An Empirical Examination of Potential Energy Minimization Using the Well-Determined Structure of the Protein Crambin

Marc Whitlow and M. M. Teeter*

Contribution from the Department of Biochemistry, Boston University, Boston, Massachusetts 02118, and Department of Chemistry, Boston College, Chestnut Hill, Massachusetts 02167. *Received February 18, 1986*

Abstract: Empirical potential energy functions for proteins may be used to study protein stability and motion. However, it is difficult to evaluate these functions because most protein crystal structures are not accurate enough to act as test cases. We have empirically examined how well potential energy minimization can model the high resolution crystal structure (0.945 Å) of the hydrophobic protein Crambin (Hendrickson, W. A.; Teeter, M. M. *Nature (London)* **1981**, *290*, 107-113). Over 70 minimizations have been performed by using the program AMBER (Assisted Model Building with Energy Refinement) written by Kollman and Weiner (*J. Comp. Chem.* **1981**, *2*, 287-303) and the parameters from Weiner et al. (*J. Am. Chem. Soc.* **1984**, *106*, 765-784). We have examined the effects of (1) the form of the electrostatic potential energy, i.e., constant vs. distance-dependent, (2) the value of the dielectric constant, (3) the size of the "united atom" van der Waals radii, (4) the nonbonding cutoff, (5) the inclusion of the HC_α, and (6) the addition of crystal environment and of explicit water. Empirically, the best overall conditions for minimization are employing Jorgensen's van der Waals radii (*J. Am. Chem. Soc.* **1981**, *103*, 335-340), a distance-dependent dielectric constant of 4.0*r* and a 6.0-Å residue-based or 9.0-Å atom-based nonbonding cutoff.

Potential energy functions for proteins have proven useful in investigations of protein stability and motion. Minimization of the calculated potential energy has provided descriptions of protein-ligand interactions,[1] refinement of an initial structure after model building,[2,3] and constraints for crystallographic refinement of proteins.[4] The parameterization of potential energy functions has been based on small molecule X-ray crystallography, infrared and Raman spectroscopy, quantum mechanical calculations, and a large number of other types of information.[5-8]

One problem encountered in the use of potential energy functions with proteins is the lack of published data needed to empirically evaluate the agreement between these functions and observed structures of proteins. A comparison of different potential energy functions has been made for several small, crystalline cyclic peptides.[9] However, small peptides are not the best test cases for potential energy functions applied to proteins because of the following: (1) their crystalline conformations are often different from their solution conformations; (2) they are too small to have extensive secondary structure; and (3) their potential energy surface is less complicated than those of proteins. Nuclear magnetic resonance (NMR) studies of small peptides have shown that their solution conformations can be different from their conformations in the crystalline state.[10] Simulations of peptides in both their crystalline environment and in vacuo also may result in different conformations.[11,12] Crystal structures of small peptides have a large number of intermolecular interactions, which may play a major role in determining their crystalline conformations. The smaller proportion of these interactions in proteins suggests that they have a minor role in determining the conformation of a protein in its crystal structure. Arguments that support the idea that proteins have similar crystalline and solution conformations include the following: (1) the similarity of independently determined tertiary structures of the same or related proteins crystallized in different space groups and in the presence of different solvents and (2) the fact that a number of enzymes are catalytically active in the crystalline phase, albeit with reduced efficiency in some cases.[13]

We present here an empirical evaluation of one potential energy function for a well-determined protein structure. We have completed over 70 minimizations, changing various parameters (Table I). Our goals were (1) to describe the effects of changing these factors in the minimization on the resulting structure, (2) to quantify these effects and stress which are most important for obtaining reliable conformations, and (3) to empirically determine which set of potential parameters resulted in the best agreement with the crystal structure at a biologically relevant temperature (say 20 °C). The results will enable the users of potential energy functions to better understand the parameters which can be varied. They elucidate the current limitations of potential energy functions, which need to be addressed in the future.

AMBER (Assisted Model Building with Energy Refinement)[6,14,15] has been employed for potential energy minimization on the high resolution structure of the protein crambin. Crambin was chosen for this study for four reasons: (1) it is a small hydrophobic protein with 46 residues and 327 non-hydrogen atoms. The larger protein bovine panceatic typsin inhibitor (BPTI, 58 residues and 454 non-hydrogen atoms),[16-18] which has traditionally been used in molecular mechanics studies, would have taken more computational time for each calculation and is not as well-determined as crambin (R = 20.0% for BPTI[18] vs. R = 11.3% for crambin). (2) Crambin has a variety of secondary structure types, i.e., two α-helices, a small anti-parallel β-sheet, an extended chain region, and five turns (Figure 1). (3) It has a high resolution X-ray structure. Crystals of crambin scatter to interplanar d spacings of at least 0.88 Å.[19] The published structure[20] included diffraction data to 1.5 Å and was refined with Konnert and Hendrickson's restrained least squares. In this study, we have used the structure refined against 0.945 Å data (unpublished results of Hendrickson

(1) Blaney, J.; Weiner, P.; Dearing, A.; Kollman, P.; Jorgensen, E.; Oatley, S.; Burnridge, J.; Blake, C. *J. Am. Chem. Soc.* **1982**, *104*, 6424-6434.
(2) Whitlow, M.; Teeter, M. M. *J. Biomol. Struct. Dynam.* **1985**, *2*, 831-848.
(3) Levitt, M. *J. Mol. Biol.* **1983**, *170*, 723-764.
(4) Jack, A.; Levitt, M. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1978**, *A34*, 931-935.
(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. M. *J. Comp. Chem.* **1983**, *4*, 65-86.
(6) Weiner, S. P.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
(7) Hermans, J.; Ferro, D. *Biopolymers* **1971**, *10*, 1121-1138.
(8) Monany, F.; McGuire, R.; Burgess, A.; Scheraga, H. *J. Phys. Chem.* **1975**, *79*, 2361-2381.
(9) Hall, D.; Pavitt, N. *J. Comp. Chem.* **1984**, *5*, 441-450.
(10) Tonelli, A. E.; Brewster, A. I. *J. Am. Chem. Soc.* **1972**, *94*, 2851-2854.
(11) Go, N.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 525-535.
(12) Go, N.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 552-559.
(13) Matthews, B. W. *Ann. Rev. Phys. Chem.* **1976**, *27*, 493-523.
(14) Weiner, P.; Kollman, P. *J. Comp. Chem.* **1981**, *2*, 287-303.
(15) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129-145.
(16) Huber, R.; Kukla, D.; Ruhlmann, A.; Steigemann, W. *Cold Spring Harbor Symp. Quant. Biol.* **1971**, *36*, 141-150.
(17) Deisenhofer, J.; Steigemann, W. W. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1975**, *B31*, 238-250.
(18) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301-329.
(19) Teeter, M. M.; Hendrickson, W. A. *J. Mol. Biol.* **1979**, *127*, 219-223.
(20) Hendrickson, W. A.; Teeter, M. M. *Nature (London)* **1981**, *290*, 107-113.

0002-7863/86/1508-7163$01.50/0

Table I. Summary of Minimizations on Crambin[a]

ϵ	nonbonding	VDW	HC_α	PE	no. of cycles	FEG	rms, Å	% CRG^3	% CV	energy
1.0	99.0	A			3960	0.095	1.096	−13.9	−10.4	−1592.2
2.0	99.0	A			4735	0.099	0.986	−10.4	−6.8	−850.9
4.0	99.0	A			1186	0.100	0.500	−7.4	−4.3	−476.6
8.0	99.0	A			1522	0.077	0.504	−7.2	−4.4	−306.8
1.0*R*	99.0	A			1957	0.084	0.689	−11.3	−6.1	−674.0
2.0*R*	99.0	A			1308	0.091	0.523	−8.3	−4.5	−386.2
4.0*R*	99.0	A			2028	0.097	0.816	−7.4	−5.0	−265.9
8.0*R*	99.0	A			2668	0.092	0.886	−5.3	−3.5	−209.5
1.0	99.0	J			2181	0.084	0.686	−7.3	−1.8	−1366.3
2.0	99.0	J			1642	0.100	0.508	−3.2	−0.3	−616.0
4.0	99.0	J			2160	0.091	0.513	−2.1	−0.9	−290.1
8.0	99.0	J			1462	0.100	0.403	−0.9	−0.3	−118.4
1.0*R*	99.0	J			1919	0.096	0.599	−5.2	−0.4	−462.4
2.0*R*	99.0	J			1412	0.092	0.464	−2.9	0.2	−174.8
4.0*R*	99.0	J			887	0.088	0.349	−1.1	0.4	−44.6
8.0*R*	99.0	J			808	0.096	0.331	−0.4	0.2	11.0
1.0	10.0 R	J			2860	0.097	0.420	−3.8	−0.3	−1345.7
2.0	10.0 R	J			2386	0.100	0.352	−2.1	0.2	−621.9
4.0	10.0 R	J			1785	0.096	0.269	−0.5	0.5	−275.1
8.0	10.0 R	J			1848	0.098	0.301	−0.2	0.9	−113.7
1.0*R*	10.0 R	J			2932	0.098	0.431	−3.7	0.2	−465.9
2.0*R*	10.0 R	J			1832	0.099	0.261	−1.0	0.1	−176.5
4.0*R*	10.0 R	J			1582	0.098	0.261	−0.3	0.7	−59.0
8.0*R*	10.0 R	J			2134	0.098	0.285	0.0	0.8	−5.2
8.0	13.0 A	A			3473	0.098	0.652	−7.9	−5.6	−316.7
8.0	11.0 A	A			1411	0.096	0.562	−4.5	−3.6	−300.0
8.0	9.0 A	A			1776	0.099	0.482	−6.2	−3.5	−299.0
8.0	7.0 A	A			1363	0.094	0.431	−3.7	−3.5	−295.6
1.0*R*	13.0 A	A			1665	0.097	0.653	−10.8	−5.2	−693.6
1.0*R*	11.0 A	A			1852	0.089	0.642	−9.4	−4.9	−689.5
1.0*R*	9.0 A	A			1091	0.110	0.559	−9.3	−4.8	−674.0
1.0*R*	7.0 A	A			1778	0.084	0.563	−7.1	−4.5	−635.2
8.0*R*	13.0 A	A			916	0.094	0.383	−5.1	−3.6	−202.6
8.0*R*	11.0 A	A			863	0.099	0.413	−5.6	−4.0	−204.1
8.0*R*	9.0 A	A			962	0.090	0.447	−6.3	−4.1	−194.7
8.0*R*	7.0 A	A			934	−0.087	0.405	−5.1	−4.0	−203.5
4.0*R*	13.0 A	J			1642	0.090	0.276	−0.2	0.5	−59.6
4.0*R*	11.0 A	J			1934	0.096	0.260	−0.2	0.7	−58.1
4.0*R*	9.0 A	J			1881	0.096	0.252	0.0	0.6	−56.6
4.0*R*	7.0 A	J			1409	0.097	0.255	1.1	1.1	−55.0
4.0*R*	14.0 R	J			1684	0.099	0.265	−0.2	0.6	−59.0
4.0*R*	12.0 R	J			1881	0.090	0.263	−0.3	0.7	−58.8
4.0*R*	10.0 R	J			1582	0.098	0.261	−0.3	0.7	−59.0
4.0*R*	8.0 R	J			1336	0.093	0.264	−0.3	0.7	−59.1
4.0*R*	6.0 R	J			1934	0.100	0.245	0.1	0.8	−57.2
1.0	99.0	A	A		2497	0.084	0.814	−10.2	−5.6	−1722.3
1.0	99.0	A	E		3597	0.095	1.177	−12.0	−5.8	−1746.2
8.0	99.0	A	A		1918+	0.099	0.885	−5.2	−3.8	−336.1
8.0	99.0	A	E		1493	0.084	0.495	−4.6	−3.5	−321.6
1.0*R*	99.0	A	A		3388	0.097	1.147	−11.3	−5.8	−742.7
1.0*R*	99.0	A	E		3576	0.093	0.848	−10.5	−5.7	−760.4
8.0*R*	99.0	A	A		1366	0.093	0.469	−4.5	−3.2	−197.6
8.0*R*	99.0	A	E		1548	0.087	0.498	−4.7	−3.7	−202.7
1.0	99.0	A		H	200	0.091	0.112	−1.2	−0.5	−1600.3
1.0	99.0	A		H	511	0.030	0.141	−1.8	−0.6	−1607.2
1.0	99.0	A		C	771	0.426	0.433	−5.3		−3098.0
1.0	99.0	A	E	C	1382	0.273	0.438	−3.7		−3647.0
1.0	99.0	A		W	1751	0.469	0.552	−9.3	−4.1	−3494.0
1.0	10.0 R	J		W	7500	0.337	0.298	−2.3	−0.5	−3740.5
1.0	10.0 R	J		W	7500	0.272	0.297	−2.1	−0.7	−3768.8

[a] Abbreviations: ϵ is the dielectric constant used. Nonbonding is the nonbonding cutoff used, with R for residue-based and A for atom-based cutoffs. VDW is the "united atom" van der Waals radii used, where A is the AMBER's and J is for Jorgensen's. HC_α is A if the alanyl dipeptide charge were used and E if the extended atom charges were used. PE is the protein environment, where H is for intermolecular hydrogen bonds, C is for the 4.0-Å crystal shell, and W is for the 5.0-Å water shell. FEG is the final energy gradient in kcal/(mol·Å). Rms is the rms coordinate shifts between the minimized structure and crambin's crystal structure for all atoms. % CGR^3 is the percent change in the cube of the radius of gyration. % CV is the percent change in the volume. The energy is the total energy after minimization in kcal/mol.

and Teeter). (4) Crambin is electrically neutral. It has only four charged groups of which two are involved in an intramolecular salt bridge. Thus counterions can be omitted from these calculations.

Methods

Sequence Chosen. There are two sites of microheterogeneties found in crambin. Residue 22 can be Pro or Ser and residue 25 can be Leu or Ile. All four forms exist in the seed of *Crambe abyssinica* from which crambin is isolated.[21] The Pro_{22}–Ile_{25} is the most abundant form found in the crystal structure of crambin,[20] and it was used in these calculations.

N-Terminal Threonine. The AMBER version we used did not include charged N-terminus residues. Thus the N-terminal Thr for crambin was

(21) Teeter, M. M.; Mazer, J. A.; L'Italien, J. J. *Biochemistry* **1981**, *20*, 5437–5443.

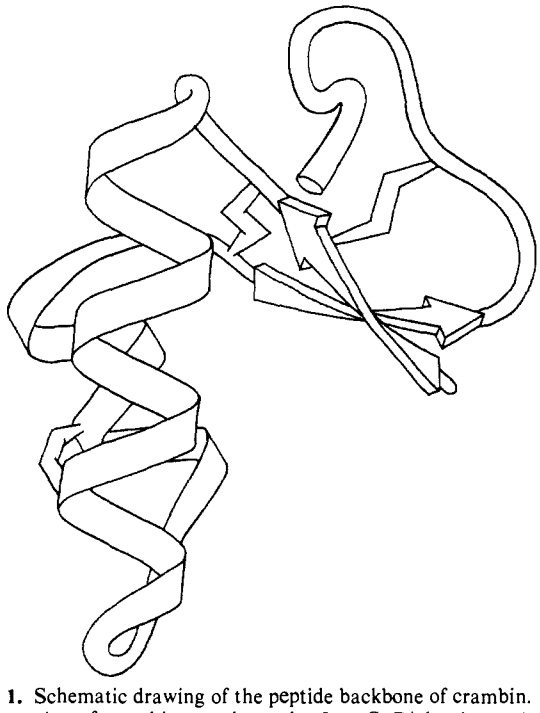


Figure 1. Schematic drawing of the peptide backbone of crambin. This representation of crambin was drawn by Jane S. Richardson. Arrows depict strands. The disulfides are drawn as "lightning flashes".

built from the N_z of lysine and a Thr residue. The charge scheme used for the N-terminal NH_3^+ was from lysine N_z and from threonine for the rest of the residue. The C_α charge was adjusted such that the N-terminal had a formal net charge of +1.00 e (0.263 e for the N-terminal Thr vs. 0.268 e for an internal Thr).

Potential Energy Minimizations. All the potential energy minimizations were run on either a VAX 11/750 or VAX 11/780. The potential energy force field (eq 1) in AMBER has a number of valuable features. It uses harmonic bond and angle stretching. More than one set of dihedral parameters can be used for any given proper or improper torsion angles, giving the potential the ability to describe very complicated torsion barriers. Atom-centered monopoles are used in the electrostatic interactions, except for sulfur where lone pairs are also included. Two forms of the electrostatic potential (the distance-dependent dielectric and the constant dielectric) can be employed. A Lennard-Jones 6-12 function is used for the nonbonding van der Waals potential. A hydrogen bond 10-12 potential is used to fine tune the hydrogen bond lengths. The hydrogen bond energy is dominated by its electrostatic interactions.

A "united atom" force field was used, which includes all atoms explicitly with the exception of hydrogens bonded to carbon. Counterions were omitted from all minimizations. The parameters used were from Weiner et al.[6] A minimization was judged to have converged when the final root-mean-square (rms) energy gradient was less the 0.1 kcal/(mol·Å). Each minimization was unconstrained unless otherwise specified.

$$U = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 +$$

$$\sum_{\text{dihedrals}} \frac{K}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r^6_{ij}} + \frac{q_iq_j}{\epsilon r_{ij}} + \sum_{\text{H bonds}} \frac{B'_{ij}}{r_{ij}^{12}} - \frac{A'_{ij}}{r_{ij}^{10}}$$

The minimizations performed varied different aspects of the potential. In particular, the following parameters and environments were changed systematically.

(1) Form of the electrostatic potential energy: AMBER can employ either a distance-dependent dielectric potential or a constant dielectric electrostatic potential energy

$$\sum_{i<j} \frac{q_iq_j}{\epsilon(r_{ij})r_{ij}} \qquad \sum_{i<j} \frac{q_iq_j}{\epsilon r_{ij}}$$

The constant dielectric electrostatic potential is the classical coulomb potential. The distance-dependent dielectric electrostatic potential mimics the polarization effect in attractive interactions, with closer interactions weighted more heavily. It also helps compensate for the lack of explicit solvation in vacuum calculations by implicitly damping the longer range charge interactions.[1] Warshel has shown that the effective dielectric constant for short-range ionic interactions in water is smaller than the bulk dielectric and increases (roughly linearly) with charge separation.[22]

(2) Dielectric constant: different potentials had dielectric constants of 1.0, 2.0, 4.0, and 8.0 for constant dielectric and 1.0r, 2.0r, 4.0r, and 8.0r for the distance-dependent dielectric electrostatic potential.

(3) Nonbonding cutoffs: the nonbonding cutoff in AMBER applies to both the electrostatic and the van der Waals interactions. Two types of nonbonding cutoffs were tried. Atom-based cutoffs were used, in which any atom within the cutoff would be included in the nonbonding interactions. A third-order polynomial smoothed the electrostatic potential between the nonbonding cutoff and 2.0 Å before the cutoff. The atom-based cutoff used a neutral sphere correction, in which the nonbonding pairs list is created such that an atom is in a net neutral sphere. This avoids the splitting of dipoles. The atom pairs list was calculated at distances 1.0 Å longer than the nonbonding cutoff and was updated every 25th cycle. Nonbonding cutoffs of 13.0, 11.0, 9.0, and 7.0 Å were used for the atom-based cutoff.

A residue-based nonbonding cutoff was also used. If any part of a residue was within the nonbonding cutoff distance then that complete residue would be included in the nonbonding interactions. The atom pairs list for the residue-based cutoff was also updated every 25th cycle. Cutoffs of 14.0, 12.0, 10.0, 8.0, and 6.0 Å were used for the residue-based cutoff.

A number of minimizations were run without a nonbonding cutoff (nonbonding cutoff = 99.0 Å). The nonbonding pairs list was generated once at the beginning of the minimization for those.

(4) Addition of C_α hydrogens: the C_α hydrogens were added to the "united atom" topology files (for the program PREP). Two different charge schemes were used with the C_α hydrogens. One charge scheme was derived from the "united atom" charge scheme in AMBER and the other from the all-atom charge scheme[23] (Table II). The van der Waals radius for the C_α was changed from the "united atom" radius to that of a carbon atom with no hydrogens.

(5) "United atom" van der Waals radii: In addition to the original "united atom" van der Waals radii for carbon, those of Jorgensen[24] were also used (Table III). All of the original "united atom" radii are smaller than those of Jorgensen's. The latter "united atom" radii decrease as the number of hydrogens bound to a carbon increases, whereas the original radii increase.

(6) Protein environment: four different environments were employed around crambin. TIP3P[25] water molecules were used in all calculations involving water. (A) In vacuo minimizations had no intermolecular interactions. (B) The four intermolecular protein-hydrogen bonds in the crystal structure of crambin were the only intermolecular interactions

Table II. Charges for the Backbone Atoms

atom	extended atom	alanyl dipeptide	AMBER + HC_α
N	-0.520	-0.463	-0.520
H	0.248	0.252	0.248
C_α	0.215	0.036	0.075
HC_α		0.048	0.140
C	0.526	0.616	0.526
O	-0.500	-0.504	-0.500

Table III. van der Waals Radii (Å)

atom	original[14]	Jorgensen[24]
C	1.850	1.850
CH	1.850	2.385
CH2	1.925	2.235
CH3	2.000	2.165
O	1.650	1.650
N	1.750	1.750

Table IV. Intermolecular Hydrogen Bonds

hydroxyl of Ser_6	carbonyl of Leu_{18}, translation along the c axis
guanidiniuim of Arg_{17}	carboxyl of Asp_{43}, translation along the c axis
amide of Ala_{38}	carbonyl of Ala_{45}, on a screw axis-related molecule translated along the a and c axes
	hydroxyl of Thr_{39}, on a screw axis-related molecule translated along the a and c axes

(22) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640-1652.
(23) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, in press.
(24) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335-340.

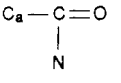
Table V. Fragments of Symmetry-Related Crambin Peptides Used in 5.0-Å Crystal Environment Shell

Pro_5–Ser_6	Ser_{11}–Asn_{12}
Asn_{14}–Val_{15}	Val_{15}–Cys_{16}
Arg_{17}–Gly_{20}	Leu_{18}–Pro_{19}
Thr_{21}–Ser_{22}	Ala_{24}–Ile_{25}
Thr_{28}–Tyr_{29}	Pro_{36}–Thr_{39}
Ala_{45}–Asn_{46}	

Table VI. The Crystal Structure Geometry of Crambin

mean distance to the center of mass	9.140 Å
radius of gyration	9.569 Å
volume	5262 Å^3
accessible surface area	3022 Å^2
hydrophobic accessible surface area	2063 Å^2
hydrophilic accessible surface area	959 Å^2
mean torsion angles	
top α-helix	
ϕ	−62.5°
ψ	−39.1°
bottom α-helix	
ϕ	−66.2°
ψ	−37.4°
β-sheet strand 1	
ϕ	−120.2°
ψ	141.1°
β-sheet strand 2	
ϕ	−122.1°
ψ	150.7°
mean deviation from planarity for ω	4.34°
average hydrogen bond geometry	
α-helices	
N···O	2.955 Å
O···H—N	152.89°
C=O···H	149.42°
β-sheet	
N···O	2.909 Å
O···H—N	155.08°
C=O···H	151.87°

included in these minimizations (Table IV). The symmetry-related segments involved in the intermolecular hydrogen bonds were represented as small fragments of the molecule. For example, a backbone carbonyl acceptor was represented as



Because of the small size of the symmetry-related fragments, they were harmonically constrained to their crystal positions, by using 100.0 and 10.0 kcal/(mol·Å^2) constraints on each atom. (C) A 4.0-Å shell of the crystal environment of crambin was used. This calculation contained 11 peptides (Table V) and 131 water molecules, and it was run in two steps. In the first step, all the non-hydrogen atoms were harmonically constrained to their initial position at 100.0 kcal/(mol·Å^2). In the final step, only the ends of the peptides were harmonically constrained at 10.0 kcal/(mol·Å^2). (D) Crambin was placed in a water box from Jorgensen's Monti Carlo simulation.[25] Waters closer than 2.8 Å and further than 5.0 Å were removed. The dipoles of the water molecules were in one case aligned with the electric field vectors of crambin, and in another case they were not aligned. Each simulation was run in two steps. In the first step, all the non-hydrogen atoms in crambin were harmonically constrained to their crystal positions at 100.0 kcal/(mol·Å^2). In the final step, no constraints were applied.

Structural Analysis of Minimized Models. After each calculation, the minimized structure of crambin was examined for the following structural information and compared with the crystal structure (Table VI). (1) rms coordinate shift: The rms coordinate shifts between the crystal structure and the minimized model were calculated for all non-hydrogen atoms, for just backbone atoms (N, C_α, C, and O), and for local secondary structure elements such as the top α-helix. One set of coordinates is rotated and translated with respect to the other set of coordinates, such that a minimum in the rms coordinate shifts is obtained for the atoms being compared. (2) Mean distance to the center of mass: the mean distance to the center of mass was calculated for the backbone atoms N, C_α, C, and O only. (3) Volume: the volume was calculated, by using the standard Voronoi polyhedral procedure to determine the face of a polyhedron surrounding each atom, with Lee and Richards' program VOLUME.[26] The percent change in volume from the crystal structure of crambin was also calculated. (4) Radius of gyration: the radius of gyration was calculated for the backbone atoms N, C_α, C, and O only. The percent change in the cube of the radius of gyration between the minimized model and the crystal structure of crambin was also calculated. A comparison of the percent volume change and the percent change in the cube of the radius of gyration enables one to separate volume compaction from changes in the overall shape of the molecule during minimization. (5) Accessible surface area: the accessible surface areas for the hydrophilic atoms, hydrophobic atoms, and the total accessible surface area were calculated by using Lee and Richards' programs ACCESS and ACCSUMS.[26] A water molecule, with a 1.4-Å probe radius, was used as a probe. The percent changes in the accessible surface areas from the crystal structure of crambin were then determined. Chothia showed that the accessible surface area has a strong correlation with the entropy of folding.[27] Novotny et al. showed that accessible surface area is an important indicator of how well a predicted structure has been folded.[28]

(6) Torsion angles: the mean deviation in torsion angles between the minimized model and the crystal structure of crambin were calculated. Mean deviations in ϕ and ψ from the crystal structure of crambin were calculated for the top helix (residues 8–18 for ϕ and 7–17 ψ), bottom helix (residues 24–29 for ϕ and 23–28 for ψ), first strand of the β-sheet (2–4 for ϕ and 1–3 for ψ), and second strand of the β-sheet (33–35 for ϕ and 32–34 for ψ). (7) Rms deviation from planarity of the peptide bond: the rms deviation from 180° for the peptide bond (ω) was calculated. (8) Mean hydrogen bond geometry: the mean N···O distance, O···H–N and C=O···H angles, were calculated for the helices (residues 7–17 in the top helix and 23–31 in the bottom helix) and the β-sheet (residues 1–3, 33–35, and 46). (9) Number of hydrogen bonds: a hydrogen bond was defined as having a donor-to-acceptor distance of 2.5–3.4 Å, a hydrogen-to-acceptor distance of less than 2.5 Å, and a donor–hydrogen–acceptor angle of greater than 120.0°.

Table VII. Magnitude of the Potential Energy Terms (kcal/mol) for Minimized Crambin at Dielectrics of 1.0 and 8.0 with No Nonbonding Cutoff and with Jorgensen's van der Waals Radii

	ϵ = 1.0	ϵ = 8.0
electrostatic	−3091.9	−359.6
1,4 electrostatic	1612.9	198.1
van der Waals	−310.6	−334.0
1,4 van der Waals	252.2	231.9
hydrogen bond 10–12	−24.3	−15.7
torsion angles	95.4	72.4
angles	80.2	68.7
bonds	19.5	19.8
total	−1366.3	−118.4

Results

Value of the Dielectric Constant. The electrostatic energy contributes the most to the overall potential energy of a system. This can be easily seen by examining the potential energy terms for crambin (Table VII). At low dielectric (ϵ = 1.0), the electrostatic terms are an order of magnitude larger than the next largest term (near neighbor van der Waals interactions). Whereas, at high dielectric (ϵ = 8.0), the electrostatic terms are of equal magnitude to the others. This large effect of the dielectric constant on the potential energy of a system makes its selection extremely important. Crambin is a hydrophobic, water-insoluble protein, and the electrostatic effects for a hydrophilic protein would undoubtedly be larger.

Decreasing the dielectric constant causes larger rms coordinate shifts between the crystal structure of crambin and the minimized structure for minimizations with no nonbonding cutoff (Figure 2). A large proportion of the rms coordinate shift appears to be due to a change in shape during minimization. Comparing the percent change in the cube of the radius of gyration with the percent change in the volume (Figure 3), one sees that, at the

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(26) Richards, F. M. *Methods in Enzymology*; Wyckoff, H. W., Timasheff, S., Eds.; New York, 1985; Vol. 115, pp 440–464.

(27) Chothia, C. *Nature (London)* **1975**, *254*, 304–308.

(28) Novotny, J.; Bruccoleri, R.; Karplus, M. *J. Mol. Biol.* **1984**, *177*, 787–818.

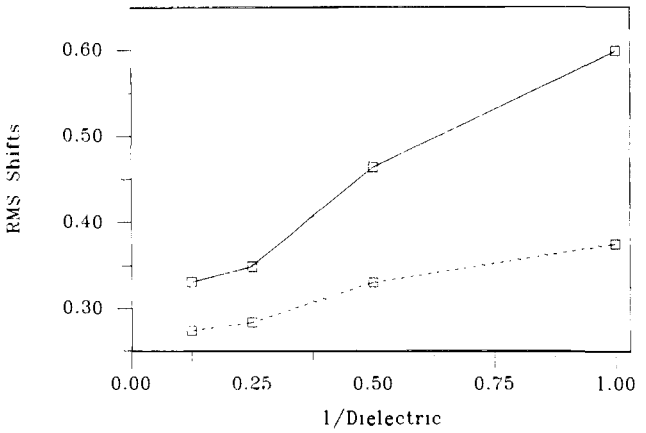


Figure 2. Effects of different dielectric constants on rms coordinates shifts between the minimized structure and the crystal structure of crambin. These shifts are for minimizations run with distance-dependent dielectrics of 1.0r, 2.0r, 4.0r, and 8.0r, by using Jorgensen's "united atom" van der Waals radii. Overall rms coordinate shifts are the solid line, and the rms coordinate shift for the top helix is the dashed line.

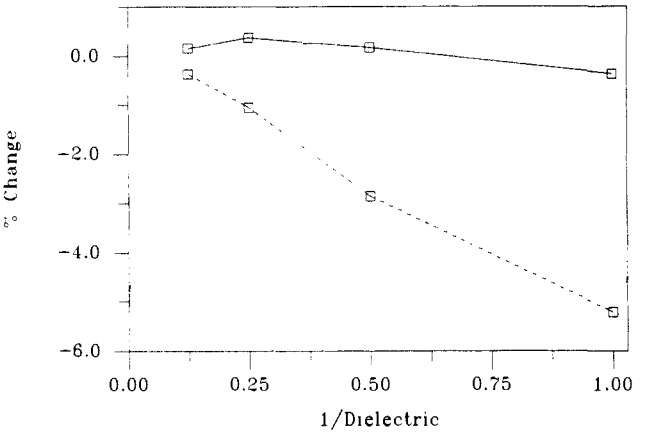


Figure 3. Effects of different dielectric constants on the percent volume change (solid line) and the percent change in the cube of radius of gyration (dashed line). These calculations are from minimizations run with distance-dependent dielectrics of 1.0r, 2.0r, 4.0r, and 8.0r.

lowest dielectric constants, there is as much as a 5% decrease in the cube of the radius of gyration, whereas there is little change in the volume. Crambin thus becomes more spherical in minimizations with lower dielectric constant. The rms coordinate shifts from the crystal structure for a local structural element (the top helix) are smaller but show the same trend as the overall rms coordinate shifts (Figure 2). Thus, decreasing the dielectric constant results in both a proportional change in the shape of crambin and in larger rms deviations from the local crystal structure.

Changes in the composition of the solvent accessible surface area of crambin upon minimization suggest one possible reason for the local deviations from the crystal structure. The accessible surface area of the hydrophilic atoms decreases while the accessible surface area of the hydrophobic atoms increases with decreasing dielectric constant (Figure 4). The overall accessible surface area decreases when compared to the crystal structure as the dielectric constant is decreased. In Figure 5, the decrease in the accessible surface area of the hydrophilic atoms from the crystal structure can be clearly seen. In crambin, the accessible surface area of the hydrophilic atoms is 32% of the total accessible surface area. For strong electrostatic interactions (ϵ = 1.0), the accessible surface area of the hydrophilic atoms can be reduced by 21%. This loss in the hydrophilic atoms' accessible surface area may be due to the lack of solvent interactions for the hydrophilic atoms.

The local deviations from the crystal structure directly affect the backbone torsion angles. The distribution of ϕ and ψ deviations from the crystal structure are disturbingly systematic, as seen in Figure 6. ϕ shifts to more negative values and ψ shifts to more

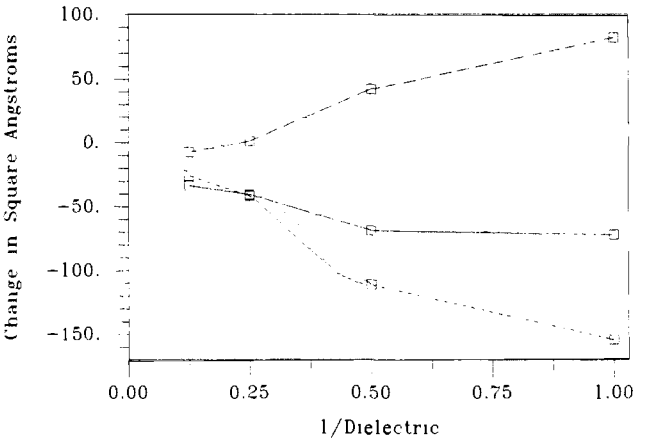


Figure 4. Changes in the overall accessible surface area (solid line), the accessible surface area of the hydrophilic atoms (small dashed line), and the accessible surface area of the hydrophobic atom (large dashed line) of crambin. These calculation are for minimizations run with distance-dependent dielectric constants of 1.0r, 2.0r, 4.0r, and 8.0r, with Jorgensen's "united atom" van der Waals radii.

positive values. There is a strong negative correlation between the shifts in ϕ_n and ψ_{n-1}, which has also been seen in the dynamics of BPTI.[29] Due to this strong negative correlation, the overall change in the conformation of the protein appears to be small. The largest deviations in ϕ and ψ are found in turn regions (1) at Pro_5 and Ser_6, a turn between the first strand of the β-sheet and the top helix; (2) at Pro_{19} and Gly_{20}, a turn between the two α-helices at the bottom of the Γ[20] in Figure 1; (3) at Gly_{31}, a turn between the bottom helix and the second strand of the β-sheet; (4) at Ala_{38}, a large turn just after the second β-strand on the right hand side of Γ in Figure 1.

The helices show increasing deviations in ϕ and ψ from the crystal structure as the dielectric constant is lowered, where the β-sheet shows complex changes with dielectric constant. The β-sheet is located between the two helices and the extended arm. During minimization with a strong electrostatic potential, the helices and the extended arm move toward each other (Figure 5), causing distortions in the β-sheet. If the change in molecular shape was removed from these calculation, we suspect that the strongest electrostatic potential would cause largest deviations in the ϕ and ψ angles of the β-sheet as well.

The rms deviation from planarity (180°) for the peptide torsion angle ω increases from 4.5 to 7.4° in going from a constant dielectric of 8.0–1.0. The planarity of ω is maintained by its torsion barrier and distorted by other interactions. Thus, as we increase the electrostatic forces, we see an increase in the deviation from planarity for ω.

In AMBER the hydrogen bond potential is dominated by the electrostatic term. The hydrogen bond 10–12 term is only used to fine tune the potential.[1] Thus, as the electrostatic potential gets stronger by decreasing the dielectric constant, the hydrogen bond energy increases and the hydrogen bond distances become smaller (Figure 7). The average hydrogen bond lengths for the α-helices and the β-sheet in the crystal structure of crambin are 2.96 and 2.91 Å, respectively. For a dielectric constant of 1.0r, these averages shortened to 2.83 and 2.86 Å.

The stronger electrostatic potential also results in formation of new hydrogen bonds. For a dielectric constant of 1.0, 16 new hydrogen bonds are formed for a total of 44. Most of the new hydrogen bonds are involved in side chain-to-side chain or side chain-to-backbone interactions. One example of this type is shown in Figure 8. In the crystal structure, the guanidinium of Arg_{17} is hydrogen bonded through a water molecule to Glu_{23} and directly to the carboxyl group of Asp_{43} in a neighboring molecule (Table IV). Upon minimization, the guanidinium of Arg_{17} forms a direct hydrogen bond to the carboxyl of Glu_{23}. Attempts have been made

(29) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585–590.

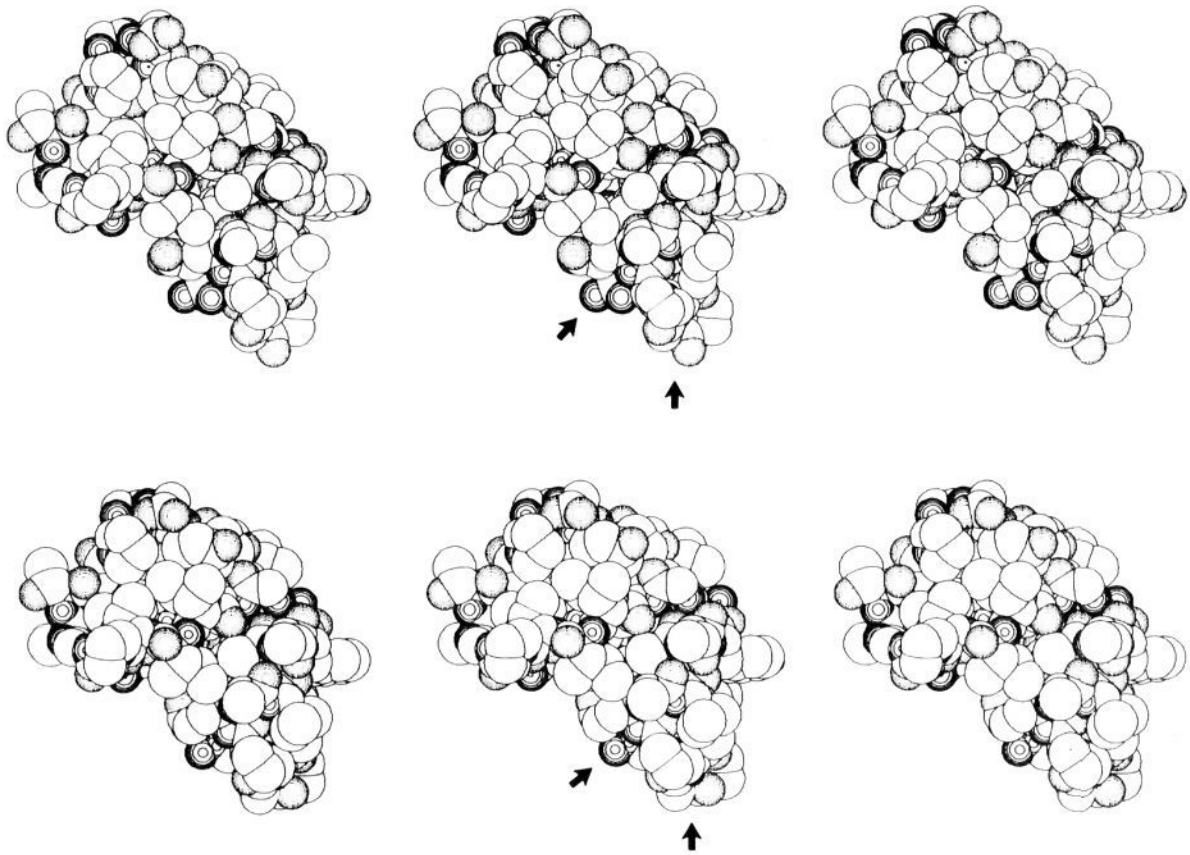


Figure 5. Stereo van der Waals surface diagrams for the crystal structure of crambin (top) and crambin minimized by using dielectric constant of 1.0 (bottom). Hydrophilic atoms are indicated by solid circles (nitrogen) or dotted circles (oxygen). Compare changes in regions near the arrows. The left pair are for divergent stereo viewing and the right pair are for convergent viewing.

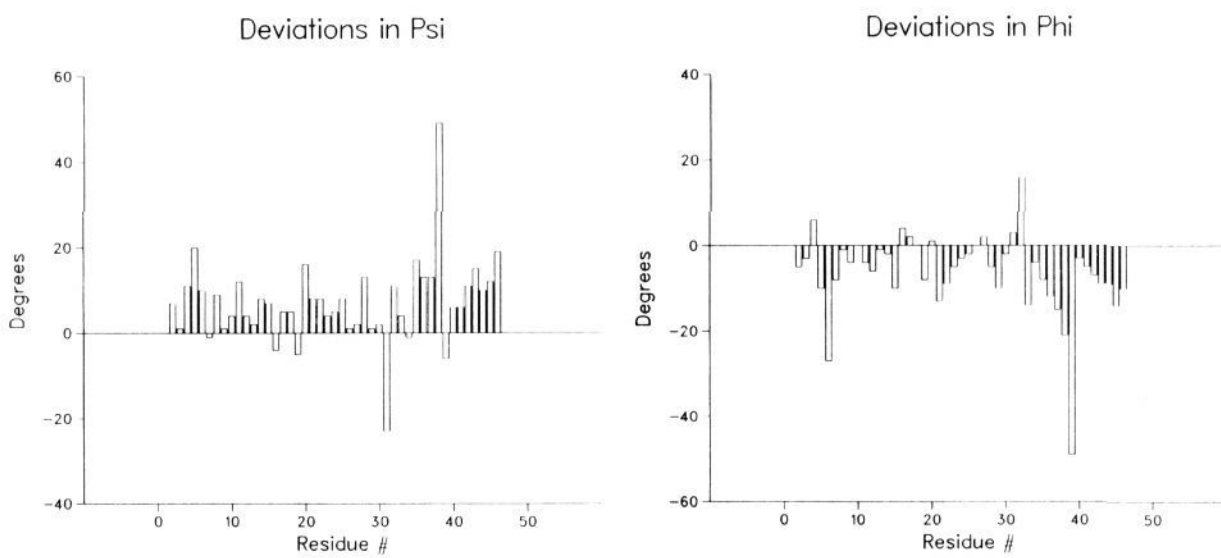


Figure 6. Deviation in ϕ and ψ between the crystal structure and the minimized structure for a constant dielectric of 2.0 by using Jorgensen's van der Waals "united atom" radii: left—deviations in ψ, and right—deviations in ϕ.

to place a water molecule found in the crystal structure between the Arg_{17} and Glu_{23}, with no success. Arg_{17} also forms a new hydrogen bond between its $N_{\eta 2}$ and its carbonyl.

Another example of a typical hydrogen bond change is the 11–15 hydrogen bond of the top helix. The top helix of crambin is slightly kinked, and the hydrogen bond distance between residues 11 and 15 of the top helix in the crystal structure is 3.16 Å. As the dielectric constant is lowered in the minimizations, the 11–15 hydrogen bond approaches a hydrogen bond length of 2.9 Å. This results in a straightening of the top helix. A kinked or bent helix typically may have a water molecule hydrogen bonded to the carboxyl oxygen and a long helix hydrogen bond in the crystal.[30] We have run calculations with the crystal environment discussed below, and they did not stop the minimization from removing the kink in the top helix. An electrostatic potential that decreases as a function of the distance to the solvent may better model the top helix of crambin, for the 11–15 hydrogen bond is on the surface

(30) Blundell, T.; Barlow, D.; Borkakoti, N.; Thornton, J. *Nature (London)* **1983**, *306*, 281–283.

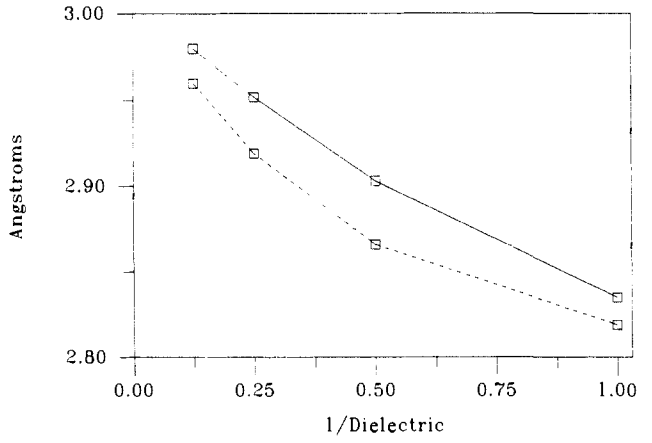


Figure 7. The mean hydrogen bond distances for α-helices (solid line) and the β-sheet (dashed line) as a function of 1/dielectric constant. These calculations are from minimizations run with distance-dependent dielectrics of 1.0r, 2.0r, 4.0r, and 8.0r. In crambin crystal structure the mean hydrogen bond distance for α-helices and β-sheet are 2.96 and 2.91 Å, respectively.

of crambin and its electrostatic forces would be greatly reduced.

Form of the Electrostatic Potential. Changing from a constant dielectric to a distance-dependent dielectric electrostatic potential is in many ways similar to raising the dielectric constant. Some of the basic effects can be understood as decreasing the strength of the electrostatic forces. The rms coordinate shifts between the crystal structure and the minimized structure are decreased for the distance-dependent dielectric. Rms coordinate shifts for the top helix are also lower with the distance-dependent dielectric. The radius of gyration for the distance-dependent dielectric minimization increases relative to that of the constant dielectric but still decreases from that of the crystal structure. Thus the overall molecular shape is closer to that of the crystal structure for the distance-dependent dielectric. The accessible surface area shows less deviation from the crystal structure with the distance-dependent dielectric. The rms deviation from planarity for the peptide bond ω decreases for each dielectric constant by up to 1.1° in changing from a constant to a distance-dependent dielectric electrostatic potential.

There are significant differences between the constant and distance-dependent dielectric electrostatic potentials that cannot be understood as a decrease in the strength of the electrostatic forces but as a decrease in the strength of the electrostatic forces at longer distances. The mean N···O distance in the backbone hydrogen bonds is shorter by 0.01 Å (ϵ = 8.0 and 8.0r) to 0.08 Å (ϵ = 1.0 and 1.0r) for the distance-dependent dielectric than for the constant dielectric electrostatic potential. A decrease in the N···O electrostatic repulsion relative to the H···O attraction for the distance-dependent dielectric accounts for these shorter N···O distances. The best agreement in these distances with the crystal structure of crambin are for ϵ = 2.0 (constant) and for ϵ = 4.0r (distance-dependent) dielectric constants.

For both electrostatic potentials, the helices show increasing deviations in ϕ and ψ from the crystal structure as the dielectric constant is lowered, whereas the β-sheet shows a complex change as a function of the dielectric constant. The constant dielectric changes more in ϕ and ψ as a function of dielectric constant for the β-sheet than does the distance-dependent ϵ.

Nonbonding Cutoffs. A nonbonding cutoff reduces and removes the long distances electrostatic interactions. This results in less distortion of the overall shape of the molecule and in smaller rms coordinate shifts from the crystal structure (Figure 9). The effects of not using a nonbonding cutoff are as strong as the effects of the choice of dielectric constant and stronger than the choice of the form of electrostatic potential (see below).

1. Atom-Based Nonbonding Cutoff. The atom-based nonbonding cutoff shows smaller backbone rms coordinate shifts from

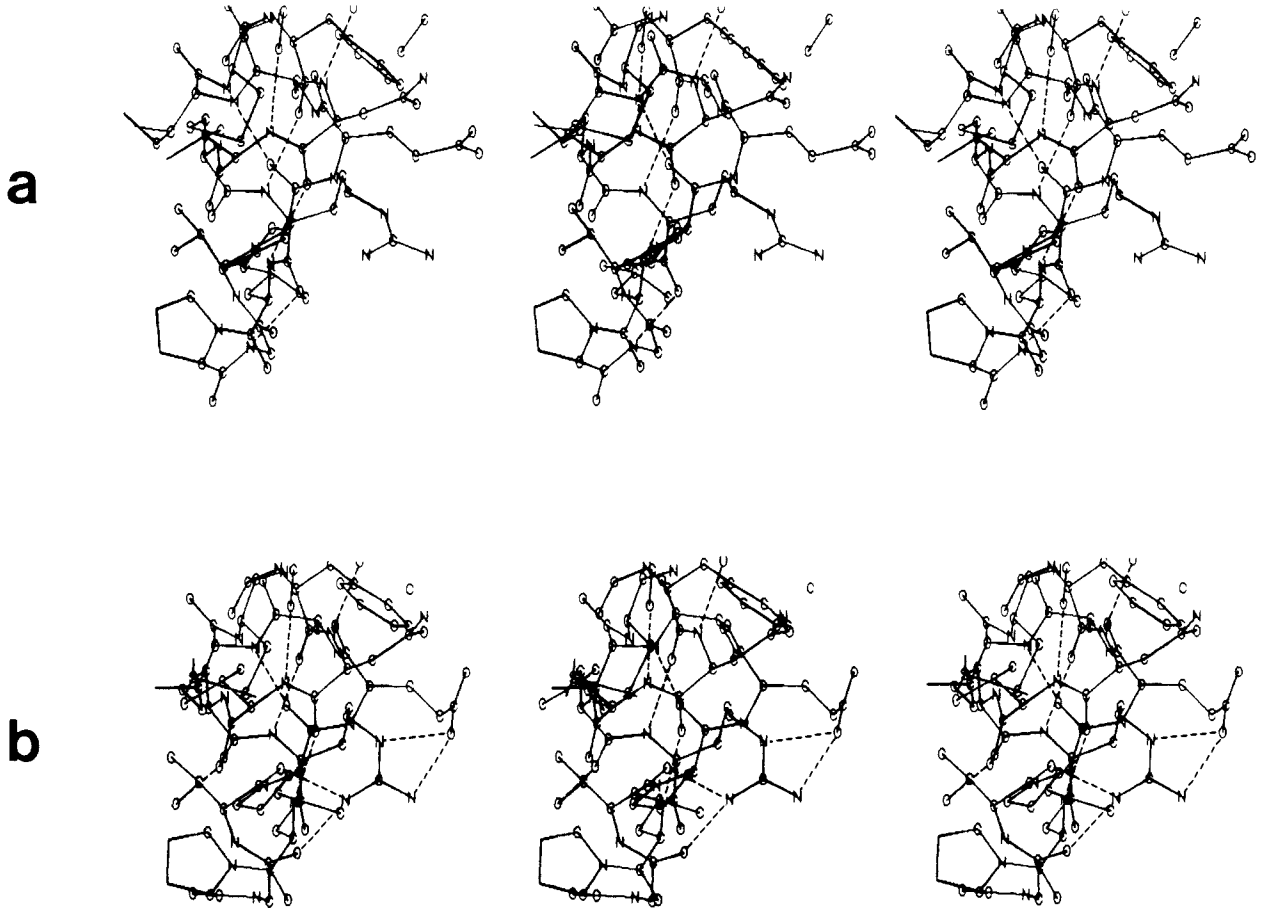


Figure 8. Stereo diagram showing the formation of a new salt bridges between Arg_{17} and Glu_{23}: (a) the crystal structure and (b) the model minimized with a distance-dependent dielectric constant of 1.0r. The left pair are for divergent stereo viewing and the right pair are for convergent viewing.

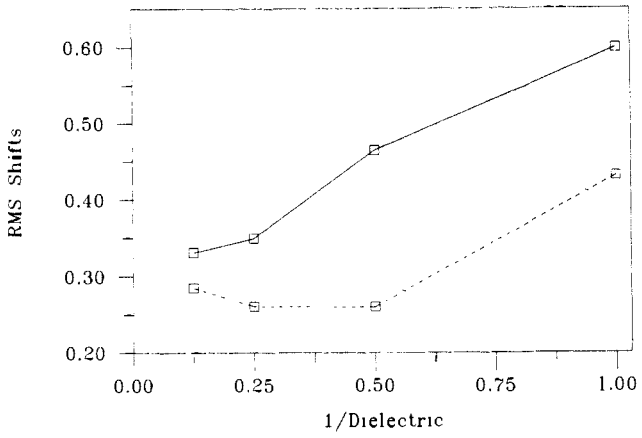


Figure 9. Effect of a nonbonding cutoff as a function of dielectric constant on the rms coordinate shifts from the crystal structure. Both the run without the nonbonding cutoff (solid line) and the run with a 10.0-Å residue-based nonbonding cutoff (dashed line) had distance-dependent dielectric constants and Jorgensen's van der Waals radii.

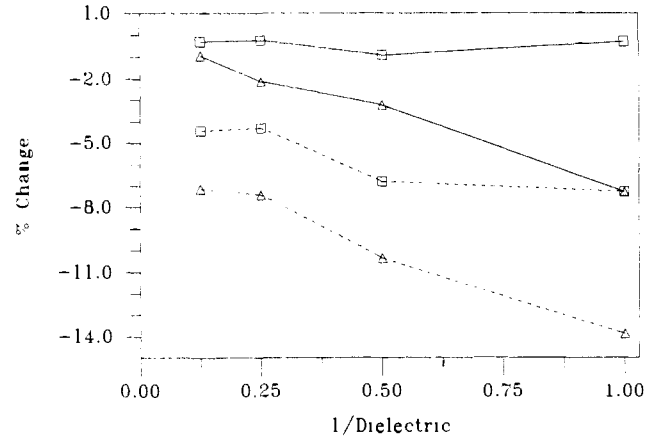


Figure 10. Effects of the "united atom" van der Waals radii and dielectric constant on the percent volume change (squares) and the percent change in the cube of radius of gyration (triangles). Both the original (dashed lines) and Jorgensen's van der Waals radii (solid lines) were run with distance-dependent dielectric constants.

the crystal structure as the nonbonding cutoff was decreased. The non-hydrogen rms coordinate shift shows a shallow minimum at a cutoff of 9.0 Å. Both the radius of gyration and the volume increased as the nonbonding cutoff was decreased. The accessible surface area for the hydrophilic atoms shows improvement with shorter nonbonding cutoffs, but the improvement is not as large as the improvement shown when changing from a constant dielectric to a distance-dependent dielectric electrostatic potential. Short-range electrostatic interactions must play a larger role in the hydrophilic atoms' decrease in accessible surface area. Therefore, they are not affected as strongly by the nonbonding cutoff as are the molecular shape parameters.

There was better agreement with the crystal structure for the mean ϕ and ψ as the nonbonding cutoff was decreased. The α-helices show their best agreement with the crystal structure ϕ and ψ values for a nonbonding cutoff of 9.0 Å and a strong electrostatic potential ($\epsilon = 1.0r$) but show little change for a weak electrostatic potential ($\epsilon = 4.0r$). The β-sheet shows complicated changes as the nonbonding cutoff is decreased. The mean N···O hydrogen bond distance increases by 0.01 Å with a nonbonding cutoff of 7.0 Å from that with no cutoff. For a strong electrostatic potential ($\epsilon = 1.0r$), the number of new hydrogen bonds formed and the number lost due to minimization becomes less as the nonbonding cutoff is decreased.

2. Residue-Based Nonbonding Cutoff. The residue-based nonbonding cutoff shows the same trends as the atom-based cutoff. Unlike the atom-based cutoff, the rms coordinate shifts for the shortest nonbonding cutoff tried, 6.0 Å, was the lowest of all residue-based cutoffs. The volume, radius of gyration, accessible surface area, and mean deviations from the crystal structure for ϕ and ψ all show far less sensitivity to the residue-based cutoff chosen. Both the helices and the β-sheets' mean ϕ and ψ angles seem to be unchanged by the residue-based nonbonding cutoff chosen, for a weak electrostatic potential ($\epsilon = 4.0r$).

van der Waals Radii. Two van der Waals (VDW) radii schemes were tried: the original VDW radii[14] and those of Jorgensen[24] (Table III). Jorgensen's larger VDW radii produce no volume compaction, whereas the original VDW radii show between 4% ($\epsilon = 8.0r$) and 7% ($\epsilon = 1.0$) volume compaction during minimization. The cube of the radius of gyration decreases with decreasing dielectric constant for both van der Waals schemes (Figure 10). For Jorgensen's VDW radii, there was a smaller radius of gyration decrease. Since there is no volume compaction for minimizations with Jorgensen's VDW radii, the cube of the radius of gyration represents the change in the shape of crambin during minimization. With the original VDW radii, the cube of the radius of gyration has contributions from both the decrease in the volume and from the change in molecule shape. The accessible surface area decreases from 1% ($\epsilon = 8.0r$) to 5% ($\epsilon = 1.0$) less for Jorgensen's van der Waals radii compared to the original van der Waals radii. This improvement in the accessible surface area decrease may be due to the smaller volume change. A 0.06 Å ($\epsilon = 2.0r$) to 0.55 Å ($\epsilon = 8.0r$) decrease in the rms coordinate shifts and a smaller deviation for the mean ϕ and ψ angles of the β-sheet from the crystal structure are a result of the smaller change in the molecular shape (radius of gyration) when using Jorgensen's larger VDW. Although the mean deviation from the crystal structure for ϕ and ψ increased, the mean deviation from planarity of ω shows less dependence on the dielectric constant. The average hydrogen bond distance becomes 0.04 Å ($\epsilon = 1.0r$) to 0.10 Å ($\epsilon = 4.0$) longer for Jorgensen's than the original VDW radii. Thus, Jorgensen's VDW radii result in better agreement with the crystal structure of crambin.

C_α Hydrogens. It was postulated that the systematic deviations in the backbone torsion angles ϕ and ψ were due to the "united atom" approximation for the C_α hydrogen (HC_α). The HC_α is the only backbone atom not used in the calculation. There were some improvements in the deviation between the crystal structure and the minimized structure for ϕ and ψ because of the incorporation of the HC_α, but it was not a consistent improvement. The modified extended atom charge scheme was better that the alanyl dipeptide charge scheme, but was no better than the models without HC_α. There was better agreement with the crystal structure for the volume and the radius of gyration with the HC_α for minimizations with the original VDW radii when the HC_α was added. Jorgensen's larger VDW radii compensates for the HC_α atom originally omitted.

Protein Environment. The previous set of minimizations were carried out without solvent. First, the effect of introducing the intermolecular protein hydrogen bonds will be described. Second, the effects of adding a 4.0-Å shell of the crystal environment around the molecule are examined. Finally, the effect of placing a 5.0-Å shell of TIP3P waters around the protein will be discussed. Each of these calculations was done with both the original VDW radii and Jorgensen's VDW radii. Only the results with Jorgensen's VDW will be discussed.

1. Intermolecular Hydrogen Bonds. There are four intermolecular protein hydrogen bonds found in the crystal structure of crambin. The symmetry-related fragments used to create the intermolecular hydrogen bonds were constrained to their initial positions. The minimization employed the strongest electrostatic potential ($\epsilon = 1.0$), and it converged very rapidly (200 cycles). Deviations from the crystal structure were small (0.112 Å) when compared with the minimizations previously discussed (Figure 2).

One of the most affected regions of the protein includes the amide hydrogen of Ala_{38} and the side chain of Thr_{39}. Here in the extended arm of crambin, the backbone temperature factors from the crystal are almost double the value of the rest of the molecule (unpublished results of Teeter and Hendrickson). In

vacuum dynamics simulations of crambin, the extended arm is mobile (Teeter, unpublished results). During minimization without the intermolecular hydrogen bonds, the ψ of Ala_{38} and the ϕ of Thr_{39} changed by 37° and −45°, respectively. With the intermolecular hydrogen bonds, they only move 8° and −9°, respectively. The intermolecular hydrogen bonds maintain a local crystal conformation, which is one of several formed in solution.

2. 4.0-Å Shell of the Crystal Environment. The minimization with the crystal environment had very similar results to those described above for the intermolecular hydrogen bonds. This calculation takes approximately 10 times longer than the intermolecular hydrogen bonds alone.

3. 5.0-Å Shell of Water. Often one does not have or does not want to use the crystal environment in a minimization calculation. Instead one might place a shell of water around the molecule. We have done just that. Crambin was placed in Jorgensen's Monti Carlo generated water box.[25] The water dipoles were aligned with the electronic vectors of the protein in one minimization. In a second, the water was not aligned. Both minimization used a constant dielectric of 1.0 and 10.0-Å residue-based nonbonding cutoff. The minimizations did not converge. After 5000 cycles with the protein constrained and 2500 cycles unconstrained, the minimizations were stopped. The final rms energy gradients were 0.33 kcal/(mol·Å) and 0.27 kcal/(mol·Å) for the aligned and unaligned water dipoles, respectively.

The inclusion of a 5.0-Å shell of water resulted in a much better agreement with the crystal structure for the hydrophilic atoms' accessible surface area. Aligning the water dipoles along the protein electronic vectors showed the best agreement with only a 2% decrease in the hydrophilic atoms' accessible surface area. The unaligned minimization showed a 5% decrease, and the minimization without waters showed a 14% decrease.

Both types of a 5.0-Å shell of water around crambin resulted in a smaller rms coordinate shift from the crystal structure (0.29 Å) than the minimization run without waters (0.42 Å). However, the final gradients were larger for the water shell calculations. The improvement in the cube of the radius of gyration and the mean deviations of ϕ and ψ from the crystal structure may account for the improvement in the rms coordinate shifts. The minimizations with the water shells showed a 2% decrease in their cube of the radius of gyration, whereas without the water shell they showed a 4% decrease. The backbone torsion angles of the α-helices show no improvement, but the β-sheet did show minor improvement with the 5.0-Å shell of water. In the crystal structure, there is a small pocket of water just below the β-sheet. The average hydrogen bond N···O distance with the water shell increased by 0.04 Å for the α-helices and decreased by 0.01 Å for the β-sheet from the minimized structure without a water shell. The 11–15 hydrogen bond distance was reduced to 2.99 Å with the water shell and from 3.16 Å for the crystal structure. Without water shell, its value is 2.92 Å. Thus, the kink in the top helix was not maintained.

Discussion and Conclusions

During minimization, there can be both volume compaction and a molecular shape change. It is necessary to calculate both the radius of gyration and the volume to understand and separate these. The percent change in the radius of gyration is affected by changes both in volume and in the molecular shape. The percent change in volume must be subtracted from the percent change in the cube of the radius of gyration. Thus, the change in the cube of the radius of gyration does not approximate the change in volume alone.

By using this analysis, we were able to show (1) the volume was affected primarily by the choice of the "united atom" van der Waals radii (the original vs. Jorgensen's) and (2) the molecular shape was most affected by the strength of the long distance electrostatic forces. We showed that Jorgensen's "united atom" van der Waals radii gave the best results, leading to little or no volume compaction. Weiner et al.[6] had previously shown that Jorgensen's VDW results in less volume compaction. Increasing the electrostatic forces resulted in crambin becoming more spherical by moving the α-helices closer to the extended arm. As crambin became more spherical, there were larger deviations from the crystal structure in the β-sheet, located between the α-helices and the extended arm. Minimizations with intermolecular protein hydrogen bonds show that crambin is held in the crystal structure by them, and one might expect some change in shape in going from the crystal environment to a vacuum or to solution.

The intermolecular protein hydrogen bonds hold higher energy crystal conformations in position. Without the hydrogen bonds, these regions can move to a lower energy minimum. This suggests that (1) the intermolecular protein hydrogen bonds stabilize local higher energy conformations found less often in solution and (2) during minimization without these intermolecular hydrogen bonds, the regions take many cycles to refine. While this occurs, other parts of the molecule shift to lower energy gradients, and the imperfections of the force field show up. The minimization with the intermolecular hydrogen bonds converges rapidly, and the imperfections in the force field do not show up.

The negatively correlated shift of the backbone torsion angles ϕ and ψ are a function of both the electrostatic and van der Waals interactions. We have shown that increasing the electrostatic potential by lowering dielectric constant, by changing from distance-dependent to constant dielectric potential, or by increasing the nonbonding cutoff results in an increase in the deviation of ϕ and ψ from those of the crystal structure. The hydrophilic atoms become less accessible to the surface with stronger electrostatic potentials. We postulate that this is due to the lack of explicit solvent in the in vacuo calculations. When explicit solvent was included in the minimization, the accessible surface area showed better agreement with the crystal structure. Berendsen and van Gunsteren[31] had previously shown that explicit water improves their simulations of proteins. Increasing the van der Waals radii from the original radii to those of Jorgensen's also results in increased deviation of the mean values of ϕ and ψ. When intermolecular protein hydrogen bonds were included, the ϕ and ψ deviations did not occur suggesting that these deviations have low potential energy gradients, and they are not seen unless some other part of the structure is converging slowly.

We postulated that these systematic deviations in ϕ and ψ were due to the "united atom" approximation for the C_α hydrogen (omission of HC_α). The HC_α is the only backbone atom not used in the calculation. The results of the minimizations with the HC_α did not consistently improve the deviations in ϕ and ψ. Thus, the use of HC_α in potential energy minimizations is not recommended. Further examination of this problem will be required. The effect of using an all-atom potential should be examined.

The mean N···O hydrogen bond distance decreased as the dielectric constant decreased. The distance-dependent dielectric electrostatic potential resulted in shorter N···O distances than the constant dielectric electrostatic potential because it decreased the N···O electrostatic repulsion relative to the H···O attraction. The longer N···O distance has weaker interaction energy. The larger van der Waals radii of Jorgensen lengthened the mean N···O hydrogen bond distance by 0.04 to 0.10 Å. Consequently, the best agreement with the crystal structure was with Jorgensen's van der Waals radii for $\epsilon = 2.0$ (constant dielectric) and $\epsilon = 4.0r$ (distance-dependent dielectric).

In order to determine the best potential energy parameters to be used without the intermolecular hydrogen bonds from the crystal structure, we must examine structural parameters that are independent of the local region movements. We have chosen to look at the volume, the hydrogen bond geometry, rms deviation from planarity for ω, and the backbone torsion angles ϕ and ψ for the α-helices (since the β-sheet is affected by shape changes). Jorgensen's van der Waals radii gave the best overall results. For it, there was little or no volume change.

Other structural features varied with regard to the optimal electrostatic potential. The mean N···O hydrogen bond distance showed the best agreement for $\epsilon = 2.0$ (constant dielectric) and

(31) Berendsen, H. J. C.; van Gunsteren, D. F. *Ann. N.Y. Acad. Sci.* **1986**, in press.

$\epsilon = 4.0r$ (distance-dependent dielectric). The rms deviation in ω from planarity agreed best with the crystal for weak electrostatic potentials ($\epsilon = 4.0$ and $4.0r$). Similar results were found by Vedani and Dunitz.[32] The best results used a 6.0-Å residue-based nonbonding cutoff with a distance-dependent dielectric constant of $4.0r$. Similar results could be obtained with a 9.0-Å atom-based nonbonding cutoff. We found that the results with explicit solvent or large portions of the crystal environment did not justify the extra computational time needed to run them.

Thus, empirically the best overall condition for minimization with AMBER are Jorgensen's "united atom" VDW, $\epsilon = 4.0r$, and 6.0-Å residue-based or a 9.0-Å atom-based nonbonding cutoff. A dielectric constant (ϵ) of $4.0r$ is a weak electrostatic potential. These appear to be necessary in minimizations in order to correct for the omissions of solvent and of atomic motion (i.e., to model the room temperature structure). Inclusion of solvent explicitly results in better agreement for the surface of the protein but does not remove the need for a weaker electrostatic potential. Our results suggest that further improvements in potential functions are still needed to model high resolution structures and to ultimately predict protein structure accurately.

(32) Vedani, A.; Dunitz, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7653-7658.

Acknowledgment. This research was supported by NIH Grant number GM 30165 and NSF Grant PCM 8303024 to M.M.T. M.W. acknowledges assistance from GM 30165. Surface diagrams were drawn by Dr. Gary J. Quigley's PLT1 program, which was run on a VAX 11/750. We would like to thank Dr. Scott Weiner for sending us the most current version of AMBER and Jorgensen's Monti Carlo simulation of water.[25]

The Structures and Energetics of Fluorine-Substituted Phosphonium Ylides

David A. Dixon* and Bruce E. Smart*

Contribution No. 4047 from E. I. du Pont de Nemours and Company, Central Research and Development Department, Experimental Station, Wilmington, Delaware 19898.
Received March 3, 1986. Revised Manuscript Received August 16, 1986

Abstract: The electronic structures and energies of the model ylides H_3PCXY and their isomeric phosphines H_2PCHXY (X = Y = H, F, CF_3 and X = H, Y = F) have been determined from ab initio molecular orbital theory at the SCF level. The calculations were performed with a double-ζ basis set augmented by polarization (d) functions on both the P atom and the C atom bonded to P and a set of diffuse s and p functions on the C atom bonded to P. The H_3PCH_2, H_3PCHF, and $H_3PC(CF_3)_2$ ylides exhibit varying degrees of P-C multiple bonding and zwitterionic character. The ylidic carbon in H_3PCHF is more pyramidal than that in H_3PCH_2, but it is planar in $H_3PC(CF_3)_2$. The structure of H_3PCF_2 is entirely different, and it closely resembles the isolated H_3P and $:CF_2(^1A_1)$ species separated by 3.54 Å. The energies for ylide to phosphine isomerizations were calculated to assess ylide stabilities and to derive ylide heats of formation. From the calculated energies of ylide to lowest singlet state carbene dissociations, the H_3PCH_2, H_3PCHF, $H_3PC(CF_3)_2$, and H_3PCF_2 binding energies were found to be 53.2, 16.6, 77.3, and 1.2 kcal/mol, respectively. The theoretical properties of the fluorinated ylides are compared to available experimental data and known solution chemistry.

Several recent studies[1-5] have used molecular orbital theory to characterize the structure and energetics of ylides, which are important synthetic intermediates.[6] These studies focussed on simple model structures of the form $H_3X^+CH_2^-$ (X = N, P) and $H_2X^+CH_2^-$ (X = O, S) where the substituent on carbon is always hydrogen. The model ylide $H_3P^+CH_2^-$ has been studied in detail and is expected to have partial double bond character in the P-C bond. The calculated value for r(P-C) is in reasonable agreement with that observed in X-ray structure determinations of substituted ylides.[7-10]

Fluorinated phosphonium ylides also have proved to be useful intermediates in synthesis,[11-13] but their properties vary dramatically depending upon whether fluorine is substituted α or β to the ylidic carbon. Ylides that have no α-fluorines ordinarily are isolable. For example, $(C_2H_5O)_3PC(CF_3)_2$,[14] $Ph_3PC(CF_3)_2$,[15] and $Ph_3PC(CF_3)C_2F_5$[16] have been made and fully characterized, and the crystal structure of **1** has been determined.[16,17] By

(1) (a) Eades, R. A.; Gassman, P. G.; Dixon, D. A. *J. Am. Chem. Soc.* **1981**, *103*, 1066. (b) Dixon, D. A.; Dunning, T. H., Jr.; Eades, R. A.; Gassman, P. G. *J. Am. Chem. Soc.* **1983**, *105*, 7011.
(2) (a) Bernardi, F.; Schlegel, H. B.; Whangbo, M.-H.; Wolfe, S. *J. Am. Chem. Soc.* **1977**, *99*, 5633. (b) Mitchell, D. J.; Wolfe, S.; Schlegel, H. B. *Can. J. Chem.* **1981**, *59*, 3280.
(3) (a) Lischka, H. *J. Am. Chem. Soc.* **1977**, *99*, 353. (b) Strich, H. *Nouv. J. Chim.* **1979**, *3*, 105.
(4) (a) Absar, I.; Van Wazer, J. F. *J. Am. Chem. Soc.* **1972**, *94*, 2382. (b) Hoffman, R.; Boyd, D. B.; Goldberg, S. Z. *J. Chem. Soc.* **1970**, *92*, 3929. (c) Graham, S. L.; Heathcock, C. H. *J. Am. Chem. Soc.* **1980**, *102*, 3713. (d) Kral, V.; Arnold, A. *Collect. Czech., Chem. Commun.* **1980**, *45*, 80, 92. (e) Trinquier, G.; Malrieu, J. P. *J. Am. Chem. Soc.* **1979**, *101*, 7169.
(5) (a) Starzewski, K. A. O.; Dieck, H. T.; Bock, H. *J. Organomet. Chem.* **1974**, *65*, 311. (b) Starzewski, K. A. O.; Bock, H. *J. Am. Chem. Soc.* **1976**, *98*, 8486.
(6) For reviews, see: (a) Bestmann, H. J.; Vostrowsky, D. *Top. Curr. Chem.* **1983**, *109*, 423. (b) Pommer, H.; Thieme, P. C. *Top. Curr. Chem.* **1983**, *109*, 165. (c) Johnson, A. W. *Ylide Chemistry*; Academic: New York, 1966. (d) Trost, B. M.; Melvin, L. S., Jr. *Sulfur Ylids*; Academic: New York, 1975.
(7) Bart, J. C. J. *J. Chem. Soc. B* **1969**, 350.
(8) (a) Wheatley, P. J. *J. Chem. Soc.* **1965**, 5785. (b) Stephens, F. S. *J. Chem. Soc.* **1965**, 5640, 5658. (c) Mak, T. C. W.; Trotter J. *Acta Crystallogr.* **1965**, *18*, 81. (d) Kennard, O.; Motherwell, W. D. S.; Coppola, J. C. *J. Chem. Soc. C* **1971**, 2461.
(9) Chioccola, G.; Daly, J. J. *J. Chem. Soc. A* **1968**, 568.
(10) Bailey, A. S.; Peach, J. N.; Cameron, T. S.; Prout, C. K. *J. Chem. Soc. C* **1969**, 2295.
(11) Burton, D. J. *J. Fluorine Chem.* **1983**, *23*, 339.
(12) Wheaton, G. A.; Burton, D. J. *J. Org. Chem.* **1983**, *48*, 917.
(13) Burton, D. J.; Greenlimb, P. E. *J. Org. Chem.* **1975**, *40*, 2796.
(14) Middleton, W. J.; Sharkey, W. H. *J. Org. Chem.* **1965**, *30*, 1384.
(15) Burton, D. J.; Inouye, Y. *Tetrahedron Lett.* **1979**, 3397.
(16) Burton, D. J.; Shinya, S.; Howells, R. D. *J. Am. Chem. Soc.* **1979**, *101*, 3689.
(17) Howells, M. A.; Howells, R. D.; Baenziger, N. C.; Burton, D. J. *J. Am. Chem. Soc.* **1973**, *95*, 5366.

0002-7863/86/1508-7172$01.50/0