Quantum mechanical map for protein-ligand binding with application to β-trypsin/benzamidine complex

Da W. Zhang, Yun Xiang, Ai M. Gao, and John Z. H. Zhang

Department of Chemistry, New York University, New York, New York 10003

(Received 14 November 2003; accepted 17 November 2003)

We report full ab initio Hartree–Fock calculation to compute quantum mechanical interaction energies for β-trypsin/benzamidine binding complex. In this study, the full quantum mechanical ab initio energy calculation for the entire protein complex with 3238 atoms is made possible by using a recently developed MFCC (molecular fractionation with conjugate caps) approach in which the protein molecule is decomposed into amino acid-based fragments that are properly capped. The present MFCC ab initio calculation enables us to obtain an “interaction spectrum” that provides detailed quantitative information on protein-ligand binding at the amino acid levels. These detailed information on individual residue-ligand interaction gives a quantitative molecular insight into our understanding of protein-ligand binding and provides a guidance to rational design of potential inhibitors of protein targets. © 2004 American Institute of Physics. [DOI: 10.1063/1.1639152]

Fully quantum mechanical computation of protein interaction energy presents a grand challenge to computational chemists. The large size of proteins with thousands of atoms makes it impractical to attempt brute force ab initio calculations using standard quantum chemistry methods. Standard applications of ab initio methods like HF (Hartree–Fock) and DFT (density functional method) are typically limited to systems of around 100 atoms on standard workstations due to steep scaling of computational time with the system size N. Efforts have been made to develop linear scaling methods to treat large biological systems over the past decade. However, current applications of these methods are limited to semi-empirical calculations for large biological systems like proteins due to huge computational cost. Recently, a number of methods have been proposed to compute interaction energy of large molecules fully ab initio, especially for computing protein interaction energies by decomposing a protein molecule into amino acid-based fragments.

In this Communication, we report benchmark ab initio computation of interaction energies for β-trypsin/benzamidine binding at Hartree–Fock level. The full ab initio computation is made possible by applying a recently developed MFCC method. The MFCC method is linear scaling, computationally efficient, and particularly suitable for calculating interaction energy of biopolymers on multi-processor computer systems. The MFCC method has been successfully applied recently to a real protein-ligand system, the streptavidin/biotin binding complex with a total of 1775 atoms. In the MFCC approach, the protein is decomposed into amino acid-based fragments that are properly capped.

The β-trypsin is one of the trypsinlike serine proteases which play critical roles in physiological processes such as blood coagulation, complement activation, digestion, fibrinolysis and so on. The discovery of the effectiveness of benzamidine as an anti-trypsin agent by Mares-Guia and Shaw in 1965 triggered extensive research on the development of benzamidine-based inhibitors of serine proteases. It is known that the amidophenyl moiety of benzamidine sits deep in the primary substrate binding pocket (the S1 pocket) by both electrostatic interaction and hydrogen bonds to the Asp189 residue of the protein at the bottom of the pocket and to other residues that form the walls of the pocket. This system has also served as an important focus for a wide variety of experimental and computational studies of ligand dissociation due to the rigid property of benzamidine. Theoretical calculations of free energies based on molecular dynamics simulation for β-trypsin/benzamidine binding have also been reported.

Using the MFCC approach, we first decompose the 223-amino acid β-trypsin into 223 fragments by cutting all the backbone peptide (C–N) bonds. At every position of cut, a pair of caps (CH₃CO- and CH₃NH-) are inserted to cap the cutoff fragments. Thus, there are also 222 cap species formed by fusing pairs of conjugate caps inserted at all positions of cuts. In addition, there are three disulfide bonds that are also cut and capped with SCH₂ group. The β-trypsin/benzamidine interaction energy (V(B-T)) is given by the expression

\[ V = \sum_{i=1}^{223} V(B-F_i) - \sum_{j=1}^{222} V(B-C_j) - \sum_{k=1}^{3} V(B-D_k), \]

where \( V(B-F_i) \), \( V(B-C_j) \), and \( V(B-D_k) \) denote, respectively, the interaction energy of benzamidine with the ith fragment of trypsin, the ith cap fragment, and the three disulfide bond caps.

The MFCC ab initio calculation at HF/3-21G level is carried out to generate intermolecular potential energies. Al-
though higher level \textit{ab initio} methods and larger basis set could be used to perform the calculation, it would be much more computationally expensive. Although the HF/3-21G calculation usually gives binding energies that are typically too large, but the energy profile or structure of the binding complex is quite reliable. In the MFCC approach, both \(\beta\)-trypsin and benzamidine are kept rigid with structures given by PDB (protein data bank) id 3ptb. Benzamidine molecule is almost planar with a tiny bending of the two NH2 groups around C\(-\)C bond toward opposite directions. Figure 1 shows the crystal structure of \(\beta\)-trypsin/benzamidine complex obtained from PDB.

In our calculation, we fix the protein in three-dimensional space, and calculate an interaction potential curve around the binding structure by moving the rigid benzamidine along the Z-axis relative to the fixed protein. At each step of the move, we optimize three Euler’s angles \((\theta, \phi, \chi)\) by energy minimization. This generates a one-dimensional potential curve shown in Fig. 2 with the starting crystal structure of the complex as the origin of the coordinate. Here the energy is normalized such that the infinite separation of the rigid protein and benzamidine has zero energy.

As shown in Fig. 2, the quantum \textit{ab initio} and AMBER force field\textsuperscript{15} calculations give similar minimum energy profiles around the crystal structure. This must be related to the fact that \(\beta\)-trypsin/benzamidine binding is dominated by a strong ionic interaction between a negatively charged Asp group and the positively charged benzamidine as will be discussed in more detail later. We note here that the binding energy from the HF/3-21G calculation is about 40 kcal/mol lower than that from the force field calculation.

The MFCC method provides an easy means to extract molecular insight of protein-ligand binding by explicitly calculating the interaction energy between individual residues and the ligand. These individual residue-ligand interaction energies provide detailed quantitative information about specific residue interaction with the ligand that should be extremely useful to our understanding of the molecular nature of protein-ligand binding. Figure 3 shows such a quantum

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure1.png}
\caption{Structure of \(\beta\)-trypsin/benzamidine binding complex from protein data bank (PDB id 3ptb).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure2.png}
\caption{The \(\beta\)-trypsin/benzamidine interaction energy curves along the Z direction. The origin of the abscissor corresponds to the crystal structure from PDB. (a) Results from AMBER force field. (b) Results from MFCC HF/3-21G calculation.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure3.png}
\caption{Interaction spectrum of \(\beta\)-trypsin/benzamidine binding complex. The solid lines represent interaction energies between individual capped residues and the ligand (benzamidine) and the short dashed lines represent interaction energies between individual cap pairs and the ligand.}
\end{figure}
mechanical “interaction spectrum” or map. If the caps don’t have significant contributions to the interaction energy as is usually the case as shown in Fig. 3, the interaction spectrum gives quantitative interaction energy between individual residue and the ligand benzamidine. As we can see from Fig. 3, the ionic interaction between the negatively charged Asp189 of the protein and the positively charged benzamidine is a dominant mechanism of binding. It is useful to point out that the labeling of residue in Fig. 3 is by sequence starting from one, which is different from that given by PDB id 3ptb whose first residue starts at 16. Thus the label Asp189 and other residue labels to be mentioned below are adopted from that given in PDB entry.

To assess the accuracy of the HF/3-21G calculation, we performed ab initio calculations to calculate the individual interaction energy between the capped Asp189 and the benzamidine at different levels of ab initio theory and basis sizes as shown in Fig. 4. The results in Fig. 4 shows that inclusion of polarization functions does not have much effect on calculated energy. In contrast, the inclusion of diffuse functions lowers the binding energy by about 10 kcal/mol in DFT/B3LYP calculation and by about 5 kcal/mol in MP2 calculation. Both DFT-B3LYP and MP2 calculations including diffuse function give the binding energy within 1 kcal/mol of each other. For comparison, the binding energy from HF/3-21G calculation is larger by about 18 kcal/mol in comparison.

In molecular force field calculations, the partial charges on atoms are fixed asymptotically and no polarization effect is included. Since β-trypsin/benzamidine binding is dominated by strong ionic interaction as shown in Fig. 3, it is interesting to investigate how strong the charge redistribution or polarization effect is. For that purpose, we carried out Mulliken population analysis to examine the change of partial charges from isolated Asp189 and benzamidine to the Asp189/benzamidine binding complex. The result from HF/6-31G* calculation given in Table I shows that the charges on Asp189 is strongly redistributed.

In addition to the dominant Asp189 residue, there are other protein residues that have strong attractive interactions with the ligand. Figure 5 shows geometries of 12 residues that make the largest contributions to benzamidine binding according to the interaction spectrum in Fig. 3. These detailed quantitative information on individual residue-ligand interaction energy gives insight into molecular mechanism in protein-ligand binding and provides useful and practical tool in rational design of protein inhibitors.

It is useful to point out that the present ab initio calculation is for gas-phase environment and does not include solvent effect. Although the solvent (water) will modify the protein-ligand interaction energy by weakening some charge interaction due to charge screening by polar water molecules, we expect the main quantitative feature of the present calculation remains valid in liquid environment. It will be desirable to extend current ab initio calculation to include solvent effect in future studies.

A few words are in order in regard to the computational aspect of the MFCC method. A single point energy calculation for β-trypsin/benzamidine interaction energy at the HF/3-21G level takes about 10 hours on a single processor Intel Pentium 4 running linux. Since the MFCC method is linear.

TABLE I. Mulliken population analysis for Asp189 and benzamidine. The $O_1^-$ and $O_2^-$ refer to the oxyanions in Asp189 and $C^+$ refers to the carbonium ion in amidinophenyl moiety of benzamidine.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$O_1^-$</th>
<th>$O_2^-$</th>
<th>$C^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp189+benzamidine</td>
<td>-0.101</td>
<td>-0.237</td>
<td>0.806</td>
</tr>
<tr>
<td>Asp189/benzamidine</td>
<td>-0.773</td>
<td>-0.773</td>
<td>0.766</td>
</tr>
</tbody>
</table>

FIG. 4. Interaction energies of Asp189/benzamidine at different levels of ab initio calculations.

FIG. 5. Geometries of 12 residues that contribute most attractive interactions to β-trypsin/benzamidine binding.
scaling and easily parallelizable, computation on multi-
processor or multi-node computer systems would make the
MFCC calculation extremely efficient.