

Experiment Design

International Edition: DOI: 10.1002/anie.201705721
German Edition: DOI: 10.1002/ange.201705721

Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates

Vasilios Duros⁺, Jonathan Grizou⁺, Weimin Xuan, Zied Hosni, De-Liang Long, Haralampos N. Miras, and Leroy Cronin*

Abstract: The discovery of new gigantic molecules formed by self-assembly and crystal growth is challenging as it combines two contingent events; first is the formation of a new molecule, and second its crystallization. Herein, we construct a workflow that can be followed manually or by a robot to probe the envelope of both events and employ it for a new polyoxometalate cluster, $\text{Na}_6[\text{Mo}_{120}\text{Ce}_6\text{O}_{366}\text{H}_{12}(\text{H}_2\text{O})_{78}]\cdot 200\text{H}_2\text{O}$ (**1**) which has a trigonal-ring type architecture (yield 4.3% based on Mo). Its synthesis and crystallization was probed using an active machine-learning algorithm developed by us to explore the crystallization space, the algorithm results were compared with those obtained by human experimenters. The algorithm-based search is able to cover ca. 9 times more crystallization space than a random search and ca. 6 times more than humans and increases the crystallization prediction accuracy to $82.4 \pm 0.7\%$ over $77.1 \pm 0.9\%$ from human experimenters.

Understanding the supramolecular self-assembly of complex inorganic molecules poses a difficult problem since it relies on two contingent events.^[1] To make a discovery the conditions under which the building blocks assemble have to be found and then the conditions under which the product aggregates into crystals to be isolated and characterized need to be identified. The vast number of combinations of the experimental conditions and the coordination modes of the transition metals taking part in the building blocks means that a full exploration of the chemical space of any given compound would be impossible.^[2] For these reasons, the intuition of experienced chemists is required to design the appropriate experiments to determine the right conditions for the isolation of any new products.^[3] But intuitions can be biased by both the current knowledge of the field and the

frame of mind of the experimenter—making important discoveries difficult to achieve.

Herein, we design and investigate a new approach for probing the envelope of both the synthesis and the crystallization process of a new polyoxometalate compound with the formula $\text{Na}_6[\text{Mo}_{120}\text{Ce}_6\text{O}_{366}\text{H}_{12}(\text{H}_2\text{O})_{78}]\cdot 200\text{H}_2\text{O}$ (**1**) $\{\text{Mo}_{120}\text{Ce}_6\}$ (Figure 1). Our method is drawn from recent

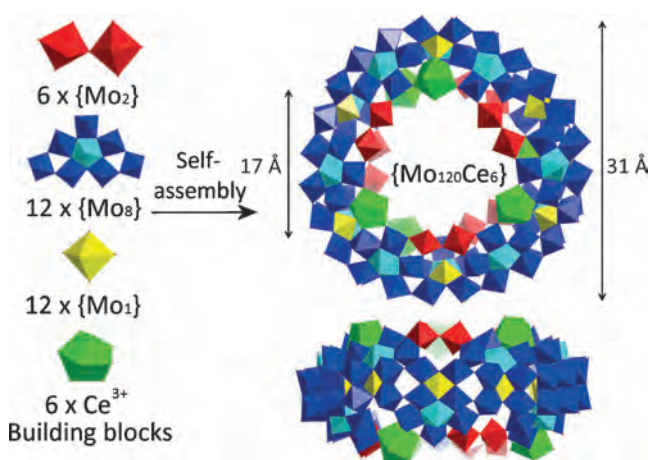


Figure 1. Schematic representation of the self-assembly of the $\{\text{Mo}_{120}\text{Ce}_6\}$ wheel from basic building blocks in polyhedron mode. Coloring code: $\{\text{Mo}_2\}$ red; $\{\text{Mo}_8\}$ blue with central atom in cyan; $\{\text{Mo}_1\}$ yellow; Ce green.

advances for active data acquisition in the field of machine learning, known as active learning.^[4] Active learning consists of methodologies able to decide what experiments to perform next in order to optimally improve the understanding of the system at hand. We compare our algorithmic method with a random screening process in the exploration and modelling of the crystallization conditions of compound (**1**). Importantly, we study how human experimenters approached this specific problem and compare their strategies and performance to our machine-learning approach.

So far, work in this area has been mainly focused on simulations and only a few studies have involved real experiments.^[5] For example, recently, Wicker and Cooper^[6] applied machine learning methods to draw a map of crystallinity according to the size of a molecule and its number of rotatable bonds. Similarly, Oliynyk et al.^[7] used machine learning to predict structures of inorganic binary compounds of the general formula AB by considering various atomic and physical properties in their calculations. Of

[*] V. Duros,^[‡] Dr. J. Grizou,^[‡] Dr. W. Xuan, Z. Hosni, Dr. D.-L. Long, Dr. H. N. Miras, Prof. L. Cronin
WEST Chem, School of Chemistry
University of Glasgow
University Avenue, Glasgow, G12 8QQ (UK)
E-mail: lee.cronin@glasgow.ac.uk
Homepage: <http://www.croninlab.com>

[‡] These authors contributed equally in this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.201705721>.

© 2017 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

particular interest, Norquist et al.^[8] made use of data from unsuccessful syntheses to predict reaction outcomes of vanadium compounds and compared the efficiency of their algorithms with the typical strategies that human chemists apply.

Our machine learning approach actively defines new experiments to perform with an aim to improve its model of the system. Such targeted data acquisition strategy allows a reduction in the number of experiments needed to attain the same model quality, thus saving time and financial resources. To our knowledge, it is the first time that such an active data acquisition strategy is applied in this context and compared with human experimenters. Machine learning methods have previously been used as a tool of optimization^[9] and a faster data mining technique for extensive databases.^[10–14] It is important to note that our approach should not be mistaken for high-throughput screening as it uses machine learning techniques capable of abstracting problems rather than a brute force increase of processing speed. We instead suggest this approach should be viewed as “intelligent throughput” since not all the possible experiments are done, and only those chosen by the algorithm are explored and the system effectively learns as the experiment continues similar to how an expert chemist would work.

We first introduce the compound that was discovered, the reaction conditions from which it can be isolated and characterized. We then compare our machine-learning approach against random screening and human experimenters in terms of performance and methodologies for the exploration of the crystallization boundaries (see Figure 2).

The new polyoxometalate cluster belongs to the family of lanthanide-doped molybdenum blues.^[15–21] Compound **1** is isostructural to the reported^[22] $\text{Na}_6[\text{Mo}_{120}\text{Pr}_6\text{O}_{366}(\text{H}_2\text{O})_{78}\text{H}_{12}] \cdot \text{ca. } 200\text{H}_2\text{O}$, but notably was first discovered automatically by our automated chemical robot, see Supporting Information: Experimental Section, Method A. In later experiments we also reproduced the synthesis and crystallization of the compound on the bench, see Supporting Information: Experimental Section, Method B. Compound **1** was characterized by elemental analysis, single-crystal X-ray structure analysis, bond valence sum (BVS) calculations, IR

and visible-NIR spectroscopy, redox titrations and thermogravimetry.

The single-crystal X-ray structure analysis reveals four of the dodecameric ring-shaped clusters **1** in the unit cell, packed parallel to the crystallographic *bc* plane giving rise to 1D channels occupied by guest water molecules (Figure S3 in the Supporting Information). The framework of **1** consists of 12 sets of basic building blocks $\{\text{Mo}_8\}$, $\{\text{Mo}_2\}$ and $\{\text{Mo}_1\}$ units, which are well-defined in Mo Blue clusters such as the archetypal $\{\text{Mo}_{154}\}$,^[22] $\{\text{Mo}_{176}\}$,^[23] and $\{\text{Mo}_{368}\}$,^[24] with 6 $\{\text{Mo}_2\}$ units substituted by 6 Ce^{III} ions. On the whole, the architecture of **1** is constructed from 12 $\{\text{Mo}_8\}$ units, 6 $\{\text{Mo}_2\}$ units, 12 $\{\text{Mo}_1\}$ units and 6 $\{\text{Ce}(\text{H}_2\text{O})_5\}$ units (Figure 1). The coordination configuration of the two distinct types of Ce^{III} can be described as a distorted monocapped square antiprism, built from four $\mu_2\text{-O}$ atoms and five H_2O molecules that is, $\{\text{Ce}(\text{H}_2\text{O})_5\}$. Bond lengths of molybdenum atoms coordinated to terminal oxo groups have a $\text{Mo}=\text{O}$ bond length in the range of 1.554(12)–1.702(9) Å. The symmetric arrangement of 3 Ce^{III} ions on both the upper and lower surfaces of $\{\text{Mo}_{120}\text{Ce}_6\}$ greatly reduces the symmetry of **1** to D_3 as compared with the parent $\{\text{Mo}_{154}\}$ (D_{7d} point group). As a result, the wheel displays an irregular ring-shaped structure with an outer ring diameter of about 31 Å and an inner ring diameter of about 17 Å. A further characteristic of the structure of **1** is the large number of protons resulting from the 24 e^- reduction. The overall reduction state of **1** was confirmed using three independent techniques: UV/Vis spectroscopy, redox titration and bond valence sum calculations (BVS), [see Supporting Information for details]. BVS calculations^[25] are carried out on all the Mo and O centers (Table S2). A careful analysis of the BVS result reveals 12 singly and 78 doubly protonated oxygen atoms. Taking into consideration the obtained information from the above calculations along with elemental analysis and redox titrations, it is possible to determine the overall building-block scheme and overall charge for compound **1**: $[\{\text{Mo}_2\}_6\{\text{Mo}_1\}_{12}\{\text{Mo}_8\}_{12}\{\text{Ce}_6\}]^{\equiv}[\{\text{Mo}^{\text{VI}}_2\text{O}_5(\text{H}_2\text{O})_2\}_6\{\text{Mo}^{\text{VI/V}}_8\text{O}_{26}(\mu_3\text{-O})_2\text{H}(\text{H}_2\text{O})_3\text{Mo}^{\text{VI/V}}\}_{12}\{\text{Ce}^{\text{III}}(\text{H}_2\text{O})_5\}_6]^{6-}$.

To explore the synthetic and crystallization process it is important to define the process of the reaction accurately as shown in Figure 2. By describing an abstract method we could then turn this into a concrete procedure and then output the precise set of experiments to perform, determined by either a human or the algorithm-driven robot using three methods; robot-algorithm; human; and robot-random as the control method. For the experimental conditions to be defined and explored three distinct pieces of information must be provided: 1) the chemicals involved in the synthesis, 2) an experimental method for the synthesis and crystallization process, and 3) an initial set of data consisting of successful and failed crystallization experiments, that is, the starting information used to decide what experiments to perform next, see Figure 3. Next, to compare the methods using a commonly calibrated and therefore robust experimental test, we developed an automated platform (Figure S8) able to consistently perform the crystallization experiments given a list of parameters such as the number of reagents and their corresponding volumes (see Supporting Information, part 7).

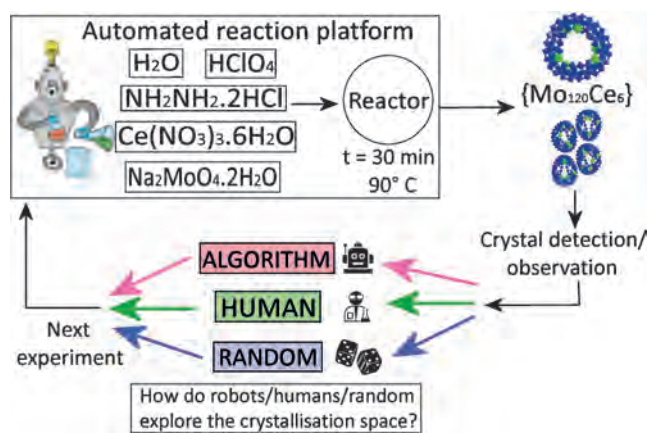


Figure 2. Representation of the experimental method showing how the automated and bench work was done. Structure: Mo blue; Ce green.

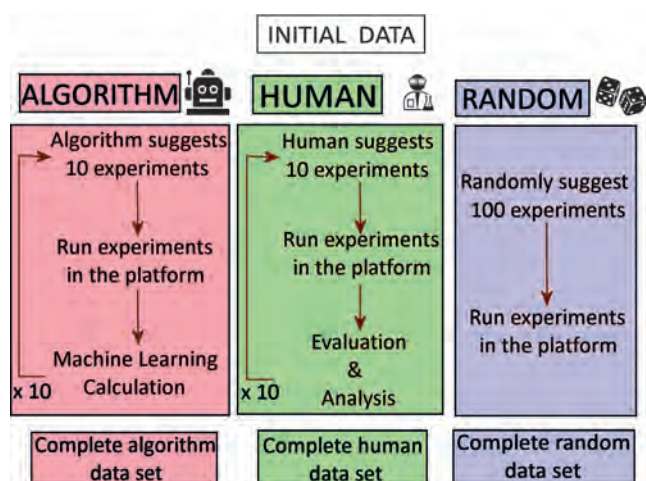


Figure 3. Schematic diagram of the exploration methods used in our studies comparing the algorithmic approach with that of the human experimenter and a random approach. Both the random and algorithmic approaches used a purpose-built liquid handling and crystallization robotic platform.

For the reactions, aqueous stock solutions of $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ 1 M, $\text{Ce}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ 0.1 M, $\text{NH}_2\text{NH}_2 \cdot 2\text{HCl}$ 0.25 M and HClO_4 1 M were prepared and used as described in the Supporting Information: Experimental Section, Method A. During the experiment both $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ and $\text{Ce}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ were always added in a 1:1 volume ratio (molar ratio 10:1 respectively). The automated platform mixes, in a reactor, the stock solutions in configurable ratios to a total volume of 15 mL, and allows them to react at 90 °C for 30 min. We then collect a 9 mL sample for crystallization and then use a cleaning protocol to reset the system before the next experiment begins. The crystallization method consists of waiting overnight to allow for crystals to form. Finally, the presence or absence of crystals is checked under illumination with a white light emitting diode (3300–3500 lux at a distance of 5 cm) and the information is added in a database of experiments. It is important to note that under these particular conditions, no other product crystallizes in a time frame longer than one month, see Supporting Information, Table S12. Following this process the initial data set was obtained from previous experiments performed in the platform and is shown in Figure 4, also see Supporting Information, section 7, Table S3. Single-crystal X-ray diffraction analysis confirmed that the main product is cluster (1). The resulting data set, consisting of 89 points, was provided to each method and served as the initial database and training ground for their subsequent exploration.

As described above, we studied three different methods: a machine-learning algorithm approach, human experimenters and random experiments as a baseline (Figure 3). Each method followed the same protocol: 1) analyze the dataset of the previous experiments, 2) specify 10 new experiments to execute, 3) receive a crystal/no-crystal information for each of the requested experiments. The process is repeated 10 times for a total of 100 experiments. At each iteration, all data collected previously are integrated in the decision process for

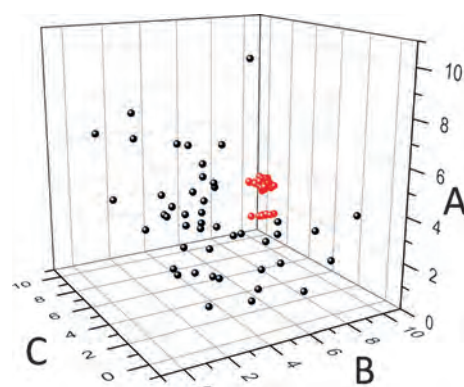


Figure 4. 3D graph of the initial set of data. A) $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ 1 M and $\text{Ce}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ 0.1 M (mL); B) HClO_4 1 M (mL); C) $\text{NH}_2\text{NH}_2 \cdot 2\text{HCl}$ 0.25 M (mL). Crystals red; non-crystals black.

generating the next set of 10 experiments. All experiments for all methods were executed on the platform under similar conditions. Each method was then evaluated in terms of strategies and overall exploration of the experimental space. The change in the ratio of the chemicals not only provides us with information of the experimental conditions that a given compound crystallizes in, which is the thermodynamic outcome of the most favorable conformation, but it also provides information on which combinations of reagents are not successful in the formation of compound (1). The latter can lead to the observation of other known compounds [see Supporting Information, section 8] or it can lead to the discovery of new chemical species. Figure 5 shows the difference in the crystal quality as we move from the central cluster of the initial data shown in Figure 4 to the outer regions of the crystallization boundaries.

The machine-learning approach is based on established active learning algorithm in classification problems^[26]

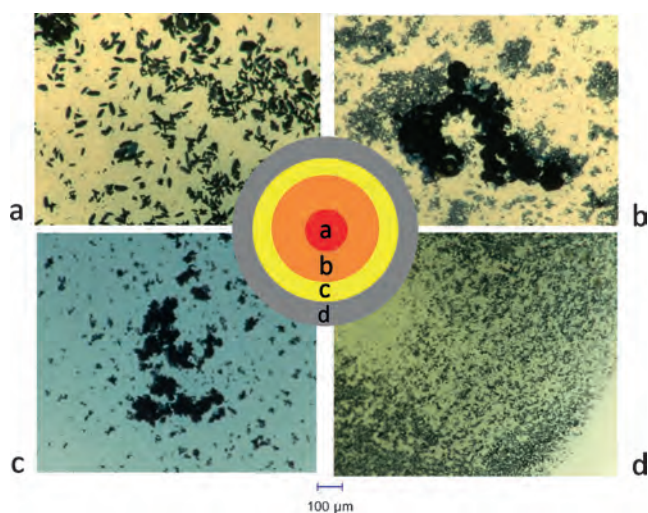


Figure 5. Change in the crystal quality of the crystallization sphere as we move from the initial data set (a), to the middle of the boundaries (b), and the outer edges of the boundaries (c). d) shows the precipitate which is observed when moving further away from the initial data set.

whereby the classification is the process of assigning labels (e.g. crystal/no-crystal) to regions of a parameters space (e.g. a range of experimental parameters) given only a few examples of known label-parameters instances. The quality of such model depends on the quality of the training data, the complexity and non-linearity of the process studied as well as its stochasticity. Methodologies have been developed to decide which experiments to perform in order to improve the model faster, called active learning procedures.^[4] Our algorithm is strongly inspired by such methods but adapted to the particular problem under investigation, see Supporting Information, part 6. Human experimenters were volunteers among PhD students in our group, all familiar with inorganic chemistry synthesis, and hence could be considered already to be “experts”. For the needs of this study, they were aware of the chemical formula of compound (**1**), the reagents, the reaction conditions, the platform and the initial set of data. They were not aware of the overall aim of comparing strategies among methods. Each human experimenter was instructed to develop their own strategy given the objective to identify the range of experimental conditions where compound (**1**) can be isolated. A baseline method is used as control, it consists of selecting experiments at random in the chemical space. This method is thus blind to both the initial and the subsequently collected crystallization information. To determine the differences between methods and their respective overall data acquisition performance we qualitatively analyzed trends in exploration strategies between each method. Then we quantified the number of experiments leading to crystals, as well as the extent of chemical space that was explored by each method. Using this data, it is then possible to compute the effectiveness of respective crystallization models and their predictive power.

The difference between human experimenters and the algorithm is illustrated in Figure S28. We can observe the stepwise exploration of human experimenters, starting from the known core of initial data provided and expanding outwards (S28, c). The limiting factor in their exploration is that when they perform an experiment that yields no-crystals, they stop after trying only a handful of experiments further than this point. Additionally, human experimenters tend to construct investigations that are more “Cartesian” shown in Figure S28,c, which reflects practical constraints. On the other hand, the algorithm follows a more “polar” approach around the initial set of data (Figure S28,b) and is not “disturbed” when crystals are not formed in the experiments, allowing a wider area to be covered.

Following the results from our experiments, in Figure 6 we plot the average explored volume of the experimental space as a function of the number of experiments performed. For the volume calculation, we compute the volume of the convex envelope of the experiments leading to crystals, see Supporting Information, part 10.2.2. We observe a large difference between algorithm and human experimenters. This can be explained because the algorithm is agnostic to the chemical environment and untied with prior chemical knowledge. Additionally, the algorithm is more “adventurous”, performing “jumps” in the chemical space straight into the believed boundaries between crystal and no-crystal. On the contrary,

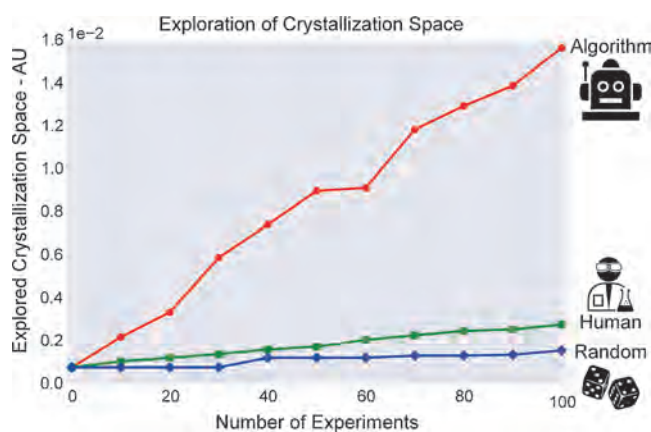


Figure 6. Explored crystallization space by the three methods. The exploration is computed as the volume of the convex envelope of the experiments leading to crystals [see Supporting Information, part 10.2.2].

human experimenters have drastically varied strategies depending on personal perceptions and biases of the particular chemistry involved. Figures S35–S38 provide a visual representation of the experiments selected by the two human experimenters. Other exploration metrics confirm the increased exploration of our active learning approach as well as the high variability between human experimenters, see Supporting Information, part 11.2.3.

Additionally, human experimenters can be baited by the absolute number of crystallization points they discover, disregarding how conservative or not their strategy can be. This point is important because a conservative strategy with many small “steps” of exploration can lead to many crystallization points but limits a wider exploration of the chemical space. For example, the second human experimenter performed an impressive 47 experiments leading to crystallization, out of 100 (Table 1, run 2). But to characterize our

Table 1: Total number of crystal points found for all runs of the three methods applied.

Method	Run 1	Run 2
Algorithm	27	32
Human experimenter	26	47
Random	4	2

system the breadth of exploration is more important than the absolute numbers of crystal formulation found. On the other hand, the algorithm, despite finding only 32 crystal experiments, revealed more about the chemical landscape. As shown in Figures S43 and S45, it was able to discover and explore a third of the crystallization region and spend a significant number of experiments exploring non-intuitive formulations.

Finally, the main interest of using a machine learning approach is that the data acquisition process can be informed and coupled with the objective—in this case, building an accurate crystallization map. We tested that hypothesis by

computing and testing a model of our crystallization system at each iteration and for each method (average of both runs) [see Supporting Information, section 10]. The quality of the prediction, that is, the percentage of time a crystal prediction is accurate, is expected to increase as more data are collected. Figure 7 shows that the machine-learning algorithm was able

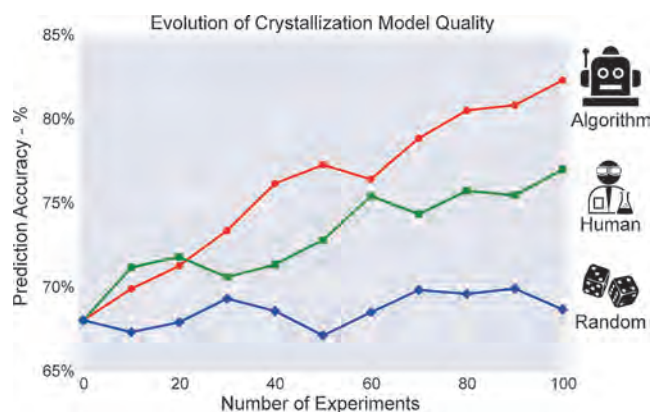


Figure 7. Average for the prediction accuracies between the classes of crystals and non-crystals for the three methods, using a RandomForest classifier [see Supporting Information, part 11.3].

to collect much better quality data and improved its classification accuracy from 68.1% (i.e. the initial prediction quality based on the initial database provided to all methods) to $82.4 \pm 0.7\%$. Whereas the humans showed a less significant improvement (from 68.1% to $77.1 \pm 0.9\%$) and the random method did not improve in accuracy (from 68.1% to $68.7 \pm 1.4\%$). The fact that the model computed using the human method improves less should be considered in light of the new data acquired, human experimenters simply did not collect as useful information as the algorithm method. This is even more striking with the random method that provided no additional information. These results were computed using a different classifier than the one used within the algorithm method, in order to verify that the data collected were not tied to the underlying assumptions of the algorithm. Results with other classifiers are presented in Supporting Information, section 10.3 and confirm the trends from Figure 7.

In previous studies the data used to characterize and model crystallization processes were extracted from databases of experiments intended for other research purposes. Here, we coupled the data acquisition process with the modeling of our system, with the aim of characterizing the crystallization boundaries of a new polyoxometalate cluster (compound **1**) in real time. Our “intelligent-throughput” is powerful since it combines both intuition from machine learning and reliable liquid handling, allowing the system to develop “chemical-intuition”; we hypothesize this could be a first step to developing a new approach we term chemical intelligence which uses machine learning to explore complex chemical systems. Using this approach we could observe significant differences in the strategies not only between the algorithm and the human experimenters but also between the two human experimenters. These differences can have a sig-

nificant impact in the ability of exploration and are heavily dependent on the personal and chemical biases of the individual. In the future, we aim to explore how to combine the intuition of the chemists with chemical intelligence to use human–machine teams to identify new phenomena and characterize new chemical systems. We have placed the code online^[27] in the hope that this will help others who wish to use machine learning in crystal chemistry.

Acknowledgements

We would like to thank Daniel Salley and Dr. Qi Zheng for their help with providing us the exploration data as the two human experimenters. Also, we would like to thank Dr. Laia Vilà-Nadal for the fruitful discussions and comments on early versions of this manuscript and Naomi A. B. Johnson for her artistic depiction of the robot in Figure 2. We gratefully acknowledge financial support from the EPSRC (Grant Nos EP/H024107/1, EP/I033459/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/K038885/1, EP/L015668/1, EP/L023652/1), the EC (projects 610730 EVOPROG, 611640 EVOBLISS, 318671 MICREAGENTS), ERC (project 670467 SMART-POM).

Conflict of interest

The authors declare no conflict of interest.

Keywords: cluster compounds · crystallization · human strategies · machine-learning · polyoxometalates

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 10815–10820
Angew. Chem. **2017**, *129*, 10955–10960

- [1] H. N. Miras, C. J. Richmond, D.-L. Long, L. Cronin, *J. Am. Chem. Soc.* **2012**, *134*, 3816–3824.
- [2] H.-Y. Zang, A. R. de la Oliva, H. N. Miras, D.-L. Long, R. T. McBurney, L. Cronin, *Nat. Commun.* **2014**, *5*, 3715.
- [3] R. J. Davey, S. L. M. Schroeder, J. H. ter Horst, *Angew. Chem. Int. Ed.* **2013**, *52*, 2166–2179; *Angew. Chem.* **2013**, *125*, 2220–2234.
- [4] B. Settles in *Active learning literature survey*, University of Wisconsin, Madison, **2010**, *52*, pp. 55–66.
- [5] S. L. Price, *Chem. Soc. Rev.* **2014**, *43*, 2098–2111.
- [6] J. G. P. Wicker, R. I. Cooper, *CrystEngComm* **2015**, *17*, 1927–1934.
- [7] A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig, A. Mar, *Chem. Mater.* **2016**, *28*, 7324–7331.
- [8] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73–76.
- [9] M. A. Korany, M. A. A. Ragab, R. M. Youssef, M. A. Afify, *RSC Adv.* **2015**, *5*, 6385–6394.
- [10] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. Anatole von Lilienfeld, K.-R. Müller, A. Tkatchenko, *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.
- [11] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, *Nat. Mater.* **2013**, *12*, 191–201.

- [12] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B* **2014**, *89*, 094104.
- [13] M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland, B. Meredig, *APL Mater.* **2016**, *4*, 053213.
- [14] M. H. S. Segler, M. P. Waller, *Chem. Eur. J.* **2017**, *23*, 5966–5971.
- [15] A. Müller, C. Beugholt, H. Bögge, M. Schmidtman, *Inorg. Chem.* **2000**, *39*, 3112–3113.
- [16] W. Xuan, A. J. Surman, H. N. Miras, D.-L. Long, L. Cronin, *J. Am. Chem. Soc.* **2014**, *136*, 14114–14120.
- [17] A. Müller, Y. Zhou, L. Zhang, H. Bögge, M. Schmidtman, M. Dressel, J. van Slageren, *Chem. Commun.* **2004**, *18*, 2038–2039.
- [18] L. Cronin, C. Beugholt, E. Krickemeyer, M. Schmidtman, H. Bögge, P. Kögerler, T. K. K. Luong, A. Müller, *Angew. Chem. Int. Ed.* **2002**, *41*, 2805–2808; *Angew. Chem.* **2002**, *114*, 2929–2932.
- [19] T. Yamase, E. Ishikawa, Y. Abe, Y. Yano, *J. Alloys Compd.* **2006**, *408–412*, 693–700.
- [20] E. Ishikawa, Y. Yano, T. Yamase, *Materials* **2010**, *3*, 64–75.
- [21] L. Zhang, Y. Zhou, R. Han, *Eur. J. Inorg. Chem.* **2010**, *17*, 2471–2475.
- [22] A. Müller, E. Krickemeyer, J. Meyer, H. Bögge, F. Peters, E. Diemann, S. Dillinger, F. Nommenbruch, M. Randerath, C. Menke, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2122–2123; *Angew. Chem.* **1995**, *107*, 2293–2295.
- [23] A. Müller, E. Krickemeyer, H. Bögge, M. Schmidtman, C. Beugholt, P. Kögerler, C. Lu, *Angew. Chem. Int. Ed.* **1998**, *37*, 1220–1222; *Angew. Chem.* **1998**, *110*, 1278–1281.
- [24] A. Müller, E. Beckmann, H. Bögge, M. Schmidtman, A. Dress, *Angew. Chem. Int. Ed.* **2002**, *41*, 1162–1167; *Angew. Chem.* **2002**, *114*, 1210–1215.
- [25] I. D. Brown in *Structure and Bonding in Crystals, Vol. II* (Eds.: M. O’Keeffe, A. Navrotsky), Academic Press, New York, **1981**.
- [26] D. D. Lewis, W. A. Gale in *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, Springer, New York, **1994**, pp. 3–12.
- [27] The code developed for this work is available online here: https://github.com/croningp/crystal_active_learning for modification and development.

Manuscript received: June 5, 2017

Accepted manuscript online: June 25, 2017

Version of record online: August 3, 2017