USACH. Santiago, Chile 5 - 9 de Junio, 2023

Continuous symmetry measures

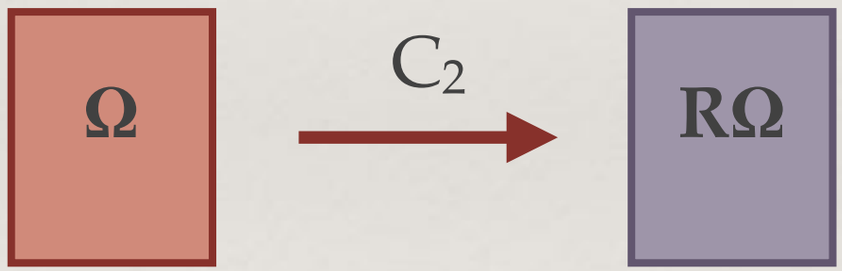
Pere Alemany
Universitat de Barcelona

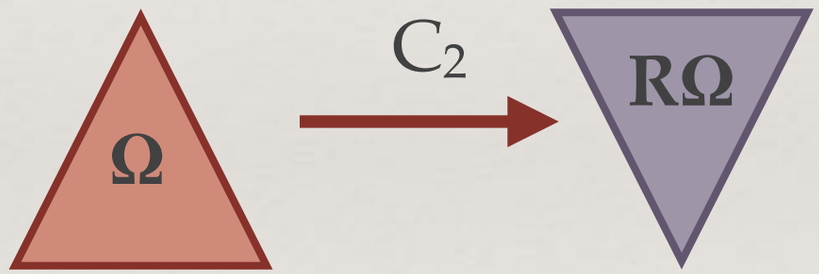
How do we detect symmetry in a molecule?

Given a spatial configuration $\mathfrak{F}$, those **automorphisms of space** which leave $\mathfrak{F}$ unchanged form a group *G*, and this group describes exactly the symmetry possessed by $\mathfrak{F}$.

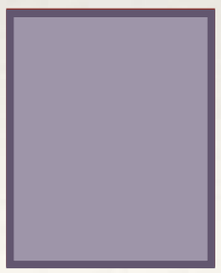
Isometries (distance preserving automorphisms):

- Translations
- Rotations
- Improper Rotations

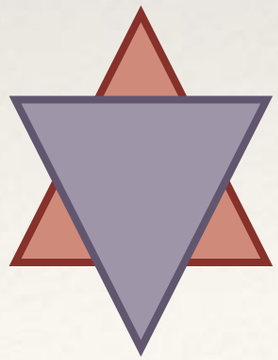




$\langle \Omega \mid R\Omega \rangle = 1$

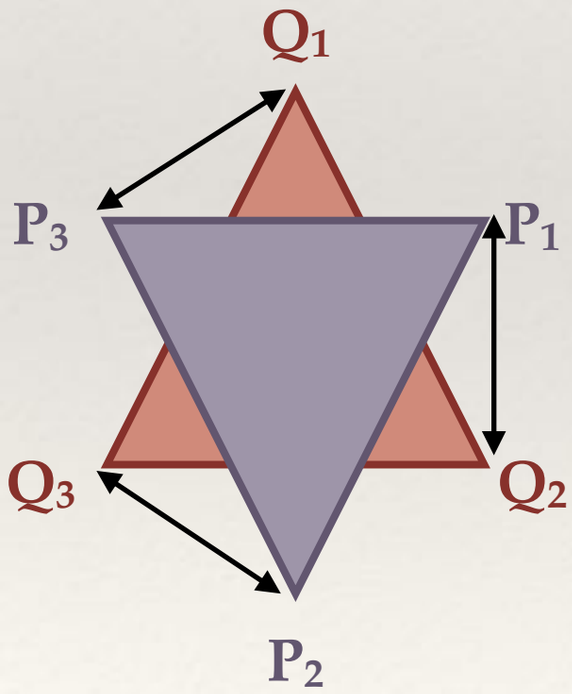


$0 \leq \langle \Omega \mid R\Omega \rangle \leq 1$



Overlap between structures

The definition of the overlap between two structures is not univocally defined, and depends on how we define the structure **Ω**. For structures defined as a set of vertices we may use the distances between vertices to define the overlap



Search for label permutation giving optimal overlap

$$< Q | P > = 1 - min \frac{\sum_{i=1}^{N} | Q_i - pP_i |^2}{\sum_{i=1}^{N} | Q_i - Q_0 |^2}$$

Normalizaton factor for size

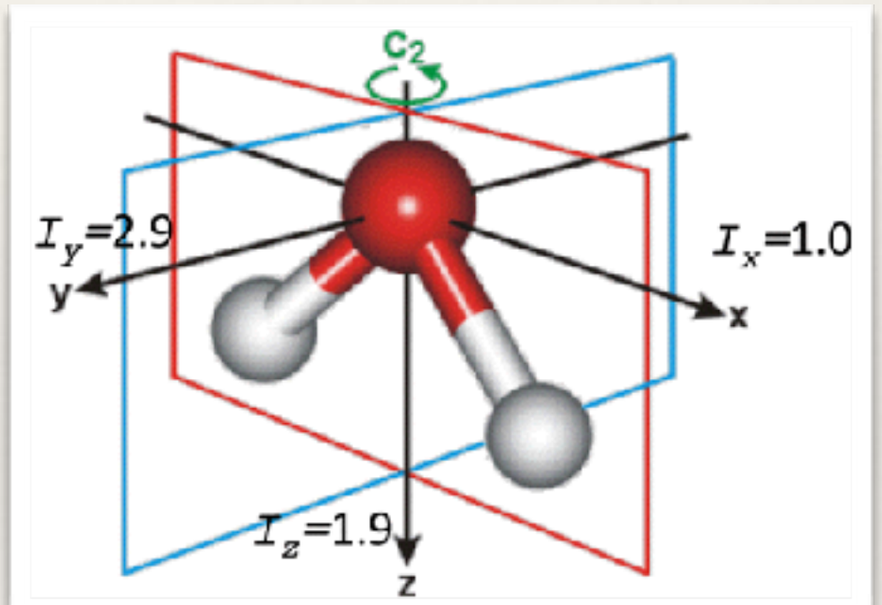
Geometric center for Q

Inertia tensor

To analyze the symmetry of a molecule (geometric shape) it is convenient to translate the **center of mass** (geometric center) and to align it with its **principal axes of inertia**

$$x_{\mathrm{CM}} = \frac{1}{m_{\mathrm{CM}}} \sum_i m_i x_i, \quad y_{\mathrm{CM}} = \frac{1}{m_{\mathrm{CM}}} \sum_i m_i y_i, \quad z_{\mathrm{CM}} = \frac{1}{m_{\mathrm{CM}}} \sum_i m_i z_i,$$

$$m_{\mathrm{CM}} = \sum_i m_i$$

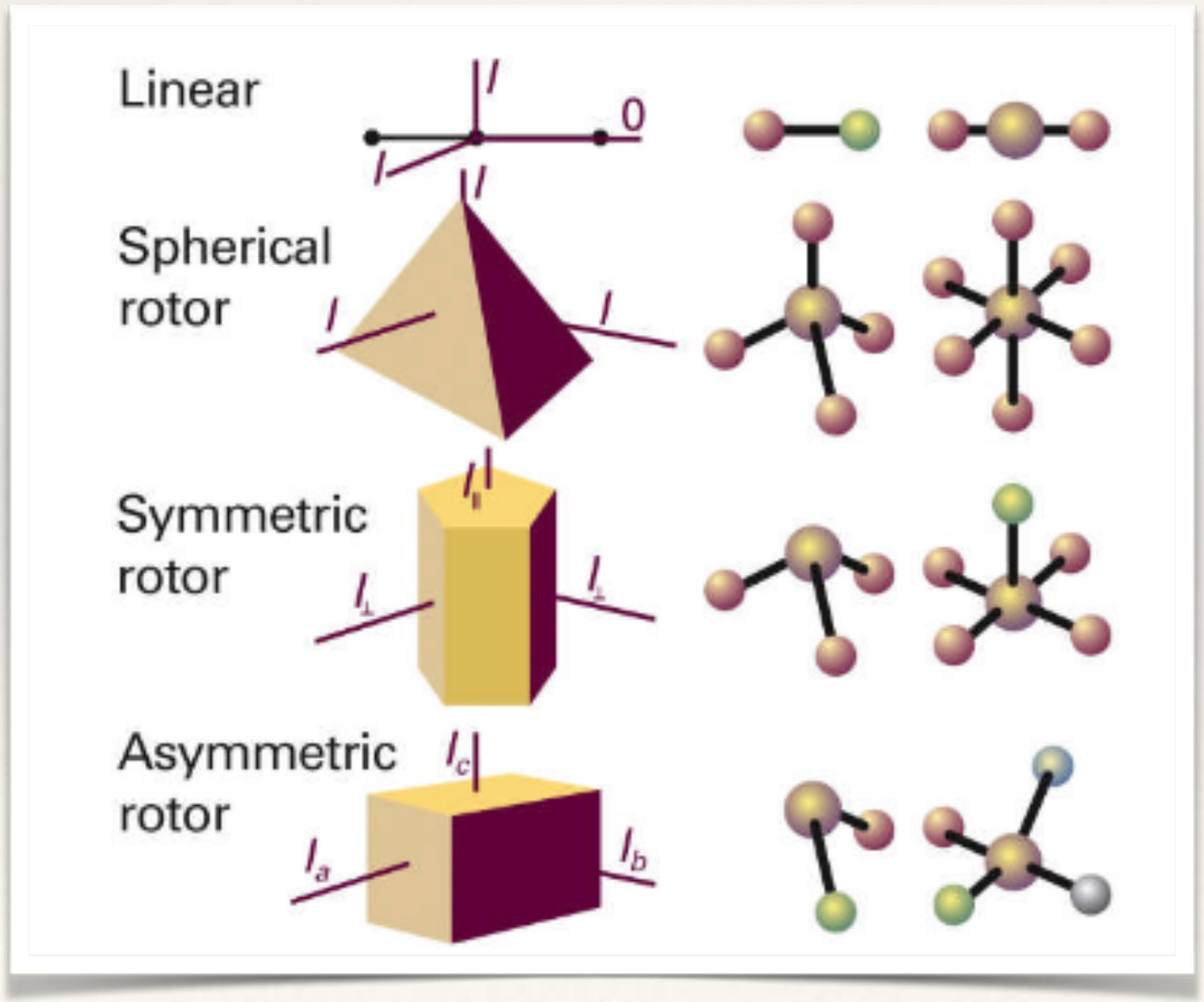


$$I_{xx} = \sum_i m_i \left(y_i^2 + z_i^2\right), \quad I_{xy} = I_{yx} = -\sum_i m_i x_i y_i,$$
$$I_{yy} = \sum_i m_i \left(z_i^2 + x_i^2\right), \quad I_{xz} = I_{zx} = -\sum_i m_i z_i x_i,$$
$$I_{zz} = \sum_i m_i \left(x_i^2 + y_i^2\right), \quad I_{yz} = I_{zy} = -\sum_i m_i y_i z_i.$$

$$\begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{bmatrix} = \begin{bmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{bmatrix} \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix}$$

Moments of inertia and symmetry

Molecules can be classified into four different groups according to their principal moments of inertia



Linear rotors: $I_a = I_b \neq I_c = 0$

$C_{\infty v}$, $D_{\infty h}$

Spherical rotors: $I_a = I_b = I_c$

T_d, O_h, I_h

Symmetric rotors: $I_a = I_b \neq I_c$

C_n, D_n, S_n, C_{nv}, C_{nh}, D_{nh}, and D_{nd} if $n \geq 3$

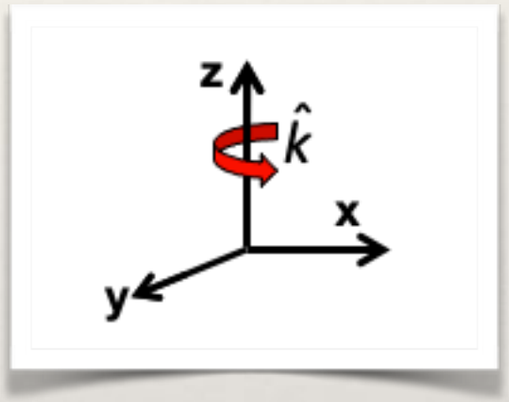
Asymmetric rotors: $I_a \neq I_b \neq I_c$

How do we apply a rotation to a molecule?

If our molecular structure is defined by a set of points (positions of the nuclei) with the origin at the center of mass (geometric center), any point symmetry operation can be executed by a simple matrix multiplication:

$$\begin{pmatrix} x'_1 & x'_2 & \dots & x'_N \\ y'_1 & y'_2 & \dots & y'_N \\ z'_1 & z'_2 & \dots & z'_N \end{pmatrix} = \begin{pmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \end{pmatrix}$$

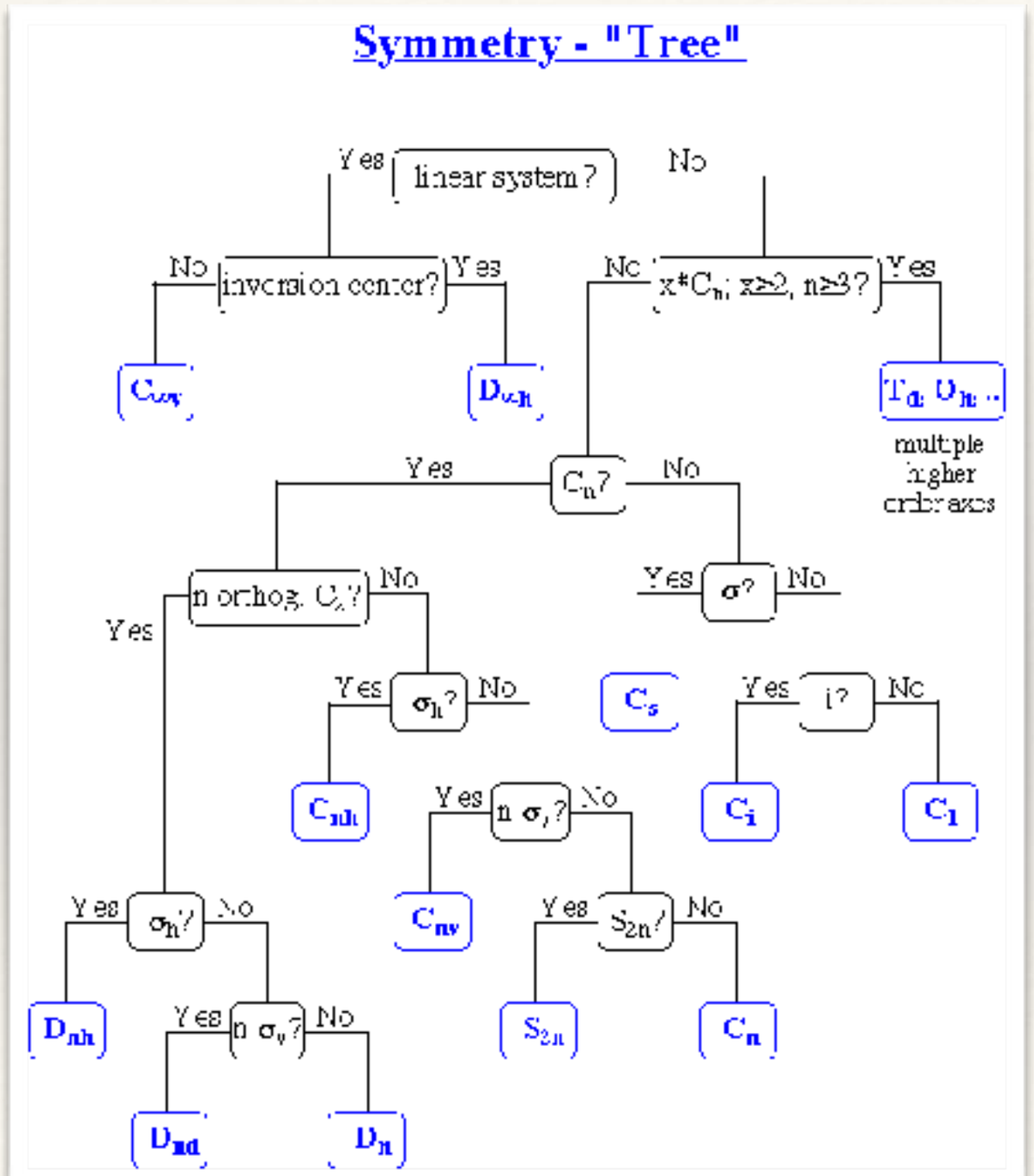
Rotated coordinates | Rotation matrix | Original coordinates



Rodriguez's formula: rotation matrix for an arbitrary axis (n_1, n_2, n_3):

$$R(\hat{\boldsymbol{n}}, \theta) = \begin{pmatrix} \cos\theta + n_1^2(1-\cos\theta) & n_1 n_2(1-\cos\theta) - n_3\sin\theta & n_1 n_3(1-\cos\theta) + n_2\sin\theta \\ n_1 n_2(1-\cos\theta) + n_3\sin\theta & \cos\theta + n_2^2(1-\cos\theta) & n_2 n_3(1-\cos\theta) - n_1\sin\theta \\ n_1 n_3(1-\cos\theta) - n_2\sin\theta & n_2 n_3(1-\cos\theta) + n_1\sin\theta & \cos\theta + n_3^2(1-\cos\theta) \end{pmatrix}$$

How to find the point group?



Once the molecule is centered with the CM in the origin and we know the principal moments of inertia we just need to use the decision tree looking for the presence of symmetry some key operations checking if:

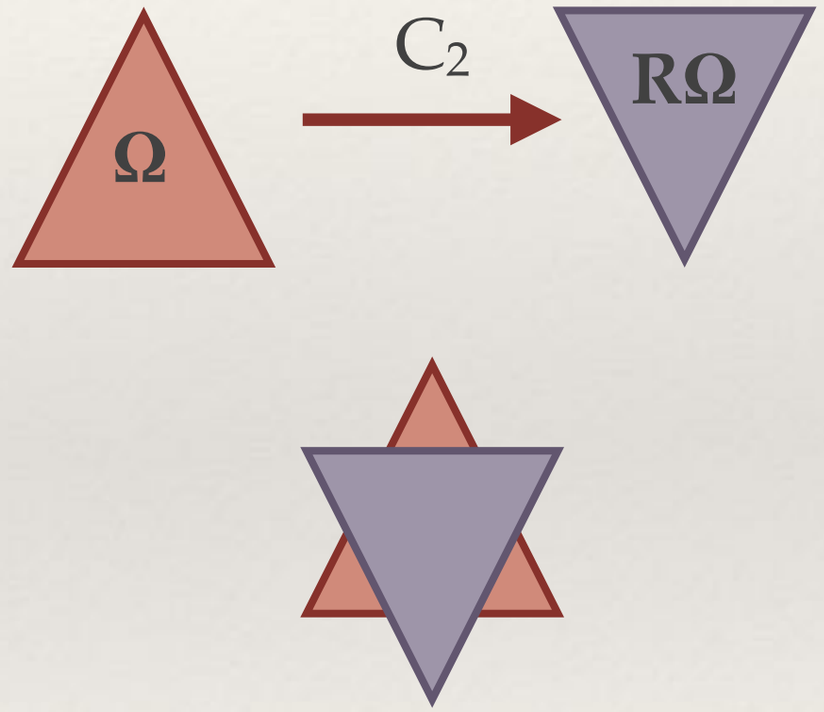
$$< Q | P > = 1 - min \frac{\sum_{i=1}^{N} | Q_i - pP_i |^2}{\sum_{i=1}^{N} | Q_i - Q_0 |^2}$$

is 1 or not.

In the program we allow for a tolerance factor ε and consider that a symmetry operation is present when $<Q|P> -1 \leqslant \varepsilon$

Continuous symmetry measures

When a given transformation is **not a symmetry operation** for the **Ω** object we will get an overlap between 0 and 1.



In this case we can use this **overlap to measure** how far the object is from having the desired symmetry by taking into account the overlaps calculated for **all operations** in a given group and minimizing them with respect of the **orientation** of the symmetry elements

$$0 \leq \langle \Omega \mid R\Omega \rangle \leq 1$$

Continuous Symmetry Measures

Let us consider an arbitrary object $|\Omega\rangle$ in a metric space and a point-symmetry group G with h R_i operations

$$S(\Omega,\boldsymbol{G}) = 100 \cdot \min\left(1 - \left\langle \hat{G} \right\rangle\right) = 100 \cdot \min\left(1 - \frac{\sum_{i=1}^{h} \left\langle \Omega \middle| \hat{R}_i \Omega \right\rangle}{h \left\langle \Omega \middle| \Omega \right\rangle}\right)$$

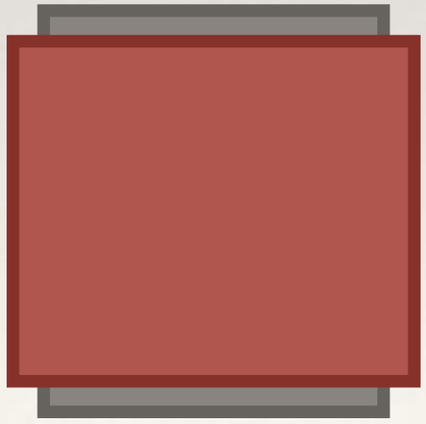
minimization with respect to orientation of symmetry operations & vertex permutations

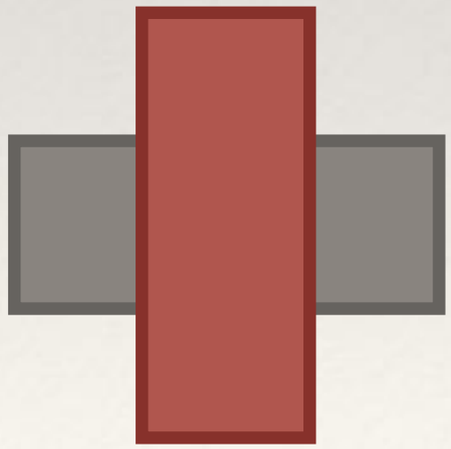
average over all h symmetry operations in the group

overlap between the original and the transformed object

$$\left\langle \Omega \middle| \hat{R}_i \Omega \right\rangle \approx 1$$

$$S(\Omega,\boldsymbol{G}) \approx 0$$

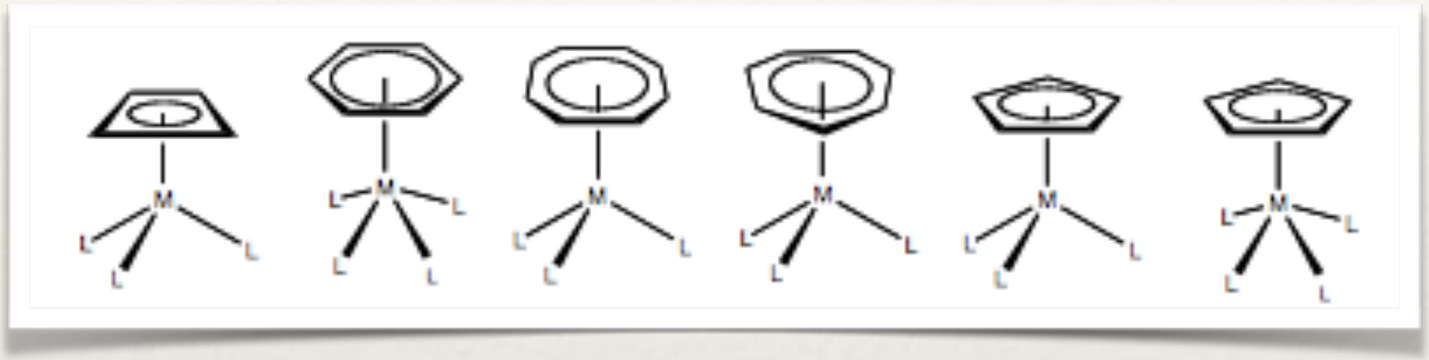




$$\left\langle \Omega \middle| \hat{R}_i \Omega \right\rangle < 1$$

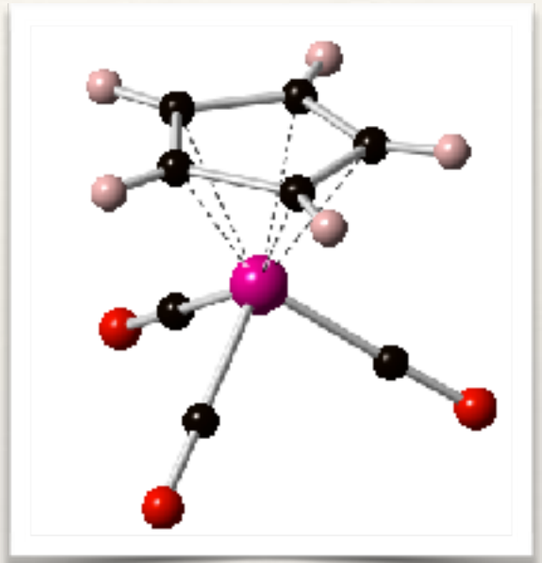
$$S(\Omega,\boldsymbol{G}) > 0$$

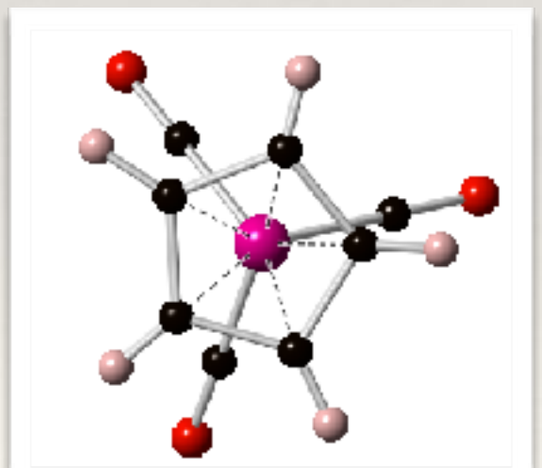
Piano Stool Compounds



m/n	Data	$S(C_m)$	$S(C_n)$	$\alpha(^o)$
4/3	32	0.06	0.04	2.0
6/4	23	0.10	0.20	1.0
8/3	18	0.03	0.40	2.0
7/3	30	0.20	0.80	3.0
5/3	489	0.04	0.10	2.0
5/4	20	0.04	0.10	2.0

Despite the flexibility of one part of the molecule, the two collinear incompatible rotation axes are maintained





Continuous Chirality Measures

Let us consider an arbitrary shape $|R\rangle$ and its enantiomeric shape $|S\rangle$

$$CCM(R) = 100 \cdot \min\left(1 - \frac{\langle R|S\rangle}{\langle R|R\rangle}\right)$$

overlap between R and S

normalization factor

minimization with respect to relative position, orientation, and size of R and S

$$CCM(R) = CCM(S)$$

Achiral shape

$$\frac{\langle R|S\rangle}{\langle R|R\rangle} = 1$$

$$CCM(R) = 0$$

Chiral shape

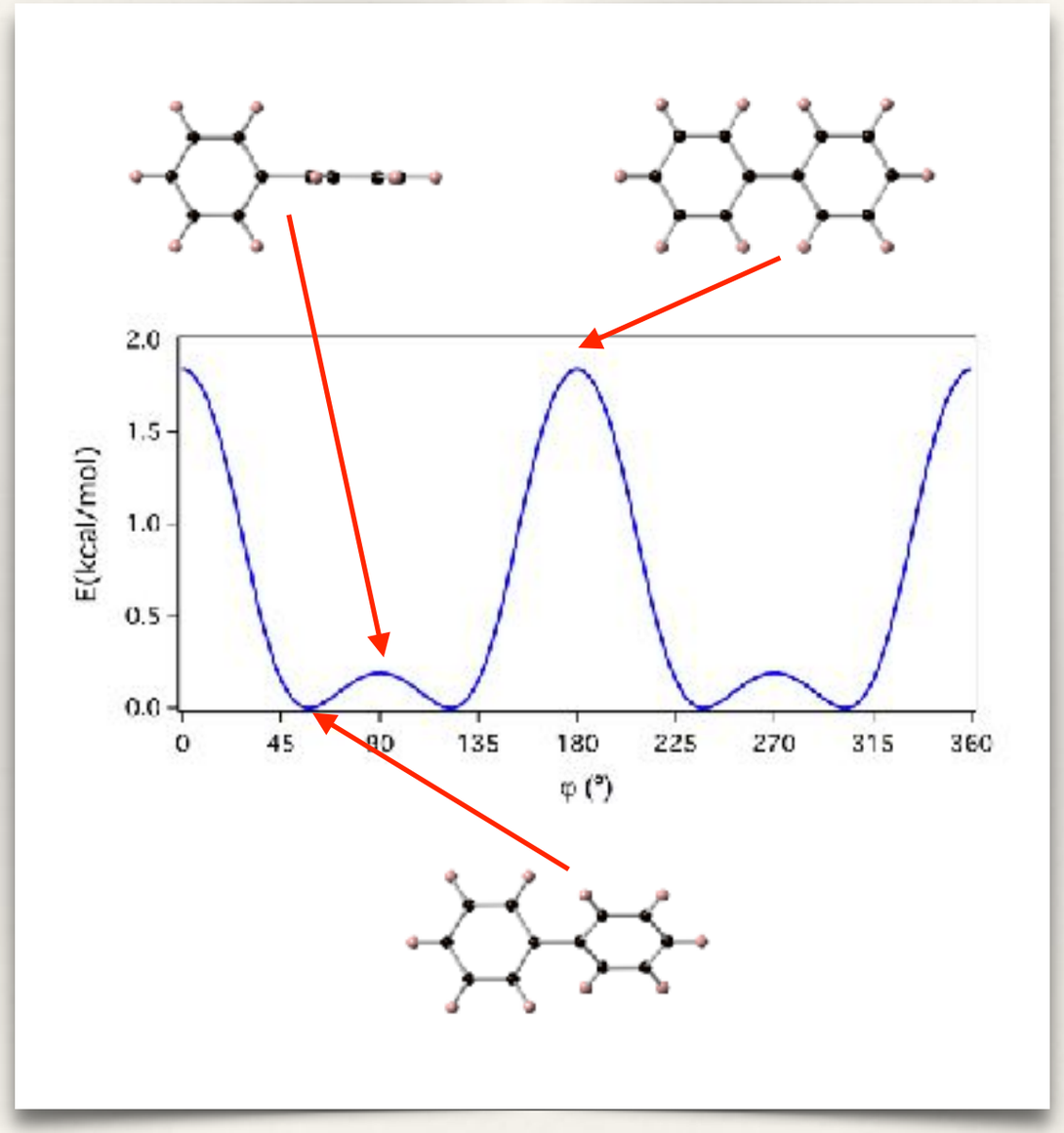
$$\frac{\langle R|S\rangle}{\langle R|R\rangle} < 1$$

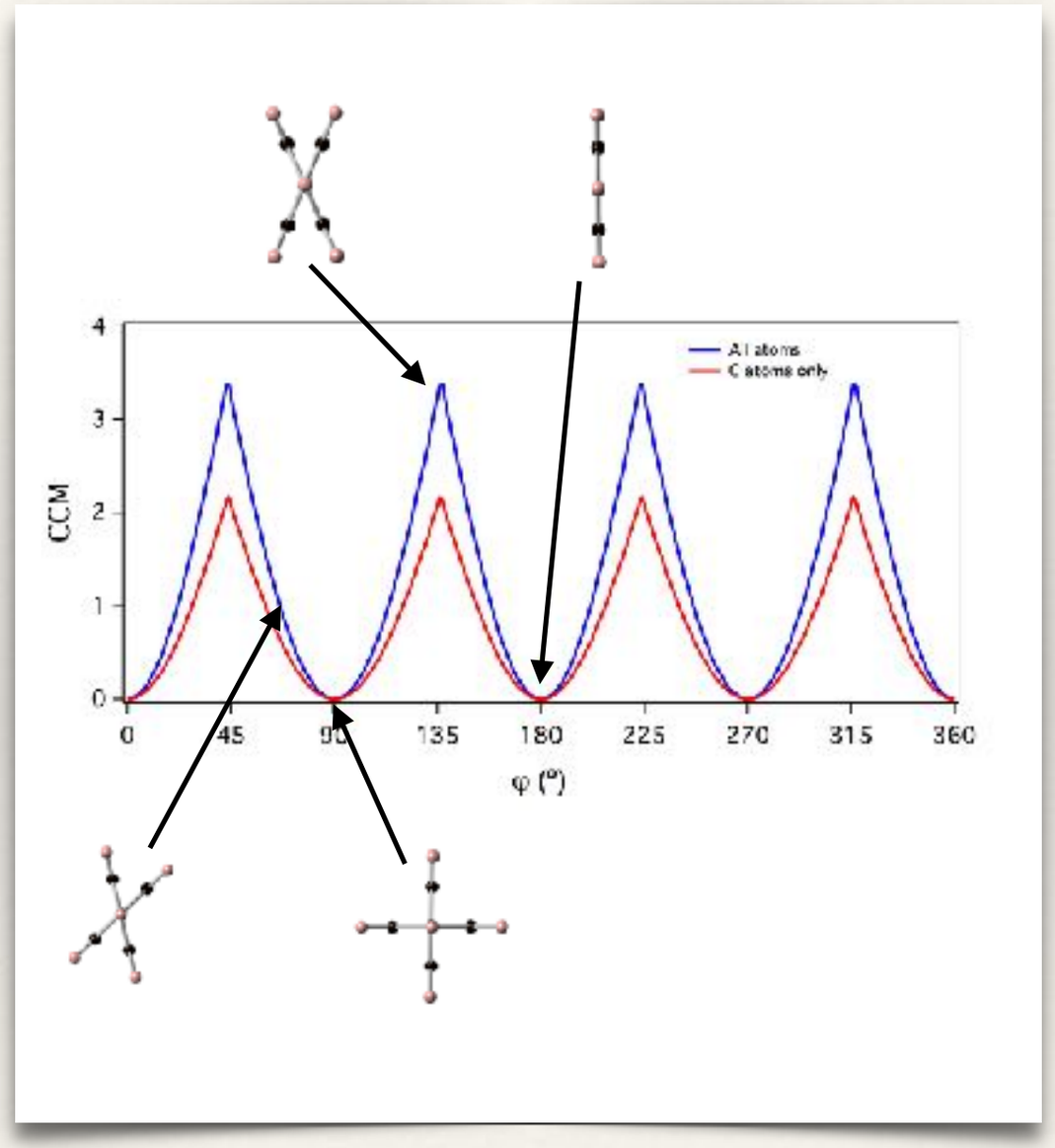
$$CCM(R) > 0$$

$$0 < CCM(R) < 100$$

$$CCM(R) = \min\{S(R,S_1), S(R,S_2), S(R,S_4)...\}$$

Floppy molecules: chirality of biphenyl





PM6 calculations

Continuous Shape Measures

Let us consider an arbitrary shape $|Q\rangle$ and a reference (ideal) shape $|P\rangle$

$$S(Q,P) = 100 \cdot \min\left(1 - \frac{\langle Q|P\rangle}{\langle Q|Q\rangle}\right)$$

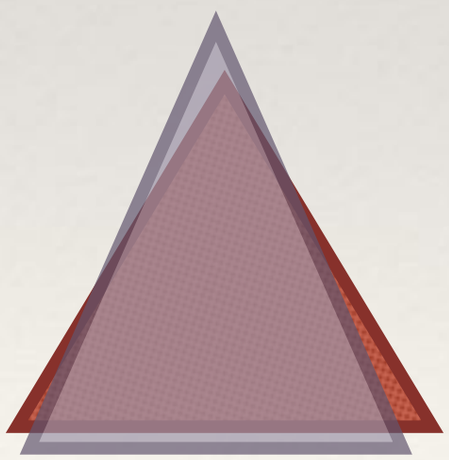
overlap between P and Q

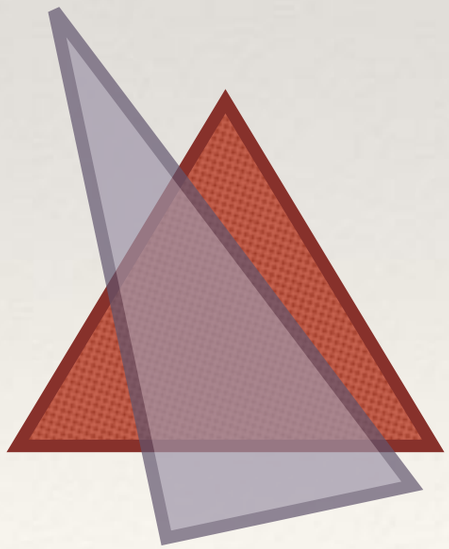
normalization factor

minimization with respect to relative position, orientation, and size of Q and P

$$\frac{\langle Q|P\rangle}{\langle Q|Q\rangle} \approx 1$$

$$S(Q,P) \approx 0$$





$$\frac{\langle Q|P\rangle}{\langle Q|Q\rangle} < 1$$

$$S(Q,P) > 0$$

$$0 < S(Q,P) < 100$$

Similarity Between Shapes

The basic ingredient to calculate a CShM is the overlap between the problem shape $|Q\rangle$ and a reference (ideal) shape $|P\rangle$

$$S(Q,P) = 100 \cdot \min\left(1 - \frac{\langle Q|P\rangle}{\langle Q|Q\rangle}\right)$$

Measures the difference (dissimilarity) between $|Q\rangle$ and $|P\rangle$

$$<Q|P> = <Q|Q>$$ Same shape

$$<Q|P> < <Q|Q>$$ Different shapes

If $|Q\rangle$ and $|P\rangle$ are represented by a distribution function (positive definite):

$$|Q> = \rho_Q(x,y,z)$$
$$|P> = \rho_P(x,y,z)$$

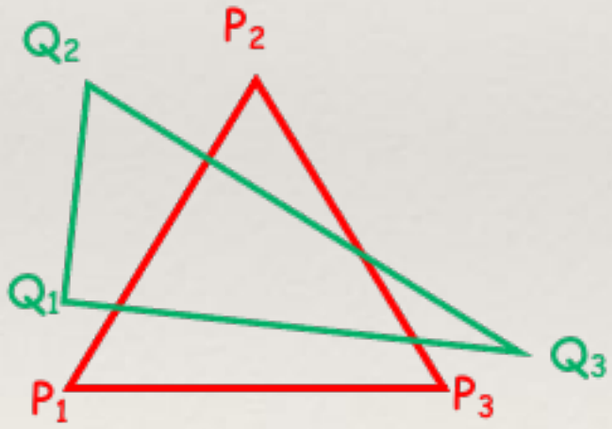
$$\longrightarrow \quad <Q|P> = \iiint_{x,y,z} \rho_Q(x,y,z)\rho_P(x,y,z)dxdydz$$

Discrete structures

Shapes are often described as a discrete set of points (vertices) in Euclidean space

$$Q = \left\{ \left(x_1, y_1, z_1\right), \ldots, \left(x_n, y_n, z_n\right) \right\}$$

The CShM can be obtained by minimizing distances between vertices:



$$S_Q(P) = \min\left\{ \frac{\sum_{i=1}^{N} \left| \vec{Q}_i - \vec{P}_i \right|^2}{\sum_{i=1}^{N} \left| \vec{Q}_i - \vec{Q}_o \right|^2} \cdot 100 \right\}$$

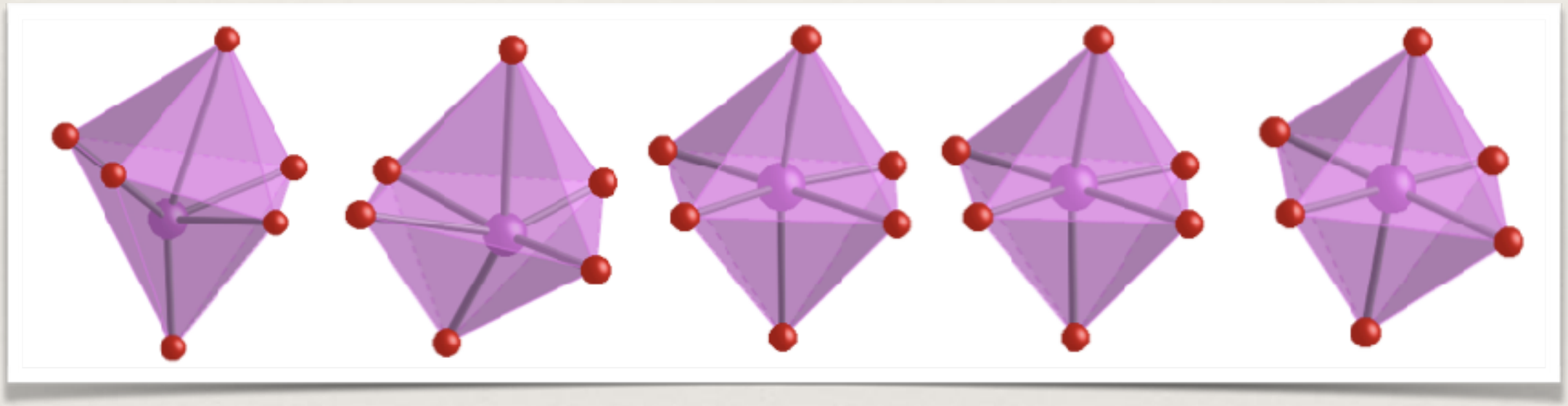
minimization with respect to relative position, orientation, and size and labelling of vertices in Q and P

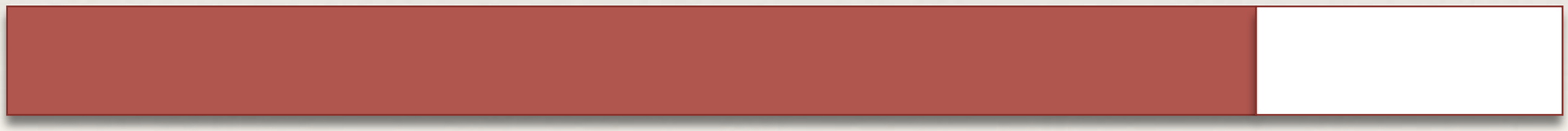
H. Zabrodsky, S. Peleg, D. Avnir, *J. Am. Chem. Soc.* **1992**, *114*, 7843

Measuring (A)symmetry

Strongly distorted **Moderately distorted** **Slightly distorted** **Octahedron**

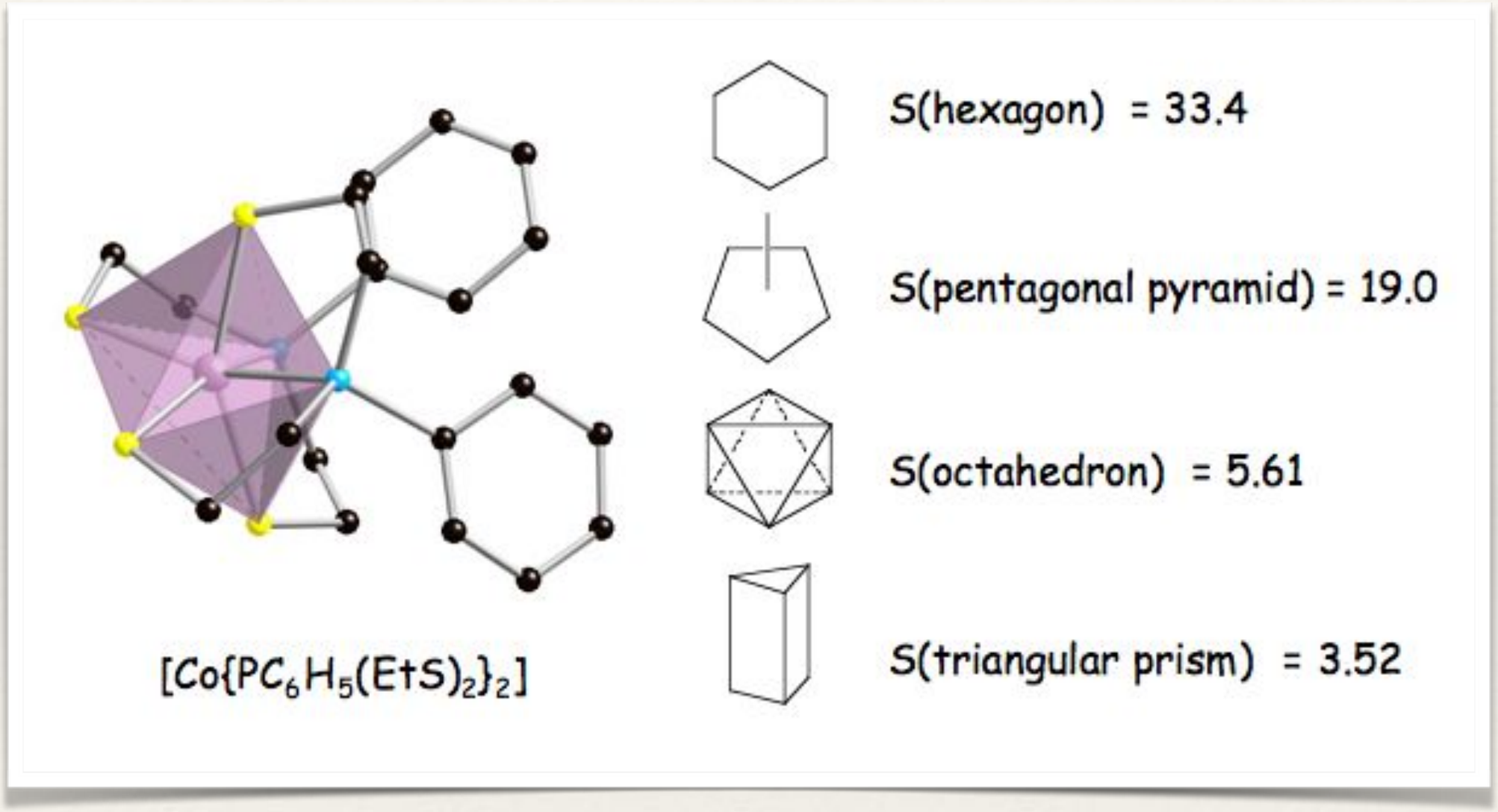
5.65	1.70	0.68	0.12	0.00





Non-octahedral **Octahedral**

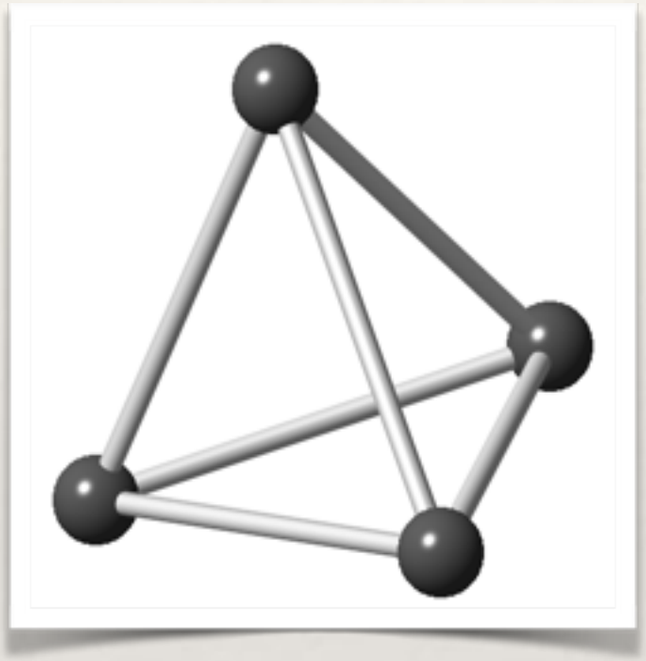
Which Coordination Geometry?



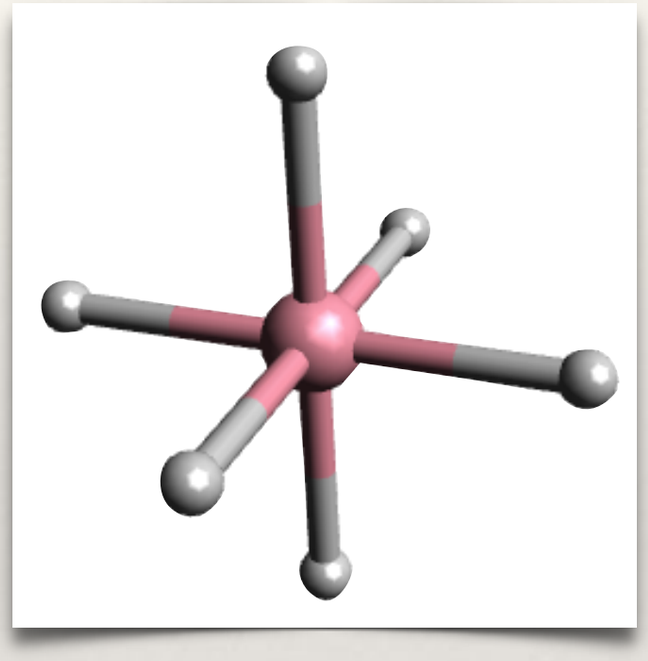
Coord. Chem. Rev. **249** (2005), 1693-1708

Polyhedra in CShM

It is important to distinguish between regular and centered structures:



L_n polyhedron

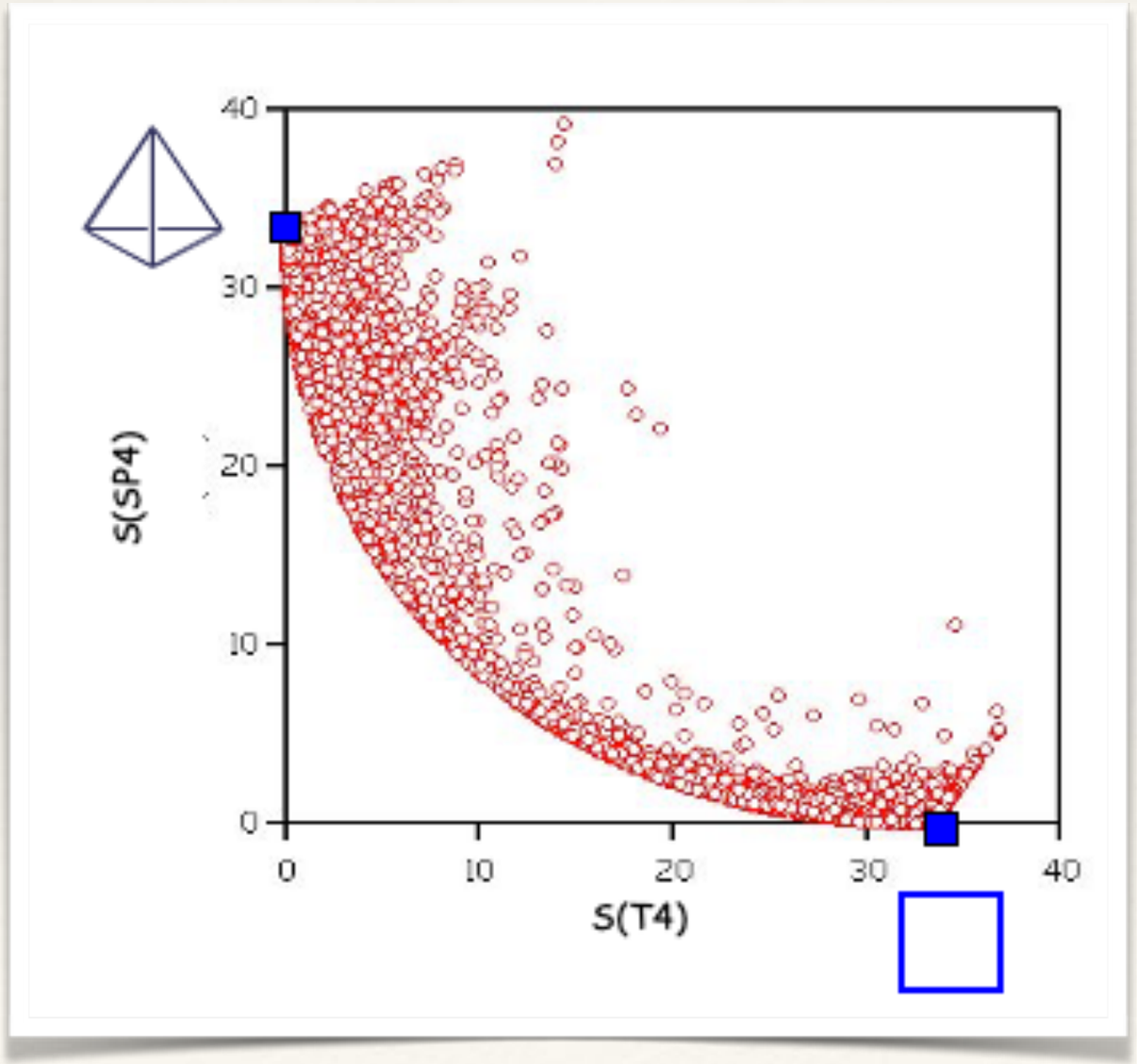


centered ML_n polyhedron

Permutations between M and L are not allowed in the optimization process.
For centered polyhedra we must explicitly indicate which vertex is M.

Shape Maps

Evaluate two different shape measures for taking different polyhedra P as a reference: $S_Q(P_1)$, $S_Q(P_2)$ and plot the results as a $S_Q(P_2)$ vs $S_Q(P_1)$ graph

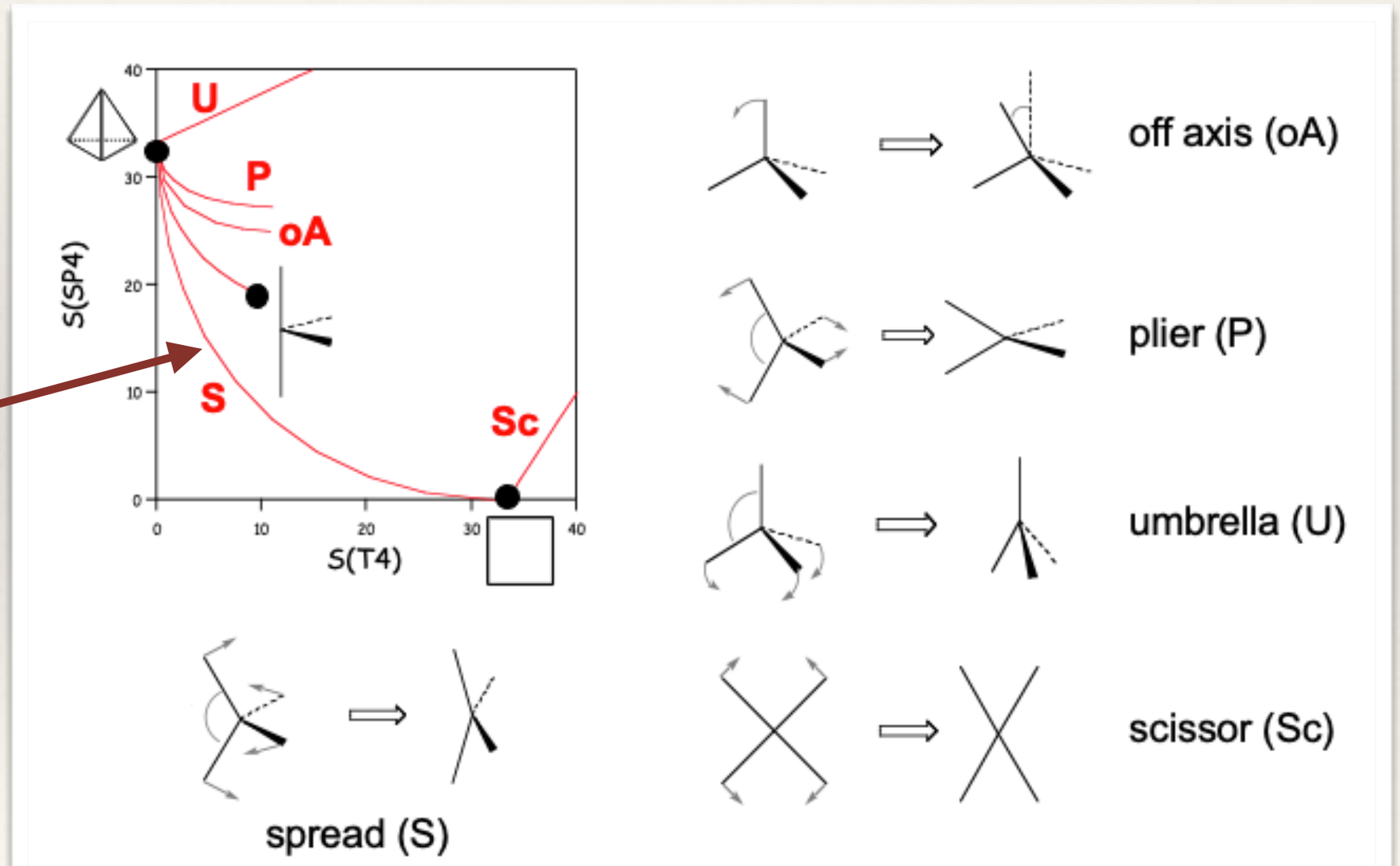


Tetrahedron – square shape map for 10,000 random generated ML4 structures with equal M-L distances

Distortion paths

Shape maps are useful in distinguishing between different types of distortions for a given ML_n structure

Minimal distortion path:
How to go from P_1 to P_2 loosing the minimal P_1 shape along the path



Tetrahedral or Square planar?