Journal of Physical & Theoretical Chemistry
Islamic Azad University of Iran 3 (3)
(2006)
Science and Research Campus
ISSN: 1735-2126

Ab initio study of the second virial coefficient protein – protein on the basis of intermolecular potential energy surface

K.Zare[1,3], M.Khaleghian[2], H.Monajemi[3], M.Monajjemi[4*] .

1. Department of Chemistry,Shahid Beheshty University, Tehran,Iran
2. Department of Chemistry, Young Researchers Club ,Varamin Campus, Islamic Azad University, Tehran,Iran
3. Department of Physics , Karaj Branch , Islamic Azad University , Karaj , Iran
4. Department of Chemistry, Science and Research Campus, Islamic Azad University, Tehran,Iran

ABSTRACT

Intermolecular potential energy surface (IPS) for protein – protein has been examined using RHF, DFT-B_3LYP and MP_2 levels of theory with 6-31G, 6-31G* basis sets. A number of basis sets were used in order to evaluate the basis set effects, at all three levels of theory, basis sets has significant effects on the calculated potential energy curves (including position, depth and width of the potential well). Counterpoise (CP) correction has been used to show the extent of the basis set superposition error (BSSE) on the potential energy curves obtained for protein – protein system. The deepest BSSE-corrected potential well have been obtained at B_3LYP level of theory with 6-31G basis set. The second virial coefficients calculated this way are fitted to the initial coefficients B_2 varying ε and r_0, eventually some other parameters.

---

* . Corresponding author: E-mail: m_monajjemi @Yahoo.com

INTRODUCTION

Knowledge of the intermolecular interaction potential is basic for understanding the properties of gases, liquids and solids. In principle, the evaluation of a macroscopic property like the second virial coefficient of a moderately dense gas is straightforward if the intermolecular pair potential is accurately known. The behavior of molecular clusters linked by hydrogen bonds is of special interest with a view to understanding a wide variety of chemical and biochemical problems. Theoretical calculations provide detailed information about some aspects of the molecular interaction and the most likely structures for the clusters, which can rarely be accessed experimentally. Ab initio quantum mechanical calculations offer a way to obtain intermolecular potentials of molecules. This approach can be used to extract detailed information of the potential energy surface, which is sometimes difficult or practically impossible by other methods. Nevertheless, the quality of the potential is sensitive to the level of theory used for the calculation of the interaction energies. Also the basis set superposition error (BSSE) has a significant effect on the calculated interaction potential and therefore it should be corrected for. Ab initio methods for calculating intermolecular potential energy surfaces (IPS) have been reviewed by van der Avoird et al.and van Lenthe et al. Theoretical studies of van der Waals complexes and intermolecular forces have been reviewed by Buckingham et al. . In this work, basis set effects on the calculated IPS of the ALA-METH system have been investigated. In this study, RHF, MP2 and DFT-B3LYP methods have been used with various basis sets to find the most appropriate basis set(s), which is (are) suitable for the derivation of the ALA-METH IPS.

Characteristics

Classification: Protein-protein interactions can be arbitrarily classified based on the proteins involved (structural or functional groups) or based on their physical properties (weak and transient, "non-obligate" vs. strong and permanent). Protein interactions are usually mediated by defined domains, hence interactions can also be classified based on the underlying domains.

Universality: All of molecular biology is about protein-protein interactions (Alberts et al. 2002, Lodish et al. 2000). Protein-protein interactions affect all processes in a cell: structural proteins need to interact in order to shape organelles and the whole cell, molecular machines such as ribosomes or RNA polymerases are hold together by protein-protein interactions, and the same is true for multi-subunit channels or receptors in membranes.

Specificity distinguishes such interactions from random collisions that happen by Brownian motion in the aqeous solutions inside and outside of cells. Note that many proteins are known to interact although it remains unclear whether certain interactions have any physiological relevance.

Protein-protein interactions and protein complexes: Most protein-protein interactions are detected as interacting pairs or as components of protein complexes. Such complexes may contain dozens or even hundreds of protein subunits (ribosomes, spliceosomes etc.). It has even been proposed that all proteins in a given cell are connected in a huge network in which certain protein interactions are forming and dissociating constantly.

Structural features of protein-interaction sites

Hundreds of protein complexes have been analyzed by X-Ray crystallography and other methods. Data about the structures of proteins and complexes are available from the Protein Databank (PDB, http://www.rcsb.org/). The following statements about the geometry and energetics of protein interactions have been drawn from the analysis of several dozens to about a 100 protein pairs and complexes that have been crystallyzed.

Forces that mediate protein-protein interactions include electrostatic interactions, hydrogen bonds, the van der Waals attraction and hydrophobic effects.

The average protein-protein interface is not less polar or more hydrophobic than the surface remaining in contact with the solvent. Water is usually excluded from the contact region. Non-obligate complexes tend to be more hydrophilic in comparison, as each component has to exist independently in the cell.

It has been proposed that hydrophobic forces drive protein-protein interactions and hydrogen bonds and salt bridges confer specificity.

Van der Waals interactions occur between all neighbouring atoms, but these interactions at the interface are no more energetically favourable than those made with the solvent. However, they are more numerous , as the tightly packed interfaces are more dense than the solvent and hence they contribute to the binding energy of association. Hydrogen bonds between protein molecules are more favourable than those made with water.

Ab initio calculations of the interaction energy in the system protein-protein

Initially, structures were fully optimized with the HF method in order to locate the stationary points on the potential surface for the system. Ab initio calculations are necessary to throw some light on this subject. Our estimates are only approximate but interesting, nevertheless.

The interaction energy, V, for two A and B systems can simply be given as

$$V = E_{AB}\,(A\cdots B) - E_{AB}\,(A + B) \qquad (1)$$

where the arguments in parenthesis indicate the basis set being used. $E_{AB}(A\cdot\cdot\cdot B)$ is the energy of the $A\cdot\cdot\cdot B$ system at the r distance while $E_{AB}(A+B)$ is the energy of the two isolated components, at infinity ($r = \infty$). In ab initio calculations the basis set superposition error (BSSE) is of paramount importance [25]. This error can be eliminated to some extent by using the counterpoise method (CP) . In this method both the physicochemical compound $A\cdot\cdot\cdot B$ and the A and B components at $r=\infty$ are calculated by using the full basis set for the $A\cdot\cdot\cdot B$, hence

$$V = E_{AB}\,(A\cdots B) - E_{AB}\,(A + B) + \Delta E_{CP} \quad (2)$$

Where

$$\Delta E_{CP} = [E_A\,(A+B) - E_A\,(A\cdots B)] + [E_B\,(A+B) - E_B\,(A\cdots B)] \qquad (3)$$

Our calculations were performed by using the program package GAUSSIAN 98 exclusively.

Instructions for carrying a counterpoise correction.

1. For the calculation of the dimmer AB, check the basis set output box. Then the basis set will be written into the output file after the line
Basis set in the form of general basis input:
2. As preparation for the calculation of monomer A replace all atoms of monomer B in the Z-matrix of AB by ghost atoms. Choose basis set "GEN". If the dimmer basis set contains 6 functions in one set of d-functions, you must check the corresponding box. Choose a new filename and generate the program input.
3. Copy the dimmer basis set to the end of the program input that is automatically generated. There must be exactly one blank line before and at least two blank lines after the basis set section of the input.
4. Repeat step 2 and 3 for monomer B.

In step 3 you have to copy a basis set .One way to do this is:

- Select the basis set with the mouse.
- Copy the selected text into the clipboard by pressing "Copy" in the Edit menu of your browser.
- Click at the end of automatically generated input in the text area of the

molecular input form. Paste the basis set by pressing "Paste" in the Edit menu.

Effects of basis sets

To evaluate the effect of basis sets, we calculated the intermolecular interaction energies of the ALA- METH system at RHF, DFT-B3LYP and MP2 levels of theory with 6-31G, 6-31G* basis sets. The ALA- METH intermolecular potential energy interactions obtained at RHF, B3LYP and MP2 levels of theory with the basis set 6-31G are plotted in the Figs.1 as functions of R; the distance between ALA and METH. It can be seen from these figures, that at all three levels of theory, basis sets has significant effects on the calculated potential energy curves (including position, depth and width of the potential well).Furthermore, the following trends with basis set can be deduced for the depth (D_e= $-E_{int}$ (R_e) = $-E_{min}$) of the potential well of the calculated IPSs. Numerical values of D_e are given in Table 1. As can be seen from Figs. 2 and Table 1, the potential energy curves with the largest value of D_e are obtained with 6-31G basis set, at B3LYP level of theory. The calculated IPSs can further be compared based on the values of the position of the minimum point (R_e) of the potential curves. As is evident from this table, these quantities are very sensitive to the basis set used in the computations.In ab initio calculations the basis set superposition error (BSSE) is of paramount importance . This error can be eliminated to some extent by using the counterpoise method (CP) . BSSE-corrected ALA-METH intermolecular potential energy curves corresponding to those calculated at RHF, B3LYP and MP2 levels of theory are plotted against R in Figs. 2.

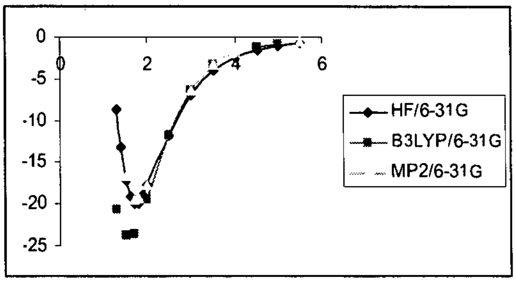


Fig.1. The ALA- METH intermolecular potential energy interactions obtained at RHF, B3LYP and MP2 levels

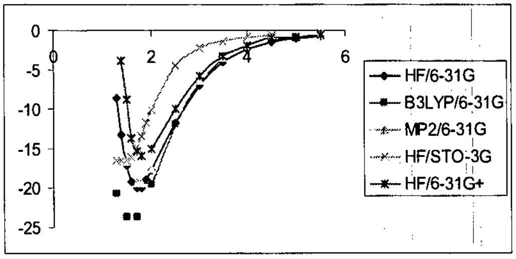


Fig.2.The ALA-METH intermolecular potential energy interactions obtained at RHF, B3LYP and MP2 levels of theory with 6-31G, 6-31G* , STO-3G basis sets.

In this work, to estimate the interaction energy U (r) in the system we use the Kihara formula.

Table1. Numerical values of D_e and R_e

Method	**Basis set**	**D_e**	**R_e**
MP_2	6-31G	-19.360	1.7
B_3LYP	6-31G	-23.699	1.5
RHF	6-31G	-19.996	1. 7
	6-31G*	-15.932	1.8
	STO-3G	-16.529	1.3

This has three adjustable parameters

$U(r)=4\varepsilon[(\sigma-2a/r-2a)^{12}-(\sigma-2a/r-2a)^{6}]$.

where a is the radius of the molecular coreat which U (r) =∞ (Kihara, T.Rev. mod.phys. 25, 831 (1953)).

Some of the parameters are treated as constants, taken from the literature, and some parameters are varied. The geometrical parameters for the system , the depths ε and the locations σ of the interacting centers are treated as constant .The set of necessary parameters is listed in Table 2. Assuming a given set of parameters, we estimate theoretically second virial coefficients for the system, Eq. (4). The form for the second virial coefficient, derived using statistical mechanics for the nonspherical symmetric surfaces of interaction energy, can be expressed as

$$B_2 = \frac{1}{(8\pi^2)^2}\frac{1}{2}N_A \int_0^{\pi}\sin(\vartheta)d\vartheta \int_0^{2\pi} d\varphi \int_0^{\pi}\sin(\beta_1)d\beta_1 \int_0^{2\pi} d\alpha_1 \int_0^{2\pi} d\gamma_1 \int_0^{\pi}\sin(\beta_2)d\beta_2 \int_0^{2\pi} d\alpha_2 \int_0^{2\pi} d\gamma_2 \int_0^{\infty}\left\{1-\exp\left[-\frac{V_{NB}}{RT}\right]\right\} r^2 dr \qquad (4)$$

where N_A is the Avogadro constant. The symbols $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2$ are the Euler angles describing the orientation of the system of coordinates connected rigidly with the first and second reactant, respectively. The spherical coordinates r, ϑ, φ describe the mutual location of the reactant centers of mass.

Table2. Adjustable parameters used in the Kihara formula.

Adjustable parameters	**Values**
ε / k	19.31732949
σ	1.137795301
a	-1.01190983

These values can be defined as described in Ref. [12]. Integration over angles require the use of the normalization factor $(8\pi^2)^2$. To estimate the multiple integral (4) by the Monte-Carlo method we have to fix the number of necessary random points N and also the upper limit of the integral, r_{max}. Both these values are strictly linked with the analyzed system and the temperature range. The simplest way to determine these values is as follows. For a given r_{max}, we search for N starting from which the first three digits of integral (4) are fixed. Similarly, r_{max} is selected so that its further increase does not result in any changes in integral (4).

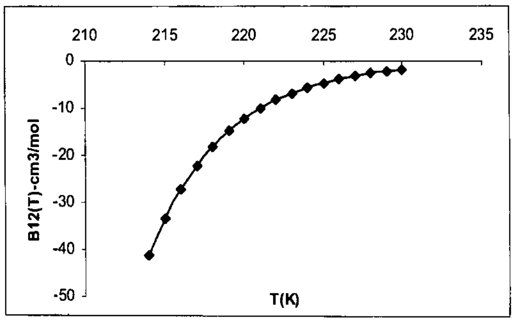
Table3. second virial coefficients for selected temperatures.

T(k)	B_2(cm^3/mol)
214	-41.19
215	-33.44
216	-27.19
217	-22.16
218	-18.09
219	-14.8
220	-12.13
221	-9.954
222	-8.186
223	-6.744
224	-5.566
225	-4.601
226	-3.81
227	-3.16
228	-2.626
229	-2.185
230	-1.822

The hard sphere approximation is very important in chemical kinetics. It is associated closely with average interaction energy between reactants that interact at different distances and at different orientations over some region of temperatures. Using this approximation we can obtain simple estimates of second virial coefficients, collisional frequencies, statistical sums and other parameters referring to the real reagents, even though these reagents are not spheres.

The form for the function describing the spherically symmetrical potential, V(r), the depth of the minimal interaction, referred to by the symbol ε (or D_0), and its location, r_0, are all very important in the hard sphere approximation.

Fig..3. Temperature dependence of second virial coefficients .



Virial coefficients B_2 obtained either experimentally or estimated theoretically using the multidimensional interaction energy surface V_{NB}, Eq. (4), can be used to determine the magnitude of these parameters. Assuming the known form for V(r) the hard sphere approximation leads to the following expression for the estimation of second virial coefficients :

$$B_2 = 2\pi N_A \int_0^\infty \left\{1 - \exp\left[-\frac{V(r)}{RT}\right]\right\} r^2\, dr \qquad (5)$$

The second virial coefficients calculated this way are fitted to the initial coefficients B_2 varying ε and r_0, eventually some other parameters. It was thought for many years that the values for the second virial coefficients do not depend on the shape of the curve V(r) for the energy interaction but only on the integral that corresponds to the area restricted by this curve. It is now known that this is not completely true . Still, the above-mentioned dependence of the second virial coefficient mostly on the integral over r remains the main drawback of the method that uses second virial coefficients to determine molecular interaction energy V(r). Especially, since we assume the semi-rigid reagent structure, the method described above cannot yield unique solutions. B_2 estimated for different V(r) and different ε and r_0 may be very close, nevertheless.To circumvent this deficiency we suggest introduction of additional criterion for curve fitting. The inclusion of the elements of the curve in the form of independently assessed points restricts to a large extent the arbitrariness in the choice of the shape of the curve for the energy interaction V(r).

Although V(r) in the hard sphere approximation refer to idealized model of spherically symmetrical molecules, some elements of this potential over the region of long-range interactions, are interlinked closely with the interactions of real nonspherical molecules.

To determination of virial coefficients B2, Eq. (4), using VAB obtained earlier over the range of temperatures of interest (Table 3). The B2 coefficient can, of course, be taken from experiment, if the data are available. Calculated results are plotted in Fig.3 and are shown for selected temperatures in Table 3.

REFERENCES

1. Bouanich JP.Site-site Lennard-Jones potential parameters for N2,O2,H2, CO and CO2.JQSRT 1992;47:243–50.
2. Gray CG, Gubbins KE.Theory of molecular uid.Oxford: Clarendon, 1984.p.33– 4.
3. Robert D, Bonamy J.Short range force eects in semi-classical molecular line broadening calculations.J Phys (Paris) 1979;40:923–43.
4. Gamache RR, Lynch R, Neshyba SP.New developments in the theory of pressure broadening and pressure-shifting of spectral lines of H2O: the complex Robert-Bonamy formalism.JQSRT 1998;59:319–35.
5. Bouanich JP, Lambot D, Blanquet G, Walrand.N2-and O2-broadening coecients of C2H2 IR lines.J Mol Spectrosc 1990;140:195 –213.
6. Lynch R, Gamache RR, Neshyba SP.Full complex implementation of the Robert–Bonamy formalism: half widths and line shifts of H2O broadened by N2.J Chem Phys 1996;105:5711–21.
7. Gamache RR, Lynch R.Argon-induced halfwidths and line shifts of water vapor transitions.JQSRT 2000;64:439–56.

8. Colmont JM, Priem D, Wlodarczak G, Gamache RR.Measurements and calculations of the halfwidths of two rotational transitions of water vapor perturbed by N2,O2 and air.J Mol Spectrosc 1999;193:233–43.
9. Lu Z, Tabisz GC, Ulivi L.Temperature dependence of the pure rotational band of HD: interference, widths, and shifts.Phys Rev A 1993;47:1159–73.
10. Dymond JH, Smith EB.The virial coecients of gases: a critical compilation.Oxford: Clarendon, 1969.
11. J.R. Stallcop, H. Partridge, Chem. Phys. Lett. 281 (1997) 212.
A. van der Pol, A. van der Avoird, P.E.S. Wormer, J. Chem. Phys. 92 (1990) 7498.
12. P. Diep, J. Johnson, J. Chem. Phys. 112 (2000) 4465.
13. F.R.W. McCourt, Private Communication via e-mail, Department of Chemistry, University of Waterloo.
14. M. Head-Gordon, J.A. Pople, M.J. Frisch, Chem. Phys. Lett. 153 (1988) 503.
15. M.J. Frisch, M. Head-Gordon, J.A. Pople, Chem. Phys. Lett. 166 (1990) 275.
16. M.J. Frisch, M. Head-Gordon, J.A. Pople, Chem. Phys. Lett. 166 (1990) 281.
17. A.D. Becke, Phys. Rev. A 38 (1988) 3098.
18. A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
19. C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785.
20. B. Miehlich, A. Savin, H. Stoll, H. Preuss, Chem. Phys. Lett. 157 (1989) 200.