Ab Initio Geometry Determinations of Proteins. 1. Crambin

Christian Van Alsenoy,*,† Ching-Hsing Yu,‡ Anik Peeters,† Jan M. L. Martin,§ and Lothar Schäfer‡

Department of Chemistry, Universitaire Instelling Antwerpen, B2610 Wilrijk, Belgium, Department of Chemistry, University of Arkansas, Fayetteville, Arkansas 72701, and Department of Organic Chemistry, Weizmann Institute of Science, IL-76100 Rehovot, Israel

Received: November 19, 1997; In Final Form: January 19, 1998

The geometry of crambin, a protein with 46 residues, was determined by ab initio HF/4-21G geometry optimization. The results are compared with the crystal structure of the compound and with HF/4-21G ϕ,ψ-conformational geometry maps calculated for the model dipeptide *N*-acetyl-*N*′-methylalaninamide. Root-mean-square (rms) deviations between calculated and crystallographic backbone structural parameters are 1.5° for N−C(α)−C′ and 0.013 and 0.017 Å, respectively, for N−C(α) and C(α)−C′. In the case of N−C(α)−C′ the rms deviations are small compared to the observed range of values, which is from <108° to >118°, confirming a definite conformational dependence of peptide backbone structural parameters on ϕ and ψ. In contrast, the deviations in bond lengths are of the same magnitude as the overall variations. The considerable nonplanarity of the peptide units found in the crystal structure is well reproduced by the calculations. When the calculated and crystal structures are superimposed, the rms positional deviation is 0.6 Å for the heavy atom framework and 0.4 Å for the backbone chain. The phenomenon of helix compression is confirmed that is found in elongated helical chains compared to isolated residues or smaller oligomers.

Introduction

Ab initio geometry optimizations of peptides are a valuable source of information on the structural properties of peptides and proteins. Even the earliest studies[1,2] of this kind, involving HF/4-21G[3] gradient[4,5] geometry optimizations of the *N*-acetyl-*N*′-methyl derivatives of glycine[1] and alanine[2] revealed unexpected details of the flexibility of peptide geometries in close agreement with the crystal structures of oligopeptides[6,7] and proteins.[8−10] In the meantime many similar studies have been performed[11−50] employing various basis sets and ab initio computational techniques.

While the earliest ab initio geometry optimizations typically involved model dipeptides,[1,2,11−42] only much later followed by oligopeptides,[43−50] recent advances in both computer hardware and software make it now possible to refine the structures of entire proteins. As a first example the complete HF/4-21G geometry refinement of crambin is reported in this paper. Crambin is a small hydrophobic protein with 46 residues that was selected for this study because its crystal structure has been determined[51−53] with high resolution. It is the purpose of this paper to describe details of the computations and to present the results of comparisons between the ab initio structural trends and the experimental structure[51−53] of crambin, and between the protein and HF/4-21G structural geometry maps recently calculated for the model dipeptide, *N*-acetyl-*N*′-methylalaninamide.[8]

Methods

The structure of crambin (642 atoms) was refined using the MIA approximation[54,55] for the SCF procedure and a standard[4] approach for the calculation of the gradients. In the MIA procedure[54,55] products of two basis functions, as they occur in the SCF formalism, are expanded in terms of a set of auxiliary functions. Applying this expansion to the charge distribution in a two-electron integral reduces the formal N^4 dependence of the Fock matrix to N^3, where N is the number of basis functions. Since the expansion in terms of auxiliary functions is not exact, those electron repulsion integrals for which the concomitant error exceeds a preset threshold are systematically corrected, so that the final results are identical with those obtained from the conventional SCF procedure. When the MIA approximation is implemented in combination with the direct SCF approach,[56] the resulting method scales linearly with system size[57] in building the Fock matrix. Linear scaling was previously achieved only by a quantum mechanical tree code[58,59] or an algorithm based on fast multipole methods.[60]

The geometry optimization of crambin was performed using a modified version of the normal coordinate force relaxation (NCFR) procedure by Sellers et al.[61] Based on the normal coordinate program by Gwinn[62] and its modification by Sellers et al.,[63] which both were designed for the specific purpose of using redundant internal coordinates to generate normal coordinates, the NCFR is the prototype of geometry optimization schemes in which redundant internal coordinates are used to relax forces along uncoupled coordinates that can be constructed in a fully automatic way. The advantages of using redundant internal coordinates were recently rediscovered by Pulay and Fogarasi,[64] who significantly improved the procedure by implementing GDIIS techniques and curvilinear displacement coordinates.[64]

In the current calculations the 4-21G basis set[3] was used for all first-row elements, and the 3-321G basis[65] for sulfur, yielding a computational problem with 3597 basis functions. To our knowledge crambin is the largest system whose structure has

† Universitaire Instelling Antwerpen.
‡ University of Arkansas.
§ Weizmann Institute of Science.

S1089-5639(98)00260-6 CCC: $15.00
Published on Web 02/28/1998

been refined at the ab initio level. Each cycle of the geometry optimization required approximately 80 h of CPU time on a DEC/Alpha Station 600 (5/266) with 256 MB memory, and 79 cycles were completed for the results of this study. A larger single-point SCF/3-21G energy calculation was recently performed on P53[58] with 698 atoms and 3836 basis functions. Since P53 has an extended ribbonlike structure while that of crambin is globular, the latter is burdened by a larger number of nonnegligible overlap distributions compared to the former.

Atomic coordinates of crambin taken from the PDB1cnr file[51] of the Brookhaven Protein Data Bank[66] were used to start the geometry optimizations. At the current level of refinement, all internal coordinates of crambin are essentially relaxed. Changes between two successive cycles amount to 0.0002 Å for bond lengths and <0.5° for valence and torsional angles. Changes in hydrogen bond lengths amount to 0.007 Å, and 0.5° and 1.5° for associated bond and torsional angles, respectively. These criteria of relaxation are less stringent than those that we customarily apply to small molecules. However, the resulting structure is sufficiently optimized to allow for meaningful comparisons with other structural data. The structural parameters needed for these comparisons were calculated from the Cartesian coordinates using the MSI/BIOSYM molecular modeling software suite.[67]

Results and Discussion

Structural parameters characteristic of the HF/4-21G geometry of crambin are listed in Table 1. They include for each residue *i*, where $i = 2-45$, the backbone parameters N−C(α), C(α)−C′, N−C(α)−C′, the torsional angles ϕ(N−C(α)), ψ(C(α)−C′), and ω(N−C′). The full set of optimized Cartesian coordinates is included in Table 2, which is deposited as Supporting Information.

Some time ago Cao et al.[6−9] described a program that calculates the HF/4-21G backbone bond lengths and angles of the model dipeptide *N*-acetyl-*N*′-methylalaninamide at any arbitrary point in its ϕ,ψ-space. Using this program, the HF/4-21G dipeptide values of N−C(α), C(α)−C′, and N−C(α)−C′) were calculated for the current study at the ϕ and ψ angles of the residues of crambin obtained by the HF/4-21G geometry refinement of the entire molecule. Both sets, the HF/4-21G dipeptide results and the HF/4-21G whole-molecule crambin results, are compared in Figures 1a−3a.

Similar comparisons, but between the HF/4-21G parameters of crambin and the crystal structure,[51−53] are given in Figures 1b to 3b. The pdb1cnr file was selected from the Brookhaven Protein Data Bank[66] for these comparisons, even though more recent crystallographic studies of crambin are now available,[52,53] because the starting geometry of the ab initio calculations was constructed from this molecular form. Finally, a comparison between the ab initio and crystallographic ω(N−C′) torsional angles is given in Figure 4 and a superposition of the HF/4-21G and crystallographic coordinates is shown in Figure 5.

Helix Compression. For the α_R-helical regions, residues 7−16 and 23−29, the average value of the N−C(α)−C′ angles in the HF/4-21G dipeptide structures (112.6°) is 1.8° larger than the average angle (110.8°) in the HF/4-21G structure calculated for the molecule as a whole. The two values demonstrate helix compression[6−10] in the polymer chain compared to the single (dipeptide) residue. That is, due to cooperative effects, the N−C(α)−C′ angles in elongated α-helices are compressed compared to single residues. Indeed, the average value of N−C(α)−C′ in the same α_R-helical regions in the crystallographic structure[51−53] is 111.9°, i.e., less than the dipeptide value (112.6°), but not less than the average HF/4-21G value calculated for crambin as a whole (110.8°). The special feature of the α_R-helical regions is also apparent from the fact that the average values of the N−C(α)−C′ angles in the nonhelical parts of crambin are 110.7°, 111.2°, and 112.1° for the HF/4-21G dipeptide, HF/4-21G crambin, and the crystal structure, respectively. That is, the opposite trend is found between the dipeptide and protein geometries.

TABLE 1: Structural Parameters of the ab Initio HF/4-21G Optimized Geometry of Crambin[a]

res	ϕ	ψ	N−C(α)−C′	N−C(α)	C(α)−C′	ω
2	−107.71	152.24	108.56	1.4546	1.5184	179.86
3	−134.81	131.04	105.64	1.4557	1.5357	−176.03
4	−118.33	151.94	110.02	1.4563	1.5337	−175.73
5	−81.51	−21.32	112.28	1.4777	1.5195	−178.72
6	−149.32	164.67	109.95	1.4480	1.5179	−177.00
7	−65.23	−40.25	108.80	1.4662	1.5272	−179.78
8	−59.52	−46.30	109.61	1.4632	1.5306	178.91
9	−57.79	−44.55	110.10	1.4606	1.5322	179.31
10	−61.54	−43.79	110.56	1.4656	1.5246	179.38
11	−56.77	−47.03	110.14	1.4566	1.5305	−177.51
12	−63.72	−40.69	110.46	1.4586	1.5357	174.82
13	−55.85	−48.33	109.97	1.4623	1.5267	−176.84
14	−59.47	−45.46	110.42	1.4550	1.5322	178.14
15	−59.30	−47.38	109.82	1.4651	1.5343	−179.42
16	−53.91	−36.52	111.88	1.4587	1.5310	−176.25
17	−79.05	−11.89	114.59	1.4535	1.5293	−173.13
18	−56.90	−38.11	114.25	1.4637	1.5356	178.09
19	−92.91	11.93	113.07	1.4691	1.5308	175.21
20	101.16	5.99	113.87	1.4464	1.5229	−171.51
21	−52.08	132.63	109.46	1.4478	1.5232	176.25
22	−56.99	144.22	112.19	1.4594	1.5243	−173.39
23	−60.19	−32.80	111.55	1.4604	1.5187	−175.72
24	−57.30	−41.43	111.81	1.4655	1.5333	−179.72
25	−75.12	−34.44	110.92	1.4659	1.5278	175.72
26	−66.74	−30.13	110.88	1.4497	1.5408	176.95
27	−64.92	−50.38	110.62	1.4663	1.5290	−178.00
28	−68.72	−23.23	113.12	1.4554	1.5369	174.02
29	−69.12	−35.46	112.37	1.4613	1.5365	−164.95
30	−113.79	−18.88	117.22	1.4515	1.5281	−166.69
31	82.41	7.35	113.82	1.4497	1.5306	169.91
32	−77.15	155.82	109.13	1.4453	1.5421	−176.82
33	−125.63	160.82	109.47	1.4519	1.5273	179.87
34	−115.15	129.82	106.69	1.4670	1.5298	174.69
35	−130.30	155.73	108.91	1.4495	1.5244	178.12
36	−81.89	−17.11	113.30	1.4714	1.5268	175.80
37	−81.34	−177.00	111.59	1.4423	1.5317	−176.10
38	−115.38	6.98	113.00	1.4689	1.5356	−178.97
39	−123.61	124.75	107.00	1.4569	1.5247	177.03
40	−88.09	155.56	109.38	1.4517	1.5391	168.09
41	−76.72	166.46	109.01	1.4623	1.5319	−171.53
42	−68.56	−23.49	115.69	1.4605	1.5329	178.38
43	−81.89	−12.30	112.47	1.4546	1.5364	179.66
44	−133.34	50.90	107.55	1.4498	1.5312	−176.89
45	−103.71	11.12	113.01	1.4597	1.5331	166.44

[a] For each residue the torsional angles ϕ(N−C(α)), ψ(C(α)−C′), and ω(N−C′) and the backbone parameters N−C(α)−C′, N−C(α), and C(α)−C′ are given. All lengths are in Å; all angles in deg. Column "res" lists the residue number of the crystal structure.[51−53]

In previous evaluations[68,69] of HF/4-21G structures, geometrical trends in organic functional groups of the kind considered here (i.e., conformational differences between parameters of the same type) were found accurate at the level of several tenths of a degree. Thus, the helix compression effect, ∼1.5°,[9,10] is significantly above the error limits of the computational procedures. A similar phenomenon, β-expansion, is also found in polypeptide systems,[6−9] but crambin does not have a sizable fragment in this exact conformational region that would clearly illustrate this effect.

In general, when the HF/4-21G dipeptide results are compared with the HF/4-21G crambin results (Figures 1a−3a), it is seen

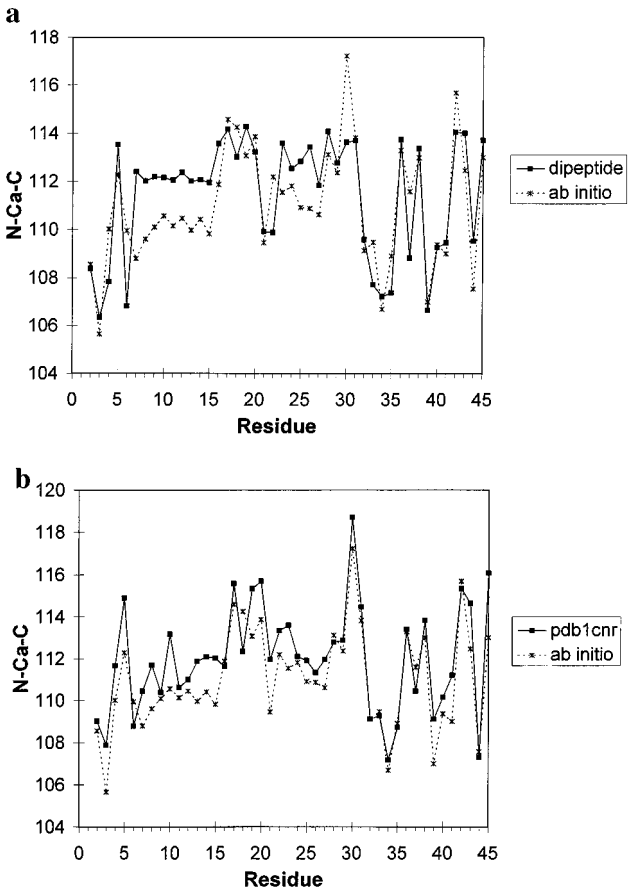

Figure 1. (a) Comparison of HF/4-21G optimized dipeptide and crambin N−C(α)−C′ backbone angles. The dipeptide values were calculated at the HF/4-21G ϕ,ψ-torsions of crambin as described in the text. (b) Comparison of crystallographic (pdb1cnr)[51] and HF/4-21G optimized N−C(α)−C′ angles of crambin.

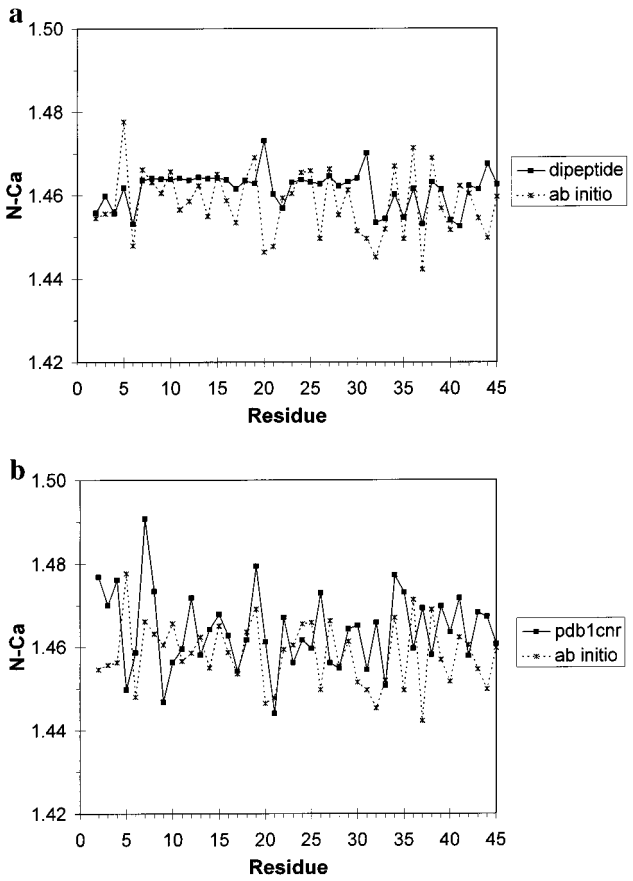

Figure 2. (a) Comparison of HF/4-21G optimized dipeptide and crambin N−C(α) bond lengths. The dipeptide values were calculated at the HF/4-21G ϕ,ψ-torsions of crambin as described in the text. (b) Comparison of crystallographic (pdb1cnr)[51] and HF/4-21G optimized N−C(α) bond lengths of crambin.

that the N−C(α)−C′ angles in the two sets follow rather similar trends, in contrast to the bond lengths, N−C(α) and C(α)−C′. The same is found (Figure 1b) when the whole-molecule HF/4-21G parameters of crambin are compared with the crystallographic values.[51−53] The root-mean-square deviation between the N−C(α)−C′ angles in the HF/4-21G dipeptide and whole-molecule crambin structures is 1.7°, and it is 1.5° between the latter and the crystallographic structure. These deviations are small compared to the full range of values observed for N−C(α)−C′,[51−53] which is from <108° to >118°. The results confirm the conclusion put forth previously[1,2,6−10] that there is a definite conformational dependence of peptide backbone structural parameters on ϕ and ψ.

The trends found for the various sets of the N−C(α) and C(α)−C′ bond lengths (Figures 2a, 3a, 2b, and 3b) are not in close agreement. The rms deviations are 0.009 and 0.007 Å, respectively, between the HF/4-21G whole-molecule crambin and dipeptide parameters and 0.013 and 0.017 Å, respectively, between the HF/4-21G crambin and crystallographic results. In these cases the deviations are large compared to the overall parameter changes, and moreover, as seen from Figures 2a, 3a, 2b, and 3b, the different sets display entirely diverging trends. We take these results to mean that the bond lengths are more sensitive than the bond angles to changes in electronic effects which are encountered on going from a single residue to a polymer, and from an isolated molecule to a system in the crystal environment. The difference, specifically, between the dipeptide and the polymer points to delocalization effects which are active in the elongated chain but not in the single residue. Since the bond lengths are among the most optimized parameters of the ab initio structure of crambin (largest residual force in any bond is 0.003 mdyn, corresponding to a further improvement of ~0.0005 Å), they are essentially relaxed at the chosen point in conformational space. Thus, with near certainty the discrepancies noted above are not an artifact of refinement. At the same time, comparisons with the crystal structure are affected by experimental error and by the fact that the ab initio bond lengths are refined at torsional angles that are not exactly the same as those of the crystal structure.

Nonplanarity of Peptide Groups. In a recent review[70] MacArthur and Thornton described a survey of peptide ω(N−C′) torsional angles taken from crystallographic data. They found that substantial deviations from planarity can be found that arise both from pyramidalization at the amino nitrogen atom and from a twist about the peptide bond. It is seen from Figure 4 that the significant deviations of ω angles from planarity that are found in the crystal structure of crambin[51−53] are well reproduced by the HF/4-21G calculations.

HF/4-21G calculations are not the most accurate means of simulating deviations from planarity for peptide groups. Subtle pyramidalization effects at the nitrogen atom or small twists of the ω angle from 180° are without any doubt affected by polarization functions and electron correlation effects. However, when the deviations from planarity are significant, HF/4-21G geometry optimizations provide a good first-order estimate.

Indeed, it is seen from Figure 4 that the calculated and crystallographic ω angles are generally in good agreement, with exceptional deviations in a small number of isolated places, i.e.,

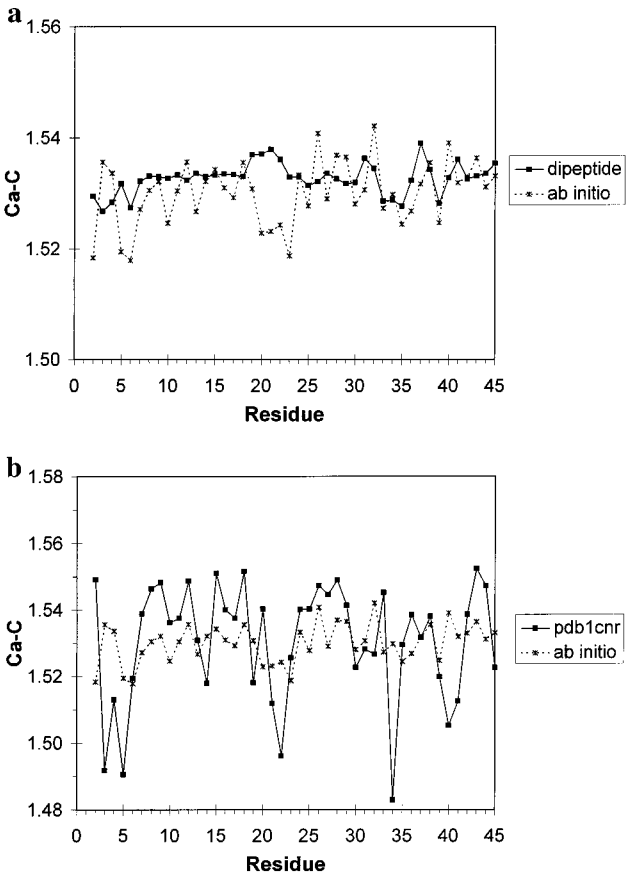

Figure 3. (a) Comparison of HF/4-21G optimized dipeptide and crambin C(α)−C′ bond lengths. The dipeptide values were calculated at the HF/4-21G ϕ,ψ-torsions of crambin as described in the text. (b) Comparison of crystallographic (pdb1cnr)[51] and HF/4-21G optimized C(α)−C′ bond lengths of crambin.

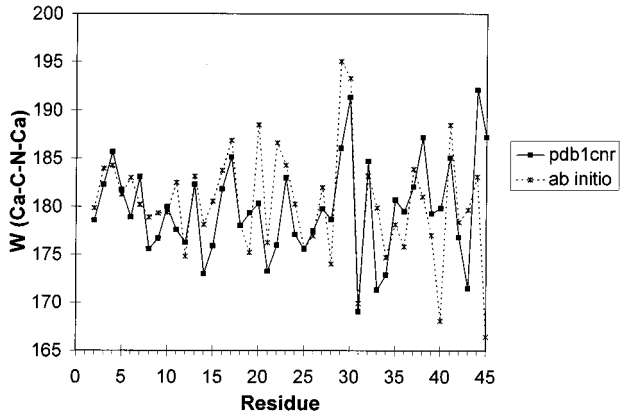

Figure 4. Comparison of crystallographic (pdb1cnr)[51] and HF/4-21G optimized ω-torsional angles of crambin.

in the vicinity of residues 19−22, 29, and 40−45. The latter are located in a part of the molecule, beyond residue 34, where the main chain temperature factors are at a peak[51] so that the molecular structure is less well defined. We will exclude this part of the molecule from our further analysis.

The deviations at residues 20 and 29 may be related to the fact that both are at the end of an area of maximum solvent accessibility.[51] In particular, the six-membered ring of Tyr-29 is highly exposed to the crystal environment, and it is possible that ω_{29} is significantly affected by intermolecular interactions.

Residues 19−21 form one of the five turns of crambin.[51] Hydrogen bond-like interactions exist between N20−H and O═C17 (∼2 Å) and between N25−H and C22═O (∼2.3 Å),

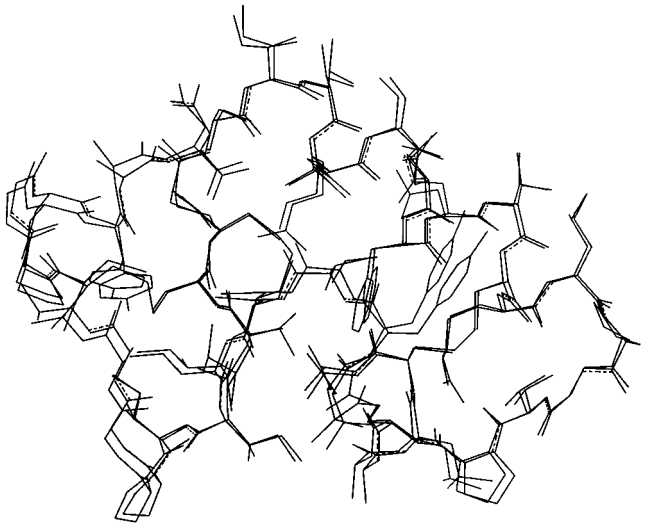

Figure 5. Superposition of the crystallographic (pdb1cnr)[51] and HF/4-21G optimized heavy-atom framework of crambin.

and each of the corresponding bends involves a proline residue (19 and 22). It is possible that such a complex structural sequence is not correctly modeled by the HF/4-21G calculations. Interestingly, unusual discrepancies around residue 20 are also found for N−C(α) and C(α)−C′, comparing the dipeptide and protein values (Figures 2a and 3a) and the ab initio and crystallographic parameters (Figures 2b and 3b). The mere presence of proline very likely is not the cause of the discrepancies noted above, since the calculated and experimental ω angles are in good agreement (Figure 4) in the vicinity of Pro-5.

Apart from these deviations, the calculated and experimental ω angles follow a rather similar pattern. At the beginning of the protein chain, the values start in the vicinity of 180°, proceeding to <180° in the vicinity of residue 15, and rising above 180° at residue 17. Most striking is the turn from >190° to <170° in the vicinity of residue 30, which is displayed by both the calculated and the experimental structures. Overall it seems that the calculated values are somewhat more flexible in either direction than the experimental deviations from planarity, which is in agreement with the absence of polarization functions in the HF/4-21G basis set.

In the vicinity of disulfide bridges (between Cys-16 and Cys-26, Cys-3, and Cys-40, and between Cys-4 and Cys-32), the calculated and experimental ω angles are in close agreement except for the high-amplitude end of the molecule.

In agreement with the closeness of calculated and experimental parameters, the overall topologies of the experimental and calculated structures of crambin are rather similar. A superposition of the HF/4-21G and crystallographic heavy atom frameworks is shown in Figure 5. The rms positional deviation is 0.6 Å for the heavy atom framework and 0.4 Å for the backbone chain.

Conclusions

During the early 1980s ab initio geometries were for the first time calculated with sufficient accuracy to make the resulting structures useful in experimental conformational analyses, mainly by gas electron diffraction and microwave spectroscopy,[71−76] providing efficient constraints of data analysis in cases in which the experimental observations did not afford the complete resolution of all structural parameters. In current protein crystallography, too, it is frequently not possible to obtain

atomic resolution diffraction data, and restraints taken from libraries of ideal geometries[77] are an integral part of data analysis. In this area of application it is an advantage that ab initio geometries can now be determined of fragments which are large enough to provide information that can be specific to a given case, in contrast to information taken from standard peptides, and because they allow for defining conformationally flexible geometrical parameters, in contrast to the rigid restraints that are currently in use.[77]

It is a particular advantage of the calculations that fragments of variable sizes can be compared in a consistent way, thus allowing the identification of those trends that emerge in growing chains and are specific for proteins. The helix compression found in crambin is an example. The divergence found between the dipeptide unit and crambin in the vicinity of residues 20−22 may point to similar effects characteristic of bends, and ongoing studies are aimed at further exploring this feature.

Acknowledgment. C.V.A. thanks the Fonds voor Wetenschappelijk Onderzoek (FWO)−Vlaanderen for appointment as Research Director (Onderzoeksdirecteur). J.M.L.M. acknowledges support as a Yigal Allon Fellow and as a Honorary Research Associate (Onderzoeksleider in eremandaat) of the Flemish FWO. The authors gratefully acknowledge partial support of this work by U.S. Department of Agriculture National Research Incentive Grant 97-35107-4362. This text presents results obtained under GOA-BOF-UA no. 23. Thanks are due to MSI for the use of the BIOSYM software. Thanks are due to Professor Peter Pulay (Arkansas), on whose suggestion the MIA method was developed.

Supporting Information Available: Table 2 of optimized Cartesian coordinates of crambin (11 pages). Ordering information is given on any current masthead page.

References and Notes

(1) Schäfer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys*. **1982**, *76*, 1439.
(2) Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schäfer, L.; Momany, F. A. *J. Am. Chem. Soc*. **1983,** *105*, 3438.
(3) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc*. **1979b**, *101*, 2550.
(4) Pulay, P. *Mol. Phys*. **1969**, *17*, 197.
(5) Pulay, P. *Theor. Chim. Acta* (*Berlin*) **1979a**, *50*, 299.
(6) Jiang, X.; Yu, C. H.; Cao, M.; Newton, S. Q.; Paulus, E. F.; Schäfer, L. *J. Mol. Struct*. **1997**, *403*, 83.
(7) Jiang, X.; Cao, M.; Newton, S. Q.; Schäfer, L.; Paulus, E. F. *Electron. J. Theor. Chem.* **1995b**, *1*, 11.
(8) Schäfer, L.; Cao, M.; Meadows, M. J. *Biopolymers* **1995a**, *35*, 603.
(9) Jiang, X.; Cao, M.; Teppen, B. J.; Newton, S. Q.; Schäfer, L. *J. Phys. Chem*. **1995a**, *99*, 10521.
(10) Karplus, P. A. *Protein Sci*. **1996**, *5*, 1406.
(11) Barone, V.; Fraternali, F.; Cristinziano, P. L. *Macromolecules* **1990**, *23*, 2038.
(12) Böhm, H. J.; Brode, S. *J. Am. Chem. Soc*. **1991**, *113*, 7129.
(13) Cheam, T. C. *J. Mol. Struct*. **1993**, *295*, 259.
(14) Cheam, T. C. *J. Mol. Struct*. **1992**, *274*, 289.
(15) Cheam, T. C.; Krimm, S. *J. Mol. Struct.* **1990**, *206*, 173.
(16) Cheam, T. C.; Krimm, S. *J. Mol. Struct*. **1989a**, *193*, 1.
(17) Cheam, T. C.; Krimm, S. *J. Mol. Struct*. **1989b**, *188*, 15.
(18) Cheam, T. C.; Krimm, S. *J. Mol. Struct*. **1986**, *146*, 175.
(19) Cheam, T. C.; Krimm, S. *J. Chem. Phys*. **1985**, *82*, 1631.
(20) Dive, G.; Dehareng, D.; Ghuysen, J. M. *J. Am. Chem. Soc*. **1994**, *116*, 2548.
(21) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc*. **1992**, *114*, 5369.
(22) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc*. **1994**, *116*, 9250.
(23) Gould, I. R.; Hillier, I. H. *J. Chem. Soc., Chem. Commun*. **1993**, 951.
(24) Gould, I. R.; Kollman, P. A. *J. Phys. Chem*. **1992**, *96*, 9255.
(25) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *J. Am. Chem. Soc*. **1991**, *113*, 5989.
(26) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *Int. J. Quantum Chem., Quantum Biol. Symp*. **1989**, *16*, 311.
(27) Jiao, D.; Barfield, M.; Hruby, V. J. *J. Am. Chem. Soc*. **1993**, *115*, 10883.
(28) Klimkowski, V. J.; Schäfer, L.; Momany, F. A.; Van Alsenoy, C. *J. Mol. Struct*. **1985**, *124*, 143.
(29) McAllister, M. A.; Perczel, A.; Csaszar, P.; Viviani, W.; Rivail, J. L.; Csizmadia, I. G. *J. Mol. Struct*. **1993b**, *288*, 161.
(30) Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. *J. Am. Chem. Soc*. **1991a**, *113*, 6256.
(31) Perczel, A.; Farkas, O.; Csizmadia, I. G. *J. Comput. Chem*. **1996**, *17*, 821.
(32) Perczel, A.; Kajtar, M.; Marcoccia, J. F.; Csizmadia, I. G. *J. Mol. Struct*. **1991b**, *232*, 291.
(33) Ramek, M.; Kelterer, A. M.; Teppen, B. J.; Schäfer, L. *J. Mol. Struct*. **1995**, *352/353*, 59.
(34) Sapse, A. M.; Daniels, S. B.; Erickson, B. W. *Tetrahedron* **1988**, *44*, 999.
(35) Sapse, A. M.; Fugler, L. M.; Cowburn, D. *Int. J. Quantum Chem*. **1986**, *29*, 1241.
(36) Sapse, A. M.; Mallah-Levy, L.; Daniels, S. B.; Erickson, B. W. *J. Am. Chem. Soc*. **1987**, *109*, 3526.
(37) Schäfer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; Van Alsenoy, C. *Biopolymers* **1984**, *23*, 2335.
(38) Siam, K.; Klimkowski, V. J.; Van Alsenoy, C.; Ewbank, J. D.; Schäfer, L. *J. Mol. Struct*. **1987**, *152*, 261.
(39) Siam, K.; Kulp, S. Q.; Ewbank, J. D.; Schäfer, L. *J. Mol. Struct*. **1989**, *184*, 143.
(40) Viviani, W.; Rivail, J. L.; Csizmadia, I. G. *Theor. Chim. Acta* **1993a**, *85*, 189.
(41) Viviani, W.; Rivail, J. L.; Perczel, A.; Csizmadia, I. G. *J. Am. Chem. Soc*. **1993b**, *115*, 8321.
(42) Weiner, S. J.; Singh, U. C.; O'Donnell, T. J.; Kollman, P. A. *J. Am. Chem. Soc*. **1984**, *106*, 6243.
(43) Böhm, H. J.; Brode, S. *J. Comput. Chem*. **1995**, *16*, 146.
(44) Böhm, H. J. *J. Am. Chem. Soc*. **1993**, *115*, 6152.
(45) McAllister, M. A.; Perczel, A.; Csaszar, P.; Csizmadia, I. G. *J. Mol. Struct*. **1993a**, *288*, 181.
(46) Perczel, A.; McAllister, M. A.; Csaszar, P.; Csizmadia, I. G. *Can. J. Chem*. **1994**, *72*, 2050.
(47) Perczel, A.; McAllister, M. A.; Csaszar, P.; Csizmadia, I. G. *J. Am. Chem. Soc*. **1993**, *115*, 4849.
(48) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc*. **1993**, *115*, 272.
(49) Van Alsenoy, C.; Cao, M.; Newton, S. Q.; Teppen, B.; Perczel, A.; Csizmadia, I. G.; Momany, F. A.; Schäfer, L. *J. Mol. Struct*. **1993**, *286*, 149.
(50) Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. *J. Am. Chem. Soc*. **1993**, *115*, 10812.
(51) Teeter, M. M.; Roe, S. M.; Heo, N. H. *J. Mol. Biol*. **1993**, *239*, 292.
(52) Lobb, L.; Stec, B.; Kantrowitz, E. K.; Yamano, A.; Stojanoff, V.; Markman, O.; Teeter, M. M. *Protein Eng*. **1996**, *9*, 1233.
(53) Yamano, A.; Heo, N. H.; Teeter, M. M. *J. Biol. Chem*. **1997**, *272*, 9597.
(54) Van Alsenoy, C. *J. Comput. Chem*. **1988**, *9*, 620.
(55) Van Alsenoy, C.; Peeters, A. *J. Mol. Struct*. **1993**, *286*, 19.
(56) Almlöf, J.; Faegri Jr., K.; Kosell, K. *J. Comput. Chem*. **1982**, *3*, 385.
(57) Van Alsenoy, C. In preparation.
(58) Challacombe, M.; Schwegler, E. *J. Chem. Phys*., in press.
(59) Schwegler, E.; Challacombe, M.; Head-Gordon, M. *J. Chem. Phys*., in press.
(60) White, C. A.; Head-Gordon, M. *J. Chem. Phys*. **1996**, *105*, 5061.
(61) Sellers, H. L.; Klimkowski, V. J.; Schäfer, L. *Chem. Phys. Lett*. **1978**, *58*, 541.
(62) Gwinn, W. D. *J. Chem. Phys*. **1971**, *55*, 477.
(63) Sellers, H. L.; Sims, L. B.; Schäfer, L.; Lewis, D. E. *J. Mol. Struct*. **1977**, *41*, 149.
(64) Pulay, P.; Fogarasi, G. *J. Chem. Phys*. **1992**, *96*, 2856.
(65) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc*. **1982**, *104*, 2797.
(66) Protein Data Bank; Chemistry Department, Building 555; Brookhaven National Laboratory, P.O. Box 5000; Upton, NY 11973-5000. E-mail: pdb@bn1.gov.
(67) *MSI, InsightII User Guide, Version 4.0.0*; Molecular Simulations, Inc.: San Diego, CA, 1996.
(68) de Smedt, J.; Vanhouteghem, F.; Van Alsenoy, C.; Geise, H. J.; Schäfer, L. *J. Mol. Struct.* **1992**, *259*, 289.
(69) Schäfer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Mol. Struct*. **1982,** *86*, 349.
(70) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1996**, *264*, 1180
(71) Boggs, J. E. *J. Mol. Struct.* **1983**, *97*, 1.

(72) Geise, H. J.; Pyckhout, W. In *Stereochemical Applications of Gas-Phase Electron Diffraction*; Hargittai, I., Hargittai, M., Eds.; VCH Publishers: New York, 1988; Vol. A, Chapter 10.
(73) Schäfer, L. *J. Mol. Struct.* **1983**, *100*, 51.
(74) Schäfer, L.; Siam, K.; Ewbank, J. D.; Caminati, W.; Fantoni, A. C. In *Modeling of Structures and Properties of Molecules*; Maksic, Z. B., Ed.; E. Horwood, Publ. Co.: Chichester, England, 1987; Chapter 4.
(75) Schäfer, L.; Ewbank, J. D.; Siam, K.; Chiu, N. S.; Sellers, H. L. In *Stereochemical Applications of Gas-Phase Electron Diffraction*; Hargittai, I., Hargittai, M., Eds.; VCH Publishers: New York, 1988; Vol. A, Chapter 9.
(76) Chiu, N. S.; Sellers, H. L.; Schäfer, L.; Kohata, K. *J. Am. Chem. Soc.* **1979**, *101*, 5883.
(77) Engh, R. A.; Huber, R. *Acta Crystallogr*. **1991**, *A47*, 392.