Szilagyi, L., & Jardetzky, O. (1989) *J. Magn. Reson. 83*, 441-449.
Trewhella, T., Blumenthal, D. K., Rokop, S. E., & Seeger, P. A. (1990) *Biochemistry 29*, 9316-9324.
Veitch, N. C., Concar, D. W., Williams, R. J. P., & Whitford, D. (1988) *FEBS Lett. 238*, 49-55.
Wagner, G., Pardi, A., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 5948-5949.
Waltho, J. P., Feher, V. A., Lerner, R. A., & Wright, P. E. (1989) *FEBS Lett. 250*, 400-404.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167-7175.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Wüthrich, K., Billeter, M., & Brown, W. (1984) *J. Mol. Biol. 180*, 715-740.

Solution Structure of Murine Epidermal Growth Factor Determined by NMR Spectroscopy and Refined by Energy Minimization with Restraints[†,‡]

Gaetano T. Montelione,[§] Kurt Wüthrich,[∥] Antony W. Burgess,[⊥] Edward C. Nice,[⊥] Gerhard Wagner,[#] Kenneth D. Gibson,[▽] and Harold A. Scheraga*[,▽]

Center for Advanced Biotechnology and Medicine, and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854-5635, Institut für Molekularbiologie und Biophysik, Eidenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland, Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, Victoria 3050, Australia, Department of Biological Chemistry and Molecular Pharmacology, Harvard University, 240 Longwood Avenue, Boston, Massachusetts 02115, and Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received June 4, 1991; Revised Manuscript Received September 18, 1991

ABSTRACT: The solution structure of murine epidermal growth factor (mEGF) at pH 3.1 and a temperature of 28 °C has been determined from NMR data, using distance geometry calculations and restrained energy minimization. The structure determination is based on 730 conformational constraints derived from NMR data, including 644 NOE-derived upper bound distance constraints, constraints on the ranges of 32 dihedral angles based on measurements of vicinal coupling constants, and 54 upper and lower bound constraints associated with nine hydrogen bonds and the three disulfide bonds. The distance geometry interpretation of the NMR data is based on previously published sequence-specific ^{1}H resonance assignments [Montelione et al. (1988) *Biochemistry 27*, 2235-2243], supplemented here with individual assignments for some side-chain amide, methylene, and isopropyl methyl protons. The molecular architecture of mEGF is the same as that described previously [Montelione et al. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5226-5230], but the structure is overall more precisely determined by a more extensive set of NMR constraints. Analysis of proton NMR line widths, amide proton exchange rates, and side-chain $^3J(H^{\alpha}-H^{\beta})$ coupling constants provides evidence for internal motion in several regions of the mEGF molecule. Because mEGF is one member of a large family of homologous growth factors and protein domains for which X-ray crystal structures are not yet available, the atomic coordinates resulting from the present structure refinement (which we have deposited in the Brookhaven Protein Data Bank) are important data for understanding the structures of EGF-like proteins and for further detailed comparisons of these structures with mEGF.

Epidermal growth factor (EGF)[1] is a small mitogenic protein containing 53 amino acids and three disulfide bonds (Cohen, 1962; Savage et al., 1972, 1973). EGF and EGF-like proteins are thought to play important roles in wound healing (Burgess, 1989) and oncogenesis (Sporn & Todaro, 1980; Sporn & Roberts, 1985; Heldin & Westermark, 1990). Although no crystal structure is available for EGF, or for any proteins homologous to EGF, solution structures derived from NMR measurements are available for human EGF (Carver et al., 1986; Cooke et al., 1987), murine EGF (Montelione et al., 1986, 1987; Kohda & Inagaki, 1988; Kohda et al., 1988), and rat EGF (Mayo et al., 1989), as well as for the homologous human type α transforming growth factor (TGFα)

[†] This work was supported by research grants from the National Institute of General Medical Sciences (GM-24893), the American Cancer Society, the Schweizerischer Nationalfonds (project 3.198.85), the Damon Runyon Walter Winchell Cancer Research Foundation (DRG-920), and the Searle Scholars Progam/The Chicago Community Trust. Support was also received from the National Foundation for Cancer Research.

[‡] The atomic coordinates have been deposited in the Brookhaven Protein Data Bank.

* Author to whom correspondence should be addressed.
[§] Rutgers University.
[∥] Institute für Molekularbiologie und Biophysik.
[⊥] Ludwig Institute for Cancer Research.
[#] Harvard University.
[▽] Cornell University.

[1] Abbreviations: EGF, epidermal growth factor; mEGF, murine EGF; hEGF, human EGF; TGF-α, type α transforming growth factor; hTGFα, human TGFα; 2D, two dimensional; COSY, 2D correlated spectroscopy; 2QF-COSY, 2D two-quantum filtered COSY; COSY-30, COSY in which the mixing pulse has a 30° flip angle; DISMAN, program used to calculate molecular structures from internuclear distance constraints; E-COSY, 2D exclusive COSY; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; RMSD, root-mean-square deviation; ECEPP/2, empirical conformational energy program for peptides.

0006-2960/92/0431-236$03.00/0

(Brown et al., 1989; Campbell et al., 1989; Kohda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Kline et al., 1990) and the EGF-like domain from bovine coagulation factor X (Selander et al., 1990). For all of these molecules, the backbone structures are very similar to that first described for murine EGF (Montelione et al., 1986), with a three-stranded antiparallel β-sheet in the N-terminal two-thirds of the sequence and a small "double hairpin" struture in the C-terminal third of the sequence. These two polypeptide segments of the protein are oriented with respect to one another by side-chain to side-chain and side-chain to backbone interactions. In this paper, we extend the earlier description of the polypeptide backbone fold of mEGF (Montelione et al., 1987) and describe the complete three-dimensional structure determined with distance geometry calculations using as input an extensive set of NMR-derived conformational constraints and refined by restrained energy minimization.

Materials and Methods

NMR Measurements. Protein samples for NMR spectroscopy were prepared at 5–6 mM concentration, pH 3.1. NOESY data sets were recorded as described elsewhere (Montelione et al., 1986, 1987, 1988). To select multiplet components of connected transitions, COSY spectra were recorded with a mixing pulse of $\beta \ll 90°$ (Aue et al., 1976; Bax & Freeman, 1981; Müller, 1987). A flip angle of $\beta = 30°$ was found to provide a good compromise between sensitivity and the desired selection of connected transitions. The COSY-30 spectrum was recorded at 500 MHz using 1024 data points in ω_1 and 4096 data points in ω_2, with zero-filling to 2048 and 8192 data points, respectively, prior to two-dimensional Fourier transformation. All NMR data were obtained at 28 ± 1 °C.

Stereospecific Assignments of β-Methylene Protons. The diastereotopic β-methylene protons of amino acid side chains can be distinguished by combined measurements of $^3J(H^\alpha–H^\beta)$ coupling constants and $H^N–H^\beta$ nuclear Overhauser effects (Hyberts et al., 1987), interpreted in terms of rotational isomeric states. $H^N–H^\beta$ nuclear Overhauser effects were estimated from cross-peak intensities in a NOESY spectrum recorded with a mixing time of 65 ms, which is sufficiently short to minimize contributions from spin diffusion. These protons are separated by four bonds, and the NOESY cross peaks are not affected by scalar coupling. In addition, $^3J(H^\alpha–H^\beta)$ coupling constants were measured using COSY-30 spectra. In these measurements, use was made of the fact that a small flip angle (e.g., 30°) of the mixing pulse in a simple COSY sequence restricts magnetization transfer to occur primarily between connected transitions (Aue et al., 1976; Bax & Freeman, 1981). Although optimal selection for connected transitions would in principle be achieved with $\beta = 0°$, good selectivity can in practice be obtained with small pulse widths that also yield workable sensitivity. In our experience, a flip angle of $\beta = 30°$ provides a good compromise, which results in a simplified COSY cross-peak pattern similar to that of E-COSY (Griesinger et al., 1985). The COSY-30 experiment, however, is significantly more sensitive than E-COSY. The passive $^3J(H^\alpha–H^\beta)$ splittings in these simplified COSY cross peaks are fairly insensitive to the proton line width and, in the absence of strong coupling effects, can be used to estimate the corresponding coupling constant. $^3J(H^\alpha–H^{\beta 2})$ and $^3J(H^\alpha–H^{\beta 3})$ vicinal coupling constants were thus estimated from the ($\omega_1 = H^{\beta 3}$, $\omega_2 = H^\alpha$) and ($\omega_1 = H^{\beta 2}$, $\omega_2 = H^\alpha$) cross peaks, respectively, of a COSY-30 spectrum.

Stereospecific Assignments of Valine Isopropyl Methyl Groups. Individual assignments for $C^\gamma H_3$ methyl groups of valine side chains can be determined by combined measurements of $^3J(H^\alpha–H^\beta)$ coupling constants and intraresidue NOEs (Zuiderweg et al., 1985; Hyberts et al., 1987). However, we could not make reliable estimates of these Val $^3J(H^\alpha–H^\beta)$ coupling constants from COSY and 2QF-COSY spectra of mEGF because of the convolution of the proton resonance line shape with $^3J(H^\alpha–H^\beta)$ splittings. Instead, we compared the relative intensities of $H^N–C^\gamma H_3$ and $H^\alpha–C^\gamma H_3$ NOEs. Assuming that the valine side chain adopts one of three low-energy rotamer conformations (i.e., g^+, t, g^-) about the $C^\alpha–C^\beta$ bond, these local constraints were used to determine stereospecific methyl assignments and restricted ranges of the side-chain dihedral angle χ^1. The NOESY cross peaks used in this assignment procedure involve protons separated by four or five bonds and therefore have little contribution from scalar coupling. A short mixing time of $\tau_m = 65$ ms was used to minimize effects of spin diffusion.

Individual Assignments of Side-Chain Amide Protons. The IUPAC identifiers for the side-chain amide protons of asparagine and glutamine are $H^{\delta 1}$ and $H^{\epsilon 1}$ for protons cis with respect to the amide carbonyl oxygen atom and $H^{\delta 2}$ and $H^{\epsilon 2}$ for the trans protons. Individual assignment for these side-chain amide protons were obtained by comparing the relative intensities of the pairs of $H^\beta–H^\delta$ and $H^\gamma–H^\epsilon$ NOEs, respectively, in asparagine and glutamine side chains (Montelione et al., 1984). The internuclear distances between the H^β's in asparagine (or H^γ's in glutamine) and $H^{\delta 1}$ (or $H^{\epsilon 1}$) are shorter than the corresponding distances to $H^{\delta 2}$ (or $H^{\epsilon 2}$) for any value of the intervening dihedral angle χ^2 (or χ^3). To minimize effects of spin diffusion, a short NOESY mixing time of $\tau_m = 65$ ms was used. Contributions to these cross peaks from coherence transfer by scalar coupling are negligible.

Calibration of NOESY Data. Internuclear distance constraint lists were derived from the three NOESY data sets of Table I. An additional NOESY spectrum recorded with a mixing time of 200 ms in H_2O was used only to verify the identities of weak cross peaks in the 65-ms NOESY spectrum in H_2O.

Upper bound distance constraints were derived from the 65-ms NOESY data set on the basis of the following considerations. In polypeptides, intraresidue H^N to H^α and sequential H^α to H^N distances range from 2.2 to 2.9 Å and from 2.2 to 3.6 Å, respectively (Billeter et al., 1982). Accordingly, the weakest H^N to H^α NOEs [$d^{max}(H^N \cdots H^\alpha) \approx 2.9$ Å] and strongest sequential NOEs [$d^{min}(H^\alpha \cdots H^N) \approx 2.2$ Å] were used as reference points for calibrating an NOE intensity vs distance relation, assuming that $I \propto d^{-6}$. With this calibration, it was found that the weakest cross peaks in the 65-ms NOESY spectrum arise from distances of ca. 4.0 Å when the data are plotted down to the white noise following two-dimensional baseline correction. The cross peaks were therefore divided into five categories corresponding to upper bound constraints of 2.5, 2.7, 3.0, 4.0, and 6.0 Å, as shown in Table I. This calibration is the same as we have used previously for mEGF (Montelione et al., 1987), except that the 6.0-Å limit has been added to include very weak NOEs observed in carefully baseline-corrected spectra. The identities of these weak NOEs were verified by using an additional NOESY data set recorded with a mixing time of 200 ms. This calibration was used to establish internuclear upper bound distances between H^N, H^α, and H^β atoms from the 65-ms NOESY data. All other NOEs with more peripheral side-chain protons, which are more prone to dynamic effects, were assigned upper bound distance constraints of 6.0 Å. There is no evidence that distances longer than 5.0 Å can be observed as direct NOEs (Wüthrich, 1986).

Table I: Calibrations for Upper Bound Distance Constraints from NOESY Data of mEGF at 28 °C and pH 3.1

spectrum no.	solvent	mixing time (ms)		calibration	
1	H_2O	65	all intraresidue NOEs and all NOEs between H^N, H^α, and H^β atoms[a]	very strong	<2.5 Å
				strong	<2.7 Å
				medium	<3.0 Å
				weak	<4.0 Å
				very weak	<6.0 Å
			all other NOEs		<6.0 Å
2	^2H_2O	100	all interresidue NOEs between H^α and H^β atoms[a,b]	strong	<2.5 Å
				medium	<3.0 Å
				weak	<4.0 Å
				very weak	<6.0 Å
			all other NOEs		<6.0 Å
3	^2H_2O	250	all NOEs[b]		<6.0 Å

[a] Also includes NOEs between (H^N, H^α, H^β) and Pro H^δ. [b] Intraresidue NOEs were not considered in the NOESY data sets recorded with τ_m = 100 ms and τ_m = 250 ms.

However, we decided to be conservative and use a loose upper bound of 6.0 Å for these constraints.

A similar approach was used to calibrate the 100-ms NOESY spectrum recorded in ^2H_2O, using the fixed distances between the geminal glycine α-, cystine β-, and proline δ-methylene protons and between juxtaposed α-protons of the β-sheets (Wüthrich et al., 1984) as reference points. For this second NOESY data set, the NOEs among H^α and H^β protons were assigned four distance limits (Table I), and all of the remaining NOEs were interpreted as upper bound distances of 6.0 Å. For the 250-ms ^2H_2O NOESY data set (Table I), only NOEs between protons three or more residues apart, which were not already identified in the shorter mixing-time data, were added to the NOE constraint list. In this 250-ms mixing-time data set, we did not attempt to calibrate internuclear distances because of the increasing contributions of spin diffusion, and all of these additional NOEs were interpreted as upper bound constraints of 6.0 Å to allow for the possibility of spin diffusion along intraresidue pathways.

Where appropriate, pseudoatom corrections (Wüthrich et al., 1983) were added to the upper bound distance constraints derived from the NOE data. In the case in which both NOEs between a given proton and the two protons of a single methylene group could be identified, the longer one of the two distance constraints was used for both methylene protons.

Reiterative Evaluation of NOESY Data. The interpretation of two-dimensional NOESY data is limited by the fact that many cross peaks cannot initially be assigned to a unique pair of interacting spins because of chemical shift degeneracies in one dimension. In the previous description of the solution structure of mEGF (Montelione et al., 1987), only NOESY cross peaks for which both the ω_1 and ω_2 resonance frequencies could be uniquely assigned were used in the distance constraint list. However, by reference to the three-dimensional structures calculated from these data, additional NOESY cross-peak assignments were made on the basis of structural consistency [e.g., Kline et al. (1988)]. In this way, the NOESY spectra of mEGF were reiteratively evaluated in several rounds of assigning NOESY cross peaks and calculating a set of structures consistent with these constraints.

Structure Calculation and Refinement. The structural interpretation of the NMR data was carried out with the DISMAN program (Braun & Gō, 1985), modified to include individually assigned methylene protons, side-chain amide protons, and isopropyl methyl groups (Braun, 1987). DISMAN uses a variable target function algorithm in which shorter range constraints are satisfied before considering longer range constraints. Bond lengths, bond angles, and peptide group dihedral angles are fixed at standard values (Némethy et al., 1983), and the dihedral angles are treated as independent variables. A penalty function is minimized with respect to the variable dihedral angles by the method of conjugate gradients (Abe et al., 1984; Braun & Gō, 1985; Braun, 1987). In principle, this penalty function is zero when all conformational constraints corresponding to NOEs, spin–spin coupling constants, hydrogen bonds, disulfide bonds, and steric constraints are satisfied. For each atom type, hard-sphere atomic radii used in calculating the steric contribution to the penalty function are defined as half the internuclear distance at which the ECEPP/2 energy (Némethy et al., 1983) of nonbonded Lennard-Jones interactions is 3 kcal/mol. DISMAN never assesses a steric contribution to the penalty function from any pair of atoms that might possibly be involved in a hydrogen bond, according to the criterion used in ECEPP/2 (Momany et al., 1975). The DISMAN computations were carried out on a Silicon Graphics 4D-120 workstation.

Structures generated with the DISMAN program were refined by restrained energy minimization, in order to remove the most serious atomic overlaps. The restraints used were of the form

$$f = \sum_{ub}[(|\mathbf{r}_{ij}| - d_{ij})|_+]^2 + \sum_{lb}[(d_{ij} - |\mathbf{r}_{ij}|)|_+]^2 + \sum_{i<j}[(s_i + s_j - |\mathbf{r}_{ij}|)|_+]^2 \quad (1)$$

where $\mathbf{r}_{ij}$ is the vector from (pseudo)atom i to (pseudo)atom j, d_{ij} is a distance constraint for this pair of (pseudo)atoms, and s_i and s_j are the atomic radii for atoms i and j as defined by Braun & Gō (1985). The symbol $|_+$ indicates that the quantity in parentheses is included only if it is positive. The first sum is taken over all upper bound (ub) constraints, the second over all lower bound (lb) constraints, and the third over all pairs of atoms. Constraints on dihedral angles derived from spin–spin couplings were not considered during the energy calculations. Restrained energy minimization was carried out by minimizing the function

$$g = \alpha f + (1 - \alpha)E \quad (2)$$

where E is the ECEPP/2 energy (Momany et al., 1975; Némethy et al., 1983; Sippl et al., 1984) and $0 \leq \alpha < 1$. All ionizable side chains (Asp, Glu, His, Arg) were taken to be un-ionized, and no attempt was made to model the effect of solvation. Peptide bond dihedral angles ω were allowed to vary in the restrained minimization. In all cases, the minimization routine (Gay, 1983) converged after $1.2N$ to $2N$ function evaluations, where N (= 297) is the number of variable dihedral angles. These energy computations were carried out on an IBM 3090/600 supercomputer at the Cornell National

Table II: Stereospecific $C^{\beta}H_2$ Assignments for mEGF at 28 °C and pH 3.1

amino acid	chemical shifts (ppm) H^N	H^{α}	H^{β}	$^3J_{\alpha\beta}$ (Hz)	H^N–H^{β} NOE[a]	χ^1	stereospecific assignments of H^{β}
Tyr-3	8.34	4.90	3.08	6	400	nonstandard or multiple rotamer states	
			2.73	9	600		
Cys-14	8.73	4.39	2.58	12	600	−60° ± 30°	β2
			2.26	3	300		β3
Asn-16	8.98	4.03	1.98	10	100	−60° ± 30°	β2
			1.37	2	50		β3
Cys-20	8.86	5.01	3.36	4	1200	180° ± 30°	β2
			3.13	12	1200		β3
His-22	8.79	5.16	3.19	12	500	insufficient data	
			2.91	4	ovlp[b]		
Leu-26	6.92	4.41	1.55	5	500	180° ± 30°	β2
			1.44	10	500		β3
Asp-27	8.02	4.34	3.17	8	250	nonstandard or multiple rotamer states	
			2.69	7	400		
Ser-28	7.19	4.60	3.65	4	250	+60° ± 30°	β2
			3.52	4	400		β3
Tyr-29	8.43	5.24	2.48	3	150	−60° ± 30°	β3
			2.28	12	500		β2
Cys-31	8.64	5.29	2.77	12	500	−60° ± 30°	β2
			2.59	3	200		β3
Asn-32	9.41	5.01	2.98	7	500	nonstandard or multiple rotamer states	
			2.76	7	ovlp		
Asp-46	8.50	4.62	2.75	7	400	nonstandard or multiple rotamer states	
			2.48	7	800		
Arg-48	8.19	3.82	1.43	10	ovlp	insufficient data	
			1.30	5	ovlp		
Trp-49	7.58	4.21	3.22	7	1000	nonstandard or multiple rotamer states	
			3.12	7	1000		
Trp-50	7.14	4.25	3.05	6	ovlp	nonstandard or multiple rotamer states	
			2.77	5	ovlp		

[a] Peak intensity in τ_m = 65-ms NOESY spectrum, in arbitrary units. [b] ovlp, overlap. The peak from which the passive coupling or NOE would be measured is overlapped with another peak or is otherwise inaccessible in the COSY-30 or τ_m = 65-ms NOESY spectrum.

Supercomputer Facility. Optimized superpositions of atomic coordinates were calculated using the algorithm of Roterman et al. (1989).

In eq 2, α is a weighting factor determining the relative weights to be allotted to the restraints and the ECEPP energy; thus, if $\alpha = 1$, only the restraint function is minimized, whereas if $\alpha = 0$, only the energy is minimized. By starting with a value of α close to 1 and gradually decreasing α, it should be possible at some stage to obtain structures in which the majority of restraints are still satisfied while at the same time most of the interatomic electrostatic and nonbonded interactions are energetically favorable.

Results

A nearly complete set of sequence-specific resonance assignments for mEGF at pH 3.1 has been presented elsewhere (Montelione et al., 1988). These assignments are in excellent agreement with those published subsequently for mEGF at pH 2.0 (Kohda & Inagaki, 1988), with the exception of the H^{γ} and H^{δ} methylene protons of Arg-41, which are unusually broad and difficult to identify reliably in 2QF-COSY and TOCSY spectra. For the structure refinement described here, additional stereospecific proton resonance assignments were obtained for mEGF at pH 3.1.

Stereospecific Assignments of $C^{\beta}H_2$ Methylene Protons. Of the 53 amino acid residues in mEGF, 41 have C^{β} methylene protons. Of these 41 side chains, 14 have nearly degenerate $C^{\beta}H_2$ chemical shifts ($\Delta\delta < 0.1$ ppm), viz., Asn-1, Ser-2, Ser-8, Ser-9, Asp-11, Tyr-13, Met-21, Glu-24, Ser-25, Tyr-37, Ser-38, Asp-40, Arg-45, and Leu-52, which precluded their stereospecific assignment. For all of these 14 side chains, the degeneracy of the methylene proton resonances was proven by identification of remote two-quantum cross peaks at [ω_1 = ($H^{\beta 1}$ + $H^{\beta 2}$), ω_2 = H^{α}] (Wagner & Zuiderweg, 1983; Montelione et al., 1988). In addition, for 10 of the remaining 25 non-proline residues, the COSY-30 cross peaks between H^{α} and one or both H^{β}'s were not sufficiently well resolved from other cross peaks to provide reliable estimates of the vicinal coupling constants. For the remaining 15 residues, well-resolved H^{α}–H^{β} cross peaks were obtained. Representative expansions of these COSY-30 cross peaks are shown in Figure 1, and the data obtained from their analysis are summarized in Table II.

In our analysis, we assumed that the protein side chains adopt one of the three staggered conformations with χ^1 = +60°, 180°, or −60°. A recent analysis of protein crystal structures (Ponder & Richards, 1987) reveals that about 90% of the buried side chains in globular proteins satisfy this assumption. In mEGF there are also many surface side chains which may adopt multiple conformations that interconvert rapidly on the NMR time scale. For this reason, special care is needed to avoid interpreting ensemble-averaged coupling constants arising from multiple side-chain conformations in terms of single conformational states. Therefore, $^3J(H^{\alpha}$–$H^{\beta})$ vicinal coupling constants were used as conformational constraints only if *both* $^3J(H^{\alpha}$–$H^{\beta})$ values were consistent with one of the staggered rotamer states for χ^1 (Nagayama & Wüthrich, 1981). When these criteria were fulfilled, intra-

Table III: Individual Side-Chain Amide Proton Assignments for mEGF at 28 °C and pH 3.1

amino acid	chemical shifts (ppm) H^{β}	H^{γ}	H^{δ}	H^{ϵ}	H^{β}–H^{δ} NOE[a]	H^{γ}–H^{ϵ} NOE[a]	stereospecific assignment
Asn-1	2.86[b]		7.58		600		δ2
			6.99		<100		δ1
Asn-16	1.98		8.03		600		δ2
	1.37				200		
	1.98		7.21		150		δ1
	1.37				<100		
Asn-32	2.98		7.23		400		δ2
	2.76				400		
	2.98		6.80		ovlp[c]		δ1
	2.76				<100		
Gln-43		2.18		7.64		300	ε2
		1.77				600	
		2.18		6.89		<250[d]	ε1
		1.77				<250[d]	

[a] Peak intensity in the τ_m = 65-ms NOESY spectrum, in arbitrary units. [b] The two H^{β} protons are degenerate in chemical shift. [c] ovlp, overlap. The peak which corresponds to this NOE overlaps with another peak or is otherwise inaccessible in the τ_m = 65-ms NOESY spectrum. [d] The corresponding NOESY cross peak partially overlaps with another cross peak in the τ_m = 65-ms NOESY spectrum. The intensity reported represents an upper bound for the actual NOE intensity.

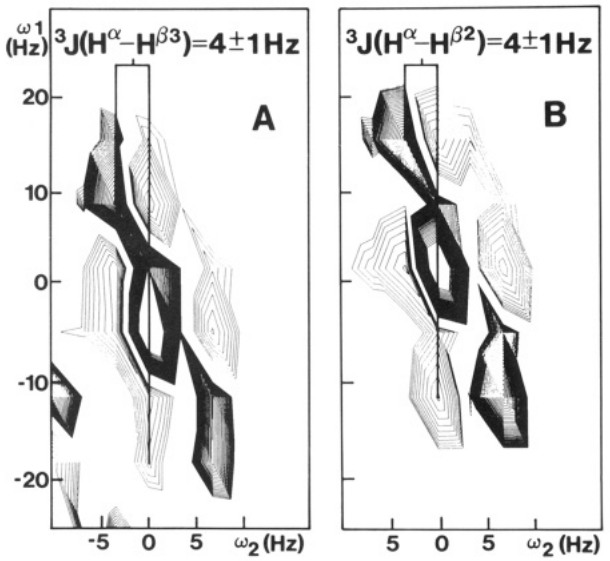

FIGURE 1: Expansions of COSY-30 cross peaks of mEGF at pH 3.1, 28 °C, and 1 mM protein concentration. (A) H^{α}–$H^{\beta 2}$ COSY-30 cross peak of Ser-28 manifesting $^3J(H^{\alpha}–H^{\beta 3})$. (B) H^{α}–$H^{\beta 3}$ COSY-30 cross peak of Ser-28 manifesting $^3J(H^{\alpha}–H^{\beta 2})$.

residue NOE data were used to obtain the individual $C^{\beta}H_2$ methylene assignments and to identify a unique staggered rotamer conformation (Table II). For the standard staggered χ^1 rotamer states of amino acid side chains, the H^{α} and H^{β} atoms are either gauche or trans, with $^3J(H^{\alpha}–H^{\beta})$ vicinal coupling constants of 3 ± 2 Hz and 12 ± 2 Hz, respectively (Bystrov, 1976). For 9 of the 15 mEGF side chains listed in Table II, i.e., Cys-14, Asn-16, Cys-20, His-22, Leu-26, Ser-28, Tyr-29, Cys-31, and Arg-48, the measured values of $^3J(H^{\alpha}–H^{\beta})$ were within these ranges. For seven of these side chains, the H^N–H^{β} NOE data then provided stereospecific assignments and a determination of χ^1 (Table II). For His-22 and Arg-48, the cross peaks with $C^{\beta}H_2$ were not sufficiently well resolved in the NOESY spectra.

For the remaining 6 of the 15 residues for which both $^3J(H^{\alpha}–H^{\beta})$ coupling constants were available, i.e., Tyr-3, Asp-27, Asn-32, Asp-46, Trp-49, and Trp-50, the coupling constants were inconsistent with any single staggered rotamer state (Nagayama & Wüthrich, 1981). These $^3J(H^{\alpha}–H^{\beta})$ coupling constants (Table II) may arise either from fixed nonstaggered χ^1 conformations or from rapid averaging between at least two staggered rotamers. With the exception of Asn-32, all of these side chains have high solvent accessibility.

Stereospecific Assignments of Valine Isopropyl Methyl Groups. On the basis of intraresidue NOEs with H^N and H^{α}, the upfield and downfield methyl resonances of Val-34 were assigned as $C^{\gamma 1}H_3$ and $C^{\gamma 2}H_3$, respectively, and χ^1 was found to be +60 ± 30°. For Val-19 the ^1H chemical shifts of the methyl resonances are degenerate.

Individual Assignments of Side-Chain Amide Protons. There are three asparagines and one glutamine in mEGF. The individual side-chain amide protons of these residues have distinct chemical shifts because the exchange by amide-bond isomerization is slow on the chemical shift time scale. On the basis of the data summarized in Table III, individual assignments were made for all eight of these amide protons. In all four side chains, the upfield and downfield amide protons are assigned as $H^{\delta 1}$ (or $H^{\epsilon 1}$) and $H^{\delta 2}$ (or $H^{\epsilon 2}$), respectively.

Experimental Conformational Constraints. Five kinds of experimental conformational constraints were used in the input for the structure determination: (i) NOE-derived upper bound ^1H–^1H distance constraints; (ii) constraints on ranges of backbone dihedral angles ϕ from vicinal $^3J(H^N–H^{\alpha})$ coupling constants; (iii) constraints on ranges of side-chain dihedral angles χ^1 determined by combined analysis of COSY-30 and NOESY data; (iv) constraints accounting for the previously identified interstrand hydrogen bonds in the β-sheet (Montelione et al., 1986); (v) disulfide-bond constraints. A complete list of the conformational constraints used as input for the structure calculations has been deposited in the Brookhaven Protein Data Bank, together with the atomic coordinates.

The NOE-derived upper bound distance constraints were obtained as described under Methods and Materials (Table I). Backbone vicinal coupling constants $^3J(H^{\alpha}–H^N)$ were estimated from ω_2 cross-sections of H^N–H^{α} cross peaks of two-quantum spectra of mEGF recorded as described elsewhere (Montelione et al., 1986, 1987, 1988). Special care was taken to account for cancellation within the antiphase cross peaks, which results in artificially enlarged splittings for coupling constants smaller than the line widths (Neuhaus et al., 1985; Wüthrich, 1986). In mEGF spectra recorded at 28 °C, most H^N–H^{α} cross peaks have line widths of ca. 7.0 Hz, and the smallest apparent values of $^3J(H^{\alpha}–H^N)$ were ca. 5.0 Hz. Therefore, only cross peaks with apparent $^3J(H^{\alpha}–H^N)$ couplings >8.0 Hz were used in the structure determination,

Table IV: Numbers of Experimental Distance Constraints Used in Determining the Solution Structure of Murine Epidermal Growth Factor

NOE-derived upper bound constraints			dihedral angle constraints		hydrogen-bond constraints	disulfide constraints
intraresidue	sequential	longer range	backbone	side chain		
93	203	348	24	8	36[a]	18[b]

[a] There are two upper bound and two lower bound distance constraints for each hydrogen bond. [b] There are three upper bound and three lower bound constraints for each disulfide bond.

with the corresponding dihedral angles ϕ restricted to the range (−160°, −80°) (Pardi et al., 1984). The following residues were thus constrained: Ser-2, Cys-6, Tyr-10, Cys-14, Leu-15, Val-19, Cys-20, Met-21, His-22, Ile-23, Leu-26, Ser-28, Tyr-29, Thr-30, Cys-31, Asn-32, Val-34, Tyr-37, Ser-38, Arg-45, Asp-46, Trp-50, Glu-51, and Leu-52. The constraints on the ranges of the side-chain dihedral angles χ^1 are tabulated in Table II.

Interstrand hydrogen bonds in the β-sheet structures were identified by locating the slowly exchanging amide protons of mEGF in the secondary structures (Montelione et al., 1986) found on the basis of NOE pattern recognition (Wüthrich et al., 1984). Hydrogen-bond constraints were used only if the pattern of interstrand NOEs was sufficient to identify the acceptor carbonyl oxygen, and in addition the donor amide proton had an exchange half-life greater than several hours at pH 3.1 and 28 °C (Montelione et al., 1988). The following nine hydrogen-bonded pairs were identified on this basis: Val-19 H^N···Asn-32 O′; Met-21 H^N···Thr-30 O′; Ile-23 H^N···Ser-28 O′; Thr-30 H^N···Met-21 O′; Asn-32 H^N···Val-19 O′; Val-34 H^N···Tyr-37 O′; Tyr-37 H^N···Val-34 O′; Ser-38 H^N···Thr-44 O′; Thr-44 H^N···Ser-38 O′. Upper bound and lower bound H^N···O and N···O distance constraints for each of these nine hydrogen bonds, as well as the S···S and C^β···S distance constraints which fix bond lengths and bond angles in the three disulfide bonds (Savage et al., 1973), were defined as described elsewhere (Williamson et al., 1985; Wüthrich, 1986).

In summary, for the calculations of the final structures used to represent the solution conformation of mEGF, 730 conformational constraints were used (Table IV), including 644 NOE-derived upper bound distance constraints, 32 dihedral angle constraints, and 54 upper and lower bound distance constraints associated with the nine hydrogen bonds and the three disulfide bonds. A survey of the medium-range and long-range NOE-derived upper distance constraints is presented in Figure 2.

Determination of the Solution Structure. The distribution of NOE-derived upper distance constraints (Figure 2) indicates that the mEGF structure is divided into two slightly overlapping subdomains (Montelione et al., 1986) corresponding to residues 1–34 and 30–53, respectively. The relative orientation of the two subdomains is determined by the conformation of the polypeptide segment Thr-30–Val-34, which contributes to both subdomains, and by 57 NOE-derived constraints between the polypeptide segments Tyr-13–Cys-31 and Cys-33–Gln-43.

For each amino acid residue, the conformational space of ϕ, ψ, and χ^1 was initially restricted to those regions which satisfy the intraresidue and sequential NOEs and the spin–spin coupling constraints (Wüthrich, 1986). A total of 200 starting conformations of the entire 53-residue mEGF molecule were generated with values for the dihedral angles selected randomly *within these allowed regions of* ϕ, ψ, *and* χ^1. All peptide bonds were kept fixed in the planar trans conformation and the DISMAN target function (Braun & Gō, 1985) for each structure was minimized. Of the resulting structures, the 16 with the smallest residual values of the target function at level 53 were selected for analysis and further energy refinement.

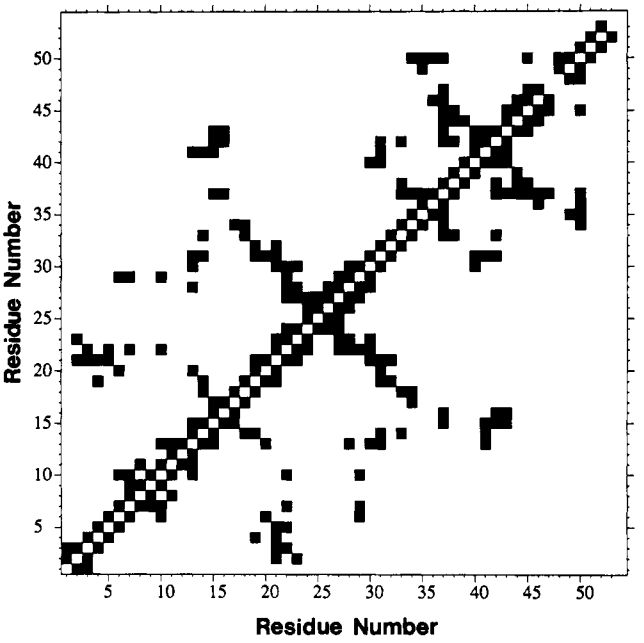

FIGURE 2: Survey of NOE-derived distance constraints for mEGF in a diagonal plot. Each axis is labeled with the amino acid sequence numbers of mEGF, and squares connect pairs of residues linked by one *or more* NOE constraints.

Restrained energy minimization was carried out as described under Materials and Methods, starting with conformations generated from the dihedral angles of these 16 structures. The outermost side-chain dihedral angles of all Arg residues, which were not specified in the DISMAN output, were set to $\chi^5 = 0°$, $\chi^6 = 180°$, and $\chi^7 = 0°$. Three steps of restrained energy minimization were carried out as follows. (i) The restrained energy function g (see eq 2) was minimized with $\alpha = 0.975$. (ii) In 9 of the 16 minimized conformations from step 1, χ^3 of Asn-1 was close to 0° rather than 180°. Since this interchanges the $H^{\delta 1}$ and $H^{\delta 2}$ atoms and thus compromises the NOE assignment involving the $H^{\delta 2}$ atom, χ^3 of Asn-1 was adjusted by adding or subtracting 180° in those conformations. The restrained energy function g was then minimized again, with $\alpha = 0.95$, starting from the (adjusted) dihedral angles from step 1. (iii) The final dihedral angles from step 2 were used as starting points for minimization of the function g with $\alpha = 0.9$.

A summary of the residual constraint violations and conformational energies of the final 16 mEGF structures before and after restrained energy minimization is presented in Table V. Although there are few constraint violations greater than 1 Å in these structures, there are many residual violations of less than 1 Å. These small (<1 Å) constraint violations may be due to many causes including incorrect constraints which are tighter than the true experimental distances or the presence of multiple conformations of surface residues which give rise to inconsistent NOEs.

The ECEPP/2 minimization lowers the potential energies of the DISMAN structures significantly without seriously violating the experimental constraints (Table V). When comparing the results for the DISMAN structures before and after

Table V: Summary of Residual Constraint Violations and Conformational Energies for the Final 16 mEGF Conformers Used to Represent the Solution Conformation

structure[a]	no. of lower-bound constraint violations		no. of upper bound constraint violations		no. of steric constraint violations		no. of dihedral angle constraint violations >20°		ECEPP/2 energy (kcal/mol)
	0.2–0.5 Å	>0.5 Å	0.5–1.0 Å	>1.0 Å	0.1–0.3 Å	>0.3 Å	ϕ	χ^1	
1D	3 (0.47)	0	30 (0.96)	0	13 (0.27)	0	0	3	4.39×10^3
1R	0	0	26	1 (1.20)	5	2 (0.38)	0	3	-1.89×10^2
2D	5	2 (0.99)	21 (0.94)	0	10 (0.22)	0	0	5	3.96×10^3
2R	0	0	19 (0.68)	0	10	2 (0.33)	0	3	-1.84×10^2
3D	2	1 (0.77)	16 (0.89)	0	11	1 (0.36)	1	3	2.11×10^5
3R	0	0	21 (0.96)	0	15	2 (0.35)	1	3	-1.66×10^2
4D	2	1 (1.02)	23 (0.87)	0	9 (0.21)	0	1	3	1.12×10^3
4R	0	0	20	1 (1.11)	8	1 (0.31)	0	3	-1.97×10^2
5D	1	2 (0.91)	30	1 (1.01)	8 (0.20)	0	1	4	1.41×10^3
5R	0	0	20	1 (1.01)	9 (0.29)	0	1	3	-1.72×10^2
6D	4	1 (1.08)	24 (0.87)	0	9 (0.22)	0	0	3	3.13×10^3
6R	0	0	28 (0.92)	0	8	1 (0.33)	0	3	-1.99×10^2
7D	5	1 (0.97)	23	2 (1.18)	14	1 (0.41)	1	5	1.22×10^9
7R	0	0	21	1 (1.04)	14 (0.30)	0	0	3	-2.09×10^2
8D	3 (0.43)	0	26 (0.93)	0	5 (0.18)	0	1[b]	3	9.62×10^3
8R	0	0	23 (0.88)	0	8	1 (0.31)	1[b]	3	-1.98×10^2
9D	9	1 (1.07)	20 (0.95)	0	5	1 (0.36)	1	2	1.05×10^3
9R	0	0	22 (0.98)	0	6	1 (0.35)	0	2	-1.92×10^2
10D	3	2 (1.09)	19 (0.94)	0	11 (0.25)	0	0	2	1.29×10^3
10R	0	0	26	1 (1.16)	9 (0.29)	0	0	1	-2.11×10^2
11D	3	1 (0.59)	26 (0.95)	0	17 (0.29)	0	1	3	5.41×10^4
11R	0	0	22	1 (1.08)	5	3 (0.33)	1	3	-2.06×10^2
12D	3	1 (0.60)	22 (0.96)	0	10	2 (0.43)	1	3	1.03×10^3
12R	0	0	35 (0.98)	0	6	1 (0.34)	1	3	-1.90×10^2
13D	3	1 (1.16)	23	2 (1.16)	11	1 (0.34)	1	2	1.73×10^3
13R	0	0	21	3 (1.12)	7	2 (0.34)	0	2	-1.53×10^2
14D	5 (0.47)	0	22	2 (1.12)	8 (0.17)	0	0	3	1.35×10^3
14R	0	0	23	1 (1.03)	5	1 (0.31)	0	3	-2.13×10^2
15D	2 (0.45)	0	22 (0.97)	0	5 (0.23)	0	2[c]	1	2.59×10^2
15R	0	0	22 (0.99)	0	8 (0.26)	0	1[c]	1	-1.79×10^2
16D	5 (0.44)	0	18	2 (1.22)	14 (0.26)	0	1	3	1.12×10^3
16R	0	0	26	1 (1.35)	6	1 (0.34)	1	3	-1.59×10^2

[a] D, residual constraint violations for the DISMAN structures before energy minimization; R, residual constraint violations after restrained energy minimization determined with the ECEPP/2 program. [b] $\phi > 0$ for Asp-46. [c] $\phi > 0$ for Tyr-10.

restrained energy minimization, it should be remembered that *any* steric overlap of two atoms that might possibly be involved in a hydrogen bond (e.g., backbone –H···O) is *never* computed by DISMAN; by contrast, the penalty function used in computing the refined structures included all these overlaps. Most of the steric overlaps exceeding 0.3 Å, and nearly all the violations of lower bound constraints exceeding 0.2 Å, were removed in step 1 of the energy refinement, but this step was not very effective at removing violations of dihedral-angle constraints. At the end of step 2, all violations of lower bound constraints exceeding 0.2 Å had been removed, and many of the dihedral-angle violations had disappeared. In 10 out of the 16 conformations, the number of upper bound violations exceeding 0.5 Å had decreased; in 5 out of the 16 conformations, the maximum violation of upper bound constraints had increased somewhat (up to 0.25 Å). Step 3 produced a further reduction in the number of steric overlaps, at the expense of a significant increase in the number of violations of upper bound constraints. Because of the increase in the number of violations of upper bound constraints at step 3, structural analysis was carried out on the conformations obtained at the end of step 2; these are referred to as the "refined structures".

Table VI: Comparison of the 16 mEGF Conformers of Table V, Which Are Used To Represent the Solution Structure, before and after Energy Minimization

polypeptide segment[a]	mean RMSD	
	backbone atoms[b]	heavy atoms[c]
Asn-1–Cys-33 (D)	1.16 (0.62–1.72)	1.70 (1.12–2.44)
Asn-1–Cys-33 (R)	1.01 (0.37–1.52)	1.49 (0.99–2.22)
Asn-32–Leu-47 (D)	1.25 (0.51–2.23)	2.18 (1.32–3.64)
Asn-32–Leu-47 (R)	1.14 (0.53–2.26)	2.14 (1.34–3.38)
Asn-1–Leu-47 (D)	1.51 (1.00–2.28)	2.15 (1.62–3.12)
Asn-1–Leu-47 (R)	1.33 (0.67–2.16)	1.92 (1.37–2.52)
Asn-1–Trp-50 (D)	1.94 (1.23–3.14)	2.87 (1.86–4.00)
Asn-1–Trp-50 (R)	1.90 (0.97–3.42)	2.75 (1.74–4.08)

[a] D, DISMAN structure; R, structure after restrained energy minimization. [b] Average (range) of the pairwise RMSDs calculated for the N, C^α, and C′ backbone atoms and all pairwise combinations of the 16 structures. [c] Average (range) of the pairwise RMSDs calculated for all heavy atoms and all pairwise combinations of the 16 structures.

Superpositions of the C^α coordinates for the 16 DISMAN structures before and after energy minimization are shown in Figure 3, and statistics for the superpositions of backbone and

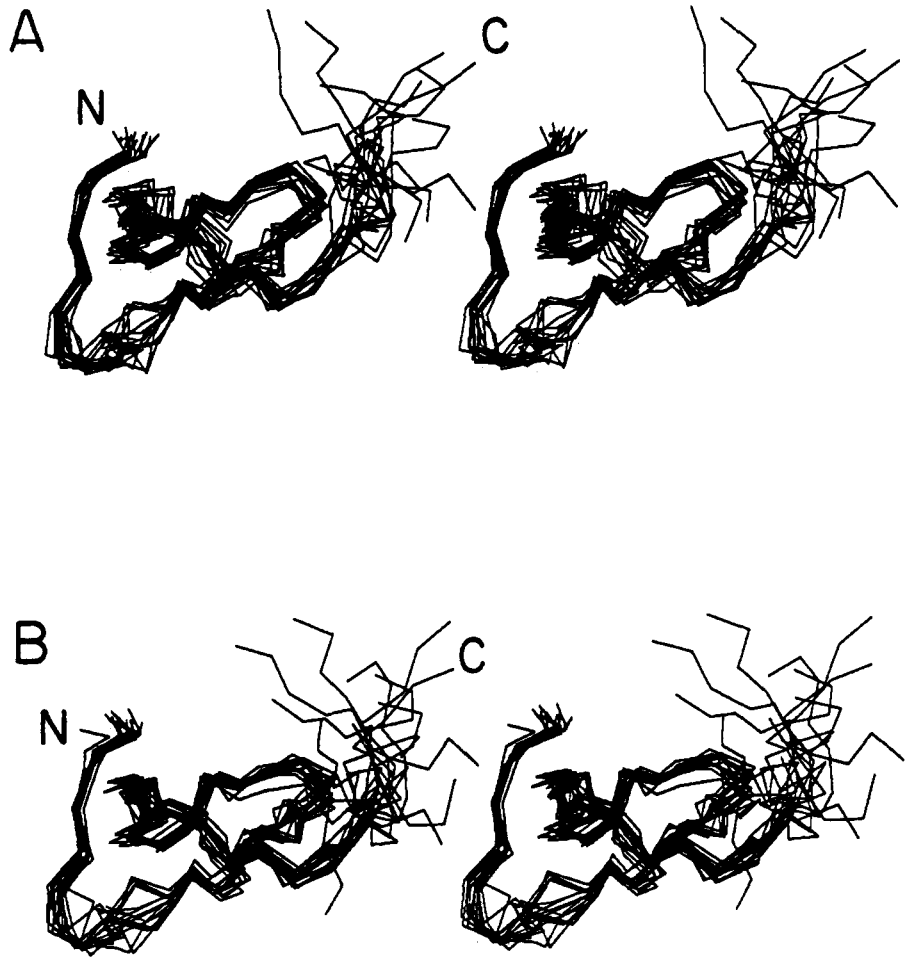

FIGURE 3: Stereoview showing superpositions of the polypeptide backbones of the 16 conformers for residues 1–53. Each set of superpositions was generated using conformer 9 (Table V) as the reference and orienting the other conformers so as to minimize the pairwise RMSD for the backbone atoms of residues 2–47. Only the C^{α} atoms are shown. (A) DISMAN conformers; (B) energy-refined conformers.

heavy atoms for these structures are presented in Table VI. For residues 1–47, the pairwise RMSDs for N, C^{α}, and C′ atoms range from 1.0 to 2.3 Å (1.5 Å averaged over all pairs) and 0.7 to 2.2 Å (1.3 Å averaged over all pairs) before and after energy minimization, respectively. For the same set of residues, the corresponding RMSD values for all heavy atoms range from 1.6 to 3.1 Å (2.2 Å averaged over all pairs) for the DISMAN structures and 1.4 to 2.5 Å (1.9 Å averaged over all pairs) for the energy-minimized structures. For residues 1–50 and 1–53, the backbone and heavy atom RMSDs are significantly higher since the positions of atoms in residues 48–53 are not well determined from the available NMR data (Figure 3). Stereopairs of a typical complete mEGF DISMAN structure before and after energy refinement are shown in Figure 4. For residues 1–47, the RMSDs (averaged over all pairs) between structures before and after energy minimization are 1.7 and 2.2 Å for the backbone and all heavy atoms, respectively. Coordinates for these 16 mEGF DISMAN conformers before and after energy refinement have been deposited in the Brookhaven Protein Data Bank.

The quality of the structure determination within the N-terminal and C-terminal subdomains is better than for the overall molecule. For the polypeptide segments Asn-1 to Cys-33 and Asn-32 to Leu-47 the average backbone RMSDs either before (1.2 and 1.3 Å, respectively) or after (1.0 and 1.1 Å, respectively) restrained energy minimization are significantly lower than the corresponding values for the entire segment of residues Asn-1–Leu-47 (1.5 Å before and 1.3 Å after energy minimization). Stereodiagrams showing superpositions of the backbone atoms for the two subdomains of mEGF are shown in Figure 5.

ϕ–ψ plots provide a useful tool for evaluating the quality of a protein structure determination, since amino acid residues in proteins generally adopt backbone conformations corresponding to low-energy minima of amino acid residues in small peptides (Roterman et al., 1989). Composite ϕ–ψ plots for residues 2–47 of all 16 DISMAN conformers before and after constrained energy minimization are shown in Figures 6 and 7, respectively. In every DISMAN conformer, before or after energy minimization, at least 3 Gly residues and between 3 and 10 non-Gly residues have positive ϕ values. A summary of the frequency of positive ϕ values in these conformers is presented in Table VII. In all 16 conformers, Asp-27, Gly-36, and Gly-39 adopt positive ϕ values. In addition, in 15 of the refined conformers Asn-16 and Gly-17 also have positive ϕ values. None of these residues have large $^3J(H^N–H^{\alpha})$ coupling constants, which would indicate a negative ϕ dihedral angle, so that it appears that the residues Asn-16, Gly-17, Asp-27, Gly-36, and Gly-39 have positive ϕ values in the solution structure of mEGF. For the remaining residues listed in Table VII, there remains some uncertainty regarding the backbone ϕ conformation. This is largely due to the scarcity of local conformational constraints. Although the global polypeptide fold is quite well defined, concerted variations of two or several dihedral angles may thus be accommodated.

DISCUSSION

Description of the Solution Structure of mEGF. Insofar

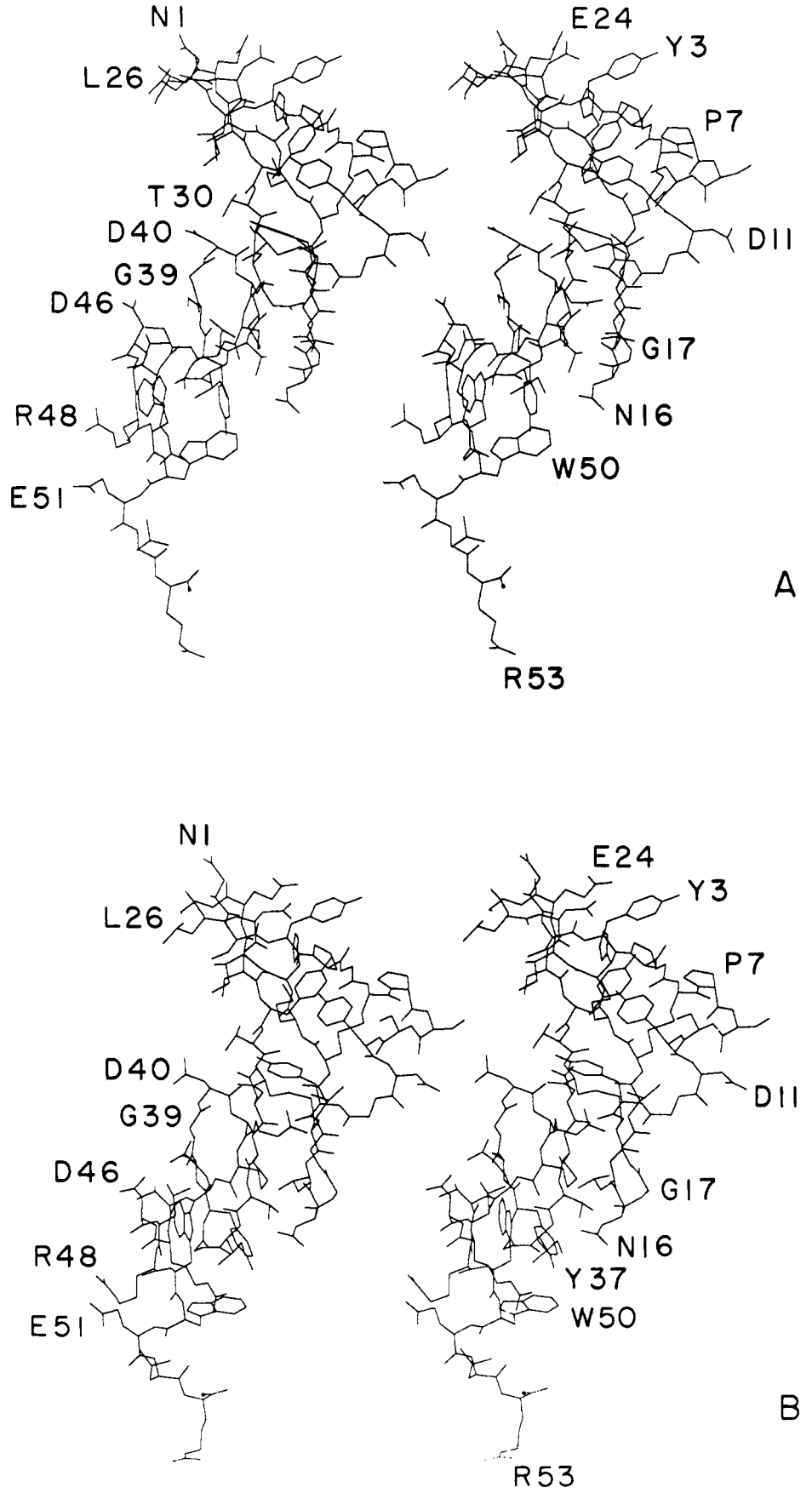

FIGURE 4: Stereoview of conformer 9 (Table V) of mEGF. (A) DISMAN conformer; (B) energy-refined conformer, oriented so as to minimize the RMSD for the backbone atoms of residues 1–50 with respect to the conformer in panel A.

as the results of the two investigations can be compared, the solution structure of mEGF described here is similar to that described earlier (Montelione et al., 1987), which was based on DISMAN calculations with a much smaller set of conformational constraints. The mEGF structure contains two overlapping subdomains: the N-terminal subdomain (residues

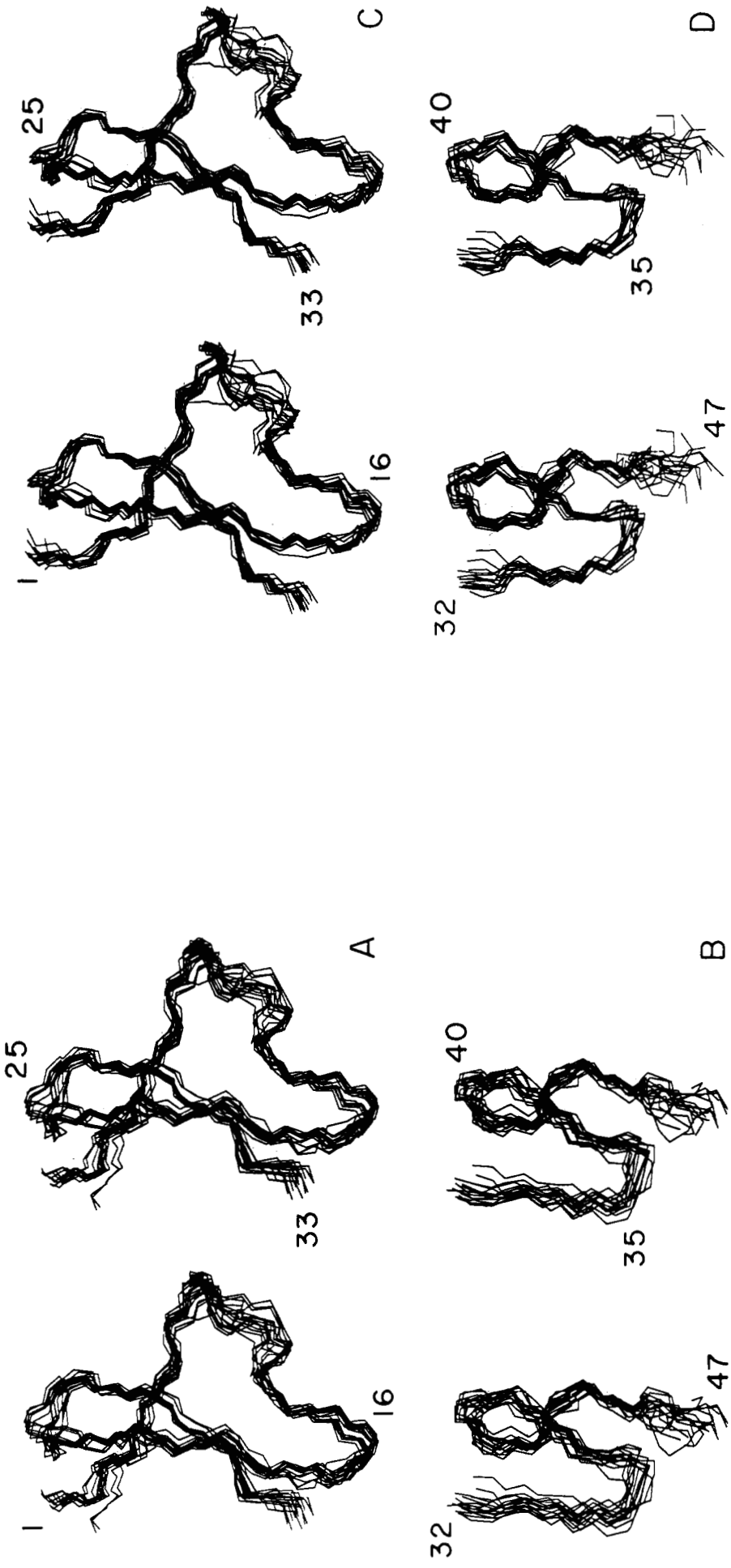

FIGURE 5: Stereoview showing superpositions for the polypeptide backbones of the 16 conformers for (i) residues 1–33 and (ii) residues 32–47. Each set of superpositions was generated using conformer 9 (Table V) as the reference and orienting the other conformers so as to minimize the pairwise RMSD for the backbone atoms of either residues 1–33, or residues 32–47. The N, C^{α}, and C′ backbone atoms are shown. (A) residues 1–33, DISMAN conformers; (B) residues 32–47, DISMAN conformers; (C) residues 1–33, energy-refined conformers; (D) residues 32–47, energy-refined conformers.

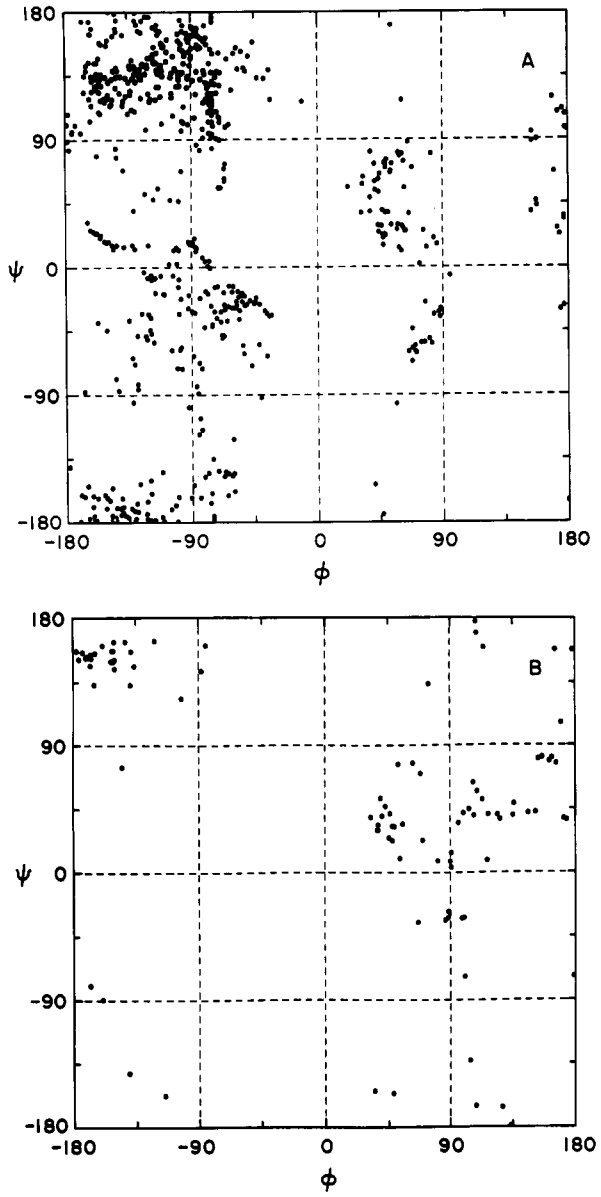

FIGURE 6: Composite ϕ–ψ map for residues 2–47 of the 16 DISMAN conformers of mEGF. (A) Non-glycine residues. (B) Glycine residues.

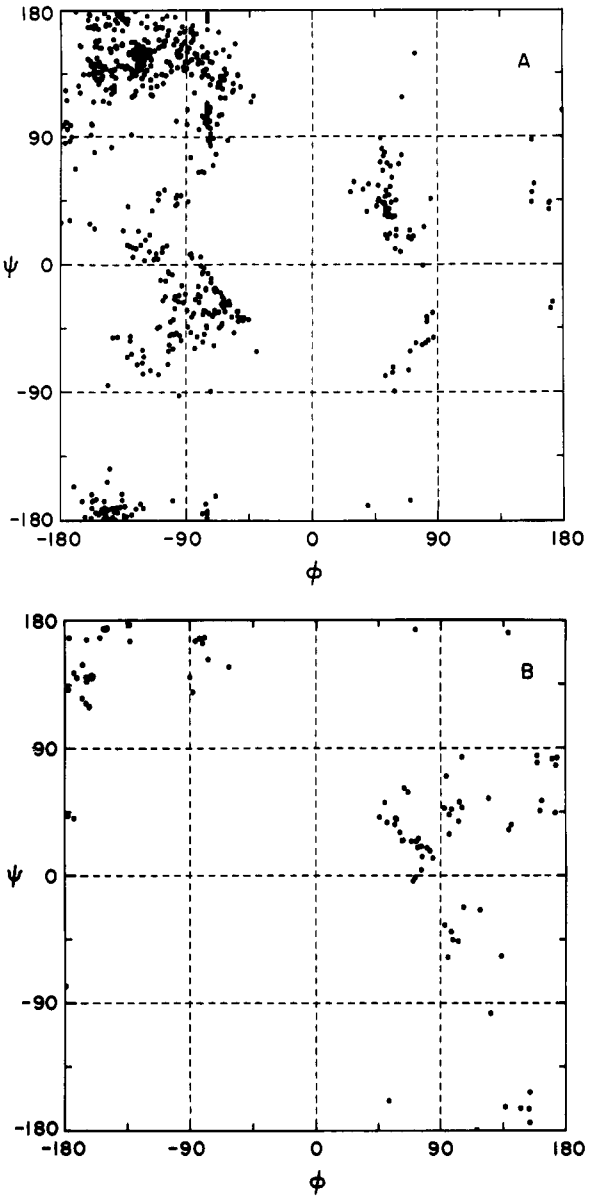

FIGURE 7: Composite ϕ–ψ map for residues 2–47 of the 16 energy-refined conformers of mEGF. (A) Non-glycine residues. (B) Glycine residues.

Asn-1–Val-34) containing the A (Cys-6–Cys-20) and B (Cys-14–Cys-31) disulfide loops, and the C-terminal subdomain (Thr-30–Arg-53) containing the C loop (Cys-33–Cys-42). The core structure of the N-terminal subdomain is a small antiparallel β-sheet involving residues Gly-18–Ile-23 and Ser-28–Cys-33. Residues Glu-24–Asp-27 form the chain reversal of this hairpin β-sheet, with a type I β-bend at Ser-25–Leu-26 in 10 out of the 16 energy-refined conformers. Residues Tyr-3–Pro-4 form a small stretch of a third strand of this antiparallel β-sheet, interacting with the polypeptide segment Met-21–His-22 in the predominant solution structure. This triple-stranded β-sheet structure has an overall right-handed twist, like that observed in most antiparallel β-sheets in globular proteins. The remaining polypeptide segment of the N-terminal domain, Cys-6–Cys-14, forms a multiple-bend structure, reminescent of a severely distorted α-helix. The backbone conformation in this polypeptide segment is better defined in the presently described structure of mEGF than in previous structures of mEGF (Montelione et al., 1987) or hEGF (Cooke et al., 1987). It is folded against the central β-sheet by side-chain to side-chain interactions, including the Cys-6–Cys-20 and Cys-14–Cys-31 disulfide bonds. The central β-sheet positions residues Cys-20 and Cys-31 opposite one another in antiparallel strands, with their side chains pointing in the same direction, to make disulfide bonds at two ends of the Cys-6–Cys-14 segment. Considering the unique role of these disulfide bonds in the polypeptide chain fold, it is not surprising that the sequence positions of these cysteine residues are conserved within the extensive family of EGF-like molecules (Campbell et al., 1989). Packing interactions between the Cys-6–Cys-20 polypeptide segment and the triple-stranded β-sheet include a partially solvent-inaccessible cluster formed by the Tyr-3, Tyr-10, Tyr-13, His-22, and Tyr-29 side chains. In superpositions of heavy atom atomic coordinates, the positions of most side chains in these residues are quite well determined.

The intensities of the backbone–backbone NOEs beteen the polypeptide segments Tyr-3–Pro-4 and Met-21–His-22 are much weaker than those of other antiparallel strand interactions in the protein, despite the fact that the proton resonance line widths of residues Asn-1–Cys-6 are relatively sharp. In addition, the hydrogen bonds between this third strand and the central β-strand do not have slowly exchanging amide

Table VII: Incidence of Positive ϕ Values among the 16 Conformers of mEGF of Table V

residue	no. of conformations for which $\phi > 0$ before energy refinement	after energy refinement
Ser-8	1	0
Ser-9	1	1
Tyr-10	1	1
Asp-11	8	8
Gly-12	16	12
Tyr-13	4	4
Asn-16	15	15
Gly-17	13	15
Gly-18	2	1
Glu-24	3	3
Ser-25	6	6
Asp-27	16	16
Gly-36	16	16
Gly-39	16	16
Asp-40	7	7
Arg-41	5	5
Cys-42	15	12
Asp-46	1	1

protons (Montelione et al., 1988) as do most hydrogen bonds in the other β-sheets. These observations indicate that the three-stranded β-sheet conformation of mEGF is in dynamic equilibrium at pH 3 and 28 °C with partially unfolded molecules which differ with respect to the conformation of polypeptide segment Asn-1 to Cys-6. A similar dynamic structure has been proposed for the third strand of the homologous β-sheet in hEGF (Cooke et al., 1987) and human TGFα (Brown et al., 1989; Kohda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Kline et al., 1990). Many 2D NMR cross peaks for residues Ser-2–Cys-6 and Cys-20–Ile-23 are split into major (90%) and minor (10%) forms, which could not be removed by chromatographic repurification of the proteins and are also observed in recombinant mEGF (Moy et al., 1989), suggesting slow conformational exchange processes among different conformers of the polypeptide backbone. Although these data (including sharp proton line widths, rapid amide proton exchange, and weak interstrand NOEs) provide good evidence for a loose association of this third strand with the rest of the protein, these uncertainties in the structure determination are not realistically displayed in superpositions of the mEGF DISMAN conformers either before or after energy refinement (Figures 3 and 5).

The core structure of the C-terminal subdomain (residues Thr-30–Arg-53) has a "double hairpin" architecture (Montelione et al., 1986), in which the polypeptide segments Cys-33–Val-34, Tyr-37–Ser-38, and Thr-44–Asp-46 form short strands of a small three-stranded antiparallel β-sheet. This β-sheet is stabilized by interstrand hydrogen bonds between backbone atoms and many side-chain to side-chain interactions, including hydrophobic interactions between the side chains of Val-34, Tyr-37, and the aliphatic portion of Arg-45. The first chain reversal in the C-terminal subdomain is a well-defined type II β-bend at residues Ile-35–Gly-36 in all 16 energy-refined conformers with $\phi_{35} = -35°$ to $-77°$, $\psi_{35} = 100°$ to $160°$, $\phi_{36} = 60°$ to $120°$, and $\psi_{36} = -40°$ to $40°$. The preference for glycine at position 2 of a type II β-bend may thus explain why Gly-36 is very strongly conserved in EGF-like proteins (Campbell et al., 1989). The second chain reversal in this "double hairpin" architecture is a polypeptide loop from Gly-39 to Gln-43, which forms an unusual multiple bend and is tethered back to the protein by the Cys-33–Cys-42 disulfide bond. In order to form this disulfide bond, the polypeptide chain is forced to form one turn of a left-handed helix (Montelione et al., 1987). This conformation requires the presence of one or more residues with positive ϕ dihedral-angle values. Residues Gly-39, Asp-40, Arg-41, and Cys-42 have positive ϕ values in 16, 7, 5, and 12 of the 16 energy-refined DISMAN conformers, respectively. The variations in the local backbone conformations of this polypeptide segment among the 16 energy-refined conformers is compensated by dihedral-angle differences in other residues, resulting in a consistent overall chain fold for the polypeptide segment of residues Thr-30–Leu-47 (Figure 5B,D). Residues Arg-41 and Leu-47, which appear to be involved in interactions between EGF and its membrane-bound receptor (Moy et al., 1989; Engler et al., 1990), are located in this C-terminal subdomain.

The conformation for the residues Arg-48–Arg-53 is not uniquely defined by the available NMR data (Figure 2). In addition to having few medium-range and long-range NOEs, these six C-terminal residues have sharp proton line widths and H^{α} and H^{N} chemical shift values close to the statistical-coil values (Wüthrich, 1986), suggesting that they are probably in a flexible form under these solution conditions.

The relative orientations of the N-terminal and C-terminal subdomains are determined by 57 long-range distance constraints derived from one or more NOEs between each of the residue pairs Tyr-13 and Arg-41, Cys-14 and Cys-33, Cys-14 and Arg-41, Leu-15 and Tyr-37, Leu-15 and Arg-41, Leu-15 and Cys-42, Leu-15 and Gln-43, Asn-16 and Tyr-37, Asn-16 and Cys-42, Asn-16 and Gln-43, Gly-17 and Val-34, Gly-18 and Cys-33, Gly-18 and Val-34, Thr-30 and Asp-40, Cys-31 and Asp-40, Cys-31 and Arg-41, and Cys-31 and Cys-42 (Figure 2). Since many of these NOEs are between side-chain protons, the relative orientations of the backbones of the two subdomains can vary by up to ±30° with coordinated changes in the conformations of the intervening side-chain dihedral angles. Although this uncertainty in the NMR data does not necessarily indicate motion between the domains, it is worth noting that the polypeptide loop Gly-39–Gln-43 at the interface between the two subdomains gives rise to proton resonances with broader line widths than the other regions of the protein. It is interesting that Arg-41, a residue that is important for receptor recognition (Engler et al., 1990) is within this polypeptide loop. Broadening of these resonances could be caused by exchange between multiple side-chain conformations at the subdomain interface. More detailed studies of nuclear relaxation might help to substantiate this preliminary interpretation of the proton resonance line broadening in this region of the molecule.

Comparison with Previously Published EGF Structures. The present structure of EGF provides a more detailed picture of the side-chain conformations (see, for example, Figure 8) than has been available from earlier work (Montelione et al., 1987; Kohda et al., 1988). Previous studies have focussed on the polypeptide chain fold. The chain fold of the present structure of mEGF is similar to the mEGF structure described by Kohda et al. (1988), which was based on mechanical model building using NMR constraints measured at pH 2. In particular, within the uncertainties described above, the relative orientation of the N- and C-terminal subdomains in the current structure is similar to those described previously (Montelione et al., 1987; Kohda et al., 1988). There is a 0.9-Å RMSD between the average backbone atom coordinates of residue 1–47 of the original 5 structures (Montelione et al., 1987) and the average backbone atom coordinates of the 16 structures described here. The present mEGF structure is also similar to that for human EGF (Cooke et al., 1987), which was based on distance geometry calculations and energy minimization.

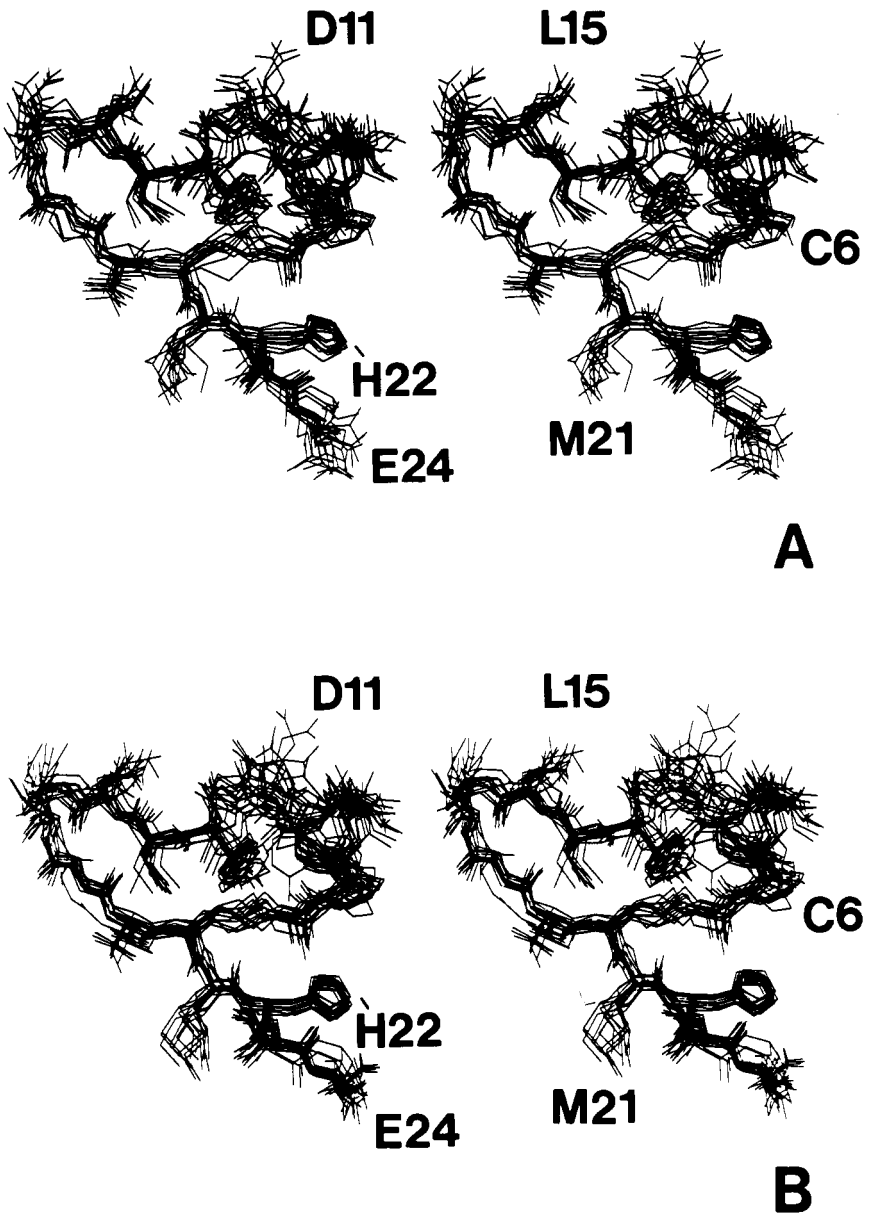

FIGURE 8: Superpositions for all heavy atoms of the polypeptide segment Cys-6–Glu-24 of mEGF. (A) 16 DISMAN conformers. (B) 16 energy-refined conformers.

However, in the absence of atomic coordinates, it is not possible to compare these structures in detail. In particular, from the limited information available in published stereodiagrams of these protein structures (Cooke et al., 1987; Kohda et al., 1988), we cannot evaluate whether or not there are small differences in the relative orientations of the two subdomains. The rat EGF structure (Mayo et al., 1989) also has the same basic molecular architecture as mEGF, including small antiparallel β-sheets in the N- and C-terminal subdomains, but a complete three-dimensional model of this protein has not yet been presented.

Comparison of the mEGF Structure with That of Human TGFα and Other Homologous Proteins. mEGF is one member of a large family of homologous growth factors and protein domains (Russell et al., 1984; Patthy, 1985; Campbell et al., 1989), including type α transforming, vaccinia virus, shopes virus, and myoxoma virus growth factors, amphiregulin, EGF-like domains of several blood-clotting proteins, and domains of the low density lipoprotein (LDL) receptor. The amino acid sequences of these proteins and domains exhibit strong (but not complete) conservation of all six cysteines and the two glycines at positions 36 and 39 of the mEGF sequence. The conservation of the relative positions of cystines suggests conservation of the mEGF chain fold in these EGF-like molecules. The selection for glycine at positions 36 and 39 further suggests that these residues may also have positive ϕ values (which are more energetically acceptable for Gly than for other amino acids) in the homologous proteins. This conclusion is supported by the recently described NMR structures of hTGFα (Brown et al., 1989; Campbell et al., 1989; Kohda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Kline et al., 1990) and the EGF-like domain from bovine coagulation factor X (Selander et al., 1990), both of which have the same overall chain fold as mEGF and hEGF. For hTGFα, the relative orientation of the N-terminal and C-terminal domains is also similar to that of mEGF (Kline et al., 1990). Comparisons of the solution structures of hTGFα and the EGF-like domain of bovine coagulation factor X with the previously published structure of mEGF (Montelione et al., 1987) have been described elsewhere (Kline et al., 1990; Selander et al., 1990).

As far as can be judged from the literature, all structure determinations of mEGF-like proteins presented so far are based on less complete NMR data sets than the one used in

the present determination of the complete solution structure of mEGF. The description of the mEGF structure in this paper, and the availability of the atomic coordinates in the Brookhaven Protein Data Bank, provide a basis for future detailed comparisons of the mEGF structure with those of homologous proteins and might thus lead to deeper insights into structure–function relations in this family of proteins.

ACKNOWLEDGMENTS

We gratefully acknowledge J. Bumby, G. Elkins, M. Olszyk, and R. Prasad for their technical contributions in the structure analysis of mEGF and F. J. Moy for valuable comments on the manuscript. Some of the computations were carried out on the Cornell National Supercomputer Facility, a resource of the Cornell Theory Center, which receives major funding from the National Science Foundation and the IBM Corp. with additional support from New York State and members of its Corporate Research Institute.

Registry No. EGF, 62229-50-9.

REFERENCES

Abe, H., Braun, W., Noguti, T., & Gō, N. (1984) *Comput. Chem. 8*, 239–247.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Bax, A., & Freeman, R. (1981) *J. Magn. Reson. 44*, 542–561.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Braun, W. (1987) *Q. Rev. Biophys. 19*, 115–157.

Braun, W., & Gō, N. (1985) *J. Mol. Biol. 186*, 611–626.

Brown, S. C., Mueller, L., & Jeffs, P. W. (1989) *Biochemistry 28*, 593–599.

Burgess, A. W. (1989) *Br. Med. Bull. 45*, 401–424.

Bystrov, V. F. (1976) *Prog. Nucl. Magn. Reson. Spectrosc. 10*, 41–81.

Campbell, I. D., Cooke, R. M., Baron, M., Harvey, T. S., & Tappin, M. J. (1989) *Prog. Growth Factor Res. 1*, 13–22.

Carver, J. A., Cooke, R. M., Esposito, G., Campbell, I. D., Gregory, H., & Sheard, B. (1986) *FEBS Lett. 205*, 77–81.

Cohen, S. (1962) *J. Biol. Chem. 237*, 1555–1562.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature (London) 327*, 339–341.

Engler, D. A., Montelione, G. T., & Niyogi, S. K. (1990) *FEBS Lett. 271*, 47–50.

Gay, D. M. (1983) *Assoc. Comput. Mach. Trans. Math. Software 9*, 503–524.

Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1985) *J. Am. Chem. Soc. 107*, 6394–6396.

Heldin, C. H., & Westermark, B. (1990) in *Growth Factors, Differentiation Factors, and Cytokines* (Habenicht, A., Ed.) pp 267–278, Springer Verlag, Berlin.

Hyberts, S. G., Märki, W., & Wagner, G. (1987) *Eur. J. Biochem. 164*, 625–635.

Kline, A. D., Braun, W., & Wüthrich, K. (1988) *J. Mol. Biol. 204*, 675–724.

Kline, T. P., Brown, F. K., Brown, S. C., Jeffs, P. W., Kopple, K. D., & Mueller, L. (1990) *Biochemistry 29*, 7805–7813.

Kohda, D., & Inagaki, F. (1988) *J. Biochem. (Tokyo) 103*, 554–571.

Kohda, D., Gō, N., Hayashi, K., & Inagaki, F. (1988) *J. Biochem. (Tokyo) 103*, 741–743.

Kohda, D., Shimada, I., Miyake, T., Fuwa, T., & Inagaki, F. (1989) *Biochemistry 28*, 953–958.

Mayo, K. H., Cavalli, R. C., Peters, A. R., Boelens, R., & Kaptein, R. (1989) *Biochem. J. 257*, 197–205.

Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361–2381.

Montelione, G. T., Arnold, E., Meinwald, Y. C., Stimson, E. R., Denton, J. B., Huang, S.-G., Clardy, J., & Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 7946–7958.

Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8594–8598.

Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5226–5230.

Montelione, G. T., Wüthrich, K., & Scheraga, H. A. (1988) *Biochemistry 27*, 2235–2243.

Montelione, G. T., Winkler, M. E., Burton, L. E., Rinderknecht, E., Sporn, M. B., & Wagner, G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1519–1523.

Moy, F. J., Scheraga, H. A., Liu, J.-F., Wu, R., & Montelione, G. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9836–9840.

Mueller, L. (1987) *J. Magn. Reson. 72*, 191–196.

Nagayama, K., & Wüthrich, K. (1981) *Eur. J. Biochem. 115*, 653–657.

Némethy, G., Pottle, M. S., & Scheraga, H. A. (1983) *J. Phys. Chem. 87*, 1883–1887.

Neuhaus, D., Wagner, G., Vasak, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257–273.

Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.

Patthy, L. (1985) *Cell 41*, 657–663.

Ponder, J. W., & Richards, F. M. (1987) *J. Mol. Biol. 193*, 775–791.

Roterman, I. K., Lambert, M. H., Gibson, K. D., & Scheraga, H. A. (1989) *J. Biomol. Struct. Dyn. 7*, 421–453.

Russell, D. W., Schneider, W. J., Yamamoto, T., Luskey, K. L., Brown, M. S., & Goldstein, J. L. (1984) *Cell 37*, 577–585.

Savage, C. R., Jr., Inagami, T., & Cohen, S. (1972) *J. Biol. Chem. 247*, 7612–7621.

Savage, C. R., Jr., Hash, J. H., & Cohen, S. (1973) *J. Biol. Chem. 248*, 7669–7672.

Selander, M., Persson, E., Stenflo, J., & Drakenberg, T. (1990) *Biochemistry 29*, 8111–8118.

Sippl, M. J., Némethy, G., & Scheraga, H. A. (1984) *J. Phys. Chem. 88*, 6231–6233.

Sporn, M. B., & Todaro, G. J. (1980) *N. Engl. J. Med. 303*, 878–880.

Sporn, M. B., & Roberts, A. B. (1985) *Nature (London) 313*, 745–747.

Tappin, M. J., Cooke, R. M., Fitton, J. E., & Campbell, I. D. (1989) *Eur. J. Biochem. 179*, 629–637.

Wagner, G., & Zuiderweg, E. R. P. (1983) *Biochem. Biophys. Res. Commun. 113*, 854–860.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.

Zuiderweg, E. R. P., Boelens, R., & Kaptain, R. (1985) *Biopolymers 24*, 601–611.