Published in final edited form as:

Theochem. 2010 December ; 962(1-3): 72–79. doi:10.1016/j.theochem.2010.09.017.

The Kernel Energy Method: Construction of 3 & 4 tuple Kernels from a List of Double Kernel Interactions

Lulu Huang¹ and Lou Massa^{2,*}

¹ Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 20375-5341

² Hunter College and the Graduate School, City University of New York, New York, NY 10065

Abstract

The Kernel Energy Method (KEM) provides a way to calculate the *ab-initio* energy of very large biological molecules. The results are accurate, and the computational time reduced. However, by use of a list of double kernel interactions a significant additional reduction of computational effort may be achieved, still retaining *ab-initio* accuracy. A numerical comparison of the indices that name the known double interactions in question, allow one to list higher order interactions having the property of topological continuity within the full molecule of interest. When, that list of interactions is unpacked, as a kernel expansion, which weights the relative importance of each kernel in an expression for the total molecular energy, high accuracy, and a further significant reduction in computational effort results. A KEM molecular energy calculation based upon the HF/STO3G chemical model, is applied to the protein insulin, as an illustration.

Keywords

Quantum Calculation; KEM; Biological Molecules; Interaction Energies; Insulin

INTRODUCTION

It is of current interest to apply the Schrodinger equation to the description of ever larger molecules, especially within the realm of biological systems. The *ab-initio* calculations have a numerical difficulty which famously grows with a high power of the number of atoms in the system. In recent years, to overcome the growth of such computational difficulty, many groups have hit upon the idea of using a variety of methods, which fragment a whole molecule, and calculations are carried out upon the fragments. Subsequently the properties of the fragments are summed, in a manner to deliver the properties of the whole molecule. Of course fragmentation reduces considerably the computational cost of calculation, inasmuch as fragments are chosen to be much smaller than the full molecule. The attraction of such fragment methods is that often near linear scaling of the computational difficulty can be achieved together with good quantum mechanical accuracy. The literature of fragment

© 2010 Elsevier B.V. All rights reserved.

* Corresponding author: lmassa@hunter.cuny.edu Hunter College, 695 Park Ave. New York, NY 10065 (212) 772-4016 (office), (212) 772-5332 (fax) .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

methods is quite large. We mention here the work of Yang [1-3] divide & conquer, Exner and Mezy [4,5], adjustable density matrix assembler (**ADMA**), Kitaura [6-8], fragment molecular orbital (**FMO**), Imamura [9-10] the elongation method, Gadre et al (**MTA**) [11-13] Das et al [14] *ab-initio* fragment orbital based theory, Zhang et al [15-17] Molecular fractionation with conjugated cap (**MFCC**), Bettens & Lee [18] Molecular fragmentation, Deev and Collins [19,20] systematic molecular fragmentation, Salahub et al [21] multilevel fragment-based approach (**MFBA**), as well as our own kernel energy method (**KEM**) [22-30]. We have reviewed elsewhere the relationship of other methods to our work in KEM [31,32].

It is possible to calculate the *ab-initio* quantum mechanics of very large biological molecules. Two things are helpful viz., (1.) advances in parallel supercomputers, and (2.) a quantum formalism called quantum crystallography & the use of quantum kernels which are perfectly adapted to parallel computation. The Kernel Energy Method (KEM), has been used to calculate the quantum mechanical *ab-initio* molecular energy of peptides, proteins (insulin and collagen), DNA, RNA, and the interaction of drugs with their biochemical molecular targets. The computational time is reduced by employing the KEM, and the accuracy obtained is satisfactory, as shown in our previous work.

In the KEM the use of atomic coordinates from X-ray crystallography leads to a reduction of computational effort. Central to the KEM is the concept of the kernels. These are the quantum pieces into which the full molecule is mathematically broken. The resulting dangling bonds are saturated with hydrogen caps. All quantum calculations are carried out on kernels & multiple kernels. (Often below, the word “kernels” is used to mean all kernels, whether they be single or multiple in nature. If no confusion is likely to occur in the context of the usage this is viewed as a more economic and easier flowing manner of speech.) Because the kernels are chosen to be smaller than a full biological molecule, the computational time is reduced. Subsequently, the properties of the full molecule are reconstructed from those of the kernels. Thus a quantum realization of the aphorism that the whole is the sum of its parts is obtained.

With known atomic coordinates, the molecule is mathematically broken into tractable pieces called kernels. The total molecular energy is reconstructed by summing the kernels and multiple-kernels up to the highest order of interaction invoked. The effect of the hydrogen caps tends to vanish in the summation over kernels. The formulas for invoking the KEM, for a molecule divided into n single kernels (S), up to orders of approximation including double (D), triple (T), and quadruple (Q) kernel interactions are displayed as equations 1, 2, and 3, respectively.

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

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

$$E_n^{\text{total}} = \left(\sum_{i=1}^{n-3} \sum_{i \langle j \langle k \langle l} E_{ijkl} \right) - (n-4) \left(\sum_{i=1}^{n-2} \sum_{i \langle j \langle k} E_{ijk} \right) + \left(\sum_{i=1}^{n-4} i \right) \left(\sum_{i=1}^{n-1} \sum_{i \langle j} E_{ij} \right) - \left(\sum_{j=1}^{n-4} \sum_{i=1}^j i \right) \left(\sum_{i=1}^n E_i \right) \quad (3)$$

These equations represent a large simplification in the problem of solving the *ab-initio* Schrodinger equation of the full molecule [29]. The point of this paper is to consider how these three equations are themselves simplified further, when only a subset of the kernel interactions, are considered [32]. Our considerations begin with knowledge of a set of single kernels which define the full molecule, and in addition a set of double kernel interactions among them, representing in the sense of perturbation theory, the principal additional components of the energy which affect the single kernels. From this, a simple algebraic rule allows construction of the 3-tuple and 4-tuple interactions which form additional perturbations to the total molecular energy. Having constructed the complete sub-list of kernel tuples of order 4,3,2,& 1 each of their contributions to the total molecular energy is summed together. The identification of the weighting coefficients, which determine the relative energy contributions of each kernel tuple is a result described in this paper. The names of the double kernel interactions, which are assumed known at the outset, are either postulated on theoretical grounds [33], or else are obtained experimentally from the crystal structure using inter-particle distance considerations [34], or the bond path information inherent in the experimental electron density [35,36].

One must recognize in what follows that higher order kernels are associated with lower order kernels. For example, the 4 indices, which define a particular Q give rise to 4 T, 6D, and 4 S combinations of kernels. Similarly, the 3 indices which define a particular T, give rise to 3 D, and 3 S combinations of kernels. And, the 2 indices, which define a particular D give rise to 2 S kernels. One may convince oneself of these associations of higher order kernels with combinations of lower order kernels by referring to the kernel energy expansion in terms of interaction energies [29], or simply by careful examination of Eq.s 1, 2, 3, above. There it may be seen in Eq. 1, that for every D kernel defined by a particular choice of indices i, j, there always must occur two S kernels of names i and j. An analogous examination of Eq.s 2 and 3 reveals the numbers of lower order kernels associated with Q and T kernels, as stated above.

RESULTS

1. Rule for creating 3-and 4 tuple kernels from given 1 and 2 tuple kernels

In Eq.s 1-3 above the total number of single kernels, n, determines the total number of interactions possible among the kernel tuples, and they are all included in the summations

indicated. So the total number of double kernels is $\frac{n(n-1)}{2}$ in Eq. (1), the total number of

triple kernels is $\frac{n(n-1)(n-2)}{6}$ in Eq. (2), and the total number of quadruple kernels is

$\frac{n(n-1)(n-2)(n-3)}{24}$ in Eq. (3). However suppose given all single kernels, S, a subset of interactions, D, between double kernels is most important, the others being ignorable. From these S and D one may construct all the corresponding 4-tuple interactions Q, and 3-tuple interactions T, using a simple numerical rule, discussed next.

1.1. 2 and 2 makes 3—Rule: For constructing the set of 3-tuple interactions T from the 2-tuples D.

Create a list of the 2-tuple interactions arranged according to ascending numerical order. Compare each 2-tuple to all items in the list. In the event the comparison 2-tuple indices have only one index in common with that of the list item, the 3 different indices, among the total of four, are retained as indicating a 3-tuple.

Sample:

In Fig. 1 we see that double kernel 12 + double kernel 13 create triple kernel 123, 12 + 14 create 124, 12 + 23 also create 123, and 12 + 25 create 125, so we obtain 3 unique triple kernels 123, 124, 125.

1.2. 2 and 3 makes 4—Rule: For constructing the set of 4-tuple interactions Q from the 2-tuples D, and 3-tuples T.

Create a list of the 2-tuple and 3-tuple interactions arranged according to ascending numerical order. Compare each 2-tuple to all items in the 3-tuple list. In the event that 2-tuple indices have only one index in common with a 3-tuple, the 4 different indices, among the total of five, are retained as indicating a 4-tuple.

Sample:

In Fig. 2 we see that double kernel 12 + triple kernel 134 create quadruple kernel 1234, 12 + 147 create 1247, 12 + 234 also create 1234, and 12 + 235 create 1235, so we obtain 3 unique quadruple kernels 1234, 1235, 1247. In creating the list of 3-tuple and 4-tuple interactions, a given tuple can be created in multiple comparisons as in the rules above, but a unique index is retained once only in the final list of newly created 3-tuples and 4-tuples.

Application of the above rules, involves only a comparison of the numerical indices which define the kernel energy interactions, and results in a complete list of all of the topologically continuous interactions. The number of these interactions is very much smaller than the number of all possible interactions, as considered in Eq.s 1, 2, and 3. Moreover it is reasonable to assume that the topologically continuous interactions of each order are those which are most important.

Using the crystal structure of insulin, in a later section we will give an explicit example of creation of a list of higher order interactions based upon double kernels of the molecule, and then show how they can be used to calculate the total energy of the molecule. The energy calculation yields results of high *ab-initio* accuracy, increasingly so with each higher order of approximation added.

2. KEM formulas for a subset of the total kernel interactions

The modifications of the equations (1-3) above, which result from retaining fewer than all possible multiple kernels in the representation of the molecular interactions are as follows:

$$E^{2^{nd}} = \sum E_D - \sum C_s^{(2)} E_s \quad (4)$$

$$E^{3^{rd}} = \sum E_T - \sum C_D^{(3)} E_D + \sum C_s^{(3)} E_s \quad (5)$$

$$E^{4^{\text{th}}} = \sum E_Q - \sum C_T^{(4)} E_T + \sum C_D^{(4)} E_D - \sum C_S^{(4)} E_S \quad (6)$$

where superscripts 2nd, 3rd and 4th mean respectively second, third, and fourth order of KEM approximation. The summations are over all elements in the lists of single (S), double (D), triple (T) and quadruple (Q) kernels. The C_S , C_D , C_T in the above equations, are each lists of coefficients. The subscripts indicate a type of kernel (S, single; D, double; T, triple, and Q, quadruple), and the superscripts indicate the order of KEM approximation (2, second order; 3, third order; and 4, fourth order). Elements of the coefficient lists weight the relative importance of individual kernel energies in the lists E_S , E_D , E_T and E_Q . The element coefficients are obtained by a count of the number of times a specific kernel appears in the list of all interaction kernels, in a certain way that we now discuss.

Equations 1,2,3 are a special case of equations 4,5,6, since equations 1, 2, 3 are only used for counting in all of the possible kernel interactions, and equations 4, 5, 6 are not only used for that but also for counting when only the most important kernel interactions are considered. When only a subset of kernel interactions are considered, the number of terms in equations 1,2,3 are reduced, but the overall form of the equations, including the signs (+ and -) of the various terms remains the same. Although the coefficients which weight the relative importance of every kernel in equations 1,2,3 are determined by n, the total number of single kernels, this will no longer hold when some of the interactions are neglected. In the latter case the coefficients are determined by counting the number of times each kernel appears in the equations, for each of the orders of approximation 2nd, 3rd, 4th. In following sections we take these up one at a time.

Attention must be called to the pattern of alternating signs (+ and -) of successive terms in each of these three equations 4,5,6. The coefficient of each kernel in these equations is simply the number of times that kernel appears in each term of the equation, care being taken to include the mathematical sign of each term in the overall summation. For example, in Eq. 6, single kernel contributions S arise from all kernels in the lists Q,T,D,S. One must count the number of times a single kernel arises in each list and sum them all together, as indicated in the equation, using the appropriate sign for the single kernel from each list. Table 1 displays the mathematical signs associated with each kernel type arising from another kernel of a given type. Thus, as table 1 shows, single kernels arising from the lists S,D,T,Q have respectively the mathematical signs +, -, +, -. Similarly, double kernels arising from the lists D, T, Q have respectively the mathematical signs +, -, +, and so forth. Also, it is helpful to recognize that the count of a given kernel type at a higher level of approximation can be summed together with counts from a lower level of approximation, again care being taken to attend to the correct signs of the terms from each level of approximation. Suppose for example, the coefficient of a single kernel S is known in 3rd approximation. Suppose the number of occurrences, for that single kernel, in the list of quadruples Q, is counted. Then, the coefficient of the single kernel in 4th order is just the sum of its count from the list Q minus its coefficient from 3rd order (i.e., minus, because single kernels from Q and T have opposite signs, as indicated in table 1).

2.1. Coefficients of equation 4—In equation 4 each double kernel to be counted with unit coefficient is read directly from the list of double kernel interactions which is given, and which defines the double kernel molecular interactions under consideration. The coefficient of any single kernel energy E_S is determined by counting the number of times it is represented in the double kernel list, and subtracting the number of times it occurs in the single kernel list (which equals 1). Thus,

$$C_s^{(2)} = n_s^{(2)}(D) - C_s^{(1)} \quad (7)$$

where $C_s^{(2)}$, in the total energy expression, is the coefficient of a single kernel S, in KEM second order approximation, $n_s^{(2)}(D)$ is the number of times the single kernel S appears in the set of double kernels D in 2nd order KEM, and $C_s^{(1)}$ is the coefficient of a single kernel in the first order approximation KEM ($C_s^{(1)} \equiv 1$, for each single kernel that appears in the list of kernels which define the full molecule).

2.2. Coefficients of equation 5—In equation 5 each triple kernel to be counted with unit coefficient is read directly from the list of triple kernels created from double kernels using the rule described above. The double kernels are of two types, D_B (bonded) and D_{NB} (non-bonded). The former type D_B , arise from two lists of kernels, T and D, but the latter type D_{NB} arise from only the list T. Thus,

$$C_{D_B}^{(3)} = \{n_{D_B}^{(3)}(T) - C_D^{(2)}\} \quad (8)$$

$$C_{D_{NB}}^{(3)} = \{n_{D_{NB}}^{(3)}(T)\} \quad (9)$$

where $C_D^{(3)}$ in the total energy expression is the coefficient of a double kernel D in KEM 3rd order approximation, $n_{D_B}^{(3)}(T)$ is the number of times the double kernel (bonded) D_B , appears in the set of triple kernels T in 3rd order KEM, and $C_D^{(2)}$ is the coefficient of a double kernel in the 2nd order approximation KEM ($C_D^{(2)} \equiv 1$, for each (bonded) double kernel that appears in the original list of double kernels. The coefficient of any single kernel in equation 5, in like manner, is

$$C_s^{(3)} = \{n_s^{(3)}(T) - C_s^{(2)}\} \quad (10)$$

where $C_s^{(3)}$ in the total energy expression is the coefficient of a single kernel S, in KEM 3rd order approximation, $n_s^{(3)}(T)$ is the number of times the single kernel S, appears in the set of triple kernels T, in 3rd order KEM, and $C_s^{(2)}$ is the coefficient of a single kernel in the 2nd order approximation KEM.

2.3. Coefficients of equation 6—In equation 6 each quadruple kernel Q to be counted with unit coefficient is read directly from list of quadruple kernels create from D and T kernels using the rule described above. In a notation analogous to that of the previous remarks,

$$C_{T_B}^{(4)} = \{n_{T_B}^{(4)}(Q) - C_T^{(3)}\} \quad (11)$$

$$C_{T_{NB}}^{(4)} = \left\{ n_{T_{NB}}^{(4)}(Q) \right\} \quad (12)$$

$$C_{D_B}^{(4)} = \left\{ n_{D_B}^{(4)}(Q) - C_{D_B}^{(3)} \right\} \quad (13)$$

$$C_{D_{NB}}^{(4)} = \left\{ n_{D_{NB}}^{(4)}(Q) - C_{D_{NB}}^{(3)} \right\} \quad (14)$$

$$C_S^{(4)} = \left\{ n_S^{(4)}(Q) - C_S^{(3)} \right\} \quad (15)$$

3. Application to the protein Insulin

To illustrate the discussion of the previous section we give below an example of the numerical procedure for obtaining the various lists of coefficients associated with the double interaction appropriate for the case of the protein insulin.

In Fig. 3(bottom) we show the topological double connections among the kernels of insulin. Insulin is a small protein containing 783 atoms, assembled within 51 amino acids, which form two polypeptide chains connected together by two sulfur bridges. For the calculations described here, insulin has been broken into 9 single kernels: chain A K1(Gly-Ile-Val-Glu-Gln-), K2(-Cys-Cys-Ala-Ser-Val-Cys-Ser-), K3(-Leu-Tyr-Gln-Leu-), K4(-Glu-Asn-Tyr-Cys-Asn-), Chain B K5(Phe-Val-Asn-Gln-His-Leu-Cys-), K6(-Gly-Ser-His-Leu-Val-Glu-), K7(-Ala-Leu-Tyr-Leu-Val-Cys-), K8(-Gly-Glu-Arg-Gly-Phe-Phe-), K9(-Tyr-Thr-Pro-Lys-Ala). K2 and K5, K4 and K7 connected by s-s bonds. There are several nonneglectable hydrogen bonds between K1 and K3, K1 and K4.

3.1. Second order of KEM, E^{2nd}

Each double kernel to be counted with unit coefficient is read directly from the original list of double kernels for Insulin. Accordingly the double kernels are $E_{12}, E_{13}, E_{14}, E_{23}, E_{25}, E_{34}, E_{47}, E_{56}, E_{67}, E_{78}, E_{89}$ a total of 11 kernels E_D . The coefficient of any single kernel energy E_S is determined by counting the number of times it is represented in the double kernel list, and reducing that number by $C_S^{(1)}$, ($C_S^{(1)} \equiv 1$), the number of occurrences in the list S. Accordingly we obtain for $C_S^{(2)}$ the set of coefficients { 2, 2, 2, 2, 1, 1, 2, 1, 0 }. The total energy E^{2nd} may be calculated using Eq. (4).

3.2. Third order of KEM, E^{3rd}

For the case of insulin, application of the rule for creating the triple kernels of Eq. (5) from the list of double kernel interactions is illustrated in Fig 4.

Insulin has 11 double kernels as Fig. 4 indicated in described above. Each double kernel compared with all double kernels, by following the rule, delivers 14 triple kernels E_T ($E_{123}, E_{124}, E_{125}, E_{134}, E_{147}, E_{234}, E_{235}, E_{256}, E_{347}, E_{467}, E_{478}, E_{567}, E_{678}, E_{789}$). The triple list is a source of both bonded and non-bonded double kernels. Each triple kernel corresponds to 3 double kernels, so 14 triple kernels E_T are associated with 42 double

kernels E_D , of which only 22 double kernels are uniquely named. There are 11 of the 22 double kernels of the bonded type and 11 of the non-bonded type. The coefficient of any double kernel energy E_D is determined by counting the number of times it is represented in the triple kernel list, and reducing that number by $C_D^{(2)}$ ($C_D^{(2)} \equiv 1$ for bonded type, and $C_D^{(2)} \equiv 0$ for non-bonded type).

Each triple kernel corresponds to 3 single kernels, so 14 triple kernels E_T are associated with 42 single kernels E_S , and of these only 9 single kernels are unique. The coefficient of any single kernel in equation (5) is the number of single kernels contained in the set of triple kernels T reduced by the coefficient of the single kernels in the 2nd order approximation. These coefficients are listed in table 3. The total KEM energy E^{3rd} may be calculated according to Eq. (5).

3.3. Fourth order of KEM, E^{4th}

Application of the rule for creating quadruple kernels in Eq. (6) from double and triple kernels is illustrated in Fig 5.

Insulin has 11 double kernels, and 14 triple kernels pictured in fig 5. Each double kernel compared with all triple kernels following the rule for creating Q-kernels, delivers as a net result 19 unique quadruple kernels E_Q (E_{1234} , E_{1235} , E_{1245} , E_{1247} , E_{1256} , E_{1347} , E_{1467} , E_{1478} , E_{2345} , E_{2347} , E_{2356} , E_{2567} , E_{3467} , E_{3478} , E_{4567} , E_{4678} , E_{4789} , E_{5678} , E_{6789}).

Each quadruple kernel corresponds to 4 triple kernels, so 19 quadruple kernels E_Q form 76 triple kernels E_T , of which 45 triple kernels are unique. There are 14 of the 45 triple kernels which are bonded, and their count in the list Q, in calculating $C_T^{(4)}$, is reduced by the coefficient $C_T^{(3)} \equiv 1$. Similarly for each non-bonded triple kernel, the number of times it occurs in the list of quadruple kernels, is reduced by the coefficient $C_T^{(3)} \equiv 0$. The triple kernels and their coefficients are indicated in table 4.

Each quadruple kernel corresponds to 6 double kernels, so 19 quadruple kernels E_Q form 114 double kernels E_D , of which only 31 double kernels are unique. Of these, 22 of the 31 double kernels appear in the triple formula (i.e., Eq. 5). Their coefficient in Eq. (6) is obtained by counting the number of times they occur in the set Q, reduced by the coefficient, $C_D^{(3)}$. The remaining 9 double kernels do not appear in the triple formula, and their coefficients are the count of their appearances in the list Q.

Each quadruple kernel corresponds to 4 single kernels, so 19 quadruple kernels E_Q are related to 76 single kernels E_S , of which only 9 single kernels are unique. The coefficient of each single kernel in 4th order is the count of the single kernels corresponding to the list Q reduced by the coefficient of S in the 3rd order approximation, $C_S^{(3)}$. These single kernels and their coefficients are listed in table 6. The total energy E^{4th} may be calculated according to Eq. (6).

3.4. Calculation of the kernel energies and total energies of 2nd, 3rd, 4th order of KEM

In table 7 are listed the calculated individual kernel energies needed for evaluation of Eq.s (4), (5) and (6). The chemical model used is HF/STO-3G. Table 8 shows the summary of coefficients used in the Eq.s (4), (5) and (6).

The total energy is calculated using equation (4), (5) and (6) together with the kernel energies displayed in the Table 7 and coefficients in Table 8. The results are given in the Table 9.

DISCUSSION and CONCLUSIONS

We are concerned here with enhancing the computational efficiency of the KEM formalism so that only the most important kernels and interactions are taken into account. The foundation for this consideration is the list of assumed double kernel interactions. This represents all the most important kernel pair interactions. The question then is how to represent all topologically continuous interacting 3-tuples and 4-tuples associated with the list of double kernel interactions. A simple rule answers the question. All the required interacting 3-tuples are obtained by 2-tuple comparisons which are retained on condition they share only one index. All interacting 4-tuples are obtained by comparing the 2-tuples to 3-tuples retaining those which have one index in common. When a given symbol occurs multiple times from such comparisons it is retained only once. Thus one builds up a complete set of lists of kernels S , D , T , & Q . The energies corresponding to these symbols can be used to calculate the KEM energies to the various orders of approximation, as in Eq.s 4, 5, & 6. In these equations the coefficients of each kernel are the count of the number of times they occur in the list of kernel energies. For example, a single kernel s can occur in any of the lists S , D , T , & Q . Similarly, a given d can occur in D , T , & Q , and a particular t can occur in T , & Q , while a particular q occurs in the list Q . We have shown how the counts delivering the coefficients in the lists C_S , C_D , C_T , and C_Q can be carried out, based upon the number of times a given kernel index occurs in the lists of S , D , T , & Q , and we have listed their values in table 8.

When the various kernels are used in equations 4, 5, & 6 to calculate the energy, results of high accuracy occur. The accuracy is seen to increase with each additional increase in the level of approximation.

The entire procedure extracts from a list of double kernel interactions indices, all the indices naming higher order topologically continuous interacting kernels, as in Eq.s 4, 5, and 6. The procedure described, is a numerical one, based upon comparisons of the indices that emanate from the list of double kernels. Once the complete list of interactions is obtained, it is unpacked in the counting of how many times each kernel is represented in the kernel expansion of the total energy, as in Eq.s 4, 5, 6. Again, the unpacking of the interactions is a numerical procedure based upon analysis of the indices that name the various interactions.

That a substantial saving of computational time will follow from judicious use of kernel interaction diagrams is evident from a simple counting of the number of kernels that need not be calculated in such circumstances. For example in the case of the insulin calculation of this paper, the total number of kernels of types S , D , T , and Q is 9, 36, 84, 126 respectively, which would be used in the eqs (1), (2), and (3). In contrast, based upon the list of double kernel interactions for insulin, the number of kernels of types S , D , T , and Q is 9, 31, 45, 19 respectively, which would be used in the eqs (4), (5), and (6), requiring many fewer kernel calculations, which results in a dramatic decrease in computational cost, while retaining very high numerical accuracy in the KEM calculations. Notice that the greatest computational saving occurs for the largest kernels in the order $Q > T > D > S$.

In a previous paper [26], we have given detailed timings for the way 2nd order (chemically bonded kernels) calculations compare to calculation of a full molecule. For most problems the 2nd order calculations would be sufficiently accurate. However, this paper has been concerned also with 3rd and 4th order of accuracy. In the calculations of this paper the Insulin full molecule calculation time may be taken to be of unit magnitude, i.e., equal to 1.000. Using that unit of time, the average relative times for calculation of single, double, triple, and quadruple kernels are, 0.015, 0.065, 0.100, and 0.141, respectively.

One may also discuss in a more general way the computational timings of KEM calculations as follows. In the simplest terms, the complexity of an HF calculation is roughly proportional to n^4 , where n is the number of atoms. If this corresponds to 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. Each multiple kernel would require computational time proportional to an order of $(n/n_K)^4$. Thus, the total computational time, 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. We note that use of quadruple kernels would allow the computation of extremely large molecules with high accuracy, that would not otherwise be possible if the memory space required to store the large number of energy integrals were insufficient. For example, in an HF calculation where the number of integrals to be stored would grow as n^4 , the corresponding memory space required in the case of individual quadruple calculations would only grow as $(n/n_K)^4$.

The rationale for the particular choice of the 1-tuples for insulin used in this paper follows from recognition that the choice of kernels may be chosen with a high degree of arbitrariness. The basic constraints of choice are simply that the bonds broken should be single bonds, that aromatic ring systems and multiple bonds should be preserved, and that no atom should occur in more than one kernel. It is also true that accuracy of the full KEM calculation tends to increase with the size of kernels. Within these heuristic rules we have generally found it computationally convenient to choose kernels of magnitude including roughly 100 atoms. In the case of insulin treated here, kernels of that size dictate that the total number would be about 9 kernels, as we have indicated in our calculations. We have grouped the 5 to 7 residues which occur in our chosen kernels so that each kernel would contain approximately 100 atoms, and also conform to the heuristic rules just previously enunciated.

While the energy accuracy of the results, of even the second-order KEM are quite impressive, and the higher order approximations are even more so, the total energies that we display in this paper are not the only properties available via the KEM. For example, other uses of KEM would include that of obtaining energy optimized geometries. One may convince oneself of this practicality by observing that all the geometrical parameters of the full molecule are each represented in the KEM approximation thereof. It follows that the best least squares approximation of KEM to the full molecule optimized energy is obtained by variationally optimizing the KEM energy with respect to the common geometrical parameters. This may well be useful in the problem of folding large biological molecules, including proteins. We mention also that including the effects of solvents upon molecular interactions is consistent with the ideas which underlie KEM. We have published elsewhere the KEM treatment of explicit solvent molecules [37-38], and have also treated the case of molecules in a variety of continuous background solvents [39].

Acknowledgments

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. (LM) thanks NIH for grants (NIGMS MBRS SCORE5 S06GM606654, and RR-03037 the National Center For Research Resources). We thank a referee for insightful suggestions which have improved our manuscript.

REFERENCES

- [1]. Lee C, Yang W. J. Chem. Phys 1992;96:2408.
- [2]. Yang W. J. Mol. Struc. THEOCHEM 1992;255:46.

- [3]. Yang, W. Condensed-Matter Theories. Blum, L.; Malik, FB., editors. Plenum Press; Berlin: 1993. p. 367-372.
- [4]. Exner TE, Mezey PG. J. Phys. Chem. A 2002;106:11791.
- [5]. Exner TE, Mezey PG. J. Comput. Chem 2003;24:1980. [PubMed: 14531052]
- [6]. Kitaura K, Ikey E, Asada T, Nakano T, Uebayasi M. Chem. Phys. Lett 1999;313:701.
- [7]. Kitaura K, Sugiki S-I, Nakano T, Komeiji Y, Uebayasi M. Chem. Phys. Lett 2001;336:163.
- [8]. Inadomi Y, Nakano T, Kitaura K, Nagashima U. Chem. Phys. Lett 2002;364:139.
- [9]. Imamura A, Aoki Y, Maekawa K. J. Chem. Phys 1991;95:5419.
- [10]. Korchowiec J, Gu FL, Imamura A, Kirtman B, Aoki Y. Int. J. Quantum Chem 2005;102:785.
- [11]. Babu K, Gadre SR. J. Comput. Chem 2003;24:484. [PubMed: 12594791]
- [12]. Gadre SR, Shirsat RN, Limaye AC. J. Phys. Chem 1994;98:9165.
- [13]. Gadre SR, Ganesh V. J. Theor. Comput. Chem 2006;5:835.
- [14]. Das GP, Yeates AT, Dudis DD. Int. J. Quantum Chem 2004;120:6832.
- [15]. Li SH, Li W, Fang T. J. Am. Chem. Soc 2005;127:7215. [PubMed: 15884963]
- [16]. Zhang DW, Zhang JZH. J. Chem. Phys 2003;119:3599.
- [17]. Zhang DW, Xiang Y, Zhang JZH. J. Phys. Chem. B 2003;107:12039.
- [18]. Bettens RPA, Lee AM. J. Phys. Chem. A 2006;110:8777. [PubMed: 16836440]
- [19]. Deev V, Collins MA. J. Chem. Phys 2005;122:154102. [PubMed: 15945620]
- [20]. Collins MA, Deev V. J. Chem. Phys 2006;125:104104. [PubMed: 16999512]
- [21]. Rezac J, Salahub DR. J. Chem. Theory. Comput 2010;6:91.
- [22]. Massa L, Huang L, Karle J. International Journal of Quantum Chemistry: Quantum Chemistry Symposium 1995;29:371.
- [23]. Huang L, Massa L, Karle J. International Journal of Quantum Chemistry: Quantum Chemistry Symposium 1996;30:1691.
- [24]. Karle J, Huang L, Massa L. G., Tsoucaris NATO Science Serise C: Mathematical and Physical Sciences 1999;519:1.
- [25]. Huang L, Massa L, Karle J. IBM J. RES. & DEV 2001;45(3/4):409.
- [26]. Huang L, Massa L, Karle J. Int. J. Quantum Chem 2005;103:808.
- [27]. Huang L, Massa L, Karle J. J. Int. J. Quantum Chem 2006;106:447.
- [28]. Huang L, Massa L, Karle J. J. PNAS 2009;106:1731.
- [29]. Huang L, Massa L, Karle J. J. PNAS 2008;105:1849.
- [30]. Huang, L.; Massa, L.; Karle, J.; Matta, CF., editors. Quantum Biochemistry, Electronic Structure and Biological Activity. WILEY-VCH Verlag GmbH & Co. KGaA; Weinheim: 2010. p. 3-60.
- [31]. Huang, L.; Massa, L.; Karle, J. Encyclopedia of Computational Chemistry. von Schleyer, P., editor. John Wiley & Sons; New York: 1998. p. 1457-1470.
- [32]. Stewart WN, Huang L, Massa L. Journal of Computational Chemistry. (Published Online: Jun 16 2010 5:45AM) DOI: 10.1002/jcc.21584.
- [33]. Huheey, JE. Inorganic Chemistry: Principles of Structure and Reactivity. third ed. Wasserman, Malvina, editor. Cynthia L. Indriso, Harper & Row, Publishers; New York: 1983. p. 256-283. Sponsoring ISBN 0-06-042987-9
- [34]. Dunitz, JD. X-Ray Analysis and the Structure of Organic Molecules, 2nd corrected reprint. Kisakürek, M. Volkan, editor. VCH Publishers, Inc.; NY (USA): 1995. p. 391-417. Verlag Helvetica Chimica Acta, Basel (Switzerland), VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany) ISBN 3-906390-14-4
- [35]. Bader, RFW. Atoms in Molecules: A Quantum Theory. Oxford University Press; Oxford: 1990.
- [36]. Matta, CF.; Boyd, RJ., editors. The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design. WILEY-VCH; Weinham: 2007.
- [37]. Huang L, Massa L, Karle J. Biochemistry 2006;44:16747. [PubMed: 16342965]
- [38]. Huang L, Massa L, Karle J. PNAS 2007;104:4261. [PubMed: 17360512]
- [39]. Huang, L.; Massa, L. (unpublished)

Fig 1.
The double kernel-double kernel comparison implies triple kernels.

Fig 2.

The double kernel-triple kernel comparison implies quadruple kernels.

Fig 3.
(top) Insulin has 2 chains, red and blue. (bottom) Chain A divided into 4 single kernels, and chain B divided into 5 single kernels.

Fig 4.
Application of the rule for creating KEM insulin triple kernels.

Fig 5.
Application of the rule for creating KEM Insulin quadruple kernels.

Table 1

Indicated, are the mathematical signs (+ or -) of a given kernel as it occurs in the lists of kernels. The symbols s, d, t, & q are of elements of the lists of kernels S, D, T, & Q, respectively.

	S	D	T	Q
s	+1	-1	+1	-1
d		+1	-1	+1
t			+1	-1
q				+1

Table 2

Double kernels E_D and corresponding coefficients C_D in equation (5), 3rd order of KEM.

	E_D bonded	$n_D^{(3)}(T) - 1$	$C_D^{(3)}B$		E_D non-bonded	$C_D^{(3)}NB$
1	E_{12}	3 - 1	2	12	E_{15}	1
2	E_{13}	2 - 1	1	13	E_{17}	1
3	E_{14}	3 - 1	2	14	E_{24}	2
4	E_{23}	3 - 1	2	15	E_{26}	1
5	E_{25}	3 - 1	2	16	E_{35}	1
6	E_{34}	3 - 1	2	17	E_{37}	1
7	E_{47}	4 - 1	3	18	E_{46}	1
8	E_{56}	2 - 1	1	19	E_{48}	1
9	E_{67}	3 - 1	2	20	E_{57}	1
10	E_{78}	3 - 1	2	21	E_{68}	1
11	E_{89}	1 - 1	0	22	E_{79}	1

Table 3

Single kernels E_S and their coefficients in equation (5), 3rd order of KEM.

	E_S	$n_S^{(3)}(T) - C_S^{(2)}$	$C_S^{(3)}$
1	E_1	5 – 2	3
2	E_2	6 – 2	4
3	E_3	5 – 2	3
4	E_4	7 – 2	5
5	E_5	4 – 1	3
6	E_6	4 – 1	3
7	E_7	7 – 2	5
8	E_8	3 – 1	2
9	E_9	1 – 0	1

Table 4

Triple kernels E_T and their coefficients in equation (6), 4th order KEM.

	E_T bonded	$n_T^{(4)}(Q) - 1$	$C_{TB}^{(4)}$		E_T non-bonded	$C_{TNB}^{(4)}$		E_T non-bonded	$C_{TNB}^{(4)}$
1	E_{123}	2 - 1	1	15	E_{126}	1	31	E_{345}	1
2	E_{124}	3 - 1	2	16	E_{127}	1	32	E_{346}	1
3	E_{125}	3 - 1	2	17	E_{135}	1	33	E_{348}	1
4	E_{134}	2 - 1	1	18	E_{137}	1	34	E_{356}	1
5	E_{147}	4 - 1	3	19	E_{145}	1	35	E_{367}	1
6	E_{234}	3 - 1	2	20	E_{146}	1	36	E_{378}	1
7	E_{235}	3 - 1	2	21	E_{148}	1	37	E_{456}	1
8	E_{256}	3 - 1	2	22	E_{156}	1	38	E_{457}	1
9	E_{347}	4 - 1	3	23	E_{167}	1	39	E_{468}	1
10	E_{467}	4 - 1	3	24	E_{178}	1	40	E_{479}	1
11	E_{478}	4 - 1	3	25	E_{236}	1	41	E_{489}	1
12	E_{567}	3 - 1	2	26	E_{237}	1	42	E_{568}	1
13	E_{678}	3 - 1	2	27	E_{245}	2	43	E_{578}	1
14	E_{789}	2 - 1	1	28	E_{247}	2	44	E_{679}	1
				29	E_{257}	1	45	E_{689}	1
				30	E_{267}	1			

Table 5

Double kernels E_D , and their coefficients in equation (6) 4th order of KEM.

	E_0	$n_D^{(4)}(q) - c_D^{(3)}$	$c_D^{(4)}$		E_0	$c_D^{(4)}$
1	E_{12}	5 - 2	3	23	E_{16}	2
2	E_{13}	3 - 1	2	24	E_{18}	1
3	E_{14}	6 - 2	4	25	E_{27}	3
4	E_{23}	5 - 2	3	26	E_{36}	2
5	E_{25}	6 - 2	4	27	E_{38}	1
6	E_{34}	6 - 2	4	28	E_{45}	3
7	E_{47}	10 - 3	7	29	E_{49}	1
8	E_{56}	5 - 1	4	30	E_{58}	1
9	E_{67}	7 - 2	5	31	E_{69}	1
10	E_{78}	6 - 2	4			
11	E_{89}	2 - 0	2			
12	E_{15}	3 - 1	2			
13	E_{17}	4 - 1	3			
14	E_{24}	5 - 2	3			
15	E_{26}	3 - 1	2			
16	E_{35}	3 - 1	2			
17	E_{37}	4 - 1	3			
18	E_{46}	4 - 1	3			
19	E_{48}	4 - 1	3			
20	E_{57}	3 - 1	2			
21	E_{68}	3 - 1	2			
22	E_{79}	2 - 1	1			

Table 6

Single kernels E_S and their coefficients in equation 6, 4th order of KEM.

	E_S	$n_S^{(4)}(Q) - C_S^{(3)}$	$C_S^{(4)}$
1	E_1	8 - 3	5
2	E_2	9 - 4	5
3	E_3	8 - 3	5
4	E_4	13 - 5	8
5	E_5	8 - 3	5
6	E_6	9 - 3	6
7	E_7	13 - 5	8
8	E_8	6 - 2	4
9	E_9	2 - 1	1

Table 7

List of quadruple, triple, double and single kernel energies

	E _Q		E _T		E _D		E _S	
	Kernel	Energy [au]	kernel	Energy [au]	kernel	Energy [au]	kernel	Energy [au]
1	E ₁₂₃₄	-9141.0035	E ₁₂₃	-6605.3993	E ₁₂	-4898.4858	E ₁	-1796.6145
2	E ₁₂₃₅	-9708.6010	E ₁₂₄	-7435.0891	E ₁₃	-3504.5522	E ₂	-3102.8732
3	E ₁₂₄₅	-10538.293	E ₁₂₅	-8001.6895	E ₁₄	-4333.2163	E ₃	-1707.9520
4	E ₁₂₄₇	-9893.6338	E ₁₂₆	-7029.3493	E ₁₅	-4900.6684	E ₄	-2536.6120
5	E ₁₂₅₆	-10131.538	E ₁₂₇	-7357.8857	E ₁₆	-3927.4785	E ₅	-3104.0541
6	E ₁₃₄₇	-8498.6958	E ₁₃₄	-6040.1544	E ₁₇	-4256.0142	E ₆	-2130.8665
7	E ₁₄₆₇	-8921.6508	E ₁₃₅	-6608.6037	E ₁₈	-4131.7840	E ₇	-2459.4030
8	E ₁₄₇₈	-9125.9391	E ₁₃₇	-5963.9482	E ₂₃	-4809.7998	E ₈	-2335.1693
9	E ₂₃₄₅	-10448.613	E ₁₄₅	-7437.2705	E ₂₄	-5639.4863	E ₉	-1947.3434
10	E ₂₃₄₇	-9803.9580	E ₁₄₆	-6464.0809	E ₂₅	-6206.0748		
11	E ₂₃₅₆	-10042.849	E ₁₄₇	-6791.7607	E ₂₆	-5233.7389		
12	E ₂₅₆₇	-10794.352	E ₁₄₈	-6668.3992	E ₂₇	-5562.2767		
13	E ₃₄₆₇	-8831.9995	E ₁₅₆	-7030.5177	E ₃₄	-4243.5635		
14	E ₃₄₇₈	-9036.2859	E ₁₆₇	-6385.9037	E ₃₅	-4812.0038		
15	E ₄₅₆₇	-10228.090	E ₁₇₈	-6590.1789	E ₃₆	-3838.8179		
16	E ₄₆₇₈	-9459.2308	E ₂₃₄	-7345.4126	E ₃₇	-4167.3513		
17	E ₄₇₈₉	-9275.6520	E ₂₃₅	-7912.9992	E ₃₈	-4043.1210		
18	E ₅₆₇₈	-10026.497	E ₂₃₆	-6940.6647	E ₄₅	-5640.6664		
19	E ₆₇₈₉	-8869.7716	E ₂₃₇	-7269.1999	E ₄₆	-4667.4791		
20			E ₂₄₅	-8742.6881	E ₄₇	-4995.1597		
21			E ₂₄₇	-8098.0344	E ₄₈	-4871.7946		
22			E ₂₅₆	-8335.9255	E ₄₉	-4483.9556		
23			E ₂₅₇	-8665.4753	E ₅₆	-5233.9062		
24			E ₂₆₇	-7692.1680	E ₅₇	-5563.4543		

	E _Q		E _T		E _D		E _S	
	Kernel	Energy [au]	kernel	Energy [au]	kernel	Energy [au]	kernel	Energy [au]
25			E ₃₄₅	-7347.6158	E ₅₈	-5439.2235		
26			E ₃₄₆	-6374.4299	E ₆₇	-4589.2952		
27			E ₃₄₇	-6702.1082	E ₆₈	-4466.0351		
28			E ₃₄₈	-6578.7461	E ₆₉	-4078.2078		
29			E ₃₅₆	-6941.8552	E ₇₈	-4793.5676		
30			E ₃₆₇	-6297.2424	E ₇₉	-4406.7462		
31			E ₃₇₈	-6501.5154	E ₈₉	-4281.4838		
32			E ₄₅₆	-7770.5190				
33			E ₄₅₇	-8099.2112				
34			E ₄₆₇	-7125.0523				
35			E ₄₆₈	-7002.6609				
36			E ₄₇₈	-7329.3378				
37			E ₄₇₉	-6942.5031				
38			E ₄₈₉	-6818.1092				
39			E ₅₆₇	-7692.3321				
40			E ₅₆₈	-7569.0748				
41			E ₅₇₈	-7897.6189				
42			E ₆₇₈	-6923.4602				
43			E ₆₇₉	-6536.6363				
44			E ₆₈₉	-6412.3490				
45			E ₇₈₉	-6739.8817				

Table 8

The coefficients of the kernels in the Equations 4,5,6.

	E_8	$C_S^{(2)}$	$C_S^{(3)}$	$C_S^{(4)}$	E_D in eq (5)	$C_D^{(3)}$	E_D in eq (6)	$C_D^{(4)}$	E_T in eq (6)	$C_T^{(4)}$
1	E_1	2	3	5	E_{12}	2	E_{12}	3	E_{123}	1
2	E_2	2	4	5	E_{13}	1	E_{13}	2	E_{124}	2
3	E_3	2	3	5	E_{14}	2	E_{14}	4	E_{125}	2
4	E_4	2	5	8	E_{15}	1	E_{15}	2	E_{126}	1
5	E_5	1	3	5	E_{17}	1	E_{16}	2	E_{127}	1
6	E_6	1	3	6	E_{23}	2	E_{17}	3	E_{134}	1
7	E_7	2	5	8	E_{24}	2	E_{18}	1	E_{135}	1
8	E_8	1	2	4	E_{25}	2	E_{23}	3	E_{137}	1
9	E_9	0	1	1	E_{26}	1	E_{24}	3	E_{145}	1
10					E_{34}	2	E_{25}	4	E_{146}	1
11					E_{35}	1	E_{26}	2	E_{147}	3
12					E_{37}	1	E_{27}	3	E_{148}	1
13					E_{46}	1	E_{34}	4	E_{156}	1
14					E_{47}	3	E_{35}	2	E_{167}	1
15					E_{48}	1	E_{36}	2	E_{178}	1
16					E_{56}	1	E_{37}	3	E_{234}	2
17					E_{57}	1	E_{38}	1	E_{235}	2
18					E_{67}	2	E_{45}	3	E_{236}	1
19					E_{68}	1	E_{46}	3	E_{237}	1
20					E_{78}	2	E_{47}	7	E_{245}	2
21					E_{79}	1	E_{48}	3	E_{247}	2
22					E_{89}	0	E_{49}	1	E_{256}	2
23							E_{56}	4	E_{257}	1
24							E_{57}	2	E_{267}	1

	E_S	$C_S^{(2)}$	$C_S^{(3)}$	$C_S^{(4)}$	E_D in eq (5)	$C_D^{(3)}$	E_D in eq (6)	$C_D^{(4)}$	E_T in eq (6)	$C_T^{(4)}$
25							E_{58}	1	E_{345}	1
26							E_{67}	5	E_{346}	1
27							E_{68}	2	E_{347}	3
28							E_{69}	1	E_{348}	1
29							E_{78}	4	E_{356}	1
30							E_{79}	1	E_{367}	1
31							E_{89}	2	E_{378}	1
32									E_{456}	1
33									E_{457}	1
34									E_{467}	3
35									E_{468}	1
36									E_{478}	3
37									E_{479}	1
38									E_{489}	1
39									E_{567}	2
40									E_{568}	1
41									E_{578}	1
42									E_{678}	2
43									E_{679}	1
44									E_{689}	1
45									E_{789}	1

Table 9

Calculation of the total KEM energy for the protein insulin.

	2 nd order KEM	3 rd order KEM	4 th order KEM
Full molecule energy [a.u.]	-21112.11241	-21112.11241	-21112.11241
KEM energy [a.u.]	-21112.10555	-21112.11216	-21112.11240
diff (full-KEM) [a.u.]	-0.00686	-0.00025	-0.00002
diff (full-KEM) [kcal/mol]	-4.30436	-0.15835	-0.00965