

Rationalization of Synthesis-Structure Relationships of Nanoporous Materials using Aggregated SHAP Analysis

Elton Pan¹, Soonhyoung Kwon¹, Zach Jensen¹, Mingrou Xie¹, Rafael Gómez-Bombarelli¹, Manuel Moliner², Yuriy Roman¹, Elsa Olivetti¹

¹ Massachusetts Institute of Technology

² Instituto de Tecnología Química, Universitat Politècnica de València-Consejo Superior de Investigaciones Científicas
{eltonpan, s1105hk, zjensen, mrx, rafagb, yroman, elsao}@mit.edu, mmoliner@itq.upv.es

Abstract

Zeolites, crystalline aluminosilicate materials with well-defined porous structures, have emerged as versatile materials with applications in various fields, including catalysis and gas separation. Hydrothermal synthesis is a widely used method for zeolite production, offering control over crystallinity and pore size. However, the intricate interplay of synthesis parameters necessitates a comprehensive understanding to optimize the synthesis process. We train a supervised classification machine learning model on ZeoSyn (a dataset of zeolite synthesis routes) to predict the zeolite framework product given a synthesis route. Subsequently, we leverage SHapley Additive Explanations (SHAP) to reveal key synthesis-structure relationships in zeolites. To that end, we introduce an aggregation SHAP approach to extend such analysis to explain the formation of composite building units (CBUs) of zeolites. Analysis at this unprecedented scale sheds light on key synthesis parameters driving zeolite crystallization.

Introduction

Zeolites are nanoporous, crystalline aluminosilicate materials with a wide range of industrial applications including catalysis, separations, and ion exchange (Davis 2002). The crystalline structure and corresponding porous network is crucial in determining a zeolite's suitability for a target application (Weitkamp 2000). While thousands of potential zeolite structures are thought to be thermodynamically accessible (Popale, Cheeseman, and Deem 2011), only about 250 have been synthesized (bae 2021) highlighting a synthesis bottleneck to zeolite discovery and deployment. The synthesis of zeolites synthesis is intricate, with numerous variables influencing the resultant zeolite structure (Corma and Davis 2004). These factors include types and amounts of framework elements, the presence of inorganic and organic cations, structure-directing agents, mineralizing agents and hydrothermal conditions. (Davis 2002; Cundy and Cox 2005; Corma and Davis 2004; Lobo, Zones, and Davis 1995; Moliner, Rey, and Corma 2013)

Many studies have examined parts of the zeolite synthesis space including compositional gel ratios (Si/Al, Na/Si, OSDA/Si, H₂O/Si, etc) (Simon-Masseron et al. 2007; Blackwell et al. 2003; Shvets et al. 2010; Corma et al. 2006a; Mo-

XAI4Sci: Explainable machine learning for sciences, AAAI-24
(xai4sci.github.io)

liner et al. 2005), aging conditions (Ginter, Bell, and Radke 1992; Alfaro et al. 2007; Wu, Ren, and Wang 2008), crystallization conditions (Zhang et al. 2013a,b; Güray et al. 1999), and precursor selection (Li et al. 2018; Martín, Moliner, and Corma 2015; Kumar, Li, and Rimer 2016) for specific OSDA system (Csicsery 1984). However, knowledge of the holistic interplay between these factors across the entire field is lacking. Machine learning has the potential in generalizing some of these relationships(Corma et al. 2006b; Serra et al. 2007; Jensen et al. 2019; Muraoka et al. 2019) but have been limited to subsections of the zeolite design space due to a lack of data, which implies that larger data sets may generalize learning across the entire zeolite space.

We leverage ZeoSyn (Fig. S4), a comprehensive dataset of 23,925 zeolite synthesis routes for >200 unique zeolite frameworks (covering >80% of synthesized frameworks till date) and nearly 1000 unique OSDAs, which constitutes an order of magnitude larger than all previously published zeolite synthesis datasets. Each unique synthesis route in ZeoSyn is *comprehensive*, composing gel composition, reaction conditions, inorganic precursors, OSDAs, resulting zeolite structure, and select zeolite properties extracted from the scientific literature. To examine relationships between hydrothermal variables, OSDAs, and resulting zeolite structures, we train a supervised classification machine learning model on ZeoSyn to predict zeolite framework products given a synthesis route. We employ SHapley Additive Explanations (SHAP) to reveal key synthesis parameters driving the formation of over 200 zeolite frameworks and their constituent composite building units (CBUs), and show potential applications in phase-selective and intergrowth synthesis. Analysis at this unprecedented level of scale is a step toward an improved understanding of key synthesis parameters driving zeolite crystallization, which could potentially guide and accelerate the discovery of new zeolite frameworks.

Methods

Zeolite and OSDA featurization

The zeolite structural properties (eg. ring sizes, largest included sphere) are obtained from the IZA database (bae 2021). Zeolite frameworks are visualized using the 3dt software. Composite building unit visualizations are taken

from the IZA database (bae 2021). We featurize the OSDA using its physicochemical descriptors (eg. molecular volume and 2D shape descriptors) of the organic molecule (Schwalbe-Koda et al. 2021; Jensen et al. 2021). The full list of OSDA features and their descriptions can be found in Table S1.

Zeolite framework prediction model

We train a supervised classification model using random forest to predict a zeolite framework product given a synthesis recipe. The model takes in a 43-dimensional vector as input where each element corresponds to either gel composition (eg. Si, Al, P, etc), reaction condition (eg. crystallization time), or OSDA descriptor (eg. molecular volume). The model predicts (1 out of 220 classes) zeolite framework (three-letter code eg. **KFI** in Fig. 1a). A 80/20 random train/test split is employed. Since the focus is on the subsequent SHAP analysis, we train the model with default parameters.

SHAP analysis of zeolite formation

To analyze the outcomes of the classification model (depicted in Fig. 1a, we employ SHAP (Lundberg and Lee 2017), which is a generalized measure for the impact of features. This approach uses Shapley values from game theory to compute the contribution made by each feature to the model prediction. Features are likened to participants in a "game" representing the prediction task, and the SHAP values measure how much prediction is attributed to these features. These values signify the relative importance of a specific feature and its impact on classification. For example, as shown in Fig. 1c, SHAP values reveal how altering the value of a feature, either increasing or decreasing it, will affect the model output. We calculate SHAP values at two levels: **1) Framework-level SHAP** quantifies the impact of synthesis parameters on the formation of a zeolite framework, based on the predicted probabilities by the classifier. **2) CBU-level SHAP** quantifies the impact of synthesis parameters on the formation of a composite building unit (CBU). To obtain CBU-level SHAP values, we employ an aggregation approach as follows:

Aggregated SHAP Let $S_f \in \mathbb{R}^{n \times m}$ be the framework-level SHAP matrix of framework f with n observations and m features. The CBU-level SHAP matrix S_{CBU} is given by aggregating framework-level SHAP matrices:

$$S_{CBU} = \sum_{f \in F_{CBU}} S_f \quad (1)$$

where F_{CBU} is the set of synthesized frameworks containing a specific CBU according to the IZA database(bae 2021). For example, to obtain CBU-level SHAP matrix S_{sod} corresponding to the *sod* CBU, we determine the set of frameworks containing *sod*, $F_{sod} = \{FAU, SOD, LTA\}$ (Note: For the sake for brevity, only 3 *sod*-containing frameworks are listed as more exist). Subsequently, the CBU-level SHAP is given by $S_{sod} = S_{FAU} + S_{SOD} + S_{LTA}$. Intuitively, by summing up S_f corresponding to frameworks containing the CBU, this aggregation approach amplifies SHAP values corresponding to *common* features that highly impact

(positively or negatively) CBU formation, while suppressing SHAP values corresponding to the features that do not have much impact. As such, this effectively shifts the SHAP analysis from a framework-centric to a CBU-centric view, allowing for an understanding of factors driving the building units that make up zeolites.

Results and discussions

Zeolite framework prediction model

Given the high-dimensional space with synthesis parameters spanning from inorganic elements, organic templates, and reaction conditions, visualization in low dimension may not fully capture complex structure-synthesis relationships in zeolite crystallization. In data-driven synthesis planning, reaction product prediction is a crucial task in informing the outcome of a reaction (crystalline or amorphous phases) given a set of synthesis parameters.

Model performance The model is evaluated on held-out, unseen test syntheses on the framework prediction task, with a model accuracy of 0.73. This performance is surprisingly strong, as our model has the challenging task of predicting 1 out of 220 possible classes. This performance is comparable to the reported accuracy of 0.82 reported by (Muraoka et al. 2019) whose model predicts only 1 out of 23 (almost 10 times fewer classes than ours) under OSDA-free conditions (lower complexity of input).

The confusion matrix shown in Fig. 1b highlights classification performance on 1 representative small (**CHA**), medium (**MFI**), large (***BEA**), and extra-large (**UTL**) pore framework, with "Others" referring to all other frameworks aggregated together, and "Failed" referring to dense/amorphous phases. Evidently, the majority of the predictions lie along the diagonal of the matrix, meaning that the majority of predictions (0.68 – 0.88) are correct. As shown by the high-intensity off-diagonal elements, the majority of the errors made by the model are misclassifications as another framework ("Others") or dense/amorphous dense ("Failed"). The model also shows strong performance in discriminating different pore sizes as shown in Fig. S10 with high accuracies of 0.78 – 0.86.

Rationalization of synthesis-structure relationships in zeolites

The synthesis knowledge learned by the classification model can be analyzed to rationalize the impact of synthesis parameters on the formation of a specific zeolite framework. As such, SHapley Additive exPlanations (SHAP) analysis, a game-theoretic approach to explain the output of ML models through optimal credit allocation with local explanations (Lundberg and Lee 2017; Karpovich et al. 2023), is employed on the classification model to reveal synthesis-structure relationships in zeolites. For each prediction, SHAP values are calculated to determine the impact of each synthesis parameter on the probability of forming a specific zeolite framework (Fig. 1c). For instance, the first row uncovers a physically-grounded trend that low crystallization temperatures (blue points) have positive SHAP values (increases probability of **LTA** formation), while high

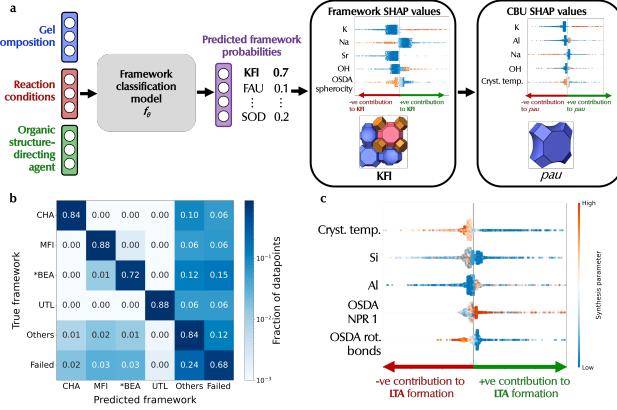


Figure 1: Interpretable ML framework for explaining synthesis-structure relationships in zeolites (a) Schematic of zeolite phase predictor model. Given synthesis parameters, the model f_θ predicts the resultant framework (eg. **KFI**). Additionally, if a dense or amorphous phase is expected, the model predicts a "Failed" class. The predicted framework probabilities are used to calculate framework-level SHAP values. In addition, CBU-level SHAP values of composite building units (CBUs) are obtained by aggregating framework-level SHAP values, allowing for CBU-level analysis of synthesis parameters. (b) Normalized confusion matrix of phase predictor model. Here, we have selected 1 representative small (**CHA**), medium (**MFI**), large (***BEA**), and extra-large pore (**UTL**) framework. "Others" refers to all other frameworks while "Failed" refers to amorphous/dense phases. (c) An example of a framework-level SHAP analysis quantifying the positive/negative impact of synthesis parameters on the probability of **LTA** framework formation.

temperatures (orange points) have negative SHAP values (decreases probability of **LTA** formation). This would agree with the fact that **LTA** is a small-pore zeolite with a relatively low framework density hence requiring low crystallization temperatures. We quantify the impact of synthesis parameters at two different levels of zeolite structure:

1. **Framework-level SHAP** shows the positive/negative impact of a synthesis parameter on the probability of crystallizing a specific zeolite framework (eg. **KFI** in Fig. 1a)
2. **CBU-level SHAP** shows the positive/negative impact of a synthesis parameter on the probability of forming a structure that contains a specific composite building unit (CBU) (eg. **pau** cage in Fig. 1a)

Framework-level SHAP Framework-level SHAP identifies the most important synthesis parameters driving the formation of a specific zeolite framework. Larger positive/negative SHAP values correspond to larger positive/negative changes in the probability of obtaining a specific framework given the synthesis parameter. Here, we consider all 43 inputs into the model f_θ and show only the top 10 most important synthesis parameters (in descending order) for specific

frameworks as shown in Fig. 2a. This ordering of synthesis parameters is determined by the mean absolute value of SHAP values of the corresponding to the parameter.

We note the two different types of synthesis parameters: 1) inorganic, which relate to composition of the inorganic components of the synthesis gel (eg. Si, Al, OH, F etc) 2) OSDA, which relate to the organic template (eg. OSDA volume, OSDA rotatable bonds etc) as shown in Table S1. Consequently, this allows us to categorize the formation of a specific zeolite framework as 1 out of 3 main types of synthesis based on its top synthesis parameters as shown in Fig. 2a: 1) Gel-dominated 2) OSDA-dominated synthesis 3) "Balanced" synthesis. An exhaustive list of framework-level SHAP for 190 framework topologies reported on IZA has been included in Fig. S16–S19.

Gel-dominated synthesis These frameworks have syntheses where inorganic components play a more crucial role, with few (≤ 3 out of top 10) OSDA-related parameters. Fig. 2a shows two of such frameworks (**CAN**, **KFI**). In terms of the gel composition, **CAN** and **KFI** share the common trend that both are favored by high levels of mineralizing agent OH. However, beyond that many gel components have vastly different impacts on these two frameworks. For instance, such analysis reveals **CAN** formation seems to be favored by high Na and low K (Barnes, Addai-Mensah, and Gerson 1999). Conversely, **KFI** formation follows the opposite trend, where it appears to be favored by low Na and high K (Han et al. 2021). In terms of reaction conditions, high and low crystallization temperatures favor **CAN** (due to high framework density) and **KFI**, respectively (Dusseiller and Davis 2018).

OSDA-dominated synthesis These frameworks have syntheses where OSDA features are more important. As shown in Fig. 2a, both **ISV** and **ITE** have all of their top synthesis parameters related to the OSDA. One notable exception is the high amount of F driving **ISV** formation due to the presence of the *d4r* CBU in the framework (Villaescusa et al. 2007). One can immediately observe that OSDAs favoring these two frameworks have low asphericity (indicating the need for a spherical OSDA), high volume, and few number of rotatable bonds (indicating rigidity). However, differences do exist; **ITE** formation is associated with high values of OSDA NPR 1 (first normalized principal moment of inertia ratio) with the orange points clearly on the right hand side, while this effect is not present in **ISV** formation where orange and blue points overlap one another. Moreover, unlike **ITE**, **ISV** requires higher amounts of OSDA. We hypothesize that physicochemical OSDA insights could be used to guide the design of optimal OSDAs that target a specific framework.

Balanced synthesis These frameworks have syntheses being driven by a balance of inorganic and OSDA components. In Fig. 2a, it is evident from the gel composition parameters that high Ge promotes **IWW** formation, which can be rationalized by Ge's role in stabilizing the *d4r* cage (Corma et al. 2003). In contrast, **RUT** requires high Si content, which could be expected considering its dense structure. In addition, high Na disfavors both frameworks, albeit the impact of the Na parameter is ranked much lower. In-

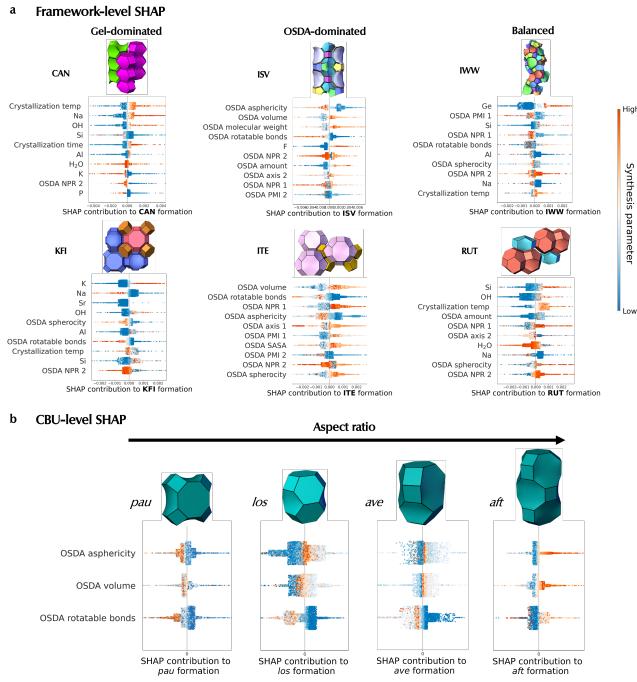


Figure 2: (a) Framework-level SHAP analysis revealing the top 10 (out of 43) most important synthesis parameters favoring the formation of specific frameworks. Each framework belongs to 1 out of 3 types of synthesis based on its top synthesis parameters: 1) Gel-dominated synthesis (CAN, KFI) where most top parameters are inorganic-related, 2) OSDA-dominated synthesis (ISV, ITE) where most top parameters are OSDA-related, and 3) balanced synthesis (IWW, RUT) where even attribution is given to inorganic and OSDA parameters. Every point is an individual synthesis colored by the value of synthesis parameter (orange and blue colors indicate high and low values, respectively). (b) CBU-level SHAP analysis (obtained from aggregated SHAP) of large CBUs showing OSDA parameters favoring their formation.

spection of the OSDA spherocity reveals an opposing trend: **IWW** and **RUT** are favored by low and high OSDA spherocity, respectively. This could be explained by the large spherical cavity present in **RUT** (see Fig. S11a), while **IWW** mainly consists of long channels (see Fig. S11b) that require longer, less spherical OSDAs.

CBU-level SHAP Zeolites adopt a hierarchical structure where CBUs combine to form the zeolite frameworks themselves. To obtain CBU-level SHAP values of a specific CBU, we employ an aggregation SHAP approach of summing the SHAP matrices of all known zeolite frameworks that contain that CBU as described in the Methods section, allowing for rationalization of synthesis-structure at a more fundamental level. An exhaustive list of CBU-level SHAP for all CBUs reported on IZA has been included in Fig. S14 (small CBUs) and Fig. S15 (large CBUs).

Small CBUs We uncover the most important inorganic

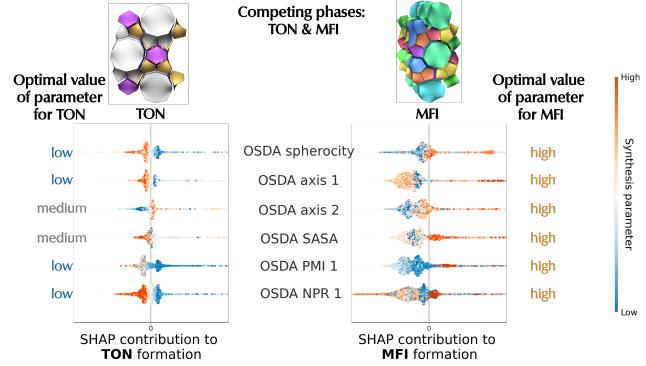


Figure 3: Application of framework-level SHAP on competing phases (TON and MFI). The left- and right-most columns describe the optimal value of OSDA parameter for maximizing formation probability of TON and MFI, respectively. For example, the first row shows *opposing effects* of OSDA spherocity: High OSDA spherocity promotes MFI formation while suppressing TON (and vice versa).

parameters driving the formation of a selection of 4 small CBUs in Fig. S12. Evidently, the synthesis of small CBUs shown are all inorganic/gel-dominated instead of OSDA-dominated with the top 5 parameters relating to the inorganic components, which is expected. We observe the well-established fact that high Ge and F are ranked as the top parameters contributing to *d4r* formation. Furthermore, this analysis reveals a less obvious relationship where low crystallization temperature also positively influences *d4r* formation. Similarly, *d8r* is favored by low crystallization temperatures, but is mainly promoted by high K^+ and Cs^+ cations (Asselman et al. 2022). *can* is driven by high K and requires large amounts of OH as a mineralizing agent. Lastly, high Na and low crystallization temperatures favor *gme* formation (Dusselier et al. 2017).

Large CBUs In contrast to small CBUs, the formation of large CBUs are influenced by OSDA parameters due to the need for a structure-directing effect by OSDAs. Fig. 2b shows a series of large CBUs (≥ 30 T sites) with an increasing aspect ratio (*pau* < *los* < *ave* < *aft*). Interestingly, in the first row, CBU-level SHAP discovers a clear relationship between aspect ratio of the CBU and OSDA asphericity (a measure of the deviation from sphere). For *pau*, low OSDA asphericity gives rise to positive SHAP values, indicating the need for a spherical OSDA. Indeed, this is due to the symmetrical shape of the *pau* cage. As we transition to a CBU with an even higher aspect ratio (*aft*), now only high levels of OSDA asphericity (orange) are needed to drive its formation, indicating the increasing need for longer, more asymmetric molecules to template CBUs with increasing aspect ratio. In the last row SHAP unravels a rather surprising trend: The first three CBUs (*lta*, *los*, *ave*) are favored by very low number of OSDA rotatable bonds, which suggests the need for rigid molecules. However, surprisingly the opposite trend exists for *aft*, where there is a need for a more flexible template with long dimensions (e.g. hexamethonium) (Xie 2021).

Applications of SHAP analysis We suggest the utility of the aforementioned SHAP analysis on an important application in zeolite synthesis: **Competing phases** We consider the most common pair of competing phases in the ZeoSyn dataset, **TON & MFI** (Fig. S9), where these 2 frameworks are frequently formed in the same synthesis. **MFI** is a framework that often appears as a competing phase due to its ease of synthesis and wide synthesis window. Here, we consider achieving phase-selective of **TON** in the absence of **MFI**. Fig. 3a shows the framework-level SHAP for **TON** and **MFI** frameworks. In order to achieve a phase-selective synthesis of **TON**, one may inspect the impact of OSDA spherococity (first row) on the two frameworks, which reveals *opposing* effects on the frameworks: Clearly, an OSDA with low spherococity promotes **TON** formation while suppressing **MFI** as indicated by the rightmost column. In the same vein, the other factors relating to OSDA, such as axis 1, axis 2, solvent-accessible surface area (SASA), principal moment of inertia (PMI 1) and normalized principal moment of inertia ratio (NPR 1) all show *opposing* effects for the 2 frameworks. As such, this showcases framework-level SHAP as a powerful tool for identifying promising synthesis "knobs" and recommends the appropriate direction to tune these "knobs" for phase-selective synthesis.

Conclusion

In this work, we leverage SHAP analysis to uncover the impact of the key synthesis parameters for a zeolites. Furthermore, we introduce an aggregated SHAP approach to extend this analysis to the building unit level, allowing understanding of synthesis parameters at a more fundamental level. Furthermore, this approach has been shown to be useful for the rational design of synthesis parameters for phase-selective synthesis. It is hoped that such rationalization would pave the way for data-driven discovery of zeolitic materials.

References

- 2021. Database of zeolite structures. <http://www.iza-structure.org/databases/>.
- Alfaro, S.; Rodriguez, C.; Valenzuela, M.; and Bosch, P. 2007. Aging time effect on the synthesis of small crystal LTA zeolites in the absence of organic template. *Materials Letters*, 61(23–24): 4655–4658.
- Asselman, K.; Vandenabeele, D.; Pellens, N.; Doppelhamer, N.; Kirschhock, C. E.; and Breynaert, E. 2022. Structural Aspects Affecting Phase Selection in Inorganic Zeolite Synthesis. *Chemistry of Materials*, 34(24): 11081–11092.
- Barnes, M. C.; Addai-Mensah, J.; and Gerson, A. R. 1999. The mechanism of the sodalite-to-cancrinite phase transformation in synthetic spent Bayer liquor. *Microporous and Mesoporous Materials*, 31(3): 287–302.
- Blackwell, C. S.; Broach, R. W.; Gatter, M. G.; Holmgren, J. S.; Jan, D.-Y.; Lewis, G. J.; Mezza, B. J.; Mezza, T. M.; Miller, M. A.; Moscoso, J. G.; et al. 2003. Open-Framework Materials Synthesized in the TMA+/TEA+ Mixed-Template System: The New Low Si/Al Ratio Zeolites UZM-4 and UZM-5. *Angewandte Chemie International Edition*, 42(15): 1737–1740.
- Corma, A.; and Davis, M. E. 2004. Issues in the Synthesis of Crystalline Molecular Sieves: Towards the Crystallization of Low Framework-Density Structures. *ChemPhysChem*, 5(3): 304–313.
- Corma, A.; Díaz-Cabanas, M. J.; Moliner, M.; and Martínez, C. 2006a. Discovery of a new catalytically active and selective zeolite (ITQ-30) by high-throughput synthesis techniques. *Journal of Catalysis*, 241(2): 312–318.
- Corma, A.; Moliner, M.; Serra, J. M.; Serna, P.; Díaz-Cabañas, M. J.; and Baumes, L. A. 2006b. A new mapping-/exploration approach for HT synthesis of zeolites. *Chemistry of materials*, 18(14): 3287–3296.
- Corma, A.; Rey, F.; Valencia, S.; Jordá, J. L.; and Rius, J. 2003. A zeolite with interconnected 8-, 10-and 12-ring pores and its unique catalytic selectivity. *Nature Materials*, 2(7): 493–497.
- Csicsery, S. M. 1984. Shape-selective catalysis in zeolites. *Zeolites*, 4(3): 202–213.
- Cundy, C. S.; and Cox, P. A. 2005. The hydrothermal synthesis of zeolites: Precursors, intermediates and reaction mechanism. *Microporous and mesoporous materials*, 82(1–2): 1–78.
- Davis, M. E. 2002. Ordered porous materials for emerging applications. *Nature*, 417(6891): 813–821.
- Dusselier, M.; and Davis, M. E. 2018. Small-pore zeolites: synthesis and catalysis. *Chemical reviews*, 118(11): 5265–5329.
- Dusselier, M.; Kang, J. H.; Xie, D.; and Davis, M. E. 2017. CIT-9: A Fault-Free Gmelinite Zeolite. *Angewandte Chemie*, 129(43): 13660–13663.
- Ginter, D.; Bell, A.; and Radke, C. 1992. The effects of gel aging on the synthesis of NaY zeolite from colloidal silica. *Zeolites*, 12(6): 742–749.
- Güray, I.; Warzywoda, J.; Bac, N.; and Sacco Jr, A. 1999. Synthesis of zeolite MCM-22 under rotating and static conditions. *Microporous and mesoporous materials*, 31(3): 241–251.
- Han, S.; Tang, X.; Wang, L.; Ma, Y.; Chen, W.; Wu, Q.; Zhang, L.; Zhu, Q.; Meng, X.; Zheng, A.; et al. 2021. Potassium-directed sustainable synthesis of new high silica small-pore zeolite with KFI structure (ZJM-7) as an efficient catalyst for NH₃-SCR reaction. *Applied Catalysis B: Environmental*, 281: 119480.
- Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z.; Roman-Leshkov, Y.; Moliner, M.; Corma, A.; and Olivetti, E. 2019. A machine learning approach to zeolite synthesis enabled by automatic literature data extraction. *ACS central science*, 5(5): 892–899.
- Jensen, Z.; Kwon, S.; Schwalbe-Koda, D.; Paris, C.; Gómez-Bombarelli, R.; Román-Leshkov, Y.; Corma, A.; Moliner, M.; and Olivetti, E. A. 2021. Discovering Relationships between OSDAs and Zeolites through Data Mining and Generative Neural Networks. *ACS central science*, 7(5): 858–867.

- Karpovich, C.; Pan, E.; Jensen, Z.; and Olivetti, E. 2023. Interpretable Machine Learning Enabled Inorganic Reaction Classification and Synthesis Condition Prediction. *Chemistry of Materials*, 35(3): 1062–1079.
- Kumar, M.; Li, R.; and Rimer, J. D. 2016. Assembly and evolution of amorphous precursors in zeolite L crystallization. *Chemistry of Materials*, 28(6): 1714–1727.
- Li, R.; Chawla, A.; Linares, N.; Sutjianto, J. G.; Chapman, K. W.; Martínez, J. G.; and Rimer, J. D. 2018. Diverse Physical States of Amorphous Precursors in Zeolite Synthesis. *Industrial & Engineering Chemistry Research*, 57(25): 8460–8471.
- Lobo, R. F.; Zones, S. I.; and Davis, M. E. 1995. Structure-direction in zeolite synthesis. *Journal of inclusion phenomena and molecular recognition in chemistry*, 21(1-4): 47–78.
- Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Martín, N.; Moliner, M.; and Corma, A. 2015. High yield synthesis of high-silica chabazite by combining the role of zeolite precursors and tetraethylammonium: SCR of NO x. *Chemical Communications*, 51(49): 9965–9968.
- Moliner, M.; Rey, F.; and Corma, A. 2013. Towards the rational design of efficient organic structure-directing agents for zeolite synthesis. *Angewandte Chemie International Edition*, 52(52): 13880–13889.
- Moliner, M.; Serra, J.; Corma, A.; Argente, E.; Valero, S.; and Botti, V. 2005. Application of artificial neural networks to high-throughput synthesis of zeolites. *Microporous and Mesoporous Materials*, 78(1): 73–81.
- Muraoka, K.; Sada, Y.; Miyazaki, D.; Chaikittisilp, W.; and Okubo, T. 2019. Linking synthesis and structure descriptors from a large collection of synthetic records of zeolite materials. *Nature communications*, 10(1): 1–11.
- Pophale, R.; Cheeseman, P. A.; and Deem, M. W. 2011. A database of new zeolite-like materials. *Physical Chemistry Chemical Physics*, 13(27): 12407–12412.
- Schwalbe-Koda, D.; Kwon, S.; Paris, C.; Bello-Jurado, E.; Jensen, Z.; Olivetti, E.; Willhammar, T.; Corma, A.; Román-Leshkov, Y.; Moliner, M.; et al. 2021. A priori control of zeolite phase competition and intergrowth with high-throughput simulations. *Science*, eabh3350.
- Serra, J. M.; Baumes, L. A.; Moliner, M.; Serna, P.; and Corma, A. 2007. Zeolite synthesis modelling with support vector machines: a combinatorial approach. *Combinatorial chemistry & high throughput screening*, 10(1): 13–24.
- Shvets, O. V.; Kasian, N.; Zukal, A.; Pinkas, J.; and Cejka, J. 2010. The role of template structure and synergism between inorganic and organic structure directing agents in the synthesis of UTL zeolite. *Chemistry of Materials*, 22(11): 3482–3495.
- Simon-Masseron, A.; Marques, J.; Lopes, J. M.; Ribeiro, F. R.; Gener, I.; and Guisnet, M. 2007. Influence of the Si/Al ratio and crystal size on the acidity and activity of HBEA zeolites. *Applied Catalysis A: General*, 316(1): 75–82.
- Villaescusa, L.; Díaz, I.; Barrett, P.; Nair, S.; Lloris-Cormano, J.; Martínez-Mañez, R.; Tsapatsis, M.; Liu, Z.; Terasaki, O.; and Cambor, M. 2007. Pure silica large pore zeolite ITQ-7: synthetic strategies, structure-directing effects, and control and nature of structural disorder. *Chemistry of Materials*, 19(7): 1601–1612.
- Weitkamp, J. 2000. Zeolites and catalysis. *Solid state ionics*, 131(1-2): 175–188.
- Wu, Y.; Ren, X.; and Wang, J. 2008. Effect of microwave-assisted aging on the static hydrothermal synthesis of zeolite MCM-22. *Microporous and mesoporous materials*, 116(1-3): 386–393.
- Xie, D. 2021. Rational design and targeted synthesis of small-pore zeolites with the assistance of molecular modeling, structural analysis, and synthetic chemistry. *Industrial & Engineering Chemistry Research*, 60(43): 15403–15415.
- Zhang, H.; Xie, B.; Meng, X.; Müller, U.; Yilmaz, B.; Feyen, M.; Maurer, S.; Gies, H.; Tatsumi, T.; Bao, X.; et al. 2013a. Rational synthesis of Beta zeolite with improved quality by decreasing crystallization temperature in organotemplate-free route. *Microporous and mesoporous materials*, 180: 123–129.
- Zhang, X.; Tang, D.; Zhang, M.; and Yang, R. 2013b. Synthesis of NaX zeolite: Influence of crystallization time, temperature and batch molar ratio SiO₂/Al₂O₃ on the particulate properties of zeolite crystals. *Powder Technology*, 235: 322–328.

Appendix

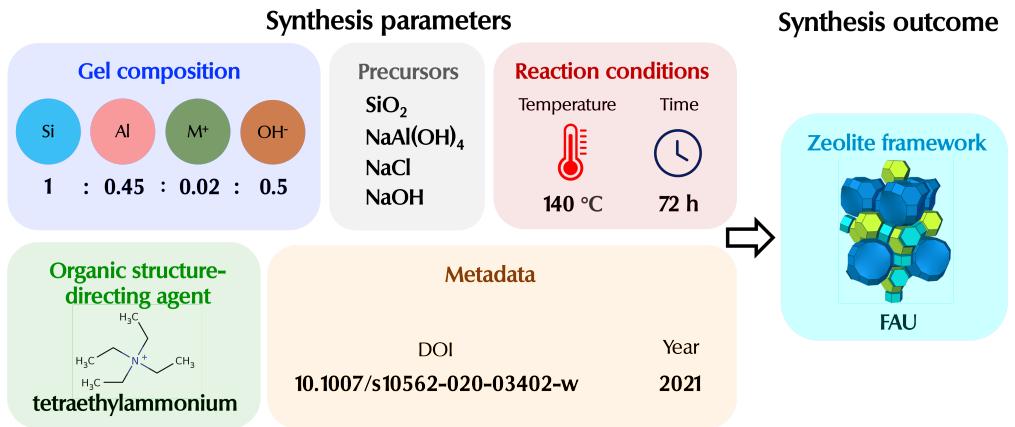


Figure 4: The ZeoSyn dataset.

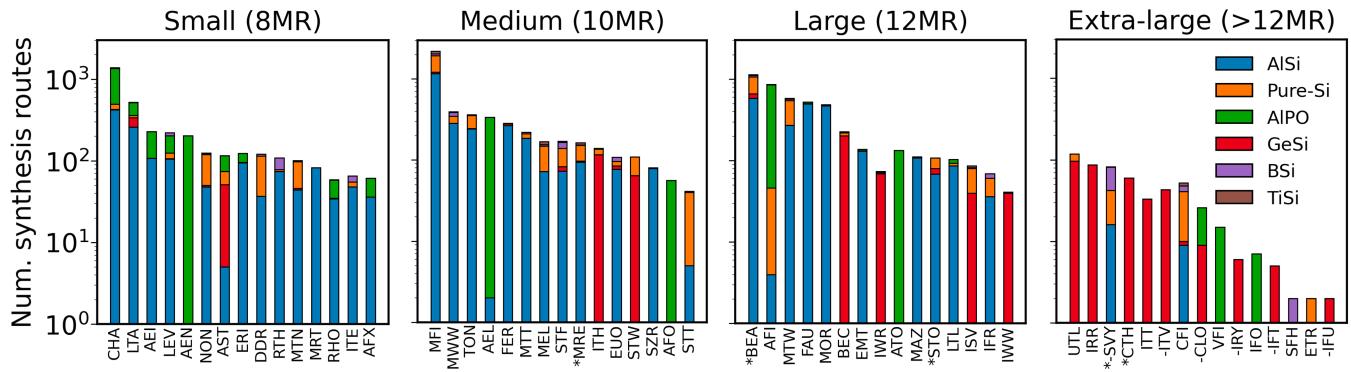


Figure 5: Number of synthetic routes for small, medium, large and extra-large pore frameworks in the dataset. Each framework is further broken down into its constituent zeotypes by color.

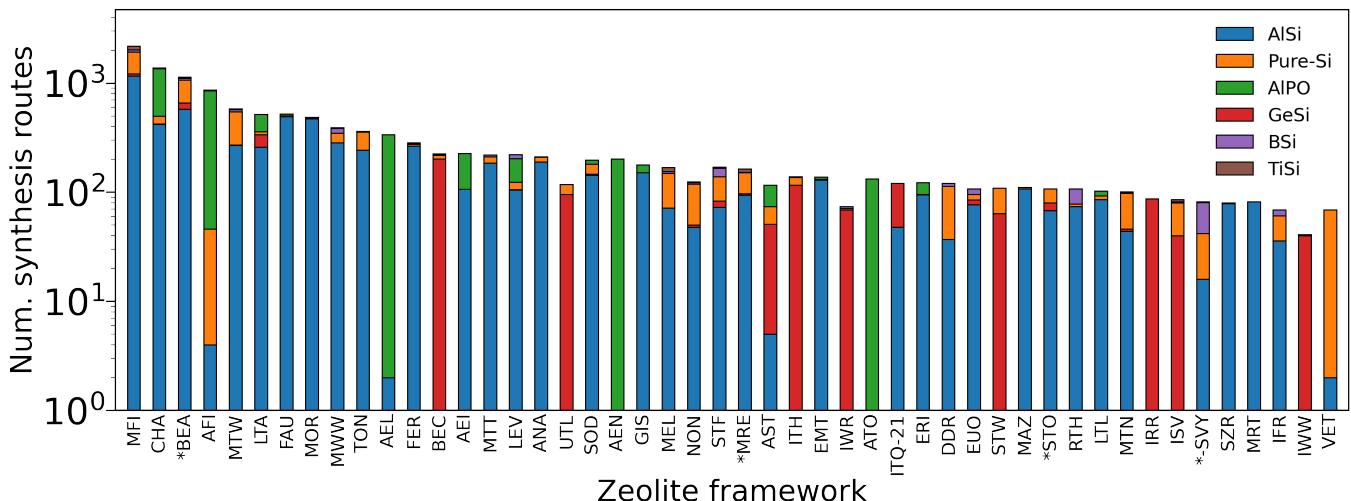


Figure 6: Number of synthesis routes of the 50 most frequent frameworks in the dataset.

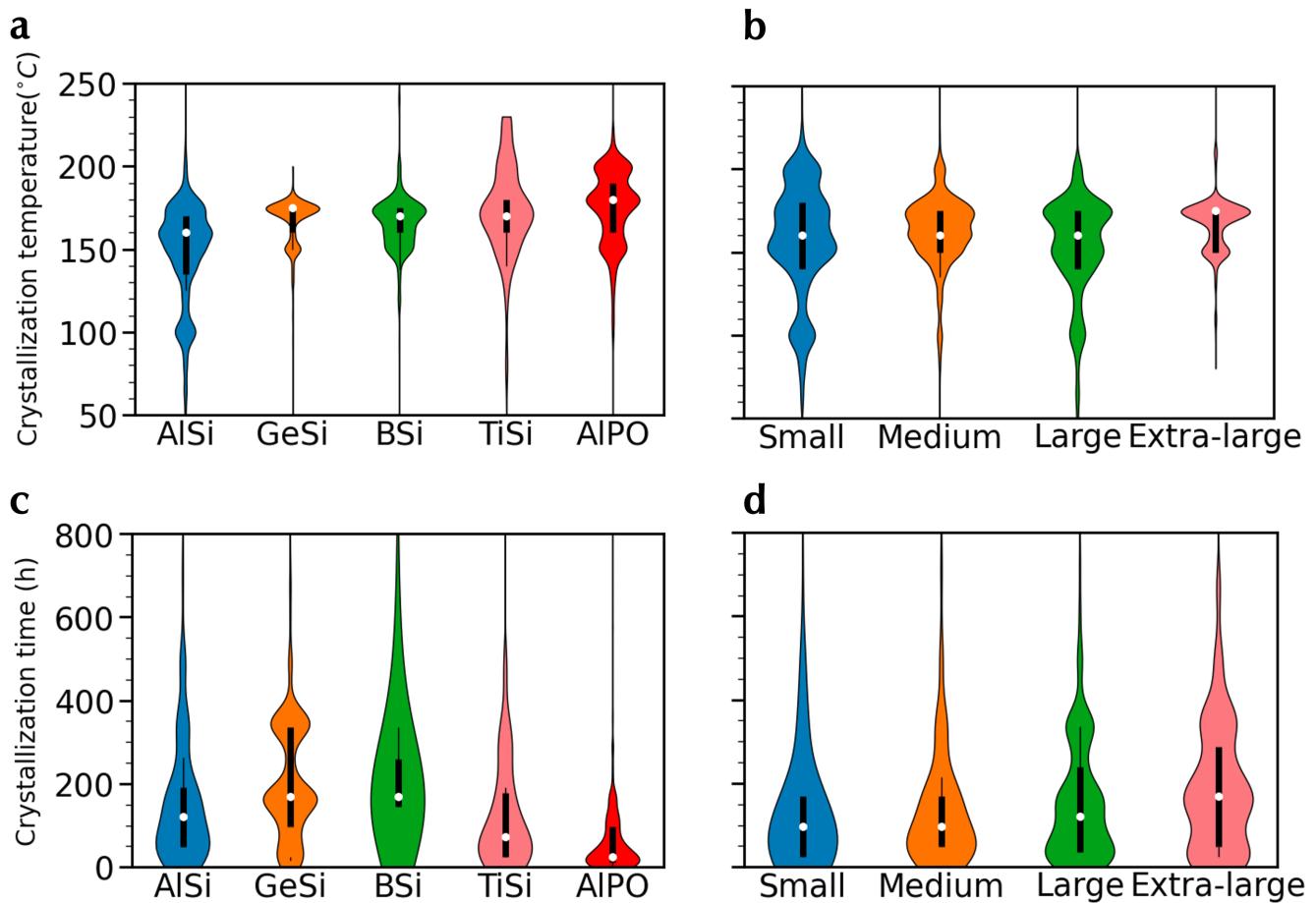


Figure 7: Distributions of reaction conditions in the dataset: crystallization temperatures (a/b) and times (c/d) of different zeotypes and pore-sizes in the dataset.

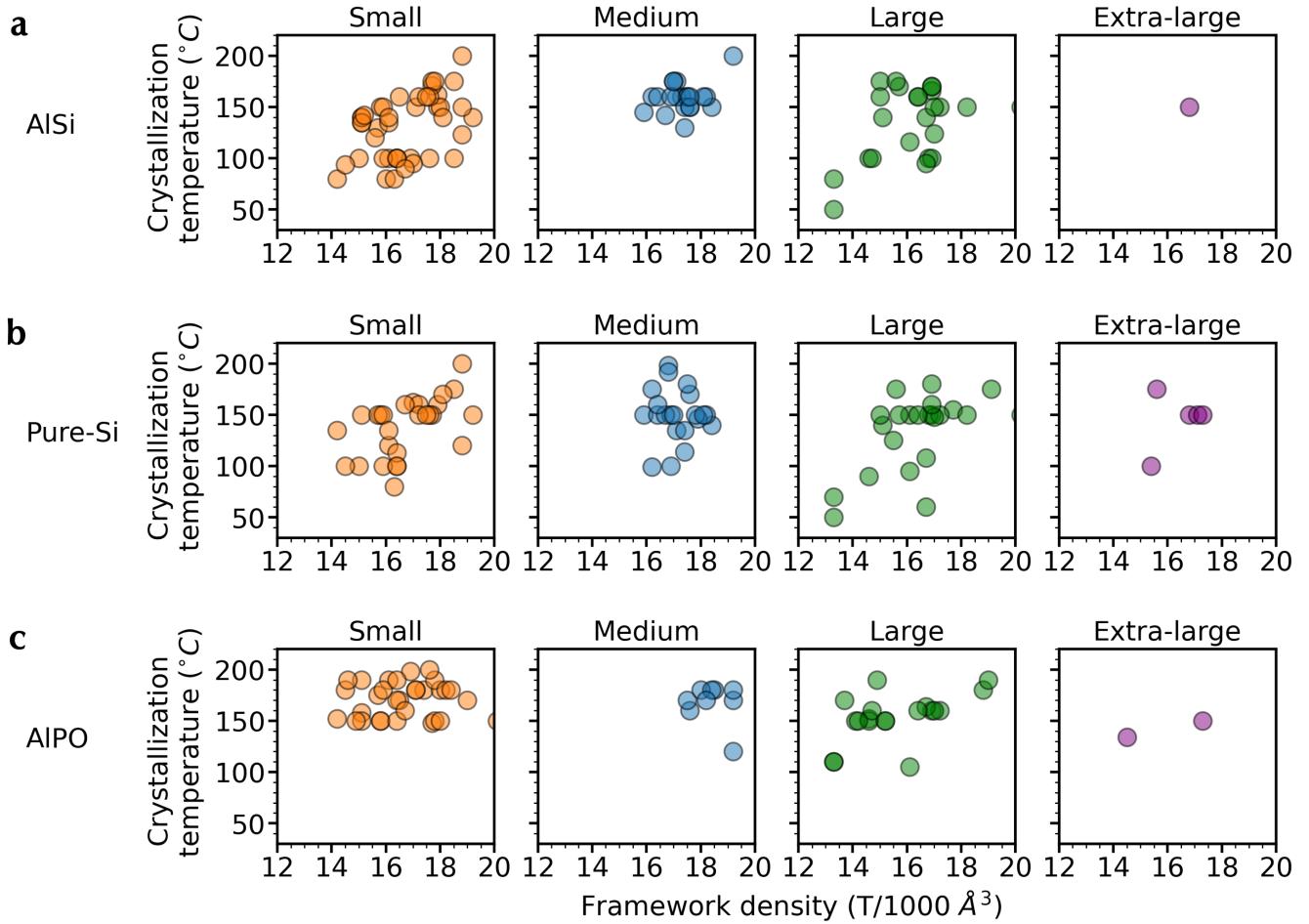


Figure 8: Relationship between crystallization temperature vs. framework density of zeolite for (a) aluminosilicate (b) pure-Si and (c) aluminophosphate frameworks across different pore sizes. Since multiple synthesis routes exist for a single framework type, for each framework we plot the crystallization temperature that corresponds to the 20th percentile for that framework in the dataset.

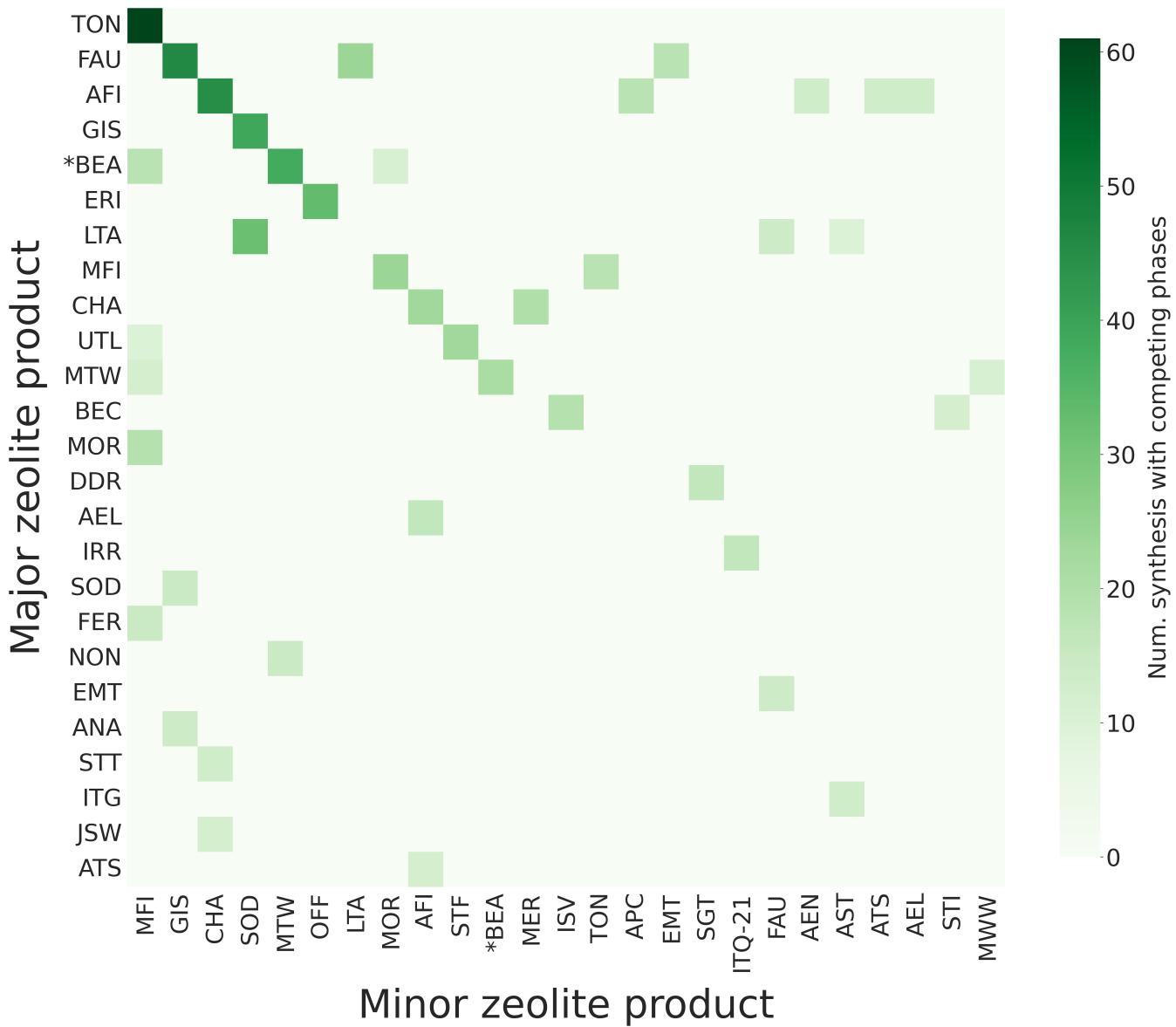


Figure 9: Number of synthesis routes with competing phases. y-axis is the major zeolite product, x-axis is the minor zeolite product.

Table 1: Physicochemical descriptors of OSDAs.

OSDA descriptor	Description
Asphericity	An anisometry descriptor for the deviation from the spherical shape
Axis 1	Two-dimensional (2D) shape descriptors of molecule calculated by projecting the atomic coordinates into a 2D space based on a principal component analysis (PCA) of the positions. The range of the distribution of points in each principal component is reported as the axis of the conformer. Axis 1 is reported as the larger axis, whereas Axis 2 is the smaller axis
Axis 2	See above
Charge	Formal charge of molecule
SASA	Solvent-accessible surface area (SASA) is the surface area of a molecule that is accessible to a solvent
Molecular weight	Molecular mass of molecule
NPR 1	Normalized principal moments ratio (I_1/I_3) where I is principal moment of inertia
NPR 2	Normalized principal moments ratio (I_2/I_3) where I is principal moment of inertia
Rotatable bonds	Number of rotatable bonds in the molecule. A measure of molecular flexibility.
PMI 1	Principal moments of inertia (PMI) are physical quantities related to the rotational dynamics of a molecule.
$I = \sum_{i=1}^A m_i \cdot r_i^2 \quad (2)$	
where A is the number of atoms, and m_i is the atomic mass and r_i is the perpendicular distance from the chosen axis of the i th atom of the molecule	
PMI 2	See above
PMI 3	See above
Spherocity	Spherocity index of molecule
Volume	Molecular volume calculated by using a grid-encoding of the molecular shape using a grid spacing of 0.2 Å and 2.0 Å of margin for the boxes

Table 2: Number of synthesis routes of intergrowths.

Intergrowth	Number of synthesis routes
ISV/BEC	167
ERI/OFF	74
MFI/MEL	47
TON/MTT	43
FAU/EMT	7
AFX/CHA	6
BEA/BEC	5
CHA/AEI	7
STF/SFF	2
RUT/RTH	2
MEL/ZSM-55	1
MOR/MFI	1

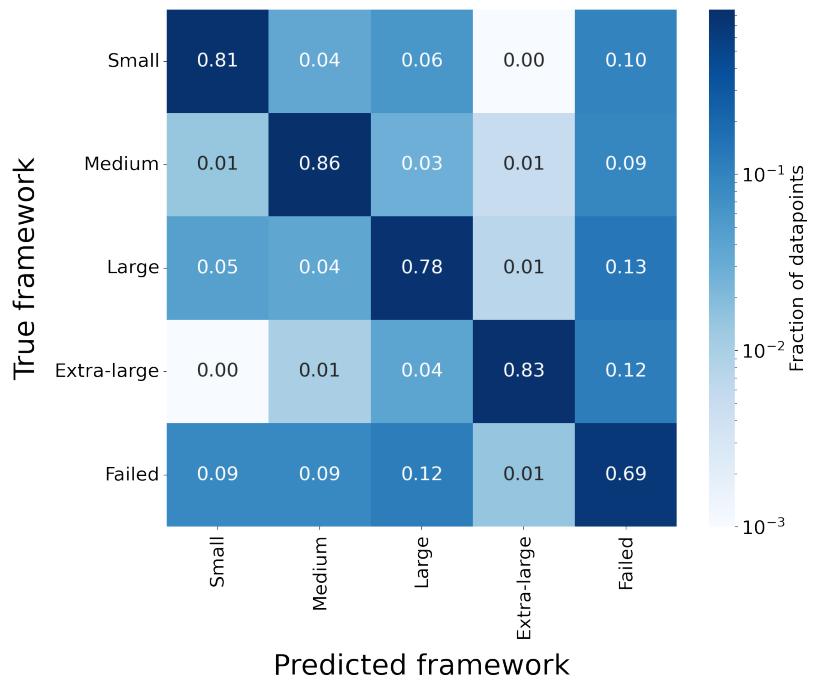


Figure 10: Normalized confusion matrix of phase predictor model. Here, we aggregate frameworks according to small, medium, large and extra-large pore frameworks. "Failed" refers to amorphous/dense phases.

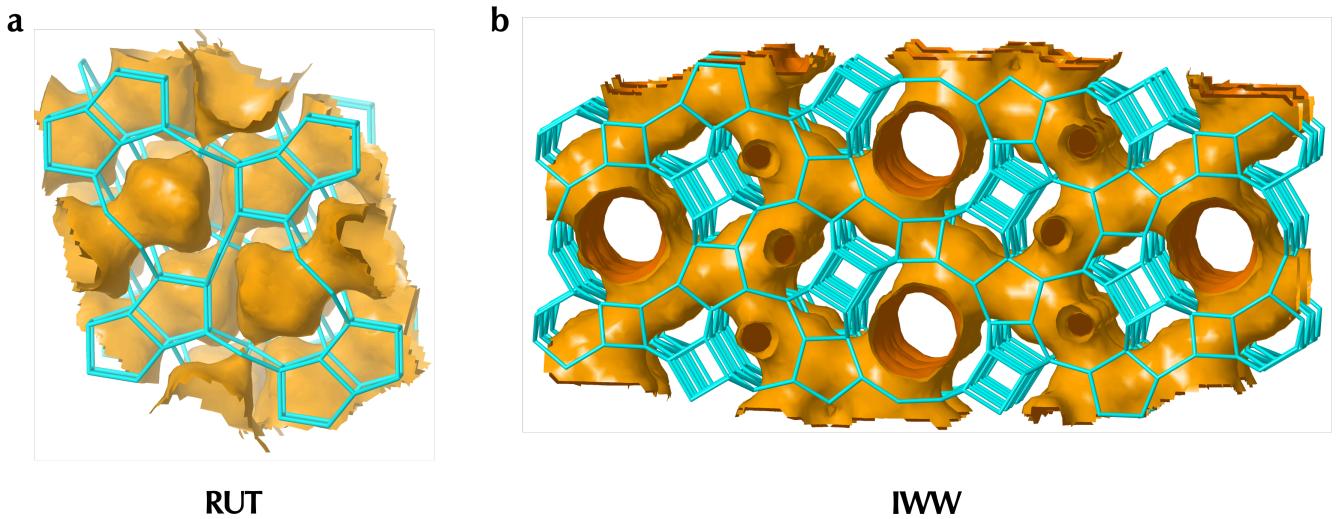


Figure 11: The (a) spherical cavities in **RUT** (b) long channels in **IWW** explain the difference in OSDA spherocity favoring the two frameworks: **IWW** is favored by spherical OSDAs while **RUT** is favored by longer OSDAs.

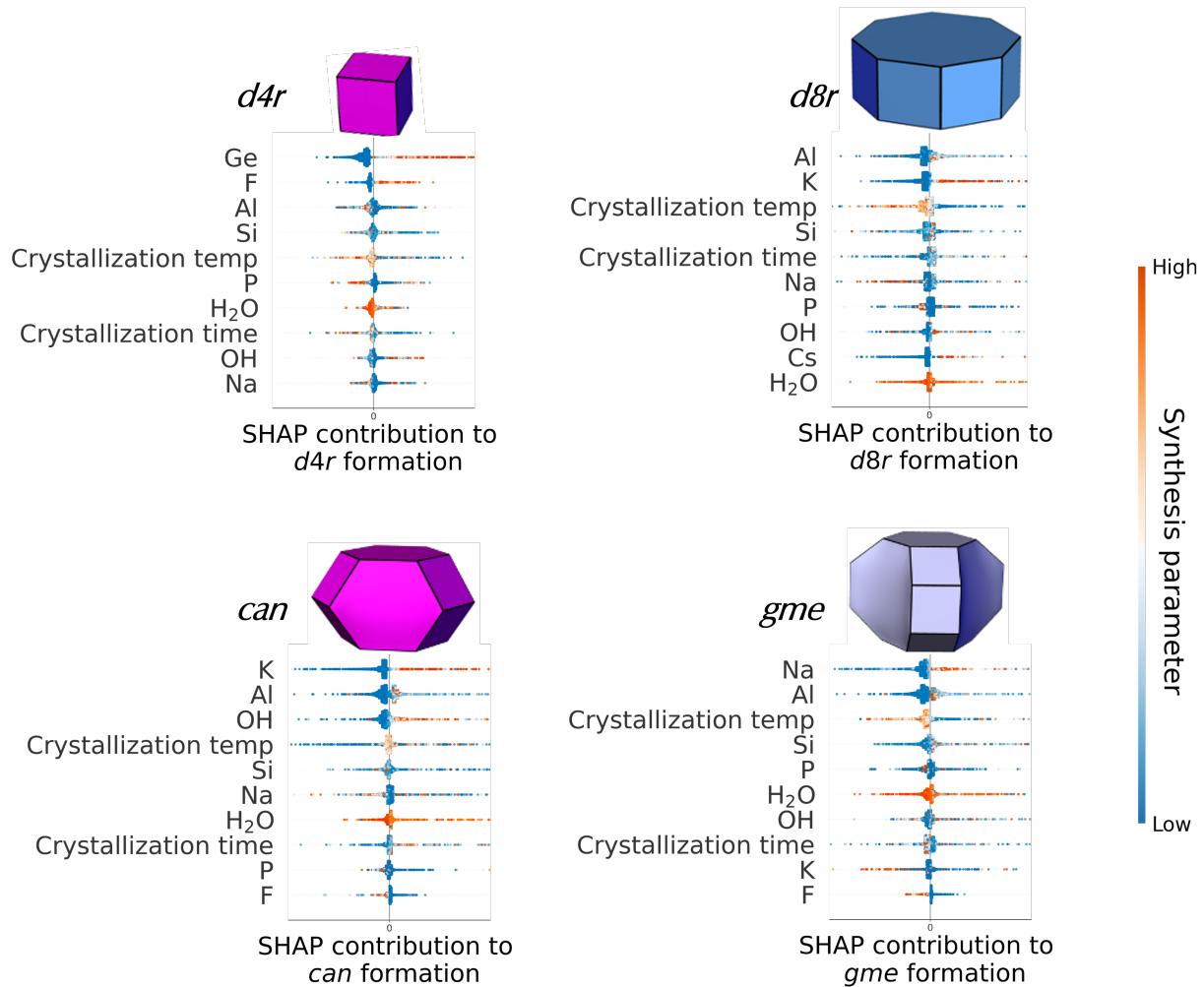


Figure 12: CBU-level SHAP analysis of small CBUs showing top 10 most important inorganic parameters (y-axis) contributing to their formation.

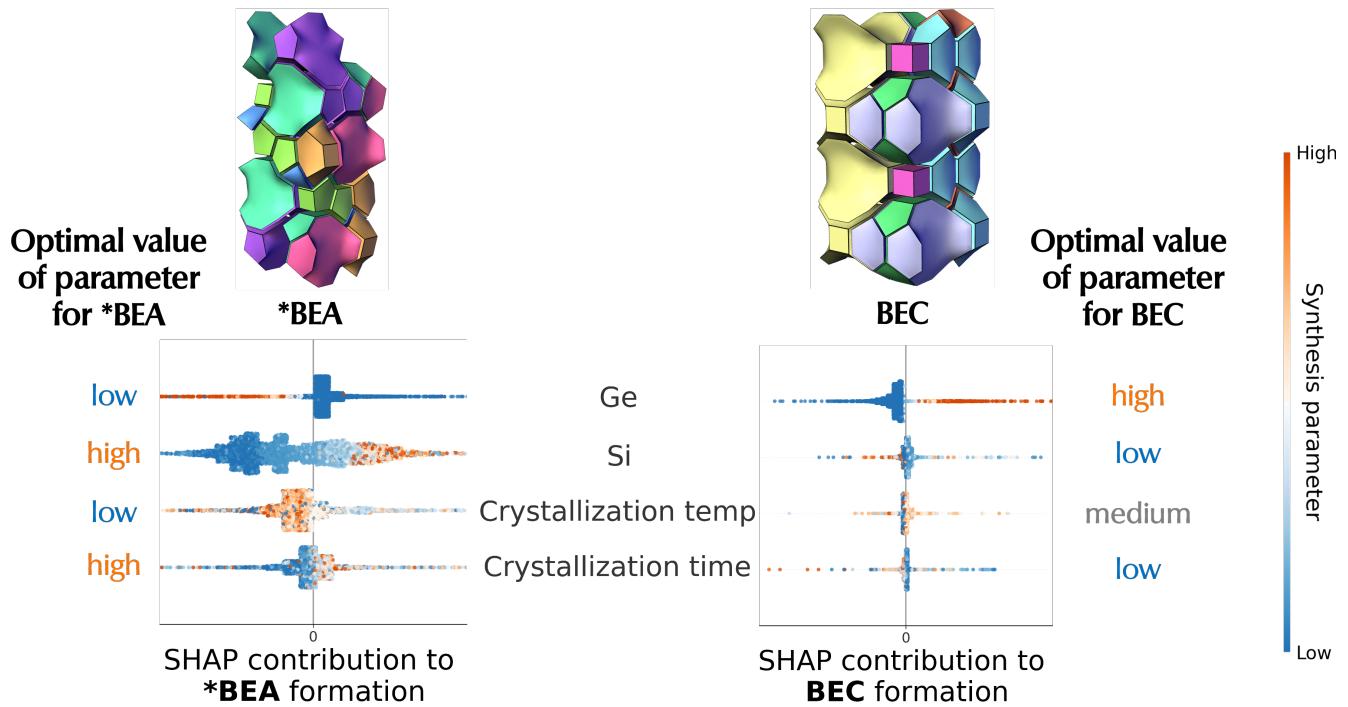


Figure 13: Application of framework-level SHAP on competing phases (***BEA** and **BEC**). The left- and right-most columns describe the optimal value of OSDA parameter for maximizing formation probability of ***BEA** and **BEC**, respectively.

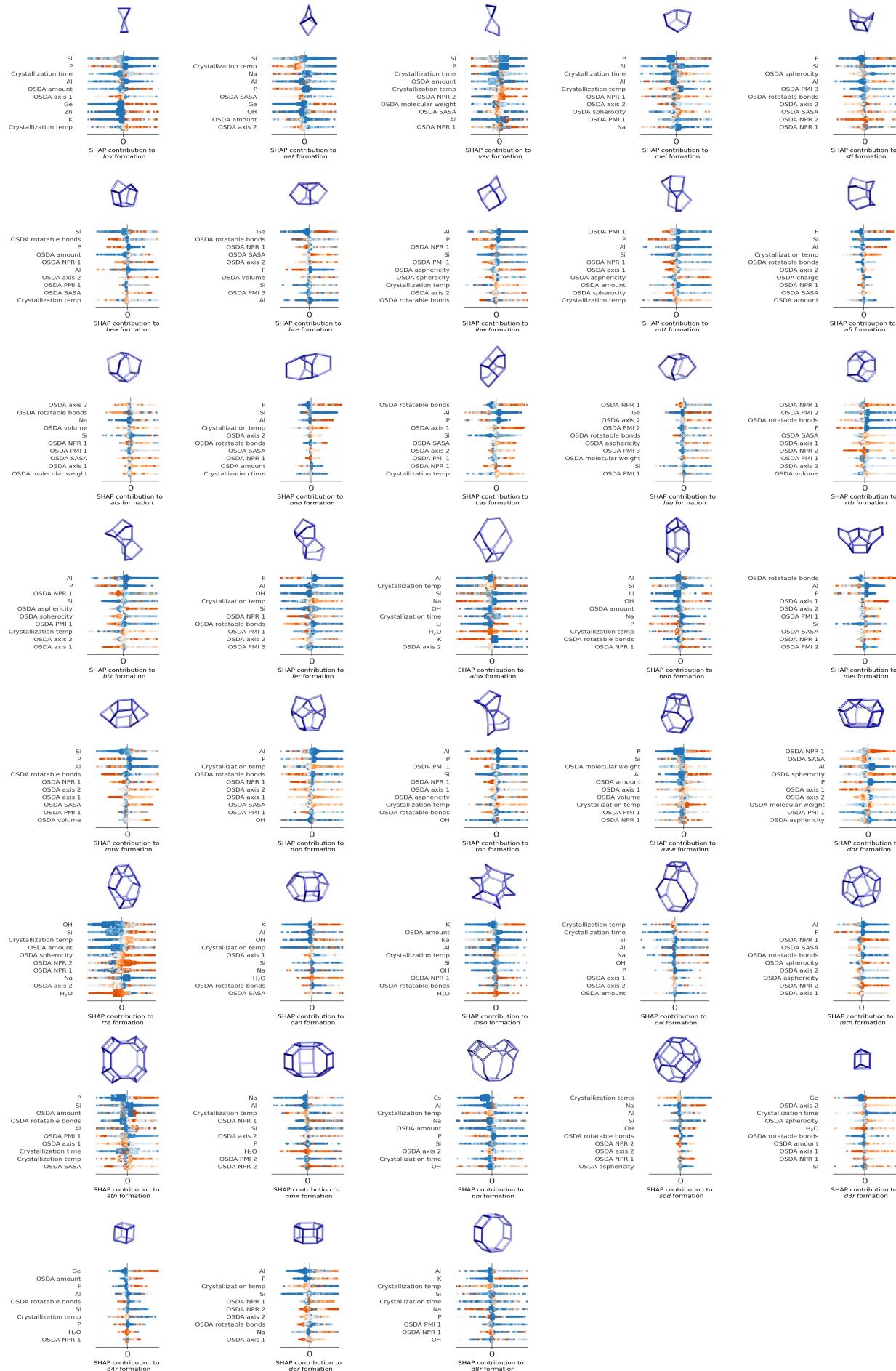


Figure 14: CBU-level SHAP of all small CBUs on IZA.

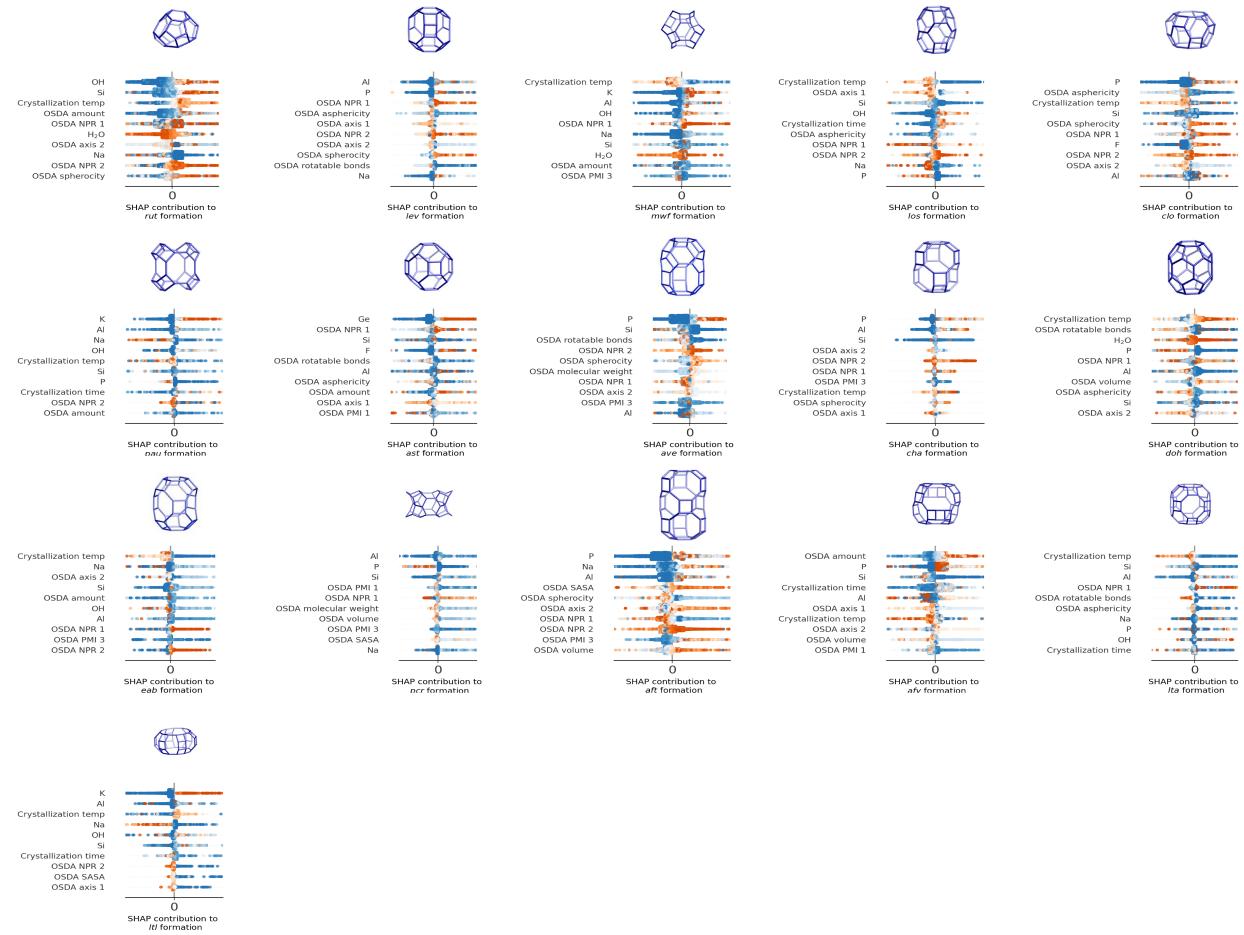


Figure 15: CBU-level SHAP of all large CBUs on IZA.



Figure 16: Framework-level SHAP of zeolite topologies (disordered and interrupted frameworks), and topologies with IZA code starting with A-C.

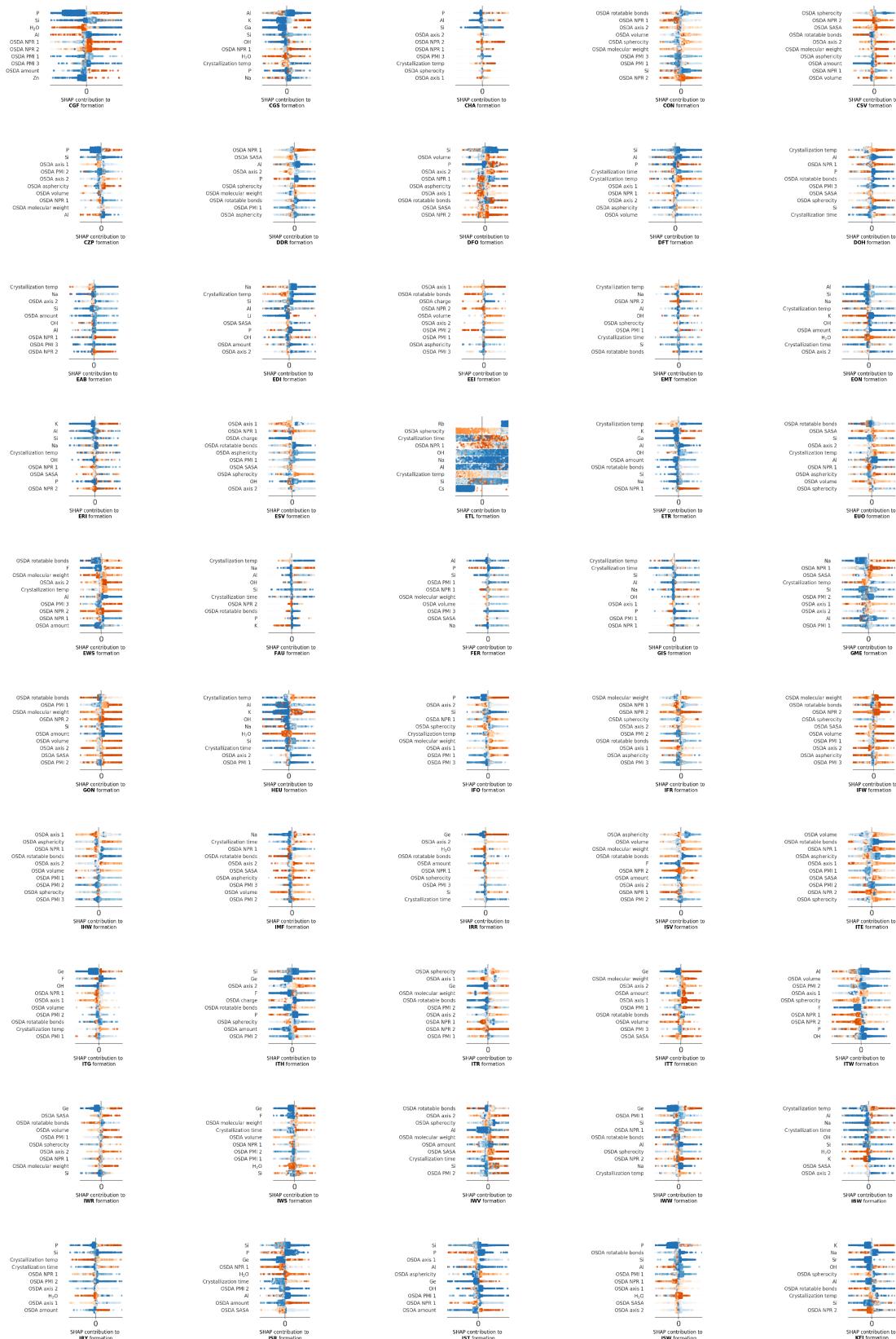


Figure 17: Framework-level SHAP of zeolite topologies with IZA code starting with C-K.



Figure 18: Framework-level SHAP of zeolite topologies with IZA code starting with L-S.

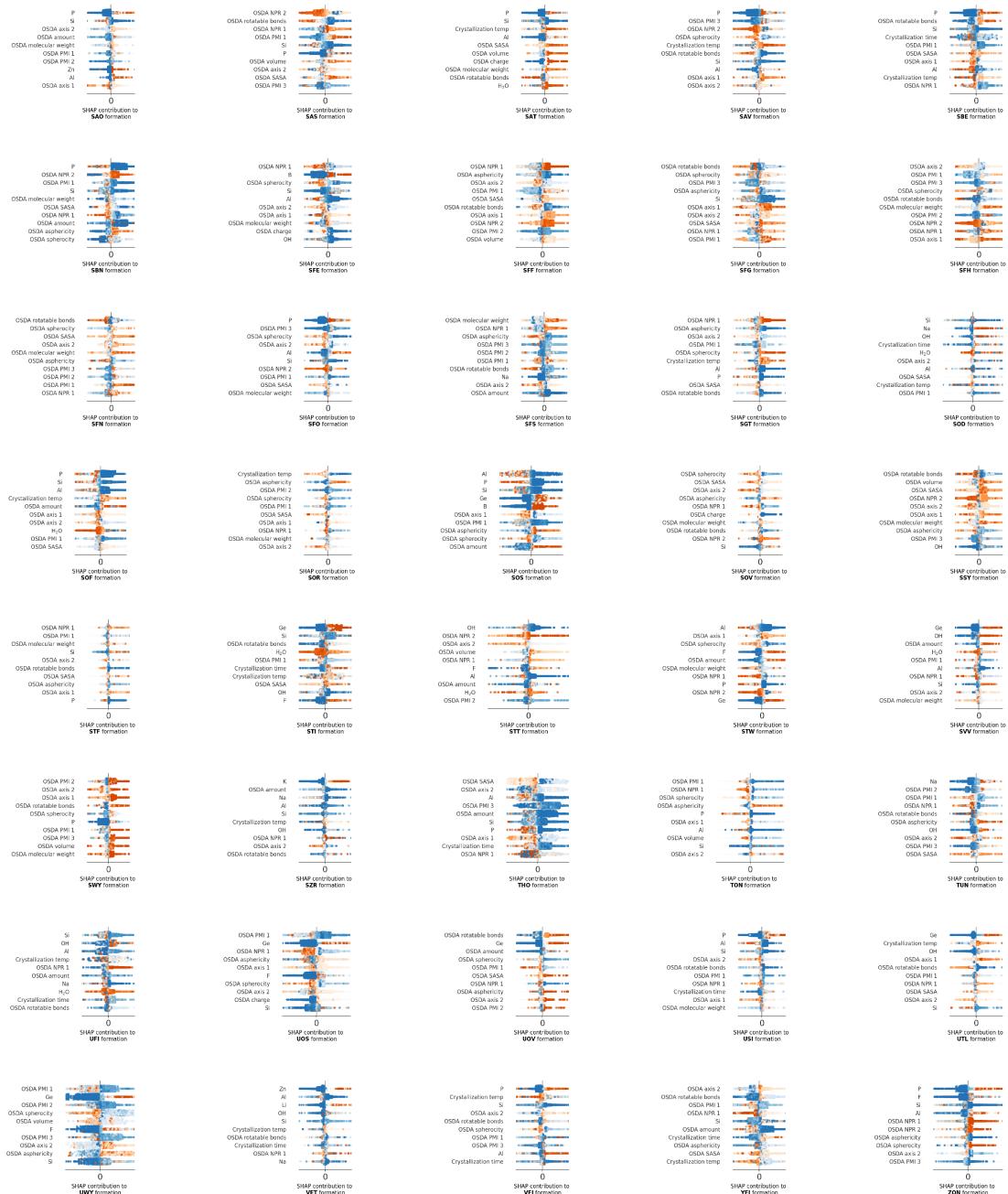


Figure 19: Framework-level SHAP of zeolite topologies with IZA code starting with S-Z.