Characterization of a comparative model of the extracellular domain of the epidermal growth factor receptor

ROBERT N. JORISSEN,^{1,4} V. CHANDANA EPA,² HERBERT R. TREUTLEIN,^{1,4}
THOMAS P.J. GARRETT,^{2,4} COLIN W. WARD,^{3,4} AND ANTONY W. BURGESS^{1,4}

¹Ludwig Institute for Cancer Research, Post Office Box 2008, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia

²Biomolecular Research Institute, 343 Royal Parade, Parkville, Victoria 3052, Australia

³The Commonwealth Scientific and Industrial Research Organization Division of Health Sciences and Nutrition, 343 Royal Parade, Parkville, Victoria 3052, Australia

⁴Cooperative Research Centre for Cellular Growth Factors, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia

(RECEIVED October 1, 1999; FINAL REVISION December 6, 1999; ACCEPTED December 7, 1999)

Abstract

The Epidermal Growth Factor (EGF) receptor is a tyrosine kinase that mediates the biological effects of ligands such as EGF and transforming growth factor alpha. An understanding of the molecular basis of its action has been hindered by a lack of structural and mutational data on the receptor. We have constructed comparative models of the four extracellular domains of the EGF receptor that are based on the structure of the first three domains of the insulin-like growth factor-1 (IGF-1) receptor. The first and third domains of the EGF receptor, L1 and L2, are right-handed beta helices. The second and fourth domains of the EGF receptor, S1 and S2, consist of the modules held together by disulfide bonds, which, except for the first module of the S1 domain, form rod-like structures. The arrangement of the L1 and S1 domains of the model are similar to that of the first two domains of the IGF-1 receptor, whereas that of the L2 and S2 domains appear to be significantly different. Using the EGF receptor model and limited information from the literature, we have proposed a number of regions that may be involved in the functioning of the receptor. In particular, the faces containing the large beta sheets in the L1 and L2 domains have been suggested to be involved with ligand binding of EGF to its receptor.

Keywords: comparative modeling; epidermal growth factor receptor

The epidermal growth factor (EGF) receptor mediates the biological actions of several structurally related ligands such as epidermal growth factor and transforming growth factor alpha (TGF- α) (Carpenter & Cohen, 1990; Ullrich & Schlessinger, 1990; Groenen et al., 1994). These proteins are involved in the regulation of cell proliferation and maturation in a number of cell lineages (reviewed in Pimentel, 1994). Increased or constitutive activity of the EGF receptor through mutation (Huang et al., 1997), gene amplification (Merlino et al., 1984; Ullrich et al., 1984), overexpression (King et al., 1985; Gullick et al., 1986), or increased amounts of an EGF receptor ligand (Smith et al., 1987) has been observed in a number of tumors including brain (Malden et al., 1988), head and neck

(Grandis & Tweardy, 1993), lung (Haeder et al., 1988), and pancreatic (Korc et al., 1992) and colorectal tumors (Messa et al., 1998). A mutant form of the receptor, in which 43% of the extracellular domain (exons 2–7, which encode for residues 6–273) is deleted, is constitutively active and has been observed in many glioblastoma and breast and ovarian cancers (Moscatello et al., 1995).

The EGF receptor (or ErbB-1) is a cell surface glycoprotein consisting of a single polypeptide chain of 1,186 amino acids (Ullrich et al., 1984). Glycosylation accounts for 35 kDa of the protein's 170 kDa mass (Lax et al., 1990). Ligand binding to the EGF receptor's extracellular domain (or ectodomain) induces dimerization of the receptor (Yarden & Schlessinger, 1987). Subsequent activation of the protein's cytoplasmic tyrosine kinase domain initiates the signal transduction pathways, leading to a variety of biological responses (Ullrich & Schlessinger, 1990). EGF binds to both low-affinity ($K_D = 1\text{--}2\text{ nM}$) and high-affinity ($K_D = 10\text{--}50\text{ pM}$) sites on EGF receptor-expressing cells (King & Cuatrecasas, 1983). Studies using the soluble ectodomain of the EGF receptor have shown that this fragment of the protein retains

Reprint requests to: Robert N. Jorissen, Ludwig Institute for Cancer Research, P.O. Box 2008, Royal Melbourne Hospital, Victoria 3050, Australia; e-mail: robert.jorissen@ludwig.edu.au.

Abbreviations: CRD, cysteine-rich domain (of IGF-1R); EGF, epidermal growth factor; IGF-1R, insulin-like growth factor-1 receptor; RMSD, root-mean-square deviation; TGF- α , transforming growth factor alpha; TNF, tumor necrosis factor.

its ligand binding and dimerization properties, although the affinity of the ectodomain for EGF has been measured to be 100–500 nM (Lemmon et al., 1997, and references therein). The measured stoichiometry of the dimerized receptor ectodomain to ligand is 2:2 (Lemmon et al., 1997; Odaka et al., 1997; Domagala et al., 2000).

Three homologues of the EGF receptor have been identified in humans, ErbB-2 (or neu) (Schechter et al., 1985), ErbB-3 (Kraus et al., 1989), and ErbB-4 (Plowman et al., 1993). The overall sequence identity between the EGF receptor and its three human homologues is between 40 and 50%, but is variable over the different domains with the tyrosine kinase domains being the most conserved at 80% identity (Peles & Yarden, 1993). EGF motif-containing proteins such as TGF- α (Anzano et al., 1983; Derynck et al., 1984) and the neuregulins (Peles & Yarden, 1993; Burden & Yarden, 1997) can induce the formation of specific receptor homodimers and heterodimers, which consist of two EGF receptor family members (Alroy & Yarden, 1997; Riese & Stern, 1998). The binding affinity of the ligand and its ability to induce dimerization depend on the identity of each of the receptors. For example, while EGF can induce receptor dimers that include at least one EGF receptor molecule, it cannot induce the formation of ErbB-3 or ErbB-4 homodimers or the ErbB-3/ErbB-4 heterodimer. The molecular basis for these specificities and affinities is yet to be determined; however, a detailed understanding of the structures of the receptor and receptor-ligand complexes should lead to a solution of this phenomenon.

On the basis of internal sequence conservation and identity among the EGF receptor family members, the ectodomain has been classified into four domains, the L1, S1, L2, and S2 domains (Ullrich et al., 1985; Bajaj et al., 1987; Ward et al., 1995). The S1 (residues 166–312) and S2 (residues 482–621) are characterized by a large number of cysteine residues, 22 and 20, respectively. The L2 domain has been demonstrated to be the major ligand binding domain (Lax et al., 1989). Although the EGF receptor ectodomain has been crystallized, it has so far resisted direct structural determination (Günther et al., 1990; Degenhardt et al., 1998).

The recent report of the three-dimensional structure of the L1, cysteine-rich (CRD), and L2 domains of the insulin-like growth factor 1 receptor (IGF-1R) from X-ray diffraction studies (Garrett et al., 1998) provides a substantial basis for constructing a model of the ectodomain of the EGF receptor. These first three domains of the insulin-like growth factor-1 receptor (Morgan et al., 1986), the L1, cysteine-rich, and L2 domains, are homologous to the first three domains of the EGF receptor (Ullrich et al., 1985; Bajaj et al., 1987; Ward et al., 1995). The sequence identities between the first three domains of the EGF receptor and those of the IGF-1R are ~25%.

The three-dimensional structure of the first three domains of the IGF-1R was used to build a comparative (or homology) model of the first four domains of the EGF receptor ectodomain. Due to the low sequence identity between the two receptors, construction of the models represented a nontrivial problem for the comparative modeling method (Tramontano, 1998). Our models were evaluated with reference to published experimental data on the EGF receptor.

Results

Models were constructed separately for the L1 and S1 domains and the L2 and S2 domains of the EGF receptor. The sequence align-

ments shown in Figure 1 were used in conjunction with the MOD-ELLER program (Šali & Blundell, 1993) to generate the EGF receptor models as described in Materials and methods. Diagrams of the L1–S1 and L2–S2 models are shown in Figure 2.

Construction and description of models of the L1 and L2 domains

To aid with the identification of residues in the β -helical L1 and L2 domains of the IGF-1R (Garrett et al., 1998), the four sides of these were numbered from 1 to 4. The first three faces contain the first (colored blue in Garrett et al., 1998), second (green), and third (yellow) β -sheets of the IGF-1R structure. Several residues that mediate interactions between L1 and the CRD of IGF-1R are conserved in the EGF receptor family. In particular, Trp176 of the CRD of IGF-1R is inserted between the last two strands of the fourth face of the L1 domain and surrounded by a number of hydrophobic residues including Trp127. These two residues are conserved in the S1 and S2 domains and in the L1 and L2 domains, respectively, of the EGF receptor family (Fig. 1).

The alignments of the L1 domain of the IGF-1R to the L1 (Fig. 1A) and L2 (Fig. 1B) domains of the EGF receptor were constructed manually. Except for two small fragments, the L2 domain of IGF-1R was not used as template for constructing the models because the backbone of the last two strands of the fourth face of the L2 domain differ significantly from those of the L1 domains that surround Trp176 from the CRD of IGF-1R. As the L1 and L2 domains of IGF-1R form β -helical structures with a defined core, a factor in constructing the alignment was to place most of the hydrophobic residues that are conserved within the EGF receptor family in positions where their side chains point toward the core of the β -helix. Asn116 and Asn142 of IGF-1R in the fourth and fifth strands of the third face form a short asparagine ladder (Yoder et al., 1993), where their side chains make hydrogen bonds with main-chain atoms in the interior of the β -helix. Several asparagine residues in the EGF receptor sequence also take up the same position within a given helical turn (marked with asterisks in Fig. 1). Other IGF-1R residues conserved in the alignments of the L1 and L2 domains of the EGF receptor are the four cysteine residues and Trp127.

Two fragments of the IGF-1R L2 domain were used as additional templates in the modeling process, and these sequences were included in the alignments (Fig. 1). The structures of the L1 and L2 domains of IGF-1R contain a loop that is between the first and second faces of the fourth helical turn. While structurally similar, the loop from L2 is one residue longer than that of L1 and is a more suitable template for the equivalent putative loop in the L1 and L2 domains of the EGF receptor (Fig. 1). A second template from the L2 domain of IGF-1R (residues 357–379) contains a number of hydrophobic residues that, in the fourth face of the IGF-1R L2 β -helix, point to the core of the β -helix. Use of this template was more consistent with the presence of conserved hydrophobic residues in Leu52–Leu55 of the EGF receptor than was using the corresponding region of IGF-1R from the L1 domain.

A number of structural features in the IGF-1R L1 domain remain in the homology-built models of the L1 and L2 domains of the EGF receptor (Fig. 2). The short asparagine ladder observed in third face of IGF-1R (Garrett et al., 1998) is retained in the L1 domain (residues 100, 129, and 158) and the L2 domain (residues 442 and 469). A helix in the first strand of the fourth face and the loop in the fourth strand between the first and second faces of the

IGF-1R L1 domain are also present in the two models (Fig. 2). Glycine residues are present in the first five helical turns of L1 and L2 domain models (residues 9, 39, 63, 85, and 122 in L1 and 315, 343, 379, 404, and 435 in L2). These residues are present in the corner between the face 1 and face 2 of each of the β -helices. Most of the corresponding residues in IGF-1R domain 1 are also glycines (Fig. 1).

A number of regions of sequence in the alignments of the L1 and L2 domains of the EGF receptor were not aligned to any part of the IGF-1R sequence and were modeled as loops in the structural models. Both of the L1 and L2 domain models of the EGF receptor contain such an insertion between the first strand of the second face of each of the β -helices and the N-terminal helix (Leu17–

Asp22 in L1 and residues Phe321–Ile327 in L2). In the model of the L1 domain, there is an insertion, modeled as a loop, present in the fourth strand of the third face of the L1 domain model (Asp102–Thr106). A proximal, one residue deletion occurs in the third helical turn, between the second and third faces of the model. There are two insertions, Phe352–Glu367 and Glu388–Thr391, which are located around the corners of the third and fourth faces of the second and third helical turns of the L2 model, respectively. A one-residue deletion occurs between Leu419 and Asn420 in the third face of the fourth helical turn. The effect of these insertions and deletions is to reduce the amount of secondary structure in the third face of both the L1 and L2 domains of the models of the EGF receptor compared to the domains in IGF-1R.

A L1 domain

Fig. 1. Sequence alignments of the L1 and cysteine rich domains of the IGF-1R sequence and the EGF receptor for (A) L1 and S1 domains and (B) L2 and S2 domains. Sequences of the insulin receptor and human EGF receptor homologues were included to improve alignment for both families of proteins. Additional fragments of the IGF-1R sequence are also included in the alignment. Residues in lower case were not aligned to residues in the sequence alignments used to generate models; they are shown as aligned for clarity. Residues that are underlined were used to generate additional distance restraints for the modeling process as described in Materials and methods. Shaded residues are part of secondary structure elements according to the assignment of Garrett et al. (1998), and the relevant secondary structure element is shown schematically above the appropriate residues. The number of the helical turn and topographical face (labeled F1 to F4) occupied each residue in the main IGF-1R L1 domain sequence are shown above the sequences. The disulfide bonds of models constructed from these alignments are shown as lines joining the two relevant cysteine residues. Selected conserved residues are shown in bold typeface. Asterisks are shown under selected asparagine residues that, in the relevant EGF receptor model, form part of an asparagine ladder (Yoder et al., 1993). The modules of the (A) S1 and (B) S2 domains of the EGF receptor models constructed using the alignments are also shown. In A, the numbering of the modules also applies to the cysteine-rich domain of the IGF-1R. (Figure continues on facing page.)

Structural features of cysteine-rich domains

The cysteine-rich domain (CRD) of IGF-1R is considered to consist of eight small modules that each contain one or two disulfide bonds (Garrett et al., 1998). Short linker regions, usually of two or three residues in length, join the modules together. Thorough examination of the modules and linker regions of the CRD of IGF-1R, the tumor necrosis factor (TNF) receptor (Naismith et al., 1995) and the third, fourth, and fifth laminin-type EGF-like modules of the mouse laminin gamma-1 chain (Stetefeld et al., 1996) revealed several correlations between sequence and structure of the CRD that have not been described previously and were subsequently exploited in the construction of the sequence alignments of the S1 and S2 domains of the EGF receptor (Fig. 1). These correlations may assist in the modeling of other cysteine-rich regions of proteins such as furin-like proteases (Ward et al., 1995).

The first three modules are so-called C2 modules (Ward et al., 1999), each of which contain four cysteine residues with a Cys1-Cys3 and Cys2-Cys4 disulfide bond pairing arrangement. The cysteine C_{α} atoms of the three C2 modules of IGF-1R can be superimposed with RMSD <1 Å (Fig. 3A). The distances between the C_{α} atoms of the first and second cysteine residues in the second and third modules of the CRD, which are separated by three residues in both modules, are 6.7 and 6.6 Å, respectively. These

values are different to the corresponding residue spacings and distances in the C2 modules of the TNF receptor (two residues and average distance of 5.6 ± 0.4 Å) and the laminin gamma-1 chain fragment (one residue and average distance of 9.2 ± 0.5 Å). This comparison suggests that where the number of residues is small, a relation exists between the number of residues between the first two cysteine residues of C2 modules and the distance between the C_{α} atoms of the cysteines. The sequence of the third module of the CRD can be aligned to three regions of sequence in each of the S1 and S2 domains of the EGF receptor (Fig. 3B), each of which have three residues between the first two cysteines.

The fourth, fifth, sixth, and seventh modules of the CRD of IGF-1R are S-shaped modules, called β -finger motifs (Garrett et al., 1998). Each is a C1 module (Ward et al., 1999), bounded by two cysteine residues that form a disulfide bond. Despite the lack of obvious sequence similarities, some structural similarities were observed. The N-terminal cysteine residue, seven C-terminal residues of the fourth and sixth modules (which we term C1^C modules), and the four residues (three linker residues and the cysteine of the next module), which follow each of these modules, superimpose with RMSD 0.73 Å (Fig. 3C). Allowing for a one-residue insertion, the C-terminus of the fifth module of the IGF-1R CRD can also be superimposed on to those of the fourth and sixth modules (Fig. 3D). The first eight residues and the last cysteine

B L2 domain

S2 domain

Fig. 1. Continued.

The eighth module of the CRD of IGF-1R is a small disulfide-linked bend that has no equivalent in the TNF receptor or laminin gamma-1 chain fragment. The equivalent region of the EGF receptor can be aligned to the sequence of the module of the CRD without gaps (Fig. 1A). The resultant alignment of the EGF receptor S1 and S2 domains to equivalent modules of the IGF-1R CRD yielded a unique disulfide bonding arrangement. All of the cysteine residues were predicted to form the same disulfide bonds as those determined experimentally by a combination of N-terminal

The entire sequence of the CRD of the IGF-1R was aligned with the sequence of the S1 domain of the EGF receptor (Fig. 1A). The positions of the cysteine residues were conserved because the disulfide bonding arrangements were the same in domains of both of the proteins. The side chains of two residues in IGF-1R, Arg77 in the L1 domain and Gln182 in the CRD, make hydrogen bonds with main-chain atoms in the adjoining domain, suggesting that the relative orientations of the two domains is conserved. These residues are conserved in the L1 and S1 domains, respectively, of the EGF receptor family as are the IGF-1R residues Trp126 and Trp176 (Fig. 1A). The alignment was constructed using the alignment of modules described in the previous sections, but using additional regions of sequence from the IGF-1R CRD as required. The arrangement of modules indicated by the sequence alignment is C2-C2-C2-C1^C-C1^C-C1^C-C1^N-C1, where the last C1 module is the five-residue, disulfide-linked bend, rather than a β -finger. While the fifth module of the CRD of IGF-1R is a C1^N module, the sequence of fifth module of the EGF receptor is followed by three

linker residues, and then another β -finger module and was assigned as a C1^C module. Consequently, its C-terminal rather than N-terminal residues were aligned with those of the sequence of the fifth module of the CRD of IGF-1R (Fig. 1A).

While the modeled structure of the EGF receptor S1 domain (Fig. 2A) is similar to that of the CRD of IGF-1R, there are several regions of difference between the two structures. The first (C2) module of the S1 domain (residues 166–183) is smaller than that of IGF-1R and the fifth module of the EGF receptor S1 model (residues 240–267) is significantly larger than that of the IGF-1R CRD. The N-terminal portion of the sixth module (residues 271–283) is smaller than the corresponding module of the IGF-1R CRD, and so there is less protrusion of this module into the cavity formed by L1, S1, and L2.

Interaction between the L1 and S1 domains of the EGF receptor model has two main components. The first includes the insertion of Trp176 into the L1 domain and hydrogen bonds between the side chain of Gln184 and main-chain atoms of the L1 domain. The second interface consists of the interaction of the first face of the L1 β -helix with the second to fifth modules of the S1 domain (Fig. 2A). The solvent accessible surface area of this interface is $\sim 1,800 \text{ \AA}^2$. The contribution of the first face of the L1 domain to the interface includes residues Glu35, Glu60, and Glu118, whereas that of the S1 domain includes the residues Arg198, Arg228, and Arg231, suggesting that electrostatic interactions are important for stabilizing the interface. This notion is supported by calculations of electrostatic potentials of each of the L1 and S1 domains (Gilson et al., 1988; Nicholls & Honig, 1991) (results not shown). Also present in this interface are hydrogen bonds made between the side chain of the conserved residue, Arg84 in the first face of the L1 domain, and the main-chain oxygen atoms of Cys227 and Arg228 in the fourth module of the S1 domain.

From the alignment of the sequences of the putative modules of the EGF receptor S2 domain (Fig. 3), as described in the previous section, the assignment of modules that make up the S2 domain was deduced to be C2-C1^C-C1^N-C2-C1^C-C1^N-C2. This yields a disulfide bonding arrangement different to those of the CRD of IGF-1R and the S1 domain of the EGF receptor. Sequences of the third, fourth, and fifth modules of the IGF-1R CRD (residues 201–246) were aligned to each set of C2-C1^C-C1^N modules in the EGF receptor S2 domain (residues 482–531 and 534–593) (Fig. 1B). The seventh module of the CRD of IGF-1R plus the five residues that follow it (two linker residues plus the first three residues of the eighth module) were also aligned with each of the two C1^N modules of the S2 domain plus their succeeding residues (Fig. 1B). The last two residues of this CRD template were used to provide overlap between the template and the next template, although there was no structural basis for this. The sequence of the third module of the CRD of IGF-1R was also aligned to the last C2 module of the EGF receptor S2 domain (residues 596–612). The most notable conserved residue, other than the cysteine residues, was Trp176 from the first module of the CRD of IGF-1R (Fig. 1B).

The resultant model of the S2 domain of the EGF receptor is a rod-like structure (Fig. 2B), similar in appearance to the laminin gamma-1 chain fragment, and having the same arrangement of disulfide bonds (Stetefeld et al., 1996). Another consequence of the altered modular arrangement to the S1 domain and CRD is that the S2 domain projects out from its N-terminal L-domain differently to S1 and the CRD (Fig. 2B; cf. Fig. 2A). The interaction between the L2 and S2 domains is primarily limited to the insertion of

Trp492 between the fourth and fifth strands of the fourth face of the L2 domain. There are two long loops in the third and sixth modules (residues 521–530 and 577–592) of the S2 domain (Fig. 2B), both of which are the result of insertions in the sequence alignment of the domain (Fig. 1B).

Evaluation of structure

The two EGF receptor models were analyzed using PROCHECK (Laskowski et al., 1993) to evaluate their stereochemical quality. For the L1–S1 and L2–S2 models, 76.4 and 84.9% of (non-Gly, Pro) residues were in the most favored regions of the Ramachandran plot, 21.7 and 12.8% in the additionally allowed regions, 1.4 and 1.6% in the generously allowed region and 1 and 2 residues in the disallowed region, which is roughly comparable to the corresponding statistics for the structure of the first three domains of the IGF-1R (results not shown). Pro112 and Pro565 were judged to be in unfavorable positions in the ϕ - ψ plot for proline residues. The Profiles-3D (Bowie et al., 1991; Lüthy et al., 1992) and ProsaII (Hendlich et al., 1990; Sippl, 1993) programs were also used to assess the quality of the protein fold of the EGF receptor models (Fig. 4). The Profiles-3D scores of the two EGF receptor models were 82 and 80% (observed score/expected score) for the L1–S1 and L2–S2 models, respectively. This is far above the threshold score of 45%, which indicates a misfolded protein (Lüthy et al., 1992). By comparison, the Profiles-3D score for the first two domains of the IGF-1R was 102%. Using a standard window size of 51 amino acids, none of the ProsaII graphs rise above the zero axis, consistent with the observation for a number of experimental structures (Sippl, 1993), although the graph of the L1 and CRD of IGF-1R has a more favorable average score (pseudo-energy) than do the two models (Fig. 4). Using a smaller window size of 11 residues in the ProsaII plot, several areas become apparent as giving rise to a poor score. These areas in the two models with values greater than zero in the ProsaII plot are residues 13–16, 20–21, 105, 154–164, 181–189, 239–250, 353–356, 358, 368, 408, and 584–587 (Fig. 4).

The residues of the models were assigned into one of three categories that reflect our confidence in their main-chain coordinates. In Fig. 2, residues were colored in green, which were judged to have been modeled with high confidence. These residues have been aligned to the IGF-1R sequence with few or no gaps (Fig. 1). The residues colored orange have a higher possibility of being in error. Such regions tend to include amino acids, which have been aligned with the IGF-1R sequence, but where some uncertainty exists such as the positions of several deletions in the L1 and L2 domains and the regions near insertions in the alignments. In the IGF-1R structure (Garrett et al., 1998), the number of residues (two for L1 and one for L2) appears to influence the position of the start of the N-terminal helix. Therefore, there is uncertainty about the position of the corresponding helices in the L1 and L2 domains of the EGF receptor, and so these regions were also colored orange in Figure 2. The coordinates of residues colored red are judged to be unable to be modeled with any confidence and tend to be part of large gaps in the alignment. Such insertions can be grouped into the N- and C-termini of the EGF receptor ectodomain, large loops (usually in the S1 or S2 domains) that are not expected to interact with a hydrophobic protein core and parts of the L1 and L2 domains, which are expected to contain regions that interact with the hydrophobic core of these domains.

Discussion

Analysis of the EGF receptor ectodomain models

We have constructed molecular models of the two halves of the EGF receptor ectodomain using comparative modeling techniques. Evaluation of the model showed that it is consistent with having a correct fold. However, we recognize that a comparative model is, by its nature, less accurate than an experimentally determined structure, and any interpretation of the behavior of the EGF receptor in terms of its structure must be tempered by the knowledge that parts of the model will be inaccurate. Never-

theless, many regions of the model appear to be informative and will assist the determination of the regions of the receptor that are involved in the ligand binding and dimerization properties of the EGF receptor.

Three sources of structural uncertainty in the coordinates of the EGF receptor models can be identified. The first uncertainty is the relative orientation of the L1-S1 EGF receptor model to that of the model of the L2 and S2 domains. One possible orientation is achieved by superimposing the first three domains of the EGF receptor models onto those of the IGF-1R structure (Garrett et al., 1998), as shown in Figure 5A. There is little contact between the

Fig. 3. See caption on facing page.

Fig. 4. Evaluation of the structure IGF-1R L1 and cysteine-rich domains, EGF receptor model for L1 and S1 model and EGF receptor model for L2 and S2 model by ProsaII. The thin line in the plots is the trace for a window size of 11, and the thick line is for a window size of 51.

C-terminal portion (residues 299–301) of the CRD and the L2 domain in the structure of the first three domains of IGF-1R, and this region has been proposed to be a “hinge” about which rotation of the L2 domain relative to the other two domains may possibly occur (Garrett et al., 1998). This region is conserved in the EGF receptor (residues 310–312) and, thus, the relative orientation of the two halves of the EGF receptor in Figure 5A may not represent that found under physiological conditions.

While the L1, S1, and L2 domains of the EGF receptor were modeled using at least one template from the IGF-1R structure that spanned the length of each domain, the largest template of IGF-1R used in the modeling of the S2 domain spanned only three modules (residues 201–246; modules 3–5 of the CRD) (Fig. 1B). Thus, the positions of the fourth and seventh domains relative to the third and sixth modules, respectively, are not well defined. Furthermore, only superimposition of the first module of the S2 domain provided information about the orientation of this domain relative to the L2 domain. Thus the relative orientation of some of the modules in the modeled structure of the S2 domain is not well defined. This lack of confidence in the coordinates is not represented by the coloring of the C_α trace of the protein (Fig. 2).

The third source of structural uncertainty arises from regions that are difficult to model accurately due to uncertainty in the sequence alignment or the lack of a structural template. It is this uncertainty that the color scheme used in Figure 2 attempts to represent. Most of the regions of the two EGF receptor models

whose residues receive the poorest score (high pseudo-energy) in the ProsaII plot (Fig. 4) have been colored red, indicating a low confidence in the modeled coordinates, or are close to such regions (Figs. 2, 4). This suggests that a major part of the inferior ProsaII plot and low Profiles 3D score, compared to those of the first two domains of IGF-1R, are due to uncertainties in these regions. One exception is the region of residues 154–164, which were colored orange (indicating moderate confidence in the main-chain coordinates) in Figure 3A. A number of models with varying alignments in this region were constructed, but none resulted in a favorable ProsaII plot for the region (results not shown).

Comparison of the models with the reported properties of the EGF receptor

Averaged electron microscopy images of the EGF receptor ectodomain showed a structure with four lobes, suggestive of individual domains (Lax et al., 1991b), which is consistent with the EGF receptor four modeled domains adopting separate structures. The calculated radius of gyration for the two models with relative orientations as shown in Figure 5A is 38 Å and compares well with the experimentally determined figure of 36 Å (Lemmon et al., 1997). (The calculated radii of gyration for the L1–S1 and L2–S2 models are 22 and 29 Å, respectively.) Ignoring the first four residues of the L1–S1 model and the last nine residues of the

Fig. 3. (*facing page*) Superimposition of modules of the cysteine rich domain of the IGF-1R structure (**A**, **C**, **D**, and **F**) and associated sequence alignments (**B**, **E**, and **G**). Sequence alignments of EGF receptor sequences are also shown. (**A**) First (colored light gray), second (colored gray), and third (colored black with some residues labeled) modules of the IGF-1R superimposed by their cysteine residues. C_α atoms and the cysteine side-chain atoms are shown as sticks. The linker residues that follow the third module are also shown. (**B**) Aligned sequences of the C2 modules shown in **A** based on structural superimposition. As with Figure 1, residues shown in lower case do not align with other residues. Sequences of putative C2 modules and subsequent linker residues in the EGF receptor are also shown. (**C**) Fourth (colored black and labeled) and sixth (colored gray) ($C1^C$) modules of the IGF-1R cysteine-rich domain and the four residues that follow them. Main-chain atoms and atoms that are part of disulfide bonds are shown as sticks. Hydrogen bonds between main-chain atoms are represented by dashed lines. This representation is also used in **D** and **F**. (**D**) Fifth (colored black and labeled) and fourth modules (colored gray) of the IGF-1R cysteine-rich domain and the next two amino acids each. (**E**) Sequence alignment of the fourth, fifth, and sixth modules of IGF-1R based on structural superimposition. Linker residues that follow these modules that structurally superimpose are also included in the alignment. Putative EGF receptor $C1^C$ modules and linker residues are included in this alignment. (**F**) Seventh (colored black and labeled) and fifth (colored gray) ($C1^N$) modules of the IGF-1R cysteine-rich domain. Amino acids 45–66 of the laminin gamma-1 chain (colored light gray) are also shown. As well as the residues that make up the module, the next three residues of the seventh IGF-1R module and the module from the laminin gamma chain structure are shown. (**G**) Sequence alignment of the fifth and seventh modules of IGF-1R based on structural superimposition together with similar sequences of modules and linkers from the TNF receptor and the laminin gamma-1 chain. Putative EGF receptor $C1^N$ modules and linker residues are included in this alignment. Panels **A**, **C**, **E**, and **G** were created using MOLSCRIPT (Kraulis, 1991). In **B**, **E**, and **G**, disulfide bonds within each module are shown as lines, and residues with asterisks below them are not part of the relevant modules but are part of either the linker residues after the modules or the first cysteine residue of the next module.

Fig. 5. C α trace representation of the models for the EGF receptor highlighting regions of interest. Heavy atoms of side chains of selected residues are shown as spheres, and those of cysteine residues are shown in stick representation. The residues of interest, which are described in Discussion, are labeled in this figure. **A:** Model of the L1 and S1 domains and of the L2 and S2 domains of the EGF receptor superimposed onto the IGF-1R structure using the first three domains of each. The section of the trace corresponding to residues 351–364, which make up LA22 monoclonal antibody binding region described by Wu et al. (1989), are colored in magenta. Residues 302–503 of the trace (with the exception of residues 351–364) are colored in yellow. This range of residues corresponds to the proteolytic fragment of the EGF receptor described by Kohda et al. (1993). The residues 520–603, which are deleted in the type II EGF receptor mutation (Humphrey et al., 1991), are colored in violet. A C α trace representation of EGF (Kohda & Inagaki, 1992) is shown to scale for comparison with the cleft formed by the first three domains of the modeled EGF receptor and is not intended to represent a proposed mode of ligand binding. **B:** Another view of the L2 domain and part of the S2 domain. Residues in this figure that are not shown in A are found in the first face of the L2 domain and are highly conserved across EGF receptor homologues and orthologues.

S2–L2 model, which are not aligned to any part of the IGF-1R sequence (Fig. 1), the longest dimension of the composite EGF receptor ectodomain model is 117 Å, which compares favorably with the experimental value of 110 Å. However, this agreement may be fortuitous, because the relative orientation of the two models cannot be predicted with confidence, and the model does not include any glycosylation.

The models of the S1 and S2 domains are qualitatively similar to the those constructed by Abe et al. (1998) by virtue of identical disulfide assignments. These models were constructed using the laminin gamma-1 chain (Stetefeld et al., 1996) as a template, which is evolutionarily less related to the EGF receptor than is IGF-1R. Thus, the model of the S1 and S2 domains described in this paper

is expected to be closer to their experimental (and as yet unknown) structures in the EGF receptor than those of Abe et al. (1998). Also, there was only one linker region in the S1 domain (and none in the S2 domain), which contained residues not aligned to a template, unlike the Abe et al. model (1998).

The structures of the EGF receptor L1–S1 and L2–S2 models suggest that there are at least two requirements for the proper folding of EGF receptors with mutations in their ectodomains: an intact β -helical L1 and/or L2 domain and the presence of the first module of the next S1 or S2 domain. The latter module contains the tryptophan residue (Trp176 or Trp492), which is inserted in to the L-domain. A proteolytically generated fragment of the EGF receptor includes all of the L2 domain and the first module of the

S2 domain (residues 302–503, Fig. 5) and is capable of binding one molecule of EGF or TGF- α (Kohda et al., 1993), indicating that it is correctly folded. Two deletion mutants of the EGF receptor that have been observed in glioblastoma and are expressed on the cell surface lack residues 520–603 from the S2 domain (colored in violet in Fig. 5A), or residues 6–273, which consists of almost all of the L1 domains including the first cysteine residue to approximately two-thirds into the S1 domain (Sugawa et al., 1990; Humphrey et al., 1991). Two EGF receptor truncation mutants missing the first 76 and 124 residues, respectively, of the L1 domain were expressed on the cell surface only in the presence of the full length, wild-type EGF receptor (Lax et al., 1990). Finally, a series of EGF receptor ectodomain mutants that contain insertions of four or five amino acids at particular positions in the L1 or L2 domains were expressed, but the expression level of most of the mutants were too low to allow their ligand binding properties to be measured (Harte & Gentry, 1995).

Potential sites of ligand binding to the EGF receptor

On the basis of the IGF-1R receptor structure and mutational data for the insulin receptor, Garrett et al. (1998) proposed that the cavity formed by the L1, cysteine-rich, and L2 domains of receptors for insulin and insulin-like growth factor-1 form a part of the binding site for their respective ligands. By analogy, the cleft in the EGF receptor ectodomain (Fig. 5) is a possible site of ligand binding. The presence of such a cleft is consistent with a proposition that the ligand can bind between domains L1 and L2 (Ullrich & Schlessinger, 1990). A naturally occurring mutant form of the EGF receptor, which is missing residues 520–603 in its S2 domain, has been observed to bind EGF and TGF- α in a similar manner to their binding to the wild-type receptor (Humphrey et al., 1991), indicating that the intact S2 domain is not required for ligand binding. The approximate dimensions of EGF and TGF- α are $40 \times 20 \times 20$ Å (Kline et al., 1990; Kohda & Inagaki, 1992). Because the distance between the L1 and L2 domains that border the cavity is ~ 30 Å when in the orientation of the IGF-1R structure (Fig. 5A; Garrett et al., 1998), there is sufficient space in the cavity to accommodate a ligand molecule. Figure 5A shows that an EGF molecule can fit into the cleft mainly; no implication of the correct binding orientation of the ligand should be taken from this figure.

Three residues in close proximity to EGF-EGF receptor interaction sites have been identified by cross-linking studies. The N-terminus of EGF has been cross-linked to the residue Tyr101 (Woltjer et al., 1992), which is on the fourth strand of the third face of the L1 β -helix and faces the large cavity (Fig. 5A). The side chain of Asn1 of EGF has also been chemically cross-linked to Lys336 (Wu et al., 1990) in the N-terminal helix in the L2 domain (Fig. 5A) and is on the same side of the domain as residue Ser323 in the insulin receptor whose mutation leads to reduced insulin binding (Roach et al., 1994). With the domain orientation as shown in Fig. 5A, this residue faces the cavity formed by the L1, S1, and L2 domains. The residue Lys465, in the last strand of the β -sheet in the second face of the modeled L2 domain (Fig. 5), has been cross-linked to residue 45 of a triple mutant of EGF and Lys465 of its receptor (Summerfield et al., 1996). The position of this amino acid would not point toward the cavity without rotation of the L2 domain about the putative hinge region of the EGF receptor ectodomain.

There are several residues in the modeled EGF receptor L1 and L2 domains where the side chain is hydrophobic, solvent-exposed

(Fig. 6A,B), and conserved in two or more of the human EGF receptor homologues (Fig. 6C). As hydrophobic residues such as Phe, Tyr, Leu, and Val are more likely than not to be part of a protein–protein interface than exposed to the solvent (Tsai et al., 1997), we propose that second face of each of the L1 and L2 domains contain candidate ligand binding sites. Both of these regions are near sites of ligand–receptor cross-linking (Tyr101 and Lys465). Further support for the second face of L2 acting as a binding surface in preference to the region near Lys336 comes from the observation that both halves of this domain are required for ligand binding (Lax et al., 1991a).

The residues Phe352–Glu367 in the L2 domain, which have been modeled as a large loop, encompass most of the epitope (residues 351–364) to the ligand competitive monoclonal antibody, LA22 (Wu et al., 1989). In the model, these residues are in the third helical turn, between the third and fourth faces of the L2 domain (Figs. 2B, 5). These residues are on the other side of the L2 domain to the second face that contains Lys465, which was cross-linked to bound ligand (Summerfield et al., 1996). One explanation for the behavior of the LA22 antibody is that ligand competition occurs by steric hindrance of the bound antibody to a site away from the antibody's epitope.

Four EGF receptor ectodomain mutants, each of which contains insertions of four or five amino acids, have been reported to exhibit reduced binding of TGF- α but not EGF, compared to the unmodified EGF receptor ectodomain (Harte & Gentry, 1995). The insertions were in positions 162 (residue before the last cysteine of L1), 169 (in the first module of S1), 174 (also in the first module of S1), and 220 (in the third module of S1). The positions of these residues are on regions of the L1 and S1 domains, which are distant from the second face (which faces the cavity formed by the first three EGF receptor domains) (Fig. 5A), which we have proposed to be a ligand binding site. These four EGF receptor residues are also spaced too far apart in the L1–S1 model to form a single binding site for a molecule the size of TGF- α . It should be noted that the EGF receptor ectodomains used in the experiments were not purified and that chemical cross-linking was used to quantify the amount of bound ligand (Harte & Gentry, 1995).

Cells that express the EGF receptor mutant, R497K, retain low affinity binding of TGF- α but lose high affinity binding (Moriai et al., 1994). In the EGF receptor L2–S2 model, Arg497 is present in the second module of the S2 domain and is proximal to the fourth face of the L2 β -helix. This position is distant from the second face of the L2 domain, which is the domain's proposed site of ligand binding (Fig. 5). It has been suggested that the loss of high-affinity binding of TGF- α to this EGF receptor mutant may be due to impaired dimerization rather than impaired binding of ligand to the uncomplexed receptor (Moriai et al., 1994).

Potential sites involved in EGF receptor action

Because few studies have been performed on the EGF receptor to examine the role of individual residues, the models were examined to suggest regions that potentially have a role in the normal functioning of the EGF receptor ectodomain. Several loops in the S1 and S2 domains of the EGF receptor ectodomain models are larger than the templates they were modeled from, suggesting that they are longer than is required for structural integrity of the protein. These loops are in the fifth module of the S1 domain (residues 241–259) and in the third (residues 521–530), fourth (residues

Fig. 6. Highly conserved, hydrophobic residues in the first face of the L1 and L2 domains of EGF receptor models. **A,B:** Connolly surfaces of the lower faces of (A) the L1 domain and (B) the L2 domain. **C:** Sequence alignment of residues in the lower β -sheet of the IGF-1R structure and the EGF receptor model. Residues in red are hydrophobic residues that point to the exterior of the β -helix and are conserved in most of the EGF receptor orthologues and homologues. Arg125 is shown in blue, because it is not hydrophobic but the equivalent residues in ErbB-2, ErbB-3, and ErbB-4 are hydrophobic. Tyr101 and Lys465, which have been cross-linked to bound EGF (Woltjer et al., 1992; Summerfield et al., 1996), are colored magenta. In **C**, the residues colored in green are those whose side chains are expected to point to the interior of the β -helix. (Figure continues on facing page.)

539–546), and sixth (residues 577–592) modules of the S2 domain. Each of the loops exhibits a degree of sequence conservation across the EGF receptor family (Fig. 1). The residues Tyr101–Gly107 in the fourth strand of face three of the L1 domain form a loop that, according to the model, projects out from the L1 β -helix. In the corresponding portion of the IGF-1R structure, there is no such loop (Garrett et al., 1998), suggesting that this loop region is also not required for maintenance of the correctly folded structure of the L1 domain. The length of the loop is seven or eight amino acids in the EGF receptor, ErbB-3, ErbB-4 (Fig. 1A), and their orthologues, whereas the length of the insertion in ErbB-2 and its homologues is 17 or 18 residues (Fig. 1A).

The first face of the L2 domain of the EGF receptor model contains a number of residues whose side chains are solvent exposed and are highly conserved among the EGF receptor family. These include Thr339 (Lys in ErbB-2 and its orthologues), Thr373, Lys375 (Lys or Arg but Glu in ErbB-2 orthologues), Glu376, Arg405, Lys430 (Lys or Arg), Glu431, and Ser433 (Gly in ErbB-2 orthologues). Except for Glu431, all of these residues are conserved in the *Drosophila* orthologue of the EGF receptor (Schejter et al., 1986). Another conserved region is the putative hinge region at the end of the S1 domain (residues Arg310–Val312), which consists of three residues between the last cysteine of the S1 domain and the first cysteine of the L2 domain. This pattern of residues is highly conserved in both the EGF receptor and insulin receptor families (Ward et al., 1995), including orthologues of lower organisms such as *Drosophila* (Schejter et al., 1986; Fernandez et al., 1995). In most of these proteins, the second residue is a lysine, although this residue is arginine in ErbB-2 orthologues.

Two mechanisms for EGF receptor dimerization have been proposed, both of which are able to be accommodated by the models presented in this paper. In the first proposal, ligand binding to the EGF receptor induces a conformational change that exposes a previously obscured dimerization site (Lemmon et al., 1997). The putative hinge region, Arg310–Val312, is a candidate region about which the conformational change (Greenfield et al., 1989) may occur. Alternatively, the change may be localized to one of the long loops in the receptor. In the second hypothesis, the bound ligands are bivalent and each is in contact with both receptors in the 2:2 ligand–EGF receptor complex (Lemmon et al., 1997; Tzahar et al., 1997). The two putative ligand binding surfaces proposed in this paper occur in separate domains (Fig. 6) and may correspond to binding sites for separate ligand molecules in the ligand–receptor complex. A hypothesis of one site-determining ligand specificity and the second site being a common site for all EGF receptor family ligands would account for the ability of receptors to heterodimerize upon stimulation with the ligand of one of the receptors of the dimer (Alroy & Yarden, 1997). EGF and TGF- α are able to bind to a fragment of the EGF receptor, which consists of all of the L2 domains and only small regions of the S1 and S2 domains (residues 302–503) (Kohda et al., 1993), indicating that this domain probably contains the site of initial ligand binding. EGF binds this fragment with a similar affinity (440 nM) as the whole ectodomain, and it has been suggested that interactions of EGF to the L2 domain will dominate binding of EGF to monomeric EGF receptor (Lemmon et al., 1997). This scheme of binding would be similar to one proposed for the binding of granulocyte colony-stimulating factor to its receptor to form a complex with 2:2 stoichiometry (Fig. 6 in Layton et al., 1999).

C Strand 2:				
IGF-1R L1	27	G-----Y	LHILL	33
InsR L1	31	G-----H	LQILL	37
EGFR L1	39	G-----N	LEITY	45
ErbB-2 L1	37	G-----N	LEITY	43
ErbB-3 L1	42	G-----N	LEIVL	48
ErbB-4 L1	41	G-----N	LEITS	47
EGFR L2	343	G-----D	LHILP	349
ErbB-2 L2	351	G-----S	LAFLP	357
ErbB-3 L2	340	G-----N	LDFLI	346
ErbB-4 L2	344	G-----N	LIFLV	350
Strand 3:				
IGF-1R L1	53	E-----Y	LLLF	59
InsR L1	59	D-----Y	LLLF	65
EGFR L1	63	G-----Y	VLTA-	68
ErbB-2 L1	61	G-----Y	VLTA-	66
ErbB-3 L1	66	G-----Y	VLVA-	71
ErbB-4 L1	65	G-----Y	VLVA-	70
EGFR L2	379	G-----F	LLIQA	385
ErbB-2 L2	387	G-----Y	LVISA	393
ErbB-3 L2	376	G-----Y	LNIQS	382
ErbB-4 L2	380	G-----F	LNIQS	386
Strand 4:				
IGF-1R L1	78	G	WKLFY-NYALVIFE	91
InsR L1	84	G	SRLFF-NYALVIFE	97
EGFR L1	85	G	NMYYENSYALAVLSNY	101
ErbB-2 L1	83	G	TQLFEDNYALAVLDNG	99
ErbB-3 L1	88	G	TQVYDGFALFVMLNY	104
ErbB-4 L1	87	G	TKLYEDRYALAFILNY	103
EGFR L2	404	G	RTKQHGQFSLAVVS	418
ErbB-2 L2	412	G	RLHNGAYSLTLQG	426
ErbB-3 L2	401	G	RSLYNRGFSLLIMK	415
ErbB-4 L2	405	G	RVLYS-GLSLILK	418
Strand 5:				
IGF-1R L1	108	R	G-----AIRTEK	115
InsR L1	114	R	G-----SVRIEK	121
EGFR L1	121	H	G-----AVRFSN	128
ErbB-2 L1	129	K	G-----GVLIQR	136
ErbB-3 L1	125	S	G-----GVYIEK	132
ErbB-4 L1	124	N	G-----GVYVDQ	131
EGFR L2	434	D	G-----DVIISG	441
ErbB-2 L2	442	S	G-----LALTIH	449
ErbB-3 L2	432	A	G-----RIYISA	439
ErbB-4 L2	435	A	G-----NIYITD	442

Fig. 6. Continued.

In summary, we have constructed models of the four domains that make up the EGF receptor ectodomain. Comparison of the models with the forthcoming experimental structure of the protein (Degenhardt et al., 1998) will be useful in assessing the comparative modeling method used in this paper. While the models themselves do not have the same accuracy as experimentally determined structures, they do provide a basis for improving our understanding of the EGF receptor. Furthermore, they provide a basis for the construction of site-directed and truncation mutants of the EGF receptor. Using the models, we have made several proposals about the location of sites of the receptor that may be involved in ligand binding. Experimental testing of the hypotheses described in this paper should lead to a fuller understanding of the molecular determinants of ligand binding and dimerization of the EGF receptor.

Materials and methods

Structure analysis and sequence alignments

The structures of the human IGF-1R domains 1–3 (Garrett et al., 1998), the third, fourth, and fifth laminin-type EGF-like modules of the mouse laminin gamma-1 chain (Protein Data Bank (PDB) code 1klo) (Stetefeld et al., 1996), and the human tumor necrosis receptor (PDB code 1ext) (Naismith et al., 1995) were obtained from the Protein Data Bank (Bernstein et al., 1977). Amino acid sequences of the human orthologues of the insulin receptor (INSR_HUMAN), EGF receptor (EGFR_HUMAN), ErbB-2 (ERB2_HUMAN), ErbB-3 (ERB3_HUMAN), and ErbB-4 (ERB4_HUMAN) were obtained from the SWISS-PROT database. The human ErbB-4 amino acid sequence was obtained from the Genpept database using the code HUMRETYKIN_1 prior to its appearance in the SWISS-PROT database. Both databases were accessed using the sequence retrieval system (SRS) (Etzold & Argos, 1993). Superimposition of structural fragments and construction of sequence alignments were performed using the InsightII program (Molecular Simulations Inc., San Diego, California). The C_{α} atoms of residues were used in the superimpositions.

Due to the low sequence identity between the EGF receptor and the IGF-1R, the sequence alignments presented in this paper were constructed predominantly by manual means. Several short regions of ungapped sequence alignment between insulin receptor homologues and EGF receptor homologues were identified by the computer program MACAW (Schuler et al., 1991).

Model building

Comparative models were generated using the MODELLER program (Šali & Blundell, 1993) as implemented in the InsightII software package (Molecular Simulations Inc.). In MODELLER, spatial and stereochemical restraints derived from the template structures and the sequence alignment are combined to form an objective function; model structures are created by optimizing this function. Parts of the IGF-1R (Garrett et al., 1998) were used as templates for the construction of the model of the EGF receptor L1 and S1 domains and of the L2 and S2 domains, as shown in Figure 1.

Preliminary models for the L1 and S1 domains of the EGF receptor showed deviation of the backbone from the IGF-1R structure between modules 5 and 8 of the CRD when the additional templates were used in the modeling. To constrain the backbone of the model to follow its major template, additional C_{α} – C_{α} distance restraints were generated as follows. The smaller IGF-1R templates were superimposed onto the main IGF-1R template structure by structural superimposition of the C_{α} atoms from residues that were aligned in the sequence alignments (Fig. 1A). From the residues of the templates that are shown as underlined in Figure 1A, coordinates for the C_{α} atoms of the analogous EGF receptor residues were generated using the Homology module of InsightII. A MODELLER script was written to generate restraints for each pair of C_{α} – C_{α} atoms spaced less than 50 Å apart. Each restraint function was of the form of a Gaussian function, whose peak occurs at the C_{α} – C_{α} distance of the given atom pair and had a σ value of 2 Å. These restraints were used in addition to the restraints generated by MODELLER. (This script will be made available on the World Wide Web at <http://modelling.ludwig.edu.au/>.)

Additional restraints were also required in the EGF receptor ectodomain L2–S2 model to constrain its third and sixth modules

and their successive linkers to adopt backbone structures similar to those of their templates. Restraints for EGF receptor modules 1–3 of the S2 domain were generated by using the same procedure as described above with the IGF-1R residues underlined in Figure 1B. A second set of restraints was generated for modules 4–6 for the residues also underlined in Figure 1B. For both sets of restraints, the functional form was that of a Gaussian curve with a σ value of 1 Å, and whose peak occurred at the C $_{\alpha}$ –C $_{\alpha}$ distance for the given residue pair. These restraints were used in addition to those generated by MODELLER for the L2–S2 model.

From each of the two sequence alignments, the MODELLER program generated 50 structures. An initial structure was generated from the sequence alignment. For each of the 50 resultant structures, the x-, y-, and z-coordinates of each atom were perturbed by the addition of a random number between –4 and 4 Å as taken from a uniform distribution. The structure was then subject to optimization of the objective function using the “refine1” option in the MODELLER program. Of the 50 models, the two structures with the lowest objective function score were each subjected to 20 rounds of further optimization using the “refine1” option. The structure from each of the two sequence alignments (Fig. 1) with the lowest objective function score was retained.

The side-chain conformation of Arg84 in the L1–S1 model significantly deviated from its analogous residue in the IGF-1R structure, Arg77, and failed to make hydrogen bonds with the relevant main-chain carbonyl oxygen atoms in the S1 domain. The “loop” routine was used to adjust the side chain of Arg84. Distance and dihedral restraints were used to adjust the conformation of this residue to conserve the hydrogen bonds it makes to the S1 domain. The dihedral restraints were of the form of Gaussian functions with a σ value of 10°. The values of the angles were chosen to be the multiple of 30° nearest to the equivalent dihedral angle in Arg77 of the IGF-1R. Two distance restraints in the form of right-handed half-Gaussian functions were used to restrain the hydrogen bond distances from Arg84.NH1 to Cys227.O and Arg84.NH2 to Arg228.O to values similar to those in the IGF-1R structure. The σ value of these functions was 0.05 Å.

Construction of insertion regions

Most of the insertion regions of the EGF receptor models were constructed using the “loop” routine of version 4 of MODELLER. Coordinates for each insertion were built using one of the two models obtained in the previous section as a scaffold. The regions of sequence for which coordinates were built in this manner were 1–5, 8–12, 16–23, 46–51, 101–107, 145–148, 184–191, 241–262, 319–328, 522–530, 540–546, 578–600, and 612–621. Coordinates for residues 351–368 and 387–393 were constructed simultaneously due to the proximity of these regions in the model of the L2 domain. For each insertion, 50 models were constructed. In cases where the generated loops with the lowest values of the objective function had similar backbone structures, the loop building process was considered to have converged and the coordinates of the loop replaced those of the same residues on the refined model. Where the loop structures did not converge, the structures with the three lowest MODELLER loop scores were evaluated using PROCHECK, ProsaII, and Profiles-3D. For several of these loops, the one with the second lowest MODELLER score was selected, as it had a more favorable Profiles-3D and ProsaII plot.

To retain certain secondary structures, additional restraints were used in the construction of some of the loops. Restraints with the

form of a right-handed half-Gaussian function with a σ value of 0.05 Å were used to hold selected main-chain N–O distances to 3.0 Å or less. The atom pairs for which this additional restraint was added were: Gln139.N–Gln184.OE1, Val268.N–Tyr261.O, Val268.O–Tyr261.N, Ser506.N–Ser529, Ile562.N–His591.O, and Glu578.N–Val592.O.

Structure validation

The two final models and many of the preliminary models were subject to evaluation by one or more computer programs. PROCHECK 3.0 (Laskowski et al., 1993) was used to examine the stereo-chemical quality of the models. Profiles-3D (Bowie et al., 1991; Lüthy et al., 1992) as implemented in InsightII and ProsaII (Hendlich et al., 1990; Sippl, 1993) were used to evaluate the validity of the protein structure. A window size of 10 was used for Profiles-3D, and window sizes of 11 and 51 were used for ProsaII.

Calculation of structural properties

Each of the models was protonated prior to the calculations using the HBUILD routine of CHARMM (Brooks et al., 1983). Standard protonation states at pH 7 were used for all side chains. Radii of gyration and accessible surface areas were calculated using the CHARMM program (Brooks et al., 1983).

Acknowledgments

We gratefully acknowledge Assoc. Prof. Ed Nice and Ms. Teresa Domagala for many fruitful discussions about the ligand binding and dimerization of the EGF receptor. We thank Dr. Peter Colman for providing his encouragement for this project, his suggestion of a confidence level for the residues of the models and for access to computer facilities, and Mr. John McFarlane for computer maintenance. One of us (R.N.J.) was supported in part by an Australian Postgraduate Research Award.

References

- Abe Y, Odaka M, Inagaki F, Lax I, Schlessinger J, Kohda D. 1998. Disulfide bond structure of human epidermal growth factor receptor. *J Biol Chem* 273:11150–11157.
- Alroy I, Yarden Y. 1997. The ErbB signaling network in embryogenesis and oncogenesis: Signal diversification through combinatorial ligand-receptor interactions. *FEBS Lett* 410:83–86.
- Anzano MA, Roberts AB, Smith JM, Sporn MB, De Larco JE. 1983. Sarcoma growth factor from conditioned medium of virally transformed cells is composed of both type α and type β transforming growth factors. *Proc Natl Acad Sci USA* 80:6264–6268.
- Bajaj M, Waterfield MD, Schlessinger J, Taylor WR, Blundell T. 1987. On the tertiary structure of the extracellular domains of the epidermal growth factor and insulin receptors. *Biochim Biophys Acta* 916:220–226.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bowie JU, Lüthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164–170.
- Brooks B, Brucoleri R, Olafson B, States D, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and molecular dynamics calculations. *J Comput Chem* 4:187–217.
- Burden S, Yarden Y. 1997. Neuregulins and their receptors: A versatile signaling module in organogenesis and oncogenesis. *Neuron* 18:847–855.
- Carpenter G, Cohen S. 1990. Epidermal growth factor. *J Biol Chem* 265:7709–7712.
- Degenhardt M, Weber W, Eschenburg S, Dierks K, Funari SS, Rapp G, Betzel C. 1998. Crystallization and preliminary X-ray crystallographic analysis of the EGF receptor ectodomain. *Acta Crystallogr D* 54:999–1001.

- Derynck R, Roberts AB, Winkler ME, Chen EY, Goeddel DV. 1984. Human transforming growth factor- α : Precursor structure and expression in *E. coli*. *Cell* 38:287–297.
- Domagala T, Konstantopoulos N, Smyth F, Jorissen RN, Fabri L, Geleick D, Lax I, Schlessinger J, Sawyer WH, Howlett GJ, et al. 2000. Stoichiometry, kinetics and binding equilibria of the interaction between epidermal growth factor (EGF) and the extracellular domain of the EGF receptor. *Growth Factors*. Forthcoming.
- Etzold T, Argos P. 1993. SRS an indexing and retrieval tool for flat file data libraries. *Comput Appl Biosci* 9:49–57.
- Fernandez R, Tabarini D, Azpiazu N, Frasch M, Schlessinger J. 1995. The *Drosophila* insulin receptor homolog: A gene essential for embryonic development encodes two receptor isoforms with different signaling potential. *EMBO J* 14:3373–3384.
- Garrett TP, McKern NM, Lou M, Frenkel MJ, Bentley JD, Lovrecz GO, Elleman TC, Cosgrove LJ, Ward CW. 1998. Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* 394:395–399.
- Gilson MK, Sharp KA, Honig BH. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J Comp Chem* 9:327–335.
- Grandis JR, Twardy DJ. 1993. TGF- α and EGFR in head and neck cancer. *J Cell Biochem Suppl* 17F:188–191.
- Greenfield C, Hiles I, Waterfield MD, Federwisch M, Wollmer A, Blundell TL, McDonald N. 1989. Epidermal growth factor binding induces a conformational change in the external domain of its receptor. *EMBO J* 8:4115–4123.
- Groenen LC, Nice EC, Burgess AW. 1994. Structure-function relationships for the EGF/TGF- α family of mitogens. *Growth Factors* 11:235–257.
- Gullick WJ, Marsden JJ, Whittle N, Ward B, Bobrow L, Waterfield MD. 1986. Expression of epidermal growth factor receptors on human cervical, ovarian, and vulvar carcinomas. *Cancer Res* 46:285–292.
- Günther N, Betzel C, Weber W. 1990. The secreted form of the epidermal growth factor receptor. Characterization and crystallization of the receptor-ligand complex. *J Biol Chem* 265:22082–22085.
- Haeder M, Rotsch M, Bepler G, Hennig C, Havemann K, Heimann B, Moelling K. 1988. Epidermal growth factor receptor expression in human lung cancer cell lines. *Cancer Res* 48:1132–1136.
- Harte MT, Gentry LE. 1995. Mutations within subdomain II of the extracellular region of epidermal growth factor receptor selectively alter TGF α binding. *Arch Biochem Biophys* 322:378–389.
- Hendlich M, Lackner P, Weitzkus S, Floeckner H, Froschauer R, Gottsbacher K, Casari G, Sippl MJ. 1990. Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J Mol Biol* 216:167–180.
- Huang HS, Nagane M, Klingbeil CK, Lin H, Nishikawa R, Ji XD, Huang CM, Gill GN, Wiley HS, Cavenne WK. 1997. The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *J Biol Chem* 272:2927–2935.
- Humphrey PA, Gangarosa LM, Wong AJ, Archer GE, Lund-Johansen M, Bjerkvig R, Laerum OD, Friedman HS, Bigner DD. 1991. Deletion-mutant epidermal growth factor receptor in human gliomas: Effect of type II mutation on receptor function. *Biochem Biophys Res Commun* 178:1413–1420.
- King AC, Cuatrecasas P. 1983. Resolution of high and low affinity epidermal growth factor receptors. Inhibition of high affinity component by low temperature, cycloheximide, and phorbol esters. *J Biol Chem* 257:3053–3060.
- King CR, Kraus MH, Williams LT, Merlino GT, Pastan IH, Aaronson SA. 1985. Human tumor cell lines with EGF receptor gene amplification in the absence of aberrant sized mRNAs. *Nucleic Acids Res* 13:8477–8486.
- Kline TP, Brown FK, Brown SC, Jeffs PW, Kopple KD, Mueller L. 1990. Solution structures of human transforming growth factor α derived from ^1H NMR data. *Biochemistry* 29:7805–7813.
- Kohda D, Inagaki F. 1992. Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions. *Biochemistry* 31:11928–11939.
- Kohda D, Odaka M, Lax I, Kawasaki H, Suzuki K, Ullrich A, Schlessinger J, Inagaki F. 1993. A 40-kDa epidermal growth factor/transforming growth factor α -binding domain produced by limited proteolysis of the extracellular domain of the epidermal growth factor receptor. *J Biol Chem* 268:1976–1981.
- Korc M, Chandrasekar B, Yamanaka Y, Friess H, Buchier M, Beger HG. 1992. Overexpression of the epidermal growth factor receptor in human pancreatic cancer is associated with concomitant increases in the levels of epidermal growth factor and transforming growth factor α . *J Clin Invest* 90:1352–1360.
- Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Kraus MH, Issing W, Miki T, Popescu NC, Aaronson SA. 1989. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: Evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci USA* 86:9193–9197.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
- Lax I, Bellot F, Honegger AM, Schmidt A, Ullrich A, Givol D, Schlessinger J. 1990. Domain deletion in the extracellular portion of the EGF-receptor reduces ligand binding and impairs cell surface expression. *Cell Regul* 1:173–188.
- Lax I, Bellot F, Howk R, Ullrich A, Givol D, Schlessinger J. 1989. Functional analysis of the ligand binding site of EGF-receptor utilizing chimeric chicken/human receptor molecules. *EMBO J* 8:421–427.
- Lax I, Fischer R, Ng C, Segre J, Ullrich A, Givol D, Schlessinger J. 1991a. Noncontiguous regions in the extracellular domain of EGF receptor define ligand-binding specificity. *Cell Regul* 2:337–345.
- Lax I, Mitra AK, Ravera C, Hurwitz DR, Rubinstein M, Ullrich A, Stroud RM, Schlessinger J. 1991b. Epidermal growth factor (EGF) induces oligomerization of soluble, extracellular, ligand-binding domain of EGF receptor. A low resolution projection structure of the ligand-binding domain. *J Biol Chem* 266:13828–13833.
- Layton JE, Shimamoto G, Osslund T, Hammacher A, Smith DK, Treutlein HR, Boone T. 1999. Interaction of granulocyte colony-stimulating factor (G-CSF) with its receptor. Evidence that Glu¹⁹ of G-CSF interacts with Arg²⁸⁸ of the receptor. *J Biol Chem* 274:17445–17451.
- Lemmon MA, Bu Z, Ladbury JE, Zhou M, Pinchasi D, Lax I, Engelman DM, Schlessinger J. 1997. Two EGF molecules contribute additively to stabilization of the EGFR dimer. *EMBO J* 16:281–294.
- Lüthy R, Bowie JU, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.
- Malden LT, Novak U, Kaye AH, Burgess AW. 1988. Selective amplification of the cytoplasmic domain of the epidermal growth factor receptor gene in glioblastoma multiforme. *Cancer Res* 48:2711–2714.
- Merlino GT, Xu YH, Ishii S, Clark AJ, Semba K, Toyoshima K, Yamamoto T, Pastan I. 1984. Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells. *Science* 224:417–419.
- Messa C, Russo F, Caruso MG, Di Leo A. 1998. EGF, TGF- α and EGF-R in human colorectal adenocarcinoma. *Acta Oncol* 37:285–289.
- Morgan DO, Jarnagin K, Roth RA. 1986. Purification and characterization of the receptor for insulin-like growth factor I. *Biochemistry* 25:5560–5564.
- Moriai T, Kobrin MS, Hope C, Speck L, Korc M. 1994. A variant epidermal growth factor receptor exhibits altered type α transforming growth factor binding and transmembrane signaling. *Proc Natl Acad Sci USA* 91:10217–10221.
- Moscatello DK, Hologado-Madruga M, Godwin AK, Ramirez G, Gunn G, Zolnick PW, Biegel JA, Hayes RL, Wong AJ. 1995. Frequent expression of a mutant epidermal growth factor receptor in multiple human tumors. *Cancer Res* 55:5536–5539.
- Naismith JH, Devine TQ, Brandhuber BJ, Sprang SR. 1995. Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor. *J Biol Chem* 270:13303–13307.
- Nicholls A, Honig BH. 1991. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comp Chem* 12:435–445.
- Odaka M, Kohda D, Lax I, Schlessinger J, Inagaki F. 1997. Ligand-binding enhances the affinity of dimerization of the extracellular domain of the epidermal growth factor receptor. *J Biochem* 122:116–121.
- Peles E, Yarden Y. 1993. Neu and its ligands: From an oncogene to neural factors. *BioEssays* 15:815–824.
- Pimentel E. 1994. *Handbook of growth factors. Volume II: Peptide growth factors*. Boca Raton: CRC Press.
- Plowman GD, Culouscou JM, Whitney GS, Green JM, Carlton GW, Foy L, Neubauer MG, Shoyab M. 1993. Ligand-specific activation of HER4/p180^{erbB4}, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci USA* 90:1746–1750.
- Riese DJ II, Stern DF. 1998. Specificity within the EGF family/ErbB receptor family signaling network. *BioEssays* 20:41–48.
- Roach P, Zick Y, Formisano P, Accili D, Taylor SI, Gorden P. 1994. A novel human insulin receptor gene mutation uniquely inhibits insulin binding without impairing posttranslational processing. *Diabetes* 43:1096–1102.
- Šali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
- Schechter AL, Hung MC, Vaidyanathan L, Weinberg RA, Yang-Feng TL, Francke U, Ullrich A, Coussens L. 1985. The neu gene: An erbB-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science* 229:976–978.

- Schejter ED, Segal D, Glazer L, Shilo BZ. 1986 Alternative 5' exons and tissue-specific expression of the *Drosophila* EGF receptor homolog transcripts. *Cell* 46:1091–1101.
- Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins* 9:180–190.
- Sippl MJ. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins* 17:355–362.
- Smith JJ, Derynck R, Korc M. 1987. Production of transforming growth factor α in human pancreatic cancer cells: Evidence for a superagonist autocrine cycle. *Proc Natl Acad Sci USA* 84:7567–7570.
- Stetefeld J, Mayer U, Timpl R, Huber R. 1996. Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin γ 1 chain harboring the nidogen binding site. *J Mol Biol* 257:644–657.
- Sugawa N, Ekstrand AJ, James CD, Collins VP. 1990. Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. *Proc Natl Acad Sci USA* 87:8602–8606.
- Summerfield AE, Hudnall AK, Lukas TJ, Guyer CA, Staros JV. 1996. Identification of residues of the epidermal growth factor receptor proximal to residue 45 of bound epidermal growth factor. *J Biol Chem* 271:19656–19659.
- Tramontano A. 1998. Homology modeling with low sequence identity. *Methods* 14:293–300.
- Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. 1997. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci* 6:53–64.
- Tzahar E, Pinkas-Kramarski R, Moyer JD, Klapper LN, Alroy I, Levkowitz G, Shelly M, Henis S, Eisenstein M, Ratzkin BJ, et al. 1997. Bivalence of EGF-like ligands drives the ErbB signaling network. *EMBO J* 16:4938–4950.
- Ullrich A, Bell JR, Chen EY, Herrera R, Petruzzelli LM, Dull TJ, Gray A, Coussens L, Liao YC, Tsubokawa M, et al. 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature* 313:756–761.
- Ullrich A, Coussens L, Hayflick JS, Dull TJ, Gray A, Tam AW, Lee J, Yarden Y, Libermann TA, Schlessinger J, et al. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 309:418–425.
- Ullrich A, Schlessinger J. 1990. Signal transduction by receptors with tyrosine kinase activity. *Cell* 61:203–212.
- Ward CW, Hoyne PA, Flegg RH. 1995. Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins* 22:141–153.
- Ward CW, Garrett TPJ, McKern, NM, Lawrence, LJ. 1999. Structure of the insulin receptor family: Unexpected relationships with other proteins. *Today's Life Sci* 11:26–32.
- Woltjer RL, Lukas TJ, Staros JV. 1992. Direct identification of residues of the epidermal growth factor receptor in close proximity to the amino terminus of bound epidermal growth factor. *Proc Natl Acad Sci USA* 89:7801–7805.
- Wu D, Wang L, Chi Y, Sato GH, Sato JD. 1990. Human epidermal growth factor receptor residue covalently cross-linked to epidermal growth factor. *Proc Natl Acad Sci USA* 87:3151–3155.
- Wu DG, Wang LH, Sato GH, West KA, Harris WR, Crabb JW, Sato JD. 1989. Human epidermal growth factor (EGF) receptor sequence recognized by EGF competitive monoclonal antibodies. Evidence for the localization of the EGF-binding site. *J Biol Chem* 264:17469–17475.
- Yarden Y, Schlessinger J. 1987. Self-phosphorylation of epidermal growth factor receptor: Evidence for a model of intermolecular allosteric activation. *Biochemistry* 26:1434–1442.
- Yoder MD, Keen NT, Jurnak F. 1993. New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science* 260:1503–1507.