Towards a comprehensive electronic database of polycyclic aromatic hydrocarbons and its application in constraining the identities of possible carriers of the diffuse interstellar bands

Xiaofeng Tan*

X Scientific, Inc., 1000 Kiely Boulevard Unit 90, Santa Clara, CA 95051, United States

ABSTRACT

A theoretical approach is developed to pre-select individual polycyclic aromatic hydrocarbons (PAHs) as possible carriers of the diffuse interstellar bands (DIBs). In this approach, a computer program is used to enumerate all PAH molecules with up to a specific number of fused benzene rings. Fast quantum chemical calculations are then employed to calculate the electronic transition energies, oscillator strengths, and rotational constants of these molecules. An electronic database of all PAHs with up to any specific number of benzene rings can be constructed this way. Comparison of the electronic transition energies, oscillator strengths, and rotational band contours of all PAHs in the database with astronomical spectra allows one to constrain the identities of individual PAHs as possible carriers of some of the intense narrow DIBs. Using the current database containing up to 10 benzene rings we have pre-selected 8 closed-shell PAHs as possible carriers of the famous λ6614 DIB.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The diffuse interstellar bands (DIBs) are absorption features spanning the near ultraviolet (UV) to near infrared (IR) spectral range seen in the spectra of stars obscured by diffuse interstellar clouds. Since their discovery in 1922 by Heger [1], the identification of the carriers of these bands has been a long-standing challenge for scientists of the 20th century. In the past two decades, the hypothesis that free gas-phase polycyclic aromatic hydrocarbons (PAHs) may carry some of the DIBs has become a consensus—which is known as the PAH-DIB hypothesis [2–6]. Pieces of evidence that support this hypothesis include the rich interstellar elemental abundances of carbon and hydrogen, the high photo stability of PAHs, the identification of PAHs in meteorites [7], and most importantly, the astonishing similarity between the IR signatures of PAHs and the unidentified infrared (UIR) emission bands [8–14]. Unfortunately, due to the insensitivity of IR spectroscopy to the overall structure of the molecule, no individual PAHs can be identified by the comparison of astronomical spectra with IR spectra of PAHs.

To assess the validity of the PAH–DIB hypothesis, electronic spectra of various neutral and ionized PAHs have been measured under conditions that mimic those found in the interstellar clouds where the DIBs originate. Earlier experimental data were primar-

ily obtained with matrix-isolation spectroscopy (MIS) due to its capability of trapping neutral and ionized PAHs in cryogenic inert-gas solid matrices [15,16]. The spectral features recorded with MIS, however, are shifted and broadened due to the interaction of the trapped species with the solid lattice. This shortcoming was overcome by the application of cavity ring-down spectroscopy (CRDS) [17–20]. Only a limited number of PAHs, however, have been studied by CRDS. This is due to the difficulty of bringing these low-vapor pressure molecules into the gas phase in sufficient amounts to allow detection under the laboratory conditions that are relevant for astrophysical studies [21–29]. Notwithstanding all these efforts, no individual PAHs have ever been identified as DIB carriers.

Recent high-resolution astronomical observations have shown partially resolved rotational band contours of several strong DIBs, including the λ5797, λ5850, λ6196, λ6379, and λ6614 DIBs [30–32]. The rotational band contours of these DIBs place additional strict constraints and provide crucial information on the properties of the carriers of these DIBs. In theory, one can survey the electronic spectra of all PAHs to identify possible matches with the DIB spectra provided that these DIBs are indeed caused by PAHs. In practice, however, this approach is not yet possible since only a limited number of PAHs have been studied either experimentally or theoretically. As a matter of fact, only several dozens of PAHs with up to 10 fused benzene rings have been studied, which accounts for only ca. 0.1% of all possible PAHs within the size range. Essentially, there can be infinite number of PAHs out there. But for a certain diffuse band, only a limited number of PAHs are relevant—the size range

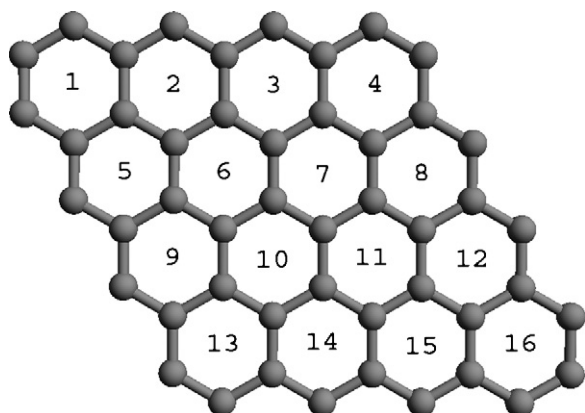* Tel.: +1 408 244 3568.
E-mail address: x.tan@jhu.edu.

Fig. 1. A 4 × 4 block honeycomb grid. The grid points are not shown on the figure but labeled numerically from left to right and from top to bottom.

of the relevant PAHs can be constrained from the rotational band contours. A big question therefore is: "how can one determine all relevant PAHs for a certain diffuse band and pre-select promising candidates that should be studied first?" In order to address this question, a systematic method of enumerating all relevant PAHs and calculating their spectroscopic properties (electronic transition energies, oscillator strengths, and rotational constants) is thus highly desirable. In this paper, we present a theoretical method – which involves computer enumeration of PAHs, fast calculations of the electronic transition energies, oscillator strengths, and rotational constants of these PAHs, and pre-selection of possible DIB carriers through the search of the generated electronic database of PAHs – to address this difficult problem.

This paper is organized as follows: in Section 2, a method of enumerating all PAHs with up to a specific number of benzene rings is given. In Section 3, the methods for fast calculations of the electronic transition energies, oscillator strengths, and ground-state rotational constants of the generated PAHs are presented. An electronic database of PAHs can be generated this way. In Section 4, we pre-select 8 closed-shell PAHs as possible carriers of the λ6614 DIB. Further discussions are given in Section 5.

2. Computer enumeration of PAHs

The prerequisite for building a comprehensive database of PAHs is to enumerate all PAHs with a specific number of fused benzene rings or "cells". To do so, one needs an abstract way to represent the carbon skeleton of a PAH. The ideal carbon skeleton of a PAH is referred as a polyhex [33]. By "ideal", we assume all benzene rings in a PAH are identical regular hexagons. Given a polyhex, one can easily add the missing hydrogen atoms to form a PAH molecule and then optimize its geometry by quantum chemical calculations.

Several algorithms exist for enumerating planar polyhexes [34–38]. The method we use is based on the encoding of planar polyhexes in a two-dimensional honeycomb grid. A 4 × 4 block honeycomb grid is shown in Fig. 1. In this grid, each cell is represented by one grid point located in its center of mass. Each cell can be identified by its unique cell number or equivalently its row and column indices. A polyhex can be encoded by its constituent cells in the honeycomb grid. If two cells are joined together by a common side, we call the two cells or the corresponding two grid points "connected". Each cell can connect up to six cells in an indefinite honeycomb grid. A straightforward way of enumerating all planar polyhexes with h cells is to start with all polyhexes with $h - 1$ cells and add to each of them a connecting cell to produce polyhexes with h cells. This cell-growth approach, however, produces duplicated polyhexes. We use

the symmetry property of the honeycomb grid to eliminate enumeration duplications. For an infinite two-dimensional honeycomb grid, there are six rotation operations and six vertical mirror reflections that transform a cell into another cell. In addition to these twelve rotation and reflection symmetries, there is a translation symmetry which translate a cell in the grid to another. To eliminate duplications, each time a new polyhex with h cells is generated, one transforms this polyhex using each of the 12 rotation and reflection operations. The derived 12 polyhexes are then compared to the existing polyhexes with h cells. If there exists a polyhex with h cells that differs from one of the 12 derived polyhexes by a translation operation, then the newly generated polyhex has already been generated before and is discarded; otherwise, it is a new polyhex. To ensure fast comparison, all enumerated polyhexes are stored in a binary search tree.

As an example, all 7 unique polyhexes with 4 cells are shown in Fig. 2 in which the polyhexes are denoted using their constitution cells in the honeycomb grid—for example, C(1, 2, 3, 4) is the polyhex consisting of cells 1, 2, 3, and 4 in the 4 × 4 block honeycomb grid as shown in Fig. 1. The fifth polyhex C(1, 2, 3, 6) contains an odd number of carbon atoms while the others contain even numbers of carbon atoms. The neutral PAH of the fifth polyhex is an open-shell radical while its first cation is a closed-shell molecule. Species like this are believed to be significant intermediates in combustion processes and possible interstellar species [39,40]. We list in Table 1 the number of polyhexes with up to 12 cells and the corresponding computational time on an Intel Pentium 4 3.33 GHz computer. It is shown from the table that the number of polyhexes and the corresponding computational time increase exponentially with the number of cells. This algorithm based on the encoding of honeycomb grid is found to have about the same efficiency as the algorithm based on the Dualist Angle-restricted Spanning Tree code [35].

3. Fast generation of electronic database of PAHs

The most important step of building a PAH database is to calculate the electronic transition energies, oscillator strengths, and ground-state rotational constants (or equilibrium geometries) of the generated PAHs quickly enough without losing much accuracy. The initial geometries of PAHs can be easily generated from the geometric data of polyhexes. Missing hydrogen atoms are added back to those carbon atoms that are connected with two and only two other carbon atoms. We set the H—C—C angles to $2\pi/3$, the C—C bond lengths to 1.397 Å, and C—H bond lengths to 1.084 Å. These PAH bond lengths are used in the semi-empirical MO/8E force field program [41]. The initial PAH geometries obtained this way are called the "model" geometries. The model geometries are then optimized using quantum chemical methods to obtain the optimized geometries.

Considering the large number of polyhexes (i.e. 38,959 for h up to 10), the methods used for geometry optimization and the calculation of electronic transition energies need to be very efficient. *Ab initio* calculations are too computationally expensive for these PAHs and thus not feasible. Instead, semi-empirical methods are more desirable for this purpose. For geometry optimization, we use the Austin Model 1 (AM1) Hamiltonian [42–44]—which is based on the Neglect of Diatomic Differential Overlap (NDDO) approximate Hamiltonian; for electronic excitation energies and oscillator strengths, we use the Zerner's Intermediate Neglect of Diatomic Differential Overlap (ZINDO) method [45–47]. It is found, based on the testing of several PAHs, that the rotational constants derived from the AM1 method are usually 2–5% larger than those obtained from the model geometries. The vertical transition energies calculated