

Structure function relationships in EGF, TGF- α and IGFI

IAIN D. CAMPBELL AND ROBERT M. COOKE*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Summary

The solution structures of the homologous growth factors hEGF and hTGF- α , have been determined independently from high resolution nuclear magnetic resonance (NMR) data. A model of the insulin-like growth factor structure based on insulin coordinates (Blundell *et al.* (1978) *Proc natn. Acad. Sci. U.S.A.* **75**, 180–184), has also been refined using molecular dynamics simulations with NMR-determined restraints. Knowledge of these structures, together with known sequences of other homologous proteins and experiments with site-specific residue changes, allows predictions to be made about growth factor residues which might be involved in the receptor–ligand interfaces.

Introduction

There is considerable pharmaceutical interest in mapping the parts of growth factors which interact with their receptors; this, in principle, could lead to the rational design of growth factor agonists and antagonists. The ultimate aim would be a complete structure of the growth factor–receptor complex but this is not likely to be available for some time. An alternative is to combine knowledge of growth factor structure with experiments on the receptor-binding properties of growth factor analogues, *e.g.* peptide fragments and various site-specific mutations. Often, however, there is no knowledge about the structure of the variant protein produced, although it is clearly important to know whether any observed change in biological activity is brought about by changes in local or global protein structure.

We have been studying the structures of three growth factors which have not proved amenable to X-ray crystallography because they do not crystallize readily. We have used high resolution nuclear magnetic resonance (NMR) to study recombinant human epidermal growth factor (hEGF), its homologue, transforming growth factor- α (hTGF- α) and insulin-like growth factor (IGFI). We are also beginning to investigate the effect of site-specific changes on the conformation of the proteins. Some of our recent work on hEGF, hTGF- α and IGFI will be briefly described and the implications for predictions and experiments to identify receptor–ligand interfaces will be outlined.

* Present address: Glaxo Group Research, Greenford Road, Greenford, Middlesex UB6 OHE, UK.

Key words: EGF, TGF- α , IGF, NMR, solution structure, site specific mutagenesis, receptor binding.

Structures

a) hEGF and hTGF- α

EGF and TGF- α are members of a family of homologous polypeptides with three disulphide bonds, which bind to the EGF receptor (Burgess, 1989). Many sequences of this family are known, with an overall amino-acid sequence homology of about 30%. In addition, there are many modules or domains of extracellular proteins which have sequences homologous to EGF (for recent surveys of the EGF family see, for example, Campbell *et al.* 1989, 1990; Burgess, 1989; Shoyab *et al.* 1989).

The structures of hEGF and hTGF- α have been determined independently in this laboratory (Cooke *et al.* 1987; Tappin *et al.* 1989; Campbell *et al.* 1989, 1990) using high resolution ^1H NMR and computer-based methods which have become established in recent years (Wüthrich, 1989; Cooke and Campbell, 1988). The molecule can be considered to consist of two domains, an N-terminal domain (1–32) and a C-terminal domain (32–53). The dominant motif is a double stranded β -sheet formed between residues 18 to 33. Three disulphide bonds radiate (up) from one face of this platform. The N-terminal strand is weakly associated to the main sheet. There is also a short anti-parallel β -sheet in the C-terminal domain. There are a number of loops and turns and intimate contacts between the N- and C-terminal domains, especially between the loop around positions 13–16 and the turn around positions 40–43.

There is good agreement between the various NMR studies which have been carried out on human, mouse and rat EGF and TGF- α (Cooke *et al.* 1987; Tappin *et al.* 1989; Campbell *et al.* 1989, 1990; Montelione *et al.* 1987; Mayo *et al.* 1989; Kohda *et al.* 1988). There is some evidence that the structures of the EGF family are relatively mobile compared to inhibitor proteins like bovine trypsin inhibitor. In both hEGF and TGF- α , few nuclear Overhauser effects (NOEs) are observed for the N- and C-terminal residues. This lack of information leads to a wider variation within the families of calculated structures in these regions. Much of this variation is probably a reflection of protein mobility, and this is supported by the observation of relatively narrow resonances from these parts of the molecule. These results, together with relatively fast amide exchange rates, suggest that the EGF family are relatively flexible molecules which are susceptible to conformational interconversions on a millisecond time scale.

b) IGFI

IGFI, which is highly homologous to insulin and IGFII, has 70 residues. The high degree of homology with insulin led to the production of a model of IGFI based on the X-ray structure of insulin (Blundell *et al.* 1978).

In a recent study we have assigned the NMR spectrum of IGFI and 344 distant restraints were obtained from NOE data. The NOE data were incorporated into restrained molecular dynamics simulations of the IGFI structure. This resulted in a refined structure which, although largely similar to the model, has some

significant differences especially in the side chain orientations. Like the model, residues 3–6 adopt a β -strand-like conformation, the first helix is between residues 8–17, residues 23–26 are defined but extended, the second helix encompasses residues 44–49 and the third helix runs from residues 54–59. The protein has three disulphide bridges, Cys6–Cys48 is on the surface, while Cys18–Cys61 and Cys47–52 are buried in a hydrophobic protein core.

Some parts of the protein are well defined by the NMR restraints although other parts are not. The average root mean square deviation (RMSD) for the backbone atoms of residues 3–19, 22–26 and 43–61 was 0.17 nm in the various molecular dynamics simulations; if all parts of the protein are considered, this figure rises to 0.51 nm. As with the EGF family there is quite strong evidence, from the NMR spectra, that the ill-defined regions undergo considerable conformational flexibility.

The receptor binding surfaces

Now that structural information about these growth factors is available, it is useful to combine this with information from sequences and receptor-binding studies on analogues to predict the parts of the growth factors which might be in contact with their receptors.

a) EGF and TGF- α

By comparing the sequences of different polypeptides which are known to bind to the EGF receptor with those which form the growth factor-like modules in a wide variety of mosaic proteins, we have previously predicted residues on the EGF molecule which might bind to the receptor (Campbell *et al.* 1989, 1990). This was done on the assumption that the structures of all these members of the EGF family are similar and that the EGF modules do not bind to the receptor. These assumptions have been borne out to some extent by our experiments, since although EGF and TGF- α only share 40% of residues their NMR structures are very similar. In addition, recent structural and calcium-binding studies on an EGF module from factor IX produced by recombinant techniques (Handford *et al.* 1990) show that while the structures of EGF and the EGF module are similar, the receptor binding activity was less than 1000 times that observed for EGF itself.

Residues 6, 14, (16), 18, 20, 31, 33, 36, 37, 39, 42 and (43) were predicted to be important for the integrity of the EGF structure while residues 13, 15, 41 and 47 were at the EGF/EGF receptor interface (the brackets indicate uncertainty about structural or functional roles) (Campbell *et al.* 1989, 1990). Although Y13, L15 and R41 are in different loops they are close to each other in space at the domain interface. L47, on the other hand, is significantly separated from the other three.

These predictions can be checked since a large number of studies have been carried out on the binding and mitogenic activity of variants of the EGF and TGF- α structures (Heath and Merrifield, 1986; Engler *et al.* 1988; Defeo-Jones *et*

al. 1988). These studies are largely, although not entirely, consistent with the identification of residues 13, 15, 41 and 47 as interface residues. One difficulty is that in some cases a mutation might affect the overall structure of the molecules rather than a local change. One way to check this is to use high resolution NMR. We chose to do this with hEGF and have been producing the 1–52 wild type molecule with several mutations at the ‘structural’ and ‘interface’ sites defined above. One of the best studied residues in the EGF family is L47. Several studies have shown that deletion or change at this site seriously affects receptor binding. We have recently carried out receptor binding and NMR studies of four L47 mutants (V, A, D and E) (Dudgeon *et al.* 1990). In receptor binding assays, comparisons with wild type hEGF showed that L47 V bound approximately seven times more weakly while L47A, L47E and L47D bound approximately 50 times more weakly. These data are consistent with those of other groups (e.g. Defeo-Jones *et al.* 1988; Engler *et al.* 1988; Moy *et al.* 1989). We also carried out detailed 1D and 2D NMR studies and observed no major changes in structure, although minor effects were observed in the C-terminal region of the molecule. These results confirm the notion that L47 is a receptor interface residue and that it is not very important for the overall structure of the EGF molecule. In other cases, however, e.g. R41H, significant structural as well as receptor binding changes have been observed (Hommel, U., Cooke, R. M., Dudgeon, T. and Campbell, I. D., unpublished data).

b) IGF1

The IGFs are peculiar in that they bind to more than one receptor. Two distinct IGF receptors, type 1 and type 2, are found in many cell lines. In addition IGF1 binds to the insulin receptor and a number of serum-binding proteins (Czech, 1989). Structural information about IGF1 is sparse although much is known about its homologue insulin (Baker *et al.* 1988).

A number of modified IGF molecules have been investigated. The most extensive recent studies have been carried out by Cascieri and Bayne (1990). They investigated the binding of IGF1, and many variants, to the type 1 IGF receptor (from human placenta) the type 2 receptor (from rat liver), the insulin receptor (from human placenta) and human serum-binding proteins. They found analogues which retained binding to some of these receptors but lost activity in others. This implies that the structure of these proteins is not significantly changed, at least in some regions, by the amino acid substitutions.

It is of interest to identify, on our refined structure, the regions of IGF1 which these studies suggest to be involved in binding to the various receptors. In summary these are: residues 1–3 and 49–51 for the binding proteins; 1, 2, 8, 9, 12, 49–51 for the type 2 receptor; 21, 23–25, 42–44, 46, 60, 62 for the insulin receptor (this list is also based on insulin results, Baker *et al.* 1988). There is overlap between the type 2 receptor and the binding protein surface and considerable overlap between the insulin and type 1 receptors. These two regions of overlap are on opposite sides of the IGF1 molecule. Now that the NMR spectrum of IGF1 has

been assigned we are in a position to investigate, in more detail, some of the structural properties of the variant IGFs that are available. This should allow these regions to be defined more precisely.

Conclusions

The NMR method for determining protein structure has been shown to be viable for the growth factors EGF, TGF- α and IGFI. The structures of all three seem to be relatively flexible compared, for example, to a trypsin inhibitor, and sensitive to solution conditions such as pH and temperature. The growth factor-receptor interfaces also seem to involve relatively large patches on the surface of the growth factors with residues from widely different parts of the sequence. These observations imply that it may be rather difficult to make a stable small agonist or antagonist for these growth factors.

This is a contribution from the Oxford Centre for Molecular Science which is supported by SERC and MRC. We also thank ICI Pharmaceuticals, British Biotechnology and Monsanto for their financial and technical support. The work owes much to many colleagues in Oxford, including Martin Baron, Jonathan Boyd, Tim Harvey, Tim Dudgeon, Uli Hommel and Mike Tappin.

References

- BAKER, E. N., BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., CROWFOOT-HODGKIN, D. M., HUBBARD, R. E., ISAACS, N. W., RENOLDS, C. D., SAKABE, K., SAKABE, N. AND VIJAYAN, N. M. (1988). The structure of 2Zn pig insulin crystals at 1.5Å resolution. *Phil. Trans. R. Soc. Lond.* **319**, 369-456.
- BLUNDELL, T. L., BEDARKER, S., RINDERKNICHT, E. AND HUMBEL, R. E. (1978). Insulin-like growth factor: a model for tertiary structure accounting for immunoreactivity and receptor binding. *Proc. natn. Acad. Sci. U.S.A.* **75**, 180-184.
- BURGESS, A. W. (1989). Epidermal growth factor and transforming growth factor- α . *Br. Med. Bull.* **45**, 401-424.
- CAMPBELL, I. D., COOKE, R. M., BARON, M., HARVEY, T. S. AND TAPPIN, M. J. (1989). The solution structures of EGF and TGF- α . *Prog. in Growth Factor Research* **1**, 13-22.
- CAMPBELL, I. D., BARON, M., COOKE, R. M., DUDGEON, T. J., FALLON, A., HARVEY, T. S. AND TAPPIN, M. J. (1990). Structure function relationships in EGF and TGF- α . *Biochem. Pharmac.* in press.
- CASCIERI, M. A. AND BAYNE, M. L. (1989). Identification of the domains of IGFI which interact with the IGF receptors and binding proteins. In *Molecular and Cellular Biology and Insulin-like Growth Factors and their Receptors* (ed. leRoith, D. and Raizada, M. K.) pp. 285-296, Plenum Press, New York.
- COOKE, R. M. AND CAMPBELL, I. D. (1988). Protein structure determination by NMR. *Bioessays* **8**, 52-56.
- COOKE, R. M., WILKINSON, A. J., BARON, M., PASTORE, A., TAPPIN, M. J., CAMPBELL, I. D., GREGORY, H. AND SHEARD, B. (1987). The solution structure of human epidermal growth factor. *Nature* **327**, 339-341.
- CZECH, M. P. (1989). Signal transmission by the insulin-like growth factors. *Cell* **59**, 235-238.
- DEFEO-JONES, D., TAI, J. Y., WEGRZYN, R. J., VUOCOLO, G. A., BAKER, A. E., PAYNE, L. S., GARSKY, V. M., OLIFF, A. AND RIEMEN, M. W. (1988). Structure-function analysis of synthetic and recombinant derivatives of TGF- α . *Molec. cell Biol.* **8**, 2999-3007.
- DUDGEON, T. J., BARON, M., COOKE, R. M., CAMPBELL, I. D., EDWARDS, R. M. AND FALLON, A. (1990). Structure and function of hEGF: receptor binding and NMR. *FEBS Lett.* **261**, 392-396.
- ENGLER, D. A., MATSUNAMI, R. K., CAMPION, S. R., STRINGER, C. D., STEVENS, A. AND NIYOGI, S.

- K. (1988). Cloning of authentic human epidermal growth factor as a bacterial secretory protein and its initial structure-function analysis by site directed mutagenesis. *J. biol. Chem.* **263**, 12 384–12 390.
- HANDFORD, P., BARON, M., MAYHEW, M., WILLIS, A., BEESLEY, T., BROWNLEE, G. G. AND CAMPBELL, I. D. (1990). The first EGF domain of Factor IX has a high affinity calcium binding site. *EMBO J.* **9**, 475–480.
- HEATH, W. F. AND MERRIFIELD, R. B. (1986). A synthetic approach to structure-function relationships in the murine EGF molecule. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6367–6371.
- KOHDA, D., SHIMIDA, I., MIYAKE, T., FUWA, T. AND INAGAKI, F. (1988). Polypeptide chain fold of hTGF- α analogous to those of mouse and human epidermal growth factors as studied by 2D ^1H NMR. *Biochemistry* **28**, 953–958.
- MAYO, K. H., CABALLI, R. C., PETERS, A. R., BOELENS, R. AND KAPTEIN, R. (1989). Sequence specific ^1H NMR assignments and peptide backbone conformation in rat EGF. *Biochem. J.* **257**, 197–205.
- MONTELIONE, G. T., WÜTHRICH, K., NICE, E. C., BURGESS, A. W. AND SHERAGA, H. A. (1987). Solution structure of murine EGF; determination of the polypeptide backbone chain-fold by NMR and distance geometry. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5226–5230.
- MOY, F. L., SHERAGA, H. A., LIU, J.-F., WU, R. AND MONTELIONE, G. T. (1989). Conformational characterization of a single-site mutant of murine epidermal growth factor (EGF) by ^1H NMR provides evidence that leucine-47 is involved in the interactions with the EGF receptor. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9836–9840.
- SHOYAB, M., PLOWMAN, G. D., McDONALD, V. L., BRADLEY, J. G. AND TODARO, G. J. (1989). Structure and function of human amphiregulin: a member of the EGF family. *Science* **243**, 1074–1076.
- TAPPIN, M. J., COOKE, R. M., FITTON, J. AND CAMPBELL, I. D. (1989). A high resolution H NMR study of hTGF- α : structure and pH dependent conformational interconversion. *Eur. J. Biochem.* **179**, 629–637.
- WÜTHRICH, K. (1989). Protein structure determination in solution by NMR. *Science* **243**, 45–50.