Searching similar clusters of polyhedra in crystallographic databases

Hans-Joachim Klein and Christian Mennerich
Institut für Informatik
Universität Kiel, 24098 Kiel, Germany
{hjk,chm}@is.informatik.uni-kiel.de

Abstract

A graph-based method is described for searching and ranking clusters of polyhedra in large crystallographic databases. It is shown how topologically equivalent substructures can be determined for a given target cluster based upon a graph representation of polyhedral networks. A mathematical modeling of geometric embeddings of polyhedra graphs is provided which can be used to define geometric similarity of polyhedral clusters. For a special kind of similarity, an algorithm for solving the problem of absolute orientation is applied in order to rank topologically equivalent clusters appropriately.

Keywords: Crystallographic databases, polyhedral clusters, polyhedra graphs, similarity search, ranking

1 Introduction

In recent years, large databases have been built in organic as well as in inorganic chemistry [2], [3]. These systems offer query facilities for searching compounds given publication data, kinds of atoms, symmetry information, etc. For databases storing information on organic and metal-organic crystal structures, it is also possible to search for certain patterns of combinations of atoms [4]. Similarity searching in general has received considerable attention in the field of molecular structures modeled as simple undirected graphs [5]. Rigid substructures can be distinguished and used to build indexes for fast search.

At present, such a kind of searching at the level of substructures is not offered for inorganic crystallographic databases. Whereas for organic compounds, search can be built upon a set of substructures of reasonable size this approach is less meaningful for inorganic compounds. Here a large variety of chemical elements and patterns can be observed. The symmetry is generally higher than in organic and many inorganic molecules leading to a rich diversity of geometric configurations. Local arrangements of atoms play an important role for the description and understanding of structures.

In order to deal with this situation, an approach has been presented in [6] which is based upon a description of inorganic crystal structures at the level of coordination polyhedra. Infinite networks formed by connections of polyhedra can be represented by finite periodic graphs. This modeling allows to build an indexation of polyhedral networks by chains, which can be used for the efficient determination of topologically equivalent substructures. These structures can have a quite different geometry. Hence a method is needed to check for geometric similarity.

It has been argued that to determine geometric transformations first and then to test for preservation of topology is more efficient in connection with geometric graph isomorphism [7]. However, when using rotations, translations, and scaling to investigate geometric isomorphism the problem arises that substructures with great similarity up to a sharp difference at a single position cannot be found in principle.

In order to be flexible with respect to the definition of similarity and to allow the user to decide which differences in the geometry are tolerable, a two-step approach for determining similar structures is applied. In the first step, all substructures in a given set of model structures are determined, which are candidates for the result because they are topologically equivalent to the given search structure. By taking the symmetries of structures into account, the set of candidates can be reduced to symmetrically non-equivalent substructures. In the second step, geometric similarity is checked for all candidates resulting in a ranking. This ranking can be used for presenting the search results. Furthermore, the concrete values of the similarity test provide information about the relationship between the structures.

This paper is an extended version of [1]. It is organized as follows. We start with describing some methods for determining coordination polyhedra as they are implemented in our system. Then we review the graph representation of clusters of polyhedra and the definition of topological equivalence. We discuss some properties of the ordered face representation of polyhedra and show how the embedding of subgraphs in model graphs can be done quite effi-

ciently in many cases. In the fourth section, the problem of geometric similarity is discussed more generally. A modeling of polyhedral clusters as joint structures is developed and some forms of similarity based upon this modeling are proposed. In Section 5, for one such form which is based upon point sets the implementation in our system POLYSEARCH is described and some results are presented. We conclude with remarks concerning the usage of the system and future work.

2 Graph representation and topological equivalence of polyhedral clusters

Graphs are often used to describe structural aspects of chemical compounds. In this section we show how the bonding between atoms in crystals can be represented by special forms of labeled graphs.

2.1 Coordination polyhedra

Inorganic crystal structures are often modeled using coordination polyhedra as components [8]. The vertices of these convex polyhedra represent the atoms which are considered as ligands of a fixed central atom. In case of small sets of ligands and regular forms (e.g. tetrahedra in silicates) the determination of coordination polyhedra is rather straightforward. If the number of ligands and their distance to the central atom grows, polyhedra may become deformed and their determination is less obvious. A generally applicable formal definition of coordination polyhedra in crystal structures seems to be impossible. It is therefore advisable to apply different methods for determining coordination polyhedra and to have a closer look at the results if they show differences. The following methods seem to be suitable into that regard and are offered by our system POLYSEARCH [9]:

- *Search within a given maximal distance.*

All atoms within the given distance are considered as ligands. It is checked whether they determine a convex polyhedron.

- *Determination of the maximal convex hull.*

A convex hull algorithm is used in order to determine the coordination polyhedron (the gift wrapping algorithm [10] is appropriate since the number of vertices of coordination polyhedra is small).

- *Determination of the maximal convex hull restricted by a given number of atoms.*

The convex hull algorithm stops when the given number of atoms is reached.

- *Search for a polyhedron with a specified number n of ligands.*

Sets of *n* atoms around the central atom are built and checked for convexity. Atoms not contained in such a set may not have a distance to the central atom smaller than the distance to the central atom of any atom of the set.

- *Search for a maximal gap (linear).*

Circumscribing spheres are defined by specifying an upper limit ϵ for the differences between the distances of neighbouring atoms to the central atom. Atoms a, a' belong to the same sphere if there is a sequence $a = a_1, ..., a_n = a'$ such that the distances of a_i and a_{i+1} to the central atom are less than or equal to ϵ for $i = 1, ..., n-1$. A gap between two spheres S, S' with S' circumscribing S is the difference between the minimal distance of an atom in S' and the maximal distance of an atom in S. The gap Δ between S and S' is maximal if all gaps between spheres circumscribed by S and the gap between S' and the next sphere are smaller than Δ.

- *Search for a maximal gap (volume).*

A similar proceeding as in the linear case is applied. Instead of the differences of distances the differences of the volumes of the spheres determined by the gaps are considered.

The linear gap method is often used in the literature [11] and applied in systems such as Pearson's Crystal Data [12] for determining coordination polyhedra. Because of problems which may arise in determining the maximal gap, Pearson's Crystal Data also offers a maximal convex hull algorithm. The volume method seems to be more appropriate in connection with the use of distances between atoms if no assumptions can be made about the arrangement of ligands around the central atom in space.

Further conditions for the search of coordination polyhedra in a given structure can be the kind of central atoms and the kind of ligands. If only homogeneous polyhedra are of interest, for example, a single kind of atoms can be allowed as ligands.

In many structures coordination polyhedra with a small number of ligands (up to six) in near neighbourhood to the central atom are quite regular and have the symmetries of the corresponding ideal forms (tetrahedra, octahedra). Such regularity can, however, often not be found if the number of ligands and their distance to the central atom increases. Then deformations are frequent and the problem of how to measure polyhedral distortion arises [13]. POLYSEARCH compares a polyhedron found by one of the methods presented above with the description of a set of ideal polyhedra. This comparison is done on the basis of the adjacency matrices of the graphs representing the adjacency of the ligands.

For the representation of polyhedra and clusters of polyhedra, a graph form has been introduced in [6]. It uses a unique numbering of vertices and a description of faces by ordered sequences of numbers for coordination polyhedra. Depending on the purpose of their usage, these graphs can be augmented by coordinates to get a complete description

of the geometry of clusters or they can be reduced to pure topological information. Two views of coordination polyhedra are distinguished in [6] using the well-known correspondence between convex polyhedra and three-connected planar graphs (Theorem of Steinitz).

Definition: The *geometrical view* of a coordination polyhedron P is a vertex-labeled simple three-connected planar graph $(V \cup \{c\}, E, pos)$. The vertex set $V \cup \{c\}$ represents the ligands of P and the central atom, respectively; the set of edges is determined by the adjacency relationship of the ligands of P and the function $pos : V \cup \{c\} \rightarrow At \times \mathbb{R}^3$ assigns to every element of the vertex set the element symbol and the coordinates of the corresponding atom of P.

Information on the symmetries of the polyhedron is not included in the definition since it may be derived from V and pos.

In the first step of our similarity check we look for all clusters in a given set of model structures which are topologically equivalent to a given search cluster. For this search, position data of atoms and the kind of atoms involved are not needed. Hence the following view of polyhedra is suitable:

Definition: The *pure topological view* of a coordination polyhedron P is a simple three-connected planar graph (V, E) with V representing the ligands of P and E the adjacency relationship of the ligands.

Since there is no information on locations of atoms, the central atom must not be represented by a vertex. The topological view of a polyhedron provides no information on its symmetry group. If symmetry properties are of interest, the symmetry group of the polyhedron or subgroups of it may be added.

When polyhedra are considered as elements of clusters of polyhedra some of their vertices become fixed as connecting points. In case no geometric information shall be used in order to refer to these vertices, some characterization at the topological level is necessary. The following representation by faces has shown to be useful in that regard [6]:

Definition: Let P be a coordination polyhedron. An *ordered face representation* of P can be obtained in the following way: Use elements of $\{1, ..., n\}$ to number the ligands $\{l_1, ..., l_n\}$ in a unique way. For every face f, the numbers of its vertices are arranged into a unique sequence as follows: P is viewed from outside and the vertices of f are collected clockwise starting with the smallest element.

Figure 1 shows a regular octahedron and an ordered face representation of it. For any polyhedron with the same topological view the same ordered face representation can be obtained by using an appropriate numbering for its vertices. For a polyhedron P with n vertices there are $n!$ different ways to number these vertices; the number of different ordered face representations, however, is $\frac{n!}{r}$. r is the order

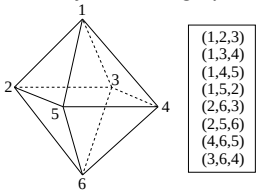

Figure 1: Ordered face representation.

of the rotation group of that ideal polyhedron which has the same topological view as P and highest symmetry. This follows from the definition of an ordered face representation, which does not refer to the geometry of a polyhedron.

Consider the topological view of a polyhedron P. An ordered face representation of P induces directions for the edges of the graph. For a digraph, rotations are automorphisms that do not change the direction of edges. As a consequence, a given ordered face representation of P is not changed as well by rotations. Consider the octahedron in Figure 1. Apply the rotation given by the permutation $(1563)(2)(4)$. For the resulting numbering of vertices we get the same ordered face representation of the octahedron.

The rotation groups of the Platonic solids are well known. For arbitrary polyhedra they can be determined using finding algorithms for geometric automorphism groups [14]. n-geometric automorphism groups of a graph can be displayed as symmetries of a drawing of the graph in n dimensions. Though the problem to determine whether a graph has a nontrivial geometric automorphism in two dimensions is NP-complete [15], it has been demonstrated that using a group-theoretic method is very efficient in practice for finding all 2- and 3-geometric automorphism groups of a graph [14]. Since the set of different topological views of coordination polyhedra in inorganic structures is finite, we can assume that the rotation group is given for every polyhedron under consideration.

2.2 Polyhedral clusters

Two coordination polyhedra of a structure can be connected by sharing common ligands. Three different forms of connection are of interest: vertex-, edge-, and face-sharing, which means that the polyhedra share one, two, or more common ligands, respectively. Connecting edges and faces must be edges and faces of both polyhedra and partial overlapping is not allowed. It is also possible that more than two polyhedra share a common vertex or edge.

Figure 2 shows face-sharing octahedra in sodium chloride and edge- as well as vertex-sharing octahedra and vertex-sharing tetrahedra in the silicate jadeite.

The ordered face representation allows to distinguish the

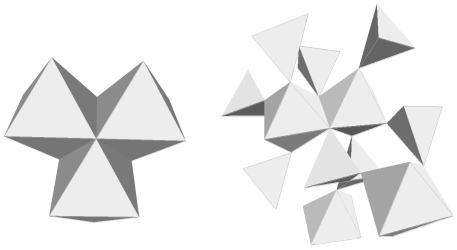

Figure 2: Coordination polyhedra in sodium chloride and the silicate jadeite.

relative orientation of polyhedra in clusters under certain conditions. Consider Figure 3 showing two pairs of square pyramids connected by an edge. There is no numbering scheme for the pyramids such that the ordered face representations and the labels of the connecting vertices become identical. In Figure 4 a) two tetrahedra of a chain

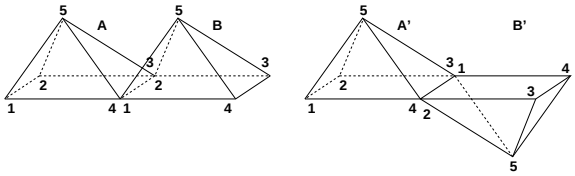

Figure 3: Non-equivalent pairs of pyramids.

of polyhedra are connected by a cube with connecting vertices belonging to the same face of the cube (a so-called trans-edge). In Figure 4 b) the connection is by vertices of different faces of the cube. There is no possibility to choose vertex numberings for the polyhedra such that the corresponding polyhedra of the chains have the same face oriented representation and the connecting vertices have the same number in both chains. This means that the face oriented representation allows to distinguish both chains.

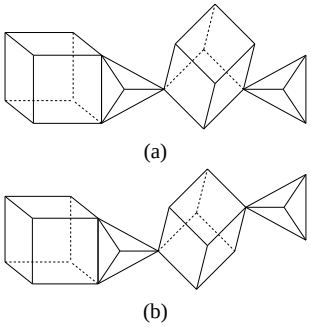

Figure 4: Non-equivalent chains of polyhedra.

Clusters of polyhedra can be described by graphs with nodes representing the polyhedra and edges representing their connections. The information added to the nodes and edges as labels depends on the usage of the graphs. In our approach we make use of topological as well as geometrical information on polyhedra. So we add an ordered face representation for each polyhedron together with the geometrical view as label to every node. The edges are labeled with one or more pairs of natural numbers identifying the polyhedra vertices involved in the corresponding connections. The following definition is from [6].

Definition: Let $\mathcal{P}$ be a cluster of polyhedra with an ordered face representation given for every polyhedron. A *polyhedra graph* for $\mathcal{P}$ is a graph $G_{\mathcal{P}} = (N_{\mathcal{P}}, E_{\mathcal{P}}, \lambda)$ with:
(1) $N_{\mathcal{P}}$ represents the polyhedra in $\mathcal{P}$; every node is labeled with the geometrical view and the ordered face representation of the corresponding polyhedron.
(2) $E_{\mathcal{P}} \subseteq N_{\mathcal{P}} \times N_{\mathcal{P}}$ is the set of directed edges representing every connection between polyhedra in $\mathcal{P}$ in both directions.
(3) $\lambda : E_{\mathcal{P}} \rightarrow 2^{\mathbb{N}\times\mathbb{N}}$ is a labeling function determining for every edge the pairs of vertex numbers involved in the connection between the polyhedra represented by the edge.

Obviously, with every edge a unique inverse is determined. For simplicity we do not remove this redundancy.

The ordered face representation of a polyhedron depends upon the chosen numbering scheme. Hence a polyhedra graph is unique up to these schemes. Figure 5 shows an example with edge- and vertex-sharing.

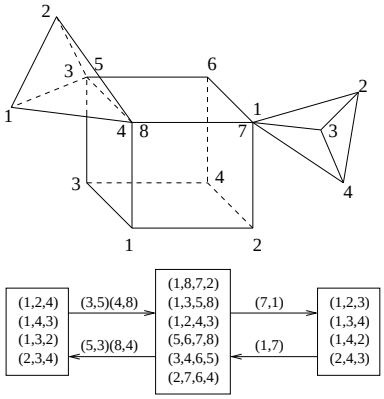

Figure 5: A cluster of polyhedra and its graph.

For labeled graphs to be isomorphic, all labels have to be preserved by the mapping. For topological equivalence the following definition is appropriate:

Definition: Let $G = (N_{\mathcal{P}}, E_{\mathcal{P}}, \lambda)$ and $G' = (N_{\mathcal{P}'}, E_{\mathcal{P}'}, \lambda')$ be two polyhedra graphs. Let $planes$ be a function which assigns the corresponding face representation to every polyhedron.
G and G' are *topologically isomorphic*, written $G \cong G'$, if the following holds:
Let S_i denote the symmetric group of degree i.

There are a bijection $\varphi : N_{\mathcal{P}} \longrightarrow N_{\mathcal{P}'}$ and a mapping $\pi : N_{\mathcal{P}} \longrightarrow \bigcup_{i=1}^{\infty} S_i, n \mapsto \pi_n$, such that

$$
\begin{aligned}
&[\forall n \in N_{\mathcal{P}} : n, \varphi(n) \text{ are isomorphic}] \wedge \\
&[\forall n, n' \in N_{\mathcal{P}}\ \forall (i,j) \in \mathbb{N}^2 : \\
&(n,n') = e \in E_{\mathcal{P}} \wedge (i,j) \in \lambda(e) \\
&\Leftrightarrow \\
&(\varphi(n), \varphi(n')) = e' \in E'_{\mathcal{P}} \wedge (\pi_n(i), \pi_{n'}(j)) \in \lambda'(e')] \wedge \\
&[\forall n \in N_{\mathcal{P}} : f = (i_1, \ldots, i_m) \in \textit{planes}(n) \Leftrightarrow \\
&(f' = (\pi_n(i_k), \ldots, \pi_n(i_m), \pi_n(i_1), \ldots, \pi_n(i_{k-1})) \\
&\in \textit{planes}(\varphi(n)) \wedge \\
&\pi_n(i_k) = \min\{\pi_n(i_1), \ldots, \pi_n(i_m)\})]
\end{aligned}
$$

This definition makes use of the ordered face representation of polyhedra and therefore takes the relative orientation of polyhedra into account.
Two clusters of polyhedra are *topologically equivalent* if their polyhedra graphs are topologically isomorphic.

The geometry of topologically equivalent clusters may differ strongly. Consider Figure 6 showing two rings of square pyramids sharing opposite edges of their square faces. In one of the rings the apices of the pyramids are directed to the outside and in the other ring they show to the inside. Since in both rings the number of pyramids is the same, they are topologically equivalent.

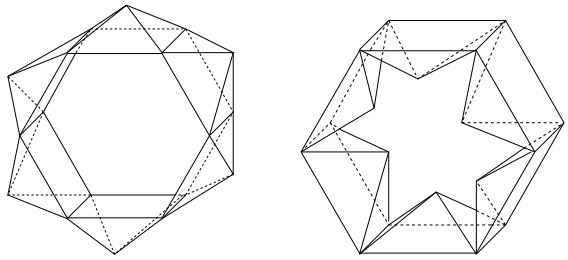

Figure 6: Topologically equivalent clusters.

Despite of this diversity of possible geometric realizations, the definition is practically useful since it does not exclude clusters from the set of candidates in the search for similar substructures when they differ with respect to the embedding in space or the form of their polyhedra but are identical with respect to the type of the polyhedra and their connections. Small deviations in the angles of connected polyhedra may sum up to quite large differences in the overall geometry, but it is difficult to provide a limit for differences being allowed. Furthermore, differences in the relative positioning of two neighbouring polyhedra in otherwise strongly similar clusters would lead to great differences when using methods like root mean square. Hence these differences are analysed in a second step and are not used for keeping structures out of the result.

An efficient method for determining topologically equivalent clusters of polyhedra in a given set of model structures is described in [6]. It uses a special index precomputed for the model structures in order to avoid the well-known complexity problems in connection with subgraph isomorphism. The method is applicable to crystal structures in general, i.e. to the search for finite clusters in infinite but periodic structures, and has been implemented as a web application [9]. In the following, we concentrate on the problem how to determine the geometric similarity of the target substructure and substructures of the infinite model structures; hence we only deal with finite clusters. For the modeling of infinite periodic crystals in general see [6].

3 The embedding problem

For the determination of equivalent substructures in model structures it is sufficient to find one mapping of the polyhedra of the search structure to polyhedra of the substructure such that the corresponding graphs are isomorphic. In order to check for geometric similarity, it is in general not sufficient to consider a single mapping for two graphs but all possible non-equivalent mappings have to be taken into account. The equivalence of mappings can result from symmetries in the model structures or from the existence of alternative permutations of the vertices of polyhedra.

3.1 Possible mappings

Consider the five-membered chain of tetrahedra shown in Figure 7. There are two ways to map the chain C considered as a cluster onto itself: mapping tetrahedron i to itself or - as indicated by C' - to tetrahedron $(4-i)'$, for $i = 0, .., 4$. Obviously, the second mapping does not fit to the geometry of the chain indicated in the figure.
Consider the starlike cluster S of tetrahedra shown in Figure 7. It can be mapped to the cluster S' in three different ways. Only by using the mapping $i \rightarrow i', i = 0, ..., 3$, the cluster S can be moved exactly on S'.

The number of possible mappings can be exponential in the size of the given cluster. Consider the cycle in Figure 8 a). Let it represent a cycle of identical polyhedra with a single kind of connection. Assume a fixed embedding of the cycle into the graph shown in Figure 8 b) to be given. There exist seven further mappings which differ from the chosen mapping only with respect to the alternatives provided by the three 4-membered cycles. It is easy to see that in general for such a cycle with $5 \times n$ nodes there are 2^n possible mappings constructable in this way when the model structure has the same form as the structure in Figure 8 b) with

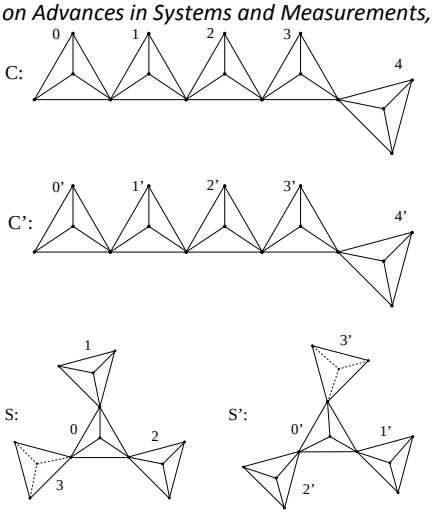

Figure 7: The embedding problem.

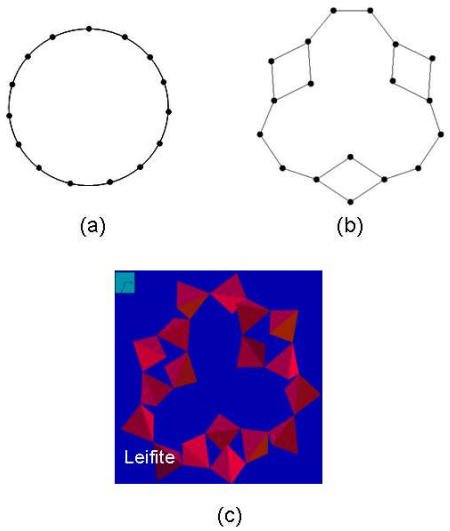

Figure 8: Exponential complexity.

$6 \times n$ nodes. Figure 8 c) demonstrates that this situation occurs in real crystal structures like the silicate leifite. There exist 315 non-translationally equivalent 15-membered rings of corner-sharing tetrahedra in this crystal (for a definition of rings in such crystal structures see [16]).

The number of mappings to consider can often be restricted by taking the symmetries of the target structure into account. It is sufficient to choose a single representative for each class of symmetrically equivalent substructures in the target structure. In leifite there are 44 different classes of 15-membered rings. Furthermore, when looking for embeddings given a representative of such a class only embeddings have to be considered which are not symmetrically equivalent. Consider again Figure 8. In the symmetry group (space-group) of leifite there is a threefold rotation axis in the center of the substructure shown in Figure 8 c). Let a concrete mapping of a 15-membered ring into the shown substructure be given. Then four additional mappings using the same tetrahedra of the substructure have to be considered instead of fourteen if no symmetry would be applicable.

3.2 Uniqueness criteria

To check two clusters for topological equivalence it is necessary to find for every polyhedron in one of the corresponding polyhedra graphs a suitable permutation of the vertices such that both graphs become identical (up to the geometric information). Though the number of permutations for a polyhedron to consider can be restricted by taking the rotation group into account, the resulting set can still be quite large. Therefore, all possibilities to restrict this set further should be applied. The following properties of a convex polyhedron P and an ordered face representation of it are helpful into that regard:

- Every edge of P belongs to exactly two faces.
- Every face of P has at least three edges.
- The representation of faces implies that every edge shows up as a pair (i, j) in one face and as (j, i) in the other face.

We get as an immediate consequence: Let P and P' be two polyhedra in ordered face representation. Let ϕ be an isomorphism between P and P' and $e = (i, j)$ an edge of P. Then there is exactly one permutation π of the vertices of P' such that the following holds: $(\pi(\phi(i)), \pi(\phi(j))) = (i, j)$ and P and P' have the same ordered face representation.

The uniqueness of the permutation π follows from the following observation: e and $\phi(e)$ shall have the same numbers; hence $\pi(\phi(i)) := i$ and $\pi(\phi(j)) := j$. There are exactly two faces F_1 and F_2 of P which are incident with e. Without loss of generality it can be assumed that e occurs in the representation of F_1 in the order i, j and in the representation of F_2 in the order j, i. For every sequence i, j there is exactly one representation of a face in the ordered faced representation of a polyhedron. This means that the number of every vertex h in $\phi(F_1)$ has to be mapped by π to the number of $\phi^{-1}(h)$ in order to get the same sequence of vertex numbers as it occurs in the representation of F_1. The same holds for $\phi(F_2)$. So π is uniquely determined for the vertices of $\phi(F_1)$ and $\phi(F_2)$. Since there are now further edges of P' determined, the permutation is uniquely fixed for all vertices of P'. The vertices of an edge being involved in a connection of a cluster of polyhedra must be numbered identically to the corresponding edge in a topologically equivalent cluster. Hence in case of edge- or face-sharing the ordered face representation guarantees that the vertex numbering in one of the corresponding poly-

hedra graphs determines uniquely the vertex numbering in the other graph. A similar situation is given when two polyhedra are linked by the two vertices of an edge e of another polyhedron P. The permutation of the polyhedron P' in the isomorphic graph is determined uniquely since the vertices of e and the corresponding edge in the isomorphic graph have to be identical.

Example: Consider the two chains of tetrahedra in Figure 9 a). They are topologically equivalent. A permutation π of P' of a topological isomorphism for the two chains must map 2 to 3 and 4 to 1. The edge $e = (3, 1)$ of P taken in the given order determines the face $(1, 2, 3)$ of P. The face in P' with $(2, 4)$ occuring in this order in its representation is $(2, 4, 1)$. We therefore get $\pi(1) = 2$ and $\pi(3) = 4$.

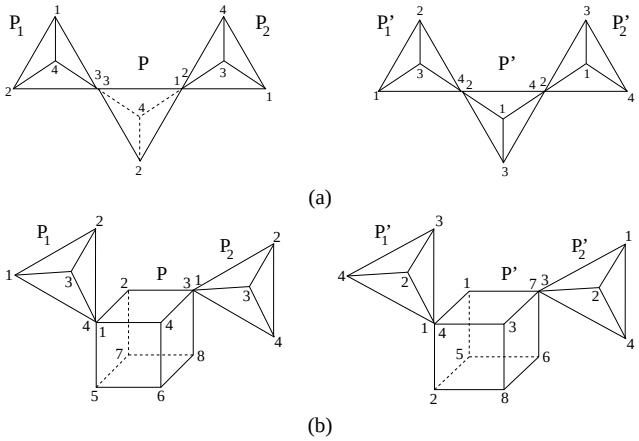

Figure 9: Connections fixing a permutation.

For the uniqueness of the permutation of a polyhedron P' it is even sufficient when the vertices of P involved in the connection are vertices of the same face F: The vertices of the corresponding face F' in P' are uniquely determined by the representation of F since no pair of vertices of a face which are not the endpoints of an edge can both be vertices of a further face. Otherwise, the second face would intersect F. Hence there is a single entry with both vertices in the representation of P. For an example, see Figure 9 b). Vertices 1 and 3 of P occur together only in face $(1, 2, 3, 4)$. From the connections follows that the vertices 4 and 7 of P' have to become 1 and 3, respectively. This implies the change of 1 to 2 and of 3 to 4.

In case the permutation of a polyhedron cannot be fixed by its neighbourhood, a representative of each class of permutations resulting in the same ordered face representation has to be considered. Take a single polyhedron, an octahedron, for example, in its regular form. There are 6! possible ways to label its vertices with different numbers. There are, however, much less ordered face representations of an octahedron, namely 30. This results from the properties of its symmetry group $m\bar{3}m$. It contains three perpendicular fourfold rotations which are automorphisms.

Coordination polyhedra may have distortions which can be used to restrict the set of geometrically reasonable permutations when only a single vertex is fixed by a connection. This is a consequence of missing symmetries (the topological view of polyhedra could be augmented by symmetry information such that it is not only sufficient to have identical face representation but the symmetry groups must fulfill conditions as well).

For a given cluster there can exist many isomorphic subgraphs in the same model graph. Hence the problem arises which ones to present in the result of a query. Clusters with isomorphic polyhedra graphs can have quite different geometries. We need a method to rank results which is fast to compute and which nonetheless is suited to measure the differences in the geometries of clusters since the number of qualifying model structures can be very large as well.

4 Geometric similarity of embeddings

Geometric similarity of three-dimensional structures has been defined in various different forms depending on the underlying applications. For the comparison of clusters of polyhedra we need a definition of similarity which takes into account that coordination polyhedra represent strong bonds and that connections between polyhedra are given by the sharing of common atoms.

4.1 The problem

For molecular structures, operations such as rotations, translations, and scaling as well as the editing of molecules including the deletion or insertion of atoms are sometimes applied to check for a possible matching [17]. Other approaches look for maximal subgraphs allowing distances of atoms to vary in fixed ranges [5] or identify backbone structures in order to find rigid motions for solving the matching problem for substructures [18].

For inorganic structures “the use of the term ’similar’, in the definition of configurational and crystal chemical isotypism, arises from the inherent difficulty in defining a priori limits on the similarity of geometrical configurations or physical/chemical characteristics” [19]. This means that flexibility has to be taken into account like for organic structures.

Since for inorganic structures the notion of coordination polyhedra is fundamental [8], it should be possible to define and analyze similarity at the level of polyhedral networks. A single coordination polyhedron characterizes the bonding structure for a distinguished central atom and neighbouring atoms. Mostly, these local bonds are strong compared to other bonds. The form of a polyhedron depends upon the

kind of the central atom, the number and kind of its ligands, and other atoms in the vicinity. In a typical silicate, for example, four oxygen atoms with strong bonds to a silicon atom form a regular tetrahedron. Sodium and chlorine atoms are forming regular octahedra in sodium chloride whereas the octahedra formed by lithium and oxygen atoms in some silicates are quite irregular.

Distortions of polyhedra from the perfect form can be measured [13], [20], [21], but it is also argued that no unique index can be defined for measuring the size of the distortion since measures are never completely model-free [22], [13]. Therefore, when comparing substructures of inorganic compounds at the level of coordination polyhedra, a distinction between the placement of polyhedra in the structure and the shape of polyhedra seems to be appropriate. The placement of a polyhedron can be described by the coordinates of its central atom. A difference in the coordinates of the central atoms of two corresponding polyhedra implies that the coordinates of the ligands are different as well or that there are differences in the shape of the two polyhedra.

The reference coordinate axes of two structures cannot be assumed to be identical. Hence rotations and translations are normally necessary to check for geometric isomorphism. The most restrictive definition of geometric similarity requires a bijection to exist between the two given point sets such that the coordinates of a point and its image in the other set are allowed to differ only in the range of a given small tolerance. This tolerance is necessary in connection with experimental data. A weakened form of this definition assumes polyhedra to be rigid bodies according to the fact that they model strong bonds. It considers vertex- and edge-sharing connections of polyhedra in clusters as 'ball-and-sockets' and 'hinges'. Whereas a face-sharing allows no flexibility, a vertex-sharing and an edge-sharing of two polyhedra allow the central atoms to move on the surface of a sphere and on circles, respectively, depending on the overall flexibility of the cluster. For two clusters it should be possible to apply appropriate motions such that the resulting clusters are geometrically similar according to the definition above. A distinction can be made whether motions with intermediate interpenetrating polyhedra are allowed or not.

Figure 10 shows an example where two structures become identical if an appropriate hinge motion is applied. It should be mentioned that these two rings of tetrahedra are normally considered as strictly distinct clusters.

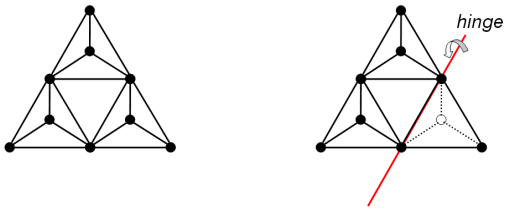

Figure 10: Identification by hinge motion.

4.2 Polyhedral clusters as joint structures

For the mathematical modeling of geometric embeddings of polyhedra graphs the algebra presented in [23] is well suited. It uses projective geometry as underlying theory and allows to deal elegantly with motions and geometric transformations. We first repeat some definitions of [23] (for an introduction into projective geometry see [24], for example).

Since we are only interested in embeddings into the Euclidean space it is sufficient to consider projective spaces over the real numbers.

Let $n \in \mathbb{N}$. For $x, y \in \mathbb{R}^{n+1} \setminus \{\mathbf{0}\}$ define the equivalence relation

$$x \sim y \qquad \text{:iff} \qquad \text{there is a } \lambda \in \mathbb{R} \text{ with } x = \lambda y.$$

The set of equivalence classes

$$\mathbb{PR}^n := (\mathbb{R}^{n+1} \setminus \{\mathbf{0}\})/\sim$$

is called *the projective space of dimension* n *over* $\mathbb{R}$; the elements of $\mathbb{PR}^n$ are called *projective points*.

Let $x \in \mathbb{PR}^n$ with $x = (x_0, \ldots, x_n)$. Then x and λx describe the same projective point for all $\lambda \neq 0$. The equivalence class of x is usually denoted as $(x_0 : \cdots : x_n)$ and $(x_0 : \cdots : x_n)$ are called *homogeneous coordinates* of the point x.

For $x \in \mathbb{R}^3$ with $x = (x_1, x_2, x_3)$ it is convenient to use the point

$$\bar{x} := (x, 1) := (x_1, x_2, x_3, 1) \in \mathbb{PR}^3$$

as representative of its equivalence class.

If $\varphi : M \longrightarrow \mathbb{R}^3$ is a mapping of an arbitrary set M into $\mathbb{R}^3$, then $\bar{\varphi}$ shall denote the continuation of φ on the homogeneous coordinates of the images, i.e.

$$\bar{\varphi}(m) := \overline{\varphi(m)}$$

for all $m \in M$.

For the description of all lines of $\mathbb{R}^3$ properties and notions of the Grassman-Cayley algebra are used. A general multiplication on the points of $\mathbb{PR}^n$ is defined as follows:

Let $a, b \in \mathbb{PR}^3$ with $a = (a_1, a_2, a_3, a_4)$ and $b = (b_1, b_2, b_3, b_4)$. Consider the matrix

$$D_{a,b} := \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix}.$$

The *outer product* is defined as

$$\vee : \mathbb{PR}^3 \times \mathbb{PR}^3 \longrightarrow \mathbb{R}^6,$$

$$a \vee b := (d_{14}, d_{24}, d_{34}, d_{23}, d_{31}, d_{12}),$$

where d_{ij} denotes the 2×2-minor obtained from the ith and jth column of $D_{a,b}$.
The 6-tupel $a \vee b$ is called 2-*extensor* of the points a and b; it is also written in the form ab.
The set of all such 6-tupels for pairs of points is denoted by $\mathbf{L}^2$.
Let $a, b \in \mathbb{R}^3$ and $\bar{a}$, $\bar{b}$ be the equivalent projective points. Then the following holds:
$\bar{a}\bar{b} = \bar{a} \vee \bar{b} = (a - b, a \times b)$.
$(a - b, a \times b)$ are called the *Plücker coordinates* of a and b. Not all elements of $\mathbb{R}^6$ correspond to pairs of points but every such 6-tuple represents a *screw* in $\mathbb{R}^3$. This results from the fact that every element of $\mathbb{R}^6$ can be obtained as sum of two 2-extensors. The set of all 6-tuples or screws is usually denoted by Λ^2.

Consider the polyhedra graph $G_{\mathcal{P}} = (N_{\mathcal{P}}, E_{\mathcal{P}}, \lambda)$ for a polyhedral cluster $\mathcal{P}$ together with the geometrical views of the polyhedra. The edges $E_{\mathcal{P}}$ of the graph can be partitioned according to their corresponding type (vertex, edge, or face):
$E_{\mathcal{P}} = E_v \dot{\cup} E_e \dot{\cup} E_f$.
Define the following mapping

$$\mathbf{H} : E_e \longrightarrow \mathbf{L}^2,$$

$$e = (P_i, P_j) \mapsto \overline{pos(L_{e_1})} \vee \overline{pos(L_{e_2})} =: L_{i,j},$$

where $\{L_{e_1}, L_{e_2}\} = V_{P_i} \cap V_{P_j}$ for every edge $(P_i, P_j) \in E_e$; V_{P_i}, V_{P_j} denote the sets of vertices of the polyhedra $P_i, P_j \in \mathcal{P}$, respectively.
Since for every edge $(P_i, P_j) \in E_e$ we have

$$L_{i,j} = -L_{j,i},$$

it follows that $\mathbf{H}$ is a hinge motion [25].

We are now ready to define when a polyhedral cluster is flexible and when two given polyhedra graphs can be considered as geometrically transformable into each other by motions which respect the connections of the polyhedra.

Definition: Let $G = (N_{\mathcal{P}}, E_{\mathcal{P}}, \lambda)$ be a polyhedra graph with geometrical views given for all polyhedra. Let

$$\mathbf{S} : N_{\mathcal{P}} \longrightarrow \Lambda^2, P_i \mapsto \mathbf{S}_i$$

be a mapping such that the following holds:

- for every connection by vertices $(P_i, P_j) \in E_v$ with $V_{P_i} \cap V_{P_j} = \{L\}$ we have

$$\mathbf{S}_i pos(L) = \mathbf{S}_j pos(L),$$

- for every connection by edges $(P_i, P_j) \in E_e$ there exists a scalar $\lambda_{i,j} \in \mathbb{R}$, such that

$$\mathbf{S}_i - \mathbf{S}_j = \lambda_{i,j} L_{i,j},$$

- for every connection by faces $(P_i, P_j) \in E_f$ we have

$$\mathbf{S}_i = \mathbf{S}_j.$$

$\mathbf{S}$ is a *joint motion* of the configuration induced by pos. $\mathbf{S}$ is called *trivial* if $\mathbf{S} \equiv D$ for some $D \in \Lambda^2$.
G is called *flexible* if it admits a non-trivial joint motion $\mathbf{S}$.

Since the geometrical views of the polyhedra of a polyhedra graph determine a configuration of the graph, it is possible to define the congruency of two polyhedra graphs in analogy to general graphs. Based on this definition, conditions for the possibility to transform the graphs into each other by motions can be given.

Definition: Let $G_1 = (N_1, E_1, \lambda_1)$ and $G_2 = (N_2, E_2, \lambda_2)$ be two polyhedra graphs with functions pos_1 and pos_2 assigning the coordinates to the vertices of the polyhedra of G_1 and G_2, respectively.
G_1 und G_2 are *congruent* if there exists a topological isomorphism $\varphi : G_1 \longrightarrow G_2$ and a $\mathbb{R}^3$-isometry T such that

$$T \circ pos_1 = pos_2 \circ \varphi.$$

Let $SE(3)$ denote the set of proper rigid motions of $\mathbb{R}^3$. Let $\mathbf{M} : \Lambda^2 \longrightarrow SE(3)$ be a mapping assigning the homogeneous representation to every screw. If a motion $\mathbf{S}$ of G_1 exists such that G_1 after applying the (homogeneous representation) of the motion $\mathbf{S}$ and G_2 are congruent with respect to pos_1 and pos_2, i.e., if there exists an $\mathbb{R}^4$-isometry T such that for all $n_i \in N_1$

$$(T \circ \mathbf{M}(\mathbf{S}_i) \circ \overline{pos_1})(n_i) = (\overline{pos_2} \circ \varphi)(n_i)$$

holds, then $\mathbf{S}$ is called a *geometric transformation* of G_1 into G_2 and G_1 is called *geometrically transformable into G_2 by joint motions*.

This definition fixes when two polyhedra graphs can be transformed into each other geometrically. The underlying assumptions are that polyhedra are rigid bodies and that only joint motions can be applied.

Consider a set of topologically isomorphic polyhedra graphs. Assume that corresponding polyhedra in the graphs are congruent. The question arises whether for any pair of graphs (G_1, G_2) in the set, G_1 is geometrically transformable into G_2 by joint motions. Look again at the rings of polyhedra in Figure 6. Applying hinge motions and moving the pyramids through one another allows to transform one ring into the other. However, if we add two handles to each of the rings with edge-sharing pyramids and with connections to the ring at opposite pyramids again by sharing edges (a kind of 'crown' is built), the resulting clusters are topologically equivalent but no geometric transformation is possible. The reason for the missing of a transformation is that the pyramids of the handles as well as those of the ring

have to be moved through one another in order to change the direction of their apices. Applying such a motion to one of the handles results in a squeezing of the other handle since each handle and the ring can only be moved perpendicular to their hinges. This follows from the fact that the complete structure has only connections by edges.

Two non-flexible topologically isomorphic clusters which are not geometrically transformable into one another can be constructed as follows: Take a cube and place a pyramid on each of its faces; remove the cube from the cluster. The resulting configuration of edge-sharing pyramids is topologically equivalent to the cluster which we obtain when the apices of the pyramides are all pushed in. Since the two clusters are not flexible, no geometric transformation is possible.

The notion of geometric transformation can be weakened in order to get a definition of similarity which takes into account that in crystal structures we do not deal with ideal polyhedra. One idea is to restrict the pos-functions in the definition of a geometric transformation to vertices involved in connections and to allow differences in the coordinates of corresponding vertices up to some fixed limit. A second possibility is to restrict these functions to the central atoms of the polyhedra with the same relaxation with respect to the coincidence of coordinates.

When two structures are geometrically similar to one another according to one of these definitions they can be further analysed for the similarity of corresponding polyhedra. Symmetries can be checked, for example, the kinds of atoms can be compared, or measures of distortions can be used. In the following, we concentrate on the first step. Furthermore, we do not consider non-trivial joint motions to change the geometry of structures. We rather use the root mean square method to measure the 'distance' of two structures with respect to the positions of the central atoms of their polyhedra.

5 Implementation and results

The problem of minimization of distances for two sets of points in three-dimensional space is sometimes called 'problem of absolute orientation'. Two well-known methods for solving this problem have been investigated: the algorithm of B.K.P. Horn [26], a closed-form solution of the problem using unit quaternions to represent rotations, and the algorithm of W.A. Dollase [27] using infinitesimal rotations. We have given preference to the algorithm of Horn since it avoids the potentiation of numeric instabilities, which may arise in connection with orthonormalization in the algorithm of Dollase.

Let $G_{\mathcal{P}} = (N_{\mathcal{P}}, E_{\mathcal{P}}, \lambda_{\mathcal{P}})$ and $G_{\mathcal{P}'} = (N_{\mathcal{P}'}, E_{\mathcal{P}'}, \lambda_{\mathcal{P}'})$ be two finite polyhedra graphs with $N_{\mathcal{P}} = \{P_1, \ldots, P_n\}$ and $N_{\mathcal{P}'} = \{P'_1, \ldots, P'_n\}$.

Let $\varphi : \mathcal{P} \longrightarrow \mathcal{P}', P_i \mapsto P'_i$, be an isomorphism and let $C_{\mathcal{P}} = \{c_1, \ldots, c_n\}$ and $C_{\mathcal{P}'} = \{c'_1, \ldots, c'_n\}$ be the sets of coordinates of the central atoms of $\mathcal{P}$ and $\mathcal{P}'$, respectively (referring to a common Cartesian coordinate system), i.e. $c_i = pos_{\mathcal{P}}(P_i)$ and $c'_i = pos_{\mathcal{P}'}(P'_i)$, for $i = 1, \ldots, n$. Here pos is the function from polyhedra to the coordinates of their central atoms derived from the geometrical views of the polyhedra.
φ induces a mapping

$$\psi : C_{\mathcal{P}} \longrightarrow C_{\mathcal{P}'},$$

$$pos_{\mathcal{P}}(c_P) = c_i \mapsto c'_i = pos_{\mathcal{P}'}(c_{\varphi(P)})$$

that sends the coordinates of the central atom of every polyhedron of $\mathcal{P}$ to the coordinates of the central atom of its image under φ.

The sets $C_{\mathcal{P}}$ and $C_{\mathcal{P}'}$ are now considered as rigid subsets of $\mathbb{R}^3$, which shall be moved such that the best possible matching results. This means that a motion $T \in SE(3)$ is looked for solving the following least-squares problem:

$$U := \sum_{P \in N_{\mathcal{P}}} ||(pos_{\mathcal{P}'} \circ \varphi)(P) - (T \circ pos_{\mathcal{P}})(P)||_2^2 = \min .$$

With the notations from above we get the following equivalent formulation:

$$U := \sum_{i=1}^{n} ||c'_i - T(c_i)||_2^2 = \min .$$

An optimal rigid Euclidean motion T moves the centroid of $C_{\mathcal{P}}$ to the centroid of $C_{\mathcal{P}'}$ [26]. By placing centroids into the origin of the common coordinate system, the least-squares problem reduces to the determination of a rotation T solving the equation above for $C_{\mathcal{P}}$ and $C_{\mathcal{P}'}$ relative to their centroids.

For measuring the similarity of two structures $\mathcal{P}$ and $\mathcal{P}'$ the root mean square

$$\epsilon := \frac{\sqrt{U}}{n}$$

is taken. The similarity increases with decreasing ϵ. What we have to keep in mind is that polyhedra distortions and rotations allowed by their connections may lead to differences in the shape of otherwise similar clusters.

For our problem it is sufficient to consider quaternions as unit vectors of $\mathbb{R}^4$. Each unit quaternion uniquely represents a rotation in three-dimensional space. Furthermore, scaling of point sets is not wanted since motions shall be rigid. Therefore, we assume the scaling factor in the algorithm of Horn to be 1.
Algorithm:
Input: Two finite topologically isomorphic clusters of polyhedra C_1 and C_2 with sets of polyhedra $\{P_1, ..., P_n\}$ and

$\{P'_1, ..., P'_n\}$, respectively, and a topological isomorphism φ such that $P'_i = \varphi(P_i)$, for $i = 1, ..., n$.

Output: Root mean square of the deviation of C_1 and C_2 after an optimal distance minimizing movement of the central atoms according to φ in a common coordinate system.

// `Initialization`

Transform the coordinates of all atoms of C_1 and C_2 into Cartesian coordinates; generate the sets $A_{C_1} = \{c_1, ..., c_n\}$ and $A_{C_2} = \{c'_1, ..., c'_n\}$ of coordinates of the central atoms of C_1 and C_2, resp.

Centralize A_{C_1} and A_{C_2}:

$\bar{c} := \frac{1}{n} \sum_{i=1}^{n} c_i$;

$\bar{c'} := \frac{1}{n} \sum_{i=1}^{n} c'_i$;

for $i = 1, ..., n$ **do**

$c_i := c_i - \bar{c}$;

$c'_i := c'_i - \bar{c'}$;

end

// `Solution of the balancing problem`

Compute with the algorithm of Horn the rotation matrix R_q, which moves A_{C_1} optimally on A_{C_2}.

// `Computation of the result and parameter`

// `extraction (rotation angle` Θ_{res} `and axis`

// $(l_{res}, m_{res}, n_{res}) \in \mathbb{R}^3$)

$U := \sum_{i=1}^{n} \|c'_i - R_q(c_i)\|_2^2$;

$\epsilon := \frac{\sqrt{U}}{n}$;

Compute $\Theta_{res}, l_{res}, m_{res}, n_{res}$ given $R_{res} = R_q$.

The implementation of the algorithm in C++ has been integrated into our system POLYSEARCH [6]. A graphical interface of POLYSEARCH allows to mark a substructure of a chosen crystal for search. Topologically equivalent substructures in a set of model structures are determined based upon a representation of the crystals by periodic graphs. When building the result, the number of substructures of the same crystal structure being geometrically equivalent to each other is reduced by exploiting the symmetries of the structure. The information on the subgraphs corresponding to the remaining substructures is used to determine the ranking, which can be done fast (cp. [28]). The execution time of the search for equivalent substructures mainly depends on the selectivity of the search structure relative to the given set of model structures [6]. The extension of our web application [9] by an appropriate graphical interface for the representation of the geometric similarity of substructures is in progress.

Figure 11 shows screenshots of a search structure in the silicate aminoffite and three substructures of the result with their root mean square values (RMS). In this example, the structures consist of tetrahedra with nearly identical geometry. The different RMS values mainly result from the differences in the orientation of the tetrahedra in the clusters.

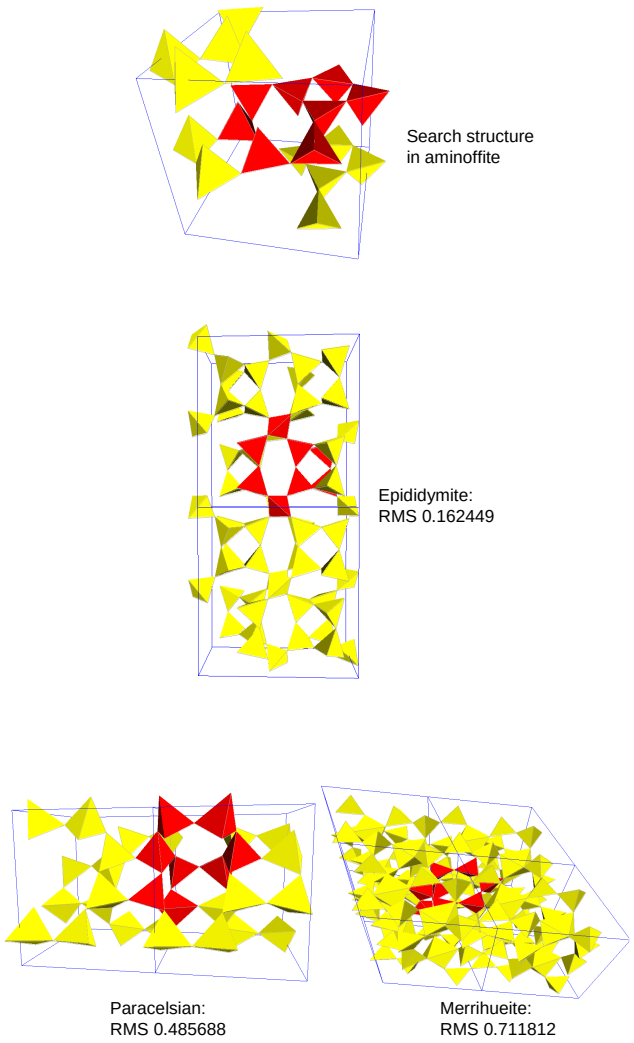

Figure 11: Search structure and three similar substructures.

6 Conclusions

In this paper, the main focus has been to show how geometric similarity of clusters of coordination polyhedra can be defined and computed by an appropriate algorithm. The integration of the algorithm into our system for the retrieval of isomorphic substructures in large sets of model structures allows to search for geometrically similar substructures in large inorganic crystal structure databases. The search can be combined with other information about the target structure such as publication data, symmetries, or the assignment to a special class.

A further application can be seen in the field of structure prediction. In recent years, the enumeration of hypothetical inorganic structures has attracted much attention [29], [30].

Geometric embeddings are computed for graphs which are generated in an enumeration process. It is very helpful to analyse these hypothetical structures with respect to their similarity with real structures.

As mentioned above, our system only checks for the similarity of point sets built from the coordinates of central atoms. In the future, we intend to implement methods for taking all information about polyhedra into account and to provide information about motions, which can be applied to get a better matching of two topologically isomorphic structures. A further idea is to apply methods developed in the field of maximal subgraph search in order to be able to compare structures without referring to a target substructure.

Recently, a graphical interface has been implemented in POLYSEARCH offering operations for constructing polyhedral clusters artificially. A collection of different kinds of polyhedra as often found in crystal structures is available and methods for constructing clusters and modifying them by applying joint motions have been realized. For the generated clusters, embeddings into polyhedral networks of real crystal structures can be determined and ranked.

References

[1] H.-J. Klein and Ch. Mennerich. *Searching and ranking similar clusters of polyhedra in crystal structures*, Proc. 2nd Int. Conf. ADVCOMP, Valencia, Spain, IEEE, pp. 235-240, 2008

[2] F.H. Allen. *The Cambridge Structural Database: a quarter of a million crystal structures and rising*, Acta Cryst. B58, pp. 380-388, 2002

[3] A. Belski, M. Hellenbrandt, V.L. Karen, and P. Liksch. *New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design*, Acta Cryst. B58, pp. 364-369, 2002

[4] J.A. Chisholm and S. Motherwell. *A new algorithm for performing three-dimensional searches of the Cambridge Structural Database*, J. of Appl. Cryst. 37, pp. 331-334, 2004

[5] J.W. Raymond and P. Willett. *Similarity searching in databases of flexible 3D structures using smoothed bounded distance matrices*, J. Chem. Inf. Comput. Sci. 43, pp. 908-916, 2003

[6] H.-J. Klein. *Retrieval of isomorphic substructures in crystallographic databases*, Proc. 16th Int. Conf. SSDBM, Santorini, Greece, IEEE, pp. 255-264, 2004

[7] M. Kuramochi and G. Karypis. *Discovering frequent geometric subgraphs*, Inform. Systems 32, pp. 1101-1120, 2007

[8] M. O'Keeffe and B.G. Hyde. *Crystal Structures: 1. Patterns and Symmetry*, Mineralogical Society of America, 1996

[9] http://www.is.informatik.uni-kiel.de/∼hjk/crystana

[10] D.R. Chand and S.S. Kapur. *An algorithm for convex polytopes*, J. of the ACM 17(1), pp. 78-86, 1970

[11] E. Prince (Ed.). *International Tables for Crystallography - Vol. C*, Part 9, online edition, Springer, 2006

[12] P. Villars and K. Cenzual. *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds*, Release 2008/9, ASM International, 2008

[13] E. Makovicky and T. Balić-Zunić. *New measure of distortion for coordination polyhedra*, Acta Cryst. B54, pp. 766-773, 1998

[14] D. Abelson, S.-H. Hong, and D.E. Taylor. *Geometric automorphism groups of graphs*, Discr. Appl. Mathematics 155, pp. 2211-2226, 2007

[15] A. Lubiw. *Some NP-complete problems similar to graph isomorphism*, SIAM J. Comput. 10(1), pp. 11-21, 1981

[16] K. Goetzke and H.-J. Klein. *Properties and efficient algorithmic determination of rings in finite and infinite polyhedral networks*, J. of Non-Crystalline Solids 127, pp. 215-220, 1991

[17] X. Wang and J.T.L. Wang. *Fast similarity search in three-dimensional structure databases*, J. Chem. Inf. Comput. Sci. 40, pp. 442-451, 2000

[18] L.P. Chew, D. Huttenlocher, K. Kedem, and J. Kleinberg. *Fast detection of common geometric substructure in proteins*, Proc. 3rd ACM RECOMB, pp. 104-112, 1999

[19] J. Lima-de-Faria et al. *Nomenclature of inorganic structure types*, Acta Cryst. A46, pp. 1-11, 1990

[20] T.B. Balić-Zunić and E. Makovicky. *Determination of the centroid or 'the best centre' of a coordination polyhedron*, Acta Cryst. B52, pp. 78-81, 1996

[21] E. Lalik. *Shannon information as a measure of distortion in coordination polyhedra*, Appl. Cryst. 38, pp. 152-157, 2005

[22] I.D. Brown. *On measuring the size of distortions in coordination polyhedra*, Acta Cryst. B62, pp. 692-694, 2006

[23] W. Whiteley. *Rigidity of molecular structures: generic and geometric analysis*, in: P. Duxbury and M. Thorpe (eds.), *Rigidity and Applications*, Kluwer, 1999

[24] H.S.M. Coxeter. *Projective Geometry*, 2nd ed., Springer, 2003

[25] H. Crapo and W. Whiteley. *Statics of frameworks and motions of panel structures, a perspective geometric introduction*, Structural Topology 6, pp. 43-82, 1982

[26] B.K.P. Horn. *Closed-form solution of absolute orientation using unit quaternions*, J. Optical Soc. of America A4, pp. 629-642, 1987

[27] W.A. Dollase. *A method on determining the distortion of coordination polyhedra*, Acta Cryst. A30, pp. 513-517, 1974

[28] A. Lorusso, D.W. Eggert, and R.B. Fisher. *A comparison of four algorithms for estimating 3-D rigid transformations*, Proc. Brit. Conf. on Machine Vision, pp. 237-246, 1995

[29] H.-J. Klein. *Systematic generation of models for crystal structures*, Proc. 10th Int. Conf on Mathem. and Comp. Modelling and Scientific Comp., Boston, 1995, in: Mathematical Modelling and Scientific Computing 6, 325-330, 1996

[30] M.D. Foster, M.M.J. Treacy. *A database of hypothetical zeolite structures*, http://www.hypotheticalzeolites.net