The Solution Structure of Motilin from NMR Distance Constraints, Distance Geometry, Molecular Dynamics, and an Iterative Full Relaxation Matrix Refinement†

S. Edmondson, N. Khan, and J. Shriver*
Department of Medical Biochemistry, School of Medicine, and Department of Chemistry and Biochemistry, College of Science, Southern Illinois University, Carbondale, Illinois 62901
J. Zdunek and A. Gräslund*
Department of Medical Biochemistry and Biophysics, University of Umeå, S-901 87 Umeå, Sweden
Received April 4, 1991; Revised Manuscript Received August 13, 1991

ABSTRACT: A model of the structure of the 22 amino acid residue gastrointestinal peptide hormone motilin in 30% hexafluoro-2-propanol has been obtained by using distance constraints obtained from two-dimensional nuclear Overhauser enhancements. A set of initial structures have been generated by using the distance geometry program DIANA, and 10 of these structures have been refined by using restrained molecular dynamics (AMBER). The resulting structures are virtually indistinguishable in terms of constraint violations and energies and display less than 0.5-Å root mean square deviations (RMSD) of the backbone atom positions from Tyr7 to Lys20. A comparison of back-calculated and experimental NOE intensities indicates that RMSD's are not the best indicators of the goodness of fit or of the precision with which the structure is defined. The structure was further refined by fitting the experimental NOE data using an iterative full relaxation matrix analysis. The mean error between the observed and calculated backbone NOE intensities for the final refined structure was 0.23 for the full length of the molecule, 0.18 for the region from Glu9 to Lys20, and 0.29 for the region from Phe1 to Gly8. *R* factors for the same regions were 0.27, 0.19, and 0.43, respectively. All of the NOE-determined structures consistently display an α-helix which extends from Glu9 to Lys20. Considerable lack of definition of structure exists at the amino and carboxyl ends of the molecule and also in the vicinity of Thr6–Tyr7–Gly8. A tendency to form a wide turn appears to exist over the sequence Pro3–Ile4–Phe5–Thr6, but the structure in this region is not well defined by the NOE data.

Motilin is a small peptide hormone of 22 amino acid residues found in the gut and possibly brain tissue (Itoh, 1990). Its primary function in the gut appears to be to regulate the migrating motor complex (MMC), a wave of contractions which occurs between digestive periods and probably functions to cleanse the gut during periods of low activity (Vantrappen & Peeters, 1989). Its role in the brain is not known. Receptors of motilin in the gut have been shown to be membrane bound and are localized in smooth muscle in the gastrointestinal tract (Bormans et al., 1986; Peeters et al., 1988b, 1990). Itoh, Peeters, and co-workers have shown that the antibiotic erythromycin, a 14-membered macrolide, has motilin-like activity and can displace motilin from its receptor (Depoortere et al., 1988; Inatomi et al., 1989; Itoh et al., 1984, 1985; Kondo et al., 1988; Peeters et al., 1989; Satoh et al., 1990).

Both human and porcine motilin consist of a single peptide chain with the sequence FVPIFTYGELQRMQEKERNKGQ (Brown et al., 1973; Schubert & Brown, 1974). The sequence appears to be unrelated to any other gut peptide hormone. No crystal structure has been reported for motilin. The N-terminal end of the peptide chain is strongly hydrophobic, while the Tyr7–Gln22 segment is hydrophilic. Studies on synthetic fragments of motilin, namely, 6–22, 7–22, and 9–22, indicate that the amino terminus is required for activity (Segawa et al., 1976; Yajima et al., 1977). However, fragment 1–6 is inactive, although fragment 1–5 has some ability to compete with intact hormone for receptor sites (Peeters et al., 1988a). It has been argued that Phe1 is important since substitution with either Lys or Ser results in almost complete loss of activity (Kitigawa et al., 1985; Peeters et al., 1988a). Methionine 13 has been substituted with norleucine with no effect on activity (Strunz et al., 1975, 1976). Substitution of Gln14 with glutamate also had no effect (Strunz et al., 1976). Comparative studies of canine and porcine motilin have been of interest since both function in in vitro assays using rabbit muscle. The two differ at five positions in the middle of the molecule: at position 7 (His in canine and Tyr in porcine), position 8 (Ser in canine and Gly in porcine), position 12 (Lys in canine and Arg in porcine), position 13 (Leu in canine and Met in porcine), and position 15 (Arg in canine and Gln in porcine) (Poitras et al., 1983). Despite these differences, both molecules are equally potent in the rabbit (Peeters et al., 1986), but only canine motilin is able to stimulate canine tissue in vitro (Poitras et al, 1987). One study seems to indicate that the C-terminus is also important for activity (Ueda et al., 1977).

We have previously reported the results from an NMR[1] study on porcine motilin where sequence specific ^{1}H NMR assignments were given and initial delineation of the secondary structure was described (Khan et al., 1990). A qualitative analysis of the results indicated that in an aqueous solvent

†This work was supported by a research grant from the Swedish Natural Science Research Council, Magn. Bergwalls Foundation, and Southern Illinois University School of Medicine. J.W.S. is a recipient of a National Institutes of Health Research Career Development Award (AR01788) and a Visiting Scientist's Fellowship from the Swedish Natural Science Research Council.
* Authors to whom correspondence should be directed.

[1] Abbreviations: DG, distance geometry; FIRM, iterative full relaxation matrix refinement program; HFP, hexafluoro-2-propanol; MD, molecular dynamics; MM, molecular mechanics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy.

0006-2960/91/0430-11271$02.50/0

containing 30% hexafluoro-2-propanol, residues 9–20 were folded into an α-helix. No particular secondary structure could be assigned to the N-terminal end of the oligopeptide other than an indication of a reverse turn involving residues 2–5. The present study describes a quantitative analysis of the NOE cross-peak intensities using distance geometry calculations, restrained molecular dynamics, NOE back-calculations, and an iterative refinement using a full relaxation matrix treatment.

Materials and Methods

Two-dimensional NOESY spectra of motilin were collected on a Varian VXR-500 instrument, and representative spectra and sequential assignments have been presented elsewhere (Khan et al., 1990). The data were collected at 21 °C in 30% hexafluoro-2-propanol (HFP) to stabilize secondary structure (Khan et al., 1990). Due to limited quantities of peptide and the appearance of additional NMR resonances in the spectra after a few days exposure of motilin to HFP, only a single NOE experiment was available for this analysis (i.e., NOE buildup rates have not been determined). The small size of the peptide (2700 Da) and the relatively large line widths observed in HFP (see below) required the use of a 450-ms mixing time to obtain correlation peak intensities of sufficient magnitude to integrate reliably. The error introduced by using a single NOESY spectrum with a long mixing time to obtain initial distance constraints has been estimated to be about 0.5 Å from simulations using a full relaxation matrix analysis of various possible spin configurations with distances appropriate for random coil and helical secondary structure (Khan et al., 1990). In the final step of the analysis reported here, the effects of spin diffusion are included by using a full relaxation matrix refinement of the data. A correlation time of 2 ns was estimated by using the relative intensities of selected NOESY cross-peaks and corresponding diagonal peaks [see Khan et al. (1990) for more details].

NOE cross-peak volume integrals were evaluated by using the integration software routine supplied by Varian with a rectangular box delimiting the peak of interest. Identical box dimensions were used for all integrals. The effects of imperfections in the baseline of the NOESY spectra were removed by integrating neighboring regions adjacent to the peak which contained only noise.

Interproton Distance Constraints. For generation of initial structures, cross-peak intensities were converted to proton–proton distances by using a two-spin approximation. Errors in the measured distances due to spin diffusion was included in the distance constraints by increasing each upper bound by 0.5 Å (see above). Due to the large line widths observed in 30% HFP, no stereospecific assignments were possible, so that pseudoatoms were used at all pro-chiral centers, on all aromatic rings, and for all methyl groups. The maximum possible error due to the introduction of pseudoatoms was incorporated into the constraints by adding additional correction terms as described by Wüthrich et al. (1983) with modifications as indicated in the DIANA manual. A total of 292 NOE cross-peaks were integrated to obtain upper limit distance constraints, of which 180 were determined to be useful by DIANA. Of these DIANA constraints, 86 were interresidue constraints and 36 were $i,i+2$ or longer interactions. These constraints were modified slightly for AMBER after closer scrutiny of the NOESY data (three were deleted) and to make more efficient use of pseudoatoms. A total of 135 distance constraints were used for AMBER restrained dynamics calculations, of which 73 were interresidue and 29 were $i,i+2$ or greater constraints. A lower limit distance constraint of 1.5 Å was also used for all proton pairs.

Distance Geometry Calculations. Distance geometry calculations were performed on a Silicon Graphics 4D25G computer using the program DIANA (Güntert et al., 1991), which minimizes a target function by varying dihedral angles. Due to the lack of three-bond coupling constants, only upper and lower limits of proton–proton distances were used as input. The minimization parameters used were those supplied in the DIANA manual. Essentially this allowed for nine stages of minimization, with each stage comprised of 100 iterations. Initial stages were devoted to minimization using only local constraints, with each stage allowing increasing use of longer sequential interactions. A total of 100 structures were generated using 100 different random starting structures. Of these, 10 were chosen for further study: namely, the five "best" structures (i.e., the five with the lowest final target function) and five additional structures chosen at random from the remaining 95 structures.

Energy Minimization of the DIANA *Structures.* Energy minimization of the distance geometry structures was performed with and without NMR constraints by using CHARMM with the Powell minimization algorithm (Brooks et al., 1983). The NOE constraints were incorporated into the total potential function using the following additional terms:

$$E(r) = K_1(r - r_o)^2 \qquad r < r_o$$

$$E(r) = K_2(r - r_o)^2 \qquad r > r_o$$

with $K_1 = S\,(kT/2)/DL^2$ and $K_2 = S(kT/2)DH^2$, where DL and DH are the deviations (below and above the constraint distance r_o) at which the energy is $kT/2$. The scale factor S was typically set to 3.0, and the temperature was set to 300 K. The DH and DL values were chosen so that the force constants K_1 and K_2 were about 5 kcal/(mol·Å)^2 with S equal to 3.0.

Molecular Dynamics. Molecular dynamics was performed on the 10 selected structures generated by DIANA by using AMBER 3.0 rev. A (Weiner & Kollman, 1981; Weiner et al., 1984). The charges on the side chains were reduced to approximately 20% of the AMBER-defined values. This was accomplished by reducing the atomic partial charges of all side-chain atoms of Asp, Glu, Lys, Arg, and His residues. The united-atom data base was modified to allow explicit inclusion of pseudoatoms. Methyl (CH_3), methylene (CH_2), amine (NH_2), and propyl [the $C(CH_3)_2$ groups of Val and Leu] peudoatoms were placed at the geometric center of the protons they represented. The protons on all chiral carbons (including α-protons) were also represented by pseudoatoms in the united-atom potential. The pseudoatoms were given zero charge and no nonbonded interaction terms. The bond length, angle, and dihedral force constants for the pseudoatoms were chosen to keep the pseudoatoms at their proper places while minimizing their effects on the standard AMBER united-atom potential.

A flat-well potential (J. Metz and D. Gorenstein, personal communication) was incorporated into AMBER for describing the energy contributions of NOE distance violations:

$$E_{noe} = 1/2K_{noe}(r - r_u)^2 \qquad \text{for } r > r_u$$

$$E_{noe} = 1/2K_{noe}(r - r_l)^2 \qquad \text{for } r < r_l$$

$$E_{noe} = 0 \qquad \text{for } r_l < r < r_u$$

where r is the interproton distance, r_l is the lower distance constraint limit, r_u is the upper distance constraint limit, and K_{noe} is the force constant for the NOE violation.

All AMBER MM and MD calculations were performed with the pseudo/united-atom force field described above, a dis-

tance-dependent dielectric constant, and a nonbonded atom cutoff distance of 8 Å. The 1–4 van der Waals and 1–4 electrostatic interactions were scaled down by a factor of 2. The nonbonded pairs were updated after every 25 cycles/steps. All MD simulations were performed with SHAKE using a step size of 0.001 ps. Translational and rotational motions were removed every 50 steps, and a temperature bath coupling constant of 0.1 was used throughout.

The MD refinement strategy involved the following steps: (1) initial energy minimization of the distance geometry structures using AMBER, (2) a 20-ps MD run while gradually increasing the NOE force constant, and finally (3) minimizing the energy of the final MD structure. The first stage of MM/MD refinement consisted of 100 cycles of conjugate-gradient MM using the united/pseudo-atom potentials with no constraints, 1000 cycles of MM with K_{noe} = 1 kcal/Å^2, and then 500 cycles of steepest descent MM with K_{noe} = 1 kcal/Å^2 and with SHAKE. The resulting structure was then subjected to a total of 20 ps of restrained MD as outlined below. The energy of the final MD structure was then quenched by 1000 cycles of conjugate-gradient MM with K_{noe} = 100 kcal/Å^2. The energies of the various models reported under Results were computed with the standard AMBER united-atom force field (i.e., with no NOE constraints and full charges).

The MD simulation was started at an elevated temperature to facilitate better sampling over conformational space and then gradually "annealed" to lower temperatures and higher NOE force constants. The first 5 ps of restrained MD were performed at 800 K with K_{noe} = 1kcal/Å^2. This was followed by 1 ps of MD at 800 K at each of the following K_{noe} values: 2, 5, 10, 15, 25, 50, and 100 kcal/Å^2. The NOE force constant was then held constant at K_{noe} = 100 kcal/Å^2 for 1 ps at each of the following temperatures: 650, 550, 475, and 425 K. The simulation was continued for 1 ps each at temperatures of 375, 325, 300, and 300 K with corresponding K_{noe} values of 200, 300, 400, and 500 kcal/Å^2. The relatively high K_{noe} values used during the last 5 ps were used to place the highest emphasis on "fitting" the NMR data. Bond angle and dihedral angle energies continued to decrease during this time, indicating that large distortions were not introduced by the heavy emphasis on the NOE potential. The final MD structure was quenched by MM with K_{noe} = 100 kcal/Å^2. The energies of the resulting structures are reported without including the NOE contribution.

Calculated structures were visualized on a Silicon Graphics 4D25G graphics workstation using the programs HYDRA (Polygen Corp., Waltham, MA, and R. Hubbard, University of York, U.K.) and QUANTA (Polygen Corp., Waltham, MA).

NOESY Back-Calculations and Iterative Refinement. NOESY cross-peak and diagonal intensities were calculated for one of the MD-refined structures (structure 4 in Table II) by using a full relaxation matrix approach (Keepers & James, 1984). The relaxation matrix was defined by assuming isotropic rotation with no cross-correlation contribution to relaxation. A rotational correlation time of 2 ns was used throughout (Khan et al., 1990). No attempt was made to include position dependent rotational correlation times. The NOE intensity matrix at a given mixing time τ_m, $\mathbf{a}(\tau_m)$, was obtained as follows:

$$\mathbf{a}(\tau_m) = \chi \cdot e^{-\lambda \tau_m} \cdot \chi^{-1}$$

where χ and λ are the matrices of eigenvectors and eigenvalues, respectively, of the relaxation rate matrix **R**. The diagonal and off-diagonal elements of **R** represent the longitudinal and cross-relaxation rate constants, respectively, for each of the spins. The cross-relaxation rates of equivalent protons (methyl rotors, methylene groups, and aromatic protons) were averaged (using $1/r^6$ weighting) as described by Koning et al. (1990).

Structural refinement was accomplished by iteratively modifying the model relaxation rate matrix until the calculated and observed NOE's were in as close agreement as possible using the refinement program FIRM. FIRM has been written in FORTRAN 77 by one of us (S.E.) for use on a Silicon Graphics 4D25G computer. The FIRM refinement protocol is essentially identical to that defined by James and co-workers in MARDIGRAS (Borgias et al., 1990; Borgias & James, 1990) and by Gorenstein and co-workers in their hybrid relaxation matrix procedure (Nikonowicz et al., 1990; Post et al., 1990; Powers et al., 1990) [see also Boelens et al. (1989) and Gippert et al. (1990)]. Briefly, the full NOE intensity matrix was calculated for an initial starting structure (the final MD structure here), and well-defined backbone atom experimental NOE's were substituted ("merged") in place of the calculated values. An option to substitute a weighted average of the calculated and observed NOE's was included as follows:

$$\mathrm{NOE}_{merge} = (1 - x)\mathrm{NOE}_{cal} + x\mathrm{NOE}_{obs}$$

where the mixing coefficient was user defined. The merged NOE matrix was then transformed back to a corresponding relaxation rate matrix according to

$$\mathbf{R} = \chi\left(\frac{-\ln \lambda}{\tau_m}\right)\chi^{-1}$$

A new set of diagonal relaxation rates was calculated, and the resulting **R** was used to obtain a new NOE intensity matrix. Merging of the experimental NOE's was performed again, and the cycle continued until the error between the calculated and observed intensities reached a minimum. Distances were back-calculated by FIRM and used as constraints in an all-atom energy minimization using the AMBER harmonic potential with a force constant of 200 kcal/Å^2. The resulting structure was used for a second series of FIRM refinement and constrained energy minimization. A more detailed description of FIRM will be published elsewhere. Display of calculated NOESY spectra at any stage of the refinement was accomplished by using a program written in Mathematica on a Silicon Graphics IRIS 4D25G by one of us (S.E.).

An *R* factor (not to be confused with the relaxation matrix **R**) analogous to that used in X-ray crystallography was defined as follows (Baleja et al., 1990; Cantor & Schimmel, 1980; Nikonowicz et al., 1990):

$$R = \frac{\sum_{\mathrm{NOE's}} |\mathrm{NOE}_{obs} - \mathrm{NOE}_{cal}|}{\sum_{\mathrm{NOE's}} |\mathrm{NOE}_{obs}|}$$

The mean error between the observed and calculated NOE intensities was calculated as

$$\text{mean error} = \frac{1}{n}\sum_{\mathrm{NOE's}} \left|\frac{\mathrm{NOE}_{obs} - \mathrm{NOE}_{cal}}{\mathrm{NOE}_{cal}}\right|$$

where n is the total number of NOE's used in the sum.

Results and Discussion

For a short linear peptide such as motilin (MW 2700), the use of distance geometry and restrained molecular dynamics to define a solution structure gives rise to particular problems due to molecular flexibility or the existence of multiple conformers. A number of strategies have been used to investigate the existence of flexibility by NMR, including NOE distance patterns, and also agreement between NOE intensities and

conformations indicated by J coupling constants (Dyson et al., 1988; Kessler et al., 1988; Simorre et al., 1990; Smith et al., 1991; Wright et al., 1988). The use of coupling constant information is precluded in the present work because of the large line widths observed in HFP. However, a qualitative comparison of the NOE distance patterns is consistent with the existence of a persistent structure over a large portion of the sequence. We have therefore proceeded with determining a structure for motilin under the initial assumption that the NOE's are derived from a single structure. First, the two-spin approximation is used to derive distance constraints for DG and MD structure refinement. The structure is refined further with distances generated by FIRM in a multispin system, including the effects of spin diffusion. Good agreement between the calculated and experimentally observed NOE intensities is used as evidence in support of a single solution structure in defined regions of the molecule. The inability to fit the NOE data with a single structure implies that the NMR data represents an average derived from two or more structures, and the structure is flexible. No attempt is made here to quantitatively describe the indicated flexibility or the dynamics of the system.

A qualitative view of the NOE data clearly indicates that motilin folds into an α-helix that extends over more than 50% of its length in the presence of 30% HFP (Khan et al., 1990). The helical content determined by CD in the *absence* of fluorinated alcohol is about 20%, indicating that HFP stabilizes the secondary structure which is normally present in aqueous solution (Khan et al., 1990). The fact that the helical content as indicated by CD reaches a plateau at 15% HFP (Figure 1, Supplementary Material), and the agreement of the helical content obtained by a qualitative analysis of the NOE constraints with that obtained by CD indicates that predominantly a single folded form of the helical region of the hormone exists in solution under the conditions used here. The use of HFP allows the stabilization of the folded hormone and permits essentially complete NMR assignments and also the determination of NOE cross-peak intensities for solution structure determination.

For conversion of the NOE data to distances, a reference distance in the backbone in the putative helical region (namely NH–αCH of Glu9) was used since the side chains might be expected to be considerably more flexible than the backbone. In addition, we considered the most generally useful and reliable reference to be an intermediate distance (i.e., approximately 2.75 Å). For example, the use of the ring protons of tyrosine for a reference distance did not give physically meaningful distances between backbone protons, presumably due to significant differences in the rotational correlation times of the side chain and backbone. Indeed, two tyrosine rotational correlation times which differ by an order of magnitude have been observed in preliminary fluorescence polarization anisotropy decay measurements, probably corresponding to global motion of the whole hormone and local motions of the tyrosine (B. M. Backlund and A. Gräslund, to be published).

We have compared distances along the backbone over the region where the existence of an α-helix is supported by $i,i+3$ interactions. The intraresidue NH(i)–αCH(i) and interresidue αCH(i)–NH($i+1$) distances should be approximately 2.75 and 3.5 Å for an ideal α-helix, respectively. This is observed from residue Glu9 to residue Arg18. Unfortunately, a series of overlapping NOE peaks occurs in the center of the putative helical region making a complete investigation impossible by this criterion alone. Even so, the data indicate that the distances are reliable over most of the region expected to be in a helical conformation. Close to the C-terminus, the pattern of distances indicates increased flexibility, as one might expect.

Another validation of the distance constraints relies on the $i,i+3$ NOE's which indicate helix (Supplementary Material). A persistent pattern of these constraints clearly indicates α-helix, and the fact that the distances are close to those expected supports their reliability (flexibility would be expected to increase them relative to that expected for a rigid helix). It should also be noted that $i,i+3$ interactions overlay regions where sequential NOE's cannot be quantified due to overlap. As will be shown below, there is very good agreement between the observed NOE intensities for $i,i+3$ interactions and the calculated intensities using the refined structure, indicating that significant averaging due to flexibility does not appear to occur at least over the helical region spanned by residues 9–20.

Table I: Summary of the DIANA-Generated Structures Used for Further Refinement

structure[a]	target function	no. of violations[b]	max violation (Å)	sum of violations[b] (Å)
1	6.42	15	1.27	9.7
2	6.45	19	1.21	10.1
3	6.47	19	1.11	10.1
4	6.48	17	1.15	10.0
5	6.69	17	1.19	10.3
29	8.24	17	1.32	12.4
37	8.62	17	1.17	11.6
56	9.71	27	1.30	15.2
83	14.2	30	1.16	16.7
99	19.2	38	1.34	22.6

[a] The structure number indicates the rank of the structure after ordering according to final target function magnitude. [b] Only those constraint violations which exceed the upper bounds by 0.2 Å are included in this tabulation.

Structures Generated with DIANA. A set of 100 structures consistent with the experimental NOE distance constraints were generated by using the distance geometry program DIANA. Table I gives an overview of the results of these calculations showing the results for the five structures with the smallest target functions and five additional structures chosen at random from the remaining 95 structures. The "worst" structure in these 10 had a final target function only 3 times that of the "best", and a little over twice as many distance constraint violations. The maximum distance constraint violation in the "worst" structure was 1.34 Å, essentially the same as that seen in the "best" structure.

Figure 1 shows a superposition of the five best DG structures (i.e., those with the smallest target function) matched for best fit over the entire sequence. The overlayed structures were also matched for best fit either in the 8–20 residue region (Figure 2, top) or in the 1–6 residue region (Figure 2, bottom). It is clear that the NMR data are sufficient to constrain the structure in both of these regions, wheres the region of the structure in the vicinity of residue 8 is less well defined. Residues 9–20 form a rather well-defined α-helix, and residues 1–6 form a wide nonclassical turn structure. The turn in the amino-terminal region is dictated by only a few constraints, e.g., Pro3 HA to Thr6 γCH3, Val2 βH to Ile4 NH, and Ile4 HA to Tyr7 βH. The Val–Pro peptide bond is trans due to a strong NOE between the Val αH and the two Pro3 δH's.

The "goodness of fit" of the NOE data obtained by DG can be evaluated by comparing the experimental NOE intensities with those calculated for the model structures using a full relaxation matrix analysis as described under Materials and Methods (Table III). A single representative structure (structure 4 in Table I; similar results were obtained with the

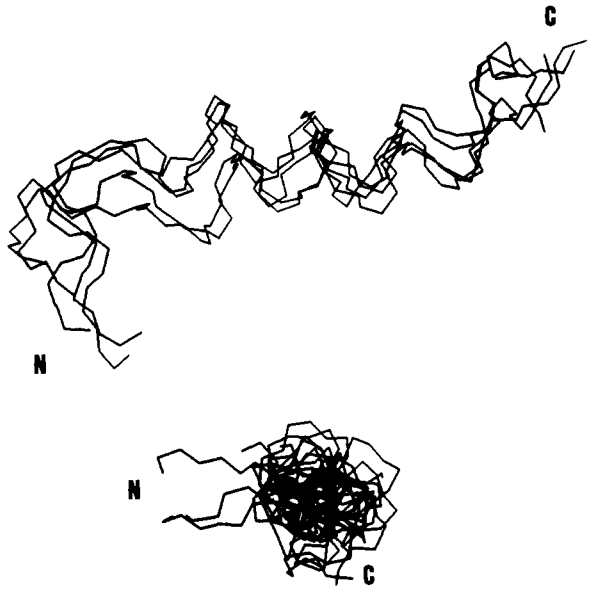

FIGURE 1: Two different orientations of an overlay of the five best structures of motilin obtained by using the distance geometry program DIANA. The structures are positioned to optimize the match of the positions of all backbone atoms.

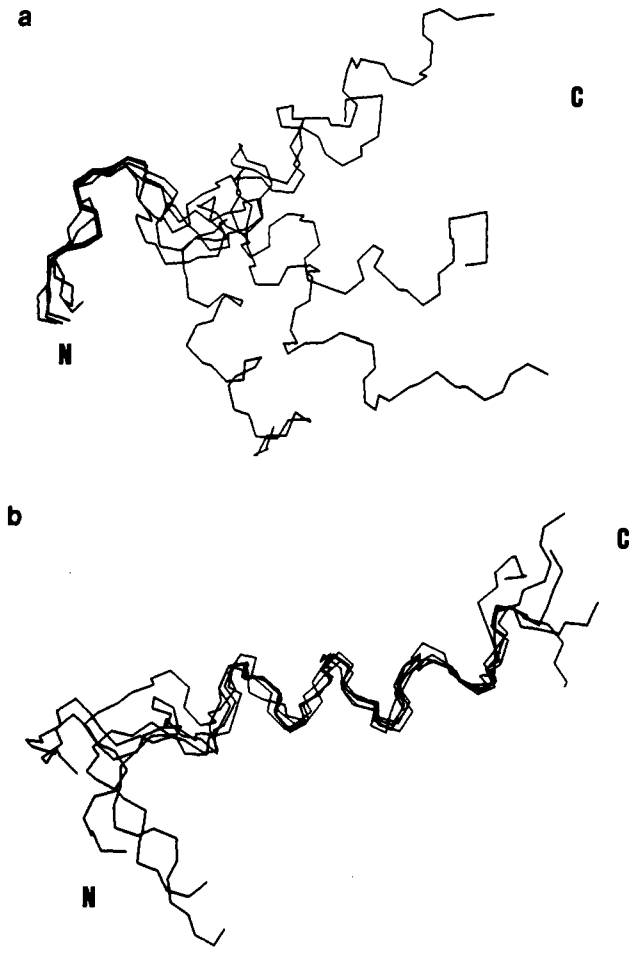

FIGURE 2: The five best DIANA generated structures of motilin are positioned to optimize the match of the backbone atoms of the amino terminus (i.e., the first six amino acid residues) in the upper view and the backbone atoms of the helical region (residues 9–20) in the lower view.

other four "best" structures) was chosen for comparison here, and only the backbone NOE's (and a few involving β-protons) which were well defined (nonoverlapping) and could be reliably integrated were used. (A complete listing of the NOE's used is included in the Supplementary Material). The mean error between the observed and calculated NOE intensities for the complete length of the molecule was 0.97. An NOE R factor of 0.81 was obtained for the complete length of the molecule. For the helical region of the molecule (Glu9–Lys20, inclusive), the mean error in NOE intensities was 0.33 and the R factor was 0.40, while for the amino terminus (Phe1 to Gly 8, inclusive) the mean error was 1.92 and the R factor was 1.38.

Table II: Summary of the Results Obtained after Structural Refinement Using Restrained Molecular Dynamics

structure[a]	energy[b] (kcal/mol)	no. of violations[c]	max violation (Å)	sum of violations[c] (Å)
1	−300	32	0.16	1.13
2	−270	41	0.16	1.78
3	−310	36	0.19	1.48
4	−330	35	0.17	1.23
5	−270	30	0.17	1.17
29	−300	39	0.15	1.64
37	−290	33	0.14	1.28
56	−320	30	0.15	1.31
83	−240	39	0.14	1.50
99	−240	44	0.16	1.77

[a] The structure number corresponds to the distance geometry structure numbers in Table I used as starting structures for the MD. [b] The energy reported is calculated without an NOE constraint contribution using the standard united-atom force field. [c] Note that *all* violations exceeding the upper constraint limit are counted here. None of the violations after restrained molecular dynamics exceeded 0.2 Å.

The 10 "best" DG structures, i.e., the 10 with the smallest target functions, were energy minimized using the program CHARMM with and without the NMR distance constraints. The RMSD's of the backbone atoms were approximately the same before and after energy minimization both with and without NOE constraints (see Supplementary Material). The resulting energies were, however, significantly lower after these calculations and reach similar low levels for all structures in both kinds of calculations (data not shown). This would indicate that the energy-minimized structures are physically better models than the original DG structures and that the potential energy surface contains many local minima close to each original DG structure. Energy minimization alone, therefore, does not appear to give a better defined set of structures [cf. Wüthrich (1989)].

Structures Generated with Restrained Molecular Dynamics. Restrained molecular dynamics using AMBER was performed using the 10 structures listed in Table I as starting structures, and the results are summarized in Table II. The best final structures from DG are not always the best starting structures for restrained MD. However, all 10 structures functioned well as starting structures, and all converged to structures with essentially the same number of distance constraint violations, maximum violation, sum of violations, and energies (Table II). It would be difficult to dismiss any of the final MD structures using any of these criteria. A comparison between the DG generated structures and those obtained after MD calculations (Tables I and II) shows that the MD procedure leads to a significant decrease in the maximum distance violation. It should be noted that the number of violations in Table I includes only those exceeding a threshold of 0.2 Å greater than the upper distance constraint, a common criterion for listing NOE constraint violations. Following MD, no distance exceeded an upper constraint by more than 0.2 Å.

The quality of the definition of the MD structures is demonstrated by the ϕ and ψ angle distributions in the family of restrained molecular dynamics structures and also by the root

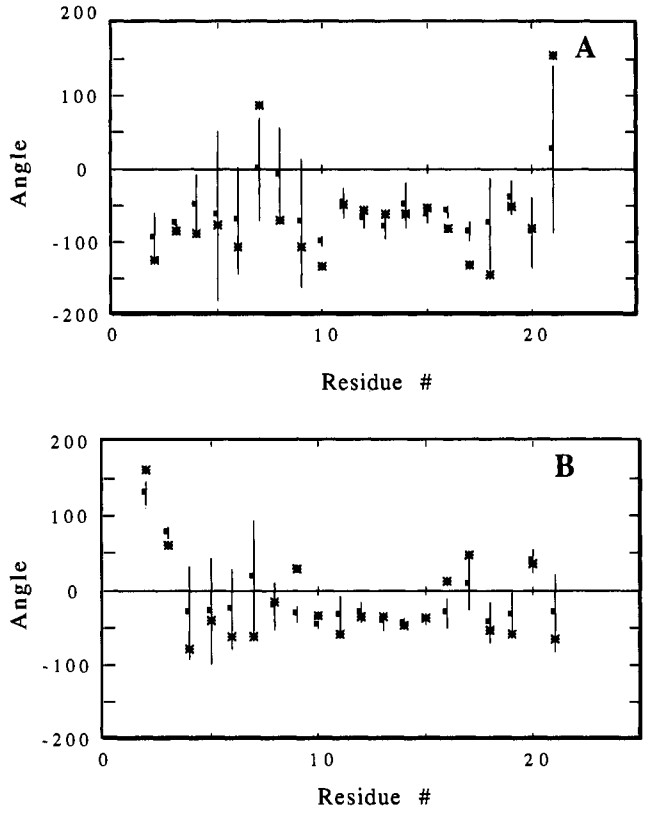

FIGURE 3: ϕ (A) and ψ (B) angles of the 10 structures obtained after applying restrained molecular dynamics to the 10 DG generated structures listed in Table I. The mean is indicated by a solid square, and the vertical bar indicates one standard deviation. In addition, the ϕ and ψ angles are represented by asterisks for the final FIRM-refined structure.

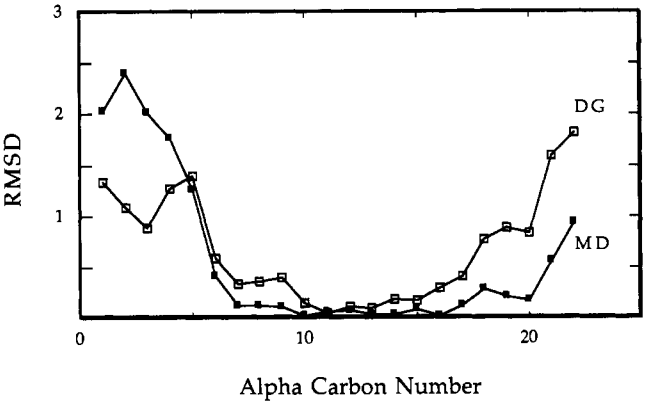

FIGURE 4: Root mean square deviations (RMSD) in angstroms of the α-carbon atom positions of the 10 structures generated by DG (□) and restrained MD (■). The α-carbon atoms of residues 10–15 were superimposed.

mean square deviations of the backbone carbon atom positions in the family of structures. Figure 3 shows the ϕ and ψ values for five of the MD structures derived from the five best DG structures. The structures appear to be well defined over a short stretch in the amino-terminal region and also from residue Glu9 to Lys20. The latter is nicely demonstrated by the low RMSD's (less than 0.5 Å) for the α-carbon atom coordinates from Tyr7 to Lys20 (Figure 4). The imprecision of definition of the structure at the ends of the peptide and in the vicinity of Phe5 to Gly8 is evident in both Figures 3 and 4. Figure 4 also includes the corresponding RMSD values obtained for the DG structures. Except for the five N-terminal residues, they are generally about twice as large as those obtained after restrained MD.

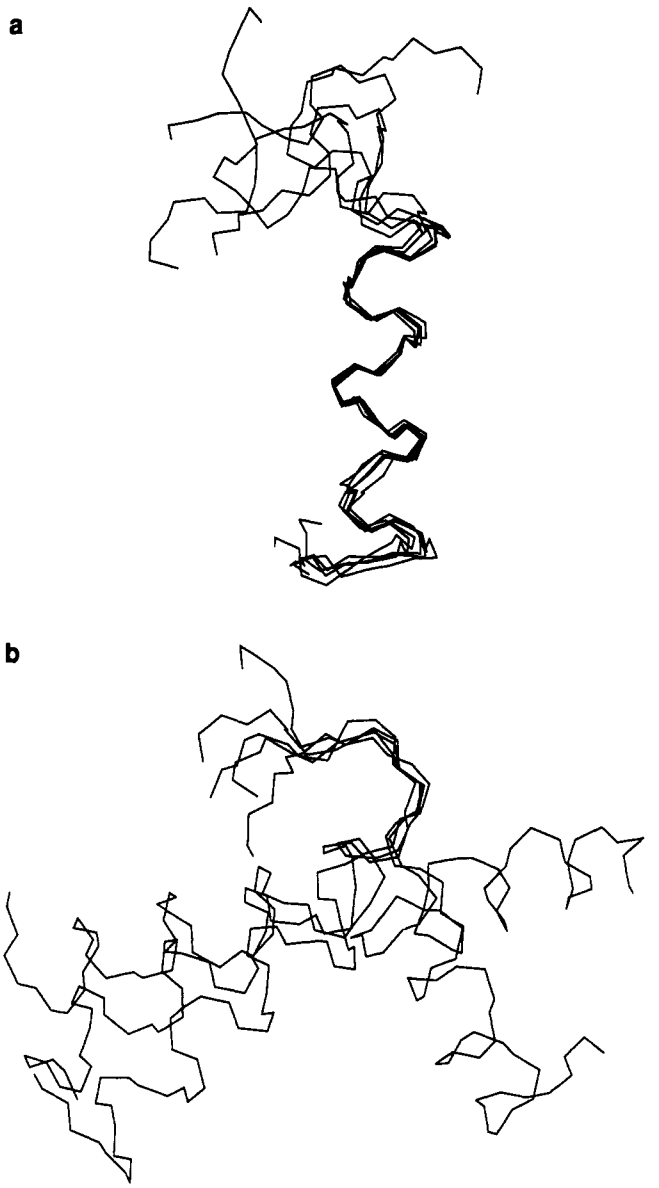

FIGURE 5: Overlay of five MD-refined structures of motilin. The structures are superimposed over the α-carbon atoms of residues 10–15 (A, top) and 3–6 (B, bottom). The N-terminus is at the top in both figures.

An overlay of the MD-refined structures resulting from starting with the best five DG structures is shown in Figure 5A. These have been aligned using the backbone atoms of residues 10–15. Clearly the α-helix is observed to extend over a large portion of the carboxy-terminal half of the molecule (from Glu9 to Lys20). In a couple of the other MD structures the helix breaks up further from the carboxyl end, e.g., at Lys16 (not shown). In the amino terminus, a turn is observed as indicated in Figure 5B where the same five structures are overlaid from residue 3 to 6. In this view, the folded helices can be seen extending outward in five different directions.

The goodness of fit or *precision* with which the structures are defined by the restrained molecular dynamics can be evaluated by comparison of the experimental and calculated NOE intensities (Table III). A comparison is made here using a single arbitrary structure (structure 4 in Table II; similar results were obtained with structures 1–5). For the full length of the molecule, the mean error between the observed and calculated NOE intensities was 1.42, and the *R* factor was 1.01 (using the same backbone atoms as above for the DG com-

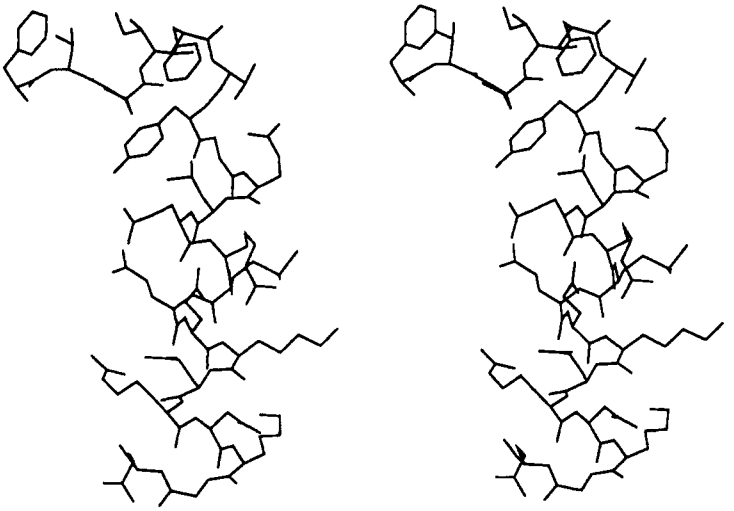

FIGURE 6: Stereodrawing of the solution structure of motilin obtained from an iterative full relaxation matrix refinement of the two-dimensional NOESY data using FIRM.

Table III: Comparison of the Mean Errors and *R* Factors between Observed and Calculated NOE Intensities for the Distance Geometry, Restrained Molecular Dynamics, and Iterative Full Relaxation Matrix Refined Structures[a]

	mean errors in NOE intensities		
	region		
refinement strategy	1–8	9–20	1–22
distance geometry	1.92	0.33	0.97
molecular dynamics	3.01	0.40	1.42
iterative full relaxation matrix refinement	0.29	0.18	0.23
	R factor		
	region		
refinement strategy	1–8	9–20	1–22
distance geometry	1.38	0.40	0.81
molecular dynamics	1.78	0.59	1.01
iterative full relaxation matrix refinement	0.43	0.19	0.27

[a] The *R* factors and mean errors were calculated using 57 well-defined NOE intensities for backbone protons (and a few selected interresidue backbone to β-proton NOE's) over three regions of the molecule: residues Phe1–Gly8, Glu9–Lys20, and the full length from Phe1 to Gln22. (A listing of the calculated and observed NOE intensities for the FIRM structure is included in Table III of the Supplementary Material.)

parisons and listed in the Supplementary Material). For the helical region (residues 9–20, inclusive), the mean error was 0.40 and the *R* factor was 0.59, while for the amino terminus (residues 1–8, inclusive) the mean error was 3.01 and the *R* factor was 1.78. Although the MD appears to give a more precisely defined set of structures than DG based on backbone RMSD's, the agreement with the experimental NOE data is not as good.

The low RMSD's obtained for the MD structures are due to the strong emphasis placed on the NMR constraints during the later stages of the MD refinement. The choice of how large K_{noe} should be in the refinement is a subjective decision which certainly affects the apparent precision of the final structure. We feel that since we are attempting to arrive at a structure based on experimental results, primary emphasis should be placed on the data. Inaccuracies that result are due in part to spin diffusion effects which are not included in the two-spin approximation used for obtaining the distance constraints. If shorter mixing times and NOE buildup rates had been available, it is expected that the MD refinement would have been more reliable.

NOESY Iterative Refinement. The most reliable criterion for evaluating the precision of an NOE defined structure is the extent of agreement between the observed NOE intensities and those calculated from the final structure. As others have indicated, "a highly converged set of very similar structures...may merely (be) a set of consistently wrong structures" (Nerdal et al., 1989). We have therefore undertaken an iterative full relaxation matrix refinement in an attempt to fit the experimental NOE data, rather than use the data to define a range of constraints. The full relaxation matrix approach includes the effects of spin diffusion and therefore is the best approach with limited data. Such a refinement can provide an overall evaluation of the reliability of the final structure.

An iterative full relaxation matrix refinement of the NOE data was performed by using the program FIRM with an arbitrary initial structure chosen from the family of MD structures (structure 4 in Table II). A number of cycles of merging the experimental NOE's into the calculated NOE matrix were performed as described under Materials and Methods. It was found that the back-transformation of the merged matrix behaved best if the calculated intensities were replaced by a weighted sum of the experimental and calculated intensities (e.g., 50% calculated and 50% experimental) rather than simply replacing the calculated with the observed values. In addition, we were not successful in obtaining a refined structure if we used all of the NOE's, including those involving side-chain protons. This is presumably due to flexibility of the side chains, leading to averaging of the NOE intensities. We therefore limited the NOE set to the backbone NOE's. A total of 57 well-resolved NOE's involving backbone protons were used in the FIRM refinement including inter- and intraresidue NOE's between αCH and NH protons, and two additional NOE's between the β-proton of Val 2 and the NH of Ile4, and between the αCH of Ile4 and the β-protons of Tyr7.

A stereodrawing of the FIRM-refined structure is shown in Figure 6. Since the NOE in a multispin environment depends on neighboring spins, the geometry of the side chains is included in Figure 6 even though experimental side-chain NOE's were not merged in the refinement procedure. The energy of the FIRM structure (calculated with the standard AMBER united-atom potential, no constraints) was −310 kcal/mol, comparable to energies of the structure refined by MD under the two-spin approximation. The ϕ and ψ angles of the FIRM structure are displayed in Figure 4 to allow comparison with the MD structure. The RMSD of the backbone atoms was

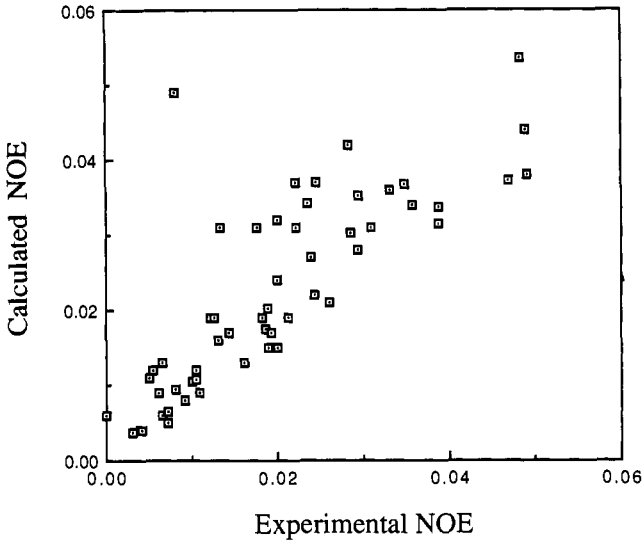

FIGURE 7: Comparison of 57 experimentally observed backbone (plus a few selected β-proton) NOE intensities with the calculated NOE's based on the final FIRM-refined structure. A listing of the NOE intensities is provided in the Supplementary Material.

1.4 Å relative to the starting MD-refined structure, with some backbone atoms moving by more than 2 Å in the N-terminal region.

Very good agreement between the calculated and observed NOE's was obtained by using the FIRM structure as a basis for calculating the NOE intensities (see Figure 7). The mean error between the calculated and observed NOE intensities was 0.23 (using the 57 backbone NOE's only) for the entire sequence and 0.18 using the backbone NOE's from Glu9 to Lys20 (Table III). For the full length of the backbone, an *R* factor of 0.27 was obtained, and for the sequence from Glu9 to Lys20 the *R* factor was 0.19. In the amino terminus (Phe1–Gly8), the *R* factor was 0.43 and the mean error between calculated and experimental NOE's was 0.29. As found with the two-spin approximation, the structure in the amino-terminal region does not fit the NOE data as well as the rest of the molecule. This may be due to flexibility (or coexistence of different conformers) and/or a different dynamic behavior in this part of the molecule.

The experimental NOE intensities in the N-terminal region of the molecule tend to be significantly less than those calculated from the FIRM structure. This supports the idea of a decreased rotational correlation time, possibly correlated with increased flexibility in this region of the molecule. For example, the calculated and experimental NOE's between Val2 βH and Ile4 NH are 0.39 and 0.19, respectively; between Thr6 NH and Tyr7 NH they are 0.049 and 0.008; between the NH and αH of Phe5 they are 0.011 and 0.005; between the NH and αH of Thr6 they are 0.031 and 0.018; and between the NH and αH of Val2 they are 0.006 and 0.000. A notable exception is between Pro3 αH and Thr6 γCH3, where the calculated NOE is less than the experimental, viz., 0.006 and 0.24, respectively. Some of this difference may be due to the inability to adequately treat cross-relaxation to a methyl group in our present refinement programs. At present the three methyl protons are treated as three exchangeable protons in a rotor with $\langle r^{-6} \rangle$ averaging (Koning et al., 1990). A clear indication of the imprecision in defining the structure is indicated by an *R* factor of 0.43 for the amino terminus (residues Phe1–Gly8) and a mean error between calculated and observed NOE intensities of 0.29.

None of the structures obtained here explain the upfield shift of the Ile4 NH (Khan et al., 1990). However, its presence would appear to imply that some tendency for secondary structure exists in the amino terminus. We hope to investigate this feature in future studies.

Summary. Models of the solution structure of motilin have been obtained by using NOE-defined constraints in conjunction with DG, restrained MD, and an iterative full relaxation matrix refinement. The quality of the structures has been evaluated at each stage of the analysis in terms of satisfaction of the constraints, the magnitude of target functions and energies, root mean square deviations of atomic positions in a family of structures, and agreement between calculated and observed NOE intensities. It is demonstrated that low RMS deviations of atomic positions in a family of structures is not the best evidence for a precisely defined structure in the absence of spectral back-calculations. In the work presented here, one could be led to the false conclusion that the MD structure was "better" than that obtained from DG using the criteria of total energy, constraint violations, and RMS deviations of atomic positions and bond angles. In the end, the best evidence of a well-defined structure must be the demonstration of agreement between experimental and calculated NOE intensities.

The quantitative analysis of the NOESY data presented here supports our previous conclusion that two types of secondary structure are present in motilin, an ill-defined turn in the hydrophobic amino terminus and an α-helix extending over the hydrophilic sequence from Glu9 to Lys20. These two regions appear to be joined by a "hinge-like" region from Thr6 to Gly8.

The two structural regions in motilin are apparently both important for hormonal activity. Since the amino-terminal fragment 1–5 alone can bind to the motilin receptor (Peeters et al., 1988), it appears that the hydrophobic portion of the hormone is of particular importance for receptor recognition. The results reported here indicate that the hormone in this region is most likely flexible and a unique conformer may not exist in solution.

ACKNOWLEDGMENTS

We acknowledge the assistance of Drs. David Gorenstein and Ed Nikonowicz in initiating AMBER molecular dynamics calculations on the computing facilities at the NIH Research Resource and the NSF National Biological Facility for Biomolecular NMR, Structure and Design at the Purdue University Biochemical Magnetic Resonance Laboratory. We thank Dr. Lennart Nilsson and Dr. Anders Ehrenberg for stimulating discussions and help.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing the mean residue ellipticity at 222 nm of motilin as a function of the concentration of hexafluoro-2-propanol and three tables giving selected intermediate NOE-derived distances in motilin, the effect of energy minimization on the average RMSD's of the atomic positions of the backbone atoms using CHARMM on the 10 best distance geometry structures, and a comparison of the calculated and experimental backbone NOE intensities at the completion of the FIRM refinement of the structure of motilin (6 pages). Ordering information is given on any current masthead page.

Registry No. Motilin, 9072-41-7.

REFERENCES

Baleja, J., Moult, J., & Sykes, B. (1990) *J. Magn. Reson. 87*, 375–384.

Boelens, R., Koning, T., van der Marel, G., van Boom, J., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290–308.

Borgias, B., & James, T. (1990) *J. Magn. Reson. 87*, 475–487.

Borgias, B., Gochin, M., Kerwood, D., & James, T. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 83–100.
Bormans, V., Peeters, T. L., & Vantrappen, G. (1986) *Regul. Pept. 15*, 143–153.
Brooks, B. R., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187.
Brown, J. C., Cook, M. A., & Dryburgh, J. R. (1973) *Can. J. Biochem. 51*, 533–537.
Cantor, C., & Schimmel, P. (1980) *Biophysical Chemistry*, p 763, W. H. Freeman, San Francisco, CA.
Depoortere, I., Peeters, T. L., Matthijs, G., & Vantrappen, G. (1988) *Hepato-Gastroenterology 35*, 198.
Depoortere, I., Peeters, T. L., & Vantrappen, G. (1990) *Gastroenterology 99*, 652–658.
Dyson, H. J., Rance, M., Houghten, R., Wright, P., & Lerner, R. (1988) *J. Mol. Biol. 201*, 201–217.
Gippert, G., Yip, P., Wright, P., & Case, D. (1990) *Biochem. Pharmacol. 40*, 15–22.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517–530.
Inatomi, H., Satoh, H., Maki, Y., Hashimoto, H., Itoh, Z., & Omura, S. (1989) *J. Pharmacol. Exp. Ther. 251*, 707–712.
Itoh, Z. (1990) *Motilin*, Academic Press, Inc., San Diego, CA.
Itoh, Z., Nakaya, M., Suzuki, T., Arai, H., & Wakabayashi, K. (1984) *Am. J. Physiol. 247*, G688–G694.
Itoh, Z., Suzuki, T., Nakaya, M., Inoue, M., Arai, H., & Wakabayashi, K. (1985) *Am. J. Physiol. 248*, G320–G325.
Keepers, J., & James, T. (1984) *J. Magn. Reson. 57*, 404–426.
Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W., & Berendsen, H. (1988) *J. Am. Chem. Soc. 110*, 3393.
Khan, N., Gräslund, A., Ehrenberg, A., & Shriver, J. (1990) *Biochemistry 29*, 5743–5751.
Kitigawa, K., Yoneto, K., Kiyarna, S., Ando, K., Kawamoto, T., Akita, T., Inoue, A., & Segawa, T. (1985) *Chem. Pharm. Bull. 33*, 3307–3316.
Kondo, K., Torii, T., Omura, S., & Itoh, Z. (1988) *Biochem. Biophys. Res. Commun. 150*, 877–882.
Koning, T., Boelens, R., & Kaptein, R. (1990) *J. Magn. Reson. 90*, 111–123.
Nerdal, W., Hare, D., & Reid, B. (1989) *Biochemistry 28*, 10008–10021.
Nikonowicz, E., Meadows, R., & Gorenstein, D. (1990) *Biochemistry 29*, 4193–4204.
Peeters, T. L., Bormans, V., Depoortere, I., Matthijs, G., Kitagawa, K., & Vantrappen, G. (1988a) *Biomed. Res. 9*, 361–366.
Peeters, T. L., Bormans, V., & Vantrappen, G. (1988b) *Regul. Pept. 23*, 171–182.
Peeters, T. L., Matthijs, G., Depoortere, I., Cachet, T., Hoogmartens, J., & Vantrappen, G. (1989) *Am. J. Physiol. 257*, G470–G474.
Peeters, T. L., Vantrappen, G., & Depoortere, I. (1990) in *Motilin* (Itoh, Z., Ed.) Academic Press, Inc., San Diego, CA.
Poitras, P., Reeve, J., Hunkapillar, M., Hood, L., & Walsh, J. (1983) *Regul. Pept. 5*, 197–208.
Poitras, P., Lahaie, R. G., St-Pieere, S., & Trudel, L. (1987) *Gastroenterology 92*, 658–662.
Post, C., Meadows, R., & Gorenstein, D. (1990) *J. Am. Chem. Soc. 112*, 6796–6803.
Powers, R., Jones, C., & Gorenstein, D. (1990) *J. Biomol. Struct. Dyn. 8*, 253–294.
Satoh, T., Inatomi, N., Satoh, H., Marui, S., Itoh, Z., & Omura, S. (1990) *J. Pharmacol. Exp. Ther. 254*, 940–944.
Schubert, H., & Brown, J. C. (1974) *Can. J. Biochem. 52*, 7–8.
Segawa, T., Nekano, M., Kai, Y., Kawatani, H., & Yajima, H. (1976) *J. Pharm. Pharmacol. 28*, 650–651.
Simorre, J., Genest, D., Caille, A., & Ptak, M. (1990) *Eur. Biophys. J. 18*, 309–316.
Smith, L., Sutcliffe, M., Redfield, C., & Dobson, C. (1991) *Biochemistry 30*, 986–996.
Strunz, U., Domschke, W., Mitznegg, P., Domschke, S., Schubert, E., Wunsch, E., Jaeger, E., & Demling, L. (1975) *Gastroenterology 68*, 1485–1491.
Strunz, U., Domschke, W., Domschke, S., Mitznegg, P., Wunsch, P., Jaeger, E., & Demling, L. (1976) *Scand. J. Gastroenterol. 11*, 199–203.
Ueda, K., Kitagawa, K., & Akita, T. (1977) *Chem. Pharm. Bull. 25*, 2123–2126.
Vantrappen, G., & Peeters, T. L. (1989) in *Handbook of Physiology*, Vol. II, Section G, pp 545–558, American Physiological Society, Bethesda, MD.
Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.
Wright, P., Dyson, H. J., & Lerner, R. (1988) *Biochemistry 27*, 7167–7175.
Wüthrich, K. (1989) *Science 243*, 45–50.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.
Yajima, H., Kai, Y., Ogawa, H., Kubota, M., Mori, Y., & Koyama, K. (1977) *Gastroenterology 72*, 793–796.