Crystal Structure of Human Epidermal Growth Factor and Its Dimerization*

Received for publication, April 2, 2001, and in revised form, June 27, 2001
Published, JBC Papers in Press, July 3, 2001, DOI 10.1074/jbc.M102874200

He-Shu Lu‡, Ji-Jie Chai‡, Ming Li‡, Bing-Ren Huang§, Cun-Heng He§, and Ru-Chang Bi‡¶

From the ‡Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China and the ¶Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100055, China

Epidermal growth factor (EGF) is a typical growth-stimulating peptide and functions by binding to specific cell-surface receptors and inducing dimerization of the receptors. Little is known about the molecular mechanism of EGF-induced dimerization of EGF receptors. The crystal structure of human EGF has been determined at pH 8.1. There are two human EGF molecules A and B in the asymmetric unit of the crystals, which form a potential dimer. Importantly, a number of residues known to be indispensable for EGF binding to its receptor are involved in the interface between the two EGF molecules, suggesting a crucial role of EGF dimerization in the EGF-induced dimerization of receptors. In addition, the crystal structure of EGF shares the main features of the NMR structure of mouse EGF determined at pH 2.0, but structural comparisons between different models have revealed new detailed features and properties of the EGF structure.

Human epidermal growth factor (hEGF) is a polypeptide of 53 amino acids with three internal disulfide bridges. As a mitogen, it first binds with high affinity to specific cell-surface receptors and then induces their dimerization, which is essential for activating the tyrosine kinase in the receptor cytoplasmic domain, initiating a signal transduction that results in DNA synthesis and cell proliferation (1, 2). Although EGF is a typical growth-stimulating peptide, little is known about the molecular mechanism of EGF-induced receptor dimerization. EGF was found to exist predominantly as a monomeric species in solution, and based on analyses of binding of EGF to the extracellular domain of its receptor and of the resulting dimerization of the receptor, some models have been proposed for EGF-induced dimerization of receptors (3). However, all these models have yet to be verified by structural studies, the most important of which is the structural determination of the complex of EGF with its receptor. Considerable attention has been paid to the structural elucidation of EGF for clarification of the structure/function relationship. Although the solution structure of EGF has been determined using NMR methods by several groups (4, 5, 6, 7, 8, 9), most of these NMR studies were performed at very acidic pH, where EGF completely loses its binding activity and biological potency (10). On the other hand, it has proved to be very difficult to grow good quality EGF crystals, despite the publication of a few relevant crystallization notes (11, 12). After numerous experiments to refine crystallization conditions, using a C-terminally truncated hEGF variant, we have obtained better quality EGF crystals in two different crystal forms at near physiological pH (13). Here we report the crystal structure of hEGF determined by the multiple isomorphous replacement method. There are two hEGF molecules in the asymmetric unit of the crystals that are in close end-to-end contact with each other. Analyses of the crystal structure and comparisons with NMR solution structures have revealed new details of the features and properties of the hEGF structure.

EXPERIMENTAL PROCEDURES

Materials—The C-terminally truncated hEGF was prepared as described (14). The hEGF gene encoding 51 amino acids was chemically synthesized based on preferred code usage in yeast. The gene was under the control of the alcohol oxidase promoter and the α-factor leader sequence including the 85-amino acid coding sequence. A multicopy insert construct was constructed as a part of the expression plasmid. The yeast Pichia pastoris was transformed by the expression plasmid, and a Mut+ His+ cell line was screened. High cell density culture of the cell line was carried out, and the cells were induced with methyl alcohol. The human epidermal growth factor with biological activity was secreted into the medium and was purified through three chromatographic steps. The final yield was 100 mg per liter of cell culture with ~98% homogeneity. All columns were purchased from Amersham Pharmacia Biotech.

Crystallization and Data Collection—Crystallization of hEGF was performed as described (13). The hEGF concentration was about 50 mg/ml. After successive rounds of crystallization refinements using the hanging-drop vapor-diffusion method, larger hEGF crystals were grown from a solution containing 0.9 M MgCl2, 3.5 mM CYMAL-3 (cyclohexylpropyl-b-n-maltoside) and 0.1 M Bicine (pH 8.1) at 291 K over a period of about two months. The hEGF crystals have a typical size of 0.4 × 0.3 × 0.3 mm3, and can eventually reach a size of 0.5 × 0.5 × 0.6 mm3. These crystals belong to the space group P21, (a = b = 61.4 Å, c = 87.0 Å). They could diffract x-ray to 3.0 Å resolution at Argonne Station of synchrotron radiation (Native 1 in Table I), and to 3.2 Å on a MarResearch IP detector, using Cu Kα x-ray from Rigaku RU-200 rotating-anode generator operating at 40 kV and 100 mA (Native 2). There are two EGF molecules (denoted by molecules A and B in the text) in the asymmetric unit of the trigonal crystals, giving a Vm of 3.82 Å3/Da (15) and a corresponding solvent content of 67.6%. The weak diffractability of hEGF crystals may be related to the higher solvent content. The difficulty in growing good quality EGF crystals may be caused by the marked conformational flexibility of the EGF structure, which will be discussed in detail in the text.

Several heavy-atom derivatives were prepared by soaking the native crystals for 3 to 7 days at 293 K in storage solution containing an appropriate concentration of dissolved heavy atom compound. Intensity data for the heavy-atom derivatives were collected at room temperature on the MarResearch IP detector, using Cu Kα x-ray from Rigaku RU-200 rotating-anode generator operating at 40 kV and 100 mA. All
The multiple isomorphous replacement (MIR) was used to determine the crystal structure of human epidermal growth factor (hEGF). Molecular replacement studies with NMR models of mouse EGF (mEGF) and several EGF-like domains as the search model were carried out with both data sets, Native 1 and Native 2. However, all these efforts failed, perhaps due to the difference between the NMR model and the crystal structure to be discussed in the text.

The multiple isomorphous replacement (MIR) was used to determine the hEGF structure, and all calculations were performed using corresponding programs in the CCP4 package (17) and Native 2 data. The heavy atoms in the derivative crystals were found by difference Patterson method and difference Fourier maps. Heavy atom positions were refined, and phases were calculated using the program MLPHARE with reflections of F(Fo) > 2.0, resulting in an initial figure of merit of 0.423 for data up to 4.0 Å resolution and contourled at 1.5σ. Carbon, oxygen, and nitrogen atoms are colored yellow, red, and blue, respectively.

It is displayed in stereo using TURBO-FRODO, computed at 3.0 Å resolution and contoured at 1.5σ. Carbon, oxygen, and nitrogen atoms are colored yellow, red, and blue, respectively.

**Table I**

A summary of crystallography data

<table>
<thead>
<tr>
<th>Data set</th>
<th>Native1</th>
<th>Native2</th>
<th>K₃PtCl₆</th>
<th>Hg(Ac)₂</th>
<th>UO₂(NO₃)₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P3₁2₁2₀</td>
<td>P3₁2₁2₀</td>
<td>P3₁2₁2₀</td>
<td>P3₁2₁2₀</td>
<td>P3₁2₁2₀</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>100–3.0</td>
<td>30–3.2</td>
<td>20–4.4</td>
<td>20–4.0</td>
<td>20–4.3</td>
</tr>
<tr>
<td>Number of molecules in asymmetric unit</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Observations/Unique</td>
<td>21097/4040</td>
<td>26521/3257</td>
<td>10710/1095</td>
<td>11950/1599</td>
<td>14001/1754</td>
</tr>
<tr>
<td>φ/θ (highest shell)</td>
<td>99.0</td>
<td>96.8</td>
<td>89.7</td>
<td>91.2</td>
<td>98.5</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>10.4</td>
<td>10.5</td>
<td>10.9</td>
<td>9.3</td>
<td>12.3</td>
</tr>
<tr>
<td>Rmpj (acentric/centric)</td>
<td>0.78/0.70</td>
<td>0.82/0.79</td>
<td>0.93/0.93</td>
<td>0.61/0.49</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>23.1</td>
<td>28.3</td>
<td>4.4</td>
<td>3.7</td>
<td>3.7</td>
</tr>
</tbody>
</table>

R-factor = \( \sum |F_o| - |F_c| / \sum |F_o| \), where \( F_o \) and \( F_c \) are the observed and calculated structure factors, respectively. 

R-free was calculated from the 90% of reflections used in refinement, and Rmerge was calculated from the remaining 10%.

**Fig. 1.** A 2Fo-Fc electron density map surrounding residues 21–29 and Tyr²⁴. It is displayed in stereo using TURBO-FRODO, computed at 3.0 Å resolution and contoured at 1.5σ. Carbon, oxygen, and nitrogen atoms are colored yellow, red, and blue, respectively.
RESULTS AND DISCUSSION

Model Quality—A summary of the structural analysis including data collection, phasing and refinement statistics for the hEGF crystals is presented in Table I. The values of $R_{cryst}$ and $R_{free}$ for the refined hEGF structure are 23.1 and 28.3%, respectively for the 8.0–3.0 Å data with $F$/$\sigma(F)$. Amino acid residues 1–5 in both molecules A and B and the residues 48–51 in molecule A are disordered, as these regions are poorly defined in the electron density map. 98.6% of the remaining residues have appropriate backbone torsion angles in the most favorable and additionally allowed regions of the Ramachandran plot.

Except the two terminal segments, most main-chains in both hEGF molecules have well defined electron densities when contoured at 1σ level (Fig. 1). Most side chains are also unambiguously located in the density map, whereas some polar residues on the molecular surface have poor densities showing their conformational disorder to some extent. The three disulfide bridges, Cys6–Cys20, Cys14–Cys31, and Cys33–Cys42, are also located in clearly defined electron densities.

hEGF Structure—The crystal structure of hEGF (Fig. 2a) shares main features of the NMR solution structures available. It is a structure consisting of an N-domain (residues 1–32) and a C-domain (residues 33–53). The N-domain has an irregular N-terminal peptide segment with residues 1–12 and an anti-parallel β-sheet (residues 19–23/28–32). The C-domain contains a short anti-parallel β-sheet (residues 36–38/44–46) and a C-terminal segment with residues 48–53, which are probably disordered in isolation.

Despite structural similarity between molecules A and B in the asymmetric unit, large local differences are found in peptide segments 6–12, 22–29, and 48–51, respectively (Fig. 2b). The most obvious difference is located at the N-terminal residues up to residue Gly12. Besides the disordered N-terminal segment with residues 1–5, residues 9–11 are well defined in molecule A, but not in molecule B. This difference may be caused by their different crystallographic environments. There are some intermolecular contacts between molecule A and a neighboring EGF molecule. They include three hydrogen bonds between residues Pro7, Ser9 of molecule A, and Val35, Cys38 of a neighboring molecule B in another asymmetric unit. However, there are no such kind of interactions between the N-terminal residues 6–11 of molecule B and any neighboring EGF molecules in the crystal. Another structural difference occurs at the surface turn with the residues 23–28 connecting the two anti-parallel β-strands of N-domain. The difference in the C-terminal peptide segment correlates with different exhibitions of electron densities for both molecules. In molecule B, residues 48–51 have clearly defined electron density, likely due to the intramolecular interactions between the residue 49 and other residues, such as the hydrogen bonds between Trp49-O and Arg44-NH1, Trp49-N and Asp46-O, whereas no electron density could be observed for residues 48–51 in molecule A.

Based on electron density map and structural comparisons between molecules A and B, several segments with rigid or flexible conformations could be defined in the EGF structure. Fig. 2b and Table II show that the rigid segments include residues 13–21 and 30–47. The RMSD for Co atoms of these residues between molecules A and B is 0.517 Å. Another indication of inherent rigidity of these regions is that the average B-values of these residues are 22.14 and 26.04 Å$^2$ for main and side-chain atoms, respectively, compared with 30.07 and 36.59 Å$^2$ for the whole molecule. Two disulfide bridges, Cys14–Cys31 and Cys33–Cys42, along with the highly conserved Gly18 and Gly29 play an important role for formation of the rigid region of the structure. The N-terminal segment (residues 1–12) and the
residues 23–28 concerning the irregular \( \beta \)-turn between the two \( \beta \)-strands in the major \( \beta \)-sheet, which are adjacent to each other on the head of the EGF molecule, and the C-terminal segment (residues 48–53) belong to the flexible regions of the EGF structure.

**Comparisons with NMR Structures**—Comparisons by least-square superposition based on Cα atoms between the crystal and the NMR structures of EGF (9) were performed, and the results are shown in Table II. The whole molecule with residues 6–47 (without the disordered terminal segments), N-domain (6–32), and C-domain (33–47) were compared separately.

The comparisons of the crystal structure with the NMR structure at pH 2.0 have revealed that large structural differences are mainly located in the terminal segments, surface turns or loops, including residues 6–11, 24–27, 34–35, 39–40, and 45–47.

The flexible peptide segments (residues 1–12, 22–29, and 48–53) were excluded from more accurate comparisons between the NMR structure at pH 2.0 and the crystal structure. The overall RMSD based on the whole molecule (residues 13–21 and 30–47) for both molecules A and B are about 1.7 Å. However, the RMSD values based on a single domain are much lower, namely, about 0.7 Å for residues 13–21 and 30–32, and about 1.07 Å for residues 33–47 (Table II). This difference may result from a relative rotation of the two domains around a hinge residue between them. The structural variation might be caused by acidic pH, thus resulting in a very different arrangement of the C-terminal segment in the EGF solution structure. Compared with the positions in the crystal structures of molecules A and B, the Cα atom of Leu47, a crucial residue for receptor binding (22), has moved by more than 2.75 Å in the NMR structure at pH 2.0.

Regarding the comparisons with the NMR structure at pH 6.8, despite the closer pH values, the RMSD values are larger than those calculated with the NMR structure at pH 2.0 (Table II). This implies that the NMR structure at pH 6.8 may not be sufficiently accurate due to the much lower number of distance constraints used compared with calculation of the pH 2.0 NMR structure (9).

**Potential Dimer**—Molecules A and B in the asymmetric unit are in close end-to-end contact and are related by a non-crystallographic 2-fold axis. The buried area between the two molecules is \(-690\, \text{Å}^2\), which might be enough to maintain two such small molecules together to form a dimer under certain circumstances. The dimerization interface concerns the minor \( \beta \)-sheets from the C-domains of the two molecules, forming a short four-stranded anti-parallel \( \beta \)-sheet. The non-crystallographic 2-fold axis passes the center of this four-stranded \( \beta \)-sheet (Fig. 2a).

Leu15, His16, Tyr37, Arg41, Gln43, Tyr44, Arg45, and Leu47 from both hEGF molecules are involved in the intermolecular interface of the potential EGF dimer, although partial surfaces of side chains for some of these residues may be accessible. Besides the hydrogen bonds formed by atoms Gln43-O and Tyr37-OH in the dimer, there are additional intermolecular hydrogen bonds between atoms His16-NE2 and Tyr37-OH in the dimer.

**Conclusions and Implications**—The crystal structure of hEGF determined at near physiological pH shares the main features of the NMR structure of mEGF determined at pH 2.0, but structural comparisons between different models revealed further details of the EGF structure. The structural differences of hEGF molecules A and B have shown detailed flexibility of the residues 22–29 in the hEGF structure. The structural comparison with the NMR structure at pH 2.0 may provide the first indication of the existence of relative movement between the N-domain and the C-domain of the EGF molecule.

The most important finding is that the dimerization of EGF molecules can occur under certain conditions. Notably, nearly all residues known to be crucial for EGF activity (23, 24), i.e., residues Leu15, His16, Tyr37, Arg41, Gln43, and Leu47 (Fig. 3), are involved in the intermolecular interface of the potential EGF dimer. It has been reported that the EGF-receptor complex contains two EGF molecules (3). Taking into account the...

---

**Table II**

<table>
<thead>
<tr>
<th>Models compared</th>
<th>Residues 6–47</th>
<th>6–32</th>
<th>RMS Deviations (Å)</th>
<th>33–47</th>
<th>13–21, 30–32</th>
<th>13–21, 30–47</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &amp; B</td>
<td>3.266</td>
<td>3.257</td>
<td>0.332</td>
<td>0.529</td>
<td>0.517 (0.252)</td>
<td>1.694 (2.751)</td>
</tr>
<tr>
<td>A &amp; N2</td>
<td>2.542</td>
<td>1.996</td>
<td>1.067</td>
<td>0.757</td>
<td>1.754 (2.832)</td>
<td></td>
</tr>
<tr>
<td>B &amp; N2</td>
<td>3.952</td>
<td>3.312</td>
<td>1.078</td>
<td>0.702</td>
<td>2.224 (6.497)</td>
<td></td>
</tr>
<tr>
<td>A &amp; N6.8</td>
<td>2.423</td>
<td>2.234</td>
<td>1.841</td>
<td>1.466</td>
<td>2.313 (6.681)</td>
<td></td>
</tr>
<tr>
<td>B &amp; N6.8</td>
<td>3.769</td>
<td>3.235</td>
<td>1.799</td>
<td>1.558</td>
<td>2.315 (6.681)</td>
<td></td>
</tr>
</tbody>
</table>

a A = molecule A; B = molecule B; N6.8 = mouse NMR structure at pH 6.8; N2 = mouse NMR structure at pH 2.0.

b The values in parentheses are distances between Cα atoms of the residue Leu47.

---

**Fig. 3. Space-filling model of the EGF molecule.** The figure shows the distribution of some surface residues known to be important for EGF binding to its receptor: Tyr23, Leu15, His16, Tyr37, Arg41, Gln43 (in green), Ile27, Ala25, Leu26 (in yellow), and Leu47 (in red).
importance of ligand oligomers in the ligand-induced dimerization of receptors in some cases (25, 26, 27, 28), this potential EGF dimer might be biologically relevant and play a special role in the dimerization of EGF receptors. This suggestion is inconsistent with the models proposed by Lemon et al. (3) in 1997, where any EGF aggregation is excluded from the molecular details of the EGF-induced dimerization of receptors. If the hypothesis concerning EGF dimerization is correct, and the consequence of the relative movement of the N- and C-terminal domains may impede formation of the potential EGF dimer; this may then account for the inactivation of EGF at acidic pH (10).

The important residues such as Leu^{15}, His^{16}, Tyr^{37}, Arg^{41}, Gln^{43}, and Leu^{47} are thought to be in the site of EGF binding to its receptor (23, 24). However, according to our hypothesis they are mainly involved in the intermolecular interface of the potential EGF dimer. So there must be other receptor binding sites in the EGF molecule, which are involved in formation of a potential EGF dimer. Mutation and chimera studies of sites in the EGF molecule, which are involved in formation of a potential EGF dimer; these changes of these residues do not alter the type of scaffold in the complex with the receptor (25). It is thought that they are mainly involved in providing a low affinity EGF receptor antagonist. Its peptide segment residues are mainly involved in the interaction of EGF with its receptor. Support for this comes from a structural comparison of EGF with the 39-amino acid potato carboxypeptidase inhibitor. The latter is a scaffold of the receptor complex. To verify the above hypothesis, it is crucial to show the different requirements for EGF-induced heterodimerization of ErbB receptors, which is involved in the binding interface and undergoes a conformational change in the complex with the receptor (25). Therefore, important structural changes might occur in the flexible region involving residues 22–29 during EGF binding to its receptor.

In addition, if the above speculation is correct, this has implications for the heterodimerization of ErbB receptors, where there exists a 1:2 complex of ligand with receptors. It would suggest that the EGF dimerization might not play a role in the heterodimerization of ErbB receptors, and residues Leu^{15}, His^{16}, Tyr^{37}, Arg^{41}, Gln^{43}, and Leu^{47} might play other roles in this case. Thus, further mutational studies of EGF are needed to show the different requirements for EGF-induced ErbB-1 homodimerization and for EGF-induced ErbB-1/ErbB-2 heterodimerization. However, in the process of heterodimeric formation, it could not be excluded that first a 2:2 ErbB-1 homodimer has to be formed, and subsequently ErbB-2 is involved, giving rise to a 2:4 complex, where ligand dimerization may be also of relevance for the formation of the heterodimeric receptor complex. To verify the above hypothesis, it is crucial to determine the structure of the complex of EGF with its receptor.

Acknowledgments—We thank Dr. Yu Luo and Prof. Ming Luo, Department of Microbiology, Center for Biophysical Science and Engineering, University of Alabama at Birmingham for collecting the 3.0 Å resolution data at Argonne Station. We thank Prof. E. J. J. Van Zoelen, Department of Cell Biology, University of Nijmegen, The Netherlands, for his helpful discussion about the molecular mechanism of ligand-induced dimerization of receptors. We also thank Prof. Yigong Shi, Department of Molecular Biology, Princeton University, for his reading the manuscript.

REFERENCES