Biochemical Pharmacology, Vol. 40, No. 1, pp. 35–40, 1990.
Printed in Great Britain.

0006-2952/90 $3.00 + 0.00

STRUCTURE–FUNCTION RELATIONSHIPS IN EPIDERMAL GROWTH FACTOR (EGF) AND TRANSFORMING GROWTH FACTOR-ALPHA (TGF-α)

IAIN D. CAMPBELL,* MARTIN BARON, ROBERT M. COOKE, TIMOTHY J. DUDGEON, ANTHONY FALLON,† TIMOTHY S. HARVEY and MICHAEL J. TAPPIN

Department of Biochemistry, University of Oxford, Oxford OX1 3QU; and †British Biotechnology, Cowley, Oxford OX4 5LY, U.K.

Abstract—The solution structures of the homologous growth factors human epidermal growth factor (hEGF) and human transforming growth factor-alpha (hTGF-α), as determined by high resolution NMR and various computational methods, are described. Knowledge of these structures and the sequences of other homologous proteins leads to predictions about growth factor residues which may be involved in the receptor/ligand interface. Recent experiments designed to check these predictions are described briefly. These involve site-specific mutagenesis, receptor binding assays and high resolution NMR studies.

Epidermal growth factor (EGF‡) and transforming growth factor-alpha (TGF-α) are members of a family of homologous polypeptides with three disulphide bonds, which bind to the EGF receptor and cause cell division. Many sequences of this family are now known including those of EGFs from human, rat, mouse and guinea pig; TGF-αs from human and rat; several viral proteins; and human amphiregulin. The overall amino-acid sequence homology among this family is about 30%. There are also many modules or domains of extracellular proteins which have sequences homologous to EGF (for recent surveys of the EGF family see, for example, Refs. 1–3) (see also Fig. 1).

It has proven difficult to crystallize members of this family, but a number of structural studies have been carried out recently using high resolution NMR combined with computer-based techniques [4–10].

EGF has clinical potential for wound healing and anti-tumour treatment. Thus, there have been numerous attempts to define the residues involved at the receptor/growth factor interface with a view to designing agonists or antagonists. A large number of variants of EGF and TGF-α have been tested for activity; these include various peptide fragments which have been synthesized and produced by cleavage as well as various site-specific mutations of recombinant growth factors [11–16]. A general problem with most of these studies is lack of knowledge about the structure of the variant protein produced. It is clearly important to know whether any observed change in biological activity or binding is brought about by changes in local or global protein structure.

In this paper some of our recent work on the solution structures of hEGF and hTGF-α will be described briefly. Recently we have produced hEGF and several mutants using a yeast expression system. Preliminary experiments on the structure and receptor binding properties of hEGF and its variants will be discussed.

Structures of hEGF and TGF-α

The structures of hEGF and hTGF-α have been determined independently in this laboratory using high resolution ^{1}H NMR and computer-based methods which have become established in recent years [17, 18].

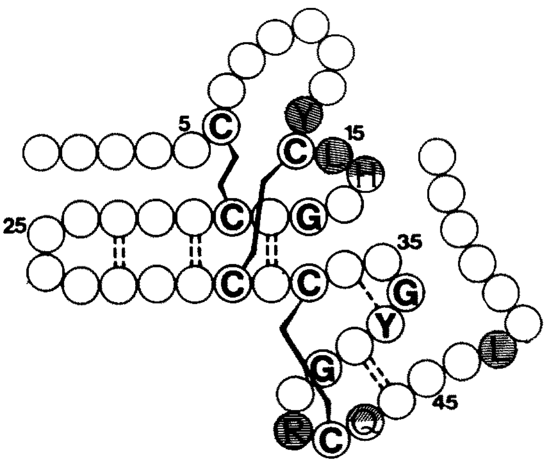
Some of the features of the EGF structure are shown in a schematic way in Fig. 1a, and they can be described as follows: The dominant motif is a double-stranded β-sheet formed between residues 18 and 33 (the EGF numbering will be used throughout this paper; translation to other numbering schemes can be made by referring to Fig. 1, b and c). 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 antiparallel β-sheet in the C-terminal domain. There are a number of loops and turns: an Ω loop between residues 6 and 13, a turn between residues 15 and 18, a type I β-hairpin around residues 23–28 which points down from the main sheet, and a type II β-turn in the vicinity of residue 35. The molecule can be considered as consisting of two domains, an N-terminal domain (1–32) and a C-terminal domain (32–53). There are intimate contacts between these domains, especially between the loop around positions 13–16 and the turn around positions 40–43.

The NMR structural restraints are mainly in the form of nuclear Overhauser effects (NOEs) although coupling constants and amide exchange rates can sometimes be incorporated [17, 18]. We used the programs DISGEO [19] and DISMAN [20] to calculate initial structures, refining these using restrained molecular dynamics with the program GROMOS [21]. Such calculations give families of structures consistent with the NMR data. In the hEGF studies, 234 NOEs were used in the structure

* Correspondence: Dr. Iain Campbell, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

‡ Abbreviations: EGF, epidermal growth factor; TGF-α, transforming growth factor-alpha; hEGF, human epidermal growth factor; hTGF-α, human transforming growth factor-alpha; and NOE, nuclear Overhauser effect.

a)



b)

1 10 20 30 40 50

hEGF:	NSDSE	C	PLSHDG	YCLH	D	G	V	C	MYIEA---LDKYA	C	N	C	VV	GY	I	G	E	RCQ	YRD	L	KWWELR
mEGF:	NSYPG	C	PSSYDG	YCLN	G	G	V	C	MHIES---LDSYT	C	N	C	VI	GY	S	G	D	RCQ	TRD	L	RWWELR
rEGF:	NSNTG	C	PPSYDG	YCLN	G	G	V	C	MYVES---VDRYV	C	N	C	VI	GY	I	G	E	RCQ	HRD	L	R
gpEGF:	QDAPG	C	PPSHDG	YCLH	G	G	V	C	MHIES---LNTYA	C	N	C	VI	GY	V	G	E	RCE	HQD	L	DWE
hTGFα:	VVSHFND	C	PDSHTQ	FCFH	-	G	T	C	RFLVQ---EDKPA	C	V	C	HS	GY	V	G	A	RCE	HAD	L	LA
rTGFα:	VVSHFNK	C	PDSHTQ	YCFH	-	G	T	C	RFLVQ---EEKPA	C	V	C	HS	GY	V	G	V	RCE	HAD	L	LA
VVP:	..DIPAIRL	C	GPEGDG	YCLH	-	G	D	C	IHARD---IDGMY	C	R	C	SH	GY	T	G	I	RCQ	HVV	L	LVDYQRS
SFVP:	..IVKHVKV	C	NHDYEN	YCLN	N	G	T	C	FTIALDNVSITPF	C	V	C	RI	NY	E	G	S	RCQ	FIN	L	VTY
MVP:	..IIKRIKL	C	NDDYKN	YCLN	N	G	T	C	FTVALNNVSLNPF	C	A	C	HI	NY	V	G	S	RCQ	FIN	L	ITIK

c)

hFIX:	DGDQ	C	ESNP	C	LNG	G	S	C	KDDI--NSYE	C	W	C	PF	GF	E	G	KN	C	EL
hFX:	DGDQ	C	ETSP	C	QNQ	G	K	C	KDGL--GEYT	C	T	C	LE	GF	E	G	KN	C	EL
hFVII:	DGDQ	C	ASSP	C	QNG	G	S	C	KDQL--QSYI	C	F	C	LP	AF	E	G	RN	C	ET
Notch:	DIDE	C	QSNP	C	LND	G	T	C	HDKI--NGFK	C	S	C	AL	GF	T	G	AR	C	QI
uEGF-1:	NIDE	C	ASAP	C	QNG	G	I	C	IDGI--NGYT	C	S	C	PL	GF	S	G	DN	C	EN
hFXII:	ASQA	C	RTNP	C	LHG	G	R	C	LEVE--GHRL	C	H	C	PV	GY	T	G	PF	C	DV
hTPA:	PVKS	C	SEPR	C	FNG	G	T	C	QQALYFSDFV	C	Q	C	PE	GF	A	G	KC	C	EI

Fig. 1. (a) Schematic representation of the structure of the EGF family. The dashed lines represent observed secondary structure H-bonds. Residues which are conserved in both the growth factors and all the growth factor modules are shown with their single letter code in unshaded circles. Residues which are changed conservatively throughout the sequences are shown half-shaded and residues which are only conserved in the growth factor sequences are shown in shaded circles. (b) Some of the known sequences which bind to the EGF receptor. (c) Some of the growth factor modules found in mosaic proteins (see Ref. 2 for further details of these sequences).

calculations, whereas 383 were used for TGF-α. After calculation of the structures from these NMR restraints the overall fold of the molecules is observed to be very similar, although the comparative lack of data for hEGF means that the relative orientations of the N- and C-domains are less well defined than for TGF-α. Our structure of TGF-α is now very good in many regions. The typical r.m.s. deviation between families for the backbone residues of regions 5–9 and 15–47 is 0.09 nm.

It is desirable to have independent checks, whenever possible, of the structure. One way of doing this is to consider the rate of exchange of amide hydrogens with solvent. It is possible to correlate observed amide exchange rates with the fraction of time H-bonds are observed during a restrained molecular dynamics trajectory. (H-bond restraints were not incorporated in the restrained molecular dynamics simulations.) In hEGF, for example, the H-bonds with the longest lifetime are those between the second and third strands of the main β-sheet, and this correlates well with the observed rates of exchange from amides in this region. Another interesting example is the amide hydrogen of L15 in hEGF which is found to exchange slowly (also observed for mouse [5] and rat [10] EGF) but which cannot be accounted for by the secondary structure shown in Fig. 1a. The restrained molecular dynamics simulations show that the main H-bond formed by the L15 amide is to the R41 carbonyl, i.e. between the N- and C-terminal domains.

It is interesting to compare the results of the various NMR studies which have been carried out on human, mouse and rat EGF and hTGF-α [4–10]. On the whole there is very good agreement, and most discrepancies can be ascribed to differences in solution conditions. TGF-α, for example, undergoes a pH-dependent structural change below pH 6 in which, amongst a number of other effects, the N-terminal strand of the main β-sheet becomes dissociated from the other two strands [7]. Most of the assignments are in agreement between different

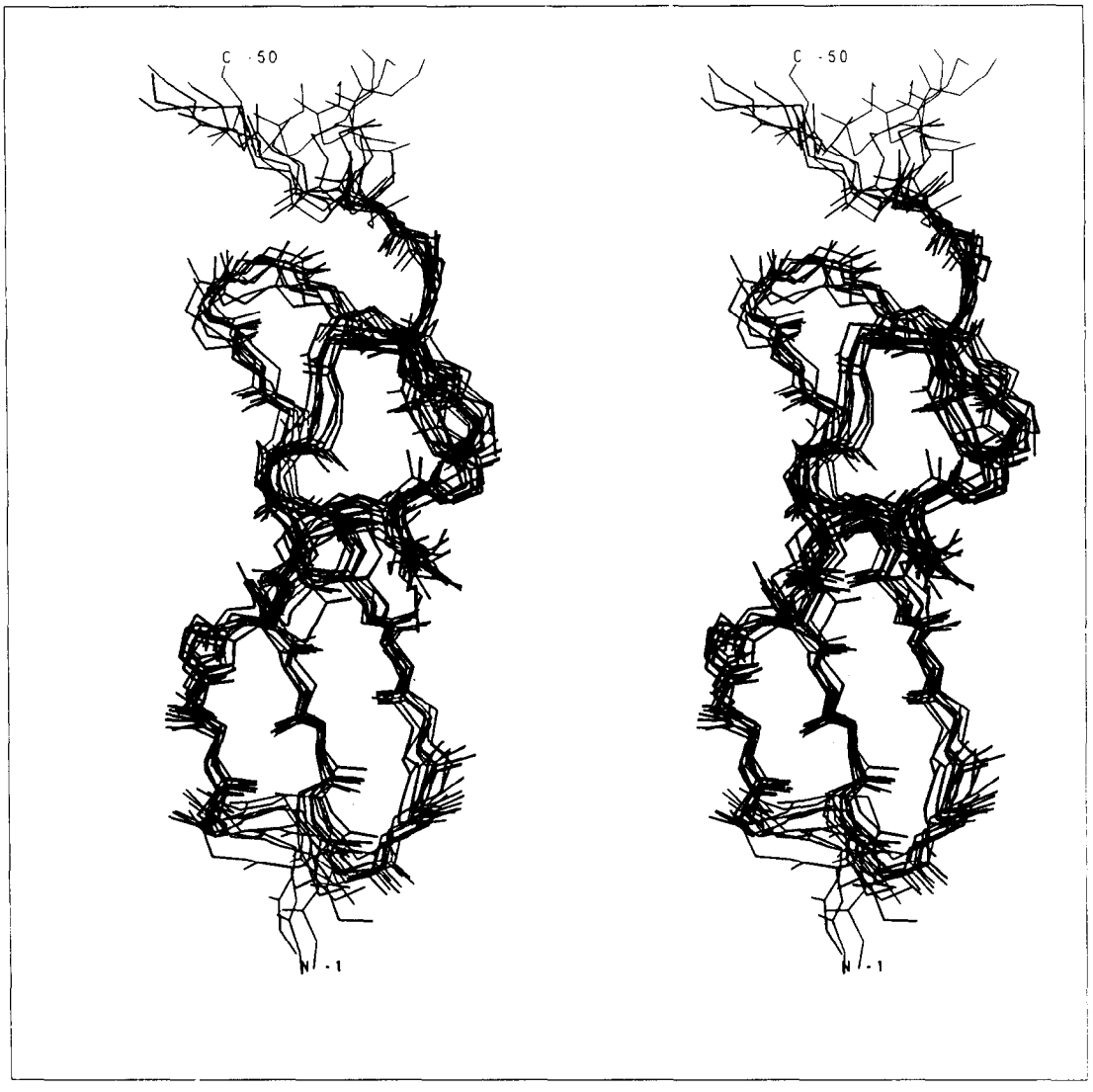


Fig. 2. Stereo view of ten overlaid structures, observed at 1 psec intervals in a restrained molecular dynamics simulation of hTGF-α. The program GROMOS was used, and the NMR distance restraints were applied with a relatively mild force constant (1000 kJ mol^{-1} nm^2).

studies although there has been some disagreement about the resonances of the R41β and γ protons, which have unusual chemical shifts in all the proteins studied. It is interesting to note that several residues have very similar and unusual shifts in all the growth factors. These shifts are probably a result of the protein fold, e.g. those of R41 mentioned above and the NH proton resonance of E/Q43 [7].

There is some evidence that the structures of the EGF family are relatively mobile compared to inhibitor proteins like bovine trysin inhibitor. In both hEGF and TGF-α, for example, few 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. This is illustrated in Fig. 2 where ten TGF-α structures, sampled from a restrained molecular dynamics trajectory, are overlaid. 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.

It was also mentioned above that the structure of TGF-α was slightly different at different pH values. We investigated this by completely assigning the resonances at pH values of 3.8, 6.5 and 9.4 [7]. Severe broadening of several amide and α-CH resonances is observed at low pH almost certainly because of a relatively slow exchange between different conformations. Several resonances from protons in the vicinity of position 42 are also observed to be broad in TGF-α [7], hEGF and EGF [5] at all pH values. These results together with relatively fast amide exchange rates suggest that the EGF family consists of relatively flexible molecules that are susceptible to conformational interconversions on a millisecond time scale. These mobility effects could

explain why the molecules have been difficult to crystallize.

Prediction of the receptor binding surface

Now that structural information is available, it is useful to combine this information with sequence information on the EGF family in an attempt to predict the part of the EGF structure which is in contact with the receptor.

Figure 1b shows some of the known amino acid sequences of the EGF growth factor family. Detailed binding studies have not been done for all of these proteins but many of them appear to bind to the EGF receptor with similar affinity, although there are considerable variations among the sequences, especially at the N- and C-termini.

Figure 1c shows some of the EGF-like sequences that have been observed in a wide variety of mosaic proteins, i.e. those which are built up from various modules [22]. There are two main types of growth factor module in mosaic proteins; only the type of module most similar to the growth factor structure is listed in Fig. 1c. Extensive testing of receptor binding by these modules has not been carried out but most of the intact proteins in which they appear do not have growth factor activity. A module from Factor IX, produced by peptide synthesis [23], was shown recently to have no detectable growth factor activity. We have carried out preliminary structural and calcium binding studies on a similar Factor IX EGF module [24] produced by recombinant techniques. These studies strongly suggest that the structures of EGF and the module are similar in the presence of calcium and that calcium does bind at physiological concentrations to the aspartate rich site which we predicted previously [4].

Making the assumptions that the structures of the modules are the same as the growth factor and that the modules do not bind to the receptor allows predictions about important residues at the growth factor/receptor interface to be made by comparing the sequences in Fig. 1, b and c [2].

Study of the sequences reveals that some residues are completely conserved throughout the series, namely the six cysteines and the glycines at positions 18 and 39. There are also some residues which are conservatively substituted, namely H/N at position 16, G/N/A at position 36, F/Y at position 37 and Q/E/D at position 43. (It should be noted that if other types of growth factor module are included in the comparisons, then some residues would be deleted from this list, e.g. residues 16 and 43 [2].) It is thus possible to consider residues 6, 14, (16), 18, 20, 31, 33, 36, 37, 39, 42 and (43) as important for the structural integrity of the EGF-like structure (the parentheses indicate uncertainty about structural or functional roles).

Comparison of the sequences reveals four further residues which are conserved in Fig. 1b but not in 1c: Y/F at position 13, L/F at position 15, R at position 41, and L at position 47. It seems plausible that these residues, shaded in Fig. 1a, may be involved at the receptor/growth factor interface. It is interesting to note that 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.

This model, in which positions 13, 15, 41 and 47 are at the growth factor/receptor interface, must be checked since it depends on certain assumptions about structural conservation between growth factor modules and growth factors which may be unjustified.

Effect of mutations on structure and receptor binding

A large number of studies have been carried out on the binding and mitogenic activity of variants of the EGF and TGF-α structures [11–16]. These studies are largely, although not entirely, consistent with the identification of residues 13, 15, 41 and 47 as interface residues. One of the problems is that, so far, only binding and mitogenic assays have been done; these assays are only reliable if done with considerable care, and some of the early results must be regarded as suspect. It is also easy to imagine that in some cases a mutation would affect the overall structure of the molecules rather than a local change. For example, in one of the more complete studies, Defeo-Jones *et al.* [16] showed that non-conservative changes at the residues equivalent to 13, 37 and 46 in TGF-α cause decreases in receptor binding. Position 13 is predicted as an interface residue by the above arguments. Position 37 is one of the residues predicted to be structurally important; thus it is perhaps not surprising that a change here should cause disruption of the overall structure. Position 46 is more difficult to explain since by the above arguments this was not classified as either structural or interface. This residue is conserved as an aspartate in EGFs and TGF-αs, but is not in the viral proteins. It is, however, likely that this residue is involved in the minor β-sheet of the C-terminal domain; thus it could be important for orienting L47 properly. In addition, the results obtained by different groups on modifying this residue are not entirely consistent (cf. Refs. 12 and 16).

From the above example, it is clear that it is important to check for structural changes produced by mutations. One way to do 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. A yeast alpha factor secretion system was used [25] to produce the protein. Synthetic hEGF genes were joined to the yeast alpha factor leader as described previously [26] in a yeast vector which was a derivative of pMA91 [27]. Cultures were harvested after 2–3 days of growth at 30°, and the hEGF was purified from cell free broth by reversed phase and ion exchange HPLC. Typically 2–4 mg of 95% pure hEGF was produced per litre of culture. As noted previously [28], we found that this secretion system produces the 1–52 form of hEGF but this protein was found to have the same affinity for the receptor as the 1–53 form. The radioreceptor assay was carried out with ^{125}I-labelled murine EGF on a monkey Vero cell line and A431 cells.

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

[e.g. Refs. 11–13, 16]. No structural studies have been carried out so far on these variant molecules. Recently we carried out receptor binding and NMR studies of four L47 mutants (V, A, D and E) [29].

In receptor binding assays, comparisons with wild type hEGF showed that L47V bound approximately seven times more weakly, while L47A, L47E and L47D bound approximately fifty times more weakly. These data are consistent with those of other groups [e.g. Ref. 11]. 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.

Conclusions

The NMR method for determining protein structure [17] has been shown to be viable for the EGF family since many groups around the world have found very similar structures for different members of this family. The structures seem to be relatively flexible and sensitive to solution conditions such as pH and temperature.

Knowledge of the structures, combined with receptor binding studies on a large number of fragments, homologous proteins and proteins with site-specific changes is allowing increasingly accurate predictions of the residues which form the interface between the growth factors and the receptor. The possibility that this knowledge can be applied to rational drug design is an interesting challenge for research.

Acknowledgements—This is a contribution from the Oxford Centre for Molecular Science which is supported by SERC and MRC. We also thank ICI and Monsanto for financial support.

REFERENCES

1. Burgess AW, Epidermal growth factor and transforming growth factor-α. *Br Med Bull* **45**: 401–424, 1989.
2. Campbell ID, Cooke RM, Baron M, Harvey TS and Tappin MJ, The solution structures of EGF and TGF-α. *Prog Growth Factor Res* **1** 13–22, 1989.
3. Shoyab M, Plowman GD, McDonald VL, Bradley JG and Todaro GJ, Structure and function of human amphiregulin: A member of the epidermal growth factor family. *Science* **243**: 1074–1076, 1989.
4. Cooke RM, Wilkinson AJ, Baron M, Pastore A, Tappin MJ, Campbell ID, Gregory H and Sheard B, The solution structure of human epidermal growth factor. *Nature* **327**: 339–341, 1987.
5. Montelione GT, Wüthrich K, Nice EC, Burgess AW and Scheraga HA, Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. *Proc Natl Acad Sci USA* **84**: 5226–5230, 1987.
6. Kohda D, Shimada I, Miyake T, Fuwa T and Inagaki F, Polypeptide chain fold of human transforming growth factor α analogous to those of mouse and human epidermal growth factors as studied by two-dimensional ^{1}H NMR. *Biochemistry* **28**: 953–958, 1988.
7. Tappin MJ, Cooke RM, Fitton J and Campbell ID, A high-resolution ^{1}H-NMR study of human transforming growth factor α: Structure and pH-dependent conformational interconversion. *Eur J Biochem* **179**: 629–637, 1989.
8. Brown SC, Mueller L and Jeffs PW, ^{1}H NMR assignment and secondary structural elements of human transforming growth factor α. *Biochemistry* **28**: 593–599, 1989.
9. Montelione GT, Winkler ME, Burton LE, Rinderknecht E, Sporn MB and Wagner G, Sequence specific ^{1}H NMR assignments and identification of two small antiparallel β-sheets in the solution structure of recombinant hTGF-α. *Proc Natl Acad Sci USA* **86**: 1519–1523, 1989.
10. Mayo KH, Cavalli RC, Peters AR, Boelens R and Kaptein R, Sequence specific ^{1}H-n.m.r. assignments and peptide backbone conformation in rat epidermal growth factor. *Biochem J* **257**: 197–205, 1989.
11. Engler DA, Matsunami RK, Campion SR, Stringer CD, Stevens A and Niyogi SK, 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**: 12384–12390, 1988.
12. Lazar E, Watanabe S, Dalton S and Sporn MB, TGF-α: Mutation of aspartic acid 47 and leucine 48 results in different biological activities. *Mol Cell Biol* **8**: 1247–1252, 1988.
13. Burgess AW, Lloyd CJ, Smith S, Stanley E, Walker F, Fabri L, Simpson RJ and Nice EC, Murine epidermal growth factor: Structure and function. *Biochemistry* **27**: 4977–4985, 1988.
14. Han KH, Ferretti JA, Niu CH, Lokeshwar V, Clarke R and Katz D, Conformational and receptor binding properties of hEGF and TGF-α second loop fragments. *J Mol Recognition* **1**: 116–123, 1988.
15. Heath WF and Merrifield RB, A synthetic approach to structure–function relationships in the murine EGF molecule. *Proc Natl Acad Sci USA* **83**: 6367–6371, 1986.
16. Defeo-Jones D, Tai JY, Wegrzyn RJ, Vuocolo GA, Baker AE, Payne LS, Garsky VM, Oliff A and Riemen MW, Structure–function analysis of synthetic and recombinant derivatives of transforming growth factor alpha. *Mol Cell Biol* **8**: 2999–3007, 1988.
17. Wüthrich K, Protein structure determination in solution by NMR. *Science* **243**: 45–50, 1989.
18. Cooke RM and Campbell ID, Protein structure determination by NMR. *Bioessays* **8**: 52–56, 1989.
19. Havel T and Wüthrich K, An evaluation of the combined use of NMR and distance geometry for the determination of protein conformations in solution. *J Mol Biol* **182**: 281–294, 1985.
20. Braun W and Go N, Calculation of protein conformations by proton-proton distance constraints. *J Mol Biol* **186**: 611–626, 1985.
21. Kaptein R, Boelens R, Scheek RM and van Gunsteren WF, Protein structures from NMR. *Biochemistry* **27**: 5389–5395, 1988.
22. Doolittle RF, The geneology of some recently evolved vertebrate proteins. *Trends Biochem Sci* **10**: 233–237, 1985.
23. Huang LH, Ke X-H, Sweeney W and Tam JP, Calcium binding and putative activity of the epidermal growth factor domain of blood coagulation Factor IX. *Biochem Biophys Res Commun* **160**: 133–139, 1989.
24. Handford P, Baron M, Mayhew M, Willis A, Beesley T, Brownlee GG and Campbell ID, The first EGF-like domain of human Factor IX has a high affinity calcium binding site. *EMBO J* **9**: 475–480, 1990.
25. Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P and Barr PJ, α-Factor directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **81**: 4642–4646, 1984.

26. Baron M, Kingsman AJ, Kingsman SM and Campbell ID, Towards the structure of mosaic proteins: Use of protein expression and NMR techniques. In: *Protein Production in Biotechnology* (ed. Harris TJR), pp. 49–60. Elsevier, Amsterdam, 1989.
27. Mellor J, Dobson MJ, Roberts NA, Tuite MF, Emtage JS, White S, Lowe PA, Patel T, Kingsman AJ and Kingsman SM, Efficient synthesis of enzymatically active calf chymosin in *Saccharomyces cerevisiae*. *Gene* **24**: 1–14, 1983.
28. George-Nascimento C, Gyenes A, Halloran SM, Merryweather J, Valenzuela P, Steimer KS, Masiarz FR and Randolph A, Characterization of recombinant human epidermal growth factor produced in yeast. *Biochemistry* **27**: 797–802, 1988.
29. Dudgeon TJ, Baron M, Cooke RM, Fallon A and Campbell ID, Structure and function of hEGF: Receptor binding and NMR studies of site specific mutants. *FEBS Lett*, in press.