Journal of Biomolecular NMR, 5 (1995) 161–172
ESCOM

J-Bio NMR 230

Combined use of ^{13}C chemical shift and $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ heteronuclear NOE data in monitoring a protein NMR structure refinement

Bernardo Celda[a,b], Clelia Biamonti[a,c], Maria Jose Arnau[b], Roberto Tejero[a,b] and Gaetano T. Montelione[a,c,d,*]

[a]*Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, U.S.A.*
[b]*Departamento de Quimica Fisica, F. Quimicas, Universidad de Valencia, cl Dr. Moliner 50, E-46100 Burjassot (Valencia), Spain*
[c]*Graduate Program in Biochemistry and Molecular Biology, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854, U.S.A.*
[d]*Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, U.S.A.*

Received 5 July 1994
Accepted 9 September 1994

Keywords: Epidermal growth factor; Nuclear relaxation; Molecular dynamics; Pulsed-field gradients

Summary

A large portion of the ^{13}C resonance assignments for murine epidermal growth factor (mEGF) at pH 3.1 and 28 °C has been determined at natural isotope abundance. Sequence-specific ^{13}C assignments are reported for 100% of the assignable C^{α}, 96% of the C^{β}, 86% of the aromatic and 70% of the remaining peripheral aliphatic resonances of mEGF. A good correlation was observed between experimental and back-calculated C^{α} chemical shifts for regions of regular β-sheet structure. These assignments also provide the basis for interpreting $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ heteronuclear NOE (HNOE) values in mEGF at natural isotope abundance. Some of the backbone polypeptide segments with high internal mobility, indicated by these $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ HNOE measurements, correlate with locations of residues involved in the putative mEGF-receptor binding site. Using four families of mEGF structures obtained over the last few years, we demonstrate that standard deviations between experimental and back-calculated $\Delta\delta C^{\alpha}$ values can be used to monitor the refinement of this protein's structure, particularly for β-sheet regions. Improved agreement between calculated and observed values of $\Delta\delta C^{\alpha}$ is correlated with other measures of structure quality, including lowered values of residual constraint violations and more negative values of conformational energy. These results support the view that experimental conformation-dependent chemical shifts, $\Delta\delta C^{\alpha}$, can provide a reliable source of information for monitoring the process of protein structure refinement and are potentially useful restraints for driving the refinement.

Introduction

Murine epidermal growth factor (mEGF) is a small protein of 53 amino acids. EGF and EGF-like proteins play important roles in wound healing (Burgess, 1989) and in oncogenesis (Burgess, 1989; Guerin et al., 1989). The EGF-like sequence module is an important structural motif, occurring in over 300 known protein sequences (Campbell and Bork, 1993). Although an X-ray crystal structure is not yet available for mEGF, its three-dimensional structure has been determined by solution NMR spectroscopy (Montelione et al., 1987,1992; Kohda and Inagaki, 1988,1992a,b). Similar three-dimensional chain folds (see Fig. 1) have also been determined by solution NMR for the homologous human EGF (hEGF) (Cooke et al., 1987; Hommel et al., 1992) and human type-α transforming growth factor (hTGFα) (Brown et al., 1989; Kohda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Kline et al., 1990; Harvey et al., 1991; Moy et al., 1993) proteins, and for the EGF-like domains of human coagulation factor IX (Baron et al., 1992), bovine coagulation factor X (Selander et al., 1990; Ullner et al., 1992) and the urokinase-type plasminogen activator (Hansen et al., 1994). X-ray crystal structures have also been deter-

*To whom correspondence should be addressed.

Abbreviations: HSQC, heteronuclear single-quantum coherence spectroscopy; PFG, pulsed-field gradient; TOCSY, ^{1}H-^{1}H total correlation spectroscopy; EGF, epidermal growth factor; mEGF, murine EGF; hEGF, human EGF; hTGFα, human type-α transforming growth factor; DIPSI, spin-locking pulse sequence; NOE, nuclear Overhauser effect; HNOE, heteronuclear Overhauser effect.

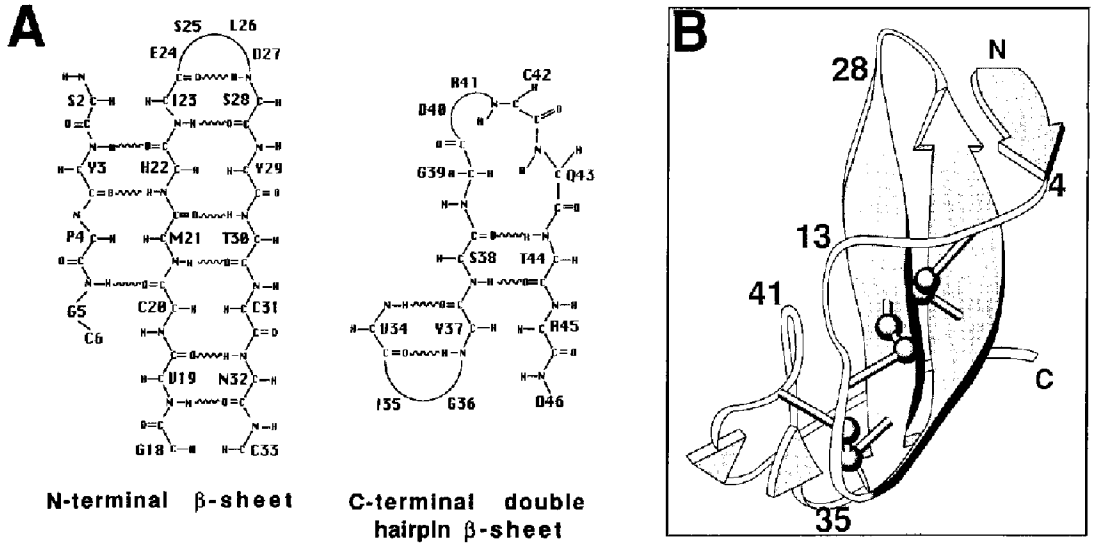

Fig. 1. (A) Regular β-sheet backbone structure in murine EGF (Montelione et al., 1987,1992). (B) Overall chain fold of EGF. The ribbon arrows represent regions of the peptide backbone which constitute the β-sheet backbone structures.

mined recently for the homologous EGF-like domains of human E-selectin (Graves et al., 1994) and human coagulation factor Xa (Padmanabhan et al., 1993).

While the general structural features of EGF and EGF-like domains are reasonably well understood, relatively little information is available about internal motions and molecular dynamics for this important family of polypeptide chain folds. Some dynamic information, in the form of $^{1}H/^{2}H$ amide exchange rates and ^{1}H linewidths, is available for mEGF (Montelione et al., 1986, 1987,1988,1992), hEGF (Cooke et al., 1987; Hommel et al., 1992) and hTGFα (Brown et al., 1989; Kohda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Harvey et al., 1991; Moy et al., 1993). Although detailed studies of ^{15}N nuclear relaxation have been published recently for hTGFα (Li, 1994; Li and Montelione, 1995), no heteronuclear relaxation studies have been reported to date for other EGF-like molecules, including mEGF.

^{13}C NMR can provide important structural and dynamic information on proteins and nucleic acids (Wüthrich, 1976; Wagner, 1989). Carbon resonance frequencies are sensitive to bond geometries, orbital hybridizations and local chemical environments (Wishart et al., 1991; de Dios et al., 1993a,b; Laws et al., 1993), while ^{13}C nuclear relaxation rates provide information about overall tumbling and internal motions (Allerhand et al., 1971; Nirmala and Wagner, 1988; Wagner, 1989,1993; Palmer et al., 1991, 1993). ^{13}C NMR parameters also can provide information that is useful for protein structure determination and refinement (Spera and Bax, 1991; Wishart et al., 1991; de Dios et al., 1993a,b; Laws et al., 1993). For example, conformation-dependent chemical shifts of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei (i.e., $\Delta\delta C^{\alpha}$ and $\Delta\delta C^{\beta}$, respectively) can be used to identify regular α-helical and β-sheet conformations of proteins and peptides (Spera and Bax, 1991; Wishart et al., 1991,1992; Lee et al., 1992; Reily et al., 1992; Thanabal et al., 1994; Wishart and Sykes, 1994). Such secondary structure analyses are especially useful in the initial stages of a protein structure determination by NMR (see, for example, Metzler et al., 1993; Anglister et al., 1994; Garrett et al., 1994). ^{13}C chemical shifts can also be used to characterize dynamic equilibria in partially folded polypeptides (Reily et al., 1992). It may also be possible to use comparisons between calculated and observed ^{13}C chemical shift values as a driving force for structure refinement (de Dios et al., 1993a,b; Laws et al., 1993), or as an independent measure of the accuracy of a protein structure determined by NMR or X-ray crystallography.

In this paper we demonstrate that conformation-dependent ^{13}C chemical shift data, measured at natural isotope abundance, can be used to monitor an NMR structure refinement. An extensive set of sequence-specific ^{13}C resonance assignments is presented for mEGF and used to interpret backbone $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ heteronuclear Overhauser effects (HNOEs). Some of the backbone polypeptide segments with high internal mobility, indicated by these $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ HNOE measurements, correlate with locations of residues involved in the putative EGF-receptor binding site. A good correlation is observed between negative values of the conformation-dependent $^{13}C^{\alpha}$ chemical shifts and the locations of extended β-sheet backbone structures in mEGF, corroborating existing evidence (Spera and Bax, 1991; Wishart et al., 1991) that $^{13}C^{\alpha}$ chemical shifts provide useful information for identifying regular β-sheet backbone conformations in proteins. Using ensembles of mEGF structures obtained over the last few years, we demonstrate that standard deviations between experimental and empirically estimated $\Delta\delta C^{\alpha}$ values can also be used to monitor the refinement of the structure of this protein, particularly for β-sheet regions.

Improved agreement between calculated and observed values of $\Delta\delta C^{\alpha}$ is correlated with other measures of structure quality, including lowered values of residual NOE and dihedral angle constraint violations, and conformational energies. These results demonstrate that experimental conformation-dependent chemical shifts, $\Delta\delta C^{\alpha}$, can provide a reliable source of information for monitoring the process of protein structure refinement and are potentially useful restraints for driving the refinement.

Materials and Methods

The preparation and chemical analyses of natural mEGF from mouse submaxillary gland have been reported elsewhere (Montelione et al., 1988). Protein NMR samples were prepared in 99.9% D_2O solution at 5 mM concentration, pH* = 3.1 ± 0.1 (the uncorrected pH meter readings of D_2O solutions are reported as pH*). The sample was not isotopically enriched.

All NMR spectra were recorded on a Varian Unity 500 NMR spectrometer at 28 °C. ^{13}C chemical shifts were determined from the published sequence-specific proton resonance assignments (Montelione et al., 1988) by combined analysis of proton-detected HSQC (Bodenhausen and Ruben, 1980) and HSQC-TOCSY (Moy et al., 1993) 2D NMR spectra. HSQC spectra were recorded with the carbon carrier frequency centered in the C^{β} and aromatic carbon regions. Each of these spectra required ~24 h of NMR measuring time. Two HSQC-TOCSY spectra, each requiring about ~60 h, were recorded. In the first HSQC-TOCSY data set the carbon carrier frequency was placed in the center of the α-carbon resonances and in a second set it was moved to the center of the β-carbon resonances. In both of these data sets isotropic mixing was performed with the DIPSI (Shaka et al., 1988) multipulse sequence, using a range of mixing times between 9 and 32 ms, which were then co-added in the frequency domain as described elsewhere (Celda and Montelione, 1993). The recycle times were 1.83 s.

^{1}H-^{13}C HNOEs were determined using 2D PFG-HNOE (Li and Montelione, 1994). Steady-state HNOEs on ^{1}H-detected ^{13}C resonance intensities were measured with and without 4.0 s of aliphatic proton saturation, sufficiently long to ensure >95% steady-state NOE for all $^{13}C^{\alpha}$ nuclei with longitudinal relaxation times $T_1 < 800$ ms. Although ^{13}C T_1 relaxation rates have not been measured for mEGF, values in proteins of similar size are typically < 500 ms (Nirmala and Wagner, 1988; Wagner, 1993). The total recycle times were 5.2 s. Each HNOE spectrum, with and without aliphatic proton presaturation, was recorded in duplicate, requiring a total of ~40 h of NMR instrument time.

Conformational energy calculations were carried out on a Silicon Graphics Indigo 2 workstation using the CHARMm energy potential function of the CONGEN program (Brooks et al., 1983; Bruccoleri and Karplus, 1987). Details of these calculations and their application in the refinement of the solution structures of mEGF and hTGFα will be presented elsewhere (Tejero, R., Bassolino-Klimas, D., Bruccoleri, R.E. and Montelione, G.T., manuscript in preparation).

Results

Sequence-specific ^{13}C resonance assignments

An expansion of the aliphatic C^{α}-H^{α} region from an HSQC spectrum of mEGF is shown in Fig. 2. Using the published proton chemical shifts (Montelione et al., 1988), most of the C^{α} and C^{β} resonances could be assigned unambiguously. However, proton chemical shift degeneracies precluded the unique determination of several ^{13}C assignments, including the C^{α} resonances of the following groups of amino acids: Asn[1], Val[19], Asp[27] and Asp[40]; Pro[4] and Arg[41]; Pro[7], Ser[28] and Asp[46]; Ser[8] and Ser[25]; Ser[9], Tyr[10], Arg[45], Glu[51] and Arg[53]; and Cys[20], Thr[30] and Asn[32] (see Fig. 2). Most of these degeneracies could be overcome using 2D HSQC-TOCSY. By recording a set of HSQC-TOCSY spectra with isotropic mixing times ranging between 9–32 ms, many two-bond heteronuclear connectivities could be established (e.g. C^{α}-H^{β}, C^{β}-H^{α}, C^{β}-H^{γ}, etc.). In these spectra, the symmetry-related two-bond C^{i}-H^{j} and C^{j}-H^{i} (e.g. C^{α}-H^{β} and C^{β}-H^{α}) HSQC-TOCSY cross peaks provide complementary information. Ambiguities in interpreting the HSQC spectrum due to degeneracies of H^{i} protons can generally be resolved using C^{i}-H^{j} cross peaks, while degeneracies of C^{i} carbons are resolved using C^{j}-H^{i} cross peaks.

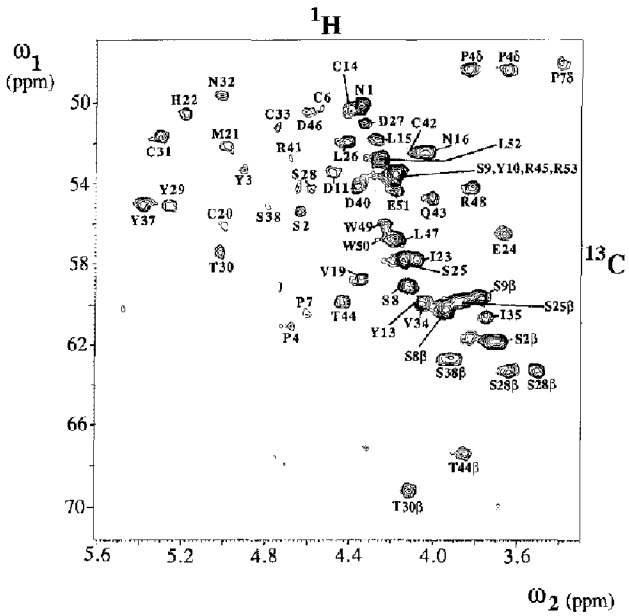

Fig. 2. Aliphatic C^{α}-H^{α} region (excluding glycine residues) of an HSQC spectrum of 5 mM mEGF in 99.9% D_2O at pH* 3.1 and 28 °C. Sequence-specific resonance assignments of these heteronuclear correlation peaks are designated with the one-letter code for the amino acids and residue number.

TABLE 1
SEQUENCE-SPECIFIC RESONANCE ASSIGNMENTS FOR MURINE EPIDERMAL GROWTH FACTOR[a]

Residue	C^{α}	H^{α}	C^{β}	H^{β}	Others
Asn[1]	52.6	4.34	37.9	2.86, 2.86[b]	
Ser[2]	57.7	4.63	64.3	3.71, 3.68	
Tyr[3]	55.6	4.90	39.5	3.08, 2.72	H^{δ} 7.14; H^{ε} 6.80; C^{δ} 133.6; C^{ε} 118.1
Pro[4]	63.6	4.68	32.3	2.27, 2.13	H^{γ} 2.07, 2.00; H^{δ} 3.82, 3.63; C^{γ} 17.3; C^{δ} 50.9
Gly[5]	45.0	4.23, 4.04			
Cys[6]	53.0	4.52	42.4	2.98, 2.81	
Pro[7]	63.0	4.59	32.7	2.43, 2.08	H^{γ} 1.93, 1.85; H^{δ} 3.38, 2.72; C^{γ} 27.4; C^{δ} 50.6
Ser[8]	61.5	4.11	62.7	3.93, 3.93[b]	
Ser[9]	56.9	4.17	62.3	3.77, 3.77[b]	
Tyr[10]	56.2	4.17	39.5	3.28, 2.67	H^{δ} 6.44; H^{ε} 5.98; C^{δ} 132.4; C^{ε} 117.4
Asp[11]	56.2	4.49	37.6	2.93, 2.93[b]	
Gly[12]	45.5	3.94, 3.94[b]			
Tyr[13]	62.5	4.07	38.2	3.03, 3.03[b]	H^{δ} 6.98; H^{ε} 6.76; C^{δ} 127.0; C^{ε} 118.3
Cys[14]	53.0	4.39	38.1	2.58, 2.26	
Leu[15]	54.5	4.26	44.0	1.61, 1.25	H^{γ} 1.64; H^{δ} 0.73; C^{γ} 27.3; C^{δ} 25.4
Asn[16]	55.1	4.03	37.1	1.98, 1.36	
Gly[17]	45.0	4.00, 3.58			
Gly[18]	45.5	4.33, 3.45			
Val[19]	61.4	4.33	35.9	2.00	H^{γ} 1.01; C^{γ} 21.4
Cys[20]	58.8	4.99	40.3	3.37, 3.13	
Met[21]	54.6	4.97	36.1	1.90, 1.90[b]	H^{γ} 2.45; H^{ε} 1.89; C^{γ} 31.9; C^{ε} 17.3
His[22]	53.0	5.17	30.7	3.19, 2.94	H^{δ} 6.54; H^{ε} 8.53
Ile[23]	60.4	4.07	37.3	1.81	H^{γ} 1.27, 0.93; $H^{\gamma m}$ 0.81; H^{δ} 0.66; C^{γ} 27.1; C^{δ} 12.4
Glu[24]	59.0	3.66	28.1	1.98, 1.98[b]	H^{γ} 2.34; C^{γ} 33.4
Ser[25]	60.4	4.14	62.4	3.87, 3.87[b]	
Leu[26]	54.6	4.40	43.4	1.55, 1.44	H^{γ} 1.52; H^{δ} 0.82, 0.80
Asp[27]	53.5	4.32	37.5	3.17, 2.67	
Ser[28]	56.7	4.59	65.9	3.64, 3.50	
Tyr[29]	57.7	5.26	39.9	2.48, 2.28	H^{δ} 6.82; H^{ε} 6.26; C^{δ} 133.4; C^{ε} 116.7
Thr[30]	59.9	5.01	71.5	4.11	H^{γ} 1.09; C^{γ} 21.0
Cys[31]	54.0	5.30	43.2	2.77, 2.59	
Asn[32]	52.4	5.01	39.0	2.98, 2.76	
Cys[33]	54.0	4.74	37.9	3.28, 2.64	
Val[34]	62.5	4.04	32.8	2.07	H^{γ} 1.15, 1.00; C^{γ} 21.9, 21.9
Ile[35]	63.0	3.74	37.8	1.65	H^{γ} 1.35, 1.01; $H^{\gamma m}$ 0.68; H^{δ} 0.80; C^{γ} 28.6; $C^{\gamma m}$ 17.1
Gly[36]	44.5	4.10, 3.18			
Tyr[37]	57.7	5.38	41.9	2.97, 2.97[b]	H^{δ} 6.88; H^{ε} 6.65; C^{δ} 133.9; C^{ε} 117.6
Ser[38]	57.7	4.79	65.2	3.92, 3.92[b]	
Gly[39]	43.4	4.82, 3.82			
Asp[40]	56.7	4.35	43.3	3.05, 3.05[b]	
Arg[41]	55.1	4.67		1.06	H^{γ} 2.30, 1.27; H^{δ} 2.91, 2.78; C^{γ} 31.0; C^{δ} 44.0
Cys[42]	55.1	4.09	36.7	3.59, 3.16	
Gln[43]	57.2	4.00		2.18, 1.77	H^{γ} 2.64, 2.55; C^{γ} 32.3
Thr[44]	62.5	4.42	69.9	3.85	H^{γ} 1.08; C^{γ} 22.2
Arg[45]	56.8	4.15	42.2	1.62, 1.56	H^{γ} 1.21; H^{δ} 2.67; C^{γ} 27.9; C^{δ} 43.4
Asp[46]	53.0	4.60	39.4	2.75, 2.48	
Leu[47]	59.3	4.18		1.59, 1.47	H^{γ} 1.59; H^{δ} 0.88, 0.79
Arg[48]	56.7	3.80	29.7	1.43, 1.33	H^{γ} 1.38; H^{δ} 3.03; C^{γ} 22.3; C^{δ} 43.3
Trp[49]	59.0	4.20	29.3	3.22, 3.12	$H^{\delta 1}$ 7.01; $H^{\varepsilon 3}$ 7.09; $H^{\zeta 2}$ 7.38; $H^{\zeta 3}$ 6.72; $H^{\eta 2}$ 7.09; $C^{\zeta 2}$ 114.4
Trp[50]	59.3	4.26	28.9	3.05, 2.80	$H^{\delta 1}$ 7.06; $H^{\zeta 2}$ 7.38; $H^{\zeta 3}$ 7.06; $H^{\eta 2}$ 7.19; $C^{\zeta 2}$ 114.4
Glu[51]	56.9	4.17	29.2	1.97, 1.80	H^{γ} 2.19, 2.09; C^{γ} 33.4
Leu[52]	55.3	4.24	42.4	1.56, 1.56[b]	H^{γ} 1.56; H^{δ} 0.82, 0.78; C^{γ} 27.0
Arg[53]	55.2	4.19	31.1	1.80, 1.65	H^{γ} 1.46; H^{δ} 2.94; C^{γ} 27.2; C^{δ} 43.3

T = 28 °C; pH 3.1.

[a] ^{1}H and ^{13}C assignments are reported relative to the methyl resonance of internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at 0 ppm and the carbon resonances of external 1,4-dioxane in H_2O at 69.4 ppm, relative to the carbon methyl resonance of trimethylsilapentanoic acid (TSP) in a coaxial capillary tube at 0 ppm (Thanabal et al., 1994).

[b] The reported chemical shift degeneracies of methylene protons were confirmed by two-quantum spectroscopy (Montelione et al., 1987).

Using the HSQC-TOCSY data, the C^{α} chemical shifts for Ser[8] and Ser[25], having similar H^{α} values (Fig. 2), were distinguished using a H^{β}-C^{α} HSQC-TOCSY cross peak for Ser[8]. The H^{α} chemical shifts of residues Asn[1], Val[19], Asp[27] and Asp[40] are also very similar, between 4.34 and 4.36 ppm, but could be assigned using C^{α}-H^{β} cross peaks

for the spin systems of Asn[1], Val[19] and Asp[27]. The remaining Asp[40] C^{α}-H^{α} HSQC peak was then assigned by elimination. The Cys[20], Thr[30] and Asn[32] C^{α}-H^{α} cross peaks, which have degenerate H^{α} resonance frequencies, were distinguished using C^{α}-H^{β} HSQC-TOCSY connectivities for Thr[30] and Asn[32], allowing assignment of the Cys[20] HSQC peak by elimination. The C^{α} chemical shifts for Ser[9], Tyr[10], Arg[45], Glu[51] and Arg[53] were assigned using C^{α}-H^{β} connectivities of the latter four spin systems; the remaining Ser[9] C^{α}-H^{α} HSQC peak was assigned by elimination. Finally, of the spin systems Pro[7], Ser[28] and Asp[46], which have similar H^{α} resonance frequencies, the C^{α}-H^{α} HSQC peak of residue Ser[28] was assigned using two C^{α}-H^{β} HSQC-TOCSY peaks.

The remaining pairs of C^{α}-H^{α} HSQC cross peaks with similar proton chemical shifts, namely Pro[7]/Asp[46] and Pro[4]/Arg[41], were distinguished using the characteristic ^{13}C chemical shifts of proline C^{α} resonances (Richarz and Wüthrich, 1978; Kessler et al., 1990; Thanabal et al., 1994) which are approximately 7 ppm downfield of the C^{α} resonance frequencies of arginine and aspartic acid. In addition, assignments for many C^{α}-H^{α} peaks which were uniquely identified in the HSQC spectrum were confirmed with HSQC-TOCSY data.

The spin systems of residues Asn[16] and Cys[42] have similar C^{α} chemical shifts which were verified by the identification of a Cys[42] C^{α}-H^{β} HSQC-TOCSY peak, together with all of the expected Asn[16] and Cys[42] C^{β}-H^{β} HSQC peaks. Similarly, the C^{α} resonances of residues Tyr[13], Val[34] and Thr[44] and the C^{β} resonance of Ser[25] have nearly identical frequencies. These carbon assignments

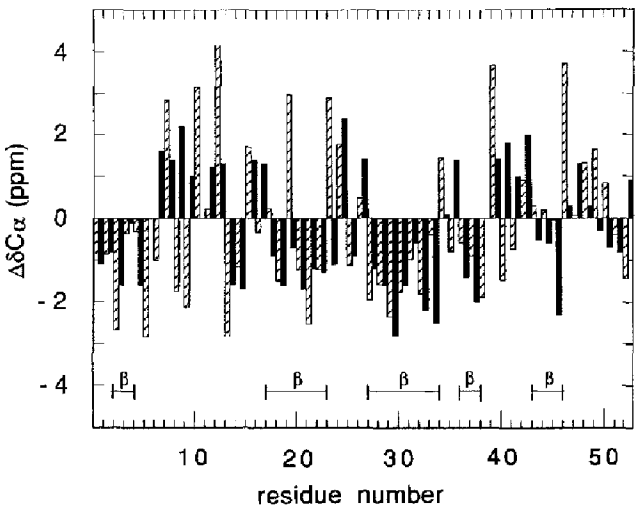

Fig. 3. Experimental (hatched bars) and back-calculated (solid bars) conformation-dependent C^{α} chemical shifts, $\Delta\delta C^{\alpha}$, in mEGF. The value of each bar indicates the conformation-dependent chemical shift ($\Delta\delta C^{\alpha} = \delta_{observed} - \delta_{random\ coil}$) for each of the C^{α} carbons of mEGF (see text). Back-calculated values of $\Delta\delta C^{\alpha}$ were estimated using an ensemble of energy-refined mEGF structures (Montelione et al., 1992; 3EGF) and the empirical ϕ-ψ correlation map of Spera and Bax (1991), as described in the text.

were corroborated by identification of Tyr[13] C^{α}-H^{β}, Cys[42] C^{α}-H^{β} and Thr[44] C^{α}-H^{β}, as well as Thr[44] C^{β}-H^{α} HSQC-TOCSY peaks.

Using a similar analysis strategy, the combined HSQC and HSQC-TOCSY data were then used to determine assignments of a large portion of the remaining aliphatic and aromatic carbon resonances. The side-chain ^{13}C resonance assignments for the five tyrosine residues and a few for the two tryptophan amino acids were also determined from these data. Altogether, 100% of the assignable C^{α}, 96% of the C^{β}, 86% of the aromatic and 70% of the remaining peripheral aliphatic resonances were identified from this analysis. These carbon resonance assignments are listed in Table 1.

Conformation-dependent chemical shifts

It is well established that C^{α} and C^{β} chemical shifts can be correlated with regular secondary structure in proteins (Spera and Bax, 1991; Wishart et al., 1991,1992; Lee et al., 1992; Reily et al., 1992; Wishart and Sykes, 1994). The secondary shift for each carbon atom was computed as the difference between its observed chemical shift (δ_{obs}) (Table 1) and the random coil shifts (δ_{rc}) of amino acid residues in H_2O solution (Thanabal et al., 1994), referenced externally to dioxane at 69.4 ppm:

$$\Delta\delta = \delta_{obs} - \delta_{rc} \qquad (1)$$

The observed conformation-dependent C^{α} chemical shifts ($\Delta\delta C^{\alpha}$) of mEGF are shown as hatched bars in Fig. 3.

Many amino acids of mEGF have negative $\Delta\delta C^{\alpha}$ values, as is expected for a structure comprising mainly β-sheet (Spera and Bax, 1991; Wishart et al., 1991,1992; Wishart and Sykes, 1994). The three-dimensional structure of EGF (Fig. 1) is composed of a triple-stranded antiparallel β-sheet in the N-terminal subdomain and a small 'double hairpin' structure (i.e., a very short triple-stranded antiparallel β-sheet) in the C-terminal subdomain (Montelione et al., 1987,1992; Kohda and Inagaki, 1988,1992b). As seen in Fig. 1A, the N-terminal antiparallel β-sheet of mEGF is formed by residues Gly[18] to Ile[23] (β-strand 2) and Ser[28] to Cys[33] (β-strand 3). There is a type-I β-bend in the polypeptide segment Glu[24] to Asp[27] (Montelione et al., 1992). Residues Tyr[3] to Pro[4] form a small stretch of a third strand of this antiparallel β-sheet (β-strand 1) which has been proposed to interact transiently with the polypeptide segment Cys[20] to His[22] of β-strand 2 (Montelione et al., 1987,1988,1992; Kohda and Inagaki, 1992b). The remaining polypeptide segment of the N-terminal subdomain, Cys[6] to Cys[14], forms a multiple-bend structure, reminiscent of a severely distorted α-helix. The C-terminal subdomain has a 'double hairpin' structure in which residues Cys[33] to Val[34], Tyr[37] to Ser[38] and Thr[44] to Asp[46] form short strands of a small three-stranded antiparallel β-sheet. The first chain reversal in

this C-terminal subdomain occurs at residues Ile[35] to Gly[36] (a type-II β-bend), while the second chain reversal is in a polypeptide loop of residues Gly[39] to Gln[43] (a multiple-bend backbone conformation).

A good correlation between experimental $\Delta\delta C^{\alpha}$ values (hatched bars) and β-strand backbone conformations (β) is evident in the data presented in Fig. 3. Almost all of the residues involved in β-sheet backbone conformation (i.e., Tyr[3], Pro[4], Gly[18], Val[19], Met[21], His[22], Ile[23], Ser[28], Tyr[29], Thr[30], Cys[31], Asn[32], Cys[33], Val[34], Tyr[37], Ser[38], Thr[44], Arg[45] and Asp[46]) show negative, or very small positive, $\Delta\delta C^{\alpha}$ values. The one significant exception within the β-strands of mEGF is residue Cys[20], which has a positive $\Delta\delta C^{\alpha}$ value, +2.9 ppm. The β-bend and multiple-bend backbone conformations exhibit both positive and negative $\Delta\delta C^{\alpha}$ values. Although the theoretical basis of ^{13}C chemical shifts is not completely understood and important deviations may arise from local electric fields (de Dios et al., 1993a,b), sequence-specific $\Delta\delta C^{\alpha}$ values for mEGF appear to provide an accurate description of its β-sheet secondary structure. In particular, extended β-strand conformations of mEGF can be identified reliably by searching for stretches of three or more negative, or nearly zero, $\Delta\delta C^{\alpha}$ values.

Comparison of experimental and back-calculated conformation-dependent chemical shifts

We also compared the experimentally measured values of $\Delta\delta C^{\alpha}$ with values calculated from the 3D structure of mEGF using the empirical correlation of backbone ϕ and ψ values which has been published by Spera and Bax (1991). Ensemble-averaged back-calculated values of $\Delta\delta C^{\alpha}$ for 16 mEGF conformers (Brookhaven Protein Data Bank (PDB) entry 3EGF) were computed from:

$$\langle \Delta\delta C^{\alpha}_{k} \rangle = N^{-1} [\Sigma_i \Delta\delta C^{\alpha}_{k} (\phi_i, \psi_i)] \qquad (2)$$

where k is the residue number in the sequence, i is the conformation number and N is the number of conformations. The value $\Delta\delta C^{\alpha}$ is the secondary shift for each conformation computed from the pair of backbone dihedral angles ϕ_i,ψ_i using expanded contour plots of the empirical relationship between $\Delta\delta C^{\alpha}$ and backbone conformation (i.e., Fig. 1 of Spera and Bax (1991)). These back-calculated values of $\Delta\delta C^{\alpha}$ are plotted for each residue in the mEGF sequence as solid bars in Fig. 3.

In general, there is good agreement between back-calculated and observed values of $\Delta\delta C^{\alpha}$, particularly in regions of regular β-sheet structure. For example, most $\Delta\delta C^{\alpha}$ values for polypeptide segments Asn[1] to Cys[6], Cys[14] to Asn[16], Val[19] to Ile[23], Ser[25] to Cys[33] and for Tyr[37] and Ser[38] are in good qualitative agreement with the experimentally derived secondary shifts. Again, the only significant exception in the β-strands is residue Cys[20]. For backbone polypeptide segments of mEGF containing turns (Cys[6] to Cys[14], Glu[24] to Asp[27], Ile[35] to Gly[36] and Gly[39] to Cys[42]), the observed and calculated $\Delta\delta C^{\alpha}$ values are generally not very consistent (Fig. 3). This may be due, in part, to the fact that amino acids involved in regular β-sheet secondary structures share stretches of common ϕ_i,ψ_i values, which is not the case for residues involved in turns.

An additional comparison between experimental and back-calculated $\Delta\delta C^{\alpha}$ values was carried out using as input the ϕ_i,ψ_i values from a single structure obtained by averaging the ϕ and ψ values for each residue in the set of 16 energy-refined (3EGF) conformers. The resulting values of $\Delta\delta C^{\alpha}$, calculated from the *average* structure, fit to the experimental data as well as the ensemble-averaged values (Eq. 2). This is not very surprising, since the 3D solution NMR structure of mEGF is generally well defined (Montelione et al., 1992) and the mathematically averaged backbone coordinates are very similar to each of the 16 energy-refined conformers.

Although α-helix and β-sheet regular structures are less reliably defined by $\Delta\delta C^{\beta}$ values (Spera and Bax, 1991), the secondary shifts for C^{β} resonances may be an additional structural probe, complementary to other parameters such as NOE interproton distance constraints, vicinal coupling constant measurements and $\Delta\delta C^{\alpha}$ values. For mEGF the $\Delta\delta C^{\beta}$ values also show some agreement with back-calculated values (Spera and Bax, 1991) in β-sheet regions (data not shown), although to a lesser degree than the $\Delta\delta C^{\alpha}$ values.

Heteronuclear NOE measurements

Fast internal motions in proteins can be characterized by ^{13}C T_1, T_2 and HNOE measurements (Allerhand et al., 1971; Nirmala and Wagner, 1988; Wagner, 1989,1993; Palmer et al., 1991,1993). For molecules of the size of mEGF and at a field strength of 125.7 MHz, ^{1}H-^{13}C HNOEs are quite sensitive to sub-nanosecond internal motions (Heinz et al., 1992), and relatively insensitive to internal motions on longer time scales. Although ^{1}H-^{13}C HNOE measurements alone cannot be used to quantitatively analyze internal molecular dynamics, they provide a sensitive approach for identifying large-amplitude internal motions with correlation times significantly shorter than the overall tumbling time τ_c (Heinz et al., 1992). Recent studies of ^{15}N relaxation rates in hTGFα show good correlations between T_1, HNOE and motional order parameters S^2, even when there are significant contributions of chemical exchange linebroadening to the transverse relaxation time T_2 (Li, 1994; Li and Montelione, 1995).

Duplicate ^{1}H-^{13}C PFG-HNOE spectra were obtained with and without aliphatic proton saturation. HNOEs for each $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ site were then calculated from:

$$\text{HNOE} = 1 + \eta = I_{sat} / I_{eq} \qquad (3)$$

where I_{sat} and I_{eq} are the ^{1}H-detected ^{13}C polarizations in spectra recorded with and without saturation of aliphatic protons during the preparation period. The quantity η is the HNOE enhancement, which for ^{13}C nuclei is always ≥ 0. The results of these measurements (HNOE = $1 + \eta$) are presented in Table 2. In mEGF at a temperature of 28 °C and pH* 3.1, these ^{1}H-^{13}C HNOE values range between 0.96–1.96 units, with estimated uncertainties of 0.05–0.20 units. As expected, most HNOE values for mEGF are slightly larger than 1.0, with an average value of 1.33 units for the whole set of residues. For several sites, HNOE values could not be determined because of peak overlaps in the $^1H^{\alpha}$-$^{13}C^{\alpha}$ region of the HNOE spectrum (cf. Fig. 1).

A large subset of residues exhibit $^1H^{\alpha}$-$^{13}C^{\alpha}$ HNOE values of 1.1 ± 0.2 units, typical of an isotropically tumbling, highly ordered small protein (Palmer et al., 1991). These include the C^{α} sites of residues Asn[1], Cys[6], Pro[7], Ser[8], Asp[11], Cys[20], His[22], Ile[23], Glu[24], Ser[25], Leu[26], Asp[27], Ser[28], Tyr[29], Cys[31], Cys[33], Tyr[37], Ser[38], Asp[40], Gln[43], Thr[44] and Arg[48]. Within this subset of backbone atoms, the average HNOE value is 1.18 units. Assuming rigid isotropic reorientation of the $^1H^{\alpha}$-$^{13}C^{\alpha}$ bond vector with no internal motions, negligible contributions from chemical shift anisotropy (Palmer et al., 1991), and the dominance of $^1H^{\alpha}$ dipolar relaxation of $^{13}C^{\alpha}$ methine nuclei, this corresponds to an approximate overall rotational correlation time $\tau_m = 3.5$ ns (Palmer et al., 1991) for mEGF at pH* 3.1 and 28 °C. This value for mEGF (molecular weight 6.0 kDa) is similar to the overall correlation time determined from ^{15}N relaxation rate measurements for the homologous 5.5 kDa hTGFα molecule, $\tau_m = 3.8$ ns at pH 6.5 and 30 °C (Li, 1994; Li and Montelione, 1995).

TABLE 2
METHINE ^{1}H-^{13}C HNOEs FOR mEGF, DETERMINED AT NATURAL ISOTOPE ABUNDANCE

Carbon site	HNOE (1 + η)	Carbon site	HNOE (1 + η)
Asn[1] C^{α}	1.26 ± 0.08[a]	Tyr[29] C^{α}	1.01 ± 0.01
Ser[2][b] C^{α}	1.46 ± 0.04	Thr[30] C^{α}	1.45 ± 0.12
Tyr[3] C^{α}	1.41 ± 0.10	Thr[30] C^{β}	1.31 ± 0.12
Pro[4] C^{α}	1.96 ± 0.21	Cys[31] C^{α}	1.09 ± 0.02
Cys[6] C^{α}	1.14 ± 0.07	Asn[32] C^{α}	1.74 ± 0.11
Pro[7] C^{α}	1.15 ± 0.11	Cys[33] C^{α}	1.20 ± 0.08
Ser[8] C^{α}	0.96 ± 0.04	Ile[35] C^{α}	1.62 ± 0.12
Asp[11] C^{α}	1.27 ± 0.09	Tyr[37] C^{α}	1.22 ± 0.08
Cys[14] C^{α}	1.74 ± 0.10	Ser[38] C^{α}	1.10 ± 0.01
Leu[15] C^{α}	1.48 ± 0.05	Asp[40] C^{α}	1.05 ± 0.08
Val[19] C^{α}	1.54 ± 0.12	Arg[41] C^{α}	1.81 ± 0.09
Cys[20] C^{α}	1.13 ± 0.07	Cys[42] C^{α}	1.40 ± 0.07
Met[21] C^{α}	1.89 ± 0.13	Gln[43] C^{α}	1.01 ± 0.04
His[22] C^{α}	1.11 ± 0.03	Thr[44] C^{α}	1.27 ± 0.07
Ile[23] C^{α}	1.18 ± 0.02	Thr[44] C^{β}	1.33 ± 0.01
Glu[24] C^{α}	1.18 ± 0.03	Asp[46] C^{α}	1.80 ± 0.12
Ser[25] C^{α}	1.22 ± 0.09	Leu[47] C^{α}	1.61 ± 0.01
Leu[26] C^{α}	1.00 ± 0.07	Arg[48] C^{α}	1.21 ± 0.01
Asp[27] C^{α}	0.99 ± 0.12	Trp[49] C^{α}	1.33 ± 0.01
Ser[28] C^{α}	1.01 ± 0.06	Leu[52] C^{α}	1.89 ± 0.14

T = 28 °C; pH* = 3.1.
[a] Errors represent the range of values obtained in two independent measurements.
[b] Methine C^{α} nuclei with HNOE values ≥ 1.3 are underlined.

Other backbone ^{13}C methine sites exhibit significant positive deviations from the values (1.1–1.3 units) expected for a rigidly tumbling $^1H^{\alpha}$-$^{13}C^{\alpha}$ bond vector with a rotational correlation time of 3.5–3.8 ns. These HNOE values, ranging from 1.33 ± 0.01 to 1.96 ± 0.21 units, are indicative of internal motions on the sub-nanosecond time scale for $^1H^{\alpha}$-$^{13}C^{\alpha}$ bonds of residues in the N-terminal β-strand 1 (i.e. Ser[2], Tyr[3] and Pro[4]), in the solvent-exposed β-bend between the second and third cystine residues (i.e. Cys[14], Leu[15], Val[19]), within β-strands 2 and 3 of the main β-sheet structure (i.e. Met[21] and Thr[30]), in the hinge region between the N- and C-terminal β-sheets (i.e. Asn[32]), in the solvent-exposed chain reversal of the C-terminal double hairpin structure (i.e. Ile[35], Arg[41] and Cys[42]), and in the solvent-exposed C-terminal polypeptide backbone (i.e. Asp[46], Leu[47], Trp[49] and Leu[52]). Most of these more mobile sites are in surface-exposed N- or C-terminal polypeptide segments, or in surface loops, except for sites Met[21], Thr[30] and Asn[32] which are in the middle of β-strands (cf. Fig. 1). Interestingly, residues Pro[4], Met[21] and Thr[30] are spacially adjacent to one another in β-strands 1, 2 and 3, respectively, while residues Val[19] and Asn[32] are also juxtaposed in β-strands 2 and 3. The relaxation behavior of these sites may be related to internal motions of the Cys[6]-Cys[20] and/or Cys[14]-Cys[31] disulfide bonds or to internal flexing within the N-terminal β-sheet (see Discussion). Residue Asn[32] has been proposed to act as conformational hinge between the N- and C-terminal subdomains of human EGF (Campion et al., 1993) and the homologous residue (Val[33]) appears to act as a conformational hinge between the subdomains of hTGFα (Li, 1994; Li and Montelione, 1995).

Applicability of ^{13}C chemical shift constraints for monitoring protein structure refinement

Table 3 shows the standard average deviation (σ) values between experimental and back-calculated conformation-dependent $^{13}C^{\alpha}$ chemical shifts ($\Delta\delta C^{\alpha}$) for four different ensembles of mEGF structures calculated over the last few years. These are designated as coordinate sets: DISMAN-1, five structures calculated with the DISMAN (Braun and Gō, 1985) program using 333 NMR constraints (Montelione et al., 1987); DISMAN-2, 16 structures calculated with DISMAN using 730 NMR constraints (Montelione et al., 1992; 1EGF); ECEPP, 16 structures calculated with DISMAN using 730 NMR constraints and energy minimized with the ECEPP/3 (Némethy et al., 1992) potential energy function (Montelione et al., 1992; 3EGF); and CONGEN, 16 structures calculated using simulated annealing with molecular dynamics using 730 NMR constraints and the CHARMm

TABLE 3
COMPARISON OF THE AVERAGE STANDARD DEVIATIONS BETWEEN EXPERIMENTAL AND BACK-CALCULATED $\Delta\delta C^{\alpha}$ VALUES FOR mEGF WITH OTHER MEASURES OF THE QUALITY OF THE NMR STRUCTURE DETERMINATION

	Average σ value[b] (ppm)			Average vdW energy[c] (kcal/mol)	Average conf. energy[d] (kcal/mol)	Average number of constraint violations per structure[e]			Backbone rmsd (Å)[f]
						NOE		Dihedral	
Refinement level[a]	ALL	BETA	BETA-RIGID			0.1–0.3 Å	>0.3 Å	>10°	
DISMAN-1	1.95	1.51	1.31	256 ± 60	4495 ± 224	37.6 ± 6.9	41.2 ± 2.5	5.2 ± 1.6	0.64
DISMAN-2 (1EGF)	1.76	1.16	1.13	77 ± 41	3003 ± 377	31.1 ± 4.5	31.4 ± 3.2	5.6 ± 1.1	0.75
ECEPP (3EGF)	1.72	1.17	1.03	−120 ± 14	2617 ± 319	27.5 ± 4.5	31.7 ± 3.8	5.5 ± 1.4	0.69
CONGEN	1.39	1.17	1.00	−186 ± 15	−673 ± 23	1.1 ± 1.1	0	0.1 ± 0.1	0.93

[a] DISMAN-1: ensemble of five NMR structures, generated from 333 constraints using DISMAN (Montelione et al., 1987); DISMAN-2: ensemble of 16 DISMAN structures, using 730 constraints (Montelione et al., 1992); ECEPP: ensemble of 16 energy-minimized DISMAN structures, using 730 constraints and ECEPP/3 (Montelione et al., 1992); CONGEN: ensemble of 16 refined structures, using 730 constraints and simulated annealing with molecular dynamics using the CONGEN computer program (Tejero, R., Bassolino-Klimas, D., Bruccoleri, R. and Montelione, G.T., manuscript in preparation). The PDB identity codes for the DISMAN-1 and ECEPP structures are given in parentheses.

[b] Standard deviations between calculated and observed $\Delta\delta C^{\alpha}$ values (σ) were computed for the set of structures at each refinement level using the following subsets of residues: ALL, all C^{α} atoms; BETA, only C^{α} atoms in β-sheet structures shown in Fig. 1A (see text); BETA-RIGID, only C^{α} atoms in β-sheet structures that have HNOE values < 1.3. For BETA and BETA-RIGID, data for residue Cys^{20} were not included.

[c] The van der Waals (vdW) energy was computed for all structures in the same way, using the Lennard-Jones part of the CHARMm potential energy function of the CONGEN computer program (Bruccoleri and Karplus, 1987).

[d] The total conformational (conf.) energy was computed for all structures in the same way, using the full CHARMm potential function of the CONGEN computer program with a distance-dependent dielectric ε = r (Bruccoleri and Karplus, 1987) and an energy cutoff of 10 Å.

[e] The NMR constraint violations were evaluated for all structures using the same complete set of 730 NOE and dihedral angle constraints. For this reason, the residual constraint violations for the DISMAN-1 structures are more numerous than those summarized in the original paper (Montelione et al., 1987).

[f] Backbone rmsd values are reported for the N, C^{α} and C' atoms in the core of the mEGF structure (i.e., residues 2–6, 18–23, 26–38 and 42–45) relative to the mathematical average structure in each ensemble.

potential energy function (Brooks et al., 1983) of the CONGEN computer program (Bruccoleri and Karplus, 1987).

Three sets of σ values were calculated for each of these four ensembles of structures: set ALL, including C^{α} chemical shift data for all residues in the estimate of σ; set BETA, including C^{α} chemical shift data only for amino acid residues in the β-sheet structures of mEGF (i.e., Ser^{2}, Tyr^{3}, Pro^{4}, Gly^{18}, Val^{19}, Met^{21}, His^{22}, Ile^{23}, Ser^{28}, Tyr^{29}, Thr^{30}, Cys^{31}, Asn^{32}, Cys^{33}, Val^{34}, Tyr^{37}, Ser^{38}, Thr^{44}, Arg^{45} and Asp^{46}), except site Cys^{20}; and set BETA-RIGID, including only those sites of set BETA with HNOE values ≤ 1.3 (Table 2). The van der Waals (vdW) energies (including only the Lennard-Jones portion of the CHARMm potential energy function) and conformational energies (including all of the terms of the CHARMm potential energy function) listed in Table 3 were calculated in the same way for all structures using CONGEN. Comparing values among these sets of structures (Table 3), it was observed that lower values of σ correlate with lower values of the average vdW or conformational energies, and with lower residual violations of the NOE and dihedral angle constraints. Interestingly, while σ values and energies were lower for the CONGEN structures, atomic root-mean-square deviations (rmsd) are higher. This can be attributed to improved sampling of the solution space by the simulated annealing protocol (Tejero, R. et al., manuscript in preparation).

These results suggested that it might be safe to use $\Delta\delta C^{\alpha}$ data as constraints to drive the structure refinement of mEGF. In order to evaluate this possibility, experimental values of $\Delta\delta C^{\alpha}$ were used to generate 34 dihedral angle constraints on φ and ψ (Table 4). The experimental $^{13}C^{\alpha}$ chemical shift data were interpreted in terms of dihedral angle ranges using the empirical relationship summarized in Fig. 1A of Spera and Bax (1991). Ranges on

TABLE 4
$\Delta\delta C^{\alpha}$ DIHEDRAL ANGLE CONSTRAINTS IN mEGF

Residue	φ (°)	ψ (°)
Ser^{2}	−100 ± 30	+110 ± 30
Tyr^{3}	−130 ± 30	+150 ± 30
Pro^{4}	−90 ± 30	+50 ± 30
Val^{19}	−110 ± 30	+130 ± 30
Met^{21}	−110 ± 30	+130 ± 30
His^{22}	−130 ± 30	+150 ± 30
Ile^{23}	−110 ± 30	+130 ± 30
Ser^{28}	−130 ± 30	+150 ± 30
Tyr^{29}	−120 ± 30	+130 ± 30
Thr^{30}	−130 ± 30	+150 ± 30
Cys^{31}	−120 ± 30	+150 ± 30
Asn^{32}	−100 ± 30	+110 ± 30
Tyr^{37}	−90 ± 30	+120 ± 30
Ser^{38}	−110 ± 30	+110 ± 30
Thr^{44a}	−90 ± 40	+50 ± 90
Arg^{45a}	−90 ± 40	+50 ± 90
Asp^{46a}	−90 ± 40	+50 ± 90

[a] In cases where two distinct regions of the φ-ψ map were consistent with the measured value of $\Delta\delta C^{\alpha}$, dihedral angle constraints were constructed to span both regions of conformational space.

these ϕ and ψ dihedral constraints (typically $\pm 30°$) were conservatively estimated from the spacings between contour lines of constant $\Delta\delta C^{\alpha}$ values in this plot (Spera and Bax, 1991). In cases where more than one region of the ϕ-ψ map was consistent with the measured values of $\Delta\delta C^{\alpha}$, the corresponding dihedral angle constraint was constructed using upper and lower bounds spanning both possible ranges of ϕ and ψ. Of these 34 backbone constraints, 14 ϕ constraints were consistent with, though generally tighter than, constraints which had been determined previously from estimates of $^3J_{HNH\alpha}$ coupling constants (Montelione et al., 1992). None of the constraints on ϕ and ψ determined from the $\Delta\delta C^{\alpha}$ data (Table 4) were inconsistent with dihedral angle constraints determined from the previously published $^3J_{HNH\alpha}$ coupling constant values (Montelione et al., 1992).

The 16 CONGEN structures described above were then used as starting points for a further 8 ps restrained molecular dynamics trajectory at 300 K using CONGEN either with (^{13}C-restrained CONGEN) or without (Control-CONGEN) the additional $\Delta\delta C^{\alpha}$ dihedral angle constraints. Contributions of $\Delta\delta C^{\alpha}$ dihedral angle constraint violations to the CONGEN target function were computed using a flat-bottomed harmonic penalty function, with variable weights beginning at 10 kcal mol^{-1} rad^{-2} at the start of the trajectory, and increased first to 50 kcal mol^{-1} rad^{-2} and then to a final value of 100 kcal mol^{-1} rad^{-2} toward the end of the trajectory. For each trajectory, the average coordinates over the last 2 ps were computed and minimized with respect to energy and constraint violations. The two resulting ensembles of structures were then used to compute σ values, average vdW and conformational energies, residual NOE and dihedral angle constraint violations, and backbone rmsd values for these mEGF structures (Table 5). Both the Control- and ^{13}C-restrained trajectories reduce the average conformational energy, largely by relaxation of some unfavorable electrostatic interactions. Although the ^{13}C-restrained CONGEN structures exhibit average energies and residual constraint violations slightly higher than the Control-CONGEN structures, the differences are relatively small and within the ranges of values observed among the 16 structures (Table 5). Overall, this test of applying experimental $\Delta\delta C^{\alpha}$ values to drive the structure refinement results in molecular structures which are consistent with NOE and other NMR data and which exhibit good vdW and conformational energies.

Discussion

^{13}C resonance assignments and HSQC-TOCSY data for mEGF

Carbon resonance assignments at natural abundance have been reported for a few other small proteins, including pancreatic trypsin inhibitor (Wagner and Brühwiler, 1986; Hansen, 1991) and the α-amylase inhibitor Tendamistat (Kessler et al., 1990). For mEGF, it was relatively straight-forward to obtain sequence-specific carbon assignments at natural isotope abundance. The ^{13}C-resolved HSQC-TOCSY spectra were especially useful for determining ^{13}C resonance assignments from the published proton assignments. In our analysis, 20 C^{α}-H^{β} connectivities and 34 C^{β}-H^{α} connectivities were observed for the 47 spin systems. It is interesting to note that 70% more C^{β}-H^{α} peaks than C^{α}-H^{β} peaks were observed in the HSQC-TOCSY spectrum. This may reflect the fact that $^{13}C^{\alpha}$

TABLE 5
SUMMARY OF RESTRAINED MOLECULAR DYNAMICS CALCULATIONS FOR mEGF USING $\Delta\delta C^{\alpha}$ DIHEDRAL ANGLE CONSTRAINTS

Ensemble[a]	Average σ value[b] (ppm)			Average vdW energy[c] (kcal/mol)	Average conf. energy[d] (kcal/mol)	Average number of constraint violations per structure[e]			Backbone rmsd (Å)[f]
						NOE		Dihedral	
	ALL	BETA	BETA-RIGID			0.1–0.3 Å	>0.3 Å	>10°	
Control-CONGEN	1.48	1.26	1.02	−186 ± 15	−765 ± 24	5.5 ± 1.3	0	0.1 ± 0.1	0.94
^{13}C-restrained CONGEN	1.44	0.91	0.96	−182 ± 15	−728 ± 22	6.7 ± 1.8	0	0.1 ± 0.1 (0.5 ± 0.5)	0.88

[a] Control-CONGEN: ensemble of 16 NMR structures refined by restrained molecular dynamics using 730 NMR constraints; ^{13}C-restrained CONGEN: ensemble of 16 NMR structures refined by molecular dynamics using 730 constraints plus 34 $\Delta\delta C^{\alpha}$ dihedral angle constraints on 17 pairs of backbone dihedral angles ϕ and ψ.

[b] Standard deviations between calculated and observed $\Delta\delta C^{\alpha}$ (σ) values were computed for the set of structures at each refinement level using the subsets of residues defined in footnote b of Table 3.

[c] The van der Waals (vdW) energy was computed for all structures as described in Table 3.

[d] The total conformational (conf.) energy was computed for all structures as described in Table 3.

[e] The NMR constraint violations were evaluated for all structures using the same set of 730 NOE and dihedral angle constraints. Small residual violations of the additional 34 $\Delta\delta C^{\alpha}$ dihedral angle constraints are not included in these statistics. The average number of violations of these additional dihedral constraints are indicated in parentheses.

[f] Backbone rmsd values are reported for the N, C^{α} and C' atoms in the core of the mEGF structure (i.e., residues 2–6, 18–23, 26–38 and 42–45) relative to the mathematical average structure within each ensemble.

nuclei generally exhibit relatively faster relaxation rates than $^{13}C^{\beta}$ nuclei.

Another interesting feature of these HSQC-TOCSY data was observed for spin systems with non-degenerate H^{β} protons. In nearly all such cases, only one of the two possible C^{α}-H^{β} cross peaks was observed. These relative cross-peak intensities are correlated with the values of the $^{3}J_{H^{\alpha}H^{\beta}}$ coupling constants for these methylene protons (Montelione et al., 1992) measured by homonuclear E.COSY. The HSQC-TOCSY data could also be used to estimate relative values of homonuclear vicinal coupling constants which were not obtained in the homonuclear E.COSY spectra, providing useful information for the determination of stereospecific assignments of some β-methylene protons (Constantine et al., 1994; Celda, B. and Montelione, G.T., manuscript in preparation).

Our experience with these natural abundance ^{1}H-^{13}C correlation spectra of mEGF indicates that such data also provide useful information for identifying amino acid spin systems and for verifying resonance assignments. The $\Delta\delta C^{\alpha}$ values themselves provide rough constraints on some backbone ϕ,ψ values, which can be combined with NOE and other scalar coupling data to restrict the local conformations of the backbone and side chains. In addition to being useful for prediction of regular backbone structure, ^{13}C resonance assignments can provide a useful monitor of the quality of a protein NMR structure. The significant amount of information which can be extracted from natural abundance ^{1}H-^{13}C correlation spectra certainly justifies the cost in instrument time expended in running these relatively insensitive experiments.

Heteronuclear NOEs and internal motions

HNOEs generally provide a reliable, though qualitative, identification of polypeptide backbone segments exhibiting sub-nanosecond internal motions. While most mEGF C^{α} nuclei exhibit HNOE values consistent with a rigid backbone on this time scale, the sites exhibiting larger HNOE values are clustered in regions of the structure (cf. Table 2 and Fig. 1B): in the N-terminal polypeptide segment Asn[1]-Pro[4], in the surface polypeptide chain reversal Cys[14]-Val[19], in the surface polypeptide loop including residues Arg[41]-Cys[42] and in the C-terminal polypeptide segment Asp[46]-Leu[52]. Detailed ^{15}N relaxation measurements on ^{15}N-enriched hTGFα at pH 7.1 (Li, 1994; Li and Montelione, 1995), including HNOE, T_1 and T_2 measurements and analysis of motional order parameters (Lipari and Szabo, 1982), reveal rapid internal motions in the corresponding homologous polypeptide segments. Internal flexibility in these regions of EGF and hTGFα is also consistent with anecdotal interpretations of proton resonance line widths and $^{1}H/^{2}H$ exchange data (Montelione et al., 1986,1987,1988,1989,1992; Brown et al., 1989; Tappin et al., 1989; Hommel et al., 1992; Kohda and Inagaki, 1992b; Moy et al., 1993).

In addition to the information obtained for the locations summarized above, these ^{1}H-^{13}C HNOE measurements in mEGF also indicate internal motions along a line through residues Pro[4], Met[21] and Thr[30], perpendicular to the principal axis of the β-sheet (see Fig. 1A), in surface β-bend residue Ile[35], and for the backbone of residue Asn[32], which forms a conformational hinge between the N- and C-terminal subdomains (Fig. 1). In hTGFα, there is no evidence for fast (i.e. sub-nanosecond) internal motions in the β-sheet or in the homologous hinge residue, though slower motions on the millisecond time scale are indicated by chemical exchange line broadening of many backbone amide ^{15}N resonances, including those of the homologous residues Met[21], Ala[30] and Val[32] (mEGF numbering). Unfortunately, it is not yet possible to better characterize the internal dynamics at these sites of mEGF by additional T_1 and T_2 ^{13}C relaxation measurements, since the sensitivities of these experiments are too low at natural ^{13}C abundance, and an expression system for mEGF is not presently available to us.

In hTGFα, chemical exchange line broadening of some ^{15}N resonances has been attributed (Moy et al., 1993; Li, 1994; Li and Montelione, 1995) to slow interconversion between conformations which differ with respect to the relative orientations of the N- and C-terminal subdomains (Fig. 1A). This interconversion involves flexing about the 'hinge' residue located between the second and third disulfide bonds (i.e. Val[32]). Hinge-bending motions in hTGFα were first suggested from comparisons of the distributions and intensities of inter-subdomain NOEs in these growth factors (Moy et al., 1993). The identification of *fast* internal motions at the corresponding Asn[32] 'hinge site' in mEGF indicates that similar hinge-bending motions may indeed occur in mEGF, though on a much faster time scale than in hTGFα. Interestingly, the principal normal mode for internal motions in mEGF is predicted to involve sub-nanosecond hinge-bending between the N- and C-terminal subdomains (Ikura and Gō, 1993).

The regions of mEGF exhibiting larger $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ HNOE values (Table 2) include residues Leu[15], Arg[41] and Leu[47], each of which have been implicated in the EGF-receptor binding epitope by site-directed mutagenesis studies. Substitutions at, or near, these positions in murine and human EGF reduce receptor affinity by 10–1000-fold without significantly disrupting the protein's structure (Moy et al., 1989; Hommel et al., 1991; Engler et al., 1992; Campion et al., 1993; Tadaki and Niyogi, 1993). The fact that these sites exhibit some internal mobility in the unbound state of mEGF, and that their relative positions may depend on the dynamic orientations of the N- and C-terminal subdomains (Fig. 1B), suggests that changes in internal dynamics may contribute to the free energy (Akke et al., 1993) of the interaction between mEGF and the EGF receptor.

Combined use of ^{13}C chemical shift and HNOE data in monitoring the structure refinement

One of the most interesting results of this study is the observation that the refinement of mEGF from an initial structure (Montelione et al., 1986,1987), through the incorporation of additional NOE and scalar coupling data (Montelione et al., 1992), the inclusion of conformational energy minimization (Montelione et al., 1992), and finally the improvement in the optimization protocol by using simulated annealing with molecular dynamics (Tejero, R. et al., manuscript in preparation) could be monitored by comparing calculated and observed values of $\Delta\delta C^{\alpha}$. This measure of the 'goodness' of the structure parallels more standard measures including van der Waals energy, total conformational energy (including hydrogen bond and electrostatic effects), and residual violations of NOE and dihedral angle constraints (Table 3). It is also possible to use some $\Delta\delta C^{\alpha}$ values to construct dihedral angle constraints which complement NOE and vicinal coupling data, although more work will be required before such chemical shift data can be used routinely in protein structure determination. The use of a dihedral angle penalty function derived from deviations between back-calculated and observed $\Delta\delta C^{\alpha}$ values appears to work best when residues exhibiting evidence for internal motions or conformational averaging are excluded from the analysis.

Conclusions

In summary, the results presented for mEGF support the view (de Dios et al., 1993a,b; Laws et al., 1993) that it is possible to use differences between calculated and observed conformation-dependent $^{13}C^{\alpha}$ chemical shifts to monitor an NMR protein structure refinement. Indeed, initial results described in this paper demonstrate that structures refined using restrained molecular dynamics with $\Delta\delta C^{\alpha}$ dihedral angle constraints exhibit good energies and are consistent with other NMR-derived data. When using structure generation programs which fit experimental NMR data to a single structure, it may be preferable to omit $\Delta\delta C^{\alpha}$ dihedral angle constraints for regions of the protein exhibiting high mobility, since the corresponding dynamically averaged chemical shift values may not be consistent with a single low-energy structure. Furthermore, although all residues exhibit improved σ values as the refinement proceeds (see the ALL data set in Table 3) we have used $\Delta\delta C^{\alpha}$ dihedral angle constraints in the restrained molecular dynamics calculations only in the β-strand regions of mEGF, since we cannot yet trust back-calculations of $\Delta\delta C^{\alpha}$ values in bend and loop regions. Nonetheless, our results strongly encourage further refinement of relationships between ^{13}C chemical shifts and polypeptide backbone conformations, as these appear to provide useful data, accessible even at natural isotope abundance, for generating and evaluating protein NMR structures.

Acknowledgements

We thank Y.-C. Li, S. Lieberman, C. Rios and G.V.T. Swapna for helpful discussions and comments on the manuscript. We also thank Molecular Simulations, Inc. (NMR Compass) and Tripos Associates, Inc. (Triad) for providing NMR data processing software. This work was supported by Grants from the Searle Scholars Program, the National Institutes of Health (GM-47014), the National Science Foundation (MCB-9407569), a National Science Foundation Young Investigator Award (MCB-9357526 to G.T.M.) and a Movilidad del Personal Investigador Award (PB88-0491, PB90-0010-C02-02 to B.C.). C.B. was supported by an N.I.H. Biotechnology Graduate Fellowship (5-T32-GM08339-05) and by a Graduate Fellowship from the New Jersey Commission on Cancer Research.

References

Akke, M., Brüschweiler, R. and Palmer, A.G. (1993) *J. Am. Chem. Soc.*, **115**, 9832–9833.

Allerhand, A., Doddrell, D. and Komoroski, R. (1971) *J. Chem. Phys.*, **55**, 189–198.

Anglister, J., Grzesiek, S., Wang, A.C., Ren, H., Klee, C.B. and Bax, A. (1994) *Biochemistry*, **33**, 3540–3547.

Baron, M., Norman, D.G., Harvey, T.S., Handford, P.A., Mayhew, M., Tse, A.G.D., Brownlee, G.G. and Campbell, I.D. (1992) *Protein Sci.*, **1**, 81–90.

Bodenhausen, G. and Ruben, D. (1980) *Chem. Phys. Lett.*, **69**, 185–189.

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

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) *J. Comput. Chem.*, **4**, 187–217.

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

Bruccoleri, R.E. and Karplus, M. (1987) *Biopolymers*, **26**, 137–168.

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

Campbell, I.D. and Bork, P. (1993) *Curr. Opin. Struct. Biol.*, **3**, 385–392.

Campion, S.R., Biamonti, C., Montelione, G.T. and Niyogi, S.K. (1993) *Protein Eng.*, **6**, 651–659.

Celda, B. and Montelione, G.T. (1993) *J. Magn. Reson. Ser. B*, **101**, 189–193.

Constantine, K.L., Friedrichs, M.S. and Mueller, L. (1994) *J. Magn. Reson. Ser. B*, **104**, 62–68.

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

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993a) *Science*, **260**, 1491–1496.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993b) *J. Am. Chem. Soc.*, **115**, 9768–9773.

Engler, D.A., Campion, S.R., Hauser, M.R., Cook, J.S. and Niyogi, S.K. (1992) *J. Biol. Chem.*, **267**, 2274–2281.

Garrett, D.S., Lodi, P.J., Shamoo, Y., Williams, K.R., Clore, G.M. and Gronenborn, A.M. (1994) *Biochemistry*, **33**, 2852–2858.

Graves, B.J., Crowther, R.L., Chandran, C., Rumberger, J.M., Li, S., Huang, K.-S., Presky, D.H., Familletti, P.C., Wolitzky, B.A. and Burns, D.K. (1994) *Nature*, **367**, 532–538.

Guerin, M., Gabillot, M., Mathieu, M.-C., Travagli, J.-P., Spielmenn, M., Andrieu, N. and Riou, G. (1989) *Int. J. Cancer*, **43**, 201–208.
Hansen, A.P., Petros, A.M., Meadows, R.P., Nettesheim, D.G., Mazar, A.P., Olejniczak, E.T., Xu, R.X., Pederson, T.M., Henkin, J. and Fesik, S.W. (1994) *Biochemistry*, **33**, 4847–4864.
Hansen, P.E. (1991) *Biochemistry*, **30**, 10457–10466.
Harvey, T.S., Wilkinson, A.J., Tappin, M.J., Cooke, R.M. and Campbell, I.D. (1991) *Eur. J. Biochem.*, **198**, 555–562.
Heinz, D.W., Hyberts, S.G., Peng, J.W., Priestle, J.P., Wagner, G. and Grütter, M.G. (1992) *Biochemistry*, **31**, 8755–8766.
Hommel, U., Dudgeon, T.J., Fallon, A., Edwards, R.M. and Campbell, I.D. (1991) *Biochemistry*, **30**, 8891–8898.
Hommel, U., Harvey, T.S., Driscoll, P.C. and Campbell, I.D. (1992) *J. Mol. Biol.*, **227**, 271–282.
Ikura, T. and Gō, N. (1993) *Proteins*, **16**, 423–436.
Kessler, H., Schmieder, P. and Bermel, W. (1990) *Biopolymers*, **30**, 465–475.
Kline, T.P., Brown, F.K., Brown, S.C., Jeffs, P.W., Kopple, K.D. and Mueller, L. (1990) *Biochemistry*, **29**, 7805–7813.
Kohda, D. and Inagaki, F. (1988) *J. Biochem.*, **103**, 554–571.
Kohda, D. and Inagaki, F. (1992a) *Biochemistry*, **31**, 677–685.
Kohda, D. and Inagaki, F. (1992b) *Biochemistry*, **31**, 11928–11939.
Kohda, D., Shimada, I., Miyake, T., Fuwa, T. and Inagaki, F. (1989) *Biochemistry*, **28**, 953–958.
Laws, D.D., de Dios, A.C. and Oldfield, E. (1993) *J. Biomol. NMR*, **3**, 607–612.
Lee, M.S., Palmer, A.G. and Wright, P.E. (1992) *J. Biomol. NMR*, **2**, 307–322.
Li, Y.-C. (1994) Ph.D. Thesis, Rutgers University, Piscataway, NJ.
Li, Y.-C. and Montelione, G.T. (1994) *J. Magn. Reson. Ser. B*, **105**, 45–51.
Li, Y.-C. and Montelione, G.T. (1995) *Biochemistry*, in press.
Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4559.
Metzler, W.J., Constantine, K.L., Friedrichs, M.S., Bell, A.J., Ernst, E.G., Lavoie, T.B. and Mueller, L. (1993) *Biochemistry*, **32**, 13818–13829.
Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W. and Scheraga, H.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8594 8598.
Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W. and Scheraga, H.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5226–5230.
Montelione, G.T., Wüthrich, K. and Scheraga, H.A. (1988) *Biochemistry*, **27**, 2235–2243.
Montelione, G.T., Winkler, M.E., Burton, L.E., Rinderknecht, E., Sporn, M.B. and Wagner, G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1519–1523.
Montelione, G.T., Wüthrich, K., Burgess, A.W., Nice, E.C., Wagner, G., Gibson, K.D. and Scheraga, H.A. (1992) *Biochemistry*, **31**, 236–249.
Moy, F.J., Scheraga, H.A., Liu, J.-F., Wu, R. and Montelione, G.T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9836–9840.
Moy, F.J., Li, Y.-C., Winkler, M.E., Rauenbuehler, P., Scheraga, H.A. and Montelione, G.T. (1993) *Biochemistry*, **32**, 7334–7353.
Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992) *J. Phys. Chem.*, **96**, 6472–6484.
Nirmala, N.R. and Wagner, G. (1988) *J. Am. Chem. Soc.*, **110**, 7557–7558.
Padmanabhan, K., Padmanabhan, K.P., Tulinsky, A., Park, C.H., Bode, W., Huber, R., Blankenship, D.T., Cardin, A.D. and Kisiel, W. (1993) *J. Mol. Biol.*, **232**, 947–966.
Palmer, A.G., Rance, M. and Wright, P.E. (1991) *J. Am. Chem. Soc.*, **113**, 4371–4380.
Palmer, A.G., Hochstrasser, R.A., Millar, D.P., Rance, M. and Wright, P. (1993) *J. Am. Chem. Soc.*, **115**, 6333–6345.
Reily, M.D., Thanabal, V. and Omecinsky, D.O. (1992) *J. Am. Chem. Soc.*, **114**, 6251–6252.
Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.
Selander, M., Persson, E., Stenflo, J. and Drakenberg, T. (1990) *Biochemistry*, **29**, 8111–8118.
Shaka, A.J., Lee, C.J. and Pines, A. (1988) *J. Magn. Reson.*, **77**, 274–293.
Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
Tadaki, D.K. and Niyogi, S.K. (1993) *J. Biol. Chem.*, **268**, 10114–10119.
Tappin, M.J., Cooke, R.M., Fitton, J.E. and Campbell, I.D. (1989) *Eur. J. Biochem.*, **179**, 629–637.
Thanabal, V., Omecinsky, D.O., Reily, D.O. and Cody, W.L. (1994) *J. Biomol. NMR*, **4**, 47–59.
Ullner, M., Selander, M., Persson, E., Stenflo, J., Drakenberg, T. and Teleman, O. (1992) *Biochemistry*, **31**, 5974–5983.
Wagner, G. (1989) *Methods Enzymol.*, **176**, 93–113.
Wagner, G. (1993) *J. Biomol. NMR*, **3**, 375–385.
Wagner, G. and Brühwiler, D. (1986) *Biochemistry*, **25**, 5839–5843.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647 1651.
Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.
Wüthrich, K. (1976) *NMR in Biological Research: Peptides and Proteins*, North Holland, Amsterdam.