

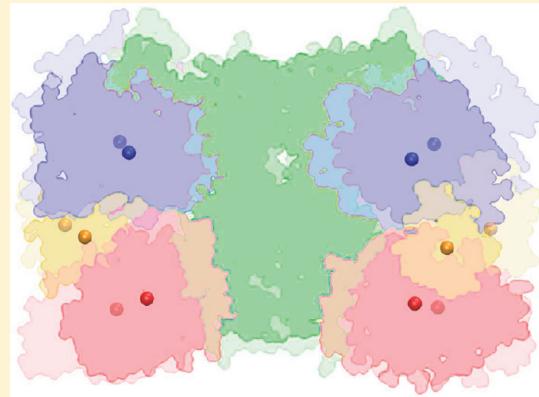
Polymorphism of the Epidermal Growth Factor Receptor Extracellular Ligand Binding Domain: The Dimer Interface Depends on Domain Stabilization

Zhiyong Zhang[†] and Willy Wriggers^{*‡}

The University of Texas School of Biomedical Informatics, 7000 Fannin Street, Houston, Texas 77030, United States

 Supporting Information

ABSTRACT: Epidermal growth factor receptors (EGFRs) and their cytoplasmic tyrosine kinases play important roles in cell proliferation and signaling. The EGFR extracellular domain (sEGFR) forms a dimer upon the binding of ligands, such as epidermal growth factor (EGF) and transforming growth factor α (TGF α). In this study, multiple molecular dynamics (MD) simulations of the 2:2 EGF·sEGFR3–512 dimer and the 2:2 TGF α ·sEGFR3–512 dimer were performed in solvent and crystal environments. The simulations of systems comprising up to half a million atoms reveal part of the structural dynamics of which sEGFR dimers are capable. The solvent simulations consistently exhibited a prominent conformational relaxation from the initial crystal structures on the nanosecond time scale, leading to symmetry breaking and more extensive contacts between the two sEGFR monomers. In the crystal control simulation, this symmetry breaking and compaction was largely suppressed by crystal packing contacts. The simulations also provided evidence that the disordered domain IV of sEGFR may act as a stabilizing spacer in the dimer. Thus, the simulations suggest that the sEGFR dimer can take diverse configurations in solvent environments. These biologically relevant conformations of the EGFR signal transduction network can be controlled by contacts among the structural domains of sEGFR and its ligands.



The epidermal growth factor receptor (EGFR) system has been studied with growing interest for more than 30 years. Signal transduction by the EGFR family is associated with various forms of cancer, which makes it an excellent therapeutic target.^{1,2} The EGFR family of receptor tyrosine kinases (RTK) consists of four members: EGFR (HER1/ErbB1), ErbB2 (Neu/HER2), ErbB3 (HER3), and ErbB4 (HER4).³ They can be activated by a dozen potential ligands.^{4,5}

The human EGFR (1186 residues) is one of the members of this family.⁶ Similar to other RTKs, it comprises five regions: an extracellular ligand binding region (residues 1–620), a transmembrane domain (residues 621–643), an intracellular juxtamembrane region (residues 644–685), a tyrosine kinase domain (residues 686–953), and a C-terminal regulatory region (residues 954–1186).³ The extracellular region consists of four subregions (Figure 1a): I (residues 1–165), II (residues 166–310), III (residues 311–480), and IV (residues 481–620).^{7,8} Domains I and III are the homologous ligand binding domains (37% sequence identity), whereas domains II and IV are homologous cysteine-rich domains.⁹ The human EGFR is activated by EGF family ligand proteins, such as human EGF and transforming growth factor α (TGF α).⁵

EGF induces dimerization of EGFR by binding to the extracellular region of the receptor to form a 2:2 EGF·EGFR complex.^{10,11} This event causes the autophosphorylation of the

cytoplasmic tyrosine kinase, which in turn activates a complex network of downstream signaling pathways that regulate cell proliferation and differentiation.^{12,13} In 2002, two essential structures at the heart of this mechanism were determined^{14,15} that confirmed the so-called receptor-mediated model of dimerization.³ In this model, each ligand can only bind to one EGFR inducing a conformational change that promotes the dimerization.¹¹ Both structures contain the soluble extracellular region of EGFR (sEGFR). One is a 2:2 EGF·sEGFR complex (determined at 3.3 Å resolution), in which the sEGFR comprises residues 3–512.¹⁵ The other is a 2:2 TGF α ·sEGFR complex (determined at 2.5 Å resolution), in which the sEGFR is truncated (residues 1–501).¹⁴ The (EGF or TGF α) ligand binding sites are located between domains I and III (Figure 1a) and are distant from the dimer interface. The dimer contacts are mediated by sEGFR domain II, particularly by a β -hairpin termed the dimerization arm.¹⁶

On the basis of the known experimental structures of sEGFR, it was proposed that sEGFR could exist in a dynamic equilibrium and sample multiple conformational states.¹⁷ An unliganded sEGFR is in a tethered configuration in which the dimerization

Received: August 23, 2010

Revised: January 27, 2011

Published: January 28, 2011

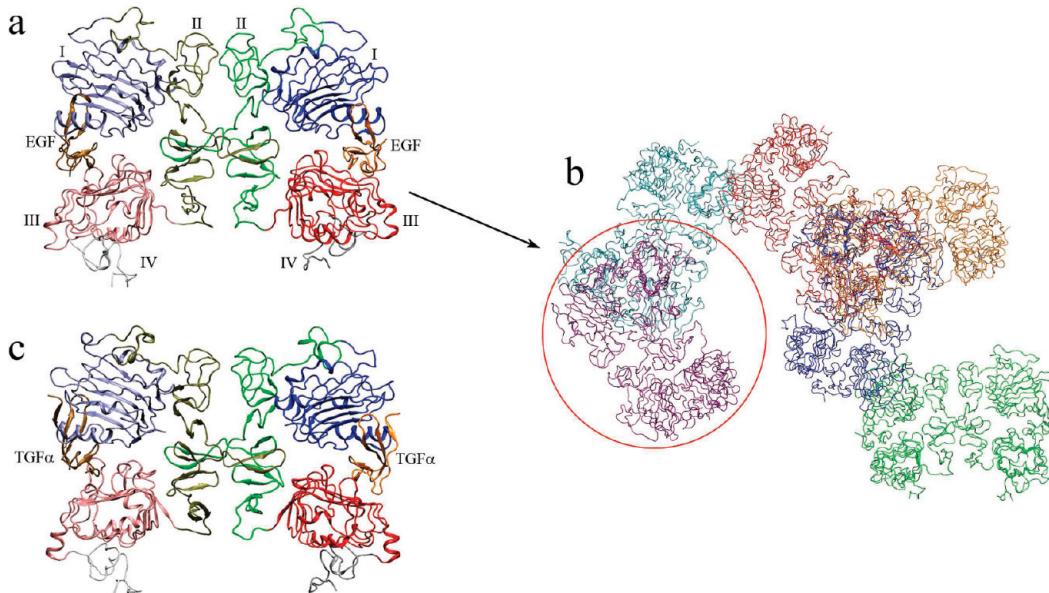


Figure 1. Start structures used for the MD simulations. (a) The 2:2 EGF·sEGFR complex (PDB entry 1IVO). Domains I–IV and the EGF ligands (in different colors) are labeled. (b) The crystal unit cell of the 2:2 EGF·sEGFR complex in space group $P3_121$. (c) The 2:2 TGF α ·sEGFR complex [based on PDB entry 1MOX (see Methods)]. Domains I–IV are colored as in panel a, and the TGF α ligands are labeled. All molecular graphics in this work were created with VMD.³⁸

arm is buried.^{18,19} Binding of the ligand between domains I and III would bring the two domains closer together and expose the dimerization arm in domain II (to create an extended sEGFR).^{20,21} Domain II establishes contacts with domain II in another extended sEGFR, forming the dimer.²² This conformational change from the tethered to the extended sEGFR upon ligand binding would require a large rigid body movement of domains I and II (130° rotation and a 2 nm translation),³ establishing a possible second dimerization contact between the two domain IV forms in the Ferguson model of the sEGFR dimer (Figure 2). Little is known about domain IV in the dimer as it was incomplete or missing in the crystal structures. The structural experiments²¹ did not reveal the atomic details of the mechanism.

Computer simulation techniques, such as molecular dynamics (MD) simulations, can reveal functional dynamics of proteins at the atomic level.^{23,24} In this work, we investigated the structure and dynamics of 2:2 sEGFR·ligand dimers by MD simulations. With the available computational power, a typical MD simulation of a system comprising several hundreds of thousands of atoms could possibly be conducted on a nanosecond or, at best, microsecond time scale.^{25,26} The extensive conformational transition of sEGFR from the tethered to the extended structure cannot be accessed by standard MD, so we did not simulate this transition directly in this work. Instead, we performed multiple MD simulations, starting with the crystal structures of 2:2 EGF·sEGFR¹⁵ and 2:2 TGF α ·sEGFR¹⁴ dimer complexes in different environments. There is a growing body of evidence that supports the notion that the intrinsic dynamics of proteins may play an important role in determining the conformational changes induced by ligands.²⁷ Therefore, it is meaningful to perform limited MD simulations of sEGFR dimers in the ligand-bound state. The observed polymorphism and dynamical modes of motion provide insights into the general structural variability of the system.

This paper is organized as follows. Methods introduces the techniques used in this work, in particular the MD protocols used for 2:2 EGF·sEGFR (in solvent and in crystal) and 2:2 TGF α ·sEGFR (in solvent) complexes. The polymorphism of sEGFR configurations

was explored by principal component analysis. As described in Results and Discussion, our first simulation of the 2:2 EGF·sEGFR complex in solvent revealed a fast relaxation from the initial crystal structure resulting in a significant compaction of the system. The apparent instability of the crystal structure was unexpected, and we subsequently conducted a number of control simulations to improve our understanding of the origin of this instability and to rule out the effects of modeling and of biophysical structure determination. We also report detailed time-resolved changes related to symmetry breaking of the dimer interface. Conclusion summarizes how the results of this investigation provide a consistent and biologically relevant explanation for the observed changes.

METHODS

Solvent Simulations of the 2:2 EGF·sEGFR Complex. The crystal structure of a 2:2 complex of human EGF and sEGFR (PDB entry 1IVO), with a resolution of 3.3 Å, was used as the starting structure.¹⁵ The majority of domain IV was missing because it was not well ordered to give interpretable electron density. In the simulated system (Figure 1a), we included 47 residues (5–51) of the two EGF molecules and 510 residues (3–512) of the two sEGFR molecules (domains I–III and a small part of domain IV).

The standard MD simulation was performed with a parallel implementation of the GROMACS package (version 3.3.1),^{28,29} using a GROMOS-96 force field (43a1) with a united-atom model.³⁰ The proteins (with 79 water molecules at crystallographic sites) were placed in a rectangular box. The minimal distance between the solute and the box boundary was 1.2 nm. The box was then filled with SPC water molecules from an equilibrated cubic box containing 216 water molecules.³¹ The system, protein, and water were energy-minimized using the steepest descent method, until the maximal force on the atoms was less than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. To compensate for the net negative charges on the protein, two Na $^+$ ions were added via replacement of water molecules with the most favorable electrostatic potential. The system (protein, two ions, and

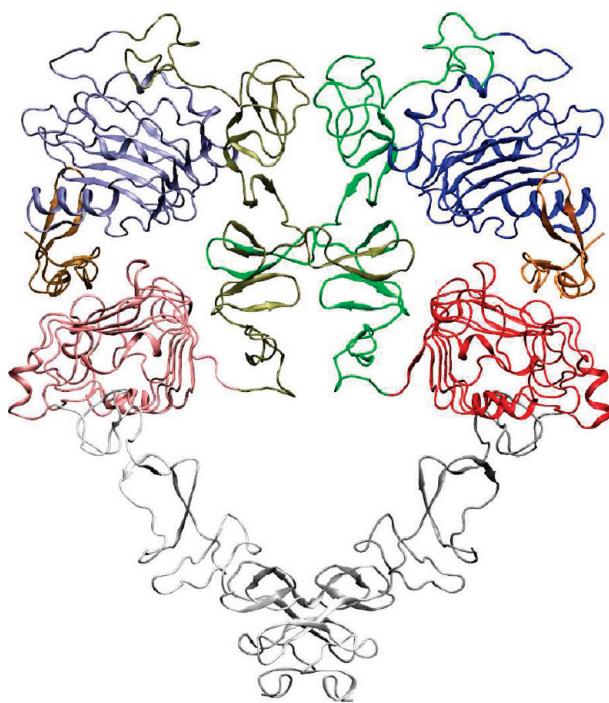


Figure 2. Ferguson model of the sEGFR dimer with a hypothetical domain IV structure. A 130° rotation and a 2 nm translation are required to bring domain I from its position in the tethered structure²⁰ to its proposed location in the dimer.³ Domain IV maintains the same domain III–domain IV relationship as in the tethered structure. Domains I–III are from PDB entry 1IVO and are shown in the same orientation as in Figure 1a,c.

53281 water molecules, with 170909 atoms in total) was energy minimized again using the conjugate gradient algorithm with a force tolerance of 200 kJ mol⁻¹ nm⁻¹, and a steepest descent step was done every 100 steps to make the minimization more efficient.³² After energy minimization, two positional-restraint simulations (force constant of 1000 kJ mol⁻¹ nm⁻²) were performed (100 ps each) with different initial velocities (generated according to a Maxwell distribution at 310 K with different random seeds). Two separate production runs (12 ns each) at 310 K were conducted to sample the conformational space.

The Verlet integration scheme (with a 2 fs time step)³³ and an isothermal–isobaric simulation algorithm³⁴ were used. The three groups (protein, ions, and solvent) were coupled separately to a temperature bath at 310 K, with a relaxation time of 0.1 ps. The pressure was adjusted to 1 bar with a relaxation time of 0.5 ps, and the compressibility was 4.5×10^{-5} bar⁻¹. Covalent bonds in the protein were constrained using the LINCS algorithm.³⁵ A twin-range cutoff was used for the van der Waals interactions: interactions within 0.9 nm were updated every step, while those within 1.4 nm were updated every five steps, together with the pair list. The long-range electrostatic interactions were treated by the PME algorithm,³⁶ with a tolerance of 10^{-5} and an interpolation order of 4.

Crystal Simulation of the 2:2 EGF·sEGFR Complex. The native crystal of the EGF·sEGFR complex belongs to space group $P3_121$, with the following unit cell parameters: $a = b = 22.016$ nm, $c = 11.312$ nm, $\alpha = \beta = 90^\circ$, and $\gamma = 120^\circ$. The unit cell of a $P3_121$ crystal contains six symmetry-related molecules (Figure 1b), which were obtained by applying the $P3_121$ transformation using SwissPDBViewer.³⁷

The simulation was performed with the GROMACS package^{28,29} as described previously, using the same force field as in the solvent simulations. The same parameters as in the solvent simulations were used here, except for those mentioned below. The six EGF·EGFR complexes (including 474 crystal waters) were placed into a triclinic box that was the exact size of the unit cell and hydrated with SPC waters. After a first energy minimization using the steepest descent method, 12 Na⁺ ions were added to compensate for the net negative charges of the system. After the second energy minimization using the conjugate gradient method, a 100 ps MD simulation with positional restraints was performed under a NPT condition. Because the crystals were grown at 293 K, the initial velocities were generated at this temperature. After the 100 ps NPT simulation, the box size was changed back to the original size of the unit cell, and water molecules were added or removed accordingly. The energy minimization described above and the NPT simulation with positional restraints were repeated until the box was equilibrated at its original unit cell size. The final system consists of 447366 atoms in total (six subunits of the 2:2 EGF·sEGFR complex, 12 Na⁺ ions, and 126990 water molecules). A production NVT run at 293 K was performed for 12 ns.

Solvent Simulations of the 2:2 TGF α ·sEGFR Complex. A crystal structure of the 2:2 TGF α ·sEGFR complex, at 2.5 Å resolution (PDB entry 1MOX), was used as the starting structure.¹⁴ There were only two truncated sEGFR molecules (residues 1–501) in the PDB file. For comparison with the EGF·EGFR complex, the first two residues were removed in sEGFR, and residues 502–512 were grafted from the 2:2 EGF·sEGFR complex to be consistent among the simulations. Missing residues were added from the 2:2 EGF·sEGFR complex, and the missing side chain atoms were placed with SwissPDBViewer.³⁷ Therefore, the sEGFR atoms in the simulated 2:2 TGF α ·sEGFR complex were the same as those in the 2:2 EGF·sEGFR complex (residues 3–512); 48 residues (3–50) from the two TGF α molecules were included in the simulation (Figure 1c).

The 2:2 TGF α ·sEGFR complex was simulated only in the solvent environment, and the force field and parameters were the same as those in the solvent simulations of the 2:2 EGF·sEGFR complex. The system was comprised of 160447 atoms in total (protein and 49807 water molecules). Two independent production runs (12 ns each) at 310 K were performed, which was similar to the case for the 2:2 EGF·sEGFR complex.

Principal Component Analysis. All MD trajectory analysis in this work was conducted with the tools in the GROMACS package^{28,29} and with VMD.³⁸ Most of the techniques are self-explanatory, but it is worth mentioning principal component analysis (PCA) in more detail.

PCA is a statistical method that can be used to analyze a MD trajectory to extract large-amplitude collective motions of the protein out of its small, random internal motions.^{39,40} Technically, a covariance matrix of internal atomic fluctuations is built from the MD trajectory after elimination of the overall translational and rotational motion of the protein. The covariance matrix is then diagonalized to yield eigenvectors and eigenvalues. An eigenvector represents a dynamical mode of motion in the protein (PCA mode), and the corresponding eigenvalue is the amplitude of fluctuation along this mode. Generally, only the first few PCA modes with the largest fluctuations dominate the major protein dynamics, defining a so-called essential subspace.⁴¹ The fluctuations in the essential subspace, termed essential dynamics, may describe functionally important protein dynamics.

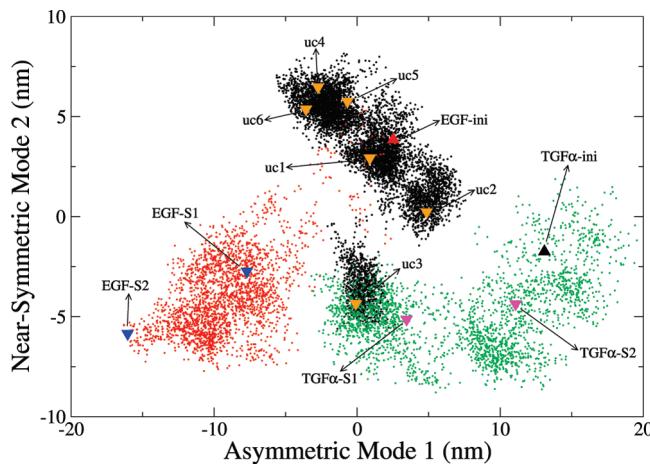


Figure 3. Projections of sEGFR dimer trajectories onto a 2D subspace spanned by the first two PCA modes. The C α atoms of domains I–III of the sEGFR dimer were used to perform PCA on the combined meta-trajectory. The initial structure of the 2:2 EGF·sEGFR complex is marked by a red upward-pointing triangle. The six subunits of the crystal simulation (uc1–uc6) are colored black, and their final structures are marked by orange downward-pointing triangles. Trajectories EGF-S1 and EGF-S2 are colored red, and their final structures are marked by blue downward-pointing triangles. The initial structure of the 2:2 TGF α ·sEGFR complex is marked by a black upward-pointing triangle. Trajectories TGF α -S1 and TGF α -S2 are colored green, and their final structures are marked by magenta downward-pointing triangles.

To illustrate the dynamics observed along individual PCA modes, it is customary to project the MD trajectory along the modes. By projecting the MD trajectory along the first two principal components, one can embed the high-dimensional conformational space linearly within the two-dimensional (2D) subspace of the largest positional fluctuations. This 2D embedding reveals how the system explores its conformational space during the MD simulation. It also reveals how the system populates conformational states and transits between them. If specific modes are sufficiently well-sampled, the one-dimensional (1D) projections along individual modes exhibit relaxation times that are shorter than the simulation length, and the corresponding autocorrelation function decays to (near) zero.^{42,43} Therefore, 1D projections along PCA modes and their autocorrelation function can reveal whether the MD simulation is long enough for sampling the specific modes.

RESULTS AND DISCUSSION

Figure 3 illustrates the 2D embedding of all trajectories afforded by PCA. After the trajectories had been concatenated (uc1–uc6, EGF-S1, and EGF-S2 for the 2:2 EGF·sEGFR complex; TGF α -S1 and TGF α -S2 for the 2:2 TGF α ·sEGFR complex; 120 ns in total), PCA was performed on the combined meta-trajectory, and the PCA modes with the largest amplitudes were investigated. The embedding in a 2D essential subspace in Figure 3 provides a general “road map” of the observed structural variability.

The first two PCA modes with the largest amplitudes contributed \sim 54% fluctuation of the meta-trajectory. All observed and simulated conformations were projected onto this 2D plane (Figure 3). The sampled regions of conformational space are well-separated, and the plot shows that the configurations and different trajectories of the sEGFR dimer span a large range of the

Table 1. Average rmsd Values (in nanometers) of the sEGFR Dimer in Different Simulation Systems

| | EGF-SOL ^a | EGF-CRY ^b | TGF α -SOL ^c |
|--------|--|--|---|
| EGF-S1 | 0.55 ^d (0.03 ^e) | uc1 0.35 ^d (0.01 ^e) | TGF α -S1 0.62 ^d (0.05 ^e) |
| EGF-S2 | 0.65 ^d (0.06 ^e) | uc2 0.39 ^d (0.01 ^e) | TGF α -S2 0.41 ^d (0.02 ^e) |
| | | uc3 0.41 ^d (0.01 ^e) | |
| | | uc4 0.35 ^d (0.01 ^e) | |
| | | uc5 0.36 ^d (0.01 ^e) | |
| | | uc6 0.44 ^d (0.01 ^e) | |

^a Solvent MD simulations of the 2:2 EGF·sEGFR complex. Two independent simulations (EGF-S1 and EGF-S2) were performed. ^b Crystal MD simulation of the 2:2 EGF·sEGFR complex, which has six subunits (termed uc1–uc6) in the unit cell. ^c Solvent MD simulations of the 2:2 TGF α ·sEGFR complex. Two independent simulations (TGF α -S1 and TGF α -S2) were performed. ^d The average rmsd value was calculated from the last 2 ns of each trajectory. ^e Standard deviation. Only domains I–III in the sEGFR dimer were used to compute the rmsd.

Table 2. Average Radii of Gyration (R_g , in nanometers) of the sEGFR Dimer in Different Simulation Systems

| | EGF-SOL | EGF-CRY | TGF α -SOL |
|--------|--|--|---|
| EGF-S1 | 3.43 ^a (0.01 ^b) | uc1 3.50 ^a (0.01 ^b) | TGF α -S1 3.46 ^a (0.02 ^b) |
| EGF-S2 | 3.39 ^a (0.02 ^b) | uc2 3.54 ^a (0.01 ^b) | TGF α -S2 3.51 ^a (0.02 ^b) |
| | | uc3 3.45 ^a (0.01 ^b) | |
| | | uc4 3.59 ^a (0.01 ^b) | |
| | | uc5 3.55 ^a (0.01 ^b) | |
| | | uc6 3.51 ^a (0.01 ^b) | |

^a The average R_g was calculated from the last 2 ns trajectory in each simulation. ^b Standard deviation. Only domains I–III in the sEGFR dimer were used to compute the R_g .

essential subspace. It is worth noting that the sign of modes 1 and 2 was chosen such that a positive direction corresponds to a larger radius of gyration (R_g). In other words, the systems were less compact in the direction of the diagonal. Figure 3 demonstrates that all solvent simulations exhibit a significant compaction of the system (despite the ligand-dependent differences in the initial conformation), whereas crystal subunits (with the exception of uc3) remain close to the initial structure. Figure 3 contains many more details that we will use to explain the following descriptions of the individual trajectories.

Solvent Simulations of the 2:2 EGF·sEGFR Complex. Historically, trajectory EGF-S1 was the first simulation conducted in this work. To test the reproducibility of the result, we performed a second independent simulation, EGF-S2, after the discovery of the significant compaction (Figure 3). The conformational changes were evaluated by root-mean-square deviations (rmsd) from the initial structure. Because of the incomplete and disordered domain IV, only domains I–III (residues 3–480) were used to calculate the rmsd. The two independent simulations, EGF-S1 and EGF-S2, both exhibit a large conformational change with an \sim 0.6 nm rmsd from the initial structure (Table 1). To quantify the compaction, the radius of gyration (R_g) of the sEGFR dimer (domains I–III) was also calculated for each trajectory of the 2:2 EGF·sEGFR complex. The initial structure had an R_g of 3.70 nm, which decreased to 3.43 nm in EGF-S1 and 3.39 nm in EGF-S2 (Table 2). Upon comparison of the initial and final structures of EGF-S2 in panels a and d of Figure 4, respectively, the compaction is very prominent and involves large displacements between 0.34 and 1.09 nm for

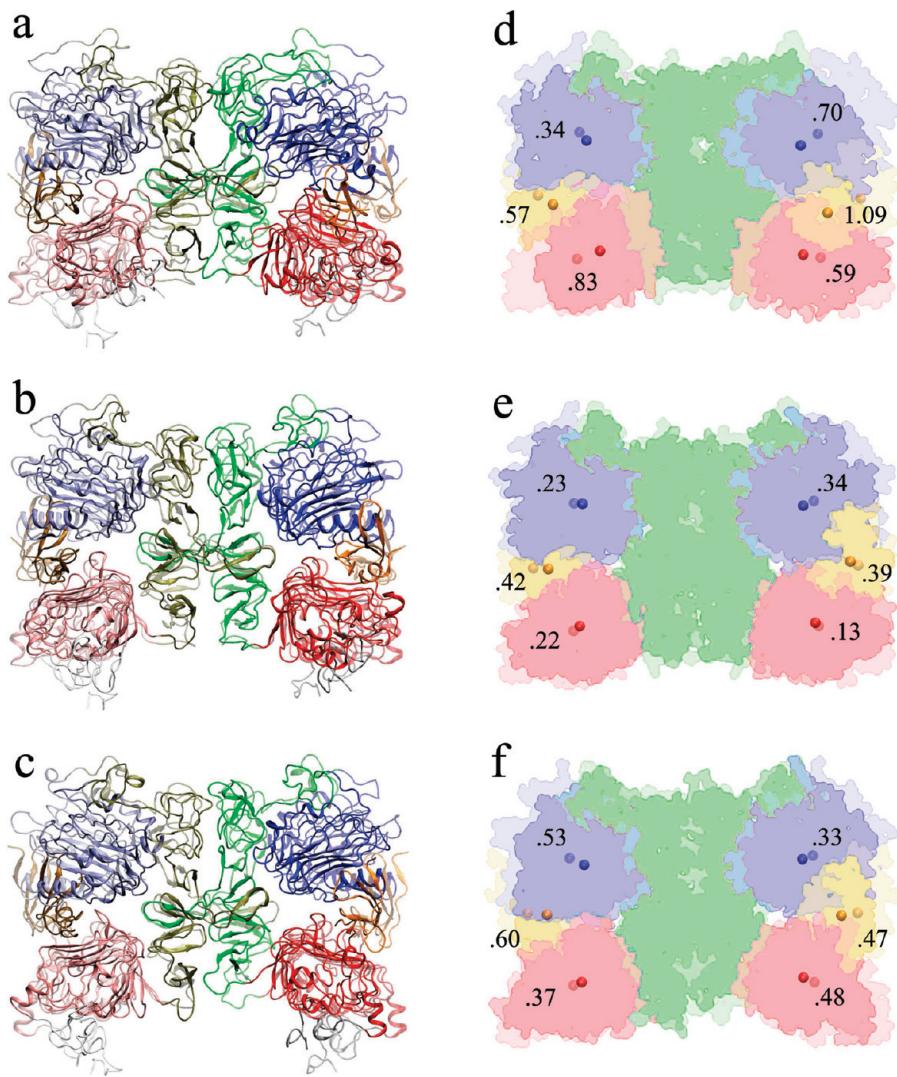


Figure 4. Conformational changes observed in the MD simulations of the 2:2 EGF·sEGFR complex. (a and d) Initial and final structures of trajectory EGF-S2, respectively. (b and e) Initial and final structures of one subunit (uc1) in the crystal simulation, respectively. (c and f) Initial and final structures of trajectory TGF α -S1, respectively. Panels a–c show a ribbon representation of the structures. The domains are colored in the same manner as in Figure 1; the initial structure is transparent, and the final one is solid. Panels d–f show a nonphotorealistic rendering of the molecular surfaces of domains I–III and the ligands. A layer blend shows the initial structure at 34% and the final structure at 66% transparency, and the center of mass locations of domains I and III, and ligands are marked by spheres, with displacements labeled in units of nanometers. The structures are superimposed with domains I–III of the sEGFR dimer.

domains I, III, and the ligand. The surprisingly large conformational changes in the solvent simulations called for additional control simulations to better explore the origin of the observed compaction.

Crystal Simulations of the 2:2 EGF·sEGFR Complex. One possible explanation for the instability of the 2:2 EGF·sEGFR crystal structure was the absence of crystal packing constraints when placed into solvent. To test this hypothesis, we simulated an entire crystallographic unit cell with periodic boundaries reflecting the symmetry space group (see Methods and Figure 1b). Indeed, the PCA embedding (Figure 3) illustrates that (with the exception of subunit uc3) the crystal subunits exhibit behavior markedly different from that of the solvent simulations with EGF ligand. Five of the six subunits remain very close to the initial structure. rmsd analysis (Table 1) confirms that average rmsd values (≈ 0.4 nm) and standard deviations are significantly lower for the crystal subunits than for the solvent simulations. As one would expect in the presence of crystal

packing,^{44–46} the system exhibits both smaller deviations from the start structure and smaller fluctuations. Comparing the initial and final structures of uc1 in Figure 4e, one observes only a very minor compaction of domains I and III (between 0.13 and 0.34 nm displacements), whereas the ligand appears to be more flexible (0.4 nm displacement).

The R_g analysis (Table 2) reveals that crystal packing does not fully restrain the crystal subunits. The R_g values for five of six subunits range from 3.50 to 3.59 nm, except for that of the uc3 outlier, which has a final R_g of 3.45 nm (Table 2). These values occupy an intermediate range between the R_g of the initial structure (3.70 nm) and the R_g of the solvent simulations (3.39 and 3.43 nm). It is interesting to note that the conformational states sampled by the crystal subunits (except uc3) are anticorrelated in Figure 3. Because the R_g is increasing in the diagonal direction, this means that the R_g of five subunits is stabilized by a compensation among the two modes. For example, a mode 1 extension (opening), as in uc2, is

compensated by a mode 2 contraction (compaction). The stabilization of the R_g appears to be an effect of crystal packing, because no other trajectory exhibited such a pronounced anticorrelated pattern (Figure 3).

Although crystal packing effects can explain much of the behavior of the EGF simulations, such as the increased stability in the crystal unit cell, they do not fully explain our observations. The R_g results indicate that the simulated sEGFR dimers in the crystal are systematically more compact than the initial structure (although not as compact as in solvent). This residual compaction cannot be explained by crystal packing alone and is especially pronounced for outlier uc3. We note that domain IV did exist in the experimental 2:2 EGF·sEGFR complex expressed for crystallization. However, it was too disordered to yield full atomic coordinates in the PDB structure and was thus not complete in our system. Mutation analysis and the Ferguson model for the sEGFR dimer including domain IV (Figure 2) suggest that, in addition to domain II, domain IV may contribute to EGFR dimerization.²² Therefore, it is likely that the initial crystal structure of the sEGFR dimer is stabilized by domain IV, and the absence of parts of this domain in our system could explain the residual compaction in the crystal simulation. To test this hypothesis, we have conducted two simulations of the Ferguson model (Figure 2) in solution. These additional control simulations demonstrated that the compaction of the system [which occurs within 2 ns (see below)] can be delayed for 8 ns by the presence of the full domain IV. Eventually, the modeled domain IV forms will slip past each other (because of an observed instability of the speculative domain IV interface in that model), and the system will be compacted. The instability of domain IV forms can be expected because the speculative interface between the domains (derived from the stable tethered conformation) is probably different from that in the untethered dimer (where the domain was too disordered to be determined by crystallography). It is therefore worth noting that we observed a significant delay of the compaction while the speculative domain IV contacts remained intact, in support of the proposed role of domain IV in stabilizing the 2:2 EGF·sEGFR complex crystal structure.

In a true biological system, domain IV is tethered to the membrane by a transmembrane helix. If this helix were included in the Ferguson model, it would stabilize domain IV and somewhat prevent compaction. Unfortunately, we could not address this issue because of the lack of structural information. The extracellular domain by itself is quite important, which is evident from the recent crystal structure published for the *Drosophila* sEGFR⁴⁷ that did not include the transmembrane helices.

In summary, the results of the crystal simulation indicate that the dimer is significantly stabilized (compared to the solvent simulations) by crystal packing effects. In addition, there is indirect evidence that points to domain IV's role as a spacer in stabilizing the 2:2 EGF·sEGFR complex crystal structure.

Solvent Simulations of the 2:2 TGF α ·sEGFR Complex. The resolution of the electron density of the 2:2 EGF·sEGFR complex was 3.3 Å. At such a low resolution, it is possible during crystallographic refinement that artifacts are introduced into the determined structure. To rule out any effect of the relatively low resolution, we conducted two simulations of the 2:2 TGF α ·sEGFR complex structure that was determined at 2.5 Å resolution. TGF α is another member of the EGF family, which has structural and functional characteristics similar to those of EGF. Starting from the crystal structure of the 2:2 TGF α ·sEGFR

complex,¹⁴ we performed two independent MD simulations (termed TGF α -S1 and TGF α -S2) in solvent.

The initial structures of the sEGFR dimer in the 2:2 EGF·sEGFR complex (red upward-pointing triangle in Figure 3) and the 2:2 TGF α ·sEGFR complex (black upward-pointing triangle in Figure 3) are distant on the embedding plane, with a rmsd of 0.44 nm. Several factors may contribute to this significant difference. (1) Domain IV was absent in the specimen expressed for crystallization of the 2:2 TGF α ·sEGFR complex but does exist in the 2:2 EGF·sEGFR complex (although it is not well ordered). Therefore, the 2:2 TGF α ·sEGFR complex may have an initial configuration different from that of the 2:2 EGF·sEGFR complex, because of the absence or presence of domain IV. (2) The crystal of the 2:2 EGF·sEGFR complex belongs to space group $P3_121$ (Figure 1b), whereas that of the 2:2 TGF α ·sEGFR complex belongs to space group $P2_1$. Different space groups may generate different crystallographic contacts that affect the configurations in the crystal unit cell. (3) The sEGFR dimer may take different configurations upon binding of different ligands.

A large compaction from the start structure was observed in the TGF α simulations (Figure 3). The rmsd from the initial structure was 0.62 nm for TGF α -S1 and 0.41 nm for TGF α -S2 (Table 1). The R_g of the initial structure is 3.72 nm, which decreased to 3.46 nm in TGF α -S1 and 3.51 nm in TGF α -S2 (Table 2). Upon comparison of the initial and final structures of TGF α -S1 in Figure 4f, the compaction is very prominent (although not as dramatic as in EGF-S2) and involves large displacements between 0.33 and 0.60 nm for domains I and III and the ligand.

The results demonstrate that the sEGFR dimer also becomes compact in the simulations of the 2:2 TGF α ·sEGFR complex. Because the resolution of the 2:2 TGF α ·sEGFR crystal structure is fairly high (2.5 Å) and all of domain IV was absent in the specimen expressed for crystallization, the large compaction of the solvent simulations of the 2:2 TGF α ·sEGFR complex can be explained only by a relaxation from the crystal packing effects.

Principal Component Analysis of Dominant Motions. We have evidence of two distinct functional motions in our simulations: relaxation from crystal packing and relaxation from domain IV stabilization. Could these two motions be related to the two dominant PCA modes in Figure 3? Our results suggest so. Solvent simulations TGF α -S1 and TGF α -S2 sample a broad region mainly along mode 1 (green dots and magenta downward-pointing triangles in Figure 3). As mentioned above, domain IV is absent in the crystal structure of the 2:2 TGF α ·sEGFR complex, which implies that mode 1 exclusively corresponds to the relaxation due to the loss of the crystallographic contacts. On the other hand, in the crystal simulation of the 2:2 EGF·sEGFR complex, five subunits sample only a limited region of conformational space, except that uc3 escapes from the initial state along mode 2 (Figure 3). Because uc3 is still restrained by crystal packing, the motion along mode 2 appears to originate exclusively in the relaxation because of incomplete domain IV. Consistent with this interpretation is the observation that the initial structure and simulation trajectories of the TGF α dimer are roughly on the same mode 2 elongation level as EGF dimer simulations EGF-S1, EGF-S2, and uc3 after their proposed relaxation from domain IV stabilization. The observation that, without both crystal packing and domain IV constraints, EGF-S1 and EGF-S2 move diagonally to both lower mode 1 and mode 2 elongation is also consistent with this interpretation.

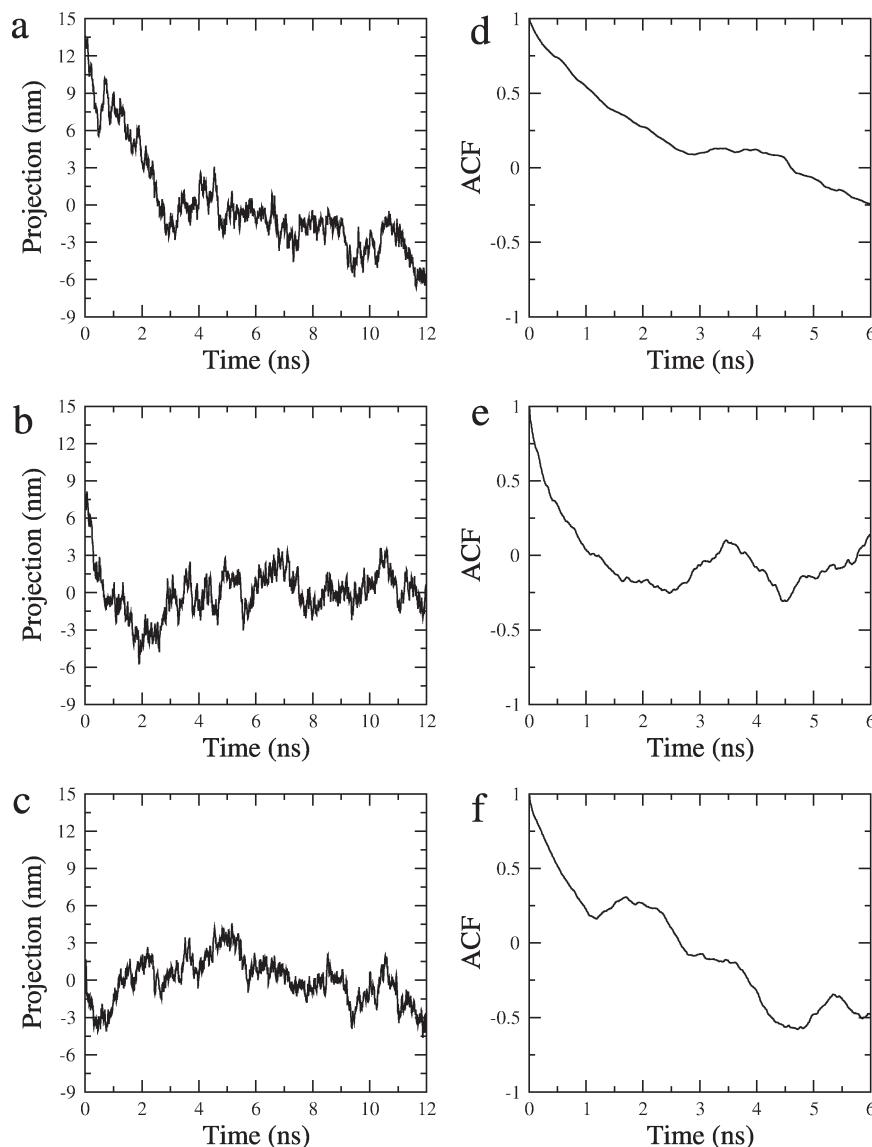


Figure 5. Projections of the EGF-S2 trajectory along its first three PCA modes (a–c) and their autocorrelation functions (d–f): (a and d) first PCA mode, (b and e) second PCA mode, and (c and f) third PCA mode. We note that modes 1 and 2 are not identical but very close to the modes in Figures 3 and 6. Here, PCA was performed on EGF-S2 only to maximize the variability of the projection along the modes (as is customary in PCA). An autocorrelation analysis based on projection along the PCA modes from the combined meta-trajectory (as in Figure 3) gave nearly identical results for modes 1 and 2 (results not shown).

The first simulations described in this work were conducted several years ago when nanosecond time scale MD simulations were state of the art. Over time, the opportunity to perform multiple control simulations and to model larger systems consisting of nearly half a million atoms arose. This expansion in computational complexity (Figure 3) required that the running length of individual trajectories be kept short. To demonstrate a sufficient simulation time and to analyze the relaxation times of the observed collective motions, we performed an autocorrelation analysis of the PCA modes. Figure 5 shows projections of the EGF-S2 trajectory onto the three largest eigenvalue PCA modes and their autocorrelation functions. The results show that the projections along the first two modes exhibit a fast relaxation from the initial conformation, with relaxation times (2–3 ns) that are much shorter than the simulation length of 12 ns. Also, the autocorrelation of the first two modes drops to within

± 0.25 with a 6 ns lag time (for the dynamics to converge, the autocorrelation function should decay to near zero within lag times on the order of the sampling time). The results confirm that the two-dimensional subspace spanned by the first two modes is reasonably well-converged. In contrast, the third mode exhibits no apparent relaxation and fluctuates about the initial value. The autocorrelation of the third mode remains at -0.5 with a 6 ns lag time, indicating insufficient sampling of this mode. This result further justifies our use of two (converged) dimensions for the embedding in Figure 3.

In the following two sections, we will explore the first two PCA modes derived from the meta-trajectory in more detail and relate them to biological function.

Asymmetric Mode 1: Relaxation from Crystal Packing Effects. Figure 6a shows that PCA mode 1 corresponds to an asymmetric motion between the two sEGFR monomers. In the

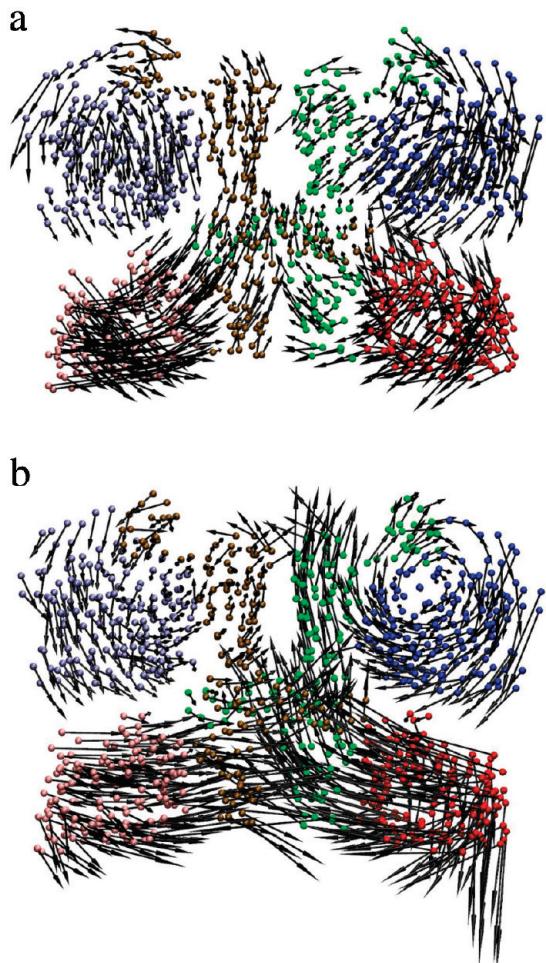


Figure 6. Important relaxation modes of the sEGFR dimer: (a) first (asymmetric) PCA mode and (b) second (near-symmetric) PCA mode. Arrows denote the direction of motion and the maximal amplitude of the displacement. Arrows point from the most positive to the most negative projection observed in the combined trajectory along each mode. Ca atoms of domains I–III are shown for the most positive projection, colored in the same manner as in Figure 1.

left sEGFR monomer (Figure 6a), domain I (ice blue) comes into the plane of the page as domain III (pink) moves out of the plane, and vice versa. The sEGFR monomer on the right performs a similar but antiparallel motion relative to the left monomer. That is to say, its domain I (blue) comes out of the plane of the page while domain III (red) moves into the plane, and vice versa. In the crystal unit cell, the motions of domains I and III are constrained by crystallographic contacts (Figure 1b), which are relaxed in the solvent simulations. This antiparallel motion described by the first PCA mode twists the two sEGFR monomers, and thus, the dimer becomes more compact. The two dimerization arms in the domain II forms (ocher and green) are tightened by this twisting motion, which indicates stronger contacts for sEGFR dimerization.

After completion of this work, the structure of a ligand-induced dimer of *Drosophila* sEGFR (d-sEGFR) that exhibited a surprisingly asymmetric conformation was reported.⁴⁷ The timely discovery of this asymmetric structure allowed us to compare the MD simulation results of (human) sEGFR symmetry breaking to those observed in an actual asymmetric crystal structure. We chose trajectory TGF α -S1 as a model for the

symmetry breaking because of its dominant movement along the asymmetric mode 1 (Figure 3). In Figure 7, we compare the motions observed in TGF α -S1 with the new asymmetric d-sEGFR structure. There is a striking similarity in the overall domain arrangement and in two of the three dimer interface contact points (Figure 7b,d). Moreover, the time evolution of the conformation provides information about the sequence of symmetry breaking events. The rmsd between TGF α -S1 and d-sEGFR (Figure 7c) exhibits a funnel-like pattern in which the structure quickly relaxes from an initial 0.48 nm rmsd to 0.28 nm within the first 1.5 ns of simulation time. Detailed inspection of the trajectory reveals that contact point I breaks first, after 0.3 ns. This contact point involves interactions between Q194 and P204 [corresponding to Q189 and P200 in d-sEGFR (see Figure 5 in ref 47)]. Subsequently, after 0.5 ns, the central dimerization arms move out of register. This motion is due to the breaking of a stabilizing contact involving D279 and H280 (not shown), two residues that were found to be important for dimerization by mutation analysis.²² We note that the misalignment ranging from contact point I to the dimerization arm appears quite similar between TGF α -S1 and d-sEGFR. However, an additional asymmetric contact point formed after 1.4 ns (labeled II in Figure 7 and comprised by E306 and K301) that was observed only in the simulation. It is apparent from the rmsd values (Figure 7c) that the simulation first approached d-sEGFR, but after 1.5 ns, the simulated system relaxes even further, establishing an asymmetric contact point II (which remains symmetric in d-sEGFR).

Nearly Symmetric Mode 2: Relaxation from Domain IV Stabilization. The motion described by PCA mode 2 is nearly symmetric, and it essentially takes place within the plane of the page (Figure 6b). As discussed above, this mode appears to describe a relaxation caused by the absence (or incompleteness) of domain IV that acts as a spacer. The domain III forms are becoming closer, and the dimerization interface (domain II) can interact more tightly than in the initial structure. In other words, the loss of putative dimerization contacts in domain IV (Figure 2) is compensated by an increase in the number of contacts between the two domain II forms. Among our trajectories, uc3 best mimics this particular mode (Figure 3). Because uc3 is still restrained by crystal packing, it does not exhibit the pronounced symmetry breaking of mode 1. The motion along mode 2 in uc3 appears to originate exclusively in the relaxation because of the incomplete domain IV.

It has been reported that deletion of domain IV has an only minor effect on dimerization strength²² despite of its obvious importance (e.g., for ligand binding). We propose that domain IV may play a dual role in EGFR dimerization. Contacts between the two domain IV forms contribute to the dimerization. On the other hand, domain IV may act as a stabilizing spacer that can buttress domain III. The two domain III forms both move inward without the support of the domain IV forms and further squeeze the two domain II forms together (Figure 6b), which seems to strengthen dimerization. This dual role of domain IV may explain why its deletion has little effect on dimerization,²² because contacts are merely shifted from domain IV to domain II. In mode 2, the ligand-binding sites between domains I and III are closed as well (Figure 6b), which may enhance ligand binding. The result could help rationalize how deletion of domain IV increases ligand binding affinity 24-fold,²² but the detailed mechanism is possibly more complex³ because it might also involve changes in the tethered conformation that is believed to be in a dynamic equilibrium with the untethered (dimer) conformation.

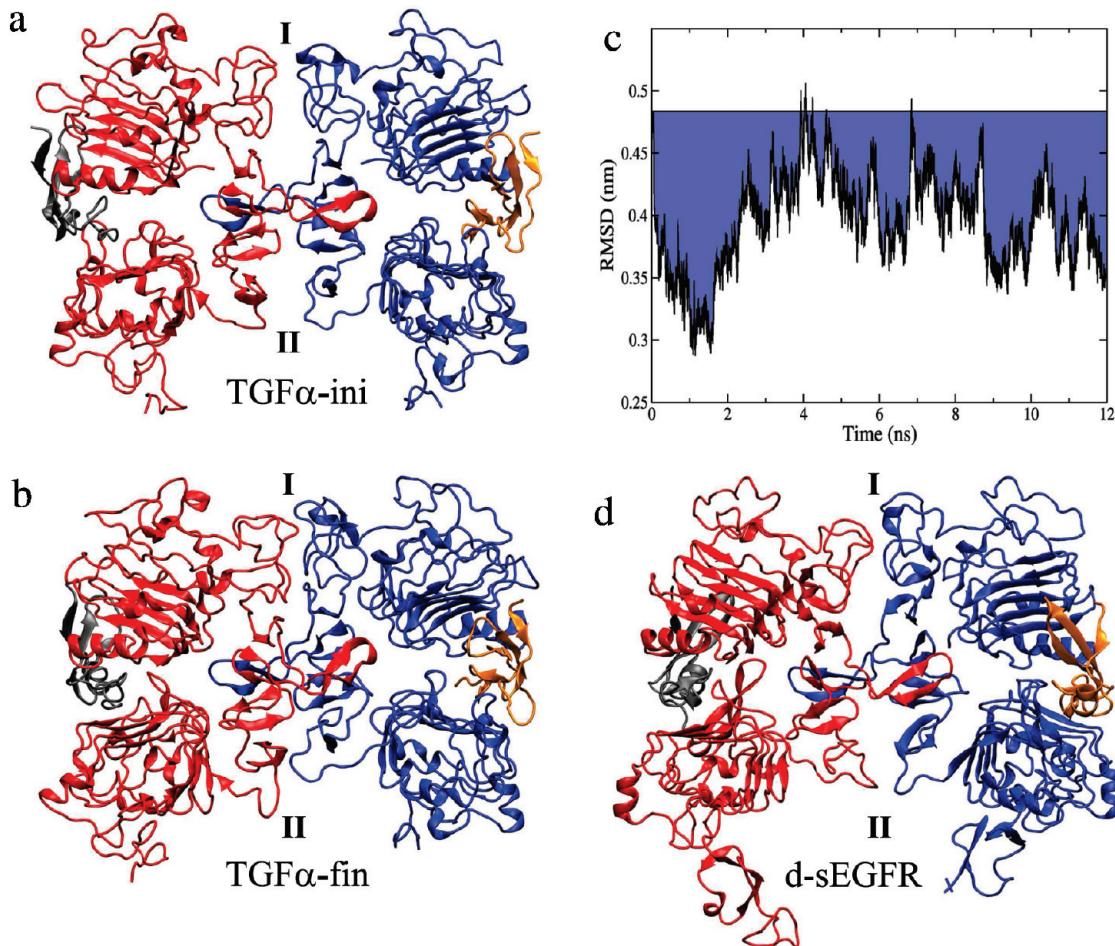


Figure 7. Trajectory TGF α -S1 as a model for symmetry breaking. (a and b) Initial (TGF α -ini) and final (TGF α -fin) structures of TGF α -S1, respectively. (c) rmsd of C α atoms of TGF α -S1 from corresponding C α atoms of the d-sEGFR structure (820 of 956 C α atoms of domains I–III were matched using the “Multiseq” program of VMD³⁸); the blue drop from the baseline emphasizes the difference from the initial rmsd. (d) d-sEGFR structure.⁴⁷ There are three symmetry-breaking contacts in the dimer interface: contact I (near domain I), the central dimerization arm (not labeled), and contact II (see the text for details).

Structural Characterization of sEGFR Dimer Compaction. The collective motions observed in our simulations have major implications for the dimerization interface of sEGFR. It is worth noting that most trajectories move along both PCA modes 1 and 2. In Figure 7, we show the symmetry-breaking motions observed in trajectory TGF α -S1 that approximately mimics mode 1, but even greater relaxation motions occurred in EGF-S2 that moves along both dominant PCA modes. In this section, we describe functionally relevant details of the compaction resulting from the combined effect of modes 1 and 2 in the EGF-S2 trajectory.

Figure 8a shows the time evolution of secondary structures in the EGF-S2 trajectory. The number of residues that belong to any kind of secondary structure, α -helix, or β -sheet is fairly stable during the simulation (Figure 8a), suggesting that the major relaxation in solvent is not caused by the forming or breaking of any secondary structure. rmsd values of different domains in one sEGFR monomer for EGF-S2 were computed to determine which domain contributes most to the structural relaxation (Figure 8b). The rmsd values of domain I (blue in Figure 8b, ~ 0.15 nm on average) and domain III (red in Figure 8b, ~ 0.2 nm on average) are small, whereas domain II (green in Figure 8b) exhibits a rmsd of > 0.3 nm. Interestingly, the rmsd values of the sEGFR monomer, excluding domain IV (black in

Figure 8b), are correlated with the rmsd of domain II. These results indicate that domains I and III act like two rigid bodies linked by domain II (see also Figure 4d). The relaxation in the sEGFR dimer in EGF-S2 (which involves both PCA modes 1 and 2) originates in a collective motion between domains I and III, in which they approach each other in each sEGFR monomer (Figure 4d) and push its domain II (green) toward the other monomer. Thus, the dimerization arms between the two sEGFR monomers intertwine more extensively, and the sEGFR dimer becomes compact.

Structural data¹⁵ and mutational analysis²² have shown that domain III may buttress the C-terminus of domain II for dimerization. The side chain of N274 in module 6 of domain II forms hydrogen bonds with many residues in domain III (Figure 9). The number of these hydrogen bonds increases in the EGF-S2 simulation (Figure 10a). Also, the side chain of E293 in module 7 of domain II forms a salt bridge with the side chain of R405 of domain III (Figure 9). The distance between the side chains of E293 and R405 decreases in EGF-S2 (Figure 10b). These results indicate that the buttressing interactions between domains II and III are strengthened in the simulation. In EGF-S2, without domain IV acting as a stabilizing spacer, the two domain III forms both move toward the center of the dimer interface as

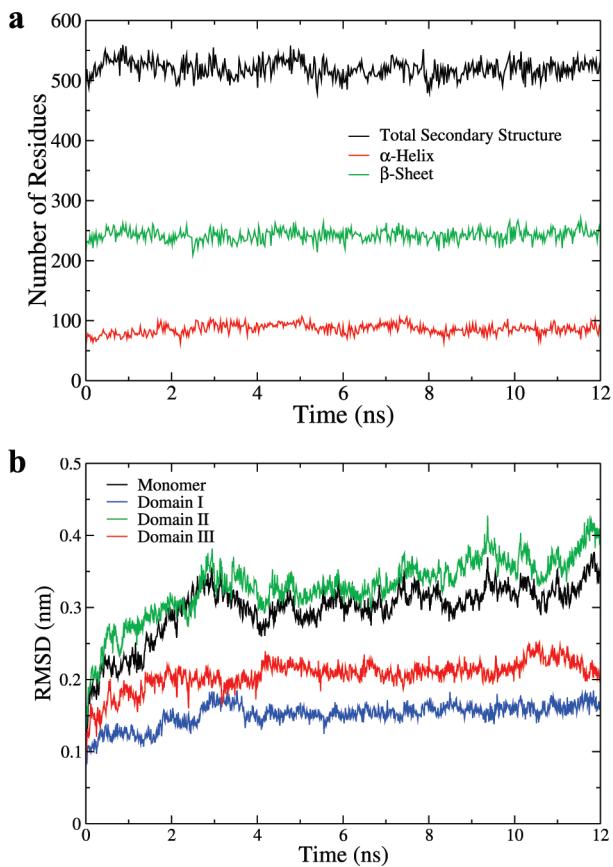


Figure 8. Structural characteristics of trajectory EGF-S2. (a) Time evolution of secondary structure. The black plot shows the number of residues that belong to any kind of secondary structure as detected by DSSP;⁵⁸ the red plot shows the number of residues that belong to α -helices, and the green plot shows the number of residues that belong to β -sheets. (b) rmsd values from the initial structure plotted for one sEGFR monomer and its domains: black for the sEGFR monomer without domain IV, blue for domain I, green for domain II, and red for domain III.

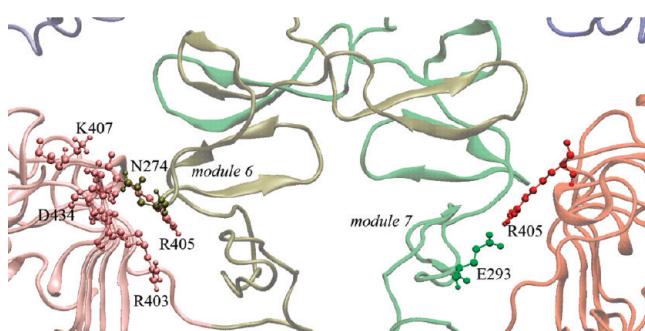


Figure 9. Buttressing interactions between domains II and III in the 2:2 EGF·sEGFR complex (PDB entry 1IVO). Hydrogen bonds between N274 in domain II and residues in domain III are shown in the left sEGFR monomer, and a salt bridge between E293 and R405 is shown in the right sEGFR monomer.

described by the second PCA mode (Figure 6b). Through those strengthened buttressing interactions between domains II and III (Figures 9 and 10), the two domain II forms (especially modules 6 and 7) are effectively pushed farther into dimer interface so that they may promote sEGFR dimerization.

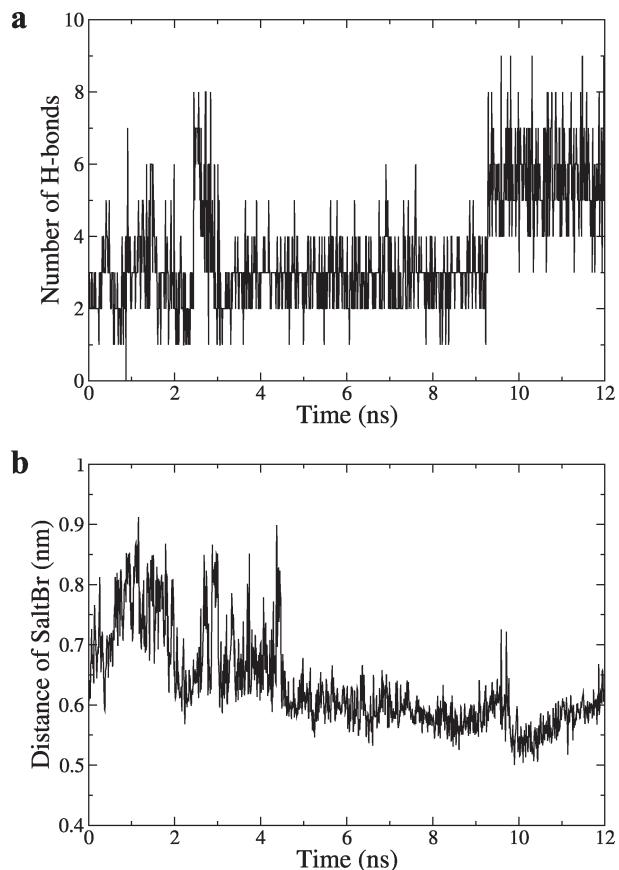


Figure 10. Time evolution of buttressing interactions between domains II and III in trajectory EGF-S2. (a) Number of hydrogen bonds between N274 in domain II and residues in domain III. A hydrogen bond is defined by a hydrogen acceptor distance of <0.35 nm. (b) Length of the salt bridge between E293 and R405 (measured as the distance between the centers of mass of the side chains of the two residues).

CONCLUSIONS

In this work, multiple MD simulations of the 2:2 EGF·sEGFR and 2:2 TGF α ·sEGFR complexes were performed in different (solvent or crystal) environments. The goal of the simulations was to explore the configurations of the sEGFR dimer in the presence of EGF family ligands and to investigate its dynamical modes of motion and relate them to the dimerization of sEGFR. We observed a significant (≈ 0.5 nm rmsd) structural relaxation in solvent that strengthens the dimerization contacts of domain II, especially in the two dimerization arms between the two sEGFR monomers. The observed structural instability and large-scale conformational change are quite exceptional for nanosecond time scale simulations. The compaction in solvent can be attributed to the loss of crystal contacts (both EGF and TGF α ligands) and to a relaxation from stabilization by domain IV (EGF ligand). There is consistent evidence that the two relaxation modes are identical to the first two PCA modes that dominate the fluctuations of the sEGFR dimer in the combined meta-trajectory of all simulations (Figure 6). The first (asymmetric) PCA mode is due to the relaxation from the crystal packing, and the second (near-symmetric) mode reveals a possible stabilizing role of domain IV (which was incomplete in our simulations).

The extent of the relaxation of sEGFR dimers in solvent, for both EGF and TGF α ligands, appeared to be quite reasonable when compared to experimental values and interpreted by

function. First, the relaxation along mode 1 was on the order of the conformational difference between the initial crystal structures, with any remaining differences between EGF and TGF α simulations possibly attributable to the type of ligand. We note that such a significant compaction in solvent after relaxation from crystal packing is not unusual. A well-known historic example of this behavior is the structure of calmodulin, which was first determined in an elongated conformation with X-ray crystallography⁴⁸ but shown to compact in solution by small-angle X-ray scattering (SAXS) experiments,⁴⁹ a result that remained controversial until the compaction was eventually confirmed by NMR.⁵⁰ Second, we found the relaxed EGF-bound structures (EGF-S1, EGF-S2, and crystal subunit uc3) at a level of mode 2 similar to the TGF α simulations (which were based on a crystal structure that had domain IV missing), consistent with our interpretation of mode 2 relating to the absence of domain IV stabilization. Third, we observed a striking similarity between symmetry breaking motions in our asymmetric trajectory TGF α -S1 and conformational differences in the novel d-sEGFR structure.⁴⁷ The asymmetric relaxation in d-sEGFR has a completely different biophysical origin and thereby provides an independent validation of our interpretation. Interestingly, our simulated system first approaches d-sEGFR, but after 1.5 ns, the system relaxes even further, breaking the remaining symmetric dimer interface in d-sEGFR. Although we do not expect a complete analogy between our asymmetric simulation of human sEGFR and *Drosophila* sEGFR, one possible explanation for this discrepancy is that the asymmetric contact point II (Figure 7) is stabilized by residual crystal packing effects in the d-sEGFR structure that are free to relax in the simulation.

Because of the necessary assumptions of the models and the simulation protocols, we point out possible limitations of our work. First, we were unable to address the issue of glycosylation. The extracellular region of EGFR is \sim 20% carbohydrate by mass, but there are very few (10 and 14) solvent-exposed sugars present in the PDB structures (1IVO and 1MOX). Without complete atomic information, we did not include any residual sugars in the models studied here. These studies demonstrate mainly the water-based polymorphism of which the structures are capable; we cannot rule out the possibility that oligosaccharide-mediated interactions could help stabilize the crystal packing and the conformation of the dimer in the extracellular region. Second, the importance of domain IV for ligand binding arises here from its proposed role in domain II–domain III buttressing in the untethered dimer structure (observed in the nearly symmetric mode 2). It is believed that effects of deletion of domain IV on the tethered structure (in a dynamic equilibrium with untethered states) will have a significant influence on ligand binding affinity.³ Our observations of compaction in the untethered dimer alone can play only a part in rationalizing how deletion of domain IV alters ligand binding.

In summary, our simulations predict that the sEGFR dimer in solution, in the absence of domain IV, is significantly more compact than the known crystal structures. Our predictions are consistent with the asymmetric crystal structure of d-sEGFR and with a number of experiments. The results could inspire future experimental research and be further verified experimentally by structural probes and techniques such as FRET and SAXS. Unfortunately, the published SAXS experiments included the stabilizing domain IV, and also they were focused mainly on the larger transition from the tethered to the extended conformation.²¹ It is hoped that the experimental sensitivities will soon be sufficient for confirmation of our computational predictions.

Some aspects of this work could also inspire future computational research. The EGF and TGF α systems did not fully converge to the same end conformation. This could be due to the short simulation time (undersampling), so we did not discuss any ligand-specific differences in structural dynamics. Longer simulations are required to study why certain antibodies can selectively inhibit binding by one of these ligands but not the other. One could also investigate the dimerization affinity between sEGFR upon binding of different ligands by binding free energy calculations, some of which are reasonably efficient for a system of this size.^{51–53} Finally, the conformational transition from the tethered to the extended sEGFR may be studied directly by taking advantage of targeted or enhanced sampling techniques in MD.^{54–57}

■ ASSOCIATED CONTENT

S Supporting Information. Final structures of the simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: (212) 403-8131. Fax: (646) 873-2131. E-mail: wriggers@biomachina.org.

Present Addresses

[†]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China Hefei, Anhui 230026, People's Republic of China.

[‡]D. E. Shaw Research, 39th Floor, 120 W. 45th St., New York, NY 10036.

Funding Sources

This work was supported in part by grants from the National Institutes of Health (1R01GM62968), the Human Frontier Science Program (RGP0026/2003), and the Alfred P. Sloan Foundation (BR-4297).

■ ACKNOWLEDGMENT

We are grateful to Stefan Birnmanns for providing continuing computational resources and support at the University of Texas School of Biomedical Informatics after both authors moved to new positions. We also thank Mili Shah (Rice University, Houston, TX) for helpful discussions in the early stages of this work.

■ ABBREVIATIONS

EGF, epidermal growth factor; EGFR, EGF receptor; sEGFR, EGFR extracellular domain; d-sEGFR, dimer of *Drosophila* sEGFR; TGF α , transforming growth factor α ; MD, molecular dynamics; RTK, receptor tyrosine kinase; PDB, Protein Data Bank; NPT, isothermal–isobaric ensemble; NVT, canonical ensemble; PCA, principal component analysis; rmsd, root-mean-square deviation; R_g , radius of gyration; NMR, nuclear magnetic resonance; FRET, Förster resonance energy transfer; SAXS, small-angle X-ray scattering.

■ REFERENCES

- (1) Yarden, Y. (2001) The EGFR family and its ligands in human cancer. Signalling mechanisms and therapeutic opportunities. *Eur. J. Cancer* 37 (Suppl. 4), S3–S8.

- (2) Grandal, M. V., and Madshus, I. H. (2008) Epidermal growth factor receptor and cancer: Control of oncogenic signalling by endocytosis. *J. Cell. Mol. Med.* 12, 1527–1534.
- (3) Ferguson, K. M. (2008) Structure-based view of epidermal growth factor receptor regulation. *Annu. Rev. Biophys.* 37, 353–373.
- (4) Yarden, Y., and Sliwkowski, M. X. (2001) Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* 2 (2), 127–137.
- (5) Harris, R. C., Chung, E., and Coffey, R. J. (2003) EGF receptor ligands. *Exp. Cell Res.* 284 (1), 2–13.
- (6) Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D., and Seuberg, P. H. (1984) Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 309, 418–425.
- (7) Bajaj, M., Waterfield, M. D., Schlessinger, J., Taylor, W. R., and Blundell, T. (1987) On the tertiary structure of the extracellular domains of the epidermal growth factor and insulin receptors. *Biochim. Biophys. Acta* 916 (2), 220–226.
- (8) Lax, I., Johnson, A., Howk, R., Sap, J., Bellot, F., Winkler, M., Ullrich, A., Vennstrom, B., Schlessinger, J., and Givol, D. (1988) Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor α . *Mol. Cell. Biol.* 8 (5), 1970–1978.
- (9) Ward, C. W., Hoyne, P. A., and Flegg, R. H. (1995) Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins: Struct., Funct., Genet.* 22, 141–153.
- (10) Brown, P. M., Debanne, M. T., Grothe, S., Bergsma, D., Caron, M., Kay, C., and O'Connor-McCourt, M. D. (1994) The extracellular domain of the epidermal growth factor receptor. Studies on the affinity and stoichiometry of binding, receptor dimerization and a binding-domain mutant. *Eur. J. Biochem.* 225, 223–233.
- (11) Lemmon, M. A., Bu, Z., Ladbury, J. E., Zhou, M., Pinchasi, D., Lax, I., Engelman, D. M., and Schlessinger, J. (1997) Two EGF molecules contribute additively to stabilization of the EGFR dimer. *EMBO J.* 16, 281–294.
- (12) Carpenter, G., and Cohen, S. (1990) Epidermal growth factor. *J. Biol. Chem.* 265, 7709–7712.
- (13) Ullrich, A., and Schlessinger, J. (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* 61, 203–212.
- (14) Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H. J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α . *Cell* 110, 763–773.
- (15) Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J. H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110, 775–787.
- (16) Schlessinger, J. (2002) Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* 110 (6), 669–672.
- (17) Burgess, A. W., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W., and Yokoyama, S. (2003) An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* 12 (3), 541–552.
- (18) Mattoon, D., Klein, P., Lemmon, M. A., Lax, I., and Schlessinger, J. (2004) The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc. Natl. Acad. Sci. U.S.A.* 101 (4), 923–928.
- (19) Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M., and Leahy, D. J. (2005) The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc. Natl. Acad. Sci. U.S.A.* 102 (42), 15024–15029.
- (20) Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H. S., Leahy, D. J., and Lemmon, M. A. (2003) EGF activates its receptor by removing interactions that auto-inhibit ectodomain dimerization. *Mol. Cell* 11, 507–517.
- (21) Dawson, J. P., Bu, Z., and Lemmon, M. A. (2007) Ligand-induced structural transitions in ErbB receptor extracellular domains. *Structure* 15 (8), 942–954.
- (22) Dawson, J. P., Berger, M. B., Lin, C. C., Schlessinger, J., Lemmon, M. A., and Ferguson, K. M. (2005) Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol. Cell. Biol.* 25 (17), 7734–7742.
- (23) Brooks, C. L., III, Karplus, M., and Pettitt, B. M. (1988) Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics. *Advances in Chemical Physics*, Vol. LXXI, John Wiley & Sons, New York.
- (24) Karplus, M. (1990) Molecular dynamics: Applications to proteins. In *Modelling of Molecular Structures and Properties* (Rivail, J.-L., Ed.) Studies of Physical and Theoretical Chemistry Series, Vol. 71, pp 427–461, Elsevier Science Publishers, Amsterdam.
- (25) Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652.
- (26) Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615.
- (27) Bakan, A., and Bahar, I. (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. U.S.A.* 106 (34), 14349–14354.
- (28) Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) Gromacs: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56.
- (29) Lindahl, E., Hess, B., and van der Spoel, D. (2001) Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7, 306–317.
- (30) van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular simulation: The GROMOS96 manual and user guide*, BIOMOS b. v., Groningen, The Netherlands.
- (31) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for water in relation to protein hydration. In *Intermolecular Forces* (Pullman, B., Ed.) pp 331–342, Reidel, Dordrecht, The Netherlands.
- (32) van der Spoel, D., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Hess, B., Feenstra, K. A., Lindahl, E., van Drunen, R., and Berendsen, H. J. C. (2001) *Gromacs User Manual*, version 3.1, University of Groningen, Groningen, The Netherlands.
- (33) Verlet, L. (1967) Computer ‘experiments’ on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159, 98–103.
- (34) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690.
- (35) Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472.
- (36) Essman, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8592.
- (37) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723.
- (38) Humphrey, W. F., Dalke, A., and Schulten, K. (1996) VMD: VisualMolecularDynamics. *J. Mol. Graphics* 14, 33–38.
- (39) Kitao, A., and Go, N. (1999) Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9, 164–169.
- (40) Berendsen, H. J. C., and Hayward, S. (2000) Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10, 165–169.
- (41) Amadei, A., Linnse, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins. *Proteins: Struct., Funct., Genet.* 17, 412–425.
- (42) Balsera, M. A., Wriggers, W., Oono, Y., and Schulten, K. (1996) Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100 (7), 2567–2572.

- (43) Zhang, Z., and Wriggers, W. (2006) Local feature analysis: A statistical theory for reproducible essential dynamics of large macromolecules. *Proteins: Struct., Funct., Bioinf.* 64, 391–403.
- (44) Rapp, C. S., and Pollack, R. M. (2005) Crystal packing effects on protein loops. *Proteins: Struct., Funct., Bioinf.* 60 (1), 103–109.
- (45) Anselmi, M., Brunori, M., Vallone, B., and Di Nola, A. (2008) Molecular dynamics simulations of the neuroglobin crystal: Comparison with the simulation in solution. *Biophys. J.* 95, 4157–4162.
- (46) Kondrashov, D. A., Zhang, W., Aranda, R., IV, Stec, B., and Phillips, G. N., Jr. (2008) Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. *Proteins: Struct., Funct., Bioinf.* 70, 353–362.
- (47) Alvarado, D., Klein, D. E., and Lemmon, M. A. (2010) Structural basis for negative cooperativity growth factor binding to an EGF receptor. *Cell* 142, 568–579.
- (48) Babu, Y. S., Bugg, C. E., and Cook, W. J. (1988) Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* 204, 191–204.
- (49) Heidorn, D. B., and Trewella, J. (1988) Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry* 27, 909–915.
- (50) Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: The central helix is flexible. *Biochemistry* 31, 5269–5278.
- (51) Kollman, P. A., 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, D., and Cheatham, T. E., III (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897.
- (52) Zoete, V., Meuwly, M., and Karplus, M. (2005) Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins: Struct., Funct., Bioinf.* 61 (1), 79–93.
- (53) Wang, J., Deng, Y., and Roux, B. (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* 91 (8), 2798–2814.
- (54) Schlitter, J., Engels, M., Krüger, P., Jacoby, E., and Wollmer, A. (1993) Targeted molecular dynamics simulation of conformational change: Application to the T \leftrightarrow R transition in insulin. *Mol. Simul.* 10 (2–6), 291–308.
- (55) Zhang, Z., Shi, Y., and Liu, H. (2003) Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* 84, 3583–3593.
- (56) He, J., Zhang, Z., Shi, Y., and Liu, H. (2003) Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *J. Chem. Phys.* 119, 4005–4017.
- (57) Zhang, Z., Boyle, P. C., Lu, B.-Y., Chang, J.-Y., and Wriggers, W. (2006) Entropic folding pathway of human epidermal growth factor explored by disulfide scrambling and amplified collective motion simulations. *Biochemistry* 45, 15269–15278.
- (58) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–637.