Protein Science (1998), 7:1738–1749. Cambridge University Press. Printed in the USA.

Role of the 6–20 disulfide bridge in the structure and activity of epidermal growth factor

KEVIN J. BARNHAM,[1] ALLAN M. TORRES,[1,4] DIANNE ALEWOOD,[2] PAUL F. ALEWOOD,[2] TERESA DOMAGALA,[3] EDOUARD C. NICE,[3] AND RAYMOND S. NORTON[1]

[1]Biomolecular Research Institute, 343 Royal Parade, Parkville, Victoria 3052, Australia
[2]Centre for Drug Design and Development, University of Queensland, Queensland 4072, Australia
[3]Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, PO Box 2008, Royal Melbourne Hospital, Victoria 3050, Australia

(RECEIVED January 22, 1998; ACCEPTED April 8, 1998)

Abstract

Two synthetic analogues of murine epidermal growth factor, [Abu6, 20] mEGF4-48 (where Abu denotes amino-butyric acid) and [G1, M3, K21, H40] mEGF1-48, have been investigated by NMR spectroscopy. [Abu6, 20] mEGF4-48 was designed to determine the contribution of the 6–20 disulfide bridge to the structure and function of mEGF. The overall structure of this analogue was similar to that of native mEGF, indicating that the loss of the 6–20 disulfide bridge did not affect the global fold of the molecule. Significant structural differences were observed near the N-terminus, however, with the direction of the polypeptide chain between residues four and nine being altered such that these residues were now located on the opposite face of the main β-sheet from their position in native mEGF. Thermal denaturation experiments also showed that the structure of [Abu6, 20] mEGF4-48 was less stable than that of mEGF. Removal of this disulfide bridge resulted in a significant loss of both mitogenic activity in Balb/c 3T3 cells and receptor binding on A431 cells compared with native mEGF and mEGF4-48, implying that the structural changes in [Abu6, 20] mEGF4-48, although limited to the N-terminus, were sufficient to interfere with receptor binding. The loss of binding affinity probably arose mainly from steric interactions of the dislocated N-terminal region with part of the receptor binding surface of EGF. [G1, M3, K21, H40] mEGF1-48 was also synthesized in order to compare the synthetic polypeptide with the corresponding product of recombinant expression. Its mitogenic activity in Balb/c 3T3 cells was similar to that of native mEGF and analysis of its ^{1}H chemical shifts suggested that its structure was also very similar to native.

Keywords: disulfide connectivities; murine epidermal growth factor; nuclear magnetic resonance; protein structure

Native protein structures are stabilized by large numbers of weak interactions, such as hydrogen bonds and hydrophobic, electrostatic and dipole-dipole interactions. Covalent cross-links also contribute in many proteins, with disulfide bridges being the most important. The stabilizing influence of disulfide bridges in proteins has traditionally been explained by their ability to reduce the conformational degrees of freedom of the polypeptide chain in its unfolded state, thereby decreasing the entropy of the unfolded state and destabilizing it relative to the folded state (Flory, 1956; Richardson, 1981; Thornton, 1981). Doig and Williams (1991) suggested, however, that the dominant effect of disulfide bridges on stability was enthalpic, with the presence of disulfide cross-links destabilizing the unfolded protein by reducing the effectiveness of the hydrogen bonding network relative to an unfolded state lacking the disulfides. More recently, Betz (1993) has suggested that disulfide bridges stabilize proteins by enthalpic or entropic contributions or a combination of both.

There have been several studies in which the role of disulfide bridges in small proteins (less then 100 residues) and polypeptides has been probed by removal of one or more of the native disulfides. When the 30 to 51 disulfide bridge was deleted in bovine pancreatic trypsin inhibitor by substitution of the half-cystines with Ala, the resulting structural changes were smaller than the conformational variability found among three different crystal structures of the molecule (Eigenbrot et al., 1990), although there were some

Reprint requests to: R.S. Norton, Biomolecular Research Institute, 343 Royal Parade, Parkville, Victoria 3052, Australia; e-mail: ray.norton@molsci.csiro.au.

[4]Present address: Department of Biochemistry, University of Sydney, NSW 2006, Australia.

Abbreviations: Abu, amino butyric acid; DMF, dimethylformamide; DQF-COSY, double-quantum-filtered scalar-correlated spectroscopy; E-COSY, exclusive correlation spectroscopy; EGF, epidermal growth factor; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; RMSD, root-mean-square deviation; RP-HPLC, reversed phase high-performance liquid chromatography; 1D, one-dimensional; 2D, two-dimensional.

movements to compensate for a packing defect associated with the missing sulfur atoms. Deletion of the 5 to 55 disulfide bridge caused small changes near the N-terminus (residues 1–7) and C-terminus (residues 55–58), but otherwise the protein had a native-like structure (van Mierlo et al., 1991). The conformation of this analogue was less native-like than those of analogues with only the 5 to 55 bridge or the 5 to 55 and 14 to 38 bridges intact, indicating that the 5 to 55 bridge is the most important of the three disulfides in stabilizing the native conformation (van Mierlo et al., 1991). An analogue with only the 14 to 38 disulfide intact (the other four half-cystines having been replaced by α-amino-n-butyric acid (Abu), an isosteric analogue of Cys with similar conformational preferences) retained two strands of antiparallel β-sheet at low temperature, but the rest of the molecule sampled a range of conformations (Barbar et al., 1995). Fully reduced bovine pancreatic trypsin inhibitor has been described as a "molten coil," characterized by extensive non-random structure that includes many non-native interactions among side chains (Pan et al., 1995). In a smaller trypsin inhibitor, EETI II, loss of the 2 to 19 disulfide caused mainly local structural changes in the vicinity of residues 2 and 19 (Le-Nguyen et al., 1993).

In insulin-like growth factor I, replacement of the 47 to 52 disulfide bridge residues by either Cys or Ala caused local unfolding of a nearby α-helix, but no other significant structural changes (Hua et al., 1996). In the potassium channel blocker charybdotoxin, removal of the 13 to 33 disulfide by replacement with Abu caused some structural changes, but the RMS difference between the 10 lowest energy structures for the analogue and native charybdotoxin was only 1.44 Å for the backbone heavy atoms (Song et al., 1997). It should be noted, however, that the solution structure of the analogue was determined at a much lower temperature (5 °C, compared with 45 °C for the native toxin). The analogue had an eightfold lower affinity for the potassium channel than native charybdotoxin. In another scorpion-derived potassium channel blocker with three disulfides, leiurotoxin I, an analogue lacking the 3 to 21 bond was fully active and had a very similar CD spectrum to that of the native toxin (Sabatier et al., 1996). Similar findings were reported for the RGD-containing disintegrin, echistatin γ, where deletion of the last four residues and one of the four disulfides (by replacement of Cys8 and Cys37 with Ala) caused only a slight reduction in activity and minor structural changes (Chuang et al., 1996), although in this protein the residues essential for activity are in less ordered parts of the structure, which might be less sensitive to conformational perturbations elsewhere in the molecule. Finally, in heat-stable enterotoxin B from *Escherichia coli*, which is a 48-residue protein with two disulfide bridges, loss of either disulfide abolished toxicity (Arriaga et al., 1995). One of the single-disulfide analogues retained native secondary structure, as monitored by CD, but the other showed significant changes.

In general, it appears that, in small proteins with multiple disulfides, the loss of a disulfide bridge has a greater impact on biological activity than on the overall structure, and that at least part of any loss of activity can be attributed to the greater flexibility of the analogue compared with the native protein. Moreover, in any given protein, some disulfides are, not surprisingly, more important than others in stabilising the native fold.

EGF is a small mitogenic peptide of 53 amino acid residues that was first characterized and sequenced in 1972 (Savage et al., 1972). EGF and EGF-like proteins such as TGF-α are thought to play roles in wound healing and oncogenesis (Burgess, 1989; Campbell & Bork, 1993). Since the discovery of EGF, more than 300 EGF-like sequences have been identified, mostly as domains of larger proteins. These proteins have been associated with a diverse range of functions, including blood coagulation, fibrinolysis, neural development, and cell adhesion (Campbell & Bork, 1993).

Structures have been determined for a large number of these EGF and EGF-like molecules, including solution structures for human EGF (Cooke et al., 1987; Hommel et al., 1992), murine EGF (Montelione et al., 1986, 1988, 1992; Kohda & Inagaki, 1992a, 1992b; Tejero et al., 1996), and TGF-α (Harvey et al., 1991; Tejero et al., 1996). EGF and EGF-like molecules contain three disulfide bridges (Fig. 1), typically paired in a 1-3/2-4/5-6 pattern (although recently White et al. (1997) have identified an EGF-like domain of thrombomodulin with a 1-2/3-4/5-6 disulfide bonding pattern). Other common structural features are two antiparallel β-sheets, consisting of a larger triple-stranded β-sheet where residues at the N-terminus contribute a small third strand, and a smaller two-stranded β-sheet near the C-terminus of the molecule. The three disulfide bridges anchor the amino-terminal strand (disulfide bridge 1-3) and loop (2-4), as well as the C-terminal domain (5-6) containing the smaller sheet, to the surface of the major β-sheet.

Given the highly conserved nature of the disulfide bridging pattern in EGF and EGF-like modules, it is likely that they play a key role in maintenance of the common EGF structure. In this paper, we describe the structural changes associated with deletion of the first disulfide bridge (Cys6–Cys20), achieved by substituting the two half-cystine residues at positions 6 and 20 with Abu. This analogue was also truncated at both the N-terminus (first three residues deleted) and C-terminus (last five residues deleted), but these parts of the molecule have been shown previously to be unimportant for biological activity in EGF (Simpson et al., 1985; Burgess et al., 1988; Groenen et al., 1994). The residues deleted at the C-terminus were also disordered in the solution structure of mEGF (Montelione et al., 1988). Another analogue, [G1, M3, K21, H40] mEGF1-48, has also been synthesized and investigated by NMR; its structure was very similar to that of native EGF.

Results

Resonance assignments

Sequence-specific ^{1}H NMR resonance assignments were made for both analogues from 2D NOESY and TOCSY spectra (Wüthrich, 1986); the C$^\alpha$H-NH region of the NOESY spectrum of [Abu6, 20] mEGF4-48 in water is shown in Figure 2. Plots of the deviation of the C$^\alpha$H and NH chemical shifts for [Abu6, 20] mEGF4-48[5] and [G1, M3, K21, H40] mEGF1-48 from those of native mEGF reported by Montelione et al. (1988) are shown in Figure 3. The chemical shift differences from the native molecule for [G1, M3, K21, H40] mEGF1-48 are very small and are limited to the sites of mutation, suggesting that the structures are very similar. As a result, no detailed structure determination was undertaken for this molecule. Chemical shift differences from native were more pronounced for [Abu6, 20] mEGF4-48, suggesting that there may be some significant structural differences between these two polypeptides, and a full structure determination of [Abu6, 20] mEGF4-48 was therefore carried out. Figure 4 summarizes the sequential and

[5]The residue numbering system of full-length mEGF is used for both analogues.

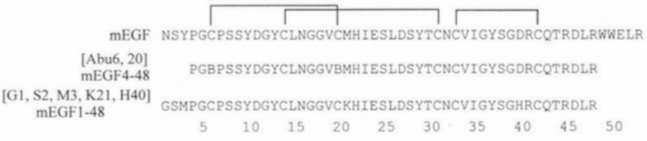

Fig. 1. Amino acid sequences of mEGF, [Abu6, 20] mEGF4-48 (where Abu denotes α-amino-butyric acid; single letter code B) and [G1, M3, K21, H40] mEGF1-48. The locations of the three disulfide bridges in mEGF are indicated above the sequence, and the numbering system is that of native mEGF.

medium-range NOE connectivities, $^3J_{NHC\alpha H}$ coupling constants, and slowly exchanging amide protons for [Abu6, 20] mEGF4-48 in aqueous solution at 300 K.

NOESY and TOCSY spectra of synthetic mEGF1-48 were also recorded. The chemical shifts were very similar to those of full-length mEGF (Montelione et al., 1988), the largest differences being for the NH and $C^{\alpha}H$ resonances of Gly36 (0.31 and 0.29 ppm downfield, respectively) and $C^{\alpha}H$ of Ile35 (0.23 ppm downfield). All other peaks were within 0.1 ppm of native values except for the new C-terminus, Arg48. These observations confirm that the C-terminal five residues of mEGF have very little interaction with the bulk of the molecule.

Structure determination

Structures were calculated using 953 upper-bound distance constraints inferred from NOEs, made up of 332 intra-residue, 234 sequential, 118 medium-range ($1 < |i - j| \leq 4$) and 269 long-range NOEs (Fig. 5A), as well as 93 lower bound constraints based on the absence of NOEs from the NOESY spectrum (Manoleras & Norton, 1994). In addition, 31 backbone dihedral angle constraints based on spin-spin coupling constants were included, as well as 5 χ^1 side chain constraints for stereo-specifically assigned, non-degenerate, geminal $C^{\beta}H$ resonances. Structures were calculated

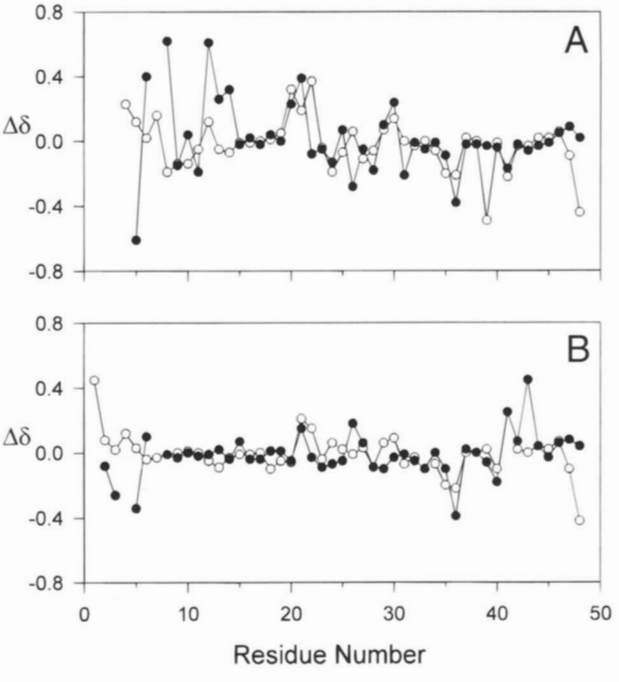

Fig. 3. Chemical shift differences (in ppm) of mEGF analogues from native mEGF (Montelione et al., 1988). **A:** [Abu6, 20] mEGF4-48. **B:** [G1, M3, K21, H40] mEGF1-48. ●, NH chemical shift; ○, $C^{\alpha}H$ chemical shift.

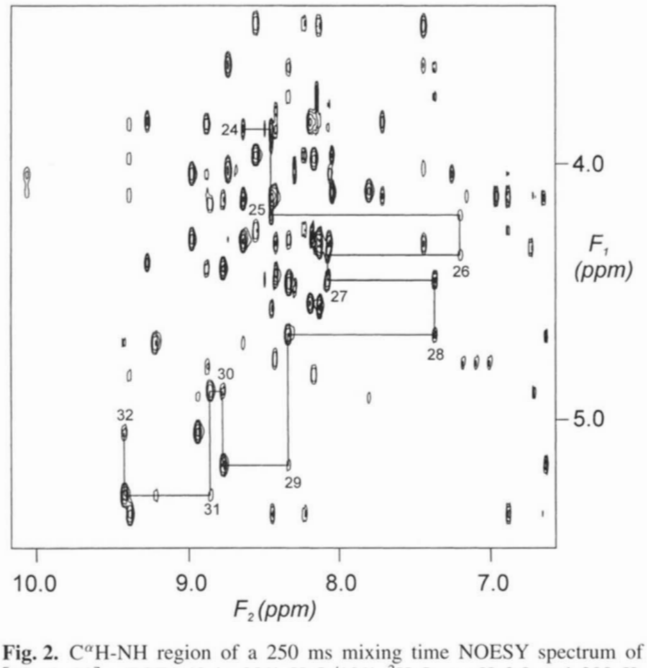

Fig. 2. $C^{\alpha}H$-NH region of a 250 ms mixing time NOESY spectrum of [Abu6, 20] mEGF4-48 in 90% H_2O/10% ^2H_2O at pH 2.8 and 300 K. Sequential connectivities from residue 24 to 32 are indicated, with intra-residue $d_{N\alpha}$ cross-peaks numbered.

Fig. 4. Summary of NMR data for [Abu6, 20] mEGF4-48 in 90% H_2O/10% ^2H_2O at pH 2.8 and 300 K. The intensities of $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ connectivities are represented as strong, medium, or weak by the height of the bars. The shaded bar indicates a $d_{\alpha\delta}$ connectivity to Pro7. Asterisks indicate that the presence of a NOE could not be confirmed unambiguously due to peak overlap. Values of $^3J_{NHC\alpha H} < 6$ Hz are indicated by ↓, those >8 Hz by ↑. Those left blank could not be measured due to overlap, or were between 6 and 8 Hz. The relative exchange rates of backbone NH protons are indicated in the row labeled NH, based on the strength of cross-peaks in a ^2H_2O exchange TOCSY experiment; slowly exchanging amides are indicated by filled circles, while open circles indicate intermediate exchange. NH protons with intermediate exchange rates were identified as those that had peaks present in the TOCSY spectra after 10 h but not after 24 h. The bottom plot is a chemical shift index (CSI) analysis (Wishart et al., 1992) for the $C^{\alpha}H$ resonances, where CSI is shown as positive if the $C^{\alpha}H$ resonance is >0.1 ppm downfield from its random coil value (Wishart et al., 1995a), negative if >0.1 ppm upfield, or 0 if within 0.1 ppm of its random coil value.

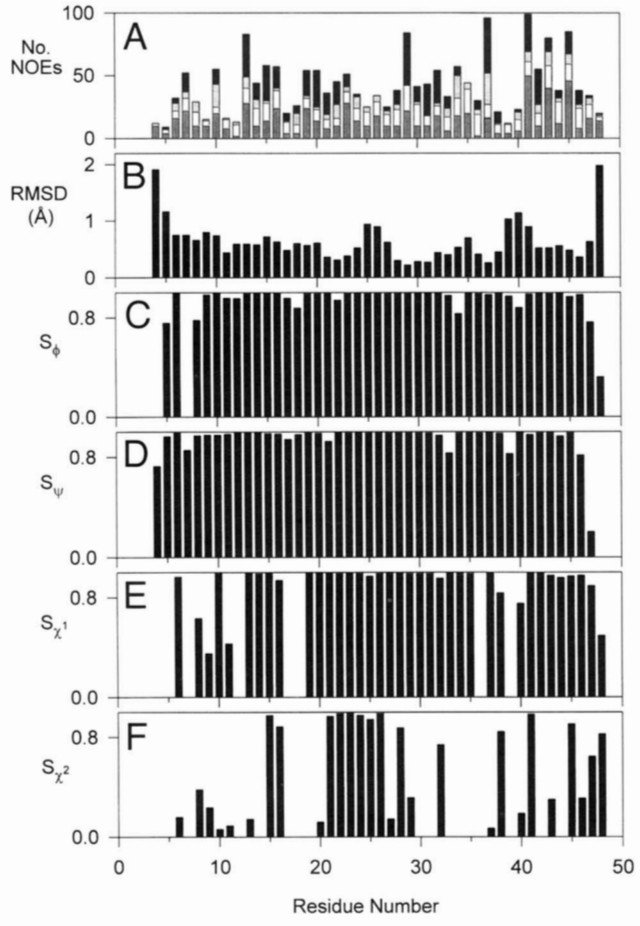

Fig. 5. Parameters characterizing the final 20 structures of [Abu6, 20] mEGF4-48 in water, plotted as a function of residue number. **A:** Upper-bound distance restraints used in the final round of structure refinement; long range ($i - j \geq 5$), medium-range ($2 \leq i - j \leq 4$), sequential and intra-residue NOEs are shown, respectively, in black, cross-hatched, white, and hatched shading. NOEs are counted twice, once for each proton involved. **B:** RMSDs from the mean structure for the backbone heavy atoms (N, C^α, C) following superposition over the whole molecule. **C–F:** Angular order parameters (S) (Hyberts et al., 1992; Pallaghy et al., 1993) for the backbone (ϕ and ψ) and side chain (χ^1 and χ^2) dihedral angles. Gaps in the χ^1 plot are due to Gly and Ala residues. Gaps in the χ^2 plot, in addition to Gly and Ala, are due to Ser, Pro, Val, Cys, and Thr residues.

Table 1. *Structural statistics for the 20 energy-minimized structures of [Abu6, 20] mEGF4-48 from X-PLOR*[a]

RMSDs from experimental distance restraints Å (1,046)[b,c]	0.020 ± 0.001
RMSDs from experimental dihedral restraints (deg) (36)[b]	0.36 ± 0.07
RMSDs from idealized geometry	
Bonds (Å)	0.0105 ± 0.0003
Angles (deg)	2.69 ± 0.05
Impropers (deg)	0.32 ± 0.01
Energies (kcal mol^{-1})	
E_{NOE}	17.9 ± 1.1
E_{cdih}	0.29 ± 0.10
$E_{L\text{-}J}$	−72.6 ± 5.9
$E_{bond} + E_{angle} + E_{improper}$	144.9 ± 6.2
E_{elec}	−547.6 ± 13.1

Mean pairwise RMSD (Å)	Backbone heavy atoms	All heavy atoms
Residues 4–48	0.62 ± 0.21	1.14 ± 0.19
Residues 6–46 ($S\phi,\psi > 0.8$)	0.42 ± 0.15	0.91 ± 0.15

[a]The best 20 structures after energy minimization in the distance geometry force-field were subsequently energy minimized in the CHARMM force-field, using a distance-dependent dielectric and neutralized side chains, as described under Materials and methods.

[b]The numbers of restraints are shown in parentheses. None of the structures had distance violations >0.3 Å or dihedral angle violations >5°.

[c]Distance restraints consisted of 953 upper-bound restraints (332 intra-residue, 234 sequential, 118 medium-range ($1 < |i - j| \leq 4$) and 269 long-range) and 93 lower-bound restraints (see text).

initially using DYANA, then refined by simulated annealing in X-PLOR and finally energy minimized in X-PLOR with the CHARMM force field. A summary of geometric and energetic parameters for these structures is given in Table 1.

Structural analysis and description

Analysis of the backbone angular order parameters (S) (Hyberts et al., 1992; Pallaghy et al., 1993) of the final 20 structures indicates that residues 6–46 are well defined, with $S > 0.8$ for both ϕ and ψ angles (Fig. 5C,D). The backbone RMSD from the mean structure is plotted as a function of residue number in Figure 5B, which confirms that the structure is well defined over most of the molecule, except for residues near the N- and C-termini. The overall conformation of [Abu6, 20] EGF4-48 is shown in Figure 6, where the backbone heavy atoms of the 20 best structures (those with the lowest overall energies, excluding the electrostatic term) have been superimposed over residues 6–46. Several of these residues have well-defined χ^1 angles and some also have well-defined χ^2 angles (Fig. 5E,F).

The absence of the 6 to 20 disulfide bond in [Abu6, 20] mEGF4-48 did not cause any gross changes to the structure of murine EGF, its overall fold being similar to that of native mEGF (Fig. 7A). The only major difference was at the N-terminus of the molecule, where there was a significant change in the direction of the polypeptide chain. The effect of the 6 to 20 disulfide in mEGF is to pull the N-terminus across the loop (residues 20–31) that contributes the two main strands of the first β-sheet, allowing the first four residues to form an extra strand of the sheet. In the absence of the 6 to 20 disulfide, the polypeptide chain continued on the opposite side of the loop and was not involved in any secondary structural elements (Fig. 7B,C). The structures diverged from one another around Ser9, the ϕ angles of which differed by ca. 160° (−31° for [Abu6, 20] mEGF4-48 and 131° for mEGF (Montelione et al., 1992). The RMS differences between the positions of the C^α atoms in the N-terminal regions of the two families of structures were as follows (mean ± SD, in Å): Pro4, 17.9 ± 0.6; Gly5, 12.5 ± 0.4; Cys/Abu6, 8.5 ± 0.5; Pro7, 5.4 ± 0.5; Ser8, 6.3 ± 0.5; Ser10, 1.8 ± 0.9. The mean pairwise RMS differences between the two families were 2.0 and 3.8 Å over the backbone heavy atoms (N, C^α, C) of residues 9–48 and 4–48, respectively.

NOE differences at N-terminus

As a result of the structural change at the N-terminus, there were several NOE peaks observed in the NOESY spectrum of [Abu6,

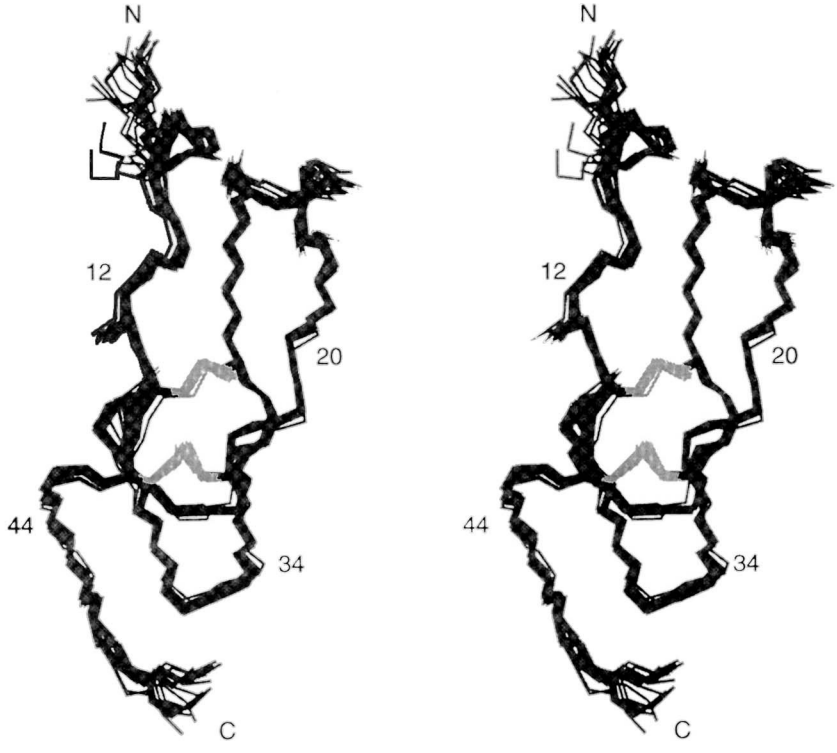

Fig. 6. Stereo view of the backbone heavy atoms and disulfide bonds of the final 20 structures of [Abu6, 20] mEGF4-48 in water, superimposed over the backbone heavy atoms (N, C^{α}, C) of the well-defined (S_{ϕ} and $S_{\psi} > 0.8$) region of the molecule, encompassing residues 6–46.

20] mEGF4-48 that were not present in the spectrum of mEGF (PDB accession no. 1EGF; Bernstein et al., 1977; Montelione et al., 1988, 1992), and vice versa. Unique NOEs observed in the spectrum of [Abu6, 20] mEGF4-48 include Gly5 $C^{\alpha}H$ to His22 C(2)H and Thr30 $C^{\beta}H$, Abu6 NH to Ser28 $C^{\alpha}H$, Abu6 $C^{\beta}H_2$ to Thr30 $C^{\beta}H_2$, Pro7 $C^{\beta}H_2$ to Tyr13 C(2,6)H, Pro7 $C^{\gamma}H_2$ to Glu24 $C^{\gamma}H_2$, and Pro7 $C^{\delta}H_2$ to Ser28 $C^{\beta}H_2$. Conversely, NOEs in native mEGF between Pro4 $C^{\alpha}H$ and Val19 $C^{\gamma}H_3$ and Met21 $C^{\alpha}H$, $C^{\gamma}H_2$ and $C^{\epsilon}H_3$, Gly5 $C^{\alpha}H_2$ and Met21 $C^{\alpha}H$, and Cys6 NH and Cys20 $C^{\beta}H_2$ were not seen in [Abu6, 20] mEGF4-48. While there were differences in the long-range NOEs for residues near the N-terminus of [Abu6, 20] mEGF4-48 and the native molecule, there were also similarities, with NOEs observed in both molecules between Gly5 and His22, Cys(Abu)6 and Tyr10, Tyr29 and Pro7 and Tyr29, and between Tyr13 and Cys(Abu)20.

Coupling constants

The coupling constants observed from the DQF-COSY spectrum of [Abu6, 20] mEGF4-48 (Fig. 4) were in good agreement with those observed by Montelione et al. (1986). There were a few exceptions, where we observed low coupling constants ($\leq$6 Hz) but high coupling constants ($\geq$8 Hz) were reported by Montelione et al. (1986), viz. Tyr13, Glu24, and Ile35.

NH exchange rates

Most of the slowly exchanging amides observed in mEGF by Montelione et al. (1988) were also observed for [Abu6, 20] mEGF4-48 (Fig. 4), but there were some subtle differences. The most striking was the NH exchange rate of Ile35, which was the slowest in native mEGF (still present in spectra well after 20 h) but had an intermediate exchange rate in [Abu6, 20] mEGF4-48 (no longer present in spectra after 20 h). The other difference was Abu20 which had an intermediate NH exchange rate in [Abu6, 20] mEGF4-48 (a weak peak was still observed in the TOCSY spectrum after 20 h), even though its counterpart in native mEGF, Cys20, was not observed in a TOCSY spectrum 10 h after dissolution in ^2H_2O (Montelione et al. 1986).

Temperature and pH dependence

Figure 8 shows the effect of temperature on 1D spectra of mEGF1-48 (with all three disulfide bridges intact) and [Abu6, 20] mEGF4-48, both in ^2H_2O. These spectra clearly show that mEGF1-48 is more stable to thermal denaturation than [Abu6, 20] mEGF4-48. Peaks due to the low field-shifted $C^{\alpha}H$ and some aromatic protons (e.g., from Tyr10 and Tyr29, which are close to one another in the structure) in the spectra of mEGF1-48 began to broaden at 333 K, and those of Tyr37 began to broaden at 343 K, but at 353 K the aromatic peaks of Tyr3 and Tyr13 were still relatively sharp and some low field $C^{\alpha}H$ resonances (5–5.6 ppm) were still observed, indicating that the molecule retained some structure even at this temperature. The low field-shifted $C^{\alpha}H$ peaks in the 1D spectra of [Abu6, 20] mEGF4-48 began to broaden at 323 K, while the aromatic peaks began to broaden at 333 K. By 343 K [Abu6, 20] mEGF4-48 was largely denatured. Because the transition to a fully unfolded form was not complete in the of case mEGF1-48 at the highest temperature examined, it was not possible to specify the mid-point of the thermal denaturation process. However, it was clear that [Abu6, 20] mEGF4-48 had a thermal denaturation temperature more than 10° lower than that of mEGF1-48.

The pK_a values of His22 were determined to be 7.03 and 7.11 for [Abu6, 20] mEGF4-48 and mEGF1-48, respectively, in ^2H_2O.

Biological activity

The mitogenic activities of [Abu6, 20] mEGF4-48, [G1, M3, K21, H40] mEGF1-48 and mEGF1-48 were compared with native mEGF by observing the uptake of tritium-labeled thymidine in Balb/c 3T3 cells. The activities of both synthetic and recombinant [G1, M3, K21, H40] mEGF1-48 and mEGF1-48 were similar to that of native mEGF (EC_{50} 50–100 pM), whilst [Abu6, 20] mEGF4-48 was considerably less active (EC_{50} 30–50 nM). All the derivatives gave a similar stimulation index (around sevenfold), with essentially parallel titration curves. To confirm that, as reported previously (Simpson et al., 1985; Burgess et al., 1988; Groenen et al., 1994), the first three N-terminal residues were not essential for biological activity, the [G1, M3, K21, H40] mEGF1-48 analogue was cleaved with cyanogen bromide (M3 is the only methionine residue in this analogue) and the product was purified to homogeneity by RP-HPLC and analyzed by mass spectrometry and amino acid analysis prior to bioassay. The EC_{50} of the [K21, H40] mEGF4-48 generated in this way (150–160 pM) confirmed that removal of the three N-terminal residues had little effect on biological activity.

The ability of [Abu6, 20] mEGF4-48 to bind to the EGF receptor was also tested in a competition binding assay using A431 cells. In this assay, [Abu6, 20] mEGF4-48 clearly showed concentration-dependent inhibition of binding, with an IC_{50} of 230 nM. By comparison, both mEGF and mEGF4-48 had IC_{50} values of approximately 1 nM in the same assay, in good agreement with values reported previously (Burgess et al., 1988, Groenen et al., 1994).

Discussion

A high-resolution solution structure of [Abu6, 20] mEGF4-48, an analogue of mEGF lacking the 6 to 20 disulfide bridge, has been determined at 300 K and pH 2.8. The deletion of this disulfide bond caused significant local changes in the structure but did not alter the overall fold of the molecule. This indicates that this disulfide bridge is not essential for the EGF structural motif, even though it is well conserved across the family of EGF sequences (Campbell & Bork, 1993; Groenen et al., 1994). [Abu6, 20] mEGF4-48 was not as thermodynamically stable as the native molecule, as shown by its lower thermal denaturation temperature, but it was possible to obtain a well-defined solution structure under similar conditions of pH and temperature to those used for the native structure (Montelione et al., 1992).

As summarized in the Introduction, in many small proteins containing multiple disulfides, the deletion of a single disulfide caused only localized changes in the structure, leaving the overall fold largely unaffected. The biological activity was usually more sensitive than the conformation to the loss of a disulfide. As it was often the case, analogues lacking individual disulfides were less stable than their native counterparts (as reflected in their greater susceptibility to denaturation), it is reasonable to ascribe at least some of this loss of activity to greater conformational entropy in the unbound analogue, with a corresponding reduction in the free energy of binding to its target receptor. In the case of [Abu6, 20] mEGF4-48, some of its loss of activity can be attributed to this factor, as it was clearly less stable than native mEGF. However, as discussed below, some of the local structural changes in [Abu6, 20] mEGF4-48 had a significant effect on functionally important residues, and these changes are probably the major cause of the activity loss.

Deletion of the 6 to 20 disulfide bridge in mEGF had no significant effect on the smaller two-stranded β-sheet near the C-terminus, consistent with their spatial separation in the structure (Fig. 7C). The two major strands of the first antiparallel β-sheet (residues 19–23 and 28–32) were also largely unaffected, presumably because of the large number of backbone hydrogen bonds and side-chain to side-chain interactions stabilizing this region. In the native structure, the first strand of this sheet (residues 2–4) is loosely anchored to the second strand by two hydrogen bonds between residues 3 and 22. Truncation of the first three residues in [Abu6, 20] mEGF4-48 effectively eliminates the N-terminal strand of this sheet, but, as has been noted previously (Simpson et al., 1985; Burgess et al., 1988; Groenen et al., 1994), and confirmed again in these studies, does not significantly affect the biological activity of mEGF. It, therefore, follows that the integrity of this strand per se is not necessary for activity. The finding that [G1, M3, K21, H40] mEGF1-48 has full mitogenic activity despite changes in two of the first three residues further supports this conclusion. The significant losses of receptor binding affinity and mitogenic activity on Balb/c 3T3 cells for [Abu6, 20] mEGF4-48, therefore, must be a direct result of the conformational changes associated with deletion of the disulfide bridge. This result is in keeping with a recent NMR study of the interaction of the structurally similar TGF-α with the EGF receptor (McInnes et al., 1996), which showed that the three N-terminal residues (which also formed the third strand of the major β-sheet) were not involved in receptor binding.

The structures of the polypeptide backbones of [Abu6, 20] mEGF4-48 and mEGF differ significantly up to Ser8, beyond which they superimpose quite well (with the exception of some residues in the loops containing reverse turns). In principle, the structural change at the N-terminus could result from the N-terminal truncation or the loss of the disulfide (or both), but the fact that N-terminal truncation did not cause any loss of activity makes it likely that the loss of the disulfide is the major cause of the perturbation. It appears that the disulfide anchors the N-terminal region to one face of the main β-sheet in mEGF and in its absence the remaining interactions are too weak to maintain the native structure. The loss of hydrogen bonding and other interactions involving the three N-terminal residues in mEGF undoubtedly makes some contribution to the loss of native structure in [Abu6, 20] mEGF4-48, but it is unlikely that their presence would have overcome the destabilisation associated with loss of the disulfide. Following Ser9, hydrophobic interactions involving a small cluster of aromatic residues (Tyr10, Tyr13, and Tyr29) and the presence of the 14 to 31 disulfide bridge stabilise the structure in its native fold.

There are some minor structural differences in [Abu6, 20] mEGF4-48 in the region of Ile35. For example, the exchange rate of the backbone amide of Ile35 was the slowest of any in mEGF (Montelione et al., 1988), but in [Abu6, 20] mEGF4-48 it had an intermediate exchange rate (Fig. 4). There were also some chemical shift differences from native in this region in [Abu6, 20] mEGF4-48 and [G1, M3, K21, H40] mEGF1-48 (Fig. 3), as well as mEGF1-48 (data not shown). As Ile35 is located on a loop some distance away from the 6 to 20 disulfide (Fig. 7), and as the chemical shift changes are evident in all three C-terminally truncated analogues, these changes must reflect the loss of the C-terminal residues rather than the disulfide. Thus, even though the C-terminus of mEGF is unstructured, there

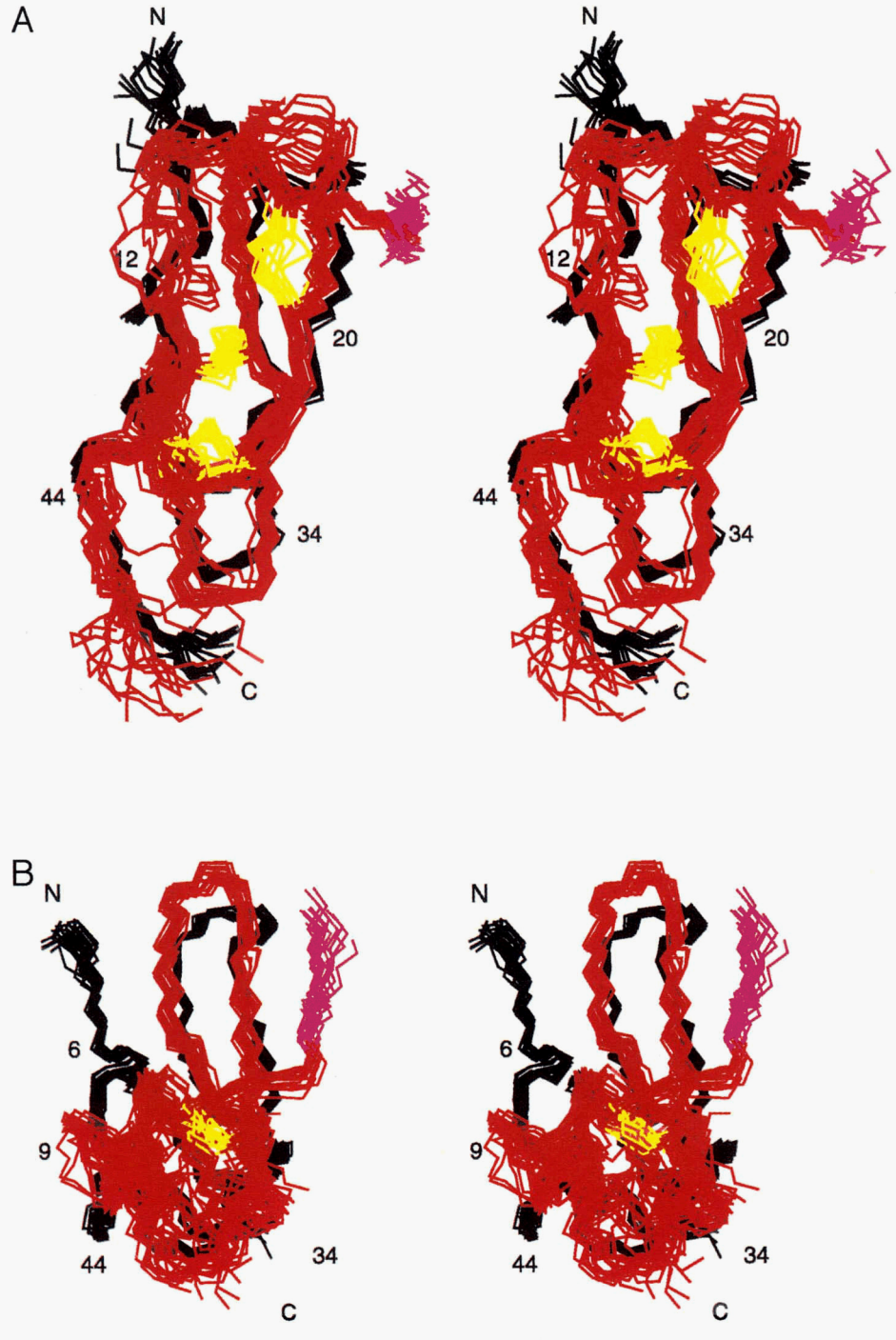

Fig. 7. A: Stereo views of the final 20 structures of [Abu6, 20] mEGF4-48 (black), and the 16 structures of native mEGF (red) from Montelione et al. (1992) (Protein Data Bank accession no. 3EGF), superimposed over the backbone heavy atoms (N, C^{α}, C) of the well-defined (S_{ϕ} and $S_{\psi} > 0.8$) region of the molecule, encompassing residues 6–46. In mEGF the first three residues, for which there are no counterparts in [Abu6, 20] mEGF4-48, are colored pink and the last five residues, which are highly disordered, are not shown. This view is identical to that shown in Figure 6. **B:** An alternative view highlighting the differences in the direction of the polypeptide chain at the N-termini of the two molecules. **C:** Ribbon diagrams of native mEGF (left) and [Abu6, 20] mEGF4-48 (right). The β-sheet strands, disulfides, and the side chain of the functionally important Arg41 are highlighted in different shades of grey. In each case the individual structure closest to the average over residues 6–46 (the well-defined region of [Abu6, 20] mEGF4-48) is shown. (*Figure continues on facing page.*)

is apparently some weak interaction with the region around Ile35, which has the effect of reducing its solvent exposure in mEGF (see Fig. 7).

Given that [G1, M3, K21, H40] mEGF1-48 and mEGF1-48 are essentially equipotent with native mEGF in mitogenic and receptor binding assays, what causes the loss of activity associ-

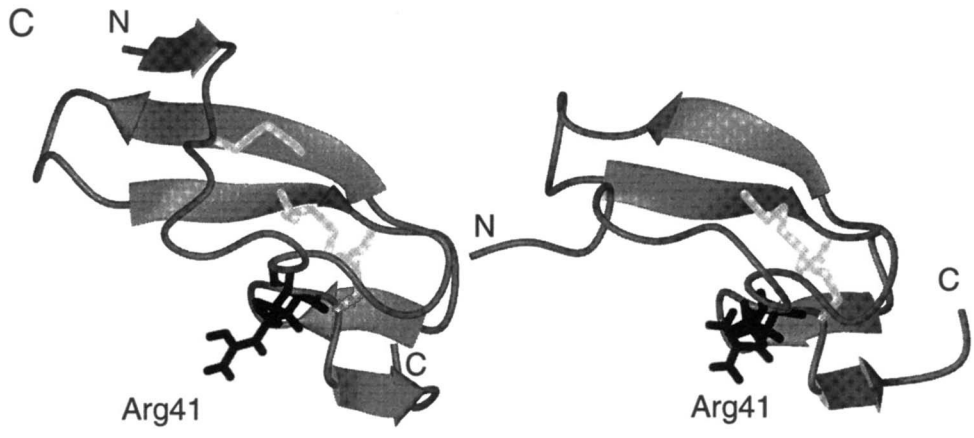

Fig. 7. *Continued.*

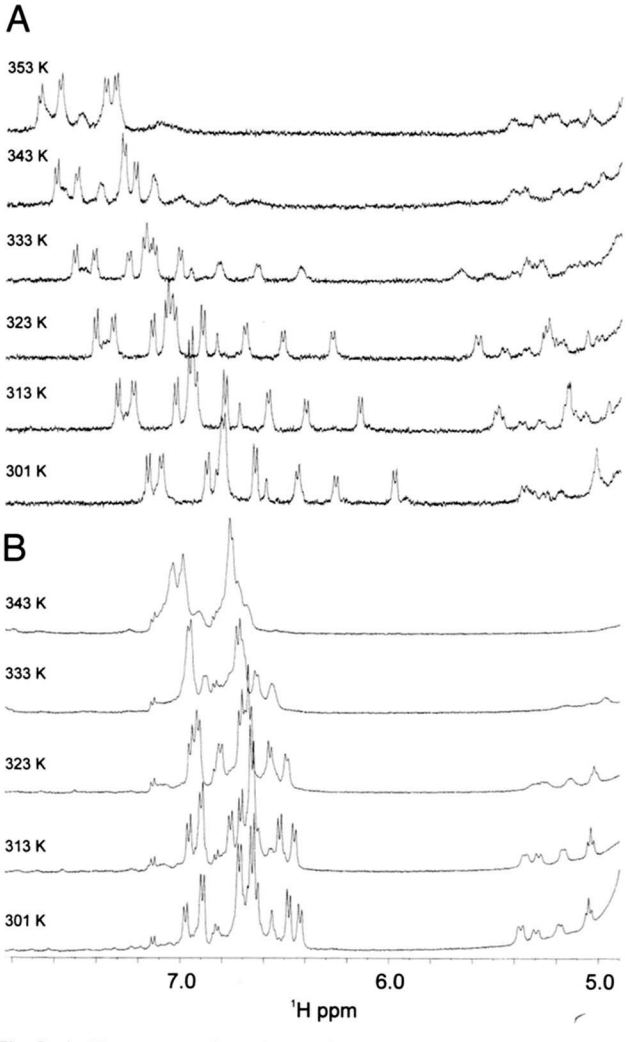

Fig. 8. A: Temperature dependence of one-dimensional spectra recorded at 500 MHz of mEGF1-48 (20 μM) at pH 3.4 in ^2H_2O. These spectra demonstrate that thermal denaturation of mEGF1-48 begins by 333 K but that some native-like structure is still present at 353 K. **B:** Temperature dependence of one-dimensional spectra at 500 MHz of [Abu6, 20] mEGF4-48 (50 μM) at pH 3.0 in ^2H_2O. These spectra indicate that the thermal denaturation of [Abu6, 20] mEGF4-48 begins by 323 K and is complete by 343 K.

ated with deletion of the 6 to 20 disulfide? Structure-function studies on mEGF and related proteins (Groenen et al., 1994) suggest that almost the entire EGF structure is required for effective receptor binding, although some residues such as Arg41 and Leu47 have been identified as being particularly important. Our structure for [Abu6, 20] mEGF4-48 is quite well defined at the C-terminus, but the corresponding region of the mEGF structure is less ordered (Fig. 7), so detailed comparisons of the local environment around Leu47 are not warranted. In the case of Arg41, however, it is clear that the dislocation of the N-terminal region in [Abu6, 20] mEGF4-48 may alter its local environment. Perhaps more importantly, the proximity of the N-terminal region of [Abu6, 20] mEGF4-48 to the receptor binding surface in the vicinity of Arg41 may interfere with the receptor binding of this analogue.

It is now generally accepted that the binding interface of EGF and TGFα encompasses a large part of the surface of the molecule, and that these growth factors might be anchored to the receptor by multiple domains (Groenen et al., 1994; McInnes et al., 1996). One of these binding surfaces appears to comprise one complete side of the molecule, with contributions from residues of both the N- and C-terminal domains (Tyr13, Ile23, Thr30, and Arg41 in EGF). The major focus of point mutation studies has been the strongly conserved residues Tyr/Ile13, Tyr37, Arg41, and Leu47, as well as residues 19–32, which comprise the main β-sheet. Changes in the eight most N-terminal residues or truncation of the first three residues of EGF, have, with the exception of Cys6, not significantly affected biological activity. In TGFα, deletion of residues 1–7 (Tam et al., 1991) or His/Lys substitution at position 4 (Defeo-Jones et al., 1989) reduced the biological activity significantly (to $<$10%). However, antibodies recognizing the N-terminus of TGFα are not neutralizing (Hoeprich et al., 1989), implying that these residues do not make contact with the receptor, and in recent NMR studies on the TGFα/EGFR (Hoyt et al., 1994; McInnes et al., 1996) the N-terminal residues were shown to have little direct involvement in receptor binding.

In conclusion, the results reported here extend our understanding of the relationship between structure and function in EGF. It is clear that the 6 to 20 disulfide bridge is critical in maintaining the biologically active structure of EGF even though the N-terminal residues themselves are not essential for activity. The role of this structural restraint will need to be taken into account in any

attempt at polypeptide minimisation (Cunningham & Wells, 1997), undertaken with the goal of reducing the receptor binding surface of EGF to its bare minimum without loss of potency.

Materials and methods

Materials

The resin used for EGF syntheses was Boc-Arg(Tos)-OCH_2-Pam resin (0.58 mmol/g), supplied by Applied Biosystems (Foster City, CA). Protected Boc amino acid derivatives were from the Peptide Institute (Osaka, Japan) or Applied Biosystems. The following side-chain protected amino acids were used: Arg(Tos), Asp(OcHex), Cys (4-MeBzl), Glu(OcHex), His(DNP), Lys(ClZ), Ser(Bzl), Thr(Bzl), and Tyr(BrZ). All other Boc-amino acids used were side chain unprotected. DMF, dichloromethane, N,N-diisopropyl ethylamine, TFA, dicyclohexylcarbodiimide, and hydroxybenzotriazole were all peptide synthesis grade from Auspep (Melbourne, Australia). HPLC-grade methanol was supplied by Millipore Waters, USA. Acetonitrile (Hipersolv-Far UV grade) was from BDH (Poole, UK). Water was obtained from a tandem Millipore Milli-RO–Milli-Q system. p-Cresol and p-thiocresol were from Fluka (Germany). HF was supplied by BOC Gases (Brisbane, Australia). Guanidinium hydrochloride (99+%) and Trizma base (reagent grade, 99.9+%) were from Sigma. ^2H_2O, NaO_2H, and ^2HCl for NMR were obtained from Cambridge Isotope Laboratories (Andover, MA).

Peptide synthesis

Machine-assisted stepwise chain assembly of mEGF analogues was performed on Pam resin (0.5 mmol scale) using the Boc/benzyl strategy with an ABI 430A protein synthesizer that was modified by removal of in-line filters to the top and bottom of the reaction vessel. Symmetric anhydride or active ester chemistries were used to couple Boc-protected amino acids to the resin. The Boc protecting group was removed using 100% TFA; DMF was used as both the coupling solvent and for flow washes throughout the cycle. Each residue was routinely double-coupled and the progress of the chain assembly evaluated by quantitative ninhydrin monitoring. Where necessary (less than 99.5% coupling), the synthesis was halted and a further coupling was carried out manually using the HBTU (or HATU) in situ neutralization method (Schnolzer et al., 1992). If this repeat coupling was less than 99.5% complete, acetylation of the remaining amino groups was carried out using acetic anhydride (87 μL/mL) in DMF.

Deprotection and cleavage

The fully protected peptide-resin was first subjected to thiolysis (20% β-mercaptoethanol, 10% N,N-diisopropyl ethylamine in DMF, 2 × 30 min) to remove histidine-DNP groups. The N-α-Boc group was then removed (TFA, 2 × 1 min), and the peptide-resin was washed with DMF, dichloromethane, and dried under nitrogen. Cleavage from the resin and simultaneous deprotection of side chains was carried out in an HF cleavage apparatus (Peptide Institute, Japan) using liquefied HF in the presence of scavengers (9.0:0.5:0.5 by volume, HF:p-cresol:p-thiocresol) for 1.5 h at −7° to −2 °C. After removal of HF under vacuum, the peptide was precipitated with cold, oxygen-free, ethyl acetate. The resultant precipitate and residual resin were collected on a filter, and washed thoroughly with more cold, oxygen-free, ethyl acetate to remove traces of scavenger and scavenger adducts. The crude, reduced peptide mixture was dissolved in 50% aqueous acetic acid and collected by filtration. The solution was then diluted to 5% (v/v) aqueous acetic acid before lyophilization.

Folding and oxidation

Crude reduced peptide was dissolved in 6 M guanidinium hydrochloride/100 mM Tris (pH 8.3), then diluted to 2–3 M guanidinium hydrochloride with water. The solution (final peptide concentration ca. 2 mg/mL) was then stirred at ambient temperature in air. Samples were taken periodically and analyzed by RP-HPLC. Eluant fractions were collected for electrospray mass spectrometric analysis. The presence of free sulfhydryl groups was confirmed using Ellman's reagent. The oxidation was stopped by acidification with TFA when the HPLC profile was unchanged between two runs, and both MS analysis and the Ellman test indicated that disulfide bond formation was complete.

Chromatographic purification

Analytical HPLC was carried out on a Brownlee RP300 (C8, 300 Å) analytical cartridge (100 mm × 4.6 mm ID) on a Waters 600 HPLC system. Samples (5 μL, 1 mg/mL peptide dissolved in 45% aqueous acetonitrile/0.1% TFA) were injected via a Waters 715 WISP autoinjector. Chromatography was carried out using a linear gradient of 0–67% B at 1 mL/min over 60 min, where buffer A was 0.15% aqueous TFA and buffer B was 90% aqueous acetonitrile/0.125% TFA. Detection was at 215 nm.

Preparative HPLC was performed using a Vydac C4 300 Å preparative column (30 cm × 2.2 cm ID) on a Waters Delta Prep chromatography system with detection at 230 nm. The solution containing the oxidized peptide was adjusted to pH 2 using TFA, before being loaded directly onto the column. The column was washed with 0.1% aqueous TFA until the oxidizing buffer components were completely eluted. Retained peptides were recovered using a linear 80 min gradient at a flow rate of 8 mL/min between 0.1% aqueous TFA and 90% aqueous acetonitrile/0.09% TFA. Fractions were collected at 30 s intervals for MS analysis.

Mass spectrometry

Mass spectrometric measurements were performed on a Sciex *APIIII* Ion Spray mass spectrometer at an orifice potential of 80 V. Data were collected in the positive ion mode by accumulation of data in the mass range 400–2,100 amu from several scans using a scan step of 0.1 amu, and a delay time of 0.3 s. Peptides were dissolved at a concentration of 1 mg/mL in 45% aqueous acetonitrile containing 0.1% TFA. HPLC fractions were used without further treatment. Samples were delivered to the orifice via a glass capillary by direct injection (5–20 μL) using a Rheodyne injector into a 30–50 μL/min solvent flow of 50% aqueous acetonitrile containing 0.05% TFA. The resultant data were subjected to deconvolution (Hypermass-MacSpec 3.2, SCIEX, Canada) to determine the M_r of the observed protonated species.

Mitogenic and receptor binding assays

Analogues for analysis were purified by micro-preparative RP-HPLC as described previously (Nice, 1996). Aliquots were taken

simultaneously for quantitative amino acid analysis, mass spectrometric, receptor binding, and mitogenic assays (Groenen et al., 1994).

Mitogenic assays were performed using Balb/c 3T3 mouse fibroblasts. Briefly, the cells were grown in 175 cm^2 tissue culture flasks in Dulbecco's Modified Eagles Medium (DMEM) and 10% newborn calf serum (NBS) at 37 °C in a water jacketed incubator equilibrated with 5% CO_2. Cells were washed with mouse tenicity phosphate buffered saline, trypsinized, and resuspended in 10 mL DMEM + 10% NBS. Cells were counted in a haemocytometer and cell number adjusted to 50,000 cells/mL.

Tissue culture plates (96 well) were seeded with 200 μL of cell suspension using a Biomek 2000 automated laboratory workstation (Beckman, California) and grown for 3 days until confluent. The medium was changed to DMEM + 0.5% NBS (200 μL) and the cells incubated overnight. EGF containing samples for analysis (10 μL) were made up to 200 μL with media and added in duplicate to the first wells of the plate. A 200 μL aliquot was then taken from these wells and serially titrated across the plate. The plates were then incubated for a further 20 h before the addition of 10 μL (0.2 μCi) of ^{3}H-thymidine per well. After a further 2 h incubation, the cells were lysed with 0.5 M NaOH and the contents harvested onto glass filter mats using a LKB Wallac (Turku, Finland) 1295-001 automatic cell harvester. The filter mats were then allowed to dry, inserted into sample bags, scintillation fluid (10 mL) was added and incorporated ^{3}H-thymidine determined using an LKB 1205 liquid scintillation counter.

Receptor binding assays were performed on the human epidermoid carcinoma cell line A431 (American Type Culture Collection) using the protocol described previously for Balb/c 3T3 fibroblasts (Burgess et al., 1988; Groenen et al., 1994).

NMR spectroscopy

The pH of the NMR samples was adjusted with small additions of 0.5 M NaO^2H or ^2HCl. Reported pH values were measured at room temperature and were not corrected for isotope or solvent effects. The ^{1}H chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate at 0 ppm, via the chemical shift of the H_2O resonance (Wishart et al., 1995b) or an impurity at 0.15 ppm.

Spectra were recorded on Bruker AMX-600 and AMX-500 spectrometers. Unless stated otherwise, all spectra were recorded at 300 K and pH 2.8, with probe temperatures calibrated according to the method of Van Geet (1970). All two-dimensional spectra were recorded in phase-sensitive mode using time-proportional phase incrementation (Marion & Wüthrich, 1983). Solvent suppression was achieved by selective, low-power irradiation of the water signal during the relaxation delay (typically 2 s) and during the mixing time in NOESY experiments.

Two-dimensional homonuclear NOESY spectra (Anil-Kumar et al., 1980; Macura et al., 1981) were recorded with mixing times of 50 and 250 ms. TOCSY spectra (Braunschweiler & Ernst, 1983) were recorded using the DIPSI-2 spin-lock sequence (Rucker & Shaka, 1989) with spin-lock times of 70–80 ms. DQF-COSY (Rance et al., 1983) and E-COSY (Griesinger et al., 1987) spectra were also recorded. Typically, spectra were acquired with 400–600 t_1 increments, 32–128 scans per increment, and 4,096 data points. The ^{1}H sweep width was 7,575.6 Hz at 600 MHz. Spectra were processed using UXNMR-941001.4 (Bruker) and analyzed using XEASY-1.3.7 (Bartels et al., 1995). Sine-squared window functions, phase shifted by 60°–90°, were applied in both dimensions prior to Fourier transformation.

The $^3J_{NHC\alpha H}$ coupling constants were measured from a DQF-COSY spectrum. The appropriate rows were extracted from the spectrum, inverse Fourier transformed, zero-filled to 32 K data points, and multiplied by a Gaussian window function prior to Fourier transformation. The dispersive peak shapes were simulated to take account of the effect of broad line widths on small coupling constants, using an in-house program COUPLING. Slowly exchanging amide protons were identified by dissolving the lyophilized protein in ^2H_2O at pH 2.9 and 300 K and recording a series of one-dimensional and TOCSY spectra immediately after dissolution and over the next 24 h. $^3J_{C\alpha HC\beta H}$ coupling constants were measured from an E-COSY spectrum.

Structural constraints

NOESY cross-peak volumes measured from a 250-ms mixing time spectrum were used to calculate upper bound distance restraints. Peaks from the upper side of the diagonal were used except where peaks from the lower side were better resolved. Peak volumes were calibrated using an average of the volumes of six well-resolved geminal $C^\beta H$ cross-peaks (distance 1.79 Å). Distance constraints were calculated using volumes proportional to r^{-6} for backbone/backbone constraints and r^{-4} for all others, in the program CALIBA (Güntert et al., 1991), which corrects automatically for degenerate and pseudo-atoms. Where cross-peak volumes could not be estimated reliably, an upper bound of 5 Å was assigned. Corrections of 0.5 and 1.0 Å were added to distance constraints involving only backbone protons and at least one side-chain proton, respectively, to allow for conformational averaging and errors in volume integration.

A small number of lower bound restraints >1.8 Å (Manoleras & Norton, 1994) was also used for NH, $C^\alpha H$, and $C^\beta H$ atoms in the final structure calculations. These were obtained with an in-house program that checked the initial (DYANA) structures for distances <3.5 Å that were not represented in the NOE restraint list. Distances identified in this way were compared with the experimental NOESY spectra to confirm that cross-peaks could have been observed had they been present. Where a cross-peak was clearly absent, a lower bound restraint of 3.5 Å was added to the restraint list; for all other NH, $C^\alpha H$, and $C^\beta H$ atoms, the lower limit was 1.79 Å.

Backbone dihedral angle constraints were inferred from $^3J_{NHC\alpha H}$ values as follows: $^3J_{NHC\alpha H} < 5$ Hz, $\phi = -60° \pm 30°$; 5 Hz $< {}^3J_{NHC\alpha H} < 6$ Hz, $\phi = -60° \pm 40°$; $^3J_{NHC\alpha H} > 8$ Hz, $\phi = -120° \pm 40°$ (Wüthrich, 1986). Where 6 Hz $< {}^3J_{NHC\alpha H} < 8$ Hz, and the possibility of positive ϕ angles had been excluded by the NOESY spectrum (Ludvigsen & Poulsen, 1992), ϕ was constrained to $-120° \pm 60°$; otherwise ϕ angles were not constrained. Where possible, $^3J_{C\alpha HC\beta H}$ coupling constants were measured from passive couplings as displacements in E-COSY spectra or peaks splittings in DQF-COSY spectra. The relative intensities of intraresidue $d_{\alpha\beta}(i,i)$ and $d_{N\beta}(i,i)$ NOEs were measured in a 50-ms mixing time NOESY spectrum. The $^3J_{C\alpha HC\beta H}$ coupling constants, together with the NOE intensities, were used to determine if side chains could be placed in one of the three staggered side-chain rotamer conformations ($\chi^1 = -60°$, 60°, or 180°) and to make stereospecific assignments (Wagner et al., 1987).

Structure calculations

Initial structures were generated with the torsion angle dynamics program DYANA, version 1.0 (Güntert et al., 1997), using dihedral

angle constraints derived from coupling constant data and distance constraints derived from NOE cross-peaks assigned unambiguously in both chemical shift dimensions. No hydrogen-bonding restraints were used at any stage in the calculations. Ambiguous NOESY cross-peak assignments were resolved, where possible, using these initial structures. An assignment was accepted if in all structures the appropriate inter-proton distance was <5 Å, and the distance between alternative pairs was >7 Å. A small number of lower bound restraints >1.8 Å was also introduced prior to the final structure calculations.

Once the final set of restraints had been obtained, a new family of structures was generated using DYANA, and the 50 structures with the lowest penalty functions were refined by simulated annealing in X-PLOR, version 3.1 (Brünger, 1992). Simulated annealing (Nilges et al., 1988) was performed using 20,000 steps at 1,000 K and 10,000 steps as the molecule was gradually cooled to 300 K. A time step of 1 fs was employed throughout. The 50 structures were then subject to further simulated annealing in which they were gradually cooled from 300 to 0 K in 20,000 steps, and then energy minimized using 100 steps of Powell conjugate gradient minimization. For each structure, this procedure was carried out 10 times, and the best of these 10 in terms of total energy and NOE energies was selected. The structures were then energy minimized in the empirical CHARMM force field (Brooks et al., 1983), with all explicit charges neutralized (Monks et al., 1995) and with a distance-dependent dielectric instead of explicit water molecules. The 20 best structures based on their stereochemical energies (i.e., the sum of all contributions to the calculated energy except the electrostatic term) and NOE energies were chosen for structural analysis. These structures and the restraints from which they were generated have been submitted to the PDB (accession number 1a3p).

Structures were analyzed using MOLMOL (2.1.0) (Koradi et al., 1996). Hydrogen bonds were identified in MOLMOL, using a maximum C-N distance of 2.4 Å and a maximum angle of 35°. Structural figures were generated using Insight II and MOLMOL.

Acknowledgments

We thank Paul Pallaghy and Mark Hinds for assistance and advice, Marilyn Olliff and John MacFarlane for assistance with computing, David Craik for provision of the program COUPLING, Peter Güntert for provision of the program DYANA-1.0, Maureen Nerrie for bioassay of the EGF analogues, Francesca Walker for advice on the receptor binding assays, and Tony Burgess for helpful advice and discussion. This work was supported in part by the Cooperative Research Centre for Cellular Growth Factors.

References

Anil-Kumar, Ernst RR, Wüthrich K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun 95*:1–6.

Arriaga Y, Harville BA, Dreyfus LA. 1995. Contribution of individual disulfide bonds to biological action of *Escherichia coli* heat-stable enterotoxin B. *Infect Immun 63*:4715–4720.

Barbar E, Barany G, Woodward C. 1995. Dynamic structure of a highly ordered β-sheet molten globule: Multiple conformations with a stable core. *Biochemistry 34*:11423–11434.

Bartels C, Xia T-H, Billeter M, Güntert P, Wüthrich K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR 5*:1–10.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, 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.

Betz SF. 1993. Disulfide bonds and the stability of globular proteins. *Protein Sci 2*:1551–1558.

Braunschweiler L, Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson 53*:521–528.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comput Chem 4*:187–217.

Brünger AT. 1992. *X-PLOR Version 3.1. A system for X-ray crystallography and NMR*. New Haven, CT: Yale University.

Burgess AW. 1989. Epidermal growth factor and transforming growth factor alpha. *Br Med Bull 45*:401–424.

Burgess AW, Lloyd CJ, Smith S, Stanley E, Walker F, Fabri L, Simpson RJ, Nice EC. 1988. Murine epidermal growth factor: Structure and function. *Biochemistry 27*:4977–4985.

Campbell ID, Bork P. 1993. Epidermal growth factor-like modules. *Curr Opin Struct Biol 3*:385–392.

Chuang L-C, Chen P-Y, Chen C-P, Huang T-H, Wang K-T, Chiou S-H, Wu S-H. 1996. Structural analysis of a biologically active echistatin analogue des(46-49)-[Ala(8,37)]-echistation γ with three disulfide bonds by 2D-NMR and computer graphics. *Biochem Biophys Res Commun 220*:246–254.

Cooke RM, Wilkinson AJ, Baron M, Pastore A, Tappin MJ, Campbell ID, Gregory H, Sheard B. 1987. The solution structure of human epidermal growth factor. *Nature 327*:339–341.

Cunningham BC, Wells JA. 1997. Minimized proteins. *Curr Opin Struct Biol 7*:457–462.

Defeo-Jones D, Tai JY, Vuocolo GA, Wegrzyn RJ, Schofield TL, Riemen MW, Olliff A. 1989. Substitution of lysine for arginine at position 42 of human transforming growth factor-alpha eliminates biological activity without changing internal disulfide bonds. *Mol Cell Biol 9*:4083–4086.

Doig AJ, Williams DH. 1991. Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. *J Mol Biol 217*:389–398.

Eigenbrot C, Randall M, Kossiakoff AA. 1990. Structural effects induced by removal of a disulfide-bridge: The X-ray structure of the C30A/C51A mutant of basic pancreatic trypsin inhibitor at 1.6 Å. *Protein Eng 3*:591–598.

Flory PJ. 1956. Theory of elastic mechanisms in fibrous proteins. *J Am Chem Soc 78*:5222–5235.

Griesinger C, Sørenson O, Ernst RR. 1987. Practical aspects of the E-COSY technique. Measurement of scalar spin-spin coupling constants in peptides. *J Magn Reson 75*:474–492.

Groenen LC, Nice EC, Burgess AW. 1994. Structure-function relationships for the EGF/TGF-α family of mitogens. *Growth Factors 11*:235–257.

Güntert P, Braun W, Wüthrich K. 1991. Efficient computation of three-dimensional protein structures in solution from NMR data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol 217*:517–530.

Güntert P, Mumenthaler C, Wüthrich K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol 273*:283–298.

Harvey TS, Wilkinson AJ, Tappin MJ, Cooke RM, Campbell ID. 1991. The solution structure of human transforming growth factor α. *Eur J Biochem 198*:555–562.

Hoeprich PD Jr, Langton BC, Zhang JW, Tam JP. 1989. Identification of immunodominant regions of transforming growth factor alpha. Implications of structure and function. *J Biol Chem 264*:19086–19091.

Hommel U, Harvey TS, Driscoll PC, Campbell ID. 1992. Human epidermal growth factor. High resolution solution structure and comparison with human transforming growth factor α. *J Mol Biol 227*:271–282.

Hoyt DW, Harkins RN, Debanne MT, O'Connor-McCourt M, Sykes BD. 1994. Interaction of transforming growth factor alpha with the epidermal growth factor receptor: Binding kinetics and differential mobility within the bound TGF-alpha. *Biochemistry 33*:15283–15292

Hua Q-X, Narhl L, Jia W, Arakawa T, Rosenfeld R, Hawkins N, Miller JA, Weiss MA. 1996. Native and non-native structure in a protein-folding intermediate: Spectroscopic studies of partially reduced IGF-1 and an engineered alanine model. *J Mol Biol 259*:297–313.

Hyberts SG, Goldberg MS, Havel TF, Wagner G. 1992. The solution structure of eglin C based on measurement of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci 1*:736–751.

Kohda D, Inagaki F. 1992a. Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions. *Biochemistry 31*:11928–11939.

Kohda D, Inagaki F. 1992b. Structure of epidermal growth factor bound to perdeuterated dodecylphosphocholine micelles determined by two-dimensional NMR and simulated annealing calculations. *Biochemistry 31*:677–685.

Koradi R, Billeter M, Wüthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics 14*:51–55.

Le-Nguyen D, Heitz A, Chiche L, El Hajji M, Castro B. 1993. Characterization and 2D NMR study of the stable [9–21, 15–27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II. *Protein Sci 2*:165–174.

Ludvigsen S, Poulsen FM. 1992. Positive ϕ-angles in proteins by nuclear magnetic resonance and distance geometry. *J Biomol NMR 2*:227–233.

Macura S, Huang Y, Suter D, Ernst RR. 1981. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J Magn Reson 43*:259–281.

Manoleras N, Norton RS. 1994. Three-dimensional structure in solution of neurotoxin III from the sea anemone *Anemonia sulcata*. *Biochemistry 33*:11051–11061.

Marion D, Wüthrich K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun 113*:967–974.

McInnes C, Hoyt DW, Harkins RN, Pagila RN, Debanne MT, O'Connor-McCourt M, Sykes BD. 1996. NMR study of the transforming growth factor-α (TGF-α)-epidermal growth factor receptor complex. *J Biol Chem 271*:32204–32211.

Monks SA, Pallaghy PK, Scanlon MJ, Norton RS. 1995. Solution structure of the cardiostimulant polypeptide anthopleurin-B and comparison with anthopleurin-A. *Structure 3*:791–803.

Montelione GT, Wüthrich K, Burgess AW, Nice EC, Wagner G, Gibson KD, Scheraga HA. 1992. Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization and restraints. *Biochemistry 31*:236–249.

Montelione GT, Wüthrich K, Nice EC, Burgess AW, Scheraga HA. 1986. Identification of two anti-parallel β-sheet conformations in the solution structure of murine epidermal growth factor by proton magnetic resonance. *Proc Natl Acad Sci USA 83*:8594–8598.

Montelione GT, Wüthrich K, Scheraga HA. 1988. Sequence-specific ^{1}H NMR assignments and identification of slowly exchanging amide protons in murine epidermal growth factor. *Biochemistry 27*:2235–2243.

Nice EC. 1996. Micro-preparative HPLC of proteins and peptides: Principles and applications. *Biopolymers Peptide Sci 40*:319–341.

Nilges M, Clore GM, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett 229*:317–324.

Pallaghy PK, Duggan BM, Pennington MW, Norton RS. 1993. Three-dimensional structure in solution of the calcium channel blocker ω-conotoxin. *J Mol Biol 234*:405–420.

Pan H, Barbar E, Barany G, Woodward C. 1995. Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry 34*:13974–13981.

Rance M, Sørenson OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun 117*:479–485.

Richardson J. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–339.

Rucker SP, Shaka AJ. 1989. Broadband homonuclear cross-polarisation in 2D NMR using DIPSI-2. *Mol Phys 68*:509–517.

Sabatier J-M, Lecomte C, Mabrouk K, Darbon H, Oughideni R, Canarelli S, Rochat H, Martin-Eauclaire M-F, Van Rietschoten J. 1996. Synthesis and characterization of leiurotoxin I analogs lacking one disulfide bridge: Evidence that disulfide bridge 3–21 is not required for full toxin activity. *Biochemistry 35*:10641–10647.

Savage CR Jr, Inagami T, Cohen SJ. 1972. The primary structure of epidermal growth factor. *J Biol Chem 247*:7612–7621.

Schnolzer M, Alewood P, Jones A, Alewood D, Kent SBH. 1992. In situ neutralization in Boc-chemistry solid phase peptide synthesis. *Int J Peptide Protein Res 40*:180–193.

Simpson RJ, Smith JA, Moritz RL, O'Hare MJ, Rudland PS, Morrison JR, Lloyd CJ, Grego B, Burgess AW, Nice EC. 1985. Rat epidermal growth factor: Complete amino acid sequence. *Eur J Biochem 153*:629–637.

Song J, Gilquin B, Jamin N, Drakopoulo E, Guenneugues M, Dauplais M, Vita C, Menez A. 1997. NMR solution structure of a two-disulfide derivative of charybdotoxin: Structural evidence for conservation of scorpion toxin α/β motif and its hydrophobic side chain packing. *Biochemistry 36*:3760–3766.

Tam JP, Lin YZ, Liu W, Wang DX, Ke XH, Zhang JW. 1991. Mapping the receptor-recognition site of human transforming growth factor-alpha. *Int J Pept Protein Res 38*:204–211

Tejero R, Bassolino-Klimas D, Bruccoleri RE, Montelione GT. 1996. Simulated annealing with restrained molecular dynamics CONGEN: Energy refinement of the NMR solution structures of epidermal and type-a transforming growth factors. *Protein Sci 5*:578–592.

Thornton JA. 1981. Disulfide bridges in globular proteins. *J Mol Biol 151*:261–287.

Van Geet AL. 1970. Calibration of methanol nuclear magnetic resonance thermometer at low temperature. *Anal Chem 42*:679–680.

van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE. 1991. (14–38, 30–51) Double-disulfide intermediate in folding of bovine pancreatic trypsin inhibitor: A two-dimensional ^{1}H nuclear magnetic resonance study. *J Mol Biol 222*:353–371.

Wagner G, Braun W, Havel TF, Schaumann T, Gō N, Wüthrich K. 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol 196*:611–639.

White CE, Hunter MJ, Meininger DP, Garrod S, Komives EA. 1997. The fifth epidermal growth factor-like domain of thrombomodulin does not have an epidermal growth factor-like disulfide bonding pattern. *Proc Natl Acad Sci USA 93*:10177–10182.

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995a. ^{1}H, ^{13}C, and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbour effects. *J Biomol NMR 5*:67–81.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson J, Oldfield E, Markley JL, Sykes BD. 1995b. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR 6*:135–140.

Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*:1647–1651.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: Wiley.