Interpretable Deep-Learning Unveils Structure-Property Relationships in Polybenzenoid Hydrocarbons

Tomer Weiss,[†] Alexandra Wahab,[‡] Alex M. Bronstein,[†] and Renana Gershoni-Poranne[*,¶,‡]

[†]*Department of Computer Science, Technion - Israel Institute of Technology, Haifa, Israel*
[‡]*Laboratory for Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zurich, Switzerland*
[¶]*Schulich Faculty of Chemistry, Technion, Haifa, Israel - Israel Institute of Technology*

E-mail: rporanne@technion.ac.il

Abstract

In this work, interpretable deep learning was used to identify structure-property relationships governing the HOMO-LUMO gap and relative stability of polybenzenoid hydrocarbons (PBHs). To this end, a ring-based graph representation was used. In addition to affording reduced training times and excellent predictive ability, this representation could be combined with a subunit-based perception of PBHs, allowing chemical insights to be presented in terms of intuitive and simple structural motifs. The resulting insights agree with conventional organic chemistry knowledge and electronic structure-based analyses, and also reveal new behaviors and identify influential structural motifs. In particular, we evaluated and compared the effects of linear, angular, and branching motifs on these two molecular properties, as well as explored the role of dispersion in mitigating torsional strain inherent in non-planar PBHs. Hence, the observed regularities and the proposed analysis contribute to a deeper understanding of the behavior of PBHs and form the foundation for design strategies for new functional PBHs.

Introduction

Polybenzenoid hydrocarbons (PBHs) are molecules comprising multiple fused benzene rings. Such molecules are prevalent in chemistry and in materials science; they play key roles in various areas, from human health to functional materials, to the study of interstellar space.[1,2] They are known for their detrimental impact as carcinogens[3,4] and pollutants,[5–7] as well as for their advantageous impact as tunable organic semiconductors.[8–11] Important advances, such as the design of novel functional PBHs or development of degradation pathways for harmful compounds, rely on understanding their properties and reactivity. Therefore, it is no surprise that PBHs have been the subject of extensive investigation for several decades, nor that they continue to hold the attention of chemists and material scientists.

The study of PBHs extends through both theoretical and experimental research and mainly focuses on understanding their intrinsic reactivity,[12] developing new synthetic pathways,[13–15] and uncovering their structure-property relationships.[16–19] Whereas the first two avenues of research enable the preparation and application of new PBHs, the latter can inform us of which PBHs are worth synthesizing – i.e., which will

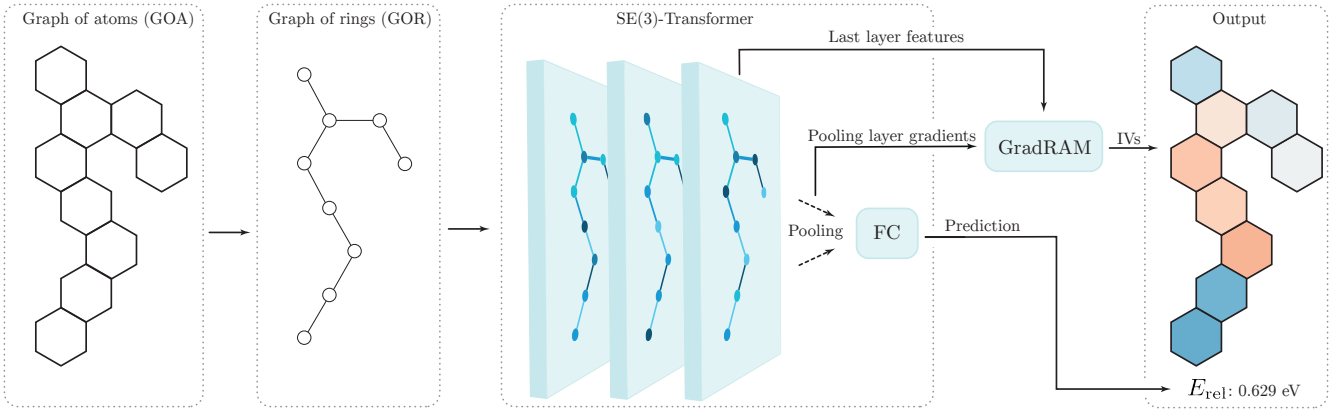

Figure 1: Schematic illustration of our prediction and interpretability pipeline: the graph of atoms is reduced to the graph of rings, which is the input for the SE(3)-Transformer model. The output is the predicted property. The GradRAM algorithm uses the last layer features and the pooling layer gradient with respect to the output to produce the importance value (IV) of each ring.

have desirable properties – and how we can tune these properties *via* structural modification.

Over the past few years, machine learning (ML) and, in particular, deep learning (DL) have become popular tools for performing such (quantitative) structure-property/activity relationship investigation.[20–25] Indeed, many successful models (wherein success is indicated by the model's out-of-training-sample predictive performance) have been implemented in chemistry and related fields. Nevertheless, these models are usually 'black-boxes', meaning that one cannot determine the nature of the learned relationship, or even judge whether it is a physically valid one. The recent introduction of interpretable models[26,27] has now made it possible to 'open' the 'black box' and examine what the model has learned, leading to substantiation of known physical and chemical concepts and to the discovery of new ones.[28–30] Such approaches have been applied within the chemistry domain to learn molecular properties and derive chemical insight. For example, by implementing an explanation mask that localizes on the most impactful characters, Goh *et al* demonstrated the interpretability of the SMILES2vec model in predicting solubility.[31] Xiong *et al* showed that inverting the AttentiveFP graph neural network and extracting the weights allows for interpretation, such as identification of aromatic subunits and intramolecular hydrogen bonding, which can aid molecular discovery.[32]

Methods

Data

The COMPAS-1D dataset of the COMPAS Project,[33] [1] was used. The dataset contains the density functional theory calculated structures and properties of ∼9K *cata*-condensed PBHs comprising 1–10 rings. In the current study, we focused on two molecular properties: the HOMO-LUMO gap (HLG) and the relative energy (E_{rel}). E_{rel} was defined as the difference in energy between each molecule and its respective lowest-energy isomer, where 'energy' refers to the total energy calculated at the optimized geometry using the B3LYP/def2-svp combination with Grimme's D3 dispersion scheme[34] and Becke-Johnson damping,[35,36] without zero-point correction. To probe the role of dispersion, we also used $E_{\text{rel}}^{\text{SCF}}$ – which is defined similarly, but does not include the dispersion correction (SCF stands for self-consistent field).

Prediction Model

We used the SE(3)-Transformer[37] as a prediction model for the electronic properties of the molecules. The model is a variant of the graph attention network (GAT),[38] which is, by construction, invariant under continuous 3D rotations and translations. As location and orienta-

[1], https://gitlab.com/porannegroup/compas

tion of input structures are arbitrary, they are treated as equivalence classes under the action of the special Euclidean group SE(3).

In our setting, the model receives as the input a graph $\mathcal{G}$ representing the chemical structure, with three-dimensional coordinate vectors $\mathbf{x}_i$ at each node i. In our experiments, nodes represent either individual atoms or the centroids of individual benzene rings. The nodes can be further endowed with feature tensors collectively denoted as $\mathbf{h}_i$. For the purpose of predicting properties for the entire structure, a permutation-invariant pooling layer is added at the end aggregating all node features into a single feature vector. A detailed explanation of the SE(3)-Transformers and how it achieves the desired invariance can be found in Section S1 of the SI.

SE(3)-invariant models, including the SE(3)-Transformer, have been developed specifically for modeling molecular structures. The unprecedentedly accurate data-driven prediction of protein folding achieved by AlphaFold2[39] is, undoubtedly, a triumph of this family of models.

GradRAM Interpretability

Neural networks are often referred to as 'black-box' models, which do not allow interpretation of their decision-making process. Interpretability is crucial for extracting chemical insight from a trained model. Over the last few years, methods have been introduced to help with the interpretability of graph neural networks.[40]

Gradient-weighted Regression Activation Mapping (GradRAM)[41] is the less-known variant of the popular Gradient-weighted Class Activation Mapping (GradCAM)[42] for regression tasks. The intuition behind both methods is to inspect the gradients with respect to the output in the last convolution layers (before the pooling), to identify the contribution of each node to the model's decision. While GradRAM was previously proposed only for regular convolutional neural networks, in this work we have adapted the method to graph neural networks in the following way: the importance value (IV) of each node i is calculated as the average of the node features in the last convolution layer weighted by the feature gradient in the pooling layer,

$$\mathrm{IV}(i) = \sum_k \beta_k h_{ki}^{(L)} \tag{1}$$

where $h_{ki}^{(L)}$ is the k-th feature of the last convolution layer, and β_k is the average gradient of the feature k in the pooling layer,

$$\beta_k = \frac{1}{n} \sum_i \frac{\partial y}{\partial h_{ki}^{(L)}} \tag{2}$$

with y denoting the model output. This method assigns to each node the magnitude and sign of its relative contribution to the output. A positive (negative) IV indicates that the node contributes to an increase (decrease) in the predicted property value. The IVs are unitless.

Molecular Representation

Most approaches applying graph neural networks in the field of chemistry use a molecular graph, in which the atoms are the nodes and the bonds are the edges, as their representation (graph of atoms – GOA). In this work, we tested the standard GOA as well as a graph of rings (GOR) representation. In the GOR, each node represents a benzene ring (the coordinates of the node are the centroid of the ring) and each edge is the fused bond of two adjacent rings. This representation can be extended to heteroatom-containing systems by setting the ring type as a node feature.

Using the GOR affords the GradRAM interpretability at the level of rings, which are our focus, as they are the basic building blocks of PBHs. Whereas in many organic compounds, individual atoms or groups of atoms may be perceived as building blocks, in PBHS a single atom reveals very little about the molecule, even if its nearest neighbors are considered. We previously demonstrated that, for PBHs, ring- or subunit-based perspectives reveal structure-property relationships governing multiple electronic properties in an intuitive and simple manner.[43–46] The ring perspective presents a clear advantage for molecular design as well: the GOR provides the IV of each ring, i.e., how

it affects the property of interest (e.g., raising or lowering the HLG). Based on this, one can formulate local modifications to the molecular structure (e.g., adding or removing rings) in order to tune the property in a desired direction. That said, future work will focus on poly(hetero)cyclic aromatic systems, in which changes to specific atoms (e.g., C to N), will be relevant and will be explored appropriately.

Using the GOR representation also reduces the complexity of the graph, while still providing an adequate representation of the molecule. This alleviates the required computational resources substantially.

Results and Discussion

Over the past few years, we have been studying the structure-property relationships of PBHs (and polycyclic aromatic systems, PASs, in a more general sense) using the approach of deconstructing them into smaller components. With this perspective, we discovered that the magnetic properties of larger *cata*-condensed PASs can be predicted with an additivity scheme,[43–45] using tricyclic components (depicted in Figure 2A). Further, we revealed that the annulation sequence (see Figure 2B) of PBHs enables prediction of a whole host of molecular properties, including the singlet-triplet gap, the location of the spin density in the first triplet excited state, and the aromatic character in both ground and triplet excited states.[46]

One of the main conclusions from this body of work was that many of the characteristics of *cata*-condensed PBHs are encoded in the sequence of tricyclic annulations – linear or angular. This entails that the representation of such molecules can be reduced dramatically, and even encoded as a text sequence. We recently reported on the development of this representation, named LALAS, and its performance with several ML models.[47] In this work, we reduced the dimensionality of the molecular representation using a different approach: the GOR representation (see Molecular Representation for further details). Such graph representations

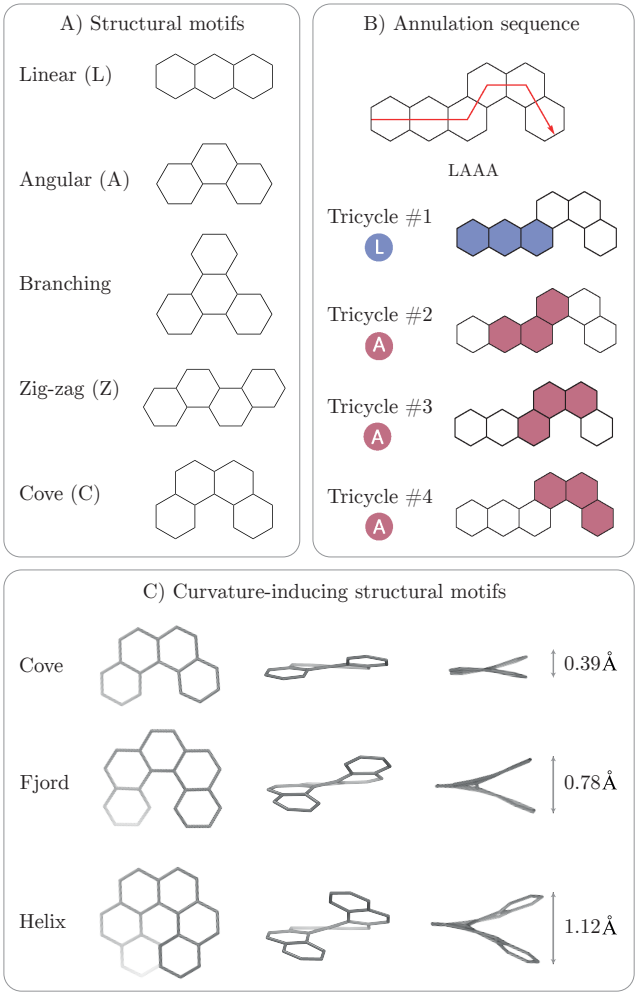

Figure 2: A) Structural motifs. B) Illustration of the annulation sequence for a hexacyclic PBH example. C) Motifs that induce deviation from planarity (from left: top, front, and side views). The σ_z for each structure is given in Å units. Double bonds and hydrogens omitted for clarity.

have been successfully employed previously in the investigation of PBHs. Balaban,[48,49] Gutman,[50] Aihara,[51] and others[52] have derived a great deal of chemical insight using such representations within graph-theoretical constructs. Yet, to the best of our knowledge, this is the first attempt to use the GOR in conjunction with interpretable DL techniques.

In the following sections we discuss the two main aspects of this approach: a) predictive performance, and b) interpretability and chemical insight.

Predictive Performance

For this study, we chose to focus on two molecular properties: HOMO-LUMO gap (HLG) and relative energy (E_{rel}). To begin, we compared the performance of our SE(3)-Transformer in predicting these two properties to two state-of-the-art methods: a) CHEM-BERT,[53] a natural language processing (NLP)-based method that uses SMILES as the input; and b) AttentiveFP,[32] a GAT-based method that uses (molecular) graphs as the input. The results (Table 1) show that our method out-performed these two methods in predicting the HLG by a significant margin: the mean absolute error (MAE) for both SE(3) models was smaller than the CHEM-BERT MAE by a factor of ∼3 and smaller than the MAE of AttentiveFP by a factor of ∼5. Though both types of representation – GOA and GOR – performed comparably for predicting the HLG, we note that the latter required a markedly shorter training time (∼6 hours) than the former (∼43 hours). For prediction of E_{rel}, the SE(3)-GOA model showed MAEs ∼7 and ∼10 times smaller, and the SE(3)-GOR model showed MAEs ∼4 and ∼5 times smaller than CHEM-BERT and AttentiveFP, respectively.

Some chemical insight could already be gleaned at this stage, simply from comparing the performance of the two types of graph inputs. We observed that the SE(3)-GOA and the SE(3)-GOR models had essentially identical predictive performance for the HLG (0.34% and 0.38% error relative to the mean HLG value, respectively) but more noticeably different error values for the E_{rel} (0.62% and 1.29% error relative to the mean E_{rel} value, respectively). This suggests that the structural information contained in the GOA and not contained in the GOR – i.e., the specific attributes of the C–C and C–H bonds – influences the prediction of the E_{rel}. In other words, the model is sensitive to variations in bond lengths and angles. Such variations result from different annulation topologies and can incur strain energy, which contributes to the E_{rel} value. In contrast, the observation that the GOR representation performed comparably well to the GOA representation for prediction of the HLG suggests that this property is much less sensitive to such structural variations. Overall, these results demonstrate that the SE(3) models are successful at predicting the molecular properties and indicate that the models are sensitive to specific (and subtle) features that affect the structure-property relationships.

Interpretability

For the interpretability investigation, we chose to continue with the GOR representation. The rationale underpinning this choice was simple: we sought the minimal unit that has chemical meaning. As explained in the Molecular Representation section, for PBHs, individual rings are meaningful chemical moieties, while individual atoms are not. Note that in this text we use the annulation sequence notation (introduced in Ref. 46 and shown in Figure 2B), which describes PBHs using 'L', 'A', '(' and ')' characters for linear, angular, and branching annulations, respectively (Figure 2A).

Validity of the model – interpretability of the HLG prediction.

Our previous work uncovered a set of regularities pertaining to the relationship between the singlet-triplet energy gap and structural features.[46] In particular, we found that the structural feature that determines the singlet-triplet gap is the Longest L feature, i.e., the longest sequence of linearly-annulated rings. We hypothesized that this would also be true for the HLG. Indeed, the kernel density estimates (KDEs) of the various subsets of data show that the value of the HLG decreases as the Longest L increases (Figure 3). The effect is marked enough that the individual substrings show distinct ranges in the distribution plots.

As a test of the reliability of the interpretability model, we used these observations to perform a 'reality-check', which involved training new models on subsets of the data. First, we excluded all molecules with an L-motif, which led to very poor predictive performance. Subsequently, we excluded all molecules with in-

Table 1: **Comparison of different models for prediction of molecular properties.**

Model	HUMO-LOMO gap		Relative energy	
	MAE(eV)	Relative	MAE(eV)	Relative
CHEM-BERT	0.033±0.015	1.04±0.49%	0.037±0.020	5.11±2.68%
AttentiveFP	0.053±0.052	1.66±0.63%	0.053±0.068	7.13±5.13%
SE(3)-GOA	**0.011±0.004**	**0.34±0.13%**	**0.005±0.002**	**0.62±0.33%**
SE(3)-GOR	0.012±0.002	0.38±0.05%	0.010±0.003	1.29±0.36%

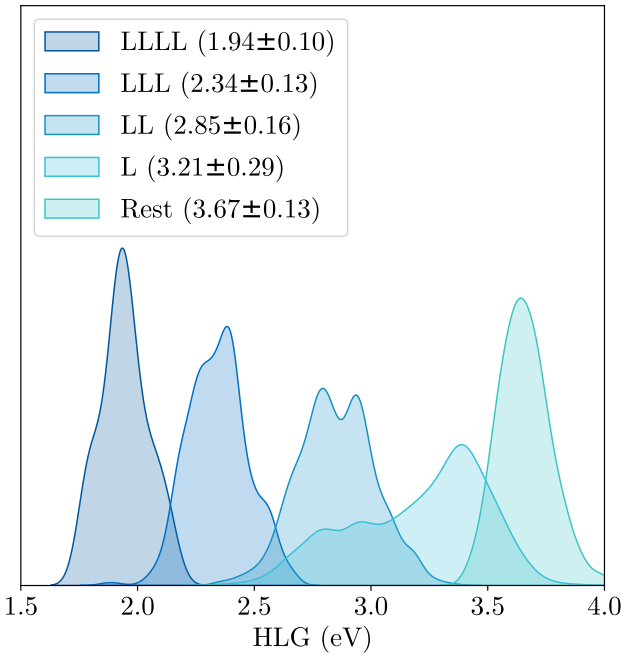

Figure 3: KDEs of the distributions of the HLG for subsets of the PBH class containing L, LL, LLL, and LLLL motifs.

creasingly longer Longest L-motifs (i.e., LL, LLL, LLLL, etc.). We observed that the lower boundary of the predicted HLG decreased in a 'step-wise' manner according to the Longest L sequence included (in line with the distribution ranges shown in Figure 3). In contrast, excluding all molecules of a particular number of rings did not adversely affect the predictive ability (see Section 1 and Figure S1 in the Supporting Information for further details).

We then conducted a series of paired tests, in which we removed specific rings from PBH structures and compared the HLG of the parent and daughter molecules (Figure 4A and B). We observed that, indeed, removal of rings with negative IVs resulted in increased HLGs, while removing rings with positive IVs resulted in decreased HLGs. In addition to this visual inspection, we also performed a statistical analysis that substantiated this relationship (Figure 4C). Thus, these results verified that the model was capturing the correct physical relationships.

Insights into structure-property relationships for HLG

Having confirmed that the model is capable of 'rediscovering' chemically valid structure-property relationships, we queried it further to substantiate additional findings and to seek out new connections. As the first step, we plotted the distributions of IVs for the middle rings in L, A, and branching motifs (Figure 5A). These plots reveal that the L motif has predominantly negative IVs and the angular/branching motifs have predominantly negligible IVs that tend towards positive values (mean IVs of −0.063, 0.016 and 0.023, for linear, angular, and branching, respectively). This suggests that, in general, linear stretches tend to decrease the HLG more strongly than the angular/branching points tend to increase it. Following these general conclusions, we performed a more in-depth analysis, focusing on each of these structural features in turn. These conclusions are described below.

Linear motifs. Initially, we focused on the importance of the linear stretches and how their influence is manifested. We performed visual inspections of a random selection of compounds containing linear stretches of various lengths and observed that, indeed, the rings contained within these structural motifs have the most dominant IVs in each molecule (representative examples in Figure 5C). As outlined above, the middle rings of L motifs generally exhibit a neg-

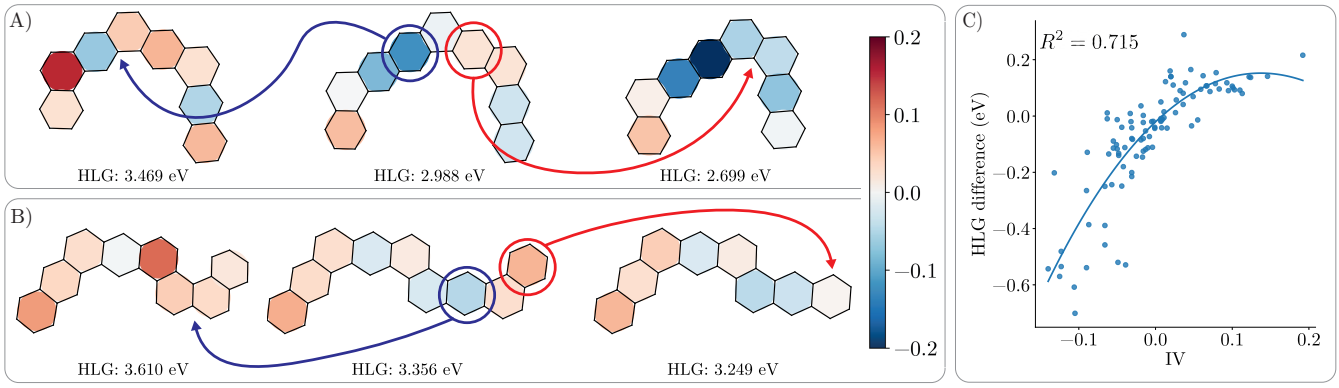

Figure 4: Paired-tests demonstrating the effect of removing rings with different IVs. A), B) Representative examples of a parent system (center) and daughter systems following removal of a positive-IV ring (right) or a negative-IV ring (left). The sign and magnitude of the IVs are denoted with color, according to the color bar. The HLG of the daughter systems decreases and increases accordingly. The circled rings are those removed. The arrow points to the location in the daughter molecule where the ring has been removed. C) Correlation between the IV value of the removed ring and the change in HLG value.

ative IV. Our visual inspection further revealed that these values become more negative as the L-sequence elongates, and also concentrate on the central rings of the linear stretch. Moreover, the IVs show that for molecules with several multi-L motifs, the longest one is the dominant one (e.g., Figure 5C4). This is yet another conclusion that is in line with our previous findings for triplet-state PBHs[46] and is also in agreement with the well-documented relationship between elongation of polyacenes and the decrease in their HLG.[54–56]

Angular motifs. We then turned to consider the effect of angular annulations (representative examples in Figure 5D). We distinguish between molecules that have only A motifs (i.e., 'all-A') and those that have a mixture of L and A motifs. We note that, Whereas L motifs have only a single direction, A motifs can form in two directions (clockwise and counterclockwise). Therefore, the number of possible isomers increases substantially with each additional annulation.[46] As a result, there are multiple all-A PBHs, but only one all-L for each isomer family. Among these all-A compounds, polyphenacenes (i.e., 'zig-zag'-type structures; Figure 2A) and 5D1) are a specific and well-known subset and have been quite extensively studied.[57–60] To the best of our knowledge, however, a systematic study has not been reported on the structure-property relationships of other all-A PBHs or on molecules containing mixtures of A and L motifs.

For the all-A PBHs, we observed that the IVs are almost always positive and are in the higher range of IVs. Small negative IVs are seen only in terminal rings (Figure 5B). The most influential rings are those that are part of a 'zig-zag' motif, rather than a 'cove' motif (i.e., C and Z motifs; Figure 2A). If there is more than one Z motif, the one that is farther from the terminus of the molecule and does not overlap with a C motif contains the more influential rings (e.g., Figure 5D2). These observations echo our previous findings of a hierarchy of structural motifs for determining the spin density distribution in the triplet excited state.[46]

For molecules comprising both A and L motifs, we observed that the A subunit(s) that 'break(s)' the linear stretch generally has a stronger positive IV (e.g., Figure 5D2-D4). However, overall, the IVs for the A rings are small. To summarize, the middle rings of A motifs have high IVs only when there are no L motifs in the PBH. When the structure contains both A and L motifs, the middle rings of the A motifs generally have small IVs (in absolute value), indicating that, in those cases, their contribution to the HLG value is relatively minor.

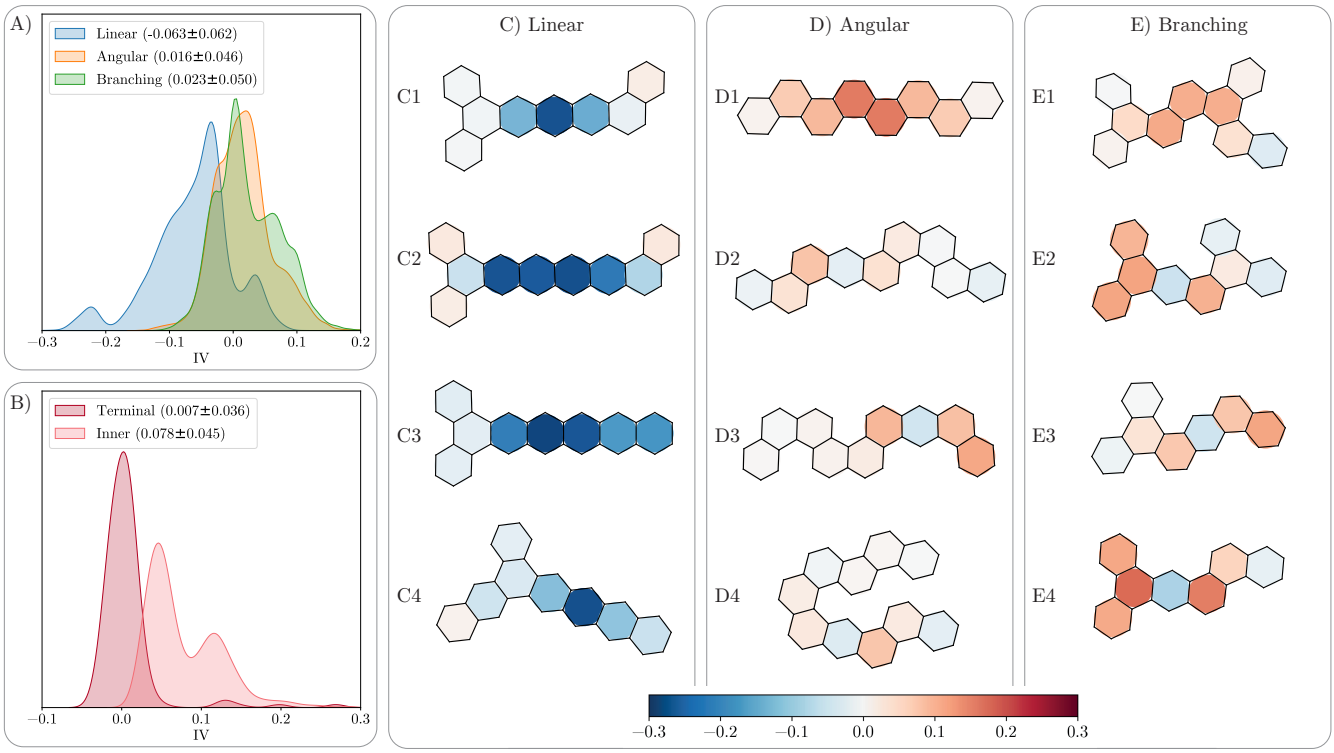

Figure 5: IVs obtained from the HLG-prediction model. A) KDE distributions of the IVs of the linear, angular, and branching motifs. B) KDE distributions of the angular motifs at terminal and inner locations for the all-A PBH isomers. C) Examples of PBHs demonstrating regularities of linear motifs. D) Examples of PBHs demonstrating regularities of angular motifs. E) Examples of PBHs demonstrating regularities of branching motifs.

Branching motifs. Whereas linearly and angularly annulated PBHs have been quite broadly investigated (in some cases exhaustively so, e.g., polyacenes and polyphenacenes), branched PBHs appear to have received less attention. This is perhaps due to the greater complexity of such structures and the difficulty of disentangling branching points from the L/A structural motifs. Based on the success of the interpretable model in finding the structure-property relationships for the unbranched PBHs (*vide supra*), we envisioned it might allow us to reveal new insights about the effects of branching. The distribution of IVs for the branching motifs shows that, for the most part, these have negligible IVs (Figure 5A). However, in some cases, their IVs become noticeably larger. Visual inspection revealed a few regularities, including that the relative size of the IVs depends on the type of motif next to the branching point: if it is adjacent to an A motif, the branching motif will have relatively low IVs, and even negative values in the peripheral rings (e.g., Figure 5E1); if it is adjacent to an L motif, the branching motif will have relatively high positive IVs (e.g., Figure 5E2-E4). These patterns indicate that the branching, in and of itself, does not have a dominant effect on the HLG. Rather, its contribution to the HLG is reliant on the absence/presence of L motifs. This is corroborated by the distribution of HLGs showing minimal or no dependence on the presence/number of branches (Figure S2 in the Supporting Information).

Extension to E_{rel}

We next turned to examine a different molecular property – relative energy. In particular, we focused on the effect of structural motifs and their role in enabling stabilizing dispersion interactions. To this end, we compared two different types of energy: E_{rel} and $E_{\mathrm{rel}}^{\mathrm{SCF}}$, where the only difference between the two terms is that the former contains the dispersion correc-

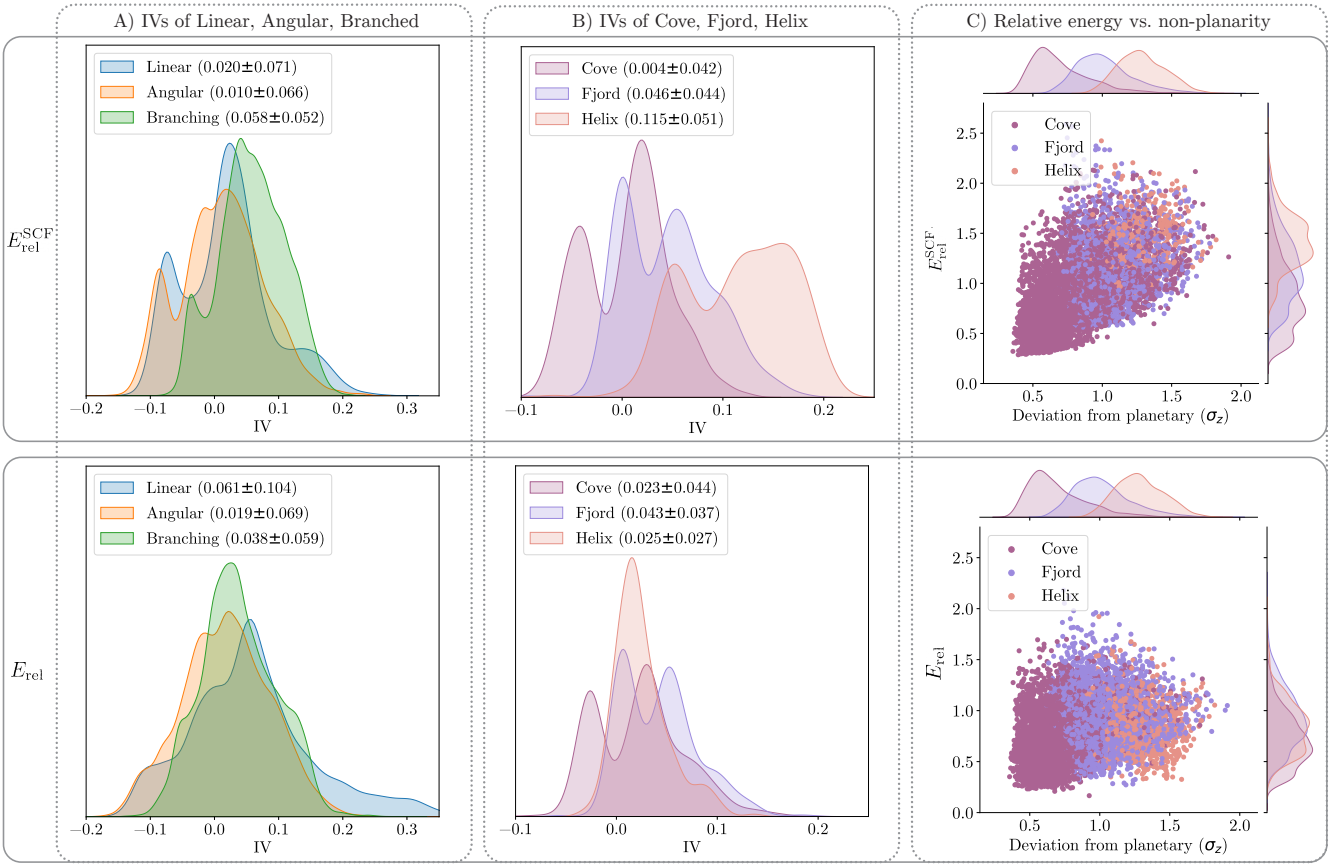

Figure 6: Comparison of results for the $E_{\text{rel}}^{\text{SCF}}$ (top) and E_{rel} (bottom) models. A) KDE distributions of the IVs of the linear, angular, and branched motifs. B) KDE distributions of the IVs of the cove, fjord, and helix motifs. C) Scatter plot of the relative energy against the deviation from planarity, colored by the cove, fjord, and helix structural motifs.

tion and the latter does not (see Data section for the definitions of these two energies).

Again, we began by plotting the distribution of IVs of the middle rings in each of the three structural motifs. In contrast to the HLG case, here we observed very small differentiation between the structural motifs and a large degree of overlap and a trend of *angular* < *branching* < *linear* (Figure 6A bottom). The model trained on $E_{\text{rel}}^{\text{SCF}}$ yielded a different trend of *angular* < *branching* < *linear* (Figure 6A, top). This was the first indication that the dispersion correction leads to a qualitatively different interpretation, i.e., might be overriding other inherent structure-related behaviors. Accordingly, the absence of dispersion corrections may bring these effects to light. Thus, we sought to extract insight into the effects of the various structural motifs on the relative stability of the PBH isomers and on the role of dispersion in this context.

Linear motifs. Visual inspection of the IVs for the E_{rel}-prediction model showed that the most strongly positive IVs are in the linear stretches and these become increasingly positive as the linear stretch elongates (see Figures S3 and S4 in the Supporting Information). This is in marked contrast to the HLG prediction, where the same motifs contributed the most strongly negative IVs. Interestingly, for the $E_{\text{rel}}^{\text{SCF}}$-based model, the same rings have even higher IVs (Figure 7A; Figures S3 and S4). In other words, the linear stretches contribute to higher relative energy of the PBH isomers and longer linear stretches have reduced dispersion interactions.

The distribution of the molecular properties verified these observations: we observed that the relative energy increases with the elonga-

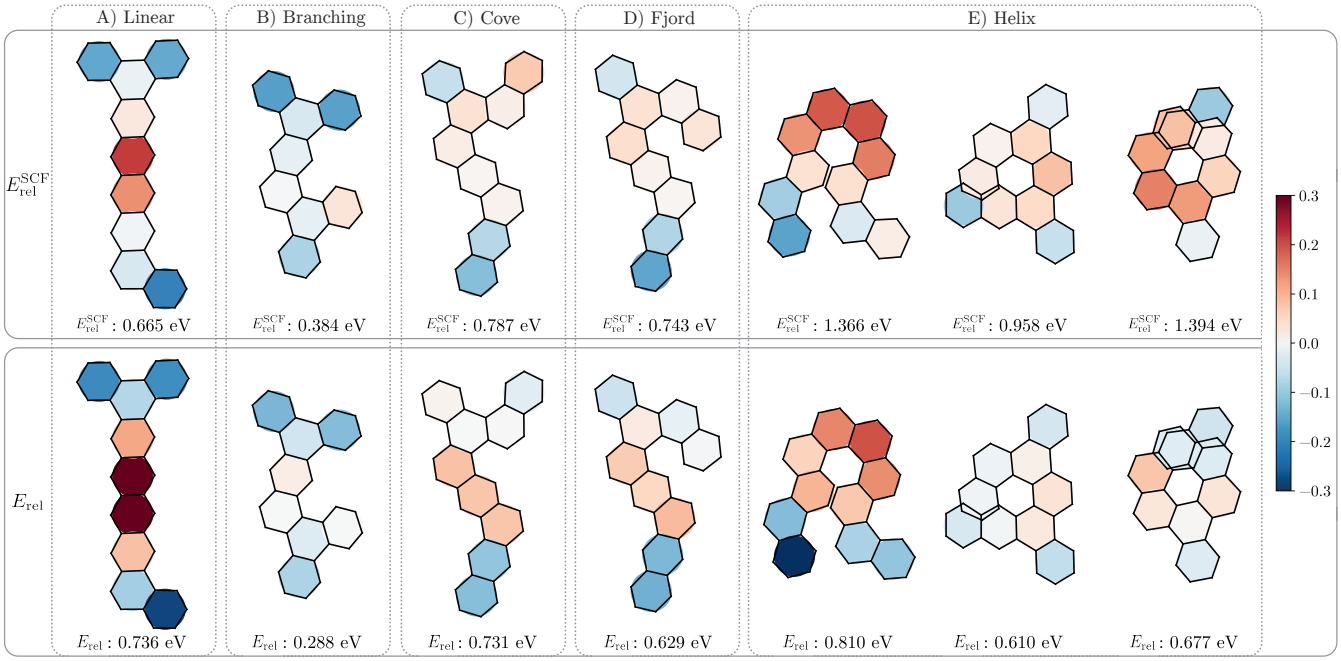

Figure 7: Visual depictions of IVs obtained from the E_{rel}^{SCF} (top) and the E_{rel} (bottom) models. Examples containing a A) linear motif; B) branching motif; C) cove motif; D) fjord motif; E) helix motif. Colors within the ring correspond to IV values, as shown on the color bar (right).

tion of the linear stretch. For L, LL, LLL, and LLLL, respectively, we found mean E_{rel} values of 0.71, 0.76, 0.94, 1.15 eV and mean $E_{\text{rel}}^{\text{SCF}}$ values of 0.83, 0.80, 0.91, and 1.14 eV (Figure S5 in the Supporting Information). Notably, the difference between the two types of relative energy changes as the L-sequence elongates ($E_{\text{rel}} - E_{\text{rel}}^{\text{SCF}} = -0.12$, -0.04, 0.03, and 0.01 eV for L, LL, LLL, and LLLL, respectively). In other words, as the interpretable model indicated, the dispersion interactions stabilize molecules with short linear stretches but do not stabilize molecules with long ones. This can be rationalized in the following way: for unsubstituted PBHs, the expected type of stabilizing dispersion interactions is $\pi - \pi$ stacking. Such interactions can only occur if the geometry of the molecule enables spatial overlap of individual rings. Molecules with longer linear stretches are expected to be (mostly) planar and therefore are not expected to have significant stabilizing $\pi - \pi$ dispersion interactions.

Angular motifs. Visual inspection of all-A PBHs showed that the terminal rings tend toward negative IVs while inner rings tend toward positive IVs. The distributions of IVs for the terminal and inner rings corroborated this observation (Figure S6 in the Supporting Information). They also revealed bi-modalities, which are caused by the rings belonging to either C or Z motifs (Figure S7 in the Supporting Information). Unsurprisingly, the effect is more noticeable for the terminal rings, where there is no overlap with additional motifs. We observed that the IVs of terminal rings in C motifs are slightly positive when dispersion is omitted and slightly negative when it is included. In contrast, the IVs of rings in Z motifs tend to be negative for both cases (Figure S7 in the Supporting Information). In other words, C motifs are inherently destabilizing and this effect is mitigated by dispersion interactions, whereas Z motifs are inherently stabilizing. Interestingly, this is similar to the hierarchy of stabilization we found in the first excited triplet state.[46]

To probe this further, we considered the *cove*, *fjord*, and *helix* series of motifs, which are all made of sequential angular annulations in the same direction with an increasing number of rings (four, five, and six rings, respectively; Figure 2C). The elongation of the motif leads to an increase in curvature and in deviation from planarity (σ_z values displayed in Figure 2C), and therefore also increases the inherent torsional/helical strain. Accordingly, all other con-

ditions being equal, it is expected that the IVs of rings should follow the trend *cove* < *fjord* < *helix*. Indeed, for the model trained on $E_{\text{rel}}^{\text{SCF}}$, this is the trend that is observed (Figure 6B, top). However, for the model trained on E_{rel}, the trend in IVs changes to *cove* < *helix* < *fjord* (Figure 6B, bottom). The shift in trends indicates that the three motifs have different extents of stabilization via dispersion interactions.

The change in IVs shows that inclusion of the dispersion correction mainly affects the helix motif, and visual inspection revealed that this is primarily due to the rings at the ends of the helical motifs. We observed that, for the $E_{\text{rel}}^{\text{SCF}}$-based model, all of the rings of the helix motifs have positive IVs (e.g., Figure 7E, top). However, in the E_{rel}-based model, the end-rings that overlap (or partially overlap) in space display negative IVs (e.g., Figure 7E, bottom middle and right examples). We hypothesized that this is due to the participation of these end-rings in $\pi - \pi$ stacking interactions, which have a stabilizing effect and therefore lower the E_{rel}. Additional corroboration can be seen in counter-examples, one of which is displayed in Figure 7E, left example. This molecule contains a helix motif in which the end-rings do not overlap and, notably, the IVs of the helix end-rings remain positive even when dispersion interactions are included. Thus, the interpretability confirms that the helix motif itself is inherently destabilizing – which we attribute to helical strain – but the extent of its destabilization is mitigated by dispersion interactions, and specifically by the $\pi - \pi$ stacking of the end-rings.

To show the generality of these conclusions for the entire data set, Figure 6C displays scatter plots of the relative energy against σ_z (which is a measure of deviation from planarity), colored according to cove, fjord, and helix subsets. The plot for $E_{\text{rel}}^{\text{SCF}}$ (top) shows that the relative energy generally increases as the deviation from planarity increases, which follows along the same trend of *cove* < *fjord* < *helix* (mean $E_{\text{rel}}^{\text{SCF}}$ values are 0.746, 1.116, and 1.363 eV for cove, fjord, and helix, respectively). The plot for E_{rel} (bottom) shows that, indeed, the subset of data most influenced by addition of dispersion corrections is that of the helix-containing molecules (mean E_{rel} values are 0.709, 0.923, 0.842 eV). These trends and the relative effect on the individual motifs are in agreement with the results observed for the IVs themselves.

Thus, the interpretability of the two models and the comparison between them enable identification of the two main components influencing the E_{rel} in non-planar PBHs and demonstrate the cost of helical strain and the role of dispersion in mitigating this cost.

For molecules that contain both L and A motifs, the behavior is in line with the conclusions above. Due to their relatively strong destabilizing effects, the $E_{\text{rel}}^{\text{SCF}}$-model assigns the larger positive IVs to A motifs. Inclusion of dispersion corrections mitigates the strain of the curved motifs, to the extent that their overall destabilizing effect lessens and, therefore, the E_{rel}-based model assigns higher positive IVs to the L motifs.

Branching motifs. The presence and number of branching motifs clearly influence the E_{rel}, as seen in Figure 8A. Increasing the number of branches in the molecule leads to an increase in E_{rel} (mean E_{rel} values of 0.64, 0.75, 0.93, 1.24 eV for 0, 1, 2, and 3 branching points, respectively). This is in line with the IVs of branching motifs having a positive mean value (Figure 6A). When the dispersion correction is omitted, the mean value of the IV increases, which indicates that, on average, branching motifs are stabilized through dispersion effects.

Following our analysis on the behavior of the angular motifs (*vide supra*), we hypothesized that the trends observed for the branching motifs may be due not to the branching, *per se*, but rather to the relationship between branching points and the formation of curved structures. As shown in Figure 8B, the formation of the various curved motifs depends on the number of branching points, n: as n increases, so do the percentages of fjord and helix motifs in the molecules. Hence, the increase in n is linked to greater deviation from planarity and higher strain, i.e., higher relative energy. This is also in line with the observation that the effect of branching is dampened when dispersion is taken into account. Note that, while

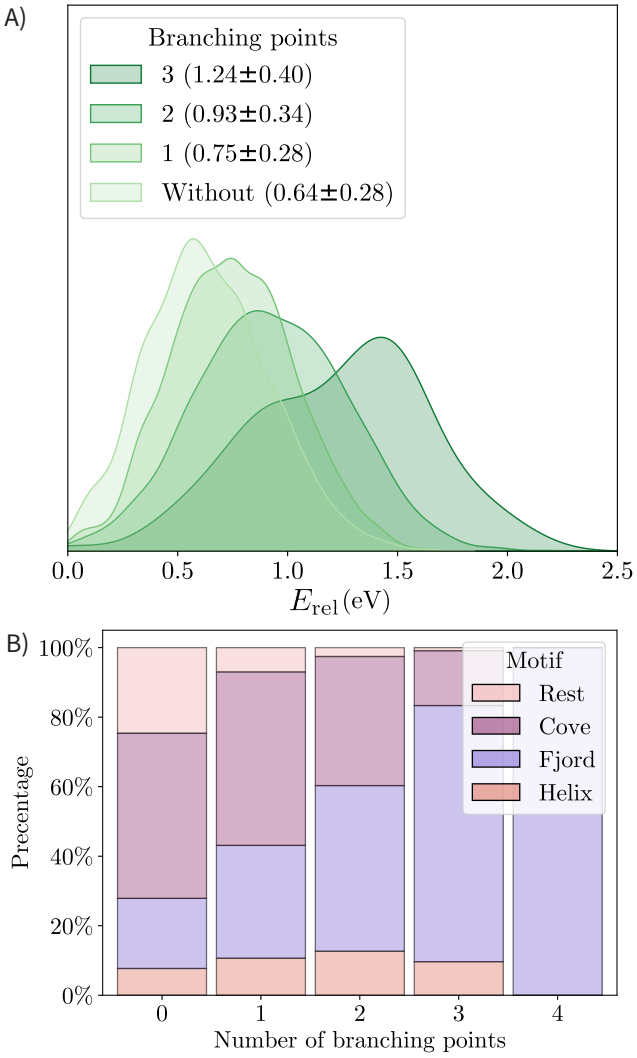

Figure 8: A) KDEs of the distribution of E_{rel} for subsets of the PBH class containing 0, 1, 2, and 3 branching points. B) Percentage of cove, fjord, and helix motifs in molecules containing 0, 1, 2, 3, and 4 branching points.

the percentage of fjord motifs increases steadily, the helix motifs decrease at $n = 3$ and disappear completely at $n = 4$. This is because our dataset contains molecules of up to 10 rings, and having four branching points precludes a single branch of six rings (necessary for forming a helix). For larger isomers, it is expected that the percentage of helices would continue to increase.

Conclusions

In this work, we reported on the application of interpretable deep-learning models to identify structure-property relationships in the archetypal polycyclic aromatic systems, PBHs. In principle, one may apply such methods to any type of molecules. However, the ability to gain meaningful understanding hinges on defining representations and motifs that are amenable to interpretation. Our GOR reductive representation enabled us to extract meaningful chemical insight while also substantially reducing training times. By assigning IVs to individual rings, we could combine the DL interpretability model with our PBH-specific structural motif-based analysis, which highlights patterns that are intuitive to grasp. Importantly, the relationships uncovered relate directly to the unique characteristics of the studied molecules.

The investigation focused on two individual properties, the HLG and the E_{rel}, and their relationship to the structural motifs. For the HLG, the main findings were:

1. the HLG is primarily determined by the Longest L sequence in the molecule – the longer the linear stretch, the smaller the HLG;

2. angular motifs increase the HLG;

3. in the absence of linear motifs, the Z motif is more influential than the C motif in determining the HLG;

4. branching points do not have a great impact, in and of themselves;

5. the effect of branching points depends on the adjacent motif.

For the E_{rel}, the main findings were:

1. linear stretches increase the E_{rel} – the longer the linear stretch, the greater the increase in E_{rel};

2. torsional/helical strain destabilizes of PBH isomers following the trend $cove < fjord < helix$;

3. dispersion interactions stabilize the motifs, along the trend $cove < fjord < helix$;

4. as a result of these opposing effects, the apparent trend in E_{rel} is $cove \leq helix \leq fjord$;

5. the helix motifs are stabilized most strongly due to $\pi - \pi$ interactions in overlapping rings;

6. the apparent relationship between E_{rel} and the number of branching points is due to the formation of longer (i.e., more distorted) curving motifs, and not the branching itself.

We note that, in a previous study that used interpretable ML models[47] and text-based molecular representations, the same motifs were identified as being influential, however, the direction of their effect was unknown. Therefore, it was not possible to develop a rationalization for the structure-property relationships. In this work, the interpretability of the models allowed identification the influential motifs as well as evaluation of the magnitude and direction of their effect. These results made it possible to construct a chemically-informed rationalization, which is both intuitive and informative.

In summary, this work has demonstrated application of the GradRAM interpretable DL method for obtaining chemical insight and advancing molecular design. We have shown the advantages of combining this method with ring-based molecular graph representation and with the perception of large PBHs as sequences of smaller structural motifs. The regularities we identified not only give a deeper understanding of the effects governing the behavior of PBHs, but also provide guiding principles for the rational design of PBHs. We are currently working on expanding our investigation to additional types of PASs, such as poly(hetero)cyclic aromatic systems and *peri*-condensed PBHs.

Data Availability

All data used in this project was obtained from the COMPAS Project,[33] dataset COMPAS-1D, a freely available data repository[2]. All code used to train, test, and analyze the models and the data is provided free of charge at https://gitlab.com/porannegroup/PBHs-interpretability.

Acknowledgement A.W. and R.G.P. express their deep gratitude to Prof. Dr. Peter Chen for his scientific and financial support and for his mentorship. A.W. and R.G.P. also thank the Branco Weiss Fellowship for supporting this work under a Society in Science fellowship. R.G.P. is a Branco Weiss Fellow and a Horev Fellow. A.M.B. and T.W. are supported by the ERC StG EARS.

Supporting Information Available

The following files are available free of charge.

- Supporting Information PDF file: all additional figures mentioned in this text, including textual description and relevant discussion.
- GitLab repository.

References

(1) Herbst, E. The chemistry of interstellar space. *Angewandte Chemie International Edition in English* **1990**, *29*, 595–608.

(2) Li, A. Spitzer's perspective of polycyclic aromatic hydrocarbons in galaxies. *Nature Astronomy* **2020**, *4*, 339–351.

(3) Straif, K.; Baan, R.; Grosse, Y.; Secretan, B.; El Ghissassi, F.; Cogliano, V. Carcinogenicity of polycyclic aromatic hydrocarbons. *The lancet oncology* **2005**, *12*, 931–932.

(4) Kim, K.-H.; Jahan, S. A.; Kabir, E.; Brown, R. J. A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environment international* **2013**, *60*, 71–80.

[2] https://gitlab.com/porannegroup/compas

(5) Haritash, A., et al. A comprehensive review of metabolic and genomic aspects of PAH-degradation. *Archives of Microbiology* **2020**, *202*, 2033–2058.

(6) Manzetti, S. Polycyclic aromatic hydrocarbons in the environment: environmental fate and transformation. *Polycyclic Aromatic Compounds* **2013**, *33*, 311–330.

(7) Shuttleworth, K. L.; Cerniglia, E. Environmental aspects of PAH biodegradation. *Applied biochemistry and biotechnology* **1995**, *54*, 291–302.

(8) Aumaitre, C.; Morin, J.-F. Polycyclic aromatic hydrocarbons as potential building blocks for organic solar cells. *The Chemical Record* **2019**, *19*, 1142–1154.

(9) Anthony, J. E. Functionalized acenes and heteroacenes for organic electronics. *Chemical reviews* **2006**, *106*, 5028–5048.

(10) Lee, C. W.; Kim, O. Y.; Lee, J. Y. Organic materials for organic electronic devices. *Journal of Industrial and Engineering Chemistry* **2014**, *20*, 1198–1208.

(11) Figueira-Duarte, T. M.; Mullen, K. Pyrene-based materials for organic electronics. *Chemical reviews* **2011**, *111*, 7260–7314.

(12) Fernández, I. Understanding the reactivity of polycyclic aromatic hydrocarbons and related compounds. *Chemical Science* **2020**, *11*, 3769–3779.

(13) Reizer, E.; Viskolcz, B.; Fiser, B. Formation and growth mechanisms of polycyclic aromatic hydrocarbons: A mini-review. *Chemosphere* **2021**, 132793.

(14) Feng, X.; Pisula, W.; Müllen, K. Large polycyclic aromatic hydrocarbons: synthesis and discotic organization. *Pure and Applied Chemistry* **2009**, *81*, 2203–2224.

(15) Harvey, R. G. Advances in the synthesis of polycyclic aromatic compounds. *Current Organic Chemistry* **2004**, *8*, 303–323.

(16) Jiang, H. Organic Ambipolar conjugated molecules for electronics: synthesis and structure–property relationships. *Macromolecular rapid communications* **2010**, *31*, 2007–2034.

(17) Uy, R. L.; Price, S. C.; You, W. Structure-Property Optimizations in Donor Polymers via Electronics, Substituents, and Side Chains Toward High Efficiency Solar Cells. *Macromolecular rapid communications* **2012**, *33*, 1162–1177.

(18) Kroon, R.; Mengistie, D. A.; Kiefer, D.; Hynynen, J.; Ryan, J. D.; Yu, L.; Müller, C. Thermoelectric plastics: from design to synthesis, processing and structure–property relationships. *Chemical Society Reviews* **2016**, *45*, 6147–6164.

(19) Blouin, N.; Leclerc, M. Poly (2, 7-carbazole) s: structure- property relationships. *Accounts of chemical research* **2008**, *41*, 1110–1119.

(20) Hessler, G.; Baringhaus, K.-H. Artificial intelligence in drug design. *Molecules* **2018**, *23*, 2520.

(21) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative structure–property relationship modeling of diverse materials properties. *Chemical reviews* **2012**, *112*, 2889–2919.

(22) Mitchell, J. B. Machine learning methods in chemoinformatics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 468–481.

(23) Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* **2019**, *5*, 1–36.

(24) Ekins, S. The next era: deep learning in pharmaceutical research. *Pharmaceutical research* **2016**, *33*, 2594–2603.

(25) Tang, W.; Chen, J.; Wang, Z.; Xie, H.; Hong, H. Deep learning for predicting toxicity of chemicals: A mini review. *Journal of Environmental Science and Health, Part C* **2018**, *36*, 252–271.

(26) Kovalerchuk, B.; Ahmad, M. A.; Teredesai, A. Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. *Interpretable artificial intelligence: A perspective of granular computing* **2021**, 217–267.

(27) Roscher, R.; Bohn, B.; Duarte, M. F.; Garcke, J. Explainable machine learning for scientific insights and discoveries. *Ieee Access* **2020**, *8*, 42200–42216.

(28) Rodriguez-Perez, R.; Bajorath, J. Explainable Machine Learning for Property Predictions in Compound Optimization: Miniperspective. *Journal of medicinal chemistry* **2021**, *64*, 17744–17752.

(29) George, J.; Hautier, G. Chemist versus machine: Traditional knowledge versus machine learning techniques. *Trends in Chemistry* **2021**, *3*, 86–95.

(30) Friederich, P.; Krenn, M.; Tamblyn, I.; Aspuru-Guzik, A. Scientific intuition inspired by machine learning-generated hypotheses. *Machine Learning: Science and Technology* **2021**, *2*, 025027.

(31) Goh, G. B.; Hodas, N. O.; Siegel, C.; Vishnu, A. Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties. *arXiv preprint arXiv:1712.02034* **2017**,

(32) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H., et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry* **2019**, *63*, 8749–8760.

(33) Wahab, A.; Pfuderer, L.; Paenurk, E.; Gershoni-Poranne, R. The COMPAS Project: A Computational Database of Polycyclic Aromatic Systems. Phase 1: cata-condensed Polybenzenoid Hydrocarbons. **2022**,

(34) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics* **2010**, *132*, 154104.

(35) Johnson, E. R.; Becke, A. D. A post-Hartree–Fock model of intermolecular interactions. *The Journal of chemical physics* **2005**, *123*, 024101.

(36) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry* **2011**, *32*, 1456–1465.

(37) Fuchs, F.; Worrall, D.; Fischer, V.; Welling, M. SE(3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems* **2020**, *33*, 1970–1981.

(38) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903* **2017**,

(39) Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589, DOI: 10.1038/s41586-021-03819-2.

(40) Pope, P. E.; Kolouri, S.; Rostami, M.; Martin, C. E.; Hoffmann, H. Explainability methods for graph convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019; pp 10772–10781.

(41) Wang, Z.; Yang, J. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. Workshops at the thirty-second AAAI conference on artificial intelligence. 2018.

(42) Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision. 2017; pp 618–626.

(43) Gershoni-Poranne, R. Piecing it together: An additivity scheme for aromaticity using NICS-XY scans. *Chemistry–A European Journal* **2018**, *24*, 4165–4172.

(44) Finkelstein, P.; Gershoni-Poranne, R. An additivity scheme for aromaticity: the heteroatom case. *ChemPhysChem* **2019**, *20*, 1508–1520.

(45) Paenurk, E.; Feusi, S.; Gershoni-Poranne, R. Predicting bond-currents in polybenzenoid hydrocarbons with an additivity scheme. *The Journal of Chemical Physics* **2021**, *154*, 024110.

(46) Markert, G.; Paenurk, E.; Gershoni-Poranne, R. Prediction of Spin Density, Baird-Antiaromaticity, and Singlet–Triplet Energy Gap in Triplet-State Polybenzenoid Systems from Simple Structural Motifs. *Chemistry–A European Journal* **2021**, *27*, 6923–6935.

(47) Fite, S.; Wahab, A.; Paenurk, E.; Gross, Z.; Gershoni-Poranne, R. Revealing Structure-Property Relationships in Polybenzenoid Hydrocarbons with Interpretable Machine-Learning. *ChemRxiv* **2022**, DOI: `10.26434/chemrxiv-2022-6dd6n`.

(48) Balaban, A. Chemical graphs—VII: Proposed nomenclature of branched catacondensed benzenoid polycyclic hydrocarbons. *Tetrahedron* **1969**, *25*, 2949–2956.

(49) Balaban, A. T. Challenging problems involving benzenoid polycyclics and related systems. *Pure and Applied Chemistry* **1982**, *54*, 1075–1096.

(50) Gutman, I. Topological properties of benzenoid systems. *Theoretica chimica acta* **1977**, *45*, 309–315.

(51) Aihara, J.-i. Graph theory of aromatic stabilization. *Bulletin of the Chemical Society of Japan* **2016**, *89*, 1425–1454.

(52) El-Basil, S. Applications of caterpillar trees in chemistry and physics. *Journal of mathematical chemistry* **1987**, *1*, 153–174.

(53) Chithrananda, S.; Grand, G.; Ramsundar, B. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885* **2020**,

(54) Yang, Y.; Davidson, E. R.; Yang, W. Nature of ground and electronic excited states of higher acenes. *Proceedings of the National Academy of Sciences* **2016**, *113*, E5098–E5107.

(55) Baldo, M.; Piccitto, G.; Pucci, R.; Tomasello, P. Semiconductor-like structure of infinite linear polyacene. *Physics Letters A* **1983**, *95*, 201–203.

(56) Korytár, R.; Xenioti, D.; Schmitteckert, P.; Alouani, M.; Evers, F. Signature of the Dirac cone in the properties of linear oligoacenes. *Nature Communications* **2014**, *5*, 1–7.

(57) Malloci, G.; Cappellini, G.; Mulas, G.; Mattoni, A. Electronic and optical properties of families of polycyclic aromatic hydrocarbons: A systematic (time-dependent) density functional theory study. *Chemical Physics* **2011**, *384*, 19–27.

(58) Pino-Rios, R.; Baez-Grez, R.; Sola, M. Acenes and phenacenes in their lowest-lying triplet states. Does kinked remain more stable than straight? *Physical Chemistry Chemical Physics* **2021**, *23*, 13574–13582.

(59) Zwadlo, M.; Hagara, J.; Duva, G.; Hagenlocher, J.; Gerlach, A.; Zaluzhnyy, I.;

Hodas, M.; Hinderhofer, A.; Siffalovic, P.; Schreiber, F. Structure of Thin Films of [6] and [7] Phenacene and Impact of Potassium Deposition. *Advanced Optical Materials* **2021**, *9*, 2002193.

(60) Okamoto, H.; Hamao, S.; Eguchi, R.; Goto, H.; Takabayashi, Y.; Yen, P. Y.-H.; Liang, L. U.; Chou, C.-W.; Hoffmann, G.; Gohda, S., et al. Synthesis of the extended phenacene molecules,[10] phenacene and [11] phenacene, and their performance in a field-effect transistor. *Scientific reports* **2019**, *9*, 1–11.