JCTC Journal of Chemical Theory and Computation

Pseudospectral Local Second-Order Møller−Plesset Methods for Computation of Hydrogen Bonding Energies of Molecular Pairs

George A. Kaminski*

Department of Chemistry, Central Michigan University, Mt. Pleasant, Michigan 48859

Jon R. Maple, Robert B. Murphy, and Dale A. Braden

Schrödinger, Inc., 120 West 45th Street, Tower 45, 32nd Floor,
New York, New York 10036

Richard A. Friesner

Department of Chemistry, Columbia University, New York, New York 10027

Abstract: We present a methodology for computing the binding energy of molecular dimers based on extrapolation of pseudospectral local second-order Moller−Plesset (MP2), or PS-LMP2, energies to the basis set limit. The extrapolation protocol is based on carrying out PS-LMP2 calculations with the Dunning cc-pVTZ (-f) and cc-pVQZ (-g) basis sets and then using a simple two-parameter function to compute the final basis set limit results. The function is parametrized to ultralarge basis set MP2 calculations for 5 molecular pairs taken from the literature and then tested by calculating results for a set of formamide dimers for which such calculations have also been carried out. The results agree to within ca. 0.2 kcal/mol with the conventional MP2 large basis set calculations. A specialized, but relatively simple, protocol is described for eliminating noise due to overcompleteness of the basis set. Timing results are presented for the LMP2 calculations, and comparisons are made with the LMP2 methodology of the QChem program. CPU time required by each of the methods scales as N^3, where N is the number of the basis functions, with the PS-LMP2 approach displaying a 2- to 3-fold advantage in the prefactor. We also discuss one set of test cases for which the PS-LMP2 results disagree with those obtained from an alternative type of MP2 calculation, *N*-methyl acetamide (NMA) dimers, and show that the results for liquid-state simulations using polarizable parameters derived by fitting to the PS-LMP2 binding energies appear to produce better results when compared with experimental data. The convergence issues associated with the alternative MP2 formulation remain to be investigated.

I. Introduction

The computation of hydrogen bonding energies between pairs of molecular fragments is an important problem in theoretical chemistry. Hydrogen bonding strengths are crucial for a wide range of chemical processes in both the gas phase and in

condensed phase. Our own interests lie particularly in the areas of molecular mechanics force field development and structure based drug design; for these endeavors, errors on the order of 0.5 kcal/mol can be significant. At the same time, pharmaceutical compounds span a very broad range of chemical functionality, the great majority of which is not accessible to experimental characterization of the inter-

* Corresponding author e-mail: kamin1ga@cmich.edu.

molecular interactions in a straightforward fashion. To address such problems, a computational approach is required that can be applied to hundreds or thousands of molecular pairs, many of significant size, while at the same time routinely providing accuracy to better than 0.5 kcal/mol. A description of such a method is the objective of the present paper.

Over the past 5 years, it has become apparent that second-order Moller−Plesset perturbation theory (MP2) provides answers of the quality specified above (at least for hydrogen bonded pairs of organic compounds, on which we shall focus our attention in the present paper) provided that the basis set limit of the calculation can be obtained.[1] The problem is that very large basis sets are required to achieve convergence, and this in turn necessitates the expenditure of large amounts of computation time if standard algorithms for MP2 calculations are employed. For treating pair hydrogen bond interactions between molecular fragments of the size relevant to force field development, a key determinant of efficiency is scaling with basis set size; in particular, the use of distance cutoffs is less relevant than, for example, if one is interested in methods for treating a linear chain of carbon atoms. At the same time, a high degree of precision in relative energies is required, which may pose problems for approximate methods, for example those relying on basis set expansions of product charge distributions.

The localized pseudospectral methods (PS-LMP2)[2] that we have developed and implemented in the Jaguar suite of ab initio electronic structure programs[3] are well suited to computations of this type. These calculations can be shown to scale as $N_b{}^2$, where N_b is the number of basis functions per atom and no cutoffs are assumed. The overall scaling of the method with system size for relatively small molecules is on the order of $N_{atom}{}^3$. These scaling behaviors imply that tractable CPU times can be obtained for the accurate hydrogen bonding calculations that are the focus of the present paper.

With a highly efficient method in hand for carrying out large basis set LMP2 calculations, one can employ extrapolation techniques to reach the basis set limit, many of which are now reported in the literature.[5] We use here a simple approach based on the Dunning correlation consistent basis sets[6] and a two-parameter fit to an exponential functional form. The extrapolation method is "trained" on a series of dimers for which converged, large basis set results have been reported by Tsuzuki and co-workers[1a] and then tested via calculation of the binding energies of a number of formamide dimers. The agreement for the test set, on the order of a few tenths of a kcal/mol in absolute binding energy for all five complexes, suggests that a reliable approach to the computation of hydrogen bonding energetics has been produced.

Over the past several years, we have extensively utilized these extrapolated LMP2 calculations to develop a polarizable molecular mechanics force field that is capable of predicting condensed phase properties accurately without the use of adjustable parameters. The development of parameters for a number of small molecules, and subsequent prediction of liquid state heats of vaporization and densities, is discussed in ref 8. Heats of vaporization are generally predicted with

an accuracy of ∼0.5 kcal/mol; this quantity is very sensitive to the hydrogen bonding energy (which controls the fitting of the short range van der Waals parameters of the polar atoms in the molecule), and hence the results provide a further confirmation of the accuracy of our LMP2 extrapolation protocol (as well as a validation of the proposition that extrapolated MP2 provides accurate hydrogen bonding energies). A case of particular interest is the NMA molecule, where our results differ substantially from those in ref 1b, computed by conventional quantum chemical methods, despite the fact that for the five formamide dimers mentioned above, the results of the two methods agree to within a few tenths of a kcal/mol. We discuss our construction of polarizable liquid-state models using both sets of hydrogen bonding energies and show that our LMP2 results yield excellent agreement with experimental data, whereas reliance on the ref 1b results produces significant overbinding in the condensed phase simulations.

The remainder of the paper is organized as follows. In Section II, we describe our computational methods, briefly reviewing the relevant aspects of the PS-LMP2 approach (which has been presented in detail in previous work) and discussing the extrapolation protocol. There are technical issues associated with overcompleteness of the basis set, which we have addressed and successfully surmounted, and these are presented in some detail. Section III presents results for the various systems that we have investigated, demonstrating that high accuracy can be achieved with PS-LMP2 methods. Timing results are also presented and compared with those obtained from the LMP2 methodology in the Q-Chem program.[7] Finally, in the conclusion, we summarize our results and briefly discuss directions of future research.

II. Computational Methods

A. Pseudospectral LMP2 Technique. We have described PS methods for carrying out local second-order Moller Plesset (LMP2) perturbation theory calculations in previous papers.[2,4] Briefly, the use of a numerical grid at a key stage in the evaluation of two electron integrals allows the usual four index transform to be eliminated and replaced by an algorithmic step which, when localized orbitals are used, has superior scaling properties with system and basis set size. This key step involves the assembly of the exchange integrals via a sum over grid points (g)

$$K_{ij}^{pq} = \sum_g Q_i(g)R_p(g)A_{jq}(g) \qquad (1)$$

over local occupied orbitals (*ij*) and local virtual orbitals. $Q_i(g)$ is the least-squares fitting operator for orbital *i*, $R_p(g)$ is the physical space representation of virtual orbital *p*, and $A_{jq}(g)$ is the three center, one electron orbital over molecular orbitals *j,q*.

The scaling of this calculation is $n_{occ}{}^2 * N_b{}^2 * N_g$, where n_{occ} is the number of occupied orbitals in the molecule, N_b is the number of orbitals in the local space, and N_g is the number of grid points on the mesh.[2] Since N_b is in principle invariant to the size of the molecule, the scaling of this step with system size (roughly represented by the number of atoms N_a) is proportional to $N_a{}^3$. Of equal importance for the present

work is the scaling with the size of the atomic basis set. For ultralarge basis sets, it is not necessary to continue to increase the size of the grid beyond a certain point. Then, the computational effort scales as N_b^2 with the number of atomic basis functions in the localized space (in essence the size of the atomic basis sets). In contrast, the initial step of the traditional four index transform scales as N_b^4. Thus, we would expect the advantage of the PS-LMP2 methods, as compared with the standard canonical MP2 calculations, to increase as the basis set size is increased. There are other methods that have been developed to reduce the computational expense of MP2 methods (e.g., refs 9 and 10); the question of which approach yields the best performance for a given problem, in terms of speed and accuracy, requires direct comparison of performance for the same systems. We present below a comparison of this type with the LMP2 methodology in the Q-Chem program, where quantitative (although not qualitative) advantage in efficiency is demonstrated. However, the results presented below can be used by others if desired to make such comparisons themselves.

B. Extrapolation Methods. Initial studies suggested that localized perturbation methods might converge to the basis set limit more quickly than canonical MP2 methods (even when augmented by counterpoise corrections) due to the reduction of basis set superposition error arising from truncation of the virtual space for each electron pair.[11] However, over the past several years, it has been demonstrated via ultralarge basis set computations that extrapolation to the basis set limit is necessary, even if localized methods are employed, if accurate intermolecular pair energetic properties are to be computed.[1,12] On the other hand, these same studies also show that, for typical hydrogen bonded dimer structures, reaching the MP2 basis set limit provides results within a few tenths of a kcal/mol in accuracy; corrections from higher order correlation effects (e.g. at the CCSD (T) level) are very small due to cancellation of several different terms (note that this conclusion does not necessarily hold for other strong nonbonded interactions, such as aromatic ring π stacking interactions). Hence, our goal here is to develop an extrapolation approach based on PS-LMP2 methods, which agrees with the benchmark studies quoted above, while at the same time provides significant reductions in computational effort, as compared to alternative protocols for achieving this level of accuracy.

After considerable experimentation, we have settled upon the following simple methodology. PS-LMP2 calculations are carried out for the pair and for the separated monomers using the cc-pVTZ (-f) and cc-pVQZ (-g) basis sets. Counterpoise BSSE corrections are determined for the Hartree−Fock part of the energy only. This approach is motivated by the idea that LMP2 based approaches, of the type that we have implemented, have a relatively low level of BSSE present, due to the use of localized virtual spaces. Other researchers suggested extrapolation schemes, including the one with the two-parameter exponential functional form to be used for the basis set extrapolation parameters.[13] We obtain the final energy by using our data with the following extrapolation scheme, i.e.

$$E_{bind} = [a \cdot E_{bind}(\text{cc} - \text{pvqz}) - b \cdot E_{bind}(\text{cc} - \text{pvtz})]/(a - b) \tag{2}$$

where $E_{bind}(\text{cc} - \text{pvqz})$ and $E_{bind}(\text{cc} - \text{pvtz})$ are the cc-pVQZ (-g) and cc-pVTZ (-f) counterpoise corrected binding energies, respectively. The values of the parameters a and b are determined by fitting to a "training set" of molecular dimers, for which benchmark results have been obtained using conventional MP2 methods and ultralarge basis sets. The extrapolation protocol is then tested by taking a new set of dimers and again comparing with high level calculations, this time without adjustment of a and b.

In the results presented below, we optimize the dimer geometries at the LMP2/cc-pVTZ (-f) level of theory, unless explicitly noted otherwise. Such optimizations are relatively expensive, and it is of interest to ask whether the use of smaller basis sets in the LMP2 optimization, or even optimization using DFT methods, would yield equivalent results. We do not consider this question in the present paper; it will be explored in future publications.

We note that the use of the specific basis sets suggested above (which are modified from their normal composition by the elimination of g functions from the QZ basis and f functions from the TZ basis) constitutes in and of itself a substantial increase in efficiency as compared to the use of larger basis sets for extrapolation of hydrogen bond energies. For the particular property of interest here, the higher angular momentum basis functions are not as important as when one is for example trying to compute covalent bond energies, although we have found that it was not possible to simply eliminate f functions completely from the calculations. Further reduction in the basis set size may be possible; we have not investigated this in the present paper.

C. Problems Due to Basis Set Overcompleteness. While Gaussian basis sets have many compelling advantages for electronic structure calculations, they suffer from one principal disadvantage, nonorthogonality of the basis set, leading in the limit of a large number of basis functions to overcompleteness and the associated numerical instabilities. This problem is particularly an issue when using approximate numerical methods, such as pseudospectral methods, to compute integrals. If one is not careful, small errors in the integrals can translate into large variation in the energy of the highly oscillatory orbitals generated from linear combinations of strongly overlapping basis functions. While we have spent a great deal of effort to ensure that the PS method correctly handles problems of this type, the use in the present calculations of QZ basis sets and the desire to achieve precision in the calculations on the order of a few tenths of a kcal/mol, renders this endeavor more challenging than in our previously published works. Note that this same problem could conceivably arise in other schemes involving numerical approximations to the two electron integrals, such as resolution of the identity (RI-MP2) approaches.

The simplest and most effective way of dealing with overcompleteness is singular value decomposition. The overlap matrix is diagonalized, and eigenvectors corresponding to small eigenvalues are eliminated from the basis set. This methodology directly removes the highly oscillatory

Table 1. Comparison of Five Small Molecule ab Initio Dimerization Energies in kcal/mol

dimer	MP2[a]	CCSD(T) limit[a]	LMP2[b]	LMP2/MP2 geometry[c]
$H_2O - MeOH$	4.99	4.90	4.95	4.93
$H_2O - Me_2O$	5.70	5.51	5.68	5.48
$H_2O - H_2CO$	5.21	5.17	5.21	5.10
MeOH − MeOH	5.58	5.45	5.79	5.46
HCOOH − HCOOH	13.79	13.93	16.18	13.90

[a] Reference 1a. [b] Plus extrapolation. This work. Optimized geometry. [c] Plus extrapolation. This work. Geometries from reference 1a.

functions and hence avoids errors in the energy of the type discussed above. Standard thresholds for the smallest allowable eigenvalue have been developed and successfully used for many years.

This method is usable for the present problem, but there is one caveat; the number of functions deleted from the monomer and dimer calculations must be consistent, if an accurate energy difference is to be computed. This is a straightforward protocol to implement. First, a cutoff for the monomer eigenvalues is specified; after some experimentation on a wide range of dimers, we have settled on a value of 3.5×10^{-4}, which gives robust results for all of the cases tested to date. Then, the number of orbitals used in the monomer calculations is saved, and the number employed in the dimer calculations is forced to be equal to the sum of the numbers for the monomers. One can imagine that in some cases, a more elaborate method, in which correspondence between the orbitals deleted in the dimer and monomer calculations (in terms of spatial composition, energy, etc.) was checked, might be necessary; so far, however, the simple approach described above appears to be adequate. It is important to emphasize, however, that failure to use an appropriate cutoff, or allowing inconsistent numbers of orbitals between the dimer and monomer calculations, can lead to errors as large as $1-2$ kcal/mol in the binding energy. The present approach, on the other hand, contributes deviations on the order of $0.1-0.2$ kcal/mol, a crucial order of magnitude improvement.

III. Results

A. Training Set. Tsuzuki et al. have studied five molecular dimers using conventional MP2 with counterpoise correction methods for basis sets up to cc-pV5Z and also at the CCSD (T) level, along with large basis set MP2 level geometry optimizations.[1a] Their final results for binding energies in the MP2 basis set limit and with the inclusion of higher levels of electron correlation are summarized in Table 1. As discussed above, the MP2 results are within a few tenths of a kcal/mol of CCSD (T) results. As the CCSD (T) methodology is at present clearly too computationally intensive for our ultimate objectives (e.g. generation of hundreds or thousands of molecular pair energies so as to design a molecular mechanics force field with broad coverage of chemical space), we limit our objectives to the achievement of accuracy on the scale of the MP2 basis set limit results— i.e. better than 0.25 kcal/mol with an average error on the order of 0.1 kcal/mol.

We have carried out extrapolated LMP2 calculation using our own LMP2/cc-pVTZ (-f) optimized geometries (LMP2 column in Table 1) and at the MP2/6-311G** geometries supplied in ref 1a (we refer to the canonical MP2 calculations in ref 1a as "MP2" in what follows). These results are presented in Table 1. Agreement of the extrapolated LMP2 results with those of ref 1a at the geometries specified therein (these data are what are used to train our extrapolation method) is on the order of ~0.1 kcal/mol.

Parameter values obtained from the fitting to the training set (defined in eq 2) are as follows:

$$a = \exp(-A), b = \exp(-B), A = 1.8, B = 2.7 \quad (3)$$

The uniformly small size of the deviations in Table 1 for every training set molecule, combined with the fact that there are five data points and only two adjustable parameters, implies that the fitting protocol is extremely stable and robust. Note also that the difference between the LMP2 and MP2 protocols is generally less than or comparable to the difference between the MP2 and CCSD (T) protocols and that in fact the standard deviation of the LMP2 and CCSD (T) results is no worse than that between MP2 and CCSD (T). While there is no guarantee that this would be true for every molecular pair, the results again suggest that our protocol is a more or less equivalent replacement for large basis set extrapolated MP2 calculations.

For four of the five dimers, our LMP2 optimized geometries and energies are very close to those reported in ref 1a. However, for the fifth, a carboxylic acid dimer, there is a substantial disagreement, with a significantly larger binding energy being computed at the LMP2 optimized geometry. Given the very close agreement of the energies at the common geometry, there are two possible explanations for this discrepancy:

(1) The geometry optimizations have located different local minima (or, possibly, that one of the reported geometries is not in fact a local minimum, but on a plateau), due to differences in the geometry optimization algorithms in the two programs used for the calculations.

(2) The potential surfaces on which the geometry optimizations are carried out are different, leading to different locations for the minimum. This is in fact possible because we employ a different basis set, and we are using LMP2 as opposed to canonical MP2 methods (so that, for example, a substantial fraction of the basis set superposition error is not present even in the geometry optimization, as opposed to the final energy evaluation when counterpoise corrections are incorporated).

We cannot distinguish between these two possibilities without access to the code used by Tsuzuki and co-worker to obtain the results in ref 1a, which we do not have. We can, however, compare the energies of the two geometries using a common method (our own LMP2 extrapolation protocol). When these values are compared, the total electronic energy of the LMP2-optimized dimer is 2.28 kcal/mol lower than that of the MP2-optimized dimer. This fact, together with the very close energy agreement between our LMP2 calculations and those obtained by Tsuzuki and co-workers when the same geometries are used, suggests that we have, in fact, managed to locate a lower energy minimum than the one used in ref 1a, and we conclude that the correct

Table 2. Comparison of Formamide ab Initio Dimerization Energies in kcal/mol

dimer	CCSD(T) limit[a]	LMP2[b]
fm__1	14.35	14.23
fm__2	9.70	9.62
fm__3	7.34	7.05
fm__4	6.76	6.87
fm__5	5.02	5.23

[a] Reference 1b. [b] Plus extrapolation. This work. LMP2/cc-pVTZ(-f) optimized geometry.

Table 3. Wall Clock Timing Results (s) for LMP2 Calculations Using the Jaguar and QChem Programs[a]

dimer	basis fcns	QChem	Jaguar	ratio (QChem/Jaguar)
ala2	936	113203	41675	2.72
gly2	820	61551	34198	1.80
nma-dimer2	496	15439	4467	3.46
nma-acetone	452	10004	4699	2.13
nma-dimer1	496	14200	4656	3.05
nma-dme	422	8465	2999	2.82
nma-meoh	364	4835	1657	2.92

[a] nma = *N*-methyl acetamide, dme = dimethyl ether, gly2 = glycine dimer, ala2 = alanine dimer, nma-dimer1 and nma-dimer2 are two different nma dimers.

equilibrium gas-phase electronic binding energy for the formic acid dimer is 16.18 kcal/mol.

B. Test Set. As a test set, we use sets of five formamide molecular dimers recently studied in ref 1b using conventional MP2, CCSD (T), and extrapolation techniques. It can be seen in Table 2 that the extrapolated LMP2 results again are in excellent agreement, this time without any fitting to the high level data. This test further confirms the validity of the protocol we have developed.

C. Computational Effort. Table 3 presents the computational effort for PS-LMP2 calculations using the cc-pVTZ basis set for a number of small molecule dimers as a function of molecular size. All calculations were carried out on a 1.4GHz Pentium III PC. It can be seen that the scaling of computational effort with the size of the molecule is approximately in the N^3 range in accordance with previous studies.

Table 3 also compares PS-LMP2 timings with those obtained from the LMP2 methodology in the Q-Chem program; the PS-LMP2 results are 2−3 times faster in wall clock time than those from Q-Chem, version 2.0. The LMP2 method used in Q-Chem was in the TRIM (triatomics in molecules) formulation.[14] It is different in its details from that in Jaguar; one major difference is that the local virtual space is somewhat larger, leading to recovery of a higher fraction of the canonical MP2 correlation energy. It is far from clear, however, whether this leads to greater accuracy in computation of properties such as binding affinity. The key issue is convergence of energy differences to the extrapolated basis set limit. The present paper demonstrates unambiguously that our definition of the local virtual space[2,4] leads to highly accurate extrapolated results using relatively small basis set calculations as an input, at least across the range of test cases that we have considered here. The

Table 4. Comparison of NMA ab Initio Dimerization Energies in kcal/mol

dimer	CCSD(T) limit[a]	LMP2[b]	dimer	CCSD(T) limit[a]	LMP2[b]
nm__6	17.18	14.96	nm__8	10.76	8.61
nm__7	12.37	10.24	nm__9	9.67	19.61

[a] Reference 1b. [b] Plus extrapolation. This work. The same geometry as in reference 1b.

Table 5. NMA ab Initio Dimerization Energies Computed Using QChem, kcal/mol

dimer	cc-pVTZ(-f)[a]	cc-pVQZ(-g)[a]
nm__6	13.30	15.36
nm__7	8.72	10.72
nm__8	7.84	9.43
nm__9	18.98	20.95

[a] The same geometry as in reference 1b.

extrapolation properties of alternative local space definitions remain to be investigated.

D. Interaction Energies for *N*-Methyl Acetamide Dimers. The work described in ref 1b also included calculations for four *N*-methyl acetamide dimers, in a number of different hydrogen bonding geometries. These calculations were performed at the cc-pVTZ level, and a BSSE correction was then applied; the results are shown in Table 4. Unlike the case of formamide, it was not possible to more rigorously determine the basis set MP2 limit binding energy, due to the larger size of the molecules involved.

Table 4 also compares our extrapolated PS-LMP2 results for the NMA dimers with these results. It can be seen that there is a substantial discrepancy even when identical geometries are used to carry out the calculation (there are also some differences in the optimized geometries as well); in general, the PS-LMP2 binding affinities are ∼2 kcal/mol smaller. This result was unanticipated, because one would ordinarily expect the BSSE corrected results to exhibit a lower binding energy than the converged basis set limit calculations. Thus, there is a discrepancy between our PS-LMP2 results and the results of ref 1b, which cannot be explained by a simple difference in geometry.

To investigate the problem further, we ran the local MP2 methodology in the Q-Chem program, using the TRIM approach, which is supposed to more closely reproduce full MP2 calculations. Results for cc-PVTZ(-f) and cc-PVQZ(-g) calculations, including a BSSE correction, are given in Table 5. It can be seen that the TZ results are much smaller than those of ref 1b and that the QZ results are reasonably close to our extrapolated LMP2 results. Extrapolation of the TRIM results cannot be performed using our LMP2-specific formulation; there is no particular reason to believe that the formulas should have significant similarity. Thus, all that can be said at this point is that the extrapolated PS-LMP2 and TRIM results are not clearly in disagreement, based on the results reported here. However, the possibility that the TRIM results when extrapolated would exceed the PS-LMP2 extrapolated results cannot be ruled out, and if this is the case, the resulting discrepancy is one that also would require further investigation.

Hydrogen Bonding Energies from Pseudospectral LMP2

J. Chem. Theory Comput., Vol. 1, No. 2, 2005 **253**

As was discussed in the Introduction, a central application of high accuracy MP2-based calculations of binding energies is in the development of molecular mechanics force fields. In ref 8, we describe a protocol for developing a polarizable force field based on calculated dimer binding energies. The protocol employs electrostatic and polarization parameters fit to large basis set DFT calculations, universal $1/r^6$ dispersion parameters for each atom to describe the long-range component of the van der Waals interaction, and fitting of the short-range atom—atom pair potential to the binding energy of molecular dimers. We have carried out this protocol for NMA, using the two different sets of dimer binding energies for nm_6 in Table 4, but employing otherwise identical values for the dispersion parameters (which were fit to simulations of other small molecules, as is discussed in detail in ref 8), and the electrostatic model, which is determined from large basis set DFT calculations independent of the dimer binding energies. Thus, the models are identical apart from the short-range atom—atom pair terms. We then perform liquid-state simulations in an NPT ensemble, using methods described in ref 8, and determine the heat of vaporization and the density of the model for the liquid.

The results obtained are as follows. For the LMP2-derived parameters, the predicted heat of vaporization is 13.2 kcal/mol as compared to the experimental value of 13.4 kcal/mol; the density is 0.876 g/mol as compared to the OPLS-AA value of 0.897 (the OPLS-AA result was used as a reference, as no experimental data was available, and the other OPLS-AA properties of liquid NMA are in close agreement with experiment[15]). These results are entirely consistent with the other data presented in ref 8 for 16 molecular liquids, in which the average errors in the heat of vaporization were on the order of 0.2—0.3 kcal/mol, and errors in the density were under 5%. Note that no parameters in the calculation are fit to experimental data on NMA, so the results constitute a true test of the predictive capability of the methodology in the condensed phase. In contrast, if the NMA dimer nm_6 binding energy from ref 1b is used to fit the short-range nonbonded parameters, the simulations fail to converge and the heat of vaporization observed during the simulation period reaches ca. 40 kcal/mol, which is in qualitative disagreement with the experimental data. These condensed phase tests therefore provide further confirmation of the robustness of our PS-LMP2 extrapolation protocol. It is unclear at present what the explanation is of the problematic results reported in ref 1b. However, our liquid state results, in conjunction with the previously discussed small molecule comparisons, suggest that the PS-LMP2 protocol has produced an accurate binding affinity for NMA. Further exploration of the discrepancies noted above remains to be undertaken.

IV. Conclusion

In this paper we have shown that a PS-LMP2 based approach to the calculation of intermolecular interaction energy is both highly accurate and very efficient in terms of computational efficiency as compared to alternatives. Achieving both speed and high accuracy together is nontrivial. DFT methods, for example, would be considerably faster but have not yet been shown to yield the level of precision demonstrated above. When used to develop polarizable force fields for liquid-state simulations, this level of accuracy is inadequate to predict heats of vaporization to the desired level, whereas the PS-LMP2 values predict this and other condensed phase properties with good precision and reliability.

We are presently using this methodology to develop a complete polarizable force field for proteins and also for a wide range of organic chemical functional groups. However, there are some limitations which need to be addressed. First, as has been shown by Sherrill and co-workers,[16] extrapolated LMP2 methods are not adequate to achieve high accuracy for stacked benzene dimers; here, the CCSD (T) level corrections are significant. Fortunately, it appears as though these corrections can be performed with relatively small basis sets. Experiments need to be done to determine what other types of interactions require these higher-level correction. Examples of interactions that need to be investigated include those involving charged groups, different types of ring structures, and interactions involving a transition metal containing species.

References

(1) (a) Tsuzuki, S.; Uchimaru, T.; Matsumara, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906−11910. (b) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem.* **2001**, *105*, 4963−4968 and references therein.

(2) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481−1490.

(3) Jaguar v5.0, Schrödinger, Inc. Portland, OR, 2003.

(4) Murphy, R. B.; Pollard, W. T.; Friesner, R. A. *J. Chem. Phys.* **1997**, *106*, 5073−5084.

(5) See for example (a) Petersson, G. A.; Frisch, M. J. *J. Phys. Chem. A* **2000**, *104*, 2183−2190. (b) Park, S. Y.; Lee, J. S. *J. Chem. Phys.* **2002**, *116*, 5389−5394. (c) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802−11808. (d) Huh, S. B.; Lee, J. S. *J. Chem. Phys.* **2003**, *118*, 3035−3042.

(6) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(7) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y. H.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532−1548.

(8) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621−627.

(9) Werner, H. J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149−8160.

(10) (a) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660−3671. (b) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691−5705. (c) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592−3601. (d) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H. J. *J. Chem. Phys.* **2000**, *113*, 9443−9455. (e) Saebø, S.; Pulay, P. *J. Chem. Phys.* **2001**, *115*, 3975−3983. (f) Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772−8785. (g) Walter, D.; Szilva, K. N. A. B.; Carter, E. A. *J. Chem. Phys.* **2002**, *117*, 1982−1993.

(11) Schütz, M.; Rauhut, G.; Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997−6003.

(12) (a) Fomine, S.; Tlenkopatchev, M. Martinez, S.; Fomina, L. *J. Phys. Chem. A* **2002**, *106*, 3941−3946. (b) Huang, N.; MacKerell, A. D. *J. Phys. Chem. A* **2002**, *106*, 7820−7827.

(13) See for example (a) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104−6114. (b) Martin, J. M. L. *J. Chem. Phys.* **1994**, *100*, 8186−8193. (c) Wilson, A. K.; Dunning, T. H. *J. Chem. Phys.* **1997**, *106*, 8718−8726. (d) Valeev, E. F.; Allen, W. D.; Hernandez, R.; Sherrill, C. D.; Schaefer, H. F. *J. Chem. Phys.* **2003**, *118*, 8594−8610.

(14) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592−3601.

(15) Jorgensen, W. L.; Maxwell D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *113*, 11225−11236.

(16) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.