

Research Article

An Assessment of Alternative Low Level Calculation Methods for the Initial Selection of Conformers of Diastereomeric Esters

Andrus Metsala,¹ Sven Tamp,¹ Kady Danilas,¹ Ülo Lille,¹ Ly Villo,¹ Sirje Vija,² Tõnis Pehk,² and Omar Parve¹

¹ Department of Chemistry, Tallinn University of Technology, Ehitajate Tee 5, 19086 Tallinn, Estonia

² National Institute of Chemical Physics and Biophysics, Akadeemia Tee 23, 12618 Tallinn, Estonia

Correspondence should be addressed to Andrus Metsala; metsala@chemnet.ee

Received 21 January 2014; Revised 25 May 2014; Accepted 27 May 2014; Published 23 June 2014

Academic Editor: Qingfeng Ge

Copyright © 2014 Andrus Metsala et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Critical assessment of performance of alternative molecular modeling methods depending on a specific object and goal of the investigation is a question of continuous interest. This prompted us to demonstrate the origin of the guidelines we have used for a rational choice and use of a proper low level calculation method (LLM) for an initial geometry optimization of generated conformers, with the aim of selecting a set for further optimization. What was performed herein was a comparison of LLMs: MM3, MM+, UFF, Dreiding, AM1, PM3, and PM6 on the optimization of conformers' geometry of α -methoxyphenylacetic acid (MPA) 2-butyl esters as a set of typical diastereomeric esters of a chiral derivatizing agent. This set of esters calculated represents only compounds of this certain type in the current work. The LLM conformer energies were correlated with benchmark energies found by using higher level reference method B3LYP/6-311++G** on the geometries gained previously by optimization with LLMs. In an alternative treatment, the energy range to be covered and corresponding number of LLM optimized conformers obligatory for submitting to further optimization using a high level optimization cascade were considered on the basis of determination of the cut-off conformer (COFC).

1. Introduction

Critical assessment of performance of different alternative molecular modeling methods depending on the particular objects and specific goals is a question of continuous interest [1–7]. The aim of the current work is to demonstrate how the guidelines are derived for a rational choice of a proper low level calculation method (LLM) for the initial geometry optimization of generated conformers in the total conformational analysis of diastereomeric esters. The LLM chosen should allow selecting of a reliable set of conformers for further higher level optimization. Having previously followed such guidelines in the total conformational analysis of several esters with up to 11 dihedrals in their structure, the initial optimization of thousands of conformers all together have afforded reliable sets of conformers. The following higher level optimization of the selected conformers has led to the results that, in turn, have allowed calculation of the molecular

shielding models in exceptional accordance with differential shielding effects in the NMR spectra [8].

The properties of organic compounds are dependent on their three-dimensional structures. Knowing the prevailing conformer [1] of certain compounds or population of reactive conformers of molecules of another type is often a key to understanding their spectral properties or stereochemistry of the reactions under various experimental conditions, respectively. The most frequently used approach is to study the conformers' geometry and relative energy as well as their barriers to the rotational interconversion.

The theoretical conformational analysis enables a reliable estimation of the geometrical structure and energy of each conformer in a set of conformers and from such results various thermochemical parameters and spectral characteristics (NMR coupling constants and chemical shifts, IR vibration modes, etc.) can be derived. Density functional theory (DFT) [9] calculations have proved successful in

numerous studies based on theoretical conformational analysis [10]. Although geometrical parameters such as bond lengths, angles and dihedrals can be reasonably accurately predicted on molecular mechanical and semiempirical levels of theory, the electron correlation effects are important for quantitative considerations. These effects significantly affect the conformer energies as well as rotational barriers between various conformers and should therefore certainly be taken into account in the total conformational analysis.

DFT and the hybrid B3LYP functional have allowed accurate computation of the geometries concerning the type of molecules under study [8]. This, in turn, has allowed calculating NMR and IR spectral characteristics in very good agreement with the experimental data [8, 11, 12]. The DF method using the B3LYP hybrid functional has been proven to give a good estimate of the electron correlation for some molecules, in particular, in conformational studies, when it is used with the 6-31G* or 6-311++G** basis set [13, 14].

Semiempirical methods (AM1, PM3) have been shown to overestimate the number of minima on the potential energy surface (PES) and underestimate the internal rotational barriers for substituted hydrazines [13]. Several force fields have been assessed by comparing calculated conformational energies in relation to MP2/aug-cc-pVTZ level benchmark energies for compounds with varying polarity. The quality of the results degraded along with increasing polarity of the molecule. This effect was attributed to insufficient description of the electrostatic term by using fixed partial charges in force fields [15].

The attempts to simplify conformational analysis are always problematic. In particular, limiting the number of conformations or setting certain geometrical constraints on the molecules under study can give a poor description of the PES leading to erroneous results. Therefore, if the subject molecule is conformationally flexible and contains many dihedrals with low potential energy barriers, a systematic search for conformers should be undertaken.

Identification of and involving all low energy conformers is important in a sense that compounds' certain physico-chemical properties, as a rule, depend on the whole conformational ensemble, not just on the conformer corresponding to the global minimum on the PES. However, investigation of only the energetically most favored conformer may give good results in some cases [1]. An example of this approach is reproducing NMR spectral parameters (chemical shifts, coupling constants) of flexible organic compounds at the DFT level using the Boltzmann distribution [16]. A systematic search for conformers and also taking into account the Boltzmann weighting of individual conformers have allowed the calculation of the circular dichroism spectra of macrocyclic lactones (by using the time dependent DFT method) in a very good correlation with experimental spectra [17]. The same approach has allowed an accurate estimation of the yield of hydroxybenzophenones in the Fries rearrangement of hydroxyphenyl benzoates [14].

Usual procedure of conformational analysis involves the generation of conformers by automated means using a conformational search algorithm (stochastic Monte Carlo, simulated annealing, genetic algorithm, etc.), followed by

the initial optimization of the geometry of the generated structures with the help of molecular mechanical or a semiempirical method. The further refinement of conformers is carried out on higher theoretical levels, at the same time discarding higher energy conformers. After optimization of all structures at a certain level of theory, a geometrical analogy comparison is used to identify duplicates. This is carried out by RMS values of geometrical deviation of the positions of the atoms of one conformer in relation to other conformers [18]. The cascade of calculations is usually finalized by using a higher level DFT method.

The purpose of the present study is to show how the choice of the initial low level optimization method and the details of its use can influence the outcome of the total conformational analysis consisting of a cascade of optimizations, including relatively high-level calculations at the final stage. More specifically, the aim is to elucidate which initial method permits reducing most efficiently the number of conformers necessary to be optimized at the higher level. Another aim is to provide certain criteria that could be relied on in order to guarantee obtaining reliable results along with an optimized feasibility of calculations. In this regard, the key feature is the estimation of the conformer energy range for a certain initial calculation method marking the conformers to be selected for further calculations. This should guarantee that the analyzer does not miss any important conformer at the initial stage of optimization.

As to the contents, this paper will firstly tackle the question of main correlational parameters such as slope, intercept, and the coefficient of correlation between the results of tested LLMs and the benchmark energies calculated using higher level reference method (RefM1). A satisfactory correlation is obligatory for an LLM to be used in the initial stage of conformational analysis. However, these correlational parameters are still, by themselves, insufficient for a correct evaluation of the prospective quality of calculation results to be obtained just by using the LLM. Secondly, the estimation of the energy ranges for the initially LLM-optimized conformers to be selected for further optimization is considered.

In addition, a critical consideration is given to the approach within the systematic theoretical conformational analysis, according to which decisive conclusions about the relative energies and corresponding distribution of conformers are drawn on the basis of results obtained using merely an LLM. The accuracy of the LLMs for calculation of conformational energies as compared to the benchmark values has been assessed.

For the current research, α -methoxyphenylacetic acid (MPA) 2-butyl esters (Figure 1) were chosen as subject compounds. They are diastereomeric esters of the well-known chiral derivatizing agent for NMR spectroscopic analysis [8, 19–21]. In the current work they are mere model compounds for calculation (Figure 1). The MPA esters are not too elementary in conformational sense; yet the optimization of their conformers is possible by using higher level methods, such as B3LYP/6-311++G** included into the cascade protocol for optimization (RefM2) in the current work.

The calculations have been performed in gas phase. This is justified by very good accordance achieved in between

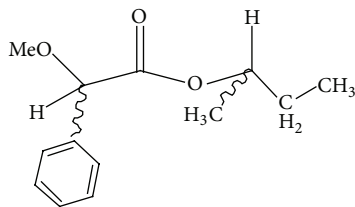


FIGURE 1: 2-Butyl esters of MPA.

the shielding models calculated in this environment and the NMR data measured in the CDCl_3 [8].

2. Theoretical Approach

2.1. Abbreviations

PES: Potential energy surface

MPA: α -Methoxyphenylacetic acid

LLM: Low level calculation method

RefM1: The first reference method is “the single point calculation of conformer energy using B3LYP/6-311++G** method” (Step 3, Figure 2); this is the single point calculation of conformer energy for the conformers’ geometries optimized with LLMs (Steps 1 and 2; Figure 2)

RefM2: The second reference method is “the high level conformational analysis using cascade procedure for optimization of conformers” (Step 4, Figure 2)

RefM3: B97D Grimme’s functional including dispersion; the basis set is the same as RefM2 (6-311++G**)

COFC: “Cut-off conformer”; this is a conformer characterized by the highest conformer energy in a series of LLM optimized conformers selected for further calculation (Steps 1 and 2, Figure 2); these conformers have been selected as corresponding to the lower energy conformers (up to 1.00 kcal/mol) that were found by optimization of all conformers using the high level RefM2 (Step 4, Figure 2). The COFC energy determines an energy range for conformers of an LLM potentially involving all lower energy conformers to be submitted for further optimization.

2.2. Computational Strategy. In order to achieve the goals of the research, the following five-step computational procedure was performed (Figure 2).

(1) The initial task was to find all the local minima over all configurations of the stereocenters of the MPA diastereomeric esters (Figure 1). The conformational landscape of the PES has been explored using the program Tinker (Figure 2) [22] (module SCAN). This program performs a general conformational search for the entire PES by using a basin hopping algorithm [23] which is able to overcome potential energy barriers. The basin hopping algorithm explores the PES along various normal modes from the initial local

minimum. When all minima on the search list have been subjected to the normal mode activation without locating additional new minima, the program terminates. Thus, the scanning time for this program is not unlimited and it also minimizes the energy of each conformer after it has been generated. The MM3 force field chosen for this minimization is known to be well parameterized for many organic compounds [24].

(2) All the conformers generated by the Tinker program were optimized with MM3 (Step 1, Figure 2). The conformers obtained were further optimized also with all the other LLMs under test: molecular mechanical, MM+, UFF, Dreiding, and semiempirical methods, AM1, PM3, PM6 (Step 2, Figure 2). The initial optimization of the conformers in this work with MM3 may, in principle, bias the results towards this method. This is because conformations may be lost in this first reminimization step which would be included if separate conformation searches were done with each force field. However, this assumption is not valid. It has been proven by exceptional accordance of experimental NMR shielding data [8] with the results of the theoretical calculations performed ignoring the matter of the above assumption. Consequently, the MM3 conformer energies might be comparable without restrictions with the conformer energies obtained using other LLMs.

(3) The quality of the LLMs was initially evaluated (the conformer energies correlated) on the basis of single point DFT B3LYP/6-311++G** calculation of conformer energies (RefM1) (Step 3, Figure 2) on the previously optimized force field and semiempirical (LLM) geometries.

(4) The conformers resulting from the above conformational search (from Step 1, Figure 2) were further geometry optimized in a sequential treatment (Step 4, Figure 2) with the help of the Gaussian 09 program [25]. The initial computational level used was the Hartree-Fock method with an STO-3G basis set. These *ab initio* calculated structures were used as input structures for the density functional method using the hybrid B3LYP exchange correlation functional and 3-21G basis set (i.e., B3LYP/3-21G). Similarly, the output of the latter method was used as an input for the B3LYP/6-31G method. The optimization cascade was finalized at the B3LYP/6-311++G** level. This cascade methodology for conformational analysis is referred to in the current text as RefM2.

Every optimization step was followed by a geometrical comparison of conformers with the aim of extracting unique structures based on a geometrical RMS criterion ($<2.0 \text{ \AA}$). Progressing towards higher levels of calculation, some of the conformers converge on each other resulting in substantially reduced number of conformers to be optimized at the highest level.

Additionally, the frequency analysis was performed at the highest optimization level. This was done in order to verify that every optimized conformer corresponds to a local minimum on the PES, that is, to a stationary point with no imaginary frequencies.

(5) All the conformers obtained by means of the above optimization cascade (Step 4, Figure 2) that fall into energy range of 0–1.00 kcal/mol were subjected to the optimization

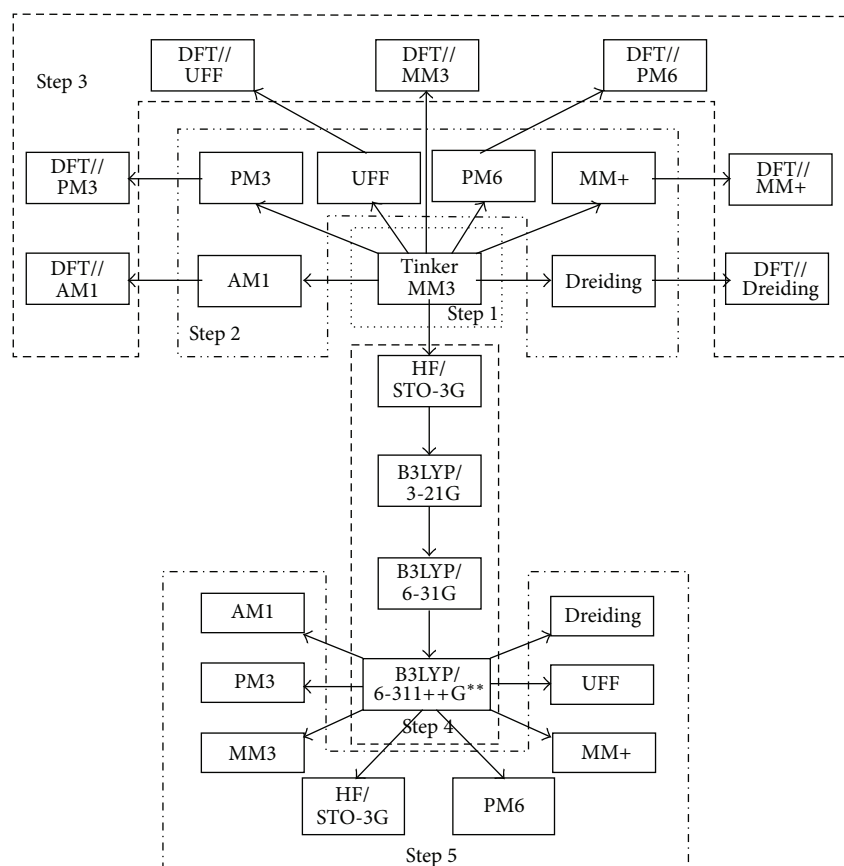


FIGURE 2: Flow chart of the calculations (Step 1: generation of conformers followed by the initial optimization using MM3 method; Step 2: optimization of the conformers by using LLMs; Step 3: single-point calculation of the conformer energies by using the reference method DFT B3LYP/6-311++G** (RefM1); Step 4: the sequential optimization cascade of conformational analysis (RefM2); Step 5: optimization of the refined conformers for their identification by using the LLMs) performed.

with all the LLMs under study (Step 5, Figure 2). It should be emphasized that Step 5 is necessary for identification of the conformers optimized in Steps 1 and 2 among geometries obtained finally in Step 4.

(6) In order to explore the accuracy of the methodological treatment presented on scheme (Figure 2) with respect to other high level reference methods, an additional method, such as B97D (Grimme's functional including dispersion) [26], is added using the same basis set (6-311++G**) as in RefM2. B97D method uses density functional constructed with a long-range dispersion correction. We label the latter reference method as RefM3 in our text.

For low level calculation using semiempirical methods [27] AM1 [28, 29], PM3 [30], and PM6 [31], molecular mechanical methods UFF [32] (UFF uses charges calculated by the method of charge equilibration [33]), and Dreiding [34], the Gaussian 09 program [25] was used while for the MM+ method [35] the Hyperchem program [36] and for the MM3 the Tinker Program [22] were used instead.

3. Results

Steps 1 and 2. The test set of conformers of MPA 2-butyl esters (including both the RR and RS diastereomers) were

generated and initially optimized with MM3 method by using the Tinker program (Step 1, Figure 2). The MM3 optimized conformers (geometries) were firstly further subjected to optimization with other LLMs under test (Step 2, Figure 2).

Step 3 and Correlation of the Conformer Energies. LLMs are assessed by correlating their conformer energies (from Steps 1 and 2, Figure 2) with the energy values resulting from single point DFT B3LYP/6-31* calculations (Step 3, Figure 2) on the same LLM optimized geometries (Table 1, Figures 3, 4, 5, 6, 7, and 8). In this case the MM3 method yields the best correlation and also correctly predicts the lowest energy conformer.

Step 4. The results of the optimization cascade (Step 4, Figure 2) and the energies of the low energy conformers in the energy range of 0 to 1.00 kcal/mol are summarized in Table 2 along with conformer energies calculated for the same conformers by optimization with LLMs (Steps 1 and 2, Figure 2) and the HF/STO-3G method (Step 4, the first level, Figure 2).

The energies of the low energy conformers from all of the test set of conformers (of MPA 2-butyl esters including both the RR and RS diastereomers) were found by using

TABLE 1: Correlation of conformer energies resulted from minimization with LLMs (Steps 1 and 2, Figure 2) with single point energies calculated for the geometries optimized with LLMs by using high level RefM1 (Step 3, Figure 2).

Parameter ^a	UFF ^b	Dreiding ^b	AM1 ^b	PM3 ^b	PM6	MM3 ^b	MM+ ^b
R^2	0.76	0.14	0.60	0.04	0.436	0.89	0.65
RMS (kcal/mol)	4.17	5.38	3.80	7.31	3.54	2.78	3.32
Slope	1.17	0.24	0.46	0.04	0.49	1.12	0.56
Intercept	0.38	2.31	0.49	1.64	1.91	-0.15	1.42

^a R^2 -correlation coefficient.

RMS-root mean square deviation of the energy values.

Slope and intercept characterized systematic deviation of the correlated energy values.

Ideally they should be 1.00 and 0.00, respectively.

^bCorrelation diagrams are presented in Figures 3–8.

TABLE 2: Determination of the cut-off conformer (COFC) energy^a (given in bold typeface) within a series of optimized with LLMs and HF/STO-3G conformers (Steps 1 and 2 and Step 4 (the first level), resp., Figure 2).

Conformer	RefM2	RefM3	HF	UFF	Dreiding	AM1	PM3	PM6	MM3	MM+
a	0.00	0.40	0.09	0.80	0.99	1.43	0.35	3.37	0.03	0.42
b	0.09	0.00	0.09	0.60	0.64	1.31	1.56	3.24	0.00	0.32
c	0.56	0.41	0.27	1.13	0.95	1.40	0.60	4.29	3.05	0.44
d	0.57	0.57	0.89	2.36	1.87	1.43	1.48	3.27	0.90	1.09
e	0.59	0.73	0.86	2.13	2.03	1.39	1.48	3.16	0.89	0.91
f	0.60	0.52	0.24	1.31	1.44	1.26	1.50	4.18	3.09	0.78
g	0.67	0.91	0.91	2.20	2.16	1.45	1.45	3.16	0.99	0.68
h	0.69	1.00	0.88	2.32	2.23	1.35	1.44	3.27	0.96	0.58
i	0.72	0.74	0.01	0.29	1.06	2.11^a	1.56	4.81	2.99	0.20
j	0.82	0.87	0.00	0.00	0.00	2.00	1.22	4.13	0.29	0.88
k	0.85	0.57	0.03	0.89	0.65	2.00	1.25	4.12	0.32	1.21
RMS ^b	0.00	0.20	0.46	1.09	1.01	0.94	0.79	3.44	1.30	0.31

^aThe COFC energy is the highest conformer energy in the series of conformers optimized with an LLM and selected as corresponding to more populated conformers in the energy range of 0–1.00 kcal/mol found by optimization with the cascade optimization method RefM2. The COFC energy determines the conformer energy range that potentially involves all the set of significant LLM conformers to be subjected to optimization at higher levels of theory.

^bThe RMS values indicate deviations of LLM energies from the energies calculated by using the RefM2.

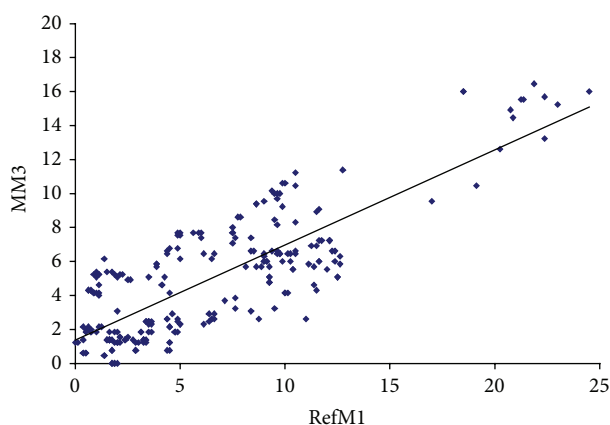


FIGURE 3: Correlation between energies (kcal/mol) of MM3 optimized conformers (Step 1, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

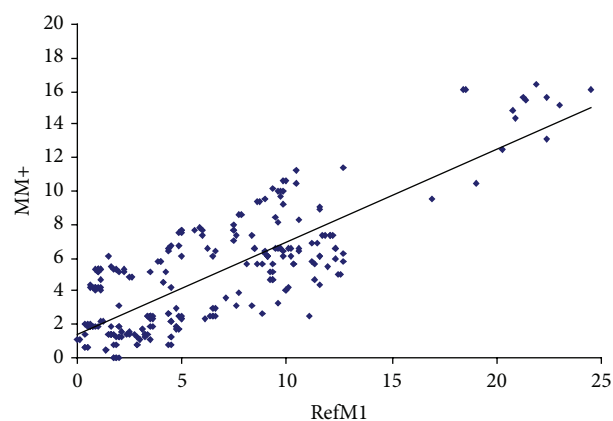


FIGURE 4: Correlation between energies (kcal/mol) of MM+ optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

the cascade of calculations (RefM2; Step 4, Figure 2) and are shown in Figure 9 in comparison with the energies resulting from optimization of just the same conformers with the LLMs

under test (Steps 1 and 2, Figure 2) and also with HF/STO-3G (Step 4, the first level, Figure 2). The latter method is included in the RefM2 as the first level of calculations, thus

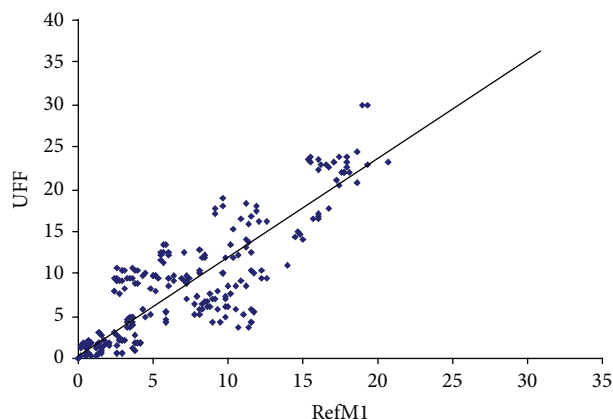


FIGURE 5: Correlation between energies (kcal/mol) of UFF optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

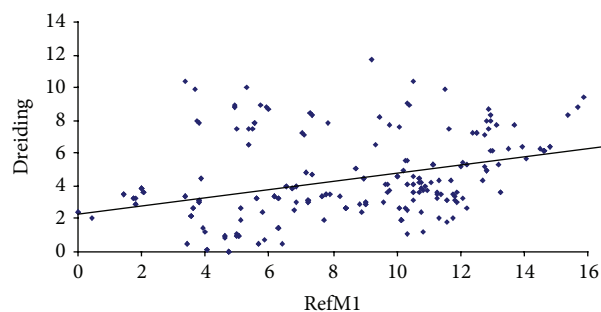


FIGURE 6: Correlation between energies (kcal/mol) of the Dreiding optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

the separated results for this method allow estimating the impact of the higher levels to the final result.

As noted, several initial conformers with different energies are converging to one certain conformer upon the sequential high level optimization procedure. The total number of conformers resulting from generation and initial optimization by using the Tinker program (Step 1, Figure 2) was 262. The further geometry optimization on subsequent higher levels reduces this number; for example, performing an HF/STO-3G step after the Tinker calculations (Figure 2) reduces it to 187. This tendency is observed for all higher level methods used.

The numbers of conformers obtained by using different LLMs (Steps 1 and 2, Figure 2) and HF/STO-3G (Step 4, first level) belonging to certain energy ranges in comparison with RefM2 are presented in Table 3.

In the case of semiempirical methods (AM1, PM3, and PM6) the number of conformers in the lower energy region is larger than that found using the other methods. For instance, by optimization with MM3, 11 conformers within the energy range from 0 to 1.00 kcal/mol have been identified whereas AM1 and PM3 yielded 32 and 41 conformers, respectively.

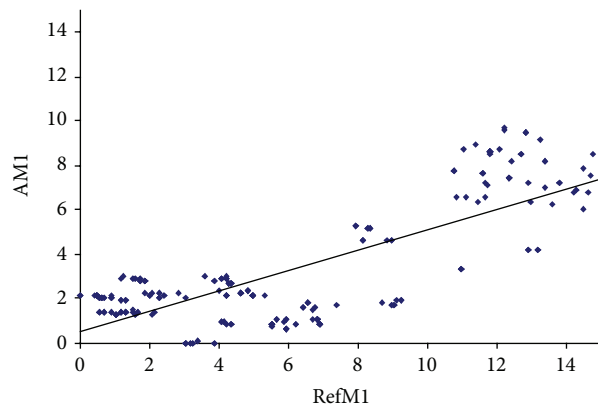


FIGURE 7: Correlation between energies (kcal/mol) of AM1 optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

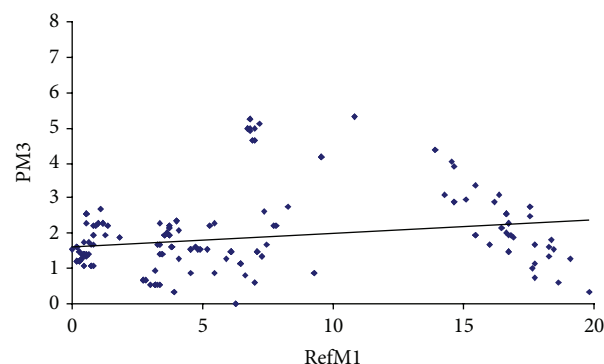


FIGURE 8: Correlation between energies (kcal/mol) of PM3 optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

The semiempirical methods tend to overestimate the number of conformers.

Step 5. After performing the calculations of Step 5 (Figure 2), it is possible to directly compare the conformational energies of identified conformers (Figure 9), that is, of conformers optimized by the LLMs and HF/STO-3G with these minimized by RefM2 (Table 2). The energy level of the lowest energy conformer is set to 0 kcal/mol, as usual. It has been noticed that quite a good similarity between energies of the entire RefM2 and HF/STO-3G occurs (Figure 10).

4. Discussion

It is an unavoidable problem in the conformational analysis in general that the number of conformers initially generated is too high and its quality too low for justified optimization of all of them on time-consuming higher calculation levels. A significant part of conformers initially generated turn out to be high energy ones or saddle points on a PES. This shortcoming has to be overcome by using an initial optimization procedure with a less accurate but also less resource consuming LLM.

TABLE 3: The number of conformers found by optimization with LLMs and HF/STO-3G (Steps 1 and 2, Step 4 (the first level), resp., Figure 2) as well as with the RefM2 (entire Step 4, Figure 2) within certain energy ranges.

Energy range from 0.00 to X [kcal/mol]	RefM2	HF	UFF	Dreiding	AM1	PM3	PM6	MM3	MM+
COFC energy ^a	11	11	35	97	87	94	76	55	37
2,50	46	54	35	111	109	159	28	39	71
2,00	43	42	23	81	78	132	27	29	64
1,50	33	42	10	49	60	89	14	26	51
1,00	11	11	6	17	32	41	10	11	26
0,75	9	4	4	13	18	26	10	4	16

^aThe COFC energy values for LLMs are given in Table 2.

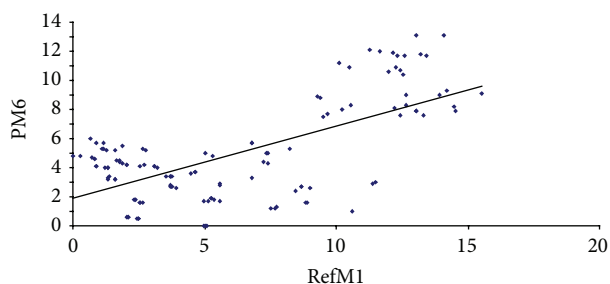


FIGURE 9: Correlation between energies (kcal/mol) of PM6 optimized conformers (Step 2, Figure 2) and energies calculated with RefM1 (Step 3, Figure 2) for the same geometries.

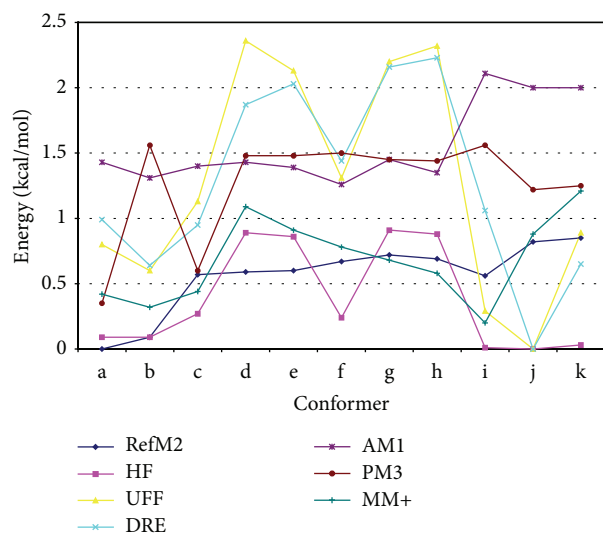


FIGURE 10: The LLM and HF conformer energies (Steps 1, 2, and 4 (the first level), resp., Figure 2) compared to the energies refined by the RefM2 (entire Step 4, Figure 2).

The rational choice of a LLM for this purpose is an important question for refining the generated conformers in order to select a reliable set for further optimization using an *ab initio* or density functional method. The failure at the initial stage of systematic conformational analysis would lead to erroneous results that cannot usually be corrected in subsequent steps by using better methods. A common way to minimize the number of conformers for further optimization

is neglecting the high energy conformers. However, this kind of simplification of the process may be unjustified if the choice of an LLM for the initial optimization as well as (or) the choice of the corresponding cut-off conformer is not right. Still, it has been difficult to guarantee that conformers which later upon optimization with RefM2 will turn out to be low energy conformers are not discarded.

In order to investigate the approach, none of the higher energy conformers has been discarded at any of the calculation steps (Figure 2) in the current work.

In sum, there are two requirements to the use of a LLM at the initial stage to guarantee obtaining high computational efficiency along with reliable results. Both of them depend on the quality of the LLM as well as on a rational protocol for performing the selection.

The LLMs that correlate with the RefM1 regarding the conformer energies are suitable for the initial stage of a conformational analysis. The suitable LLM should also be characterized by low (preferably the lowest) cut-off conformer energy.

In addition, the situation where the LLM results in a low energy conformer but a high level method gives a high energy value for the same conformer is not disastrous (i.e., it does not bring upon erroneous results), but rather makes the LLM less selective and therefore the number of the conformers to be optimized still remains too high, which is just impractical. A preferable situation would be if a low and a high level method would correlate to a satisfactory extent.

The quality of the conformer energies minimized with different LLMs (Steps 1 and 2, Figure 2) is evaluated on the basis of single point DFT B3LYP/6-311++G** calculation (RefM1; Step 3, Figure 2) of energies corresponding to the molecular geometries optimized with LLMs under test (Steps 1 and 2, Figure 2) (correlation diagrams are presented in Figures 3–8). Table 1 presents correlation parameters between these two types of energy values.

The conformer energies are best reproduced with MM3 force field, followed by UFF (R^2 values of 0.89 and 0.76, resp.) though both of them tend to slightly overestimate conformational energies, given by the slope value above 1.0 of the least square fit line. The MM+ force field with an RMS value of 3.3 kcal/mol tends to underestimate the conformational energies. In the series of LLMs under test, the Dreiding force field proved to be rather problematic

with its very poor correlation and severe underestimation of conformational energies.

AM1 systematically underestimates the energies that are more pronounced in the middle-range DFT energy values. Although the AM1 method gives a fairly good regression fit ($R^2 = 0.6$), PM3 shows no correlation with DFT relative energy values. Besides, the slope value close to zero indicates an underestimation of conformer energies. However, regarding energy differences, herein the PM3 method has correctly identified the energetically most favorable conformer, that is, a very valuable result. In particular, the use of the semiempirical methods for the conformational analysis of complex compounds may give valuable information. For instance, AM1 has been used for the conformational analysis of prostaglandins providing results (based on identification of the prevailing conformers) that qualitatively correlated with experimental characteristics [28].

Based on the above reasoning, the MM force fields can thus be suggested as quick preliminary methods for energy minimization, yielding satisfactory semiquantitative results regarding the energies of conformers. In addition, the examination of the results of conformational analysis with entire RefM2 indicates that molecular mechanics MM3 force field is able (as PM3 did) to predict correctly the lowest energy conformers (Figure 3; conformers a and b) obtained by this higher level procedure.

The semiempirical methods usually give less accurate results. This is because they are parameterized near the local minimum of the PES. At more prominent distances from these minima, the semiempirical methods are not able to reproduce correct energy values. Thus, optimizing the structure from this region of a PES might result in an incorrect optimized structure.

The COFC energies were found as follows. Table 2 lists the low energy (<1.00 kcal/mol) conformers obtained with a high level method RefM2. These conformer energies are compared with energies found for the same conformers by optimization with various LLMs. For instance, the COFC found with MM3 has energy of 3.09 kcal/mol, which corresponds to a conformer labeled with min this paper (Table 2). In the same way the UFF method yields the COFC, namely, n, with the energy of 2.36 kcal/mol. Repeating this procedure for the rest of the methods yielded all the COFCs for the methods tested herein and the results are presented in Table 2 (the COFC energies are in bold typeface).

The number of significant LLM conformers appeared to be different in case of different LLMs. The question related to computational efficiency in an attempt to find all significant low energy conformers (0.00–1.00 kcal/mol) is the following: which initial LLM gives a minimum number of conformers necessary to be optimized with a high level method? The numbers of conformers corresponding to the energy ranges are given in Table 3. All the conformers in a series within energy range up to COFC's energy value (which are included in Table 3) have to be considered. On the basis of the results provided in Table 3 it can be concluded that the most efficient LLM is the UFF which permits to limit the amount of conformers to 35. The MM+ and MM3 methods select 37 and 55

conformers, respectively. Semiempirical methods AM1, PM3, and PM6 yield 87, 94, and 76 conformers, respectively, while Dreiding force field selects 97 conformers to be subjected to further calculation with a high level method. Replacing the high level reference method from RefM2 to RefM3 and performing a similar cascade like optimization scheme as performed in case of RefM2 (Figure 2, Step 4), no significant changes occur as can be seen from Table 4. The UFF selects 39 conformers; the other low level methods collect the numbers which are approximately at the same line as compared to RefM2.

Special attention should be paid to the (low level *ab initio*) Hartree-Fock HF/STO-3G method. It could be positioned in between the low and high level computational methods. In this work HF/STO-3G is included into RefM2 as the first step of the cascade of higher level calculations (Step 4; Figure 2) and therefore comparison of the results of RefM2 with these of HF/STO-3G can only be illustrative. The accordance between the HF/STO-3G method and the entire RefM2 is pretty good; the energy of the corresponding COFC is the lowest found for LLMs, only 1.07 kcal/mol. As mentioned, HF/STO-3G is assumed to belong to LLMs in some respect and corresponding characteristics have been included in the Tables 1 and 2. This method yields only 11 conformers in the energy range up to the COFC energy.

In addition to evaluation of the quality of an LLM for the calculation of conformer energy, it is important to create and follow criteria for the use of an LLM. In this regard it is reasonable to pick up the RefM2 optimized conformers appearing within an energy range, for example, <1.00 kcal/mol (Figure 3). Approximately, this range is justified because of the Boltzmann law, according to which the higher energy conformers are less significant due to low populations. Among these selected conformers one should identify the conformer showing the highest LLM energy (Table 2). Herein, we label this conformer as a cut-off conformer (COFC) for this LLM and its energy defines the energy range involving the LLM conformers that have to be subjected to further consideration.

In order to assess the LLMs, we have to determine the COFCs and corresponding energies for all of the methods under test. The best LLM in this respect is characterized by the lowest COFC energy value. Thus, the LLMs can be ranked according to the COFC energies. However, this ranking should be very critically evaluated taking into account also the conformer energy correlation results (Table 1) and the number of conformers to be submitted to optimization on the higher levels of theory.

In sum, it is assumed that the following guidelines should be taken into account in assessing the suitability of the LLM.

- (1) Conformer energies found by using the high and low level method should correlate.
- (2) A COFC of an LLM selected for use should be characterized by the low energy in the manifold of LLMs, indicating probably a more preferable method. The COFC energy should be taken as a cut-off value when discarding high energy conformers after initial optimization of conformers generated.

TABLE 4: The same as Table 3, but the high level reference method is replaced from B3LYP/6-311++G** (RefM2) to B97D/6-311++G** (RefM3).

Energy range from 0.00 to X [kcal/mol]	RefM3	HF	UFF	Dreiding	AM1	PM3	PM6	MM3	MM+
COFC energy	15	36	39	84	73	112	84	44	55
2,50	61	52	42	68	76	110	27	39	76
2,00	47	43	32	47	49	93	26	29	61
1,50	38	43	13	21	39	59	13	26	37
1,00	11	33	10	6	22	23	9	11	16
0,75	9	8	7	3	13	17	9	4	13

- (3) It would be preferable if the lowest energy conformer calculated with a high level method would also appear as the lowest energy conformer obtained by using the chosen LLM.
- (4) The selectivity of the LLM should be sufficient, that is, the number of conformers remaining for the optimization on higher levels should be acceptable (as compared with other LLMs).

5. Summary

A comparative demonstration of the LLMs' performance based on optimization of conformer geometries of the 2-butyl esters of MPA was carried out. Special focus was on the rational selection of the LLM for the initial optimization of generated conformers in total conformational analysis of esters. The methods assessed were MM3, MM+, UFF, Dreiding, AM1, PM3, and PM6. The performance of the low level *ab initio* calculation method HF/STO-3G was studied in the role of a lower step of the optimization cascade. The single point calculation of the conformer benchmark energies for the LLM optimized geometries with DFT B3LYP/6-311++G** was performed. The LLM conformer energies were correlated with the higher level benchmark single point energies; corresponding diagrams along with the correlation parameters are presented. In this approach, the MM3 force field appeared to yield the best correlation values.

It is also shown that the above correlation parameters, obligatory for the assessment of the quality of a LLM regarding the energy calculation, are neither sufficient to assess fully the quality of a LLM nor do they provide exact criteria for applying of a more proper LLM to the initial optimization of geometry of conformers. For the adequate assessment of an LLM, in addition to the energy correlation quality, the number of conformers that need to be fed to the further optimization in order to obtain reliable results should be determined and critically assessed in comparison with the other methods. This is the number of conformers up to the cut-off conformer (COFC) energy value; the lower these numbers are for a certain LLM, the better its quality is. And in reverse, high number of such conformers is indicative of poor selectivity of the LLM. Various methods have yielded different thresholds for the energy ranges (the COFC energies) and the number of conformers to be considered further.

Taking into account the assessment criteria, the MM3 force field has been chosen for the initial optimization of generated conformers in our investigations of esters [8].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank the Estonian Ministry of Education and Science (Grants no. SF0140133s08, SF0690034s09, ETF8880, and IUT19-32) and EU European Regional Development Fund (3.2.0101.08-0017) for the financial support. The authors are also grateful to Tallinn University of Technology for the financial support through Grant no. B35.

References

- [1] R. Aav, T. Pehk, S. Tamp et al., "Theoretical prediction and assignment of vicinalH-H coupling constants of diastereomeric 3-alkoxy-6,7-epoxy-2-oxabicyclo[3.3.0]octanes," *Magnetic Resonance in Chemistry*, vol. 49, no. 2, pp. 76–82, 2011.
- [2] K. W. Sattelmeyer, J. Tirado-Rives, and W. L. Jorgensen, "Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules," *Journal of Physical Chemistry A*, vol. 110, no. 50, pp. 13551–13559, 2006.
- [3] K. Gkionis, J. G. Hill, S. P. Oldfield, and J. A. Platts, "Performance of Becke's half-and-half functional for non-covalent interactions: energetics, geometries and electron densities," *Journal of Molecular Modeling*, vol. 15, no. 9, pp. 1051–1060, 2009.
- [4] R. S. Paton and J. M. Goodman, "Hydrogen bonding and π -stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions," *Journal of Chemical Information and Modeling*, vol. 49, no. 4, pp. 944–955, 2009.
- [5] G. L. Warren, C. W. Andrews, A.-M. Capelli et al., "A critical assessment of docking programs and scoring functions," *Journal of Medicinal Chemistry*, vol. 49, no. 20, pp. 5912–5931, 2006.
- [6] R. Wang, Y. Lu, and S. Wang, "Comparative evaluation of 11 scoring functions for molecular docking," *Journal of Medicinal Chemistry*, vol. 46, no. 12, pp. 2287–2303, 2003.
- [7] T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang, "Comparative assessment of scoring functions on a diverse test set," *Journal of Chemical Information and Modeling*, vol. 49, no. 4, pp. 1079–1093, 2009.
- [8] S. Tamp, K. Danilas, M. Kreen et al., "A total conformational analysis of diastereomeric esters and calculation of their conformational shielding models," *Journal of Molecular Structure: THEOCHEM*, vol. 851, no. 1–3, pp. 84–91, 2008.

- [9] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, UK, 1989.
- [10] Ö. Alver, M. F. Kaya, M. Bilge, and C. Parlak, "Vibrational spectroscopic investigation and conformational analysis of methacrylamidoantipyrine: a comparative density functional study," *Journal of Theoretical Chemistry*, vol. 2013, Article ID 386247, 10 pages, 2013.
- [11] Z. Weiqun, L. Baolong, C. Yang, Z. Yong, and L. L. Y. Xujie, "The structure and conformation analysis of N-2-fluorobenzoyl-N'-2-methoxyphenylthiourea," *Journal of Molecular Structure: THEOCHEM*, vol. 715, no. 1-3, pp. 117-124, 2005.
- [12] M. W. Hill, K. B. Wiberg, H. Castrejon, W. F. Bailey, J. Ochterski, and J. M. Seminario, *Recent Developments and Applications of Modern Density Functional Theory*, Elsevier, London, UK, 1996.
- [13] O. Tšubrik, P. Burk, T. Pehk, and U. Mäeorg, "Conformational analysis of 1-acetyl-2-methylhydrazine," *Journal of Molecular Structure: THEOCHEM*, vol. 546, pp. 119-125, 2001.
- [14] A. Metsala, E. Usin, I. Vallikivi, L. Villo, T. Pehk, and O. Parve, "Quantum chemical evaluation of the yield of hydroxybenzophenones in the fries rearrangement of hydroxyphenyl benzoates," *Journal of Molecular Structure: THEOCHEM*, vol. 712, no. 1-3, pp. 215-221, 2004.
- [15] T. D. Rasmussen and F. Jensen, "Force field modelling of conformational energies," *Molecular Simulation*, vol. 30, no. 11-12, pp. 801-806, 2004.
- [16] M. Stahl, U. Schopfer, G. Frenking, and R. W. Hoffmann, "Conformational analysis with carbon-carbon coupling constants. A density functional and molecular mechanics study," *Journal of Organic Chemistry*, vol. 62, no. 11, pp. 3702-3704, 1997.
- [17] E. Voloshina, J. Fleischhauer, and P. Kraft, "Conformational analysis and CD calculations of methyl-substituted 13-tridecano-13-lactones," *Helvetica Chimica Acta*, vol. 88, no. 2, pp. 194-209, 2005.
- [18] A. Leach, *Molecular Modeling: Principles and Applications*, Prentice Hall, New York, NY, USA, 2nd edition, 2001.
- [19] B. M. Trost, J. L. Belletire, S. Godleski et al., "On the use of the O-methylmandelate ester for establishment of absolute configuration of secondary alcohols," *Journal of Organic Chemistry*, vol. 51, no. 12, pp. 2370-2374, 1986.
- [20] T. Pehk, E. Lippmaa, M. Lopp, A. Paju, B. C. Borer, and R. J. K. Taylor, "Determination of the absolute configuration of chiral secondary alcohols; new advances using carbon-13 and 2D-NMR spectroscopy," *Tetrahedron: Asymmetry*, vol. 4, no. 7, pp. 1527-1532, 1993.
- [21] J. M. Seco, E. Quiñoá, and R. Riguera, "The assignment of absolute configuration by NMR," *Chemical Reviews*, vol. 104, no. 1, pp. 17-117, 2004.
- [22] J. W. Ponder, Program TINKER, (version 5.0), 2007, <http://dasher.wustl.edu/tinker/>.
- [23] R. V. Pappu, R. K. Hart, and J. W. Ponder, "Analysis and application of potential energy smoothing and search methods for global optimization," *Journal of Physical Chemistry B*, vol. 102, no. 48, pp. 9725-9742, 1998.
- [24] D. C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*, John Wiley & Sons, New York, NY, USA, 2001.
- [25] M. J. Frisch, G. W. Trucks, H. B. Schlegel et al., Gaussian 03 (Revision E. 01), Gaussian Inc., Wallingford, Conn, USA, 2009.
- [26] S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *Journal of Computational Chemistry*, vol. 27, no. 15, pp. 1787-1799, 2006.
- [27] J. J. P. Stewart, "Optimization of parameters for semiempirical methods. II. Applications," *Journal of Computational Chemistry*, vol. 109, pp. 221-264, 1998.
- [28] M. J. S. Dewar, E. G. Zoebisch, and E. F. Healy, "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model," *Journal of the American Chemical Society*, vol. 107, no. 13, pp. 3902-3909, 1985.
- [29] L. P. Davis, R. M. Guidry, J. R. Williams, M. J. S. Dewar, and H. S. Rzepa, "MNDO calculations for compounds containing aluminum and boron," *Journal of Computational Chemistry*, vol. 2, pp. 433-445, 1981.
- [30] J. J. P. Stewart, "Optimization of parameters for semiempirical methods. I. Method," *Journal of Computational Chemistry*, vol. 10, pp. 209-220, 1989.
- [31] J. J. P. Stewart, "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements," *Journal of Molecular Modeling*, vol. 13, no. 12, pp. 1173-1213, 2007.
- [32] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American Chemical Society*, vol. 114, no. 25, pp. 10024-10035, 1992.
- [33] A. K. Rappé and W. A. Goddard III, "Charge equilibration for molecular dynamics simulations," *Journal of Physical Chemistry*, vol. 95, no. 8, pp. 3358-3363, 1991.
- [34] S. L. Mayo, B. D. Olafson, and W. A. Goddard III, "DREIDING: a generic force field for molecular simulations," *Journal of Physical Chemistry*, vol. 94, no. 26, pp. 8897-8909, 1990.
- [35] Hyperchem manual.
- [36] HYPERCHEM ver. 7, Hypercube Inc., <http://www.hyper.com>.

