JOURNAL

OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 110, NUMBER 6 MARCH 16, 1988

The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin

William L. Jorgensen* and Julian Tirado-Rives

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received January 26, 1987*

Abstract: A complete set of intermolecular potential functions has been developed for use in computer simulations of proteins in their native environment. Parameters are reported for 25 peptide residues as well as the common neutral and charged terminal groups. The potential functions have the simple Coulomb plus Lennard-Jones form and are compatible with the widely used models for water, TIP4P, TIP3P, and SPC. The parameters were obtained and tested primarily in conjunction with Monte Carlo statistical mechanics simulations of 36 pure organic liquids and numerous aqueous solutions of organic ions representative of subunits in the side chains and backbones of proteins. Bond stretch, angle bend, and torsional terms have been adopted from the AMBER united-atom force field. As reported here, further testing has involved studies of conformational energy surfaces and optimizations of the crystal structures for four cyclic hexapeptides and a cyclic pentapeptide. The average root-mean-square deviation from the X-ray structures of the crystals is only 0.17 Å for the atomic positions and 3% for the unit cell volumes. A more critical test was then provided by performing energy minimizations for the complete crystal of the protein crambin, including 182 water molecules that were initially placed via a Monte Carlo simulation. The resultant root-mean-square deviation for the non-hydrogen atoms is still ca. 0.2 Å and the variation in the errors for charged, polar, and nonpolar residues is small. Improvement is apparent over the AMBER united-atom force field which has previously been demonstrated to be superior to many alternatives.

Computer simulations are undoubtedly destined to become an increasingly important means for investigating the structures and dynamics of biomolecular systems.[1] At the heart of such theoretical calculations are the force fields that describe the interatomic interactions and the mechanics of deformations of the molecules.[2] There is also little doubt that there will be a continual evolution in force fields with added complexity and improved performance paralleling the availability of computer resources. Our own efforts in this area over the last few years have resulted in the OPLS potential functions for proteins whose development and performance are summarized here. These potential functions have a simple form and they have been parametrized directly to reproduce experimental thermodynamic and structural data on fluids. Consequently, they are computationally efficient and their description of proteins in solution or crystalline environments should be superior to many alterantives that have been developed with limited condensed-phase data. The latter point is pursued here primarily through calculations on the crystal structures for four cyclic hexapeptides, a cyclic pentapeptide, and the protein crambin. Improvements are apparent in comparison to the AMBER united-atom force field[3] which has previously been shown to be superior to many alternatives.[4]

Parametrization

The peptide residues of proteins contain readily identifiable organic subunits such as amides, hydrocarbons, alcohols, thioethers, etc. In view of this and since data are available on the corresponding pure organic liquids, our approach to developing a force field for proteins was to build it up from parameters demonstrated to yield good descriptions of organic liquids. Ultimately, the force field would need to treat both intramolecular terms for bond stretches, angle bends, and torsions, as well as the intermolecular and intramolecular nonbonded interactions. The latter are generally accepted to be the most difficult part of the problem and have been our focus.[3] A simple, computationally efficient form was chosen to represent the nonbonded interactions through Coulomb and Lennard-Jones terms interacting between sites centered on nuclei (eq 1). Thus, the intermolecular interaction energy between molecules a and b is given by the sum of interactions between the sites on the two molecules. The nonbonded contribution to the intramolecular energy is evaluated with the same expression for all pairs of sites separated by more than three bonds. In the OPLS (optimized potentials for liquid simulations) model, each atomic nucleus has an interaction site, except CH_n groups are treated as united atoms centered on the carbon. It is important to note that in this model *no special functions were*

$$\Delta E_{ab} = \sum_{i}^{\text{on a}} \sum_{j}^{\text{on b}} (q_i q_j e^2 / r_{ij} + A_{ij}/r_{ij}^{12} - C_{ij}/r_{ij}^{6}) \qquad (1)$$

(1) Beveridge, D. L., Jorgensen, W. L., Eds. *Ann. N.Y. Acad. Sci.* **1986**, *482*.

(2) For reviews, see: (a) Levitt, M. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 251. (b) McCammon, J. A. *Rep. Prog. Phys.* **1984**, *47*, 1.

(3) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(4) Hall, D.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 411.

0002-7863/88/1510-1657$01.50/0

Table I. Liquids Simulated with the OPLS Potential Functions

liquid	T (°C)	ref	liquid	T (°C)	ref
$HCONH_2$	25	5	pyrrole	25	9
$HCON(CH_3)_2$	25, 100	5	pyridine	25	9
$CH_3CONHCH_3$	100	5	CH_4	−161	10
CH_3OH	25	6	C_2H_6	−89	10
C_2H_5OH	25	6	C_3H_8	−42, 25	10
n-C_3H_7OH	25	6	n-C_4H_{10}	−0.5, 25	10
i-C_3H_7OH	25	6	i-C_4H_{10}	25	10
t-C_4H_9OH	25	6	n-C_5H_{12}	25	10
CH_3SH	6	7	i-C_5H_{12}	25	10
C_2H_5SH	25	7	neo-C_5H_{12}	25	10
$(CH_3)_2S$	25	7	c-C_5H_{10}	25	10
$C_2H_5SCH_3$	25	7	n-C_6H_{14}	25	10
$(C_2H_5)_2S$	25	7	$CH_3CH_2CH{=}CH_2$	25	10
CH_3SSCH_3	25	7	t-$CH_3CH{=}CHCH_3$	25	10
$(CH_3)_2O$	−25	8	C-$CH_3CH{=}CHCH_3$	25	10
$C_2H_5OCH_3$	25	8	$(CH_3)_2C{=}CH_2$	25	10
$(C_2H_5)_2O$	25	8	benzene	25	10
THF	25	8	$CH_3CO_2CH_3$	25	8

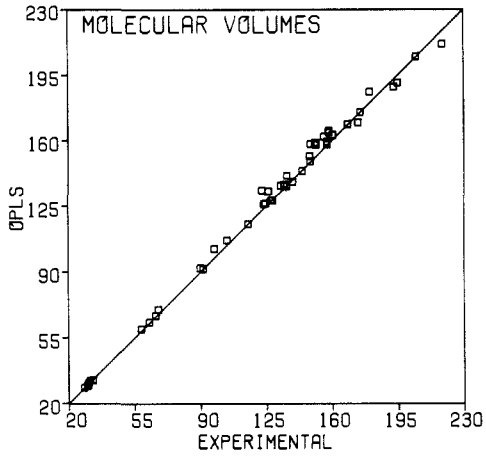


Figure 1. Comparison of computed and experimental volumes per molecule in Å^3 for the liquids in Table I and TIP4P water.

found to be needed to describe hydrogen bonding and there are no additional interaction sites for lone pairs. Another important point is that standard combining rules are used for the Lennard-Jones interactions such that $A_{ij} = (A_{ii}A_{jj})^{1/2}$ and $C_{ij} = (C_{ii}C_{jj})^{1/2}$. The A and C parameters may also be expressed in terms of Lennard-Jones σ's and ϵ's as $A_{ii} = 4\epsilon_i\sigma_i^{12}$ and $C_{ii} = 4\epsilon_i\sigma_i^6$.

The OPLS parameters for the 20 neutral peptide residues reported here were obtained primarily via Monte Carlo simulations for the 36 organic liquids listed in Table I.[5-10] Standard geometries were used for the molecules with fixed bond lengths and bond angles, though torsional motion was included, as described in detail elsewhere.[5-10] Particular emphasis was placed on reproducing the experimental densities and heats of vaporization for the liquids. In view of the simplicity of the functional form (eq 1), the accord with the experimental data is remarkable as illustrated in Figures 1 and 2; the average deviation between the experimental data and the theoretical results is less than 3%. The structural results for the liquids were also shown to be in accord with available experimental data including vibrational spectroscopy and diffraction data for formamide, dimethylformamide (DMF), methanol, ethanol, 1-propanol, 2-methyl-2-propanol, methane, ethane, neopentane, and benzene. The hydrogen bonding in the alcohols, thiols, and amides is well-represented by the OPLS potential functions. It should be noted that the number of unique parameters has been kept to a minimum.[5-10] Thus, only 12 different CH_n groups are used to describe all alkanes, alkenes,

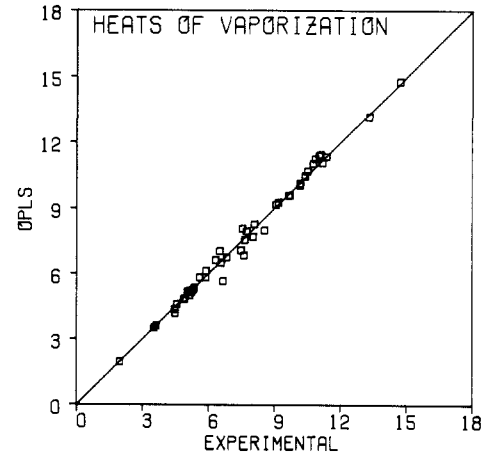


Figure 2. Comparison of computed and experimental heats of vaporization in kcal/mol for the liquids in Table I and TIP4P water.

and benzene,[10] and, for example, the parameters for the OH groups in all alcohols[6] and the carbonyl groups in all amides are the same.[5]

The parametrization for the neutral residues also entailed careful consideration of the interactions between the organic fragments and a water molecule. The water model used in conjunction with the OPLS potentials was TIP4P,[11,12] though the TIP3P[11] or SPC[13] models yield very similar results. For most purposes, these three alternatives may be considered to be interchangeable, though the slightly more complicated TIP4P model gives a better description of the angular variation of hydrogen bond energies. Complexes of a water molecule with amides, ethers, esters, alcohols, thiols, sulfides, azoles, and azines were studied with the OPLS potentials as well as ab initio molecular orbital calculations primarily with the 6-31G(d) basis set.[14] The trends in the ab initio findings for the hydrogen bond strengths and geometries are well reproduced by the OPLS results.[5-9,15] Furthermore, Monte Carlo simulations were carried out for dilute aqueous solutions of formamide,[15] N-methylacetamide (NMA),[15] DMF,[15] methanol,[16] and seven alkanes.[17] For the amides, experimental structural data are limited; however, the computed numbers of amide–water hydrogen bonds are reasonable and the computed heats of hydration, ca. −20 kcal/mol, are in the correct range.[15] Similarly, the hydration of methanol appears reasonable and the computed difference in free energies of hydration for methanol and ethane, 6.75 ± 0.2 kcal/mol, is in excellent accord with the experimental value, 6.93 kcal/mol.[16] The free energy calculations are a powerful diagnostic tool, but very demanding on computer resources.[16] The results for the hydrophobic hydration of the alkanes also revealed no aberrations and yielded pleasing correlations between numbers of water molecules in the first hydration shells and experimental enthalpies and entropies of hydration.[17]

The parametrization for the five charged protein residues, Asp, Glu, Hip (protonated His), Lys, and Arg, and terminal ammonium and carboxylate groups required a somewhat different approach. Since corresponding pure organic liquids cannot be construed in these cases, the emphasis was placed on comparisons with ab initio results for ion–molecule complexes and on Monte Carlo simulations for hydrated ions. Specifically, parameters for Lys, Glu, Asp and the charged terminal groups were developed through a

(5) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569.
(6) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.
(7) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 6379.
(8) Jorgensen, W. L.; Briggs, J. M., to be published.
(9) Jorgensen, W. L.; Contreras, L., to be published.
(10) Jorgensen, W. L.; Madura, M. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(12) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.
(13) Berendsen, H. J. C.; Postma, J. P. M.; von Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1981; p 331.
(14) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1983**, *77*, 3054.
(15) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.
(16) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(17) Jorgensen, W. L.; Gao, J.; Ravimohan, C. *J. Phys. Chem.* **1985**, *89*, 3470.

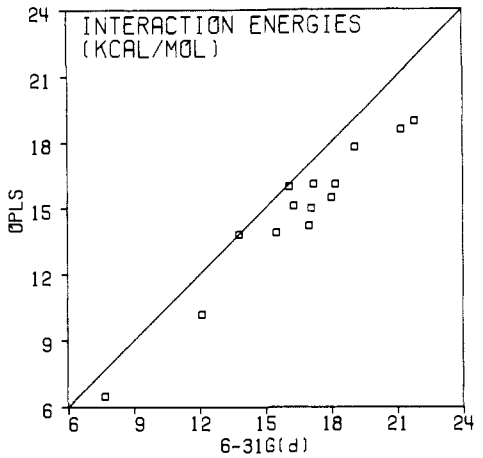
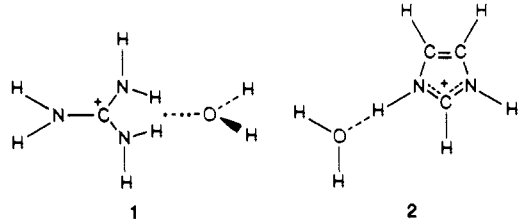


Figure 3. Comparison of interaction energies (kcal/mol) for ion–water complexes obtained with the OPLS potential functions and ab initio 6-31G(d) calculations.

general study of the hydration of ammonium and carboxylate ions.[18] Ab initio calculations were carried out with the 6-31G(d) basis set for low-energy forms of complexes between water and NH_4^+, $CH_3NH_3^+$, and $HCOO^-$.[18,19] The OPLS parameters were chosen to reproduce the resultant optimal geometries and interaction energies, which are also in good accord with gas-phase experimental data.[18,19] In addition, the OPLS parameters were required to yield good agreement with experimental heats of hydration for NH_4^+, $CH_3NH_3^+$, $(CH_3)_4N^+$, $HCOO^-$, and CH_3COO^-.[18] This was demonstrated through Monte Carlo simulations for the five ions in dilute aqueous solution.[18] The structural results were also shown to mirror experimental estimates of hydration numbers for the ammonium and carboxylate groups in Lys, Glu, and Asp from NMR studies of frozen polypeptide solutions.[18,20]

Recently, the OPLS parameters for Arg and Hip have been obtained by fitting to ab initio 6-31G(d) results for complexes of water with guanidinium ion and protonated imidazole.[21] The principal concern was the charge distributions for the ions since the Lennard-Jones parameters were adopted from standard values for nitrogen and carbons (all explicit hydrogens have $\sigma = \epsilon = 0$ in the OPLS potentials). The accord between the OPLS and 6-31G(d) results for low-energy geometries is uniformly good. For example, the OPLS optimal interaction energy and CO distance for **1** are 16.1 kcal/mol and 3.33 Å, whereas the 6-31G(d) values with fixed water and guanidinium geometries are 18.2 kcal/mol and 3.41 Å. And, for **2**, the OPLS predictions for the interaction energy and NO distance are 16.0 kcal/mol and 2.72 Å versus the



1 **2**

6-31G(d) values of 16.1 kcal/mol and 2.85 Å. In general, the accord between the OPLS and 6-31G(d) results is good as illustrated in Figures 3 and 4 for 14 low-energy geometries of water with NH_4^+, $CH_3NH_3^+$, $HCOO^-$, guanidinium ion, and protonated imidazole. The OPLS interaction energies are deliberately designed to be less than the 6-31G(d) results, since the latter are typically somewhat greater than the limited experimental data.[18,19] At this time, fluid simulations have not been executed for guanidinium ion or protonated imidazole in water. Experimental thermodynamic data do not appear to be available in these cases.

The OPLS parameters obtained in this way for 25 common peptide residues and both neutral and charged terminal residues

(18) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.
(19) Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784.
(20) Kuntz, I. D. *J. Am. Chem. Soc.* **1971**, *93*, 514.
(21) Jorgensen, W. L.; Gao, J., unpublished results.

Table II. OPLS Atom and Group Assignments for Proteins[a]

residue	atom or group	type	residue	atom or group	type
		Main Chains			
Gly	N	3	Ala	N	3
	H(N)	4		H(N)	4
	CH_2^α	5		CH^α	6
	C	1		C	1
	O	2		O	2
Pro	N	3	Aib	N	3
	CH^α	14		H(N)	4
	C	1		C^α	64
	O	2		C	1
				O	2
		Side Chains			
Ala	CH_3^β	7	Val	CH^β	8
Aib	CH_3^β	65		CH_3^γ	7
Pro	CH_2^β	9	Leu	CH_2^β	9
	CH_2^γ	9		CH^γ	8
	CH_2^δ	15		CH_3^δ	7
Ile	CH^β	8	Phe	CH_2^β	9
	CH_2^γ	9		C^γ	11
	CH_3^γ	7		CH^δ	11
	CH_3^δ	10		CH^ϵ	11
				CH^ζ	11
Ser	CH_2^β	22	Cys	CH_2^β	31
	O^γ	23		S^γ	32
	H^γ	24		H^γ	33
Thr	CH^β	25	Met	CH_2^β	9
	O^γ	23		CH_2^γ	34
	$H^\gamma(O)$	24		S^δ	35
	CH_3^γ	7		CH_3^ϵ	36
Tyr	CH_2^β	9	Cystine	CH_2^β	37
	C^γ	11		S^γ	38
	CH^δ	11	Hyp	CH_2^β	9
	CH^ϵ	11	(Pro-OH)	CH^γ	25
	C^ζ	26		CH_2^δ	15
	O^η	23		O^δ	23
	H^η	24		$H^\delta(O)$	24
Asn	CH_2^β	9	Gln	CH_2^β	9
	C^γ	1		CH_2^γ	9
	O^δ	2		C^δ	1
	N^δ	12		O^ϵ	2
	$H^\delta(N)$	13		N^ϵ	12
				$H^\epsilon(N)$	13
Asp	CH_2^β	16	Glu	CH_2^β	9
	C^γ	17		CH_2^γ	16
	O^δ	18		C^δ	17
				O^ϵ	18
His	CH_2^β	9	Hip	CH_2^β	9
	C^γ	45	($His\text{-}H^+$)	C^γ	49
	N^δ	40		N^δ	46
	$H^\delta(N)$	41		$H^\delta(N)$	47
	CH^δ	44		CH^δ	49
	CH^ϵ	43		CH^ϵ	48
	N^ϵ	42		N^ϵ	46
				$H^\epsilon(N)$	47
Trp	CH_2^β	9	Arg	CH_2^β	9
	C^γ	50		CH_2^γ	57
	CH^δ	45		CH_2^δ	56
	C^δ	50		N^ϵ	54
	N^ϵ	40		$H^\epsilon(N)$	55
	$H^\epsilon(N)$	41		C^ζ	53
	C^ϵ	45		N^η	51
	CH^ϵ	11		$H^\eta(N)$	52
	CH^ζ	11	Hyl	CH_2^β	9
	CH^η	11	(Lys-OH)	CH_2^γ	9
Lys	CH_2^β	9		CH^δ	25
	CH_2^γ	9		O^ϵ	23
	CH_2^δ	9		$H^\epsilon(O)$	24
	CH_2^ϵ	19		CH_2^ϵ	19
	N^ζ	20		N^ζ	20
	$H^\zeta(N)$	21		$H^\zeta(N)$	21

[a] Nomenclature for atoms: ref 22.

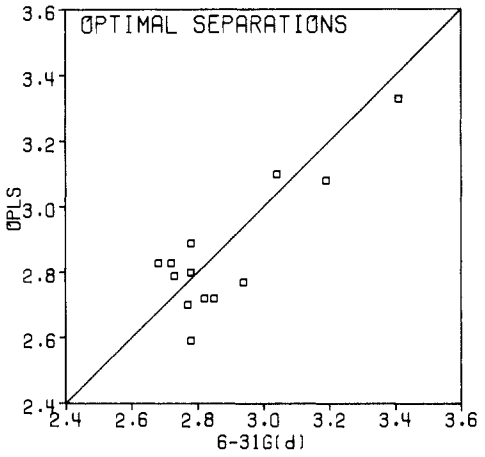


Figure 4. Comparison of optimal separations in Å for ion–water complexes obtained with the OPLS potential functions and ab initio 6-31G(d) calculations.

Table III. OPLS Atom and Group Assignments for Terminal Residues

residue	atom or group	type
	Charged Termini	
$H_3N^+CHRC{=}O$	N	20
	H(N)	21
	CH^{α}	29
	CH_2^{α} (R=H)	27
	C	1
	O	2
$NHCHRCO_2^-$	N	3
	H(N)	4
	CH^{α}	30
	CH_2^{α} (R=H)	28
	C	17
	O	18
	Neutral Termini	
$NHCHRC(O)OCH_3$	N	3
	H(N)	4
	CH^{α}	60
	CH_2^{α} (R=H)	61
	C	58
	O	59
	$O(CH_3)$	62
	CH_3	63
$CH_3C(O)(NHCHRC(O))$	CH_3	7
	C	1
	O	2
$(NHCHRC(O))NHCH_3$	N	3
	H(N)	4
	CH_3	39

are summarized in Tables II–IV. The atom and CH_n group type assignments are given in Tables II and III with use of standard notation,[22] while the actual charges and Lennard-Jones parameters are in Table IV. In all, 65 unique atom and group types are designated, though the number of unique sets of Lennard-Jones parameters is only 19. For reference, the parameters for the TIP4P, TIP3P, and SPC models for water are provided in Table V with use of consistent units. It should be noted that the side chains are each charge balanced to a net charge of 0, +1 or −1. The only charged side chains are for Asp, Glu, Hip, Lys, and Arg. Also, all residues use the Ala backbone except Gly, Pro, and Aib.

Further testing of the OPLS potentials then ensued after incorporation into the AMBER program.[3]

Merger with AMBER

In order to provide a complete energetic description of biomolecular systems, the intramolecular terms for bond length and bond angle variations as well as the torsions and nonbonded terms need to be included. Since substantial work has been done on the former items by others,[2,3] merger of the OPLS nonbonded potential functions and the local vibration and torsional functions from another force field could be considered. AMBER[3] was chosen because it is widely used and because of its documented success in comparison to 15 other force fields for calculations of the crystal structures of 3 cyclic hexapeptides, though we recognize that the test was limited since only Gly and Ala residues were represented.[4]

The bond stretch and angle bend terms in AMBER are quadratic, while the torsional potentials consist of a cosine term

(22) IUPAC–IUB Commission on Biochemical Nomenclature: *Biochemistry* **1970**, *9*, 3471.

Table IV. OPLS Parameters for Proteins

type	q	σ, Å	ϵ, kcal/mol
1	0.500	3.750	0.105
2	−0.500	2.960	0.210
3	−0.570	3.250	0.170
4	0.370	0.0	0.0
5	0.200	3.800	0.118
6	0.200	3.800	0.080
7	0.0	3.910	0.160
8	0.0	3.850	0.080
9	0.0	3.905	0.118
10	0.0	3.905	0.175
11	0.0	3.750	0.110
12	−0.850	3.250	0.170
13	0.425	0.0	0.0
14	0.285	3.800	0.080
15	0.285	3.800	0.118
16	−0.100	3.905	0.118
17	0.700	3.750	0.105
18	−0.800	2.960	0.210
19	0.310	3.905	0.118
20	−0.300	3.250	0.170
21	0.330	0.0	0.0
22	0.265	3.905	0.118
23	−0.700	3.070	0.170
24	0.435	0.0	0.0
25	0.265	3.850	0.080
26	0.265	3.750	0.110
27	0.310	3.800	0.118
28	0.100	3.800	0.118
29	0.310	3.800	0.080
30	0.100	3.800	0.080
31	0.180	3.905	0.118
32	−0.450	3.550	0.250
33	0.270	0.0	0.0
34	0.235	3.800	0.118
35	−0.470	3.550	0.250
36	0.235	3.800	0.170
37	0.300	3.800	0.118
38	−0.300	3.550	0.250
39	0.200	3.800	0.170
40	−0.570	3.250	0.170
41	0.420	0.0	0.0
42	−0.490	3.250	0.170
43	0.410	3.750	0.145
44	0.100	3.750	0.145
45	0.130	3.750	0.145
46	−0.540	3.250	0.170
47	0.460	0.0	0.0
48	0.500	3.750	0.145
49	0.330	3.750	0.145
50	−0.055	3.750	0.145
51	−0.800	3.250	0.170
52	0.460	0.0	0.0
53	0.640	2.250	0.050
54	−0.700	3.250	0.170
55	0.440	0.0	0.0
56	0.310	3.905	0.118
57	0.070	3.905	0.118
58	0.550	3.750	0.105
59	−0.450	2.960	0.210
60	0.250	3.800	0.080
61	0.250	3.800	0.118
62	−0.400	3.000	0.170
63	0.250	3.800	0.170
64	0.200	3.800	0.050
65	0.0	3.960	0.145

Table V. Parameters for Water Models

model	geometry	site	q	σ, Å	ϵ, kcal/mol
TIP4P[a]	r(OH) = 0.9572 Å	O	0.0	3.15365	0.1550
	r(OM) = 0.1500 Å	H	0.520	0.0	0.0
	∠HOH = 104.52°	M	−1.040	0.0	0.0
TIP3P[a]	r(OH) = 0.9572 Å	O	−0.834	3.15061	0.1521
	∠HOH = 104.52°	H	0.417	0.0	0.0
SPC[b]	r(OH) = 1.0000 Å	O	−0.820	3.16557	0.1554
	∠HOH = 109.47°	H	0.410	0.0	0.0

[a] Reference 11. [b] Reference 13.

Table VI. Relative Energies for Conformations of Butane[a]

method	gauche	cis	ref
AMBER/OPLS	1.03	7.08	this work
AMBER—normal	0.89	6.97	this work
AMBER—big	0.37	5.56	this work
AMBER—all atom	0.58	4.57	24
MM2	0.88	4.73	25
MP3/6-311G** + ZPE	0.7	6.0	26
Raman, gas phase	0.89	4.52	27
IR, gas phase	0.97		28
ED, gas phase	0.65	3.6	29

[a] Energies relative to the trans conformer in kcal/mol.

plus the 1,4-nonbonded interaction, both Coulombic and Lennard-Jones. Thus, the torsional potentials are affected by the choice of nonbonded parameters. Furthermore, the 1,4-nonbonded interactions are scaled in AMBER by dividing by factors SCNB and SCEE for the Lennard-Jones and Coulombic terms, respectively. The default value for SCEE is 2.0 and has been used in all calculations reported here. The default value for SCNB is also 2.0 when the "normal" AMBER nonbonded parameters are used.[3] However, in the note added in proof in ref 3, an alternative set of "big" parameters was proposed for CH, CH_2, and CH_3 united atoms adopted from the TIPS potentials.[23] In this case, the recommended SCNB is 8.0.[3] For the purpose of merging the OPLS and AMBER force fields in an uncomplicated manner, it was necessary to readdress the best choices for SCNB and SCEE. This was done by choosing values that gave reasonable agreement between results for conformational surfaces with AMBER/OPLS and "normal" AMBER. These tests are summarized in the next section, followed by more significant tests of the two force fields on crystal structures.

The calculations were executed by using a modified version of AMBER 2.0 on a Microvax II computer in our laboratory. Complete geometry optimizations were carried out with the conjugate gradients procedure.[3] All of the calculations employed a dielectric constant of 1 for evaluating the electrostatic energy. This is the proper choice since the OPLS parameters have been derived in this way and are intended for use on condensed-phase systems.

Conformational Results

Conformational energy surfaces were computed for butane, methyl ethyl ether, and two dipeptides. These calculations indicated that for AMBER/OPLS acceptable choices for SCEE and SCNB are 2.0 and 8.0, i.e., the same as for "big" AMBER.[3] All results for AMBER/OPLS reported here use these values.

For butane, the energies of the gauche and cis conformers relative to trans are listed in Table VI. The AMBER/OPLS and normal AMBER results are similar; the gauche − trans energy difference is on the high side of the range of experimental values[27-30] and of the best available ab initio result.[26]

Table VII. Relative Energies for Conformations of Methyl Ethyl Ether[a]

method	gauche	cis	ref
AMBER/OPLS	1.5	8.7	this work
AMBER—normal	1.6	9.4	this work
AMBER—big	1.6	8.9	this work
AMBER—all atom	1.4	5.3	24
MM2	1.8	4.5	31
4-31G	2.0	7.3	32
IR, gas phase	1.5		33
ED, gas phase	1.2		34

[a] Energies relative to the trans conformer in kcal/mol.

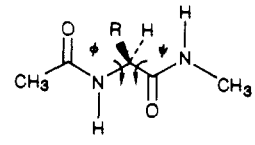
Table VIII. Relative Energies and Torsional Angles (Φ, Ψ) for Conformations of *N*-Acetylglycine *N*-Methylamide

method	C_7	C_5[b]	α	ref
AMBER/OPLS	0.0 (82, −67)	1.7		this work
AMBER—normal	0.0 (77, −64)	3.2	4.1 (66, 35)	3
AMBER—all atom	0.0 (75, −65)	3.3	4.1 (60, 39)	3
UNICEPP	0.0 (83, −76)	0.9	1.2 (71, 52)	35
ECEPP/2	0.0 (79, −73)	1.2	1.2 (73, 74)	36
4-21G	0.0 (83, −71)	0.8		37
PCILO	0.0 (80, −40)	2.0		38
IR, NMR (CCl_4)	(75, −50)			39
X-ray (crystal)	(109, −21)			40

[a] Energies in kcal/mol, angles (Φ and Ψ) in deg. [b] The C_5 conformation has $\Phi = \Psi = 180°$.

The corresponding results for methyl ethyl ether are summarized in Table VII. The AMBER/OPLS and normal AMBER results are again similar; the predicted gauche − trans energy differences are also close to the experimental findings.[33,34]

The two standard dipeptides that were studied are *N*-acetylglycine *N*-methylamide (GA) and *N*-acetylalanine *N*-methylamide (AA). Rough energy maps were constructed by varying Φ and ψ in 30° intervals between −180° and 180°. The local energy



GA: R=H
AA: R=CH_3

(23) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335.
(24) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(25) (a) Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 5757. (b) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(26) Ragavachari, K. *J. Chem. Phys.* **1984**, *81*, 1383.
(27) Compton, D. A. C.; Montero, S.; Murphy, W. F. *J. Phys. Chem.* **1980**, *84*, 3587.
(28) Verma, A.; Murphy, W.; Bernstein, H. *J. Chem. Phys.* **1974**, *60*, 1540.
(29) Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1959**, *32*, 748.
(30) Kanesaka, I.; Snyder, R. G.; Strauss, H. L. *J. Chem. Phys.* **1986**, *84*, 395.
(31) Burkert, U. *J. Comput. Chem.* **1980**, *1*, 285.
(32) Jorgensen, W. L.; Ibrahim, M. *J. Am. Chem. Soc.* **1981**, *103*, 3976.
(33) Kitagawa, T.; Miyazawa, T. *Bull. Chem. Soc. Jpn.* **1968**, *41*, 1976.
(34) Oyanagi, K.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 2237.
(35) Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2609.
(36) Vazquez, M.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.
(37) (a) Schafer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys.* **1982**, *76*, 1439. (b) Klimkowski, V. J.; Schafer, L.; Momany, F. A.; Van Alsenoy, C. *J. Molec. Str. (THEOCHEM)* **1985**, *124*, 143.
(38) Maigret, B.; Pullman, B.; Dreyfus, J. *J. Theor. Biol.* **1970**, *26*, 231.
(39) Cung, M. T.; Marraud, M.; Neel, J. *Ann. Chim. (Paris)* **1972**, *7*, 183.
(40) Iwasaki, F. *Acta Crystallogr., Sect. B* **1974**, *B30*, 2503.
(41) Schafer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; Van Alsenoy, C. *Biopolymers* **1984**, *23*, 2335. Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schafer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438.
(42) Hossain, M. B.; van der Helm, D. *J. Am. Chem. Soc.* **1978**, *100*, 5191.
(43) Karle, I. L.; Gibson, J. W.; Karle, J. *J. Am. Chem. Soc.* **1970**, *92*, 3755.
(44) Kostansek, E. C.; Thiessen, W. E.; Schomburg, D.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1979**, *101*, 5811.
(45) Karle, I. L. *J. Am. Chem. Soc.* **1978**, *100*, 1286.

Table IX. Relative Energies and Torsional Angles (Φ, Ψ) for Conformations of *N*-Acetylalanine *N*-Methylamide

method	C_7^{eq}	C_5	α_R	α_L	C_7^{ax}	ref
AMBER/OPLS	0.0 (−84, 70)	1.5 (−150, 162)			2.5 (67, −56)	this work
AMBER—normal	0.0 (−79, 69)	2.3 (−150, 154)	3.0 (−69, −29)	4.6 (55, 35)	0.8 (68, −58)	3
AMBER—all atom	0.0 (−76, 66)	3.2 (−161, 169)	3.6 (−61, −41)	4.3 (54, 42)	0.6 (69, −64)	3
UNICEPP	0.0 (−83, 81)	0.7 (−152, 147)	1.2 (−72, −44)	3.5 (55, 57)		35
ECEPP/2	0.0 (−80, 76)	0.7 (−155, 157)	0.8 (−74, −35)	2.3 (54, 46)	7.3 (76, −65)	36
4-21G	0.0 (−85, 73)	1.4 (−166, 167)	6.0 (−78, −26)	6.7 (61, 41)	2.6 (75, −62)	41
PCILO	0.3 (−78, 40)	1.7 (−171, 164)	2.4 (−29, −59)		0.0 (75, −40)	38
IR, NMR (CCl_4)	(−75, 50)	(−160, 170)				39

[a] Energies in kcal/mol, angles (Φ, Ψ) in deg.

Table X. Experimental Data on Cyclic Peptide Crystals

abbrev	peptide	no. of water[a]	space group	Z^b	ref
CP1	cyclo-(Ala-Ala-Gly-Gly-Ala-Gly)	1	$P2_1$	2	42
CP2	cyclo-(Ala-Ala-Gly-Ala-Gly-Gly)	2	$P2_12_12_1$	4	42
CP3	cyclo-(Gly-Gly-D-Ala-D-Ala-Gly-Gly)	3	$P2_12_12_1$	4	43
CP4	cyclo-(Gly-Pro-Gly-Gly-Pro-Gly)	4	$P2_1$	2	44
CP5	cyclo-(Gly-Pro-Gly-D-Ala-Pro)	0	$P2_12_12_1$	4	45

[a] Number of water molecules per peptide in the crystal. [b] Number of peptide molecules in the unit cell.

minima were then located in unconstrained optimizations starting from conformations in the low energy regions. The results are summarized in Tables VIII and IX where the relative energies of the C_7 (1–7 H-bonded), C_5 (extended), and α-helical forms are reported along with the Φ and ψ values for the minima.

For GA, there is agreement that the C_7 conformer is lowest in energy with Φ and ψ near 80° and −70°. There is scatter in the predicted energies for the C_5 form with the AMBER/OPLS value in the middle of the range. The α-helical conformer is at still higher energy with AMBER. A minimum could not be found in this case with AMBER/OPLS; all attempts at optimization collapsed to the C_7 conformer.

The reduced symmetry in AA leads to more possibilities for distinct energy minima (Table IX). There is now general accord that the equatorial C_7 form is lowest in energy with Φ and ψ near −80° and 70°. The energy for the extended C_5 conformation from AMBER/OPLS is in the middle of the tabulated range. The α helical conformers are again not found as energy minima by several computational methods including AMBER/OPLS. Their existence as minima was previously found to be sensitive to the scaling of the 1,4-interactions in AMBER.[3] In the absence of more definitive experimental data, the main conclusion from these comparisons is that the AMBER/OPLS predictions for conformational energies are reasonable.

Polypeptide Crystals

The structures for the five cyclic polypeptide crystals listed in Table X were also calculated with the normal AMBER and AMBER/OPLS force fields to obtain a stringent test of the representation of the intermolecular interactions. Similar computations were the basis of the recent evaluation of force fields by Hall and Pavitt that proved very favorable for AMBER.[4] They used the first three polypeptides in Table X.

Version 2.0 of AMBER did not include the code necessary for the crystal calculations. Thus, additions were made to allow energy minimizations for a realistic representation of the crystalline environment. In our modified version of the minimization procedure, the intramolecular interactions are calculated in the standard fashion over all atoms in one asymmetric unit. For evaluating the intermolecular interactions, the unit cell is first completed by generating coordinates for all remaining atoms. This entails reflection and/or translation of the original asymmetric unit. The full crystalline environment is then provided by periodic boundary conditions using translated images of the unit cell in all directions, i.e., the unit cell is effectively surrounded by 26 images of itself. These spatial transformations were made relative to the current dimensions of the unit cell at each cycle of the minimization. Two types of calculations were performed. In one, the unit cell dimensions were fixed at the experimental values, while in the other, they were optimized with use of the simplex method.[46] The latter calculations were relatively time-consuming since complete energy minimization for the contents of the asymmetric unit was performed with the AMBER program between each simplex cycle.

The energy and forces were calculated for an asymmetric unit by including the interactions with all image atoms within the cutoff range. Specifically, a residue based cutoff was used such that if any atoms of two residues were within 11 Å, the interactions between all atoms in the two residues were included. All of the crystal calculations employed a dielectric constant of 1 for the electrostatic interactions and the same 1,4-scale factors as in the conformational studies. The water molecules in the crystals were represented by the TIP3P model.

The procedure described above limits the calculations to crystals with monoclinic or orthorhombic units cells. In particular, the program was set up to handle the symmetry operations for the following space groups: P_1, $P_{\bar{1}}$, $P2$, $P2_1/C$, $C2/C$, and $P2_12_12_1$. Besides conforming to this restriction, the chosen test cases in Table X have several other desirable features. The crystal structures are known to high precision; the average uncertainties in the positions of the non-hydrogen atoms are less than ca. 0.02 Å. The crystals also do not contain any solvent molecules or residues for which parameters are unavailable in either force field. And, the sizes of the systems are computationally manageable. The disadvantage is that only Gly, Ala, and Pro residues are represented, so the protein crambin has also been examined as described in the next section.

The unit cells for the peptide crystals, which have been designated CP1 to CP5 in Table X, are illustrated in Figures 5–9. CP1 has two peptide and two water molecules in the unit cell. The peptides feature two 4 → 1 intramolecular hydrogen bonds corresponding to β turns of types I and II′.[42] The unit cell for CP2 is larger with four peptide and eight water molecules. The two 4 → 1 β turns in this case are of types I and the relatively rare I′.[42] Both crystals show extensive hydrogen bonding networks with each NH and C=O group participating in the β turns, interpeptide hydrogen bonds, or hydrogen bonds with water. CP3 also has four peptide molecules in the unit cell, but now there are twelve water molecules. Two 4 → 1 β turns are again present along with hydrogen bonds for each NH and C=O group.[43] The water molecules in CP1–CP3 each participate in three hydrogen bonds.

CP4 and CP5 contain the Pro residues that all have trans peptide bonds. CP4 has two protein and eight water molecules in the unit cell, while CP5 has four peptide molecules and is anhydrous. A variety of intrapeptide hydrogen bonds are present; CP4 has 4 → 1 β(I) and β(II) turns, whereas the pentapeptide CP5 has a β(II) turn with the clear transannular hydrogen bond in Figure 9 and a 3 → 1 γ bend involving the NH of Gly-1 and

(46) Nedler, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308.

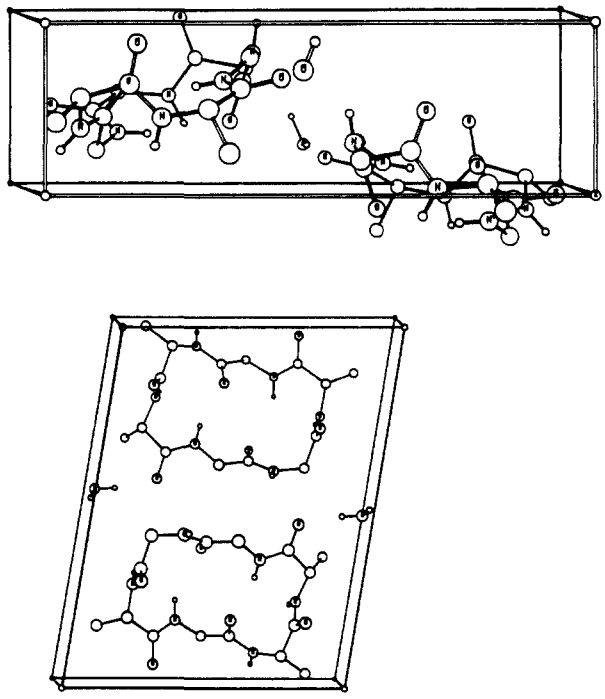


Figure 5. Experimental unit cell for cyclo-(Ala-Ala-Gly-Gly-Ala-Gly), CP1.

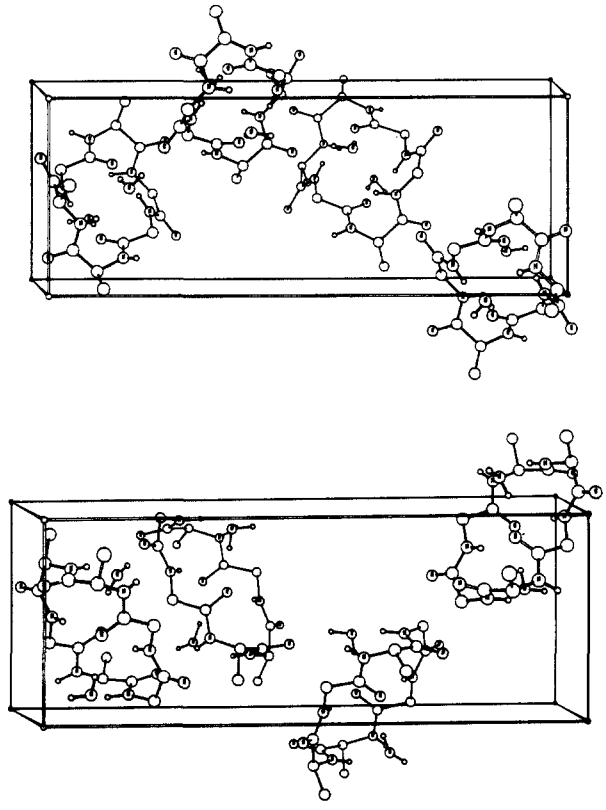


Figure 6. Experimental unit cell for cyclo-(Ala-Ala-Gly-Ala-Gly-Gly), CP2.

the CO of Ala-4 that straddle the Pro-5 side chain. Overall, a substantial structural range is represented by these five cyclic peptide crystals with a great variety of hydrogen bonding and crystal packing.

Some key results from the energy minimizations are reported in Table XI, in particular, the root-mean-square deviations between the computed and experimental positions for the atoms in the crystals. The deviation is the average for all atoms except for CP3 in which case only the non-hydrogen atoms are considered. The positions of 21 of the 26 hydrogens for CP3 were only established

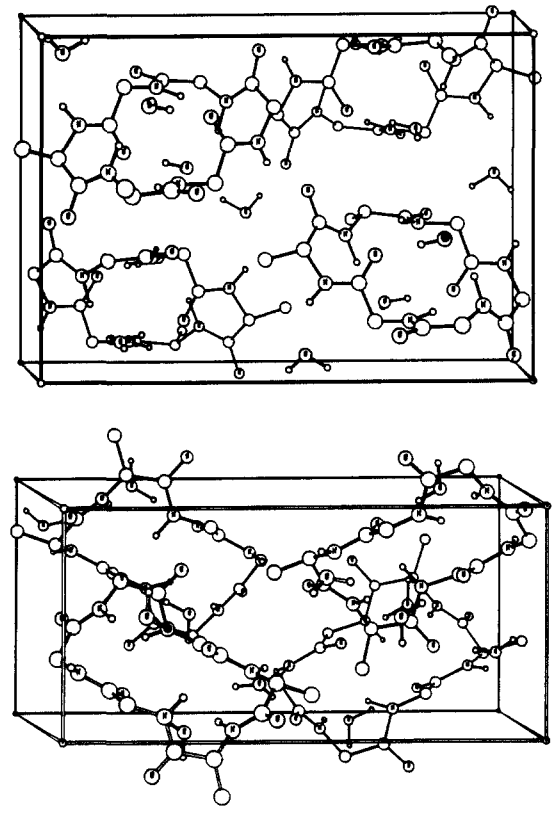


Figure 7. Experimental unit cell for cyclo-(Gly-Gly-D-Ala-D-Ala-Gly-Gly), CP3.

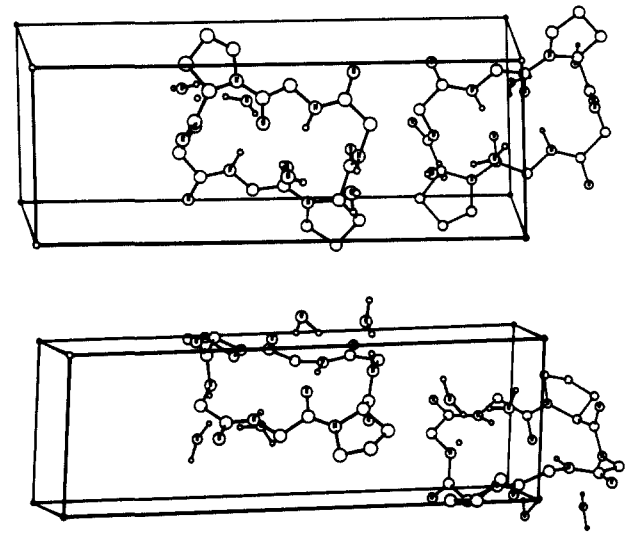


Figure 8. Experimental unit cell for cyclo-(Gly-Pro-Gly-Gly-Pro-Gly), CP4.

approximately in the X-ray study.[43] The unit cell dimensions were also not optimized for CP3 which is structurally similar to CP2. The change in the cell volume upon optimization of the lattice dimensions is given by V_{rel}, which is the ratio of the computed to experimental volumes.

The root-mean-square deviations are improved with the AMBER/OPLS potentials. With use of the experimental cell dimensions, the deviations average 0.16 Å for AMBER and 0.14 Å for AMBER/OPLS, while they are 0.22 and 0.17 Å when the cell dimensions are optimized. The optimization causes greater compaction with the normal AMBER parameters. Less compaction was obtained for the one entry in Table XI that used the "big" AMBER CH_n parameters. However, the root mean square error for CP2 was actually worsened slightly by this modification from 0.23 to 0.26 versus 0.14 for AMBER/OPLS. Some compaction is expected since the energy minimization procedure

Table XI. Computed Results for Cyclic Peptide Crystals

peptide	cell[a]	AMBER—normal E^b	AMBER—normal rms[c]	AMBER—normal $V_{rel}{}^d$	AMBER/OPLS E^b	AMBER/OPLS rms[c]	AMBER/OPLS $V_{rel}{}^d$
CP1	exptl.	-219.5	0.11	(1.00)	-282.0	0.09	(1.00)
CP1	min.	-222.0	0.11	0.94	-283.0	0.09	0.97
CP2	exptl.	-213.0	0.22	(1.00)	-260.4	0.14	(1.00)
CP2	min.	-213.8	0.23	0.90	-260.7	0.14	1.00
CP2[e]	min.	-221.0	0.26	0.98			
CP3	exptl.	-235.4	0.07[f]	(1.00)	-283.7	0.09[f]	(1.00)
CP4	exptl.	-212.2	0.31	(1.00)	-257.7	0.23	(1.00)
CP4	min.	-224.0	0.36	0.90	-259.3	0.25	0.96
CP5	exptl.	-121.1	0.08	(1.00)	-172.2	0.16	(1.00)
CP5	min.	-128.3	0.17	0.83	-173.8	0.18	0.95
av	exptl.		0.16	(1.00)		0.14	(1.00)
av	min.		0.22	0.89		0.17	0.97

[a] Exptl.: Unit cell dimensions fixed at experimental values. Min: cell dimensions determined by energy minimization. [b] Total energy in kcal/mol for one asymmetric unit after energy minimization. [c] Root-mean-square deviation between computed and experimental positions for all atoms in an asymmetric unit in Å. [d] Ratio of calculated to experimental volume of the unit cell. [e] Calculated using the AMBER—"big" non-bonded parameters. [f] Hydrogen atoms not included in calculated RMS; their experimental positions were not all reported in ref 43.

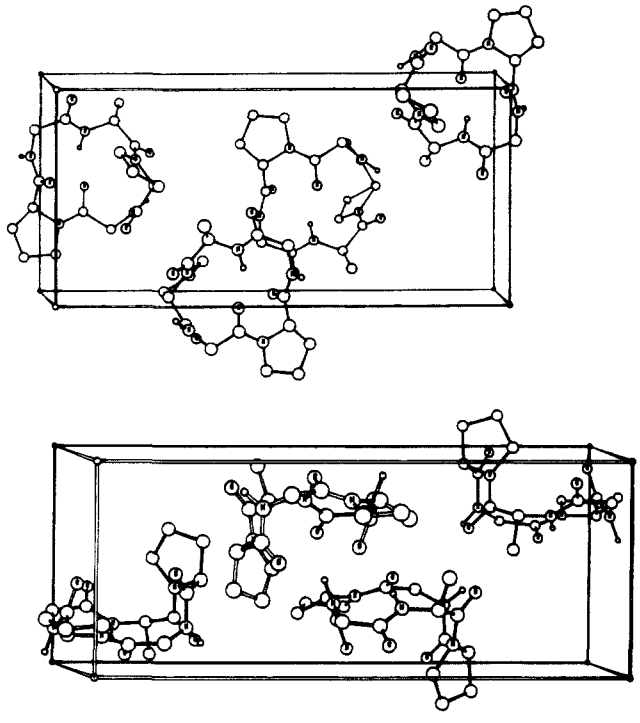


Figure 9. Experimental unit cell for cyclo-(Gly-Pro-Gly-D-Ala-Pro), CP5.

corresponds to $T = 0$ K, whereas the crystal structures appear to have been determined between −135 °C and room temperature.[42-45] In view of the performance for liquid densities (Figure 1), the OPLS potentials may be anticipated to make reasonable predictions for the compaction. Thus, the 10% and 17% volume reductions predicted by AMBER for CP4 and CP5 are probably excessive and may be associated with the more elaborated hydrocarbon side chains in these cases.

The root-mean-square deviations are decomposed into separate values for the peptide hydrogens, other protein atoms, and the water oxygens in Table XII. This shows that the errors are largest for the peptide hydrogens whose positions have the greatest experimental uncertainty. The average root-mean-square deviations for the C, N, and O atoms of the peptides are only 0.10–0.11 Å for the AMBER/OPLS force field and 0.13–0.17 Å for AMBER. It is also apparent that there is much improvement in the placement of the water molecules when AMBER/OPLS is used. Some further assessment can be made by comparing the predicted and experimental values for the N···O separations in the intrapeptide hydrogen bonds for CP1–CP5. As shown in Table XIII, the predictions from both force fields are in good accord with the experimental data; the average errors are 0.08 and 0.09 Å for AMBER/OPLS and AMBER, respectively, though the largest deviation is much greater for AMBER (0.36 Å) than AMBER/OPLS (0.18Å).

Table XII. Computed Root-Mean-Square Deviations for Atomic Positions in the Peptide Crystals[a]

force field	cell[b]	peptide H	peptide C, O, N	water O
AMBER	exptl.	0.25	0.13	0.27
AMBER/OPLS	exptl.	0.26	0.10	0.18
AMBER	min.	0.29	0.17	0.38
AMBER/OPLS	min.	0.28	0.11	0.17

[a] rms deviations in Å. Results for the five peptide crystals using the normal AMBER or AMBER/OPLS parameters. [b] See footnote *a*, Table XI.

Table XIII. Intramolecular Hydrogen Bond Lengths[a]

peptide	residue C=O	residue N—H	calculated AMBER	calculated AMBER/OPLS	exptl
CP1	Gly-6	Gly-3	3.02	2.98	2.92
CP1	Gly-3	Gly-6	3.71	3.49	3.35
CP2	Gly-6	Gly-3	2.89	3.10	3.10
CP2	Gly-3	Gly-6	2.85	2.88	2.99
CP3	Gly-5	Gly-2	3.14	3.22	3.04
CP3	Gly-2	Gly-5	3.13	3.13	3.16
CP4	Gly-4	Gly-1	2.91	2.95	2.91
CP5	Gly-1	D-Ala-4	2.87	2.94	2.87
CP5	D-Ala-4	Gly-1	2.86	2.81	2.92

[a] N···O separations in Å. Experimental unit cell dimensions used.

Overall, these results show that the description of the peptide crystals is improved by use of the OPLS nonbonded parameters. The reduction in the root-mean-square errors by 30–110% in Table XII is particularly notable since AMBER had previously been established as probably the best available force field for proteins,[4] and because the AMBER/OPLS force field is simpler than AMBER. Specifically, there are no lone pairs and the 10–12 function for describing the hydrogen bonds has been eliminated. The improvement with AMBER/OPLS for the crystal structures is reasonable owing to the greater emphasis placed on condensed-phase data for development of the OPLS parameters. Nevertheless, further testing on a more complex system including polar residues is desirable.

Crambin

Crambin is a small plant seed protein with 46 residues and 326 non-hydrogen atoms. It has fifteen different residues whose occurrence is given in parentheses: Cys (6), Thr (6), Ala (5), Ile (5), Gly (4), Pro (4), Asn (3), Ser (3), Arg (2), Val (2), Tyr (2), Asp (1), Glu (1), Leu (1), and Phe (1). The secondary structure features two regions of α-helix and a small antiparallel β-sheet. The crystal's structure has been resolved to high resolution by Hendrickson and Teeter.[47] For the present calculations, their

(47) Hendrickson, W. A.; Teeter, M. M. *Nature (London)* **1981**, *290*, 107.

Table XIV. Results of Energy Minimizations for the Crambin Crystal

property	force field	
	AMBER	AMBER/OPLS
final energy, kcal/mol	−6516.1	−9015.1
rms—protein, Å	0.22	0.17
rms—backbone, Å	0.19	0.14
rms—side chains, Å	0.25	0.20
rms—Φ, deg	7.2	6.1
rms—ψ, deg	7.9	5.6
rms—ω, deg	4.1	4.6
rms—χ, deg	10.9	11.5

structure refined to 0.945 Å[48] has been used and corresponds to the Pro_{22}–Ile_{25} variant.[48,49] This same structure was employed by Whitlow and Teeter in a comprehensive study of energy minimizations mostly for an isolated crambin molecule.[50] They found improved accord with the X-ray structure when the electrostatic interactions were highly damped and when either a 5 Å shell of water was placed around the protein or the interprotein hydrogen bonds in the crystal were included.[50] For the present purposes, the minimizations have been carried out for the entire unit cell with its periodic images just as for the peptide crystals. The unit cell with space group *P*21 contains 2 protein molecules, four ethanols near the protein surface,[51] and an estimated 182 water molecules based on the density and cell dimensions (a = 40.96 Å, b = 18.65 Å, c = 22.52 Å, and β = 90.77°).[48] Many of the water molecules have been located experimentally;[47,51] however, starting coordinates were needed for all atoms in the unit cell for the present energy minimizations. They were obtained via a Monte Carlo (MC) calculation as described next.

Our standard statistical mechanics program was modified to handle the non-orthogonal unit cell and its contents. The positions of all C, N, O, and S atoms of the protein and ethanols were taken from the X-ray data. The hydrogens on the heteroatoms were added with use of standard geometrical parameters and the asymmetric unit was then repeated according to the symmetry operation for the P21 space group. The unit cell was next filled with water molecules taken from a simulation of pure TIP3P water.[11] Any water molecules with oxygens within ca. 2.15 Å of a non-hydrogen atom of the protein were removed which left the desired 182 water molecules. It was verified during the initial stages of the MC simulation and by graphical display that none of the remaining water molecules were in improbable locations, i.e., trapped inside the protein.

The Monte Carlo simulation was carried out in the NVT ensemble at 25 °C with Metropolis sampling and periodic boundary conditions. Only the water molecules were allowed to move. The OPLS potentials were used, including the TIP3P model for the water molecules. Interactions were included between a water molecule and the nearest images of all residues from the unit cell that had any atom within 8.5 Å of the water's oxygen. The system was well equilibrated after 10^6 configurations. The last configuration was then used as the starting point for the energy minimizations with the modified AMBER program. The MC run was also extended as the basis for an analysis of the hydration of crambin.[52]

The energy minimizations were carried out with the normal AMBER and AMBER/OPLS force fields. Since the unit cell contained 1356 explicit atoms (Figure 10), several months were required on the Microvax II to reach an acceptable level of convergence. After ca. 1100 cycles of energy minimization, the root-mean-square gradient averaged over all atoms reached

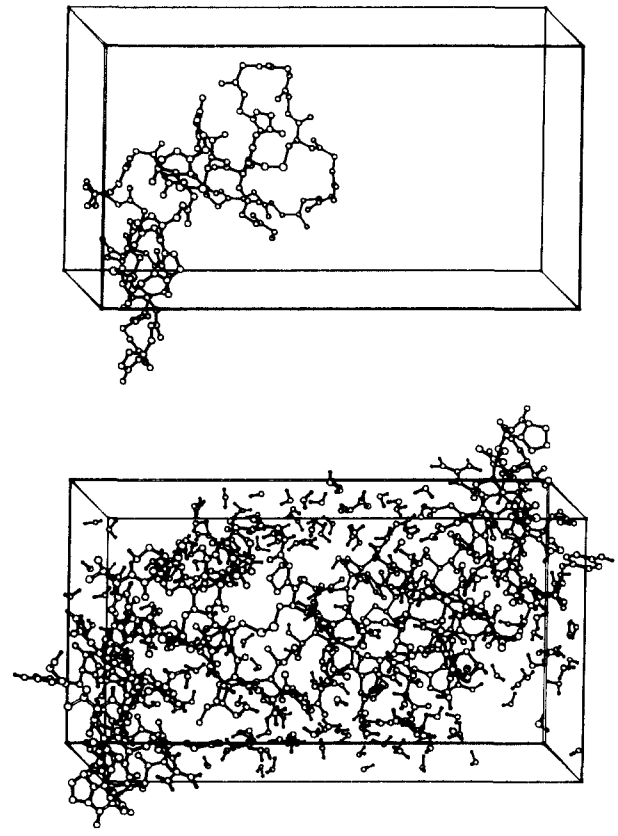


Figure 10. Top: The backbone of a crambin molecule in the unit cell. Bottom: The complete unit cell with 2 protein, 4 ethanol, and 182 water molecules. The seemingly vacant regions in the corners are filled in by atoms from the adjacent unit cells.

0.79–0.80 kcal/(mol·Å) for both calculations and the change in total energy between cycles was 0.05–0.06 kcal/mol. Additional computation did not lead to systematic reduction of the gradient, though in the calculations on the peptide crystals and on an isolated crambin molecule convergence of the gradient to less than 0.1 kcal/(mol·Å) was achieved in each case. The difference in behavior is likely due to the large amount of water in the crambin crystal which makes the potential energy surface relatively "soft", i.e., without sharp minima.

The results from the two energy minimizations are summarized in Table XIV. First, it may be noted that the total energy is ca. 40% lower with the OPLS parameters. This is primarily an artefact of the scaling of the 1,4-nonbonded interactions with a factor (SCNB) of 2 for AMBER and 8 for AMBER/OPLS, as discussed above. The key result is that the average root-mean-square deviations for the protein and ethanol C, N, O, and S atoms are 0.22 Å from AMBER and 0.17 Å from AMBER/OPLS. As shown in Table XIV, the root-mean-square deviations are also lower with AMBER/OPLS for the backbone and side-chain atoms, when they are considered separately, as well as for the main chain dihedral angles, Φ and ψ. The largest single deviations in a Φ or ψ from the X-ray data are 28.4° for AMBER and 24.6° for AMBER/OPLS. The average deviations for the side chain dihedral angles are greater at 10.9° for AMBER and 11.5° for AMBER/OPLS. However, the largest single deviations of 28.2° and 24.7°, respectively, indicate that there were no substantial conformational changes during the minimizations.

The root-mean-square deviations for the atomic positions were also decomposed by residue type in order to ascertain if either force field has particular difficulties with any specific residues. It turns out that the variation is slight with AMBER/OPLS; the smallest errors are 0.13–0.14 Å for Gly, Ala, Leu, Cys, Arg, Glu, and Phe, while the largest errors are 0.19–0.20 Å for Thr, Tyr, and Asn. The variation for AMBER is greater with Ala, Ile, Cys, and Phe having the smallest errors of 0.14–0.17, while errors of 0.24–0.29 occur for Pro, Ser, Thr, Asp, and Asn. The consistency

(48) Teeter, M. M.; Hendrickson, W. A., unpublished results.
(49) Vermeulen, J. A. W. H.; Lamerichs, R. M. J. N.; Berliner, L. J.; De Marco, A.; Llinas, M.; Boelens, R.; Alleman, J.; Kaptein, R. *FEBS Lett.* **1987**, *219*, 426.
(50) Whitlow, M.; Teeter, M. M. *J. Am. Chem. Soc.* **1986**, *108*, 7163.
(51) Teeter, M. M.; Whitlow, M. D. *Trans. Am. Crystallogr. Assoc.* **1986**, *22*, 75. Teeter, M. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014.
(52) Jorgensen, W. L.; Tirado-Rives, J.; Teeter, M. M., to be published.

of the AMBER/OPLS results and the small errors for the charged residues are particularly reassuring.

Overall, these results provide important support for the quality of both force fields, though some increases in the root-mean-square deviations may be anticipated with more complete convergence. In fact, the minimization with AMBER/OPLS was extended to 1978 cycles at which point the gradient was 0.495 and the root-mean-square deviation for the protein atoms had only increased to 0.22 Å. Further testing of both force fields in molecular dynamics simulations for proteins in water is planned.

Conclusion

A wide range of computations on organic liquids, dilute aqueous solutions, hydrogen bonding, and ion-water complexes has provided a set of functions to describe the nonbonded interactions for proteins in crystals or aqueous solutions. In all, parameters have been reported for 25 amino acid residues and various terminal groups. The basis for the potential functions is comparatively sound and their form is attractively simple. Combination with bond stretch, angle bend, and torsional terms from the AMBER force field provides a complete model for proteins that is now available for application to innumerable problems addressing the structure, dynamics, and function of biomolecular systems. Initial tests on crystals of cyclic peptides and the protein crambin have demonstrated that the AMBER/OPLS force field yields root-mean-square errors of only 0.1-0.2 Å for the positions of non-hydrogen atoms.

Acknowledgment. Gratitude is expressed to the National Institutes of Health (GM32136) for support of this research, to Jiali Gao for computational assistance, and to Professor Martha M. Teeter for unpublished data and consultation.

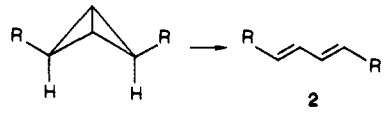
A Theoretical Investigation of the Thermal Ring Opening of Bicyclobutane to Butadiene. Evidence for a Nonsynchronous Process†

Philip B. Shevlin* and Michael L. McKee*

Contribution from the Department of Chemistry, Auburn University, Auburn, Alabama 36849.
Received July 14, 1987

Abstract: An ab initio investigation of the ring opening of bicyclobutane (**1a**) to butadiene (**2a**) with geometries optimized at the HF/3-21G level predicts a two-step process with an intermediate 3-butenylidene (**5**). Since this prediction is in conflict with experimental results, the reaction surface was reinvestigated with geometries optimized at the MP2/3-21G level. This procedure leads to a transition state for the $\sigma 2_s + \sigma 2_a$ ring opening which lies 43.6 kcal/mol above bicyclobutane ([MP4SDTQ/6-31G*]//MP2/3-21G + ZPC//3-21G) in good agreement with the known barrier for the thermal ring opening of **1a** to **2a** of 40.6 kcal/mol. The geometry of the MP2/3-21G transition state, with one C-C bond lengthened by 0.783 Å and the other increasing by only 0.088 Å, indicates a nonsynchronous reaction. The disrotatory ring opening of **1a** (based on a C_2 transition state) has a predicted barrier of 97.0 kcal/mol at the [MP4SDQ/6-31G*] + ZPC//3-21G level of theory.

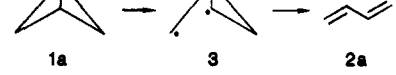
The bicyclo[1.1.0]butane ring system, with its strain energy of over 60 kcal/mol,[1] has been the subject of numerous theoretical and experimental investigations. Of particular interest have been studies of the ring opening to 1,3-butadienes. When the ring opening of bicyclo[1.1.0]butane (**1a**) to 1,3-butadiene (**2a**) is carried out thermally, the central bond remains intact while two opposite peripheral C-C bonds are broken in going to the product.[2-4] Studies of 1,3-dimethylbicyclobutanes have demonstrated that the reaction is highly stereoselective with exo substituents at C_2 and C_4 becoming cis,trans in the product butadiene in what has been characterized as a stereoselective $\sigma 2_s + \sigma 2_a$ reaction (eq 1).[3] Mechanistic interpretations of this ring opening have been



1a, R = H
b, R = CH_3 **2** (1)

somewhat problematic.[5-10] A disrotatory process is predicted to be forbidden from an analysis of a correlation diagram where orbitals can be assigned with respect to a C_2 axis that is maintained during the reaction. However, the conrotatory ring opening, which is often interpreted as the allowed $[\sigma 2_s + \sigma 2_a]$ reaction, cannot be followed by any element of symmetry higher than C_1. Assuming that the reaction can be followed by an approximate correlation diagram may lead to erroneous conclusions.[6]

Dewar[8] has argued that most reactions involving the breaking or forming of two bonds are not synchronous and many are not concerted. MINDO/3 calculations[9] have found that the ring opening of bicyclobutane is a two-step reaction involving the intermediacy of the cyclopropylcarbinyl biradical, **3**. The stereochemistry of the reaction is maintained due to slow interconversion of the "biradicaloids" compared to further reaction to form



1a **3** **2a**

† Dedicated to Professor Michael J. S. Dewar on the occasion of his 70th birthday.

(1) Wiberg, K. B. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 312-322.
(2) Blanchard, E. P., Jr.; Cairncross, A. *J. Am. Chem. Soc.* **1966**, *88*, 487-495.
(3) Closs, G. L.; Pfeffer, P. E. *J. Am. Chem. Soc.* **1968**, *90*, 2452-2453.
(4) Wiberg, K. B.; Lavanish, J. M. *J. Am. Chem. Soc.* **1966**, *88*, 5272-5275.
(5) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Verlag Chemie, Academic Press: New York, 1970.
(6) Mulder, J. J. C. *J. Am. Chem. Soc.* **1977**, *99*, 5177-5178.
(7) Aihara, J. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 1788-1792.
(8) Dewar, M. J. S. *J. Am. Chem. Soc.* **1984**, *106*, 209-219.
(9) Dewar, M. J. S.; Kirschner, S. *J. Am. Chem. Soc.* **1975**, *97*, 2931-2932.
(10) For the application of the principle of least motion to the rearrangement of **1a** see: Altmann, J. A.; Tee, O. S.; Yates, K. *J. Am. Chem. Soc.* **1976**, *98*, 7132-7138.

0002-7863/88/1510-1666$01.50/0