Quantum crystallography, a developing area of computational chemistry extending to macromolecules

by L. Huang
L. Massa
J. Karle

We describe the concept of quantum crystallography (QCr) and present examples of its potential as a technique for facilitating computational chemistry, particularly, applications of quantum mechanics. Structural information has been used to facilitate quantum-mechanical calculations for several decades. Recent advances in theory and computational facilities have led to research opportunities that could be considered only in the past several years. We focus on the feasibility of applications of quantum mechanics to macromolecules. The approach used involves the concept of calculations based on fragments of molecules. The method for constructing fragments, their composition, and how they are assembled to form a projector matrix are discussed without the introduction of mathematical detail. Papers that provide the theoretical basis for QCr and our method for making fragment calculations are referenced, and some initial calculations are described here.

Introduction

The term *quantum crystallography* (QCr) refers to the combination of structural, mainly crystallographic, information with quantum-mechanical theory. The objective is to facilitate the theoretical calculations and thereby enhance the information that may be derived from a crystallographic experiment. This concept has a long history and in recent years has been finding increased attention because of the advances in theory and computational techniques. Examples of the additional information that becomes available are more accurate electron density distributions, difference electron densities showing the nature of the bonding, charges on atoms,

0018-8646/01/$5.00 © 2001 IBM

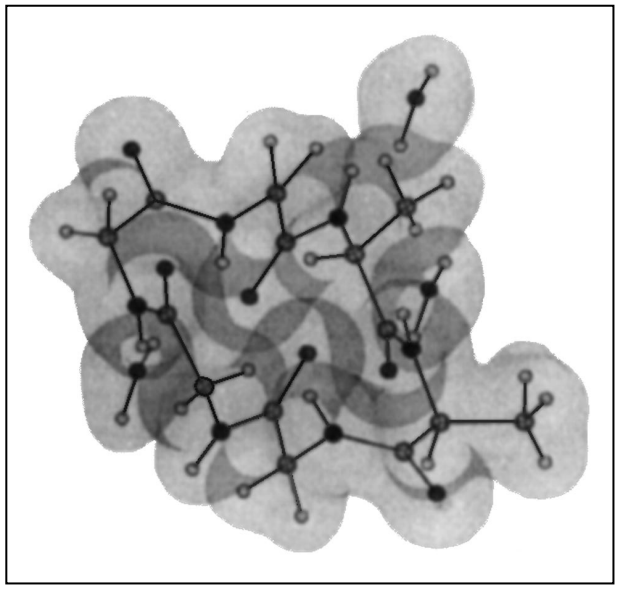

Figure 1

Isodensity surface of 0.023 $e/Å^3$ for the cyclic hexapeptide trihydrate c[Gly-Gly-D-Ala-D-Ala-Gly-Gly] · $3H_2O$, as obtained from a Hartree–Fock calculation. This figure is indistinguishable in appearance from that based on our fragment calculation.

electrostatic potentials, and various energies. The feasibility of carrying this out on macromolecules has already been demonstrated, and is discussed next.

Quantum-mechanical calculations have been made on molecular moieties in a variety of ways. Yang and Lee [1, 2], for example, solved the Kohn–Sham [3] equations for partial structures and obtained a representation of a full molecule by the use of spatial partition functions to weight a sum of partial structures. Later, Yang and Lee [4] generalized the "divide and conquer" density method [2] to one based on the density matrix. Their density matrix for a full molecule is obtained from subsystem contributions in a fashion consistent with Mulliken population analysis. Zero elements of the density matrix also occur in a way that is analogous to the definitions used in an investigation of ours [5] and one by Walker and Mezey [6]. The method of Yang and Lee has been applied in a density-functional context to polypeptides and some proteins [7, 8].

A density-functional/Wannier function formalism was presented by Kohn [9] with an objective similar to that of Lee and Yang [2] but "differs by the central role of the systematic construction of the generalized Wannier function."

Bader and co-workers [10–13] developed a method for splitting the known electron density of a molecule into densities for atomic groupings that have sharply defined borders. The borders are fixed by a condition concerning the vanishing of the density gradient. Such atomic groupings with their known electron densities are used to compose electron density distributions for molecules whose electron densities are unknown. The assumption is made that the electron density distributions of specific moieties change little as the molecules, in which they appear, differ.

Walker and Mezey [6] presented an alternative method for determining the electron density distribution of groups of atoms for use in composing electron density distributions for larger molecules as large as proteins. They make a database of the electron densities of the various atomic groupings required for larger molecules of interest by calculating the electron densities in small molecules having the moieties of interest and, in addition, atomic neighborhoods surrounding the moieties that resemble the neighborhoods that occur in the larger molecules. By use of a rule analogous to that of Mulliken population analysis of charge, Walker and Mezey [6] extracted the electron density of an atomic grouping computed along with the electron density of a suitable small molecule. The electron density for the moiety is stored as numerical information over a three-dimensional grid of points. Such moieties are called "lego" pieces. The electron density for a molecule of interest is obtained by placing each lego piece into its appropriate position and orientation in the molecule and summing the numerical densities associated with all of the lego pieces. Good accuracy has been reported [6], and the calculations, e.g., applications to macromolecules, proceed very rapidly.

Li, Nunes, and Vanderbilt [14] discuss a procedure for obtaining a projector matrix by a method whose computational time increases linearly with the number of atoms in the structure.

The theoretical background for our approach to quantum crystallography may be found in References [5, 15–17] and additional references therein. This paper is devoted to a general description of the concepts involved and some initial applications. The latter point the way to future computational developments and potentially worthwhile types of investigations.

A fragment method

There are several approaches to fragment methods that are suitable for use with macromolecules. In our approach, each fragment of the structure, with known atomic coordinates, is composed of a kernel and a neighborhood. The combined total of the kernels forms the complete structure. The neighborhoods surrounding the kernels include all atoms that are located within some specified distance from all of the atoms in the corresponding kernels. In calculations thus far, the distance has been set at about 5 Å. Beyond that distance,

the interaction of the neighborhood atoms with the kernel atoms would be quite small. In our approach to the calculation of electron density distributions, we generate a matrix of special type, called a projector matrix, which plays a central role in this kind of calculation. The fragment method is used when the calculation of the entire projector matrix at once is prohibitively large.

Two applications, with the objective of testing the fragment method, have been carried out in our laboratory. The substances studied were a cyclic hexapeptide trihydrate, c[Gly-Gly-D-Ala-D-Ala-Gly-Gly] · $3H_2O$ [5], and leu[1]-zervamicin, Ac-Leu-Ile-Gln-Iva-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol [15].

As noted above, the purpose of fragment calculations is to obtain projector matrices when it is not feasible to do so with the entire molecule or when it may turn out to be more efficient to do so. Since a structure of interest is assumed to have known coordinates from, for example, a diffraction experiment, there is no problem in defining a choice of kernels and corresponding neighborhoods. To obtain an electron balance, it may be necessary to attach hydrogen atoms to some of the atoms in a neighborhood. Generally, there is a large variety of ways in which a molecule can be divided into kernels with their corresponding neighborhoods. One rule, however, must be obeyed, which is that all of the atoms in a molecule must be a member of some kernel once and only once. This rule plays a valuable role in defining how the fragment calculations are to be combined in forming a complete matrix that is preliminary to the formation of the desired projector matrix. The contributions from the kernels are included with a weight of 1, whereas the contributions from the neighborhoods are added with a weight of 1/2. Because of the above-mentioned rule concerning membership in a kernel, there are only two circumstances in which a particular pair of kernel and neighborhood atoms can interact, i.e., when one is the kernel atom and the other is the neighborhood atom, and vice versa. Thus, entering each of these two interactions with the weight of 0.5 is quite appropriate.

The fragments for the hydrated cyclic hexapeptide were selected by defining the kernels as the six peptide residues in the ring with each of the three water molecules associated with the appropriate residues, determined by proximity. In this case, the neighborhoods of the kernels were provided by the two peptide residues and associated water molecules, if any, adjoining a particular kernel; e.g., residues 2 and 4 were the neighborhood for residue 3 when treated as a kernel. With this characterization in place, the electron density calculations were carried out.

Isodensity surfaces were calculated for the entire hydrated hexapeptide molecule and also by means of the fragment method. The results were indistinguishable, and hence one figure can suffice for either method of calculation. An example of a calculation at an isodensity surface of 0.023 $e/Å^3$ is given in **Figure 1**. There are, actually, very small differences between the two calculations. In order to obtain a more quantitative insight into the magnitude of the differences, a series of difference isodensity surfaces were calculated at various levels, making it possible to observe the level at which residual differences essentially disappeared, as can be seen in **Figure 2**. The isodensity difference studies on this small polypeptide indicate that fragment calculations have the potential for affording the opportunity to apply quantum mechanics to even larger structures.

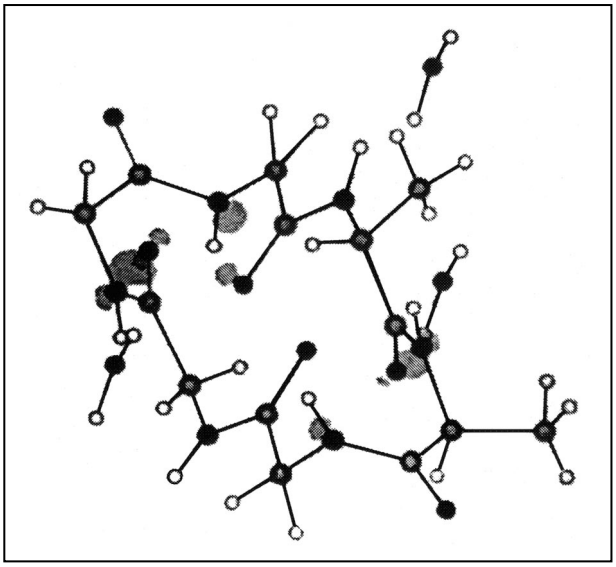

Figure 2

Difference isodensity surface of 3×10^{-3} $e/Å^3$ for the cyclic hexapeptide trihydrate. The small fuzzy regions in the vicinity of the ring represent the difference isodensity surface. They are close to disappearing, enclosing a very small fraction of the molecular volume. A difference isodensity surface at the level of 5×10^{-3} $e/Å^3$ would cause a significant decrease in the fuzzy regions.

Encouraged by these results, we investigated the applicability of the fragment technique with the hexadecapeptide, hydrated leu[1]-zervamicin. As a consequence, it was demonstrated that a projector matrix could be produced from fragment calculations for this structure which is essentially indistinguishable from one obtained from direct calculation with the entire molecule. The input information was a set of atomic coordinates obtained from a crystal structure investigation. The fragments were again composed of kernels and their neighborhoods. In this case, the neighborhoods were chosen for each kernel according to the 5-Å rule. Since the coordinates of the structure are available, it is possible to calculate which atoms would occur within a particular

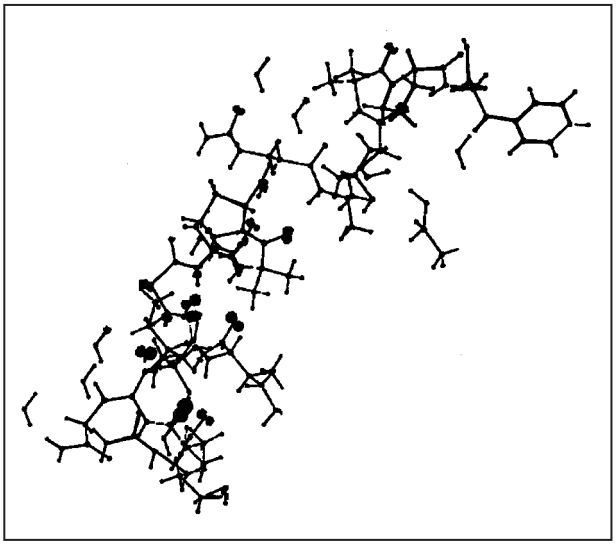

Figure 3

Difference isodensity surface of 1.0×10^{-3} $e/\text{Å}^3$ for the hydrated leu[1]-zervamicin molecule. Small fuzzy regions represent the remaining difference. They involve a small molecular volume and are evidently small in magnitude.

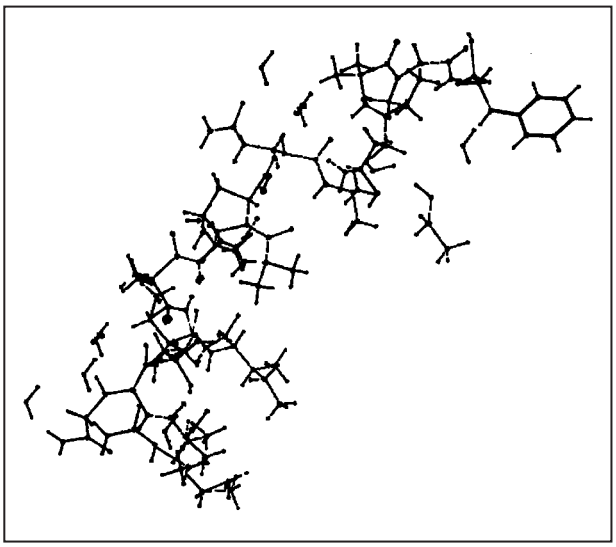

Figure 4

Difference isodensity surface of -1.2×10^{-3} $e/\text{Å}^3$ for the hydrated leu[1]-zervamicin molecule. As in Figure 3, the small fuzzy regions represent the remaining difference, and involve a small molecular volume and are evidently small in magnitude.

chosen distance from each of the atoms in a kernel. In order to effect a pairing of all electrons and an even number of electron pairs, it was necessary to attach hydrogen atoms to atoms belonging to neighborhoods or a moiety that is part of the molecule. This extended the outer limits of the fragment somewhat. Because of the denser packing of peptide residues in leu[1]-zervamicin as compared to the cyclic hexapeptide, more residues were required to form the neighborhoods of each kernel.

The fragment calculations for leu[1]-zervamicin were performed by defining 19 kernels, based on 16 peptide residues, two clusters of water molecules, and a cluster consisting of a water molecule and an ethanol molecule. As in the case for the hexapeptide, isodensity surfaces were calculated for the hydrated leu[1]-zervamicin molecule with the use of a projector matrix generated from the entire molecule and also from the fragments. The two calculations gave electron densities that were quite similar in appearance. In order to obtain a more quantitative insight into the similarity of both types of isodensity calculation, a series of difference isodensity surfaces were calculated in which differences that did not exceed increasingly larger values were omitted. Such a calculation can be used to determine and locate the largest differences between the electron densities. A difference isodensity surface is shown at the values of 1×10^{-3} $e/\text{Å}^3$ in **Figure 3** and -1.2×10^{-3} $e/\text{Å}^3$ in **Figure 4**. It is seen that in rather small regions there are differences as large or larger than the values of the difference isodensities shown. They disappear at slightly larger difference isodensities. The high accuracy of the fragment calculation is again quite clear here and could, in fact, be increased by including more atoms in the neighborhoods of the kernels.

Features of interest in zervamicin

Zervamicin is an antibiotic which is known to transport potassium ions across cell membranes. Experimental observations have indicated that helical peptides, such as zervamicin, which are ionophores, assemble to form channels that allow ions to pass through. An examination of the crystal structures of such membrane-active peptides can reveal the number and packing of the peptides involved in forming a channel and a variety of detailed structural features related to the function of the channels. Crystal-structure investigations of zervamicin indicate that the channels are formed from a cluster of three molecules and that individual molecules can be involved in forming more than one channel. Zervamicin forms a type of helix that is called amphiphilic. 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 the side chain of residue 11, glutamine. This side chain is attached on the hydrophobic side of the peptide. **Figure 5** shows the side chain of glutamine 11 in two different orientations, as

determined from a crystal-structure investigation by Karle et al. [18]. These two orientations are of particular interest because one closes the channel and the other opens it.

In the investigation, it was possible to simulate the passage of a K^+ ion through the channel. In the course of the simulation, it was found that the gate that was formed by the side chain had to first open and then close in order for the K^+ ion to proceed through the channel. Apparently, this structural feature of glutamine 11, as it occurs in the crystal, requires the gate to open and close each time an ion goes through. Experiment shows [19] that a potential is needed for the K^+ ions to be transported. 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.

Some additional evidence for the latter conjecture also comes from a related crystallographic investigation. The crystal structure of antiamoebin closely resembles that of zervamicin. The two substances are chemically similar. In a crystal structure formed jointly with antiamoebin and normal octanol [20], the hydrophobic sides of the antiamoebin molecules share an attraction with the octanol molecules in a fashion that could simulate the attraction within a cell membrane.

A role for quantum crystallography

It would be of interest to examine glutamine 11 and the area around it with the tools of QCr. Toward that objective, we discuss in this section how such an investigation might be carried out. The nature of the calculations for one type of approach is outlined here, with further details given in Reference [17]. An alternative approach may be found in Reference [16], wherein the calculation is simplified by use of the crystallographic data in the form of atomic coordinates. This permits "single-point" quantum-mechanical calculations to be carried out for the determination of various properties of the region of interest.

In Reference [17], it is shown how kernel R_{ik} matrices are computed; the subscript i labels the various numbered kernels, and the subscript k is simply an abbreviation for kernel. It is also shown how to obtain from them their corresponding partial structure factors F_{hik} with and without thermal effects and form from them total structure factors with and without thermal effects. With this accomplished, it is possible to put the experimental and theoretically calculated data on the same scale, remove thermal effects from the experimental data, correct the experimental data for systematic errors, estimate the experimental values for the partial structure factors associated with glutamine 11, and, finally, perform quantum crystallographic calculations on the residue, glutamine 11.

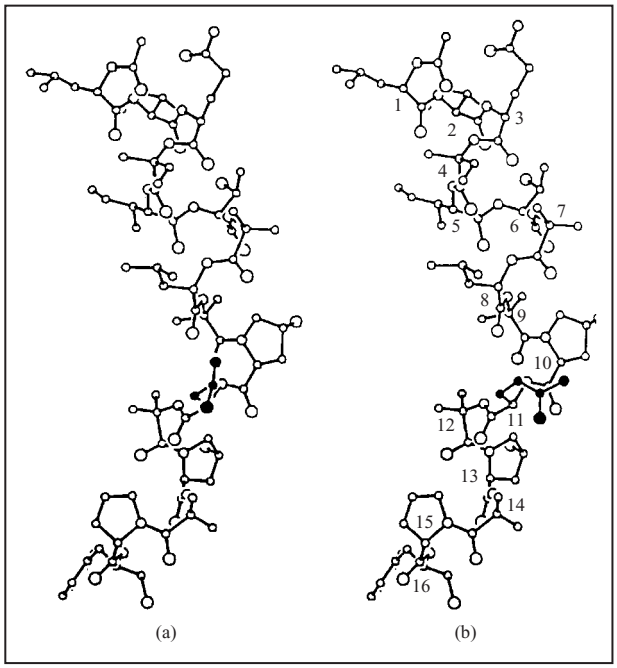

Figure 5

Open (a) and closed (b) forms of the side chain of glutamine 11 in leu[1]-zervamicin.

Comparable calculations which could be of interest concern active sites in protein structures and other regions of macromolecules that have important effects on the functions of these molecules. Of interest would be atomic charges, electron density distributions, electrostatic potentials, certain energies, for example, associated with the attraction of substrates, and the potential for enhancing structural accuracy by quantum-mechanical means.

The treatment of crystallographic information

In this section, we continue to discuss various aspects of handling X-ray diffraction data with the objective of facilitating quantum-mechanical calculations that give information which goes beyond that obtained from the structure analysis of the X-ray data. In our approach to QCr, we have used molecular orbitals and a single-determinant density matrix formalism to develop a quantum-mechanical model. The diffraction data from a crystal of maleic anhydride [16] were used to examine various aspects of the treatment of the experimental data with the objective of optimizing both the experimental information available and the procedures for its use.

One approach has involved the adjustment, by the least-squares method, of the elements of the projector matrix

and some other variables that occur in the quantum-mechanical model in order to optimize the fit between the experimental structure factor magnitudes and the values of those same magnitudes obtained from the quantum-mechanical model. In the application to the crystal of maleic anhydride, it was found to be worthwhile to correct the diffraction data for systematic errors that can arise from gross features in the shape of a crystal and also from characteristics of the diffraction equipment [16]. A statistical method developed from quantum-mechanical theory was found to be quite suitable for making the necessary corrections on the data from a crystal of maleic anhydride. This method appears to have potential for general application and merits further study in this regard.

It was noticed in the course of this investigation that the values of the atomic coordinates determined by the X-ray diffraction experiment were highly insensitive to the changes that were made to the structure factor magnitudes by, for example, corrections for systematic errors. This was probably due to the fact that the coordinates were obtained in a statistical fashion, as a consequence of having information about all of the experimental structure factor magnitudes and their accompanying phase angles. In this case, at least, the corrections for systematic errors resulted in only extremely small changes in the atomic coordinates. This suggested an alternative path for performing QCr which could be more efficient than the method suggested earlier in which the projector matrix was adjusted to effect an optimal match between the calculated and experimental structure factor magnitudes. As mentioned above, the alternative would be to use the crystallographic coordinates in single-point calculations to facilitate the quantum-mechanical calculations of interest.

Since use can be made of quantum-mechanical calculations to perform a geometric optimization that gives values to the coordinates of a structure, the question arises concerning what role the experimental coordinates from crystal structure analysis might play. The following comments address this question:

1. We note that it is necessary to know the crystal structure in order to determine whether the free-molecule approximation is valid. Aside from that, as molecules of interest become larger, even if the free-molecule approximation could be applied, the optimization calculations would soon become prohibitively time-consuming.
2. Crystallography can furnish coordinates and packing when, in the usual case, interaction between molecules can make the free-molecule approximation unacceptable.
3. Crystallography can either validate or throw doubt on the mode or basis set used for the quantum-mechanical calculations.

A summary outline for proceeding with fragment calculations for macromolecules

1. The process starts with, for example, a protein structure that has been determined by use of X-ray diffraction.
2. The structure is then divided into kernels, so chosen that the kernels add up to one complete molecule.
3. Attach to each a neighborhood of atoms that occur about 5 Å or more from any atom in the accompanying kernel. Some additional H atoms or parts of the structure may also have to be attached in order to effect pairing of all electrons and an even number of electron pairs.
4. The structure of the fragment may require geometric optimization by means of quantum-mechanical calculations because it may have been obtained from low-resolution diffraction data. In these circumstances, the optimization, if feasible, would evidently be quite desirable.
5. After optimization, the fragment would be ready for quantum-mechanical calculations of interest.
6. The fragments that would perhaps be of particular interest would be the ones associated with active sites or ones having special influence on the behavior of the protein.

Acknowledgments

L. M. acknowledges an IBM Shared University Research (SUR) grant and a NASA JOVE/JAG grant. L. H. and J. K. wish to acknowledge support by the Office of Naval Research and by Public Health Service Grant GM-30902.

References

1. W. Yang, "Direct Calculation of Electron Density in Density-Functional Theory," *Phys. Rev. Lett.* **66,** 1438–1441 (1991).
2. C. Lee and W. Yang, "The Divide-and-Conquer Density-Functional Approach: Molecular Internal Rotation and Density of States," *J. Chem. Phys.* **96,** 2408–2411 (1992).
3. W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev. A* **140,** 1133–1138 (1965).
4. W. Yang and T.-S. Lee, "A Density-Matrix Divide-and-Conquer Approach for Electronic Structure Calculations of Large Molecules," *J. Chem. Phys.* **103,** 5674–5678 (1995).
5. L. Huang, L. Massa, and J. Karle, "Quantum Crystallography and the Use of Kernel Projector Matrices," *Int. J. Quantum Chem.: Quantum Chem. Symp.* **29,** 371–384 (1995).
6. P. D. Walker and P. G. Mezey, "Molecular Electron Density Lego Approach to Molecule Building," *J. Amer. Chem. Soc.* **115,** 12423–12430 (1993).
7. J. P. Lewis, C. W. Carter, Jr., J. Hermans, W. Pan, T.-S. Lee, and W. Yang, "Active Species for the Ground-State Complex of Cytidine Deaminase: A Linear-Scaling Quantum Mechanical Investigation," *J. Amer. Chem. Soc.* **120,** 5407–5410 (1998).

8. T.-S. Lee, J. P. Lewis, and W. Yang, "Linear-Scaling Quantum Mechanical Calculations of Biological Molecules: The Divide-and-Conquer Approach," *Computational Mater. Sci.* **12,** 259–277 (1998).
9. W. Kohn, "Density-Functional Wannier Function-Theory for Systems of Very Many Atoms," *Chem. Phys. Lett.* **208,** 167–172 (1993).
10. R. F. W. Bader and T. T. Nguyen-Dang, "Quantum-Theory of Atoms in Molecules—Dalton Revisited," *Adv. Quantum Chem.* **14,** 63–124 (1981).
11. R. F. W. Bader, "Atoms in Molecules," *Acc. Chem. Res.* **18,** 9–15 (1985).
12. R. F. W. Bader, M. T. Carroll, J. Chessman, and C. Chang, "Properties of Atoms in Molecules—Atomic Volumes," *J. Amer. Chem. Soc.* **109,** 7968–7979 (1987).
13. R. F. W. Bader, *Atoms in Molecules—A Quantum Theory*, Oxford University Press, Oxford, England, 1990.
14. X.-P. Li, R. W. Nunes, and D. Vanderbilt, "Density-Matrix Electronic-Structure Method with Linear System-Size Scaling," *Phys. Rev. B* **47,** 10891–10894 (1993).
15. L. Huang, L. Massa, and J. Karle, "Kernel Projector Matrices for Leu^1-Zervamicin," *Int. J. Quantum Chem.: Quantum Chem. Symp.* **30,** 1691–1700 (1996).
16. L. Huang, L. Massa, and J. Karle, "Quantum Crystallography Applied to Crystalline Maleic Anhydride," *Int. J. Quantum Chem.* **73,** 439–450 (1999).
17. J. Karle, L. Huang, and L. Massa, "Quantum Crystallography, a Technique for Extending the Concept of Structure," *Pure & Appl. Chem.* **70,** 319–324 (1998).
18. I. L. Karle, J. L. Flippen-Anderson, S. Agarwalla, and P. Balaram, "Conformation of the Flexible Bent Helix of Leu^1-Zervamicin in Crystal C and a Possible Gating Action for Ion Passage," *Biopolymers* **34,** 721–735 (1994).
19. P. Balaram, K. Krishna, M. Sukumar, I. R. Mellor, and M. S. P. Sansom, "The Properties of Ion Channels Formed by Zervamicins," *Eur. Biophys. J.* **21,** 117–128 (1992).
20. I. L. Karle, M. A. Perozzo, V. K. Mishra, and P. Balaram, "Crystal Structure of the Channel Forming Polypeptide Antiamoebin in a Membrane Mimetic Environment," *Proc. Natl. Acad. Sci. USA* **95,** 5501–5504 (1998).

Received June 7, 2000; accepted for publication August 22, 2000

Lulu Huang *Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375 (huang@harker.nrl.navy.mil).* Dr. Huang is a Senior Scientist at the Naval Research Laboratory, Washington, D.C., and is an employee of Geo-Centers, Inc. She received a B.S. degree from the Shanghai Institute of Chemical Technology, and M.A. and Ph.D. degrees from the City University of New York. Her current interests include quantum mechanics and computational chemistry.

Lou Massa *Hunter College and the Graduate School, City University of New York, New York, New York 10021 (lmassa@hejira.hunter.cuny.edu).* Dr. Massa is Professor of Chemistry and Physics at Hunter College and the Graduate School, City University of New York. He received a B.S. degree from Lemoyne College, an M.A. degree from Clarkson University, and a Ph.D. degree from Georgetown University. Subsequently he had a postdoctoral appointment at Brookhaven National Laboratory. He is a Visiting Professor at the Naval Research Laboratory, Washington, D.C. His principal interest is quantum mechanics.

Jerome Karle *Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375 (williams@harker.nrl.navy.mil).* Dr. Karle is Chief Scientist of the Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. He received a B.S. degree at the City College of New York, an M.A. degree at Harvard University, and M.S. and Ph.D. degrees in physical chemistry at the University of Michigan. He was awarded a Nobel Prize in Chemistry in 1985. His current interests include structural and theoretical chemistry.