

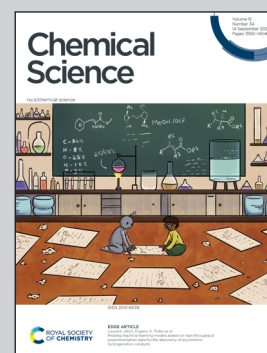
Showcasing research from Professor Senge's laboratory, School of Chemistry, Trinity College Dublin, Ireland and the Institute for Advanced Study, Technical University Munich, Germany.

Quantifying near-symmetric molecular distortion using symmetry-coordinate structural decomposition

SCSD is a new method and online toolset for analyzing crystal structures of molecules distorted from regular shape. Separating atom movements into symmetry modes, molecular conformation can be quantified and compared easily across databases, offering quantitative insight into the "molecular origami" approach to functional materials. Structural features can also be visualized by the program 'painting' the molecular conformations in the artistic De Stijl of Piet Mondrian.

Cover design by Ella Marushchenko.

As featured in:



See Christopher J. Kingsbury and Mathias O. Senge, *Chem. Sci.*, 2024, 15, 13638.

Cite this: *Chem. Sci.*, 2024, 15, 13638 All publication charges for this article have been paid for by the Royal Society of ChemistryReceived 11th March 2024  
Accepted 1st August 2024

DOI: 10.1039/d4sc01670j

rsc.li/chemical-science

# Quantifying near-symmetric molecular distortion using symmetry-coordinate structural decomposition†

Christopher J. Kingsbury \*<sup>a</sup> and Mathias O. Senge \*<sup>ab</sup>

We imagine molecules to be perfect, but rigidified units can be designed to bend from their ideal shape, discarding their symmetric elements as they progress through vibrations and larger, more permanent distortions. The shape of molecules is either simulated or measured by crystallography and strongly affects chemical properties but, beyond an image or tabulation of atom-to-atom distances, little is often discussed of the accessed conformation. We have simplified the process of shape quantification across multiple molecular types with a new web-accessible program – SCSD – through which a molecular subunit possessing near-symmetry can be dissected into symmetry coordinates with ease. This parameterization allows a common set of numbers for comparing and understanding molecular shape, and is a simple method for database analysis; this program is available at <https://www.kingsbury.id.au/scsd>.

## Introduction

To understand a molecule, we must understand its shape. Chemistry, after all, is the study of nuclei and electrons, with electronic interactions and energy levels impacted by residence within molecular geometries. This is especially true of molecular components or rigid groups which are distorted from their usual arrangement, exhibiting stretching, twisting or bending from regular structure.<sup>1,2</sup> Those molecules exhibiting irregular shape have distinct chemical properties from their regular counterparts, especially in interaction with light and with reaction partners. In crystal structure reports, distortion-induced change is often represented in qualitative terms which are difficult to compare across multiple structures.

Shape change in metal ion coordination environments is well known to affect measured properties, especially spectral, magnetic, and biological properties.<sup>3</sup> Interaction with visible light is conceptually linked to d-block metal complexes and polycyclic, predominantly carbon molecules with distributed  $\pi$ -systems, each with electronic states separated by an energy compatible with the incident photon. First-row transition metal complexes are often three-dimensional, six-coordinate species near to the ideal octahedral coordination environment, with

a history of numeric analysis derived from ligand effects and corresponding alteration of coordination environment, such as Jahn and Teller's work on spontaneous symmetry-breaking.<sup>4</sup> Algorithms such as bond valence sum<sup>5</sup> and the continuous symmetry measure<sup>6</sup> allow for the digestion of the distortion of these coordination polyhedra into single numerical terms which assist the correlation of structure to properties. Symmetry-coordinate analysis as previously described by Murray-Rust, Bürgi and Dunitz<sup>7</sup> for the simple enumerated inorganic polyhedra is perhaps the most sophisticated of these analyses, though has seemingly not found a place in routine crystal structure reporting. Small changes in inorganic polyhedra, especially in crystalline minerals, can be crucial to understanding photochemical properties.

Polycyclic aromatic molecules are often similarly approximately symmetric, though with far greater chemical diversity lack a similar common coordinate set which can be referenced. In aromatic molecules, chemical and electronic modification, steric demand and mechanical manipulation can be used to disrupt planarity, or indeed any of the symmetry elements; this distortion is usually described in qualitative terms.<sup>8</sup> The quantification of distortion usually relies on the reporting of individual bond torsions or deviations from mean planes, but these numbers can quickly become unwieldy in large examples or conflate multiple structural relationships.<sup>9</sup> Normal-coordinate studies<sup>10–12</sup> which require calculated symmetry-gated vector tables are not readily available for molecules which may only have a few reported examples. Similarly, results generated without shared coordinates and vector tables are difficult to compare. Conformation in the solid state can be a particularly confounding problem for *ab initio* analysis, even for small molecules, given the multiple demands that inform shape upon

<sup>a</sup>School of Chemistry, Chair of Organic Chemistry, Trinity College Dublin, The University of Dublin, Trinity Biomedical Sciences Institute, 152-160 Pearse Street, Dublin, D02R590, Ireland. E-mail: ckingsbury@ccdc.cam.ac.uk; sengem@tcd.ie

<sup>b</sup>Institute for Advanced Study (TUM-IAS), Technical University of Munich, Lichtenberg-Str. 2a, 85748 Garching, Germany. E-mail: mathias.senge@tum.de

† Electronic supplementary information (ESI) available: SCSD program description; implemented models; test cases of SCSD analysis; Mondrian typing. See DOI: <https://doi.org/10.1039/d4sc01670j>



SCSD for refcode JEZXEJ retrieved from the anthracene dataset using  $D_{2h}$  symmetry

Table 1: SCSD parameters derived from the atom positions

Symm	Sum (SCSD, Å)	PCA Err	PCA (1)	PCA (2)
$A_g$	1.8084	0.409	-0.581	-1.836
$B_{1g}$	0.4109	0.087	0.386	0.093
$B_{2g}$	0.0382	0.005	0.026	-0.031
$B_{3g}$	0.1097	0.000	-0.110	0.002
$A_u$	5.6757	0.019	5.670	-0.338
$B_{1u}$	0.6875	0.076	-0.628	0.113
$B_{2u}$	0.0626	0.052	0.032	0.028
$B_{3u}$	0.0780	0.048	0.054	-0.005

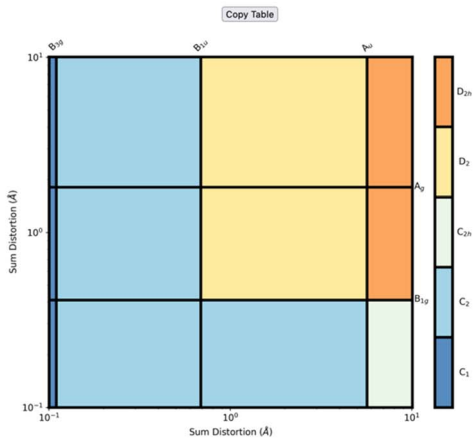


Figure 1: Neoplastic representation of the SCSD parameters

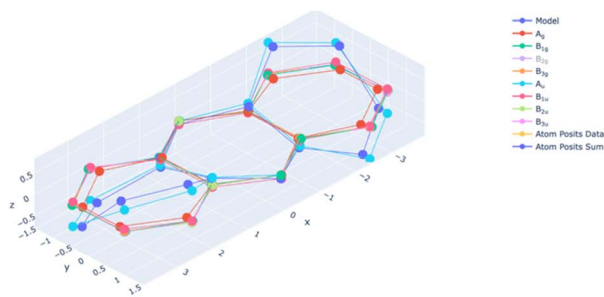


Figure 2: Atom positions and symmetric decomposition modes

Table 2: Comparison with precalculated databases

CCDC link	scsd link	$\sqrt{\sum(S(1)-S(2))^2}$
JEZXEJ	JEZXEJ	0.0 Å
SEWZEA	SEWZEA	0.391 Å
JEZXIY	JEZXIY	0.416 Å
JEYGEC	JEYGEC	0.505 Å
JEZPIQ	JEZPIQ	0.62 Å

Copy Table

```

show SCSD values
mode_name : Ag
sum_dist : 1.8084
dev_mat : [[[-7.19700000e-02 -1.98725000e-01 2.52283670e-16]
[-5.32050000e-02 -1.96767500e-01 2.23291535e-16]
[-2.47675000e-02 -2.55500000e-02 1.16279142e-16]
[ 8.73142552e-17 2.65000000e-03 8.73142552e-17]
[-7.19700000e-02 1.98725000e-01 1.88996985e-16]
[-5.32050000e-02 1.96767500e-01 7.46399262e-17]
[-2.47675000e-02 2.55500000e-02 3.09151369e-17]
[ 2.47675000e-02 -2.55500000e-02 -3.09151369e-17]
[ 2.47675000e-02 2.55500000e-02 -1.16279142e-16]
[ 5.32050000e-02 1.96767500e-01 -2.23291535e-16]
[ 7.19700000e-02 -1.98725000e-01 2.52283670e-16]
[ 7.19700000e-02 -1.98725000e-01 -1.88996985e-16]
[ 5.32050000e-02 -1.96767500e-01 -7.46399262e-17]
[-8.73142552e-17 -2.65000000e-03 -8.73142552e-17]]]
dev_mat_plus_total : [[ [ 3.60333000e+00 5.16775000e-01 2.52283670e-16]
[ 2.43279500e+00 1.21383250e+00 2.23291535e-16]
[ 1.20193250e+00 6.97050000e-01 1.16279142e-16]
[ 8.73142552e-17 1.42529500e+00 8.73142552e-17]
[ 3.60333000e+00 -5.16775000e-01 -1.88996985e-16]

```

Document prepared by the SCSD online tool – kingsbury.id.au/scsd  
 Dr. C. J. Kingsbury, Prof. Dr. M. O. Senge (2020-21) – ckingsbu@tcd.ie

Fig. 1 An example symmetry-coordinate structural decomposition output for a 'twisted' anthracene unit. The largest mode,  $A_u$ , is the principal 'bent' shape, but small deviations from this  $D_2$  symmetry also contribute to the solid-state conformation.

crystallization. A mechanism to relate chemical structure to measured atom positions in a way which is quantitative, extensible, and predictive – while still being generally approachable – would allow a shared, unambiguous language to discuss shape. We present our approach to that parameterization here.

Herein we present *Symmetry-Coordinate Structural Decomposition (SCSD)*, a program which decomposes 3D molecular coordinates into sets of symmetry coordinates, quantified and symmetry-delineated vectors in selected point groups ( $C_{2v}$ ,  $C_{2h}$ ,  $D_{2h}$  and  $D_{4h}$  among others). An example of the program output from crystallographic data of a substituted anthracene<sup>13</sup> is provided in Fig. 1. The mathematics behind this decomposition are relatively simple – each symmetry-coordinate group is the application of the symmetry elements of an irreducible representation, and their sum is the original atom positions. Symmetry-coordinate analysis is not commonly encountered in routine structural studies, perhaps due to difficulty in access to programs. SCSD provides simple program access online as a narrative aid for individual structure reports that can be useful for understanding how and why a particular molecule may be distorted in a single crystal structure. As with Hirshfeld analysis for intermolecular interactions,<sup>14</sup> or packing diagrams illustrating crystal lattices, a distortional profile quantifies organic molecules' accessed shape by simple inspection of a table or Mondrian symmetry plots,<sup>15</sup> and allows discussion of shape in quantified terms.

## An example of SCSD analysis

Four outputs are returned to the investigator of a single structure, and these are shown in Fig. 1 for an exemplar non-planar aromatic – the anthracene core of 9,11,20,22-tetraphenyl-10,21-ditolyl-tetrabenzo(*a,c,l,n*)pentacene as the chloroform solvate (CSD: JEZXEJ).<sup>13</sup>

The first table shows the magnitude of the symmetry-coordinate decomposition of the identified atoms. In this example, the twisted anthracene unit is most influenced by its  $A_u$  distortion, the largest of the deformations, as given by the "Sum (Å)" value. The  $A_u$  value is representative of the out-of-plane deformation which preserves the three  $C_2$  axes of symmetry of  $D_{2h}$ , but not its other symmetry elements. This is determined by comparing the molecule positions with  $C_2$ ,  $C_2'$  and  $C_2''$  symmetry operations applied, *versus* those with the  $C_2$ ,  $C_2'$ ,  $C_2''$ ,  $\sigma_{xy}$ ,  $\sigma_{xz}$  and  $\sigma_{yz}$  operations applied (with  $A_g$  symmetry) to develop a vector set. The sum of the magnitudes of these vectors is a value of 5.68 Å, substantial for a small molecule and indicates a noticeable and persistent distortion to the molecular species.

$A_u$  distortion is, however, only perpendicular to the mean plane of this species. The second-largest value of this column is the  $A_g$  distortion of 1.81 Å – encountered when comparing the  $A_g$  representation of the molecular unit with a model compound – anthracene itself. Bond distances are only altered by less than 0.04 Å; the majority of this distortion is an in-plane shadow of



the  $A_u$  out-of-plane distortion. This is equivalent to how a right-angle triangle with a fixed hypotenuse will contract on the adjacent edge and increase on the opposite edge, as the angle is increased from zero. Additional distortional elements ( $B_{1u}$ ,  $B_{1g}$ ,  $B_{3g}$ ) result from the crystal packing.

Principal component analysis (PCA) is explained in more detail below, but is simply a set of vector components which explain the most common deformation in molecules of each type, with each set of symmetry operations. As one can expect that molecules deform along their lowest-energy pathways, these resemble the vibrational normal modes which can be calculated for a small molecule but are derived from many crystal structures. Each of these normal modes is orthogonal to all other modes of that representation and can be fitted to our measurement data using linear summation; this table indicates that the first two principal components of  $A_u$  distortion ( $A_{u1}$ : 5.67 Å,  $A_{u2}$ : -0.34 Å), model this distortion with 99.6% accuracy (residual: 0.02 Å). Principal components can be explored in more detail through the 'model' page, which is accessible through the hyperlink at the head of the page.

The Mondrian diagram shows how out symmetry coordinates relate to the point group symmetry of the molecule.<sup>15</sup> In any given crystal structure, atom positions will result from distortions from multiple intra- and inter-molecular forces at different magnitudes, and this diagram shows how these may relate to each other and the symmetry.

Large distortions will exert greater influence on properties and contribute more to our perception and narrative of the molecule in the solid state, and so influence more of the diagram's area. What most would recognize as the point-group symmetry of this molecule is an area center to upper-right of the diagram, a color region here indicative of  $D_2$  symmetry. The  $C_1$  symmetry imposed by the space group is shown in the lower left.

The third diagram is interactive and can be accessed online;<sup>15</sup> with this viewer, one can explore in more detail the shapes resulting from symmetry-coordinate decomposition. An interactive 3D viewer is essential to understanding the shape of unfamiliar three-dimensional objects, and different modes can be shown or hidden by clicking on the legend.

A second table indicates the most similar conformations observed in the reference database, the CCDC CSD. These are calculated by minimum sum-square-distance of SCSD magnitudes and generally show compounds with similar chemistry, as well as shape. In this example, the most similar is the structure itself, then the same molecule as the ethyl acetate solvate, then three similar structures from the same paper.

More information is available in a click-box at the bottom of this page, including all the raw data that is used in the generation of the above plots. In all, these presentations show different aspects of molecular shape, and provide the investigator with quotable information on the deformation and symmetry.

## Program description

A Symmetry-Coordinate Structural Decomposition (SCSD) of a set of molecular coordinates decomposes this model into a set

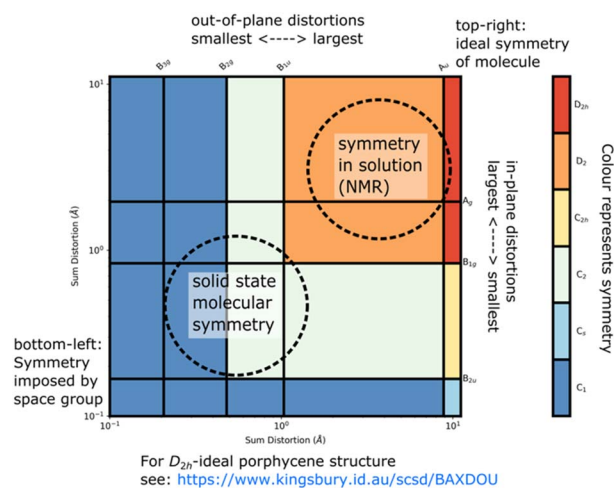


Fig. 2 An annotated Mondrian symmetry diagram for an example saddle-shaped porphycene (CCDC: BAXDOU).<sup>16</sup>

of modes aligned to the irreducible representations of the point group. The sum of SCSD modes is identical to the input atom positions, whether from a single crystal X-ray diffraction (SC-XRD) experiment or simulation of molecular shape. These data can be presented as the simple magnitudes of these separated modes, the symmetry resulting from combined distortions as a Mondrian diagram (see Fig. 2),<sup>16</sup> and an interactive visualization of these distortions in 3D space, each instructive in understanding accessed shape. From our previous studies on porphyrins,<sup>12,15</sup> and for the examples in Table 1, we have found the magnitude (e.g.,  $B_{2u} > 6$  Å) sufficient to indicate a shape, such as *saddle*, such that a name and number can be used mostly interchangeably.

When smaller deformations appear, these may indicate favorable intermolecular interactions or vibrations from this principal shape. Data for case studies listed in Table 1 is available online,<sup>17</sup> with data available from the authors. The SCSD program reports values as a real space magnitude, in Å, and can be reported and compared between series of similar molecules (for a flowchart of the procedure see Fig. 3). The magnitude is simply the summation of how much all atoms in the framework are deforming in a certain way. In more technical terms, this results in a set of vectors indicating deviation from co-kernel symmetry which obeys the operations of a given irreducible representation – one of the rows of a point group table. A larger number means more distortion away from a higher symmetry. Extracted coordinates can be used for further computational analysis through Principal Coordinate Analysis (PCA);<sup>18</sup> this is best suited for comparison of large datasets. Individual structure analysis is designed to be as seamless to the investigator as possible, does not require any knowledge of the molecule except ideal point group symmetry and, optionally, a minimally distorted comparison example for determining totally-symmetric ( $A_g$ ,  $A_1$ ,  $A_{1g}$ ) distortion and faster computation.

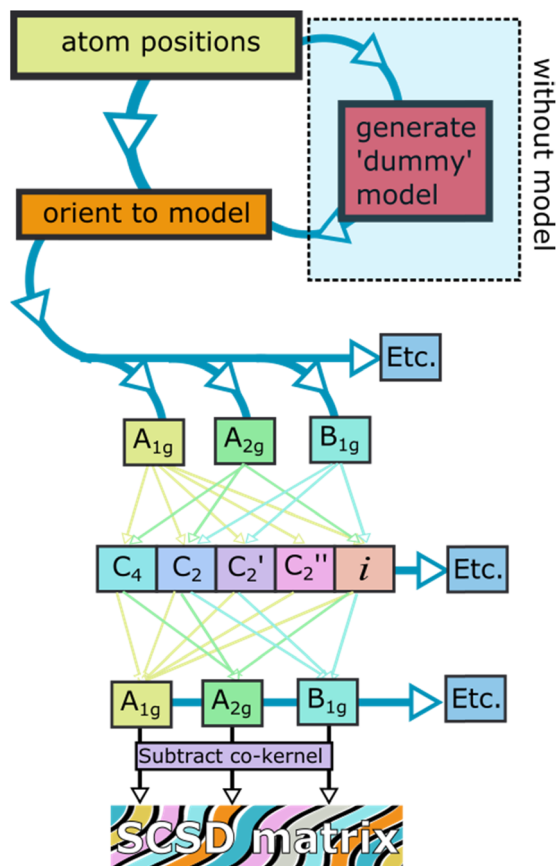
Symmetry-coordinate analysis is performed in stages outlined in Fig. 3, and all are performed by the program automatically.



**Table 1** Exemplar pre-calculated molecular databases. All online available databases are listed at [https://www.kingsbury.id.au/scsd\\_models\\_table](https://www.kingsbury.id.au/scsd_models_table)

Molecular type	Name of example
Aromatics	Anthracene
	Pyrene
	Coronene <sup>a</sup>
Fluorophores	BODIPY
	Xanthene
Molecular electronics	Tetrathiafulvalene
	Tetracyano- $\alpha,\alpha'$ -quinodimethane
	Dihydroxy- <i>p</i> -benzoquinone
	Naphthalenediimide
Porphyrinoids	Phthalocyanine
	Cucurbit[6]uril <sup>a</sup>
	Corrole
	Norcorrole
	Corrphycene
	Chlorin
Other	Phthalocyanine
	Cucurbit[6]uril <sup>a</sup>
	Carbamazepine
	Salen
	$\beta$ -Carotene
	Curcumin

<sup>a</sup> Additional approximations are used for non-perpendicular E-groups in  $D_{6h}$ .



**Fig. 3** A flowchart showing the SCSD algorithm in  $D_{4h}$ .

Firstly, the atoms and connections from the model are identified in the query structure and isolated, with the molecule moved and rotated to be a least-squares distance from the model. Secondly, the symmetry operations of each of the rows of the point group table are imposed on the atom positions separately, taking the mean of a position and its symmetry-equivalent atom yields coordinates which conform to that symmetry. For example, in  $C_{2v}$ , we will generate four sets of coordinates, those corresponding to  $A_1$ ,  $A_2$ ,  $B_1$  and  $B_2$  – and these coordinate sets will have  $C_{2v}$ ,  $C_2$ ,  $C_s(\perp x)$  and  $C_s(\perp y)$  symmetry respectively. Thirdly, these coordinate sets are subtracted from their co-kernel symmetry, which is the next-most-symmetric representation of the molecule – in  $C_{2v}$ , each of the co-kernels of  $A_2$ ,  $B_1$  and  $B_2$  is the  $A_1$  representation, and the  $A_1$  is compared to the model. This subtraction generates vectors which relate the model to its deviation in each symmetry representation, the decomposition; the summation of these decompositions is the original atom positions after the first step. Finally, the vector magnitudes are summed – this is what is reported as the symmetry-coordinate magnitude and can be understood as the amount to which the whole molecule is deformed in this manner.

Principal Component Analysis is an optional additional step for database analysis, in which vector sets from step 3 in multiple crystal structure collections are compared, with their most prominent common features identified. This generates a set of perpendicular vectors which are similar to normal modes, and a normal-coordinate decomposition can be performed. A tutorial is available at <https://github.com/cjkingsbury/scsd/> that walks an investigator through the process.

Descriptions of similar algorithms have been reported which assess the ideal point-group assignment of calculated and experimentally-determined molecules<sup>19</sup> – a similar process is used in SCSD to orient the molecule towards the relevant symmetry operations when no model is provided. SCSD analysis is conceptually the inverse of the continuous symmetry measure; splitting the positional data into constituent modes can allow each symmetry mode to be classed by its contribution to – rather than deviation from – any individual point group assignment.

The assignment of a value to the distortion of each molecule allows for structural chemistry databases to be interrogated as a function of shape, where similarly shaped molecules will presumably form clusters. This method allows more information to be determined with only a small number of measurements as orthogonal representations have minimal influence on one another. PCA, cluster identification and other methods to develop quantitative three-dimensional quantitative structure-activity relationships (3D-QSAR) are included in the library; analysis can be performed through a python notebook interface, such as Jupyter.<sup>20</sup> Symmetry coordinate decomposition followed by PCA is found to be a good method of parameter reduction in structural chemistry, potentially useful for constraining crystallographic refinement. PCA is discussed further below.



Finally, this program addresses a niche for comparative structural analysis – simplifying (pseudo-)normal-coordinate analysis for a wide range of chemical substructures. Generalized structure prediction frameworks attempt to model structural variations while providing little in the way of insight, individualized structure description can be over-long and inconsistent between individual reports. We believe SCSD is suited for human-understandable structure comparison and a numerical approach to literature comparison in one publication, or to allow for insight and review of medium-sized sets (*ca.* 100–10 000 structures) derived from a crystallographic database. As such, one can succinctly summarize the conformational landscape of crystal structures of a single component. Our goals were easier generation of data, understanding and interpretation of shape. For this reason, we have provided an online interface – including interactive plots – so that the methodology is immediately understandable and within reach for all. Lastly, calculations can be performed in a notebook format; an easy and extensible procedure for many-structure investigation is included. SCSD can be installed with the command “pip install scsdp”. This software is distributed under a licensing scheme that allows only non-commercial scientific and educational use.

With this program, we first sought to validate our assumptions around how molecules might deform from their symmetric state. From initial studies on exemplar molecules detailed in Table 1 (see ESI 2†), the first principal component mode describes 90–95% of the variance in the dataset, with the second PCA mode increasing this to ~99%. Beyond the second PCA mode, residual distortion is normally below the usual atom positional accuracy in crystallographic analysis. A simplified model of distortion for a molecule [*e.g.*, boron dipyrromethene (BODIPY) distorts as *aileron* ( $A_2$ ) or *butterfly* ( $B_2$ )] can be used as a convenient shorthand for the total distortion by symmetry arguments, for the first PCA mode  $A_2(1)$ , and for the first normal mode  $A_2(1)$ , all having similar shape. *Aileron* and *butterfly* are simple names that can relate these movements to familiar objects; equivalent designations like “twisting” or “folding” might be equally appropriate. No more than two parameters from each dissymmetric basis are required to generate a model as accurate as the measurement. This mirrors the way *saddle* and *ruffle* shapes (towards  $D_{2d}$ ) of porphyrin derivatives<sup>9a,12,21</sup> have a semantic, approximate meaning coincident with the symmetry coordinates and a fixed normal coordinate vector description. Across the structures indicated in Table 1, no meaningful relationships were found from chemical patterning to molecular shape beyond the first two parameters of each irreducible representation, excepting  $A_{1g}$ ,  $A_1$  or  $A_g$ . When different strategies are employed to distort a molecule – steric hindrance, fusion of an additional aromatic ring or exocyclic strapping, for example – new examples of molecules could still be modelled satisfactorily with combinations of these same two parameters. This finding is in line with our earlier findings on porphyrin macrocycles.<sup>12</sup>

Symmetry-coordinate magnitude is a simple, measured shape descriptor that aligns with how shape is often discussed in structural chemistry. This simple version can compare the

structural features of a limited crystal series as the shape may change with, *e.g.*, temperature, pressure or in an electronic excited state. Structures with similar SCSD values will be similar, constrained by related chemical demands; similarly shaped molecules will hold similar SCSD values and display similar shape-determined properties. If many molecules with a certain chemical patterning access the same shape, it is reasonable to assume that this shape persists to the chemical in solution, for example.

Just as the distortion is important, so the symmetry that is accessed is an important feature. These symmetry coordinates are how a chemist may describe shape when wishing to relate that a structure is perhaps *both bent and twisted* – it is sometimes hard to relate that this combination results in  $C_2$  symmetry. The presence or absence of symmetry can affect the allowed or forbidden optical transitions, as in Laporte's rule.<sup>4d</sup>

Further investigation of a dataset requires that a set of normal modes is generated; these can be implemented from external sources, such as molecular mechanics simulation, or calculated from provided data through PCA. As such, this program can be used to ascertain the accessed conformational space which best approximates the data, with components suggestive of the calculated lowest frequency vibrational modes for a particular subset.

With these tools, data from thousands of structure reports can be decomposed using only a handful of expressions and summarized using these small parameter sets. Examples of structure comparison are provided below; databases and notebooks are available upon request. These data are an excellent beginning point for structural review of each individual molecular component in quantitative terms.

The SCSD tool (see ESI 1† for a description of the program) is available to use online at <https://www.kingsbury.id.au/scsd>, or as a standalone program for academic use through GitHub or by contacting the authors.<sup>22</sup>

## Discussion

The interest in writing an easy-to-use general form of normal-coordinate analysis came from our investigations surrounding the “Mondrian” symmetry graphs (see Fig. 2), in which the dissymmetric modes of a molecule are presented as a pseudo-artwork pastiche.<sup>15</sup> Mondrian plots are a method which demonstrates the approximate molecular point group at multiple levels of scrutiny to the atom positions. Where a molecule appears, for example, in  $Z' \geq 1$  structures, the molecule may have no symmetry which relates its constituent atoms; nonetheless, approximate symmetry relationships, and their measurement, are often readily observable. This plot style, described in detail in our previous paper displays the magnitude of symmetry coordinates on logarithmic axes as vertical and horizontal lines; these lines each act to separate two symmetry point groups – with, and without the consideration of that dissymmetric movement. As we progress from upper right to lower left, we cross more of these boundaries and take on more and finer-grained information of the molecule, and its distortional profile becoming ever closer to the measured. At



the same time, we progress from an idealized representation to the measured one, often with  $C_1$  representation. In reverse, from lower left to upper right, we would be acquiring symmetry in the sequence of smallest distortions, conceptually making the molecule more symmetrical by removal of distortion until arriving at our idealized model. We intuit molecules as somewhere between these two extremes – the precisely measured “real” and the assumed perfect “forms”.

The emergent features of these graphs were intriguing, especially the L-shaped correlations between the in-plane and out-of-plane features, often observed for a porphyrin distorted from regular structure.<sup>12,21</sup> Similarly, the regularity of logarithmic molecular distortion was both aesthetically pleasing and implied that there may be a deeper relationship between the elements of molecular structural distortion. As part of projects aimed at using conformational design for the molecular engineering of functional porphyrin scaffolds,<sup>23</sup> we have investigated the structural underpinnings of porphyrin-based chiral sensors and organocatalysts,<sup>24</sup> where large and small distortions from structure both play an important role. ‘Artworks’ such as the ‘Mondrian plots’ present a view into an obfuscated world of picometer-scale distortion and fine control, allowing a simple additive metric for understanding this molecular chirality. The additive nature is important, as any chiral photochemical or templating effect can be expected to grow in proportion to the dissymmetry imposed on the molecule.

Investigating the intricate relationships between solid-state structure, symmetry, and substitution, moving beyond porphyrins to apply general insights to molecules required these new tools. We found that, owing to the tendency for similar molecules to bend in similar ways, using a normal-coordinate framework and its underlying pre-computed modes was not necessary for generating meaningful insight from the artwork or when writing about shape. The general solution of point-group-derived symmetry coordinates was compelling as an analysis routine for individual structures as all it required was the atom positions.

Crystallography for small molecules generates a lot of precise data;<sup>25</sup> how these molecules are reported is at the discretion of the crystallographer, relying on a wealth of experience in determining the important structural features and close contacts. In routine organic communications, however, often little space is reserved for discussing the molecular shape or packing beyond structure validation. The conformation of an individual molecular unit is perhaps of little interest to the synthetic chemist – modulated by the molecules’ immediate crystal environment, polymorphism and the assumptions or ‘fixes’ made by the crystallographer in solving the structure, including space group, features not necessarily representative of the structure in solution. The influence of these constraints can be downplayed by looking at multiple structures of the same unit, understanding molecular shape through careful comparison of multiple samples of a conformational landscape; not usually pursued except where polymorphism is indicated. SCSD, intended as an extra analysis routine in the tool-belt of a crystallographer, allows one to rapidly assess, contextualize

and compare varied crystallographic data on symmetry and concerted atom movement terms, and simplifies the process of normal-coordinate analysis. For example, where one would report a mean deviation from a plane (a convolution of any number of distortional types), one can report the components of distortion in irreducible representation terms. This interface is designed to function on an unmodified *.pdb* or *.sd* input file and can be performed and understood without any specialist knowledge in computational chemistry.

A semantic divide can be observed in the distortional modes – between those which are intended (or designed), and therefore contribute to a “canonical” symmetry representation (how the molecule may be described), and those deviations which are below this arbitrary threshold. Canonical representation is not something fixed; this may be a conformation profile consistent with spectroscopic results, observed in solution-state conformational studies or in most crystal structures, or simply a common conformational cliché common to a structure type. Smaller modes can be thought as components of a vibrational structure upon crystallization; while potentially useful in the assignment of vibrational and vibronic spectra, these also indicate how the conformational compromise to minimize lattice energy is achieved. These modes and deviances are typically ignored in structural discussion. Structural variance under polymorphism would indicate multiple local minima of this combined function – either small (in disorder) or large – as two or more distinct self-catalyzed precipitation processes, such as (2,3,7,8,12,13,17,18-octaethylporphyrinato)nickel(II) – which exhibits distinctly different near-planar and ruffled polymorphic examples with the same chemical composition.<sup>26</sup>

SCSD-generated magnitudes can be used in a similar manner to how Normal-coordinate Structural Decomposition (NSD) has been used for porphyrin analysis – to determine the effectiveness of any distortion-inducement strategy, such as metalation, oxidation state change or peripheral substitution, and to correlate these strategies with photophysical attributes and biological functions in small datasets.<sup>10–12,15,21,27</sup> With this program, a large database of conformational information can be generated for structural review and cluster analysis on any near-symmetric molecular subunit, an easy method of surveying the conformational landscape of a component, *e.g.*, those substructures indicated in Fig. 4. Comparing structures generated by screening methods or through serial crystallography often requires the selection of choice attributes of the crystal, like a select bond distance; symmetry coordinates are an easy, fast global parameterization, and normal coordinates are ideal for turning this insight into bidirectional quantitative relationships through curve fitting.

Normal-coordinate analysis is a worthwhile endeavor for the emerging field of automated and machine learning-driven solid-state analysis,<sup>28</sup> allowing for parameter reduction while sacrificing little in the way of atom-positional accuracy. SCSD parameterization allows for symmetry modes to be derived directly from the symmetry-banded data, rather than from precalculated modes, and therefore removing at least one layer of computational model complexity. Similarly, compared with the simple direct PCA of crystallographic data, SCSD



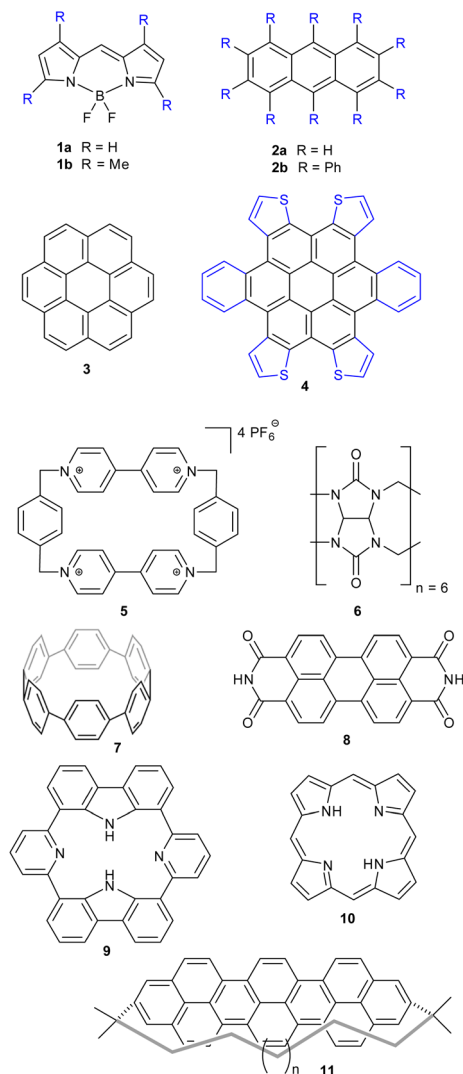


Fig. 4 Examples of classes of compounds used for illustration.

parameterization allows a significantly smaller data set to produce chemically meaningful results. Using refinement constrained to these measured parameters, more accurate solutions to poor quality data from, *e.g.*, cryo-electron microscopy might be made, or approximation of solution state structure by analysis of symmetric concerns independent of the constraints of crystallographic confinement, from spectroscopy.<sup>29</sup>

Machine learning approaches to chemical reasoning (*e.g.* ANI-1)<sup>30</sup> can be incredibly useful for understanding interaction and conformation of individual units, though may fail to accurately gauge aromaticity or interaction profiles and tend to exclude metal centers. We believe the questions of interest to organic and coordination chemists, however, are often much smaller in scope – usually constrained to only one structure type, and deeply rooted in similar structure comparison rather than global modelling. SCSD allows one to reduce a query on distortional predilection imposed by a particular substitution pattern on a shared, symmetrical frame to a statistical relationship. All tools required to make those assessments are

included, with this runtime efficient enough to run on laptop hardware in minutes for large data sets.

Principal Component Analysis, from scikit-learn and implemented by this module, allows one to determine the best-fitting vector set for a large database of symmetry-banded vectors, showing similarity to the low-frequency vibrational modes which can be calculated. PCA can be a method of relating the distortional magnitude to atom positions bidirectionally. SCSD PCA is included in *scsd.py*; this taps a rich vein of structural insight when considering multiple structures, allowing the investigator a method of clustering and sampling similar molecular shapes. For example, in a comparison of BODIPY (boron dipyrromethene) molecular compounds (**1**),<sup>31</sup> substitution of a 1,3,5,7-tetramethyl-moiety (**1b**) may seem to have little effect on the structure compared to the unsubstituted comparators (Fig. 4). However, these two motifs can be separated into two clusters by a two-component model in  $A_1$  symmetry of the BODIPY core (Fig. 5 and ESI 3.1†). Real insight can be gained by numerical analysis of many structures, and, hopefully, these tools bring that analysis into the realm of the non-specialist. Should we subscribe to the axiom that each of the crystallographically-determined positions for a given component are a near-random sampling of the vibrations accessed by the canonical structure, these principal components accurately represent the measured vibrational structure of molecules, something unattainable by direct measurement previously. In some irreducible representations, symmetric molecular components exhibit a clear signal for spontaneous dissymmetry; investigation of this behavior is ongoing. Space group symmetry can clash with approximate point group near-symmetry of crystallographically-inequivalent fragments; the constraints imposed upon measurement should be considered in the reporting of these values, especially those that would impose strict coplanarity of components, and corresponding zero SCSD values.

The point groups available in the first iteration of the SCSD algorithm are  $C_{2v}$ ,  $C_{2h}$ ,  $D_{2d}$ ,  $D_{2h}$ ,  $D_{3h}$ ,  $D_{4h}$  and  $D_{6h}$ , covering common multiple-symmetry chromophore molecules of interest. As for simpler cases, in  $C_1$ , this algorithm becomes

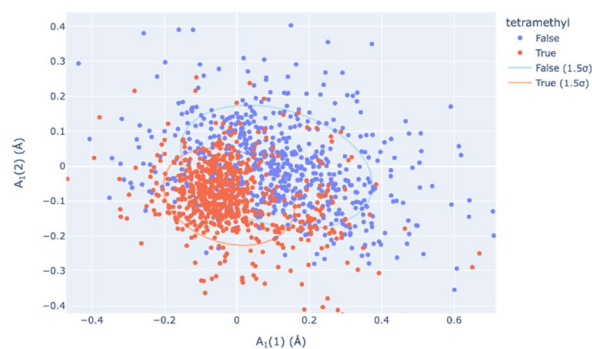


Fig. 5 The  $A_1(1)$  and  $A_1(2)$  principal components of structural decomposition in BODIPY compounds (**1**) – two distinct but overlapping clusters of symmetry coordinate vectors represent those with a 1,3,5,7-tetramethyl substructure (**1b**, blue) from those without this substructure, including **1a** (red).





a simple sum of vectors between an aligned fragment and model. With only one non-identity symmetry operation, the trivial case result is analogous to an implementation of the “symmetrize” routine and the “S” value of the parent symmetry.<sup>19</sup> While these structures are possible to probe with the SCSD routine, we believe the utility of this routine comes from disentangling multiple modes, and more conventional comparisons may be more appropriate in these cases.

A Mondrian-style diagrammatic representation<sup>15</sup> of these symmetry concerns can illuminate synchronistic dissymmetry by the same methodology presented previously for porphyrin molecules; these are generated as part of the SCSD runtime by use of a predetermined symmetry lookup table. These diagrams are an elegant manner of presenting symmetric information at multiple levels, as a symmetry mode is clearly visible both in magnitude and effect on the total symmetry of the idealized molecule. From here, choice of what is considered ‘important’ symmetry is left to the investigator.

## Examples of SCSD analysis

We have endeavored here to give useful and instructive examples of the analysis of distorted molecules across different areas of chemistry, mostly planting the idea of what can be found using these investigative tools (see ESI 3† for full descriptions). Of course, these examples are not exhaustive, nor is SCSD limited to just these frameworks.

Anthracene (**2a**) is a very common motif in organic chemistry. The distortion from regular planar structure resulting from decaphenyl – (**2b**)<sup>32</sup> or similar substitution around the anthracene periphery has  $A_u$  character – and predictably yields a near- $D_2$  conformation where the aromatic frame is twisted (ESI 3.2†).<sup>32,33</sup> Other types of side groups (*e.g.*, sulfur decoration) or other distortion-inducing patterns (alkyl strapping) intermolecular interactions, chemical modifications or inclusions in large cycles, can influence atom displacement in distinct ways, each with a characteristic SCSD footprint. A researcher may compare their novel determination with all anthracene-containing structures in the CCDC-CSD<sup>25</sup> to determine most-similar structures, for example, or to draw a substitution-distortion relationship.

Larger aromatics, such as substituted coronene (**3**) materials, are useful as model compounds for graphene, as well as nonplanar polycyclic aromatic hydrocarbons being interesting molecular electronics components. As an example, *a,d,j,m*-tetrathienyl-dibenzocoronene (**4**) has  $D_{2h}$  symmetry on paper – steric clash between adjacent units bifurcate the molecular modelling into two distinct conformations, with  $C_{2h}$  and  $C_{2v}$  near-symmetry (ESI 3.3†).<sup>34</sup>

Larger still, a molecular fragment such as Stoddart’s “Blue Box” (**5**),<sup>35</sup> a mainstay of host–guest chemistry, can be quantified for its response upon inclusion of other fragments. It is an important early example of supramolecular design, and how molecular machines can be controlled by redox state. The cycle exists as a strained near- $D_{2h}$  symmetry compound and has 225 examples in the CSD. One obvious separate cluster can be identified in the data of the first two  $A_g$  principal components, where

a distinct cluster of data points (the radical, reduced form) is evident from the kernel density estimation (KDE, Fig. 6, upper right), as well as a bisected main cluster [naphthalene-including and non-naphthalene (TTF, hydroquinone) including]. This is one method of showing the clear structural distinction of host–guest chemistry on the molecular structure. Clearly, the structural variation from different inclusion compounds, chemical structures and crystallization conditions introduces noise, but clear signals can be obtained, and features identified and summarized succinctly. Because these derived parameters are not dependent on any individual atom, this can be a robust method of assigning oxidation state in organic molecules.

Even larger groups, such as the cucurbit[6]uril toroidal host (**6**),<sup>36</sup> can be studied in this automatic fashion, illuminating the mechanisms of host response towards different intercalants (ESI 3.5†). *E.g.*, threading of an alkyl chain promotes expansion; with  $NH_2$  units, as in 1,5-diaminopentane, a contraction can be observed. A flat aryl group can be seen to promote an elliptical deformation.

Other test cases reported in the ESI† are for cycloparaphenylenes (*e.g.*, **7**) as nanotube fragments and example of structures with rigidification by angular strain (ESI 3.6†);<sup>37</sup> perylene-3,4,9,10-tetracarboxylic diimide (**8**, PDI) (ESI 3.7†), esp. distorted PDI compounds as circularly polarized luminescent materials;<sup>38</sup> dipyrityldicarbazoles (**9**) as model compounds for investigating shape-assisted self-assembly (ESI 3.8†);<sup>39</sup> a selection (ESI 3.9†) of porphyrinoids (porphyrin **10**) comprising Sessler’s amethyrin as an example of expanded porphyrinoids,<sup>40</sup> Brückner and Meyer’s Siamese twin porphyrin,<sup>41</sup> nonplanar phthalocyanines from Kobayashi’s group,<sup>42</sup> a triply-fused diporphyrin from Osuka and coworkers,<sup>43</sup> and corroles.<sup>44,45</sup> Other examples are a study of teropyrenes (**11**, ESI 3.10†) developed by Bodwell’s group allowing assumptions about the likely shape of molecules in the series,<sup>46</sup> and an analysis of the application of time-resolved crystallography to elucidate structural changes upon photoexcitation in inorganic complexes (ESI 3.11†).<sup>47</sup> Lastly, a brief list on Mondrian<sup>15</sup> types is given (ESI 4†).<sup>48</sup>

These individual data sets are all available online ([https://www.kingsbury.id.au/scsd\\_models\\_table](https://www.kingsbury.id.au/scsd_models_table)<sup>49</sup>) and at *e.g.* [https://www.kingsbury.id.au/scsd\\_model\\_ext/subporphyrin](https://www.kingsbury.id.au/scsd_model_ext/subporphyrin), or by contacting the authors. A worked example of the analysis of a phthalocyanine molecule with a ruffle deformation (CSD:

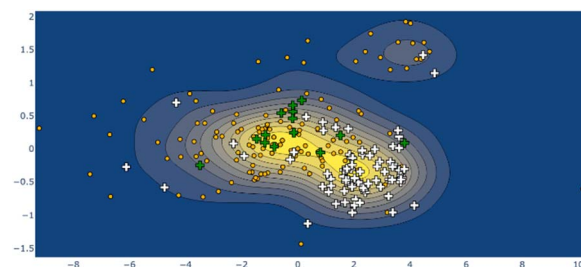


Fig. 6 KDE plot of the ‘blue box’  $A_g$  distortion; naphthalene-containing structures are shown as white crosses, tetrathiafulvalene-containing structures as green crosses, other structures as orange dots. The ‘diradical’ structure cluster is shown in the upper right of frame.



FOLDET)<sup>50</sup> is provided in ESI 1.7;† an example of using a notebook interface is included in the module available from github.<sup>51</sup>

Different approaches are being explored to using this information in chemical and protein database analysis, serial crystallography interpolation with pressure and temperature variations and ultrafast X-ray free electron laser (XFEL) studies, for crystallographic data validation and error-checking. This algorithm is potentially applicable to all pseudo-rigid groups or near-symmetry molecules, as the program is designed to be chemistry-agnostic in this regard.

## Conclusions

The SCSD routine has been described and offers a method to quantify perturbations of simple molecular components using symmetry-specific vector sets. We have shown previously that distortional vectors encode important chemical and photo-physical information, and here are demonstrated to act as a simple method of parameter reduction and fitting for molecular shapes adopted by near-symmetric molecules. A numerical structure description is of immense value to the crystallographer, assisting in understanding and relaying structural concepts across structural sets and series. The story that can be told with a small set of numerical values, in the same manner as a constructed set of normal-coordinate vectors, allows for the non-specialist to see deeper into the shape of molecules and place their shape in the proper context.

Assigning a fixed symmetry point group to a dynamic molecule is an oversimplification in many cases. SCSD is a method of quantifying that compromise between diagrammatic symmetry and measured positions in crystal structures. The notion of sum-of-vectors is a comfortable one for a chemical crystallographer, we believe, given familiarity with vibrational modes in physical chemistry. Extension of this parameterization proves just as effective in ascribing shape to simple aromatic systems, macrocycles and larger molecular groups as it has been for denoting the shape of coordination polyhedra. Finally, this is an efficient means of sampling an accessible conformational space from minimal observations, often simply those that have been determined by prior researchers.

We invite our colleagues to use SCSD terms in multi-structural analysis and review – this is a structural comparison technique which can yield genuine insight into chemical properties and the efficacy of strategies toward dissymmetry. Given that often individual crystallography papers can offer only limited context for their reported structures, a numerical basis allows one to formalize these insights succinctly and with statistical power. We hope that this conformational control, in concert with chemical modification, can elicit new modes of reactivity and tuning of the innate properties of organic and coordination compounds.

## Experimental section

The SCSD module is written to be compatible with python 3.8, with numpy,<sup>52</sup> scipy, networkx, matplotlib<sup>53</sup> and plotly<sup>54</sup> libraries, using Flask<sup>55</sup> for web templating. A copy of this

program is available on request from the authors for academic use, and at <https://github.com/cjkingsbury/scsd>.<sup>51</sup>

The alignment of point-group symmetry operations with a cluster of points in three-dimensional space is made by a minimization of the sum distortion attributable to the various symmetry operations – a least sum of pairs of the vectors  $V$  and  $V \times [A]$ , representing the original and transformed, assigned structures. As any residual rotational and transformational displacement will contribute to dissymmetric distortion, this correct alignment of the atomic structure is important to attain consistent values. The existence of separated ‘basins’ in even simple analyses requires either a basin-hopping algorithm or choice of multiple perpendicular starting points – the use of six perpendicular orientations was found to be sufficient to achieve the global minima in most cases.

The use of a quaternion parameterization allows for an approximately uniform minimization space.<sup>56</sup> The symmetric operations rotation, improper rotation, and mirror plane can be defined as operations on the unit sphere, by spherical coordinates ( $\phi$  and  $\theta$ ) converted into the unit vector  $[x,y,z]$ ; rotations are defined as a rotational unit quaternion  $[x,y,z,w]$  ( $w = 2/n$ ) applied to the coordinates, mirror planes as twice the dot product of the initial coordinates and the vector  $[x,y,z]$  subtracted from the coordinate matrix, and an improper rotation as the sequential application of reflection and rotation; inversion is simply  $[-a,-b,-c]\dots$ .<sup>12</sup> Atoms are assigned to their symmetry-opposite equivalents by using a Hungarian method solution to the assignment problem.<sup>57</sup> After alignment of the molecule to the symmetry operations, a dummy symmetric model is generated by iteratively averaging the atom positions with their symmetry equivalents through each operation of the point group; in non-2-fold axes, this is repeated. An assignment matrix for each of these operations is cached in a *scsd\_model* object. This first step is skipped if a precomputed model is provided.

The query atoms are aligned by least-squares minimization or rotation and translation either to a predefined model or to this dummy model. A molecular graph-based method for the assignment of query atoms to the model is also available for very distorted structures (*by\_graph*). There exists an upper limit to distortion where a structure with normal connectivity may be incorrectly assigned by the simple program, the *by\_graph* option can assist some of these edge-cases.

Symmetric vectors attributable to each of the irreducible representations are determined by subtraction of the “symmetrized” molecule with the symmetry operations defined by one row of an expanded point group table (*i.e.*, without grouping of equivalent operations), from the co-kernel symmetric representation – the totally symmetric representation of the class in most cases. The totally symmetric query component ( $A_g, A_1, A_{1g}$ , etc.) is compared to the model to attain an indication of the totally symmetric distortion, equal to zero when no model is used.

SCSD values obtained by this method are the sum magnitude of the individual displacement vectors which comprise the eigenmodes, equivalent to the sum of normal coordinates for a well-constructed set. Eigenmodes are shown in an interactive visualization, with details in the raw data output.



Formation of a symmetric, aligned model is performed at [https://www.kingsbury.id.au/scsd\\_new\\_model](https://www.kingsbury.id.au/scsd_new_model), or equivalently by the *yield\_model* subroutine in notebook interface, and is stored in *scsd\_models\_user.py*. This is the first step in generating a dataset; where axes are interchangeable, models are aligned such that variance  $\text{var}(x) > \text{var}(y) > \text{var}(z)$  for  $D_{2h}$ , so that near-planar molecules will generally lie in the  $x$ - $y$  plane, and  $\text{var}(x) > \text{var}(y)$  for  $C_{2v}$  ( $x$ - $z$  plane); in this way, interconvertible mode descriptors are generally consistent between measurements. Symmetric models for the compounds in Table 1 have been pre-computed and can be used with the web interface to accelerate computation, and new models can be uploaded and used for individual structure comparison.

Some point-group specific concerns are necessary when dealing with E- or T-symmetric modes; for example, in  $D_{4h}$  the number of possible  $E_g$ -mode orientations is 4, on  $x$ ,  $y$ ,  $x + y$  and  $x - y$  ( $2C'$  and  $2C''$ ) axes, where only two are necessary for a complete description of the system. The sum of these modes is multiplied by  $\frac{1}{2}$  to arrive at the correct atom positions, and the Mondrian diagram uses only the greatest magnitude vector and its orthogonal counterpart to describe the symmetry of the molecular component. This is achieved by using the derived eigenmodes as input vector sets in a *pseudo*-normal coordinate analysis, which is the *simple\_scsd* output of the *scsd\_matrix* object.

Model generation is the only computationally expensive step; large-scale database analysis is rapid even for large queries. Symmetry coordinate analysis becomes most useful for comparisons between large number of structures; this can be formulated by use of an extracted database (in *.sd* format) and a chosen model; the commands for dealing with large numbers of structures are contained within the *scsd\_collection* object.

Installing the *scsd* module is most easily performed with *pip install scsdpy*. The simplest method of interaction is by using *scsd\_mercury.py* from the program examples,<sup>51</sup> through Mercury, which identifies those models which have been previously defined and runs a symmetry-coordinate analysis based on SMARTS matching of the substructure. Additional interfaces are available to perform database and multi-structure analysis.

## Data availability

The data supporting this article have been included as part of the ESI.† The SCSD tool is available to use online at <https://www.kingsbury.id.au/scsd> or as a standalone program for academic use through GitHub (<https://github.com/cjkingsbury/scsd/>).

## Author contributions

C. J. K.: conceptualization, investigation, methodology, software, formal analysis, data curation, writing – original draft, review & editing. M. O. S.: investigation, methodology, funding acquisition, resources, supervision, writing – original draft, review & editing.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

The authors would like to thank Dr Karolis Norvaiša, Ashley H. Peak and Dr Matthew Guberman-Pfeffer for helpful discussions. We are especially grateful to Dr Alexandru A. Moldovan for technical expertise. This work was prepared with the support of the Technical University of Munich—Institute for Advanced Study through a Hans Fischer Senior Fellowship and has received funding from the European Union's Horizon 2020 research and innovation program under the FET-OPEN grant agreement no. 828779 and Science Foundation Ireland (awards 21/FFP-A/9469, 21/FFP-A/9214).

## Notes and references

- 1 J. D. Dunitz, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 14260–14266.
- 2 C. Kratky, R. Waditschatka, C. Angst, J. E. Johansen, J. C. Plaquevent, J. Schreiber and A. Eschenmoser, *Helv. Chim. Acta*, 1985, **68**, 1312–1337.
- 3 (a) M. F. Perutz, *Nature*, 1970, **228**, 726–734; (b) J. L. Hoard, *Science*, 1971, **174**, 1295–1302.
- 4 (a) H. A. Jahn and E. Teller, *Proc. R. Soc. London, Ser. A*, 1937, **161**, 220–235; (b) M. A. Halcrow, *Chem. Soc. Rev.*, 2013, **42**, 1784–1795; (c) I. B. Bersuker, *Chem. Rev.*, 2021, **121**, 1463–1512; (d) P. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, 6th edn, 1998, pp. 501–502.
- 5 I. D. Brown, *Chem. Rev.*, 2009, **109**, 6858–6919.
- 6 H. Zabrodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1993, **115**, 8278–8289.
- 7 P. Murray-Rust, H. B. Bürgi and J. D. Dunitz, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1978, **34**, 1793–1803.
- 8 M. Ball, Y. Zhong, Y. Wu, C. Schenck, F. Ng, M. Steigerwald, S. Xiao and C. Nuckolls, *Acc. Chem. Res.*, 2015, **48**, 267–276.
- 9 (a) W. R. Scheidt and Y. J. Lee, in *Structure and Bonding*, ed. J. W. Buchler, Springer, Berlin, Heidelberg, 1987, vol. 64, pp. 1–70; (b) H. L. Kee, C. Kirmaier, L. Yu, P. Thamyongkit, W. J. Youngblood, M. E. Calder, L. Ramos, B. C. Noll, D. F. Bocian, W. R. Scheidt, R. R. Birge, J. S. Lindsey and D. Holten, *J. Phys. Chem. B*, 2005, **109**, 20433–20443.
- 10 W. Jentzen, X.-Z. Shi and J. A. Shelnutt, *J. Phys. Chem. B*, 1997, **101**, 1684–1699.
- 11 J. Krumsieck and M. Bröring, *Chem.–Eur. J.*, 2021, **27**, 11580–11588.
- 12 C. J. Kingsbury and M. O. Senge, *Coord. Chem. Rev.*, 2021, **431**, 213760.
- 13 R. G. Clevenger, B. Kumar, E. M. Menuey, G.-H. Lee, D. Patterson and K. V. Kilway, *Chem.–Eur. J.*, 2018, **24**, 243–250.
- 14 M. A. Spackman and D. Jayatilaka, *CrystEngComm*, 2009, **11**, 19–32.
- 15 C. J. Kingsbury and M. O. Senge, *Angew. Chem., Int. Ed.*, 2024, **63**, e202403754.



- 16 T. Hayashi, Y. Nakashima, K. Ito, T. Ikegami, I. Aritome, A. Suzuki and Y. Hisaeda, *Org. Lett.*, 2003, **5**, 2845–2848.
- 17 C. J. Kingsbury and M. O. Senge, SCSD models table, 2022, can be found under [https://www.kingsbury.id.au/scsd\\_models\\_table](https://www.kingsbury.id.au/scsd_models_table), accessed October 2023.
- 18 J. C. Gower, *Biometrika*, 1966, **53**, 325–338.
- 19 R. J. Largent, W. F. Polik and J. R. Schmidt, *J. Comput. Chem.*, 2012, **33**, 1637–1642.
- 20 T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing and Jupyter Development Team, in *Positioning and Power in Academic Publishing: Players, Agents, and Agendas*, ed. F. Loizides and B. Schmidt, IOS Press, Amsterdam, 2016, pp. 87–90.
- 21 M. O. Senge, *Chem. Commun.*, 2006, 243–256.
- 22 C. J. Kingsbury and M. O. Senge, Symmetric-Coordinate Structural Decomposition (SCSD) - an online tool for molecular analysis, Dublin, 2021, see <https://www.kingsbury.id.au/scsd>, accessed 4 July 2024.
- 23 M. Kielmann and M. O. Senge, *Angew. Chem., Int. Ed.*, 2019, **58**, 418–441.
- 24 (a) M. Roucan, M. Kielmann, S. J. Connon, S. S. R. Bernhard and M. O. Senge, *Chem. Commun.*, 2018, **54**, 26–29; (b) K. Norvaiša, K. J. Flanagan, D. Gibbons and M. O. Senge, *Angew. Chem., Int. Ed.*, 2019, **58**, 16553–16557.
- 25 C. R. Groom, I. J. Bruno, M. O. Lightfoot and S. C. Ward, *Acta Crystallogr.*, 2016, **B72**, 171–179.
- 26 (a) E. F. Meyer Jr, *Acta Crystallogr.*, 1972, **B28**, 2162–2167; (b) D. L. Cullen and E. F. Meyer Jr, *J. Am. Chem. Soc.*, 1974, **96**, 2095–2102; (c) T. D. Brennan, W. R. Scheidt and J. A. Shelnut, *J. Am. Chem. Soc.*, 1988, **110**, 3919–3924.
- 27 (a) J. A. Shelnut, X.-Z. Song, J.-G. Ma, S.-L. Jia, W. Jentzen and C. J. Medforth, *Chem. Soc. Rev.*, 1998, **27**, 31–42; (b) A. B. Graves, M. T. Graves and M. D. Liptak, *J. Phys. Chem. B*, 2016, **120**, 3844–3853; (c) M. O. Senge, S. A. MacGowan and J. M. O'Brien, *Chem. Commun.*, 2015, **51**, 17031–17063; (d) C. J. Kingsbury, K. J. Flanagan, H.-G. Eckhard, M. Kielmann and M. O. Senge, *Molecules*, 2020, **25**, 3195; (e) C. J. Kingsbury, M. Kielmann, B. Twamley and M. O. Senge, *Molecules*, 2022, **27**, 7060.
- 28 (a) J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, *npj Comput. Mater.*, 2019, **5**, 83; (b) K. Ryan, J. Lengyel and M. Shatruk, *J. Am. Chem. Soc.*, 2018, **140**, 10158–10168.
- 29 M. O. Senge, M. W. Renner, W. W. Kalisch and J. Fajer, *J. Chem. Soc., Dalton Trans.*, 2000, 381–385.
- 30 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.
- 31 A. A. Buglak, A. Charisiadis, A. Sheehan, C. J. Kingsbury, M. O. Senge and M. A. Filatov, *Chem.–Eur. J.*, 2021, **27**, 9934–9947.
- 32 X. Qiao, M. A. Padula, D. M. Ho, N. J. Vogelaar, C. E. Schutt and R. A. Pascal, *J. Am. Chem. Soc.*, 1996, **118**, 741–745.
- 33 (a) E. Trzop and I. Turowska-Tyrk, *Acta Crystallogr.*, 2008, **B64**, 375–382; (b) C. J. Brown and M. Ehrenberg, *Acta Crystallogr.*, 1984, **C40**, 1059–1060; (c) I. I. Schuster, L. Craciun, D. M. Ho and R. A. Pascal, *Tetrahedron*, 2002, **58**, 8875–8882.
- 34 C.-Y. Chiu, B. Kim, A. A. Gorodetsky, W. Sattler, S. Wie, A. Sattler, M. Steigerwald and C. Nuckolls, *Chem. Sci.*, 2011, **2**, 1480–1486.
- 35 P. L. Anelli, P. R. Ashton, R. Ballardini, V. Balzani, M. Delgado, M. T. Gandolfi, T. T. Goodnow, A. E. Kaifer, D. Philip, M. Pietraszkiewicz, L. Prodi, M. V. Reddington, A. M. Z. Slawin, N. Spencer, J. F. Stoddart, C. Vincent and D. J. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 193–218.
- 36 (a) R. Behrend, E. Meyer and F. Rusche, *Liebigs Ann. Chem.*, 1905, **339**, 1–40; (b) W. A. Freeman, W. K. Mock and N.-Y. Shih, *J. Am. Chem. Soc.*, 1981, **103**, 7367–7369; (c) R. Aav and K. A. Mishra, *Symmetry*, 2018, **10**, 98.
- 37 (a) R. Jasti, J. Bhattacharjee, J. B. Neaton and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2008, **130**, 17646–17647; (b) S. Kammermeier, P. G. Jones and R. Herges, *Angew. Chem., Int. Ed.*, 1996, **35**, 2669–2671; (c) Y. Segawa, A. Yagi, K. Matsui and K. Itami, *Angew. Chem., Int. Ed.*, 2016, **55**, 5136–5158; (d) J. Xia and R. Jasti, *Angew. Chem., Int. Ed.*, 2012, **51**, 2474–2476; (e) T. Iwamoto, E. Kayahara, N. Yasuda, T. Suzuki and S. Yamago, *Angew. Chem., Int. Ed.*, 2014, **53**, 6430–6434.
- 38 (a) P. Singh, A. Hirsch and S. Kumar, *Trends Anal. Chem.*, 2021, **136**, 116237; (b) W. Jiang and Z. Wang, *J. Am. Chem. Soc.*, 2022, **144**, 14976–14991; (c) E. M. Sánchez-Garnerero, A. R. Agarrabeitia, F. Moreno, B. L. Maroto, G. Muller, M. J. Ortiz and S. de la Moya, *Chem.–Eur. J.*, 2015, **21**, 13488–13500; (d) S. E. Penty, M. A. Zwijnenburg, G. R. F. Orton, P. Stachelek, R. Pal, Y. Xie, S. L. Griffin and T. A. Barendt, *J. Am. Chem. Soc.*, 2022, **144**, 12290–12298.
- 39 (a) L. Arnold, H. Norouzi-Arasi, M. Wagner, V. Enkelmann and K. Müllen, *Chem. Commun.*, 2011, **47**, 970–972; (b) J. F. Woods, L. Gallego, P. Pfister, M. Maaloum, A. V. Jentzsch and M. Rickhaus, *Nat. Commun.*, 2022, **13**, 3681.
- 40 (a) J. L. Sessler, S. J. Weghorn, Y. Hisaeda and V. Lynch, *Chem.–Eur. J.*, 1995, **1**, 56–67; (b) S. Hannah, D. Seidel, J. L. Sessler and V. Lynch, *Inorg. Chim. Acta*, 2001, **317**, 211–217; (c) S. J. Weghorn, J. L. Sessler, V. Lynch, T. F. Baumann and J. W. Sibert, *Inorg. Chem.*, 1996, **35**, 1089–1090; (d) Y. Ishimaru, H. Shiba and T. Fujihara, *X-Ray Struct. Anal. Online*, 2014, **30**, 33–34; (e) J. T. Brewster II, D. N. Mangel, A. J. Gaunt, D. P. Saunders, H. Zafar, V. M. Lynch, M. A. Boreen, M. E. Garner, C. A. P. Goodwin, N. S. Settineri, J. Arnold and J. L. Sessler, *J. Am. Chem. Soc.*, 2019, **141**, 17867–17874.
- 41 (a) L. K. Frensch, K. Pröpper, M. John, S. Demeshko, C. Brückner and F. Meyer, *Angew. Chem., Int. Ed.*, 2011, **50**, 1420–1424; (b) L. K. Blusch, K. E. Craigo, V. Martin-Diaconescu, A. B. McQuarters, E. Bill, S. Dechert, S. DeBeer, N. Lehnert and F. Meyer, *J. Am. Chem. Soc.*, 2013, **135**, 13892–13899; (c) A. Vogel, S. Dechert, C. Brückner and F. Meyer, *Inorg. Chem.*, 2017, **56**, 2221–2232.
- 42 T. Furuyama, K. Satoh, T. Kushiya and N. Kobayashi, *J. Am. Chem. Soc.*, 2014, **136**, 765–776.



- 43 A. Tsuda, H. Furuta and A. Osuka, *J. Am. Chem. Soc.*, 2001, **123**, 10304–10321.
- 44 (a) C. Di Natale, C. P. Gros and R. Paolesse, *Chem. Soc. Rev.*, 2022, **51**, 1277–1335; (b) S. Bhowmik, M. Kosa, A. Mizrahi, N. Fridman, M. Saphier, A. Stanger and Z. Gros, *Inorg. Chem.*, 2017, **56**, 2287–2296.
- 45 C. M. Lemon, A. J. Nissley, N. R. Latorraca, E. C. Wittenborn and M. A. Marletta, *RSC Chem. Biol.*, 2022, **3**, 571–581.
- 46 (a) K. S. Unikela, B. L. Merner, P. G. Ghasemabadi, C. C. Warford, C. S. Qiu, L. N. Dawe, Y. Zhao and G. J. Bodwell, *Eur. J. Org. Chem.*, 2019, 4546–4560; (b) K. S. Unikela, P. G. Ghasemabadi, V. Houska, L. N. Dawe, Y. Zhao and G. J. Bodwell, *Chem.–Eur. J.*, 2021, **27**, 390–400; (c) B. L. Merner, K. S. Unikela, L. N. Dawe, D. W. Thompson and G. J. Bodwell, *Chem. Commun.*, 2013, **49**, 5930–5932; (d) B. L. Merner, L. N. Dawe and G. J. Bodwell, *Angew. Chem., Int. Ed.*, 2009, **48**, 5487–5491.
- 47 (a) H. Tao, T. K. Allison, T. W. Wright, A. M. Stooke, C. Khurmi, J. van Tilborg, Y. Liu, R. W. Falcone, A. Belkacem and T. J. Martinez, *J. Chem. Phys.*, 2011, **134**, 244306; (b) I. I. Vorontsov, T. Graber, A. Y. Kovalevsky, I. V. Novozhilova, M. Gembicky, Y.-S. Chen and P. Coppens, *J. Am. Chem. Soc.*, 2009, **131**, 6566–6573; (c) P. Coppens, J. Sokolov, E. Trzop, A. Makal and Y. Chen, *J. Phys. Chem. Lett.*, 2013, **4**, 579–582; (d) Y. Nobuhiro, U. Hidehiro and O. Yuji, *Bull. Chem. Soc. Jpn.*, 2004, **77**, 933–944.
- 48 (a) J. Oddershede and S. Larson, *J. Phys. Chem. A*, 2004, **108**, 1057–1063; (b) V. P. Andreev, P. S. Sobolev and V. A. Tafeenko, *Russ. J. Gen. Chem.*, 2017, **87**, 1572–1579; (c) A. J. Golder, D. C. Povey, J. Silver and Q. A. A. Jassim, *Acta Crystallogr.*, 1990, **C46**, 1210–1212; (d) Y.-C. Hsieh, T.-C. Wu, J.-Y. Li, Y.-T. Chen, M.-Y. Kuo, P.-T. Chou and Y.-T. Wu, *Org. Lett.*, 2016, **18**, 1868–1871.
- 49 C. J. Kingsbury and M. O. Senge, SCSD Models, 2021, can be found under [https://www.kingsbury.id.au/scsd\\_models\\_table](https://www.kingsbury.id.au/scsd_models_table), accessed 4 July 2024.
- 50 (a) S. R. Kidd, W. Zhou, J. J. Warren and D. B. Leznoff, *Dalton Trans.*, 2024, **53**, 6537–6546; (b) C. Hunt, M. Peterson, C. Anderson, T. Chang, G. Wu, S. Scheiner and G. Ménard, *J. Am. Chem. Soc.*, 2019, **141**, 2604–2613.
- 51 C. J. Kingsbury and M. O. Senge, scsd – a tool for symmetry-coordinate structural decomposition analysis, 2024, can be found under <https://github.com/cjkingsbury/scsd/>, accessed 4th June 2024.
- 52 S. Van der Walt and M. Aivazis, *Comput. Sci. Eng.*, 2011, **13**, 22–30.
- 53 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 99–104.
- 54 Plotly Technologies Inc., *Plotly, Collaborative Data Science*, Montreal, Canada, 2015.
- 55 M. Grinberg, *Flask Web Development: Developing Web Applications with Python*, O'Reilly Media Inc., Sebastopol, USA, 2018.
- 56 A. J. Hanson, *Acta Crystallogr.*, 2020, **A76**, 432–457.
- 57 H. W. Kuhn, *Nav. Res. Logist. Q.*, 1955, **2**, 83–97.

