

Understanding structural adaptability: a reactant informatics approach to experiment design

Rosalind J. Xu, Jacob H. Olshansky, Philip D. F. Adler, Yongjia Huang, Matthew D. Smith, Matthias Zeller,[†] Joshua Schrier and Alexander J. Norquist

Department of Chemistry, Haverford College, Haverford PA 19041

[†] Department of Chemistry, Purdue University, West Lafayette IN

Abstract

The structural and electronic adaptability ranges of a $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework found in organically templated vanadium selenites were determined using a three step approach, informed by cheminformatics descriptors, involving (i) the extraction of the most important reaction parameters from historical reaction data, (ii) a fractional factorial design on those parameters to better explore chemical space and (iii) decision tree construction on organic molecular properties to determine the factors governing framework formation. This process enabled the elucidation of both the structural and electronic adaptability ranges and provided the context to extract chemical understanding from the structural features that give rise to these respective ranges. This work resulted in synthesis and structural determination of five new compounds.

· Author to whom correspondence should be addressed.

Haverford College
370 Lancaster Avenue
Haverford PA, 19041 USA
tel (610) 896 2949
fax (610) 896 4963
email anorquis@haverford.edu
<http://www.haverford.edu/chem/Norquist/>

Design, System, Application paragraph

Materials design is a problem of both “what to make” and “how to make it”—and given a set of reactions, it is usually unclear “why it works”. Furthermore, in the case of exploratory materials synthesis it is often not clear “where to start”. In this paper, we demonstrate a hybrid approach combining machine learning on small datasets and statistical design of experiments that addresses these problems, using the synthesis of organically-templated vanadium selenite compounds as a case study. Constructing a decision tree model on a small historical dataset yields informed, unbiased choices for the key variables governing crystal formation. Statistical design of experiments selects a systematic and optimal exploratory set of reactions probing these key variables. Finally, a decision tree constructed to predict the observed structural outcomes in terms of the reactant physicochemical properties explains the factors governing structure formation. As a specific example, we study organically-templated vanadium selenite compounds. Using this approach we discovered five novel compounds, and demonstrate that the structural outcomes can be explained in terms of a hierarchy of reactant properties.

Introduction

Scientific discovery is often conceived of as a process of extracting explanatory rules from observations, and then validating these explanations by testing falsifiable predictions. Historically, these scientific explanations are expressed as empirical “rules” (e.g., the octet rule) or as mathematical theories (e.g., quantum mechanics). The advent of digital computers has enabled numerical simulations to become another way of generating “observations” and explaining observed experimental phenomena.¹ More recently, informatics-based approaches to gather insights from digital datasets have become increasingly important in chemistry and materials science. For example, the launch of the Materials Genome Initiative for Global Competitiveness in 2011,² stimulated researchers to explore many new ways of using informatics approaches to discover new materials.^{3,4}

Most materials informatics studies have focused on identifying synthetic targets by simulating *properties* to eliminate poor performers. This has led to an extensive body of work on developing structure-property relationships of materials using both experimental and simulation data.^{4,5} However, the problem of understanding and predicting the *syntheses* of new materials still relies upon traditional Edisonian trial-and-error type approaches to determine appropriate reaction conditions. Although the experimental campaigns can be systematized by the use of statistical design-of-experiments⁶ and accelerated by the use of high-throughput experimentation,^{7,8} this strategy still typically relies upon human expertise to determine the important variables to search.

Unlike the analogous problem of informatics-based *organic* synthesis planning,⁹ *materials* synthesis planning is much less mature. Informatics-driven approaches for materials experimental designs have primarily focused on control of processing conditions. Representative

examples include work in optimizing gun power, tilt, and substrate height in magnetron sputtering of thin films,¹⁰ optimizing localized heating conditions to control microstructure formation during additive manufacturing,¹¹ and optimizing cooling rates for batch crystallization.¹² The synthesis of materials with novel compositions is more limited, and strategies have generally consisted of one of the following four approaches. The first approach uses unsupervised and semisupervised learning algorithms to identify boundaries in composition-property space (e.g. catalytic activity,¹³ and mapping of phase diagrams¹⁴). However, this strategy only determines regions of interest for further experimental attention using the existing precursors and processing conditions ranges, rather than identifying novel synthetic routes. The second approach involves computed thermodynamic considerations about equilibrium ground states¹⁵ or allowed metastability criteria¹⁶ to select feasible candidates. However, this strategy only determines if the material's existence is plausible, but does not provide a specific synthetic route. The third approach is to train machine learning models using synthesis data extracted from the published literature¹⁷ or from unpublished laboratory notebooks.¹⁸ However, this strategy relies upon having a large dataset of past experimental outcomes to construct the models, and thus is generally not applicable to new chemical systems (with the notable exception of ion-substitution similarity.¹⁹ A fourth approach is to use active machine learning methods to optimally select new exploratory synthetic experiments.^{20, 21} However, this strategy has only been used to optimize reaction conditions and stoichiometry, but not to provide insight about the physicochemical properties of the reactants that determine experimental success.

This paper includes an exploration of a hybrid approach for planning experiments and gaining insight into the factors governing the formation of different structural outcomes for amine-templated vanadium selenites of a $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework. Starting from a small

(75 discrete reactions) set of historical reaction data, we constructed a decision tree based on information gain to select the important reaction conditions, reactant masses, and reactant properties to study, and thus avoided inserting experimental biases into the choice of variables. We then used these reaction conditions to design a fractional-factorial experiment studying the effects of these variables in a more systematic and efficient way than the typical “one-variable-at-a-time” approach. This process resulted in the synthesis of five novel compounds. An explanation of the structural outcomes was generated in terms of the reactant properties by training a decision tree on predicted reactant physicochemical properties. The resulting model captures and ranks “the hierarchy of structural influences” previously proposed for these materials in a quantitative way.

Experimental

Materials. NH_4VO_3 (99 %), NaVO_3 (anhydrous, 99.9 %), $\text{VOSO}_4 \cdot x\text{H}_2\text{O}$ (97 %), SeO_2 (99.4 %), $\text{N,N}'$ -dimethylethylenediamine (dmed, 99 %), 1,2-bis(3-aminopropylamino)ethane (bape, ≥ 99 %), diethyltriamine (deta, 99.0 %) and piperazine (pip, 99 %) were purchased from Sigma-Aldrich. All reagents were used as received. Deionized water was used in these syntheses. A full list of amines used in this study is provided in the Electronic Supplementary Information.

Syntheses. All reactions were conducted in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels. Initial reaction pHs were controlled by the addition of 4 M HCl and 4 M NaOH. The reactions were heated to a set temperature and allowed to soak. The reactions were then cooled to room temperature at a rate of $6\text{ }^\circ\text{C h}^{-1}$ to promote the growth of large single crystals. Autoclaves and bottles were opened in air, and products were recovered via vacuum filtration.

The reactions shown below describe the optimized conditions under which no additional crystalline or amorphous reaction products were observed, for the purposes of bulk characterization. These are distinct from the factorial design experiments described in the text

[C₄H₁₄N₂]₂[VO(SeO₃)(HSeO₃)]₆·5H₂O (1) was synthesized as black-green crystals through the reaction of 0.179 g (8.04×10^{-4} mol) VOSO₄, 1.332 g (1.20×10^{-2} mol) SeO₂, 0.104 g (1.18×10^{-3} mol) dmed, and 6.00 g (3.33×10^{-1} mol) H₂O. The reaction was heated at 110 °C for 24 h in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels; initial pH was set to 1.

[C₈H₂₆N₄][VO(SeO₃)(HSeO₃)]₆·6H₂O (2) was synthesized as black-green crystals through the reaction of 0.0982 g (8.05×10^{-4} mol) NaVO₃, 1.334 g (1.20×10^{-2} mol) SeO₂, 0.141 g (8.11×10^{-4} mol) bape, and 6.06 g (3.36×10^{-1} mol) H₂O. The reaction was heated at 110 °C for 24 h in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels; initial pH was set to 1.

[C₄H₁₆N₂]₂[VO(SeO₃)(HSeO₃)]₈·9.333H₂O (3) was synthesized as black-green crystals through the reaction of 0.110 g (9.43×10^{-4} mol) NH₄VO₃, 1.233 g (1.11×10^{-2} mol) SeO₂, 0.0758 g (7.35×10^{-4} mol) deta, and 2.01 g (1.12×10^{-1} mol) H₂O. The reaction was heated at 110 °C for 24 h in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels; initial pH was set to 3.

[C₄H₁₂N₂][(VO)₃(SeO₃)(HSeO₃)₄]·H₂O (4) was synthesized as blue crystals through the reaction of 0.222 g (1.36×10^{-3} mol) VOSO₄, 0.888 g (8.00×10^{-3} mol) SeO₂, 0.0858 g (9.96×10^{-4} mol) pip, and 6.01 g (3.33×10^{-1} mol) H₂O. The reaction was heated at 110 °C for 24 h in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels; initial pH was set to 1.

[C₄H₁₂N₂][(VO)₂O₂(SeO₃)₂] (5) was synthesized as yellow crystals through the reaction of 0.113 g (9.64×10^{-4} mol) NH₄VO₃, 1.096 g (9.87×10^{-3} mol) SeO₂, 0.761 g (8.83×10^{-4} mol)

pip, and 6.26 g (3.47×10^{-1} mol) H₂O. The reaction was heated at 110 °C for 24 h in 23 mL poly(fluoro-ethylene-propylene) lined pressure vessels; initial pH was set to 1.

Single Crystal X-ray Diffraction. Data were collected using a Bruker AXS Smart Apex CCD, ApexII CCD or Quest CMOS diffractometers with Mo-K α radiation ($\lambda = 0.71073$ Å). The Smart Apex and ApexII instruments featured fine focus sealed tube X-ray sources with graphite monochromators. The Quest CMOS instrument is an I μ S microsource with a laterally graded multilayer (Goebel) mirror for monochromatization. A single crystal was mounted on a Mitegen micromesh mount using a trace of mineral oil and cooled *in-situ* to 100(2) K for data collection. Frames were collected, reflections were indexed and processed, and the files scaled and corrected for absorption using APEX2.²² The heavy atom positions were determined using SIR92.²³ All other non-hydrogen sites were located from Fourier difference maps. All non-hydrogen sites were refined using anisotropic thermal parameters using full matrix least squares procedures on F_o^2 with $I > 3\sigma(I)$. Hydrogen atoms were placed in geometrically idealized positions. All calculations were performed using Crystals v. 14.23c.²⁴ Relevant crystallographic data are listed in Table 1.

Powder X-ray Diffraction. Powder diffraction patterns were recorded on a GBC-Difftch MMA powder diffractometer. Samples were mounted on glass plates. Calculated powder patterns were generated from single crystal data using ATOMS v. 6.0.²⁵ Powder X-ray diffraction patterns were consistent with patterns predicted from the refined structures of **1** - **5**. No evidence of additional phases was observed.

Infrared Spectroscopy. Infrared measurements were obtained using a Perkin Elmer FT-IR Spectrum 1000 spectrophotometer. Samples were diluted with spectroscopic grade KBr and pressed into pellets. Scans were collected over the range of 400 – 4000 cm⁻¹.

Bond Valence Sums. Bond valence sums²⁶ calculations were performed using parameters compiled by Brese and O’Keeffe.²⁷ Complete tables of bond valence sums for compounds **1** – **5** are available in the Electronic supplementary information.

Decision tree generation. J48 type decision trees, which is a java implementation of the C4.5 decision tree algorithm,¹⁵ were generated using Weka.¹⁴ The algorithm was provided with the full set of descriptors and selected the descriptors that produced the ‘best’ split of the data. Heuristically, the best split is the one that most accurately separates the greatest number of results. Weka’s default parameters for the C4.5 decision tree algorithm were used, which includes a tolerance of two instances per leaf, a confidence factor of 0.25 for pruning. No artificial discretization was imposed prior to the processing by the C4.5 (J48) model. At each node, the normalized information gain is evaluated for splitting from each of the available criteria. The criteria with the highest information gain is chosen to split the tree. Training and testing were performed on both the historical and the factorial data set with 10-fold cross-validation, and the trees were pruned. The ChemAxon descriptors, which were used to analyze the factorial design outcomes, were calculated using the ChemAxon Calculator Plugin²⁸; see Table 4 for a list of ChemAxon descriptors. The historical descriptors are provided in Supplementary Data. Confusion matrices for each decision tree are available as Tables S12-S14 in the Electronic Supplementary Information, in addition to a series of calculated measures in Table S15.

Results and discussion

Our experimental efforts are often focused on the formation of new materials through exploratory syntheses. Traditionally, such work in new chemical systems is accomplished

through two stages. Initial exploratory reactions are generally based upon related reports in the literature, with wide variations introduced into many experimental parameters such as time, temperature, pH and reactant choice and relative reactant concentrations. As an example, our initial efforts in synthesizing novel templated vanadium selenites were based upon both our work on templated vanadium tellurites²⁹⁻³² and on previously reported vanadium selenites.³³⁻³⁹ Success in these initial exploratory efforts provide the foundation for more focused exploitation reactions. In this stage, a wide range of organic amines are utilized in parallel in reactions that directly mimic the successes identified during the exploratory stage. The work presented was based upon a series of exploitation reactions designed to mimic the conditions used in the preparation of $[\text{C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ and $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$.⁴⁰

A dataset of 75 historical reactions was compiled from past entries found in our laboratory notebooks. This dataset, provided in the Electronic supplementary information, contains a range of information including reactant quantities (amounts of vanadium, selenium, the amine and solvent water), reaction conditions (temperature, initial pH), and calculated geometric and property descriptors for the amine present. The initial set of 24 descriptors was selected from variables identified as important in our past work on templated vanadium selenites.^{18, 41} Inorganic descriptors are focused on the vanadium source, where both NH_4^+ and Na^+ are possible counter ions. Selenium dioxide was used in every reaction, and so no selenium source descriptors are included. The amine descriptors include structural (C:N ratio, chain length, molecular weight, nitrogen count, presence of primary and/or secondary ammonium sites, cyclic and/or spherical structures (bicycles), and phase at 25 °C (solid or liquid)), acidity (pK_a values), and inverse-charge density matching^{29, 31, 42-44} (maximal projection area / N) parameters. Note that inverse charge density was used to maximize the separation between values for

different amines. The amine areas were calculated using ChemAxon.²⁸ Finally, the last two columns describe the reaction outcome. The outcome field describes if an individual reaction resulted in a $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework structure, either type-1 or type-2, vide infra. A purity field is used to denote if a reaction resulted in a single solid phase or multiple phases. A full list of these descriptors and their respective definitions is provided in the Electronic supplementary information.

The set of historical reactions described above was compiled to exploit past successes in the formation of new organically templated vanadium selenites with framework structures. A ‘pruned’ decision tree⁴¹ was created to identify the most important descriptors in the historical dataset, essentially acting as a feature selector, and indicating the boundaries for the ‘-1’ and ‘+1’ levels for a fractional factorial design. This decision tree is shown in Figure 1. Of the 6 features identified in the decision tree, 5 were selected for inclusion in a fractional factorial design. Two features identified in the decision tree, minimum pK_a and pK_a 1, were merged and included as minimum pK_a 1. To capture composition effects not contained in our historical dataset, we augmented this with four stoichiometric descriptors (moles of Se, moles of V, moles of amine, mass of water) and a vanadium oxidation state descriptor (VOSO_4 , the only V^{4+} source used in these experiments). The resulting set of 10 descriptors was used to generate 128 individual reactions by a $2^{(10-3)}$ resolution V fractional factorial design. See Table 2. The purpose of this fractional factorial design study was to most effectively explore chemical space. Such a study enables one to systematically and efficiently probe the effects of many variables in complex systems, and allows for the avoidance of a “one-variable-at-a-time” approach. This experimental design is not intended to find ‘correct’ or ‘incorrect’ outcomes, but rather to generate a uniform dataset for generating subsequent machine learning models. Its function is to create unique

experiments that uniformly sample the impact of individual reaction parameters for each amine. Moreover, the identification of phase stability boundaries is essential in the determination of structural adaptability ranges. All 128 individual reactions generated in this fractional factorial analysis were conducted and their respective products were determined. This reaction set constitutes a more effective and systematic exploration of the chemical space under consideration. The desired framework was produced in 27 of the 128 individual reactions performed in this fractional factorial analysis. In 8 of these ‘successful’ reactions, the amines decomposed and were no longer present in the final compound, these amines were 1-(2-aminoethyl)piperazine and tris(2-aminoethyl)amine. As such, these amines are denoted as unable to be found in the desired framework. The role of decomposition in this study is discussed below.

The dataset containing the 75 historical reactions was augmented with the 128 reactions from the fractional factorial analysis. A decision tree created from this larger dataset is shown in Figure S5 of the Electronic Supplementary Information. The two decision trees share strong commonalities, with the presence of secondary ammonium sites, cyclic amine structure and amine pK_a values being present in both trees. Additionally, the observation of reaction stoichiometry, specifically the number of moles of the amine, in the second tree validates the inclusion of these descriptors in the fractional factorial analysis design.

Five new compounds were formed during this study. The inorganic components in compounds **1** – **5** are constructed from similar primary building units. $[VO_6]$ and $[VO_5]$ coordination polyhedra are observed. Both V^{4+} and V^{5+} vanadium centers are found in the $[VO_6]$ octahedra, while the $[VO_5]$ polyhedra contain V^{4+} exclusively. The vanadium oxide bonds in the $[V^{5+}O_6]$ octahedra are generally shorter than the corresponding bonds in the $[V^{4+}O_6]$ octahedra,

with the following observed ranges (Å): $V^{5+} - O_{\text{terminal}}$ 1.590(3) – 1.6254(13), $V^{5+} - O_{\text{bridging}}$ 1.6700(13) – 2.2418(13), $V^{4+} - O_{\text{terminal}}$ 1.5982(14) – 1.6159(14), $V^{4+} - O_{\text{bridging}}$ 1.9686(13) – 2.248(3). All selenium sites in these compounds are 4+ and exist in trigonal pyramidal coordination geometries with a stereoactive lone pairs. Compounds **1** – **4** contain both unprotonated $[\text{SeO}_3]$ and protonated $[\text{HSeO}_3]$ moieties, while **5** is comprised of $[\text{SeO}_3]$ alone. The $\text{Se} - O_{\text{bridging}}$ bonds range between 1.648(3) and 1.769(3) Å, while the $\text{Se} - \text{O}(\text{H})$ bonds range between 1.746(3) and 1.7919(14) Å. Protonated organic amines are found in each compound, creating extensive hydrogen-bonding networks.

The inorganic components in compounds **1** – **3** exhibit the same connectivity. See Figure 2. These structures are constructed from the same one-dimensional ladder secondary building unit (SBU), containing $[\text{VO}_6]$ and $[\text{SeO}_3]$ groups. This SBU has been observed in other vanadium selenites^{34, 35, 37, 39, 40, 45} and zinc phosphates.⁴⁶ Adjacent chains are connected through bridging $[\text{HSeO}_3]$ moieties, resulting in the framework connectivity shown in Figure 2. Note that a distinct vanadium selenite framework has already been reported,⁴⁰ which is similar but not identical to the one reported here. The framework designated here as type-1⁴⁰ is observed in $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[\text{C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[(S)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ and $[(R)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, while the type-2 variation is observed in compounds **1** – **3**. The SBUs in these two framework types are identical, as are the connections between one-dimensional SBUs. However, the hydrogen-bonding interactions involving pendant $[\text{HSeO}_3]$ donor groups differ. The hydrogen bond acceptors in the type-1 variation are located within the central SBU backbone, while the acceptors in the type-2 variation are exclusively non-framework occluded water molecules. This results in different $[\text{HSeO}_3]$ orientations and channel metrics. See Figure 3. The channels in the

type-1 framework are nearly cylindrical while those in the type-2 framework are better described as elliptical cylinders.

The data generated in the fractional factorial design were used to better understand the formation of type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework structures. This framework type was specifically chosen for two reasons. First, this framework connectivity (both type-1 and type-2) was unreported before our initial exploratory reactions were conducted, making this study maximally useful as it did not just replicate a current understanding of a well explored system. Second, a range of vanadium charge ordering states are observed in compounds **1** – **3**, suggesting the possibility of interesting physical properties. In order to focus the fractional factorial design study on the formation of type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework structures, results were encoded on an amine specific level. Experimental results were collated for each of the 23 individual amines. An amine was tagged as ‘successful’ if it was *ever* found to be present in the target framework. In contrast, ‘unsuccessful’ amines were those that were *never* found in the target structure in any individual reaction.

Three different amines were used in reactions that resulted in the desired type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework; 2-piperazinoethylamine (aep), 1,4-bis(3-aminopropyl)piperazine (bapp) and tris(2-aminoethyl)amine (tren). These amines, however, were tagged as ‘unsuccessful’ because the resulting compounds did not contain the initial amines. Instead, the amines decomposed during reaction and the resulting products contained distinctly different organic structures (diethyltriamine (deta) in the case of aep and tren, or 1,2-bis(3-aminopropylamino)ethane (bape) in the case of bapp). The amines in question and their decomposition products are listed in Table 3. The decomposition of these amines precludes marking them as successful because the type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework was never observed to contain them in their original

form. Additionally, three amines were found to be crystallographically disordered and two amines resulted in twinned crystals. The structures containing triethylenetetramine (teta), tetraethylenepentamine (tepa) and pentaethylenehexamine (peha) contained well-resolved inorganic frameworks with crystallographically disordered amines. The structures containing ethylenediamine (en) and N-methylethylenediamine (n-meda) were both badly twinned. While identification of both the inorganic and organic structures was possible, the twinning present in each crystal precluded a stable refinement. See Table 3. Nevertheless, en, n-meda, tetra, tepa, and peha are considered successful amines because they were indeed present in the type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework, despite crystallographic complications.

To elucidate the dependence of these different framework types on the amine properties, we constructed models based on physicochemical descriptors describing the hierarchy of influences that we have proposed to most strongly affect the nature of reaction products in this family of compounds. These influences include reactant concentrations,^{40, 47-55} charge density matching,^{40, 42-44, 54, 56-60} hydrogen-bonding,^{31, 40, 41, 61, 62} sterics,^{30, 60} and symmetry.^{63, 64} This hierarchy is based upon postulated formation mechanisms for organically templated metal oxides.^{42, 43, 65} As the nature of this study is amine-centric, our descriptor set is focused on the amine properties found in our hierarchy. Descriptors for amine shape (surface areas, volume, projection sizes), charge and charge distribution (pK_a , polar surface area), and hydrogen-bonding (number of N-H bonds) have been included. See Table 4. This complete dataset is available in the Electronic supplementary information. The decision tree generated with these data, shown in Figure 4, selects two features (minimum projection distance and polar surface area) as the most informative of the reaction outcome.

Plotting the reaction outcome as a function of both amine polar surface area and minimum projection distance reveals an interesting dependence. See Figure 5. Note that solid symbols in Figure 5 represent the initial 23 amines. The dark blue squares represent the amines that can form type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ frameworks. The collection of amines that result in type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework structures can be described as having high polar surface areas for their minimum projection distance. In short, this means that the amines are either small (low projection distances), or linear with a larger number of ammonium sites (higher polar surface areas). As the geometry of the channels in type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ frameworks is quite constrained, this region of phase stability makes sense. Amines that either cannot fit into a narrow channel or exhibit low charge densities^{29, 31, 42-44} are unable to result in the desired structure type. Note that 2-methylpiperazine (2-mpip) and 2,5-dimethylpiperazine (2,5-dmpip) both exist in type-1 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ frameworks. However, their polar surface areas ($33.2 \text{ \AA}^2$) and minimum projection distances ($6.5 - 6.7 \text{ \AA}$) do not allow for these amines to fit in type-2 channels. These amines are marked by cyan squares in Figure 5. The channels are larger in a type-1 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework, which enables the formation with 2-mpip or 2,5-dmpip.

The influence of charge density matching^{29, 31, 42-44} is also observed in the selection of these two descriptors. Charge density matching in a geometrically constrained three-dimensional framework that contains a channel structure is effectively reduced to the linear channel charge density (positive charge per unit length), which must balance the charge on the inorganic framework. As such, the combination of polar surface area (amount of the molecule that is charged) and minimum projection distance (the footprint of the molecule) suggests that charge density matching remains important in this study.

The outcome of the factorial experiment—comprising 23 different organic amines—was used to create the decision tree shown in Figure 4. Amine selection for the exploratory synthetic efforts for the formation of new templated metal oxides is generally based upon several criteria, including commercial availability, cost, stability and past history of success in related chemical systems. As such, a small set of organic amines is disproportionately represented in the literature.¹⁸ In an attempt to avoid the introduction of historical bias in this dataset and analysis, an additional 19 amines were chosen randomly from our Dark Reactions database¹⁸ and used in a supplemental set of 32 reactions. The results of these reactions were plotted alongside the original amines. See Figure 5. These data are consistent with the original hypothesis concerning the structural adaptability ranges for type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework. The accuracy of the decision tree model, shown in Figure 4, was calculated using the additional 19 amines as a test set. Of these 19 amines, 15 true negatives were observed (both predicted and observed to not form the framework in question) and 4 false positives (predicted to form the framework, but observed to be unable) were found, resulting in an accuracy of 78.9 %. These points are shown in Figure 5 as open symbols. The 4 false positive are clustered at low minimum projection distances, between 5.5 and 6 Å, where the model had no prior data. A new decision tree model using the full data set of 42 amines was generated, shown in Figure 6. Once again, the polar surface area and minimum projection distance appear as nodes in the decision tree, but inclusion of the 4 formerly false positive amines refines the boundary chosen for the minimum projection distance (as shown in Figure 5).

The work described above details the exploitation of the structural adaptability of type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework, in that the generalized amine properties that enable or preclude framework formation are identified and described. An additional electronic adaptability also

exists in this $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework system (considering both type-1 and type-2 analogs), from which the bounds of charge density matching can be determined for this system. All vanadium centers in the type-1 framework compounds exist in the +4 oxidation state, while each type-2 framework contains both V^{4+} and V^{5+} . The vanadium oxidation states in these compounds were determined using bond valence sums calculations.^{26, 27} See Table 2. Full tables of calculated bond valence sums for compounds **1** – **5** are provided in the Electronic supplementary information. The fraction of vanadium sites in the V^{4+} state in compounds **1**, **2**, and **3** are 0.667, 0.667 and 0.75. As noted above, the location of V^{4+} and V^{5+} sites can be determined readily from experimental bond distances using bond valence sums. Moreover, the colors of these materials support the assigned vanadium oxidation states. The type-1 framework compounds; $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[\text{C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[(S)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ and $[(R)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, contain V^{4+} exclusively and exist as light blue crystals, owing to the d^1 configuration on the V^{4+} sites. Compounds **1** – **3**, in contrast are all very dark green, the result of an intervalence charge transfer band between adjacent V^{4+} and V^{5+} sites.

The vanadium charge ordering schemes for compounds **1** – **3**, $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$,⁴⁰ $[\text{C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$,⁴⁰ $[(S)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ ⁴⁰ and $[(R)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ ⁴⁰ are shown in Figure 7. Green and orange octahedra represent V^{4+} and V^{5+} sites respectively. The variability in vanadium oxidation state concentration and location suggests a compensation mechanism to achieve charge density matching in these compounds that does not affect the connectivity within the framework. The amount of positive charge provided by the organic amines is dictated by length and protonation state. Additionally, the distance between amines is

controlled largely by sterics. As such, the amount of positive charge per Å in the framework channels varies between compounds and needs to be matched by the negative charge on the framework.

The amines found in $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[\text{C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$, $[(S)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ and $[(R)\text{-C}_5\text{H}_{14}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$ (2,5-dimethylpiperazine and either racemic or enantiomerically pure 2-methylpiperazine) are largely cyclic and compact. As such, they are able to pack closely, resulting in higher channel charge densities. In contrast, the linear amines in **1** – **3** are far less compact. The channel charge densities in these compounds are significantly lower. These related frameworks are able to compensate for differences in channel charge density through the inclusion of both V^{4+} and V^{5+} . The relationship between fraction of V^{4+} sites and channel charge density is essentially linear. See Figure 8. The electronic adaptability of this framework is complementary to the connectivity component of the structural adaptability described above.

The negative charge on the $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework and the positive charge in the respective channels must perfectly balance. Small, highly charged amines such as 2-mpip and 2,5-dmpip can pack together closely, resulting in a higher positive charge density on the channels. In contrast, longer, lower charged amines such as dmed result in lower positive charge density channels. These channel charge densities could conceivably exist over a wide range, from small cations such as NH_4^+ to quite long amines with extended carbon backbones. However, the framework charge is more stringently bounded. Both V^{4+} and V^{5+} centers can be incorporated into these structures, but V^{3+} sites would require more substantial structural rearrangement, and of course V^{6+} is not chemically accessible. Therefore, an all- V^{4+} framework

minimizes the positive charge on the vanadium sites and results in a higher negative framework charge. The introduction of V^{5+} then lowers the amount of negative charge on the framework, and an all- V^{5+} framework would represent the lower bound on framework charge. Channel charge densities can neither exceed that of an all- V^{4+} framework nor fall below an all- V^{5+} framework. Exceptionally small cations such as NH_4^+ could possibly be incorporated if the inclusion of solvent allowed for the cations to be properly spaced without the inclusion of voids. The opposite, however, is not possible. Organic cations with exceptionally low charge densities could never be incorporated because mechanisms to increase their charge density do not exist.

The utility of our systematic exploration of chemical space described is twofold. First, it was used to successfully elucidate the structural adaptability ranges for the vanadium framework type in question. Second, such a systematic exploration is likely to also result in the formation of previously unknown compounds because chemical space is more evenly investigated. Two such compounds were discovered in this exploration using the amine piperazine. Compounds **4** and **5** are quite different in terms of composition and structure, a reflection of the diverse reagent choices and experimental parameters that gave rise to their formation.

The inorganic framework in $[C_4H_{12}N_2][(VO)_3(SeO_3)(HSeO_3)_4] \cdot H_2O$ (**4**) is constructed from $[VO_5]$, $[VO_6]$, $[HSeO_3]$ and $[SeO_3]$ primary building units. All vanadium sites in **4** exist in the +4 oxidation state. See Figure 9. $[VO(SeO_3)_2(HSeO_3)]$ chains extend along the $[0\ 0\ 1]$ direction. Such chain connectivities have been observed in other metal selenites.^{45, 66, 67} These chains are connected by $[VO_5]$ bridging units in two directions. The resulting framework is related to $[C_5H_{14}N_2][(VO)_3(SeO_3)_2(HSeO_3)_4]$, $[(R)-C_5H_{14}N_2][(VO)_3(SeO_3)_2(HSeO_3)_4]$ and $[(S)-C_5H_{14}N_2][(VO)_3(SeO_3)_2(HSeO_3)_4]$.⁴⁵ The orientations of the $[VO_5]$ coordination polyhedra differ between these reported structures and compound **4**, in that the alternation in their

respective orientations is observed in **4** alone. A figure containing the $[\text{VO}(\text{SeO}_3)_2(\text{HSeO}_3)]$ chains and the $[\text{VO}_5]$ orientations is available in the Electronic supplementary information.

$[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ (**5**) contains both $[(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ layers and protonated piperazinium dications. The lone vanadium site in compound **5** exists in the +5 oxidation state. The $[(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ layers are constructed from $[\text{V}_2\text{O}_{10}]$ dimers and bridging $[\text{SeO}_3]$ groups. See Figure 10. The connectivity within the $[(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ layers has been observed previously in $[\text{C}_6\text{H}_{14}\text{N}_2][\text{V}_2\text{O}_4(\text{SeO}_3)_2] \cdot 1.25\text{H}_2\text{O}$,³⁸ however the orientations of the bridging $[\text{SeO}_3]$ groups differ distinctly. The piperazinium dications reside between inorganic layers, forming an extensive hydrogen-bonding network.

The reaction mixtures from which these materials were formed dictate their compositions and structures. $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_3(\text{SeO}_3)(\text{HSeO}_3)_4] \cdot \text{H}_2\text{O}$ (**4**) and $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ (**5**) were both synthesized using piperazine as the amine, using vanadium sources. Compound **4** was synthesized using VO_2SO_4 , while compound **5** was formed from NH_4VO_3 . The vanadium oxidation states in the products mirror those in their respective vanadium sources, +4 or +5. Such behavior is consistent with a hypothesis regarding the role of piperazine and piperazine-like amines in the formation of templated vanadium selenites.¹⁸ Moreover, compounds **4** and **5** were synthesized from reaction mixtures with approximate reactant ratios of 1 pip : 6 V : 1 Se and 1 pip : 10 V : 9 Se. The increased amine concentration in the reaction mixture from which **5** was formed is reflected in this compounds stoichiometry with respect to **4**. Compound **5** is amine rich, with composition of 1 pip : 2 V : 2 Se, in contrast to the amine deficient **4**, whose composition is 1 pip : 3 V : 5 Se. The combination of vanadium oxidation state and relative amine concentration is responsible for the differences observed between compounds **4** and **5**.

Conclusions

The structural and electronic adaptability ranges for a specific vanadium selenite framework were determined using a combined cheminformatics and machine learning approach. Historical reaction data were coupled with a fractional factorial design to efficiently explore the effects of amine structure and reactivity. This approach led to the identification of structural and electronic adaptability ranges in the target structure type, elucidation of the physiochemical properties most responsible for these adaptability ranges and the formation of five novel materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge support from the NSF (Award No. DMR-1307801 and DMR-1709351), A.N. and J.S. acknowledge the Henry Dreyfus Teacher-Scholar Awards Program. M.Z. acknowledges support for the purchase of a diffractometer from the NSF grant 1337296, the Ohio Board of Regents grant CAP-491 and from Youngstown State University.

Electronic Supplementary Information

Crystallographic data for the structure reported in this paper have been deposited with the Cambridge Crystallographic Data Centre as supplementary publication numbers 1588069 – 1588073. Packing figures for $[\text{C}_4\text{H}_{14}\text{N}_2]_2[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_6 \cdot 5\text{H}_2\text{O}$ (1) and $[\text{C}_8\text{H}_{26}\text{N}_4][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_6 \cdot 6\text{H}_2\text{O}$ (2), a structural figure for

$[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_3(\text{SeO}_3)(\text{HSeO}_3)_4]\cdot\text{H}_2\text{O}$ (**4**), tables of descriptors, amine specifications, reaction outcomes, and bond valence sums are available.

Table 1. Crystallographic data for compounds **1** – **5**.

Compound	[C ₄ H ₁₄ N ₂] ₂ [VO(SeO ₃)(HSeO ₃)] ₆ ·5H ₂ O (1)	[C ₈ H ₂₆ N ₄][VO(SeO ₃)(HSeO ₃)] ₆ ·6H ₂ O (2)	[C ₄ H ₁₆ N ₂] ₂ [VO(SeO ₃)(HSeO ₃)] ₈ ·9.333H ₂ O (3)	[C ₄ H ₁₂ N ₂][(VO) ₃ (SeO ₃)(HSeO ₃) ₄]·H ₂ O (4)	[C ₄ H ₁₂ N ₂][(VO) ₂ O ₂ (SeO ₃) ₂] (5)
Formula	C ₈ H ₄₃ N ₄ O ₄₇ Se ₁₂ V ₆	C ₈ H ₄₄ N ₄ O ₄₈ Se ₁₂ V ₆	C ₈ H _{58.67} N ₆ O _{65.33} Se ₁₆ V ₈	C ₄ H ₁₈ N ₂ O ₂₂ Se ₆ V ₃	C ₄ H ₁₂ N ₂ O ₁₀ Se ₂ V ₂
fw	2200.60	2217.60	2955.43	1072.77	507.95
Space-Group	<i>P</i> 2 ₁ / <i>c</i> (No. 14)	<i>C</i> 2/ <i>c</i> (No. 15)	<i>P</i> 2 ₁ / <i>c</i> (No. 14)	<i>P</i> -1 (No. 2)	<i>P</i> 2 ₁ / <i>c</i> (No. 14)
<i>a</i> / Å	24.713(2)	18.816(5)	17.9990(7)	10.0027(6)	10.0879(4)
<i>b</i> / Å	7.9212(8)	7.934(5)	15.6580(6)	10.4765(6)	6.2047(2)
<i>c</i> / Å	34.399(2)	34.439(5)	26.0344(10)	12.2924(7)	10.6765(4)
α / °	90	90	90	74.478(3)	90
β / °	130.376(4)	95.850(5)	113.803	72.643(3)	116.5657(17)
γ / °	90	90	90	79.707(3)	90
<i>V</i> / Å ³	5129.8(9)	5114(4)	6713.1(5)	1177.80(7)	597.71(2)
<i>Z</i>	4	4	4	2	2
ρ_{calc} / g cm ⁻³	2.849	20880	2.924	3.025	2.822
λ / Å	0.71073	0.71073	0.71073	0.71073	0.71073
<i>T</i> / K	100(1)	100(1)	100(1)	100(1)	100(1)
μ / mm ⁻¹	9.688	9.720	9.875	10.540	7.709
<i>R</i> ₁ ^a	0.0375	0.0508	0.0423	0.0180	0.0190
<i>wR</i> ₂ ^b	0.0903	0.1167	0.1134	0.0396	0.0521

$$^a R_1 = \Sigma \|F_o\| - \|F_c\| / \Sigma \|F_o\| . \quad ^b wR_2 = [\Sigma w(F_o^2 - F_c^2)^2 / [\Sigma w(F_o^2)^2]]^{1/2}.$$

Table 2. Fractional factorial design variables and values.

Variable	Name	-1 level	+1 level
A	min pK _a 1 value	> 5.18	≤ 5.18
B	contains a secondary ammonium site	no	yes
C	contains a ring	no	yes
D	NaVO ₃ was used	no	yes
E	NH ₄ VO ₃ was used	no	yes
F	VOSO ₄ was used	no	yes
G	number of moles of Se	0.005	0.010
H	number of moles of the amine	0.0005	0.001
J	number of moles of V	0.0004	0.0008
K	mass of water	3	6

Table 3. Summary of amine decomposition, disorder and twinning occurrence.

Initial amine	Observed organic structure	Note
aep	deta	decomposition
bapp	bape	decomposition
tren	deta	decomposition
teta	teta	disorder
tepa	tepa	disorder
peha	peha	disorder
en	en	twinning
n-meda	n-meda	twinning

Table 4. Organic molecule descriptors

Name	Description
amine_abbreviation	abbreviation for the amine used in the reaction
mw	molecular weight (g / mol)
polar_surface_area	Three-dimensional polar surface area, computed using van der Waals radii of all nitrogen and oxygen atoms ($\text{\AA}^2$)
3D_surface_area	Three-dimensional molecular surface area, computed using van der Waals radii of the atoms ($\text{\AA}^2$)
van_der_Waals_volume	Three-dimensional volume, computed using van der Waals radii of the atoms ($\text{\AA}^3$)
max_proj_d	maximum value of the projection distance ($\text{\AA}$)
max_proj_A	maximum value of the projection area ($\text{\AA}^2$)
min_proj_d	minimum value of the projection distance ($\text{\AA}$)
min_proj_A	minimum value of the projection area ($\text{\AA}^2$)
pKa1	amine pK_a 1
pKa2	amine pK_a 2
pKa3	amine pK_a 3, if applicable
pKa4	amine pK_a 4, if applicable
min_pKa	minimum pK_a value
max_N-H_bonds	total number N-H bonds

Table 5. Framework charge density data.

Compound	Fraction V^{4+}	Amine – amine distance (Å)	Channel charge (e^-)	Channel charge density ($e^- / \text{\AA}$)
$[C_4H_{14}N_2]_2[VO(SeO_3)(HSeO_3)]_6 \cdot 5H_2O$ (1)	0.667	37.728	8	0.2120
$[C_8H_{26}N_4][VO(SeO_3)(HSeO_3)]_6 \cdot 6H_2O$ (2)	0.667	37.523	8	0.2132
$[C_4H_{16}N_2]_2[VO(SeO_3)(HSeO_3)]_8 \cdot 9.333H_2O$ (3)	0.75	24.970	6	0.2403
$[C_6H_{16}N_2][VO(SeO_3)(HSeO_3)]_2 \cdot 2H_2O^{40}$	1.0	12.716	4	0.3146
$[C_5H_{14}N_2][VO(SeO_3)(HSeO_3)]_2 \cdot 2H_2O^{40}$	1.0	12.667	4	0.3454
$[(S)-C_5H_{14}N_2][VO(SeO_3)(HSeO_3)]_2 \cdot 2H_2O^{40}$	1.0	12.674	4	0.3156
$[(R)-C_5H_{14}N_2][VO(SeO_3)(HSeO_3)]_2 \cdot 2H_2O^{40}$	1.0	12.682	4	0.3158

Figure captions.

Figure 1. Historical reaction decision tree. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

Figure 2. Three-dimensional packing of $[\text{C}_4\text{H}_{16}\text{N}_2]_2[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_8 \cdot 9.333\text{H}_2\text{O}$ (**3**). Green octahedra represent $[\text{VO}_6]$ while purple, red, blue, white and gray spheres represent selenium oxygen, nitrogen, carbon and hydrogen atoms, respectively. Selected hydrogen atoms have been removed for clarity.

Figure 3. Channel dimensions in the (a) type-1 and (b) type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework. . Selected distances are shown (Å). Green octahedra represent $[\text{VO}_6]$ while purple, red, and gray spheres represent selenium oxygen and hydrogen atoms, respectively.

Figure 4. Original $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ type-II layer formation decision tree. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

Figure 5. Plot of reaction outcome as a function of both amine polar surface area and minimum projection distance. Symbols for the original 23 amines are solid, while the symbols for the additional 19 amines are open. Blue and red lines represent Dashed lines represented polar surface area and minimum projection distance decision boundaries, respectively. Dashed lines correspond to the decision boundaries for the original 23 amines only, while dotted lines represent decision boundaries for the full set of amines.

Figure 6. $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ type-II layer formation decision tree, including the results from reactions involving all 42 amines. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

Figure 7. Charge ordering schemes in (a) $[\text{C}_4\text{H}_{14}\text{N}_2]_2[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_6 \cdot 5\text{H}_2\text{O}$ (**1**), (b) $[\text{C}_8\text{H}_{26}\text{N}_4][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_6 \cdot 6\text{H}_2\text{O}$ (**2**), (c) $[\text{C}_4\text{H}_{16}\text{N}_2]_2[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_8 \cdot 9.333\text{H}_2\text{O}$ (**3**) and (d) $[\text{C}_6\text{H}_{16}\text{N}_2][\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_2 \cdot 2\text{H}_2\text{O}$. Green and orange octahedra represent $[\text{V}^{4+}\text{O}_6]$ and $[\text{V}^{5+}\text{O}_6]$, respectively. Purple, red and gray spheres represent selenium, oxygen and hydrogen, respectively.

Figure 8. Plot of V^{4+} fraction versus linear channel charge density.

Figure 9. Three-dimensional packing of $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_3(\text{SeO}_3)(\text{HSeO}_3)_4] \cdot \text{H}_2\text{O}$ (**4**). Green polyhedra represent $[\text{VO}_6]$ and $[\text{VO}_5]$ while purple, red, blue, white and gray spheres represent selenium oxygen, nitrogen, carbon and hydrogen atoms, respectively. Selected hydrogen atoms have been removed for clarity.

Figure 10. (a) Layer connectivity and (b) three-dimensional packing of $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ (**5**). Green polyhedra represent $[\text{VO}_6]$ while purple, red, blue, white

and gray spheres represent selenium oxygen, nitrogen, carbon and hydrogen atoms, respectively. Hydrogen atoms have been removed for clarity.

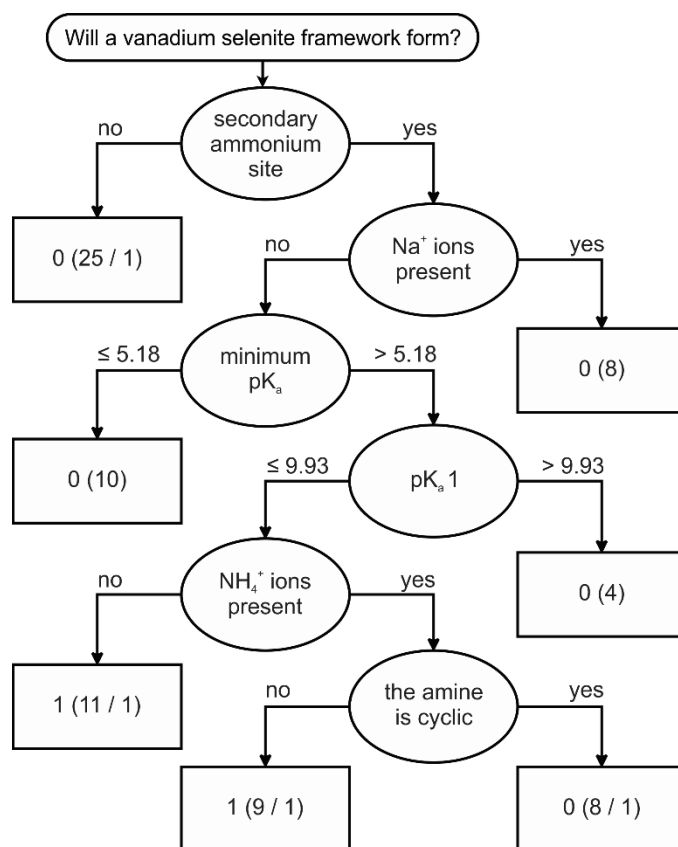


Figure 1. Historical reaction decision tree. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

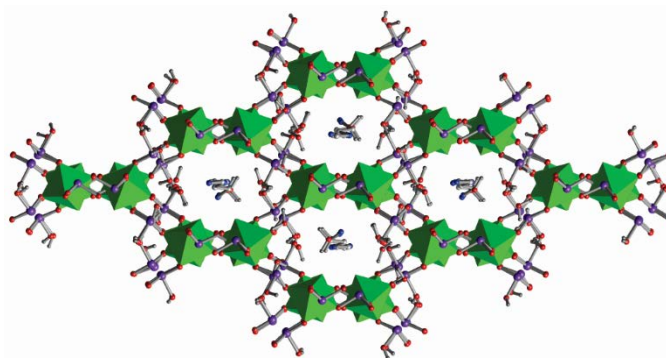


Figure 2. Three-dimensional packing of $[\text{C}_4\text{H}_{16}\text{N}_2]_2[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]_8 \cdot 9.333\text{H}_2\text{O}$ (**3**). Green octahedra represent $[\text{VO}_6]$ while purple, red, blue, white and gray spheres represent selenium, oxygen, nitrogen, carbon and hydrogen atoms, respectively. Selected hydrogen atoms have been removed for clarity.

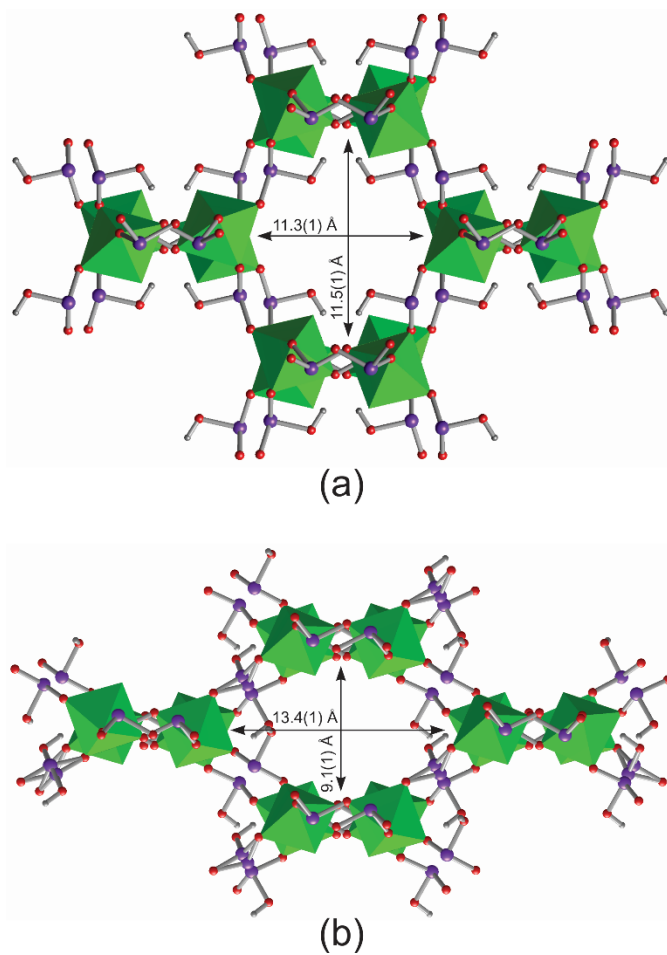


Figure 3. Channel dimensions in the (a) type-1 and (b) type-2 $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ framework. . Selected distances are shown (Å). Green octahedra represent $[\text{VO}_6]$ while purple, red, and gray spheres represent selenium oxygen and hydrogen atoms, respectively.

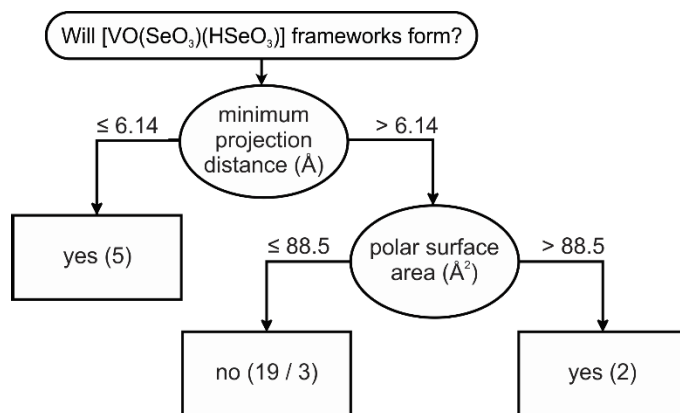


Figure 4. Original $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ type-II layer formation decision tree. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

Figure 5. Plot of reaction outcome as a function of both amine polar surface area and minimum projection distance for the original 23 amines. Vanadium selenite structure type identification for each reaction is provided in the legend. Dashed lines represented polar surface area (blue) and minimum projection distance (red) decision boundaries.

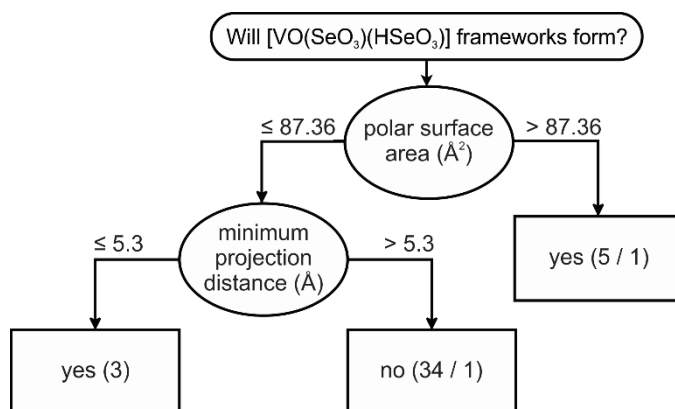


Figure 6. $[\text{VO}(\text{SeO}_3)(\text{HSeO}_3)]$ type-II layer formation decision tree, including the results from reactions involving all 42 amines. Each reaction bin contains a specific outcome value and number of reactions correctly and incorrectly assigned to that bin, respectively.

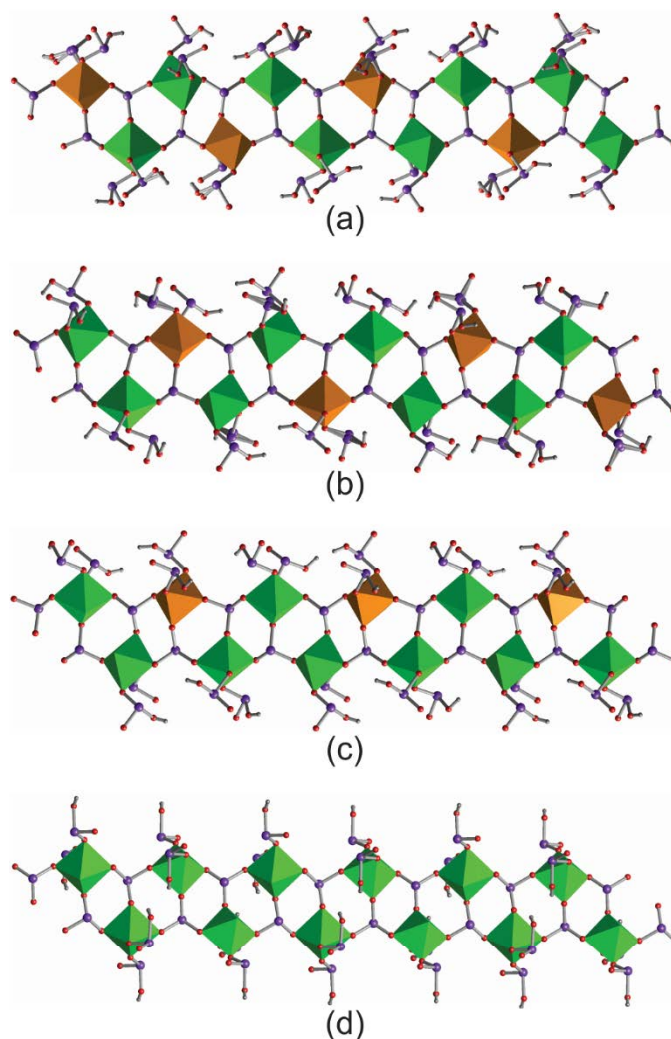


Figure 7. Charge ordering schemes in (a) $[C_4H_{14}N_2]_2[VO(SeO_3)(HSeO_3)]_6 \cdot 5H_2O$ (**1**), (b) $[C_8H_{26}N_4][VO(SeO_3)(HSeO_3)]_6 \cdot 6H_2O$ (**2**), (c) $[C_4H_{16}N_2]_2[VO(SeO_3)(HSeO_3)]_8 \cdot 9.333H_2O$ (**3**) and (d) $[C_6H_{16}N_2][VO(SeO_3)(HSeO_3)]_2 \cdot 2H_2O$. Green and orange octahedra represent $[V^{4+}O_6]$ and $[V^{5+}O_6]$, respectively. Purple, red and gray spheres represent selenium, oxygen and hydrogen, respectively.

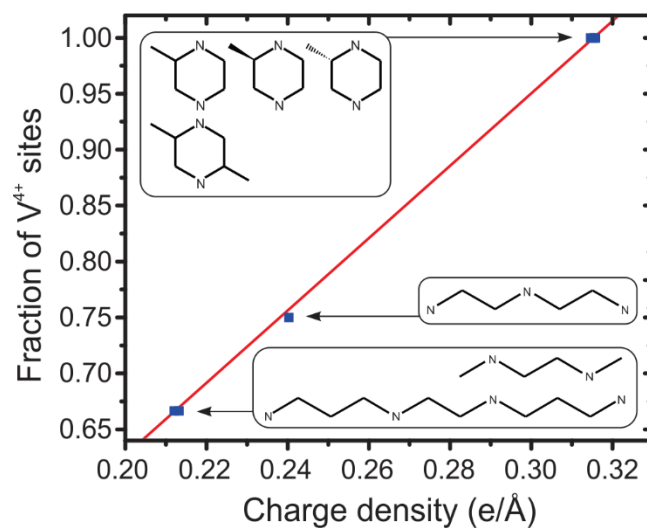


Figure 8. Plot of V^{4+} fraction versus linear channel charge density.

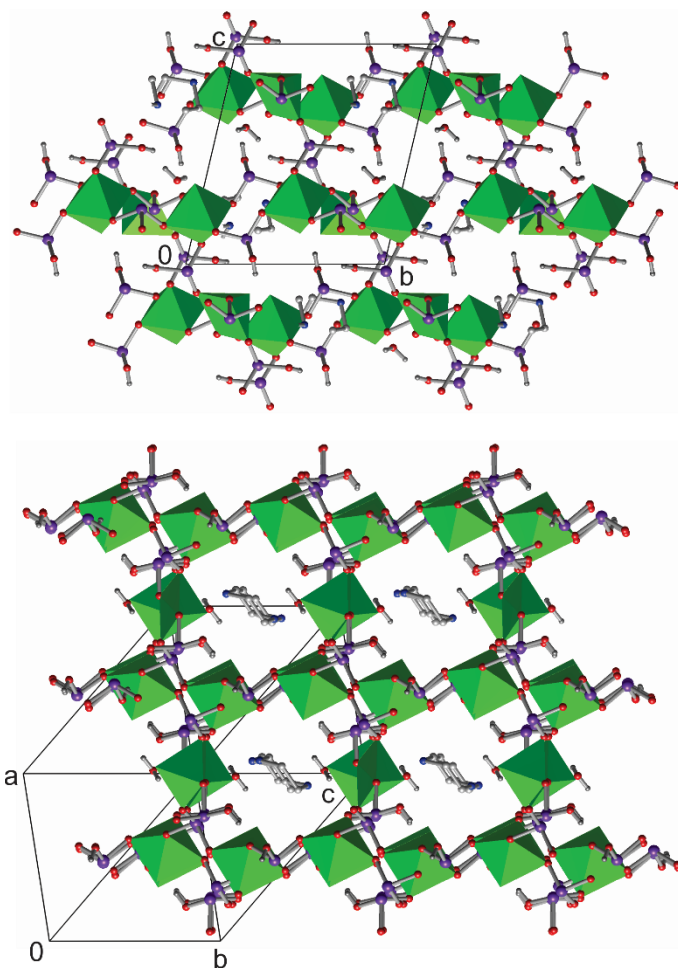


Figure 9. Three-dimensional packing of $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_3(\text{SeO}_3)(\text{HSeO}_3)_4]\cdot\text{H}_2\text{O}$ (**4**). Green polyhedra represent $[\text{VO}_6]$ and $[\text{VO}_5]$ while purple, red, blue, white and gray spheres represent selenium oxygen, nitrogen, carbon and hydrogen atoms, respectively. Selected hydrogen atoms have been removed for clarity.

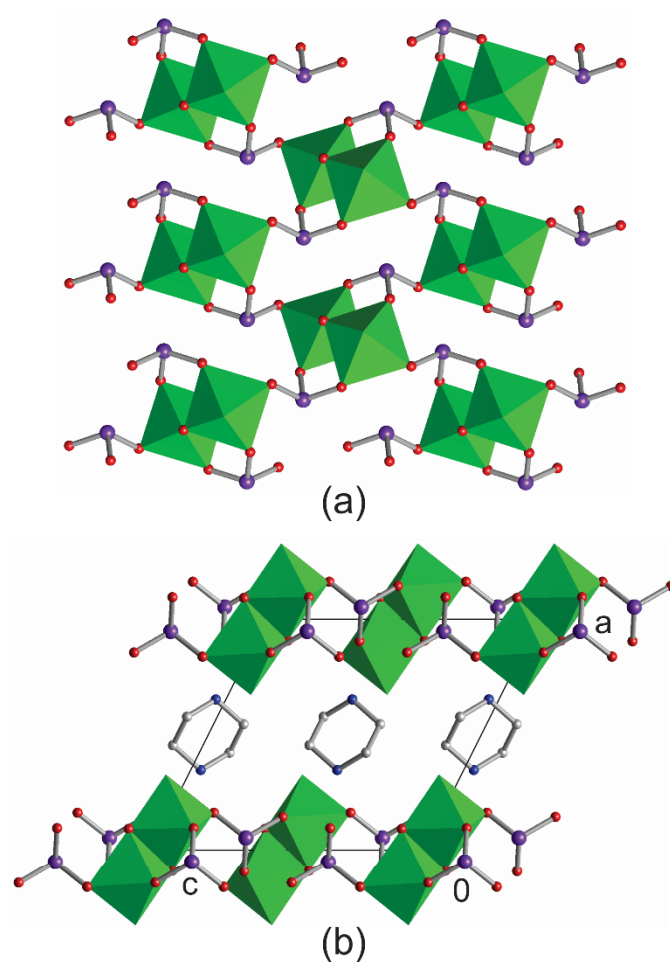
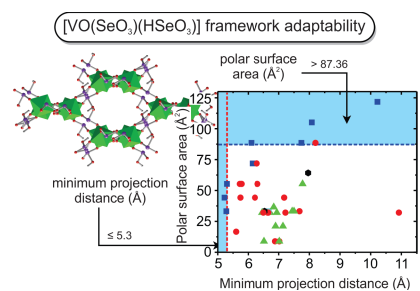


Figure 10. (a) Layer connectivity and (b) three-dimensional packing of $[\text{C}_4\text{H}_{12}\text{N}_2][(\text{VO})_2\text{O}_2(\text{SeO}_3)_2]$ (5). Green polyhedra represent $[\text{VO}_6]$ while purple, red, blue, white and gray spheres represent selenium, oxygen, nitrogen, carbon and hydrogen atoms, respectively. Hydrogen atoms have been removed for clarity.

For Graphical Abstract use only



The structural and electronic adaptability of a vanadium selenite framework is determined using cheminformatics data and machine learning algorithms.

1. E. Winsberg, *Science in the age of computer simulation*, University of Chicago Press, Chicago, 2010.
2. J. P. Holdren, *Material Genome Initiative Strategic Plan*, National Science and Technology Council, 2014.
3. K. Rajan, *Annual Review of Materials Research*, 2015, 45, 153-169.
4. T. Lookman, F. J. Alexander and K. Rajan, Springer International Publishing, Switzerland, 2016.
5. Y. Liu, T. Zhao, W. Ju and S. Shi, *Journal of Materiomics*, 2017, 3, 159-177.
6. I. A. Rauf, *J Phys Chem Biophys*, 2015, 5, 1000e1129.
7. R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm and H. Lam, *ACS Combinatorial Science*, 2011, 13, 579-633.
8. M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. V. Duren and A. Zakutayev, *Applied Physics Reviews*, 2017, 4, 011105.
9. S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk and B. A. Grzybowski, *Angew. Chem. Int. Ed.*, 2016, 55, 5904-5937.
10. J. K. Bunn, R. Z. Voepel, Z. Wang, E. P. Gatzke, J. A. Lauterbach and J. R. Hattrick-Simpers, *Industrial & Engineering Chemistry Research*, 2016, 55, 1236-1242.
11. E. Popova, T. M. Rodgers, X. Gong, A. Cecen, J. D. Madison and S. R. Kalidindi, *Integrating Materials and Manufacturing Innovation*, 2017, 6, 54-68.
12. D. J. Griffin, M. A. Grover, Y. Kawajiri and R. W. Rousseau, *Industrial & Engineering Chemistry Research*, 2016, 55, 1361-1372.
13. S. K. Suram, J. A. Haber, J. Jin and J. M. Gregoire, *ACS Combinatorial Science*, 2015, 17, 224-233.
14. J. R. Hattrick-Simpers, J. M. Gregoire and A. G. Kusne, *APL Materials*, 2016, 4, 053211.
15. A. R. Akbarzadeh, V. Ozoliņš and C. Wolverton, *Adv. Mater.*, 2007, 19, 3233-3239.
16. W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson and G. Ceder, *Science Advances*, 2016, 2.
17. E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder and E. Olivetti, *Chem. Mater.*, 2017, 29, 9436-9444.
18. P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature (London, U. K.)*, 2016, 533, 73-76.
19. E. Kim, K. Huang, S. Jegelka and E. Olivetti, *npj Computational Materials*, 2017, 3, 53.
20. V. Duros, J. Grizou, W. Xuan, Z. Hosni, D.-L. Long, H. N. Miras and L. Cronin, *Angew. Chem. Int. Ed.*, 2017, 56, 10815-10820.
21. P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto and B. Maruyama, *Npj Computational Materials*, 2016, 2, 16031.
22. , Bruker AXS Inc., Madison (WI), USA, 2009.
23. A. Altomare, G. Cascarano, C. Giacovazzo and A. Guagliardi, *J. Appl. Crystallogr.*, 1993, 26, 343-350.
24. P. W. Betteridge, J. R. Carruthers, R. I. Cooper, K. Prout and D. J. Watkin, *J. Appl. Crystallogr.*, 2003, 36, 1487.
25. E. Dowty, Shape Software, TN, USA,, 2002.
26. I. D. Brown and D. Altermatt, *Acta Crystallogr. Sect. B*, 1985, 41, 244-247.
27. N. E. Brese and M. O'Keeffe, *Acta Crystallogr. Sect. B*, 1991, 47, 192-197.

28. , Marvin 14.10.20.0, ChemAxon (<http://www.chemaxon.com>), 2014 (last accessed Dec. 17, 2014).
29. K. B. Chang, D. J. Hubbard, M. Zeller, J. Schrier and A. J. Norquist, *Inorg. Chem.*, 2010, 49, 5167-5172.
30. E. C. Glor, S. M. Blau, J. Yeon, M. Zeller, P. Shiv Halasyamani, J. Schrier and A. J. Norquist, *J. Solid State Chem.*, 2011, 184, 1445-1450.
31. M. D. Smith, S. M. Blau, K. B. Chang, M. Zeller, J. Schrier and A. J. Norquist, *Cryst. Growth Des.*, 2011, 11, 4213-4219.
32. K. B. Chang, M. D. Smith, S. M. Blau, E. C. Glor, M. Zeller, J. Schrier and A. J. Norquist, *Cryst. Growth Des.*, 2013, 13, 2190-2197.
33. H. Nakano, T. Ozeki and A. Yagasaki, *Inorg. Chem.*, 2001, 40, 1816-1819.
34. Z. Lian, C. Huang, H. Zhang, H. Yang, Y. Zhang and X. Yang, *J. Chem. Crystallogr.*, 2004, 34, 489-494.
35. Z. Dai, Z. Shi, G. Li, X. Chen, X. Lu, Y. Xu and S. Feng, *J. Solid State Chem.*, 2003, 172, 205-211.
36. Z. Dai, G. Li, Z. Shi, W. Fu, W. Dong, J. Xu and S. Feng, *Solid State Sci.*, 2004, 6, 91-96.
37. I. Pasha, A. Choudhury and C. N. R. Rao, *Inorg. Chem.*, 2003, 42, 409-415.
38. Z. Lian, J. Zhang, Y. Gu, T. Wang and T. Lou, *J. Mol. Struct.*, 2009, 919, 122-127.
39. Z. Dai, G. Li, Z. Shi, X. Liu and S. Feng, *Inorg. Chem. Commun.*, 2005, 8, 890-893.
40. J. H. Olshansky, T. Thao Tran, K. J. Hernandez, M. Zeller, P. S. Halasyamani, J. Schrier and A. J. Norquist, *Inorg. Chem.*, 2012, 51, 11040-11048.
41. P. D. F. Adler, R. Xu, J. H. Olshansky, M. D. Smith, K. C. Elbert, Y. Yang, G. M. Ferrence, M. Zeller, J. Schrier and A. J. Norquist, *Polyhedron*, 2016, 114, 184-193.
42. G. Ferey, *J. Fluorine Chem.*, 1995, 72, 187-193.
43. G. Ferey, *Chem. Mater.*, 2001, 13, 3084-3098.
44. H. S. Casalongue, S. J. Choyke, A. Narducci Sarjeant, J. Schrier and A. J. Norquist, *J. Solid State Chem.*, 2009, 182, 1297-1303.
45. J. H. Olshansky, K. J. Wiener, M. D. Smith, A. Nourmahnad, M. J. Charles, M. Zeller, J. Schrier and A. J. Norquist, *Inorg. Chem.*, 2014, 53, 12027-12035.
46. A. A. Ayi, A. Choudhury, S. Natarajan, S. Neeraj and C. N. R. Rao, *J. Mater. Chem.*, 2001, 11, 1181-1191.
47. P. Halasyamani, M. J. Willis, C. L. Stern and K. R. Poeppelmeier, *Inorg. Chim. Acta*, 1995, 240, 109-115.
48. A. J. Norquist, K. R. Heier, C. L. Stern and K. R. Poeppelmeier, *Inorg. Chem.*, 1998, 37, 6495-6501.
49. P. M. Thomas, A. J. Norquist, M. B. Doran and D. O'Hare, *J. Mater. Chem.*, 2003, 13, 88-92.
50. T. R. Veltman, A. K. Stover, A. Narducci Sarjeant, K. M. Ok, P. S. Halasyamani and A. J. Norquist, *Inorg. Chem.*, 2006, 45, 5529-5537.
51. J. H. Nelson, A. R. Johnston, A. Narducci Sarjeant and A. J. Norquist, *Solid State Sci.*, 2007, 9, 472-484.
52. A. J. Norquist, M. B. Doran, P. M. Thomas and D. O'Hare, *Dalton Transactions*, 2003, 1168-1175.
53. A. K. Stover, J. R. Gutnick, A. Narducci Sarjeant and A. J. Norquist, *Inorg. Chem.*, 2007, 46, 4389-4391.

54. D. J. Hubbard, A. R. Johnston, H. Sanchez Casalongue, A. Narducci Sarjeant and A. J. Norquist, *Inorg. Chem.*, 2008, 47, 8518-8525.
55. S. J. Choyke, S. M. Blau, A. A. Lerner, A. Narducci Sarjeant, J. Yeon, P. S. Halasyamani and A. J. Norquist, *Inorg. Chem.*, 2009, 48, 11277-11282.
56. A. Monnier, F. Schuth, Q. Huo, D. Kumar, D. Margolese, R. S. Maxwell, G. D. Stucky, M. Krishnamurty, P. Petroff, A. Firouzi, M. Janicke and B. F. Chmelka, *Science*, 1993, 261, 1299-1303.
57. Q. Huo, D. I. Margolese, U. Ciesla, P. Feng, T. E. Gier, P. Sieger, R. Leon, P. M. Petroff, F. Schueth and G. D. Stucky, *Nature*, 1994, 368, 317-321.
58. S. H. Tolbert, C. C. Landry, G. D. Stucky, B. F. Chmelka, P. Norby, J. C. Hanson and A. Monnier, *Chem. Mater.*, 2001, 13, 2247-2256.
59. J. El Haskouri, M. Roca, S. Cabrera, J. Alamo, A. Beltran-Porter, D. Beltran-Porter, M. D. Marcos and P. Amoros, *Chem. Mater.*, 1999, 11, 1446-1454.
60. J. H. Koffer, J. H. Olshansky, M. D. Smith, K. J. Hernandez, M. Zeller, G. M. Ferrence, J. Schrier and A. J. Norquist, *Cryst. Growth Des.*, 2013, 13, 4504-4511.
61. J. H. Olshansky, S. M. Blau, M. Zeller, J. Schrier and A. J. Norquist, *Cryst. Growth Des.*, 2011, 11, 3065-3071.
62. A. Nourmahnad, M. D. Smith, M. Zeller, G. M. Ferrence, J. Schrier and A. J. Norquist, *Inorg. Chem.*, 2015, 54, 694-703.
63. J. R. Gutnick, E. A. Muller, A. Narducci Sarjeant and A. J. Norquist, *Inorg. Chem.*, 2004, 43, 6528-6530.
64. E. A. Muller, R. J. Cannon, A. Narducci Sarjeant, K. M. Ok, P. S. Halasyamani and A. J. Norquist, *Cryst. Growth Des.*, 2005, 5, 1913-1917.
65. C. N. R. Rao, S. Natarajan, A. Choudhury, S. Neeraj and A. A. Ayi, *Acc. Chem. Res.*, 2001, 34, 80-87.
66. B. Engelen, K. Boldt, K. Unterderweide and U. Baeumer, *Z. Anorg. Allg. Chem.*, 1995, 621, 331-339.
67. Z. Micka, I. Nemec, P. Vojtisek and J. Ondracek, *J. Solid State Chem.*, 1996, 122, 338-342.