Stubbs, G., Warren, S., & Holmes, K. (1977) *Nature (London) 267*, 216–221.
Turner, D. R., Mondragon, A., Fairall, L., Bloomer, A. C., Finch, J. T., VanBoom, J. H., & Butler, P. J. G. (1986) *Eur. J. Biochem. 157*, 269–274.
Vogel, D., DeMarcillac, G. D., Hirth, L., Gregori, E., & Jaenicke, R. (1979) *Z. Naturforsch., C: Biosci. 34C*, 782–792.

Solution Structure of Apamin Determined by Nuclear Magnetic Resonance and Distance Geometry†

Joseph H. B. Pease and David E. Wemmer*
Department of Chemistry and Chemical Biodynamics Division, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720
Received May 9, 1988; Revised Manuscript Received July 1, 1988

ABSTRACT: The solution structure of the bee venom neurotoxin apamin has been determined with a distance geometry program using distance constraints derived from NMR. Twenty embedded structures were generated and refined by using the program DSPACE. After error minimization using both conjugate gradient and dynamics algorithms, six structures had very low residual error. Comparisons of these show that the backbone of the peptide is quite well-defined with the largest rms difference between backbone atoms in these structures of 1.34 Å. The side chains have far fewer constraints and show greater variability in their positions. The structure derived here is generally consistent with the qualitative model previously described, with most differences occurring in the loop between the β-turn (residues 2–5) and the C-terminal α-helix (residues 9–17). Comparisons are made with previously derived models from NMR data and other methods.

Apamin is a small neurotoxic peptide component of honey bee venom. Like many other peptide neurotoxins, apamin has a high cystine content and a high basicity, but apamin is different from most peptide toxins in its unusual ability to cross the blood brain barrier and act on the central nervous system (Habermann, 1972). Apamin is known to block calcium-dependent potassium fluxes, possibly by binding to a Ca^{2+}-dependent potassium channel (Banks et al., 1974). Additionally, apamin serves as a model for understanding various aspects of peptide folding and amide proton exchange. To determine the structural basis of its activity and for interpretation of other experimental results, it is important to have a good understanding of the structure of this peptide. Apamin contains 18 amino acids with two disulfides (see Figure 1), which give it an extremely stable structure with respect to temperature, pH, and denaturants, not unfolding completely even at 70 °C in 6 M guanidinium hydrochloride (Miroshnikov et al., 1978; N. V. Kumar, personal communication). During the past few years there have been a number of studies aimed at determination of apamin's structure, using energy refinements (Freeman et al., 1986; Hider & Ragnarsson, 1981), circular dichroism (CD)[1] (Miroshnikov et al., 1978), and nuclear magnetic resonance spectroscopy (NMR) (Wemmer & Kallenbach, 1983; Okhanov et al., 1980; Bystrov et al., 1978). Here we use two-dimensional proton NMR data with distance geometry to obtain an improved picture of apamin's solution structure.

Previously, two-dimensional NMR was used to obtain almost complete assignments of apamin's proton NMR spectrum (Wemmer & Kallenbach, 1983).[2] As with many other peptides, the sequential assignment pattern and identification of slowly exchanging amides gave information about the secondary structure of the molecule. Using this information, Wemmer and Kallenbach (1983) derived a model that was consistent with the previously reported CD data (Miroshnikov et al., 1978) and ϕ angles from $^3J_{NH-\alpha H}$ coupling constants (Bystrov et al., 1980). Hider et al. (1986) have also developed structural models for apamin based on energy refinement methods, and Bugg and co-workers[3] have used coordinates from their crystal structure of a scorpion toxin, for which one section seemed to be structurally similar to apamin, together with modeling and energy refinement to develop a structural model. In this work we describe the detailed structure obtained from further NMR measurements and our distance geometry calculations and compare the previously derived structural models with it.

Materials and Methods

NMR Spectroscopy. All two-dimensional nuclear Overhauser effect (NOESY) spectra (Anil-Kumar et al., 1980; Jeener et al., 1979) were recorded on a 500-MHz General

† Supported by the Office of Energy Research, Office of Health and Environmental Research, Health Effects Research Division of the U.S. Department of Energy under Contract DE AC03-76SF00098 and through instrumentation grants from the U.S. Department of Energy, DE FG05-86ER75281, and the National Science Foundation, DMB 86-09035.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; TPPI, time proportional phase incrementation; NOE, nuclear Overhauser effect; CD, circular dichroism; standard single-letter abbreviations for amino acids (A = alanine, C = cysteine, E = glutamic acid, H = histidine, K = lysine, L = leucine, N = asparagine, P = proline, Q = glutamine, R = arginine, T = threonine).

[2] Some of the previously reported assignments are corrected in the present work. These are the Cys 1 β-protons (3.12 and 2.74 ppm), the Pro 6 α-proton (4.66 ppm), the Pro 6 β-protons (1.90 and 1.75 ppm), and the Glu 7 β-protons (2.25 and 2.08 ppm). Chemical shifts are indirectly referenced to TSP through assignment of $\delta(H_2O)$ = 4.75 ppm at 30 °C.

[3] C. E. Bugg, A. Zell, M. Carson, M. Epps, and J. Hermans, personal communication.

0006-2960/88/0427-8491$01.50/0

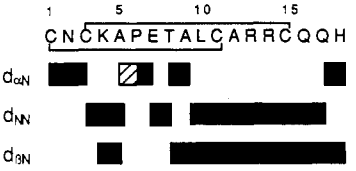

FIGURE 1: Sequence of apamin with a summary of the sequential resonance assignments. The lines connecting residues 1 and 11 and 3 and 15 represent apamin's disulfides. The solid bars indicate the presence of that connectivity between connected residues, and the slashed bar indicates Ala 5 αH to Pro 6 δH connectivities.

Electric GN-500 spectrometer using time proportional phase incrementation (TPPI) to obtain phase-sensitive spectra (Drobny et al., 1979; Bodenhausen et al., 1984). Each NOESY spectrum had 512 t_1 points, spectral width of 5000 Hz, and 1024 complex t_2 data points. Suppression of the residual water peak was achieved by using continuous low-power irradiation during the relaxation delay for the D_2O spectra and during both the relaxation delay and the mixing time for the H_2O spectrum. A skewed sine-bell apodization (Hare et al., 1985) was used in both dimensions with a skew of 0.6 and a phase shift of 50°. Data were zero filled in the second dimension to yield 1024 × 1024 point real matrices giving 5 Hz/point digital resolution in both dimensions. The first row of the (t_1, ω_2) matrix was multiplied by 0.5 before the t_1 transform to suppress t_1 ridges (Otting et al., 1986). One D_2O buildup curve was done at 30 °C with mixing times of 150, 300, and 450 ms. NOE intensities were significantly higher upon lowering the temperature to 5 °C, due to the lengthening of the correlation time, moving away from $\omega\tau_c \approx 1$. Mixing times of 200 and 450 ms were used at this lower temperature. The 5 °C spectra also had the advantage that the residual water peak moved downfield of all of the α-protons. One 30 °C 90% H_2O/10% D_2O NOESY was collected with a mixing time of 450 ms. All spectra were obtained from an approximately 20 mM sample of apamin at pH 2 in 400 μL of either D_2O or 90% H_2O/10% D_2O to include amide resonances. All data processing was done by using the FTNMR program (D. Hare, unpublished results) on either a Micro Vax II or a Vax 11/785 computer. NOE cross-peaks on both sides of the diagonal were integrated by using the volume integration routine in FTNMR. The average intensity of the cross-peaks on either side of the diagonal was used except when a streak was present, in which case the peak without the streak was used. Most cross-peak intensities were calibrated by using the geminal distance between the β-protons of residues 3 and 7, assuming a distance of 1.75 Å between them. Some NOEs between methyl groups and individual protons were calibrated by using α-proton to methyl intraresidue NOE of Ala 5 and Ala 12, assuming an effective distance of 2.38 Å between them. The uncertainty in all calculated distances was ±0.25 Å, estimated from uncertainties in the slope of the mixing time dependence of NOEs. Figure 3 shows sample buildup curves for both short- and long-distance NOEs. Distances derived from the water spectrum were treated as semiquantitative: 1.9–2.5 Å for strong cross-peaks; 1.9–3.0 Å for medium cross-peaks; and 2.5–3.5 Å for weak cross-peaks. The maximum proton–proton distance that had an observable cross-peak was <3.5 Å, on the basis of NOE intensities of intraresidue αH–NH peaks for Asn 2 and Cys 3, which are in an extended conformation. Stereospecific assignments were used for the side-chain protons of Pro 6. The assignments were established by using the short distances between the protons on either side of the ring, beginning with the αH. The αH is closer to βH than to β'H and is also close to γH; γH is closer to δH than to δ'H. This puts αH, βH, γH, and δH on the same side of the proline ring, with a similar argument for β'H, γ'H, and δ'H on the opposite side.

Distance Geometry. Distance geometry calculations were done by using the DSPACE program (D. Hare, unpublished results) on a Micro Vax II computer. The program uses the metric matrix approach as described by Crippen (1981). The basic protocol for obtaining embedded structures is generally similar to that in the program DISGEO (Havel & Wüthrich, 1984), but DSPACE uses all the atoms when embedding and does not use a subembedding step as is used in DISGEO. Most aspects of the subsequent refinement are quite different. Distance geometry is a general method for computing a random sampling of molecular conformations that are consistent with a set of input distance constraints. A bounds matrix was generated that included constraints from the NOE data, hydrogen bonds deduced from the slowly exchanging amides, and the disulfide bonds. The NMR constraints used are given in Table I. NOE buildup rates were determined to obtain more accurate distances than would be obtained from a single NOESY spectrum and to determine whether spin diffusion was occuring (Anil-Kumar et al., 1981). A sample NOESY spectrum is pictured in Figure 2. Although relatively long mixing times were used, the small size of apamin makes its rotational correlation time quite short and thus the NOE buildup rather slow (see Figure 3), minimizing the importance of spin diffusion except for the longest mixing times at low temperature.

Several specific hydrogen bonds were defined corresponding to regions of defined secondary structure in apamin. Apamin's small size allows solvent accessibility to most of its residues, and therefore slow amide exchange is most likely due to hydrogen bonding and not from the amide being buried within the peptide (Englander & Kallenbach, 1984). The amides of residues 4, 5, 8, and 11–16 are relatively slowly exchanging (Bystrov et al., 1980; Wemmer & Kallenbach, 1983; Dempsey, 1986). Residue 5's amide hydrogen bonds with the backbone carbonyl of 2, forming a type I β-turn. This is supported by the connectivity pattern NH–NH–NH for residues 3–5, by the slow exchange of residue 5's amide, and the values of coupling constants for these residues. Residue 4's amide is in a position where it can form a hydrogen bond with the side-chain carbonyl of Asn 2, which would stabilize the β-turn. The preference for asparagine at this position in a β-turn and its ability to hydrogen bond to the backbone amide are well-known (Richardson, 1981). This hydrogen bond has little effect on the position of the backbone atoms, and if not included the side chain of Asn 2 would take on a more random conformation. The amides of residues 12–17 form an α-helix, which frays significantly beyond residue 15. The connectivities for residues 9–17 are all NH to NH and NH to βH type, consistent with a helix. This helix is further confirmed by tertiary contacts of the αH(i) to βH(i + 3) and αH(i) to NH(i + 3) types (Wüthrich et al., 1984). The observed NOEs were between αH(9) and βH(12), αH(10) and βH(13), αH(11) and βH(14), αH(12) and βH(15), αH(9) and NH(12), αH(11) and NH(14), and αH(13) and NH(16). The connectivity between the αH(10) and the βH(13) was clearly present as a shoulder on the αH(13)–βH(13) cross-peak in some spectra, but could not be quantified due to near degeneracy of the αH chemical shifts of residues 10 and 13. This degeneracy also makes it hard to tell if the αH(10)–NH(13) NOE is present. Although the exchange of the amides of residues 8 and 11 is somewhat slower than that of other residues that are fully exposed to solvent, there were no clear hydrogen-bond acceptors that could be identified, and hence no hydrogen-bond

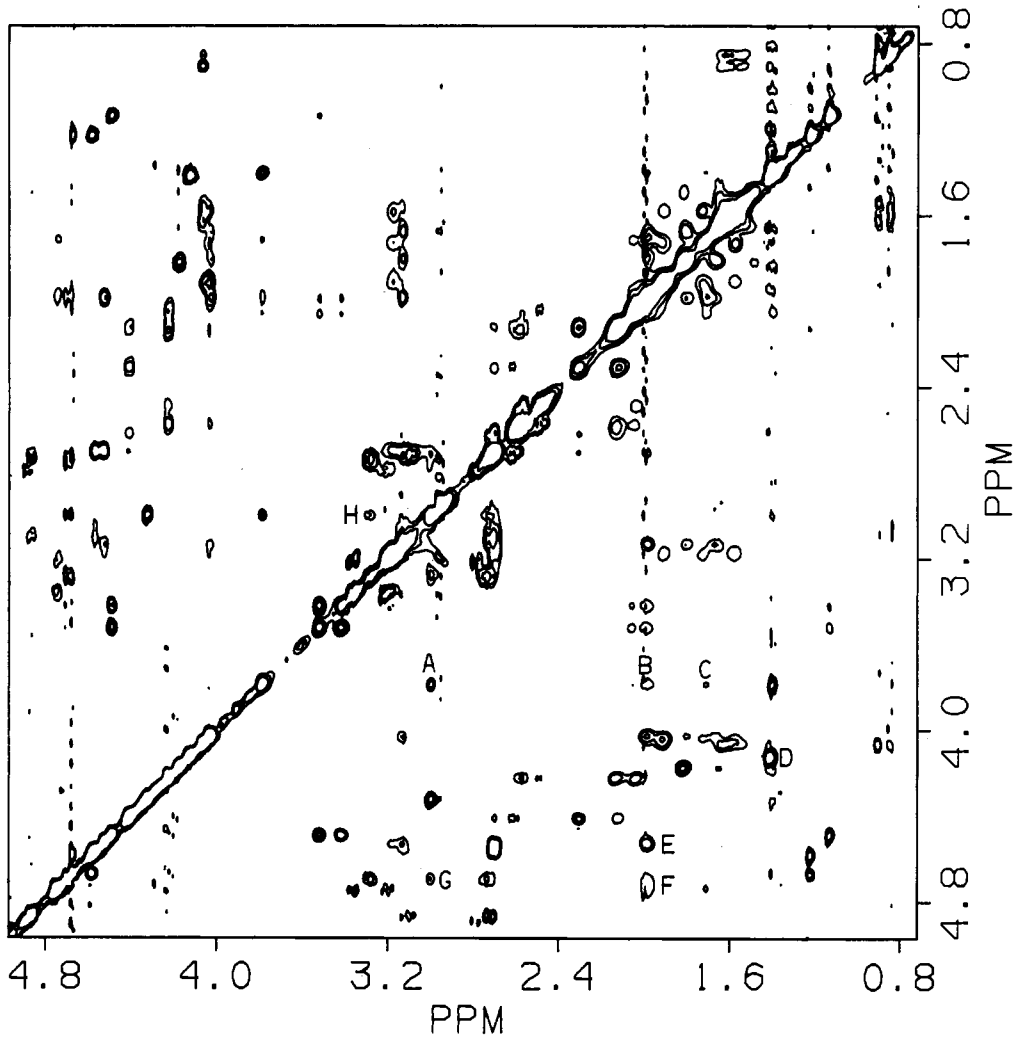

FIGURE 2: Part of the NOESY spectrum in D_2O solution, pH 2, taken at 5 °C with a mixing time of 450 ms. NOEs that represent secondary and tertiary structure-dependent contacts are indicated. (A) αH(12)–$\beta\beta'$H(15); (B) αH(12)–γH,βH(6); (C) αH(12)–β'H(6); (D) αH(9)–CH_3(12); (E) αH(11)–$\beta\beta'$H(14); (F) αH(3)–γH,βH(6); (G) αH(3)–$\beta\beta'$H(15); (H) βH(3)–$\beta\beta'$H(15).

constraints were used for these residues. Hydrogen bonds were included for the helix and the turn as listed in Table I and were defined to be 2.0 Å from amide proton to acceptor with the donor to acceptor distances defined by the program so that the hydrogen bonds are nearly linear. There is clearly extensive fraying of the last two residues, and the lack of NOE constraints leaves this region relatively poorly defined.

The program also uses templates derived from crystallographic data to set bond lengths, geminal distances, and chirality constraints and to maintain the planarity of peptide bonds and aromatic rings. An additional distance constraint was put in to define the distance between the β-carbons in a disulfide bond so that it adopts a standard conformation, making a Cβ–S–S–Cβ dihedral angle of approximately $\pm(90 \pm 10°)$ (Richardson, 1981) as found in essentially all high-resolution crystal structures. All unknown distances were given an initial upper bound of 1000 Å (effectively no constraint) and a lower bound of the sum of the van der Waals radii. Pseudoatoms were used only for methyl groups (Wüthrich et al., 1983); all other protons were included explicitly. The distance constraints within the bounds matrix were then extended by using the triangle inequality. The resulting smoothed matrix was used both to generate the starting structures and to evaluate the penalty function during refinement of the embedded structures.

Twenty starting structures were generated by using correlated trial distances between the upper and lower bounds. Correlated trial distances were generated by selecting an initial random distance and then setting other distances, within their allowed range, to be consistent with the bounds matrix and the initially chosen distance. Although this process takes a considerable amount of time, the embedded structures are clearly better than those with all distances chosen randomly. The penalty function used during refinement included independently weighted terms that correspond to chirality; templates (planarity); geminal; bond; van der Waals distances; experimental distances; hydrogen bonds; and constraints derived from smoothing. Two types of refinement algorithm were used. The first minimized the penalty function by using standard conjugate gradient routines, with an adjustable cutoff distance for testing violations. This was done either locally, for specific groups or residues in the molecule, or for the entire molecule at once with a cutoff of 5 Å. The second method used a molecular dynamics approach, assigning a random velocity term to each atom (with adjustable average magnitude and force constant for maintaining bounds) and then letting the system evolve dynamically by using classical equations of motion, maintaining constant total error instead of energy. It is not a true dynamics refinement since it does not use real potential energy terms, but it serves a very similar function in moving the molecule out of local minima in the error function. When desired, a damping term was added that gradually reduces the velocities and hence the total error. A cycle of assigning random velocities, letting the molecule evolve

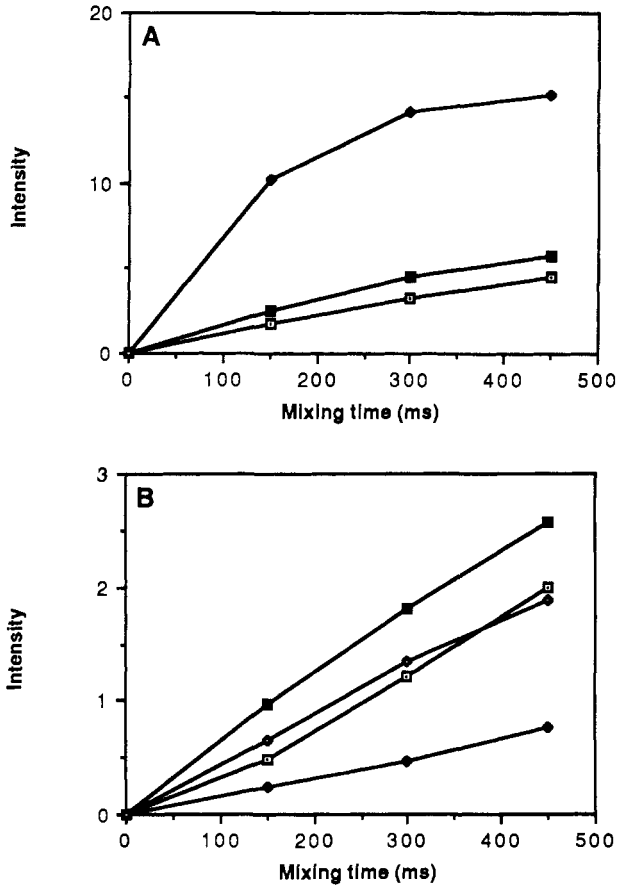

FIGURE 3: NOE buildup curves from NOESYs taken at 30 °C. (A) Short fixed distances: 1.75 Å [◆, βH(3)–β′H(3)] and 2.4 Å [⊡, αH(12)–βM(12); ■, αH(5)–βM(5)]. (B) Longer distances: 2.4–3.5 Å [⊡, αH(12)–ββ′H(15); ◆, βH(6)–αH(12); ■, αH(11)–ββ′H(14); ◆ αH(12)–β′γ′H(6)]. Note the difference in vertical scales.

dynamically, and then adding damping to gradually dissipate the velocity, which was repeated several times with lower average velocity in each cycle, worked well to refine the structures. When these cycles were no longer effective at further reducing the total error, conjugate gradient refinement was done until no further reduction of error could be achieved. The ability to independently weight each element of the penalty function (and to change these weights during refinement) also provides a mechanism for avoiding local minima during refinement with either method.

Full refinements were done with two different approaches. The first initially refined the local structure using the conjugate gradient routine on individual amino acids and then subsequently the whole molecule using both conjugate gradient and dynamics routines. The alternate approach refined the whole molecule at once, occasionally using local refinements to fix chirality. The dynamics refinement algorithm was both faster to converge and better at getting out of local minima than the conjugate gradient routine. Both refinement algorithms gave similar final results. Of the 20 embedded structures, 6 refined to very low total error, giving total residual errors [sums of absolute values of (distance-bound)] of 2.68–3.66 Å. Table II shows the rms differences in non-hydrogen backbone atom positions of the six structures. These rms differences do not correlate with the total error; i.e., a higher total error does not neccessarily mean a larger rms difference. Approximately 24 h of CPU time was required to refine each structure to very low residual error.

Results and Discussion

Although apamin is rather small, $M_r \approx 2000$, giving an $\omega\tau_c$ value near 1, negative NOEs are observed, and by using somewhat longer than typical mixing times quantitative estimates of distances were obtained. There is some uncertainty in these distances due to the uncertainty in the slope of the mixing time dependence of NOEs. Some of the distances derived from NOE cross-peaks, particularly those that were intraresidue, were not used for the refinement since after experimental uncertainties were added they did not additionally constrain the molecule. NOEs that contribute significantly to the definition of the structure include sequential NOEs that define local backbone conformation, secondary structure-dependent NOEs, such as the $\alpha H(i)$–$\beta H(i + 3)$ in the helix, and tertiary NOEs between side chains, the most significant of

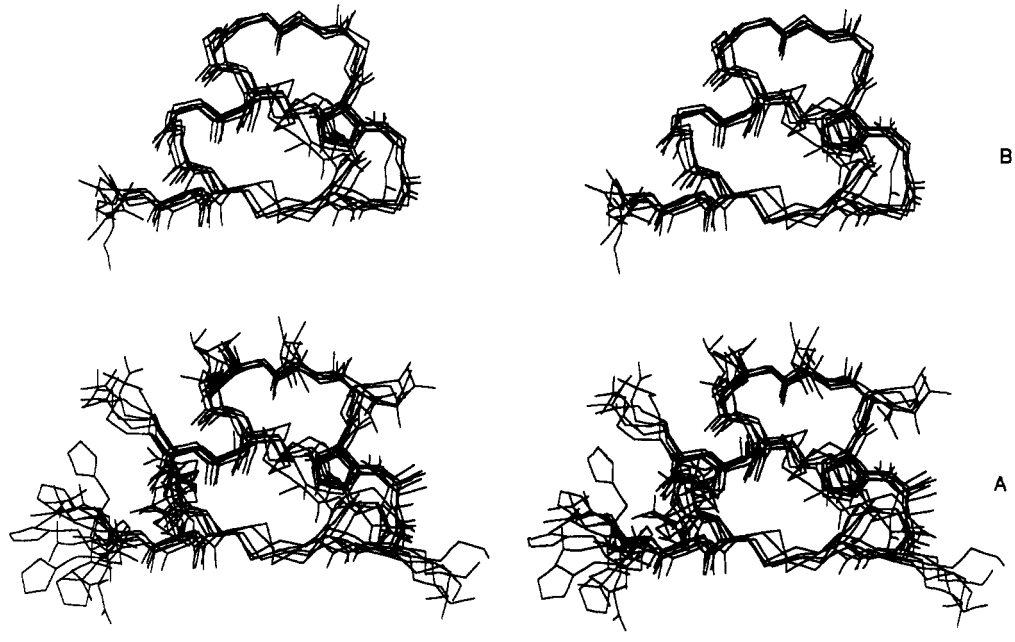

FIGURE 4: Stereoview of the six lowest error DSPACE structures showing (A) all atoms except hydrogens and (B) backbone atoms, disulfides, and the proline side chain. All structures were superimposed by using a least-squares fit of the backbone atoms to the lowest error DSPACE structure.

Table I: Distance Constraints for Apamin Obtained from NOESY Spectra[a,b]

sequential backbone		other		sequential backbone		other	
C-1				T-8			
αH(1)–NH(2)	1.90, 3.00			αH(8)–NH(9)	1.90, 3.00	αH(8)–γM(8)	2.71, 3.73[f]
N-2				βH(8)–NH(9)	1.90, 3.00	βH(8)–NH(10)	1.90, 3.00
αH(2)–NH(3)	1.90, 3.00	NH(2)–βH(2)	1.90, 3.00	γM(8)–NH(9)	1.95, 5.20[g]		
		NH(2)–β′H(2)	1.90, 3.00	A-9			
		αH(2)–βH(2)	2.41, 2.86[c]	NH(9)–NH(10)	1.90, 3.00	NH(9)–βM(9)	1.95, 2.90[f]
		αH(2)–β′H(2)	2.41, 2.86[c]			αH(9)–βM(12)	1.95, 5.50[g]
C-3						αH(9)–NH(12)	2.50, 3.50
NH(3)–NH(4)	2.50, 3.50	NH(3)–βH(3)	1.90, 3.50	L-10			
		NH(3)–β′H(3)	1.90, 3.50	NH(10)–NH(11)	2.50, 3.50		
		αH(3)–γH(6)	2.50, 6.00[d]	C-11			
		αH(3)–β′H(6)	2.50, 6.00	NH(11)–NH(12)	2.50, 3.50	αH(11)–βH(14)	2.43, 4.68
		βC(3)–βH(15)	2.20, 5.45[e]			αH(11)–β′H(14)	2.43, 4.68
		βC(3)–β′H(15)	2.20, 5.45[e]			αH(11)–NH(14)	2.50, 3.50
		αH(3)–βH(15)	2.50, 5.25[d]	A-12			
		αH(3)–β′H(15)	2.50, 5.25[d]	NH(12)–NH(13)	2.50, 3.50	NH(12)–βM(12)	1.95, 3.00[f]
K-4						αH(12)–βH(15)	2.55, 4.80
NH(4)–NH(5)	1.90, 2.50					αH(12)–β′H(15)	2.55, 4.80
A-5				R-13			
NH(5)–βH(6)	2.50, 3.50			NH(13)–NH(14)	2.50, 3.50	αH(13)–NH(16)	2.50, 3.50
αH(5)–βH(6)	2.25, 2.75			R-14			
αH(5)–β′H(6)	1.90, 2.46			NH(14)–NH(15)	2.50, 3.50		
P-6				C-15			
αH(6)–NH(7)	1.90, 3.00	γH(6)–αH(12)	2.55, 6.10	NH(15)–NH(16)	2.50, 3.50		
		γH(6)–αH(12)	3.03, 3.53	Q-16			
		βH(6)–αH(12)	3.04, 3.54	NH(16)–NH(17)	2.50, 3.50	αH(16)–βH(16)	2.44, 2.87[c]
		βH′(6)–αH(12)	2.55, 6.10			αH(16)–β′H(16)	2.44, 2.87[c]
E-7				Q-17			
NH(7)–NH(8)	2.50, 3.50			αH(17)–NH(18)	1.90, 3.00		

hydrogen bonds						disulfide constraints[h]	
NH(4)–βO(2)	2.00	NH(13)–O(9)	2.00	NH(16)–O(12)	2.00	βC(1)–βC(11)	3.72, 3.99
NH(5)–O(2)	2.00	NH(14)–O(10)	2.00	NH(17)–O(13)	2.00	βC(3)–βC(15)	3.72, 3.99
NH(12)–O(8)	2.00	NH(15)–O(11)	2.00				

[a] Atom names: αH = α-proton; NH = amide proton; βH and β′H = β-protons; γH and γH = γ-protons; δH and δ′H = δ-protons; γM and βM = methyl pseudoatoms; βC = β-carbon; and O = carbonyl oxygen. [b] Note: Not all NOESY peaks have a constraint listed since the resulting constraint covers the entire range of possible values. Constraints with amide protons are broken up into three categories: strong, 1.90–2.50 Å; medium, 1.90–3.0 Å; and weak, 2.5–3.5 Å. Van der Waals distance between two protons is 1.90 Å and between a methyl group and a proton is 1.95 Å. [c] Largest possible value. [d] Semiquantitative. [e] Constraint to βC is used since the NOE is to one β-proton. Upper bound = NOE distance constraint + βH–βC bond length (1.0 Å). [f] NOE distance constraint referenced to αH–βM distance from Ala 5 and Ala 12. [g] Treated as pseudoatom constraint: added 1.0 Å to upper bound (Wüthrich et al., 1983). [h] Distance constraint to give a χ(βC–S–S–βC) angle of ±90 ± 10° (Richardson, 1981).

which for apamin is the contact between the Pro 6 side chain and the αH of Ala 12.

The structures obtained from distance geometry by using the constraints listed in Table I are pictured in Figure 4. The structures were superimposed by choosing the lowest error structure as the reference and then least-squares minimizing deviations between backbone non-hydrogen atom positions. Total errors and rms differences of the six structures are given in Table II. None of the structures contained individual residual violations of greater than 0.05 Å for any of the constraints. It is clear that the backbone of the peptide is fairly well defined, especially within the α-helix. This helix is highly constrained by the combination of NOEs, hydrogen bonds, and disulfide bonds. The remaining backbone is not as precisely defined, but all of the structures have very similar structural features. There is significant variation in the geometry of the disulfide bonds, which leads to rearrangement of the position of the backbone atoms in the β-turn region as well. The degeneracy of the Cys 15 β-protons makes the NOEs to the Cys 3 α- and β-protons relatively weak constraints. The chemical shifts of the β-protons of Cys 1 and Cys 11 are so similar that even if NOEs were present they could not be assigned. Thus, the exact conformations of the disulfides are relatively poorly determined. The positions of other amino acid side chains could not be determined accurately since there were few NOEs involving them. This is not surprising since the side chains are most likely mobile in the solution. More

Table II: Errors and rms Differences in Non-Hydrogen Backbone Atom Positions

DSPACE structure	error[a] (Å)	rms difference (Å) for DSPACE structure 1	2	3	4	5
1	2.68					
2	3.66	0.95				
3	2.78	0.95	0.98			
4	2.68	0.98	1.05	1.00		
5	2.91	1.04	1.23	1.03	1.34	
6	3.36	1.06	0.91	0.95	0.98	1.02

[a] Sum of the absolute value of all deviations from the smoothed bounds matrix.

NOEs were seen at low temperature, possibly because of an increase in rotational correlation time for the whole molecule and from a decrease in conformational averaging, but all were still within single sidechains.

The structures in Figure 4 are generally consistent with the previously proposed model (Wemmer & Kallenbach, 1983), with the biggest differences occurring between the β-turn and the α-helix, residues 6–8. It was previously suggested that the amide of residue 8 was hydrogen bonded to the carbonyl of residue 5, explaining the slow exchange of 8. While this was stereochemically reasonable from models, the new constraints involving residues 6 and 12 cannot be met and simultaneously have this hydrogen bond present. Attempts to include both constraints [the NH(8)–O(5) hydrogen bond was defined in

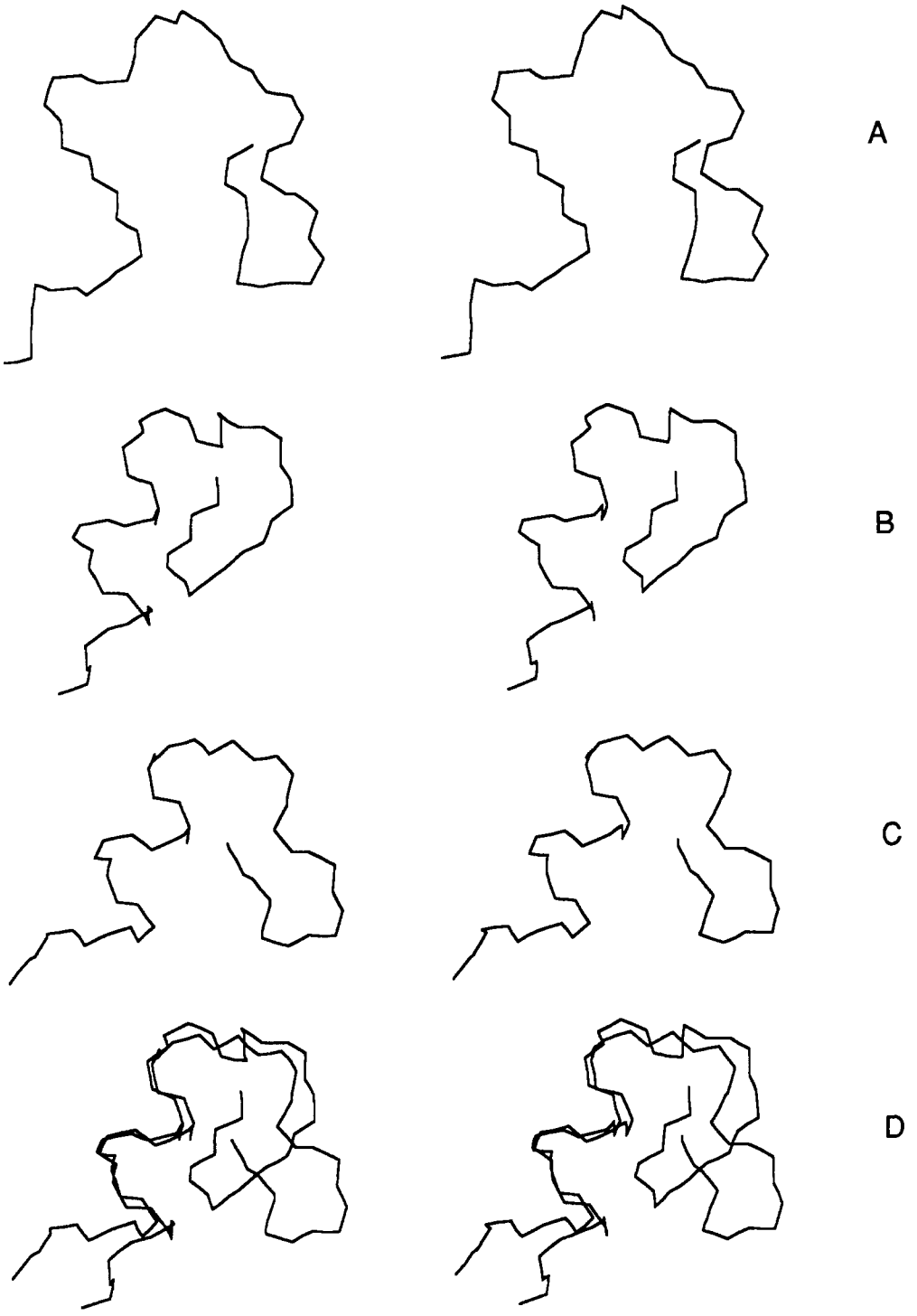

FIGURE 5: Comparison of energy-refined structures (Freeman et al., 1986) to the lowest error DSPACE structure showing only backbone atoms. The starting model for (A) is from ab initio calculations (Popov & Melnikov, 1979, 1980). The starting model for (B) is from Hider and Ragnarsson (1980). (C) is the lowest error DSPACE structure. (D) is model B superimposed on model C by using only the backbone atoms of residues 9–15 (α-helical region).

the same way as the other hydrogen bonds] during refinement always led to much higher residual error, specifically arising from violations in this segment of the molecule. When the hydrogen-bond constraint was removed, the error was significantly reduced and was approximately equally distributed throughout the molecule. Dempsey (1986) has suggested that the amide of residue 11 is included in the α-helix hydrogen bonding at low pH, thus explaining its relatively low exchange rate. However, the NOEs involving residue 7 (the hydrogen-bond acceptor required to extend the helix to include the amide of 11) and the NOEs in the 6–8 loop cannot be satisfied if this hydrogen bond is present. It appears more likely that the side chain of Thr 8 is in a position to interact with the amide of 11 and could slow its exchange. At higher pH some interaction with the side chain of Glu 7 could explain the change in exchange rate of this residue. Dempsey also suggested that a different structure exists at higher pH. We tested this idea by running a NOESY at pH 4 under conditions similar to those used at pH 2 (30 °C, t_m = 450 ms). The sequential connectivity pattern was the same, and the side chain of Pro 6 was still in close proximity to αH of Ala 12. This suggests the structure is generally the same as at pH 2

and no major structural change occurs at high pH. However, the amide protons of Glu 7 and Asn 2 shift downfield dramatically at the higher pH, indicating some sort of change. The salt bridge proposed by Okhanov and co-workers (1980) between the side-chain carboxylate of Glu 7 and the α-amino group could be present in the calculated structures since the two groups are in fairly close proximity. However, there are not a sufficient number of NOEs involving protons on the side chains of Glu 7 and Cys 1 to definitively demonstrate that it is present. The strongest evidence for it is the titration behavior of amide chemical shifts for residues 2 and 7, as described by Bystrov et al. (1980).

We have also compared the distance geometry structures with the models based on energy refinement proposed by Freeman et al. (1986) (see Figure 5). The coordinates for these models were calculated from the reported angles. These structures were developed by using energy refinement of a number of different starting models. The fact that these models converge locally but do not evolve into the same structure is a demonstration of the well-known problem of finding the global energy minimum in such calculations. Among the structures proposed, the refined model of Hider and Ragnarson (1980) is in best agreement with the distance geometry structures (3.42 ± 0.18 Å rms difference between backbone atom positons), but several differences remain. The largest differences occur in the N-terminal half of the molecule. Their model has a turn that involves residues 1–4, instead of 2–5, and then an extended region that puts the Pro 6 side chain too far away from the αH of Ala 12 to be consistent with our data. The C-terminal α-helix is very similar to our structure (see Figure 5D).

Arginines 13 and 14, which have been found to be important for activity (Granier et al., 1978; Cosand & Merrifield, 1977), are both in the helix, but again the lack of NOEs involving side-chain protons made it impossible to determine the exact conformation of their side chains. In fact, the lack of NOEs strongly suggests that both side chains are mobile in solution. The lack of activity for the reduced peptide with blocked disulfides suggests that some aspect of the positioning of these groups is important.

The distance geometry models were also checked against the previously reported $^3J_{NH-\alpha H}$ and $^3J_{\alpha H-\beta H}$ coupling constants (Bystrov et al., 1980). This was done for $^3J_{NH-\alpha H}$ by defining distance constraints between the amide protons and α-protons and between the amide proton and carbonyl carbon, consistent with the ϕ angles derived from the Karplus relationship (Wüthrich, 1986). The distance geometry structures that had been previously refined without these constraints essentially did not change upon their addition. There were only very minor refinements required to lower the residual error to its previous value. This is not surprising since the sequential assignments define helical (small $^3J_{NH-\alpha H}$) and extended regions (large $^3J_{NH-\alpha H}$) very well, making coupling constant information redundant. The $^3J_{\alpha H-\beta H}$ coupling constant was only useful for residues 2, 3, and 6 since the β-protons of all the other residues are degenerate or near degenerate. Distances were calculated between the α-proton and β-protons by using a Karplus relationship (De Marco et al., 1978), and all were found to be consistent with the refined structures.

As has been found with other distance geometry structures of peptides, it is the NOEs between residues far apart in the sequence that have the largest effect in determining the global fold of the peptide (Havel & Wüthrich, 1985). The secondary structure of peptides can be determined quite accurately by using sequential assignment patterns and observation of slowly exchanging amides (Wüthrich, 1986; Wüthrich et al., 1984; Wemmer & Kallenbach, 1983). Information from coupling constants has little effect on the global structure but can be used to help define local structure. In this case, as has been found for other proteins, it is unlikely that more accurate distances would significantly improve the structure. Havel and Wüthrich (1985) have found that the global fold of BPTI can be defined using a large number of relatively inaccurate medium- and long-range distance constraints. However, in this case, the small size of apamin means that there are relatively few side-chain contacts; probably all have already been identified.

The structure of apamin defined here through extensive NMR measurements and distance geometry calculations should be quite useful in testing a variety of modeling calculations, understanding the amide exchange from segments of secondary structure, and understanding the stabilization of such structures by disulfide bonds.

ACKNOWLEDGMENTS

We thank Dr. Dennis Hare of Hare Research, Woodenville, WA, for providing copies of the DSPACE program and for his help in using it for this work. We also thank Prof. Neville Kallenbach, New York University, for providing the sample of apamin used in this work and for his continued enthusiasm for pursuing these studies.

REFERENCES

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

Anil-Kumar, Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654–3658.

Banks, B. E. C., Brown, C., Burgess, G. M., Burnstock, G., Claret, M., Cocks, T. M., & Jenkinson, D. H. (1974) *Nature (London) 282*, 415–417.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Bystrov, V. F., Arseniev, A. S., & Gavrilov, Yu. D. (1978) *J. Magn. Reson. 30*, 151–184.

Bystrov, V. F., Okhanov, V. V., Miroshnikov, A. I., & Ovchinnikov, Yu. A. (1980) *FEBS Lett. 119*, 113–117.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 47–148.

Cosand, W. L., & Merrifield, R. B. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2771–2775.

Crippen, G. M. (1981) *Distance Geometry and Conformational Calculations*, Research Studies (Wiley), New York.

De Marco, A., Llimas, M., & Wüthrich, K. (1978) *Biopolymers 17*, 637.

Dempsey, C. E. (1986) *Biochemistry 25*, 3904–3911.

Drobny, G., Pines, A., Sinton, S., Weitekamp, D., & Wemmer, D. (1979) *Symp. Faraday Soc. 13*, 49–55.

Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521–655.

Freeman, C. M., Catlow, C. R. A., Hemmings, A. M., & Hider, R. C. (1986) *FEBS Lett. 197*, 289–296.

Granier, C., Podroso Muller, E., & Van Rietschoten, J. (1978) *Eur. J. Biochem. 82*, 293–299.

Habermann, E. (1972) *Science (Washington, D.C.) 177*, 314–322.

Hare, D. R., Ribeiro, N. S., Wemmer, D. E., & Reid, B. R. (1985) *Biochemistry 24*, 4300–4306.

Havel, T. F., & Wüthrich, K. (1984) *Bull. Math. Biol. 46*, 673–698.

Hider, R. C., & Ragnarsson, U. (1980) *FEBS Lett. 111*, 189–193.

Hider, R. C., & Ragnarsson, U. (1981) *Biochim. Biophys. Acta 677*, 197-208.
Jeener, J., Meier, B. H., Bachmann, P., & Wüthrich, K. (1979) *J. Chem. Phys. 71*, 4546-4553.
Miroshnikov, A. I., Elyakova, E. G., Kudelin, A. B., & Senyavina, L. B. (1978) *Sov. J. Bioorg. Chem. (Engl. Transl.) 4*, 746-752.
Okhanov, V. V., Afanasev, V. A., Gurevich, A. Z., Elyakova, E. G., Miroshnikov, A. I., Bystrov, V. F., & Ovchinnikov, Yu. A. (1980) *Sov. J. Bioorg. Chem. (Engl. Transl.) 6*, 840-860.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187-193.
Popov, E. M., & Melnikov, P. N. (1979) *Bioorg. Khim. 5*, 828-847, 1011-1024, 1471-1593.
Popov, E. M., & Melnikov, P. N. (1980) *Bioorg. Khim. 6*, 21-30.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167-339.
Wemmer, D., & Kallenbach, N. R. (1983) *Biochemistry 22*, 1901-1906.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949-961.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715-740.

Origin of DNA Helical Structure and Its Sequence Dependence†

A. Sarai,‡ J. Mazur,§ R. Nussinov,‡,∥ and R. L. Jernigan*,‡

Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Advanced Scientific Computing Laboratory, Program Resource Inc., NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, and Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, 69978 Israel

Received June 3, 1988; Revised Manuscript Received July 7, 1988

ABSTRACT: Conformational analysis of DNA shows that the origin of the B-form double helix can be attributed in large part to the atomic charge pattern in the base pairs. The charge patterns favor specific helical stacking of the base pairs. Base pairs alone—without backbones—have a strong tendency to form helix, indicating that the backbones play a rather passive role in determining the basic helical structure of DNA. It is mainly the electrostatic interactions determined by the charge pattern on base pairs that stabilize a particular helical conformation. The charge pattern in the base pairs appears to be responsible for much of the sequence dependence of DNA conformation, rather than steric clashes.

Double-helical DNA plays the essential role of storing genetic information in a stable form, and its stability affects the control of gene expression. The classical B-form DNA consists of a right-handed double helix with two polynucleotide strands twisted about each other. The successive base pairs stack on top of each other with a relative rotation near 36°. The double-stranded structure is stabilized by complementary hydrogen bonding, by stacking the flat surfaces of the base pairs, and by exposing the polar edges and phosphates to solvent. Although the structures of DNA have been studied in some detail (Arnott, 1976; Saenger, 1984; Dickerson et al., 1985; Shakked & Kennard, 1985), it is not well understood why DNA is stabilized in a particular helical conformation. Hydrogen bonding between bases leads to double-stranded forms but not to a specific type of helix. Within the major families, B, A, Z, etc., and depending upon the base sequence, DNA exhibits smaller scale local conformational variations (Dickerson et al., 1985; Shakked & Kennard, 1985), from small base pair orientational changes to DNA bending (Koo et al., 1986; Hagerman, 1986). Implications for biological function have recently been a focus of intensive research (Bossi & Smith, 1984; Ryder et al., 1986; Snyder et al., 1986; Zhan & Blattner, 1987). However, the physical reasons for this conformational polymorphism are not clearly understood, partly because interactions in such a macromolecule are very complex. To date, the most popular model of the sequence dependence of B-form variants was proposed by Calladine (1982), who stated that the steric clash between base pairs may be responsible for the conformational variation. In order to identify the important interactions responsible for the stabilization of particular conformations of a complex macromolecule like DNA, it is useful to first consider separate components, base pairs and backbones. In this paper, we analyze the interactions between base pairs to investigate the origin of DNA conformations and its sequence dependence in terms of chemical structure of base pairs and their interactions, and show that the DNA helix and its sequence-dependent variation are a direct consequence of the atomic charge pattern in the base pairs, and their interactions.

Method of Calculation

Our method is a combination of energy calculations and conformational statistics, developed especially for studying DNA conformations (Sarai et al., 1988). By generating large numbers of conformations, we estimate the conformational free energy. This method is particularly suitable for DNA, which exhibits unusually flat energy potentials with many minima, a problem that is not amenable to treatment with usual methods.

Consider the two base pairs shown in Figure 1. The coordinate system we have used to describe their geometry is as follows. The X axis connects atoms C8 and C6 in the base pair and is oriented as shown. The Y axis points along the helix direction from the center of the line between C8 and C6.

† Research sponsored, at least in part, by the NCI, DHHS, under Contract NO1-CO-74102 with PRI.
‡ National Cancer Institute.
§ NCI-Frederick Cancer Research Facility.
∥ Tel-Aviv University.