The NASA Ames PAH IR Spectroscopic Database: Computational Version 3.00 with Updated Content and the Introduction of Multiple Scaling Factors

Charles W. Bauschlicher, Jr.¹, A. Ricca^{2,3}, C. Boersma³, and L. J. Allamandola³ ¹NASA Ames Research Center, MS 230-3, Moffett Field, CA 94035-0001, USA; Charles.W.Bauschlicher@nasa.gov

Carl Sagan Center, SETI Institute, 189 Bernardo Avenue, Suite 200, Mountain View, CA 94043, USA

³NASA Ames Research Center, MS 245-6, Moffett Field, CA 94035-0001, USA

Received 2017 September 25; revised 2017 December 5; accepted 2017 December 5; published 2018 February 5

Abstract

Version 3.00 of the library of computed spectra in the NASA Ames PAH IR Spectroscopic Database (PAHdb) is described. Version 3.00 introduces the use of multiple scale factors, instead of the single scaling factor used previously, to align the theoretical harmonic frequencies with the experimental fundamentals. The use of multiple scale factors permits the use of a variety of basis sets; this allows new PAH species to be included in the database, such as those containing oxygen, and yields an improved treatment of strained species and those containing nitrogen. In addition, the computed spectra of 2439 new PAH species have been added. The impact of these changes on the analysis of an astronomical spectrum through database-fitting is considered and compared with a fit using Version 2.00 of the library of computed spectra. Finally, astronomical constraints are defined for the PAH spectral libraries in PAHdb.

Key words: astrochemistry – astronomical databases: miscellaneous – ISM: lines and bands – methods: numerical – molecular data – techniques: spectroscopic

1. Introduction

In 2010, the inaugural version (Bauschlicher et al. 2010) of the NASA Ames PAH IR Spectroscopic Database⁴ (PAHdb), containing a large collection of theoretical and experimental infrared (IR) spectra of polycyclic aromatic hydrocarbons (PAHs), was made publicly available. The library of computed PAH spectra featured 583 density functional theory (DFT) spectra. In 2013, PAHdb underwent a major update in the form of a version 2.00 release (Boersma et al. 2014). This version extended the library of computed PAH spectra to a total of 700 theoretical spectra, which featured larger PAHs, and included advanced tools for the analysis of the IR spectra, such as, for example, the non-negative least squares fitting of imported astronomical spectra. This paper reports on the version 3.00 update to the library of computed PAH spectra that adds 2439 spectra and revises the treatment of scale factors.

The vibrational frequencies are computed using the harmonic approximation. To account for limitations in the level of theory and for anharmonic corrections, the computed frequencies are scaled to improve the agreement with the experimental fundamentals. Since the C-H stretching anharmonicities are known to be larger than those of the other modes, Bauschlicher & Langhoff (1997) initially used two scale factors to fit the naphthalene results to the band positions from two matrixisolation experiments (Szczepanski & Vala 1993; Hudgins & Sandford 1998). Using the B3LYP (Stephens et al. 1994) hybrid (Becke 1993) functional in conjunction with a 4-31G basis set (Frisch et al. 1984), it was found that the two scale factors were essentially identical, and that the agreement with the experimental fundamentals was very good. However, Bauschlicher & Ricca (2010) noted that for some PAHs, the B3LYP functional could fail due the presence of low-lying excited states. This failure was easy to detect due to unreasonably large IR intensities. In some cases, using a

Using the B3LYP or BP86 functional in conjunction with the 4-31G basis set allowed the study of numerous PAHs. However, this level of theory was not without its limitations. The basis set is too small for systems containing oxygen atoms. In addition, a comparison of the computed IR data for strained systems like C₆₀, C₇₀, and benzyne, as well as a series of acridine molecules (Mattioda et al. 2017) with experiment showed somewhat larger errors for strained or nitrogencontaining systems than found for planar PAH molecules containing only carbon and hydrogen atoms. Clearly, these systems required a higher level of theory. Bauschlicher & Ricca (2010) showed that switching to a different hybrid functional did not improve the results over B3LYP, nor did changing the GGA used improve the results compared with the BP86. This showed that the choice of functional was not the root of the problem and therefore larger basis sets were needed to improve the description of these systems. However, using basis sets larger than 4-31G with only two scale factors actually made the agreement with experiment worse (Bauschlicher & Langhoff 1997; Bauschlicher & Ricca 2010). A clue to the problem with larger basis sets was provided by a comparison of the B3LYP/4-31G and the BP86/4-31G results, where there was a systematic difference in the $11-15 \,\mu m$ region, with B3LYP better describing the very strong band near 11 μ m. This suggests that the C-H stretch, the 6-9 μ m, and 9-15 μ m regions should be scaled separately. Switching from two to three scale factors improved the accuracy of the BP86/4-31G



generalized gradient approximation (GGA), i.e., a non-hybrid functional, such as BP86 (Perdew 1986; Becke 1988), could circumvent the failure. As with B3LYP, the two scale factors were essentially identical and only one factor was used. However, when both functionals worked, the scaled B3LYP frequencies were in better agreement with experiment than the scaled BP86 results. Thus, the computational part of the database included results obtained using the B3LYP/4-31G approach when available or the BP86/4-31G level of theory when the B3LYP/4-31G results were unavailable.

⁴ http://www.astrochemistry.org/pahdb/



Figure 1. Pie charts showing the distribution of charge (left) and composition (right) for all species in the computational part of the database, version 3.00.

PAH spectra, bringing it into better agreement with experiment and with the B3LYP/4-31G results. Using three scaling factors also improved the agreement with experiment as the basis set quality was increased. Thus, using three scale factors allowed the inclusion of results obtained using larger basis sets into the database. Note that even three scale factors can be insufficient. As noted previously (Bauschlicher & Ricca 2013), the C-C triple bond requires yet a different scale factor.

In Section 2 the newly added species to the computational library are discussed. Section 3 features a discussion on scale factors. Section 4 considers the impact of the changes on the analysis of an astronomical spectrum through database-fitting and introduces the concept of the astronomical PAH subset to the database. This paper closes with a summary and conclusions in Section 5.

2. Added Species and Composition

PAHs containing five- and seven-membered rings (Bauschlicher 2015), including C_{60} and C_{70} , have been added with molecular UIDs ranging from 717 to 737. Circum-circum-circumanthracene $(C_{190}H_{34})$, circum-circum-circumtetracene $(C_{210}H_{36})$, and those PAHs containing bay regions described in Peeters et al. (2017) have been added with molecular UIDs ranging from 738 to 774. The other oval PAH molecules mentioned in Peeters et al. (2017) were included in version 2.00. PAHs with side groups have been assigned molecular UIDs ranging from 775 to 812 (Bauschlicher 2016). The naphtho-benzo-pyrene anion from the study of Mattioda et al. (2014) that was computed using a 6-31G^{*} basis set was added with molecular UID 813 and hexacene, heptacene, and octacene have molecular UIDs 814 to 822; the IR spectra for the other species described in Mattioda et al. (2014) were computed using the B3LYP/4-31G approach and were already in the database. PAHs with molecular UIDs from 823 to 828 were some unpublished C54H18 species; one had an extra hydrogen added, while the others had one or more five-membered rings. Perylene (Langhoff 1996) was added with molecular UID 829. The species from our recent B3LYP/6-31G* study (Mattioda et al. 2017) of some neutral acridine species and their PAH parents have been added (molecular UIDs 830 to 832) or replaced species that were computed using the 4-31G basis set. Species with molecular UIDs that range from 833 to 836 are the most stable species from our study (Bauschlicher & Ricca 2014) of the loss of C₂H₂. Species with molecular UIDs from 837 to 2357 are the neutral $C_n H_v$ species of Mackie et al. (2015), while their neutral dehydrogenated C_n species have molecular UIDs ranging from 2358 to 3155.

Figure 1 shows the distribution of charge and composition for all of the PAHs included in the computational part of version 3.00 of the database. Clearly, most of the species are neutrals and have a formula of the form C_xH_y , i.e., they are pure PAHs. A much more detailed breakdown of the molecules included in version 2.00 and 3.00 of the database is given in Table 1. From this table it is clear that most of the 2439 added species are neutral PAHs containing less than 50 carbon atoms. This is because Mackie et al. (2015) studied all of the neutral C_xH_y and C_x PAHs with eight or fewer fused rings. However, there is a near doubling of the species with more than 100 carbons. Few of the new PAH species contain nitrogen and none of the new PAH species contain oxygen, magnesium, or iron.

The breakdown of the molecules included in the theoretical part of the database by number of carbon atoms is shown in Figure 2. Most species of the database involve molecules with between 21 and 50 carbon atoms and most of the species are "pure" PAHs. In fact, the database should now include all neutral PAHs with 8 or fewer fused rings. Version 2.00 had about equal numbers of neutrals and cations, with a reasonable population of multiply charged cations and anions of the larger species. For version 3.00, this charge balance is no longer true, as most of the added species are neutrals. Figures 2(c) and (d)are analogous to Figures 2(a) and (b), but exclude the dehydrogenated " C_x " species. The insights that can be deduced from Figures 2(a) and (b) also apply to Figures 2(c) and (d). Figures 2(e) and (f) show the composition for the C_x species in the database. They are essentially all neutral and contain between 21 and 50 carbons.

3. Scale Factors

As described in the introduction, the decision was made to switch to three scale factors with ranges as follows: (1) C–H stretching bands (greater than 2500 cm⁻¹; less than 4 μ m), (2) bands between 2500 and 1111.1 cm⁻¹ (i.e., between 4 and 9 μ m), and (3) bands between 1111.1 and 0 cm⁻¹ (i.e., greater than 9 μ m). Note that in our original study, we determined the scale factors by minimizing $\sum_{i=1}^{n} |s(i) * \omega_e(i) - \nu(i)|$, where *n* is the number of IR bands for which there are experimental data, *s*(*i*) is the scale factor, which is a function of the harmonic frequency, ω_e , that is obtained from DFT, and ν is the experimental fundamental value. In this work we switch to $\sum_{i=1}^{n} (s(i) * \omega_e(i) - \nu(i))^6$ since we are willing to increase the small errors somewhat to reduce the maximum error.

We scaled our harmonic DFT frequencies to better reproduce the experimental fundamentals, therefore we needed one or more experimental spectra to determine the scale factors. Ideally, we would have used the spectra of many large PAH molecules. Since the experimental library of the database contains 75 PAHs, with the largest being C₅₀H₂₂, this might have seemed straightforward. However, a comparison of the computed and experimental spectra for the large PAHs showed that assignments were not easy to make. The experimental IR spectra commonly have weak bands in regions where theoretical spectra have no allowed bands; these are presumably overtones or combination bands that are not included in the harmonic approximation. Even after discarding these "extra" experimental bands, lining up the computed and experimental bands was not easy. In some cases the band positions suggested one assignment, while the band intensities suggest another. For example, consider two nearby theory bands A(theory) and B(theory) and two nearby experimental bands A(exp) and B(exp), where band positions might have

	Breakdown	of the Molecules Ir	ncluded in Version	3.00 of the Computation	ational Database by	Charge, Composi	tion, and Size				
	Number of Carbon Atoms										
	1–10	11–20	21-30	31–50	51-70	71–100	101–386	Total			
Charge			olecules								
all	76(76)	211(195)	839(74)	1663(104)	140(104)	114(98)	96(49)	3139(700)			
neutral	39(39)	115(98)	795(37)	1590(35)	43(27)	37(31)	35(19)	2654(286)			
anion –	3(3)	10(11)	8(4)	22(20)	16(13)	21(20)	27(14)	107(85)			
cation +	34(34)	86(86)	29(26)	45(43)	65(48)	51(42)	34(16)	344(295)			
++/+++	0(0)	0(0)	7(7)	6(6)	16(16)	5(5)	0(0)	34(34)			
	PAHs Without Substitutions										
all	26(26)	121(107)	814(50)	1648(89)	108(72)	100(84)	96(49)	2913(477)			
neutral	12(12)	70(55)	784(27)	1589(34)	41(25)	37(31)	35(19)	2568(203)			
anion –	3(3)	8(9)	8(4)	22(20)	16(13)	21(20)	27(14)	105(83)			
cation +	11(11)	43(43)	18(15)	31(29)	41(24)	37(28)	34(16)	215(166)			
++/+++	0(0)	0(0)	4(4)	6(6)	10(10)	5(5)	0(0)	25(25)			
	PAHs with Nitrogen										
all	10(10)	48(46)	18(17)	9(9)	18(18)	8(8)	0(0)	111(108)			
neutral	5(5)	24(22)	11(10)	0(0)	0(0)	0(0)	0(0)	40(37)			
anion –	0(0)	2(2)	0(0)	0(0)	0(0)	0(0)	0(0)	2(2)			
cation +	5(5)	22(22)	7(7)	9(9)	12(12)	8(8)	0(0)	63(63)			
++/+++	0(0)	0(0)	0(0)	0(0)	6(6)	0(0)	0(0)	6(6)			
	PAHs with Oxygen										
all	40(40)	42(42)	0(0)	0(0)	2(2)	0(0)	0(0)	84(84)			
neutral	22(22)	21(21)	0(0)	0(0)	0(0)	0(0)	0(0)	43(43)			
anion –	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)			
cation +	18(18)	21(21)	0(0)	0(0)	2(2)	0(0)	0(0)	41(41)			
++/+++	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)			
	PAHs with Magnesium or Iron										
all	0(0)	0(0)	7(7)	6(6)	12(12)	6(6)	0(0)	31(31)			
neutral	0(0)	0(0)	0(0)	1(1)	2(2)	0(0)	0(0)	3(3)			
anion –	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)			
cation +	0(0)	0(0)	4(4)	5(5)	10(10)	6(6)	0(0)	25(25)			
++/+++	0(0)	0(0)	3(3)	0(0)	0(0)	0(0)	0(0)	3(3)			
	PAHs without Hydrogen										
all	0(0)	10(0)	295(2)	495(0)	10(4)	3(3)	0(0)	813(9)			
neutral	0(0)	10(0)	294(1)	495(0)	4(2)	1(1)	0(0)	804(4)			
anion –	0(0)	0(0)	0(0)	0(0)	2(0)	0(0)	0(0)	2(0)			
cation +	0(0)	0(0)	1(1)	0(0)	4(2)	1(1)	0(0)	6(4)			
++/+++	0(0)	0(0)	0(0)	0(0)	0(0)	1(1)	0(0)	1(1)			
/		· · ·	5 Z		5 Z		· · ·				

 Table 1

 Breakdown of the Molecules Included in Version 3.00 of the Computational Database by Charge Composition, and Siz

Note. The number of species in each class in version 2.00 is given in parentheses.

suggested assigning A(theory) with A(exp) and B(theory) with B(exp), but with A(theory) and B(exp) having larger intensities than B(theory) and A(exp). Thus, on the basis of the intensities, one would have reversed the assignment. Uncertainties in the band assignments for large PAHs have led us to take a different approach, namely we have used naphthalene, for which the assignments of the bands are definitive and there are several experimental bands to determine the scale factors. We have then used these scale factors to compute the spectra of larger PAH molecules. The good agreement of the (scaled) computed and experimental spectra for larger species has been used to validate this approach.

In version 3.00 we used the gas-phase experimental spectrum of naphthalene instead of the average of two matrix-isolation spectra (Szczepanski & Vala 1993; Hudgins

& Sandford 1998). The scale factors were obtained by fitting to the 17 allowed IR bands determined by Pirali et al. (2009; this includes the 16 bands in Table 1 of their paper and the band at 1601 cm^{-1} shown in their Figure 2) the 4 A_g and 2 B_{2g} bands summarized by Behlen et al. (1981) and the one A_u and one B_{1g} bands summarized by Cané et al. (1998). Figure 3 compares all of the available laboratory-determined matrix-isolated band positions in PAHdb with those determined using either a single or multiple scale factors. The figure shows that there is only a very slight (0.02%) increase in spread when moving to multiple scale factors based on gas-phase experiments.

The average and maximum errors in the band position for the B3LYP functional, as a function of the basis set, are given in Table 2. The results are given for the Pople-style (Frisch et al. 1984) basis sets in the top part of the table and for the



Figure 2. Breakdown of the species in the computational part of the database, version 3.00, by number of carbon atoms: (a) by composition, (b) by charge, (c) by composition excluding the C_x species, (d) by charge excluding the C_x species, (e) by composition for the C_x species, and (f) by charge for the C_x species.

correlation-consistent valence polarized (cc-pV) basis sets of Dunning (1989) in the bottom of the table. Note that the B3LYP/4-31G errors are different from those reported previously, since we are now comparing with a different, much larger, set of experimental values. As the basis sets improve, for both the Pople-style and correlation-consistent style sets, the average and maximum errors tend to decrease. Note that the cc-pVQZ basis set actually has a slightly larger error than the cc-pVTZ basis set and the 6-311G(2d, 2p) basis set has somewhat larger errors than the 6-311G^{**} basis set; that is, all of the largest basis sets have about the same errors. We also note that the same experimental bands have the largest error for all basis sets and computing the anharmonic frequencies actually increases the maximum error. The problem bands in naphthalene are the a_u band at 195, the b_{2u} band at 1135.5, and the a_g band at 1379.9 cm⁻¹. None of these bands appear in the older matrix-isolation experiments. It would be interesting to understand the origin of the difference between theory and experiment for these three bands. The gas-phase frequencies are in good agreement with the average of matrixisolation results, so the change in the experimental data used for the scale factor fits makes only small changes in the scale factors. For example, the single scale factor for the B3LYP/ 4-31G changed from 0.958 to 0.954. While this switch might slightly degrade the agreement between theory and matrixisolation experiments, it will hopefully improve the agreement



Bauschlicher et al.

Figure 3. Comparison of the distributions (red line) of the ratio of PAH band positions established from matrix-isolation experiments over the DFT-calculated band positions using either a single scale factor (top) or multiple scale factors (bottom). Shown is a fitted Gaussian (blue dotted line) and indicated are the mean of the fitted Gaussian (μ) and the width of the fitted Gaussian (σ).

 Table 2

 Scale Factors and the Average and Maximum Errors (in cm⁻¹) as a Function of Basis Sets for the B3LYP Functional

Basis Set	S	cale Facto	rs	Avg. Error	Max Error	
	1	2	3			
4-31G	0.956	0.952	0.960	7.8	26.9	
6-31G	0.961	0.953	0.957	8.0	23.6	
6-31G*	0.979	0.969	0.960	6.9	15.1	
6-31G**	0.979	0.973	0.961	6.2	11.8	
6-311G**	0.982	0.980	0.966	4.8	11.9	
6-311G(2d, 2p)	0.982	0.982	0.965	5.4	13.8	
cc-pVDZ	0.983	0.980	0.963	7.2	15.9	
cc-pVTZ	0.980	0.979	0.965	4.8	11.7	
cc-pVQZ	0.982	0.981	0.966	5.0	12.3	

Note. Scale factor 1 is for 0 to 1111. \overline{I} cm⁻¹, 2 is for 1111. \overline{I} to 2500 cm⁻¹, and 3 is for greater than 2500 cm⁻¹.

with observations, which correspond to gas-phase molecules; see Section 4.

As noted in the introduction, the B3LYP/6-31G^{*} harmonic frequencies with three scale factors were in better agreement with experiment than were the B3LYP/4-31G results for PAHs containing heteroatoms, such as acridine species. Also note that even larger basis sets can be used and reduce the errors even further. With current programs and computers it is possible to use the cc-pVTZ set on a system as large as $C_{150}H_{30}$. Thus, using the three scale factors allows an improved treatment for

smaller species where a larger basis set can be used and puts the BP86/4-31G approach on more equal footing with the B3LYP/4-31G approach. The naphthalene results suggest that most of our scaled bands should be within $5-10 \text{ cm}^{-1}$ of the experimental band positions and a few could have an error as large as 15 cm^{-1} . This level of accuracy is sufficient to gain insight into astronomical observations.

4. Astronomical Considerations

In version 2.00 of the PAHdb (Boersma et al. 2014), the capability to directly fit an astronomical spectrum was added. Since then, the database-fitting approach has proven to be a valuable addition to the astronomer's toolkit (e.g., Andrews et al. 2015; Zhang & Kwok 2015; Boersma et al. 2016). To consider the astronomical impact of the updates to the library of computed spectra, the database-fitting approach is applied to the same Spitzer-IRS (Houck et al. 2004; Werner et al. 2004) spectrum from Boersma et al. (2014) to decompose the PAH emission into contributing PAH subclasses, i.e., charge and size. This spectrum, reproduced from Boersma et al. (2013; position II), probes emission from the diffuse region near the northwest photodissociation region (PDR) of NGC 7023. Before fitting, any molecular hydrogen lines were removed, the spectrum was corrected for extinction and a warm dust continuum was subtracted. The reader is directed to Boersma et al. (2013) for details.

In Boersma et al. (2014), to illustrate the fitting approach that was then made publicly available via PAHdb, no constraints were placed on the PAHs that should be considered for inclusion or exclusion in the fit. This was done to highlight the general inconsistencies that arise with our current understanding of the PAH model, specifically those that arise when comparing with our first fit results that were presented in Boersma et al. (2013). That exercise emphasized the need to properly constrain the PAHs to be considered for inclusion in a fit by taking into account the physical conditions in the astronomical environment.

Here, full advantage is taken of that knowledge and only PAHs in the computational part of the database meeting the following astronomical constraints are considered for a fit:

- 1. contain more than 20 carbon atoms;
- 2. are either a "pure" PAH or PANH;
- 3. are not super-hydrogenated;
- 4. have no aliphatic side groups;
- 5. are not fully dehydrogenated, excluding C_{60} and C_{70} , which are included in the fit.

The size limit (1) is imposed as PAHs with less than 20 carbon atoms are quickly destroyed due to their instability with respect to photodissociation in the emission region (Allamandola et al. 1989; Puget & Léger 1989). Only "pure" PAHs—those containing only carbon and hydrogen—and PANHs—those containing carbon, hydrogen, and nitrogen—are considered (2) as other heteroatom inclusion is deemed less significant⁵ (Hudgins et al. 2005). Super-hydrogenated and PAHs containing

aliphatic side groups (3) are not included at this stage because at best they add only minor substructure to the 5–15 μ m region of the spectrum. This will change with the availability of JWST spectra, which will include the 2.5–3.5 μ m region where the CHstretch aromatic and aliphatic signatures are clearcut (Geballe et al. 1989; Joblin et al. 1996). The possible contribution of emission from fully dehydrogenated PAHs (5) to the astronomical PAH spectrum is intriguing; however, it is highly unlikely that a significant fraction of the PAHs will be fully dehydrogenated in the emission region considered here (Mackie et al. 2015). However, fullerenes (C_{60} and C_{70}) have not been excluded, as there is clear observational evidence for their presence in harsh radiation environments (e.g., (Cami et al. 2010; Sellgren et al. 2010; Boersma et al. 2012; Berné et al. 2013). Imposing these constraints results in a pool of 283 and 1877 spectra to be considered when using the library of computed PAH spectra in version 2.00 and 3.00, respectively, of the PAHdb.

A PAH emission model must be applied to transform the integrated PAH absorption cross sections available in the database into PAH emission spectra. Here, each PAH is excited by a 7 eV photon, the computed band positions are redshifted 15 cm^{-1} to simulate anharmonic effects; each transition is convolved with a Gaussian emission profile with a FWHM of 15 cm^{-1} ; and the model takes the full temperature cascade into account (see Bauschlicher et al. 2010; Boersma et al. 2013).

The AmesPAHdbIDLSuite is utilized to perform the database fit. This suite is a collection of IDL^6 object classes (a set of programs utilizing IDL's object-oriented programming capabilities) and can be obtained from the PAHdb website⁷ or from the GitHub development repository.⁸ Figure 4 presents the results of the fit, directly comparing the results obtained using versions 2.00 (left), 3.00 (middle), and 3.00' (right). Version 3.00' contains the same collection of PAHs as in version 3.00, but instead of using multiple scale factors, it uses a single scale factor (0.958 for B3LYP and 0.985 for BP86 as in version 2.00, and 0.978 for the fullerenes as in version 3.00). This allows one to separate the effects of adding more species from those produced by introducing multiple scale factors.

The top row of the figure shows that the fits are good and that they are very comparable to each other. The second row of the figure shows the breakdown of the spectrum in terms of size. The contribution from large PAHs ($n_{carbon} > 50$) and small PAHs ($n_{carbon} \leq 50$) is about 50–50 in all three panels. The bottom row of Figure 4 shows the breakdown of the spectrum by charge. Again, the differences between the left, middle, and right panels are generally small.

Figure 5 presents an overview of the chemical structures involved in each of the fits, ordered according to their contribution to the fit in terms of flux. The figure shows that a variety of PAH structures is needed, including those with nitrogen and with defects, as well as fullerenes. The number of PAHs required for the fit is about the same when using either version 2.00 (29), 3.00 (27), or 3.00' (27) of the computed library.

We compute the error in the fits as the area of the absolute value of the residual over the area of the astronomical spectra. This integration is done in cm^{-1} space since it is linear in

⁵ Note that while oxygen substitution for carbon atoms within the hexagonal carbon skeleton is considered highly unlikely due to stability arguments, it has been shown that H addition, as well as replacement of peripheral H atoms with OH and = O (forming aromatic alcohols and ketones), readily occurs in PAH/ water ice experiments simulating photoprocessing in dense clouds (Bernstein et al. 1999; Cook et al. 2015; de Barros et al. 2017). These species will be released from the ices at the edge of the cloud and will be part of the PAH family feedstock into the PDR. Perusal of Table 1 shows these species are currently underrepresented in the library of computed spectra.

⁶ http://www.harrisgeospatial.com/ProductsandSolutions/GeospatialProducts/ IDL.aspx

⁷ http://www.astrochemistry.org/pahdb/downloads/

⁸ http://github.com/PAHdb/AmesPAHdbIDLSuite



Figure 4. The 5–15 μ m Spitzer-IRS spectrum of position II from Boersma et al. (2013) of the reflection nebula NGC 7023 decomposed using the spectra from version 2.00 (left column), 3.00 (middle column), and 3.00' (right column) using of the computed libraries of the NASA Ames PAH IR Spectroscopic Database. Top row: observed spectrum ("×"), fit (red). Middle row: size breakdown of the spectrum into contributions from small ($N_C \le 50$; red) and large ($N_C > 50$; green) PAHs. The observed spectrum and total fit are shown as "×"-symbols and a black line, respectively. Bottom row: charge breakdown of the spectrum into contributions from PAH anions (red), neutrals (green), and cations (blue). The observed spectrum and the total fit are shown as "×"-symbols and a black line, respectively. See Section 4 for details.

energy. The errors are given in Figure 4 and they show an increase when moving from the computational database at version 2.00 to version 3.00, but a reduction when using version 3.00'. It should be noted that the contribution to the fit of the molecules treated using the BP86 approach is reduced when using version 3.00' compared to using version 2.00. Part of the motivation for introducing three scale factors was to prevent the BP86 results from being used to incorrectly fit the red wing of the 11.2 μ m band. However, it appears that new molecules in version 3.00 have replaced the BP86 contribution to the spectra. Not surprisingly the BP86 contribution to the fit using version 3.00 is also very small. About half of the difference in the fit between version 3.00 and 3.00' comes from the PANH molecules. When three scale factors are used, the very small reduction in the scale factor in the 6–9 μ m range reduces the contribution to the fit from the PANHs. In the long run, the computational spectra for the nitrogen-containing molecules should be replaced with those obtained using larger basis sets. The slightly larger error obtained when using version 3.00 compared to using 3.00' is not because the 3.00' version is necessarily better, but because version 3.00 has reduced the

importance of species included in the 3.00' fit due to errors in the band position associated with a single scale factor. That is, the difference between versions 3.00 and 3.00' indicates that the following areas need improvement: (i) important classes of molecules are underrepresented and need to be included, (ii) a more thorough computational treatment that includes anharmonicity, and (iii) a superior treatment of the emission process that includes anharmonicity is needed. These limitations force the inclusion of incorrect species in the fit due to small errors in the older scale factor.

Table 3 cross-references the PAHs in each fit and presents the chemical formula of each PAH and its contribution in terms of flux (f) and number density (a). In all three cases, about 35% of the emission is produced by the same two PAHs. Of the remaining PAHs, most contributing at $\leq 5\%$, there are eight (8) unique to version 2.00 and one (1) to version 3.00'. Between versions 2.00 and 3.00 there are eleven (11) new PAHs required for the fit.

While the quality of the overall fit is generally good, it has difficulties matching some of the emission. Notable are the blue side of the $6.2 \mu m$ PAH feature and the PAH emission



Figure 5. Overview of the PAH structures contributing to the decomposition of the 5–15 μ m *Spitzer*-IRS spectrum of position *II* from Boersma et al. (2013) of the reflection nebula NGC 7023 decomposed using the spectra from version 2.00 (top), 3.00 (middle), and 3.00' (bottom) using the computed libraries of the NASA Ames PAH IR Spectroscopic Database. The structures are presented in order of their contribution to the fit in terms of flux. Table 3 shows, for each structure, the chemical formula, its contribution to the total number of PAHs (a) and its contribution to the total flux (f) in percent. See Section 4 for details.

originating between 10 and 15 μ m. In addition, the contribution from PAH anions to the spectrum seems unreasonably large (see the discussion in (Boersma et al. 2012).

These difficulties are mostly due to the limited variation within the appropriate class of PAH species in PAHdb. It has been suggested that the blue side of the 6.2 μ m band can be

attributed to PAHs containing nitrogen, PANHs (e.g., Peeters et al. 2002; Hudgins et al. 2005, class A). Version 3.00 of the computed library holds 55 PANHs meeting the astronomical criteria listed above. Of these, most form a series where the loci of the nitrogen atom is permuted, not different large molecular structures containing nitrogen.

The Astrophysical Journal Supplement Series, 234:32 (10pp), 2018 February

	Version 2.00				Version 3.00				Version 3.00'		
Ref.	Formula	f [%]	a [%]	Ref.	Formula	f [%]	a [%]	Ref.	Formula	f [%]	a [%]
2.1 = 3.1 = 3.1'	$C_{66}H_{20}^{-}$	20.0	20.4	3.1 = 3.1' = 2.1	$C_{66}H_{20}^{-}$	21.3	21.6	3.1' = 2.1 = 3.1	$C_{66}H_{20}^{-}$	19.3	19.5
2.2 = 3.2 = 3.2'	$C_{52}H_{18}N_2^{+2}$	16.6	15.5	3.2 = 3.2' = 2.2	$C_{52}H_{18}N_2^{+2}$	11.5	10.7	3.2' = 2.2 = 3.2	$C_{52}H_{18}N_2^{+2}$	16.0	14.8
2.3 = 3.3 = 3.5'	C ₆₆ H ₂₀	8.7	9.1	3.3 = 3.5' = 2.3	C ₆₆ H ₂₀	9.2	9.6	3.3' = 2.5 = 3.4	$C_{42}H_{22}^{-}$	6.8	6.4
2.4 = 3.10 = 3.4'	$C_{31}H_{14} N^+$	6.0	5.9	3.4 = 3.3' = 2.5	$C_{42}H_{22}^{-}$	6.8	6.4	3.4' = 2.4 = 3.10	$C_{31}H_{14}N^+$	6.3	6.1
2.5 = 3.4 = 3.3'	$C_{42}H_{22}^{-}$	5.6	5.3	3.5 = 3.9'	$C_{52}H_{16}^+$	5.0	4.7	3.5' = 2.3 = 3.3	C66H20	5.7	6.0
2.6 = 3.11 = 3.6'	$C_{48}H_{18}^+$	4.8	4.6	3.6 = 3.17' = 2.7	$C_{66}H_{20}^+$	4.9	4.6	3.6' = 2.6 = 3.11	$C_{48}H_{18}^+$	4.9	5.4
2.7 = 3.6 = 3.17'	$C_{66}H_{20}^+$	3.7	4.6	3.7 = 3.13' = 2.8	$C_{32}H_{14}^+$	4.8	4.6	3.7' = 2.10 = 3.8	$C_{23}H_{12}N$	4.8	5.3
2.8 = 3.7 = 3.13'	$C_{32}H_{14}^+$	3.7	3.8	3.8 = 3.7' = 2.10	$C_{23}H_{12}N$	3.9	4.4	3.8' = 2.9 = 3.9	C36H16	4.2	4.5
2.9 = 3.9 = 3.8'	C36H16	3.6	3.5	3.9 = 3.8' = 2.9	C36H16	3.6	4.4	3.9' = 3.5	$C_{52}H_{16}^+$	3.5	3.3
2.10 = 3.8 = 3.7'	$C_{23}H_{12}N$	3.4	3.4	3.10 = 3.4' = 2.4	$C_{31}H_{14}N^+$	3.4	3.2	3.10' = 2.13 = 3.12	$C_{90}H_{24}^+$	2.8	2.9
2.11 = 3.24 = 3.15'	$C_{96}H_{22}^{-}$	3.3	3.2	3.11 = 3.6' = 2.6	$C_{48}H_{18}^+$	3.4	3.2	3.11' = 2.12 = 3.13	$C_{52}H_{18}N_2^+$	2.8	2.7
2.12 = 3.13 = 3.11'	$C_{52}H_{18}N_2^+$	2.7	2.8	3.12 = 3.10' = 2.13	$C_{90}H_{24}^+$	3.3	3.1	3.12' = 3.27	$C_{146}H_{30}^{-}$	2.8	2.7
2.13 = 3.12 = 3.10'	$C_{90}H_{24}^+$	2.5	2.4	3.13 = 3.11' = 2.12	$C_{52}H_{18}N_2^+$	2.5	2.6	3.13' = 2.8 = 3.7	$C_{32}H_{14}^+$	2.7	2.5
2.14 = 3.14	$C_{42}H_{22}^+$	2.5	2.4	3.14 = 2.14	$C_{42}H_{22}^+$	2.4	2.2	3.14' = 3.17	$C_{34}H_{20}^{-}$	2.7	2.5
2.15	$C_{98}H_{28}^+$	2.4	2.3	3.15 = 3.22'	$C_{34}H_{20}^+$	2.0	2.1	3.15' = 2.11 = 3.24	$C_{96}H_{22}^{-}$	2.3	2.2
2.16	$C_{36}H_{20}^{-}$	1.8	2.0	3.16 = 3.16'	C_{60}^{-}	2.0	2.0	3.16' = 3.16	C_{60}^{-}	2.1	2.2
2.17 = 3.19 = 3.19'	$C_{32}H_{14}^{-}$	1.7	1.7	3.17 = 3.14'	$C_{34}H_{20}^{-}$	1.9	1.8	3.17' = 2.7 = 3.6	$C_{66}H_{20}^+$	2.1	2.0
2.18	$C_{150}H_{30}^{-}$	1.7	1.7	3.18 = 3.26'	$C_{26}H_{16}^{-}$	1.8	1.8	3.18' = 2.26	$C_{170}H_{32}^{-}$	1.9	1.9
2.19 = 3.27'	$C_{32}H_{18}^+$	1.2	1.2	3.19 = 3.19' = 2.17	$C_{32}H_{14}^{-}$	1.7	1.8	3.19' = 2.17 = 3.19	$C_{32}H_{14}^{-}$	1.7	1.9
2.20 = 3.23	$C_{24}H_{12}^{-}$	0.8	1.1	3.20	$C_{128}H_{28}^+$	1.6	1.5	3.20' = 2.29	$C_{32}H_{14}^{-}$	1.3	1.6
2.21	C ₂₃ H ₁₂ N	0.7	0.8	3.21 = 3.21'	C ₃₄ H ₂₀	1.0	1.3	3.21' = 3.21	C ₃₄ H ₂₀	1.2	1.5
2.22	$C_{102}H_{26}^+$	0.7	0.7	3.22	$C_{28}H_{14}$	0.8	1.1	3.22' = 3.15	$C_{34}H_{20}^+$	0.7	0.6
2.23	$C_{24}H_{12}^{+2}$	0.6	0.5	3.23 = 2.20	$C_{24}H_{12}^{-}$	0.5	0.6	3.23' = 2.24	$C_{112}H_{26}^{-}$	0.4	0.4
2.24 = 3.23'	$C_{112}H_{26}^{-}$	0.5	0.5	3.24 = 3.15' = 2.11	$C_{96}H_{22}^{-}$	0.3	0.3	3. 24′	$C_{80}H_{20}^+$	0.3	0.3
2.25	$C_{24}H_{14}^{+}$	0.3	0.3	3.25	$C_{40}H_{16}^{-}$	0.3	0.3	3.25' = 2.27	$C_{22}H_{14}^+$	0.3	0.3
2.26 = 3.18'	$C_{170}H_{32}^{-}$	0.2	0.2	3.26	C ₃₀ H ₁₆	0.1	0.1	3.26' = 3.18	$C_{26}H_{16}^{-}$	0.3	0.3
2.27 = 3.25'	$C_{22}H_{14}^+$	0.1	0.1	3.27 = 3.12'	$C_{146}H_{30}^{-}$	0.1	0.1	3.27' = 2.19	$C_{32}H_{18}^+$	0.3	0.2
2.28	$C_{82}H_{24}^+$	0.1	0.1		50				10		
2.29 = 3.20'	$C_{32}H_{14}^{-}$	0.1	0.1								

 Table 3

 Overview of the PAHs Contributing to the Decomposition of the 5–15 µm Spitzer-IRS Spectrum of Position II from Boersma et al. (2013) of the Reflection Nebula NGC 7023 Decomposed Using the Spectra from Version 2.00 (left), 3.00 (Middle), and 3.00' (Right) of the Computed Libraries of the NASA Ames PAH IR Spectroscopic Database

Note. The PAHs are presented in order of their contribution to the fit in terms of flux. For each PAH we provide the chemical formula, its contribution to the total number of PAHs (a), and its contribution to the total flux (f) as a percent. Figure 5 presents the chemical structure for each PAH. See Section 4 for details.

 ∞

The emission between 10 and 15 μ m is attributed to the C-H out-of-plane (CH_{oop}) vibrational motion of peripheral hydrogen atoms. Each of the distinct bands in the 10–15 μ m region is associated with a particular hydrogen adjacency class-the number of neighboring hydrogen atoms protruding a single aromatic ring, namely, the 11.2, 12.0, 12.7, and 13.5 μ m PAH bands are associated with the solo, duo, trio, and quartet adjacency classes, respectively (e.g., Hony et al. 2001). Currently, the computed library is limited in the number of large PAHs with varying adjacency classes. Most of the larger PAHs ($n_{carbon} > 50$) have condensed structures and are part of a series where one characteristic is permuted, e.g., the location of a seven-membered ring defect (Bauschlicher 2015). Hence the need for the acenes in the fit to add trio and guartet CH outof-plane bending bands.

Although the presence of PAH anions in space has been suggested (e.g., Bregman & Temi 2005; Bauschlicher et al. 2009), the uncannily large contribution here is attributed to improper treatment of anharmonicity. One of the effects of anharmonicity is that it introduces a red wing into the emission profile. Since the PAH band positions of PAH anions systematically fall to the red of their neutral and cationic counterparts, they tend to fill in the red wings instead.

The way forward to overcome these difficulties is straightforward and includes (1) the computation of the spectra of a large variety of PANHs and adding them to PAHdb; (2) the computation of a large variety of spectra from large PAHs covering all hydrogen adjacency classes and adding them to PAHdb; and (3) extending PAHdb to include the computed PAH spectra that treat the effects of anharmonicity and/or enhance the employed PAH emission model to better treat anharmonicity.

5. Summary and Conclusions

The computational part of the database has been modified to support multiple scale factors. While the B3LYP/4-31G results are slightly changed, the BP86/4-31G results are improved by the use of three scale factors, namely the systematic difference between the B3LYP and BP86 for the band at $11.2 \,\mu m$ has been reduced. More importantly, the use of multiple scale factors reduces the error for the larger basis set to be similar to or smaller than that found for the 4-31G basis set. This allows the addition of spectra computed using a larger basis set to the PAHdb. This yields improved results for species containing oxygen and nitrogen and for strained systems. We can now deal with strained triple bonds like those found in benzyne.

The database now contains C_{60} and C_{70} neutral and ions computed using the 6-31G* basis set and includes a special scale factor for strained systems in this basis set. In addition to changing the handling of scale factors, 2439 species have been added. Most of these are the neutral species from Mackie et al. (2015). However, the computational part of database version 3.00 has almost twice as many molecules with more than 100 carbons than contained in version 2.00. This includes species with bay regions, which has been attributed to the band at 8.2 μ m (Peeters et al. 2017).

A subset of PAH spectra from the updated version 3.00 library of computed spectra meeting astronomically relevant criteria were used in combination with a PAH emission model to fit and analyze an astronomical spectrum originating in the northwest PDR of the reflection nebula NGC 7023. These results were compared to those obtained from a fit using version 2.00 of the library and using the updated library with a

single scale factor instead (version 3.00'). Overall, the results are on a par with each other, especially when considering emission originating from the different PAH subclasses, i.e., charge and size. The noticeable reduction in the error when assuming a single scale factor (version 3.00') over multiple scale factors (version 3.00) is interpreted as the latter emphasizing the currently limited variation within each of the PAH subclasses (i.e., charge, size, structure, composition) in the PAHdb spectral library, as the use of multiple scale factors is considered appropriate from a molecular spectroscopic standpoint.

Some of the remaining significant limitations of the computational spectroscopic library of PAHdb in matching astronomical PAH emission spectra can be overcome by (1) the computation and addition of compact and irregular large PANH spectra in each charge state; (2) the computation and addition of spectra from large PAHs covering all hydrogen adjacency classes and charge states; and (3) the computation and addition of PAH spectra for large PAHS that treat anharmonicity and/or employ a PAH emission model that better treats anharmonicity.

Lastly, those who make use of the data and tools of the NASA Ames PAH IR Spectroscopic Database are kindly asked to refer to this paper, Boersma et al. (2014) and A. L. Mattioda et al. (2017, in preparation).

A.R. acknowledges support from NASA's APRA program (NNX17AE71G). C.B. is grateful for an appointment at NASA Ames Research Center through the San Jose State University Research Foundation (NNX14AG80A and NNX17AJ88A) and acknowledges support from NASA's ADAP program (NNH14ZDA001N). L.J.A. gratefully acknowledges support from NASA's ADAP program (NNH16ZDA001N) and is thankful for an appointment at the NASA Ames Research Center through the Bay Area Environmental Research Institute (80NSSC17M0014).

Facility: Spitzer (IRS).

Software: MPFIT (Markwardt 2009), IDL Astronomy User's Library (Landsman 1993), AmesPAHdbIDLSuite (Boersma et al. 2014).

ORCID iDs

A. Ricca https://orcid.org/0000-0002-3141-0630

C. Boersma ^(b) https://orcid.org/0000-0002-4836-217X

References

Allamandola, L. J., Tielens, G. G. M., & Barker, J. R. 1989, ApJS, 71, 733 Andrews, H., Boersma, C., Werner, M. W., et al. 2015, ApJ, 807, 99

- Bauschlicher, C. W. 2015, CP, 448, 43
- Bauschlicher, C. W. 2016, CPL, 665, 100
- Bauschlicher, C. W., Boersma, C., Ricca, A., et al. 2010, ApJS, 189, 341 Bauschlicher, C. W., & Langhoff, S. R. 1997, AcSpA, 53, 1225
- Bauschlicher, C. W., Peeters, E., & Allamandola, L. J. 2009, ApJ, 697, 311
- Bauschlicher, C. W., & Ricca, A. 2010, MolPh, 108, 2647
- Bauschlicher, C. W., Jr., & Ricca, A. 2013, ApJ, 776, 102

Bauschlicher, C. W., Jr., & Ricca, A. 2014, Theor. Chem. Acc., 133, 1479 Becke, A. D. 1988, PhRvA, 38, 3098

- Becke, A. D. 1993, JChPh, 98, 5648
- Behlen, F. M., McDonald, D. B., Sethuraman, V., & Rice, S. A. 1981, JChPh, 75, 5685
- Berné, O., Mulas, G., & Joblin, C. 2013, A&A, 550, L4
- Bernstein, M. P., Sandford, S. A., Allamandola, L. J., et al. 1999, Sci, 283, 1135
- Boersma, C., Bauschlicher, C. W., Ricca, A., et al. 2014, ApJS, 211, 8
- Boersma, C., Bregman, J., & Allamandola, L. J. 2016, ApJ, 832, 51
- Boersma, C., Bregman, J. D., & Allamandola, L. J. 2013, ApJ, 769, 117

- Boersma, C., Rubin, R. H., & Allamandola, L. J. 2012, ApJ, 753, 168
- Bregman, J., & Temi, P. 2005, ApJ, 621, 831
- Cami, J., Bernard-Salas, J., Peeters, E., & Malek, S. E. 2010, Sci, 329, 1180
- Cané, E., Palmieri, P., Tarroni, R., Trombetti, A., & Handy, N. 1998, JChPh, 106, 9004
- Cook, A. M., Ricca, A., Mattioda, A. L., et al. 2015, ApJ, 799, 14
- de Barros, A. L. F., Mattioda, A. L., Ricca, A., Cruz-Diaz, G. A., & Allamandola, L. J. 2017, ApJ, 848, 112
- Dunning, T. H., Jr. 1989, JChPh, 90, 1007
- Frisch, M. J., Pople, J. A., & Binkley, J. S. 1984, JChPh, 80, 3265
- Geballe, T. R., Tielens, A. G. G. M., Allamandola, L. J., Moorhouse, A., & Brand, P. W. J. L. 1989, ApJ, 341, 278
- Hony, S., Van Kerckhoven, C., Peeters, E., et al. 2001, A&A, 370, 1030
- Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, ApJS, 154, 18
- Hudgins, D. L., & Sandford, S. A. 1998, JPCA, 102, 329
- Hudgins, D. M., Bauschlicher, C. W., & Allamandola, L. J. 2005, ApJ, 632, 316
- Joblin, C., Tielens, A. G. G. M., Allamandola, L. J., & Geballe, T. R. 1996, ApJ, 458, 610
- Landsman, W. B. 1993, in ASP Conf. Ser. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco, CA: ASP), 246

- Langhoff, S. R. 1996, JPhCh, 100, 2819
- Mackie, C. J., Peeters, E., Bauschlicher, C. W., Jr., & Cami, J. 2015, ApJ, 799, 131
- Markwardt, C. B. 2009, in ASP Conf. Ser. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler (San Francisco, CA: ASP), 251
- Mattioda, A. L., Bauschlicher, C. W., Bregman, J. D., et al. 2014, AcSpA, 130, 639
- Mattioda, A. L., Bauschlicher, C. W., Ricca, A., et al. 2017, AcSpA, 181, 286 Peeters, E., Bauschlicher, C. W., Jr., Allamandola, L. J., et al. 2017, ApJ, 836, 198
- Peeters, E., Hony, S., Van Kerckhoven, C., et al. 2002, A&A, 390, 1089
- Perdew, J. P. 1986, PhRvB, 33, 8822
- Pirali, O., Vervloet, M., Mulas, G., Malloci, G., & Joblin, C. 2009, PCCP, 11, 3443
- Puget, J. L., & Léger, A. 1989, ARA&A, 27, 161
- Sellgren, K., Werner, M. W., Ingalls, J. G., et al. 2010, ApJL, 722, L54
- Stephens, P. J., Devlin, F. J., Chabalowski, C. F., & Frisch, M. J. 1994, JPhCh, 98, 11623
- Szczepanski, J., & Vala, M. 1993, ApJ, 414, 646
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1
- Zhang, Y., & Kwok, S. 2015, ApJ, 798, 37