Protein Science (1996), 5:578-592. Cambridge University Press. Printed in the USA.

Simulated annealing with restrained molecular dynamics using CONGEN: Energy refinement of the NMR solution structures of epidermal and type-α transforming growth factors

ROBERTO TEJERO,[1,3] DONNA BASSOLINO-KLIMAS,[2] ROBERT E. BRUCCOLERI,[2] AND GAETANO T. MONTELIONE[1]

[1] Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-5638
[2] Department of Macromolecular Structure, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543
[3] Departamento de Quimica Fisica, Facultad de Ciencias Quimicas, Universidad de Valencia, Dr. Moliner, 50 46100-Burjassot (Valencia) Spain

(RECEIVED September 19, 1995; ACCEPTED January 15, 1996)

Abstract

The new functionality of the program CONGEN (Bruccoleri RE, Karplus M, 1987, *Biopolymers 26*:137-168; Bassolino-Klimas D et al., 1996, *Protein Sci 5*:593-603) has been applied for energy refinement of two previously determined solution NMR structures, murine epidermal growth factor (mEGF) and human type-α transforming growth factor (hTGFα). A summary of considerations used in converting experimental NMR data into distance constraints for CONGEN is presented. A general protocol for simulated annealing with restrained molecular dynamics is applied to generate NMR solution structures using CONGEN together with real experimental NMR data. A total of 730 NMR-derived constraints for mEGF and 424 NMR-derived constraints for hTGFα were used in these energy-refinement calculations. Different weighting schemes and starting conformations were studied to check and/or improve the sampling of the low-energy conformational space that is consistent with all constraints. The results demonstrate that loosened (i.e., "relaxed") sets of the EGF and hTGFα internuclear distance constraints allow molecules to overcome local minima in the search for a global minimum with respect to both distance restraints and conformational energy. The resulting energy-refined structures of mEGF and hTGFα are compared with structures determined previously and with structures of homologous proteins determined by NMR and X-ray crystallography.

Keywords: conformational energy; conformational pinning; NMR structure refinement

Epidermal growth factor is a small mitogenic protein containing 53 amino acids and three disulfide bonds (Cohen, 1962). It has both sequence (~30%) and structural homology with type-α transforming growth factor (Simpson et al., 1985; Campbell & Bork, 1993). EGF and TGFα compete in binding the same membrane-bound receptor (Carpenter et al., 1983). TGFα and other EGF-like proteins play central roles in the molecular basis of wound healing (for a review see Burgess, 1989) and in the etiology of some human cancers (Burgess, 1989; Guerin et al., 1989). More than 300 EGF-like sequences have been identified as domains of larger proteins (Campbell & Bork, 1993) and many of these probably have chain folds similar to those of EGF and TGFα. Medium-resolution NMR solution structures have been described for human EGF (Cooke et al., 1987; Hommel et al., 1992) and murine EGF (Montelione et al., 1986, 1987, 1992; Kohda & Inagaki, 1992a, 1992b), and for several homologous proteins including hTGFα (Kline et al., 1990; Harvey et al., 1991; Moy et al., 1993) and the EGF-like domains from bovine coagulation factor X (Selander-Sunnerhagen et al., 1992;

Reprint requests to: Gaetano T. Montelione, Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854-5638; e-mail: montelione@nmrlab.cabm.rutgers.edu.

Abbreviations: Conf. E., total conformational energy, including electrostatic effects, computed from the CHARMM potential function; EGF, epidermal growth factor; mEGF, murine EGF; TGFα, type-α transforming growth factor; hTGFα, human TGFα; NOESY, 2D NOE spectroscopy; RMSD, RMS deviation; SAMD, simulated annealing with restrained molecular dynamics; VDW E., van der Waals energy, computed from the Lennard-Jones portion of the CHARMM potential function.

Ullner et al., 1992), human factor IX (Huang et al., 1991; Baron et al., 1992), human tissue-type plasminogen activator (Smith et al., 1994), and human urokinase-type plasminogen activator (Hansen et al., 1994). X-ray crystal structures of the EGF-like domains of human E-selectin (Graves et al., 1994; Weis, 1994) and human factor Xa (Padmanabhan et al., 1993) have also been determined recently.

CONGEN (Bruccoleri & Karplus, 1987) is a computer program for molecular structure generation. In the companion paper (Bassolino-Klimas et al., 1996), new CONGEN penalty functions are described for the representation of NMR-derived distance, dihedral angle, and vicinal scalar coupling (J) constraints. These new functionalities of CONGEN were tested with good success (Bassolino-Klimas et al., 1996) in the structure determination of the 46-amino acid protein crambin, using a large number of distance, dihedral, and J constraints computed from its X-ray crystal structure (Hendrickson & Teeter, 1981).

The previously described structures of EGF and TGFα have either taken minimal account of conformational energy (Kline et al., 1990; Harvey et al., 1991; Hommel et al., 1992; Kohda & Inagaki, 1992a, 1992b) or have attempted to refine NMR structures using a fairly complete potential energy function with simple conjugate-gradient minimization (Montelione et al., 1992; Moy et al., 1993). In this paper, we describe energy refinement of hTGFα and mEGF NMR structures by simulated annealing of molecular dynamics, using the CHARMM potential energy function (Brooks et al., 1983) and the CONGEN (Bruccoleri & Karplus, 1987) computer program. The results demonstrate the use of CONGEN for determining 3D protein structures with real NMR data. These energy-refined structures of mEGF and hTGFα are compared with previous structure determinations of EGF, TGFα, and homologous EGF-like proteins.

Results

Experimental conformational constraints

Four kinds of experimental conformational constraints were used as input for structure calculations with CONGEN: (1) NOE-derived upper-bound ^{1}H-^{1}H distance constraints; (2) constraints on ranges of dihedral angles ϕ and χ^1 determined by analysis of COSY and NOESY data; (3) upper- and lower-bound constraints for hydrogen bonds in the β-sheets based on amide ^{1}H/^{2}H exchange analysis; and (4) disulfide-bond constraints. Distance, dihedral-angle, and hydrogen bond constraints were derived from experimental NMR data for mEGF at pH 3.1 and temperature of 28 °C (Montelione et al., 1988, 1992) and for hTGFα at pH 6.5 and temperature of 30 °C (Moy et al., 1993), and were filtered to remove nonconstraining entries in the constraint list. Summaries of the numbers and distributions of NMR-derived constraints for mEGF and hTGFα are presented in Table 1. There are 13.7 and 8.5 constraints per residue for mEGF and hTGFα, respectively. A detailed description of considerations used in converting NMR data into input distance and dihedral constraints for CONGEN, together with definitions for hydrogen bond and disulfide constraints, is presented in the Materials and methods.

Table 1. *Summary of input data for structure calculations of mEGF and hTGFα*

Data for structure calculations	mEGF	hTGFα
Total number of NMR constraints	730	424
Total number of NOE constraints	644	357
Intraresidue constraints [$i = j$]	93	58
Interresidue constraints [$(i - j) = 1$]	203	131
Backbone-backbone constraints	60	56
Backbone-side-chain constraints	31	7
Side-chain-side-chain constraints	112	68
Interresidue constraints [$1 < (i - j) \leq 5$]	132	46
Backbone-backbone constraints	12	11
Backbone-side-chain constraints	24	2
Side-chain-side-chain constraints	96	33
Interresidue constraints [$(i - j) > 5$]	216	122
Hydrogen bond constraints	36	24
Disulfide constraints	18	18
Dihedral angle constraints	32	25

Variables in protocols for SAMD

Details of protocols used for SAMD are presented in the Materials and methods. In order to optimize the convergence and sampling properties of these protocols, we explored different ways of handling the experimental constraints, weighting energy and constraint terms in the target function, and generating starting points. We first studied the effects of "relaxing" experimental upper-bound NMR distance constraints. In CONGEN calculations carried out with simulated constraints described in the companion paper, we observed improved convergence and lower energy structures when each of the exact distance constraints generated from the X-ray crystal structure of crambin was loosened by 5–10% (Bassolino-Klimas et al., 1996). In addition, our calibration of the relationship between NOE intensities and internuclear distances (see the Materials and methods) assumes uniform isotropic tumbling with an identical autocorrelation function for every proton–proton pair and error-free intensity measurements, resulting in both over- and underestimates of some internuclear distances. For these reasons, simulated annealing calculations were compared for different constraint sets generated by multiplying each distance constraint by a factor $1.0 + (\underline{\text{relax}}/100)$. We also examined the effects of different relative weights on experimental restraints and conformational energy in the total target function. Because there is no general agreement about how to best choose relative weights for energy and experimental-penalty terms in NMR structure calculations, we compared two different weighting schemes for the final value of the NOE energy term, denoted by the terms hard (i.e., 100 kcal mol^{-1} Å^{-2}) and soft (i.e., 20 kcal mol^{-1} Å^{-2}). The third variable in the different protocols examined was the choice of starting conformation. We use the designations DISMAN, when the starting points were the result of previously published DISMAN calculations (Montelione et al., 1992; Moy et al., 1993), and extended, when starting from a fully extended chain with random velocities for each atom chosen from a Maxwell-Boltzmann distribution at the starting temperature of the simulation.

Solution structure of mEGF determined by SAMD

Different simulated annealing schemes were used to generate five ensembles of 16 mEGF structures from the 730 constraints sum-

marized in Table 1. These protocols are named: i (**hard, 0% relax, DISMAN**), in which the refinement was started from a set of 16 previously published DISMAN structures (Montelione et al., 1992; PDB ascension number 3EGF), allowed no relaxation of the distance constraints calibrated from the NOESY spectra, and used a high (i.e., hard) weighting factor of 100 kcal mol^{-1} $Å^{-2}$ for the final value of the NOE energy term in the empirical potential energy; ii (**soft, 10% relax, DISMAN**), a similar protocol beginning with the same set of DISMAN structures with "relaxation" of 10% for every internuclear distance constraint, and using a low weight (i.e., soft) of 20 kcal mol^{-1} $Å^{-2}$ for the final value of the NOE energy term in the empirical potential energy; iii (**hard, 10% relax, DISMAN**), a similar protocol using a high weighting factor, 100 kcal mol^{-1} $Å^{-2}$, for the final NOE energy term; iv (**hard, 15% relax, DISMAN**), the same protocol, but compensating the increase in the NOE weight with a larger amount of relaxation in the distance constraints; and v (**hard, 15% relax, extended**), a similar protocol using as the starting points fully extended polypeptide chains (all dihedral angles equal to 180°) with random initial velocities for each atom. A summary of structural statistics and RMS deviations from ideal polypeptide geometries, including statistics for bond lengths and bond angles involving hydrogen atoms, is presented in Table 2. A summary of the residual violations of the experimental constraints and conformational energies for these five ensembles of structures is presented in Table 3.

The structures obtained with protocols ii–v exhibit good covalent and peptide bond geometries (Table 2), with RMS bond length deviations of ~0.01 Å, RMS bond angle deviations of ~2.8°, and average peptide bond dihedral angles of 180 ± 2.5°. These average values, and the corresponding *maximum* bond length (0.10 Å), bond angle (22°), and peptide bond dihedral angle (5.9°) deviations, observed in these mEGF structures, are similar to those observed using perfect simulated NMR data for crambin (Bassolino-Klimas et al., 1996), and to values observed in high-resolution crystal structures (see Discussion). On the other hand, the structures generated with protocol i (**hard, 0% relax, DISMAN**) exhibit significantly larger deviations from ideal polypeptide geometries (Table 2), with RMS bond angle deviation of 4.0° and RMS bond length deviation of 0.015 Å.

Residual distance and dihedral-angle violations quantify how well the computed structures fit the experimental data. The residual constraint violations in these structures of mEGF, summarized in Table 3, were all computed using the same reference set of distance constraints (i.e., the 15% relaxed set), so we could compare the qualities of convergence obtained with the different protocols. Average values of the total conformational energy, including electrostatic effects, and the van der Waals energy computed from the Lennard-Jones portion of the CHARMM potential function, are also presented for each ensemble of structures in Table 3. In these energy calculations, nonbonded interactions were included up to a cutoff of 10 Å and electrostatic interactions were evaluated using a distance-dependent dielectric equal to the interatomic distance (see the Materials and methods for details). In our experience, with these definitions of cutoff and electrostatic potential, a good rule of thumb is that energies should range from −4 to −5 kcal/mol-residue for VDW E. and −10 to −15 kcal/mol-residue for Conf. E.

All five protocols result in structures that satisfy the experimental constraints, with few distance violations greater than 0.3 Å or dihedral-angle violations greater than 10° (Table 3).

Table 2. *Summary of RMSDs from ideal geometries for the five sets of 16 mEGF structures determined by simulated annealing with restrained molecular dynamics*

i. hard, 0% relax, DISMAN[a]	
Bond lengths[b]	0.015 Å
Bond angles[b]	3.96°
Peptide bonds (impropers)	1.21°
Peptide bonds (omega)	2.56°
ii. soft, 10% relax, DISMAN	
Bond lengths	0.010 Å
Bond angles	2.76°
Peptide bonds (impropers)	0.81°
Peptide bonds (omega)	2.29°
iii. hard, 10% relax, DISMAN	
Bond lengths	0.011 Å
Bond angles	3.01°
Peptide bonds (impropers)	0.90°
Peptide bonds (omega)	2.32°
iv. hard, 15% relax, DISMAN	
Bond lengths	0.011 Å
Bond angles	2.79°
Peptide bonds (impropers)	0.83°
Peptide bonds (omega)	2.32°
v. hard, 15% relax, extended	
Bond lengths	0.011 Å
Bond angles	2.84°
Peptide bonds (impropers)	0.83°
Peptide bonds (omega)	2.34°

[a] Soft and hard refer to the final weight of NOE term in the composite constraint and potential energy function. Hard means that the final NOE weight was 100 kcal/mol-$Å^2$, soft means that the final NOE weighting was 20 kcal/mol-$Å^2$. Designations 0, 10, and 15% relax refer to the amount by which the internuclear distance constraints were relaxed in the set of constraints. In cases i, ii, iii, and iv, the starting points were 16 structures previously obtained with the DISMAN program (Braun & Gō, 1985). In case v, the starting structures were fully extended backbone and side-chain conformations (all dihedral angles equal to 180°) with randomly assigned initial atomic velocities.

[b] Bond length and bond angle deviations are reported for all atoms, including hydrogen atoms.

However, mEGF structures with significantly lower energies were obtained with the relaxed protocols. The total conformational energies and van der Waals energies are much more negative in the four relaxed protocols ii–v (average VDW E. = −196 ± 7 kcal/mol, average Conf E. = −727 ± 69 kcal/mol) than in protocol i (average VDW E. = −112 ± 16 kcal/mol; average Conf. E. = −300 ± 53 kcal/mol). These results demonstrate that, although structures can be generated that satisfy the unrelaxed experimental distance and dihedral-angle constraints for mEGF, these structures do not exhibit very good CHARMM energies.

This requirement for constraint relaxation indicates that, on the average, the NOEs are poorly calibrated in this set of mEGF constraints. This may be due in part to internal motions of the mEGF molecule (Celda et al., 1995). Nonetheless, by relaxing the distance constraints by 10–15% (corresponding to a loosening of 0.3–0.5 Å, depending on the constraint) and using a "hard" weight for distance-constraint violations, structures are generated that fit the experimental data quite well and exhibit very good conformational energies.

Although protocols iii, iv, and v exhibited similar convergence with respect to distance and dihedral-angle constraint violations,

Table 3. *Summary of distance and dihedral angle violations and final energies for mEGF and hTGFα structure refinements*

Refinement level[a]	Average VDW energy[b] (kcal/mol)	Average Conf. energy[c] (kcal/mol)	Average number of constraint violations per structure[d] NOE 0.1–0.3 Å	NOE >0.3 Å	Dihedral >10°	Backbone RMSD (Å)[e]
mEGF						
DISMAN-1	256 ± 60	4,495 ± 224	28.6 ± 6.0	27.2 ± 4.8	5.2 ± 1.6	0.64
DISMAN-2 (1EGF)	77 ± 41	3,003 ± 377	23.7 ± 3.6	19.7 ± 2.6	5.6 ± 1.1	0.75
ECEPP (3EGF)	−120 ± 14	2,617 ± 319	16.4 ± 3.1	21.3 ± 3.0	5.5 ± 1.4	0.69
CGEN i, hard 0% relax DISMAN	−112 ± 15	−300 ± 53	0.8 ± 0.8	0.1 ± 0.1	0.2 ± 0.2	0.71
CGEN ii, soft 10% relax DISMAN	−203 ± 11	−707 ± 34	9.7 ± 2.1	1.1 ± 0.9	0	0.88
CGEN iii, hard 10% relax DISMAN	−187 ± 15	−674 ± 22	0.2 ± 0.2	0	0.1 ± 0.3	0.93
CGEN iv, hard 15% relax DISMAN	−198 ± 13	−734 ± 37	2.6 ± 1.3	0	0.1 ± 0.3	1.03
CGEN v, hard 15% relax extended	−194 ± 14	−835 ± 40	3.1 ± 2.1	0	0	1.01
hTGFα						
REDAC	100 ± 28	631 ± 106	18.1 ± 3.0	19.8 ± 3.6	1.2 ± 0.5	0.76
ECEPP	−102 ± 14	147 ± 31	20.6 ± 2.9	11.3 ± 2.9	7.3 ± 1.5	0.77
CGEN vi, hard 0% relax DISMAN	−169 ± 11	−429 ± 40	0.6 ± 0.5	0	0	0.72
CGEN vii, hard 10% relax DISMAN	−181 ± 9	−528 ± 33	2.0 ± 1.3	0	0	0.86
CGEN viii, hard 10% relax extended	−181 ± 9	−641 ± 35	1.5 ± 1.4	0	0	0.93

[a] For EGF: 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; ECEPP, ensemble of 16 energy-minimized DISMAN structures, using 730 constraints and ECEPP; CGEN, ensembles of 16 energy-refined CONGEN structures described in the text. For TGFα: REDAC, ensemble of 16 structures obtained using DISMAN and redundant dihedral angle constraints (REDAC) (Moy et al., 1993); ECEPP, ensemble of the 16 structures above energy-minimized using ECEPP; CGEN, ensembles of 16 energy-refined CONGEN structures described in the text.

[b] van der Waals' energy was computed for all structures in the same way, using the Lennard-Jones part of the CHARMM potential energy function in the CONGEN computer program (Bruccoleri & Karplus, 1987), with a cutoff of 10.0 Å.

[c] Total conformational energy (Conf. E.) was computed for all the structures in the same way, using the full CHARMM potential function in the CONGEN computer program with a distance-dependent dielectric $\epsilon = r$ and a cutoff of 10.0 Å.

[d] NMR constraint violations were evaluated for all structures using the same complete set of 730 NOE (15% relaxed) and dihedral angle constraints for the mEGF, and the 424 NOE (10% relaxed) and dihedral angle constraints for TGFα. For this reason, the residual constraint violations of the older stages of the refinement are more numerous than those summarized in the original papers.

[e] Backbone RMSDs are reported for the N, C^{α}, and C′ atoms in the core for each case. In mEGF, structure core is composed of residues 2–6, 18–23, 26–38, and 42–45; in TGFα, structure core is composed of residues 19–24, 29–35, 38–39, 44–46.

conformational energies were significantly better for the structures generated with protocol v. The average total conformational (and van der Waals) energies for structures generated by protocols ii, iii, iv, and v are −707 ± 34 (−203 ± 11), −674 ± 22 (−187 ± 15), −734 ± 37 (−198 ± 13), and −835 ± 40 (−194 ± 14) kcal/mol, respectively. Accordingly, the structures generated from extended starting structures exhibited better conformational energies than structures generated from folded DISMAN structures of mEGF.

Another important criterion used in evaluating the quality of NMR solution structures is the RMSDs between several sets of optimally superimposed atomic coordinates. The average, minimum, and maximum RMSD values for each ensemble of conformers are reported in Table 4. The mEGF molecule is composed of two partially overlapping subdomains (Montelione et al., 1986, 1987), the N-terminal subdomain (residues Asn-1–Cys-33) and the C-terminal subdomain (residues Asn-32–Leu-47). The last few residues (Arg-48–Arg-50) are poorly defined by the NMR data. For these reasons, RMSDs are reported in Table 4 separately for the entire molecule (residues 1–50), the entire molecule minus the C-terminal tail (residues 1–47), the N-terminal subdomain (residues 1–33), the C-terminal subdomain (residues 32–47), and the β-sheet core (residues 2–6, 18–23, 26–38, and 42–45), excluding surface loops. Stereo diagrams of superimposed backbone atoms of two mEGF ensembles are shown in Figure 1A and B.

Comparing the different protocols, a strong correlation was observed between larger values of atomic RMSDs and lower values of the conformational energy. A similar observation was made using simulated NMR constraint data for crambin (Bassolino-Klimas et al., 1996). In all cases, the lowest RMSDs for the mEGF ensembles were obtained for protocol i, which includes the highest energy structures, whereas the largest RMSDs are observed for protocols iv and v, which include the lowest energy structures (compare values reported in Tables 3 and 4). For example, considering the backbone atoms of core residues, the RMSD values in Table 4 are 0.71, 0.88, 0.93, 1.03, and 1.01 Å for protocols i–v, respectively. Protocols using larger amounts of constraint relaxation together with "hard" NOE weights better span the conformational solution space consistent with both the NMR data and the conformational energy potential. Although protocol v, beginning with extended structures, results in individual structures with the largest backbone and heavy-atom RMSD values, these structures satisfy the experimental constraints reasonably well, and exhibit the lowest conformational energies. Unfortunately, this means that our

Table 4. *RMSDs for superposed mEGF structures*[a]

Residue range	Protocol				
	i. hard, 0% relax, DISMAN	ii. soft, 10% relax, DISMAN	iii. hard, 10% relax, DISMAN	iv. hard, 15% relax, DISMAN	v. hard, 15% relax, extended
1-50					
Backbone	1.39 (0.94-1.96)	1.58 (1.05-2.11)	1.58 (1.12-2.37)	1.63 (1.08-2.30)	1.78 (1.26-2.17)
All heavy	1.95 (1.49-2.58)	2.15 (1.71-2.68)	2.18 (1.81-3.00)	2.21 (1.80-2.86)	2.47 (1.92-2.83)
1-47					
Backbone	1.02 (0.71-1.41)	1.19 (0.90-1.45)	1.23 (0.81-1.89)	1.30 (0.96-1.80)	1.37 (1.14-1.74)
All heavy	1.44 (1.05-1.90)	1.64 (1.43-1.88)	1.67 (1.23-2.23)	1.74 (1.43-2.15)	1.94 (1.64-2.43)
1-33					
Backbone	0.80 (0.53-1.12)	0.97 (0.72-1.21)	0.88 (0.52-1.31)	1.03 (0.77-1.30)	1.21 (0.95-1.41)
All heavy	1.11 (0.78-1.51)	1.35 (1.00-1.63)	1.26 (0.85-1.84)	1.41 (1.19-1.77)	1.82 (1.60-2.06)
32-47					
Backbone	0.95 (0.51-1.35)	0.98 (0.49-1.77)	1.11 (0.75-1.60)	1.03 (0.71-1.44)	1.00 (0.84-1.78)
All heavy	1.63 (1.27-2.08)	1.64 (1.37-2.30)	1.73 (1.33-2.27)	1.73 (1.34-2.14)	1.67 (1.31-2.84)
Core[b]					
Backbone	0.71 (0.52-0.89)	0.88 (0.67-1.14)	0.93 (0.61-1.50)	1.03 (0.56-1.40)	1.01 (0.65-1.45)
All heavy	1.10 (0.84-1.42)	1.31 (1.15-1.68)	1.36 (1.00-1.97)	1.43 (1.08-1.87)	1.54 (1.20-1.97)

[a] All RMSDs of atomic coordinates are computed in Ångstroms relative to the average coordinates computed separately for each of the five protocols. The range of RMSDs is given in parentheses.

[b] Core region is defined by residues 2-6, 18-23, 26-38, and 42-45.

best energy-refinement procedures result in less tightly superimposed bundles than those obtained from ensembles of high-energy structures.

The dihedral angle order parameter $S(\theta)$ (Hyberts et al., 1992) provides another useful measure of conformational similarity within an ensemble of structures. Plots of S-values of backbone ϕ and ψ dihedral angles versus sequence number for the mEGF structures generated with the two best protocols, iv (hard, 15% relax, DISMAN) and v (hard, 15% relax, extended), are shown in Figure 2. For both protocols, the local polypeptide backbone is reasonably well defined with S values > 0.85. The less well-defined regions ($S < 0.85$ for phi or psi) are localized primarily in surface loops including polypeptide segments Ser-8–Tyr-13, Ile-23–Ser-25, Ser-38–Arg-41, and Arg-45–Arg-53.

Solution structure of human TGFα determined by SAMD

A total of 424 NMR-derived distance and dihedral constraints were used for SAMD calculations on hTGFα (Table 1). For hTGFα, there are fewer constraints available than for mEGF. This is attributed to conformational dynamics between the two subdomains of hTGFα, which appear to attenuate many of the expected intersubdomain NOEs (Moy et al., 1993; Li & Montelione, 1995).

Based on our experience with mEGF, only three different simulated annealing protocols were used in the hTGFα calculations: vi **(hard, 0% relax, DISMAN)**, starting from a set of 16 previously published DISMAN structures (Moy et al., 1993), with no relaxation of the distance constraints calibrated from the NOESY spectra, and using a "hard" weighting factor of 100 kcal mol^{-1} $Å^{-2}$ for the final value of the NOE energy term in the empirical potential energy; vii **(hard, 10% relax, DISMAN)**, a similar protocol beginning with the same set of DISMAN structures with relaxation of 10% for every internuclear distance constraint; and viii **(hard, 10% relax, extended)**, a similar protocol using as the starting points fully extended polypeptide chains (all dihedral angles equal to 180°) with random initial velocities for each atom. A summary of structural statistics and RMSDs from ideal polypeptide geometries for these ensembles of structures is presented in Table 5, and a summary of the residual violations of the experimental constraints and the conformational energies is shown in Table 3.

The structures obtained with relaxed protocols vii and viii exhibit good covalent and peptide bond geometries (Table 5), with RMS bond length deviations of ~0.01 Å, RMS bond angle deviations of ~2.8°, and average peptide bond dihedral angles 180 ± 2.3°. These average values, and the corresponding maximum bond length (0.06 Å), bond angle (16°), and peptide bond dihedral angle (4.5°) deviations, observed in these hTGFα structures, are similar to those obtained for mEGF (described above) and for simulated data sets derived from the crystal structure of crambin (Bassolino-Klimas et al., 1996). However, the structures generated without constraint relaxation (protocol vi [hard, 0% relax, DISMAN]) exhibit somewhat larger deviations from ideal polypeptide geometries (Table 5): with RMS bond length deviations of 0.012 Å and RMS bond angle deviations of 3.2°. As with mEGF, somewhat better covalent geometry was obtained using the relaxed protocols.

Residual constraint violations in these structures of hTGFα are summarized also in Table 3. As with mEGF, these violations of the experimental constraints were all computed using the same set of upper-bound distance (i.e., the 10% relaxed set) and dihedral-angle constraints as references. All three protocols result in structures that satisfy the experimental constraints, with no distance violations greater than 0.2 Å or dihedral-angle violations greater than 5°. However, hTGFα structures with signifi-

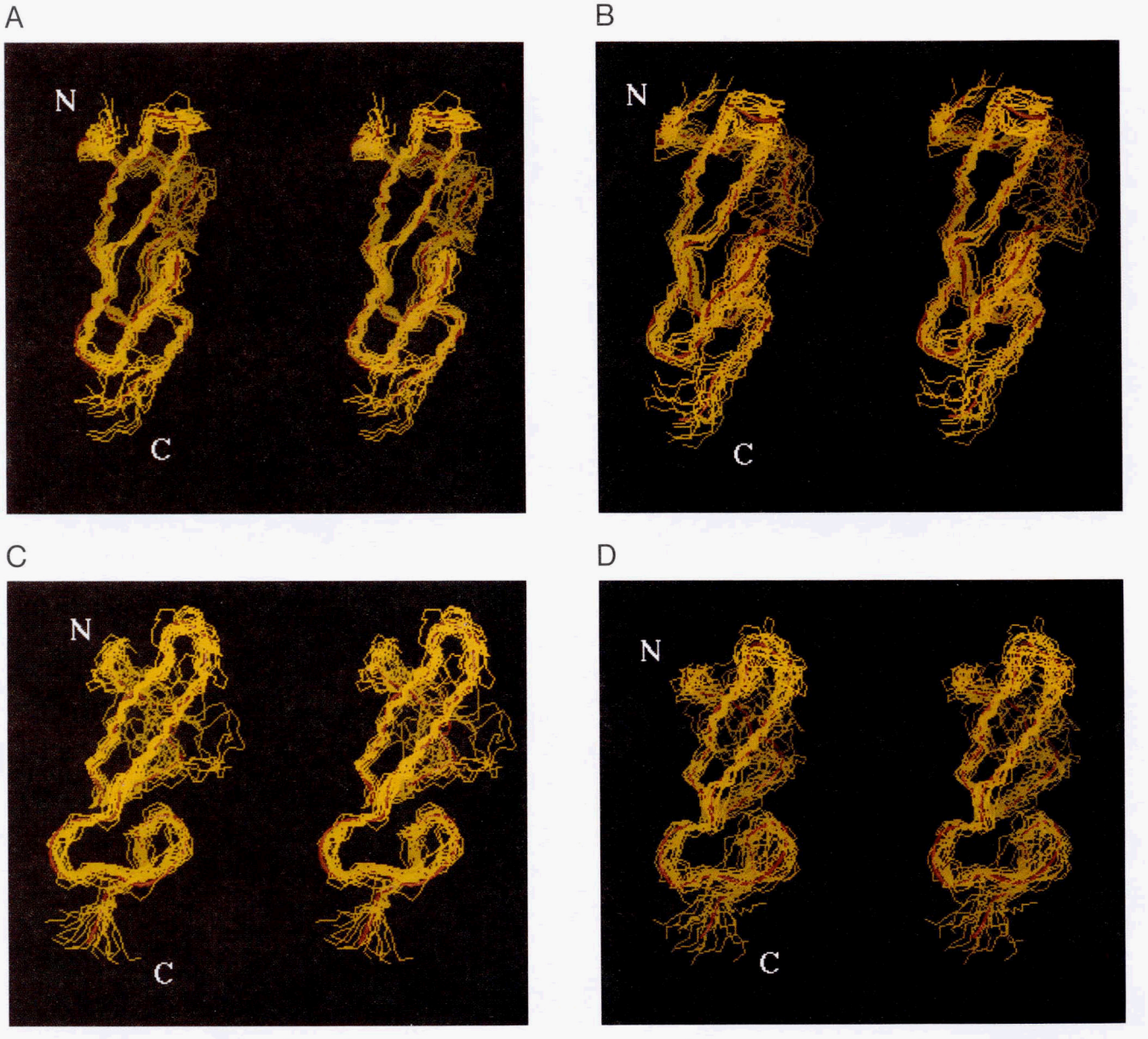

Fig. 1. Stereo drawings of 16 energy-refined backbone conformers of mEGF and hTGFα generated using the (**A**) hard, 0% relax, DISMAN mEGF, (**B**) hard, 15% relax, extended mEGF, (**C**) hard, 0% relax, DISMAN hTGFα, and (**D**) hard, 10% relax, extended hTGFα protocols.

cantly lower energies were obtained with the relaxed protocols. The total conformational and van der Waals energies are much more negative in the two relaxed protocols vii and viii (average VDW E. = −181 ± 9 kcal/mol, average Conf E. = −528 ± 33 kcal/mol) than in protocol vi (average VDW E. = −169 ± 11 kcal/mol; average Conf. E. = −429 ± 40 kcal/mol). As was observed in the SAMD calculations on mEGF, structures can be generated that satisfy the unrelaxed distance and dihedral-angle constraints for hTGFα, but these structures do not exhibit CHARMM energies as favorable as for those generated with relaxed protocols. Although the differences in Conf. and VDW energies for structures generated with and without distance-constraint relaxation were not as great for hTGFα as for mEGF, the requirement for constraint relaxation in hTGFα indicates that, on average, the NOEs are poorly calibrated in this set of constraints. This may be due in part to internal motions of the hTGFα molecule (Li & Montelione, 1995). Although protocols vi, vii, and viii exhibited similar convergence with respect to distance and dihedral-angle constraint violations, the lowest conformational energies were obtained with protocol viii using 10% constraint relaxation, hard weights, and extended starting conformations.

Atomic RMSD values for each of these three ensembles of hTGFα structures are presented in Table 6. Like mEGF, the hTGFα molecule is composed of two partially overlapping subdomains, the N-terminal subdomain (residues Val-1–Cys-33) and the C-terminal subdomain (residues Val-33–Asp-47). The first 14 residues (Val-1–Gln-14) and last few residues (Leu-48–Ala-50)

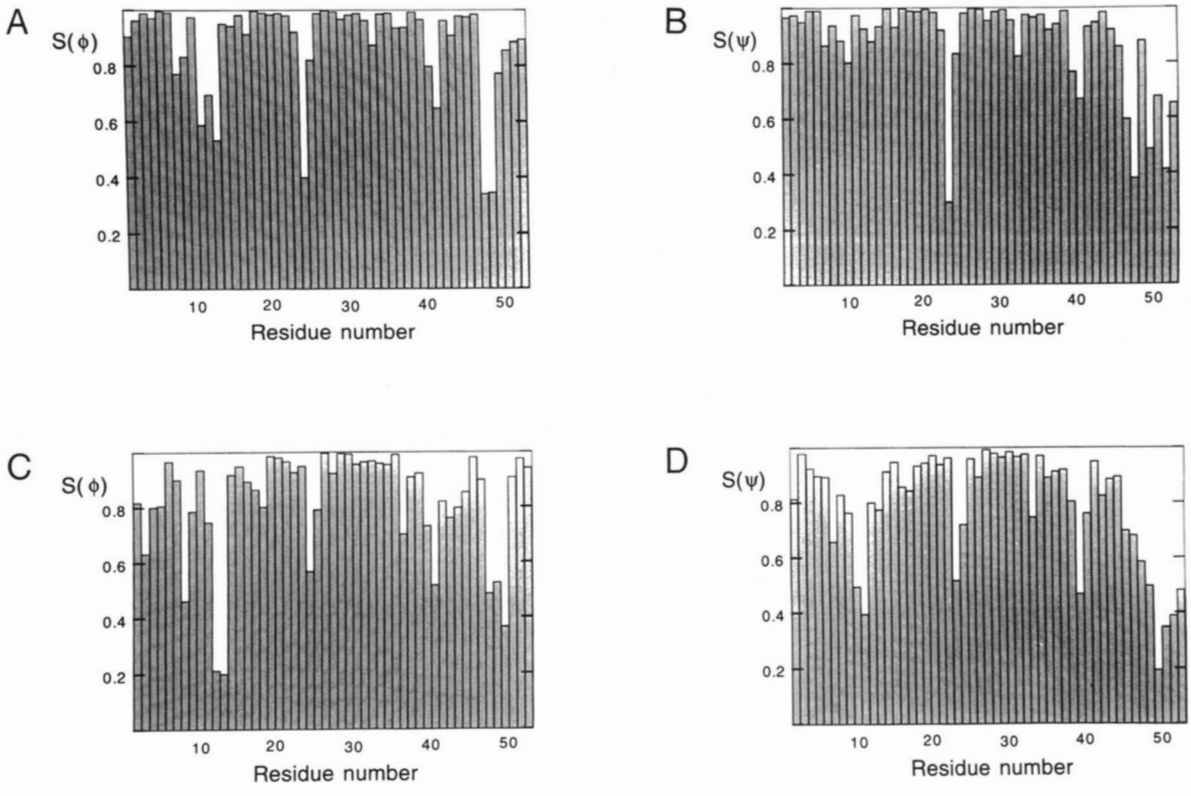

Fig. 2. Plot of backbone dihedral-angle order parameters $S(\phi)$ and $S(\psi)$ as a function of the residue number for energy-refined structures of mEGF. **A:** $S(\phi)$ for hard, 15% relax, DISMAN. **B:** $S(\psi)$ for hard, 15% relax, DISMAN. **C:** $S(\phi)$ for hard, 15% relax, extended. **D:** $S(\psi)$ for hard, 15% relax, extended.

Table 5. *Summary of structural statistics for three sets of 16 TGFα structures refined by simulated annealing with molecular dynamics*[a]

vi. hard, 0% relax, DISMAN	
Bond lengths[b]	0.012 Å
Bond angles[b]	3.21°
Peptide bonds (impropers)	0.81°
Peptide bonds (omega)	2.37°
vii. hard, 10%, DISMAN	
Bond lengths	0.010 Å
Bond angles	2.80°
Peptide bonds (impropers)	0.76°
Peptide bonds (omega)	2.28°
viii. hard, 10%, EXTENDED	
Bond lengths	0.010 Å
Bond angles	2.74°
Peptide bonds (impropers)	0.71°
Peptide bonds (omega)	2.25°

[a] Hard means that the final NOE weight was 100 kcal/mol-Å^2. The designations 0 and 10% relax refer to the amount by which the internuclear distance constraints were relaxed in the set of constraints. In cases vi and vii, the starting point was the family of 16 structures previously obtained with the DISMAN program (Braun & Gō, 1985), whereas for case viii, the starting structures were fully extended backbone and side-chain conformations (all dihedral angles equal to 180°) with randomly assigned initial atomic velocities.

[b] Bond length and bond angle deviation are reported for all atoms, including hydrogen atoms.

Table 6. *RMSDs for superposed TGF-alpha structures*[a]

	Protocol		
Residue range	vi. hard, 0% relax, DISMAN	vii. hard, 10% relax, DISMAN	viii. hard, 10% relax, extended
6–47			
Backbone	1.58 (1.13–2.11)	1.70 (1.17–2.18)	1.68 (1.12–2.22)
All heavy	2.29 (1.76–2.79)	2.41 (1.89–2.91)	2.26 (1.81–2.90)
15–47			
Backbone	0.98 (0.65–1.89)	1.20 (0.69–1.83)	1.31 (0.76–2.07)
All heavy	1.54 (1.18–2.46)	1.74 (1.31–2.38)	1.79 (1.19–2.61)
15–34			
Backbone	0.66 (0.38–1.26)	0.80 (0.56–1.14)	0.79 (0.45–1.42)
All heavy	1.41 (1.01–2.08)	1.50 (1.13–2.03)	1.49 (1.07–2.18)
33–47			
Backbone	0.80 (0.52–1.30)	0.85 (0.50–1.10)	1.18 (0.61–3.10)
All heavy	1.27 (0.91–2.01)	1.39 (1.03–1.93)	1.61 (1.03–2.81)
Core[b]			
Backbone	0.72 (0.45–1.11)	0.86 (0.40–1.40)	0.93 (0.50–1.50)
All heavy	1.21 (0.83–1.71)	1.36 (0.97–1.90)	1.45 (0.99–2.21)

[a] All RMSDs of atomic coordinates are computed in Ångstroms relative to the average coordinates computed separately for each of the three protocols. The range of RMSDs are given in parentheses.

[b] Core region is defined by residues 19–24, 29–35, 38–39, and 44–46.

are poorly defined from the NMR data at pH 6.5. For these reasons, RMSDs are reported in Table 6 separately for two reasonably well-defined regions of the entire molecule (polypeptide segments of residues 6–47 and 15–47), the well-defined portion of the N-terminal subdomain (residues 15–34), the C-terminal subdomain (residues 33–47), and for the β-sheet core (residues 19–24, 29–35, 38–39, and 44–46). Stereo diagrams showing the complete backbone structures of hTGFα are shown in Figure 1C and D. As described above for mEGF, for hTGFα we again found a correlation between larger values of atomic RMSDs and lower values of the conformational energy; the lowest RMSDs were obtained for protocol vi, which includes the highest-energy structures, whereas the largest RMSDs are observed for protocols vii and viii, which include the lowest-energy structures. As with mEGF, our best energy-refinement procedures for hTGFα exhibit less tightly superimposed bundles than those obtained from the unrelaxed ensemble of higher-energy structures.

Dihedral angle order parameters $S(\theta)$ were also computed for these hTGFα structures. Plots of S-values of backbone ϕ and ψ dihedral angles versus sequence number for the structures generated with the two best protocols, iv (hard, 10% relax, DISMAN) and v (hard, 10% relax, extended), are shown in Figure 3. For both protocols, the local polypeptide backbone is reasonably well defined with S values > 0.85. The less well-defined regions ($S < 0.85$ for phi or psi) are localized primarily in the N-terminal polypeptide segment (Val-1–Phe-15) and in surface loops including polypeptide segments Cys-34–Ser-36, Val-39–Cys-43, and Ala-46–Ala-50.

Discussion

Protocols for energy refinement using SAMD

In carrying out these SAMD calculations, our two goals were to demonstrate the utility of the CONGEN computer program for energy refinement of NMR solution structures, and to generate better structures of mEGF and hTGFα. Our results demonstrate with real NMR data that SAMD with CONGEN can be used either for energy refinement of a previously determined set of NMR structures (e.g., a set of DISMAN or DIANA structures) or for generating molecular structures from starting conformations with extended conformations. Our experience with these two NMR data sets revealed that, when the starting points are fully extended polypeptide chains with random atomic velocities, the resulting ensemble of minimized structures spans more conformational space consistent with the experimental constraints than structures generated from folded starting points. This conclusion is based on the fact that the conformers generated from extended structures have larger RMSDs, but similar qualities of constraint satisfaction compared to conformers generated from previously calculated DISMAN structures. Even when using identical sets of relaxed constraints, the structures generated from extended starting points exhibited lower final energies than SAMD conformers generated from DISMAN structures.

The best protocols we have tested involve relaxation of the upper-bound distance constraints and high weights on the

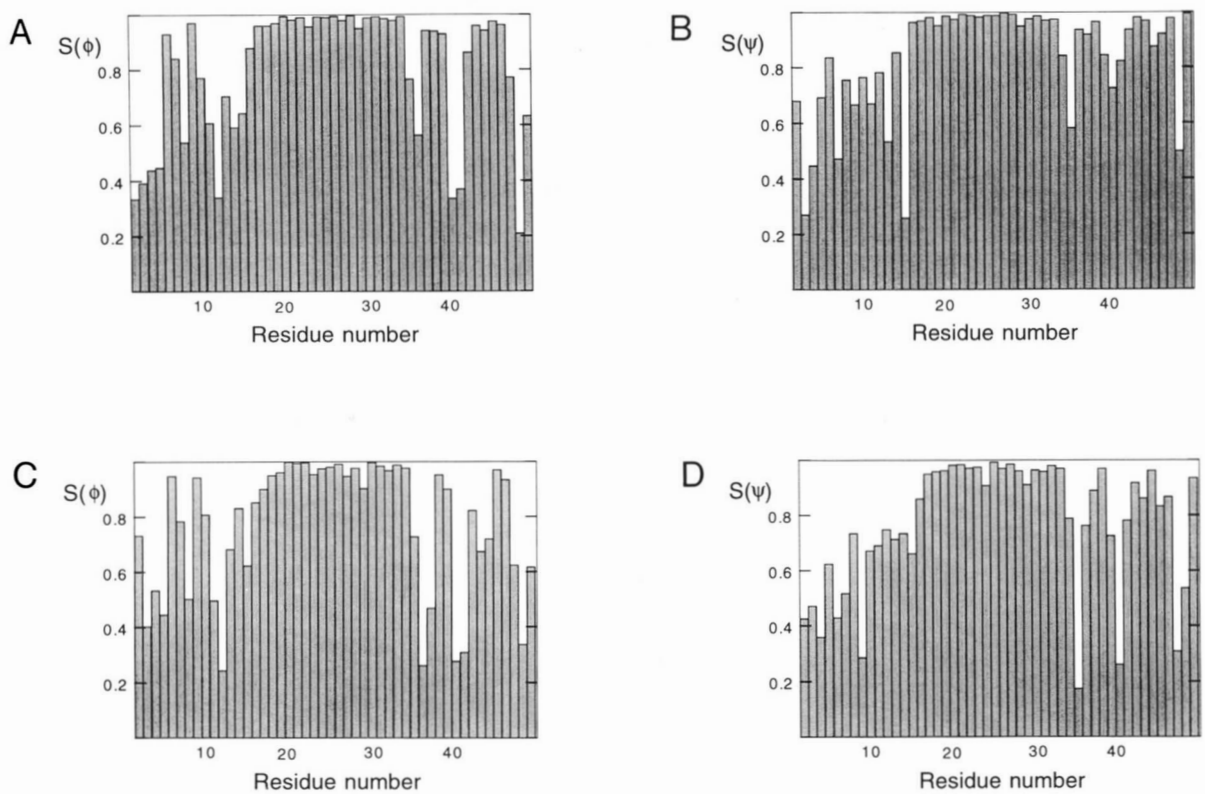

Fig. 3. Plot of backbone dihedral-angle order parameters $S(\phi)$ and $S(\psi)$ as a function of the residue number for energy-refined structures of hTGFα. **A:** $S(\phi)$ for hard, 10% relax, DISMAN. **B:** $S(\psi)$ for hard, 10% relax, DISMAN. **C:** $S(\phi)$ for hard, 10% relax, extended. **D:** $S(\psi)$ for hard, 10% relax, extended.

distance-restraint term of the penalty function relative to the potential energy. These "hard" protocols with relaxed constraints generate structures that satisfy the experimental constraints and have lower conformational energies than those generated with "soft" protocols. However, although constraint loosening provides lower-energy structures, it also results in ensembles of structures with larger RMSD values.

Using both simulated (Bassolino-Klimas et al., 1996) and real (in this work) distance constraints, we see improved energies in structures computed with relaxed constraints. Comparing the total conformational energy and van der Waals component of the total conformational energy in Table 3, it is clear that a major part of energetic improvement is due to improved electrostatics in the "relaxed" structures. Although significant gains are obtained in both van der Waals energies and bond angle and bond length strain (Tables 2, 5) by introducing some (~10%) constraint relaxation (e.g., compare results for protocols i with ii-v, or vi with vii-viii), further constraint relaxation results primarily in improved electrostatics. Some of this energetic relaxation may be due to the lack of a realistic description of the effects of solvent on these electrostatics.

Table 3 also provides a comparison of conformational energies and residual constraint violations for sets of mEGF and hTGFα structures calculated by our laboratory over the last few years. For mEGF, these are designated as coordinate sets: DISMAN-1, five structures calculated with the DISMAN program using 333 NMR constraints (Montelione et al., 1987); DISMAN-2, 16 structures calculated with DISMAN using 730 NMR constraints (Montelione et al., 1992; PDB ascension number 1EGF); ECEPP, 16 structures calculated with DISMAN using 730 NMR constraints and energy minimized with the ECEPP (Némethy et al., 1983, 1992) potential energy function (Montelione et al., 1992; PDB ascension number 3EGF); and five ensembles of CONGEN structures from this work, each consisting of 16 structures calculated with CONGEN using 730 NMR constraints and protocols i-v. All of these statistics were computed using the same set of mEGF NMR constraints (i.e., the complete set of 730 constraints with 15% relaxation of the distance constraints) and the CHARMM potential function of CONGEN. Whereas there is a clear correlation between improved convergence with respect to residual constraint violations, VDW energies, and Conf. energies in these families of mEGF structures (Table 3), there is an inverse correlation with respect to backbone RMSD values. Indeed the "worst" structures with respect to energies and constraint violations exhibit the smallest values of RMSD. Similar results for several families of hTGFα structures are also shown in Table 3.

Conformational "pinning" in NMR structure determination

It is a general aim of the NMR spectroscopist to generate ensembles of structures with the tightest superpositions of atomic coordinates; i.e., to determine the most precise structure possible. Accordingly, RMSD values are sometimes presented as a simple, objective measure of the quality of the structure determination, and are often treated as if they also provide a rough estimate of the precision of the structure. Neither of these assumptions is generally valid. For example, very tightly defined (i.e., low RMSD) structures can arise when inconsistent constraints "pin" the polypeptide chain or side chain between two different conformations. Such inconsistent constraints can arise from inaccurate interpretation of NOE cross peak intensities, from mis-assigned NOESY cross peaks, or from multiple conformations that are in rapid dynamic equilibrium relative to the chemical-shift timescale. Inconsistencies can also arise between the NMR constraints and the potential energy function if these two components of the target function have different global minima or if there is conformational averaging of the NMR data. Conformational pinning can result in apparently precise (i.e., low RMSD) but inaccurate structures. As is illustrated in Figure 4, molecular structures generated from such inconsistent constraints (or inconsistencies between the experimental restraint and energy terms of the target function) are characterized by somewhat higher conformational energies and small residual violations of distance and/or dihedral-angle constraints. These small constraint violations may sometimes be considered to be acceptable, even though the structure is highly strained and, at least locally, incorrect. The resulting local structures do not correspond to any real low-energy conformation, but rather to inaccurate structures with apparent, but misleading, high precision.

In this work, we have used the CHARMM conformational energy potential as a guide in identifying locally inconsistent structures. The simulated annealing procedures that we have developed are very powerful, and easily satisfy all of the exper-

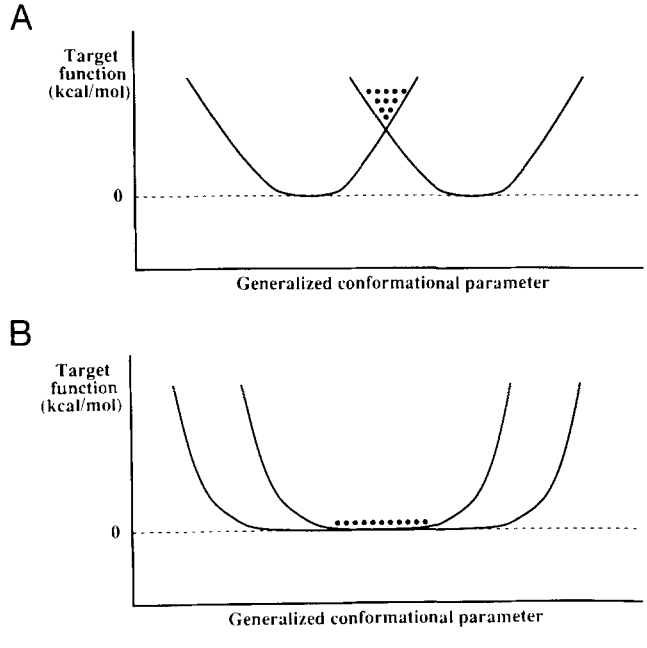

Fig. 4. A: Schematic diagram showing an example of "conformational pinning" resulting from inconsistent terms of the target function. In this generalized example, there are two restraints, arising from inconsistent experimental data or inconsistencies between the global minimum of the experimental restraint and potential energy terms of the target function. In an attempt to simultaneously satisfy both of these inconsistent restraints, the conformers become "pinned" at a cusp point. The resulting structures appear to be tightly defined with small RMSD values, but are, in fact, inaccurate. **B:** By loosening the upper- and/or lower-bounds on the experimental constraints and simultaneously strongly penalizing further deviations from these bounds (i.e., using steeper slopes on the restraint wells), structures can be generated that better sample the solution space consistent with the experimental data and potential energy function and that do not exhibit unrealistically small RMSD values.

imental distance constraints to within 0.2–0.3 Å and the dihedral angle constraints to within 5–10°. However, this is sometimes accomplished at the expense of the conformational energy, particularly the bond-angle bending and van der Waals energy (i.e., Lennard–Jones component) terms of the CHARMM potential function. Accordingly, the energy values in the CONGEN-minimized structures are a useful probe of conformational "pinning" or "strain," where the composite target function of CONGEN attempts to satisfy the experimental NMR constraints by forcing atoms too close to one another or by distorting the covalent geometry. These energy values can also be used to identify errors in the input constraint list that do not result in significant constraint violations. For example, in another protein structure refinement in progress in the laboratory (M. Tashiro & G.T. Montelione, unpubl. results), high-energy values for the CONGEN-minimized structures provided key clues for identifying incorrectly packed α-helices due to several mis-assigned NOESY cross peaks, even though these incorrect structures satisfied all of the input constraints.

The requirement for constraint loosening may arise from one or more of the following factors: (1) inaccuracies in the derivation of NMR constraints from the NOESY data; (2) inaccuracies in the potential energy function; and (3) omission of the effects of rapid conformational averaging on the the NMR data. Spin diffusion and differential relaxation effects corrupt the simple relationship between NOE intensity and internuclear distances unless efforts are made to extract distance information from full relaxation matrix analysis. Different interproton pairs can also exhibit different correlation times due to internal motions. Even if distance constraints are accurate, the absence of an explicit solvent term in the composite energy and constraint function is a serious shortcoming of the current implementation of CONGEN. The effects of conformational averaging can be accounted for using CONGEN, as has been demonstrated recently using carefully calibrated NOE data for a small peptide (Constantine et al., 1995a, 1995b). No attempt was made to do this with the relatively imprecise distance constraints available for mEGF and hTFGα. Even when simulated exact distance constraints are derived from a unique and static structure, it is our experience that constraint loosening provides improved convergence in the search for the global minimum of the composite target function by SAMD (Bassolino-Klimas et al., 1996).

Relative weighting of conformational-energy and experimental-constraint terms of the target function

In this study, we have attempted to evaluate optimal ways of implementing restraint relaxation for NOESY data interpreted using the simple two-spin approximation and without efforts to compute ensemble-averaged NOEs. Two approaches to constraint relaxation were explored: (1) use of low weights (i.e., soft) on the distance restraint terms of the composite penalty function together with strict interpretation of NOE intensities (i.e., no relax); and (2) loosening or relaxation of the entire set of constraints together with high weights on the distance restraint terms in the penalty function. Our results on both mEGF and hTGFα show that the best results (both in terms of constraint satisfaction and good energetics) are obtained by combining moderate restraint relaxation (i.e., 10–15%) with high weight on further violations of these constraints. The effects of this constraint loosening with steeper target-function wells on conformational pinning is illustrated in Figure 4; the resulting structures are generally more relaxed in the force field, but also exhibit higher RMSD values.

Deviations from ideal polypeptide geometry in structures generated by SAMD

A convenient indication of how "relaxed" a structure is in the force field of choice is to evaluate the residual RMSDs of bond lengths, bond angles, and peptide bond geometries from ideal values. Tables 2 and 5 summarize these structural statistics for both mEGF and hTGFα. These deviations from idealized geometries for the calculated structures are similar to those reported for high quality NMR and X-ray crystallography structures (for references see Bassolino-Klimas et al., 1996), particularly considering that the statistics reported here include bond lengths and bond angles involving hydrogen atoms. These results show that the structures generated with loosened constraints are reasonably "relaxed" in the CHARMM (Brooks et al., 1983) force-field, and that the composite CONGEN target function for both NMR-derived restraint terms and CHARMM energy terms is reasonably self-consistent.

Comparison of solution NMR structures of mEGF and hTGFα

Ribbon diagrams of representative energy-refined solution structures of mEGF and hTGFα, determined in this paper from NMR data recorded at pH 3.1 and pH 6.5, respectively, are shown in Figure 5. At acidic pH, the structure of hTGFα is much more dynamic and poorly defined, whereas near neutral pH, EGF exhibits aggregation and poor spectral properties. CD and 1D NMR studies indicate little or no structural change in mEGF over the pH range 3–7 (unpubl. results). An important

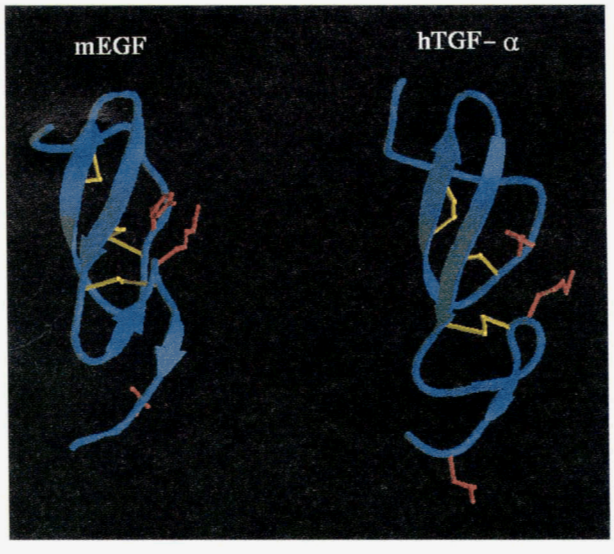

Fig. 5. RIBBONS (Carson, 1991) diagrams of the backbone structures (blue) of mEGF at pH 3.1 (left) and hTGFα at pH 6.5 (right), showing locations of disulfide bonds (yellow) and of side chains (red) of residues Tyr-13, Arg-41, and Leu-47 in mEGF or Phe-15, Arg-42, and Leu-48 hTGFα, which are critical for receptor-binding activities (Groenen et al., 1994).

feature of the comparison shown in Figure 5 is the relative orientation of the C-terminal subdomain with respect to the β-sheet of the N-terminal subdomain. The relative orientations of these two subdomains differ by about 45° between these two structures. This difference arises from the presence of several intersubdomain NOEs in the spectra of mEGF that are not observed in the spectra of hTGFα (Moy et al., 1993), and is attributed to attenuation of some NOEs in hTGFα at pH 6.5 due to slow hinge-bending motions and chemical-exchange line broadening of resonances at the subdomain interface (Moy et al., 1993; Li & Montelione, 1995). Although many intersubdomain NOEs are observed for hTGFα at pH 6.5, at pH 3.5 no intersubdomain NOEs are observed at all despite the fact that the structures of the individual subdomains are well defined by the NMR data (Moy et al., 1993). Intersubdomain motions that are already present at pH 6.5 (Li & Montelione, 1995) appear to be enhanced at acidic pH due to titration of crucial intersubdomain hydrogen bonds between residues His-19 and His-35 (Tappin et al., 1989; Moy et al., 1993).

Comparisons with previously published structures of EGF and TGFα

Differences in relative subdomain orientations in EGF-like molecules are further documented in Figure 6, in which the backbone structures of 10 EGF and EGF-like structures determined by solution NMR methods are superimposed with respect to the β-sheet of the N-terminal subdomain. This family of structures exhibits a wide range of relative orientations of the two subdomains, spanning from the orientation observed in our structures of mEGF to the orientation in our structure of hTGFα. Studies of proton linewidths (Montelione et al., 1992), nitrogen-15 (Li & Montelione, 1995) and carbon-13 (Celda et al., 1995) relaxation rates, molecular dynamics simulations (Fadel et al., 1995), and normal mode analysis (Ikura & Gō, 1993) for TGFα and/or EGF all suggest that, in any one EGF-like molecule, there are, in fact, multiple orientations of these two subdomains in dynamic equilibrium. Moreover, the intersubdomain NOEs that are observed in these systems are generally quite weak, and probably involve conformational averaging. Accordingly, the density of intersubdomain NOE data, and therefore the apparent orientations of the subdomains for any one EGF-like molecule, will depend critically on the amplitude and frequencies of these motions under the particular conditions of measurement.

Fig. 6. Superpositions of backbone structures reported here and in the Brookhaven Protein Data Bank for some EGF and EGF-like molecules. The backbone atoms of the core β-sheet in the N-terminal subdomains are superimposed, allowing analysis of the distribution of relative orientations between the N- and C-terminal subdomains. The structures shown are: white, mEGF at pH 3.1 (protocol v from this paper); yellow, mEGF at pH 3.1 (3EGF; Montelione et al., 1992); pink, mEGF at pH 2.0 (1EPG; Kohda & Inagaki, 1992a); red, mEGF at pH 6.8 (1EPI; Kohda & Inagaki, 1992a); light blue, hTGFα at pH 6.5 (protocol viii from this paper); green, hTGFα at pH 6.5 (2TGF; Harvey et al., 1991); dark blue, hTGFα at pH 6.3 (4TGF, Kline & Mueller; 1992); brown, Ca^{+2}-free form of the N-terminal EGF-like domain of bovine blood coagulation factor X at pH 5.8 (1APO; Ullner et al., 1992); purple, Ca^{+2}-bound form of the N-terminal EGF-like domain of bovine blood coagulation factor X at pH 5.8 (1CCF; Selander-Sunnerhagen et al., 1992); orange, first EGF-like domain of human blood clotting factor IX at pH 4.5 (1IXA; Baron et al., 1992).

Conclusions

The results in this paper show that the program CONGEN can be used successfully in structure determination and/or refinement using real NMR data. In order to minimize conformational pinning effects due to inconsistencies within the set of experimentally derived constraints or between these constraints and the conformational energy function, we have explored the use of upper-bound constraint loosening combined with high weights on further violations of the experimental data. The resulting structures are generally more relaxed in the force field, but also exhibit higher RMSD values. The requirement for constraint loosening in the EGF/TGFα system appears to reflect combined effects of conformational averaging of NMR parameters by internal motions, inaccuracies in the potential energy function, and inaccuracies in the interpretation of NMR data as conformational constraints, all of which can contribute to conformational pinning and unrealistically small values of RMSDs. In comparing different protocols for simulated annealing, we also observed that somewhat better results were obtained using extended starting conformations with random initial velocities rather than using previously calculated three-dimensional structures. This protocol also has the advantage of avoiding an initial structure determination prior to the SAMD calculations, except insofar as it may be useful to generate one family of structures with minimal assumptions about the potential energy function in addition to the energy-refined structures computed with CONGEN. These conclusions provide the basis for using "hard," 5–10% relax, extended protocols in our future NMR structure determinations with CONGEN.

Materials and methods

Experimental NMR data

For mEGF, the experimental constraints were derived primarily from four homonuclear NOESY data sets (mixing times of 65 ms, 100 ms, 200 ms, and 250 ms) in H_2O and D_2O at pH 3.1 and temperature of 28 °C (Montelione et al., 1992). The NOESY spectrum recorded with a mixing time of 200 ms was used only to verify the identities of weak cross peaks in the 65-ms NOESY spectrum in H_2O. For hTGFα, internuclear distance constraint

lists were derived from five homonuclear and two ^{15}N-edited NOESY spectra at pH 6.5 and temperature of 30 °C (Moy et al., 1993). The ^{15}N-edited 2D HSQC-NOESY and 3D NOESY-HSQC spectra with mixing times of 80 ms were recorded to resolve NOEs that were overlapped in the homonuclear NOESY spectrum. From these experiments, internuclear distances were estimated using standard calibration procedures (see for example, Montelione et al., 1992). Stereospecific assignments of some methylene and isopropyl methyl groups were determined as explained previously (Montelione et al., 1992; Moy et al., 1993). The resulting distance list was filtered to remove entries that did not actually constrain at least one dihedral angle, and upper-bound constraint lists for CONGEN were constructed as explained in the next section. Vicinal ^{3}J(H$^\alpha$-H^N) and ^{3}J(H$^\alpha$-H$^\beta$) coupling constants were also used, together with NOE data to determine upper- and lower-bound constraints on dihedral angles ϕ and χ_1 as outlined previously (Montelione et al., 1992; Moy et al., 1993).

Conversion of experimental NOE data into CONGEN distance constraints

Even if spin diffusion effects are minimized or explicitly accounted for, the NOESY cross peak intensities do not generally correspond to a simple interaction between two protons. For example, methyl groups, some methylene groups, and symmetry-related sites on aromatic rings generally exhibit degenerate proton resonance frequencies, resulting in NOESY cross peaks with contributions from two or more NOE interactions. For this reason, the NMR constraint term implemented in CONGEN uses an effective distance for the constraint, defined either as the average effective distance:

$$r_{ij,eff}^{av} = \left\{ \frac{1}{N_i N_j} \sum_{k=1}^{n} \frac{1}{r_{ij,k}^6} \right\}^{-1/6}, \qquad (1A)$$

or as a summed effective distance:

$$r_{ij,eff}^{sum} = \left\{ \sum_{k=1}^{n} \frac{1}{r_{ij,k}^6} \right\}^{-1/6}, \qquad (1B)$$

where r_{ij} is the distance from proton j in one group to proton i in the other, N_i and N_j are the total number of protons in each interacting group, and $n = N_i * N_j$. The use of an effective distance, $r_{ij,eff}$, is appropriate when the protons i contributing one frequency to the $i-j$ NOESY crosspeak interconvert rapidly on the chemical-shift timescale (e.g., methyl protons or Hδ and Hε protons of rapidly rotating aromatic rings) or when the several protons of a single methylene or isopropyl methyl group have accidentally degenerate resonance frequencies.

Our procedures for interpreting distance data derived from NOE measurements as upper-bound distance constraints for CONGEN are summarized for the following specific cases. The choice of using the averaged (Equation 1A) or summed (Equation 1B) definitions depends on how the information is extracted from the NMR spectra; CONGEN allows for both methods. In the work described in this paper, the summed definition was used both in extracting distance data from the NMR spectra and in the CONGEN calculations. In all the cases, it is assumed that the distance is between the group of protons discussed and one reference proton.

Methyl groups

Methyl groups have three equivalent protons that interconvert very rapidly on the NMR timescale. Six common amino acid residues have methyl groups: Ala, Met, Ile, Thr, Val, and Leu. Constraints involving these methyl groups are handled in the input to CONGEN as follows.

Case A. Methyl groups of Ala, Met, Ile, and/or Thr residues. The full NOE intensity (or volume) is used together with the sum definition (Equation 1B) for r_{eff} to compute an effective distance. Composite atom types are defined in the CONGEN constraint list as Ala HB*, Met HE*, Ile or Thr HG2*, and Ile HD*.

Case B. Isopropyl methyl groups of Val and/or Leu residues. Four subcases can be distinguished.

B1. NOE's to both methyls are observed and stereospecific assignments for these isopropyl methyl groups are available. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff} to compute an effective distance. Composite atom types are defined in the CONGEN constraint list as Val HG1*, HG2*, and Leu HD1*, HD2*, where the stereospecific labels 1 and 2 follow standard IUPAC nomenclature.

B2. NOE's to both methyls are observed, but no stereospecific assignments are available. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. The larger of the two r_{eff} values calibrated for each methyl group is assigned to both methyl groups. Composite atom types are defined in the CONGEN constraint list as Val HG1*, HG2*, and Leu HD1*, HD2*.

B3. Degenerate isopropyl methyl groups. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. Composite atom types are defined in the CONGEN constraint list as Val HG* and Leu HD*.

B4. The two isopropyl methyl groups are known to be non-degenerate, but only one of these exhibits a resolved NOESY cross peak (e.g., the second cross peak is overlapped with other cross peaks). The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. Composite atom types are defined in the CONGEN constraint list as Val HG* and Leu HD*.

Methylene protons

Most amino acid residues have one or more methylene group(s). Four specific cases can be distinguished.

Case A. Stereospecific methylene proton assignments are available. The full NOE intensity is used and the uniquely assigned atom types are defined in the CONGEN constraint list as HB1, HB2, HG1, HG2, HD1, HD2, HE1, HE2, etc. These correspond to HB2, HB3, HG2, HG3, etc. designators, respectively, in standard IUPAC notation.

Case B. No stereospecific assignments are available, but NOE's are observed to both methylene protons. The weakest of the two NOE intensities is used to calculate an upper-bound distance to both sites, and this same upper-bound constraint is assigned to both the HB1 and HB2 (or HG1 and HG2, etc.) constraints.

Case C. The methylene protons are degenerate. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. Composite atom types are defined in the CONGEN

constraint list as HB*, HG*, HD*, HE*, or HA* (Gly), depending on the assignment of the methylene group.

Case D. The two methylene protons are known to be nondegenerate, but only one of these exhibits a resolved NOESY cross peak (e.g., the second cross peak is overlapped with other cross peaks). The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. Composite atom types are defined in the CONGEN constraint list as HB*, HG*, HD*, HE*, or HA* (GLY), depending on the assignment of the methylene group.

Degenerate protons of aromatic rings

Two different residues have potentially degenerate aromatic protons: Tyr and Phe. Three cases can be distinguished.

Case A. Hδ's (or Hϵ's) are degenerate. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. Composite atom types are defined in the CONGEN constraint list as HD* or HE*.

Case B. All four aromatic Hδ's and Hϵ's are degenerate. The full NOE intensity is used together with the sum definition (Equation 1B) for r_{eff}. The composite atom type is defined in the CONGEN constraint list as (HD* AND HE*), selecting the four atoms for the sum.

Case C. Slow chemical exchange results in four separate resonances. The full NOE intensity is used for each one. Uniquely assigned atom types are defined in the CONGEN constraint list as HD1, HD2, HE1, and HE2.

Side-chain amide and guanido protons

Two residues have side-chain amide protons: Asn and Gln. The sites *cis* and *trans* with respect to the side-chain carbonyl oxygen are designated HD22 (HG22) and HD21 (HG21) for Asn (and Gln), respectively. These generally can be uniquely assigned from the relative intensities of NOEs to C$^{\beta}$H or C$^{\gamma}$H protons (Montelione et al., 1984, 1992). In the unusual case of degenerate side-chain amide protons, the composite atom is defined in the CONGEN constraint list as HD2* or HG2*. Similar composite atom types can be defined for uniquely assigned or degenerate arginine guanido protons.

As a final step in NMR constraint generation for the CONGEN calculation, a variable relaxation distance was applied to these upper-bound distance constraints. Lower bounds were generated as the sum of the van der Waals radii (1.8 Å), except for disulfide and hydrogen bonds, which used standard values of upper- and lower-bound constraint values (Wüthrich, 1986).

Simulated annealing with restrained molecular dynamics

Structure refinements were performed using simulated annealing with the molecular dynamics routines implemented in CONGEN (Brooks et al., 1983; Bruccoleri & Karplus, 1987). All calculations were run on Silicon Graphics 4D/240 or Indigo 2 computer workstations. Nonbonded interactions were included up to a cutoff of 10.0 Å and the list updated every 20 steps. Electrostatic effects were included and employed a distance-dependent dielectric equal to the atom-atom distance. Integration of the equations of motion was performed using a Verlet integration algorithm (Verlet, 1967) with initial velocities randomly assigned according to a Maxwell-Boltzmann distribution corresponding to the temperature of the simulation. The time step of the integrator was 1 fs in all the stages.

The SAMD protocol begins with adopted basis Newton-Raphson (ABNR) energy minimization (Brooks et al., 1983) of the starting coordinates (from a previous DISMAN calculation or from a fully extended chain), in order to avoid instabilities in the following dynamic simulation. Random velocities are then assigned to each atom at a temperature of 1,000 K and a restrained MD simulation is carried out at 1,000 K. Over this 8-40-ps trajectory, the weight for the NOE-penalty function, K_{NOE}, is increased from 1.0 (for extended starting structures) or 10.0 (for DISMAN starting structures) kcal mol^{-1} Å^{-2} to 20.0 (for **soft** protocols) or 100.0 (for **hard** protocols) kcal mol^{-1} Å^{-2}. During this "weight annealing" process, the weights on experimental distance constraints (i.e., the NOE, disulfide bond, and hydrogen bond restraints) are increased while maintaining constant values on most of the conformational energy terms. In order to maintain correct covalent geometry at 1,000 K, the force constants for bond length and bond angle deviations are increased to $K_{bl} = 600.0$ kcal mol^{-1} Å^{-2} and $K_{ba} = 500.0$ kcal mol^{-1} rad^{-2}, respectively. In addition, the peptide bond dihedral angles ω are restrained to planar conformations ($K_{dih} = 500.0$ kcal mol^{-1} rad^{-2}) in order to prevent *cis/trans* peptide bond flips, and improper constraints on peptide bonds are increased to $K_{inp} = 500.0$ kcal mol^{-1} rad^{-2} to maintain planarity of the peptide bonds. An additional constraint file is used to maintain correct chirality of backbone Cα sites and side-chain Cβ sites of Ile and Thr. The effect of weight annealing is to selectively "cool" the experimental constraint terms of the potential function while maintaining a highly fluid potential energy surface. Once the weight K_{NOE} reaches its maximum value (i.e., 20.0 or 100.0 kcal mol^{-1} Å^{-2}), dihedral-angle and disulfide-bond topology constraints are added to the composite penalty function, and a 3-ps restrained SAMD trajectory is computed at 1,000 K.

The details of the weight annealing protocol were somewhat different depending on the starting conformation. In runs beginning with a previously computed DISMAN structure, we began with weight $K_{NOE} = 10.0$ kcal mol^{-1} Å^{-2}. The values of F_{max} and F_{slope} that define the flexible NOE penalty function (Bassolino-Klimas et al., 1996) are then increased from 10.0 kcal mol^{-1} Å^{-2} and 0.50 kcal mol^{-1} Å^{-2}, respectively, to 240.0 kcal mol^{-1} Å^{-2} and 40.0 kcal mol^{-1} Å^{-2} during 8 ps of dynamics. On the other hand, when starting from a fully extended chain, we first created the local structure during 40 ps of dynamics in which K_{NOE} is set to 1.0 kcal mol^{-1} Å^{-2} and F_{max} and F_{slope} are increased from 1.0 kcal mol^{-1} Å^{-2} and 0.01 kcal mol^{-1} Å^{-2} to 6.0 kcal mol^{-1} Å^{-2} and 0.05 kcal mol^{-1} Å^{-2}, respectively. Small values of F_{max} and F_{slope} result in small forces on large distance violations, effectively enhancing the effects of small distance violations in driving structure formation. Next, these local structural elements are folded into tertiary structures by increasing K_{NOE} from 5.0 kcal mol^{-1} Å^{-2} up to its final value of 100.0 kcal mol^{-1} Å^{-2} over 18 ps of restrained molecular dynamics, whereas the F_{max} and F_{slope} vary from 30.0 kcal mol^{-1} Å^{-2} and 1.0 kcal mol^{-1} Å^{-2} to 600.0 kcal mol^{-1} Å^{-2} and 30.0 kcal mol^{-1} Å^{-2}. These large values allow incorporation of the whole set of constraints through the subsequent thermal annealing procedure.

In these high-temperature restrained molecular dynamics calculations, it was necessary to employ velocity rescaling (Bassolino-Klimas et al., 1996) in order to avoid failures in the numerical

integration. This is achieved using the TLIMIT option in CONGEN. The limit was chosen to correspond to an atomic temperature five times the temperature of the simulation. This option proved to be invaluable when starting from fully extended chains because it prevents atoms from moving too fast in one molecular dynamics step. This usually tends to happen at the beginning of the simulation because the forces on each atom can be quite large due to the fully extended chain conformation.

Following the weight annealing procedure, simulated thermal annealing was carried out by slowly cooling the system to 300 K. In order to favor low-temperature conformations, the cooling protocol is faster at the beginning than at the end. Initially, the temperature T was decreased in 150-K steps following each 1,000 steps of SAMD, whereas toward the end of the thermal annealing, T was decreased in steps as small as 5 K for each 1,000 steps of molecular dynamics. At this point, standard CHARMM energy parameters (Brooks et al., 1983) for bond length and bond angle force constants are reloaded and the system is equilibrated at 300 K for 5 ps. A simulation of the system at 300 K was carried out for an additional 10 ps while writing coordinates to the computer disk each 100 steps. For each SAMD trajectory, an average structure was then calculated from the coordinates stored in the last 3 ps. This average structure was then energy-minimized with constraints using the indicated NOE weights (20 or 100 kcal mol^{-1} $Å^{-2}$) for a maximum of 4,000 steps (or until a local minimum was reached) using the ABNR minimizer. The final coordinates were then saved on the computer disk for further analysis.

Statistical analysis of the structures

Statistical analysis on each set of structures was performed using the ANALYSIS facility in CONGEN and using the computer program PDBSTAT (R. Tejero & G. T. Montelione, unpubl. software). These computer programs are available from the authors. Dihedral-angle order parameters $S(\theta)$ were computed as defined by Hyberts et al. (1992). The structures computed in this study, together with the corresponding constraint files, will be deposited in the Brookhaven Protein Data Bank.

Acknowledgments

We gratefully acknowledge B. Celda, M. Tashiro, and L. Snyder for helpful discussions and comments on the manuscript. This work was supported by The National Institutes of Health (GM-47014), The National Science Foundation (MCB-93557526), a National Science Foundation Young Investigator Award (to G.T.M.), and a Dreyfus Teacher-Scholar Award (to G.T.M.). R.T. was also supported in part by a fellowship from Pharmacia, A.B., Stockholm, Sweden.

References

Baron N, Norman DG, Harvey TS, Handford PA, Mayhew M, Tse AGD, Brownlee GG, Campbell ID. 1992. The three-dimensional structure of the first EGF-like module of human factor IX: Comparison with EGF and TGF-alpha. *Protein Sci 1*:81–90.

Bassolino-Klimas D, Tejero R, Krystek SR, Metzler WJ, Montelione GT, Bruccoleri RE. 1996. Simulated annealing with restrained molecular dynamics using a flexible restraint potential: Theory and evaluation with simulated NMR constraints. *Protein Sci 5*:593–603.

Braun W, Gō N. 1985. Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm. *J Mol Biol 186*:611–626.

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

Bruccoleri RE, Karplus M. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers 26*:137–168.

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

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

Carpenter G, Stoscheck CM, Preston YA, DeLarco JE. 1983. Antibodies to the epidermal growth factor receptor block the biological activities of sarcoma growth factor. *Proc Natl Acad Sci USA 80*:5627–5630.

Carson M. 1991. RIBBONS 2.0. *J Appl Cryst 24*:958–961.

Celda B, Biamonti C, Arnau MJ, Tejero R, Montelione GT. 1995. Combined use of ^{13}C chemical shift and $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ heteronuclear NOE data in monitoring a protein NMR structure refinement. *J Biomol NMR 5*:161–172.

Cohen S. 1962. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J Biol Chem 237*:1555–1562.

Constantine KL, Friedrichs MS, Mueller L, Bruccoleri RE. 1995a. J-coupling restraint potentials for non-stereospecifically assigned methylene protons and ensemble average calculations. *J Magn Reson Ser B 108*:176–184.

Constantine KL, Mueller L, Anderson NH, Tong H, Wandler CF, Friedrichs MS, Bruccoleri RE. 1995b. Structural and dynamical properties of a β-hairpin-forming linear peptide. 1. Modeling using ensemble-averaged constraints. *J Am Chem Soc 117*:10841–10854.

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

Fadel AR, Jin DQ, Montelione GT, Levy RM. 1995. Crankshaft motions of the polypeptide backbone in molecular dynamics simulations of human type-α transforming growth factor. *J Biomol NMR 6*:221–226.

Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, Presky DH, Familletti PC, Wolitzky BA, Burns DK. 1994. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature 367*:532–538.

Groenen LC, Nice EC, Burgess AW. 1994. Structure-function relationships for the EGF/TGFα family of mitogens. *Growth Factors 11*:235–237.

Guerin M, Gabillot M, Mathieu MC, Travagli JP, Spielmenn M, Andrieu N, Riou G. 1989. Structure and expression of c-erbB-2 and EGF receptor genes in inflammatory and non-inflammatory breast cancer: Prognostic significance. *Int J Cancer 43*:201–208.

Hansen AP, Petros AM, Meadows RP, Nettesheim DG, Mazar AP, Olejniczak ET, Xu RX, Pederson TM, Henkin J, Fesik SW. 1994. Solution structure of the amino-terminal fragment of urokinase-type plasminogen activator. *Biochemistry 33*:4847–4864.

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

Hendrickson WA, Teeter MM. 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. *Nature 290*:107–113.

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

Huang LH, Cheng H, Pardi A, Tam JP, Sweeney WV. 1991. Sequence-specific ^{1}H NMR assignments, secondary structure, and location of the calcium binding site in the first epidermal growth factor like domain of blood coagulation factor IX. *Biochemistry 30*:7402–7409.

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

Ikura T, Gō N. 1993. Normal mode analysis of mouse epidermal growth factor: Characterization of the harmonic motion. *Proteins Struct Funct Genet 16*:423–436.

Kline TP, Brown FK, Brown SC, Jeffs PW, Kopple KD, Mueller L. 1990. Solution structures of human transforming growth factor alpha derived from ^{1}H NMR data. *Biochemistry 29*:7805–7813.

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

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

Li YC, Montelione GT. 1995. Human type-α transforming growth factor undergoes slow conformational exchange between multiple backbone con-

formations as characterized by nitrogen-15 relaxation measurements. *Biochemistry 34*:2408–2423.

Montelione GT, Arnold E, Meinwald YC, Stimson ER, Denton JB, Huang SG, Clardy J, Scheraga HA. 1984. Chain-folding initiation structures in ribonuclease A: Conformational analysis of *trans* Ac-Asn-Pro-Tyr-NHMe and Ac-Tyr-Pro-Asn-NHMe in water and in the solid state. *J Am Chem Soc 106*:7946–7958.

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

Montelione GT, Wüthrich K, Nice EC, Burgess AW, Scheraga HA. 1986. Identification of two anti-parallel beta-sheets structures in the solution conformation of murine epidermal growth factor by ^{1}H-NMR. *Proc Natl Acad Sci USA* 83:8594–8598.

Montelione GT, Wüthrich K, Nice EC, Burgess AW, Scheraga HA. 1987. Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. *Proc Natl Acad Sci USA 84*:5226–5230.

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

Moy FJ, Li YC, Rauenbuehler P, Winkler ME, Scheraga HA, Montelione GT. 1993. Solution structure refinement of human type-α transforming growth factor determined by heteronuclear NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry 32*:7334–7353.

Némethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HA. 1992. Energy parameters in polypeptides. 10: Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with applications to proline-containing peptides. *J Phys Chem 96*:6472–6484.

Némethy G, Pottle MS, Scheraga HA. 1983. Energy parameters in peptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J Phys Chem 87*:1883–1887.

Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W, Huber R, Blankenship DT, Cardin AD, Kisiel W. 1993. Structure of human des(1-45) factor Xa at 2.2 Å resolution. *J Mol Biol 232*:947–966.

Selander-Sunnerhagen M, Ullner M, Persson M, Teleman O, Stenflo J, Drakenberg T. 1992. How an epidermal growth factor (EGF)-like domain binds calcium: High resolution NMR structure of the calcium form of the amino-terminal EGF-like domain in coagulation factor X. *J Biol Chem 267*:19642–19649.

Simpson RJ, Smith JA, Moritz RL, O'Hare MJ, Rudland PS, Morrison JR, Lloyd CJ, Grego B, Burgess AW, Nice EC. 1985. Rat epidermal growth factor: Complete amino acid sequence. Homology with the corresponding murine and human proteins; isolation of a form truncated at both ends with full in vitro biological activity. *Eur J Biochem 153*: 629–637.

Smith BO, Downing AK, Dudgeon TJ, Cunningham M, Driscoll PC, Campbell ID. 1994. Secondary structure of fibronectin type 1 and epidermal growth factor modules from tissue-type plasminogen activator by nuclear magnetic resonance. *Biochemistry 33*:2422–2429.

Tappin MJ, Cooke RM, Fitton JE, Campbell ID. 1989. A high-resolution ^{1}H-NMR study of human transforming growth factor alpha. Structure and pH-dependent conformational interconversion. *Eur J Biochem 179*: 629–637.

Ullner M, Selander M, Persson E, Stenflo J, Drakenberg T, Teleman O. 1992. Three-dimensional structure of the apo form of the N-terminal EGF-like module of blood coagulation factor X as determined by NMR spectroscopy and simulated folding. *Biochemistry 31*:5974–5983.

Verlet L. 1967. Computer experiments on classical fluids. I. Thermodynamic properties of Lennard–Jones molecules. *Phys Rev 159*:98–103.

Weis WI. 1994. Lectins on a roll: The structure of E-selectin. *Structure 2*: 147–150.

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