A high resolution ^1H NMR study of the solution structure of human epidermal growth factor

J.A. Carver+, R.M. Cooke, G. Esposito, I.D. Campbell*, H. Gregory† and B. Sheard†

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU and †*ICI Pharmaceuticals Division, Mereside Alderley Park, Macclesfield SK10 4TG, England*

Received 26 June 1986

500 MHz ^1H NMR studies of human epidermal growth factor are described. The backbone resonances of the 1–48 derivative of hEGF have been assigned using two-dimensional techniques. Analysis of the type and magnitude of the observed sequential nuclear Overhauser effects and the NH-αCH spin-spin coupling constants allowed prediction of the secondary structure. Aspects of the tertiary structure are also identified. A pair of antiparallel β-sheets involving residues 18–23 and 28–34 is a dominant feature of the solution structure.

Growth factor *(Human EGF)* *NMR* *Protein structure*

1. INTRODUCTION

Human epidermal growth factor (hEGF), also known as urogastrone, is a protein of 53 amino acids with three disulphide bridges [1]. EGF inhibits gastric acid secretion [2] and is thought to play an important part in the growth and differentiation of cells [3]. It is homologous to epidermal growth factors isolated from other mammals [4,5] and to other growth factors [6]. EGF-like sequences have also been observed to form domains in several other extracellular proteins [7]. No detailed crystal structure is available for EGF or any other member of this family of related proteins.

The solution structures of small proteins can now be determined from NMR data [8]. Some NMR studies of murine EGF have been published [9,10], but relatively few assignments were made and conclusions about the protein structure were limited. We describe here NMR studies of the 1–48 fragment of hEGF which retains virtually complete biological activity in vivo. A complete assignment of the observable NH and αCH resonances has been achieved and will be reported in detail in a later paper. In this paper we present an analysis of the observed nuclear Overhauser effects (NOE) and the NH-αCH spin-spin coupling constants. Slowly exchanging hydrogens have also been identified and related to hydrogen bonding patterns. These procedures, which have been discussed in detail elsewhere [11], have allowed several features of the secondary and tertiary structure of hEGF to be identified.

2. MATERIALS AND METHODS

hEGF, with a polyarginine tail at the C-terminus, was produced in *E. coli* by G.D. Searle using recombinant DNA techniques [12]. The hEGF and the 1–48 derivative were obtained by trypsin treatment of this material and then purified by ion-exchange and gel-exclusion chromatography [13]. The lyophilised powder was dissolved in 10% D_2O/90% H_2O or in 100% D_2O to

* To whom correspondence should be addressed
+ Present address: Department of Biochemistry, University of Adelaide North Terrace, Adelaide, S. Australia 5000

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

give a protein concentration of about 8 mM. The 1–53 derivative was found to be less soluble and solutions of around 2 mM were studied. The solutions were titrated with DCl to a pH of 3.0–3.1 (uncorrected meter reading).

^{1}H NMR spectra were recorded at 30°C and the shifts are quoted with reference to 3-trimethylsilyl[2,2,3,3-^{2}H]-propionate. spectra were acquired with a Bruker AM500 spectrometer or a 500 MHz spectrometer assembled in this laboratory equipped with a Nicolet/GN 1280 computer and an Oxford Instruments magnet. Two-dimensional NOESY [14] and COSY [15] spectra were acquired in the phase sensitive mode with quadrature detection in F_1 achieved by the method of States et al. [16] or by TPPI [17]. Relayed coherence transfer (RCT) spectra [18] were acquired in the absolute value mode with the phase cycling scheme of Bax and Drobny [19] and delays of 30 and 50 ms between the second and third 90° pulses.

3. RESULTS AND DISCUSSION

This study was mainly carried out using a derivative of the native hEGF; it is thus necessary to consider the relevance of its structure. High resolution 500 MHz 1D NMR spectra of the 1–48 and 1–53 peptides were carefully compared. Although the deleted pentapeptide contains two tryptophans, only small changes in shift were observed for all the remaining methyl group and aromatic resonances. This suggests that the 49–53 pentapeptide is not very intimately associated with the remainder of the protein. Previous photo-CIDNP experiments on murine EGF [20] also indicated that Trp-49 and Trp-50 are solvent exposed. In addition, experiments on the biological activity of the 1–47 peptide [13] and the 1–48 peptide (H. Gregory, unpublished) show that, while these peptides are less potent than hEGF as mitogens, the effects on gastric acid secretion are comparable. There is thus good reason to suppose that structural conclusions about the 1–48 fragment will have validity for the native protein.

Determination of the secondary structure of a protein from NMR data requires the backbone NH and αCH resonances to be assigned [21]. The procedure used to identify the spin systems of the amino acids was to collect and analyse COSY and RCT spectra, with some additional assistance from NOESY spectra. In fig.1 the COSY cross peaks between the NH and αCH resonances of residues 29–35 are indicated.

Spin systems of residues which are adjacent in the protein chain were identified by NOEs from backbone NH resonances of one residue to resonances of the NH, αCH and βCH groups of the preceding residue [21]. A compilation of this information leads to sequence specific assignment for the spin system of each amino acid. Some of the observed sequential NOEs observed for residues 29–34 are shown in fig.1.

Sequential NOEs are classified as $d_{\alpha N}$ (between αCHI$_i$ and NH$_{i+1}$), d_{NN} (between NH$_i$ and NH$_{i+1}$) and $d_{\beta N}$ (between βCH$_i$ and NH$_{i+1}$). A list of sequential NOEs observed in the 1–48 fragment of hEGF is given in fig.2. Examination of this figure shows that one or more connectivity is observed for all the residues except where overlap of the Ser-4, Cys-6 and Tyr-22 αCH resonances with the H_2O resonance prevented direct observation of sequential NOEs to Glu-5, Pro-7 and Ile-23. (In the case of Cys-6, the sequential NOE would be to the δCH resonances of Pro-7.)

Slowly exchanging amide protons and approximate values for several NH-αCH coupling constants ($^3J_{NH\text{-}\alpha CH}$) are also listed in fig.2. The coupling constants were measured from the antiphase peak separation in high resolution phase sensitive COSY cross sections parallel to F_2 with a digital resolution of 1.2 Hz/point. No allowance was made for possible distortions arising from peak overlap [23].

The observed NOE patterns and $^3J_{NH\text{-}\alpha CH}$ values are diagnostic for protein secondary structure [11,12]. For example β-sheet or extended conformations lead to strong $d_{\alpha N}$ NOEs and large $^3J_{NH\text{-}\alpha CH}$ values, while α-helix leads to strong d_{NN} NOEs and small values of $^3J_{NH\text{-}\alpha CH}$. Many of the $^3J_{NH\text{-}\alpha CH}$ values listed in fig.2 are in the range 8–10 Hz and $d_{\alpha N}$ NOEs are prevalent, suggesting that much of the protein is in an extended or β-sheet conformation.

Features of the tertiary structure can also be deduced from the pattern of slowly exchanging amide hydrogens and longer range NOEs. For example NOEs are observed between the αCH resonances of Cys-31 and Cys-20, and between those of Cys-33 and Gly-18. At 30°C, the Tyr-22

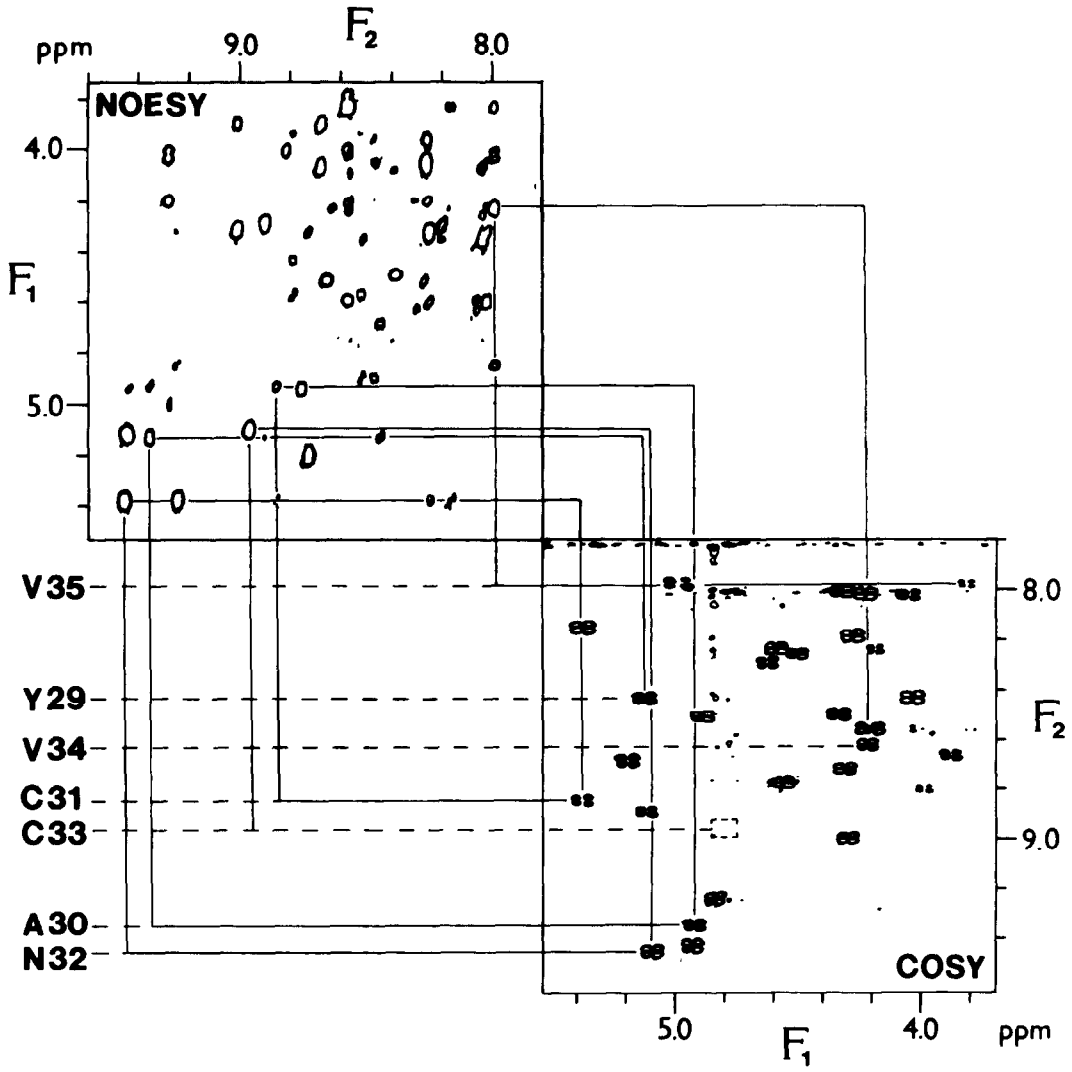

Fig.1. The 'fingerprint' regions of the NOESY (mixing time = 140 ms) and COSY spectra of 1-48 hEGF in 90% H_2O/10% D_2O. The NH-αCH COSY cross peaks for residues 29-35 are indicated and $d_{\alpha N}$ connectivities between adjacent residues are marked in the NOESY spectrum. At this temperature (30°C) the NH-αCH COSY peak for Cys-33 is obscured by the H_2O resonance. This connectivity was observed in a COSY experiment recorded at 50°C and its position is indicated with the dashed box.

αCH resonance is obscured by irradiation of the solvent peak but it is visible at 26°C in D_2O, when an NOE is observed to the αCH resonance of Tyr-29. This indicates that these six residues are involved in an antiparallel β-sheet [11]. Residues in such sheets alternate sequentially between those in which the αCH faces the opposite strand and those in which the backbone NH is hydrogen bonded to a carbonyl of the neighbouring strand. Residues for which αCH-αCH interstrand NOEs are observed should, thus, be preceded and followed by residues with slowly exchanging backbone amide protons. Since slowly exchanging protons are observed at positions 19, 21, 23, 28, 30, 32 and 34 (fig.2), this confirms the antiparallel β-sheet involving residues 18-23 and 28-34.

The pairing of residues 23 and 28 forces residues 24-27 to adopt a 'hairpin' turn [24] and leads to

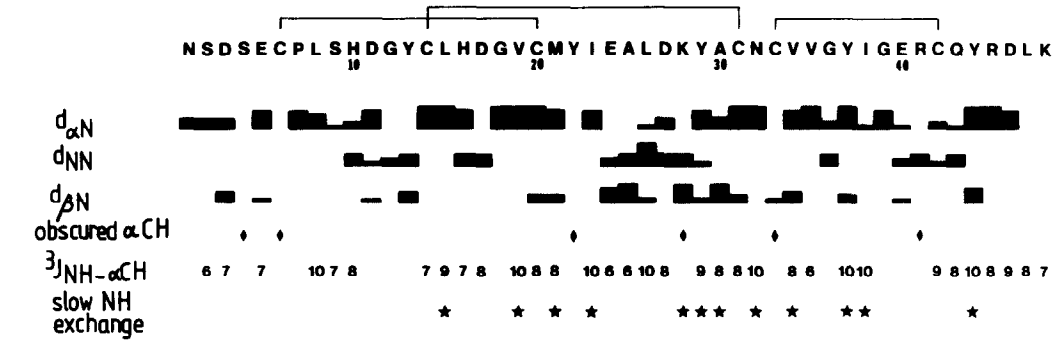

Fig.2. A summary of the NMR data relevant to the structure of 1–48 hEGF. At the top of the diagram is shown the hEGF sequence with the location of the disulphide bonds. The observable $d_{\alpha N}$, d_{NN} and $d_{\beta N}$ connectivities are marked directly underneath. The height of the bars indicates the approximate magnitude of the NOEs. Below this, the residues for which the αCH resonances are obscured at 30°C by irradiation of H_2O are indicated (♦). The values of $^3J_{NH-\alpha CH}$ (in Hz) are then listed. Finally the locations of the backbone amide protons which are slowly exchanging in D_2O are indicated (★).

the d_{NN} NOEs observed for these residues (fig.2).

A type II turn [24] is located between residues 34 and 37, characterised [11] by a strong $d_{\alpha N}$ connectivity between Val-35 and Gly-36, a small $^3J_{NH-\alpha CH}$ for Val-35, a d_{NN} connectivity between Gly-36 and Tyr-37 and the involvement of the Tyr-37 amide proton in a hydrogen bond. Confirmatory evidence comes from the NOEs between resonances of the side chains of Val-34 and Tyr-37.

The location of another turn between residues 41 and 44 is suggested by the sequential NOEs in this region (fig.2). This would allow the Tyr-44 amide proton to be hydrogen bonded to the Arg-41 carbonyl and account for its slow exchange rate. It is not yet possible, however, to classify unambiguously this turn. Further turns are expected between the disulphide linked residues 6 and 20, possibly in the vicinity of Gly-12. There is no evidence for helix. The N- and C-terminal regions appear to be extended with the N-terminus relatively mobile as judged by its relatively sharp resonances.

4. CONCLUSIONS

^{1}H NMR spectra of hEGF have been analysed and some aspects of the solution structure identified. The major structural feature of hEGF is the antiparallel β-sheet, involving residues 18–23 and 28–34. It is interesting to note that one residue from each disulphide bridge is located in this β-sheet region. The location of the β-sheet is in good agreement with the structure prediction of Holladay et al. [25] for murine EGF, although there are important differences; residues 34–37, for example, were predicted to be in a sheet while we find this region is involved in a type II turn. A complete determination of the solution structure of human EGF using distance constraints from NMR is now required. We are beginning to apply both the restrained molecular dynamics [26] and the distance geometry approaches [27] to this task.

ACKNOWLEDGEMENTS

We wish to thank the MRC, the SERC and G.D. Searle for their financial support. We also thank Dr J. Boyd, M. Baron and N. Soffe for their assistance.

REFERENCES

[1] Gregory, H. and Preston, B.M. (1977) Int. J. Peptide Protein Res. 9, 107–118.
[2] Gregory, H. (1975) Nature 257, 325–327.
[3] Schlessinger, J., Schreiber, A.B., Levi, A., Lax, I., Libermann, T. and Yarden, Y. (1983) CRC Crit. Rev. Biochem. 14, 93–111.
[4] Cohen, S. and Carpenter, G. (1975) Proc. Natl. Acad. Sci. USA 72, 1317–1321.

[5] Simpson, R.J., Smith, J.A., Moritz, R.L., O'Hare, M.J., Rudland, P.S., Morrison, J.R., Lloyd, C.J., Grego, B., Burgess, A.W. and Nice, E.C. (1985) Eur. J. Biochem. 153, 629–637.
[6] Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y. and Goeddel, D.V. (1984) Cell 38, 287–297.
[7] Doolittle, R.F. (1985) Trends Biochem. Sci. 9, 233–237.
[8] Williamson, M.P., Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 295–315.
[9] Mayo, K.H. (1984) Biochemistry 23, 3960–3973.
[10] Mayo, K.H. (1985) Biochemistry 24, 3783–3794.
[11] Wüthrich, K., Billeter, M. and Braun, W. (1984) J. Mol. Biol. 180, 715–740.
[12] Sassenfeld, H.M. and Brewer, S.J. (1984) Biotechnology Jan., 76–81.
[13] Hollenberg, M.D. and Gregory, H. (1980) Mol. Pharmacol. 17, 314–320.
[14] Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) J. Chem. Phys. 71, 4546–4553.
[15] Bax, A. and Freeman, R. (1981) J. Magn. Reson. 44, 542–561.
[16] States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) J. Magn. Reson. 48, 286–292.
[17] Drobny, G., Pines, A., Sinton, S., Weitekamp, D.P. and Wemmer, D. (1979) Symp. Faraday Soc. 13, 49–53.
[18] Eich, G., Bodenhausen, G. and Ernst, R.R. (1982) J. Am. Chem. Soc. 104, 3731–3732.
[19] Bax, A. and Drobny, G. (1985) J. Magn. Reson. 61, 306–320.
[20] De Marco, A., Menegatti, E. and guarneri, M. (1983) FEBS Lett. 159, 201–206.
[21] Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) J. Mol. Biol. 155, 311–319.
[22] Pardi, A., Billeter, M. and Wüthrich, K. (1984) J. Mol. Biol. 180, 741–751.
[23] Neuhaus, D., Wagner, G., Vasak, M., Kägi, J.H.R. and Wüthrich, K. (1985) Eur. J. Biochem. 151, 257–273.
[24] Richardson, J.S. (1981) Adv. Protein. Chem. 34, 167–339.
[25] Holladay, L.A., Savage, C.R., Cohen, S. and Puett, D. (1976) Biochemistry 15, 2624–2632.
[26] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) J. Mol. Biol. 182, 179–182.
[27] Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 281–294.