JCTC Journal of Chemical Theory and Computation

Kernel Energy Method: The Interaction Energy of the Collagen Triple Helix

Lulu Huang,[†] Lou Massa,[‡] and Jerome Karle*,[†]

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375-5341, and Hunter College and the Graduate School, City University of New York, New York, New York 10021

Received March 16, 2007

Abstract: There is a rapid growth in computational difficulty with the number of atoms when quantum mechanics is applied to the study of biological molecules. This difficulty may be alleviated in two different ways. One is the advance of parallel supercomputers. And the second is the use of a quantum crystallographic formalism based upon quantum kernels. The kernel methodology is well suited for parallel computation. Recently published articles have applied these advances to calculate the quantum mechanical ab initio molecular energy of peptides, protein (insulin), DNA, and RNA. The results were found to have high accuracy. This paper shows that it is possible to use the full power of ab initio quantum mechanics to calculate the interaction of long chain molecules of biological and medicinal interest. Such molecules may contain thousands or even tens of thousands of atoms. In the approach presented here the computational difficulty of representing a molecule increases only modestly with the number of atoms. The calculations are simplified by representing a full molecule by smaller "kernels" of atoms. The general case is illustrated by a specific example using an important protein, viz., a triple helix collagen molecule of known molecular structure. In order for such a molecule to be a stable helix, the overall interactions among the chains must be attractive. The results show that such interactions are accurately represented by application of the KEM to this triple helix.

I. Introduction

The Kernel Energy Method (KEM) calculates the quantum mechanical molecular energy by the use of the parts of a whole molecule, called kernels. The kernels are chosen to be much smaller than a full biological molecule. Thus the calculations of kernels and double kernels are in a practical way, doable. The kernel contributions are summed to obtain the energy of a whole molecule. In this way the task of calculating a quantum mechanical energy is simplified. Also, the computational time is much reduced. Previous work has shown that the accuracy obtained appears to be satisfactory.

The first applications of the KEM[1] referred to above involved a number of peptides. Good accuracy was retained throughout a wide range of basis functions and all of the most commonly used computational methods[2] that were studied. It was also found that good results were obtained in application of the KEM to the protein, insulin,[3] and also to A, B, and Z DNA,[4] RNA,[5] and the rational design of drug.[6] Theoretical background for the application of quantum mechanics to known molecular structures may be found in refs 7−14. References, that review the quantum mechanical methods related to computing the properties of large molecules from fragments, may be found in two articles referenced in this paper.[7,9]

This paper combines a collagen molecule of given structure[15] with quantum-mechanical KEM calculations to obtain the energies and interaction energies of a triple helix.

* Corresponding author e-mail: Jerome.karle@nrl.navy.mil.
[†] Naval Research Laboratory.
[‡] City University of New York.

10.1021/ct7000649 CCC: $37.00
Published on Web 06/07/2007

It is knowledge of such energetics which allows one to understand the stability of known structures and the rational design of new protein interacting chains. It is shown that the kernel energy method accurately represents the energies and interaction energies of each of the chains separately and in combinations with one another. This is a challenging problem for the case of large molecular protein chains. But here the computational chemistry calculations are simplified, and the information derived from the atomic coordinates of the structure is enhanced by quantum mechanical information extracted there from. For the sake of completeness, the main ideas of the KEM are reviewed in the next section.

II. Review of the Kernel Energy Method

In the KEM, the knowledge of atomic coordinates is combined with quantum mechanics. Central to the KEM is the concept of the kernel. These are the quantum pieces which, when summed together, represent the whole molecule. Quantum calculations are carried out on kernels and double kernels only. All properties of the full molecule may be reconstructed from those of the kernels and double kernels. Given a known molecular structure, a molecule may be mathematically broken into tractable pieces called kernels, all of whose atomic coordinates are known. Each atom occurs in only one kernel. The total molecular energy is calculated in this paper by summation over the energy contributions of all double kernels reduced by those of any single kernels which have been over counted in the sum over double kernels.

In connection with the kernels and double kernels, we mention that in the KEM, the fragment calculations are carried out on double kernels and single kernels whose ruptured bonds have been mended by the attachment of H atoms. In the summation of energies the contribution of hydrogen atoms introduced to saturate the broken bonds tends to zero, on the assumption that the energy added by hydrogen atoms is transferable among the kernels and double kernels. The energy of the hydrogen atoms added to the double kernels effectively cancels that of the hydrogen atoms added to the pure single kernels, which enter with opposite sign. This cancellation of the mending hydrogen atom energy effects contributes to the accuracy achieved by the KEM.

The total energy is

$$E_{\text{total}} = \sum_{m=1}^{n-1} \Big(\sum_{\substack{i=1 \\ j=i+m}}^{n-m} E_{ij} \Big) - (n-2) \sum_{i=1}^{n} E_i \qquad (1)$$

where E_{ij} = energy of a double kernel of name *ij;* E_{i} = energy of a single kernel of name *i; i, j, m* = running indices; and n = number of single kernels.

The validity of this approximation, in the case of a variety of peptides, proteins, DNA, RNA, and drug structures, has been shown in previous works.[1,3−6] In this paper we depend upon the known ab initio accuracy of the KEM to show how it may be applied to a triple helix collagen molecule, 1A89,[15] of given molecular structure, to obtain its relevant interaction energies, defined next.

III. The Interaction Energy

The definition of the interaction energy between any pair of kernels is

$$I_{ij} = E_{ij} - E_i - E_j \qquad (2)$$

where the subscript indices name the pair of kernels in question, I_{ij} is the pair interaction energy, E_{ij} is the energy of a double kernel, and E_i and E_j are each the energies of a single kernel. The sign of the interaction energy, I_{ij}, indicates whether the kernels *i* and *j* attract (negative *I)* or repel (positive *I*). The total interaction energy is a sum of the pair interaction energies of the individual double kernels. The magnitude of a given pair interaction energy I_{ij} determines its relative importance to the total molecular interaction energy.

A generalization of eq 2 gives the interaction energy for a particular pair of molecular chains, *a* and *b*, as

$$I_{ab} = E_{ab} - E_a - E_b \qquad (3)$$

where the subscript indices name the pair of protein chains in question, I_{ab} is the chain pair interaction energy, E_{ab} is the energy of a chain pair, and E_a and E_b are each the energies of a single protein chain. The sign of the interaction energy, I_{ab}, indicates whether the protein chains, *a* and *b*, attract (negative *I*) or repel (positive *I*). An equation analogous to eq 3 applies for the other chain pairs *ac* and *bc*.

The interaction energy among a triplet of protein chains is generalized to

$$I_{\text{abc}} = E_{abc} - (E_a + E_b + E_c) \qquad (4)$$

where the subscript indices name the triplet of protein chains in question, I_{abc} is the triplet chain interaction energy, E_{abc} is the energy of a triplet of chains, and E_a, E_b, and E_c are each the energies of a single protein chain. Again importantly, the sign of the interaction energy, I_{abc}, indicates whether the triplet of protein chains *a*, *b*, and *c* altogether attract (negative *I*) or repel (positive *I*). The magnitude of the interaction energies flows naturally from implementation of the KEM. The KEM delivers the ab initio quantum mechanical interaction energy between and among protein chains. And, this may be envisioned to be computationally practical for molecular structures containing thousands or even tens of thousands of atoms.

IV. Collagen

Collagen is a protein, essential to the physical structure of the animal body. The molecule is made of three peptide chains forming a triple helix. These are incorporated in a vast number of ways to create structure. Collagen molecular cables provide strength in tendons, resilience to skin, support to internal organs, and a lattice structure to the minerals of bones and teeth. A repeated sequence of three amino acids forms the chains out of which the collagen triple helix is composed. Every third amino acid is glycine. Remaining positions in the chain often contain proline and hydroxyproline.

We have selected for study a particular collagen molecule whose molecular structure is known, 1A89,[15] and whose

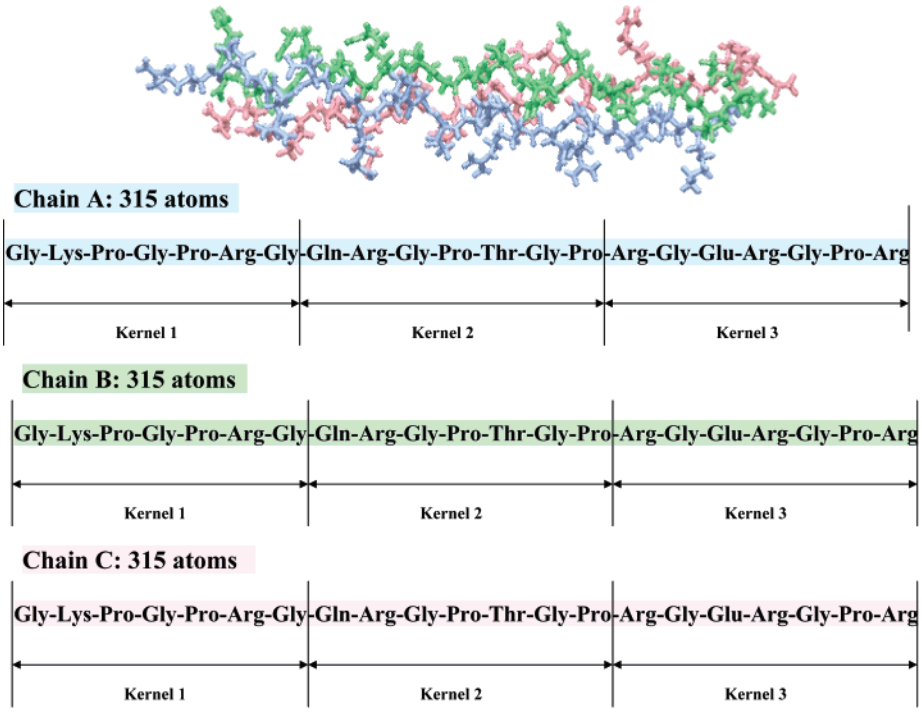

Figure 1. A picture of the collagen triple helix, 1A89, and the primary structure of each of its individual protein chains broken into kernels.

Table 1. Energy Calculations of Collagen Triple Helix by HF/STO-3G

chain	atoms	kernels	E_{HF} [au]	E_{KEM} [au]	$E_{HF}-E_{KEM}$[au]	$E_{HF}-E_{KEM}$ [kcal/mol]
A	315	3	−7381.8557	−7381.8557	0.0000	0.0047
B	315	3	−7382.1621	−7382.1621	0.0000	0.0260
C	315	3	−7382.8332	−7382.8330	−0.0002	−0.1027
triple helix	945	9	−22146.9171	−22146.9112	−0.0059	−3.7332

atomic coordinates are readily available in the Protein Data Bank. The atomic coordinates are the starting information from which the KEM proceeds. From the structural role that collagen plays in the animal body, it is clear that it must be a stable molecule, with the chains of the triple helix structure adhering to one another. We apply the KEM to the molecular structure, 1A89, to see if the approximation is sufficiently accurate to reveal the expected adhesion of the collagen triple chains.

V. Results

Figure (1) shows a triple helix of protein chains that make up the collagen molecule under study. Also shown is the primary structure of the 3 identical protein chains that make up the triple helix, and each protein chain is broken into 3 kernels. The total triplex contains 945 atoms, each chain contains 315 atoms, with kernels 1, 2, and 3 containing 96, 98, and 121 atoms, respectively.

Table 1 contains the KEM calculations for each of the protein chains considered as a single entity. All calculations of this paper are of quantum mechanical Hartree Fock type, using an STO-3G limited basis of atomic orbitals. An "exact" result refers to the Hartree Fock calculation of an entire molecule, including all of its atoms together, without use of the kernel approximation. The KEM calculated energies are meant to approximate the "exact" results. The difference between the two types of calculation are listed in both [au] and [kcal/mol]. One may conclude that the KEM calculation well represents the "exact" result. The percentage difference between the two types calculation is small. For the single chains A, B, and C the percentage differences are 1.0×10^{-7}%, 5.6×10^{-7}%, and 2.2×10^{-6}%, respectively. Notice also that the percentage difference for the entire triple helix is only 2.7×10^{-5}%. This level of accuracy accords with our previous experiences.[1−6]

In Table 2 we list the calculation results for the triplex protein chains considered in pairs. The rows and columns are arranged as in Table 1, except that a new quantity, the interaction energy between the chains of the pairs, is also listed. As before the accuracy of the KEM energies is as expected, with differences for pairs AB, AC, and BC of approximately 2.6×10^{-5}%, 2.2×10^{-5}%, and 2.8×10^{-5}%, respectively. Notice especially that not only do we obtain the chain pair interaction energies but also, as expected, the interaction is attractive.

In Table 3 we list the calculation results for the full triple helix of the collagen structure. As indicated above, the KEM result for the total energy is accurate. The HF and KEM interaction energies of the triple helix are also listed.

Table 2. Interaction Energy Calculations[a] of Chain Pairs by HF/STO-3G

chains	atoms/ kernels	E_{HF} [au]	E_{KEM} [au]	I_{HF} [kcal/mol]	I_{KEM} [kcal/mol]	$I_{HF}-I_{KEM}$ [kcal/mol]
AB	630/6	−14764.0547	−14764.0508	−23.1488	−20.7075	−2.4413
AC	630/6	−14764.7100	−14764.7067	−13.2896	−11.2950	−1.9946
BC	630/6	−14765.0091	−14765.0050	−8.6151	−6.2123	−2.4028

[a] Interaction energies, $I_{ab} = E_{ab} - E_a - E_b$.

Table 3. Interaction Energy Calculations[a] of Collagen Triple Helix (945 Atoms and 9 Kernels) by HF/STO-3G

$E_{HF(abc)}$ [au]	$E_{HF(a+b+c)}$ [au]	$E_{KEM(abc)}$ [au]	$E_{KEM(a+b+c)}$[au]	I_{HF} [kcal/mol]	I_{KEM} [kcal/mol]	$I_{HF}-I_{KEM}$ [kcal/mol]
−22146.9171	−22146.8510	−22146.9112	−22146.8508	−41.4778	−37.9010	−3.5768

[a] Interaction energies, $I_{abc} = E_{abc} - E_{a+b+c}$, $E_{a+b+c} = E_a + E_b + E_c$.

VI. Discussion and Conclusions

A limited basis (STO-3G) was chosen simply to make the energy calculations as convenient as possible, for a protein structure of this size. Previous numerical experience has shown that the KEM can be applied to a wide variety of molecules with good accuracy, and such expectations were realized in this instance.

We have shown how to begin with a known molecular structure and obtain there from quantum mechanical information not otherwise known from the structure alone. With collagen, such information includes the energy of the individual protein chains and their combinations in pairs and as a triplex. Importantly, the interaction energy between chains of a pair or among those of a triplex are well represented, by the KEM. Notably, the KEM approximation is sufficiently accurate to reveal the expected adhesion which must prevail among the collagen triple chains. This forms the basis of an understanding of the structure of collagen in particular but more generally of a rational design of protein chain interactions.

The advantageous contribution which derives from the KEM is the interaction energy between and among protein chains when the molecular structure might contain tens of thousands of atoms. In such a case, if an ab initio quantum mechanical description of the interaction is to be obtained, then an approximation such as that of the KEM is indicated. Such calculations have typically been computationally impractical. The use of the KEM alleviates much of the computational difficulty by dividing a system into kernels, each smaller than the whole. Computations with each of the kernels can be assigned individually to separate nodes of a parallel processor. Thus, two advantages accrue to the KEM, since calculations are smaller and may be computed in parallel. The entire molecular structure is reconstituted from a sum over kernels. What has been shown by the calculations of this paper is that the KEM may be applied for purposes of obtaining the interaction energy between protein chains for understanding of known molecular structures and rational design of proposed structures of considerable size.

Acknowledgment. The collagen structure, 1A89, used in this article has been taken from the Protein Data Bank (PDB). The research reported in this article was supported by the Office of Naval Research. One of us (L.M.) wishes to thank the U.S. Navy Summer Faculty Research Program administered by the American Society of Engineering Education for the opportunity to spend summers at NRL. L.M. thanks NIH for grants (NIGMS MBRS SCORE5 S06GM606654 and RR-03037 the National Center For Research Resources) and NSF for CREST grant support.

References

(1) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method Illustrated with Peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808.

(2) Huang, L.; Massa, L.; Karle, J. The Kernel Energy Method: Basis Fuctions and Quantum Methods. *Int. J. Quantum Chem.* **2006**, *106*, 447.

(3) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method: Application to Insulin. *PNAS* **2005**, *102*, 12690.

(4) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method: Application to DNA. *Biochemistry* **2005**, *44*, 16747.

(5) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method: Application to a tRNA. *PNAS* **2006**, *103*, 1233.

(6) Huang, L.; Massa, L.; Karle, J. Drug-Target Interaction Energies by the Kernel Energy Method: Aminoglycoside Drugs and Ribosomal A-site RNA Target. *PNAS* **2007**, *104*, 4261.

(7) Massa, L.; Huang, L.; Karle, J. Quantum Crystallography and the Use of Kernel Projector Matrices. *Int. J. Quantum Chem. Quantum Chem. Symp*. **1995**, *29*, 371.

(8) Huang, L.; Massa, L.; Karle, J. Kernel Projector Matrices for Leu[1]- Zervamicin. *Int. J. Quantum Chem. Quantum Chem. Symp.* **1996**, *30*, 1691.

(9) Huang, L.; Massa, L.; Karle, J. Kernel Projector Matrices: Application to Leu[1]- Zervamicin. *Encyclopedia of Computational Chemistry*, 1st ed.; von Schleyer, P., Ed.; John Wiley & Sons: New York, 1998; pp 1457−1470.

(10) Karle, J.; Huang, L.; Massa, L. Quantum Crystallography, a Technique for Extending the Concept of Structure. *J. Pure Appl. Chem.* **1998**, *70*, 319.

(11) Huang, L.; Massa, L.; Karle, L. Quantum Crystallography Applied to Crystalline Maleic Anhydride. *Int. J. Quantum Chem.* **1999**, *73*, 439.

(12) Karle, J.; Huang, L.; Massa, L. Quantum Crystallography: Features and Application, in Current Challenges on Large Supramolecular Assemblies. *NATO Sci. Ser., Ser. C* **1999**, *519*, 1.

(13) Huang, L.; Massa, L.; Karle, J. Quantum Crystallography, a Developing Area of Computational Chemistry extending to Macromolecules. *IBM J. Res. Dev.* **2001**, *45*, 409.

(14) Karle, J.; Huang, L. The glue that holds crystals together: A review. *J. Mol. Struct.* **2003**, *746*, 9.

(15) Delacoux, F.; Fichard, A.; Geourjon, C.; Garrone, R.; Ruggiero, F. Molecular Features of the Collagen V Heparin Binding Site Theoretical Model. *J. Biol. Chem.* **1998**, *273*, 15069.

CT7000649