

Validation of the General Purpose Tripos 5.2 Force Field

Matthew Clark,* Richard D. Cramer III, and Nicole Van Opdenbosch

Tripos Associates, 1699 S. Hanley Road, St. Louis, MO 63144

Received 22 November 1988; accepted 22 March 1989

A molecular mechanics force field implemented in the Sybyl program is described along with a statistical evaluation of its efficiency on a variety of compounds by analysis of internal coordinates and thermodynamic barriers. The goal of the force field is to provide good quality geometries and relative energies for a large variety of organic molecules by energy minimization. Performance in protein modeling was tested by minimizations starting from crystallographic coordinates for three cyclic hexapeptides in the crystal lattice with rms movements of 0.019 angstroms, 2.06 degrees, and 6.82 degrees for bond lengths, angles, and torsions, respectively, and an rms movement of 0.16 angstroms for heavy atoms. Isolated crambin was also analyzed with rms movements of 0.025 angstroms, 2.97 degrees, and 13.0 degrees for bond lengths, angles, and torsions respectively, and an rms movement of 0.42 angstroms for heavy atoms. Accuracy in calculating thermodynamic barriers was tested for 17 energy differences between conformers, 12 stereoisomers, and 15 torsional barriers. The rms errors were 0.8, 1.7, and 1.13 kcal/mol, respectively, for the three tests. Performance in general purpose applications was assessed by minimizing 76 diverse complex organic crystal structures, with and without randomization by coordinate truncation, with rms movements of 0.025 angstroms, 2.50 degrees, and 9.54 degrees for bond lengths, angles and torsions respectively, and an average rms movement of 0.192 angstroms for heavy atoms.

INTRODUCTION

Success in computer-aided molecular design requires accuracy in calculating geometries and relative energies of a wide variety of compounds. While good geometries can be obtained by rule based methods,^{1,2} the most common method of molecular modeling is presently energy minimization using a molecular mechanics force field.

The scientific value of a force field is derived from both the variety and quality of the parametrization. Early work on parametrizing force fields for hydrocarbons led to spectacular successes in agreements between calculated and observed structures; however, this ability was limited to a very narrow range of structures.³

In response to the need for modeling in specific areas, many parameter sets and functional forms have been developed to accurately describe narrow classes of compounds. MM2 has become the standard for modeling small monofunctional molecules,⁴ while Amber/OPLS,⁵ Amber,⁶ ECEPP,⁷

CHARMM,⁸ and CFF/VFF⁹ are prominent in studies of biopolymers, particularly proteins. There is still the need, however, for a general force field, which although it may not equal the best force fields for specific chemical classes, will give uniform quality results for an extremely wide variety of compounds. Examples of force fields for drug design by modeling small-molecule protein interactions are the YETI force field,¹⁰ and the Tripos 5.2 force field of the Sybyl¹¹ molecular modeling package.

A systematic problem with the molecular mechanics method is an unbiased evaluation of the quality of the results for comparisons among force fields. While most force fields are accompanied by a few examples of their quality, the literature is lacking in methods for general statistical comparison of molecular structures. While the rms deviation of atom positions can be a useful measurement of distortion of overall shape, it can be misleading in the case of a small movement at the base of a chain causing large movements at the end of the chain due to lever action.

The choice of experimental structures for evaluation of force fields is problematic. The

*To whom all correspondence should be addressed.

most obvious source is from X-ray crystallography. A disadvantage of X-ray structures is that many structural details are products of the interactions of the molecule with its neighbors, which usually do not exist in the environment of interest. A second disadvantage is that there is latitude in the quality of crystal structures, even at low R factors.¹² A third is that if the structure is not corrected for rigid body motions the bond lengths can be too short by 0.005 to 0.015 angstroms. The molecular mechanics minimization of single molecule from a crystal structure should therefore not be expected to reproduce the structure exactly. This is especially true in considering the conformations of side chains.

In the following study, the model geometries produced by the Tripos force field are assessed by minimizing crystallographic structures: crambin, three cyclic hexapeptides in the crystal lattice, and 76 diverse organic structures. The protein results are compared with published work by Whitlow and Teeter,¹³ by Hall and Pavitt,¹⁴ and by Jorgensen and Tirado-Rives,⁵ using the two specialized force fields, Amber and the very recently introduced Amber/OPLS.⁵ Second, the relative energies produced by the Tripos force fields are assessed by calculation of 29 conformational or stereochemical energy differences and 15 torsional barriers. These energy differences are compared with experimental values and also with those reported by Burkert and Allinger⁴ using the MM2 force field, and other MM2 calculations.

In this work the authors propose a measure of force field accuracy by statistical deviation of internal coordinates. An evaluation of the motions of the bond distances, angles, and torsions has two advantages. First, it alleviates the problem of slight conformational changes having a large weighting in the rms analysis due to lever effects. Second, it allows an examination of the errors in internal coordinates, the same coordinates used in the force field terms.

METHODS

The Tripos force field terms and parameters used throughout this work, an adaptive evolution of those originally proposed by White,¹⁶ modified by Vinter,¹⁵ by Nicole

van Opdenbosch and Jan Labanowski, and also in this work, are shown in Tables I through VI.¹⁷

The force field equations, as given in Tables I through VI, consist of harmonic bond stretching and angle bending terms. The force field uses an out-of-plane bending terms that depends harmonically on the distance of the central atom from the plane defined by its three attached neighbors. The torsional function consists of a single cosine term. The bond stretching and torsional parameters used depend on the bond type to correctly represent both double and single bonds between sp^2 hybridized atoms. The nonbonded interaction terms are an electrostatic term with either a constant or distance dependent dielectric function, as well as a 6-12 Lennard-Jones potential. Since the electrostatic term is extremely large and negative when the interatomic distance is very small, at a cutoff of 0.5 angstrom a linear extrapolation is used to prevent the energy causing a numeric overflow. No explicit hydrogen bond terms are included, but hydrogens attached to atom types designated as hydrogen bond donors are given a radius of zero in the 6-12 nonbonded term for their interactions with atom types designated as a hydrogen bond acceptor.⁹

Some bond stretching parameters not present in the force field tables were automatically derived from the starting position as detailed in Tables II and III. This was necessary for 6 of the 1483 bonds in the sample: the C.3-Br bond of ABAXES (2alpha-bromo-17beta-acetoxy-9-methyl-5alpha,9beta,10alpha-estran-3-one), the S.o2—N.2 bond of ABZTCX (3-dimethylamino-4,4-dimethyl-5,6-dihydro-4H-1,2,5-benzothiazocin-6-one 1,1-dioxide), the N.am—S.o2 bond of ACRAMS (4'-(acridin-9-ylamine) methanesulfonanilide hydrochloride), the C.ar—N.1 bond of ADMOPM (adenosine-5'-O-methylphosphate methanol solvate), the N.ar—C.3 bond of AFCYDP (2,2'-anhydro-1-beta-D-arabinofuranosyl-cytosine-3',5'-diphosphate), and the S.o2—N.am and N.2—S.o2 bonds of AFUTDZ10 (7-amino-4H-furazano(3,4-*d*)-1,2,6-thiadiazine-1,1-dioxide). Default generic values of $k = 0.2 \text{ kcal/mol}^{-1} \text{ degree}^{-2}$ and $s = 3$ for the torsional constants were used for torsions involving rotations about the S.o2—N.2,

Table I. Tripos 5.2 force field atom types.

Symbol	Val. ^a	Geom. ^b	Hbd ^c	Hba ^d	LP ^e	Comment
Al	4	TH	NO	NO	0	
Br	4	TH	NO	NO	3	
C.1	2	L2	NO	NO	0	<i>sp</i> carbon
C.2	3	TG	NO	NO	0	<i>sp</i> ² carbon
C.3	4	TH	NO	NO	0	<i>sp</i> ³ carbon
C.ar	3	TG	NO	NO	0	aromatic
Ca	2	L2	NO	NO	0	
Cl	4	TH	NO	NO	3	
Du	0	f	NO	NO	0	dummy, ignored
F	4	TH	YES	YES	3	
H	1	L1	NO	NO	0	
I	4	TH	NO	NO	3	
K	1	L1	NO	NO	0	
Li	1	L1	NO	NO	0	
LP	1	L1	NO	NO	0	lone pair
N.1	2	L2	NO	YES	1	<i>sp</i> nitrogen
N.2	3	TG	YES	YES	1	<i>sp</i> ² nitrogen
N.3	4	TH	YES	YES	1	<i>sp</i> ³ nitrogen
N.4	4	TH	YES	NO	0	cationic
N.am	3	TG	YES	NO	0	amide nitrogen
N.ar	3	TG	NO	YES	1	aromatic
N.pl3	3	TG	YES	NO	0	<i>sp</i> ³ planar
Na	1	L1	NO	NO	0	
O.2	3	TG	NO	YES	2	<i>sp</i> ² oxygen
O.3	4	TH	YES	YES	2	<i>sp</i> ³ oxygen
P.3	4	TH	NO	NO	0	<i>sp</i> ³ phosphorus
S.2	3	TG	NO	NO	2	<i>sp</i> ² sulfur
S.3	4	TH	NO	NO	2	<i>sp</i> ³ sulfur
S.o	4	TH	NO	NO	1	sulfoxide
S.o2	4	TH	NO	NO	0	sulfone
Si	4	TH	NO	NO	0	<i>sp</i> ³ silicon

^avalence, including lone pairs.^bgeometry, TH, tetrahedral; TG, trigonal planar; L2, linear; L1, singly bonded; * no assumed geometry.^cYes if atom is a hydrogen bond donor, No if not.^dYes if atom is a hydrogen bond acceptor, No if not.^eNumber of lone pairs.^fNo assumed geometry.

S.o2—N.am and N.ar—C.3 atoms, as detailed in the formula of Table V. All other force field parameters were present.

The standard Sybyl energy minimizer, MAXIMIN2, was used under the following conditions. If the force on any atom exceeded 5 millidynes, an atom-by-atom Simplex minimization was performed until all forces fell below this threshold. After this threshold conjugate gradient minimization proceeded until the convergence criterion of an rms gradient over all atoms of less than 0.1 kcal/mol-angstrom was reached. Nonbonded interactions between substructures (residues) more than 8.0 angstroms apart were neglected. The nonbonded interaction list was reevaluated every 10 iterations.

Although the electrostatic potential can have the largest magnitude of all the force field terms, the authors have chosen to omit

it in this work. Despite the intuitive importance of electrostatics there is ambiguity in the calculation of the charges, the dielectric functions and the attenuation functions, or cutoffs. The choices for any of these may have impact on the results of the calculations, and the basis for such decisions is also unclear. In this work it was found that results were superior when the electrostatic term was not used.

The MM2 program as distributed by QCPE and Tripos Associates was used to calculate energy differences for comparisons.^{17,18}

The crambin structure used in this work was taken from the Brookhaven data base entry 1CRN.¹⁹ No solvent molecules were present. Hydrogen atoms were added to complete the valences of all atoms. The structure was then minimized as described above. Structures for the three cyclic hexapeptides,

Table II. Bond stretching parameters for the Tripos 5.2 force field.

Atom <i>i</i>	Atom <i>j</i>	Bond type ^a	$d_{i,j}^{0\ b}$	$k_{i,j}^c$	Source
C.1	C.1	3	1.204	1400	JL_EST
C.1	C.1	1	1.38	700	MC_88
Br	C.2	1	1.89	500	EXP
C.1	C.2	1	1.44	1340	*
C.1	C.2	2	1.44	1340	*
C.2	C.2	2	1.335	1340	WHITE_77
C.2	C.2	1	1.47	700	MC_88
C.1	C.3	1	1.458	640	*
C.2	C.3	1	1.501	639	WHITE_75
C.3	C.3	1	1.54	633.6	*
Br	C.ar	1	1.85	500	*
C.1	C.ar	1	1.44	1340	*
C.2	C.ar	1	1.51	1340	*
C.3	C.ar	1	1.525	640	*
C.ar	C.ar	ar	1.395	1400	*
C.ar	C.ar	1	1.480	1000	M7/22/88
C.2	Cl	1	1.75	520	*
C.3	Cl	1	1.767	600	*
C.ar	Cl	1	1.75	513.36	*
C.2	F	1	1.33	1200	*
C.3	F	1	1.36	600	*
C.ar	F	1	1.33	500	*
*	H	1	1.008	700	*
C.1	H	1	1.056	700	JL_EST
C.2	H	1	1.089	692	*
C.3	H	1	1.1	662.4	WHITE_77
C.ar	H	1	1.084	692	*
C.ar	I	1	2.05	490	*
C.1	N.1	3	1.158	1600	*
C.1	N.2	1	1.33	1300	*
C.1	N.2	2	1.33	1300	*
C.2	N.2	2	1.27	1305.94	*
C.2	N.2	1	1.444	1300	MC_88
C.3	N.2	1	1.44	760.2	*
C.ar	N.2	1	1.346	1305.94	*
N.2	N.2	2	1.346	1305.94	*
N.2	N.2	1	1.418	1300	MC_88
C.2	N.3	1	1.33	1300	EXP
C.3	N.3	1	1.47	760	*
C.ar	N.3	1	1.41	720	EXP
H	N.3	1	1.08	692	*
C.2	N.4	1	1.33	1300	*
C.3	N.4	1	1.47	760	*
C.ar	N.4	1	1.41	720	*
C.2	N.am	am	1.345	870.1	*
C.2	N.am	1	1.345	870.1	*
C.3	N.am	1	1.45	677.6	*
C.ar	N.am	1	1.416	1090.08	*
H	N.am	1	1	700	*
N.2	N.am	1	1.44	667.6	*
N.am	N.am	1	1.45	744.48	*
C.ar	N.ar	ar	1.346	1305.94	*
N.ar	N.ar	ar	1.33	1400	*
C.2	N.pl3	1	1.3	1200	*
C.3	N.pl3	1	1.45	676	*
C.ar	N.pl3	1	1.35	1306	*
H	N.pl3	1	1.03	692	*
N.2	N.pl3	1	1.35	1305.94	EXP
C.2	O.2	2	1.22	1555.2	*
N.am	O.2	1	1.24	1120	*
N.pl3	O.2	2	1.21	680	EXP
C.2	O.3	1	1.33	699.84	M7/22/88
C.3	O.3	1	1.43	618.9	M7/22/88
C.ar	O.3	1	1.39	700	M7/22/88
H	O.3	1	.95	1007.5	*
N.2	O.3	1	1.405	1200	*

Table II. (continued)

Atom <i>i</i>	Atom <i>j</i>	Bond type ^a	d_{ij}^0 ^b	k_{ij} ^c	Source
N.pl3	O.3	1	1.4	620	EXP
O.3	O.3	1	1.48	1172.16	*
C.3	P.3	1	1.83	407.6	*
O.2	P.3	1	1.49	1400	M7/22/88
O.2	P.3	2	1.49	1400	M7/22/88
O.3	P.3	1	1.6	800	*
C.2	S.2	2	1.71	400	*
C.3	S.2	1	1.8	381.6	*
C.ar	S.2	1	1.74	700	*
C.2	S.3	1	1.78	360	*
C.3	S.3	1	1.817	381.6	*
C.ar	S.3	1	1.77	360	*
N.3	S.3	1	1.625	360	*
N.4	S.3	1	1.625	360	*
O.2	S.3	2	1.45	600	*
S.3	S.3	1	2.03	600	EXP
C.2	S.o	1	1.71	360	TRIPOS_86
C.3	S.o	1	1.8	381.6	TRIPOS_86
O.2	S.o	2	1.45	600	*
O.3	S.o	1	1.5	600	*
C.3	S.o2	1	1.8	381.6	TRIPOS_86
O.2	S.o2	2	1.45	600	TRIPOS_86
O.3	S.o2	1	1.5	600	*

^aBond types: 1, single bond; 2, double bond; 3, triple bond; ar, aromatic bond; am, amide bond.

^bAngstroms.

^ckcal/mol-angstrom²

The energy associated with bond stretching and compression is given by the expression:

$$E_{ij} = k_{ij} * (d_{ij} - d_{ij}^0)^2$$

where d_{ij} is the actual bond length, and k and d^0 are taken from the table above. If no bond parameters are found the starting bond length is used for d_{ij} with a stretching constant of 600 kcal/mol⁻¹ angstrom⁻². An asterisk denotes a wildcard parameter.

cyclo-(L-Ala-L-Ala-Gly-Gly-L-Ala-Gly),²⁰ *cyclo*-(L-Ala-L-Ala-Gly-L-Ala-Gly-Gly),²⁰ and *cyclo*-(Gly-Gly-D-Ala-D-Ala-Gly-Gly)²¹ were taken from the literature, hydrogens being added as necessary. For these structures the crystallographic neighbors were generated using the crystallographic functions of Sybyl so that the reference peptide was completely surrounded by a lattice including all molecules with atoms within 6 angstroms of any atom in the reference molecule, as was done in other benchmark tests.¹³ The lattices were then optimized as described above except that the optimization affected only the geometry of the reference peptide unit cell.

Conformer structures for energy difference and torsional barrier calculations were built using the Sybyl molecular modeling program. Metastable intermediates at the top of torsional barriers were stabilized by a single torsional penalty function with the same form as the torsional potential given in Table V of 100 kcal/deg-mol, defined by the two central atoms and the heaviest of the attached atoms.

The minimizations were carried out as described except that the convergence criterion was lowered to 0.001 kcal/mol-angstrom rms force over all atoms to ensure stable energy values.

For a general test of the force field 76 molecules were extracted from a subset of the Cambridge Structural Database.²² The entries included a variety of organic functions, each molecule with an R factor less than 0.05. The structures were retrieved from the database using the Sybyl interface with connectivity and atom types determined algorithmically. All atom and bond types were rechecked by individual examination and corrected if necessary. All solvents and counterions were removed and hydrogens were added as necessary, although the structures had the majority of hydrogen positions determined. The structures are shown in Figure 1. All 76 structures were minimized as described above. As an additional test to determine the sensitivity of the ending structure to the starting position, the coordinates

Table III. Angle bending parameters for the Tripos 5.2 force field.

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Theta	<i>k</i> ^a	Source
*	C.1	*	180	0.04	*
C.1	C.1	C.2	180	0.04	JL_EST
C.2	C.1	N.1	180	0.04	*
C.3	C.1	N.1	180	0.04	*
C.ar	C.1	N.1	180	0.04	*
N.1	C.1	O.3	180	0.04	*
*	C.2	*	120	0.024	WHITE_77
Br	C.2	Br	120	0.02	EXP
Br	C.2	C.2	120	0.036	EXP
C.2	C.2	C.2	121.7	0.018	*
C.1	C.2	C.3	120	0.024	JL_EST
C.2	C.2	C.3	121	0.024	WHITE_77
C.3	C.2	C.3	116.4	0.046	WHITE_77
C.1	C.2	C.ar	120	0.024	*
C.2	C.2	C.ar	120	0.026	*
C.3	C.2	C.ar	120	0.024	EXP
C.ar	C.2	C.ar	120	0.024	*
C.2	C.2	Cl	120	0.036	*
C.ar	C.2	Cl	120	0.036	*
Cl	C.2	Cl	122	0.03	*
*	C.2	H	120	0.012	JL_EST
C.1	C.2	N.2	123	0.07	*
C.2	C.2	N.2	120	0.024	*
C.3	C.2	N.2	118	0.02	*
C.ar	C.2	N.2	120	0.04	*
C.2	C.2	N.3	120	0.024	*
C.3	C.2	N.3	118	0.04	EXP
N.2	C.2	N.3	121.8	0.03	EXP
N.3	C.2	N.3	116.4	0.03	EXP
C.2	C.2	N.am	120	0.024	*
C.3	C.2	N.am	117	0.02	WHITE_75
C.ar	C.2	N.am	120	0.04	*
N.2	C.2	N.am	123	0.07	*
N.am	C.2	N.am	120	0.03	*
C.2	C.2	N.pl3	120	0.024	*
C.3	C.2	N.pl3	117	0.02	EXP
N.2	C.2	N.pl3	123	0.07	EXP
C.1	C.2	O.2	120	0.06	JL_EST
C.2	C.2	O.2	120	0.026	*
C.3	C.2	O.2	120	0.026	M7/22/88
C.ar	C.2	O.2	120	0.026	*
N.3	C.2	O.2	120	0.026	EXP
N.am	C.2	O.2	123	0.03	WHITE_75
N.pl3	C.2	O.2	123	0.03	EXP
C.2	C.2	O.3	120	0.072	*
C.3	C.2	O.3	114	0.03	M7/22/88
C.ar	C.2	O.3	120	0.03	*
N.am	C.2	O.3	110.5	0.014	*
O.2	C.2	O.3	120	0.03	*
N.2	C.2	S.3	125.6	0.028	*
N.am	C.2	S.3	111.5	0.03	*
O.2	C.2	S.3	125	0.016	*
*	C.3	*	109.5	0.02	WHITE_77
C.2	C.3	C.2	109.5	0.018	*
C.1	C.3	C.3	109.5	0.024	*
C.2	C.3	C.3	109.5	0.018	WHITE_75
C.3	C.3	C.3	109.5	0.024	*
C.2	C.3	C.ar	109.47	0.018	*
C.3	C.3	C.ar	109.5	0.024	*
C.ar	C.3	C.ar	109.5	0.018	*
C.3	C.3	Cl	109.5	0.02	*
Cl	C.3	Cl	109.5	0.02	*
C.ar	C.3	F	110	0.024	JL_EST
F	C.3	F	109.5	0.04	*
*	C.3	H	109.5	0.016	WHITE_77

Table III. (continued)

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Theta	<i>k</i> ^a	Source
C.2	C.3	H	110	0.016	WHITE_75
H	C.3	H	109.5	0.024	WHITE_75
C.3	C.3	N.2	109.5	0.018	*
C.2	C.3	N.3	109.5	0.018	*
C.3	C.3	N.3	109.5	0.024	*
C.ar	C.3	N.3	109.5	0.018	*
C.2	C.3	N.am	109.5	0.022	*
C.3	C.3	N.am	109.5	0.018	*
C.ar	C.3	N.am	109.5	0.02	EXP
H	C.3	N.am	110	0.02	WHITE_75
N.2	C.3	N.am	109.5	0.02	EXP
N.am	C.3	N.am	109.5	0.04	*
F	C.3	N.ar	109.5	0.04	*
C.2	C.3	N.pl3	109.5	0.018	*
C.3	C.3	N.pl3	109.5	0.02	*
C.ar	C.3	N.pl3	109.5	0.02	*
C.2	C.3	O.3	109.5	0.022	*
C.3	C.3	O.3	109.5	0.022	*
C.ar	C.3	O.3	109.5	0.018	*
N.am	C.3	O.3	109.5	0.02	*
O.3	C.3	O.3	109.5	0.02	*
C.3	C.3	P.3	112	0.014	*
C.2	C.3	S.2	109.5	0.018	*
C.3	C.3	S.2	109.5	0.018	*
N.am	C.3	S.2	109.5	0.04	*
C.2	C.3	S.3	107.8	0.018	*
C.3	C.3	S.3	107.8	0.018	*
C.ar	C.3	S.3	107.8	0.018	*
N.am	C.3	S.3	109.5	0.024	JL_EST
O.3	C.3	S.3	107.8	0.02	*
*	C.ar	*	120	0.024	*
Br	C.ar	C.ar	120	0.036	*
C.2	C.ar	C.ar	120	0.024	*
C.3	C.ar	C.ar	120	0.024	*
C.ar	C.ar	C.ar	120	0.024	*
C.ar	C.ar	Cl	120	0.036	*
C.ar	C.ar	F	120	0.036	*
C.ar	C.ar	I	120	0.036	*
C.3	C.ar	N.2	120	0.04	*
C.ar	C.ar	N.2	120	0.04	*
N.2	C.ar	N.2	120	0.04	*
C.ar	C.ar	N.3	120	0.062	EXP
C.ar	C.ar	N.am	120	0.062	*
N.2	C.ar	N.am	118	0.04	*
N.am	C.ar	N.am	120	0.03	*
C.2	C.ar	N.ar	120	0.04	EXP
C.3	C.ar	N.ar	120	0.04	EXP
C.ar	C.ar	N.ar	120	0.024	EXP
N.am	C.ar	N.ar	118	0.04	EXP
C.3	C.ar	N.pl3	120	0.04	*
C.ar	C.ar	N.pl3	120	0.04	*
N.2	C.ar	N.pl3	120	0.04	*
N.am	C.ar	N.pl3	120	0.04	*
N.pl3	C.ar	N.pl3	120	0.04	*
C.3	C.ar	O.3	120	0.04	*
C.ar	C.ar	O.3	120	0.062	*
C.ar	C.ar	S.2	120	0.062	*
C.ar	C.ar	S.3	120	0.062	*
*	N.1	*	180	0.08	*
*	N.2	*	120	0.04	*
C.1	N.2	C.2	120	0.04	*
C.2	N.2	C.2	123	0.08	EXP
C.2	N.2	C.3	110	0.082	*
C.2	N.2	C.ar	123	0.08	*
C.3	N.2	C.ar	110	0.082	*
C.ar	N.2	C.ar	120	0.04	*

Table III. (continued)

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Theta	<i>k</i> ^a	Source
C.2	N.2	N.2	112	0.044	*
C.3	N.2	N.2	118	0.04	*
C.ar	N.2	N.2	118	0.04	*
C.2	N.2	N.am	120	0.044	EXP
C.2	N.2	N.pl3	120	0.044	EXP
C.2	N.2	O.3	105	0.044	EXP
*	N.3	*	109.5	0.04	*
C.2	N.3	C.3	110	0.04	EXP
C.3	N.3	C.3	109.5	0.018	*
C.3	N.3	C.ar	118	0.04	EXP
C.ar	N.3	C.ar	118	0.04	EXP
C.3	N.3	S.2	109.5	0.04	*
*	N.4	*	109.5	0.01	TRIPOS_86
C.3	N.4	C.3	109.5	0.018	*
*	N.am	*	120	0.02	*
C.2	N.am	C.2	120	0.018	*
C.2	N.am	C.3	118	0.044	*
C.3	N.am	C.3	122	0.04	*
C.2	N.am	C.ar	120	0.052	*
C.3	N.am	C.ar	118	0.044	*
C.ar	N.am	C.ar	120	0.044	*
C.2	N.am	H	119	0.016	WHITE_75
C.3	N.am	H	117	0.02	WHITE_75
C.2	N.am	N.2	120	0.018	EXP
C.3	N.am	N.2	120	0.024	*
C.ar	N.am	N.2	109.5	0.044	*
C.2	N.am	N.am	120	0.018	*
C.3	N.am	N.am	120	0.024	*
C.ar	N.am	N.am	120	0.052	*
C.2	N.am	O.2	120	0.024	*
C.3	N.am	O.2	120	0.02	*
C.ar	N.am	O.2	120	0.024	*
O.2	N.am	O.2	120	0.02	*
*	N.ar	*	120	0.02	*
C.ar	N.ar	C.ar	120	0.04	EXP
*	N.pl3	*	120	0.04	*
C.2	N.pl3	C.2	120	0.04	EXP
C.2	N.pl3	C.ar	120	0.04	EXP
C.3	N.pl3	C.ar	120	0.04	*
C.ar	N.pl3	C.ar	120	0.04	*
C.2	N.pl3	N.2	120	0.018	EXP
C.ar	N.pl3	N.2	120	0.018	EXP
C.2	N.pl3	O.2	112	0.04	EXP
O.2	N.pl3	O.2	127	0.06	EXP
*	O.3	*	109.5	0.02	M7/22/88
C.2	O.3	C.2	110	0.02	M7/22/88
C.2	O.3	C.3	109.5	0.044	M7/22/88
C.3	O.3	C.3	109.5	0.044	M7/22/88
C.2	O.3	C.ar	110	0.02	M7/22/88
C.3	O.3	C.ar	110	0.02	M7/22/88
C.ar	O.3	C.ar	110	0.02	M7/22/88
C.2	O.3	N.2	108.5	0.044	EXP
C.3	O.3	O.3	103.9	0.094	*
C.3	O.3	P.3	120	0.01	*
*	P.3	*	109.5	0.02	*
O.2	P.3	O.2	109.5	0.02	M7/22/88
Du	P.3	O.3	109.5	0.014	M7/22/88
O.2	P.3	O.3	109.5	0.02	M7/22/88
O.3	P.3	O.3	109.5	0.02	M7/22/88
*	S.2	*	110	0.04	*
C.3	S.2	N.3	111	0.04	*
C.ar	S.2	N.3	111	0.04	*
*	S.3	*	97	0.02	*
C.2	S.3	C.3	94.3	0.022	*
C.3	S.3	C.3	98	0.02	*
C.ar	S.3	C.ar	97.5	0.062	*

Table III. (continued)

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Theta	<i>k</i> ^a	Source
C.3	S.3	S.3	102.9	0.06	EXP
*	S.o	O.2	107	0.04	TRIPOS_86
*	S.o2	O.2	107	0.04	TRIPOS_86
O.2	S.o2	O.2	118	0.04	TRIPOS_86

^akcal/mol-degree²^bThe angle bending term is given by the formula

$$E = k_{i,j,k} * (\text{actual angle} - \text{theta})^2$$

Wild cards entries, denoted by an asterisk, will be used if there is not a specific match. If no entry is found, the minimizer informs the user and substitutes $k = 0.02 \text{ kcal/mol}^{-1} \text{ degree}^{-2}$, and the equilibrium angle is assumed to be the starting angle for the minimization.

of all atoms in all 76 molecules were truncated to one decimal place and the resulting structures minimized.

The resulting structural data was analyzed by comparison of internal coordinates, bond lengths, bond angles, and bonded torsion angles, of the original and the minimized structures. The figures reflect the optimized value minus the value from the crystal structure. In each case only bond lengths and angles not involving hydrogens were considered. In addition, subsets of the torsion angles were selected for the analysis to investigate the contribution of side chain conformational changes during energy minimization. For the peptides and crambin, the backbone angles and the C_α-C_β torsions were analyzed separately. For the molecules from the Cambridge Structural Database the set of all bonds to atoms in ring structures was selected for separate torsional analysis.

Table IV. Out-of-plane bending terms for the Tripos 5.2 force field.

Atom	<i>k</i> ^a	Source
C.2	480	TRIPOS_85
C.ar	480	TRIPOS_85
N.2	120	TRIPOS_85
N.am	120	TRIPOS_85
N.ar	120	TRIPOS_85
N.pl3	120	TRIPOS_85

^akcal/mol-angstrom²

The term used for out-of-plane bending energy is

$$E = k * d^2$$

where *k* is the out-of-plane bending force constant for the atom, and *d* is the distance from the atom to the plane defined by its three attached atoms.

Atoms not found in this table make no contributions to the out-of-plane bending energy.

The statistical analysis included calculating the maximum deviations, the standard deviations, the average deviations, the rms deviations, and the coefficient of skewness of the internal coordinates. Large deviations were flagged in a file for individual inspection. The average deviations are an indicator of systematic over or underestimation of a parameter; in the best case these should be zero. The coefficient of skewness is the third moment of the deviations divided by the standard deviation.

RESULTS AND DISCUSSION

Cyclic Peptides

Table VII summarizes force field studies of the Tripos, Amber,^{5,13} and Amber/OPLS⁵ force fields based on three cyclic hexapeptides in the crystal lattice. The calculations with the Tripos force field were performed with each peptide in a fixed surrounding. Workers with the Amber and Amber/OPLS force fields implemented crystallographic boundary conditions so that each member of the cell moved in unison,⁵ while Hall and Pavitt leave this point unspecified.¹³ This makes an exact comparison among the studies difficult.

For the method used in this work it appears that the Tripos force field is comparable to the other two more specialized force fields in this test. The Tripos rms backbone torsional deviations are greater than those produced by the Amber united atom force field according to the study of Hall and Pavitt¹³ at 9.80 and 9.97 degrees for phi and psi, respectively as compared to 3.7 degrees for phi and psi combined. The average deviation in phi and psi angles for the Tripos force

Table V. Torsional parameters for the Tripos 5.2 force field.

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Atom <i>l</i>	Bond Type ^a	<i>k</i> ^b	<i>s</i>	Comment
*	C.1	C.1	*	3	0	1	JL_EST
*	C.1	C.1	*	1	0	1	*
*	C.1	C.2	*	1	0	1	*
*	C.1	C.2	*	2	0	1	*
*	C.2	C.2	*	2	12.5	-2	WHITE_77
*	C.2	C.2	*	1	1.424	-2	M7/22/88
*	C.1	C.3	*	1	0.	1	*
*	C.2	C.3	*	1	0.12	-3	*
*	C.2	C.3	C.2	1	0.126	3	WHITE_77
*	C.2	C.3	C.3	1	0.126	3	WHITE_77
*	C.2	C.3	H	1	0.274	3	WHITE_77
O.2	C.2	C.3	C.3	1	0.7	-3	JL_EST
*	C.3	C.3	*	1	0.2	3	*
C.2	C.3	C.3	C.2	1	0.04	3	WHITE_77
C.2	C.3	C.3	C.3	1	0.126	3	WHITE_77
C.3	C.3	C.3	C.3	1	0.5	3	MC_88
*	C.3	C.3	H	1	0.32	3	MC_88
*	C.1	C.ar	*	1	0	1	*
*	C.2	C.ar	*	1	1.6	-2	*
*	C.3	C.ar	*	1	0.12	-3	*
*	C.ar	C.ar	*	ar	2.0	-2	M7/22/88
*	C.ar	C.ar	*	1	0.6	-2	M7/22/88
*	C.1	N.2	*	1	0	1	*
*	C.1	N.2	*	2	0	1	*
*	C.2	N.2	*	2	12	-2	*
*	C.2	N.2	*	1	12	-2	*
*	C.3	N.2	*	1	0.4	-3	*
*	C.ar	N.2	*	1	1.6	-2	*
*	N.2	N.2	*	2	1.6	-2	*
*	N.2	N.2	*	1	1.6	-2	*
*	C.2	N.3	*	1	0.12	-3	*
*	C.3	N.3	*	1	0.2	3	*
*	C.ar	N.3	*	1	0.12	-3	*
*	N.3	N.3	*	1	0.2	3	*
*	C.2	N.am	*	am	6.46	-2	*
*	C.2	N.am	*	1	6.46	-2	*
*	C.3	N.am	*	1	0.2	3	*
*	C.ar	N.am	*	1	1.6	-2	*
*	N.2	N.am	*	1	1.6	-2	*
*	N.3	N.am	*	1	0.12	-3	*
*	N.am	N.am	*	1	1.6	-2	*
*	C.ar	N.ar	*	ar	1.6	-2	EXP
*	C.2	N.pl3	*	1	12	-2	*
*	C.3	N.pl3	*	1	0.4	-3	*
*	C.ar	N.pl3	*	1	1.6	-2	*
*	N.2	N.pl3	*	1	1.6	-2	*
*	N.pl3	N.pl3	*	1	1.6	-2	*
*	C.2	O.3	*	1	5.8	-2	*
*	C.3	O.3	*	1	1.2	3	*
*	C.ar	O.3	*	1	1.2	-2	*
*	N.2	O.3	*	1	1	2	EXP
*	N.3	O.3	*	1	0.2	3	*
*	C.2	P.3	*	1	1	-2	*
*	C.3	P.3	*	1	0.4	3	*
*	C.ar	P.3	*	1	1	3	*
*	O.3	P.3	*	1	0.4	3	*
*	C.2	S.2	*	2	1	-2	*
*	C.3	S.2	*	1	0.4	3	*
*	C.ar	S.2	*	1	1	3	*
*	N.3	S.2	*	1	0.4	3	*
*	C.2	S.3	*	1	1	-2	*
*	C.3	S.3	*	1	0.4	3	*
*	C.ar	S.3	*	1	1	3	*
*	S.3	S.3	*	1	4	3	EXP

Table V. (continued)

Atom <i>i</i>	Atom <i>j</i>	Atom <i>k</i>	Atom <i>l</i>	Bond Type ^a	<i>k</i> ^b	<i>s</i>	Comment
C.2	C.2	C.3	C.2	1	0.126	-3	WHITE_77
C.2	C.2	C.3	H	1	0.273	-3	WHITE_77
C.3	C.2	C.3	C.2	1	0.126	3	WHITE_77
C.3	C.2	C.3	C.3	1	0.126	3	WHITE_77
C.3	C.2	C.3	H	1	0.274	3	WHITE_77
H	C.2	C.3	C.2	1	0.274	3	WHITE_77
H	C.2	C.3	C.3	1	0.274	3	WHITE_77
H	C.2	C.3	H	1	0.274	3	WHITE_77
C.2	C.2	C.3	*	1	0.126	-3	WHITE_77
C.3	C.2	C.3	*	1	0.126	3	WHITE_77
H	C.2	C.3	*	1	0.274	3	WHITE_77
C.2	C.2	C.3	C.3	1	0.126	-3	WHITE_77

^aBond types: 1, single bond; 2, double bond; 3, triple bond; ar, aromatic bond; am, amide bond.

^bkcal/mol-degree²

The torsional energy associated with four consecutively bonded atoms *i, j, k, l* is given by the function

$$E = k_{i,j,k,l} * (1 + s/|s| * \cos(|s|B_{i,j,k,l}))$$

Where *B* is the torsion angle between atoms *i, j, k*, and *l*, and the values of *k* and *s* are taken from the table above. A torsional contribution is made by every bonded quartet in the molecule.

Wild card entries will be substituted if an exact match is not found for the outer atoms of the torsion. If there are no entries that match the inner two atoms, the minimizer emits a message and uses the values 0.2 kcal/mol⁻¹ degree⁻² for *k* and 3 for *s*.

field was -3.19 and 3.08 degrees, respectively. The rms movement of atoms is greater than that of the Amber force field at 0.16 angstroms, but better than all but the ECEP2 and LEVB force fields tested by Hall and Pavitt.¹⁴ Although there are no explicit hydrogen bond terms, and no electrostatic attraction between hydrogen and oxygen, the rms deviation in intramolecular hydrogen bond length is less for the Tripos force field than the others at 0.09 angstroms.

The Amber force field is parametrized specifically for proteins and nucleic acids and has been shown to be considerably more accurate than other biopolymer force fields. In a comparison to the study of Hall and Pavitt¹³ on the same three cyclic peptides, the Tripos force field values are superior to the others with the exception of Amber, ECEP2, and the LEVB force fields as used by those workers.

Crambin

Table VIII summarizes studies comparing the same three force fields in minimization studies of crambin. Again there is the complication that the presence or absence of the crystal lattice during minimization influences the resulting structure. The most di-

rect comparison is between the Tripos force field and the work done using Amber without crystallographic boundary conditions.¹² The work done by Whitlow and Teeter¹² used the Kollman united atom force field with both a variety of dielectric constants, solvation models, and nonbonded cutoffs. They found that the protein volume depended on both the dielectric function and the solvation model with a deviation of up to 13%. The authors' work without electrostatics or solvation gave a volume change of 0.6% from an original volume of 4130 angstroms³ as calculated from the van der Waals surface.¹¹ The maximum error in bond lengths was a lengthening of 0.1 angstroms, the maximum angle deviation -9.7 degrees, and the maximum torsion deviation 53.82 degrees. This last value refers to the N-Ca-C-O angle of Ala₃₈; this residue is also responsible for the largest rms deviation due to the carbonyl oxygen. Ala₃₈ lies in a large turn just after the second beta chain. Upon crystallographic expansion of the lattice it is revealed that this residue is tightly sandwiched between the Thr₁ from its own chain and Asn₄₆ from a neighboring strand, with a closest contact of 4.1 angstroms from the beta carbon of Ala₃₈ to the nitrogen of Asn₄₆. Whitlow and Teeter found this residue to have the largest

Table VI. 6-12 potential parameters for the Tripos 5.2 force field.

Atom	r^a	k^b	Comment
C.3	1.7	0.107	*
C.2	1.7	0.107	*
C.ar	1.7	0.107	*
C.1	1.7	0.107	*
N.3	1.55	0.095	*
N.2	1.55	0.095	*
N.1	1.55	0.095	*
O.3	1.52	0.116	*
O.2	1.52	0.116	*
S.3	1.8	0.314	*
N.ar	1.55	0.095	*
P.3	1.8	0.314	*
H	1.5	0.042	M7/22/88
Br	1.85	0.434	*
Cl	1.75	0.314	*
F	1.47	0.109	*
I	1.98	0.623	*
S.2	1.8	0.314	*
N.pl3	1.55	0.095	*
LP	0	0.0	*
Na	1.2	0.400	*
K	1.2	0.400	*
Ca	1.2	0.600	*
Li	1.2	0.400	*
Al	1.2	0.042	*
Du	0	0.0	*
Si	1.2	0.042	*
N.am	1.55	0.095	*
S.o	1.7	0.314	*
S.o2	1.7	0.314	*
N.4	1.55	0.095	*

^aAngstroms.^bkcal/mol.

The nonbonded term is associated with any pair of atoms which are neither directly bonded to a common atom, or belong to substructures (residues) more than a specified cutoff distance away. The function used is

$$E = k_{ij} * (1.0/a^{12} - 2.0/a^6)$$

where a is the distance between the two atoms divided by the sum of their radii, and k is the geometric mean of the k constants associated with each atom.

Hydrogens attached to hydrogen bond donors have a constant, k , of zero in terms involving hydrogen bond acceptors.

movement, but they also noted large movements for Pro₅, Ser₆, Pro₁₉, Gly₂₀, and Gly₃₁. In this work these residues show statistically large deviations from the crystal structure upon minimization. The expansion of the crystal lattice indicates that each of these residues are contact points with neighboring molecules. Thus, the causes of these geometric discrepancies between minimized and crystal structures is not a weakness in either the Amber or Tripos force fields, but is due to the external influence of the crystal neighbors.

The average errors for the phi and psi angles using the Tripos force field are -1.5 and 1.3 degrees, respectively. This is similar, but much less pronounced, to the trend in the Amber force field where the average movements are -12 and 10 degrees.¹² This would indicate that although the magnitudes of the deviations are similar, the Tripos force field phi and psi averages are closer to the crystal structure. A similar trend is apparent in the study of the cyclic peptides discussed above. Interestingly, the coefficient of skewness is positive for both phi and psi angles.

Figure 2 superimposes the crystal and calculated structures of crambin in the region of Ala₃₈, showing the influence of the residues from the neighboring protein in the lattice.

Relative Conformational and Stereochemical Energies

Tables IX, X, and XI compare energy differences between different rotameric and stereochemical forms of various molecules, as measured experimentally, as calculated by the MM2 force field, and as calculated using the Tripos force field. Statistics summarizing the differences between experimental and either the Tripos or the MM2 force field values appear at the end of each table.

Summarizing the energy differences in Table IX, the Tripos energy differences have an rms error of 0.8 kcal/mol, while the MM2 calculated values have an average rms error of 0.5 kcal/mol.

The Tripos force field is comparable to MM2 in these cases. The figures for MM2 disregard the unusually high barrier calculated for the rotation of the inner bond of 1,3 butadiene, since the MM2 force field used in this study did not include pi electron calculations and therefore is not applicable to conjugated systems. The currently available MM2 force fields use pi electron calculations to treat conjugated systems accurately.²³ The Tripos force field handles conjugation without pi electron calculations by having different parameters for single, double, and triple bonds between a given atom pair.

The Tripos force field calculates the preferred conformation of propanal to exhibit an oxygen-hydrogen eclipsing interaction; the oxygen-methyl eclipsing conformation is slightly destabilized by steric interaction.

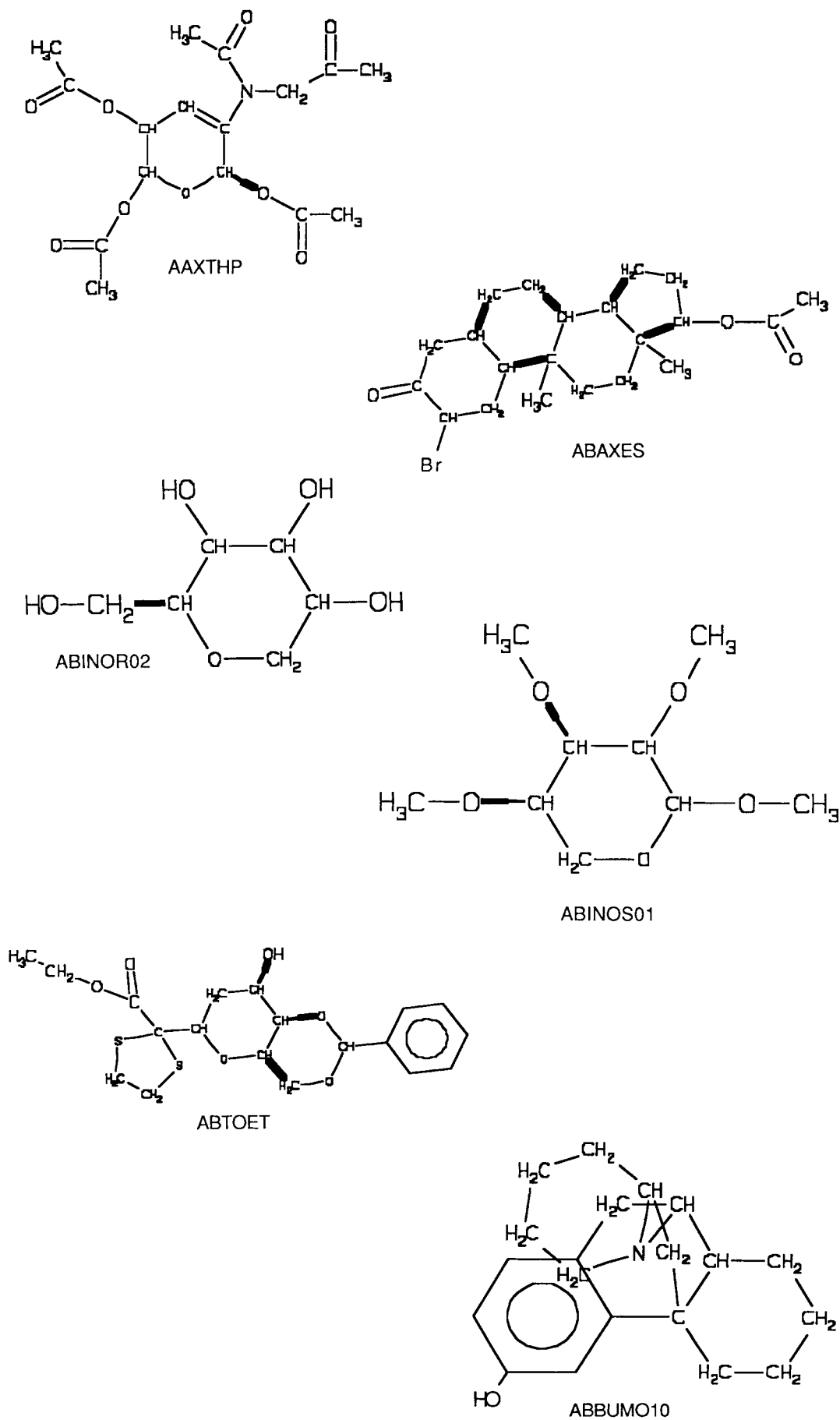


Figure 1. Organic molecules used in force field tests, with Cambridge Structural Database reference codes.

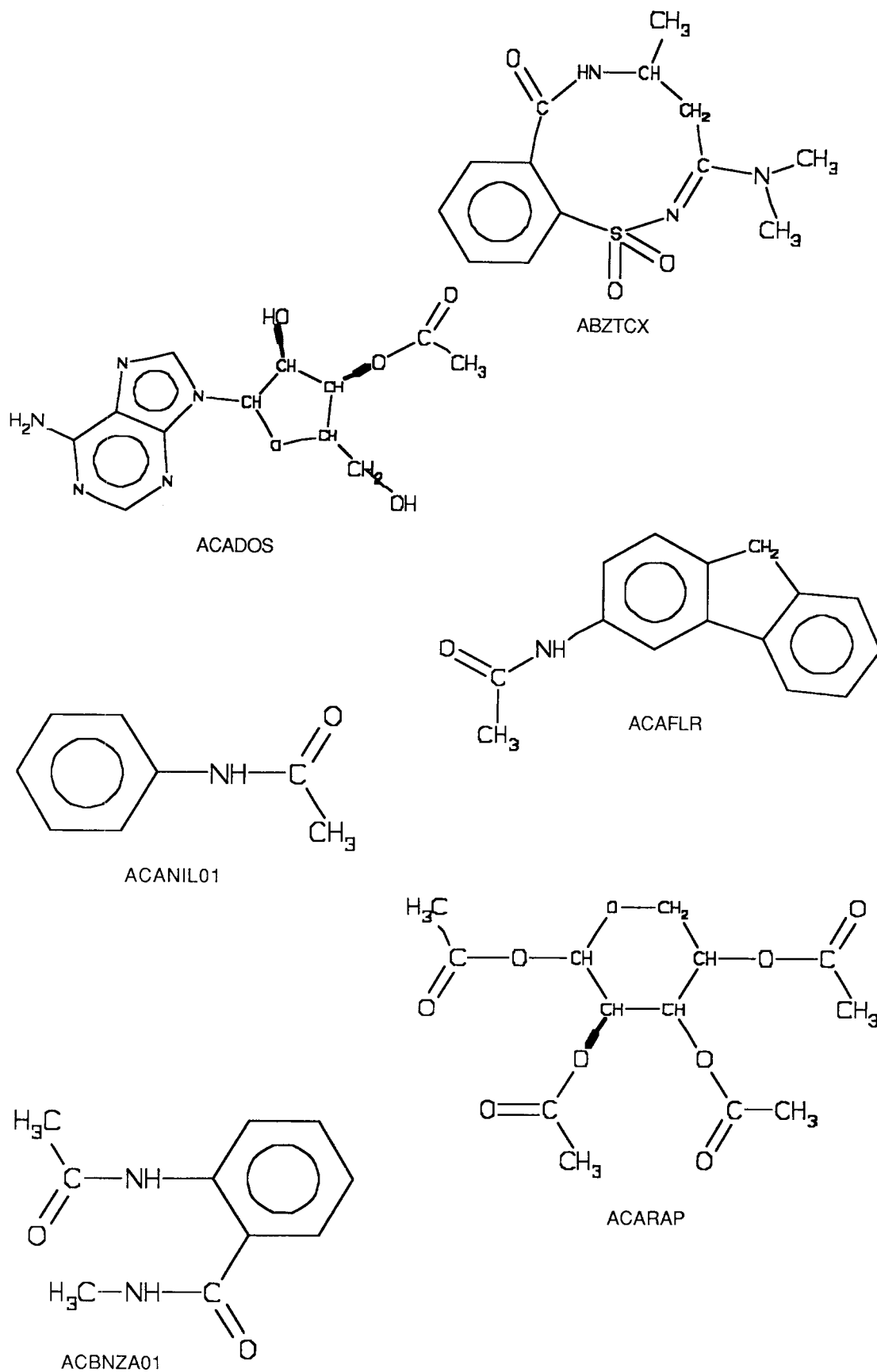


Figure 1. (continued)

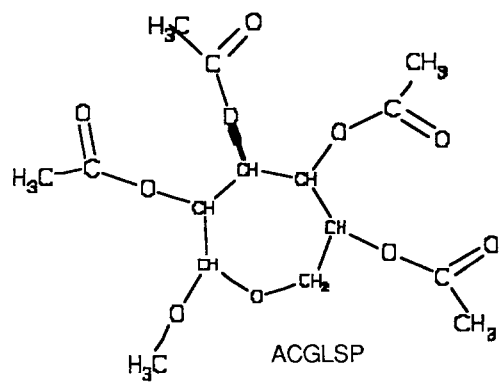
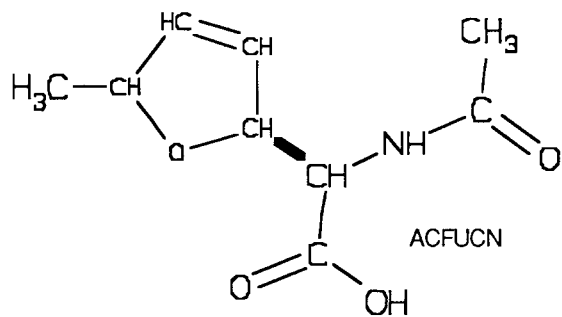
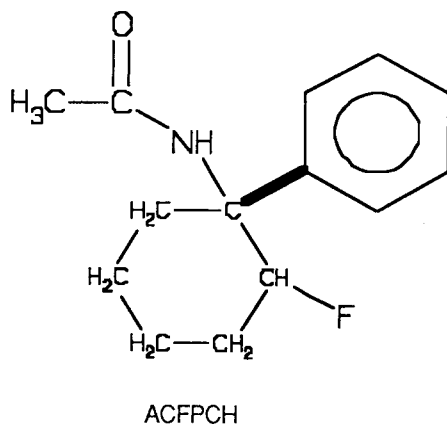
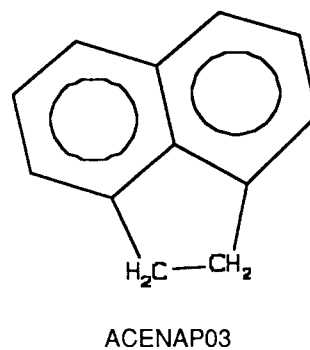
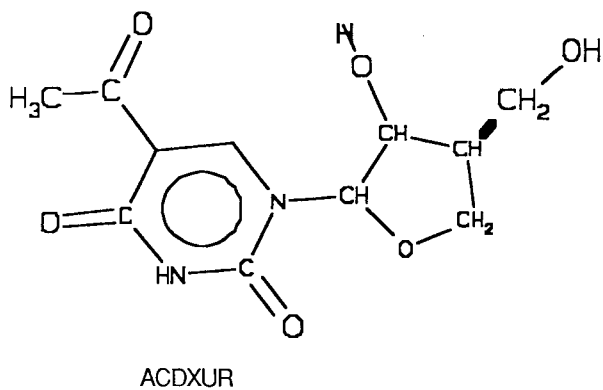
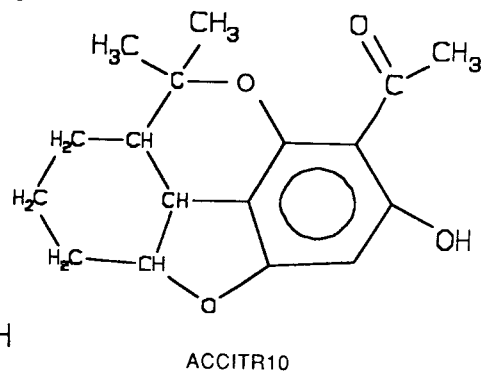
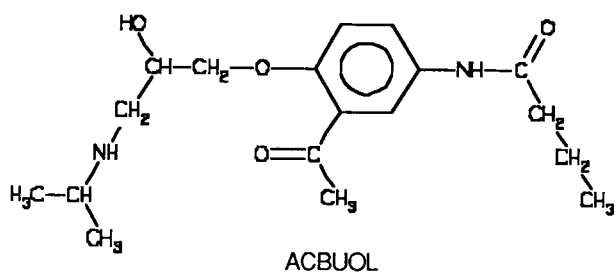


Figure 1. (continued)

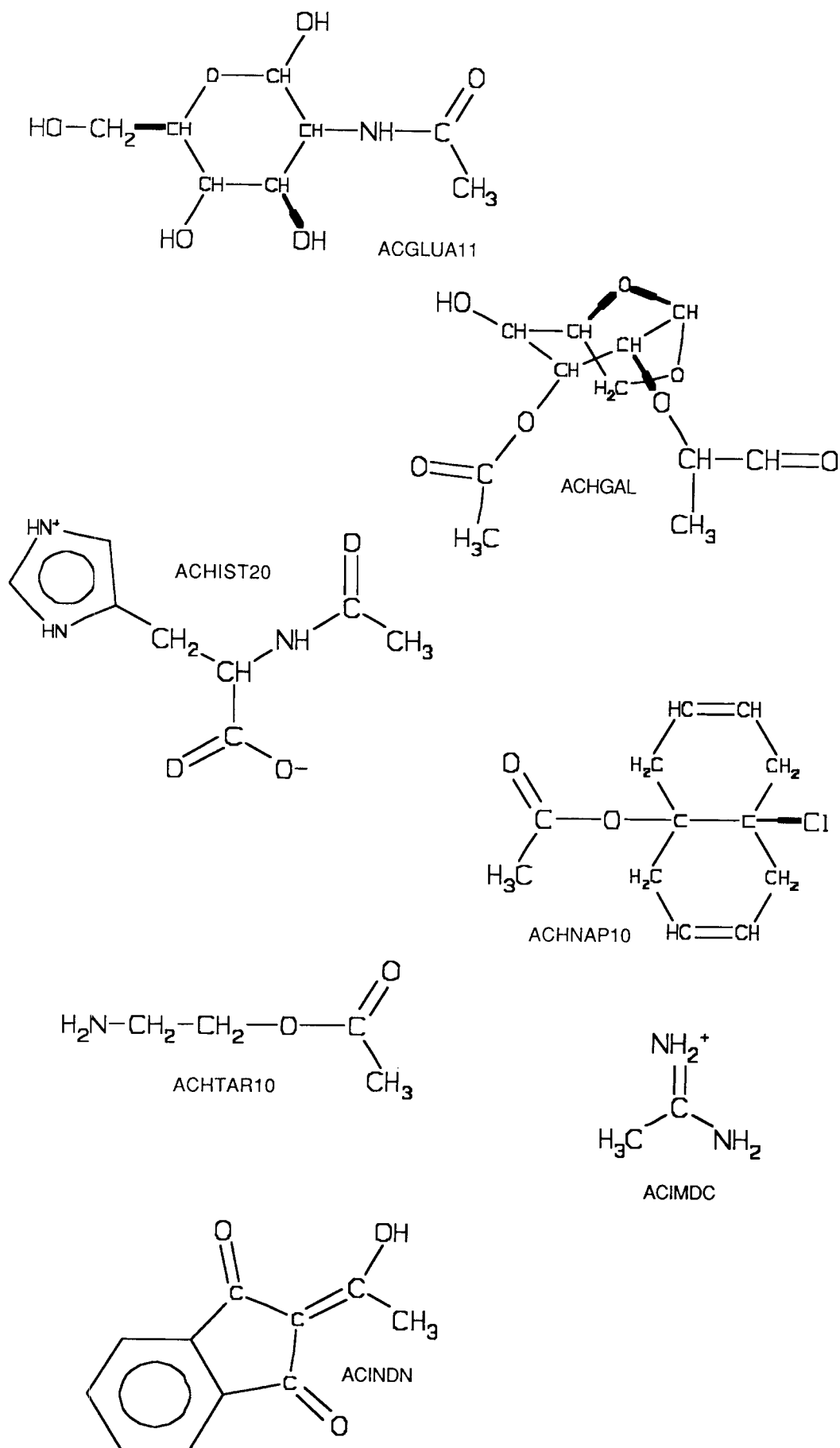


Figure 1. (continued)

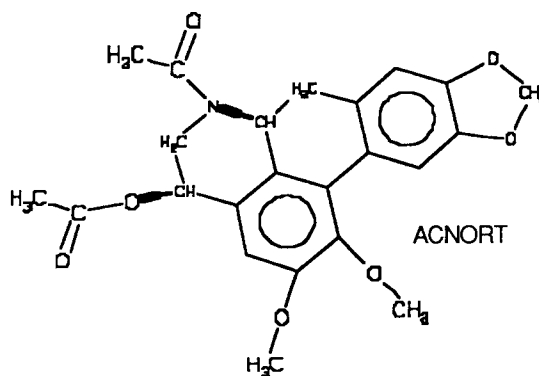
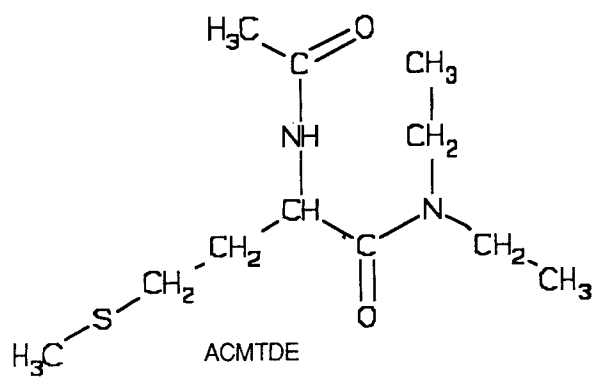
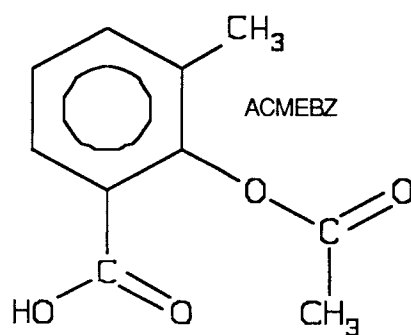
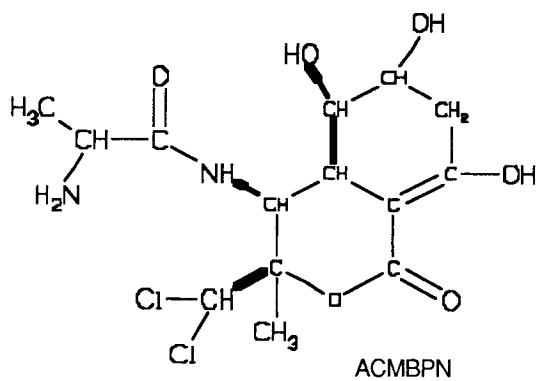
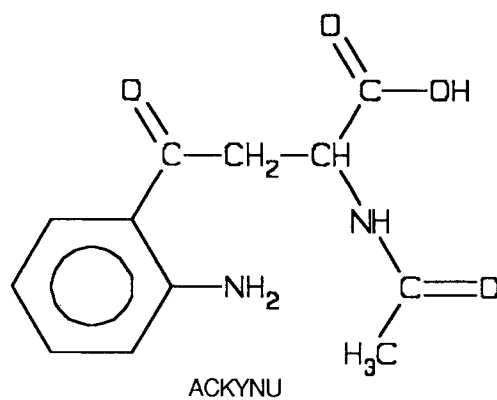
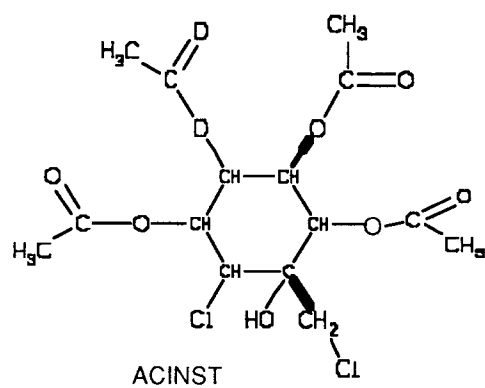


Figure 1. (continued)

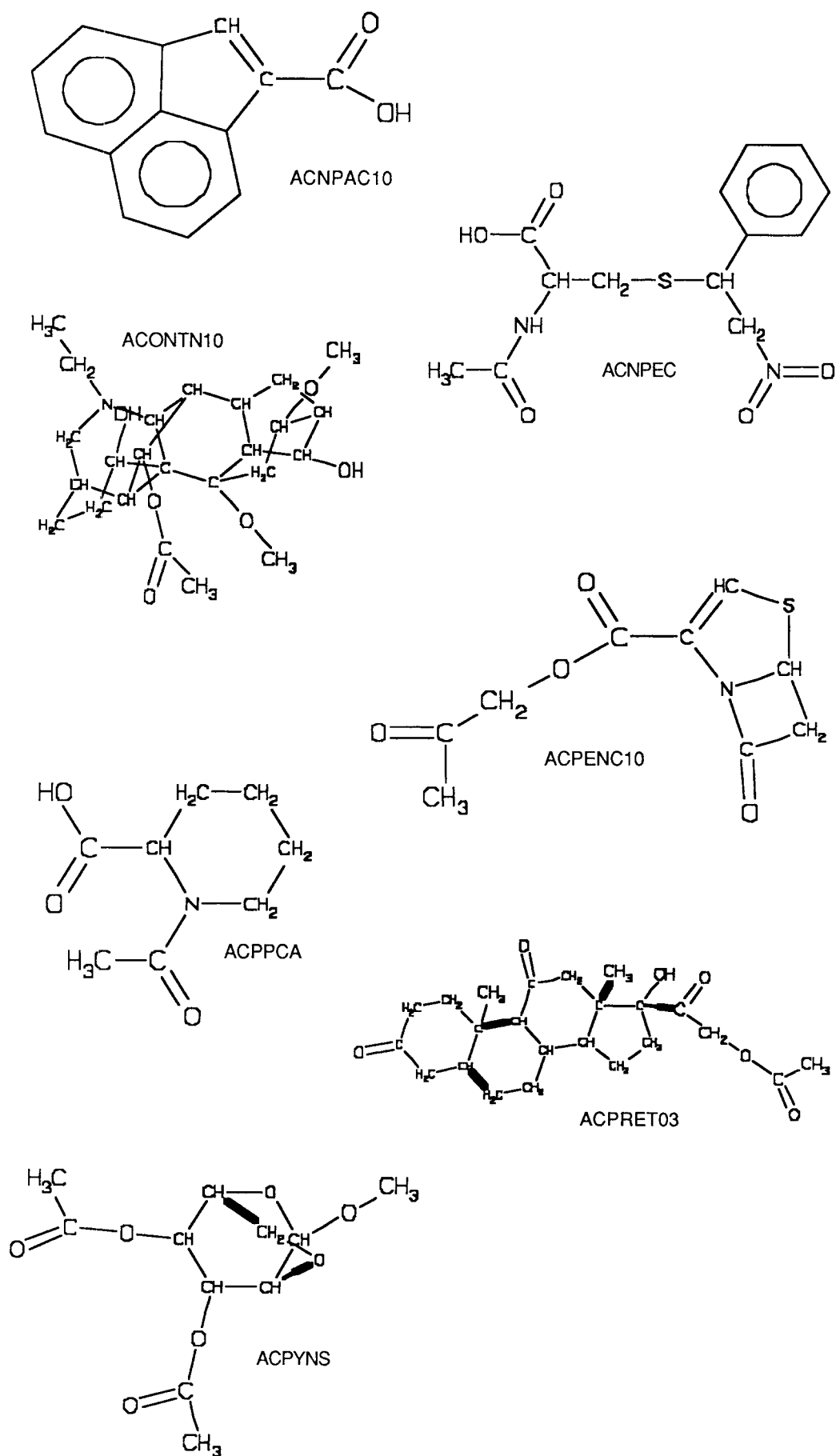


Figure 1. (continued)

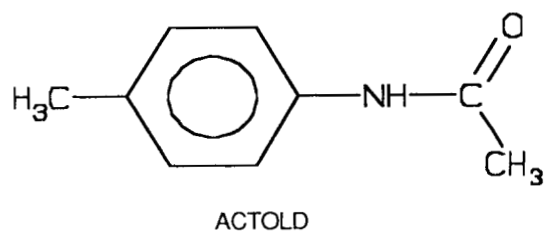
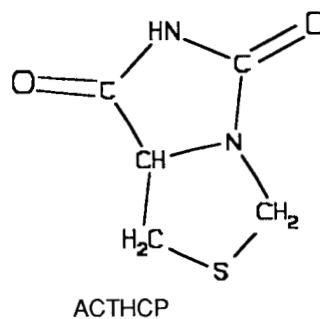
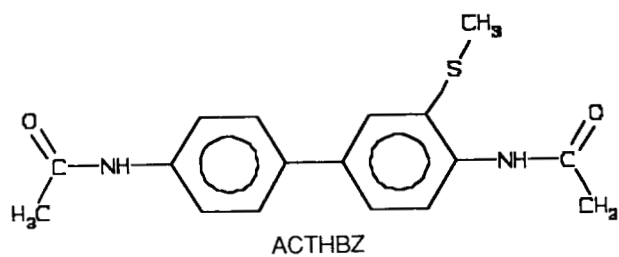
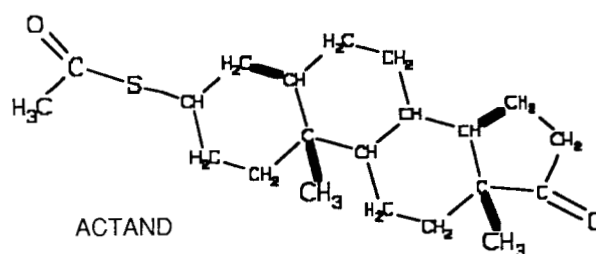
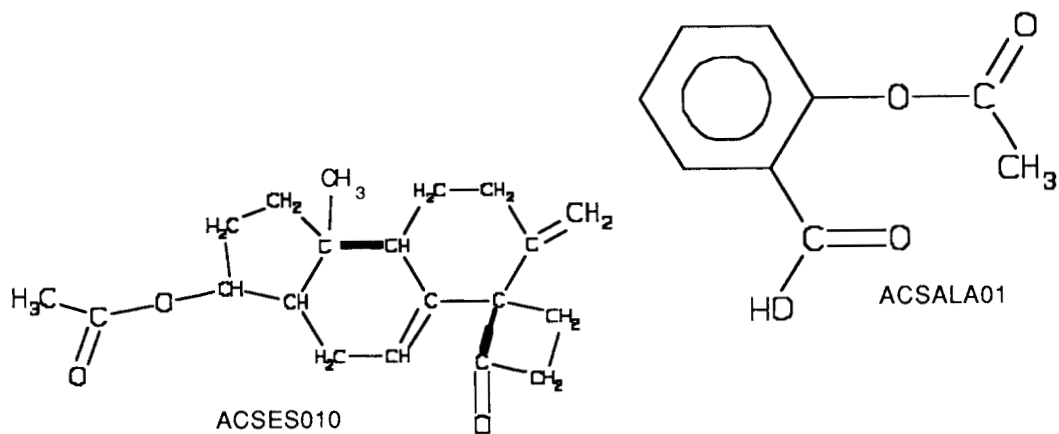
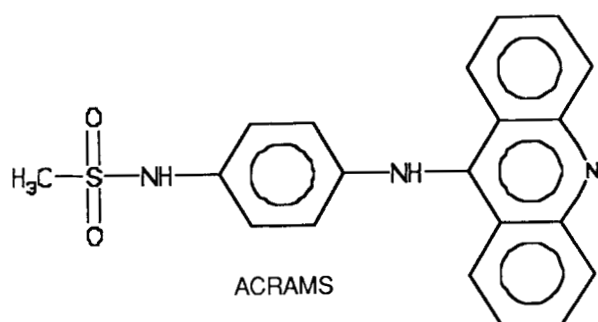


Figure 1. (continued)

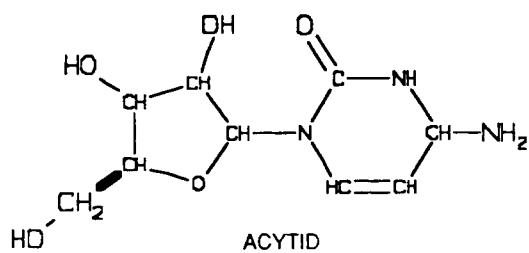
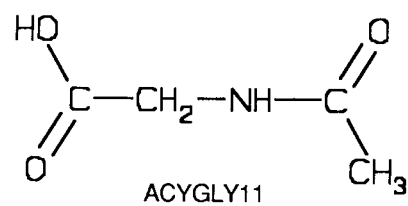
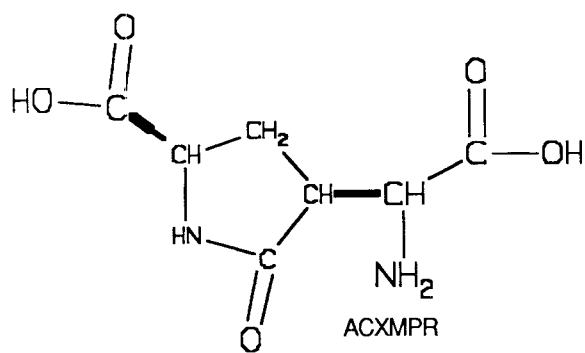
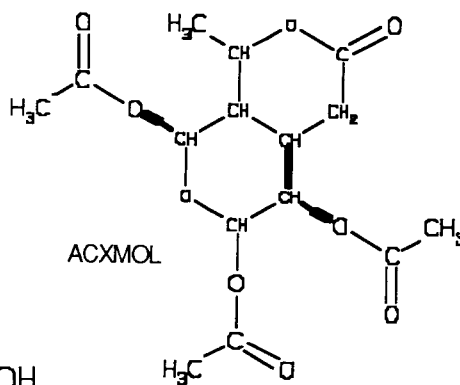
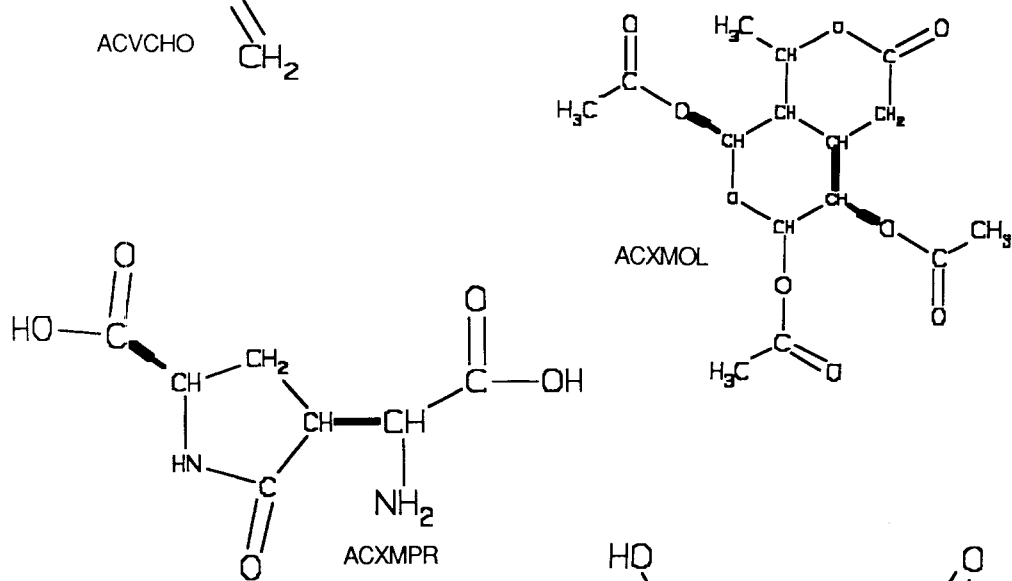
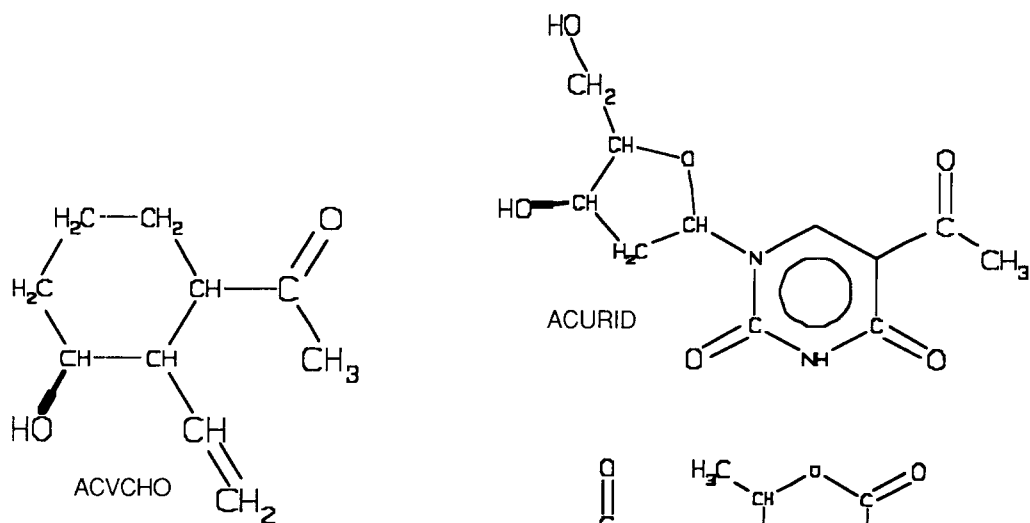
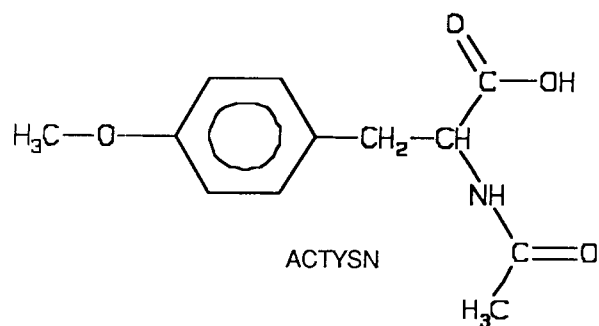


Figure 1. (continued)

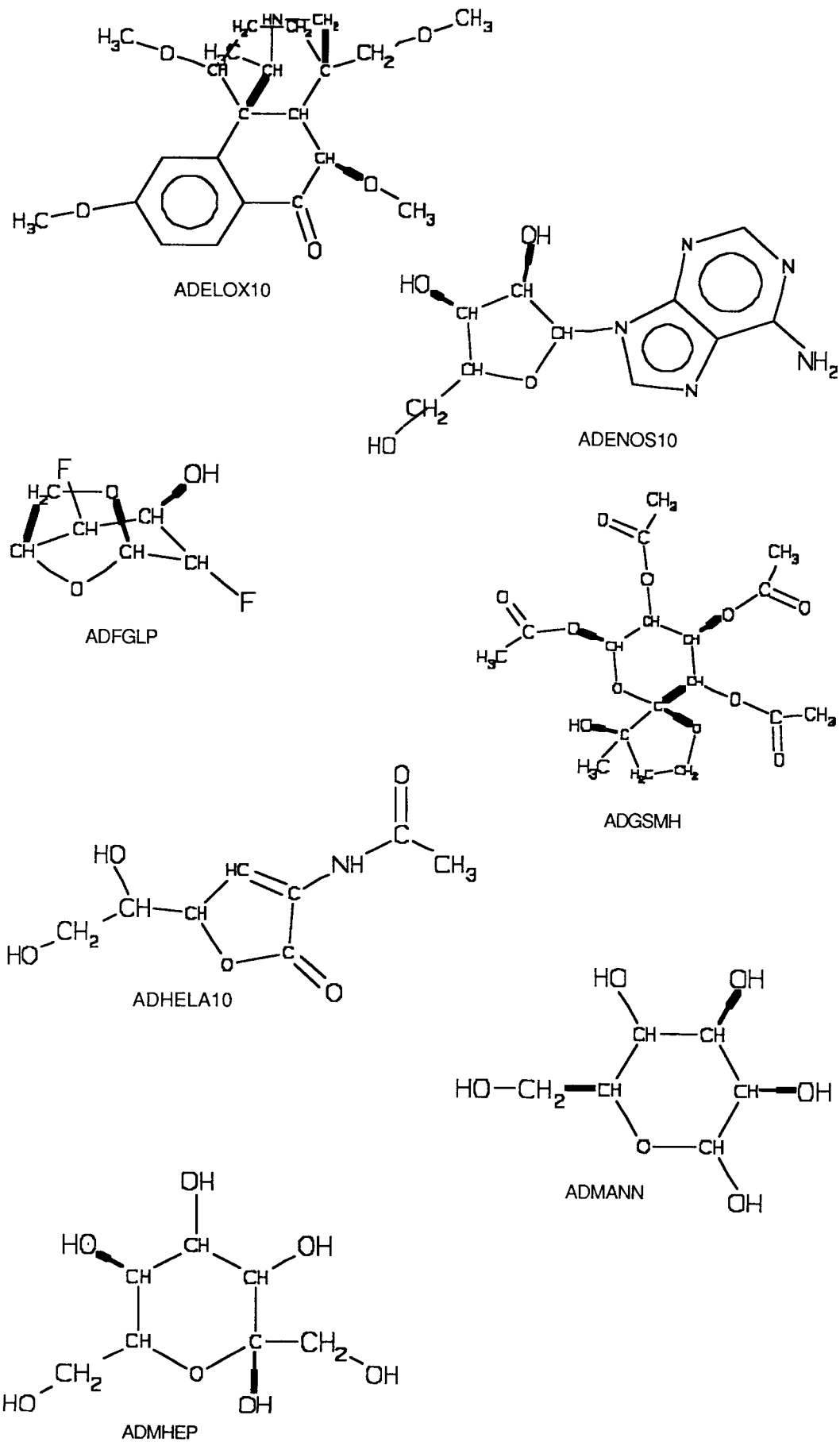
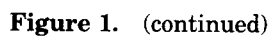


Figure 1. (continued)



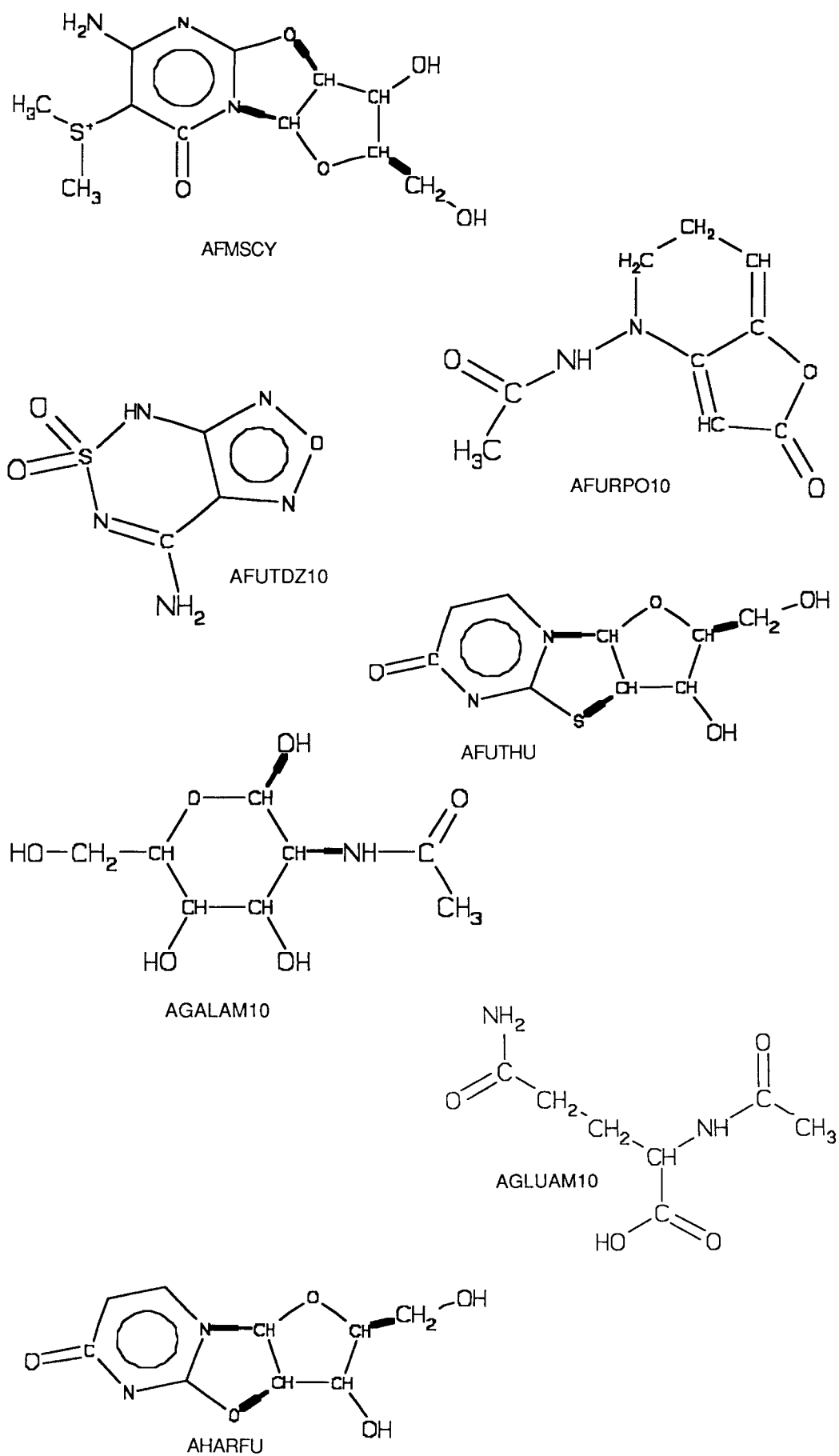


Figure 1. (continued)

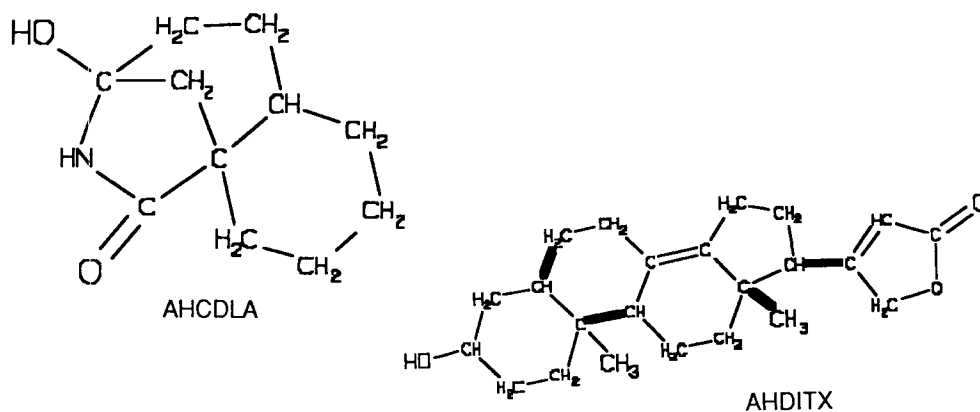


Figure 1. (continued)

The MM2 force field correctly calculates that the oxygen-methyl eclipsed conformation to be the preferred conformation, since MM2 can superimpose one, two, and threefold torsional barriers to stabilize this conformation.

Table X compares Tripos values with MM2 values for a series of stereochemical energy differences. With both Tripos and MM2 force fields, note that the energy difference between the *trans-syn-trans* (TST) and *trans-anti-trans* (TAT) forms of perhydroanthracene is overestimated. The TAT stereochemistry constrains the middle ring into the high energy twist-boat forms. Note that the overestimation of this value is the major contributor to the deviations for both the Tripos and MM2 force fields. The energy difference for axial and equatorial nitrocyclohexane could not be calculated with the normal MM2 program for the lack of parameters. The summary statistics indicate that

the Tripos force field is slightly worse than MM2 for the rest of the examples in Table X, but both the Tripos and MM2 calculated errors are quite small.

Table XI presents data for torsional barriers as calculated with the Tripos force field using constraints to keep the minimizer from altering the appropriate torsion angle, while the rest of the molecule was allowed to relax. The rms error of the barriers is 1.7 kcal/mol, the deviations skewed to overestimating the barriers, especially for carbon-heteroatom bonds.

Organic Molecules

The summary statistics for the minimization of the 76 diverse organic crystal structures are presented in Table XII. The average deviations in internal coordinates not involving hydrogen are presented along with the

Table VII. Statistics for differences of minimized structures of three peptides from the crystal structure (angstroms and degrees).

Feature	<i>n</i>	Ave.	TRIPOS ^a rms	Max	AMBER ^b rms	AMBER/OPLS ^c rms
Bonds	80	0.003	0.019	0.096	—	—
Angles	106	-0.410	2.06	-5.59	—	—
Torsions	114	0.280	6.82	17.9	—	—
Phi	18	-3.19	9.80	27.1	3.7 ^d	—
Psi	18	3.08	9.97	17.4	3.7 ^d	—
Omega	18	0.34	3.78	8.0	2.1	—
Coords ^e	3	0.15	0.16	0.20	0.09	0.10
H-Bonds ^f	6	-0.05	0.09	-0.21	0.12	0.11

^aTripos 5.2 force field, no electrostatics.

^bReference 6.

^cReference 5.

^dStatistics reported for rms deviations of phi and psi combined.

^erms deviation by atoms on a per compound basis, not atom by atom basis.

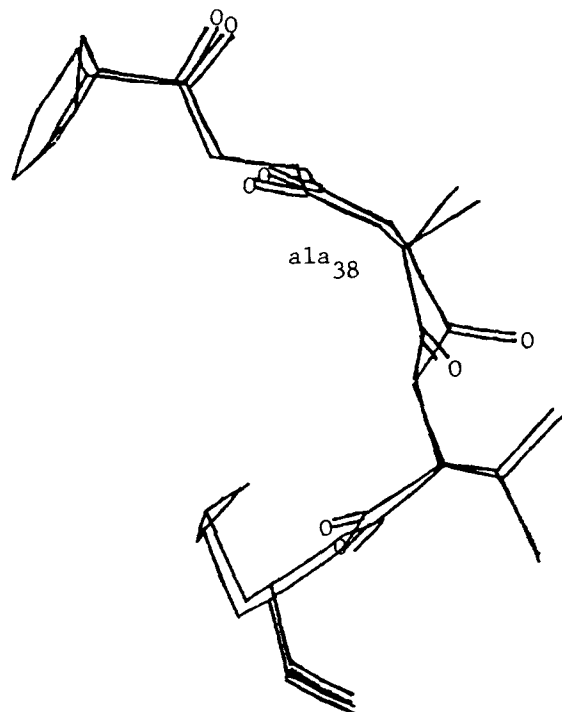
^fIntramolecular hydrogen bonds.

Table VIII. Statistics for difference between minimized and crystallographic structures of crambin (Angstroms and degrees).

Feature ^c	n	No crystal lattice TRIPOS 5.2 ^a			AMBER ^b		Crystal lattice AMBER ^c AMBER/O ^d	
		Ave.	rms	Max.	Ave.	rms	rms	rms
Bonds	337	-0.002	0.025	0.098	—	—	—	—
Valences	466	-0.018	2.97	-9.76	—	—	—	—
Torsions ^f	482	0.060	13.4	57.4	—	—	—	—
Torsions ^g	568	0.102	13.0	57.4	—	—	—	—
Phi	45	-1.508	17.3	50.7	-12	17	7.2	6.1
Psi	45	1.347	15.5	47.6	10	16	7.9	5.6
Omega	45	-0.0219	4.2	-13.8	—	—	4.1	4.6
Coord.	327	0.352	0.42	1.47	—	0.35	0.22	0.17

^aTripos 5.2 force field, no electrostatics, no crystal environment.^bReference 6, electrostatics with dielectric of 2.0, 10 angstrom cutoff, Jorgensen radii. Other conditions produced rms in coordinates from 0.24 to 1.18 angstroms.^cReference 6, Amber force field.^dReference 5, Amber/OPLS force field.^eGeometries involving hydrogens are excluded.^fTorsions involving only backbone atoms.^gAll torsions.

rms deviations, the coefficient of skewness, and the maximum movement for all molecules. The bond lengths from the crystal structures are, on the average, overestimated by 0.011 angstroms. This error is expected since crystal structures uncorrected for rigid body motion at room temperature may exhibit bond lengths that are 0.005 to 0.015

**Figure 2.** Superimposed drawings of crambin in the region of highest internal coordinate deviations, before and after energy minimization.

too short. The average error of angles is -0.128 degrees. When all torsions involving non-hydrogens are considered the average deviation is 0.069; when only ring atom torsions are considered, 0.045 degrees. The rms deviation of torsion angles decreases from 9.54 to 6.96 degrees when going from all torsions to only torsions involving atoms in rings. Thus, for the most part the minimization did not alter the conformation. The largest torsional deviation was 56.26 degrees, in the molecule ACPENC10 ((5S)-acetylpenem-3-carboxylate). In this case there was no conformational change; the fused five- and four-membered ring system angles containing sulfur and nitrogen were distorted. The largest bond angle deviation of 12.24 degrees also occurs in this molecule, as do other deviations in the range of 8 to 11 degrees. When this example is removed from the analysis the rms deviation of torsion angles falls to 6.5 degrees.

Even though all crystal structures had R factors of less than 0.05, some still had bond lengths that were questionable. For example the molecule AENLAN10 (3beta-acetoxy-6,7-epidithio-19norlanosta-5,7,9,11-tetraene) had a bond between sp^3 carbons of 1.426 angstroms. This accounted for the largest deviation in crystal and minimized bond lengths, 0.127 angstroms, since the force field lengthened the bond to a reasonable 1.555 angstroms. This and other mole-

Table IX. Comparison of experimental conformation and stereochemical energy differences with values calculated with the Tripos 5.2 and MM2 force fields. (kcal/mol)

System (higher energy conformer first)	Exp. ^a	MM2	TRIPOS
Butane, <i>gauche-anti</i>	0.8	0.9	0.6
2-methylbutane, <i>gauche-anti</i>	0.6	0.7 ^b	0.4
1-butene C—C—C=C torsion 120-0 degrees	0.2	-0.5	1.0
Butadiene torsion, 40-180 degrees	2.1	79.4	3.4
Methylcyclohexane <i>axial-equatorial</i>	1.9	1.6 ^b	1.4
Phenylcyclohexane <i>axial-equatorial</i>	3.0	1.9	4.2
Fluorocyclohexane <i>axial-equatorial</i>	0.2	0.2	0.0
Chlorocyclohexane <i>axial-equatorial</i>	0.4	0.4	0.2
Bromocyclohexane <i>axial-equatorial</i>	0.5	0.5	0.3
Nitrocyclohexane <i>axial-equatorial</i>	1.2	^c	0.6
Cyclohexanol <i>axial-equatorial</i>	0.5	0.9	0.0
1,4 dichlorocyclohexane <i>axial-equatorial</i>	0.2	0.3 ^b	0.4
Methyl ethyl ether <i>gauche-anti</i>	1.5	1.7 ^b	0.8
Propanal C—C—C=O angle 120-0 degrees	0.9	0.7	-0.1
Cyclohexane, <i>twist-boat-chair</i>	6.0	5.4	7.7
Indolizidine <i>cis-trans</i>	2.4	2.6 ^b	1.6
4,4,5-methyl-1,3-dioxolane <i>axial-equatorial</i>	1.3	0.8	0.7
Rms error		0.5 ^d	0.8

^aJ. P. Lowe, *Prog. Phys. Org. Chem.* **6**, 1 (1968); ref. 3.^bMM2 values from reference 3.^cNo MM2 parameters available for this molecule.^dvalue does not reflect the barrier for butadiene.

cules with smaller apparent defects were not eliminated from the set thus insuring that the test would not be biased towards molecules that the Tripos force field handled well.

In order to investigate the sensitivity of the resulting structures to the starting point the coordinates of all atoms were truncated to one decimal place and the resulting molecules re-minimized. Table XIII compares the structures which were minimized from the truncated starting coordinates to the original crystal coordinates. These deviations are

very similar to those shown in Table XII. Table XIV compares the minimized structures starting from the crystal structure with the minimized structures starting from the truncated coordinates. In general, the molecules minimized to the same ending structures. However, in a few cases the molecules minimized into slightly different conformations. The largest change, in comparing the optimized structures from the crystal starting point and the truncated-coordinate starting point, is a torsion rotation of 94 de-

Table X. Comparison of stereochemical energy differences in kcal/mol⁻¹ among experimental, MM2, and Tripos 5.2 values (kcal/mol).

System	Exp. ^a	MM2	TRIPOS
1,3-dimethylcyclobutane, <i>cis-trans</i>	-0.3	-0.2	-0.2
1,2-dimethylcyclohexane, <i>cis-trans</i>	1.9	1.6 ^b	1.1
1,3-dimethylcyclohexane, <i>cis-trans</i>	-2.0	-1.8	-1.4
1,4-dimethylcyclohexane, <i>cis-trans</i>	1.9	1.8	1.4
1,1,3,5-tetramethylcyclohexane, <i>cis-trans</i>	-3.7	-5.4	-4.8
Bicyclo[3.3.0]octane, <i>cis-trans</i>	-6.8	-7.0 ^b	-8.8
Hydrindane, <i>cis-trans</i>	1.0	1.2 ^b	0.5
Decalin, <i>cis-trans</i>	2.2	2.7 ^b	2.2
Perhydroanthracenes, relative to <i>trans-syn-trans</i>			
<i>Cis-trans</i>	2.8	2.6 ^b	2.2
<i>Trans-anti-trans</i>	4.1	5.9 ^b	9.1
<i>Cis-anti-cis</i>	5.6	5.6 ^b	4.8
<i>Cis-syn-cis</i>	8.7	8.1 ^b	8.2
rms error		0.8	1.7

^aJ. P. Lowe, *Prog. Phys. Org. Chem.*, **6**, 1 (1968).^bReference 3.

Table XI. Comparison of experimental torsional barriers^a and barriers calculated with the Tripos 5.2 force field. (kcal/mol)

Molecule	Experimental ^a	TRIPOS
Ethane	2.9	3.8
Propane	3.4	4.0
Butane, methyl-methyl eclipsed	6.0	5.9
Butane, methyl-hydrogen eclipsed	3.4	5.1
Isobutane	3.9	4.5
Neopentane	4.7	4.7
Methylsilane	1.7	1.9 ^b
Methylamine	2.0	2.8
Dimethylamine	3.6	4.9
Fluoroethane	3.3	3.8
Chloroethane	3.7	4.0
Bromoethane	3.7	4.0 ^b
Iodoethane	3.2	4.1 ^b
Methanol	1.1	4.1
Dimethyl ether	2.7	4.2
rms error		1.13

^aJ. P. Lowe, *Prog. Phys. Org. Chem.*, **6**, 1 (1968).^bTorsional parameter automatically estimated using procedures noted in Tables I–VI.**Table XII.** Summary statistics for difference between crystallographic and Tripos 5.2 energy minimized geometries for 76 organic molecules (angstroms and degrees).

Feature	<i>n</i>	Average	Skew ^a	rms	Maximum
Bond lengths	1483	0.011	−612.0	0.025	0.127
Bond angles	2174	−0.128	387.8	2.50	12.24
Torsion angles ^b	2251	0.045	−364.6	6.96	56.26
Torsion angles ^c	2888	0.069	−368.6	9.54	56.26
Coordinates	76	0.192		0.25	1.088

^acoefficient of skewness^bStatistics only for torsions in rings and directly connected to rings.^cStatistics for all torsion angles.**Table XIII.** Summary statistics for difference between crystallographic and Tripos 5.2 energy minimized geometries for 76 organic molecules, with the starting coordinates truncated to one decimal place (angstroms and degrees).

Feature	<i>n</i>	Average	Skew ^a	rms	Maximum
Bond lengths	1483	0.011	−595.4	0.025	0.127
Bond angles	2174	−0.132	392.8	2.50	12.24
Torsion angles ^b	2251	0.045	−357.1	6.95	56.01
Torsion angles ^c	2888	0.140	−887.7	9.65	65.87
Coordinates	76	0.191		0.25	1.123

^acoefficient of skewness^bStatistics only for torsions in rings and directly connected to rings.^cStatistics for all torsion angles.**Table XIV.** Summary statistics for difference between Tripos 5.2 energy minimized geometries for 76 organic molecules, with crystal and truncated starting coordinates (angstroms and degrees).

Feature	<i>n</i>	Average	Skew ^a	rms	Maximum
Bond lengths	1483	0.000	−17752.9	0.003	0.073
Bond angles	2174	−0.004	137.7	0.24	5.83
Torsion angles ^b	2251	0.010	−2489.1	0.570	4.97
Torsion angles ^c	2888	0.071	−89698.3	2.65	94.40
Coordinates	76	0.019		0.03	0.543

^aCoefficient of skewness^bStatistics only for torsions in rings and directly connected to rings.^cStatistics for all torsion angles.

grees. This occurs in an acetyl side chain of ACONTN10 (6- α -acetoxy-1 α -hydroxy-8 β -methoxy-4 β -methyl delphisine hydroiodide) attached to the five membered ring. This and similar conformational changes are not surprising, since the crystal structure represents only one of many energy minima. If the analysis considers only changes in rings and torsions to atoms directly connected to rings, the maximum difference between bonded torsion angles in the sets of minimized structures is 4.97 degrees.

The overall rms differences in atom positions after least-squares superposition for each of the 76 compounds is summarized in Table XV. The worst structure, as measured by the rms movement of atoms after least-squares superposition is ADMOPM (adenosine-5'-*O*-methylphosphate), which has an rms heavy-atom movement of 1.088, nearly twice the next highest value. However, this molecule consists of two-ring systems joined by a rotationally flexible bond, which is the source of the rms deviations due to movement of the ring moieties. The molecule with the largest deviation of ring atoms is ACRAMS (4'-(acridin-9-ylamine)methanesulfonanilide hydrochloride). This molecule consists of two-ring systems joined by a nitrogen; while the geometries of the ring systems are essentially unchanged, torsion about the nitrogen results in large movements of the ring systems and thus large rms movements with only small internal coordinate deviations. This is an example of the weakness of rms movement as a lone measure of quality of force field geometries.

The largest deviations in individual bond lengths are associated either with questionable experimental data, as in AENLAN10, or conjugated heterocyclic rings as in the five-membered ring of purines. While the crystal-line environments of the molecules certainly influenced the structures it is not practical or instructive to comment on how the packing

affects each molecule individually. Figure 3 superimposes the crystal and calculated structures of ACRAMS, the molecule with the largest rms movement of ring atoms.

The data indicates that the Tripos force field has a systematic error in overestimating the bond lengths of atoms in small rings; these bonds are usually shortened by electronic effects. The largest valence angle deviations are all in four-membered rings. This small ring problem has been addressed by Allinger by adding atom types specific for three- and four-membered rings to the MM2 force field.⁴

CONCLUSIONS

The Tripos force field has been shown to produce molecular geometries close to those of crystal structures for a diverse selection of molecules. Comparison of rms movements of heavy atoms after least-squares superposition and analysis of deviations of the internal coordinates show that both must be examined to measure the relationships between starting and energy minimized structures since the rms movement can be a misleading indicator of structural changes.

Many problems require the structure of a ligand bound to a protein. These structures are difficult to model as there have been few parameter sets suitable for modeling both the protein and the arbitrary ligand structure. The results presented here suggest that the Tripos force field is well suited for overcoming these difficulties. The statistical analysis has shown that the Tripos force field performs with even quality with both peptides and organic molecules. While the performance with these two classes of molecules is not equal to the best specialized force fields, the performance overall is superior.

The authors thank the scientific staff of Tripos Associates for their comments and suggestions, and the Cambridge Crystallographic Data Center for access to the structural database.

Table XV. Rms movements in angstroms for heavy atoms, and for ring atoms only for 76 organic molecules from the Cambridge Structural Database.

Compound		rms ^a	rms ^b
AAXTHP	1,4,6-Tri- <i>O</i> -acetyl-2-(<i>N</i> -acetylacetamido)-2,3-dideoxy- α -D-threo-hex-2-enopyranose	0.500	0.064
ABAXES	2- α -Bromo-17- β -acetoxy-9-methyl-5- α , 9- β , 10- α -estran-3-one	0.123	0.072
ABBUM010	16- α , 17-Butanomorphan-3-ol	0.124	0.122
ABINOR02	β -DL-Arabinose	0.125	0.036
ABINOS01	β -L-Arabinose	0.036	0.023
ABTOET	Ethyl 3,7-anhydro-6,8- <i>O</i> -benzylidene-4-deoxy-2-ethylenedithio-D-talo-2-octulosonate	0.351	0.262
ABZTCX	3-Dimethylamino-4,4-dimethyl-5,6-dihydro-4H-1,2,5-benzothiazocin-6-one 1,1-dioxide	0.480	0.170
ACADOS	3'- <i>O</i> -Acetyladenosine	0.264	0.125
ACAFLR	2-Acetylaminofluorene	0.144	0.040
ACANIL01	Acetanilide	0.250	0.011
ACARAP	1,2,3,4-Tetra- <i>O</i> -acetyl- α -D-arabinopyranose	0.463	0.055
ACBNZA01	<i>o</i> -Acetamidobenzamide	0.155	0.033
ACBUOL	2-(2-Hydroxy-3-isopropylamino-propoxy)-5-butyrylamino-acetophenone hydrochloride	0.422	0.025
ACCITR10	Acetylcitrin	0.091	0.074
ACDXUR	α -5-Acetyl-2'-deoxyuridine	0.246	0.120
ACENAP03	1,2-Dihydro-acenaphthylene	0.020	0.020
ACFPCH	<i>cis</i> -1-Acetamido-2-fluoro-1-phenylcyclohexane	0.277	0.184
ACFUCN	<i>N</i> -Acetyl-furanomycin	0.425	0.122
ACGLSP	Methyl-2,3,4,5-tetra- <i>O</i> -acetyl- α -D-galactoseptanoside	0.448	0.056
ACGLUA11	<i>N</i> -Acetyl- α -D-glucosamine	0.092	0.030
ACHGAL	2,3-Di- <i>O</i> -acetyl-1,6-anhydro- β -D-galactopyranose	0.297	0.057
ACHIST20	L- <i>N</i> -Acetylhistidine monohydrate	0.117	0.020
ACHNAP10	<i>trans</i> -4a-Acetoxy-8a-chloro-1,4,4a,5,8,8a-hexahydronaphthalene	0.055	0.036
ACHTAR10	(Acetoxyethyl)-trimethylammonium hydrogen (+ -)-tartrate	0.139	— ^c
ACIMDC	Acetamidinium chloride	0.019	— ^c
ACINDN	2-Acetyl-indan-1,3-dione	0.072	0.039
ACINST	DL-1,4,5,6-Tetra- <i>O</i> -acetyl-3-chloro-2-C-(chloromethyl)-epi-inositol	0.548	0.048
ACKYNU	<i>N</i> -Acetyl-kynurenine	0.420	0.031
ACMBPN	2-Amino- <i>N</i> -(3-dichloromethyl-3,4,4a,5,6,7-hexahydro-5,6,8-trihydroxy-3-methyl-1-oxo-1H-2-benzopyran-4-yl)-propanamide hydrobromide dihydrate	0.151	0.060
ACMEBZ	2-Acetoxy-3-methylbenzoic acid	0.175	0.020
ACMTDE	<i>N</i> -Acetyl-D,L-methionine-diethylamide	0.175	— ^c
ACNORT	<i>N,O</i> -Diacyl-4-hydroxy-nornantenine	0.215	0.118
ACNPAC10	Acenaphthylene-1-carboxylic acid	0.047	0.025
ACNPEC	(<i>R,R</i>)- <i>N</i> -Acetyl-S-(2-nitro-1-phenylethyl)-L-cysteine	0.931	0.028
ACONTN10	6- α -Acetoxy-1- α -hydroxy-8- β -methoxy-4- β -methyl-delphisine hydroiodide	0.270	0.063
ACPENC10	(5 <i>S</i>)-Acetonyl-penem-3-carboxylate	0.511	0.201
ACPPCA	(+ -)-(E)- <i>N</i> -Acetyl-piperidine-2-carboxylic acid	0.140	0.040
ACPRET03	21-Acetoxy-17- α -hydroxy-pregn-4-ene-3,11,20-trione	0.202	0.073
ACPYNS	Methyl-3,4-di- <i>O</i> -acetyl-2,6-anhydro- α -D-altropyranoside	0.280	0.050
ACRAMS	4'-(Acridin-9-ylamine)methanesulfonanilide hydrochloride	0.515	0.447
ACSALA01	2-(Acetyloxy)-benzoic acid	0.075	0.018
ACSESO10	(10 <i>S</i>)-17- β -Acetoxy-3,10-cyclo-3,4-seco-4,9(11)-estradien-1-one	0.208	0.114
ACTAND	α -Acetylthio-5- α -androstan-17-one	0.257	0.070
ACTHBZ	<i>N,N'</i> -Diacyl-3-methylthiobenzidine	0.466	0.220
ACTHCP	7-Thia-1,3-diazabicyclo(3.3.0)octa-2,4-dione	0.093	0.093
ACTOLD	<i>p</i> -Acetotoluidine	0.231	0.025
ACTYSN	<i>N</i> -Acetyl-L-tyrosine	0.448	0.012
ACURID	β -5-Acetyl-2'-deoxyuridine	0.268	0.134
ACVCHO	(+ -)-(1 <i>R</i> , 2 <i>R</i> , 3 <i>R</i>)-3-Acetyl-2-vinylcyclohexan-1-ol	0.071	0.028
ACXMOL	(-)-1- <i>O</i> -Acetyl-xylomollin	0.412	0.088
ACXMPR	3 <i>R</i> -(1' <i>S</i>)-Aminocarboxymethyl-2-pyrrolidone-5(<i>S</i>)-carboxylic acid	0.203	0.033
ACYGLY11	<i>N</i> -Acetyl-glycine	0.048	— ^c
ACYTID	α -Cytidine	0.120	0.067
ADELOX10	Delphinine aromatic acid oxalate	0.165	0.167
ADENOS10	Adenosine	0.109	0.080
ADFGLP	1,6-Anhydro-2,4-deoxy-2,4-difluoro- β -D-glucopyranose	0.163	0.079

Table XV. (continued)

Compound		rms ^a	rms ^b
ADGSMH	<i>trans</i> -2,3,4,6-Tetra- <i>O</i> -acetyl-1-deoxy-D-glucopyranoside-1,2'-spiro(3'-methyl-3'-hydroxy-tetrahydrofuran)	0.382	0.082
ADHELA10	2-Acetamido-2,3-dideoxy-D-threo-hex-2-enono-1,4-lactone	0.107	0.043
ADMANN	alpha-D-Mannopyranose	0.103	0.029
ADMHEP	alpha-D-Manno-2-heptulose	0.113	0.034
ADMINA	3-Amino-1,4-dimethyl-5H-pyrido(4,3-b) indole acetate	0.066	0.044
ADMOPM	Adenosine-5'- <i>O</i> -methylphosphate methanol solvate	1.088	0.333
ADRTAR	(-)-Adrenaline hydrogen(+)-tartrate	0.094	0.015
ADYPNL	alpha-Dypnopinacoline	0.314	0.321
AEBDOD10	DL-4-beta,10-beta-Dimethyl-6-ethylamino-4-hydroxycarbonyl-2,3,5-beta,10-tetrahydro-phenanthr-1-one lactam	0.076	0.054
AENLAN10	3-beta-Acetoxy-6,7-epidithio-19-norlanosta-5,7,9,11-tetraene	0.466	0.158
AFCYDP	2,2'-Anhydro-1-beta-D-arabinofuranosyl-cytosine-3',5'-diphosphate	0.527	0.173
AFMSCY	2,2'-Anhydro-1-beta-D-arabinofuranosyl-5-dimethylsulfonio-6-oxo-cytosine chloride	0.209	0.117
AFURPO10	<i>N</i> -Acetyl-5,6-dihydrofuro(2,3-b) pyrid-2-one	0.094	0.075
AFUTDZ10	7-Amino-4H-furazano(3,4-d)-1,2,6-thiadiazine-1,1-dioxide	0.209	0.109
AFUTHU	2,2'-Anhydro-1-beta-D-arabinofuranosyl-2-thio-uracil	0.210	0.112
AGALAM10	<i>N</i> -Acetyl-alpha-D-galactosamine	0.071	0.026
AGLUAM10	<i>N</i> -Acetyl-L-glutamine	0.297	— ^c
AHARFU	2,2'-Anhydro-1-beta-D-arabinofuranosyl-uracil	0.375	0.251
AHCDLA	3-Amino-3-hydroxy- <i>trans</i> -bicyclo(4.4.0) decane-1-carboxylic acid lactam	0.064	0.040
AHDITX	delta-8,14-Anhydro-digitoxigenin	0.153	0.139
Mean		0.245	0.090

^aAll heavy atoms.^bHeavy atoms in rings only.^cMolecule contained no ring systems.

References

1. T. Wipke and M. Hahn, Evans and Sutherland, personal communication.
2. R. Pearlman, University of Texas, Concord program (1987).
3. E. M. Engler, J. D. Andose, and P. v. R. Schleyer, *J. Am. Chem. Soc.*, **95**, 8005 (1973).
4. U. Burkert and N. L. Allinger, *Molecular Mechanics*, ACS Monograph No. 177, American Chemical Society, Washington, DC, 1982; and references within.
5. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
6. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984); *J. Comp. Chem.*, **2**, 287 (1981).
7. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
8. B. R. Gelin and M. Karplus, *Biochemistry*, **18**, 1256 (1979); B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).
9. S. Lifson, A. T. Hagler and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979); A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
10. A. Vedani, *J. Comp. Chem.*, **9**, 269 (1988).
11. Tripos Associates, 1699 S. Hanley Road, Suite 303, St. Louis, MO, 63144, SYBYL 5.2 (1988).
12. P. G. Jones, *Chem. Soc. Rev.*, **13**, 157 (1984).
13. M. Whitlow and M. M. Teeter, *J. Am. Chem. Soc.*, **108**, 7163 (1986).
14. D. Hall and N. Pavitt, *J. Comp. Chem.*, **5**, 441 (1984).
15. J. G. Vinter, A. Davis, and M. R. Saunders, *J. Comp.-Aided Mol. Des.*, **1**, 31 (1987).
16. D. N. J. White and M. J. Bovill, *J. Chem. Soc., Perkin II*, **12**, 1610 (1977).
17. The key to the references in the tables is as follows:

*	Adapted from reference 15.
JL_EST	Jan Labanowski, March 1983
WHITE_77	White, D. N. J., <i>JCS Perkin II</i> , (1977) 1610.
EXP	Experimental value.
WHITE_75	White, D. N. J., <i>JCS Perkin II</i> , (1975) 43.
TRIPOS_86	January, 1986, Tripos

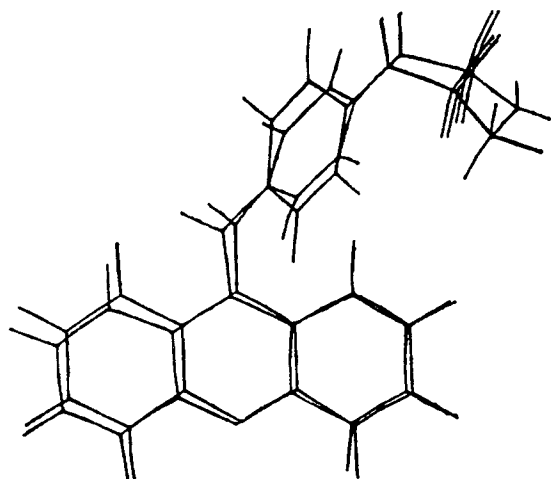


Figure 3. Superimposed drawings of ACRAMS before and after energy minimization.

- | | | |
|-----------|--|---|
| CRC_56 | Jan Labanowski, April 1983; <i>CRC Handbook of Chemistry and Physics</i> , CRC Press, 56th ed. p. F-213. | 19. W. A. Hendrickson and M. M. Teeter, <i>Nature</i> , 290 , 107 (1981). |
| TRIPOS_85 | Tripos Scientific Staff, June 1985. | 20. M. B. Hossain and D. J. van der Helm, <i>J. Am. Chem. Soc.</i> , 100 , 5191 (1978). |
| MC_88 | Matthew Clark, Tripos, 1988. | 21. I. L. Karle, J. W. Gibson, and J. J. Karle, <i>J. Am. Chem. Soc.</i> , 92 , 3755 (1980). |
| M7/22/88 | This work | 22. F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, <i>Acta. Crystallogr., Sect B</i> , B35 , 2331 (1979). |
18. N. Allinger and Y. Yuh, QPCE 395; this version includes a Steffensen iteration scheme for the minimizer: Kalman, Barry, Washington University Technical Memo 46, May 1982.
23. T. Liljefors, J. C. Tai, S. Li, and N. L. Allinger, *J. Comp. Chem.*, **8**, 1051, (1985).