The kernel energy method of quantum mechanical approximation carried to fourth-order terms

Lulu Huang[†], and Lou Massa[‡], and Jerome Karle^{*§}

[†]Laboratory for the Structure of Matter, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20375-5320; and [‡]Departments of Chemistry and Physics, Hunter College and the Graduate School, City University of New York, 695 Park Avenue, New York, NY 10021

Contributed by Jerome Karle, December 3, 2007 (sent for review November 1, 2007)

It is now possible to calculate the *ab initio* quantum mechanics of very large biological molecules. Two things lead to this perspective, namely, (i) the advances of parallel supercomputers, and (ii) the discovery of a quantum formalism called quantum crystallography and the use of quantum kernels, a method that is well suited for parallel computation. The kernel energy method (KEM) carried to second order has been used to calculate the quantum mechanical *ab initio* molecular energy of peptides, protein (insulin and collagen), DNA, and RNA and the interaction of drugs with their biochemical molecular targets. The results were found to have good accuracy. In this article, the accuracy of the KEM is investigated up to an approximation including fourth-order interactions among kernels. Remarkable accuracy is achieved in the calculation of the energy of the ground state of the important biological molecule Leu¹-zervamicin, whose crystal structure is known and used in the calculations.

molecular energy | n-body cluster | quantum crystallography

The kernel energy method (KEM) combines structural crystallographic information with quantum mechanical theory. KEM may be described as the determination of the quantum mechanical molecular energy by the use of the parts of a whole molecule, which are referred to as kernels. Because the kernels are much smaller than a full biological molecule, the calculations of kernels and multiple kernels are practicable. Subsequently, kernel contributions are summed in a manner affording an estimate of the energy for the whole molecule. Thus, the task of obtaining a quantum mechanical energy has been simplified for biological molecules as large as those containing thousands of atoms. The computational time is much reduced by employing the KEM, and the accuracy obtained appears to be quite satisfactory, as shown in previous work.

The first applications of the KEM (1) referred to above, involved a large number of peptides of various shapes and sizes, giving good accuracy, which was retained throughout a wide range of basis functions and computational methods (2). The KEM has been applied to the protein insulin (3), various A, B, and Z DNA (4), RNA (5), drug-target and ribosomal A-site RNA (6), and a collagen triple helix (7) with favorable results. The overall theoretical background for the application of quantum mechanics with crystallography may be found in refs. 8–15. References that review the quantum mechanical methods related to the KEM may be found in ref. 10.

This article is devoted to calculating the KEM energy to an order of approximation including terms up to a fourth order of interaction among the kernels. Given that extending the accuracy of the KEM is a subject of interest, the problem that presents itself is the derivation of the approximation to higher orders and its application, to test whether an expected improvement in accuracy results from computation. That is the problem addressed here.

To review, in the KEM, the use of atomic coordinates from x-ray crystallography leads to a reduction of computational effort and an extraction of quantum information from the crystallography. Central to the KEM is the concept of the kernel.

Kernels are the quantum pieces into which the full molecule is mathematically broken. All quantum calculations are carried out on kernels and multiple kernels. Because the kernels are chosen to be smaller than a full biological molecule, the calculations are accomplished efficiently, and the computational time is much reduced. Subsequently, the properties of the full molecule are reconstructed from those of the kernels and multiple kernels. Thus, a quantum realization of the aphorism that the whole is the sum of its parts is obtained.

It is assumed that the crystal structure is known for a molecule under study. With known atomic coordinates, the molecule is mathematically broken into tractable pieces called kernels. The kernels are chosen such that each atom occurs in only one kernel. Schematically defined kernels, double, triple, and quadruple kernels, are shown in supporting information (SI) Fig. 1, and only these objects are used for all quantum calculations. The total molecular energy is reconstructed therefrom, by summation over the contributions of the kernels and multiple kernels up to the highest order of interaction to be imposed. In our previous work, only an approximation including up to double kernels had been invoked, and in that case, the total energy is calculated according to the formula $E_n^{\text{total}} = \sum_{1 \leq i < j \leq n} E_{ij} - (n - 2) \sum_{1 \leq i \leq n} E_i$.

In this article, we extend the KEM to higher orders of approximation. The purpose of this, of course, is to increase the accuracy of the KEM calculations. We anticipate that this higher order of accuracy can be applied to obtain the energy of enormous biological molecules when it is not computationally feasible otherwise to treat the entire molecule as a whole because of its size. When a structure of interest has known crystallographic coordinates, one may easily define kernels, which altogether represent the entire composite molecule. The use of single kernels and multiple kernels is an approximation that is made to obtain a simplification in the quantum calculation. The validity of this approximation, up to the use of double kernels, for a variety of large biological molecules has been mentioned above. In this article, we derive an extension of the KEM approximation, to higher orders of interaction, and show how increased accuracy may be obtained in its application. Remarkable accuracy, as we indicate below, can be achieved.

Results for the Total Molecular Energy Expanded in Contributions of Kernels and Their Interactions

Molecular Energy as a Sum Over Kernel Energies. The formulas for invoking the KEM up to orders of approximation including double, triple, and quadruple energies are displayed as Eqs. 1, 2, and 3, respectively.

Author contributions: L.H., L.M., and J.K. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

[§]To whom correspondence should be addressed. E-mail: jerome.karle@nrl.navy.mil.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711297105/DC1.

© 2008 by The National Academy of Sciences of the USA

$$E_n^{\text{total}} = \sum_{1 \leq i < j \leq n} E_{ij} - (n-2) \sum_{1 \leq i \leq n} E_i. \quad [1]$$

$$E_n^{\text{total}} = \sum_{1 \leq i < j < k \leq n} E_{ijk} - (n-3) \sum_{1 \leq i < j \leq n} E_{ij} + \left(\sum_{1 \leq i \leq n-3} i \right) \cdot \left(\sum_{1 \leq i \leq n} E_i \right). \quad [2]$$

$$E_n^{\text{total}} = \left(\sum_{1 \leq i < j < k < l \leq n} E_{ijkl} \right) - (n-4) \left(\sum_{1 \leq i < j < k \leq n} E_{ijk} \right) + \left(\sum_{1 \leq i \leq n-4} i \right) \left(\sum_{1 \leq i < j \leq n} E_{ij} \right) - \left(\sum_{1 \leq j \leq n-4} \sum_{1 \leq i \leq j} i \right) \cdot \left(\sum_{1 \leq i \leq n} E_i \right). \quad [3]$$

KEM Molecular Energy Approximation, Obtained from the Kernel Interaction Energies, up to Fourth Order. Using the notion of interaction energies, we may decompose the total energy of a molecule composed of n kernels into its single, double, triple, and quadruple kernel contributions as follows (16–18):

$$E_n^{\text{total}} = \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} \Delta E_{ij} + \sum_{1 \leq i < j < k \leq n} \Delta E_{ijk} + \sum_{1 \leq i < j < k < l \leq n} \Delta E_{ijkl}.$$

The various terms ΔE in this last equation are the interaction energies between and among kernels. We will define each of these interaction energies explicitly as they are used in the derivations below.

Double-kernel interactions. The formula for the total energy, calculated up to only an inclusion of the interactions between double kernels is

$$E_n^{\text{total}} = \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} \Delta E_{ij}, \quad [1a]$$

where E_i is a single-kernel energy and ΔE_{ij} is the interaction energy between two single kernels (i and j) defined as $\Delta E_{ij} = E_{ij} - (E_i + E_j)$.

By using this definition of the double-kernel interaction energy, the total energy becomes

$$E_n^{\text{total}} = \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} [E_{ij} - (E_i + E_j)]. \quad [1b]$$

In the above expression, and in all of the derivations below, we shall find that collecting coefficients of similar terms is simplified by using average quantities. We emphasize that such simplification is the purpose of introducing average quantities. We now define the average quantities associated with Eq. 1b. The single-kernel energies $E_1, E_2, E_3, \dots, E_n$ have an average energy that we call E_i^* , defined as $E_i^* = \sum_{1 \leq i \leq n} E_i / n$. The double-kernel energies, $E_{12}, E_{13}, E_{14}, \dots, E_{(n-1)n}$ will also have an average energy value that we call E_{ij}^* . The number of double kernels is given by the binomial coefficient

$$\binom{n}{2} = \frac{n!}{(n-2)!2!} = \frac{n(n-1)}{2}.$$

This means the average of double-kernel energies is given by the expression

$$E_{ij}^* = \frac{\sum_{1 \leq i < j \leq n} E_{ij}}{\frac{n(n-1)}{2}},$$

and similarly the average of the double-kernel interaction energies is

$$\Delta E_{ij}^* = \frac{\sum_{1 \leq i < j \leq n} \Delta E_{ij}}{\frac{n(n-1)}{2}}.$$

This last average can also be written as $\Delta E_{ij}^* = E_{ij}^* - (E_i^* + E_j^*)$. Summing the double-kernel interaction energies, we have

$$\sum_{1 \leq i < j \leq n} \Delta E_{ij} = \sum_{1 \leq i < j \leq n} [E_{ij} - (E_i + E_j)] = \frac{n(n-1)}{2} [E_{ij}^* - (E_i^* + E_j^*)].$$

By using average values, Eq. 1b becomes

$$E_n^{\text{total}} = nE_i^* + \frac{n(n-1)}{2} [E_{ij}^* - (E_i^* + E_j^*)] = nE_i^* + \frac{n(n-1)}{2} E_{ij}^* - \frac{n(n-1)}{2} (E_i^* + E_j^*). \quad [1c]$$

Noting that E_i^* is the average energy of $E_i (i = 1 \dots n)$, and E_j^* is the average energy of $E_j (j = 1 \dots n)$, so that $E_i^* = E_j^*$ and $E_i^* + E_j^* = 2E_i^*$, the total energy becomes

$$E_n^{\text{total}} = \frac{n(n-1)}{2} E_{ij}^* - (n-2)nE_i^*. \quad [1d]$$

Reverting back from average values, to their expressions over sums of energies, we have the desired result (see Eq. 1).

Triple-kernel interactions. The formula for the total energy, calculated up to only an inclusion of the interactions among triple kernels is

$$E_n^{\text{total}} = \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} \Delta E_{ij} + \sum_{1 \leq i < j < k \leq n} \Delta E_{ijk}, \quad [2a]$$

where, as before, E_i is the energy of a single kernel, ΔE_{ij} is the interaction energy between two single kernels, and ΔE_{ijk} is the interaction energy among three single kernels. The interaction energy for a given triple kernel is obtained from its total energy by subtracting the energy of its single kernels and all of its double-kernel interaction energies. What remains after subtraction is the triple-kernel interaction energy, $\Delta E_{ijk} = (E_{ijk} - E_i - E_j - E_k) - \Delta E_{ij} - \Delta E_{ik} - \Delta E_{jk}$. Using this in Eq. 2a, we have for the total energy of the full molecule,

$$E_n^{\text{total}} = \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} (E_{ij} - E_i - E_j) + \sum_{1 \leq i < j < k \leq n} \{(E_{ijk} - E_i - E_j - E_k) - \Delta E_{ij} - \Delta E_{ik} - \Delta E_{jk}\}$$

$$\begin{aligned}
& - [E_{ij} - (E_i + E_j)] - [E_{ik} - (E_i + E_k)] \\
& - [E_{jk} - (E_j + E_k)] \}. \quad [2b]
\end{aligned}$$

As before, by using the definition of the average quantities E_i^* , E_{ij}^* , ΔE_{ij}^* , and $\Delta E_{ij}^* = E_{ij}^* - (E_i^* + E_j^*)$, so too here, it will be useful to define the analogous generalizations associated with the triple kernels. The number of triple kernels E_{123} , E_{124} , E_{134} , ..., $E_{(n-2)(n-1)n}$, is given by the binomial coefficient

$$\binom{n}{3} = \frac{n!}{(n-3)!3!} = \frac{n(n-1)(n-2)}{6},$$

and the average of the triple-kernel energies and interaction energies is therefore

$$E_{ijk}^* = \frac{\sum_{1 \leq i < j < k \leq n} E_{ijk}}{\frac{n(n-1)(n-2)}{6}}, \quad \text{and} \quad \Delta E_{ijk}^* = \frac{\sum_{1 \leq i < j < k \leq n} \Delta E_{ijk}}{\frac{n(n-1)(n-2)}{6}},$$

respectively. Also, $\Delta E_{ijk}^* = (E_{ijk}^* - E_i^* - E_j^* - E_k^*) - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] - [E_{jk}^* - (E_j^* + E_k^*)]$.

Now, by using average quantities, the contribution of the triple-kernel interaction energies to the total molecular energy becomes

$$\begin{aligned}
\sum_{1 \leq i < j < k \leq n} \Delta E_{ijk} &= \frac{n(n-1)(n-2)}{6} \{ (E_{ijk}^* - E_i^* - E_j^* - E_k^*) \\
&\quad - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] \\
&\quad - [E_{jk}^* - (E_j^* + E_k^*)] \}.
\end{aligned}$$

Altogether, we have for the total molecular energy, expressed entirely in the average expressions for single-, double-, and triple-kernel contributions, that Eq. 2b becomes

$$\begin{aligned}
E_n^{\text{total}} &= nE_i^* + \frac{n(n-1)}{2} [E_{ij}^* - (E_i^* + E_j^*)] \\
&\quad + \frac{n(n-1)(n-2)}{6} \{ (E_{ijk}^* - E_i^* - E_j^* - E_k^*) \\
&\quad - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] \\
&\quad - [E_{jk}^* - (E_j^* + E_k^*)] \}. \quad [2c]
\end{aligned}$$

Using $E_i^* = E_j^* = E_k^*$ and $E_{ij}^* = E_{ik}^* = E_{jk}^*$, we have

$$\begin{aligned}
E_n^{\text{total}} &= nE_i^* + \frac{n(n-1)}{2} (E_{ij}^* - 2E_i^*) \\
&\quad + \frac{n(n-1)(n-2)}{6} (E_{ijk}^* - 3E_{ij}^* + 3E_i^*). \quad [2d]
\end{aligned}$$

Collecting coefficients of similar terms, gives

$$\begin{aligned}
E_n^{\text{total}} &= \frac{n(n-1)(n-2)}{6} E_{ijk}^* - \frac{n(n-1)}{2} (n-3)E_{ij}^* \\
&\quad + \frac{n^2 - 5n + 6}{2} nE_i^*. \quad [2e]
\end{aligned}$$

Now, reverting back from the average quantities in this last equation to the equivalent expressions in sums over individual energies that define the averages, we obtain

$$\begin{aligned}
E_n^{\text{total}} &= \sum_{1 \leq i < j < k \leq n} E_{ijk} - (n-3) \sum_{1 \leq i < j \leq n} E_{ij} \\
&\quad + \frac{(n-2)(n-3)}{2} \sum_{1 \leq i \leq n} E_i. \quad [2f]
\end{aligned}$$

The reader may be convinced that $[(n-2)(n-3)]/2 = \sum_{1 \leq i \leq n-3} i$ by evaluating the first few terms of the above equation (say for $n = 4, 5, 6, 7, 8, \dots$). Thus, we obtain the desired result (see Eq. 2).

Quadruple-kernel interactions. The formula for the total energy, calculated up to only an inclusion of the interactions among quadruple kernels is

$$\begin{aligned}
E_n^{\text{total}} &= \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} \Delta E_{ij} + \sum_{1 \leq i < j < k \leq n} \Delta E_{ijk} \\
&\quad + \sum_{1 \leq i < j < k < l \leq n} \Delta E_{ijkl}, \quad [3a]
\end{aligned}$$

where in the equation above the various symbols retain their previously used meanings, and ΔE_{ijkl} is the interaction energy among four single kernels (i, j, k, l). ΔE_{ijkl} is defined as the total energy of the quartet of kernels after subtracting the energy of its single kernels, and the interaction energies of its double kernels and its triple kernels. The quadruple interaction energy is written as $\Delta E_{ijkl} = E_{ijkl} - E_i - E_j - E_k - E_l - \Delta E_{ij} - \Delta E_{ik} - \Delta E_{il} - \Delta E_{jk} - \Delta E_{jl} - \Delta E_{kl} - \Delta E_{ijk} - \Delta E_{ijl} - \Delta E_{ikl} - \Delta E_{jkl}$. The contribution of such terms to the total molecular energy would be obtained by evaluation of their sum $\sum_{1 \leq i < j < k < l \leq n} \Delta E_{ijkl} = \sum_{1 \leq i < j < k < l \leq n} \{ E_{ijkl} - E_i - E_j - E_k - E_l - [E_{ij} - (E_i + E_j)] - [E_{ik} - (E_i + E_k)] - [E_{il} - (E_i + E_l)] - [E_{jk} - (E_j + E_k)] - [E_{jl} - (E_j + E_l)] - [E_{kl} - (E_k + E_l)] - \{ (E_{ijk} - E_i - E_j - E_k) - [E_{ij} - (E_i + E_j)] - [E_{ik} - (E_i + E_k)] - [E_{jk} - (E_j + E_k)] \} - \{ (E_{ijl} - E_i - E_j - E_l) - [E_{ij} - (E_i + E_j)] - [E_{il} - (E_i + E_l)] - [E_{jl} - (E_j + E_l)] \} - \{ (E_{ikl} - E_i - E_k - E_l) - [E_{ik} - (E_i + E_k)] - [E_{il} - (E_i + E_l)] - [E_{kl} - (E_k + E_l)] \} - \{ (E_{jkl} - E_j - E_k - E_l) - [E_{jk} - (E_j + E_k)] - [E_{jl} - (E_j + E_l)] - [E_{kl} - (E_k + E_l)] \} \}$. When this sum is added to the contributions of single, double, and triple kernels, one obtains the rather lengthy expression for the total molecular energy,

$$\begin{aligned}
E_n^{\text{total}} &= \sum_{1 \leq i \leq n} E_i + \sum_{1 \leq i < j \leq n} [E_{ij} - (E_i + E_j)] + \sum_{1 \leq i < j < k \leq n} \\
&\quad \cdot \{ (E_{ijk} - E_i - E_j - E_k) - [E_{ij} - (E_i + E_j)] \\
&\quad - [E_{ik} - (E_i + E_k)] - [E_{jk} - (E_j + E_k)] \} \\
&\quad + \sum_{1 \leq i < j < k < l \leq n} \{ E_{ijkl} - E_i - E_j - E_k - E_l \\
&\quad - [E_{ij} - (E_i + E_j)] - [E_{ik} - (E_i + E_k)] \\
&\quad - [E_{il} - (E_i + E_l)] - [E_{jk} - (E_j + E_k)] \\
&\quad - [E_{jl} - (E_j + E_l)] - [E_{kl} - (E_k + E_l)] \\
&\quad - \{ (E_{ijk} - E_i - E_j - E_k) - [E_{ij} - (E_i + E_j)] \\
&\quad - [E_{ik} - (E_i + E_k)] - [E_{jk} - (E_j + E_k)] \} \\
&\quad - \{ (E_{ijl} - E_i - E_j - E_l) - [E_{ij} - (E_i + E_j)] \\
&\quad - [E_{il} - (E_i + E_l)] - [E_{jl} - (E_j + E_l)] \} \\
&\quad - \{ (E_{ikl} - E_i - E_k - E_l) - [E_{ik} - (E_i + E_k)] \\
&\quad - [E_{il} - (E_i + E_l)] - [E_{kl} - (E_k + E_l)] \} \\
&\quad - \{ (E_{jkl} - E_j - E_k - E_l) - [E_{jk} - (E_j + E_k)] \\
&\quad - [E_{jl} - (E_j + E_l)] - [E_{kl} - (E_k + E_l)] \} \}.
\end{aligned}$$

$$- \{(E_{jkl} - E_j - E_k - E_l) - [E_{jk} - (E_j + E_k)] \\ - [E_{jl} - (E_j + E_l)] - [E_{kl} - (E_k + E_l)]\}. \quad [3b]$$

As before, the use of average energy quantities will allow a simplification of the energy expression above so that the individual contributions of single, double, triple, and quadruple kernels can be gathered together and separated from one another.

We retain the previous meanings of the expressions for averages E_i^* , E_{ij}^* , $\Delta E_{ij}^* = E_{ij}^* - (E_i^* + E_j^*)$, E_{ijk}^* , ΔE_{ijk}^* , and $\Delta E_{ijkl}^* = (E_{ijk}^* - E_i^* - E_j^* - E_k^*) - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] - [E_{jk}^* - (E_j^* + E_k^*)]$. E_{ijkl}^* is the average energy of the quadruple kernels E_{1234} , E_{1245} , E_{1345} , $\dots$, $E_{(n-3)(n-2)(n-1)n}$. The number of such quadruple kernels is given by the binomial coefficient

$$\binom{n}{4} = \frac{n!}{(n-4)!4!} = \frac{n(n-1)(n-2)(n-3)}{24}.$$

Therefore, the average energy for the quadruple kernels is

$$E_{ijkl}^* = \frac{\sum_{1 \leq i < j < k < l \leq n} E_{ijkl}}{\frac{n(n-1)(n-2)(n-3)}{24}}.$$

Now, introducing into Eq. 3b the averages of all kernel and multiple kernel energy terms, we obtain

$$E_n^{\text{total}} = nE_i^* + \frac{n(n-1)}{2} [E_{ij}^* - (E_i^* + E_j^*)] \\ + \frac{n(n-1)(n-2)}{6} \{ (E_{ijk}^* - E_i^* - E_j^* - E_k^*) \\ - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] \\ - [E_{jk}^* - (E_j^* + E_k^*)] \} + \frac{n(n-1)(n-2)(n-3)}{24} \\ \cdot \{ E_{ijkl}^* - E_i^* - E_j^* - E_k^* - E_l^* - [E_{ij}^* - (E_i^* + E_j^*)] \\ - [E_{ik}^* - (E_i^* + E_k^*)] - [E_{il}^* - (E_i^* + E_l^*)] \\ - [E_{jk}^* - (E_j^* + E_k^*)] - [E_{jl}^* - (E_j^* + E_l^*)] \\ - [E_{kl}^* - (E_k^* + E_l^*)] - \{ (E_{ijk}^* - E_i^* - E_j^* - E_k^*) \\ - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{ik}^* - (E_i^* + E_k^*)] \\ - [E_{jk}^* - (E_j^* + E_k^*)] \} - \{ (E_{ijl}^* - E_i^* - E_j^* - E_l^*) \\ - [E_{ij}^* - (E_i^* + E_j^*)] - [E_{il}^* - (E_i^* + E_l^*)] \\ - [E_{jl}^* - (E_j^* + E_l^*)] \} - \{ (E_{ikl}^* - E_i^* - E_k^* - E_l^*) \\ - [E_{ik}^* - (E_i^* + E_k^*)] - [E_{il}^* - (E_i^* + E_l^*)] \\ - [E_{kl}^* - (E_k^* + E_l^*)] \} - \{ (E_{jkl}^* - E_j^* - E_k^* - E_l^*) \\ - [E_{jk}^* - (E_j^* + E_k^*)] - [E_{jl}^* - (E_j^* + E_l^*)] \\ - [E_{kl}^* - (E_k^* + E_l^*)] \} \}. \quad [3c]$$

By using $E_i^* = E_j^* = E_k^* = E_l^*$, $E_{ij}^* = E_{ik}^* = E_{il}^* = E_{jk}^* = E_{jl}^* = E_{kl}^*$, $E_{ijk}^* = E_{ijl}^* = E_{ikl}^* = E_{jkl}^*$, Eq. 3c simplifies to

$$E_n^{\text{total}} = nE_i^* + \frac{n(n-1)}{2} (E_{ij}^* - 2E_i^*)$$

$$+ \frac{n(n-1)(n-2)}{6} (E_{ijk}^* - 3E_{ij}^* + 3E_i^*) \\ + \frac{n(n-1)(n-2)(n-3)}{24} \\ \cdot (E_{ijkl}^* - 4E_{ijk}^* + 6E_{ij}^* - 4E_i^*). \quad [3d]$$

Gathering the coefficients of similar terms, yields the expression

$$E_n^{\text{total}} = \frac{n(n-1)(n-2)(n-3)}{24} E_{ijkl}^* + \frac{n(n-1)(n-2)}{6} \\ \cdot [1 - (n-3)] E_{ijk}^* + \frac{n(n-1)}{2} \frac{(n-3)(n-4)}{2} E_{ij}^* \\ + \frac{-n^3 + 9n^2 - 26n + 24}{6} nE_i^*. \quad [3e]$$

Reverting back from the average energies in the last equation to their expressions as sums over the individual energy terms gives the total molecular energy as

$$E_n^{\text{total}} = \sum_{1 \leq i < j < k < l \leq n} E_{ijkl} - (n-4) \sum_{1 \leq i < j < k \leq n} E_{ijk} \\ + \frac{(n-3)(n-4)}{2} \sum_{1 \leq i < j \leq n} E_{ij} - \frac{n^3 - 9n^2 + 26n - 24}{6} \sum_{1 \leq i \leq n} E_i. \quad [3f]$$

The reader may be convinced that

$$\frac{(n-3)(n-4)}{2} \equiv \sum_{1 \leq i \leq n-4} i, \text{ and} \\ \frac{n^3 - 9n^2 + 26n - 24}{6} \equiv \sum_{1 \leq j \leq n-4} \sum_{1 \leq i \leq j} i$$

by evaluating the first few terms of these equations for various values of n . Using these last two equations in Eq. 3f gives the desired result (see Eq. 3).

Application to Leu¹-zervamicin of the Fourth-Order Approximation of KEM. We test the accuracy achievable with the above formulas by application to the important biological molecule Leu¹-zervamicin, shown in SI Fig. 2. It is an antibiotic that transports potassium ions across cell membranes. Groups of zervamicin molecules assemble to form channels that serve to allow ion passage. The molecule has hydrophobic side chains extending from the peptide residues on one side and polar side chains extending from other peptide residues on the other side. A side chain of particular interest is a side chain of (residue 11) glutamine. This side chain is attached on the hydrophobic side of the peptide, as determined from a crystal-structure investigation by Karle *et al.* (19). The side chain acts as a gate that allows K⁺ ions to proceed through a channel membrane. Glutamine 11, by a swinging motion, acts to open and close the channel each time an ion goes through it. If the structural arrangement of the zervamicin molecules in a cell membrane is similar to that in the crystal, it is possible that the zervamicin crystal structures offer a model of a gating mechanism on the level of atomic resolution. The results of calculation related to Leu¹-zervamicin are listed in Table 1.

The program GAUSSIAN was used to do the calculations,

Table 1. The HF STO-3G energy calculated up to fourth-order approximation of KEM

	Single	Double	Triple	Quadruple
E_{KEM} , au	−5,857.8663	−5,851.5469	−5,851.5686	−5,851.5703
E_{exact} , au	−5,851.5703	−5,851.5703	−5,851.5703	−5,851.5703
ΔE , au	6.2960	−0.0234	−0.0017	0.0000
ΔE per atom, kcal/mol	14.9	-5.5×10^{-2}	-4.0×10^{-3}	3.3×10^{-7}

related to each of the formulas 1, 2, and 3 applied to the Leu¹-zervamicin molecule. The results are listed in Table 1 in order of increasing accuracy, reading across the table, for the double-, triple-, and quadruple-interaction approximations. For each of these approximations, the result listed as the exact energy is the brute force Hartree–Fock (HF) calculation on the full molecule using the limited basis of functions called STO-3G. The table also lists the corresponding energy calculated by means of the KEM equations. Eq. 1, 2, and 3 are used to obtain the results listed for the double-, triple-, and quadruple-kernel approximations, respectively. The differences between the brute-force and the KEM calculations are listed in atomic units (au). Finally the results are listed on the basis of energy difference per atom in units of Kcal/mol. As expected, the difference between the exact and the KEM results decrease as the order of the KEM approximations taken in to account increase from double to triple to quadruple interactions. The numerical differences are −0.0234 au, −0.0017 au, and 0.0000 au for the double, triple, and quadruple interactions, respectively.

Discussion and Conclusions

The KEM has successfully combined x-ray structures with quantum mechanical theory to obtain the molecular energy of a variety of large biological molecules. Because the kernels are much smaller than a full biological molecule, the calculations of kernels and multiple kernels are practicable, and the task of obtaining a quantum mechanical energy has been simplified for molecules containing thousands of atoms. The question arises, what are the limits of accuracy based on the idea of quantum kernels. To answer this question, this article is devoted to calculating the KEM energy to an order of approximation including terms up to a fourth order of interaction among the kernels. In this first set of test calculations, a limited basis set of atomic orbitals and the HF approximation were used. However, as we have indicated previously, the KEM (2) applies to any basis set and to all of the usual quantum mechanical methods based on orbitals.

We have derived the energy equations that allow the generalization of the kernel approximation to higher orders of interaction. Eq. 1 takes into account the double-kernel interactions, and this is the equation we have used for most of the results published heretofore (1, 3–7). The generalizations of this equation to include triple- and quadruple-kernel interactions are new and are given by Eqs. 2 and 3, respectively.

The numerical test of these equations using the molecule Leu¹-zervamicin indicate that remarkable accuracy is achievable. This molecule contains 265 atoms and was broken into seven kernels as indicated in SI Fig. 3. All three of the equations were applied to calculate the energy of the molecule by using these seven kernels. The principal guideline for choosing kernel boundaries is simply to sever only single bonds between amino acids. We consider a kernel to be of convenient and optimal size if it contains ≈ 100 atoms.

The standard of accuracy for these calculations was the brute-force HF calculation of the energy for the full molecule in the same basis. As the results of Table 1 show, the accuracy of the KEM increases with each order of approximation. For

example, by using Table 1 consider the calculation difference on the basis of difference per atom. At the level of double-kernel interactions (Eq. 1) the magnitude of difference is 10^{-2} Kcal/mol per atom, an already small error. If triple-kernel interactions (Eq. 2) are invoked the magnitude of difference is reduced by an order of magnitude to 10^{-3} Kcal/mol per atom. Finally, the use of quadruple interactions (Eq. 3) induces a further reduction in difference by an additional four orders of magnitude to 10^{-7} Kcal/mol per atom. For this molecule at least, and those similar to it, one need not contemplate going beyond the quadruple level of accuracy in the KEM approximation. The results here allow one to conclude that *ab initio* accuracy is obtainable for biological molecules within the KEM, carrying the approximation up to quadruple interactions. For large enough molecules, for which brute force calculations are not feasible, the KEM calculations will still be practicable, because the kernels and multiple kernels can be chosen to be very much smaller than the full molecule.

In previous papers, we have shown that the KEM provides an effective simplification in the task of solving the Schrödinger equation for the case of large biological molecules. The point of this article, then, is to demonstrate that the accuracy of KEM can be vastly increased by incorporating higher-order interactions among the kernels. To make this point, we have calculated the higher-order interactions associated with the molecule Leu¹-zervamicin listed in Table 1. Moreover, because the KEM constructs a whole from the sum of its parts, it is especially suitable for parallel computation.

It is important to mention the computational complexity of the method that has yielded the results of Table 1. In the simplest terms, one may state that the complexity of an HF calculation is proportional to n^4 , where n is the number of atoms (and in a rough way this is proportional to the number of basis functions used). If a system of n atoms is broken into n_K single kernels, then the number of double kernels, triple kernels, and quadruple kernels would be of the order of n_K^2 , n_K^3 , and n_K^4 , respectively, and each individual multiple kernel would require computational effort of the order of $(n/n_K)^4$. Thus, the total computational effort, in the case of using at most double, triple, or quadruple kernels, would be of the order of (n^4/n_K^2) , (n^4/n_K) , and n^4 , respectively. It is clear that a great saving in time would be associated with the use of Eq. 1 and Eq. 2 (which incorporate, at most, double kernels and triple kernels) because of the way a power of n_K enters the denominator in the description of their computational effort. Because Eq. 1 and Eq. 2 are capable of quite good results, as may be seen from Table 1, in most cases there would not be a need to go beyond their level of accuracy. Eq. 3 (which incorporates quadruple kernels) is even more accurate but at the cost of the increased computational effort noted above. Reasons to expend such cost would include the advantage of parallel computation organized by the assignment of the kernels and multiple kernels to individual parallel nodes of computation, the possibility of an organized omission of certain quadruples bound to be negligible based on a criterion of distances among kernels, and obtaining separate interaction energies of second, third, and fourth order. An HF calculation on a full molecule, for example, would not yield the separate interaction energies of each order. Most importantly, the use of