



UNIVERSITAT<sub>DE</sub>  
BARCELONA

**Cosymlib: a Python library  
for continuous symmetry measures and its application  
to problems in structural chemistry**

Efrem Bernuz Fito



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution 4.0. Spain License.**



UNIVERSITAT<sup>DE</sup>  
BARCELONA

FACULTAT DE QUÍMICA

PROGRAMA DE DOCTORAT DE QUÍMICA TEÒRICA I MODELITZACIÓ  
COMPUTACIONAL

# Cosymlib: a Python library for continuous symmetry measures and its application to problems in structural chemistry

Tesi doctoral presentada per l'**Efrem Bernuz Fito** per optar al títol de doctor de la  
universitat de Barcelona.

Director i tutor:

Pedro Alemany Cahner  
(Universitat de Barcelona)

Signatura del doctorand:

**Efrem Bernuz Fito**



## Agraïments

Aquesta secció esta dedicada a totes les persones que han fet possible el trajecte fins a arribar a la conclusió d'aquesta tesi i que tot i que son moltes persones i alguna altra em deixaré, faré el millor possible per deixar-hi un record en aquesta tesi.

Primer li vull dedicar a la meva família que tant en bons i mals moments han estat des del principi aquí i m'han ajudat no sols a aprendre, sinó a créixer i motivar-me per les coses que m'han agradat al llarg de la meva vida. Els hi podria dedicar molts moments, però especialment els hi vull agrair el moment en el qual vaig mudar-me a Barcelona i vaig començar els meus estudis del grau de Química, que m'animessin a que seguís i no defallís en un moment prou complicat per a mi.

Agrair-li als meus avis que han estat sempre per allò que necessites fins i tot en els moments en que ells necessitaven més ajuda.

També a tota aquella gent que ha estat durant el transcurs d'aquesta tesi aportant el seus coneixements, ensenyant-me i orientant-me en tot aquest procés d'aprenentatge. Al meu director de tesi Pere, t'agraeixo la teva paciència i la teva genialitat per veure coses que en el seu moment jo era incapaç de veure. Al Miquel per ajudar-me a entendre conceptes de la estructura molecular amb tanta facilitat. A l'Abel per ensenyar-me tots aquells conceptes sobre programació que desconeixia i fer-me un millor programador. També per acollir-me tant a Kyoto com a Donostia en les meves estades doctorals.

No voldria passar per alt als meus amics de Ponts, que des de que vaig mudar-me allà amb 10 anys han estat sempre i m'han donat moltes alegries, malgrat que en els últims anys haguem vist menys degut a les distancies físiques que ens separen.

Una part important de la meva vida, actualment la meitat d'aquesta l'he viscut en l'entorn universitari a Barcelona i per això no puc deixar d'agrair-li agrair-li a tota aquella

gent que m'ha acompanyat durant aquesta etapa. El Jordi, l'Enric, l'Adrià, l'Alex, etc. que entre partida i partida de rol, alguna cerveseta que altra, les nostres mítiques barbacoes i moltes altres coses m'han ajudat a fer aquí camí molt més planer.

A la gent del màster, d'on hi van sortir algunes amistats i la meva actual parella. A la Mirel·la Díaz, l'olox, l'ajolote i l'Alba menja patates. Tot i que el doctorat se'ns ha fet difícil puc dir que al final tots l'hem superat i aconseguit arribar on volíem estar.

Als castellers del Poble Sec que m'han fet descobrir a una tardana època una de les aficions més maques i que he gaudit molt durant els darrers 5 anys. Es fa difícil explicar amb paraules tots els moments d'emocions plegats que hem viscut, ni nomenar a la gent que ho ha fet possible (bàsicament perquè sou més d'un centenar i he d'alleugerir), però per tot gràcies.

M'agradaria també fer especial incís en la gent que m'ha ajudat a alleugerir el pes de la tesi. Als Tonis, el Guty, la Vera, la Claudia i la Montse, pels seus dinars en família, fent que no tingues ni una, ni dues, sinó tres famílies.

També voldria agrair altra gent de Poble sec moments inoblidables. A la Clara A., la Clara M., la LP, la Mimi, l'Arnau, el Toni, el Dani, l'Elena, la Maria.

També a la Irene, per ajudar-me a desconstruir-me una mica més i ensenyar-me que hi ha altres visions de la vida.

I finalment, agrair-li especialment a l'Alba. Perquè no només ha patit les dificultats d'una tesi en persona, sinó que també m'ha ajudat a portar la meva amb més tranquil·litat. Ningú com ella pot entendre el dia a dia del que significa una tesi, i per això no li podré agrair prou tot el que ha fet per mi en aquests darrers 5 anys, per moltes pàgines que aquí escrigui. GRACIES.



## Summary

In modern chemistry, any property or transformation of a substance may be traced back to the interactions that atoms, and more specifically their nuclei and electrons, undergo when approaching each other. In this sense, the molecular structure, that is the spatial arrangement of atoms within a molecule, is of the utmost importance to understand the properties of a molecule and the ways in which this molecule interacts with its surrounding. It has been demonstrated that a set of atoms, when forming a given molecule, adopt a preferred spatial distribution which leads to the concept of molecular shape, and that this shape depends normally on its close chemical environment and on external conditions such temperature and pressure.

For many years symmetry has been a useful concept in the study of the spatial organization of atoms in molecules or solids. The presence of symmetry elements in a given molecular structure gives a valuable information about its properties and chemical behaviour. However, it has been demonstrated that most molecules in nature tend to adopt shapes which are not fully symmetric, presenting small distortions from the ideal symmetric model structures used to rationalize the stereochemical knowledge. Thus, the quantification of the asymmetry of a given molecular structure, or in other words, the quantification of the departure of a structure from a given reference shape or symmetry, has been a quite fruitful approach to analyse the changes in properties observed for a set of related molecules. In this sense, the classic "black or white" point of view that considers if a given structure is symmetric or not, may become obsolete and replaced by a more quantitative approach where the presence or absence of symmetry has been replaced by a continuous scale of grey shades.

Continuous symmetry measures (CSMs) were developed precisely to quantify the amount of asymmetry of a given object by comparing a distorted structure with an ideal symmetric reference. This methodology has been very helpful to classify, for example, the shape of the coordination environment of transition metal atoms in several coordination complexes.

Additionally, CSMs can be extended to the analysis of the presence / absence of symmetry in the electronic structure of molecules. In this regard it is now possible to perform a pseudosymmetry analysis, that is, to classify the molecular orbitals in a molecule, not with regard of the irreducible representations of the actual symmetry group of the molecule, but using instead a different pseudosymmetry group with a higher degree of symmetry. The result is the mixing of irreducible representations of the pseudosymmetry group that leads to the lower symmetry of the actual molecular orbitals, an information that may be fruitfully used to relate the properties of related similar molecules that exhibit, however, different symmetries.

In order to apply the formalism of CSMs to stereochemical problems, it is however necessary to use computer programs to calculate first the CSMs, which are expensive to compute, especially for large structures. A few efficient algorithms have been developed in the past separately for computing different types of CSMs that apply to different types of objects of interest in chemistry such as the shape of a molecule described either by the geometrical arrangement of its nuclei or by its electron density. However, the lack of a unified computational software to calculate different CSMs for a same molecule using a common format to treat both the geometrical symmetry and that of the electronic structure has in the past prevented the interested users to familiarize with the possibilities of applying an integrated continuous symmetry formalism to stereochemical problems.

In the present thesis, we present an overview of the formalism of CSMs, describing the computational methodologies that had been developed in the past. The main aim of the present thesis is the development of *Cosymlib*, a Python library englobing all previous algorithms within a unified computational framework that allows a seamless computation of different CSMs for a given molecule using a unified format. Extensive discussion of the advantage of implementing modern programming techniques such as object-oriented programming in the development of a unified computational approach to CSMs will be given in a second methodological chapter. Afterwards, the use of the different tools included in *Cosymlib* for the symmetry analysis of the molecular structure will be illustrated by applying it to different stereochemical problems related to organometallic coordination complexes, the effect of temperature on the shape of

several polyhedral cage molecules and, the effect of temperature and the crystal environment on the shape of phosphate anions.

Specifically, this thesis is organized in chapters as follows:

1. **The mathematical formalism of Continuous Symmetry Measures**
2. ***Cosymlib*: a Python library for the computation of Continuous Symmetry Measures**
3. **Shape and symmetry of organometallic piano-stool compounds**
4. **Effect of temperature on the shape and symmetry of molecules and solids**
5. **Effect of temperature and the crystal environment on the shape of phosphate ions**



# Index

<b>Chapter 1. The mathematical formalism of Continuous Symmetry Measures .....</b>	<b>1</b>
Introduction .....	1
Shape and symmetry.....	3
Continuous shape measures .....	5
Methodology.....	6
Shape Maps.....	7
Continuous symmetry measures .....	11
Methodology.....	12
Symmetry operation measures .....	12
Methodology.....	13
Algorithms for obtaining a CShM, CSM, or SOM .....	14
Pseudosymmetry analysis for the electronic structure .....	16
Methodology.....	16
Algorithm .....	21
<b>Objectives.....</b>	<b>25</b>
<b>Chapter 2. CosymLib: a Python library for the computation of Continuous Symmetry Measures .....</b>	<b>27</b>
Introduction .....	27
Object-oriented programming.....	30
Why Python? .....	33
CosymLib.....	37
Code development.....	39
CosymLib classes.....	42
CosymLib class .....	44
Molecule class.....	44
Geometry class.....	45
Electronic structure class .....	45

---

Shape class .....	46
Symmetry class.....	47
Pointgroup class .....	47
Simulation class.....	47
Beyond CosymLib.....	51
Symgroupy .....	52
Wfnsympy .....	53
Huckelpy.....	54
Future scope .....	56
Summary .....	57
<b>Chapter 3. Shape and symmetry of organometallic piano-stool compounds .....</b>	<b>61</b>
Introduction .....	61
Methodology.....	65
Shapes of piano-stool compounds in crystals .....	68
Computational models.....	71
Electronic structure.....	80
Td/Oh pseudosymmetry analysis .....	98
Conclusions .....	103
Appendix .....	104
<b>Chapter 4. Effect of temperature on the shape and symmetry of</b>	
<b>molecules and solids .....</b>	<b>109</b>
Introduction .....	109
Methodology.....	110
P <sub>4</sub> molecule .....	112
Ethane .....	118
Polyhedral molecules.....	123
Tetrahedral solids .....	127
Conclusions .....	132

---

<b>Chapter 5. Effect of temperature and the crystal environment on the shape of phosphate ions .....</b>	<b>135</b>
Introduction .....	135
Molecular simulations for isolated phosphate anions .....	137
CSM analysis for experimental phosphate crystal structures .....	147
Effect of the temperature and the environment on the symmetry of phosphate anions: Molecular dynamics simulations for crystal structures.....	161
Pseudosymmetry analysis.....	171
Conclusions .....	177
Appendix .....	178



# Chapter 1. The mathematical formalism of Continuous Symmetry Measures

## Introduction

Symmetry has been employed for centuries in science as a tool to rationalize different aspects of our world, where complex patterns arise in a natural way from apparently simple causes. In the ancient era, symmetry was associated to the ideas of “beauty” or “perfection”, and the early civilizations used these words in order to make reference to highly symmetrical objects. When they talked about beauty, they were referring to something that is proportioned or balanced in physical space, a definition that can be more or less easily translated into a mathematical description of the aspect of objects. Probably, Plato was one of the first to use the concept of beauty to describe highly symmetrical objects, such as the famous Platonic solids<sup>1</sup>, polyhedra that are composed by regular polygons in their faces and having equal angles in all of their vertices.

The ancient concept of balance can be related to the modern concept of bilateral symmetry, that describes how two points in space are related with respect to a reflection plane<sup>2</sup>. However, reflection, as opposed to beauty, is a precise geometrical descriptor which can be obtained by a mathematical operation. Furthermore, aside from reflection, other mathematical transformations related to symmetry such as rotations, translations, etc. were introduced, and nowadays all of them belong to the modern mathematical theory of symmetry, based on the abstract algebraic concept of symmetry groups. In chemistry, physics, and biology, the study of symmetry has been applied mainly using group theory, analysing whether an object belongs to a given symmetry group by checking if all the transformations included in the group leave the object invariant or not. However, this modern concept of symmetry based on a binary interpretation that is only able to tell us if something is symmetric or not, entails some difficulties in the description of symmetry in natural objects which usually do not

exactly belong to highly symmetric groups, although presenting a high degree of regularity or, in other words, which are only approximately symmetric. For many years scientists have developed highly symmetric ideal models, such as the tetrahedral carbon atom in chemistry, which are considered to be equally valid for real quasi-symmetric objects that present some degree of distortion with regard to the ideal model. The general validity found when describing low symmetry objects with theories based on symmetric models implies that symmetry may be described with something more than just a "yes or no" spectrum and the search for a way to quantify the amount of symmetry, or in other words, how far an object is from the ideal symmetry, has been discussed in several studies<sup>3-8</sup>.

One such approach is based on the description of the distortions of a molecule with respect to an idealized structure by using the symmetry of normal modes, for example to describe  $\text{MX}_4$  distorted structures by means of the normal modes for an ideal  $\text{MX}_4$  structure with tetrahedral symmetry<sup>9</sup>. Another example is the quantification of chirality for a set of different enantiomers<sup>10</sup>, that has been claimed to offer a generalized description of this property. A common feature in all these approaches is the introduction of a metric, that is, a distance which measures the departure of an object from a given ideal symmetry. In general, the distance of an object from this reference symmetry can be defined by how similar, or properly speaking, how dissimilar this object is from the closest object having the desired symmetry. It should be noted that although it is equally valid to ask for either the similarity or the dissimilarity with respect to the reference object, the two approaches lead to two different types of measures, where the similarity is equal to 1 (in a scale between 0 and 1) when two objects are identical while their dissimilarity is equal to 0, with both points of view giving us exactly the same information<sup>3,11</sup>.

In 1992 Avnir et al. reported a general procedure for the quantification of symmetry which is specially well suited for problems in stereochemistry. They developed what they called *Continuous Symmetry Measures* (CSM), a set of measures each quantifying how far an object is from having a particular symmetry. By definition, "The continuous symmetry measure quantifies the minimal distance movement that the points of an

object have to undergo in order to be transformed into a shape of the desired symmetry"<sup>12</sup>.

The modification of the usual binary concept of symmetry by replacing it by a continuous spectrum of symmetry content, gave a huge advantage in the comparison of molecular structures considered as merely geometrical objects. The mathematical framework of continuous symmetry is wide and can be easily extended to other geometrical properties aside from symmetry such as, for example, the shape of an object, that can be also described by a continuous measure, a *Continuous Shape Measure* (CShM) in this case. Moreover, the concept of a CSM has been lately generalized to other more abstract objects of interest in stereochemistry such as the electronic structure of a molecule that can also be studied using CSMs which have been specifically developed to describe not geometrical objects made from a discrete set of points but functions such as the electron density or the molecular orbitals or even other more abstract mathematical objects such as quantum mechanical operators like the Hamiltonian.

In this chapter the different methodologies developed in our group that were implemented in the *Cosymlib* library will be reviewed in some detail. This includes their definition and mathematical description, as well as the algorithms used in the computational programs that implement them. For a more detailed description of the mathematical framework of CSMs the reader is invited to consult the primary bibliography<sup>12-17</sup>.

## Shape and symmetry

Before giving an actual mathematical description of what continuous symmetry measures are, it is necessary to explain the differences between symmetry and shape, as these two concepts will appear frequently all along the thesis and in some cases confusion may arise because often the same labels may be used for the two descriptors in cases where shape and symmetry are equivalent. To clarify the differences, one may focus on the example in figure 1, where three Archimedean polyhedra with 24 vertices

are shown. Just from a simple eyesight it is possible to see that the three polyhedra present distinct arrangements of vertices and faces, which correspond to three qualitatively distinct shapes for 24-vertices polyhedra: the truncated cube, the truncated octahedron, and the rhombicuboctahedron. However, a closer inspection shows that all three have exactly the same symmetry, in other words, all three belong to the same symmetry group, the octahedral symmetry group  $O_h$  with 48 different unique symmetry operations in this case. This example shows that, the shape of an object is a more restrictive property than its symmetry, since a set of qualitatively different polyhedra may all have the same symmetry, but a given shape is only related to one unique polyhedron. An exception to this situation is, however, found for the highly symmetric platonic solids, where symmetry and shape are strictly equivalent. For example, this is the case for the tetrahedron, because is the only possible four-vertex polyhedron exhibiting tetrahedral symmetry, and there is no possible choice of different shapes compatible with this symmetry. So, the closest  $T_d$ -symmetric object for any polyhedron with four vertices will be a tetrahedron and the shape and symmetry measures will be exactly equivalent.

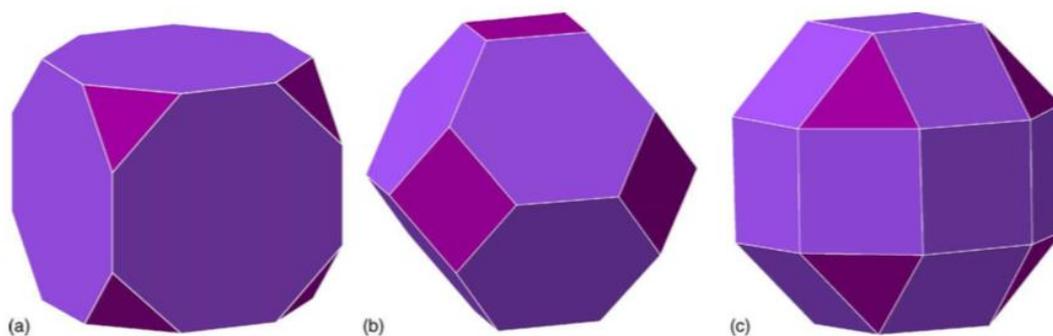


Figure 1. Three 24-vertices Archimedean polyhedra with the same symmetry ( $O_h$ ), but different shapes: a) the truncated cube, b) the truncated octahedron, and c) the rhombicuboctahedron.

## Continuous shape measures

Any object has a certain shape, which distorted or not, is usually defined by comparing it with a reference object, such as a square, a circle, a sphere, etc. Note that although usually we use highly symmetric objects as references, this is not strictly necessary, and any object may be used to define a shape. These relationships are established because humans tend to classify any object in nature with a certain criterion in mind. In science, in order to avoid subjective interpretations, the classification has been unified to normalize the rules that should be followed, for example, to describe the “shape” of the arrangement of atoms in a solid, or the organization of ligands around a metal ion.

A continuous shape measure (CShM) is a tool to describe the spatial organization in chemical structures (molecules or groups of atoms) and to find if it is possible to classify them into any particular shape. In a nutshell, a CShM consists in the quantification of the dissimilarity of a structure A (described by a set of points in space) with respect to a reference structure B, that should share the same number of vertices with A. In other words, it is a measure of how far A is from B. The reference structure can be either a regular polyhedron such as the octahedron or the tetrahedron, but it is not restricted to symmetric polyhedrons and it is possible to use any general, non-symmetrical object as a reference as far as it has the same number of vertexes as the problem structure. The classification of molecules according to their geometrical shapes has proven to be very useful to categorize structural features of molecules such as the geometry of the coordination environment of transition metal atoms in coordination complexes. The cases of  $ML_n$  ( $n=4-10, 12, 20$ ) coordination compounds have been analysed in detail<sup>18-29</sup> and the preference for different shapes has been reported, finding in many cases a strong correlation between the shape and the metal d-electron count as shown later in chapter 3 for piano-stool complexes. A general conclusion derived from these studies is that the metal coordination environments in most coordination compounds present a more or less evident deviation from their ideal geometries. In fact, most of the compounds can have shapes that do not exactly coincide with those of highly symmetrical models, but rather lie along a distortion pathway between a pair of reference structures, such as for example, the tetrahedron and the square<sup>30</sup>.

## Methodology

Mathematically, the distance between a problem and a reference structure is calculated as shown in figure 2, by minimizing the sum of squared distances of the vertices of a structure Q, defined by N points in the space with position vectors  $\{\mathbf{q}_k\}$ , from those of a reference polyhedron P with position vectors  $\{\mathbf{p}_k\}$ ,

$$S_Q(P) = \min \frac{\sum_{k=1}^N |\mathbf{q}_k - \mathbf{p}_k|^2}{\sum_{k=1}^N |\mathbf{q}_k - \mathbf{q}_0|^2} \times 100 \quad (1)$$

where  $\mathbf{q}_0$  is the coordinate vector of the geometrical centre of structure Q.  $S_Q(P)$  is the continuous shape measure of structure Q with respect to the reference shape P. Note that in the following, if there is no possible confusion about the problem and the reference structures, to avoid cumbersome subscripts we will use the simplified S(P) notation to indicate the shape measure with respect, reference structure P for any given, unspecified structure.

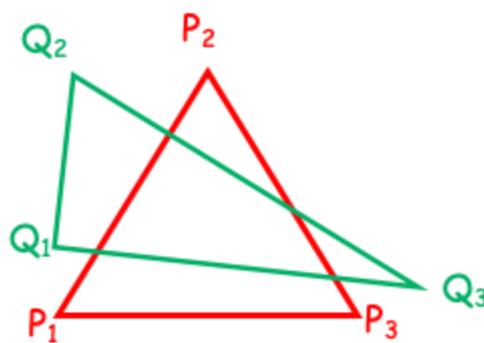


Figure 2. Comparison between a problem structure Q (green) and with a reference structure P (red). The CShM is evaluated using equation (1) for the relative orientation/scaling of the two structures that minimizes the distance between the  $\mathbf{q}_k$  and  $\mathbf{p}_k$  points.

The minimization of equation (1) should consider that two objects with different size, position and/or orientation may have the same shape since shape is invariant to translations, rotations and scaling. Thus, to compute the CShM one should translate the objects to a common position and change their relative orientation and size to minimize the distance between their vertices, taking into account the pairing of vertices (or permutation) that minimizes the measure. From the expression in equation (1) it becomes apparent that the minimization should lead to a measure in the bounds of  $0 \leq S_Q(P) \leq 100$  where  $S_Q(P) = 0$  means that the structure Q has the same shape as the reference structure P, while an increase of this value means a distancing from the reference shape.

In order to use CShMs in stereochemistry we need first to establish what can be considered as a chemically relevant distortion in the continuous shape spectrum. Although a CShM can vary between 0 to 100, almost all the structures analysed so far do not have shape values above 50 which correspond to physically unfeasible structures where, for instance, two or more vertices collapse on a same point in space. Experience shows also that measures between 0.1 and 3 are sufficiently relevant from a chemical point, where values greater than 3 are considered to point towards highly distorted structures.

## Shape Maps

Sometimes, especially when dealing with a large family of compounds, a single CShM is not sufficient to describe their shape, as they usually present structures that are intermediate between a pair of reference polyhedra. To synthesize this information, shape maps were developed as a tool to represent two CShM values for a single structure in a two-dimensional plot<sup>15</sup>. A structure A might be interconverted into another structure B by the successive application of different distortions. The path that joins a structure A with a structure B such that all intermediate structures X between A and B have been obtained by distortions that preserve the maximal shape with respect to A is called the minimal distortion path from A to B.

An example of a shape map is shown in figure 3, where the minimal path for the interconversion between a tetrahedral and a square planar  $ML_4$  structure is represented as the curve on the lowest part of the diagram where the different structures are displayed as black dots according to their tetrahedral and square planar CShMs. For any other continuous transformation between A and B the line will lie above the minimal distortion path in the corresponding shape map. Note also that since  $S_A(B) = S_B(A)$  by definition, the minimal distortion path is symmetrical around the  $S_X(A) = S_X(B)$  line.

In a shape map each axis corresponds to a single reference and the lower limit for geometrically feasible structures is given by the minimal distortion path. An example of the utility of shape maps can be found in the work of Cirera et al.<sup>18</sup>, where they prove that the majority of tetracoordinate transition metal complexes can be identified to have structures with shapes intermediate between tetrahedral (T-4) and square planar (SP-4). Moreover, they showed a way to identify trends on the distortion paths that can be shown in a shape map. It should be noted that from now on, the labels of the shape references will be used for continuous shape measures, while the symmetry group labels will be only utilized for continuous symmetry measures that will be introduced later. For example, for the tetrahedral shape measure, the label **T-4** will be used while, the  $T_d$  label will be reserved for the tetrahedral symmetry measure. Although in the specific case of the tetrahedron this may be somewhat misleading because both measures are equivalent, for other cases where symmetry and shape do not necessarily coincide, the distinction between shape and symmetry measures is important and we will adhere to this notation even in cases such as the tetrahedron where it is not strictly necessary.

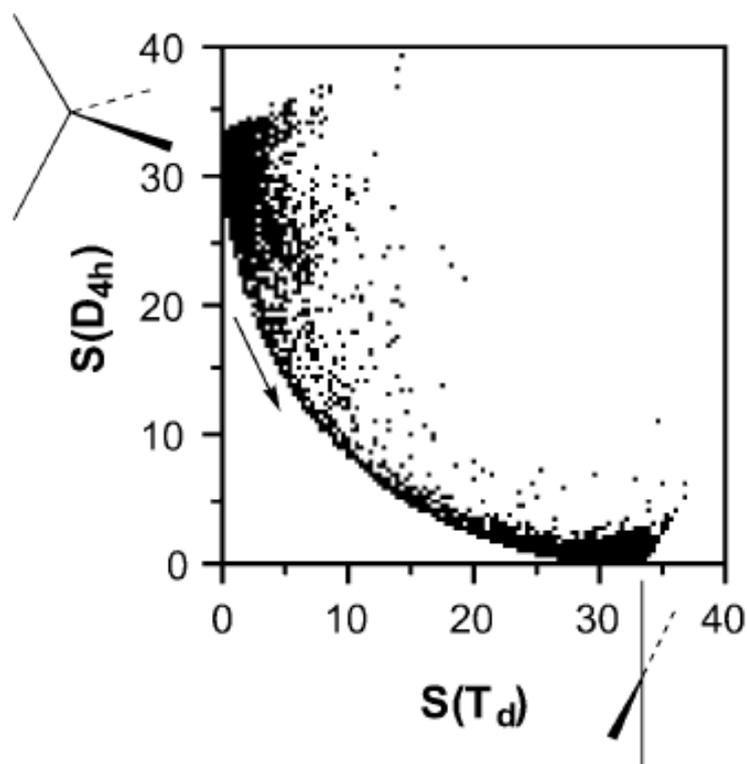


Figure 3. Shape map for the interconversion pathway of the tetrahedron ( $T_d$ ) and the square planar ( $D_{4h}$ ) shape structures. The black dots are the positions of the tetracoordinate transition metal complexes. Figure from David et al. work<sup>30</sup>.

The lack of experimental structures shown in the upper left and lower right limits in figure 3 is not uniquely characteristic of this kind of molecules as it can appear for any pair of  $n$ -vertices shapes. Thus, the limits of the minimal distortion interconversion pathways are intrinsic of the definition of CShMs and consequently common for all shape maps.

Sometimes, it may be useful to calculate the distance of a structure to the minimal distortion path through the path deviation function,

$$\Delta_i(P, T) = \frac{1}{\theta_{PT}} \left[ \arcsin \frac{\sqrt{S_i(P)}}{10} + \arcsin \frac{\sqrt{S_i(T)}}{10} \right] \quad (2)$$

where the symmetry angle  $\theta_{PT}$  is a constant for each pair of reference polyhedra, P and T, and must obey the following criteria for each structure X along the minimal distortion path<sup>30</sup>,

$$\arcsin \frac{\sqrt{S_X(P)}}{10} + \arcsin \frac{\sqrt{S_X(T)}}{10} = \theta_{PT} \quad (3)$$

Sometimes it is useful to plot other distortion paths so that multiple paths can be found in a single shape map. This is useful to categorize if a certain collection of structures show all a particular type of distortion or if they are in between different distortions. An example is shown in figure 4, where the previous minimum distortion path from figure 3, the tetrahedral to square planar ( $T_d - D_{4h}$ ) path, and other distortion paths like the  $C_3$  preserving umbrella distortion path, or the the off-axis and sawhorse paths are also displayed.

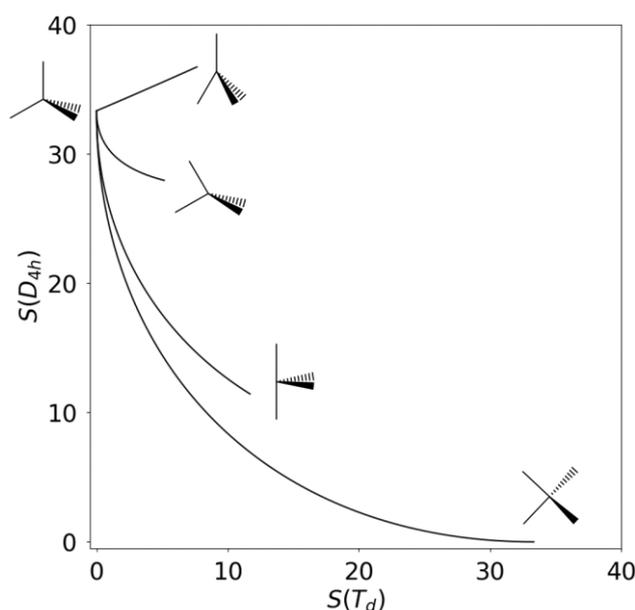


Figure 4. Distortion paths from a perfect tetrahedral  $ML_4$  structure to four different distorted structures. From top to bottom: umbrella, off-axis, sawhorse and spread distortions. The path for the spread distortion joining the tetrahedron and the square planar structure corresponds to the minimal distortion path in the  $S(T_d) - S(D_{4h})$  map.

## Continuous symmetry measures

Continuous symmetry measures (CSM) for an object  $Q$  can be defined following the same idea as in CShM, but taking as the reference structure  $P$ , the closest structure with the desired symmetry having the same number of vertices as  $Q$ . From the mathematical point of view, the definition of a CSM is analogous to that of a CShM<sup>13</sup>, and the way of computing it is also very similar since it implies a minimization of the distances between pairs of points. However, there is a fundamental difference between both types of measures. While in one case we compare the problem object  $Q$  with a known reference structure  $P$ , in the case of CSM the reference structure, that is, the closest structure with the desired symmetry, is in general unknown and must be found during the minimization process. In some particular cases such as the tetrahedral symmetry for a 4-vertex polyhedron, the reference structure is known a priori, because the tetrahedron is the only possible 4-vertex polyhedron with tetrahedral symmetry. This situation is, however, rather an exception and, in general, when computing a CSM we must not only determine the actual value of the measure but also the closest symmetric reference structure which is often not obvious.

For the measure of symmetry of a structure, different algorithms can be used to generate the reference symmetric structure. One of them is the folding – unfolding algorithm<sup>12</sup> which gives directly the  $G$ -symmetric reference structure and the CSM. The other one is by applying all symmetry operations of a group to a given object and averaging the result. This much simpler algorithm is the one that has been implemented in *Cosymlib*. Nevertheless, both algorithms lead to a minimization problem where one must find the optimal orientation of the symmetry elements of the group that result in the smallest value of the distance between the problem and the reference structures. The continuous symmetry measure corresponds by definition to this minimum distance.

Summarizing the main point in the definition of CSMs and CShMs, the feature that links both concepts is that in both cases the measure is evaluated as the distance of a

structure Q with respect to a reference structure P in a minimization procedure to find the matching of the two structures leading to the minimal distance..

### Methodology

As it has been explained before, the equation that defines the CSM is the same as for the CShM, a measure of the distance given by the squared differences between the points in the two objects, with the particularity that the reference structure P must belong to a given symmetry point group G,

$$S(Q, e) = \min \frac{\sum_{k=1}^N |\mathbf{q}_k - \mathbf{p}_k|^2}{\sum_{k=1}^N |\mathbf{q}_k - \mathbf{q}_0|^2} \times 100 \quad (4)$$

In this case, the minimization does not require to translate or resize the structures as they are not intrinsic of symmetry operations belonging to a point group, however it is convenient to place the geometric centre of the Q structure at the origin  $\mathbf{q}_0 = \mathbf{o}$ . Any symmetry operation will leave this centre invariant as, by definition, any symmetry element must pass through the geometrical centre of Q. Since this point is invariant, the centre of P must also be located precisely at this point and then only the orientation and the possible permutations between vertices are minimized to find the CSMs.

### Symmetry operation measures

As mentioned above a way to quantify the amount of symmetry in a certain structure is by using the so-called symmetry operation measures (SOMs), which is a slightly different approximation to the CSM than that in the folding-unfolding algorithm. In this case, the contents of a certain symmetry are measured by comparing the dissimilarity

between a structure and the same structure transformed by a single symmetry operation  $R$ , not the whole group. The advantage of using SOMs is that the reference structure for each symmetry operation is trivially calculated, just by operating on the initial structure with a given operation  $R$ . By combining the SOMs for all operations in a group (actually not all operations are always needed, only those associated with the generators of a group are necessary) it is possible to calculate both the CSM and the reference  $G$ -symmetric structure.

## Methodology

The expression that defines SOM is analogous to that for either the CSMs or CShMs, with the particularity that now the reference structure is  $RQ$ , the initial structure  $Q$  transformed by applying the symmetry operation  $R$ . The expression for a set of points  $\{\mathbf{q}_k\}$  is,

$$\Omega(Q, R) = \min \frac{\sum_{k=1}^N |\mathbf{q}_k - \hat{R}\mathbf{q}_k|^2}{4 \sum_{k=1}^N |\mathbf{q}_k - \mathbf{q}_0|^2} \times 100 \quad (5)$$

It can be demonstrated, that the factor 4 in the denominator must be included to compare SOMs with the previously introduced CSM values<sup>13</sup>. The rest of the denominator is the same normalization factor of the CSM and CShM that is related to the size of the structure.  $\Omega(Q, R)$  is minimized with respect to all possible orientations of the symmetry element associated to the operation  $\hat{R}$ , as well as to all possible pairings of vertices between the original structure  $Q$  and its transformed image  $\hat{R}Q$ .

## Algorithms for obtaining a CShM, CSM, or SOM

Either CSM, CShM and SOM, despite their differences, are based on the same methodology. For a set of  $N$  points with coordinates  $\mathbf{q}_k$  that forms the structure  $Q$ , the continuous measure is defined as the distance to a reference structure  $P$  defined by the set of points  $\mathbf{p}_k$ . The main difference between the different types of measures is in how this reference structure is defined. The calculation of CShMs requires a predefined reference structure  $P$  to be compared. This reference structure is arbitrary, it may be a well-known polyhedra or it can be any arrangement of points, for instance, the position of atoms in a molecule as obtained by minimizing its energy. When computing a CSM, the reference structure  $P$  will be obtained during the process of calculating the CSM as it is done in the folding-unfolding algorithm, that is not going to be explained in this thesis, but can be found in previous works<sup>12</sup>. The other way of obtaining a reference structure for a CSM is by using SOMs where a reference structure  $P_R$  is calculated for each symmetry operation in the group just by applying  $R$  to the structure  $Q$ . The reference structure for the CSM is obtained afterwards by averaging over all  $P_R$  structures preserving the relative orientations of all  $R$  operations in the group.

Once the reference structure is set, the basics of the algorithm are all the same and work as follow:

1. CShMs and CSMs for point-symmetry groups do not consider translation operations. Thus, the structure problem  $Q$  and the reference structure  $P$  must share the same geometric centre  $q_0$  which is located at the coordinate origin. Any symmetry element in the group must contain this geometric centre which is mapped into itself by all symmetry operations in the group.
2. Normalize  $Q$ . In principle the two structures,  $Q$  and  $P$ , may not have the same size and since by definition shape and symmetry are independent of the size of an object, it is advisable to normalize  $Q$  to give  $N$ . The normalization scales the

- size of the object  $Q$  so that the longest distance from the centre  $\mathbf{q}_0$  to any vertex of  $Q$  and  $P$  is one.
3. The distance between  $N$  and  $Q$  is calculated as the square of the distances between the points of  $N$  and  $P$  and minimized with respect to rotations and a scaling factor (eq. 1, 4, 5). This distance is, however, not the final measure since we must minimize also among all possible vertex pairings of  $N$  and  $P$ .
  4. For this minimization, a permutation operator  $\hat{P}$  is constructed. This operator applied to  $N$  generates all possible permutations of vertices for the normalized problem structure  $N$  and the distance is calculated as in step 3 for all permuted versions of  $N$ .
  5. The permutation that minimizes the distance gives the optimal pairing between  $Q$  and  $P$  and the distance calculated for this permutation is, by definition, the continuous shape or symmetry measure.

Although the algorithm may seem simple enough at a first look, it presents a huge computational obstacle in the minimization process, which is the bottleneck in this kind of calculations. The computational cost due to the minimization by comparing all permutations for pairs of coordinates means that the calculation scales with a factor of  $n!$  where  $n$  is the number of vertices in structure  $Q$ . This means that even though mathematically it is possible to compute a CShM or CSM for any possible structure described as a set of vertices, the time needed to find the optimal permutation may be prohibitively long. For large structures that present only small distortions from a known reference structure, computation of CShMs is only computationally feasible by fixing a permutation (or a small set of permutations) for the calculation, considering that one of these permutations is indeed the permutation that actually minimizes the measure. An example of this type of procedure can be found in calculation of the CSM for fullerene ( $C_{60}$ ) molecules done by Echeverría et al.<sup>21</sup>.

## Pseudosymmetry analysis for the electronic structure

In the previous sections, we have been discussing the different ways to analyse the geometry of any kind of structure, either with CShMs or CSMs. While these measures can be quite useful in order to classify, for example, a large collection of molecules, in quantum chemistry it may be useful to study the symmetry of the electronic structure of molecules to, for example, understand the nature of molecular orbitals. In chemistry, it has been particularly useful to analyse the orbital interactions between atoms and molecules as they can explain different chemical phenomena. The use of symmetry has been advantageous in the description of organic complexes in the classification of their molecular orbitals such as the  $\sigma$ - or  $\pi$ -orbitals. It has also been important to analyse the d-molecular orbitals of metal coordination complexes in terms of their tetrahedral or octahedral symmetry ( $e/t_2$  or  $e_g/t_{2g}$  orbitals). However, as is usually happens when molecular properties are described in terms of point-group theory, real molecules tend to present distorted versions of the idealized structures used in the theoretical models. One particular case is found in four-coordinate complexes, where the geometrical distortion induces a mixture of their d-orbitals which produces a change in their molecular orbital irreducible representations<sup>31</sup>. Therefore, a method was developed to quantify the symmetry of the wave function when the actual symmetry is lower than that of the idealized model: the pseudosymmetry analysis first developed by Casanova et al.<sup>17</sup>. This section will cover some details of the pseudosymmetry analysis and its implementation as the method will be used in the different studies that are in the following chapters.

### Methodology

If a molecule with  $N$  nuclei and  $n$  electrons is considered under the Born-Oppenheimer approximation, it can be proven that any geometrical transformation that leaves the

nuclear framework invariant belongs to a symmetry point group  $G$  in which case the structure is said to be  $G$ -symmetric or that it has  $G$  symmetry.

Any small displacement of its nuclei can remove some symmetry elements from  $G$ , thus the molecule becomes a lower symmetry object  $G'$  that, from group theoretical considerations, must be a subgroup of  $G$ , and in the worst case scenario become the non-symmetric  $C_1$  group with the identity as its only element. In the case of a small displacement of the  $G$ -symmetry, the  $G'$ -symmetric molecule is considered to be quasi  $G$ -symmetric. The quasi-symmetry of the electronic structure refers to the quantification of the distance between the  $G$ - and  $G'$ -symmetric structures and is closely related to the CShM and CSM concepts seen before. However, and from now on, in the present thesis, the concept that will be used to define the CSM of the electronic structure of a molecule is pseudosymmetry. Pseudosymmetry, although being related to quasi-symmetry, is the analysis of a given structure using a  $G_0$  symmetry group that is not necessarily the actual  $G$  symmetry for the structure. Thus, for a  $ML_4$  structure that is derived from a tetrahedron by an umbrella-type distortion, the actual symmetry of the molecule will be  $G = C_{3v}$ , but we may wish use the  $G_0 = T_d$  pseudosymmetry group to analyse the symmetry of the molecular orbitals. Note that we can always do such a pseudosymmetry analysis, for structures that are only slightly distorted tetrahedra as well as for structures which are quite far away from the ideal one. From a mathematical point of view the pseudosymmetry analysis (using  $G_0$  instead of  $G$  to classify molecular orbitals) is valid in all cases, although it gives only sensible results for cases with small departures from the ideal symmetry. If the  $G$  symmetric structure differs too much from  $G_0$  symmetry, the mixing of different irreducible representations becomes normally a mess which is hard to understand in simple terms.

In any case, the previous definition not only allows to quantify the symmetry content of the Hamiltonian or the electron density, but also to analyse the symmetry of the molecular orbitals or the wave function. To understand how a symmetry measure is calculated for each case, let us consider the domain of functions  $\varphi$  defined by the set of variables  $\tau$ . The general definition of the CSM of a function  $\varphi$  for a group  $G$  is defined in the following expression,

$$S_G(\varphi) = \min \left[ 1 - \frac{\sum_{i=1}^h \int_{-\infty}^{\infty} \varphi^* \hat{g}_i \varphi d\tau}{h \int_{-\infty}^{\infty} \varphi^* \varphi d\tau} \right] \quad (6)$$

where  $\hat{g}_i$  is the operator that transforms the domain of  $\varphi$  according to the symmetry operation  $g_i$  of the group  $G$ .

Sometimes it may be interesting to analyse the pseudosymmetry contents of the electron density, as it is an observable that can be measured through X-ray diffraction. The expression for the CSM of the electron density is,

$$S_G(\rho) = \min \left[ 1 - \frac{\sum_{i=1}^h \int_{-\infty}^{\infty} \rho(\mathbf{r})^* \hat{g}_i \rho(\mathbf{r}) d\mathbf{r}}{h \int_{-\infty}^{\infty} \rho(\mathbf{r})^2 d\mathbf{r}} \right] \quad (7)$$

It should be noted that contrarily to what happens to the wave function, the denominator  $\int_{-\infty}^{\infty} \rho(\mathbf{r})^2 d\mathbf{r}$  is, in general, not equal to 1. In fact, in most cases this value is greater than 1 and this normalization factor is called the density's self-similarity.

Additionally, applying the CSM concept to the molecular orbitals to their analysis of symmetry properties requires the introduction of a projection operator, defined for each irreducible representation (IR)  $\Gamma_n$  of the point group  $G$ ,

$$\widehat{P}_\Gamma = \frac{1}{h} \sum_{i=1}^h \chi_\Gamma^*(g_i) \hat{g}_i \quad (8)$$

where the scalar  $\chi_\Gamma^*(g_i)$  for non-degenerate irreducible representations is the complex conjugate of the character corresponding to the  $\hat{g}_i$  operator. As it was demonstrated<sup>16</sup>, the CSM equation (6) can be generalized to compute the symmetry with respect a given irreducible representation,

$$S_G(\varphi) = \min \left[ 1 - \frac{\sum_{i=1}^h \chi_\Gamma^*(g_i) \int_{-\infty}^{\infty} \varphi^* \hat{g}_i \varphi d\tau}{h \int_{-\infty}^{\infty} \varphi^* \varphi d\tau} \right] \quad (9)$$

In the case of degenerate irreducible representations equation (8) can be describe as follows,

$$\widehat{P}_{\Gamma_{rs}} = \frac{n_\Gamma}{h} \sum_{i=1}^h \left[ D_{g_i^{-1}}(\Gamma) \right]_{rs} \hat{g}_i \quad (10)$$

where  $\left[ D_{g_i^{-1}}(\Gamma) \right]_{rs}$  are the  $n_\Gamma^2$  different elements of the elements of the matrix representing the  $g_i$  operator in the irreducible representation  $\Gamma$ . For  $n_\Gamma > 1$  it is sufficient to only use one  $n_\Gamma$  of those operators, that can be used to calculate the expectation value of the operator  $P$

$$\langle \widehat{P}_\Gamma \rangle = \sum_{i=1}^{n_\Gamma} \langle \widehat{P}_{\Gamma_r} \rangle \quad (11)$$

Once we have established how to calculate the CSM with respect to an irreducible representation, it is possible to explore the symmetry behaviour of the molecular orbitals, a well-known concept in chemistry. To quantify the amount of symmetry of a molecular orbital  $\varphi$ , first it is necessary to express the molecular orbitals as a linear combination of a set  $\{\phi_\nu^\mu\}$  of functions adapted to the symmetry of a pseudosymmetry group  $G_o$ ,

$$\varphi = \sum_{\mu=1}^{N_{IR}} \sum_{\nu=1}^{n_{\Gamma^{\mu}}} c_{\nu}^{\mu} \phi_{\nu}^{\mu} \quad (12)$$

where  $N_{IR}$  is the number of IRs of  $G_0$  and  $n_{\Gamma^{\mu}}$  is the dimension of the  $\mu$ th IR. If the pseudosymmetry  $G_0$  coincides with the symmetry  $G$  of the molecule, the molecular orbitals can be expressed as a symmetry-adapted linear combination of atomic orbitals<sup>32-34</sup>. The weight  $\omega$  of the  $\mu$ th IR ( $\Gamma^{\mu}$ ) in a molecular orbital is calculated by the square of the set of coefficients  $\{c_{\nu}^{\mu}\}$ ,

$$\omega(\Gamma^{\mu}) = \sum_{\nu=1}^{n_{\Gamma^{\mu}}} |c_{\nu}^{\mu}|^2 \quad (13)$$

From here, it is easy to calculate the  $\omega(\Gamma^{\mu})$  for a  $G_0$  pseudosymmetry group for a set of molecular orbitals once we fix a given orientation for the operations in the  $G_0$  group. A good orientation is key to get the right answer as the irreducible representations of a group  $G$  are associated with an invariant kernel representation<sup>35</sup>. For each orbital, the symmetry operation expectation values (SOEVs) defined as  $\{\langle \varphi | \hat{g}_i | \varphi \rangle\}$  must be determined for the  $h$  symmetry operations of  $G_0$ . The expression for each  $\omega(\Gamma^{\mu})$  is,

$$\omega(\Gamma^{\mu}) = \frac{n_{\Gamma^{\mu}}}{h} \sum_{\nu=1}^{n_{\Gamma^{\mu}}} \sum_{i=1}^h D_{\nu\nu}^{\mu}(g_i)^* \langle \varphi | \hat{g}_i | \varphi \rangle \quad (14)$$

where  $h$  is the order of the group  $G_0$ ,  $D_{\nu\nu}(R)$  is the  $\nu$ th diagonal term of the  $n_{\Gamma^{\mu}} \times n_{\Gamma^{\mu}}$  representation matrix of  $\Gamma^{\mu}$ .

## Algorithm

The algorithm that measures the weight of the IRs of a set of molecular orbitals is implemented as follows:

1. Obtain the set of molecular orbitals  $\{\phi\}$  from an electronic structure calculation.
2. Select the pseudosymmetry group  $G_0$  and a proper orientation.
3. Compute the overlap of the atomic orbitals that define the basis set.
4. Obtain the set of SOEVs for a single operation  $R$  that belongs to a pseudosymmetry group  $G_0$ .
5. Repeat 3 and 4 for each non-equivalent operation of  $G_0$ .
6. Obtain the different weights  $\omega$  of an IR by multiplying the SOEVs and the representation matrix of that IR.
7. Repeat 6 for all possible operations  $R \in G_0$ .

## Bibliography

1. Atiyah, M. & Sutcliffe, P. Polyhedra in Physics, Chemistry and Geometry. *Milan J. Math.* **71**, 33–58 (2003).
2. Weyl, H. *Symmetry*. (Princeton University Press, 1952).
3. Carbó, R., Leyda, L. & Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **17**, 1185–1189 (1980).
4. Hodgkin, E. E. & Richards, W. G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem.* **32**, 105–110 (1987).
5. Petke, J. D. Cumulative and discrete similarity analysis of electrostatic potentials and fields. *J. Comput. Chem.* **14**, 928–933 (1993).
6. Petitjean, M. Geometric molecular similarity from volume-based distance minimization: Application to saxitoxin and tetrodotoxin. *J. Comput. Chem.* **16**, 80–90 (1995).
7. Robert, D. & Carbó-Dorca, R. A formal comparison between molecular quantum similarity measures and indices. *J. Chem. Inf. Comput. Sci.* **38**, 469–475 (1998).
8. Grimme, S. Continuous symmetry measures for electronic wavefunctions. *Chem. Phys. Lett.* **297**, 15–22 (1998).
9. Murray-Rust, P., Bürgi, H. B. & Dunitz, J. D. Distortions of  $\text{MX}_4$  molecules from  $T_d$  symmetry. II. Analysis of  $\text{PO}_4$ ,  $\text{SO}_4$  and  $\text{AlCl}_4$  species. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **34**, 1793–1803 (1978).
10. Buda, A. B., der Heyde, T. A. & Mislow, K. On Quantifying Chirality. *Angew. Chemie Int. Ed. English* **31**, 989–1007 (1992).
11. Carbó-Dorca, R. Scaled Euclidian distances: A general dissimilarity index with a suitably defined geometrical foundation. *J. Math. Chem.* **50**, 734–740 (2012).
12. Zabrodsky, H., Peleg, S. & Avnir, D. Continuous Symmetry Measures. *J. Am. Chem. Soc.* **114**, 7843–7851 (1992).
13. Pinsky, M. *et al.* Symmetry Operation Measures. *J. Comput. Chem.* **29**, 190–197 (2008).
14. Pinsky, M., Dryzun, C., Casanova, D., Alemany, P. & Avnir, D. Analytical Methods for Calculating Continuous Symmetry Measures and the Chirality Measure. *J. Com* **29**, 2712–2721 (2008).
15. Alvarez, S. *et al.* Shape maps and polyhedral interconversion paths in transition metal chemistry. *Coord. Chem. Rev.* **249**, 1693–1708 (2005).
16. Alemany, P., Casanova, D. & Álvarez, S. Continuous symmetry measures of

- irreducible representations: Application to molecular orbitals. *Phys. Chem. Chem. Phys.* **14**, 11816–11823 (2012).
17. Casanova, D., Alemany, P., Falceto, A., Carreras, A. & Alvarez, S. Pseudosymmetry analysis of molecular orbitals. *J. Comput. Chem.* **34**, 1321–1331 (2013).
  18. Cirera, J., Alemany, P. & Alvarez, S. Mapping the Stereochemistry and Symmetry of Tetracoordinate Transition-Metal Complexes. *Chem. - A Eur. J.* **10**, 190–207 (2004).
  19. Keinan, S. & Avnir, D. Continuous symmetry analysis of tetrahedral/planar distortions. Copper chlorides and other AB<sub>4</sub> species. *Inorg. Chem.* **40**, 318–323 (2001).
  20. Ruiz-Martínez, A. & Alvarez, S. Stereochemistry of compounds with coordination number ten. *Chem. - A Eur. J.* **15**, 7470–7480 (2009).
  21. Echeverría, J., Casanova, D., Lluell, M., Alemany, P. & Alvarez, S. Molecules and crystals with both icosahedral and cubic symmetry. *Chem. Commun.* 2717–2725 (2008) doi:10.1039/b719615f.
  22. Cirera, J., Ruiz, E. & Alvarez, S. Stereochemistry and spin state in four-coordinate transition metal compounds. *Inorg. Chem.* **47**, 2871–2889 (2008).
  23. Alvarez, S. & Lluell, M. Continuous symmetry measures of penta-coordinate molecules: Berry and non-Berry distortions of the trigonal bipyramid. *J. Chem. Soc. Dalt. Trans.* 3288–3303 (2000) doi:10.1039/b004878j.
  24. Alvarez, S., Avnir, D., Lluell, M. & Pinsky, M. Continuous symmetry maps and shape classification. The case of six-coordinated metal compounds. *New J. Chem.* **26**, 996–1009 (2002).
  25. Cremades, E., Echeverría, J. & Alvarez, S. The trigonal prism in coordination chemistry. *Chem. - A Eur. J.* **16**, 10380–10396 (2010).
  26. Casanova, D., Alemany, P., Bofill, J. M. & Alvarez, S. Shape and symmetry of heptacoordinate transition-metal complexes: Structural trends. *Chem. - A Eur. J.* **9**, 1281–1295 (2003).
  27. Casanova, D., Lluell, M., Alemany, P. & Alvarez, S. The rich stereochemistry of eight-vertex polyhedra: A continuous shape measures study. *Chem. - A Eur. J.* **11**, 1479–1494 (2005).
  28. Alvarez, S. The gyrobifastigium, not an uncommon shape in chemistry. *Coord. Chem. Rev.* **350**, 3–13 (2017).
  29. Ruiz-Martínez, A., Casanova, D. & Alvarez, S. Polyhedral structures with an odd number of vertices: Nine-coordinate metal compounds. *Chem. - A Eur. J.* **14**, 1291–1303 (2008).
  30. Casanova, D. *et al.* Minimal Distortion Pathways in Polyhedral Rearrangements. *J. Am. Chem. Soc.* **126**, 1755–1763 (2004).

31. Falceto, A., Casanova, D., Alemany, P. & Alvarez, S. Pseudo-symmetry Analysis of the d-block Molecular Orbitals in Four-Coordinate Complexes. *Inorg. Chem.* **52**, 6510–6519 (2013).
32. Bishop, D. M. *Group theory and chemistry*. (Dover Publications, 1993).
33. Cotton, F. A. *Chemical Applications of Group Theory*. (1990).
34. McWeeny, R. *Symmetry: An Introduction to Group Theory and Its Applications*. (Dover Publications, 2002).
35. Murray-Rust, P., Bürgi, H. -B & Dunitz, J. D. Description of molecular distortions in terms of symmetry coordinates. *Acta Crystallogr. Sect. A* **35**, 703–713 (1979).

## Objectives

After the previous introduction to the formalism of the continuous symmetry concept developed in the past, this thesis will cover the development of a Python library. Additionally, the enhancements that the new library provides over the past programs of continuous symmetry measures will be explained. Finally, a collection of three different representative applications of the continuous symmetry measures will be discuss.

Specifically, the main objectives of the thesis are list below:

- Develop a new Python library that collects the previous methodologies and improve the interaction with their algorithms by creating a uniform format.
- Enhance the capabilities of the Python library to improve the accessibility to anyone around the world
- Stablish a common design to facilitate future development and implementations on the actual library.
- Study the shape and symmetry of a group of organometallics called piano-stools
- Figure out if there is any connection between geometrical changes in the shape of piano-stools with their electronic structure.
- Find a connection between the geometrical and electronic structure symmetry
- Analyse the effect of temperature in different molecules and solids
- Understand the effect of the temperature and molecular environment in the tetrahedral shape of a set of phosphate anions



## Chapter 2. *Cosymlib*: a Python library for the computation of Continuous Symmetry Measures

### Introduction

The development of efficient computer models to analyse the behaviour of matter at the level of atoms and molecules has led to a notable breakthrough in science and the numerical implementation of the complex mathematical models of theoretical chemistry allows us today to safely simulate chemical and physical phenomena that cannot be reproduced in real experiments, either due to the lack of precise technology or due to an unaffordable material cost. The list of fields where computer simulations have become routine tools is huge: physics, biology, chemistry, economics, sociology, etc. and the number of computational models to simulate complex systems has increased during the past decades. Software development is not only necessary to apply existing mathematical models such as the Monte Carlo method, first applied in the Manhattan project<sup>1</sup>, or the quantum mechanical Hartree-Fock method<sup>2</sup> to study the electronic structure of atoms and molecules, but also to lower the computational cost of available computer simulations allowing their extension to larger systems or to longer time intervals. Lately, the assembly of large collections of information, also known as “big data” and their processing by machine learning algorithms, has become a helpful source to gain richer insights in science and business<sup>3</sup>. Therefore, computational models are being continuously developed with the aim not only to obtain more accurate results, but also to process, store, and further represent the available information in a more efficient ways<sup>4</sup>.

Since the development of quantum mechanics in the early 1920's, scientists interested in the structure and properties of matter at the atomic scale have sought new ways to solve complex differential equations in many variables that usually involve cumbersome numerical calculations and, consequently, the aid of efficient computer

technologies to perform complex calculations has been inseparable in this field of science since the early development of modern computers.

At the early stages of the development of modern computers, programming languages consisted in machine level languages which interacted directly with the hardware of the computer. Besides that, the resulting codes were difficult to understand, the major problem associated with these languages was that they differed significantly from computer to computer, preventing the sharing of codes between different developers. Assembly languages, or second-generation languages, were the first major development in programming languages as they simplified the machine language by utilizing low level instructions, coded using letters and numbers which were afterwards converted to the binary code of the computer. The biggest improvement in programming languages arrived, however, with the third generation of programming languages, which, even though they had slower execution speeds than second-generation languages, implemented several advantages for the users, basically a simple human-readable syntax and, therefore, an easy readability, as well as being able to generate portable and machine independent codes<sup>5</sup>. One of the first widely used languages was *FORTTRAN* (FORmula TRANslation), specifically designed for mathematical and scientific computations, which was developed before any modern text editor and instead of the modern input devices, the “keyboard” was a collection of 80-column punched cards that were given to a reader to introduce the code into the computer. This highly efficient language has been widely adopted by scientists for writing numerically intensive programs and is still used today in many research applications.

Later, third-generation languages involved the emergence of an object-oriented programming paradigm, which is based on the division of tasks into objects rather than imperative instructions, allowing them to structure the algorithm by specialized areas instead of procedurally solving the problems. The first object-oriented programming (OOP) language was Simula 67<sup>6</sup> which contained several concepts included in modern object-oriented languages such as Java or C++.

All these programming languages, plus fourth- and fifth-generation languages, are usually classified depending on how they face a computational problem, called

program paradigms. Although, sometimes it is difficult to classify them in a single paradigm as they present multiple features belonging to different paradigms. There are several programming paradigms which differ essentially in their complexity when dealing with an algorithm. However, all paradigms can be split into two major groups: imperative and declarative paradigms. The former command the program to perform multiple orders in a sequence of steps that modify the state, while the later cannot modify the state given that there are neither modifiable variables nor a semantic memory function<sup>7</sup>. An example of imperative and declarative programs written in *Python* is shown in figure 1 where a list of random number is analysed to extract those that are smaller than 50. Note that even if the final result is exactly the same, both programs are formally totally different.

```
from random import randint

random_list = sample(range(0, 100), 1000)

# Imperative

small_numbers = []
for number in random_list:
    if number < 50:
        small_numbers.append(number)

# Declarative

small_numbers = [x for x in random_list if x < 50]
```

Figure 1. Example of imperative and declarative paradigms giving the same result.

However, object-oriented programming is hard to be analysed in terms of the imperative or declarative paradigms as most of object-oriented languages are multi-paradigm and, consequently, it is usually treated as a separate programming paradigm with its own features and goals.

In our group, we have been developing computer codes in the past to compute continuous symmetry and shape measures (CSMs and CShMs, respectively) implementing the algorithms detailed in chapter 1. However, since the programs grew with the necessity of incorporating new types of tasks or when new algorithms were developed, some parts of these programs were not very well structured and with time it became more difficult to implement new features. The available codes had also a few limitations in their computational implementation that needed to be addressed to improve their efficiency and applicability. In addition, the codes were not adapted to work well in combination with each other as they had different interfaces. For these reasons we decided to develop a new open-source library written in *Python* to merge the previous existing programs under a common interface using an object-oriented design composed of intuitive object abstractions, such as molecules, which should help the users to rapidly familiarize with the program. This library has been named as *CosymLib*, which stands for Continuous Symmetry Library.

In this chapter, we include an overview of object-oriented languages and their advantages, plus an introduction of the most prominent features of the *Python* language that will be presented before describing the new *CosymLib Python* library in some detail.

## Object-oriented programming

Object-oriented programming (OOP) focuses on objects, an abstract data type that can include variables, functions and data structures. There are four principles that describe the main features of OOP: encapsulation, abstraction, inheritance, and polymorphism.

- Encapsulation

Encapsulation divides the code in small packages of information stored in private objects. The conjunction of attributes of an object is inaccessible and the only way to access this information is through methods. Technically, the

users can still access this information intentionally, but they are prevented to not do it, as it may result in a bad behaviour within that object.

- Abstraction

The abstraction principle focuses on mimicking the features and behaviour of real objects using abstract computational objects. This allows the users to ignore the complexity of abstract objects, as they are representations of well-known real objects with similar attributes.

- Inheritance

Through inheritance, an object may inherit all the methods and properties of another object. Inheritance wants to solve the problem of duplicate information between similar objects that share common attributes. In other words, this implies that a “child” object inherits all the methods and properties from a “parent” object, while it can additionally implement its own different methods.

- Polymorphism

Polymorphism is the capability of a method to do different things based on the object that it is acting upon. For example, a circle and a square have the same area attribute in common, but the way to calculate them vary entirely.

An example on how the Bird class is shown below the classes definition. In the example, the user creates two objects. The first one, `pidgey`, is an instance of the class `Bird` with `name = Trump` and `age = 2`. In object-oriented programming, an instance is a concrete occurrence of any object, existing usually during the runtime of a computer program. When the methods in the class are called for a given instance, the results inside each

method of the class are returned. For example, `pidgey.age()` returns the age of `pidgey` and `pidgey.fly()` its ability to fly.

The second example shows how the inheritance and polymorphism work in *Python*. The `auk` object this time is created through the `Penguin` class. This class inherits all the methods of the `Bird` class, thus when the attribute `age` is called for the `auk` object, the result is the same as in the `pidgey` example, returning however the value for the age indicated when was `auk` created. However, when we use the `fly()` method for a the `Penguin` object to ask if it can fly, the method was changed in the definition of the `Penguin` class to adapt to the reality of penguins in real life. Then, the method returns that the penguin cannot fly. It is interesting to note that the method `fly()` is called in the exact same way in both cases, even if the result is completely different, a characteristic of methods that is called polymorphism.

```
class Bird:

    def __init__(self, name, age):
        self._name = name
        self._age = age

    def age(self):
        return print("{} is {} years old".format(self._name, self._age))

    def fly(self):
        return print("{} is flying".format(self._name))

    def eat(self):
        return print("{} is eating".format(self._name))

class Penguin(Bird):

    def fly(self):
        return print("{} cannot fly, he is a penguin".format(self._name))
```

```
=====
>> pidgey = Bird("Trump", 2)
>> pidgey.age()
Trump is 2 years old
>> pidgey.fly()
Trump is flying

>> auk = Penguin("Angus", 4)
>> auk.age()
Angus is 4 years old
>> auk.fly()
Angus cannot fly, he is a penguin
```

Figure 2. A *Python* example of the main features in object-oriented programming and an example of how inheritance and polymorphism work.

Working with objects, with the benefits that are intrinsic from them, such as encapsulation, abstraction, inheritance and polymorphism make it easy to think why OOP is so powerful when writing code involving complex heterogeneous data structures. After the *Simula 67* release, several object-oriented languages were developed like *Java*, *C++*, etc. One of the most popular nowadays and the language that will be used to develop our symmetry program is *Python*.

### Why Python?

While other programming languages such as *Java* or *JavaScript* became more popular with the growth of the Internet, as they are considered OOP for webs, *Python* has been well-received during the last years. The number of questions asked in the *Stack Overflow* website, a question-and-answer site for programmers, reveals that *Python* has grown to become one of the most searched topics in the past years (figure 3). The reason behind this popularity is hard to explain and it might be related to the recent popularity of data mining and machine learning techniques, but most people agree that

*Python* is a language that is relatively easy to learn and understand, which helps new programmers to get used to its structure quickly and for this reason it has been adopted as the language used for teaching many introductory programming courses in colleges and universities.

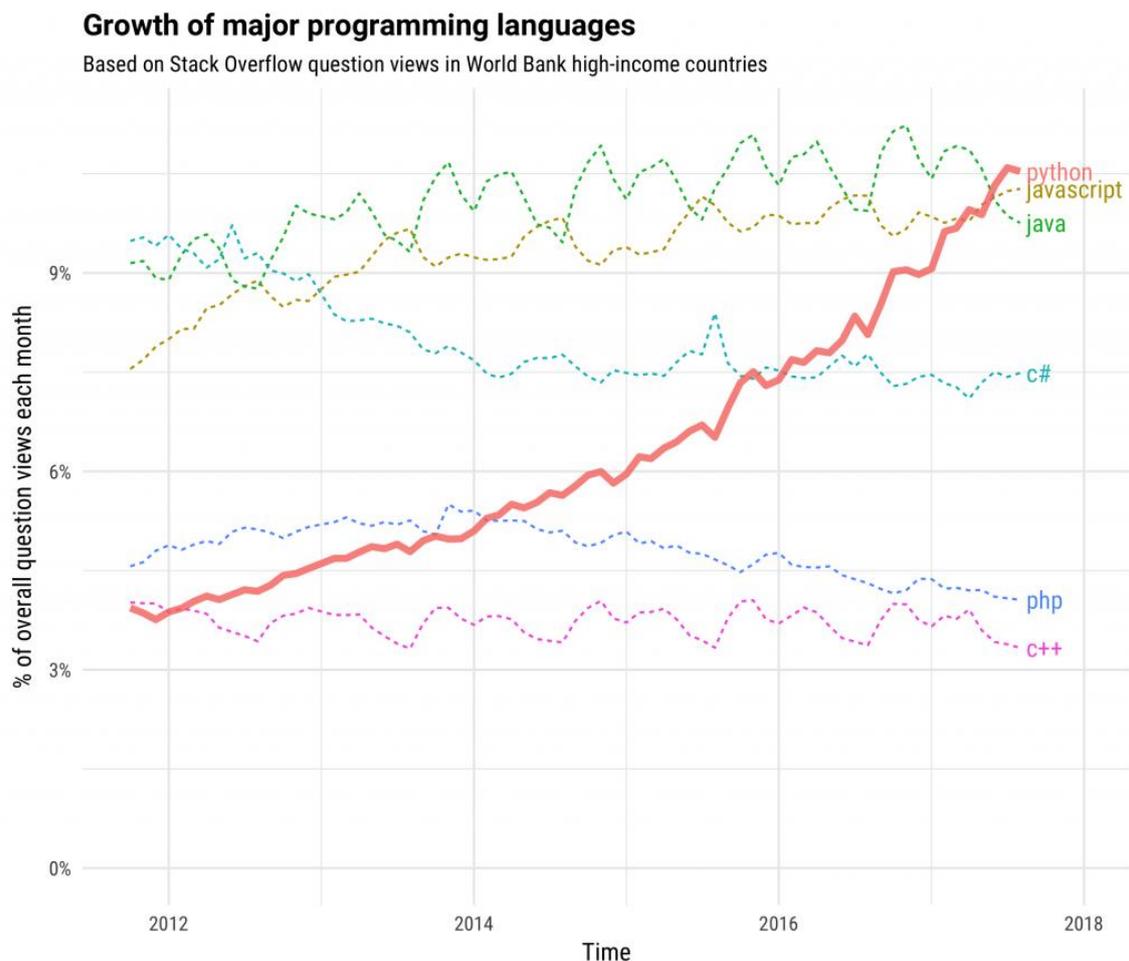


Figure 3. Growth of the questions about the major programming languages in the *Stack Overflow* webpage in the 2012-2018 period. Data obtained from the *Stack Overflow* webpage.<sup>8</sup>

*Python's* benefits do not only affect the programmer during software development, but also allow the users to have a clearer understanding of what the program is doing.

Until now, the intrinsic OOP features and the simplicity of understanding of the *Python* language have been discussed, however there are many other advantages derived from its use, and the more relevant will be listed below:

- **Multi-platform supported**

*Python* is not only supported by the most common platforms like Windows, MacOS and Linux, but also by other specialized operative systems <sup>9</sup>.

- **Large programming community**

The increase in popularity also creates an extensive community that supports and interacts with other *Python* programmers. This means that when you face a *Python*-related problem it is highly probable that someone else encountered that problem before you, documenting the solution he found, making thus the development of the programs faster and better.

- **Large libraries and packages**

The number of available *Python* libraries and packages is huge. According to the principal *Python* package index (PyPI), there are more than 300,000 projects available<sup>10</sup> and every day there are more and more packages and libraries around internet. This helps developers and users quite a lot. Why do you want to program your own linear algebra library when *Numpy*<sup>11</sup> will probably be more powerful and well supported?

In the present, a lot of scientific libraries such as *Numpy*, *Scipy*, *Matplotlib* and many others are easily available for *Python* developers. These libraries are also used in other specific scientific programs like *Pandas*, *Biopython*, etc. A

summary of the most popular libraries and their scientific field is shown in figure

4.

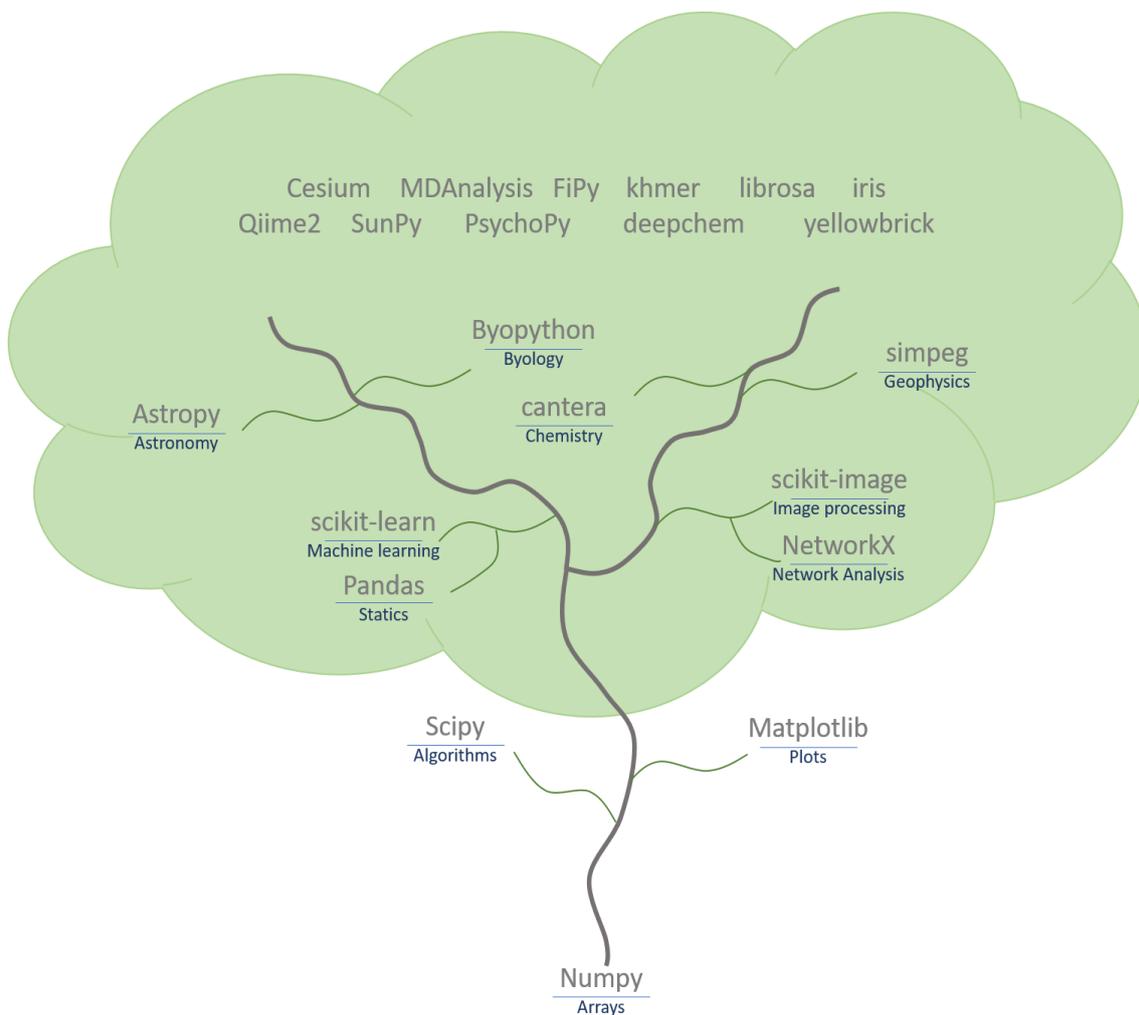


Figure 4. A Python scientific library tree of some of the most relevant Python packages and how they are related. Going up on the tree the complexity and specificity of the programs increases.

- **Interoperability**

*Python* can be extended with code written in other languages such as C, C++ or *FORTRAN*, in order to take advantage of combining the execution speed given

by compiled languages with the simplicity and flexibility of the *Python* syntax. This allows *Python* to interface with programs written in other languages saving a lot of time in file i/o (input & output).

- **Open source**

The development of programs using a free programming language makes *Python* not only cheaper for commercial and academic purposes, but also accessible for anyone who is willing to install it.

With all these features in mind, we decided that the *Cosymlib* suite of programs should be written in *Python* to take advantage of all the attributes described above.

## Cosymlib

*Cosymlib* is a *Python* library for computing continuous symmetry & shape measures (CSMs & CShMs). It has been developed as a powerful open-source program that could help scientists around the world to analyse stereochemical problems without the need of a thorough knowledge of the details of the underlying mathematical formalism and the algorithms used to compute the measures efficiently. Thus, a combination of a *Python* interface with our previous programs that implement efficient algorithms to compute symmetry measures has been arranged into the new continuous symmetry library *Cosymlib*. One of the problems that we had to tackle with the previous programs was that they were independent and did not keep a unique format, hampering thus the transfer of information between them. The use of the *Python* interpreter allowed us to include the core code of these algorithms into the main program, and therefore the development was split into several libraries with a sharing of data in a common format.

Nowadays, *Cosymlib* is an open-source software available at the *GitHub* webpage (DOI:[10.5281/zenodo.4925766](https://doi.org/10.5281/zenodo.4925766))<sup>12</sup> accessible to anyone with a *Python* interpreter and the

documentation of the program is available online<sup>13</sup>. The advantages of working with *GitHub* is that the software can be easily accessed by simply downloading a link from the *CosymLib* project and then installed in any computer. Additionally, other possible ways to install *CosymLib* are available, like downloading the code from the *Python* packages repository (PyPI).

It is important to note that *CosymLib* may work with input data in different formats that are usual in many computational chemistry programs. This question was especially important in the development of *CosymLib* since usually the transformation of molecular data obtained from other sources into a suitable input file for a specific software is quite tedious and time-consuming. Therefore, we decided that *CosymLib* should be able to read and handle information coded in five different standard input formats.

The most typical use of the procedures included in *CosymLib* is through scripts written in *Python* that are available from *GitHub* and are installed in the user's computer when the *CosymLib* is downloaded. These scripts use command line options to indicate the different types of calculations that may be done once a chemical structure (or a set of structures) is provided. The advantage of using these general-purpose scripts over specific user-written scripts calling the methods included in the *CosymLib* is that the syntax of the scripts included in the *CosymLib* distribution is simple but allowing to perform most of the tasks included in the *CosymLib* software without the need of having advanced programming skills in *Python*.

Moreover, the different scripts are specific for a given task, so for example, if the user wants just to analyse the shape of a structure given by a set of vertices he only needs to use the *shape* script, and if he wants to study the symmetry using CSMs for the same structure he will use the *gsym* script, with no need to know anything else about other functions in the library such as those dedicated to compute CSMs for the electronic structure for which additional information besides the geometry of the nuclear framework is needed. Intermediate-level users that want to combine different aspects of *CosymLib* without the need of writing their own scripts may use the general purpose *cosym* script which will allow them to access all features included in the *CosymLib* via

simple command line options. Some examples of using these general-purpose scripts will be shown later in this chapter.

## Code development

Since the *Cosymlib* was split into different parts, some part has been developed as an independent library such as the, *symgruopy*, *wfnsympy*, and *huckelpy*, allowing to decouple the development and debugging tasks for each of the libraries, making in this way the overall maintenance of the code much simpler and easier.

In order to run *Cosymlib* a set of very common *Python* libraries is required. These are listed below:

1. **NumPy**: This is the fundamental package for scientific computing in *Python*. It is a library that provides a multidimensional array object and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random number generation and much more.
2. **Matplotlib**: A cross-platform data visualization and graphical plotting library for *Python* and its numerical extension *NumPy*.
3. **PyYAML**: The *Python* parser of YAML, a data format designed for human readability and interaction, that will be used to store chemical properties such as information contained in the periodic table or the coordinates for the set of ideal reference shape structures.

Additionally, a couple of *FORTRAN* libraries are necessary in order to compile some part of the algorithm:

1. **Blas & Lapack**: Two *FORTRAN* libraries for linear algebra calculations

As the development of the different libraries included in *CosymLib* was carried out by different people at different places, we decided to employ GitHub, an internet hosting platform for software development and version control using *git* which provides many collaboration features such as bug tracking, feature requests, task management, and continuous integration among others. Version control is a system to manage changes in computer codes and other documents. This is accomplished by generating checkpoints of your progress during the development process. By the construction of a repository, all files and documents that are part of the software are stored and by pushing changes, the repository saves a check point of the current state of the code, which allows to have control over the differences between the current and past state of the files within the repository.

Additionally, *git* allows also to create different branches of a project. A branch is a copy of the code that can be modified, allowing experimentation and radical changes without affecting the code in other branches. It enables the possibility of working on future improvements without adding bugs to the main branch (usually called master). Moreover, it helps to include multiple features by creating several branches that can be developed by one or multiple programmers. Figure 5 shows schematically how the main branch (master) is split into two branches that evolve separately, merging them in a later moment of progress with the master. The scheme shows the coexistence in time with an additional feature2 branch that was created after and merged before the feature1 branch.

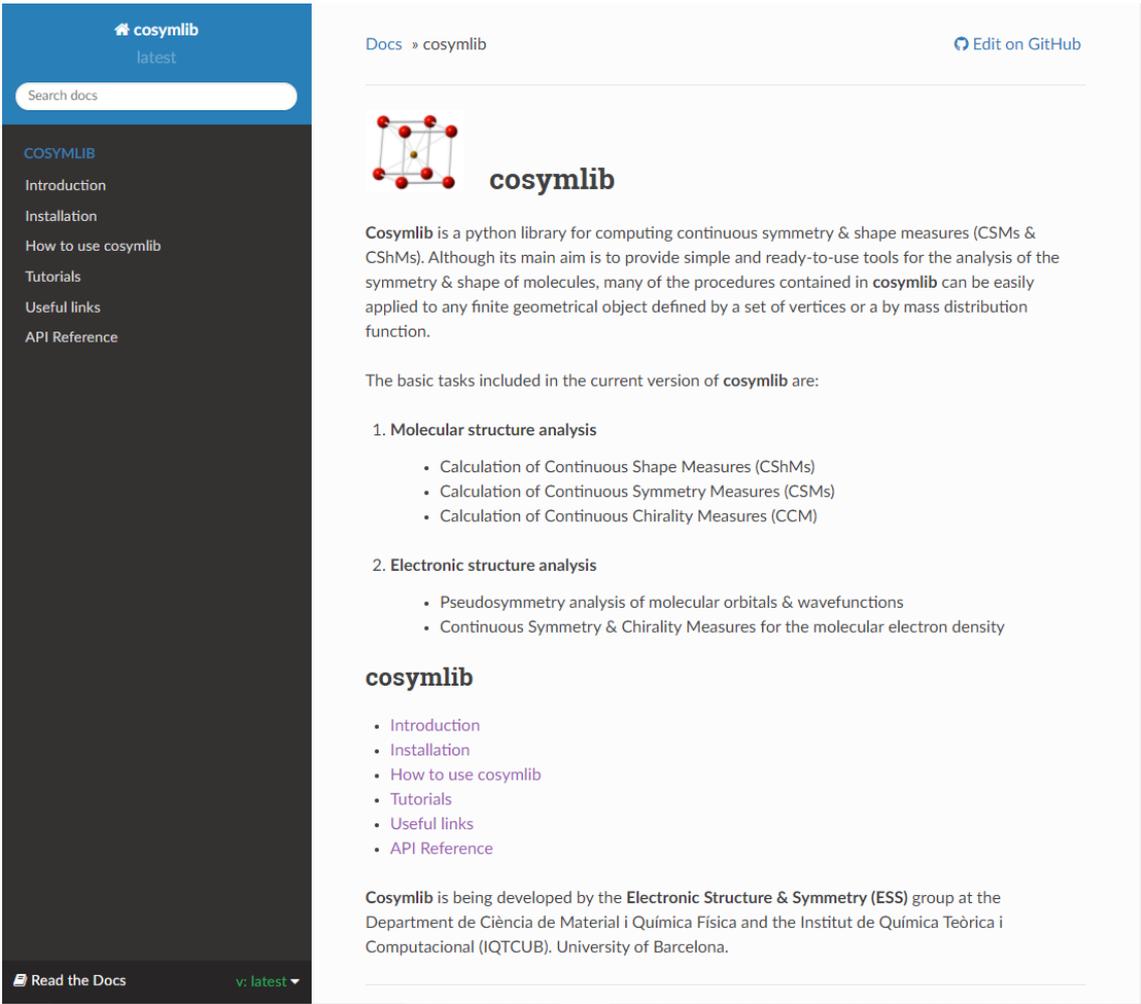


Figure 5. Timeline for the creation of multiple development branches and the merging (red dashed arrows) of the improvements introduced in the feature1 and feature2 branches into the master version.

The development was not only separated into different branches, but it was also continuously integrated at each small step. Continuous integration (CI) allows to automatically build and test each new implementation which helps to quickly detect errors as the code is integrated in small steps. For this purpose, the *Travis CI* service was chosen to build the code and run the tests automatically when changes are pushed into the GitHub repository. Each time a developer includes new changes, the *Travis CI* service installs the program into a virtual machine, runs the tests set by the programmers and informs about the issues or warnings (if any) that happened during the test's execution.

Additionally, this service includes a feature that is very convenient for *Python* packages. The deployment of software is available for those developers that want to distribute their software not only in *GitHub* but also to upload it to the *PyPI* repository. *PyPI* is a popular repository of *Python* packages that can be installed through the package installer for *Python* (*pip*), an easy to use package manager that simplifies the installation of *Python* packages. We use this feature to upload *Cosymlib* to *PyPI* since most installations of *Python* come with *pip* already preinstalled.

Finally, as the *Cosymlib* grew it was necessary to create a detailed documentation for the software. Documentation is an essential part of the development of a software project. It should provide all the necessary information about the use of the software to the users in a simple and readable way, since a well-done documentation may be the point making the difference between people using a software or not. *Read the Docs* provides an online service to host software documentation by automating building, versioning, and hosting a public website containing your manual pages. This service is also integrated within *GitHub* which implies that it is automatically triggered every time a new change in is pushed into the repository. The front page of the documentation in the webpage is shown in figure 6.



The screenshot shows the documentation page for the `cosymlib` Python library on the Read the Docs platform. The page is organized into a sidebar on the left and a main content area on the right. The sidebar, titled "cosymlib latest", contains a search bar and a list of navigation links: "COSYMLIB", "Introduction", "Installation", "How to use cosymlib", "Tutorials", "Useful links", and "API Reference". At the bottom of the sidebar, there is a "Read the Docs" button and a version selector set to "v: latest". The main content area, titled "Docs » cosymlib", features a "cosymlib" logo (a ball-and-stick molecular model) and a brief description: "Cosymlib is a python library for computing continuous symmetry & shape measures (CSMs & CShMs). Although its main aim is to provide simple and ready-to-use tools for the analysis of the symmetry & shape of molecules, many of the procedures contained in `cosymlib` can be easily applied to any finite geometrical object defined by a set of vertices or a by mass distribution function." Below this, it lists the basic tasks included in the current version: "1. Molecular structure analysis" (with sub-points for CShMs, CSMs, and CCM) and "2. Electronic structure analysis" (with sub-points for pseudosymmetry and symmetry/chirality measures). A "cosymlib" section provides a list of links: "Introduction", "Installation", "How to use cosymlib", "Tutorials", "Useful links", and "API Reference". At the bottom, it states that the library is developed by the Electronic Structure & Symmetry (ESS) group at the University of Barcelona.

Figure 6. *Cosymlib* documentation in the *Read the Docs* webpage<sup>43</sup>.

## Cosymlib Classes

*Cosymlib* is based in object abstractions written in classes, that represent concepts from chemistry, physics, and symmetry. Each class contains attributes and methods that try to mimic the real object's characteristics. For example, a geometry object contains the positions of the atoms or vertices of a structure as a set of cartesian coordinates. Figure 6 shows how *Cosymlib* is organized in folders that contain the blueprint for each class or set of functions in the library.

Any input information that should be analysed by the program, either given manually or through external files, should be transformed into one of two principal objects in *Cosymlib*: `geometry` or `molecule`. A `geometry` object (an instance of the

`Geometry` class) contains the geometry of a molecule (given as a set of cartesian coordinates indicating the nature and position of the different nuclei) and methods that depend only on this geometry. An example of one such method would be the square planar continuous shape measure for the  $\text{XeF}_4$  molecule is given below. In figure XX, the coordinates of the molecule and the symbol of the elements of each atom are stored in two lists called `coordinates` and `elements`, respectively. Calling the `Geometry` class, this generates an instance of that class that is stored as a `geometry` object, which in this case is called `xef4`. If a method of the object is called, for example the shape measure for the square planar shape, the object returns the value of the measure, which then can be saved in a variable `cshm` in the example.

```
coordinates = [[ 0.00000,  0.00000,  0.00000],
               [ 1.95000,  0.00000,  0.00000],
               [ 0.00000,  1.95000,  0.00000],
               [ 0.00000, -1.95000,  0.00000],
               [-1.95000,  0.00000,  0.00000]]
elements = ['Xe', 'F', 'F', 'F', 'F']
xef4 = Geometry(coordinates, elements)
cshm = xef4.get_shape_measure('SP-4', central_atom=1)
```

A `molecule` object (instance of the `Molecule` class) contains an instance of the basic `Geometry` class combined with additional information about the electronic structure, can also sustain CSM calculations for the electronic structure of a molecule, not only for its structure.

These two objects contain all the information required for the user to obtain symmetry and shape measures of a given molecule. However, *CosymLib* aims to facilitate the data processing as well as to give readable results for the general user. Consequently, the `cosymLib` class was created to store a set of methods that handle all the information and transform it into customized outputs. The attributes of these three main classes are explained below.

- CosymLib class

The `CosymLib` class is the main class in the *CosymLib* library. It only requires a list of structures that are encoded as `geometry` or `molecule` type objects. However, all methods in this class work with `molecule` type objects, in consequence, all structures are transformed into this type of object if needed. This class collects all possible outputs from the different symmetry and shape analysis included in the different methods of the *CosymLib*.

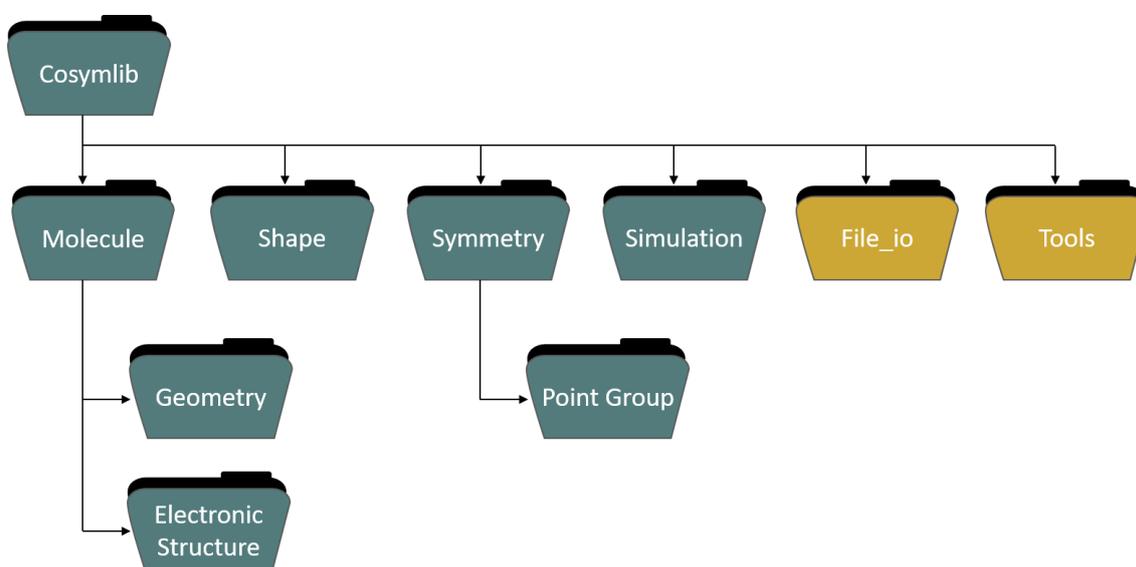


Figure 6. *CosymLib* program scheme in coloured folders. Each folder represents a directory with the class associated to the same name (green) or a directory containing a set of functions available for the user (yellow).

- Molecule class

As stated before, the `Molecule` class represents the abstract concept of a molecule in real world. It is composed by the `Geometry` and the `Electronic structure` classes and contains a set of methods that return the symmetry and shape information of a molecule. Additionally, if the electronic structure of a molecule is not provided by the user, an automated process begins to compute the

molecular orbitals at the Extended-Hückel<sup>14</sup> level by calling *HuckelPy*, an external *Python* library developed by us.

- [Geometry class](#)

The `Geometry` class is the central class that is necessary to perform a geometrical analysis of a structure in terms of continuous symmetry or shape measures. It may contain chemical information such as the nature of the atoms located at each coordinate or simply represent the vertices of a purely geometrical structure. The methods included in this class do not only return the geometrical information of the structure, but also call the shape and symmetry classes to obtain the continuous symmetry analysis of the given structure.

Additionally, if the object contains the symbols of the atoms in the structure, it is able to calculate the connectivity between them by using the covalent radii of the atoms. This connectivity may be used to accelerate symmetry calculations by lowering the number of permutations to be analysed. Nevertheless, the user can also set a given connectivity manually or generate a new one considering a threshold to detect bonds different from the one used by default.

- [Electronic structure class](#)

This class handles the data related to the electronic structure of a molecule. All the information regarding the basis set, the molecular orbital coefficients, the molecular orbital energies, the occupancies of the molecular orbitals, and the number of alpha and beta electrons is read from a previous calculation (or generated at the extended Hückel level) and stored. Three methods are available for the users which allow to set the number of alpha, beta, or both types of electrons in order to perform a symmetry analysis for the electronic structure considering an alternative electron occupation were electrons can be removed, added or changed

for a set of frozen molecular orbitals to explore symmetry changes for different electronic configurations.

The main function of the three previous classes is to store data for a molecule and request shape and symmetry information. The two following classes that will be presented below are mainly focused on the calculation of shape and symmetry measures of a structure by calling the algorithms presented in chapter 1.

- [Shape class](#)

The `Shape` class handles the CShMs calculations. It requires a geometry object to initialize an instance. To perform any calculation, a reference structure is also necessary. The reference structure may be one of the available predefined reference shapes in the program, stored in a *yaml* file, or be given directly by the user as an input structure. In both cases, the problem and the reference geometries are transferred to the compiled *shape* core function written in *FORTRAN* that performs a minimization to find the relative orientation leading to the minimum distances between the vertices of both structures and compute the CShM. It is also possible to fix a desired permutation that cancels the sweeping of the minimization process over all possible permutations, optimizing only the distance between the vertices of the two structures with respect to rotations and their relative size. In addition, it is possible to get coordinates for the problem structure that is closest to the reference structure obtained after the minimization process where the CShM is calculated. Since determining the closest structure is more computationally demanding than the calculation of the CShM, this option must be explicitly activated if the optimal structure is required for further analysis since by default only the CShM is being evaluated.

- Symmetry class

There are different procedures in *Cosymlib* to compute continuous symmetry measures, depending on the object that is being analysed. One way is by computing the purely geometrical CSM for a structure described by a set of vertices that may correspond to atoms or not. Alternatively, it is possible to perform a CSM analysis for different aspects of the electronic structure of a molecule such as the electron density, the wavefunction or the individual molecular orbitals. In the `Symmetry` class both aspects are considered, however the specific algorithms that perform the continuous symmetry analysis of the structure and the symmetry analysis of the electronic structure are separated in two *Python* libraries: *symgroupy* and *wfnscopy*.

In fact, the `Symmetry` class calls and stores the results of *symgroupy* and *wfnscopy* calculations into class methods that are accessible for the user.

- PointGroup class

Although the main aim of the *Cosymlib* is to compute continuous symmetry measures, a point group analyser has been implemented to obtain the closest point symmetry group that fits a given structure. An instance of this class can be called for `molecule` type object by using the `get_pointgroup()` method.

- Simulation class

The `Simulation` class is designed with the idea of offering the user multiple simulation methods to obtain the geometry or the electronic structure of a molecule. At this moment, the only procedure implemented in this class is to compute the molecular orbitals for a given geometry using the Extended-Hückel method implemented in the external *Huckelpy* library.

*CosymLib* also supports other secondary tasks not strictly related to the calculation of shape and symmetry properties. An example is the `file_io` file. As its name indicates, it includes procedures to control input and output and helps in the reading of data in different input file formats transforming them into suitable `molecule` or `geometry` objects. The available input formats are the standard `.xyz`<sup>15</sup>, the COORD format file from the CCDC<sup>16</sup>, the formatted checkpoint file `.fchk` created by *Gaussian*<sup>17</sup>, the molden type file<sup>18</sup> and the protein data base `.pdb` file<sup>19</sup>. A general reading function is also given to select which kind of input file was given. Additionally, there are functions to transform *CosymLib* data contained in a `geometry` or `molecule` type object into files written using well-known formats such as `.xyz` or `.fchk`.

Finally, as explained above, the *CosymLib* package includes a set of *Python* scripts that can be executed from the command line. They are written using the *argparse* library, and are intended to be a user-friendly command-line interface<sup>20</sup>, that contains general options for the calculation of continuous shape and symmetry measures. Each script is written to handle a specific type of measure. These scripts are *shape* (for shape measures), *shape\_map* (to create shape maps), *gsym* (for geometric symmetry measures) and *esym* (for symmetry measures of the electronic structure), and finally *cosym* (for general shape and symmetry measures). A couple of examples are given below.

First, suppose that we have a `crcl6.xyz` file containing the geometrical data of a Chromium hexachloride molecule ( $\text{CrCl}_6$ ) as indicated below,

```
7
crcl6
24  0.000000  0.000000  0.000000
17  2.229061  0.000000  0.000000
17 -2.229061  0.000000  0.000000
17  0.000000  2.229061  0.000000
17  0.000000 -2.229061  0.000000
17  0.000000  0.000000  2.229061
17  0.000000  0.000000 -2.229061
```

If someone wants to analyse how octahedral is  $\text{CrCl}_6$ , it is necessary to run the command,

```
~$ shape -m OC-6 -c 1 crcl6.xyz
```

Where *shape* is the name of the script, -m (short version of --measure) is the shape measure flag, OC-6 is the label for the octahedral shape, -c is the flag for the position of the central atom, 1 is where is located the central atom in the input file, and crcl6.xyz is the input file. The previous command line gives the continuous shape measure result in the following custom format,

```
-----  
COSYMLIB v0.9.5  
Electronic Structure Group  
Institut de Quimica Teorica i Computacional (IQTC)  
Universitat de Barcelona  
-----
```

```
Structure      OC-6  
  
crcl6,        0.000
```

```
-----  
                        End of calculation  
-----
```

The script displays the computed CShM value for the  $\text{CrCl}_6$  structure on the screen. Since we took a perfect octahedral arrangement of six chlorine atoms around the chromium, the computed CShM value is just  $S(\text{OC-6}) = 0$ . If we edit the .xyz file and replace, for example two chlorine atoms by two fluorine atoms, after the optimization at B3LYP/6-31G(d,p) level, the chromium chlorine distance is enlarged to 2.234027 Å while the chromium fluorine distances are as short as 1.718106 Å. Repeating the

calculation will get  $S(\text{OC-6}) = 1.372$ , indicating the departure of the distorted structure from the perfect octahedral shape.

Another usual task in the shape analysis is the construction of a shape map. Imagine that we have an `example.xyz` file containing the geometry for a set of tetracoordinate  $\text{ML}_4$  structures, where the central atom M is always in the first position of the coordinate list for each structure. If we want to compute the tetrahedral (T-4) and square planar (SP-4) shape measure for all structures and plot the corresponding shape map we need only to call the *shape\_map* script already installed in your computer by invoking the following instruction:

```
~$ shape_map -m_1 T-4 -m_2 SP-4 example.xyz
```

where `-m_1` (short version of `--measure_1` option) is the shape measure argument that will act as the x axis in the shape map and the `-m_2` (short version of `--measure_2` option) will act as the y axis.

After executing the command above, the code opens a figure window with the results in a customized shape map plot, using the *matplotlib* library (figure 7). *CosymLib* calculates first the minimum distortion path between the tetrahedron and the square planar structure (black line) and then calculates the position in the map for each structure contained in the input file. In this case, the positions of the tetracoordinated structures were generated using a random distortion over the four vertices and fill the accessible areas of the plot more or less uniformly.

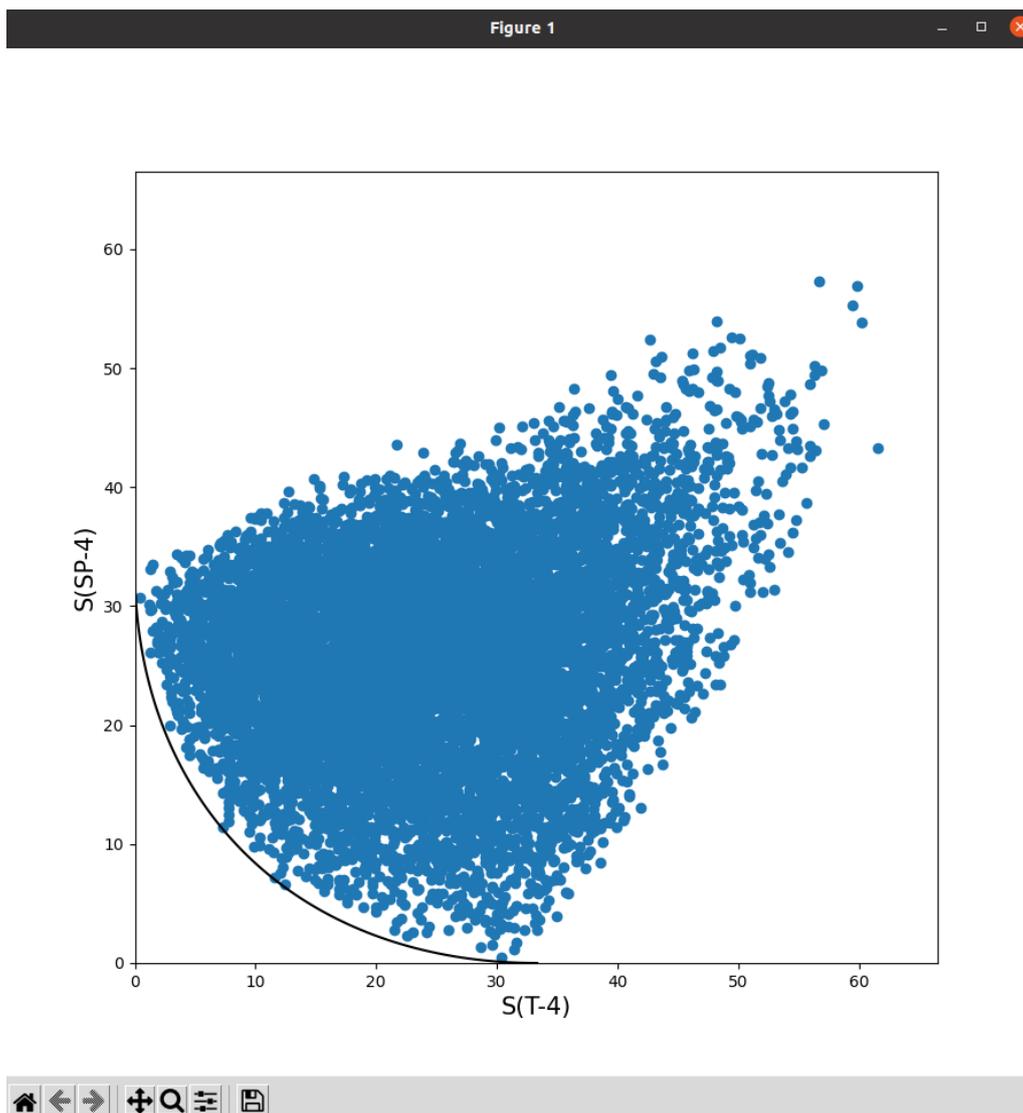


Figure 7. Shape map for the tetrahedron (T-4) and square planar (SP-4) shapes. The minimum distortion path between both structures is shown as a black line. The blue dots are the positions of a set of 10000 random generated tetracoordinate structures which do not maintain the M-L distance.

## Beyond CosymLib

Once the *CosymLib Python* library has been described, the external *Python* libraries that were developed by our group will be briefly described in this section. The three libraries are called *Symgroupy*, *Wfnsympy* and *Huckelpy*.

## Symgroupy

*Symgroupy* is a *Python* library to perform continuous symmetry calculations for a molecular structure. This library contains a main class called *Symgroupy* which requires the coordinates of a structure and the label for a symmetry group for which the CSM will be computed. Alternatively, the program can calculate instead single symmetry operation measures if the user wishes to. Then, these variables are sent to the *FORTTRAN* compiled core that performs the calculation of the corresponding continuous symmetry measure.

After the value for the CSM is obtained, the results are post-processed to become properties of the `Symgroupy` class, so for example, the `csm` property returns the CSM value of a calculation. Also, similarly as for the `Shape` class, the nearest structure with the desired symmetry can be obtained together with the optimum orientation of the symmetry group and the optimum permutation that minimizes the measure are given. Additionally, the user can ask the class to store the *n*th local minimum values and their optimum orientations. An example of how CSM calculation works using the CosymLib library is given below.

```
~$ gsym -m C4 xef4.xyz
```

The last command is necessary to analyze, for example the xenon tetrafluoride previous example molecule for the  $C_4$  symmetry group, giving the following results.

```
-----  
COSYMLIB v0.9.5  
Electronic Structure Group  
Institut de Química Teòrica i Computacional (IQTC)  
Universitat de Barcelona  
-----
```

```
Evaluating symmetry operation : C4
```

```
xef4          0.000
```

```
-----  
                          End of calculation  
-----
```

As for now, the available groups in *Symgroupy* are restricted to the pure and improper rotation groups  $C_n$  and  $S_n$  due to the simplicity in the analytical solutions. However, the implementation of new approximations to calculate continuous measures for other point-symmetry groups as well as to decrease the minimization process that has problems with the scalability of the system's size were discussed during the development of the library and their implementation is projected for a near future.

### [Wfnscopy](#)

While the *Symgroupy* library focuses on the geometrical symmetry of a structure, the *wfnscopy* library is related to the symmetry of the electronic structure of a molecule. This library was developed to calculate symmetry measures for the electronic density, the wavefunction, and the molecular orbitals of molecules. To do this it requires the coordinates and chemical elements of the molecule, the electronic structure defined by a basis set of atomic orbitals and the coefficients of the molecular orbitals expressed in this basis set, and a symmetry point group. Any gaussian-type basis set can be used,

either with defined using spherical or Cartesian  $d$  and  $f$  orbitals. If molecular orbitals coefficients for beta electrons are given, the electronic structure is considered as the result of an open-shell calculation, otherwise it will be treated as a closed shell case. Additionally, the orientation of the symmetry group operations with respect to the molecular structure, necessary to perform a pseudo-symmetry analysis using a given symmetry group, can be given by fixing the centre and the direction of two axes. If no orientation is given, the program will find the optimum orientation by minimizing the symmetry measure for the molecular structure or for the electron density, locating the centre for the symmetry group operations at the geometrical centre or at the centre of charge, respectively. Depending on the size and the AO types included in the basis set, the process of optimizing the orientation of the symmetry group for the electron density may be computationally expensive due to the high number of integrals needed to compute the CSM for each orientation. The occupation of the molecular orbitals, which the user can arbitrarily change, is required to compute symmetry measures for the wavefunction and the electron density.

The computations of CSMs for the density and wavefunctions are divided into two different *FORTRAN* codes that implement the algorithms described in chapter 1. The results are returned as properties defined in the `WfnSymPy` main class. The symmetry overlap expectation values (SOEVs) for alpha and beta molecular orbitals are accessed through `mo_SOEVs_a` and `mo_SOEVs_b`. The irreducible representation of each molecular orbital for the given point symmetry group is stored in the `mo_IRd_a` (and `mo_IRd_b`) properties as well as the mixture of irreducible representations for the total occupied wavefunction. Finally, the orientation and centre for the symmetry group used in the calculations are also stored, either if the user manually introduced them or if they were obtained in a minimization process.

## HuckelPy

The *HuckelPy Python* library was developed to obtain a reasonable electronic structure at a low computational cost, choosing the extended Hückel method<sup>14</sup> for this purpose.

Despite using a semi-empirical method to obtain the valence molecular orbitals of a molecule, it is considered good enough to qualitatively describe their main features, especially those related to molecular symmetry. During the thesis some examples will be presented using this methodology. The idea behind developing our own extended Hückel (eH) code, although other programs which perform the same type of calculation are already available, is that having a more reliable control of the attributes and methods used to obtain the MOs will be helpful to handle further developments in the *Cosymlib* library. An extended Hückel calculation requires only the input of atomic coordinates and symbols for the atoms in the molecule as well as the charge for the system. With this information, the `ExtendedHuckel` main class performs a diagonalization of the Hamiltonian matrix to obtain the eigenvectors and eigenvalues, which represent the molecular orbital coefficients and energies, respectively.

The Hamiltonian matrix elements ( $H_{ij}$ ) are obtained by the Wolfsberg-Helmholz approximation<sup>21</sup>,

$$H_{ij} = K \frac{h_i + h_j}{2} S_{ij}$$

where each element of the Hamiltonian matrix is computed from the overlap matrix elements ( $S_{ij}$ ), the Wolfsberg–Helmholz constant ( $K = 1.75$ ) and the valence orbital ionization energies  $h_i$  for each atomic orbital. These are taken from the set obtained by Vela et al.<sup>22</sup>. The overlap matrix is computed for a valence only basis set of STO-3G atomic orbitals, which are encoded in an atomic orbital object that includes the position, coefficients and exponents defining each STO-3G orbital obtained by the works of Huzinaga et. al.<sup>23–26</sup>, as well as the atom nature and the valence orbital ionization energy. We are currently working now on the option to change these values to a custom set of coefficients, exponents and energies given by the user.

## Future scope

The development of *Cosymlib* and its external libraries will be of a great help to future implementations of new algorithms and procedures as the main structure of the library is already finished. For example, new methodologies to compute the electronic structure could be easily integrated in the simulation class or additional input formats could be added to the `file_io` file. Also, other symmetry groups and approximations can be added to the *symgroupy*, as well as new optimization approaches to find the optimum orientation of a symmetry group could be implemented in the *wfnscopy* library.

## Summary

In this chapter an introductory historical overview of the importance of computational simulations was given to offer a description of the main features of computer simulations nowadays. Also, the properties of object-oriented languages were briefly introduced to justify the benefits of choosing *Python* as the main programming language to develop the *Cosymlib* library. A brief introduction of the usage of this library was given, as well as a more detailed picture of the objects included in the library and how they interact with each other.

Additionally, pieces of the code were separated into different libraries that perform specific calculation so that the development and maintenance can be simplified. Also, the description of the different classes in the library was detailed to explain the procedures that each individual class follows to compute the continuous symmetry measures.

## Bibliography

1. Benov, D. M. The Manhattan Project, the first electronic computer and the Monte Carlo method. *Monte Carlo Methods Appl.* **22**, 73–79 (2016).
2. Froese Fischer, C. General Hartree-Fock program. *Comput. Phys. Commun.* **43**, 355–365 (1987).
3. Provost, F. & Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **1**, 51–59 (2013).
4. Zomaya, A. Y. & Sakr, S. *Handbook of big data technologies. Handbook of Big Data Technologies* (2017). doi:10.1007/978-3-319-49340-4.
5. O'Regan, G. Computer Programming Languages. in *A Brief History of Computing* (Springer, 2008).
6. Dahl, O.-J., Myhrhaug, B. & Nygaard, K. Some Features of the Simula 67 language. 29–31 (1968).
7. Gabbrielli, M. & Martini, S. *Programming Languages: Principle and Paradigms. Journal of Chemical Information and Modeling* vol. 53 (2010).
8. The Incredible Growth of Python | Stack Overflow. <https://stackoverflow.blog/2017/09/06/incredible-growth-python/>.
9. Python for Other Platforms | Python.org. <https://www.python.org/download/other/>.
10. PyPI · The Python Package Index. <https://pypi.org/>.
11. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
12. CosymLib. <https://github.com/GrupEstructuraElectronicaSimetria/cosymlib>.
13. CosymLib documentation. <https://cosymlib.readthedocs.io/en/latest/>.
14. Hoffmann, R. An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* **39**, 1397–1412 (1963).
15. XYZ file format. [https://en.wikipedia.org/wiki/XYZ\\_file\\_format](https://en.wikipedia.org/wiki/XYZ_file_format).
16. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
17. formchk | Gaussian.com. <https://gaussian.com/formchk/>.
18. MOLDEN a visualization program of molecular and electronic structure. <https://www3.cmbi.umcn.nl/molden/>.
19. RCSB PDB. <https://www.rcsb.org/>.
20. argparse — Parser for command-line options, arguments and sub-commands.

<https://docs.python.org/3/library/argparse.html>.

21. Wolfsberg, M. A. X. & Helmholz, L. The spectra and electronic structure of the tetrahedral ions  $\text{MnO}_4^-$ ,  $\text{CrO}_4^-$ , and  $\text{ClO}_4^-$ . *J. Chem. Phys.* **20**, 837–843 (1952).
22. Vela, A. & Gázquez, J. L. Extended Hückel parameters from density functional theory. *J. Phys. Chem.* **92**, 5688–5693 (1988).
23. Huzinaga, S. Gaussian-type functions for polyatomic systems. I. *J. Chem. Phys.* **42**, 1293–1302 (1965).
24. Huzinaga, S. & Sakai, Y. Gaussian-type functions for polyatomic systems. II. *J. Chem. Phys.* **50**, 1371–1381 (1969).
25. Huzinaga, S. & Arnau, C. GaussianType Functions for Polyatomic Systems. III. *J. Chem. Phys.* **52**, 2224 (1970).
26. Huzinaga, S. & Arnau, C. Gaussian-type functions for polyatomic systems. IV. *J. Chem. Phys.* **53**, 348–351 (1970).



## Chapter 3. Shape and symmetry of organometallic piano-stool compounds

### Introduction

The use of polyhedral models to rationalize stereochemical information was introduced, independently, by Le Bel<sup>1</sup> and Van't Hoff<sup>2</sup> in 1874. In order to explain the existence of optical isomers in organic compounds they proposed that the bonds of a carbon atom linked to four other atoms should be directed towards the vertices of a tetrahedron<sup>3</sup>.

A few years later, Werner introduced the polyhedral model in coordination chemistry, where a central transition metal is coordinated to a set of neutral or anionic ligands<sup>4</sup>. In this sense, aside from structural organic chemistry, where  $sp^3$  atoms are surrounded by four atoms in a tetrahedral geometry, metals can exhibit coordination numbers from two to fourteen. In contrast to what happens in organic chemistry, in coordination compounds it is possible to find different polyhedral structures for a given type of  $ML_n$  compounds, for instance the tetrahedron and the square for  $ML_4$  compounds. The determination of the number of atoms coordinated to a central metal in a molecule or solid have been a central primary focus in x-ray diffraction and other crystal structure determination techniques, thus giving the necessity for a clear description and classification of the coordination environments in terms of their shapes and symmetries. However, the common yes or no classification according to point group symmetry is only valid for idealized geometries where the approximation tends to describe a highly symmetric static system that poorly describes the details of polyhedral geometries found in real crystal structures.

The problem in unambiguously classifying coordination sphere in  $ML_n$  compounds by its shape, where the number of ligands  $n$  is above 6 presents a challenge, especially for low symmetry structures due to the large number of similar alternative polyhedral

structures for large values of  $n$ . Another difficult case emerges when the central metal atom is bonded to carbon atoms in organic ligands as in organometallic compounds.<sup>5</sup> In these case the metal coordination number may change depending on the coordination mode. This can be illustrated by the well-known ferrocene molecule (figure 1), where two cyclopentadiene ligands surround the iron atom in a sandwich-like structure. At first, the assumption was to assign a single covalent bond of a saturated carbon atom from each cyclopentadiene to the iron atom<sup>6</sup>. However, nowadays it is well-accepted that the aromatic cyclopentadienyl ligands share the twelve  $\pi$ -electrons with the  $\text{Fe}^{2+}$  atom.

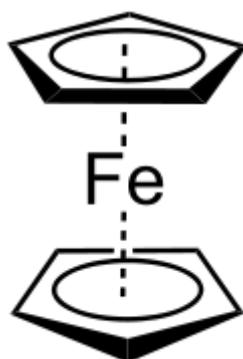


Figure 1. Ferrocene molecule. The cyclopentadiene ligands share all the  $\pi$ -electrons with the iron atom, a bonding mode represented by a dashed-line joining the metal atom with the centroid of the aromatic rings.

From the stereochemical point, the difficulty in defining a polyhedral model for the metal's coordination sphere in organometallic compounds arises when one has to decide which set of points should be included for  $\pi$ -donor ligands. Contrarily to what happens for  $\sigma$ -donor ligands that are simply described by assigning a single vertex of the polyhedron for each donor atom. An example of the difficulty found for  $\pi$ -donor ligands is displayed in figure 2: is a  $\pi$ -bonded ethylene ligand coordinated to a metal atom M better represented by two points associated with the C atoms, or just by a single point in middle of the double C-C bond?

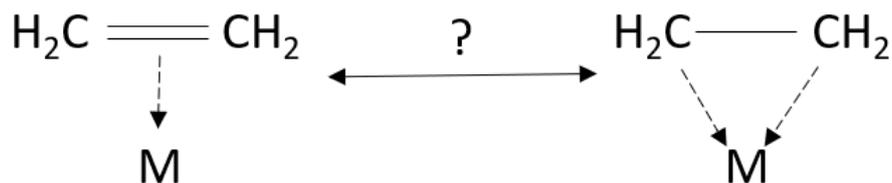


Figure 2. Different coordination possibilities for ethylene. The ethylene double bond can act as a single coordination site located at the center of the C-C bond (left) or as two coordinated sites where the carbon atoms act as two vertices (right).

One particularly complex case is that of cyclic ligands such as benzene, an aromatic ring with three double bonds delocalized along the ring of carbon atoms. Consequently, a benzene ligand can act as a  $\sigma$ -donor via a saturated carbon, a  $\pi$ -donor through one double bond or by coordinating two or the three double bonds with the central atom. From a geometrical point of view, ferrocene and other sandwich compounds with ligands such as benzene have been often described as a single coordination site which usually is located at the centroid of the cyclic molecule. From an electronic structure perspective, however, the metal centred d-type molecular orbitals are often associated to the two over three  $e_g/t_{2g}$  splitting pattern found for octahedral  $ML_6$  complexes, which geometrically implies to consider that each cyclic ligand should be represented by 3 vertices, even though the symmetry of the whole complexes is just  $D_{nh}$  or  $D_{nd}$ , where  $n$  is the number of atoms in the rings which are, in general, not even subgroups of  $O_h$ , except for a few cases such as  $D_{2h}$ ,  $D_{2d}$ ,  $D_{4h}$  or  $D_{3d}$ .

Following the success of using continuous shape measures (CShMs) to characterize the shape and symmetry of the coordination environment of  $ML_n$  coordination compounds, J. Cirera et. al.<sup>7</sup> extended the study to coordination compounds with  $\pi$ -bonded ligands such as olefins, alkynes and allyls, finding that considering the centroid of the double or triple bond as a single vertex in the coordination polyhedron yields useful polyhedral models for the coordination sphere around the metal atoms in these complexes. A specially interesting case included in this study is that of piano-stools complexes<sup>8</sup>, organometallic half-sandwich compounds with one cyclic ligand such as cyclopentadiene (Cp) and three atomic donor ligands such as Cl. From the point of view

of symmetry these complexes are specially interesting since they feature two incompatible rotational symmetries,  $C_3$  and  $C_5$ , along the same axis a case already studied in a previous work<sup>9</sup> (figure 3). In this aspect, they considered two different approaches to the geometrical description of the metal coordination sphere in piano-stools, one considering just the centroid of the cyclopentadiene as a single coordination site and another by taking into account three coordination sites located at the centres of the two double bonds and at the unsaturated carbon of the aromatic ring. As highlighted in figure 3, these two alternatives lead to a tetrahedral and an octahedral polyhedral model, respectively. In order to estimate which of the two models would yield a best description for the stereochemistry of the known piano-stool complexes, Cirera et al. centred their attention in the  $ML_3$  fragment in these compounds using CShMs to compare the shape of this fragment with either that of an ideal vacant tetrahedron or a fac-trivacant octahedron, the two shapes that the  $ML_3$  fragment is expected to adopt in the tetrahedral and in the octahedral model, respectively. These two shapes correspond to a pyramidal  $ML_3$  shape with all three equal M-L distances and three equal L-M-L angles with values of  $109.4^\circ$  and  $90.0^\circ$  for the vacant tetrahedron and the fac-trivacant octahedron, respectively.

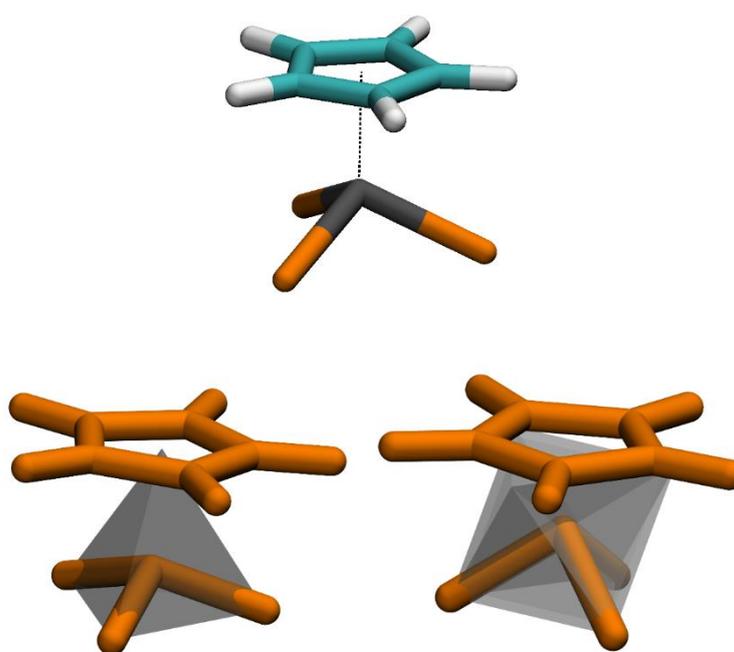


Figure 3. CpMCl<sub>3</sub> molecule and the tetrahedral and octahedral possible shape descriptions

This study shows, however, an intriguing result: the  $ML_3$  fragments in the known piano stool complexes present a change in their shape, that evolves from tetrahedral to octahedral as the electronic configuration of the d orbitals changes from  $d^0$  to  $d^6$ . This purely geometrical finding leads to some interesting questions that were not touched in ref. 7. On one hand it would be desirable to confirm the observed trend using first principles quantum chemical calculations for a series of representative complexes in order to study the influence of the type of ligand as well as the size of the rings in the geometry of the  $ML_3$  fragment. On the other hand, if the trend is confirmed by these calculations it would be interesting to unveil the driving force behind this structural change and to check if this tetrahedral to octahedral transition implies also deep changes in the electronic structure of the compounds, switching from the 3 over 2  $t_2/e$  splitting pattern in tetrahedral symmetry to the 2 over 3  $e_g/t_{2g}$  of octahedral complexes as the number of d electrons increases.

In order to give some answers to these questions, in the present chapter we will discuss the shape and symmetry of piano-stool compounds, including a pseudosymmetry analysis of their d-type orbital. For this purpose, we will be using the tools described in chapter 1, in order to establish if there is a relation between the orbital nature and the geometrical configuration of these molecules. Besides the piano-stools containing a Cp ligand considered originally in ref. 7, we will include in this study other half-sandwich compounds with different aromatic rings in order to explore the effect of different combinations of rotational symmetries around a same axis on the local electronic structure around the transition metal atom.

## Methodology

The search for experimental structures in the Cambridge Structural Database (CSD)<sup>10</sup> was constrained to organometallic molecules that contain one cyclic carbon fragment, a central transition metal atom, and three non-organic ligands that may be not equivalent. To simplify the study, the aromatic rings (Ar) and the three ligands (L) were

restrained to present only substituents as bulky as an ethylene molecule to avoid possible steric effects that could modify the geometrical structure.

As stated above, to understand how piano stools can be classified in terms of shape and symmetry, the participation in geometric terms of the cyclic ligand to the coordination polyhedron should be defined. Based on previous studies, the two cases that are of interest, are the one that represents the whole ligand ring as a single coordination site or the one where two double bonds and one atom act as three coordination sites, resulting in two possible polyhedral arrangements, the tetrahedron and the octahedron, respectively. If the  $ML_3$  fragment is considered, the tetrahedral model should be associated to that of a vacant tetrahedron (vT-3) and for the octahedral model we should consider the fac-trivacant octahedron (fvOC-3).

Additionally, a theoretical model was constructed by computing a set of neutral piano-stool molecules with different number of electrons in the d-type molecular orbitals. For example, if the ligands considered are cyclopentadienyl and three chlorine anions, the central metal should have a formal charge of +4, thus the titanium  $CpTiCl_3$  compound was considered to represent a  $d^0$  electronic configuration (0 electrons in the d-type molecular orbitals). On the other side, Ni was considered to form a  $d^6$  compound. Then, the series of titanium, vanadium, chromium, manganese, iron, cobalt and nickel  $CpMCl_3$  compounds represent simplified models for compounds with  $d^0$  to  $d^6$  electronic configurations, respectively. For each electronic configuration, the different spin states were analysed. The results conclude that in all cases the ground state corresponds to the situation where the three energetically lower d-type MOs are filled while the two other remaining d-type MOs remain empty. This means that for each electronic configuration from  $d^1$  to  $d^3$  the spin multiplicity will be  $2S+1 = 2, 3$  and  $4$ , respectively. Contrarily, the  $d^4, d^5$  and  $d^6$  configurations will prefer to pair electrons instead of increasing the multiplicity, giving the  $2S+1 = 3, 2, 1$  electronic configurations.

Geometry optimizations for these simplified models were carried out using different density functionals and basis sets, but the resulting structures did not vary strongly neither in energy nor geometry. The hybrid meta-GGA functional TPSSh<sup>11,12</sup> with a 6-31G(d,p) basis set was chosen for further calculations and the resulting optimized geometries that lie in the minimum of the potential energy surface were considered in

the geometric analysis. In order to simplify the analysis of the relation between geometry and the electronic structure, a symmetrical model was considered by fixing the concurrent rotational symmetries of the two fragments, an exact  $C_3$  symmetry for the  $L_3$  fragment and an exact  $C_5$  symmetry for the Cp ligand to share a common axis passing through the metal atom.

Concerning the electronic structure, the main goal was to obtain a qualitative insight in the changes in symmetry of the d-type molecular orbitals as the molecular geometry changes from “tetrahedral” to “octahedral”. For this reason we did not consider molecular orbitals provided by the general ab initio or density functional theory based methods, choosing the semi-empirical extended-Hückel<sup>13</sup> (eH) method, as it provides a cleaner and intuitive way to understand the electronic structure, allowing to change the d-electron count while keeping the same model parameters and focus only on the effect of changing the electron count. The eH MOs were obtained for the DFT optimized geometries by the *HuckelPy Python* module<sup>14</sup>, a program developed in our group, that calculates the molecular orbital energies and coefficients using a minimal valence STO-3G basis set and the valence orbital ionization energies provided by A. Vela et. al.<sup>15</sup>. The energy parameters used in these calculations can be found in table 1A of the Appendix.

To analyse the symmetry of these molecular orbitals, a pseudosymmetry analysis was performed using either the tetrahedral ( $T_d$ ) or octahedral ( $O_h$ ) symmetries together with other subgroups of these two symmetry groups. To study the pseudosymmetry measures, the *CosymLib* library was used to obtain the participation of the different irreducible representations of the pseudosymmetry group for a given molecular orbital. The orientation of the pseudosymmetry groups was fixed by forcing the  $C_3$  symmetry axis, present both in the  $T_d$  and  $O_h$  groups, to pass through the central metal and the centroid of the aromatic ring and locating the metal atom at the origin.

## Shapes of piano-stool compounds in crystals

As stated before, the first set of molecules to be analysed will be the half-sandwich or piano-stool structures that contain a cyclopentadiene ring as the aromatic ligand. The initial search yielded around 100 experimental structures in the CSD that match our case of study. In an initial search we found different non-aromatic ligands such as chlorine, iodine, carbon monoxide, cyanide or methyl. In a few cases (~11%), instead of three equal L ligands we find an  $(L_1)_n(L_2)_m$  mixture with  $n + m = 3$ . Most of the central atoms of the complexes included in the final set are first- and second-row transition metals, with electronic configurations that present different d-type orbital occupations, from 0 to 9 electrons, except for the  $d^4$  case for which no experimental structures were found. Most of the structures, however, present a  $d^6$  configuration, around 70% of the total, while the second dominant configuration is  $d^5$  (15%). In the case of cyclobutadiene (Cbd) and benzene (Bz), only 6 and 41 structures were found in CSD, respectively. The amount of structures found for both cases was considered insufficient to find any trend in these structures because almost all the molecules present a  $d^6$  electronic configuration and only a few (4) in the benzene case present a  $d^0$  configuration. Nevertheless, they were analysed along with the cyclopentadiene structures, giving similar trends in the CShM values.

The shapes of the  $ML_3$  fragments in all these structures were analysed using a continuous shape map described in previous chapters considering the vacant tetrahedron (vT-3) and the fac-trivacant octahedron (fvOC-3) as the two reference polyhedra. All experimental structures lie almost on the minimum distortion pathway with a few exceptions. Therefore, the minimum distortion path was enlarged further away from the reference structures to explain two phenomena (figure 5a).

The first one is that some structures with the Cbd and Bz aromatic rings present a high deviation from the curve because they contain non-equivalent ligands ( $ArM(L_1)_2L_2$ ) which break the  $C_3$  symmetry of the  $ML_3$  fragment. The case of the distorted cyclopentadiene complexes is harder to explain, however, by looking into the crystal structure determination (figure 4), the two highly distorted molecules (top-right dots,

CSD ID: CPHXCU01, UDOLIK) seem to include another ligand in the coordination sphere giving molecules that do not have a  $C_3$  symmetry ( $ML_4$  fragment). In the other cases with structures that do not fall close to the minimum distortion path the reason is that they present significantly different c-M-L non-equivalent angles, i. e. the structure with CSD ID: JOGLOH, with  $113^\circ$ ,  $137^\circ$ ,  $109^\circ$  values.

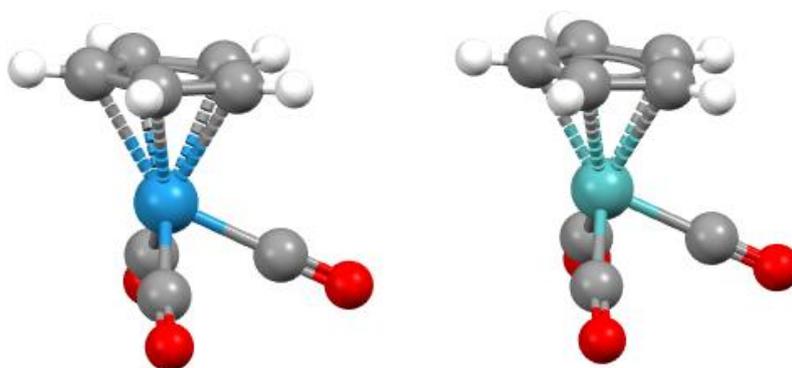


Figure 4. CPHXCU01 (left) and UDOLIK (right) structures.

The second reason of expanding the minimal distortion path beyond the two reference shapes is to explain the points on the right of the octahedral reference shape, with  $S(vT-3) > 2.7$  values. The minimum distortion path (figure 5a) is simple given by changing the angle that the ligands have with the  $C_3$  rotational axis from  $109.4^\circ$  ( $vT-3$ ) to  $125.3^\circ$  ( $fvOC-3$ ). By increasing or decreasing this angle further from the values corresponding to the ideal polyhedra we obtain a continuation of the minimal distortion path as shown in figure 5a. It can be clearly seen that structures that did not fall on the minimal distortion path lie, however, along the prolonged curve, meaning that the outsider structures are close to the octahedral shape but with X-M-L angles larger than  $125.3^\circ$ .

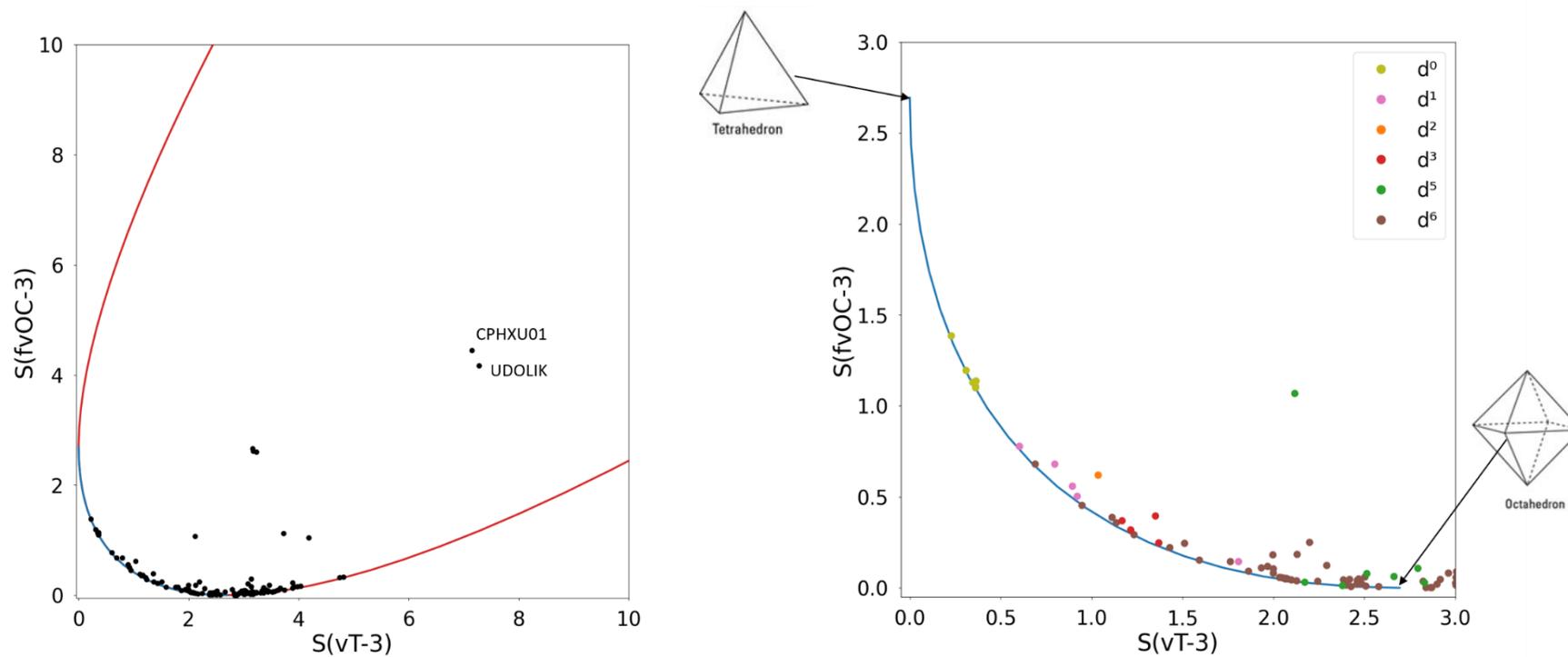


Figure 5. a) Shape map for the experimental CpML<sub>3</sub> structures (black dots). The blue line corresponds to the minimal distortion path along the distorted angle XML. b) vT-3 (vacant tetrahedron) / fvOC-3 (fac-Trivacant Octahedron) shape map for the experimental CpML<sub>3</sub> structures. The blue line corresponds to the minimal distortion path between the two reference shapes. Shape measures for different molecules are shown as solid circles, using different colors to group them according to the number of electrons in the d-type orbitals of each compound.

A zoom in the shape map (figure 5b) shows the minimum distortion path between  $vT_3$  and  $fvOC_3$  reference structures represented by a solid blue line, while the different sets of structures are represented by a scatter plot, grouped by colour according to their d-orbital electron configuration. As noted in reference 7, a general trend showing the evolution from a more tetrahedral to a more octahedral shape as the number of electrons increases, is found, even though some structures lie at some distance from the curve because the  $ML_3$  fragment is not perfectly symmetric as they present either different M-L bond lengths or c-M-L angles.

Nevertheless, these initial evidences are sufficient to manifest that the shape of these molecules is somehow controlled by the filling of the d orbitals of the central metal, which means that  $d^0$  cyclopentadiene organometallic complexes should present a tetrahedral shape while  $d^6$  compounds will exhibit a more octahedral shape.

## Computational models

In order to rationalize the structural trends that appear in figure 5, we have considered different simplified computational models to avoid possible distortions in the molecular shape due to intermolecular interactions in the experimental crystal structures and to get idealized representative structures for each electronic configuration, including also hypothetical  $d^4$  compounds

The first set of molecules whose structure was optimized were the neutral  $CpMCl_3$  organometallic complexes with  $M = Ti, V, Cr, Mn, Fe, Co$  and  $Ni$ . The resulting CShM values are shown in the shape map in Figure 6a, where the different structures almost lie on the minimum distortion path, with a similar distribution along the path as found for the experimental structures. The higher the number of electrons in the d-type MOs, the shape of the  $ML_3$  fragment changes gradually from tetrahedral to octahedral. However, the  $d^1$ ,  $d^4$  and  $d^6$  structures present a deviation from the minimum distortion path due to the loss of the  $C_3$  symmetry given by the distortion of the  $L_3$  triangle in the DFT optimization as can be seen by the values of the deviation from the path (see chapter 1) in table 1. The deviation function measures how far does a point ( $S_Q(P)$ ,

$S_Q(R)$  lie from the minimum distortion pathway between reference shapes P and R. The distortion that takes the structures away from the minimal distortion path appears due to a displacement of the cyclopentadiene ligand during the optimization, which causes that the  $C_3$  and  $C_5$  rotational axis do not coincide, thus modifying the equivalence of the X-M-Cl angles, where X is the centre of the ring.

	deviation	$X-M-Cl$ ( $^\circ$ )		
		$L_1$	$L_2$	$L_3$
$d^1$	9.05	112.60	112.60	124.57
$d^4$	10.18	120.31	120.31	118.45
$d^5$	12.08	122.21	122.21	118.02

Table 1. Path deviation values function for the optimized structures for molecules with  $d^1$ ,  $d^4$  and  $d^5$  electronic configurations and their three different X-M-Cl angles.

Freezing the  $C_3$  symmetry of the  $L_3$  fragment as well as the  $C_5$  symmetry of the cyclopentadiene by z-matrix construction in the optimization process gives structures with energies very close to the minimum energy structures (table 2) that reflect, however, the trend in the shape change much better since now all structures lie on the minimum distortion path (figure 6b).

d electrons	$E [C_3, C_5 \text{ opt.}] - E [\text{full opt.}]$ (eV)
0	0.05
1	0.13
2	0.05
3	0.05
4	0.10
5	0.11
6	0.08

Table 2. Energy difference (eV) between  $CpMCl_3$  molecules with and without freezing angles.

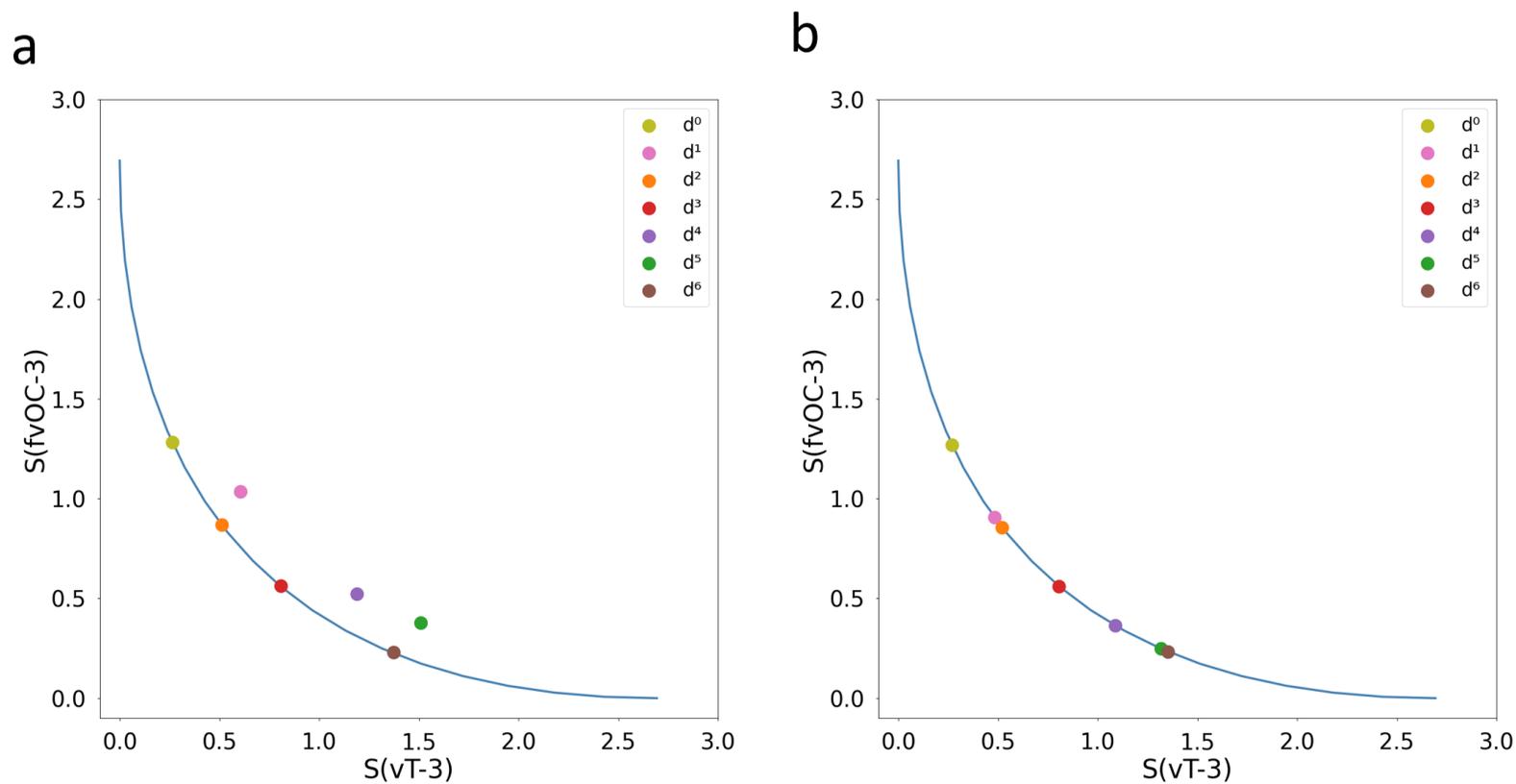


Figure 6. vT-3 (vacant tetrahedron) / fvOC-3 (fac-Trivacant Octahedron) shape maps for the optimized  $\text{CpMCl}_3$  structures. The blue line corresponds to the minimal distortion path between the two reference shapes. Shape measures for different molecules are shown as solid circles, using different colours to highlight the number of electrons in the d-type orbitals of each compound. (a) Shape map for fully optimized structures and (b) shape map for structures optimized fixing the  $C_5$  and  $C_3$  rotational symmetry around the same axis for the Cp and  $L_3$  ligands, respectively.

The distributions of shape measures obtained for the experimental structures are compared to those obtained for the symmetrized models in the box plots shown in figure 7. The trends of the medians for the experimental structures coincides well with that found for the model, confirming the tetrahedral to octahedral shape change as the number of electrons in the d-type MOs increases. The box plots for the experimental data sets show also that the dispersion of results is larger for compounds with a high occupation of the d-type MO's, especially for the most usual  $d^6$  systems.

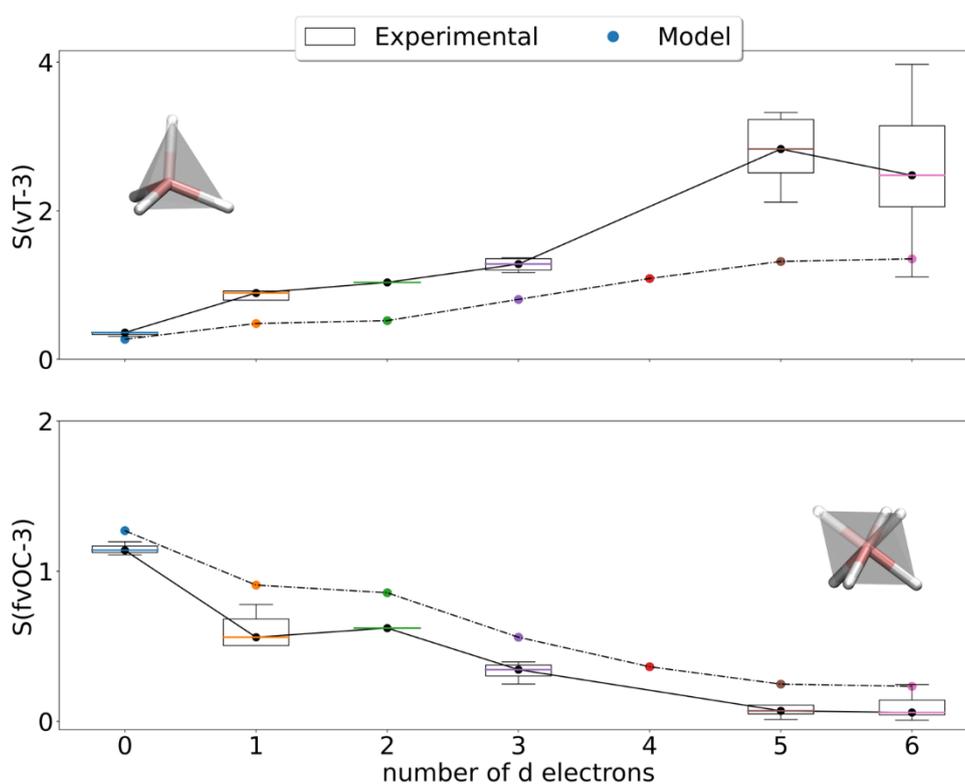


Figure 7. Distribution of vacant tetrahedral (top) and fac-trivacant octahedral CShMs for the pyramidal  $ML_3$  fragments found for experimental compounds (box diagrams) and the optimized structures for the symmetrized model (dots). Lines joining data for different electron counts are just meant as a guide to the eye to highlight the observed trends.

The effect of the ligand-field splitting, which represents the energy difference between the two degenerate sets of d-type MOs in crystal field theory, was analysed by

exploring models with different ligands in the symmetrized computational model. Considering the spectrochemical series, based on the absorption spectra's proposed by Tsuchida<sup>16</sup>, chlorine is a weak field ligand and in figure 6 shape measures for the different electron counts appear closely grouped in the middle segment of the path between the tetrahedron and the octahedron. The effect of a strong field ligand such as cyanide is to spread the measures along the whole minimum distortion path, displaying highly symmetric shapes at both ends, with practically a tetrahedral structure for the  $d^0$  case and a practically octahedral one for the  $d^6$  one (Figure 8a). Other ligands with an intermediate field strength such as methylene (Figure 8b) show an intermediate spread of the measures between the limiting weak and strong field cases of Cl and CN, respectively.

Besides the changes induced by the type of L ligand, it is also important to evaluate if the trends found for Cp also hold for other aromatic ligands. For this purpose we have included in our computational models also compounds with the cyclobutadiene (Cbd) and benzene (Bz) ligands, computing CShM values for geometries optimized considering coincident  $C_3$  and  $C_4$  or  $C_6$  symmetries for the  $L_3$  triangle and the aromatic ligands, respectively.

In both cases, we used a symmetrized model such as the one used for  $CpMCl_3$  but constraining the symmetry of the aromatic ring to fit that of the corresponding regular polygon. Additionally, to keep the neutrality of the molecule we also considered Sc to act as a  $d^0$  metal in the Bz series and Cu as the  $d^6$  metal for the Cbd one.

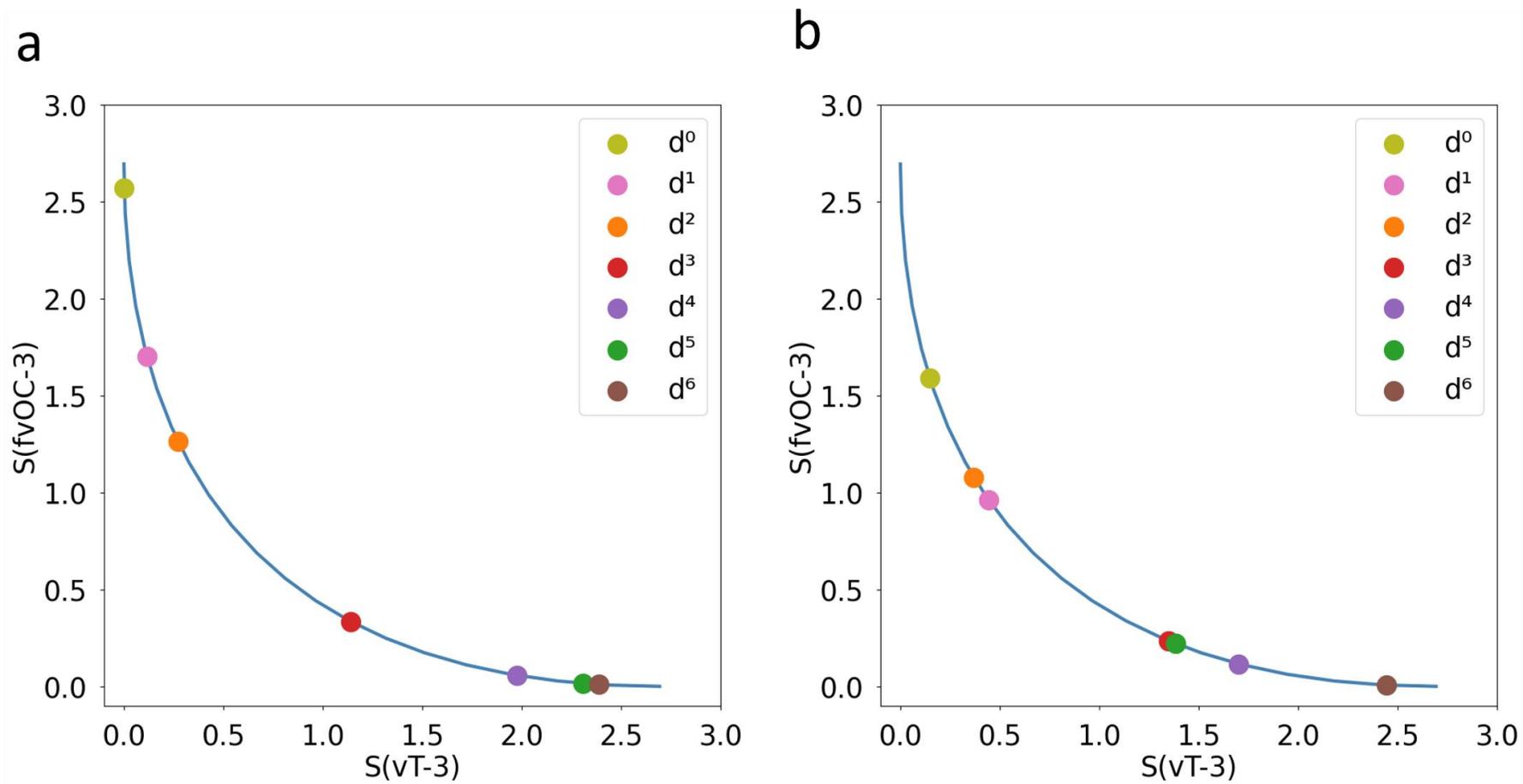


Figure 8. Shape measure comparison of the cyanide (left) and methylene ligands (right) for the vT-3 and fvOC-3 shapes.

The results for compounds with the square cyclobutadiene ligand differ somewhat from those obtained for compounds with the Cp ligand. In contrast to what we found for compounds with the Cp ligand (figure 7), for cyclobutadiene the  $d^6$  molecule is now the one with a higher tetrahedral shape content (figure 9). The trend of increasing octahedral shape with the filling of the d-block orbitals is more or less kept for structures with d-electron counts between  $d^0$  and  $d^5$  if we disregard the  $d^2$  case for which we find a shape that is more octahedral than those of the  $d^3$  and  $d^4$  compound. It is also important to point out that the octahedral shape measure in this series is never below 0.5, indicating that compounds with the Cbd ligand do not present shapes as close to the octahedron as in the case of the Cp ligand, irrespective of the filling of the d-block orbitals. This finding is, however, not surprising from a merely geometrical point of view since it is hard to imagine on how to distribute the three more or less equivalent points on a regular square that would be necessary to complete an octahedral coordination sphere.

Regarding the compounds with the six membered benzene ligand, the shape measures reveal that the trend is similar to that found for compounds with the 5-vertices Cp ring as the shape variation with the electron filling shows a gradual change towards the octahedron, where the lower  $d^0$  electron systems are more likely to have a tetrahedral shape while the  $d^6$  are more prone to be octahedral, although the  $d^1$  and especially the  $d^5$  cases do not follow the general trend. For complexes with the benzene ligand, for which it is easy to think of a regular disposition of three points in space to complete an octahedron with the  $L_3$  triangle we find again that at least for the  $d^6$  structures it might be sensible to consider a octahedral polyhedral model for the coordination sphere around the metal.

Although the results for these two aromatic ligands were not perfectly in line with those obtained for the more usual cyclopentadiene complexes, our model displays a relation between the shape and the electronic structure on the piano-stool molecules, with the preference between more tetrahedral/octahedral structures dictated mainly by the d-electron count. Sadly, as explained before, the number of experimental structures for these aromatic rings are insufficient to check if there is a trend similar to the computational results as no structures with configurations between 1 and 5 electrons are found.

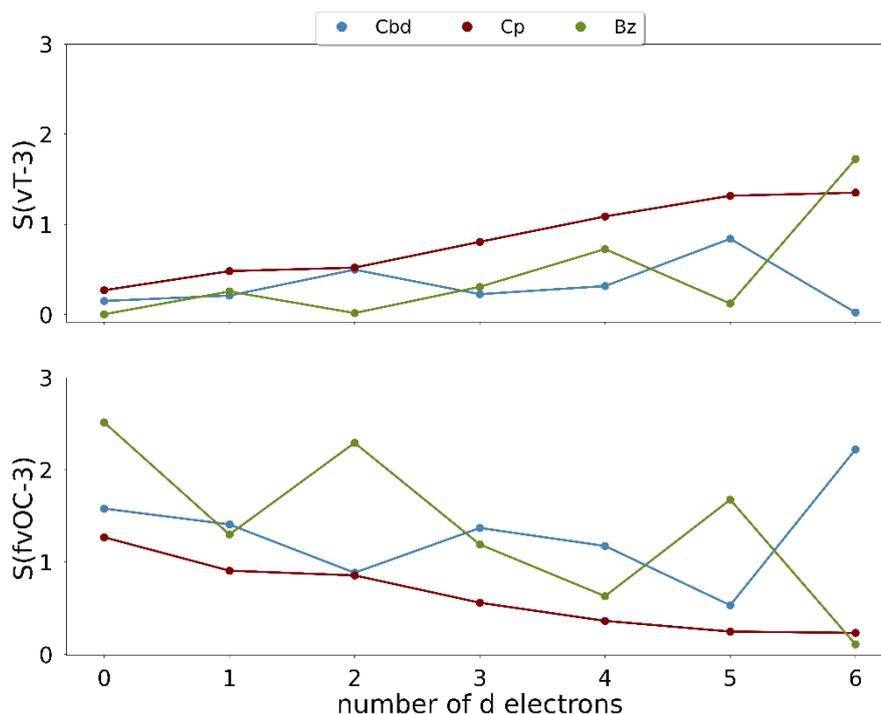


Figure 9.  $(\text{Cbd})\text{MCl}_3$  (blue),  $\text{CpMCl}_3$  (red) and  $\text{BzMCl}_3$  (green) shape measures for vT-3 and fvOC-3 ideal shapes.

Additionally, the rotation of the aromatic ring with respect to the  $\text{L}_3$  triangle was also considered for Cbd, Cp and Bz rings, finding that there is no relevant change (figure 10) on the energy,  $E < 0.05$  eV, for different relative orientations of the Cp and  $\text{L}_3$  fragments if their respective  $\text{C}_3$  and  $\text{C}_5$  axes coincide. Therefore, the rotation of the aromatic ring

was not considered significant in this study and the eclipsed conformation is adopted for all the modelled molecules.

The results obtained comparing the full optimized structures and the symmetrized model account for the origin of the small deviations from the minimum distortion path found for the experimental structures which are due to small deviations from a perfect  $C_3$  symmetry for the  $ML_3$  fragment. Since the symmetrized model with coincident  $C_5$  and  $C_3$  rotational symmetries of the Cp and the  $L_3$  fragments, respectively, is cleaner and theoretically easier to understand, while remaining still close to the minimum energy structures, we will consider only this symmetrized model in our further analysis of the electronic structure later.

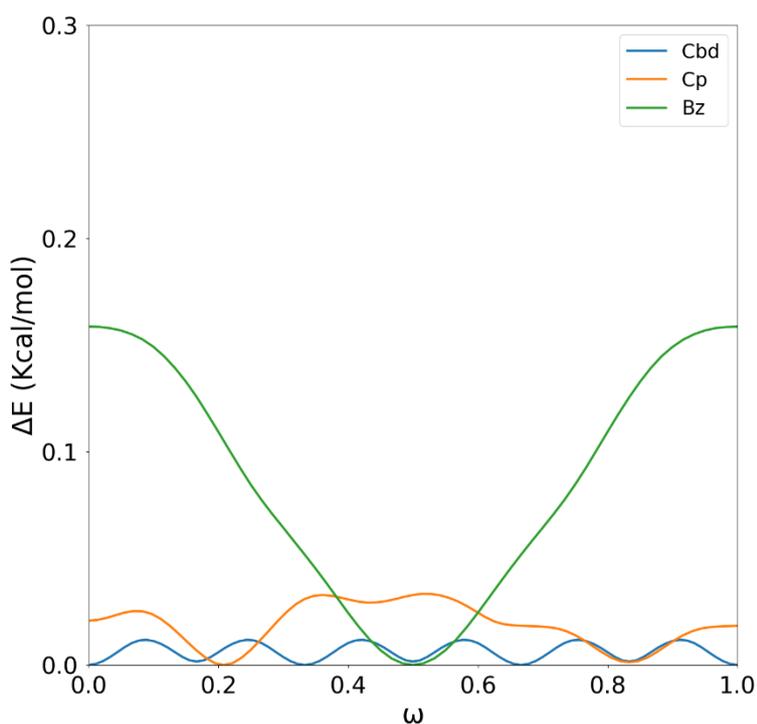


Figure 10. Variation of the rotational energy of the  $ArMCl_3$  molecules for a periodic rotation ( $\omega$ )

## Electronic structure

The previous results show for both the computational models and the experimental structures that the shape of the  $ML_3$  fragment is somehow controlled by the electronic configuration of the d-type MOs of the complex and that this feature is the key factor that guides the atom arrangement around the metal ion, as long as the  $L_3$  fragment maintains the  $C_3$  symmetry and no bonds are formed between these ligands that can difficult the free movement of the three atoms coordinated to the central metal. In order to understand why the filling of these orbitals is responsible of such a phenomenon, different aspects of the electronic structure based on molecular symmetry as well as crystal field theory were analysed.

Removing the cyclopentadiene ligand, the  $ML_3$  fragment gives a good approximation to what represents the shape of a vacant tetrahedron or a fac-trivacant octahedron and the main parameter that differentiates between the tetrahedral and octahedral shapes is the angle that the coordinated atom on each L ligand forms with the metal (M) and the centroid (X) of the cyclic ligand, with  $109.5^\circ$  and  $125.2^\circ$  being the exact values corresponding to the tetrahedral and octahedral arrangements, respectively.

Although the shape of an object is independent of its size, in the case of molecules, the distances between atoms are important to determine the overlap between atomic orbitals and, hence, the shape and energies of molecular orbitals and the whole electronic structure for molecular systems.

In Figure 11, the metal-to-ring distance and the centroid-metal-ligand angle for the optimized structures are represented for compounds with two different L ligands, chlorine and methylene, as a function of the number of electrons in the d-block MOs. Both cases show qualitatively similar trends that are independent of the nature of the ligand. As found from the shape measures for the  $ML_3$  fragment, the angle X-M-L increases with the number of electrons, moving from angles around  $115^\circ$  close to the tetrahedron to angles about  $125^\circ$  corresponding to an octahedral shape. In addition, the metal-to-ring distance, considering the centroid of the ring, gives more details of the interaction of this ligand with the central metal atom, with the ring moving closer

to the metal as the number of electrons in the d-type MOs increases. It should be noted that no significant variations ( $d \leq 0.1 \text{ \AA}$ ) were found in the M-L distances either for chlorine or methylene if we compare them with the  $\sim 0.25 \text{ \AA}$  change found for the metal to cyclopentadiene distance along the series for both Cl or  $\text{CH}_3$ .

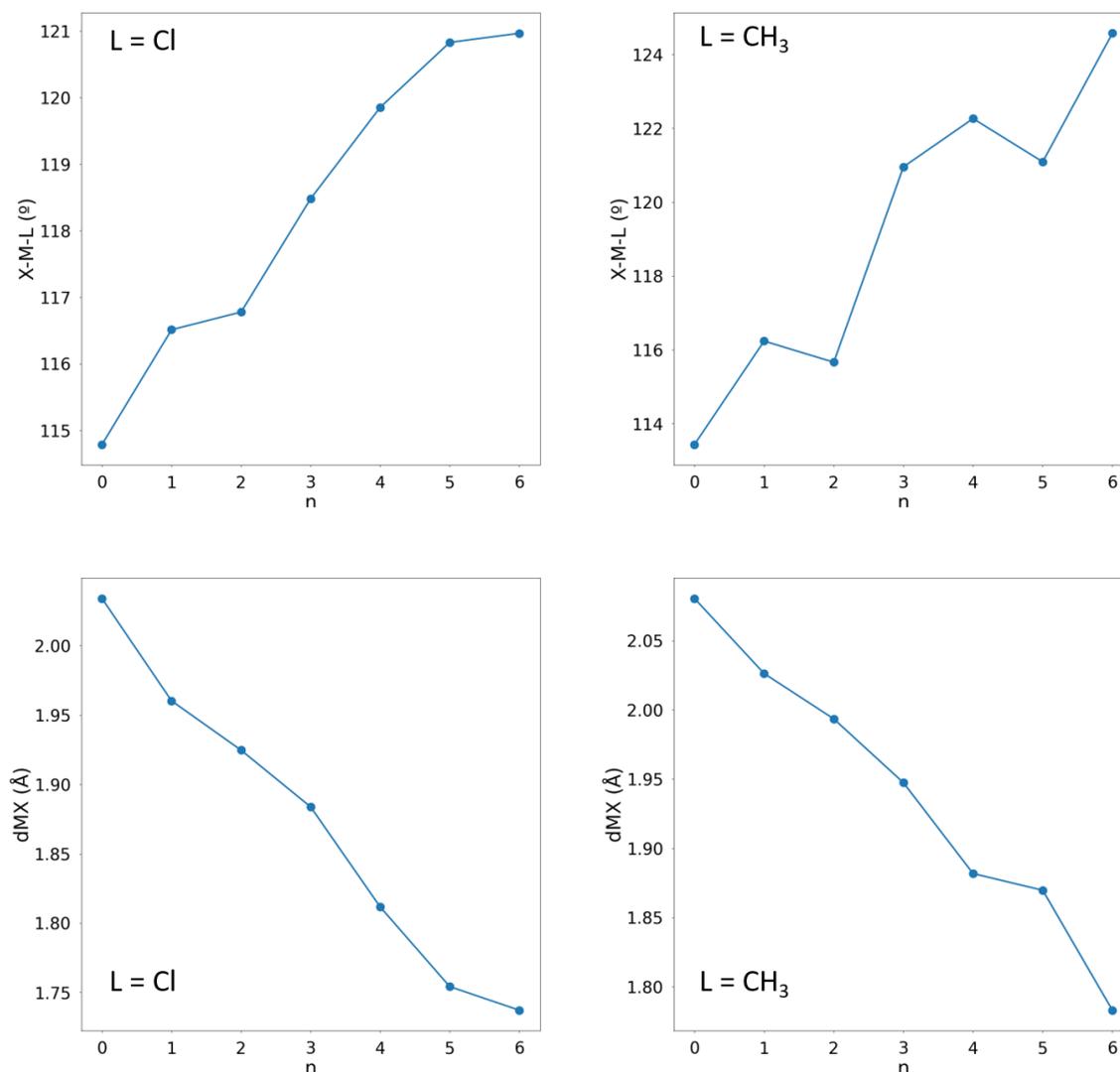


Figure 11. XML angle (top) of chlorine (left) and methylene ligands (right) and XM distance (bottom) for the piano-stools model with freezing the  $C_3/C_5$  symmetry.

Besides the purely geometrical aspects, a change from a tetrahedral to octahedral shape is also expected to induce changes in the electronic structure. The most evident differences in the d-block MOs of a tetrahedral and an octahedral complex are in the

different patterns in breaking the orbital degeneracy of an isolated transition metal atom. The five d orbitals of a transition metal atom are degenerated when the atom is completely isolated. However, if we switch on the interaction with its environment, the set of five d-type orbitals splits in different patterns, depending on the symmetry of the crystal field around the central atom (Figure 12). For a tetrahedral symmetry one finds a splitting of the d-type orbitals into two sets of 3 over 2 degenerated orbitals transforming to the  $T_2$  and  $E$  irreducible representations of the  $T_d$  group, respectively. In an octahedral field, on the contrary, we find a 2 over 3 splitting of degenerate orbitals with  $E_g$  and  $T_{2g}$  symmetries. In addition, in crystal field theory, the energy gap between both sets of d-type orbitals for a tetrahedral field ( $\Delta_{tet}$ ) is found to be roughly  $4/9$  times the gap found for an octahedral field ( $\Delta_{oct}$ )<sup>17</sup>.

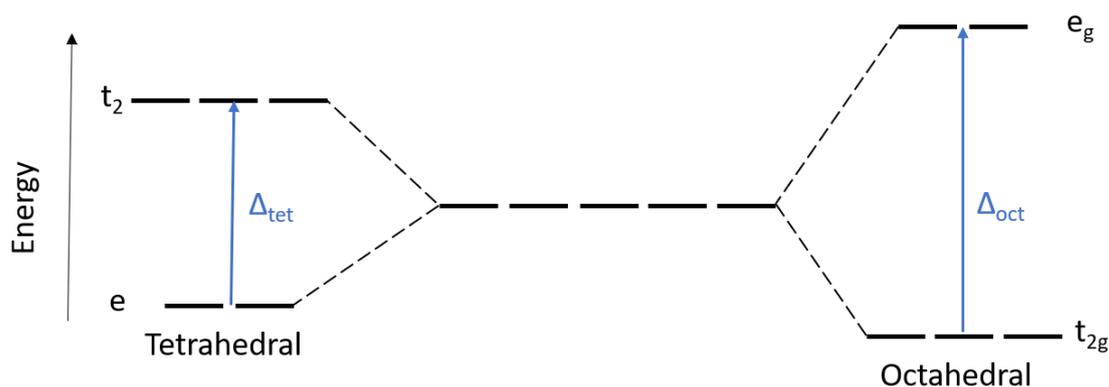


Figure 12. Tetrahedral ( $\Delta_{tet}$ ) and octahedral ( $\Delta_{oct}$ ) energy splitting of the d molecular orbitals.

The actual symmetry of the crystal field in piano-stool complexes is of course neither tetrahedral nor octahedral and the precise splitting pattern will resemble only approximately one of the two limiting cases if we consider that the aromatic ring acts either as a single coordination site or as three coordination sites. Using symmetry arguments, for  $CpML_3$  compounds, with just  $C_{3v}$  symmetry we expect the metal's d orbitals to split into five nondegenerate MOs as seen in figure 12. The d orbital energies obtained for the  $TiH_3$  at the extended Hückel level are plotted as the difference in energy with respect to the metal's d orbital energy resulting into a splitting of nearly 2

over 3 MOs similar to the octahedral splitting (figure 13). However, the addition of an aromatic ring splits the almost triply degenerated set of the lower d orbitals resulting into an approximated 2, 1, 2 splitting pattern. It should be noted that the Cbd ring results in a large breaking of the degeneracy of the  $xy / x^2-y^2$  and  $xz / yz$  orbitals making the 2, 1, 2 splitting pattern less evident. For the other two rings these pairs or orbitals are practically left forming a degenerate set even if strictly speaking. As seen in the figure, the expected 2, 1, 2 splitting pattern for the orbitals is quite different from the 2 over 3 or 3 over 2 patterns expected for tetrahedral or octahedral complexes, respectively.

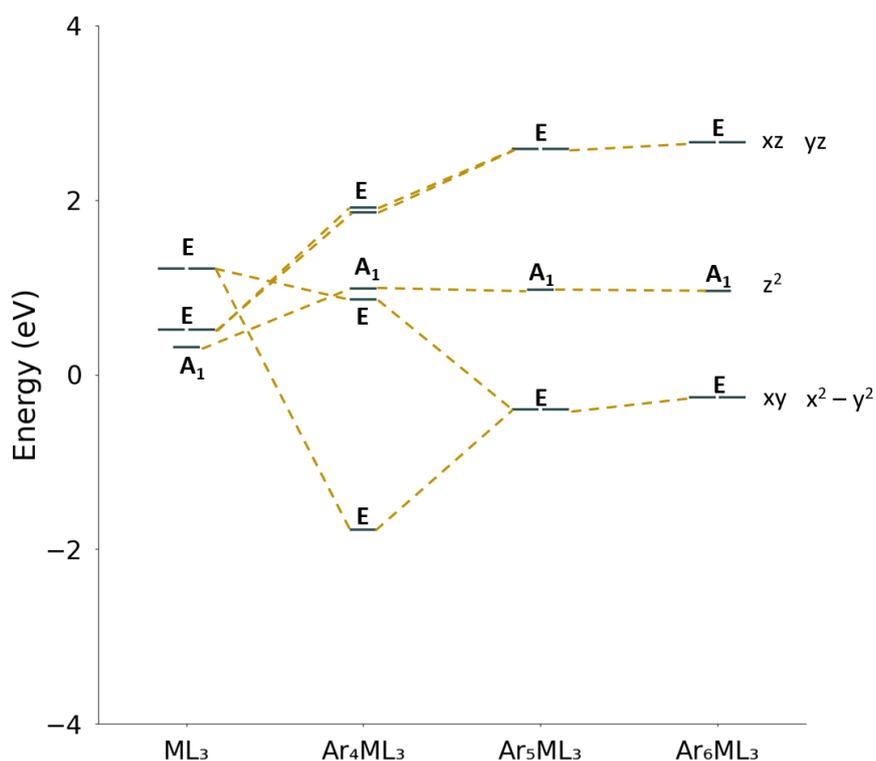


Figure 13. Splitting d orbital energy for the ML<sub>3</sub> fragment and the coordination complexes Ar<sub>n</sub>ML<sub>3</sub> with n=4, 5,6. In figure, M = Ti, and Ar<sub>4</sub>, Ar<sub>5</sub>, and Ar<sub>6</sub> are the Cbd, Cp and Bz cyclic rings, respectively. The symmetry labels are obtained from the classification of these orbitals into the C<sub>3v</sub> point group. The parameters MX, ML and XML were maintained equal in all cases with values of 2.0 Å, 2.2 Å and 114.8°, respectively.

Let us remember, however that the diagrams shown in figure 13 are merely qualitative since they have been obtained from extended Hückel calculation with the same parameters and a given set of M-L, X-M, and X-M-L distances. Changes in the parameters and geometries can lead to situations where the energy of the central  $A_1$  orbital is closer to either the lower or the higher two-orbital sets, yielding splitting patterns that are more similar to those expected for compounds with octahedral or tetrahedral symmetry, respectively.

If we come back to the DFT results for the optimized models, the energies of the five molecular d-type orbitals for  $\text{CpMCl}_3$  and  $\text{CpM}(\text{CH}_3)_3$  are shown as a function of the number of d-electrons in figure 14. For both cases, the splitting pattern is similar, with a rough separation into two sets with 2 and 3 orbitals with similar energies, a splitting more similar to the 2 over 3  $e_g/t_{2g}$  pattern of octahedral complexes than that of tetrahedral ones. Since there are not multidimensional representations in the  $C_s$  group, none of the five d-type orbitals are strictly degenerated, but the splitting in the upper 2 orbital set or in the lower three orbital set is much smaller than that between the two sets. Although the 2 over 3 “octahedral” splitting pattern is found for any number of d-electrons, it is easy to see that this pattern is much clearer for compounds with 3 or more d-electrons than for compounds with lower electron counts. For the latter the lower lying 3 orbital set is really split into 2+1 orbitals, indicating that the description based on the octahedral symmetry is more adequate the higher the number of d-electrons. Besides this, for the  $d^6$  molecule, the splitting between the two blocks is almost 4/9 times the splitting in the  $d^0$  molecule, adding an additional evidence to the tetrahedral to octahedral shift in the electronic structure of these compounds as the number of d-electrons is increased, even if the trend is not evident at all looking only at the splitting pattern. Moreover, the splitting in the cyanide case is larger than in the case of methylene, a result expected in crystal field theory as the cyanide ligand is a strong-field ligand which implies larger splitting energies between the d orbitals.

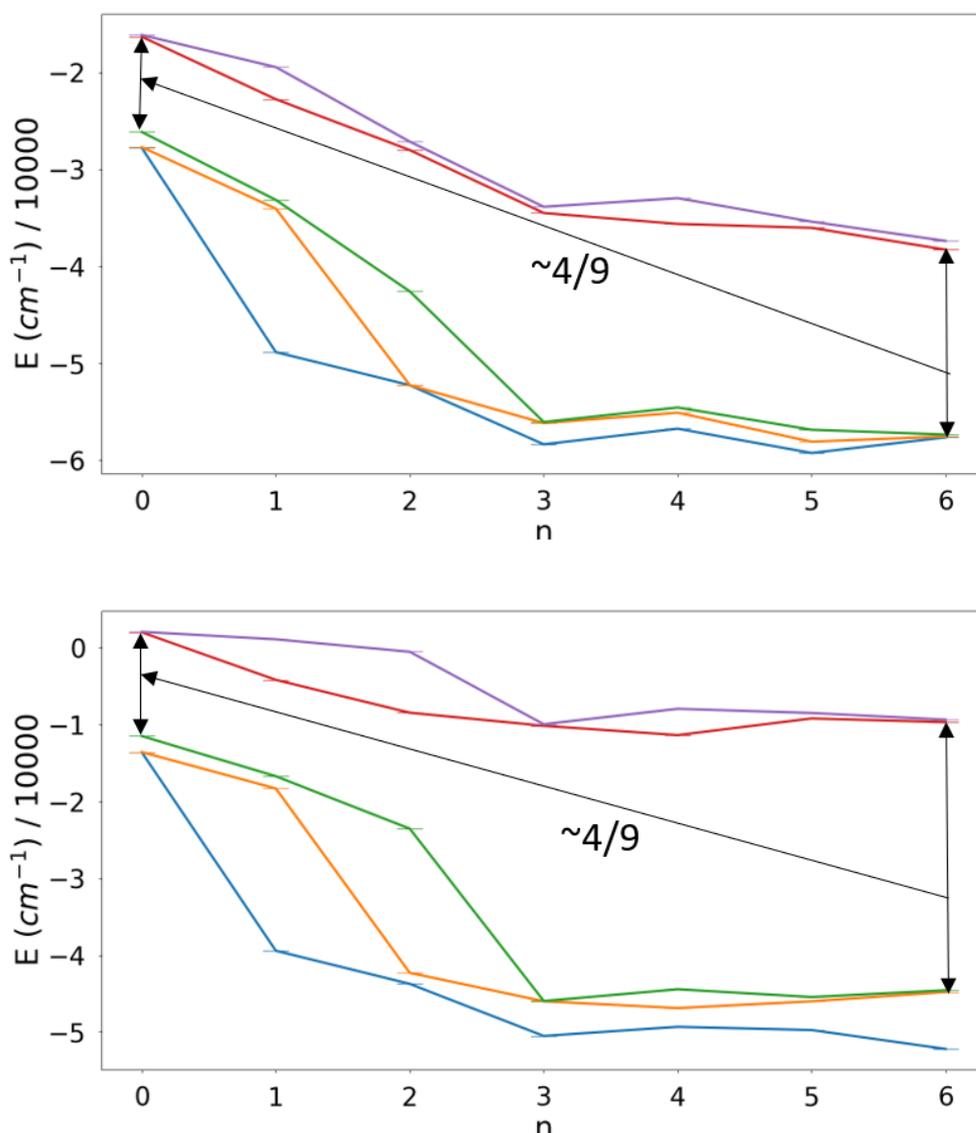


Figure 14. Molecular d-orbital energies for CpML<sub>3</sub> compounds with chlorine (top) and methyl (bottom) ligand taken from the DFT calculations for the corresponding symmetric models.

Similar trends are found for piano-stool complexes with the Cbd and Bz aromatic ligands (figure 15). Note, however, that for the Cbd case a large splitting is found between the two upper MOs in the d-block for the d<sup>6</sup> compound and as a result, the approximate 2 over 3 orbital pattern is changed to a 1 over 1 over 3 scheme for this particular case.

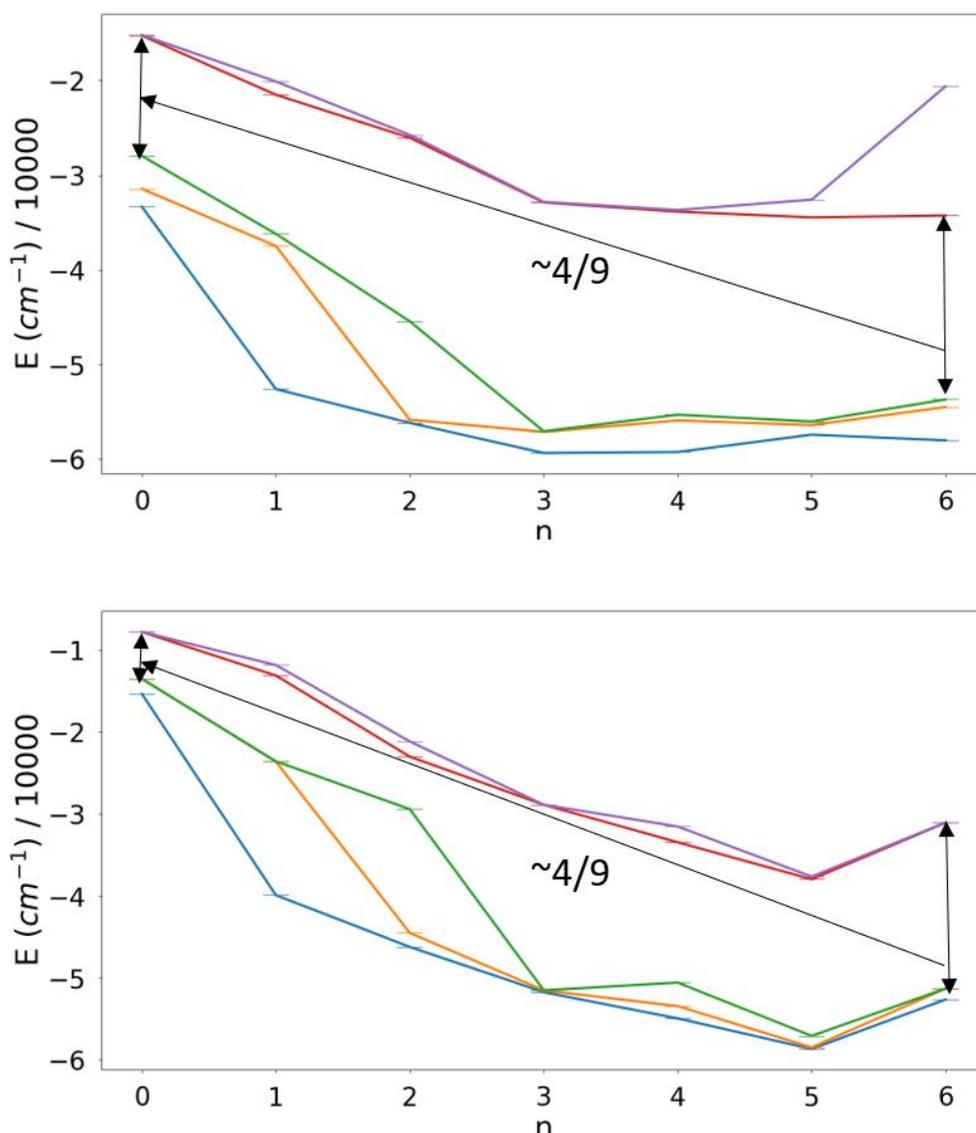


Figure 15. Molecular d orbital energies for  $(\text{Cbd})\text{MCl}_3$  (top) and  $\text{BzMCl}_3$  (bottom) compounds taken from the DFT calculations for the corresponding symmetric models.

The analysis above is based on using different transition metal atoms to change the filling of the d-type MOs, showing an overall relation between the gap separating the 2- and 3-orbital sets and the continuous shape measure for the  $\text{ML}_3$  fragment (related basically to the X-M-L angle). The use of *ab initio* calculations with different metals to preserve the overall charge in the series of complexes difficult, however, a qualitative understanding of the observed trends since changing the metal (and consequently the X-M and M-L distances) and not only the electron count introduces more sources for

different behaviour along the series. It is, for instance, difficult to compare the trends in the Ar-M and M-L interactions for molecules with different metals. Moreover, the open shell nature for most of the systems along the series originates an additional problem related to the difference in symmetry contents between alpha and beta molecular spin-orbitals, making the comparison between MOs in molecules with different occupations of the d-block orbitals quite difficult.

For this reason, in order to find qualitative general trends and to analyse the effect of the d-electron count on the interactions between the three fragments (Ar, M and  $L_3$ ), we will perform a detailed pseudosymmetry analysis using the Extended-Hückel method. In this method the molecular charge does not affect at all the calculation of the molecular orbitals so that we can construct a qualitative molecular model where the central metal is unchanged and all effects of changing the electron count arise just from filling a different number of d-type MOs while all metal – ligand interactions remain the same. In order to simplify further our model, we have used  $[\text{CpTiH}_3]^{-n}$  with  $n = 0-6$  for our initial analysis to limit the complexity of M-L interactions and possible L...L steric effects to its minimum. The geometry used in this calculation was obtained from a geometry optimization in *gaussian 09* for the  $\text{CpTiH}_3$  structure with the same DFT method and basis set as in the previous molecules, which correspond to the  $d^0$  configuration. The optimization gave M-X and M-L distances of 2.0 Å and 2.2 Å, respectively. In this case, the angle X-M-L obtained was 114.8.

The interactions between the d AOs on the metal with those on the ligands is better understood by building the orbital interaction diagram in two steps, considering first the interaction in the Cp-M fragment and later the interaction of this fragment with the  $L_3$  triangle. The differences between the “tetrahedral” or “octahedral” shapes for the whole molecule affect only the second step, that is, the interaction of the Cp-M fragment with the  $L_3$  set of ligands since it depends only on the X-M-L angle (and the L-L distances in the  $L_3$  fragments which depend on the X-M-L angle if we consider fixed M-L distances in our model).

In the molecular orbital diagram for the Cp-M interaction (figure 16), where the z axis is chosen to coincide with the X-M direction, it is easy to see how four of the five atomic d-type orbitals of the titanium atom,  $xy$ ,  $x^2-y^2$ ,  $xz$ , and  $yz$  change their energies due to

$\pi$ -type bonding and anti-bonding interactions with the ring's  $E_2$  and  $E_1$  MOs giving rise to two degenerated pairs with  $(xy, x^2-y^2)$  and  $(xz, yz)$  character respectively. The  $z^2$  orbital, which has the right symmetry to interact with the lowest lying completely bonding  $\pi$ -type MO of A symmetry of the ring, does however have a poor energy match with this orbital, so that it barely combines with the ring orbital giving an essentially non-bonding metal-centred orbital at the same energy of the d-type orbitals of the isolated metal.

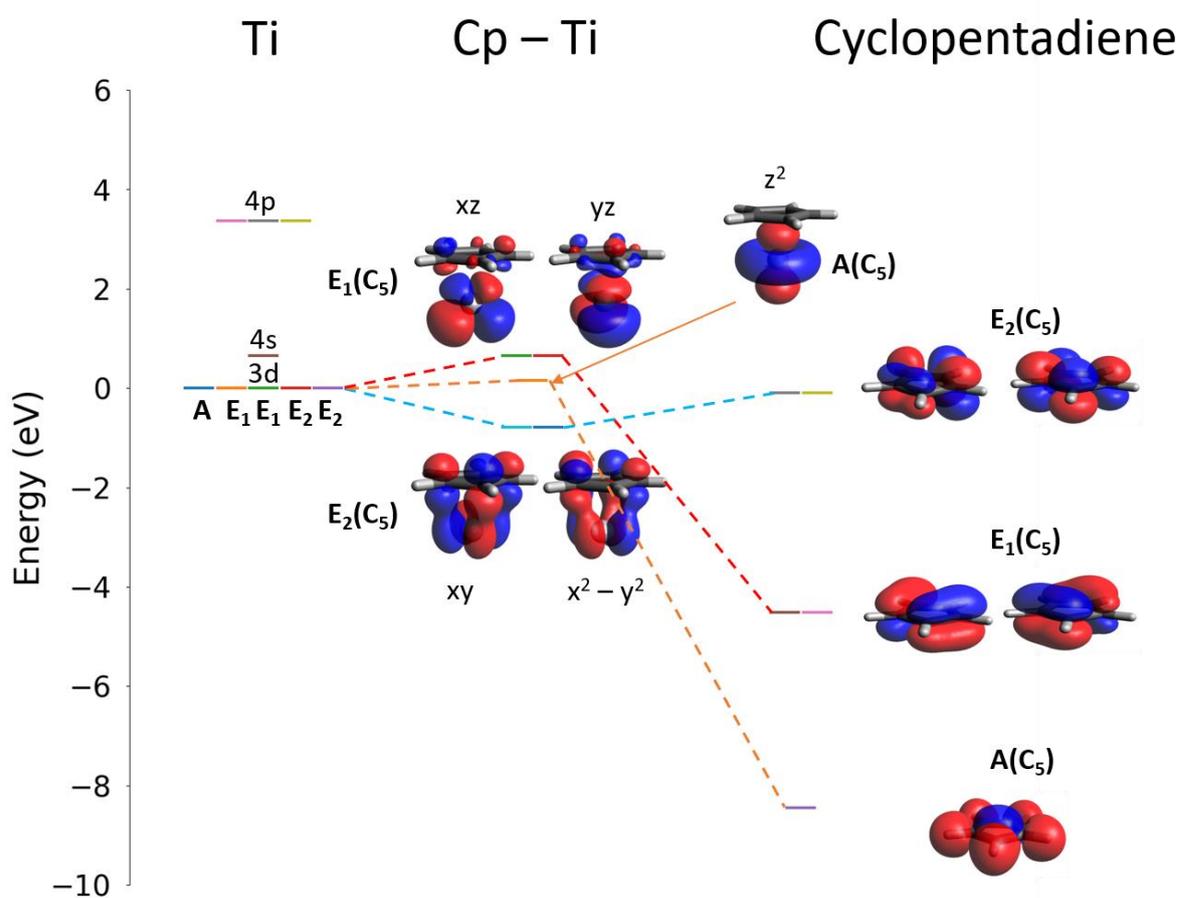


Figure 16. Orbital interaction diagram between the valence AOs on a titanium atom and the  $\pi$ -type MOs of the cyclopentadiene ligand. Symmetry labels for molecular orbitals are for the irreducible representation of the  $C_5$  symmetry group. The zero for the energy scale corresponds to the orbital energy of the d-type AOs of an isolated titanium atom.

As shown in figure 17, the five d-type orbitals of the Cp-Ti fragment interact with the three MOs obtained from the combination of the three 1s atomic orbitals of the H<sub>3</sub> fragment. These three MOs split into two sets, one containing the more stable totally bonding A orbital and the other with two degenerate H-H anti-bonding combinations of symmetry E.

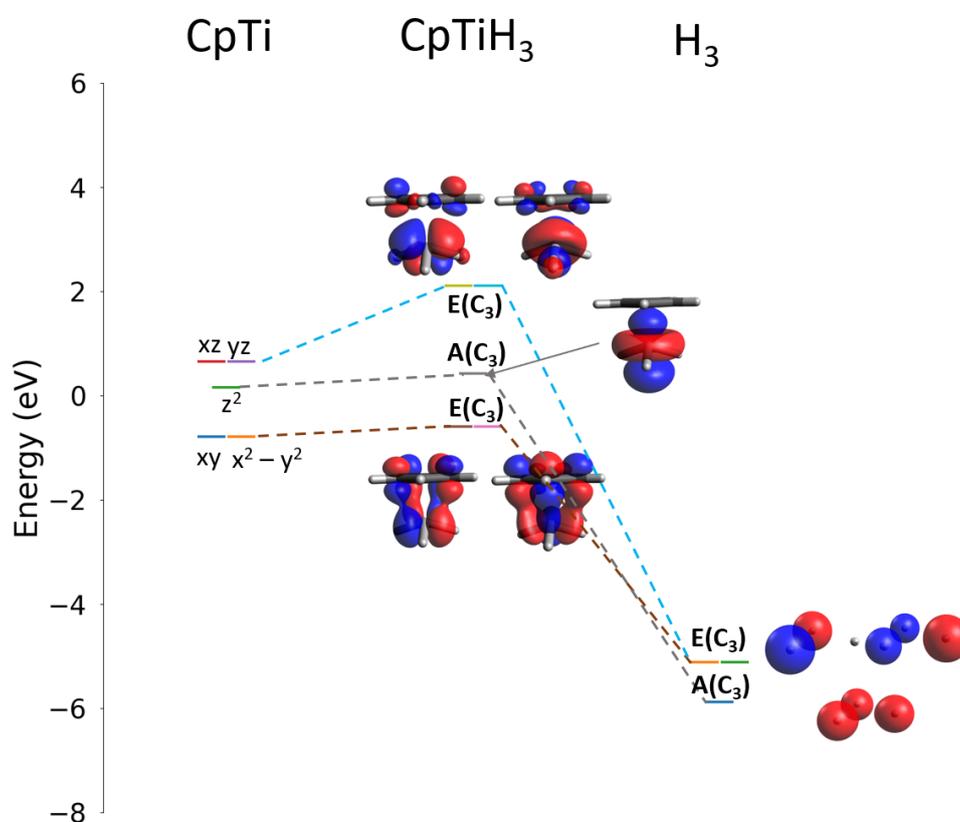


Figure 17. Molecular orbital interaction between the Cp-M d orbitals and the molecular orbitals for the H<sub>3</sub> fragment of the structure corresponding to the tetrahedral shape. The zero for the energy scale corresponds to the orbital energy of the d-type AOs of an isolated titanium atom.

Due to the lack of a common rotational symmetry around the Ti-X axis all orbitals of the two fragments are allowed to mix, although the most important interaction leads to a sizeable destabilization of the (xz,yz) MOs through interaction with the degenerate  $E(C_3)$  pair on the  $H_3$  fragment. Interaction with these  $H_3$  orbitals leads to a slightly smaller destabilization of the (xy,  $x^2-y^2$ ) pair on the Cp-Ti fragment, while the  $z^2$ -type orbital is also slightly destabilized by interacting with the totally symmetric MO of the  $H_3$  fragment.

This case is a paradigmatical example where a pseudosymmetry analysis of the molecular orbitals may help in understanding how the orbitals of fragments with different symmetries mix together. The symmetry group for the whole molecule is the  $C_s$  group containing just the identity operation and a reflection. The two interacting fragments have higher local rotational symmetries, a local  $C_3$  and a local  $C_5$  rotation axis, that are, however, incompatible between them and for this reason do not appear as global symmetry operations although sharing the same axis. Moreover, the local coordination environment around the metal atom may be analysed using an even higher symmetry group such as  $T_d$  or  $O_h$ .

To start, let us take the  $C_5$  group present in the Cp-Ti fragment as our pseudosymmetry group and determine to which irreducible representations of this symmetry group do the MOs of the triangular  $H_3$  fragment belong. Let us consider first a  $H_3$  fragment with an H-H distance of 1.72 Å, corresponding to the situation with an X-Ti-H angle of 109.5° (tetrahedral CpTiH<sub>3</sub> model). For this configuration, the main contributions to the orbitals of the triangular  $H_3$  fragment are 68%  $A(C_5)$  + 29%  $E_2(C_5)$  for the non-degenerated  $A(C_3)$  orbital and 64%  $E_1(C_5)$  + 33%  $E_2(C_5)$  for the degenerated  $E(C_3)$  set. From these figures we find that although the two rotational symmetries may seem totally incompatible from a geometrical point of view (if we consider only the position of the atoms), the MOs of the fragments actually present a quite large symmetry match. The non-degenerated  $A(C_3)$  orbital has a 68% of  $A(C_5)$  character, while the  $E(C_3)$  set presents a 64%  $E_1(C_5)$  character, allowing a good symmetry match with the  $z^2$ -type  $A(C_5)$  and the (xz,yz)-type  $E_1(C_5)$  set of the CpTi fragment, respectively. The (xy,  $x^2-y^2$ ) pair of  $E_2(C_5)$  symmetry of the CpTi fragment has the poorest match with orbitals on the  $H_3$  fragment which have only a ~30% of  $E_2$  character, explaining why the interaction

between the  $H_3$  and Cp-Ti fragments leads to a much larger energy change for the  $(xz, yz)$  set than for the  $(xy, x^2-y^2)$  one. These conclusions are not affected by the changes introduced in the H-H distances of the  $H_3$  fragment to adapt it to the octahedral model for the  $CpTiH_3$  compound, where the pseudosymmetry analysis gives a 76%  $A(C_5)$  + 22%  $E_2(C_5)$  decomposition for the non-degenerated  $A(C_3)$  orbital and 67%  $E_1(C_5)$  + 31%  $E_2(C_5)$  for the degenerated  $E(C_3)$  set, showing that the symmetry match is even better in this case, although not much different from the one in the tetrahedral model. The main conclusion from this pseudosymmetry analysis is that although the  $C_3$  and  $C_5$  symmetries may seem totally incompatible for geometrical objects, when it comes to functions that change smoothly in space, the symmetry match of the two objects is significantly better.

If we look at the changes in the orbital energies between both geometries, the effect of the variation of the Cp-Ti-X angle is only appreciable for the  $z^2$  and  $(xz, yz)$  orbitals, with the former being slightly stabilized and the latter destabilized. As explained above, due to the symmetry mismatch, the  $(xy, x^2-y^2)$  remains practically non-bonding, and consequently, its energy is practically independent from the geometry within the limit of variation between the tetrahedral and the octahedral shapes.

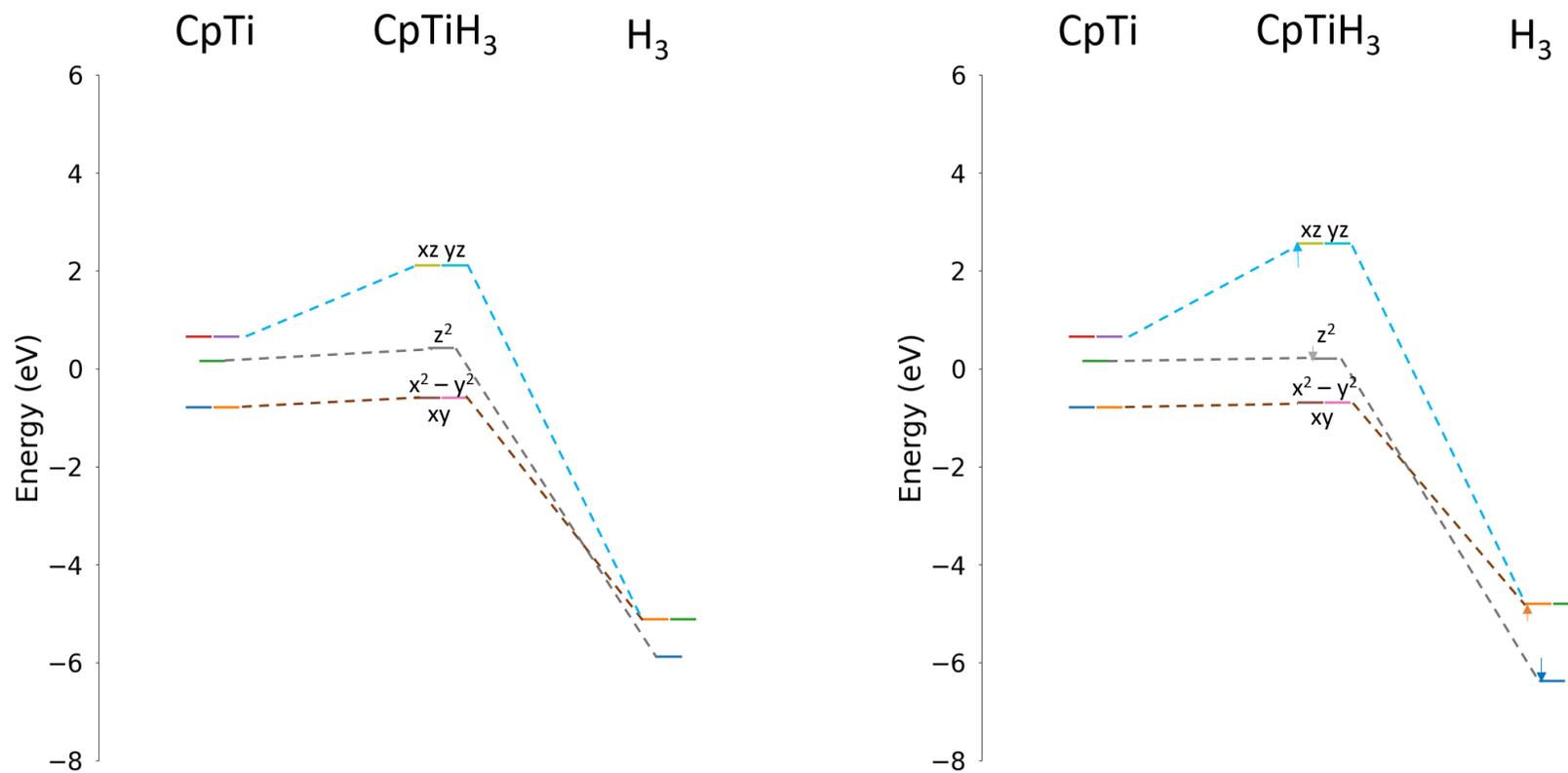


Figure 18. Molecular orbital interaction between the Cp-M d orbitals and the H<sub>3</sub> molecular orbitals for two structures, the tetrahedral shape (left) and the octahedral shape (right).

As explained above, the key geometric parameter that defines the molecular shape of piano-stool compounds as more tetrahedral- or octahedral-like is the angle between the centroid of the ring, the central metal, and one of the L atoms on the  $L_3$  fragment (X-M-L). To study the effect of changing this angle on the electronic structure, a set of different geometries for the model compounds were constructed by modifying this angle in a range between 100 to 130 degrees, covering the angles for the ideal tetrahedral ( $109.5^\circ$ ) and octahedral ( $125.2^\circ$ ) shapes, plotting the corresponding Walsh<sup>18</sup> diagram (figure 19).

To better visualize the energy changes in the molecular orbital energies induced by the change in geometry, orbital energies are given relative to their value for the initial geometry with X-M-L =  $100^\circ$ . Negative/positive values for the orbital energy represent, thus, a stabilization/destabilization of a given MO with respect to its energy for the X-M-L =  $100^\circ$  geometry. To simplify the Walsh diagrams, we did not represent the energy for each individual orbital, but joined them in different blocks, depending on their main bonding characteristics. In the CpTiH<sub>3</sub> model, the whole set of MOs can be divided into four main groups according to the interactions in which they are involved: the  $\sigma$ (C-C),  $\sigma$ (C-H),  $\pi$ (C-C), and  $\sigma$ (M-H) groups, with 5 MOs for the two former cases each, and 3 MOs in the two later. As seen in figure 19, the average energy for the  $\sigma$ (C-H) orbitals does not present any change as the angle varies, a behaviour that is fully consistent with its localized character on the Cp fragment in regions that are far away from the rest of the molecule. The sets of  $\sigma$ (C-C) and  $\pi$ (C-C) MOs, however, contribute each to an  $\sim 1$  eV stabilization of the total energy of the molecule each as the angle is opened. The energy for the  $\sigma$ (M-H) MOs, on the contrary, increases its energy upon the geometry change. The overall contribution of these four sets to the total energy for a  $d^0$  configuration (remember that in the extended Hückel method, the total energy is just the sum of each MO energy times its occupation) shows that the minimum energy geometry has an X-M-L angle around  $118^\circ$ .

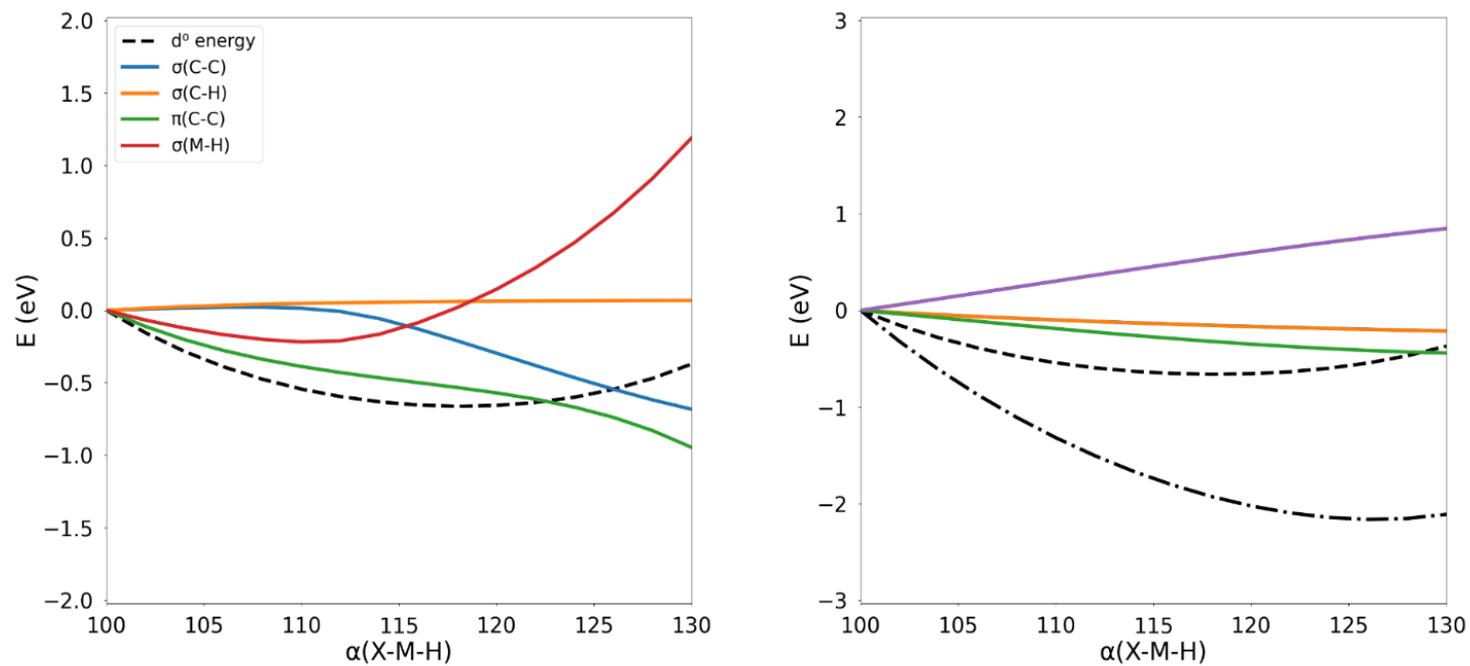


Figure 19. Walsh diagrams showing the change in the energy of the occupied molecular orbitals with respect to the  $100^\circ$  geometry as a function of the X-M-H angle (top) and for the d-type MOs (bottom). Dashed and dot-dashed lines indicate the variation of the total energy for different  $d^n$  configurations.

Color codes for the d orbitals:  $z^2$  in orange,  $(xy, x^2-y^2)$  in green, and  $(xz, yz)$  in purple.

In the same figure, the individual contribution for the five d-type orbitals is also presented. As seen in the diagram, the degenerate set of ( $xy$ ,  $x^2-y^2$ ) orbitals (in green) and the  $z^2$  orbital (in blue) are energetically stabilized at higher angles, while the degenerate ( $xz$ ,  $yz$ ) orbitals (in purple) produce the opposite effect. The net result is that by adding up to 6 electrons to the d-type metal centred orbitals to the  $\text{CpTiH}_3$  molecule will result in an opening of the X-M-L angle as shown by the change in the total energy given by the dashed line ( $d^0$ : minimum at  $\sim 118^\circ$ ) and the dash-dotted line ( $d^6$ : minimum at  $\sim 126^\circ$ ) predicting a tetrahedral to octahedral shape change when the d electron count increases. These results are in good agreement with the trend observed for the shape measures for compounds with different transition metals discussed above. Since the three lowest-lying d-type orbitals are all stabilized when the angle is opened, progressive filling of these orbitals will stabilize structures with increasing X-M-L angles, or in other words, with increasingly octahedral shapes. In the hypothetical case of  $d^n$  compounds with  $n > 6$ , our model predicts that the shape should go backwards to the tetrahedral shape again, for  $d^8$ , for instance, we predict a X-M-L angle of  $118^\circ$  as for the  $d^0$  case.

A similar analysis can be applied to other model compounds with other L ligands such as  $\text{CpTiCl}_3$ , where a similar behaviour can be found in the five d orbitals. Although the energetic stabilization of higher angle geometries found for this case is slightly smaller than for  $L = \text{H}$ , the qualitative behaviour still explains the tetrahedral to octahedral change in shape when the d orbitals are progressively filled (figure 20). Similar results are also obtained for  $\text{ArML}_3$  compounds with other aromatic ligands such as  $\text{Cbd}$  or  $\text{Bz}$  (figure 21).

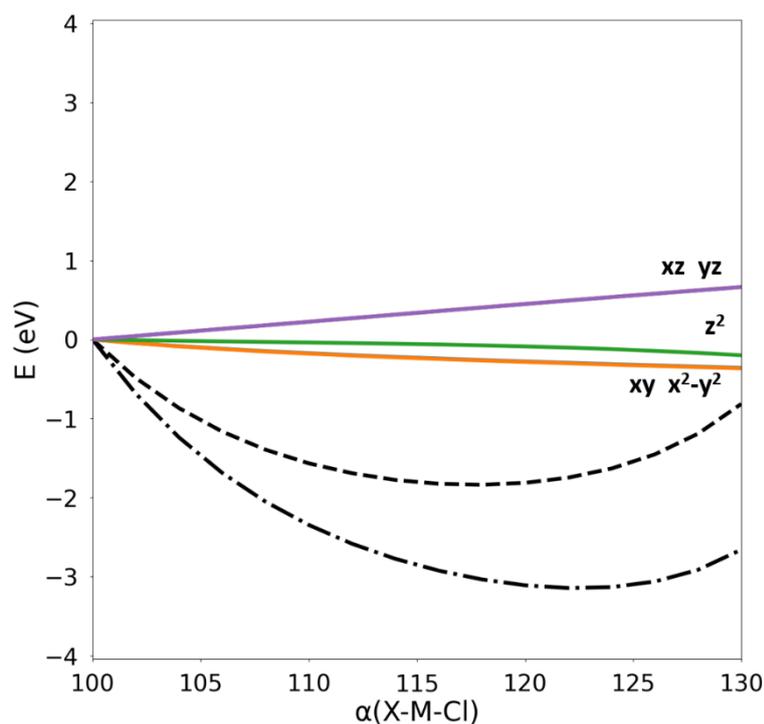


Figure 20. Walsh diagram showing the change in the energy of the d-type MOs with respect to the  $100^\circ$  geometry as a function of the X-M-H angle for the  $\text{CpTiCl}_3$  model. Dashed and dot-dashed lines indicate the variation of the total energy for different  $d^n$  configurations. Color codes for the d orbitals:  $z^2$  in orange,  $(xy, x^2-y^2)$  in green, and  $(xz, yz)$  in purple. Molecular orbital diagram showing the d orbital contributions are plotted as well. Colormap:  $z^2$  – green,  $xy$  and  $x^2-y^2$ : orange,  $xz$  and  $yz$  purple

Summarizing the results found for this simplified orbital interaction model we can say that the predicted trends agree well with those found either in the DFT calculations or for the x-ray structures recovered from the CSD. The overall conclusion is that occupying the three lower energetic d-type MOs in  $\text{ArML}_3$  piano-stool compounds should, in general, stabilize more “octahedral” geometries with higher X-M-L angles, whereas compounds with  $d^n$  configurations with  $n = 0$  to 3 electrons should present more “tetrahedral” shapes.

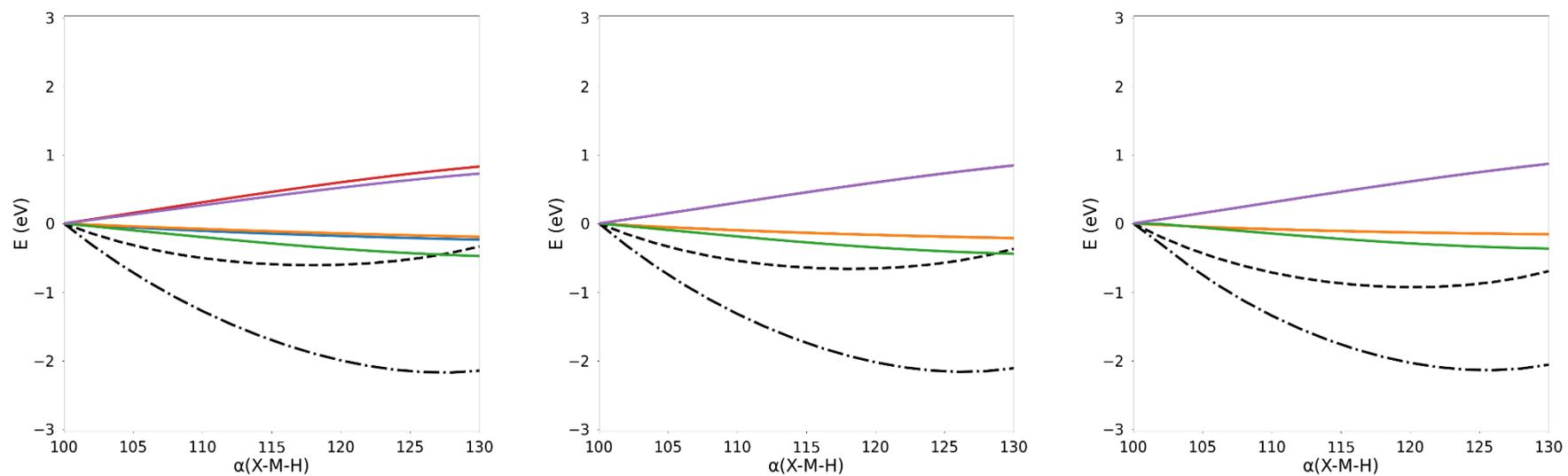


Figure 21. Walsh diagrams showing the change in the energy of the occupied molecular orbitals with respect to the  $100^\circ$  geometry as a function of the X-M-H angle for the (a) CbdtiH<sub>3</sub>, (b) CpTiH<sub>3</sub>, and (c) BzTiH<sub>3</sub> models. Dashed and dot-dashed lines indicate the variation of the total energy for different  $d^n$  configurations. Color codes for the  $d$  orbitals:  $z^2$  in orange,  $(xy, x^2-y^2)$  in green, and  $(xz, yz)$  in purple.

## Td/Oh pseudosymmetry analysis

The pseudosymmetry analysis of the MOs using the  $C_3$  and  $C_n$  rotational symmetry groups where  $n$  is the number vertices of the aromatic ring is quite useful to analyse the interaction between two fragments with coincident incompatible rotation axes. Lets remember, however, that our initial goal was to study if the geometrical change from a “tetrahedral” to an “octahedral” shape, which we have shown to be related to the filling of the d-type orbitals, is also reflected in a  $T_d \rightarrow O_h$  symmetry change for the d-type molecular orbitals.

As discussed above, we get in general a “3 over 2” splitting pattern for the energies of the d-block orbitals which hint towards a more “octahedral” symmetry for the orbitals irrespective of the geometry of the underlying nuclear framework. To delve further in this aspect, we have also performed a pseudosymmetry analysis for the five d-type orbitals for some related model compounds using either the  $T_d$  or the  $O_h$  groups as the pseudosymmetry group. As mentioned in chapter 1, an important question in a pseudosymmetry analysis is to fix a sensible orientation for the operations of the group. In this case we have chosen in all cases to set the origin for the symmetry operations at the position of the metal atom and to align one of the  $C_3$  axes of the tetrahedron or the octahedron with the  $C_3/C_n$  axis of the model compounds.

Before presenting the  $T_d/O_h$  pseudosymmetry analysis for the piano-stool complexes, let us centre on a simpler case to illustrate the results that can be obtained in such type of analysis. The ferrocene molecule is a well-known textbook example where the features of the d-block orbitals are often described using the  $O_h$  group as a reference. The idea behind this point of view is that from the electronic point of view each of the two Cp ligands in ferrocene can be seen as a triangle with three electron pairs each pointing towards the central metal atom. In order to perform our  $O_h$  pseudosymmetry analysis we orient the symmetry elements of the  $O_h$  group setting the inversion center on the metal atom and aligning one of its three  $C_3$  axes along the  $C_5$  axis of the ferrocene molecule. The  $O_h$  pseudosymmetry analysis for this orientation is shown in figure 22. The five d AOs of the metal belong to the  $T_{2g}$  and  $E_g$  irreducible representations of the

$O_h$  group. On the other hand, the five combinations of  $\pi$ -type orbitals of the Cp...Cp fragment that have the most important interactions with these d-type orbitals present a mixed character with the two leading contributions shown in the right part of figure 22. The result of the interactions of these orbitals with the metal d-block splits the five d-type AOs into two sets with an approximated 2 over 3 pattern characteristic for octahedral complexes. The pseudosymmetry analysis reveals that the main contribution to the three low lying orbitals corresponds to functions of  $T_{2g}(O_h)$  symmetry (67% for two of them and 100% for the other one), while for the upper two-orbital set we find a 64%  $E_g(O_h)$  character. The pseudosymmetry analysis shows that in this case the description of the d-block orbitals using  $O_h$  symmetry is perfectly sensible with the orbitals reproducing the  $3 \times T_{2g} + 2 \times E_g$  pattern quite satisfactorily.

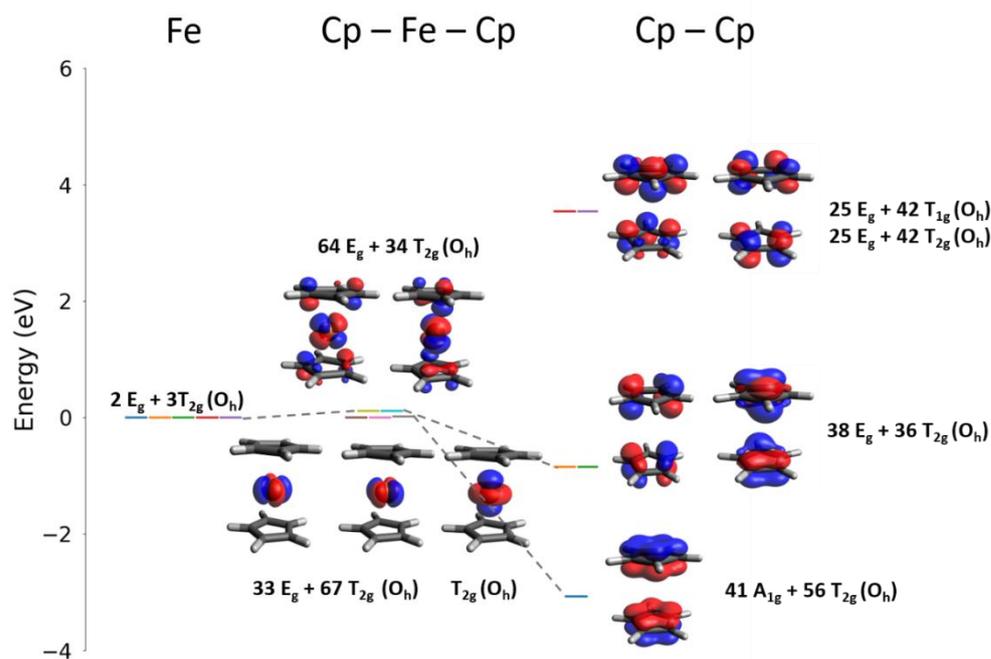


Figure 22. Ferrocene molecular orbital diagram with the pseudosymmetry analysis for the tetrahedral and octahedral (parenthesis) symmetry groups.

Let us turn now to the more complex case of piano-stool complexes where we have different ligands at each side of the metal atom. Following the same procedure as for

ferrocene, a pseudosymmetry analysis was performed for the d-block MOs of the  $\text{CpTiH}_3$  molecule, using either  $T_d$  or  $O_h$  as the pseudosymmetry groups for two alternative geometries with X-Ti-H angles corresponding to either the tetrahedral (X-Ti-H =  $110^\circ$ ) or the octahedral (X-Ti-H =  $126^\circ$ ) geometries (figure 23). In this case it is evident at a first sight that from the point of view of the MOs, the departure from either octahedral or tetrahedral symmetry is evident. First of all, as noted earlier, the splitting pattern is not close to a 2 over 3 ( $O_h$ ) or 3 over 2 ( $O_h$ ) pattern, with three sets with 2, 1, and 2 orbitals being well separated and clearly identified. If we join, however, the three lowest lying orbitals together since the central orbital is always closer in energy to the two lowest one, we find that the main character of the orbitals does not fit the  $3 T_{2g} + 2 E_g$  pattern of octahedral complexes, since when using  $O_h$  pseudosymmetry to analyse the orbitals, besides finding a considerable mixing, we find a leading 35% character for the two lower orbitals and a 87% and 2x62% character for the central and the two upper orbitals. The pseudosymmetry analysis using the  $T_d$  group reveals also an inconsistency with the three low lying orbitals having a leading  $T_2$  character and the two upper orbitals being mostly of E type. The change in geometry from a more "tetrahedral" to a more "octahedral" structure does not greatly affect the symmetry properties of the orbitals, leading to the conclusion that the observed geometrical change in shape is not reflected in the symmetry properties of the d-block orbitals and that describing them using either  $T_d$  or  $O_h$  symmetry labels is not very appropriate in this case.

We have extended the pseudosymmetry analysis to  $\text{ArML}_3$  structures with other aromatic rings, but the conclusions are similar. Figure 24 shows the case for the Bz ring with a  $C_3$  rotational symmetry compatible with either the  $T_d$  or the  $O_h$  groups, but as it is clearly seen in the figure, participation of functions belonging to different irreducible representations of the  $T_d$  or of the  $O_h$  groups is similar to what is found for the Cp case and the labelling of the orbitals using the irreducible representations of these groups does not seem appropriate. The same conclusions are reached for the case of cyclobutadiene, with the graphic showing the results of the pseudosymmetry analysis shown in the Appendix.

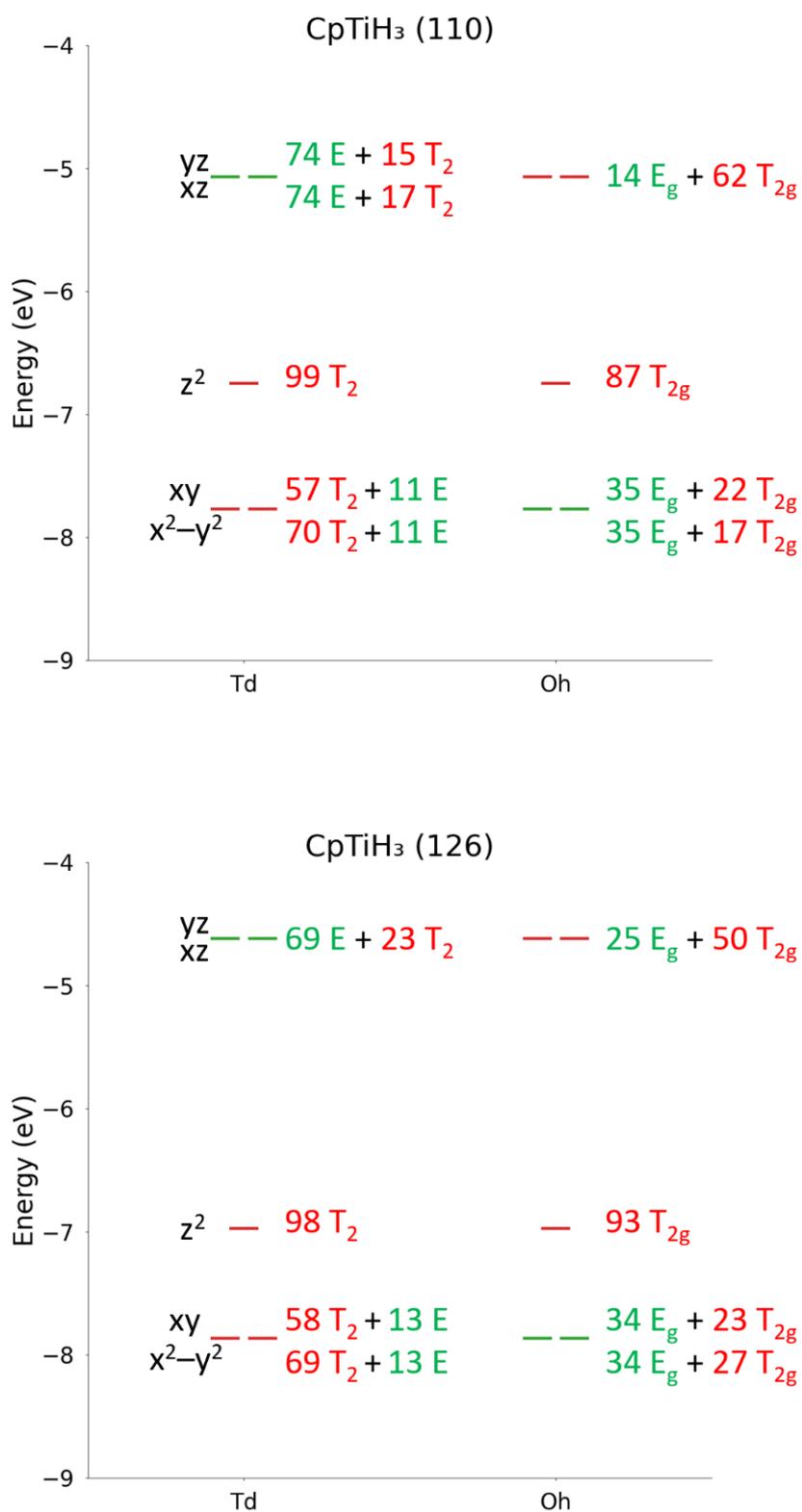


Figure 23. % of T<sub>d</sub> - E/T<sub>2</sub> (left) and O<sub>h</sub> - E<sub>g</sub>/T<sub>2g</sub> (right) character of the d-type MOs for CpTiH<sub>3</sub> with  $\alpha = 110^\circ$  (top) and  $126^\circ$  (bottom) geometries.

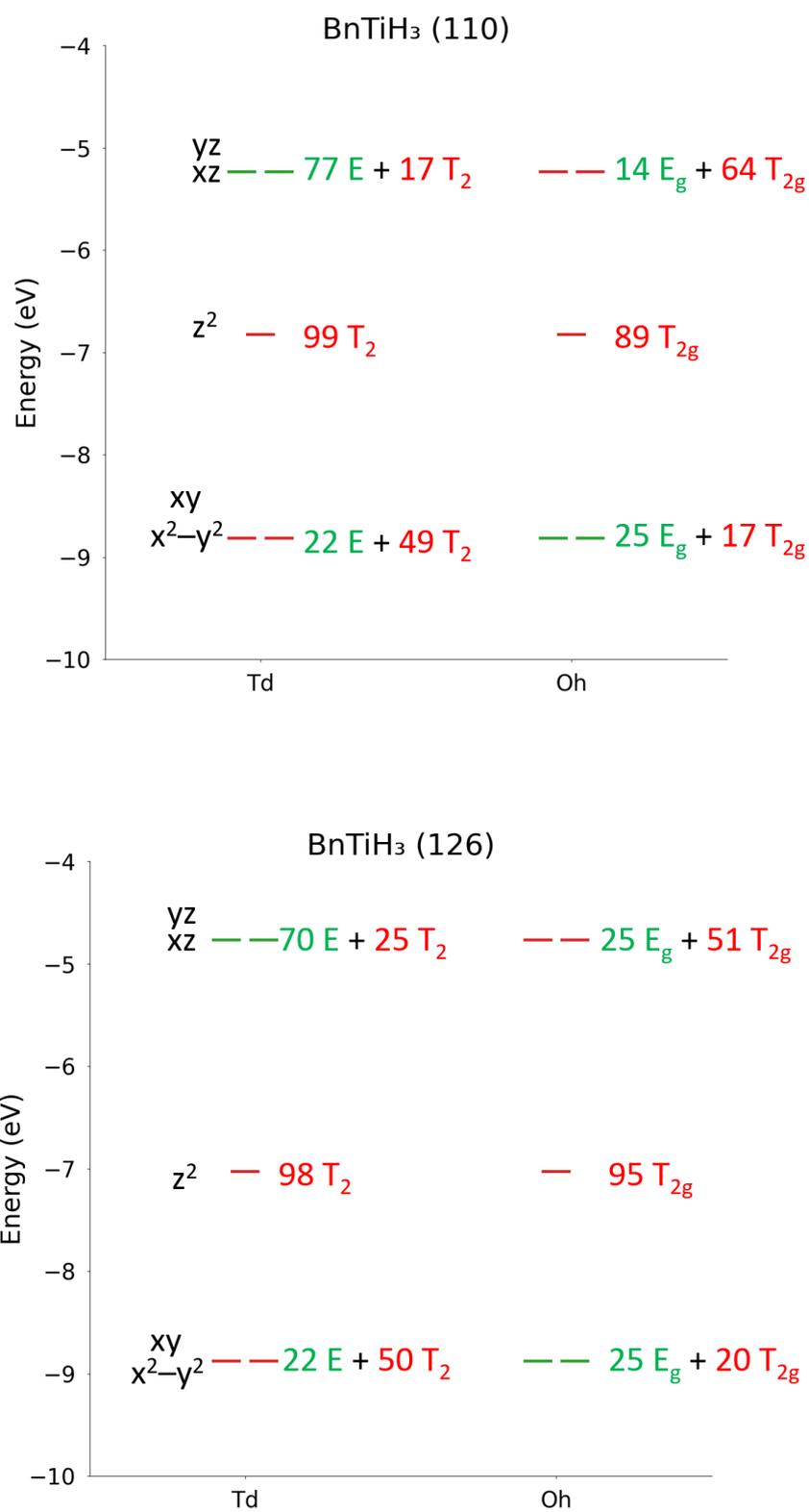


Figure 24. % of T<sub>d</sub> - E/T<sub>2</sub> (left) and O<sub>h</sub> - E<sub>g</sub>/T<sub>2g</sub> (right) character of the d-type MOs for BzTiH<sub>3</sub> with  $\alpha = 110^\circ$  (top) and  $126^\circ$  (bottom) geometries

## Conclusions

In this chapter, the analysis of piano-stool compounds leads to an interesting relation between the shape of these molecules and the number of electrons found in the molecular d-block orbitals. Filling the three lowest d orbitals will vary the shape of piano-stools from a more tetrahedral to a more octahedral shape. However, the addition of more electrons to the higher energy d-type orbitals is predicted to revert this behaviour. It has been proven, that this effect is mainly caused by the  $z^2$  orbital, as the hybridization needed to form an interaction with the  $L_3$  fragment is only significant for some angles. For the  $(xy, x^2-y^2)$  this energy stabilization only exists in some cases and is much smaller ( $\Delta E \approx 0.1$  eV). Finally, the  $xz$  and  $yz$  orbitals present an anti-bonding character with both the aromatic ring and the  $L_3$  fragment orbitals. In the latter case, as the angle increase their anti-bonding orbitals also increase in energy, giving an overall energy destabilization of the molecular orbitals of the whole molecule.

On the other hand, the pseudosymmetry analysis using either  $T_d$  or  $O_h$  pseudosymmetry reveals that the observed structural changes, although linked to the electron filling of the d-block orbitals, does not have any influence on the symmetry characteristics of these orbitals which cannot be assimilated neither to octahedral nor tetrahedral symmetry orbitals irrespective of the molecular geometry. This finding is in sharp contrast to what is found for ferrocene, where a structure with a symmetry that is totally different from the  $O_h$  symmetry presents a set of d-block orbitals which, are shown to possess a high degree of octahedral symmetry.

## Appendix

Energy (eV)		
	s	p
<i>H</i>	-12.565	-
<i>C</i>	-19.654	-11.129
<i>N</i>	-25.366	-13.900
<i>O</i>	-31.600	-16.776
<i>F</i>	-38.373	-19.769
<i>Cl</i>	-26.636	-14.217

Energy (eV)			
	s	p	d
<i>Sc</i>	-6.123	-3.664	-6.067
<i>Ti</i>	-6.520	-3.808	-7.177
<i>V</i>	-6.850	-3.910	-8.181
<i>Cr</i>	-7.141	-3.989	-9.109
<i>Mn</i>	-7.409	-4.053	-9.986
<i>Fe</i>	-7.659	-4.104	-10.822
<i>Co</i>	-7.897	-4.148	-11.625
<i>Ni</i>	-8.125	-4.184	-12.404
<i>Cu</i>	-8.345	-4.216	-13.162

Table 1A. Energy parameters for the different atomic orbitals used in the Extended-Hückel calculations done with *Huckelpy*. Values obtained from Vela et al.<sup>15</sup>

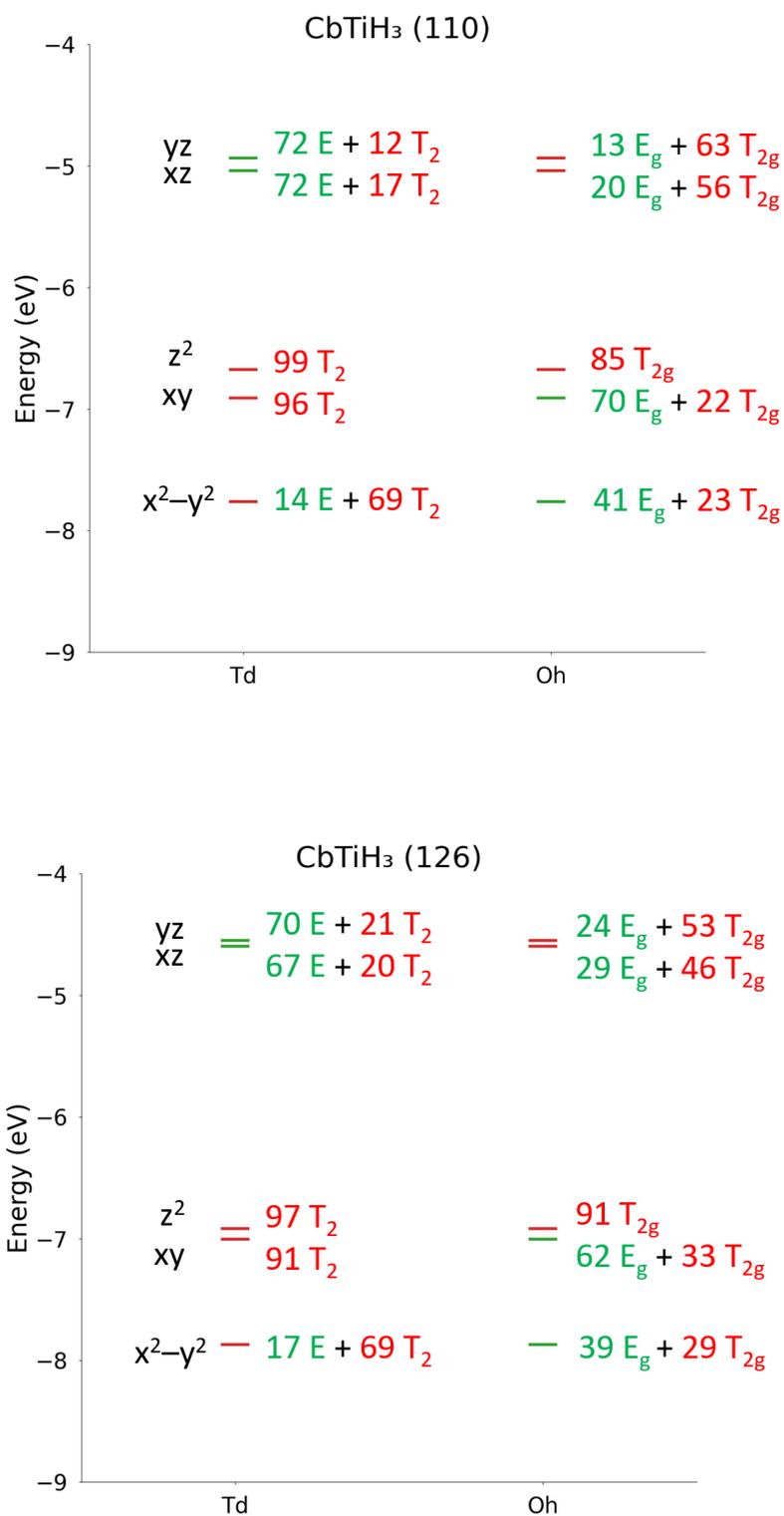


Figure 1A. Pseudosymmetry analysis of the molecular d orbitals by Td and Oh symmetry for cyclobutadiene half-sandwich compounds. The molecule at 110° and at 126° represent the tetrahedral and octahedral shape, respectively.

## Bibliography

1. J. A. Le Bel. Sur les relations qui existent entre les formules atomiques des corps organiques et le pouvoir rotatoire de leurs dissolutions. *Bull. Soc. Chim. Fr.* **22**, 337–347 (1874).
2. Hoff, J. H. van't. Voorsteel tot uitbreiding der tegenwoordig in de scheikunde gebruikte structuur-formules in de ruimte. *Arch. Neerl. des Sci. Exactes Nat.* **9**, 445–454 (1874).
3. Bel, L. & Hoff, V. Le Bel-van't Hoff rule. <https://medical-dictionary.thefreedictionary.com/Le+Bel-van%27t+Hoff+rule>.
4. Lawrance, G. A. *Introduction to Coordination Chemistry. Introduction to Coordination Chemistry* (2009). doi:10.1002/9780470687123.
5. Nič, M., Jirát, J., Košata, B., Jenkins, A. & McNaught, A. *IUPAC Compendium of Chemical Terminology. Blackwell Scientific Publications* (2009). doi:10.1016/j.egypro.2017.03.1781.
6. Werner, H. At least 60 years of ferrocene: The discovery and rediscovery of the sandwich complexes. *Angew. Chemie - Int. Ed.* **51**, 6052–6058 (2012).
7. Cirera, J., Ruiz, E. & Alvarez, S. Continuous shape measures as a stereochemical tool in organometallic chemistry. *Organometallics* **24**, 1556–1562 (2005).
8. Caliskan, B. Radical Mechanisms in the Metallocenes. *Recent Prog. Organomet. Chem.* (2017) doi:10.5772/intechopen.68952.
9. Echeverría, J., Carreras, A., Casanova, D., Alemany, P. & Alvarez, S. Concurrent symmetries: The interplay between local and global molecular symmetries. *Chem. - A Eur. J.* **17**, 359–367 (2011).
10. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
11. Staroverov, V. N., Scuseria, G. E., Tao, J. & Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **119**, 12129–12137 (2003).
12. Tao, J., Perdew, J. P., Staroverov, V. N. & Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **91**, 3–6 (2003).
13. Hoffmann, R. An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* **39**, 1397–1412 (1963).
14. Bernuz, E. huckelpy. <https://github.com/efrembernuz/ExtendedHuckel>.
15. Vela, A. & Gázquez, J. L. Extended Hückel parameters from density functional theory. *J. Phys. Chem.* **92**, 5688–5693 (1988).

16. Tsuchida, R. Absorption Spectra of Co-ordination Compounds. II. *Bull. Chem. Soc. Jpn.* **13**, 436–450 (1938).
17. Zumdahl, S. S. *Chemical Principles. The Physical Basis of Biochemistry* (Brooks/Cole, 2009).
18. Miller, C. S. & Ellison, M. Walsh Diagrams: Molecular Orbital and Structure Computational Chemistry Exercise for Physical Chemistry. *J. Chem. Educ.* **92**, 1040–1043 (2015).



## Chapter 4. Effect of temperature on the shape and symmetry of molecules and solids

### Introduction

The shape and symmetry of molecules, crystals, and molecular aggregates are crucial concepts in stereochemistry defined usually in terms of classical Euclidian geometry and group theory. The simplicity of describing molecular structures using perfect polyhedral models, such as the *tetrahedral carbon atom*, induce the wrong idea that the vast majority of molecular structures are highly symmetric, while the real situation is, however, the contrary. Any molecular structure is actually dynamic, with its nuclei and electrons in a continuous motion, actively changing their positions. Consequently, the instantaneous shape of a molecule is in reality usually distorted, with a shape that does not fully coincide with any ideal highly symmetric model. The two main sources for this shape/symmetry loss are: the thermal effect in the motion of atoms and the asymmetry in the surrounding non-bonded interactions produced by the packing in molecular crystals or aggregates. Note that these two sources of asymmetry are quite different in nature. While thermal motion leads to a succession of distorted geometries that may lead to an average structure with the shape/symmetry of an ideal model, the effect of an asymmetric environment leads to a symmetry/shape loss even for the average structure.

In this chapter, the focus will be on the effects of temperature on the shape, symmetry, or chirality of molecules or in the local coordination environment of individual atoms in solids. Although at a given temperature the average molecular geometry, as determined for instance in X-ray diffraction experiments, may have a certain regular shape / symmetry, instantaneous configurations deviate from this average structure due to the atomic motion induced by temperature (strictly speaking, due to the quantum nature of matter, these deviations do not completely disappear even in the limit of the absolute zero temperature) and it is the mean deviation from the average

shape/symmetry that we are looking at in this chapter. The effect of temperature will be analysed for a few illustrative cases using the continuous shape and symmetry measures.

We will start our discussion with the relatively simple case of the robust  $P_4$  tetrahedral molecule to illustrate the main questions related to shape/symmetry loss due to thermal motion. After this initial example, we will deal with the ethane molecule as an example where the internal rotation is able to induce large changes in the molecular shape for relatively low temperatures. We will study also the effects of temperature on highly symmetric structures such as  $B_6H_6^{2-}$  or  $C_8H_8$  and other cage molecules with the shape of the platonic polyhedra. To end, we will look at the effects of temperature on the local coordination environment of a single atom in the diamond-type tetrahedral crystal structures of the elemental solids carbon, silicon, and germanium.

## Methodology

In the following sections, the effect of temperature on the different molecular structures will be studied using the Monte Carlo method<sup>1</sup>, a statistical approach to solve expensive computational simulations that rely on the generation of a series of repeated random numbers to solve problems that might be deterministic. In problems related to molecular structure, a Monte Carlo simulation is based on the statistical analysis of a series of random molecular or crystal geometries that reproduce a mechanical system typically at a constant volume and temperature in a NVT ensemble. Supposing that if the sample is large enough, average values should reproduce the outcome of an experimental structure determination. The resulting sample represents a probability distribution for different states derived from the acceptance or rejection of calculations that satisfy a certain condition. In our case, the Monte Carlo simulations are guided by the Metropolis algorithm<sup>2</sup>, a Markov Chain Monte Carlo method. A Markov chain is a sequence of possible events where the probability of future events depends only on the present state<sup>3</sup>. The probability distribution that will be used to describe the following

systems will be derived from Maxwell-Boltzmann statistics, which describes systems with non-interacting particles in thermodynamical equilibrium.

In the present work, the simulations were carried out with the *Montemodes* program<sup>4</sup>, a classical Monte Carlo simulation program. Starting from an initial geometry, the program generates new possible structures displacing the atoms in the direction of the normal modes of the initial structure. The energy of the new structure is then calculated using an external program such as *Gaussian*<sup>5</sup> or *Tinker*<sup>6</sup>. This energy is used to calculate an acceptance/rejection probability of the proposed structure in the sequence of structures for which statistical averages are evaluated at the end of the simulation. The use of normal modes instead of small random atomic movements in cartesian coordinates can be quite useful to explore the possibility of structures with larger distortions with less steps, especially in cases where low frequency modes such as those associated with internal rotations may lead to large structural changes with a small change in energy.

The algorithm implemented in the *Montemodes* program works as follows:

Initialise

Set an initial state  $\mathbf{x}_0$  (in this case a previously optimized geometry for a given molecule) and calculate its normal modes.

Iterate

- i. Generate a new state  $\mathbf{x}'$  by displacing the atoms randomly following the normal modes of the structure
- ii. If energy of  $\mathbf{x}'$  ( $E'$ ) < energy of  $\mathbf{x}$  ( $E$ ), the new state is accepted in the sequence.
- iii. Else
  - a. Evaluate the acceptance/rejection probability (Boltzmann factor)

$$A(\mathbf{x}', \mathbf{x}) = \exp\left(\frac{E' - E}{k_B T}\right)$$

- b. Generate a random number  $\mathbf{u} \in [0,1]$ 
  - If  $A(\mathbf{x}', \mathbf{x}) < \mathbf{u}$  accept the new state
  - If  $A(\mathbf{x}', \mathbf{x}) > \mathbf{u}$  reject the new state
- iv. Repeat the process until the desired number of simulation steps is reached

For the crystals with diamond-type structures we have used an alternative approach to obtain a set of representative geometries. In this case we have used the *DL\_poly* program<sup>7</sup> to generate trajectories using classical molecular dynamics simulations in the NPT ensemble in combination of forcefield methods. The simulation generates a trajectory with 1,200,000 structures where the first 200,000 were used as equilibration steps with a timestep of 0.001. Every 1000 time steps, a snapshot with the current structure and properties is saved. Each trajectory is generated using the Verlet scheme<sup>8</sup>, a numerical method to integrate the Newton's equation of motion. Usually, molecular dynamics simulations are run in the NVE ensemble, however, in this case the NPT ensemble was used by using a Nosé–Hoover thermostat<sup>9–11</sup>, which rescales the velocities to keep the temperature constant along the trajectory.

## P<sub>4</sub> molecule

In its minimum energy geometry, the P<sub>4</sub> molecule presents a perfect tetrahedral symmetry, with a structure where each phosphorus atom occupies a vertex of a tetrahedron. Optimizing the structure for this molecule with the *Gaussian* program using the semiempirical PM6<sup>12</sup> method yields a perfect tetrahedral structure with P-P distances of 2.26 Å. Since optimization using DFT methods gives similar results (2.22 Å at the B<sub>3</sub>LYP/6-31G(d,p) level), we will use the cheap PM6 method in combination with the *Montemodes* program to generate sufficiently long representative sets of molecular geometries for each studied temperature.

The tetrahedral geometry corresponds to a minimum on the potential energy surface with six normal modes  $A_1 + 3 T_2 + 2 E$  with  $430 \text{ cm}^{-1}$ ,  $626 \text{ cm}^{-1}$ , and  $938 \text{ cm}^{-1}$  at the PM6 level, respectively. Considering an isolated molecule, excitation of these six normal modes will describe the motion of the atoms in the molecule as a function of the temperature where the four atoms suffer large enough displacements to break the ideal tetrahedral structure, except upon excitation along the totally symmetric  $A_1$  mode that will result only in a shrinkage/expansion of the perfect tetrahedron.

A rough estimation of the changes in the P-P distances induced by thermal motion can be made by considering for each atom the vibration of a mass  $m$  held in place by springs. The average energy associated with the displacement of the atom from its equilibrium position  $u$  is  $m\omega^2\langle u^2 \rangle$  where  $\omega$  is the angular frequency of the oscillatory motion and  $\langle u^2 \rangle$  the average square displacement. Equating this average displacement energy with the thermal energy  $k_B T$  gives:

$$\langle u^2 \rangle^{1/2} = \sqrt{\frac{k_B T}{m\omega^2}} \quad (1)$$

At a temperature  $T=300\text{K}$ , if we consider the movement of the P atoms along the  $T_2$  normal mode of  $626 \text{ cm}^{-1}$ , we obtain a value of  $0.024 \text{ \AA}$  for  $\langle u^2 \rangle^{1/2}$ , a small displacement considering that it is only about a 1% of the total P-P distance.

Taking into account the dependence of the displacement with the atom's mass and the vibrational frequency in equation (1), it is easy to see that larger average displacements are expected for lighter atoms and that low energy vibrational modes will produce larger displacements as well. In other words, for the same structure twisting modes will cause larger symmetry losses than the bending modes and high frequency stretching modes will give the smallest distortions of the equilibrium structure.

As mentioned before, shape and symmetry are invariant to scaling transformations, so the absolute value of the displacement is meaningless for the purposes of quantifying the shape or symmetry loss of a given structure, and the relative size of the

displacement should be taken into consideration. For example, for two compounds with tetrahedral structures having different atom-atom distances, the same value for the average atomic displacement will affect much less the shape/symmetry of the compound with the larger structure than that with the smaller one.

Another important factor that should be pointed out is that the average shape/symmetry loss due to thermal motion is not the same as the changes in shape/symmetry of the average structure. This leads to an interesting contradiction that is usually found in crystallography: raising the temperature, and hence, enlarging the displacement of individual atoms from their equilibrium positions, leads to a more symmetric structure. An example of this can be found in the alkali-metal cyanides or alkali-earth carbides such as NaCN<sup>13</sup> and CaC<sub>2</sub><sup>14</sup>. The diatomic CN<sup>-</sup> and C<sub>2</sub><sup>2-</sup> molecular ions rotate when raising the temperature leading to an average structure where each anion occupies a single position in the crystal, showing a more symmetric average structure, even though the instantaneous symmetry for the actual atomic positions at a given time is lower.

As will be shown, the simulations for a single P<sub>4</sub> molecule also reproduce this behavior, where instantaneous configurations with an average tetrahedral shape measure of 0.06 yield an average structure that is perfectly tetrahedral, with a tetrahedral shape measure of 0.

To analyze the loss of the perfect tetrahedral symmetry of the P<sub>4</sub> molecule due to the displacement of each of the P atoms from their equilibrium positions, a set of 30,500 P<sub>4</sub> structures were generated for each temperature using a Monte-Carlo Metropolis simulation where the energy for each structure was calculated using Gaussian 09<sup>5</sup> with the semiempirical PM6 method<sup>12</sup>. Since the goal of these simulations is to show the qualitative effects that affect the shape and symmetry of the molecules at different temperatures, the use of the semiempirical PM6 method to calculate the energies of the generated structures gives results in good agreement with the qualitative picture at a low computational cost, when compared with test results obtained using a more sophisticated method such the DFT based B<sub>3</sub>LYP method<sup>15,16</sup>. The starting configuration corresponds to the equilibrium geometry of P<sub>4</sub> and can vary during the simulation displacing the atoms along the normal modes of the molecule within an

acceptation to rejection rate of 50%. The first 500 configurations in each run were discarded in the shape analysis, considering them as forming part of the equilibration period. The other 30.000 structures were then analyzed in terms of their distribution of energy and shape measure values as shown in Figure 1. Both histograms fit, as it has been noted by Tuví-Arad et al.<sup>17</sup>, to a log-normal distribution,

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (2)$$

where  $\mu$  and  $\sigma$  are two parameters related to the position of the peak and the width of the distribution, respectively.

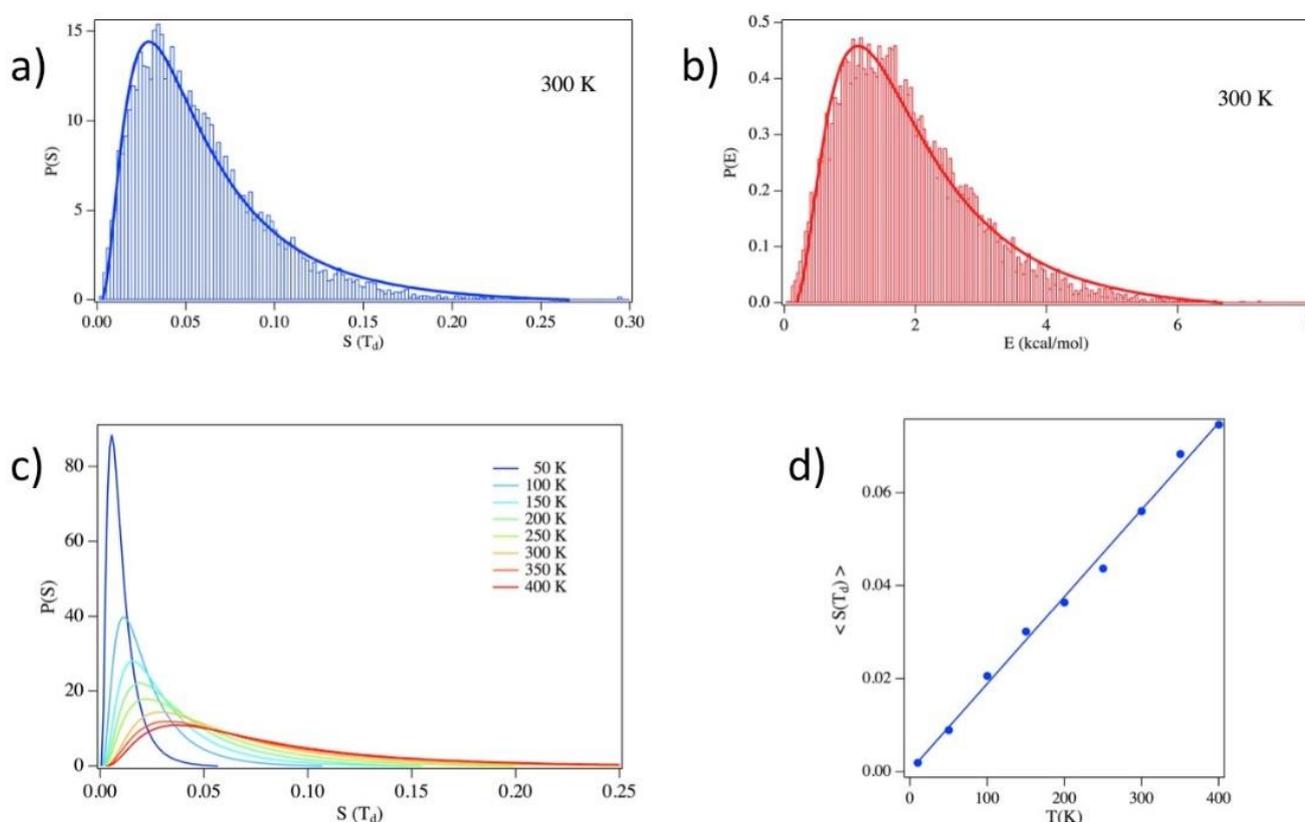


Figure 1. a) Distribution of the tetrahedral CSM for the set of 30.000  $P_4$  structures obtained in a simulation at  $T = 300$  K. b) Distribution of the energy for the same set. c) Evolution for the distribution of the tetrahedral CSM with the temperature. d) Average tetrahedral CSM as a function of the temperature.

As seen in figure 1c, the spread of the CSM values increases with temperature. A similar behavior can be found in the energy profile.

As expected for a log-normal distribution, the average value is given by

$$\langle x \rangle = \exp \left( \mu + \frac{\sigma^2}{2} \right) \quad (3)$$

Both for the CSM and energy distributions we find a linear dependence of the average values with the temperature, as shown in figure 1d for the CSM. From these two trends, one can also expect a linear dependence between the average energy and the average CSM value.

Figure 2a, shows how CSM values are related to the energy. In this case where the minimum energy structure is perfectly tetrahedral, low energy structures have higher tetrahedral symmetry ( $S(Td) \approx 0$ ) while higher energy structures imply some sort of distortion from the perfect tetrahedral geometry, even though there is a low probability of finding highly symmetric structures in the high energy region. This occurrence can be explained when the molecule is distorted along the totally symmetric  $A_1$  mode for which an increase in energy does not imply any symmetry loss. This is shown graphically in figures 2b and 2c, where the change in energy and symmetry is plotted as a function of the displacement along a given mode. In these figures, it is clear that the displacement along the  $A_1$  mode, unlike the E and  $T_2$  ones, implies an increase of the energy while the molecular symmetry is unaffected. Note also that the symmetry loss for any other mode, E and  $T_2$  in this case, leads to the same value for a given amplitude, a relation that can be shown to hold for any structure formed by a set of atoms all with the same mass.

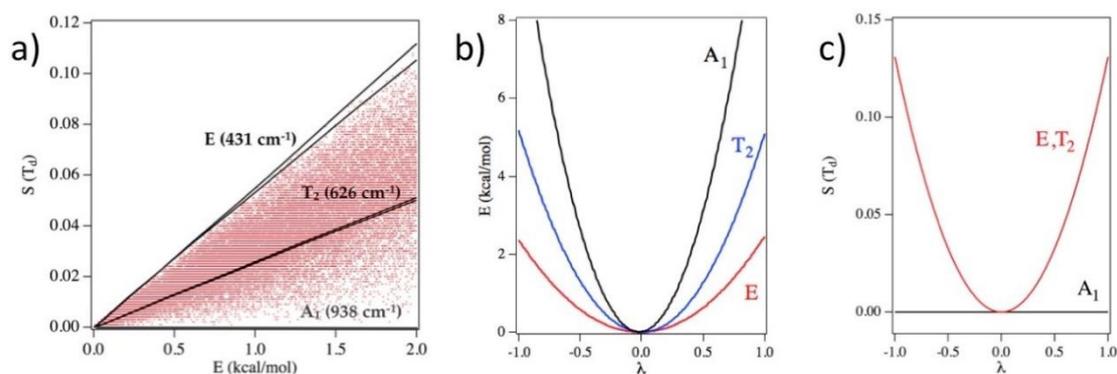


Figure 2. Tetrahedral CSM versus energy (red dots) for the set of 30.000 P<sub>4</sub> structures obtained in a simulation at  $T = 400\text{K}$ . The black lines show the variation of the CSM with the energy along the normal modes of vibration. b) Variation of the energy along the different normal modes for a single P<sub>4</sub> molecule. c) Variation of  $S(T_d)$  along the different normal modes for a single P<sub>4</sub> molecule.

An interesting feature is observed in the upper limit of figure 2a, where almost no structures are found in this region, meaning that a given increase in energy for a certain structure cannot yield an arbitrary symmetry loss beyond a certain value, with the limit in the  $S$  vs  $E$  distribution given by the softest normal mode.

The harmonic approximation assumes that the energy variation along any mode should be parabolic, which in fact, it is not the case and the energetic cost is slightly different for the positive or negative direction of the normal mode. However, the change of the symmetry loss is perfectly parabolic, and as a result of this, the linear relation between symmetry and energy differs slightly when moving forward or backward a given mode. The E mode shows this effect as two lines can be clearly observed at higher displacements, where the anharmonicity cannot be ignored. For smaller displacements the energy cost is symmetric and both lines coincide in the lower displacement region where anharmonicity can be neglected. In the case of the T<sub>2</sub> and A<sub>1</sub> modes, the former has a smaller anharmonicity effect and the two lines coincide perfectly, while in the later, since there is no symmetry loss in any direction, the anharmonicity does not play any role.

As seen until now, the shape variation of a robust tetrahedron like that of the  $P_4$  molecule is almost insignificant as the minimum on the potential energy surface is distant from other possible shapes and the possibility of populating minima corresponding to other conformers at low temperatures is negligible. A different case can be found in a more flexible molecule such as ethane, for which two alternative geometries with distinct shapes may be accessible when the temperature is increased.

## Ethane

Ethane is a simple case to evaluate the effect on shape of a temperature increase since it allows to observe important shape variations with moderate increases in the temperature. Exploring this case allows us to analyze, from a symmetry perspective, the effects of the conformational variation that occurs when exciting low energy modes that correspond to the internal rotation around C-C bonds in organic molecules. In the case of ethane, this rotation results in the interchange between the minimal energy staggered ( $D_{3d}$ ) conformation and the eclipsed ( $D_{3h}$ ) conformation, corresponding to a low energy barrier (~1 Kcal/mol at the PM6 level) that separates equivalent staggered configurations (Figure 3a). As seen in figure 3b, if we consider the position of the H atoms only, the most stable geometry, corresponding to the staggered conformation, has the shape of a distorted octahedron, elongated along one of the  $C_3$  axes that coincides with the C-C bond. The rotation of one of the triangular  $H_3$  fragments around this symmetry axis can be described as a Bailar-twist type distortion<sup>18</sup>, leading to the triangular prismatic arrangement of the H atoms corresponding to the eclipsed structure.

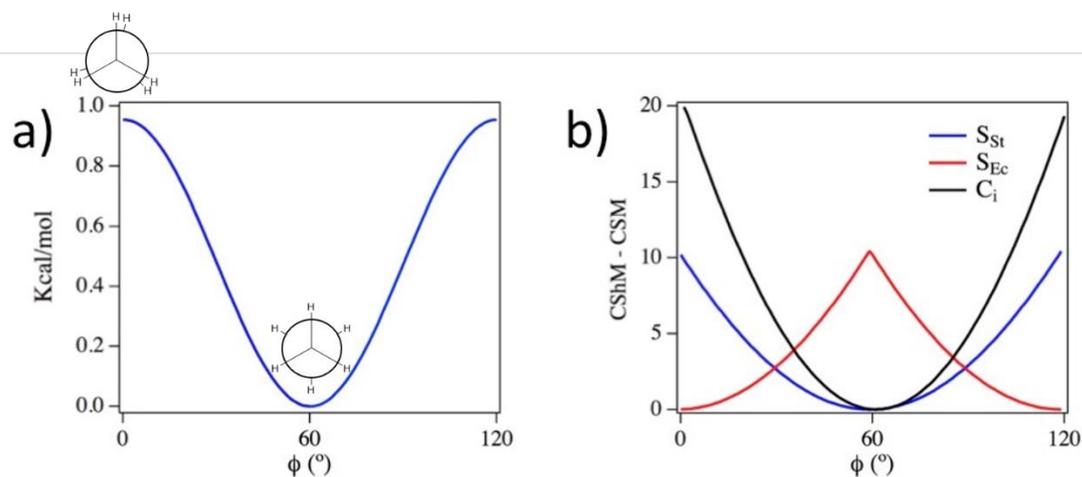


Figure 3. a) Potential energy curve for the internal rotation in ethane. b)  $S_{St}$ ,  $S_{Ec}$  and  $S_{C_i}$  as a function of the dihedral angle.

Contrarily to the previous case of the  $P_4$  molecule where we were able to use only one reference structure (the tetrahedron) to describe the shape loss due to thermal motion, for the ethane molecule we will need to consider two possible reference structures: the distorted octahedral arrangement corresponding to the staggered form and the trigonal prismatic arrangement corresponding to the eclipsed form. Although we could use the octahedron and the trigonal prism as reference shapes to describe the arrangement of the six H atoms in ethane, a more accurate description of the shape loss can be found by using directly the optimized structures for the staggered and eclipsed conformations as reference shapes. These will be referenced as St and Ec from now on, and the corresponding shape measures will be labelled correspondingly as  $S_{St}$  and  $S_{Ec}$ , respectively. An alternative way, as seen in figure 3, is to analyze the loss of a symmetry that is only present in one of the two conformations, which is the case of the inversion group  $C_i$ , a subgroup of the  $D_{3d}$  symmetry, that is only present in the staggered conformation.

As an approximation, the focus will be on distortions arising from the lowest energy mode, that corresponds to an internal rotation along the C-C bond, so that we will consider only geometries lying on the relaxed potential energy curve for this mode. The effect of other distortions due to higher energy modes is expected to be small at low

temperatures and will not be considered further in this work. The staggered minimal energy conformation will be considered as the main reference shape and the only possible distortions will oscillate between this conformation and the eclipsed one, meaning that the corresponding average CShM and CSM values will be  $\langle S_{St} \rangle \approx 0$ ,  $\langle S_{Ec} \rangle \approx 10$ , and  $\langle S_{Ci} \rangle \approx 0$  at very low temperatures where the thermal energy is not sufficient to overcome the energy barrier corresponding to the eclipsed conformations. In this case, it has been necessary to increase the number of structures generated at each temperature to 60,000 in order to obtain a representative sampling of structures with different dihedral angles.

The simplest way to analyze the shape variation between two reference structures is to use a shape map (chapter 1), where each axis corresponds to the shape measure of a structure with respect to one of the reference shapes. In this case, the values for the two measures can oscillate between 0 and  $\sim 12$  (figure 4). In this figure we can observe that structures at low temperatures (blue point cloud) have  $S_{St}$  and  $S_{Ec}$  CShMs distributed along the minimum distortion path but centered mainly on the staggered region ( $S_{Ec} \sim 0$ ). Increasing the temperature produces larger point clouds that reach the region around the Eclipsed form for temperatures between 100 and 200K. Further increase of the temperature results only in a broadening of the point cloud.

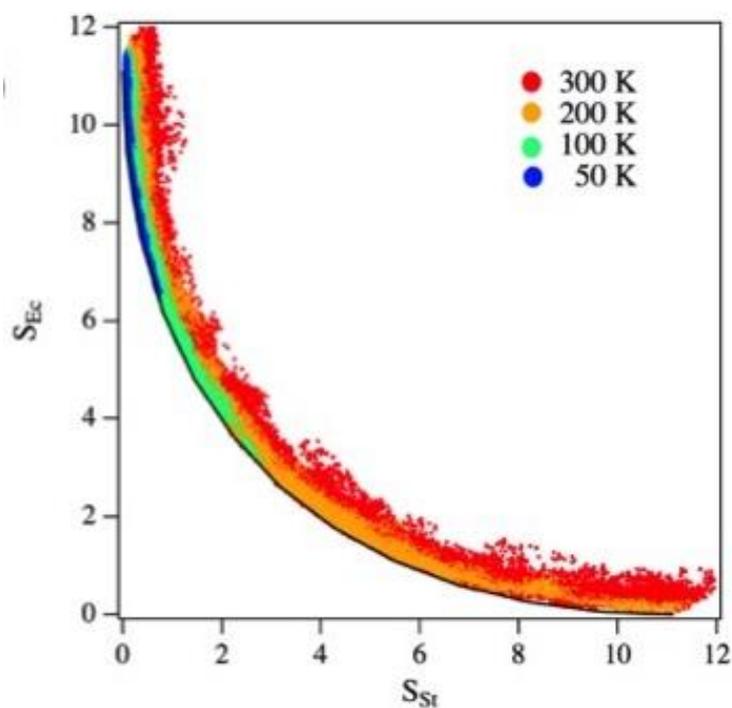


Figure 4. Shape map for the staggered ( $S_{St}$ ) and eclipsed ( $S_{Ec}$ ) CShMs for the ethane structures obtained from the Monte Carlo simulations run at different temperatures.

To clearly show how the mean values of the CShMs for the staggered and eclipsed conformations behave at different temperatures, it is convenient to fit the distributions to a log-normal function (figure 5). At very low temperatures, the ( $S_{St}, S_{Ec}$ ) values lie in the region of the initial staggered conformation geometry  $\langle S_{St} \rangle \approx 0$  and  $\langle S_{Ec} \rangle \approx 10$ , and more distorted geometries with dihedral angles in between the staggered and the eclipsed ones are found progressively as the temperature raises. Similarly, to what we found for the  $P_4$  case, a linear relation of the average shape variation with the temperature is expected, and that is the case for low temperatures up to 50 K. However, the linear behavior is lost at higher temperatures where both  $\langle S_{St} \rangle$  and  $\langle S_{Ec} \rangle$  have a more pronounced increase with temperature (figure 5). It is not until we reach very high temperatures that the average values,  $\langle S_{St} \rangle = 3.7$  and  $\langle S_{Ec} \rangle = 3.4$ , that correspond to a uniform distribution along all the possible dihedral angles, are reached. This situation is never reached at moderate temperatures, where the breaking of the molecule can be

discarded, as probably more than one source of distortion should be considered and not only the rotational mode.

Considering that for the ethane molecule the lower energy normal modes at the PM6 level correspond to the internal rotation ( $\nu = 180 \text{ cm}^{-1}$ ) and a doubly degenerated mode ( $\nu = 883 \text{ cm}^{-1}$ ). The effects of the second mode should become evident only at a vibrational temperature,  $\Theta_{\text{vib}} = \nu/k_{\text{B}}$ , of 1270 K. In the case of the first normal mode, the vibrational temperature is 260 K, lying in the range included in the figure and thus indicating that this mode is the main contribution to the geometrical distortion in the temperature range where ethane does not dissociate.

After analyzing two basic cases, the robust  $\text{P}_4$  tetrahedral molecule and the flexible  $\text{C}_2\text{H}_8$  molecule, it seems pertinent to analyze more illustrative cases that will show that sometimes counterintuitive results that at first seem to contradict some general notions on molecular structures may appear. Another interesting case, an  $\text{AB}_n$  molecule formed by a central atom A surrounded by n equivalent B atoms with different masses such as in  $\text{CH}_4$  will be studied in more detail in a following chapter, where the effect of the temperature on the shape of the tetrahedral phosphate  $\text{PO}_4^{3-}$  ion is analyzed in detail.

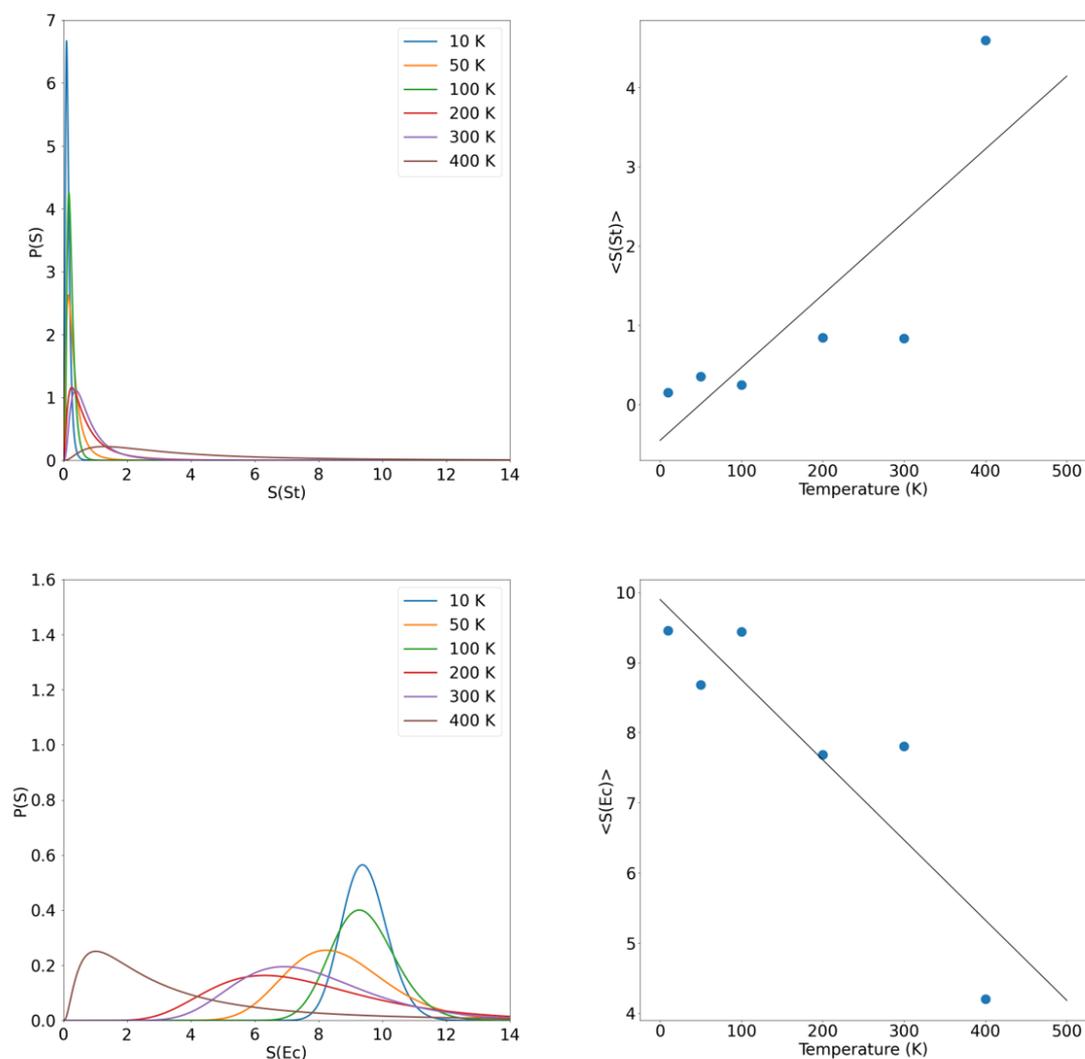


Figure 5. Distribution of CShMs for ethane alternate (top-left) and staggered (bottom-left) reference polyhedrons. At right the mean values for the log-normal distributions at each temperature are shown as a function of the temperature. For each temperature a sequence of 60,000 structures have been generated using a Monte Carlo algorithm in which the Cartesian coordinates for all 8 atoms in the molecule can change, with the energy being evaluated at the PM6 level.

## Polyhedral molecules

To extend the previous analysis, we have explored also the effects of temperature on a set of cage structures that are highly symmetric,  $C_4H_4$ ,  $B_6H_6^{2-}$ ,  $C_8H_8$ ,  $B_{12}H_{12}^{2-}$ , and  $C_{20}H_{20}$

(figure 6), that adopt the shape of the five platonic solids, the tetrahedron, octahedron, cube, icosahedron, and dodecahedron, respectively. The comparison of the effect of the thermal motion on the shape of these structures should lead to a better understanding of the role of the relative rigidity or size of the molecule, or even a competition between these two factors. Another important question that will be discussed is the presence of an outer shell of hydrogen atoms that may affect the distortion of the inner cage. For this purpose, we will compare the effects of temperature of the carbon tetrahedron in tetrahedrane with those found above for the  $P_4$  molecule. Finally, we will address the question if the different number of symmetry operations in a group is a relevant factor in the degree of symmetry loss under the influence of a random asymmetry source such as that provided by thermal motion.

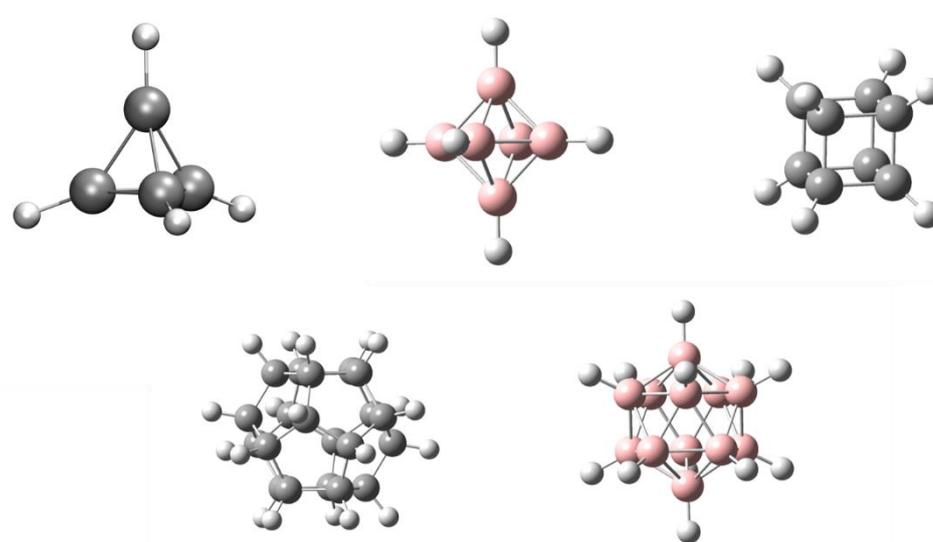


Figure 6. From left to right, top to bottom, molecular structures for  $C_4H_4$ ,  $B_6H_6^{2-}$ ,  $C_8H_8$ ,  $B_{12}H_{12}^{2-}$ , and  $C_{20}H_{20}$  with tetrahedral, octahedral, cubic, icosahedral, and dodecahedral main group atom cages.

The Monte Carlo simulations for these polyhedral molecules show that the shape and energy distributions do also follow log-normal distributions. As in the  $P_4$  case, the average values for both shape and energy have increase linearly with the temperature

(figure 7). We find that the larger the molecule, the higher is the increase in average energy, a finding that can be rationalized by the equipartition theorem,

$$\langle E \rangle = \frac{3}{2} N k_B T \quad (4)$$

Where N is the total number of atoms in a molecule. For the average polyhedral shape measure, however, this trend is no longer true since, for instance,  $B_6H_6^{2-}$  shows a higher variation than the larger  $C_{20}H_{20}$  molecule, thus other considerations must be taken into account.

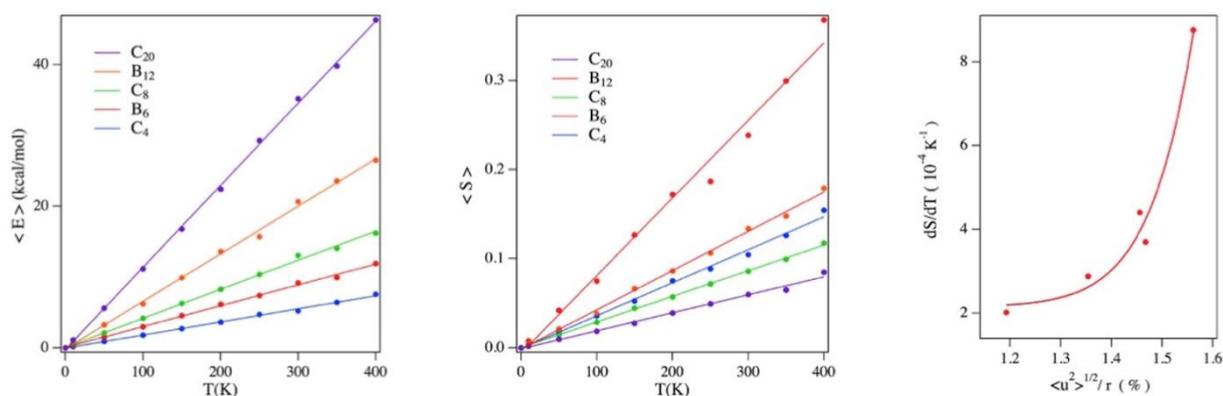


Figure 7. Average energy (a) and average polyhedral CShM (b) as a function of temperature for the series of  $C_4H_4$ ,  $B_6H_6^{2-}$ ,  $C_8H_8$ ,  $B_{12}H_{12}^{2-}$ , and  $C_{20}H_{20}$  molecules. (c) Slope of the average CSM vs T curves as a function of the relative atomic displacement at room temperature.

To explain this behavior, we must consider the motion of atoms around their equilibrium positions in the different normal modes of vibration (equation 1). As it has been explained above, the crucial factor determining the degree of change of the shape is the magnitude of the relative atomic displacement, that is, the magnitude of the average individual atomic displacements with respect to the overall size of the molecule, and not by the absolute magnitude of the displacements themselves. Since

according to equation (1) the value of the average displacement for a given atom varies depending on the mass of the atom and the frequency of the vibration under consideration, carbon-based molecules should differ significantly from boron-based ones. On one hand, the frequency of the totally symmetric breathing mode for the inner cluster gives an idea of the rigidity of the cage. Comparing tetrahedrane with dodecahedrane we find that the tetrahedron in  $C_4H_4$  is more rigid,  $\omega = 51$  THz, than the dodecahedron in  $C_{20}H_{20}$ ,  $\omega = 27$  THz. Taking into account eq. 1 this means that being the mass of the atoms equal, the average of the absolute displacement will be larger in the floppier dodecahedral cage than in the more rigid tetrahedral one. As mentioned, carbon and boron cages present different rigidity, as the former gives, for instance, a more rigid structure for  $C_8H_8$  ( $\omega = 38$  THz), than for the smaller boron octahedral cage in  $B_6H_6^{2-}$  ( $\omega = 39$  THz), hence we expect a larger average atomic displacement in  $B_6H_6^{2-}$  than in  $C_8H_8$ . Another important aspect that helps to understand the trends in the average shape variation is the radius of the polyhedra, which is important for comparing the average displacement relative to the size of the cage, which gives the key to comprehend the behavior of the average symmetry loss induced by temperature. As seen in figure 7, the larger the ratio between the average displacement of an individual atom (obtained from eq. 1) and the radius of the cage, the larger the slope on the average symmetry vs temperature curve, consequently proving that the observed trends are due to a combination of both the rigidity and the size of the cluster. This explains why similar sized molecules with different rigidities may show quite different symmetry losses upon heating.

When comparing a bare cage as in the  $P_4$  molecule with that in  $C_4H_4$  surrounded by a shell of hydrogen atoms, we find that the outer hydrogen atoms do not appear to have a significant effect on the shape and symmetry loss of the inner cage as the temperature rises. In fact, in the range of temperatures that we have studied the shape distribution of the outer hydrogen shell almost mimics that of the inner core. Considering previous statements, since the hydrogen atoms form a larger polyhedron, their average displacement relative to the size of the polyhedron will cause smaller variations to the shape and symmetry than in the inner polyhedron. Moreover, the H atoms are lighter and the C-H or B-H stretching modes have a much higher frequency,

thus we can approximately reason that in these molecules the hydrogen atoms will follow the heavier atom to which they are bonded and the result will be that they will barely influence the shape loss observed for the inner core. Note also that since there are no bonds between the H atoms in the outer shell, there is no other restriction to the movement of the H atoms than that given by the maintenance of the C-H or B-H bonds, which results again in the H atoms “following” the movement of the heavier atoms to which they are attached. According to the average CShM’s values, the hydrogen polyhedra lose more symmetry than the inner polyhedra, though the average shape loss for both polyhedra follow linear trends. For example, the carbon tetrahedron in tetrahedrane at 400K has an average tetrahedral shape measure of 0.15 while the outer hydrogen polyhedron presents an average symmetry loss that is practically the double, 0.29, a difference that can be explained considering the larger radius of the H sphere.

Finally, as all evidences seem to clarify, the magnitude of the average symmetry loss is not related to the number of operations in the symmetry group, so that at least for these cases, the intuitive idea that it should be easier to lose symmetry in a structure with a higher symmetry does, in principle, not hold and the magnitude of the symmetry loss does not depend on the number of symmetry operations in the minimal energy reference geometry since rigidity of the cage, independently of the number of the symmetry operations is the main factor to be considered in order to predict the extent of symmetry loss with temperature.

## Tetrahedral solids

In the last example that we will explore is how the temperature influences the local symmetry in a covalent crystal structure. We have analyzed a set of solids with a relatively simple crystal structure, the cubic diamond-type structure, the most prototypical one for tetrahedral elemental solids such as carbon, silicon, and germanium. In these cases, in the minimum energy configuration all atoms in the unit cell are equivalent and they present a perfect tetrahedral four-coordination. Raising the temperature will displace all atoms in the cell from their equilibrium positions, leading

to more or less distorted tetrahedra. The study of temperature effects in the shape and symmetry of crystal structures will be continued in a following chapter where we will apply the techniques developed here to the study of the symmetry of the phosphate ion in several crystals.

In contrast to the simulations presented above for single molecules, the actual simulations required for solid-structures have been performed using a different technique, molecular dynamics simulations using force fields in order to handle efficiently the sample of structures of the size required to analyze crystal structures. As stated above, the molecular dynamics simulations have been carried out using the *DLpoly* program. All simulations for C, Si, and Ge were run using a Tersoff potential<sup>19</sup> in the NPT ensemble at different temperatures and pressures. Along the simulation of 1.2 ns, a snapshot of the diamond-type super-cell containing 512 atoms was taken every 1 fs, except for the first 0.2 ns that were necessary to equilibrate the structure at the given P and T conditions. For each simulation the coordination environment was analyzed for about 51,712 individual atoms using continuous symmetry and shape measures, except in the cases at high temperatures where some atoms lost their initial four-fold coordination and were, therefore, not included in the shape analysis.

As expected, the average shape is affected by the temperature, giving more distorted tetrahedra at high temperatures and, as seen previously, we also find a linear dependence between the average shape and the temperature (Figure 8). A different outcome occurs when the pressure is raised since the effect of the pressure is contrary to that of the temperature and it prevents the molecules to distort due to the increase of the rigidity of the crystal structure with pressure.

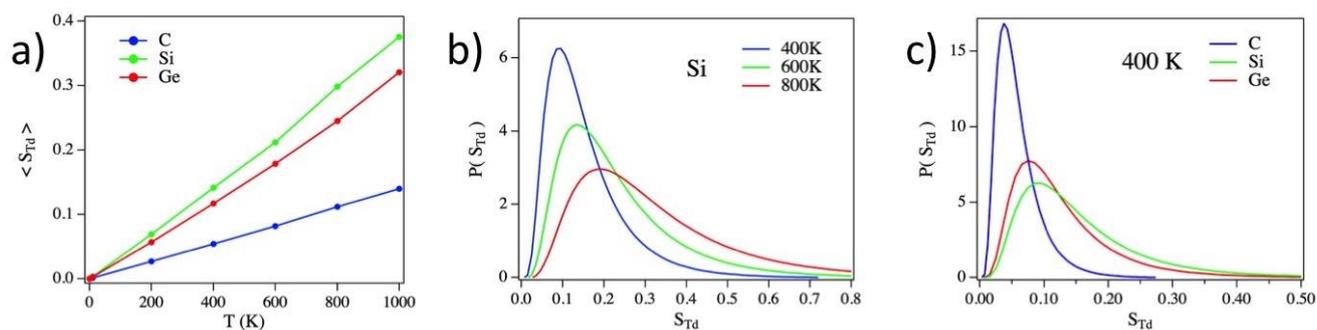


Figure 8. a) Average tetrahedral CSM for C, Si, and Ge atoms in the diamond-type crystal structure as a function of temperature. b) Distribution of tetrahedral CSM values for Si at different temperatures. c) Distribution of tetrahedral CSM values for C, Si, and Ge at  $T = 400\text{K}$ .

Concerning the effects of the temperature, the distribution of the tetrahedral CShM values at a fixed temperature again follows a log-normal distribution where the location of the maximum shifts towards higher values, meaning more distorted tetrahedra, as the temperature is increased. By comparing this effect at a given temperature for each of the three analyzed compounds, the most rigid carbon structure presents the narrower distribution with a peak closer to the perfect tetrahedron when compared to the other two other cases. However, no clear trends can be derived from the different nature of the atoms since Si and Ge present very similar distributions at the same temperature even though the intuition says that silicon should be more rigid than germanium and therefore having a lower and sharper peak in the CShM distribution. The intuition about the rigidity of the lattice increasing in the  $C > Si > Ge$  series is confirmed when the phonon spectrum is calculated giving a decrease from C to Ge for the transverse optical mode at the center of the Brillouin zone:  $\nu_{TO} = 48, 16$  and  $10$  THz for C, Si and Ge, respectively, which may be used as a qualitative measure of the crystal structure's rigidity. These findings contradict the results obtained in our simulations (figure 8) where the symmetry loss is found to be higher for Si than for Ge, so that other parameters should be considered to reach an explanation of the observed behavior. Once more, the explanation is found by looking at equation 1 since the mean atomic displacement depends not only on the frequency, but also on the mass and since the parameter determining the shape loss is not the absolute but the relative displacement,

the size of the tetrahedron should also be considered. This means that if the variation in the atom-atom distance is studied, the C-C distance (1.53 Å) is, as expected, smaller than the Si-Si (2.35 Å) and the Ge-Ge (2.45 Å) ones. Again, these factors explain why, even though the lower vibrational frequency in Ge should lead to larger displacements than in the other two cases, the heavier mass in Ge and the larger size of the tetrahedron, which means necessarily larger displacements to significantly have similar shape variations, explain why the temperature affects the tetrahedral symmetry loss considerably more in the more rigid Si structure than in Ge.

The last property that will be evaluated is the degree of local chirality induced by thermal motion in the compounds with diamond structure. As the temperature raises, the perfect tetrahedron around each atom will be distorted and, therefore, in general neither the planes nor the  $S_4$  improper rotations of the tetrahedral symmetry will be conserved. As a result, each distorted tetrahedron is in fact chiral, and the degree of chirality of each of them should increase with temperature. This is, indeed, the case for the three studied diamond-type structures. In the same way that rigidity controls the shape, it affects also the local chirality, and so the carbon structure presents smaller variations than the silicon case. Although temperature is able to induce chirality at a local scale, in the macroscopic framework this effect does not lead to any kind of optical activity. This is explained by considering that the local chirality that is induced leads to a mixture of two equally distributed distorted enantiomers, typically labeled as R and S, that lead to optically inactive macroscopic samples. We have analyzed for each temperature the ratio of R and S chiral distorted tetrahedra, following the scheme proposed by Avnir et al<sup>20</sup> to determine the handedness of a chiral distorted tetrahedron. Our results indicate that, as expected, at any given temperature, thermal motion leads to a racemic mixture of R and S tetrahedra with the calculated enantiomeric excess being always practically 0.

It should be noted that this result is only true when the equilibrium structure is achiral, that is, all local coordination environments are achiral as in the case of a perfect tetrahedron. In the case of structures where the local symmetry of the minimum energy configuration shows a certain degree of chirality, the local environments will distort and the average chirality will increase, although for small distortions the handedness of

each center will be maintained, while for larger distortions (higher temperatures) we could expect to find a switch in the handedness. The enantiomeric excess for the macroscopic structure in these cases will be governed by diverse factors such as the enantiomeric excess in the equilibrium geometry and the amount of temperature needed to induce a switch in the handedness for each different local chiral environment.

## Conclusions

The work in the present chapter illustrates the idea that continuous shape and symmetry measures are able to describe structural properties in molecules and solids, in this case derived from the temperature effect on to the molecular environment, in a concise and quite visual way. It therefore enhances the purpose of this thesis of disposing of a powerful tool such as the *CosymLib* library that allows the calculation and analysis of these measures in a simple manner.

As stated above, the previous examples show how temperature modifies the shape of a given structure with the main shape / symmetry contents, changing with temperature following a simple linear relation in which the rigidity and the size of the structure dictate the slope. The results presented may help in the future to predict the behavior that molecules and solids could exhibit as the temperature varies, giving tools to generate data to be treated by new machine learning approaches to help understand the complexity of some chemical structures. CSMs and CShMs also give the opportunity to extract the relevant structural changes from the random displacements of the individual atoms.

## Bibliography

1. Granovskii, B. L. & Ermakov, S. M. The Monte Carlo method. *J. Sov. Math.* **7**, 161–192 (1977).
2. Hastings, A. W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97–109 (1970).
3. Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L. *Handbook of Markov Chain Monte Carlo*. (Chapman & Hall/CRC, 2011).
4. MonteModes: <https://github.com/abelcarreras/MonteModes>.
5. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ort, and D. J. F. Gaussian 09, Revision D.01. (2016).
6. TinkerTools.org – Community Site for Tinker Users and Developers. <https://tinkertools.org/>.
7. Todorov, I. T., Smith, W., Trachenko, K. & Dove, M. T. DL\_POLY\_3: New dimensions in molecular dynamics simulations via massive parallelism. *J. Mater. Chem.* **16**, 1911–1918 (2006).
8. Verlet, L. Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev. Lett.* **159**, 98–103 (1967).
9. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
10. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
11. Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).
12. Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **13**, 1173–1213 (2007).
13. Sharma, G. *et al.* Electronic structure and dynamic properties of solid alkali cyanides. *J. Phys. Chem. A* **101**, 385–386 (1997).
14. Casanova, D., Alemany, P., Falceto, A., Carreras, A. & Alvarez, S. Pseudosymmetry analysis of molecular orbitals. *J. Comput. Chem.* **34**, 1321–1331 (2013).
15. Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
16. Chengteh Lee, Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*

- 37**, 785–789 (1988).
17. Tuvi-Arad, I., Rozgonyi, T. & Stirling, A. Effect of temperature and substitution on cope rearrangement: A symmetry perspective. *J. Phys. Chem. A* **117**, 12726–12733 (2013).
  18. Alvarez, S., Avnir, D., Llunell, M. & Pinsky, M. Continuous symmetry maps and shape classification. The case of six-coordinated metal compounds. *New J. Chem.* **26**, 996–1009 (2002).
  19. Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B* **39**, 5566–5568 (1989).
  20. Yogev-Einot, D., Pinsky, M. & Avnir, D. Left/right handedness assignment to chiral tetrahedral AB<sub>4</sub> structures. *Tetrahedron Asymmetry* **18**, 2295–2299 (2007).

## Chapter 5. Effect of temperature and the crystal environment on the shape of phosphate ions

### Introduction

As seen in the previous chapter, continuous shape and symmetry measures can be a handy tool to describe the effect of temperature on the atomic structure in molecules and solids. The instantaneous distortions suffered by the structure due to the vibrations at several temperatures can be analysed taking the arrangement of atoms at different times in a molecular dynamics simulation to get a picture of the average changes in shape and symmetry as the temperature is increased. Another important concept in the description of a molecular structure that is related to shape and symmetry is the coordination environment of an individual atom as it has been shown in chapters 1 and 4.

Up to now, the analysis of the coordination environment using continuous shape and symmetry measures has been focused on the metal atoms in transition metal complexes (see chapter 1), where the coordination environment is more plastic than for main group atoms or where it may be difficult to define a clear geometry for the arrangement of ligands like in the case presented in chapter 3, where even the aromatic rings can interact with the metal atoms to form structures with various possible shapes.

The case of the arrangement of bonded atoms around other non-metal or metalloid atoms is, nevertheless, of interest. The tetrahedral shape is commonly assigned to the  $sp^3$  carbon atom in  $CR_4$  molecules in organic chemistry, the  $SiO_4$  fragments in silica, silicates, and zeolites or any group 15  $AO_4$  molecule such as the  $PO_4^{3-}$  phosphate anion found in many substances as diverse as minerals or biological compounds. Although the variation in shape of the coordination environment has been not studied as much in detail as for transition metal complexes since their tetrahedral shape is assumed to be highly robust, there have been many attempts to synthesize square-planar

tetracoordinate carbon compounds<sup>1,2</sup>. The closest square-planar  $CR_4$  compounds found until now are fenestranes<sup>3</sup>, with R-C-R angles above  $130^\circ$ . However, the majority of studies on main group  $AB_4$  molecules have shown that in most cases the shape of the coordination environment is closely related to the tetrahedron and the deviations from this ideal shape have been quantified by means of indices related to angle and bond length averages that have been analysed for several cases<sup>4,5</sup>.

A well-known case is the phosphate anion that appears in the crystals of many minerals. Together with methane and the  $SiO_4$  groups in silicates, the phosphate anion ( $PO_4^{3-}$ ) is, indeed, a textbook example for the tetrahedral coordination of a main group atom and it is assumed to adopt a perfect tetrahedral geometry. However, in 1973, Baur<sup>6</sup> published a thorough crystallo-chemical analysis of the site symmetry for hundreds of these phosphate groups in crystalline solids revealing that, in general, the symmetry of the phosphate group in crystals is not tetrahedral, and to the general surprise, none of analysed structures by Baur had the P atoms located at sites with the  $-43m$  ( $T_d$ ) crystallographic point group symmetry<sup>6</sup>.

The main aim of this chapter is to revise first the departure of local tetrahedral symmetry in the set of phosphate anions analysed by Baur by means of continuous shape and symmetry measures to observe if the lack of the exact tetrahedral symmetry in these structures may be relevant from a chemical point of view or if this departure is so small that, to a good approximation, the phosphate group can be considered to be tetrahedral. Previous studies show that although the local symmetry of a given fragment may be incompatible with the overall crystallographic symmetry, these fragments retain to a high degree this shape/symmetry, as for example in sets of five bonded atoms forming a practically perfect pentagon with local  $C_5$  rotational symmetry<sup>7</sup>. Thus, in this chapter we want to analyse if  $PO_4^{3-}$  anions present a similar behaviour, with a fairly perfect local tetrahedral symmetry embedded in an environment with a lower crystallographic symmetry, or if indeed the low symmetry of the environment is in some cases able to induce significant distortions in the shape of the  $PO_4$  fragments. For this we will study on one hand the effect of the atoms surrounding the  $PO_4$  units in the distribution of continuous tetrahedral shape measures. To end, we will discuss the results of several Monte Carlo simulations at different

temperatures carried out to quantify the symmetry loss due to molecular distortions produced by the thermal motion on the phosphate anions. Moreover, a few molecular dynamics simulations for different ionic solids containing phosphate anions will be analysed. Finally, we have performed a pseudosymmetry analysis to study the effects of the observed deviations from the ideal tetrahedral shape in the symmetry of the molecular orbitals.

## Molecular simulations for isolated phosphate anions

Before analysing the set of structures comprised in Baur's report, we will look at the tetrahedral CShM for different representative isolated phosphate anions to understand the extent of the symmetry loss that can be expected from the atomic motion induced by temperature solely. In his report Baur classified the phosphate structures in two main groups, the orthophosphates where each oxygen atom is bonded to only one phosphorus atom and the condensed phosphates, where some of the oxygen atoms of a phosphate anion are bonded to more than one phosphorus atom. In Baur's study, each of these two groups is further divided into three subgroups, proper phosphates, acid phosphates (with one or more hydrogen atoms per phosphate unit) and organic phosphates (with one or more oxygen atoms bonded to carbon atoms). Within this classification, phosphates belong to the orthophosphate group, while the diphosphates (with one oxygen atom bonded to two phosphorus atoms), the ring phosphates (where three to six phosphate anions form a ring) and polyphosphates (the rest of compounds which are neither diphosphates nor ring phosphates) fit in the group of condensed structures.

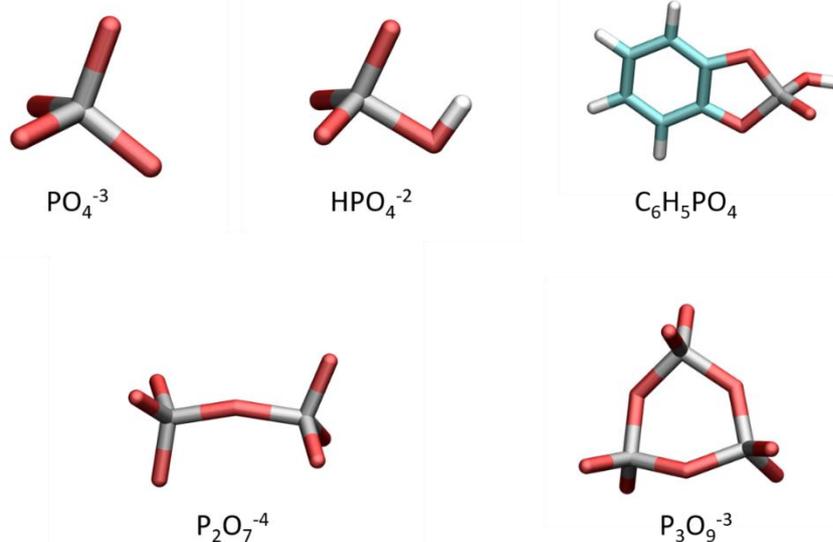


Figure 1. Optimized isolated phosphate structures representative of each Baur group. Top structures are orthophosphate molecules and bottom structures are condensed phosphate structures.

As a first approximation to the problem, several Monte Carlo simulations were run with the *MonteModes*<sup>8</sup> program to figure out the effect of temperature on the shape of the coordination environment of the phosphorus atom in the phosphate anion, the hydrogen phosphate anion and the diphosphate anion in the gas phase (Figure 1), which represent the perfect tetrahedral phosphate anion and one representative example from orthophosphate and the condensed phosphate groups, respectively. The simulations were run considering that the acceptance factor should converge to an accept/reject ratio of 50% of the generated structures. The calculation ends when 30500 molecular structures have been found in the range of the acceptance ratio, which depends on the energy variation. The energies were computed at the STO-3G/B3LYP level using Gaussian 09<sup>9</sup>, although no significant differences are expected using other computational methods since only qualitative energy differences are of interest in this study.

The calculated molecular energies differences with respect to the minimal energy configuration at different temperatures show a lognormal distribution (equation 1) as

described by Tuvi-Arad et al.<sup>10</sup> In addition, the CShM values were also computed and found also to follow a log-normal distribution (figure 2).

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (1)$$

For the isolated  $\text{PO}_4^{3-}$  anion, at room temperature the tetrahedral CShM for the different distorted structures is mainly below 0.6 and at temperatures below 100K, the phosphate anion should be considered, for practical purposes, to be tetrahedral since most of the values obtained for the CShM are  $< 0.1$ , which as discussed in chapter 1, are values indicating small distortions without chemical interest. As in the other examples shown in chapter 4, the average CShM increases linearly with temperature (figure 2), reaching a value of approximately 0.2 at room temperature.

As seen in the previous chapter, the extent of the local motion of an atom can be estimated for a given temperature considering the frequencies of the normal modes of the molecule, taking into account, however, that in this case the comparison is not as straightforward as for the  $\text{P}_4$  molecule in the previous chapter, since in the definition of the normal modes the use of mass weighted cartesian coordinates complicates the comparison with the CShM values which consider only geometric aspects of a molecular structure, irrespective of the relative masses of the different atoms that form it.

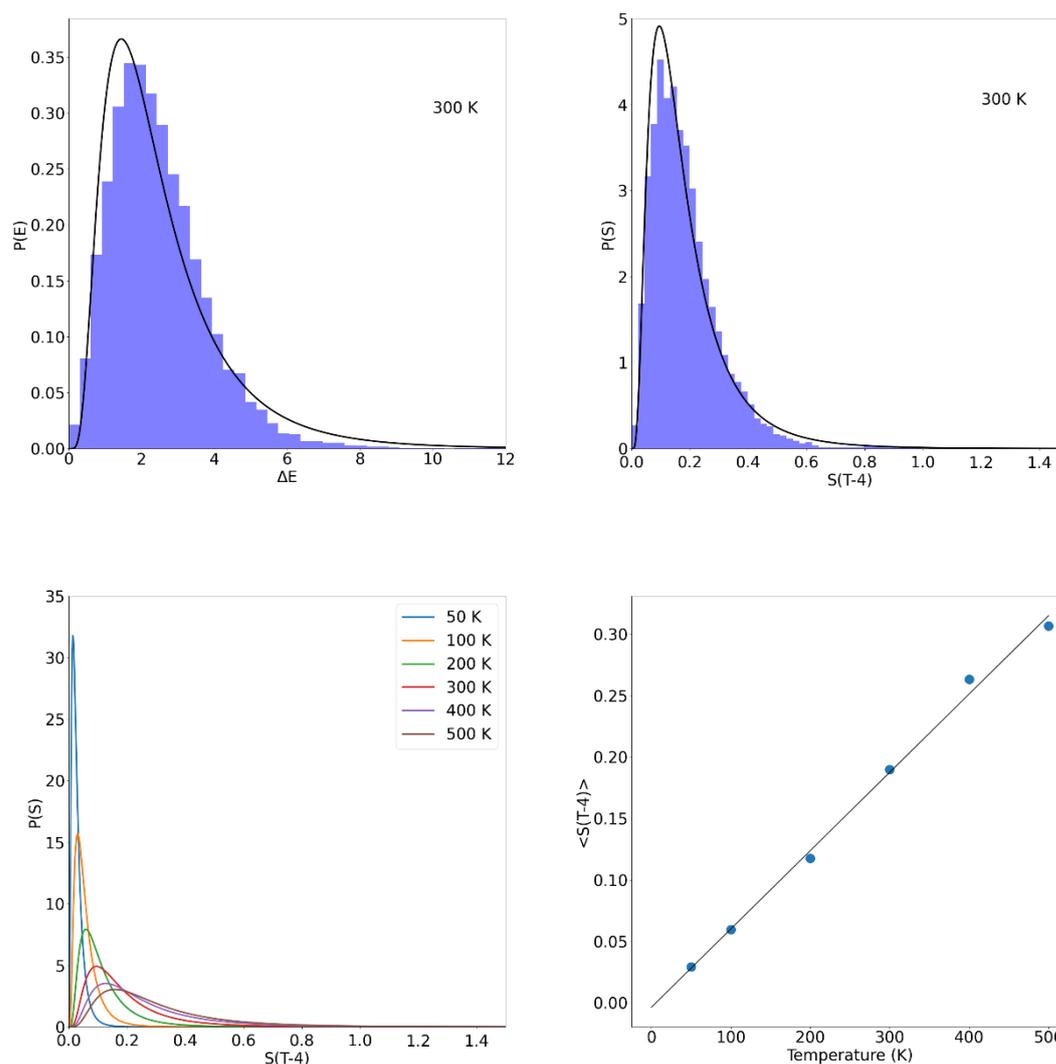


Figure 2. Probability distribution for the energy relative to the minimum energy structure (top-left) and for the tetrahedral CShM (top-right) for the isolated  $\text{PO}_4^{3-}$  anion at 300K. Log-normal distributions of the CShM at different temperatures (bottom-left) and average CShM as a function of the temperature (bottom-right) for the isolated phosphate anion.

For structures with only one type of atom such as the  $\text{P}_4$  molecule the comparison is straightforward since displacement in the normal modes is equivalent for all atoms. In the case of  $\text{PO}_4$ , however, the displacement of the heavier P atoms in the normal modes is divided by a larger factor than for the lighter O atoms, making a direct comparison between the symmetry loss and the distortions along the normal modes not straightforward and it is not possible to extract general equations relating the

symmetry loss and the frequency of a given mode. In general terms, nevertheless, we find that the distortions along the soft, low-frequency modes are those that lead to the largest symmetry loss for a given energy as shown in figure 3. Note that for this case, a strong anharmonicity is found for the E normal vibration modes which is evident in the separation of the S(T<sub>4</sub>) vs E lines for the displacement along these modes in the positive and negative directions.

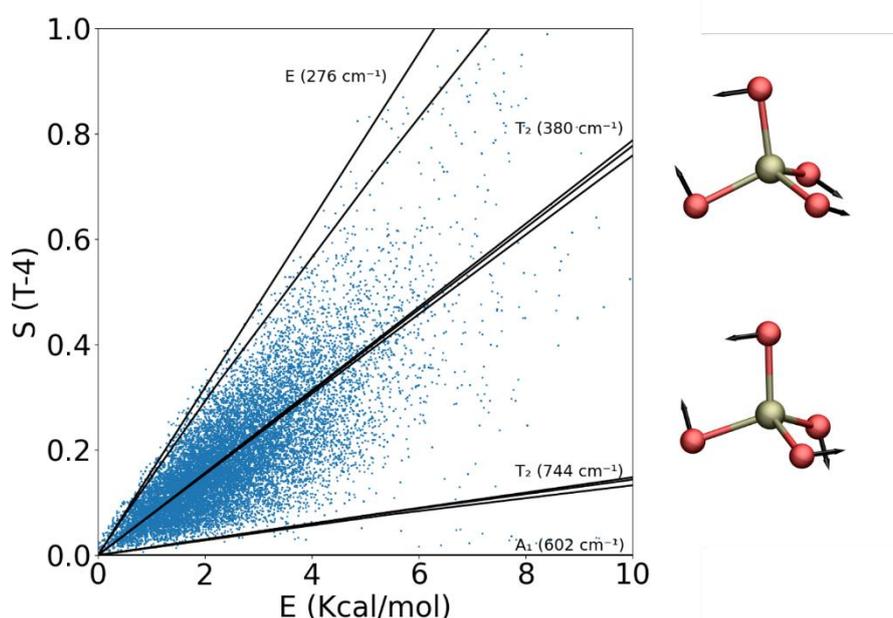


Figure 3. CShM as a function of the energy (blue dots) of the phosphate anion structures found in a Monte Carlo simulation at 300K. The change of the CShM as a function of the energy for structures distorted along the normal modes are shown in black. Displacement of the atoms in the E normal modes (black arrows) are shown at the right part of the figure.

In the case of the diphosphate anion, similar conclusions can be derived. In this case, however, the distribution of CShM values is distinct since the two phosphate fragments in the minimal energy structure of  $P_2O_7^{4-}$  present already a slight deviation from the ideal tetrahedral symmetry ( $S(T-4) = 0.54$ ). At very low temperatures thus the average CShM will be already different from 0 and the maximum of the log-normal distribution will be centred around this  $S(T-4)$  value. For the isolated phosphate anion, where the

minimum energy structure is perfectly tetrahedral, an increase in the temperature may only result in less tetrahedral distorted structures. This is, however, not the case for the diphosphate anion, where some distortions away from the minimal energy structure may counterintuitively result in more symmetric structures (figure 4).

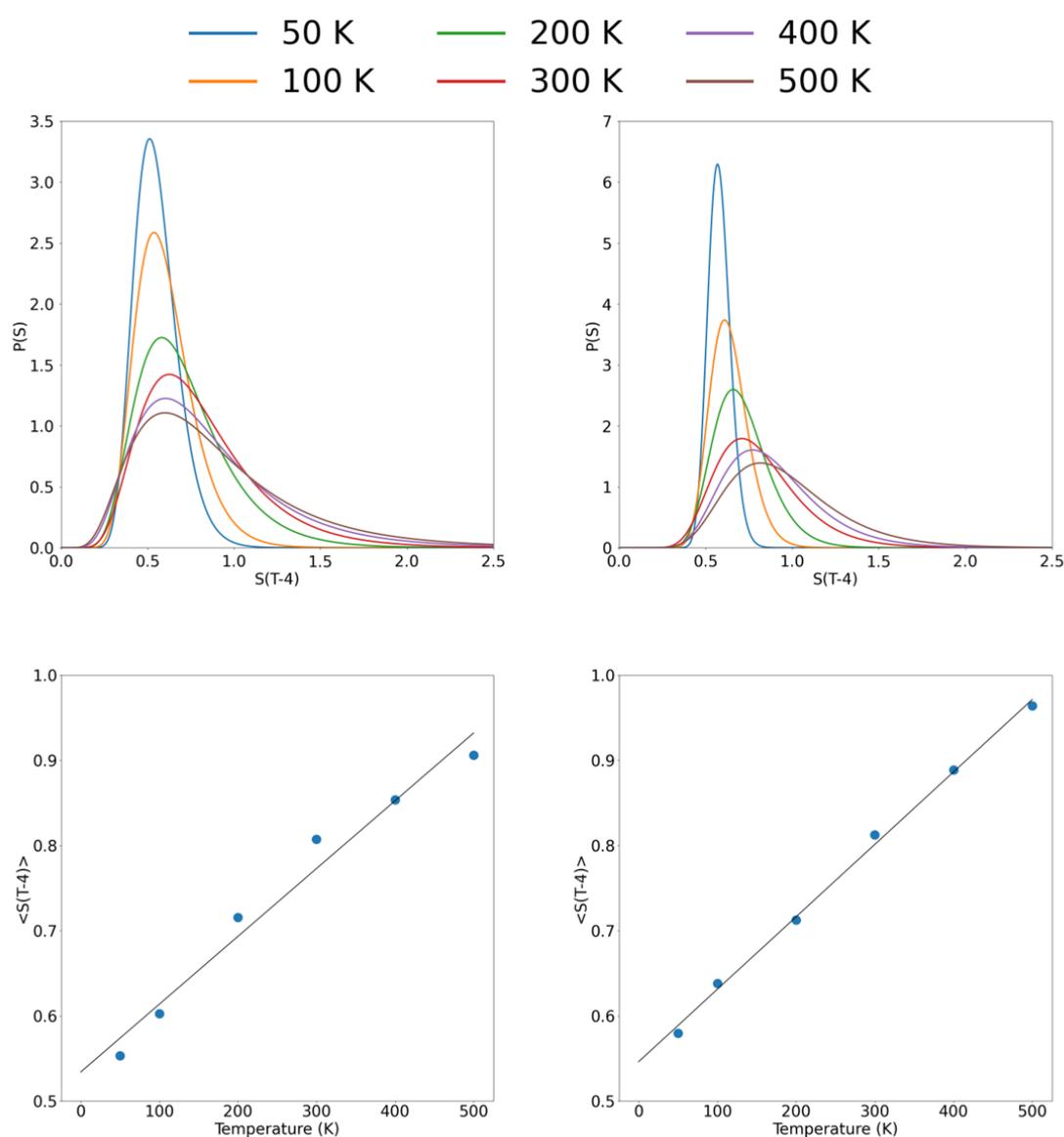


Figure 4. Distribution of the tetrahedral CShM for the diphosphate (left-top) and hydrogen phosphate (right-top) anions at different temperatures. Mean CShM values for diphosphate (bottom-left) and hydrogen phosphate (bottom-right) as a function of temperature.

Nevertheless, when the temperature is increased, the probability of having more symmetrical distortions is much lower than for those with higher values of the CShM and the net effect is that the distribution for the CShM values expands in such a way that the maximum of the distribution, and hence, the average CShM moves to higher values as the temperature is increased.

If the minimal energy geometries are compared, the  $\text{PO}_4$  unit in the hydrogen phosphate presents a shape value,  $S(\text{T-4}) = 0.55$ , close to that obtained for the two  $\text{PO}_4$  units in the diphosphate anion. If we consider only the shape of the  $\text{O}_4$  tetrahedron formed by the isolated vertices, we can, however, find significative differences between both cases. In the case of the diphosphate anion, the CShM value for the  $\text{O}_4$  fragment is 0.44, closer to the value obtained for the whole  $\text{PO}_4$  fragment. On the contrary, for the  $\text{O}_4$  fragment in the hydrogen phosphate molecule we find a significantly lower value,  $S(\text{T-4}) = 0.07$ , pointing to a negligible distortion from the perspective of shape. These results tell us that the origin of tetrahedral shape loss of the  $\text{PO}_4$  units in the two cases is different. In the case of the hydrogen phosphate anion, the main distortion is an off-center displacement of the phosphorus atom while keeping the shape of the  $\text{O}_4$  fragment more or less equal. In the diphosphate case, contrarily, the main source of asymmetry is due to angular distortions of the O-P-O angles that affect both the tetrahedral shape of the  $\text{PO}_4$  and the  $\text{O}_4$  fragments.

Despite the fundamental difference in the origin of the distortion for the minimal energy structure, when raising the temperature, the  $\text{PO}_4$  units in both anions present similar deviations from the perfect tetrahedron (figure 4), with similar average  $S(\text{T-4}, \text{PO}_4)$  values at each temperature.

If we then compare the evolution of the average  $S(\text{T-4}, \text{PO}_4)$  for all three anions,  $\text{PO}_4^{3-}$ ,  $\text{P}_2\text{O}_7^{2-}$ , and  $\text{HPO}_4^{2-}$ , the linear regression lines show how for all three structures the rates of shape loss for the  $\text{PO}_4$  tetrahedra are comparable, with only small variations in their slopes,  $6 \times 10^{-4}$ ,  $9 \times 10^{-4}$ , and  $8 \times 10^{-4}$  for the phosphate, hydrogen phosphate and diphosphate anions, respectively (figure 5).

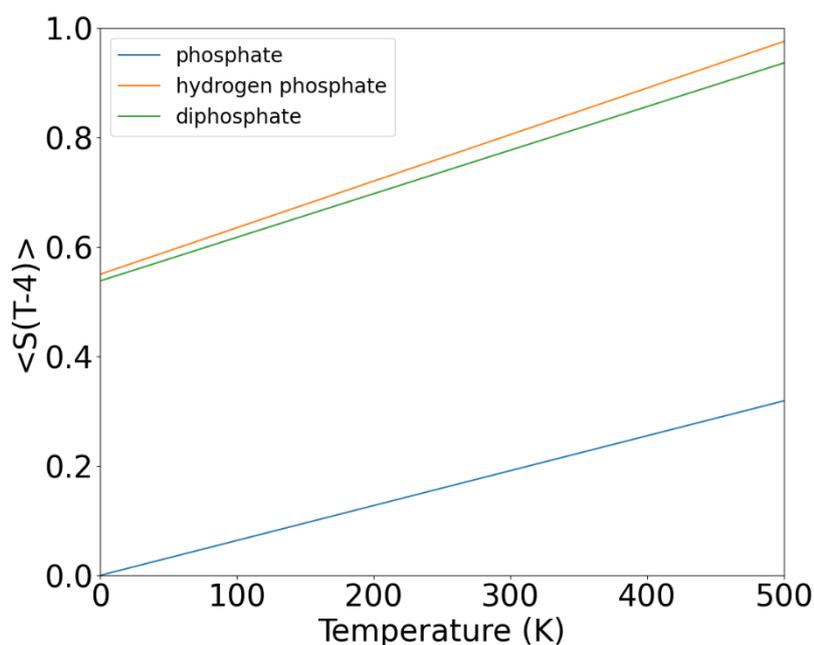


Figure 5. Phosphate Monte Carlo CShM values at different temperatures for the phosphate anion, hydrogen phosphate and diphosphate.

An interesting question when analysing the shape loss for different molecules with a common element such as the  $\text{PO}_4$  unit in phosphates is to see if the underlying mechanisms yielding low energy structures are the same. We have already seen that looking at the tetrahedral shape measure for both the  $\text{PO}_4$  and the  $\text{O}_4$  fragments it is possible to distinguish between two fundamental shape loss sources: the off-center displacement of the P atom and the angular O-P-O distortions. Another way for obtaining a more detailed view on these aspects is to compute CSMs for different symmetry groups, all of them subgroups of the  $T_d$  group. Analysing the CSM values for the  $C_{3v}$ ,  $C_{2v}$ ,  $S_4$  and  $C_s$  groups for the structures obtained in simulations at different temperatures we find also log-normal distributions for all of them. If we look at the evolution of the average CSM values with temperature for the phosphate anion we can see that the symmetry corresponding to the higher symmetry subgroups,  $C_{3v}$  and  $S_4$ , is lost much more than the lower symmetries in  $C_{2v}$  and  $C_s$ .

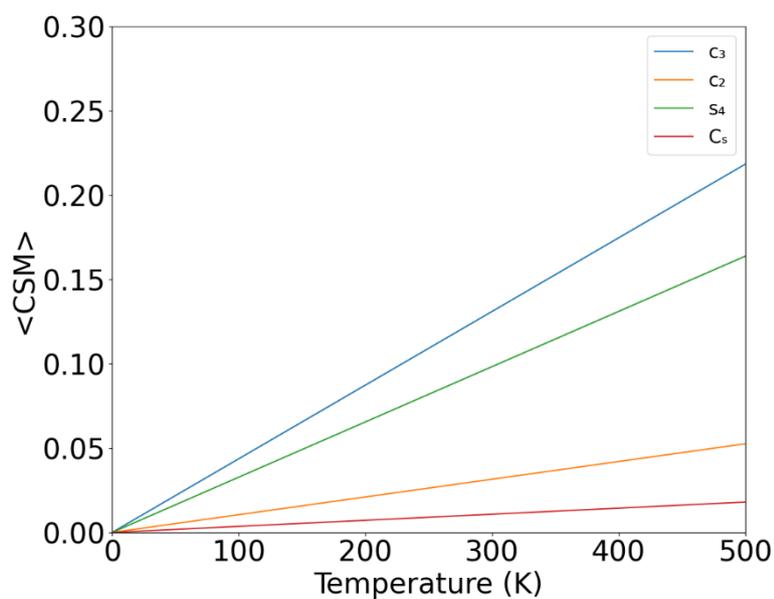


Figure 6. Phosphate continuous symmetry measures of the  $C_3$ ,  $C_2$ ,  $S_4$  and  $C_s$  symmetry groups.

However, the case of hydrogen phosphate and diphosphate is slightly different. Both structures present a higher CShM deviation than the distortion found in phosphate, thus a higher deviation in their CSM values is expected.

In the hydrogen phosphate structures, the highly distorted values are found for the  $S_4$  and  $C_2$  symmetries, while  $C_3$  and  $C_s$  symmetries remain almost unchanged. The results are similar in the case of the diphosphate anion, where the  $C_3$  and  $C_s$  changes with temperature are almost constant (figure 7).

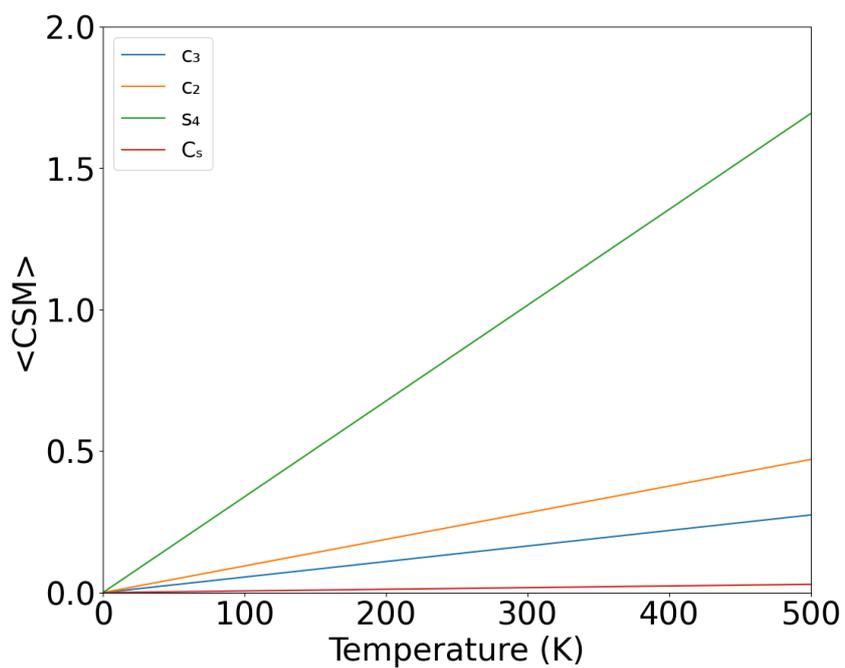
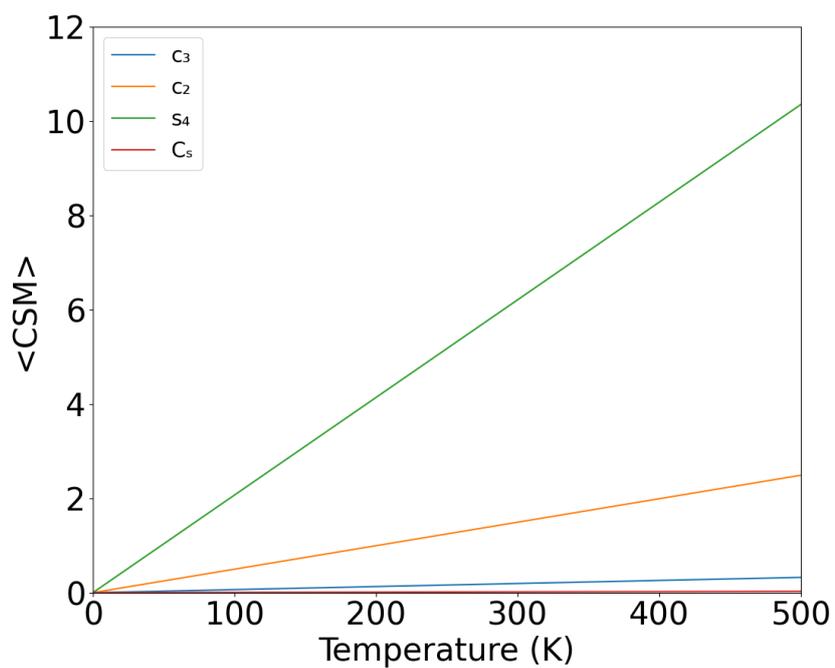


Figure 7. Hydrogen phosphate (top) and diphosphate (bottom) continuous symmetry measures of the  $C_3$ ,  $C_2$ ,  $S_4$  and  $C_s$  symmetry groups.

## CSM analysis for experimental phosphate crystal structures

Of the 211 phosphate groups found in the 129 crystal structures analysed in Baur's report<sup>6</sup>, complete structures for only 197 of them were found in literature nowadays, 155 in the ICSD Database<sup>11</sup> and 42 in CCSD<sup>12</sup> Database. The missing 14 structures still appear as unpublished data and have been discarded in our analysis. Nevertheless, the amount of available data was considered sufficient to get an idea of the behaviour of phosphate molecules in crystal structures comparing Baur's results with the information obtained from a CShM based analysis. In the original study, Baur measured the amount of deviation from the ideal tetrahedron by introducing three different distortion indexes based on structural parameters such as interatomic distances or angles. The three distortion indexes are calculated for the P-O distances (bond distance), O-P-O angles (tetrahedral angle) and O-O distances (edge distance) by the following equations,

$$DI (PO) = \left( \sum_{i=1}^4 \frac{|PO_i - PO_m|}{4 * PO_m} \right) \quad (2)$$

$$DI (OPO) = \left( \sum_{i=1}^6 \frac{|OPO_i - OPO_m|}{6 * OPO_m} \right) \quad (3)$$

$$DI (OO) = \left( \sum_{i=1}^6 \frac{|OO_i - OO_m|}{6 * OO_m} \right) \quad (4)$$

where  $PO_i$ ,  $OPO_i$ , and  $OO_i$  are the individual P-O, O-P-O, and O-O distances or angles of a given phosphate structure and  $PO_m$ ,  $OPO_m$ , and  $OO_m$  are the corresponding average or mean values. Note that, by definition, the three indices are exactly 0 in the case of a perfect  $O_4$  tetrahedron with the P atom located at its centre.

If the distortion indexes are calculated for the phosphate structures included in Baur's report (figure 8), the major deviation found for each individual group of structures is in the DI(OPO) angular distortion index, while almost no distortion is found for the D(OO) edge's distortion index. This trend is found for all subgroups, except for the proper phosphates, indicating that attaching other atoms to the oxygen atoms induces a heavier distortion in the P-O bonds and O-P-O angles than in the O-O edges of the tetrahedron. In other words, the  $O_4$  tetrahedron is left practically unchanged while the phosphorus atom is moved away from its centre.

An alternative approach to the previous study can be achieved by calculating the tetrahedral shape measure for each phosphate anion. These  $S(T_d)$  values quantify the amount of distortion of each  $PO_4$  fragment by estimating how far the phosphate anions are from the ideal  $ML_4$  tetrahedral shape. The distribution of  $S(T_d)$  values for each of the subsets of phosphates included in Baur structures are shown in figure 9. We have normalized the histograms by dividing by total number of structures in each set, so that for a set with all fragments with the same  $S(T_d)$  value we would get a single bin of unit height. In this way the different distributions can be compared regardless of the number of structures in each set. A first look at the histograms in figure 9 tells us that all sets present, in average, larger deviations from the tetrahedral symmetry than the set of proper phosphates.

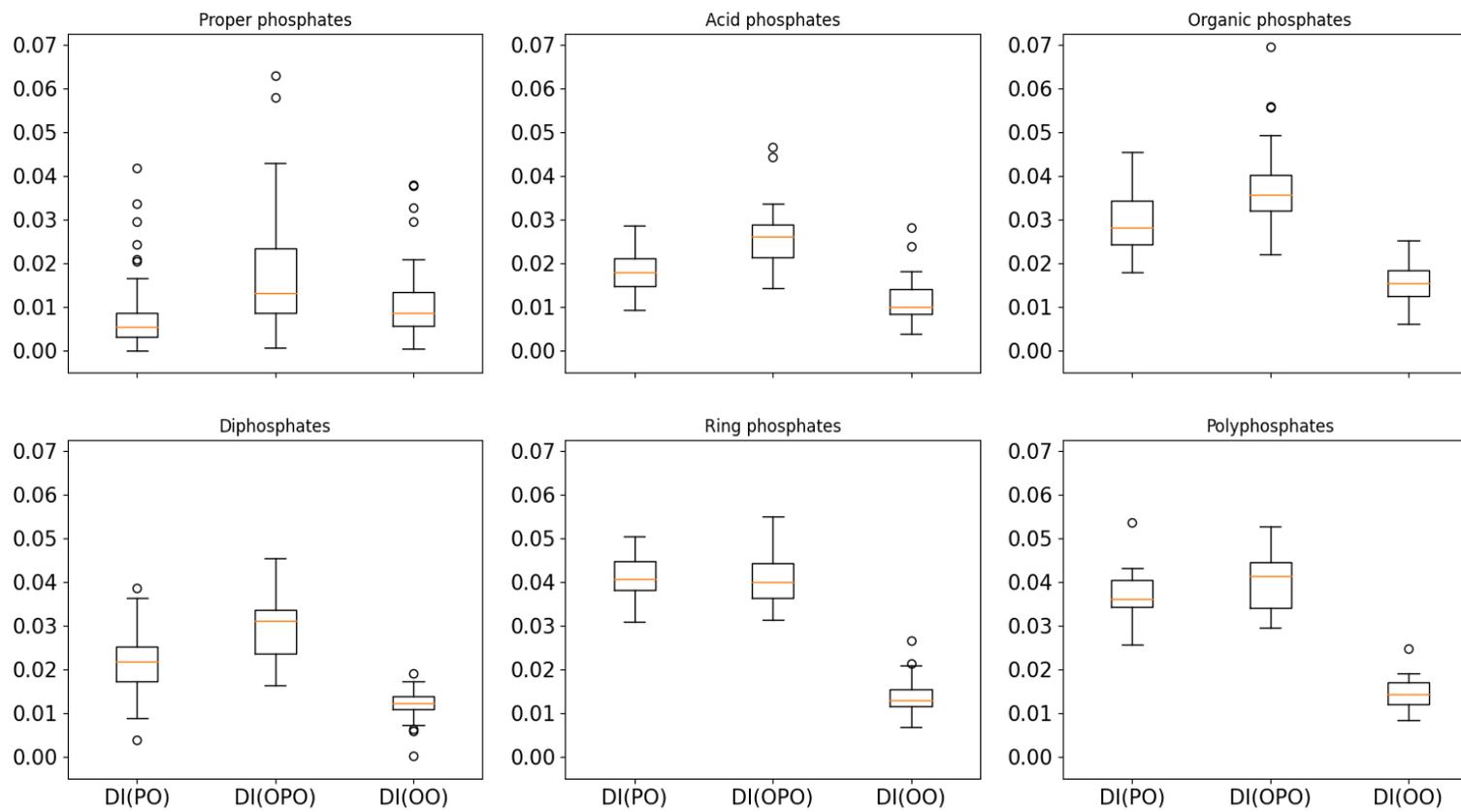


Figure 8. Box plots for the DI(PO), D(OPO), and D(OO) distortion indices for the different sets of phosphate structures included in Baur's report<sup>6</sup>.

It should be noted that the highly distorted structures included by Baur in the proper phosphate set correspond to the  $\gamma$ -zinc orthophosphate and the leucophosphate crystals (ICSD number = 34303 and 22048, respectively) where the  $\text{PO}_4$  tetrahedra are strongly distorted due to the sharing of oxygen atoms of the phosphate  $\text{PO}_4$  anion with an  $\text{MO}_6$  octahedrally coordinated metal ion. This type of structures form a set of its own in the extensive classification of mineral phosphates by Huminicki & Hawthorne<sup>13</sup>, where this type of structures are classified in different subsets depending on the participation of the phosphate anions in the coordination sphere of the octahedral metal atoms. The distortion of the  $\text{PO}_4$  tetrahedron arises from the fact that the metal ions force the oxygen atoms to arrange close to an octahedron and the phosphate anion is accordingly distorted to accommodate to the steric requirements of the  $\text{MO}_6$  polyhedron, with the final structure being determined by a balance of the energies needed to distort the  $\text{PO}_4$  and the  $\text{MO}_6$  polyhedra, with metals giving very rigid  $\text{MO}_6$  octahedra forcing larger distortions in the adjacent phosphate groups.

As discussed above, the distribution of  $S(\text{T}_d)$  values alone do not allow us to differentiate between the distortion produced by an off-center displacement of the P atom from the distortions associated to the  $\text{O}_4$  tetrahedron. To do this we can compare the  $\text{ML}_4$  tetrahedral shape measure of the whole  $\text{PO}_4$  phosphate anion with the tetrahedral  $\text{L}_4$  shape measure for the  $\text{O}_4$  fragment. In figure 10, the distribution of both measures for the set of structures included in Baur's report are shown using box plots, where it is evident that, in general,  $S(\text{T}-4, \text{PO}_4) > S(\text{T}-4, \text{O}_4)$  for all sets except in that of the proper phosphates. These results yielding, of course, to the same conclusion obtained comparing the distributions for the  $\text{DI}(\text{PO})$  and  $\text{DI}(\text{OO})$  distortion indices, that is, the phosphate group tends to adapt to the asymmetry in its close environment by displacing the P atom out of the central position rather than by distorting its external tetrahedral  $\text{O}_4$  hull.

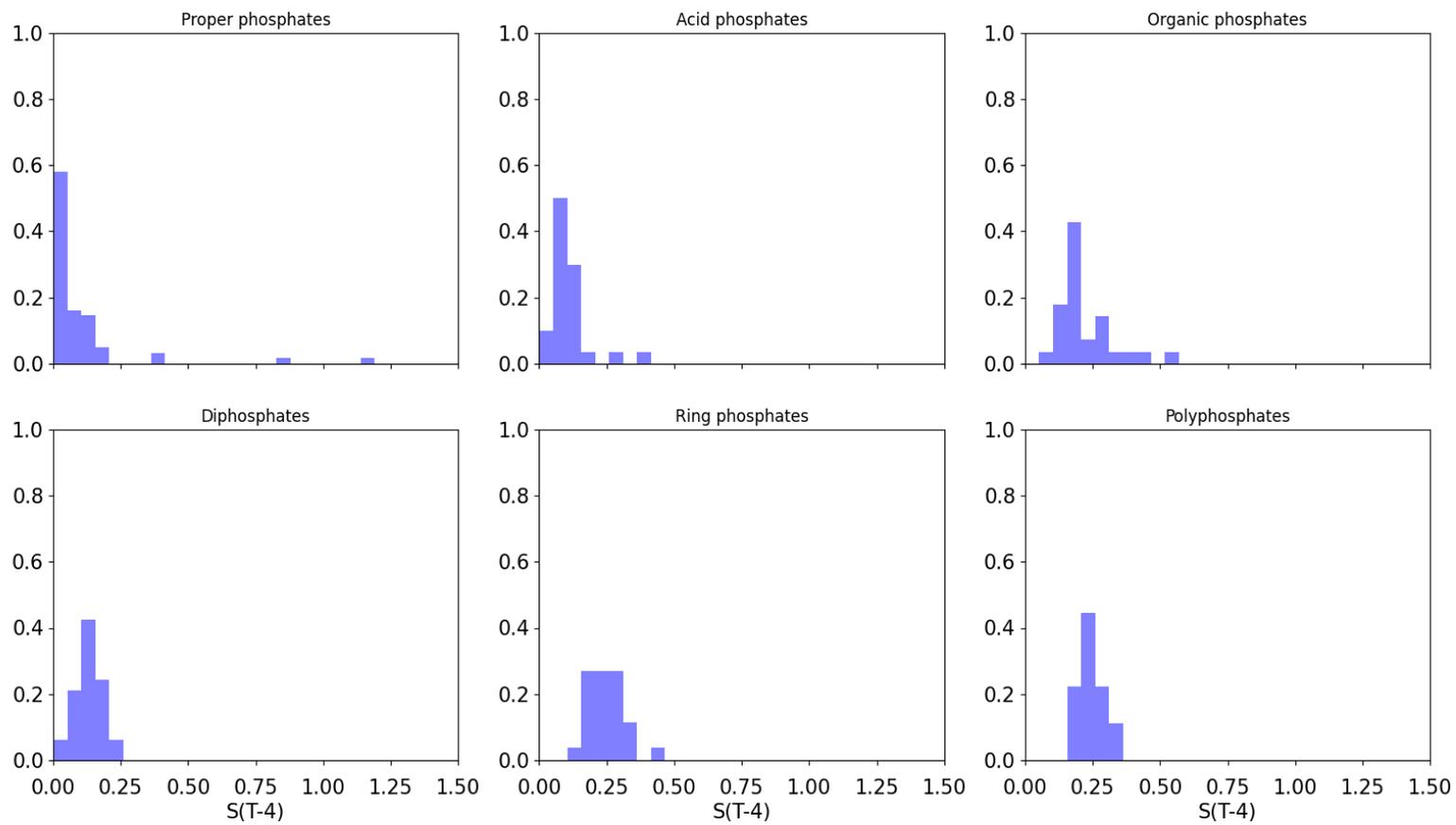


Figure 9. Distribution of the  $S(T-4, PO_4)$  shape measures for the different sets of phosphate structures included in Baur's report<sup>6</sup>.

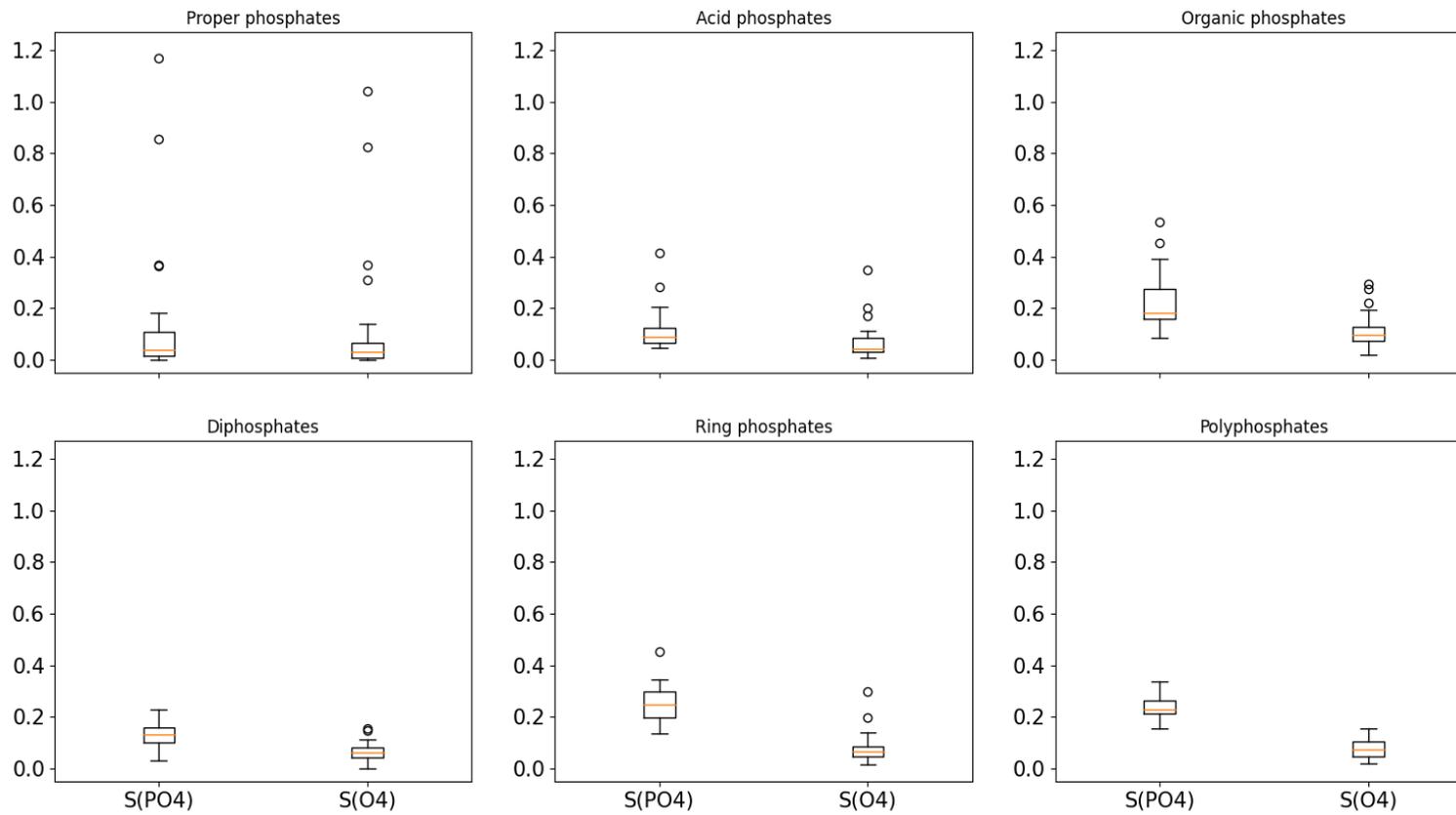


Figure 10. Box plots for the distribution of the  $S(T-4, PO_4)$  and  $S(T-4, O_4)$  shape measures for the different sets of phosphate structures included in Baur's report<sup>6</sup>.

More information on the effect of the environment on the shape of the phosphate group may be obtained by comparing the distribution of  $S(T-4, PO_4)$  values for the sets of proper phosphates, acid phosphates and diphosphates with those obtained in the Monte Carlo simulations for the corresponding isolated anions at room temperature (figure 11). At a first sight we can observe that for all three types of phosphates, the distribution of CShM measures is much narrower for the crystal structures than for structures obtained in the MC simulations. Besides this, the average CShM values are also much lower for the  $PO_4$  groups found in crystal structures than for those in the MC simulations. Although these trends are common for the three types of phosphates, the differences in the CShM distributions between crystal structures and MC simulations are much smaller for the proper phosphate group than for the other two cases. The results shown in figure 11 show clearly that the crystal environment has in reality a double-faced character with respect to the symmetry of the phosphate group. On one hand, as noted by Baur, it is a source of asymmetry, with highly symmetric  $PO_4$  fragments being an exception. On the other hand, however, the quantification of the symmetry loss using CShMs allows us to see that the crystal environment is also preserving the symmetry of the  $PO_4$  groups by limiting the extent of distortion due to thermal motion in the free ions. Similar results are found for the  $O_4$  analysis for Baur's and Monte Carlo structures, with the former still preserving the tetrahedral symmetry due to the crystal environment and the later distorting the oxygen atoms to a higher degree.

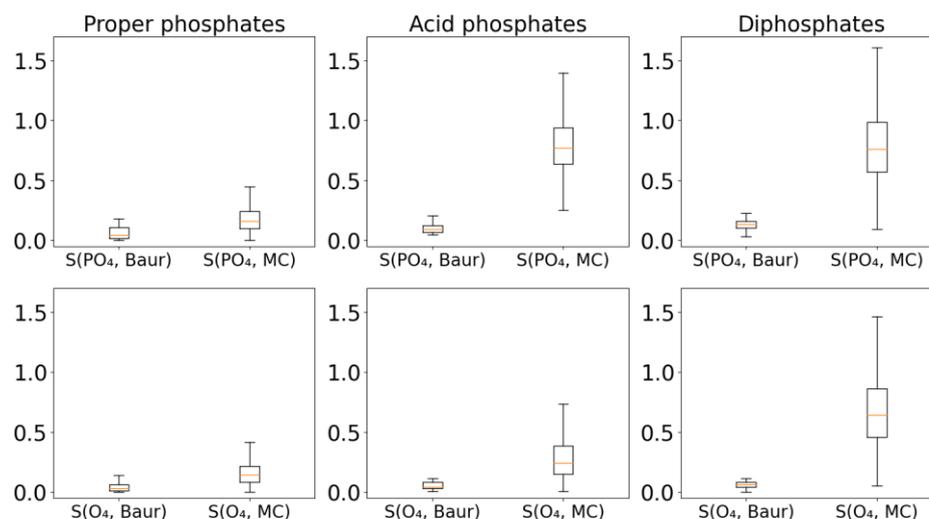


Figure 11. Comparison of the distributions of tetrahedral CShMs for the proper phosphates, acid phosphates, and diphosphates in Baur's report with those obtained in the Monte Carlo simulations at 300K. Top results are for the  $\text{PO}_4$  fragment and bottom results are related to the  $\text{O}_4$  fragment.

To have a reference of the magnitude of the symmetry loss in each set of different types of phosphate structures we have optimized the structure for a set of isolated phosphates, each belonging to one of the different sets in Baur's report. The optimized isolated phosphate anion is perfectly tetrahedral. As expected, the set of proper phosphates included in Baur's report present higher  $S(\text{T-4}, \text{PO}_4)$  since any asymmetry in the environment will cause some degree of departure from the ideal tetrahedral structure.

As seen in figure 12, the effect of bonding different atoms to the oxygen atoms of a  $\text{PO}_4$  fragment has a notable influence on its tetrahedral shape. For example, adding just a H atom as in the hydrogen phosphate increases the  $S(\text{T-4}, \text{PO}_4)$  measure to 0.55 while keeping the  $\text{O}_4$  tetrahedron practically untouched with  $S(\text{T-4}, \text{O}_4) = 0.07$ . The value of  $S(\text{T-4}, \text{PO}_4)$  for the isolated anion is much higher than the average found in the crystal structures of the acid phosphates with  $S(\text{T-4}, \text{PO}_4)$ , which are below 0.25 (figure 12).

The comparison of the  $S(\text{T-4}, \text{PO}_4)$  and  $S(\text{T-4}, \text{O}_4)$  measures for the optimized structures shows the same trend that is found for the distributions in figure 10, that is,

that the main source of distortion arises from the off-center displacement of the P atom, while keeping the tetrahedral shape for the  $O_4$  fragment mostly unaltered. The only exception for this trend is in the case of the organic phosphates, where geometric constraints in the organic part may induce large displacements of the O atoms from their positions in a perfect tetrahedron. By comparing the case of the hydrogen phosphate and the diphosphate anions it is also evident that attaching a single hydrogen atom to a  $PO_4$  group induces a much larger distortion due to the displacement of the P atom from its central position than when joining two tetrahedra that are sharing a common vertex. In the case of the diphosphate and triphosphate anions the  $S(T-4, O_4)$  measures indicate the presence of a non-negligible distortion that is not originated simply by the displacement of the P atom away from the centre.

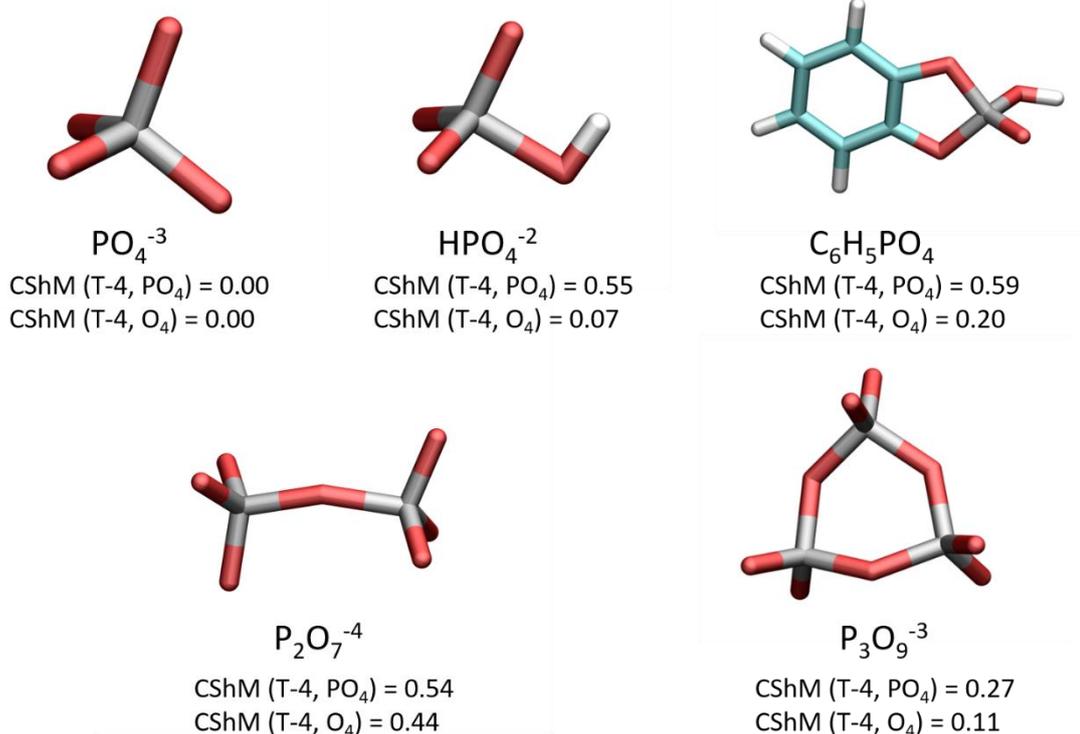


Figure 12. Optimized isolated phosphate structures representative of each Baur group and their CShM tetrahedral values for the phosphate and  $O_4$  fragment.

To further analyse the deviation from the ideal tetrahedral symmetry found for the structures included in Baur's report we have calculated the shape measures along the most usual distortion paths found for tetrahedral  $ML_4$  compounds to see if the crystal environment is modifying the structure of the phosphate anions in a certain preferred direction. For this we considered the angular distortions of a tetrahedron described in Cirera et al.<sup>14</sup> together with the distortions related to the off-centre displacement of the central atom which seem to be particularly important in the acid phosphates.

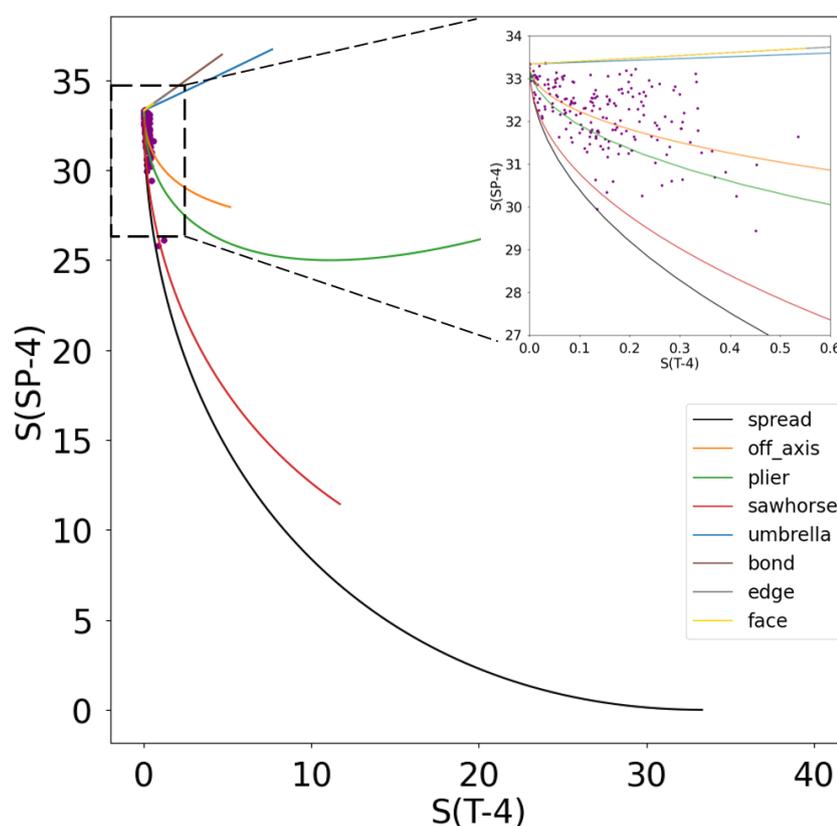


Figure 13. Tetrahedral-square planar shape map showing the principal distortion pathways (solid lines) and the position of the structures collected in Baur's report (purple dots). A zoom-in of the region close to the ideal tetrahedral shape is shown in the inside box for more detail.

The solid lines in figure 13 represent the position in the map of structures along the distortion paths shown in figure 14 as the superposition of perfect tetrahedron (shaded

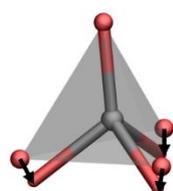
grey polyhedron plus ball representation) with a distorted structure (stick representation). In the umbrella-type distortion three O-P-O angles are simultaneously changed maintaining one of the  $C_3$  axes of the tetrahedron, with the structures along the distortion path having a global  $C_{3v}$  symmetry. For this distortion both  $S(T-4)$  and  $S(SP-4)$  increase with respect to the (0, 33.3) values found for a tetrahedron and structures along this distortion path are found along the straight blue line found in the upper part of the shape map shown in figure 13. In the sawhorse distortion, two vertices of the tetrahedron change to obtain a lower symmetric structure of  $C_{2v}$  symmetry. The distortion path is shown in figure 13 as a red curve where the  $S(T-4)$  increases while and  $S(SP-4)$  decreases. The spread distortion, corresponding to the minimal distortion path in a  $T-4 - SP-4$  shape map, is similar to the sawhorse distortion, but in this case the two pairs of vertices change simultaneously to progressively turn into a square planar structure with  $D_{4h}$  symmetry (black line). In the off-axis distortion just one of the vertices of the tetrahedron is displaced away from the  $C_3$  axis (orange line). In the plier angular distortion all four vertices are displaced simultaneously away from the  $C_3$  axes while maintaining the  $C_2$  symmetry for the whole structure.

The cases of the off-center distortions considered are the movement of the central atom towards one vertex of the  $O_4$  tetrahedron shortening the P-O bond distance, towards an **edge** of the tetrahedron, or towards a **face** of the tetrahedron formed by three oxygen atoms. The three distortions appear in the map as superimposed lines at the top-left corner, above the path for the umbrella distortion.

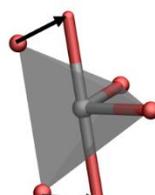
Taking a look at the whole shape map, it is evident that all phosphate structures included in Baur's report are really close to the ideal tetrahedral shape,  $S(T-4) = 0$  and  $S(SP-4) = 33.3$ , with a small spread in the  $S(T-4)$  values (0-0.5) indicating slight to moderate distortions from the tetrahedron. The spread in the  $S(SP-4)$  measures is somewhat larger (29.5-33.5), although most of the structures have  $S(SP-4)$  values smaller than that of the perfect tetrahedron. From the complete shape map it seems evident that there are no structures following the umbrella distortion with increasing  $S(SP-4)$  values, and it seems that the distribution is closer to follow the initial section of the spread path. A closer look into the region of the map around the tetrahedron (see inset in figure 13) shows, however, that this is not the case and the distribution

resembles more a structureless chaotic cloud of points with no representants neither in the upper limiting umbrella-type distortion path nor in the lower limiting spread-type path. The large dispersion of the point cloud reveals that none of the typical distortion paths described for tetrahedral  $ML_4$  coordination compounds<sup>14</sup> is particularly relevant for the description of the distortions found for the  $PO_4$  group in crystal structures which appear to be in quite random directions. The main conclusion that can be extracted from the shape map in figure 13 is that, at least for the structures included in Baur's study, the environment does not produce any specific angular deformation in the phosphate group, independently of the set in which it may be classified from the chemical point of view.

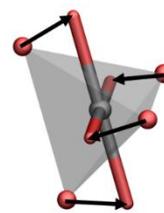
### Angular distortions



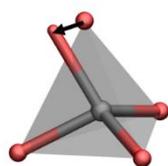
Umbrella



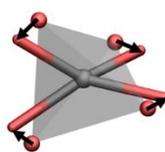
Sawhorse



Spread

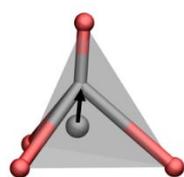


Off-axis



Plier

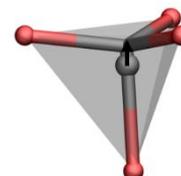
### Off-centre distortions



Bond



Edge



Face

Figure 14. Different shape molecular pathways showing the initial phosphate polyhedron (spherical atoms and grey shaded polyhedron representations) and an arbitrary phosphate structures of the path (stick representation).

Let us remind, however, that from the analysis of the distribution of distortion indices and the joint  $S(T-4, PO_4)/S(T-4, O_4)$  distributions it seems quite evident that the relevant distortions in this case are not the angular ones, but those where the P atom loses its central position while maintaining the  $O_4$  fragment practically tetrahedral. An alternative map to visualize the presence of off-center distortions is to plot the  $S(T-4, O_4)$  measure as a function of the  $S(T-4, PO_4)$  measure. In this map, shown in figure 15, the paths of angular distortions appear as straight lines with different slopes, while the three pure off-center distortions for which the  $O_4$  framework keeps its perfect tetrahedral shape coincide with the  $S(T-4, PO_4)$  axis since  $S(T-4, O_4) = 0$  for all three of them. Looking at the distribution of points calculated for the structures included in Baur's report it is evident that for most structures with small distortions, i.e. with  $S(T-4, PO_4) < 0.4$ , the distortions are neither purely angular nor off-center, with the cloud of points lying more or less in between the  $S(T-4, PO_4)$  axis and the lines corresponding to the pure angular distortions. The low  $S(T-4, O_4)$  values are, however, indicative of a strong off-center component in the overall distortion. For the scarce structures with larger distortions, i.e. with  $S(T-4, PO_4) > 0.4$  the large associated  $S(T-4, O_4)$  values are indicative of a large angular component. Summarizing all this information we may conclude that most of the phosphate groups in Baur's set present small deviations from the ideal tetrahedral symmetry, with a strong off-center component yielding structures where the  $O_4$  tetrahedron remains practically unaltered. For the larger distortions such as those found for organic phosphates or crystal structures where the  $PO_4$  tetrahedra share edges or faces with other  $MO_n$  polyhedra, the principal component are angular distortions to adapt O...O distances to the shapes of the adjacent  $MO_n$  polyhedra.

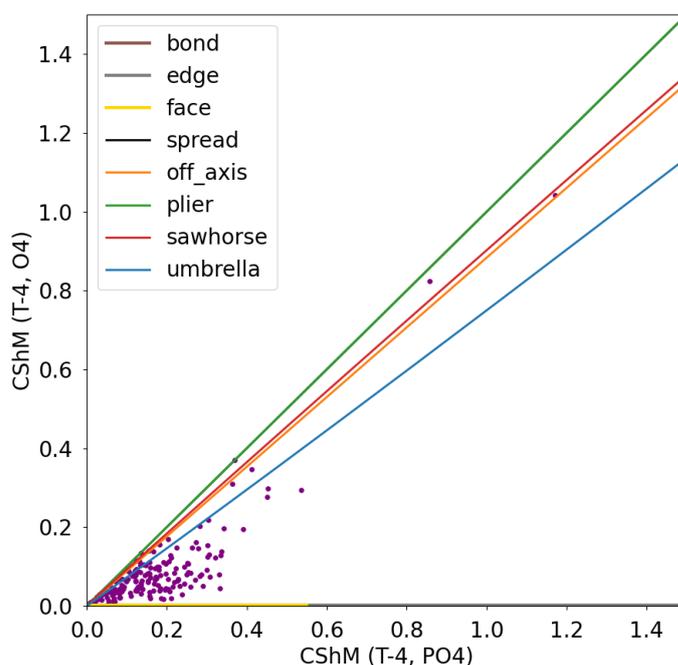


Figure 15.  $S(T-4, PO_4) - S(T-4, O_4)$  shape map showing the paths for angular and off-centre distortions (solid lines) and the structures from Baur's report (purple dots).

Another way to analyse the loss of tetrahedrality is again to compute continuous symmetry measures for subgroups of the  $T_d$  group. For each subgroup, the set of structures was analysed giving a histogram distribution of the continuous symmetry measure (CSM) values in a log-normal distribution. For each subgroup, the CSM values that do not significantly change ( $CSM < 0.1$ ) are the majority of the 196 Baur structures (table 1).

		%CSM < 0.1						
		All	Proper	Acid	Organic	Di-	Ring	Poly-
$C_3$		66	89	80	39	91	19	28
$C_2$		96	97	100	89	97	100	89
$S_4$		57	89	93	36	51	4	0
$C_s$		100	100	100	100	100	100	100

Table 1. Number of symmetric structures (in %) without a significant change in their CSM.

As expected, the greater distortions occur in higher symmetric groups,  $C_3$  and  $S_4$  in this case, where values lower than 0.1 are the 66% and 57% of the total structures, meaning that around a 40% of structures have a significant distortion in their symmetry. The other two cases, the low symmetry groups  $C_2$  and  $C_s$ , have most of the values between 0 and 0.1. Of the high distorted symmetries, the groups that distort more the  $C_3$  and  $S_4$  symmetry are the organic, ring and polyphosphates groups.

If these values are compared with the CSM values for the Monte Carlo simulations at room temperature, the greatest change is observed in the acid and diphosphate groups. For  $C_2$  and  $S_4$  symmetries, the Monte Carlo structures does not have as many structures with a CSM > 0.1 than in the Baur's experimental structures, another evidence of the influence of the crystal environment present in the crystal structures (table 2).

	<i>Baur's</i>			<i>Monte Carlo</i>		
	<i>Proper</i>	<i>Acid</i>	<i>Di-</i>	<i>Proper</i>	<i>Acid</i>	<i>Di-</i>
$C_3$	89	80	91	51	22	6
$C_2$	97	100	97	98	0	5
$S_4$	89	93	51	87	0	0
$C_s$	100	100	100	100	100	100

Table 2. Number of symmetric structures (in %) for each  $T_d$  subgroup in Baur's and Monte Carlo structures.

## Effect of the temperature and the environment on the symmetry of phosphate anions: Molecular dynamics simulations for crystal structures

After analysing the effect of temperature on the shape of isolated phosphate anions in the gas-phase and the effect of the crystal environment by looking at the crystal structures studied by Baur, it is pertinent to address the combined effects of temperature and crystal environment on the shape of phosphate anions. For this

purpose, a few molecular dynamics (MD) simulations were run in the NVT ensemble for the range of temperatures analysed before (0-500K). Three crystals containing the same representative phosphate anions chosen for the Monte Carlo simulations were analysed: calcium phosphate with isolated  $\text{PO}_4^{3-}$  anions, hydrogen calcium phosphate with isolated hydrogen phosphate  $\text{HPO}_4^{2-}$  anions, and calcium diphosphate with isolated  $\text{P}_2\text{O}_7^{2-}$  anions. Since the computational cost of these simulations at the *ab initio* level would be highly prohibitive, the MD simulations in this section were based on force fields, using the parameters fitted by Demichelis et al.<sup>15</sup>, in the two former cases, and those of Taylor and Simkiss<sup>16</sup> for the latter one. Initial structures were taken from earlier crystallographic studies:  $\alpha$ -calcium phosphate ( $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ) from Matthew et al.<sup>17</sup>, monetite ( $\text{CaHPO}_4$ ) from Catti et al.<sup>18</sup>, and  $\beta$ -calcium diphosphate ( $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ ) from Webb<sup>19</sup> (figure 16). The structures were first optimized using the GULP program<sup>20</sup> giving the optimized cell parameters gathered in table 3, where we see that the optimized cell parameters are satisfactorily close to those obtained by x-ray diffraction.

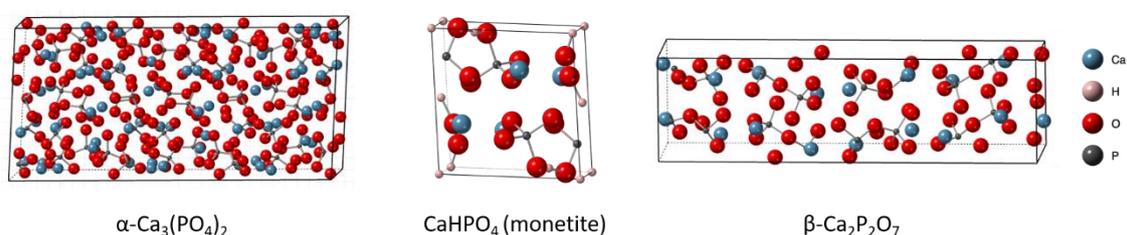


Figure 16. Cell crystal structures of  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ,  $\text{CaHPO}_4$  and  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ .

	$\alpha\text{-Ca}_3(\text{PO}_4)_2$		$\text{CaHPO}_4$ (monetite)		$\beta\text{-Ca}_2\text{P}_2\text{O}_7$	
	Exp.	Opt.	Exp.	Opt.	Exp.	Opt.
a (Å)	12.887	12.891	6.616	7.051	6.684	6.643
b (Å)	27.280	27.731	6.619	6.656	6.684	6.643
c (Å)	15.219	15.373	6.946	6.943	24.144	23.931
$\alpha$ (°)	90.00	90.00	96.18	93.98	90.00	90.00
$\beta$ (°)	126.20	126.81	103.82	104.18	90.00	90.00
$\alpha$ (°)	90.00	90.00	88.34	89.66	90.00	90.00

Table 3. Cell crystal experimental parameters and optimized parameters for  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ,  $\text{CaHPO}_4$  and  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ .

Comparing the experimental and optimized structures, the CShM found for the tetrahedral  $\text{PO}_4$  groups are slightly different in all cases, but with no relevant divergences. For example, in the case of  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ , there are five non-equivalent phosphate anions with CShM values of 0.03, 0.07, 0.11, 0.12 and 0.24 that change to 0.07, 0.09, 0.10, 0.13 and 0.20, respectively in the optimized structure. Note that both the magnitude and the ordering of the different CShM values are the same in the experimental and the optimized crystal structures (table 4). For the diphosphate anion there are a few tetrahedral phosphates that decrease their CShM values, however, the decrease is small in all cases except in one case, but the simulations were carried out with this unexpected variation as the optimized parameters were still in the order of the experimental ones.

	experimental	optimized
$\alpha\text{-Ca}_3(\text{PO}_4)_2$	0.03	0.07
	0.07	0.09
	0.11	0.10
	0.12	0.13
	0.24	0.20
$\text{CaHPO}_4$	0.07	0.21
	0.07	0.24
	0.24	0.38
	0.24	0.54
$\beta\text{-Ca}_2\text{P}_2\text{O}_7$	0.07	0.02
	0.14	0.04
	0.20	0.04
	0.20	0.11
	0.25	0.16
	0.65	0.16

Table 4. Cell crystal experimental and optimized shape measures for the different phosphate anions in  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ,  $\text{CaHPO}_4$  and  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ .

For each of the optimized structures several molecular dynamics calculations were run at different temperatures in the range 10-500K taking single cells with fixed cell parameters containing 48, 4 and 16 independent phosphate tetrahedra for  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ,  $\text{CaHPO}_4$  and  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ , respectively. All MD simulations were run for 40 ps with an initial 10 ps equilibration period with a timestep of 0.001 ps. Snapshots for each simulation were taken every 0.05 ps, resulting in a total of 800 geometries after the equilibration time of which only the last 200 steps were taken to be analysed as the pressure was found almost constant in the NVT ensemble.

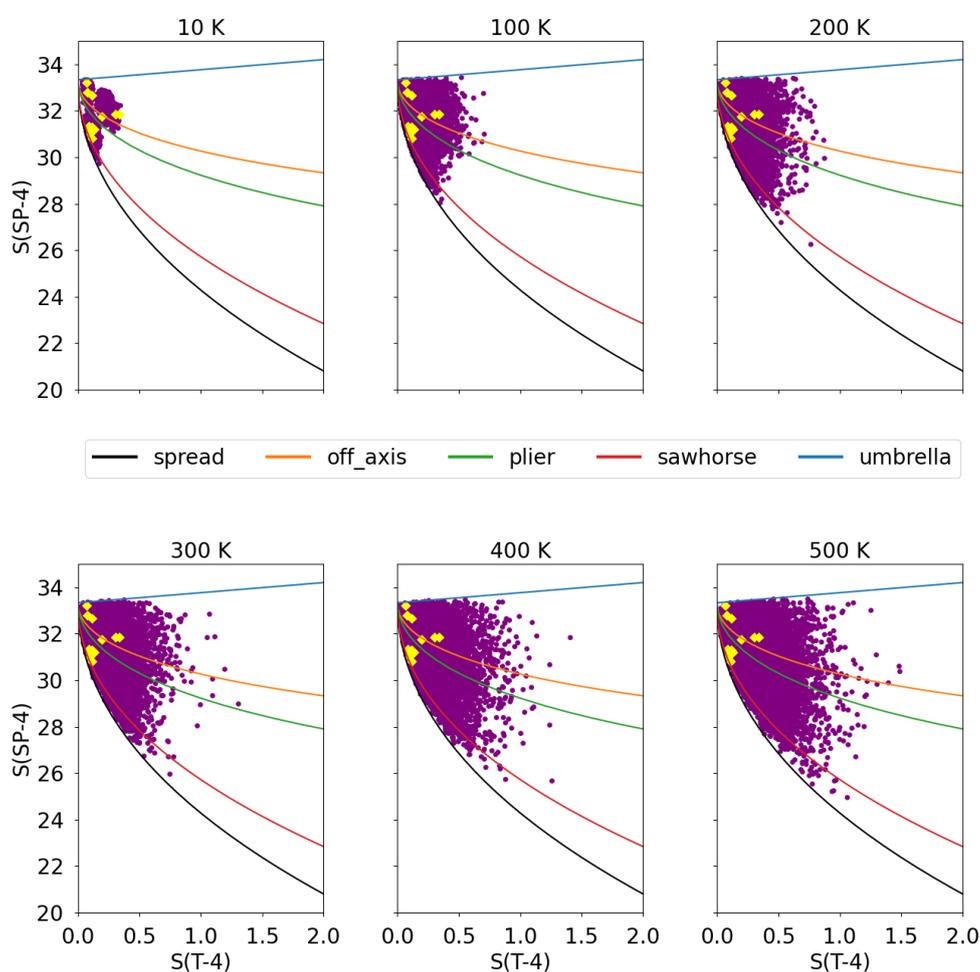


Figure 17.  $S(\text{T-4}) - S(\text{SP-4})$  shape map showing the distribution of shape measures for the phosphate anions in  $\alpha\text{-Ca}_3(\text{PO}_4)_2$  for each simulation at different temperatures (purple dots). The CShM values for the initial crystal structure are shown as yellow dots. Solid lines indicate the principal distortion paths for a tetrahedral  $\text{ML}_4$  structure.

The distribution for these structures in a shape map gives an idea of how the shape of phosphate anions is affected by the temperature. The shape maps for the three analysed crystals (figures 17 and 1,3A in the appendix) show that in all three cases the molecular distortions induced by temperature do not follow any preferred path. Comparing the results in the MD simulation for the crystal structures with those obtained for the free phosphates in the Monte Carlo simulations (figure 18 and 2,4A in the appendix) we can appreciate similar trends for all temperatures, although the area occupied by the point cloud clearly shows that the MD simulations for the crystal structure tend to explore more possible distortions at low temperatures. This result is, however not a surprise, since there are already five different phosphate anions in the crystal structure which are already distorted at very low temperatures and the temperature induced distortions for each of them can cover a larger area in the map even at low temperatures. At high temperatures, however, there is sufficient energy available for the simulations to explore similar larger distortions in both cases.

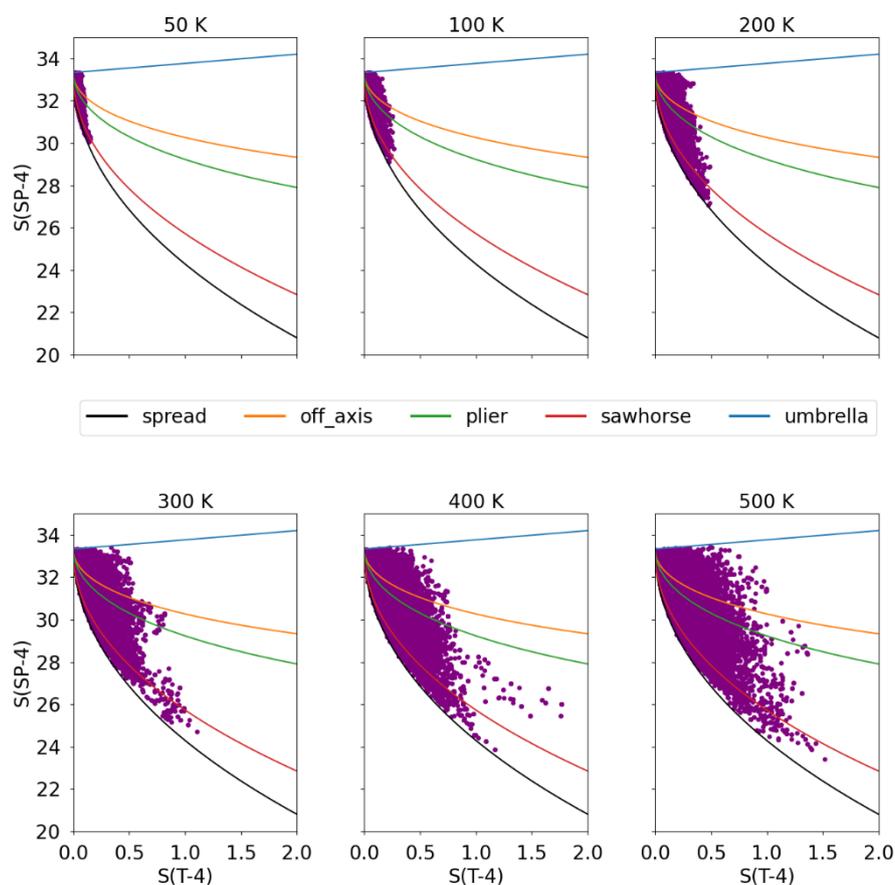


Figure 18.  $S(T-4) - S(SP-4)$  shape map showing the distribution of shape measures for the isolated phosphate anion obtained from MC simulations at different temperatures (purple dots). Solid lines indicate the principal distortion paths for a tetrahedral  $ML_4$  structure.

If the distributions of CShM values at different temperatures are compared in a box plot, the structures from the MD simulations for the crystals present larger distortion values at each temperature (figure 19). However, both in the simulations for the isolated anion and the whole crystal structure there are no significant differences between the measures for the whole  $PO_4$  tetrahedron and the  $O_4$  fragment, indicating that in the case of proper phosphates, the off-center distortions does not seem to have a significant contribution to the overall distortion of its tetrahedral shape.

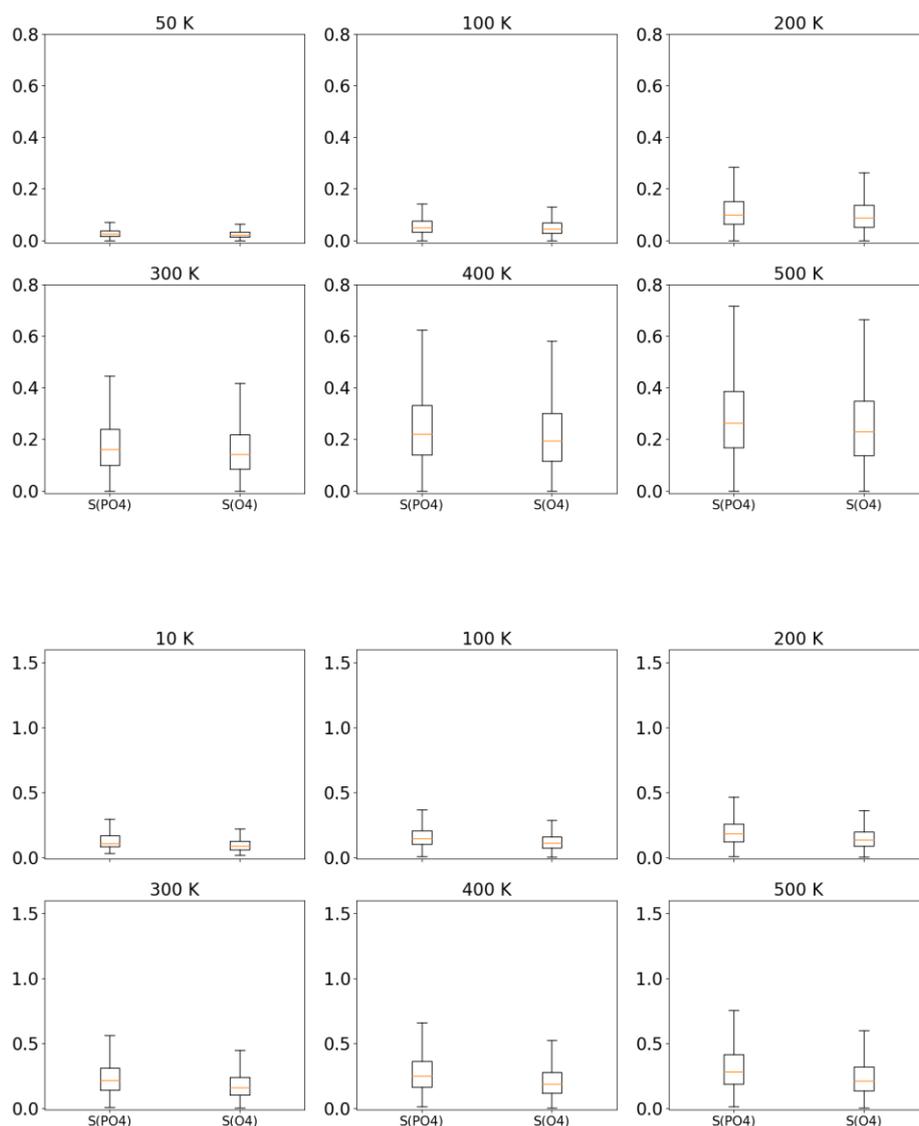


Figure 19. Phosphate anion CShM mean and dispersion values at different temperatures for the Monte Carlo simulations (top) and Molecular dynamics simulations (bottom).

In the case of the acid phosphates, the CShMs for the anions in the  $\text{CaHPO}_4$  crystal structure present an interesting difference at low temperatures. If we compare the distributions for the  $S(\text{T}_{-4}, \text{PO}_4)$  and  $S(\text{T}_{-4}, \text{O}_4)$  measures in the simulations for the crystal and those for the free hydrogen phosphate anion (figure 20) we detect a striking difference. While in the simulations for the crystal structure the distribution of the measures for the  $\text{PO}_4$  and  $\text{O}_4$  tetrahedra are quite similar, with the average  $S(\text{T}_{-4}, \text{O}_4)$  values being slightly larger than the  $S(\text{T}_{-4}, \text{PO}_4)$  ones, the two distributions are clearly

different for the isolated atom where at low temperatures the  $O_4$  fragment remains practically tetrahedral. This comparison shows that while for the isolated anion off-center distortions dominate at low temperatures, in the case of the crystal, the effect of the asymmetric environment is to distort the  $O_4$  tetrahedron already for the minimum energy structure. In this case the effect of temperature is just to broaden the distribution of CShM values, but with similar results both for the  $PO_4$  and the  $O_4$  fragments.

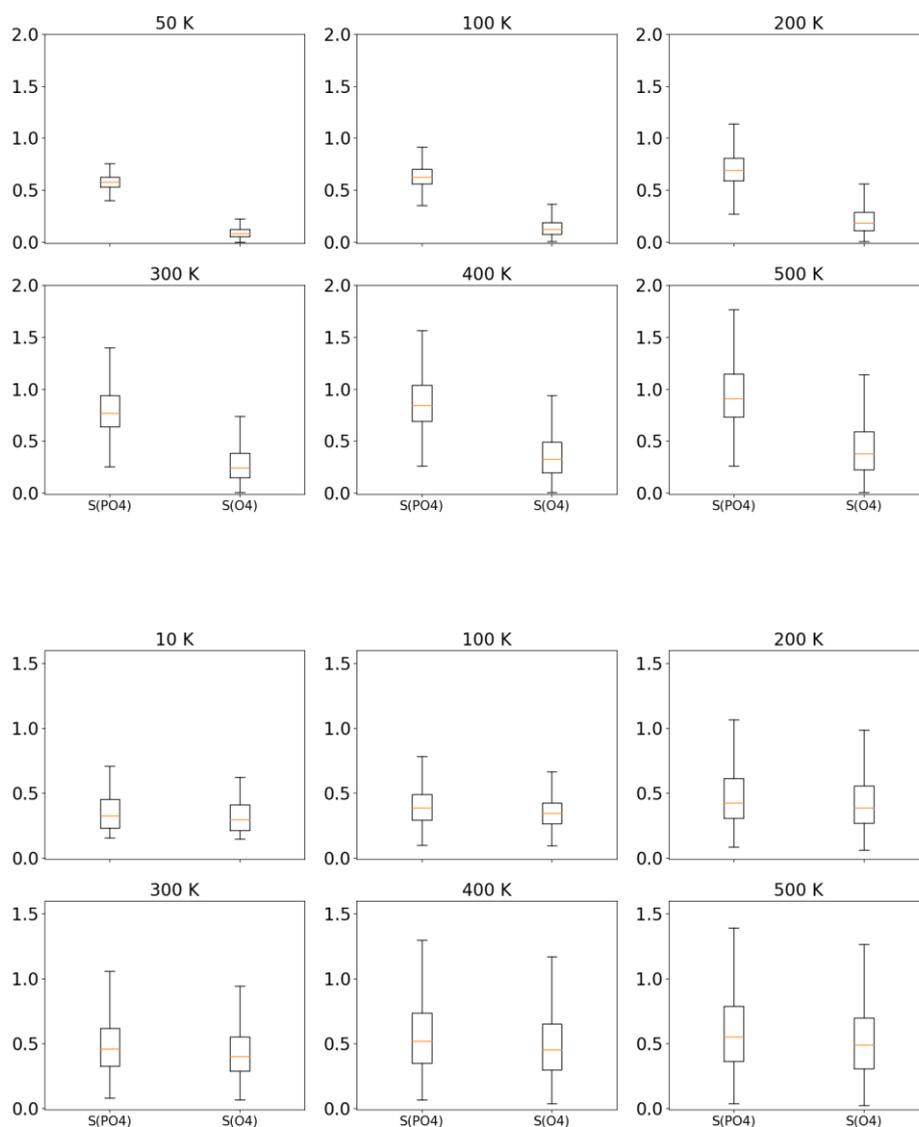


Figure 20. Hydrogen phosphate CShM mean and dispersion values at different temperatures for the Monte Carlo simulations (top) and Molecular dynamics simulations (bottom)

Finally, for the case of  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$  (figure 21) we find trends similar to the ones in the phosphate case. In this case it is interesting to see that the crystal environment reduces the increasing spread in the CShM values with temperature with much wider distributions in the simulations for the isolated diphosphate anion.

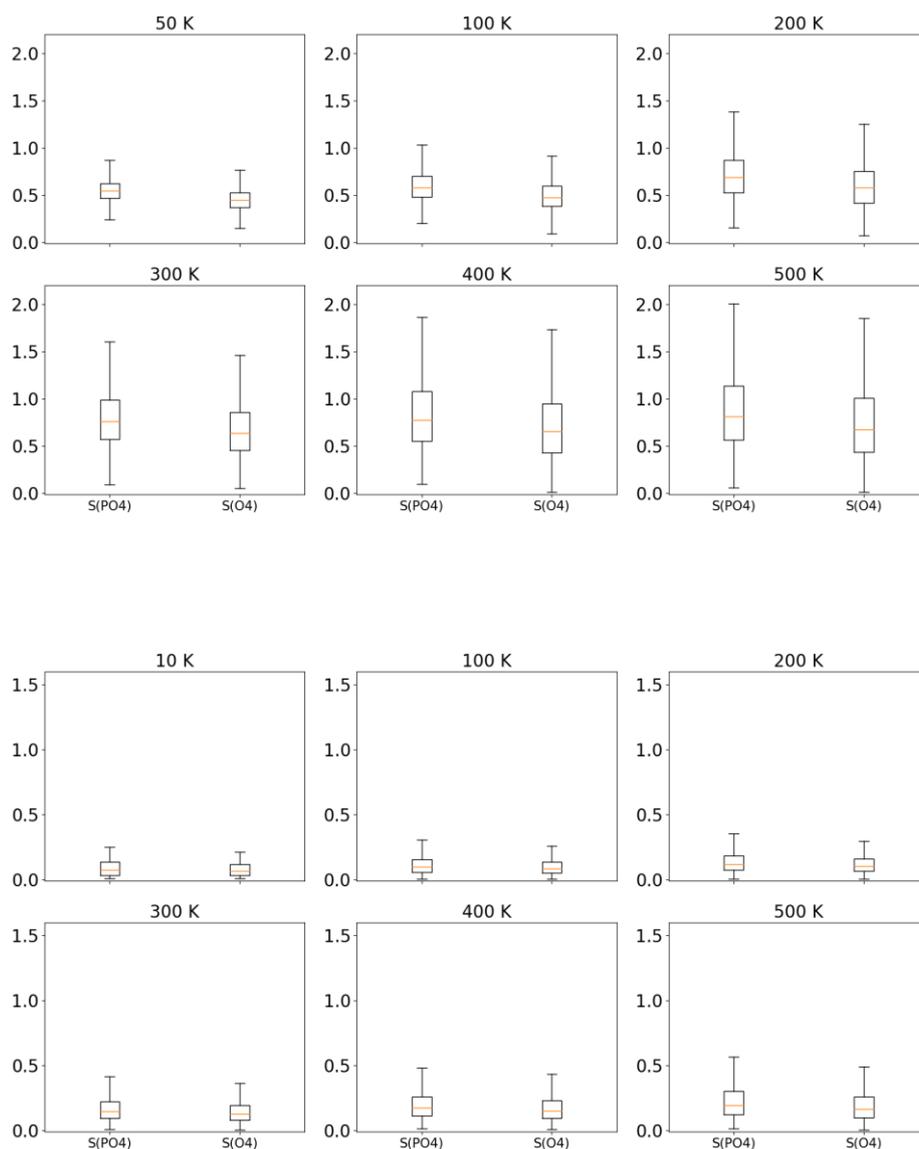


Figure 21. Diphasphate CShM mean and dispersion values at different temperatures for the Monte Carlo simulations (top) and Molecular dynamics simulations (bottom).

The CShMs values for the  $\text{PO}_4$  groups in the three structures still follows a log-normal distribution for each temperature. The average values at each temperature show a linear dependence of the CShM with the temperature (figure 22). If the mean values are inspected, the most probable tetrahedral shapes for any temperature are found in the range of the experimental crystal structures of Baur.  $\alpha\text{-Ca}_3(\text{PO}_4)_2$  gives CShM expected values in the range of 0.1 to 0.3,  $\text{CaHPO}_4$  in a range of 0.3-0.6 and  $\alpha\text{-Ca}_2\text{P}_2\text{O}_7$  values lower than 0.2, similar to the distributions as seen in figure 9. The slopes, however, are a bit different from those obtained in the MC results. In these three crystal structures, the slopes are  $4\text{e-}04$ ,  $5\text{e-}05$  and  $3\text{e-}04$  for the  $\alpha\text{-Ca}_3(\text{PO}_4)_2$ ,  $\text{CaHPO}_4$  and  $\beta\text{-Ca}_2\text{P}_2\text{O}_7$ , respectively. In the MC simulations, however, the lower slope is found for the phosphate anion ( $6\text{e-}04$ ), but the larger is still the hydrogen phosphate structure ( $9\text{e-}04$ ).

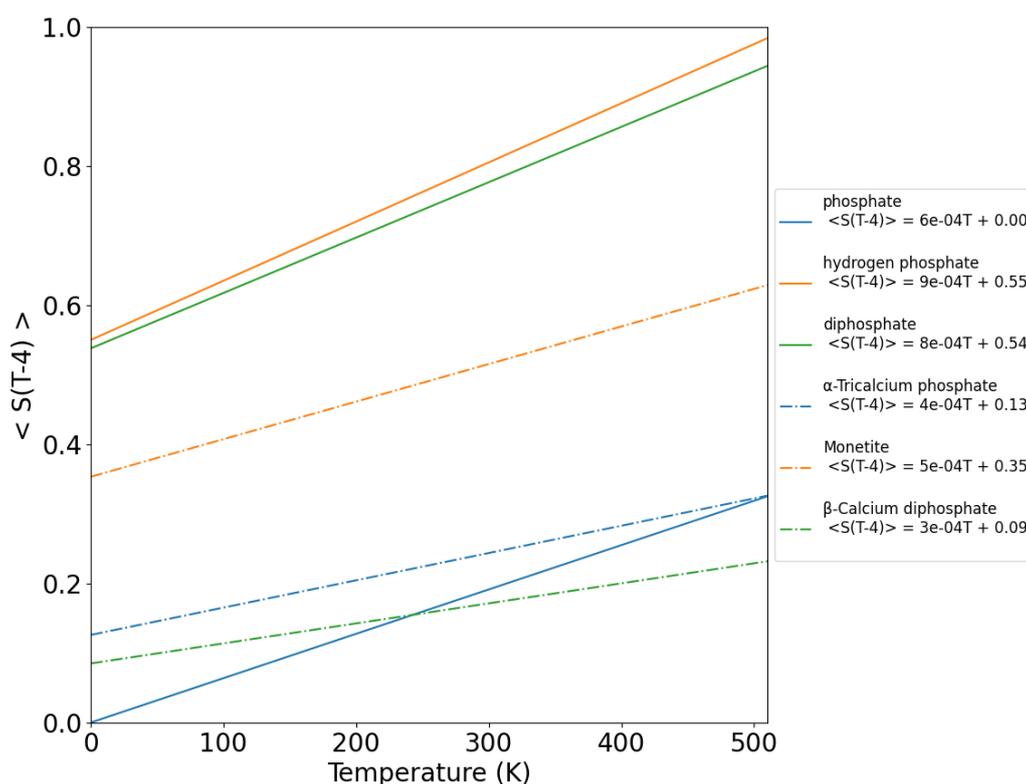


Figure 22. Linear regression fittings for the three phosphate Monte Carlo structures (solid line) and the phosphate crystals analysed (dash-dot lines).

If the deviation of these values is analysed, the dispersion values obtained in the Monte Carlo simulations present higher values than in the crystal structures case, meaning that the crystal environment affects drastically the possible distortions that can occur due to temperature (figure 23).

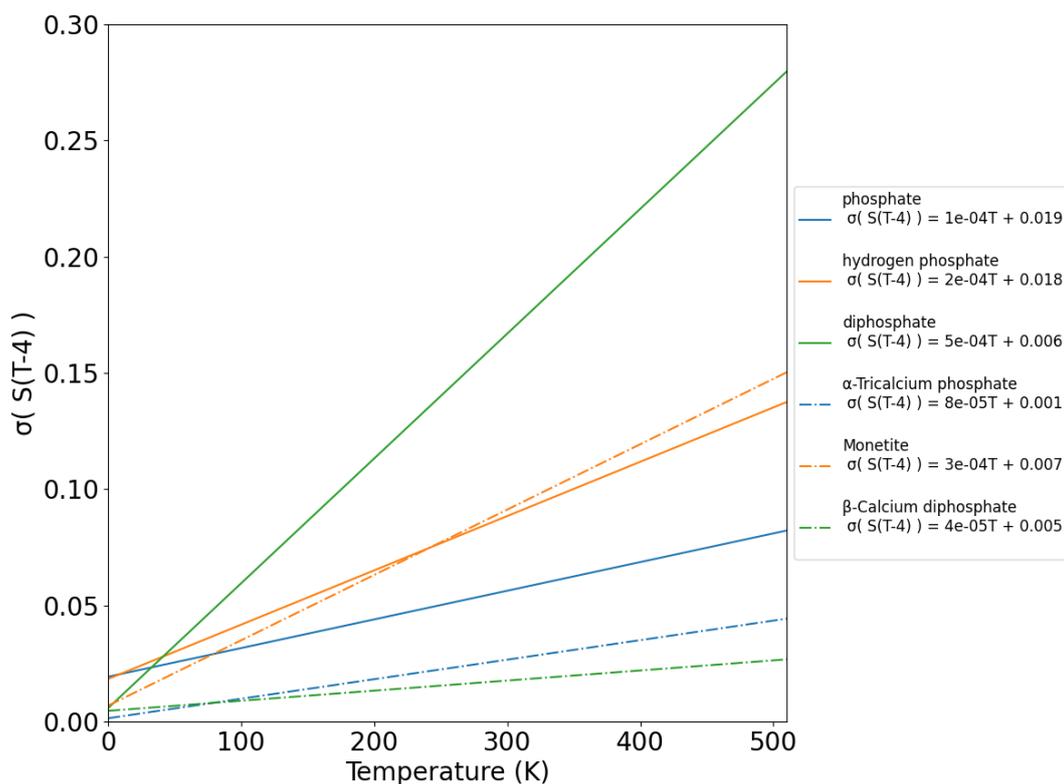


Figure 23. Linear regression fittings of the deviation of the CShM values for the three Monte Carlo structures (solid line) and the phosphate crystals analysed (dash-dot lines).

## Pseudosymmetry analysis

After analysing the shape of the different phosphate structures in purely geometrical terms, our goal is to determine if these distortions are relevant from a chemical point of view. For this purpose, we will test how the observed distortions affect the symmetry of the molecular orbitals by applying a pseudosymmetry analysis (see chapter 1). Since

our interest is just to give a qualitative insight, the molecular orbitals used in our analysis were calculated at the extended-Hückel level<sup>21</sup>, using the *huckelpy*<sup>22</sup> library, a python module developed by us that can be seamlessly connected with *CosymLib* to perform the pseudosymmetry analysis.

The interaction diagram for the  $\text{PO}_4$  fragment is shown in figure 24. For simplicity, the four lower energetic molecular orbitals, corresponding to the 2s orbitals of the O atoms, are omitted. The four s, p orbitals of the phosphorus atom mix with those of the  $\text{O}_4$  fragment, giving rise to the four  $A_1 + T_2$  P-O bonding orbitals and their respective antibonding combinations, with the later representing the unoccupied orbitals of the molecule. The four non-bonding orbitals corresponding to the lone pairs on the oxygen atoms form another quasi-degenerated set of  $A_1 + T_2$  symmetry.

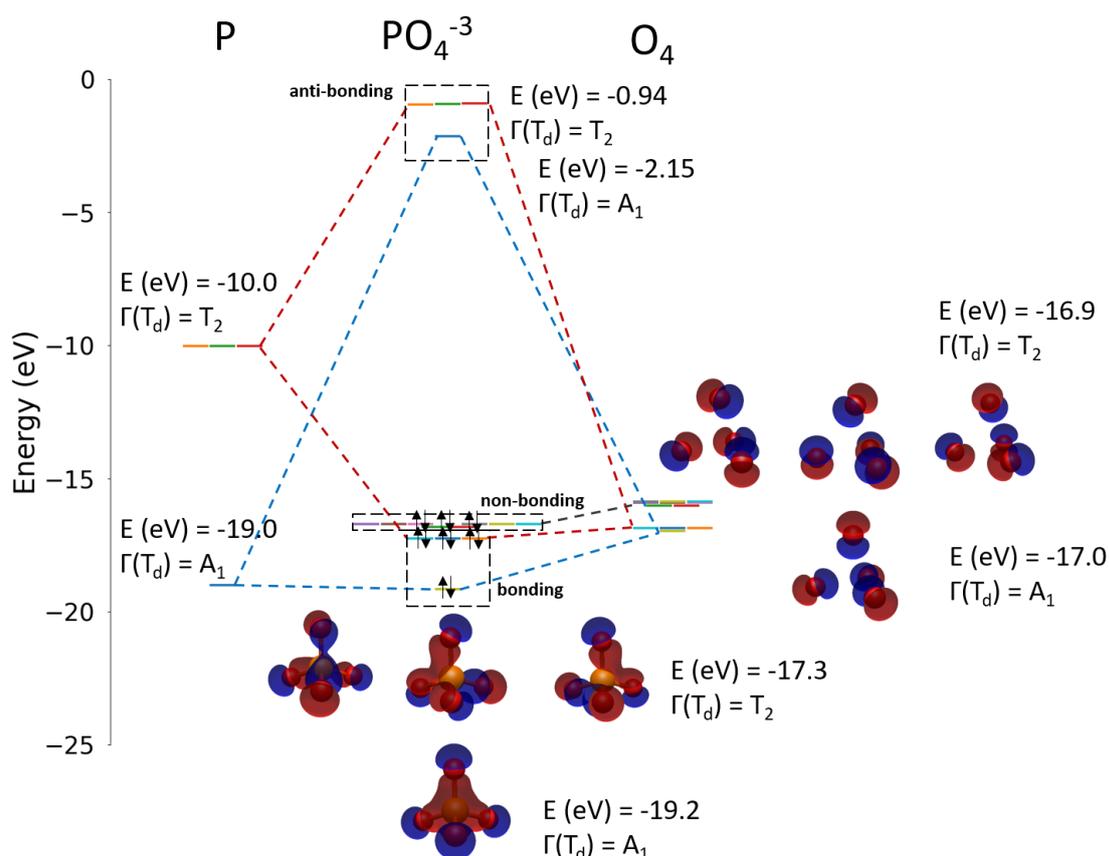


Figure 24. Molecular orbital diagram for the  $\text{PO}_4^{3-}$  ion. The molecular orbitals of the  $\text{O}_4$  fragment that interact with the AOs of the phosphorous atom and the resulting  $\text{PO}_4$  molecular orbitals are shown, as well as their energies and irreducible representations for  $T_d$

symmetry. The non-bonding molecular orbitals of  $\text{PO}_4^{3-}$  molecule were slightly destabilized in the  $\text{O}_4$  fragment for a better visualization.

What happens to the symmetry of the four orbitals involved in the four P-O bonds upon one of the experimentally found distortions? To have an idea on how these distortions may affect the symmetry features of the four P-O bonding orbitals we will focus our attention on the structure with the largest deviations from the tetrahedral symmetry found in the crystal structures studied by Baur which belongs to the proper group (ICSD code : 22048).

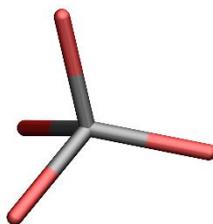


Figure 25. Higher phosphate distorted anion of Baur structures.

In this case, the angular distortions play an important role, with the bottom face of the phosphate anion in figure 25, formed by the three oxygen atoms, resulting in highly distorted angles of 94.7, 111.7 and 125.0 degrees. This angular distortion is confirmed by analysing the CShM values for the whole phosphate group and for the  $\text{O}_4$  fragment,  $S(\text{PO}_4) = 1.17$  and  $S(\text{O}_4) = 1.04$ , where the off-center distortion does not seem to play a significant role in this case. The molecular orbitals with their corresponding pseudosymmetry analysis using the  $T_d$  group as a reference are represented in figure 26, where the orientation of the principal  $C_3$  axis was along the best P-O bond that minimize the density measure of  $T_d$  group and the second axis was chosen to coincide with another P-O bond. As shown in the figure, in this case, the set orbitals on the  $\text{O}_4$

fragment which is no longer perfectly tetrahedral, retain mainly their original  $T_2$  character, although a mixture with functions belonging to other irreducible representations of the  $T_d$  group in some of the orbitals may be as high as 33%. However, when we centre our attention on the symmetry of the four P-O bonding orbitals of the  $PO_4$  anion, we find a ~90%  $A_1$  or  $T_2$  character for them, indicating that even in the case with the largest deviations from the ideal tetrahedral symmetry, properties associated to the electronic structure will behave in a practically identical way as in an undistorted, perfectly tetrahedral anion. Note however, that for the virtual P-O antibonding set the asymmetry results in a large splitting in the formally  $T_2$  set, even if these orbitals maintain their  $T_2$  character at a large extent. These empty orbitals have, however no consequences in the observable physical properties such as the electron density which determines the shape of the anion.

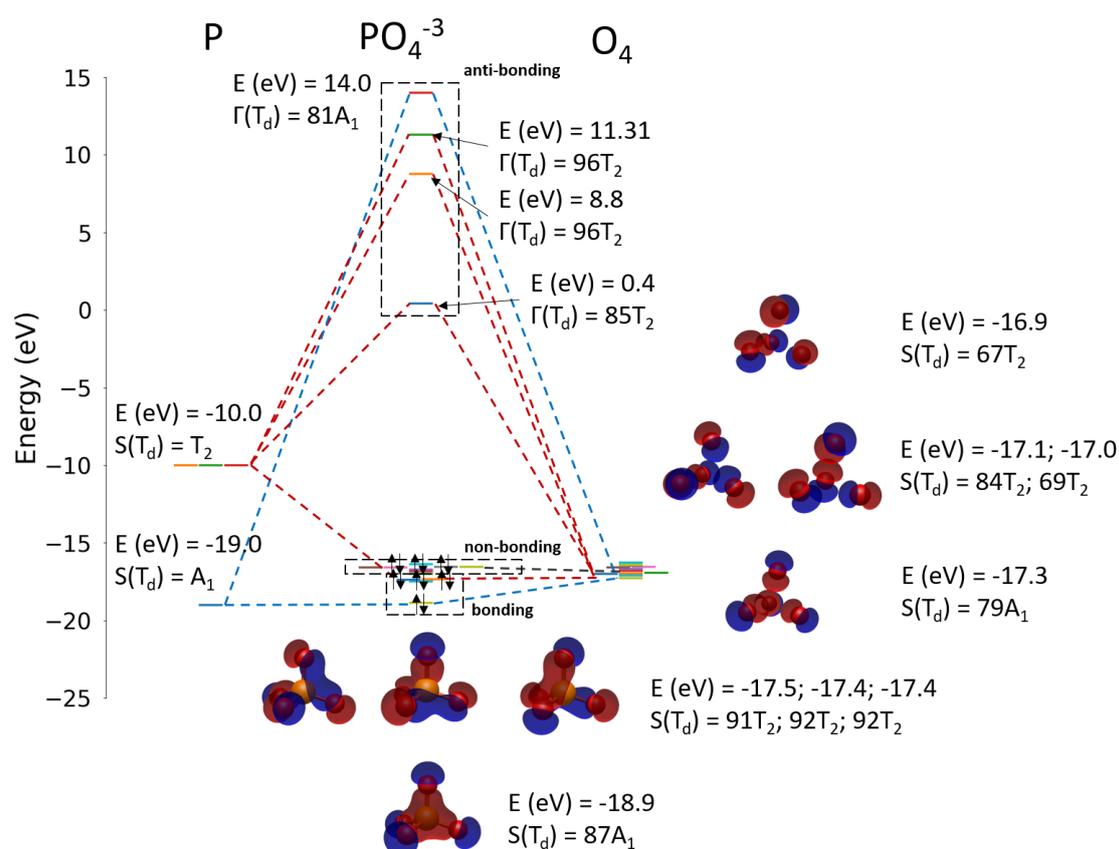


Figure 26. Molecular orbital diagram for a distorted phosphate anion tetrahedron.

A similar analysis can be performed for structures distorted according to the different distortion paths analysed in previous sections. In order to obtain comparable features, we have taken structures along each different distortion path with CShM of  $S(T-4) \sim 0.5$ . As shown in figure 27, the  $A_1$  and  $T_2$ -type P-O bonding molecular orbitals of the perfect phosphate anion tetrahedron retain their symmetry character up to a 90% in all cases, with the only consequence being a small splitting of the degenerated  $T_2$  set. The case of the non-bonding orbitals is, however, different, as the  $3T_1$  and  $3T_2$  orbitals of the ideal phosphate anion are distorted in a way that the six orbitals have a strong mixture between both irreducible representations. This mixture is explained by the type of distortion that this molecule suffers, which is an angular distortion, that modifies mainly the oxygen atoms which these orbitals depends on. In general, we can conclude that the symmetry of the four molecular orbitals involved in the P-O bonds is practically unaffected by any distortion with a geometrical shape measure close to 0.5, such as those found for almost all the crystal structures included in Baur's study, and for this reason we may conclude that although no phosphate anions with a perfect symmetry may be found in actual crystal structures, the observed deviations are, however, practically irrelevant from the chemical point of view and the  $PO_4$  anions may be safely considered to be tetrahedral for most purposes.



Figure 27. Irreducible representation for the  $T_d$  pseudosymmetry for the P-O bonding orbitals of the phosphate anion for structures with  $S(T_4, PO_4) \sim 0.5$  along each of the different paths analysed in the present chapter.

## Conclusions

Phosphate anions, even though they do not belong to the  $-43m$  ( $T_d$ ) crystallographic symmetry as stated by Baur, do present a high degree of local tetrahedral symmetry, as the continuous shape measures for their structure remain practically unaffected, giving values that are very close to those for the ideal  $PO_4$  tetrahedron.

Additionally, the molecular environment around these anions does not distort the  $PO_4$  tetrahedra in any preferred direction, reinforcing the idea that the observed deformations are equally caused by temperature and condensation effects. However, a preference for the phosphorus off-center distortion over other angular distortions has been observed in all of Baur's groups of phosphates.

However, the MC and MD simulations exhibit that the crystal environment plays an important role by difficult the phosphate anions distortion as the increase of the temperature does not affect their tetrahedrality as much as in the gas phase simulations. The dispersion values found for the CShM tetrahedral measures show that the gas phase simulations gave more flexibility to the phosphate anions to explore more distorted structures than in the crystal structure.

Finally, the pseudosymmetry analysis for the molecular orbitals shows that the largest distortions observed in the crystal structures are not able to induce any relevant symmetry loss in the molecular orbitals, confirming that for most chemical purposes, the  $PO_4$  group may be considered to have a tetrahedral structure.

## Appendix

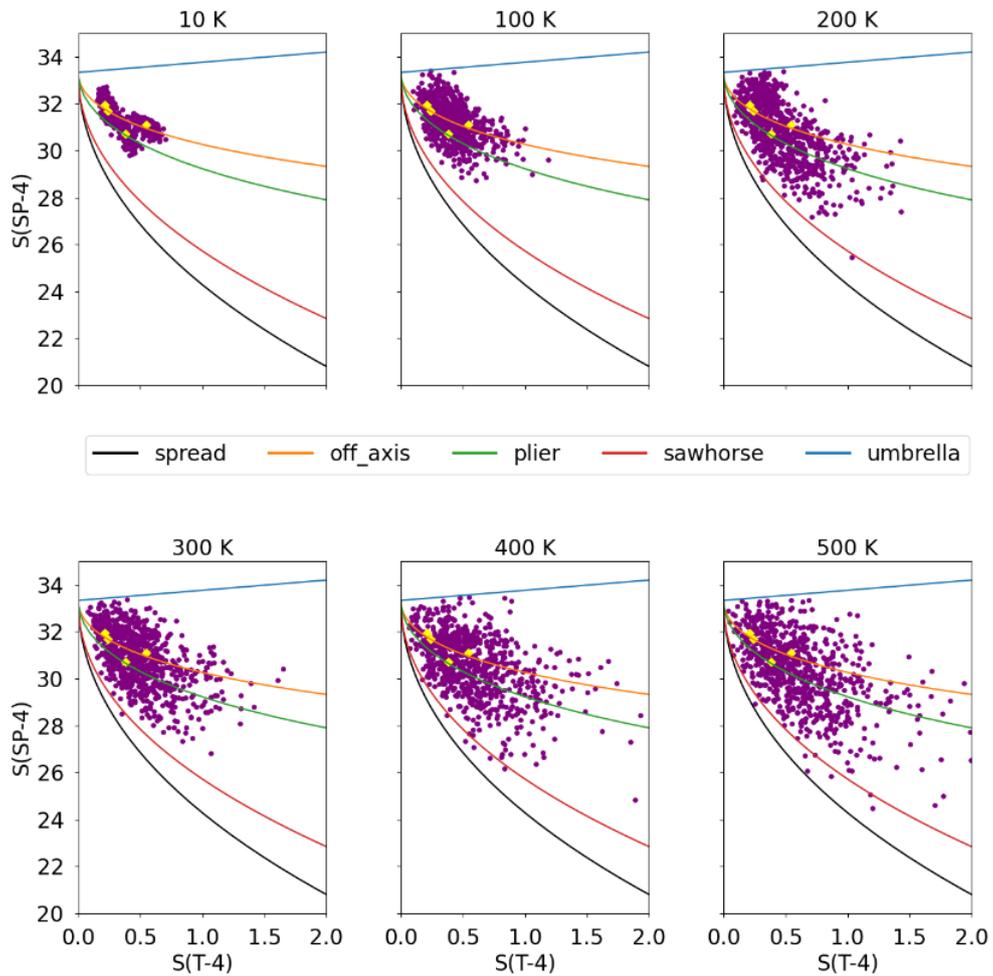


Figure 1A.  $\text{CaHPO}_4$  continuous shape measures for each temperature (purple dots) for  $S(\text{T-4})$  and  $S(\text{SP-4})$  reference shapes from the GULP simulation.

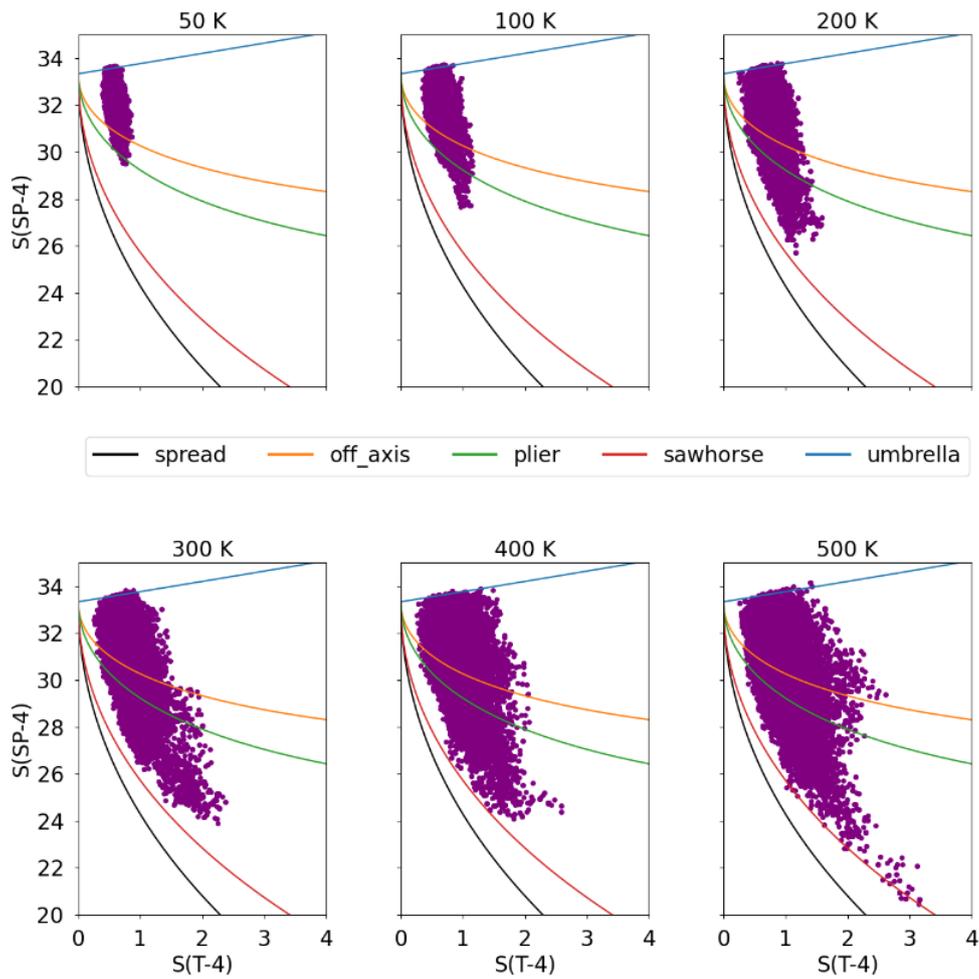


Figure 2A.  $\text{CaHPO}_4$  continuous shape measures for each temperature (purple dots) for  $S(\text{T-4})$  and  $S(\text{SP-4})$  reference shapes from the Monte Carlo simulations.

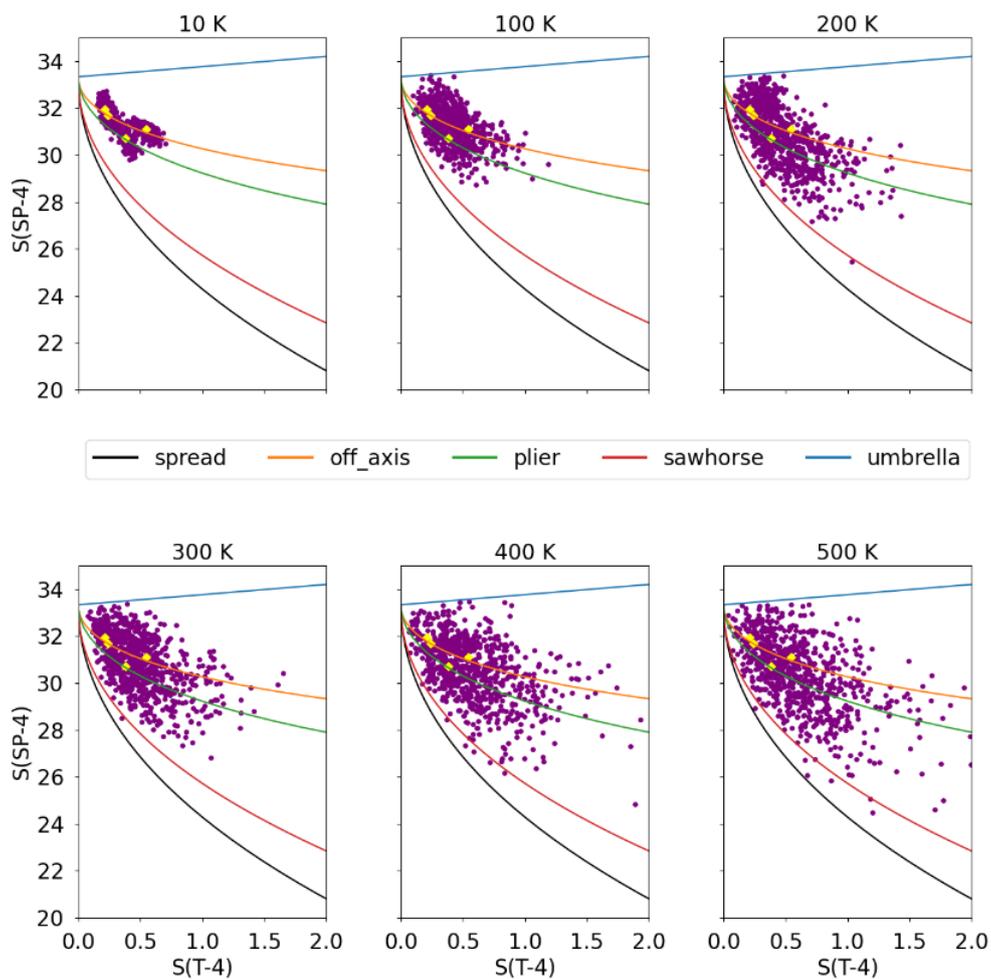


Figure 3A.  $\alpha\text{-Ca}_2\text{P}_2\text{O}_7$  continuous shape measures for each temperature (purple dots) for  $S(\text{T-4})$  and  $S(\text{SP-4})$  reference shapes from the GULP simulation.

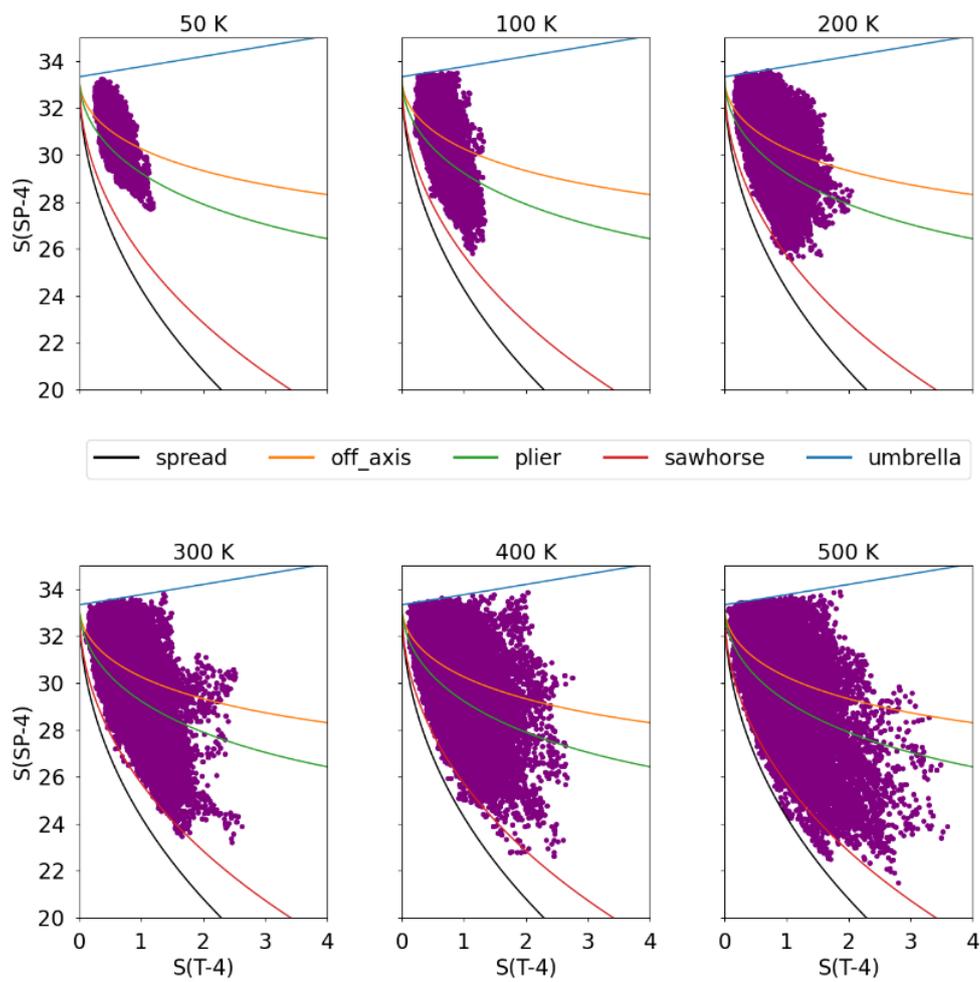


Figure 4A.  $\alpha\text{-Ca}_2\text{P}_2\text{O}_7$  continuous shape measures for each temperature (purple dots) for  $S(\text{T-4})$  and  $S(\text{SP-4})$  reference shapes from Monte Carlo simulations.

## Bibliography

1. Hoffmann, R., Alder, R. W. & Wilcox Jr., C. F. Planar tetracoordinate carbon. **92**, 4992–4993 (1970).
2. Rottger, D. & Erker, G. Compounds Containing Planar-Tetracoordinate Carbon. *Angew. Chemie - Int. Ed.* **36**, 812–827 (1997).
3. Boudhar, A., Charpenay, M., Blond, G. & Suffert, J. Fenestranes in synthesis: Unique and highly inspiring scaffolds. *Angew. Chemie - Int. Ed.* **52**, 12786–12798 (2013).
4. Muetterties, E. L. & Guggenberger, L. J. Idealized Polytopal Forms . Description of Real Molecules Referenced to Idealized Polygons or Polyhedra in Geometric Reaction Path Form. **3021**, 1748–1756 (1974).
5. Gillespie, R. J. Improving our understanding of molecular geometry and the VSEPR model through the ligand close-packing model and the analysis of electron density distributions. **197**, 51–69 (2000).
6. Baur, W. H. The geometry of polyhedral distortions. Predictive relationships for the phosphate group. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **30**, 1195–1215 (1974).
7. Echeverría, J., Casanova, D., Lluell, M., Alemany, P. & Alvarez, S. Molecules and crystals with both icosahedral and cubic symmetry. *Chem. Commun.* 2717–2725 (2008) doi:10.1039/b719615f.
8. MonteModes: <https://github.com/abelcarreras/MonteModes>.
9. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ort, and D. J. F. Gaussian 09, Revision D.01. (2016).
10. Tuvi-Arad, I., Rozgonyi, T. & Stirling, A. Effect of temperature and substitution on cope rearrangement: A symmetry perspective. *J. Phys. Chem. A* **117**, 12726–12733 (2013).
11. Bergerhoff, G., Brown, I. & Allen, F. Crystallographic databases. *Int. Union Crystallogr.* (1987).
12. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
13. Huminicki, D. M. C. & Hawthorne, F. C. The crystal chemistry of the phosphate minerals. *Phosphates Geochemical, Geobiol. Mater. Importance* **48**, 123–254 (2019).
14. Cirera, J., Alemany, P. & Alvarez, S. Mapping the Stereochemistry and Symmetry

- of Tetracoordinate Transition-Metal Complexes. *Chem. - A Eur. J.* **10**, 190–207 (2004).
15. Demichelis, R. *et al.* Simulation of Calcium Phosphate Species in Aqueous Solution: ForceField Derivation. *J. Phys. Chem. B* **122**, 1471–1483 (2018).
  16. Taylor, M. G., Simkiss, K., Rg, U. K. & Leslie, M. Computer Modelling of Phosphate Biominerals. **90**, 641–647 (1994).
  17. Mathew, M., Schroeder, L. W., Dickens, B. & Bawn, W. E. The Crystal Structure of  $\alpha$ -Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub>\*. *Acta Cryst* **B33**, 1325–1333 (1977).
  18. Catti, M. & Ferraris, G. Low-Temperature Ordering of Hydrogen Atoms in CaHPO<sub>4</sub> (Monetite): X-ray and Neutron Diffraction Study at 145 K\*. *Acta Cryst* **B36**, 254–259 (1980).
  19. Webb, N. C. The crystal structure of  $\beta$ -Ca<sub>2</sub>P<sub>2</sub>O<sub>7</sub>. *Acta Crystallogr.* **21**, 942–948 (1966).
  20. Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. **93**, 629–637 (1997).
  21. Hoffmann, R. An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* **39**, 1397–1412 (1963).
  22. Bernuz, E. huckelpy. <https://github.com/efrembernuz/ExtendedHuckel>.

