Kernel Energy Method: Application to DNA†

Lulu Huang,‡ Lou Massa,§ and Jerome Karle*,‡

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

Received August 19, 2005; Revised Manuscript Received October 20, 2005

ABSTRACT: The kernel energy method (KEM) has been used in three recent papers (*1*−*3*) to calculate the quantum mechanical *ab inito* molecular energy of peptides and the protein insulin. It was found to have good accuracy. The computational difficulty of representing a molecule increases only modestly with the number of atoms. The calculations are simplified by adopting the approximation that a full biological molecule can be represented by smaller "kernels" of atoms. In this paper, the accuracy of the KEM is tested in the application to DNA, whose basic kernels, chemical bonding, and overall molecular structure are quite different from peptides and proteins. The basic kernel in the case of peptides and proteins is an amino acid. The basic kernel in the case of DNA is a nucleotide consisting of a phosphate−sugar−base. The molecular energy is calculated for all three basic types of DNA, i.e., B, A, and Z configurations of DNA. The results give an accuracy that is comparable to that achieved with peptides and proteins. Thus, the KEM is found to be applicable to major types of biological molecules.

This paper combines structural crystallographic information with the quantum-mechanical theory. The objective is to simplify computational chemistry calculations and enhance the information that may be derived from a crystallographic experiment. This paper focuses on the calculation of the energies of DNA structures.

The kernel energy method (KEM)[1] 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 double kernels are practicable. Subsequently, kernel contributions are summed in a manner to be described, which affords an estimate of the energy for the whole molecule. Thus, the task of obtaining a quantum mechanical energy is simplified for biological molecules as large as the various configurations of DNA studied in this paper. The computational time is much reduced by employing the KEM, and the accuracy obtained appears to be quite satisfactory.

The first applications of the KEM (*1*−*3*) involved a large number of peptides and also the protein insulin. These studies showed that the KEM applied with good accuracy to insulin and a very wide variety of peptides of various shapes and sizes. The good accuracy was retained throughout a wide range of basis functions and computational methods (*2*). The overall theoretical background for the application of quantum mechanics with crystallography may be found in refs *4*−*11*. References that review the quantum mechanical methods related to this work may be found in ref *6*.

This paper is devoted to the calculation of the energies of various configurations of DNA molecules. Only one each of the basis sets and computational methods that were previously tested within the KEM was applied. All calculations employ a limited basis and the Hartree−Fock (HF) approximation.

REVIEW OF THE KEM

Each macromolecular structure, with known atomic coordinates, is composed of pieces called kernels. The set of all kernels forms the complete structure. Schematically defined kernels are shown in Figure 1. The kernels are chosen so that each atom occurs in only one kernel. This rule plays a valuable role in defining how the kernel calculations are to be combined in forming the energy of the full molecule. All of the atoms of the entire molecule are represented in the collection of all of the kernels. Figure 1 also indicates that kernels may be collected in pairs and that these are called double kernels. All quantum calculations are carried out on single and double kernels only.

The molecular energy is described by a sum over the separate contributions of double kernels reduced by the energy of those single kernels, which have been overcounted. If only the chemically bonded double kernels are considered, the total energy E is decomposed in this approximation as

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

where E_{ij} is the energy of a chemically bonded double kernel of the name ij, E_i is the energy of a single kernel of name i, i and j are running indices, and n is the number of kernels.

† This work was supported by the Office of Naval Research and, in addition, for L.M., the Navy Summer Faculty Research Program, NIH Grants (NIGMS MBRS SCORE5 S06GM606654, and RR-03037 the National Center For Research Resources), and the NSF for the CREST grant.

* To whom correspondence should be addressed. Telephone: (202) 767-3496. Fax: (202) 767-6874. E-mail: williams@harker.nrl.navy.mil.

‡ Naval Research Laboratory.

§ City University of New York.

[1] Abbreviations: KEM, kernel energy method; HF, Hartree−Fock.

10.1021/bi051655l CCC: $30.25
Published on Web 11/22/2005

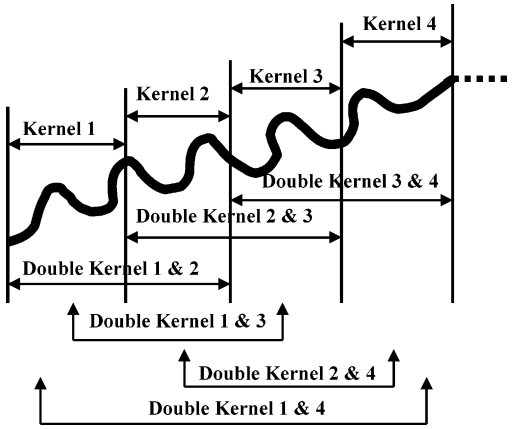

FIGURE 1: Abstract sketch of a polymer showing single and double kernels.

For better energy accuracy, all double kernels are included. In that case, the total energy E, is decomposed in this approximation as

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

where E_{ij} is the energy of a double kernel of name ij, E_i is the energy of a single kernel of name i, i, j, and m are running indices, and n is the number of single kernels.

The above formulas are applied below to the calculation of the molecular energies of a variety of DNA molecules. The purpose of the calculations is to obtain kernel contributions to the energy when it is not computationally feasible to treat the entire molecule as a whole. When a structure of interest has known crystallographic coordinates, the kernels may be readily defined, which altogether represent the entire molecule. The use of the single and double kernels indicated above is an approximation that has been made to obtain a simplification in the quantum calculation. The validity of this approximation, in the case of a variety of DNA structures, may be seen in the application to the molecules discussed below.

DNA STRUCTURES CALCULATED BY THE KEM

This paper concerns the calculation of molecular energy using the concept of kernels in application to DNA molecules. DNA molecules of known crystal structure and of biological importance have been chosen for study. Also chosen are examples that are sufficiently large to provide significant demonstrations of *ab inito* energy calculations with the use of the KEM but not so large as to prevent energy calculations of whole molecules using supercomputers. The latter cases were required to provide a standard of excellence against which the approximations using kernels could be judged. Selected for study are the group of DNA structures defined by their nucleic acid sequence shown in Table 1. The pictures of their crystal structure geometries are available in the data banks PDB and NDB.

The DNA molecule is composed of nucleotides, i.e., of phosphate−sugar−base units, which form the elemental kernels for our calculations. Base pairs across the space between the phosphate sugar backbones stabilize the double-helical structure by means of hydrogen bonding between purines and pyrimidines (A with T and G with C), and such stabilization must be accounted for in the KEM. There are three basic forms of DNA, which may be seen in the common literature called A, B, and Z. The properties and pictorial representations characteristic of these forms may also be obtained in the literature (*12*−*14*).

The fundamental biological activity of DNA molecules is dependent upon their geometrical structure, which is correlated with their energy. Thus, the molecular energy of DNA molecular conformations is fundamental to the study of their structure and their function. In this paper, it is shown how the concept of kernels allows for accurate calculation of DNA molecular energy in a variety of conformations. All of the crystal structures for the molecular sequences in Table 1 are known (*15*−*25*) and have been used in the energy calculations presented here. Quite a wide variety of DNA forms are represented there, which vary in size, shape, and function.

The first three molecules (1−3) of Table 1 are single-helix conformations of B DNA. The molecule 1 is a hexamer of nucleotides (*15*). The four central bases are typical of a B-form helix, but the terminal bases have an A-form conformation. A wide minor groove (6.9 Å) contains a double spine of hydration. The molecule 2 is a hexanucleotide−anthracycline complex (*16*). The anthracyline is bound to nonpreferred base-pair triplet sites. Anthracyclines are antibiotics and may be antitumor agents. Their biological activity is a result of their binding interactions with DNA. The case of molecule 3 is the first structure of a deoxyoligonucleotide crystallized together with zinc ions, whose presence in the crystal structure is clearly identified (*17*). The terminal bases at either end of the molecule occupy nonhelical geometries.

The next seven molecules (4−10) in Table 1 are double-helix structures of B DNA form. The molecule 4 is a hexamer having the interesting history that it has been captured in crystalline form, in each step of a sequence of steps, showing the transformation from B to A forms (*18*). Only the initial B form of the molecule is calculated. The case of molecule 5 is the first example of crystalline spermine and native B DNA (*19*). The asymmetric unit contains 1 hexamer duplex, 17 water molecules, and 1.5 spermine molecules. The B DNA minor groove is narrow (5.5 Å) and deep (7 Å). The hexamer displays non-Watson−Crick base pairing between adenine and thymine in one region of the molecule. The heptanucleotide designated as molecule 6 is the first oligonucleotide crystal of B-form DNA obtained in a cubic system (*20*). It has a crystalline organization that may be useful in the development of nano-devices based upon DNA structure. The decamer, molecule 7, contains eight Watson−Crick base pairs capped at either end of the molecule with nucleotide 5′ sticky ends (*21*). This first example of a decamer with sticky ends associates end-to-end and forms an essentially infinite helix within the crystal. The crystal structure of molecule 8 from a papilloma virus is a variant of B DNA (*22*). The molecule 9 is related to the previous molecule, as though it consisted of three of them, arranged along mutually perpendicular axes (*22*). The final structure of the B form considered is molecule 10. This is a dodecamer complexed with propamine (*23*). The propamine is bound within the AT tract of the DNA minor groove.

Table 1: DNA Sequence Definitions

number	molecules	chain	DNA seqences	resolution (Å)	water
1	251D	A	CTCGAG	1.90	32
2	110D	A	CGGCCG	1.90	50
3	1G6D	A	CGCAATTGCG	2.90	39
4	1IH1	A, B	GGCGCC, GGCGCC	2.00	no
5	206D	A, B	CGGTGG, CCACCG	2.50	17
6	1S9B	A, B	GAATTCG, GAATTCG	2.81	20
7	309D	A, B	CGACGATCGT, CGACGATCGT	2.60	68
8	425D	A, B	ACCGGTACCG GT, ACCGGTACCG GT	2.80	8
9	424D	A, B, C, D, E, F	3(ACCGGTACCG GT, ACCGGTACCG GT)	2.70	18
10	102D	A, B	CGCAAATTTG CG, CGCAAATTTG CG	2.20	73
11	1D48	A, B	CGCGCG, CGCGCG	1.00	47
12	ADH010	A, B	GGTATACC, GGTATACC	1.80	no

Table 2: Energy Calculation for DNA without Solvents[a]

energy/molecule	B-DNA 251D	B-DNA 110D	B-DNA 1G6D	B-DNA 1IH1	B-DNA 206D	B-DNA 1S9D
number of atoms	198	197	330	394	395	466
(number of kernels)	(3 kernels)	(3 kernels)	(5 kernels)	(6 kernels)	(6 kernels)	(7 kernels)
E_{HF} (atomic unit)	−8127.6777	−8143.3804	−13 614.2601	−16 287.6312	−16 270.5032	−19 082.2964
E_{KEM}[b] (eq 1) (atomic unit)	−8127.6778	−8143.3803	−13 614.2609	−16 287.6519	−16 270.4951	−19 082.2954
$E_{HF} - E_{KEM}$[b] (kcal/mol)	0.0795	−0.0802	0.4643	13.0150	−5.0828	−0.6149
E_{KEM}[c] (eq 2) (atomic unit)	−8127.6776	−8143.3804	−13 164.2600	−16 287.6324	−16 270.5044	−19 082.2963
$E_{HF} - E_{KEM}$[c] (kcal/mol)	−0.0500	−0.0328	−0.0779	0.7520	0.7515	−0.0420

energy/molecule	B-DNA 309D	B-DNA 425D	B-DNA 424D[d]	B-DNA 102D	Z-DNA 1D48	A-DNA ADH010
number of atoms	658	788	2418	790	394	528
(number of kernels)	(6 kernels)	(6 kernels)	(18 kernels)	(6 kernels)	(6 kernels)	(8 kernels)
E_{HF} (atomic unit)	−27 079.0830	−32 509.9718	−97 529.4207	−32 476.7427	−16 286.6395	−21 652.5366
E_{KEM}[b] (eq 1) (atomic unit]	−27 079.0939	−32 509.9638		−32 476.7451	−16 286.6428	−21 652.5418
$E_{HF} - E_{KEM}$[b] (kcal/mol)	6.8480	−4.9989		1.5031	2.0736	3.2426
E_{KEM}[c] (eq 2) (atomic unit)	−27 079.0842	−32 509.9774	−97 529.4212	−32 476.7480	−16 286.6268	−21 652.5389
$E_{HF} - E_{KEM}$[c] (kcal/mol)	0.7393	3.5453	0.3410	−1.2018	−7.9827	1.4089

[a] The KEM applied to DNA using eqs 1 and 2 and with HF/STO-3G. [b] The only double kernels included are those made of single-kernel pairs, which are chemically bonded to one another. [c] All double kernels are included. [d] 424D has three double-helix chains, $E_{HF} = E_{ab} + E_{cd} + E_{ef}$ (see the text).

The last two molecules (11−12) of Table 1 depict the Z and A forms of double-helix DNA. Molecule 11 has the structure of a spermine Z DNA hexamer complex (*24*). The A form of DNA is illustrated by molecule 12. Electrostatic stacking interactions between adjacent guanine and thymine bases induce systematic bending of the double helix and major-groove widening (*25*).

The wide variety of examples illustrated in Table 1 are sufficiently diverse in geometry and bonding to provide a good test for assessing the accuracy of application of the KEM to DNA molecular systems.

KEM CALCULATED RESULTS

The numerical results obtained in this work using the HF equations and a limited basis are listed in Tables 2 and 3. In Table 2, the results for each of the dozen DNA systems discussed above are displayed. For each DNA system, the number of atoms and the number of kernels involved are shown. These range from 198 atoms and 3 kernels for the smallest molecule (B-DNA 251D) up to 2418 atoms and 18 kernels for the largest molecule considered (B-DNA 424D). For each DNA molecular system (except for 424D, which, because of its size, is a special case), the full molecule HF energy E_{HF} is indicated. This number is the standard against which the accuracy of KEM results are judged. The energies listed as E_{KEM} represent the results obtained by dividing the DNA molecular systems into kernels and then calculating the total energy in the approximations formalized within eqs 1 and 2 above. The results of eqs 1 and eqs 2 are listed separately. Also listed for each molecular system are the energy differences $E_{HF} - E_{KEM}$, for both of the cases of eqs 1 and 2. (Note that the full molecular energies are listed in units of au, but the energy differences are listed in the smaller units of kcal/mol.)

As indicated above, the molecule 424D was treated as a special case. Because it contains 2418 atoms and is more than 3 times as large as any of the other molecules that were used in this paper (see Table 2), it was considered that it would not be convenient to calculate the full molecular energy of the entire molecule taken as one whole unit. As indicated in Table 1, it is seen that 424D is composed of 3 individual DNA structures (ab, cd, and ef) oriented almost at right angles to one another, much as would be the case of independent coordinate axes of a Cartesian coordinate frame. In this one case, for computational convenience, the full molecular energy E_{HF} was approximated by calculating the energy of each of the 3 separate DNA structures and then summing them together. Thus, for 424D, the full molecular energy used as a standard of comparison against the KEM result is represented as $E_{HF} = E_{ab} + E_{cd} + E_{ef}$.

A conclusion to be drawn from the results of Table 2 is that the KEM is quite accurate, as one may observe from

Table 3: Energy Calculation for DNA with Solvents[a]

energy/molecule	B-DNA 251D	B-DNA 110D	B-DNA 1G6D	B-DNA 1IH1	B-DNA 206D	B-DNA 1S9D
number of atoms	294	411	447	no solvents	507	526
(number of kernels)	(4 kernels)	(4 kernels)	(7 kernels)		(7 kernels)	(8 kernels)
E_{HF} (atomic unit)	−10 524.4991	−13 634.8546	−16 537.7036		−18 451.4126	−20 580.1620
E_{KEM}[b] (eq 1) (atomic unit)	−10 524.4996	−13 634.8541	−16 537.7057		−18 451.4038	−20 580.1623
$E_{HF} - E_{KEM}$[b] (kcal/mol)	0.2983	−0.2736	1.3104		−5.5187	0.1835
E_{KEM}[c] (eq 2) (atomic unit)	−10 524.4994	−13 634.8542	−16 537.7049		−18 451.4126	−20 580.1632
$E_{HF} - E_{KEM}$[c] (kcal/mol)	0.1688	0.2262	0.7582		0.3157	0.7563

energy/molecule	B-DNA 309D	B-DNA 425D	B-DNA 424D[d]	B-DNA 102D	Z-DNA 1D48	A-DNA ADH010
number of atoms	862	812	2472	1052	575	no solvents
(number of kernels)	(7 kernels)	(7 kernels)	(21 kernels)	(7 kernels)	(7 kernels)	
E_{HF} (atomic unit)	−32 176.1146	−33 109.5892	−98 878.6437[d]	−38 955.9682	−20 413.6870	
E_{KEM}[b] (eq 1) (atomic unit)	−32 176.1135	−33 109.5850		−38 955.9749	−20 413.7238	
$E_{HF} - E_{KEM}$[b] (kcal/mol)	5.4141	−2.6172		4.2088	23.1151	
E_{KEM}[c] (eq 2) (atomic unit)	−32 176.1135	−33 109.5986	−98 878.6420	−38 955.9706	−20 413.7078	
$E_{HF} - E_{KEM}$[c] (kcal/mol)	−0.6947	5.9270	−1.0516	1.5050	13.0589	

[a] The KEM applied to DNA using eqs 1 and 2 and with HF/STO-3G. [b] The only double kernels included are those made of single-kernel pairs, which are chemically bonded to one another. [c] All double kernels are included. [d] 424D has three double-helix chains, $E_{HF} = E_{ab}(5H_2O) + E_{cd}(9H_2O) + E_{ef}(4H_2O)$ (see the text).

the energy differences $E_{HF} - E_{KEM}$. For eq 1, the absolute magnitude of the energy differences range from a minimum of 0.0795 kcal/mol to a maximum of 13.0105 kcal/mol. These differences are relatively small and thus speak to the accuracy of the KEM as implemented in eq 1. The results of eq 2 are even more accurate. For eq 2, the absolute magnitude of the energy differences range from a minimum of 0.0328 kcal/mol to a maximum of 7.9827 kcal/mol. With one exception (B-DNA 1D48), the eq 2 results are generally more accurate than the case for eq 1, as would be expected, and often the increase in accuracy is by as much as an order of magnitude.

The format of Table 3 is an exact replicate of that for Table 2. All of the B DNA systems considered in Table 3 are the same as those of Table 2, except that now the explicit atomic positions of all solvent molecules of crystallization have been taken into account. The manner of treating the solvent molecules is to simply form one or more additional solvent kernels, over and above those representing the DNA molecular system, composed only of solvent molecules occupying their experimental positions in the crystal. Thus, the results of Table 3 concern the energies of the crystalline solvated DNA molecular systems. The meaning of the energy symbols E_{HF}, E_{KEM}, and $E_{HF} - E_{KEM}$ is retained from their use in Table 2. For each DNA system, the presence of solvent molecules increases the total number of atoms and the number of kernels involved in the calculations. These range from 294 atoms and 4 kernels for the smallest molecule (B-DNA 251D) up to 2472 atoms and 21 kernels for the largest molecule considered (B-DNA 424D).

In Table 3, in analogy with the discussion of 424D results listed in Table 2, the solvated molecule is not treated as one whole molecular unit. Again, this is because there are too many atoms (2472) for the calculation to be conveniently managed. As before, this is treated as a special case. The HF energy of each of the solvated DNA substructures is calculated independently. The collection of solvent molecules closest to each substructure is used to form a solvent kernel associated with the molecular system of that substructure. In close association with substructure ab are 5 H_2O, with cd are 9 H_2O, and with ef are 4 H_2O. After the molecular energy of each DNA substructure and its companion solvent molecules are calculated as one whole molecular unit, the total energy of the full solvated 424D is approximated as the sum of the energies of the solvated substructures; i.e., $E_{HF} = E_{ab}(5H_2O) + E_{cd}(9H_2O) + E_{ef}(4H_2O)$. Thus, for the solvated 424D, this is the full molecular energy used as a standard of comparison against the KEM result.

The calculated results show that the solvent contributions to the total molecular energy are well-represented by the KEM. For example, observing in Table 3 the energy differences $E_{HF} - E_{KEM}$, for eq 1, the absolute magnitude of the energy differences range from a minimum of 0.1835 kcal/mol to a maximum of 23.1151 kcal/mol. These energy differences range from 1.42×10^{-6} to 1.81×10^{-4}%, respectively. Again, these differences are relatively small, as a fraction of the total energy calculated, and thus again speak to the accuracy of the KEM as implemented in eq 1. As before, the results of eq 2 are generally more accurate than those of eq 1. For eq 2, the absolute magnitude of the energy differences range from a minimum of 0.1688 kcal/mol to a maximum of 13.0589 kcal/mol. These energy differences range from 2.56×10^{-6} to 1.02×10^{-4}%, respectively. Except for two cases (B-DNA 1S9D and B-DNA 425D), the eq 2 results are more accurate than the case for eq 1, as would be expected. Although including the solvent molecules in additional solvent kernels decreases the over all accuracy of the method relative to the calculation accuracy without the solvent, the accuracy of the calculations is still good. Thus, the effect of the solvent on the total energy can be simply accounted for in the KEM.

DISCUSSION AND CONCLUSIONS

The DNA molecular systems of this paper were treated within the context of the *ab initio* HF approximation. The basis set used for all cases was a limited basis of Gaussian STO-3G type. A limited basis was chosen to make the energy calculations on full molecular systems (i.e., E_{HF}) as convenient as possible. The numerical values of E_{HF} provided

the standard of comparison for the energy values obtained by the KEM. The comparisons between E_{HF} and E_{KEM} have shown that the KEM can be applied to a wide variety of DNA molecular systems with good accuracy. In particular, such calculation accuracy holds true for A, B, and Z DNA, the three main types of DNA configuration. The most common configuration of DNA, i.e., B DNA, was examined in 10 different molecular systems of variable geometry and magnitude, as judged by the number of atoms in the system, and was in each case found to be described with good accuracy by the KEM. The KEM energies for the Z-DNA hexamer duplex (both with and without the solvent present in Tables 2 and 3, respectively) exhibit the biggest deviations relative to the standard HF full-molecule values. Notice however that even these deviations, measured in units of kcal/mol, in comparison to the molecular energies measured in units of au, are very small. For example, in Table 2, the Z DNA difference magnitude is only 7.98 kcal/mol amounting to a percentage difference of 0.000 078%. In Table 3, the Z DNA difference magnitude is only 13.06 kcal/mol amounting to a percentage difference of 0.0001%. It is clear that in the one case studied, the Z DNA differences are small. They are nonetheless larger than the A and B DNA differences. However, testing with many other Z DNA molecules would undoubtedly show fluctuations of differences such that some of those might be smaller than some of the A and B DNA differences. The point is that there is nothing about the KEM that makes it inherently less applicable to Z DNA than to A and B DNA.

The fundamental physical idea upon which the KEM is based is that the energy of any given kernel is most affected by its own atoms and those of neighboring kernels. A pair of interacting kernels form a double kernel. Perhaps the most important double kernels are those formed of "chemically bonded" single kernels. In previous work with peptides and the protein insulin, the phrase "chemically bonded" was meant to signify being bonded together by covalent bonds. However, here, because of its importance in the study of DNA structures, the meaning of the phrase "chemically bonded" has been expanded to include Watson−Crick hydrogen bonding between base pairs. Good accuracy is obtained from eq 1, i.e., from consideration of only "chemically bonded" double kernels. For the best accuracy, however, all double kernels are calculated and utilized in eq 2.

With the KEM, the fragment calculations are carried out on double and single kernels whose ruptured bonds have been mended by attachment of hydrogen atoms. A satisfactory occurrence in the summation of energies is that the total contribution of hydrogen atoms introduced to saturate the broken bonds tends to 0. This happens because the effect on 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.

The KEM affords the opportunity to make *ab initio* calculations for DNA molecular systems, whose complexity exceeds the capacity of presently available computers and computer programs. The increase in computing time with N (number of atoms) is only a modestly increasing function of N because in the KEM the molecule is not calculated as a whole. It is only the kernels and double kernels that are calculated, and they are chosen to be much smaller than the whole molecule. Moreover, because the method constructs a whole from the sum of parts, it is especially suitable for parallel computation. Altogether, the results of this paper show that KEM makes practical the quantum mechanical study of DNA molecular systems.

ACKNOWLEDGMENT

The crystal structures used in this paper have all been taken from the Protein Data Bank (PDB) and Nucleic Acid Database (NDB). The research reported in this paper was supported by the Office of Naval Research. One of us (L. M.) thanks 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 the NIH for grants (NIGMS MBRS SCORE5 S06GM606654 and RR-03037 the National Center For Research Resources) and the NSF for CREST grant support.

REFERENCES

1. Huang, L., Massa, L., and Karle, J. (2005) Kernel energy method illustrated with peptides, *Int. J. Quantum Chem. 103*, 808−817.
2. Huang, L., Massa, L., and Karle, J. (2005) The kernel energy method: Basis fuctions and quantum methods, *Int. J. Quantum Chem.*, manuscript accepted for publication.
3. Huang, L., Massa, L., and Karle, J. (2005) Kernel energy method: Application to insulin, *Proc. Natl. Acad. Sci. U.S.A. 102*, 12690−12693.
4. Massa, L., Huang, L., and Karle, J. (1995) Quantum crystallography and the use of kernel projector matrices, *Int. J. Quantum Chem., Quantum Chem. Symp. 29*, 371−384.
5. Huang, L., Massa, L., and Karle, J. (1996) Kernel projector matrices for Leu[1]-zervamicin, *Int. J. Quantum Chem., Quantum Chem. Symp. 30*, 1691−1700.
6. Huang, L., Massa, L., and Karle, J. (1998) Kernel projector matrices; application to Leu[1]-zervamicin, in *Encyclopedia of Computational Chemistry* (von Schleyer, P., Ed.) pp 1457−1470, John Wiley and Sons, New York.
7. Karle, J., Huang, L., and Massa, L. (1998) Quantum crystallography, a technique for extending the concept of structure, *Pure Appl. Chem. 70*, 319−324.
8. Huang, L., Massa, L., and Karle, J. (1999) Quantum crystallography, *J. Mol. Struct. 474*, 9−12.
9. Huang, L., Massa, L., and Karle, J. (1999) Quantum crystallography applied to crystalline maleic anhydride, *Int. J. Quantum Chem. 73*, 439−450.
10. Karle, J., Huang, L., and Massa, L. (1999) Quantum crystallography: Features and application, in *Current Challenges on Large Supramolecular Assemblies* (Tsoucaris, G., Ed.) Vol. 519, pp 1−5, NATO Science Series C: Mathematical and Physical Sciences.
11. Huang, L., Massa, J., and Karle, J. (2001) Quantum crystallography, a developing area of computational chemistry extending to macromolecules, *IBM J. Res. Dev. 45*, 409−415.
12. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. (1982) The anatomy of A-, B-, and Z-DNA, *Science 216*, 475−485.
13. DNA structure, http://tigger.uic.edu/classes/phys/phys461/phys450/ANJUM04/.
14. A, B, and Z-DNA's, http://www.albany.edu/∼achm110/abzdna.html.
15. Wahl, M. C., Rao, S. T., and Sundaralingam, M. (1996) Crystal structure of the B-DNA hexamer d(CTCGAG): Model for an A-to-B transition, *Biophys. J. 70*, 2857−2866.
16. Leonard, G. A., Hambley, T. W., and Mcauleyhecht K. (1993) Anthracycline−DNA interactions at unfavourable base-pair triplet-binding sites: Strucutres of d(CGGCCG)/daunomycin and d(TGGCCA)/adriamycin complexes, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 49*, 458−467.

17. Soler-Lopez, M., Malinina, L., Tereshko, V., Zarytova, V., and Subirana, J. A. (2002) Interaction of zinic ions with d(CGCAATTGCG) in a 2.9 Å resolution X-ray structure, *J. Biol. Inorg. Chem. 7*, 533−538.
18. Vargason, J. M., Henderson, K., and Ho, P. S. (2001) A crystallographic map of the transition from B-DNA to A-DNA, *Proc. Natl. Acad. Sci. U.S.A. 98*, 7265−7270.
19. Tari, L. W., and Secco, A. S. (1995) Base-pair opening and spermine binding B-DNA features displayed in the crystal structure of a gal operon fragment: Implications for protein−DNA recognition, *Nucleic Acids Res. 23*, 2065−2073.
20. Valls, N., Uson, I., Gouyette, C., and Subiriana, J. A. (2004) A cubic arrangement of DNA double helices based on nickel−guanine interactions, *J. Am. Chem. Soc. 126*, 7812−7816.
21. Qiu, H., Dewan, J. C., and Seeman, N. C. (1997) A DNA decamer with a sticky end: The crystal structure of d-CGACGATCGT, *J. Mol. Biol. 267*, 881−898.
22. Rozenberg, H., Rabinovich, D., Frolow, F., and Hegde, R. S. (1998) Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets, *Proc. Natl. Acad. Sci. U.S.A. 95*, 15194−15199.
23. Nunn, C. M., and Neidle, S. J. (1995) Sequence-dependent drug binding to the minor groove of DNA: Crystal structure of the DNA dodecamer d(CGCAAATTTGCG)$_2$ complexed with propamidine, *Med. Chem. 38*, 2317−2325.
24. Egli, M., Williams, L. D., Gao, Q., and Rich, A. (1991) Structure of the pure-spermine form of Z-DNA (magnesium free) at 1 Å resolution, *Biochemistry 30*, 11388−11402.
25. Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B., Salisbury, S. A., and Viswamitra, M. A. (1983) Sequence-dependent conformation of an A-DNA double helix. The crystal structure of the octamer d(G-G-T-A-T-A-C-C), *J. Mol. Biol. 166*, 183−201.

BI051655L