

Data Descriptor

Not peer-reviewed version

Curated Polyoxometalate Formula Dataset

[Aleksandar Kondinski](#)^{*}, [Nadiia Gumerova](#)^{*}, [Anette Rompel](#)

Posted Date: 2 September 2024

doi: 10.20944/preprints202408.2238.v1

Keywords: polyoxometalates; chemical formula; dataset



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Curated Polyoxometalate Formula Dataset

Aleksandar Kondinski ^{1,*} , Nadiia Gumerova ^{2,*}  and Annette Rompel ² 

¹ Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK

² Universität Wien, Fakultät für Chemie, Institut für Biophysikalische Chemie, Althanstr. 14, 1090 Vienna, Austria

* Correspondence: aleksandar@kondinski.com (A.K.); nadiia.gumerova@univie.ac.at (N.G.)

Abstract: Reticular and cluster materials often feature complex formulas, making a comprehensive overview challenging due to the need to consult various resources. While datasets have been collected for metal-organic frameworks (MOFs), covalent organic frameworks (COFs), and zeolites, among others, there remains a gap in systematically organized information for polyoxometalates. This paper introduces a carefully curated dataset of 1984 polyoxometalate (POM) and related cluster metal oxide formula instances, currently connecting over 2,500 POM material instances. These POM instances incorporate 75 different chemical elements, with compositions ranging from binary to octonary element clusters. This dataset not only enhances accessibility to polyoxometalate data but also aims to facilitate further research and development in the study of these complex inorganic compounds.

Keywords: polyoxometalates; chemical formula; dataset

1. Summary

In materials science, comprehensive datasets are indispensable for investigating complex materials such as covalent organic frameworks (COFs) [1], metal-organic frameworks (MOFs) [2], and zeolites [3, 4]. These datasets facilitate the exploration of structural diversity and analysis of properties such as porosity, adsorption, and permeation. Moreover, they empower researchers to apply data intelligence in understanding trends in material properties and in predicting behaviors under various conditions, including applications in catalysis and sieving [5].

Polyoxometalates (POMs) are versatile metal-oxo clusters with diverse applications in catalysis [6], life sciences [7–11], and nanoelectronics [12,13]. Despite their structural complexity and promising potential in smart applications, there has been limited development of curated POM datasets crucial for advancing AI-driven technologies in exploring inorganic chemical spaces. The creation of these datasets is essential for accelerating POM chemistry through the understanding of POM speciation in solution *via* techniques such as nuclear magnetic resonance and mass spectrometry [14–17], utilizing POMs as building blocks in hybrid composite materials [18–20], and developing metallodrugs [11].

Over the past century, the designation of POM formulas has posed significant challenges at the forefront of inorganic chemistry [21], as crystalline POM materials are among the most information-dense [22]. The structural complexity, intricacies of coordination bonding, and charge (de)localization make common chemical identifiers like InChI (IUPAC International Chemical Identifier) inadequate for fully capturing their nuances [23,24]. Although SciFinder has developed some collections of POM information, the proprietary nature of this database limits access to the data [25]. Consequently, the development of open and curated datasets of POM formulations could significantly accelerate research and collaboration within the POM community.

In this work, we present the recently completed “Curated Polyoxometalate Formula Dataset,” now published on a Git platform under a creative common license. This dataset addresses a critical data gap by linking POM formulas with their corresponding materials and provenances, specifically through Digital Object Identifiers (DOIs) [26]. It facilitates searches based on elements, charges, molecular mass, and POM formulas. Additionally, the structure and implementation of the dataset are designed for future reusability and expansion, enabling the development of new data-driven methods for exploring POM synthesis and other related research avenues.

2. Data Description

Polyoxometalates (POMs), as cluster materials, are featured as crystallographic motifs in a variety of POM-containing or POM-based materials. For instance, the α -Wells Dawson polyoxometalate composed of P, W, and O atoms with the formula $[(PO_4)_2(WO_3)_{18}]^{6-}$, as illustrated in Figure 1, exemplifies such motifs. This specific POM, reported in diverse materials like $Li_6[(PO_4)_2(WO_3)_{18}] \cdot 28H_2O$ (DOI: 10.1039/C3DT51120K), showcases the potential for various formula descriptions within the same POM motif. The given formula, $[(PO_4)_2(WO_3)_{18}]^{6-}$, is a coordination formula easily interpreted by chemists, indicating an overall charge of 6− and a composition that can be described as $P_2W_{18}O_{62}$. This leads to possible empirical derivations like PW_9O_{31} , which can be explored further in synthetic studies.

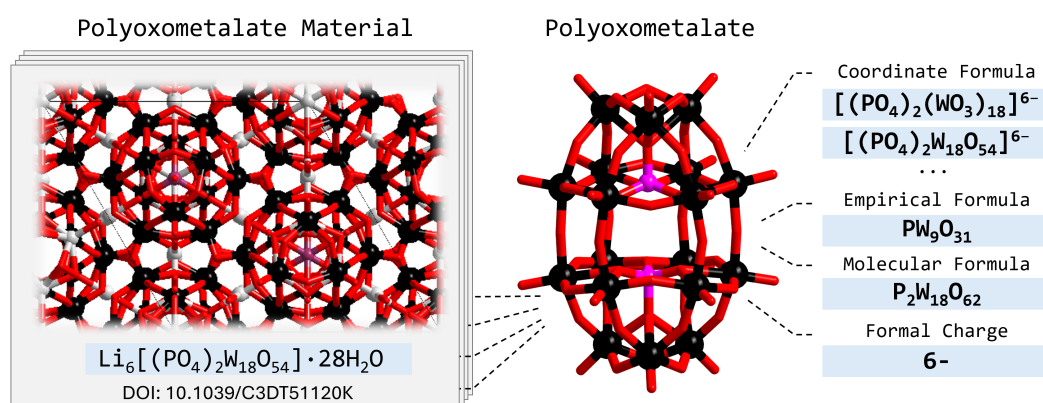


Figure 1. Illustration of an α -Wells Dawson polyoxometalate instance, showcasing its connection to different chemical formulas on the right and various crystalline materials on the left, each linked to literature (e.g., DOI: 10.1039/C3DT51120K).

Consider the connectivity outlined in Figure 2.a, which depicts a data connection schema for polyoxometalates and polyoxometalate materials. Each material is linked to its formula and a Digital Object Identifier (DOI), encapsulating a POM as a distinct entity within the class, assigned a unique identifier. This POM entity connects to various attributes such as elemental composition, from which both the molecular and POM formulas—typically serving as coordination formulas—are derived. These features, coupled with specific charge and molecular mass data, enable domain chemists to label and classify the structure effectively.

In terms of implementation (see Figure 2.b), our JSON data structure organizes information hierarchically about a specific Polyoxometalate (POM) entity, encapsulated within a top-level object. For example, we assign a unique four-letter key (e.g., "POM_GUKA") as a POM identifier, referencing the phosphotungstate Wells-Dawson structure from Figure 1. This key points to a nested JSON object detailing the POM's properties such as formulas, elemental composition, and related materials. Keys like "POM Formula" and "Molecular Formula" store formulas as strings, "Contains Elements" maps to an object listing each element's count, "Molecular Mass" is recorded as a string for precision, and "Charge" as a floating-point number. An array under "Labels" collects descriptive tags, while "POM Material" includes nested objects with unique identifiers detailing each material's formula and DOI, enhancing data retrieval for scientific analysis and database integration.

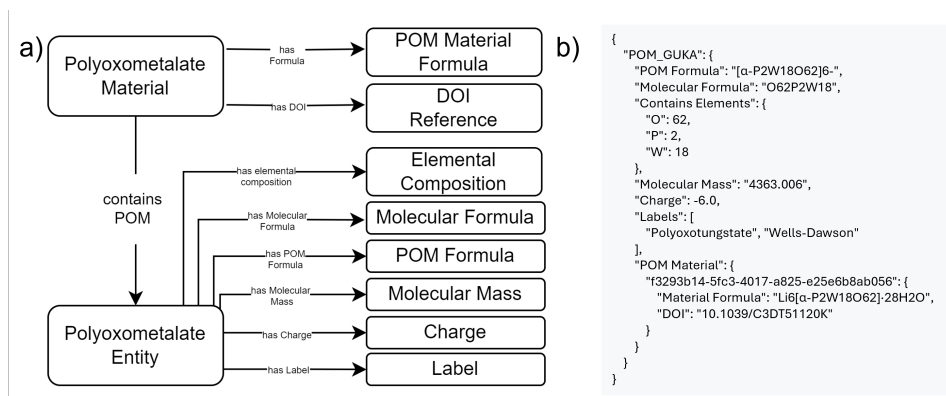


Figure 2. Hierarchical schema of polyoxometalate data. (a) Shows the relationship between polyoxometalate materials and their characteristics, including elemental composition, molecular formulas, and bibliographic references. (b) Provides a JSON example of how these attributes are digitally organized, illustrating a specific polyoxometalate entity with its associated material details.

3. Methods

For the creation of the curated polyoxometalate formula dataset, we have meticulously compiled chemical data on polyoxometalates (POMs) by analyzing over 1300 distinct research papers, each identified by a unique Digital Object Identifier (DOI). The resulting dataset currently includes 1984 unique entries, linking to more than 2500 POM-containing materials. Given the crucial role of nuclearity and charge in characterizing POMs, our data are presented in terms of overall charge plotted as a function of molecular mass, as illustrated in Figure 3.a. This panel depicts the broad distribution of molecular weights and corresponding charges, highlighting the chemical diversity within polyoxometalates. Notably, extreme examples such as the polyoxomolybdate $[Mo_{368}O_{1032}H_{16}(H_2O)_{240}(SO_4)_{48}]^{48-}$ [27], the polyoxotungstate $[(P_8W_{48}O_{184})\{(P_2W_{14}Mn_4O_{60})(P_2W_{15}Mn_3O_{58})_2\}_4]^{152-}$ [28], the dense polyoxouranate $[(UO_2)_{120}(O_2)_{120}(C_2O_4)_{90}]^{180-}$ [29], and the robust polyoxoniobate $[Nb_{288}O_{768}(OH)_{48}(CO_3)_{12}]^{180-}$ [30], are easily distinguishable, demonstrating the current extents of the POM chemical space.

A further qualitative analysis of the plot in Figure 3.b reveals that as the charge increases while the molecular mass decreases or remains constant, there are fewer or no POMs observed. This trend suggests an over-reduction of the addenda centers, which likely requires coordination changes not captured by this dataset. Nonetheless, this pattern can serve as empirical guidance to evaluate the feasibility of proposed POM formulas. Focusing on the region characterized by a charge of less than -75 and a molecular mass below 30,000 g/mol, we observe that this is where the majority of POMs are situated. By employing different colorings for analysis, it becomes apparent that many polyoxomolybdates are located in the domain of high molecular mass but relatively low overall charge. Conversely, polyoxoniobates appear as one of the most negatively charged species with lower molecular weights, highlighting the diverse charge and mass relationships within POMs.

The current set of POM entries is primarily dominated by 1023 polyoxotungstate formulas. The prevalence of tungstates is likely due to tungsten's ability to form strong, non-labile bonds in high oxidation states and to generate lacunary POM species that facilitate further chemical reactions, enhancing the diversity of tungstate-based POMs. Following tungstates are polyoxomolybdates with 482 entries, polyoxovanadates with 206, polyoxoniobates with 102, and polyoxotantalates with 30 entries; an additional 141 entries are categorized under "other," including diverse metal-oxo clusters. The dataset showcases chemical diversity with 75 unique elements across all entries, illustrating a broad range of structural combinations. Specifically, the distribution of unique elements per POM varies from two to eight, with the majority of structures, 603, containing five elements, often reflecting

the integration of heteroatoms and hybrid species involving carbon, which underscores the complex compositional variety within this dataset.

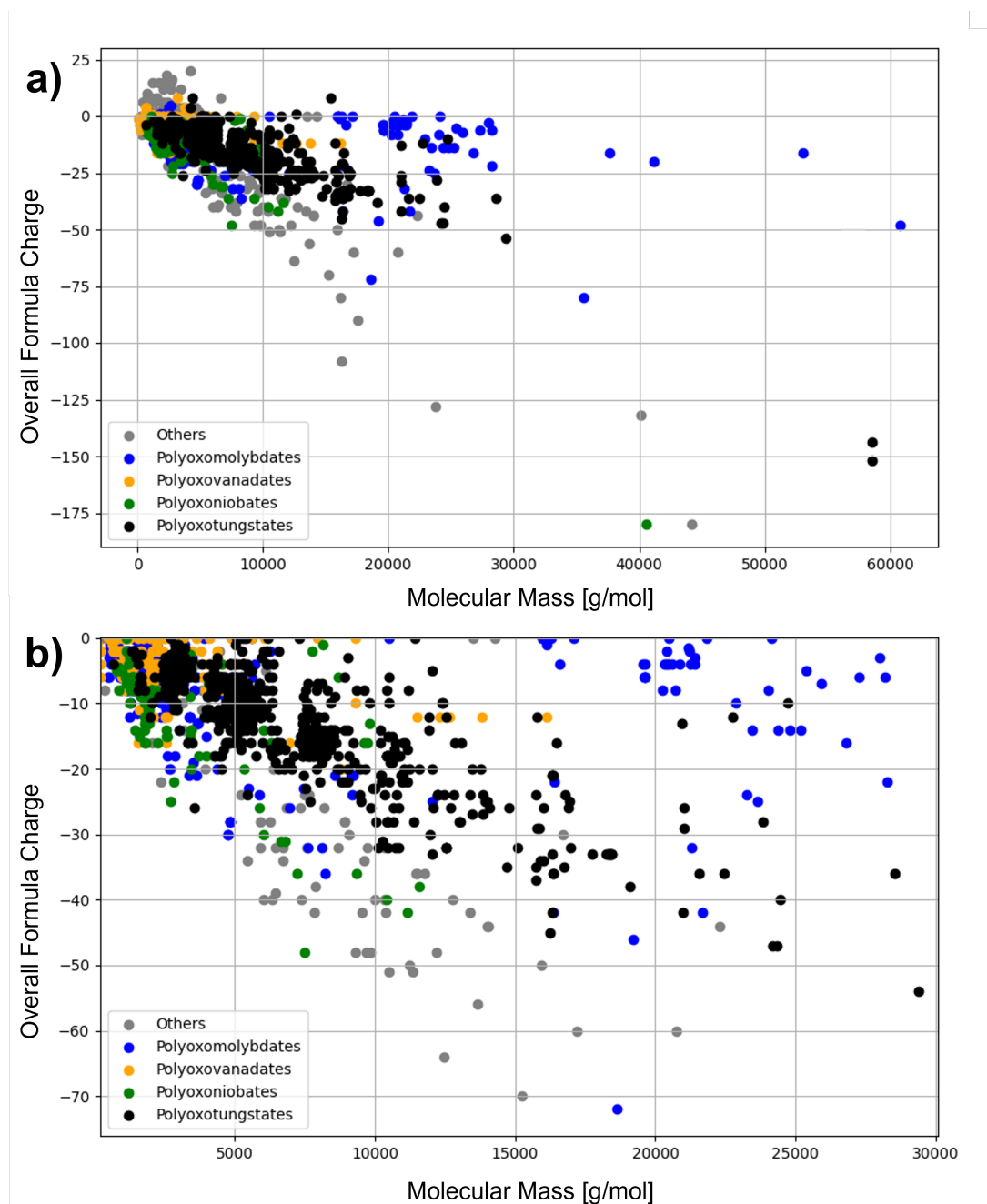


Figure 3. Charge vs. Molecular Mass in Polyoxometalates: (a) shows the overall distribution for reported polyoxometalates, highlighting broad molecular weight trends. (b) zooms into molecular masses up to 30,000 Da, detailing finer distribution characteristics. These plots explore the chemical space of polyoxometalates.

4. Conclusions

This paper presents a comprehensive dataset of nearly 2000 POMs, currently featuring details such as molecular formulas, charges, and DOI identifiers, with an aim for future expansion to include topological, synthetic, and functional properties in the near future. The dataset is designed to evolve, incorporating charge distributions and mass projections to enhance material data projects through machine learning and cheminformatics. Upcoming enhancements will also cover polyoxometalates'

speciation in various environments to assess protonation levels, isomerism, and charge distribution for automated computational analyses, as well as integrated components for retrosynthetic strategies, significantly advancing both the fundamental and applied research in polyoxometalate chemistry.

Author Contributions: AK, NG, and AR conceptualized the need for the dataset. AK designed the data structure. AK and NG were involved in data curation. NG performed validation checks to ensure the reliability of the data. The manuscript was completed with contributions from all authors.

Funding: This research was funded in part by the Austrian Science Fund (FWF) [DOI 10.55776/P33089 (to A.R.); DOI 10.55776/P33927 (to N.G.)].

Data Availability Statement: One can access the “Curated Polyoxometalate Formula” dataset using the Digital Chemistry git repository at the following URL: https://github.com/digital-chemistry/Curated_POMs (accessed on 27 August 2024).

Acknowledgments: The University of Cambridge and the University of Vienna are thanked for research support. We extend our gratitude to Ella Duvanova for her invaluable assistance with data verification.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

POM	Polyoxometalate
DOI	Digital Object Identifiers
CSV	Comma-separated values
JSON	JavaScript Object Notation

References

1. Ongari, D.; Yakutovich, A.V.; Talirz, L.; Smit, B. Building a consistent and reproducible database for adsorption evaluation in covalent–organic frameworks. *ACS Cent. Sci.* **2019**, *5*, 1663–1675. doi:10.1021/acscentsci.9b00619.
2. Moghadam, P.Z.; Li, A.; Wiggin, S.B.; Tao, A.; Maloney, A.G.; Wood, P.A.; Ward, S.C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database subset: a collection of metal–organic frameworks for past, present, and future. *Chem. Mater.* **2017**, *29*, 2618–2625.
3. Yang, S.; Lach-Hab, M.; Vaisman, I.I.; Blaisten-Barojas, E.; Li, X.; Karen, V.L. Framework-type determination for zeolite structures in the inorganic crystal structure database. *J. Phys. Chem. Ref. Data.* **2010**, *39*. doi:10.1063/1.3432459.
4. Zheng, C.; Li, Y.; Yu, J. Database of open-framework aluminophosphate structures. *Sci. Data* **2020**, *7*, 107. doi:10.1038/s41597-020-0452-4.
5. Kancharlapalli, S.; Snurr, R.Q. High-throughput screening of the CoRE-MOF-2019 database for CO₂ capture from wet flue gas: a multi-scale modeling strategy. *ACS Appl. Mater. Interfaces* **2023**, *15*, 28084–28092. doi:10.1021/acsami.3c04079.
6. Liu, R.; Streb, C. Polyoxometalate-single atom catalysts (POM-SACs) in energy research and catalysis. *Adv. Energy Mater.* **2021**, *11*, 2101120. doi:10.1002/aenm.202101120.
7. Gumerova, N.I.; Rompel, A. Interweaving disciplines to advance chemistry: Applying polyoxometalates in biology. *Inorg. Chem.* **2021**, *60*, 6109–6114. doi:10.1021/acs.inorgchem.1c00125.
8. Barba-Bon, A.; Gumerova, N.I.; Tanuhadi, E.; Ashjari, M.; Chen, Y.; Rompel, A.; Nau, W.M. All-Inorganic Polyoxometalates Act as Superchaotropic Membrane Carriers. *Adv. Mater.* **2024**, *36*, 2309219. doi:10.1002/adma.202309219.
9. Bijelic, A.; Rompel, A. Ten good reasons for the use of the tellurium-centered Anderson–Evans polyoxotungstate in protein crystallography. *Acc. Chem. Res.* **2017**, *50*, 1441–1448. doi:10.1021/acs.accounts.7b00109.
10. Bijelic, A.; Aureliano, M.; Rompel, A. The antibacterial activity of polyoxometalates: structures, antibiotic effects and future perspectives. *ChemComm* **2018**, *54*, 1153–1169.
11. Bijelic, A.; Aureliano, M.; Rompel, A. Polyoxometalates as potential next-generation metallodrugs in the combat against cancer. *Angew. Chem. Int. Ed.* **2019**, *58*, 2980–2999. doi:10.1002/anie.201812022.

12. Kondinski, A. Metal–metal bonds in polyoxometalate chemistry. *Nanoscale* **2021**, *13*, 13574–13592. doi:10.1039/D1NR03115H.
13. Kondinski, A.; Banerjee, A.; Mal, S.S. Hervé- and Krebs-Type Magnetic Polyoxometalate Dimers. *Magnetochemistry* **2022**, *8*, 96. doi:10.3390/magnetochemistry8090096.
14. Gumerova, N.I.; Rompel, A. Speciation atlas of polyoxometalates in aqueous solutions. *Sci. Adv.* **2023**, *9*, eadi0814. doi:10.1126/sciadv.adi0814.
15. Gumerova, N.I.; Rompel, A. Polyoxometalates in solution: speciation under spotlight. *Chem. Soc. Rev.* **2020**, *49*, 7568–7601. doi:10.1039/D0CS00392A.
16. Surman, A.J.; Robbins, P.J.; Ujma, J.; Zheng, Q.; Barran, P.E.; Cronin, L. Sizing and discovery of nanosized polyoxometalate clusters by mass spectrometry. *J. Am. Chem. Soc.* **2016**, *138*, 3824–3830. doi:10.1021/jacs.6b00070.
17. Kondinski, A.; Rasmussen, M.; Mangelsen, S.; Pienack, N.; Simjanoski, V.; Näther, C.; Stares, D.L.; Schalley, C.A.; Bensch, W. Composition-driven archetype dynamics in polyoxovanadates. *Chem. Sci.* **2022**, *13*, 6397–6412. doi:10.1039/D2SC01004F.
18. Vilà-Nadal, L. POMzites: A roadmap for inverse design in metal oxide chemistry. *Int. J. Quantum Chem.* **2021**, *121*, e26493. doi:10.1002/qua.26493.
19. Soria-Carrera, H.; Atrián-Blasco, E.; de la Fuente, J.M.; Mitchell, S.G.; Martín-Rapún, R. Polyoxometalate–polypeptide nanoassemblies as peroxidase surrogates with antibiofilm properties. *Nanoscale* **2022**, *14*, 5999–6006. doi:10.1039/d1nr08223j.
20. Cameron, J.M.; Guillemot, G.; Galambos, T.; Amin, S.S.; Hampson, E.; Haidaraly, K.M.; Newton, G.N.; Izzet, G. Supramolecular assemblies of organo-functionalised hybrid polyoxometalates: from functional building blocks to hierarchical nanomaterials. *Chem. Soc. Rev.* **2022**, *51*, 293–328. doi:10.1039/D1CS00832C.
21. Pope, M.T.; Jeannin, Y.; Fournier, M. *Heteropoly and isopoly oxometalates*; Vol. 8, Springer, 1983.
22. Krivovichev, S.V. Which inorganic structures are the most complex? *Angew. Chem. Int. Ed.* **2014**, *53*, 654–661. doi:10.1002/anie.201304374.
23. Bard, A.J.; Dance, I.G.; Day, P.; Ibers, J.A.; Kunitake, T.; Meyer, T.J.; Mingos, D.M.P.; Roesky, H.W.; Sauvage, J.P.; Simon, A.; others. *Bonding and charge distribution in polyoxometalates: a bond valence approach*; Springer, 1999. doi:10.1007/BFb0103378.
24. Heller, S.R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC international chemical identifier. *J. Cheminform.* **2015**, *7*, 1–34. doi:10.1186/s13321-015-0068-4.
25. Gabrielson, S.W. SciFinder. *J. Med. Libr. Assoc.* **2018**, *106*, 588. doi:10.5195/jmla.2018.515.
26. Paskin, N. Digital object identifier (DOI®) system. *Encycl. Libr. Inf. Sci.* **2010**, *3*, 1586–1592. doi:10.1081/E-ELIS3-120044418.
27. Müller, A.; Beckmann, E.; Bögge, H.; Schmidtman, M.; Dress, A. Inorganic Chemistry Goes Protein Size: A Mo₃₆₈ Nano-Hedgehog Initiating Nanochemistry by Symmetry Breaking. *Angew. Chem. Int. Ed.* **2002**, *41*, 1162–1167. doi:10.1002/1521-3773(20020402)41:7<1162::AID-ANIE1162>3.0.CO;2-8.
28. Fang, X.; Kögerler, P.; Furukawa, Y.; Speldrich, M.; Luban, M. Molecular growth of a core–shell polyoxometalate. *Angew. Chem. Int. Ed.* **2011**, *50*, 5212–5216. doi:10.1002/anie.201008225.
29. Ling, J.; Qiu, J.; Burns, P.C. Uranyl peroxide oxalate cage and core–shell clusters containing 50 and 120 uranyl ions. *Inorg. Chem.* **2012**, *51*, 2403–2408. doi:10.1021/ic202380g.
30. Wu, Y.L.; Li, X.X.; Qi, Y.J.; Yu, H.; Jin, L.; Zheng, S.T. {Nb₂₈₈O₇₆₈(OH)₄₈(CO₃)₁₂}: A Macromolecular Polyoxometalate with Close to 300 Niobium Atoms. *Angew. Chem. Int. Ed.* **2018**, *57*, 8572–8576. doi:10.1002/anie.201804088.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.