Molecular assembly indices of mineral heteropolyanions: some abiotic molecules are as complex as large biomolecules

Robert M. Hazen1, Peter C. Burns2,3, H. James Cleaves II1,4,5,6, Robert T. Downs7, Sergey V. Krivovichev8,9 and Michael L. Wong1,10

1Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA
2Department of Civil and Environmental Engineering and Earth Sciences, and 3Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA
3Earth Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan
4Blue Marble Space Institute for Science, Seattle, WA 98104, USA
5Department of Chemistry, Howard University, Washington, DC 20059, USA
6Geological Sciences, University of Arizona, Tucson, AZ 85721, USA
7Department of Crystallography, Institute of Earth Sciences, St. Petersburg State University, St. Petersburg 199034, Russia
8Nanomaterials Research Centre, Kola Science Centre, Russian Academy of Sciences, Fersmanma 14, Apatity 184209, Russia
9Sagan Fellow, NASA Hubble Fellowship Program, Space Telescope Science Institute, Baltimore, MD 21218, USA

Molecular assembly indices, which measure the number of unique sequential steps theoretically required to construct a three-dimensional molecule from its constituent atomic bonds, have been proposed as potential biosignatures. A central hypothesis of assembly theory is that any molecule with an assembly index \( \geq 15 \) found in significant local concentrations represents an unambiguous sign of life. We show that abiotic molecule-like heteropolyanions, which assemble in aqueous solution as precursors to some mineral crystals, range in molecular assembly indices from 2 for \( \text{H}_2\text{CO}_3 \) or \( \text{Si(OH)}_4 \) groups to as large as 21 for the most complex known molecule-like subunits in the rare minerals ewingite and ilmajokite. Therefore, values of molecular assembly indices \( \geq 15 \) do not represent unambiguous biosignatures.

1. Introduction
Cronin, Walker, and colleagues [1–8] have championed the assembly theory (AT) approach to characterizing complex evolving systems. In their framework, evolving systems are made from objects (e.g. molecules; automobiles) that must be assembled through a series of steps from smaller building units (e.g. chemical bonds; automobile parts). Furthermore, assembly can be imagined to occur through a minimum number of discrete steps called the ‘assembly index’. In the AT model, the products of living systems, whether biomolecules or automobiles, are distinguished from nonliving objects by two coexisting criteria: (1) the object requires numerous assembly steps and (2) the object is present in numerous copies—in ‘high abundance’. Only a directed assembly by life can overcome the combinatorial improbability of objects as complex as ribosomes or Bentleys occurring multiple times.

AT has not gone unchallenged. Some reviewers question whether a scalar assembly index can be employed to adequately discriminate between living and nonliving systems [9,10]. Other critics note similarities of the AT approach to uncited prior efforts to distinguish biotic from abiotic molecular suites by statistical or algorithmic measures [11]. In particular, Hernández-Orozco and colleagues [12] may have anticipated key conclusions of assembly theory by exploring connections among causal memory, selection, and evolution. This
hypothesis has also received criticism based on a variety of concerns, including ambiguities in the numbers of molecular copies that constitute ‘high abundance’, the optimal algorithm to calculate ‘pathway complexity’, the disconnect between the proposed theoretical assembly pathways and actual chemical processes, and the absence of kinetic and thermodynamic factors in assessing the probability for a complex molecule’s formation [9,10,13].

In the context of their ambitious research programme, the AT group has developed molecular assembly (MA) indices to characterize the number of unique chemical bonding steps necessary to build discrete molecules from their constituent atoms. Such indices range from 1 for the simplest molecules, to more than 15 for many molecules in cells, to greater than 30 for the most complex biomolecules [2–7]. Cronin, Walker, and colleagues conclude that any organic molecule that occurs abundantly (defined variously as being present in greater than a million copies or producing measurable peaks in mass spectrometric analysis) and that has MA index ≥15 is a uniquely biological signature. Such complex molecules, they explain, represent combinatoric rarities that are too unlikely to be repeatedly synthesized without highly controlled cellular processes. They extend this idea to postulate that only living systems can produce molecular structures of such complexity.

Here we test this quantitative AT hypothesis by examining some of the most complex abiotic molecule-like heteropolyanions (also known as polyoxometalates in oxide minerals) that are building blocks of inorganic mineral structures to test the intriguing conjecture that ‘One thing that discriminates living things from inanimate matter is their ability to generate similarly complex or non-random structures in a large abundance’ [1]. Specifically, Marshall et al. [3] propose that there exists a ‘threshold abiotic-biotic divide,... which can be used to unambiguously assign complex objects as biosignatures’. In this context, structural complexity is measured by ‘pathway complexity’ or ‘molecular assembly index’ — a numerical value that is defined by the number of unique steps required ‘to assemble a given object by allowing the object to be dissected into a set of basic building units and rebuilding the object using those units’. A complication arises from the use of two similar yet differing published protocols for calculating MA indices. One from Marshall et al. [2–4], called the ‘split-branch algorithm’, is simpler to implement and provides an ‘upper bound’ on the MA index of a molecule. The other more rigorous approach (‘exact MA’, which can be derived via Monte Carlo methods) is described by Cronin, Walker, and colleagues [5,6], who provide a computer program to calculate minimum MA indices [7]. In both instances, the calculation of MA indices relies on the sequential formation of links between ‘basic building units’, which are defined as chemical bonds between two atoms other than hydrogen. Typical basic building units in organic molecules include (C–C), (C=O), and (C–N).

In the case of the split-branch algorithm [2, fig. 2A] (see also [4]), each bond added to an initial unit is counted as a step; the MA index depends entirely on the number of ‘joining operations’. Thus, forming (C–C–C) in tryptophan or (N=C–N) in ATP requires only 2 assembly steps, for example to join (C–C) with (C–C), while assembly of a (C,N) group in asparagine or (CO2) carbonate group requires only 2 steps. Furthermore, assembly of (SiO2) requires only 2 steps in the exact MA algorithm: (Si–O) → 2 → (SiO2); then (SiO2) × 2 → (SiO3). Employing the exact MA method, the resulting assembly indices for asparagine and tryptophan are reduced from 7 and 12 [2] to 6 and 11, respectively [7], demonstrating that the split-branch index is usually greater than exact MA index [4,7].

In the exact MA approach [5–7], the assembly index of any molecule begins with an already formed bond; no initial step is required to join two non-H atoms (e.g. ethane (H2C2H6) or carbon monoxide (CO)). Consequently, the assembly index for these small molecules is 0, whereas a variety of tri-atomic molecules, including (CO2) and (HCN), and C2 hydrocarbons (propane, propene) have exact assembly indices of 1 [5–7]. Addition of other building units leads to progressively greater MA indices, for example 3 for cyclopentane (C5H10), 4 for glycine (H2NCH2COOH), and 5 for alanine (H2NCH(2CH3)COOH).

For larger organic molecules, which often incorporate a combination of chains, branches, and/or rings in an open configuration, assembly may proceed from one region of the molecule to the next by the addition of new bonds or small molecular subunits one at a time. Once a subunit has been synthesized, that unit can be used repeatedly in the subsequent assembly process without requiring additional steps. Cronin and colleagues [5,6] provide several examples of exact MA index calculations up to MA index ≥15, as well as a program to calculate exact MA indices for a wide range of organic molecular species [7].

A difficulty remains. While exact MA index calculations are presented in [5–7], the proposed abiotic/biotic boundary of MA index ≥15 was based on the split-branch MA indices of Marshall et al. [2]. Because the Marshall et al. method usually overestimates MA indices by at least 1 or 2 steps, it seems probable that the abiotic/biotic boundary for exact MA indices may be 13 or 14, rather than ≥15. Because of this ambiguity, in the following section we calculate both the split-branch and exact assembly indices of molecule-like mineral substructures.

Mineral structures, in contrast to most organic molecules, tend to be dense arrangements of repeating small atom clusters (i.e. cation polyhedra such as (SiO2) tetrahedra or (MgO6) octahedra) that may share corners, edges, and/or faces with adjacent polyhedra. Thousands of three-dimensional atomic arrangements in minerals have been documented [15], some of which are exceedingly complex with as many as a dozen essential elements and hundreds of symmetrically distinct atom sites in the asymmetric unit. Many pathways will exist to assemble any such three-dimensional structure starting with individual atoms, though an assembly pathway of shortest length must exist. Nevertheless, according to the conventions of assembly
theory, which only relates to monodispersed and countable copies of a molecule, neither mineral crystals nor polymers are considered no matter how complex the structures might be [8]. Therefore, in this contribution we consider only polyatomic inorganic clusters that are assumed to be aqueous precursors to minerals that directly incorporate such substructures from solution [16–18]. These minerals feature clusters of atoms in which cations of oxidation states 2+ or greater form well-defined substructures that are linked internally by relatively strong bonds. By contrast, adjacent clusters link to each other through relatively weak bonds, such as [(Na,K)-O] or hydrogen bonds.

Mineralogical glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>asymmetric unit</td>
<td>the fraction of the unit cell that, when transformed by symmetry operators, generates the complete unit cell. The asymmetric unit of minerals contains from one to more than 200 distinct atomic sites.</td>
</tr>
<tr>
<td>essential element</td>
<td>an essential element is a chemical element that is required by the International Mineralogical Association’s (IMA’s) definition of the mineral species.</td>
</tr>
<tr>
<td>heteropolyanion</td>
<td>a strongly bonded cluster of atoms with two or more kinds of cations in relatively high oxidation state bonded to oxygen atoms; a synonym of polyoxometalates.</td>
</tr>
<tr>
<td>polyhedron</td>
<td>a small atom cluster with a central cation or anion surrounded by several nearest neighbour atoms of the opposite charge, e.g. SiO4^{4-}.</td>
</tr>
<tr>
<td>polyoxometalates</td>
<td>polyoxometalates are nanoscale metal-oxide clusters that spontaneously assemble in aqueous solution; a synonym of heteropolyanion.</td>
</tr>
<tr>
<td>mineral species</td>
<td>one of approximately 6000 mineral kinds approved by the IMA’s Commission on New Minerals, Nomenclature, and Classification. Each species has a unique combination of idealized composition and crystal structure.</td>
</tr>
<tr>
<td>solid solution</td>
<td>a solid solution occurs when two or more elements occupy the same crystallographic site in varying ratios. Common examples of solid solution include (Mg-Fe^{2+}), (Al-Fe^{3+}), and (OH-F).</td>
</tr>
<tr>
<td>unit cell</td>
<td>the parallelepiped-shaped volume that repeats in the crystal structure.</td>
</tr>
<tr>
<td>Wyckoff site</td>
<td>a crystallographically unique site in the unit cell.</td>
</tr>
</tbody>
</table>

2. Molecular assembly indices for selected mineral heteropolyanions

The International Mineralogical Association’s Commission on New Minerals, Nomenclature, and Classification (IMA-CNMNC) recognizes approximately 6000 different mineral ‘species’, each defined by its unique combination of idealized chemistry and crystal structure [15]. The majority of these mineral structures are ‘space-filling’ arrangements of atoms, and thus do not feature distinct molecule-like regions amenable to assembly index calculations. Well-known examples include such high-symmetry phases as native iron (Fe), halite (NaCl), and spinel (MgAl2O4). Other minerals, most notably silicates such as quartz (SiO2), alkali feldspars [Na, KAlSi3O8], and scores of framework zeolite group phases, are highly polymerized three-dimensional arrays of corner-linked [SiO4] tetrahedra. According to conventions of molecular assembly theory, these repeating three-dimensional structures have undefined molecular assembly indices [8]. Therefore, we focus exclusively on minerals that incorporate convincing molecule-like subunits that are thought to form from aqueous precursors.

Scores of complex mineral structures feature strongly bonded polyatomic subunits that are molecular in character, that likely assemble in an aqueous or silicate melt environment prior to crystallization, and for which a molecular assembly index can be calculated unambiguously. In some cases, direct evidence exists supporting a mechanism of molecular assembly in solution prior to crystallization [16–18]. In table 1, we provide examples of minerals with such molecule-like subunits with calculated assembly indices for these subunits based on both the split-branch protocols of Marshall et al. [2–4] and the more recent ‘exact’ protocols [5–7]. We also list the chemical metric of the number of different essential chemical elements—i.e. an element that is required in the IMA’s definition of the mineral species, as well as a structural complexity metric (see below).

The simplest ‘molecules’ in minerals are polyhedra of high-field-strength cations (typically oxidation state 3+ or greater) bonded to oxygen atoms. These rigid molecule-like subunits are linked by 1+ or 2+ cations and/or hydrogen bonds in interstitial regions. Common examples of simple, rigid polyhedra include the (CO3)2− carbonate group, the (BO3)3− borate group, the (SiO4)4+ orthosilicate group, and the (SO4)2− sulfate group. In all cases, we calculate assembly indices with bonds between two non-hydrogen atoms as building blocks; however, the MA indices differ depending on which of two published protocols are employed. Using the Marshall et al. split-branch procedures [2, fig. 2a], (CO3) and (SiO2) each require 3 assembly steps. Alternatively, in the Liu et al. [4] exact approach both (CO3) and (SiO2) require only 2 assembly steps. In the case of (CO3), first combine two (C–O) bonds to form (CO2); then add a third (C–O) bond. Similarly, in the exact MA approach the formation of (SiO2) requires 2 steps: 2 (Si–O) bonds form (SiO2); then 2 (SiO2) groups combine into (SiO2)4. It is important to note that the bonds, not the atoms, are combined in these operations.

Silicate tetrahedra often link together to form larger clusters, including in various disilicates (Si2O5)3− and cyclosilicates with three-member (Si3O9)6− and six-member (Si4O12)8− rings. In the split-branch approach these molecular subunits
have assembly indices of 5, 5, and 6, respectively. Alternatively, in the exact approach the indices are 4, 4, and 5, respectively.

Additional complexity arises when silicate clusters combine with other elements. In beryl \([\text{Be}_3\text{Al}_2\text{Si}_6\text{O}_{18}]\), for example, 6-member \((\text{Si}_6\text{O}_{18})^{12-}\) rings are cross-linked by six corner-sharing \(\text{AlO}_6\) octahedra to form a three-dimensional framework. Several examples will clarify molecular assembly calculations.

### 2.1. Vesuvianite

Among the most common minerals incorporating disilicate groups is the hydrous calcium-aluminium silicate vesuvianite \([\text{Ca}_9\text{Na}_1\text{Al}_2\text{Mg}_5\text{Fe}_2\text{Si}_{22}\text{O}_{72}(\text{OH},\text{F},\text{O})_{10}]\), which occurs abundantly in Ca-rich metamorphic rocks [15]. A molecule-like \([(\text{Si}_2\text{O}_7)(\text{AlO}_4)(\text{Si}_4\text{O}_{10})^{15-}]^{-}\) subunit of vesuvianite exemplifies the steps of a molecular assembly using both split-branch (SBMA) and exact (EMA) approaches:

\[
\begin{align*}
\text{(Si–O)} × 2 & → (\text{Si}_2\text{O}_7) & \text{SBMA} = 2 \text{ steps} & \text{EMA} = 1 \text{ step} \\
\text{(Si–O)} + (\text{Si}_2\text{O}_7) & → (\text{Si}_4\text{O}_{10}) & 1 \text{ step} & 1 \text{ step} \\
(\text{Al–O}) × 2 & → (\text{AlO}_4) & 2 \text{ steps} & 1 \text{ step} \\
(\text{AlO}_4) × 2 & → (\text{AlO}_4) & 1 \text{ step} & 1 \text{ step} \\
(\text{AlO}_4) + (\text{Si}_2\text{O}_7) & → [(\text{Si}_2\text{O}_7)(\text{AlO}_4)] & 1 \text{ step} & 1 \text{ step} \\
[(\text{Si}_2\text{O}_7)(\text{AlO}_4)] + (\text{Si}_2\text{O}_7) & → (\text{Si}_4\text{O}_{10}) & 1 \text{ step} & 1 \text{ step} \\
\text{TOTAL} & & \text{SBMA} = 10 \text{ steps} & \text{EMA} = 8 \text{ steps}
\end{align*}
\]

Silicate groups bond via corner sharing \((\text{AlO}_4)^{12-}\) octahedra:

Thus, the assembly index of this \([(\text{Si}_2\text{O}_7)(\text{AlO}_4)(\text{Si}_4\text{O}_{10})^{15-}]^{-}\) molecule-like subunit is either 8 or 10 steps, depending on the MA index calculation method used—values significantly below the abiotic/biotic divide of Marshall et al. [3]. In the following sections, we consider six more complex structures containing molecule-like subunits with MA index \(≥ 10\).

### 2.2. Tourmaline group

Cyclosilicate structures often feature molecule-like subunits that extend beyond the silicate ring. The varied tourmaline group, with at least 38 species that form in both igneous and metamorphic environments [15], has a generalized formula of \(\text{XY}_2\text{Z}_n\text{TO}_{3n}(\text{BO}_3)_3\text{V}_2\text{W}\) with typical constituent elements including \(X = \text{Na}, \text{K}, \text{or} \text{Ca}; Y = \text{Si}, \text{Al}, \text{or} \text{Fe}; M_{\text{g}}, \text{T} = \text{Si}, \text{Al}; V = (\text{OH}), \text{O}; \text{and} W = (\text{OH}), \text{F}, \text{O}\). Note that the substitution of multiple elements in a given site via ‘solid solution’, while increasing a specimen’s chemical complexity, does not increase the assembly index. In the species fluor-uvite \([\text{CaMg}_3\text{Al}_5\text{Mg}(\text{Si}_6\text{O}_{18})(\text{BO}_3)_3(\text{OH})_3\text{F}]\), for example, one can consider a ‘structural island’ [14] consisting of a \((\text{Si}_6\text{O}_{18})^{12-}\) ring bonded to a group of three edge-sharing MgO octahedra in a \((\text{Mg}_3\text{O}_{12})^{18-}\) cluster and three planar \((\text{BO}_3)_3^{12-}\) molecular subunits (figure 1). This molecule-like group with composition \((\text{Mg}_2\text{Si}_3\text{O}_{10})^{33-}\) can be constructed in a minimum of 11 steps (figure 2). In this example, the minimum assembly pathway is not immediately obvious. Rather than first assemble the \((\text{Si}_6\text{O}_{18})^{12-}\) ring and expand outward, the optimal path involves assembly of a \((\text{MgBSi}_3\text{O}_{10})^{11-}\) subunit that can be repeated by tourmaline’s trigonal symmetry:

\[
\begin{align*}
\text{(Si–O)} × 3 & → (\text{Si}_2\text{O}_7) & \text{SBMA} = 3 \text{ steps} & \text{EMA} = 2 \text{ steps} \\
(\text{Si}_2\text{O}_7) × 2 & → (\text{Si}_2\text{O}_7) & 1 \text{ step} & 1 \text{ step} \\
(\text{Mg}–\text{O}) × 3 & → (\text{MgO}_4) & 3 \text{ steps} & 3 \text{ steps} \\
(\text{Si}_2\text{O}_7) + (\text{MgO}_4) & → (\text{MgSi}_3\text{O}_{10}) & 1 \text{ step} & 1 \text{ step} \\
(\text{Si}_2\text{O}_7) + (\text{MgSi}_3\text{O}_{10}) & → (\text{MgBSi}_3\text{O}_{12}) & 1 \text{ step} & 1 \text{ step}
\end{align*}
\]
At this point, three identical \((\text{MgBSi})_2\) subunits can be linked in 2 steps:

\[
\begin{align*}
\text{(MgBSi)}_2 & \rightarrow \text{2 steps} \\
\text{TOTAL} & \rightarrow \text{2 steps}
\end{align*}
\]

The assembly index for the tourmaline cluster is thus 14 or 11, depending on the method used. Note that this cluster as written has a large negative charge \((-33\text{ in this case)}\)—a situation typical of many other molecular clusters in minerals. In an aqueous environment, this charge would be compensated by the formation of numerous peripheral \((\text{O}\cdots\text{H})\) bonds, which are not considered in the calculation of assembly indices. Alternatively, the addition of apical \(\text{F}\) atoms instead of \(\text{OH}^\cdot\), as is common in tourmaline (thus forming \((\text{Mg-F})\) or \((\text{Si-F})\) bonds as part of the cluster) might increase the exact MA index to 12 or greater.

In this example, tourmaline displays a characteristic shared by many high-symmetry clusters. In spite of the fact that the cluster contains more than 100 cation-oxygen bonds, the assembly index is only 11 or 14 for exact or split-branch calculations, respectively, as a consequence of the repeated use of subunits such as \((\text{MgSi})_2\) and \((\text{BeO})_3\), which only need to be assembled once.

2.3. Eudialyte group

Complex molecular subunits often arise when silicate structures are coupled with other high-field-strength cations,
such as V$^{4+}$, V$^{5+}$, As$^{5+}$, Zr$^{4+}$, and/or U$^{6+}$ into structures called heteropolyanions. A common example is provided by the diverse eudialyte group (figure 3), with more than 30 species that form exclusively in high-temperature igneous environments associated with alkali-rich, silica-poor magmas [15,22,23]. A typical species is the eponymous eudialyte $[\text{Na}_{15}\text{Ca}_6\text{Fe}_3\text{Zr}_3\text{Si}(\text{Si}_{25}\text{O}_{73})(\text{O,OH,H}_2\text{O})_3(\text{Cl,OH})_2]$. Eudialyte incorporates at least two different molecule-like sub-assemblies. One example, shown here, includes an edge-sharing six-member ring of calcium octahedra (in orange), with three-member (Si$_3$O$_9$) rings (red) above and below and six additional (Si$_3$O$_{10}$) three-tetrahedron segments around the periphery to form a $[\text{Ca}_6(\text{Si}_3\text{O}_9)_2(\text{Si}_3\text{O}_{10})_6]^{30-}$ cluster. We also focus on the strongly bonded sandwich-like subunit of (Zr$_6$O$_{12}$) octahedra between two (Si$_9$O$_{27}$) rings $[(\text{Si}_9\text{O}_{27})_2(\text{Zr}_6\text{O}_{12})_6]^{60-}$, as represented by adjacent red and dark blue polyhedra in (a). (From Johnson & Grice [22], used with permission from the Canadian Journal of Mineralogy and Petrology.)

for calculating a minimum assembly pathway is to exploit this cluster’s 6-fold symmetry by first assembling a $[\text{Ca}(\text{SiO}_3)(\text{Si}_9\text{O}_{27})]^{6+}$ subunit. Start by considering the silicate groups:

<table>
<thead>
<tr>
<th>Reaction</th>
<th>SBMA</th>
<th>EMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\text{SiO}_4] \times 3 \rightarrow (\text{SiO}_3)$</td>
<td>3 steps</td>
<td>2 steps</td>
</tr>
<tr>
<td>$[\text{SiO}_4] \times 2 \rightarrow (\text{Si}_2\text{O}_6)$</td>
<td>1 step</td>
<td>1 step</td>
</tr>
<tr>
<td>$([\text{Si}_2\text{O}_6] + (\text{Si}-0) \rightarrow (\text{Si}<em>3\text{O}</em>{10})$</td>
<td>1 step</td>
<td>1 step</td>
</tr>
<tr>
<td>$([\text{Si}_2\text{O}_6] + (\text{Si}<em>3\text{O}</em>{10}) \rightarrow (\text{Si}<em>4\text{O}</em>{16})$</td>
<td>1 step</td>
<td>1 step</td>
</tr>
</tbody>
</table>

([3 tetrahedra chain])
These silicate groups are linked by the calcium atom:

\[
\begin{align*}
(SiO_3)_2 + (Ca_0) &\rightarrow [Ca(SiO_3)_2] \\
[Ca(SiO_3)] + (Si_0) &\rightarrow [Ca(SiO_3)(SiO_3)] \\
[Ca(SiO_3)(SiO_3)] &\rightarrow [Ca(SiO_3)(SiO_3)] \\
2 \rightarrow [Ca(SiO_3)(SiO_3)] &\rightarrow [Ca(SiO_3)(SiO_3)] \\
3 \rightarrow [Ca(SiO_3)(SiO_3)] &\rightarrow [Ca(SiO_3)(SiO_3)] \\
\end{align*}
\]

TOTAL \( SBMA = 12 \) \( EMA = 10 \) steps

Thus, the assembly index is 10 or 12, depending on the method used. Note that the final assembly steps involve the simultaneous formation of 6 or more bonds between two subunits—a situation typical for assembly of high-symmetry clusters. As in the previous examples, H atoms must decorate the apical oxygens of this cluster in solution to maintain a low net molecular charge; nevertheless, O–H bonds are not included in the assembly calculation.

Another key molecular subunit of eudialyte is \([Si_6O_{27}]^{2-} (ZrO_4)^{6-}\), which consists of a sandwich of 9-member \((Si_6O_{27})^{18-}\) silicate rings above and below, and linked by six \((ZrO_4)^{6-}\) octahedra (figure 3a). In this case, the first steps involve assembling a \([Si_6O_{27}](ZrO_4)_2^{20-}\) subunit, thus exploiting the cluster's 3-fold symmetry. Start with assembly of the \((Si_3O_9)\) and \((ZrO_4)\) polyhedral units:

\[
\begin{align*}
([ZrO_3]) + 2 \rightarrow ([ZrO_3]) &\rightarrow [ZrO_3] \\
([ZrO_3]) + 2 \rightarrow ([ZrO_3]) &\rightarrow [ZrO_3] \\
([Si_3O_9]) + 3 \rightarrow ([Si_3O_9]) &\rightarrow [Si_3O_9] \\
([Si_3O_9]) + 3 \rightarrow ([Si_3O_9]) &\rightarrow [Si_3O_9] \\
\end{align*}
\]

Combine those units in the \([Si_6O_{27}](ZrO_4)_2\) subassembly:

\[
\begin{align*}
([Si_3O_9]) + ([ZrO_3]) &\rightarrow [Si_3O_9](ZrO_3) \\
([Si_3O_9])(ZrO_3) &\rightarrow [Si_3O_9](ZrO_3) \\
\end{align*}
\]

Finally, combine 3 subassemblies into the cluster:

\[
\begin{align*}
([Si_3O_9])(ZrO_3) &\rightarrow [Si_3O_9](ZrO_3) \\
3 \rightarrow [Si_3O_9](ZrO_3) &\rightarrow [Si_3O_9](ZrO_3) \\
\end{align*}
\]

TOTAL \( SBMA = 12 \) \( EMA = 10 \) steps

Thus, the molecular assembly of this eudialyte subunit is 10 or 12, depending on the method used. Note that adding on to this subunit are sodium, calcium, and iron atoms, as well as three-member \((Si_3O_{10})^{18-}\) molecules, thus significantly increasing the structural complexity of eudialyte.

### 2.4. Vanarsite and related structures

Among the most remarkable heteropolyanions are examples of the \([As^{3+}V_{12}^{4+}As^{6+}_6O_{48}]\) cluster (figure 4a), which is found in several rare minerals, including vanarsite, packratite, gatewayite and morrisonite [24]. Consider vanarsite \([NaCa_{12}As^{3+}V_{12}^{4+}As^{6+}_6O_{48}·78H_2O]\), in which \([As^{3+}V_{12}^{4+}As^{6+}_6O_{48}]\) heteropolyanion clusters are linked to each other by weaker \((Na–O), (Ca–O), \) and hydrogen bonds. In calculating the molecular assembly index for this cluster, we treat the two different oxidation states of arsenic separately because they play distinct structural roles. As in the examples of tourmaline and eudialyte, we take the trigonal symmetry into account, thus first assembling the \([As^{3+}_2V_{12}^{4+}O_{48}]\) subcluster, which has a pair of \((As^{5+}O_2)\) tetrahedra, a triad of edge-sharing \((VO_6)\) octahedra, and a corner-sharing \((VO_6)\) octahedron (figure 4a). Begin by assembling the polyhedral units:

\[
\begin{align*}
[As^{5+}O_4] &\times 2 \rightarrow [As^{5+}O_4] \\
[VO_6] &\times 2 \rightarrow [VO_6] \\
[VO_6] &\times 2 \rightarrow [VO_6] \\
[VO_6] &\times 2 \rightarrow [VO_6] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [V octahedron] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [V octahedron] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [V octahedron] \\
\end{align*}
\]

Next, combine polyhedra into the \([As^{5+}_2V_{12}^{4+}O_{48}]\) subcluster:

\[
\begin{align*}
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [VO_6] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [VO_6] \rightarrow [As^{5+}_2V_{12}^{4+}O_{48}] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times [VO_6] \rightarrow [As^{5+}_2V_{12}^{4+}O_{48}] \\
\end{align*}
\]

Three of these \([As^{5+}_2V_{12}^{4+}O_{48}]\) subclusters form the ring-like structure:

\[
\begin{align*}
[As^{5+}_2V_{12}^{4+}O_{48}] &\times 3 \rightarrow [As^{5+}_2V_{12}^{4+}O_{48}] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times 3 \rightarrow [As^{5+}_2V_{12}^{4+}O_{48}] \\
[As^{5+}_2V_{12}^{4+}O_{48}] &\times 3 \rightarrow [As^{5+}_2V_{12}^{4+}O_{48}] \\
\end{align*}
\]

Assembly is completed in 11 or 16 steps by the addition of a central \(As^{3+}\) atom, which forms 3 \((As^{3+}–O)\) bonds in a single step:

\[
\begin{align*}
[As^{5+}_2V_{12}^{4+}O_{48}] + [As^{3+}–O] &\rightarrow [As^{5+}_2V_{12}^{4+}O_6] \\
3 \rightarrow [As^{5+}_2V_{12}^{4+}O_6] \\
\end{align*}
\]

TOTAL \( SBMA = 16 \) \( EMA = 11 \) steps

Note that in this instance the split-branch method results in a significantly higher index (16 steps) than the exact method (11 steps). More than a dozen additional assembly steps, including the numerous sodium, calcium, and hydrogen atoms of the so-called ‘interstitial structure’, are required to complete the crystal structure of vanarsite (figure 4b). Additionally, several apical oxygen atoms on both \((As^{5+}O_4)\) tetrahedra and \((V^{4+}O_6)\) octahedra are terminated by \((OH^-)\) groups—bonds that are not included according to assembly theory protocols.

### 2.5. Paddlewheelite

Heteropolyanions that spontaneously assemble in aqueous solution have been extensively studied in the laboratory [16–18,24]. More than 40 minerals have been described that contain heteropolyanions, which preserve complex signatures of geochemical conditions of formation [16,25]. Often natural heteropolyanions are highly complex [19,20,25], and thus minerals incorporating these clusters represent an excellent opportunity to test hypotheses related to the maximum possible molecular assembly indices for natural inorganic molecules.

A variety of complex heteropolyanions occur in secondary uranium-bearing minerals. One fascinating example
The magnesium-calcium-copper-uranyl carbonate mineral paddlewheelite \(\text{[MgCa}_5\text{Cu}_2\text{(UO}_2\text{)}_4\text{(CO}_3\text{)}_{12}\text{(H}_2\text{O)}_{33}\text{]} \) [19], which gets its colourful name from a unique polyanion \(\text{[(Cu}_2\text{+O)}_2\text{Ca}^2\text{+U}^{6+}_4\text{O}_{10}\text{(CO}_3\text{)}_{12}]^{18-}\) — a cluster with a (CaCu2) ‘axle’ and four paddle-like radiating \(\text{[UO}_2\text{(CO}_3\text{)}_3\text{]}\) groups

Assembly of this complex cluster begins with sub-assembly of one of the \(\text{[UO}_2\text{(CO}_3\text{)}_3\text{]}\) planar groups as follows:

\[
\begin{align*}
\text{(C–O)} \times 3 & \rightarrow \text{(CO}_3\text{)} & \text{SBMA} = 3 \text{ steps} & \text{EMA} = 2 \text{ steps} \\
\text{(U–O)} \times 2 & \rightarrow \text{(UO}_2\text{)} & 2 \text{ steps} & 1 \text{ step}
\end{align*}
\]

In the above steps, recall that we add bonds, not atoms; therefore, for example, 2 oxygen atoms are shared by U and C in the last of these steps.

\[
\begin{align*}
\text{(UO}_2\text{)} \times 2 & \rightarrow \text{(UO}_2\text{)} & 1 \text{ step} & 1 \text{ step} \\
\text{(UO}_2\text{)} + \text{(CO}_3\text{)} & \rightarrow \text{(UO}_2\text{(CO}_3\text{)}_2\text{)} & 1 \text{ step} & 1 \text{ step} \\
\text{(UO}_2\text{(CO}_3\text{)}_2\text{)} + \text{(CO}_3\text{)} & \rightarrow \text{(UO}_2\text{(CO}_3\text{)}_3\text{)} & 1 \text{ step} & 1 \text{ step}
\end{align*}
\]
The next steps add 2 (Ca–O) and a (Cu–O2) bond to the paddlewheels:

\[
\begin{align*}
(Ca-O) \times 2 + [(UO_2)(CO_3)]_3 &\rightarrow \text{3 steps} \quad \text{2 steps} \\
(Cu-O) \times 2 + (Cu_2) &\rightarrow \text{2 steps} \quad \text{1 step}
\end{align*}
\]

apical oxygen to Cu

\[
\begin{align*}
(Cu_2O_2) + [Ca[(UO_2)(CO_3)]_3] &\rightarrow \text{1 step} \quad \text{1 step} \\
([CuO][Ca[(UO_2)(CO_3)]_3]] &\rightarrow \text{1 step} \quad \text{1 step}
\end{align*}
\]

Finally, the cluster is completed by combining 4 paddlewheels in 2 steps:

\[
\begin{align*}
(CuO)(Ca)(UO_2)(CO_3)]_3 \times 2 &\rightarrow \text{1 step} \quad \text{1 step} \\
(CuO)(Ca)(UO_2)(CO_3)]_3 \times 2 &\rightarrow \text{1 step} \quad \text{1 step} \\
[(CuO)(Ca)(UO_2)(CO_3)]_3 \times 2 &\rightarrow \text{1 step} \quad \text{1 step} \\
(Total SBMA = 17 steps) &\text{EMA = 13 steps}
\end{align*}
\]

The resulting molecular assembly index for paddlewheelite is 13 or 17, depending on the method used.

2.6. Ewingite

The most complex mineral structure identified thus far is the hydrous magnesium-calcium-uranyl carbonate, ewingite \([Mg_{6}Ca_{6}(UO_2)_{24}(CO_3)_{30}O_{4}(OH)_{12}(H_2O)_{138}]\), with only six different essential elements but a remarkable 121 different atomic sites in its asymmetric unit [20]. Embedded in ewingite is a molecule-like uranyl carbonate ‘cage’ with composition \([(UO_2)_{24}(CO_3)_{30}O_{4}]^{20-}\), which in turn is made from three different polyhedral building units: four of \([U_3O_{16}]\), six of \([UO_4)(CO_3)_{2}\), and six of \([UO_2)(CO_3)]_3\) (figure 6). In spite of the overall structural complexity, the assembly of the molecular subunit is simplified by the numerous repeating (U–O) and (C–O) bonds in (UO2), (UO3), and (CO3) polyhedra, as well as its pseudo-trigonal symmetry.

Consider first the \((U_3O_{16})\) cluster (figure 6c), which requires a triad of edge-sharing \((UO_7)\) polyhedra, which can be assembled by combining one unit with 6 (U–O) bonds to 2 units with 5 (U–O) bonds.

\[
\begin{align*}
(U-O) \times 2 + [(UO_2)(CO_3)]_3 &\rightarrow \{Ca[(UO_2)(CO_3)]_3\} \quad \text{3 steps} \quad \text{2 steps} \\
(Cu-O) \times 2 &\rightarrow \{CuO\} \quad \text{add an} \quad \{CuO\} \quad \text{apical oxygen to Cu} \\
(Cu_2O_2) + [Ca[(UO_2)(CO_3)]_3] &\rightarrow \{CuO(Ca[(UO_2)(CO_3)]_3)\} \quad \text{1 step} \quad \text{1 step} \\
([CuO][Ca[(UO_2)(CO_3)]_3]] &\rightarrow \{CuO(Ca[(UO_2)(CO_3)]_3)\} \quad \text{1 step} \quad \text{1 step} \\
[(CuO)(Ca)(UO_2)(CO_3)]_3 \times 2 &\rightarrow \{CuO(Ca[(UO_2)(CO_3)]_3)\} \quad \text{1 step} \quad \text{1 step} \\
\{CuO(Ca[(UO_2)(CO_3)]_3)\} \times 2 &\rightarrow \{CuO(Ca[(UO_2)(CO_3)]_3)\}_2 \quad \text{1 step} \quad \text{1 step} \\
\{CuO(Ca[(UO_2)(CO_3)]_3)\}_2 &\rightarrow \{CuO(GaU_4O_{10}(CO_3)_{12}\} \quad \text{1 step} \quad \text{1 step}
\end{align*}
\]

TOTAL SBMA = 17 steps EMA = 13 steps

Thus, 4 steps yield the basic uranium coordination groups required for cluster 1 (figure 6c), which is further assembled by combination:

\[
\begin{align*}
(UO_2)_3 \times 2 + (UO_3) &\rightarrow \{U_2O_{16}\} \quad \text{2 steps} \quad \text{2 steps} \\
(UO_2) + (UO_3) &\rightarrow \{U_2O_{16}\} \quad \text{1 step} \quad \text{1 step} \\
(UO_2) + (UO_3) &\rightarrow \{U_2O_{16}\} \quad \text{1 step} \quad \text{1 step}
\end{align*}
\]

Four additional steps are required to assemble cluster 3, \([(UO_2)(CO_3)]_3\) (figure 6e), which incorporates a \((UO_6)\) polyhedron with two peripheral \((CO_3)\) groups:

\[
\begin{align*}
(UO_2) + (UO_3) &\rightarrow \{U_2O_{16}\} \quad \text{1 step} \quad \text{1 step} \\
(C-O) \times 2 &\rightarrow (CO_3) \quad \text{2 steps} \quad \text{1 step} \\
(UO_2) + (CO_3) \times 2 &\rightarrow \{[UO_2](CO_3)_2\} \quad \text{2 steps} \quad \text{2 steps}
\end{align*}
\]
Only one additional step is required to assemble cluster 2, \([\text{UO}_4\text{(CO}_3)_3]\) (figure 6d), which adds a third peripheral \((\text{CO}_3)\) group to cluster 3:

\[
[\text{UO}_4\text{(CO}_3)_2] + (\text{CO}_2) \rightarrow [\text{UO}_4\text{(CO}_3)_3]
\]

1 step 1 step

Note that several alternative assembly pathways might be employed to form clusters 1, 2, and 3 from the constituent bonds, but all pathways appear to require a minimum of 11 assembly steps.

Assembly of the entire cluster requires attaching one or two \([\text{UO}_4\text{(CO}_3)_2]\) and \([\text{UO}_2\text{(CO}_3)_3]\) groups to each of four \([\text{U}_3\text{O}_{16}]\) subclusters (4 steps):

\[
[\text{UO}_4\text{(CO}_3)_2] + [\text{UO}_2\text{(CO}_3)_3] \rightarrow [\text{UO}_2\text{2(CO}_3)_5]
\]

1 step 1 step

\[
[\text{UO}_2\text{2(CO}_3)_5] + [\text{UO}_4\text{(CO}_3)_2] \rightarrow [\text{UO}_2\text{3(CO}_3)_7]
\]

1 step 1 step

\[
[\text{UO}_2\text{3(CO}_3)_7] + [\text{U}_3\text{O}_{16}] \rightarrow [\text{UO}_2\text{6(CO}_3)_7\text{O}]
\]

1 step 1 step

\[
[\text{UO}_2\text{3(CO}_3)_8] + [\text{U}_3\text{O}_{16}] \rightarrow [\text{UO}_2\text{6(CO}_3)_8\text{O}]
\]

1 step 1 step

The final assembly steps involve combining 2 \([\text{UO}_2\text{6}(\text{CO}_3)_7\text{O}]\) and 2 \([\text{UO}_2\text{6}(\text{CO}_3)_8\text{O}]\):

\[
[\text{UO}_2\text{6}(\text{CO}_3)_7\text{O}] + [\text{UO}_2\text{6}(\text{CO}_3)_8\text{O}] \rightarrow [\text{UO}_2\text{12(CO}_3)_15\text{O}_4]
\]

1 step 1 step

\[
[\text{UO}_2\text{12(CO}_3)_15\text{O}_4] \times 2 \rightarrow [\text{UO}_2\text{24(CO}_3)_30\text{O}_4]
\]

1 step 1 step

Thus, it requires a total of 18 or 20 steps to assemble the \([\text{UO}_2\text{24}(\text{CO}_3)_30\text{O}_4]\) heteropolyanion.

The structure of ewingite was difficult to determine and refine owing to the quality of the available crystals. Synthetic analogues similar to this cluster have since been reported [26,27], but containing four \([\text{U}_3\text{O}_{16}]\) subclusters and 12 \([\text{UO}_2\text{(CO}_3)_3]\) subclusters. It is possible that the cluster of ewingite also contains 12 \([\text{UO}_2\text{(CO}_3)_3]\) as opposed to 6 each of \([\text{UO}_2\text{(CO}_3)_2]\) and \([\text{UO}_2\text{(CO}_3)_2]\); however, the diffraction data failed to reveal the locations of disordered carbonate groups. In the event that this is the case, a minimum of 16 steps would be needed to assemble the ewingite cluster.

### 2.7. Ilmajokite

Ilmajokite \([\text{Na}_{11}\text{KBaCe}_2\text{Ti}_{12}\text{Si}_{37.5}\text{O}_{94}(\text{OH})_{30}\cdot 29\text{H}_2\text{O}]\) is a cerium-bearing titanium silicate of exceptional structural complexity, with 236 unique atom positions in the asymmetric unit [28]. The monoclinic structure incorporates a trigonal prismatic titanosilicate (TPTS) cluster of composition \([\text{Ce(Ti}_2\text{SiO}_{12})_3(\text{Si}_{17}\text{O}_{35})]\), which is among the most complex known molecule-like entities thus far observed in minerals (figure 7a,b). The core of each cluster consists of three separate \((\text{Ti}_2\text{O}_{10})\) dimers of edge-sharing titanium octahedra that create a nine-coordinated site for one central \(\text{Ce}^{3+}\) — hence a \([\text{Ce(Ti}_2\text{O}_{10})_3]\) group. Each \((\text{Ti}_2\text{O}_{10})\) group is also bonded to individual \((\text{SiO}_4)\) tetrahedra (including tetrahedra numbered 7 and 31 in figure 7b), resulting in a \([\text{Ce(Ti}_2\text{SiO}_{12})_3]\) subassembly. The periphery of this group is decorated by 17 corner-linked silicate tetrahedra in a complexly branched \((\text{Si}_{17}\text{O}_{52})\) chain that wraps around all six titanium octahedra (including 17 corner-shared O atoms as attachment points).

The structure of this tetrahedral assembly (figure 7b) can be described by a linear 11-tetrahedron chain (beginning with tetrahedron 8 (here abbreviated ‘\(T_8\’\); figure 7b) and ending at \(T_{19}\)), with \(T_3\) and \(T_4\) branching off \(T_5\); \(T_{13}\) and \(T_{14}\) branching off \(T_{11}\); and \(T_{17}\) and \(T_{18}\) branching off \(T_{10}\). This topology is analogous to that of 3,6,9-ethylundecane, but with corner-linked silicate tetrahedra instead of single-bonded carbon atoms.

Calculation of the molecular assembly index of this 98-atom cluster begins with the central \([\text{Ce(Ti}_2\text{SiO}_{12})_3]\) subassembly, which has pseudo-trigonal symmetry:

\[
(\text{Ti}_0) \times 2 \rightarrow (\text{Ti}_0)^2 \quad \text{SBMA} = 2 \quad \text{EMA} = 1
\]

\[
(\text{Ti}_0)^2 \times 2 \rightarrow (\text{Ti}_0)^4 \quad 1 \text{ step} \quad 1 \text{ step}
\]

\[
(\text{Ti}_0)^4 \times 2 \rightarrow (\text{Ti}_0)^8 \quad 1 \text{ step} \quad 1 \text{ step}
\]

\[
(\text{Ti}_0)^8 \times 2 \rightarrow (\text{Ti}_0)^{16} \quad 1 \text{ step} \quad 1 \text{ step}
\]
Each of 3 (Ti2O10) dimers must be bound to an apical (SiO4) tetrahedron, as well as the central (CeO6) polyhedron:

\[
\begin{align*}
(SiO_4) & \rightarrow 2SiO_4) \\
(SiO_4) + (Ti_2O_{10}) & \rightarrow (Ti_2SiO_{12}) \\
(SiO_4) + (Ti_2O_{10}) & \rightarrow [Ce(Ti_2SiO_{12})] \\
(SiO_4) + (Ti_2O_{10}) & \rightarrow [Ce(Ti_2SiO_{12})] \\
\end{align*}
\]

Thus, a total of 9 or 12 assembly steps is required for the central [Ce(Ti2SiO12)3] subunit. Assembly of the branched silicate chain proceeds as follows, noting that (SiO2) has already been assembled above (refer to figure 7b):

\[
\begin{align*}
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \text{ (construction of T8)} \\
(SiO_4) \times 2 & \rightarrow (SiO_4) \text{ (linking T8 to T2)} \\
(SiO_4) \times 2 & \rightarrow (SiO_4) \text{ (linking T8-T2 to T5-T4)} \\
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \text{ (T8 to T3)} \\
\end{align*}
\]

This is a chain with 5 tetrahedra that repeats 3 times between T8 and T14:

\[
(SiO_4) \times 3 \rightarrow (Si_3O_9) \text{ [tetrahedral chain from T8 to T14]} \\
\]

To complete the silicate subcluster, add T16 and T19 (i.e. Si6O18) to T13 of the previous assembly:

\[
\begin{align*}
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \\
(Si_3O_9) + (SiO_4) & \rightarrow (Si_3O_9) \\
\end{align*}
\]

Thus, the branched silicate chain requires 8 steps. The final assembly step is to attach the branched silicate chain to the central [Ce(Ti2SiO12)3] cluster, requiring 1 additional step:

\[
(Si_3O_9) + [Ce(Ti_2SiO_{12})_3] \rightarrow [Ce(Ti_2SiO_{12})_3] \\
\]

Consequently, the molecular assembly index of the ilmajo-kite heteropolyanion is 18 or 21, depending on the method used.

3. Mineral structural complexity and assembly indices

Krivovichev et al. [21,29–34] have introduced a number of metrics to characterize the chemical and structural complexity of minerals. Of greatest relevance to this study are calculations of the structural complexity per unit cell (in bits). Krivovichev et al. [34] recorded structural complexity values for 4596 mineral species—approximately 75% of all IMA-approved mineral species.

Structural complexity is estimated as the amount of structural Shannon information per atom (\(I_{Sc}\)) and per unit cell (\(I_{Sc,cell}\)) calculated according to the following equations [21,29]:

\[
I_{Sc} = -\sum_{i=1}^{k} p_i \log_2 p_i \left( \frac{\text{bit}}{\text{atom}} \right) \\
\]

\[
\alpha I_{Sc,cell} = -\sum_{i=1}^{k} p_i \log_2 p_i \left( \frac{\text{bit}}{\text{cell}} \right) \\
\]

where \(k\) is the number of different crystallographic sites in the structure and \(p_i\) is the random choice probability for an atom from the \(i\)th crystallographic site, that is:

\[
p_i = \frac{m_i}{v} \\
\]

Thus, a total of 9 or 12 assembly steps is required for the central [Ce(Ti2SiO12)3] subunit. Assembly of the branched silicate chain proceeds as follows, noting that (SiO2) has already been assembled above (refer to figure 7b):

\[
\begin{align*}
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \text{ (construction of T8)} \\
(SiO_4) \times 2 & \rightarrow (SiO_4) \text{ (linking T8 to T2)} \\
(SiO_4) \times 2 & \rightarrow (SiO_4) \text{ (linking T8-T2 to T5-T4)} \\
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \text{ (T8 to T3)} \\
\end{align*}
\]

This is a chain with 5 tetrahedra that repeats 3 times between T8 and T14:

\[
(SiO_4) \times 3 \rightarrow (Si_3O_9) \text{ [tetrahedral chain from T8 to T14]} \\
\]

To complete the silicate subcluster, add T16 and T19 (i.e. Si6O18) to T13 of the previous assembly:

\[
\begin{align*}
(SiO_4) + (SiO_4) & \rightarrow (SiO_4) \\
(Si_3O_9) + (SiO_4) & \rightarrow (Si_3O_9) \\
\end{align*}
\]

Thus, the branched silicate chain requires 8 steps. The final assembly step is to attach the branched silicate chain to the central [Ce(Ti2SiO12)3] cluster, requiring 1 additional step:

\[
(Si_3O_9) + [Ce(Ti_2SiO_{12})_3] \rightarrow [Ce(Ti_2SiO_{12})_3] \\
\]

Consequently, the molecular assembly index of the ilmajo-kite heteropolyanion is 18 or 21, depending on the method used.

3. Mineral structural complexity and assembly indices

Krivovichev et al. [21,29–34] have introduced a number of metrics to characterize the chemical and structural complexity of minerals. Of greatest relevance to this study are calculations of the structural complexity per unit cell (in bits). Krivovichev et al. [34] recorded structural complexity values for 4596 mineral species—approximately 75% of all IMA-approved mineral species.

Structural complexity is estimated as the amount of structural Shannon information per atom (\(I_{Sc}\)) and per unit cell (\(I_{Sc,cell}\)) calculated according to the following equations [21,29]:

\[
I_{Sc} = -\sum_{i=1}^{k} p_i \log_2 p_i \left( \frac{\text{bit}}{\text{atom}} \right) \\
\]

\[
\alpha I_{Sc,cell} = -\sum_{i=1}^{k} p_i \log_2 p_i \left( \frac{\text{bit}}{\text{cell}} \right) \\
\]

where \(k\) is the number of different crystallographic sites in the structure and \(p_i\) is the random choice probability for an atom from the \(i\)th crystallographic site, that is:

\[
p_i = \frac{m_i}{v} \\
\]

Mineral assembly indices and complexity per unit cell differ significantly in their derivations; however, as might be expected the two measures display some correlations. The most complex crystal structures also tend to have the highest assembly indices.

In their list of the 20 most complex known mineral structures, all with more than 3000 bits per unit cell, Krivovichev et al. [33] reveal at least six characteristics that lead to complexity:

1. the presence of large polyoxometalate clusters, often exceeding a nanometre in diameter;
2. the presence of large clusters linked to form an even larger framework topology;
3. three-dimensional modular frameworks that incorporate multiple cage-like topologies;
4. formation of modular layers that can combine in multiple topologies;
5. high-hydration states in salts with complex heteropolyhedral units; and
6. formation of ordered superstructures of relatively simple structure types.

Most of these conditions lead to a high number of different local coordination topologies—factors that appear to be representative of minerals with molecule-like substructures with MA index ≥10 (table 1). Only in the latter instance (criterion 6) of superstructures based on simple subunits, with the same coordination states appearing over and over again, will the assembly indices be low compared to the structural complexity.

A survey of all 4596 known mineral structures [33,34] reveals that 1600 minerals have relatively simple structures with ≤100 bits per unit cell—a value typical for minerals with an assembly index ≤6 (table 1). Another approximately 1600 minerals have between 100 and 500 bits per unit cell, typically corresponding to crystalline phases with assembly indices up to 10 in our study. By contrast, only about 750 of the 4596 minerals surveyed (16%) of all minerals have >500 bits per unit cell, some of which correspond to phases in our calculations with assembly indices >10. Of these more complex minerals, only 369 of 4596 (8%) display >1000 bits per unit cell—most of which have assembly indices significantly greater than 10.

It is important to note that of these most complex minerals, a significant number, including paddlewheelite,
ewigite and vanarsite, arise as secondary oxidized and hydrated phases of precursor primary phases. Therefore, it might be argued that biology (principally owing to atmospheric oxygenation via photosynthesis) played a role in their formation [23]. However, in theory, these oxidized minerals could form on lifeless terrestrial worlds whose near-surface environments have been oxidized via photochemical processes and atmospheric escape of hydrogen [35–37]; thus, their formation may not unambiguously reflect the presence of a biosphere. Importantly, other minerals with complex molecular substructures, such as ilmajokite and members of the eudialyte, vesuvianite and tourmaline groups, form exclusively as primary igneous or metamorphic phases—i.e. with no biological influence.

4. Conclusion

We draw several conclusions from this investigation of the MA indices of naturally occurring inorganic molecules:

— Some naturally occurring heteropolyanion clusters, including those in ewigite and ilmajokite, have assembly indices greater than the proposed abiotic/biotic cutoff of 15, thus invalidating the claim that only biological processes can produce molecules with assembly indices ≥15.

— Heteropolyanion clusters of complexity greater than those of ewigite and ilmajokite likely occur in nature. All of the 5 most complex structures listed in table 1 were determined since 2016. The structures of more than 1000 known minerals, many of them extremely complex, have yet to be fully refined [34]. Comparable examples of natural complex heteropolyanions with molecular assembly indices exceeding 14 steps for strongly bonded molecular subunits of mineral structures are possibly among the minerals containing heteropolyanions reviewed by Krivovichev [25].

— Hystad and colleagues [38] have demonstrated that more than 3000 additional mineral species on Earth exist, but have yet to be discovered and described. New mineral species are being discovered at a rate of approximately 100 species per year [39], and many of those display complex, previously unknown structures [34,40].

A principal objective of this study is to test the hypothesis of Marshall et al., who suggest that ‘There is a critical value of pathway complexity above which all artefacts must be biologically derived’ [1], while ‘complex molecules found in high abundance are universal biosignatures’ [3]. They suggest that only living cells generate significant concentrations of structures with MA index ≥15. Our evaluation of this central thesis of assembly theory is direct and quantitative. The present study of mineral assembly indices reveals that, while most common molecule-like structures in minerals have MA indices ≤ 10, the most complex mineral species have molecule-like subunits with MA indices ≥ 21. What does this result imply in the context of the search for agnostic biosignatures? One of two conclusions might be drawn.

First, it is possible that the complex molecular heteropolyanions incorporated into mineral crystals are fundamentally different from organic molecules in their assembly rules. Therefore, the fact that some inorganic minerals have molecule-like subunits that can achieve assembly indices greater than 15 might be irrelevant in the search for life. In this case, the rhetoric associated with molecular assembly theory should be clarified to exclude exclusively inorganic species. However, such an exclusion would seem to violate the very premise of assembly theory’s claims regarding unambiguous biosignatures based on molecular complexity.

Alternatively, given the plausibility of the assembly theory contention that only life can produce multiple copies of extremely complex objects, it is possible that a value (or range of values) of assembly index greater than 15 may represent a more realistic, if ‘fuzzy’, divide between abiotic and biotic structures. In this regard, an intriguing topic for speculation relates to the potential for ‘open-ended evolution’ of different natural systems [41–43]. We suggest that minerals display bounded evolution, with the implication that there exists an upper limit to the molecular assembly index for naturally occurring heteropolyanions—perhaps a limit not much greater than 21. Life, by contrast, may display unbounded evolution, with the possibility that objects with ever-increasing assembly indices may emerge, initially through biological evolution and subsequently through technological evolution.

In either case, we conclude that significant structural complexity of molecules is not the unique province of biochemistry and that natural inorganic chemistry has the potential to generate significant local populations of molecular structures with MA indices greater than 15. In that case, the possibility should be entertained that abiotic processes might also produce organic crystals of great complexity on a carbon-rich, abiotic planet or moon, especially given billions of years of abiotic organic mineral evolution in the absence of life. For example, Maynard-Casley et al. [44, table 2] and Cable et al. [45, table 1] have proposed the existence of complex organic co-crystals on Titan. Examples of these compounds, all of which reflect laboratory synthesis under plausible Titan conditions, include the bimolecular co-crystals (1:4) 1,3-butanediol:nitrogen and (1:1) acetonitrile:3,4-dihydroxybenzoic acid, as well as the tri-molecular co-crystals (1:1:1) acetonitrile:cytc:urea and (1:2:1) acetonitrile:3,4-dihydroxybenzoic acid:2,3,5,6-tetramethylpyrazine. The estimated molecular assembly indices of these proposed Titan organic minerals range from 5 to 10. If such abiotic compounds do exist on Titan, then, by analogy with the multi-stage congruent evolution of minerals on Earth [46,47], local pockets of much greater abiotic molecular complexity are likely to persist, as well.

An important conclusion of prior studies is that planets and moons, given sufficient time and chemical processing, display remarkable degrees of mineral evolution, which include striking increases in the average diversity and complexity of minerals [33,34,46,47]. On Earth, the earliest minerals formed by high-temperature condensation display an average complexity per unit cell of approximately 83 bits, with no minerals incorporating molecule-like subunits with MA indices >6. Following the advent of plate tectonics, that average structural complexity value increased to 275 bits per unit cell, with tourmaline and eudialyte group minerals displaying molecule-like subunits with MA indices ≥10. By contrast, the average complexity of mineral species formed in tandem with today’s terrestrial biosphere exceeds 370 bits per unit cell, with MA indices of some heteropolyanionic clusters significantly exceeding 15. These trends suggest that planets and moons can generate chemical compounds of significant structural complexity through purely abiotic
processes. Furthermore, the processes leading to mineral complexification are congruent—i.e. each stage of mineral evolution provides new phases that have the potential for further complexification [46–48].

We have demonstrated that abiotic chemical processes have the potential to form crystal structures of great complexity—values exceeding the proposed abiotic/biotic divide of MA index = 15 [3]. In this context of dynamic, evolving planetary environments, the absence of concentrations of complex abiotic organic molecules on Earth is not surprising. The ubiquity of complex cellular life would certainly mask an abiotic suite of complex molecules, assuming that such molecules had not been long since consumed by a voracious biosphere. However, no such constraints would seem to affect Titan and other organic-rich abiotic worlds, where complex molecular species in the form of co-crystals may abound.

In conclusion, while the proposal of a biosignature based on a molecular assembly index of 15 is an intriguing and testable concept, the contention that only life can generate molecular structures with MA index ≥ 15 is in error. Furthermore, in spite of amusing speculations in the literature of science fiction [49], it is unlikely that the definition of any universal phenomenon as complex and diverse as ‘life’ can be reduced to a scalar.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

References

15. Downs RT. 2023 The muff database. See https://muff.info/maa.
24. Kampf AR, Hughes JM, Nash BP, Marty J. 2016 Vananite, packrinite, morrisonite, and gatewayite: four new minerals containing the $\text{As}^{3+11}V_{12}^{3+}$ heteropolyanion, a novel