Journal of Mathematical Chemistry 18 (1995) 141–168

Macromolecular density matrices and electron densities with adjustable nuclear geometries

Paul G. Mezey

Mathematical Chemistry Research Unit,
Department of Chemistry and Department of Mathematics and Statistics,
University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9

Received 12 May 1995

Based on the additive fuzzy electron density fragmentation principle introduced earlier within the *ab initio* Hartree-Fock quantum chemical computational framework, two new methods are introduced for the construction of geometry-adjustable, *ab initio* quality macromolecular electron densities. Both methods are designed for the computation of *ab initio* quality electron densities and other properties for macromolecules of arbitrary reference nuclear geometry, as well as for the rapid computation of approximate electron densities and other molecular properties for nuclear geometries slightly distorted with respect to the reference geometry. This latter feature is expected to improve the description of some of the vibrational and dynamic properties of macromolecules. The first of the two techniques, the Adjustable Local Density Assembler, or ALDA method, generates geometry-adjusted macromolecular electron densities directly, using Mulliken-Mezey fragment density matrices, basis set information, and nuclear coordinates. The method requires an ALDA fragment electron density matrix database. The second technique, the Adjustable Density Matrix Assembler or ADMA method, is introduced for the generation of *ab initio* quality approximate *density matrices* for macromolecules. The method assembles Mulliken-Mezey fragment density matrices designed to fulfill a macromolecular compatibility condition. The ADMA method generates macromolecular density matrices without requiring the computation of a macromolecular wavefunction. The ADMA method allows one to apply most of the density matrix techniques of conventional quantum chemistry to macromolecules such as proteins.

1. Introduction

Conventional applications of the Hartree-Fock-Roothaan-Hall molecular orbital methods [1–4] and more advanced techniques based on them are not ideally suited for the study of macromolecules. Without special selection of the non-negligible molecular integrals arising in the direct application of the molecular orbital methodology within the LCAO framework, the computational complexity for large molecules grows with the fourth power of the number of AO basis functions. Whereas with judicious selection of integrals and other improvements the

fourth power dependence can be reduced, nevertheless, a conventional *ab initio* LCAO computation for a protein is not possible with the supercomputer hardware of today.

One of the motivating factors in the recent development of macromolecular quantum chemical approaches has been the need for reliable local shape analysis and local similarity measures suitable for the quantification of the similarities of local regions of molecules. For small molecules, the global similarity measures based on the Quantum Similarity Measure techniques of Carbó [5–11], the related similarity indices of Richards [12–14], the Fourier analysis methods of Bywater [15,16], the momentum space electron density analysis of Cooper and Allan [17,18], the method of Cioslowski [19–21] and the algebraic-topological Shape Group Method of Mezey [22–28] provide the basis for a variety of computational approaches. The local similarity measures based on the algebraic shape groups [22–25], as applied for electron density fragments, complement the quantum similarity measures proposed by Carbó [5–11]. However, until recently, the local similarity analysis of *macromolecules* has been hampered by the high degree of complexity and the associated difficulties involved in the conventional computation of reasonably accurate macromolecular electron densities.

A natural approach to the treatment of complex problems involves two steps: (i) the decomposition of the problem into smaller ones which can be solved, (ii) followed by an assembly of the partial solutions into a global solution of the original problem. The importance of this general principle for the quantum chemical description of macromolecules has been recognized early, and a variety of computational approaches have been proposed following the general scheme. One early approach of Christoffersen and Maggiora [29–31] was based on linear combinations of fragment orbitals where the approximations of the MOs and total energies of large systems were obtained within a framework based on orbitals of smaller molecular fragments.

The general principle of decomposition followed by an additive assembly of fuzzy molecular components appears to hold the key to macromolecular quantum chemistry. In this contribution we describe two new methods for the construction of macromolecular representations of electronic densities, based on mutually interpenetrating, fuzzy electron density fragments. The relevant background and earlier, related approaches are briefly reviewed below.

Applying a fuzzy decomposition method to molecular electron densities and constructing representations of fuzzy molecular fragments have been proposed as tools for local shape analysis of large molecules [32]. In this scheme, mutually interpenetrating isoproperty surfaces define truncation patterns that can be used for both the definition of molecular fragments and for the actual shape analysis of the fuzzy electron density of such fragments. For detailed shape analysis of both complete molecules and molecular fragments, the Shape Group Methods have been used [22–28]. Another related technique is the Density Domain approach, proposed for the study of fuzzy, mutually interpenetrating molecular electron density

fragments within molecules, where the "corresponding concept of chemical bonding is three-dimensional and refers to parts of the molecular body" [33]. One of the goals of the local shape analysis of molecular electron density fragments was the development of molecular similarity measures. High degree of similarity between local, fuzzy electron density domains in different molecules indicate similar chemical properties [34], and also serve as justification [35] for using local, fuzzy, mutually interpenetrating electron density fragments for the construction of electron densities for large molecules.

Recently, an independent, density functional approach has been proposed by Yang [36,37] for a fuzzy fragmentation of electron densities of large molecules, where local Hamiltonians and local electron density contributions are defined within an iterative scheme. Mutual compatibility between fragment densities within a given molecular environment and self-consistency between the electron density fragments and the local potential needed for their determination require an iterative solution.

The earlier fuzzy electron density fragment models of the Density Domain [33] and fragment shape analysis [32] approaches were formulated within the Hartree-Fock-Roothaan-Hall LCAO molecular orbital *ab initio* framework. Several methods were proposed based on this fuzzy fragment model; the essence of these techniques can be described using the conventional density matrix approach. These MO-based, additive, fuzzy density fragmentation methods avoid artificial fragment boundaries and provide *local molecular fragments fully analogous to complete molecules*. These techniques allow the direct application of rigorous shape analysis methods, such as the Shape Group Methods [22–28], developed for complete molecules and suggested for molecular fragments [32,33].

Take a conventional Hartree-Fock-Roothaan-Hall SCF LCAO *ab initio* representation of a molecular wavefunction with respect to some fixed nuclear arrangement K. The electronic density $\rho(\mathbf{r})$ of the molecule, a function of the three-dimensional position variable $\mathbf{r}$, is defined in terms of a set of n atomic orbitals $\varphi_i(\mathbf{r})$, $i = 1, 2, \ldots, n$. The $n \times n$ dimensional density matrix $\mathbf{P}$ can be determined using the coefficients of atomic orbitals in the occupied molecular orbitals, and the electronic density $\rho(\mathbf{r})$ of the molecule can be written as

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \,. \tag{1}$$

The electronic charge cloud $\rho(\mathbf{r})$ provides a detailed representation of the shape of the fuzzy "body" of the molecule. It is natural to require that the local shape properties of molecular fragments are also described within the same framework. Whereas the resulting fuzzy density fragments usually do not exist as independent entities, nevertheless, the consistency of complete molecule and molecule fragment representations is advantageous, since identical shape analysis methods are applicable.

For the purposes of diagnosing the distorting effects of molecular surroundings on the electronic density within a given molecular moiety, a non-additive, fuzzy "pseudo-density" was proposed by Walker [38]. The "pseudo-densities" assigned to a given collection s of atomic nuclei present in different molecules can be compared and the influence of the different molecular surroundings can be diagnosed. The pseudo-density matrix ${}^*\mathbf{P}^s$ of a formal molecular moiety containing a subset s of the nuclei of a molecule is defined by

$$\begin{aligned} {}^*P^s_{ij} &= P_{ij} \quad \text{if AO } \varphi_i(\mathbf{r}) \text{ or } \varphi_j(\mathbf{r}) \text{ is centered on a nucleus} \\ &\qquad \text{that belongs to the subset } s, \\ &= 0 \quad \text{otherwise}. \end{aligned} \tag{2}$$

The resulting pseudo-density ${}^*\rho^s(\mathbf{r})$ for the formal molecular moiety involving the subset s of the nuclei is calculated as

$${}^*\rho^s(\mathbf{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} {}^*P^s_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) . \tag{3}$$

Walker's pseudo-density ${}^*\rho^s(\mathbf{r})$ of a formal molecular moiety involves an enhanced contribution from the surroundings of the local molecular neighborhood, and it is a sensitive diagnostic tool for the detection of shape differences induced by conformational or other changes at different locations within the molecules [38]. The pseudo-density ${}^*\rho^s(\mathbf{r})$ exaggerates the shape differences and ${}^*\rho^s(\mathbf{r})$ is distorted when compared to the actual local shape of the molecule; nevertheless, the detection of shape changes is enhanced. These pseudo-densities are not additive and are not intended for the local description of the actual shapes of molecules; if pseudo-densities for several nuclear families s of a given molecule are determined, then these pseudo-densities ${}^*\rho^s(\mathbf{r})$ do not add up to the molecular density. Their role is in the diagnosis of shape changes.

It is possible, however, to define additive fuzzy density contributions to the molecular electron density which are representatives of local molecular shapes and, by their additivity, are also suitable for the construction of *ab initio* quality electron densities for large molecules [38–46].

A family of methods designed for these tasks is based on the *additive fuzzy density fragmentation (AFDF) principle* of Mezey [27,28,39]. Both additivity and fuzziness appear essential. Electron densities of complete molecules have no discontinuous features, they have no formal boundary surfaces; similarly, the additive fuzzy electron density fragments have no discontinuous features or boundary surfaces. The additive, fuzzy electron density fragmentation methods are motivated by a preference for a choice of representations of molecular fragments which are fully analogous with the representations of complete molecules [27,28]. These methods are less sensitive for diagnostic purposes than the pseudo-density technique of Walker, however, they provide faithful shape representations and also serve as the basis for building *ab initio* quality electron densities for large molecules.

The simplest implementation of the additive fuzzy electron density fragmentation principle for actual density fragmentation, the Mulliken-Mezey method, was proposed by the author, however, its origins can be found in Mulliken's population analysis [47,48]. This scheme can be regarded as an "atom-group population analysis without integration". Although more advanced AFDF methods have also been developed [27,28], the Mulliken-Mezey scheme is sufficient to provide *ab initio* quality results. Formulated within the framework of the Hartree-Fock-Roothaan-Hall molecular orbital *ab initio* method, the additive fuzzy fragmentation method was first implemented as a tool for building electron densities from fragment densities [39]; generalized versions of this scheme are described in refs. [27,28].

The Mulliken-Mezey scheme is the basis of the *Molecular Electron Density Lego Assembler* (MEDLA) method of Walker and Mezey [38,39,40–43], as well as for the *Adjustable Local Density Assembler* (ALDA) and the *Adjustable Density Matrix Assembler* (ADMA) methods of Mezey [44–46]. The MEDLA method has provided the first possibility for the calculation of *ab initio* quality electron densities for proteins [38,41,42]. The ALDA method [44,46] generates macromolecular electron densities that can be adjusted for small nuclear geometry changes, whereas the ADMA method [45,46] generates macromolecular density matrices, suitable for the computation of macromolecular electron densities as well as other molecular properties.

According to this fragmentation scheme, the set of nuclei of a molecule M is divided into m mutually exclusive groups, denoted by

$$f_1, f_2, \ldots, f_k, \ldots f_m . \tag{4}$$

These *nuclear families* serve as AO reference locations when generating the corresponding m *density fragments,*

$$F_1, F_2, \ldots, F_k, \ldots F_m , \tag{5}$$

of *fragment density functions*

$$\rho^1(\mathbf{r}), \rho^2(\mathbf{r}), \ldots, \rho^k(\mathbf{r}), \ldots \rho^m(\mathbf{r}) , \tag{6}$$

defined in terms of the AO set of the molecule M and the family of *fragment density matrices*

$$\mathbf{P}^1, \mathbf{P}^2, \ldots, \mathbf{P}^k, \ldots \mathbf{P}^m , \tag{7}$$

respectively.

An additive fuzzy density fragment, AFDF $\rho^k(\mathbf{r})$ of the electron density $\rho(\mathbf{r})$ of molecule M is specified by an arbitrary subset k of nuclei and their "share" $\mathbf{P}^k$ of the density matrix $\mathbf{P}$ of the molecule. There is no inherent restriction on the choices of nuclear families; in particular, a fuzzy density fragment $\rho^k(\mathbf{r})$ does not have to refer to a set of nuclei which are interrelated by formal bonds. In practice, however, it is advantageous to select nuclear families where the nuclei within a family are

near one another. Based on the simplest version of the additive fuzzy fragmentation method, the kth fuzzy electron density fragment $\rho^k(\mathbf{r})$ is calculated in terms of Mezey's *additive fragment density matrix, AFDM* $\mathbf{P}^k$, defined as follows:

$$\begin{aligned} P^k_{ij} &= P_{ij} && \text{if both } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ are AO's centered on nuclei of the } k\text{th fragment,} \\ &= 0.5P_{ij} && \text{if only one of } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ is centered on a nucleus of fragment } k, \\ &= 0 && \text{otherwise.} \end{aligned} \tag{8}$$

In this definition the motivating influence of Mulliken's population analysis method is clearly recognizable.

Both the density matrix $\mathbf{P}$ of the complete molecule, and the additive fragment density matrix, AFDM $\mathbf{P}^k$ of the kth fragment have the same $n \times n$ dimensions.

In terms of the full AO set of the molecule and the fragment density matrix $\mathbf{P}^k$, the electron density of Mezey's kth *additive fuzzy density fragment, AFDF* $\rho^k(\mathbf{r})$ is defined as

$$\rho^k(\mathbf{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} P^k_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) . \tag{9}$$

If the nuclear families $f_1, f_2, \ldots, f_k, \ldots f_m$ are mutually exclusive, and if they collectively contain all the nuclei of molecule M, then eq. (8) defining the matrix elements P^k_{ij} implies that the sum of the fragment density matrices $\mathbf{P}^k$ is equal to the density matrix $\mathbf{P}$ of molecule M:

$$P_{ij} = \sum_{k=1}^{m} P^k_{ij} \tag{10}$$

and

$$\mathbf{P} = \sum_{k=1}^{m} \mathbf{P}^k . \tag{11}$$

That is, these fragment density matrices are additive. Furthermore, the *linearity* of the electron density expressions (1) and (9) in the matrix elements P_{ij} and P^k_{ij} of the molecular density matrix $\mathbf{P}$ and fragment density matrices $\mathbf{P}^k$ implies that the sum of fragment densities $\rho^k(\mathbf{r})$ is equal to the density $\rho(\mathbf{r})$ of molecule M:

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r}) . \tag{12}$$

Consequently, at any given *ab initio* HF-LCAO level, the Mulliken-Mezey electron density decomposition scheme is an exactly additive, fuzzy electron density fragmentation scheme.

Whereas the additive fuzzy density fragmentation scheme and its generalized versions [27,28] provide new approaches to local shape analysis [46], the scheme has been used mostly in new algorithms designed for building reasonably accurate

electron densities for large molecules. For many of these large molecules, a conventional *ab initio* electron density calculation is impossible at present using current computer hardware. One important feature of the additive fuzzy density fragmentation scheme is the fact that building macromolecular electron densities from such fragments requires computational effort that grows only linearly with molecular size. This is an important advantage over more conventional quantum chemical methods, which have cubic or quartic size dependence.

The Mulliken-Mezey additive fuzzy electron density fragmentation scheme, eqs. (8)–(12), is the basis of the Molecular Electron Density Lego Assembler (MEDLA) method of Walker and Mezey [39,40]. This method is the simplest of the techniques designed to build *ab initio* quality electron densities for large "target" molecules from "custom-made" fragments obtained from *ab initio* HF-LCAO electron density calculations for smaller "parent" molecules, M_1, M_2, ..., M_k, ... M_m, artificially distorted to match the local nuclear arrangement and surroundings in the target molecule.

The MEDLA technique uses a numerical electron density MEDLA databank, containing pre-calculated electron density fragments obtained from calculations of smaller "parent" molecules containing the "custom-made" nuclear geometry required for the fragment and a sufficiently large molecular neighborhood surrounding these nuclei, ensuring that the local interactions affecting the fragment within the target molecule are well reproduced within the smaller parent molecule. For example, an NH_2 group may occur with many different nuclear geometries and within many different local surroundings; accordingly, the fragment density of the NH_2 group is stored in several "custom-made" versions in the MEDLA database, each obtained from a different parent molecule. In the MEDLA database used for the first *ab initio* quality electron density calculations for proteins [38,41,42], each fuzzy MEDLA fragment has been previously obtained from a 6-31G** *ab initio* calculation for a small, "custom-made" parent molecule. Detailed tests have shown that the MEDLA electron densities are more accurate than the standard 3-21G *ab initio* results (using the 6-31G** basis set results as benchmark), and are virtually indistinguishable from the results of standard *ab initio* 6-31G** *ab initio* calculations. The additive, fuzzy electron density fragmentation method, as implemented within the Walker-Mezey MEDLA scheme, generates *ab initio* quality electron densities for macromolecules.

2. The adjustable local density assembler (ALDA) method for the computation of geometry-adjusted macromolecular electron densities

The additive fuzzy electron density fragmentation method, in particular, its simplest realization, the Mulliken-Mezey fragmentation scheme given by eqs. (8)–(12), can be used for the computation of macromolecular electron densities without relying on a numerical electron density fragment database.

A direct application of the additive fuzzy density fragmentation (AFDF) principle is the basis of the *Adjustable Local Density Assembler* (ALDA) method [44,46]. The AFDF scheme is used only to generate the fragment density matrices AFDM, and the actual fragment density contributions are computed only when they are needed. No numerical electron density database is generated, hence, there is no need for the storage of electron density values at several million grid points for each fragment. Instead, the ALDA method uses a much smaller ALDA database that stores the actual fragment density matrix elements for each AFDM $\mathbf{P}^k$, as well as the associated nuclear geometry and basis set information. Evidently, this requires much less memory than a MEDLA database generating comparable electron densities.

The actual calculation of ALDA electron densities for a macromolecule requires the evaluation of the sum of eq. (9) for each fragment, AFDF $\rho^k(\mathbf{r})$. In principle, this evaluation is required for each point $\mathbf{r}$ of the space for the given target molecule, however, the fragment densities fall off rapidly with the distance from the nearest nucleus of the nuclear family f_k. Consequently, in practice, a cutoff for density contributions implies a distance cutoff, similar to that used in the construction of the MEDLA fragments. That is, the fragment densities are generated "on the spot", for each location within the target molecule, using fragment density matrix information and local basis set and nuclear geometry information, all stored in the ALDA databank.

The fragment densities are added up at any desired set of grid points according to eq. (12), resulting in the ALDA electron density $\rho(\mathbf{r})$ of the target macromolecule. If the same set of grid points and the same set of parent molecules $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_k, \ldots \mathbf{M}_m$ are used, then the ALDA method gives results exactly equivalent to the MEDLA results.

In the ALDA method actual density evaluations are needed instead of just looking up numerical density values in a databank. Consequently, the ALDA method is slower than the MEDLA method. For each fragment, the time requirement of the ALDA method is quadratic with respect to the number of atomic orbitals contained in the parent molecule of the fragment. However, this number is limited by the feasible size of parent molecules, hence there is a constant upper bound for the actual time requirement for each fragment. The computer time requirement is determined by the number of fragments. The ALDA method is also linear in the number of fragments, consequently, the overall computer time requirement of the ALDA method grows linearly with the molecular size.

The disadvantage of the slower, but still linear, performance of the ALDA method is compensated by several advantages.

(a) The ALDA database is smaller then a MEDLA database, since the ALDA database contains only fragment density matrices, basis set information, and nuclear coordinates for the parent molecules $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_k, \ldots \mathbf{M}_m$.

(b) In the computation of an ALDA numerical electron density of a target macromolecule, arbitrary grid types can be used, for example, one may use a more

detailed grid near the nuclei, or grids with enhanced resolution at some interesting location of the target molecule. This feature has importance in the calculation of molecular properties strongly dependent on the electron density near the nuclei, and in local shape analysis of active sites of macromolecules, such as the pocket regions of enzymes.

(c) Another important advantage is a versatility in the rapid, approximate computation of macromolecular electron densities for nuclear arrangements slightly distorted with respect to the arrangements found in the ALDA database. Although it is always possible to generate a new, "custom-made" fragment density matrix for any given nuclear geometry, this step is not always required if an approximate electron density of the macromolecule is sufficient. If the nuclear geometry for a fragment already existing in the ALDA database is similar enough to the nuclear geometry of the required fragment, then a good approximation can be obtained using the same fragment density matrix. By simply changing the nuclear locations for the AO basis functions, and using the same fragment density matrix, small deviations from the exact fragment geometries can be treated in a simple and efficient manner.

As long as the nuclear geometry variations are small, the electronic density "follows" closely the nuclear geometry variations. The inherent nonlinearity of the density deformations associated with small changes in nuclear geometries is well represented by the ALDA method. For a small adjustment of the nuclear geometry, an important contribution to the corresponding adjustment of the electronic density can be approximated by taking the same fragment density matrix, AFDM $\mathbf{P}^k$, and using it with the same set of atomic orbitals located at a set of slightly displaced nuclear locations.

The ALDA method is suitable for the rapid calculation of approximate electron densities for any small range of nuclear geometries of a macromolecule. This adjustibility of the nuclear geometry appears useful in several applications. Small amplitude vibrations involve such minor geometry changes. Adjustable macromolecular electron densities are likely to be useful in the structure refinement process of X-ray structure determination. Another area is the study of protein folding processes. These processes usually involve large geometry changes, however, the ALDA method may provide a tool for breaking up such folding paths to segments where each segment can be treated using simple geometry adjusted ALDA electron densities. Some of the related applications of the ALDA method will be discussed elsewhere [49].

3. The adjustable density matrix assembler (ADMA) method for the generation of macromolecular density matrices

Various implementations of the additive fragment density matrix approach [27,28], including the simplest, Mulliken-Mezey scheme, eqs. (8) and (10), can also

be used for the generation of density matrices for the target macromolecules. A direct, algebraic application of the additive fuzzy electron density fragmentation principle, based on additive fragment density matrices, AFDMs, defined within a consistent framework of AO representations for the fragments, is the basis of a method for the construction of *ab initio* quality approximate density matrices for macromolecules.

The approach taking advantage of this possibility offered by the fragmentation scheme is referred to as the *Adjustable Density Matrix Assembler* (ADMA) technique [45,46]. The ADMA technique provides a link between the additive fuzzy fragmentation approach and mainstream quantum chemistry. If a macromolecular density matrix is available, then most of the routine quantum chemical computational techniques, including expectation value computations for various property evaluations are applicable.

The actual ADMA macromolecular density matrix constructed from the fragment density matrices represents the same level of accuracy as the MEDLA and ALDA methods. In particular, the ADMA method reproduces the effects of interactions between local fragment representations to the same level of accuracy as the MEDLA and ALDA methods. The ADMA density matrix technique also has provisions for the adjustability of the calculated electron density with respect to small nuclear geometry changes of the macromolecule, a feature similar to that of the ALDA method.

The ADMA macromolecular density matrix $\mathbf{P}$ is obtained by combining appropriately defined, *mutually compatible*, additive fragment density matrices (MC-AFDM) $\mathbf{P}^k$. Mutual compatibility involves two conditions:

(a) AO basis set orientation constraints,
(b) fragment choices fulfilling a compatible target - parent fragmentation condition.

The ADMA method uses a fragment density matrix database, similar to that of the ALDA method, however, these fragment density matrices fulfill the second of the above two compatibility conditions. By a suitable transformation, the fragment density matrices can always be converted to physically equivalent fragment density matrices defined with respect to AO basis sets fulfilling condition (a).

3.1. BASIS SET ORIENTATION CONSTRAINT

The compatibility between the atomic orbital basis functions of the fragment density matrices and those of the final, macromolecular density matrix is a natural requirement. The final, macromolecular density matrix must refer to a single (although rather large) set of basis orbitals. This condition requires that all the fragment density matrices should refer to local coordinate systems where the coordinate axes are oriented the same way as the reference axes of a common, macromolecular coordinate system.

If the fragment density matrices stored in the ADMA density matrix database refer to AO basis sets defined within local coordinate systems with different orientations, then local coordinate transformations can be carried out on each fragment density matrix, changing the orientations of atomic orbitals to those in the common, macromolecular coordinate system.

Consider the kth fragment density matrix, obtained from an *ab initio* calculation for the parent molecule $\mathbf{M}_k$ within a local coordinate system. Vector $\varphi^{(k)}(\mathbf{r})$ represents the set of atomic orbitals for the parent molecule $\mathbf{M}_k$ where the orientations of AOs are chosen with respect to this local coordinate system. Accordingly, the $\mathbf{P}^k(\varphi)$ notation is used for the kth fragment density matrix defined in the local coordinate system. Vector $\psi^{(k)}(\mathbf{r})$ represents the same sequence of atomic orbitals at the same nuclear centers in local coordinate systems with axes aligned with the axes of the common, macromolecular coordinate system. An orthogonal matrix transformation $\mathbf{T}^{(k)}$ interrelates these two representations:

$$\psi^{(k)}(\mathbf{r}) = \mathbf{T}^{(k)}\varphi^{(k)}(\mathbf{r}) . \tag{13}$$

Matrix $\mathbf{T}^{(k)}$ is block-diagonal, assembled from the one- dimensional identity matrix for each of the s-orbitals, the ordinary three-dimensional rotation matrix for each triple of p- orbitals, the standard five-dimensional conversion matrix for each set of five orthonormalized d-orbitals, the seven-dimensional conversion matrix for each set of seven orthonormalized f-orbitals, and so on. If non-orthonormal AOs, for example, sets of six non-orthonormal d-functions are used, then an appropriately modified transformation matrix $\mathbf{T}^{(k)}$ is used.

A simple similarity transformation,

$$\mathbf{P}^k = \mathbf{T}^{(k)}\mathbf{P}^k(\varphi)\mathbf{T}'^{(k)} , \tag{14}$$

is used to convert the local representation $\mathbf{P}^k(\varphi)$ of the kth fragment density matrix into the actual fragment density matrix $\mathbf{P}^k$ used in the process of building the macromolecular ADMA density matrix $\mathbf{P}$.

3.2. COMPATIBLE TARGET-PARENT FRAGMENTATION CONDITION

The Mulliken-Mezey density matrix fragmentation scheme and the alternative, generalized schemes [27,28] are defined in terms of nuclear families $f_1, f_2, \ldots, f_k, \ldots f_m$. The ADMA method uses a mutually consistent choice of nuclear families f_k for the fragmentation of the large target molecule M and for the fragmentation of the parent molecules $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_k, \ldots \mathbf{M}_m$ used to generate the fragment density matrices $\mathbf{P}^k$.

The following condition must be fulfilled:

> *If the nuclei of the target molecule* M *are classified into* m *families, then each parent molecule* $\mathbf{M}_k$ *may contain only complete nuclear families* $f_{k'}$ *from the target molecule* M.

A parent molecule M_k either contains a given nuclear family $f_{k'}$ in full as part of the surroundings for the actual nuclear set f_k of the fragment density matrix $\mathbf{P}^k$, or M_k does not contain any part of the nuclear family $f_{k'}$. The only exception to this rule are some peripheral H nuclei (or, possibly other nuclei) used to tie off dangling bonds in parent molecule M_k, where these extra nuclei are at large distances from the actual nuclear set f_k of the fragment density matrix $\mathbf{P}^k$. By chance, such a peripheral nucleus might also appear at the same location as a part of another nuclear family.

In both the target molecule M and the parent molecule M_k, the nuclear family f_k is usually surrounded by several other nuclear families $f_{k'}$, where (within the Mulliken-Mezey scheme) approximately half of the interaction components of electron densities associated with orbitals of these latter families, on the one hand, and with orbitals of the central family f_k of nuclei, on the other hand, are assigned to fragment F_k.

As implied by the definition of matrix elements P^k_{ij}, interaction matrix elements for orbital pairs associated with the nuclear family f_k of the parent molecule M_k contribute fully to the kth fragment density matrix $\mathbf{P}^k$, generating the fragment density matrix elements $P^k_{ij} = P_{ij}$ where both $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are AOs centered on nuclei of the nuclear family f_k. The additional nuclear families $f_{k'}$ occurring in parent molecule M_k, the corresponding fragment density matrices $\mathbf{P}^{k'}$, and the electron density fragments associated with them have their own parent molecules $M_{k'}$. Nevertheless, the consistency condition of the construction principle for building proper macromolecular density matrices requires that any such nuclear family $f_{k'}$ appears in full or not at all within the surroundings of nuclear family f_k in any parent molecule M_k.

The "dangling bonds", formally belonging to the peripheral nuclei of nuclear families $f_{k'}$ representing surroundings for nuclear family f_k in the parent molecule M_k, are attached to H atoms, or, possibly, to other small functional groups. More complicated groups may be needed if conjugative or other interactions are to be extended beyond the usual "coordination shell" provided by the nuclear families $f_{k'}$ surrounding the actual nuclear family f_k of the fragment density matrix $\mathbf{P}^k$.

If the centrally located nuclear family f_k of parent molecule M_k is surrounded by a sufficiently thick layer of additional nuclear families $f_{k'}$, then the additional H (or other) nuclei are positioned far from the nuclei of the actual family f_k, and the density matrix elements between the orbitals of these additional atoms and the orbitals centered on nuclei of central family f_k are negligibly small. Consequently, within the ADMA method only those density matrix elements $\mathbf{P}^k_{ij}$ of each parent molecule M_k are involved in the construction of the final, macromolecular density matrix P, which

(a) fulfill the selection condition in the defining equation (8) (or the alternative, more general conditions of refs. [27,28]) of fragment density matrix $\mathbf{P}^k$, and
(b) do not involve the peripheral extra H (or other) nuclei of the parent molecules.

Evidently, such consistent choices of nuclear families f_k and the corresponding choices of parent molecules M_k can always be obtained for any macromolecule M.

3.3. THE MAIN STEPS OF THE ADMA MACROMOLECULAR DENSITY MATRIX METHOD

The application of the ADMA method for an actual macromolecule involves several steps, listed below.

(i) Identify nuclear families $f_1, f_2, \ldots, f_k, \ldots f_m$ of the target macromolecule M, and denote the number of AOs in these families by

$$n_1, n_2, \ldots, n_k, \ldots n_m , \tag{15}$$

respectively. Define

$$n_0 = 0 , \tag{16}$$

and define an index

$$f(k,i) = i + \sum_{v=0}^{k-1} n_v . \tag{17}$$

Within the macromolecular density matrix $\mathbf{P}$, $f(k,i)$ is the serial index assigned to basis orbital $\varphi_i(\mathbf{r})$ of nuclear family f_k.

(ii) Unless appropriate fragment density matrices are already available in the ADMA database, define parent molecules $M_1, M_2, \ldots, M_k, \ldots M_m$. Each parent molecule M_k contains the corresponding central nuclear family f_k, additional nuclear families $f_{k'}$ surrounding family f_k in a formal "coordination shell", as well as additional, peripheral nuclei attached to dangling bonds linked to some nuclei in families $f_{k'}$. The nuclear families $f_{k'}$ and the associated AOs within this coordination shell have the role of reproducing all local interactions of the electron density dominated by nuclear family f_k with the surrounding electron density cloud within the target molecule. Note that a given nuclear family f_k occurs in precisely one parent molecule M_k as the central family, however, the same family f_k may occur in several other parent molecules $M_{k'}$ as part of their "coordination shell", reproducing local interactions for the corresponding central nuclear families $f_{k'}$.

(iii) Introduce consistent labeling of nuclear families within the target macromolecule and the various parent molecules $M_1, M_2, \ldots, M_k, \ldots M_m$.

For each pair $(k,k'), k,k' = 1,2,\ldots,m$, define

$$c_{k'k} = \begin{cases} 1 & \text{if nuclear family } f_{k'} \text{ contributes to parent molecule } M_k , \\ 0 & \text{otherwise} , \end{cases} \tag{18}$$

and set

$$c_{0k} = 0, \quad \text{for each } k = 1, 2, ..., m. \tag{19}$$

(iv) Carry out *ab initio* computations for each parent molecule M_1, M_2, ..., M_k, ... M_m, and determine the fragment density matrices $\mathbf{P}^1$, $\mathbf{P}^2$, ..., $\mathbf{P}^k$, ... $\mathbf{P}^m$, unless a suitable fragment density matrix is already available in the ADMA databank.

(v) If required, transform the AO basis of each fragment density matrix $\mathbf{P}^k$ to an AO basis defined with respect coordinate axes parallel to the axes of the common, macromolecular coordinate system, using eq. (14) and the appropriate transformation matrix $\mathbf{T}^{(k)}$.

(vi) If required, carry out permutations of the AO basis of each fragment density matrix $\mathbf{P}^k$ and the corresponding row and column permutations of the fragment density matrix $\mathbf{P}^k$ itself in order to generate a block structure for $\mathbf{P}^k$ where

(a) the blocks correspond to AOs associated to specific nuclear families $f_{k'}$,
(b) within all fragment density matrices $\mathbf{P}^k$ all blocks corresponding to a given nuclear family $f_{k'}$ follow the same ordering of the AOs centered on the nuclei of family $f_{k'}$,
(c) the block-row and block-column indices in each fragment density matrix $\mathbf{P}^k$ follow a monotonically increasing subsequence from the sequence of nuclear families $f_1, f_2, \ldots, f_k, \ldots f_m$ of the target macromolecule M, and
(d) the AOs centered on the additional nuclei used to "tie off" peripheral "dangling bonds" in each parent molecule M_k are listed at the end of the sequence of AOs for each fragment density matrix $\mathbf{P}^k$.

Step (vi) may be rendered redundant by designing the parent molecule representations in conformity with these conditions; if, however, the ADMA databank already contains a suitable fragment density matrix generated for a different but locally similar macromolecule, then step (vi) ensures that this fragment density matrix can be "recycled" and used again for the current calculation.

(vii) Expand each fragment density matrix $\mathbf{P}^k$ to a fragment density matrix $\mathbf{P}^k(\mathrm{M})$ of the target macromolecule M, by inserting blank rows and blank columns of blocks corresponding to AO sets of all nuclear families $f_{k''}$ *not contributing* to the fragment density matrix $\mathbf{P}^k$. This is only a symbolic transformation, since in practice, a simple index transformation can be used to identify the x, y index pair of element P_{xy} of the macromolecular density matrix $\mathbf{P}$ where a given element P^k_{ij} of a fragment density matrix $\mathbf{P}^k$ contributes.

Define the following quantity:

$$t(k, k', i) = c_{k'k}\left[i + \sum_{v=0}^{k'-1} c_{vk} n_k\right]. \tag{20}$$

Index $t(k, k', i)$ is the actual serial index of orbital $\varphi_i(\mathbf{r})$ of nuclear family $f_{k'}$ in the fragment density matrix $\mathbf{P}^k$.

In order to construct the macromolecular density matrix $\mathbf{P}$, the transformed

row and column indices of each fragment density matrix $\mathbf{P}^k$ are determined and the corresponding matrix element is added to the appropriate matrix element of $\mathbf{P}$, according to the following scheme:

If

$$t(k,k',i)t(k,k'',j) \neq 0\,, \tag{21}$$

then set

$$P_{f(k',i),f(k'',j)} = P_{f(k',i)f(k'',j)} + P^k_{t(k,k',i),t(k,k'',j)}\,. \tag{22}$$

By carrying out this procedure for each nonzero element for each of the fragment density matrices $\mathbf{P}^k$, the corresponding ADMA density matrix $\mathbf{P}$ of the target macromolecule is obtained.

An alternative description of the ADMA algorithm, given below, follows more closely the actual computational process.

An individual AO $\varphi(\mathbf{r})$ can be assigned three indices, depending on the context. The notation $\varphi_{b,k'}(\mathbf{r})$ indicates that this is the bth AO within the set

$$\{\varphi_{a,k'}(\mathbf{r})\}_{a=1}^{n_{k'}} \tag{23}$$

of AOs associated with the nuclear family $f_{k'}$. The notation $\varphi_j^k(\mathbf{r})$ can also be used for the same AO if it is the jth AO within the basis set

$$\{\varphi_i^k(\mathbf{r})\}_{i=1}^{n_{Pk}} \tag{24}$$

used in the definition of the kth fragment density matrix $\mathbf{P}^k$. Here the number n_{P^k} of AOs involved in fragment density matrix $\mathbf{P}^k$is calculated as

$$n_{P^k} = \sum_{k'=1}^{m} c_{k'k} n_{k'}\,. \tag{25}$$

The notation $\varphi_y(\mathbf{r})$ is also used for the same AO if its serial index is y within the basis set

$$\{\varphi_x(\mathbf{r})\}_{x=1}^{n} \tag{26}$$

used in the definition of the macromolecular density matrix $\mathbf{P}$.

These indices are interrelated. If $\varphi_{a,k'}(\mathbf{r}) = \varphi_x(\mathbf{r})$, then index x can be determined from indices a and k' by

$$x = x(k',a,f) = a + \sum_{b=1}^{k'-1} n_b\,, \tag{27}$$

where f in the argument of index function $x(k',a,f)$ indicates that indices k' and a originate from a nuclear family. For each index k of fragment density matrices $\mathbf{P}^k$, if $\varphi_i^k(\mathbf{r}) = \varphi_x(\mathbf{r})$, then index x can be determined from indices i and k by the following procedure. For all those nuclear families $f_{k''}$ for which

$$c_{k''k} \neq 0\,, \tag{28}$$

define

$$a'_k(k'', i) = i + \sum_{b=1}^{k''} n_b c_{bk}\,, \tag{29}$$

$$k' = k'(i,k) = \min\{k'' : a'_k(k'', i) \leq 0\}\,, \tag{30}$$

and

$$a_k(i) = a'_k(k', i) + n_{k'}\,. \tag{31}$$

Using the index function $x(k', a, f)$ defined in eq. (27) and taking k' given in eq. (30), index x assigned to indices i and k is given by

$$x = x(k, i, P) = x(k', a_k(i), f)\,, \tag{32}$$

where **P** in the argument of the index function $x(k, i, P)$ indicates that indices k and i refer to a fragment density matrix.

With these index assignments, the macromolecular density matrix **P** is obtained by identifying each nonzero element P^k_{ij} of each fragment density matrix $\mathbf{P}^k$ and by setting

$$P_{x(k,i,P),y(k,j,P)} = P_{x(k,i,P),y(k,j,P)} + P^k_{ij}\,. \tag{33}$$

The ADMA algorithm described above generates a macromolecular density matrix **P** from fragment density matrices $\mathbf{P}^1, \mathbf{P}^2, \ldots, \mathbf{P}^k, \ldots \mathbf{P}^m$ calculated for parent molecules $M_1, M_2, \ldots, M_k, \ldots M_m$, designed to fulfill the compatibility conditions with one another and with the macromolecule M.

The ADMA macromolecular density matrix **P** and the macromolecular AO basis set $\varphi_x(\mathbf{r})_{x=1,..n}$ provide a detailed, *ab initio* quality quantum chemical description of macromolecule M. In terms of this density matrix **P**, the macromolecular electron density can be computed using relation (1), and other molecular properties can also be evaluated using standard density matrix methodology [50].

Note that for each fragment density matrix $\mathbf{P}^k$ the ranges of AO indices are limited by the size of the AO basis set used in the *ab initio* calculation for the parent molecule M_k. Since this size itself is limited, the computer time requirement for the entire index reassignment and for the matrix element assignment procedure for each fragment density matrix to the macromolecular density matrix is bounded by a constant. Consequently, the overall time requirement of the ADMA computation scales linearly with the number of fragment density matrices, that is, with the size of the macromolecule.

The negligible interaction matrix elements between AOs of the central nuclear family f_k and the peripheral AOs located at the additional nuclei beyond the "coor-

dination shell" in each parent molecule M_k are ignored within the ADMA scheme. These matrix elements belong to AO pairs separated by distances which are beyond the distance cutoff of the MEDLA numerical density fragments stored in the MEDLA database. Consequently, the ADMA macromolecular density matrix construction method reproduces the accuracy of the electron densities obtained with the MEDLA method, where the latter approach has been shown to generate *ab initio* quality electron densities. Consequently, the ADMA macromolecular density matrix construction method is suitable for the generation of *ab initio* quality electron densities.

By selecting sufficiently large nuclear families and parent molecules with large enough "coordination shells", the accuracy of the ADMA method, with respect to a direct, traditional *ab initio* computation, is limited only by the feasibility of large parent molecule computations. The ADMA method is based on the Mulliken-Mezey fragmentation scheme or on more general related schemes [27,28], consequently, ADMA accounts for the interfragment interactions in a consistent way and the macromolecular density matrix **P** constructed by the ADMA method provides a good approximation. However, if only small coordination shells are used in the parent molecules, and if the macromolecular density matrix is built from many fragment density matrices, then the neglect of the small "phantom" contributions from the additional, peripheral H atomic orbitals and accumulated numerical errors may cause the constructed density matrix **P** to deviate slightly from the condition of perfect charge conservation and the condition of idempotency. For proper charge conservation, the scaling method described in reference [38] can be applied directly on the density matrix **P**. Idempotency is the condition of projectors, where the projected image of a projected image remains the original projected image. Accurate density matrices must fulfill the condition of idempotency:

$$\mathbf{P} \cdot \mathbf{P} = \mathbf{P} . \tag{34}$$

Here the product operation · is interpreted as the matrix product **PSP** where **S** is the overlap matrix for a given nonorthogonal AO basis. However, by a small modification of the macromolecular density matrix **P**, the idempotency property can be restored. The idempotency condition is equivalent to requiring that a multidimensional vector is stretched to a point of a multidimensional unit sphere, that can be achieved by a simple transformation. Methods and numerical techniques for such transformations are discussed elsewhere [49].

Small nuclear geometry variations can be introduced in the same manner as it is done in the ALDA technique. The electronic density has maxima at the nuclei and the major effect of nuclear geometry variation on the electron density is dominated by the motion of the electronic density near the nuclei, essentially following the nuclear motion. If the geometry change is small, then an important contribution to the corresponding change in the electronic density can be obtained by taking the same macromolecular density matrix **P** obtained by the ADMA technique, and

using it with a new set of atomic orbitals located at a set of slightly displaced nuclear locations.

Small nuclear geometry variations change the relative positions of the AOs, change their overlap matrix **S**, and introduce small changes in both the idempotency property of **P** and the calculated total charge. If required, these changes can be accounted for by readjusting the macromolecular density matrix P, using the same charge-scaling and idempotency-restoring procedure outlined above.

The adjustability of the ADMA macromolecular density matrix **P** and the calculated electronic density to accommodate small geometry variations is advantageous in approximate representations of small molecular distortions, such as small amplitude vibrations and other minor geometry changes. This is likely to find applications in the structure refinement process of X-ray structure determination, in the study of minor conformational changes of polymers and protein folding processes.

The ADMA method can be regarded as a tool for obtaining *ab initio* quality density matrices **P** for large molecules **M** without first determining a wavefunction for M. Since within the Hartree-Fock framework the first-order density matrix **P** fully determines all higher-order density matrices, and expectation values for any one-electron and two-electron operators can be expressed in terms of the first-order and second-order density matrices [50], the ADMA method also provides a tool for calculating approximate expectation values for various macromolecular properties, including energy. In this context, the ADMA method extends the applicability of many quantum chemical computational techniques to macromolecules. Some of the suggested applications of the ADMA method are discussed elsewhere [49].

3.4. EXAMPLE AND ILLUSTRATION OF THE COMPATIBILITY CONDITION OF THE ADMA METHOD

The compatibility condition for selecting nuclear families of fragments within a target macromolecule and a corresponding set of parent molecules is illustrated by a simple example. The nuclei of the formal "macromolecule" used in the example are classified into 16 families,

$$f_1, f_2, \ldots, f_k, \ldots f_{16}\,. \tag{35}$$

Different nuclear families may contain different numbers of nuclei.

The macromolecular density matrix **P** generated by the ADMA method is defined in terms of a set of atomic orbitals belonging to the nuclear families 1, 2, 3, ..., m, where in the example, $m = 16$ was used. Within each nuclear family the ordering of atomic orbitals is fully specified, hence, by following the serial indices of the nuclear families, the entire set of atomic orbitals of the macromolecule is well ordered.

In order to simplify the visualization of the interactions between fragments, we consider a "macromolecule" rather "thin" along the z direction; note that the general case of more globular macromolecules can be treated the same way. The nuclear families arranged in any arbitrary way in three dimensions can always be assigned sequential numerical indices. Evidently, the interactions between fragments arranged in any arbitrary way in three dimensions can be represented by density matrix blocks and elements arranged within a two-dimensional matrix.

We assume that the 16 nuclear families of the "macromolecule" have a spatial arrangement approximately represented by the pattern

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{array} \tag{36}$$

Note that the nuclear arrangement is *not* assumed planar; pattern (36) merely indicates that there are strong interactions, for example, between the electron density clouds near the nuclei in family f_1 and those in family f_2, but there is little or negligible interaction between the electron density contributions near nuclei of family f_5 and near those of family f_{12}.

As an example, consider the construction of a fragment density matrix for the nuclear family f_6. If the nuclear families are large enough, then it is sufficient to take a parent molecule M_6 represented by the same arrangement of nuclear family f_6 and the corresponding nuclear arrangement of a "single layer deep" "coordination shell" generated by the additional nuclear families f_k of serial numbers 1, 2, 3, 5, 7, 9, 10, and 11, as shown by the following pattern, a sub-pattern of pattern (36):

$$\begin{array}{ccc} 1 & 2 & 3 \\ 5 & \mathbf{6} & 7 \\ 9 & 10 & 11 \end{array} \tag{37}$$

The nuclear set of parent molecule M_6 is completed by adding H nuclei (or, if needed, those of larger groups) to link up with the peripheral "dangling bonds" associated with nuclear families f_k of serial numbers 3, 7, 9, 10, and 11. In the actual example only these nuclear families, 3, 7, 9, 10, and 11, have neighboring families *not* included in the given parent molecule M_6. If a conventional *ab initio* computation is carried out for parent molecule M_6, then the complete density matrix $\mathbf{P}(M_6)$ of parent molecule M_6 is determined. Based on the density matrix $\mathbf{P}(M_6)$, the additive fragment density matrix (AFDM) $\mathbf{P}^6$ is determined, using eq.(8) or one of the alternative definitions [27,28]. This fragment density matrix $\mathbf{P}^6$ involves only those (possibly scaled) elements of $\mathbf{P}(M_6)$ which fall within the pattern of index ranges of AOs of the nuclear families schematically given below by the *fragment interaction pattern matrix* $\mathbf{F}^6$, using boldface number and asterisk:

$$
\begin{array}{ccccccccccc}
1 & & & & * & & & & & & \\
 & 2 & & & * & & & & & & \\
 & & 3 & & * & & & & & & \\
 & & & 5 & * & & & & & & \\
* & * & * & * & \mathbf{6} & * & * & * & * & & \\
 & & & & * & 7 & & & & & \\
 & & & & * & & 9 & & & & \\
 & & & & * & & & 10 & & & \\
 & & & & * & & & & 11 & &
\end{array}
\qquad (38)
$$

Each symbol in this fragment interaction pattern matrix $\mathbf{F}^6$ corresponds to a *block* of density matrix elements, where the row and column dimensions of each block depend on the number n_k of atomic orbitals associated with the nuclear family f_k of the given serial index k specified along the diagonal. The atomic orbitals belonging to the extra H atoms are specified at the end of the orbital list of parent molecule M_6. The orbitals on these H atoms have negligible interactions with the orbitals centered on the nuclei of the actual fragment, these extra orbitals do not contribute to the macromolecular density matrix $\mathbf{P}$, and no corresponding columns and rows on the right hand side and at the bottom of the diagram are indicated.

All those matrix elements of density matrix $\mathbf{P}(M_6)$ which involve only atomic orbitals belonging to nuclear family f_6 are included in full; their block is symbolized by the boldface number **6** in the diagram. Symbol * represents off- diagonal blocks of density matrix $\mathbf{P}(M_6)$ for orbital pairs where in each pair only one orbital is centered on a nucleus of family f_6 and the other orbital belongs to another family f_k in the interacting "coordination-shell", where k is 1, 2, 3, 5, 7, 9, 10, or 11. For all these off-diagonal blocks, each of the original matrix elements of $\mathbf{P}(M_6)$ is scaled: if eq. (8) is applied, then each of these matrix elements is multiplied by a factor of 0.5; if one of the more general schemes [27,28] is used, then the appropriate multiplying factors are applied.

Although the atomic orbitals centered on nuclei of family f_7 contribute to the fragment density matrix $\mathbf{P}^6$, this involvement is restricted to a representation of approximately half of the local interfragment electron density interactions. For the construction of a fragment density matrix $\mathbf{P}^7$ for fragment 7, a different parent molecule, M_7, is used. Parent molecule M_7 is symbolically represented by a subpattern of pattern (36), as shown below:

$$
\begin{array}{ccc}
2 & 3 & 4 \\
6 & \mathbf{7} & 8 \\
10 & 11 & 12
\end{array}
\qquad (39)
$$

The "coordination shell" in parent molecule M_7 is generated by the nuclear families f_k of serial numbers 2, 3, 4, 6, 8, 10, 11, and 12. The "dangling bonds" of the

peripheral nuclei of these families are linked to extra H atoms (or, if needed, to larger groups). These extra nuclei are located far from the central nuclear family f_7, therefore, the interaction density matrix elements between orbitals of family f_7 and the orbitals centered on the extra, peripheral nuclei are assumed to be negligible. The density matrix $\mathbf{P}(M_7)$ of the parent molecule M_7 is used for the construction of the fragment density matrix $\mathbf{P}^7$ according to eq.(8) or one of the alternative definitions [27,28], that can be represented symbolically by the following fragment interaction pattern matrix $\mathbf{F}^7$:

2				*							
	3			*							
		4		*							
			6	*							
*	*	*	*	**7**	*	*	*	*			(40)
				*	8						
				*		10					
				*			11				
				*				12			

The interaction matrix elements between orbitals centered on nuclei in family f_6 and family f_7 occur only in the two parent molecules M_6 and M_7, as indicated in the two fragment interaction pattern matrices (38) and (40), respectively. Accordingly, these interfragment interactions are represented only in the corresponding two fragment density matrices, $\mathbf{P}^6$ and $\mathbf{P}^7$.

There are two main strategies to ensure a high level of incorporation of interfragment interactions in the final macromolecular density matrix $\mathbf{P}$.

According to the first, simple approach, one may select large nuclear families f_k for each fragment. The compatibility requirement of the fragmentation used in the ADMA method implies that the fragments representing the "coordination shell" in each of the parent molecules are also large. Consequently, within each of the fragment density matrices $\mathbf{P}^k$ the interactions from a large part of the actual target molecule M are well represented, and a faithful representation of the macromolecular density matrix $\mathbf{P}$ is obtained when one combines the fragment density matrices $\mathbf{P}^k$.

According to the second approach, a larger number of smaller nuclear families are used, where each parent molecule M_k involves a large number of such nuclear families. If a high level of accuracy is required for the representation of interfragment interactions in a macromolecule M, then relatively small nuclear families are still applicable for each fragment, if the formal "coordination-shell" in each parent molecule M_k involves two (or more) layers of nuclear families surrounding the central nuclear family f_k. If the "coordination shell" involves several layers of nuclear families, a construction method similar to that described for the simplest, single-layer "coordination-shell" applies.

As an illustration, we shall consider the same "macromolecule" M, with the modification of taking two layers of nuclear families for a "coordination shell" in each of the parent molecules. As examples, the corresponding patterns of parent molecules M_5 and M_9, the associated fragment interaction pattern matrices $\mathbf{F}^5$ and $\mathbf{F}^9$, and the respective fragment density matrices $\mathbf{P}^5$ and $\mathbf{P}^9$ of nuclear families f_5 and f_9 are specified below.

The approximate arrangements of nuclear families in parent molecules M_5 and M_9 are

$$\begin{array}{cccc} & 1 & 2 & 3 \\ & \mathbf{5} & 6 & 7 \\ & 9 & 10 & 11 \\ & 13 & 14 & 15 \end{array} \tag{41}$$

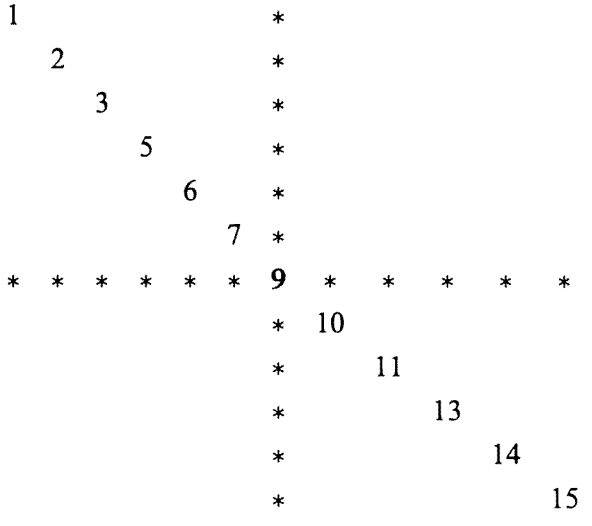
and

$$\begin{array}{cccc} & 1 & 2 & 3 \\ & 5 & 6 & 7 \\ & \mathbf{9} & 10 & 11 \\ & 13 & 14 & 15 \end{array} \tag{42}$$

respectively, where the "dangling bonds" associated with the peripheral nuclei of families f_k of indices 3, 7, 11, and 15 are linked to additional H atoms. Due to the simplicity of the target "macromolecule" M, the same parent molecule, $M_5 = M_9$, can be used for both fragment density matrices $\mathbf{P}^5$ and $\mathbf{P}^9$. The block-pattern of the fragment density matrix $\mathbf{P}^5$ is specified by the respective fragment interaction pattern matrix $\mathbf{F}^5$

$$\begin{array}{cccccccccccc} 1 & & & * & & & & & & & & \\ & 2 & & * & & & & & & & & \\ & & 3 & * & & & & & & & & \\ * & * & * & \mathbf{5} & * & * & * & * & * & * & * & * \\ & & & * & 6 & & & & & & & \\ & & & * & & 7 & & & & & & \\ & & & * & & & 9 & & & & & \\ & & & * & & & & 10 & & & & \\ & & & * & & & & & 11 & & & \\ & & & * & & & & & & 13 & & \\ & & & * & & & & & & & 14 & \\ & & & * & & & & & & & & 15 \end{array} \tag{43}$$

The fragment interaction pattern matrix $\mathbf{F}^9$ describes the corresponding block pattern of the fragment density matrix $\mathbf{P}^9$:



(44)

In the process of combining the fragment density matrices $\mathbf{P}^k$ when building the macromolecular density matrix $\mathbf{P}$, blank rows and columns of zero blocks are included for those nuclear families which do not contribute to the parent molecule $\mathbf{M}_k$ of the given fragment density matrix $\mathbf{P}^k$.

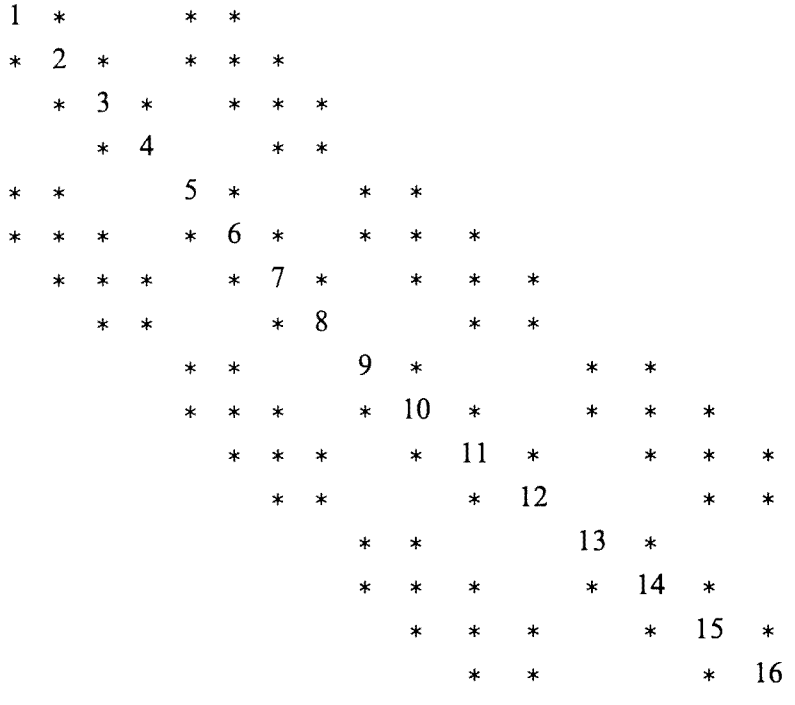
In the example, if single-layer coordination shells are used for each parent molecule $\mathbf{M}_k$, then the fragment density matrices $\mathbf{P}^6$ and $\mathbf{P}^7$ are transformed by inserting rows and columns of zero blocks corresponding to atomic orbitals associated with the "non-participating" nuclear families of indices 4, 8, 12, 13, 14, 15, and 16 for $\mathbf{P}^6$ and 1, 5, 9, 13, 14, 15, and 16, for $\mathbf{P}^7$, respectively. Similarly, if two layers of nuclear families are taken for a coordination shell in each parent molecule $\mathbf{M}_k$, then the corresponding double-layer fragment density matrix $\mathbf{P}^5$ is transformed by inserting rows and columns of zero blocks corresponding to atomic orbitals associated with the "non-participating" nuclear families of indices 4, 8, 12, and 16. In the given example, both nuclear families f_5 and f_9 are located within the same peripheral region of the "macromolecule" $\mathbf{M}$, and by coincidence, the transformation of the corresponding double-layer fragment density matrix $\mathbf{P}^9$ requires the insertion of rows and columns of zero blocks corresponding to atomic orbitals associated with the *same* set of nuclear families, 4, 8, 12, and 16.

The inclusion of rows and columns of zero blocks into the fragment density matrices $\mathbf{P}^k$ corresponds to the inclusion of blank rows and columns into the respective fragment interaction pattern matrices $\mathbf{F}^k$. The resulting density matrix block patterns are called the *augmented fragment interaction pattern matrices*.

The macromolecular density matrix $\mathbf{P}$ can be assembled from fragment density

matrices $\mathbf{P}^k$ appropriately augmented with rows and columns of zero blocks. This construction is symbolized by a *composite fragment interaction pattern matrix* $\mathbf{F}$ obtained as the sum of the augmented interaction pattern matrices of all the individual fragment density matrices $\mathbf{P}^k$.

In the example of the single-layer coordination shell, the composite fragment interaction pattern matrix $\mathbf{F}$ for the "macromolecular" density matrix $\mathbf{P}$ has the following form:



(45)

Each block of fragment density matrix elements along the diagonal of this composite fragment interaction pattern matrix $\mathbf{F}$ originates from the fragment density matrix $\mathbf{P}^k$ of the serial index k indicated in the diagonal of $\mathbf{F}$. By contrast, there are two contributions to each block represented by an off-diagonal asterisk in this composite fragment interaction pattern matrix $\mathbf{F}$. If the asterisk is located in row i and column j, then the off-diagonal blocks of the ith and jth augmented fragment density matrices $\mathbf{P}^i$ and $\mathbf{P}^j$ with contributions from atomic orbitals belonging to the corresponding jth and ith nuclear families are added together to form the corresponding off-diagonal block of the final, macromolecular density matrix $\mathbf{P}$.

If each coordination shell involves two layers of nuclear families, as illustrated by patterns (43) and (44) for the corresponding fragment density matrices $\mathbf{P}^5$ and $\mathbf{P}^9$, then the associated composite fragment interaction pattern matrix $\mathbf{F}$ for a more accurate macromolecular density matrix $\mathbf{P}$ has the following form:

$$
\begin{array}{cccccccccccccccc}
1 & * & * & & * & * & * & & * & * & * & & & & & \\
* & 2 & * & * & * & * & * & * & * & * & * & * & & & & \\
* & * & 3 & * & * & * & * & * & * & * & * & * & & & & \\
 & * & * & 4 & & * & * & * & & * & * & * & & & & \\
* & * & * & & 5 & * & * & & * & * & * & & * & * & * & \\
* & * & * & * & * & 6 & * & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & 7 & * & * & * & * & * & * & * & * & * \\
 & * & * & * & & * & * & 8 & & * & * & * & & * & * & * \\
* & * & * & & * & * & * & & 9 & * & * & & * & * & * & \\
* & * & * & * & * & * & * & * & * & 10 & * & * & * & * & * & * \\
* & * & * & * & * & * & * & * & * & * & 11 & * & * & * & * & * \\
 & * & * & * & & * & * & * & & * & * & 12 & & * & * & * \\
 & & & & * & * & * & & * & * & * & & 13 & * & * & \\
 & & & & * & * & * & * & * & * & * & * & * & 14 & * & * \\
 & & & & * & * & * & * & * & * & * & * & * & * & 15 & * \\
 & & & & & * & * & * & & * & * & * & & * & * & 16
\end{array}
\qquad (46)
$$

One may regard this composite fragment interaction pattern matrix **F** as the blueprint for the construction of the macromolecular density matrix **P** involving two layers of fragments for the coordination shell of each parent molecule M_k.

The interpretation of this pattern is analogous to that for the simpler, single-layer case. Each block of fragment density matrix elements along the diagonal of **F** originate from the fragment density matrix $\mathbf{P}^k$ of the serial index k shown in the diagonal. Each block of the final, macromolecular density matrix **P** represented by an off-diagonal asterisk in row i and column j of the composite fragment interaction pattern matrix **F** is a sum of two contributions: the row j and column i off-diagonal blocks of the ith and jth augmented fragment density matrices $\mathbf{P}^i$ and $\mathbf{P}^j$, containing interaction contributions from atomic orbitals belonging to the corresponding jth and ith nuclear families are added together.

In practice, no actual augmentation of matrices is carried out; a simple re-indexing of nonzero matrix elements is a more efficient alternative.

The sizes of target and parent molecules are typically very different, implying that both the composite fragment interaction pattern matrix **F** and the resulting macromolecular density matrix **P** obtained using the ADMA method are rather sparse. This fact can be exploited when storing these matrices in a database. In this context, the actual pattern obtained in our second example, indicating many contributions, is somewhat misleading, since our "macromolecule" is not much larger than the parent molecules M_k.

4. Summary

Two new applications of the additive fuzzy electron density fragmentation principle are reported. The techniques are formulated within the *ab initio* Hartree-Fock quantum chemical computational framework.

The first of the two techniques, the Adjustable Local Density Assembler or ALDA method, generates geometry-adjusted *ab initio* quality macromolecular electron densities directly, using fragment density matrices, basis set information, and nuclear coordinates. The method requires an ALDA fragment electron density matrix database.

The second method, the Adjustable Density Matrix Assembler or ADMA method, is introduced for the generation of *ab initio* quality approximate *density matrices* for macromolecules. The method assembles fragment density matrices designed to fulfill a macromolecular compatibility condition. The ADMA method is suitable for generating macromolecular density matrices without the computation of a macromolecular wavefunction. The ADMA method extends the applications of the density matrix techniques of conventional quantum chemistry to macromolecules such as proteins. The simple treatment of small nuclear geometry variations is expected to improve the description of some of the vibrational and dynamic properties of macromolecules.

Acknowledgement

The support of the Natural Sciences and Engineering Research Council of Canada is acknowledged.

Note added in proof. A new method for the construction of projector matrices based on X-ray diffraction data has been developed recently by Massa, Huang and Karle [51].

References

[1] D.R. Hartree, Proc. Cambridge Phil. Soc. 24 (1928) 111,426, *ibid.* 25 (1929) 225,310.
[2] V. Fock, Z. Physik 61 (1930) 126.
[3] C.C. Roothaan, Rev. Mod. Phys. 23 (1951) 69, ibid. 32 (1960) 179.
[4] G.G. Hall, Proc. Roy. Soc. London A205 (1951) 541.
[5] R. Carbó, L. Leyda and M. Arnau, Int. J. Quantum Chem. 17 (1980) 1185.
[6] R. Carbó and Ll. Domingo, Int. J. Quantum Chem. 32 (1987) 517.
[7] R. Carbó and B. Calabuig, Comput. Phys. Commun. 55 (1989) 117.
[8] R. Carbó and B. Calabuig, Int. J. Quantum Chem. 42 (1992) 1681.
[9] R. Carbó and B. Calabuig, Int. J. Quantum Chem. 42 (1992) 1695.
[10] R. Carbó, E. Besalú, B. Calabuig and V. Vera, Adv. Quant. Chem. 25 (1994) 253.

[11] E. Besalú, R. Carbó, J. Mestres and M. Solá, Foundations and recent developments on molecular quantum similarity, in: *Topics in Current Chemistry, Vol. 173, Molecular Similarity*, ed. K. Sen (Springer-Verlag, Heidelberg, 1995).
[12] E.E. Hodgkin and W.G. Richards, J. Chem. Soc. Chem. Commun. 1986 (1986) 1342.
[13] E.E. Hodgkin and W.G. Richards, Int. J. Quantum Chem. 14 (1987) 105.
[14] A. Good and W.G. Richards, J. Chem. Inf. Comp. Sci. 33 (1992) 112.
[15] S. Leicester, R. Bywater and J.L. Finney, J. Mol. Graph. 6 (1988) 104.
[16] R. Bywater, in: *Molecular Similarity and Reactivity: from Quantum Chemical to Phenomenological Approaches*, ed. R. Carbó (Kluwer, Dordrecht, The Netherlands, 1995).
[17] D.L. Cooper and N.L. Allan, in: *Molecular Similarity and Reactivity: from Quantum Chemical to Phenomenological Approaches*, ed. R. Carbó (Kluwer, Dordrecht, The Netherlands, 1995).
[18] N.L. Allan and D.L. Cooper, in: *Advances in Molecular Similarity*, eds. R. Carbó and P.G. Mezey (JAI Press, Greenwich, USA, to be published).
[19] J. Cioslowski, S.T. Mixon, Can. J. Chem. 70 (1992) 443.
[20] J. Cioslowski and A. Nanayakkara, J. Am. Chem. Soc. 115 (1993) 11213.
[21] J. Cioslowski and B.B. Stefanov, Mol. Phys. 84 (1995) 707.
[22] P.G. Mezey, Int. J. Quant. Chem. Quant. Biol. Symp. 12 (1986) 113.
[23] P.G. Mezey, J. Comput. Chem. 8 (1987) 462.
[24] P.G. Mezey, J. Math. Chem. 2 (1988) 299.
[25] P.G. Mezey, J. Math. Chem. 2 (1988) 325.
[26] P.G. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology* (VCH Publishers, New York, 1993).
[27] P.G. Mezey, Density domain bonding topology and molecular similarity measures, in: *Topics in Current Chemistry, Vol. 173, Molecular Similarity*, ed. K. Sen (Springer-Verlag, Heidelberg, 1995).
[28] P.G. Mezey, Methods of molecular shape-similarity analysis and topological shape design, in: *Molecular Similarity in Drug Design*, ed. P.M. Dean (Chapman & Hall - Blackie Publishers, Glasgow, U.K., 1995).
[29] R.E. Christoffersen and G.M. Maggiora, Chem. Phys. Lett. 3 (1969) 419.
[30] R.E. Christoffersen, L.L. Shipman and G.M. Maggiora, Int. J. Quantum Chem. 5 (1971) 143.
[31] R.E. Christoffersen, D. Spangler, G.G. Hall and G.M. Maggiora, J. Am. Chem. Soc. 95 (1973) 8526.
[32] P.G. Mezey, Int. J. Quant. Chem. Quant. Biol. Symp. 14 (1987) 127.
[33] P.G. Mezey, Molecular surfaces, in: *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz and D.B. Boyd (VCH Publ., New York, 1990).
[34] P.G. Mezey, Three-dimensional topological aspects of molecular similarity, in: *Concepts and Applications of Molecular Similarity*, eds. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990).
[35] G.A. Arteca, N.D. Grant and P.G. Mezey, J. Comput. Chem. 12 (1991) 1198.
[36] W. Yang, Phys. Rev. Lett. 66 (1991) 1438.
[37] C. Lee and W. Yang, J. Chem. Phys. 96 (1992) 2408
[38] P.D. Walker and P.G. Mezey, J. Math. Chem. 17 (1995) 203.
[39] P.D. Walker and P.G. Mezey, J. Am. Chem. Soc. 115 (1993) 12423.
[40] P.D. Walker and P.G. Mezey, *Program MEDLA 93* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1993).
[41] P.D. Walker and P.G. Mezey, J. Am. Chem. Soc. 116 (1994) 12022.
[42] P.D. Walker and P.G. Mezey, Canad. J. Chem. 72 (1994) 2531.
[43] P.D. Walker and P.G. Mezey, J. Comput. Chem. 16 (1995) 1238.
[44] P.G. Mezey, *Program ALDA 95* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1995).

[45] P.G. Mezey, *Program ADMA 95* (Mathematical Chemistry Research Unit, University of Saskatchewan, Saskatoon, Canada, 1995).
[46] P.G. Mezey, Local shape analysis of macromolecular electron densities, in: *Computational Chemistry: Reviews and Current Trends*, ed. J. Leszczynski (World Scientific, Singapore, in press).
[47] R.S. Mulliken, J. Chem. Phys. 23 (1955) 1833,1841,2338,2343.
[48] R.S. Mulliken, J. Chem. Phys. 36 (1962) 3428.
[49] P.G. Mezey, to be published.
[50] F.L. Pilar, *Elementary Quantum Chemistry* (McGraw-Hill, New York, 1968).
[51] L. Massa, L. Huang and J. Karle, to be published.