Can Deep Learning Search for Exceptional Chiroptical Properties? The Halogenated [6]Helicene Case

Rafael G. Uceda,[†] Alfonso Gijón,[‡*] Sandra Míguez-Lago,[†] Carlos M. Cruz,[†] Víctor Blanco,[†] Fátima Fernández-Álvarez,[†] Luis Álvarez de Cienfuegos,[†‖] Miguel Molina-Solana,[‡] Juan Gómez-Romero,[‡] Delia Miguel[⸸], Antonio J. Mota,[§*] and Juan M. Cuerva[†*]

† Department of Organic Chemistry, Faculty of Sciences, Unidad de Excelencia de Química (UEQ), University of Granada. Avda. Fuente Nueva s/n, 18071 Granada, Spain.

‡ Department of Computer Science and Artificial Intelligence, School of Technology and Telecommunications Engineering, University of Granada. Calle Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain.

⸸ Department of Physical Chemistry, Faculty of Pharmacy, UEQ, University of Granada. Avda. Profesor Clavera s/n, 18071 Granada, Spain.

§ Department of Inorganic Chemistry, Faculty of Sciences, UEQ, University of Granada. Avda. Fuente Nueva s/n, 18071 Granada, Spain.

‖ Instituto de Investigación Biosanitaria. Avda. Madrid, 15, 18016 Granada, Spain.

Email: alfonso.gijon@ugr.es , mota@ugr.es , jmcuerva@ugr.es

ABSTRACT: In this work we predict, among more than a billion possibilities, the best candidates of halogenated [6]helicenes in order to obtain excellent chiroptical properties in terms of the rotatory strength (R). We have used DFT calculations to randomly create derivatives from 1 to 16 halogen atoms, that were then used as a data set to train different deep neural network models. It is worth noting that the simplest model affords a parametrization that allows to easily predict the value of R for any hexahalogenated [6]helicene. The correlation between calculated and predicted data increases together with the complexity of the model. The results show that some positions and halogens are preferred to increase the R value. In this sense, we have also synthesized the derivatives with the higher predicted R, obtaining excellent correlation among the values obtained experimentally, by DFT-calculations and machine learning predictions.

INTRODUCTION

[n]Helicenes are prototypical helical structures composed by *n ortho*-fused phenyl rings.[1–3] They also present high racemization barriers (when n>5) and show interesting chiroptical properties.[4,5] Remarkably, such properties can be now predicted with high confidence using DFT calculations with relatively low computational cost for the smallest members of the family.[6–10] However, although the chiroptical properties are codified in the intrinsic physics of the molecule, it is not easy to extract any structure-property relationship (apart from absolute configuration) from such kind of calculations. Even in the case that any correlation would exist, a huge volume of examples or data should be necessary for its understanding. The situation becomes more complex if we consider substitutions in the [n]helicene core. As an example, if we consider multiple halogen substitution in any of the sixteen positions in [6]helicene (Figure 1a), the challenge is intimidating. For instance, if we consider the mono-halogenation case only 32 derivatives can exist, according to Eq.1:

$$N_k = \frac{1}{2}\binom{16}{k}4^k \quad \textbf{Eq. (1)}$$

$$N_1 = 32 \quad N_2 = 960 \quad N_3 = 17920$$

$$N_4 \approx 2.3x10^5 \quad N_5 \approx 2.2x10^6 \quad N_6 \approx 1.6x10^7$$

This expression gives the number of different compounds that can be obtained with k substituents considering also the reflexion symmetry dividing the general expression by 2. Di-halogenated [6]helicenes are a larger family composed of 960 members. In that case, the DFT calculations are costly but continue to be affordable in a reasonable period of time. Nevertheless, the in-depth analysis of the resulting data begins to be daunting. For tetra-halogenated [6]helicenes the numbers go up to almost 2.33×10^5 possibilities, and for hexadeca-halogenated [6]helicenes the variation gives an astonishing number of 2.15×10^9 different compounds (Eq. 1). In those cases, neither theoretical calculations nor the analysis becomes viable. Globally, considering from mono- to hexadeca-halogenated [6]helicenes, 7.63×10^{10} structures should be evaluated.

The problem is even more complex considering that the chiroptical properties are diverse and the corresponding magnitude of the response can be defined in many ways. In this work we have focused our attention on electronic circular dichroism (ECD), one of the prototypical chiroptical techniques employed.[6] In this case, rotational strength (R_{0j}) is a good indicator,[11] which is associated with each ground to excited state transition (*0* to *j*). This scalar value represents the intensity of the chiroptical response, being the shape of the ECD spectra characterized by the most intense ones. A complete set of R_{0j} values can be extracted from theoretical calculations (Figure 1c).

Considering the above-mentioned framework, the answer to the question, "*which is the maximum value for a rotatory strength in a (poly)halogenated [6]helicene?*", is beyond the human interpretative capabilities using standard approaches. As an alternative, machine learning techniques have succeeded in many cases to extract hidden patterns and develop predictions for complex problems only from data points of the observed phenomenon.[12] Specifically, neural networks –the computational model behind deep learning– have shown efficiency in Chemistry[13–15] as to classify organic reaction mechanisms,[16–23] to accelerate DFT calculations,[24,25] and to predict molecular properties[26–33] and antibacterial activities.[34] Indeed, despite machine learning methodologies have been applied to achiral nanomaterials, there is no examples including chirality in these structures.[35] It is also worth noting that the approach in the search of new materials with improved properties must meet two important requirements: i) to be able to extrapolate values for the extreme cases, where exceptional materials are, and ii) to be synthetically viable.[36] Within this context, we wondered if the starting point question can be accomplished using deep learning approaches, searching for exceptional responses. To that end,

predicting chiroptical properties in [6]helicenes in a faster way than typical DFT simulations is required to estimate the maximum R_{0j} values (R_{max}) of the vast number of interrogated systems (Figure 1c).

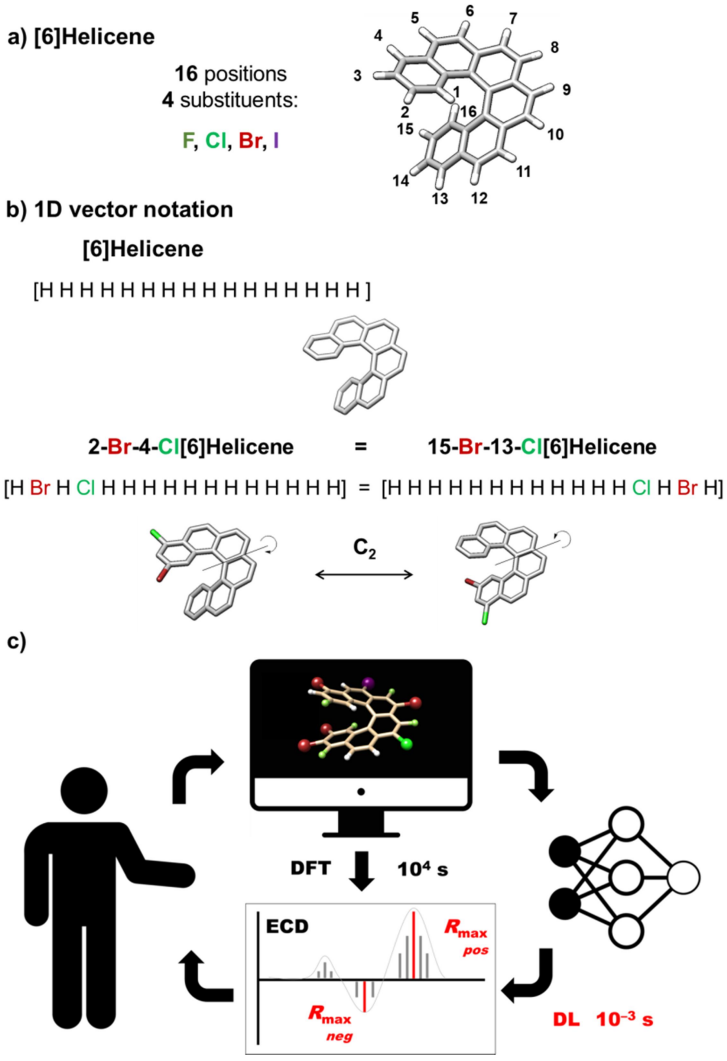

Figure 1. a) [6]helicene structure. Geometry optimized at the M06/TZVP level of calculation (PCM dichloromethane). b) 1D vector notation examples employed for differently halogenated [6]helicenes. c) Interplay between DFT calculations and Deep Learning (DL) based training and prediction of chiroptical properties.

Beyond scientific curiosity and gaining fundamental knowledge, this question is relevant owing to [n]helicenes have been proposed as promising candidates in terms of chiroptical responses and their consequent applications into devices.[37–40] Therefore, the first question to be addressed is if there is a limit in the R_{max} and, in case, which is its value. Thus, knowing the limit, we can either be impelled to explore those boundaries or, on the other hand, to turn to other chiral entities. Here we have designed and trained a neural network which is able to estimate R_{max} of halogenated [6]helicenes with a minimal computational cost. In particular, the neural network produced results regarding the structure-property relationships for systems ranging from mono- to hexadeca-halogenated [6]helicenes (Figure 1a). Such results were then compared with the prediction of two simpler and physically interpretable models. R_{max} values can be deconvoluted as a linear combination of coefficients depending on the relative position of 1 to 2 halogens on the structure. While the approach is less precise than DFT calculations, the developed models allow a rationalization of the results by two main reasons: i) the neural network is able to

produce a big amount of data that additionally fits with those obtained through calculations and ii) simpler models give interpretable physical information of the behaviour of the system. The former allows the creation of a full database in which the selection of the better substitution pattern is straightforward even for blind interpretative models. Considering the latter, we have interestingly found that better values do not correspond to randomly distributed structures but indicate that some positions and atoms are preferred. Under this circumstance, a kind of parameterization can be made, assigning different weights to each halogen for each position. With significant differences, this situation resembles the concept of free energy linear relationships developed from seminal studies by Hammet.[41] That is, primary positions of the halogens establish a relative weight using hydrogen atom substitution as a reference. Vicinal substitutions are then considered as secondary corrections.

We have organized the discussion presenting sequential training of models (up to hexahalogenated [6]helicenes) for which parameterizations seem to be robust, allowing a confident prediction about R_{max}. For highly substituted systems, the model has been statistically checked. Optimal weighting factors determine that the best response in terms of R_{max} can be found for some tetra-substituted [6]helicenes among thousands of millions of potential structures. Our full analysis suggests that 2,3,14,15-tetrabromo [6]helicene **1** is the best candidate to achieve the highest rotatory strength. It is worth noting that R_{max} value for this compound has been predicted, being outside of the training dataset. Besides, the product has been synthetized and its chiroptical properties experimentally determined, being in excellent agreement with those theoretically predicted

RESULTS AND DISCUSSION

Dataset and model training. The success of deep learning approaches relies on the capabilities of neural networks to approximate functions from a number of sample points. Thus, we decided to use data samples as pairs <X, Y>, where X is the representation of the molecule and Y is the property we calculate by DFT (in this case R_{max}) for a given X. Since all the input molecules share the same carbo[6]helicene skeleton, we decided to represent the helicene as a 1D vector constituted by 16 elements representing the substitution of the molecule (Figure 1b). To this regard, the combination of a simple vector containing the hydrogen position to be exchanged (1 to 16) and the nature of the saturating atom (0 = H, 1 = F, 2 = Cl, 3 = Br, and 4 = I) is enough for the complete description of the structure. Furthermore, all models are built to respect the reflexion symmetry of the molecule, being the position n equivalent to the position 17-n, with n=1,...,16 (Figure 1b).

Despite the simplicity of the representation, all hidden contributions of any geometrical distortions (bond lengthening, resonance/inductive effects, etc) are codified in the calculated rotatory strengths. For the training, examples dealing only with the *P* configuration in the helix were selected. By symmetry, conclusions derived from the study are applicable to the opposite *M* helical configuration, just by changing the sign of the R_{0j} values. At this point it is worth noting that R_{0j} values can be positive and negative for each configuration. We have analysed both situations (positive and negative R_{max}) independently. The model is then trained to fit Y for X, and more interestingly, to make a meaningful estimation of Y for a given X. To this end, it is desirable to train the neural network with the most diverse and accurate available data. Hereof, theoretical calculations of a randomly selected family of [6]helicenes provided the dataset, including molecules with low and high R_{max} values.[42,43] It should be also noted that all the results are indispensable in every machine learning protocol.[44,45]

Although neural networks are suitable regressors to capture complex non-linear relations between input molecular representations and the target magnitude, they are black-box models and suffer a lack of interpretability. Therefore, we decided to accompany this approach with two simpler alternatives, so-called 1- and 2-body models, where the interpretation of the underlying physics is more easily achieved. In the simplest 1-body model, the maximum rotatory strength of a molecule can be obtained directly by adding the 16 contributions of each atomic position. The contribution of each substituent depends on its relative position in the [6]helicene and the

chemical species at that position. Accordingly, the model has 5×8=40 free parameters, coming from the number of different atoms (hydrogen plus four halogens) multiplied by the number of non-equivalent positions, taking into account the system symmetry (Figure 2a, Eq. 2 and Figure 3a). Furthermore, the 2-body model considers the previous 1-body term plus an additional contribution accounting for adjacent interactions between first neighbours (Figure 2b, Eq. 3). While the 1-body term was defined by a 1D vector of size 5 for each position, the 2-body term is defined by a 5×5 matrix (Eq. 3). This increases the number of free parameters of the 2-body model up to 240. From such on-site and neighbouring parameters general conclusions about the physics of the system can be extracted. Following this terminology, the neural network model could be considered a many-body model, hereafter called N-body model (Figure 2c, Eq. 4), where contributions of single positions are mixed by a multilayer perceptron (MLP) to obtain the final output (Figure 3b). Obviously, the accuracy of the N-body neural network model, with 9257 parameters, is better than that of the 1- and 2-body models. However, the possibility of easily parameterize the response together with the good correlations also obtained by simpler models, make them very attractive and powerful strategies. All the models have an adding constant (R_0), which is set to the mean rotatory strength of the whole dataset.

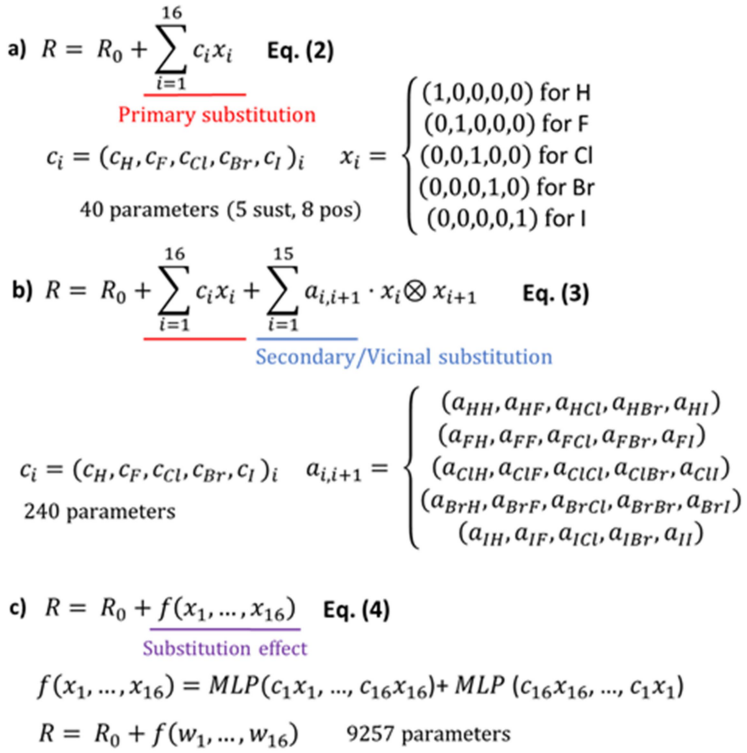

Figure 2. a) 1-body model equation with the definition of the main parameters. Note that the 1-body contributions from positions 9 to 16 are equal to those of positions 8 to 1, to respect the symmetry. b) 2-body model equation. c) N-body model equation for obtaining rotational strength (R) values, where f is a neural network-based function constructed to impose the symmetry

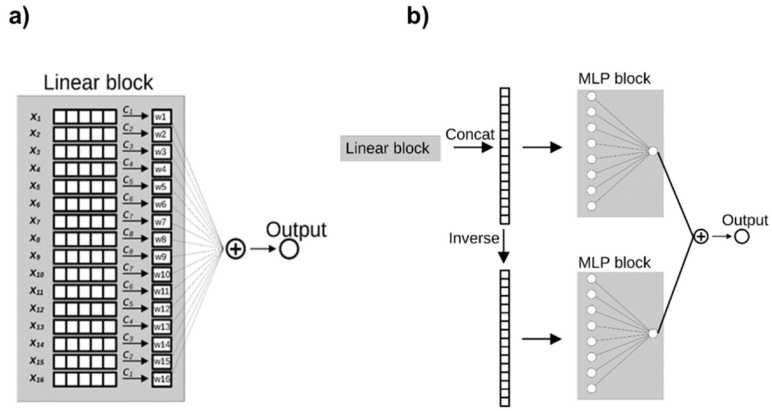

Figure 3. Process diagram of a) 1-body and b) 2-body model. After applying the linear block, the same MLP (multilayer perceptron) is applied to the original and inverted features vector to approximate a function respecting the symmetry of the system.

Our dataset was composed of 32 mono-halogenated [6]helicenes and randomly selected families of di (150), tri (200), tetra (200), penta (200) and hexa-halogenated [6]helicenes (400) examples, constituting 1182 examples in total. For each molecule, the ECD spectra as a set of R_{0j} values versus absorption wavelengths were calculated using DFT methods as implemented in Gaussian 09 (see SI for details).[46] Most of calculated positive R_{max} values are around 400-700 10^{-40} cgs units with minor subsets with examples presenting low (100-300 10^{-40} cgs units) and high (800-900 10^{-40} cgs units) ones (Figure 4d). Negative R_{max} values present a mean absolute value of 250 10^{-40} cgs units (R_0) with a very minor subset beyond 600 10^{-40} cgs units. Owing to this different behaviour we analysed the two scenarios independently.

The dataset was then split into 80% of the molecules for training and 20% for testing. The robustness of the predictive models was tested by means of a 10-times repeated random sub-sampling validation. All the models were implemented and trained using TensorFlow.[47] The employed data and code are available (see SI). The prediction of $R_{\max}$ is treated as a regression task, and we employ evaluation metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage (MAPE), Mean Squared Error (MSE) and coefficient of determination R^2, to assess the performance of the proposed methods. Among these metrics, MAE represents the mean of the absolute differences between the predicted values and the actual values. It assigns equal weight to all error values, thereby mitigating the impact of outliers. MSE calculates the mean of the squared differences between the predicted values and the actual values, amplifying the influence of larger errors, thus more susceptible to outliers.

Case 1. Positive R_{max} values for tetrahalogenated *P*-[6]helicenes. The number of potential tetrahalogenated [6]helicenes is 232960. Despite that individual DFT calculations are easily affordable with actual computers, such a volume of required individual calculations is too expensive to be practical. Therefore, we used this case to evaluate the feasibility of the DL approach to provide reasonable estimations for those compounds. As mentioned before, a set of 582 individual DFT calculations (the [6]helicenes up to four substituents) was used as training and test dataset. Figure 4 shows the correlation results using the three models, expanding from very different R_{max} values. It can be observed that the correlation improves with the number of bodies together with a decrease of data dispersion. In this sense, the N-body model also presents a reasonable Mean Absolute Error (MAE) of 17 and 30 10^{-40} cgs for the train and test datasets, respectively. The MAE remains similar independently of the substitution degree which also evidences the reliability of the model for the considered substitution (Figure S7). With the confidence that the model is suitable at this level of substitution, we then estimated the rest of the members of the tetrahalogenated family (Figure 4d). The prediction yielded an R_{max} distribution very similar to that obtained with pure DFT dataset (Figure 4d, orange), spanning

mainly from 200 to 800 10^{-40} cgs units. On the other hand, the DFT-calculated R_{max} value for the parent [6]helicene is 698 10^{-40} cgs and the predicted one 696 10^{-40} cgs units, which suggests that the N-body model is getting the underlying physics of the system and that in general the substitution is detrimental, minimizing the R_{max} values.

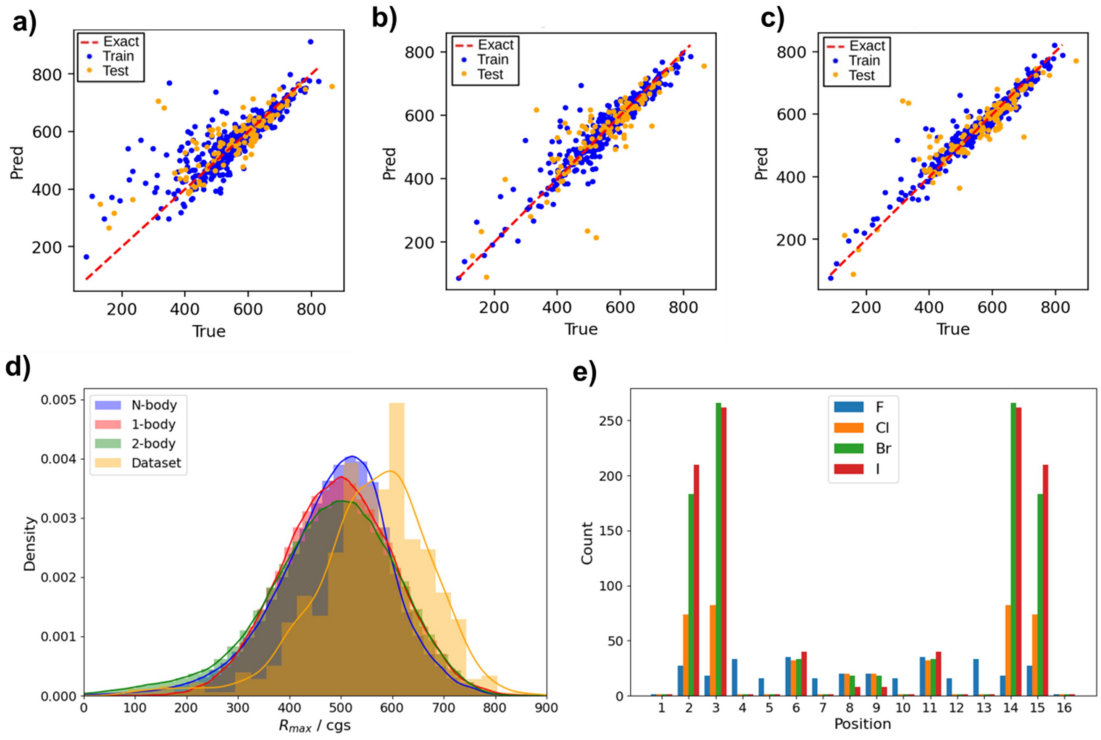

Figure 4. Correlation between model-predicted (y axis) vs DFT-calculated (x axis) rotational strength (R, 10^{-40} cgs units) values for halo[6]helicenes up to 4 halogen atoms employing a) 1-body, b) 2-body, c) N-body models. d) Distributions of positive R_{max} obtained from DFT, 1-, 2-, and N-body models for tetrahalogenated[6]helicenes. e) Location of halogens in molecules with high R_{max} from N-body model for tetrahalogenated[6]helicenes

Table 1. Statistic parameters for the three 1-, 2-, and N-body models within the tetrahalogenated[6]helicene.

	1-body		**2-body**		***N*-body**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
MAE[a]	37±2	39±6	22±1	37±4	17±1	30±4
RMSE[b]	63±3	62±11	39±2	59±8	31±3	47±9
MAPE[c]	9.4± 0.5	9±2	4.8± 0.3	8±1	3.8± 0.3	7±1
R^2 score	0.69±0.03	0.70±0.10	0.88±0.01	0.70±0.10	0.92±0.01	0.80±0.1

[a] Mean Absolute Error, [b] Root Mean Squared Error, [c] Mean Absolute Percentage. All values are in 10^{-40} cgs units.

To achieve a better understanding of the origin of the R_{max} values, we analysed derivatives with R_{max}>800 10^{-40} cgs units, finding a substantial preference (between 650 and 750 possibilities) when bromine and iodine atoms (Figure 4e, green and red bars respectively) are placed in the

2,3,14 and 15 positions. It clearly shows that some positions and specific halogens are consistently favoured. This intriguing preference could be rationalized simplifying the model and invoking a kind of parameterization. That is, the R_{max} value could be obtained by simple addition of individual contributions of substituents to an initial R_0 value. Not surprisingly, the use of the 1-body approach and the original N-body simulation look similar (Figure 4d, blue and red), pointing out that the position, the halogen, and the R_{max} value are in fact closely related in an apparently systematic way. Thus, the extracted parameterization data (See Table 2) are relevant to rationalize the previous findings obtained at the level of the N-body model. Hydrogen substitution by a halogen is in general disfavoring the maximum R_{max} value except in the 2,3,14 and 15 positions when a bromine or iodine atom is placed. Nevertheless, the spatially close 1,4,13 and 16 positions are intriguingly highly disfavoring. In any case, these 1-body simulations must be only used to look for tendencies owing to the predicted R_{max} values are systematically higher than DFT ones. In fact, the 1-body model showed us 15 privileged candidates for high R_{max} values (878 – 914 10^{-40} cgs units) (Table S12), all of them possessing bromine and iodine atoms in the 2,3,14 and 15 positions. It is worth noting that the model is able to extrapolate values beyond the ones used in the training dataset. This is in fact critical for our purpose. Their DFT R_{max} values were then calculated. Although resulting values were usually lower than the predicted ones (Table S12), it is especially relevant the case of 2,3,14,15-tetrabromo [6]helicene **1** case (Figure 5). For it DFT calculations give an astonishing R_{max} value for a small molecule of 942 10^{-40} cgs units.

Differences between 1- and N-body simulations come, most likely, from the inability of the simple model to describe secondary interactions between bulky halogens placed in contiguous positions. If the assumption is true, an improved model including such contributions should present a better correlation. Consequently, a 2-body model (Figure 4b) showed a better profile than the 1-body model (Figure 4a). It is worth noting that these new models have been built to rationalize the results, remaining the N-body one of the most reliable in terms of predictions. Nevertheless, the fact that they perform in acceptable way is important, showing that machine learning treatment can help suggesting a physical interpretation. The primary parameterization table (Table S3) remains similar to the previous one with similar contributions of positions and halogens. Secondary contributions correct the initial values using a new 5x5 matrix for each position (Tables S4-S11). A close inspection of the secondary corrections shows that very few combinations can increase R_{max}, being the global value controlled by the primary parameterization. Although the parameterization is not perfect, general trends can be observed. Hydrogen atoms are the most efficient substituents for high R_{max} values except for 2,3,14 and 15 positions in which bromine and iodine atoms are the best ones. On the other hand, no synergies seem to appear by the presence of vicinal halogens. On the contrary, the observed trend points at an increase in the number of halogens is always detrimental to achieve high R values.

Trying to rationalize the results from a photophysical perspective, we analysed in more detail some prototypical examples. Symmetric structures with two halogens (1,16-, 2,15-, 3,14-, … Figure S8) were then calculated by DFT to find any potential structure-property relationship. R_{0j} is described as the scalar product of the electric (μ_{0j}) and magnetic (m_{0j}) transition dipole moments for a certain transition, $R_{ij} = \mu_{0j}m_{0j} = |\mu_{0j}|\cdot|m_{0j}|\cdot\cos\theta$. It is maximum when the electric (μ_{0j}) and magnetic (m_{0j}) transition dipole moments are maximized, and the mutual orientation is parallel (or antiparallel). The module of vector m_{0j} is usually several orders of magnitude smaller than the module of μ_{0j}. Being a component of the equation, very low $|m_{0j}|$ values result in very poor chiroptical response. This value is maximized when the transition involves the extended helicene π-orbitals,[48] while localized π-orbitals usually give low $|m_{0j}|$ values. On the other hand, the angle θ is also critical. Small changes in the substitution pattern can result in very different cosine values. The corresponding parameters for dibrominated [6]helicenes and the parent [6]helicene, for comparison, are presented in Table 3.

As it can be seen, the best R_{max} value appears as a result of an optimization of $|m_{0j}|$ and the corresponding angle θ. Employing the Multiwfn software package[49] we were able to visualize the transition magnetic dipole moment density graphs. It can be observed that the magnetic

transition extends to the bromine atoms for some favoured (2,15 and 3,14) positions, thus creating a better electron circulation during the transition (Figure 5b). This phenomenon was in fact evidenced by the models in their predictions. The combination of such $|m_{0j}|$ increase with a slightly better angle is the base of the improved R_{max} values. A similar analysis can be done for the iodine substitution (See Table S16). On the other hand, in compounds with fluorine and chlorine substitution $|m|$ is not improved and at the same time the angle is also worse than the one in parent helicene (Tables S13-S14). Consequently, such substitution is detrimental for exceptional R_{max} values.

Table 2. Parametrized coefficients for each atom in each position[a] obtained with the 1-body model[b]

Position	H	F	Cl	Br	I
1	35.66	-90.05	-81.33	-55.88	-7.11
2	-21.49	-52.21	6.91	42.68	41.32
3	-23.53	-25.70	4.88	25.39	26.16
4	28.73	6.98	-21.90	-26.65	-102.83
5	32.47	2.77	-49.25	-58.93	-96.61
6	16.09	-16.64	-27.80	-24.26	-62.44
7	14.41	-7.56	-29.08	-46.16	-73.10
8	6.89	-14.32	-23.95	-26.53	-35.49

[a] *n* position is equivalent to *17-n* position (e.g. position 3 and position 14).

[b] Green shading corresponds to the most favoring substitution and red shading to the worst one. $R_0 = 508.89 \times 10^{-40}$ cgs units.

Table 3. Electric and magnetic transition dipole moments, angle between them (θ), angle cosine, and rotational strength DFT-calculated values for selected dibromo[6]helicenes.[a]

Bromine Position	10^{20} |μ| / esu cm	10^{20} |m| / erg G^{-1}	θ/ °	cos θ	$10^{40} R_{max}$ /cgs units
1,16	447	3.59	**68**	**0.37**	606
2,15	473	**3.98**	**62**	**0.47**	**866**
3,14	**564**	**4.42**	**70**	**0.34**	**847**
4,13	**563**	**3.91**	75	0.26	541
5,12	**652**	**3.74**	78	0.21	506
6,11	515	3.20	**68**	**0.37**	611
7,10	**596**	2.90	**70**	**0.34**	569
8,9	542	2.49	**64**	**0.44**	569
[6]helicene	556	3.65	70	0.34	698

[a] Parameters with equal/better values than parent [6]helicene are shown in bold.

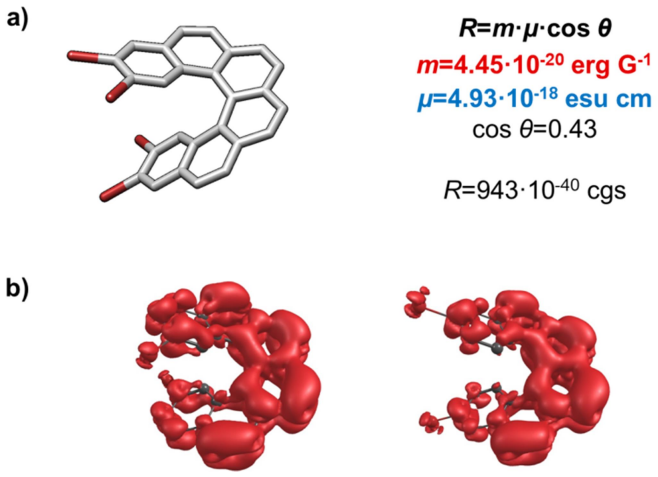

Figure 5. a) Structure of 2,3,14,15-tetrabromo[6]helicene **1**. b) Transition magnetic dipole moment density map for the best substitutions, corresponding to (2,15-) and (3,14-)dibromo [6]helicene.

Case 2. Positive R_{max} values for hexahalogenated (*P*)-[6]helicenes. At this point we tested a more complex case, the prediction of the highest positive R_{max} values for the hexahalogenated [6]helicene family. It should be noted that in the previous case the models have been developed to predict the response of around 200,000 molecules, while including helicenes with 5 and 6 halogens means increasing the number of possible molecules to almost 19 million. The critical point is how to train the new model at the same level than for the Case 1 owing to the notable increase of the sample size. Moreover, the model should at the same time successfully predict the response of the previous less substituted structures. If such kind of concordance is possible, similar physical processes must be at the core of the prediction. In this case we used up to 1,182 randomly selected molecules containing one to six halogens as training examples. Such number was the result of a steady increase of examples until the *N*-body correlation remained statistically stable with a MAE of 27 10^{-40} cgs for train and 36 10^{-40} cgs for test (see Figure S3 and Table S17). We observed that with the new model the predicted R_{max} values for halogenated[6]helicenes (Figure 6c), although reasonable, were slightly smaller than in Case 1 owing to training values for hexahalogenated [6]helicenes are, in general, smaller. Such full data set presents a R_{mean} value of 508 10^{-40} cgs units (Figure 6d) in which very few cases with R_{max} beyond 800 10^{-40} cgs units appear.

We were then curious about unravelling if parameterization process would continue being valid in Case 2. 1- (Figure 6a) and 2-body (Figure 6b) models were created, both presenting a reasonable correlation. This fact suggests that a kind of parameterization is again present in the physics of the system. Coefficients of the 1-body and 2-body models for both cases present common main features (SI, Tables S2-S11 for 1-body model and Tables S18-S27 for 2-body model). Among them, it is worth highlighting that no synergies were detected in the secondary parameters, showing that the higher the substitution, the lower the R_{max} value. Furthermore, the 2-body model presented a better correlation than in case 1, which is reasonable since the database now includes more adjacently substituted helicenes. That is, examples where halogens are more likely to occupy adjacent positions.

With the models in hand, the R_{max} value distribution for the total 1.64×10^7 hexahalogenated[6]helicenes structures were predicted. Distributions obtained for the three models are quite similar, being the *N* body one slightly narrower than the others, and also properly fitting the DFT R_{max} value distribution (Figure 6d). Again, the substitution increase seems to be detrimental for high R_{max} values. This observed trend is in qualitative agreement with the underlying physics extracted from case 1. The introduction of additional substituents is

disfavouring the electronic circulation of the π conjugated system during the transition, minimizing $|m_{0j}|$ values. If that assumption is correct, higher substitution numbers from hepta- to hexadecahalogenated [6]helicenes would always yield poor chiroptical responses. The model suggested good hexahalogenated [6]helicene candidates (ca. 20) with high values of R_{max} (>910 10^{-40} cgs units) (Table S28). All have bromine/iodine atoms in positions 2,3(14,15), supporting the conclusions from Case 1, and also fluorine/chlorine atoms in positions 8-9 (Figure 6e). The simplified models show, as in case 1, that vicinal interactions are mainly detrimental to R_{max} value. Therefore, it is reasonable that the fifth and sixth halogen atoms, being fluorine and chlorine (the smallest ones), lie on the 8,9 positions (the furthest ones) (Figure 7). All the suggested candidates for R_{max} values were then evaluated by DFT (Table S28). Nevertheless, neither resulted in R_{max} values higher than that obtained for privileged compound **1**.

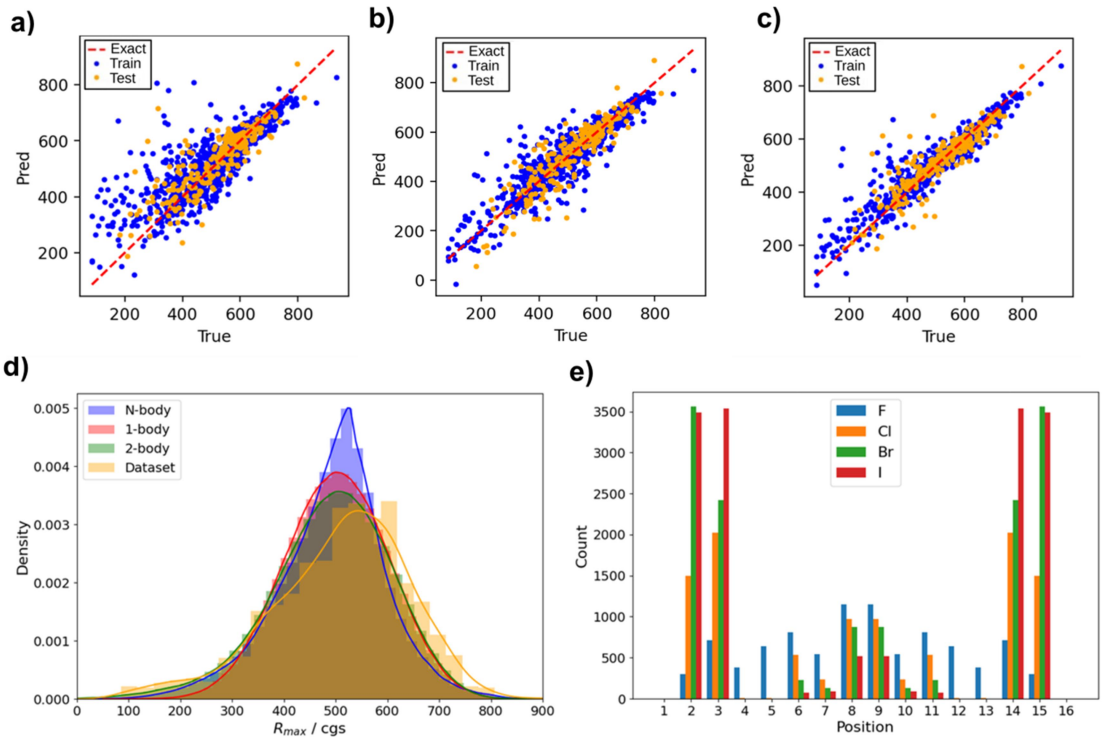

Figure 6. Correlation between model-predicted (y axis) vs DFT-calculated (x axis) rotational strength (R) values for halo[6]helicenes up to 6 halogen atoms employing a) 1-body, b) 2-body, c) N-body models. d) Distributions of positive R_{max} obtained from DFT, 1-, 2-, and N-body models for hexahalogenated[6]helicenes. e) Location of halogens in molecules with high $R_{\max}$ from N-body model for hexahalogenated[6]helicenes.

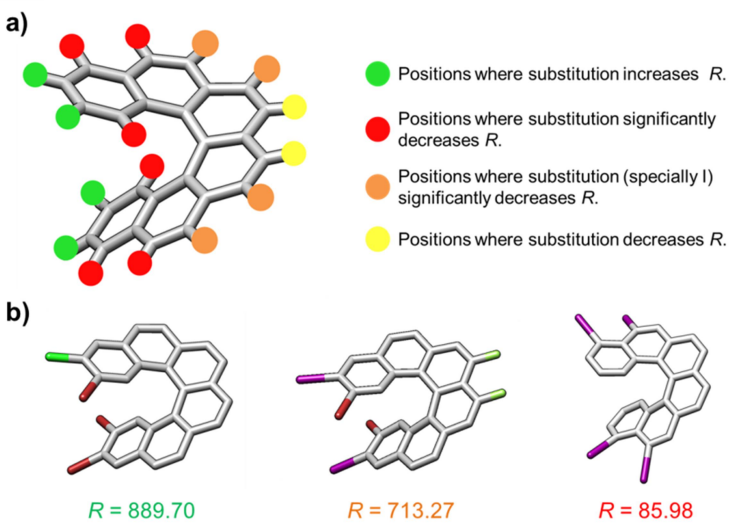

Figure 7. a) Role of position and nature of the substitution in [6]helicenes and b) Example of derivatives with high, normal and low R_{max} values. Color code: iodine, violet; bromine, brown; chlorine, green; fluorine, pale green.

Case 3. Positive R_{max} values from hepta- to hexadecahalogenated (*P*)-[6]helicenes.

The previous cases suggest that an increase in the number of halogens is detrimental for the R_{max} values. To check if the model developed in Case 2 remains valid for higher substitutions we tested a dataset of 1000 randomly selected DFT calculated examples from hepta- to hexadecahalogenated [6]helicenes (100 samples each). They were estimated by the three models and the results summarized Figures S10-S19. Figure 8a shows the DFT R_{max} values and those predicted by the N-body model for the 100 heptahalogen[6]helicenes. Regarding DFT examples it can be clearly seen that the diminishing of R_{max} value is consistent with the increasing number of halogens (Figures S10-S19). *N*-body model seems to remain valid, reporting suitable values for higher halogenations degrees even being trained using only with up to six halogens. 1-body model, despite the simplicity, gives a reasonable agreement, although worse than the N-body one. The 2-body model, which was successful in cases 1 and 2, becomes invalid (e.g. Figure S19). Higher halogenation degree is related with an increase in the halogen close contacts and the model is unable to evaluate the simultaneous interactions with second neighbours and beyond. It is consistent with the training using only six maximum halogens and therefore low halogen contact, being the model biased to very low values. The negative predictions are in fact an artefact from the undertraining dataset. On the other hand, the non-lineal *N*-body approach, although less interpretable, deals with such multiple interactions owing to the nature of the model, considering all the potential contacts at any distance, and thus overpassing the limitations of the 2-body model. The relevant thing here is that the *N*-body and 1-body predictions remain essentially valid for any kind of substitution. For the latter one, it seems to catch the main feature of maximizing $|m_{0j}|$ value without disturbing the electron circulation during the transition, which is essentially dependent of the halogen position.

Additionally, if the corresponding distributions are considered representative and a Gaussian-type curve is assumed, an estimation of the expected values beyond some critical number can be done (Tables S29-S31). For example, the possibility of finding a heptahalogenated[6]helicene with a R_{max} higher than 1000 is 3.79 10^{-06} (0.000379%). Despite being such small probability, the enormous number of candidates means that approximately 300 helicenes with R_{max} higher than 1000 10^{-40} cgs units are statistically predicted. We evaluated R_{max} values for the entire family of heptahalogenated [6]helicenes, composed by 9.4×10^7 molecules. The maximum value among them all using the N-body model was 846 10^{-40} cgs units, in line with previous findings. For higher substitutions 10^6 examples of each family were evaluated to have a better description

of the phenomena. As the number of halogens increases, the R_{mean} value diminishes (Figure 8b), hampering the existence of compounds with exceptional R_{max} values (Figure 8c displays the largest value of R_{max} for each type of substitution). With the above-mentioned distributions (Figure 8b), the expectation for R_{max} values beyond 1000 10^{-40} cgs units is spurious and the Gaussian-type distributions do not reveal any candidate with an R_{max} above 1150 10^{-40} cgs units (Table S32). Basically, almost no halogen substitution beyond four halogen atoms in privileged positions allows a reinforcement of the optimal rotatory strength of the system.

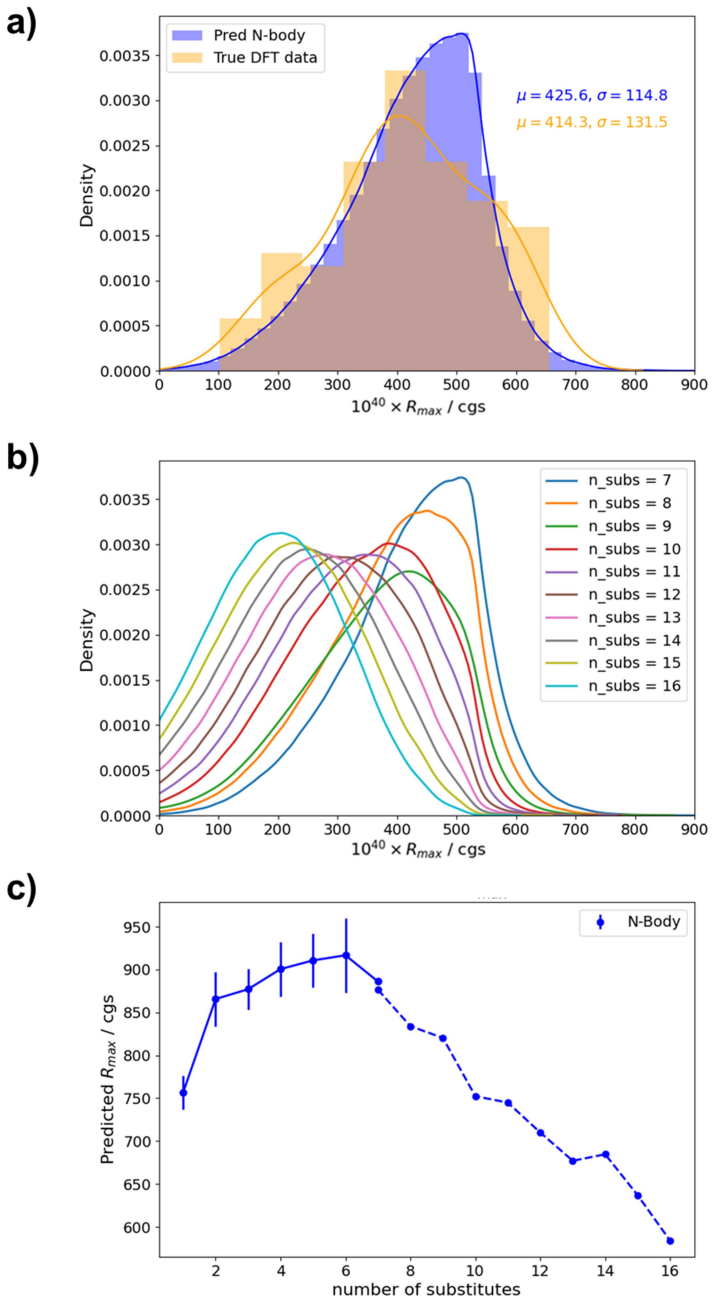

Figure 8. a) Distributions of positive R_{max} obtained from DFT and N-body model for heptahalogenated [6]helicenes including mean (μ) and standard deviation (σ). b) Predicted $R_{\max}$ distributions using N-body model for [6]helicenes ranging from hepta- to hexadeca-halosusbtituted ones. c) Evolution of $R_{\max}$ values for the whole series of substitutions using N-body model. Solid line = prediction employing all possible molecules. Dashed line = prediction employing 10^6 selected candidates of each family.

Case 4. Negative R_{max} values from mono- to hexadecahalogenated (*P*)-[6]helicenes.

Negative R_{max} values were also evaluated using a similar approach. With the knowledge acquired with positive ones we prepared the corresponding N-, 1- and 2-body models using the same DFT-based data set. At first glance it can be seen that the correlation within the N-body model is reasonably good, being the simpler models less reliable (See SI, Figure S21). The second thing is that the R_{max} value span is more limited, being the R_{mean} smaller than in the case of positive ones. The expectation for exceptional negative R_{max} values is then smaller. To be more confident with this assumption, the full family of hexahalogenated *P*-[6]helicenes was simulated (Figure 9b). Very few examples with R_{max} values beyond -850 10^{-40} cgs units were predicted (Table S34). Its evaluation using DFT calculations also supported that negative values are smaller in absolute value than the positive ones. Nevertheless, a geometric analysis of the data points out once again that some positions are preferred (3,14- and 5,12-). In this case, the associated transitions are located along the C2 axis of the [6]helicene core, being the magnetic contribution associated to an electronic circulation following that axis. The analysis of the corresponding involved transition dipole moments suggests that the success of the preferred positions comes from an elongation of the helicene electron density to the substituents and consequently the increase of the involved momenta (Table S35 and Figure S22).

For higher substitutions we followed a similar reasoning than in case 3 and created a 10^6-element simulation for each family. Again, we observed that the R_{max} values strongly diminish with the substitution and no values around 900 10^{-40} cgs units can be obtained (Figure 9c). This case concludes that positive R_{max} values are higher than negative ones in absolute value for the *P*-enantiomer.

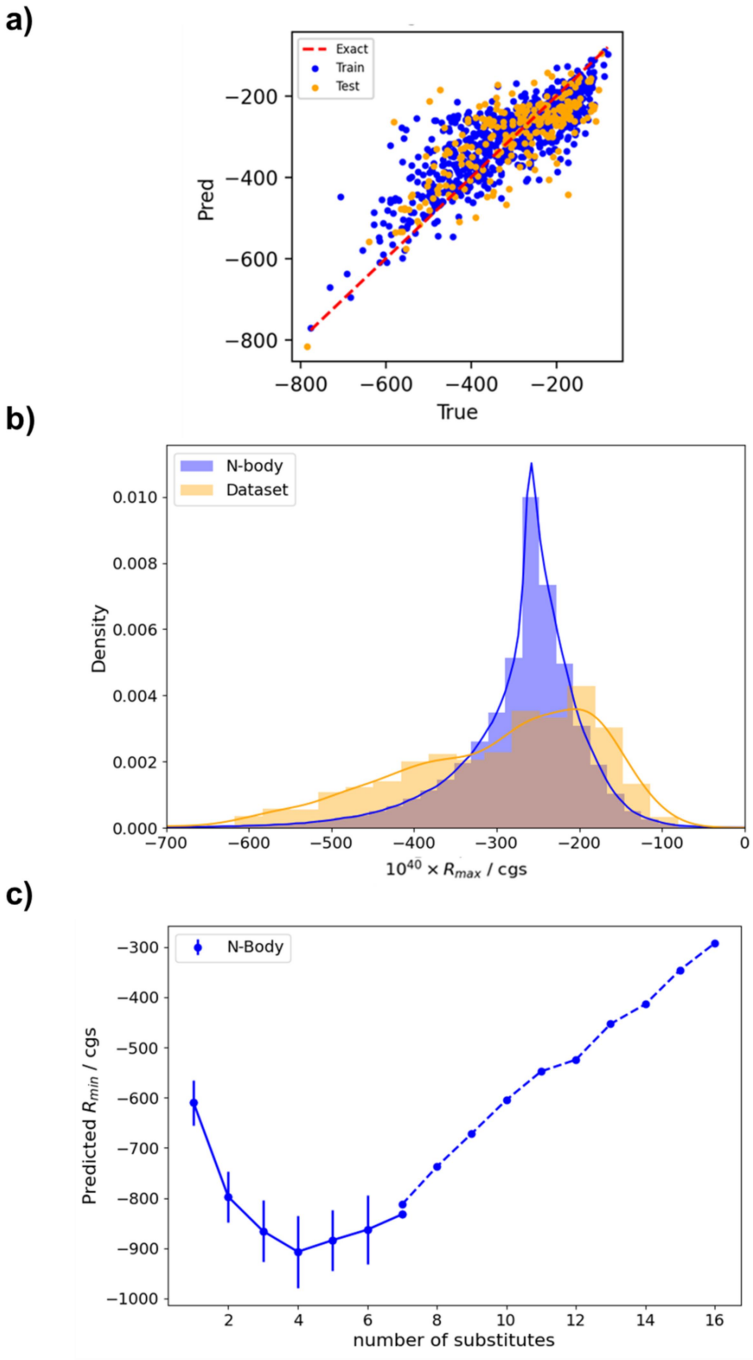

Figure 9. a) Correlation between model-predicted (y axis) vs DFT-calculated (x axis) maximum negative rotational strength (R) values for halo[6]helicenes up to 6 halogen atoms employing N-body model. b) Distribution of negative R_{max} for halo[6]helicenes up to 6 halogen atoms employing N-body model (blue) and DFT (orange). c) Evolution of negative R_{max} values for the whole series of substitutions using N-body model. Solid line = prediction employing all possible molecules. Dashed line = prediction employing selected 10^6 candidates of each family.

Synthesis of selected examples with exceptional chiroptical properties. Although machine learning approaches are considered valuable for exploring the extrema of desired properties, the subsequent validation of the predictions is highly infrequent. In our case, the evaluation of the fit between the predictions of the model and the reality was carried out. The most outstanding candidate in terms of chiroptical properties, compound **1** was synthesized (See SI for details). In addition, other related compound, 2,15-bromo[6]helicene (**2**), also proposed as candidate by the models, was also prepared for comparison following a described procedure.[50] Structural assignment was carried out by usual NMR techniques but also by single crystal X-ray diffraction of suitable crystals for compound **1**.

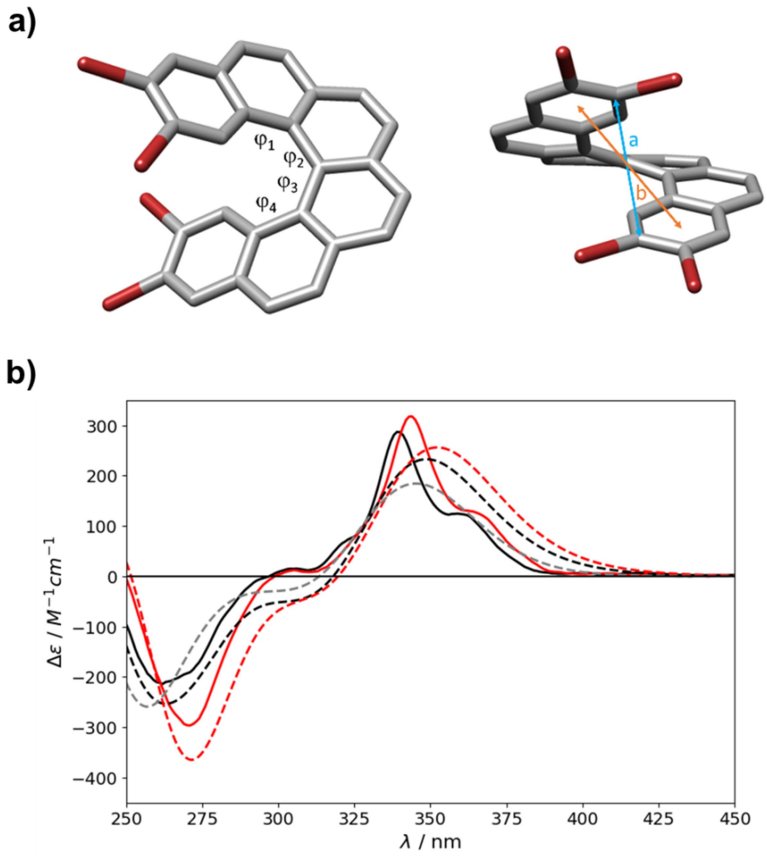

Figure 10. a) Top (left) and front (right) views of the crystal structure of tetrabromo[6]helicene **1**. Angles and distances: $\varphi_1 = \varphi_4$, 7.2°; $\varphi_2 = \varphi_3$, 28.7°; a, 4.10 Å; b, 4.41 Å. H atoms have been omitted for clarity.[51] b) Theoretical (dashed lines) and experimental (solid lines) ECD spectra of parent [6]helicene (grey) and di- (black) (**2**) and tetrabromo (red) (**1**) substituted ones.

The results obtained from the refinement of the diffraction data confirmed the proposed structure of the compound. Its analysis revealed that the main geometrical features are similar to those of the unsubstituted [6]helicene.[52] The higher distortion is observed in the central aromatic rings, with dihedral angles of 28.7° (φ_3) while those on the more peripheral rings are 7.2° (φ_4). The mean value of the dihedral angle is, therefore, 17.9°. The dihedral angles in the crystal structure of the [6]helicene are, however, slightly larger, with values of 11.1°, 30.1°, 31.2° and 15.1° (mean value, 21.9°).[53] As a result, the C2-C15 distance and that between the centroids of the outer aromatic rings are 4.10 Å and 4.41 Å, respectively. The corresponding distances in the solid-state structure of the [6]helicene are 4.64 Å and 4.49 Å. The angles between the mean planes of the outer rings of the helicene moiety are 52.2° for **1** and 59.6° for [6]helicene. Nevertheless, these subtle differences might arise from the different packing observed in both

crystal structures. Thus, for the [6]helicene there is a higher inclusion of the outer ring of another helicene molecule within the pitch, while for **1**, this inclusion, which involves now the central part of the helicene, is less pronounced (see SI for details).

We then studied the chiroptical properties, particularly the ECD, not reported for compound **2**. For parent [6]helicene the values are also available for the comparison with the precedent ones.[6] It is worth noting that R_{max} value is not directly extracted from the experimental ECDs. They usually register the ΔA or the molar circular dichroism Δε at any wavelength. Those values result from the summatory of contributions of different S_0 to S_j transitions with different magnitudes and signs. Considering the same envelop function for all the transitions in the different compounds, DFT simulated ECD gives a suitable comparison scenario (Figure 10). Experimental ones for compounds **1** and **2** are in excellent agreement with the expected ones. Molar circular dichroism for tetrabromo[6]helicene **1** (317 $M^{-1}cm^{-1}$) is higher than for the dibromo derivative (287 $M^{-1}cm^{-1}$), matching as well their relative intensities. Those are higher than the reported for parent [6]helicene (259 $M^{-1}cm^{-1}$).[6] Overall, this final experimental work, validates the accuracy of the predicting models, according to their corresponding trainings. Thus, the reliability of the developed deep learning, turns it into a perfect tool on the rapid elucidation of optimal synthetic targets in order to maximize chiroptical properties.

CONCLUSIONS

Taking advantages of deep learning techniques, we have developed a neutral network to predict the R_{max} values of billions of halogenated [6]helicenes, from one to the full hexadecahalogenated derivatives, with a minimal computation cost. We have built three different models with increasing complexity (1-body, 2-body and N-body respectively), whose predictions reasonably correspond with the DFT-calculated values. Although the best correlation is always obtained with the N-body model it is worth noting that a parametrization of R_{max} acquire evident physical meaning when simpler 1- and 2-body models are used in derivatives with up to six halogen atoms. It has also been observed that increasing the number of halogens above four promotes a diminish of R_{max}. More interestingly, we have found a structure-properties relationship, as there are favoured positions and halogen atoms that increase its value, mainly bromine and iodine in 2,3 and 14,15 positions. An exhaustive analysis of data has been done, considering both positive and negative values of rotational strength, presenting these last lower values. Finally, the predictions have been experimentally supported by the synthesis of the two best candidates predicted by the network, confirming the optimal ECD values in excellent agreement with the predicted by the deep learning approach.

ASSOCIATED CONTENT

Data Availability Statement

All scripts utilized for training the models, along with the entire dataset, are accessible at the repository https://github.com/alfonsogijon/Helicenes_NNs.

Accession Codes

CCDC 2341175 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallo-graphic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033.

ACKNOWLEDGMENT

Financial support is acknowledged. This project has received funding from: Grant PID2020-113059GB-C21 funded by MICIU/AEI/10.13039/501100011033; Grant PID2021-125537NA.I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU; Grant PID2022-137403NA-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU; ERDF/Junta de Andalucía (D3S project P21.00247). R.G.U. also acknowledges for his FPU contract (FPU20/03582). C.M.C. thanks Junta de Andalucía for a post-doctoral grant

(POSTDOC_21_00139). S. M. L. thanks Junta de Andalucía for a postdoctoral grant (DOC_01165). We acknowledge Centro de Servicio de Informática y Redes de Comunicaciones (CSIRC), Universidad de Granada for providing the computing time. Craiyon.com was used to generate the robot of the graphical table of contents image using the prompt: “cute little robot character".

REFERENCES

(1) Shen, Y.; Chen, C.-F. Helicenes: Synthesis and Applications. Chem Rev 2012, 112 (3), 1463–1535. https://doi.org/10.1021/cr200087r.

(2) Gingras, M. One Hundred Years of Helicene Chemistry. Part 1: Non-Stereoselective Syntheses of Carbohelicenes. Chem. Soc. Rev. 2013, 42 (3), 968–1006. https://doi.org/10.1039/C2CS35154D.

(3) Gingras, M.; Félix, G.; Peresutti, R. One Hundred Years of Helicene Chemistry. Part 2: Stereoselective Syntheses and Chiral Separations of Carbohelicenes. Chem. Soc. Rev. 2013, 42 (3), 1007–1050. https://doi.org/10.1039/C2CS35111K.

(4) Abbate, S.; Longhi, G.; Mori, T. Chiroptical Properties of Helicenes. In Helicenes; Wiley, 2022; pp 373–394. https://doi.org/10.1002/9783527829415.ch11.

(5) T. J. J. Müller; U. H. F. Bunz. Functional Organic Materials; Müller, T. J. J., Bunz, U. H. F., Eds.; Wiley, 2006. https://doi.org/10.1002/9783527610266.

(6) Nakai, Y.; Mori, T.; Inoue, Y. Theoretical and Experimental Studies on Circular Dichroism of Carbo[n]Helicenes. J Phys Chem A 2012, 116 (27), 7372–7385. https://doi.org/10.1021/jp304576g.

(7) Johannessen, C.; Blanch, E. W.; Villani, C.; Abbate, S.; Longhi, G.; Agarwal, N. R.; Tommasini, M.; Lightner, D. A. Raman and ROA Spectra of (−)- and (+)-2-Br-Hexahelicene: Experimental and DFT Studies of a π-Conjugated Chiral System. J Phys Chem B 2013, 117 (7), 2221–2230. https://doi.org/10.1021/jp312425m.

(8) Kubo, H.; Hirose, T.; Nakashima, T.; Kawai, T.; Hasegawa, J.; Matsuda, K. Tuning Transition Electric and Magnetic Dipole Moments: [7]Helicenes Showing Intense Circularly Polarized Luminescence. J Phys Chem Lett 2021, 12 (1), 686–695. https://doi.org/10.1021/acs.jpclett.0c03174.

(9) Mahato, B.; Panda, A. N. Effect of Terminal Fluorination on Chiroptical Properties of Carbo[5–8]Helicenes: A Systematic Computational Study at the RI-ADC(2) Level. J Phys Chem A 2023, 127 (10), 2284–2294. https://doi.org/10.1021/acs.jpca.2c08474.

(10) Furche, F.; Ahlrichs, R.; Wachsmann, C.; Weber, E.; Sobanski, A.; Vögtle, F.; Grimme, S. Circular Dichroism of Helicenes Investigated by Time-Dependent Density Functional Theory. J Am Chem Soc 2000, 122 (8), 1717–1724. https://doi.org/10.1021/ja991960s.

(11) Warnke, I.; Furche, F. Circular Dichroism: Electronic. WIREs Computational Molecular Science 2012, 2 (1), 150–166. https://doi.org/10.1002/wcms.55.

(12) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

(13) Margraf, J. T. Science-Driven Atomistic Machine Learning. Angew Chem Int Ed 2023, 62 (26). https://doi.org/10.1002/anie.202219170.

(14) Karthikeyan, A.; Priyakumar, U. D. Artificial Intelligence: Machine Learning for Chemical Sciences. Journal of Chemical Sciences 2022, 134 (1), 2. https://doi.org/10.1007/s12039-021-01995-2.

(15) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. J Chem Inf Model 2019, 59 (6), 2545–2559. https://doi.org/10.1021/acs.jcim.9b00266.

(16) Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep Learning for Chemical Reaction Prediction. Mol Syst Des Eng 2018, 3 (3), 442–452. https://doi.org/10.1039/C7ME00107J.

(17) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. J Am Chem Soc 2022, 144 (11), 4819–4827. https://doi.org/10.1021/jacs.1c12005.

(18) Pereira, A.; Trofymchuk, O. S. Machine Learning Prediction of High-Yield Cobalt- and Nickel-Catalyzed Borylations. J Phys Chem C 2023, 127 (27), 12983–12994. https://doi.org/10.1021/acs.jpcc.3c01704.

(19) Fitzner, M.; Wuitschik, G.; Koller, R.; Adam, J.-M.; Schindler, T. Machine Learning C–N Couplings: Obstacles for a General-Purpose Reaction Yield Prediction. ACS Omega 2023, 8 (3), 3017–3025. https://doi.org/10.1021/acsomega.2c05546.

(20)Singh, S.; Sunoj, R. B. Molecular Machine Learning for Chemical Catalysis: Prospects and Challenges. Acc Chem Res 2023, 56 (3), 402–412. https://doi.org/10.1021/acs.accounts.2c00801.

(21) Zhang, S.; Xu, L.; Li, S.; Oliveira, J. C. A.; Li, X.; Ackermann, L.; Hong, X. Bridging Chemical Knowledge and Machine Learning for Performance Prediction of Organic Synthesis. Chem Eur J 2023, 29 (6). https://doi.org/10.1002/chem.202202834.

(22) Tu, Z.; Stuyver, T.; Coley, C. W. Predictive Chemistry: Machine Learning for Reaction Deployment, Reaction Development, and Reaction Discovery. Chem Sci 2023, 14 (2), 226–244. https://doi.org/10.1039/D2SC05089G.

(23) Burés, J.; Larrosa, I. Organic Reaction Mechanism Classification Using Machine Learning. Nature 2023, 613 (7945), 689–695. https://doi.org/10.1038/s41586-022-05639-4.

(24) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. J Chem Theory Comput 2017, 13 (11), 5255–5264. https://doi.org/10.1021/acs.jctc.7b00577.

(25) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. Chem Sci 2017, 8 (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.

(26) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. J Phys Chem Lett 2015, 6 (12), 2326–2331. https://doi.org/10.1021/acs.jpclett.5b00831.

(27) Pinheiro, G. A.; Mucelini, J.; Soares, M. D.; Prati, R. C.; Da Silva, J. L. F.; Quiles, M. G. Machine Learning Prediction of Nine Molecular Properties Based on the SMILES Representation of the QM9 Quantum-Chemistry Dataset. J Phys Chem A 2020, 124 (47), 9854–9866. https://doi.org/10.1021/acs.jpca.0c05969.

(28) Collins, E. M.; Raghavachari, K. A Fragmentation-Based Graph Embedding Framework for QM/ML. J Phys Chem A 2021, 125 (31), 6872–6880. https://doi.org/10.1021/acs.jpca.1c06152.

(29) Bhat, V.; Sornberger, P.; Pokuri, B. S. S.; Duke, R.; Ganapathysubramanian, B.; Risko, C. Electronic, Redox, and Optical Property Prediction of Organic π-Conjugated Molecules through a Hierarchy of Machine Learning Approaches. Chem Sci 2023, 14 (1), 203–213. https://doi.org/10.1039/D2SC04676H.

(30) Nguyen, T. H.; Le, K. M.; Nguyen, L. H.; Truong, T. N. Atom-Based Machine Learning Model for Quantitative Property–Structure Relationship of Electronic Properties of Fusenes and Substituted Fusenes. ACS Omega 2023, 8 (41), 38441–38451. https://doi.org/10.1021/acsomega.3c05212.

(31) Weiss, T.; Wahab, A.; Bronstein, A. M.; Gershoni-Poranne, R. Interpretable Deep-Learning Unveils Structure–Property Relationships in Polybenzenoid Hydrocarbons. J Org Chem 2023, 88 (14), 9645–9656. https://doi.org/10.1021/acs.joc.2c02381.

(32) Karuth, A.; Casanola-Martin, G. M.; Lystrom, L.; Sun, W.; Kilin, D.; Kilina, S.; Rasulev, B. Combined Machine Learning, Computational, and Experimental Analysis of the Iridium(III) Complexes with Red to Near-Infrared Emission. J Phys Chem Lett 2024, 15 (2), 471–480. https://doi.org/10.1021/acs.jpclett.3c02533.

(33) Sigmund, L. M.; Sowndarya, S.; Albers, A.; Erdmann, P.; Paton, R. S.; Greb, L. Predicting Lewis Acidity: Machine-Learning the Fluoride Ion Affinity of P-Block-Atom-based Molecules. Angew Chem Int Ed 2024. https://doi.org/10.1002/anie.202401084.

(34) Orsi, M.; Shing Loh, B.; Weng, C.; Ang, W. H.; Frei, A. Using Machine Learning to Predict the Antibacterial Activity of Ruthenium Complexes. Angew Chem Int Ed 2024, 63 (10). https://doi.org/10.1002/anie.202317901.

(35) Kuznetsova, V.; Coogan, Á.; Botov, D.; Gromova, Y.; Ushakova, E. V.; Gun'ko, Y. K. Expanding the Horizons of Machine Learning in Nanomaterials to Chiral Nanostructures. Adv Mater 2024. https://doi.org/10.1002/adma.202308912.

(36) Schrier, J.; Norquist, A. J.; Buonassisi, T.; Brgoch, J. In Pursuit of the Exceptional: Research Directions for Machine Learning in Chemical and Materials Science. J Am Chem Soc 2023, 145 (40), 21699–21716. https://doi.org/10.1021/jacs.3c04783.

(37) Rodríguez, R.; Naranjo, C.; Kumar, A.; Matozzo, P.; Das, T. K.; Zhu, Q.; Vanthuyne, N.; Gómez, R.; Naaman, R.; Sánchez, L.; Crassous, J. Mutual Monomer Orientation To Bias the Supramolecular Polymerization of [6]Helicenes and the Resulting Circularly Polarized Light and Spin Filtering Properties. J Am Chem Soc 2022, 144 (17), 7709–7719. https://doi.org/10.1021/jacs.2c00556.

(38) Dhbaibi, K.; Abella, L.; Meunier-Della-Gatta, S.; Roisnel, T.; Vanthuyne, N.; Jamoussi, B.; Pieters, G.; Racine, B.; Quesnel, E.; Autschbach, J.; Crassous, J.; Favereau, L. Achieving High Circularly Polarized Luminescence with Push–Pull Helicenic Systems: From Rationalized Design to Top-Emission CP-OLED Applications. Chem Sci 2021, 12 (15), 5522–5533. https://doi.org/10.1039/D0SC06895K.

(39) Kettner, M.; Maslyuk, V. V.; Nürenberg, D.; Seibel, J.; Gutierrez, R.; Cuniberti, G.; Ernst, K.-H.; Zacharias, H. Chirality-Dependent Electron Spin Filtering by Molecular Monolayers of Helicenes. J Phys Chem Lett 2018, 9 (8), 2025–2030. https://doi.org/10.1021/acs.jpclett.8b00208.

(40) Kiran, V.; Mathew, S. P.; Cohen, S. R.; Hernández Delgado, I.; Lacour, J.; Naaman, R. Helicenes—A New Class of Organic Spin Filter. Adv Mater 2016, 28 (10), 1957–1962. https://doi.org/10.1002/adma.201504725.

(41 Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. J Am Chem Soc 1937, 59 (1), 96–103. https://doi.org/10.1021/ja01280a022.

(42) Lee, A.; Sarker, S.; Saal, J. E.; Ward, L.; Borg, C.; Mehta, A.; Wolverton, C. Machine Learned Synthesizability Predictions Aided by Density Functional Theory. Commun Mater 2022, 3 (1), 73. https://doi.org/10.1038/s43246-022-00295-7.

(43) Huang, B.; von Rudorff, G. F.; von Lilienfeld, O. A. The Central Role of Density Functional Theory in the AI Age. Science (1979) 2023, 381 (6654), 170–175. https://doi.org/10.1126/science.abn3445.

(44) Taniike, T.; Takahashi, K. The Value of Negative Results in Data-Driven Catalysis Research. Nat Catal 2023, 6 (2), 108–111. https://doi.org/10.1038/s41929-023-00920-9.

(45) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. Angew Chem Int Ed 2022, 61 (29). https://doi.org/10.1002/anie.202204647.

(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian09. Revission B.01. Gaussian 09. Revission B.01. Inc.: Wallingford CT 2010.

(47) M. Abadi; A. Agarwal; P. Barham; E. Brevdo; Z. Chen; C. Citro; G. S. Corrado; A. Davis; J. Dean; M. Devin; S. Ghemawat; I. Goodfellow; A. Harp; G. Irving; M. Isard; Y. Jia; R. Jozefowicz; L. Kaiser; M. Kud- lur; J. Levenberg; D. Mané; R. Monga; S. Moore; D. Murray; C. Olah; M. Schuster; J. Shlens; B. Steiner; I. Sutskever; K. Talwar; P. Tucker; V. Vanhoucke; V. Vasudevan; F. Viégas; O. Vinyals; P. Warden; M. Wattenberg; M. Wicke; Y. Yu; X. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXIiv:1603.04467 2016.

(48) Uceda, R. G.; Cruz, C. M.; Míguez-Lago, S.; de Cienfuegos, L. Á.; Longhi, G.; Pelta, D. A.; Novoa, P.; Mota, A. J.; Cuerva, J. M.; Miguel, D. Can Magnetic Dipole Transition Moment Be Engineered? Angew Chem Int Ed 2024, 63 (4). https://doi.org/10.1002/anie.202316696.

(49) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. J Comput Chem 2012, 33 (5), 580–592. https://doi.org/10.1002/jcc.22885.

(50) Schulte, T. R.; Holstein, J. J.; Clever, G. H. Chiral Self-Discrimination and Guest Recognition in Helicene-Based Coordination Cages. Angew Chem Int Ed 2019, 58 (17), 5562–5566. https://doi.org/10.1002/ange.201812926.

(51) Deposition Number CCDC 2341175 Contains the Supplementary Crystallographic Data for This Paper. These Data Are Provided Free of Charge by the Joint Cambridge Crystallographic Data Centre and Fachinformationszentrum Karlsruhe Access Structures Service Www.Ccdc.Cam.Ac.Uk/Structures.

(52) Dračínský, M.; Storch, J.; Církva, V.; Císařová, I.; Sýkora, J. Internal Dynamics in Helical Molecules Studied by X-Ray Diffraction, NMR Spectroscopy and DFT Calculations. Phys Chem Chem Phys 2017, 19 (4). https://doi.org/10.1039/c6cp07552e.

(53) We Considered for Comparison the Crystal Structure Obtained from Data Collected at the Same Temperature (100 K) (CSD Reference: HEXHEL02).