inhibitors on epidermal growth factor receptor (EGFR). *Med Chem* 2011;7:24–31.
 27. Balius TE, Rizzo RC, Quantitative prediction of fold resistance for inhibitors of EGFR. *Biochemistry* 2009;48:8435–8448.
 28. Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE, Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* 2012;149:860–870.
 29. Du Y, Yang H, Xu Y, Cang X, Luo C, Mao Y, Wang Y, Qin G, Luo X, Jiang H, Conformational transition and energy landscape of erbb4 activated by neuregulin1 beta: one microsecond molecular dynamics simulations. *J Am Chem Soc* 2012;134:6720–6731.
 30. Gallicchio E, Levy RM, Advances in all atom sampling methods for modeling protein ligand binding affinities. *Curr Opin Struct Biol* 2011;21:161–166.
 31. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M, Validation and use of the mm-PBSA approach for drug discovery. *J Mol Chem* 2005;48:4040–4048.
 32. Stoica I, Sadiq SK, Coveney PV, Rapid and accurate prediction of binding free energies for Saquinavir–Bound HIV–1 proteases. *J Am Chem Soc* 2008;130:2639–2648.
 33. Massova I, Kollman PA, Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. *J Am Chem Soc* 1999;121:8133–8143.
 34. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;33:889–897.
 35. Gohlke H, Case DA, Converging free energy estimates: MM-PB(GB)SA studies on the protein protein complex Ras Raf. *J Comput Chem* 2004;25:238–250.
 36. Dudgeon T, Cooke R, Baron M, Campbell I, Edwards R, Fallon A, Structure-function analysis of epidermal growth factor: site directed mutagenesis and nuclear magnetic resonance. *FEBS Lett* 1990;261:392–396.
 37. Matsunami RK, Campion SR, Niyogi SK, Stevens A, Analogs of human epidermal growth factor which partially inhibit the growth factor-dependent protein-tyrosine kinase activity of the epidermal growth factor receptor. *FEBS Lett* 1990;264:105–108.
 38. Engler DA, Montelione GT, Niyogi SK, Human epidermal growth factor: distinct roles of tyrosine 37 and arginine 41 in receptor binding as determined by site-directed mutagenesis and nuclear magnetic resonance spectroscopy. *FEBS Lett* 1990;271:47–50.
 39. Engler DA, Campion SR, Hauser MR, Cook JS, Niyogi SK, Critical functional requirement for the guanidinium group of the arginine 41 side chain of human epidermal growth factor as revealed by mutagenic inactivation and chemical reactivation. *J Biol Chem* 1992;267:2274–2281.
 40. Campion SR, Matsunami RK, Engler DA, Niyogi SK, Biochemical properties of site-directed mutants of human epidermal growth factor: the importance of solvent-exposed hydrophobic residues of the amino-terminal domain in receptor binding. *Biochemistry* 1990;29:9988–9993.
 41. Koide H, Muto Y, Kasai H, Kohri K, Hoshi K, Takahashi S, Tsukumo K, Sasaki T, Oka T, Miyake T, Fuwa T, Kohda D, Inagaki F, Miyazawa T, Yokoyama S, A site-directed mutagenesis study on the role of isoleucine–23 of human epidermal growth factor in the receptor binding. *Biochim Biophys Acta* 1992;1120:257–261.
 42. Koide H, Yokoyama S, Katayama Y, Muto Y, Kigawa T, Kohno T, Takusari H, Oishi M, Takahashi S, Tsukumo K, Sasaki T, Miyake T, Fuwa T, Kawai G, Miyazawa T, Receptor-binding affinities of

- human epidermal growth factor variants having unnatural amino acid residues in position 23. *Biochemistry* 1994;33:7470–7476.
43. Li H, Robertson AD, Jensen JH, Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 2005;61:704–721.
 44. Bas DC, Rogers DM, Jensen JH, Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins* 2008;73:765–783.
 45. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res* 2004;32:W665–W667.
 46. Nose S, A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* 1984;81:511–519.
 47. Hoover WG, Canonical dynamics: equilibrium phase–space distributions. *Phys Rev A* 1985;31:1695–1697.
 48. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K, Scalable molecular dynamics with NAMD. *J Comput Chem* 2005;26:1781–1802.
 49. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.
 50. Mackerell AD, Feig M, Brooks CL, Extending the treatment of backbone energetics in protein force fields: Limitations of gas–phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 2004;25:1400–1415.
 51. Beglov D, Roux B, Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J Chem Phys* 1994;100:9050–9063.
 52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
 53. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA, Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J Am Chem Soc* 1998;120:9401–9409.
 54. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217.
 55. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M, CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;30:1545–1614.
 56. Loeffler HH, available at: <http://www.stfc.ac.uk/CSE/randd/cbg/software/25247.aspx>. 2010.
 57. Sitkoff D, Sharp KA, Honig B, Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 1994;98:1978–1988.
 58. Nina M, Beglov D, Roux B, Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J Phys Chem B* 1997;101:5239–5248.
 59. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA, Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci* 2001;98:10037–10041.
 60. Bouyain S, Longo PA, Li S, Ferguson KM, Leahy DJ, The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc Natl Acad Sci USA* 2005;102:15024–15029.
 61. Grossfield A, Feller SE, Pitman MC, Convergence of molecular dynamics simulations of membrane proteins. *Proteins* 2007;67:31–40.
 62. Lee MS, Olson MA, Calculation of absolute protein–ligand binding affinity using path and endpoint approaches. *Biophys J* 2006;90:864–877.
 63. Shivakumar D, Deng Y, Roux B, Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J Chem Theory Comput* 2009;9:919–930.
 64. Lemmon MA, Bu Z, Ladbury JE, Zhou M, Pinchasi D, Lax I, Engelman DM, Schlessinger J, Two EGF molecules contribute additively to stabilization of the EGFR dimer. *EMBO J* 1997;16:281–294.
 65. Gallicchio E, Levy RM, Recent theoretical and computational advances for modeling protein ligand binding affinities. In: C Christov, editor, *Computational chemistry methods in structural biology*, Vol. 85 of *advances in protein chemistry and structural biology*, Academic Press, 2011, pp 27–80.
 66. Yang CY, Sun H, Chen J, Nikolovska-Coleska Z, Wang S, Importance of ligand reorganization free energy in protein–ligand binding–affinity prediction. *J Am Chem Soc* 2009;131:13709–13721.
 67. Kongsted J, Ryde U, An improved method to predict the entropy term with the MM/PBSA approach. *J Comp-Aided Mol Des* 2009;23:63–71.
 68. Fuentes G, Scaltriti M, Baselga J, Verma CS, Synergy between trastuzumab and pertuzumab for human epidermal growth factor 2 (Her2) from colocalization: an in silico based mechanism. *Breast Cancer Res* 2011;13:R54.
 69. Hommel U, Dudgeon TJ, Fallon A, Edwards RM, Campbell ID, Structure–function relationships in human epidermal growth factor studied by site–directed mutagenesis and proton NMR. *Biochemistry* 1991;30:8891–8898.
 70. Liu P, Bouyain S, Eigenbrot C, Leahy DJ, The ErbB4 extracellular region retains a tethered–like conformation in the absence of the tether. *Prot Sci* 2012;21:152–155.
 71. Liu P, Cleveland TE, Bouyain S, Byrne PO, Longo PA, Leahy DJ, A single ligand is sufficient to activate EGFR dimers. *Proc Natl Acad Sci* 2012;109:10861–10866.
 72. Luo C, Xu L, Zheng S, Luo X, Shen J, Jiang H, Liu X, Zhou M, Computational analysis of molecular basis of 1:1 interactions of NRG–1 *beta* wild–type and variants with ErbB3 and ErbB4. *Proteins* 2005;59:742–756.
 73. Tsuda T, Ikeda Y, Taniguchi N, The Asn–420-linked sugar chain in human epidermal growth factor receptor suppresses ligand-independent spontaneous oligomerization: possible role of a specific sugar chain in controllable receptor activation. *J Biol Chem* 2000;275:21988–21994.
 74. Zhen Y, Caprioli RM, Staros JV, Characterization of glycosylation sites of the epidermal growth factor receptor. *Biochemistry* 2003;42:5478–5492.
 75. Loeffler HH, Kitao A, Collective dynamics of periplasmic glutamine binding protein upon domain closure. *Biophys J* 2009;97:2541–2549.
 76. Walker R, Steinbrecher T, McGee D, Miller III B, Swails J, *Amber Advanced Tutorial 3*. 2012.
 77. Lemmon MA, Schlessinger J, Cell signaling by receptor tyrosine kinases. *Cell* 2010;141:1117–1134.
 78. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995;117:5179–5197.
 79. Brice AR, Dominy BN, Analyzing the robustness of the MM/PBSA free energy calculation method: application to DNA conformational transitions. *J Comput Chem* 2011;32:1431–1440.
 80. Genheden S, Ryde U, How to obtain statistically converged MM/GBSA results. *J Comput Chem* 2010;31:837–846.
 81. Genheden S, Ryde U, A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations. *J Comput Chem* 2011;32:187–195.
 82. Swanson JM, Henchman RH, McCammon JA, Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys J* 2004;86:67–74.