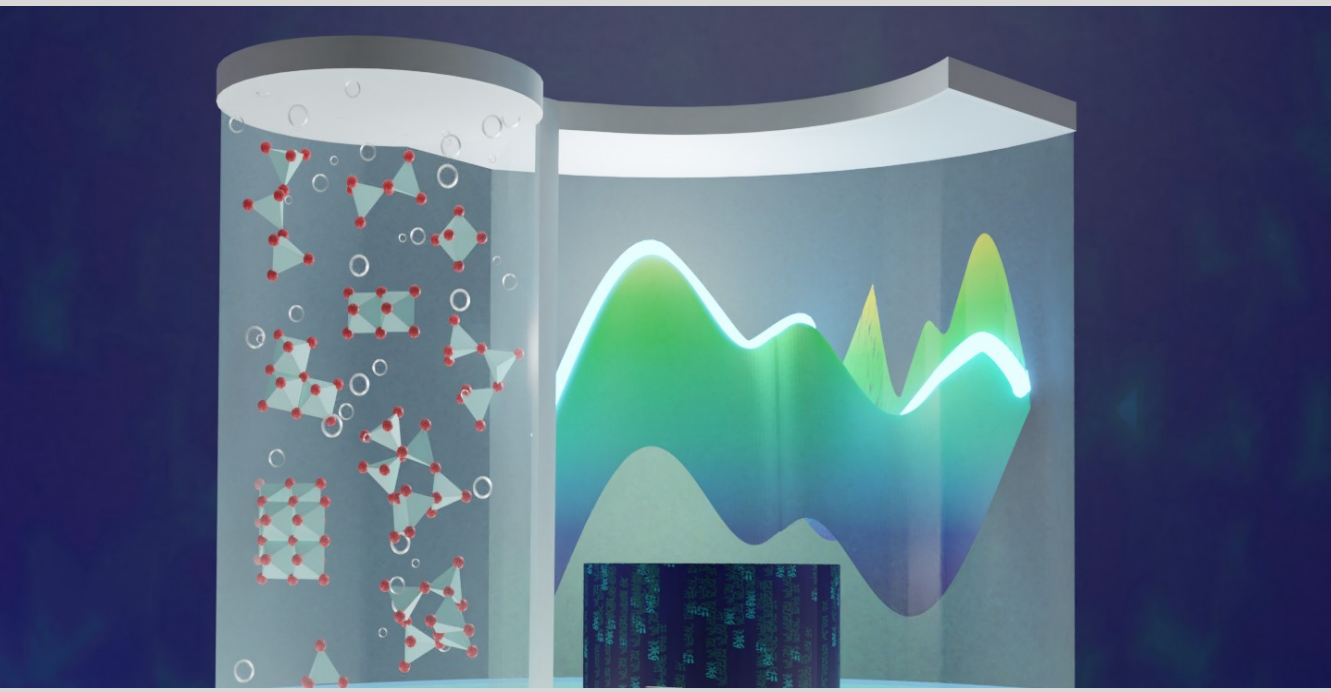
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES

Enric Petrus Pérez

UNIVERSITAT ROVIRA i VIRGILI

POMSimulator: A method for Understanding the Multi-Equilibria and Self-Assembly Processes of Polyoxometalates

ENRIC PETRUS PÉREZ



DOCTORAL THESIS
2022

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Enric Petrus Pérez

POMSimulator: A method for Understanding the Multi-Equilibria and Self-Assembly Processes of Polyoxometalates

Ph.D. Thesis

Supervised by Dr. Carles Bo Jané

Tarragona, 2022

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

UNIVERSITAT ROVIRA i VIRGILI

Dr. Carles Bo i Jané, investigador de l'ICIQ i professor de la Universitat Rovira i Virgili,

FAIG CONSTAR que aquest treball titulat "POMSimulator: A method for Understanding the Multi-Equilibria and Self-Assembly Processes of Polyoxometalates", que presenta Enric Petrus Pérez per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció a l'Institut Català d'Investigació Química (ICIQ) i que compleix els requisits per poder optar a la Menció Internacional.

Signat per Carles Bo Jané
23 Aug 2022 11:24:56 CEST
Certificat emès per: EC-SectorPublic
Número de serie:
837732019611299763846668177670313345572
Fundació Institut Català d'Investigació

El director de la tesi doctoral Dr. Carles Bo i Jané
Tarragona, 23 d'agost 2022

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Well done is that action of doing which one repents not later, and the fruit of which one, reaps with delight and happiness.

— Buddha, The Dhammapada [Verse 68]

Dedicada a la meva família

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Acknowledgements

I would like to thank my supervisor, Prof. Carles Bo, for giving me the chance of pursing my PhD in his research group. It has been a very enriching and pleasant experience. I warmly remember when I came for the first time to ICIQ as an undergraduate student six years ago. Back then I knew (practically) nothing related to academic research, and even computer programming. Thanks to Carles I had the opportunity to learn many new skills which I think they will be useful in the coming future. Moltes gràcies de tot cor.

Also, I am very grateful of having met many smart and friendly workmates. Their insights have made the difference on countless occasions. Moreover, I acknowledge the work of the anonymous reviewers who have suggested multiple improvements to our publications. On the emotional part, I wish to thank Buddhism for teaching me to remain calm and serene in my daily life.

Finally, I would like to thank my family for their unconditional support. Without their help it would have been definitely impossible to write this thesis. As the Spanish philosopher Ortega y Gasset stated: *"yo soy yo y mis circunstancias"*.[a]

[a] José Ortega y Gasset, Meditaciones del Quijote, Madrid, 1914.

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Funding Agencies

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Publications

Petrus, E.; Segado, M.; Bo, C. Nucleation Mechanisms and Speciation of Metal Oxide Clusters. *Chem.Sci.*, **2020**, 11, 8448-8456. doi.org/10.1039/D0SC03530K; Preprint Available at chemrxiv.12639269.v1

Petrus, E.; Bo, C. Unlocking Phase Diagrams for Molybdenum and Tungsten Nanoclusters and Prediction of their Formation Constants. *J.Phys.Chem.A.*, **2021**, 125, 23, 5212-5219. doi.org/10.1021/acs.jpca.1c03292; Cover Art Included. Preprint Available at chemrxiv.14292119

Petrus, E.; Segado, M.; Bo, C. Computational Prediction of Speciation Diagrams and Nucleation Mechanisms: Molecular Vanadium, Niobium and Tantalum Oxide Nanoclusters in Solution. *Inorg.Chem.*, **2022**, 61, 35, 13708–13718. doi.org/10.1021/acs.inorgchem.2c00925; Preprint Available at chemrxiv-2022-2pt95

Martin, N. P.; Petrus, E.; Segado, M.; Arteaga, A.; Zakharov, L. N.; Bo, C.; Nyman, M. Strategic Capture of the $\{Nb_7\}$ Polyoxometalate. *Chem.-A.Eur.J.*, **2019**, 10580–10584. doi.org/10.1002/chem.201902770

Rahman, T.; Petrus, E.; Segado, M.; Martin, Nicolas P.; Palys, Lauren N.; Rambaran, Mark A.; Ohlin, C Andre.; Bo, C.; and Nyman, M. Predicting the Solubility of Inorganic Ion Pairs in Water. *Angew.Chem.Int.Ed.*, **2022**, 19, 61, e202117839

Abstract

In this thesis we have developed *POMSimulator*, a novel computational method that predicts the aqueous speciation of molecular oxoclusters as well as the formation mechanism. In our first work we tested this methodology with a model system: the octamolybdate, $[Mo_8O_{26}]^{4-}$.[1] We reported the speciation diagrams of molybdates with high accuracy, and we proposed a reaction mechanism considering the acid-base equilibria. Next, we extended this methodology to all the isopolyoxomolybdates and -tungstates.[2] We described the formation of larger clusters such as: $[H_{32}Mo_{36}O_{128}]^{8-}$, $[W_{12}O_{42}]^{12-}$, $[W_{12}O_{40}]^{8-}$, and $[W_{10}O_{32}]^{4-}$. Besides, we introduced phase diagrams in our methodology to obtain a better overview of the speciation at different concentrations. Recently, we have applied *POMSimulator* to polyoxovanadates, -niobates, and tantalates.[3] We showed that accurate formation constants could be computed for the three metal-oxo systems. We proposed a vanadate intermediate, $[V_5O_{14}]^{3-}$, which could be involved in the interconversion of the decavanadate, $[V_{10}O_{28}]^{6-}$, and the metavanadates, $[V_nO_{3n}]^{n-}$ (n=4,5,6). Furthermore, we reported the speciation diagram of niobates including paramount clusters such as $[H_9Nb_{24}O_{72}]^{15-}$ and $[Nb_7O_{22}]^{9-}$. In con-

trast, we found no evidence of the formation of the tantalum analogues. Nonetheless, we observed the formation of the decatantalate, $[Ta_{10}O_{28}]^{6-}$, even though it has not been synthesized yet due to its poor solubility in water. Overall, we believe that our method can find a very promising synergy with experimental chemistry.

Contents

1 Introduction **1**
1.1 Polyoxometalates 1
1.2 Applications 5
1.3 Objectives 13

2 Theoretical Background and Method Development **15**
2.1 Motivation 15
2.2 Graph Theory 18
2.2.1 Molecular Graphs 20
2.2.2 Isomorphism 27
2.2.3 Computational Complexity 28
2.2.4 Construction of the Chemical Reaction Network 34
2.3 Chemical Equilibrium 44
2.3.1 Speciation Models 47
2.3.2 Benchmark of Solvers 53
2.3.3 Thermodynamic Activity 55
2.3.4 Combinatorial Explosion 62
2.3.5 Batch Parallelization 69
2.3.6 Accuracy Threshold 70

2.3.7 Raw Speciation Diagrams 73
2.4 Linear Scaling 76
2.4.1 Formation Constants 78
2.4.2 Formation Mechanism 90
2.4.3 Acid Dissociation Constants 93
2.4.4 Universal Scaling 96
2.5 Speciation Parameters 100
2.5.1 Total Concentration 101
2.5.2 Ionic Strength 104
2.6 Summary 106

3 Molybdenum and Tungsten Isopolyoxoanions 111
3.1 Introduction 111
3.2 Speciation Results 114
3.3 Mechanistic Insights 123
3.4 Summary 128

4 Vanadium, Niobium and Tantalum Isopolyoxoanions 133
4.1 Introduction 133
4.2 Speciation Results 136
4.3 Mechanistic Insights 152
4.4 Summary 161

5 Conclusions 165

A Computational Details 169

B Accuracy Benchmark 171

C Chemical Reactions **175**

Bibliography **191**

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Chapter 1

Introduction

> Meaning is context-bound, but context is boundless.
>
> — Jonathan Culler, Literary Theory 2007

1.1 Polyoxometalates

Polyoxometalates (POMs) are molecular metal-oxide anions of early d^0 metals such as molybdenum (Mo^{6+}), tungsten (W^{6+}), vanadium (V^{5+}), niobium (Nb^{5+}), and tantalum (Ta^{5+}). This class of species was first described by Berzelius in 1826 who observed the formation of a pale yellow compound produced by the addition of ammonium molybdate to phosphoric acid.[4] Equation 1.1.1 shows the balanced reaction leading to the formation of the ammonium 12-molybdophosphate, $(NH_4)_3[PMo_{12}O_{40}]$, as written by the author. Almost forty years after Jean Charles Galissard de Marignac found and determined the composition of the 12-tungstosilicic acid ana-

logue, $H_4[SiW_{12}O_{40}]$. He described it according to equation 1.1.2.[5]

$$12[MoO_4]^{2-} + H_3PO_4 + 21H_3O^+ + 3NH_4^+ \rightleftarrows (NH_4)_3[PMo_{12}O_{40}] + 33H_2O \quad (1.1.1)$$

$$12[WO_4]^{2-} + SiO_3^{2-} + 26H_3O^+ \rightleftarrows H_4[SiW_{12}O_{40}] + 37H_2O \quad (1.1.2)$$

Although chemists succeeded in synthesizing the 12-tungstosilicic and 12-molybdophosphoric compounds, their respective molecular structures were not determined yet. It was in 1929 when Linus Pauling found that the central atom had to be phosphorus or silicon based on the ionic radii. The internal atom adopted a tetrahedral coordination whereas the external metal atoms adopted an octahedral coordination.[6] Few years later, H. F. Keggin finally established the structure of the 12-tungstophosphoric acid with a modern method at that time: X-ray diffraction (powder method).[7]

Figure 1.1 shows five structures of emblematic polyoxometalates with different range of sizes. For example, the smallest polyoxometalate is the Lindqvist cluster, $[M_6O_{19}]^{n-}$. It was discovered in 1952, and it consists of a superoctahedral structure with six metal centers (Figure 1.1a).[9] Then, the decametalate (Figure 1.1b) is also similar to the Lindqvist with the difference that it has four additional metal atoms. This type of cluster is characteristic of Group V polyoxometalates. The first examples were found for the decavanadate,[10] $[V_{10}O_{28}]^{6-}$, and decaniobate,[11] $[Nb_{10}O_{28}]^{6-}$, in the 1980s. In contrast, the decatantalate, $[Ta_{10}O_{28}]^{6-}$, was not discovered until 2008.[12] It is

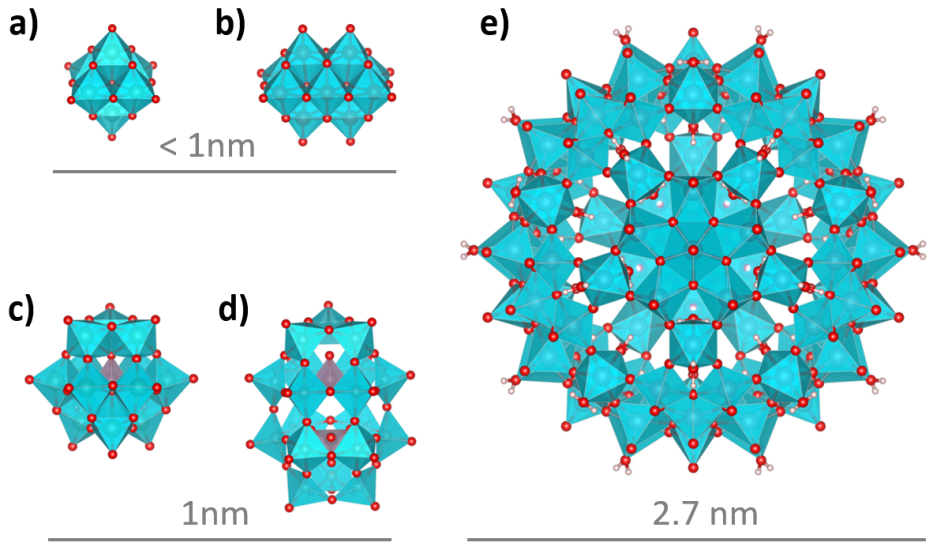


Figure 1.1: Polyhedral structures of five POM clusters (M: blue, O: red, X: yellow). a) Lindqvist anion $[M_6O_{19}]^{n-}$ b) Decametalate $[M_{10}O_{28}]^{n-}$ c) The Keggin structure, $[XO_4M_{12}O_{36}]^{n-}$ d) Wells-Dawson structure $[XO_{42}M_{18}O_{54}]^{n-}$ e) Keplerate-type structure, $\{Mo_{72}^{VI}Mo_{60}^{V}\}$ or in its compressed form $\{Mo_{132}\}$. Approximate diameters for the clusters are showed to provide a picture of the range of scales. Polyhedra generated with VESTA package.[8]

noteworthy that both the Lindqvist and the decametalate clusters represent the smallest polyoxometalates, with an approximate diameter of less than 1 nanometer. Next, we find the Keggin cluster mentioned in the paragraph above (Figure 1.1c). Another relevant cluster which is based on the Keggin structure is the Well-Dawson polyoxometalate. It was first synthesized in the early 1950s and it consists of two half-Keggin structures fused together (Figure 1.1d).[13]

It is remarkable that unlike the Lindqvist and decametalate, the Keggin and the Wells-Dawson have addenda atoms in their respect-

ive cores. Usually, these atoms are phosphorus, silicon, and germanium. This difference is commonly employed to separate polyoxometalates in two groups: isopolyoxoanions (IPAs) and heteropolyoxoanions (HPAs). This classification is well-established in the field, and in fact Michael Pope published in 1983 a quintessential book precisely entitled *Heteropoly and Isopoly Oxometalates*.[14] While IPAs include clusters composed of a unique metal oxide framework, HPAs comprise clusters with internal heteroatoms or heteroanions. Another notable family of polyoxometalates are the Keplerates. These spherical clusters with pentagonal motifs, $[\{(Mo)Mo_5\}_{12}\{L\}]$, receive the name of 'Keplerate' due to their similarity to Kepler's early model of the universe, named *Mysterium Cosmographicum*.

This type of POMs was first reported by Achim Müller and coworkers at the end of the past century.[15] Figure 1.1e shows the tridimensional structure of a Keplerate, $\{Mo_{132}\}$. An important feature is the mixture of molybdenum atoms at different oxidation states: Mo^{V} (d^1) and Mo^{VI} (d^0). This results in $\{Mo_{132}\}$ appearing in a dark brown color (aka., Mo-brown) because of the intervalence charge transfer between the Mo^{V} and Mo^{VI} centers. The ability to combine metal atoms with different oxidation states unlocks the formation of large clusters. The maximum exponent is the $\{Mo_{368}\}$ commonly referred to as *Blue Lemon* due to its elyptic shape.[16] It has a diameter size of 5.4 nm, thus being within the range of proteins and macromolecular science. Even so, the formation mechanism of highly-reduced POMs is very complex. As a consequence, multiple disciplines such as density functional theory, UV-vis, electron paramagnetic resonance spectroscopy, and small-angle X-ray scattering

spectra have been recently employed to study this phenomena.[17]

1.2 Applications

Polyoxometalates' broad range of topologies represents a major asset in strategic applications such as catalysis, biochemistry, energy materials and technology. An example of the current auge of POMs is reflected into the publication of numerous reviews in the recent years.[18–23] Figure 1.2 presents a general overview of POMs experimental research, as well as the computational techniques employed.

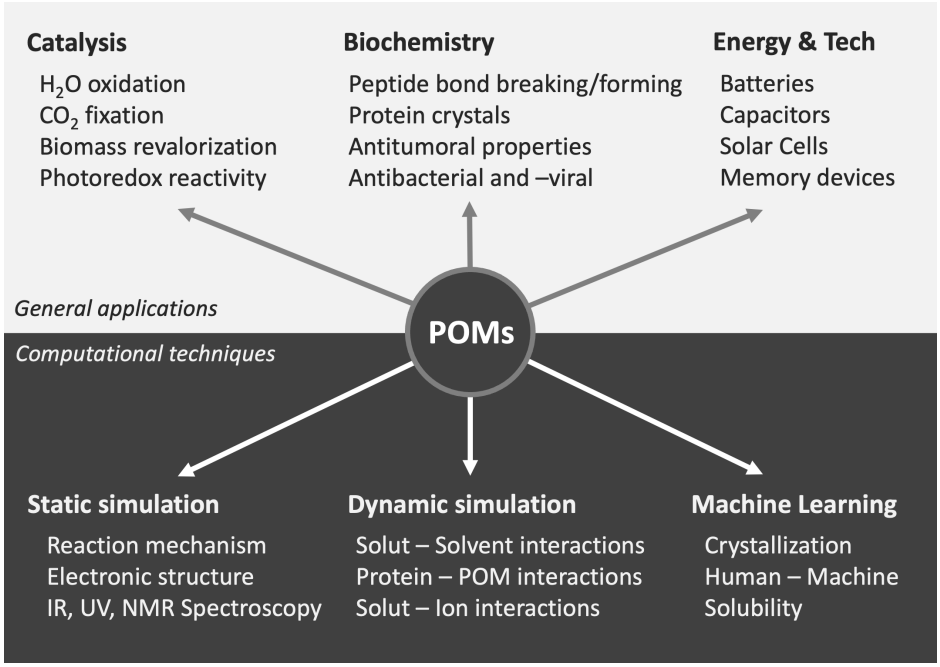


Figure 1.2: Top: general overview of polyoxometalates' main applications. Bottom: current computational techniques employed in this field.

The first key application is related to catalysis where POMs have been known to accelerate several important reactions. One common application is related to oxidation reactions as recent examples show for sulfur-containing compounds[24] and alkylbenzenes[25]. Within this application, water oxidation reactions are one of the most notorious examples. Blasco and coworkers reported a cobalt-phosphotungstate polyoxometalate that outperformed the catalytic activity of the state-of-the-art IrO_2 catalyst.[26] POMs also do catalyse CO_2 fixation reactions thanks to their base-catalyzed properties.[27] For example, the tetra-n-butylammonium tungstate, $TBA_2[WO_4]$, was reported to catalyze the formation of cyclic carbonates.[28]. Another relevant application of POMs' catalytic properties is found in biomass revalorization. For instance, polyoxometalates are able to break the β-1,4-glycosidic bond of cellulose through successive hydrolysis reactions. This results in the production of glucose with high yields.[29] Apart from ground-state reactivity, POMs have also been employed in photocatalytic reactions. Arguably the most popular POM-based photocatalyst is the decatungstate, $[W_{10}O_{32}]^{4-}$. It is employed in multiple transformations such as, hydrogen atom transfer reactions,[30] C-H functionalization,[31] and sulfinylations.[32] Nonetheless, there are other POM structures such as Keggin derivatives which have demonstrated to improve the photocatalytic H_2 evolution activity.[33]

The second area of application resides in biochemical transformations. For instance, POMs perform direct amide bond formation,[34], hydrolysis of a fully insoluble protein[35] and Adenosine 5-Triphosphate (ATP)[36]. Also related to this application, polyoxometalates have facilitated X-ray structure analysis of proteins.[37] As a matter of fact,

the Nobel Prize in Chemistry of 2009 was partly awarded to the structure characterization of ribosomal subunits with Wells-Dawson anions.[38] Furthermore, polyoxometalates also have antitumoral properties. For example, photothermal therapy relies on the cytotoxic heat generated by irradiating a polyoxometalate in blood stream to ablate tumor cells. Molybdenum-based POMs have been used as photothermal agents thanks to their ability to self-assemble into tens of nanometers and perform intratumoral accumulation.[39] Moreover, the decavanadate, $[V_{10}O_{28}]^{6-}$, has been described to inhibit the growth of several bacteria such as: *Mycobacterium tuberculosis*[40] and *Escherichia coli*[41]. Interestingly, the biocidal activity of POMs have also been used to remove *cyanobacteria* and *microalgae* responsible for the deterioration of art pieces located in humid and subterranean environments.[42] In addition, POMs have been successfully tested with several viruses, even though in less extent compared to bacteria and tumors. For example, polyoxotungstates have showed antiviral properties towards the Human Immunodeficiency Virus (HIV) through protease inhibition.[43] It is worth mentioning that *in vivo* studies of POMs' antiviral properties remain scarce. This is because of the inherent toxicity of polyoxometalates. Nevertheless, some studies have been favorably performed in mouses.[44]

The third field of application is connected to energy and technological materials. In this context we find energy storage devices such as batteries. Polyoxometalates have very fitting properties for undergoing the redox reactions associated to the charge and discharge cycles. As we observed in the previous section 1.1, metal centers can easily change their oxidation valence. One model example of POMs' capa-

bilities of storing electrons was published in 1976 when Jean Pierre Launay succeeded in reducing the metatungstate ion, $[H_2W_{12}O_{40}]^{6-}$, with 32 electrons and 32 protons.[45] More recently, molybdenum Keggin compound was reduced with 24 electrons and its structure was successfully determined with quantum mechanical methods.[46] Vanadium polyoxometalates have also found several applications in redox flow batteries as well.[47] Besides, POMs can also act as capacitors, and in fact their performance has been boosted when coupled to metal-organic frameworks.[48] Another energy-related application of polyoxometalates is perovskite solar cells. POMs act as low-cost, efficient and highly flexible chemical dopants for improving the performance of these devices.[49] Recently, a fully reduced $\{Mo^{V}_{60}\}$ synthesized at hydrothermal conditions also offered very promising light-absorbing capabilities.[50] In addition, POMs present interesting applications in memory devices. One of the pioneer articles was published by Busche and coworkers in 2014.[51] The authors proposed a Wells-Dawson analogue $[W_{18}O_{54}(SeO_3)_2]^{4-}$ as a candidate for realistic nanoscale flash memory. This application has gathered more attention in the last years due to the constraints in the miniaturization of the classical metal-oxide-semiconductor devices.[52,53]

In parallel to POMs' experimental research, computational chemistry has emerged as a reliable ally to explore and predict several polyoxometalates' properties.[54] This is a direct consequence of the growth in computational power, as well as the refining of quantum mechanics algorithms. Indeed, almost one century ago the famous quantum physicist Paul Dirac foretold the following: "*The underlying physical laws necessary for the mathematical theory of a large part*

of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation".[55] Sixty-nine years later Walter Kohn and John A. Pople received the Nobel Prize in Chemistry for precisely: *"the development of the density-functional theory and computational methods in quantum chemistry"*.[56] More concretely, the *approximate practical methods* foretold by Dirac became the Density Functional Theory (DFT) methods historically introduced by Walter Kohn and Lu Jeu Sham in their article published in 1965.[57]

Figure 1.3 illustrates three representative examples of computational techniques employed in the field of polyoxometalates. The first example corresponds to the exploration of the potential energy surface (PES) by means of static simulations. This technique has been the workhorse of quantum chemistry for the last decades.[58] Gibbs free energies of transition states (TS) and intermediates (Int) are calculated by solving the Schrödinger equation. Figure 1.3a shows a fictitious PES where the reaction path has been highlight with a glowing white curve. This same technique has been successfully applied to a broad variety of studies such as: formation of $[W_6O_{19}]^{2-}$and $[Mo_8O_{26}]^{4-}$,[59,60] disassembly of polyoxoniobate clusters,[61] and protein hydrolysis by Zr(IV)-containing metal.[62] Apart from reaction mechanisms, static simulations have also been favorably applied to determine the electronic structures of polyoxometalates. The

first electronic structures were computed for the decametalate and Lindqvist anions in the 1990s.[63,64] From then on, many publications have successfully addressed the electronic properties of inorganic crown ethers[65] and Keplerates[66] among many others. In addition, computational chemistry has reproduced and predicted the results of different types of spectroscopies. For example, ^{17}O Nuclear Magnetic Ressonance,[67] Infrared,[68] and Ultraviolet-visible.[69]

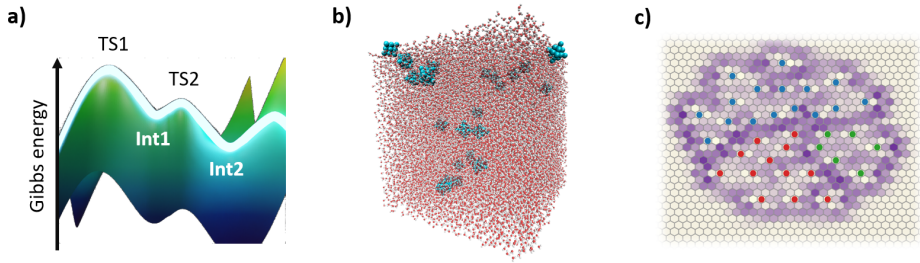


Figure 1.3: Computational techniques employed in the field of polyoxometalates: a) exploration of the potential energy surface b) molecular dynamic simulation c) Kohonen neural network.

Although the exploration of PES is very convenient, it does not account explicitly for the reaction time. For some properties the dependence on time is paramount, and thus another type of theoretical treatment is required. Molecular dynamics is an appropriate tool for studying the evolution of a given system in a time span. Figure 1.3b shows a common picture of a molecular dynamic simulation; water molecules solvate explicitly the Lindqvist anions. It usually consists of a cubic space, filled with molecules under study. Molecular dynamics can either apply *classical* or *quantum* mechanics, or a combination of both. Classical molecular dynamics have proved to appropriately describe polyoxometalate-protein interactions,[70] and

the solubility trends of molecular oxo-clusters (including the effect of the cation).[71] Furthermore, quantum molecular dynamics (aka. first-principle molecular dynamics, FPMD) and its mechanical molecular extensions (QM/MM) have also experienced very good results. For instance, by predicting: chemical redox potentials,[72] formation of Lindqvist derivatives,[73] and hydration structures of metal ions.[74]

Lately, machine learning applications have also found success in the field of polyoxometalates. The research group of Lee Cronin has been a pioneer in applying ML to the synthesis of POMs. More concretely, they trained a classifier algorithm (Support Vector Machine) which determined if crystallization would occur based on experimental parameters.[75] Once trained, the algorithm was able to explore a much larger crystallization space (6 times more than humans), and increase the crystallization prediction accuracy. More recently, the same research group proposed that the performance of the algorithm could be overcome if combined with human intuition.[76] This intuition-enabled ML algorithm outperformed (75.6%) the accuracy prediction of the algorithm (71.8%) and humans (66.3%). Another example of ML application resulted from our collaboration with May Nyman's and Andree Ohlin's research groups. In this case the algorithym focused on predicting the solubility trends of polyoxometalates.[77] We modeled compounds as σ-profiles[a] and then we employed a machine learning algorithm (Kohonen Neural Network) which performed the clustering. Figure 1.3c shows a conceptual plot of the neural network. Similar compounds are grouped together thus cre-

[a]The definition of a σ-profile is an histogram that quantifies the distribution of molecular surface segments that have a specific charge density named sigma.

ating three clusters depicted in blue, red and green colors. Once trained, the algorithm not only agreed with the experimental data, but also predicted the solubility behaviour of unreported compounds.

Despite the extensive and successful research carried out by computational chemists, one topic in polyoxometalates' field has been understudied in the recent years: the aqueous speciation. On the one hand, the resolution of multi-species chemical equilibria is a well-studied topic. There are renown software packages such as *Visual MINTEQ* which has been extensively applied to solutions in natural waters.[78] Furthermore, new Python-coded packages have also been developed for similar purposes in the recent years. For example, *PyEqulon*,[79] *pyEQL*,[80] and *Pytzer*[81]. On the other hand, these codes are exclusively focused on natural water systems. Therefore, their application to polyoxometalates is not straightforward. The main challenge is the tendency of POMs to nucleate into larger clusters. This feature is in clear contrast to other typical oxo-anions such as sulphate, carbonate, and nitrate. Consequently, modeling polyoxometalates' speciation involves the consideration of not only acid-base reactions, but nucleation reactions as well.

1.3 Objectives

The main goal of this thesis is to develop a computational method for studying the aqueous speciation and nucleation mechanism of polyoxometalates. This goal can be divided in five subgoals which are the following:

- Development of an algorithm for constructing a chemical reaction network.
- Development of an algorithm for setting up and solving speciation models.
- Testing the method with a model system.
- Fine-tuning the parameters of the method.
- Application of the method to polyoxometalate systems.

In terms of structure, this thesis is divided in two blocks. The first part (Chapter 2) deals with the theoretical development of the method. We describe the application of Graph Theory to chemical compounds and we detail the systems of equations related to the multi-species chemical equilibria. Besides, we introduce a common but important procedure in this methodology: the linear scaling of the equilibrium constants. The second part of this thesis collects the results of applying *POMSimulator* to two families of isopolyoxometalates. In Chapter 3 we focus on molybdates and tungstates, and in Chapter 4 we center on vanadates, niobates and tantalates. Both chapters have the same structure, which consists in a short

introduction, and the discussion of the speciation and mechanistic results. Next, we will detail the specific objectives and contents of each chapter.

Chapter 2 presents the development of the method. First, we introduce Graph Theory and its applications to chemistry. We explain the usage of molecular graphs and isomorphisms for constructing the reaction network. Next, we describe the implementation of speciation models and the challenges associated with the activity functions and the factorial growth of models. Then, we disclose the importance of the linear scaling and its potential universality for polyoxometalates.

Chapter 3 collects the results of applying *POMSimulator* to isopolyoxomolybdates and -tungstates. We describe the speciation and phase diagrams of medium sized clusters such as: $[H_{32}Mo_{36}O_{128}]^{8-}$, $[W_{12}O_{42}]^{12-}$, $[W_{12}O_{40}]^{8-}$, and $[W_{10}O_{32}]^{4-}$. Additionally, we provide formation constants for reactive intermediates such as the pentagonal molybdate, $\{\mathrm{Mo_6Mo}_n\}$, and the $[W_{11}O_{40}]^{14-}$. Moreover, we detail the formation mechanism of these clusters through reaction maps.

Chapter 4 gathers the results of applying *POMSimulator* to isopolyoxovanadates, -niobates, and -tantalates. We present the phase and speciation diagrams of paramount compounds such as: $[V_{10}O_{28}]^{6-}$, $[V_4O_{12}]^{4-}$, $[H_9Nb_{24}O_{72}]^{15-}$, $[Nb_7O_{22}]^{9-}$, $[Nb_{10}O_{28}]^{6-}$, $[Ta_6O_{19}]^{8-}$, and $[Ta_{10}O_{28}]^{6-}$. In addition, we present a unified reaction map where we compare the different reaction patterns of these three families of polyoxometalates.

Chapter 2

Theoretical Background and Method Development

> Scientists have to know how to hone into the right problem, and what the right complexity of problem is to tackle next.
>
> — Demis Hassabis, Game Changer 2018

2.1 Motivation

The self-assembly of polyoxometalates is still a subject under study, both from the experimental and computational standpoint. The inherent complexity of this topic stems from the large number of simultaneous chemical equilibria in solution. All the oxo-clusters are interconnected with other species, either by protonation, or by nucleation reactions. To make matters worse, the nucleation process strongly depends on several parameters: pH, ionic force, total con-

centration, temperature and pressure. As a matter of fact, there are oxo-clusters which only assemble in acid media, whereas others are insoluble at low pH. Thus, slight changes in the former parameters might result in the formation of different products. Nevertheless, in the past decades experimental researchers have made numerous breakthroughs in the field.[82] For example, synthesising giant oxo-clusters,[83] discovering niobium and tantalum reactivity,[84,85] and applying POMs as efficient catalysts[86] and batteries[87]. Analogously, computational studies, rooted in quantum mechanics, have also increased the chemical understanding of polyoxometalates.[54] For instance, by elucidating reaction mechanisms using the potential energy surface,[88] predicting the electronic structures,[66] and reproducing X-ray disorders with Boltzmann distributions[89]. However, few to no computational studies address the mechanism and the speciation simultaneously. Therefore, it was necessary to develop a new methodology that went beyond the classical energy-profile approach.

We have named this novel method *POMSimulator* because it is designed to simulate the nucleation of polyoxometalates. It has three cornerstone theories embedded in its foundations. Figure 2.1 shows a schematic overview of the program. First and foremost, we employed Quantum Mechanics (QM) to describe the physics of metal-oxo clusters. We relied on the Schrödinger equation and the electron density to optimize the geometries of the clusters, calculate their thermochemical energies, and analyze their electron density. The computational details are described in Appendix A. In second place, we had to construct the Chemical Reaction Network (CRN) in order to predict the nucleation processes. To ensure that the network

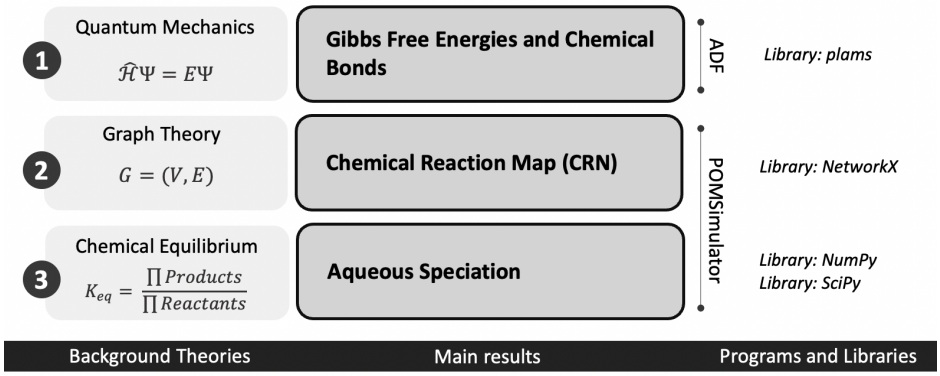


Figure 2.1: Overview of the theories, results, and libraries employed in *POMSimulator*. 1) Quantum mechanics results are generated with ADF 2) Graph Theory is employed to create a chemical reaction network 3) Chemical Equilibrium theory is applied to set up and solve the speciation models.

was robust and scalable, we employed Graph Theory.[a] This allowed a convenient simplification of molecules to graphs, and a systematic construction of the network. Last but not least, we had to model the multi-species chemical equilibria that underwent in aqueous solution. Thus, we implemented the Chemical Equilibrium equations that define the speciation models. Because these models are complex systems of equations, we depended on powerful numerical solvers.[b]

[a] NetworkX[90] library was used for constructing the CRN.

[b] *NumPy*[91] was used to ease the array manipulation, and *SciPy*[92] to solve the non-linear equations.

2.2 Graph Theory

The very first manuscript about graph theory was authored by the Swiss mathematician Leonhard Euler in 1736.[93] A graph *G = (N,E)* is defined as a set N of nodes (aka. vertices) and a set E of pairs of elements of N, which are the edges. This theory was initially developed to answer a popular conundrum in the Prussian city of Koningsberga. Locals wondered whether it would be feasible to make a journey through the city crossing all of the bridges, but without using any of the bridges twice (in graph theory terms, performing an Euler path). Figure 2.2 (left) shows a map of the city at that time; bridges have been highlighted with green rectangles. The city was organized in four islands and six bridges. In contrast, Figure 2.2 (right) shows a graph representation of Koningsberga. Bridges have been converted to edges, and islands to nodes. This transformation has the benefit of expressing the problem in a simple, yet convenient manner. According to the results extracted from the graph interpretation, Euler concluded that is was not possible to make the aforementioned journey. To perform an Euler Path, the graph must only have *two* odd nodes (i.e., a pair of nodes with an odd number of edges). However, there are *four* nodes that have an odd number of edges, therefore, it is not feasible to draw an Euler path. Even so, graph theory did not vanish after solving this problem, but kept on expanding until the present day.

In this thesis we employ the branch of graph theory which is strictly related to chemical systems: *Chemical graph theory*. Its first application dates back to 1758, when the Scottish chemist, William Cullen, employed it for his lectures. Under the strong influence of

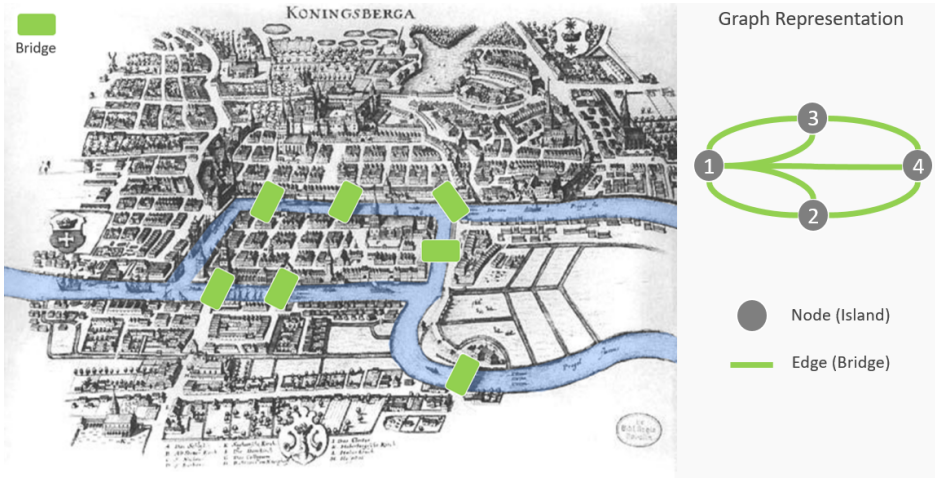
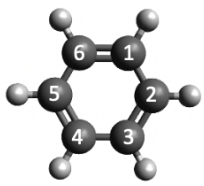


Figure 2.2: Left: Map of the city of Koningsberga (actual Kalingrad). Bridges have been marked with green rectangles. Right: Graph representation of Koningsberga where nodes and edges correspond to islands and bridges, respectively.

Newtonian physics, he used graphs to represent *(fictitious)* gravitational attractions between substances. Although it had no physical basis, he was a pioneer in applying Euler's recent theory to a chemical problem. Regrettably, the lecture notes were strictly used for teaching and thus, none were ever published.[94] From then on, the development of chemical graph theory was not very steady. There were some periods of flourishing ideas, followed by drops in interest. It was not until the twentieth century, after the breakthrough of quantum mechanics, when chemical graph theory regained popularity. One remarkable example, which relates both fields, was found by Ruedenberg in 1954.[95] He proved that the Hückel's determinant of benzene was isomorphic to the adjacency matrix of the molecular graph. Figure 2.3 shows the molecular structure of benzene, along

with the two matrices. The α variable is the Coulomb integral for a carbon atom, the β variable corresponds to the resonance integral for a pair of carbon atoms, and the E variable is the electronic energy. The eigenvectors of both matrices are practically identical, and only differ by a normalizing factor. Although widely recognized at present, this discovery showed that the topology of a molecule determined its properties. In the 1990s, chemical graph theory was already well-established in the chemistry paradigm. Several books were published by proficient chemists: Bonchev and Rouvray,[96] Trinajstic,[97] Gutman and Polansky.[98] Nowadays, the most general formalism is the *chemical graph*, which serves as a convenient model for representing chemical objects.[97] There are two main chemical entities which are frequently represented as chemical graphs: molecules and chemical reactions. We will see both in section 2.2.1 and 2.2.4, respectively.



$$\begin{pmatrix} \alpha-E & \beta & 0 & 0 & 0 & 0 \\ \beta & \alpha-E & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha-E & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha-E & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha-E & \beta \\ 0 & 0 & 0 & 0 & \beta & \alpha-E \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Hückel Matrix **Adjacency Matrix**

Figure 2.3: Benzene molecule with the corresponding Hückel, and Adjacency matrices (only carbon atoms are considered).

2.2.1 Molecular Graphs

A *molecular graph* is the graph interpretation of a molecular structure. Thus, atoms are assimilated to nodes, and chemical bonds

are converted to edges. Figure 2.4 shows two case examples. First, the monomer $[H_2MO_3]^{2+}$ is transformed to its graph representation.[c] The molecular graph can either be depicted in its two-dimensional form, or in a pair of vectors: nodes (N) and edges (V). The 2D-representation unveils an interesting feature: spatial coordinates are disregarded. Therefore, graphs can be plotted in indistinct forms, as long as the set of nodes and edges are kept constant. This is a very committing simplification since the spatial position of the atoms is paramount in quantum mechanics. Indeed, atom coordinates are the main input in the Schrödinger's equation. Nevertheless, in this thesis we do not employ graphs to predict electronic properties, but to take advantage of their topological, and morphological properties. The second example in Figure 2.4 also shows a common small oxo-cluster: the dimer $H_2M_2O_7$. The procedure for generating the molecular graph is the same as for $H_2M_2O_7$. However, the second case shows that the size of the edges array scales rapidly. Thus, for the largest compounds, the amount of pair of nodes becomes huge. This fact has important consequences when calculating topological properties. We will discuss this issue in section 2.2.2.

Although molecular graphs are one of the most popular options to encode molecules, there are other worth-mentioning alternatives. Nowadays, there is a broad variety of representations in the literature, and far from slowing down, it increases every year.[99] This is mainly due to the rise of machine learning algorithms (neural networks, in special) whose performance strongly depends on the nature of the chemical representation. We will briefly mention *three* examples

[c] Unless otherwise stated, M refers to a general metal of valence +6.

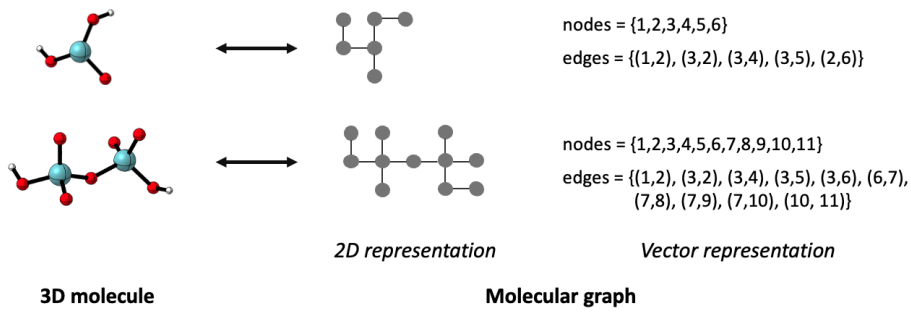


Figure 2.4: Relationship between three-dimensional molecules and molecular graphs.

which are within the state-of-the-art. First, the simplified molecular-input line-entry system (SMILES), is a minimalist but powerful representation to describe chemical species using short ASCII (American Standard Code for Information Interchange) strings.[100] SMILES have been the workhorse of cheminformatics for several decades. Nonetheless, researchers have recently showed that they have excellent prospects as descriptors for neural network architectures.[101] Secondly, the electron density is also employed as a descriptor for machine learning algorithms.[101] Molecule atoms are voxelized into discretized cubic grids, and used as training data.[102] Frequently, data follows a process of featurization to optimize the performance of the predictive algorithm. The third and final example concerns the Coulomb matrix; a tensor consisting of atomic energies and internuclear Coulomb repulsion operator. This representation has been successfully used to predict molecular atomization energies using a nonlinear statistical regression.[103] Overall, the choice of the molecular representation strongly depends on the purpose of the study. Herein, we focus

on determining chemical reactions, and thus, the graphs' topological properties fit perfectly with our needs.

Returning the discussion to molecular graphs, these representations are applied to: property prediction, molecule generation and automated-discovery of chemical reaction networks. If we focus on the first application, researchers have repeatedly proved that graph-based neural networks (GNNs) offer excellent results. In fact, GNNs have predicted a variety of properties, such as: bond dissociation energies for charged molecules,[104] solvation free energies,[105], acid constants for organic compounds,[106] and chemical reactivity[107]. Molecular graphs are also employed for mapping molecules in databases, specially in the field of drug discovery.[108] Training sets are huge, with data sets overpassing by far the million molecules.[109] Last but not least, graphs are also employed for generating and manipulating reaction networks.[110] Indeed, graph theory grants powerful tools to tackle complex reaction pathways, such as sugar-related reactions.[111]

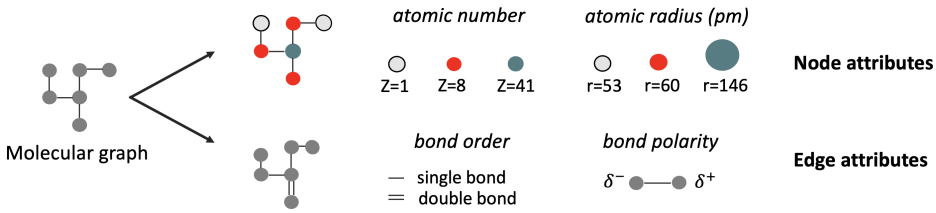


Figure 2.5: Attributes for a molecular graph. Top: node attributes such as atomic number and atomic radius. Bottom: edge attributes such as bond order and bond polarity.

So far, we have restricted to the basic definition of a molecular graph. Nonetheless, graphs can encode more information apart from the default in nodes and edges. The additional features of a mo-

lecular graph are called *attributes*. Figure 2.5 depicts that a single graph can have several attributes both for nodes and edges, respectively. Figure 2.5 (top) shows the representation of $[H_2MO_3]^{2+}$ with coloured nodes. Because we can set an atomic number to a node, the graph can adopt a more chemical-looking aspect. Oxygen and hydrogen nodes are represented in red and grey colours respectively, whereas the metal node is showed in a greenish colour. Beyond making graphs more readable, atomic numbers are key attributes when applying morphological properties. We want to make sure that we are not only looking for topological resemblance, but chemical as well (i.e., atom types). In the current implementation of *POMSimulator*, the atomic numbers are introduced as attributes in the molecular graphs. Additionally, the Gibbs free energy, the enthalpy, the charge, and an internal index (for keeping track of each graph) are also stored as attributes. These data will be used for calculating the reaction energies, and the activity coefficients. Still, there are other chemical properties which might be useful in a different context. For example, if molecular graphs have to regard steric effects, atomic radius should be included. A similar analogy could be made with the oxidation state of the metal atoms. Herein, there are no reduction/oxidation reactions, therefore it is not necessary to specify the oxidation state. Furthermore, edges can also encode important attributes. As stated before, the molecular graphs that we have employed do not consider any edge attribute. Despite that, there are some cases which might require a different choice. For example, if the system undergoes hydrogenation or elimination reactions. In this instance, bond orders should be regarded to track the transformation

of simple, double, and triple bonds. Moreover, if we want to study bond dissociation energies, the bond polarity should be regarded as well. This attribute would encode information of the electron density distribution throughout chemical bonds. Overall, attributes are an excellent feature for improving graphs models, and adapt them to the requirements of the problem.

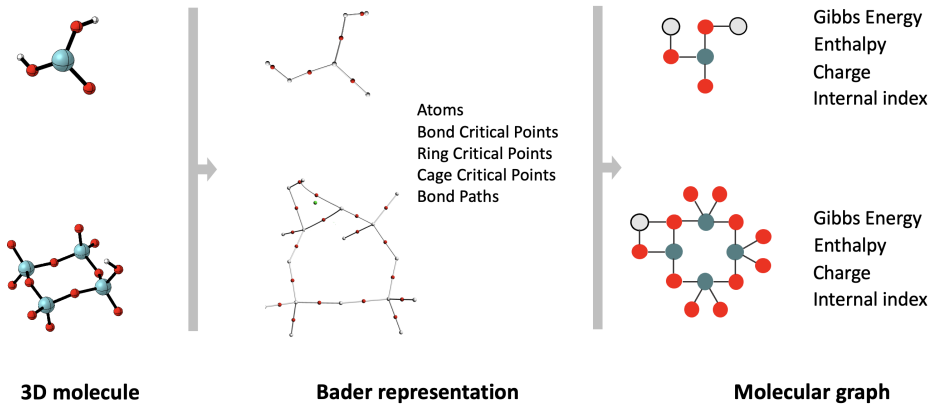


Figure 2.6: Molecular graphs based on the Bader representations. Attributes are listed on the right: Gibbs energy, enthalpy, charge and internal index.

We have detailed how to define molecular graphs, but there is an important concern left to discuss. The most crucial aspect of graphs are the edges, since they define the topology of the object. Thus, when converting chemical bonds to edges, we have to make sure that the chemical connectivity is completely right. For most of the cases, the transformation is simple and straightforward, however, there are some critical examples. For instance, hydrogen bonds are not regarded (explicitly) in the connectivity of a molecule. This is a relevant challenge since polyoxometalates are very pH-dependent,

resulting in abundant hydrogen bond interactions. To adress this aspect, we have relied on *Bader's Quantum Theory of Atoms in Molecules* (QTAIM).[112] This theory has been broadly applied to determine atom-atom interactions, whether covalent, or non-covalent. The basic motive of QTAIM is to analyse the topology of the electronic density of molecules. By means of this analysis, it is possible to extract information related to the chemical bonding. Figure 2.6 shows two examples where QTAIM has been applied to determine the chemical connectivity. On the top, there is the niobium oxide example. We performed the Bader topological analysis, and located all the critical points (i.e., maxima, minima and saddle points in the electronic density). A bond critical point indicates that there is a chemical bond between two atoms. If we compare the 3D molecule with the Bader representation, we can observe a perfect match. Thus, we would proceed by adding the node- and edge-attributes to obtain the final version of the molecular graph. On the other hand, Figure 2.6 bottom shows a problematic case. The connectivity of the 3D molecule does not match with its Bader representation. Actually, the Bader analysis proves that there are two chemical bonds which are disregarded in the 3D representation. The first missing bond is a metal-oxygen interaction, while the second case involves a hydrogen bond. Thus, it can be observed that QTAIM provides an accurate picture of the atom connectivity. That is why our program generates a first guess of the molecular graph taking advantage of Bader's chemical bond matrix.

2.2.2 Isomorphism

Isomorphism is a mathematical property which establishes morphological relationships between graphs. The word derives from the Ancient Greek: *isos* "equal", and *morphe* "form or shape". Thus, two graphs, g_i and g_j are isomorphic if there is a complete match between both structures. Instead, if there is a partial match, the exact definition would be *subgraph isomorphism*. In general terms, isomorphism evaluates the resemblance between edges of g_i and g_j. Besides, it also accounts for the attributes of the molecular graphs. This detail is specially relevant because nodes contain key data: the atomic numbers. Thus, when evaluating the isomorphisms, the mapping must not only match the edges, but also the nodes' attributes. We relied on the implementation developed by Cordella and coworkers, addressed to match large graphs efficiently.[113,114]

Figure 2.7 shows four practical examples to illustrate the isomorphism property. The target structure is depicted in the center, and labeled as g_3. There are four additional molecular graphs which will be matched to g_3 in order to determine if they are isomorphic. In first place, g_1 is a subgraph of the target structure: g_3. Actually, g_1 can be exchanged by the top-left part of g_3, leading to the same structure. Secondly, g_2 is not a subgraph of g_3, even though edgewise would fulfill the subgraph criteria. However, the nodes' attributes are different; g_2 has a tantalum atom (Z=73), and g_3 has niobium atom (Z=41). This example shows the importance of the nodes' attributes. In third place, g_4 it is not a subgraph of g_3 either. In this case, there are two protons in g_4, whereas there is only one proton in g_3. Finally, g_5 is isomorphic to g_3. Indeed, g_5 can also be exchanged by

the top-left part of g_3, and be rotated either clock- or anticlockwise.

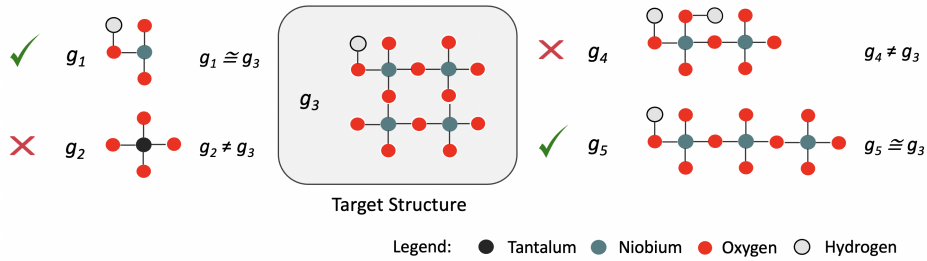


Figure 2.7: Relationships of isomoprhisms between five graphs. Color legend: black (tantalum), green (niobium), red (oxygen), grey (hydrogen).

2.2.3 Computational Complexity

Evaluating the isomorphism property is extremely useful. It unlocks the possibility of establishing chemical relationships in an automatic manner. However, this feature has one major drawback: the high computational cost. From a hardware perspective, the computational power has increased exponentially during the past decades. This evidence is very well illustrated by *Moore's law*, named after the CEO of the Intel company: Gordon Moore. It roughly estimates that the number of transistors, in a dense integrated circuit, doubles every two years. Figure 2.8 shows this steady increase during the past five decades.[115] For example, if we focus on the number of transistors, we observe that it has raised seven orders of magnitude. This stunning growth has boosted computational applications in a unprecedented manner. Even the frequency, power density, and cores which had remained constant, have also increased considerably in the recent years. In section 2.3.5 we will show how *POMSimulator* benefits

from running on multiple threads. Actually, it employs up to 28 cores, thus accelerating the resolution of the speciation models.

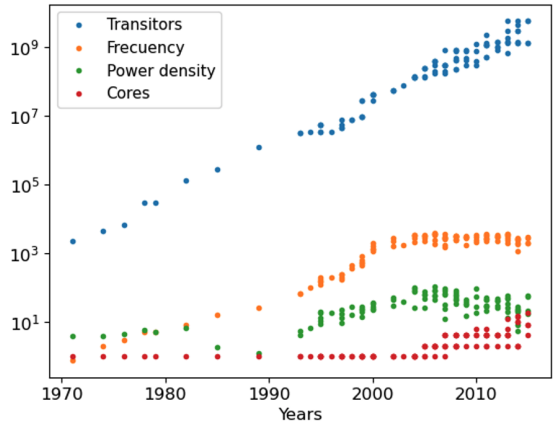


Figure 2.8: Moore's law highlighting the increase of transistors, frequency, power density and cores throughout the last five decades; data extracted from[115] (y-axis in logarithmic units).

Despite the exponential growth of computational power, computer scientists soon realized that brute-force was not always enough. In fact, there are some computer programs which scale very quickly with the size of the system. Thus, it is not possible to assume that Moore's law will eventually provide the sufficient computer power. Consequently, problems started to be sorted according to their complexity, giving rise to a classification called *NP*, which stands for *nondeterministic polynomial time*. This classification aims to organize problems according to the computational time required to solve them. Also, it studies how the solving time escalates with the size of the system. This latter aspect is key, specially for large systems. Figure 2.9 left shows a simplified Euler diagram with different *NP*

classes. The easiest problems are classified as *P*, which stands for being solvable, deterministically, in polynomial time. Within this class there are relevant examples, such as: multiplication, determining if a number is prime, and calculating the greatest common divisor (GCD). Figure 2.9 right shows the growth of the computation time respect to the system size. For instance, calculating the greatest common divisor is performed in less than 1 µs, regardless of the system size. Analogously, determining if a number is prime shows a similar trend but one order of magnitude above GCD. Next, multiplication shows a slight increase of the computation time respect to the system size. Even so, the time scale is still within the 1 µs regime.

The second class corresponds to *NP*. It groups problems whose solutions can be *verified* in polynomial time. In this class we find the subgraph isomorphism problem. Because the isomorphicity can be verified, but not solved in polynomial time, the computational cost rapidly increases with the system size. For example, calculating the isomorphism properties for small ($[MO_4]^{n-}$) and medium ($[M_8O_{26}]^{n-}$) compounds is affordable. However, finding subgraphs for large structures ($[H_{32}M_{36}O_{128}]^{n-}$) becomes expensive, and troublesome.

To illustrate this point, Figure 2.9 right shows the computation time for calculating isomorphisms of random-generated graphs. It should be said that these random graphs do not include node attributes. Thus, they can be regarded as a simplified version of molecular graphs. Even so, the time scale is substantially larger than the other operations depicted in Figure 2.9. The mismatch is several orders of magnitudes higher, moving between 1 ms and 1 s. Besides, the sys-

tem size has a strong impact on the computational cost. As a matter of fact, isomorphism has the steepest slope, whereas multiplication, prime search, and GCD are almost constant (note that the y-axis is in logarithmic units). Fortunately, our program does not deal with molecular graphs larger than 250 nodes. Thereby, the average computation time would be approximately 10 ms for each isomorphism. Although 10 ms might seem low, it becomes an issue if this process has to be repeated a large number of times. As a rough estimation, lets assume we have 120 molecular graphs. To apply the isomorphism property, we have to create pair of molecular graphs, and compare them. However, chemical reactions might involve more than two species. To also consider bimolecular reactions, a third molecular graph must be included. If we consider all possible combinations of graphs, this operation has to be repeated at maximum $120^3 = 1,728,000$ times. Provided that each isomorphism is solved in ≈ 10 ms, the computation cost would be equal to ≈ 5 hours. It is not an unbearable amount of time, but it shows how easily the computation time scales when considering such systems.

Although calculating isomorphisms is already expensive, there are other problems which are even more computationally-demanding. Still within the *NP* framework, we find the *NP-complete* group. This class includes the difficult cases of *NP*. It is widely acknowledged that *NP-complete* problems cannot be solved in polynomial time. A popular example is the resolution of Sudokus. While the classical 9x9 Sudoku can be easily handled, it becomes exponentially difficult for larger dimensions. Despite this, *NP-complete* cases can still be verified in polynomial time. For instance, solving a Sudoku of 12x12

can result time-consuming, but it is straightforward to check if there is some number repeated. Indeed, this duality between solving, and verifying, is by no means trivial. In computer science is referred to as: *"the P versus NP question"*. In fact, it is widely acknowledged as the most groundbreaking problem in computer science.[116] For instance, it is one of the seven Millenium Price Problems chosen by the Clay Mathematics Institute, with a bounty of $1,000,000$ (US) prize for the first correct solution.[117] Solving this question would mean that problems which are *verified* in polynomial time, can be *solved* in polynomial time. Notwithstanding, the current state-of-the-art suggests that there is no such relationship. Therefore, problems are harder to compute than to verify, commonly expressed as P $\neq$ NP.[118]

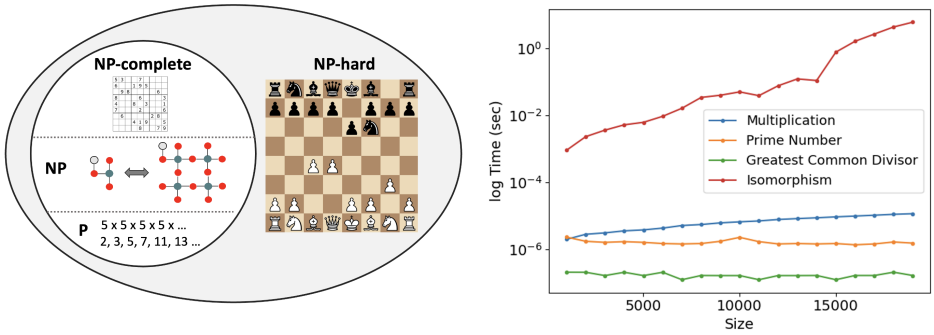


Figure 2.9: Left: Euler diagram with polynomial and non-polynomial type of classes. Right: Computation time versus system size for multiplication, prime search, greatest common divisor, and isomorphism (y-axis in logarithmic units).

This brings us to the last group: *NP-hard*. This class contains all problems which are not only difficult to solve, but also to verify. A classical example is determining the best move in chess. Although today's chess programs come up with excellent moves,[119] it is yet

unclear how to verify if those moves are the *absolute* best. In addition, the search space is practically boundless. More than 70 years ago, Claude Shannon (one the founders of computer science, along with Alan Turing) demonstrated that there were at least 10^{120} possible chess games.[120] To illustrate the overwhelming magnitude of the problem, let's compare it to the number of atoms in the universe, 10^{82}. The latter number was estimated thanks to data collected by Planck's satellite, published in 2016.[121] If we compare both numbers, we observe a difference of 38 orders of magnitude, $10^{120-82} = 10^{38}$. This simple example provides a glimpse of the complex nature of *NP-hard* problems.

Shannon and Turing believed that games were convenient models for testing and expanding the boundaries of computer science. It was not until the end of the past century, in 1997, when a computer program named Deep Blue (developed by the IBM Company) beat Garry Kasparov[d].[122] This achievement showed the resourcefulness of computer scientists in optimizing both the software and the hardware. As a consequence, tree search algorithms (employed in Deep Blue), were later used in many other applications.[123] Two decades later, board games still remain a research topic. In 2018, Deep Mind (an artificial intelligence company) developed a program to beat the best Go player: Lee Sedol.[124] This breakthrough had a big impact on the scientific community, leading many researchers to develop similar AI architectures. For instance, AI applications in chemistry has increased notably in the last years.[125]

[d] Garry Kasparov was the 13^{th} World Chess Champion (1985-2000), and he is widely regarded as the best chess player in the history.

2.2.4 Construction of the Chemical Reaction Network

Determining the chemical reaction network (CRN) is a central piece to predict the self-assembly of polyoxometalates. To construct the CRN, we split the process in several steps. First, we defined an artificial grid, M_{grid}, to guide the search of chemical reactions. This grid consisted of a lower triangular matrix, of NxN dimensions, where N is equal to the number of species. Next, we defined a vector which contained all the molecular graphs, V_{graphs}. These graphs are numerically sorted according to the number of metals, oxygen and hydrogen atoms, respectively. Therefore, the first elements correspond to the smaller clusters, and viceversa. Figure 2.10 shows a simplified example of M_{grid} and V_{graphs}. Note that it only includes three monomers, and three dimers, at different protonation states. As a reference, results presented in the following chapters, employ matrices from 40x40, up to 120x120. In order to find all the chemical reactions, we started by mapping M_{grid}. If the matrix element, M_{ij}, is equal to 1, then it calls the searching functions. Otherwise, if M_{ij} is equal to 0, it does not apply any function, and jumps to the next element. In this manner, we narrow the search, and the usage of computational resources. Importantly, this simplification does not disregard any reaction because of the reversibility principle. Since all the oxo-clusters are in equilibrium, finding a direct reaction implies finding its reverse counterpart. For example, an association reaction will have the same absolute energy as its dissociation, but with an opposite sign. Consequently, by simply changing the sign we can obtain the reverse reaction, without increasing the search time.

Figure 2.10: Search grid matrix, M_{grid}, and molecular graph vector, V_{graphs}. Example shows a simplified case for monomers and dimers at different protonation states. If M_{ij} is equal to 1, we apply the reaction search functions. Otherwise, the program does not call the search functions.

The searching functions consist of two independent functions which evaluate if there is a chemical reaction between a pair (g_i,g_j), or a trio (g_i,g_j,g_k) of molecular graphs. First, a function assesses if there is an isomorphic relationship between a set of molecular graphs. Then, another paramount condition has to be considered: the reaction stoichiometry. Indeed, more than two centuries ago Antoine Lavoisier[e] introduced this quintessential concept. It is iconically embedded in the sentence: *"Nothing is lost, nothing is created, everything is transformed"*.[126] Thus, chemical transformations have to fulfill an equal balance between the atoms of reactants and products. Thus, there is a second function which assesses the stoichiometry rule. Moreover, the order in which this conditions are evaluated is not

[e] Antoine Lavoisier (1743-1794) was a French chemist who is widely considered as "the father of modern chemistry"; he discovered the role of oxygen in combustion and respiration, proved the law of conservation, reformed the chemical nomenclature, and named hydrogen.

Reaction Type	**Graph Reaction**	**Atom Balance**
Acid-Base	$g_j + g_{H^+} \rightarrow g_i$	[0,0,1]
Condensation 1	$g_j + g_k \rightarrow g_i + g_{H_2O}$	[0,1,2]
Condensation 2	$g_j + g_k \rightarrow g_i + 2\ g_{H_2O}$	[0,2,4]
Condensation 3	$g_j + g_k \rightarrow g_i + 3\ g_{H_2O}$	[0,3,6]
Condensation 4	$g_j + g_k \rightarrow g_i + 4\ g_{H_2O}$	[0,4,8]
Addition	$g_j + g_k \rightarrow g_i$	[0,0,0]
Dimerization	$g_j + g_j \rightarrow g_i$	[0,0,0]
Hydrolysis	$g_j + g_{H_2O} \rightarrow g_i + g_{H^+}$	[0,1,1]
Hydration	$g_j + g_{H_2O} \rightarrow g_i$	[0,1,1]

Table 2.1: Reaction types considered in the program, molecular graph formula, and atom balance. Balances refer to the difference of atoms only between g_i, g_j, and g_k (e.g., $[\Delta M, \Delta O, \Delta H]_{ijk}$).

irrelevant. Computing the stoichiometry is cheaper than calculating the isomorphism. As we discussed in the previous section, evaluating isomorphism can easily become the bottleneck of the method. Therefore, to construct the chemical reaction network, we first check the stoichiometry. If the latter leads to a stoichiometrically-feasible reaction, then the isomorphism is check. Finally, if a isomorphic relationship is found between the molecular graphs, the chemical reaction is added to the network.

The type of chemical reactions are predefined by the user in the function that assesses the stoichiometry. Table 2.1 collects the seven types of chemical reactions considered in this thesis. Each chemical reaction is related to an atom balance. We have chosen this set of reactions based on the reactivity patterns of polyoxometalates. It is widely accepted that the self-assembly of molecular oxides is dictated by acid-base, and nucleation reactions.[82] Actually, there

are some polyoxometalate families that only aggregate at acid media, whereas other families strictly form at alkaline conditions. Regarding the nucleation reactions, we have defined three types: condensations, additions and dimerizations. Within the condensation group, there are four types according to the number of water molecules released. These reactions can be regarded as a trade-off between protons and cluster size. The more protons an oxo-cluster has, the higher the degree of condensation it can undergo. Besides, we have also considered addition reactions, which are similar to condensation reactions, but without releasing water molecules. A special case of addition is the dimerization reaction. This latter becomes relevant for larger clusters. In fact, the largest polyoxometalates tend to display high degrees of symmetry. Therefore, it is reasonable to assume that most of them will form through the dimerization of two smaller parts. Next, we have included hydrolysis reactions to ensure that we reproduce alkaline conditions. At basic pH there is a high concentration of hydroxide anions, consequently, the formation of metal hydroxide will be favored. The last type of transformation, depicted in Table 2.1, is the hydration reaction. Because the self-assembly of these oxo-clusters happens in aqueous media, water molecules play the role of solvent and reactants. The hydration reaction contemplates the latter case. A water molecule adds to the metal center of a polyoxometalate, thus expanding its coordination sphere. In this manner, oxo-clusters can display tetrahedral, trigonal bipyramid, and octahedral geometries.

It is worth pointing out that the function that evaluates the stoichiometry has been designed in a modular approach. The code will

only consider the reactions that are predefined in Table 2.1. Thereby, the number of reaction types can be easily modified by just adding or removing rows in Table 2.1. Ideally, the reaction network should consider all the chemical reactions possible. However, this approach would lead to very complex reaction networks. In fact, in the next section we will show why large networks become unmanageable when solving speciation models. Therefore, it is reasonable to provide tools for controlling the size of the network. For example, the oxo-clusters regarded in this thesis do not undergo redox reactions. That is why, Table 2.1 only collects nucleation, and acid-base reactions. Nevertheless, polyoxometalates are redox active compounds. Thus, if there are oxidants in the reaction medium, oxo-clusters could be reduced. It is up to the user to adjust Table 2.1 according to the chemistry of the system.

To illustrate how the program checks the stoichiometry, let's see a practical example. In first place, it looks for a non-zero element in M_{grid}, for instance M_{11}.[f] This element has an i index of 1, and thus, it corresponds to molecular graph $g_{[HMO_4]^-}$. Next, it screens the same row M_{1j}, and looks for the non-zero elements. In the present example, it would find M_{10}, which corresponds to $g_{[MO_4]^{2-}}$. Once it has found two molecular graphs, it calculates an atom balance, employing equation 2.2.1. The result of this balance is then compared to the cases depicted in Table 2.1.

$$g_i - g_j = [M_i - M_j, O_i - O_j, H_i - H_j] \tag{2.2.1}$$

[f] *POMSimulator* has been developed with Python, and thus lists are 0-indexed. The first element is 0, second is 1, and so on. This convention is kept throughout the thesis.

In this example, the atom balance would be equal to [0,0,1]. Consequently, these two molecular graphs, $g_{[HMO_4]^-}$ and $g_{[MO_4]^{2-}}$, could be related by an acid-base reaction. This *modus operandis* is kept during the reaction search, with one variation. We contemplate the inclusion of a third molecular graph, g_k, in the atom balance. For example, take element g_i as M_{44}, which corresponds to $g_{[HM_2O_7]^-}$. The algorithm would do a second screen of row M_{4j}, and would find $g_{[MO_4]^{2-}}$ as a possible candidate. However, the balance between $g_{[HM_2O_7]^-}$ and $g_{[MO_4]^{2-}}$ would result in a non-tabulated vector: [1,3,1]. Despite that, the program would do a third screen of row M_{4j}, taking into account an additional molecular graph, g_k. Next, it would use equation 2.2.2 to calculate the atom balance. If g_k is $g_{[HMO_4]^-}$, the atom balance would result in [0,1,2], leading to a tabulated case: a condensation reaction. Next, the program would call the isomorphism function to verify if the isomorphic relationship is fulfilled. If it is the case, the reaction would be accepted in the network.

$$g_i - (g_j + g_k) = [M_i - M_j - M_k, O_i - O_j - M_k, H_i - H_j - M_k] \quad (2.2.2)$$

Figure 2.11 shows tridimensional molecular models with the corresponding reaction types and balanced stoichiometric coefficients. The first type corresponds to the protonation reaction between $[HMO_4]^-$ and H_2MO_4. Despite the simplicity of the reaction, it is extremely relevant. All acid-base equilibria play a major role in favoring some species over others. For this reason, every nuclearity[g]

[g] Nuclearity refers to clusters which have the same number of oxygen and metal atoms.

must have different number of protons to ensure that they can appear at distinct ranges of pH. Next, Figure 2.11 also shows two examples of condensation reactions. First, condensation 1 involves $MO_3(H_2O)$ and $[MO_4]^{2-}$ to yield $[M_2O_7]^{2-}$, and one equivalent of H_2O. This type of condensation is the most usual, because it does not require reactants with high states of protonation. Therefore, this transformation can both happen at acid, neutral, and even slightly basic pH. On the other hand, condensation 2 involves $[HMO_3(H_2O)_2]^{+}$ and $[MO_4]^{2-}$ to yield $[HM_2O_7]^{-}$, and two equivalents of H_2O. This reaction is less frequent than condensation 1, but it is also crucial, specially at low pH. When the medium is very acid, condensations tend to release more water molecules to yield larger clusters. For example: $[H_{32}Mo_{36}O_{128}]^{8-}$, or $[H_2W_{12}O_{42}]^{10-}$.

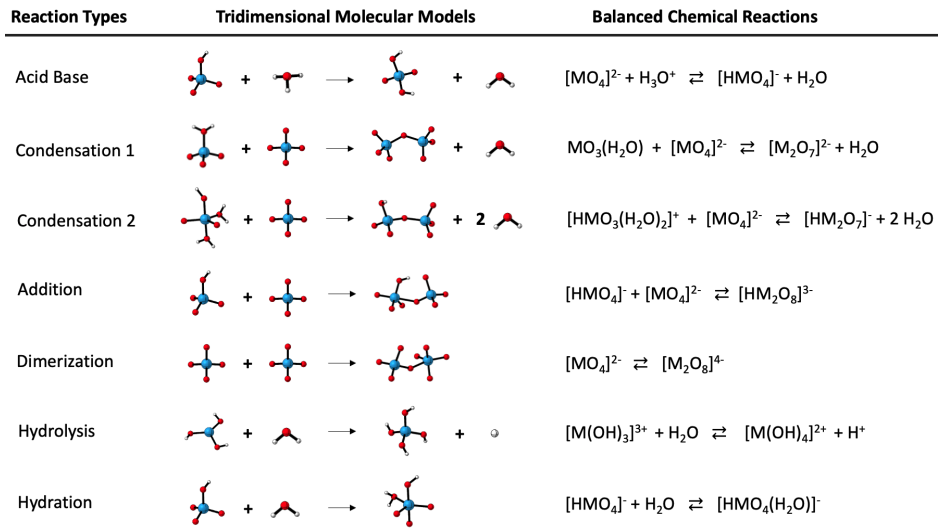


Figure 2.11: Reaction types, tridimensional molecular models, and stoichiometrically-balanced reactions.

The next following type of nucleation reactions are the additions. They resemble to condensations reactions but with the single difference that water is not formed as a byproduct. Figure 2.11 shows the transformation of $[HMO_4]^-$ and $[MO_4]^{2-}$ to $[HM_2O_8]^{3-}$. Because there are no water molecules released, the product has a higher oxygen metal ratio. For instance, the condensation product is $[M_2O_7]^{2-}$, whereas the addition product is $[HM_2O_8]^{3-}$. This evidence is relevant because not all the polyoxometalates have the same metal oxygen ratio. Therefore, depending on the nuclearity of the product, a combination of condensation and addition reactions will be required. The last type of nucleation reaction is depicted as the dimerization. Figure 2.11 shows a simple example of two equivalents of $[MO_4]^{2-}$ which yield $[M_2O_8]^{4-}$. This type of reaction is specially relevant for larger clusters. The most straightforward example would be the synthesis of $[H_{32}Mo_{36}O_{128}]^{8-}$ which occurs through the dimerization of two symmetric $[H_{16}Mo_{18}O_{64}]^{8-}$ units.

Metal oxo-clusters can also react directly with water, leading to two types of transformations: hydrolysis and hydration. The former, as it name suggests, involves the splitting of water to form metal hydroxides. Figure 2.11 shows the hydrolysis reaction of $[M(OH)_3]^{3+}$ to $[M(OH)_4]^{2+}$. This transformation is specially important for polyoxoniobates and -tantalates, because they have a neutral-alkaline speciation. Therefore, the abundance of metal hydroxides is higher than for polyoxomolybdates or -tungstates. Finally, Figure 2.11 shows the hydration reaction. It relates $[HMO_4]^-$ and H_2O to form $[HMO_4(H_2O)]^-$. This transformation expands the coordination sphere without inflicting any change to the molecular charge.

In fact, water molecules are very labile, which means that polyoxometalates have a dynamic equilibrium with this ligand. For instance, molybdic acid can expand its tetrahedral geometry to octahedral by adding two consecutive water molecules.

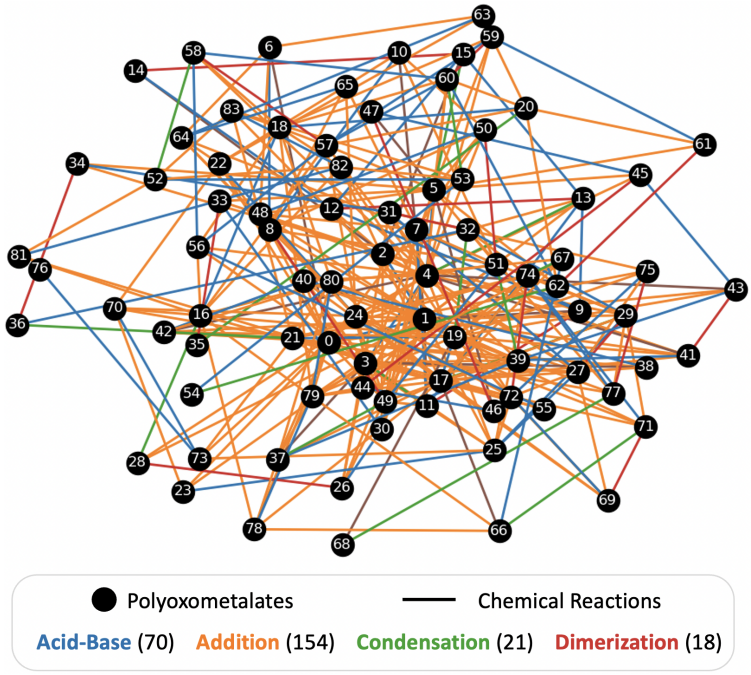


Figure 2.12: Chemical reaction network constructed automatically by *POMSimulator*. Nodes are oxo-clusters, edges are chemical reactions. Colors indicate reaction types.

The next step is to arrange all the chemical reactions in a unique network. Indeed, chemical reaction networks are a paramount asset in computational chemistry. This is so because they are necessary for computing speciation and microkinetic models. Figure 2.12 shows a CRN constructed automatically following the set of heuristic rules

explained previously. Black nodes represent metal-oxo clusters, and links represent chemical reactions. This network depicts the self-assembly of polyoxotungstates in aqueous medium. Note that nodes are sorted randomly, in a circular pattern. Nonetheless, it is also possible to represent the nodes according to their nuclearity, or connectivity. In this case, Figure 2.12 collects a complex network with a total of 263 chemical reactions. In first place, acid-base reactions form a considerable part (27%) of the network. Indeed, these percentage is constant for most of the networks studied in this thesis. The percentage left is formed by the different types of nucleation reactions: condensations, additions, and dimerizations.

Finally, the current CRN has two relevant issues that must be discussed. First, the lack of molecular searches. The simulation will only consider compounds introduced previously by the user. This is an important limitation because it restricts the chemical space. However, polyoxometalates have been extensively studied with multiple experimental techniques.[82] Thereby, there is plenty of data describing oxo-cluster intermediates. These experimental evidences are excellent guidelines for building the molecular set. Therefore, this first issue does not pose a insurmountable problem. The second issue concerns the absence of transition states in the CRN. The network is exclusively set with stable (aka., minima) structures. One the one hand, this is not a problem for solving speciation models. Because speciation does not depend on time, transition states are not required. On the other hand, microkinetic simulations can only be performed with the presence of transition states; or more precisely, with activation energies. Determining TS for polyoxometalates is troublesome be-

cause most POMs have almost barrierless reactions.[59] For instance, there are fewer transition states for POMs than for other homogeneous systems.[54] Even so, in the present thesis we focus on systems in chemical equilibrium, and thus no TS are required.

2.3 Chemical Equilibrium

Chemical equilibrium is a cornerstone concept in chemistry. It is ubiquitous in every chemical process; from microgram synthesis to large scale industrial processes. According to the IUPAC[h], chemical equilibrium is the point where the rates in both chemical directions (direct, and reverse) are identical. Therefore, the chemical system has reached a static composition. From a historical perspective, it was the french chemist, Claude Louis Berthollet, who laid the first foundations in his *Essay on chemical statics*, in 1803.[127] According to historians, Berthollet was accompanying the expedition of Napoleon Bonaparte in Egypt. At the beginning of the *Napoleonic wars* (1803-1815), Berthollet found some surprising evidences in the lake Natron, located in the north-east of Tanzania. This lake had sodium carbonate in its shores.[i] Berthollet noticed that it could come from the reverse reaction of the dissolution of calcium carbonate. Equation 2.3.1 shows the aforementioned reaction, where both pair of carbonate and chloride salts are in equilibrium.

[h] International Union of Pure and Applied Chemistry

[i] From[127] p. 405: " (...) *la décomposition du muriate de soude continue donc jusqu'à ce qu'il se soit formé assez de muriate de chaux, parce que l'acide muriatique devant se partager entre les deux bases en raison de leur action, il arrive un terme où leurs forces se balancent.*"

$$Na_2CO_3 + CaCl_2 \rightleftarrows 2NaCl + CaCO_3 \downarrow \tag{2.3.1}$$

Chemical equilibrium is measured experimentally with the *equilibrium constant* (K_{eq}). For a general bimolecular reaction (equation 2.3.2), the equilibrium constant takes the formula expressed in equation 2.3.3. It is a quotient between the multiplication of chemical products (C and D), and the multiplication of reactants (A and B). Besides, all the species are powered to their respective stoichiometric coefficients. Usually, equilibrium constants consist of large numbers due to the small values in the denominator. Because working with large values is not ideal, chemists also express equilibrium constants in logarithmic form. Furthermore, the absolute value of the constant also informs about the reaction yield. For example, if the constant is equal to a large number, it indicates that the equilibrium is shifted towards products. On the other hand, if the constant has a low value, the equilibrium is displaced towards reactants.

$$\nu_A A + \nu_B B \rightleftarrows \nu_C C + \nu_D D \tag{2.3.2}$$

$$K_{eq} = \frac{[A]^{\nu_A}[B]^{\nu_B}}{[C]^{\nu_C}[D]^{\nu_D}} \tag{2.3.3}$$

There are multiple parameters that affect the chemical equilibrium: change of concentrations, ionic force, temperature and pressure. To take into account these parameters, the french chemist Henry Louis Le Chatelier studied their effect on the equilibrium state. In general terms, he concluded that: "*any system at equilibrium when subjected to a change will evolve to a new equilibrium,*

by trying to counteract the applied change". For example, if concentration of $[C]$ or $[D]$ (in equation 2.3.2) decreases, the system will shift towards products. The same logic applies if the concentration of reactants lowered. This casuistic is rather common when some molecule changes its physical state, either because of precipitation or evaporation. This feature is used to promote the formation of a desired compound. For instance, a recent work showed that mechanochemistry favored the synthesis of polyoxometalates.[128] This is so because POMs are formed by condensation reactions, and thus, water molecules are generated as byproducts. Thus, if water is removed from the equilibrium as it occurs in a dry set up, the reaction will be constantly shifted towards products. Analogously, a similar effect happens when a variation in the temperature is performed. Let's assume that equation 2.3.2 corresponds to an exothermic reaction ($\Delta H_r < 0$). Then, if we increase the temperature of the system, the equilibrium will shift towards reactants. In this manner, the reaction is compensating the excess of heat. In a similar way, if we lowered the temperature, the equilibrium would be shifted towards products. Thereby, Chatelier's law provides a rule for predicting the system reaction to external changes. Despite that, Chatelier's law cannot provide an estimation of the equilibrium constant at different physical conditions. That is why, in 1884 the Dutch chemist, Jacobus Henricus van't Hoff, described a relationship to predict equilibrium constants at different temperatures.[129] The expression is shown below:

$$ln\left[\frac{K_T}{K_{T_0}}\right] = -\frac{\Delta H_r^0}{R}\left[\frac{1}{T} - \frac{1}{T_0}\right] \tag{2.3.4}$$

Equation 2.3.4 requires the equilibrium constant at standard temperature, K_{T_0}, the reaction enthalpy at standard conditions, ΔH_r^0, and the ideal gas constant, R. In this manner, K_T can be isolated, and expressed as function of known parameters. Consequently, it is possible to study chemical systems where non-standard temperatures are employed. For instance, polyoxomolybdates are also synthesized at hydrothermal conditions.[130] Thereby, if we wanted to study the synthesis of POMs at for instance $40°C$, we could employ equation 2.3.4 to estimate the equilibrium constants. On the other hand, equation 2.3.4 only provides an estimation of K_T.[131] This limitation is due to two important assumptions: first, ΔH_r^0 is considered constant respect to the temperature. However, ΔH_r^0 and ΔS_r^0 do vary with temperature. The second assumption is closely related; it assumes that ΔH_r^0 is equal to its value at standard temperature. Despite these simplifications, equation 2.3.4 provides reasonable estimations.

2.3.1 Speciation Models

To study the chemical equilibria of a system, we need to determine the distribution of species in the reaction media. To do so, we rely on speciation models. This model should contain all the chemical reactions present in solution. Consequently, the model should predict the most abundant compounds at a given set of conditions. Chemists employ analytical techniques such as mass spectrometry and nuclear magnetic ressonance to determine the speciation *in-situ*.[84,85] Notwithstanding, in this thesis we aim at determining the speciation *in-silico*. Therefore, we do not rely on experimental techniques, but on quantum mechanical calculations. To start we will define the

equations that conform the speciation model.

$$\sum_{i=1}^{N}[M]_i - M_{t_0} = 0 \tag{2.3.5}$$

$$K_{eq} = e^{-\Delta G_r/(RT)} \tag{2.3.6}$$

The first one is equation 2.3.5, which depicts the mass balance constraint. The sum of the concentrations of all the species must be equal to the initial concentration. With the parameter M_{t_0} we can choose the working concentration. This aspect is relevant because some POMs only form at diluted conditions, and viceversa. For example, the largest isopolyoxomolybdate, $\{Mo_{36}\}$, forms at concentrations higher than 1 M. On the other hand, if more diluted solutions are employed, the most abundant compounds would be the octamolybdate, $\{Mo_8\}$, and finally the molybdic acid, H_2MoO_4. Therefore, having an exact control of the metal concentration is a valuable asset for predicting the formation of these oxo-clusters.

Next, we need an equation which relates theoretical Gibbs reaction energies (ΔG_r) with the concentration of the chemical species. This relationship is given by Equation 2.3.6. The equilibrium constant is equal to the exponential powered to the $-\Delta G_r$ and divided by the ideal gas constant, R, and temperature, T. Because we had defined the expression of the equilibrium constant previously, we can combine equation 2.3.6 and 2.3.3 to obtain equation 2.3.7.

$$\frac{[A]^{\nu_A}[B]^{\nu_B}}{[C]^{\nu_C}[D]^{\nu_D}} = e^{-\Delta G_r/(RT)} \tag{2.3.7}$$

$$[C]^{\nu_C}[D]^{\nu_D}e^{-\Delta G_r/(RT)} - [A]^{\nu_A}[B]^{\nu_B} = 0 \tag{2.3.8}$$

Equation 2.3.7 relates the concentration of the reacting species, and the reaction free energy. We can estimate ΔG_r with quantum mechanics, thus the only unknown variables of the system are the concentrations of the chemical species. Additionally, equation 2.3.7 can be displayed in a similar way as equation 2.3.5. This transformation is done in equation 2.3.8, and it is a convenient rearrangement when introducing the equations to the solver. Then, to calculate the concentrations, we have to combine equations 2.3.5 and 2.3.8. A textbook example is the speciation diagram of carbonic acid. To obtain such diagram, the speciation model must include the acid-base reactions which relate H_2CO_3, $[HCO_3]^-$, and $[CO_3]^{2-}$. Equations 2.3.9 and 2.3.10 show the protonation reactions of $[CO_3]^{2-}$ and $[HCO_3]^-$, respectively.

$$[CO_3]^{2-} + H_3O^+ \rightleftarrows [HCO_3]^- + H_2O \tag{2.3.9}$$

$$[HCO_3]^- + H_3O^+ \rightleftarrows H_2CO_3 + H_2O \tag{2.3.10}$$

With the additional equation related to the mass balance, we are able to set up a determined system of equations. It has three equations, and three unknown variables which correspond to the concentration of each carbonic derivate:

$$\left.\begin{aligned}[HCO_3^-]\cdot[H_2O]\cdot e^{-\Delta G_r/(RT)} - [CO_3^{2-}]\cdot[H_3O^+] = 0\\ [H_2CO_3]\cdot[H_2O]\cdot e^{-\Delta G_r/(RT)} - [HCO_3^-]\cdot[H_3O^+] = 0\\ [CO_3^{2-}] + [HCO_3^-] + [H_2CO_3] - M_{t_0} = 0\end{aligned}\right\} \quad (2.3.11)$$

Analogously, equation 2.3.11 can also be expressed in matricial form, as depicted below. Therefore, solving the speciation model involves a straightforward diagonalization to obtain the eigenvalues.

$$\begin{bmatrix} 0 & [H_2O]\cdot e^{\frac{-\Delta G_r}{(RT)}} & -[H_3O^+] \\ [H_2O]\cdot e^{\frac{-\Delta G_r}{(RT)}} & -[H_3O^+] & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} [H_2CO_3] \\ [HCO_3^-] \\ [CO_3^{2-}] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ M_{t_0} \end{bmatrix}$$

Apart from the carbonic species, there are two additional concentrations which deserve attention: $[H_2O]$ and $[H_3O^+]$. First, $[H_2O]$ is usually fixed to a value several orders of magnitude above M_{t_0}. In this manner, we account for the fact that water is much more abundant than the rest of species. Secondly, $[H_3O^+]$ is the independent variable of the speciation model. The system of equations is solved at multiple pH values to study the behaviour of the dependent variables. Figure 2.13 left shows the speciation diagram associated to the carbonic acid. Notice that the plot contains dots at the pH values at which the speciation model was solved. It can be easily observed that carbonic acid predominates at acid pH. Then, at pH between 4 and 6 the hydrogencarbonate is the most abundant compound. Finally, from neutral pH to more alkaline conditions the carbonate is

the most dominant anion.

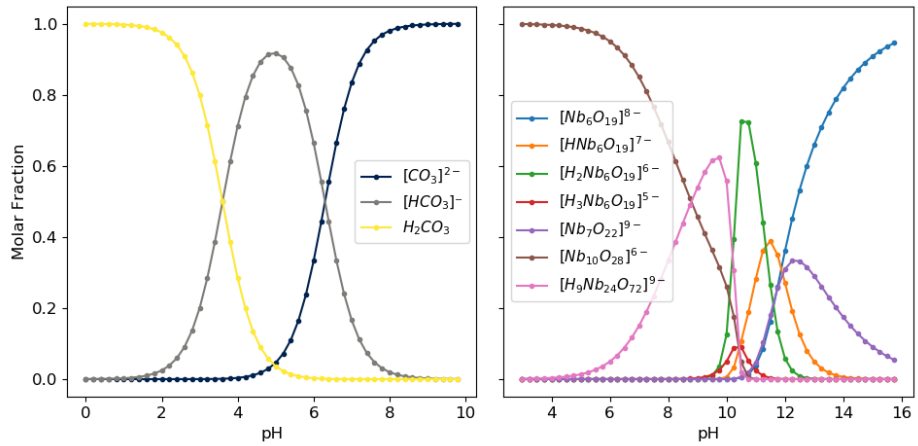


Figure 2.13: Left: speciation diagram for carbonic acid. Right: speciation diagram for isopolyoxoniobates. Dots indicate at which pH values the speciation models were solved.

The former example only regarded acid-base reactions for a small oxoacid. Unlike carbonic acid, polyoxometalates present a more complex challenge. Apart from the acid-base properties, POMs also polymerize yielding a broad variety of structures in shape and size. Figure 2.13 right shows the speciation diagram of polyoxoniobates which we first described in a recent work.[3] We will discuss in more detail niobates speciation in chapter 4. For the time being, let's focus on the crucial difference between diagrams in Figure 2.13. The diagram on the right shows four nuclearities, whereas the left one depicts a unique nuclearity at different protonation states. This evidence shows that some oxo-acids do polymerize (e.g., polyoxometalates) while some others do not (e.g., carbonic, sulfuric, or nitric oxoacids). When working with POMs, pH not only affects the protonation state

of the clusters, but also their size. This feature adds another layer of complexity to the speciation model. To take into account the growth of these oxo-clusters, we have to include nucleation reactions in the model. As an example, we will introduce the pair of dimers ($[M_2O_7^{2-}]$ and $[HM_2O_7^{-}]$) in the speciation model. Equation 2.3.12 show the condensation reaction between two equivalents of a hydrogenmetalate, to yield a dimer, and a water molecule.

$$2[HMO_4]^- \rightleftarrows [M_2O_7]^{2-} + H_2O \qquad \text{Nucleation reaction} \quad (2.3.12)$$

$$\left.\begin{aligned}
[HMO_4^-] \cdot [H_2O] \cdot e^{-\Delta G_r/(RT)} - [MO_4^{2-}] \cdot [H_3O^+] = 0 \\
[H_2MO_4] \cdot [H_2O] \cdot e^{-\Delta G_r/(RT)} - [HMO_4^-] \cdot [H_3O^+] = 0 \\
[HM_2O_7^-] \cdot [H_2O] \cdot e^{-\Delta G_r/(RT)} - [M_2O_7^{2-}] \cdot [H_3O^+] = 0 \\
[M_2O_7] \cdot [H_2O] \cdot e^{-\Delta G_r/(RT)} - [HMO_4^-] \cdot [HMO_4^-] = 0 \\
[MO_4^{2-}] + [HMO_3^-] + [H_2MO_4] + [M_2O_7^{2-}] + [HM_2O_7^-] - M_{t_0} = 0
\end{aligned}\right\} \quad (2.3.13)$$

Equation 2.3.13 shows the speciation model with the inclusion of the condensation reaction shown above. At first glance, the speciation model is very similar to the one shown in equation 2.3.11. Both share alike expressions for the acid-base reactions, and a mass balance. However, the speciation model containing the condensation reaction has a subtle but crucial difference. The discrepancy between both is highlighted in equation 2.3.13 with orange colour. The variable $[HMO_4^-]$ is multiplying itself, and thus we no longer

have a linear equation. Consequently, it is not possible to express this system of equations in matricial form. In mathematical terms, the model has to be considered as a system of non-linear equations. In fact we deduced that speciation models with nucleation reactions cannot be solved by diagonalization.

2.3.2 Benchmark of Solvers

To solve the systems of non-linear equations, our program employs the optimizers implemented in the SciPy library.[92] This library offers five solvers for finding the root of multivariate functions. To decide which algorithm worked best for our system, we made a comparative study. Figure 2.14 shows a performance benchmark for five different solvers. The first option is *hybr* that calls the MINPACK's *hybrd* and *hybrj* routines, which employ a modified Powell method.[132] Secondly, *lm* is the abbreviation for the least squares with Levenberg-Marquardt.[133] In third place, *anderson* keyword refers to the Anderson acceleration, also known as Anderson mixing.[134] Finally, the *exciting* and *linear mixing* contain random effects in addition to the usual fixed effects.[135] We employed three types of polyoxometalates to demonstrate which solver was the most efficient for solving the NLE. Besides, NLEs were solved at different pH values to ensure consistency.

Molybdates are the system which show a greater variance regarding the solver. For example, *hybr* and *lm* solved the molybdate speciation model in $\approx 10^{-1}$ seconds, whereas the *anderson* and *mixing* took $> 10^2$ seconds. There is a difference of almost three magnitudes, which is a huge time-gap. For the vanadates system happens

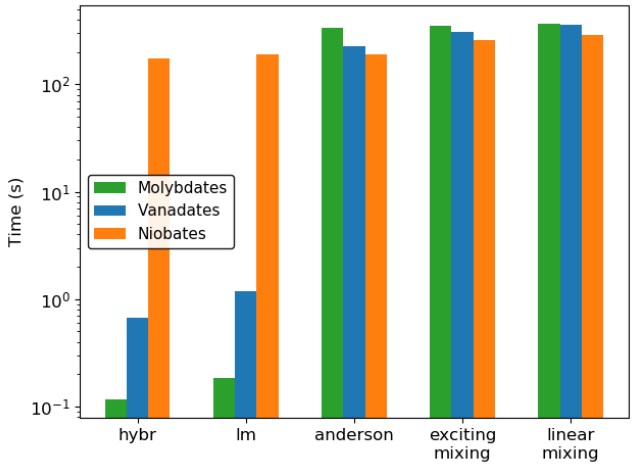


Figure 2.14: Benchmark times with five algorithms for solving speciation models of molybdates (green), vanadates (blue), and niobates (orange). Models were solved at a pH range between 1 and 14, with a 0.5 step. Calculation run in a Intel(R) Core(TM) i7-3820 CPU @ 3.60GHz.

a similar situation. The *hybr* and *lm* algorithms solved the models in $\approx 10^0 seconds$, whereas the others did it in $> 10^2 seconds$. In this case, the time-gap is one magnitude lower, $\Delta t = 10^2 seconds$, but still rather large. Thus, *hybr* and *lm* still are the preferred option. For the niobates' system, the time differences are not so clear compared to molybdates and vanadates. In general terms, the same trend is observed where *hybr* and *lm* are the best options. Nonetheless, the time-gap is no longer greater than one magnitude. With this results at hand, we chose *hybr* as the default algorithm for solving the systems of non-linear equations.

None of the methods mentioned above required previous computation of the Jacobian or Hessian. Nonetheless, they did request

an initial guess, which in our case consisted of an array of zeros; as many zeros as polyoxometalates were considered in the system of equations. Alternatively, we could have provided an array with random values, or with a reasonable estimation of the hypothetical concentrations. However, these options had considerable drawbacks as well. First, employing random values might cause the solver to be more unstable. In other words, the speciation diagrams might change depending on the random seed employed. Ideally, the solver should provide the same solution regardless of the initial conditions. This would be fulfilled if the algorithm found the absolute minimum of the multivariate function. However, solvers depicted in Figure 2.14 are designed to locate local minima. Therefore, different initial guesses might lead to different local minima for the same multivariate function. Alternatively, we could have employed initial guesses that consisted of reasonable estimations of the oxo-clusters concentrations. For example, using the concentrations of the previous pH step as the initial guess. Nonetheless, this option would have introduced a significant bias to the calculation. If there was a sudden change in concentrations, the solver would be already biased due to the initial guess. That is why we chosen the zero-array, to ensure a robust and unbiased resolution.

2.3.3 Thermodynamic Activity

We had to consider the effect of one important variable: the ionic strength (I). This variable measures the concentration of ions that coexist in a solution. The concept of ionic strength was developed by Lewis and Randall in 1921 in parallel to the description of the activity

coefficients of strong electrolytes.[136] Determining the ionic strength is important because it affects the equilibrium constants. For instance, salts such as sodium chloride, $NaCl$, are usually employed to stabilize the formation of polyoxometalates, as well as to control the pH. As a consequence, the synthesis of POMs occurs in high ionic strength conditions. Equation 2.3.14 shows the mathematical expression of I, as a summation of the products between the concentration, c_i, and chemical charge of each species, z_j.

$$I = \frac{1}{2}\sum_{i=1}^{N} c_i \cdot z_j^2 \qquad c_i \text{ in (mol/L solution)} \tag{2.3.14}$$

In fact, to account for the effect of the ionic strength, *POMSimulator* must work ultimately with activities instead of raw concentrations. In chemical thermodynamics, activity, a, is a measure of the *effective concentration* of species in solution. Ions have a non-innocent role, and thus, they interact with the oxo-clusters. This is due to the mismatch between the *real* concentration, and the *apparent* concentration. The term *activity* in this sense was coined by G. N. Lewis in 1907.[137] Equation 2.3.15 shows the mathematical expression where the activity is depicted as the product between the the activity coefficient, γ_c, and the quotient between the concentration, c_i, and the concentration at standard state of 1 M, c_0.

$$a = \gamma_c \cdot \frac{c}{c_0} \qquad c_0 \text{ is the standard state at 1 M} \tag{2.3.15}$$

In order to estimate the activity, it is necessary to calculate the activity coefficient, γ_c. There are several ways of estimating this para-

meter depending on the ionic strength conditions. Figure 2.15 shows the five types of functions discussed in this thesis; sodium chloride is employed for the five cases. The most basic formula for estimating γ_c is the Debye-Hückel expression, depicted in equation 2.3.16. This expression was developped by the chemists Peter Debye and Erich Hückel in 1923.[138] According to this equation, the activity coefficient is a function of the ionic strength instead of the electrolyte concentration. Besides, at standard conditions in water, the A parameter is equal to 0.5085 $M^{-1/2}$. Figure 2.15 left shows that the a^{DH} decays too rapidly.[j] That is why, we have also included Figure 2.15 right, which presents the x-axis in logarithmic units. This plot shows that at very low ionic strengths, $\approx 10^{-3}$, the Debye-Hückel function provides accurate results. However, at ionic strengths higher than $1 \cdot 10^{-2.3}$ M, a^{DH} tends to zero. This would implicate that whatever the concentration of the solute was, the activity would be null. Therefore, equation 2.3.16 cannot be employed in these circumstances. To increase the range of applicability of a^{DH}, the Extended Debye-Hückel expression depicted in equation 2.3.17 was also developed.[138] The A parameter is the same as in the former equation, and the B parameter is equal to 3.281 $M^{-1/2}nm^{-1}$. Figure 2.15 shows that for very low values of I, the denominator in the expression 2.3.17 approximates to one. Thus, $\gamma_{\pm}^{EDH}$ is proportional to the square root of the ionic strength (aka, Debye–Hückel limiting law). However, at higher I values, a^{EDH} does not decay to zero completely. In this manner, $\gamma_{\pm}^{EDH}$ can be used for ionic strengths up to 0.1 M.

[j] Activities, a, and activity coefficients, γ, share the same superscript.

$$\log_{10}\gamma_{\pm}^{DH} = -A \cdot z_j^2 \cdot \sqrt{I} \qquad \text{Debye-Hückel} \qquad (2.3.16)$$

$$\log_{10}\gamma_{\pm}^{EDH} = -A \cdot z_j^2 \cdot \frac{\sqrt{I}}{1 + B \cdot a_0 \cdot \sqrt{I}} \qquad \text{Extended Debye-Hückel} \qquad (2.3.17)$$

Notwithstanding, most chemical systems work with I greater than 0.1 M. Therefore, even the Extended Debye-Hückel expression is not enough to model more complex solutions. As a consequence, other alternatives have been developed to model conditions at greater ionic strengths. Most of them are empirical extensions of the Debye-Hückel theory. At low concentrations, they follow the γ^{DH} equation, and at higher ionic strength conditions more terms (at some power of I) are added to fit with the experiments. The main extensions are the Daviesj and the Pitzer equations.

$$\log_{10}\gamma_{\pm}^{DV} = -A \cdot z_j^2 \left(\frac{\sqrt{I}}{1 + \sqrt{I}} + B \cdot I \right) \qquad \text{Davies Equation} \qquad (2.3.18)$$

The Davies expression is defined in equation 2.3.18, and it was originally published by Cecil W. Davies in 1938.[139] The expression was developed with the aid of experimental data. The B parameter is equal to -0.3, while the A remains unchanged respect to the value shown before. Figure 2.15 right shows that at low concentrations, equation 2.3.18 reduces to the Extended Debye–Hückel equation (2.3.17). This is so because when the term $-0.3I$ tends

to zero, the ionic strength also approximates to zero. On the other hand, as concentration increases, the second term becomes increasingly important. In this manner, the Davies equation can be used for solutions which are too concentrated for the Debye–Hückel equation. In fact, $\log_{10} \gamma_{\pm}^{DV}$ can be used for solutions up to an ionic strength of $0.5M$. Another relevant advantage of the Davies equation lies in its simplicity.

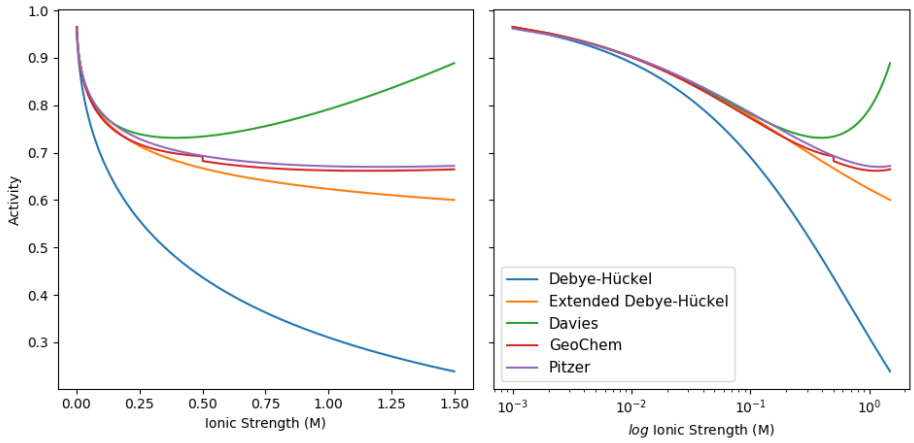


Figure 2.15: Five activity models for sodium chloride: Debye-Hückel, Extended Debye-Hückel, Davies, GeoChem and Pitzer. Notice that the right plot presents the ionic strength in logarithmic units.

Although an ionic strength of 0.5 M is already high, there are still many experiments which are performed at more concentrated solutions. Consequently, the Davies expression cannot be employed in such cases. In fact, Figure 2.15 shows that a^{DV} increases dramatically for I values greater than 0.5 M. This unusual increase is an artifact, and the consequent activities a^{DV} are misestimated. Because several formation constants of POMs are computed at high ionic conditions,

the Davies equation was not always convenient. Nevertheless, the *Geochem* software employed a modification of equation 2.3.18 which increased its range of applicability.[140,141] This modification consisted in splitting the activity equation in two conditions, as shown in equation 2.3.19. Essentially, the function would remain the same but it would only change two parameters: a_0 and B_0. For instance, if $I \leq$ 0.5 M, a_0 would be equal to 1/B, whereas B_0 would be equal to the product of 0.2, B_0 and I. Instead, if $I > 0.5M$, a_0 would be equal to $(3 + |z|) \cdot 10^{-8}$, and B_0 to a constant value of 0.041. In this manner, the second term would not increase exponentially if greater values of I were employed. Figure 2.15 shows the evolution of a^{GC} for the same range of ionic strength. Note that there is the discontinuity at $I = 0.5$ which marks the switch of the a_0 and B_0 parameters. As a consequence, at high ionic strength conditions a^{GC} reaches a plateau which is in full agreement with the experiments.

$$\log_{10} \gamma_{\pm}^{GC} = -A \cdot z_j^2 \frac{\sqrt{I}}{1 + B \cdot a_0 \cdot \sqrt{I}} \cdot B_0 \cdot I$$

$$\begin{array}{llll} if & I \leq 0.5M & a_0 = \dfrac{1}{B} & B_0 = 0.2 \cdot A \cdot z^2 \\ if & I > 0.5M & a_0 = (3 + |z|) \cdot 10^{-8} & B_0 = 0.041 \end{array} \tag{2.3.19}$$

If even greater values of ionic strength are required (e.g., I $\gg$ 1 M) the GeoChem modification might fail as well. Fortunately, the physical chemist Kenneth Pitzer developed an expression to describe the activity coefficients at concentrated solutions. Equation 2.3.20 depicts the general formula of the Pitzer expression. First, we can

observe that Pitzer also includes the Extended Debye-Hückel as the principal term. Nonetheless, there are two additional terms which correspond to the interaction between the cations and anions for a general salt of formula: M_pX_q. More concretely, the interaction of the pair ion correspond to B^{γ}_{MX}, and the interaction of the three ions corresponds to C^{γ}_{MX}. Equation 2.3.21 depicts the expression for the two-ionic interaction, where $\beta^{(0)}_{MX}$, $\beta^{(1)}_{MX}$, and α are parameters derived from experiments. Analogously, equation 2.3.22 depicts the formula for the tri-ions interactions, where μ_{MMX} and μ_{MXX} are parameters determined experimentally as well. Figure 2.15 shows that a^{PZ} works well both at diluted, and concentrated conditions.[k]

Employing the Pitzer equation should be the ideal choice. However, the strong parametrization of this model hinders its applicability. While for common salt pairs ($NaCl$, CaF_2, $LiBr$ etc) there are well-documented databases, it is not the case for polyoxometalates. In fact, there are few to none Pitzer parameters for most of molybdates, tungstates and vanadates salts. Even if recent articles have worked on this direction,[143] they are still far from providing a complete set of parameters. Specially, taking into account that metal-oxo clusters form many intermediates which should be considered as well. Therefore, from a practical point of view *POMSimulator* cannot apply the Pitzer equations yet. Despite this inconvenient, the GeoChem implementation is very close to the ideal case, as Figure 2.15 demonstrates. That is why, in this thesis we employed a^{GC} to estimate the activities of the molecular metal-oxo clusters.

[k] To calculate a^{PZ}, we have employed the *PyEQL* Python package, developed by Ryan Kingsbury.[142]

$$\log_{10}\gamma^{PZ} = \log_{10}\gamma^{EDH} + m\left(\frac{2pq}{p+q}\right)B^{\gamma}_{MX} + m^2\left[2\frac{(pq)^{3/2}}{p+q}\right]C^{\gamma}_{MX} \tag{2.3.20}$$

$$B^{\gamma}_{MX} = \beta^{(0)}_{MX} + \beta^{(1)}_{MX}e^{-\alpha\sqrt{I}} \tag{2.3.21}$$

$$C^{\gamma}_{MX} = \left[\frac{3}{\sqrt{pq}}\right](p\mu_{MMX} + q\mu_{MXX})\,. \tag{2.3.22}$$

2.3.4 Combinatorial Explosion

In the previous section we showed the definition of a speciation model which considered one condensation reaction. It consisted of a system of five equations and five unknown variables. Nonetheless, speciation models are typically larger and include from tens to hundreds of reactions. Figure 2.16 shows a bar plot with the number of reactions for the five types of polyoxometalates considered in this thesis. Polyoxomolybdates are the system with the largest number of reactions, followed by tantalum, tungsten, vanadium and niobium. Ideally, a speciation model should consider all the reaction network for a given family of polyoxometalates. However, this practice is not feasible. For example, if we take the number of reactions for molybdenum metal-oxo clusters, there are 96 possible transformations. Thus, if we wanted to include all these reactions in a single speciation model, we would need 96 equations. Nonetheless, there are only 73 different molybdenum clusters. Thereby, we have more reactions than compounds ($96 - 73 = 23$ reactions left), or alternatively, more equations

than variables. The same problems occurs for the rest of the polyoxometalates. The amount of chemical reactions is greater than the number of polyoxometalates. This scenario is referred to as an *overdetermined system*. As a consequence, it is not possible to find a consistent solution.

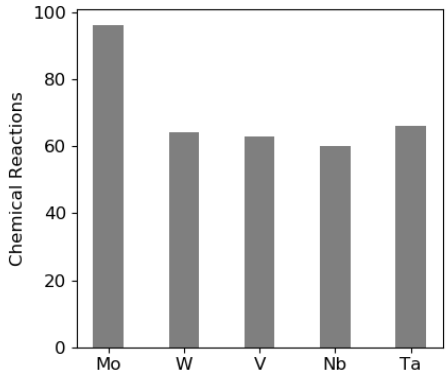


Figure 2.16: Total number of chemical reactions for molybdenum, tungsten, vanadium, niobium and tantalum polyoxometalates.

To address this problem, we divided the overdetermined speciation model in smaller systems of non-linear equations. The size of these smaller models will be fixed by the number of polyoxometalates. For instance, there are 73 molybdates, thus, the model will be formed by 73 equations and 73 variables. Next, it is necessary to rely on a algorithm which considers the different combinations leading to distinct models. In other words, how many speciation models can be set up with a given number of chemical reactions. To answer this question we relied on the binomial coefficient. Equation 2.3.23 shows the formula where the numerator corresponds to the unknown variables, and the denominator to the number of equations. In chemical

terms, n would be the different metal-oxo clusters, and k the number of chemical reactions. The result of equation 2.3.23 produces all the possible combinations of reactions. Then, each combination corresponds to a speciation model with a square system of non-linear equations. In this manner, we have exchanged an unsolvable model, for a sequence of smaller and solvable models.

$$Speciation\ Models = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad n(reactions), k(POMs) \tag{2.3.23}$$

Applying equation 2.3.23 is very convenient, but it has a relevant drawback: the computational cost. Due to the large amount of speciation models, the computation time can increase very easily. Figure 2.17 left shows the number of speciation models for polyoxomolybdates, -tungstates, -vanadates, -niobates, and-tantalates, according to equation 2.3.23. Due to the factorial growth of equation 2.3.23, the y-axis is in logarithmic units. Even so, the growth for the five families of polyoxometalates is overwhelming. Again, polyoxomolybdates are the system with the highest number of speciation models due to its larger molecular dataset. Nonetheless, the rest of polyoxometalates also increase factorially, reaching more than 10^{20} models with just one hundred reactions. Because the sheer number of models does not inform about the computational cost, let's assume that one speciation model takes 10 seconds to be solved. Then, solving all the speciation models (for a network of 100 nodes) would cost the unbearable amount of 10^{21} seconds or $\approx 3^{13}$ days. Because it is not feasible to rely on the latter time scale, the computational cost must

be reduced. To do so, we can minimize n and k.

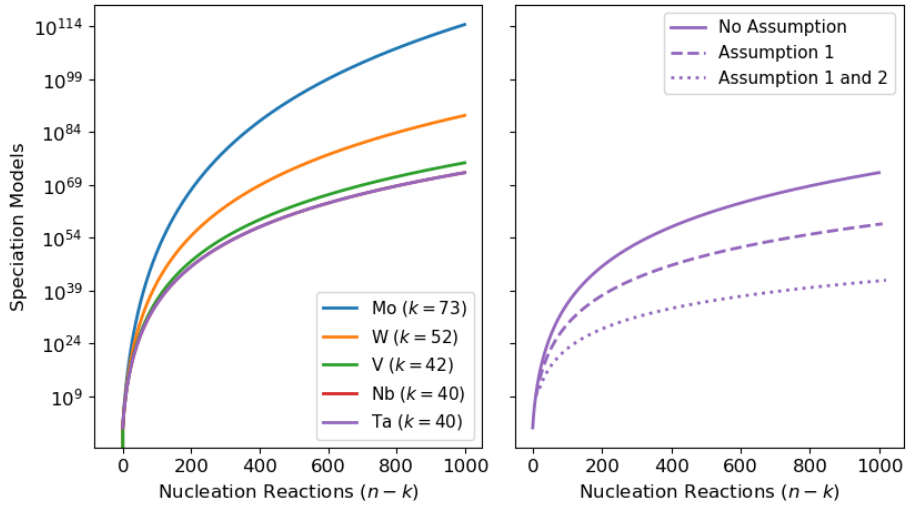


Figure 2.17: Left: number of speciation models according to the POM system. Right: effect of the assumptions on the number of speciation models.

In first place, k is related to the initial number of polyoxometalates. Less polyoxometalates will mean simpler speciation models, because there will be less equations. Thereby, it is extremely important to invest time in selecting the building blocks for the initial molecular set. The choice of compounds not only affects the chemical reaction network construction (see section 2.2.4), but also the number of speciation models. Figure 2.17 depicts two plots showing the effect of the number of reactions in the number of speciation models. For the sake of clarity, we have normalized the number of reactions with the number of compounds of each system $(n-k)$. In this manner, all the curves start on the same point and thus the comparison is more

straightforward. Figure 2.17 left shows that nucleation reactions have a striking influence in the factorial growth of equation 2.3.23. Small increases in the x-axis lead to large increases in the number of models. Molybdates have the largest set ($k = 73$), and thus, the number of speciation models is approximately 10^{28} greater than the second largest group. Moreover, the molecular set of niobium and tantalum have the same size which explains that their respective curves overlap. Bearing this in mind, the molecular set should be large enough to account for all the crucial intermediates, but also small enough to avoid overwhelming the NLE solver.

Secondly, n can also be reduced in order to decrease the number of speciation models. In this case, the restriction is related to the number of chemical reactions. *A priori*, it is not obvious which reactions are more important than others. Nonetheless, some constraints had to be imposed to avoid investing CPU-time in poor speciation models. Examples of bad speciation models would be those which lack relevant reactions. Thereby, we have made two paramount assumptions in order to both disregard poor models and decrease the computational cost.

The first assumption states as follows: *a speciation model must always contain all the possible acid-base reactions*. Therefore, the set of acid-base reactions will be fixed for all the speciation models. In contrast, nucleation reactions will vary according to the binomial coefficient. This assumption is made based on the crucial role of acid-base properties of polyoxometalates. It is well-reported that different pH values cause the formation of distinct metal-oxo clusters.[82] As a consequence, this reaction type must be always present in the spe-

ciation models. With this assumption, the number of reactions n is reduced to the number of nucleation reactions. Figure 2.17 shows that applying this first assumption causes a notable decrease the number of speciation models.

The second assumption states as follows: *a speciation model must always contain nucleation reactions that include all the nuclearities.* For instance, if we have fifteen nuclearities and a model only considers fourteen, the formation of the fifteenth will not take place; this case must be prevented. The formation of metal-oxo clusters should be driven by system parameters (pH, reaction energy, concentration etc) instead of arbitrary factors. With the second assumption we address this potential problem. Notice that once a model includes a nucleation reaction with a given nuclearity, the acid-base reactions will connect it with its protonated and deprotonated states. Figure 2.17 also depicts the effect of this second assumption has a relevant impact on the number of speciation models. Thus, the gap between *No Assumption* and *Assumption 1 and 2* should be equal to speciation models that are incorrect. In this way, we are disregarding 10^{30} models which are most likely inaccurate.

To set a practical example, lets suppose that we have twelve compounds split in four nuclearities: $[H_xMO_4]^{x-2}$, $[H_xM_2O_7]^{x-2}$, $[H_xM_3O_{10}]^{x-2}$, and $[H_xM_4O_{13}]^{x-2}$. Each nuclearity is found in three protonation states. Thereby, we have twelve compounds, thus we need twelve equations. One equation is always reserved for the mass balance. Consequently, there are eleven equations which will be related to chemical reactions.

$$
\begin{aligned}
[MO_4]^{-2} + H_3O^+ &\rightleftarrows [HMO_4]^- + H_2O && \textit{Acid} - \textit{Base} \\
[HMO_4]^- + H_3O^+ &\rightleftarrows H_2MO_4 + H_2O && \textit{Acid} - \textit{Base} \\
[M_2O_7]^{-2} + H_3O^+ &\rightleftarrows [HM_2O_7]^- + H_2O && \textit{Acid} - \textit{Base} \\
[HM_2O_7]^- + H_3O^+ &\rightleftarrows H_2M_2O_7 + H_2O && \textit{Acid} - \textit{Base} \\
[M_3O_{10}]^{-2} + H_3O^+ &\rightleftarrows [HM_3O_{10}]^- + H_2O && \textit{Acid} - \textit{Base} \\
[HM_3O_{10}]^- + H_3O^+ &\rightleftarrows H_2M_3O_{10} + H_2O && \textit{Acid} - \textit{Base} \\
[M_4O_{13}]^{-2} + H_3O^+ &\rightleftarrows [HM_4O_{13}]^- + H_2O && \textit{Acid} - \textit{Base} \\
[HM_4O_{13}]^- + H_3O^+ &\rightleftarrows H_2M_4O_{13} + H_2O && \textit{Acid} - \textit{Base} \\
[HMO_4]^- + H_2MO_4 &\rightleftarrows [HM_2O_7]^- + H_2O && \textit{Condensation} \\
[MO_4]^{2-} + H_2M_2O_7 &\rightleftarrows [M_3O_{10}]^{-2} + H_2O && \textit{Condensation} \\
H_2MO_4 + H_2M_3O_{10} &\rightleftarrows H_2M_4O_{13} + H_2O && \textit{Condensation}
\end{aligned}
\tag{2.3.24}
$$

Equation 2.3.24 shows a set of chemical reactions which defines a speciation model. The choice of these eleven reactions was made according to the assumptions explained before. For instance, with four nuclearities and three protonation states, the maximum number of acid-base equilibria is eight. Therefore, all these acid-base reactions (depicted in blue color) are considered in the model. Indeed, they represent a constant set in any binomial combination. Then, there are three reactions left which are kept for condensations, additions, and dimerizations. These reactions have to represent all the four nuclearities present in the molecular dataset. Equation 2.3.24 shows that the three reactions meet the condition of the second assumption. Notice that these reactions define one model combination

out of many. For example, if there were 20 nucleation reactions, the binomial coefficient $\binom{20}{3}$ would produce 1,140 different speciation models.

2.3.5 Batch Parallelization

So far, we had described measures to minimize the computational cost from the software point of view. Nonetheless, we can also optimize the usage of the hardware. The resolution of all the speciation models is an appropriate case where this idea can be performed. As a matter of fact, this step is the main bottleneck of the program due to the large amount of speciation models.

The standard approach is to solve one speciation model after the other. This procedure is depicted as sequential mode, and it is shown in the top part of Figure 2.18 left. For the sake of simplicity, the scheme only shows six speciation models (grey squares), even though in reality there are thousands more. Additionally, the hexagon represents a six-core CPU. In the sequential mode, the CPU only employs one core while the five left remain inactive. This mode is not the most efficient because the resources of the CPU are not optimized. However, because the models are collected in a long array and they are independent between them, the famous rule *Divide et Impera* can be applied. Figure 2.18 left depicts an example of parallel computing. It is named batch parallelization because we create separate batches of data. Next, they are distributed to all the cores of a same CPU. If there are six speciation models, all of them would be solved simultaneously. In this manner, we could employ the 100% of the computational resources of a single CPU.

To show the effect of this technique in the performance of the program, we did benchmark considering: the number of cores and the calculation time. Figure 2.18 right depicts the times involved in solving the 1024 speciation models of the octamolybdate system. In this case we have employed a CPU with 12 cores. The time for solving all the models with one core is 17.3 hours. This case would correspond to the sequential mode. Instead, with already two cores the calculation time descends almost by half, to 8.1 hours. Analogously, with 4 cores the time also descends by roughly the half, 4.1 hours. If we analyze the case with the most cores, the time is 1.5 hours, which is substantially less than results obtained with the sequential mode. In fact, there is a factor of 11.53 (=17.3/1.5) between the worst and best case scenarios. Additionally, we have plotted a red line to showcase the descending trend with the number of cores.

POMSimulator can work either in sequential or in parallel mode depending on the purpose of the simulation. If some testing is being carried out in the local computer, we encourage to run it in sequential mode. Otherwise, the program will grab all the resources of the CPU and it will disable the computer. On the other hand, if *POMSimulator* runs on a High-Performance-Computer environment (as in Figure 2.18), we strongly recommend using the parallel mode.

2.3.6 Accuracy Threshold

From this point on, it was already possible to solve the speciation models, and obtain the respective activities. Nevertheless, we decided to solve each speciation model with two different solvers. This decision was made based on these two factors:

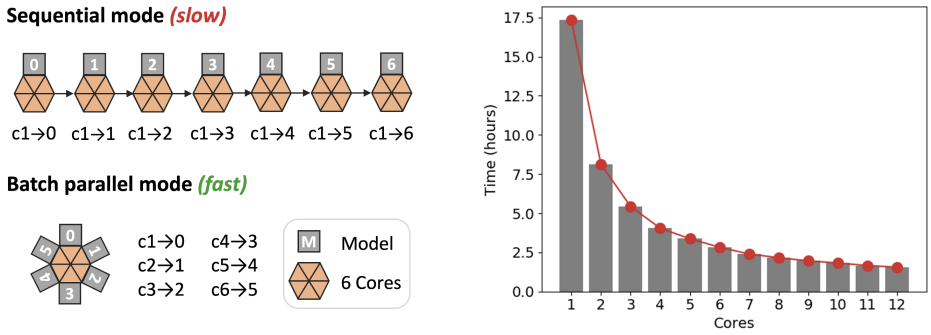


Figure 2.18: Left: simplified diagram of two running modes: sequential and batch parallel. Right: Bar plot with the calculation time of 1024 speciation models (octamolybdate system) as function of the number of cores. We have employed an Intel(R) Xeon(R) CPU @ 2.40GHz.

(i) After splitting the overdetermined problem into multiple solvable models, there is no obvious way to pinpoint which are the best models. Therefore, a criteria had to be designed to discriminate the quality of each solution.

(ii) Speciation models are formed by complex multivariate functions. To ensure that the Powell's method found a reasonable local minimum, we doublechecked the solution with a second algorithm.

The choice of the second algorithm is made according to the benchmark depicted in Figure 2.14. We chose *LS* as the algorithm for solving (for the second time) each speciation model.[l] More concretely, we compared the two solutions of the Powell and *LS* algorithms with a linear regression.

[l] We have employed the *LS* implementation available in SciPy in the library *scipy.optimize.root()*.

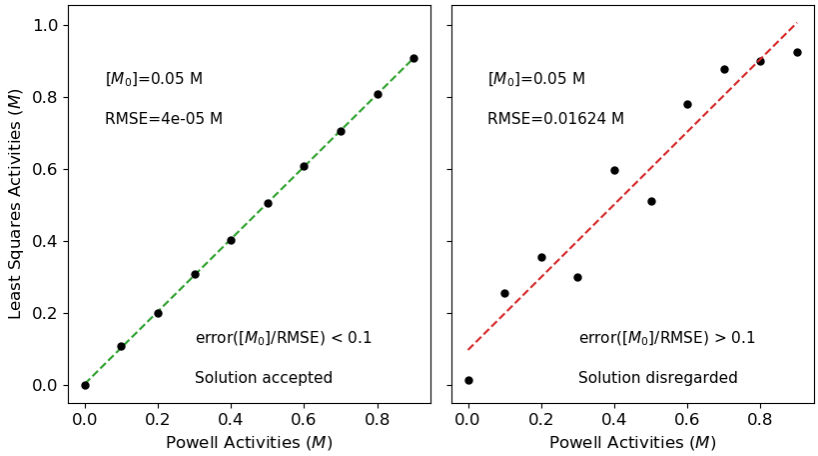


Figure 2.19: Linear regressions between Powell's and LS' solutions for two different speciation models. Left: solution accepted. Right: solution rejected.

Then, we normalized the root mean square error (RMSE) of the regression with the initial concentration. In this manner, the error was directly related to concentration units. Then, we had to ensure that the error was lower than the initial concentration. Otherwise, the uncertainty of the results would be too large. For example, if we started with a solution of 0.5 M, and the RMSE was 0.5, we would have activities with huge errors ($error(M_0/RMSE) = 1$). Thereby, it was necessary to determine a quality threshold, even though the choice was not trivial. We performed an acuracy benchmark where we employed threshold values ranging from 1.0 to 0.1 units. We found that the lower the threshold the greater the accuracy, thus we chose a threshold of 0.1 units. For further discussion of this topic, we refer the reader to Appendix B.

Figure 2.19 shows two linear regressions where the solutions of Powell and LS are compared. On the x-axis we depict the activities found with Powell's method, and on the y-axis we show the solution using the Least Squares. Figure 2.19 left depicts an scenario where the activities computed with Powell's method are accepted. This is so because the RMSE is under the threshold of 0.1. On the other hand, Figure 2.19 right shows a case where the activities were disregarded. In this case, the error is greater than 0.1, and consequently the activities are not accepted. The program would proceed with the next pH value until completing all the predefined range for a given speciation model.

2.3.7 Raw Speciation Diagrams

Once the speciation models have been formulated according to the two assumptions, solved with Powell's and LS methods, and passed the quality threshold, we finally get the raw activities of each oxo-cluster. To illustrate some preliminary results, we have selected the small molecular set of the octamolybdate, $[H_xMo_8O_{26}]^{x-4}$. We have used the octamolybdate because it is a simple but convenient example. It is formed by thirty oxo-clusters and it has 40 reactions and 1024 speciation models.

Figure 2.20 shows the speciation diagrams of six different models. It is worth mentioning that we have not plot oxo-clusters which have very low activities ($< 10^{-8}M$). Moreover, we have chosen separated models (i.e, model 0, 200, 400, 600, 800, and 1000) to evaluate the effect of the nucleation reactions. Notice that the more separated two models' indexes are, the more different their reaction set will be. Even

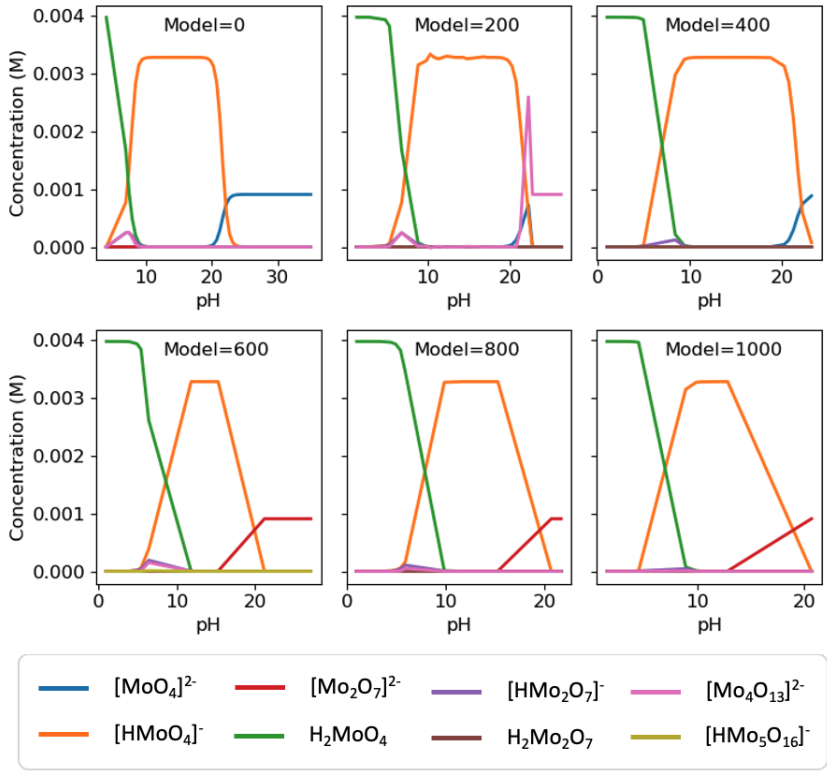


Figure 2.20: Speciation diagrams for the octaomolybdate molecular set. Raw activities of models 0, 200, 400, 600, 800, and 1000.

so, Figure 2.20 demonstrates that the six models have alike speciation diagrams in spite of having different reactions. This evidence shows the strong weight of the acid-base reactions in the models. Moreover, it is a good sign that there is not a huge variance between models. It indicates the soundness of the chemical assumptions, as well as the robustness of the models.

There are several interesting points concerning the speciation diagrams in Figure 2.20. Firstly, we can observe that the *relative* pK_a

of the species seems to be consistent. For instance, H_2MoO_4 appears at more acidic conditions than $[HMoO_4]^-$. Analogously, $[HMoO_4]^-$ predominates at more acidic conditions than $[MoO_4]^{2-}$. This trend is consistent for the six cases collected in Figure 2.20. However, none of the six diagrams shows the formation of the octamolybdates, $[H_xMo_8O_{26}]^{x-4}$, and the heptamolybdates, $[H_xMo_7O_{24}]^{x.-6}$. In fact, most of the models show the tiny formation of the tetramolybdate, $[Mo_4O_{13}]^{2-}$, even though it is not a major compound in the experiments.

Moreover, Figure 2.20 depicts another problem: the pH scale of the diagrams. Notice that the pH axis is roughly between 10 and 25, which is in disagreement with the experiments. Indeed, experimental results show that polyoxomolybdates form at very acid conditions (0-5 pH). Therefore, our speciation models are not describing the *absolute* values of the acidity constants. In fact, the speciation of these oxo-clusters is anomaly predicted in more alkaline conditions. Fortunately, this error in the acid-base constants is not an isolated but common problem in quantum mechanics. The mismatch between theory and experiments stems from the poor description of the solvated proton, H^+.[144] Because polyoxometalates are very pH-dependent, this error has strong implications in the raw concentrations, as shown in Figure 2.20. Therefore, it was necessary to come up with a solution. In the following section we will show the approach that we have chosen to overcome this problem.

2.4 Linear Scaling

In the previous section we highlighted a common challenge in DFT based methods: reproducing experimental equilibrium constants. The speciation diagrams depicted in Figure 2.20 indicated that the formation constants of the oxo-clusters were overestimated. To face this problem, there are two alternatives: (i) performing a linear scaling with experimental data (ii) setting up a thermodynamic cycle. Figure 2.21 summarizes these two options. The first option consists in calculating a linear regression between the experimental and theoretical equilibrium constants. Then, the slope and intercept parameters are employed to rescale the original constants. This approach is the most common, and it has been applied for different systems such as: small peptides,[145] macrocyclic clusters,[146] and lantanide complexes.[147] However, whether the scaling succeeds depends on finding a high determinant coefficient (R^2) and a low standard deviation error. Because the comparison is usually made with the logarithmic form of equilibrium constants, the linear scaling is very sensitive. Even small errors might lead to huge mismatches with the experimental data.

The other alternative is to set up a thermodynamic cycle to circumvent the problem related to the solvated energy of the proton. Figure 2.21 right shows an schematic overview of this method. The main idea is to express the Gibbs reaction energy in water, ΔG_r^{water}, as function of the reaction energy in gas phase, ΔG_r^{gas}, and the solvation energies, ΔG_r^{solv}. Then, the energy of the proton, ΔG_{r2}^{solv}, is estimated with experimental techniques. Notice that it is not possible to compute the energy of the proton with quantum mechanical methods due to the lack of electrons. According to the experimental

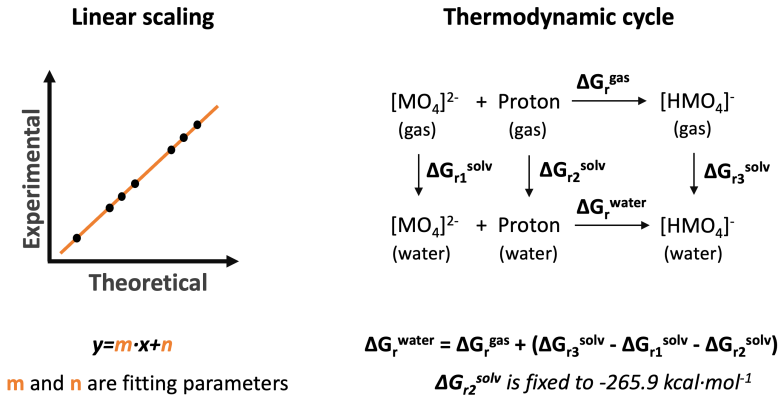


Figure 2.21: Two methods for solving the poor description of the Gibbs energy of the solvated proton. Left: performing a linear scaling. Right: setting up a thermodynamic cycle.

work of Tissandier et al. ΔG_{r2}^{solv} corresponds to -265.9 kcal·mol^{-1}.[148] With this energy it is possible to estimate the reaction energy in water, ΔG_{r}^{solv}.

In this thesis we have employed the linear scaling approach for correcting the theoretical constants. We made this choice based on two reasons. First, there is a considerable amount of literature describing experimental constants of polyoxometalates. In fact, the literature about this topic started relatively early. Pioneer articles were published in the 1960s,[149] but it not was until the 1980s when researchers determined several formation constants for polyoxotungstates,[150] molybdates,[151] and vanadates.[152] On the other hand, the thermodynamic cycle presents a couple of disadvantages. In second place, the cyclic approach requires that every geometry must be computed twice: in gas and solution state. Thus, the computational cost related to the quantum mechanical calculations would be higher. Also,

the absolute solvation energy of the proton is still under study.[153] Therefore, we decided to employ the linear scaling approach because of the extensive literature available.

2.4.1 Formation Constants

The speciation data of polyoxometalates is usually expressed as formation constants. These constants are defined by the formation reactions. On the reactants side there are p equivalents of the reference cluster, $[H_xM_pO_m]^{n-}$, and q equivalents of protons. The reference cluster is generally chosen based on its abundance in solution. Preferably, such compound should appear at any range of pH. Otherwise, the quotient of the formation constant would have a zero in the denominator. That is why different polyoxometalates systems employ distinct reference clusters. For example, molybdenum and tungsten employ the metal tetraoxide, $[MO_4]^{2-}$, as reference. It is a very abundant compound in alkaline conditions, and it is still present at neutral and acid media. In contrast, niobium and tantalum monomeric oxides are unstable and thus they cannot be selected as the reference compounds. Instead, niobium and tantalum employ the Lindqvist structure, $[M_6O_{19}]^{8-}$. This hexametalate compound is very stable both in alkaline and neutral conditions.

Equations 2.4.1 and 2.4.2 show the general formation reaction and the corresponding formation constant. Notice that formation reactions are a useful convention for establishing stoichiometric relationships for each product. However, formation reactions do not explain the self-assembly mechanism of these clusters. For instance, the formation of the largest molybdenum cluster, $[H_{32}Mo_{36}O_{128}]^{32-}$,

does not take place with the reaction of 36 equivalents of $[MoO_4]^{2-}$ and 32 equivalents of H^+. Instead, it forms through the dimerization of two $[H_{16}Mo_{18}O_{65}]^{16-}$ units. Thereby, formation constants should be strictly understood as a useful convention for expressing speciation data.

$$p[MO_4]^{2-} + qH^+ \rightleftarrows [H_xM_pO_m]^{n-} + H_2O \tag{2.4.1}$$

$$K_f = \frac{a_{[H_xM_pO_m^{n-}]}a_{[H_2O]^{\nu_B}}}{a_{[MO_4^{2-}]^p}a_{[H^+]^q}} \tag{2.4.2}$$

To calculate the formation constant, K_f, we require the raw activities of $[MO_4]^{2-}$, H^+, H_2O, and the product under consideration, $[H_xM_pO_m]^{n-}$. *POMSimulator* solves each speciation model multiple times; as many as pH values were predefined. Thus, we get an array of activities at each pH value which we convert to an array of formation constants using equation 2.4.2. This process is repeated throughout the whole range of pH. As a result, we obtain as many arrays of formation constants as pH values were considered. Because formation constants are invariant to the pH, we averaged the constants of each compound to obtain a unique value.

It might seem contradictory to invest time in computing formation constants at different pH values. However, we made this choice to minimize the numerical instabilities that appear in the calculation of these constants. Because some compounds might present extremely low activities, the numerical precision might consider them zero. As a consequence, the value of the formation constant may change considerably. To reduce the effect of this error, we solved the NLE at

different pH values to ensure that the relative activities were not always the same. For instance, $[Mo_2O_7]^{2-}$ is practically inexistent at pH 1, but it is abundant at pH 6. By calculating the formation constants at different stages, the final value is more robust.

One consequence of calculating the constants for every model is that we obtain a large amount of data. To represent and study this data adequately we employ on distribution plots. More concretely, we use a box plot to collect the formation constants of each compound. In this manner we can observe the average value, the standard deviation and the outliers for each oxo-cluster. Figure 2.22 shows the distributions of formation constants based on 1029 speciation models. The x-axis depicts the label of oxo-cluster while the y-axis shows the absolute value of the formation constants. From Figure 2.22 we can extract that there is a significant variance for each cluster. While some POMs have small deviations from the average value, others present major fluctuations. In fact, we observed the consequences of this variance in Figure 2.20. Speciation diagrams changed considerably depending on the model employed.

Notwithstanding, the constants collected in Figure 2.22 are in disagreement with the experiments. To illustrate this discrepancy we have compared some formation constants computed using *POMSimulator* with the values reported in the literature.[154] Table 2.2 summarizes the constants determined experimentally and theoretically. Apart from the reference value ($logK_f^{[MoO_4]^{2-}} = 0$), the rest of theoretical constants are systematically greater than the experimental ones. For instance, the constant predicted for $[HMoO_4]^-$ is almost 20 units greater than the reported in the experiments. This

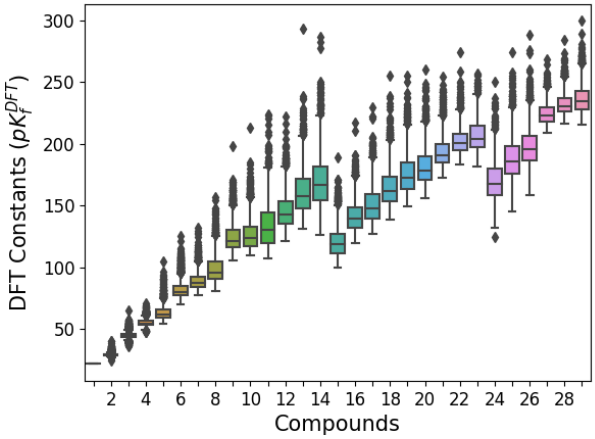


Figure 2.22: Box plots of the DFT formation constants for each compound (comprised between 1 to 29). Data obtained from the resolution of the 1024 speciation models.

trend is even more accentuated for the largest clusters. For example, octamolybdate constants differ more than 150 units. If we take into account that these constants are in logarithmic form, the mismatch between both sets of constants is overwhelming.

To solve this issue we followed the linear scaling approach presented previously. We performed six linear regressions for the six speciation models shown before in Figure 2.20. As a result we obtained six linear scalings collected in Figure 2.23. It is clear that the six regressions show a high degree of linearity. This evidence is reinforced by the high values of the determinant coefficients. For instance, for the six cases R^2 is greater than 0.9700. Model 200 has the highest determinant coefficient ($R^2 = 0.9964$), and thus it is a good candidate for performing the scaling. It is worth emphasizing the relevance

Compound	$logK_f^{Exp}$	$logK_f^{DFT}$
$[MoO_4]^{2-}$	0	0
$[HMoO_4]^{-}$	3.39	22.09
H_2MoO_4	7.35	29.53
$[Mo_7O_{24}]^{6-}$	52.42	166.85
$[HMo_7O_{24}]^{5-}$	57.23	195.93
$[H_2Mo_7O_{24}]^{4-}$	60.78	192.64
$[Mo_8O_{26}]^{4-}$	71.62	226.54
$[HMo_8O_{26}]^{3-}$	73.38	230.16

Table 2.2: Experimental and theoretical formation constants for the set of the octamolybdate. Formation constants employ the molybdenum tetraoxide, $[MoO_4]^{2-}$, as the reference cluster.

of the linear regressions. The fact the theoretical and experimental constants are very well correlated is a major milestone in the applicability of *POMSimulator*.

With the linear scalings in Figure 2.23 not only we can correct the eight raw constants employed in the regression, but the rest of the dataset as well. The 30 formation constants determined by *POMSimulator* can be scaled using the slope, and intercept parameters found in the regressions. This aspect is of great significance because some constants might be difficult, or even impossible to determine experimentally. Problems such as insolubility, or very low activities may occur in experiments. Therefore, we open the possibility of predicting unreported constants through the interpolation of the linear scalings.

Moreover, Figure 2.23 shows an intriguing feature: the slope parameter. For the six cases, the slope is equal to 0.3. Although the intercept shows greater variance, it can be stated that the linear scalings

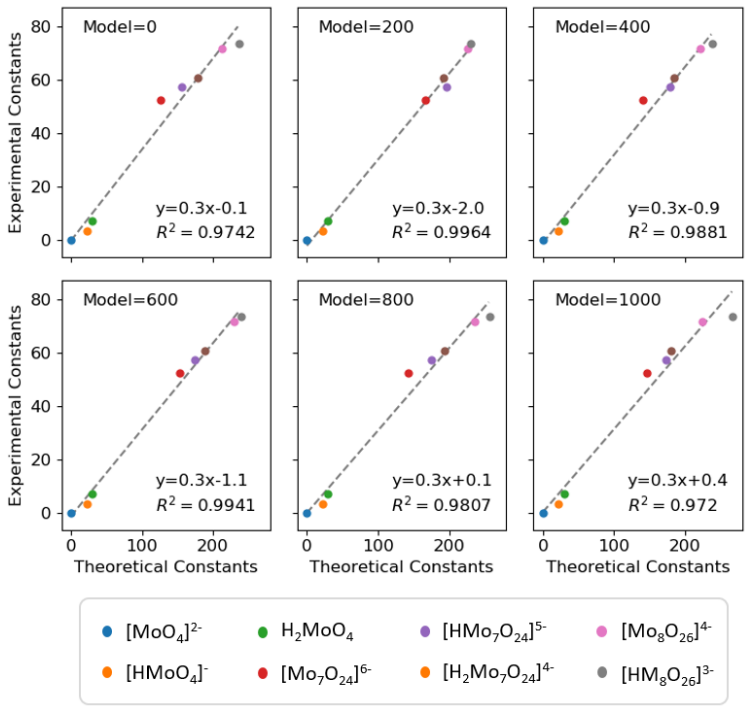


Figure 2.23: Linear scalings between experimental and theoretical constants for the six speciation models: 0, 200, 400, 600, 800, and 1000. Each scaling depicts the determinant coefficient, R^2, and the linear equation.

are strikingly alike. This means that the theoretical constants are generally overestimated by an order of three times respect to the experiments. In section 2.4.4 we discuss in more detail the potential universality of the linear scaling.

Once we have found a good linear scaling, we proceeded to the correction of the raw theoretical constants. To do so, we fit these constants into the linear scaling of each speciation model and we interpolate the formation constants. Next, we employed these new constants to calculate the speciation diagrams. However, there is a

technical difference between computing diagrams with reaction energies (as in Figure 2.20) and with formation constants. So far, speciation diagrams were generated through the resolution of models based on acid-base and nucleation reactions. In order to employ formation constants, we had to set up speciation models with formation reactions. In this manner, we could employ the scaled constants to plot the speciation diagrams.

$$\begin{aligned} [MO_4]^{2-} + H_3O^+ &\rightarrow [HMO_4]^- + H_2O \\ [MO_4]^{2-} + 2H_3O^+ &\rightarrow H_2MO_4 + 2H_2O \\ 2[MO_4]^{2-} + H_3O^+ &\rightarrow [M_2O_7]^{2-} + H_2O \\ 2[MO_4]^{2-} + 2H_3O^+ &\rightarrow [HM_2O_7]^- + 2H_2O \end{aligned} \tag{2.4.3}$$

$$\left.\begin{aligned} [HMO_4^-] \cdot [H_2O] \cdot K_f^{HMO_4^-} - [MO_4^{2-}] \cdot [H_3O^+] &= 0 \\ [H_2MO_4] \cdot [H_2O]^2 \cdot K_f^{H_2MO_4} - [MO_4^{2-}] \cdot [H_3O^+]^2 &= 0 \\ [M_2O_7^{2-}] \cdot [H_2O] \cdot K_f^{M_2O_7^{2-}} - [MO_4^{2-}]^2 \cdot [H_3O^+] &= 0 \\ [HM_2O_7^-] \cdot [H_2O] \cdot K_f^{HM_2O_7^-} - [MO_4^{2-}]^2 \cdot [H_3O^+]^2 &= 0 \\ [MO_4^{2-}] + [HMO_3^-] + [H_2MO_4] + [M_2O_7^{2-}] + [HM_2O_7^-] - M_{t_0} &= 0 \end{aligned}\right\} \tag{2.4.4}$$

To illustrate this point, we have taken the small speciation model defined in equation 2.3.13. It consisted of three acid-base and one nucleation reactions. In first place, we define the formation reactions for each oxo-cluster. Equation 2.4.3 collects the four formation reactions, where the product is formed from the reference compound and the hydronium cation. The next step is to represent these reactions in a speciation model. Equation 2.4.4 depicts an speciation

model defined by formation reactions. Furthermore, the system of equations is still non-linear since the reference compound is powered to two. Next, we fit the scaled constants to equation 2.4.4, and we solve the NLE with the Powell algorithm. In this case we do not apply the Least Squares algorithm to doublecheck because this type of speciation model is exclusively employed for plotting purposes.

Figure 2.23 depicts the six speciation diagrams for the same models mentioned before but with scaled constants. Indeed, the speciation is very different from the previous examples. In first place, the pH range is no longer between 0 and 30 but between 0 and 10. This scale is more reasonable since polyoxomolybdates mainly form at acid pH. In second place, the relative pK_a values are still preserved; the molybdic acid appears at more acid conditions than the hydrogenmolybdic acid and than the molybdate anion. Finally, model 800 shows the formation of the octamolybdate, $[H_2Mo_8O_{26}]^{4-}$, at acid pH. This evidence is closer to the experiments as this cluster is one of the most predominant compounds in solution.

Despite the good results collected in Figure 2.24, the speciation diagrams are still substantially different from the experimental data. For example, none of the six models shows the formation of the heptamolybdate clusters. Additionally, model 800 does not form the molybdic acid at acid pH, which is in disagreement with the experiments. Although the linear scalings have corrected most of the error present in the theoretical formation constants, there is still some further improvement to be made.

Because formation constants are in logarithmic units, small errors have a large impact on the final prediction. For example, an error

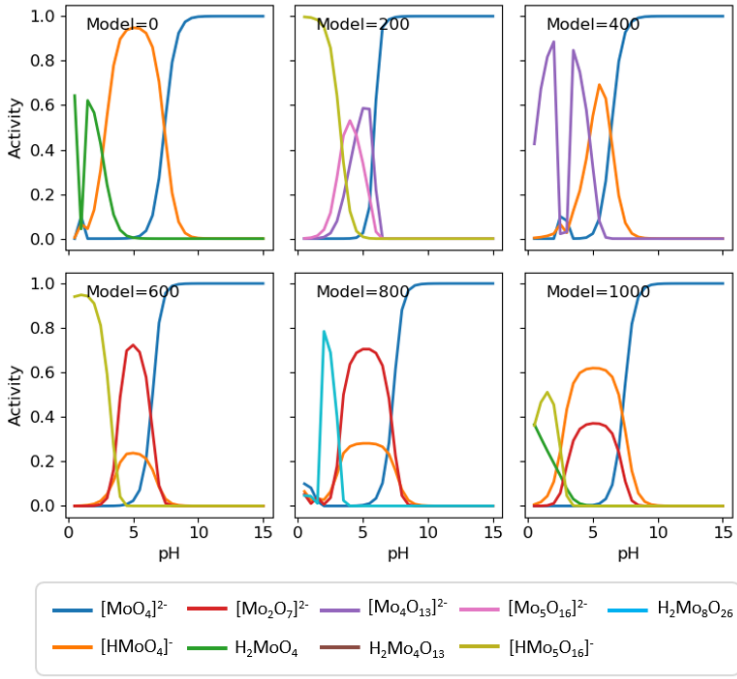


Figure 2.24: Speciation diagrams for the octaomolybdate molecular set. Scaled formation constants of models 0, 200, 400, 600, 800, and 1000.

of 1.36 kcal·mol^{-1} in the deprotonation Gibbs energy causes an error of 1 pK_a unit.[144] Therefore, it is not possible to rely purely on the R^2 for determining the best linear scaling. Instead, we must consider the root mean square error as an strict and accurate measure of the scalings' standard deviation. To determine the linear scaling with the highest accuracy, we calculated the RMSE for the 1024 speciation models. Next, we had to select the model with the lowest RMSE value. Figure 2.25A shows that the model with the lowest RMSE is the number 793 with a RMSE of 1.0 and a R^2 of 0.9990. The next step is to employ this linear scaling to calculate the new scaled constants.

Then, we set up the formation speciation model with these scaled constants, and we solve the system of NLE with Powell's algorithm.

Figure 2.25B shows the speciation diagram using the scaled constants. Notwithstanding, the result it not optimal yet. Firstly, the most stable oxo-clusters, such as the octamolybdate and heptamolybdate, do not appear at any value of pH. More importantly, there are numerical issues in the resolution of the speciation model: (i) the activities of each compound display discontinuities throughout the pH range, (ii) some oxo-clusters such as, $[Mo_3O_{10}]^{2-}$, present negative activities, (iii) the activities of some compounds (depicted in grey color) are erratic. Even if we are employing the best scaling, when we employ the 30 scaled constants the model does not perform well.

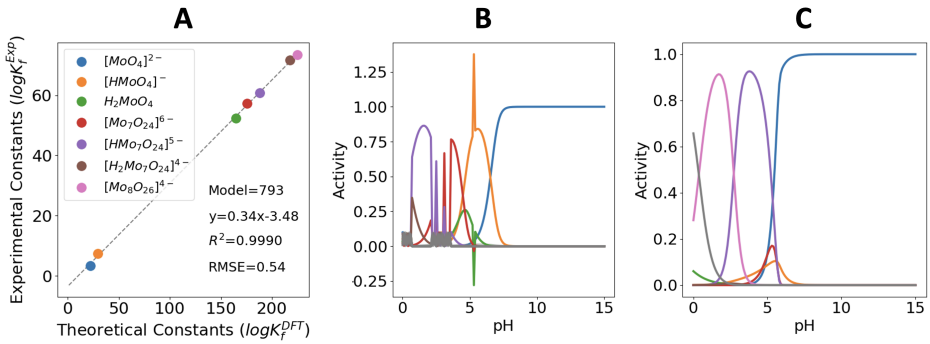


Figure 2.25: Left: linear scaling (A) with the lowest RMSE for the octamolybdate dataset. Center: speciation diagram (B) computed with the 30 scaled constants. Right: speciation diagram (C) computed with eight scaled constants. Color legend depicted in A is common for the three plots.

Despite calculating the scaled constants for the 30 oxo-clusters, it might not be possible to employ them all for building a robust

speciation diagram. A similar challenge occurred to Cruywagen and coworkers when determining the molybdate speciation with potentiometry techniques.[155] They proposed models with different clusters and studied which provided the best fittings. These experimental models contained the order of ten compounds. This narrow but precise selection made the speciation diagrams more reliable. Following a similar approach, we employed a smaller selection of scaled constants to plot the speciation diagram of the octamolybdate. These selection consisted in seven oxo-clusters: $[MoO_4]^{2-}$, $[HMoO_4]^{-}$, $[H_2MoO_4]$, $[Mo_7O_{24}]^{6-}$, $[HMo_7O_{24}]^{5-}$, $[H_2Mo_7O_{24}]^{4-}$, $[Mo_8O_{26}]^{4-}$, and $[HMo_8O_{26}]^{3-}$. Figure 2.25C shows the speciation using the scaled constants of the compounds mentioned above. In this case the speciation is very well described. Notice that the RMSE is below the unit (0.54). In general terms, the methodology works best when the $RMSE < 1.0$. Otherwise, the error in the speciation diagrams are too large. When we analyze Figure 2.25C we can observe that pH is within the correct range. Moreover, the hepta- and octamolybdates form at reasonable activities. Finally, the monomeric species have significant activity values. These different aspects demonstrate the reliability and accuracy of the chosen scaling.

We have shown that the RMSE of the linear scaling is a good indicator of *POMSimulator's* accuracy. Therefore, the RMSE can be regarded as a measure for optimizing our methodology. For example, the quantum mechanical method can be improved based on the results of the scaling. Indeed, the QM method has a crucial influence in the reaction energies. More concretely, one key component of the QM method is the DFT functional. Thus, we have compared the effect of

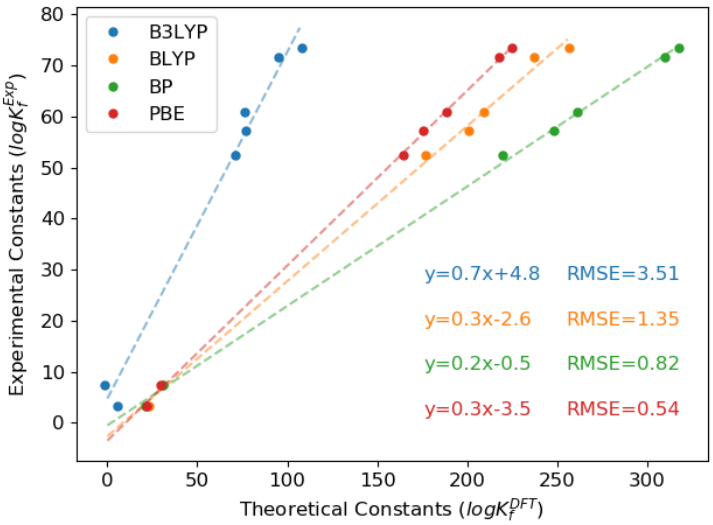


Figure 2.26: Linear scalings for four different DFT functionals with the lowest RMSE scalings.

four functionals: B3LYP, BLYP, BP, and PBE.[m] Figure 2.26 depicts the most accurate regression for each type of functional. In first place, we observe that the family of Generalized Gradient Approximation (GGA), formed by BLYP, PBE and BP are clustered together. Instead, the Hybrid GGA B3LYP has a significant higher slope. This difference can be related simply to the class of functional, and thus to the Hartree-Fock exchange. Moreover, the slope parameter of B3LYP (0.7) is close to one which is in stark contrast to the rest of functionals. This evidence suggests that B3LYP does not overestimate the formation constants as much as the rest of functionals do. Secondly, we observe that B3LYP performs very poorly, with an RMSE over

[m] More information about the computational details can be found in Appendix A.

the unit (3.51). Consequently, the corresponding scaled constants would rend useless due to the large variance error. Then, BLYP also provides unsatisfactory results with an RMSE of 1.35 even though not as poor as B3LYP. On the other hand, BP performs very well with an RMSE of 0.82. Thereby, this functional could be employed for calculating reaction energies since it would ensure high accuracy. Finally, the functional that provides the lowest error is PBE with a RMSE of 0.54 units.

2.4.2 Formation Mechanism

Hitherto, we have focused on the elucidation of the speciation diagrams. Nonetheless, there is another paramount aspect to be addressed: the self-assembly mechanism. Previous works succeeded in studying this process with static and dynamic calculations.[73,156] However, in this thesis we have approached the formation mechanism from a different angle. Instead of deducing the nucleation process with metadynamic calculations, we have employed a distinct approach. We rely on the hypothesis that we can infer the mechanism from the most accurate speciation model. In this manner we can consider the effect of not only the DFT reaction energies, but also the pH, ionic force, and concentration effects.

Figure 2.27 shows an schematic overview of this approach. In first place, we observe that a given model (in this case, model 793) provides the best linear scaling. Consequently, we employ this scaling to correct the formation constants of the model. In case of numerical instabilities we make a selection of clusters for predicting the speciation diagram. Then, we analyse the diagram and we compare it to

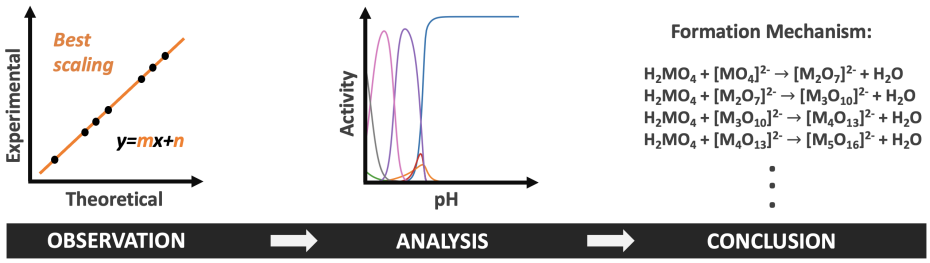


Figure 2.27: Schematic overview of the process for inferring the self-assembly mechanism from the speciation model.

the results reported in the literature. If the speciation agrees with the experimental evidences, we regard the model as the most accurate. Once we have determined the best speciation model, we infer the formation mechanism from the model itself. To do so we concluded that: *the self-assembly mechanism is defined by the speciation model that best agrees with the experiments.* Therefore, there is a direct relationship between the reactions that define a speciation model, and the formation mechanism.

Acid-base reactions are included in every speciation model, so they are always present in the reaction mechanism. On the other hand, the nucleation reactions depicted by the best speciation models correspond to the formation mechanism. Equation 2.4.5 shows that there are four condensations and five additions reactions (reaction energies are in kcal· mol^{-1}). Notice that the larger clusters are formed by additions whereas the smaller ones by condensations. This reaction mechanism is not the first one published for the formation of the octamolybdate. Indeed, Steffler et al. performed a very similar study where they made a meticulous analysis with static calculations. They concluded that the molybdenum tetraoxide, $[MoO_4]^{2-}$, was the

main building block, and thus it appeared in the most favourable reactions. However, they also pointed out that a more accurate modelling was required for considering parameters such as temperature, pH and ionic strength.

$$
\begin{aligned}
H_2MoO_4 + H_2MoO_4 &\rightleftarrows H_2Mo_2O_7 + H_2O && \Delta G_r = 1.00 \\
H_2MoO_4 + H_2Mo_2O_7 &\rightleftarrows H_2Mo_3O_{10} + H_2O && \Delta G_r = 1.49 \\
[HMoO_4]^- + [HMo_3O_{10}]^- &\rightleftarrows [Mo_4O_{13}]^{-2} + H_2O && \Delta G_r = -9.53 \\
[HMoO_4]^- + [HMo_4O_{13}]^- &\rightleftarrows [Mo_5O_{16}]^{2-} + H_2O && \Delta G_r = 10.38 \\
[MoO_4]^{2-} + [Mo_4O_{13}]^{2-} &\rightleftarrows [Mo_5O_{17}]^{4-} && \Delta G_r = 11.95 \\
[HMoO_4]^- + [HMo_5O_{16}]^- &\rightleftarrows [H_2Mo_6O_{20}]^{2-} && \Delta G_r = 4.06 \\
[HMoO_4]^- + [HMo_6O_{20}]^{3-} &\rightleftarrows [Mo_7O_{23}]^{4-} && \Delta G_r = -0.48 \\
[MoO_4]^{2-} + [Mo_6O_{20}]^{4-} &\rightleftarrows [Mo_7O_{24}]^{6-} && \Delta G_r = 25.36 \\
H_2MoO_4 + [H_2Mo_7O_{23}]^{2-} &\rightleftarrows [H_2Mo_8O_{26}]^{2-} && \Delta G_r = -7.79
\end{aligned}
\tag{2.4.5}
$$

If we compare the mechanism reported by Steffler et al.[60] and the one proposed in this thesis, we observe that $[H_mMoO_4]^{2-}$ appears in all the chemical reactions. This evidence is in very good agreement with the previous theoretical work. Nonetheless, the mechanism reported in this thesis considers the molybdenum tetraoxide in three distinct protonation states. This feature makes the mechanism more realistic because it accounts for the effect of the pH. Furthermore, the reaction energies in equation 2.4.5 are mostly endergonic. It seems counter-intuitive that the best formation mechanism has a positive reaction energy. We think that it is due to the weight of other parameters such as the ionic strength, pH, and concentration

effects. Consequently, it is not possible to rely exclusively on reaction energies for predicting the most likely mechanism. This feature is applied in chapters 3 and 4 when we discuss the reaction mechanism of each polyoxometalate family.

2.4.3 Acid Dissociation Constants

The calculation of the linear scaling is not restricted to formation constants. In fact, there is another type of equilibrium constant which is frequently employed in polyoxometalates: acid dissociation constants. In this case, the speciation models are more simple because they only consider acid-base reactions. For instance, the speciation model described for the carbonic acid (equation 2.3.11) fits into this definition. The procedure would be very similar to the one we described for formation constants: setting up the system of linear equations, diagonalizing the matrix, and expressing the activities as acid-base constants. Next, these theoretical constants are compared with the experimental ones, and scaled with the appropriate linear scaling.

Figure 2.28 left shows the linear scaling of the acid-base constants for a set of 13 polyoxometalates (marked with circles). Notice that there are oxo-clusters with different metal centers and shapes. For instance, there are five types of metal centers: molybdenum, tungsten, vanadium, niobium and tantalum metal-oxo clusters. Additionally, the set also contains distinct topologies such as the decavanadate, octamolybdate, heptametalate, and hexametalate. In similar manner than for formation constants, the theoretical acid dissociation constants are overestimated as well. While the ex-

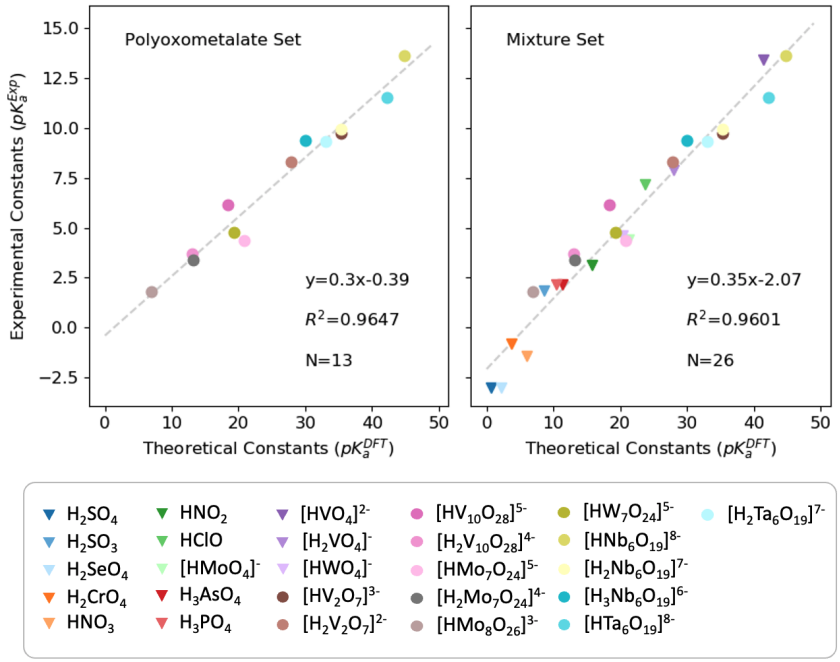


Figure 2.28: Linear scalings using acid dissociation constants. Left: regression with a set of polyoxometalates (marked with circles). Right: regression with set of polyoxometalates and monomeric oxoacids (marked with triangles).

perimental constants are between 1 and 13 of pK_a, the theoretical ones are within 0 and 45. Nonetheless, the relative pK_a are well reproduce. For example, $[HMo_8O_{26}]^{3-}$ is the strongest acid oxocluster whereas $[HNb_6O_{18}]^{8-}$ is the weakest one. Besides, Figure 2.28 shows that POMs' pK_a can be roughly classified in two groups: acids ($pK_a^{Exp} \approx 2.5 - 6.5$) and alkalines ($pK_a^{Exp} \approx 7.0 - 13.0$). In the acid group we find molybdenum and tungsten clusters, in good agreement with their aqueous speciation.[154] In contrast, niobium and tantalum acids only appear in the alkaline group. This evidence

fits very well with the poor acidity and strong basicity of these clusters.[157] Vanadates are an intermediate case because there are clusters in the acid group, $[H_xV_{10}O_{28}]^{x-6}$, as well as in the alkaline group, $[H_xV_2O_7]^{x-4}$. In fact, this broad range of pH is a trademark of polyoxovanadates.[158]

Although Figure 2.28 left includes different clusters, we found a very good linear relationship ($R^2 = 0.9647$). Furthermore, the linear scaling is very similar to the scaling calculated for six formation constants in Figure 2.23. The fact that both scalings are so alike is remarkable. We attribute this similarity to presence of the proton species in their respective reactions. Thereby, it is reasonable to assume that the linear scaling parameters do not depend on the type of equilibrum constants but on the presence of the proton species. Not only that, but also the scaling does not depend on the metal center either. Despite the fact that there are five different types of metals, the linear scaling is still very good.

Motivated by the excellent results depicted in Figure 2.28 left, we decided to test whether this linear trend was maintained for other monomeric oxo-acids. More concretely, we focused on classical inorganic acids. As a matter of fact, they are also oxo-compounds even if they do not form large clusters. We have considered the following oxo-acids: sulfuric, H_2SO_4, sulfurous, H_2SO_3, selenic, H_2SeO_4, cromic, H_2CrO_4, nitric, HNO_3, nitrous, HNO_2, hypoclorous, $HClO$, arsenic, H_3ArO_4, and phosphoric, H_3PO_4. We have also included in this test the monomeric forms of the polyoxometalates: $[HMoO_4]^-$, $[HVO_4]^{2-}$, $[H_2VO_4]^-$, and $[HWO_4]^-$.

Figure 2.28 right shows the linear scaling including the constants

of both the 13 polyoxometalates and the 13 small oxo-acids (marked with triangles). The first consequence of these additions is the pK_a range has been expanded due to the presence of strong oxo-acids such as the sulfuric, nitric, and cromic acids. While the theoretical constants are close to zero, the experimental ones have a negative value. The rest of small acids are homogeneously distributed along the pK_a range. More importantly, Figure 2.28 right shows that there still is a very good linear relationship ($R^2 = 0.9601$). Thus, the addition of the small oxo-acids did not worsen the scaling. Even so, the linear parameters have changed; the slope is greater than Figure 2.28 left ($m = 0.35$) whereas the intercept has a lower value ($b = -2.07$). This difference is mainly because of the strongest oxo-acids depicted in the lower part of Figure 2.28. Notice that the small inorganic monomers are systematically more acid than the rest of POMs. Even so, the regression is still robust and accurate.

2.4.4 Universal Scaling

Hitherto, we have observed some remarkable trends concerning the linear scalings. In first place, formation constants shared a common slope for the seven scalings considered in Figure 2.23. Moreover, the best scaling depicted in Figure 2.25 shared the same m parameter. In second place, acid dissociation constants also depicted a very alike behaviour. In fact, Figure 2.28 went a step further since it defined a sole linear equation for a set of different polyoxometalates. Besides, the slope parameter remained the same as for the formation constants: $m = 0.3$. We decided to analyze this sequence of evidences to determine if they were random or there was a logical pattern

underneath.

Figure 2.28 left demonstrated that acid dissociation constants could be scaled regardless of the metal type. Thus, we found appropriate to evaluate this remarkable trend for the formation constants. To do so we have collected the linear scalings reported in chapters 3 and 4. Before diving into the results it is worth highlighting two technical aspects: (i) each polyoxometalate family has a different amount of speciation models. For instance, polyoxovanadates have over one million models whereas polyoxomolybdates have over one hundred thousand models. Thereby, we must normalize the number of models of each family to enable the comparison between them. (ii) There is a huge amount of data because of the combinatorial nature of the speciation models. Consequently, we represented the data in continuous plots instead of a discrete ones.

Figure 2.29 left shows the slope distributions for the five systems. We can observe that the five functions are mostly centered at 0.3. This evidence is paramount since it reinforces the results obtained in the previous sections. The fact that Figures 2.23 and 2.25 depicted the same slope parameter was not fortuitous. Next, Figure 2.29 right collects the distribution of intercept values. In this case there is a greater variance compared to the slope parameter. Molybdenum and tungsten share the same average value, $b \approx 18$, vanadium is equal to $b \approx 5$, and niobium and tantalum are equal to $b \approx -1$.

At present we have two hypothesis for explaining the variance of the intercept values.

1. Dependence on the chemical reactivity: initially we thought that this variance could be related to the type of chemistry

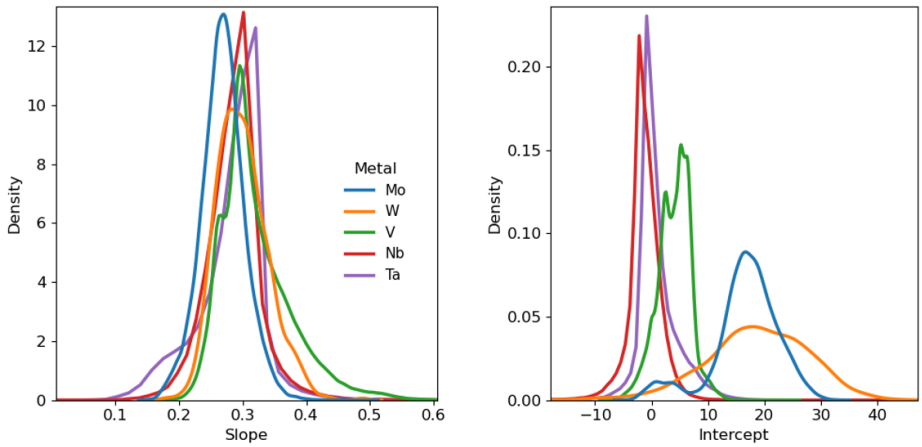


Figure 2.29: Left: density histograms for the slope values. Right: density histogram for the intercept values. Color indicates the polyoxometalate system.

of each system. For instance, tungsten and molybdenum have similar aqueous speciations. They share common monomeric species and the heptametalate structure. Then, vanadium is an intermediate case between molybdenum/tungsten and niobium/tantalum. On one hand, it has a rich acid chemistry in similar manner to molybdenum and tugnsten. On the other hand, it forms the decametalate structure, which only exists in the niobium and tantalum systems. Therefore, it seemed reasonable that vanadium had an intermediate intercept value. Finally, niobium and tantalum have almost identical intercept values. This is also reasonable because they have similar alkaline aqueous speciations.

2. Dependence on the reference compound: formation constants

employ different reference compounds depending on the chemical system. This means that formation reactions are not always the same. On the contrary, acid dissociation constants are always defined by the acid and the conjugated base. This subtle but relevant difference might have an effect on the linear scaling. For example, molybdenum and tungsten employ the tetraoxide, $[MO_4]^{2-}$, vanadium employs the hydrogentetraoxide, $[HVO_4]^{2-}$, and niobium and tantalum employ the Lindqvist, $[M_6O_{19}]^{9-}$. Interestingly, the reference clusters are correlated to the pattern observed in Figure 2.29.

Between the two hypothesis detailed above, we believe that the second one is the most likely. We already observed in Figure 2.28 that the chemical reactivity did not affect the scaling. For instance, notice that the acid dissociation scaling included acid and basic polyoxometalates, but the linear scaling remained the same for all the oxo-clusters. Furthermore, formation constants also consider the proton species (as acid dissociation constants) so it is incongruous that they distinguish intercept values based on the pH reactivity. Consequently, this first hypothesis does not explain the difference behaviour observed in the intercept distributions. On the other hand, the second hypothesis offers a better explanation of the variance in the intercept value. The formation reactions differ due to the reference cluster, and thus it seems reasonable that the intercept might change as well.

The discussion developed in this section has a transcendental importance in the present and future of *POMSimulator*. If a unique linear scaling could be found for formation constants of different sys-

tems, we would have attained a universal scaling. This would be a major milestone for our methodology. We would not need experimental constants for each system, thus reducing the dependence on experimental data. So far, results indicate that there is potential for a universal scaling, but further work is required. Although it is beyond the scope of this thesis, it would be very interesting to test the universality of the scaling with other polyoxometalates systems. For instance, heteropolyoxometalates are a different type of clusters which include heterometals such as phosphor, arsenic and silicon. This molecule set would challenge the applicability and universality of the whole protocol.

2.5 Speciation Parameters

In the past sections we have explained in detail the development of *POMSimulator*. We showed the construction of the reaction network, the resolution of the speciation models, and the scaling of the equilibrium constants. With these steps we succeeded in building a robust methodology for studying the chemical speciation. In the present section we are going to employ this methodology to study the effect of two parameters in the speciation of the octamolybdate system. In first place, we analyse the effect of the total concentration in the speciation. To do so, we implemented the calculation of phase diagrams. In second place, we also investigate the effect of the ionic strength in both the scaling and the speciation.

2.5.1 Total Concentration

The total concentration plays a very important role in the speciation of polyoxometalates. For instance, at low concentrations the formation of large clusters is hindered. Thereby, the speciation consists mainly on metal monomers at different protonation states. This lack of nucleation is due to diffusion effects. Because the concentration is low ($< 10^{-4}M$), the probability of two metal clusters reacting is unlikely.[154] In contrast, at higher concentrations ($> 10^{-3}M$) the nucleation is already favored. Despite the relevance of the concentration parameter, computational studies have often disregarded this parameter. This is because energy profiles cannot consider this effect. Nonetheless, we can examine the dependence of this feature relying on the resolution of the speciation models.

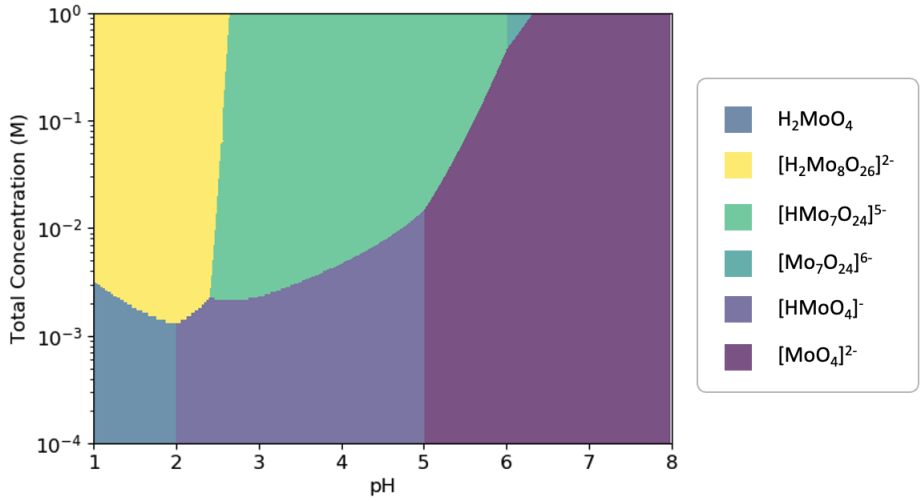


Figure 2.30: Phase diagram for the octamolybdate system computed with the scaled constants at 298.15K, 1 atm, and 0.6 M of NaCl.

To study the effect of the total concentration we must solve the speciation models at multiple concentrations. This would result in many speciation diagrams where we would have to analyse the relative abundances of every oxo-cluster. This approach has two drawbacks: (i) it generates huge amounts of data since the concentrations have to be kept at different values of pH and concentrations (ii) interpreting the results from multiple diagrams is not straightforward, specially because the concentration scale might differ by several orders of magnitude. Therefore, it is necessary to represent the data in a unified plot which summarizes the data from multiple speciation diagrams. This type of plot is called phase diagram, and it represents the most abundant compound at a given pH and concentration. It is similar to the Pourbaix diagrams, but the y-axis represents redox potential instead of total concentration. Phase diagrams are less frequent in the literature, and in fact we have only found a complete diagram for polyoxovanadates.[159] Even so, our methodology offers a workflow for predicting the phase diagrams of any system in an automated manner.

Figure 2.30 collects the phase diagram for the octamolybdate set. We have employed the formation constants collected in Table 2.2. Notice that the diagram represents the total concentration in logarithmic scale. We have used this scale to examine the speciation both at high and low concentrations. Otherwise, it would not be possible to observe the abundance of the monomeric clusters. However, the direct consequence of using this scale is that we have to rely on a very precise grid. More concretely, we have employed 10^4 values of pH and concentrations. As a result, the phase diagram relies on the

resolution of 10^{16} systems of NLE.

Figure 2.30 shows that below a total concentration of $10^{-3}M$ the formation of large clusters does not take place. In these conditions, we only observe the molybdenum tetraoxide at different protonation states: H_2MoO_4 (1-2 pH), $[HMoO_4]^-$ (2-5 pH), and $[MoO_4]^{2-}$, (5-8 pH). In contrast, at concentrations greater than 10^{-3} M, the nucleation is favored. For example, at very acid pH (1-2) the most abundant compound is the octamolybdate, $[H_2Mo_8O_{26}]^{2-}$. Then, at an still acid pH (2-6) the protonated form of the heptametalate, $[HMo_7O_{24}]^{5-}$ predominates. Then, the formation of the non-protonated heptametalate, $[Mo_7O_{24}]^{6-}$, is favored at high concentration and pH close to neutral. Finally, from neutral pH to more alkaline conditions the most abundant compound is the molybdenum tetraoxide. Thereby, molybdates' nucleation is favored by high concentrations and acid conditions.

In our opinion, phase diagrams are very useful for carrying out an overview of the aqueous speciation. Moreover, they can serve for guiding experiments in systems which are less studied. Nonetheless, we must highlight that these diagrams only inform about the most abundant compounds. Therefore, the rest of minority species are disregarded even though they might be in considerable concentrations. The extreme case would be if the activity of one compound is 0.51. In this case, only this compound would appear in the diagram in spite of having another compound with an activity of 0.49. To avoid misinterpretations, it is essential to combine the information of phase diagrams with speciation diagrams. In this manner, we can obtain both a general and concrete picture of the speciation.

2.5.2 Ionic Strength

The aqueous speciation of polyoxometalates is also affected by the ionic strength (I). We saw in section 2.3.3 several models for calculating the effect of I in the activities. It is worth noting that formation constants do depend on the ionic strength at which they were calculated. This is a relevant difference compared to the effect of the total concentration. While for phase diagrams we could employ the same set of formation constants, it is not possible to do so with solutions at different ionic strengths. Therefore, to study the effect of this parameter in the speciation we must employ experimental constants reported at different I conditions.

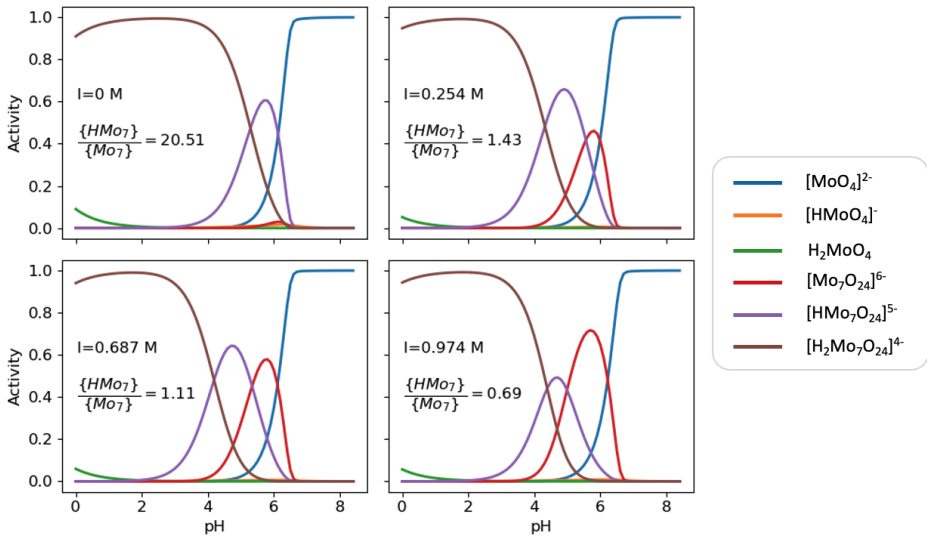


Figure 2.31: Speciation diagrams with scaled constants at four different ionic strengths: 0 M, 0.254 M, 0.687 M, and 0.974 M NaCl. Ratios between the maximum activities of the monoprotonated $\{HMo_7\}$ and the nonprotonated $\{Mo_7\}$ hetptamolybdates.

Francesco Crea et al. published an article where they reported formation constants of mono- and heptamolybdates at different I conditions.[143] Thus, we have employed these constants as reference in the scaling of our methodology. In first place we set up four simulations at ionic strengths of 0 M, 0.254 M, 0.687 M and 0.974 M NaCl. Notice that all four cases fit within the range of applicability of *Geochem* activity function (I < 1.0 M). We compute the formation constants at each ionic strength and then we rescale them with the experimental data.

Figure 2.31 shows four speciation diagrams calculated with the scaled constants. The diagrams are very similar to the ones reported by Francesco Crea et al. thus proving the high accuracy of the scaling. Essentially, the four diagrams are rather similar: at alkaline conditions (pH > 6.5) the molybdenum tetraoxide is the most abundant species. In contrast, at acid conditions (pH < 3) the diprotonated heptamolybdate $[H_2Mo_7O_{24}]^{4-}$ predominates. Despite these similarities, the ionic strength causes some fluctuations in the pH range between 4 and 6. More concretely, the activity ratio between the monoprotonated {HMo_7} and nonprotonated {Mo_7} heptamolybdate changes substantially. To illustrate this variation we have calculated the ratio between the maximum peak activities of each cluster. For instance, at I of 0 M, the ratio is 20.51 which indicates that {HMo_7} is twenty times greater than {Mo_7}. An alike trend is observed at I 0.254 M and 0.687 M; {HMo_7} has a greater activity than {Mo_7} and thus the ratio is greater than 1. On the other hand, at I 0.74 M the activity of {Mo_7} is greater than {HMo_7}. Consequently, the ratio is under the unit (0.69).

We have observed that I has a direct effect on the activity ratio between the monoprotonated and diprotonated heptamolybdate. In fact, anionic species tend to be more stable at higher ionic strength conditions. For example, $\{Mo_7\}$ has a formal charge of -6 whereas $\{HMo_7\}$ of -5. Thus, the former is more stable than the latter at I of 0.974 M. Finally, we have proved that our methodology also contemplates the effect of the ionic strength in the aqueous speciation.

2.6 Summary

In the first part of this chapter we developed an algorithm which constructed a chemical reaction network based only on the atomic number and chemical connectivity of every molecule. This reaction network consisted only in minima calculations, and thus no transition states were included. While the atomic numbers' translation to graphs did not pose any problem, it was not the case for the chemical connectivity. Chemical bonds can result difficult to characterize, specially when clusters present multiple metal geometries (e.g., tetrahedral, pentagonal pyramid, and octahedral). Therefore, we employed Bader's theory of Atoms in Molecules to determine the chemical bonds. This approach led to some ambiguities but it offered a good guess of the connectivity. To disregard any uncorrect bond, we checked the final connectivity of the molecules before proceeding with the application of Graph Theory properties. Then, we evaluated the isomorphic property between all the molecular graphs. This task was computationally-demanding due to inherent complexity of searching recursively topological patterns. To minimize the poten-

tial bottleneck-effect of this step, we simplified the searching space. Instead of calculating first the isomorphism for a pair (g_i, g_j), or a trio (g_i, g_j, g_k) of graphs, we computed the stoichiometric balance. If the resulting balance corresponded to a tabulated reaction (eg., condensation, addition, acid base), we proceeded with the evaluation of the isomorphic property. Otherwise, we did not apply this topological property thus reducing the computational cost. We conclude that the actual construction of the network is efficient and reliable, thus ensuring that it can be employed for evaluating the aqueous speciation.

The second part of this chapter was focused on the prediction of the polyoxometalates' aqueous speciation. We developed an algorithm which set up speciation models based on the chemical reactions collected in the network. Because there were more reactions (equations) than compounds (variables), we faced an overdetermined system. To approach this problem, we had to set up all the possible combinations of models based on the binomial coefficient. Due to the factorial growth related to the number of models, we defined two chemical assumptions. These aimed at reducing the number of combinations while maintaining chemically sound models. The first assumption stated that every model had to contain all the acid-base reactions. The second assumption stated that every model had to contain all the nuclearities. With these two assumptions we succeeded in both reducing the computational cost, and disregarding poor models. Moreover, we implemented a variation of the Davies equation for calculating the activities in high ionic strengths (> 1 M NaCl). We did not chose more sophisticated functions such as

the Pitzer equations due to the lack of experimental parameters. To solve the speciation models we employed Powell's minimizing method. Additionally, we created an accuracy threshold based on the comparision between the solution of Powell's and Least Squares algorithm. We found out that including this threshold led to better speciation models and thus more accurate formation constants. Besides, we implemented a batch paralelization for the resolution of the speciation models. In this manner, the method benefited from the availability of multi-thread hardware. This resulted in a speed-up factor of up to ≈ 28 times. The speciation algorithm solved the models successfully, but the constants were biased towards alkaline pH. Despite that the relative pK_a of the compounds were reasonable, the absolute values did not correspond to the experiments.

In the third part of this chapter we approached and solved the problem related the bias in the theoretical formation constants. We found out that the theoretical constants were linearly related to the experimental ones. In fact, the accuracy of the linear scaling was exceptionally high (e.g., RMSE < 0.5). This excellent results unlocked the possibility of not only reproducing the experimental constants, but also predicting unreported ones. We showed that thanks to the linear scaling, the speciation diagrams fit excellently with the experimental data. Moreover, we made an important assumption: we established a direct relationship between the most accurate speciation model and the formation mechanism. Therefore, the set of nucleation reactions that provide the best fitting with the experiments would be the reactions involved in the mechanism. In this chapter, we compared the chemical reactions of the best molybdenum speciation

model to the previous theoretical work. We found that there was a good agreement in the condensation and additions reactions that yielded the octamolybdate. Next, we found that the acid dissociation constants were also very accurately related to the experimental ones. In fact, we proved that this relationship did not only involve oxo-clusters but small oxo-acids as well. For example, strong inorganic acids such as sulfuric, nitric, selenic, and cromic. Thus, the fact that one linear regression describes several oxo-compounds of different size is a major achievement. Besides, it provides a solid evidence that the method and its implicit assumptions are chemically-sound. Last but not least, the linear regression parameters' were surprisingly similar (specially the slope, m≈0.3) for the five families of polyoxometalates. This led us to propose a potential universal scaling for correcting the theoretical constants. If this hypothesis was true, the scaling would no longer depend on experimental data. As a consequence, the scope of the method application would be greatly enlarged. To test this hypothesis it would require a wide and systematic study which is beyond the scope of this thesis. However, we envision that this future work could become very useful for the synergy between experimental and theoretical chemistry.

The fourth and last part of this chapter focused on testing the effect of two relevant parameters: the total concentration, and the ionic strength. To study the effect of the initial concentration of the precursor we implemented speciation phase diagrams. These diagrams were very useful for determining the most concentrated specie at a given pH and concentration. The effect of this parameter cannot be underestimated since below 10^{-3} M nucleation is largely hindered

and thus only monomeric compounds are present. Moreover, phase diagrams are a useful tool for performing an overview of the aqueous speciation of a system. With a simple diagram, we can obtain a rather accurate picture of the most important clusters. In addition, we also studied the effect of the ionic strength. We found out that higher ionic strengths led to the stabilization of more anionic species. For example, $\{Mo_7\}$ was more abundant at I=0.974 M NaCl than at I=0 M NaCl. In constrast, $\{HMo_7\}$ was more abbundant at low ionic strength conditions. This behavior agreed with the experimental data, therefore proving the correct modeling of this key parameter.

Chapter 3

Molybdenum and Tungsten Isopolyoxoanions

> The highest activity a human being can attain is learning for understanding, because to understand is to be free.
>
> — Baruch Spinoza

3.1 Introduction

Group VI polyoxometalates consists of molybdenum and tungsten molecular oxo-clusters. Both families of polyoxometalates are very relevant in the development of this research field. Polyoxomolybdates have applications in photochromism and electrochromism,[160] water splitting processes,[161] and catalysis.[162] Then, polyoxotungstates have also been applied to electrochemistry[163] and photochemistry[164] (e.g, the decatungstate $[W_{10}O_{32}]^{4-}$ is a renowned photocata-

lyst). Moreover, both molybdenum and tungsten have a rather similar speciation behaviour. For example, both metal-clusters yield a solid precipitate when subjected to very acid conditions. Molybdenum forms a yellow oxide, $MoO_3 \cdot 2H_2O$, whereas tungsten forms a white oxide, $WO_3 \cdot 2H_2O$. Furthermore, both metals present a stable monomeric specie with a tetrahedral geometry: $[H_xMO_4]^{x-2}$.

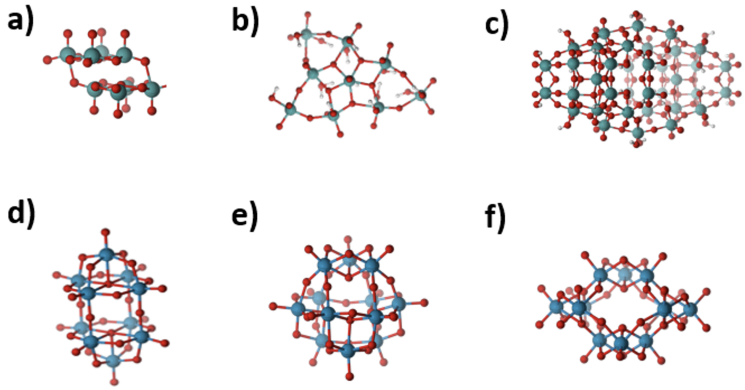


Figure 3.1: Representative structures for molybdenum and tungsten oxo-clusters. Ball sticks representation of a) octamolybdate $[Mo_8O_{26}]^{4-}$ ($\{M_4\}$), b) pentagonal derivate $[H_xM_7O_y]^{z-}$ $\{M_6M_1\}$, c) $[H_{32}M_{36}O_{128}]^{8-}$ $\{M_{36}\}$, d) $[M_{10}O_{32}]^{4-}$ $\{M_{10}\}$, e) α-Keggin, metatungstate, $[M_{12}O_{40}]^{8-}$ $\{\alpha\text{-}M_{12}\}$ f) paratungstate-B $[M_{12}O_{42}]^{12-}$ $\{M_{12}\}$.

The speciation of molybdenum in acidified solutions has been studied since the 1960s. Several experimental techniques such as potentionetry, NMR spectroscopy, and ESI-MS have been employed to determine formation constants of polyoxomolybdates. Pioneer researchers on this field were Tytko and Glemser,[165] Pettersson,[166] Maksimovskaya,[167] Cruywagen[155] among others. For instance, relevant constants were discovered regarding the heptamolybdate, ($[H_xMo_7O_{24}]^{6-x}$), the octamolybdate ($[H_xMo_8O_{26}]^{4-x}$; Figure

3.1A), and the largest isopolyoxomolybdate: $[H_{32}Mo_{36}O_{128}]^{8-}$ (Figure 3.1C). Because of the extensive work carried out on this field, more recent studies have been focused on the formation mechanism. For example, computational studies have addressed this topic employing static[60] and metadynamic[73] calculations. Both works showed that the self-assembly process of polyoxomolybdates is thermodynamically favorable, in good agreement with the experiments. However, recent studies have usually overlook the role of the speciation. This is largely because of the difficulty attached to the description of the complex equilibria reactions involved. For instance, the pentagonal structure and its derivatives (e.g, Figure 3.1B) have been known for several decades but no formation constant has been reported so far.

Tungsten oxo-clusters are generally more stable than their molybdenum analogues. This difference is due to the enthalpy factor of the condensation reactions when the tetrahedral metal centers expand to octahedral geometries. This energy difference is explained based on the larger force constants of the tungsten-oxygen bonding compared to the molybdenum-oxygen bonding.[168] As a consequence, the dynamism in the multi-equilibria of tungsten oxides is more limited. This makes more difficult to study the speciation of isopolyoxotungstates. Despite this, Cruywagen[150] and Rozantsev[169] succeeded in calculating the formation constants of the most important tungsten clusters. Concretely, they determined the constants of: the tungstate-Y ($[H_xW_{10}O_{32}]^{x-4}$; Figure 3.1D), the paratungstate A $[H_xW_7O_{24}]^{x-6}$, paratungstate B ($[H_xW_{12}O_{42}]^{x-12}$; Figure 3.1F) and α-Keggin metatungstate ($[H_xW_{12}O_{40}]^{x-10}$; Figure 3.1E). Analogously to the molybdenum system, there are numerous computational studies focusing

on the reaction mechanism. For example, Vilà-Nadal et al. have reported several works focusing on: the monomer expansion from tetrahedral to octahedral geometry,[170] the formation of the Lindqvist $[W_6O_{19}]^{2-}$,[156] and the initial steps of the heteropolyoxotungstates' nucleation.[171]

3.2 Speciation Results

We have defined two molecular sets for polyoxomolybdates and -tungstates. Our molybdenum molecular set is formed by 22 different topologies, and the total number of molybdates is 72. It is noteworthy that this molecular set is larger than the one employed in chapter 2. On the one hand, it shares the heptamolybdate ($[H_xMo_7O_{24}]^{6-m}$) and the octamolybdate ($[Mo_8O_{26}]^{4-}$; $\{Mo_8\}$, Figure 3.1A). On the other hand, it also includes pentagonal structures ($\{Mo_6Mo_x\}$) such as nonamolybdate ($[H_xMo_9O_y]^{z-}$; $\{Mo_6Mo_3\}$, Figure 3.1B). This type of oxo-clusters are relevant for the formation of the largest isopolyoxomolybdates. For example both the $\{Mo_{36}\}$ (Figure 3.1C) and its half analogue $[H_{16}Mo_{18}O_{65}]^{6-}$ ($\{Mo_{18}\}$) contain pentagonal motifs in their structures.

Then the molecular set of tungstates is formed by 17 nuclearities and 50 chemical species. It shares similar structures with the molybdenum set, specially for the smallest clusters. For example, the monomers, $[H_xMO_4]^{x-2}$, dimers, $[H_xM_2O_7]^{x-2}$, trimers, $[H_xM_3O_{10}]^{x-2}$, tetramers, $[H_xM_4O_{13}]^{x-2}$, and pentamers, $[H_xM_5O_{16}]^{x-2}$. Nonetheless, the tungsten molecular set also includes species that are characteristic of this type of polyoxometalates. For

example, the tungstate-Y ($[H_xW_{10}O_{32}]^{x-4}$; Figure 3.1D), the paratungstate A $[H_xW_7O_{24}]^{x-6}$, paratungstate B ($[H_xW_{12}O_{42}]^{x-12}$; Figure 3.1F) and α-Keggin metatungstate ($[H_xW_{12}O_{40}]^{x-10}$; Figure 3.1E). To determine the protonation sites we have employed the electrostatic potentials (ESP) as well as literature evidences. Moreover, each nuclearity is considered in at least two protonation states to account for the acid-base properties of the metal-oxo clusters. The molecular geometries have been optimized using a DFT methodology that is described in the appendix A. A dataset collection is available[172] in the ioChem-BD repository.[173]

In the first place, we computed the CRN for molybdenum and tungsten oxides, which are formed by 97 and 75 chemical reactions, respectively. We convert molecules to graphs and then we map the isomorphic property to decide which molecules can be related by a chemical reaction. We have defined four types of reactions: acid-base equilibria, condensation, addition and dimerization. The complete lists of reactions, with their corresponding free energies, are collected in Appendix C. Figure 3.2 shows an example of the reaction map involving the 50 polyoxotungstates. Most of the reactions present a reaction energy of approximately -10 kcal·mol^{-1}. Therefore, we can conclude that the formation of larger clusters is already favorable from a thermodynamic point of view. Furthermore, the monomeric tungstate, $[H_xWO_4]^{2-}$, plays a key role in the self-assembly process. Notice that $[HWO_4]^-$ has a high degree of connectivity with other molecular graphs. This evidence showcases that it has an active role in the formation of the oxo-clusters.

Despite the chemical likeness of molybdenum and tungsten, their

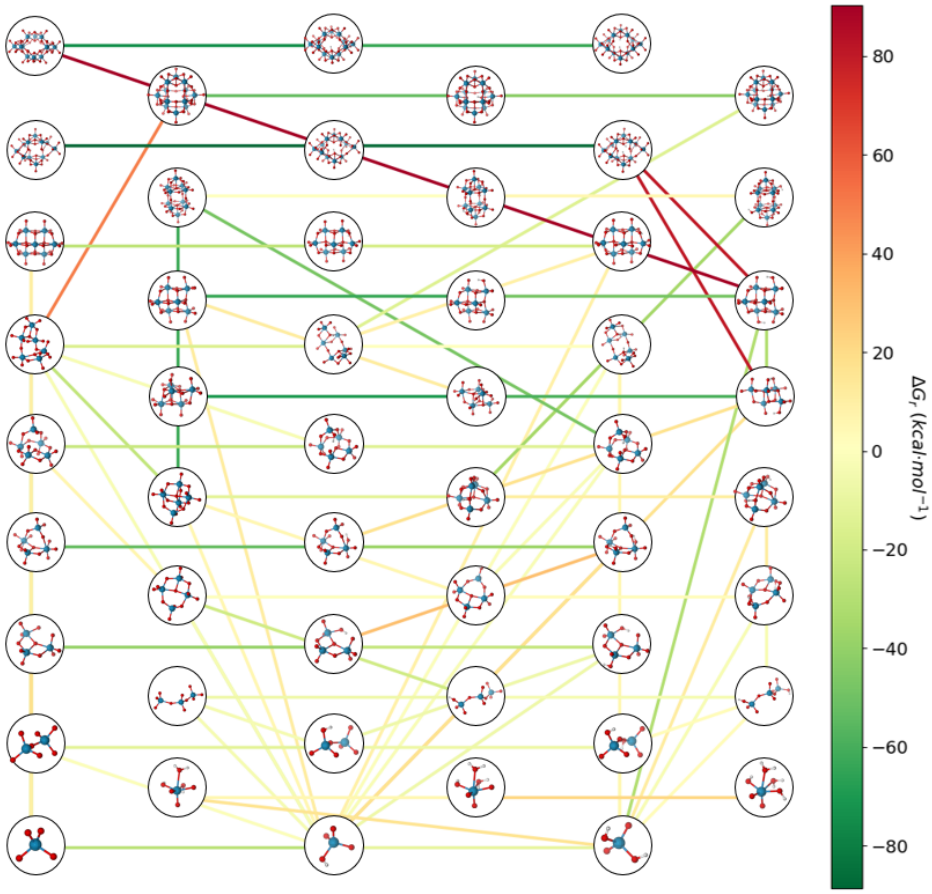


Figure 3.2: Chemical reaction map for the polyoxotungstate system. Edges represent chemical reactions and the color the ΔG_r in kcal·mol^{-1}. Reaction energies should be read from bottom-up and left-right directions.

speciations 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 3.3 shows the DFT formation constants of these structures (in box plots). Green and orange colors correspond to the molybdenum and tungsten systems, respectively. Additionally, there is an inset plot that shows the tridimensional geometry of the heptamer $[M_7O_{24}]^{6-}$. It is worth noting that molybdates formation constants show a larger variance than tungstates 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.

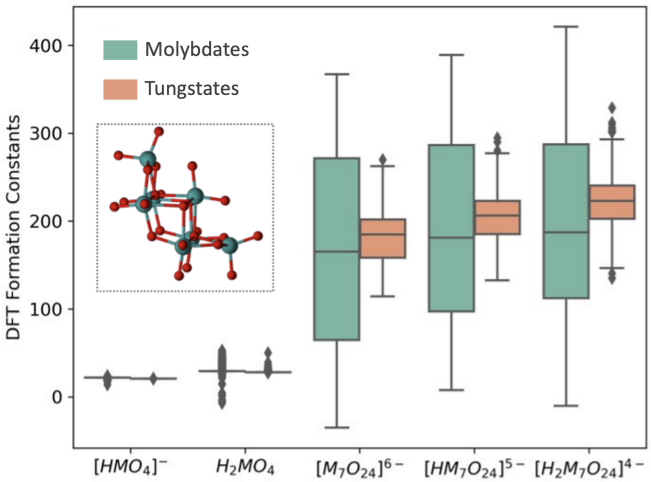


Figure 3.3: Box plots of the DFT formation constants for the monomer and heptamer. Inset plot of the $[M_7O_{24}]^{6-}$ tridimensional structure.

Interestingly, the median values in Figure 3.3 vary if we compare tungsten and molybdenum oxides. *POMSimulator* predicts that the median values for $[H_nW_7O_{24}]^{n-6}$ are greater than those for

$[H_nMo_7O_{24}]^{n-6}$. The larger the formation constants, the more stable the 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.[174] The chemical explanation of this phenomenon is attributed to the larger strength of the tungsten-oxygen bond compared to the molybdenum-oxygen bond.[168] Even if this energy difference is relatively small, the accumulative effect in medium and large clusters gains importance.

The next step is to evaluate the correlation between the theoretical and the experimental constants. Figure 3.4 shows that *POMSimulator* predicts the equilibrium constants of two different metal oxide systems with extraordinary accuracy. The experimental formation constants employed were determined by Cruywagen et al.[154] and Rozantsev et al.[155] at ionic strengths of 1.0 and 0.25 M, respectively. Firstly, Figure 3.4A presents the linear correlation for molybdenum formation constants. The x-axis corresponds to the constants computed by our method, while the y-axis corresponds to the experimental constants reported by Cruywagen.[154] The molybdenum compounds used for the calibration are: $[HMoO_4]^-$, 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 the largest clusters in the set, $[H_{32}Mo_{36}O_{128}]^{8-}$. The eight molybdenum clusters are represented as circles in the plot. 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 with high accuracy.[169]

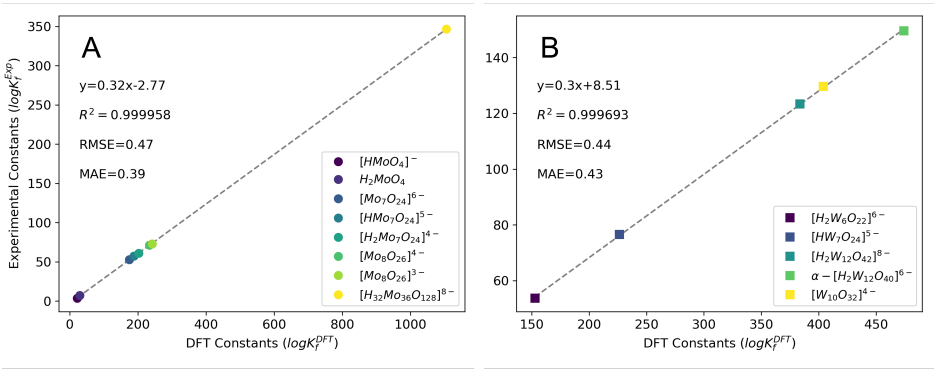


Figure 3.4: Linear regressions between $\log K_f^{Exp}$ and $\log K_f^{DFT}$, with the lowest root mean squared error (RMSE) and mean average (MAE) values at 298.15 K and 1 atm. Left: polyoxomolybdates (circles) at 1.0 M NaCl. Right: polyoxotungstates (squares) at 0.25 M NaCl.

Secondly, Figure 3.4B shows the linear correlation for tungsten formation constants. The set of tungsten oxides used for the calibration are as follows: $[H_2W_6O_{22}]^{6-}$, $[HW_7O_{24}]^{5-}$, $[H_2W_{12}O_{42}]^{8-}$, α-$[H_2W_{12}O_{40}]^{6-}$, and $[W_{10}O_{32}]^{4-}$. Figure 3.4B presents the best linear regression for the tungsten oxide system. Tungsten oxide calibration offers outstanding accuracy with a correlation of R^2=0.9997 and prediction errors of RMSE=0.44 and MAE=0.43. These excellent results imply that *POMSimulator* can predict tungsten formation constants with high accuracy as well. It is worth noting the similarities in the slope parameter of both regressions. We already pointed out in section 2.4.4 the potential universality of the linear scaling.

Once we have ensured that the linear scalings offer very high accuracy, we can proceed with the calculation of the aqueous speciation. Figure 3.5 shows the speciation phase diagrams for molybdenum ox-

ides 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. Under 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, $[HMoO_4]^-$, and molybdic acid, H_2MoO_4.

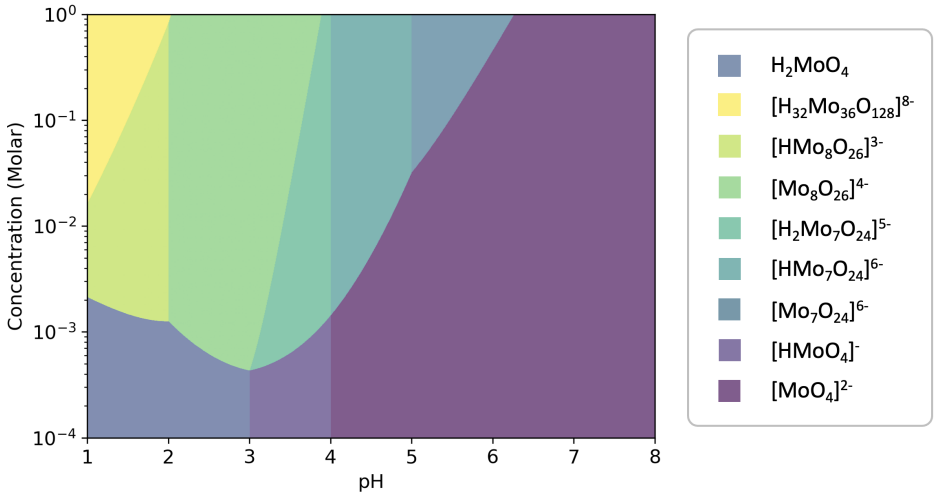


Figure 3.5: Computed speciation phase diagram for polyoxomolybdates at an ionic strength of 0.6 M NaCl, 298.15 K and 1 atm.

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 and 3.5. Under more neutral conditions, between 4 and 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 the number of species and topologies.

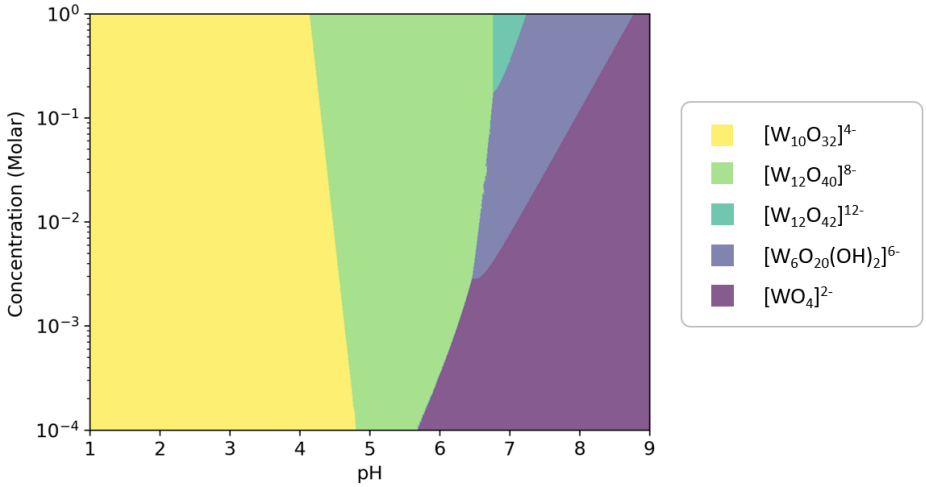


Figure 3.6: Computed speciation phase diagram for polyoxotungstates at an ionic strength of 0.25 M NaCl, 298.15 K and 1 atm.

Likewise, the speciation phase diagram for polyoxotungstates, using the rescaled formation constants, is shown in Figure 3.6B. Tungsten monoxide is the most abundant compound under relatively alkaline conditions (pH > 9). In contrast to molybdenum, the concentrations for tungstates do not have a sharp disturbing effect on the relative abundances. Despite that, in very diluted solutions, the polymerization process does not take place either. Tungstate-Y $\{W_{10}\}$, is the predominant oxide at pH lower than 5. Between pH 5 and 6.5, the most abundant species are dodecatungstates. The most pre-

Molybdates	$logK_f^{Res}$	**Tungstates**	$logK_f^{Res}$
$[Mo_2O_7]^{2-}$	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	$[W_5O_{19}]^{8-}$	22.01
$[Mo_7O_{23}]^{4-}$	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		

Table 3.1: Rescaled formation constants for the molybdate (I=1.0 M) and tungstate (I=0.25 M) systems at 298.15 K and 1 atm.

dominant one is the metatungstate $\{\alpha\text{-}W_{12}\}$, which comprises most of the phase diagram in this range. Nonetheless, the paratungstate B $\{W_{12}\}$, also appears at greater concentrations ($> 0.1M$). Then, at more basic conditions the formation of the paratungstate-A derivate $[H_2W_6O_{22}]^{6-}$ is formed. This is the last oligomeric cluster before disembling to the tungstate tetroxide.

Table 3.1 collects the predicted values for the formation constants for some relevant molybdenum and tungsten isopolyoxometalates. To estimate these constants, the linear regressions depicted in Figure 3.4 are used to calibrate the raw formation constants. Based

on the excellent fitting, we can obtain accurate constants of unreported metal-oxo compounds. This sample includes from small clusters (e.g., $[Mo_2O_7]^{2-}$, $[Mo_4O_{13}]^{2-}$) to pentagonal-like structures (e.g., $[H_6Mo_5O_{22}]^{8-}$, $[H_{14}Mo_9O_{38}]^{8-}$). Furthermore, Table 3.1 shows the 10 most representative samples of polyoxometalates. 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 noting that even with this newly calculated data, all tungsten oxide formation constants are larger than the molybdenum oxide analogues, thus indicating higher stability in accordance with experimental evidences.

3.3 Mechanistic Insights

The self-assembly mechanisms of molybdates and tungstates have been studied from the experimental and computational point of view. For instance, a joint work between mass-spectometry and metadynamic calculations provided insights in the fragmentation of the molybdenum Lindqvist, $[Mo_6O_{19}]^{2-}$.[73] More recently, Cronin and coworkers have investigated the kinetics involved in the formation of the $\{Mo_{36}\}$. They combined empiric results with an stochastic algorithm to predict self-assembly process.[175] It is noteworthy that $\{Mo_{36}\}$ plays a crucial role in the formation of Mo-blue structures since it acts as a molecular template.[83] On the other hand, the formation mechanism of isopolyoxotungstates has been less studied. This is because of the slow kinetics involved in the nucleation process. Nonetheless, in 2009 a collaborative work between experiments and

theory led to the first mechanistic proposal of the Lindqvist formation.[156] Three years later a more detailed reaction profile, including transition states, was published.[59] In this section we will provide more insights on the respective mechanisms with the data obtained from the speciation models.

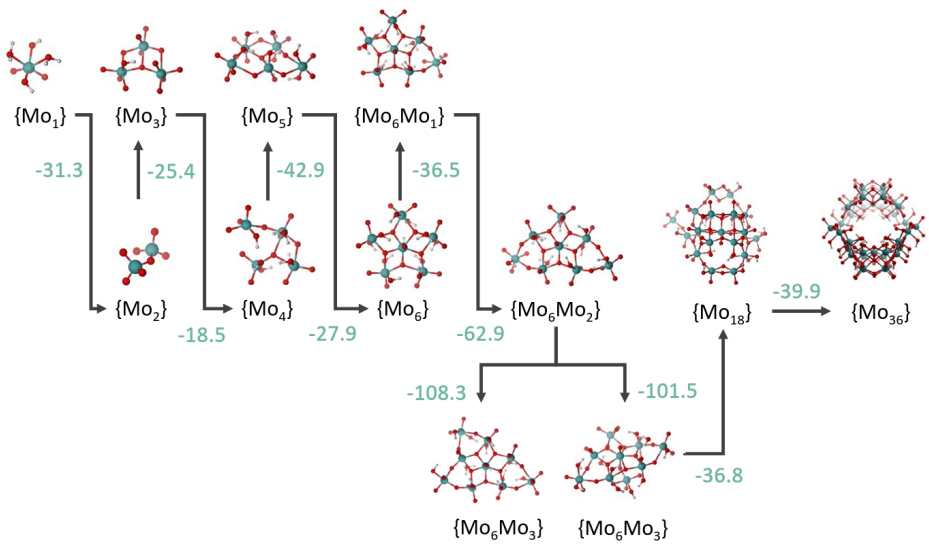


Figure 3.7: Simplified formation mechanism of $\{Mo_{36}\}$ with the reaction energies extracted from the best speciation model.

Figure 3.7 shows the formation mechanism of the largest isopolyoxomolybdate, $\{Mo_{36}\}$, associated to the best speciation model. A first inspection of the mechanism already indicates that the process is spontaneous. All the reaction free energies are negative, even the steps where there is a loss of entropy (e.g., 2 $\{Mo_{18}\} \rightarrow \{Mo_{36}\}$). The mechanism can be divided in three parts: first, the formation of the pentagon, $\{Mo_6\}$, second, the formation of pentagon derivates, $\{Mo_6Mo_x\}$, and third, the formation of the $\{Mo_{36}\}$. All the

chemical reactions and the corresponding reaction energies can be found in Appendix C. The first part of the mechanism starts with the dimerization of two $[HMoO_4]^-$ units with a reaction energy of -31.2 kcal·mol^{-1}. Then, $H_2Mo_2O_7$ gets deprotonated and reacts with the octahedral monomer $[MoO_4(H_2O)_2]^{2-}$ to yield $[H_8Mo_3O_{14}]^{2-}$ (-25.4 kcal·mol^{-1}). Again, the previous cluster is deprotonated and together with $[MoO_4(H_2O)_2]^{2-}$ forms $[H_8Mo_4O_{18}]^{4-}$ spontaneously (-18.5 kcal·mol^{-1}). Next, the tetramolybdate reacts with $[MoO_4(H_2O)_2]^{2-}$ to generate $[H_8Mo_5O_{22}]^{6-}$ in an exergonic manner (-42.9 kcal·mol^{-1}). Then, $[H_8Mo_5O_{22}]^{6-}$ is deprotonated and reacts with $[HMoO_4(H_2O)_2]^-$ to form the pentagonal motif, $[H_{12}Mo_6O_{27}]^{7-}$ $\{Mo_6\}$.

The second part of the formation mechanism is more direct. The pentagonal motif expands up to $\{Mo_6Mo_3\}$, forming structures that have been proposed in the literature.[175,176] The nucleation starts with the reaction of $[H_{11}Mo_6O_{27}]^{7-}$ with $[HMoO_4(H_2O)_2]^-$ to yield $[H_{14}Mo_7O_{31}]^{6-}$ $\{Mo_6Mo_1\}$ (-36.5 kcal·mol^{-1}). Next, $\{Mo_6Mo_1\}$ is deprotonated and together with $[HMoO_4(H_2O)_2]^-$ generate $[H_{15}Mo_8O_{35}]^{7-}$ $\{Mo_6Mo_2\}$ with a reaction energy of -62.9 kcal·mol^{-1}. Once $\{Mo_6Mo_2\}$ is formed, the next step if the generation of the two $\{Mo_6Mo_3\}$ isomers. Herein, we propose these two different structures even though they have not been synthesized yet. The formation of $[H_{14}Mo_9O_{37}]^{6-}$ a-$\{Mo_6Mo_3\}$ happens through the reaction of $[H_{14}Mo_8O_{35}]^{6-}$ and $H_2MoO_4(H_2O)_2$ (-108.3 kcal·mol^{-1}). Then, the formation of the other isomer $[H_{15}Mo_9O_{38}]^{6-}$ b-$\{Mo_6Mo_3\}$ takes place via the addition of $H_2MoO_4(H_2O)_2$ to $[H_{13}Mo_8O_{35}]^{5-}$ (-101.5 kcal·mol^{-1}). Thus, the formation of the pentagonal motifs is

clearly exergonic.

The third and last part of the mechanism involves the generation of the largest molybdenum isopolyoxoanion, $\{Mo_{36}\}$. The two $\{Mo_6Mo_3\}$ units condensate to yield spontaneously the $[H_{16}Mo_{18}O_{65}]^{6-}$ (-36.8 kcal·mol^{-1}). Finally, two units of $[H_{16}Mo_{18}O_{65}]^{6-}$ undergo a condensation reaction to form $[H_{32}Mo_{36}O_{128}]^{8-}$ spontaneously (-39.9 kcal·mol^{-1}). It is worth noting that all the reaction steps are exergonic, and thus the formation of the $\{Mo_{36}\}$ is extremely favourable.

Figure 3.8 depicts the formation mechanism of the most important tungsten isopolyoxoanions, according to the best speciation model. The nucleation process can be divided in two parts: firstly, the generation of $\{W_5\}$, and secondly the formation of the main clusters $\{W_7\}$, $\{W_{12}\}$, α-$\{W_{12}\}$, and $\{W_{10}\}$. The nucleation starts with the spontaneous dimerization (-0.9 kcal·mol^{-1}) of two $[HWO_4]^-$ to yield $[W_2O_7]^{2-}$. Next, the dimer gets diprotonated and reacts with $H_2W_2O_7$ to generate $H_2W_3O_{10}$ with a similar reaction energy than before (-3.7 kcal·mol^{-1}). The self-assembly process proceeds with the exergonic formation of $[W_4O_{13}]^{2-}$ through the reaction of $[HW_3O_{10}]^-$ and $[HWO_4]^-$. Then, the tetramer is protonated and reacts with $[HWO_4]^-$ to yield $[W_5O_{16}]^{2-}$ even though the reaction is slightly endergonic (5.2 kcal·mol^{-1}). The formation of $\{W_5\}$ is paramount since it sets the divergence point of for the formation of the rest of relevant clusters. For example, the protonated Y-tungstate $[H_2W_{10}O_{32}]^{2-}$ forms spontaneously through the dimerization of two $[HW_5O_{16}]^-$ units. Figure 3.8 contains two dashed arrows which lead to two different hexametalate clusters. We have drawn these dashed

arrows because these compounds form from pentametalates which are not included in the plot. Figure 3.8 is a simplified scheme of the overall reaction network showing only the most relevant molecules for the sake of clarity. Nonetheless, in the text we are going to specify the exact reactions, and they can also be found in Appendix C.

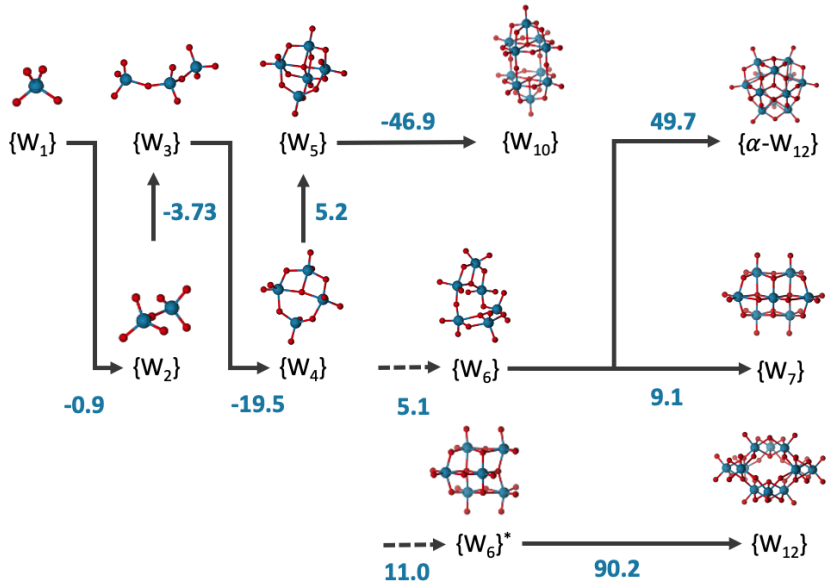


Figure 3.8: Simplified formation mechanism of polyoxotungstates with the reaction energies corresponding to the best speciation model.

The formation of $[H_2W_6O_{20}]^{2-}$, $\{W_6\}$, happens through the reaction of $[H_2W_5O_{17}]^{2-}$ and H_2WO_4 (5.1 kcal·mol^{-1}). This cluster is crucial for the formation of the paratungstate derivates. Then, the hexatungstate gets deprotonated $[HW_6O_{20}]^{3-}$ and reacts with $[HWO_4]^-$ to yield the paratungstate-A $[H_2W_7O_{24}]^{4-}$ (9.1 kcal·mol^{-1}). Furthermore, metatungstate α-$[W_{12}O_{40}]^{8-}$ forms via the dimerization of two $[W_6O_{20}]^{4-}$. The associated reaction energy is clearly ender-

gonic (49.7 kcal·mol^{-1}) but the subsequent protonation reactions are very exergonic (-52.2 and -42.4 kcal·mol^{-1}). Indeed, this evidence is in good agreement with experiments since the isolation of the $[W_{12}O_{40}]^{8-}$ has not been reported. Instead, its protonated forms have been synthesized successfully.[177] Next, there is another hexatungstate intermediate which plays an important role in the formation of the paratungstate-B. Unlike the previous hexatungstate, this new structure has two more oxygen atoms, $[H_2W_6O_{22}]^{6-}$. It also dimerizes to yield the paratungstate-B $[W_{12}O_{42}]^{12-}$ even though the reaction energy is very endergonic as well (90.2 kcal·mol^{-1}). Analogously to the previous tungsten cluster, if we consider the protonation reactions of $[W_{12}O_{42}]^{12-}$ (-72.9 and -62.5 kcal·mol^{-1}) the overall reaction energy becomes clearly exergonic (-45.2 kcal·mol^{-1}).

3.4 Summary

This chapter focused on the application of *POMSimulator* to polyoxomolybdates and -tungstates. In the first part we centered on determining the aqueous speciation of these two systems. To do so, we calculated the formation constants of 72 molybdenum and 50 tungsten oxo-clusters, respectively. Interestingly, our method predicted that tungstate formation constants had greater values than molybdate ones, in good agreement with the experiments. This implied that tungstates are more stable clusters, and thus more reactively inert. The theoretical constants of $\{H_xMo_1\}$, $\{H_xMo_7\}$, $\{H_xMo_8\}$, $\{Mo_{18}\}$ and $\{Mo_{36}\}$ were used for the linear scaling step. The high determinant coefficient ($R^2 = 0.99995$) and low root mean squared

error (RMSE=0.47) indicated the high accuracy of the scaling. Analogously, we applied the same procedure to tungstates where we set up the scaling with the formation constants of $\{H_2W_6\}$, $\{HW_7\}$, $\{H_2W_{12}\}$, α-$\{H_2W_{12}\}$, and $\{W_{10}\}$. Again, the accuracy of the scaling was excellent, with a determinant coefficient $R^2 = 0.99969$ and low root mean squared error RMSE=0.44. Therefore, in both metal-oxo systems the scaling step did provide exceptional results. Next, we employed the rescaled constants to study the behaviour of these clusters in solution. Molybdates have a richer speciation compared to tungstates, as it is observed in Figures 3.5 and 3.6. Regarding the molybdenum speciation, we reproduced the total concentration effect that favors $\{Mo_{36}\}$ at very high concentrations, $\{Mo_8\}$ at medium concentrations, and $\{H_2Mo_1\}$ at low concentrations. This trend is also fulfilled at more alkaline conditions where $\{H_xMo_7\}$ exist at high concentrations and $\{H_xMo_1\}$ (x=0,1) at low concentrations. Concerning the tungstates speciation, we described the behaviour of the most relevant tungsten oxo-clusters. The acid part of the phase diagram was entirely dominated by Y-tungstate $\{W_{10}\}$. This was coherent with the fact that Y-tungstate is regarded as a stable photocatalyst. Then, between pH 4.5 and 6.5 the metatungstate predominated the phase diagram regardless of the total concentration (like Y-tungstate at more acid pH). Then, at pHs greater than 6 the concentration did have an effect on the speciation. At high concentrations ($> 10^{-1}$ M) the formation of the paratungstate-B was favorable. At similar pH and slightly more diluted concentrations the formation of the $\{W_6\}$ was preferred. Finally, at more alkaline and diluted concentrations the speciation is entirely dominated by the monomeric oxide, $\{W_1\}$.

The second part of this chapter was dedicated to the formation mechanism of molybdenum and tungsten isopolyoxoanions. We proposed a reaction pathway for the nucleation of $\{Mo_{36}\}$. The mechanism started with several addition and condensation reactions which led to the formation of the pentagonal motif, $\{Mo_6\}$. All the reaction steps were spontaneous, with energies greater than -18 kcal·mol^{-1} (in absolute values). Then, $\{Mo_6\}$ formed larger pentagonal structures up to the isomers a- and b-$\{Mo_6Mo_3\}$. These units are very important in large molybdenum clusters since they appear in several structures. In this case, $\{Mo_6Mo_x\}$ are present in the $\{Mo_{36}\}$ cluster. Next, the two $\{Mo_6Mo_3\}$ isomers condense to yield $\{Mo_{18}\}$, and the latter dimerizes to form $\{Mo_{36}\}$. All the reaction energies of this mechanism are strongly exergonic, showing the spontaneity of the self-assembly. Furthermore, we proposed a reaction mechanism for the formation the most relevant tungsten oxo-clusters. The nucleation started from the monomeric specie $\{W_1\}$ and grew up to the pentatungstate, $\{W_5\}$. From this point on, there were several reaction routes which led to different clusters. For instance, the straighforward dimerization of $\{W_5\}$ yielded the tungstate-Y, $\{W_{10}\}$ with a clear exergonic energy (-46.9 kcal·mol^{-1}). Besides, $\{W_5\}$ could also generate $\{W_6\}$ which is a relevant intermediate in this mechanism. Its dimerization led to the formation of the metatungstate, α-$\{W_{10}\}$. However, the reaction energy of the addition was very endergonic (49.7 kcal·mol^{-1}). We attributed this high value to the fact that α-$\{W_{12}\}$ did not actually exist but its protonated forms did. Thus, if we accounted the reaction energies of the protonations, the formation became exergonic. Besides, $\{W_6\}$ could also form the paratungstate-

A $\{W_7\}$ via the reaction with a monomeric species. Next, there was another hexatungstate intermediate which had an important role in the nucleation process: $\{W_6\}^*$. The latter compound dimerized to yield the paratungstate-B, $\{W_{12}\}$ even though the reaction energy was 90.2 kcal·mol^{-1}. Analogously to the previous example, the reaction energy was very endergonic because the cluster was not protonated. If we considered the two subsequent protonations of the internal oxygen atoms, the process became spontaneous.

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Chapter 4

Vanadium, Niobium and Tantalum Isopolyoxoanions

> Originality is nothing but judicious imitation.
>
> — François Voltaire

4.1 Introduction

Group V polyoxometalates consist of vanadium, niobium and tantalum molecular oxo-clusters. Despite sharing the same group of the periodic table, they show different speciation and self-assembly processes. On the one hand, a bottom-up approach is also followed to obtain polyoxovanadates. For instance, increasing the pH of a solution of the vanadium precursor yields larger structures, such as the decavanadate ($[V_{10}O_{28}]^{6-}$; $\{V_{10}\}$, Figure 4.1f). On the other hand, niobates and tantalates have a top bottom route because of

the lack of a stable monomeric precursor. This is due to the larger size and lower charge-densities of Nb^{5+} and Ta^{5+}, compared to V^{5+}, Mo^{6+} and W^{6+}.[178] Thus, niobium and tantalum oxides only form the Lindqvist ion ($[M_6O_{19}]^{8-}$; $\{M_6\}$ M=Nb/Ta, Figure 4.1a) in highly alkaline solution.[179]

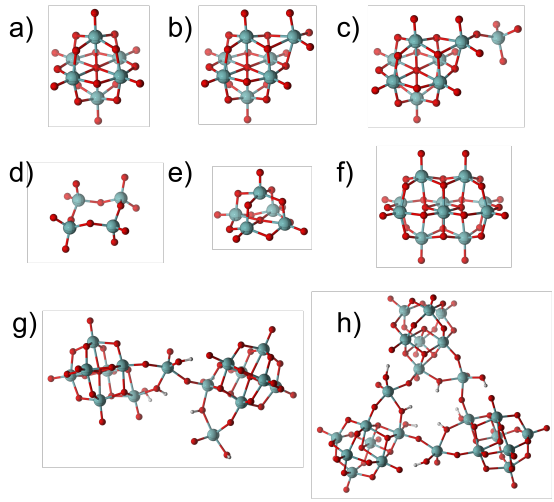


Figure 4.1: Representative structures for vanadium, niobium, and tantalum oxo-clusters. Ball sticks representation of a) Lindqvist $[M_6O_{19}]^{8-}$ ($\{M_6\}$), b) $[M_7O_{22}]^{9-}$ $\{M_7\}$, c) $[M_8O_{25}]^{10-}$ $\{M_8\}$, d) $[M_4O_{12}]^{4-}$ $\{\text{m-}M_4\}$, e) $[M_5O_{14}]^{3-}$ $\{M_5\}$ f) decametalate $[M_{10}O_{28}]^{6-}$ $\{M_{10}\}$, g) $[H_8M_{16}O_{49}]^{10-}$ $\{M_{16}\}$, h) $[H_9M_{24}O_{72}]^{15-}$ $\{M_{24}\}$.

Vanadium aqueous speciation was first described thirty years ago thanks to the direct accessibility to ^{51}V NMR and potentiometric titration studies.[180,181] It is well-established that oligomers with one to five vanadium atoms are formed in solutions of pH $>$ 6. At lower pHs only the decavanadate, $\{V_{10}\}$, and $[VO_2]^+$ have been reported. Polyoxovanadates are the lightest class of molecular oxides available. They display high gravimetric energy densities, which is

a key factor for electron mobility applications.[182] Additionally, vanadium has multiple oxidization states (V^{5+}/V^{4+} and V^{4+}/V^{3+}), which enables rich redox chemistry. These features combined make vanadium oxo-clusters very promising compounds for batteries.[183,184] Decavanadate has been the main prototype for electrochemical studies. To cope with thermal degradation of the metal-oxo cluster, Streb and coworkers prepared a supramolecular crystal that stabilized the decavanadate in a lattice.[47] Moreover, polyoxovanadates are sensitive to the visible light region, which makes them suitable for photoredox catalysis.[185] For example, a molecular manganese vanadium oxide cluster exhibits water oxidation in the natural photosystem-II.[186] Vanadium oxoclusters also play important roles in biological processes.[40,187] There are three remarkable isopolyoxovanadates that have demonstrated antibacterial properties: vanadium oxide, $[VO_4]^{3-}$, tetravanadate ($[V_4O_{12}]^{4-}$; $\{V_4\}$, Figure 4.1d) and decavanadate $\{V_{10}\}$.[188] Nonetheless, the decavanadate is the most prominent cluster due to its high biological activity and antitumor properties.[189,190]

Despite vanadium, niobium and tantalum belonging to the same group in the periodic table, niobates and tantalates have a different behavior in aqueous solution, compared to polyoxovanadates. In fact, niobium and tantalum oxo-clusters only form in neutral and alkaline conditions. In addition, niobium and tantalum molecular oxides display anomalous solubility trends in the presence of alkaline ions. For instance, cesium niobate and tantalate salts are more soluble than their lithium analogs.[191] Niobium oxo-cluster chemistry has experienced great success in the past decade and new compounds be-

yond $\{Nb_6\}$ and $\{Nb_{10}\}$ have been characterized.[192] In fact, just the list of isopolyoxoniobates has increased considerably: ($[Nb_7O_{22}]^{9-}$; $\{Nb_7\}$, Figure 4.1b),[89] ($[H_9Nb_{24}O_{72}]^{15-}$; $\{Nb_{24}\}$, Figure 4.1h),[193] $[HNb_{27}O_{76}]^{16-}$; $\{Nb_{27}\}$),[194] $[Nb_{32}O_{96}]^{32-}$,[195] $[Nb_{52}O_{150}]^{40-}$,[196] and $[Nb_{54}O_{151}]^{32-}$.[197] These discoveries have important implications in expanding the main applications of niobium oxo-clusters in radioactive waste management[198] and catalysis.[199,200] On the other hand, polyoxotantalate chemistry remains underexplored, partly due to slow reaction kinetics and poor solubility in water.[201] This lack of development is exemplified by the low number of reported tantalate species.[85] Only the Lindqvist $\{Ta_6\}$ is observed for isopolyoxotantalates compounds in aqueous solution.[202,203] Furthermore, the decatantalate $\{Ta_{10}\}$ has only been isolated in non-aqueous media.[204] Despite that, controlling polyoxotantalate speciation is a promising asset for synthesizing solid materials. Indeed, tantalum oxides have excellent applications in photocatalysis[205] and technological materials such as transistors.[206]

4.2 Speciation Results

We have defined three molecular sets for polyoxovanadates, -niobates and –tantalates. Our vanadium molecular set is formed by 16 different topologies. If we consider all the protonation states, the final number of species is 40 metal-oxo clusters. It has the classical decavanadate, $\{V_{10}\}$ as well as the Lindqvist $\{V_6\}$. Furthermore, we have included cyclic structures (also known as metavanadates): $[V_3O_9]^{3-}$, $[V_4O_{12}]^{4-}$, $[V_5O_{15}]^{5-}$, $[V_6O_{18}]^{6-}$, $[V_7O_{21}]^{7-}$, $[V_8O_{24}]^{8-}$ and

$[V_9O_{27}]^{9-}$. These cyclic structures will be denoted as {m-V_n}; m refers to "metavanadate" and n to the number of metal atoms. Secondly, niobium and tantalum molecular sets share the same 42 structures, just differing in the metal centers, plus monomeric hydroxides, $[M(OH)_x]^{(5-x)}$, condensed forms of the Lindqvist structures, $\{M_7\}$, $\{M_8\}$, $[M_9O_{27}]^{9-}$ $\{M_9\}$, and medium-large clusters such as $\{M_{16}\}$ and $\{M_{24}\}$. Although some structures are unique for each system, there are common structures in the three sets of polyoxometalates. For example, vanadium, niobium and tantalum share the same decametalates, and Lindqvist structures. In this manner, we ensure that we are not biasing the simulation towards a specific product. In addition, we have included cyclic compounds in the niobium and tantalum sets, despite there is no experimental evidence that they exist in solution. Thereby, we can test whether our method detects the inherently different chemistry of each metal automatically. Additionally, we have included non-reported species such as $\{M_5\}$ and $\{M_{16}\}$ to investigate the existence of possible transient intermediates. Furthermore, we have employed the electrostatic potential (ESP) and literature evidence to build the most stable proton isomers. Regarding the protonation degree, each nuclearity is considered in at least two protonation states. In this manner, we can handle the acid/base properties of the metal-oxo cluster. Finally, a dataset collection[2] of the main results is available in the ioChem-BD repository.[173]

In the first place, we computed the CRN for vanadium, niobium, and tantalum oxides, which are formed by 75, 65 and 65 chemical reactions, respectively. We started by converting the molecules to molecular graphs. Then, we mapped the isomorphism property to

decide which molecules could be potentially related by a chemical reaction. We have defined five types of reactions: acid/base equilibria, hydrolysis, condensation, addition and dimerization. The complete lists of reactions, with their corresponding free energies, are collected in Appendix C. Moreover, Figure 4.2 shows an example of the reaction map involving the 40 polyoxovanadates. Most of the reactions present a reaction energy of approximately -20 kcal·mol^{-1}. Thus, the formation of larger clusters is already favorable from a thermodynamic point of view. It is also worth showing that the monomeric species play a central role in the nucleation process. In fact, the $[H_xVO_4]^{x-3}$ present a high degree of connectivity as the number of edges in Figure 4.2 demonstrates.

Next, we moved to the resolution of the multi-species chemical equilibria (MSCE). We set up $1.1 \cdot 10^6$, $4.7 \cdot 10^5$ and $5.2 \cdot 10^5$ speciation models for vanadium, niobium, and tantalum, respectively. Then, we solved every speciation model at different pH values, thus obtaining new sets of formation constants at each pH. Although formation constants are formally independent of the pH, there are small numeric fluctuations. Therefore, to decrease the margin of error, we increased the sample size by solving the speciation models multiple times. More precisely, we solved $7.5 \cdot 10^7$, $6.9 \cdot 10^7$ and $8.6 \cdot 10^7$ systems of non-linear equations (NLE) and then we estimated the equilibrium constants of all the metal-oxo compounds.

The resulting theoretical formation constants already provide meaningful information in terms of chemical properties. Polyoxoniobates have larger formation constants than tantalates. Therefore, we deduced that niobium oxide clusters are more stable than tantalum

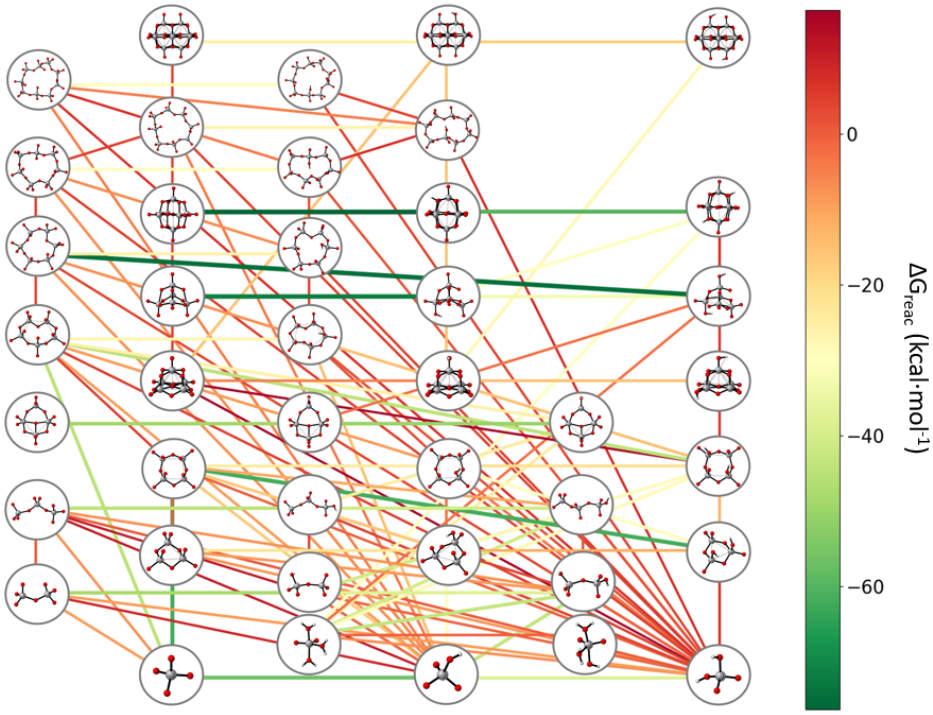


Figure 4.2: Chemical reaction map for the polyoxovanadate system. Edges represent chemical reactions and the color the ΔG_r in kcal·mol^{-1}. Reaction energies should be read from bottom-up and left-right directions.

oxide clusters. This evidence is in good agreement with the enthalpy data reported experimentally.[207] However, DFT formation constants cannot be directly used to predict the aqueous speciation. This is so because theoretical constants are overestimated with respect to the experimental values.[1,2] For molybdenum and tungsten oxides, we have recently found a strong linear dependence between DFT and empirical formation constants (Chapter 3, Figure 3.4), thus granting the application of a systematic linear scaling.[2] Therefore, the first

milestone for predicting Group V aqueous speciation depended on the accuracy of the linear scaling.

We chose as references eleven polyoxovanadates formation constants reported by Elvingson et al.,[158] three polyoxoniobates constants reported by Etxebarria et al.,[157] and three polyoxotantalates constants reported by Deblonde and coworkers.[179] However, the number of experimental data points for niobium and tantalum is clearly insufficient. In fact, we could only find formation constants for the three Lindqvist structures: $\{H_3M_6\}$, $\{H_2M_6\}$, and $\{HM_6\}$ (where M=Nb and Ta). Due to the small size of the sample, we could not rely exclusively on the RMSE and R^2 to decide the best linear scaling. Therefore, we performed a deeper analysis of the regressions to estimate the best linear scaling.

We have just showcased that polyoxoniobates and -tantalates have few formation constants reported in the literature. In both cases, there are only three constants, corresponding to the Lindqvist derivatives: $\{H_xNb_6\}$ and $\{H_xTa_6\}$. Thus, the linear scaling can only be performed with three data points. Due to the reduced and specific size of the sample, the best scaling might fail to correct the constants of different topologies, such as $\{H_xM_7\}$, $\{H_xM_{10}\}$ and $\{M_{24}\}$. Alternatively stated, there is no guarantee that the best linear scaling, according to RMSE and R^2 parameters, leads to the optimal results. We observe this feature in Figure 4.3 where we show the speciation diagrams with the best scaling for niobium and tantalum oxides, respectively. Neither of them reproduces correctly the speciation of the Lindqvists at alkaline pH. Furthermore, the $\{Nb_{10}\}$ and $\{Nb_{24}\}$ do not appear in the niobium diagram either. To overcome this in-

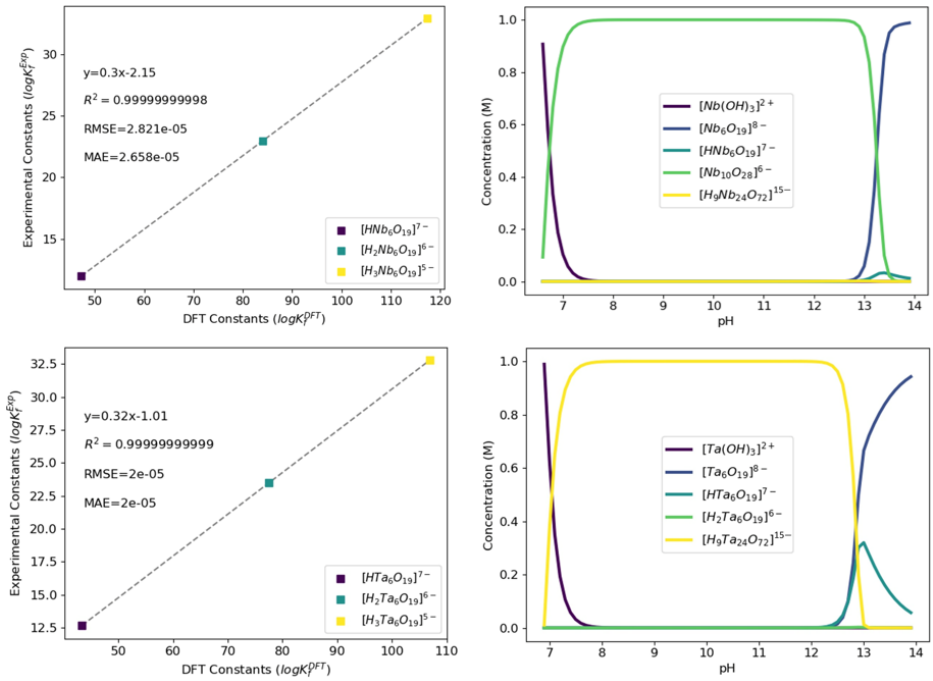


Figure 4.3: Linear scalings with the lowest RMSE and the corresponding speciation diagrams. Top: polyoxoniobates at 0.25 M NaCl. Bottom: polyoxotantalates at 1.0 M NaCl.

convenient in the scaling step we made the following hypothesis: the optimal scaling must have an RMSE value between the minimum showed in Figure 4.3 and 0.5 units. We chose 0.5 as the upper error threshold because of the linear scalings obtained in our previous work (for molybdenum 0.47, and for tungsten 0.44).[2] We screened all the speciation models that fulfilled the RMSE condition. There were 20,763 models for niobium oxides, and 27,650 models for tantalum oxides. Additionally, we developed a function which checked if the speciation of each model agreed with the experiments. More con-

cretely, it analysed if there was significant concentrations of major species such as: $\{H_xNb_{10}\}$, $\{H_xNb_{24}\}$, $\{H_xNb_7\}$, and $\{H_xTa_6\}$. In this manner we could determine which was the best scaling relying not only on the RMSE but also on experimental evidences.

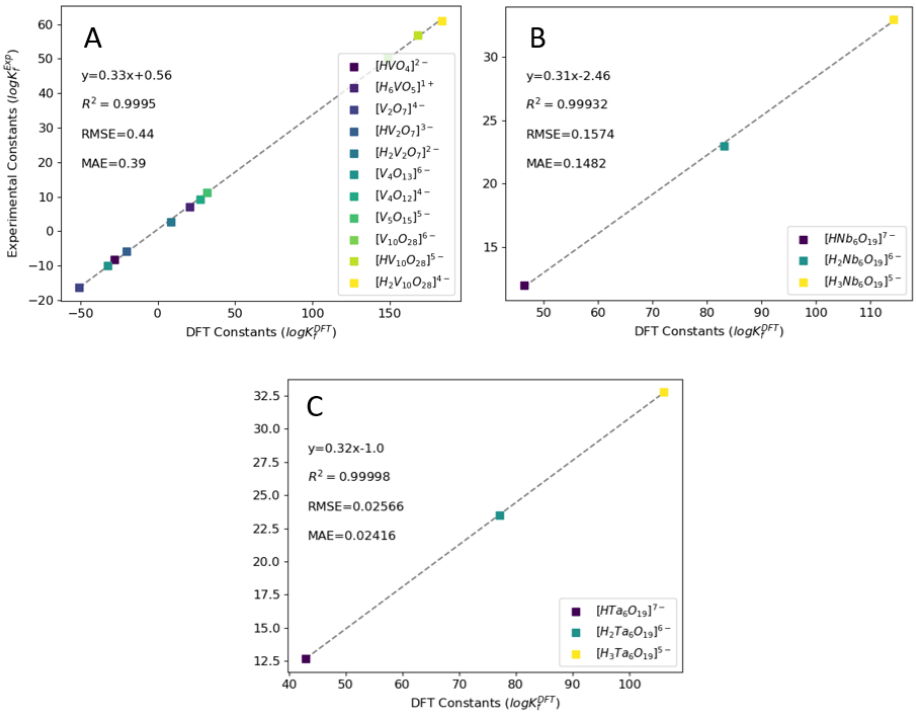


Figure 4.4: Linear regressions between $logK_f^{Exp}$ and $logK_f^{DFT}$. Lowest root mean squared error (RMSE) and mean average (MAE) values at 298.15 K and 1 atm for polyoxovanadates (A), -niobates (B) and -tantalates (C) at 0.25 M, 0.1 M and 1.0 M NaCl ionic strengths, respectively.

Figure 4.4A shows the linear scalings that provide the best results. The most straightforward example corresponds to the linear regression of polyoxovanadates. It presents a robust regression of eleven

points, a determinant coefficient of 0.9995 and a RMSE of 0.44. Notice that that there are different topologies: monomeric compounds, $[VO_4]^{3-}$, $[HVO_4]^{2-}$, $[VO_2{\cdot}3H_2O]^{+}$, cyclic structures {m-V_4}, {m-V_5}, and decavanadates, $\{H_2V_{10}\}$, $\{HV_{10}\}$, $\{V_{10}\}$. Furthermore, Figure 4.4 also includes the linear scalings that best describe the speciation of polyoxoniobates and tantalates. Figure 4.4B depicts the regression for niobium oxo-clusters with a R^2 of 0.9993 and a RMSE of 0.1574. Analogously, Figure 4.4C collects the calibration for tantalum oxo-clusters, with a R^2 of 0.99998 and a RMSE of 0.0256.

Figure 4.5 shows the potential of *POMSimulator* for supplying a complete picture of the POMs' speciation. To do so, we have considered not only the free reaction energy at a given temperature, but also the pH, ionic force and concentration effects. The left part of Figure 4.5 depicts phase speciation diagrams, which pinpoint the most abundant oxo-clusters depending on both the pH and the total metal concentration. These plots are extremely helpful for guiding experiments since they specify the conditions to synthesize a given compound. Then, the right part of Figure 4.5 shows simple speciation diagrams, which collect information of all relevant species at a fixed metal concentration but variable pH. They are useful for detecting minor species and transient intermediates involved in the formation of larger polyoxometalates. Overall, both types of plots have an excellent synergy, since they reveal the intricate effect of pH and concentration on the oxo-clusters formation.

First, we focused on the speciation of polyoxovanadates. It is worth noting that this is the only polyoxometalate family for which an experimental phase speciation diagram exists. It was reported

by Sadoc and coworkers,[159] and is in excellent agreement with the one computed in this work (Figure 4.5A left). One general but remarkable feature of vanadium oxo-clusters is their wide pH range. The left and right parts of Figure 4.5A show that they aggregate within acid and neutral pH. This fact is in stark contrast to niobium and tantalum polyoxometalates, which polymerize at alkaline conditions. At low pH, the most abundant compound is the solvated cation $[VO_2 \cdot 3H_2O]^+$.[159] Figure 4.5A left shows that the presence of $[VO_2 \cdot 3H_2O]^+$ decreases as the metal concentration increases. This is so because of the competitive reaction to form the decavanadate, $\{HV_{10}\}$. The higher the metal concentration, the more favorable the polymerization reactions are. This evidence is supported by the diffusion effect, which is a recurrent theme in polyoxometalates' chemistry.

Figure 4.5A right shows that $\{HV_{10}\}$ is the most abundant species at 0.5 M and pH 1-2. Indeed, the decavanadate compound is a paradigmatic structure. For example, $\{V_{10}\}$ is the largest isopolyoxovanadate in vanadium chemistry. Analogously to Group VI oxo-clusters, the largest polyoxomolybdate, $\{Mo_{36}\}$, and -tungstate, $\{W_{12}\}$, are formed at very acid conditions. This similarity between different group metals is explained by their charge size ratio; lower-charge clusters are generally more stable at lower pH, and vice versa. Figure 4.5A shows that $\{HV_{10}\}$ can also be found in its deprotonated form, $\{V_{10}\}$, at a pH closer to 4. At a pH between 6 and 8 a change in the aqueous speciation occurs. The decavanadate decomposes to form metavanadates. These compounds consist of cyclic vanadium oxo-clusters, such as: {m-V_4} and {m-V_5}. They are characteristic compounds within the broad spectra of species in the

polyoxovanadates family. In fact, such cyclic structures are not found either in niobium or in tantalum oxoclusters. For example, the generic {m-V_4} structure, even though included in polyoxoniobates and -tantalates, is not predicted as a major species.

If we draw our attention to Figure 4.5A left at pH between 6 and 8, we can see a nice example of the concentration effects. By keeping the pH constant and varying the concentration, it is possible to selectively obtain: the pentavanadate (> 0.1M), the tetravanadate ($0.1 - 0.01$ M), the divanadate ($0.01 - 0.001$ M), or the vanadic acid ($<$ 0.001M). Given that the only cluster that has antibacterial properties is the tetravanadate, $[V_4O_{12}]^{4-}$, it is of great interest to know its exact formation conditions. Figure 4.5A right shows that both {m-V_4} and {m-V_5} are abundant species, but at 0.05 M the tetravanadate is the most abundant one. Then, if we decrease the concentration further, the polymerization is discouraged in gradual manner. The vanadium dimer is the most abundant cluster at 0.005 M, and if we go to more diluted solutions, nucleation is interrupted.

Next, the speciation of polyoxovanadates between neutral and alkaline pH is less crowded. At high concentrations and pH ≈ 9, we find another tetravanadate, $[V_4O_{13}]^{6-}$, which is more alkaline than the earlier one, {m-V_4}. This evidence is in good agreement with ^{51}V and ^{17}O NMR results which confirm the formation constant of this compound.[208] In the most alkaline region of the phase diagram, the most abundant compounds are the monomers in two protonation states: $[HVO_4]^{2-}$, and $[VO_4]^{3-}$. The speciation diagram also shows some residual concentration of $[V_2O_7]^{4-}$ ({V_2}). Nonetheless, the polymerization region is found at acid and neutral pH, analogous-

ly to molybdenum and tungsten polyoxometalates. Therefore, our method reproduces the proton-driven nucleation of polyoxovanadates very well.

Secondly, we focused on the speciation of polyoxoniobates. Initially, the aqueous behavior of polyoxoniobates was very much restricted to Lindqvist structures: $\{H_xNb_6\}$. This was mainly due to the poor solubility of niobate salts in water, especially at low pH. Despite the poor solubility and the unusual alkaline speciation, experimental techniques have succeeded in detecting new polyoxoniobates in solution over the past ten years.[189] The chemistry of niobium is vastly different from its neighbors in the periodic table. As we mentioned before, the charge size ratio has a strong influence on the speciation of polyoxometalates. This fact involves a dramatic change in the aqueous speciation of niobium, compared to classical polyoxometalates, such as molybdates, tungstates, and vanadates. Figure 4.5B shows the predicted phase and speciation diagram of the old and new niobium oxo-clusters (at an ionic force of 0.1 M of NaCl). To the best of our knowledge, the phase speciation diagram in Figure 4.5B left is the first that has oxo-clusters that have been discovered over this decade. In other words, the diagrams in Figure 4.5B include compounds that go beyond the Lindqvist limit.

If we analyze Figure 4.5B, we see that our method reproduces the alkaline character of polyoxoniobates' chemistry (pH 5 – 14) automatically. This is a noteworthy feature, since we did not introduce any modification to readjust the program for a completely different pH behavior. This evidence reinforces our belief that the method used here is a general purpose tool for predicting the speciation of

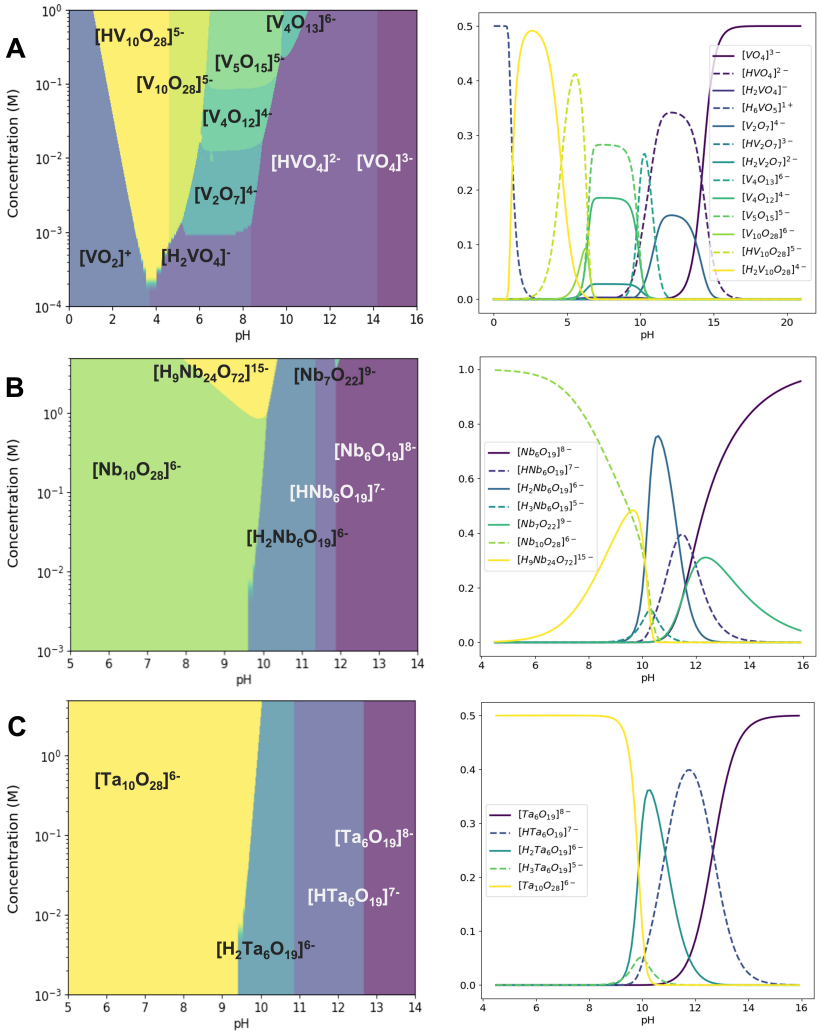


Figure 4.5: Left: speciation phase diagrams employing scaled ΔG^{DFT} values, at 298.15K and 1 atm. Right: speciation diagrams at 0.5 M, 1.0 M and 1.0 M metal concentration for vanadium, niobium and tantalum respectively. Polyoxovanadates (A) at I=0.25 M NaCl, polyoxoniobates (B) at I=0.1 M NaCl and polyoxotantalates (C) at I=1.0 M NaCl (ionic force value fixed according to the experimental data).

POMs. The phase speciation diagram in Figure 4.5B reveals that $\{Nb_{10}\}$ is the most abundant compound at acid, neutral, and slight alkaline pH. Comparing Figures 4.5A and 4.5B, we can also deduce that $\{Nb_{10}\}$ is more alkaline than its vanadium analog $\{V_{10}\}$, in good agreement with the experiments.[209]

Figure 4.5B left also reveals that, at the similar pH but higher concentrations, $\{Nb_{24}\}$ is formed. The chemical relationship between $\{Nb_{10}\}$ and the $\{Nb_{24}\}$ was reported in our earlier work in collaboration with May Nyman's research group.[61] Figure 4.5B left reproduces well the pH range (between 8 and 11) and the concentration (2 M) at which $\{Nb_{24}\}$ was synthesized. Note that the phase diagram only shows the most abundant compounds. Therefore, $\{Nb_{24}\}$ can also be detected at more alkaline conditions, even though it is not the most principal compound. The formation of $\{Nb_{24}\}$ led to the discovery of another relevant niobium cluster: the heptaniobate.[89] $\{Nb_7\}$ has a particularly significant role in the formation of larger clusters such as the $\{Nb_{10}\}$ and $\{Nb_{24}\}$. In fact, our phase diagram also predicts the appearance of the heptaniobate structure, thus supplying the experimental conditions to isolate this compound.

In Figure 4.5B left $\{Nb_7\}$ appears at a more alkaline region, far from the $\{Nb_{10}\}$ and $\{Nb_{24}\}$ areas. At first glance, it is counterintuitive because $\{Nb_7\}$ is regarded as a key intermediate in the formation of large clusters. We attribute this to the thermodynamic stability of $\{Nb_7\}$ in aqueous solution. For instance, the $\{Nb_{10}\}$ and $\{Nb_{24}\}$ are the most abundant and stable compounds at neutral pH. Therefore, within this pH range, $\{Nb_7\}$ acts as a transient intermediate. That is why it does not appear in a phase diagram since it rapidly reacts

to form either the $\{Nb_{10}\}$ or $\{Nb_{24}\}$. On the other hand, if we are interested in obtaining $\{Nb_7\}$ as the main product, then we must move to more alkaline conditions. To reinforce this idea, we followed an earlier work by Casey and coworkers[12] in which they performed ^{17}O NMR experiments starting from a solution of $\{Nb_{10}\}$ at pH 12.5 and they detected heptaniobate derivates ($\{Nb_{7+x}\}$) throughout the decomposition of $\{Nb_{10}\}$. Thus, there is experimental evidence that $\{Nb_7\}$ is detected at high pH values, in good agreement with our prediction in Figure 4.5B.

According to our phase diagram, $\{Nb_7\}$ can be only isolated at remarkably high concentrations ($\approx$ 2M). Nonetheless, Figure 4.5B right shows that $\{Nb_7\}$ is already present at 1 M, even though it is not the most abundant compound. Given that it is in equilibrium with the very stable Lindqvist structures, it is reasonable to assume that $\{Nb_7\}$ will hardly become the most dominant cluster. Additionally, we believe it is consistent to find $\{Nb_7\}$ close to $\{Nb_6\}$, since their respective structures are extremely similar (see Figures 4.1a and 4.1b). Then, the compounds left in the phase diagram correspond to the well-known Lindqvists. They predominate in most of the alkaline region and are found in different protonation states: $\{H_2Nb_6\}$, $\{HNb_6\}$, and $\{Nb_6\}$. Figure 4.4 already shows that Lindqvist's theoretical formation constants are in excellent agreement with the experimental data. Moreover, note that $\{H_3Nb_6\}$ does not appear in Figure 4.5B left, but it does in Figure 4.5B right. Consequently, the triprotonated state of the Lindqvist is regarded as a minority compound, but detectable in solution. The cases of $\{Nb_7\}$ and $\{H_3Nb_6\}$ show the importance of providing speciation diagrams along with

phase diagrams. In this way, we can obtain a general, and yet detailed picture of the aqueous speciation, without disregarding any oxocluster.

Finally, we focused on the speciation of polyoxotantalates. The aqueous behavior of tantalum oxo-clusters is the most limited in terms of pH range and number of species. The main reason lies in the poor solubility of tantalate salts in aqueous solution (even poorer than niobates). In fact, the only species that have been detected experimentally in water are the Lindqvist-derivate compounds, $\{H_xTa_6\}$. At pH lower than 8, the aqueous speciation is hindered due to the formation of the Ta_2O_5 precipitate. Despite the difficulties, $\{Ta_{10}\}$ was synthesized in non-aqueous media almost ten years ago.[85] This evidence confirmed that the decametalate structure was also present in tantalum as it is in vanadium and niobium. Further work was done in this direction by Casey and coworkers. They showed that if one or two tantalum atoms were exchanged by titanium, the solubility in water (at neutral pH) increased.[203] This evidence suggested that the formation of $\{Ta_{10}\}$ is deprived because of solubility instead of thermodynamic reasons. This conclusion was further supported by Hatchett who highlighted that insolubility should not imply instability.[210]

Figure 4.5C left shows the most abundant clusters in a pH range between 5 and 14, at acid, neutral and slightly alkaline pH; our simulation predicts that the dominant compound is $\{Ta_{10}\}$. Although there is no experimental evidence of the formation of this compound, we justify our results based on Hatchett's assumption. Therefore, our program provides for the first time the formation cons-

Compounds	$logK_f^{Res}$ **(M=Nb)**	$logK_f^{Res}$ **(M=Ta)**
$[M(OH)_3]^{2+}$	-12.52	1.45
$[M(OH)_4]^{+}$	0.84	5.30
$M(OH)_5$	4.69	6.74
$[M_6O_{19}]^{8-}$	-2.45	-0.99
$[HM_6O_{19}]^{7-}$	11.87	12.66
$[H_2M_6O_{19}]^{6-}$	23.21	23.52
$[H_3M_6O_{19}]^{5-}$	32.80	32.75
$[M_7O_{22}]^{9-}$	3.95	-0.18
$[M_{10}O_{28}]^{6-}$	79.40	79.09
$[H_9M_{24}O_{72}]^{15-}$	185.29	174.95

Table 4.1: Computed formation constants for polyoxoniobates and tantalates at an ionic strength of 0.1 M and 1.0 M NaCl.

tant ($logK_f$=70.09) of the decatantalate. Notice that Figure 4.5 not only supplies information of the compounds that do appear, but also of the oxo-clusters that are not present. The values collected in Table 4.1 were used for calculating the phase diagrams depicted in Figure 4.5B and 4.5C.

For example, Figure 4.5C left does not show the formation of $\{Ta_{24}\}$, even though it was included in the simulation. Table 4.1 reveals that the formation constant of $\{Nb_{24}\}$ is greater than the one for $\{Ta_{24}\}$. Thus, the nucleation of the latter is less favored, which is consistent with the results depicted in Figure 4.5B and Figure 4.5C. A similar phenomenon occurs for the heptatantalate, $\{Ta_7\}$, which does not appear either in the phase nor in the speciation diagrams. Again, if we compare the formation constants of the niobium and tantalum heptametalates (Table 4.1), we see that the former is greater than the latter. Consequently, none of these compounds will be formed.

Next, the phase diagram at pH values greater than 9 included the tantalate Lindqvist's. In an analogous manner to niobates, Figure 4.5C left shows the presence of $\{H_2Ta_6\}$, $\{HTa_6\}$, and $\{Ta_6\}$. However, $\{H_3Ta_6\}$ is also a minor species, and so, it only appears in the speciation diagram (Figure 4.5C right).

4.3 Mechanistic Insights

The self-assembly mechanisms to form large polyoxovanadates, -niobates, and -tantalates is far from obvious. Although their aqueous speciation is well-known (to a lesser extent for niobium and tantalum), their formation mechanisms are still unclear. Elucidating the nucleation pathways of polyoxometalates is paramount because it would improve the selectivity of their synthesis. However, a purely energy-based evaluation of the self-assembly mechanism can be misleading. In fact, reaction free energies do not account for the effect of the ionic force, pH, and concentration. Nonetheless, our simulator offers the remarkable possibility to consider the latter variables in the mechanistic study because each speciation model must be solved at a fixed value of pH, ionic force, and concentration. This point is crucial, and it implies that reaction energies in Figure 4.6 are not the only factor for determining the most likely path.

For instance, even though the reaction energy of some steps is endergonic, the subsequent larger clusters assemble through energetically favorable processes overall. Additionally, each speciation model is defined by a unique set of chemical reactions. In this study, we are considering the following reaction types: protonation (acid/base),

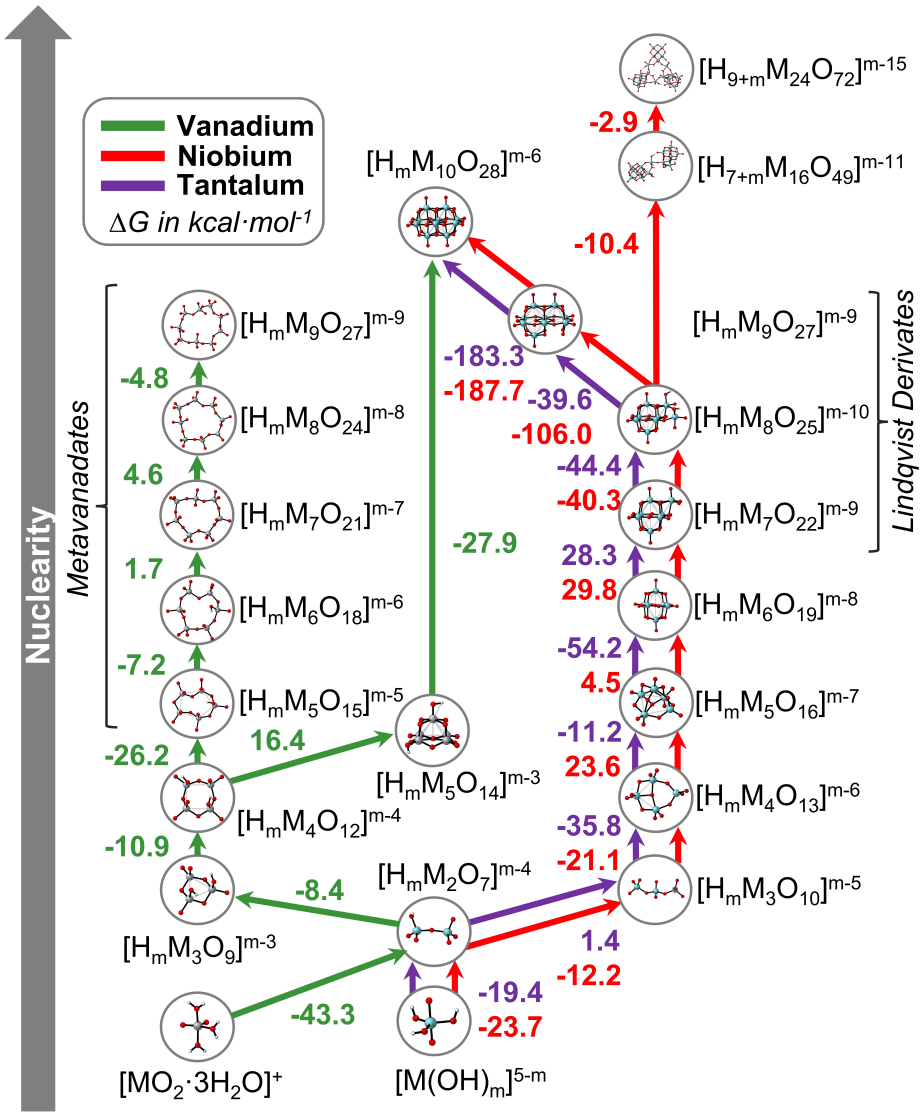


Figure 4.6: Nucleation mechanisms of polyoxovanadates (green), -niobates (red) and -tantalates (purple) according to the best speciation models.

addition, condensation, hydrolysis and dimerization reactions. Then, if a successful speciation model is found, we automatically infer the nucleation mechanism. In the previous section we determined the speciation models that provided the best agreement with the experimental data. Figure 4.5 showed that polyoxovanadates had a different aqueous behavior compared to –niobates and –tantalates. Thus, the corresponding nucleation pathways will be distinct as well.

Figure 4.6 depicts a schematic reaction map of the aggregation mechanisms for these three systems. Notice that the three families of POMs share multiple structures, the decametalate being the most iconic. Despite that, the nucleation mechanisms leading to the formation of $\{M_{10}\}$ can be split into two classes: one class including the decavanadate, and a second class including the decaniobate and decatantalate.

On the left part of Figure 4.6, we see the nucleation pathway of polyoxovanadates. The polymerization starts with the smallest oxo-cluster in solution: vanadium monoxide. It can be found either in its tetrahedral form, $[H_xVO_4]^{x-3}$, or in its triagonal bipyramid form, $[VO_2 \cdot 3H_2O]^+$. It is worth reminding that the process, even though exergonic, only happens at medium-high concentrations. Vanadium monoxide dimerizes spontaneously (-43 kcal·mol^{-1}) to form the dimeric specie $\{V_2\}$. The nucleation proceeds in an exergonic manner up to the formation of the cyclic tetravanadate, $\{\text{m-}V_4O_{12}\}$. Actually, $\{\text{m-}V_4O_{12}\}$ is the first bifurcation point in the reaction map of vanadium since it can either undergo the formation of larger metavanadates (e.g. $\{\text{m-}V_5O_{15}\}$, $\{\text{m-}V_6O_{16}\}$) or generate a newly proposed intermediate: $\{V_5\}$. Both paths lead to the formation of

stable clusters, but it is ultimately the pH and the total concentration that dictate the outcome. For example, at a pH between 6 and 8, the metavanadates are the most abundant species, while at a pH between 2 and 6 the decavanadate is the predominant compound. The formation of metavanadates is exergonic for the tetra- (-10.9 kcal·mol^{-1}), penta- (- 26.2 kcal·mol^{-1}) and hexavanadate (-7.2 kcal·mol^{-1}). However, the nucleation of larger metavanadates is thermodynamically discouraged, in good agreement with the experimental data.[187]

Then, we proposed a reactive intermediate, $\{V_5\}$, as a transient species to form $\{V_{10}\}$. It is a very reactive cluster as its formation energy shows (+16.4 kcal·mol^{-1}). However, the dimerization reaction of $\{V_5\}$ leading to the decavanadate is strongly exergonic (-27.9 kcal·mol^{-1}). If we calculate the energy balance between both reactions, the outcome is negative (-9.9 kcal·mol^{-1}). Consequently, the overall formation of the decavanadate is spontaneous. Figure 4.5 can be also analyzed in combination with Figure 4.6 to further prove the existence of the $\{V_5\}$. For example, Figure 4.5A left shows that the regions including the decavanadates and the metavanadates are remarkably close. Therefore, it is reasonable to assume that there is a transient intermediate, such as $\{V_5\}$, which connects both regions. The satisfactory results obtained in Figure 4.5 support our proposal about the role of $\{V_5\}$ as a key intermediate for yielding the $\{V_{10}\}$.

On the right part of Figure 4.6, we describe the nucleation pathway of polyoxoniobates and -tantalates. A first inspection of the reaction map suggests considerable differences between niobium and tantalum with respect to vanadium. For instance, POMSimulator predicts that the formation of the decametalate structure follows dis-

tinct paths depending on the metal center. From a chemical point of view, this is reasonable due to the divergence in pH behavior. Moreover, this evidence fully agrees with the experimental work published by Aureliano and coworkers.[209]

We will start first by describing the formation of the niobium oxo-clusters. Figure 4.6 shows that the monomeric hydroxides are the smallest oxo-clusters in the niobium reaction network. These hydroxides display different geometries: trigonal pyramid $[Nb(OH)_3]^{2+}$, tetrahedral $[Nb(OH)_4]^{+}$, trigonal bipyramid $Nb(OH)_5$, and octahedral $[Nb(OH)_6]^{-}$. The dominance of hydroxide-species is explained by the alkaline character of these polyoxometalates. High concentrations of hydroxide anions favor the likelihood of hydrolysis reactions. Although monomeric hydroxides do not appear as dominant species in Figure 4.5, this does not imply that they do not play a key role in the mechanism. In fact, they are actively involved in increasing the size of the oxo-clusters. Chemical reactions depicted in Appendix C demonstrate that these small hydroxides appear in almost every condensation and addition reaction.

The mechanism continues spontaneously with a first dimerization forming the $[Nb_2O_7]^{4-}$ ($\{Nb_2\}$) dimer. Then a further condensation reaction takes place yielding the trimer $[Nb_3O_{10}]^{5-}$ ($\{Nb_3\}$). From now on, the polymerization follows the so-called Lindqvist condensation pattern: $\{M_mO_{3m+1}\}$. Thus, $\{Nb_3\}$ condensates spontaneously to $[Nb_4O_{13}]^{6-}$ ($\{Nb_4\}$) with a reaction free energy of -21.1 kcal·mol^{-1}. Next, the formation step for $[Nb_5O_{16}]^{6-}$ ($\{Nb_5\}$) is slightly endergonic (23.6 kcal·mol^{-1}). Analogously, the generation of $\{Nb_6\}$ is also endergonic, but the overall reaction energy is still exergonic. There-

fore, the formation of $\{Nb_6\}$ is thermodynamically favored. This evidence is directly seen in Figure 4.5B where the phase diagrams are dominated by the Lindqvist compounds, at different protonation rates. If we move beyond the Lindqvist structures, we find $\{Nb_{10}\}$ and $\{Nb_{24}\}$.

Our research group performed a first study[61] of the formation of not only the $\{Nb_{10}\}$, but also of the $\{Nb_{24}\}$ with the aim of honing these previous results by using our simulator. Thanks to having direct access to the speciation, we can also account for the effect of main variables such as: pH, total concentration, and ionic force. In our previous work, we reported the decomposition mechanism of $\{Nb_{10}\}$ to $\{Nb_7\}$, promoted by Cs^+ and Li^+ cations. Additionally, we also highlighted that the cation had a non-innocent role in the reaction pathway. In the present study, we have employed Na^+ in our simulation since the ionic force was set to 0.1 M of NaCl. Therefore, we expect a similar, but not identical mechanism. Note that the effect of Na^+ is not considered in the DFT models but it is included in the Davies equation.

Figure 4.6 shows that the $\{Nb_{10}\}$ and $\{Nb_7\}$ are also related by condensation reactions, in agreement with our previous work. However, the Cs^+ mechanism involved an intramolecular reaction that yielded a cationic trimer, $[Nb_3O_6(OH)_3)]^{2+}$, as a byproduct. Because the preference of this mechanism decays according to the alkali size ($Cs^+ > K^+ > Na^+ > Li^+$), we predict that the reaction will undergo a different path. In contrast to the formation of $\{V_{10}\}$, *POMSimulator* does not predict as favorable the dimerization of $[Nb_5O_{14}]^{3-}$ ($\{Nb_5^*\}$) to form $\{Nb_{10}\}$, thus confirming the experimental evidence

that the $\{V_{10}\}$ and $\{Nb_{10}\}$ follow different nucleation mechanisms.[209] In our previous work, we also reported a reaction equation for the formation of the $\{Nb_{24}\}$. Three equivalents of $\{Nb_{10}\}$ were needed to yield one equivalent of $\{Nb_{24}\}$. However, two equivalents of a cationic niobate trimer, $[Nb_3O_6(OH)_3)]^{2+}$, were also formed. In the present study we plan an improvement of our previous suggestion. Instead of three equivalents of $\{Nb_{10}\}$, we propose three equivalents of $\{Nb_8\}$, as depicted in Figure 4.6.

In this manner, we bypass the generation of unfavorable cationic niobates. The first equivalent would form the dimer $\{Nb_{16}\}$, and the second one would yield $\{Nb_{24}\}$. Both condensation reactions are spontaneous (-10.4 and -2.9 kcal·mol^{-1}) showing their thermochemical preference. Additionally, it is reasonable to proffer the existence of the $\{Nb_{16}\}$ dimer following experimental precedents as the formation of the dimer $[Nb_{14}O_{40}(O_2)H_3]^{14-}$ reported by Nyman's group.[89] Finally, this mechanism scheme would fit with the speciation results depicted in Figure 4.5. $\{Nb_{10}\}$ is most stable oxo-cluster, in concordance with its large negative reaction free energy. If concentration effects are considered, the reaction can be shifted to the formation of $\{Nb_{24}\}$ despite being less thermodynamically favored.

Interestingly, the sum of reaction energies from $\{Nb_4\}$ to $\{Nb_7\}$ is 57.9 kcal·mol^{-1}. This result suggests that it is physically impossible to undergo this pathway. However, we must consider the acid-base reactions as well. Notice that the products in (1) and (2) are non-protonated (energies in kcal·mol^{-1}). Speciation models always contain all the acid-base equilibria, thus we must consider these reactions as well. For instance, the protonation reactions to form $\{HNb_5\}$ (3)

and $\{HNb_7\}$ (4) are very favorable:

(1) $[NbO_2(OH)_3]^{2-} + [HNb_4O_{13}]^{5-} \rightleftarrows [Nb_5O_{16}]^{7-} + 2H_2O$ (23.64)

(2) $[NbO_2(OH)_3]^{2-} + [HNb_6O_{19}]^{7-} \rightleftarrows [Nb_7O_{22}]^{9-} + 2H_2O$ (29.84)

(3) $[Nb_5O_{16}]^{7-} + H_5O_2^+ \rightleftarrows [HNb_5O_{16}]^{6-} + 2H_2O$ (-73.01)

(4) $[Nb_7O_{22}]^{9-} + H_5O_2^+ \rightleftarrows [HNb_7O_{22}]^{8-} + 2H_2O$ (-54.24)

In a simplified manner, if we sum (1) + (3) = (5), the reaction energies from $\{Nb_4\}$ to $\{Nb_5\}$ become exergonic with a release of –49.37 kcal·mol^{-1}. Analogously, the formation of $\{Nb_7\}$ becomes exergonic by summing (2) + (4) = (6). In both cases, only single-protonated products are considered. Nonetheless, the inclusion of one protonation reaction inverts the thermodynamics. This stabilization can also be observed in the formation reaction of $[H_mNb_8O_{25}]^{m-10}$ in Figure 4.6. Because the reaction yields a protonated form of $\{H_xNb_8\}$, the reaction energy (7) is clearly exergonic.

(5) $[NbO_2(OH)_3]^{2-} + [HNb_4O_{13}]^{5-} + H_5O_2^+ \rightleftarrows [HNb_5O_{16}]^{6-} + 4H_2O$ (-49.37)

(6) $[NbO_2(OH)_3]^{2-} + [HNb_6O_{19}]^{7-} + H_5O_2^+ \rightleftarrows [HNb_7O_{22}]^{8-} + 4H_2O$ (-24.4)

(7) $[NbO(OH)_3] + [HNb_7O_{22}]^{8-} \rightleftarrows [H_2Nb_8O_{25}]^{8-} + H_2O$ (-40.32)

Figure 4.6 also summarizes the nucleation mechanism of polyoxotantalates. In this case, there is far less experimental data available. Notwithstanding, we suggest a formation pathway that leads

to $\{Ta_6\}$ and the unreported $\{Ta_{10}\}$. Despite the fact that the latter cluster has not been detected in aqueous solution, we predict the mechanism in case the insolubility of $\{Ta_{10}\}$ is ever overcome. Analogously to polyoxoniobates, the smallest tantalum oxo-clusters consist of monomeric hydroxides with the following geometries: trigonal pyramid $[Ta(OH)_3]^{2+}$, tetrahedral $[Ta(OH)_4]^{+}$, trigonal bipyramid $Ta(OH)_5$, and octahedral $[Ta(OH)_6]^{-}$. The monomer dimerizes to yield $[Ta_2O_7]^{4-}$ ($\{Ta_2\}$) spontaneously (-19.4 kcal·mol^{-1}). The formation of the trimer $[Ta_3O_{10}]^{5-}$ ($\{Ta_3\}$) is less favored (1.4 kcal·mol^{-1}) but the next reactions are exergonic. For instance, the generation of the tetramer, $[Ta_4O_{13}]^{6-}$ ($\{Ta_4\}$), and pentamer, $[Ta_5O_{16}]^{7-}$ ($\{Ta_5\}$) show large negative reaction free energies: -35.8 and -11.2 kcal·mol^{-1}. Next, $\{Ta_5\}$ condensates to form the main cluster in polyoxotantalates' chemistry: the Lindqvist $\{Ta_6\}$. Figure 4.5 shows that $\{H_xTa_6\}$ dominates most of the phase diagram. If we acidify the media, $\{Ta_6\}$ condensates to $\{Ta_7\}$ even though it does not appear as the most abundant compound at any range of pH. This reveals the important effect of the pH and concentration.

Although the reaction free energy for $\{Ta_7\}$ and $\{Nb_7\}$ is remarkably similar, only the niobium analogue becomes abundant. That is why a strict energetic picture of the nucleation can lead to wrong conclusions. In contrast, equilibrium constants, such as the ones reported in Table 4.1, do capture other important variables such as ionic force and concentration effects. To generate the decatantalate, $\{Ta_7\}$ undergoes three consecutive condensations: first to yield $\{Ta_8\}$ (-44.4 kcal·mol^{-1}), then $\{Ta_9\}$ (-106.0 kcal·mol^{-1}), and finally $\{Ta_{10}\}$ (-187.7 kcal·mol^{-1}). The full process is remarkably spontan-

eous, with a large negative reaction free energy. However, there are no experimental studies that report the formation of $\{Ta_{24}\}$, either in aqueous or non-aqueous media. This evidence is also in good agreement with the prediction of our simulator. Although tantalum and niobium share the same type of molecular set, our method automatically assesses that $\{Nb_{24}\}$ is more stable than $\{Ta_{24}\}$.

It is worth mentioning that the synthesis of both polyoxoniobates and -tantalates employs the Lindqvist structure as starting material. Instead, Figure 4.6 shows a bottom-up nucleation pathway. It begins with the monomeric hydroxides and goes up to the decametalates, and even further for -niobates. Thus, it seems as if there is a discrepancy between experiments and theory. Notwithstanding, Figure 4.6 should not be understood as a linear pathway to one product, but as an ensemble of chemical equilibria. If the synthesis starts with the Lindqvist, Figure 4.6 should be read starting on the hexametalate structure. Depending on the reaction conditions, the formation of some clusters will be favored above others. Thus, Figure 4.6 provides a reaction map where the most relevant chemical transformations are depicted. In this manner, our method offers the possibility of proposing the underlying self-assembly reactions which are hard to detect experimentally.

4.4 Summary

This chapter focused on the application of *POMSimulator* to polyoxovanadates, -niobates, and -tantalates. In the first part we centered on determining the aqueous speciation of these three systems. We

calculated the formation constants of 40 vanadium, 42 niobium, and 42 tantalum oxo-clusters, respectively. We ensured that these three systems shared common structures (such as the decametalate, $\{M_{10}\}$) to study their different chemical patterns. Then, we compared the theoretical constants to the experimental ones in order to perform the linear scaling. For vanadates we performed a regression with eleven data points (the largest in this thesis) thanks to the numerous experimental data. Even so, the regression parameters were excellent, with a determinant coefficient of 0.9995 and a RMSE of 0.44 units. In contrast, niobium and tantalum polyoxometalates lacked of an extensive database of formation constants. In fact, the most studied constants corresponded to the Lindqvist anion in its three protonation states, $\{H_xM_6\}$ (M=V,Nb,Ta and x=0,1,2,3). This posed an important problem in the linear scaling step. Performing regressions with only three data points was far from ideal. In fact, we tested the speciation with the best scaling and the result was unsatisfactory. Consequently, we explored all the regressions with a minimum RMSE of 0.5. We chose this threshold based on the previous values of RMSE obtained for the tungsten and molybdenum systems. According to this criterion, there were 20,763 models for niobium, and 27,650 models for tantalum. Next, we developed a function that checked the speciation for every model and determined the one that fit the best with experiments. Once the best linear scalings were settled, we proceeded with the study of the aqueous speciation. The vanadium system presented the richest speciation, with the presence of up to ten species in the phase diagram. Despite the complexity, we succeeded in reproducing the previous reported phase diagram

of vanadium. The concentration effects showed a relevant effect on the speciation, specially in the formation of the metavanadates structures. Regarding niobium, we proposed a phase diagram including relatively new niobate species, such as $\{Nb_7\}$ and $\{H_xM_6\}$. Unlike vanadium, the speciation of niobium (and tantalum) is centered at neutral and basic pH. Niobium phase diagram showed that the formation of $\{Nb_{24}\}$ only occurred at high concentration and medium-high pH. Moreover, $\{Nb_7\}$ also formed at high concentrations but more alkaline conditions; close to the Lindqvist structures. This evidence was interesting since $\{Nb_7\}$ formed at more acid conditions as well, but only as a transient intermediate in the formation of $\{Nb_{24}\}$. Concerning tantalum, the speciation is much more limited. We found that the most abundant species are the Lindqvists at different protonation states $\{H_xTa_6\}$ (x=0,1,2), in good agreement with the experiments. Nonetheless, we determined the formation constants and speciation of the decatantalate. This compound has only been synthesized in non-aqueous conditions due to its poor solubility in water. Nonetheless, insolubility does not imply instability, and indeed our method indicated that $\{Ta_{10}\}$ is an stable cluster. Therefore, we believe that if the insolubility factor could be overcome, the formation of $\{Ta_{10}\}$ could become possible.

In the second part of this chapter we focused on the self-assembly mechanism of vanadium, niobium, and tantalum polyoxometalates. More concretely, we compared the reaction mechanisms of these three systems since they share a paramount structure: the decametalate. We proposed an intermediate, $[V_5O_{14}]^{3-}$ ($\{V_5\}$), which dimerizes to yield $\{V_{10}\}$. Additionally, the same intermediate $\{V_5\}$ results from

the condensation of a monomer to {m-V_4}. Therefore, {V_5} is connected to the decametalate part of the phase diagram (pH 1-5), and to the metavanadate part (pH 5-8.5). Consequently, {V_5} can be regarded as key transient intermediate which forms in most part of the acid speciation. On the other hand, the formation of {Nb_{10}} and {Ta_{10}} did not occur through the dimerization of ({M_5}) (M=Nb, Ta) compound. Indeed, the self-assembly happened through the monomeric growth of the Lindqvist species. Interestingly, Figure 4.6 suggested that the formation of {Nb_7} had a reaction energy of 22.0 kcal·mol^{-1}. However, acid-base reactions had to be considered as well. Because POMs are anionic species, the protonation led to a substantial stabilization of the clusters. Furthermore, niobium speciation also includes the formation of {Nb_{24}}. In a previous work, a description of the condensation reaction of {Nb_{24}} was made. With our method we have refined the chemical reaction taking into account the pH, ionic strength, and concentration. We determined that three equivalents of {Nb_8} were needed to yield {Nb_{24}}. Concerning tantalum, the self-assembly pattern is similar to niobium since the assembly of {Ta_{10}} happens in the same manner. However, there is one relevant exception connected to formation of {Ta_{24}}. Indeed, if we compare the constants of {Nb_{24}} and {Ta_{24}} there is a gap of 10.3 units. Given that constants are in logarithmic units, the difference explains very well why {Ta_{24}} is less stable than {Nb_{24}}.

Chapter 5

Conclusions

> Nothing in life is to be feared. It is only to be understood. Now is the time to understand more, so that we may fear less.
>
> — Marie Curie

In this thesis we have developed and applied a new computational method for investigating the multi-species chemical equilibria processes of polyoxometalates. We have named it *POMSimulator* because it simulates the formation of POMs. This new method was initially tested with the octamolybdate system. Once we succeeded in optimizing the method, we applied it to five chemical systems: isopolyoxomolybdates, -tungstates, -vanadates, -niobates, and -tantalates. Based on the results obtained, we concluded that this method could become a useful tool for exploring the chemistry of polyoxometalates. Specially, because it takes into consideration experimental parameters such as pH, ionic force, precursor concentration, temperature and pressure. Besides, we can calculate equilibrium

constants which can guide experiments when determining unreported compounds. Last but not least, the universal scaling observed in the linear relationships might become a important asset in this methodology. If the slope parameter proved constant (≈ 0.3), we would not require experimental data for scaling the theoretical constants thus expanding the scope. Below we summarize the specific conclusions that we have drawn for each chapter.

- Conclusion 1: to study the multi-species chemical equilibria involved in formation of octamolybdate, we have developed a numerical method that we refer to as *POMSimulator*.

- Conclusion 2: from the many challenges encountered in the formulation and solution of the mathematical equations of the speciation models, we have concluded that the most critical issue is the linear scaling of the formation constants. Also, we have found very good linear correlations between the experimental and computed values which have enabled a perfect description of the pH scale.

- Conclusion 3: the speciation diagrams computed in this thesis are in very good agreement with experimental data available and enable the study of the stability and reactivity as function of the pH.

- Conclusion 4: the phase speciation diagrams (% vs pH) computed for the first time in this thesis allow the consideration of the total concentration and pH simultaneously thus giving a complete vision of the aqueous speciation.

- Conclusion 5: we have successfully applied our method to isopolyoxomolybdates and -tungstates, and we have revealed the different chemical nature of both systems. In particular, we have demonstrated that molybdates form large species $\{Mo_{36}\}$ whereas tungstates generate medium-sized compounds $\{W_{12}\}$.

- Conclusion 6: once the most accurate speciation model has been determined, it is possible to predict equilibrium constants for a large number of species which are not available in experiments.

- Conclusion 7: we have confirmed our initial hypothesis that the best speciation model provided the reaction mechanism for the formation of the metal-oxo clusters. Indeed, we have suggested tinny differences between molybdenum and tungsten that explain the different shapes of these compounds.

- Conclusion 8: our method has perfectly described the different chemistry of vanadium, niobium and tantalum isopolyoxometalates. The computed phase diagrams reveal that while vanadium clusters nucleate at acid-neutral pH, niobium and tantalum aggregate at neutral-alkaline pH.

- Conclusion 9: the speciation diagram of vanadium explains the diversity of small and medium clusters (metavanadates) as well as the largest clusters such as the decavanadate. However, niobium and tantalum do not show these same trends and even grow much larger aggregates $\{Nb_{24}\}$.

- Conclusion 10: the analysis of the reactions included in the speciation model highlights the distinct mechanisms to form

the decametalates depending on the metal framework.

- Conclusion 11: the comparison of the linear scaling equations for Mo, W, V, Nb, Ta as well as other oxoanions suggest the existence of an universal procedure to correct the DFT computed constants. Currently, this issue is left as an open question and further work will be performed to fully verify it.

Appendix A

Computational Details

Molecular geometries for all species were fully optimized using a DFT method by employing ADF package (SCM ADF version 2019.1).[211] We have used several functionals: PBE,[212,213] PBE0,[214] BLYP,[215,216] B3LYP-D3,[215,217,218] and BP86.[212,213,216] Unless otherwise stated, we have used PBE functional as reference. Relativistic corrections were taken into account by the scalar-relativistic zero-order regular approximation (ZORA)[219,220] at the TZP basis set level. Solvent effects were introduced using the continuous solvent model COSMO with Klamt radii for water.[221] Stationary points were characterized with analytic frequency calculations. Ground state free energies were computed at 298 K and 1 atm. No standard state corrections have been considered so as not to introduce distorting effects which could affect the linear scaling of the formation constants. Dataset collections of the computational results are available in the ioChem-BD repository[173] and can be accessed via references [222], [172], and [3]. A prototype version of *POMSimulator* is available in reference [223].

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Appendix B

Accuracy Benchmark

In chapter 2 section 2.3.6 we introduced the accuracy threshold implemented in *POMSimulator*. We computed the RMSE between the solution using Powell's and Least Squares' algorithms, and we normalized it with the total concentration of the system ($M_0/RMSE$). This performance parameter has a crucial impact on the outcome of the method, and thus we have performed a benchmark to determine the optimum value. In this section we have discussed the chosen threshold and the reasons that led us to this choice.

To measure the accuracy of the method we have checked the best linear scaling out of all the speciation models (see the linear scaling details in section 2.4). More concretely, the constants with the lowest RMSE (respect to the experiments) mark the final accuracy of the simulation. Therefore, we have performed ten simulations with threshold errors ranging from 1.0 to 0.1 units, and then we have determined the best scaling for each case. Figure B.1 left depicts a bar plot showing the relationship between the performance of the method

and the threshold value. The higher the threshold, the poorer the model. For instance, if no threshold is imposed (1.0 case) the RMSE of the scaling is equal to 0.88 units. In fact, this error is approximately constant for lower thresholds such as 0.9, 0.8, 0.7, 0.6 and 0.5. However, for the 0.4 threshold we observe an increase of accuracy (RMSE=0.81). This trend follows for lower thresholds, up to the minimum value where the accuracy is 0.54 (as in Figure 2.25).

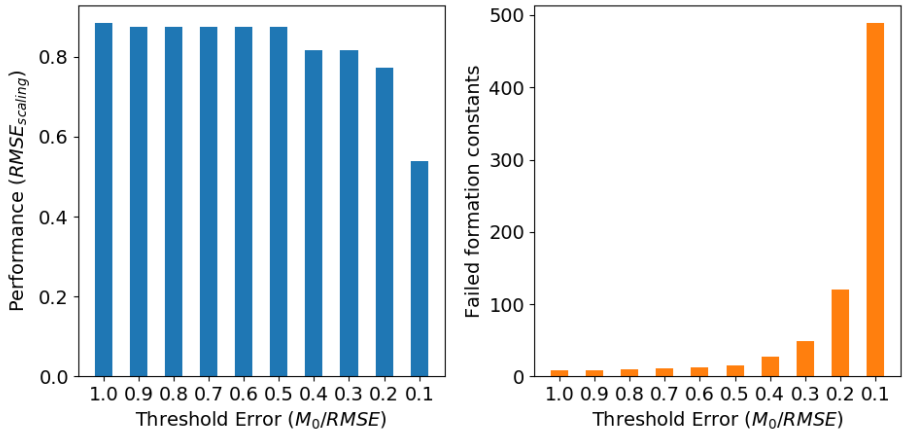


Figure B.1: Left: bar plot indicating the RMSE of the best scaling respect to the threshold error. Right: bar plot indicating the number of failed constants respect to the treshold error.

It seems that the lower the threshold, the greater the performance of the method. However, there is a relevant drawback related with small thresholds. This inconvenient is observed in Figure B.1 right. When no threshold is imposed there are no failed formation constants.[a] By failed we refer to constants that have not been com-

[a] In the output file this is expressed as *None* which is the Python key to define a null value, or no value at all.

puted at any range of pH. Figure B.1 right clearly shows that more strict thresholds lead to more failed formation constants. For lowest threshold there are 489 failed constants. It may not seem a large problem given that the method had computed 1,024 models with 30 constants each, thus a total of 30,720 constants. Consequently, the most accurate case only involves a marginal error of 1.6%. Notwithstanding, this benchmark has been performed on a model system (octamolybdate) where complexity was severely reduced. More complex systems (such as HPA) might require a less strict threshold in order to find a good balance between accuracy and number of computed constants.

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Appendix C

Chemical Reactions

In this section we collect all the chemical reactions that we found for the five polyoxometalate systems. Reaction energies are expressed in kcal·mol^{-1}.

Table C.1: Chemical reactions for the polyoxomolybdate system. Energies in kcal·mol^{-1}.

Reactions	Energies
$[MoO_4]^{2-} + H_5O_2^+ + \rightleftarrows [HMoO_4]^- + 2H_2O$	-29.4
$[HMoO_4]^- + H_5O_2^+ + \rightleftarrows H_2MoO_4 + 2H_2O$	-9.79
$H_6MoO_6 + H_5O_2^+ \rightleftarrows [H_7MoO_6]^+ + 2H_2O$	2.5
$[H_7MoO_6]^+ + H_5O_2^+ \rightleftarrows [H_8MoO_6]^{2+} + 2H_2O$	20.2
$[Mo_2O_7]^{2-} + H_5O_2^+ \rightleftarrows [HMo_2O_7]^- + 2H_2O$	-12.95
$[HMo_2O_7]^- + H_5O_2^+ \rightleftarrows H_2Mo_2O_7 + 2H_2O$	-5.81
$[H_6Mo_2O_{10}]^{2-} + H_5O_2^+ \rightleftarrows [H_7Mo_2O_{10}]^- + 2H_2O$	-17.29
$[H_7Mo_2O_{10}]^- + H_5O_2^+ \rightleftarrows H_8Mo_2O_{10} + 2H_2O$	-8.81
$[Mo_3O_{10}]^{2-} + H_5O_2^+ \rightleftarrows [HMo_3O_{10}]^- + 2H_2O$	-8.17
$[HMo_3O_{10}]^- + H_5O_2^+ \rightleftarrows H_2Mo_3O_{10} + 2H_2O$	-5.42

Reactions	Energies
$[H_6Mo_3O_{14}]^{4-} + H_5O_2^+ \rightleftarrows [H_7Mo_3O_{14}]^{3-} + 2H_2O$	-46.48
$[H_7Mo_3O_{14}]^{3-} + H_5O_2^+ \rightleftarrows [H_8Mo_3O_{14}]^{2-} + 2H_2O$	-10.03
$[Mo_4O_{13}]^{2-} + H_5O_2^+ \rightleftarrows [HMo_4O_{13}]^{-} + 2H_2O$	0.51
$[HMo_4O_{13}]^{-} + H_5O_2^+ \rightleftarrows H_2Mo_4O_{13} + 2H_2O$	4.45
$[H_6Mo_4O_{18}]^{6-} + H_5O_2^+ \rightleftarrows [H_7Mo_4O_{18}]^{5-} + 2H_2O$	-43.99
$[H_7Mo_4O_{18}]^{5-} + H_5O_2^+ \rightleftarrows [H_8Mo_4O_{18}]^{4-} + 2H_2O$	-23.96
$[Mo_5O_{16}]^{2-} + H_5O_2^+ \rightleftarrows [HMo_5O_{16}]^{-} + 2H_2O$	-13.55
$[HMo_5O_{16}]^{-} + H_5O_2^+ \rightleftarrows H_2Mo_5O_{16} + 2H_2O$	25.82
$[Mo_5O_{17}]^{4-} + H_5O_2^+ \rightleftarrows [HMo_5O_{17}]^{3-} + 2H_2O$	-27.96
$[HMo_5O_{17}]^{3-} + H_5O_2^+ \rightleftarrows [H_2Mo_5O_{17}]^{2-} + 2H_2O$	-9.23
$[H_6Mo_5O_{22}]^{6-} + H_5O_2^+ \rightleftarrows [H_7Mo_5O_{22}]^{5-} + 2H_2O$	-55.85
$[H_7Mo_5O_{22}]^{5-} + H_5O_2^+ \rightleftarrows [H_8Mo_5O_{22}]^{4-} + 2H_2O$	-36.04
$[Mo_6O_{20}]^{4-} + H_5O_2^+ \rightleftarrows [HMo_6O_{20}]^{3-} + 2H_2O$	-13.82
$[HMo_6O_{20}]^{3-} + H_5O_2^+ \rightleftarrows [H_2Mo_6O_{20}]^{2-} + 2H_2O$	-9.44
$[H_{11}Mo_6O_{27}]^{7-} + H_5O_2^+ \rightleftarrows [H_{12}Mo_6O_{27}]^{6-} + 2H_2O$	-34.24
$[H_{12}Mo_6O_{27}]^{6-} + H_5O_2^+ \rightleftarrows [H_{13}Mo_6O_{27}]^{5-} + 2H_2O$	-51.3
$[H_{13}Mo_6O_{27}]^{5-} + H_5O_2^+ \rightleftarrows [H_{14}Mo_6O_{27}]^{4-} + 2H_2O$	62.67
$[Mo_7O_{23}]^{4-} + H_5O_2^+ \rightleftarrows [HMo_7O_{23}]^{3-} + 2H_2O$	-15.16
$[HMo_7O_{23}]^{3-} + H_5O_2^+ \rightleftarrows [H_2Mo_7O_{23}]^{2-} + 2H_2O$	4.54
$[Mo_7O_{24}]^{6-} + H_5O_2^+ \rightleftarrows [HMo_7O_{24}]^{5-} + 2H_2O$	-28.25
$[HMo_7O_{24}]^{5-} + H_5O_2^+ \rightleftarrows [H_2Mo_7O_{24}]^{4-} + 2H_2O$	-17.82
$[H_{12}Mo_7O_{31}]^{8-} + H_5O_2^+ \rightleftarrows [H_{13}Mo_7O_{31}]^{7-} + 2H_2O$	-41.21
$[H_{13}Mo_7O_{31}]^{7-} + H_5O_2^+ \rightleftarrows [H_{12}Mo_7O_{31}]^{6-} + 2H_2O$	-26.63
$[Mo_8O_{26}]^{4-} + H_5O_2^+ \rightleftarrows [HMo_8O_{26}]^{3-} + 2H_2O$	-9.31
$[HMo_8O_{26}]^{3-} + H_5O_2^+ \rightleftarrows [H_2Mo_8O_{26}]^{2-} + 2H_2O$	0.1
$[H_{13}Mo_8O_{35}]^{9-} + H_5O_2^+ \rightleftarrows [H_{14}Mo_8O_{35}]^{8-} + 2H_2O$	-24.04
$[H_{14}Mo_8O_{35}]^{8-} + H_5O_2^+ + \rightleftarrows [H_{15}Mo_8O_{35}]^{7-} + 2H_2O$	-37.45
$[H_{14}Mo_9O_{37}]^{6-} + H_5O_2^+ \rightleftarrows [H_{15}Mo_9O_{37}]^{5-} + 2H_2O$	-26.65

Reactions	Energies
$[H_{15}Mo_9O_{37}]^{5-} + H_5O_2^+ \rightleftarrows [H_{16}Mo_9O_{37}]^{4-} + 2H_2O$	-23.15
$[H_{16}Mo_9O_{37}]^{4-} + H_5O_2^+ \rightleftarrows [H_{17}Mo_9O_{37}]^{3-} + 2H_2O$	-4.17
$[H_{17}Mo_9O_{37}]^{3-} + H_5O_2^+ \rightleftarrows [H_{18}Mo_9O_{37}]^{2-} + 2H_2O$	-3.38
$[H_{18}Mo_9O_{37}]^{2-} + H_5O_2^+ \rightleftarrows [H_{19}Mo_9O_{37}]^{-} + 2H_2O$	3.1
$[H_{14}Mo_9O_{38}]^{10-} + H_5O_2^+ \rightleftarrows [H_{15}Mo_9O_{38}]^{9-} + 2H_2O$	-30.37
$[H_{15}Mo_9O_{38}]^{9-} + H_5O_2^+ \rightleftarrows [H_{16}Mo_9O_{38}]^{8-} + 2H_2O$	-33.47
$[H_{16}Mo_9O_{38}]^{8-} + H_5O_2^+ \rightleftarrows [H_{17}Mo_9O_{38}]^{7-} + 2H_2O$	-10.96
$[H_{17}Mo_9O_{38}]^{7-} + H_5O_2^+ \rightleftarrows [H_{18}Mo_9O_{38}]^{6-} + 2H_2O$	-15.67
$[H_{18}Mo_9O_{38}]^{6-} + H_5O_2^+ \rightleftarrows [H_{19}Mo_9O_{38}]^{5-} + 2H_2O$	-9.27
$[H_{16}Mo_{18}O_{65}]^{6-} + H_5O_2^+ \rightleftarrows [H_{17}Mo_{18}O_{65}]^{5-} + 2H_2O$	-33.45
$[H_{17}Mo_{18}O_{65}]^{5-} + H_5O_2^+ \rightleftarrows [H_{18}Mo_{18}O_{65}]^{4-} + 2H_2O$	-11.96
$[H_{32}Mo_{36}O_{128}]^{8-} + H_5O_2^+ \rightleftarrows [H_{33}Mo_{36}O_{128}]^{7-} + 2H_2O$	19.07
$[H_{33}Mo_{36}O_{128}]^{7-} + H_5O_2^+ \rightleftarrows [H_{34}Mo_{36}O_{128}]^{6-} + 2H_2O$	-42.12
$H_2MoO_4 + H_2O \rightleftarrows H_6MoO_6$	18.07
$[H_7MoO_6]^+ + [H_7Mo_5O_{22}]^{5-} \rightleftarrows [H_{12}Mo_6O_{27}]^{6-} + H_2O$	-27.87
$[H_8MoO_6]^{2+} + [H_8Mo_5O_{22}]^{4-} \rightleftarrows [H_{14}Mo_6O_{27}]^{4-} + H2O$	-0.66
$H_6MoO_6 + H_6MoO_6 \rightleftarrows H_8Mo_2O_{10} + 2H_2O$	15.58
$H_6MoO_6 + [H_6Mo_2O_{10}]^{2-} \rightleftarrows [H_8Mo_3O_{14}]^{2-} + 2H_2O$	-25.44
$H_6MoO_6 + [H_6Mo_3O_{14}]^{4-} \rightleftarrows [H_8Mo_4O_{18}]^{4-} + 2H_2O$	-18.48
$H_6MoO_6 + [H_6Mo_4O_{18}]^{6-} \rightleftarrows [H_8Mo_5O_{22}]^{4-} + 2H_2O$	-42.87
$H_6MoO_6 + [H_{11}Mo_6O_{27}]^{7-} \rightleftarrows [H_{13}Mo_7O_{31}]^{7-} + 2H_2O$	-36.55
$[H_7MoO_6]^+ + [H_{11}Mo_6O_{27}]^{7-} \rightleftarrows [H_{14}Mo_7O_{31}]^{6-} + 2H_2O$	-65.68
$[H_7MoO_6]^+ + [H_{12}Mo_7O_{31}]^{8-} \rightleftarrows [H_{15}Mo_8O_{35}]^{7-} + 2H_2O$	-62.89
$[HMoO_4]^- + [H_7MoO_6]^+ \rightleftarrows H_2Mo_2O_7 + 3H_2O$	-31.34
$[H_7MoO_6]^+ + [HMo_2O_7]^- \rightleftarrows H_2Mo_3O_{10} + 3H_2O$	-26.88
$[H_7MoO_6]^+ + [HMo_3O_{10}]^- \rightleftarrows H_2Mo_4O_{13} + 3H_2O$	-17.33
$[H_7MoO_6]^+ + [HMo_4O_{13}]^- \rightleftarrows H_2Mo_5O_{16} + 3H_2O$	9.89
$[H_7MoO_6]^+ + [HMo_5O_{17}]^{3-} \rightleftarrows [H_2Mo_6O_{20}]^{2-} + 3H_2O$	-24.75

Appendix C. Chemical Reactions

Reactions	Energies
$[H_7MoO_6]^+ + [HMo_6O_{20}]^{3-} \rightleftarrows [H_2Mo_7O_{23}]^{2-} + 3H_2O$	-23.86
$[H_7MoO_6]^+ + [HMo_7O_{23}]^{3-} \rightleftarrows [H_2Mo_8O_{26}]^{2-} + 3H_2O$	-25.8
$[H_7MoO_6]^+ + [H_{13}Mo_8O_{35}]^{9-} \rightleftarrows [H_{14}Mo_9O_{38}]^{10-} + 3H_2O$	-41.45
$[H_8MoO_6]^{2+} + [H_{13}Mo_8O_{35}]^{9-} \rightleftarrows [H_{15}Mo_9O_{38}]^{9-} + 3H_2O$	-92.02
$[H_8MoO_6]^{2+} + [H_{14}Mo_8O_{35}]^{8-} \rightleftarrows [H_{16}Mo_9O_{38}]^{8-} + 3H_2O$	-101.45
$[HMoO_4]^- + [HMoO_4]^- \rightleftarrows [Mo_2O_7]^{2-} + H_2O$	-1.8
$H_2MoO_4 + H_2MoO_4 \rightleftarrows H_2Mo_2O_7 + H_2O$	-0.98
$[HMoO_4]^- + [HMo_2O_7]^- \rightleftarrows [Mo_3O_{10}]^{2-} + H_2O$	-2.51
$H_2MoO_4 + H_2Mo_2O_7 \rightleftarrows H_2Mo_3O_{10} + H_2O$	-0.5
$[HMoO_4]^- + [HMo_3O_{10}]^- \rightleftarrows [Mo_4O_{13}]^{2-} + H_2O$	-11.51
$H_2MoO_4 + H_2Mo_3O_{10} \rightleftarrows H_2Mo_4O_{13} + H_2O$	8.66
$[HMoO_4]^- + [HMo_4O_{13}]^- \rightleftarrows [Mo_5O_{16}]^{2-} + H_2O$	8.4
$H_2MoO_4 + H_2Mo_4O_{13} \rightleftarrows H_2Mo_5O_{16} + H_2O$	26.01
$[HMoO_4]^- + [HMo_5O_{17}]^{3-} \rightleftarrows [Mo_6O_{20}]^{4-} + H_2O$	9.29
$H_2MoO_4 + [H_2Mo_5O_{17}]^{2-} \rightleftarrows [H_2Mo_6O_{20}]^{2-} + H_2O$	5.05
$[HMoO_4]^- + [HMo_6O_{20}]^{3-} \rightleftarrows [Mo_7O_{23}]^{4-} + H_2O$	-2.46
$H_2MoO_4 + [H_2Mo_6O_{20}]^{2-} \rightleftarrows [H_2Mo_7O_{23}]^{2-} + H_2O$	6.15
$[HMoO_4]^- + [HMo_7O_{23}]^{3-} \rightleftarrows [Mo_8O_{26}]^{4-} + H_2O$	-5.81
$H_2MoO_4 + [H_2Mo_7O_{23}]^{2-} \rightleftarrows [H_2Mo_8O_{26}]^{2-} + H_2O$	-9.77
$[H_8MoO_6]^{2+} + [H_2Mo_7O_{24}]^{4-} \rightleftarrows [H_2Mo_8O_{26}]^{2-} + 4H_2O$	-86.13
$[H_8MoO_6]^{2+} + [H_{14}Mo_8O_{35}]^{8-} \rightleftarrows [H_{14}Mo_9O_{37}]^{6-} + 4H_2O$	-108.32
$[MoO_4]^{2-} + [Mo_4O_{13}]^{2-} \rightleftarrows [Mo_5O_{17}]^{4-}$	11.95
$[HMoO_4]^- + [HMo_4O_{13}]^- \rightleftarrows [H_2Mo_5O_{17}]^{2-}$	3.65
$[MoO_4]^{2-} + [Mo_5O_{16}]^{2-} \rightleftarrows [Mo_6O_{20}]^{4-}$	-15.63
$[HMoO_4]^- + [HMo_5O_{16}]^- \rightleftarrows [H_2Mo_6O_{20}]^{2-}$	4.06
$[MoO_4]^{2-} + [Mo_6O_{20}]^{4-} \rightleftarrows [Mo_7O_{24}]^{6-}$	25.36
$[HMoO_4]^- + [HMo_6O_{20}]^{4-} \rightleftarrows [H_2Mo_7O_{24}]^{4-}$	22.51
$[H_{18}Mo_9O_{37}]^{2-} + [H_{18}Mo_9O_{37}]^{2-} \rightleftarrows [H_{16}Mo_{18}O_{65}]^{6-} + 10H_2O$	-36.79

Reactions	Energies
$[H_{19}Mo_9O_{37}]^- + [H_{19}Mo_9O_{38}]^{5-} \rightleftarrows [H_{18}Mo_{18}O_{65}]^{4-} + 10H_2O$	-76.03
$[H_{18}Mo_{18}O_{65}]^{4-} + [H_{18}Mo_{18}O_{65}]^{4-} \rightleftarrows [H_{32}Mo_{36}O_{128}]^{8-} + 2H_2O$	-39.91

Table C.2: Chemical reactions for the polyoxotungstate system. Energies in kcal·mol^{-1}.

Reactions	Energies
$[WO_4]^{2-} + H_5O_2^+ \rightleftarrows [HWO_4]^- + 2H_2O$	-27.53
$[HWO_4]^- + H_5O_2^+ \rightleftarrows H_2WO_4 + 2H_2O$	-9.95
$H_6WO_6 + H_5O_2^+ \rightleftarrows [H_7WO_6]^+ + 2H_2O$	5.49
$[H_7WO_6]^+ + H_5O_2^+ \rightleftarrows [H_8WO_6]^{2+} + 2H_2O$	21.26
$[W_2O_7]^{2-} + H_5O_2^+ \rightleftarrows [HW_2O_7]^- + 2H_2O$	-11.29
$[HW_2O_7]^- + H_5O_2^+ \rightleftarrows H_2W_2O_7 + 2H_2O$	-6.91
$[W_3O_{10}]^{2-} + H_5O_2^+ \rightleftarrows [HW_3O_{10}]^- + 2H_2O$	-6.62
$[HW_3O_{10}]^- + H_5O_2^+ \rightleftarrows H_2W_3O_{10} + 2H_2O$	-3.63
$[W_3O_{11}]^{4-} + H_5O_2^+ \rightleftarrows [HW_3O_{11}]^{3-} + 2H_2O$	-39.41
$[HW_3O_{11}]^{3-} + H_5O_2^+ \rightleftarrows [H_2W_3O_{11}]^{2-} + 2H_2O$	-24.06
$[W_4O_{13}]^{2-} + H_5O_2^+ \rightleftarrows [HW_4O_{13}]^- + 2H_2O$	-0.01
$[HW_4O_{13}]^- + H_5O_2^+ \rightleftarrows H_2W_4O_{13} + 2H_2O$	2.83
$[W_4O_{15}]^{6-} + H_5O_2^+ \rightleftarrows [HW_4O_{15}]^{5-} + 2H_2O$	-52.07
$[HW_4O_{15}]^{5-} + H_5O_2^+ \rightleftarrows [H_2W_4O_{15}]^{4-} + 2H_2O$	-37.12
$[W_5O_{16}]^{2-} + H_5O_2^+ \rightleftarrows [HW_5O_{16}]^- + 2H_2O$	-11.17
$[HW_5O_{16}]^- + H_5O_2^+ \rightleftarrows H_2W_5O_{16} + 2H_2O$	10.08
$[W_5O_{17}]^{4-} + H_5O_2^+ \rightleftarrows [HW_5O_{17}]^{3-} + 2H_2O$	-22.78
$[HW_5O_{17}]^{3-} + H_5O_2^+ \rightleftarrows [H_2W_5O_{17}]^{2-} + 2H_2O$	-13.32
$[W_5O_{19}]^{8-} + H_5O_2^+ \rightleftarrows [HW_5O_{19}]^{7-} + 2H_2O$	-69.4
$[HW_5O_{19}]^{7-} + H_5O_2^+ \rightleftarrows [H_2W_5O_{19}]^{6-} + 2H_2O$	-58.74
$[W_6O_{20}]^{6-} + H_5O_2^+ \rightleftarrows [HW_6O_{20}]^{5-} + 2H_2O$	-15.25

Appendix C. Chemical Reactions

Reactions	Energies
$[HW_6O_{20}]^{5-} + H_5O_2^+ \rightleftarrows [H_2W_6O_{20}]^{4-} + 2H_2O$	0.23
$[W_6O_{22}]^{8-} + H_5O_2^+ \rightleftarrows [HW_6O_{22}]^{7-} + 2H_2O$	-62.23
$[HW_6O_{22}]^{7-} + H_5O_2^+ \rightleftarrows [H_2W_6O_{22}]^{6-} + 2H_2O$	-47.48
$[W_7O_{24}]^{6-} + H_5O_2^+ \rightleftarrows [HW_7O_{24}]^{5-} + 2H_2O$	-26.13
$[HW_7O_{24}]^{5-} + H_5O_2^+ \rightleftarrows [H_2W_7O_{24}]^{4-} + 2H_2O$	-17.59
$[W_{10}O_{32}]^{4-} + H_5O_2^+ \rightleftarrows [HW_{10}O_{32}]^{3-} + 2H_2O$	0.27
$[HW_{10}O_{32}]^{3-} + H_5O_2^+ \rightleftarrows [H_2W_{10}O_{32}]^{2-} + 2H_2O$	5.05
$[W_{11}O_{40}]^{14-} + H_5O_2^+ \rightleftarrows [HW_{11}O_{40}]^{13-} + 2H_2O$	-88.58
$[HW_{11}O_{40}]^{13-} + H_5O_2^+ \rightleftarrows [H_2W_{11}O_{40}]^{12-} + 2H_2O$	-86.61
$\alpha - [W_{12}O_{40}]^{8-} + H_5O_2^+ \rightleftarrows [HW_{12}O_{40}]^{7-} + 2H_2O$	-52.15
$\alpha - [HW_{12}O_{40}]^{7-} + H_5O_2^+ \rightleftarrows [H_2W_{12}O_{40}]^{6-} + 2H_2O$	-42.4
$[W_{12}O_{42}]^{12-} + H_5O_2^+ \rightleftarrows [HW_{12}O_{42}]^{11-} + 2H_2O$	-72.97
$[HW_{12}O_{42}]^{11-} + H_5O_2^+ \rightleftarrows [H_2W_{12}O_{42}]^{10-} + 2H_2O$	-62.53
$H_2WO_4 + H_2O \rightleftarrows H_6WO_6$	16.05
$[HWO_4]^- + [HWO_4]^- \rightleftarrows [W_2O_7]^{2-} + H_2O$	-0.96
$H_2WO_4 + H_2WO_4 \rightleftarrows H_2W_2O_7 + H_2O$	0.74
$[HWO_4]^- + [HW_2O_7]^- \rightleftarrows [W_3O_{10}]^{2-} + H_2O$	-3.73
$H_2WO_4 + H_2W_2O_7 \rightleftarrows H_2W_3O_{10} + H_2O$	2.88
$[HWO_4]^- + [W_3O_{10}]^- \rightleftarrows [W_4O_{13}]^{2-} + H_2O$	-19.45
$H_2WO_4 + H_2W_3O_{10} \rightleftarrows H_2W_4O_{13} + H_2O$	-3.05
$[HWO_4]^- + [HW_4O_{13}]^- \rightleftarrows [W_5O_{16}]^{2-} + H_2O$	5.22
$H_2WO_4 + H_2W_4O_{13} \rightleftarrows H_2W_5O_{16} + H_2O$	11.25
$[HWO_4]^- + [HW_5O_{17}]^{3-} \rightleftarrows [W_6O_{20}]^{6-} + H_2O$	-3.19
$H_2WO_4 + [H_2W_5O_{17}]^{2-} \rightleftarrows [H_2W_6O_{20}]^{4-} + H_2O$	5.06
$[HWO_4]^- + [HW_5O_{19}]^{7-} \rightleftarrows [W_6O_{22}]^{8-} + H_2O$	11.01
$H_2WO_4 + [H_2W_5O_{19}]^{6-} \rightleftarrows [H_2W_6O_{22}]^{6-} + H_2O$	-30.01
$[WO_4]^{2-} + [HW_2O_7]^{2-} \rightleftarrows [W_3O_{11}]^{4-}$	18.21
$[WO_4]^{2-} + [HW_2O_7]^- \rightleftarrows [HW_3O_{11}]^{3-}$	-9.91

Reactions	Energies
$[HWO_4]^- + [HW_2O_7]^{2-} \rightleftarrows [HW_3O_{11}]^{3-}$	6.33
$[HWO_4]^- + [HW_2O_7]^- \rightleftarrows [H_2W_3O_{11}]^{2-}$	-6.44
$[HWO_4]^- + [W_3O_{11}]^{4-} \rightleftarrows [HW_4O_{15}]^{5-}$	27.47
$[HWO_4]^- + [HW_3O_{11}]^{3-} \rightleftarrows [H_2W_4O_{15}]^{4-}$	29.76
$[WO_4]^{2-} + [W_4O_{13}]^{2-} \rightleftarrows [W_5O_{17}]^{4-}$	6.25
$[WO_4]^{2-} + [HW_4O_{13}]^- \rightleftarrows [HW_5O_{17}]^{3-}$	-16.52
$[HWO_4]^- + [W_4O_{13}]^{2-} \rightleftarrows [HW_5O_{17}]^{3-}$	11
$[HWO_4]^- + [HW_4O_{13}]^- \rightleftarrows [H_2W_5O_{17}]^{2-}$	-2.31
$[HWO_4]^- + [W_4O_{15}]^{6-} \rightleftarrows [HW_5O_{19}]^{7-}$	23.18
$[HWO_4]^- + [HW_4O_{15}]^{5-} \rightleftarrows [H_2W_5O_{19}]^{6-}$	16.51
$[WO_4]^{2-} + [W_5O_{16}]^{2-} \rightleftarrows [W_6O_{20}]^{6-}$	-24.93
$[WO_4]^{2-} + [HW_5O_{16}]^- \rightleftarrows [HW_6O_{20}]^{5-}$	-29.01
$[HWO_4]^- + [W_5O_{16}]^{2-} \rightleftarrows [HW_6O_{20}]^{5-}$	-12.65
$[HWO_4]^- + [HW_5O_{16}]^- \rightleftarrows [H_2W_6O_{20}]^{4-}$	-1.25
$[WO_4]^{2-} + [W_6O_{20}]^{6-} \rightleftarrows [W_7O_{24}]^{6-}$	10.09
$[WO_4]^{2-} + [HW_6O_{20}]^{5-} \rightleftarrows [HW_7O_{24}]^{5-}$	-0.79
$[HWO_4]^- + [W_6O_{20}]^{6-} \rightleftarrows [HW_7O_{24}]^{5-}$	11.49
$[HWO_4]^- + [HW_6O_{20}]^{5-} \rightleftarrows [H_2W_7O_{24}]^{4-}$	9.15
$[W_5O_{16}]^{2-} + [W_5O_{16}]^{2-} \rightleftarrows [W_{10}O_{32}]^{4-}$	-61.91
$[HW_5O_{16}]^- + [HW_5O_{16}]^- \rightleftarrows [H_2W_{10}O_{32}]^{2-}$	-34.25
$[W_6O_{20}]^{6-} + [W_6O_{20}]^{6-} \rightleftarrows \alpha - [W_{12}O_{40}]^{8-}$	+49.71
$[HW_6O_{20}]^{5-} + [HW_6O_{20}]^{5-} \rightleftarrows \alpha - [H_2W_{12}O_{40}]^{6-}$	-14.34
$[H_2W_6O_{22}]^{6-} + [H_2W_6O_{22}]^{6-} \rightleftarrows [H_2W_{11}O_{40}]^{12-} + H_2O$	79.92
$[H_2W_5O_{17}]^{2-} + [H_2W_5O_{17}]^{2-} \rightleftarrows [W_{10}O_{32}]^{4-} + 2H_2O$	-46.85
$[H_2W_6O_{22}]^{6-} + [H_2W_6O_{22}]^{6-} \rightleftarrows [W_{12}O_{42}]^{12-} + 2H_2O$	90.24

Table C.3: Chemical reactions for the polyoxovanadate system. Energies in kcal·mol^{-1}.

Reaction	Energies
$[VO_4]^{3-} + H_5O_2^+ \rightleftarrows [HVO_4]^{2-} + 2H_2O$	-56.36
$[HVO_4]^{2-} + H_5O_2^+ \rightleftarrows [H_2VO_4]^- + 2H_2O$	-38.04
$H_5VO_5 + H_5O_2^+ \rightleftarrows [H_6VO_5]^+ + 2H_2O$	-0.71
$[V_2O_7]^{4-} + H_5O_2^+ \rightleftarrows [HV_2O_7]^{3-} + 2H_2O$	-48.06
$[HV_2O_7]^{3-} + H_5O_2^+ \rightleftarrows [H_2V_2O_7]^{2-} + 2H_2O$	-37.93
$[V_3O_9]^{3-} + H_5O_2^+ \rightleftarrows [HV_3O_9]^{2-} + 2H_2O$	-21.71
$[HV_3O_9]^{2-} + H_5O_2^+ \rightleftarrows [H_2V_3O_9]^- + 2H_2O$	-11.58
$[V_3O_{10}]^{5-} + H_5O_2^+ \rightleftarrows [HV_3O_{10}]^{4-} + 2H_2O$	-44.46
$[HV_3O_{10}]^{4-} + H_5O_2^+ \rightleftarrows [H_2V_3O_{10}]^{3-} + 2H_2O$	-40.34
$[V_4O_{12}]^{4-} + H_5O_2^+ \rightleftarrows [HV_4O_{12}]^{3-} + 2H_2O$	-24.48
$[HV_4O_{12}]^{3-} + H_5O_2^+ \rightleftarrows [H_2V_4O_{12}]^{2-} + 2H_2O$	-21.09
$[V_4O_{13}]^{6-} + H_5O_2^+ \rightleftarrows [HV_4O_{13}]^{5-} + 2H_2O$	-48.49
$[HV_4O_{13}]^{5-} + H_5O_2^+ \rightleftarrows [H_2V_4O_{13}]^{4-} + 2H_2O$	-50.23
$[V_5O_{14}]^{3-} + H_5O_2^+ \rightleftarrows [HV_5O_{14}]^{2-} + 2H_2O$	-5.2
$[HV_5O_{14}]^{2-} + H_5O_2^+ \rightleftarrows [H_2V_5O_{14}]^- + 2H_2O$	-14.79
$[V_5O_{15}]^{5-} + H_5O_2^+ \rightleftarrows [HV_5O_{15}]^{4-} + 2H_2O$	-27.09
$[V_5O_{16}]^{7-} + H_5O_2^+ \rightleftarrows [HV_5O_{16}]^{6-} + 2H_2O$	-72.02
$[HV_5O_{16}]^{6-} + H_5O_2^+ \rightleftarrows [H_2V_5O_{16}]^{5-} + 2H_2O$	-33.03
$[V_6O_{18}]^{6-} + H_5O_2^+ \rightleftarrows [HV_6O_{18}]^{5-} + 2H_2O$	-26.31
$[V_6O_{19}]^{8-} + H_5O_2^+ \rightleftarrows [HV_6O_{19}]^{7-} + 2H_2O$	-76.38
$[HV_6O_{19}]^{7-} + H_5O_2^+ \rightleftarrows [H_2V_6O_{19}]^{6-} + 2H_2O$	-59.78
$[V_7O_{21}]^{7-} + H_5O_2^+ \rightleftarrows [HV_7O_{21}]^{6-} + 2H_2O$	-29.38
$[V_8O_{24}]^{8-} + H_5O_2^+ \rightleftarrows [HV_8O_{24}]^{7-} + 2H_2O$	-26.61
$[V_9O_{27}]^{9-} + H_5O_2^+ \rightleftarrows [HV_9O_{27}]^{8-} + 2H_2O$	-29.47
$[V_{10}O_{28}]^{6-} + H_5O_2^+ \rightleftarrows [HV_{10}O_{28}]^{5-} + 2H_2O$	-25.04

Reaction	Energies
$[HV_{10}O_{28}]^{5-} + H_5O_2^+ \rightleftarrows [H_2V_{10}O_{28}]^{4-} + 2H_2O$	-17.59
$[H_2VO_4]^- + H_3O^+ \rightleftarrows H_5VO_5$	-7.12
$[HVO_4]^{2-} + H_5VO_5 \rightleftarrows [H_2V_2O_7]^{2-} + +2H_2O$	-43.29
$[H_2VO_4]^- + [H_2V_2O_7]^{2-} \rightleftarrows [V_3O_9]^{3-} + 2H_2O$	-8.42
$H_5VO_5 + [HV_2O_7]^{3-} \rightleftarrows [H_2V_3O_{10}]^{3-} + 2H_2O$	-40.41
$[H_2VO_4]^- + [H_2V_3O_{10}]^{3-} \rightleftarrows [V_4O_{12}]^{4-} + 2H_2O$	-10.87
$H_5VO_5 + [HV_3O_9]^{2-} \rightleftarrows [H_2V_4O_{12}]^{2-} + 2H_2O$	-28.79
$H_5VO_5 + [HV_3O_{10}]^{4-} \rightleftarrows [H_2V_4O_{13}]^{4-} + 2H_2O$	-24.87
$[H_2VO_4]^- + [H_2V_4O_{12}]^{2-} \rightleftarrows [V_5O_{14}]^{3-} + 2H_2O$	16.44
$[H_2VO_4]^- + [H_2V_4O_{13}]^{4-} \rightleftarrows [V_5O_{15}]^{5-} + 2H_2O$	-26.17
$H_5VO_5 + [HV_4O_{13}]^{5-} \rightleftarrows [H_2V_5O_{16}]^{5-} + 2H_2O$	-2.76
$[H_2VO_4]^- + [H_2V_5O_{16}]^{5-} \rightleftarrows [V_6O_{18}]^{6-} + 2H_2O$	-74.35
$H_5VO_5 + [HV_5O_{16}]^{6-} \rightleftarrows [H_2V_6O_{19}]^{6-} + 2H_2O$	-30.23
$[VO_4]^{3-} + [H_2VO_4]^- \rightleftarrows [V_2O_7]^{4-} + H_2O$	-9.22
$[HVO_4]^{2-} + [HVO_4]^{2-} \rightleftarrows [V_2O_7]^{4-} + H_2O$	9.1
$[HVO_4]^{2-} + [H_2VO_4]^- \rightleftarrows [HV_2O_7]^{3-} + H_2O$	-0.92
$[H_2VO_4]^- + [H_2VO_4]^- \rightleftarrows [H_2V_2O_7]^{2-} + H_2O$	-0.81
$[VO_4]^{3-} + [H_2V_2O_7]^{2-} \rightleftarrows [V_3O_{10}]^{5-} + H_2O$	-7.64
$[HVO_4]^{2-} + [HV_2O_7]^{3-} \rightleftarrows [V_3O_{10}]^{5-} + H_2O$	10.79
$[H_2VO_4]^- + [V_2O_7]^{4-} \rightleftarrows [V_3O_{10}]^{5-} + H_2O$	0.77
$[HVO_4]^{2-} + [H_2V_2O_7]^{2-} \rightleftarrows [HV_3O_{10}]^{4-} + H_2O$	4.26
$[H_2VO_4]^- + [HV_2O_7]^{3-} \rightleftarrows [HV_3O_{10}]^{4-} + H_2O$	4.37
$[H_2VO_4]^- + [H_2V_2O_7]^{2-} \rightleftarrows [H_2V_3O_{10}]^{3-} + H_2O$	1.96
$[VO_4]^{3-} + [H_2V_3O_9]^- \rightleftarrows [V_4O_{12}]^{4-} + H_2O$	-61.6
$[HVO_4]^{2-} + [HV_3O_9]^{2-} \rightleftarrows [V_4O_{12}]^{4-} + H_2O$	-16.82
$[H_2VO_4]^- + [V_3O_9]^{3-} \rightleftarrows [V_4O_{12}]^{4-} + H_2O$	-0.49
$[HVO_4]^{2-} + [H_2V_3O_9]^- \rightleftarrows [HV_4O_{12}]^{3-} + H_2O$	-29.72
$[H_2VO_4]^- + [HV_3O_9]^{2-} \rightleftarrows [HV_4O_{12}]^{3-} + H_2O$	-3.26

Reaction	Energies
$[H_2VO_4]^- + [H_2V_3O_9]^- \rightleftarrows [H_2V_4O_{12}]^{2-} + H_2O$	-12.77
$[VO_4]^{3-} + [H_2V_4O_{12}]^{2-} \rightleftarrows [V_5O_{15}]^{5-} + H_2O$	-44.22
$[HVO_4]^{2-} + [HV_4O_{12}]^{3-} \rightleftarrows [V_5O_{15}]^{5-} + H_2O$	-8.95
$[H_2VO_4]^- + [V_4O_{12}]^{4-} \rightleftarrows [V_5O_{15}]^{5-} + H_2O$	4.61
$[HVO_4]^{2-} + [H_2V_4O_{12}]^{2-} \rightleftarrows [HV_5O_{15}]^{4-} + H_2O$	-14.95
$[H_2VO_4]^- + [HV_4O_{12}]^{3-} \rightleftarrows [HV_5O_{15}]^{4-} + H_2O$	2.0
$[HVO_4]^{2-} + [HV_5O_{15}]^{4-} \rightleftarrows [V_6O_{18}]^{6-} + H_2O$	-7.22
$[H_2VO_4]^- + [V_5O_{15}]^{5-} \rightleftarrows [V_6O_{18}]^{6-} + H_2O$	3.73
$[H_2VO_4]^- + [HV_5O_{15}]^{4-} \rightleftarrows [HV_6O_{18}]^{5-} + H_2O$	4.51
$[H_2VO_4]^- + [H_2V_5O_{16}]^{5-} \rightleftarrows [H_2V_6O_{19}]^{6-} + H_2O$	7.24
$[HVO_4]^{2-} + [HV_6O_{18}]^{5-} \rightleftarrows [V_7O_{21}]^{7-} + H_2O$	-6.95
$[H_2VO_4]^- + [V_6O_{18}]^{6-} \rightleftarrows [V_7O_{21}]^{7-} + H_2O$	4.78
$[H_2VO_4]^- + [HV_6O_{18}]^{5-} \rightleftarrows [HV_7O_{21}]^{6-} + H_2O$	1.71
$[HVO_4]^{2-} + [HV_7O_{21}]^{6-} \rightleftarrows [V_8O_{24}]^{8-} + H_2O$	-3.99
$[H_2VO_4]^- + [V_7O_{21}]^{7-} \rightleftarrows [V_8O_{24}]^{8-} + H_2O$	4.67
$[H_2VO_4]^- + [HV_7O_{21}]^{6-} \rightleftarrows [HV_8O_{24}]^{7-} + H_2O$	7.44
$[HVO_4]^{2-} + [HV_8O_{24}]^{7-} \rightleftarrows [V_9O_{27}]^{9-} + H_2O$	-4.76
$[H_2VO_4]^- + [V_8O_{24}]^{8-} \rightleftarrows [V_9O_{27}]^{9-} + H_2O$	6.67
$[H_2VO_4]^- + [HV_8O_{24}]^{7-} \rightleftarrows [HV_9O_{27}]^{8-} + H_2O$	3.81
$[V_5O_{14}]^{3-} + [V_5O_{14}]^{3-} \rightleftarrows [V_{10}O_{28}]^{6-}$	4.38
$[V_5O_{14}]^{3-} + [HV_5O_{14}]^{2-} \rightleftarrows [HV_{10}O_{28}]^{5-}$	-15.46
$[HV_5O_{14}]^{2-} + [HV_5O_{14}]^{2-} \rightleftarrows [H_2V_{10}O_{28}]^{4-}$	-27.85

Table C.4: Chemical reactions for the polyoxoniobate system. Energies. in kcal·mol^{-1}.

Reactions	Energies
$[NbO(OH)_2]^+ + H_5O_2^+ \rightleftarrows [Nb(OH)_3]^{2+} + 2H_2O$	28.44

Reactions	Energies
$NbO(OH)_3 + H_5O_2^+ \rightleftarrows [Nb(OH)_4]^+ + 2H_2O$	6.88
$[NbO_2(OH)_3]^{2-} + H_5O_2^+ \rightleftarrows [NbO(OH)_4]^- + 2H_2O$	-28.54
$[NbO(OH)_4]^- + H_5O_2^+ \rightleftarrows Nb(OH)_5 + 2H_2O$	-22.19
$[Nb_2O_7]^{4-} + H_5O_2^+ \rightleftarrows [HNb_2O_7]^{3-} + 2H_2O$	-44.06
$[HNb_2O_7]^{3-} + H_5O_2^+ \rightleftarrows [H_2Nb_2O_7]^{2-} + 2H_2O$	-41.2
$[Nb_3O_{10}]^{5-} + H_5O_2^+ \rightleftarrows [HNb_3O_{10}]^{4-} + 2H_2O$	-42.04
$[Nb_4O_{12}]^{4-} + H_5O_2^+ \rightleftarrows [HNb_4O_{12}]^{3-} + 2H_2O$	-25.63
$[Nb_4O_{13}]^{6-} + H_5O_2^+ \rightleftarrows [HNb_4O_{13}]^{5-} + 2H_2O$	-45.08
$[Nb_5O_{14}]^{3-} + H_5O_2^+ \rightleftarrows [HNb_5O_{14}]^{2-} + 2H_2O$	-16.27
$[Nb_5O_{16}]^{7-} + H_5O_2^+ \rightleftarrows [HNb_5O_{16}]^{6-} + 2H_2O$	-73.01
$[Nb_6O_{19}]^{8-} + H_5O_2^+ \rightleftarrows [HNb_6O_{19}]^{7-} + 2H_2O$	-61.05
$[HNb_6O_{19}]^{7-} + H_5O_2^+ \rightleftarrows [H_2Nb_6O_{19}]^{6-} + 2H_2O$	-48.1
$[H_2Nb_6O_{19}]^{6-} + H_5O_2^+ \rightleftarrows [H_3Nb_6O_{19}]^{5-} + 2H_2O$	-40.73
$[Nb_7O_{22}]^{9-} + H_5O_2^+ \rightleftarrows [HNb_7O_{22}]^{8-} + 2H_2O$	-54.24
$[H_2Nb_8O_{25}]^{8-} + H_5O_2^+ \rightleftarrows [H_3Nb_8O_{25}]^{7-} + 2H_2O$	-42.88
$[H_3Nb_8O_{25}]^{7-} + H_5O_2^+ \rightleftarrows [H_4Nb_8O_{25}]^{6-} + 2H_2O$	-33.46
$[H_4Nb_8O_{25}]^{6-} + H_5O_2^+ \rightleftarrows [H_5Nb_8O_{25}]^{5-} + 2H_2O$	-27.4
$[Nb_9O_{27}]^{9-} + H_5O_2^+ \rightleftarrows [HNb_9O_{27}]^{8-} + 2H_2O$	-62.4
$[HNb_9O_{27}]^{8-} + H_5O_2^+ \rightleftarrows [H_2Nb_9O_{27}]^{7-} + 2H_2O$	-44.87
$[Nb_{10}O_{28}]^{6-} + H_5O_2^+ \rightleftarrows [HNb_{10}O_{28}]^{5-} + 2H_2O$	-19.11
$[HNb_{10}O_{28}]^{5-} + H_5O_2^+ \rightleftarrows [H_2Nb_{10}O_{28}]^{4-} + 2H_2O$	-10.91
$[H_7Nb_{16}O_{49}]^{11-} + H_5O_2^+ \rightleftarrows [H_8Nb_{16}O_{49}]^{10-} + 2H_2O$	-32.42
$[H_9Nb_{24}O_{72}]^{15-} + H_5O_2^+ \rightleftarrows [H_{10}Nb_{24}O_{72}]^{14-} + 2H_2O$	-36.61
$[NbO(OH)_2]^+ + H_6O_3 \rightleftarrows NbO(OH)_3 + H_5O_2^+$	-31.29
$[Nb(OH)_3]^{2+} + H_6O_3 \rightleftarrows [Nb(OH)_4]^+ + H_5O_2^+$	-52.85
$NbO(OH)_3 + H_6O_3 \rightleftarrows [NbO(OH)_4]^- + H_5O_2^+$	28.93
$[Nb(OH)_4]^+ + H_6O_3 \rightleftarrows Nb(OH)_5 + H_5O_2^+$	-0.14
$[Nb_4O_{12}]^{4-} + H_6O_3 \rightleftarrows [HNb_4O_{13}]^{5-} + H_5O_2^+$	43.18

Appendix C. Chemical Reactions

Reactions	Energies
$NbO(OH)_3 + [NbO_2(OH)_3]^{2-} \rightleftarrows [H_2Nb_2O_7]^{2-} + 2H_2O$	-23.7
$[NbO_2(OH)_3]^{2-} + [HNb_2O_7]^{3-} \rightleftarrows [Nb_3O_{10}]^{5-} + 2H_2O$	1.33
$[NbO_2(OH)_3]^{2-} + [H_2Nb_2O_7]^{2-} \rightleftarrows [HNb_3O_{10}]^{4-} + 2H_2O$	0.49
$[NbO(OH)_4]^{-} + [HNb_2O_7]^{3-} \rightleftarrows [HNb_3O_{10}]^{4-} + 2H_2O$	-12.17
$NbO(OH)_3 + [HNb_3O_{10}]^{4-} \rightleftarrows [Nb_4O_{12}]^{4-} + 2H_2O$	-35.38
$[Nb(OH)_4]^{+} + [HNb_3O_{10}]^{4-} \rightleftarrows [HNb_4O_{12}]^{3-} + 2H_2O$	-67.89
$[NbO_2(OH)_3]^{2-} + [HNb_3O_{10}]^{4-} \rightleftarrows [Nb_4O_{13}]^{6-} + 2H_2O$	-4.59
$[NbO(OH)_4]^{-} + [HNb_3O_{10}]^{4-} \rightleftarrows [HNb_4O_{13}]^{5-} + 2H_2O$	-21.13
$NbO(OH)_3 + [HNb_4O_{12}]^{3-} \rightleftarrows [Nb_5O_{14}]^{3-} + 2H_2O$	-33.56
$[Nb(OH)_4]^{+} + [HNb_4O_{12}]^{3-} \rightleftarrows [HNb_5O_{14}]^{2-} + 2H_2O$	-56.71
$[NbO_2(OH)_3]^{2-} + [HNb_4O_{13}]^{5-} \rightleftarrows [Nb_5O_{16}]^{7-} + 2H_2O$	23.64
$[NbO(OH)_4]^{-} + [HNb_4O_{13}]^{5-} \rightleftarrows [HNb_5O_{16}]^{6-} + 2H_2O$	-20.83
$[NbO_2(OH)_3]^{2-} + [HNb_5O_{16}]^{6-} \rightleftarrows [Nb_6O_{19}]^{8-} + 2H_2O$	4.54
$[NbO(OH)_4]^{-} + [HNb_5O_{16}]^{6-} \rightleftarrows [HNb_6O_{19}]^{7-} + 2H_2O$	-27.97
$[NbO_2(OH)_3]^{2-} + [HNb_6O_{19}]^{7-} \rightleftarrows [Nb_7O_{22}]^{9-} + 2H_2O$	29.84
$[NbO_2(OH)_3]^{2-} + [H_2Nb_6O_{19}]^{6-} \rightleftarrows [HNb_7O_{22}]^{8-} + 2H_2O$	23.7
$[NbO(OH)_4]^{-} + [HNb_6O_{19}]^{7-} \rightleftarrows [HNb_7O_{22}]^{8-} + 2H_2O$	4.14
$NbO(OH)_3 + [H_2Nb_8O_{25}]^{8-} \rightleftarrows [HNb_9O_{27}]^{8-} + 2H_2O$	-35.83
$NbO(OH)_3 + [H_3Nb_8O_{25}]^{7-} \rightleftarrows [H_2Nb_9O_{27}]^{7-} + 2H_2O$	-37.82
$[Nb(OH)_4]^{+} + [H_2Nb_8O_{25}]^{8-} \rightleftarrows [H_2Nb_9O_{27}]^{8-} + 2H_2O$	-87.58
$[NbO(OH)_2]^{+} + [H_2Nb_9O_{27}]^{7-} \rightleftarrows [Nb_{10}O_{28}]^{6-} + 2H_2O$	-114.38
$[Nb(OH)_3]^{2+} + [HNb_9O_{27}]^{8-} \rightleftarrows [Nb_{10}O_{28}]^{6-} + 2H_2O$	-187.69
$[Nb(OH)_3]^{2+} + [H_2Nb_9O_{27}]^{7-} \rightleftarrows [HNb_{10}O_{28}]^{5-} + 2H_2O$	-161.93
$[H_5Nb_8O_{25}]^{5-} + [H_8Nb_{16}O_{49}]^{10-} \rightleftarrows [H_9Nb_{24}O_{72}]^{15-} + 2H_2O$	-2.91
$[H_4Nb_8O_{25}]^{6-} + [H_5Nb_8O_{25}]^{5-} \rightleftarrows [H_7Nb_{16}O_{49}]^{11-} + H_2O$	-10.43
$[H_5Nb_8O_{25}]^{5-} + [H_5Nb_8O_{25}]^{5-} \rightleftarrows [H_8Nb_{16}O_{49}]^{10-} + H_2O$	-15.45
$[NbO(OH)_2]^{+} + [Nb_3O_{10}]^{5-} \rightleftarrows [Nb_4O_{12}]^{4-} + H_2O$	-102.75
$[NbO(OH)_2]^{+} + [HNb_3O_{10}]^{4-} \rightleftarrows [HNb_4O_{12}]^{3-} + H_2O$	-86.34

Reactions	Energies
$[NbO(OH)_2]^+ + [Nb_4O_{12}]^{4-} \rightleftarrows [Nb_5O_{14}]^{3-} + H_2O$	-84.52
$[NbO(OH)_2]^+ + [HNb_4O_{12}]^{3-} \rightleftarrows [HNb_5O_{14}]^{2-} + H_2O$	-75.16
$NbO(OH)_3 + [HNb_5O_{16}]^{6-} \rightleftarrows [H_2Nb_6O_{19}]^{6-} + H_2O$	-41.18
$NbO(OH)_3 + [HNb_7O_{22}]^{8-} \rightleftarrows [H_2Nb_8O_{25}]^{8-} + H_2O$	-40.32
$[NbO(OH)_2]^+ + [H_2Nb_8O_{25}]^{8-} \rightleftarrows [H_2Nb_9O_{27}]^{7-} + H_2O$	-106.03
$[Nb_5O_{14}]^{3-} + [Nb_5O_{14}]^{3-} \rightleftarrows [Nb_{10}O_{28}]^{6-}$	-50.81
$[Nb_5O_{14}]^{3-} + [HNb_5O_{14}]^{2-} \rightleftarrows [HNb_{10}O_{28}]^{5-}$	-53.65
$[HNb_5O_{14}]^{2-} + [HNb_5O_{14}]^{2-} \rightleftarrows [H_2Nb_{10}O_{28}]^{4-}$	-48.29

Table C.5: Chemical reactions for the polyoxotantalate system. Energies in kcal·mol^{-1}.

Reactions	Energies
$[TaO(OH)_2]^+ + H_5O_2^+ \rightleftarrows [Ta(OH)_3]^{2+} + 2H_2O$	21.45
$TaO(OH)_3 + H_5O_2^+ \rightleftarrows [Ta(OH)_4]^+ + 2H_2O$	1.77
$[TaO_2(OH)_3]^{2-} + H_5O_2^+ \rightleftarrows [TaO(OH)_4]^- + 2H_2O$	-32.33
$[TaO(OH)_4]^- + H_5O_2^+ \rightleftarrows Ta(OH)_5 + 2H_2O$	-23.62
$[Ta_2O_7]^{4-} + H_5O_2^+ \rightleftarrows [HTa_2O_7]^{3-} + 2H_2O$	-44.19
$[HTa_2O_7]^{3-} + H_5O_2^+ \rightleftarrows [H_2Ta_2O_7]^{2-} + 2H_2O$	-38.38
$[Ta_3O_{10}]^{5-} + H_5O_2^+ \rightleftarrows [HTa_3O_{10}]^{4-} + 2H_2O$	-39.73
$[Ta_4O_{12}]^{4-} + H_5O_2^+ \rightleftarrows [HTa_4O_{12}]^{3-} + 2H_2O$	-28.44
$[Ta_4O_{13}]^{6-} + H_5O_2^+ \rightleftarrows [HTa_4O_{13}]^{5-} + 2H_2O$	-56.39
$[Ta_5O_{14}]^{3-} + H_5O_2^+ \rightleftarrows [HTa_5O_{14}]^{2-} + 2H_2O$	-19.52
$[Ta_5O_{16}]^{7-} + H_5O_2^+ \rightleftarrows [HTa_5O_{16}]^{6-} + 2H_2O$	-71.14
$[Ta_6O_{19}]^{8-} + H_5O_2^+ \rightleftarrows [HTa_6O_{19}]^{7-} + 2H_2O$	-57.44
$[HTa_6O_{19}]^{7-} + H_5O_2^+ \rightleftarrows [H_2Ta_6O_{19}]^{6-} + 2H_2O$	-45.05
$[H_2Ta_6O_{19}]^{6-} + H_5O_2^+ \rightleftarrows [H_3Ta_6O_{19}]^{5-} + 2H_2O$	-37.43
$[Ta_7O_{22}]^{9-} + H_5O_2^+ \rightleftarrows [HTa_7O_{22}]^{8-} + 2H_2O$	-49.7

Reactions	Energies
$[H_2Ta_8O_{25}]^{8-} + H_5O_2^+ \rightleftarrows [H_3Ta_8O_{25}]^{7-} + 2H_2O$	-43.76
$[H_3Ta_8O_{25}]^{7-} + H_5O_2^+ \rightleftarrows [H_4Ta_8O_{25}]^{6-} + 2H_2O$	-28.13
$[H_4Ta_8O_{25}]^{6-} + H_5O_2^+ \rightleftarrows [H_5Ta_8O_{25}]^{5-} + 2H_2O$	-23.54
$[Ta_9O_{27}]^{9-} + H_5O_2^+ \rightleftarrows [HTa_9O_{27}]^{8-} + 2H_2O$	-63.19
$[HTa_9O_{27}]^{8-} + H_5O_2^+ \rightleftarrows [H_2Ta_9O_{27}]^{7-} + 2H_2O$	-55.83
$[Ta_{10}O_{28}]^{6-} + H_5O_2^+ \rightleftarrows [HTa_{10}O_{28}]^{5-} + 2H_2O$	-12.01
$[HTa_{10}O_{28}]^{5-} + H_5O_2^+ \rightleftarrows [H_2Ta_{10}O_{28}]^{4-} + 2H_2O$	-3.49
$[H_7Ta_{16}O_{49}]^{11-} + H_5O_2^+ \rightleftarrows [H_8Ta_{16}O_{49}]^{10-} + 2H_2O$	-31.75
$[H_9Ta_{24}O_{72}]^{15-} + H_5O_2^+ \rightleftarrows [H_{10}Ta_{24}O_{72}]^{14-} + 2H_2O$	-26.88
$[TaO(OH)_2]^+ + H_6O_3 \rightleftarrows TaO(OH)_3 + H_5O_2^+$	-38.22
$[Ta(OH)_3]^{2+} + H_6O_3 \rightleftarrows [Ta(OH)_4]^+ + H_5O_2^+$	-57.9
$TaO(OH)_3 + H_6O_3 \rightleftarrows [TaO(OH)_4]^- + H_5O_2^+$	25.16
$[Ta(OH)_4]^+ + H_6O_3 \rightleftarrows Ta(OH)_5 + H_5O_2^+$	-0.23
$[Ta_4O_{12}]^{4-} + H_6O_3 \rightleftarrows [HTa_4O_{13}]^{5-} + H_5O_2^+$	20.09
$TaO(OH)_3 + [TaO2(OH)_3]^{2-} \rightleftarrows [H_2Ta_2O_7]^{2-} + 2H_2O$	-19.4
$[TaO_2(OH)_3]^{2-} + [HTa_2O_7]^{3-} \rightleftarrows [Ta_3O_{10}]^{5-} + 2H_2O$	1.41
$[TaO_2(OH)_3]^{2-} + [H_2Ta_2O_7]^{2-} \rightleftarrows [HTa_3O_{10}]^{4-} + 2H_2O$	0.06
$[TaO(OH)_4]^- + [HTa_2O_7]^{3-} \rightleftarrows [HTa_3O_{10}]^{4-} + 2H_2O$	-5.99
$TaO(OH)_3 + [HTa_3O_{10}]^{4-} \rightleftarrows [Ta_4O_{12}]^{4-} + 2H_2O$	-30.77
$[Ta(OH)_4]^+ + [HTa_3O_{10}]^{4-} \rightleftarrows [HTa_4O_{12}]^{3-} + 2H_2O$	-60.98
$[TaO_2(OH)_3]^{2-} + [HTa_3O_{10}]^{4-} \rightleftarrows [Ta_4O_{13}]^{6-} + 2H_2O$	-11.78
$[TaO(OH)_4]^- + [HTa_3O_{10}]^{4-} \rightleftarrows [HTa_4O_{13}]^{5-} + 2H_2O$	-35.84
$TaO(OH)_3 + [HTa_4O_{12}]^{3-} \rightleftarrows [Ta_5O_{14}]^{3-} + 2H_2O$	-152.06
$[Ta(OH)_3]^{2+} + [HTa_4O_{13}]^{5-} \rightleftarrows [Ta_5O_{14}]^{3-} + 2H_2O$	-43.86
$[Ta(OH)_4]^+ + [HTa_4O_{12}]^{3-} \rightleftarrows [HTa_5O_{14}]^{2-} + 2H_2O$	-65.15
$[TaO_2(OH)_3]^{2-} + [HTa_4O_{13}]^{5-} \rightleftarrows [Ta_5O_{16}]^{7-} + 2H_2O$	27.59
$[TaO(OH)_4]^- + [HTa_4O_{13}]^{5-} \rightleftarrows [HTa_5O_{16}]^{6-} + 2H_2O$	-11.22
$[TaO_2(OH)_3]^{2-} + [HTa_5O_{16}]^{6-} \rightleftarrows [Ta_6O_{19}]^{8-} + 2H_2O$	-15.15

Reactions	Energies
$[TaO(OH)_4]^- + [HTa_5O_{16}]^{6-} \rightleftarrows [HTa_6O_{19}]^{7-} + 2H_2O$	-40.26
$[TaO_2(OH)_3]^{2-} + [HTa_6O_{19}]^{7-} \rightleftarrows [Ta_7O_{22}]^{9-} + 2H_2O$	28.29
$[TaO_2(OH)_3]^{2-} + [H_2Ta_6O_{19}]^{7-} \rightleftarrows [HTa_7O_{22}]^{9-} + 2H_2O$	23.64
$[TaO(OH)_4]^- + [HTa_6O_{19}]^{7-} \rightleftarrows [HTa_7O_{22}]^{8-} + 2H_2O$	10.92
$TaO(OH)_3 + [H_2Ta_8O_{25}]^{8-} \rightleftarrows [HTa_9O_{27}]^{8-} + 2H_2O$	-39.64
$TaO(OH)_3 + [H_3Ta_8O_{25}]^{7-} \rightleftarrows [H_2Ta_9O_{27}]^{7-} + 2H_2O$	-51.71
$[Ta(OH)_4]^+ + [H_2Ta_8O_{25}]^{8-} \rightleftarrows [H_2Ta_9O_{27}]^{7-} + 2H_2O$	-97.24
$[TaO(OH)_2]^+ + [H_2Ta_9O_{27}]^{7-} \rightleftarrows [Ta_{10}O_{28}]^{6-} + 2H_2O$	-106.06
$[Ta(OH)_3]^{2+} + [HTa_9O_{27}]^{8-} \rightleftarrows [Ta_{10}O_{28}]^{6-} + 2H_2O$	-183.34
$[Ta(OH)_3]^{2+} + [H_2Ta_9O_{27}]^{7-} \rightleftarrows [HTa_{10}O_{28}]^{5-} + 2H_2O$	-139.52
$[H_5Ta_8O_{25}]^{5-} + [H_8Ta_{16}O_{49}]^{10-} \rightleftarrows [H_9Ta_{24}O_{72}]^{15-} + 2H_2O$	-24.93
$[H_4Ta_8O_{25}]^{6-} + [H_5Ta_8O_{25}]^{5-} \rightleftarrows [H_7Ta_{16}O_{49}]^{11-} + H_2O_4$	-17.81
$[H_5Ta_8O_{25}]^{5-} + [H_5Ta_8O_{25}]^{5-} \rightleftarrows [H_8Ta_{16}O_{49}]^{10-} + H_2O_4$	-26.02
$[TaO(OH)_2]^+ + [HTa_3O_{10}]^{4-} \rightleftarrows [HTa_4O_{12}]^{3-} + H_2O$	-91.47
$[TaO(OH)_2]^+ + [Ta_4O_{12}]^{4-} \rightleftarrows [Ta_5O_{14}]^{3-} + H_2O$	-104.56
$[TaO(OH)_2]^+ + [HTa_4O_{12}]^{3-} \rightleftarrows [HTa_5O_{14}]^{2-} + H_2O$	-95.64
$TaO(OH)_3 + [HTa_5O_{16}]^{6-} \rightleftarrows [H_2Ta_6O_{19}]^{6-} + H_2O$	-54.19
$TaO(OH)_3 + [HTa_7O_{22}]^{8-} \rightleftarrows [H_2Ta_8O_{25}]^{8-} + H_2O$	-44.4
$[TaO(OH)_2]^+ + [H_2Ta_8O_{25}]^{8-} \rightleftarrows [H_2Ta_9O_{27}]^{8-} + H_2O$	-127.73
$[Ta_5O_{14}]^{3-} + [Ta_5O_{14}]^{3-} \rightleftarrows [Ta_{10}O_{28}]^{6-}$	-67.01
$[Ta_5O_{14}]^{3-} + [HTa_5O_{14}]^{2-} \rightleftarrows [HTa_{10}O_{28}]^{5-}$	-59.5
$[HTa_5O_{14}]^{2-} + [HTa_5O_{14}]^{2-} \rightleftarrows [H_2Ta_{10}O_{28}]^{4-}$	-43.47

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez

Bibliography

[1] Enric Petrus, Mireia Segado and Carles Bo. "Nucleation mechanisms and speciation of metal oxide clusters". In: *Chemical Science* 11.32 (2020), pp. 8448–8456 (cit. on pp. xiii, 139).

[2] Enric Petrus and Carles Bo. "Unlocking Phase Diagrams for Molybdenum and Tungsten Nanoclusters and Prediction of their Formation Constants". In: *The Journal of Physical Chemistry A* 125.23 (2021), pp. 5212–5219 (cit. on pp. xiii, 137, 139, 141).

[3] Enric Petrus, Mireia Segado and Carles Bo. *ioChem Data Collection*. 2022. URL: `https://doi.org/10.19061/iochem-bd-1-244` (cit. on pp. xiii, 51, 169).

[4] J J Berzelius. "Beitrag zur näheren Kenntniss des Molybdäns". In: *Annalen der Physik* 82.4 (1826), pp. 369–392 (cit. on p. 1).

[5] Galissard de Marignac J.-C. "Recherches sur les Acides Silicotungstiques". In: *Compt. Rendus* 58 (1864), pp. 809–812 (cit. on p. 2).

[6] Linus Pauling. "The molecular structure of the tungstosilicated and related compounds". In: *Journal of the American Chemical Society* 51.10 (1929), pp. 2868–2880 (cit. on p. 2).

[7] J. F. Keggin and William Lawrence Bragg. "The structure and formula of 12-phosphotungstic acid". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 144.851 (1934), pp. 75–100 (cit. on p. 2).

[8] Koichi Momma and Fujio Izumi. "*VESTA3* for three-dimensional visualization of crystal, volumetric and morphology data". In: *Journal of Applied Crystallography* 44.6 (2011), pp. 1272–1276 (cit. on p. 3).

[9] I. Lindqvist. "On the structure of the paratungstate ion". In: *Acta Crystallographica* 5.5 (1952), pp. 667–670 (cit. on p. 2).

[10] A. Durif, M. T. Averbuch-Pouchot and J. C. Guitel. "Structure d'un décavanadate d'hexasodium hydraté". In: *Acta Crystallographica Section B* 36.3 (Mar. 1980), pp. 680–682 (cit. on p. 2).

[11] E. J. Graeber and B. Morosin. "The molecular configuration of the decaniobate ion ($[Nb_{10}O_{26}]^{6-}$)". In: *Acta Crystallographica Section B* 33.7 (July 1977), pp. 2137–2143 (cit. on p. 2).

[12] Eric M. Villa et al. "Reaction dynamics of the decaniobate ion $[H_xNb_{10}O_{28}]^{(6-x)-}$ in water". In: *Angewandte Chemie - International Edition* 47.26 (2008), pp. 4844–4846 (cit. on pp. 2, 149).

[13] B. Dawson. "The structure of the 9(18)-heteropoly anion in potassium 9(18)-tungstophosphate, $K_6(P_2W_{18}O_{62})\cdot 14H_2O$". In: *Acta Crystallographica* 6.2 (1953), pp. 113–126 (cit. on p. 3).

[14] M. Pope, Y. Jeannin and M. Fournier. *Heteropoly and Isopoly Oxometalates*. Inorganic Chemistry Concepts. Springer, 1983 (cit. on p. 4).

[15] Achim Müller et al. "Organizational Forms of Matter: An Inorganic Super Fullerene and Keplerate Based on Molybdenum Oxide". In: *Angewandte Chemie International Edition* 37.24 (1998), pp. 3359–3363 (cit. on p. 4).

[16] Achim Müller et al. "Inorganic Chemistry Goes Protein Size: A Mo368 Nano-Hedgehog Initiating Nanochemistry by Symmetry Breaking". In: *Angewandte Chemie International Edition* 41.7 (2002), pp. 1162–1167 (cit. on p. 4).

[17] "Effective Storage of Electrons in Water by the Formation of Highly Reduced Polyoxometalate Clusters". In: *Journal of the American Chemical Society* 144.20 (2022), pp. 8951–8960 (cit. on p. 5).

[18] Aleksandar Bijelic, Manuel Aureliano and Annette Rompel. "Polyoxometalates as Potential Next-Generation Metallodrugs in the Combat Against Cancer". In: *Angewandte Chemie International Edition* 58.10 (2019), pp. 2980–2999 (cit. on p. 5).

[19] Ning Li et al. "Polyoxometalate-Based Compounds for Photo- and Electrocatalytic Applications". In: *Angewandte Chemie International Edition* 59.47 (2020), pp. 20779–20793 (cit. on p. 5).

[20] Moslem Ahmadian and Mansoor Anbia. "Oxidative Desulfurization of Liquid Fuels Using Polyoxometalate-Based Catalysts: A Review". In: *Energy & Fuels* 35.13 (2021), pp. 10347–10373 (cit. on p. 5).

[21] Hamid Khalilpour et al. "Application of Polyoxometalate-based composites for sensor systems: A review". In: *Journal of Composites and Compounds* 3.7 (2021), pp. 129–139 (cit. on p. 5).

[22] Jiawei Zhong, Javier Pérez-Ramírez and Ning Yan. "Biomass valorisation over polyoxometalate-based catalysts". In: *Green Chemistry* 23.1 (2021), pp. 18–36 (cit. on p. 5).

[23] Fei Lu et al. "Polyoxometalate-Based Nanomaterials Toward Efficient Cancer Diagnosis and Therapy". In: *Chemistry – A European Journal* 27.21 (2021), pp. 6422–6434 (cit. on p. 5).

[24] Ekaterina A. Eseva et al. "Heterogeneous Catalysts Containing an Anderson-Type Polyoxometalate for the Aerobic Oxidation of Sulfur-Containing Compounds". In: *Industrial and Engineering Chemistry Research* 60.39 (2021), pp. 14154–14165 (cit. on p. 6).

[25] Baijie Xu et al. "A Copper-Containing Polyoxometalate-Based Metal–Organic Framework as an Efficient Catalyst for Selective Catalytic Oxidation of Alkylbenzenes". In: *Inorganic Chemistry* 60.7 (2021), pp. 4792–4799 (cit. on p. 6).

[26] "Polyoxometalate electrocatalysts based on earthabundant metals for efficient water oxidation in acidic media". In: *Nature Chemistry* 10.1 (2018), pp. 24–30 (cit. on p. 6).

[27] Keigo Kamata and Kosei Sugahara. "Base Catalysis by Mono- and Polyoxometalates". In: *Catalysts* 7.11 (2017) (cit. on p. 6).

[28] Keigo Kamata et al. "Scope of chemical fixation of carbon dioxide catalyzed by a bifunctional monomeric tungstate". In: *Catalysis Today* 226 (2014), pp. 160–166 (cit. on p. 6).

[29] Manami Nakamura et al. "Microwave aided conversion of cellulose to glucose using polyoxometalate as catalyst". In: *RSC Adv.* 11 (55 2021), pp. 34558–34563 (cit. on p. 6).

[30] "Unraveling the Key Features of the Reactive State of Decatungstate Anion in Hydrogen Atom Transfer (HAT) Photocatalysis". In: *ACS Catalysis* 6.10 (2016), pp. 7174–7182 (cit. on p. 6).

[31] Davide Ravelli et al. "Site-Selective C–H Functionalization by Decatungstate Anion Photocatalysis: Synergistic Control by Polar and Steric Effects Expands the Reaction Scope". In: *ACS Catalysis* 8.1 (2018), pp. 701–713 (cit. on p. 6).

[32] Patrick J. Sarver, Noah B. Bissonnette and David W. C. MacMillan. "Decatungstate-Catalyzed C(sp3)–H Sulfinylation: Rapid Access to Diverse Organosulfur Functionality". In: *Journal of the American Chemical Society* 143.26 (2021), pp. 9737–9743 (cit. on p. 6).

[33] Qingbo Shen et al. "Improving the photocatalytic H_2 evolution activity of Keggin polyoxometalates anchoring copper-azole complexes". In: *Green Chem.* 23 (8 2021), pp. 3104–3114 (cit. on p. 6).

[34] Yujie Zhang, Francisco de Azambuja and Tatjana N. Parac-Vogt. "Zirconium oxo clusters as discrete molecular catalysts for the direct amide bond formation". In: *Catal. Sci. Technol* 12 (10 2022), pp. 3190–3201 (cit. on p. 6).

[35] Nada D. Savić, David E. Salazar Marcano and Tatjana N. Parac-Vogt. "Expanding the Scope of Polyoxometalates as Artificial Proteases towards Hydrolysis of Insoluble Proteins". In: *Chemistry – A European Journal* 28.8 (2022), e202104224 (cit. on p. 6).

[36] Francisco de Azambuja, Nele Steens and Tatjana N. Parac-Vogt. "Kinetic and Interaction Studies of Adenosine-5-Triphosphate (ATP) Hydrolysis with Polyoxovanadates". In: *Metals* 11.11 (2021) (cit. on p. 6).

[37] Anat Bashan and Ada Yonath. "The linkage between ribosomal crystallography, metal ions, heteropolytungstates and functional flexibility". In: *Journal of Molecular Structure* 890.1 (2008), pp. 289–294 (cit. on p. 6).

[38] *Nobel Prize Chemistry 2009*. `https://www.nobelprize.org/prizes/chemistry/2009/summary/`. Accessed: 2022-07-19 (cit. on p. 7).

[39] Chen Zhang et al. "A Polyoxometalate Cluster Paradigm with Self-Adaptive Electronic Structure for Acidity/Reducibility-Specific Photothermal Conversion". In: *Journal of the American Chemical Society* 138.26 (2016), pp. 8156–8164 (cit. on p. 7).

[40] Nuttaporn Samart et al. "Decavanadate Inhibits Mycobacterial Growth More Potently Than Other Oxovanadates". In: *Frontiers in Chemistry* 6 (2018), pp. 1–16 (cit. on pp. 7, 135).

[41] Juliana M. Missina et al. "Effects of Decavanadate Salts with Organic and Inorganic Cations on Escherichia coli, Giardia intestinalis, and Vero Cells". In: *Inorganic Chemistry* 57.19 (2018), pp. 11930–11941 (cit. on p. 7).

[42] Isabel Franco-Castillo et al. "New protective coatings against lampenflora growing in the Pommery Champagne cellar". In: *International Biodeterioration Biodegradation* 173 (2022), p. 105459 (cit. on p. 7).

[43] Andreas Flütsch et al. "HIV-1 protease inhibition potential of functionalized polyoxometalates". In: *Bioorganic amp; medicinal chemistry letters* 21.4 (2011), pp. 1162–1166 (cit. on p. 7).

[44] Toshihiro Yamase. "Polyoxometalates Active Against Tumors, Viruses, and Bacteria". In: *Biomedical Inorganic Polymers: Bioactivity and Applications of Natural and Synthetic Polymeric Inorganic Molecules*. Ed. by Werner E. G. Müller, Xiaohong Wang and Heinz C. Schröder. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 65–116. ISBN: 978-3-642-41004-8 (cit. on p. 7).

[45] Jean Pierre Launay. "Reduction de l'ion metatungstate: Stades eleves de reduction de $H_2W_{12}O_{40}^{6-}$, derives de l'ion $HW_{12}O_{40}^{7-}$ et discussion generale". In: *Journal of Inorganic and Nuclear Chemistry* 38.4 (1976), pp. 807–816 (cit. on p. 8).

[46] Yoshio Nishimoto et al. "Super-Reduced Polyoxometalates: Excellent Molecular Cluster Battery Components and Semipermeable Molecular Capacitors". In: *Journal of the American Chemical Society* 136.25 (2014), pp. 9042–9052 (cit. on p. 8).

[47] Simon Greiner et al. "Solid-state-stabilization of molecular vanadium oxides for reversible electrochemical charge storage". In: *Inorganic Chemistry Frontiers* 7.1 (2020), pp. 134–139 (cit. on pp. 8, 135).

[48] "Polyoxometalate-based metal-organic frameworks for boosting electrochemical capacitor performance". In: *Chemical Engineering Journal* 373.May (2019), pp. 587–597 (cit. on p. 8).

[49] Xiao Fan et al. "Chemical doping engineering by utilizing trilacunary Keggin polyoxometalates as a dopant for high performance perovskite solar cells". In: *Dalton Transactions* 50.1 (2021), pp. 279–286 (cit. on p. 8).

[50] Xue-Xin Li et al. "An unprecedented fully reduced {Mo V 60 } polyoxometalate: from an all-inorganic molecular light-absorber

model to improved photoelectronic performance". In: *Chemical Science* 13.16 (2022), pp. 4573–4580 (cit. on p. 8).

[51] Christoph Busche et al. "Design and fabrication of memory devices based on nanoscale polyoxometalate clusters". In: *Nature* 515.7528 (2014), pp. 545–549 (cit. on p. 8).

[52] "The Extraction of Polyoxometalate Flash Compact Model and Corresponding Circuit Simulation". In: *Journal of Nanoelectronics and Optoelectronics* 16.6 (2021), pp. 884–890 (cit. on p. 8).

[53] Marco Moors et al. "Insights from Adsorption and Electron Modification Studies of Polyoxometalates on Surfaces for Molecular Memory Applications". In: *Accounts of Chemical Research* 54.17 (2021), pp. 3377–3389 (cit. on p. 8).

[54] Xavier López et al. "Structure, properties and reactivity of polyoxometalates: a theoretical perspective". In: *Chem. Soc. Rev.* 41 (22 2012), pp. 7537–7571 (cit. on pp. 8, 16, 44).

[55] Paul Adrien Maurice Dirac. "Quantum mechanics of many-electron systems". In: *Proc. R. Soc. Lond. A* 123 (1929), pp. 714–733 (cit. on p. 9).

[56] *Nobel Prize Chemistry 1998*. `https://www.nobelprize.org/prizes/chemistry/1998/summary/`. Accessed: 2022-07-18 (cit. on p. 9).

[57] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Phys. Rev.* 140 (4A 1965), A1133–A1138 (cit. on p. 9).

[58] Jeremy N. Harvey et al. "Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis". In: *ACS Catalysis* 9.8 (2019), pp. 6803–6813 (cit. on p. 9).

[59] Zhong Ling Lang et al. "The self-assembly mechanism of the Lindqvist anion $[W_6O_{19}]^{2-}$ in aqueous solution: A density functional theory study". In: *Dalton Transactions* 41.37 (2012), pp. 11361–11368 (cit. on pp. 9, 44, 124).

[60] Fernando Steffler, Guilherme Ferreira de Lima and Hélio Anderson Duarte. "Polyoxomolybdate formation – A thermodynamic analysis from density functional/PCM calculations". In: *Chemical Physics Letters* 669 (2017), pp. 104–109 (cit. on pp. 9, 92, 113).

[61] Dylan Sures et al. "Alkali-Driven Disassembly and Reassembly of Molecular Niobium Oxide in Water". In: *Journal of the American Chemical Society* 140.34 (2018), pp. 10803–10813 (cit. on pp. 9, 148, 157).

[62] Albert Solé-Daura et al. "Origin of Selectivity in Protein Hydrolysis by Zr(IV)-Containing Metal Oxides as Artificial Proteases". In: *ACS Catalysis* 10 (22 2020), pp. 13455–13467 (cit. on p. 9).

[63] Jean Yves Kempf et al. "Relative basicities of the oxygen sites in $[V_{10}O_{28}]^{6-}$. An analysis of the ab initio determined distributions of the electrostatic potential and of the Laplacian of charge density". In: *Journal of the American Chemical Society* 114.4 (1992), pp. 1136–1146 (cit. on p. 10).

[64] Juan Miguel Maestre et al. "Ab Initio Study of the Relative Basicity of the External Oxygen Sites in $M_2W_4O_{19}^{4-}$ (M = Nb and V)". In: *Inorganic Chemistry* 37.12 (1998), pp. 3071–3077 (cit. on p. 10).

[65] Valérie Vallet and Ingmar Grenthe. "Structure and Bonding in Uranyl(VI) Peroxide and Crown Ether Complexes; Comparison of Quantum Chemical and Experimental Data". In: *Inorganic Chemistry* 56.24 (2017), pp. 15231–15240 (cit. on p. 10).

[66] Dolores Melgar, Nuno A.G. Bandeira and Carles Bo. "Electronic Structure Studies on the Whole Keplerate Family: Predicting New Members". In: *Chemistry - A European Journal* 23.22 (2017), pp. 5338–5344 (cit. on pp. 10, 16).

[67] Magda Pascual-Borràs et al. "^{17}O NMR chemical shifts in oxometalates: from the simplest monometallic species to mixed-metal polyoxometalates". In: *Chem. Sci.* 5 (5 2014), pp. 2031–2042 (cit. on p. 10).

[68] Soheila Mir et al. "DFT study of -Keggin, lacunary Keggin, and ironII–VI substituted Keggin polyoxometalates: the effect of oxidation state and axial ligand on geometry, electronic structures and oxygen transfer". In: *RSC Adv.* 10 (56 2020), pp. 33718–33730 (cit. on p. 10).

[69] Fernando Steffler, Guilherme Ferreira De Lima and Hélio Anderson Duarte. "The effect of the heteroatom (X=P, As, Si and Ge) on the geometrical and electronic properties of -Keggin polyoxometalates (M=Mo, W and Nb) – A DFT investigation". In: *Journal of Molecular Structure* 1213 (2020), p. 128159 (cit. on p. 10).

[70] Albert Solé-Daura et al. "Probing Polyoxometalate–Protein Interactions Using Molecular Dynamics Simulations". In: *Chemistry - A European Journal* 22 (2016), pp. 15280–15289 (cit. on p. 10).

[71] Mireia Segado, May Nyman and Carles Bo. "Aggregation Patterns in Low- And High-Charge Anions Define Opposite Solubil-

ity Trends". In: *Journal of Physical Chemistry B* 123 (49 2019), pp. 10505–10513 (cit. on p. 11).

[72] E. Falbo and T. J. Penfold. "Redox Potentials of Polyoxometalates from an Implicit Solvent Model and QM/MM Molecular Dynamics". In: *Journal of Physical Chemistry C* 124.28 (2020), pp. 15045–15056 (cit. on p. 11).

[73] "Combined theoretical and mass spectrometry study of the formation-fragmentation of small polyoxomolybdates". In: *Inorganic Chemistry* 50.16 (2011), pp. 7811–7819 (cit. on pp. 11, 90, 113, 123).

[74] Ning Zhang et al. "Hydration structures of vanadium/oxovanadium cations in the presence of sulfuric acid: A molecular dynamics simulation study". In: *Chemical Engineering Science* 195 (2019), pp. 683–692 (cit. on p. 11).

[75] Vasilios Duros et al. "Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates". In: *Angewandte Chemie International Edition* 56.36 (2017), pp. 10815–10820 (cit. on p. 11).

[76] Vasilios Duros et al. "Intuition-Enabled Machine Learning Beats the Competition When Joint Human-Robot Teams Perform Inorganic Chemical Experiments". In: *Journal of Chemical Information and Modeling* 59.6 (2019), pp. 2664–2671 (cit. on p. 11).

[77] Tasnim Rahman et al. "Predicting the Solubility of Inorganic Ion Pairs in Water". In: *Angewandte Chemie International Edition* 134.19 (2022) (cit. on p. 11).

[78] Jon Petter Gustafsson. "Visual MINTEQ 3.0 user guide". In: *KTH, Department of Land and Water Recources, Stockholm, Sweden* (2011) (cit. on p. 12).

[79] Caio Felippe Curitiba Marcellos et al. *PyEquIon: A Python Package For Automatic Speciation Calculations of Aqueous Electrolyte Solutions*. 2021. DOI: 10.48550/ARXIV.2101.07246 (cit. on p. 12).

[80] Ryan S. Kingsbury. *pyEQL*. https://github.com/rkingsbury/pyEQL. Accessed: 2022-07-08 (cit. on p. 12).

[81] Matthew P. Humphreys and Abigail J. Schiller. *Pytzer: the Pitzer model for chemical activities in aqueous solutions in Python (beta)*. Version v0.5.2. 2021. DOI: 10.5281/zenodo.5342712 (cit. on p. 12).

[82] Nadiia I. Gumerova and Annette Rompel. "Polyoxometalates in solution: speciation under spotlight". In: *Chemical Society Reviews* 49.21 (2020), pp. 7568–7601 (cit. on pp. 16, 36, 43, 66).

[83] Miras Haralampos N. et al. "Unveiling the Transient Template in the Self-Assembly of a Molecular Oxide Nanowheel". In: *Science* 327 (2010), pp. 72–74 (cit. on pp. 16, 123).

[84] Montserrat Filella and Peter M. May. "The aqueous solution thermodynamics of niobium under conditions of environmental and biological interest". In: *Applied Geochemistry* 122 (2020), p. 104729 (cit. on pp. 16, 47).

[85] Montserrat Filella and Peter M. May. "The aqueous solution thermodynamics of tantalum under conditions of environmental and biological interest". In: *Applied Geochemistry* 109 (2019), p. 104402 (cit. on pp. 16, 47, 136, 150).

[86] Nataliya V. Maksimchuk et al. "Activation of H_2O_2 over Zr(IV). Insights from Model Studies on Zr-Monosubstituted Lindqvist Tungstates". In: *ACS Catalysis* 11.16 (2021), pp. 10589–10603 (cit. on p. 16).

[87] Montaha Anjass, Grace A. Lowe and Carsten Streb. "Molecular Vanadium Oxides for Energy Conversion and Energy Storage: Current Trends and Emerging Opportunities". In: *Angewandte Chemie - International Edition* 60.14 (2021), pp. 7522–7532 (cit. on p. 16).

[88] "Unprecedented coupling reaction between two anionic species of a closo -decahydrodecaborate cluster and an Anderson-type polyoxometalate". In: *Dalton Transactions* 49.15 (2020), pp. 4685–4689 (cit. on p. 16).

[89] Nicolas P. Martin et al. "Strategic Capture of the $\{Nb_7\}$ Polyoxometalate". In: *Chemistry - A European Journal* (2019), pp. 10580–10584 (cit. on pp. 16, 136, 148, 158).

[90] Aric Hagberg, Pieter Swart and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008 (cit. on p. 17).

[91] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362 (cit. on p. 17).

[92] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272 (cit. on pp. 17, 53).

[93] Norman L. Biggs, Robin J. Wilson and E. Keith Lloyd. *Graph theory 1736-1936 / Norman L. Biggs, E. Keith Lloyd, Robin J. Wilson*. Clarendon Press Oxford [Eng.], 1976, p. 239 (cit. on p. 18).

[94] A. Thackray. *Atoms and Powers*. Harvard University Press, 1970, p. 226 (cit. on p. 19).

[95] Klaus Ruedenberg. "Free-Electron Network Model for Conjugated Systems. V. Energies and Electron Distributions in the FE MO Model and in the LCAO MO Model". In: *The Journal of Chemical Physics* 22.11 (1954), pp. 1878–1894 (cit. on p. 19).

[96] D. Bonchev and D.H. Rouvray. *Chemical Graph Theory: Introduction and Fundamentals*. Chemical Graph Theory. Taylor & Francis, 1991 (cit. on p. 20).

[97] N. Trinajstic. *Chemical Graph Theory*. CRC-Press, 1983 (cit. on p. 20).

[98] Ivan Gutman and Oskar E. Polansky. In: *SIAM Review* 30.2 (1988), pp. 348–350 (cit. on p. 20).

[99] Jonathan L. Gross and Jay Yellen. *Graph Theory and Its Applications*. 2005 (cit. on p. 21).

[100] David Weininger. "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules". In: *J. Chem. Inf. Comput. Sci.* 1 (1988) (cit. on p. 22).

[101] Zhenqin Wu et al. "MoleculeNet: a benchmark for molecular machine learning". In: *Chem. Sci.* 9 (2 2018), pp. 513–530 (cit. on p. 22).

[102] Miha Skalic et al. "Shape-Based Generative Modeling for de Novo Drug Design". In: *Journal of Chemical Information and Modeling* 59.3 (2019), pp. 1205–1214 (cit. on p. 22).

[103] Matthias Rupp et al. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning". In: *Phys. Rev. Lett.* 108 (5 2012), p. 058301 (cit. on p. 22).

[104] Mingjian Wen et al. "BonDNet: a graph neural network for the prediction of bond dissociation energies for charged molecules". In: *Chemical Science* (2021) (cit. on p. 23).

[105] Yashaswi Pathak, Sarvesh Mehta and U Deva Priyakumar. "Learning Atomic Interactions through Solvation Free Energy Prediction Using Graph Neural Networks". In: *Journal of Chemical Information and Modeling* (2021), acs.jcim.0c01413 (cit. on p. 23).

[106] Xiaolin Pan et al. "MolGpka: A Web Server for Small Molecule $p\mathrm{K}_a$ Prediction Using a Graph-Convolutional Neural Network". In: *Journal of Chemical Information and Modeling* 61.7 (2021), pp. 3159–3165 (cit. on p. 23).

[107] Liliana C Gallegos et al. "Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties". In: *Accounts of Chemical Research* 54.4 (2021), pp. 827–836 (cit. on p. 23).

[108] Dejun Jiang et al. "Could Graph Neural Networks Learn Better Molecular Representation for Drug Discovery? A Comparison Study of Descriptor-based and Graph-based Models". In: *Journal of Cheminformatics* (2021) (cit. on p. 23).

[109] Robin Winter et al. "Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations". In: *Chemical Science* 10.6 (2019), pp. 1692–1701 (cit. on p. 23).

[110] Jan P. Unsleber and Markus Reiher. "The Exploration of Chemical Reaction Networks". In: *Annual Review of Physical Chemistry* 71.1 (2020), pp. 121–142 (cit. on p. 23).

[111] Gregor N. Simm and Markus Reiher. "Context-Driven Exploration of Complex Chemical Reaction Networks". In: *Journal of Chemical Theory and Computation* 13.12 (2017), pp. 6108–6119 (cit. on p. 23).

[112] R. F. W. Bader. "Atoms in molecules". In: *Accounts of Chemical Research* 18.1 (1985), pp. 9–15 (cit. on p. 26).

[113] L. P. Cordella et al. "An improved algorithm for matching large graphs". In: *In: 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Cuen*. 2001, pp. 149–159 (cit. on p. 27).

[114] Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15 (cit. on p. 27).

[115] *Moore's Law*. `https://github.com/wallento/mooreandmore`. Accessed: 2022-03-29 (cit. on pp. 28, 29).

[116] Zulfia A. Chotchaeva. "P vs NP: P is Equal to NP: Desired Proof". In: *Global Journal of Computer Science and Technology* (2021) (cit. on p. 32).

[117] *Clay Mathematics Instiute - Millenium Price Problem*. `https://www.claymath.org/millennium-problems/p-vs-np-problem`, note = Accessed: 2022-03-29 (cit. on p. 32).

[118] Peter Bossaerts, Nitin Yadav and Carsten Murawski. "Uncertainty and computational complexity". In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 374.1766 (2019), p. 20180138 (cit. on p. 32).

[119] *Chess Engine - Stockfish*. `https://github.com/official-stockfish/Stockfish`, note = Accessed: 2022-03-29 (cit. on p. 32).

[120] Claude E. Shannon. "Programming a Computer Playing Chess". In: *Philosophical Magazine* Ser.7, 41.312 (1959) (cit. on p. 33).

[121] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: 594 (2016), A13 (cit. on p. 33).

[122] *Deep Blue beats Garry Kasparov*. `https://www.nytimes.com/1997/05/12/nyregion/swift-and-slashing-computer-topples-kasparov.html`, note = Accessed: 2022-03-29 (cit. on p. 33).

[123] Yixiang Fang et al. *A survey of community search over big graphs*. Vol. 29. 1. Springer Berlin Heidelberg, 2020, pp. 353–392 (cit. on p. 33).

[124] David Silver et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In: *Science* 362.6419 (2018), pp. 1140–1144 (cit. on p. 33).

[125] Michele Ceriotti, Cecilia Clementi and O. Anatole Von Lilienfeld. "Introduction: Machine Learning at the Atomic Scale". In: *Chemical Reviews* 121.16 (2021), pp. 9719–9721 (cit. on p. 33).

[126] A. Lavoisier. *Elements of Chemistry*. Dover Publications, 2011 (cit. on p. 35).

[127] C.L. Berthollet. *Essai de Statique Chimique (Éd.1803)*. Hachette Livre, 2012 (cit. on p. 44).

[128] Manuel Wilke and Nicola Casati. "A new route to polyoxometalates via mechanochemistry". In: *Chem. Sci.* 13 (4 2022), pp. 1146–1151 (cit. on p. 46).

[129] J.H. van 't Hoff. *Études de dynamique chimique*. Études de dynamique chimique v. 1. Muller, 1884 (cit. on p. 46).

[130] Johannes Noack et al. "Speciation of Molybdates under Hydrothermal Conditions". In: *Zeitschrift fur Anorganische und Allgemeine Chemie* 640 (14 2014), pp. 2730–2736 (cit. on p. 47).

[131] Norman C. Craig. "Entropy Diagrams". In: *Journal of Chemical Education* 73.8 (1996), p. 710 (cit. on p. 47).

[132] M J D Powell. *An efficient method for finding the minimum of a function of several variables without calculating derivatives*. 1964 (cit. on p. 53).

[133] Kenneth Levenberg. "A Method for the Solution of Certain Non-Linear Problems in Least Squares". In: *Quarterly of Applied Mathematics* 2 (2 1944), pp. 164–168 (cit. on p. 53).

[134] Donald G. Anderson. "Iterative Procedures for Nonlinear Integral Equations". In: *Journal of the ACM.* 12 (4 1965), pp. 547–560 (cit. on p. 53).

[135] N. E. Breslow and D. G Clayton. "Approximate Inference in Generalized Linear Mixed Models". In: *Journal of the American Statistical Association* 88 (421 1993), pp. 9–25 (cit. on p. 53).

[136] Gilbert Newton Lewis and M. J. Randall. "The activity coefficient of strong Electrolytes 1". In: *J. Am. Chem. Soc.* 43 (5 1921), pp. 1112–1154 (cit. on p. 56).

[137] Gilbert Newton Lewis. "Outlines of a New System of Thermodynamic Chemistry". In: *Proceedings of the American Academy of Arts and Sciences* 43.7 (1907), pp. 259–293 (cit. on p. 56).

[138] P. Debye and E. Hückel. *Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen.* Physikalische Zeitschrift, 1923, pp. 185–206 (cit. on p. 57).

[139] C. W. Davies. *Ion Association.* London: Butterworths, 1962, pp. 37–53 (cit. on p. 58).

[140] G Sposito and SV Mattigod. *Geochem: a computer program for the calculation of chemical equilibria in soil solution and other natural water systems.* 1980 (cit. on p. 60).

[141] DR Parker, LW Zelazny and TB Kinraide. "Improvements to the program Geochem". In: *Soil Sci Soc Am J* 51 (1987), pp. 488–491 (cit. on p. 60).

[142] Ryan Kingsbury. *PyEQL: a Python library for solution chemistry.* 2018. URL: `https://github.com/rkingsbury/pyEQL` (cit. on p. 61).

[143] Francesco Crea et al. "Modeling the acid-base properties of molybdate(VI) in different ionic media, ionic strengths and temperatures, by EDH, SIT and Pitzer equations". In: *Journal of Molecular Liquids* 229 (2017), pp. 15–26 (cit. on pp. 61, 105).

[144] Kristin S. Alongi and George C. Shields. "Theoretical calculations of acid dissociation constants. A review article". In: 6 (2010), pp. 113–138 (cit. on pp. 75, 86).

[145] Juan Frau, Noemí Hernández-Haro and Daniel Glossman-Mitnik. "Computational prediction of the $p\mathrm{K}_a$s of small peptides through Conceptual DFT descriptors". In: *Chemical Physics Letters* 671 (2017), pp. 138–141 (cit. on p. 76).

[146] Philipp Pracht et al. "High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pK_a values in the context of the SAMPL6 challenge". In: *Journal of Computer-Aided Molecular Design* 32.10 (2018), pp. 1139–1149 (cit. on p. 76).

[147] Donghai Yu et al. "Theoretical Study of pK_a Values for Trivalent Rare-Earth Metal Cations in Aqueous Solution". In: *Journal of Physical Chemistry A* 122.2 (2018), pp. 700–707 (cit. on p. 76).

[148] "The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data". In: *Journal of Physical Chemistry A* 102.40 (1998), pp. 7787–7794 (cit. on p. 77).

[149] J. Aveston. "Hydrolysis of Tungsten(VI): Ultracentrifugation, Acidity Measurements, and Raman Spectra of Polytungstates". In: *Inorganic Chemistry* 3.7 (1964), pp. 981–986 (cit. on p. 77).

[150] J.J. Cruywagen and Izak F. J van der Merwee. "Tungsten(VI) Equilibria: A Potentiometric and Calorimetric Investigation". In: *Dalton Trans.* 7 (1987), pp. 1701–1704 (cit. on pp. 77, 113).

[151] J. J. Cruywagen. "Potentiometric investigation of molybdenum(VI) equilibriums at 25 degree C in 1 M sodium chloride medium". In: *Inorganic Chemistry* 19.2 (1980), pp. 552–554 (cit. on p. 77).

[152] K. Pettersson et al. "Multicomponent Polyanioans. 36. Hydrolysis and Redox Equilibria of the H^+-$HVO_4{}^{2-}$ System in 0.6M NaCl. A complementary Potentiometric and ^{51}V NMR Study at Low Vanadium COncentrations in Acid Solution". In: *Acta Chemica Scandinavica A* 39 (1986), pp. 499–506 (cit. on p. 77).

[153] Alhadji Malloum, Jean Jules Fifen and Jeanet Conradie. "Determination of the absolute solvation free energy and enthalpy of the proton in solutions". In: *Journal of Molecular Liquids* 322 (2021), p. 114919 (cit. on p. 78).

[154] J.J. Cruywagen. *Protonation, Oligomerization, and Condensation Reactions of Vanadate(V), Molybdate(VI), and Tungstate(VI).* Vol. 49. 1999, pp. 127–182 (cit. on pp. 80, 94, 101, 118).

[155] J. J. Cruywagen et al. "Molybdenum(VI) equilibria in different ionic media. Formation constants and thermodynamic quantities". In: *Inorganica Chimica Acta* 331.1 (2002), pp. 322–329 (cit. on pp. 88, 112, 118).

[156] Laia Vilá Nadal et al. "Nucleation mechanisms of molecular oxides: A study of the assembly-dissassembly of $[W_6O_{19}]^{2-}$ by theory and mass spectrometry". In: *Angewandte Chemie - International Edition* 48.30 (2009), pp. 5452–5456 (cit. on pp. 90, 114, 124).

[157] "On the hydrolysis of niobium(V) and tantalum(V) in 3 mol·dm^{-3} KCl at 25°C. Part 1. Construction of a thermodynamic model for Nb(V)". In: *Journal of the Chemical Society, Dalton Transactions* 20 (1994), pp. 3055–3059 (cit. on pp. 95, 140).

[158] "Speciation in vanadium bioinorganic systems. 2. An NMR, ESR, and potentiometric study of the aqueous H^+-vanadate-maltol system". In: *Inorganic Chemistry* 35.11 (1996), pp. 3388–3393 (cit. on pp. 95, 140).

[159] Aymeric Sadoc et al. "Structure and stability of VO_2^+ in aqueous solution: A car-parrinello and static ab initio study". In: *Inorganic Chemistry* 46.12 (2007), pp. 4835–4843 (cit. on pp. 102, 144).

[160] Ashwani Kumar et al. "Engineering Multifunctionality in Hybrid Polyoxometalates: Aromatic Sulfonium Octamolybdates as Excellent Photochromic Materials and Self-Separating Catalysts for Epoxidation". In: *Inorganic Chemistry* 56.17 (2017), pp. 10325–10336 (cit. on p. 111).

[161] Chang-jiang Yang. "Polyoxometalate/Lead Composite Anode for Efficient Oxygen Evolution in Zinc Electrowinning". In: *Journal of The Electrochemical Society* 166.4 (2019), E129–E136 (cit. on p. 111).

[162] Ignacio Julian et al. "Polyoxometalates as alternative Mo precursors for methane dehydroaromatization on Mo/ZSM-5 and Mo/MCM-22 catalysts". In: *Catalysis Science and Technology* 9.21 (2019), pp. 5927–5942 (cit. on p. 111).

[163] Zhongling Lang et al. "Element table of TM-substituted polyoxotungstates for direct electrocatalytic reduction of nitric oxide to ammonia: a DFT guideline for experiments". In: *Inorg. Chem. Front.* 7 (22 2020), pp. 4507–4516 (cit. on p. 111).

[164] Jia-Qun Yu et al. "Trapping a $[W_{10}O_{32}]^{6-}$ Decatungstate Anion in an Ag44 Nanowheel". In: *Chemistry – An Asian Journal* 17.8 (2022), e202200072 (cit. on p. 111).

[165] Karl-Heinz Tytko and Oskar Glemser. "Isopolymolybdates and Isopolytungstates". In: *Advances in Inorganic Chemistry and Radiochemistry* 19 (1976), pp. 239–315 (cit. on p. 112).

[166] Oliver W. Howarth, Paul Kelly and Lage Pettersson. "Aqueous isopolymolybdates(VI): an oxygen-17 and molybdenum-95 nuclear magnetic resonance study". In: *J. Chem. Soc., Dalton Trans.* (1 1990), pp. 81–84 (cit. on p. 112).

[167] Raisa I. Maksimovskaya and Gennady M. Maksimov. "^{95}Mo and ^{17}O NMR Studies of Aqueous Molybdate Solutions". In: *Inorganic Chemistry* 46.9 (2007), pp. 3688–3695 (cit. on p. 112).

[168] Louis C.W. Baker and Diana C. Glick. "Present general status of understanding of heteropoly electrolytes and a tracing of some major highlights in the history of their elucidation". In: *Chemical Reviews* 98.1 (1998), pp. 3–49 (cit. on pp. 113, 118).

[169] G. M. Rozantsev and O. I. Sazonova. "Thermodynamic Parameters of Interconversions of Isopolyanions in Solutions of Tungsten(VI)". In: *Russ. J. Coord. Chem.* 31.8 (2005), pp. 552–558 (cit. on pp. 113, 119).

[170] Antonio Rodríguez-Fortea, Laia Vilà-Nadal and Josep M. Poblet. "Hydration of hydrogentungstate anions at different pH conditions: A car-parrinello molecular dynamics study". In: *Inorganic Chemistry* 47.17 (2008), pp. 7745–7750 (cit. on p. 114).

[171] Laia Vilà-Nadal et al. "Connecting theory with experiment to understand the initial nucleation steps of heteropolyoxometalate clusters". In: *Physical Chemistry Chemical Physics* 13.45 (2011), pp. 20136–20145 (cit. on p. 114).

[172] Enric Petrus and Carles Bo. *ioChem Data Collection*. 2021. URL: http://dx.doi.org/10.19061/iochem-bd-1-201 (cit. on pp. 115, 169).

[173] M. Álvarez-Moreno et al. "Managing the computational chemistry big data problem: The ioChem-BD platform". In: *Journal of Chemical Information and Modeling* 55.1 (2015), pp. 95–103 (cit. on pp. 115, 137, 169).

[174] Michael T. Pope and Achim Müller. "Polyoxometalate Chemistry: An Old Field with New Dimensions in Several Disciplines". In: *Angewandte Chemie International Edition* 30.1 (1991), pp. 34–48 (cit. on p. 118).

[175] "Spontaneous formation of autocatalytic sets with self-replicating inorganic metal oxide clusters". In: *Proceedings of the National Academy of Sciences* 117.20 (2020), pp. 10699–10705 (cit. on pp. 123, 125).

[176] Nathalie Leclerc-Laronze et al. "Structural Link between Giant Molybdenum Oxide Based Ions and Derived Keggin Structure: Modular Assemblies Based on the $[BW_{11}O_{39}]^{9-}$ Ion and Pentagonal $\{MM_5\}$ Units (M=W; M=Mo,W)". In: *Angewandte Chemie International Edition* 48.27 (2009), pp. 4986–4989 (cit. on p. 125).

[177] Jeremy J. Hastings and Oliver W. Howarth. "A 183W, 1H and 17O nuclear magnetic resonance study of aqueous isopolytungstates". In: *J. Chem. Soc., Dalton Trans.* (2 1992), pp. 209–215 (cit. on p. 128).

[178] May Nyman. "Polyoxoniobate chemistry in the 21st century". In: *Dalton Transactions* 40.32 (2011), p. 8049 (cit. on p. 134).

[179] Gauthier J.P. Deblonde et al. "Experimental and computational exploration of the UV-visible properties of hexaniobate and hexatantalate ions". In: *RSC Advances* 5.10 (2015), pp. 7619–7627 (cit. on pp. 134, 140).

[180] Lage Pettersson. "Equilibria of polyoxometalates in aqueous solution". In: *Molecular Engineering* 3.1 (1993), pp. 29–42 (cit. on p. 134).

[181] "New isopolyoxovanadate ions identified by electrospray mass spectrometry". In: *Inorganic Chemistry Communications* 2.10 (1999), pp. 487–489 (cit. on p. 134).

[182] J.-M. Tarascon and M Armand. "Issues and challenges facing rechargeable lithium batteries". In: *Nature* 414.6861 (2001), pp. 359–367 (cit. on p. 135).

[183] "Polyoxovanadate-Alkoxide Clusters as a Redox Reservoir for Iron". In: *Inorganic Chemistry* 56.12 (2017), pp. 7065–7080 (cit. on p. 135).

[184] "Accelerated design of vanadium redox flow battery electrolytes through tunable solvation chemistry". In: *Cell Reports Physical Science* (2021), p. 100323 (cit. on p. 135).

[185] Carsten Streb. "New trends in polyoxometalate photoredox chemistry: From photosensitisation to water oxidation catalysis". In: *Dalton Transactions* 41.6 (2012), pp. 1651–1659 (cit. on p. 135).

[186] Benjamin Schwarz et al. "Visible-light-driven water oxidation by a molecular manganese vanadium oxide cluster". In: *Angewandte Chemie - International Edition* 55.21 (2016), pp. 6329–6333 (cit. on p. 135).

[187] Manuel Aureliano and Debbie C. Crans. "Decavanadate ($V_{10}O_{28}^{6-}$) and oxovanadates: Oxometalates with many biological activities". In: *Journal of Inorganic Biochemistry* 103.4 (2009), pp. 536–546. ISSN: 01620134 (cit. on pp. 135, 155).

[188] Norio Fukuda and Toshihiro Yamase. "In Vitro Antibacterial Activity of Vanadate and Vanadyl Compounds against Streptococcus pneumoniae." In: *Biological and Pharmaceutical Bulletin* 20.8 (1997), pp. 927–930 (cit. on p. 135).

[189] Hamid Saeed Shah et al. "Cytotoxicity and enzyme inhibition studies of polyoxometalates and their chitosan nanoassemblies". In: *Toxicology Reports* 1 (2014), pp. 341–352 (cit. on pp. 135, 146).

[190] M. Aureliano and C. André Ohlin. "Decavanadate in vitro and in vivo effects: Facts and opinions". In: *Journal of Inorganic Biochemistry* 137 (2014), pp. 123–130 (cit. on p. 135).

[191] Archismita Misra et al. "Beyond Charge Balance: Counter-Cations in Polyoxometalate Chemistry". In: *Angewandte Chemie - International Edition* 59.2 (2020), pp. 596–612 (cit. on p. 135).

[192] Hai Yan Zhao et al. "State-of-the-art advances in the structural diversities and catalytic applications of polyoxoniobate-based materials". In: *Coordination Chemistry Reviews* 443 (2021), p. 213966 (cit. on p. 136).

[193] Ranko P. Bontchev and May Nyman. "Evolution of polyoxoniobate cluster anions". In: *Angewandte Chemie - International Edition* 45.40 (2006), pp. 6670–6672 (cit. on p. 136).

[194] Ryo Tsunashima et al. "The construction of high-nuclearity isopolyoxoniobates with pentagonal building blocks: $[HNb_{27}O_{76}]^{16-}$ and $[H_{10}Nb_{31}O_{93}(CO_3)]^{23-}$". In: *Angewandte Chemie - International Edition* 49.1 (2010), pp. 113–116 (cit. on p. 136).

[195] Peng Huang et al. "Self-Assembly and Photocatalytic Properties of Polyoxoniobates: $\{Nb_{24}O_{72}\}$, $\{Nb_{32}O_{96}\}$, and $\{K_{12}Nb_{96}O_{288}\}$ Clusters". In: *Journal of the American Chemical Society* 134.34 (2012), pp. 14004–14010 (cit. on p. 136).

[196] "Record High-Nuclearity Polyoxoniobates: Discrete Nanoclusters $\{Nb_{114}\}$, $\{Nb_{81}\}$, and $\{Nb_{52}\}$, and Extended Frameworks Based

on $\{Cu_3Nb_{78}\}$ and $\{Cu_4Nb_{78}\}$". In: *Angewandte Chemie - International Edition* 56.51 (2017), pp. 16288–16292 (cit. on p. 136).

[197] Yan-Lan Wu et al. "Two high-nuclearity isopolyoxoniobates containing $\{Nb_{54}O_{151}\}$-based helical nanotubes for the decomposition of chemical warfare agent simulants". In: *Chemical Communications* 29 (2022), pp. 4–7 (cit. on p. 136).

[198] A. L. Chabauty et al. "Niobium- and bismuth-silver phosphate glasses for the conditioning of radioactive iodine". In: *Journal of Non-Crystalline Solids* 510 (2019), pp. 51–61 (cit. on p. 136).

[199] Nataliya V. Maksimchuk et al. "Why Does Nb(V) Show Higher Heterolytic Pathway Selectivity Than Ti(IV) in Epoxidation with H_2O_2? Answers from Model Studies on Nb- and Ti-Substituted Lindqvist Tungstates". In: *ACS Catalysis* 9.7 (2019), pp. 6262–6275 (cit. on p. 136).

[200] Vasiliy Yu Evtushok et al. "Heterolytic alkene oxidation with H_2O_2 catalyzed by Nb-substituted Lindqvist tungstates immobilized on carbon nanotubes". In: *Catalysis Science and Technology* 11.9 (2021), pp. 3198–3207 (cit. on p. 136).

[201] Travis M. Anderson et al. "An aqueous route to $[Ta_6O_{19}]^{8-}$ and solid-state studies of isostructural niobium and tantalum oxide complexes". In: *Journal of the Chemical Society. Dalton Transactions* 9226.40 (2007), pp. 4517–4522 (cit. on p. 136).

[202] "Contrasting ion-association behaviour of Ta and Nb polyoxometalates". In: *Dalton Transactions* 43.41 (2014), pp. 15295–15299 (cit. on p. 136).

[203] Jung Ho Son and William H. Casey. "Titanium-Substituted Polyoxotantalate Clusters Exhibiting Wide pH Stabilities: $[Ti_2Ta_8O_{28}]^{8-}$ and $[Ti_{12}Ta_6O_{44}]^{10-}$". In: *Chemistry - A European Journal* 22.40 (2016), pp. 14155–14157 (cit. on pp. 136, 150).

[204] Miki Matsumoto et al. "Decatantalate - The last member of the group 5 decametalate family". In: *Inorganic Chemistry* 52.14 (2013), pp. 7825–7827 (cit. on p. 136).

[205] Hanggara Sudrajat et al. "Water-Splitting Activity of La-Doped $NaTaO_3$ Photocatalysts Sensitive to Spatial Distribution of Dopants". In: *Journal of Physical Chemistry C* 124.28 (2020), pp. 15285–15294 (cit. on p. 136).

[206] Sandeep Kumar et al. "Investigation of Ta_2O_5 as an alternative high-k dielectric for InAlN/GaN MOS-HEMT on Si". In: *IEEE Transactions on Electron Devices* 66.3 (2019), pp. 1230–1235 (cit. on p. 136).

[207] "Thermochemical measurements of alkali cation association to hexatantalate". In: *Molecules* 23.10 (2018), pp. 1–9 (cit. on p. 139).

[208] Nichola McCann, Michael Wagner and Hans Hasse. "A thermodynamic model for vanadate in aqueous solution - Equilibria and reaction enthalpies". In: *Dalton Transactions* 42.7 (2013), pp. 2622–2628 (cit. on p. 145).

[209] Manuel Aureliano et al. "Characterization of decavanadate and decaniobate solutions by Raman spectroscopy". In: *Dalton Transactions* 45.17 (2016), pp. 7391–7399 (cit. on pp. 148, 156, 158).

[210] Walter G. Klemperer and Keith A. Marek. "An ^{17}O NMR study of hydrolyzed Nb(V) in weakly acidic and basic aqueous solutions".

In: *European Journal of Inorganic Chemistry* 28.10-11 (2013), pp. 1762–1771 (cit. on p. 150).

[211] G. te Velde et al. "Chemistry with ADF". In: *Journal of Computational Chemistry* 22.9 (2001), pp. 931–967 (cit. on p. 169).

[212] J. P Perdew. "Density-functional approximation for the correlation energy of the in homogeneous electron gas". In: *Phys. Rev. B* 33.12 (1986), pp. 8822–8824 (cit. on p. 169).

[213] J. P Perdew. "Erratum: Density-functional approximation for the correlation energy of the inhomogeneous electron gas". In: *Phys. Rev. B* 34.10 (1986), pp. 7406–7406 (cit. on p. 169).

[214] Carlo Adamo and Vincenzo Barone. "Toward reliable density functional methods without adjustable parameters: The PBE0 model". In: *Journal of Chemical Physics* 110.13 (1999), pp. 6158–6170 (cit. on p. 169).

[215] Chengteh Lee, Weitao Yang and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Phys. Rev. B* 37.2 (1988), pp. 785–789 (cit. on p. 169).

[216] A. D. Becke. "Density-functional exchange-energy approximation with correct asymptotic behavior." In: *The Journal of Chemical Physics* 38 (1988), pp. 3098–3100 (cit. on p. 169).

[217] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange." In: *J. Chem. Phys.* 98 (1993), pp. 5648–5652 (cit. on p. 169).

[218] Stefan Grimme et al. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu". In: *Journal of Chemical Physics* 132.15 (2010) (cit. on p. 169).

[219] E Van Lenthe, E. J. Baerends and J. G. Snijders. "Relativistic regular two-component Hamiltonians". In: *J. Chem. Phys* 99.6 (1993), pp. 4597–4610 (cit. on p. 169).

[220] E Van Lenthe and E. J. Baerends. "Optimized Slater-type basis sets for the elements 1–118". In: *J. Comput. Chem* 24.9 (2003), pp. 1142–1156 (cit. on p. 169).

[221] A. J. Klamt. "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena". In: *J. Phys. Chem.* 99.7 (1995), pp. 2224–2235 (cit. on p. 169).

[222] Enric Petrus, Mireia Segado and Carles Bo. *ioChem Data Collection*. 2020. URL: `http://dx.doi.org/10.19061/iochem-bd-1-178` (cit. on p. 169).

[223] Enric Petrus. *POMSimulator*. `https://gitlab.com/enricpp/pomsimulator`. Accessed: 2022-09-13 (cit. on p. 169).

UNIVERSITAT ROVIRA I VIRGILI
POMSIMULATOR: A METHOD FOR UNDERSTANDING THE MULTI-EQUILIBRIA AND SELF-ASSEMBLY PROCESSES OF POLYOXOMETALATES
Enric Petrus Pérez