

RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I

GERD B. ROCHA,¹ RICARDO O. FREIRE,¹ ALFREDO M. SIMAS,¹ JAMES J. P. STEWART²

¹Departamento de Química Fundamental, CCEN, UFPE, 50590-470, Recife PE, Brazil

²Stewart Computational Chemistry, 15210 Paddington, Colorado Springs Colorado 80921-2512

Received 25 October 2005; Accepted 27 December 2005

DOI 10.1002/jcc.20425

Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Twenty years ago, the landmark AM1 was introduced, and has since had an increasingly wide following among chemists due to its consistently good results and time-tested reliability—being presently available in countless computational quantum chemistry programs. However, semiempirical molecular orbital models still are of limited accuracy and need to be improved if the full potential of new linear scaling techniques, such as MOZYME and LocalSCF, is to be realized. Accordingly, in this article we present RM1 (Recife Model 1): a reparameterization of AM1. As before, the properties used in the parameterization procedure were: heats of formation, dipole moments, ionization potentials and geometric variables (bond lengths and angles). Considering that the vast majority of molecules of importance to life can be assembled by using only six elements: C, H, N, O, P, and S, and that by adding the halogens we can now build most molecules of importance to pharmaceutical research, our training set consisted of 1736 molecules, representative of organic and biochemistry, containing C, H, N, O, P, S, F, Cl, Br, and I atoms. Unlike AM1, and similar to PM3, all RM1 parameters have been optimized. For enthalpies of formation, dipole moments, ionization potentials, and interatomic distances, the average errors in RM1, for the 1736 molecules, are less than those for AM1, PM3, and PM5. Indeed, the average errors in kcal · mol⁻¹ of the enthalpies of formation for AM1, PM3, and PM5 are 11.15, 7.98, and 6.03, whereas for RM1 this value is 5.77. The errors, in Debye, of the dipole moments for AM1, PM3, PM5, and RM1 are, respectively, 0.37, 0.38, 0.50, and 0.34. Likewise, the respective errors for the ionization potentials, in eV, are 0.60, 0.55, 0.48, and 0.45, and the respective errors, in angstroms, for the interatomic distances are 0.036, 0.029, 0.037, and 0.027. The RM1 average error in bond angles of 6.82° is only slightly higher than the AM1 figure of 5.88°, and both are much smaller than the PM3 and PM5 figures of 6.98° and 9.83°, respectively. Moreover, a known error in PM3 nitrogen charges is corrected in RM1. Therefore, RM1 represents an improvement over AM1 and its similar successor PM3, and is probably very competitive with PM5, which is a somewhat different model, and not fully disclosed. RM1 possesses the same analytical construct and the same number of parameters for each atom as AM1, and, therefore, can be easily implemented in any software that already has AM1, not requiring any change in any line of code, with the sole exception of the values of the parameters themselves.

© 2006 Wiley Periodicals, Inc. J Comput Chem 27: 1101–1111, 2006

Key words: semiempirical methods; RM1; AM1; PM3; PM5

Introduction

Twenty years ago, the landmark quantum chemical semiempirical method AM1¹ was introduced, and has since had an increasingly wide following among chemists due to its consistently good results and time-tested reliability. AM1 is presently available in many commercial and noncommercial quantum chemistry softwares, such as MOPAC,² Gaussian,³ Hyperchem (HyperChem™ Professional 7.51, Hypercube, Inc., Gainesville, FL), Spartan (Spartan '04, Wavefunction, Inc., Irvine, CA), Amber,⁴ Dynamo,⁵ VAMP (AMP, Accelrys Corporate Headquarters, San Diego, CA), Ghemical,⁶ BOSS,⁷ AMPAC (AMPAC 8.0 manual Copyright © 2004, Semichem, Inc., USA, <http://www.semichem.com>), and countless others.

Correspondence to: A. M. Simas; e-mail: simas@ufpe.br

Contract/grant sponsor: CNPq (Brazilian agency)

Contract/grant sponsor: the Instituto do Milênio de Materiais Complexos

Contract/grant sponsor: FACEPE (Programa Primeiros Projetos)

Contract/grant sponsor: Construção do Conhecimento por Agrupamento de Dados (CoCADA)

Contract/grant sponsor: NIH; contract/grant number: 2 R44 GM067327-02 (to J.J.P.S.)

This article contains Supplementary Material available at <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>

In 1989, Stewart improved the techniques of parameterization^{8,9} and published PM3,⁸ which gave lower average errors than AM1, mainly for the enthalpies of formation.

Subsequently, Stewart further improved his parameterization program and was able to optimize 37 elements simultaneously: all main groups, together with the three transition metals with fully filled d orbitals: Zn, Cd, and Hg. The resulting model, PM5, possesses average errors about 40% less than the other semiempirical models for all 37 elements. However, PM5 was not published, being available only in commercial softwares that contain MOPAC, like MOPAC 2002.¹⁰

There are other semiempirical methods and models available that may possibly be even more accurate, such as SCC-DFTB,¹¹ OM1,¹² OM2,¹² and PM3-PDDG.¹³ However, either they have not yet been parameterized for many elements, or they have not been as widely distributed as AM1 or PM3.

Semiempirical molecular orbital methods have been proven useful to chemical research in many ways. They are particularly invaluable in three situations: (1) for the calculation of large sets of molecules, as in molecular design for specific purposes, such as in the design of compounds for nonlinear optics,¹⁴ and, in general, for computational combinatorial chemistry and drug design;¹⁵ (2) for the repeated calculation of a given system, such as in dynamics simulations;¹⁶ and (3) for the calculation of very large molecules.

Recently, two techniques have been introduced, aiming at obtaining linear scaling in semiempirical calculations: MOZYME¹⁷ and LocalSCF.¹⁸ Both share the same semiempirical code, and LocalSCF is reported to have achieved true linear scaling, not only in computing time, but also in memory, and seems to be better suited for geometry optimization of proteins, whereas MOZYME allows a faster evaluation of absolute heats of formation of large molecules.

As a result, today it is perfectly possible to conduct a fully AM1-like semiempirical calculation of a protein of more than one hundred thousand atoms using a personal computer. Full molecular orbital calculations of systems such as enzymes, assemblies of proteins, DNA, polysaccharides, etc., start to appear in the literature.^{19–30} For example, the computation of the wave function and total energy of the system GroEL–GroES–Chaperonine system (119,273 atoms) at fixed geometry took approximately 4.5 h in a 2.4-GHz Pentium 4 personal computer.¹⁸ Another example was the calculation of the frontier orbitals of an enzyme, Ribonuclease T1, which revealed that they are located in its active site.¹⁹

Many other articles have appeared in the literature using semiempirical methods for the calculation of systems important for biochemistry: calculation of pKa of proteins,^{20–22} studies of enzymatic reaction mechanisms^{23–26} calculation of spectroscopic properties of proteins,³¹ among many others.^{32–38} A recent study showed a molecular dynamics calculation of the protein crambin, in solution, where the protein was treated with a semiempirical method and the solvent by a molecular mechanics force field. Results reproduced much more accurately aspects of the geometry of the protein than would be obtained by a pure force field approach.³⁹

In this work we introduce a reparameterization of the AM1 (Austin Model 1) semiempirical molecular orbital method, which by analogy, we call RM1 (Recife Model 1).

Parameterization of RM1

The accuracy of a semiempirical method is determined by a synergy of three factors: (1) its formalism and approximations, (2) the training set of molecules, and (3) the quality of the numerical techniques used for parameterization. At first, one may think that to improve a semiempirical model, it would be necessary to introduce changes in the formalisms and equations, aiming at slowly transubstantiating the method into a first principles methodology. However, we consider that the most critical part of a semiempirical model development is, instead, its parameterization. Indeed, in AM1, for all 10 elements considered in this article, there are 191 different parameters that must be optimized.

A precondition for achieving an accurate model is the quantity and quality of the experimental values used. It is simply not possible to arrive to a sensible 191 parameter set, which will define the model, without a considerable amount of atomic and molecular systems in the reference parameterization set.

Our reference parameterization set is thus composed of 1736 species, consisting mainly of atoms and molecules of importance to organic and biochemistry. Similar to AM1, we used, as references, experimental values of enthalpies of formation, dipole moments, ionization potentials, and geometrical variables (interatomic distances, bond, and dihedral angles). We refer to interatomic distances because in this category we included not only bond lengths, but also strategic distances between nonbonded atoms. We did not include all interatomic distances and angles available, to maintain equilibrium in the response function, which would, otherwise, be disturbed by overrepresented CH bond

Table 1. Description of the Parameters for the RM1 Semiempirical Method.

Parameter	Description
U_{ss}	s atomic orbital one-electron one-center integral
U_{pp}	p atomic orbital one-electron one-center integral
β_s	s atomic orbital one-electron two-center resonance integral term
β_p	p atomic orbital one-electron two-center resonance integral term
α_A	atom A core-core repulsion term
Gss	s–s atomic orbitals one-center two-electron repulsion integral
Gsp	s–p atomic orbitals one-center two-electron repulsion integral
Gpp	p–p atomic orbitals one-center two-electron repulsion integral
Gp2	p–p' atomic orbitals one-center two-electron repulsion integral
Hsp	s–p atomic orbital one-center two-electron exchange integral
a_i	Gaussian multiplier for the i^{th} Gaussian of atom A
b_i	Gaussian exponent multiplier for the i^{th} Gaussian of atom A
c_i	radial center of the i^{th} Gaussian of atom A
ζ_s	s-type Slater atomic orbital exponent
ζ_p	p-type Slater atomic orbital exponent

Table 2. Weights for the Properties, Used in the Definition of the Nonlinear Response Function, F_{resp} , for the Parameterization Procedure.

Property, q_i	Weight, w_i
Heat of formation (ΔH_f)	1 kcal ⁻¹ mol
Ionization potential (IP)	10 eV ⁻¹
Dipole moment (μ)	20 D ⁻¹
Interatomic distance (R_{ab})	100 Å ⁻¹
Bond angle (θ_i)	2/3 degree ⁻¹
Dihedral angle (θ_d)	1/3 degree ⁻¹

lengths, for example. Hence, we considered only handpicked key interatomic distances and angles, chosen to be representative and span the differing bonding situations present in molecules of these atoms. Moreover, for geometrical variables only, in some cases we included as reference data results from B3LYP/6-31G* calculations when gas-phase experimental data are not easily available, mostly for molecules of importance to biochemistry, such as amino acids, both neutral and protonated, representative examples of dipeptides with α -helix and β -sheet types of conformations, all five nitrogen bases of DNA, together with the pairs adenine–thymine, guanine–cytosine, and adenine–uracil, and the phosphates. The saccharides pentose and ribose were also included, as well as most aldoses and ketoses. We also considered the glycosidic linkage, by including a few representative disaccharides, such as lactose, sucrose and maltose.

Table 1 presents a description of the parameters to be optimized, which, of course, are the same as for AM1. The atoms differ only in the number of Gaussian functions in the core–core repulsion integral. As in AM1, oxygen, and the halogens fluorine, chlorine, bromine, and iodine, all have two Gaussians; hydrogen, nitrogen, phosphorus, and sulfur have three Gaussians; and carbon has four Gaussians.

Unlike AM1, and similarly to PM3 and PM5, all 191 parameters of RM1 have been optimized, including the one-center two-electron repulsion integrals G_{ss} , G_{sp} , G_{pp} , G_{p2} , and H_{sp} , which, in AM1, were taken from experimental data.

The parameterization procedure, the most time-consuming part of the present work, consisted in the optimization of a nonlinear response function, F_{resp} , parametrically defined in a space of 191 dimensions and computed from the deviations between the predicted and reference values according to eq. (1)

$$F_{\text{resp}} = \sum_i (q_i^{\text{calc}} - q_i^{\text{exp}})^2 w_i^2, \quad (1)$$

where q^{calc} and q^{exp} are, respectively, the calculated and experimental values of the property i for the molecules of the parameterization set. For the weights w_i , we used the same MNDO values according to Table 2.

To find the best parameter set possible, we used a mix of numerical techniques, Newton–Raphson and simplex,⁴⁰ as well as some others developed by Stewart,^{8,9,41} all coded to tackle large multidimensional problems.

Table 3. RM1 Parameters for H, C, N, and O Atoms.

Parameter	H	C	N	O
U_{ss} (eV)	−11.96067697	−51.72556032	−70.85123715	−96.94948069
U_{pp} (eV)	—	−39.40728943	−57.97730920	−77.89092978
β_s (eV)	−5.76544469	−15.45932428	−20.87124548	−29.85101212
β_p (eV)	—	−8.23608638	−16.67171853	−29.15101314
α (Å ⁻¹)	3.06835947	2.79282078	2.96422542	4.17196717
G_{ss} (eV)	13.98321296	13.05312440	13.08736234	14.00242788
G_{sp} (eV)	—	11.33479389	13.21226834	14.95625043
G_{pp} (eV)	—	10.95113739	13.69924324	14.14515138
G_{p2} (eV)	—	9.72395099	11.94103953	12.70325497
H_{sp} (eV)	—	1.55215133	5.00000846	3.93217161
a_1	0.10288875	0.07462271	0.06073380	0.23093552
b_1	5.90172268	5.73921605	4.58892946	5.21828736
c_1	1.17501185	1.04396983	1.37873881	0.90363555
a_2	0.06457449	0.01177053	0.02438558	0.05859873
b_2	6.41785671	6.92401726	4.62730519	7.42932932
c_2	1.93844484	1.66159571	2.08370698	1.51754610
a_3	−0.03567387	0.03720662	−0.02283430	—
b_3	2.80473127	6.26158944	2.05274659	—
c_3	1.63655241	1.63158721	1.86763816	—
a_4	—	−0.00270657	—	—
b_4	—	9.00003735	—	—
c_4	—	2.79557901	—	—
ζ_s (au)	1.08267366	1.85018803	2.37447159	3.17936914
ζ_p (au)	—	1.76830093	1.97812569	2.55361907

Table 4. Average Errors for the Various Properties Used in the Parameterization Procedure, for AM1, PM3, PM5, and RM1 Semiempirical Models for C, H, N, and O Atoms.

Properties	AM1	PM3	PM5	RM1	N^a
Enthalpies of formation (kcal · mol ⁻¹)	9.06	5.98	5.41	5.04	986
Dipole moment (D)	0.26	0.29	0.33	0.23	59
Ionization potential (eV)	0.48	0.55	0.44	0.42	102
Bond length (Å)	0.027	0.023	0.030	0.020	618
Angles (degree)	6.06	7.33	10.50	7.05	737

^a N refers to the number of quantities used in the comparison.

Similar to AM1, and unlike PM3 and PM5, however, we divided the parameterization problem into three parts: (1) we first considered only the leading atoms of organic chemistry: hydrogen, carbon, nitrogen, and oxygen. When we became satisfied with their parameters, we locked them, and (2) proceeded to consider phosphorus, and sulfur, thus taking into consideration most molecules of importance to life, and, again, when we became satisfied with their parameters, we also locked them; and, finally, (3) we proceeded to consider the halogens to make sure RM1 could be used in pharmaceutical research.

Results for H, C, N, O

Initially, we parameterized only the atoms H, C, N, and O. For that, we took into consideration 1094 species containing only these elements. Due to the scarceness of experimental data, not all of them had all their reference properties known. Accordingly, we considered enthalpies of formation of 986 species, dipole moments of 59, and ionization potentials of 102. For geometries, we considered 618 interatomic distances and 737 bond angles.

Table 3 presents the RM1 parameters for H, C, N, and O, and Table 4 presents the corresponding average errors of RM1 compared with AM1, PM3, and PM5 for the same molecular set of reference properties. Results from Table 4 indicate that RM1 possesses the least average errors for enthalpies of formation, dipoles, and ionization potentials. Indeed, in terms of average errors for enthalpies of formation, RM1 almost halved the average error of AM1 for the 986 species.

RM1 average errors in the dipole moments are smaller than the average errors of PM5 by almost a third. Predicting more accurate dipole moments implies that the predicted charge density is also more accurate, leading, for example, to better molecular electrostatic potential surfaces. Results from Table 4 display how the average errors for the dipole moments evolved from AM1 up to RM1, clearly indicating that the previous up trend was definitely reversed.

Predicting correct ionization potential implies that the energy of the highest occupied molecular orbital is also being correctly predicted, which is of importance for modeling chemical reactivity. Table 4 presents how the average errors for the ionization potentials evolved from AM1 up to RM1.

Correct geometries are essential for having correct charge densities and, from a density functional perspective, for having all other properties also correct. As indicated in Table 4, RM1 has the lowest average error for interatomic distances, its average error being ~67% of the average error for PM5.

On the other hand, AM1 is the model which has consistently the least average errors for bond and dihedral angles. Nevertheless, RM1 is successful in breaking the uptrend in the average error for the angles, which went from 6.06° for AM1, to 7.33° for PM3, and to 10.50° for PM5, down to 7.05° for RM1—a level closer to AM1, albeit slightly higher.

In summary, for H, C, N, and O, RM1 obtained a very low average error for the enthalpies of formation without compromising accuracy in any of the other properties.

During RM1 development, we took special care with the values of the parameters U_{ss} and U_{pp} to make sure that they reflect the electronegativities of the elements, thus guaranteeing that no imbalance of calculated charges in any of the atoms would result. Thus, the problem of the net charge in nitrogen, extant in PM3, is corrected in RM1. As is well known, PM3 net charges for amines are close to zero, whereas for nitro compounds they are too positive. Indeed, PM3 nitrogen charges for methylamine and nitromethane are, respectively, $-0.03e$ and $1.24e$, whereas the corresponding values for RM1 are $-0.44e$ and $0.40e$.

Results for P, S

Table 5 shows the RM1 parameters for P and S, and Table 6 shows the average errors of RM1 compared with AM1, PM3, and PM5 for the same set of 211 molecules of reference, from which we

Table 5. RM1 Parameters for P and S Atoms.

Parameter	P	S
U_{ss} (eV)	-41.81533184	-55.16775121
U_{pp} (eV)	-34.38342529	-46.52930422
β_s (eV)	-6.13514969	-1.95910719
β_p (eV)	-5.94442127	-8.77430652
α (Å ⁻¹)	1.90993294	2.44015636
G_{ss} (eV)	11.08059265	12.48828408
G_{sp} (eV)	5.68339201	8.56910574
G_{pp} (eV)	7.60417563	8.52301167
G_{p2} (eV)	7.40265182	7.66863296
H_{sp} (eV)	1.16181792	3.88978932
a_1	-0.41063467	-0.74601055
b_1	6.08752832	4.81038002
c_1	1.31650261	0.59380129
a_2	-0.16299288	-0.06519286
b_2	7.09472602	7.20760864
c_2	1.90721319	1.29492008
a_3	-0.04887125	-0.00655977
b_3	8.99979308	9.00000180
c_3	2.65857780	1.80060151
ζ_s (au)	2.12240118	2.13344308
ζ_p (au)	1.74327954	1.87460650

Table 6. Average Errors for the Various Properties Used in the Parameterization Procedure, for AM1, PM3, PM5, and RM1 Semiempirical Models for P and S Atoms.

Properties	AM1	PM3	PM5	RM1	N^a
Enthalpies of formation (kcal · mol ⁻¹)	12.60	14.64	9.11	7.40	167
Dipole moment (D)	0.74	0.56	0.58	0.49	16
Ionization potential (eV)	0.52	0.47	0.46	0.42	38
Bond length (Å)	0.069	0.061	0.076	0.050	84
Angles (degree)	6.75	6.39	10.18	7.77	72

^a N refers to the number of quantities used in the comparison.

compared 167 enthalpies of formation, 16 dipole moments, 38 ionization potentials, 84 interatomic distances, and 72 bond and dihedral angles.

The average errors in the enthalpies of formation for RM1 are 7.40 kcal · mol⁻¹, almost half the average error of PM3, and, again, the smallest of all models considered.

Likewise, the trend in the average errors for the dipole moments, going from AM1, PM3, and PM5 to RM1 is a descending one. The equivalent trend for the average errors for the ionization potentials is also a descending one. RM1, again, is the model with the least average errors for both properties.

The RM1 geometries of P and S compounds, mostly involving directly the P and S atoms, behave like the previous H, C, N, and O atom set. RM1 interatomic distances possess the least average error as is indicated in Table 6, correcting the upward trend from PM3 to PM5 present in this set of molecules. RM1 represents a 34% improvement in the average error for interatomic distances over PM5.

Table 8. Average Errors for the Various Properties Used in the Parameterization Procedure, for AM1, PM3, PM5, and RM1 Semiempirical Models for F, Cl, Br, and I Atoms.

Properties	AM1	PM3	PM5	RM1	N^a
Enthalpies of formation (kcal · mol ⁻¹)	16.71	10.62	6.30	7.12	327
Dipole moment (D)	0.37	0.43	0.67	0.42	52
Ionization potential (eV)	0.77	0.59	0.53	0.49	92
Bond length (Å)	0.051	0.038	0.045	0.039	202
Angles (degree)	3.94	4.88	4.61	4.43	101

^a N refers to the number of quantities used in the comparison.

On the other hand, RM1 P and S bond angles error, although correcting the large average error for PM5, is still higher than AM1 and PM3 by about 1 degree. This is similar to what happened with the H, C, N, and O set, and is probably due to the fact that we emphasized distances more than angles in the parameterization response function of RM1.

Results for F, Cl, Br, I

Table 7 presents the RM1 parameters for the halogens, and Table 8 the average errors for the properties of halogen containing compounds. For them, RM1 could only represent an overall improvement over the previous methods.

The average errors for the enthalpies of formation for the halogens as the methods evolved. RM1 could not better the average error for enthalpies of formation of PM5. For the halogens, that is the only case where RM1 loses in accuracy for PM5,

Table 7. RM1 Parameters for F, Cl, Br, and I Atoms.

Parameter	F	Cl	Br	I
U_{ss} (eV)	-134.18369591	-118.47306918	-113.48398183	-74.89997837
U_{pp} (eV)	-107.84660920	-76.35330340	-76.18720023	-51.41023805
β_s (eV)	-70.00000512	-19.92430432	-1.34139841	-4.19316149
β_p (eV)	-32.67982711	-11.52935197	-8.20225991	-4.40038412
α (Å ⁻¹)	6.00000062	3.69358828	2.86710532	2.14157092
G _{ss} (eV)	16.72091319	15.36023105	17.11563074	19.99974131
G _{sp} (eV)	16.76142629	13.30671171	15.62419251	7.68957672
G _{pp} (eV)	15.22581028	12.56502640	10.73546293	7.30488343
G _{p2} (eV)	14.86578679	9.66397083	8.86056199	6.85424614
H _{sp} (eV)	1.99766171	1.76489897	2.23512762	1.41602940
a_1	0.40302025	0.12947108	0.98689937	-0.08147724
b_1	7.20441959	2.97724424	4.28484191	1.56065072
c_1	0.81653013	1.46749784	2.00019696	2.00002063
a_2	0.07085831	0.00288899	-0.92731247	0.05914991
b_2	9.00001562	7.09827589	4.54005910	5.76111270
c_2	1.43802381	2.50002723	2.01617695	2.20488800
ζ_s (au)	4.40337913	3.86491071	5.73157215	2.53003753
ζ_p (au)	2.64841556	1.89593144	2.03147582	2.31738678

Table 9. Average Errors for All Properties for All Atomic and Molecular Species Considered in the Parameterization Procedure, for AM1, PM3, PM5, and RM1 Semiempirical Methods, for all 1736 Atomic and Molecular Systems Considered.

Properties	AM1	PM3	PM5	RM1	N^a
Enthalpies of formation (kcal · mol ⁻¹)	11.15	7.98	6.03	5.77	1480
Dipole moment (D)	0.37	0.38	0.50	0.34	127
Ionization potential (eV)	0.60	0.55	0.48	0.45	232
Bond length (Å)	0.036	0.029	0.037	0.027	904
Angles (degree)	5.88	6.98	9.83	6.82	910

^a N refers to the number of quantities used in the comparison.

though. However, like PM5, RM1 still represents a marked improvement over AM1 and PM3 for the enthalpies of formation for the halogens.

Consider the trend in the dipole moment errors for the halogen containing molecules from AM1 up to RM1. Although, unfortunately, RM1 could not beat AM1 in this important property for electrostatics, the upward trend from AM1 to PM5 has been severely reversed and RM1 presents an average error essentially equal to that of PM3, and considerably smaller than the number for PM5.

Table 8 presents the average errors for the ionization potentials, which display a monotonically downward trend from AM1 to RM1, with RM1 displaying an average error 36% smaller than that of AM1.

The situation with geometries is somewhat more entwined. Whereas PM3 represented a substantial improvement with respect to AM1, PM5 displayed larger average errors. RM1 reversed the trend downward to just above the PM3 level.

Again, AM1 strengthens itself as an excellent method in terms of angles. Although RM1 continued the descending trend from PM3 and passing through PM5, it could not lower the average error in the angles down to the AM1 level.

Nevertheless, considering the properties for the halogen-containing species altogether, RM1 represents an overall improvement over the previous methods, not displaying any distortions in any of the properties considered.

Overall Results

Having achieved parameters for all 10 atoms, now consider results for all 1736 systems as a whole. Table 9 presents average errors for 1480 enthalpies of formation, 127 dipole moments, 232 ionization potentials, 904 interatomic distances, and 910 bond and dihedral angles, for all 1736 atomic and molecular systems considered. Clearly, RM1 performs better than any other method in all properties with the exception of the angles, where only AM1 is about 1 degree better. This could have been expected from the results presented in the previous sections (Figs. 1–5).

To be able to discuss results for individual systems, let us consider the least-squares fit between calculated RM1 properties and their corresponding reference values. Figure 6 presents such

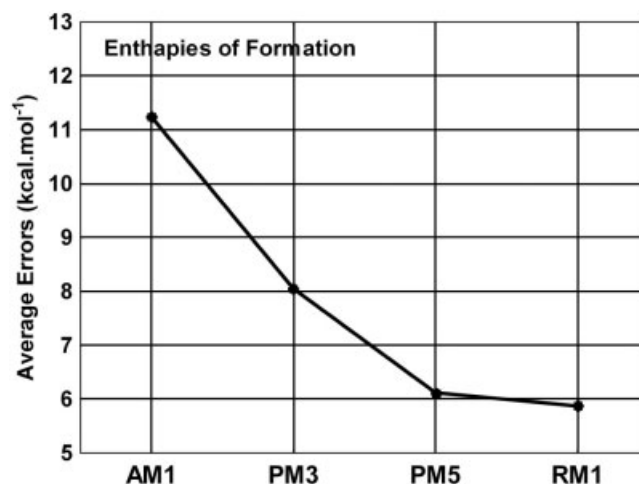


Figure 1. Average errors for enthalpies of formation for 1480 molecules, calculated with RM1, AM1, PM3, and PM5 semiempirical methods (H, C, N, O, P, S, F, Cl, Br, and I). The line which connects the points is present just to guide the eyes.

results for the enthalpies of formation, the property best described by RM1. Indeed, the angular coefficient of the fit is essentially 1, and the intercept is only 0.970 kcal · mol⁻¹. Of the 1480 enthalpies, we detect only 2, which have large errors. These systems also present large deviations for the other models used in this work as benchmarks for comparison. The one with the largest deviation, indicated as point (a) in Figure 6, is diphenyl disulfone, C₁₂H₁₀S₂O₄, which has an experimental enthalpy of formation of -115 kcal · mol⁻¹,⁴² and an RM1 value of -4 kcal · mol⁻¹. The corresponding values for AM1, PM3, and PM5, all in kcal · mol⁻¹, are, respectively, -10, -26, and -35. The second, indicated as point (b) in Figure 6, is phosphorus thiobromide, PSBr₃, which has

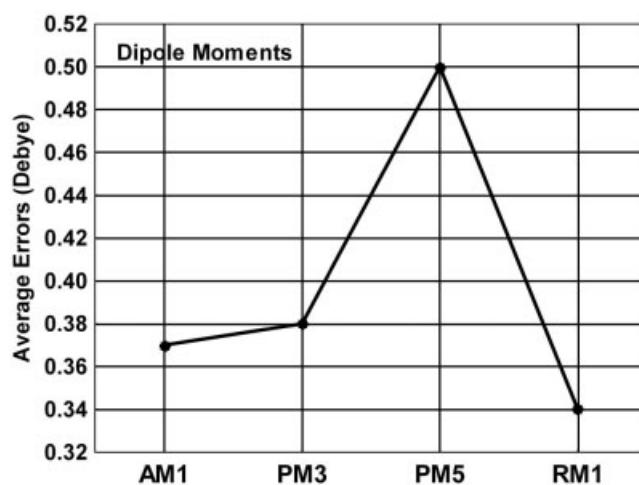


Figure 2. Average errors for dipole moments for 127 molecules, calculated with RM1, AM1, PM3, and PM5 semiempirical methods (H, C, N, O, P, S, F, Cl, Br, and I). The line that connects the points is present just to guide the eyes.

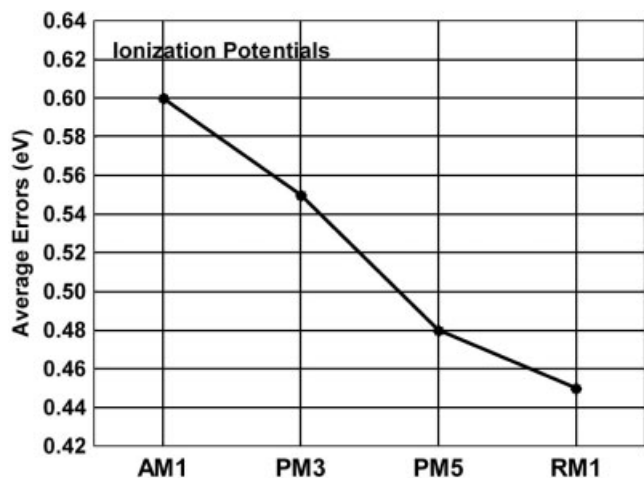


Figure 3. Average errors for ionization potentials for 232 molecules, calculated with RM1, AM1, PM3, and PM5 semiempirical methods (H, C, N, O, P, S, F, Cl, Br, and I). The line that connects the points in the graphic is present just to guide the eyes.

an experimental enthalpy of formation of $-67 \text{ kcal} \cdot \text{mol}^{-1}$,⁴³ and an RM1 value of $27 \text{ kcal} \cdot \text{mol}^{-1}$. The corresponding values for AM1, PM3, and PM5, all in $\text{kcal} \cdot \text{mol}^{-1}$, are, respectively, 45, -8, and -35. In this case, PM5 approaches much more the experimental value. In the two cases, RM1 behaves much more similarly to AM1 than to the other two models. For the other 1478 atomic and molecular systems, RM1 enthalpies of formation display admissible deviations.

Dipole moments are a much more difficult property to predict by semiempirical models. Figure 7 presents the linear fit between experimental and RM1 dipole moments for the 127 molecular systems considered. The squared correlation coefficient of 0.785 is

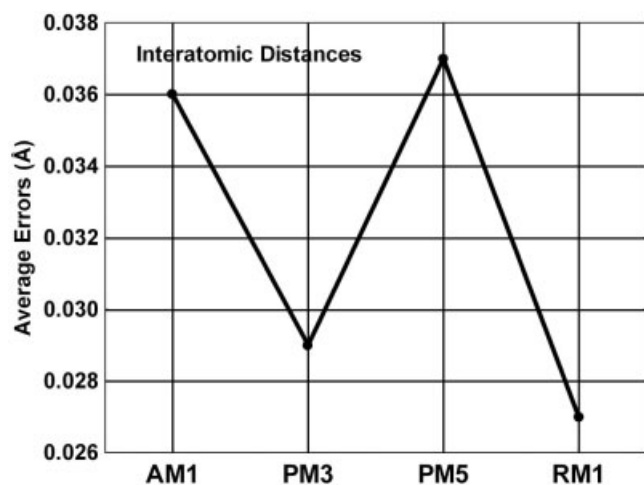


Figure 4. Average errors for interatomic distances for 904 quantities, calculated with RM1, AM1, PM3, and PM5 semiempirical models (H, C, N, O, P, S, F, Cl, Br, and I). The line that connects the points in the graphic is present just to guide the eyes.

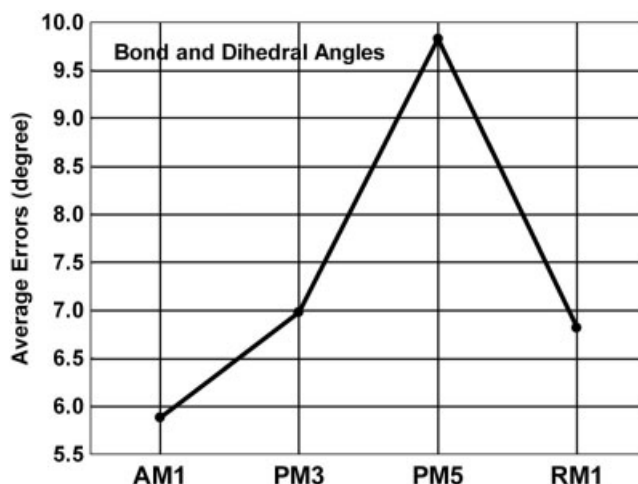


Figure 5. Average errors for bond and dihedral angles for 910 quantities, calculated with RM1, AM1, PM3, and PM5 semiempirical methods (H, C, N, O, P, S, F, Cl, Br, and I). The line that connects the points in the graphic is present just to guide the eye.

the smallest of all. The plot is much more dispersed than the plot for the enthalpies of formation. Nevertheless, we detect a single outlier that is significantly different from the others, the lone point (a) in the lower part of the plot, indicated by an arrow and corresponding to sulfur dioxide, which has an experimental dipole

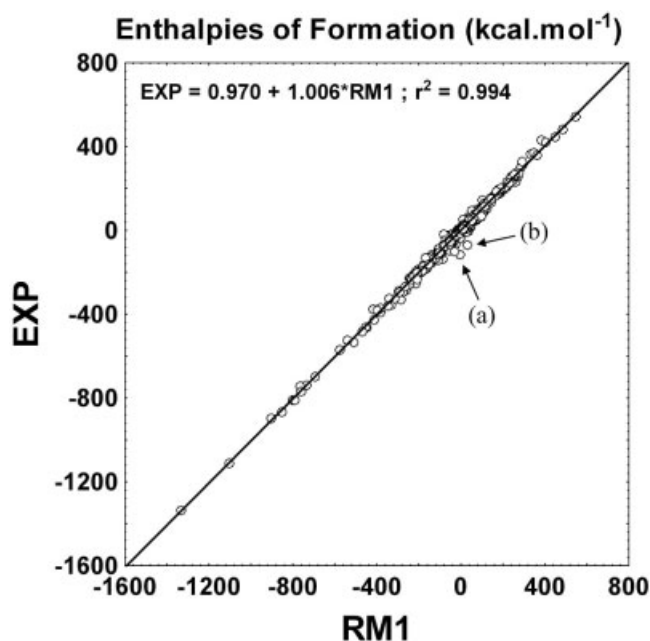


Figure 6. Linear regression between experimental enthalpies of formation and the corresponding RM1 predicted values for all 1480 atomic and molecular systems containing H, C, N, O, P, S, F, Cl, Br, and I atoms. Point (a) refers to diphenyl disulfone, $\text{C}_{12}\text{H}_{10}\text{S}_2\text{O}_4$, and point (b) to phosphorus thiobromide, PSBr_3 .

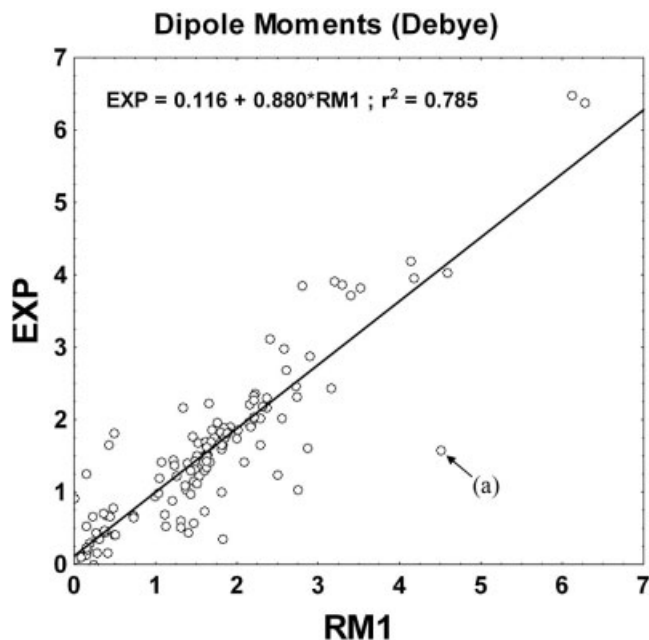


Figure 7. Linear regression between experimental dipole moments and the corresponding RM1 predicted values for all 127 molecular systems considered containing H, C, N, O, P, S, F, Cl, Br, and I atoms. Point (a) refers to sulfur dioxide, SO_2 .

moment of 1.58D⁴⁴ and the RM1 value is 4.51D. The corresponding values for AM1, PM3, and PM5, all in Debye and also overestimated are, respectively, 4.29, 3.63, and 3.73. Again, in this case, RM1 behaves more similarly to AM1.

Ionization potentials are also more disperse than enthalpies of formation as shown in Figure 8. However, we included in the set the fourth ionization potential of water: point (a) as indicated in the uppermost right part of Figure 8. This inclusion was to verify whether or not its RM1 value would be too deviant of the trend for all 232 ionization potentials considered. Not only did it not compromise the trend, but also it did not significantly leverage the regression. The slope of the regression is 0.928, close to 1, together with a squared correlation coefficient of 0.915, further evidence that RM1 ionization potentials can be truly considered as predictions of experimental ionization potentials. The RM1 ionization potential with the largest deviation from the experimental value is point (b) of Figure 8, corresponding to 1,2-dibromotetrafluoroethane, $\text{C}_2\text{Br}_2\text{F}_4$, which has an experimental ionization potential of 14.4 eV,⁴⁵ whereas the RM1 value is 10.9 eV. The corresponding values for AM1, PM3, and PM5, all in eV and also underestimated are, respectively, 11.5, 11.0, and 11.7.

Interatomic distances are the overall second best property predicted by RM1. This can be seen from the least-squares linear regression between RM1 and reference (mostly experimental) interatomic distance values. Indeed, the squared correlation coefficient is 0.968 for the 904 (mostly experimental) interatomic distances and the slope of the regression is 0.970, very close to the ideal value of 1. There are three points indicated in Figure 9 as (a), (b), and (c), which show larger deviations. Point (a), is the OH

hydrogen bond distance of hydrogen peroxide, which has an experimental length of 2.032 Å⁴⁶ and the RM1 value is 2.483 Å. The corresponding values for AM1, PM3, and PM5 are, respectively 2.094, 1.809, and 2.681 Å. Point (b), also below the regression line, is the sulfur–bromine bond length of sulfur pentafluoride bromide, SBrF_5 , which has an experimental value of 2.133 Å⁴⁷ and the RM1 value is 2.423 Å, while the corresponding values for AM1, PM3, and PM5 are, respectively 2.594, 2.426, and 2.400 Å. Finally, point (c) is the NN bond length of dinitrogen trioxide which has an experimental value of 2.08 Å⁴³ and whose RM1 value is 1.640 Å. The corresponding values for AM1, PM3, and PM5 are, respectively 1.582, 1.697, and 1.622 Å. Seemingly, for the case of bond distances and unlike what we detected for the previous properties, there are no easy patterns in interatomic distances that reproduce across the models.

As could have been expected, bond angles are mostly clustered around the 100 degrees region as displayed in Figure 10. The squared correlation coefficient is 0.908 for the 641 (mostly experimental) bond angles, and the slope of the regression is 0.922, close to the ideal value of 1. There is one clear outlier, indicated by the arrow in Figure 10, which represents the $\text{SN}=\text{C}$ bond angle of pentafluoro (isocyanato) sulfur, CNSOF_5 , which possesses the experimental value of 124.9°, while the RM1 predicted value is 180.0°. However, this deviation also happens with AM1 and PM5, where both predict this value to be 180.0°. Only PM3 predicts a better value: 152.9°.

Finally, we now examine the accuracy of the various models with respect to the properties, categorized by the atom type involved. Table 10 presents the average errors for the enthalpies of

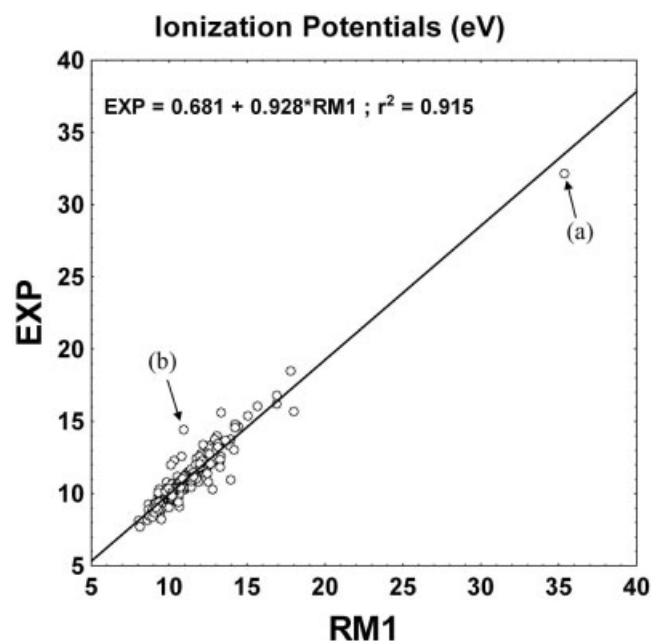


Figure 8. Linear regression between experimental ionization potentials and the corresponding RM1 predicted values for all 232 atomic and molecular systems containing H, C, N, O, P, S, F, Cl, Br, and I atoms. Point (a) refers to the fourth ionization potential of water; and point (b) to 1,2-dibromotetrafluoroethane, $\text{C}_2\text{Br}_2\text{F}_4$.

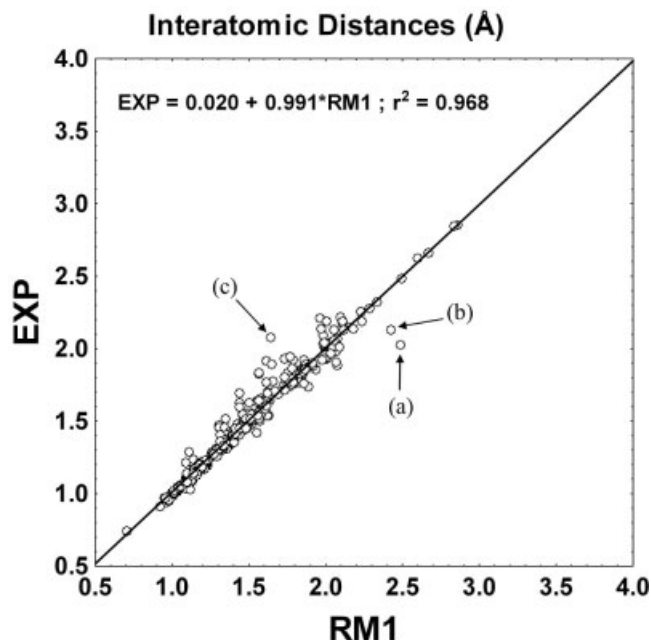


Figure 9. Linear regression between 904 reference (mostly experimental) interatomic distances and the corresponding RM1 predicted values for the molecular systems considered, containing H, C, N, O, P, S, F, Cl, Br, and I atoms. Point (a) refers to the OH distance of hydrogen peroxide; point (b) refers to the sulfur-bromine bond length of sulfur pentafluoride bromide, SBrF_5 ; and point (c) refers to the NN distance of dinitrogen trioxide.

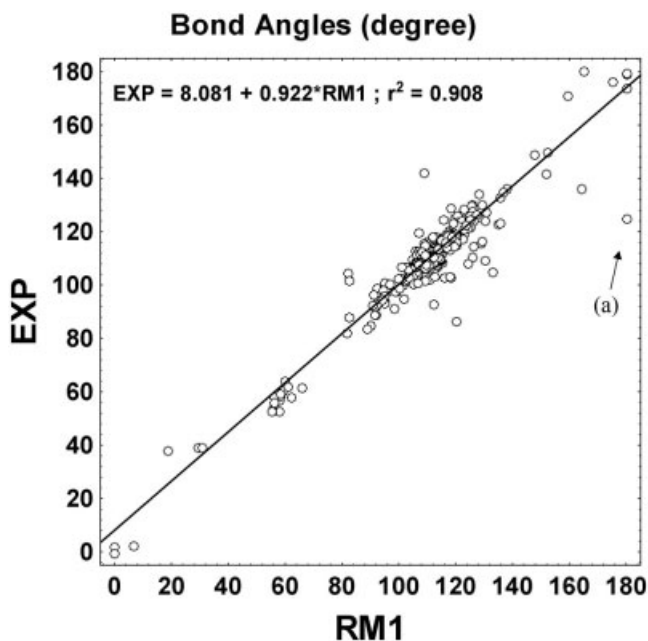


Figure 10. Linear regression between 641 reference (mostly experimental) bond angles and the corresponding RM1 predicted values involving H, C, N, O, P, S, F, Cl, Br, and I atoms. Point (a) refers to the $\text{SN}=\text{C}$ bond angle of pentafluoro (isocyanato) sulfur, CNSOF_5 .

Table 10. Average Errors for the Enthalpies of Formation, in $\text{kcal} \cdot \text{mol}^{-1}$, for AM1, PM3, PM5, and RM1 Semiempirical Methods, Categorized by Atom Type Present.

Element	AM1	PM3	PM5	RM1	N [†]
Hydrogen	8.79	6.33	5.32	4.94	1246
Carbon	9.31	6.40	5.39	5.08	1299
Nitrogen	10.68	6.70	6.00	5.49	481
Oxygen	11.44	7.80	6.19	6.02	691
Fluorine	22.35	11.34	7.56	8.44	165
Phosphorus	17.66	21.13	15.04	15.19	60
Sulfur	13.86	11.76	7.17	7.16	151
Chlorine	13.76	8.12	6.39	6.28	94
Bromine	17.83	9.35	5.64	7.76	53
Iodine	17.57	11.96	4.32	3.60	58

The lesser values are in bold.

[†]N refers to the number of atomic and molecular systems used in the comparison.

formation where RM1 presents the least average errors of all models for H, C, N, O, S, Cl, and I. Only for F, P, and Br, PM5 is more accurate than RM1. It is noticeable that RM1 average errors for enthalpies of formation, for all atoms, are smaller than those for PM3.

With respect to dipole moments, RM1 displays the smallest errors for H, C, N, O, S, and I. PM3 has the smallest error for dipoles of P containing molecules, and AM1 for F, Cl, and Br containing molecules (Table 11).

Now, considering ionization potentials as in Table 12, the average errors for all atoms for RM1 are smaller than all those for AM1 and PM3. In comparison with PM5, RM1 has the lowest average errors for H, C, N, F, P, Br, and I. Only for O, S, and Cl, PM5 presents less average errors.

For interatomic distances as displayed in Table 13, RM1 possesses the lowest average errors of all in H, C, N, P, S, and I containing compounds. For iodine compounds, RM1 displays a

Table 11. Average Errors for the Dipole Moments, in Debyes, for AM1, PM3, PM5, and RM1 Semiempirical Methods, Categorized by Atom Type Present.

Element	AM1	PM3	PM5	RM1	N ^a
Hydrogen	0.31	0.30	0.40	0.25	102
Carbon	0.30	0.30	0.44	0.24	99
Nitrogen	0.48	0.43	0.42	0.37	35
Oxygen	0.35	0.46	0.53	0.35	45
Fluorine	0.39	0.40	0.80	0.40	30
Phosphorus	0.75	0.41	0.87	0.54	8
Sulfur	0.65	0.61	0.61	0.58	12
Chlorine	0.29	0.74	0.71	0.73	6
Bromine	0.26	0.39	0.51	0.41	15
Iodine	0.45	0.40	0.70	0.40	8

The lesser values are in bold.

^aN refers to the number of molecular systems used in the comparison.

Table 12. Average Errors for the Ionization Potentials, in eV, for AM1, PM3, PM5, and RM1 Semiempirical Methods, Categorized by Atom Type Present.

Element	AM1	PM3	PM5	RM1	N [†]
Hydrogen	0.50	0.50	0.38	0.35	164
Carbon	0.52	0.48	0.39	0.36	180
Nitrogen	0.46	0.55	0.43	0.43	39
Oxygen	0.69	0.64	0.58	0.59	67
Fluorine	0.79	0.69	0.65	0.60	51
Phosphorus	0.61	0.64	0.52	0.41	14
Sulfur	0.67	0.54	0.53	0.60	44
Chlorine	0.68	0.67	0.52	0.58	26
Bromine	0.63	0.67	0.40	0.39	15
Iodine	1.11	0.30	0.42	0.22	15

The lesser values are in bold.

[†]N refers to the number of molecular systems used in the comparison.

significantly reduced average error of only 0.018 Å. PM3 has the least average errors for O and F, and, PM5, for Cl and Br. Nevertheless, in all four cases the differences between RM1 and PM3 are very small: 0.003 Å for O, and 0.006 Å for F.

For the analysis of the angles, we only considered bond angles where the atom is at the vortex of the angle, and dihedral angles where the atom is at the intersection of the planes. In this case the situation is very mixed, with RM1 possessing the least average errors for P, S, and I, as shown in Table 14. Again, for I, RM1 errors for the angles is the extremely low value of 0.02°, thus reinforcing the fact that RM1 seems to be an excellent model for iodine containing compounds. AM1 possesses the least average errors for N, Cl, and Br; PM3 for H, and O; and PM5 for C, and F.

Overall, considering Tables 10 to 14, RM1 is the most accurate model in 29 cases, PM5 in 10 cases, PM3 in 5 cases, and AM1 in 6 cases. Thus, RM1 is the most accurate model in 58% of the cases, PM5 in 20%, PM3 in 10%, and AM1 in 12% of the cases.

Table 13. Average Errors for the Interatomic Distances Categorized by Atom Type, in Angstroms, for AM1, PM3, PM5, and RM1 Semiempirical Methods.

Element	AM1	PM3	PM5	RM1	N [†]
Hydrogen	0.074	0.047	0.091	0.044	149
Carbon	0.018	0.019	0.019	0.016	672
Nitrogen	0.032	0.033	0.027	0.022	296
Oxygen	0.054	0.037	0.064	0.040	217
Fluorine	0.048	0.033	0.055	0.039	65
Phosphorus	0.050	0.047	0.059	0.047	54
Sulfur	0.067	0.053	0.045	0.039	65
Chlorine	0.066	0.043	0.039	0.064	34
Bromine	0.097	0.076	0.053	0.070	30
Iodine	0.090	0.059	0.161	0.018	12

The lesser values are in bold.

[†]N refers to the number of interatomic distances used in the comparison.

Table 14. Average Errors for Bond and Dihedral Angles, in Degrees, for AM1, PM3, PM5, and RM1 Semiempirical Methods, Categorized by Atom Type.

Element	AM1	PM3	PM5	RM1	N ^a
Hydrogen	14.21	10.37	35.21	11.87	4 ^b
Carbon	2.32	2.38	1.66	1.85	369
Nitrogen	2.30	2.36	2.83	2.98	119
Oxygen	4.75	3.62	8.81	5.23	56
Fluorine	18.16	38.98	11.35	16.26	1 ^c
Phosphorus	2.81	2.91	2.57	2.50	27
Sulfur	3.66	4.85	3.30	2.98	31
Chlorine	17.77	20.06	24.57	24.18	2 ^d
Bromine	4.92	16.45	16.01	17.85	3 ^e
Iodine	4.56	20.65	12.57	0.02	1 ^f

Only bond angles where the atom is at the vertex of the angle, or dihedral angles where the atom is at the intersection of the planes are considered. The lesser values are in bold.

^aN refers to the number of bond and dihedral angles used in the comparison.

^bWater dimer (two different OHH angles), adenine–uracyl (OHN), and adenine–thymine (OHN).

^cHydrogen fluoride dimer (HFH).

^dChlorine trifluoride (FCIF) and perchloryl fluoride (OCIF).

^eBromine trifluoride (FBrF), perbromyl fluoride (FBrO), and bromine pentafluoride (FBrF).

^fIodine pentafluoride (FIF).

Conclusion

RM1 was created as a modern reparameterization of AM1—that is, as an attempt to maintain the mathematical structure and qualities of AM1, while significantly improving its quantitative accuracy with the help of today's computers and also of the more advanced techniques available for nonlinear optimization.

However, RM1 ended up also representing an overall improvement over PM3, and being at least very competitive with PM5, which is a somewhat different model, and not fully disclosed.

Because RM1 possesses the same analytical construct and the same number of parameters for each atom as AM1, RM1 can therefore be easily implemented in any software which already has AM1, not requiring any change in any line of code, with the sole exception of the values of the parameters themselves.

Given the large amounts of existing software that run AM1 and PM3 models, RM1 may eventually prove to be of value simply by easily enhancing their quantitative accuracies, even though many qualitative problems, extant in AM1, still remain unresolved, such as the ϕ – ψ barriers in peptides. Therefore, we must also state that, from an absolute perspective, RM1 was only partially successful as an attempt to develop an improved model.

On the other hand, for example, RM1 can be easily implemented in MOZYME¹⁷ and LocalSCF¹⁸ because these softwares are coded to scale linearly semiempirical calculations under the zero differential overlap approximation. Thus, RM1 may perhaps prove to be of value for giant biomolecule modeling and genomics research—something that would not be immediately possible if we had decided, for example, to add full overlap to RM1.

Supplementary Information Available

Instructions and examples on how to use the RM1 model in MOPAC with the help of the external file RM1.rm1, which is also provided. Additional figures displaying how the average errors evolved from AM1, passing through PM3 and PM5, up to RM1. Least-squares linear regression plots for all properties considered for AM1, PM3, and PM5. Tables containing reference and calculated values for AM1, PM3, PM5, and RM1 for each of the 1480 enthalpies of formation, 127 dipole moments, 232 ionization potentials, 904 interatomic distances, and 910 bond and dihedral angles.

Part of the material in this article was presented by Alfredo M. Simas in his thesis submitted to the Open Competitive Examination for a Full Professorship Position at Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife, Brazil, November of 2003. We thank CENAPAD (Centro Nacional de Processamento de Alto Desempenho) at Campinas, Brazil, for having made available to us their computational facilities. We also acknowledge the Cambridge Crystallographic Data Centre for the Cambridge Structural Database.

References

- Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J Am Chem Soc* 1985, 107, 3902.
- Stewart, J. J. P. *MOPAC 93.00 Manual*; Fujitsu Limited: Tokyo, Japan, 1993.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
- Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*, University of California, San Francisco, 2004.
- Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J Comput Chem* 2000, 21, 1088.
- Rowley, C.; Hassinen, T. *Ghemical for GNOME*, version 1.00; University of Iowa: Iowa City, IA, 2002.
- Jorgensen, W. L. *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Editor-in-chief; John Wiley & Sons Ltd.: Athens, 1998, p. 3281, vol. 5.
- Stewart, J. J. P. *J Comput Chem* 1989, 10, 209.
- Stewart, J. J. P. *J Comput Chem* 1989, 10, 221.
- Stewart, J. J. P. *MOPAC 2002*; Fujitsu Limited: Tokyo, Japan, 1999.
- Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. *Phys Rev B* 1998, 58, 7260.
- Weber, W.; Thiel, W. *Theor Chem Acc* 2000, 103, 495.
- Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J Comput Chem* 2002, 23, 1601.
- Moura, G. L. C.; Simas, A. M.; Miller, J. *Chem Phys Lett* 1996, 257, 639.
- Caffisch, A.; Karplus, M. *Perspect Drug Discov Des* 1995, 3, 51.
- Cichos, F.; Brown, R.; Bopp, P. A. *J Chem Phys* 2001, 114, 6824.
- Stewart, J. J. P. *Int J Quantum Chem* 1996, 58, 133.
- Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Adreyev, A. M. *J. Chem Phys* 2004, 121, 1266.
- Ohno, K.; Kamiya, N.; Asakawa, N.; Inoue, Y.; Sakurai, M. *J Am Chem Soc* 2001, 123, 8161.
- Ohno, K.; Kamiya, N.; Asakawa, N.; Inoue, Y.; Sakurai, M. *Chem Phys Lett* 2001, 341, 387.
- Nakajima, S.; Ohno, K.; Inoue, Y.; Sakurai, M. *J Phys Chem B*, 2003, 107, 2867.
- Yoda, M.; Inoue, Y.; Sakurai, M. *J Phys Chem B* 2003, 107, 14569.
- Titmuss, S. J.; Cummins, P. L.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E. *Chem Phys Lett* 2000, 320, 169.
- Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. *J Comput Chem* 2002, 23, 1314.
- Hutter, M. C.; Helms, V. *Int J Quantum Chem* 2003, 95, 479.
- Yang, W.; Drueckhammer, D. G. *J Phys Chem B* 2003, 107, 5986.
- Kurita, N.; Sengoku, Y.; Sekino, H. *Chem Phys Lett* 2003, 372, 583.
- Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. *Int J Quantum Chem* 2004, 97, 747.
- Pichierri, F.; Serai, A. *Chem Phys Lett* 2000, 322, 536.
- Yang, W.; Drueckhammer, D. G. *J Phys Chem B* 2003, 107, 5986.
- Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J. K.; Spudich, J. L.; Birge, R. R. *Biochemistry* 2001, 40, 13906.
- Pichierri, F. *Biophys Chem* 2004, 109, 295.
- Brandt, W.; Dessoy, M. A.; Fulhorst, M.; Gao, W. Y.; Zenk, M. H.; Wessjohann, L. A. *Chem Biochem* 2004, 5, 311.
- Pichierri, F. *J Mol Struct (Theochem)* 2003, 664, 197.
- Linnanto, J.; Korppi-Tommola, J. *J Comput Chem* 2004, 25, 123.
- Anisimov, V. M.; Anikin, N.; Bugaenko, V.; Bobrikov, V.; Andreyev, A. *Theor Chem Acc* 2003, 109, 213.
- Cummins, P. L.; Titmuss, S. J.; Jayatilaka, D.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E. *Chem Phys Lett* 2002, 352, 245.
- Bliznyuk, A. A.; Rendell, A. P.; Allen, T. W.; Chung, S. H. *J Phys Chem B* 2001, 105, 12674.
- Liu, H. Y.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. T. *Proteins Struct Funct Genet* 2001, 44, 484.
- Nelder, J. A.; Mead, R. *Comput J* 1965, 7, 308.
- Stewart J. J. P. *Am Chem Soc Abstr* 2001, 222, 86.
- W. G. Mallard, P. J. Linstrom, Eds.; *NIST Chemistry WebBook*, NIST Standard Reference Database, No. 69; National Institute of Standards and Technology: Gaithersburg (<http://webbook.nist.gov/chemistry>).
- Chase, M. W.; Davies, C. A.; Downey, J. R.; Frurip, D. R.; McDonald, R. A.; Syverud, J. A. N. *Thermochemical Tables*, 3rd ed. *J Phys Chem* 1985, Ref. Data 14, Suppl. 1.
- McClellan, A. L. *Tables of Experimental Dipole Moments*; Rahara Enterprises: El Cerrito, 1974, vol. 2.
- Levin, R. D.; Lias, S. G. *Ionization Potentials and Appearance Potential Measurements 1971–1981*; Natl. Stand. Ref. Data Ser., Natl. Bur. Stand. 71, 1982, Cat. No. C13.48:71.
- Kuchitsu, K.; Fukuyama, T.; Marino, Y. *J Mol Struct* 1968, 1, 463.
- Callomon, J. H.; Hirota, E.; Iijima, T.; Kuchitsu, K.; Lafferty, W. J. *Structure Data on Free Polyatomic Molecules*; New Series, Group II; Landolt-Bornstein: Springer, Berlin, 1977, vol. 15.