Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5226–5230, August 1987
Biophysics

Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry

(protein structure determination/oncogenesis/molecular design/transforming growth factor)

GAETANO T. MONTELIONE†, KURT WÜTHRICH‡, EDOUARD C. NICE§, ANTONY W. BURGESS§, AND HAROLD A. SCHERAGA†

†Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301; ‡Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland; and §Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, Victoria 3050, Australia

Contributed by Harold A. Scheraga, April 17, 1987

ABSTRACT **The polypeptide backbone fold in the solution structure of murine epidermal growth factor has been determined by nuclear magnetic resonance spectroscopy and distance geometry calculations. The results are based on nearly complete sequence-specific resonance assignments and on 333 distance and dihedral-angle constraints; these were determined from nuclear Overhauser effect measurements, identification of hydrogen-bonded amide protons, the known locations of disulfide bonds, and backbone vicinal spin–spin coupling constants. The polypeptide chain of the protein is arranged into two distinct domains. The structures of these domains were determined independently in separate calculations and then combined to obtain an overall view of the protein. The backbone fold thus determined includes the regular backbone structure elements that were previously identified using different techniques for the analysis of the nuclear magnetic resonance data. The distance geometry calculations also provided additional details about the conformations of bends and loops and about the twists of the β-sheets.**

Epidermal growth factor (EGF) is a small single-chain protein containing 53 amino acids and three disulfide bonds (1–3). EGF and the homologous α-type transforming growth factor (TGF-α) appear to play an important role in the molecular mechanisms controlling mammalian cell growth, oncogenesis (4, 5), and wound healing (6). For this reason, there is much interest in structure–function studies of EGF and EGF-like molecules. Several laboratories have cloned genes for human EGF (7–9), and others have synthesized either polypeptide fragments of EGF (10–12) or the entire EGF protein (13). With this technology, it may be possible to design and engineer EGF-like molecules with desirable biological activities, such as growth factor antagonists or agonists. Knowledge of the three-dimensional structure will be indispensable for understanding the structural basis of EGF function. No crystal structure of EGF or of a homologous growth factor is presently available. This paper presents an initial description of the three-dimensional structure of murine EGF (mEGF) in aqueous solution determined by ^{1}H NMR.

Previous NMR studies of murine (14–18), rat (19), and human (20) EGF include two papers on the use of NMR techniques for obtaining complete sequence-specific proton resonance assignments and determination of the three-dimensional structure in solution (21–23). So far the principal elements of regular backbone structure have been identified for both mEGF (18) and human EGF (20). In the work described here, the structure determination of mEGF was continued by collection of a more extensive set of experimental NMR constraints and the use of distance geometry calculations for the structural analysis of the NMR data.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHODS

Experimental. mEGF (type α1) from male submaxillary glands was purified as described (24, 25) and characterized as homogeneous by both C_{18} reversed-phase and Pharmacia Mono-Q anion-exchange HPLC. These homogeneity checks were made to exclude the possibility of deamidation of the N-terminal asparagine, which occurs in neutral and alkaline solutions (26–28). Samples for NMR spectroscopy were prepared at ≈6 mM protein concentration, pH 3.0 ± 0.1, in either 99.9% ^2H_2O or 85% H_2O/15% ^2H_2O. (For ^2H_2O solutions, the "pH" refers to the pH meter reading, without correction for isotope effects.) NOESY (two-dimensional nuclear Overhauser effect spectroscopy; refs. 29 and 30) and 2QF-COSY (two-dimensional two-quantum-filtered correlation spectroscopy; ref. 31) spectra were recorded on a Brüker WM-500 spectrometer with the probe temperature regulated at 28 ± 1°C. In NOESY, t_1 ridges were suppressed by recording sine-modulated data (32). The NMR data were recorded with time-proportional phase incrementation (33, 34) and were Fourier-transformed as absorption-mode spectra. The quality of the NOESY spectra obtained under these conditions was illustrated previously (18).

Computational. The structural interpretation of the NMR data was carried out with the DISMAN program (35), in which bond lengths, bond angles, and peptide bond dihedral angles are fixed at standard ECEPP (empirical conformational energy program for peptides) values (36, 37) and the remaining dihedral angles are treated as independent variables. The algorithm uses a variable target function (35) by which shorter-range distance constraints are satisfied before incorporating longer-range constraints. This variable-target procedure aids in overcoming the multiple-minima problem inherent in minimization procedures. In subsequent stages, longer-range distances are introduced and a conformation-dependent penalty function, which is zero when all distance and steric constraints are satisfied, is calculated. This penalty

Abbreviations: EGF, epidermal growth factor; mEGF, murine EGF; ECEPP, empirical conformational energy program for peptides; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; RMSD, root-mean-square deviation; τ_m, mixing time in NOESY experiment; 2QF-COSY, two-dimensional two-quantum-filtered correlation spectroscopy; DISMAN, program used to calculate protein structures from NMR data; $d_{AB}(i,j)$, distance between proton types A and B located, respectively, in amino acid residues i and j, where N, α, and β denote the amide proton, C$^\alpha$H, and C$^\beta$H, respectively, and $d_{\alpha N} \equiv d_{\alpha N}(i, i+1)$, $d_{NN} \equiv d_{NN}(i,i+1)$, and $d_{\beta N} \equiv d_{\beta N}(i,i+1)$.

function is then minimized with respect to the variable dihedral angles by the method of conjugate gradients (38), as described in detail elsewhere (35, 39). For each atom type, atomic radii used in calculating the steric contribution to the penalty function are defined as half the internuclear distance at which the ECEPP Lennard–Jones energy is 3 kcal/mol. Since we have not yet determined stereospecific resonance assignments of prochiral methylene protons and isopropyl methyl groups, pseudo-atom representations (21, 40) were used where appropriate.

Relative orientations of different conformers for minimum root-mean-square deviations (RMSDs) of the backbone atoms were calculated by the method of Kabsch (41, 42). Except where specified otherwise, the backbone RMSDs pertain to the N, C^{α}, and C′ atoms.

All calculations were carried out with a Prime 550 minicomputer.

RESULTS

Experimental Distance Constraints. Four kinds of experimental distance information were used in the structure determinations: (*i*) NOE-derived upper-bound ^1H–^1H distance constraints; (*ii*) constraints on the backbone dihedral angles ϕ indicated by vicinal $^3J_{HN\alpha}$ coupling constants; (*iii*) constraints implied by interstrand hydrogen bonds of β-sheets; and (*iv*) disulfide-bond constraints. Altogether, in addition to the distance constraints implicit in standard ECEPP geometry and in the nonbonded steric interactions, 278 upper-bound distance constraints (from *i*, *iii*, and *iv*), 27 lower-bound distance constraints (from *iii* and *iv*), and 28 dihedral angle constraints (from *ii*) have been identified (Table 1).

The interpretation of the NMR data is based on a nearly complete set of sequence-specific proton resonance assignments at pH 3.0 and 28°C (unpublished results). By use of these proton resonance assignments, cross-peaks of NOESY spectra in 85% H_2O recorded with mixing time τ_m = 65 ms and in ^2H_2O recorded with τ_m = 100 ms were assigned to particular proton pairs. The NOESY distance constraints used for this initial structure determination represent an incomplete data set, since unambiguous assignments for numerous cross-peaks could not be made because of chemical-shift degeneracy in one or both frequency dimensions. As was discussed elsewhere (e.g., refs. 21 and 43) many of these additional data points will be amenable to unambiguous assignment by reference to the initial structures in future work toward structure refinement.

Upper-bound distance constraints were derived from the following crude calibration of NOESY cross-peak intensity vs. interproton distance. In polypeptides, the intraresidue distances $d_{N\alpha}$ and the sequential distances $d_{\alpha N}$ range from 2.2 to 2.9 Å and from 2.2 to 3.6 Å, respectively (21, 44). Accordingly, for the 65-ms NOESY data in H_2O, the weakest $d_{N\alpha}$-type NOEs ($d_{N\alpha}^{max} \approx 2.9$ Å) and the strongest $d_{\alpha N}$-type NOEs ($d_{\alpha N}^{min} \approx 2.2$ Å) were used as reference points for calibrating a NOE intensity vs. distance relation, $I \propto d^{-6}$. With this calibration, it was found that the weakest cross-peaks in the spectrum arise from distances of *ca.* 3.4 Å. Based on these considerations, the 65-ms NOESY cross-peaks were divided into four categories corresponding to upper-bound constraints of 2.5, 2.7, 3.0, and 4.0 Å, respectively. A similar approach was used to calibrate the 100-ms NOESY spectrum recorded in ^2H_2O using the fixed distances (36, 37) between glycine α-, cystine β-, and proline δ-methylene protons and between juxtaposed α-protons of the β-sheets (21, 23, 44) as reference points. For this second NOESY data set, the cross-peaks were divided into only three groups of distance constraints, with upper bounds of 2.5, 3.0, and 4.0 Å, respectively. These calibrations were used to relate the intraresidue, sequential, and long-range backbone–backbone NOEs (21) to the corresponding distance constraints. The remaining long-range NOEs involving side-chain protons were assigned a fixed upper-bound distance of 4.0 Å. Where appropriate, pseudo-atom correction factors (40) were added to the upper-bound distances. This interpretation of the NOESY spectra corresponds to that used in earlier structure determinations (21, 43). A survey of the medium-range and long-range distance constraints used in the present structure determination is provided by Fig. 1.

Backbone vicinal $^3J_{HN\alpha}$ coupling constants were estimated from ω_2 cross-sections of NH/αH cross-peaks in 2QF-COSY spectra. We were careful to consider the effect of cancellation within the antiphase 2QF-COSY cross-peaks, which results in artificially enlarged splittings for coupling constants smaller than the natural linewidth (45). In the spectra recorded at 28°C, the NH/αH cross-peaks have linewidths of *ca.* 7 Hz, and the smallest apparent values of $^3J_{HN\alpha}$ were *ca.* 6 Hz. Only cross-peaks with apparent values of $^3J_{HN\alpha} > 8.0$ Hz were used for this structure determination, with the corresponding backbone dihedral angles ϕ constrained within the limits [−160°, −80°] (43, 46).

Interstrand hydrogen bonds in the β-structures were identified by locating the slowly exchanging amide protons of mEGF in a model of the previously identified antiparallel β-sheets (18). Five hydrogen bonds were identified in the peptide segment from Val^{19} to Asn^{32}, and four more in the Cys^{33}–Asp^{46} "double-hairpin" structure (see figure 4 of ref. 18). The upper-bound and lower-bound distances used for these hydrogen bonds, as well as those for the three disulfide bonds of mEGF (3), were the same as those described elsewhere (43).

Distance Geometry Calculations. The distribution of distance constraints (Fig. 1) indicates that the mEGF structure is divided roughly into an N-terminal and a C-terminal domain. Within each of these domains there is a dense network of medium-range and long-range distance constraints. The two domains are slightly overlapped in the sense that there are some distance constraints between the linking tripeptide segment 31–33 and the core of both domains (i.e., the disulfide bond between residues 33 and 42) and NOEs between residues 19 and 31, 31 and 40, 31 and 41, 33 and 38, and 33 and 42. For this reason, the chain-folds for the two polypeptide segments 1–33 and 32–53 were calculated separately. Subsequently, the two domain structures were combined and additional cycles of DISMAN calculations were carried out with the complete polypeptide chain (see *Discussion*).

Starting conformations for each of the two domains were generated as follows (35, 47). The short-range $d_{\alpha N}$ and d_{NN}

Table 1. Numbers of experimental constraints used in determining the polypeptide chain-fold of mEGF

Polypeptide	NOE-derived upper-bound constraints			$^3J_{HN\alpha}$ constraints	Hydrogen-bond constraints*	Disulfide constraints†
	Intraresidue	Sequential	Longer-range			
Segment 1–33	36	60	61	17	20	12
Segment 32–53	37	27	29	12	16	6
mEGF (1–53)	73	86	92	28	36	18

*There are two upper-bound and two lower-bound distance constraints for each hydrogen bond.
†There are three upper-bound and three lower-bound distance constraints for each disulfide bond.

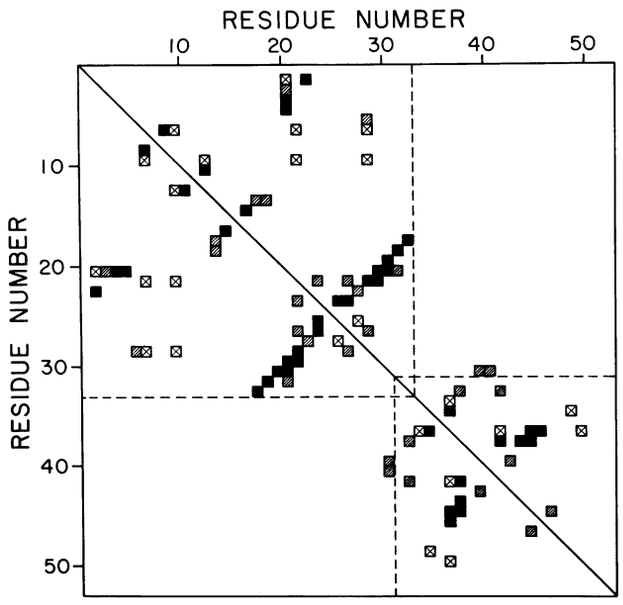

FIG. 1. Survey of medium-range and long-range NOE-derived distance constraints for mEGF. Both axes pertain to the sequence of mEGF, and squares connect pairs of residues linked by one or more NOE constraints. Filled, hatched, and crossed squares indicate backbone–backbone, backbone–side chain, and side chain–side chain NOE constraints, respectively. Where two residues are connected by more than one NOE, the NOE constraint involving the greater number of backbone protons is represented. The NOE constraints were derived from a 65-ms NOESY spectrum obtained in H_2O and from a 100-ms NOESY spectrum obtained in ^2H_2O, pH meter reading 3.1, t = 28°C. The broken lines indicate the N-terminal domain 1–33 and the C-terminal domain 32–53, which overlap in the peptide segment 32–33.

distances impose constraints on the intervening ϕ and ψ dihedral angles, which (with ω fixed at 180°) can easily be determined analytically (21, 44). Along with the $^3J_{HN\alpha}$ coupling constants discussed above, these short-range distance constraints limit the conformational space to be searched in subsequent stages of the DISMAN calculations. Accordingly, starting conformations were generated randomly within the range of ϕ and ψ values specified by these short-range constraints. Backbone dihedral angles lacking short-range constraints and side-chain dihedral angles χ^1 were generated completely randomly, and the more peripheral side-chain dihedral angles were set initially at 180°.

Fifteen and 10 starting conformations were generated for the fragments 1–33 and 32–53, respectively. The randomness of these starting polypeptide chain-folds was checked in two ways. First, backbone–backbone distance maps were generated for each starting conformation. From visual inspection of these distance maps, there is no evidence for long-range conformational correlations in either of the two sets of starting conformations, despite the short-range correlations that are inherent in our procedure for generating them. Secondly, the average RMSDs for the C^α atoms for all pairs of starting conformations were also calculated. These RMSDs had values of 10.0 ± 2.5 Å and 8.0 ± 2.1 Å and are thus in the range expected for random-coil 33- and 22-residue polypeptides (48). The distance object functions for these structures were then minimized using the DISMAN program. Of the 15 structures for the segment 1–33 and the 10 structures for the segment 32–53 calculated with DISMAN, the five structures for each segment that best satisfied the experimental distance constraints were selected for further study and comparison.

Residual constraint violations in the two separate domain structures are summarized in Table 2. The largest distance-constraint violations are ≈0.6 Å and the largest steric-constraint violations are ≈0.3 Å. Table 2 also contains values of the average violation distance per constraint for each structure (47); i.e., the sum of all distance violations divided by the total number of NMR constraints. This parameter provides a sequence-independent measure of the deviations between the experimental distance constraints and the corresponding distance violations in the calculated structures. The small ranges of average violations calculated for the five best structures for the N-terminal domain (0.037–0.040 Å) and the five best structures for the C-terminal domain (0.074–0.078 Å) indicate that any of these is an equally credible solution. In Fig. 2*A*, the five superimposed backbone structures of the N-terminal domain are shown. The mean backbone RMSD, calculated as the average over all nonidentical pairs of conformers, is 1.3 Å. In the five structures for the C-terminal domain, the hexapeptide segment Arg^{48}–Arg^{53} is nearly unconstrained (Fig. 1) and was therefore not considered in the RMSD calculations. For the Asn^{32}–Leu^{47} segment, the mean backbone RMSD is 1.4 Å. The superimposed backbone structures for this segment are shown in Fig. 2*B*.

DISCUSSION

It is clear from the preceding sections that the superposed structures of the two separate domains of mEGF in Fig. 2 represent the hard core of the information obtained, since

Table 2. Summary of residual constraint violations for the mEGF structures calculated with the DISMAN program

Structure	Number of distance constraint violations*		Number of steric constraint violations†		Number of dihedral angle constraint violations >10°	Average violation distance per constraint‡, Å
	0.3–0.5 Å	>0.5 Å	0.1–0.3 Å	>0.3 Å		
Segment 1–33						
1	6	1 (0.57 Å)	3	0	1	0.040
2	8	1 (0.55 Å)	3	1 (0.31 Å)	1	0.039
3	5	1 (0.58 Å)	2	1 (0.32 Å)	1	0.037
4	7	1 (0.57 Å)	3	0	1	0.040
5	6	2 (0.60 Å)	3	0	2	0.041
Segment 32–53						
1	8	1 (0.54 Å)	1	0	1	0.075
2	9	1 (0.58 Å)	1	0	2	0.074
3	10	1 (0.51 Å)	0	0	2	0.078
4	5	1 (0.50 Å)	2	0	1	0.075
5	9	1 (0.62 Å)	1	0	1	0.078

*The largest experimental constraint violation for each structure is given in parentheses.
†The largest steric constraint violation, if >0.3 Å, is given in parentheses.
‡Sum of all distance violations divided by the total number of NMR distance constraints.

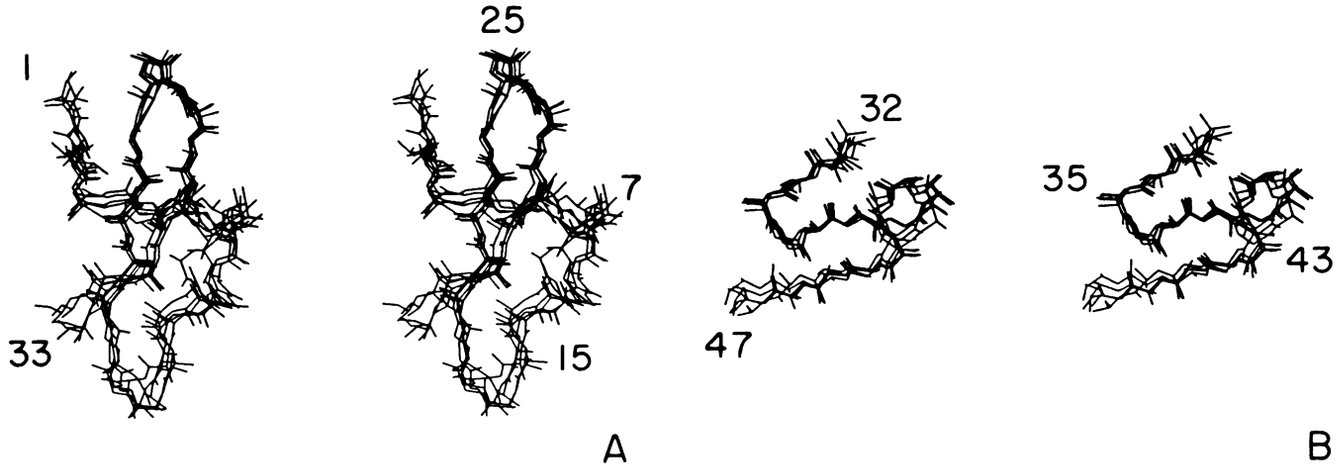

FIG. 2. (*A*) Stereo view of a minimum RMSD superposition of the five best structures for the N-terminal domain Asn[1]–Cys[33] of mEGF. The mean backbone RMSD for these five structures is 1.3 Å. (*B*) Stereo view of the five best structures for the C-terminal domain Asn[32]–Leu[47] superimposed for minimum RMSD. The mean backbone RMSD for these five structures is 1.4 Å. The N, C^{α}, C′, O′, and C^{β} atoms are drawn, and the disulfide bonds are not shown.

each domain is in itself well constrained by the experimental data (Fig. 1). The backbone conformations in Fig. 2 are in good agreement with regular backbone structures that were previously identified (18) based on qualitative NOE pattern-recognition methods (21, 23). The larger domain (residues Asn[1]–Cys[33]) consists of an antiparallel β-sheet (residues Gly[18]–Ile[23] and Ser[28]–Cys[33]) with a loosely attached antiparallel third strand (residues Ser[2]–Gly[5]). This β-sheet is slightly larger than the one we described previously (18), since the distance geometry calculations show that residues Gly[18] and Cys[33] are also part of the antiparallel strands. The antiparallel β-sheet structure has a well-defined right-handed twist (Fig. 2*A*), which is typical of β-sheets in proteins (49–51). The β-bend at Ser[25]-Leu[26] is of type I. The residues Cys[6]–Gly[17] of this domain of mEGF adopt an irregular multiple-bend structure, part of which (residues Pro[7]–Leu[15]) may be regarded as a short, irregular, right-handed helical structure (Fig. 2*A*). This domain structure may be a common feature in the whole family of homologs of mEGF, since an antiparallel β-sheet conformation for residues Gly[18]–Ile[23] and Lys[28]–Val[34] has also been identified in the solution structure of human des(49–53)-EGF (20) by using ^{1}H NMR spectroscopy.

In the smaller, C-terminal domain, the backbone conformation of residues Asn[32]–Leu[47] is best described as a "double-hairpin" structure (Fig. 2*B*), with a type II β-bend at Ile[35]-Gly[36] and a multiple-bend conformation in the segment Gly[39]–Gln[43]. The backbone conformation of segment Gly[39]–Gln[43] has a well-defined left-handed twist (Fig. 2*B*). The left-handed backbone structure arises because Gly[39] and Asp[40] adopt A* and B* backbone conformations (52), respectively. In order to determine if a right-handed twist of the Gly[39]–Gln[43] loop is also compatible with the experimental data, additional DISMAN calculations were carried out on the domain Asn[32]–Arg[53], in which the backbone dihedral angles of both Gly[39] and Asp[40] were constrained to (ϕ, ψ) values of (−57°, −47°); i.e., to the A-type backbone conformation (52). All of the resulting structures had a significantly increased number of constraint violations. In small peptides, the A* conformation of glycine corresponds to one of the lowest energy minima (53), while the B* conformation of aspartic acid is very unfavorable energetically (53). A visual examination of the three dimensional structure of mEGF suggests that the Cys[33]–Cys[42] disulfide bond may be responsible for stabilizing the unusual, left-handed structure of the Gly[39]–Gln[43] loop.

The conformation of the remaining C-terminal residues Arg[48]–Arg[53] is nearly unconstrained by the available experimental data (Fig. 1). Significantly, this part of the molecule is not required for the binding of mEGF to its receptor, since des(49–53)-mEGF, which can be prepared by treatment with trypsin, is biologically active (2). Photochemically induced dynamic nuclear polarization studies of mEGF (14, 54) indicated that the side chains of both Trp[49] and Trp[50] are accessible to an extrinsic flavin-dye probe, which has been interpreted as evidence for conformational flexibility of the C-terminal residues (14). While the present results do not allow more precise statements about either the structure or the dynamics of the peptide segment 48–53, a disordered flexible state would clearly be a possible explanation for the absence of ^{1}H–^{1}H NOEs.

To obtain an overall view of the complete mEGF molecule, five starting conformations were generated by combining the N-terminal domain structure 1 in Table 2 with the C-terminal domain structure 1, and similarly by combining the structures 2 through 5. In making these combinations of the two domains, the dihedral angles of the residues Asn[32] and Cys[33] were taken from the N-terminal and C-terminal domain structures, respectively. The complete distance object functions for these five starting conformations, including the two distance constraints not used in the separate domain structure calculations (i.e., NOEs between Cys[31] and Asp[40] and between Cys[31] and Arg[41]), were further minimized and the structures were improved by additional cycles of minimization with DISMAN. The significance of these complete mEGF structures is determined by the facts that the backbone dihedral angles of residues Ser[2]–Asn[32] and of residues Cys[33]–Leu[47] are fairly well determined from distance constraints within the N-terminal domain 1–33 and the C-terminal domain 32–53, respectively, and that there are two additional distance constraints which were not used in the separate domain calculations—i.e., those between residues 31 and 40 and between residues 31 and 41. As in the domain structures from which they are derived, the largest distance- and steric-constraint violations for these complete mEGF structures are 0.59 Å and 0.31 Å, respectively, and the average violation distance per constraint is very similar (0.05 ± 0.01 Å), indicating that each one is an equally credible structure. For residues Asn[1]–Leu[47], the mean backbone RMSD is 1.8 Å. Two representations of the backbone chain-fold of one of these mEGF structures are shown in Fig. 3. Comparison of Figs. 2 and 3 clearly indicates that the structures obtained in the separate calculations for the N- and C-terminal domains are largely preserved in the structures thus obtained for the complete mEGF molecule.

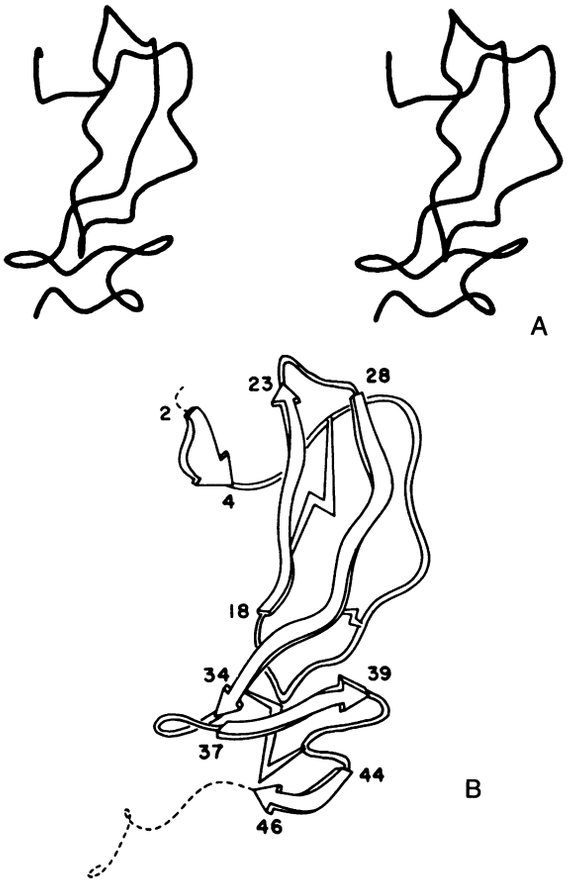

FIG. 3. (*A*) Stereo view of the backbone structure of mEGF. In this representation a spline-fit to the C^{α} coordinates of one of the mEGF structures is used to give an overall view of the polypeptide chain-fold. Only residues Ser^{2}–Leu^{47} are drawn. (*B*) Schematic drawing of the backbone polypeptide chain-fold obtained using the coordinates of the same mEGF structure. The arrowed ribbons indicate the position and direction of the β-sheets, and the dotted lines indicate the chain terminal segments, for which the backbone conformation could not be determined reliably from the available data. The disulfide bonds are indicated by "lightning bolts."

We thank Drs. W. Braun, A. D. Kline, and G. Wagner for many helpful discussions. This work was supported by research grants from the National Institute of General Medical Sciences (GM24893), the Schweizerischer Nationalfonds (3.198-9.85), and the National Science Foundation (DMB84-01811). Support was also received from the National Foundation for Cancer Research and the Cornell Biotechnology Center. We also acknowledge the support of the National Institutes of Health Research Resource for Multinuclear NMR and Data Processing at Syracuse University (RR-01317) and the Cornell Production Supercomputer Facility.

1. Cohen, S. (1962) *J. Biol. Chem.* **237,** 1555–1562.
2. Savage, C. R., Jr., Inagami, T. & Cohen, S. (1972) *J. Biol. Chem.* **247,** 7612–7621.
3. Savage, C. R., Jr., Hash, J. H. & Cohen, S. (1973) *J. Biol. Chem.* **248,** 7669–7672.
4. Sporn, M. B. & Roberts, A. B. (1985) *Nature (London)* **313,** 745–747.
5. Sporn, M. B. & Todaro, G. J. (1980) *N. Engl. J. Med.* **303,** 878–880.
6. Buckley, A., Davidson, J. M., Kamerath, C. D., Wolt, T. B. & Woodward, S. C. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7340–7344.
7. Smith, J., Cook, E., Fotheringham, I., Pheby, S., Derbyshire, R., Eaton, M. A. W., Doel, M., Lilley, D. M. J., Pardon, J. F., Patel, T., Lewis, H. & Bell, L. D. (1982) *Nucleic Acids Res.* **10,** 4467–4482.
8. Urdea, M. S., Merryweather, J. P., Mullenbach, G. T., Coit, D., Heberlein, U., Valenzuela, P. & Barr, P. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7461–7465.
9. Oka, T., Sakamoto, S., Miyoshi, K.-I., Fuwa, T., Yoda, K., Yamasaki, M., Tamura, G. & Miyake, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7212–7216.
10. Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H. & Schlessinger, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1351–1355.
11. Nestor, J. J., Jr., Newman, S. R., DeLustro, B., Todaro, G. J. & Schreiber, A. B. (1985) *Biochem. Biophys. Res. Commun.* **129,** 226–232.
12. Eppstein, D. A., Marsh, Y. V., Schreiber, A. B., Newman, S. R., Todaro, G. J. & Nestor, J. J., Jr. (1985) *Nature (London)* **318,** 663–665.
13. Heath, W. F. & Merrifield, R. B. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6367–6371.
14. DeMarco, A., Menegatti, E. & Guarneri, M. (1983) *FEBS Lett.* **159,** 201–206.
15. Mayo, K. H. (1984) *Biochemistry* **23,** 3960–3973.
16. Mayo, K. H. (1985) *Biochemistry* **24,** 3783–3794.
17. DeMarco, A., Mayo, K. H., Bartolotti, F., Scalia, S., Menegatti, E. & Kaptein, R. (1986) *J. Biol. Chem.* **261,** 13510–13516.
18. Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W. & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8594–8598.
19. Mayo, K. H., Schaudies, P., Savage, C. R., DeMarco, A. & Kaptein, R. (1986) *Biochem. J.* **239,** 13–18.
20. Carver, J. A., Cooke, R. M., Esposito, G., Campbell, I. D., Gregory, H. & Sheard, B. (1986) *FEBS Lett.* **205,** 77–81.
21. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
22. Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982) *J. Mol. Biol.* **155,** 311–319.
23. Wüthrich, K., Billeter, M. & Braun, W. (1984) *J. Mol. Biol.* **180,** 715–740.
24. Burgess, A. W., Knesel, J., Sparrow, L. G., Nicola, N. A. & Nice, E. C. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5753–5757.
25. Burgess, A. W., Lloyd, C. J. & Nice, E. C. (1983) *EMBO J.* **2,** 2065–2069.
26. Koch, J. H., Fifis, T., Bender, V. J. & Moss, B. A. (1984) *J. Cell. Biochem.* **25,** 45–59.
27. Thannhauser, T. W. & Scheraga, H. A. (1985) *Biochemistry* **24,** 7681–7688.
28. Meinwald, Y. C., Stimson, E. R. & Scheraga, H. A. (1986) *Int. J. Pept. Protein Res.* **28,** 79–84.
29. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71,** 4546–4553.
30. Kumar, A., Ernst, R. R. & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95,** 1–6.
31. Piantini, U., Sørensen, O. W. & Ernst, R. R. (1982) *J. Am. Chem. Soc.* **104,** 6800–6801.
32. Otting, G., Widmer, H., Wagner, G. & Wüthrich, K. (1986) *J. Magn. Reson.* **66,** 187–193.
33. Redfield, A. G. & Kunz, S. D. (1975) *J. Magn. Reson.* **19,** 250–254.
34. Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974.
35. Braun, W. & Gō, N. (1985) *J. Mol. Biol.* **186,** 611–626.
36. Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79,** 2361–2381.
37. Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983) *J. Phys. Chem.* **87,** 1883–1887.
38. Fletcher, R. (1980) *Practical Methods of Optimization* (Wiley, New York), Vol. 1, pp. 63–70.
39. Abe, H., Braun, W., Noguti, T. & Gō, N. (1984) *Comput. Chem.* **8,** 239–247.
40. Wüthrich, K., Billeter, M. & Braun, W. (1983) *J. Mol. Biol.* **169,** 949–961.
41. Kabsch, W. (1976) *Acta Crystallogr. Sect. A* **32,** 922–923.
42. Kabsch, W. (1978) *Acta Crystallogr. Sect. A* **34,** 827–828.
43. Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol.* **182,** 295–315.
44. Billeter, M., Braun, W. & Wüthrich, K. (1982) *J. Mol. Biol.* **155,** 321–346.
45. Neuhaus, D., Wagner, G., Vašák, M., Kägi, J. H. R. & Wüthrich, K. (1985) *Eur. J. Biochem.* **151,** 257–273.
46. Pardi, A., Billeter, M. & Wüthrich, K. (1984) *J. Mol. Biol.* **180,** 741–751.
47. Kline, A. D., Braun, W. & Wüthrich, K. (1986) *J. Mol. Biol.* **189,** 377–382.
48. McLachlan, A. D. (1984) *Biopolymers* **23,** 1325–1331.
49. Chothia, C. (1973) *J. Mol. Biol.* **75,** 295–302.
50. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
51. Chou, K.-C., Pottle, M., Némethy, G., Ueda, Y. & Scheraga, H. A. (1982) *J. Mol. Biol.* **162,** 89–112.
52. Zimmerman, S. S., Pottle, M. S., Némethy, G. & Scheraga, H. A. (1977) *Macromolecules* **10,** 1–9.
53. Vásquez, M., Némethy, G. & Scheraga, H. A. (1983) *Macromolecules* **16,** 1043–1049.
54. Mayo, K. H., DeMarco, A. & Kaptein, R. (1986) *Biochim. Biophys. Acta* **874,** 181–186.