

## On the suitability of semiempirical calculations as sources of force field parameters

C. Alemán and M. Orozco\*

*Departament de Bioquímica i Fisiologia, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1,  
08028 Barcelona, Spain*

Received 10 October 1991

Accepted 23 March 1992

*Key words.* Force fields; Force field parametrization; Molecular mechanics; Semiempirical methods; Molecular modelling

---

### SUMMARY

The suitability of Dewar's Hamiltonians as a source of bonded force field parameters is explored from the comparison analysis between up to 270 semiempirically derived force field parameters and experimentally derived values reported in some of the most popular force fields. From the statistical analysis of the results, some general conclusions about the semiempirical parametrization are formulated.

---

### INTRODUCTION

Dewar's semiempirical methods, and force field derived techniques are, because of their low computational cost, the most suitable tools to perform analysis of medium or even large molecular structures which cannot be studied by *ab initio* methods [1–12]. In this respect, molecular mechanical (MM) and dynamical methods (MD) have become useful tools for the study of the interactions of biological macromolecules [12]. All the MM and MD methods are based on the assumption that the energy of a molecular system can be described as the addition of two contributions: (i) the bonded term, which includes the energy of stretching, bending, proper and improper torsion; and (ii) the nonbonded term, where the electrostatic, hydrogen bond and van der Waals energy are considered. The set of analytical expressions used to represent these interactions is known as the force field. To date, several of them have been developed and are implemented in different computational programs [5–16], that are currently used in a wide range of chemical and biochemical problems.

All the analytical expressions used to compute both the bonded and the nonbonded terms in-

---

\* To whom correspondence should be addressed.

clude several parameters, the values of which are selected so that results derived from force field computations reproduce experimental data. Therefore, the study of any molecular system by MM or MD techniques implies the previous knowledge of all the parameters describing such a system. This leads to one of the most serious shortcomings of MM and MD methods, since even though MM and MD programs include large parameter data sets, several parameters describing a molecular system are often not available, which would preclude its study by MM or MD methods.

The determination of force field parameters from experimental data is a long and tedious process, for which a great amount of experimental data are needed. This explains the considerable research effort directed towards the development of strategies to determine force field parameters from quantum mechanical calculations [1,17–28]. In this respect, some strategies have recently been reported to compute the nonbonded force field parameters, i.e. the van der Waals and hydrogen bond constants [18,19], as well as the atomic charges and bond dipoles [24–28]. Moreover, other strategies have been developed to calculate the bonded force field parameters [17,19–23], among which Hagler's [17,19,20] and Hopfinger's [21] procedures are probably the most widely used.

Very recently, a new strategy to determine bonded force field parameters has been developed in our laboratory [1], and implemented in the computer program PAPQMD. Our strategy, described in detail elsewhere [1], is based on the rigorous fitting of the quantum and the molecular mechanical energies related to the perturbation of the equilibrium geometry. The PAPQMD program was created to allow the user a fast and suitable evaluation of the bonded force field parameters of the large molecules of biochemical and pharmacological interest.

Since PAPQMD determines the force field parameters from quantum mechanical data, it is obvious that the reliability of the parameters will mainly depend on three different factors:

- (a) The suitability of the quantum mechanical method used as reference to represent the energy variation profile derived from the perturbation of the molecular geometry. This point makes it necessary to examine, not only the reliability of the Hamiltonian (semiempirical or *ab initio*), but also the suitability of the wave-function quality (RHF, UHF or multiconfigurational).
- (b) The suitability of the force field to fit the energy variation profile obtained when the molecular geometry is perturbed. At this point, it must be emphasized that the suitability of the harmonic expressions used to compute most of the bonded energy terms in the force fields decreases as the perturbation in the molecular geometry increases, and may be lost altogether for highly perturbed molecules.
- (c) The statistical quality of the fitting of the quantum and molecular mechanical energies.

In this paper we present a study on the reliability of the semiempirical methods as reference computations for the parametrization procedure. The importance is obvious when considering that rigorous *ab initio* parametrization is not possible for most of the biologically relevant molecules, because of the computational cost of *ab initio* wave-function evaluation, even if it is determined at the RHF level, and low sophisticated basis sets are used. Thus, in practice, *ab initio* derived force field parameters are usually obtained from calculations on small molecules, which are very different to those of real interest. As a consequence, even the *ab initio* methods provide excellent force field parameters [17–20,22,23,29,30], since such parameters are obtained in a very different molecular environment, their ability to describe the large molecule which will be studied by molecular mechanical dynamical techniques can be poor. Here, our attention is focused on the study of

the reliability of the semiempirical RHF wave function computed from MINDO/3 [2], MNDO [3], and AM1 [4] Hamiltonians to correctly represent the molecular characteristics in its perturbed geometries. Moreover, the range of validity of the equations used to compute the molecular mechanical energy is also examined. Finally, in the light of the results obtained, and taking into consideration statistical criteria, a general procedure for fitting semiempirical quantum mechanical and molecular mechanical energies is suggested.

## METHODS

Semiempirical quantum mechanical computations were carried out using the standard Dewar's MINDO/3 [2], MNDO [3], and AM1 Hamiltonians [4] and considering both the RHF and UHF wave functions. Molecular mechanical energies are computed considering a general force field (see Eq. 1):

$$E_{FF} = E_{str} + E_{bnd} + E_{tor} + E_{itor} + E_{VW} + E_{ele} + E_{HB} \quad (1)$$

where  $E_{str}$ ,  $E_{bnd}$ ,  $E_{tor}$ ,  $E_{itor}$ ,  $E_{VW}$ ,  $E_{ele}$  and  $E_{HB}$  are the stretching, bending, torsion, improper torsion, van der Waals, electrostatic and hydrogen bond energies.

The stretching and bending energies are computed using a usual quadratic expression, the improper torsional term is calculated following the four widely used strategies: Allinger's [5,6], Kollman's [7,8], Karplus' [9,10], and TRIPOS [11]; the proper torsion energy is computed from a three-term Fourier expression. With respect to the nonbonded terms, the van der Waals energy is computed using the  $r^{-12} - r^{-6}$  expression, the electrostatic energy is determined by a Coulombic expression considering high-quality semiempirical electrostatic charges [28,31]; and finally, the hydrogen bond energy is computed considering a  $r^{-12} - r^{-10}$  expression. Kollman's parameters [7] for van der Waals and hydrogen bond interactions are used.

The quantum mechanical energy profiles corresponding to the stretching, bending, proper and improper torsions of different molecules were calculated in four steps: (a) the equilibrium geometry of the molecule is calculated, (b) one geometrical variable ( $-i$ ) (for instance a bond length) is monotonously modified, and the quantum mechanical energy ( $E_{QM}$ ) is obtained from a semiempirical calculation with full geometry optimization (with the obvious exception of the  $-i$  variable); (c) a corrected quantum mechanical energy ( $E_{QM}^{cor}$ ) is computed (see Eq. 1) following the procedure described in a previous paper [1]; (d) finally, the force field equation is fitted to the corrected quantum mechanical energy, as described in [1].

$$E_{QM}^{cor}(i) = E_{QM}(i) - E_{per}(i) \quad (2)$$

where  $E_{per}$  is the 'perturbation' energy, which includes the energy contribution due to the modification of all the geometrical variables of the molecule, obtained when the geometrical variable  $-i$  is monotonously varied.

The bonded force field parameters were determined by using the PAPQMD program. All the bonded force field parameters of the different molecules were simultaneously determined. The fitting was considered as satisfactory when the Pearson correlation coefficients took values greater than 0.95 and the largest difference between two successive iterations was less than 1%.

In order to determine the reliability of semiempirically determined force fields, they were compared with the empirically derived parameters used in four of the most widely used force fields: Allinger's MM2 [5,6] (for conjugated systems MMP2 derived parameters were considered instead of standard MM2), Kollman's [7,8], Karplus' [9,10], and TRIPOS [11]. The statistical evaluation of the similarity between semiempirical and empirical force parameters was carried out considering: (i) the Pearson correlation coefficient, (ii) the root mean square deviation (RMS, see Eq. 1), and (iii) the relative root mean square deviation ( $\text{RMS}^{\text{rel}}$ , see Eq. 2). For proper torsions, an additional parameter ( $\text{RMS}^{\text{path}}$ , see Eq. 3) was defined in order to determine the suitability of the torsion term in the whole range of dihedral angles.

$$\text{RMS} = \sqrt{\frac{\sum_i (x_{ij} - x_{ik})^2}{n}} \quad (3)$$

$$\text{RMS}^{\text{rel}} = \sqrt{\frac{\sum_i (x_{ij} - x_{ik})^2}{\sum_i x_{ij}^2}} \quad (4)$$

$$\text{RMS}^{\text{path}} = \frac{\sum_i \sqrt{\frac{\sum_m (x_{ijm} - x_{ikm})^2}{p}}}{n} \quad (5)$$

where:  $i$  stands for a geometrical variable;

$j$  and  $k$  represent the methods to be compared;

$n$  is the number of samples;

$m$  stands for a value of the dihedral angle defining a bond rotation;

$p$  is the number of points considered for comparison in bond rotation.

All the semiempirical calculations were carried out by using a locally modified [32] version of the MOPAC computer program [33]. The geometry of all the 'perturbed' molecular geometries was fully optimized (with the obvious exception of the geometrical parameter which is monotonously 'perturbed'). All the computations were performed on the IBM-3090/150 at the Barcelona University Computer Center.

## RESULTS AND DISCUSSION

As a previous step to the analysis of the three Dewar's Hamiltonians, we studied the ability of the RHF wave function to represent perturbed geometries, as well as the suitability of force field equations to represent such perturbed structures. From these studies we determined the optimum ranges of perturbation in the different geometrical parameters to be considered in the quantum mechanical calculations used as reference in the parametrization process.

With respect to the stretching, we studied several single, double and triple bonds. Each bond was perturbed from its equilibrium value, and the corresponding stretching profiles were computed at both RHF and UHF levels (these profiles are available upon request from the authors). The

RHF wave function was assumed to be incorrect when it was different to the UHF one (that is when birradical structures were obtained). Such a discrepancy occurs at around 0.6 Å (single bonds) and 0.2 Å (double and triple bonds). We also tested the suitability of quadratic expressions to represent perturbed bonds; for this purpose we analyzed the maximum range of perturbation in which the quadratic expression was able to reproduce within an error of 1 kcal/mol (single), 3 kcal/mol (double) and 5 kcal/mol (triple) quantum mechanical data (statistical and graphic data about such analysis are available upon request from the authors). We found that only a perturbation of  $\pm 0.1$  Å from the equilibrium bond length can be well represented by quadratic expressions. Accordingly, all the determinations of stretching parameters presented below were performed considering such a perturbation range, where both the RHF wave function and the classical quadratic expression are suitable.

A similar study was performed for the bending, and as expected, the RHF wave function was quite accurate for our purposes even for large perturbations. Moreover, the quadratic expression was able to reproduce quantum mechanical data for perturbations in the range of  $\pm 20^\circ$  with an error less than 1 kcal/mol (graphical and statistical data are available upon request from the authors). Such a perturbation range of the bond angles was considered for all the bending parametrization presented below.

The improper torsion of planar systems was also properly represented by the RHF wave function over a wide range of perturbations. Moreover, quantum mechanical data was well fitted (error less than 3 kcal/mol) to the four classical formalisms [5,7,9,11] for a perturbation range defined by elevations of up to 0.3 Å of the Sp<sup>2</sup> atom from the plane defined by the three atoms directly bound to it (statistical and graphical data are available upon request from the authors). Such a perturbation range was considered in all the improper torsion parametrizations presented below.

From our calculations it is stated that the rotation around single bonds can be well parametrized from RHF calculations, even though it is recommended to perform the whole rotation in order to verify the nonexistence of hysteresis phenomena. On the contrary, rotations around double bonds need the use of UHF wave functions, which leads to an underestimation of the torsional barrier. Such an underestimation can be corrected by using the linear method proposed by Bofill et al. [34]. The rotation of the conjugated bond was not considered in our studies, because of the well-known shortcomings of Dewar's semiempirical methods to properly describe them.

Once the optimum range of perturbation was determined (to be considered in the parametrization, see preceding paragraphs), we explored the suitability of MINDO/3 [2], MNDO [3] and AM1 [4] Hamiltonians. In spite of the fact that all the force field parameters were determined simultaneously, results are presented in four subsections corresponding to stretching, bending, and proper and improper torsion, in order to clarify the discussion.

### *Stretching*

The ability of Dewar's semiempirical methods to reproduce suitable stretching parameters was explored by the study of 35 different stretchings. Force parameters and equilibrium lengths computed from MINDO/3, MNDO and AM1 data are compared in Tables 1 (bond lengths) and 2 (force parameters) with the experimentally derived values included in several of the most widely used force fields: Allinger's [5,6], Kollman's [7,8], Karplus' [9,10] and TRIPOS [11]. The statistical

study of the relationships between the different stretching parameters is displayed in Tables 3A (bond lengths) and 3B (force parameters).

The comparison of equilibrium lengths determined from semiempirical methods with respect to values included in empirical force fields (see Tables 1 and 3A) points to the excellent performance of all the semiempirical methods. Thus, a RMS deviation of only 0.03–0.04 Å is determined be-

TABLE I  
EQUILIBRIUM LENGTHS FOR SEVERAL BONDS DETERMINED FROM SEMIEMPIRICAL AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, KARPLUS', ALLINGER'S AND TRIPOS/5 FORCE FIELDS. VALUES ARE IN Å

Bond	Molecule	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
C–O	H <sub>2</sub> CO	1.23	1.22	1.18		1.23	1.208	1.220
H–C	H <sub>2</sub> CO	1.11	1.11	1.12	1.08	1.08	1.113	1.081
C–C	H <sub>3</sub> C–CH <sub>3</sub>	1.50	1.52	1.48	1.53		1.523	1.540
H–C	H <sub>3</sub> C–CH <sub>3</sub>	1.12	1.11	1.11	1.09	1.08	1.100	1.100
N–H	H <sub>3</sub> C–NH <sub>2</sub>	1.00	1.01	1.03	1.01	0.98	1.020	1.080
N–C	H <sub>3</sub> C–NH <sub>2</sub>	1.43	1.46	1.40	1.47	1.45	1.440	1.470
C–C	H <sub>2</sub> C–CH <sub>2</sub>	1.33	1.33	1.31			1.337	1.335
H–C	H <sub>2</sub> C–CH <sub>2</sub>	1.10	1.09	1.10		1.09	1.010	1.089
C–N	H–CN	1.16	1.16	1.15			1.158	1.158
H–C	H–CN	1.07	1.05	1.09			1.090	1.056
C–C	HC–CH	1.19	1.16	1.20			1.212	1.204
C–F	H <sub>3</sub> C–F	1.37	1.35	1.37			1.392	1.360
C–C	C <sub>6</sub> H <sub>6</sub>	1.39	1.41	1.41	1.40	1.38	1.392	1.395
C–H	C <sub>6</sub> H <sub>6</sub>	1.10	1.09	1.10	1.08			1.084
C–H	CH <sub>3</sub> –OH	1.41	1.39	1.34	1.41	1.42	1.402	1.430
C–H	CH <sub>3</sub> –OH	0.96	0.95	0.95	0.96	0.96	0.942	0.950
C–C	HOC–CH <sub>2</sub> – NH <sub>2</sub>	1.51	1.53	1.50	1.52	1.53	1.509	1.501
C–N	HOC–NH <sub>2</sub>	1.37	1.41	1.33	1.33	1.33	1.368	1.345
C–O	HOC–NH <sub>2</sub>	1.24	1.22	1.21	1.23	1.23	1.223	1.220
C–O	O–(CH <sub>3</sub> ) <sub>2</sub>	1.42	1.40	1.34	1.41	1.42	1.402	1.430
C–S	S–(CH <sub>3</sub> ) <sub>2</sub>	1.75	1.72	1.75	1.81		1.815	1.817
O–O	(O–CH <sub>3</sub> ) <sub>2</sub>	1.29	1.30	1.39			1.470	1.480
S–H	CH <sub>3</sub> –SH	1.32	1.30	1.34	1.34		1.345	
S–S	S–(CH <sub>3</sub> ) <sub>2</sub>	2.11	1.94	2.09	2.04		2.024	2.030
N–N	N <sub>2</sub> H <sub>4</sub>	1.38	1.40					1.418
N–H	N <sub>2</sub> H <sub>4</sub>	1.01	1.02		1.01			1.080
C <sub>1</sub> –N <sub>2</sub>	Imidazole	1.35	1.34	1.33	1.34		1.307	1.346
N <sub>2</sub> –C <sub>3</sub>	Imidazole	1.39	1.39	1.38	1.39		1.345	1.346
C <sub>3</sub> –C <sub>4</sub>	Imidazole	1.41	1.39	1.38	1.38		1.430	1.395
C <sub>4</sub> –N <sub>5</sub>	Imidazole	1.39	1.40	1.38	1.39		1.374	1.346
N <sub>5</sub> –C <sub>1</sub>	Imidazole	1.40	1.40	1.36	1.40		1.403	1.346
C <sub>1</sub> –H <sub>6</sub>	Imidazole	1.09	1.08	1.11	1.08		1.110	1.084
C <sub>3</sub> –H <sub>7</sub>	Imidazole	1.09	1.08	1.10	1.08		1.098	1.084
C <sub>4</sub> –H <sub>8</sub>	Imidazole	1.09	1.08	1.10	1.08		1.114	1.084
N <sub>5</sub> –H <sub>9</sub>	Imidazole	0.98	0.99	1.02	1.01		1.048	1.008

tween semiempirically and empirically derived bond lengths, a difference similar to that obtained when empirical force fields are compared.

The comparison of force parameters (see Tables 2 and 3B) shows a notable discrepancy between the results obtained from different methodologies. Thus, the intercorrelations between semiempirical methods are in the range of  $r=0.94-0.97$ , but the root mean square deviations take values in the range 95–129 kcal/mol·Å<sup>2</sup>, which represent a relative root mean square deviation of

TABLE 2  
STRETCHING FORCE PARAMETER FOR SEVERAL BONDS DETERMINED FROM SEMIEMPIRICAL AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, KARPLUS', ALLINGER'S AND TRIPOS/5 FORCE FIELDS. VALUES ARE IN KCAL/MOL·Å<sup>2</sup>

Bond	Molecule	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
C-O	H <sub>2</sub> CO	1219	1332	1210		612	777	778
H-C	H <sub>2</sub> CO	364	418	443	350	350	331	346
C-C	H <sub>3</sub> C-CH <sub>3</sub>	390	394	485	310		316	316.8
H-C	H <sub>3</sub> C-CH <sub>3</sub>	374	433	465	331	300	331	331.2
N-H	H <sub>3</sub> C-NH <sub>2</sub>	440	502	429	434	405	439	346
N-C	H <sub>3</sub> C-NH <sub>2</sub>	464	454	289	367	422	367	380
C-C	H <sub>2</sub> C-CH <sub>2</sub>	805	758	887			690.6	670
H-C	H <sub>2</sub> C-CH <sub>2</sub>	383	472	518		350	330.9	346
C-N	H-CN	1670	1649	1507			1275	800
H-C	H-CN	462	537	502			424.4	350
C-C	HC-CH	1668	1649	1505			1122	700
C-F	H <sub>3</sub> C-F	616	756	372			366.9	300
C-C	C <sub>6</sub> H <sub>6</sub>	476	390	390	469	450	580.1	700
C-H	C <sub>6</sub> H <sub>6</sub>	423	455	497	340			346
C-H	CH <sub>3</sub> -OH	484	565	505	320	400	385	341
C-H	CH <sub>3</sub> -OH	505	667	620	553	450	330.9	503.8
C-C	HOC-CH <sub>2</sub> -NH <sub>2</sub>	302	375	418	335	405	316.5	319.5
C-N	HOC-NH <sub>2</sub>	569	351	599	490	471	460.4	435
C-O	HOC-NH <sub>2</sub>	1088	1240	1032	570	580	695.2	777.6
C-O	O-(CH <sub>3</sub> ) <sub>2</sub>	477	570	501	320	400	385.6	309.4
C-S	S-(CH <sub>3</sub> ) <sub>2</sub>	202	340	254	222		231.1	190.8
O-O	(O-CH <sub>3</sub> ) <sub>2</sub>	495	714	459			561.8	586.1
S-H	CH <sub>3</sub> -SH	181	372	281	274		273.4	
S-S	S-(CH <sub>3</sub> ) <sub>2</sub>	109	314	141	166		223	300
N-N	N <sub>2</sub> H <sub>4</sub>	605	665					650
N-H	N <sub>2</sub> H <sub>4</sub>	436	480		434			346
C <sub>1</sub> -N <sub>2</sub>	Imidazole	616	622	483	488		558	653
N <sub>2</sub> -C <sub>3</sub>	Imidazole	518	542	427	410		461	653
C <sub>3</sub> -C <sub>4</sub>	Imidazole	558	589	570	512		616	700
C <sub>4</sub> -N <sub>5</sub>	Imidazole	580	618	521	422		342	653
N <sub>5</sub> -C <sub>1</sub>	Imidazole	473	601	503	477		439	653
C <sub>1</sub> -H <sub>6</sub>	Imidazole	414	485	468	340		331	346
C <sub>3</sub> -H <sub>7</sub>	Imidazole	438	482	515	340		331	346
C <sub>4</sub> -H <sub>8</sub>	Imidazole	437	504	490	340		331	346
N <sub>5</sub> -H <sub>9</sub>	Imidazole	517	540	498	434		439	350

TABLE 3

STATISTICAL RESULTS OF THE COMPARISON BETWEEN THE EQUILIBRIUM BOND LENGTHS (A) AND STRETCHING FORCE PARAMETERS (B) DETERMINED FROM THE DIFFERENT METHODS<sup>a</sup>

A	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
AM1	1.00	0.99	0.99	0.99	0.99	0.98	0.98
	<b>0.00</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.04</b>	<b>0.04</b>
	0.00	2.4%	2.4%	2.4%	2.4%	3.4%	3.4%
MNDO	0.99	1.00	0.98	0.99	0.99	0.98	0.98
	<b>0.03</b>	<b>0.00</b>	<b>0.04</b>	<b>0.03</b>	<b>0.03</b>	<b>0.04</b>	<b>0.04</b>
	2.4%	0.00	3.4%	2.4%	2.3%	3.4%	3.4%
MINDO	0.99	0.98	1.00	0.99	0.98	0.99	0.98
	<b>0.03</b>	<b>0.04</b>	<b>0.00</b>	<b>0.03</b>	<b>0.04</b>	<b>0.03</b>	<b>0.04</b>
	2.4%	3.4%	0.00	2.3%	3.6%	2.4%	3.4%
Kollman	0.99	0.99	0.99	1.00	0.99	0.99	0.99
	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>	<b>0.03</b>	<b>0.03</b>
	2.4%	2.4%	2.3%	0.00	0.00	2.3%	2.3%
Karplus	0.99	0.99	0.98	0.99	1.00	0.99	0.99
	<b>0.03</b>	<b>0.03</b>	<b>0.04</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.03</b>
	2.4%	2.3%	3.5%	0.00	0.00	0.00	2.4%
Allinger	0.98	0.98	0.98	0.99	0.99	1.00	0.99
	<b>0.04</b>	<b>0.04</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.00</b>	<b>0.03</b>
	3.3%	3.3%	2.4%	2.3%	2.3%	0.00	2.3%
TRIPOS	0.98	0.98	0.98	0.99	0.99	0.99	1.00
	<b>0.04</b>	<b>0.04</b>	<b>0.04</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.00</b>
	3.4%	3.4%	3.4%	2.4%	2.5%	2.4%	0.00
B	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
AM1	1.00	0.97	0.97	0.80	0.91	0.95	0.71
	<b>0.00</b>	<b>102</b>	<b>95</b>	<b>131</b>	<b>227</b>	<b>185</b>	<b>270</b>
	0.00	15%	14%	27%	37%	27%	40%
MNDO	0.97	1.00	0.94	0.63	0.82	0.90	0.68
	<b>102</b>	<b>0.00</b>	<b>129</b>	<b>183</b>	<b>294</b>	<b>232</b>	<b>295</b>
	14%	0.00	18%	34%	43%	32%	41%
MINDO	0.97	0.94	1.00	0.73	0.83	0.91	0.67
	<b>95</b>	<b>129</b>	<b>0.00</b>	<b>137</b>	<b>234</b>	<b>184</b>	<b>257</b>
	14%	20%	0.00	27%	37%	28%	51%
Kollman	0.80	0.63	0.73	1.00	0.80	0.80	0.79
	<b>131</b>	<b>183</b>	<b>137</b>	<b>0.00</b>	<b>56</b>	<b>71</b>	<b>119</b>
	33%	46%	34%	0.00	13%	18%	26%



TABLE 3 (continued)

B	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
Karplus	0.91	0.82	0.83	0.80	1.00	0.89	0.86
	<b>227</b>	<b>294</b>	<b>234</b>	<b>56</b>	<b>0.00</b>	<b>81</b>	<b>110</b>
	<i>52%</i>	<i>67%</i>	<i>53%</i>	<i>13%</i>	<i>0.00</i>	<i>18%</i>	<i>25%</i>
Allinger	0.95	0.90	0.91	0.80	0.89	1.00	0.77
	<b>185</b>	<b>232</b>	<b>184</b>	<b>71</b>	<b>81</b>	<b>0.00</b>	<b>151</b>
	<i>35%</i>	<i>44%</i>	<i>35%</i>	<i>17%</i>	<i>18%</i>	<i>0.00</i>	<i>28%</i>
TRIPOS	0.71	0.68	0.67	0.79	0.86	0.77	1.00
	<b>270</b>	<b>295</b>	<b>257</b>	<b>119</b>	<b>110</b>	<b>151</b>	<b>0.00</b>
	<i>52%</i>	<i>57%</i>	<i>39%</i>	<i>25%</i>	<i>23%</i>	<i>30%</i>	<i>0.00</i>

<sup>a</sup> Standard font refers to the Pearson correlation coefficient, bold font refers to the root mean square deviation (RMS, see Eq. 1), and the italic font refers to the relative root mean square deviation (RMS<sup>REL</sup>, see Eq. 2).

14–20%. The statistical study in Table 3B also points out the poor intercorrelation found between the empirically derived force field parameters included in different force fields. Thus, Pearson correlation coefficients are in the range 0.77–0.89, with root mean square deviations of 56–151 kcal/mol·Å<sup>2</sup> (relative deviations: 13–28%). Semiempirical force parameters correlate quite satisfactorily with Allinger's parameters (Pearson correlation coefficients ranging from 0.90 (MNDO) to 0.95 (AM1), but the correlation is poorer with other force fields, especially with TRIPOS. Root mean square deviations vary strongly depending on the semiempirical method and the empirical force field which are compared (RMS in the range 131–295 kcal/mol·Å<sup>2</sup>); the smallest RMS values are generally obtained for the AM1-derived parameters, especially when they are compared with Kollman's and Allinger's force parameters (relative deviation of 27%), whereas MINDO/3, the oldest Dewar's method is the one that provides stretching parameters more clearly deviating from experimental data.

Finally, it is worth noting that the accuracy of semiempirical methods is not the same for all the bonds. Thus, in general, single bonds (specially the X-H bonds) are more correctly represented than double or triple bonds, where the stretching force parameter is clearly overestimated. The reason for such accuracy is not only related to the intrinsic characteristics of the semiempirical methods, but also to the greater importance of correlation effects in double and triple bonds [30].

### Bending

The ability of Dewar's semiempirical methods to reproduce bendings is explored from the study of 41 bendings. Force parameters and equilibrium bond angles computed from MINDO/3, MNDO and AM1 methods are reported in Tables 4 and 5, where the experimentally derived parameters of Allinger's, Kollman's, Karplus' and TRIPOS force fields are also displayed. The statistical analyses of the relationships between the different bending parameters are displayed in Tables 6A and 6B.

Results in Tables 4 and 6A show the notable similarity in equilibrium bond angles used in the different force fields, with a maximum difference of only 3° (relative RMS of 2%). Similar equilib-

TABLE 4

EQUILIBRIUM VALUES FOR SEVERAL BOND ANGLES DETERMINED FROM SEMIEMPIRICAL AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, ALLINGER'S AND TRIPOS/5 FORCE FIELDS. VALUES ARE IN DEGREES

Bond angle	Molecule	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
H-C-H	H <sub>2</sub> CO	115.5	113.0	106.8			120.0	120.0
H-C-O	H <sub>2</sub> CO	122.2	123.5	126.6		120.3	120.0	120.0
H-C-C	H <sub>3</sub> C-CH <sub>3</sub>	110.7	111.2	113.1	109.5	109.5	109.4	109.5
H-C-H	H <sub>3</sub> C-CH <sub>3</sub>	108.2	107.7	105.6	109.5	109.5	109.4	109.5
H-C-N	H <sub>3</sub> C-NH <sub>2</sub>	109.0	109.5	111.1	109.5		108.8	109.5
C-N-H	H <sub>3</sub> C-NH <sub>2</sub>	111.3	109.8	113.0	109.5		109.5	109.5
H-N-H	H <sub>3</sub> C-NH <sub>2</sub>	109.0	105.5	105.5	107.3	109.5	104.5	109.5
H-C-H	H <sub>2</sub> C=CH <sub>2</sub>	114.9	113.6	110.7			119.0	120.0
H-C-C	H <sub>2</sub> C=CH <sub>2</sub>	122.7	123.2	124.6			120.0	120.0
H-C-N	H-CN	180.0	180.0	180.0				180.0
H-C-C	HC-CH	180.0	180.0	180.0			180.0	180.0
H-C-F	H <sub>3</sub> C-F	109.5	110.6	107.8			110.5	109.5
O-C-C	HOC-CH <sub>2</sub> -NH <sub>2</sub>	123.5	125.5	129.8	120.4	121.5	122.5	120.5
H-C-C	HOC-CH <sub>2</sub> -NH <sub>2</sub>	114.8	113.3	109.6			116.4	120.0
C-O-C	O-(CH <sub>3</sub> ) <sub>2</sub>	112.9	119.9	125.0	109.5		106.8	109.5
H-C-O	O-(CH <sub>3</sub> ) <sub>2</sub>	110.6	112.6		109.5	109.5	106.7	109.5
C-C-C	C <sub>6</sub> H <sub>6</sub>	120.0	120.0	120.0	120.0	120.5	120.0	120.0
H-C-C	C <sub>6</sub> H <sub>6</sub>	120.0	120.0	120.0	120.0		120.0	120.0
C-O-O	(O-CH <sub>3</sub> ) <sub>2</sub>	112.1	113.5	115.0			104.0	
C-S-H	CH <sub>3</sub> -SH	99.6	102.3	106.4	96.0		96.0	97.0
H-C-S	S-(CH <sub>3</sub> ) <sub>2</sub>	112.0	112.2		109.5	109.5	112.0	109.5
H-N-N	N <sub>2</sub> H <sub>4</sub>	107.4	107.1				103.2	109.5
H-N-H	N <sub>2</sub> H <sub>4</sub>	105.9	103.2				104.5	109.5
H-C-O	CH <sub>3</sub> -OH	110.8	112.3	114.9	109.5	109.5	106.7	109.5
H-O-C	CH <sub>3</sub> -OH	107.2	111.6	110.8	108.5	109.5	106.9	109.5
H-N-C	HOC-NH <sub>2</sub>	120.6	117.7	125.7			119.0	119.0
O-C-N	HOC-NH <sub>2</sub>	121.9	121.1	126.6	122.9	121.0	124.3	120.0
H-C-N	HOC-NH <sub>2</sub>	115.0	114.4	110.0	119.8	120.0		120.0
C1-N2-C3	Imidazole	105.8	106.6	106.4	105.1			
N2-C3-C4	Imidazole	110.0	110.3	110.4	109.9			
C3-C4-N5	Imidazole	106.1	105.2	103.7	105.9			
C4-N5-C1	Imidazole	106.8	107.8	110.2	107.3			
N5-C1-N2	Imidazole	111.3	110.1	109.3	111.6			
H6-C1-N2	Imidazole	125.9	126.5	128.2	120.0			
H6-C1-N5	Imidazole	122.8	123.4	122.5	120.0			
H7-C3-N2	Imidazole	121.6	120.6	121.4	120.0			
H7-C3-C4	Imidazole	125.8	129.1	128.2	120.0			
H8-C4-C3	Imidazole	131.3	132.5	134.5	120.0			
H8-C4-N5	Imidazole	122.6	122.4	121.7	120.0			
H9-N5-C4	Imidazole	126.3	121.7	123.2	126.3			
H9-N5-C1	Imidazole	126.9	125.7	123.6	123.1			

TABLE 5

BENDING FORCE PARAMETER FOR SEVERAL BOND ANGLES DETERMINED FROM SEMIEMPIRICAL AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, KARPLUS', ALLINGER'S AND TRIPOS/5 FORCE FIELDS. VALUES ARE IN KCAL/MOL-RAD<sup>2</sup>

Bond angle	Molecule	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
H-C-H	H <sub>2</sub> CO	26	30	27			39.4	39.4
H-C-O	H <sub>2</sub> CO	47	42	38		25.9	39.4	39.4
H-C-C	H <sub>3</sub> -CH <sub>3</sub>	48	47	35	35	40	25.9	26.3
H-C-H	H <sub>3</sub> -CH <sub>3</sub>	36	33	35	35	40	23	26.3
H-C-N	H <sub>3</sub> -NH <sub>2</sub>	51	54	44	35		36	26.3
C-N-H	H <sub>3</sub> C-NH <sub>2</sub>	52	60	28	35		36	65.7
H-N-H	H <sub>3</sub> C-NH <sub>2</sub>	49	56	33	35	40	36	65.7
H-C-H	H <sub>2</sub> C=CH <sub>2</sub>	28	30	27			23	39.4
H-C-C	H <sub>2</sub> C=CH <sub>2</sub>	35	35	25			25.9	19.7
H-C-N	H-CN	32	34	22				65.7
H-C-C	HC-CH	31	34	22			25.9	65.7
H-C-F	H <sub>3</sub> C-F	61	70	23			35.2	26.3
O-C-C	HOC-CH <sub>2</sub> -NH <sub>2</sub>	65	62	40	80	85	46.8	49.2
H-C-C	HOC-CH <sub>2</sub> -NH <sub>2</sub>	79	37	30			26.6	19.7
C-O-C	O-(CH <sub>3</sub> ) <sub>2</sub>	98	116	39	60		55.6	72.2
H-C-O	O-(CH <sub>3</sub> ) <sub>2</sub>	56	60	39	35	50	38.8	26.3
C-C-C	C <sub>6</sub> H <sub>6</sub>	76	67	50	85.9	90	30.9	39.4
H-C-C	C <sub>6</sub> H <sub>6</sub>	35	36	29	35		25.9	39.4
C-O-O	(O-CH <sub>3</sub> ) <sub>2</sub>	77	88	38			44.6	
C-S-H	CH <sub>3</sub> -SH	53	57	24	44		34.1	32.8
H-C-S	S-(CH <sub>3</sub> ) <sub>2</sub>	38	43		35	40	25.9	26.3
H-N-N	N <sub>2</sub> H <sub>4</sub>	63	70				30.9	66.7
H-N-H	N <sub>2</sub> H <sub>4</sub>	51	60				36.0	66.7
H-C-O	CH <sub>3</sub> -OH	57	61	44	35	50	38.5	26.3
H-O-C	CH <sub>3</sub> -OH	69	63	34	55	50	25.2	32.8
H-N-C	HOC-NH <sub>2</sub>	43	26	21			30.2	26.3
O-C-N	HOC-NH <sub>2</sub>	70	77	48	80	85	48.2	42.7
H-C-N	HOC-NH <sub>2</sub>	35	44	38	35	40		39.4
C1-N2-C3	Imidazole	117	102	54	70			
N2-C3-C4	Imidazole	104	94	63	70			
C3-C4-N5	Imidazole	111	87	48	70			
C4-N5-C1	Imidazole	104	107	65	70			
N5-C1-N2	Imidazole	103	97	40	70			
H6-C1-N2	Imidazole	30	34	26	35			
H6-C1-N5	Imidazole	31	36	29	35			
H7-C3-N2	Imidazole	32	35	27	35			
H7-C3-C4	Imidazole	33	30	23	35			
H8-C4-C3	Imidazole	26	30	22	35			
H8-C4-N5	Imidazole	33	37	26	35			
H9-N5-C4	Imidazole	32	38	26	35			
H9-N5-C1	Imidazole	34	37	25	35			

TABLE 6  
 STATISTICAL RESULTS OF THE COMPARISON BETWEEN THE EQUILIBRIUM BOND ANGLES (A) AND  
 STRETCHING FORCE PARAMETERS (B) DETERMINED FROM THE DIFFERENT METHODS<sup>a</sup>

A	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
AM1	1.00	0.99	0.97	0.93	0.93	0.98	0.99
	<b>0.00</b>	<b>2</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>3</b>
	0.00	2%	3%	3%	2%	3%	2%
MNDO	0.99	1.00	0.98	0.86	0.87	0.96	0.97
	<b>2</b>	<b>0.00</b>	<b>3</b>	<b>4</b>	<b>3</b>	<b>5</b>	<b>4</b>
	2%	0.00	2%	4%	3%	4%	3%
MINDO	0.97	0.98	1.00	0.77	0.75	0.90	0.94
	<b>4</b>	<b>3</b>	<b>0.00</b>	<b>6</b>	<b>5</b>	<b>7</b>	<b>7</b>
	3%	2%	0.00	5%	5%	6%	6%
Kollman	0.93	0.86	0.77	1.00	0.99	0.98	0.99
	<b>3</b>	<b>4</b>	<b>6</b>	<b>0.00</b>	<b>1</b>	<b>2</b>	<b>1</b>
	3%	4%	5%	0.00	1%	2%	1%
Karplus	0.93	0.87	0.75	0.99	1.00	0.96	0.99
	<b>2</b>	<b>3</b>	<b>5</b>	<b>1</b>	<b>0.00</b>	<b>2</b>	<b>1</b>
	2%	3%	5%	1%	0.00	2%	1%
Allinger	0.98	0.96	0.90	0.98	0.96	1.00	0.99
	<b>3</b>	<b>5</b>	<b>7</b>	<b>2</b>	<b>2</b>	<b>0.00</b>	<b>3</b>
	3%	4%	6%	2%	2%	0.00	2%
TRIPOS	0.99	0.97	0.94	0.99	0.99	0.99	1.00
	<b>3</b>	<b>4</b>	<b>7</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>0.00</b>
	2%	3%	6%	1%	1%	2%	0.00
B	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
AM1	1.00	0.92	0.78	0.81	0.82	0.57	0.11
	<b>0.00</b>	<b>10</b>	<b>28</b>	<b>20</b>	<b>12</b>	<b>25</b>	<b>24</b>
	0.00	17%	46%	31%	22%	46%	43%
MNDO	0.92	1.00	0.77	0.76	0.80	0.73	0.37
	<b>10</b>	<b>0.00</b>	<b>27</b>	<b>20</b>	<b>13</b>	<b>23</b>	<b>25</b>
	17%	0.00	45%	31%	24%	41%	44%
MINDO	0.78	0.77	1.00	0.72	0.77	0.47	0.10
	<b>28</b>	<b>27</b>	<b>0.00</b>	<b>17</b>	<b>22</b>	<b>20</b>	<b>9</b>
	78%	75%	0.00	43%	31%	57	25%
Kollman	0.81	0.76	0.72	1.00	0.97	0.46	0.26
	<b>20</b>	<b>20</b>	<b>17</b>	<b>0.00</b>	<b>8</b>	<b>22</b>	<b>21</b>
	39%	38%	33%	0.00	14%	43%	41%

TABLE 6 (continued)

B	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
Karplus	0.82	0.80	0.77	0.97	1.00	0.53	0.21
	<b>12</b>	<b>13</b>	<b>22</b>	<b>8</b>	<b>0.00</b>	<b>26</b>	<b>26</b>
	<i>22%</i>	<i>23%</i>	<i>39%</i>	<i>13%</i>	<i>0.00</i>	<i>46%</i>	<i>46%</i>
Allinger	0.57	0.73	0.47	0.46	0.53	1.00	0.39
	<b>25</b>	<b>23</b>	<b>20</b>	<b>22</b>	<b>26</b>	<b>0.00</b>	<b>17</b>
	<i>56%</i>	<i>52%</i>	<i>46%</i>	<i>51%</i>	<i>66%</i>	<i>0.00</i>	<i>39%</i>
TRIPOS	0.11	0.37	0.10	0.26	0.21	0.39	1.00
	<b>24</b>	<b>25</b>	<b>9</b>	<b>21</b>	<b>26</b>	<b>17</b>	<b>0.00</b>
	<i>69%</i>	<i>73%</i>	<i>24%</i>	<i>58%</i>	<i>75%</i>	<i>49%</i>	<i>0.00</i>

<sup>a</sup> Standard font refers to the Pearson correlation coefficient, bold font refers to the root mean square deviation (RMS, see Eq. 1), and the italic font refers to the relative root mean square deviation ( $\text{RMS}^{\text{REL}}$ , see Eq. 2).

rium bond angles were obtained by the different semiempirical methods, the AM1 leading to equilibrium bond angles in best agreement with experimentally derived results (differences of only 3–4°, relative RMS 2–3%).

The comparison of force parameters (see Tables 5 and 6B) shows a great variability, either where it concerns semiempirical or experimental force field data. Thus, MNDO and AM1 results are quite similar ( $r = 0.92$ ,  $\text{RMS} = 10 \text{ kcal/mol-rad}^2$ , and relative RMS = 17%), but they differ from the force parameters obtained from computations by MINDO/3, a method that leads to a low rigidity for most bond angles. Force parameters reported in the different force fields vary considerably. Thus, with the exception of the good correlation between Kollman's and Karplus' data, poor intercorrelation coefficients were obtained ( $r$  values ranging from 0.21 to 0.53), the RMS deviations giving values in the range of 17–26  $\text{kcal/mol-rad}^2$  (relative RMS greater than 40%). The correlations between semiempirically and empirically derived force field are not as good as in the case of bond stretching, but the correlations are in general similar or better than those found when experimental parameters are compared with each other. The RMS deviation between semiempirical and empirical force parameters is in the range of 9–25  $\text{kcal/mol-rad}^2$ , which is similar to, or less than, the RMS difference between the different empirically derived force parameters. As a general rule, AM1 and MNDO methods overestimate the rigidity of most bond angles with respect to empirical force fields, especially with respect to Allinger's and TRIPOS. In contrast, MINDO/3 exhibits a more irregular profile, since the rigidity of some bond angles is underestimated, whereas it is overestimated for other bond angles when compared with empirical force field parameters.

### *Improper torsion*

The improper torsion term is used in 'all atoms' force fields as a reinforcement to the torsion term to represent the planar tendency of  $\text{sp}^2$  atoms. Consequently, the number of different improper torsion parameters in force fields is rather small, and in most cases they are only approximate values (for instance, only two different force constants are used in TRIPOS, three in MM2, and also three in Kollman's force field to represent all the different 'out-of-plane bendings'). This,

and the diversity of formalisms used to compute the improper torsion energy (see Methods) makes it difficult to perform a systematic comparison between the different improper torsion parameters. Nevertheless, in order to explore the reliability of force parameters derived from Dewar's method, three improper torsions, X-X-C=O (H<sub>2</sub>CO), X-X-Caromatic-H (C<sub>6</sub>H<sub>6</sub>) and X-X-C-C (CH<sub>2</sub>CH<sub>2</sub>), have been studied. Parameters derived from semiempirical calculations are compared to those used in Allinger's, Karplus', Kollman's and TRIPOS force fields in Table 7.

Inspection of Table 7 shows that the three semiempirical methods provide similar improper torsion force parameters, MNDO leads in general to the largest force parameters, whereas MINDO/3 parameters are the smallest. Comparison between empirical and semiempirical derived parameters is difficult owing to the small number of different improper torsions available in force fields, and the low accuracy and uncertainty of the experimental values. Nevertheless, it is clear that all the semiempirical methods lead to force parameters which are notably lower than the TRIPOS 'universal' value of 240 kcal/mol·Å<sup>2</sup>. Acceptable agreement is found between Kollman's and semiempirical parameters, especially when the MINDO/3 results are considered. With respect to Allinger's force field, good agreement is found for the 'out-of-plane bending' of H<sub>2</sub>CO, but MM2

TABLE 7  
FORCE PARAMETERS FOR DIFFERENT IMPROPER TORSIONS (OUT-OF-PLANE BENDING) DETERMINED FROM AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, KARPLUS', ALLINGER'S AND TRIPOS/5 FORCE FIELDS<sup>a</sup>

	AM1	MNDO	MINDO/3	Experimental
<b>X-X-C=O</b> (H <sub>2</sub> CO)				
Kollman	12.3	13.7	11.5	10.5
Karplus	21	24	20	100
Allinger	226	248	204	173
TRIPOS	171	190	157	240
<b>X-X-Csp<sub>2</sub>-H</b> (C <sub>2</sub> H <sub>4</sub> )				
Kollman	4.4	5.0	1.0	
Karplus	8.0	9.0	1.9	
Allinger	98	116	26	10.9
TRIPOS	74	86	19	240
<b>X-X-CA-H</b> (C <sub>6</sub> H <sub>6</sub> )				
Kollman	7.1	8.2	3.9	2.0
Karplus	11	13	6.2	
Allinger	120	137	65	10.9
TRIPOS	62	71	33	240

<sup>a</sup> The units in which values are displayed vary depending on the formalism used to represent the improper torsion. Thus for Allinger's and Karplus' algorithms the units are kcal/mol·rad<sup>2</sup>, for TRIPOS algorithms they are expressed in kcal/mol·Å<sup>2</sup>, and finally for Kollman's formalism they are in kcal/mol.

force parameters for  $\text{CH}_2\text{CH}_2$  and  $\text{C}_6\text{H}_6$  are smaller than those derived from semiempirical computations. It should be noted that the best agreement between MM2 and semiempirical force parameters is found when MINDO/3 results are considered, and that the least agreement is obtained when MNDO is considered. The low ability of MNDO to reproduce 'out-of-plane bendings' is in absolute agreement with the well-known shortcomings of this method to reproduce some planar structures, such as the nitro- and aminobenzenes.

### *Proper torsion*

Force fields represent proper torsion following two different algorithms: (i) the general parameter strategy according to which all the dihedral angles which represent a bond rotation have the same parameters; and (ii) the singular parameter strategy in which each dihedral angle has different force parameters. In spite of the loss of accuracy, most of the force field parameters use the general parameter strategy in order to reduce the number of parameters necessary to run a calculation. Since our purpose is only to explore the suitability of Dewar's methods as a source of approximate force field parameters we have always considered the general parameter strategy with the only exception of  $\text{HCOOH}$ , for which the singular parameters describing the  $\text{O}-\text{C}-\text{O}-\text{H}$  were determined. All the parameters obtained from semiempirical computations, and the empirical values of Kollman, Karplus, Allinger and TRIPOS are shown in Table 8.

Statistical results of the comparison between the different force field parameters in Table 9 demonstrate general agreement between the different parameters, the major discrepancies being obtained for the TRIPOS force field. The performance of semiempirical methods is quite satisfactory, especially that of the AM1 method, which leads to force parameters deviating by only 0.44 kcal/mol from Karplus' data, 0.52 kcal/mol from Kollman's and 0.84 kcal/mol from Allinger's. The Pearson correlation coefficients between the different semiempirically derived force field pa-

TABLE 8

TORSIONAL PARAMETERS FOR SEVERAL BONDS DETERMINED FROM SEMIEMPIRICAL AM1, MNDO AND MINDO/3 COMPUTATIONS AND VALUES REPORTED IN KOLLMAN'S, KARPLUS', ALLINGER'S AND TRIPOS/5 FORCE FIELDS. S1, S2 AND S3 STANDS FOR THE HALF OF THE TORSIONAL BARRIER ENERGY CORRESPONDING TO THE ONE-, TWO- AND THREEFOLD TORSIONAL TERM (V1, V2 AND V3). ALL THE VALUES ARE IN KCAL/MOL

Torsion	Molecule	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
Csp3-Csp3	$\text{CH}_3-\text{CH}_3$	S3=0.60	S3=0.50	S3=0.42	S3=1.30	S3=1.60	S3=1.02	S3=1.44
Csp3-Nsp3	$\text{CH}_3-\text{NH}_2$	S3=0.52	S3=0.41	S3=0.30		S3=0.60	S3=0.75	S3=0.60
Csp3-Osp3	$\text{CH}_3-\text{OH}$	S3=0.50	S3=0.26	S3=0.12	S3=0.50	S3=0.50	S3=0.60	S3=1.80
Csp3-Osp3	$\text{CH}_3-\text{O}-\text{CH}_3$	S3=0.41	S3=0.51	S3=0.55	S3=1.00		S1=0.10 S3=1.20	S3=1.80
Csp3-Ssp3	$\text{CH}_3-\text{S}-\text{CH}_3$	S3=0.28	S3=0.28	S3=0.30			S3=0.81	S3=1.20
Csp2-Csp2	$\text{CH}_2=\text{CH}_2$	S3=32.5	S3=32.5	S3=32.5			S2=30.0	S2=25.0
Csp3-Csp2	$\text{CH}_3-\text{CH}=\text{CH}_2$	S3=0.51	S3=0.40	S3=0.40			S2=0.42	S2=0.27
O-C-O-H	$\text{HCOOH}$	S1=-1.64 S2=-2.58	S1=-3.48 S2=-3.67	S1=-1.73 S2=-2.71			S1=-1.64 S2=-2.80	
Csp3-Nsp3	$\text{CH}_3-\text{NH}_3^+$	S3=0.69	S3=0.53	S3=0.54	S3=1.40	S3=0.60	S3=1.12	
Csp3-Csp2	$\text{CH}_3-\text{CHO}$	S3=0.026	S3=0.015	S3=0.032	S3=0.067	S3=0.00	S3=0.262	S3=0.137

rameters and the empirical values are also satisfactory, especially when the accurate parameters of Allinger are considered as reference. Moreover, when not only the energy barrier, but the whole energy profile for the bond rotation is considered, the suitability of semiempirical methods is clearly demonstrated, not only by the fact that all the bond periodicities were correctly reproduced, but also by the small mean variation of the  $\text{RMS}^{\text{path}}$  (see definition in Eq. 5), which take values of a few tenths of kcal/mol (see Table 9).

From the results presented above, it can be stated that semiempirical methods (especially the AM1), even though they minimize the torsional energy barriers, can be good alternatives to highly sophisticated ab initio calculations, since results reported here show deviations between semiempirical and empirical derived parameters that are within the narrow range of variability of experimental values, and in good agreement with results reported by other authors [4,35–39]. Nevertheless, it has been also reported [4] that semiempirical methods lead to bad representations of

TABLE 9  
STATISTICAL RESULTS OF THE COMPARISON BETWEEN THE TORSION PARAMETERS DETERMINED FROM THE DIFFERENT METHODS<sup>a</sup>

	AM1	MNDO	MINDO	Kollman	Karplus	Allinger	TRIPOS
AM1	1.00	0.99	0.99	0.89	0.64	0.99	0.99
	<b>0.00</b>	<b>0.45</b>	<b>0.16</b>	<b>0.52</b>	<b>0.44</b>	<b>0.84</b>	<b>2.77</b>
	<i>0.00</i>	<i>0.35</i>	<i>0.17</i>	<i>0.55</i>	<i>0.35</i>	<i>0.68</i>	<i>2.09</i>
MNDO		1.00	0.99	0.96	0.72	0.99	0.99
		<b>0.00</b>	<b>0.39</b>	<b>0.58</b>	<b>0.48</b>	<b>0.93</b>	<b>2.78</b>
		<i>0.00</i>	<i>0.27</i>	<i>0.66</i>	<i>0.43</i>	<i>0.92</i>	<i>2.14</i>
MINDO			1.00	0.91	0.59	0.99	0.99
			<b>0.00</b>	<b>0.61</b>	<b>0.53</b>	<b>0.86</b>	<b>2.8</b>
			<i>0.00</i>	<i>0.70</i>	<i>0.49</i>	<i>0.76</i>	<i>2.18</i>
Kollman				1.00	0.75	0.93	0.67
				<b>0.00</b>	<b>0.39</b>	<b>0.22</b>	<b>0.77</b>
				<i>0.00</i>	<i>0.37</i>	<i>0.28</i>	<i>0.79</i>
Karplus					1.00	0.67	0.54
					<b>0.00</b>	<b>0.38</b>	<b>0.74</b>
					<i>0.00</i>	<i>0.46</i>	<i>0.70</i>
Allinger						1.00	0.99
						<b>0.00</b>	<b>1.84</b>
						<i>0.00</i>	<i>1.35</i>
TRIPOS							1.00
							<b>0.00</b>
							<i>0.00</i>

<sup>a</sup> Standard font refers to the Pearson correlation coefficient, bold font refers to the root mean square deviation (RMS, see Eq. 1), and the italic font refers to the root mean square deviation along the rotation path ( $\text{RMS}^{\text{PATH}}$ , see Eq. 3).



some torsions, like those of conjugate bonds, for which sophisticated *ab initio* calculations, including correlation effects, are sometimes necessary. Therefore, and in spite of the good results obtained in our computations, caution is necessary when semiempirical calculations are used as reference methods to parametrize the torsional term, since these methods are not able to represent rotation around the different bonds with the same accuracy.

## CONCLUSIONS

The results presented here, which are based on the computation of up to 270 force parameters and of 270 equilibrium geometries, allow us to obtain several general conclusions about the suitability of Dewar's Hamiltonians:

1. Results demonstrate the suitability of Dewar's methods to provide equilibrium geometric values (equilibrium bond lengths, bond angles, and bond rotation periodicity). The AM1 method leads to the results closest to the experimental values, whereas the MINDO/3 gives the worst results. In general the differences obtained between the semiempirically derived equilibrium geometric values and the empirically derived ones are very small, and in practice lie within the range of the differences found between the different empirically derived values.
2. From our statistical analysis it is evident that a notable difference exists between the force parameters reported in the different force fields. This seems surprising since all the force fields were determined from experimental data. Therefore, this disagreement should mainly be related to the different sets of molecules considered during the parametrization procedure in the different force fields. Accordingly, it is clear that force parameters are not fully transferable, but rather depend on the set of molecules which was considered during the parametrization procedure, which strongly suggests that parametrization should be performed in molecules as similar as possible to that of real interest.
3. Due to the great relevance of the molecular environment in the determination of force parameters, semiempirical results reported in this paper should be considered only as approximate values, which are only valid for the molecule studied. A rigorous determination of a general force parameter would imply its determination from a large set of molecules, the results being averaged. Consequently, all comparison analyses between theoretical and experimental results only have a qualitative value, and must not be used for instance to scale semiempirical results.
4. In spite of the considerations noted in point 3, the general suitability of Dewar's Hamiltonians as a source of force parameters is demonstrated. In general, and even though caution is necessary, Dewar's semiempirical methods, specially the AM1 method, provides force parameters which are within the range of variability of the experimental values reported in several of the most used force fields. This suggests that the inexpensive semiempirical parametrization can be a good alternative for other methods when the 'missing' parameter appears during a molecular mechanical or dynamical study.

## ACKNOWLEDGEMENTS

We are indebted to Dr. S. Olivella and Dr. J.M. Bofill for making accessible their particular copy of the MOPAC program. We also thank Dr. F.J. Luque and Dr. J.M. Bofill for many helpful and encouraging discussions.

## REFERENCES

- 1 Alemán, C., Canela, E.I., Franco, R. and Orozco, M., *J. Comp. Chem.*, 12 (1991) 664.
- 2 Bingham, R.C., Dewar, M.J.S. and Lo, D.H., *J. Am. Chem. Soc.*, 97 (1975) 1285.
- 3 Dewar, M.J.S. and Thiel, W., *J. Am. Chem. Soc.*, 99 (1977) 4899.
- 4 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., *J. Am. Chem. Soc.*, 107 (1985) 3902.
- 5 Allinger, N.L., *J. Am. Chem. Soc.*, 99 (1977) 8127.
- 6 Sprague, J.T., Tai, J.C., Yuh, Y.H. and Allinger, N.L., *J. Comp. Chem.*, 8 (1977) 581.
- 7 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, P.A., *J. Comp. Chem.*, 7 (1986) 230.
- 8 Weiner, S.J., Kollman, P.A., Case, P.A., Singh, U.C., Ghio, G. and Alagona, G., *J. Am. Chem. Soc.*, 106 (1984) 765.
- 9 Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., *J. Comp. Chem.*, 4 (1983) 187.
- 10 Nilsson, L. and Karplus, M., *J. Comp. Chem.*, 7 (1986) 591.
- 11 Clark, M., Cramer III, R.D. and Opdenbosch, N., *J. Comp. Chem.*, 10 (1989) 982.
- 12 Karplus, M. and Petsko, G.A., *Nature*, 347 (1990) 631.
- 13 Allinger, N.L., Yuh, Y.H. and Lii, J.H., *J. Am. Chem. Soc.*, 111 (1989) 8551.
- 14 Mommany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., *J. Phys. Chem.*, 79 (1975) 2361.
- 15 Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F. and Postma, J.P.M., *Biopolymers*, 23 (1984) 1513.
- 16 Hagler, A.T., Huler, E. and Lifson, S., *J. Am. Chem. Soc.*, 96 (1974) 5319.
- 17 Maple, J.R., Dinur, U. and Hagler, A.T., *Proc. Natl. Acad. Sci. USA*, 85 (1988) 5350.
- 18 Dinur, U. and Hagler, A.T., *J. Am. Chem. Soc.*, 111 (1989) 5149.
- 19 Hagler, A.T., Maple, J.R., Thacher, T.S., Fitzgerald, G.B. and Dinur, U., In van Gunsteren, W.F. and Weiner, P.K. (Eds.), *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, ESCOM, Leiden, 1989, pp. 149–167.
- 20 Dinur, U. and Hagler, A.T., *J. Comp. Chem.*, 11 (1990) 1234.
- 21 Hopfinger, A.J. and Pearlstein, R.A., *J. Comp. Chem.*, 5 (1984) 486.
- 22 Fogarasi, G. and Pulay, P., *Annu. Rev. Phys. Chem.*, 35 (1984) 191.
- 23 Leonard, J.M. and Ashman, W.P., *J. Comp. Chem.*, 11 (1990) 952.
- 24 Singh, U.C. and Kollman, P.A., *J. Comp. Chem.*, 5 (1984) 129.
- 25 Cox, S.R. and Williams, D.E., *J. Comp. Chem.*, 2 (1981) 304.
- 26 Williams, D.E., *J. Comp. Chem.*, 7 (1988) 745.
- 27 Momany, F.A., *J. Phys. Chem.*, 82 (1978) 592.
- 28 Orozco, M. and Luque, F.J., *J. Comp. Chem.*, 11 (1990) 909.
- 29 Palmö, K., Pietilä, L.O. and Krimm, S., *J. Comp. Chem.*, 12 (1991) 385.
- 30 Orozco, M. and Luque, F.J., submitted.
- 31 Orozco, M. and Luque, F.J., *J. Comput.-Aided Mol. Design*, 4 (1990) 411.
- 32 Olivella, S., *QCPE Bull.*, 4 (1984) 109. (Modified by S. Olivella and J.M. Boffill, 1990.)
- 33 Stewart, J.J.P., *QCPE Bull.*, 3 (1983) 101.
- 34 Boffill, J.M., Gomez, J. and Olivella, S., *J. Mol. Struct. (THEOCHEM)*, 163 (1988) 285.
- 35 Clark, T., *Handbook of Computational Chemistry*, Wiley, New York, 1985.
- 36 Tyrell, J., *J. Am. Chem. Soc.*, 101 (1979) 3766.
- 37 Tvaroska, I., *Coll. Czech. Chem. Comm.*, 47 (1982) 3199.
- 38 van Lier, J.J.C., Smits, M.T. and Buck, H.M., *Eur. J. Biochem.*, 132 (1983) 55.
- 39 Jones, A.A., O’Gara, J.F., Inglefield, P.T., Bendler, J.T., Yee, A.F. and Ngai, K.L., *Macromolecules*, 16 (1983) 658.