Unlocking Phase Diagrams for Molybdenum and Tungsten Nanoclusters and Prediction of their Formation Constants

Enric Petrus[a] and Carles Bo*[a,b]

[a]   Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology

(BIST), Av. Països Catalans 16, 43007 Tarragona, Spain.

[b]   Departament de Química Física i Inorgànica. Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007

Tarragona, Spain

Abstract

Understanding and controlling aqueous speciation of metal oxides are key for the discovery and development of novel materials, and challenge both experimental and computational approaches. Here we present a computational method, called POMSimulator, which is able to predict speciation phase diagrams (Conc. vs pH) for multi-species chemical equilibria in solution, and which we apply to molybdenum and tungsten isopolyoxoanions (IPAs). Starting from the MO_4 monomers, and considering dimers, trimers, and larger species, the chemical reaction networks involved in the formation of $[H_{32}Mo_{36}O_{128}]^{8-}$ and $[W_{12}O_{42}]^{12-}$ are sampled in an automatic manner. This information is used for setting up $\sim 10^5$ speciation models, and from there, we generate the speciation phase diagrams, which show an insightful picture of the behavior of IPAs in aqueous solution. Furthermore, we predict the values for 107 formation constants for a diversity of molybdenum and tungsten molecular oxides. Among these species, we could include several pentagonal shaped species and very reactive tungsten intermediates as well. Last but not least, the calibration employed for correcting the *DFT* Gibbs energies is remarkably similar for both metals, which suggests that a general rule might exist for correcting computed free energies for other metals.

Introduction

Self-assembly processes of discrete metal oxide nanoclusters are driven by a combination of many factors, the pH, the ionic force, and the presence of additional species being the main players. Although in this chemistry there are not well stablished rules as in organic chemical synthesis,[1] stable anionic metal-oxo compounds have been characterized continuously for almost two centuries. One of the first polyoxometalate (POM) examples was the molybdenum α-Keggin synthesized by Berzelius in 1826[2]. Since then, POMs have extended throughout the periodic table, thus we can find examples with lanthanides[3], actinides[4] and other transition metals[5]. This chemical variety has been accompanied by a broad range of topologies: molybdenum blue wheels[6,7], keplerates[8,9] and polyperoxouranates[10,11]. Metal-oxo compounds also present different applications in relevant fields such as biochemistry[12–15], catalysis[16–21] and nuclear reprocessing[22,23].

Hitherto, polyoxometalates chemistry has largely succeeded in discovering new compounds and novel applications. However, it is becoming more and more apparent that a deeper understanding of metal-oxo speciation, and thus of the self-assembly process, is needed.[24] Currently, different techniques are used to study POMs in solution: electronic[25] and vibrational[26] spectroscopy, nuclear magnetic resonance[27], mass-spectrometry experiments[28], small-angle X-ray scattering[29] and kinetic investigations[30]. Despite that, the experimental determination of unstable metal-oxo equilibrium constants is proving to be difficult. To the best of our knowledge, there is no information yet on the thermodynamic constants of pentagonal-based units $\{M_xM_6\}$, which are a recurrent motif for providing curvature and sphericity to metal-oxo nanoclusters.

Speciation of metal oxides in solution is determined by a myriad of chemical equilibria that coexist in the reaction medium. Therefore, the cornerstone for predicting the relative abundance of species is to elucidate the chemical reaction network (CRN)[31,32] along with its dependency on physical conditions: temperature, ionic strength and concentration. Quantum mechanical calculations are consolidated as an effective and accurate method for predicting reaction mechanisms through the analysis of the potential energy surfaces (PES).[33] In fact, computing static energy profiles based on chemical intuition has been the workhorse of quantum chemistry for the past three decades.[34] For example, a group additivity method, based on ab initio calculations, can describe the aluminum cluster speciation and its Pourbaix diagrams.[35] Alternatively, metadynamics simulations have also achieved great success, for instance, in explaining the aggregation process of the Lindqvist structure in concordance with mass-spectrometry experiments.[36–38] However, both static and dynamic approaches strongly rely on human guidance in either locating the transition states (TS) or determining the collective variables. Because of this, the analysis of large-scale PES, purely based on human intuition and in a non-automated fashion, is time-consuming and error-prone.[39]

The study of coupled reactions requires a meticulous determination of all the chemical reactions involved. Indeed, how to build a robust and scalable reaction network is a subject of discussion in the literature.[40,41] First of all, it is important to choose a general-purpose molecular representation for the chemical species. Broadbelt *et al.* showed that molecular graphs are a straightforward option in which atoms are assigned as nodes and chemical bonds as edges.[42] Moreover, the parallelism between chemistry and graph theory can be further extended to chemical reaction networks.[43] Therefore, graphs provide a unified framework to both represent chemical compounds and ease the management of the CRNs. There are numerous

examples of this methodology in very different chemical processes such as: prebiotics[44], hydroformylation[45], water-shift[46] and combustion[47] reactions.

We recently introduced the basics of a new methodology aimed at dealing with the multi-species equilibria involved in the formation of metal-oxo clusters in aqueous solution from first-principles and in an automated manner.[48] In that first implementation, we considered a small set of small molybdates (up to octamolybdate) and started by capturing raw results derived from Density Functional Theory (DFT) calculations, which were transformed into molecular graphs based on the topological properties of the charge density. In the next step, our method performed heuristic searches to determine the CRNs connecting all species in the set. Finally, the reaction network was used to set up multiple speciation models, which were solved. We found that the DFT derived formation constants were systematically overestimated, so we came up with a calibration procedure that enabled constructing the speciation diagram (% species vs pH, at fixed metal concentration) .

Herein we present the first full application of our new method, called POMSimulator, towards the *in-silico* prediction of speciation phase diagrams (Conc. vs pH) of molybdenum and tungsten medium-sized IPA (isopolyanions) nanoclusters. On the one hand, we have considered a large set of molybdates, including pentagonal $\{Mo_xMo_6\}$ and medium sized $\{Mo_{36}\}$ nanoclusters. On the other hand, we have expanded this methodology to tungsten IPAs. A general overview of the workflow followed in this work is depicted in Scheme 1. We start by optimizing the geometries and compute the harmonic frequencies for all the metal-oxo compounds. Then POMSimulator, (1) creates the CRN and estimates the free reaction energy of each transformation, (2) sets up and solves all the possible speciation models, (3) calibrates the DFT free reaction energies, ΔG^{DFT}, using experimental data, and (4) predicts speciation phase diagrams and not yet reported thermodynamic constants.

Results and Discussion

We have started describing a class of anionic molecular oxides which contain a unique type of transition metals, also known as IPAs. On the one hand, the molecular set of polyoxomolybdates (POMos) is formed by a total of 72 compounds with the general formulas $[H_zMo_pO_{3p+1}]^{z-2}$, $[H_zMo_pO_{3p+2}]^{z-4}$, $[H_zMo_pO_{3p+3}]^{z-6}$, pentagonal based structures $\{Mo_xMo_6\}$ and the largest existing IPAs: $[H_{16+z}Mo_{18}O_{65}]^{z-6}$ and $[H_{32+z}Mo_{36}O_{128}]^{z-8}$, abbreviated as $\{Mo_{18}\}$ and $\{Mo_{36}\}$. On the other hand, the set of polyoxotungstanates (POTs) is formed by 50 species with the general formulas: $[H_zW_pO_{3p+1}]^{z-2}$, $[H_zW_pO_{3p+2}]^{z-4}$, $[H_zW_pO_{3p+3}]^{z-6}$. It also includes medium-sized structures, such as $[H_zW_{11}O_{40}]^{z-14}$, $[H_zW_{12}O_{42}]^{z-12}$ and $\alpha\text{-}[H_zW_{12}O_{40}]^{z-8}$, abbreviated as $\{W_{11}\}$, $\{W_{12}\}$ and $\alpha\text{-}\{W_{12}\}$. In both groups, multiple protonation states are considered, and a detailed overview of the complete molecular set can be found in Figure S1. The molecular geometries have been optimized using a DFT methodology that is described in the Supporting Information. A dataset collection[49] is available in the ioChem-BD repository.[50]
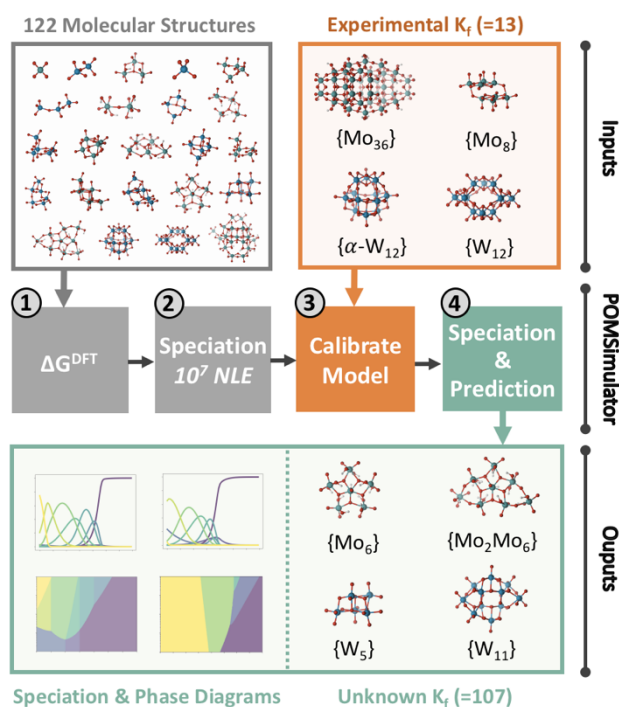
Scheme 1. General overview of POMSimulator workflow. **(1)** Compute DFT Gibbs energies of metal oxides and determine reaction free energies (195 reactions in total) through heuristic searches. **(2)** Solve systems of non-linear equations (NLE). Our parallelized code needs 48 hours in a 28-core CPU to solve $1 \cdot 10^7$ systems of equations. **(3)** Calibrate DFT intrinsic error. Eight experimental formation constants (K_f) for molybdenum (e.g., $\{Mo_{36}\}$, $\{Mo_8\}$) and five for tungsten (e.g., $\{W_{12}\}$, $\{\alpha\text{-}W_{12}\}$) were used, respectively. **(4)** Generate speciation and phase diagrams using corrected DFT energies, and predict unreported thermodynamic constants.

In the first place, POMSimulator collects the results of the DFT calculations. It starts by capturing the optimized geometries of the metal oxides and transform them to molecular graphs. Atoms are translated to nodes and bond critical points to edges. Additionally, chemical properties such as the total charge and the Gibbs free energy are stored. Once all the relevant information has been processed, POMSimulator is ready to create a CRN. To do so, the isomorphism property is used (i.e., morphological-like graphs) to define relationships between all the species. If two graphs are isomorphic, we subtract the atoms between both compounds to determine which chemical reaction could interrelate them: acid-base, condensation, addition or dimerization. Following this methodology, we automatically sample large reaction networks with their associated reaction free energies (ΔG^{DFT}). Then, the CRNs are employed to set up the speciation models in order to predict the behavior of polyoxometalates in solution. However, the larger the size of the CRN, the larger the number of multi-species chemical equilibria (MSCE). Furthermore, the MSCE consists of systems of non-linear equations, the resolution of which is slow and numerically instable. Thus, the bottleneck of POMSimulator is the resolution of the speciation models derived from the CRNs. In order to find a good balance between thoroughness and computational efficiency, we have come up with two solutions. Firstly, we have parallelized the solving step in order to run POMSimulator in a multi-core CPU. Secondly, POMSimulator applies heuristic searches to disregard meaningless reactions, which would otherwise increase the computational cost. Scheme S1 provides a complete overview of the protocol, as well as an exhaustive explanation of the implementation details.

Chemical reaction networks scale rapidly with the size of the molecular set. Because of this, CRNs tend to be dense and convoluted, as they contain an extensive number of transformations. One way of analyzing complex networks is to represent them as distributions. In our case, we have examined the thermochemistry of the CRNs employing violin plots. Figure 1A shows eight energy distribution plots, organized according to the type of reaction and IPA metal. Colors green and orange correspond to molybdenum and tungsten systems, respectively. Moreover, plots are also organized according to the reaction type: acid-base, condensation, addition or dimerization. Note that acid-base equilibria are key reactions in driving the nucleation of metal oxides. For instance, molybdenum and tungsten oxides aggregate at low pH, whereas niobium and tantalum only form large clusters at alkaline conditions. The other three types of reactions handle the growth and dissemble transformations. The median values, depicted in Figure 1A provide a first estimation of the thermochemistry within the CRNs. For example, the median rates in acid-base, condensation and dimerization are below 0 kcal·mol^{-1}, in contrast to addition reactions. Thus, acid-base, condensation and dimerization reactions are mostly spontaneous whereas additions are mostly endergonic. Notice that this trend is very similar for both metal-oxo families. Moreover, the distribution plots also provide information about the variances for each reaction type. POMos present a larger variance in condensation-type reactions whereas POTs show an even greater variance for dimerization-type reactions. Overall, Figure 1A only shows a purely energetic-based picture. Therefore, reactions which have a ΔG^{DFT} above 0 kcal·mol^{-1} might turn out to be favorable when pH or ionic strength are considered. Table S1 and Table S2 collect the complete list of chemical reactions, with their associated Gibbs reaction energies and reaction types.
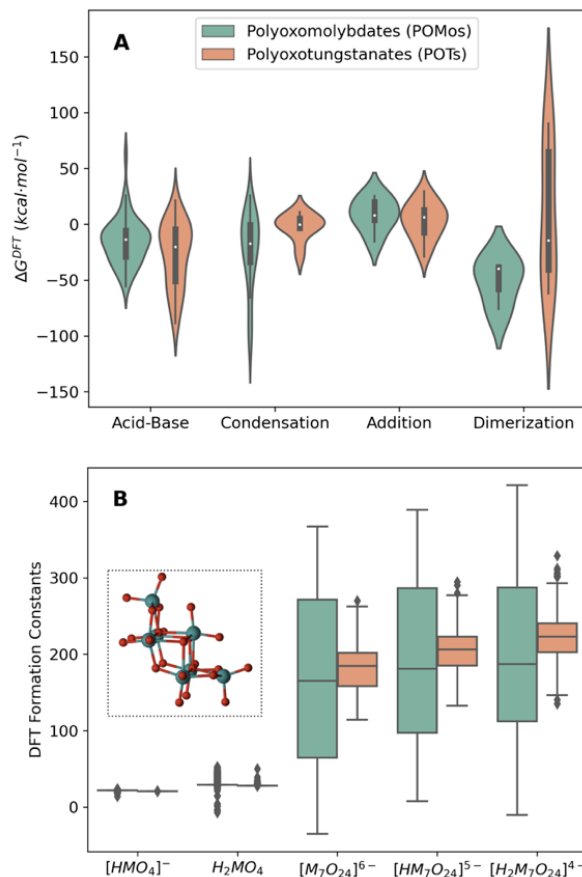
Figure 1. (A) Gibbs energy distribution plot (in kcal·mol^{-1}) for the main types of reactions present in the speciation models: acid-base, condensation, addition, and dimerization. **(B)** Box plots of the DFT formation constants for the monomer and heptamer. Samples for molybdenum and tungsten constants are of the order of $8 \cdot 10^4$ and $1 \cdot 10^3$ respectively. Inset plot of the $[M_7O_{24}]^{6-}$ tridimensional structure. Constants are referred to the formation reaction: $p[MO_4]^{2-} + q\text{H}+ \rightleftharpoons [H_zM_pO_m]^{n-} + (q/2+ p - z/2)\ H_2O$ Polyoxomolybdates are marked in green and –tungstanates in orange.

Despite the chemical likeness of molybdenum and tungsten, they have few structures in common. The six alike compounds are the monomers: $[MO_4]^{2-}$, $[HMO_4]^-$, H_2MO_4 and the heptamers: $[M_7O_{24}]^{6-}$, $[HM_7O_{24}]^{5-}$, and $[H_2M_7O_{24}]^{4-}$. Figure 1B shows the DFT formation constants of these structures (in box plots). Analogously to Figure 1A, green and orange correspond to the molybdenum and tungsten systems, respectively. Additionally, there is an inset plot which shows the tridimensional geometry of the heptamer $[M_7O_{24}]^{6-}$. The complete set of 122 formation constants computed with our software can be found in Figure S2. It is worth noting that molybdates formation constants show a larger variance than tungstanates due to the relative size of the molecular sets. A total of 116,644 speciation models were solved for molybdenum oxides, whereas 1,536 models were solved for tungsten oxides. Furthermore, POMSimulator computes an average of $8 \cdot 10^4$ constants for each molybdenum cluster (see Table S3), whereas it computes an order of magnitude less for tungsten compounds (see Table S4). Interestingly, the median values in Figure 1B vary if we compare tungsten and molybdenum oxides. POMSimulator predicts that the median values for $[H_nW_7O_{24}]^{n-6}$ are greater than for $[H_nMo_7O_{24}]^{n-6}$. The larger the formation constants, the more stable metal-oxo compounds are. At this stage, our method successfully describes the fact that formation of tungsten oxides is more favorable than molybdenum oxides. This evidence is in very good agreement with experiments.[51] The chemical explanation of this phenomenon is attributed to

the larger strength of the W-O bond compared to the Mo-O bond.[52] Even if this energy difference is relatively small, the accumulative effect in medium and large clusters gains importance.

There is big weakness in quantum mechanical assessment of the Gibbs energies of small solvated ions: the proton specie.[53] Although this error could be relatively small, the fact that the pH units are logarithmic amplifies the disturbing effect. The most common approaches to mitigate this error rely on setting up thermodynamic cycles and performing some sort of calibrations. For example, Grimme *et al.* studied 26 multi-proton organic compounds to estimate the pK_a employing a linear free-energy relationship.[54] Besides, Machuqueiro *et al.* recently developed a method to predict pK_a values of titratable residues in biomolecules.[55] These great achievements contrast with the more restricted developments in the area of inorganic chemistry.[56] We attribute this difference to the difficulties involved in treating distinctly highly anionic-charged compounds in solution in an equal manner. Within this context, our group has recently discovered that POMSimulator offers an exceptionally accurate calibration for small molybdenum oxides.[48] We found that our method computed formation constants which relate just linearly to some experimental ones. Thereby, we could correct the intrinsic DFT error with a simple linear scaling whilst retaining high accuracy (R^2=0.9997 and RMSE=0.43). Thanks to this successful calibration, POMSimulator computed speciation diagrams which were in excellent agreement with the experimental ones. In light of those promising results, we have now applied our method to a larger molybdenum oxide set and a tungsten oxide set. So far, we have observed that POMSimulator correctly predicts the relative stabilities of molybdenum and tungsten oxides (see Figure 2). Even so, the most determining aspect for assessing the overall performance is the accuracy of the calibration step.
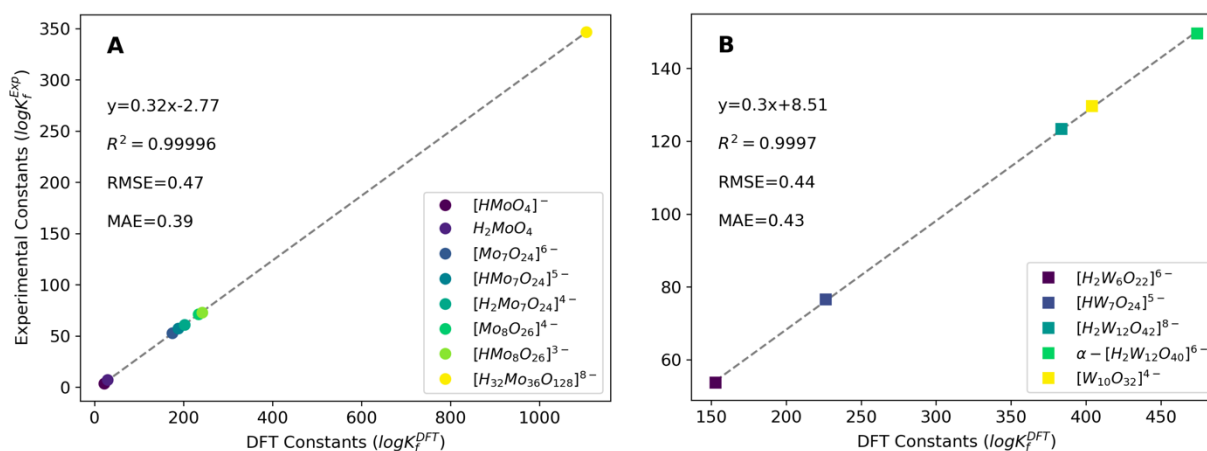

Figure 2. Linear regressions between $logK_f^{Exp}$ and $logK_f^{DFT}$, with the lowest RMSE (Root Mean Squared Error) and MAE (mean average) (mean average) values at 298.15K and 1 atm for **A** polyoxomolybdates (circles) and **B** –polyoxotungstanates (squares) at 1.0M and 0.25M NaCl ionic strengths, respectively.

Figure 2 shows that POMSimulator predicts the equilibrium constants of two different metal oxide systems with extraordinary accuracy. Figure 2A presents the linear correlation for molybdenum formation constants. The *x* axis corresponds to the constants computed with our method while the *y* axis corresponds to the experimental constants reported by Cruywagen.[57] The molybdenum compounds used for the calibration are the following: $[HMoO_4]^-$, H_2MoO_4, $[Mo_7O_{24}]^{6-}$, $[HMo_7O_{24}]^{5-}$, $[H_2Mo_7O_{24}]^{4-}$, $[Mo_8O_{26}]^{4-}$

, $[HMo_8O_{26}]^{3-}$ and $[H_{32}Mo_{36}O_{128}]^{8-}$. The eight molybdenum clusters are represented as circles in the plot. Notice that the linear regression in Figure 2A is the most accurate case out of the other 239 promising regressions depicted in Figure S3A. The selected regression presents the highest correlation coefficient, R^2=0.99996, and the lowest root mean squared, RMSE=0.47, and mean average, MAE=0.39, errors. These excellent parameters prove that POMSimulator can predict equilibrium constants of medium-size molybdenum clusters. Figure 2B shows the linear correlation for tungsten formation constants. The experimental values are extracted from the work of Rozantsev and coworkers.[58] The set of tungsten oxides used for the calibration are the following: $[H_2W_6O_{22}]^{6-}$, $[HW_7O_{24}]^{5-}$, $[H_2W_{12}O_{42}]^{8-}$, α-$[H_2W_{12}O_{40}]^{6-}$ and $[W_{10}O_{32}]^{4-}$. Figure 2B presents the best linear regression for the tungsten oxide system, even though there are ten additional regressions with an RMSE lower than 1.0 (Figure S3B). Tungsten oxides calibration offers outstanding accuracy with a correlation of R^2=0.9997 and prediction errors of RMSE=0.44 and MAE=0.43. Therefore, this calibration method not only applies to molybdenum oxides but also to tungsten oxides. In addition to this generalization, the prediction capabilities are remarkably high in both cases. Note that the values of the two slopes, which were obtained independently, are very similar (0.3 vs 0.32), but the value in the origin is not, indicating that this calibration method is metal dependent.

POMSimulator aims at providing an exhaustive picture speciation in solution. We have already demonstrated that the calibration step is close to optimal, thus we can obtain accurate equilibrium constants. Therefore, the next step is the resolution of the multi-species chemical equilibria employing the corrected thermodynamic constants. In our previous work, we presented speciation diagrams, which consist in concentration distribution curves which vary according to the pH. Nonetheless, we have noticed that this approach only shows a narrow window of the metal oxides speciation. As a matter of fact, the total metal concentration is a determining variable as well. For instance, there is no aggregation at very diluted concentrations, whereas there are clusters which only form at high concentrations. Thus, it is necessary to scan both the pH and the total concentration to build speciation phase diagrams. For the very first time, we are presenting phase diagrams for polyoxometalates aqueous speciation based on quantum mechanical calculations. Figure 3 shows the speciation phase diagrams computed by POMSimulator employing a rather fine grid. The number of systems of equations solved in Figure 3A and Figure 3B are $5 \cdot 10^9$ and $1 \cdot 10^8$, respectively. Note that not only the proton concentration (*x* axis) is in logarithmic scale but also the metal oxide concentration (*y* axis). We have done so to better represent details of the diagram at very low concentrations. Chemical species in Figure 3 are annotated with their stoichiometric formula, $[H_zM_pO_m]^{n-}$, and with the common notation in speciation studies, *(p,q)*, which refers to stoichiometric coefficients of the formation reactions.
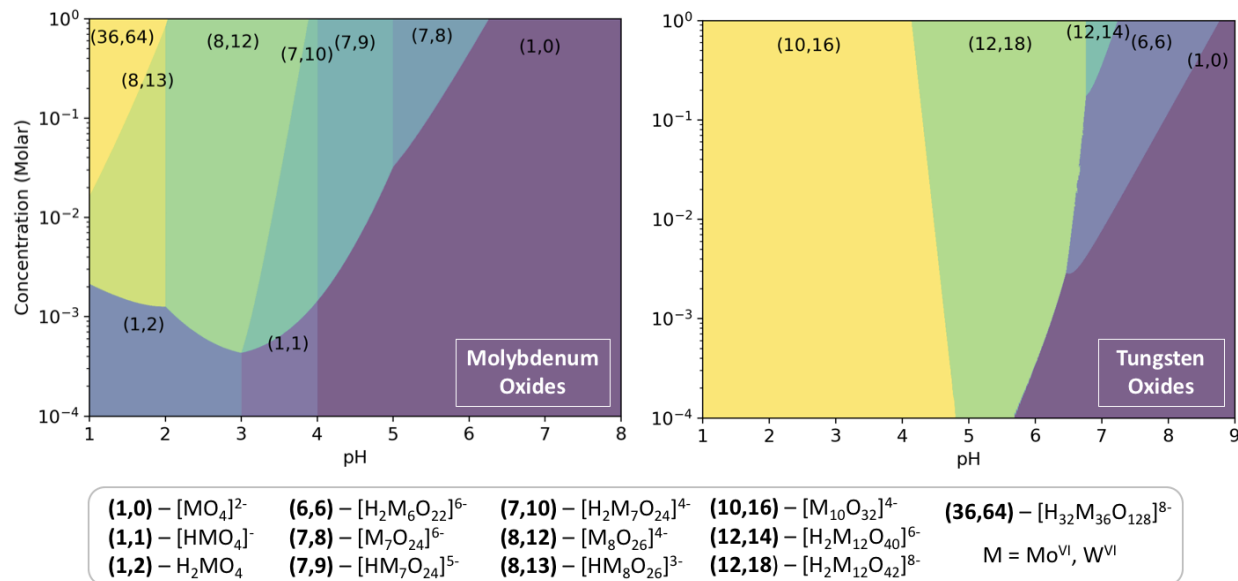
Figure 3. Speciation phase diagrams employing scaled ΔG^{DFT} values, at 298.15K and 1 atm. Left: polyoxomolybdates at I=1 M NaCl, with a grid of $5 \cdot 10^9$ solved models. Right: polyoxotungstanates at I=0.35M NaCl, with a grid of $1 \cdot 10^8$ solved models. Indices (p, q) refer to $p[MO_4]^{2-} + qH^+ \rightleftharpoons [H_zM_pO_m]^{n-} + (q/2+ p - z/2) H_2O$. Total metal molar concentration is fixed between 10^0-10^4M. Concentration axes are in logarithmic scale.

Figure 3 shows the speciation phase diagrams for molybdenum and tungsten oxides using the rescaled formation constants. At a pH higher than 7, the speciation is completely dominated by the monomer $[MoO_4]^{2-}$, regardless of the total concentration. At more acidic conditions, molybdenum monoxide either undergoes polymerization or protonation reactions. This variability depends on the total concentration: at lower values (< 10^3 M), it is less probable that two molecules will react due to the diffusion factor. Therefore, the only reactions that take place are the acid-base equilibria of the monomer to hydrogenmolybdate acid, $[HMoO_4]^-$, and molybdic acid, H_2MoO_4. However, at higher concentrations, the aggregation is favored and promotes the growth of larger clusters. In fact, the largest polyoxomolybdate, {Mo_{36}}, is only formed at concentrations larger than $5 \cdot 10^2$ M. Furthermore, the condensation zone (yellow-green area) includes other relevant oxides such as the octamolybdates $[Mo_8O_{26}]^{4-}$ and $[HMo_8O_{26}]^{3-}$, at a pH range between 1-3.5. At more neutral conditions, between 4 to 6, the heptamer oxides $[H_2Mo_7O_{24}]^{4-}$, $[HMo_7O_{24}]^{5-}$ and $[Mo_7O_{24}]^{6-}$ predominate. It can be observed that POMos speciation is extraordinarily rich both in number of species and topologies. Likewise, the speciation phase diagram for polyoxotungstanates, using the rescaled formation constants, is shown in Figure 3B. Tungsten monoxide is the most abundant compound at relative alkaline conditions (pH > 9). In contrast to POMos, the concentration for POTs does not have a sharp disturbing effect on the relative abundances. Despite that, in very diluted solutions, the polymerization process does not take place either. Decatungstate, $[W_{10}O_{32}]^{4-}$, is the predominant oxide for pH lower than 5. Between the 5 and 6.5 pH interval, the most abundant species are dodetungstanates $[H_2W_{12}O_{42}]^{8-}$ and to a lesser extend $[H_2W_{12}O_{40}]^{6-}$, which only exists at high concentrations (> 10^1M). Before dissembling to the monomer, the hexatungstanate $[H_2W_6O_{22}]^{6-}$ predominates at neutral pH.

Experimental techniques have made an immense effort to characterize metal oxides formation constants. Even so, there are many metal-oxo compounds that have proved to be difficult to isolate and characterize with current state-of-the-art techniques. For example, pentagonal molybdenum oxides, {Mo_xMo_6}, are a recurrent motif in medium and large clusters. These pentagonal structures are assigned as important building blocks for adding curvature and sphericity to keplerate architectures.[5] Despite that and as far as we know, no formation constants have been determined yet. Moreover, the formation of the $[W_{12}O_{42}]^{8-}$, through the lacunary form $[W_{11}O_{40}]^{12-}$, is still under discussion.[24] In view of this knowledge gaps, our method offers a straightforward approach to estimate equilibrium constants of transient intermediates. Table 1 collects the predicted values for the formation constants for some molybdenum and tungsten IPAs. Notice that Table 1 shows a single state of protonation for each cluster to ease the interpretation. Nevertheless, the complete set of 122 constants can be found in Table S7 and Table S8. To estimate these constants, the linear regressions, depicted in Figure 2, are used to calibrate the raw formation constants. Thanks to the excellent fitting, we can obtain accurate constants of unreported metal-oxo compounds. POMSimulator has determined 64 molybdenum equilibrium constants, but Table 1 only collects the 16 most relevant examples. This sample includes from small clusters (e.g., $[Mo_2O_7]^{2-}$, $[Mo_4O_{13}]^{2-}$) up to pentagonal alike structures (e.g., $[H_6Mo_5O_{22}]^{8-}$, $[H_{14}Mo_9O_{38}]^{8-}$). Furthermore, our method has computed 47 tungsten formation constants, but Table 1 only shows the 10 most representative samples. This set is formed by small clusters (e.g., $[W_3O_{10}]^{2-}$, $[W_5O_{16}]^{2-}$) and larger and more anionic species (e.g., $[W_5O_{19}]^{8-}$, $[W_{11}O_{40}]^{14-}$). It is worth noticing that, even with this new calculated data, all tungsten oxide formatin constants are larger than the molybdenum oxides analogues, thus indicating higher stability in accordance with experimental facts.

Table 1. DFT computed Formation Constants for polyoxomolybdates at 298.15K, 1 atm and 1M NaCl and polyoxotungstanates at 298.15K, 1 atm and 0.25M NaCl.

Molybdenum Oxides	Formation Constant [a,b]	Tungsten Oxides	Formation Constant [a,c]
$[Mo_2O_7]^{2-}$	11.68	$[W_2O_7]^{2-}$	23.42
$[H_6Mo_2O_{10}]^{2-}$	2.66	$[W_3O_{10}]^{2-}$	38.85
$[Mo_3O_{10}]^{2-}$	22.14	$[W_3O_{11}]^{4-}$	26.27
$[H_6Mo_3O_{14}]^{4-}$	3.27	$[W_4O_{13}]^{2-}$	54.48
$[Mo_4O_{13}]^{2-}$	33.59	$[W_4O_{15}]^{6-}$	19.53
$[H_6Mo_4O_{18}]^{6-}$	12.86	$[W_5O_{16}]^{2-}$	58.09
$[Mo_5O_{16}]^{2-}$	38.02	$[W_5O_{17}]^{4-}$	51.06
$[Mo_5O_{17}]^{4-}$	35.19	$[W_5O_{19}]^{8-}$	22.01
$[H_6Mo_5O_{22}]^{8-}$	18.82	$[W_6O_{20}]^{4-}$	68.21
$[Mo_6O_{20}]^{4-}$	41.84	$[W_{11}O_{40}]^{14-}$	64.00
$[H_{11}Mo_6O_{27}]^{7-}$	32.37		
$[Mo_7O_{23}]^{4-}$	55.36		
$[H_{12}Mo_7O_{31}]^{8-}$	44.37		
$[H_{13}Mo_8O_{35}]^{9-}$	51.47		
$[H_{14}Mo_9O_{37}]^{6-}$	68.61		
$[H_{14}Mo_9O_{38}]^{8-}$	62.51		

[a] Constants refer to the formation reaction: $p[MO_4]^{2-} + qH+ \rightleftharpoons [H_zM_pO_m]^{n-} + (q/2 + p - z/2) H_2O$. [b] Linear regression employed: y=0.32x-2.77 (RMSE=0.47, MAE=0.39) using Cruywagen *et al.* constants[57] for the calibration. [c] Linear regression employed: y=0.30x+8.51 (RMSE=0.44, MAE=0.43) using Rozantsev *et al.* constants[58] for the calibration.

Hitherto, we have observed that computing equilibrium constants with quantum mechanical methods involves a small but cumbersome systematic error. In our initial studies,[48] we could overcome this obstacle by applying a simple linear relationship (y=0.3x-3.1). However, we then had no evidence whether this former linearity was circumstantial or could be generalized to larger and more different systems. By comparing the slope (m) and intercept (b) parameters of the former and present (Figure 2A) works, and if we round the decimal places to the tenths, we can observe that both equations are identical. Therefore, we now have the proof that our previous regression was already a very good guess. This match shows that the method presented here can predict the simultaneous multi-equilibria of the whole family of IPA molybdates. By turning our attention to tungsten oxides, we have demonstrated that POTs formation constants can be accurately rescaled as well (Figure 2B). Remarkably, we have observed that the slope parameter (m=0.3) is the same in tungsten and molybdenum regressions. Besides, the intercept is positive for tungsten oxides (b=8.5) whereas it is negative for molybdenum oxides (b=-0.3). Thus, molybdenum and tungsten calibration equations share the same slope but differ in the intercept. This discovery is exceptionally interesting since it suggests that an heuristic could be found to further generalize that linear relationship to other metals.

Conclusions

The present work presents a robust methodology for dealing with complex multi-species chemical equilibria in solution in an automated manner, which allows predicting speciation phase diagrams and formation constants of transient species that are out of reach experimentally. This first complete application of the new method to the full set of molybdenum and tungsten IPAs demonstrates its potential for understanding and predicting metal oxides speciation in aqueous solution. Concentrations, pH and the ionic force play a key role in the nucleation and growth processes of larger clusters, and our method enables to take all of them into account.

Starting with a diverse set of molecular species at different protonation states, our method creates molecular graphs firstly using just quantum mechanical derived raw data (basically the electronic charge density and Gibbs free energies). From there, reaction networks are deduced and subsequently employed for setting up the speciation models, thus hundreds of thousands of non-linear equations that are solved relatively fast. The values of the equilibrium constants computed with our method show that polyoxotungstanates are more stable than polyoxomolybdates, which is in very good agreement with experimental facts. We predict the formation constants of, so far, 107 unreported compounds. For example, we have determined formation constants of the pentagonal structures: $\{Mo_6\}$, $\{MoMo_6\}$, $\{Mo_2Mo_6\}$ and $\{Mo_3Mo_6\}$ and reactive tungsten intermediates such as $\{W_5\}$ and $\{W_{11}\}$.

The unique capability to build speciation phase diagrams, a feature with which our code has been equipped, provides a powerful tool for guiding experiments. We expect that this kind of plots will help to optimize the experimental conditions towards the synthesis of a desired cluster. Molybdenum and tungsten IPAs, as those considered here, are well known species and thus not the holy grail in polyoxometalates chemistry. Other cations and anions play a fundamental role in the growth of larger structures for which trial and error approaches still prevail. Although the present work proves that the multi-equilibrium concept can be extended further, including kinetics equations in our simulation protocols is the next challenge.

Furthermore, and following common practice to rescale the DFT Gibbs free energies when computing acid/base equilibrium constants, we observed a strong linear dependence between computed and experimental formation constants, which allows us to calibrate the thermodynamic constants. The linear relationships exhibit exceptionally high correlations (R^2=0.9999) and low errors (RMSE=0.45), both beyond the current benchmark. These results provide a solid ground to describe the systems included in this study.

Last but not least, we have noticed some insightful aspects about the calibration method. On the one hand, the actual linear regression of POMos is remarkably close so it extends what we reported previously. This fact is highly relevant since it adds robustness to the calibration scale. Moreover, tungsten oxides present a regression with the very same slope (m=0.3) and different intercept. This trend suggests that a universal calibration might exist that could correct the ΔG^{DFT} error for various systems. More experiments are currently underway to investigate this hypothesis further and expand the applicability of POMSimulator throughout the periodic table. Since molecular metal oxides are often viewed as inorganic polymers, we envisage that this kind of treatment could be applied to other condensation and addition reactions, as those in organic polymerization reactions.

References

1   B. Mikulak-Klucznik, P. Gołębiowska, A. A. Bayly, O. Popik, T. Klucznik, S. Szymkuć, E. P. Gajewska, P. Dittwald, O. Staszewska-Krajewska, W. Beker, T. Badowski, K. A. Scheidt, K. Molga, J. Mlynarski, M. Mrksich and B. A. Grzybowski, *Nature*, 2020, **588**, 83–88.

2   J. J. Berzelius, *Ann. Phys.*, 1826, **82**, 369–392.

3   V. Das, R. Kaushik and F. Hussain, *Coord. Chem. Rev.*, 2020, **413**, 213271.

4   T. Auvray and E. M. Matson, *Dalt. Trans.*, 2020, **49**, 13917–13927.

5   L. Vilà-Nadal and H. N. Miras, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, 2020, pp. 1–37.

6   A. Müller and C. Serain, *Acc. Chem. Res.*, 2000, **33**, 2–10.

7   E. Garrido Ribó, N. L. Bell, W. Xuan, J. Luo, D.-L. Long, T. Liu and L. Cronin, *J. Am. Chem. Soc.*, 2020, **142**, 17508–17514.

8   A. Müller, E. Krickemeyer, H. Bögge, M. Schmidtmann and F. Peters, *Angew. Chemie Int. Ed.*, 1998, **37**, 3360–3363.

9   J. Lin, N. Li, S. Yang, M. Jia, J. Liu, X.-M. Li, L. An, Q. Tian, L.-Z. Dong and Y.-Q. Lan, *J. Am. Chem. Soc.*, , DOI:10.1021/jacs.0c06582.

10 P. C. Burns, K. A. Kubatko, G. Sigmon, B. J. Fryer, J. E. Gagnon, M. R. Antonio and L. Soderholm, *Angew. Chemie - Int. Ed.*, 2005, **44**, 2135–2139.

11 M. Dembowski, C. A. Colla, P. Yu, J. Qiu, J. E. S. Szymanowski, W. H. Casey and P. C. Burns, *Inorg. Chem.*, 2017, **56**, 9602–9608.

12 A. Bijelic, M. Aureliano and A. Rompel, *Angew. Chemie - Int. Ed.*, 2019, **58**, 2980–2999.

13 L. Vandebroek, H. Noguchi, K. Kamata, J. R. H. Tame, L. Van Meervelt, T. N. Parac-Vogt and A. R. D. Voet, *Chem. Commun.*, 2020, **56**, 11601–11604.

14 J. Moons, F. Azambuja, J. Mihailovic, K. Kozma, K. Smiljanic, M. Amiri, T. Cirkovic Velickovic, M. Nyman and T. N. Parac-Vogt, *Angew. Chemie*, 2020, **132**, 9179–9186.

15 G. Sciortino, M. Aureliano and E. Garribba, *Inorg. Chem.*, 2021, **60**, 334–344.

16 M. Blasco-Ahicart, J. Soriano-Lopez, J. J. Carbo, J. M. Poblet and J. R. Galan-Mascaros, *Nat. Chem.*, 2018, **10**, 24–30.

17 N. V. Maksimchuk, I. D. Ivanchikova, G. M. Maksimov, I. V. Eltsov, V. Y. Evtushok, O. A. Kholdeeva, D. Lebbie, R. J. Errington, A. Solé-Daura, J. M. Poblet and J. J. Carbó, *ACS Catal.*, 2019, **9**, 6262–6275.

18 S. Greiner, B. Schwarz, M. Ringenberg, M. Dürr, I. Ivanovic-Burmazovic, M. Fichtner, M. Anjass and C. Streb, *Chem. Sci.*, 2020, **11**, 4450–4455.

19 P. Gobbo, L. Tian, B. V. V. S. Pavan Kumar, S. Turvey, M. Cattelan, A. J. Patil, M. Carraro, M.

Bonchio and S. Mann, *Nat. Commun.*, 2020, **11**, 41.

20   M. J. Craig, R. Barda-Chatain and M. García-Melchor, *J. Catal.*, 2021, **393**, 202–206.

21   J. Zhong, J. Pérez-Ramírez and N. Yan, *Green Chem.*, 2021, **23**, 18–36.

22   T. L. Spano, A. Simonetti, L. Corcoran, P. A. Smith, S. R. Lewis and P. C. Burns, *J. Nucl. Mater.*, 2019, **518**, 149–161.

23   H. Traustason, N. L. Bell, K. Caranto, D. C. Auld, D. T. Lockey, A. Kokot, J. E. S. Szymanowski, L. Cronin and P. C. Burns, *J. Am. Chem. Soc.*, 2020, **142**, 20463–20469.

24   N. I. Gumerova and A. Rompel, *Chem. Soc. Rev.*, 2020, **49**, 7568–7601.

25   A. Ruiz de La Oliva, V. Sans, H. N. Miras, D. L. Long and L. Cronin, *Inorg. Chem.*, 2017, **56**, 5089–5095.

26   M. Marianski, J. Seo, E. Mucha, D. A. Thomas, S. Jung, R. Schlögl, G. Meijer, A. Trunschke and G. Von Helden, *J. Phys. Chem. C*, 2019, **123**, 7845–7853.

27   M. S. Centellas, M. Piot, R. Salles, A. Proust, L. Tortech, D. Brouri, S. Hupin, B. Abécassis, D. Landy, C. Bo and G. Izzet, *Chem. Sci.*, 2020, **11**, 11072–11080.

28   M. J. Hülsey, G. Sun, P. Sautet and N. Yan, *Angew. Chemie Int. Ed.*, 2020, **59**, 2–12.

29   M. Nyman, *Coord. Chem. Rev.*, 2017, **352**, 461–472.

30   H. N. Miras, C. Mathis, W. Xuan, D.-L. Long, R. Pow and L. Cronin, *Proc. Natl. Acad. Sci.*, 2020, **117**, 10699–10705.

31   G. Oster and A. Perelson, *IEEE Trans. Circuits Syst.*, 1974, **21**, 709–721.

32   D. T. Gillespie, *J. Phys. Chem.*, 1977, **81**, 2340–2361.

33   I. Funes-Ardoiz and F. Schoenebeck, *Chem*, 2020, **6**, 1904–1913.

34   J. N. Harvey, F. Himo, F. Maseras and L. Perrin, *ACS Catal.*, 2019, **9**, 6803–6813.

35   L. A. Wills, X. Qu, I.-Y. Chang, T. J. L. Mustard, D. A. Keszler, K. A. Persson and P. H.-Y. Cheong, *Nat. Commun.*, 2017, **8**, 15852.

36   L. V. Nadal, A. Rodríguez-Fortea, L. K. Yan, E. F. Wilson, L. Cronin and J. M. Poblet, *Angew. Chemie - Int. Ed.*, 2009, **48**, 5452–5456.

37   L. Vilà-Nadal, E. F. Wilson, H. N. Miras, A. Rodríguez-Fortea, L. Cronin and J. M. Poblet, *Inorg. Chem.*, 2011, **50**, 7811–7819.

38   J. M. Cameron, L. Vilà-Nadal, R. S. Winter, F. Iijima, J. C. Murillo, A. Rodríguez-Fortea, H. Oshio, J. M. Poblet and L. Cronin, *J. Am. Chem. Soc.*, 2016, **138**, 8765–8773.

39   A. L. Dewyer, A. J. Argüelles and P. M. Zimmerman, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, 1–20.

40   J. P. Unsleber and M. Reiher, *Annu. Rev. Phys. Chem.*, 2020, **71**, 121–142.

41   C. W. Coley, *Trends Chem.*, 2021, **3**, 133–145.

42   L. J. Broadbelt, S. M. Stark and M. T. Klein, *Ind. Eng. Chem. Res.*, 1994, **33**, 790–799.

43   G. Craciun and M. Feinberg, *IEE Proc. - Syst. Biol.*, 2006, **153**, 179.

44   A. Wolos, R. Roszak, A. Zadlo-Dobrowolska, W. Beker, B. Mikulak-Klucznik, G. Spólnik, M. Dygas, S. Szymkuc and B. A. Grzybowski, *Science (80-.).*, 2020, **369**, 1–12.

45   Y. Kim, J. W. Kim, Z. Kim and W. Y. Kim, *Chem. Sci.*, 2018, **9**, 825–835.

46   I. Ismail, H. B. V. A. Stuttaford-Fowler, C. Ochan Ashok, C. Robertson and S. Habershon, *J. Phys. Chem. A*, 2019, **123**, 3407–3417.

47   A. Jinich, B. Sanchez-Lengeling, H. Ren, J. E. Goldford, E. Noor, J. N. Sanders, D. Segrè and A. Aspuru-Guzik, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **117**, 32910–32918.

48   E. Petrus, M. Segado and C. Bo, *Chem. Sci.*, 2020, **11**, 8448–8456.

49   C. . Àlvarez, M.; M.; Graaf, C.; López, N.; Maseras, F.; Poblet, M, J.; Bo, *J. Chem. Inf. Model.*, 2015, **55**, 95–103.

50   E. Petrus and C. Bo, ioChem Data Collection, https://iochem-bd.iciq.es/browse/review-collection/100/28661/5a5952f3e6a7c291eaa56b38.

51   M. T. Pope and A. Müller, *Angew. Chemie Int. Ed. English*, 1991, 30, 34–48.

52   L. C. W. Baker and D. C. Glick, *Chem. Rev.*, 1998, **98**, 3–49.

53   P. G. Seybold and G. C. Shields, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2015, **5**, 290–297.

54   P. Pracht, R. Wilcken, A. Udvarhelyi, S. Rodde and S. Grimme, *J. Comput. Aided. Mol. Des.*, 2018, **32**, 1139–1149.

55   P. B. P. S. Reis, D. Vila-Viçosa, W. Rocchia and M. Machuqueiro, *J. Chem. Inf. Model.*, 2020, **60**, 4442–4448.

56   P. G. Seybold, *Mol. Phys.*, 2015, **113**, 232–236.

57   J. J. Cruywagen, in *Adv. Inorg. Chem.*, 1999, vol. 49, pp. 127–182.

58   G. M. Rozantsev and O. I. Sazonova, *Russ. J. Coord. Chem.*, 2005, **31**, 552–558.

Table of Content

POMSimulator is a software that extracts information from quantum mechanical calculations to study the aqueous speciation of polyoxometalates *in silico*.