A High-Resolution Shape-Fragment MEDLA Database for Toxicological Shape Analysis of PAHs

P. G. Mezey,*,† Z. Zimpel,† P. Warburton,† P. D. Walker,† D. G. Irvine,‡ D. G. Dixon,§ and B. Greenberg§

Mathematical Chemistry Research Unit, Department of Chemistry, and Toxicology Research Center, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan, Canada S7N 5C9, and Department of Biology, University of Waterloo, Waterloo, Ontario, Canada

Received October 19, 1995⊗

A new, high-resolution shape-fragment database has been developed for computing *ab initio* quality molecular electron densities for polyaromatic hydrocarbons (PAHs) which play a significant role as toxicants in the environment. Using the new PAH electron density fragment database and the Molecular Electron Density Lego Assembler (MEDLA) method, one can generate detailed and reliable electron densities for virtually any of the PAH molecules. Accurate electron density shape representations for these molecules is essential in the study of detailed shape-toxicity correlations. One of our goals is to investigate the potential of detailed molecular shape analysis as a predictive tool in toxicological risk assessment. In this study we report the results of the first phase of the study: the construction and testing of a high quality shape-fragment database for PAHs.

INTRODUCTION

The purpose of this study is the establishment of reliable electron density shape-fragment databases for polyaromatic hydrocarbons, PAHs, with the aim to use these databases in shape-activity correlations in quantitative toxicological risk assessment (QRA) studies. After a brief review of the fundamental justification, motivation, and the methodology of this approach, the new shape-fragment databases are described, and various test calculations establishing their quality are reported.

The high level of complexity of most molecules and the wide variety of their chemical reactions often obscure a simple fact: *a molecule contains nothing else than atomic nuclei and an electronic density cloud.* The distribution of the atomic nuclei can be deduced from the electronic density, consequently, *the electronic density contains all information about the molecule.* Specifically, if two molecules react in a similar or in a different way, these similarities or differences *must* be reflected in their electronic densities. In fact, the *complete information* about molecular similarities and differences must be present in their electronic densities, since besides the nuclei whose distribution is described by the electron density, there is nothing else left to be similar or different. This holds equally for static and dynamic properties: molecules have similar static and dynamic properties precisely because their electron densities and the nuclear distributions fully reflected by them have similar or different static and dynamic properties. Electronic densities hold the key to chemistry. A general tool that gives a complete description of the electron density cloud is *shape analysis*, involving both the high and low density ranges of the charge density clouds.

Whereas all the information that can be attributed to individual molecules is present in the electron density, the difficult questions are how to obtain reliable electron density representations and how to decode the information contained in them. If the chemical or biochemical problem is complex, such as most problems in toxicology, then our present level of knowledge and information processing capacity is seldom sufficient for a detailed understanding of the actual complex role of electron density. Instead of understanding, one is usually forced to rely on mere *correlations* between observed toxic properties or reactivities and some of the specific shape features of the electron density. Nevertheless, such correlations are of predictive value even without a full understanding of the actual chemical problem; similar features of electron density are expected to correlate with similar chemical properties.

The fundamental role of molecular similarity in modern drug design and molecular engineering[1−10] is well documented, and very similar considerations apply in toxicological risk assessment. The toxicological risk presented by a contaminant released to the environment arises from interactions between the contaminant molecules and a very complex environment, involving various biological receptors and a multitude of factors which modify these interactions. The extremely complex environment cannot be modeled accurately; however, the much simpler molecules can be modeled in detail using scientifically sound methods. Hence it is prudent to approach toxicological risk assessment from the side of the molecules which, after all, are responsible for the hazard (potential risk). A different molecule represents a different hazard, and, evidently, any risk associated with a given molecule that can be assessed at all, depends on the properties, hence, on the electron density of the molecule.

The fundamental idea behind the approach followed in this study can be summarized as follows: *The shape of the electron density cloud contains all information on the reactivity of the molecule, including all information on every*

† Department of Chemistry.
‡ Toxicology Research Center.
§ Department of Biology.
* Author for correspondence.
⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

0095-2338/96/1636-0602$12.00/0

component of the assessable toxicological risk that can be attributed to the given molecule.

The principles and methods of molecular shape analysis, as described in earlier studies primarily from the drug design perspective,[11−21] can be applied to develop a new toxicological approach. By analogy with the Quantitative *Shape*-Activity Relations (QShAR) approach of drug design, where the conventional *structure* used in the QSAR is replaced by the *shape* of electron density clouds, the toxicological adaptation of these methods is the basis of the Quantitative Shape-Toxicological Activity Relations (QShTAR) approach. Earlier QShAR computer programs used in drug design applications[20,21] can be modified for toxicological applications. The new QShTAR computer programs have the dual role of detecting and evaluating similarities within molecular families and correlating local and global shape features with known toxicological information. Correlations between specific local shape regions and toxicological activity of kown toxicants can be used as risk predictors for untested potential environmental contaminants or newly synthesized molecules with similar shape features, even before they are released to the environment and any harmful effect is detected experimentally.

The extension of the molecular shape analysis methods to macromolecules, such as proteins including receptors for toxicants, and their use in toxicological QRA, have become enhanced by a new computational method, the Molecular Electron Density Lego Assembler (MEDLA) technique for rapid computation of *ab initio* quality shapes of large molecular bodies.[22−26] The detailed local and global shapes of electron density clouds, representing the actual, real molecular bodies, can now be computed efficiently for molecules of virtually any size.[22−26] The electron densities of the parent molecules are generated by the GAUSSIAN 90 *ab initio* program package,[27] whereas the actual shape analysis program GSHAPE 90 is applied to the electron densities of larger molecular systems generated by the MEDLA 93 program.[28,29] An alternative geometry optimization approach, such as the molecular mechanics method,[30] may replace the quantum chemical geometry optimization step, or the newly developed macromolecular Adjustable Density Matrix Assembler (ADMA) technique[31−33] can be used for quantum chemical property calculations and geometry optimization. These approaches allow one to study in great detail local and global shape correlations with toxicological activity, and to provide tools for predictive toxicology and toxicological risk assessment.

The selection of the family of polycyclic aromatic hydrocarbons (PAHs) for the development of the first systematic shape fragment database was motivated by their toxicological importance. Various types of PAH toxicity are especially well documented, including extensive histopathology in fish and aquatic invertebrates and chlorosis, as well as reduced growth in aquatic and terrestrial plants. The photo-oxidation products of the PAHs are often implicated.[34−36] An important property of PAHs is that their toxicity can be altered either abiotically (*e.g.*, photochemically) or biologically (*e.g.*, cytochrome P450-mediated monooxygenation).[35,37−40] In photochemical activation, the extensive π-orbital systems of PAHs absorb light in the ultraviolet-B (UV-B; 290−320 nm) and/or the ultraviolet-A (UV-A; 320−400 nm) ranges of the solar spectrum.[35,37] Following the absorbance of a photon, the molecules become reactive, that can result in photooxidation of PAHs. These processes have been shown to occur rapidly under environmentally relevant levels of actinic radiation.[41,42] The photooxidation products are similar in structure to those produced via cytochrome P450-mediated oxygenation. Several studies have indicated that the toxicity of PAHs to microbes, animals and aquatic plants is enhanced by actinic radiation.[34,35,43] Most PAHs have high quantum yields for triplet-state formation and are active type II photosensitizers.[34,44] In this capacity, they are very efficient at promoting the formation of biologically damaging singlet oxygen in the presence of sunlight.[45] As well, recent studies with the aquatic plant *Lemna Gibba* (a duckweed) have shown that PAHs are photomodified by actinic radiation to a complex, unidentified mixture of photoproducts that are more toxic than the parent compounds.[35,40,41] The molecular aspects of physiologically-based pharmacokinetic (PBPK) modeling of PAH families in multiple species also involve molecular shape-effect correlations within dose response models. The pharmacokinetic behavior of the actual toxicants in various species, where the toxic response can be traced back to a PAH, is ultimately also influenced by the shape of the given PAH. The molecular shape features of PAHs, as the original initiators of toxicological action, influence the pharmacokinetic roles of the actual toxicants in various subgroups of multiple species. In studies involving multiple species, the trends and correlations are multidimensional. A multiple local shape feature-multiple toxicological effect correlation approach requires the identification of local molecular shape features relevant to toxicological actions in various species.

A BRIEF REVIEW OF THE SHAPE GROUP AND MEDLA METHODS

The Shape Group Method[11−13,19] is a computational technique for detailed, nonvisual shape analysis of molecular electron densities, forming the fuzzy bodies of molecules. In its simplest form, the shape group technique is applicable to the shape analysis of a family of molecular isodensity contour (MIDCO) surfaces $G(a)$, where one such surface corresponds to each density threshold value a. Each molecular contour surface $G(a)$ is partitioned into domains, usually defined by local curvature properties. If the curvature of the plane, that is, zero curvature, $b = 0$, is used as reference, then these domains are the locally convex (D_2), locally saddle type (D_1), and locally concave (D_0) domains. If a different reference curvature, $b \neq 0$, is used, for example, the curvature of a sphere of radius $1/b$, then the corresponding "relative convexity domains" $D_2(b)$, $D_1(b)$, and $D_0(b)$, are defined relative to this curvature b. The pattern and mutual relations among these domains define several topological groups, the shape groups of the molecule. The shape groups are the two-dimensional, one-dimensional, and zero-dimensional homology groups of truncated surfaces,[11−13] obtained from the original surface $G(a)$ by removing all domains of a given type [for example, $D_2(b)$, $D_1(b)$, or $D_0(b)$]. Homology groups are powerful tools of modern algebraic topology, suitable to extract the essential shape properties of various objects.[19] The ranks of the homology groups are called Betti numbers; these integers provide a concise numerical characterization of the topology of the given object. In most molecular applications of shape groups, the one-dimensional homology groups of the $D_2(b)$-type truncation of isodensity surfaces has been found to provide the most efficient

molecular shape characterization. The shape groups are algebraic groups, just like symmetry groups, but the shape groups are not determined by the point symmetry groups of the molecules. The shape groups can be obtained algorithmically for computer-generated molecular isodensity contour surfaces of molecules.

The Betti numbers of the shape groups are important topological invariants and serve as the basis of a numerical, nonvisual shape characterization. The shape groups and their Betti numbers for an isodensity contour surface $G(a)$ for some density value a and a fixed reference curvature b are invariant within some density intervals (a_i, a_{i+1}). A similar invariance is obtained if the density threshold is fixed and the reference curvature is varied: the shape groups are invariant within ranges (b_i, b_{i+1}) of the reference curvature b. There are only a finite number of the invariance intervals (a_i, a_{i+1}) and (b_i, b_{i+1}). Consequently, for the entire range of physically relevant density values a and reference curvatures b, there are only a finite number of topologically different shape groups, describing the shape of the entire charge density distribution of the molecule. A list of the one-dimensional Betti numbers for the finite number of shape groups of the molecule generates a numerical shape code.[19] Such a shape code gives a detailed and rigorous shape characterization of the entire electronic charge density of the molecule.

Most molecules are rather flexible objects: their shape is not a static property. This conformational flexibility of molecules can also be described using the shape group methods. The shape groups of a molecule are invariant to small conformational changes, hence within a range of nuclear arrangements only finite number of different shape groups occur. This leads to the dynamic shape space approach for a dynamic shape characterization of molecules within conformational domains.[19]

A numerical comparison of molecular shape codes can be used for similarity analysis: the similarity of the shapes of three-dimensional molecular bodies can be quantified and measured by comparing their numerical shape codes.[14−19] Usually, the family of one-dimensional Betti numbers is ordered into a matrix, called the (a,b)-map. The standard methods of comparing matrices can be applied to these (a,b)-maps in order to obtain a numerical measure for molecular similarity. This comparison involves no visual inspection and is carried out automatically by the computer. The subjective elements of judging similarity by human observers is avoided; the algorithmic shape-similarity evaluation by the computer is fully reproducible.

The Molecular Electron Density Lego Assembler (MEDLA) method[22−26] is a technique that can generate *ab initio* quality electron densities for large molecules of size that renders their description by conventional software of *ab initio* electron density computations unfeasible or impossible. The MEDLA method circumvents the main obstacles of conventional *ab initio* quantum chemical calculations of electron densities for macromolecules: the formidable memory and computer time requirements. In chemical, biochemical, pharmaceutical, and toxicological applications, the MEDLA technique serves as a *computational microscope*, as well as the source of detailed electron densities for molecular shape-similarity studies and QShAR analysis.

The MEDLA approach is based on a simple electron density fragment additivity principle and an algorithm described in detail elsewhere.[22−26] Using the simplest variant of the additive, fuzzy density fragmentation (AFDF) method, mutually interpenetrating molecular fragment densities are computed for smaller "parent" molecules by an *ab initio* method using a suitable basis set. In this study, a high quality, 6-31G** basis set is used. The fragment densities are stored in the MEDLA data bank and can be used as fuzzy "building blocks" for the construction of electron densities of larger "target" molecules. The building process involves the selection and arrangement of the interpenetrating fuzzy charge clouds of density fragments so that the nuclear positions match those in the target molecule. If the stored fragment densities are recombined to form the density of the "parent" molecule they are derived from, then the reconstruction is exact on the given *ab initio* level. Detailed tests have shown that the computed electronic densities are highly accurate (at the given *ab initio* level) even if the fragment densities are combined to form a new molecule.[22−26] The scope of earlier methods for the analysis of global and local shapes of molecules has been considerably extended by the introduction of the MEDLA method.

The MEDLA fragment data bank is generated by first computing several versions of each of the required electron density fragments, within a range of possible distorted nuclear geometries. These fragments are calculated as fuzzy electron density contributions within some smaller "parent" molecules, using the Gaussian 90 *ab initio* program and the 6-31G** basis set. The electron densities of the parent molecules are computed in several distorted configurations. As a result, each fragment type is represented by several versions in the MEDLA data bank; these versions differ by minor deviations in the local nuclear geometry and local surroundings. When building the electron density of the large target molecule, the best matching version of the fragment is selected from the MEDLA database. This fragment density mimics the actual nuclear arrangement and the effects of the surroundings of the given fragment in the large target molecule.

If the MEDLA data bank does not contain a suitable fragment, that is, if a new version of a fragment is required, then a conventional *ab initio* computation is carried out for another, "custom made" small parent molecule that contains the fragment with the required local geometry and surroundings. The new density fragment is then added to the MEDLA database. If one takes a large enough parent molecule for each fragment, where the local nuclear arrangement and the local surroundings match within a large enough neighborhood, then, in principle, the direct *ab initio* result of the target molecule can be reproduced at the given basis set level to any desired accuracy.

Numerical similarity measures obtained in test calculations comparing MEDLA results and conventional *ab initio* electron densities have shown that within all ranges of the electron density, including regions near the nuclei, near the chemical bonds, in hydrogen-bonded structures, and in moieties with overlap interactions in folded structures, the MEDLA densities are of *ab initio* quality.[22−26] MEDLA densities are of better quality than densities obtained by conventional 3-21G *ab initio* calculations, and the differences between the 6-31G** MEDLA and conventional 6-31G** *ab initio* electron densities are much less than those between the conventional 3-21G and conventional 6-31G** *ab initio* electron densities. In most instances, the 6-31G** MEDLA

and conventional 6-31G** *ab initio* results were visually indistinguishable. For shape analysis purposes, the conventional 6-31G** *ab initio* and 6-31G** MEDLA electron densities are of comparable quality, and form a compatible pair of computational levels for QShAR analysis.

A SHAPE-FRAGMENT DATA BANK FOR PAH MOLECULES

Since most PAH molecules are rather large, conventional *ab initio* electron density calculation methods are often unsuitable for shape analysis, especially if more than a few molecules are to be studied. The MEDLA method provides an alternative approach of comparable accuracy. Since the MEDLA technique uses a shape fragment database, this approach suggests the need for a high-resolution MEDLA electron density fragment database specifically designed for PAHs.

In the context of QShTAR, the family of PAH molecules is special for two reasons. On the one hand, shape-toxicity correlation studies of PAH molecules and oxidation products of substituted PAH molecules are of major importance in environmental toxicology. On the other hand, the shape analysis methodology is rather simple for PAHs due to their special structural features. The structural simplicity of PAHs, relative to most nonplanar molecules, implies that there are only a few dominant shape correlations between easily recognizable structural features and toxic activity. It is likely that less than a dozen types of specific local shape features account for most of the toxicological trends of PAHs, including oxy-PAHs. If this assumption is valid, then the identification of these shape features will provide important predictive tools for toxicological risk assessment.

For the development of methods of multiple local shape feature−multiple toxicological effect correlations, it is important to identify local molecular regions and their specific roles in the toxicity of compounds. The recognition and analysis of such local shape features is essential in attempts to construct valid models and propose explanations for specific toxic activity. These tasks can be approached by the MEDLA method, if a sufficiently high resolution electron density fragment database is available.

In all the computations reported here the 6-31G** Gaussian basis set was used for each parent molecule, ensuring a uniform quality for all the electron densities in the resulting PAH MEDLA database. Each electron density fragment in the PAH database contains the nucleus of either a single aromatic carbon of a ring (C fragment), or the nuclei of a carbon and a hydrogen atom attached to it (CH fragment). The effects of the various possible surroundings of these C and CH fragments are taken into account as distortions in the electron density caused by formal coordination shells around the fragments. For a classification scheme, the "depth" of a formal coordination shell, representing the local neighborhood of the fragment within the parent (and the target) molecules is used. We have generated three levels, depending on the number of nearest neighbors used to represent the distorting effects of the surroundings on the C and CH fragments.

Level 1 fragments are those where only the nearest neighbors directly bonded to the carbon atom of the fragment are considered. There are only two types of level 1 fragments (Figure 1): one containing a single carbon nucleus (C fragment), the other the nuclei of a carbon and a hydrogen atom (CH fragment).

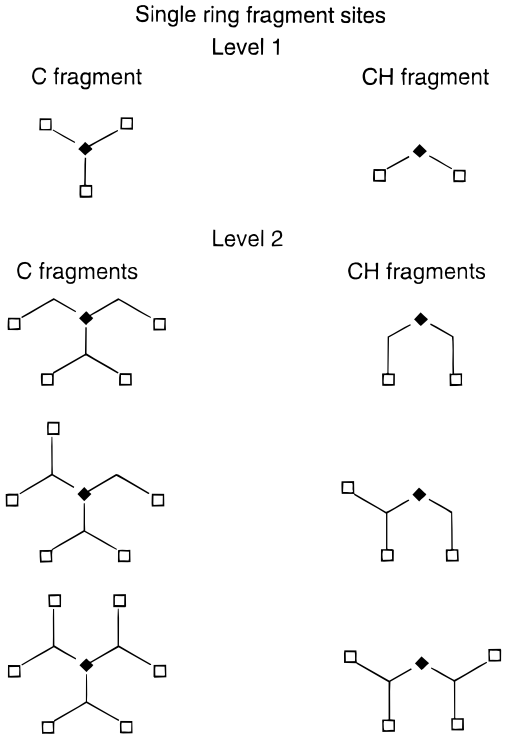

Figure 1. Definitions of "level 1" and "level 2" fragments. The diamond symbol stands for the nucleus of the carbon atom belonging to the fragment, the square symbol marks the most distant atoms explicitly included in the environment of the carbon fragment.

The level 2 fragments take into account the effects of the first and second nearest neighbors (Figure 1). The differences in second nearest neighbors increase the variety of fragment types to six.

The level 3 fragments account for the effects of the first, second and third closest neighbors. This criterion leads to further distinctions among the possible fragment types; in the level 3 database there are 58 different PAH fragments, as illustrated in Figure 2. An important feature used for dividing level 3 fragments into groups is the number of aromatic rings involving a given fragment (one, two, or three).

In Figure 3, three MIDCOs of the parent molecule phenanthrene are shown as examples for molecular isodensity contour surfaces, where the thresholds for the density contours are $a = 0.1$, 0.01, and 0.001 au, respectively. Note that the electronic density is computed for the whole range of density thresholds, not only for those displayed in the figure.

In the notations used for the fragments, the letter symbols ben, naf, ant, fen, pyr, and fyl refer to the parent molecules benzene, naphthalene, anthracene, phenanthrene, pyrene, and phenalenyl, whereas the numbers assigned to carbon atoms follow the standard rules used in Chemical Abstracts. The notations used for these fragments are displayed in Table 1, where the serial numbers of the actual fragment carbon and two additional carbons are also specified. The nuclear locations of these three "anchor" atoms are used in the positioning and orientation of fragment densities when building electron densities for PAHs.

In Figure 4, MIDCOs of two carbon fragments, fen4a and fen8a, and two CH fragments, fen2 and fen9, obtained from

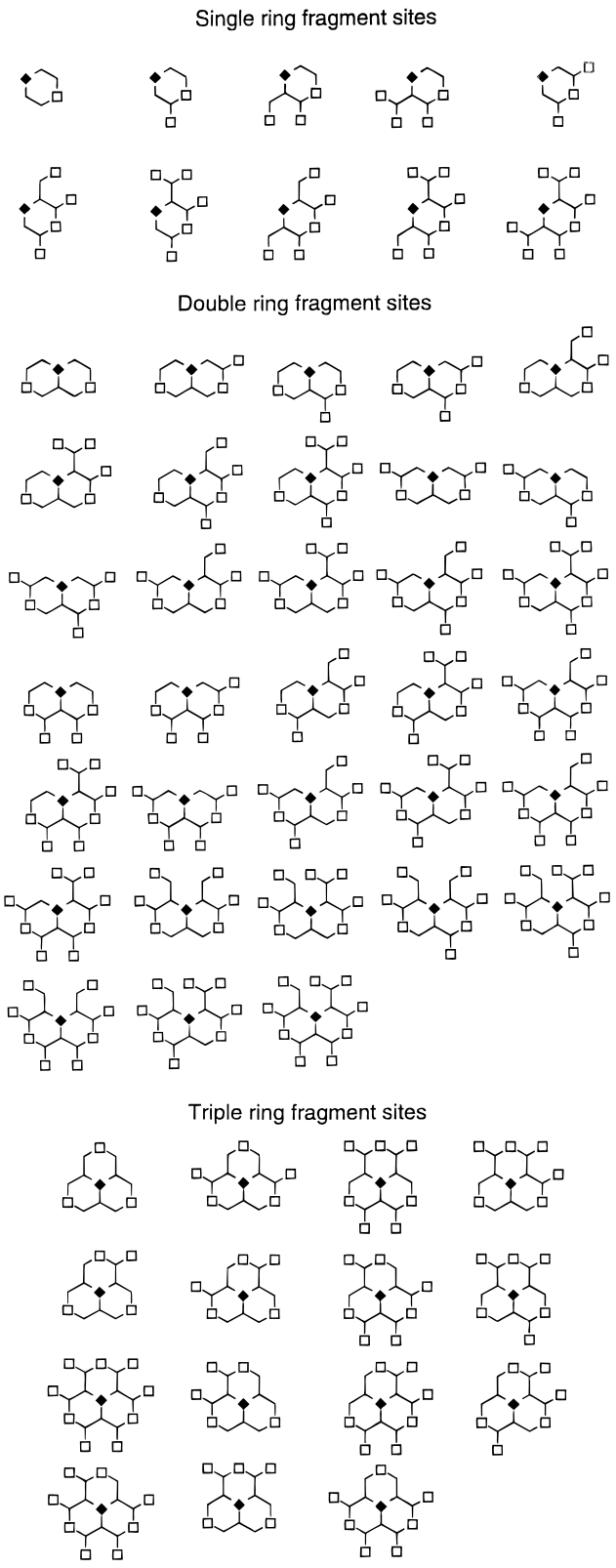

Figure 2. Definitions of the 58 "level 3" fragments, belonging to one (a), two (b) or three (c) aromatic rings. The diamond symbol stands for the nucleus of the carbon atom belonging to the fragment, the square symbol marks the most distant atoms explicitly included in the environment of the carbon fragment.

the parent molecule phenanthrene are displayed at the same three density threshold values, $a = 0.1$, 0.01, and 0.001 au. For both the fen2 and fen9 series of MIDCOs shown in Figure 4, the formal "bond" lines interconnecting the C and H nuclei are vertical, hence parallel, nevertheless, the

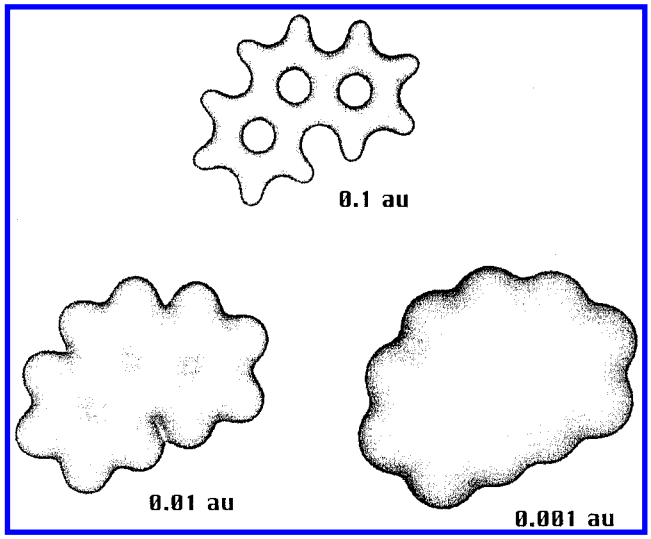

Figure 3. Three MIDCOs of parent molecule phenanthrene with density thresholds of $a = 0.1$, 0.01, and 0.001 au, respectively. The electronic density is computed for the whole range of density thresholds, not only for those displayed in the figure.

Table 1. Labels and "Anchor" Atoms of the 24 Electron Density Fragments Obtained from the Parent Molecules Benzene, Naphthalene, Anthracene, Phenanthrene, Pyrene, and Phenalenyl[a]

molecule	label	anchors
benzene	ben1	1-2-6
naphthalene	naf1	1-2-8a
	naf2	2-1-3
	naf4a	4a-4-5
anthracene	ant1	1-2-9a
	ant2	2-1-3
	ant4a	4a-9a-10
	ant9	9-8a-9a
phenanthrene	fen1	1-2-10a
	fen2	2-1-3
	fen3	3-2-4
	fen4	4-3-4a
	fen4a	4a-4-10a
	fen8a	8a-4b-8
	fen9	9-8a-10
pyrene	pyr1	1-2-10a
	pyr2	2-1-3
	pyr3a	3a-3-4
	pyr4	4-3a-5
	pyr10b	10b-3a-10a
phenalenyl	fyl1	1-2-9a
	fyl2	2-1-3
	fyl3a	3a-3-4
	fyl9b	9b-3a-6a

[a] The first anchor atom is the carbon of the fragment that may or may not carry a hydrogen; the other two anchor atoms specified are used for the positioning and orientation of the density fragment when building PAH densities. See also Figure 5.

MIDCOs at different density thresholds appear tilted with respect to one another. These differences arise as the consequence of the different surroundings within the parent molecules affecting the high and low density regions of the fragments in different ways. As such, the shapes of the two carbon fragments, as well as the shapes of the two CH fragments, are rather different. This example illustrates the subtle but important differences in the electron density distributions that depend on both the local nuclear arrangement and the surroundings. Evidently, the *location* of a given CH electron density fragment within a molecule has an important influence on the actual shape of the electron density of that fragment.

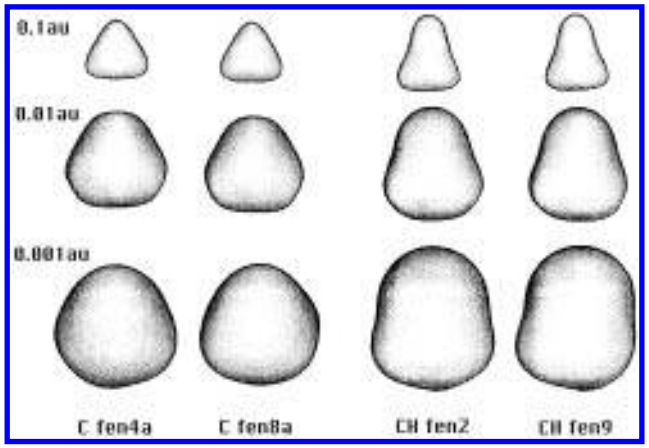

Figure 4. Three MIDCOs (displayed at the density threshold values, a = 0.1, 0.01, and 0.001 au) for each of two carbon fragments and two CH fragments obtained from the parent molecule in Figure 3. There are subtle but essential differences in the electron density distributions depending on the local nuclear arrangement and the surroundings. In the fen2 and fen9 series of MIDCOs, the CH bonds are vertical (hence parallel), but the MIDCOs appear tilted with respect to one another, showing the effects of different surroundings.

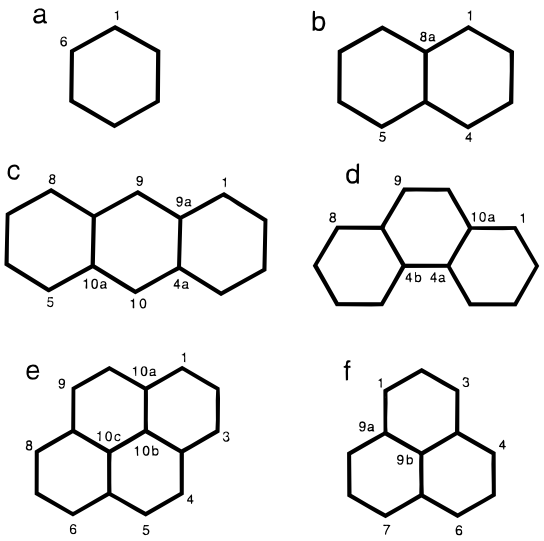

Figure 5. Serial numbers of carbon atoms in the parent molecules (a) benzene, (b) naphthalene, (c) anthracene, (d) phenanthrene, (e) pyrene, and (f) phenalenyl, used for the development of the core of the PAH shape-fragment database.

In general, one can define higher level fragments taking an arbitrary number of neighbors into account. However, the number of fragment types will rise sharply with an increase of the number of neighbors. The size of the parent molecules also increases considerably; note that for the generation of all level 3 fragments one already needs to perform some computations on six-ring PAHs. Based on pilot studies on the recommended size of parent molecules, the level 3 fragments are expected to perform well for our present purposes.

For the development of the core of the PAH shape-fragment database, we chose the following six PAHs as parent molecules: the closed shell systems of benzene, naphthalene, anthracene, phenanthrene, and pyrene, and the open shell system of the phenalenyl radical (Figure 5). The nuclear arrangements were optimized using Gaussian 90 program[27] and the high-quality 6-31G** AO basis set, employing the Gaussian 90 implementations of the RHF

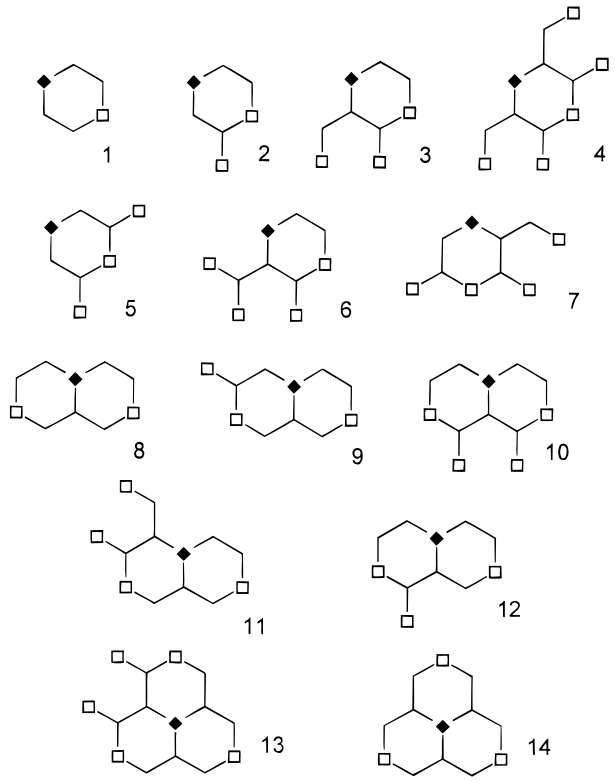

Figure 6. The 14 types of "level 3" fragments obtained from parent molecules benzene (#1), naphthalene (#2, 3, and 8), anthracene (#2, 3, 4, and 9), phenanthrene (#2, 3, 6, 7, 11, and 12), pyrene (#3, 5, 7, 10, and 13), and phenalenyl (#3, 5, 10, and 14).

(Restricted Hartree-Fock) algorithm for the closed shell systems and the UHF (Unrestricted Hartree-Fock) algorithm for the phenalenyl radical.

The six parent molecules were used to generate 24 nonequivalent fragments. Table 1 contains the full list of these fragments. Each fragment is named according to the parent molecule it was taken from and the serial number of the fragment carbon in the parent molecule. The serial numbers of the fragment carbon and two of the nearest carbons are also specified as "anchor" points, used as reference points to determine the position and orientation of the fragment within the original parent molecule. These 24 fragments fall into 14 categories according to the level 3 classification scheme (Figure 6). In Table 2, all carbons of the selected six parent compounds are characterized using the same classification scheme.

TESTS OF THE QUALITY OF THE PAH SHAPE-FRAGMENT DATABANK

In order to evaluate the quality of the computed electron density fragments of the PAH shape fragment databank, several tests involving 4-ring PAHs were performed. The first set of tests was used to compare the results of high-quality, conventional electron density calculations for PAHs and the PAH electron densities obtained from MEDLA computations based on the optimum selection of fragments from the PAH shape-fragment database. The 6-31G** *ab initio* electron densities of naphthacene, benz[*a*]anthracene, chrysene, benzo[*c*]phenanthrene, triphenylene, and benz[*de*]-anthracenyl, as calculated with the conventional *ab initio* method using GAUSSIAN 90, were compared to the electron densities obtained using the MEDLA technique and the new

Table 2. "Level 3" Fragments Derived from Six PAH Parent Molecules Benzene, Naphthalene, Anthracene, Phenanthrene, Pyrene, and Phenalenyl[a]

fragment	benzene	naphthalene	anthracene	phenanthrene	pyrene	phenalenyl
1	1,2,3,4,5,6					
2		2,3,6,7	2,3,6,7	2,3,6,7		
3		1,4,5,8	1,4,5,8	1,8	1,3,6,8	1,3,4,6,7,9
4			9,10			
5					2,7	2,5,8
6				4,5		
7				9,10	4,5,9,10	
8		4a,8a				
9			4a,8a,9a,10a			
10					3a,5a,8a,10a	3a,6a,9a
11				4a,4b		
12				8a,10a		
13					10b,10c	
14						9b

[a] Each box of the Table lists the carbons defining equivalent fragments. See also Figure 5.

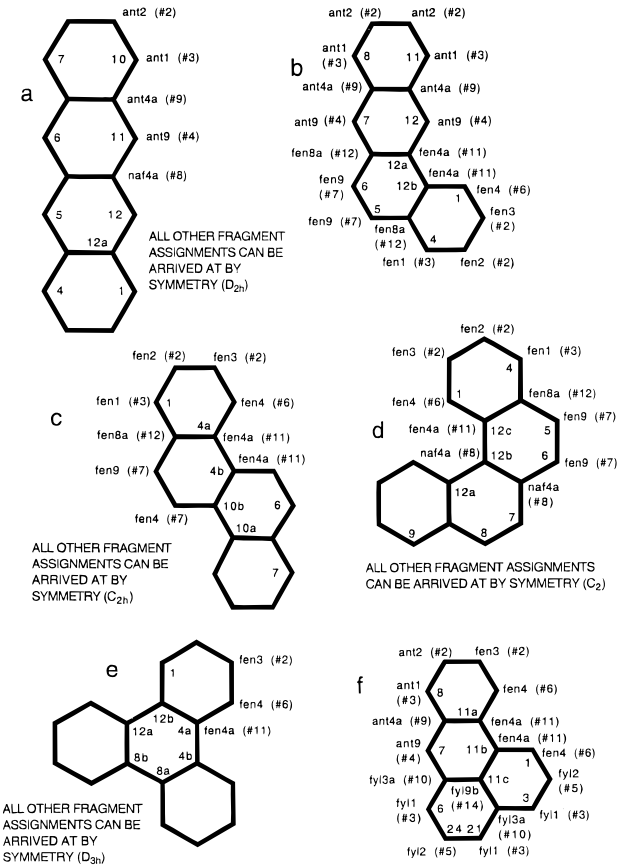

Figure 7. Choice of PAH database fragments for test calculations for (a) naphthacene, (b) benz[*a*]anthracene, (c) chrysene, (d) benzo[*c*]phenanthrene, (e) triphenylene, and (f) benz[*de*]anthracenyl. Fragment labels (with fragment type in parentheses) are specified along the periphery of each structural diagram, whereas nontrivial atom labels for each molecule are given within the hexagons.

PAH shape-fragment database. The fragments selected from the PAH shape fragment database for each of these molecules are shown in Figure 7. Three types of similarity measures were used in the comparisons; the first one involved a direct comparison of densities, the second measure was based on the topological shape codes, whereas the third similarity measure was based on the relative volumes enclosed by isodensity contours.

The simplest, direct comparison of electron densities using point-by-point comparisons of electron densities for a set of grid points and the average relative difference between the corresponding density values, generates the L-measure, described in detail earlier.[23] These similarity measures can be evaluated for any given range of electron density thresholds as well as for the whole range of the electron density. The direct similarity measures L calculated for various density ranges of the five test PAH molecules, naphthacene, benz[*a*]anthracene, chrysene, benzo[*c*]phenanthrene, triphenylene, and benz[*de*]anthracenyl, are shown in Table 3. The largest relative deviation was found for benzo[*c*]phenanthrene, a molecule where the ring structure is apparently rather different from those of the parent molecules, a finding that justifies adding further, special, custom-made fragments to the PAH shape-fragment database. Nevertheless, even for this molecule, the comparisons demonstrate that reasonably accurate molecular shapes can be generated with the core set of the PAH shape-fragment database. These comparisons indicate the level of reliability of using MEDLA densities for PAH shape analysis, confirming earlier, more general findings, where the Shape Group and MEDLA methods have been validated for the purposes of drug design.[20,21,26]

Table 3. Direct (Point-by-Point) Similarity Measures (in %) of Electron Densities of the 4-Ring PAHs Naphthacene, Benz[*a*]anthracene, Chrysene, Benzo[*c*]phenanthrene, Triphenylene, and Benz[*de*]anthracenyl (Figure 7), Derived from Conventional *ab Initio* Computations and from MEDLA Calculations for a Variety of Ranges of the Electron Density[a]

	1.−0.1	0.1−0.01	0.01−0.001
naphthacene	99.18	98.84	98.65
benz[*a*]anthracene	99.21	98.84	98.46
benzo[*c*]phenanthrene	98.24	97.09	96.29
chrysene	99.27	98.99	98.63
triphenylene	99.16	98.41	97.29
benz[*de*]anthracenyl	98.37	97.72	96.95

[a] Ranges in au, atomic units.

The second type of similarity measure was based on the topological shape codes, using (*a*,*b*)-map comparisons. This measure is very sensitive to small changes, since small shifts of the boundaries of various topological regions within the (*a*,*b*)-maps already can cause mismatches. These similarity measures usually exaggerate the effects of small shape changes of electron density. Accordingly, the results of (*a*,*b*)-map comparisons for the series of test molecules, naphthacene, benz[*a*]anthracene, chrysene, benzo[*c*]phenanthrene, triphenylene, and benz[*de*]anthracenyl, listed in Table 4,

Table 4. Similarities of Topological (*a*,*b*)-Maps of the 4-Ring PAH Molecules Naphthacene, Benz[*a*]anthracene, Chrysene, Benzo[*c*]-phenanthrene, Triphenylene, and Benz[*de*]anthracenyl (Figure 7), Derived from *ab Initio* Electron Densities and MEDLA Calculations

compound	similarity of (*a*,*b*)-maps
naphthacene	0.8086
benz[*a*]anthracene	0.8373
benzo[*c*]phenanthrene	0.7768
chrysene	0.8439
triphenylene	0.8755
benz[*de*]anthracenyl	0.8400

Table 5. MIDCO-Volume Similarity Measures (in %) of Electron Densities for Six 4-Ring PAHs (Figure 7), Derived from Conventional *ab Initio* Computations and from MEDLA Calculations for a Variety of Ranges of the Electron Density[a]

	1.−0.1	0.1−0.01	0.01−0.001
naphthacene	99.57	99.47	99.28
benz[*a*]anthracene	99.56	99.46	99.20
benzo[*c*]phenanthrene	99.04	98.67	98.28
chrysene	99.62	99.50	99.21
triphenylene	99.53	99.21	98.61
benz[*de*]anthracenyl	99.17	98.93	98.52

[a] Ranges in au, atomic units.

show significant deviations from the ideal value of 1.00 that corresponds to a perfect match. The largest deviation was again obtained for benzo[*c*]phenanthrene, a finding consistent with those involving the direct density comparisons.

The third type of similarity measure applied to the test molecules, the least sensitive of the three measures used, is based on the comparison of the relative volumes enclosed by MIDCO surfaces of identical density thresholds (Table 5). As with the first two measures, the largest deviation was obtained for benzo[*c*]phenanthrene. Whereas this measure is not as sensitive as direct density comparison, the numerical values provide an easily visualizable indication of the degree of accuracy of the MEDLA-based electron densities obtained with the new PAH shape-fragment database.

All three tests indicate that PAH electronic densities obtained using the new shape fragment databank within the MEDLA framework, and the electronic densities obtained by the conventional *ab initio* 6-31G** method are nearly equivalent. For the purposes of shape analysis, the MEDLA-based electron densities can be used to replace the electron densities calculated by standard *ab initio* 6-31G** method.

A second set of tests was used to compare MEDLA results obtained using optimum and various nonoptimum selections of density fragments from the PAH shape-fragment database. This test involved comparisons of conventional *ab initio* electron densities obtained directly with densities obtained using different fragment selections within the MEDLA framework.

The purpose of the second set of tests was to study the importance of the quality and choice of the fragments, that is, to determine the degree of sensitivity of the computed electron densities to the accuracy of the local density fragments. The molecule chosen for these tests was triphenylene, for which four different sets of fragments (medla (i) to (iv)) were selected (Figure 8). Medla(i) corresponds to the optimum fragment choice of level 3 fragments; medla-(ii) uses only three fragments, naf1, naf2, and naf4a, obtained from naphthalene as parent molecule; medla(iii) uses only two fragments, ben1, from benzene as parent molecule, and naf4a, obtained from naphthalene as parent molecule; and

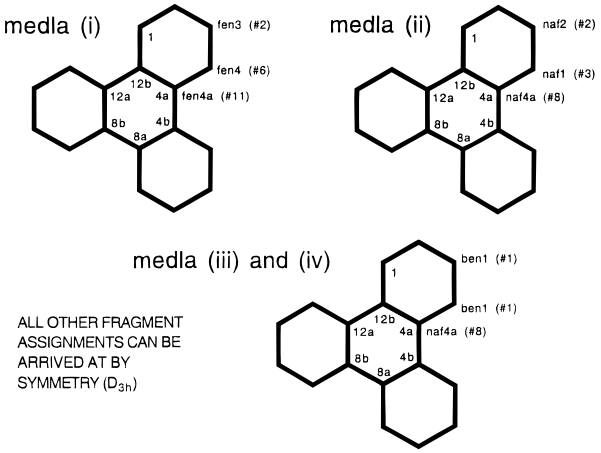

Figure 8. Three different fragment choices, medla(i)−medla (iii), for tests involving the triphenylene molecule. Medla(i): optimum level 3 fragments; medla(ii): naf2, naf3, naf8; medla(iii): ben1, naf8. The fourth choice, medla(iv), involves the same fragments as medla(iii), but with molecular mechanics geometry. Fragment labels (with fragment type in parentheses) are specified along the periphery of structural diagrams, whereas nontrivial atom labels for the molecule are given within the hexagons.

Table 6. Test of the Importance of Fragment Quality: Direct, Point-by-Point Similarity Measures (in %) of Calculated Electron Densities of Triphenylene Molecule Are Derived, Comparing Conventional *ab Initio* Computations and MEDLA Calculations Using Four Different Sets of Fragments, medla(i)−medla(iv) (Figure 8) for a Variety of Ranges of the Electron Density

	1.−0.1	0.1−0.01	0.01−0.001
medla(i)	99.16	98.41	97.29
medla(ii)	98.29	96.93	95.84
medla(iii)	98.34	96.74	95.16
medla(iv)	77.76	69.62	61.41

[a] Ranges in au, atomic units. Medla(i): optimum Level 3 fragments; medla(ii): lower quality set using only naf2, naf3, naf8; medla(iii): uses only ben1, naf8; medla(iv), uses only ben1, naf8 and molecular mechanics geometry

medla(iv) uses the same type fragments as medla(iii), but created with a nonoptimum nuclear geometry of the parent molecules obtained by the molecular mechanics algorithm.[30] These computations represent four levels of quality, from the best, medla(i) to the worst, medla(iv), suitable only for rough estimates. The three similarity measures used are the direct (L) similarity measures of the electron densities based on point-by-point comparisons of relative densities, the topological similarity measure defined by the overlap of the (*a*,*b*)-maps, and the similarity measure based on the relative volumes of MIDCOs.

The results of the direct, point-by-point comparisons of relative densities are shown in Table 6. Whereas all three of the medla(i)−medla(iii) fragment choices generate similarities with the conventional *ab initio* electron densities of at least 95% accuracy, as expected, the medla(i) choice was consistently better than any of the other choices, throughout the entire density range considered. The differences between fragment choices medla(i), medla(ii), and medla(iii) were small but significant. The effects of nonoptimum choice of the nuclear geometry, however, was the single most important factor affecting accuracy; the medla(iv) results based on molecular mechanics geometry were unsatisfactory, especially, when compared to the high quality medla(i) results.

Table 7. Test of the Importance of Fragment Quality: The Similarities of Topological (*a*,*b*)-Maps of Calculated Electron Densities of the Triphenylene Molecule Comparing Conventional *ab Initio* Computations and MEDLA Calculations Using Four Different Sets of Fragments, medla(i)−medla (iv) (Figure 8)[a]

	ab initio	medla (i)	medla (ii)	medla(iii)	medla (iv)
ab initio	1.0000	0.8755	0.8661	0.8302	0.5070
medla(i)		1.0000	0.8492	0.8367	0.5097
medla(ii)			1.0000	0.8019	0.5110
medla(iii)				1.0000	0.5199
medla(iv)					1.0000

[a] Medla(i): optimum Level 3 fragments; medla(ii): lower quality set using only naf2, naf3, naf8; medla(iii): uses only ben1, naf8; medla(iv), uses only ben1, naf8 and molecular mechanics geometry.

Table 8. Test of the Importance of Fragment Quality: MIDCO-Volume Similarity Measures (in %) of the Calculated Electron Densities of Triphenylene Molecule Comparing Conventional *ab Initio* Computations and MEDLA Calculations Using Four Different Sets of Electron Density Fragments, medla(i)−medla (iv) (Figure 8) for a Variety of Ranges of the Electron Density[a]

	1.−0.1	0.1−0.01	0.01−0.001
medla(i)	99.53	99.21	98.61
medla(ii)	98.93	98.63	98.06
medla(iii)	99.01	98.46	97.58
medla(iv)	83.64	80.82	74.58

[a] Ranges in au, atomic units. Medla(i): optimum Level 3 fragments; medla(ii): lower quality set using only naf2, naf3, naf8; medla(iii): uses only ben1, naf8; medla(iv), uses only ben1, naf8 and molecular mechanics geometry.

The second similarity measure, based on the topological (*a*,*b*)-maps, led to similar comparisons (Table 7). This, rather sensitive similarity measure also ranked medla(i) as having the highest degree of similarity with the conventional *ab initio* result, with the accuracy monotonically decreasing in the medla(i)−medla (iv) sequence. The molecular mechanics geometry used in the medla(iv) computations was clearly a disadvantage when compared to the more accurate medla-(i)−medla (iii) results (Table 7).

The results of the least sensitive test, based on the relative volumes of MIDCOs, showed the same general trends as the other two tests (Table 8).

The tests performed indicate that reasonably accurate shape description of PAHs can be achieved using the shape group methods in combination with the new PAH shape fragment databank as applied within the MEDLA technique, if the electron density construction is based on the level 3 electron density fragments and a consistent set of nuclear coordinates.

SUMMARY

A high-resolution electron density fragment databank was developed for PAH molecules. The PAH shape fragment databank is designed for applications within the QShAR (quantitative shape-activity relations) and the QShTAR (quantitative shape-toxicological activity relations) framework for toxicological risk assessment. The tests reported indicate that the shapes of PAH molecular electron densities, computed and evaluated by a combination of the Shape Group and MEDLA methods, can be represented with satisfactory accuracy using the new PAH shape fragment databank.

Most PAHs and their derivatives show a range of toxic effects on bacteria, plants, animals and humans. An important component of environmental damage caused by PAH release through oil spills, combustion of fossil fuels or industrial activity arises from the direct toxicity of PAHs or their products, particularly photooxidation products such as epoxides and orthoquinones. Many of the smaller PAHs, *e.g.*, methyl substituted naphthalenes, are the major toxicants for bacteria and plants. Conversely, pyrene and its derivatives, benzo[*a*]pyrene, as well as various substituted and fused anthracene and phenanthrene derivatives, are all important toxicants. The molecular shape of various isomeric forms strongly influences toxicity, leading to altered activities of the different molecules. It is likely that in the family of PAH molecules, shape and toxicological risk correlations are of major importance. Therefore, these molecules have been selected as subjects for the first systematic QShTAR project.

The shapes of PAHs (the shapes of their electronic charge clouds) have a fundamental role in all of the initial chemical reactions responsible for their toxicity. In particular, their chemical transformations initiated by photochemically induced electronic excitation, such as photooxidation processes, are influenced by the shape features of PAHs. Known and currently researched toxicological activities for PAHs can be correlated with results of detailed shape analysis, providing a predictive tool for assessing toxicological risks for related PAH molecules not tested yet experimentally. The construction of the PAH shape fragment database reported here will provide high quality shape information for such analysis.

ACKNOWLEDGMENT

This study has been supported by the Canadian Network of Toxicology Centers and by a research grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES AND NOTES

(1) *Concepts and Applications of Molecular Similarity;* Johnson, M. A.; Maggiora, G. M., Eds.; Wiley: New York, 1990.
(2) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Chapman & Hall Blackie Publishers: Glasgow, 1995.
(3) *Molecular Similarity, Topics in Current Chemistry, Vol. 173*; Sen, K., Ed.; Springer-Verlag: Heidelberg, 1995.
(4) *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*; Carbo, R., Ed.; Kluwer Academic Publishers: Dordrecht, 1995.
(5) Martin, Y. C. *Quantitative Drug Design: A Critical Introduction;* Dekker: New York, 1978.
(6) Richards, W. G. *Quantum Pharmacology;* Butterworths: London, 1983.
(7) Franke, R. *Theoretical Drug Design Methods;* Elsevier: Amsterdam, 1984.
(8) Dean, P. M. *Molecular Foundations of Drug-Receptor Interaction;* Cambridge University Press: New York, 1987.
(9) Liebman, M. N.; Venanzi, C. A.; Weinstein, H. *Biopolymers* **1985**, *24*, 1721.
(10) Tapia, O.; Eklund, H.; Brändén, C. I. Molecular, Electronic, and Structural Aspects of the Catalytic Mechanism of Alcohol Dehydrogenase. In *Steric Aspects of Biomolecular Interactions;* Náray-Szabó, G., Simon, K., Eds.; CRC Press: West Palm Beach, FL, 1987.
(11) Mezey, P. G. *Int. J. Quant. Chem. Quant. Biol. Symp.* **1986**, *12*, 113.
(12) Mezey, P. G. *J. Comput. Chem.* **1987**, *8*, 462.
(13) Mezey, P. G. *J. Math. Chem.* **1988**, *2*, 299.
(14) Mezey, P. G. Three-Dimensional Topological Aspects of Molecular Similarity. In *Concepts and Applications of Molecular Similarity;* Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.
(15) Mezey, P. G. Topological Quantum Chemistry. In *Reports in MolecularTheory;* Náray-Szabó, G., Weinstein, H., Eds.; CRC Press: Boca Raton, FL, 1990.

(16) Mezey, P. G. Molecular Surfaces. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publ.: New York, 1990.
(17) Mezey, P. G. Dynamic Shape Analysis of Biomolecules Using Topological Shape Codes. In *The Role of Computational Models and Theories in Biotechnology;* Bertran, J., Ed.; Kluwer Academic Publishers: Dordrecht, 1992.
(18) Mezey, P. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 650.
(19) Mezey, P. G. *Shape in Chemistry: Introduction to Molecular Shape and Topology;* VCH Publishers: New York, 1993.
(20) Walker, P. D.; Maggiora, G. M.; Johnson, M. A.; Petke, J. D.; Mezey, P. G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 568.
(21) Walker, P. D.; Mezey, P. G.; Maggiora, G. M.; Johnson, M. A.; Petke, J. D. *J. Comput. Chem.* **1995**, *16*, 1474.
(22) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 12423.
(23) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1994**, *116*, 12022.
(24) Walker, P. D.; Mezey, P. G. *Can. J. Chem.* **1994**, *72*, 2531.
(25) Walker, P. D.; Mezey, P. G. *J. Math. Chem.* **1995**, *17*, 203.
(26) Walker, P. D.; Mezey, P. G.. *J. Comput. Chem.* **1995**, *16*, 1238.
(27) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M. A.; Binkley, J. S.; González, C.; DeFries, D. J.; Fox, D. J.;Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. A. *Program GAUSSIAN 90;* Carnegie-Mellon Quantum Chemistry Publishing Unit: Pittsburgh, PA, 1990.
(28) Walker, P. D.; Arteca, G. A.; Mezey, P. G. *Program GSHAPE 90;* Mathematical Chemistry Research Unit, University of Saskatchewan: Saskatoon, Canada, 1990.
(29) Walker, P. D.; Mezey, P. G. *Program MEDLA 93;* Mathematical Chemistry Research Unit, University of Saskatchewan: Saskatoon, Canada, 1993.
(30) BIOGRAF; Biodesign, Inc.: 199 S. Los Robles Ave., Pasadena, CA 91101, 1988.
(31) Mezey, P. G. *J. Math. Chem.* **1995**, *18*, 141.
(32) Mezey, P. G. Local Shape Analysis of Macromolecular Electron Densities. In *Computational Chemstry: Reviews and Current Trends;* Leszczynski, J., Ed.; World Scientific Publishers: Singapore, 1996.
(33) Mezey, P. G. Descriptors of Molecular Shape in 3D. In *3D Molecular Shape and Graph Theory;* Balaban, A. T., Ed.; Plenum Press: New York, 1996.
(34) Newsted, J. L.; Giesy, J. P. *Environ. Toxicol. Chem.* **1987**, *6*, 445.
(35) Huang, X.-D.; Dixon, D. G.; Greenberg, B. M. *Environ. Toxicol. Chem.* **1993**, *12*, 1067.
(36) Ren, L.; Zeiler, L. F.; Dixon, D. G.; Greenberg, B. M. *Ecotox. Environ. Saf.* Submitted for publication.
(37) Nicolaou, K.; Masclet, P.; Mouvier, G. *Sci. Total Environ.* **1984**, *32*, 103.
(38) Yang, S. K.; McCourt, D. W.; Roller, P. R.; Gelboin, H. V. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2594.
(39) Harvey, R. G.; Cortez, C.; Sugiyama, T.; Ito, Y.; Sawyer, T. W.; DiGiovanni, J. *J. Med. Chem.* **1988**, *31*, 154.
(40) Ren, L.; Huang, X.-D.; McConkey, B. J.; Dixon, D. G.; Greenberg, B. M. *Ecotox. Environ. Saf.* **1994**, *28*, 160.
(41) Huang, X.-D.; Dixon, D. G.; Greenberg, B. M. *Ecotox. Environ. Saf.* **1995**, submitted
(42) Katz, M.; Chan, C.; Tosine, H.; Sakuma, T. Relative Rates of Photochemical and Biological Oxidation (*in vitro*) of Polynuclear Aromatic Hydrocarbons. In *Polynuclear Aromatic Hydrocarbons*; Jones, P. W., Leber, P., Ed.; Ann Arbor Science Publishers: Ann Arbor, MI, 1979.
(43) Ankley, G. T.; Collyard, S. A.; Monson, P. D.; Kosian, P. A. *Environ. Toxicol. Chem.* **1994**, *13*, 1791.
(44) Morgan, D. D.; Warshawsky, D.; Atkinson, T. *Photochem. Photobiol.* **1977**, *25*, 31.
(45) Foote, C. S. *Photochem. Photobiol.* **1991**, *54*, 659.

CI9501610