Sequence-Specific ^{1}H NMR Assignments and Identification of Slowly Exchanging Amide Protons in Murine Epidermal Growth Factor†

Gaetano T. Montelione,‡ Kurt Wüthrich,§ and Harold A. Scheraga*,‡

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301, and Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

Received August 18, 1987; Revised Manuscript Received October 30, 1987

ABSTRACT: Proton nuclear magnetic resonance (^{1}H NMR) assignments for the murine epidermal growth factor (mEGF) in aqueous solution were determined by using two-dimensional NMR at pH 3.1 and 28 °C. The assignments are complete for all backbone hydrogen atoms, with the exception of the N-terminal amino group, and for 46 of the 53 side chains. Among the additional seven amino acid residues, three have complete assignments for all but one side-chain proton, and between two and four protons are missing for the remaining four residues. The sequential assignments by nuclear Overhauser effect spectroscopy are consistent with the chemically determined amino acid sequence. The NMR data show that the conformations of both the Tyr3–Pro4 and Cys6–Pro7 peptide bonds are trans in the predominant solution structure. Proton–deuterium exchange rate constants were also measured for 13 slowly exchanging amide protons. The information presented here has been used elsewhere to determine the three-dimensional structure of mEGF in aqueous solution.

Epidermal growth factor (EGF)[1] is a small growth-promoting protein first isolated by Cohen (1962) from mouse submaxillary glands. Because of their role in cell growth regulation, there is much current interest in structure–function studies of EGF and EGF-like proteins. EGF stimulates growth and differentiation of various epidermal and epithelial tissues (Cohen & Elliott, 1963; Cohen, 1965; Savage & Cohen, 1973), and human EGF (urogastrone) may function physiologically to inhibit gastric acid secretion (Gregory, 1975). EGF and EGF-like proteins may also play a role in oncogenesis (Sporn & Todaro, 1980; Sporn & Roberts, 1985) and wound healing (Buckley et al., 1985). For certain fibroblast cell lines, EGF acts synergistically with type β transforming growth factor to induce phenotypic cellular transformation (Anzano et al., 1982; Roberts et al., 1982). EGF is therefore regarded as an α-type transforming growth factor (TGFα). Related TGFα's have been isolated from retrovirus-transformed fibroblasts and from human cancer cells (Todaro et al., 1980; Marquardt et al., 1983). The membrane-bound EGF receptor protein has also been purified (Downward et al., 1984) and cloned (Ullrich et al., 1984) and is homologous to both the v-*erb* B (Downward et al., 1984) and *neu* (Schechter et al., 1984; Semba et al., 1985; Coussens et al., 1985) oncogenes. Amplified expression of the *neu* oncogene or of the EGF receptor itself is observed in some human cancers (Ullrich et al., 1984; Semba et al., 1985; Slamon et al., 1987). These results suggest that EGF-like proteins or their receptors may be involved in the molecular basis of certain human cancers.

For murine EGF (mEGF), the sequence of 53 amino acids (Savage et al., 1972) has been determined by standard chemical techniques. The protein contains no lysines, phenylalanines, or alanines, and only one each of histidine, methionine, and glutamine. Biologically active synthetic mEGF has been prepared by solid-phase synthesis (Heath & Merrifield, 1986), corroborating the chemically determined sequence. Although no crystal structure is available, we have recently determined the polypeptide chain fold of mEGF using interproton distance constraints obtained from two-dimensional (2D) ^{1}H NMR spectroscopy (Montelione et al., 1986a, 1987). A three-dimensional structure of human EGF based on 2D NMR data is also available (Carver et al., 1986; Cooke et al., 1987).

In this paper, we describe how the sequence-specific resonance assignments were obtained. These resonance assignments provide the basis for determining the three-dimensional structure of mEGF in solution. They are compared with the previously reported resonance assignments for a few selected protons (Mayo, 1984, 1985; DeMarco et al., 1986; Mayo et al., 1986). In addition, the rates of proton–deuterium exchange for 13 backbone amide groups are reported; these were also used in the spatial structure determination (Montelione et al., 1986a, 1987).

† This work was supported by research grants from the National Institute of General Medical Sciences (GM-24893), 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 continued support of the NIH Resource for Multinuclear NMR and Data Processing at Syracuse University (RR-01317).

* Address correspondence to this author.
‡ Cornell University.
§ Eidgenössische Technische Hochschule-Hönggerberg.

[1] Abbreviations: DSS, 2,2-dimethyl-2-silapentane-5-sulfonate; EGF, epidermal growth factor; mEGF, murine EGF; TGFα, α-type transforming growth factor; 2D, two dimensional; MLEV-17, composite pulse sequence for spin-locking of transverse magnetization; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; pH*, pH meter reading without correction for isotope effects; RELAYED-COSY, 2D relayed coherence-transfer spectroscopy; TOCSY, 2D total correlation spectroscopy; 2Q-Spectra, 2D two-quantum spectroscopy; 2QF-COSY, 2D two-quantum-filtered correlation spectroscopy; HPLC, high-performance liquid chromatography. $d_{AB}(i,j)$ designates the distance between proton types A and B located in amino acid residues i and j, respectively, where N, α, and β denote the amide protons, αH, and βH, respectively. Examples used in the text include the sequential distances $d_{\alpha N} \equiv d_{\alpha N}(i,i+1)$, $d_{\beta N} \equiv d_{\beta N}(i,i+1)$, and $d_{NN} \equiv d_{NN}(i,i+1)$. The $d_{AB}(i,j)$ notation is also used as an adjective, e.g., $d_{\alpha N}$-type NOE, referring to the NOE associated with the $d_{\alpha N}(i,i+1)$ distance.

0006-2960/88/0427-2235$01.50/0

Table I: ^{1}H Chemical Shifts for Murine Epidermal Growth Factor at 28 °C and pH* 3.1

amino acid residue	chemical shifts (ppm)				amino acid residue	chemical shifts (ppm)			
	NH	αH	βH	others[b]		NH	αH	βH	others[b]
Asn-1		4.35	2.86, 2.86[a]	δNH$_2$ 7.58, 6.99	Val-34	8.41	4.05	2.07	γCH$_3$ 1.14, 0.99
Ser-2	8.58	4.62	3.72, 3.68		Ile-35	7.95	3.76	1.65	γCH$_2$ 1.36, 1.01
Tyr-3	8.34	4.90	3.08, 2.73	δH 7.14					γCH$_3$ 0.68
				εH 6.80					δCH$_3$ 0.80
Pro-4		4.67	2.27, 2.13	γCH$_2$ 2.07, 2.00	Gly-36	8.16	4.13, 3.19		
				δCH$_2$ 3.82, 3.63	Tyr-37	8.22	5.38	2.97, 2.97[a]	δH 6.88
Gly-5	8.07	4.24, 4.06							εH 6.65
Cys-6	8.68	4.51	2.98, 2.81		Ser-38	9.36	4.82	3.92, 3.92[a]	
Pro-7		4.58	2.43, 2.08	γCH$_2$ 1.93, 1.85	Gly-39	8.13	4.81, 3.81		
				δCH$_2$ 3.38, 2.72	Asp-40	9.22	4.37	3.05, 3.05[a]	
Ser-8	9.03	4.12	3.95, 3.95[a]		Arg-41	8.77	4.69	1.06	γCH$_2$ 2.32, 1.27
Ser-9	7.99	4.16	3.78, 3.78[a]						δCH$_2$ 2.91, 2.78
Tyr-10	8.11	4.18	3.28, 2.67	δH 6.44					εNH 7.54
				εH 5.97					ηNH$_2$ 6.75, 6.29
Asp-11	7.93	4.49	2.94, 2.94[a]		Cys-42	7.78	4.07	3.60, 3.15	
Gly-12	8.79	3.95, 3.95[a]			Gln-43	9.99	4.01	2.18, 1.77	γCH$_2$ 2.66, 2.55
Tyr-13	7.97	4.07	3.02, 3.02[a]	δH 6.99					εNH$_2$ 7.64, 6.89
				εH 6.76	Thr-44	8.85	4.43	3.85	γCH$_3$ 1.08
Cys-14	8.73	4.39	2.58, 2.26		Arg-45	8.74	4.15	1.63, 1.56	γCH$_2$ 1.21
Leu-15	8.31	4.28	1.61, 1.25	γH 1.64					δCH$_2$ 2.68
				δCH$_3$ 0.73					εNH 6.66
Asn-16	8.98	4.03	1.98, 1.37	δNH$_2$ 8.03, 7.21					ηNH$_2$ 6.32
Gly-17	8.71	4.02, 3.60			Asp-46	8.50	4.62	2.75, 2.48	
Gly-18	7.48	4.33, 3.44			Leu-47	8.21	4.18	1.59, 1.47	γH 1.59
Val-19	8.13	4.34	1.99	γCH$_3$ 0.99, 0.99					δCH$_3$ 0.88, 0.79
Cys-20	8.86	5.01	3.36, 3.13		Arg-48	8.19	3.82	1.43, 1.30	γCH$_2$ 1.38
Met-21	9.60	4.98	1.90, 1.90[a]	γCH$_2$ 2.45					δCH$_2$ 3.03
				εCH$_3$ 1.89					εNH 7.16
His-22	8.79	5.16	3.19, 2.91	δ^{2}H 6.54	Trp-49	7.58	4.21	3.22, 3.12	δH 7.01
				ε^{1}H 8.53					ε^{1}NH 10.07
Ile-23	8.38	4.08	1.79	γCH$_2$ 1.27, 0.93					ε^{3}H 7.09
				γCH$_3$ 0.81					ζ^{2}H 7.38
				δCH$_3$ 0.66					ζ^{3}H 6.72
Glu-24	8.50	3.67	1.98, 1.98[a]	γCH$_2$ 2.34					η^{2}H 7.09
Ser-25	8.51	4.13	3.88, 3.88[a]		Trp-50	7.14	4.25	3.05, 2.77	δH 7.06
Leu-26	6.92	4.41	1.55, 1.44	γH 1.52					ε^{1}NH 9.99
				δCH$_3$ 0.82, 0.80					ζ^{2}H 7.38
Asp-27	8.02	4.34	3.17, 2.69						ζ^{3}H 7.06
Ser-28	7.19	4.60	3.65, 3.52						η^{2}H 7.19
Tyr-29	8.43	5.24	2.48, 2.28	δH 6.82	Glu-51	7.63	4.20	1.97, 1.80	γCH$_2$ 2.19, 2.08
				εH 6.26	Leu-52	7.66	4.25	1.54, 1.54[a]	γH 1.56
Thr-30	9.01	5.02	4.11	γCH$_3$ 1.09					δCH$_3$ 0.82, 0.78
Cys-31	8.64	5.29	2.77, 2.59		Arg-53	7.88	4.21	1.81, 1.65	γCH$_2$ 1.48
Asn-32	9.41	5.01	2.98, 2.76	δNH$_2$ 7.23, 6.80					δCH$_2$ 2.95
Cys-33	8.88	4.75	3.28, 2.64						εNH 7.00

[a] The chemical shift degeneracy of these methylene protons was confirmed by 2Q-Spectra. [b] Protons that are not listed were not observed.

Materials and Methods

Protein Sample Preparation. Type α1 mEGF was purified from male mouse submaxillary glands (Burgess et al., 1982, 1983). Each preparation was characterized as >99% homogeneous by analytical reversed-phase (Waters) and Mono-Q (Pharmacia) anion-exchange HPLC using conditions described by Burgess et al. (1983). A typical amino acid composition has been published elsewhere (Montelione et al., 1986a). Samples were also analyzed by Mono-Q HPLC subsequent to completing the ^{1}H NMR measurements to ensure that chemical decomposition or deamidation during the data acquisition was negligible (<3%). The concentrations of the mEGF samples were determined spectrophotometrically by measuring the absorbance at 278 nm in aqueous 0.1% trifluoroacetic acid (ϵ_{278} = 19 000 cm^{-1} M^{-1}; unpublished data based on quantitative amino acid analysis). Samples for NMR spectroscopy were prepared in 5-mm tubes at 5–6 mM protein concentration and pH* 3.1 ± 0.1, where pH* refers to the direct pH meter reading without correction for isotope effects. No additional buffering reagents were used.

NMR Spectroscopy. All 2D NMR spectra were obtained on a Bruker WM-500 spectrometer with the proton probe thermostated to 28 ± 1 °C. The 2D NMR data sets were recorded with quadrature detection in both dimensions using time-proportional phase incrementation in the t_1-dimension (Redfield & Kunz, 1975; Marion & Wüthrich, 1983). Two-quantum-filtered correlation spectroscopy (2QF-COSY) (Piantini et al., 1982), nuclear Overhauser effect spectroscopy (NOESY) (Jeener et al., 1979; Anil Kumar et al., 1980), relayed coherence-transfer spectroscopy (RELAYED-COSY) (Eich et al., 1982; Wagner, 1983), two-quantum spectroscopy (2Q-Spectra (Bax et al., 1981; Wagner & Zuiderweg, 1983), and total correlation spectroscopy (TOCSY) (Braunschweiler & Ernst, 1983) were carried out as described elsewhere, except that the phase cycling for the NOESY data collection was adjusted to provide sine modulation in the t_1 dimension (Otting et al., 1986) and the transverse magnetization spin lock used in the TOCSY experiment was done through the decoupler amplifier using the MLEV-17 composite pulse sequence (Bax & Davis, 1985). The mixing times for the RELAYED-COSY and 2Q-Spectra were optimized for particular spectral features by plotting the appropriate coherence-transfer functions (Sørensen et al., 1983; Bax & Drobny, 1985). In order to allow accurate comparisons of cross-peak positions in 2QF-

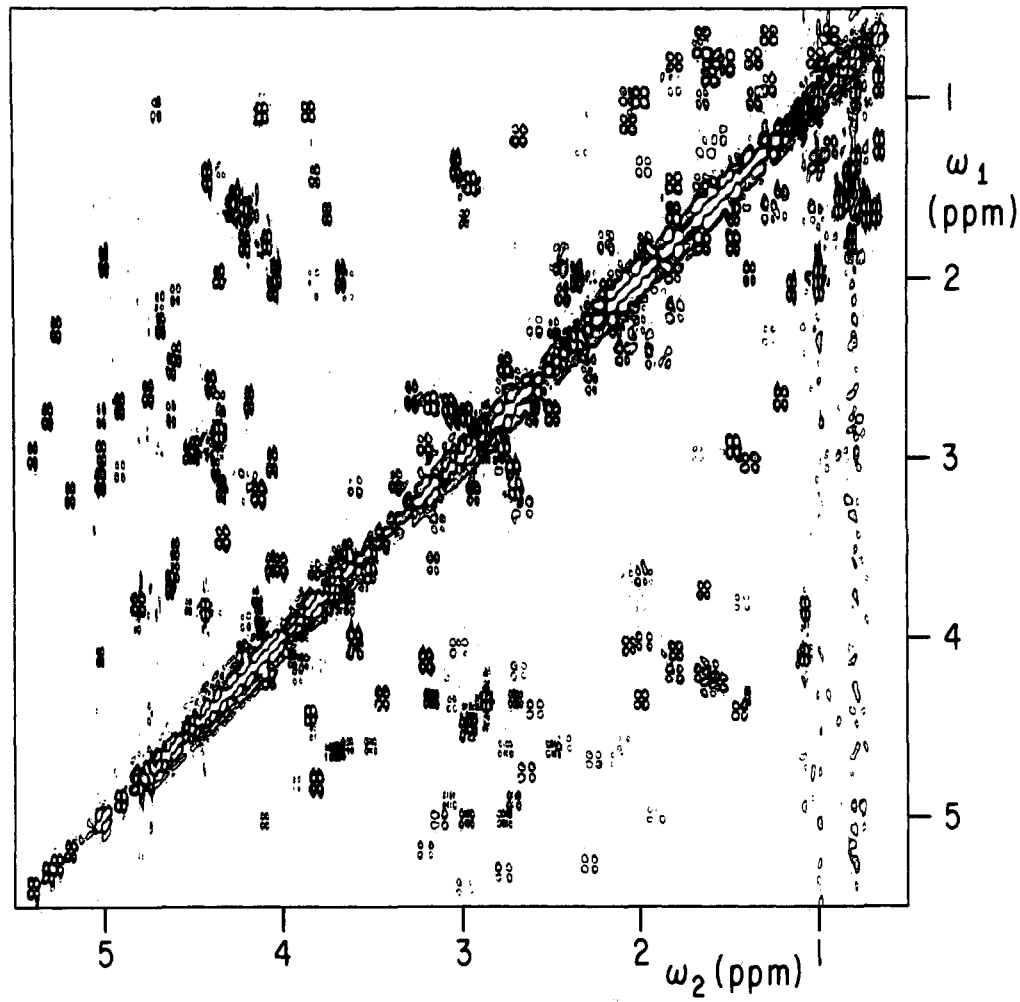

FIGURE 1: Contour plot of the aliphatic region (ω_1 = 0.5–5.5 ppm, ω_2 = 0.5–5.5 ppm) of a 500-MHz ^{1}H 2QF-COSY spectrum of 6 mM mEGF in 99.9% D_2O at pH* 3.1 and 28 °C. The sample was prepared by first exchanging the backbone amide protons for deuterium at pH* 3.1, 55 °C, for 5 min, followed by several cycles of lyophilization from D_2O. A total of 500 t_1 values were collected, each with 2048 points along t_2. Zero filling was used to extend this time domain to 2048 points in t_1 and 4096 points in t_2 prior to Fourier transformation. The resulting digital resolution is 7.36 Hz/point along ω_1 and 4.14 Hz/point along ω_2. Positive and negative peaks are plotted without distinction. The data collection for this 2QF-COSY spectrum was interleaved with data collection for a NOESY spectrum (Figure 6) of the same sample. Selected regions of these spectra are shown on an expanded scale in the documentation of spin system assignments which is available as supplementary figures (see paragraph at end of paper regarding supplementary material).

COSY and NOESY, the two data sets were acquired in an interleaved fashion (Neuhaus et al., 1985). For measurements on samples dissolved in H_2O, the solvent resonance was suppressed by selective saturation (Wider et al., 1983). The 2D NMR time domain data sets were multiplied by appropriately matched phase-shifted sine-bell window functions (DeMarco & Wüthrich, 1976) and Fourier-transformed as phase-sensitive spectra. Additional experimental details are described in the figure legends. Chemical shifts are reported in parts per million (ppm) relative to the methyl resonance of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).

RESULTS

Assignment Strategy. Sequence-specific resonance assignments for mEGF were determined by established procedures (Wüthrich et al., 1982; Wagner & Wüthrich, 1982a; Wüthrich, 1986). Briefly, this approach involves three key steps. First, the spin systems of the individual amino acid residues were identified by using 2QF-COSY, RELAYED-COSY, 2Q-Spectra, and TOCSY in D_2O and H_2O solutions of the protein. Second, segments of two or several neighboring spin systems in the sequence were identified, using NOE's between the amide proton of residue $i + 1$ and the α-proton, amide proton, or β-proton of the preceding residue i to establish sequential connectivities (Billeter et al., 1982). For sequential connections with prolines, the NOE's involving δCH_2 instead of the amide proton were used. Third, sequence-specific assignments were obtained by matching the sequences of spin systems in these sequentially assigned peptide segments with the corresponding segments in the chemically determined amino acid sequence (Savage et al., 1972), and the spin system identifications were completed for the residues with complex side chains (Wüthrich, 1983). The resulting sequence-specific resonance assignments for mEGF are listed in Table I.

Special care in completing these assignments was needed because of the presence of weak extra resonances in our mEGF spectra. It is not yet known if these minor resonances represent chemically distinct polypeptides which were not resolved in the HPLC purification or if they result from multiple molecular conformations of mEGF which interconvert slowly on the ^{1}H NMR time scale. The ratio of major to minor components measured in one-dimensional ^{1}H NMR spectra was ca. 8:1. These minor resonances did not arise from chemical decomposition of the mEGF sample during the NMR measurements, as both reversed-phase and ion-exchange analytical HPLC indicated insignificant chemical decomposition even in samples used for several different NMR measurements at

pH* 3.1. Even though the weak cross-peaks associated with these minor protein components presented a potential pitfall in determining the resonance assignments of the major mEGF species, a nearly complete set of resonance assignments which are entirely consistent with the chemically derived amino acid sequence (Savage et al., 1972) could be obtained. Sequence-specific resonance assignments could not be determined for the minor resonances of mEGF.

To illustrate the quality of the experimental data obtained with mEGF, six key 2D NMR spectra are shown in Figures 1–6. The following three sections provide a description of the identifications of the spin systems, the sequential assignments by NOESY, and the attachment of the peripheral protons in side chains with aromatic, amide, and guanidino groups (Billeter et al., 1982). Expansions of the relevant spectral regions, presented in Figures S1–S13 of the supplementary material (see paragraph at end of paper regarding supplementary material), provide a detailed documentation for the spin system identifications and sequential resonance assignments.

Spin System Identifications. Most of the spin systems were identified in the 2QF-COSY spectrum shown in Figure 1 which was obtained from a sample of mEGF dissolved in D_2O solution. The 2QF-COSY data were supplemented when necessary by RELAYED-COSY, 2Q-Spectra, and TOCSY data recorded under the same conditions of temperature and pH*.

The spin systems for five of the six glycines, both threonines, both valines, and both isoleucines in mEGF were uniquely identified in the 2QF-COSY spectrum. In addition, a weak glycine spin system associated with a minor protein species was also detected. The remaining sixth glycine spin system of the major EGF species could not be identified in the 2QF-COSY spectrum because its two protons are degenerate in chemical shift (Table I). This problem was overcome by analysis of a 2Q-spectrum recorded in H_2O solution. In this spectrum, remote glycine cross-peaks occur at the frequencies $\omega_1 = \delta_{\alpha H_1} + \delta_{\alpha H_2}$ and $\omega_2 = \delta_{NH}$, where δ_X denotes the chemical shift of proton X (Wagner & Zuiderweg, 1983) (Figure 2). Both the NH and αH chemical shifts of the sixth glycine spin system were determined from these data.

The remaining αH–βH cross-peaks in 2QF-COSY (Figure 1) were divided into two groups (Wüthrich, 1983, 1986), i.e., the AMX spin systems of the residues containing an αCH–βCH_2 fragment and the spin systems of *long* side chains in which there is additional scalar coupling to more peripheral protons. Some of these spin systems were distinguished by the characteristic fine structure of the αH–βH cross-peaks (Wüthrich, 1986; Widmer & Wüthrich, 1987). However, this approach was limited by the rather broad resonance line widths of 4–7 Hz, resulting in smeared fine structures and cancellations in cross-peaks with small active scalar couplings (Neuhaus et al., 1985). Therefore, the information available from 2QF-COSY was supplemented with data from TOCSY spectra, obtained on the same mEGF sample, which also contain αH-γH and αH–δH cross-peaks of the long side chains. With these combined data, all 27 AMX systems and 9 of the 10 spin systems of the long side chains in mEGF were identified. The αH–βH cross-peaks of the missing *long* spin system (subsequently assigned to Gln-43) were too weak to be identified in the 2QF-COSY and TOCSY spectra. However, these missing αH–βH cross-peaks were both identified in a 2Q-Spectrum recorded with a higher signal-to-noise ratio, and the corresponding βH–γH and γH–γH cross-peaks could then be identified in the 2QF-COSY spectrum. All of the assigned AMX spin systems have βH chemical shifts at lower field than 2.2 ppm, and all the long side chains are at higher field than 2.2 ppm, except for one AMX spin system (subsequently assigned to Asn-16), for which the βH chemical shifts are in the range expected for long side chains (Table I), and the two prolines, each of which has one βH shift at lower field than 2.2 ppm.

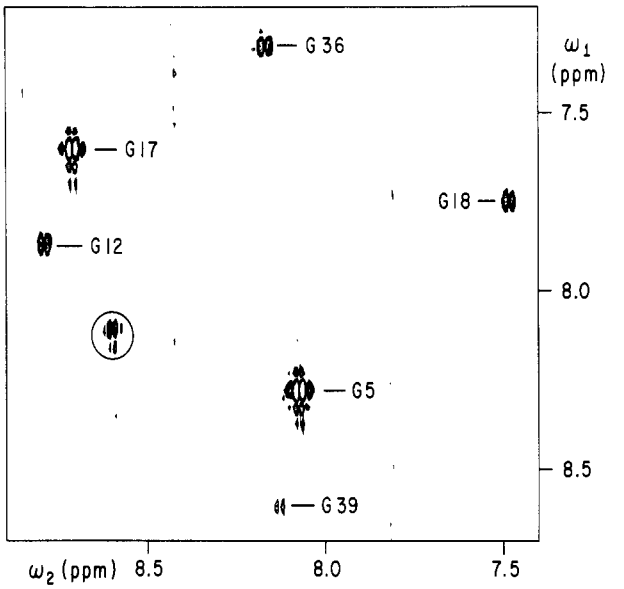

FIGURE 2: Contour plot of a 500-MHz ^1H 2Q-Spectrum of 6 mM mEGF in 85% H_2O/15% D_2O, pH* 3.1 and 28 °C. The mixing time was 25 ms. A total of 2048 t_1 values were collected, each with 2048 data points along t_2. Zero filling was used to extend this time domain to 4096 data points in t_1 and t_2 prior to Fourier transformation. The resulting digital resolution is 5.87 Hz/point along ω_1 and 3.94 Hz/point along ω_2. The region (ω_1 = 7.2–8.7 ppm, ω_2 = 7.4–8.9 ppm) contains the remote glycine αH–αH cross-peaks. Six remote glycine cross-peaks are labeled with their sequence-specific assignments. The circled cross-peak arises from a glycine residue in a minor protein species present in the mEGF sample. Its intensity is comparable to that of the other remote glycine cross-peaks because the efficiency of coherence transfer to these remote cross-peaks is strongly dependent on the conformation of the intervening backbone dihedral angle ϕ.

For some of the leucine, arginine, and proline spin systems, it was possible to make unique identifications of the amino acid type using supplementary information from TOCSY spectra recorded from samples in D_2O and from 2QF-COSY and RELAYED-COSY data obtained from samples in H_2O. Two of the four leucine spin systems, subsequently assigned to Leu-47 and Leu-52, were identified from TOCSY αH–δCH_3 cross-peaks. Since the two valine and two isoleucine spin systems were fully characterized from the 2QF-COSY data, the connection of these two αH resonances to methyl groups by TOCSY unambiguously identified them as belonging to leucine spin systems. For three of the four arginines, assigned later as Arg-45, Arg-48, and Arg-53, 2QF-COSY and RELAYED-COSY data obtained from H_2O solutions were used to trace the side-chain spin systems from αH to ϵNH. Since there are no lysines in mEGF, the presence of an ϵNH in these spin systems uniquely identified them as arginines. One of the two proline spin systems was completely identified from 2QF-COSY data alone, while for the second proline the βH–γH cross-peaks are too close to the diagonal to be identified reliably. This problem was overcome by using TOCSY data to characterize this second proline spin system fully; it was later assigned to Pro-4.

The characterization of the spin systems was completed by attaching the backbone amide protons, using 2QF-COSY, 2Q-Spectra, and RELAYED-COSY of samples in H_2O. For 26 spin systems, the backbone amide protons were uniquely

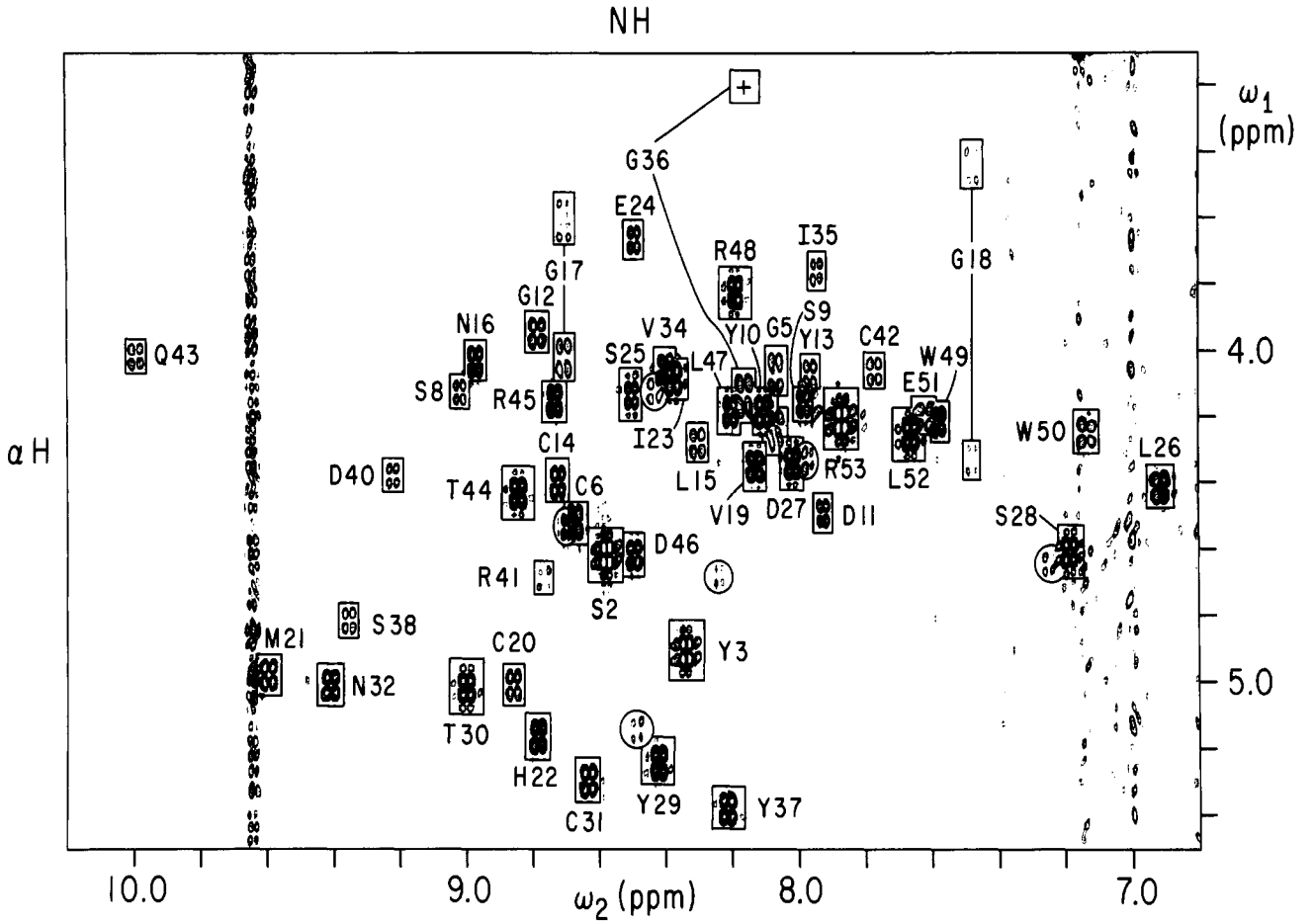

FIGURE 3: Contour plot of the region (ω_1 = 3.1–5.5 ppm, ω_2 = 6.8–10.2 ppm) of a 500-MHz ^{1}H 2QF-COSY spectrum of 6 mM mEGF in 85% H_2O/15% D_2O, pH* 3.1 and 28 °C. A total of 492 t_1 values were collected, each with 2048 data points along t_2. Zero filling was used to extend this time domain to 2048 points in t_1 and 4096 points in t_2 prior to Fourier transformation. The resulting digital resolution is 5.87 Hz/point along ω_1 and 3.54 Hz/point along ω_2. Positive and negative peaks are plotted without distinction. The NH–αH cross-peaks are enclosed in boxes and labeled with the amino acid type and the sequence position. The upfield Gly-36 NH–αH cross-peak (+) is very weak in this contour plot, but its identification is corroborated by the positions of the remote cross-peaks shown in Figure 2. Circled cross-peaks near the Tyr-3, Cys-6, Ile-23, Asp-27, Ser-28, and Tyr-29 NH–AH cross peaks arise from minor protein components present in the mEGF sample. The data collection for this 2QF-COSY spectrum was interleaved with the data collection for a NOESY spectrum (Figure 5) of the same sample.

identified from unambiguously assigned NH–αH cross-peaks of 2QF-COSY (Figure 3) or 2Q-Spectra. For the remaining residues, near-degeneracy of the αH chemical shifts precluded a unique amide proton connection from the 2QF-COSY and 2Q-Spectra data alone. These were therefore supplemented by identifying at least one NH–βH cross-peak in the RELAYED-COSY spectrum (Figure 4) for each pair of spin systems with nearly degenerate αH resonances. From these data, backbone amide protons were identified uniquely by scalar-coupling connections for all of the mEGF spin systems, except for the two previously identified prolines and one AMX spin system subsequently assigned to Asn-1.

Sequential Resonance Assignments. NOE cross-peaks identified with sequential distances $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ between amino acid residues which are adjacent in the amino acid sequence were identified by combining information from 2QF-COSY and NOESY spectra of mEGF dissolved in H_2O. Relevant spectral regions of this NOESY spectrum, recorded with a mixing time of 200 ms, are shown in Figure 5, and a survey of all observed sequential NOE's is presented in Figure 7. A rough calibration of cross-peak intensity versus interproton distances indicated an upper bound distance of ca. 4.0 Å for observable NOE's in this NOESY data set. This relatively high upper bound for NOE-observable distances obtained with the long mixing time used here makes it necessary to corroborate each $d_{\alpha N}$-type sequential NOE with at least one additional d_{NN}- or $d_{\beta N}$-type NOE for safely identifying each dipeptide connection (Billeter et al., 1982). This stringent criterion was used to establish each of the sequential connectivities described below. A survey of a similar NOESY data set recorded with a shorter mixing time (τ_m = 65 ms) is published elsewhere (Montelione et al., 1986a).

Sequential NOE's involving proline are useful both in establishing sequential resonance assignments and in distinguishing trans from cis X-Pro peptide bonds (Arseniev et al., 1983; Wüthrich et al., 1984; Montelione et al., 1986b; Wüthrich, 1986). For dipeptide sequences involving trans X–Pro peptide bonds, sequential connections were established by $d_{\alpha\delta}$- and $d_{\beta\delta}$-type NOE's. For mEGF, these sequential $d_{\alpha\delta}$-type NOE's are documented in Figure 6 for the Tyr3-Pro4 and Cys6-Pro7 sequences in a NOESY spectrum recorded with τ_m = 100 ms, and are included in the survey of Figure 7. There is no evidence for sequential X–Pro αH–αH NOE's in any of these NOESY spectra, showing that Tyr3-Pro4 and Cys6-Pro7 have *trans* peptide bonds in the major solution conformation of mEGF.

Beginning with the uniquely identified spin systems described in the previous section (i.e., 6 glycines, 2 threonines, 2 isoleucines, 2 leucines, 3 arginines, and 2 prolines), we identified sequence-specific connections from the NOESY data for the following 16 unique dipeptide pairs (Wüthrich, 1986): Pro4-Gly5, Pro7-AMX8, Gly17-Gly18, Gly18-Val19, Val19-

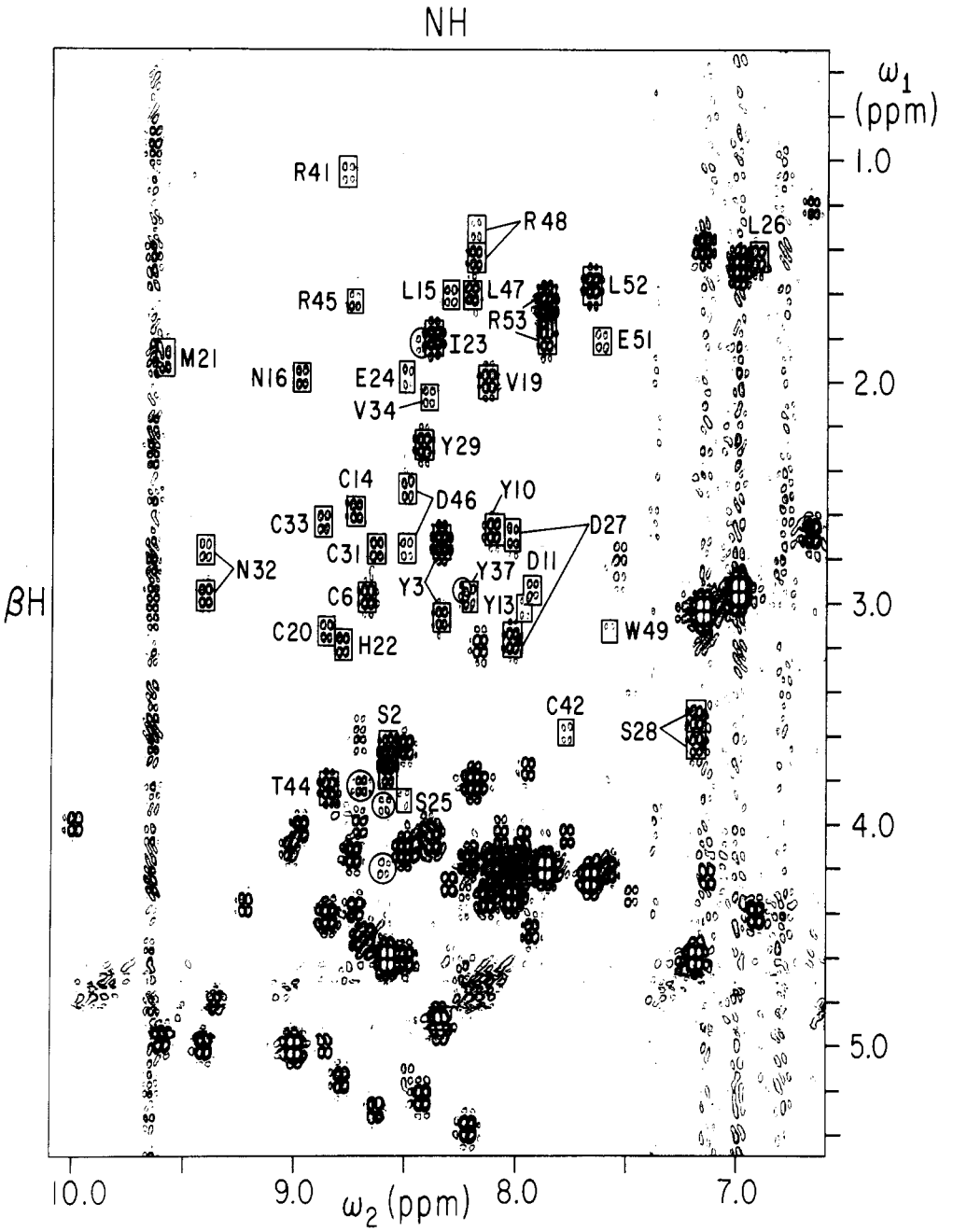

FIGURE 4: Contour plot of a 500-MHz ^{1}H RELAYED-COSY spectrum recorded with τ_m = 34 ms for the same mEGF sample as in Figure 3 at 28 °C. A total of 468 t_1 values were collected, each with 2048 data points along t_2. Zero filling was used to extend this time domain to 2048 points in t_1 and 4096 points in t_2 prior to Fourier transformation. The digital resolution is the same as in Figure 3. Positive and negative peaks are plotted without distinction. The region (ω_1 = 0.5–5.5 ppm, ω_2 = 6.6–10.1 ppm) contains both direct NH–αH and relayed NH–βH cross-peaks. The relayed peaks are framed and labeled with the amino acid type and the sequence position. The circled cross-peaks arise from minor protein components present in the mEGF sample.

AMX20, AMX22-Ile23, Ile23-*long*24, AMX29-Thr30, Thr30-AMX31, AMX33-Val34, Val34-Ile35, Ile35-Gly36, *long*43-Thr44, Thr44-Arg45, Leu47-Arg48, and Leu52-Arg53, where *long* stands for one of the long side chains other than Pro, the two identified Leu, and the three identified Arg. Then, from these starting points, the network of $d_{\alpha N}$-, $d_{\beta N}$-, and d_{NN}-type sequential NOE's was followed to complete the sequential resonance assignments. Of the 52 sequential residue pairs in mEGF, 44 are identified reliably, based on the criterion described above (Figure 7). For seven of the remaining eight dipeptide connections, at least one of the $d_{\alpha N}$-, $d_{\beta N}$-, or d_{NN}-type NOE's was identified. This sequential resonance assignment procedure identified the six AMX spin systems with lowest field βH chemical shifts as serines, which is consistent with chemical shifts in terminally blocked amino acids (Bundi & Wüthrich, 1979), and the AMX spin system lacking an assigned backbone amide proton as the N-terminal asparagine.

Connectivities to Peripheral Side-Chain Protons Using NOESY. The aromatic spin systems of five tyrosines, two tryptophans, and one histidine were separately identified in 2QF-COSY spectra (Figure S3). For the two tryptophans, NOESY was used to make the intraresidue connections between the five- and six-membered rings of the indole moiety. All eight aromatic side chains were connected to their associated AMX spin systems of αCH–βCH$_2$ by strong βH–δH NOE's (Billeter et al., 1982; Wüthrich, 1986). For the asparagine and glutamine spin systems, the side-chain amide protons identified in the NOESY spectrum of Figure 5 were attached to the corresponding β- and γ-methylene groups,

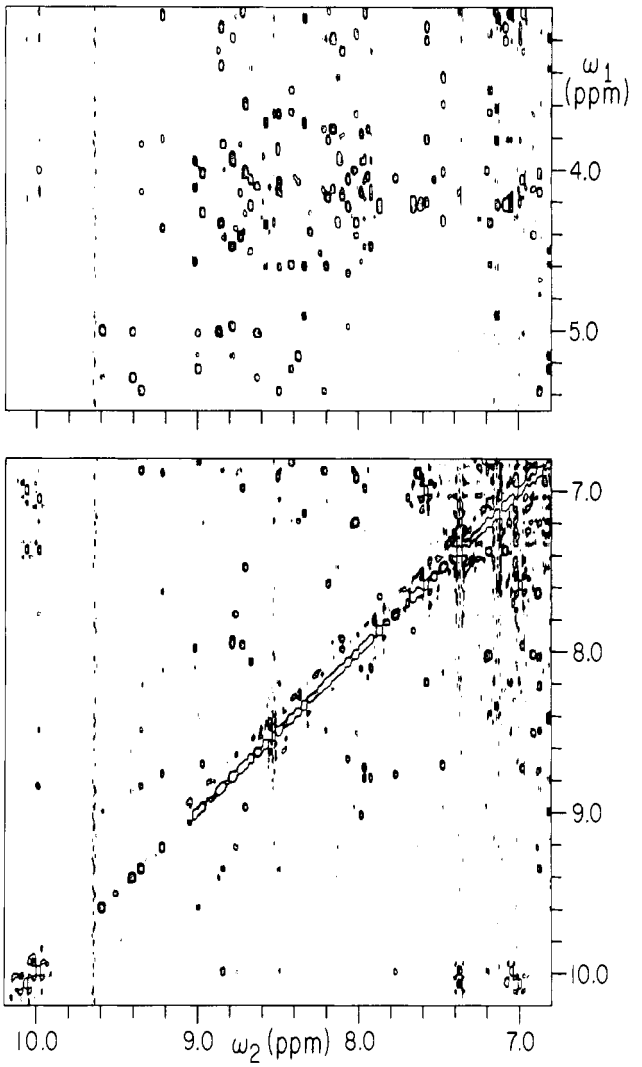

FIGURE 5: Contour plot of a 500-MHz ^{1}H NOESY spectrum recorded with a mixing time of 200 ms for the same mEGF sample as in Figure 2 at 28 °C. The data collection was interleaved with that of the 2QF-COSY spectrum of Figure 3. The digital resolution is 5.87 Hz/point along ω_1 and 3.54 Hz/point along ω_2. The upper panel shows the region (ω_1 = 3.0–5.5 ppm, ω_2 = 6.8–10.2 ppm) containing the intraresidue NH–αH and sequential αH–NH NOESY cross-peaks, and the lower panel shows the region (ω_1 = 6.8–10.2 ppm, ω_2 = 6.8–10.2 ppm) containing sequential NH–NH cross-peaks. Selected regions of this NOESY spectrum are shown in the documentation of sequential assignments available as supplementary figures (see paragraph at end of paper regarding supplementary material).

respectively, by strong δNH–βH and ϵNH–γH NOE's (Billeter et al., 1982). These data provide an independent check on the validity of the sequential resonance assignments described above in the polypeptide segments containing the five tyrosines, two tryptophans, three asparagines, and the single residues of glutamine and histidine. The ϵCH_3 singlet methyl resonance of the only methionine residue, Met-21, was identified in a one-dimensional ^{1}H NMR spectrum of mEGF. The remaining methyl resonances were attached to the Leu-15 and Leu-26 side chains using αH–δCH_3 and NH–δCH_3 NOE's. These assignments are umambiguous because Leu-15 and Leu-26 are far apart in the three-dimensional solution structure of mEGF (Montelione et al., 1987). Finally, the side-chain spin systems of Arg-41 and Arg-45 were completed by identifying intraresidue γH–δH, δH–ϵNH, and ϵNH–ηNH NOESY cross-peaks.

Measurement of Amide Proton Exchange Rates. Approximately 25% of the amide protons of mEGF have slow proton exchange rates which can be studied in ^{1}H NMR spectra of mEGF samples in D_2O (Mayo, 1985; Montelione et al., 1986a). For these protons, pseudo-first-order proton-deuterium exchange rate constants measured at pH* 3.1 and 28 °C over a period of 26 days are listed in Table II. Under these conditions, the most slowly exchanging amide proton of Val-19 has a half-life of about 38 days.

Discussion

The data presented in this paper and in the supplementary figures (see paragraph at end of paper regarding supplementary material) provide documentation for a nearly complete set of mEGF proton resonance assignments at pH 3 and 28 °C. Missing assignments include the N-terminal amino group as the only backbone protons. For several methylene groups in Met-21, Glu-24, Arg-41, Arg-45, Arg-48, and Arg-53, it could not be unambiguously established if the single observed resonance corresponds to one or both protons. For Trp-50, chemical shift degeneracies precluded a reliable assignment for the ϵ^3H resonance, and this degeneracy was not resolved by 2Q-Spectra. Finally, the guanidino protons of Arg-48 and Arg-53 and, as usual, all of the exchange-broadened carboxyl and hydroxyl protons were not observed. Overall, considering that they are in full agreement with the chemically determined amino acid sequence, the sequence-specific ^{1}H NMR assignments for mEGF are solidly ascertained by the data presented in this paper.

In the polypeptide segment Gly39-Gln43, many of the intraresidue 2D NMR cross-peaks are broader and weaker than those of residues in other parts of the protein. The Cys-42 βH and Gln-43 NH resonances also have unusually low-field chemical shifts (Table I). For Arg-41, the 2QF-COSY cross-peaks are so weak that a complete identification of the side-chain spin system could not be made until the final stage of the analysis when the corresponding NOESY data could be used reliably. Similarly, for Gln-43, neither of the two αH–βH cross-peaks was identified in the 2QF-COSY spectrum although both are observed as weak cross-peaks in 2Q-Spectra and NOESY. In the three-dimensional structure of mEGF (Montelione et al., 1987), the polypeptide segment Gly39-Gln43 forms a small loop structure, part of which is sandwiched between the N-terminal and C-terminal polypeptide domains. Inspection of this three-dimensional structure leads one to speculate that the line broadening in this loop might reflect local or interdomain dynamic processes.

It is now well established that the most slowly exchanging backbone amide protons in globular proteins are invariably involved in hydrogen bonds (Wagner & Wüthrich, 1982b; Wlodawer & Sjölin, 1982), although in rare cases the rate of amide proton exchange may also be retarded by other kinds of structural features (Englander & Kallenbach, 1984). In mEGF, 8 of the 13 slowest exchanging amide protons (Table II) including those of Val-19, Met-21, Thr-30, Asn-32, Val-34, Tyr-37, Ser-38, and Thr-44 participate in interstrand hydrogen bonds of antiparallel β-sheets, and those of Gly-18, Ile-35, Cys-42, and Gln-43 are in β-bends or in the Gly-39–Gln43 backbone loop of the solution structure (Montelione et al., 1986a, 1987).

Sequence-specific resonance assignments were previously proposed for some individual protons in mEGF (Mayo, 1984, 1985; DeMarco et al., 1986; Mayo et al., 1986), including assignments for five of the slowly exchanging amide protons (Mayo, 1985). Since these earlier assignments were obtained under different conditions of pH and temperature, a quantitative comparison of the chemical shifts is not warranted. Nonetheless, it is clear that the earlier assignments for the spin

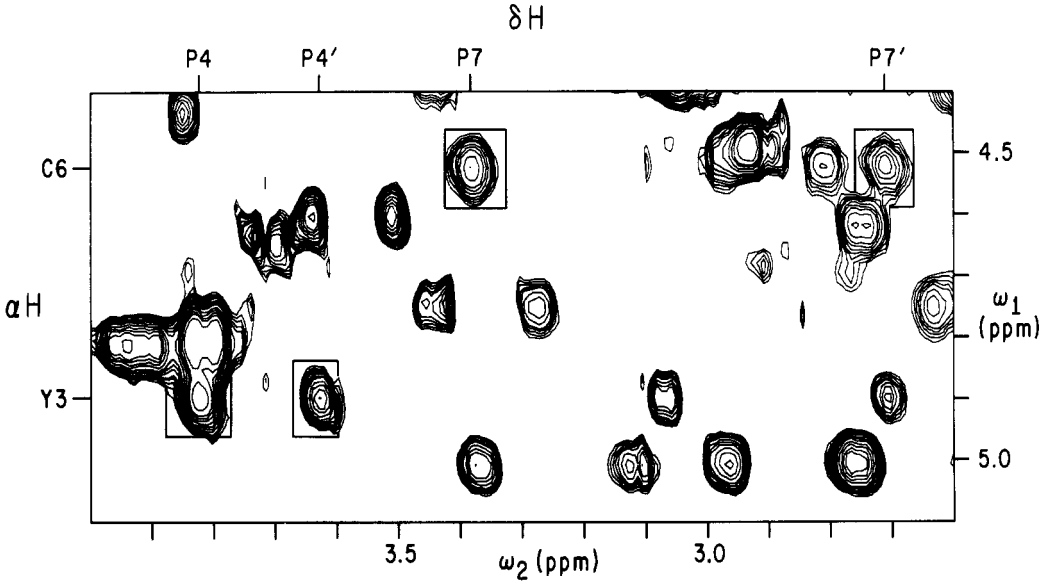

FIGURE 6: Contour plot of the region (ω_1 = 4.4–5.1 ppm, ω_2 = 2.6–4.0 ppm) of a 500-MHz ^{1}H NOESY spectrum of mEGF in D_2O solution recorded with τ_m = 100 ms for the same mEGF sample as in Figure 1. The data collection was interleaved with that of the 2QF-COSY spectrum of Figure 1, and the digital resolution is 7.36 Hz/point along ω_1 and 4.14 Hz/point along ω_2. The sequential $d_{\alpha\delta}$-type NOE's across X–Pro peptide bonds are framed, and sequence-specific assignments are indicated along the top and on the left.

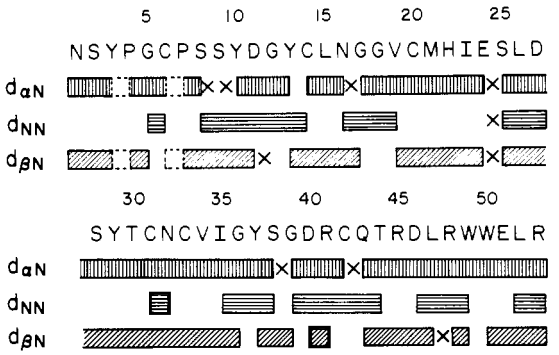

FIGURE 7: Amino acid sequence of mEGF and survey of the sequential connectivities used in establishing the sequence-specific ^{1}H NMR assignments. The vertically hatched blocks represent sequential $d_{\alpha N}$-type NOE connections, horizontally hatched blocks represent sequential d_{NN}-type connections, and diagonally hatched blocks represent sequential $d_{\beta N}$-type connections. Empty blocks drawn with dashed lines represent sequential $d_{\alpha\delta}$- and $d_{\beta\delta}$-type connections involving prolines. X denotes locations where the connections could not be identified reliably because of chemical shift degeneracy or other technical reasons. All data indicated were obtained from the NOESY spectrum described in Figure 5, except that the $d_{\alpha N}$(33,34), $d_{\beta N}$(33,34), and $d_{\alpha N}$(41,42) connections were identified in a NOESY spectrum of a freshly prepared mEGF solution in 99.9% D_2O, pH* 3.1, and 28 °C, and the $d_{\alpha\delta}$- and $d_{\beta\delta}$-type connections involving proline were identified in the NOESY spectrum shown in Figure 6.

systems of Tyr-37 and Ser-38 (Mayo, 1984, 1985) are confirmed by Table I, whereas the previously reported assignments for the aromatic protons of Tyr-13 and Tyr-29 (Mayo, 1984, 1985; DeMarco et al., 1986) must be interchanged, and the spin systems previously assigned to Gly-36, Leu-47, and Arg-48 (Mayo, 1984) appear to correspond to spin systems assigned in Table I to Val-19, Val-34, and Ile-35, respectively.

The sequential resonance assignments presented in this paper provide the basis for determining the three-dimensional structure of mEGF using 2D NMR spectroscopy. A description of the polypeptide chain fold is described elsewhere (Montelione et al., 1987). Further refinement of this structure, with incorporation of conformational energy minimization (Momany et al., 1975; Némethy et al., 1983), is now in progress.

Table II: Approximate Pseudo-First-Order Exchange Rates (k_{NH}) of 13 Slowly Exchanging Amide Protons Observed in D_2O Solution of Murine Epidermal Growth Factor at 25 °C and pH* 3.1

residue[a]	k_{NH}[b] (×10^{-6} min^{-1})	residue[a]	k_{NH}[b] (×10^{-6} min^{-1})
Leu-15	20	Ile-35	>700
Gly-18	330	Tyr-37	40
Val-19	10	Ser-38[c]	(80)[c]
Met-21	80	Cys-42	30
Thr-30	140	Gln-43	50
Asn-32[c]	(80)[c]	Thr-44	90
Val-34	20		

[a] The chemical shifts are listed in Table I. [b] Amide proton exchange was studied at 300 MHz over a period of 26 days at 25 ± 1 °C. The samples contained 0.3 mM mEGF, 150 mM NaCl, and 1 mM NaN_3 in D_2O at pH* 3.1. [c] This pair of NH resonances are partially overlapped in the 300-MHz 1D NMR spectrum.

ACKNOWLEDGMENTS

We are indebted to Dr. A. W. Burgess and Dr. E. Nice for providing the samples of mEGF. We thank Dr. W. Chazin, Dr. A. D. Kline, and Dr. G. Wagner for many helpful discussions.

SUPPLEMENTARY MATERIAL AVAILABLE

Thirteen figures documenting spin system assignments (46 pages). Ordering information is given on any current masthead page.

Registry No. EGF, 62229-50-9; H_2, 1333-74-0.

REFERENCES

Anil Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Anzano, M. A., Roberts, A. B., Meyers, C. A., Komoriya, A., Lamb, L. C., Smith, J. M., & Sporn, M. B. (1982) *Cancer Res. 42*, 4776–4778.

Arsen'ev, A. S., Kondakov, V. I., Maiorov, V. N., Volkova, T. M., Grishin, E. V., Bystrov, V. F., & Ovchinnikov, Yu. A. (1983) *Bioorg. Khim. 9*, 768–793.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Bax, A., & Drobny, G. (1985) *J. Magn. Reson. 61*, 306–320.
Bax, A., Freeman, R., Frenkiel, T. A., & Levitt, M. H. (1981) *J. Magn. Reson. 43*, 478–483.
Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Buckley, A., Davidson, J. M., Kamerath, C. D., Wolt, T. B., & Woodward, S. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7340–7344.
Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–297.
Burgess, A. W., Knesel, J., Sparrow, L. G., Nicola, N. A., & Nice, E. C. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5753–5757.
Burgess, A. W., Lloyd, C. J., & Nice, E. C. (1983) *EMBO J. 2*, 2065–2069.
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.
Cohen, S. (1965) *Dev. Biol. 12*, 394–407.
Cohen, S., & Elliott, G. A. (1963) *J. Invest. Dermatol. 40*, 1–6.
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.
Coussens, L., Yang-Feng, T. L., Liao, Y.-C., Chen, E., Gray, A., McGrath, J., Seeburg, P. H., Libermann, T. A., Schlessinger, J., Francke, U., Levinson, A., & Ullrich, A. (1985) *Science (Washington, D.C.) 230*, 1132–1139.
DeMarco, A., & Wüthrich, K. (1976) *J. Magn. Reson. 24*, 201–204.
DeMarco, A., Mayo, K. H., Bartolotti, F., Scalia, S., Menegatti, E., & Kaptein, R. (1986) *J. Biol. Chem. 261*, 13510–13516.
Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J., & Waterfield, M. D. (1984) *Nature (London) 307*, 521–527.
Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3731–3732.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521–655.
Gregory, H. (1975) *Nature (London) 257*, 325–327.
Heath, W. F., & Merrifield, R. B. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6367–6371.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Marquardt, H., Hunkapiller, M. W., Hood, L. E., Twardzik, D. R., DeLarco, J. E., Stephenson, J. R., & Todaro, G. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4684–4688.
Mayo, K. H. (1984) *Biochemistry 23*, 3960–3973.
Mayo, K. H. (1985) *Biochemistry 24*, 3783–3794.
Mayo, K. H., DeMarco, A., & Kaptein, R. (1986) *Biochim. Biophys. Acta 874*, 181–186.
Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361–2381.
Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8594–8598.
Montelione, G. T., Hughes, P., Clardy, J., & Scheraga, H. A. (1986b) *J. Am. Chem. Soc. 108*, 6765–6773.
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.
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.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.
Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M., Frolik, C. A., Marquardt, H., Todaro, G. J., & Sporn, M. B. (1982) *Nature (London) 295*, 417–419.
Savage, C. R., Jr., & Cohen, S. (1973) *Exp. Eye Res. 15*, 361–366.
Savage, C. R., Jr., Inagami, T., & Cohen, S. (1972) *J. Biol. Chem. 247*, 7612–7621.
Schechter, A. L., Stern, D. F., Vaidyanathan, L., Decker, S. J., Drebin, J. A., Greene, M. I., & Weinberg, R. A. (1984) *Nature (London) 312*, 513–516.
Semba, K., Kamata, N., Toyoshima, K., & Yamamoto, T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6497–6501.
Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., & McGuire, W. L. (1987) *Science (Washington, D.C.) 235*, 177–182.
Sørensen, O. W., Levitt, M. H., & Ernst, R. R. (1983) *J. Magn. Reson. 55*, 104–113.
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.
Todaro, G. J., Fryling, C., & DeLarco, J. E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5258–5262.
Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D., & Seeburg, P. H. (1984) *Nature (London) 309*, 418–425.
Wagner, G. (1983) *J. Magn. Reson. 55*, 151–156.
Wagner, G., & Wüthrich, K. (1982a) *J. Mol. Biol. 155*, 347–366.
Wagner, G., & Wüthrich, K. (1982b) *J. Mol. Biol. 160*, 343–361.
Wagner, G., & Zuiderweg, E. R. P. (1983) *Biochem. Biophys. Res. Commun. 113*, 854–860.
Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316–336.
Wlodawer, A., & Sjölin, L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1418–1422.
Wüthrich, K. (1983) *Biopolymers 22*, 131–138.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.