Contents

[Outline: 1](#_Toc168430164)

[Organised Outline 2](#_Toc168430165)

[Interesting Quotes and ideas from automation papers. 3](#_Toc168430166)

[Things to add 3](#_Toc168430167)

[Artificial Curiosity: A Search for Novel Chemistry and Chemical Understanding. 5](#_Toc168430168)

[Main 5](#_Toc168430169)

[Defining Chemistry 7](#_Toc168430170)

[Data Generation 11](#_Toc168430171)

[Data Exploration 13](#_Toc168430172)

[X-ray Crystallographic data 13](#_Toc168430173)

[Spared information to be compiled into a coherent document. 13](#_Toc168430174)

[References 13](#_Toc168430175)

[Supplementary Information 16](#_Toc168430176)

# Outline:

* Hypothesis: it is possible to learn new chemistry from a robotic ML platform.
* Definition of chemistry and how it defines the search space.
* Removal of bias from sampling all combination reactions.
* Bias of decision maker for single, discrete, diamagnetic supramolecular structures.
* Add reagents used for generation of sample space.
* The choice of reagents were based on: bite angle, rigidity, aromaticity, chirality, sterics, metal valency, and metal size.
* Strength of using ML systems: analysis of huge amounts of data and come to unintuitive decisions.
* Supramolecular chemistry is chosen due to its combinatorics, complexity, easy reaction conidiations to automate, reactions being in equilibrium make single discrete species making it easier to be analysed by NMR.
* The advantage of automation comes with its organisation, (and a bit of saving human effort). Humans get tired and bored and will therefore mess up making up tens of different reactions. Robotic platforms do not have this issue and follow steps to the dot.
* The choice of features for model.
* What is SHAPS and how can it be used for inference (if smiles representations are used, SHAPS will highlight what part of the molecule has the greatest say in prediction).
* The physical setup of the platform.
* How the workflow operates.
* Discovery of Novel structures via a more reasoning approach. If this proof of concept works, the paper would not only have provided insights into the chemistry, but to the fact that ML can infact learn chemical rules. The ML can then be trained on a subset of the combination space, and then predict chemical structures for the remainder. This allows for the discovery of novel compounds without neccasrly having to carry out reactions.
* Robotic platforms allow for the standardisation of reactions, helping bridge the information gap present in todays research1.

# Organised Outline

1) Introduction:

* Hypothesis: it is possible to learn new chemistry from a robotic ML platform.
* Removal of bias from sampling all combination reactions.
* Strength of using ML systems: analysis of huge amounts of data and come to unintuitive decisions.
* Supramolecular chemistry is chosen due to its combinatorics, complexity, easy reaction conidiations to automate, reactions being in equilibrium make single discrete species making it easier to be analysed by NMR.
* The advantage of automation comes with its organisation, (and a bit of saving human effort). Humans get tired and bored and will therefore mess up making up tens of different reactions. Robotic platforms do not have this issue and follow steps to the dot.
* Discovery of Novel structures via a more reasoning approach. If this proof of concept works, the paper would not only have provided insights into the chemistry, but to the fact that ML can infact learn chemical rules. The ML can then be trained on a subset of the combination space, and then predict chemical structures for the remainder. This allows for the discovery of novel compounds without neccasrly having to carry out reactions.

2) Results and discussion:

* Definition of chemistry and how it defines the search space.
* Bias of decision maker for single, discrete, diamagnetic supramolecular structures.
* Add reagents used for generation of sample space.
* The choice of reagents were based on: bite angle, rigidity, aromaticity, chirality, sterics, metal valency, and metal size.
* The choice of features for model.
* What is SHAPS and how can it be used for inference (if smiles representations are used, SHAPS will highlight what part of the molecule has the greatest say in prediction).
* The physical setup of the platform.
* How the workflow operates.
* Robotic platforms allow for the standardisation of reactions, helping bridge the information gap present in todays research1.

3) Conclusions:

# Interesting Quotes and ideas from automation papers.

# Things to add

Talk about the limitation and how the project could be expanded. Change the reaction space to new combinations, add X-ray crystallography workflow, vary temperatures (reaction conditions).

‘The adoption of the user-friendly reaction data format (SURF[41](https://www.nature.com/articles/s42004-023-01047-5#ref-CR41)), facilitated the collection of reaction data from literature sources and enabled standardized reporting of results from HTE and virtual reaction screening. Sharing reaction data in a standardized format plays a pivotal role in the effective utilization of machine learning models for predicting chemical reactivity[44](https://www.nature.com/articles/s42004-023-01047-5#ref-CR44),[45](https://www.nature.com/articles/s42004-023-01047-5#ref-CR45). By using SURF, the initi'al reaction data from three distinct sources (45 from literature, 207 from experiments, and 368 decoy reactions) became readily available for machine learning, obviating the need for manual data curation. Since both the experimental and, particularly, the literature data are predominantly comprised of positive results, incorporating decoy data from unsuccessful transformations played a crucial role in constructing a dependable prediction model.’2

Have a section describing the chemist time vs reaction time (i.e. robotic platform v human chemist time).

Talk about possible future work (this is required by UofG)

Make how parameters as acquired clearer.

Make how programme takes in chemist measured masses to calculate solvent volumes.

The majority of the chemists efforts and time went into setting up the workflow, and setting up runs.

Introduction of metal-organic supramolecular architectures.

Better link between paragraphs.

The imine constitutes as faces or edges of polyhedral structures, while metals act as glue and therefore are found at the vertices of these structures3. One can imaging such simplicity creates a diverse landscape of possible geometric structures. Fortunately, either the kinetic or thermodynamic product is favoured. These geometric structures are governed by chemical rules.

Reactivity and coordination geometry of the metal is a key component in determining structure topology4. Coordination preferences of metals results in vertices with well-defined linker directionality4.

‘In this study, we design and investigate a highly automated workflow that synergizes a high-throughput experimentation platform with a state-of-the-art active learning algorithm to significantly enhance the solubility of redox-active molecules in organic solvents.’5

‘With our automated HTE workflow, the total experimental time to finish the solubility measurement for 42 samples is ca. 27 h (~39 min/sample, less time per sample with running more samples). As shown in Fig. 2g, this is more than 13 times faster than processing samples one by one manually using the ‘excess solute’ approach, which requires approximately 525 min per sample (Supplementary Table S1). While the screening speed of our HTE workflow based on the ‘excess solute’ method is comparable to that of the automated platform proposed by Shiri et al. (20–80 min/sample)27, there are two important distinctions. First, we measured thermodynamic solubility, whereas Shiri and co-workers used the ‘excess solvent’ method for kinetic solubility measurements. Second, our workflow processes 42 or more samples at once, while Shiri et al.’s platform operates on one sample at a time.’5

Structures of coordination driven supramolecular chemistry can be simplified based on the number of binding sites on the metal, the orientation of these binding sites, the number of lewis-acid sites on the ligand, and the orientation between these sites6. These are some of the main features used to describe a supramolecular metal-organic structure, and are the same features which effect topology.

# Artificial Curiosity: A Search for Novel Chemistry and Chemical Understanding.

Random discoveries play a major role in Chemistry, revolutionising different industries. Unfortunately, this makes finding novel materials and compounds rare and unpredictable. Here we present a machine learning model that demonstrates capabilities of predicting chemical reactivity, while making its chemical decision-making process explicit. The machine learning model was trained on experimental reaction data acquired by a robotic platform capable of carrying out hundreds of reactions. To demonstrate the capabilities of combining machine learning and robotics to understand and predict chemical subspaces reactivities, the paper explores imine-based metal organic complexes with 720 different amine, imine, transition metal combinations. The ability to predict reactivity leads to a quicker screening and potential identification of novel compound candidates. Acquired data and code is made public for the research community to use. Preliminary X-ray results led to the discovery of a novel supramolecular structure.

## Main

The chemical space is huge. It is thought that the total space of organic small molecules (<500 daltons) is in the order of 1060 compounds7. Its massive!

Currently, most methods of sampling such a large space and leading to the discovery of novel compounds and chemistry is via serendipity. Such random molecular and material discoveries have aided in tacking challenges in healthcare, materials, energy storage, and sustainability. Just to name a few Nylon, ferrocene, fullerene, electrically conducting polymers, and Wittig olefination were all discovery by chance8. Rarely, have new materials been developed from theory, one must just look at the field of catalysis7. Furthermore, it has been shown that chemist have a bias towards starting materials with historically similar compositions, a further ill sampling of the chemical space9.

Therefore, if such vast chemical space can be explored more quickly in an unbiased manner or the chemistry better understood, novel reactions and materials with socio-economic impact may be screened and tested at a faster rate. Currently, self-driving labs (SDLs) are used to identify and sample interesting chemistry10. SDLs are robotic platforms that carry out reactions in a chemical space directed by computers, namely, machine learning (ML) algorithms. A more traditional method of searching and understanding the chemical space is via high-throughput experimentation (HTE), where a library of compounds is reacted as different combinations and these reactions analysed. A combination of the two allows for the exhaustive search of a space, proving SDLs conscientious chemical sampling and understanding.

HTE of the space, collects data of the space which can be used to train AI algorithms for space exploration via SDLs. A notable example of AI implementation in chemistry is Burger et al’s mobile robotic chemist11. The paper focuses on HTE chemistry to train a ML model for reactivity screening and understanding, such a model can then be implemented in a SDL platform. To do so the ML algorithm has the potential to learn a subset of the HTE space to then extrapolate reactivities to the remainder of the space - allowing for a rapid screening of potential novel molecules. When not implemented in a SDL, the same algorithm allows the chemist to efficiently learn from vast amount of data, while making the model’s conclusions reputable. Using HTE data with explainable ML algorithms, makes future use of AI in SDLs more credible.

Explainable AI are ML algorithms that have the ability to make their decision-making process more understandable to the user. One method of gaining insights into the model’s logic is via SHAP analysis. Simply put, SHAP analysis allows the chemist to understand what parameters were the most relevant for a ML model to predict reactivity. Due to SHAP’s game theory approach, SHAP analysis lies in local explanations but may be used for global explanations. Local explanations look at the chemistry of a single reaction, while global explanations look at the general chemical trend - both provide powerful insights of the explored chemistry. It’s clear that ‘explainbility’ cements the chemists understand of the explored space and their trust with the model’s extrapolative abilities. It therefore makes finding novel compounds via SDLs more robust.

Moreover, chemical space exploration vial ML can be made further more robust thanks to the combination of HTE and robotic-oriented chemistry. The first being the generation of the HTE dataset may be used to further develop alternative ML tools. Additionally, the use of robotic systems to physically explore the space allows for training on experimental data and outcompetes any ab initio explorations due to the latter’s associated computational costs12. Thirdly, robots focus on the mundane tasks, allowing the chemist to focus on the intellectual challenges. Likewise, by improving reproductivity and accuracy of results, robotic platform reduces human error10. Lastly, exploring chemistry as high throughput experiments help tackle some biased data (its limitations are further discussed in the paper), a major problem with current digital chemistry13,14. Such systems also bring a big disadvantage.

Robotic platforms for use in chemistry have limited capabilities due chemistry’s intricacies. To add perspective, a platform must be able of handling: different reagent solubilities, heterogenous mixtures, stirring, heating, quantitative workup procedures, solid dispensing, purification, side reactions, and unexpected crashes. Simple chemistry to perform via robots is therefore imperative, even though it unfortunately limits chemical space exploration. A big limitation of the platform developed in the paper.

The paper explores supramolecular chemistry for their compatibility with the paper’s platform along with the fact that they find uses in molecular recognition, chirality sensing, separations, stabilisation of reactive species and catalysis3,15. Their varied use makes novelty discovery in this space important.

Additionally, the paradigm of self-assembly is perfect for use with ML: component molecules form a define arrangement based solely on the guidance of the non-covalent interactions between components. In other words, the architecture of the product is predefined by the information present in the ligands, in the form of structural and conformational features and are formed on the basis of h-bonds, non-covalent interactions and metal-ligand bonds16. Supramolecular structures formed on the Metal-ligand architecture information is better encoded in the metal and ligand building blocks4. Therefore, predicting if a structure is formed or not a priory is possible and best done via coordination driven self-assembly. This is one of the reasons the paper focuses on imine-based metal organic complexes.

The dynamic nature of imine and metal ligand bonds allows for the selection of the thermodynamic product out of a landscape of possible species17. The simplicity of NMR and MS architecture characterisation makes product identification a simple task for the workflow. This is supported by previous research, where protons in a supramolecular complex exhibit a single chemical and magnetic environment, while ESI-MID elucidates the stoichiometries of building blocks in the supramolecular architectures16. In traditional vanilla HTE, product characterisation require some form of tags or substituent modification for analysis18. Therefore, the paper explores supramolecular chemistry due to simple automated analysis.

Furthermore, reaction simplicity cuts down the number of challenging steps the platform must be capable of handling. The ChemSpeed used for this research, is only capable of heating, stirring, and transferring volumes. The spontaneity of imine formation and coordination means the reaction may be carried out at room temperature19. Additionally, the solubility of many metals, amines, and aldehydes in acetonitrile allows for the ChemSpeed to handle reagents in the form of stock solutions with a standard solvent. Also, imine metal supramolecular structure formation is a one pot process, allowing the platform to treat a simple vial as a reaction vessel.

Combinatorial accessibility allows for the introduction of multiple metals, imines and aldehydes in a single reaction to have an effect on the structure of the product. In other words, with the same reagents, multiple combinatorial spaces with unique reactivities may be generated, there are just a greater number of participating reagents in a reaction. Plus, the diversity of commercially availabile aldehydes and amines does not make reagent selection a limiting factor in this study19. Of course, imine chelation and metal coordination sites are another combinatorial parameter imine complex have: this paper focuses on diamine-monoaldehyde combinations with metals of coordination number 4 and 6. The chemistries ability to generate multiple combinations ensure a large enough interesting chemical space to explore.

Lastly, the diversity and complexity of product structures makes it a challenge for both the chemist and ML to predict reactivity. Compared to covalent interactions, intermolecular forces are less directional, making predicting supramolecular geometries a challenge for the chemist20. The advantage of ML comes in its ability to explore complex and unintuitive data with speed and minimal computation21. Additionally, the discovery of new metal containing compounds has always been shadowed by more traditional organic chemicals, therefore it remains an underexplored area9. With a greater potential to make discoveries in relatively complex chemistry is a perfect testcase for a system aiming to bridge SDLs, HTEs and explainable AI.

How the paper intends to combine HTE, robotics, and chemical space exploration is highlighted in Figure 1 – an overview of the study.

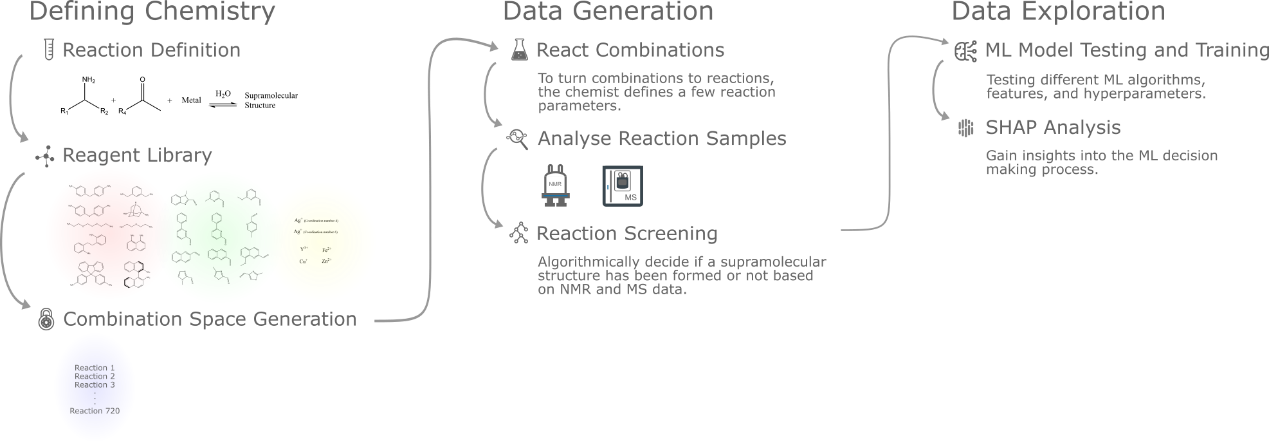


Figure 1: An overview of the study preformed in the paper, subdivided into three sections: Defining chemistry, data generation and data exploration. The first section is responsible for defining the chemical space to explore and the generation of reactions to be carried as HTE. The second section explores the generated reaction space, analysis the reaction to then label reactions as failed or not. The third section explores different machine learning models, hyperparameters, features and is responsible for data and ML model analysis.

## Defining Chemistry

The chemical space is massive and driving a robotic platform to explore it all is unrealistic, taking a subset of the space narrows down the search. Focusing on a reaction selects a space who’s compounds are derived from different reagent combinations, making it possible to fully explore a well-defined space. The process of selecting a reaction to explore and the space it encompasses is shown in Figure 2.

A reaction can be abstracted as an equation where reagents of two or more functional groups react to give a product. Identifying reagents with similar chemistries allows for the parting of the reaction reagents into subdivisions. Therefore, if combinations of all reagents in the different subdivisions are reacted, products should cover the possible chemical space of the reaction. Such a combinatorial space is still ill-defined as the number of reagents in a subdivision may be vast, and combinations may have different numbers of the same subclasses in them. For example, a combination might have one subdivision 1, two subdivision 2 and one subdivision 3 reagents, while another has two subdivision 1 one subdivision 2 and two subdivision 3 reagents. Therefore, the number of subdivision reagents in a combination must be defined, and reagents in subdivisions selected.

Defining the chemical space as combinations allows the model to learn reactions that work and reactions that won’t work which reduces bias in the dataset, making the model more reliable for identification of possible novel compounds and reactions. The model must also be generalisable, that is, able to predict reactivity with unseen chemistry. Selecting diverse reagents in subdivisions improves the generalisability. This process of selecting the chemical combination space is demonstrated in Figure 2, how it fits into the scope of the paper is shown in Figure 1.

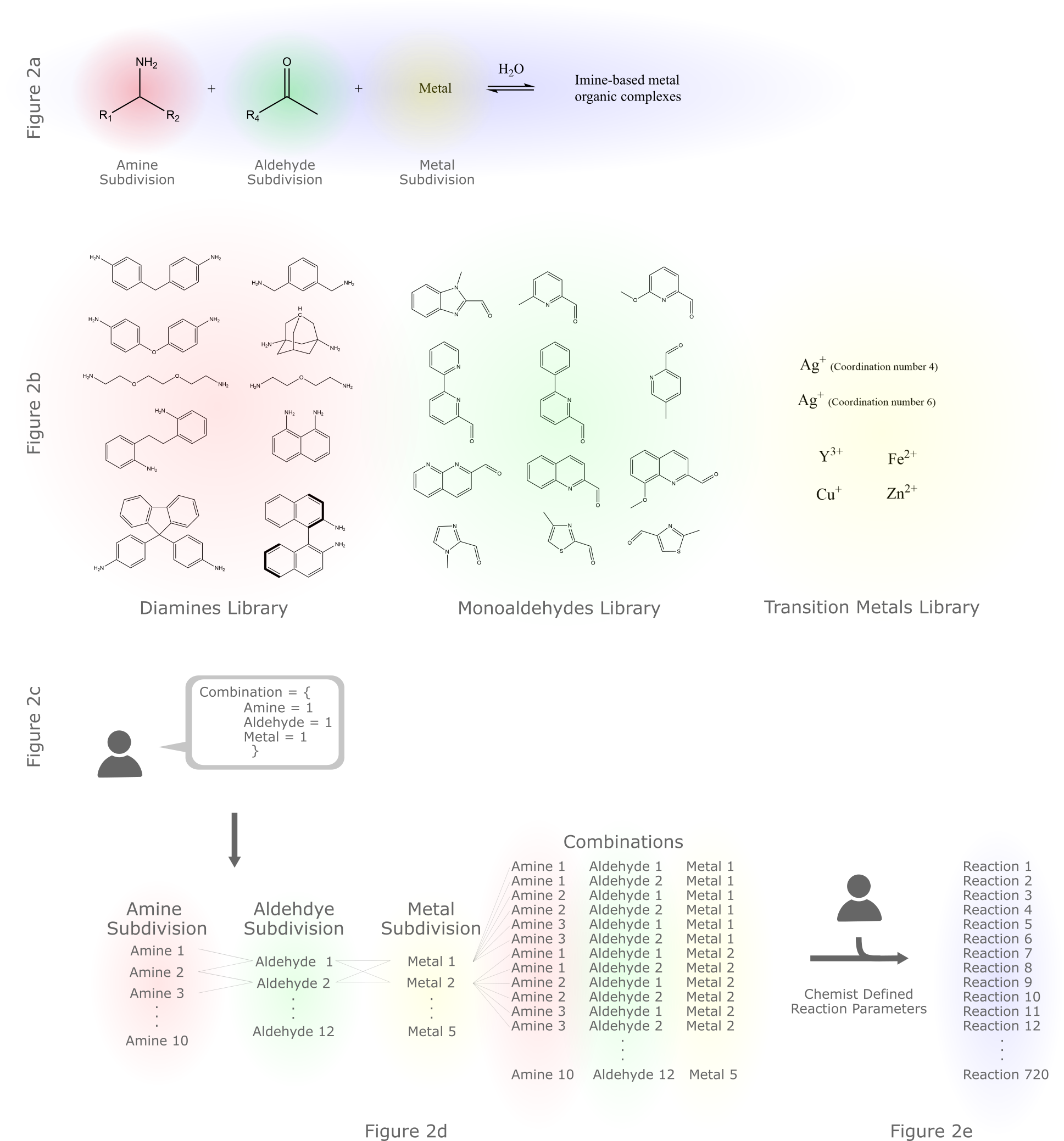
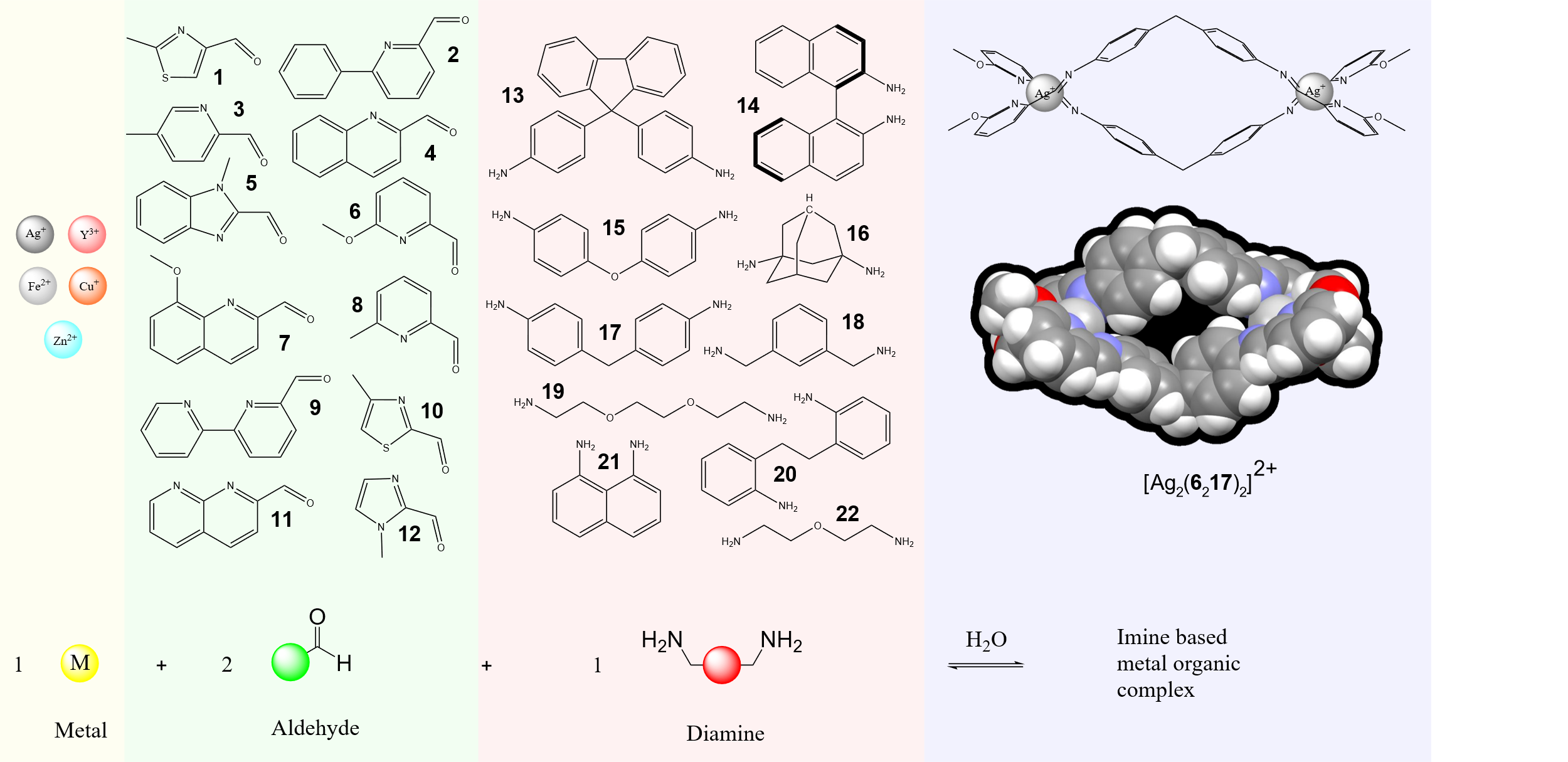


Figure 2: Generating the combinatorial space, where red, green and yellow represent subdivisions while blue represents reactions. 2a) The reaction explored in the paper. 2b) The reagent libraries used in the paper. There are three libraries for each subdivision. 2c) Stating how many reagents of a subdivision should appear in a combination. The paper uses one of each. 2d) The generation of combinations using the previously defined libraries, subdivisions and subdivisions quantities. 2e) Translation of combinations to reactions based on several different chemist defined parameters. These reactions are computer, and thus robot, readable.



The paper looks at the formation of Imine-based metal organic complexes and the subdivisions of the reaction are diamines, mono-aldehydes, and metals (Figure 2a). Diamines react with monoaldehyde to form imines which may coordinate to metals. The imine constitutes as faces or edges of polyhedral structures, while metals coordination act as connecting vertices, resulting in a supramolecular structure3. One can image such simplicity creates a diverse landscape of possible geometric structures. At its foundations, the ML is learning imine, coordination and geometric chemistry.

For the selection of subdivision libraries, 10 different diamines, 12 different monoaldehydes, and 5 different transition metals (of which silver was considered to have coordination numbers of 4 and 6) were selected based on commercial availability, a nitrogen beta to the carbonyl group and diamagnetic metals (Figure 2b). A full list of reagents and structures can be found in S 1. Beta nitrogen allows for the chelation of the imine to a metal, while diamagnetic metals ensure interpretable NMR spectra are produced. The reagents selected also encapsulate functional diversity: bite angle, rigidity, aromaticity, chirality, sterics, electronegativity, metal valency, and metal size (Figure 2b). Unfortunately, selection of reagent libraries does introduce bias, such as commercially available reagents.

After selection of reagents, the number of subdivisions in a combinations have to be selected. The paper looks at taking combinations with one reagent from each subdivision (Figure 2c), for a total space size of 720 combinations (Figure 2d). To test ML model predictability and understanding, the space is fully explored. However, the model is only trained on a subset of this space, and its predictions of the remaining space compared to the real space.

The chemist then defines reaction and reagent parameters to translate combinations to reactions (Figure 2e). Parameters include: concentration of a subgroup in a combination, chemical weights of reagents, solubility of reagents, price of reagents, coordination number, and transfer volumes. This paper takes combinations, heats them at 60oC for 40h to create reactions. Since these are user parameters, they can be adjusted by the chemist to fit the chemistry of interest.

Since the translated reactions can be in the order of thousands, a robotic platform is used to standardise and make reactions replicable. Standardisation and replication helps bridge the information gap present in literature1. Again, it is important to note that the current robotic platform does create bias towards nongaseous, stable, soluble reagent libraries. A chemical bias introduced by the limitations of robotic platforms has also been identified in the automation of medicinal chemistry22. With the chemical space defined, and a robotic platform ready to go, reactions need to be carried out and analysed. This is the data generation step of the study - Figure 1.

## Data Generation

The next few steps include carrying out reactions, taking NMR and MS spectra to then decide if a combination results in a supramolecular compound. The process of carrying out and analysis of chemical space is handled by a Chemspeed Swing platform, a Bruker benchtop NMR machine, and a Waters Acquity LCMS system (Figure 3).

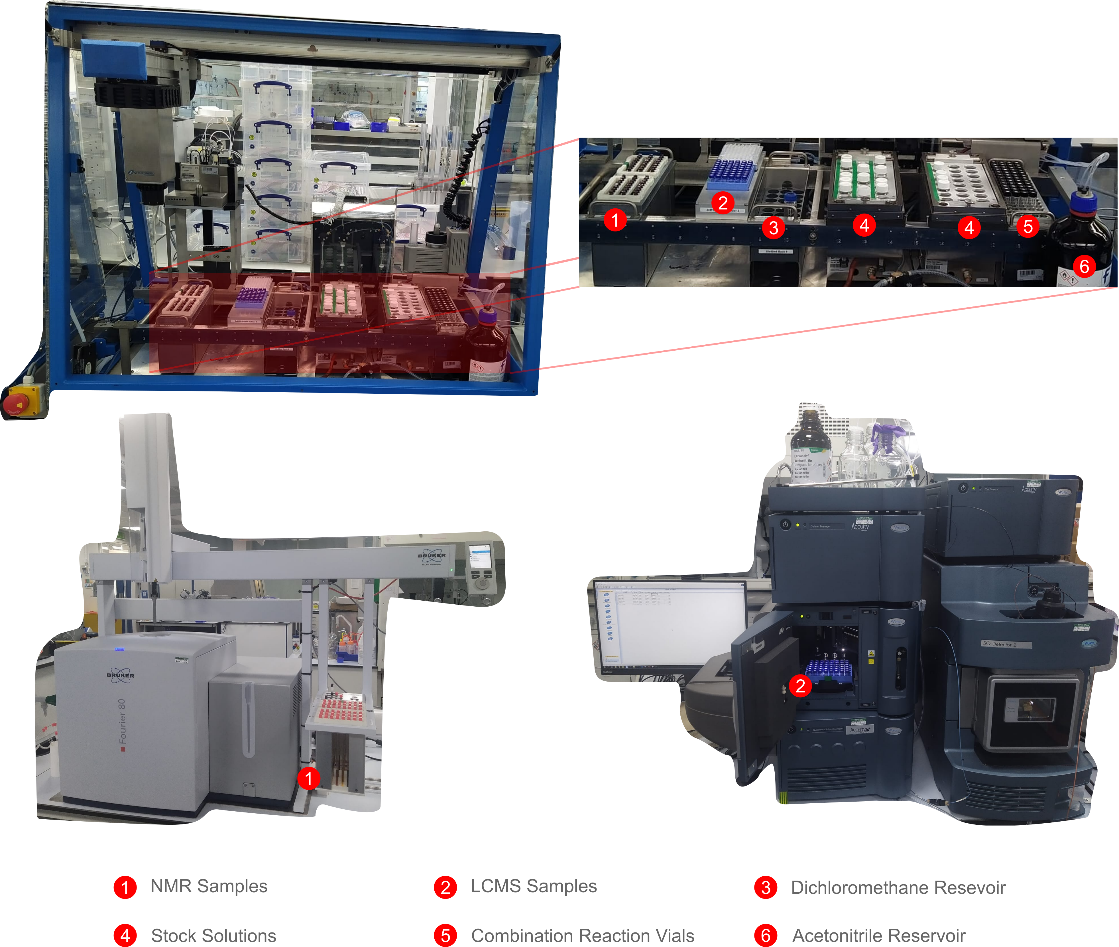


Figure 3: The three platforms used in the workflow. At the top is the Chemspeed platform, responsible for making up stock solutions, reactions scheduled by a programme and taking NMR and LCMS samples for each combination. At the bottom left is the benchtop NMR used in the workflow along with the in-house built NMR sample rack. At the bottom right is the LCMS machine used in the workflow, along with reaction samples.

Since platforms are independent of each other and are limited to 48 samples, a programme must schedule and communicate samples to make up between platforms. The same programme is also responsible for labelling combinations as failed or successful reactivities. All programmes and algorithms used in these stages were developed by the research group. Lastly, since everything is defined via an abstracted programme, expansion of the workflow, and use for other areas of chemistry is simple. The workflow the programme follows and its capabilities are shown in Figure 4.

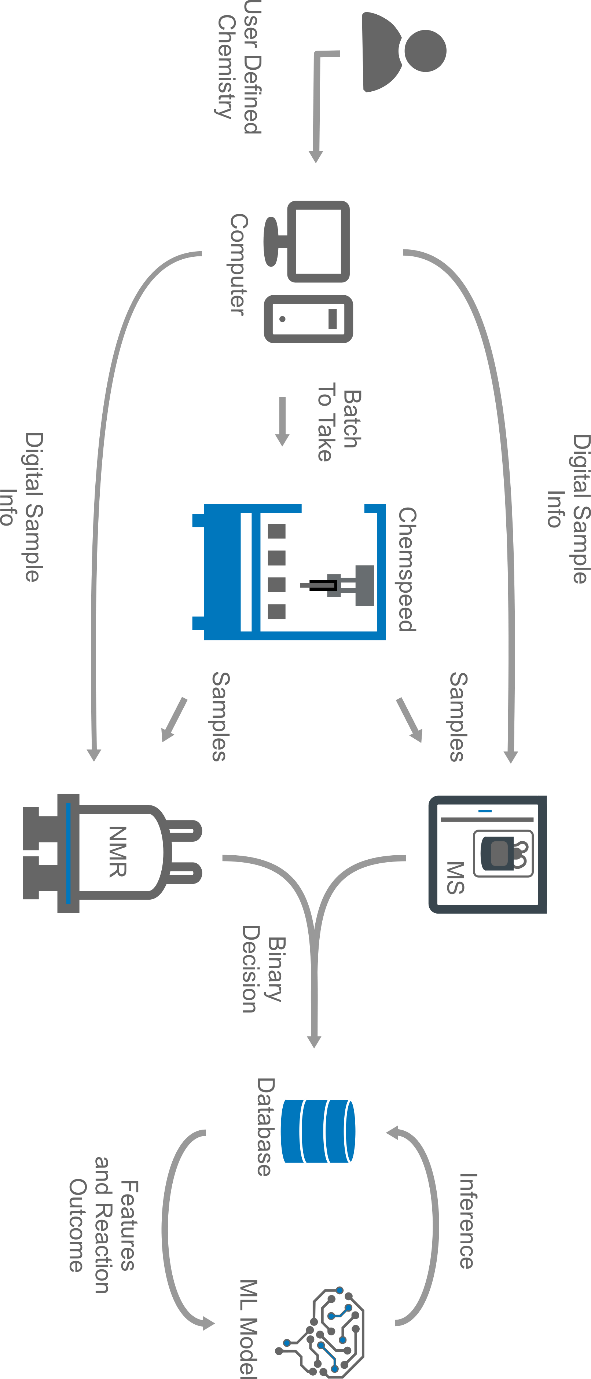
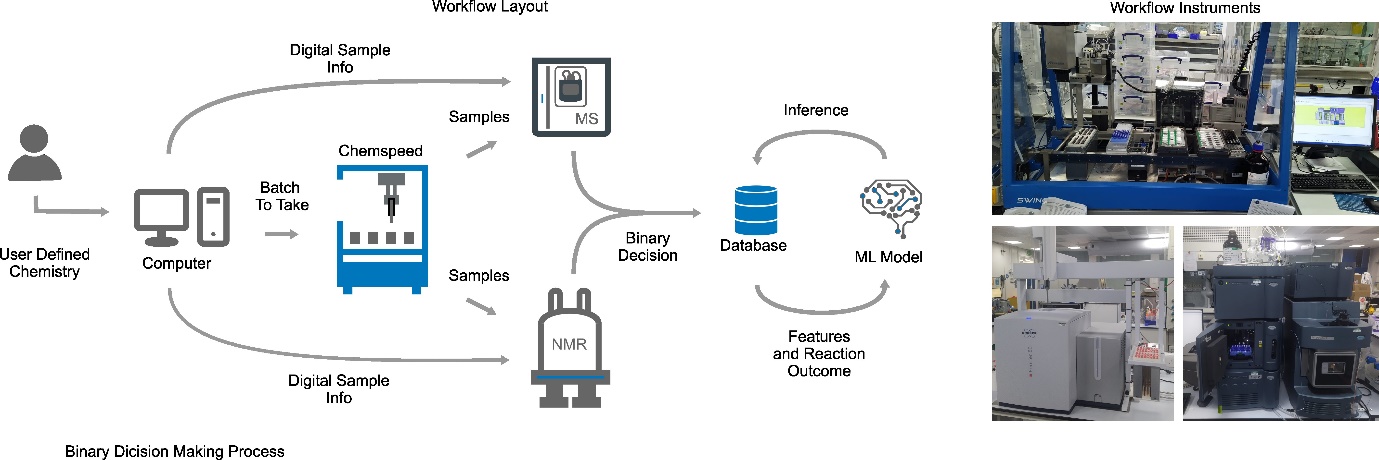


Figure 4: The overall workflow the programme manages shown as a sequence of steps. The logical order shown in the diagram is the same as the one taken out physically. There are four main pieces of equipment: a centralised computer to store data and manage tasks, a ChemSpeed platform to carry out reactions, an NMR and LCMS machine to analyse data.

Once the chemist digitally defines reaction, subdivisions, and reagents into the programme (Figure 2), the combination space to explore is generated by the programme. The space is then filtered by price and reagents. Early stages of workflows are prone to errors, hence starting with the cheapest reagents reduces project expenses. Reagent masses are measured manually, hence minimising the number of reagents to measure in a batch reduces manual labour.



After sorting the chemical space, standard reactions are added to quickly assess the quality of a batch. Only then are batches sent as digital samples to the robotic, NMR and LCMS platforms. Next, the robotic platform makes up the reactions and creates samples for the chemist to then transport to the NMR and LCMS machines, which thanks to the digital sample knows what sample corresponds to what combination. With analytical data acquired, the programme uses chemical shifts and peak positions from 1HNMR data to determine if a reaction has occurred, it then uses m/z ratios from mass spec (MS) data to determine the size of the supramolecular structure (it is possible that only imines are formed and not a complex, MS data helps identify such combinations). The final output of the algorithm is a binary classification (0,1) of what combinations yield discrete, diamagnetic, supramolecular structures.

Unfortunately, the analytical classification algorithm introduces bias to the dataset. It is only able to classify discrete, diamagnetic species. High spin iron (II) complexes, and oligomers will not be accounted for, although preliminary manually curated results demonstrated such products do exist in the paper’s chemical space. Fortunately, this is an algorithmic limitation, which could be solved by building better tools. The fundamental dataset remains unaffected. Future development would improve ML model predictions and will offer more robust insights into the dataset.

Once all reactions have been carried out and classified, it is stored in a local database to then be used for ML model training and SHAP analysis. This is the data exploration step in Figure 1.

## Data Exploration

Data still needs to be collected ☹. Talk about feature selection.

Feature selection was based on ideas from ‘edge-directed self-assembly’, that is, metals and organic ligands behave as vertices and shape edges respectively. Therefore, geometric information of the structure contained in the isolated building blocks can be used. On the other hand, to capture reactivity, more traditional features from ideas in chemistry are used, such as homo-lumo gaps. Embracing of the two allows the model to understand geometry based on chemical rules.

More specifically, since reactions contain a metal, an aldehyde, and an amine, features of reactions can be subdivided based on the reagents. Therefore, a supramolecular structure resulting from a reaction is represented by its building blocks (Figure 5) – the importance of such a paradigm has been highlighted in the introduction. The building blocks in turn can be further represented and their chemistries described. A few of these descriptors (Figure 5) are calculated via quantum mechanical means. Amines and aldehydes geometries were first optimized with B97-3c functionals and then energies were calculated with PWPB95 functionals. Both functionals were used as they are relatively less computationally expensive, while providing accurate results. This allowed calculations to be run on a mid-range laptop with Orca23. Other features and descriptors were extracted with rdkit.

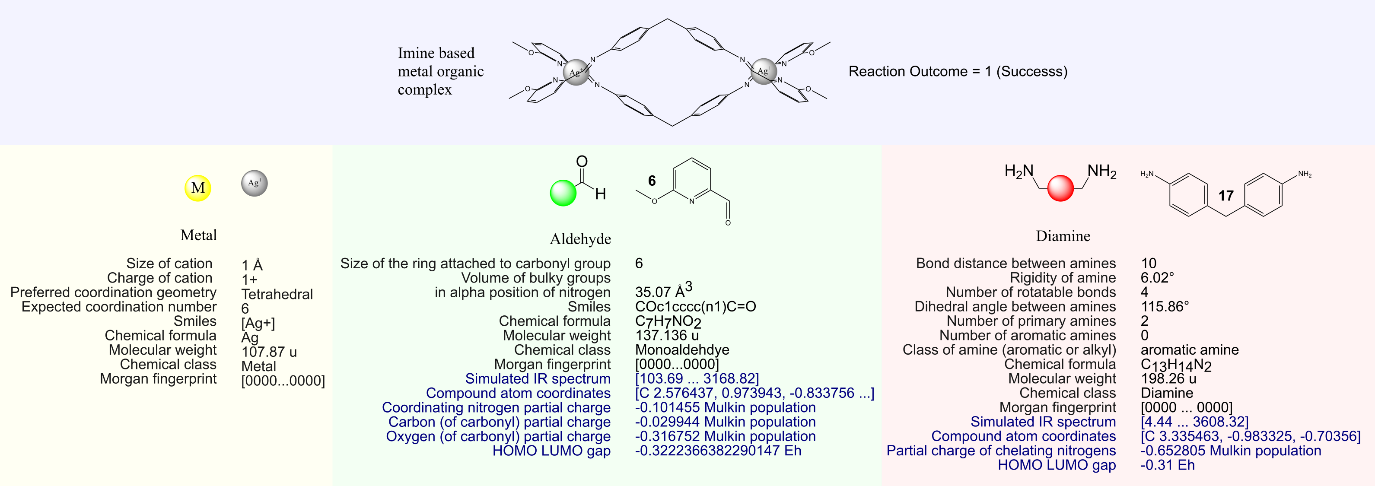


Figure 5: Decomposition of a reaction into reagents with their respective relevant features and representations used for data analysis. Example features and representations are given for reaction 109 (Silver tetrafluoroborate + 6-Methoxypyridine-2-carbaldehyde + 4,4'-Methylenedianiline). Reaction 109 was successful and the formed supramolecular structure is shown. Features in blue are calculated via quantum mechanical means in orca.

Generate table of different features, noting anything interesting about them.

Unfortunately, for this Master (MSci) project the final dataset only contains 200 datapoints, and due to SHAPs coalitions approach, limits the number of features that may be used in a SHAPLY analysis. To mitigate this, only 7 of the relevant features in Figure 5 were selected.

For metals the cation size was chosen as this indirectly contains information on the charge (greater charges result in smaller radii), and possible geometries (smaller radii limit the number of ligands and covalency). Cation sizes are expected to, therefore, affect topology and reactivity. Bennett et al. have highlighted the importance of cation radii in coordination to polypyridine ligands both in terms of chelation and preorganisation24.

For aldehydes, the thesis focuses on the partial charge on the carbon of the aldehyde, the volume of bulky groups in coordination sphere, and the ring size attached to the carbonyl. The author expects the partial charge on the carbonyl carbon to dictate reactivity with amines, where lower electron populations results in a better electrophilic centre for an attack from the amine. Furthermore, it is thought that substituent volumes constrain imine metal coordination and supramolecular vertex geometry. A bulkier group may affect preorganisation of the imine for a specific metal, introduce steric strain to two coordinating imines and if exaggerated may prevent another imine from coordinating. Lastly the authors think ring size provides further information on both supramolecular topology and imine reactivity. The size of the ring controls the number of double bonds conjugated to the imine and the carbonyl group, affecting their reactivities. A larger ring size also increases the steric presence of the imine in the coordinating sphere affecting coordination geometry.

Lastly for amines the selected feature are: the nature of the amine (if its bonded to an aromatic group or not), the enforced angles between the two amine, and the number of rotatable bonds. The authors expect the amine to behave as linkers, hence more features that encompass more geometric information we chosen. For similar reasons as the aldehydes, resonance of aromatics with amine groups affect the reactivity of both imines and amines. The dihedral angle between amines affects linker coordination direction, while the number of rotatable bonds provide information on the slack the amine can take affecting topology. For example, a completely rigid linear amine cannot possibly form a helicate. If the same linear amine has multiple rotatable bonds the compound may wrap about and possible form a helicate.

## X-ray Crystallographic data

After reaction combinations were sampled for NMR and MS, remaining solutions in the reaction vials were left to evaporate for 2 weeks. This makeshift method of crystal growth rarely worked, giving us only X-amount of samples. Supramolecular structures can be seen in FIGURE X.

# Spared information to be compiled into a coherent document.

## References

(1) Schrader, M. L.; Schäfer, F. R.; Schäfers, F.; Glorius, F. Bridging the Information Gap in Organic Chemical Reactions. *Nat. Chem.* **2024**, *16* (4), 491–498. https://doi.org/10.1038/s41557-024-01470-8.

(2) Nippa, D. F.; Atz, K.; Müller, A. T.; Wolfard, J.; Isert, C.; Binder, M.; Scheidegger, O.; Konrad, D. B.; Grether, U.; Martin, R. E.; Schneider, G. Identifying Opportunities for Late-Stage C-H Alkylation with High-Throughput Experimentation and in Silico Reaction Screening. *Commun. Chem.* **2023**, *6* (1), 1–11. https://doi.org/10.1038/s42004-023-01047-5.

(3) Zhang, D.; Ronson, T. K.; Nitschke, J. R. Functional Capsules via Subcomponent Self-Assembly. *Acc. Chem. Res.* **2018**, *51* (10), 2423–2436. https://doi.org/10.1021/acs.accounts.8b00303.

(4) Chakrabarty, R.; Mukherjee, P. S.; Stang, P. J. Supramolecular Coordination: Self-Assembly of Finite Two- and Three-Dimensional Ensembles. *Chem. Rev.* **2011**, *111* (11), 6810–6918. https://doi.org/10.1021/cr200077m.

(5) Noh, J.; Doan, H. A.; Job, H.; Robertson, L. A.; Zhang, L.; Assary, R. S.; Mueller, K.; Murugesan, V.; Liang, Y. An Integrated High-Throughput Robotic Platform and Active Learning Approach for Accelerated Discovery of Optimal Electrolyte Formulations. *Nat. Commun.* **2024**, *15* (1), 2757. https://doi.org/10.1038/s41467-024-47070-5.

(6) Cook, T. R.; Zheng, Y.-R.; Stang, P. J. Metal–Organic Frameworks and Self-Assembled Supramolecular Coordination Complexes: Comparing and Contrasting the Design, Synthesis, and Functionality of Metal–Organic Materials. *Chem. Rev.* **2013**, *113* (1), 734–777. https://doi.org/10.1021/cr3002824.

(7) Renom-Carrasco, M.; Lefort, L. Ligand Libraries for High Throughput Screening of Homogeneous Catalysts. *Chem. Soc. Rev.* **2018**, *47* (13), 5038–5060. https://doi.org/10.1039/C7CS00844A.

(8) Rulev, A. Y. Serendipity or the Art of Making Discoveries. *New J. Chem.* **2017**, *41* (11), 4262–4268. https://doi.org/10.1039/C7NJ00182G.

(9) *Exploration of the chemical space and its three historical regimes*. https://doi.org/10.1073/pnas.1816039116.

(10) Abolhasani, M.; Kumacheva, E. The Rise of Self-Driving Labs in Chemical and Materials Sciences. *Nat. Synth.* **2023**, *2* (6), 483–492. https://doi.org/10.1038/s44160-022-00231-0.

(11) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2.

(12) Caramelli, D.; Granda, J. M.; Mehr, S. H. M.; Cambié, D.; Henson, A. B.; Cronin, L. Discovering New Chemistry with an Autonomous Robotic Platform Driven by a Reactivity-Seeking Neural Network. *ACS Cent. Sci.* **2021**, *7* (11), 1821–1830. https://doi.org/10.1021/acscentsci.1c00435.

(13) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144* (11), 4819–4827. https://doi.org/10.1021/jacs.1c12005.

(14) Artrith, N.; Butler, K.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13*, 505–508. https://doi.org/10.1038/s41557-021-00716-z.

(15) Wang, K.; Jordan, J. H.; Hu, X.-Y.; Wang, L. Supramolecular Strategies for Controlling Reactivity within Confined Nanospaces. *Angew. Chem. Int. Ed.* **2020**, *59* (33), 13712–13721. https://doi.org/10.1002/anie.202000045.

(16) Ayme, J.-F.; Lehn, J.-M.; Bailly, C.; Karmazin, L. Simultaneous Generation of a [2 × 2] Grid-Like Complex and a Linear Double Helicate: A Three-Level Self-Sorting Process. *J. Am. Chem. Soc.* **2020**, *142* (12), 5819–5824. https://doi.org/10.1021/jacs.0c00896.

(17) Ayme, J.-F.; Beves, J. E.; Campbell, C. J.; Leigh, D. A. Probing the Dynamics of the Imine-Based Pentafoil Knot and Pentameric Circular Helicate Assembly. *J. Am. Chem. Soc.* **2019**, *141* (8), 3605–3612. https://doi.org/10.1021/jacs.8b12800.

(18) Troshin, K.; Hartwig, J. F. Snap Deconvolution: An Informatics Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **2017**, *357* (6347), 175–181. https://doi.org/10.1126/science.aan1568.

(19) van Dam, A.; van Schendel, R.; Gangarapu, S.; Zuilhof, H.; Smulders, M. M. J. DFT Study of Imine-Exchange Reactions in Iron(II)-Coordinated Pincers. *Chem. – Eur. J.* **2023**, *29* (61), e202301795. https://doi.org/10.1002/chem.202301795.

(20) *Supramolecular chemistry and self-assembly*. https://doi.org/10.1073/pnas.062524299.

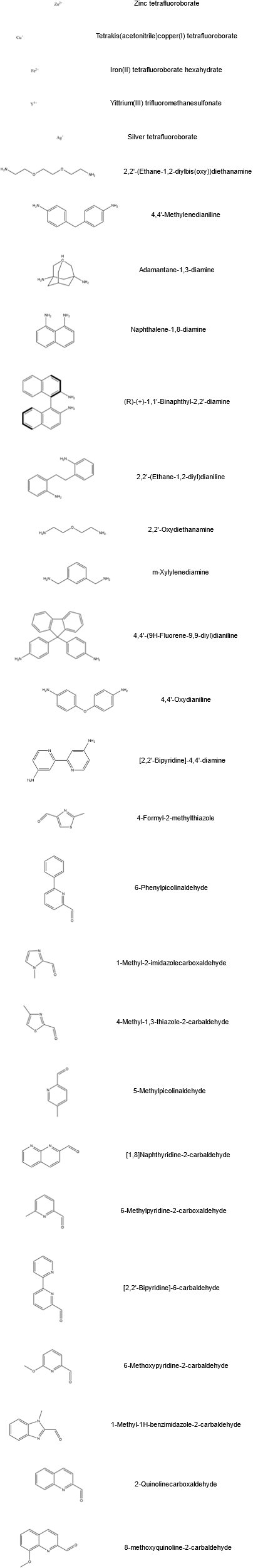
(21) Jelfs, K. E. Computational Modeling to Assist in the Discovery of Supramolecular Materials. *Ann. N. Y. Acad. Sci.* **2022**, *1518* (1), 106–119. https://doi.org/10.1111/nyas.14913.

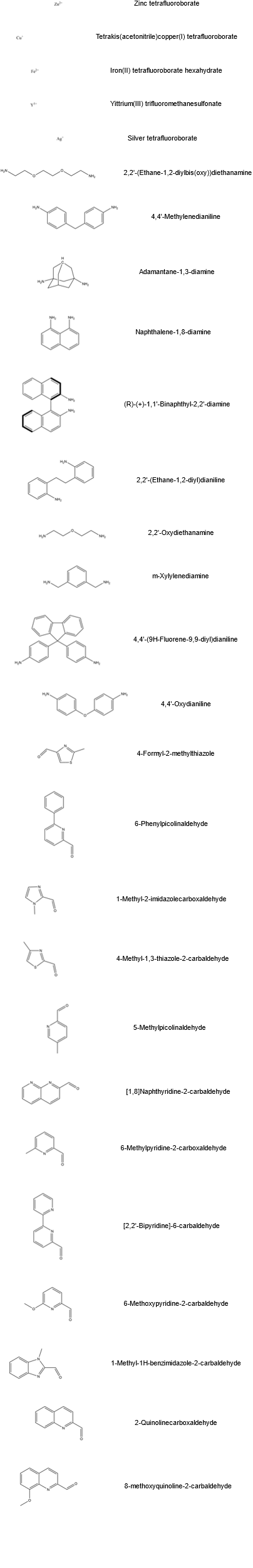
(22) Farrant, E. Automation of Synthesis in Medicinal Chemistry: Progress and Challenges. *ACS Med. Chem. Lett.* **2020**, *11* (8), 1506–1513. https://doi.org/10.1021/acsmedchemlett.0c00292.

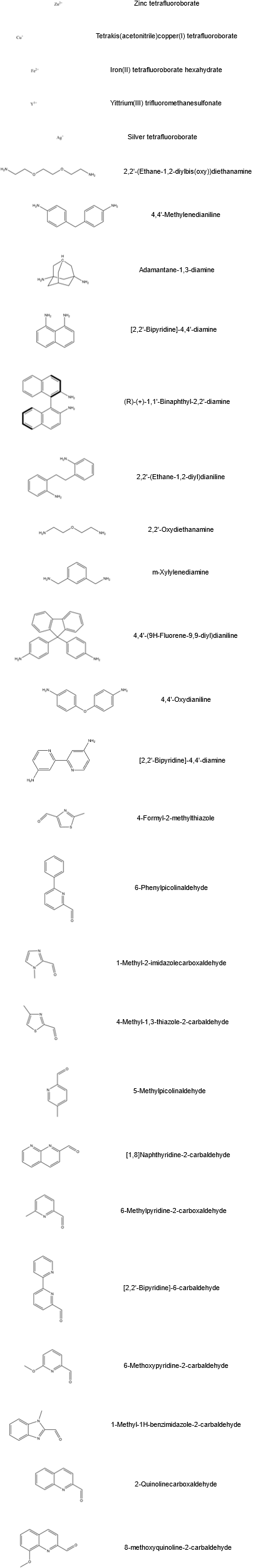
(23) Neese, F. (2012) The ORCA Program System, Wiley Interdiscip.

(24) Bennett, K. J.; Bates, A. L.; Carolan, A. N.; Jones, S. B.; Lee, H.-S.; Hancock, R. D. Complexation of Lanthanide(III) and Yttrium(III) Cations with Terpyridyl, Dipyridoacridine, and 1,10-Phenanthroline in Aqueous Solution. Importance of Covalence in Metal–Ligand Bonding in Complexes of Lanthanides with Nitrogen Donor Ligands. *Inorganica Chim. Acta* **2023**, *557*, 121669. https://doi.org/10.1016/j.ica.2023.121669.

# Supplementary Information







S 1: A table of reagent names and structures used in the study.