UNIVERSITY OF CALIFORNIA
RIVERSIDE

Computational Insights Into Rare-Earth Separations

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Tongyu Liu

December 2022

Dissertation Committee:
Dr. De-en Jiang, Chairperson
Dr. Gregory Beran
Dr. Francisco Zaera

The Dissertation of Tongyu Liu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

COPYRIGHT ACKNOWLEDGEMENT

The text and figures in Chapter 3, in part or full, are reproduced from "N-oxide ligands for selective separations of lanthanides: Insights from computation" *RSC Advances*, under review. The coauthor (Dr. De-en Jiang) directed and supervised this research.

The text and figures in Chapter 4, in part or full, are reproduced from "Advancing Rare-Earth Separation by Machine Learning", *JACS Au* 2022, 2 (6), 1428-1434. The coauthor (Dr. De-en Jiang) directed and supervised this research.

The text and figures in Chapter 7, in part or full, are reproduced from "All-Carboxylate-Protected Superatomic Silver Nanocluster with an Unprecedented Rhombohedral Ag_8 Core", *J. Am. Chem. Soc.* 2020, 142, 16905–16909. The coauthor (Dr. De-en Jiang) directed and supervised this research.

The text and figures in Chapter 8, in part or full, are reproduced from "Understanding the interaction between carboxylates and coinage metals from first principles", *J. Chem. Phys.* 2021, 155, 034301. The coauthor (Dr. De-en Jiang) directed and supervised this research.

ACKNOWLEDGEMENTS

I would like to first express my sincere gratitude to my advisor, Dr. De-en Jiang, for his invaluable mentorship and support during the past years. I have learned much from Dr. Jiang on how to conduct research with the highest standard. His teaching will continue to inspire me throughout my career.

I would like to acknowledge the collaborators for their work covered in this thesis, Dr. Santa Jansone-Popova, Dr. Alexander S. Ivanov, Dr. Ilja Popovs, and Dr. Katherine R. Johnson in Oak Ridge National Laboratory.

I would like to acknowledge my thesis committee members: Dr. Gregory Beran and Dr. Francisco Zaera for their guidance and patient support during my study. I would like to thank my committee and Dr. Vincent Lavallo and Dr. Sinisa Coh for their generous service on my oral qualifying exam.

I would like to thank my colleagues Dr. Dhileep Reddy, Dr. Lu Wang, Dr. Tao Wu, Dr. Kristen Wang Romero, Dr. Song Wang, Dr. Yangyunli Sun, Yuqing Fu, Chuanye Xiong, and Hyuna Kwon for their insightful discussion and collaboration.

I would like to thank my parents and my husband Dr. Yin Zhang for their wise counsel and wholehearted support.

ABSTRACT OF THE DISSERTATION

Computational Insights Into Rare-Earth Separations

by

Tongyu Liu

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2022
Dr. De-en Jiang, Chairperson

Constituting the bulk of rare-earth elements, lanthanides (Ln) need to be separated to fully realize their potential as critical materials in many important technologies. Solvent extraction is the most extensively used process to separate lanthanides on an industrial scale. Recently, preorganized ligands such as bis-lactam-1,10-phenanthroline show unique selectivity trends across the lanthanide series, indicating the synergistic effects of both N and O donors in complexing with lanthanides. So, we examined mixed N, O-donor ligands containing pyridinic N and N-oxide groups and evaluated their relative aqueous La(III)/Ln(III) selectivity by computing free energy changes for the exchange reaction between the designed ligands and a reference ligand via density functional theory (DFT). Three novel ligands show promise as excellent extractant agents in selectively separating trivalent lanthanides. However, no matter of our DFT investigations or other discovery of new ligands in rare-earth separations by solvent extraction, they are still largely based on trial and error, a low-throughput and inefficient approach. As a result, deep neural networks on the available experimental data of distribution coefficients measured for hundreds of

ligands for 14 Ln(III) ions have been trained to accurately and quickly predict their distribution coefficients for a given ligand and the extraction conditions. Four new-synthesized ligands were found that their predicted distribution coefficients from our trained machine-learning model match well with their experimental values. Then, this trained model was applied for some large ligand databases automatically generated by molecule generation tools based on string-based representations. Several hexalkyl-nitrilotriacetamide ligands were screened out with high potential in selective rare-earth separations. Therefore, our machine-learning approach paves the way for accelerating the discovery of new ligands for rare-earth separations. In additional to rare-earth elements, some theoretical insights towards atomically precise ligand-protected nanoclusters by first principles were performed. All works in this dissertation aim at chemically understanding the interactions between metals and organic ligands by different computational approaches.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Introduction of rare earth elements and their applications

Rare earth elements (REE) are the group of 17 chemical elements of the periodic tables, namely the fifteen lanthanides and the Yttrium and the Scandium, as defined in the International Union of Pure and Applied Chemistry (IUPAC). Scandium (Sc) and Yttrium (Y) are considered REE because they are usually found in the exactly same deposits of ore as the lanthanides, and they have similar physical and chemical properties. These 17 elements are furtherly classified into light REE (Sc and La-Eu) and heavy REE (Y and Gd-Lu).

The applications of REE and its alloys have risen significantly during the past three decades in a variety of technologies, as summarized in Table 1-1, including but not limited to autocatalytic converters, computer memory, fluorescent materials, glass additives, LED lighting, mobile phones, rechargeable batteries, superconductors, and super magnets. They are known as the vitamins of the contemporary industry since they are the most vital components of all sophisticated technology. Due to their special characteristics, including their physical, chemical, magnetic, and luminescence, they contribute to a variety of technical benefits, including a lower energy need, greater efficiency, a smaller size, and a quicker. Their need for green technology, which is more effective, lighter, and smaller equipment, has increased in recent years. Even better and smaller analytical tools are made possible by these technologies. The usage of REE will continue in the foreseeable future due to the expansion of green technology.

Table 1-1. Industrial applications of REEs.[1]

Area	Applications
Electronics	Television screens, computers, cell phones, silicon chips, monitor displays, long-life rechargeable batteries, camera lenses, light emitting diodes (LEDs), compact fluorescent lamps (CFLs), baggage scanners, marine propulsion systems.
Manufacturing	High-strength magnets, metal alloys, stress gauges, ceramic pigments, colorants in glassware, chemical oxidizing agent, polishing powders, plastics creation, as additives for strengthening other metals, automotive catalytic converters.
Medical Science	Portable X-ray machines, X-ray tubes, magnetic resonance imagery (MRI) contrast agents, nuclear medicine imaging, cancer treatment applications, and for genetic screening tests, medical and dental lasers.
Technology	Lasers, optical glass, fiber optics, masers, radar detection devices, nuclear fuel rods, mercury-vapor lamps, highly reflective glass, computer memory, nuclear batteries, high temperature superconductors.
Renewable Energy	Hybrid automobiles, wind turbines, next-generation rechargeable batteries, biofuel catalysts.

Not all REEs are found in nature in the form of pure metal. The estimated average concentration of REE in the earth's crust is between 130 and 240 μg/g, which is really much greater than the average concentration of most frequently used elements. Because of the significance of REE in applications as opposed to their erratic supply and challenging to discover alternatives, the separation of REE has been identified as one of the seven chemical separations to transform the world.

1.2 Solvent extractions in the separation of rare earth elements

High purity levels of individual rare-earth compounds or metals are needed for high-tech applications and new materials; as a result, mixed rare earths obtained from rare-earth concentrates must first be separated into pure or highly pure individual rare-earth compounds before being used in high-tech applications. It is exceedingly challenging to separate and create high-purity individual rare-earth element products from the mixed rare earths using simple fractional crystallization or precipitation because the features of the physics and chemistry of the nearby rare-earth elements are comparable. Solvent extraction-separation is the most efficient, cost-effective, and practical method of separating mixed rare earths into separate rare-earth elements with high-purity results. In lanthanide atoms, the 4f orbitals are gradually filled as the atomic number increases, although the arrangement of the valence electrons in the outermost shell is the same for all species. The remarkably comparable physical and chemical characteristics of the elements result from the screening of the 4f orbitals. The so-called "lanthanide contraction," in which the ionic radius gradually reduces from La^{3+} (1.06 Å) to Lu^{3+} (0.85 Å), is another comparable effect.

Solvent extraction is liquid-liquid extraction, which means that the solute is transferred from one solvent to another due to the difference in solubility or distribution coefficient between two immiscible or slightly soluble solvents, and after repeated extractions, so that the vast majority of solutes can be extracted to achieve enrichment. The solvent extraction method has the advantages of large processing capacity, good separation effect, high recovery rate, fast equilibrium speed, simple equipment, and high automation.[2]

The disadvantage is that the price of the extractant is high and the acid consumption is large. Since the 1970s, with the rapid development of organic chemistry, inorganic chemistry, petrochemical and other fields, it has brought a good development environment for the development and application of new, efficient and low-cost extractants. The solvent extraction method is used to separate, purify, and concentrate metal ions worldwide, and it has been widely used in the purification of single rare earth elements.[3] According to the differences in the chemical properties of the extractants, they are mainly divided into acidic extraction systems, neutral extraction systems, basic amine extraction systems and ionic liquid systems, etc.

Acidic extractant has better extraction and separation performance, so it is widely used in rare earth separation industry. Commonly used acidic extractants are: organic phosphoric acid extractants P204, P507 and Cyanex272; carboxylic acid extractants CA12, CA100 and naphthenic acid. In industry, organic phosphoric acid extractants P507, P204 and Cyanex272 are mainly used to extract, group and separate all rare earths, and organic carboxylic acid extractants CA12 and naphthenic acid are used to extract and separate yttrium. P204 is mainly used for the separation and purification of light rare earths, and P507 is mainly used for the separation and purification of heavy rare earths, but they have the problems of small separation coefficient, low selectivity and large strip acidity. Due to the steric hindrance effect of its own structure, Cyanex272 has a larger separation coefficient for rare earth extraction, so the selectivity is better than that of P204 and P507, and the stripping is easier. Cyanex272 also has its own shortcomings, such as small extraction volume and easy emulsification during extraction. At present, the separation

process of solvent extraction method is very mature, but the existing extraction system also has obvious shortcomings. Bontha et al. conducted the first study on the separation of dysprosium using PC88A in a phosphoric acid system, and found that the extraction and combination of organic phase dysprosium.[4] According to PC88A, rare earth extraction has different degrees of selectivity depending on the acidity of the solution and the concentration of the extractant.

The neutral extraction system can be divided into neutral phosphorus extractant, neutral oxygen-containing extractant, neutral sulfur-containing extractant and substituted amide extractant. Common extractants are tributyl phosphate (TBP), Cyanex921, TOPO and P350.[5] Phosphorus extractant is the most widely used in industry (Table 1-2), which has the advantages of good extraction performance and low acidity in stripping. During the extraction process, the phosphoryl group of the extractant forms coordination with rare earth, and the rare earth is separated and enriched from the rare earth leaching solution.

Amine extractants use nitrogen atoms as extraction functional groups to extract rare earths. The nature of the substituents on the nitrogen atoms determines the pH of the extractants. The substituents of such extractants are usually long-chain alkyl groups, increasing their alkalinity become stronger. At present, amine extractants that are widely used in the extraction and separation of rare earths include primary amine (N1923), secondary amine, tertiary amine and quaternary amine salts. Amine extractants have the advantages of strong extraction performance, good selectivity and large extraction capacity. However, at present, amide extractants have disadvantages such as complex

synthesis processes, small extraction load, difficult recovery, and the need for a large amount of salting-out agent, and are gradually being replaced at present.

Table 1-2. Common extractants used in rare-earth separation industry[6]

Industrial extractant	shortened	Molecular formula
Di(2-ethylhexyl)phosphoric acid	P204	$(C_8H_{17}O)_2P(O)OH$
2-Ethylhexyl 2-ethylhexylphosphonic acid	P507 or PC88A	$(C_8H_{17}O)_2P(O)OC_8H_{17}(OH)$
Cyanex 272	Cyanex 272	$R_2PO(OH)$
Tributyl phosphate	TBP	$(C_4H_9O)_3PO$
Di(1-methylhepthl)methylphosphonate	P350	$(C_8H_{17})_3PO$
Tri-n-octylamine (Alamine-336)	N235	$(C_nH_{2n+1})_3N$ n = 8–10
Aliquat 336	N263	$[CH_3\text{-}N\text{-}(C_{8-10}H_{17-21})_3]^+ Cl^-$
Primary amine N1923		$(C_nH_{2n+1})_2CHNH_2$ n = 9–11

Many extractants have the potential to be poisonous and combustible and to generate secondary contamination as shown in Figure 1-1. The current focus of solvent extraction research is on selective extractants and organic solvent optimization to increase enrichment factors and separation efficiency. In order to increase separation efficiency and enrichment factors, we think that the present focus of solvent extraction research should be on developing green extractants with high separation selectivity. Studying the methods for recovering and reusing the vast amounts of organic solvent waste liquid that the current process produces.

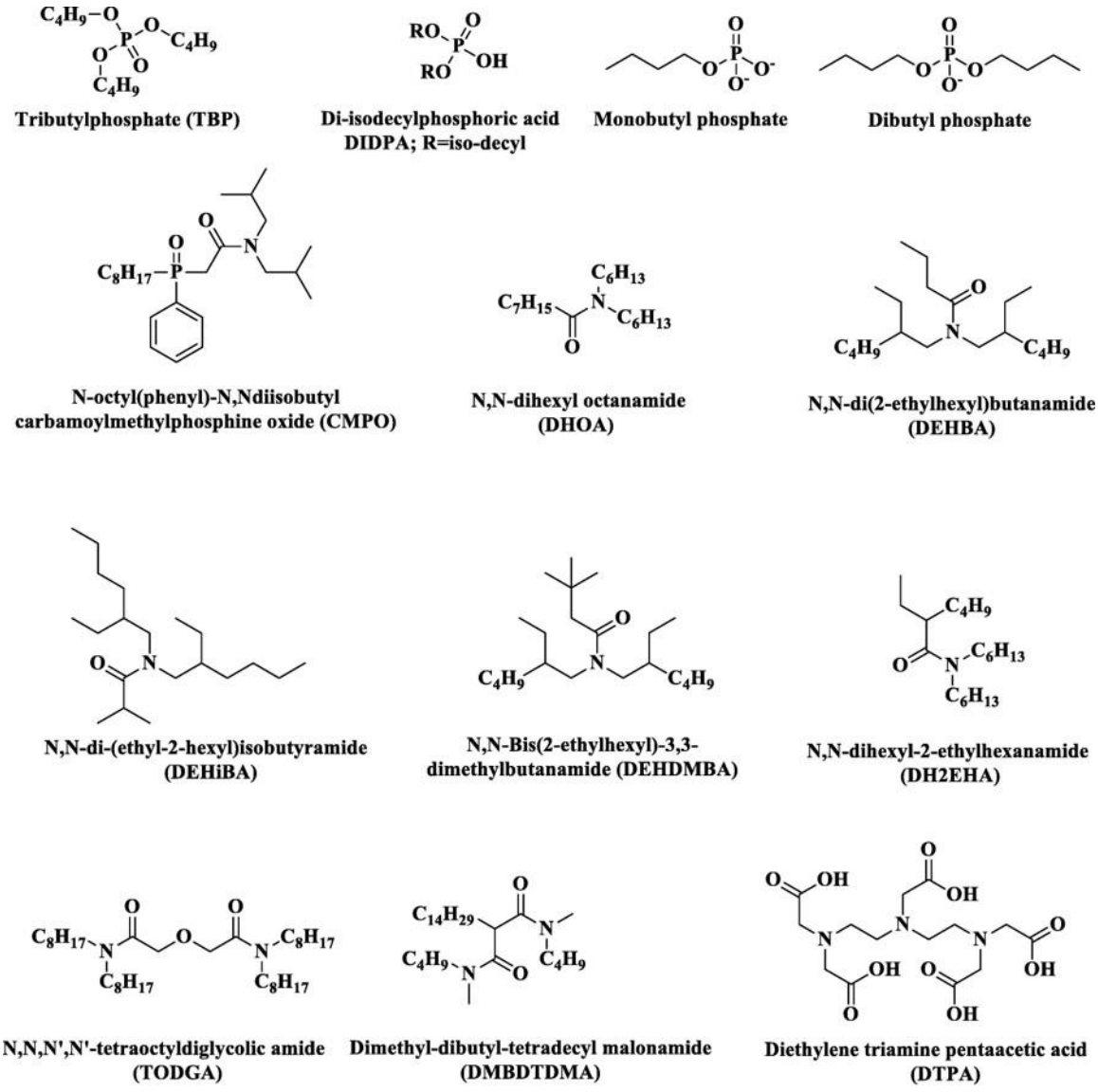


Figure 1-1. Natural ligands used in REE separations at laboratory scale in literature.[7]

1.3 Recent research on selective extractants

1.3.1 New types of extractants

A series of ligands combing soft and hard donors based on the rigid phenanthroline skeleton has been a strategy of promising extractants for actinide (An) and lanthanide (Ln) selective separation since 2014.[8-12] Due to a mount of similarities between Ln and An, especially their trivalent ions, more investigations on phenanthroline-derived ligands in Ln(III) selective separations have been performed recently. Jansone-Popova et al. reported a separation protocol that employs shape-persistent 2,9-bis-lactam-1,10-phenanthroline (BLPhen) ligands exhibiting unparalleled selectivity for light trivalent lanthanides in 2019.[13] As shown in Figure 1-2a, by using BLPhen ligand **1** in a nitric environment, an impressive extraction efficiency over the lanthanide family was obtained. Even in the presence of competing transition metals, the highly preorganized binding pockets of the ligands enabled the high-fidelity separation of lanthanides in a biphasic separation system. Notably, by changing the molecular stiffness of the extractant, the selectivity trends of the BLPhen ligands towards metal ions over the lanthanide series may be chemically controlled. In 2020, Xiao et al. reported extraction properties of tetrabutyl-(1,10-phenanthroline-2,9-diyl)bis(phosphonate) (C4-POPhen) ligand on La(III), Eu(III) and Lu(III) in different solvents as shown in Figure 1-2b.[14]

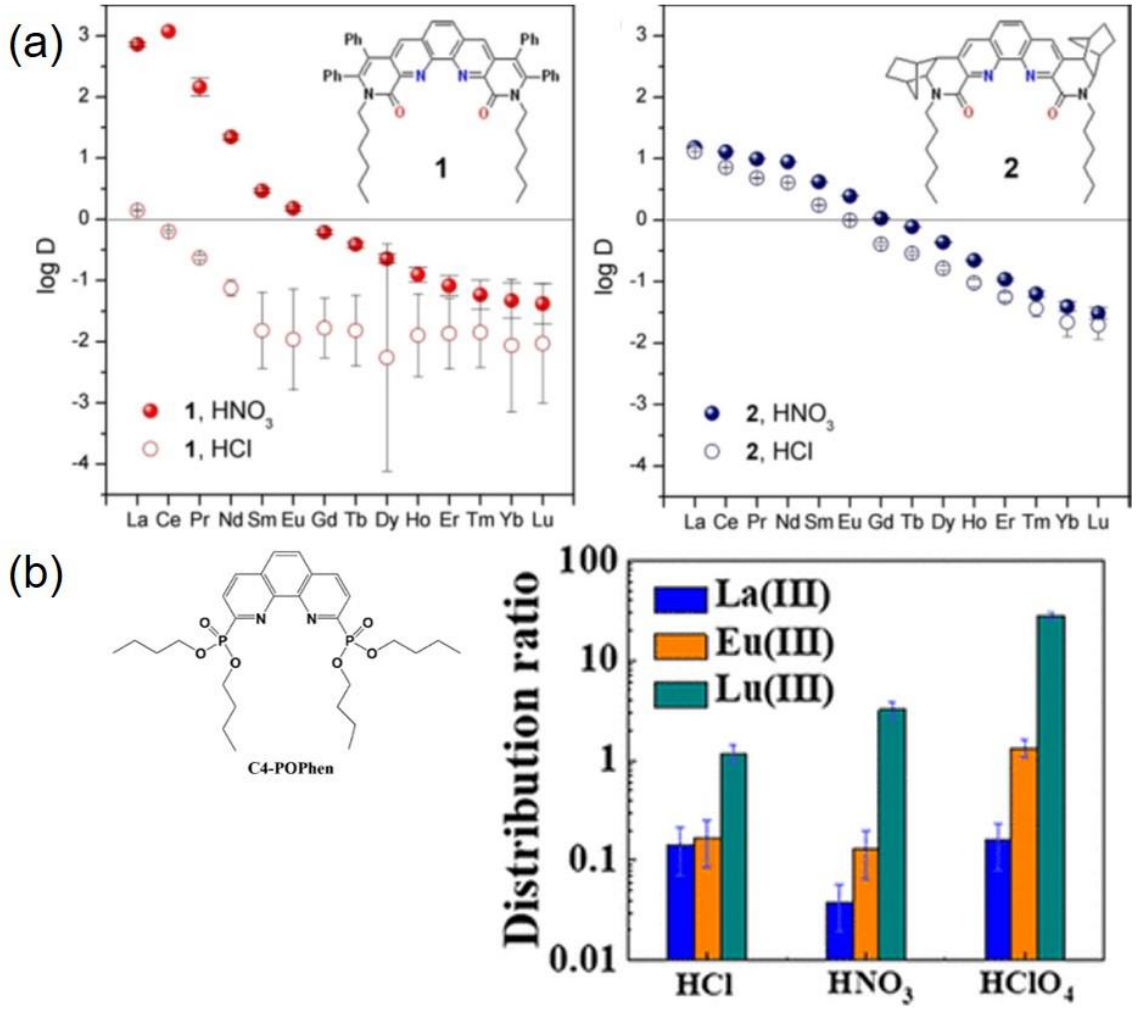


Figure 1-2. (a) Jansone-Popova et al. reported variation of logD in the extraction of Ln(III), by BLPhen ligand **1** and BLPhen ligand **2** from 0.9M HNO_3 and 0.9M HCl media into 1,2-dichloroethane at 25°C after 25 h.[13] (b) Xiao et al. reported extraction results of Ln(III) (La, Eu, and Lu) with C4-POPhen in n-octanol as a function of different inorganic acids.[14]

1.3.2 Structure-function relationship

In addition to exploring the new strategy of selective ligands in solvent extractions, understanding the structure-function relationship on the known ligands to optimize their structure is another thought for higher selectivity. Diglycolamide (DGA) ligand, a widely used ligand at the laboratory scale, has been shown that the shortening of N,N′-alkyl substituents leads to an improved extraction efficiency of lanthanides(III) according to the invitations on three n-alkyl substituted DGA ligands.[15] However, other factors such as symmetry, side chain, heteroatom, etc., have not been examined due to the limitation of synthesis speed. Recently, Jansone-Popova et al. reported the syntheses and measurements on *D* values of light Ln(III) for 15 DGA ligands (Table 1-3).[16] Their study reveals that reduced extraction of lanthanides(III) is caused by increased crowding at the metal-ion binding site, which interferes with the effective accommodation of counterions in a secondary coordination shell, when the *N,N′*-alkyl group size is increased. Importantly, the better lanthanide extraction is caused by moving the branching sites on the *N,N′*-alkyl substituents farther away from the amide functional groups (III). Nevertheless, similar study combing both experimental measurements and theoretical simulations are still limited by synthesis routes, computational cost, and other complicatedness of lanthanides.

Table 1-3. Dependence of alkyl substituents' length and size of DGA on the separation of La(III) and Sm(III).[16]

N,N-substituent	N′,N′-substituent	D_{La}	D_{Sm}	$SF_{Sm/La}$
n-octyl	n-octyl	0.0 ± 0.47	5.99 ± 0.07	90 ± 39
2-ethylhexyl	2-ethylhexyl	0.04 ± 0.27	0.11 ± 0.11	2.9 ± 0.4
methyl	n-octyl	4.13 ± 0.02	259 ± 0.2	63 ± 8
methyl	2-ethylhexyl	0.7 ± 0.04	40.9 ± 0.3	56 ± 13
methyl	4-butyldecyl	0.14 ± 0.06	2.66 ± 0.05	20 ± 0
methyl	5,9-dimethyldecyl	0.66 ± 0.04	8.37 ± 0.20	13 ± 2
ethyl	n-octyl	0.73 ± 0.04	61.9 ± 0.1	85 ± 4
propyl	n-octyl	0.34 ± 0.07	38.5 ± 0.1	114 ± 5
n-octyl, methyl	n-octyl, methyl	3.69 ± 0.04	$>3 \times 10^3$	$>1 \times 10^3$
n-octyl	n-octyl, methyl	0.34 ± 0.10	23.2 ± 0.1	69 ± 2
	2-methyldec-2anyl	1.93 ± 0.07	21.5 ± 0.0	11 ± 0
n-octyl	n-decyl	0.08 ± 0.07	11.1 ± 0.0	146 ± 14
n-octyl	3,5-trimethylhexyl	0.05 ± 0.07	1.52 ± 0.02	32 ± 2
n-octyl	4-butyldecyl	0.00 ± 0.86	0.67 ± 0.03	211 ± 103
n-octyl	5,9-dimethyldecyl	0.040 ± 0.07	4.21 ± 0.02	97 ± 4

1.4 Scientific questions and challenge

Although solvent extraction is the best commercial method for separating rare earths, its separation efficiency is limited, and many extractions must be performed in order to obtain high purity.[17-19] As a result, more efficient extractants in trivalent lanthanides selective separations are required. There are two approaches to reaching this long-term aim at extremely strong and selective ligands for trivalent lanthanides. One is designing a new type of molecule, the other is the modification on the substituent of a known ligand. An accurate simulation on the selectivity of a given ligand could significantly accelerate the exploration on new ligands that synthesis and measurement could focus on the candidate with computational high selective. In addition, plenty of simulations on ligands' selectivity according to different structures would conclude more structure-function relationships to help further design. Hence, we have two imperative questions to solve for design, evaluate and comprehend new selective ligands in lanthanide separations: 1) how do different simulation methodologies at different scales evaluate the selectivity of a ligand, using thermodynamical constants relating to the selectivity or directly simulating the extraction property. 2) how to use simulated results to expand the diversity of selective extracts.

This dissertation presents our several works related to the design of strong and selective ligands by different computational methods, new type of phenanthroline-derived ligands by density functional theory simulations, new machine learning approach on predictions of extraction property, combing machine learning model and new ligands for high-throughput screenings and new inspirations from Ln(III)/An(III) to lanthanide separations.

Reference

1. V. Balaram, *Geosci. Front.*, 2019, **10**, 1285-1303.

2. W. Wang and C. Y. Cheng, *Journal of Chemical Technology & Biotechnology*, 2011, **86**, 1237-1246.

3. E. Polyakov and A. Sibilev, *Theor. Found. Chem. Eng.*, 2016, **50**, 607-612.

4. R. Surampally, N. K. Batchu, L. K. Mannepalli and R. R. Bontha, *Journal of the Taiwan Institute of Chemical Engineers*, 2012, **43**, 839-844.

5. Z. Zhu and C. Y. Cheng, *Hydrometallurgy*, 2011, **107**, 1-12.

6. D. Qi, in *Hydrometallurgy of Rare Earths*, ed. D. Qi, Elsevier, 2018, pp. 187-389.

7. K. De Jesus, R. Rodriguez, D. L. Baek, R. V. Fox, S. Pashikanti and K. Sharma, *J. Mol. Liq.*, 2021, **336**, 116006.

8. C.-L. Xiao, C.-Z. Wang, L.-Y. Yuan, B. Li, H. He, S. Wang, Y.-L. Zhao, Z.-F. Chai and W.-Q. Shi, *Inorg. Chem.*, 2014, **53**, 1712-1720.

9. S. Jansone-Popova, A. S. Ivanov, V. S. Bryantsev, F. V. Sloop Jr, R. Custelcean, I. Popovs, M. M. Dekarske and B. A. Moyer, *Inorg. Chem.*, 2017, **56**, 5911-5917.

10. M. Alyapyshev, J. Ashina, D. Dar'In, E. Kenf, D. Kirsanov, L. Tkachenko, A. Legin, G. Starova and V. Babain, *RSC Adv.*, 2016, **6**, 68642-68652.

11. N. E. Borisova, T. B. Sumyanova, A. V. Kharcheva, P. I. Matveev, A. V. Ivanov, E. A. Razumova and S. V. Patsaeva, *Dalton Trans.*, 2018, **47**, 16755-16765.

12. L. Xu, N. Pu, Y. Li, P. Wei, T. Sun, C. Xiao, J. Chen and C. Xu, *Inorg. Chem.*, 2019, **58**, 4420-4430.

13. M. R. Healy, A. S. Ivanov, Y. Karslyan, V. S. Bryantsev, B. A. Moyer and S. Jansone-Popova, *Chem. Eur. J.*, 2019, **25**, 6326-6331.

14. X. Yang, L. Xu, Y. Hao, R. Meng, X. Zhang, L. Lei and C. Xiao, *Inorg. Chem.*, 2020, **59**, 17453-17463.

15. X. Peng, Y. Li, J. Ma, Y. Cui and G. Sun, *J. Radioanal. Nucl. Chem.*, 2017, **313**, 327-332.

16. D. Stamberga, M. R. Healy, V. S. Bryantsev, C. Albisser, Y. Karslyan, B. Reinhart, A. Paulenova, M. Foster, I. Popovs, K. Lyon, B. A. Moyer and S. Jansone-Popova, *Inorg. Chem.*, 2020, **59**, 17620-17630.

17. B. Dewulf, N. K. Batchu and K. Binnemans, *ACS Sustainable Chemistry & Engineering*, 2020, **8**, 19032-19039.

18. Y. Dong, X. Sun, Y. Wang, C. Huang and Z. Zhao, *ACS Sustainable Chemistry & Engineering*, 2016, **4**, 1573-1580.

19. J. E. Quinn, K. H. Soldenhoff and G. W. Stevens, *Hydrometallurgy*, 2017, **169**, 621-628.

Chapter 2. Computational Methods

The computational methods employed in this thesis will be briefly introduced in this chapter, along with their primary functions. In the chapters that follow, the computational methods employed in each chapter will be explored in detail.

2.1 Density functional theory (DFT)

2.1.1 Basic theories

Density Functional Theory (DFT), a popular and effective approach, has improved how electron-electron interactions are treated while building on the achievements of Hartree-Fock theory. In DFT, the electron density rather than the wavefunction itself is used to calculate the locations of atomic nuclei and the ground state energy of a system. A many-electron system's electron density in DFT is:

$$\rho(\mathbf{r}) = n \int \Psi^*(\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n}) \Psi(\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n})\, d\sigma_1 d\boldsymbol{X_n} \qquad \text{(Eq. 2.1)}$$

And the total energy of the system in terms of the electron density is:

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{ne}\rho(\mathbf{r}) + J[\rho(\mathbf{r})] + K[\rho(\mathbf{r})] \qquad \text{(Eq. 2.2)}$$

The energy is a function of a function because it depends on density, which in turn depends on the spatial coordinates of the electrons. The theory's name comes from this functional, which is referred to as the energy density functional and is shown by square brackets i.e. $E[\rho(r)]$. Instead of solving for the system with $3n$ variables, where n is the number of electrons, calculations on the density of a system, as opposed to the

wavefunction, are easier (and hence faster). This is because the electron density only depends on three variables.

Modern DFT begins with the Kohn-Sham equations, which take into account a fictional non-interacting system of electrons with a ground state density that is exactly the same as the real, fully interacting system of electrons. By taking into account this hypothetical system, the difficulty shifts from locating the universal functional to locating a hypothetical system with the same density as the system with interacting particles. In order to determine the kinetic energy, orbitals are introduced in the Kohn-Sham equations. The exchange and correlation functionals are only needed as approximations as a result.

The kinetic energy of the non-interacting system, T_{NI}, must be increased through interactions in order to transfer the energy of the non-interacting system onto the interacting system. These comprise Coulombic interactions, V_{ne}, interactions between nuclei and electrons, J, and an exchange-correlation component, E_{xc}, which contains the negligible kinetic energy left over from particle interactions and exchange-correlation interactions between electrons. The Kohn-Sham DFT energy, E_{KS}, can be expressed generally as:

$$E_{KS}[\rho(\mathbf{r})] = T_{NI}[\rho(\mathbf{r})] + V_{ne}\rho(\mathbf{r}) + J[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \qquad \text{(Eq. 2.3)}$$

The exchange-correlation term is the sole unidentified term in this equation for E_{KS}. Then, Eq. 2.3 can be resolved using Kohn-Sham equations, which are Schrödinger-like equations:

$$\left(-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\right)\varphi_i(\boldsymbol{r}) = \varepsilon_i\varphi_i(\mathbf{r}) \qquad \text{(Eq. 2.4)}$$

In these equations, ε_i is the energy associated with the orbital φ_i and V_{KS} is the Kohn-Sham potential, which is the necessary effective potential to produce an electron density that is similar to that of a system with interacting particles. The Kohn-Sham orbitals

that are the answers to these equations are denoted by the symbol $\varphi_i^{\mathrm{KS}}(\mathbf{r})$, and the sum of the square moduli of the occupied Kohn-Sham orbitals determines the overall electron density:

$$\rho(\mathbf{r}) = \sum_i |\varphi_i^{KS}(\mathbf{r})|^2 \qquad \text{(Eq. 2.5)}$$

The SCF approach, as used in Hartree-Fock theory, may be used to solve these Kohn-Sham equations using a starting set of molecular orbitals. Hartree-Fock theory and Kohn-Sham DFT are quite similar, but they vary in that Kohn-Sham DFT is an accurate technique while Hartree-Fock theory is an approximation, assuming that the precise form of E_{xc} is known and that the electron density can be described using a single electronic configuration. E_{xc} must, however, be approximated in real life.

The exchange-correlation functional (xc-functional) has several approximations that may be used to compute the ground state energies of molecular systems with varying degrees of precision. The simplest form of exchange-correlation functional, and the first rung on the Jacob's Ladder, is the Local Density Approximation (LDA),[1, 2] in which the exchange-correlation energy at a given point is equal to that of a uniform electron gas of the same electron density, for which E_{xc} is known (for all intents and purposes) exactly, as follows:

$$E_{XC}^{LDA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{XC}[\rho(\mathbf{r})]\, d\mathbf{r} \qquad \text{(Eq. 2.6)}$$

Where xc stands for the energy density of exchange-correlation. The LDA functional has a number of drawbacks, including a tendency to over bind atoms in molecules and a preference for more homogeneous electron densities than the precise density.[1, 2]

The Generalized Gradient Approximation (GGA) and meta-GGA functionals are the next two steps up from LDA because they express the exchange-correlation energy not only as a function of the density at a specific point in space but also as a function of the gradient and higher derivatives of the density. This improves LDA. GGA functionals take into account the first derivative, $\nabla\rho$, whereas meta-GGA functionals take into account the second, $\nabla^2\rho$, as follows:

$$E_{XC}^{GGA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{XC}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\,]\,d\mathbf{r} \qquad \text{(Eq. 2.7)}$$

$$E_{XC}^{meta-GGA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{XC}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \nabla^2\rho(\mathbf{r})]\,d\mathbf{r} \qquad \text{(Eq. 2.8)}$$

The exchange-correlation energy is calculated via a variety of different GGA functionals, each of which derives its parameters from first principles or semi-empirically from experimental data, such as atomization energies. Examples of such functionals include the B88P86 (BP) functional, which combines the P86 correlation functional and the B88 exchange functional developed by Perdew and Wang. Other examples include the non-empirical Perdew-Burke-Ernzerhof (PBE) functional,[3, 4] which is a refinement of earlier work by Perdew et al. and includes the BP functional, and the semi-empirical BLYP functional,[5] which is named after Becke for the B88 exchange part and Lee, Yang and Parr for the correlation part, which was parameterized using the correlation energy of the helium atom. While not as effective for other qualities, GGA functionals perform well for the prediction of structures. The meta-GGA functionals, which are on the following rung and are less frequent than the GGA functionals, provide a marginal improvement over the former at a negligible additional cost to computation. The non-empirical functional TPSS,

which bears the names Tao, Perdew, Staroverov, and Scuseria, is an illustration of a well-known meta-GGA functional.[3, 6] It was created in an effort to improve computation accuracy.

In the Hartree-Fock approximation, the exchange contribution to the exchange-correlation energy, which is approximated in the "pure" xc-functionals (such as PBE, BLYP), is accurate. A portion of this Hartree-Fock exact exchange energy is present in hybrid-GGA functionals, which are the fourth step on the ladder.

The semi-empirical B3LYP functional[5, 7-10] and the non-empirical PBE0 functional[3, 4, 11] are two examples of common hybrid functionals. Exact exchange, LDA and GGA (B88) exchange, and LDA and GGA (LYP) correlation are all included in the B3LYP functional:

$$E_{XC}^{B3LYP} = (1-a)E_X^{LDA} + aE_X^{HF} + b\Delta E_X^{B88} + (1+c)E_c^{LDA} + cE_c^{LYP} \qquad \text{(Eq. 2.9)}$$

The three parameters, *a*, *b*, and *c*, are 0.2, 0.7, and 0.8, respectively, and they are fit to experimental data. Instead, using perturbation theory, the PBE0 functional calculates the exact exchange contribution and combines exchange from the PBE functional with 25% of the Hartree-Fock exact exchange energy as follows:

$$E_{XC}^{PBE0} = 0.25E_X^{HF} + 0.75E_X^{PBE} + E_c^{PBE} \qquad \text{(Eq. 2.10)}$$

2.1.2 Relativistic effective core potentials for lanthanides

While relativistic effects on lighter atoms are minimal, they become more significant when the electron's speed approaches the speed of light as they approach c. The radial velocity for the innermost electrons rises to a considerable fraction of c in heavier

atoms, among which the lanthanides and actinides investigated in this work are most certainly counted, and these effects become noticeable. Effective core potentials (ECPs) address the valence electrons explicitly while substituting a number of an atom's core electrons with a potential field. Because lanthanides and other heavy elements have a large number of core electrons, using pseudopotentials can significantly lower the cost of computations involving these elements. The f-electrons can be treated explicitly in pseudopotentials for the f-block elements, or they can be included in the core. The latter, known as "f-in-core" ECPs,[12, 13] reduces computing cost at the price of accuracy. However, while the 5f orbitals of the actinides are more engaged in chemical bonding than the 4f orbitals of the lanthanides, which are typically thought of as 'core-like' and not directly involved in chemical bonding, care must be taken when employing these f-in-core ECPs for the lanthanides. [12, 13]

Additionally, relativistic effective core potentials (RECPs), which incorporate an implicit treatment of relativistic effects into the potential, offer an alternate approach for dealing with the impacts of relativity. Since the core electrons are primarily affected by direct relativistic effects, RECPs permit the employment of a non-relativistic Hamiltonian for the valence electrons and offer indirect relativistic effects for the valence electrons. It has been proposed that RECPs are more accurate than the majority of popular scalar relativistic techniques.[14] Two commonly used RECPs for lanthanides are large-core (LC) and Stuttgart small-core (SSC) RECPs. Large-core RECPs put 5s, 5p, 6d, and 6s shells in the valence space.[15-17] Small-core RECPs have 28 electrons for the lanthanides.[18, 19] The corresponding valence basis sets associated with small-core pseudopotentials are

(14s13p10d8f) contracted to [10s8p5d4f] for lanthanides. The basis set associated with lanthanide large-core pseudopotentials is (7s6p5d) contracted to [5s4p3d].

2.1.3 Modelling of solvent extractions

Qualitative predictions for selectivity trends throughout the lanthanide series are now hindered by the lack of a clear theoretical account of the solvation effects for trivalent metal ions. Ivanov and Bryantsev developed a new polarizable dielectric continuum model to explain the aqueous-phase selectivity for the Ln(III) series in 2016,[20] explicitly taking into account the initial coordination shell surrounding a trivalent metal ion and implicitly accounting for the impact of the outer hydration shells. They employed RECP DFT in a number of flavors starting from a model process where two Ln(III) ions swap ligands (Eq. 2.11) where m is the overall charge of the respective complexes. The biggest flaw in such models is that they only explicitly account for the initial hydration shell surrounding a multivalent ion, which leaves out the solvation-free energy of the ion. They postulate that a dense arrangement of poorly coordinated counterions can take the role of a broad sheath of water molecules that were previously helping to disseminate high cationic charge into the medium. An improved equation Eq. 2.12 was used to define the aqueous selectivity for La(III) over the other Ln(III) ions. The requirement that the total charge of each complex, |m|, is less than one determines the number of nitrate anions in equation 2.12. Their findings support the notion that the aqueous-phase selectivity produced by the model suggested in Eq. 2.12 closely tracks experimental data.[20, 21]

$$La(H_2O)_9^{3+}{}_{(aq)} + Ln(L)(H_2O)_x^{m}{}_{(aq)} \rightleftharpoons La(L)(H_2O)_x^{m}{}_{(aq)} + Ln(H_2O)_9^{3+}{}_{(aq)} \qquad \text{(Eq. 2.11)}$$

$La(NO_3)_3(H_2O)_{3(aq)} + Ln(L)(NO_3)_y(H_2O)_x{}^{m}{}_{(aq)} \rightleftharpoons La(L)(NO_3)_y(H_2O)_x{}^{m}{}_{(aq)} + Ln(NO_3)_3(H_2O)_{3(aq)}$ (Eq. 2.12)

2.2 Machine learning

With the rise of artificial intelligence technology, in the field of chemistry, traditional methods based on experiments and physical models are gradually merging with data-based machine learning paradigms. More and more representations of data for computer processing are being developed and increasingly adapted to generative-dominant statistical models.

Machine learning improves existing methods of simulating chemical environments. We have already mentioned that computational chemistry allows us to partially bypass laboratory experiments. However, the computation of computational chemistry that simulates quantum mechanical processes is poor in terms of computational cost and accuracy of chemical simulations. A central problem in computational chemistry is to solve the electron Schrödinger equation for complex molecules—that is, to compute properties of interest given the positions of the ensemble of nuclei and the total number of electrons. Exact solutions are possible only for single-electron systems, while for other systems we have to rely on "good enough" approximations. Furthermore, many popular methods for approximating the Schrödinger equation scale exponentially, making brute force solutions difficult to solve. Over the last century, many methods have been developed to speed up computation without sacrificing too much accuracy. However, even some "cheaper" methods can lead to computational bottlenecks.

One way AI can speed up these calculations is to combine them with machine learning. Another approach bypasses the modeling of physical processes entirely by directly mapping molecular characterization to desired properties. Both methods allow chemists to more efficiently examine chemical databases for various properties, such as atomic charge, ionization energy, and more. The basic workflow for building machine learning models on the basis of supervised learning can be broken down as follows.[22]

1) Data pre-processing

2) Feature selection

3) Model selection

4) Fitting the model to the training data

5) Using the model to predict values of the response variables for the test data.

Various machine learning methods are used to build mathematical models for compound prediction. These methods include linear regression models, Bayesian neural networks, random forests (RF), partial least squares, and support vector machines (SVMs). In the face of some complex molecular property prediction, traditional machine learning algorithms are still insufficient, and deep learning technology has added a new dawn to cheminformatics research with its outstanding capabilities. What are the advantages of deep learning algorithms over traditional machine learning algorithms? First of all, for traditional machine learning algorithms, there are three characteristics: features are artificially set, in the process of feature extraction or construction, different transformations and approximations are applied to the input features; simple template matching, only

considering how The use of the input feature set produces relatively satisfactory results, and does not pay attention to the characterization of the problem; the expressiveness of the model gradually becomes stronger as the number of parameters increases. In deep learning, the input features are represented at different levels through a multi-layer artificial neural network (ANN), realizing the concept of "hierarchical representation", and then optimizing the performance of the model through a variety of training techniques. The surface layer of each of these layers may also be applied to other new problems. The expressiveness of the model grows exponentially with the number of layers.[23] With such characteristics, deep learning can directly perform feature training on the basis of input raw data, and finally obtain a competitive prediction model. In short, deep learning algorithms are not only capable of building predictive models, but also have the potential to learn features automatically.

The concept of deep learning stems from artificial neural networks. Artificial neural network is based on the basic principle of neural network in biology. After understanding and abstracting the structure of the human brain and the response mechanism of external stimuli, it uses the knowledge of network topology as the theoretical basis to simulate the processing mechanism of the human brain's nervous system for complex information. A mathematical model. ANN is an operational model composed of a large number of nodes (or neurons) connected to each other. Each node represents a specific output function, called an activation function, and the line between each two nodes represents a weight. In order to minimize the prediction error, the weights in the ANN model will be adjusted with continuous training. The network consists of 3 parts: input layer, hidden layer and output

layer. The 3-dimensional input information is mapped to the 1-dimensional output value through the nonlinear combination of neurons in each layer. The performance of the ANN model depends on the multi-layer nonlinear transformation ability of the hidden layer.[23] As the number of layers increases and the layer width (the number of neurons in each layer) increases, more complex and abstract features will be constructed, and correspondingly, the model can also learn more complex and abstract representations. In this thesis, we will use ANN as the core algorithm to apply a machine-learning approach for predictions of solvent extraction properties of both trivalent lanthanides and actinides by different ligands.

Reference

1. R. V. Leeuwen and E. J. Baerends, *Int. J. Quantum Chem*, 1994, **52**, 711-730.
2. M. Gugelchuk and Y. Cui, *Journal of Molecular Structure: THEOCHEM*, 1996, **365**, 111-117.
3. J. P. Perdew and Y. Wang, *Physical review B*, 1992, **45**, 13244.
4. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
5. C. Lee, W. Yang and R. G. Parr, *Physical review B*, 1988, **37**, 785.
6. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
7. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372-1377.
8. A. D. Becke, *Physical review A*, 1988, **38**, 3098.
9. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200-1211.
10. A. D. Becke, *J. Chem. Phys.*, 1992, **96**, 2155-2160.

11. J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982-9985.

12. M. Hülsen, A. Weigand and M. Dolg, *Theor. Chem. Acc.*, 2009, **122**, 23-29.

13. A. Moritz, X. Cao and M. Dolg, *Theor. Chem. Acc.*, 2007, **117**, 473-481.

14. H. Stoll, B. Metz and M. Dolg, *J. Comput. Chem.*, 2002, **23**, 767-778.

15. M. Dolg, H. Stoll, A. Savin and H. Preuss, *Theor. Chim. Acta*, 1989, **75**, 173-194.

16. D. Feller, *J. Comput. Chem.*, 1996, **17**, 1571-1586.

17. K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045-1052.

18. X. Cao and M. Dolg, *J. Chem. Phys.*, 2001, **115**, 7348-7355.

19. X. Cao and M. Dolg, *Journal of Molecular Structure: THEOCHEM*, 2002, **581**, 139-147.

20. A. S. Ivanov and V. S. Bryantsev, *Eur. J. Inorg. Chem.*, 2016, **2016**, 3474-3479.

21. M. R. Healy, A. S. Ivanov, Y. Karslyan, V. S. Bryantsev, B. A. Moyer and S. Jansone-Popova, *Chem. Eur. J.*, 2019, **25**, 6326-6331.

22. D. Packwood, L. T. H. Nguyen, P. Cesana, G. Zhang, A. Staykov, Y. Fukumoto and D. H. Nguyen, *Machine learning with applications*, 2022, **8**, 100265.

23. M. Raghu, B. Poole, J. Kleinberg, S. Ganguli and J. Sohl-Dickstein, 2017.

Chapter 3. N-oxide Ligands for Selective Separations of Lanthanides

3.1 Abstract

Preorganized ligands such as bis-lactam-1,10-phenanthroline (BLPhen) show unique selectivity trend across lanthanide series, indicating the synergistic effects of both N and O donors in complexing with lanthanides. We hypothesized that by replacing amide functional group with N-oxide functionality would open the door to new ligand architectures with improved selectivities. To test this idea, we computationally examined mixed N,O-donor ligands containing pyridinic N and N-oxide groups and evaluated their relative aqueous La(III)/Ln(III) selectivity by computing free energy changes for the exchange reaction between the designed ligands and a reference ligand. Three novel ligands show promise as excellent extractant agents in selectively separating trivalent lanthanides. The extent of conjugation, the complex geometry, and the electron accumulations on the two O-donors of the N-oxide groups are found to be important factors in dictating the selectivity trends.

3.2 Introduction

Rare earth elements (REEs), including fifteen lanthanides (Ln), Sc, and Y, find broad applications in enabling many important technologies and industries.[1-6] However, they occur naturally together due to their similar properties and must be separated. Solvent extraction is the primary means to separate different lanthanides on an industrial scale.[7] Due to lanthanide contraction, most ligands prefer to bind heavier lanthanides than the lighter ones, because of the decreasing ion size traversing the series. Commercial

extractants employ oxygen donors such as tributyl-phosphate (TBP),[8] diglycolamide (DGA),[9-11] and bis(2-ethylhexyl) phosphoric acid (D2EHPA).[12] Ligands with N donors such as alkylated bis-triazinyl pyridines (BTP),[13] 6,6′-bis-triazinyl-2,2′-bipyridine (BTBP),[14] and 2,9-bis-triazinyl-1,10-phenanthroline (BTPhen)[15, 16] are also used.

Recently, ligands combining hard O-donor and soft N-donor atoms have been recognized as efficient extractants in that amide oxygen donors can provide stronger metal-ion binding. One such example is 2,9-bis-lactam-1,10-phenanthroline (BLPhen)[17] that shows unparalleled selectivity for light trivalent lanthanides.[18] The rigidity of the BLPhen backbone has been shown via quantum chemical calculations to be an important factor influencing the selectivity of lanthanide ions.[18] First principles molecular dynamics simulations suggested a tight binding pocket between BLPhen and Ln(III).[19] Therefore, combining O-donor and N-donor atoms could be a general strategy to design new ligands for separations of lanthanides. An innovation would be to introduce different types of N,O donors on the BLPhen framework.

N-oxide donors are common in chelate complexes of transition metals.[20] More interestingly, some pyridine-N-oxide-derived ligands have exhibited abilities to selectively coordinate to Ln(III)[21-23] or actinides[24] whose separations can provide enlightenment to Ln(III) behaviors. Moreover, computational approaches haven been increasingly used to help design new ligands, including the data-driven machine learning approach.[25] Our goal here is to computationally examine new mixed N,O-donor ligands based on phenanthroline and N-oxide functionalities for complexation across the Ln(III) series, in order to gain insights into their potential for separating Ln(III)s and correlate with their molecular

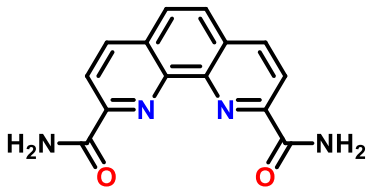
structure and electronic structure in terms of key descriptors. Below we explain our computational approach.

3.3 Computational Method

Following a previously established computational strategy,[18, 26] the relative aqueous selectivity for La(III) over the other Ln(III) ions was evaluated by computing the Gibbs free energy change, $\Delta\Delta G_{aq}$(La/Ln), of the following ligand–exchange reaction:

$[La(L_{ref})](NO_3)_{3(aq)}+[Ln(L_{target})](NO_3)_{3(aq)} \rightleftharpoons [La(L_{target})](NO_3)_{3(aq)}+[Ln(L_{ref})](NO_3)_{3(aq)}$ (1),

where L_{ref} and L_{target} represent the refence ligand and the target ligand (the designed one), respectively. A negative $\Delta\Delta G_{aq}$(La/Ln) means that La(III) prefers complexation with the target ligand and Ln(III) prefers the reference ligand. For the same La/Ln pair, the target ligand with more negative $\Delta\Delta G_{aq}$(La/Ln) than another target ligand would have higher La/Ln selectivity. Following the previous study,[18] we use 2,9-bis-amide-1,10-phenanthroline (BAPhen; Scheme 3-1) as the reference ligand.



Scheme 3-1. The 2,9-bis-amide-1,10-phenanthroline (BAPhen) ligand, used as the reference ligand to determine free-energy change for the ligand–exchange reaction between La(III) and Ln(III).

The energies of the four 1:1 ligand-metal complexes in Reaction (1) were computed at the B3LYP level of density functional theory (DFT) using the Gaussian 16 (revision

C.01) program package.[27] 6-31+G(d) basis sets were used for the main-group elements and hydrogen. The corresponding large-core (LC) relativistic effective core potentials (RECP) were used for all lanthanides elements.[28] Frequency calculations were performed to ensure real vibrational modes for the minimum ground-state structures and to provide zero-point energies (ZPEs). ZPEs and entropy contributions (T=298.15 K) calculated at the B3LYP/LC/6-31+G(d) level were added to the total energy to obtain the Gibbs free energies. The geometry optimizations started from the corresponding experimental crystal structures by appropriately modifying donor atoms, substituents, and backbone of the BLPhen ligands. All calculations were performed in aqueous environment by employing the IEF-PCM (integral equation formalism of the polarizable continuum model) implicit solvation model to obtain solvation free energies in the aqueous solution.[29] To correct the errors on the free energies of low-frequency vibrational modes from the harmonic oscillator model, frequencies lower than 60 cm^{-1} were set to 60 cm^{-1} by following the quasiharmonic approximation.[30]

3.4 Results and discussion

3.4.1 Changing amides in BLPhen to N-oxides

Our design starts with modification of the O donors on the BLPhen ligand (**1a** in Figure 3-1) by replacing the amide moieties with N-oxides (**1b**-**1d**). Figure 3-1b summarizes the computed ΔΔG trends across the Ln(III)s for the four ligands. One can see that the La(III)/Ln(III) selectivity becomes worse upon replacing one or both amides in **1a** with N-oxide functionality. Figure 3-1b also shows an interesting peak at Ce, meaning that

the target ligands (**1a**-**1d**) prefer to bind Ce(III) over La(III). This non-linear trend at Ce has also been observed in several experimental reports,[18, 31] suggesting a different chemical nature of Ce(III) from nearby Ln(III)s.

To understand the overall performance from the N-oxides (ligands **1b**-**1d**), we determined the partial charges on the O donors by the natural bond orbital (NBO) populations and found the charges of -0.68 e for **1a**, -0.64 e for **1b**, and -0.63 e for **1c** and **1d**. In other words, changing amide moieties to N-oxides does not make the O donors more negative, as initially hypothesized. This is likely due to the enhanced π conjugation from the pyridinium group in **1b**-**1d**. Hence our next strategy is to tune the conjugation to see if the performance of N-oxide-based ligands can be further optimized.

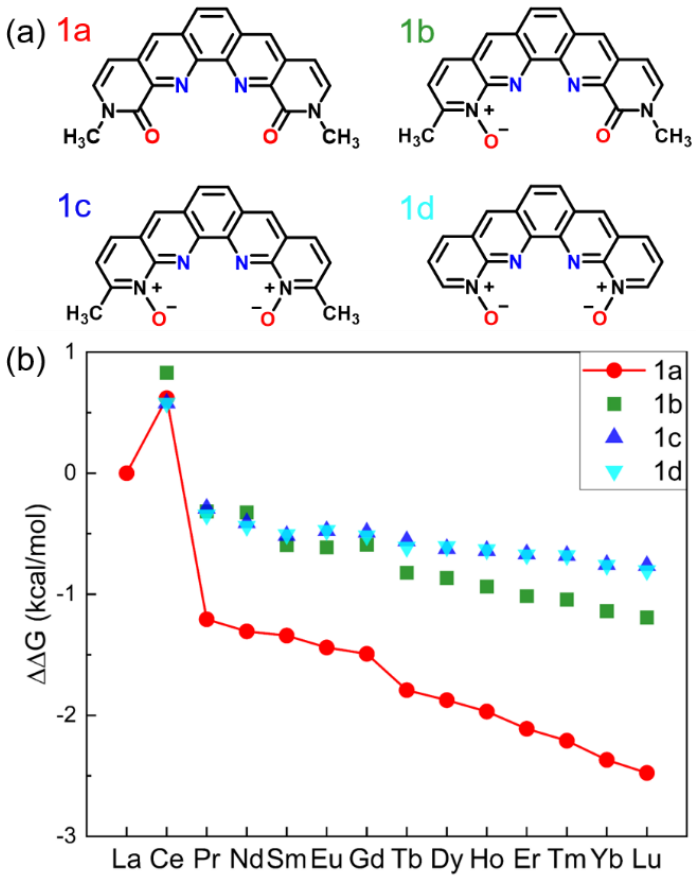


Figure 3-1. (a) Chemical structures of BLPhen (**1a**) and BLPhen-derived N-oxide ligands (**1b** - **1d**); (b) DFT-calculated relative aqueous-phase selectivity, $\Delta\Delta G_{aq}$(La/Ln), for the ligands **1a** - **1d**, with respect to the refence ligand (BAPhen).

3.4.2 Tuning the conjugation

The ligands **1a** – **1d** all have the conjugation throughout the whole molecule. Our idea was to disrupt the conjugation at the middle ring (**2a**-**2d** in Figure 3-2a). Since **1b**, **1c**, and **1d** show similar performances (Figure 3-1b), we selected **1d** as a starting ligand. Four new ligands were created based on modifications of **1d**: changing the middle top C=C double bond to a single bond (**2a**); reducing the middle 6-membered ring to a 5-membered

ring (**2b**); enlarging the middle ring to be a 7-membered one (**2d**) or eliminating the top C=C bond (**2c**). The calculated $\Delta\Delta G_{aq}$(La/Ln) values for **2a**-**2d** (Figure 3-2b) are compared to those of **1a**. Interestingly, ligand **2b** has much better performance than **1a**. On the other hand, the performances of **2a**, **2c**, and **2d** are worse than that of **1a**. Therefore, the reduced extent of conjugation in **2b**, combined with other factors, makes it a more selective ligand. To reveal those factors, we first compare the optimized geometries after complexation.

Similar to the structure of [La(BLPhen)]$(NO_3)_3$,[17] our optimized [La(**2b**)]$(NO_3)_3$ complex (Figure 3-3) also has tenfold coordination: four donors from **2b** ligand and six from three bidentate nitrates. Both [La(**2b**)]$(NO_3)_3$ and [Ln(**1a**)]$(NO_3)_3$ complexes are planar, as evidenced by the close-to-zero O1-N1-N2-O2 dihedral angle (Table 3-1 and Figure 3-2). On the other hand, ligands **2a**, **2c**, and **2d** are significantly non-planar, leading to less planar complexes. Therefore, the geometric comparison suggests that the planar geometry combined with the reduced conjugation helps improve the selectivity for **2b**, while the non-planarity of ligands **2a**, **2c**, and **2d** indicates that the whole conjugation is completely broken down into two smaller parts which might be detrimental to the complexation.

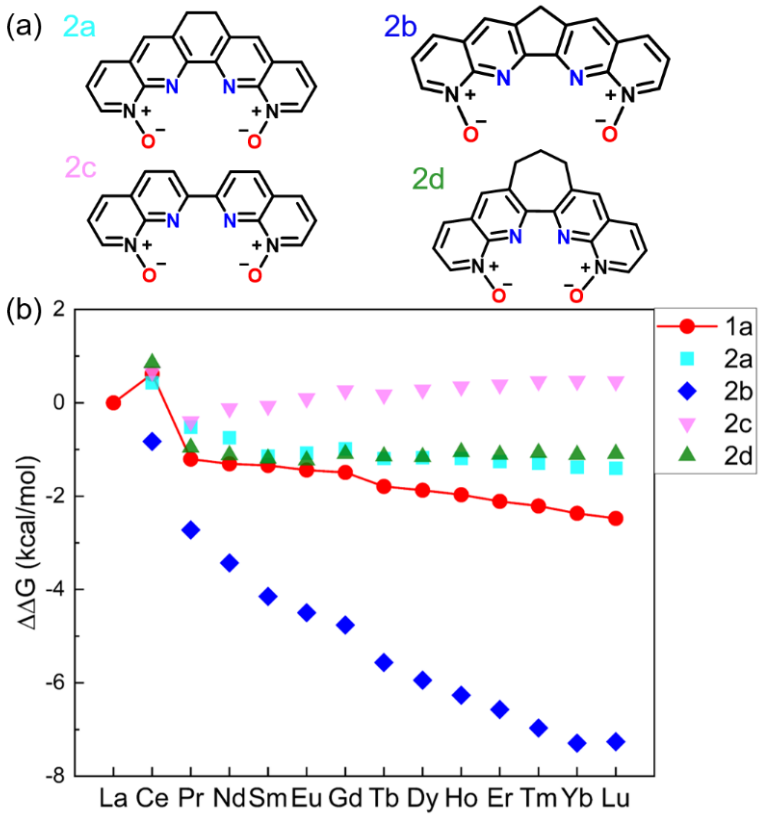


Figure 3-2. (a) Chemical structures of 1,10-phenanthroline derived N-oxides with reduced conjugation. (b) DFT-calculated relative aqueous phase selectivity ($\Delta\Delta G_{aq}$(La/Ln)).

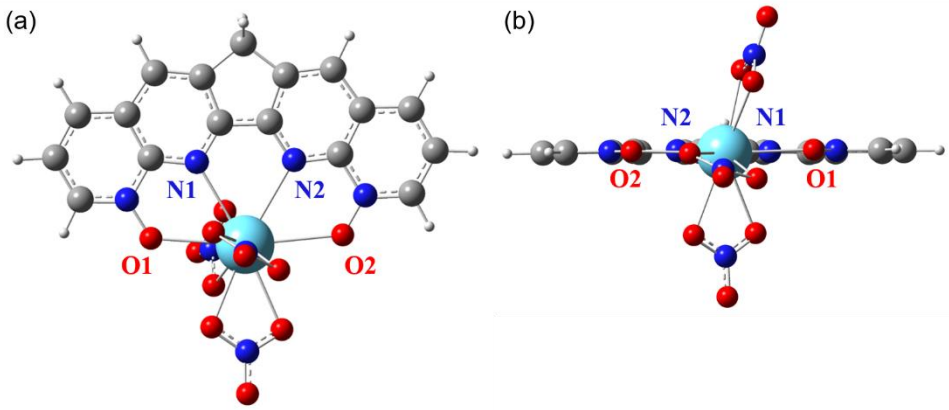


Figure 3-3. DFT-optimized [La(**2b**)]$(NO_3)_3$ complex with four donors on the **2b** ligand labelled: (a) top view; (b) side view.

Table 3-1. O1-N1-N2-O2 dihedral angles in ligands **1a**, **2a**-**2d** and corresponding La complexes (see **Figure 3-3** for the atom labels).

	1a	**2a**	**2b**	**2c**	**2d**	**3a**	**3b**	**3c**
Ligand	0.4˚	9.9˚	0.5˚	18.9˚	26.8˚	0.0˚	0.2˚	0.6˚
La(ligand)$(NO_3)_3$	0.8˚	8.3˚	1.3˚	5.0˚	14.3˚	1.4˚	2.0˚	28.0˚

Unlike other ligands considered in this study, **2b** possesses a fluorene-like moiety, where the methylene group is known to be involved in hyperconjugation.[32] Indeed, our natural bond orbital (NBO) analysis for **2b** revealed relatively strong interactions of the electrons in C-H σ-bonds with adjacent C=C π^* orbitals (Figure 3-4) with estimated second-order stabilization energies $E^{(2)}$ of ~3.4 kcal/mol. Such hyperconjugation leads to a better geometric fit for Ln(III) ions, as reflected in the shorter Ln-N bonds in the [Ln(**2b**)]$(NO_3)_3$ complexes than in the complexes of other ligands (see Table 3-2 for some comparisons). On the other hand, NBO charge analyses on ligands **2a**-**2d** do not show significant differences of charges on pyridinic nitrogen atoms or NO oxygen atoms (Table 3-3). In other words, we think that the impact of the hyperconjugation from the middle constrained 5-membered ring is more geometric than electronic.

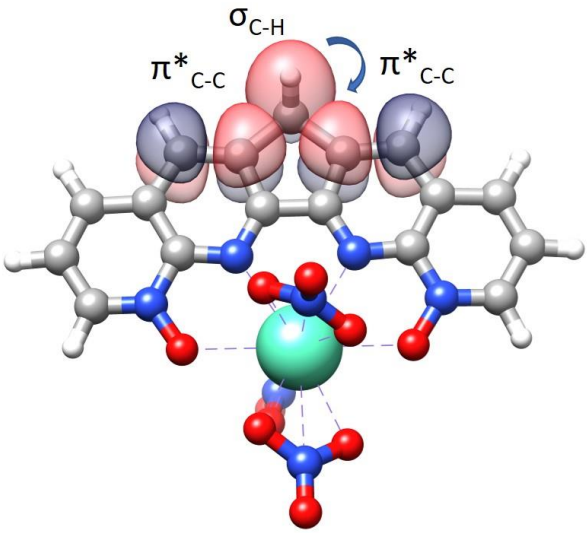


Figure 3-4. Natural bond orbital (NBO) analysis of the [Eu(**2b**)]$(NO_3)_3$ complex: $\sigma_{C\text{-}H} \rightarrow \pi^*_{C\text{-}C}$ interactions characterized by the overlap of the C-H σ-bond orbitals with adjacent C=C π^* orbitals.

Table 3-2. Average distance between Ln(III) and N-donor in selected complexes for comparison

Complexes	[Eu(**1a**)]$(NO_3)_3$	[Eu(**2a**)]$(NO_3)_3$	[Eu(**2b**)]$(NO_3)_3$	[Eu(**2c**)]$(NO_3)_3$
Eu-N distance (Å)	2.66	2.67	2.64	2.67
Complexes	[Lu(**1a**)]$(NO_3)_3$	[Lu(**2a**)]$(NO_3)_3$	[Lu(**2b**)]$(NO_3)_3$	[Lu(**2c**)]$(NO_3)_3$
Lu-N distance (Å)	2.57	2.57	2.55	2.57

Table 3-3. NBO charges of N and O atoms in ligands **2a**-**2d**

	2a	**2b**	**2c**	**2d**
N (pyridine)	-0.43	-0.43	-0.44	-0.44
O (N-Oxide)	-0.54	-0.55	-0.54	-0.55

We have further performed NBO charge analysis for [Ln(**2a**)]$(NO_3)_3$ and [Ln(**2b**)]$(NO_3)_3$ complexes (Figure 3-5) and found that the non-linear trends of Ln partial charges across the Ln series are very similar between **2a** and **2b**. The charges on Ln are lightly smaller in [Ln(**2a**)]$(NO_3)_3$ complexes than in their [Ln(**2b**)]$(NO_3)_3$ counterparts; in other words, there are more ligand charge transfers to Ln from **2a** than from **2b**. On the contrary, $\Delta\Delta G_{aq}$(La/Ln) values in Figure 3-2 show that **2b** is more selective than **2a**; the selective trend across the Ln series is rather monotonic. So, we conclude that the charge transfer is unlikely to be the key factor. This is consistent with our conclusion that the geometric factor is more important.

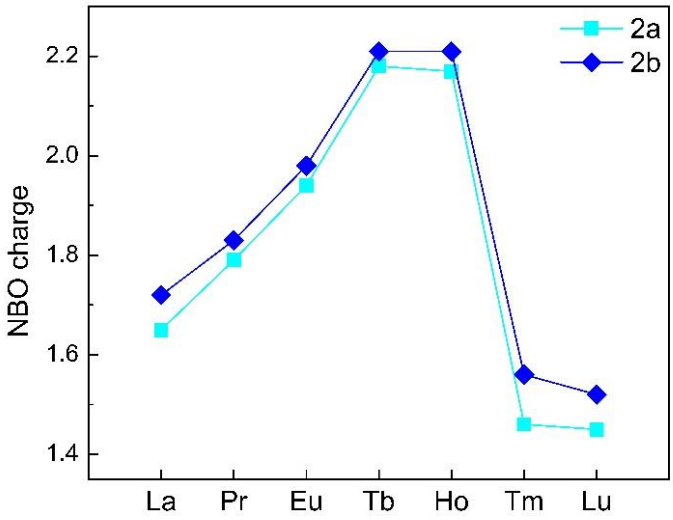


Figure 3-5. NBO charge on Ln in [Ln(**2a**)]$(NO_3)_3$ and [Ln(**2b**)]$(NO_3)_3$ complexes.

3.4.3 Further tuning of the conjugation size

Since **2b** notably enhances the La(III)/Ln(III) selectivity across the whole lanthanide series, we designed additional ligands based on **2b** by modifying the conjugation size via the following approaches as shown in Figure 3-6a: changing the two pyridiniums to five-

membered rings (**3a**); changing only one outside ring to a five-membered one (**3b**); inserting one phenyl ring on each side (**3c**). In comparison with **2b**, the calculated ΔΔG (Figure 3-6b) indicates better La(III)/Ln(III) selectivity for **3a** and much worse La(III)/Ln(III) selectivity for **3c**. Considering the contrast between **3b** and **3c** in $\Delta\Delta G_{aq}$(La/Ln) values, we think that the O-O distance is a key factor here: it is too short in **3c** (Figure 3-6a) which is very detrimental to the La(III)/Ln(III) selectivity. Besides geometric factors such as planarity and O-O distances, orbital interactions may also be important, which we analyze next.

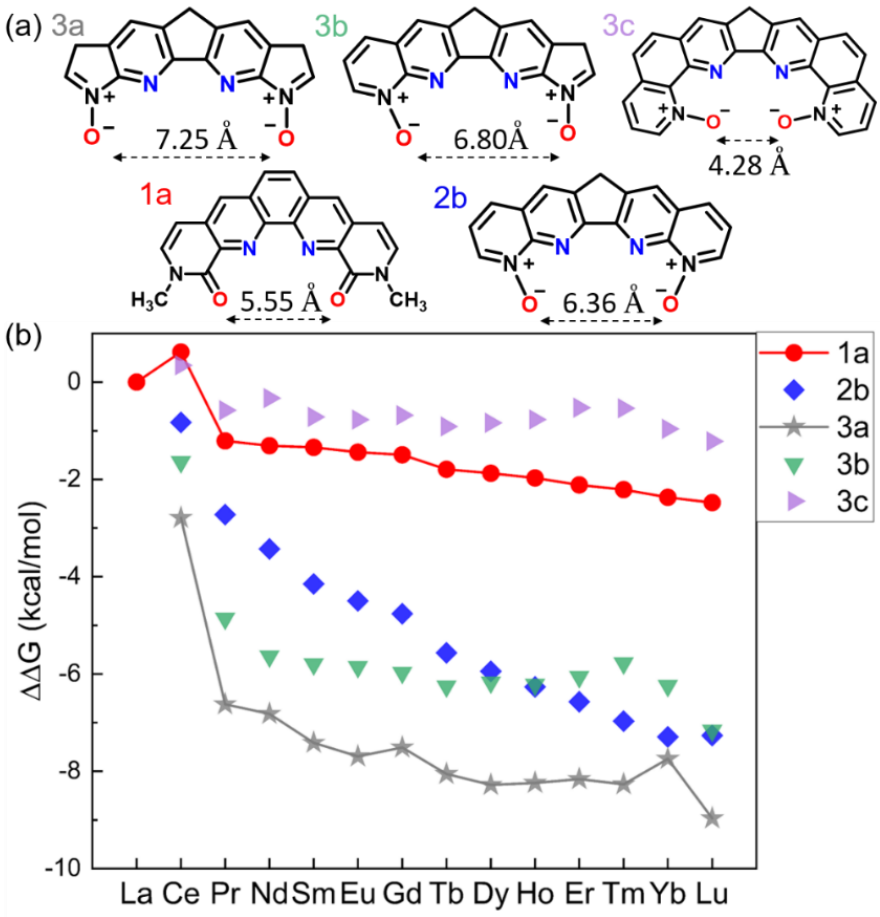


Figure 3-6. (a) Chemical structures of N-oxide ligands **3a**-**3c** derived from **2b**; ligand **1a** is also shown for comparison in terms of O-O distance. (b) DFT-calculated relative aqueous phase selectivity, $\Delta\Delta G_{aq}$(La/Ln), for ligands **3a**-**3c** in comparison with ligands **1a** and **2b**.

3.4.4 Orbital analysis of ligands with high La(III)/Ln(III) selectivity

Computed $\Delta\Delta G_{aq}$(La/Ln) for our designed mixed N, O-donor shows that the N-oxide-based ligands **2b**, **3a**, and **3b** have the potential to be more La(III)/Ln(III) selective than the BLPhen ligand **1a**. The complexation of these ligands with Ln(III) involves mainly the donation of the lone pairs of electrons from the ligands to the Ln(III) ion. Therefore, it

would be interesting to compare the HOMOs of these high-performing ligands. From Figure 3-7 one can see more electron accumulation around O donors in **2b**, **3a**, and **3b** than in **1a**, which could be a reason for their improved selectivity profile. The other interesting feature is that the conjugation is similar between **1a** and **2b**, but more fragmented in **3a** and **3b**. This causes very different electron density distribution at N donors from **1a** and **2b** to **3a** and **3b**. We think that this may be the reason for the more non-linear selectivity trends across the Ln series for **3a** and **3b** (Figure 3-6b).

Of course, there could be other important factors such as conjugation sizes and O-O distances discussed above, as the orbitals for all the ligands considered, when examined together, displays a more complicated picture. A machine-learning model that can rank all descriptors in terms of their importance in dictating the selectivity would be highly desirable. Future work is warranted.

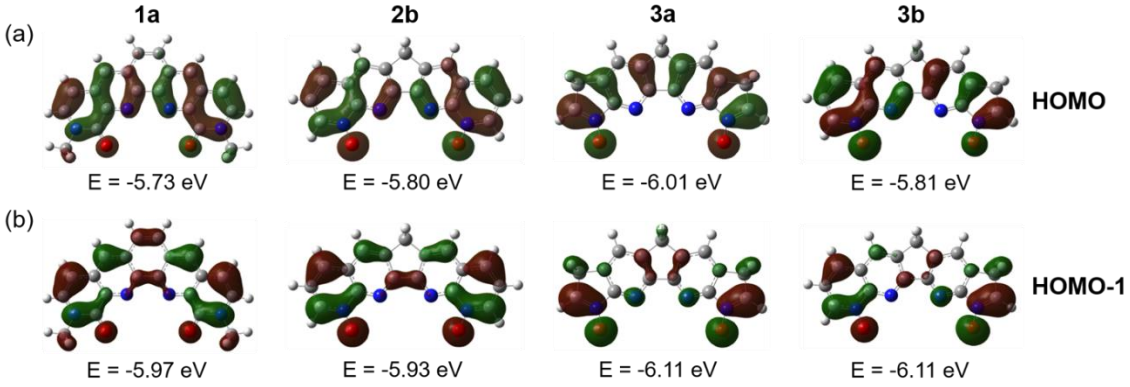
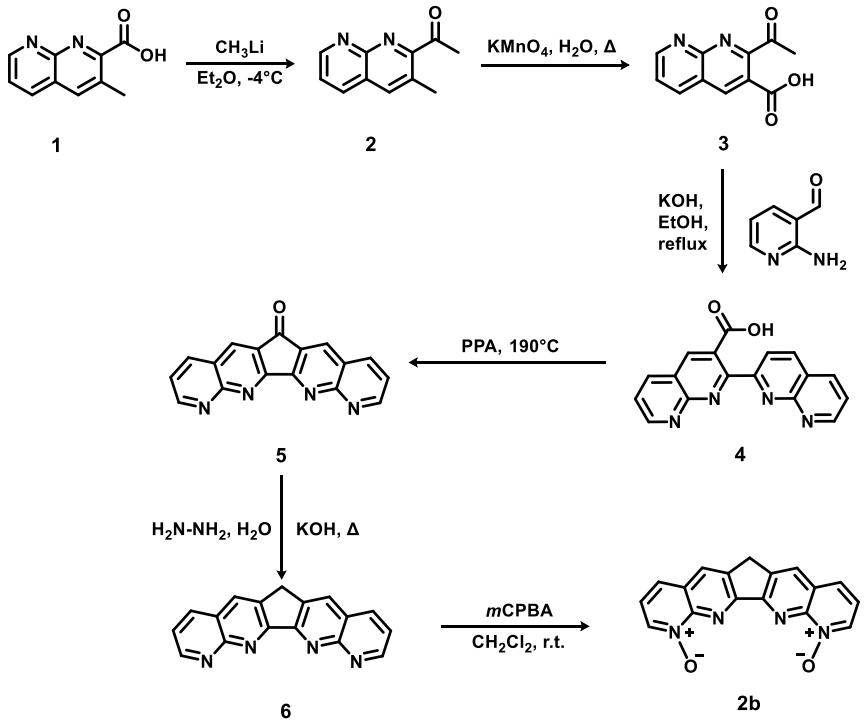


Figure 3-7. (a) HOMO and (b) HOMO-1 of ligands **1a**, **2b**, **3a** and **3b** and their energies.

3.4.5 Implications for ligand synthesis

Ligand **1a** (BLPhen) and its derivatives have been reported recently for Ln(III) separations.[17,18] Our results above suggest that the N-oxide-based **2b**, **3a**, and **3b** ligands can potentially surpass the BLPhen-based ligands. Being the simplest among the three, **2b** could be the first target and Scheme 3-2 shows the proposed route. It starts with a commercially available compound, 3-methyl-1,8-naphthyridine-2-carboxylic acid (**1**) which reacts with methyllithium to yield **2**,[33] followed by oxidation to **3**.[34] Equivalent **3** and 2-aminonicotinaldehyde react in potassium hydroxide and ethanol to yield **4**,[35] which dehydrates in polyphosphoric acid (PPA) to yield **5**,[36] followed by hydrazine reduction to **6**.[36] Finally, **6** can be selectively oxidized by reacting with 3.5 equivalent meta-chloroperoxybenzoic acid (*m*CPBA) in dichloromethane,[37] to yield **2b**. Similar routes can also be adopted to synthesize **3a** and **3b**.



Scheme 3-2. Proposed route to synthesize **2b**.

3.5 Summary and conclusions

We have computationally evaluated a new family of mixed N,O-donor ligands, derived from BLPhen and incorporating N-oxide functionalities, for their relative aqueous La(III)/Ln(III) selectivity. Three novel ligands were identified to be promising and experimentally viable targets in selective separations of trivalent lanthanides. We found that the conjugation, O-O distance, planarity of the formed complex, and the electron density on the two O atoms are important control knobs that affect ligand's selectivity for lanthanides. Our computational insights will guide the follow-up efforts towards synthesizing the top candidates.

References

1. T. Cheisson and E. J. Schelter, *Science*, 2019, **363**, 489-493.

2. T. I. Kostelnik and C. Orvig, *Chem. Rev.*, 2019, **119**, 902-956.

3. Z. Zeng, Y. Xu, Z. Zhang, Z. Gao, M. Luo, Z. Yin, C. Zhang, J. Xu, B. Huang and F. Luo, *Chem. Soc. Rev.*, 2020, **49**, 1109-1143.

4. F. Saraci, V. Quezada-Novoa, P. R. Donnarumma and A. J. Howarth, *Chem. Soc. Rev.*, 2020, **49**, 7949-7977.

5. B. Zheng, J. Fan, B. Chen, X. Qin, J. Wang, F. Wang, R. Deng and X. Liu, *Chem. Rev.*, 2022, **122**, 5519-5603.

6. F. Ortu, *Chem. Rev.*, 2022, **122**, 6040-6116.

7. F. Xie, T. A. Zhang, D. Dreisinger and F. Doyle, *Miner. Eng.*, 2014, **56**, 10-28.

8. L. K. Sinclair, J. W. Tester, J. F. H. Thompson and R. V. Fox, *Ind. Eng. Chem. Res.*, 2019, **58**, 9199-9211.

9. E. A. Mowafy and D. Mohamed, *Sep. Purif. Technol.*, 2014, **128**, 18-24.

10. R. J. Ellis, D. M. Brigham, L. Delmau, A. S. Ivanov, N. J. Williams, M. N. Vo, B. Reinhart, B. A. Moyer and V. S. Bryantsev, *Inorg. Chem.*, 2017, **56**, 1152-1160.

11. X. He, X. Wang, Y. Cui, Z. Su, G. Ye, C. Lu and L. Li, *J. Radioanal. Nucl. Chem.*, 2021, **329**, 1019-1026.

12. A. Kumari, R. Panda, J. Y. Lee, T. Thriveni, M. K. Jha and D. D. Pathak, *Sep. Purif. Technol.*, 2019, **227**, 115680.

13. A. Kovács, C. Apostolidis and O. Walter, *Inorganics*, 2019, **7**, 26.

14. M. R. Foreman, M. J. Hudson, M. G. Drew, C. Hill and C. Madic, *Dalton Trans.*, 2006, 1645-1653.

15. Y. Yang, J. Liu, L. Yang, K. Li, H. Zhang, S. Luo and L. Rao, *Dalton Trans.*, 2015, **44**, 8959-8970.

16. D. M. Whittaker, T. L. Griffiths, M. Helliwell, A. N. Swinburne, L. S. Natrajan, F. W. Lewis, L. M. Harwood, S. A. Parry and C. A. Sharrad, *Inorg. Chem.*, 2013, **52**, 3429-3444.

17. S. Jansone-Popova, A. S. Ivanov, V. S. Bryantsev, F. V. Sloop, R. Custelcean, I. Popovs, M. M. Dekarske and B. A. Moyer, *Inorg. Chem.*, 2017, **56**, 5911-5917.

18. M. R. Healy, A. S. Ivanov, Y. Karslyan, V. S. Bryantsev, B. A. Moyer and S. Jansone-Popova, *Eur. J. Chem.*, 2019, **25**, 6326-6331.

19. T. D. N. Reddy, A. S. Ivanov, D. M. Driscoll, S. Jansone-Popova and D.-e. Jiang, *ACS Omega*, 2022, **7**, 21317-21324.

20. J. Quagliano, J. Fujita, G. Franz, D. Phillips, J. Walmsley and S. Tyree, *J. Am. Chem. Soc.*, 1961, **83**, 3770-3773.

21. I. Binyamin, S. Pailloux, E. N. Duesler, B. M. Rapko and R. T. Paine, *Inorg. Chem.*, 2006, **45**, 5886-5892.

22. M. A. S. Goher and F. A. Mautner, *J. Mol. Struct.*, 2007, **846**, 153-156.

23. H.-J. Zhang, R.-H. Gou, L. Yan and R.-D. Yang, *Spectrochim. Acta A Mol.*, 2007, **66**, 289-294.

24. D. Das, M. Joshi, S. Kannan, M. Kumar, T. K. Ghanty, A. S. Pente, A. Sengupta and C. P. Kaushik, *Polyhedron*, 2021, **201**, 115166.

25. T. Liu, K. R. Johnson, S. Jansone-Popova and D.-e. Jiang, *JACS Au*, 2022, **2**, 1428-1434.

26. A. S. Ivanov and V. S. Bryantsev, *Eur. J. Inorg. Chem.*, 2016, **2016**, 3474-3479.

27. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian, Inc., Wallingford CT*, 2016.

28. M. Dolg, H. Stoll and H. Preuss, *J. Chem. Phys.*, 1989, **90**, 1730-1734.

29. J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999-3094.

30. R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2011, **115**, 14556-14562.

31. J. Dehaudt, N. J. Williams, I. A. Shkrob, H. Luo and S. Dai, *Dalton Trans.*, 2016, **45**, 11624-11627.

32. R. Baker, C. Eaborn and J. Sperry, *Journal of the Chemical Society (Resumed)*, 1962, 2382-2385.

33. G. M. Rubottom and C. Kim, *J. Org. Chem.*, 1983, **48**, 1550-1552.

34. G. K. Wagner, D. Kotschenreuther, W. Zimmermann and S. A. Laufer, *J. Org. Chem.*, 2003, **68**, 4527-4530.

35. B. Xiong, S. Zhang, H. Jiang and M. Zhang, *Org. Lett.*, 2016, **18**, 724-727.

36. M. E. Hoque, R. Bisht, A. Unnikrishnan, S. Dey, M. M. Mahamudul Hassan, S. Guria, R. N. Rai, R. B. Sunoj and B. Chattopadhyay, *Angew. Chem. Int. Ed.*, 2022, **61**, e202203539.

37. R. P. Thummel and Y. Jahng, *J. Org. Chem.*, 1985, **50**, 3635-3636.

Chapter 4. Advancing Rare-Earth Separation by Machine Learning

4.1 Abstract

Constituting the bulk of rare-earth elements, lanthanides need to be separated to fully realize their potential as critical materials in many important technologies. The discovery of new ligands for improving rare-earth separations by solvent extraction, the most practical rare-earth separation process, is still largely based on trial and error, a low throughput and inefficient approach. A predictive model that allows high-throughput screening of ligands is needed to identify suitable ligands to achieve enhanced separation performance. Here, we show that deep neural networks, trained on the available experimental data, can be used to predict accurate distribution coefficients for solvent extraction of lanthanide ions, thereby opening the door to high-throughput screening of ligands for rare-earth separations. One innovative approach that we employed is a combined representation of ligands with both molecular physicochemical descriptors and atomic extended-connectivity fingerprints, which greatly boosts the accuracy of the trained model. More importantly, we synthesized four new ligands and found that the predicted distribution coefficients from our trained machine-learning model match well with the measured values. Therefore, our machine-learning approach paves the way for accelerating the discovery of new ligands for rare-earth separations.

4.2 Introduction

Rare earth elements (REEs), including the 14 lanthanides, yttrium, and scandium, are recognized as critical materials vital to many technologies.[1-4] Due to their similar properties, REEs are difficult to separate from one another.[5] Solvent extraction is the most extensively used process to separate lanthanides on industrial scale. This process employs an organic ligand (extractant or complexing agent) in a nonpolar, water-immiscible solvent (org) to extract trivalent lanthanides, Ln(III), from an aqueous (aq) solution. The extraction performance is expressed as a distribution ratio for each Ln(III), $D = [M^{3+}]_{org}/[M^{3+}]_{aq}$. High D values indicate better extraction efficiency and imply the formation of stable Ln(III) complexes in the organic phase. Ligands that show great promise in REE separations include diglycolamides (DGA),[6-9] alkylated bis-triazinyl pyridines (BTP),[10] and 2,9-bis-lactam-1,10-phenanthroline (BLPhen),[11, 12] among others.[13-15] Extraction performance is also impacted by experimental conditions, including solvent, temperature, and volume of each phase. Organic solvents such as toluene,[6] n-dodecane,[16] 1-octanol,[17] and dichloroethane[18] are commonly used to carry out the liquid-liquid separations.

Innovation in ligand design and discovery is key to achieving more efficient separation of Ln(III)s. Knowledge-based design, followed by synthesis of new ligands, tends to be low throughput and often relies on trial-and-error to determine optimized extraction conditions. In addition, quantum chemical calculations of the ligand-metal binding are limited by the solvation model and lack solvation dynamics; usually the relative change in free energy in reference to a common ligand[19, 20] is predicted instead of directly

predicting *D* values for Ln(III) for a specific ligand. These calculations also have limited throughput due to high costs.

Data-driven machine learning (ML) approach allows high-throughput screening of much larger chemical space, and the model will continuously improve as more data are generated. This approach has been increasingly used in predicting important equilibrium properties such as solubility,[21, 22] binding affinity,[23] pKa,[24] adsorption capacities,[25, 26] and partition coefficients of molecules.[27, 28] Hence, there is an opportunity to accelerate the discovery of new ligands for Ln(III) separation using the data-driven ML approach.

Herein we have developed a predictive model that accurately predicts *D* values for a given ligand by training deep neural nets on experimental data of measured *D* values and by sufficiently representing ligands, Ln(III), and experimental conditions. The model is then tested on four new ligands synthesized and the predicted *D* values are in very good agreement with the experiment, highlighting its predictive power to enable further high-throughput screening.

4.3 Computational Method

Data collection. All 1,202 logD values of lanthanide extraction in our database were collected from the scientific literature where a single neutral ligand was the only extractant used to extract Ln(III) from the aqueous phase to the organic phase consisting of one or two different solvents. For each data point, the inputs include sequentially the

representation of the ligand, descriptors of the extraction conditions, descriptors of the lanthanide.

Representation of ligands. The first 2,048 inputs of each data point are Extended-Connectivity Fingerprints[29] (ECFP) of the ligand; the next 208 inputs are RDKit descriptors.[30] They are both generated from the simplified molecular-input line-entry system (SMILES) expression of the ligand by the DeepChem package.[31] Chirality is considered in ECFP and other parameters use default settings: radius of fingerprint = 2; length of generated bit vector = 2,048; bond order considered; feature descriptors not used. RDKit descriptors use default parameters: binary descriptors of fragments like 'fr_XXX" are returned; avg = True for the Ipc (information of polynomial coefficients) descriptor[32] to return the information content divided by the total population. The names of the 208 descriptors returned by the RDKit module are listed in the Excel file, including molecular weight, number of valence electrons, partial charges, electrotopological state indexes, etc.

Details of the deep learning model and the training process. The training of fully connected neural networks (FCNNs) is performed via the PyTorch package (version 1.9.1)[33] with L1 type loss function, SGD optimizer, and L2 regularization for weight decay. The weight initializations obey the default normal distributions. Mean-absolute-error (MAE), root-mean-square-error (RMSE), and coefficient of determination (R^2) as calculated via the scikit-learn module were used as metrics for evaluation during the training process.[34]

4.4 Results and discussion

4.4.1 Data and Machine-Learning Workflow

1,202 reported *D* values using 109 different ligands were collected from the literature and used to build the dataset. Each Ln(III) has more than 60 entries in the dataset (Figure 4-1a). The experimental *D* values span eight orders of magnitude: as shown in Figure 4-1b, log*D* ranges from -4 to +4. Many classes of ligands, including phosphine oxides, amides, and N-heterocyclic derivatives, were selected (Figure 4-1c).[35, 36] 117 data points out of 1202 for 14 Ln(III)s were randomly selected as the validation set.

The workflow of our ML approach is summarized in Figure 4-1d. The input data comprises three parts: Ln(III), ligand, and solvent extraction conditions. First, the ligand, represented by a string-based name (SMILES), is fed into RDKit[30] – a cheminformatics toolkit that automatically generates 208 descriptors for the ligand. The RDKit descriptors are then combined with the extended connectivity fingerprints (ECFPs)[29] for a more detailed representation of the ligand. Fourteen descriptors are used for each Ln element (Table 4-1); solvent extraction conditions such as temperature, concentration of the ligand, and physical properties of organic solvents are also part of the input (Table 4-2). In total, 2291 inputs are used for each output log*D* value.

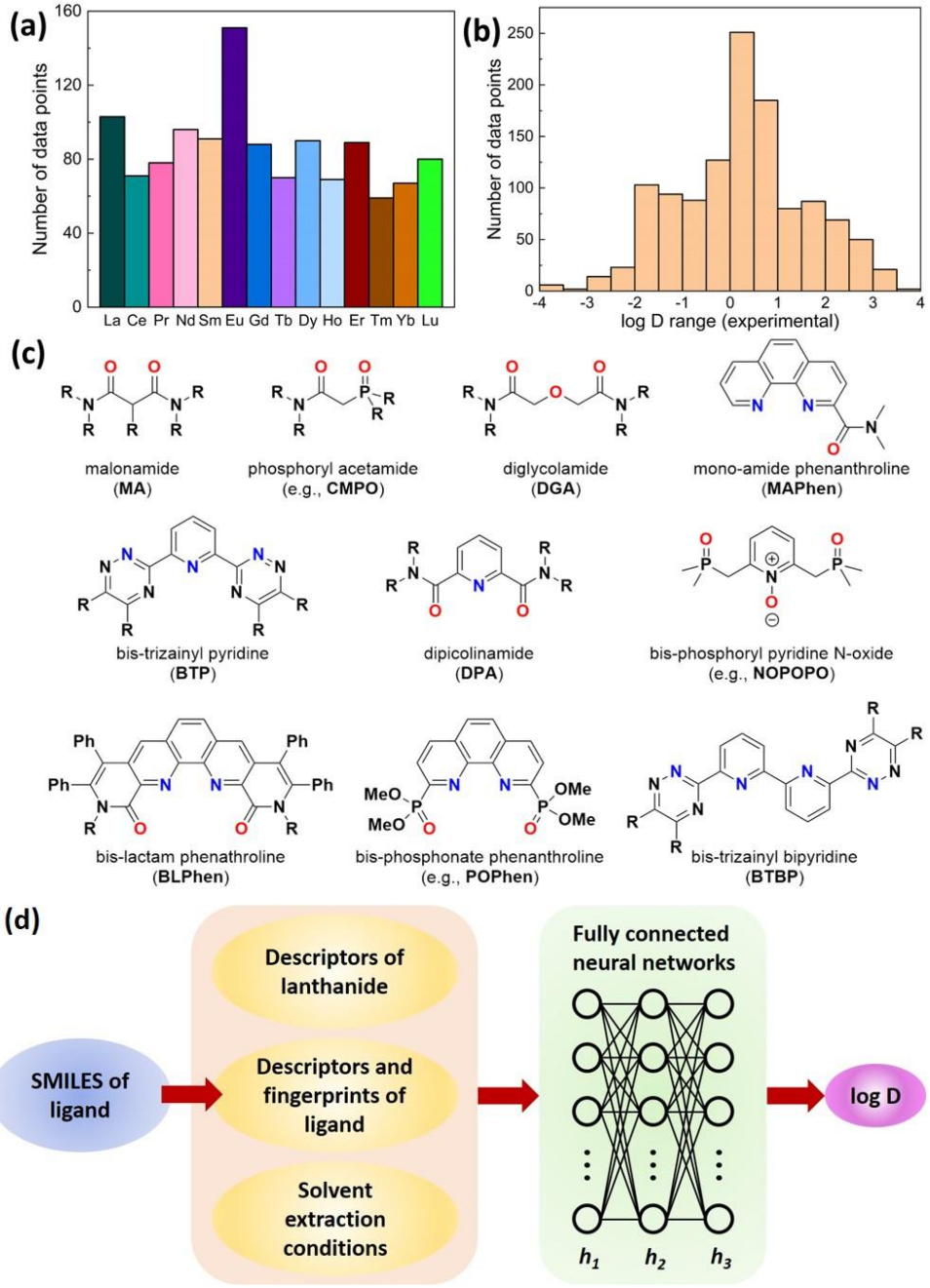


Figure 4-1. Dataset of log*D* values, typical ligands, and machine learning workflow. Distribution of the total dataset of log*D* values: (a) based on Ln(III), excluding radioactive Pm(III); (b) the value range. (c) Chemical structures of some representative ligands in the dataset. (d) The workflow of predicting log*D* of Ln(III) extracted by a ligand via fully connected neural networks with three hidden layers.

Table 4-1. Descriptors for fourteen lanthanides.

	Atomic Number	Outer shell electrons	Melting Point (K)	Boiling Point (K)	Density (g/cm^3)	First IE (kJ/mol)	Second IE (kJ/mol)
La	57	0	1193	3737	6.16	538.1	1067
Ce	58	1	1072	3716	6.77	534.4	1046.9
Pr	59	2	1204	3793	6.77	528.1	1017.9
Nd	60	3	1289	3347	7.01	533.1	1034.3
Sm	62	5	1345	2067	7.52	544.5	1068.1
Eu	63	6	1095	1802	5.24	547.1	1085.4
Gd	64	7	1586	3546	7.9	593.4	1166.5
Tb	65	8	1632	3503	8.23	565.8	1111.5
Dy	66	9	1685	2840	8.55	573.1	1125.9
Ho	67	10	1745	2973	8.8	580.9	1138.5
Er	68	12	1802	3141	9.07	589.3	1151.1
Tm	69	12	1818	2223	9.32	596.7	1162.7
Yb	70	13	1097	1469	6.9	603.4	1174.8
Lu	71	14	1936	3675	9.84	523.5	1341.1

	Third IE (kJ/mol)	Electron Affinity (kJ/mol)	Atomic Radius (Å)	Covalent Radius (Å)	Pauling Electronegativity	Ionic Radius (Å)	Standard Entropy (J·mol^{-1}·K^{-1})
La	1850.3	48	1.87	2.08	1.1	1.03	56.9
Ce	1948.8	55	2.42	1.84	1.12	1.02	72
Pr	2086.4	92.8	2.4	1.9	1.13	0.99	73.9
Nd	2132.3	184.9	2.39	1.88	1.14	0.983	71.1
Sm	2257.8	15.6	2.36	1.85	1.17	0.958	69.5
Eu	2404.4	11.2	2.35	1.83	1.2	0.947	77.8
Gd	1990.5	13.2	2.34	1.82	1.2	0.938	67.9
Tb	2113.9	112.4	2.33	1.81	1.1	0.923	73.3
Dy	2199.9	33.9	2.31	1.8	1.22	0.912	75.6
Ho	2203.7	32.6	2.3	1.79	1.23	0.901	75
Er	2194.1	30.1	2.29	1.77	1.24	0.89	73.2
Tm	2284.8	99.3	2.27	1.77	1.25	0.88	74
Yb	2416.9	-1.9	2.26	1.78	1.1	0.868	59.8
Lu	2022.3	32.8	2.24	1.74	1	0.861	51

Table 4-2. Descriptors of different organic solvents in the dataset.

	Molar mass (g/mol)	Density (g/mL)	Boiling point (K)	Melting point (K)	Dipole moment (D)	Solubility in water (g/L)	log P
CH_3Cl	50.49	1.003	249.3	175.8	1.9	5.325	0.91
Dichloromethane	84.93	1.3266	312.8	176.5	1.6	17.5	1.86
toluene	92.141	0.87	384	178	0.36	0.52	2.68
chloroform	119.37	1.489	334.3	209.7	1.15	8.09	1.97
n-dodecane	170.34	0.7495	489	263.5	0.07	0.004	6.82
1-octanol	130.231	0.83	468	257	1.68	0.3	3
nitrobenzene	123.11	1.199	383	278.8	4.28	1.9	1.85
cyclohexanone	98.15	0.9478	428.8	226	2.9	80.6	0.81
meta-nitrobenzotrifluoride	191.11	1.427	475.8	270.6	7	0	2.62
phenyl trifluoromethyl sulfone	210.17	1.249	476	315	5	0.3	0
TBP	266.32	0.9727	562	193	4	0.4	4
benzene	78.11	0.8765	353.2	278.7	0	1.8	2.13
CCl4	153.8	1.587	350	250	0	0.8	2.83
xylene	106.16	0.864	412	226	0.64	0.106	3.12
kerosene	200	0.8	523	298	0	0	0
sulfonated kerosene	300	0.82	523	298	5	0	0
Isopar L	170	0.77	471	270	0	0	0

4.4.2 Training and model performance

Fully connected neural networks (FCNNs) where every neuron in one layer is connected to every neuron in the next layer were used as the core of our approach for deep learning.[37] The training of the FCNNs was performed with the PyTorch package.[33] In each epoch, 80% of the 1085 data points were randomly selected for training. As shown in Figure 4-2, the coefficient of determination, R^2, between the predicted log D and experimental log D values of the validation set by using the combination of ECFP and RDKit for the ligands reached a higher value (~0.80) than that using only ECFP (~0.45)

or RDKit (~0.65) after 5000 epochs of training (Figure 4-2a). Likewise, the root-mean-squared error (RMSE) of the validation set for the ECFP+RDKit representation decreased more rapidly and achieved a lower value after 5000 epochs. Hence, the ECFP+RDKit representation of the ligand was used for the subsequent training.

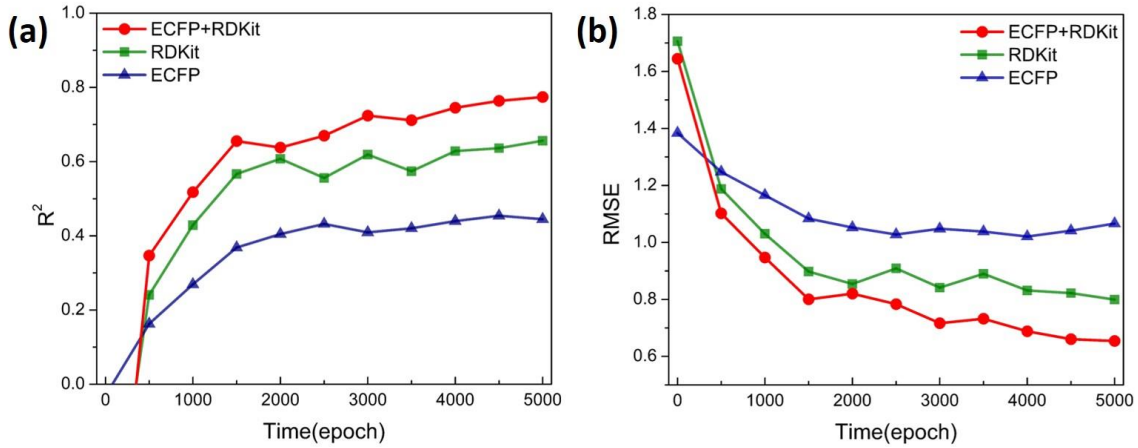


Figure 4-2. Comparing the three different approaches, RDKit, ECFP, or ECFP + RDKit, to represent ligands, based on the validation set performances of the trained FCNN for predicting log D against the experiment in the first 5000 epochs: (a) coefficient of determination, R^2, between the predicted *log D* and experimental *log D* values; (b) root-mean-square error, RMSE, between the predicted *log D* and experimental *log D* values (also measured against the standard deviation, σ, of experimental log D values of the training set, right axis). FCNN hyperparameters: 0.00001 learning rate, PReLU activation functions, 0.01 weight decay, three hidden layers, and the number of neurons on each layer = 512, 128, and 16.

Screenings of hyperparameters are listed in Table 4-3 from the evaluations of their performances on the validation set. After 5000 epochs, three-hidden-layer models showed better predictions than one or two layers; likewise, the 0.00001 learning rate (i.e., step size in the gradient descent algorithm) was better than 0.001 and 0.000001. On the other hand, different activation functions did not show great differences after 5000 epochs; the activation function introduces nonlinearity when passing inputs from one layer of neurons

to the next, mimicking the firing of a neutron for a given input. The most popular activation function is ReLU (rectified linear unit): when passing the ReLU function, the output equals to input when it is positive and zero otherwise. PReLU or parametric rectified linear unit has the same output as ReLU for a positive input but a slightly different output ($y = 0.25x$) for a negative input (x), instead of 0. We found that the highest R^2 (0.85) for the validation set was reached by the PReLU activation function after 15,000 epochs, with 0.00001 learning rate, 0.01 weight decay, three hidden layers, and the number of neurons on each layer as 512, 128, and 16.

The best FCNN model's performance is further shown in Figure 4-3 as the parity plot. For the 1085 data points used for training, the R^2 value reached 0.92 (Figure 4-3a) with RMSE of 0.40 and MAE of 0.19. More importantly, the model shows very good performance for the validation set: $R^2 = 0.85$, RMSE = 0.53, and MAE = 0.34. In other words, this trained model can predict log D values with an uncertainty of ~0.5. Of note, there are some cases with large errors in predicted log D values (Figure 4-3b), and we found that they are mainly from ligands with rare groups (such as -SR) for which we do not have a lot of data in the training set.

Table 4-3. Evaluations of the different combinations of the hyperparameters based on their performance on the validation set. The optimal set of hyperparameters are in bold.

Activation function		ELU	ELU	ELU	ELU	ELU	
Learning rate		0.00001	0.00001	0.00001	0.001	0.000001	
Weight decay		0.01	0.01	0.01	0.01	0.01	
epochs		5,000	5,000	5,000	5,000	5,000	
Neurons on hidden layers	1st	256	512	512	512	512	
	2nd		32	128	128	128	
	3rd			16	16	16	
R^2 on the validation set		0.36	0.58	0.77	0.18	0.65	
MAE on the validation set		0.68	0.66	0.41	0.85	0.52	
RMSE on the validation set		1.09	0.88	0.66	1.17	0.81	
Activation function		ReLU	PReLU	ELU	ReLU	PReLU	ELU
Learning rate		0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
Weight decay		0.01	0.01	0.01	0.01	0.01	0.01
epochs		5,000	5,000	10,000	10,000	10,000	15,000
Neurons on hidden layers	1st	512	512	512	512	512	512
	2nd	128	128	128	128	128	128
	3rd	16	16	16	16	16	16
R^2 on the validation set		0.77	0.73	0.80	0.84	0.81	0.84
MAE on the validation set		0.41	0.43	0.37	0.36	0.39	0.36
RMSE on the validation set		0.65	0.71	0.61	0.54	0.59	0.55
Activation function		ReLU	**PReLU**	PReLU	PReLU	PReLU	
Learning rate		0.00001	**0.00001**	0.00001	0.00001	0.00001	
Weight decay		0.01	**0.01**	0.01	0.01	0.01	
epochs		15,000	**15,000**	15,000	15,000	15,000	
Neurons on hidden layers	1st	512	**512**	512	256	512	
	2nd	128	**128**	64	64	128	
	3rd	16	**16**	16	8	32	
R^2 on the validation set		0.82	**0.85**	0.83	0.77	0.80	
MAE on the validation set		0.36	**0.34**	0.33	0.42	0.38	
RMSE on the validation set		0.58	**0.53**	0.52	0.65	0.61	

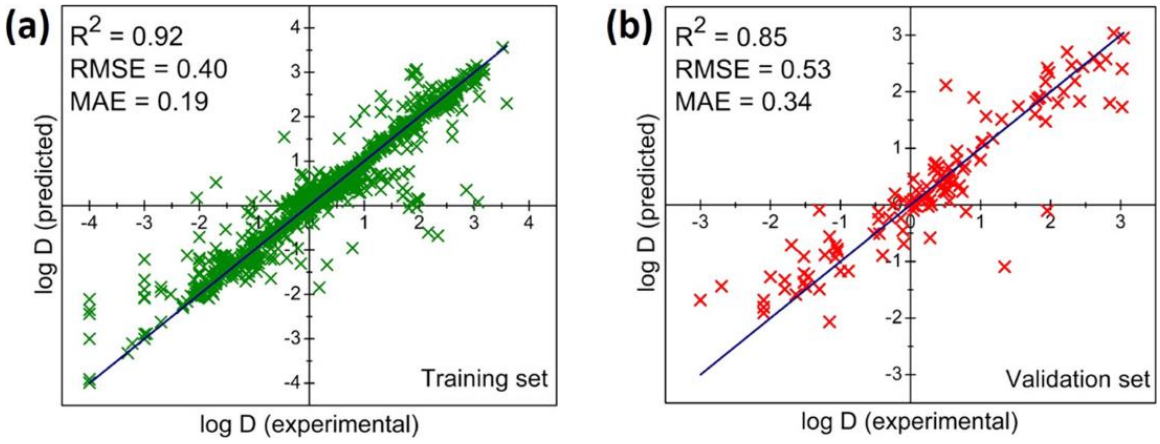


Figure 4-3. Performance of the best FCNN model. The parity plot between the predicted and experimental log D values: (a) training set and (b) validation set

4.4.3 Prediction on New Ligands

To further test our FCNN model, four new DGA ligands (**1**–**4** in Figure 4a) with different N-alkyl substituents were synthesized by our collaborators. It is known that subtle changes to the size of N,N'-alkyl groups affect DGA performance in Ln(III) separation.[8] The performance of DGAs that incorporate N,N'-alkyl substituents with branching is rather underexplored, for example, the substituents at γ (e.g., **2** and **3**) and δ (e.g., **1**) positions as opposed to α[38] and β[39] positions with respect to the amide nitrogen. Additionally, introduction of structure-rigidifying elements in DGA, such as δ-lactam motif in ligand **4**, opens new possibilities for chemically modifying the diglycolamide backbone to further alter separation behaviour. Benefits of implementing such structural modifications in DGAs are twofold: (1) extraction strength of Ln(III) can be tuned by varying the steric hindrance around the tridentate binding site; (2) the formation of third-phase in liquid-

liquid setting is more likely to be avoided due to improved hydrodynamic properties of these ligands and their Ln(III) complexes in the nonpolar solvent.

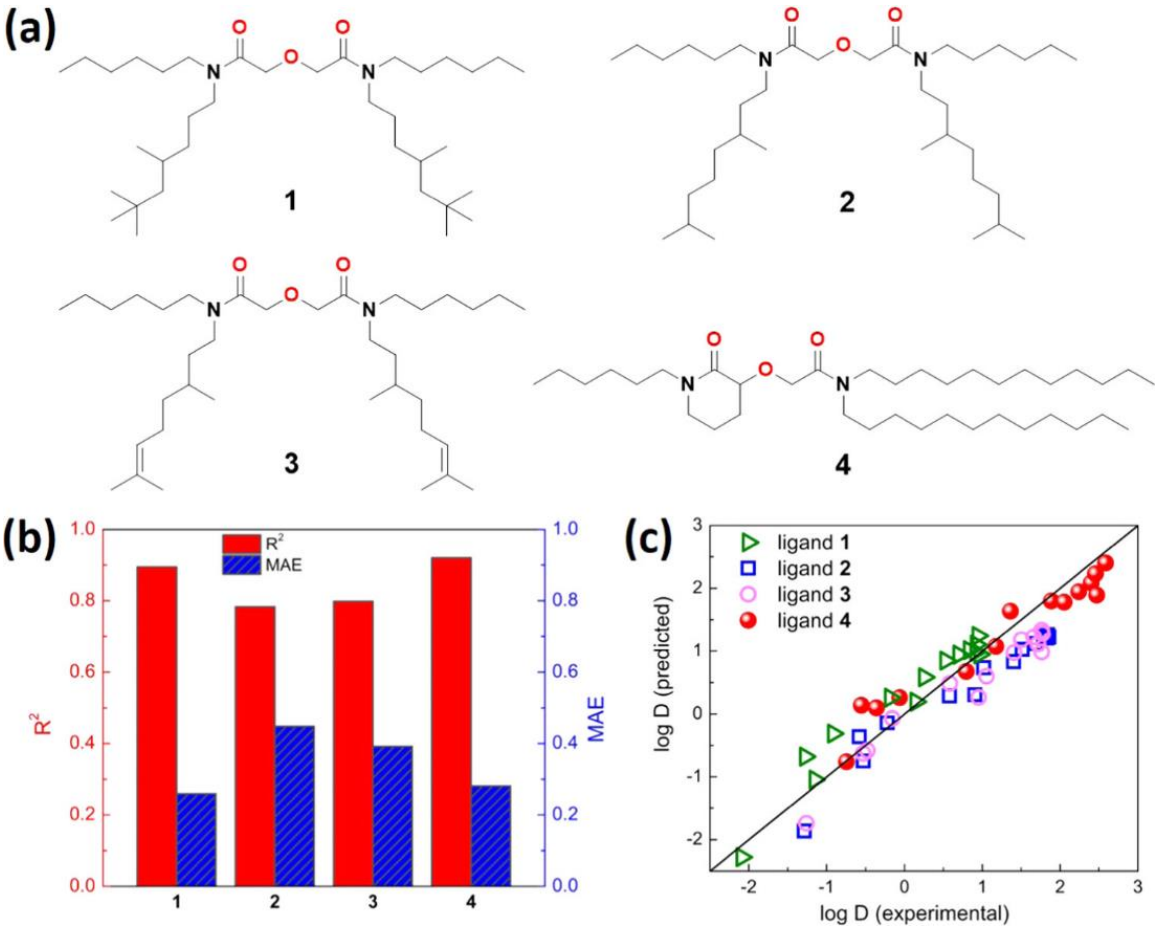


Figure 4-4. Predictions on new DGA ligands. (a) Chemical structures of new ligands **1**−**4** synthesized for Ln(III) extractions. (b) R^2 and MAE values of predicted log D for new ligands **1**−**4** in comparison with the measured values. (c) Parity plots between the predicted and experimental log D for ligands **1**−**4**; there are 14 data points for each ligand, representing 14 Ln(III)s extracted at the same conditions.

After their successful syntheses, ligands **1**−**4** were dissolved in an organic phase and contacted with mixed Ln(III) aqueous solutions in either hydrochloric or nitric acid. After phase separation, their *D* values were experimentally determined by measuring the

aqueous concentration of Ln(III) before and after extraction using inductively coupled plasma optical emission spectroscopy. To test the accuracy of our ML model to predict *log D* values, we fed these four new ligands together with their separation conditions into our well-trained FCNN model. As shown in Figure 4-4b, the predicted *log D* values are in good agreement with the experimental values, with R^2 ranging from 0.78 to 0.92; the MAE between the model predictions and experimental observations of *log D* in ligands **1**−**4** are 0.21, 0.41, 0.38, and 0.22, respectively. Even though this is a small test data set, the observed errors are similar to the validation set MAE of 0.34. This performance is consistent with the validation set shown in Figure 4-3b. The parity plot of the predicted vs experimental *log D* values for ligands **1**−**4** in Figure 4-4c highlights the very good performance of this ML model.

In addition to widely-used DGA ligands, our FCNN model also provides reliable predictions on three preorganized ligands sp-1-67, sp-1-79, and sp-1-99 (Figure 4-5). The R^2 between the model predictions and experimental observations of *log D* in these three preorganized ligands are 0.85, 0.70, and 0.74 supporting strong correlations; the MAE of predictions are 0.42, 0.58, and 0.50. It is predictive for our model to have great performance on ligands sp-1-67 and sp-1-79, since sp-1-67 is a BLPhen ligand and sp-1-79 is a BAPhen ligand which have analogs in the training dataset. Ligand sp-1-99 belongs to an extremely new classification the conjugation in the middle is reduced and no similar core structure is used for training. So, the successful prediction on sp-1-99 is proof in our model that even if only partial structure has been learned, accurate predictions can be achieved by the FCNN.

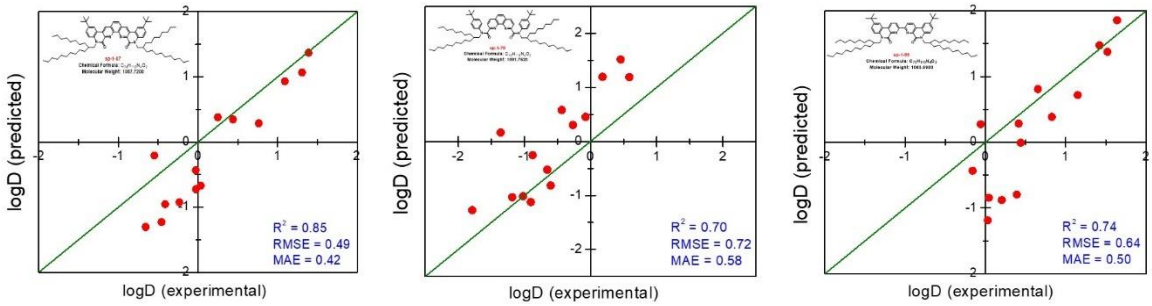


Figure 4-5. Predictions on new preorganized ligands sp-1-67, sp-1-79, sp-1-99.

Our model can be further improved by incorporating more data into the training dataset as they become available, especially for new ligand systems that are not represented in this work. This will help increase the accuracy (R^2) and lower the uncertainty (MAE) of the predicted log*D* values. More importantly, the trained model will allow us to rapidly evaluate new ligands for Ln(III) separation. Recent advances in the automatic generation of molecular structures based on string-based representations[40, 41] provide opportunities to create a large ligand database that can be fed into our ML model for high-throughput screening of new ligands for REE separations. In addition, our approach can be potentially extended to biomolecule-based ligands[42] and biogenic materials.[43]

In principle, our approach can also be used to screen extraction conditions. There are, however, some practical difficulties, with the main one being that researchers tend to report good extraction conditions while the less desirable conditions were not reported. As a result, the reported extraction conditions usually show limited coverage of the parameter space and there is insufficient data coverage in the extraction conditions in our data set. We think that high-throughput and automated experimentation of extraction conditions

would alleviate this insufficiency and make the future effort of predicting optimal extraction conditions with ML highly worthwhile.

4.5 Summary and conclusions

To advance the solvent-extraction separation of rare-earth elements, we have trained deep neural networks on the available experimental data of distribution coefficients measured for hundreds of ligands for 14 Ln(III) ions to accurately and quickly predict their distribution coefficients for a given ligand and the extraction conditions. To best represent the ligands, we found that a combination of molecular physicochemical descriptors and atomic extended-connectivity fingerprints yields the highest accuracy of the trained model on the validation set. We have further explored many combinations of hyperparameters that led to a set of optimal hyperparameters. The best trained model performed well on the validation set: R^2 = 0.85 and RMSE = 0.53. To further test our model, we synthesized four new ligands by modifying the diglycolamide (DGA) backbone and side chains and measured their log D values for Ln(III) ions; we found that the predicted distribution coefficients from our trained neural network agree well with the measured values. One can envision that our neural network can now be used to quickly predict log D values of Ln(III) ions for thousands to hundreds of thousands of ligands once they are generated. These log D values can be further evaluated to screen ligands for separation factors, that is, the ratios of log D values. Therefore, this work paves the way for further high-throughput screening of ligands to accelerate the discovery of new ligands for REE separations.

References

1. A. J. Hurd, R. L. Kelley, R. G. Eggert and M.-H. Lee, *MRS Bull.*, 2012, **37**, 405-410.

2. P. Siyushev, K. Xia, R. Reuter, M. Jamali, N. Zhao, N. Yang, C. Duan, N. Kukharchyk, A. Wieck and R. Kolesov, *Nat. Commun.*, 2014, **5**, 1-6.

3. V. Balaram, *Geosci. Front.*, 2019, **10**, 1285-1303.

4. M. Raha, S. Chen, C. M. Phenicie, S. Ourari, A. M. Dibos and J. D. Thompson, *Nat. Commun.*, 2020, **11**, 1-6.

5. J. A. Bogart, C. A. Lippincott, P. J. Carroll and E. J. Schelter, *Angew. Chem. Int. Ed.*, 2015, **54**, 8222-8225.

6. E. A. Mowafy and D. Mohamed, *Sep. Purif. Technol.*, 2014, **128**, 18-24.

7. R. J. Ellis, D. M. Brigham, L. Delmau, A. S. Ivanov, N. J. Williams, M. N. Vo, B. Reinhart, B. A. Moyer and V. S. Bryantsev, *Inorg. Chem.*, 2017, **56**, 1152-1160.

8. D. Stamberga, M. R. Healy, V. S. Bryantsev, C. Albisser, Y. Karslyan, B. Reinhart, A. Paulenova, M. Foster, I. Popovs and K. Lyon, *Inorg. Chem.*, 2020, **59**, 17620-17630.

9. X. He, X. Wang, Y. Cui, Z. Su, G. Ye, C. Lu and L. Li, *J. Radioanal. Nucl. Chem.*, 2021, **329**, 1019-1026.

10. A. Kovács, C. Apostolidis and O. Walter, *Inorganics*, 2019, **7**, 26.

11. S. Jansone-Popova, A. S. Ivanov, V. S. Bryantsev, F. V. Sloop, R. Custelcean, I. Popovs, M. M. Dekarske and B. A. Moyer, *Inorg. Chem.*, 2017, **56**, 5911-5917.

12. Y. Karslyan, F. V. Sloop, L. H. Delmau, B. A. Moyer, I. Popovs, A. Paulenova and S. Jansone-Popova, *RSC Adv.*, 2019, **9**, 26537-26541.

13. E. Mowafy and H. Aly, *Solvent Extr. Ion Exch.*, 2006, **24**, 677-692.

14. J. Sulakova, R. Paine, M. Chakravarty and K. Nash, *Sep. Sci. Technol.*, 2012, **47**, 2015-2023.

15. M. Simonnet, T. Kobayashi, K. Shimojo, K. Yokoyama and T. Yaita, *Inorg. Chem.*, 2021, **60**, 13409-13418.

16. Y. Sasaki, M. Matsumiya, M. Nakase and K. Takeshita, *Chem. Lett.*, 2020, **49**, 1216-1219.

17. X. Yang, L. Xu, Y. Hao, R. Meng, X. Zhang, L. Lei and C. Xiao, *Inorg. Chem.*, 2020, **59**, 17453-17463.

18. M. R. Healy, A. S. Ivanov, Y. Karslyan, V. S. Bryantsev, B. A. Moyer and S. Jansone-Popova, *Chem. Eur. J.*, 2019, **25**, 6326-6331.

19. G. König, F. C. Pickard, J. Huang, A. C. Simmonett, F. Tofoleanu, J. Lee, P. O. Dral, S. Prasad, M. Jones, Y. Shao, W. Thiel and B. R. Brooks, *J. Comput. Aided Mol. Des.*, 2016, **30**, 989-1006.

20. I. M. Kenney, O. Beckstein and B. I. Iorga, *J. Comput. Aided Mol. Des.*, 2016, **30**, 1045-1058.

21. J. Huuskonen, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 773-777.

22. T. Deng and G.-z. Jia, *Mol. Phys.*, 2020, **118**, e1600754.

23. S. Chaube, S. Goverapet Srinivasan and B. Rai, *Sci. Rep.*, 2020, **10**, 14322.

24. Q. Yang, Y. Li, J.-D. Yang, Y. Liu, L. Zhang, S. Luo and J.-P. Cheng, *Angew. Chem. Int. Ed.*, 2020, **59**, 19282-19291.

25. Z. Zhang, J. A. Schott, M. Liu, H. Chen, X. Lu, B. G. Sumpter, J. Fu and S. Dai, *Angew. Chem. Int. Ed.*, 2019, **58**, 259-263.

26. S. Wang, Y. Li, S. Dai and D. e. Jiang, *Angew. Chem. Int. Ed.*, 2020, **59**, 19645-19648.

27. K. Wu, Z. Zhao, R. Wang and G. W. Wei, *J. Comput. Chem.*, 2018, **39**, 1444-1454.

28. Z. Wang, Y. Su, W. Shen, S. Jin, J. H. Clark, J. Ren and X. Zhang, *Green Chem.*, 2019, **21**, 4555-4565.

29. D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742-754.

30. G. Landrum, *http://www.rdkit.org*, 2006.

31. B. Ramsundar, P. Eastman, P. Walters and V. Pande, *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more*, " O'Reilly Media, Inc.", 2019.

32. D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, 1977, **67**, 4517-4533.

33. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 8026-8037.

34. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort and J. Grobler, *arXiv preprint arXiv:1309.0238*, 2013.

35. B. J. Mincher, G. Modolo and S. P. Mezyk, *Solvent Extr. Ion Exch.*, 2009, **27**, 579-606.

36. M. K. Jha, A. Kumari, R. Panda, J. R. Kumar, K. Yoo and J. Y. Lee, *Hydrometallurgy*, 2016, **165**, 2-26.

37. Y. LeCun, Y. Bengio and G. Hinton, *nature*, 2015, **521**, 436-444.

38. G.-J. Sun, J.-H. Yang, H.-X. Yang, G.-X. Sun and Y. Cui, *Nucl. Sci. Tech.*, 2016, **27**, 75.

39. Y. Sasaki, Y. Sugo, K. Morita and K. L. Nash, *Solvent Extr. Ion Exch.*, 2015, **33**, 625-641.

40. F. Berenger and K. Tsuda, *J. Cheminformatics*, 2021, **13**, 88.

41. A. Nigam, R. Pollice, M. Krenn, G. dos Passos Gomes and A. Aspuru-Guzik, *Chem. Sci.*, 2021, **12**, 7079-7090.

42. Z. Dong, J. A. Mattocks, G. J. P. Deblonde, D. Hu, Y. Jiao, J. A. Cotruvo and D. M. Park, *ACS Cent. Sci.*, 2021, **7**, 1798-1808.

43. R. M. Pallares, M. Charrier, S. Tejedor-Sanz, D. Li, P. D. Ashby, C. M. Ajo-Franklin, C. Y. Ralston and R. J. Abergel, *J. Am. Chem. Soc.*, 2022, **144**, 854-861.

Chapter 5. Machine Learning Based high-throughput Screenings on Ligands for Ln(III) Selective Separations

5.1 Abstract

Novel and highly selective separation ligands are increasingly desired as a result of the rising demand for high-purity individual rare-earth elements. Herein, we combine the machine learning model and molecular structure automatic generation instruments to achieve high-throughput screening of new effective ligands. Four amide-trizainyl-phenanthroline ligands and several hexalkyl-nitrilotriacetamide (NTA) ligands are predicted with higher selectivity than the corresponding most selective analog reported in the literature. The more advanced investigation and evaluations of the structure-selectivity relationship on NTA ligands suggest that larger size and suitable dimethyl-hexyl group(s) at the terminal(s) of the substituents on amide groups can significantly improve the selectivity of this type of ligand. Symmetric NTA ligands with longer carbon chains could be the first set of synthesis targets given the difficulties in experimental synthesis.

5.2 Introduction

Lanthanides (Ln), as the main part of rare earth elements (REEs), have wide applications in the fuel industry, advanced materials, and life science as catalysis, [1-3] electrodes,[4] additive agents,[5] etc., leading their high purity isolations to one of the burning topics in separation chemistry. Solvent extraction is the primary chemical means to separate different trivalent lanthanides according to the slight differences in radii since their

other physical and chemical properties are highly similar. Explorations from the discovered selective ligands such as diglycolamide (DGA),[6-8] alkylated bis-triazinyl pyridines (BTP),[9] tributyl-phosphate (TBP),[10] 6,6′-bis-triazinyl-2,2′-bipyridine (BTBP),[11] bis(2-ethylhexyl) phosphoric acid (D2EHPA),[12] 2,9-bis-lactam-1,10-phenanthroline (BLPhen),[13] and 2,9-bis-triazinyl-1,10-phenanthroline (BTPhen).[14, 15] Computational evaluations on the selectivity of designed ligands can suggest synthesis targets and focus on the ligands having higher potentials. Distribution coefficient (D), the ratio of Ln(III) concentrations in the organic phase and aqueous phase after extraction, is a noteworthy quantitative indicator.

The traditional theoretical prediction on D or log D value originates from the thermodynamic theory for equilibrium constant and combines Gibbs free energy of multiple reactions, which requires great efforts on simulations for one predicted value.[16, 17] On the contrary, predictions from a well-trained machine learning (ML) model cost much shorter time and lower computational spending.[18] A large ligand database that can be fed into a machine learning model for high-throughput screening[19, 20] of new ligands for REE separations is also made possible by recent developments in the automatic generation of molecular structures based on string-based representations, such as simplified molecular-input line-entry system (SMILES).[21, 22]

Current automatic generation programs based on string-based representations are developed following two different principles: mutation and recombination. STONED-SELFIES is an example of molecule mutation generation.[22] It is a superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecule generations by using self-referencing embedded strings (SELFIES).[23] Mutations could

provide more chances to generate ligands with new core structures. However, it is more possible to detect some unstable or unusual substructures in new molecules, leading to low accuracy in predictions by the ML model because no such structural factors have been learned. Different from molecule mutations, more similarities could be found between the given ligands and the generated ones like the automatic molecule generative program used in this work: molecular generation by Fast Assembly of SMILES Fragments (FASMIFRA).[21] With help from FASMIFRA some known selective ligands (fed-in ligands) will be cut into small fragments and the reorganizations on these fragments will lead to new molecules. Reliable predictions could be achieved by the ML model because of few changes in the core structures of the fed-in ligands.

5.3 Methods

Separation factor (SF), the ratio of two distribution coefficients, is usually used to evaluate the selectivity of ligands in the experimental aspect. Especially, the ratio between La and Lu ($SF_{La/Lu}$) could work as a criterion for the ligand's general selectivity across the lanthanide series because the *D* value usually monotonously increases or decreases from La to Lu as decreasing of trivalent radii. Since our reported ML model directly outputs logD values,[18] log $SF_{La/Lu}$ is used for the following screenings. The overall workflow of ML-based high-throughput screenings for selective ligands in Ln(III) separations is shown in Figure 5-1a. Selective ligands from the lanthanide extraction database are used to generate new ligands for screening via FASMIFRA and some in-home scripts. Then, their predicted logD values on extracting La and Lu under different extraction conditions are

produced by the well-trained fully connected neural network (FCNN). Finally, their potentials in selective separations are compared by their calculated log $SF_{La/Lu}$ values.

Since seeking ligands with higher selectivity is the long-term aim in the lanthanide separation field, the best ligand belonging to different classifications reported in the literature is set as a reference ligand as shown in Figure 5-1b. According to stronger binding between light Ln(III) or heavy Ln(III), their log $SF_{La/Lu}$ values could be positive or negative. New ligands are compared to their reference own ligands based on the core structure to determine if they have the potential to the improvement of selectivity. No cross-comparison between the ligands favoring light Ln(III) and those favoring heavy Ln(III).

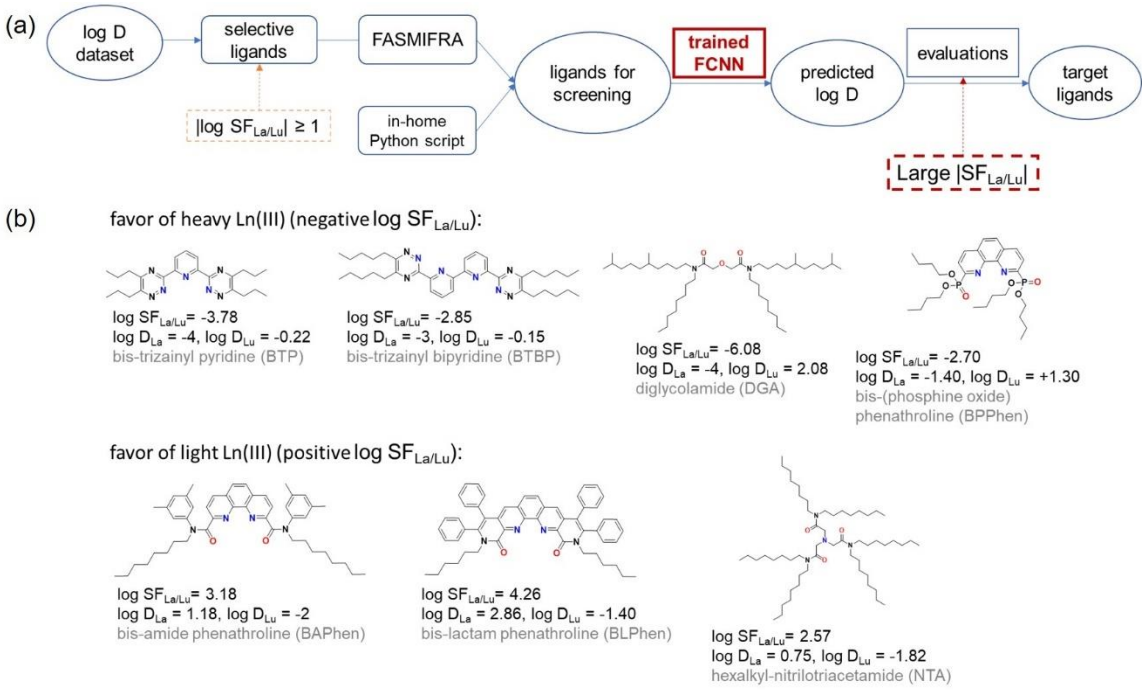


Figure 5-1. (a) The Workflow of machine learning based high-throughput screenings for selective ligands in Ln(III) separations. (b) Chemical structures and experimental extraction data of reference ligands in the lanthanide dataset.

5.4 Results and discussion

5.4.1 New database generated by FASMIFRA

According to the experimental performance on La/Lu separations, 55 ligands in the lanthanide extraction database built up in our previous work were defined as effective selective ligands since the difference between their extraction coefficients on La and Lu is higher than one magnitude ($|\log SF_{La/Lu}| \geq 1$). All these 55 ligands are fed into FASMIFRA to generate new molecules. As a result, 2562 new and distinct molecules without abnormal substituents are obtained consisting of the dataset for screening. The distributions of different types of ligands in the published selective ligand database and FASMIFRA-generated new ligand database (Figure 5-2) show more BAPhen, BLPhen, DGA and hexalkyl-nitrilotriacetamide (NTA) ligands are produced by FASMIFRA. Because of some new ligands containing two of these four structures at the same time, the total percentage in the FASMIFRA-generated database is larger than one hundred. All the new ligands within more than one core structure are fed into our well-trained neural network with different optimized extraction conditions. On the contrary, fewer DPA and MA ligands, and nearly no new BTBP and BTP ligands in the new ligand database, so further investigations are performed on new BAPhen, BLPhen, DGA and NTA ligands. However, the best DGA ligand reported has shown over 6 magnitudes on La/Lu separation (Figure 5-1b), which is already a huge difference, leading to less potential in new DGA ligands with high selectivity. Hence, the screening results on BAPhen, BLPhen, and NTA ligands are discussed separately below.

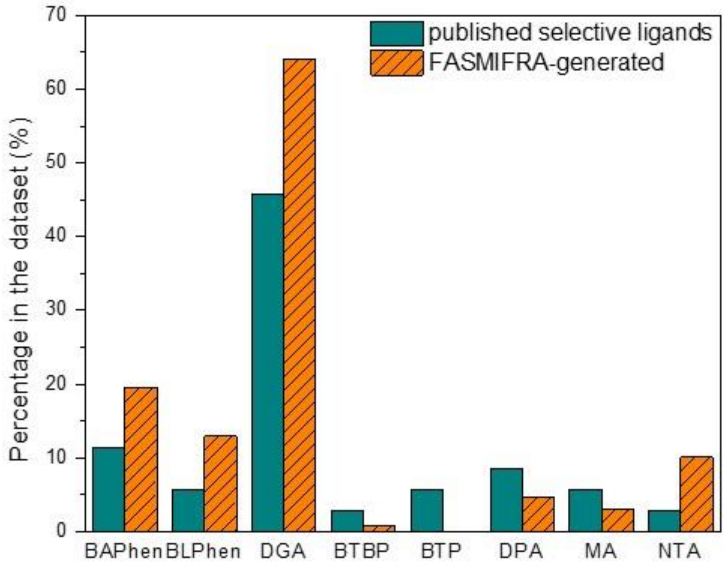


Figure 5-2. Distributions of different types of ligands in published selective ligand database (green) and FASMIFRA-generated new ligand database (orange).

5.4.2. New selective BAPhen-derivative ligands

Since BLPhen ligand is also a derivative of BAPhen ligands, all FASMIFRA-generated BAPhen, BLPhen and BAPhen derivatives are compared to the reference BAPhen ligand which has a log $SF_{La/Lu}$ as 3.18. However, none of the new ligands strictly belonging to BAPhen or BLPhen shows a log $SF_{La/Lu}$ higher than 3.18. Only four amide-trizainyl-phenanthroline ligands (Figure 5-3) show little higher selectivity (2% ~ 20%) than the reference BAPhen. There are 80 new BAPhen-derivative ligands in the FASMIFRA-generated dataset, including different replacements on one or two amide groups in the BAPhen, but only these four ligands are predicted with higher log $SF_{La/Lu}$ indicating the possible importance of trizainyl group in the selectivity. Trizainyl group provides one N-donor, so from all O-donor DGA ligands to N, O mixed BAPhen ligands, the ratio of N-

donor and O-donor further increases. Although the investigations on this new type of ligands, such as the relationship between substituents on trizainyl group and the selectivity are limited by the challenging synthesis of its asymmetric backbone, it could provide more inspiration to design new N, O mixed selective ligands.

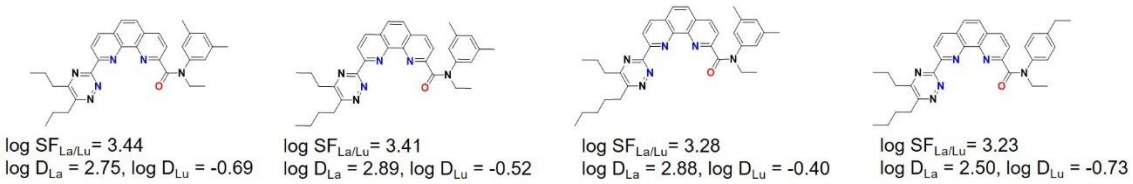


Figure 5-3. New BAPhen-derivative ligands predicted with high selectivity by neural networks and their predicted log D and log $SF_{La/Lu}$ values

5.4.3 New selective NTA ligands

More interesting than BAPhen-derivative ligands, more of the new NTA ligands produced by FASMIFRA show log $SF_{La/Lu}$ values higher than 2.57 which is the measurement of the best reported NTA. In total, 184 new NTA ligands are predicted by our well-trained neural network, and 65 of them show higher predicted selectivity, increasing our attention to further explore this type of ligand and understand the determinative factors of their structures and selectivity. The reported selective NTA ligands are both symmetric in that their six substituents on three amide groups are the same, octyl or dodecyl. However, most FASMIFRA-generated NTA ligands are asymmetric, resulting in a more difficult synthesis but a more diverse structure-property relationship. Then we rank these 184 new NTA ligands by their log $SF_{La/Lu}$ values and find that dimethyl-hexyl groups at the terminal of the substituents on amide groups significantly improve NTA ligand selectivity.

Among 65 NTA ligands with predicted higher selectivity, 20 of them have at least one dimethyl-hexyl group at the terminal, other 45 have all n-alkyl groups as the substituents on three amide groups. All these 20 NTA ligands within the dimethyl-hexyl group at the terminal are shown in Figure 5-4 together with their predicted extraction and separation data; the best 15 NTA ligands only n-alkyl groups as the substituents are shown in Figure 5-5, representatively. The new NTA ligand owning two dimethyl-hexyl groups has an extraordinary log $SF_{La/Lu}$ than all others, 50% numerical increase compared to the reference NTA. Five NTA ligands within the dimethyl-hexyl group(s), 25% of all, display higher selectivity than all NTA ligands only consisting of n-alkyl groups. In the remaining 119 new NTA ligands showing lower log $SF_{La/Lu}$ values, no dimethyl-hexyl group is detected. Thus, the dimethyl-hexyl groups at the terminal of the substituents on amide groups have an important effect on NTA selectivity. In addition to the effect of substituent, we notice that most new ligands predicted with high selectivity have more carbon atoms than the reference NTA. So, the size of the NTA ligand could be another factor of its selectivity.

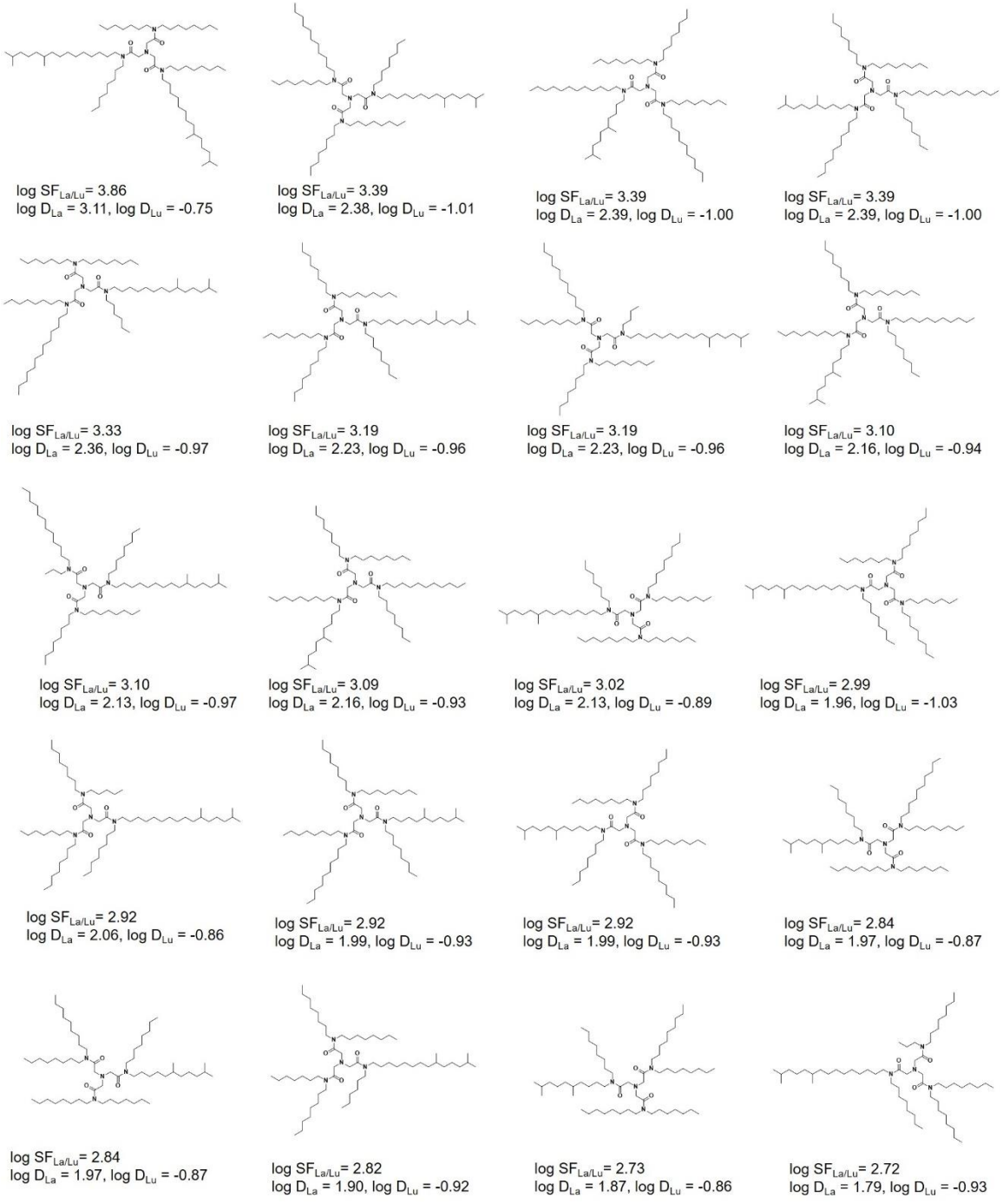


Figure 5-4. FASMIFRA-generated NTA ligands with dimethyl-hexyl group(s) at the terminal(s) of the substituents on amide groups and corresponding FCNN predicted selectivity.

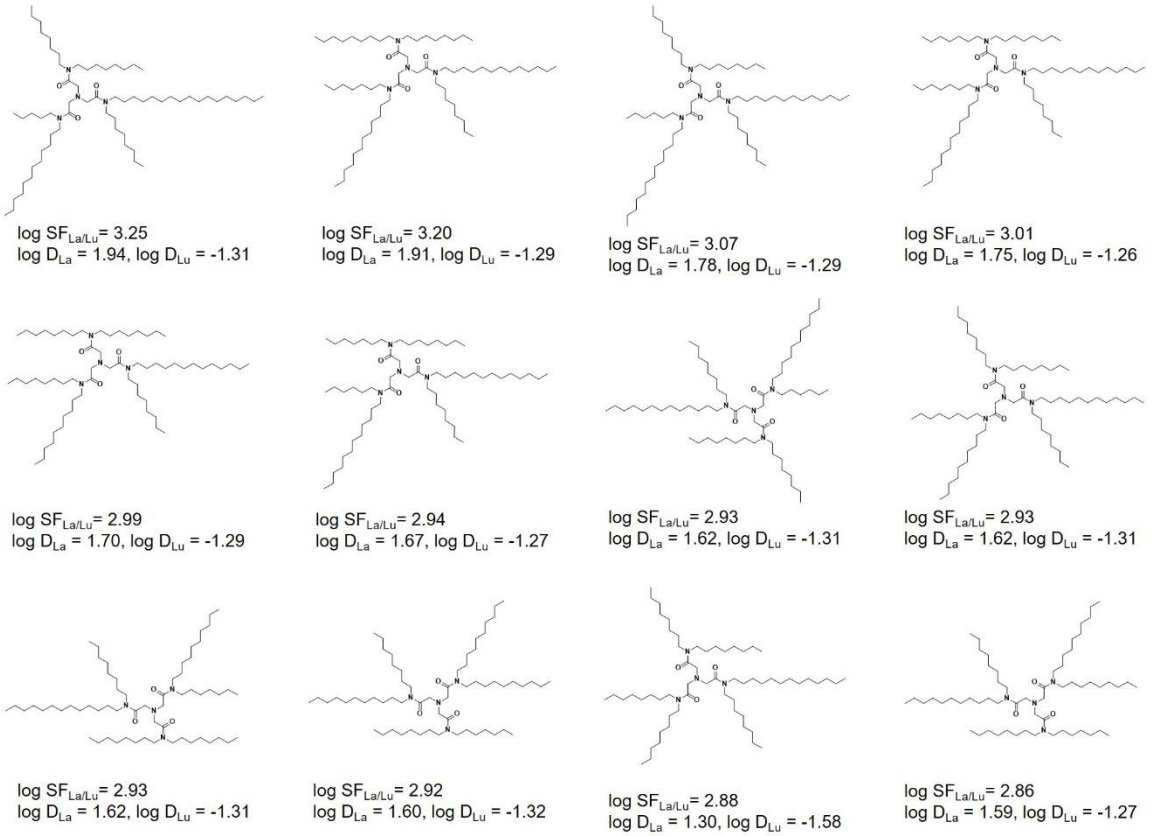


Figure 5-5. Representative FASMIFRA-generated NTA ligands with n-alkyl groups as the substituents on amide groups and corresponding FCNN predicted selectivity.

5.4.4 Tuning the selectivity of NTA ligands

Since molecule generation from FASMIFRA is controlled by the inputs and there are limited ligands in the current lanthanide database, we apply some in-home scripts for further NTA generations. According to the reported symmetric NTA ligands, three series of NTA are generated separately: NTA-C8 series in which the total carbon atoms are 54 same as the reference NTA; NTA-C10 series in which the total carbon atoms are 66 same as the symmetric decanoyl NTA; NTA-C12 series in which the total carbon atoms are 78 same as the symmetric dodecyl NTA. Hundreds of new NTA ligands within different

numbers of dimethyl-hexyl groups are generated and screened out for log $SF_{La/Lu}$ values. Based on different carbon atoms, the average log $SF_{La/Lu}$ values of NTA ligands and the highest log $SF_{La/Lu}$ in each series are compared in Table 5-1. No matter the average of the series or the highest performance, the NTA-12 series is better than the NTA-10 series better than the NTA-8 series. Their average log $SF_{La/Lu}$ have approximately one magnitudes difference in sequence, which is an appreciable distinction for selective ligands in lanthanides separations. Hence, it is conclusive that NTA ligands like larger sizes to reach higher selectivity so that the extension on carbon chains could be considered as a starting point to improve their selectivity.

Table 5-1. Selectivity of new NTA ligands according to different sizes.

	NTA-8 series	NTA-10 series	NTA-12 series
Average log $SF_{La/Lu}$	2.19 ± 0.12	3.23 ± 0.26	4.24 ± 0.23
Highest log $SF_{La/Lu}$	2.57	3.52	4.49

We further evaluate the effect of the number of dimethyl-hexyl groups on selectivity by ranking and statistics on different series of NTA ligands. In the NTA-C10 series, 91 of 95 new ligands are predicted with higher log $SF_{La/Lu}$ than the reference NTA and the highest three are shown in Figure 5-6a. The average and highest log $SF_{La/Lu}$ of ligands with different numbers of dimethyl-hexyl groups in the NTA-C10 series are compared in Table 5-2. Starting from one dimethyl-hexyl group, the average and highest log $SF_{La/Lu}$ keep decreasing, especially more than three dimethyl-hexyl groups the average only decreases by 0.04 from one to two dimethyl-hexyl groups and 0.13 from two to three but 0.26 from three to four. So, one or two dimethyl-hexyl groups are preferred by high

selective ligands in the NTA-C10 series, three is acceptable. In NTA-C12 series screenings, all new ligands show higher selectivity than the reference NTA based on FCNN predictions which double confirms the favor of large size. And the best three are shown in Figure 5-6b which can reach the magnitudes that the reference BLPhen ligand has. Then, similar statistics on the numbers of dimethyl-hexyl groups and selectivity are carried out on the NTA-12 series (Table 5-3). One or Two dimethyl-hexyl group(s) do not show an obvious difference in this series but a sharp decrease is found from three dimethyl-hexyl groups to four again: only 0.08 from two to three dimethyl-hexyl groups but 0.24 from three to four.

Table 5-2. Selectivity of new ligands in NTA-10 series with different numbers of dimethyl-hexyl groups

Numbers of dimethyl-hexyl groups	1	2	3	4	5
Average log $SF_{La/Lu}$	3.43	3.39	3.26	3.00	2.62
Highest log $SF_{La/Lu}$	3.52	3.48	3.30	3.07	2.78

Table 5-3. Selectivity of new ligands in NTA-12 series with different numbers of dimethyl-hexyl groups

Numbers of dimethyl-hexyl groups	1	2	3	4	5
Average log $SF_{La/Lu}$	4.38	4.39	4.31	4.07	3.74
Highest log $SF_{La/Lu}$	4.49	4.47	4.38	4.12	3.86

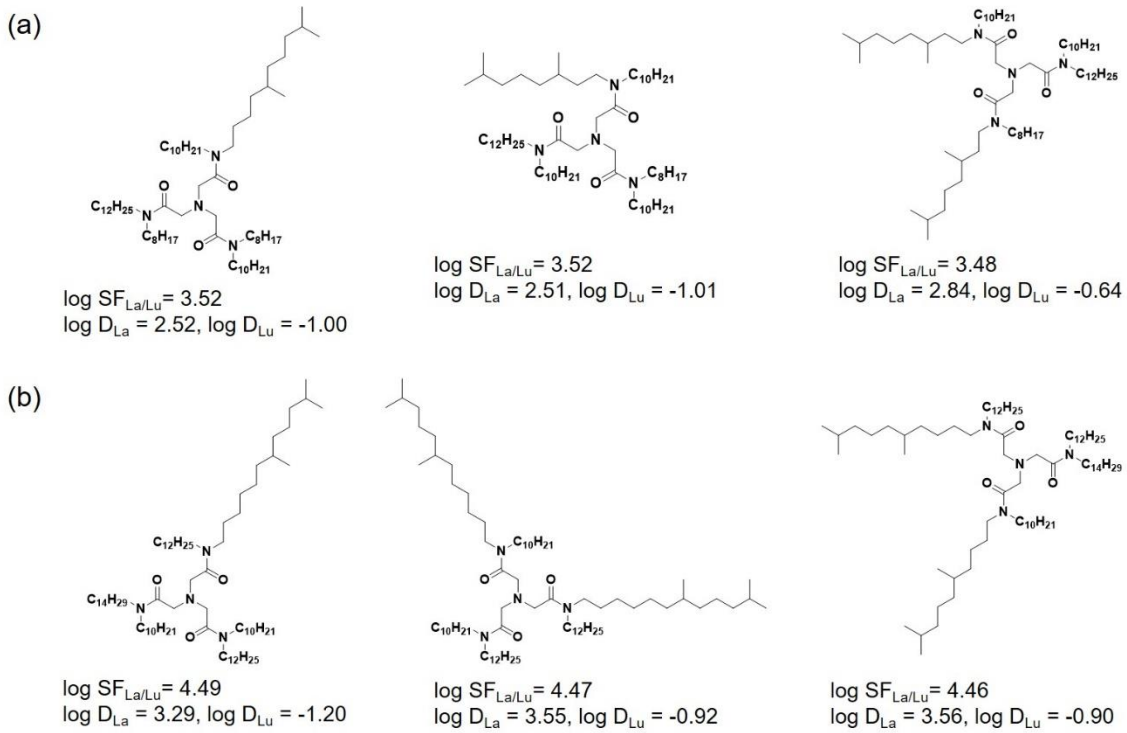


Figure 5-6. Representative high selective NTA ligands in (a) NTA-10 series (b) NTA-12 series and corresponding FCNN predicted selectivity.

Machine-learning-based high-throughput screenings on about 3,000 new ligands generated by FASMIFRA and in-home scripts suggest a few amide-trizainyl-phenanthroline ligands and amounts of NTA ligands within high potential in lanthanides selective separations. Due to the difficulties in experimental synthesis, further verifications on these proposed ligands are still in process. Another factor that may have a little influence on the accuracy of our predictions is only the reference NTA ligand has all experimental log *D* values for all Ln(III) and has been used in the training set. Other NTA ligands involved in the training set only have log *D* of Eu(III) since they are reported in Eu(III)/Am(III) separation which is a typical model of lanthanide and actinide separation.

Although our model has proved the strong capacity of great predictions on the preorganized ligands that only partial structure has been learned, the prediction accuracy still depends on the amounts of similar ligands in the training set.

5.5 Summary and conclusions

Machine learning based high-throughput screenings on thousands of ligands generated by Fast Assembly of SMILES Fragments (FASMIFRA) suggest four amide-trizainyl-phenanthroline ligands and over 60 hexalkyl-nitrilotriacetamide (NTA) ligands have a high possibility to improve the selective separations of Ln(III). Further computational investigations on hundreds of NTA ligands confirm two structure-activity relationships. One is a larger size preferred for an NTA ligand showing higher selectivity. The other is for the NTA ligands owning the same size, one or two dimethyl-hexyl group(s) at the terminal(s) of the substituents on amide groups can improve its selectivity. Considering challenges in experimental synthesis, symmetric NTA ligands with longer carbon chains could be the first series of synthesis targets.

Reference

1. M. Shibasaki and N. Yoshikawa, *Chem. Rev.*, 2002, **102**, 2187-2210.
2. X. Yu, S. Seo and T. J. Marks, *J. Am. Chem. Soc.*, 2007, **129**, 7244-7245.
3. A. R. Richard and M. Fan, *J. Rare Earths*, 2018, **36**, 1127-1135.
4. J. Kido and Y. Okamoto, *Chem. Rev.*, 2002, **102**, 2357-2368.

5. F. Lofaj, R. Satet, M. Hoffmann and A. de Arellano Lopez, *J. Eur. Ceram. Soc.*, 2004, **24**, 3377-3385.

6. E. A. Mowafy and D. Mohamed, *Sep. Purif. Technol.*, 2014, **128**, 18-24.

7. R. J. Ellis, D. M. Brigham, L. Delmau, A. S. Ivanov, N. J. Williams, M. N. Vo, B. Reinhart, B. A. Moyer and V. S. Bryantsev, *Inorg. Chem.*, 2017, **56**, 1152-1160.

8. X. He, X. Wang, Y. Cui, Z. Su, G. Ye, C. Lu and L. Li, *J. Radioanal. Nucl. Chem.*, 2021, **329**, 1019-1026.

9. A. Kovács, C. Apostolidis and O. Walter, *Inorganics*, 2019, **7**, 26.

10. L. K. Sinclair, J. W. Tester, J. F. H. Thompson and R. V. Fox, *Ind. Eng. Chem. Res.*, 2019, **58**, 9199-9211.

11. M. R. Foreman, M. J. Hudson, M. G. Drew, C. Hill and C. Madic, *Dalton Trans.*, 2006, 1645-1653.

12. A. Kumari, R. Panda, J. Y. Lee, T. Thriveni, M. K. Jha and D. D. Pathak, *Sep. Purif. Technol.*, 2019, **227**, 115680.

13. S. Jansone-Popova, A. S. Ivanov, V. S. Bryantsev, F. V. Sloop, R. Custelcean, I. Popovs, M. M. Dekarske and B. A. Moyer, *Inorg. Chem.*, 2017, **56**, 5911-5917.

14. Y. Yang, J. Liu, L. Yang, K. Li, H. Zhang, S. Luo and L. Rao, *Dalton Trans.*, 2015, **44**, 8959-8970.

15. D. M. Whittaker, T. L. Griffiths, M. Helliwell, A. N. Swinburne, L. S. Natrajan, F. W. Lewis, L. M. Harwood, S. A. Parry and C. A. Sharrad, *Inorg. Chem.*, 2013, **52**, 3429-3444.

16. G. König, F. C. Pickard, J. Huang, A. C. Simmonett, F. Tofoleanu, J. Lee, P. O. Dral, S. Prasad, M. Jones and Y. Shao, *J. Comput. Aided Mol. Des.*, 2016, **30**, 989-1006.

17. I. M. Kenney, O. Beckstein and B. I. Iorga, *J. Comput. Aided Mol. Des.*, 2016, **30**, 1045-1058.

18. T. Liu, K. R. Johnson, S. Jansone-Popova and D.-e. Jiang, *JACS Au*, 2022, **2**, 1428-1434.

19. S. Nishimura, S. D. Le, I. Miyazato, J. Fujima, T. Taniike, J. Ohyama and K. Takahashi, *Catalysis Science & Technology*, 2022, **12**, 2766-2774.

20. M. Wang, Q. Xu, H. Tang and J. Jiang, *ACS Applied Materials & Interfaces*, 2022, **14**, 8427-8436.

21. F. Berenger and K. Tsuda, *J. Cheminformatics*, 2021, **13**, 88.

22. A. Nigam, R. Pollice, M. Krenn, G. dos Passos Gomes and A. Aspuru-Guzik, *Chem. Sci.*, 2021, **12**, 7079-7090.

23. M. Krenn, F. Häse, A. Nigam, P. Friederich and A. Aspuru-Guzik, *Machine Learning: Science and Technology*, 2020, **1**, 045024.

Chapter 6. Separations of Lanthanides and Actinides via Machine Learning

6.1 Abstract

Lanthanides (Ln), which make up the majority of rare-earth elements, must be separated in order to fully fulfill their potential as vital components in several crucial technologies. The most practical rare-earth separation method, solvent extraction, still relies heavily on trial-and-error, a low throughput and inefficient method, to find novel ligands for improved rare-earth separations. It is necessary to develop a prediction model that enables high-throughput ligand screening in order to find the right ligands for improved separation performance. This work was motivated by several successful cases of ligands effective in Eu(III)/Am(III) separation could be further utilized in Ln(III) selective separations. We have trained deep neural networks to accurately and quickly predict distribution coefficients of trivalent lanthanides (Ln(III)) and actinides (Am (III)) for a given ligand and the extraction conditions. 194 ligands and measurements from over 2,000 extractions make up the experimental data on distribution coefficients that are currently accessible. For the depiction of a ligand, atomic extended-connectivity fingerprints and molecular physicochemical descriptors are combined. The best trained model, with R^2 = 0.83 and MAE = 0.45, performed well on the validation set after testing and comparing several alternative combinations of hyperparameters. The trained neural network examines the behavior of ligands in the database that are solely employed for the Eu(III)/Am(III) separation. Some bis-phosphine oxide phenanthroline, dithiophosphinic acids and bis-pyrazole phenanthroline show trustworthy high potentials in Ln(III) selective separations.

6.2 Introduction

Fifteen lanthanides (Ln) along with Sc and Y are named as rare earth elements (REEs) due to their dispersed distributions in the earth's crust and exceedingly tough exploitations at the early stage. Even after over 200 years' development, separation and purification of REEs are still difficult because their minerals are always paragenetic in nature and their physical and chemical properties are extremely similar. On the other hand, lanthanides have wide applications in the fuel industry, new materials, and life science[1, 2] as catalysis,[3-5] electrodes,[6] additive agents,[7] etc., making their high purity isolations one of the burning topics in separation chemistry. Solvent extraction is the primary chemical means to separate different lanthanides according to the different binding strengths of the trivalent lanthanides with the identical organic ligand. Certainly, more adjacent in the periodic table, more slight differences in radii of their trivalent cations, more challengeable for complete purifications. Based on functions, valid ligands could be classified into two major categories one of which is more sensitive to heavier lanthanides than the lighter, such as bis(2-ethylhexyl) phosphoric acid (D2EHPA),[8] N,N,N',N'-tetraoctyldiglycolamide (TODGA),[9-11] and trialkylphosphine oxides (e.g., Cyanex 923).[12] However, fewer ligands could extract light lanthanide(s) from the aqueous phase to the organic phase, like 2,9-bis-amide-1,10-phenathroline (BAPhen).[13-15] A supreme extractant has huge differences in its binding strengths to fifteen Ln(III) is the desire of research.

In the term of electron donors in the extractants used in industrial extractions, the most effective ones are oxygen-donor chelating ligands such tributyl-phosphate (TBP),[16] diglycolamide (DGA)[17-19] and D2EHPA. Ligands with all soft N-donor atoms such as

alkylated bis-triazinyl pyridines (BTP),[20] 6,6′ -bis-triazinyl-2,2′ -bipyridine (BTBP),[21] and 2,9-bis-triazinyl-1,10-phenanthroline (BTPhen)[22, 23] are another strategy of extractants for isolated Ln(III) in nitric acid. Recently, alternative ligands combining hard O-donor and soft N-donor atoms have been recognized as efficient separating agents and have attracted interest in that amide oxygen donors can provide stronger metal-ion binding. One example is 2,9-bis-lactam-1,10-phenanthroline (BLPhen)[24] which was reported with unparalleled selectivity for light trivalent lanthanides.[25]

In fact, a lot of selective ligands in Ln(III) separations were utilized in lanthanide and actinide (An) separation because of larger differences in properties and more urgent requirements such as the advanced reprocessing of high-level waste.[26, 27] For example, the BLPhen ligand was reported with exceptionally high efficiency to separate Am(III) over Eu(III) in 2016, then was applied for Ln(III) extractions. In other words, extractants used in separating Ln(III) and An(III) have a higher possibility than other molecules to be effective in Ln(III) separations. The first step to use the ligands in Ln(III)/An(III) separation to improve the separating efficiency of Ln(III)s is to process a computational model that could accurately predict ligand's extraction properties such as distribution coefficient (D) on both Ln(III) and An(III). Certainly, the subsequent ligand design would be more efficient if the model can obtain some structure-function relationships during the simulations. Thus, in this work, by training deep neural networks on experimental data of observed D values and by adequately encoding ligands, Ln(III) or An(III), and experimental circumstances, we have created a predictive model that successfully predicts D values for a given ligand.

6.3 Computational Method

Data collection. Lanthanide extraction data including 1,202 log D values come from our previous work.[28] Additional 821 log D values of Eu (III) and trivalent actinides (Am, Cm, Bk, Cf, Es) in the database were collected from the scientific literature where a single neutral ligand was the only extractant used to extract An(III) from the aqueous phase to the organic phase consisting of one or two different solvents. For each data point, the inputs include sequentially the representation of the ligand, descriptors of the extraction conditions, descriptors of the lanthanide or actinides.

Representation of ligands. The first 2,048 inputs of each data point are Extended-Connectivity Fingerprints[29] (ECFP) of the ligand; the next 208 inputs are RDKit descriptors.[30] They are both generated from the simplified molecular-input line-entry system (SMILES) expression of the ligand by the DeepChem package.[31] Chirality is considered in ECFP and other parameters use default settings: radius of fingerprint = 2; length of generated bit vector = 2,048; bond order considered; feature descriptors not used. RDKit descriptors use default parameters: binary descriptors of fragments like 'fr_XXX" are returned; avg = True for the Ipc (information of polynomial coefficients) descriptor[32] to return the information content divided by the total population. The names of the 208 descriptors returned by the RDKit module are listed in the Excel file, including molecular weight, number of valence electrons, partial charges, electrotopological state indexes, etc.

Descriptors of lanthanides and actinides. Nine numerical descriptors available for all 14 lanthanides and 5 actinides are used to characterize the metal in extraction: atomic number,

outer shell electrons of the atom, melting point (K) of the metal, boiling point (K) of the metal, Density (g/cm^3) of the metal, covalent radius (Å) of the atom, Pauling electronegativity of the atom, ionic radius (Å) of trivalent ions and standard entropy ($J \cdot mol^{-1} \cdot K^{-1}$) of the metal at ambient conditions.

Descriptors of the extraction conditions. These descriptors include the concentration of the ligand (mM), volume ratios of organic solvents a and b, descriptors of organic solvent a, descriptors of organic solvent b, the dipole moment of the inorganic acid in the aqueous phase (D), inorganic acid concentration (M), and Temperature (K) in sequence. For the one-component organic phase, the volume ratio of organic solvent a is 10, the volume ratio of organic solvent b is 0, and all descriptors of organic solvent b are set as 0. Descriptors of organic solvent a or b consist of seven numerical properties: molar mass (g/mol), density (g/mL), boiling point (K), melting point (K), dipole moment (D), solubility in water (g/L), and log P ($P=[a]_{n\text{-}octanol}/[a]_{water}$, partition coefficient in n-octanol/water at ambient temperature). For the extraction data which provides only the pH value of the aqueous phase, the inorganic acid concentration is determined by this pH value.

Details of the deep learning model and the training process. The training of fully connected neural networks (FCNNs) is performed via the PyTorch package (version 1.9.1)[33] with L1 type loss function, SGD optimizer, and L2 regularization for weight decay. The weight initializations obey the default normal distributions. Mean-absolute-error (MAE), and coefficient of determination (R^2) as calculated via the scikit-learn module were used as metrics for evaluation during the training process.[34]

6.4 Results and discussion

6.4.1 Data and workflow

The workflow of our ML approach is summarized in Figure 6-1a. The input data comprises three parts: Ln(III) or An(III), ligand, and solvent extraction conditions. First, the ligand, represented by a string-based name (SMILES), is fed into RDKit[30] – a cheminformatics toolkit that automatically generates 208 descriptors for the ligand. The RDKit descriptors are then combined with the extended connectivity fingerprints (ECFPs)[29] for a more detailed representation of the ligand. Nine descriptors are used for each Ln element; solvent extraction conditions such as temperature, the concentration of the ligand, and physical properties of organic solvents are also part of the input. In total, 2286 inputs are used for each output log*D* value.

In addition to 1,202 log D values of lanthanide extraction from our previous work,[28] 821 log D values of Eu (III) and trivalent actinides (Am, Cm, Bk, Cf, Es) extracted by 85 ligands are collected from literatures into our lanthanides and actinides database. More than diglycolamides (DGA),[17, 18, 35, 36] alkylated bis-triazinyl pyridines (BTP),[20] and 2,9-bis-lactam-1,10-phenanthroline (BLPhen),[24, 37] and other ligands used for Ln(III) separations,[38-40] new classes of ligands only reported for Ln(III)/An(III) separations including benzoxazole,[41] benzothiazole,[41] bipyridine dicarboxamide (BPDA),[42] tripyridine dicarboxamide (TPDA),[43] bis-pyrazole phenanthroline (BPyPhen),[44] dithiophosphinic acids,[45] pyridine phosphine oxide (POPy),[46] etc. enrich the diversity of the database. Meanwhile, more novel BAPhen and 2,9-bis-phosphineoxide-1,10-phenanthroline (BPPhen) ligands are involved in the new database than the lanthanides database because

of increasing reports on them in Eu(III)/Am(III) separation in recent year.[47-49] Better predictions on BAPhen-derivative ligands could be expected due to the larger number in the training dataset. All SMILES expressions of the 194 ligands used in this work are listed in Appendix A, where the first 109 ligands were also used for previous work.

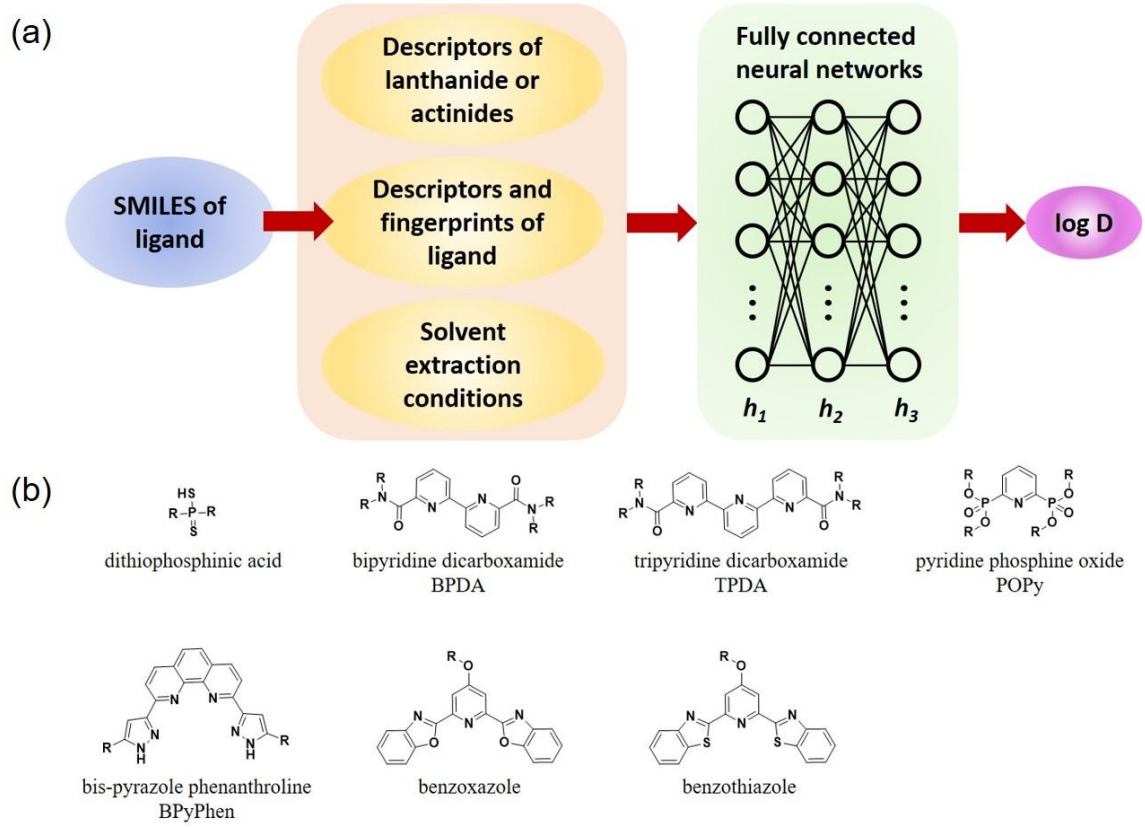


Figure 6-1. (a) The workflow of predicting log*D* of Ln(III) and selected An(III) extracted by a ligand via fully connected neural networks with three hidden layers. (b) Chemical structures of ligands only used for Ln(III)/An(III) separation in the dataset.

6.4.2 Training and model performance

Later, 200 data points randomly selected from our dataset, covering all 14 Ln(III) and including 52 Am(III) log *D* values are defined as the held-out validation set which is unknown to the neural networks until the training finishes. Fully connected neural networks (FCNNs) where every neuron in one layer is connected to every neuron in the next layer were used as the core of our approach for deep learning.[50] The training of the FCNNs was performed with the PyTorch package.[33] Plenty of FCNN models are evaluated and the best is achieved after 20,000 epochs with settings as 0.0001 learning rate, PReLU activation functions, 0.01 weight decay, three hidden layers, and numbers of neurons on each layer as 512, 128, and 16. The best FCNN model's performance is shown in Figure 6-2 as the parity plot. Although there are some outliers when experimental log *D* is relatively small or large (< -2 or > 2), the R^2 value of the training set based on 1834 data points reaches 0.90 (Figure 6-2a) and MAE on 1834 data points is 0.28, a small value. Then, this best FCNN model also shows an excellent correlation (R^2 = 0.83) between predicted log *D* and experimental log *D* on the held-out validation set (Figure 6-2b). The MAE on the validation set is 0.45 and 67% of the validation examples have prediction errors of less than 0.5 of log *D* units (in the range between cyan dash lines in Figure 6-2b). More interesting, the best FCNN model has consistent predictions on 52 Am(III) log *D* values and 148 Ln(III) log *D* values in the validation set that R^2 is 0.79 and MAE is 0.45 for Am(III) only.

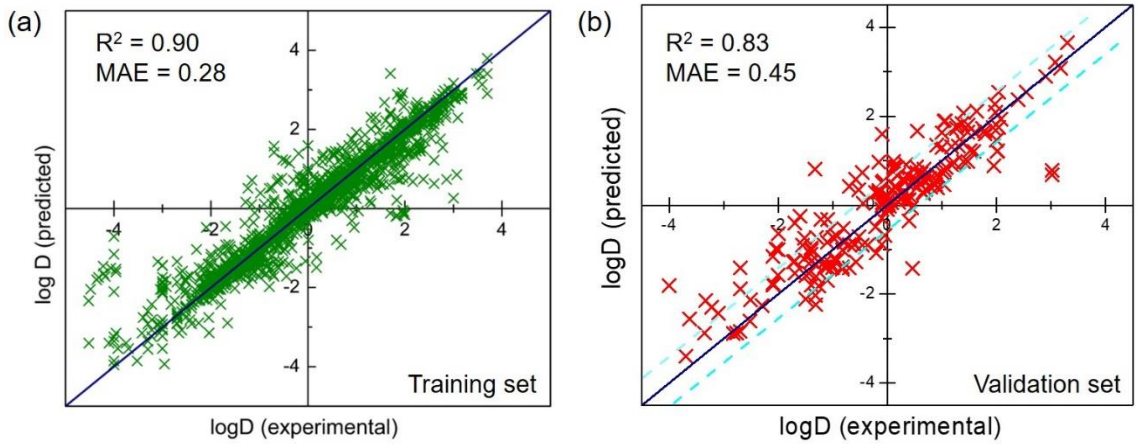


Figure 6-2. Performance of the best FCNN model. The parity plot between the predicted and experimental log D values: (a) training set and (b) validation set.

6.4.3 From Ln(III)/An(III) separations to Ln(III) selective separations

Since 85 ligands in our database were only reported on Ln(III)/An(III) separations in literature and our FCNN model has consistent reliable predictions on Am(III) and Ln(III) log D values, we use our FCNN model to provide log D_{La} and log D_{Lu} values for these 85 ligands to evaluate their potentials in Ln(III) selective separations. The extraction conditions for screenings on different classes of these 85 ligands are determined by their reported extraction conditions in Eu(III)/Am(III) separations. The separation factor, the ratio of distribution coefficients between two cations, is a measurable indicator of ligand selectivity. Since our FCNN model directly predicts log D values, log $SF_{La/Lu}$ is used for evaluation, in which $|\log SF_{La/Lu}| \geq 1$ represents an effective selective separation and larger $|\log SF_{La/Lu}|$, higher selectivity.

Some novel BPPhen ligands, as analogs of BAPhen ligands, are predicted with excellent selectivity as shown in the first row in Figure 6-3. The most selective BAPhen

ligand has been reported as a log $SF_{La/Lu}$ as 3.18,[25] and three BPPhen ligands screened out from selective ligands in Eu(III)/Am(III) separations show higher or close to that value. As a result, more modifications and expansion on BPPhen ligands could improve the Ln(III) selective separations. In addition to BPPhen ligands, two dithiophosphinic acids show log $SF_{La/Lu}$ values larger than one (second row in Figure 6-3). Since dithiophosphinic acids have never been applied for Ln(III) selective separations, fewer understandings, and investigations on their structure-selectivity relationships so more opportunities to design, improve and explore this unique type of ligand. The other significant ligand in these screenings is a BPyPhen showing a log $SF_{La/Lu}$ as 2.84. Although BPyPhen has a phenanthroline backbone as BAPhen, all electron donors are nitrogen so the hardness difference between nitrogen and oxygen could lead to a large difference in binding strength which quantitively affects ligand selectivity. As a result, although this BPyPhen does not show a higher $|\log SF_{La/Lu}|$ than the most selective BLPhen, its exclusive structure raises more inspirations of different types of phenanthroline derivatives in Ln(III) selective separations.

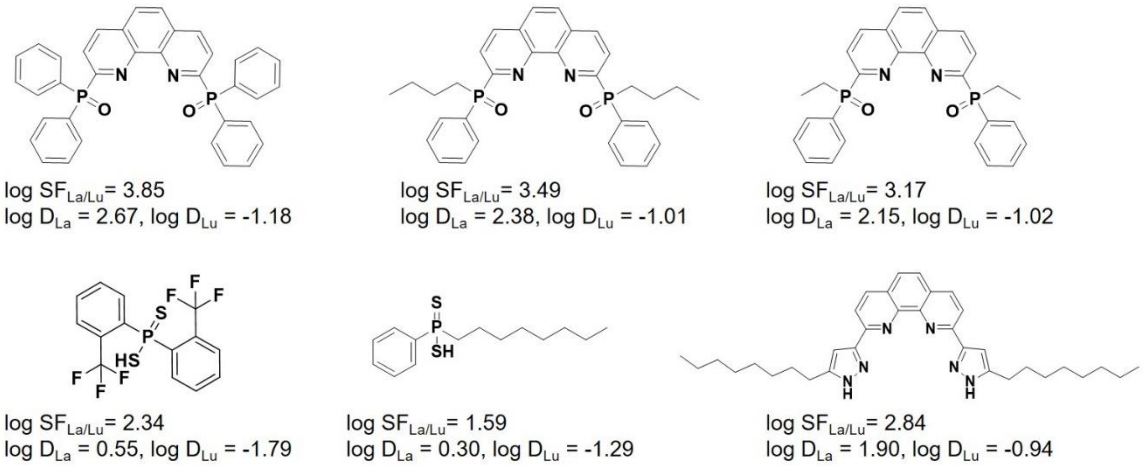


Figure 6-3. New ligands have high predictive potentials in Ln(III) selective separations which were previously reported in Eu(III)/Am(III) separations and their corresponding FCNN predicted selectivity.

6.5 Summary and conclusions

Inspired by several successful cases of ligands effective in Eu(III)/Am(III) separation could be furtherly utilized in Ln(III) selective separations, we have trained deep neural networks to accurately and quickly predict distribution coefficients of Ln(III) and Am(III) for a given ligand and the extraction conditions. The available experimental data of distribution coefficients involve 194 ligands and measurements over 2,000 extractions. A combination of molecular physicochemical descriptors and atomic extended-connectivity fingerprints is used for the representation of a ligand. After trying and comparing many different combinations of hyperparameters, the best trained model performed well on the validation set: R^2 = 0.83 and MAE = 0.45. The trained neural network screens ligands in the database that are only used for Eu(III)/Am(III) separation for their behaviors on the La(III)/Lu(III) separation. Several bis-phosphine oxide

phenanthrolines, dithiophosphinic acids, and bis-pyrazole phenanthrolines exhibit reliable high potentials in futural Ln(III) selective separations.

Reference

1. Y. He, A. Lopez, Z. Zhang, D. Chen, R. Yang and J. Liu, *Coord. Chem. Rev.*, 2019, **387**, 235-248.

2. K. Staszak, K. Wieszczycka, V. Marturano and B. Tylkowski, *Coord. Chem. Rev.*, 2019, **397**, 76-90.

3. M. Shibasaki and N. Yoshikawa, *Chem. Rev.*, 2002, **102**, 2187-2210.

4. X. Yu, S. Seo and T. J. Marks, *J. Am. Chem. Soc.*, 2007, **129**, 7244-7245.

5. A. R. Richard and M. Fan, *J. Rare Earths*, 2018, **36**, 1127-1135.

6. J. Kido and Y. Okamoto, *Chem. Rev.*, 2002, **102**, 2357-2368.

7. F. Lofaj, R. Satet, M. Hoffmann and A. de Arellano Lopez, *J. Eur. Ceram. Soc.*, 2004, **24**, 3377-3385.

8. T. Sato, *Hydrometallurgy*, 1989, **22**, 121-140.

9. Y. Sasaki, Y. Sugo, S. Suzuki and S. Tachimori, *Solvent Extr. Ion Exch.*, 2001, **19**, 91-103.

10. S. A. Ansari, P. Pathak, P. K. Mohapatra and V. K. Manchanda, *Chem. Rev.*, 2012, **112**, 1751-1772.

11. A. Wilden, P. M. Kowalski, L. Klaß, B. Kraus, F. Kreft, G. Modolo, Y. Li, J. Rothe, K. Dardenne, A. Geist, A. Leoncini, J. Huskens and W. Verboom, *Eur. J. Chem.*, 2019, **25**, 5507-5513.

12. M. Reddy, R. Luxmi Varma, T. Ramamohan, S. K. Sahu and V. Chakravortty, *Solvent Extr. Ion Exch.*, 1998, **16**, 795-812.

13. Y. A. Ustynyuk, N. Borisova, V. Babain, I. Gloriozov, A. Manuilov, S. Kalmykov, M. Y. Alyapyshev, L. Tkachenko, E. Kenf and N. Ustynyuk, *Chem. Commun.*, 2015, **51**, 7466-7469.

14. J. Dehaudt, N. J. Williams, I. A. Shkrob, H. Luo and S. Dai, *Dalton Trans.*, 2016, **45**, 11624-11627.

15. X.-Z. Li, L.-P. Zhou, L.-L. Yan, Y.-M. Dong, Z.-L. Bai, X.-Q. Sun, J. Diwu, S. Wang, J.-C. Bünzli and Q.-F. Sun, *Nat. Commun.*, 2018, **9**, 1-10.

16. L. K. Sinclair, J. W. Tester, J. F. H. Thompson and R. V. Fox, *Industrial & Engineering Chemistry Research*, 2019, **58**, 9199-9211.

17. E. A. Mowafy and D. Mohamed, *Sep. Purif. Technol.*, 2014, **128**, 18-24.

18. R. J. Ellis, D. M. Brigham, L. Delmau, A. S. Ivanov, N. J. Williams, M. N. Vo, B. Reinhart, B. A. Moyer and V. S. Bryantsev, *Inorg. Chem.*, 2017, **56**, 1152-1160.

19. X. He, X. Wang, Y. Cui, Z. Su, G. Ye, C. Lu and L. Li, *Journal of Radioanalytical and Nuclear Chemistry*, 2021, 1-8.

20. A. Kovács, C. Apostolidis and O. Walter, *Inorganics*, 2019, **7**, 26.

21. M. R. Foreman, M. J. Hudson, M. G. Drew, C. Hill and C. Madic, *Dalton Trans.*, 2006, 1645-1653.

22. Y. Yang, J. Liu, L. Yang, K. Li, H. Zhang, S. Luo and L. Rao, *Dalton Trans.*, 2015, **44**, 8959-8970.

23. D. M. Whittaker, T. L. Griffiths, M. Helliwell, A. N. Swinburne, L. S. Natrajan, F. W. Lewis, L. M. Harwood, S. A. Parry and C. A. Sharrad, *Inorg. Chem.*, 2013, **52**, 3429-3444.

24. S. Jansone-Popova, A. S. Ivanov, V. S. Bryantsev, F. V. Sloop, R. Custelcean, I. Popovs, M. M. Dekarske and B. A. Moyer, *Inorg. Chem.*, 2017, **56**, 5911-5917.

25. M. R. Healy, A. S. Ivanov, Y. Karslyan, V. S. Bryantsev, B. A. Moyer and S. Jansone-Popova, *Eur. J. Chem.*, 2019, **25**, 6326-6331.

26. P. J. Panak and A. Geist, *Chem. Rev.*, 2013, **113**, 1199-1236.

27. Z. Kolarik, *Chem. Rev.*, 2008, **108**, 4208-4252.

28. T. Liu, K. R. Johnson, S. Jansone-Popova and D.-e. Jiang, *JACS Au*, 2022, **2**, 1428-1434.

29. D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742-754.

30. G. Landrum, *http://www.rdkit.org*, 2006.

31. B. Ramsundar, P. Eastman, P. Walters and V. Pande, *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more*, " O'Reilly Media, Inc.", 2019.

32. D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, 1977, **67**, 4517-4533.

33. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 8026-8037.

34. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort and J. Grobler, *arXiv preprint arXiv:1309.0238*, 2013.

35. D. Stamberga, M. R. Healy, V. S. Bryantsev, C. Albisser, Y. Karslyan, B. Reinhart, A. Paulenova, M. Foster, I. Popovs and K. Lyon, *Inorg. Chem.*, 2020, **59**, 17620-17630.

36. X. He, X. Wang, Y. Cui, Z. Su, G. Ye, C. Lu and L. Li, *J. Radioanal. Nucl. Chem.*, 2021, **329**, 1019-1026.

37. Y. Karslyan, F. V. Sloop, L. H. Delmau, B. A. Moyer, I. Popovs, A. Paulenova and S. Jansone-Popova, *RSC Adv.*, 2019, **9**, 26537-26541.

38. E. Mowafy and H. Aly, *Solvent Extr. Ion Exch.*, 2006, **24**, 677-692.

39. J. Sulakova, R. Paine, M. Chakravarty and K. Nash, *Sep. Sci. Technol.*, 2012, **47**, 2015-2023.

40. M. Simonnet, T. Kobayashi, K. Shimojo, K. Yokoyama and T. Yaita, *Inorg. Chem.*, 2021, **60**, 13409-13418.

41. M. G. B. Drew, C. Hill, M. J. Hudson, P. B. Iveson, C. Madic, L. Vaillant and T. G. A. Youngs, *New J. Chem.*, 2004, **28**, 462-470.

42. R. Meng, L. Xu, X. Yang, M. Sun, C. Xu, N. E. Borisova, X. Zhang, L. Lei and C. Xiao, *Inorg. Chem.*, 2021, **60**, 8754-8764.

43. C. Marie, M. Miguirditchian, D. Guillaneux, J. Bisson, M. Pipelier and D. Dubreuil, *Solvent Extr. Ion Exch.*, 2011, **29**, 292-315.

44. Y. Liu, X. Yang, S. Ding, Z. Wang, L. Zhang, L. Song, Z. Chen and X. Wang, *Inorg. Chem.*, 2018, **57**, 5782-5790.

45. Z. Wang, N. Pu, Y. Tian, C. Xu, F. Wang, Y. Liu, L. Zhang, J. Chen and S. Ding, *Inorg. Chem.*, 2019, **58**, 5457-5467.

46. E. A. Konopkina, P. I. Matveev, P.-W. Huang, A. A. Kirsanova, M. G. Chernysheva, T. B. Sumyanova, K. S. Domnikov, W.-Q. Shi, S. N. Kalmykov and V. G. Petrov, *Dalton Trans.*, 2022, **51**, 11180-11192.

47. P. Lemport, M. Evsiunina, P. Matveev, V. Petrov, A. Pozdeev, E. Khult, Y. V. Nelyubina, K. Isakovskaya, V. Roznyatovsky and I. Gloriozov, *Inorganic Chemistry Frontiers*, 2022, **9**, 4402-4412.

48. X. Yang, S. Wang, L. Xu, Q. Yan, C. Xu, P. Matveev, L. Lei and C. Xiao, *Inorganic Chemistry Frontiers*, 2022, **9**, 4671-4684.

49. L. Xu, Y. Hao, X. Yang, Z. Wang, C. Xu, N. E. Borisova, M. Sun, X. Zhang, L. Lei and C. Xiao, *Chem. Eur. J.*, 2021, **27**, 10717-10730.

50. Y. LeCun, Y. Bengio and G. Hinton, *nature*, 2015, **521**, 436-444.

Chapter 7. All-Carboxylate-Protected Superatomic Silver Nanocluster with an Unprecedented Rhombohedral Ag_8 Core

7.1 Abstract

Our collaborations with Kuan-Guan Liu group from Ningxia University report the first all-carboxylate-protected superatomic silver nanocluster. The $[Ag_8(pfga)_6]^{6-}$ cluster has a rhombohedral Ag_8^{6+} core, with each of its faces protected by one dianionic perfluoroglutarate (pfga) ligand. Electronic-structure analysis from density functional theory confirms the stability of this two-electron cluster due to the shell closing of the superatomic orbital in the $(1S)^2$ configuration and explains the optical absorption of the cluster in the visible region as the transition from 1S to 1P orbital. The $[Ag_8(pfga)_6]^{6-}$ cluster emits bright green-yellow light in THF solution and bright orange light in the solid state. This work opens the door to using the widely available carboxylic acids to synthesize atomically precise Ag clusters of attractive properties.

7.2 Introduction

Atomically precise monolayer-protected coinage metal nanoclusters (CMNCs) have attracted intense interest for their structural diversity and intriguing properties.[1-5] Their formation and stabilization rely on both the electron count and the monolayers (organic/inorganic ligands), which play vital roles in determining their structures and properties.[1-9] The classic theory of hard and soft acids and bases (HSAB) rationalizes the selection of soft-base ligands, such as phosphines,[10, 11] thiolates,[12-14] selenolates,[15-17] and

alkynylides,[18-22] to stabilize the soft-acid coinage metal atoms in forming the CMNCs. Many such Cu/Ag/Au clusters have been characterized by X-ray single-crystal analysis.

In contrast, it is much more challenging to obtain atomically precise CMNCs with the hard-base ligands containing N and O donors, even though those ligands are often used in preparing larger coinage-metal nanoparticles.[23, 24] Some recent successes in synthesizing hard-base-ligand-protected CMNCs suggested promising routes to overcome the challenge. For instance, Wang et al. reported all-nitrogen-donor-protected Ag_{21} and Ag_{22} that showed interesting optical properties.[25] More excitingly, DNA-stabilized (via nitrogen and oxygen donors on the bases) Ag_8 and Ag_{16} nanoclusters with near-IR emitting properties were characterized by X-ray single-crystal analysis.[26, 27] Using inorganic donors, Mizuno and co-workers reported $[SiW_{10}O_{36}]^{8-}$ protected Ag_6 and Ag_{27} clusters,[28, 29] while Zhang et al. obtained two unprecedented Ag_6@Ti_{16}-oxo nanoclusters with Ti-oxo clusters as ligands.[30]

Despite the extensive use of inorganic oxo groups, using organic oxygen-donors such as carboxylates to direct synthesis of atomically precise CMNCs has not been reported. The great availability, affordability, and variety of carboxylic groups would enable great tunability of the CMNCs, if a synthetic route to all-carboxylate-protected CMNCs could be opened. Herein, we report the synthesis and structure determination of the first all-carboxylate-protected superatomic silver nanocluster, hence opening the door to using carboxylate groups to synthesize atomically precise Ag clusters. The key to our success is the use of a flexible dicarboxylic acid (perfluoroglutaric acid) as a multidentate protecting agent, to overcome the challenge of using the monocarboxylate ligands to

construct CMNCs. The double carboxylate groups of each perfluoroglutarate (pfga) provide four oxygen donors for binding with multiple metal atoms. Moreover, when in the gauche form, pfga can bend into a staple shape, which helps lock adjacent metal atoms in place.[31] In addition, the carboxylate O atoms are a strong hydrogen-bonding acceptor that can further help stabilizing metal clusters via supramolecular interactions.[32]

7.3 Results and discussion

7.3.1 Unique structure and stability

Our collaborators prepared the cluster by heating a dimethylformamide solution of perfluoroglutaric acid and $AgNO_3$ under alkaline conditions, yielding a single crystal. The crystal's chemical formula is determined by single-crystal X-ray diffraction as $[(CH_3)_2NH_2]_6[Ag_8(pfga)_6]\cdot H_2O\cdot 6DMF$ (**1**), containing the $[Ag_8(pfga)_6]^{6-}$ center (Figure 7-1a) with a unique rhombohedral Ag_8^{6+} core (Figure 7-1b). The Ag···Ag distances range from 2.768 to 2.868 Å, similar to those in other silver nanoclusters.[27, 33, 34] Compared with the widely reported Ag_6^{4+} nanocluster cores,[28, 33, 35-37] the Ag_8^{6+} core is rare and illuminated here for the first time. The rhombohedral Ag_8 core (D_{3d} point group) is also unprecedented among eight-atom coinage-metal clusters, as the M_8 cores usually take a bicappedoctahedral,[38, 39] cubic,[40] tetracapped tetrahedral,[41, 42] or edge-shared tetrahedron structure.[43]

In the $[Ag_8(pfga)_6]^{6-}$ anionic cluster (Figure 7-1a), each of the six pfga ligands protects the Ag_8^{6+} core in cis-μ_4-η,$^1\eta$,$^1\eta$,$^1\eta^1$ coordination mode; the dihedral angle of the fluoroalkyl backbone (highlighted in black in Figure 7-1a) conforms to the gauche form.[31,

[44] Six pfga ligands cap six diamond-shaped faces of the rhombohedral Ag_8^{6+} core to form the anionic cluster $[Ag_8(pfga)_6]^{6-}$: each pfga's two parallel carboxylate groups coordinate to two parallel edges of each face (Figure 7-1c). The use of the fluorinated pfga ligand is necessary to obtain the Ag_8 cluster; we tried glutaric acid but failed to obtain the desired Ag_8 product. In addition to the unique cis-μ_4-η,$^1\eta^1$,η^1,η^1 coordination mode, fluorine groups on pfga withdraw electrons from the negatively charged cluster and more broadly distribute the negative charges, thereby stabilizing the whole cluster. The coordination mode and the electron-withdrawing fluorine groups enabled by pfga are also key to achieving the unprecedented rhombohedral Ag_8 core.

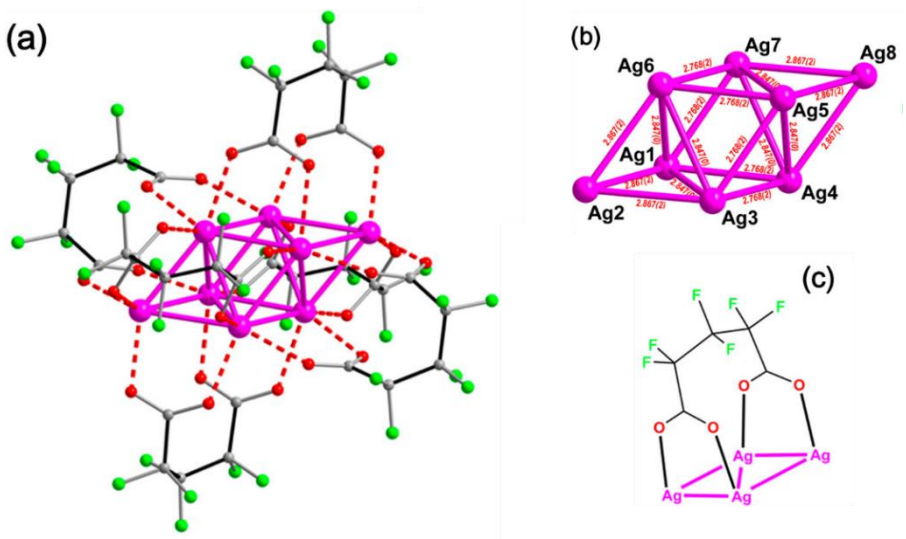


Figure 7-1. (a) Structure of the $[Ag_8(pfga)_6]^{6-}$ cluster in the crystal (**1**). Inset: Microscope photograph of the rhombohedral crystals of $[(CH_3)_2NH_2]_6[Ag_8(pfga)_6]\cdot H_2O\cdot 6DMF$ **1**. Dashed lines represent silver−ligand coordination bonds (via Ag···O interactions). (b) Rhombohedral Ag_8^{6+} core of the cluster (Ag−Ag distances in Å). (c) Binding mode of pfga on the core surface. Ag, purple; O, red; C,gray; F, green.

7.3.2 Optical properties

We next examine electronic structure and optical properties of the cluster. As shown in Figure 7-2, the UV−vis absorption spectrum of **1** in the THF solution exhibits two prominent peaks at 3.28 eV (α; 378 nm, ε_{max} = 5.5 × 10^4 L·mol^{-1} cm^{-1}) and 4.71 eV (β; 263 nm, ε_{max} = 9.2 × 10^4 L·mol^{-1} cm^{-1}), which were well reproduced in the simulated spectrum from time-dependent density functional theory showing two main peaks at 3.0 and 4.4 eV. Orbital analysis showed that the α peak corresponds to the HOMO−LUMO transition, while the β peak includes multiple transitions but mostly of HOMO to LUMO+1 transition (Figure 7-2b). $[Ag_8(pfga)_6]^{6-}$ has two valence electrons according to the superatom model: $n^* = N_{Ag} - 2 \times N_{pfga} - z = 8 - 2 \times 6 - (-6) = 2$.[45] Therefore, the superatomic electronic structure is in $(1S)^2$ configuration. The three empty 1P orbitals split (Figure 7-2b) into a nondegenerate $1P_z$ (along the long axis of the rhombohedron; LUMO) and the higher-energy double-degenerate $1P_{x,y}$ (LUMO+1). Indeed, one can see the P-type orbital character of the LUMO (Figure 7-2b), while the HOMO is not of regular S-type due to stronger hybridization of Ag states with the ligand states and the cluster's nonspherical shape. The pfga ligands influence the superatomic electron structure of the Ag_8 cluster as thiolate ligands do to the Au and Ag clusters: by localizing electrons from the metal center. Such superatomic state enhances the interaction between a conventional hard base (carboxylate) and soft acid (Ag); the fluorine groups on pfga further polarizes the carboxylate groups to have more covalent interaction with Ag (Figure 7-3).

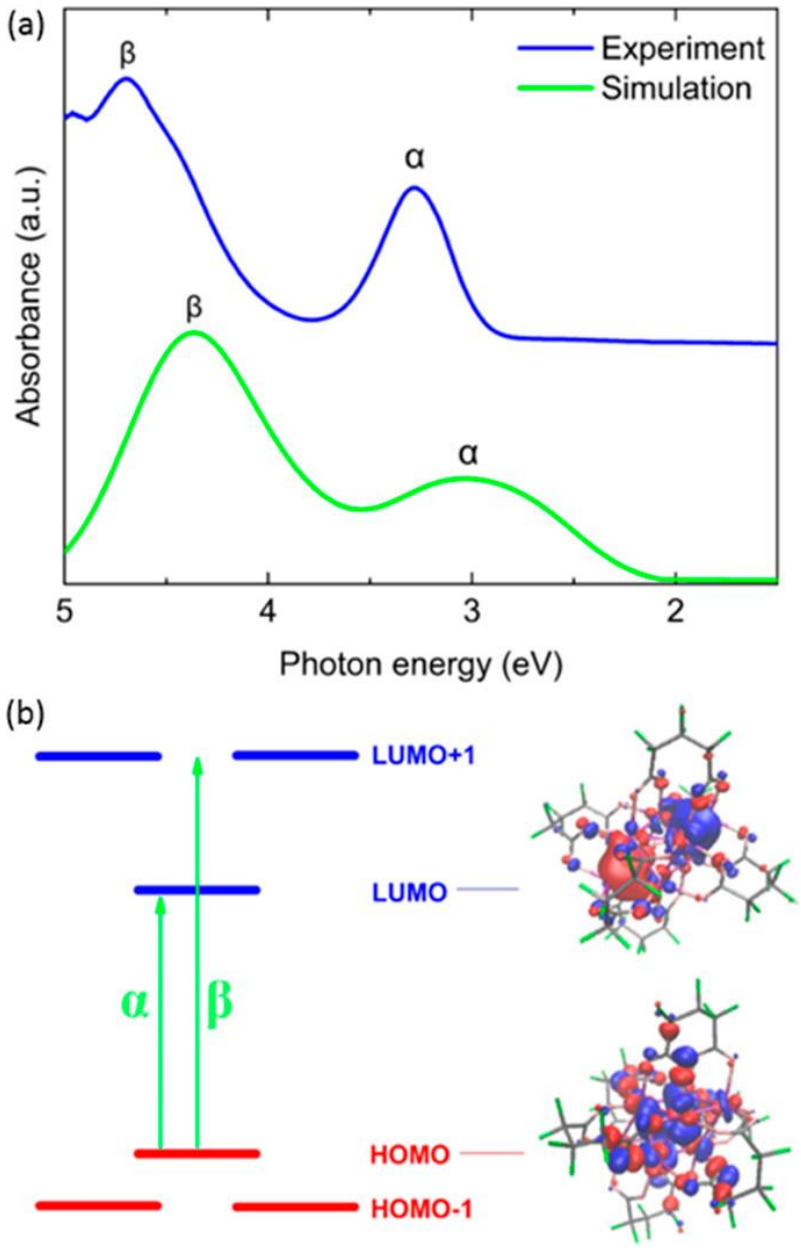


Figure 7-2. (a) Experimental UV−vis absorption spectrum (blue) of **1** in THF and simulated spectrum of $[Ag_8(pfga)_6]^{6-}$ from time-dependent density-functional theory (green). (b) Orbital diagram and frontier orbitals of $[Ag_8(pfga)_6]^{6-}$.

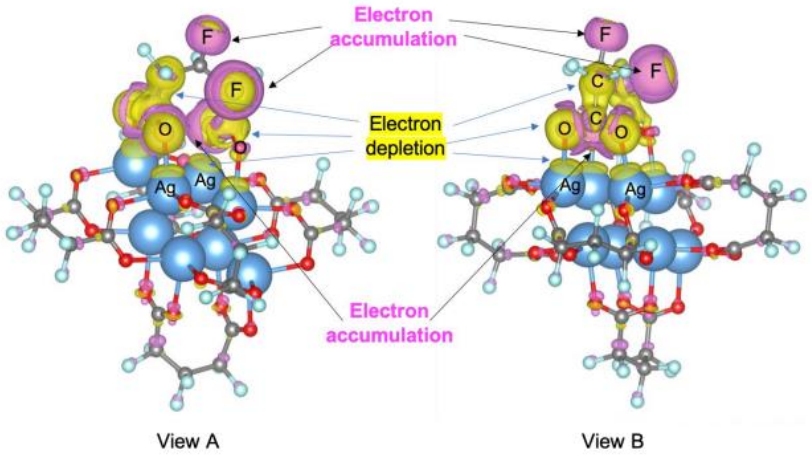


Figure 7-3. Electron-density difference (Δρ) isosurface for one pfga ligand binding on the $[Ag_8(pfga)_6]^{6-}$ cluster surface (Δρ = ρ$[Ag_8(pfga)_6]^{6-}$ - ρ$[Ag_8(pfga)_5]^{4-}$ - ρ$[pfga]^{2-}$). The magenta and yellow isosurfaces correspond to the regions of electron accumulation and depletion, respectively. Isovalue at 0.004 e/Å^3.

The successful use of carboxylic acids to synthesize an atomically precise silver nanocluster demonstrated here has blazed a path that could lead to a great vista of carboxylate-protected coinage metal nanoclusters. We envision that many other types of carboxylic acids including mono, nonfluorinated carboxylic acids and amino acids could be used and that clusters of higher electron counts such as 8 and 58 could be obtained. The high coordination number of carboxylate ligands around a surface Ag may be difficult to achieve when the cluster size becomes larger, but it may be overcome via strategies including use of a co-ligand (such as halides and phosphines), bulky carboxylates, or a core−shell construction (complexes of Ag^+ and carboxylates protecting zerovalent Ag core).

The carboxylate-protected Ag clusters could have interesting chemical/electronic/optical properties for potential applications. One advantage of this cluster is that the carboxylate ligands can be removed rather easily at ambient conditions, due to the weaker Ag−COO bond than Ag-thiolate or Au-thiolate bond. Hence, it can be used as a precursor to prepare supported atomically precise Ag nanocluster catalysts or to produce assemblies of ultrathin Ag nanowires for electronics applications.[46] Moreover, carboxylate groups could provide a key link to connect metal nanoclusters and metal−organic frameworks using organic carboxylate linkers, which might lead to interesting composite materials.

7.4 Summary and conclusions

In sum, the first example of an all-carboxylate-protected superatomic silver nanocluster, $[Ag_8(pfga)_6]^{6-}$, featuring a rhombohedral Ag_8^{6+} core and perfluoroglutarate (pfga) protecting ligands, was demonstrated. Each of the six diamond-shaped faces of Ag_8^{6+} core is "stapled" by the two parallel carboxylate groups of a pfga ligand. Optical absorption of the cluster and its electronic structure could be well understood from density functional theory and the superatomic model. The cluster is a strong luminophore both in solution and in the solid state at room temperature. This work opens up a new avenue of research in atomically precise nanochemistry that may lead to many new carboxylate-protected coinage metal nanoclusters being discovered.

References

1. R. Jin, C. Zeng, M. Zhou and Y. Chen, *Chem. Rev.*, 2016, **116**, 10346-10413.

2. J. Yan, B. K. Teo and N. Zheng, *Acc. Chem. Res.*, 2018, **51**, 3084-3093.

3. I. Chakraborty and T. Pradeep, *Chem. Rev.*, 2017, **117**, 8208-8271.

4. B. Bhattarai, Y. Zaker, A. Atnagulov, B. Yoon, U. Landman and T. P. Bigioni, *Acc. Chem. Res.*, 2018, **51**, 3104-3113.

5. S. Sharma, K. K. Chakrahari, J.-Y. Saillard and C. Liu, *Acc. Chem. Res.*, 2018, **51**, 2475-2483.

6. C. P. Joshi, M. S. Bootharaju and O. M. Bakr, *The journal of physical chemistry letters*, 2015, **6**, 3023-3035.

7. X. Kang and M. Zhu, *Chem. Soc. Rev.*, 2019, **48**, 2422-2457.

8. P. Liu, R. Qin, G. Fu and N. Zheng, *J. Am. Chem. Soc.*, 2017, **139**, 2122-2131.

9. G.-G. Luo, Q.-L. Guo, Z. Wang, C.-F. Sun, J.-Q. Lin and D. Sun, *Dalton Trans.*, 2020, **49**, 5406-5415.

10. M. S. Bootharaju, R. Dey, L. E. Gevers, M. N. Hedhili, J.-M. Basset and O. M. Bakr, *J. Am. Chem. Soc.*, 2016, **138**, 13770-13773.

11. J. Chen, Q.-F. Zhang, T. A. Bonaccorso, P. G. Williard and L.-S. Wang, *J. Am. Chem. Soc.*, 2014, **136**, 92-95.

12. Y. Negishi, K. Nobusada and T. Tsukuda, *J. Am. Chem. Soc.*, 2005, **127**, 5261-5270.

13. H. Lee, S. M. Dellatore, W. M. Miller and P. B. Messersmith, *science*, 2007, **318**, 426-430.

14. Q. Yao, T. Chen, X. Yuan and J. Xie, *Acc. Chem. Res.*, 2018, **51**, 1338-1348.

15. W.-T. Chang, P.-Y. Lee, J.-H. Liao, K. K. Chakrahari, S. Kahlal, Y.-C. Liu, M.-H. Chiang, J.-Y. Saillard and C. W. Liu, *Angew. Chem. Int. Ed.*, 2017, **56**, 10178-10182.

16. X. Kang and M. Zhu, *Small*, 2019, **15**, 1902703.

17. C. A. Hosier and C. J. Ackerson, *J. Am. Chem. Soc.*, 2018, **141**, 309-314.

18. P. Maity, H. Tsunoyama, M. Yamauchi, S. Xie and T. Tsukuda, *J. Am. Chem. Soc.*, 2011, **133**, 20123-20125.

19. P. Maity, S. Takano, S. Yamazoe, T. Wakabayashi and T. Tsukuda, *J. Am. Chem. Soc.*, 2013, **135**, 9450-9457.

20. X.-K. Wan, Q. Tang, S.-F. Yuan, D.-e. Jiang and Q.-M. Wang, *J. Am. Chem. Soc.*, 2015, **137**, 652-655.

21. Y. Wang, H. Su, C. Xu, G. Li, L. Gell, S. Lin, Z. Tang, H. Häkkinen and N. Zheng, *J. Am. Chem. Soc.*, 2015, **137**, 4324-4327.

22. Z. Lei, J.-J. Li, X.-K. Wan, W.-H. Zhang and Q.-M. Wang, *Angew. Chem. Int. Ed.*, 2018, **57**, 8639-8643.

23. T.-Y. Dong, W.-T. Chen, C.-W. Wang, C.-P. Chen, C.-N. Chen, M.-C. Lin, J.-M. Song, I.-G. Chen and T.-H. Kao, *Phys. Chem. Chem. Phys.*, 2009, **11**, 6269-6275.

24. J.-M. Song, T.-Y. Pai, K.-H. Hsieh, M.-Y. Lai, C.-N. Cheng, S.-Y. Liang, H.-Y. Lee and L.-T. Chen, *RSC Adv.*, 2016, **6**, 97449-97454.

25. S.-F. Yuan, Z.-J. Guan, W.-D. Liu and Q.-M. Wang, *Nat. Commun.*, 2019, **10**, 1-7.

26. D. J. Huard, A. Demissie, D. Kim, D. Lewis, R. M. Dickson, J. T. Petty and R. L. Lieberman, *J. Am. Chem. Soc.*, 2018, **141**, 11465-11470.

27. C. Cerretani, H. Kanazawa, T. Vosch and J. Kondo, *Angew. Chem. Int. Ed.*, 2019, **58**, 17153-17157.

28. Y. Kikukawa, Y. Kuroda, K. Suzuki, M. Hibino, K. Yamaguchi and N. Mizuno, *Chem. Commun.*, 2013, **49**, 376-378.

29. K. Yonesato, H. Ito, H. Itakura, D. Yokogawa, T. Kikuchi, N. Mizuno, K. Yamaguchi and K. Suzuki, *J. Am. Chem. Soc.*, 2019, **141**, 19550-19554.

30. S. Chen, W. H. Fang, L. Zhang and J. Zhang, *Angew. Chem. Int. Ed.*, 2018, **57**, 11252-11256.

31. E. Lee, Y. Kim and D.-Y. Jung, *Inorg. Chem.*, 2002, **41**, 501-506.

32. J.-Z. Li, F. Bigdeli, X.-M. Gao, R. Wang, X.-W. Wei, X.-W. Yan, M.-L. Hu, K.-G. Liu and A. Morsali, *Inorg. Chem.*, 2019, **58**, 5397-5400.

33. Z. Wang, H.-F. Su, M. Kurmoo, C.-H. Tung, D. Sun and L.-S. Zheng, *Nat. Commun.*, 2018, **9**, 1-8.

34. J. Yan, J. Zhang, X. Chen, S. Malola, B. Zhou, E. Selenius, X. Zhang, P. Yuan, G. Deng and K. Liu, *National Science Review*, 2018, **5**, 694-702.

35. Z. Wang, Q.-P. Qu, H.-F. Su, P. Huang, R. K. Gupta, Q.-Y. Liu, C.-H. Tung, D. Sun and L.-S. Zheng, *Science China Chemistry*, 2020, **63**, 16-20.

36. Z. Wang, F.-L. Yang, Y. Yang, Q.-Y. Liu and D. Sun, *Chem. Commun.*, 2019, **55**, 10296-10299.

37. H. Yang, J. Lei, B. Wu, Y. Wang, M. Zhou, A. Xia, L. Zheng and N. Zheng, *Chem. Commun.*, 2013, **49**, 300-302.

38. K. Nakamae, M. Tanaka, B. Kure, T. Nakajima, Y. Ura and T. Tanase, *Chem. Eur. J.*, 2017, **23**, 9457-9461.

39. J.-H. Jia and Q.-M. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 16634-16635.

40. C. Latouche, S. Kahlal, E. Furet, P.-K. Liao, Y.-R. Lin, C.-S. Fang, J. Cuny, C. Liu and J.-Y. Saillard, *Inorg. Chem.*, 2013, **52**, 7752-7765.

41. P.-K. Liao, K.-G. Liu, C.-S. Fang, C. Liu, J. P. Fackler Jr and Y.-Y. Wu, *Inorg. Chem.*, 2011, **50**, 8410-8417.

42. C. Liu, B. Sarkar, Y.-J. Huang, P.-K. Liao, J.-C. Wang, J.-Y. Saillard and S. Kahlal, *J. Am. Chem. Soc.*, 2009, **131**, 11222-11233.

43. Y. Kamei, Y. Shichibu and K. Konishi, *Angew. Chem. Int. Ed.*, 2011, **50**, 7442-7445.

44. P. Tarakeshwar and S. Manogaran, *Journal of Molecular Structure: THEOCHEM*, 1996, **362**, 77-99.

45. M. Walter, J. Akola, O. Lopez-Acevedo, P. D. Jadzinsky, G. Calero, C. J. Ackerson, R. L. Whetten, H. Grönbeck and H. Häkkinen, *Proceedings of the National Academy of Sciences*, 2008, **105**, 9157-9162.

46. A. Som, I. Chakraborty, T. A. Maark, S. Bhat and T. Pradeep, *Adv. Mater.*, 2016, **28**, 2827-2833.

Chapter 8. Understanding the Interaction Between Carboxylates and Coinage Metals from First Principles

8.1 Abstract

Carboxylate groups have recently been explored as a new type of ligand to protect superatomic copper and silver nanoclusters, but little is known of the interfacial structure and bonding. Here, we employ density functional theory to investigate the interfaces of a model carboxylate group, CH_3COO, on the coinage metal surfaces and clusters. We found that μ_2-CH_3COO is the most preferred binding mode on the three M(111) surfaces (M = Cu, Ag, and Au), while μ_3-CH_3COO is also stable on Cu(111) and Ag(111). The saturation coverage was found to be about seven CH_3COO groups per nm^2 for all surfaces. CH_3COO has the strongest binding on Cu and weakest on Au. Moving from the flat surfaces to the icosahedral M_{13} clusters, we found that the eight-electron superatomic $[M_{13}(CH_3COO)_6]^-$ nanoclusters also prefer the μ_2-CH_3COO mode on the surface. The icosahedral kernel in $[Cu_{13}(CH_3COO)_6]^-$ and $[Ag_{13}(CH_3COO)_6]^-$ was well maintained after geometry optimization, but a larger deformation was found in $[Au_{13}(CH_3COO)_6]^-$. Given the broad availability and variety of carboxylic acids including amino acids, our work suggests that carboxylate groups could be the next generation ligands to further expand the universe of atomically precise metal clusters, especially for Cu and Ag.

8.2 Introduction

Ligand-protected coinage metal nanoclusters find applications in microelectronics,[1] electrochemistry,[2] energy conversion,[3] pharmaceutical chemistry,[4] sensing,[5, 6] and

catalysis.[7, 8] Great efforts of synthesis and characterization have been devoted to establishing their structure-property relationships. Generations of ligands from phosphines[9] and thiolates[10-13] to alkynyls[14] and carbenes[15, 16] have been used in protecting coinage metal nanoclusters. Meanwhile, first principles density functional theory (DFT) has provided insights into many fundamental questions such as structure, interfaces, formation, and reactions.[17-19]

The confluence of efforts in the past decade has led to the new field of atomically precise nanochemistry, epitomized by ligand-protected coinage metal clusters. One important and common theme in this new field is the interfacial structure and bonding of ligands on the cluster surface and the impact of the curvature and cluster size on the interfacial motifs.[20] Case in point is the famous staple motif of RS−Au−SR on the gold clusters[10, 21] and surfaces.[22] The analogous RCC-Au-CCR motif was also later found,[23, 24] opening up a new avenue of research that leads to many new structural, optical, and catalytic insights.[25, 26] Hence, exploring new ligands is a long-lasting thrust in the field of atomically precise nanochemistry.

Although carboxylates as a hard base were less explored for protecting soft acid ions of coinage metals, self-assembly monolayers (SAMs) of n-alkanoic carboxylates on copper and silver surfaces were investigated in 1990s.[27-29] Recently, researchers were also interested in carboxylate chemistry on gold nanoparticles[30] and gold surfaces.[31] Lately, a new all-carboxylate-protected superatomic silver nanocluster, $[Ag_8(pfga)_6]^{6-}$, was discovered[32] that has generated great interest in using carboxylic acids including amino acids for synthesizing coinage metal nanoclusters.[33] The pfga (perfluoroglutarate) ligand

is a special chelate ligand with fluorination in the middle and two carboxylate groups at the ends. To fully take advantage of the great variety and availability of carboxylate ligands for protecting coinage metals, one wonders what the expected binding mode should be for a typical or more general carboxylate group at the interface with a coinage-metal cluster or surface.

The recent discovery of carboxylate-protected Ag nanoclusters and the previous work of carboxylates on coinage metal nanoparticles and surfaces prompted us to examine in detail the structure and energetics of the carboxylate-metal interface for both a flat surface and a nanocluster for a general carboxylate group. To this end, herein we employ first-principles density functional theory (DFT) to first explore the interfacial structures of a model carboxylate group (namely, CH_3COO group) on the (111) surfaces of Cu, Ag, and Au. Then we will probe the acetate group on icosahedral M_{13} clusters of Cu, Ag, and Au from the superatomic perspective.[34] Below we first introduce the computational method.

8.3 Computational Method

Structure and energetics of the acetate group on the (111) surfaces of Cu, Ag, and Au were studied by using the Vienna ab initio simulation package (VASP).[35] The ion−electron interaction is described with the projector augmented wave (PAW) method.[36] Electron exchange-correlation is represented by the functional of Perdew, Burke, and Ernzerhof (PBE) of generalized gradient approximation (GGA).[37] Cutoff energy of 450 eV was used for the plane-wave basis set. The convergence criteria were 10^{-4} eV in energy and 0.02 eV/Å in force for all optimizations. As shown in Table 7-1, the optimized bulk lattice

parameters of fcc Cu, Ag, and Au show good agreement with experiment. Then, the M (111) (M=Au, Ag, Cu) surface slabs were modeled in a rectangular $3 \times 2\sqrt{3}$ supercell with atomic layers (16 atoms per layer) and their parameters were summarized in Table I. The thickness of the vacuum layer between slab images was set to be 17 Å. The bottom two layers of the slab were fixed at the optimized bulk positions during structural relaxation. The Brillouin zone of the surface supercell was sampled by a Monkhorst-Pack k-point mesh of 4×4×1 grid. Bader charge analysis was done with the implementation of Henkelman et al.[38]

Table 8-1. Bulk lattice parameters of the fcc metals and the surface lattice parameters of the rectangular $3 \times 2\sqrt{3}$ supercell of their (111) surfaces.

Lattice parameters	Cu	Ag	Au
Exp. bulk (Å)	3.62	4.16	4.17
DFT bulk (Å)	3.57	4.08	4.12
DFT surface (Å)	10.27×8.89	11.81×10.23	11.79×10.27

Structural optimizations of the acetate-protected M_{13} nanoclusters were carried out via the quantum chemistry program Turbomole V6.5.[39] The TPSS (Tao, Perdew, Staroverov, and Scuseria) functional[40] was used for electron exchange and correlation, with the def2-TZVP orbital and auxiliary basis sets. Effective core potentials which have 19 valence electrons were used for Ag and Au.[41] The convergence criteria were 3×10^{-5} eV for energy and 5×10^{-3} eV/Å for force.

The van der Waals interaction was included via the DFT-D3 approach with zero damping:[42] for VASP, the coefficients used for the PBE functional are $s_6 = 1.0$, $s_{r,6} = 1.217$,

and s_8 = 0.722; for Turbomole, the coefficients used for the TPSS functional are s_6 = 1.0, $s_{r,6}$ = 1.166, and s_8 = 1.105.

8.4 Results and discussion

8.4.1 Adsorption of one CH_3COO group on the M(111) Surface

The goal of the present work is to computationally assess the preferred binding mode and its associated binding energy of the acetate group on the coinage-metal surfaces and clusters. We started with a low-coverage scenario where there is only one CH_3COO group on the surface supercell. We first computed the binding energy, defined as the energy needed to dissociate the CH_3COO group as a radical into the gas phase away from the surface; we used this energy reference to avoid the complication if one were to study desorption of a charged group from a surface with periodic boundary conditions. As shown in Table 8-2 and Figure 8-1, the μ_2 mode of the carboxylate O atoms anchoring on two neighboring metal sites is the preferred mode on all three surfaces. The μ_3 motif on a surface M_3 triangle (where one O atom of the COO moiety is at the bridge site of two neighboring metal atoms) is slightly higher in energy than μ_2 but is also a local minimum on Cu(111) and Ag(111). The μ_3 motif is not stable on Au(111) and relaxes to the μ_2 mode. Comparing just the common μ_2 mode on the three surfaces, one can see that the CH_3COO group binds to Cu(111) the strongest, followed by Ag(111) and then Au(111). The stronger CH_3COO-Cu(111) interaction is also reflected in the shorter Cu-O distance (2.00 Å; Figure 8-2a). Interestingly, the Ag-O and Au-O distances (Figure 8-2c,e) are very close, but the CH_3COO-Ag(111) interaction is 0.46 eV stronger than CH_3COO-Au(111).

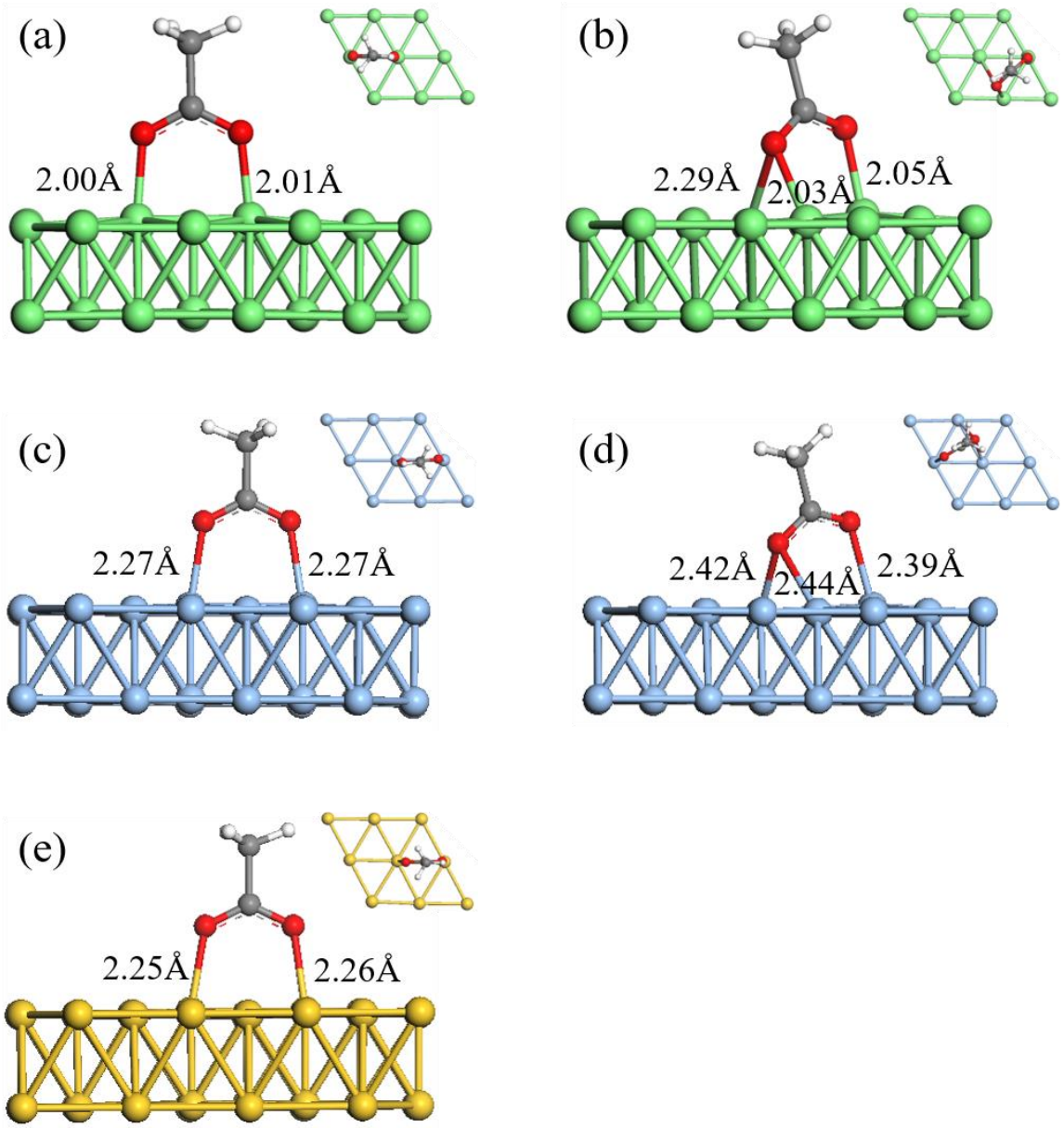


Figure 8-1. Bonding geometries of μ_2-CH_3COO (left panel) and μ_3-CH_3COO (right panel) on the (111) surfaces: (a), (b) Cu(111); (c), (d) Ag(111); (e) Au(111). Color code: H, white; C, grey; O, red; Cu, green; Ag, blue; Au, yellow.

Table 8-2. Binding energies (in eV) of a single CH_3COO group, defined as the energy needed to dissociate the CH_3COO group as a radical into the gas phase away from the surface, on the supercells of the (111) surfaces of Cu, Ag, and Au (see **Table 8-1.** for the supercell dimensions).

Binding mode	Cu	Ag	Au
μ_2	2.88	2.50	2.04
μ_3	2.76	2.44	-

To further shed light on the interfacial bonding, we analyzed the charge transfer between CH_3COO and the metal surfaces. First, we used Bader charge analysis and obtained the partial atomic charges on CH_3COO. After summing up the atomic charges, we obtained the molecular charges of μ_2-CH_3COO and μ_3-CH_3COO on the three surfaces. As one can see from Table 8-3, the molecular charge correlates with the binding energy: the stronger the bonding, the more negative the CH_3COO group. Using CH_3COO-Ag(111) as an example, the charge-density-difference plot (Figure 8-2) clearly shows that electrons move from the metal surface and the C-O bonds (yellow) to the O-Ag bonds and O atoms (magenta). In the case of μ_2-CH_3COO, the C-C bond also accepts some transferred electron (Figure 8-2a).

Table 8-3. Molecular charge (in e) of the CH_3COO group on the (111) surfaces of Cu, Ag, and Au for the two different binding modes.

Binding mode	Cu	Ag	Au
μ_2	-0.65	-0.65	-0.50
μ_3	-0.63	-0.62	

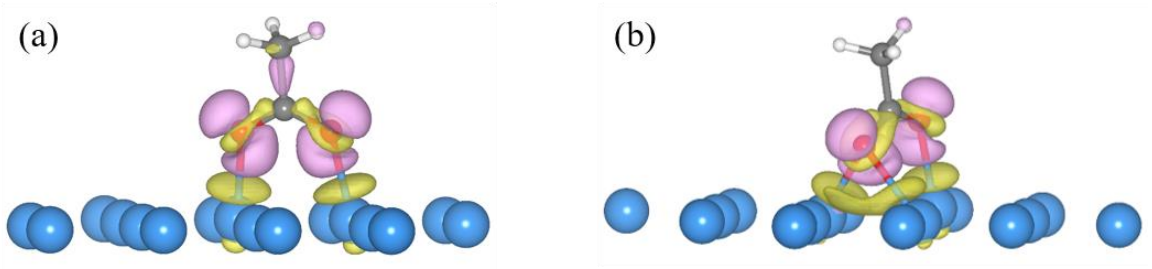


Figure 8-2. Charge density difference for (a) μ_2-CH_3COO and (b) μ_3-CH_3COO on Ag(111) (Ag atoms in blue). Magenta (electron accumulation) and yellow (electron depletion) isosurfaces are at contour levels of 0.003 e/Å^3.

The binding-energy trend can be further explained by the soft-hard acid base theory from a molecular perspective as well as by the electronegativity from the ligand-surface interaction perspective. Carboxylate is a hard base so has the weakest interaction with the softest acid of the three which is gold.[43] On the surface, the carboxylate group will gain some electron from the surface (Table 8-3 and Figure 8-2); Au has the greatest electronegativity of the three and the least tendency to give away electrons to the carboxylate group, hence the weakest interaction. Unfortunately, we have not found any previous experimental work comparing all three coinage metals for the same carboxylate ligand(s) in terms of the binding strength. One difficulty we envision is the decomposition of the carboxylate group during the desorption process. In fact, a recent experimental study found that the carboxylate groups from adsorption of pyridine dicarboxylic acid on Cu(111) undergo decarboxylation during desorption with a measured activation energy of ~ 1.9 eV.[44] This is consistent with our finding of ~ 2.8 eV binding energy of a carboxylate group on Cu(111); in other words, the decarboxylation reaction tends to happen already before the desorption of the complete ligand.

8.4.2 Change of the binding energies with the CH_3COO coverage on the M(111) Surfaces

Both SAMs and ligand-protected metal nanoclusters usually have a complete layer of ligands covering the metal surfaces. Therefore, we next examine how the binding energies and interfacial motifs vary with the CH_3COO coverage. We computed the differential binding energy as the energy needed to desorb just the newly adsorbed CH_3COO group off the surface. As shown in Figure 8-3, the binding energies of CH_3COO groups in the coverage of 1 nm^{-2} to 5 nm^{-2} are rather constant for each metal surface. The saturation coverage is similar for the three surfaces, at ~ 6.6 nm^{-2}. Beyond this coverage, the binding energies turn negative, meaning that desorption is spontaneous and downhill in energy.

The adsorption configurations of CH_3COO groups with increasing coverage on Cu(111) are shown in Figure 8-4. One can see that the μ_3 motif starts to appear when the coverage reaches about half of the saturation coverage (Figure 8-4c). At the saturation coverage, there is one μ_3 and five μ_2 CH_3COO groups. But due to steric effect, the five μ_2 CH_3COO groups are slightly tilted (Figure 8-4e).

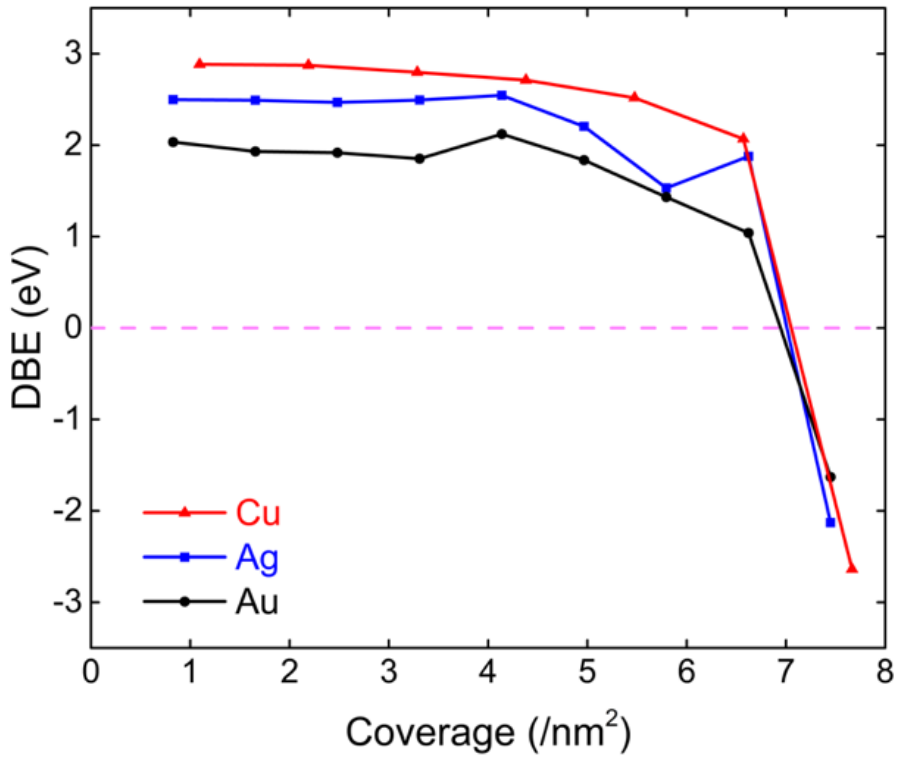


Figure 8-3. Differential binding energy (DBE) versus the coverage of CH_3COO on the (111) surfaces of Cu, Ag, and Au. DBE is defined as the energy needed to dissociate just one CH_3COO group as a radical into the gas phase away from the surface at the specific coverage.

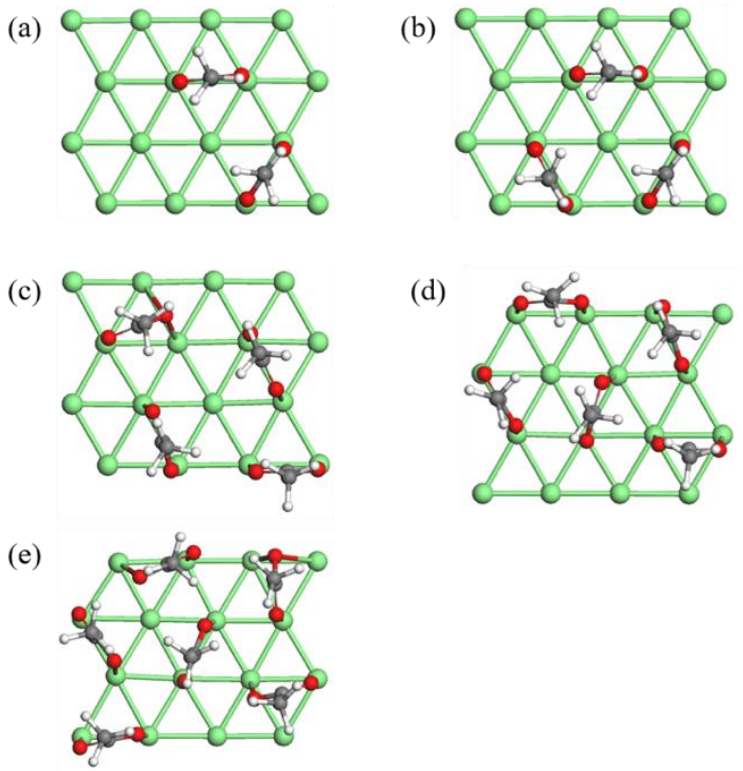


Figure 8-4. Adsorption configurations of CH_3COO with increasing coverage on Cu(111): (a) 2; (b) 3; (c) 4; (d) 5; (e) 6 CH_3COO groups per surface supercell.

The structures of CH_3COO groups on Ag(111) are shown in Figure 8-5. Interestingly, the μ_3 motifs become more frequent especially in the intermediate coverages (Figure 8-5c,d). One reason for this more frequent appearance of the μ_3 motifs is that the μ_2 and μ_3 motifs are much closer in energy on Ag(111) than on Cu(111) (Table 8-2). At the saturation coverage, however, all the carboxylate group are in the μ_2 mode on Ag(111); this is because now all the surface Ag atoms can be utilized to anchor the carboxylate group groups (Figure 8-5g). In the case of Au(111), the picture is simpler, because only μ_2 motif is stable on it, even at high coverages; in the full coverage (Figure 8-6), all surface Au atoms are bonded to carboxylates.

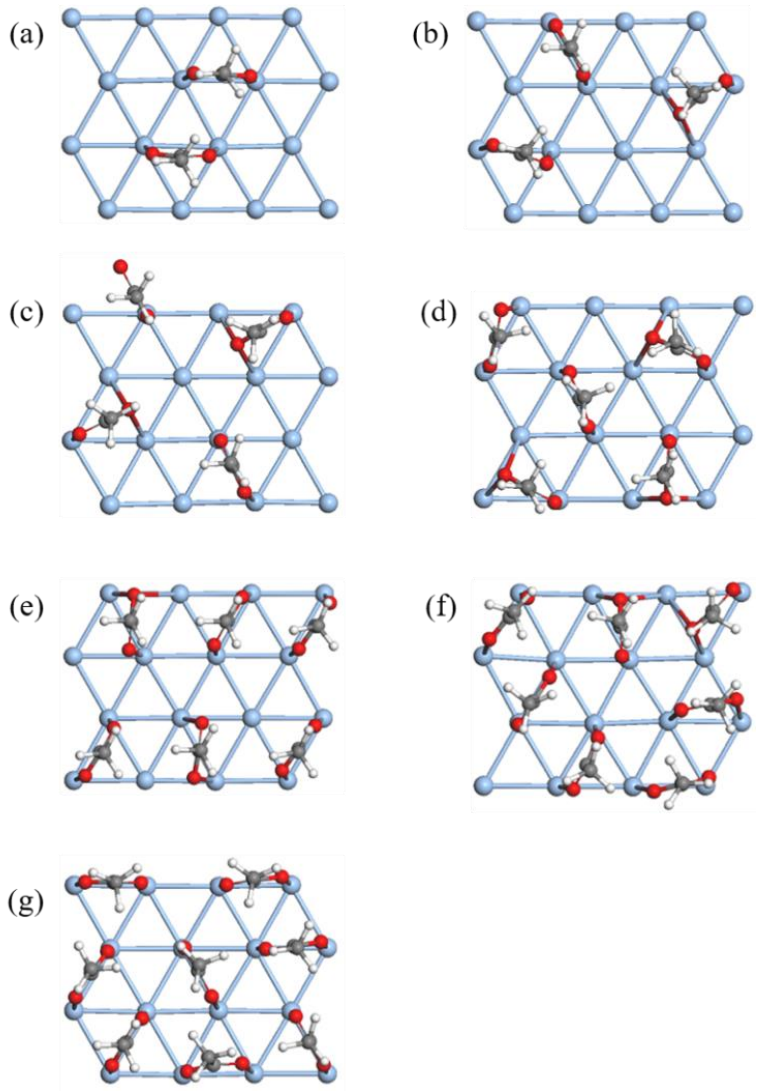


Figure 8-5. Adsorption configurations of CH_3COO with increasing coverage on Ag(111): (a) 2; (b) 3; (c) 4; (d) 5; (e) 6; (f) 7; (g) 8 CH_3COO groups per surface supercell.

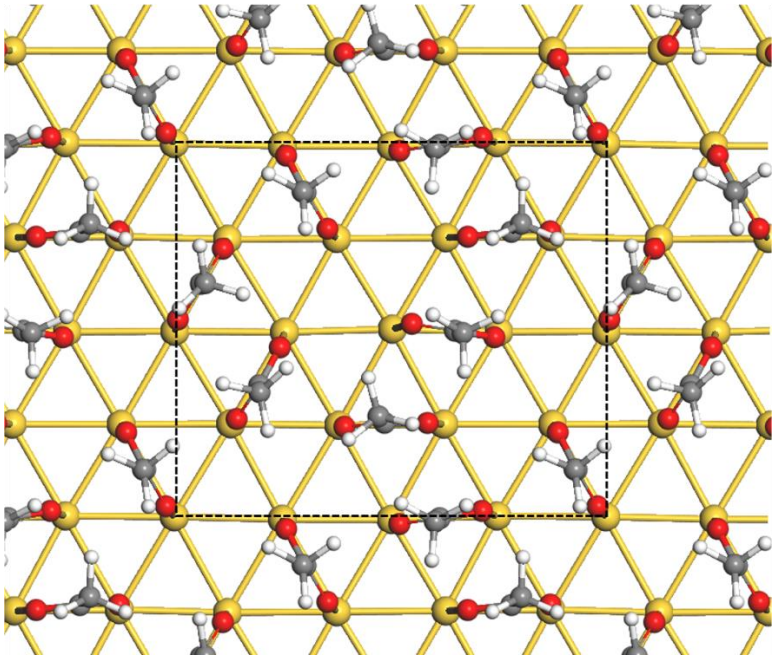


Figure 8-6. Adsorption configuration of full coverage of CH_3COO groups on Au(111). The dashed rectangle denotes the surface supercell.

8.4.3. Carboxylate-protected superatomic M_{13} nanoclusters

The interfacial structure and energetics of CH_3COO on M(111) surfaces (M=Cu, Ag, Au) explored above now provide us a basis to further investigate the chemical bonding of CH_3COO ligands on nanoclusters. The centered icosahedral M_{13} was chosen as the kernel because of its common appearance in ligand-protected metal clusters. Since there are 12 metal atoms at the surface of the icosahedral M_{13} core, six CH_3COO groups are needed to fully protect the surface metal atoms in the μ_2 mode. Then, according to the superatom model,[34] an anionic cluster of the composition of $[M_{13}(CH_3COO)_6]^-$ would be an eight-electron superatom. To test this design idea, we explored a few different arrangements of the six ligands on each M_{13} core and the most stable structures are shown

in Figure 8-7. We found that the μ_2 mode is preferred on the icosahedral M_{13} kernel of Cu and Ag; we tested cases with mixed μ_2 and μ_3 CH_3COO groups on Cu_{13} and Ag_{13} initially and found that μ_3 CH_3COO group changed to μ_2 after relaxation. The optimized $[Cu_{13}(CH_3COO)_6]^-$ and $[Ag_{13}(CH_3COO)_6]^-$ are similar in structure and their icosahedral cores are well maintained after structural relaxation. Hence, both $[Cu_{13}(RCOO)_6]^-$ and $[Ag_{13}(RCOO)_6]^-$ could be viable targets for synthesis. In contrast, the icosahedral shell of $[Au_{13}(CH_3COO)_6]^-$ opened up after geometry optimization, indicating that such a geometry is unstable.

The computed HOMO-LUMO gap of 1.06 eV for $[Cu_{13}(CH_3COO)_6]^-$ and 1.32 eV for $[Ag_{13}(CH_3COO)_6]^-$ is similar to other eight-electron superatoms in the $(1S)^2(1P)^6$ configuration.[45, 46] Indeed, the double degenerate HOMOs show the P character (the three 1P orbitals split into a nondegenerate HOMO-1 and double-degenerate HOMO), while the triple degenerate LUMOs show clear D character (Figure 8-8).

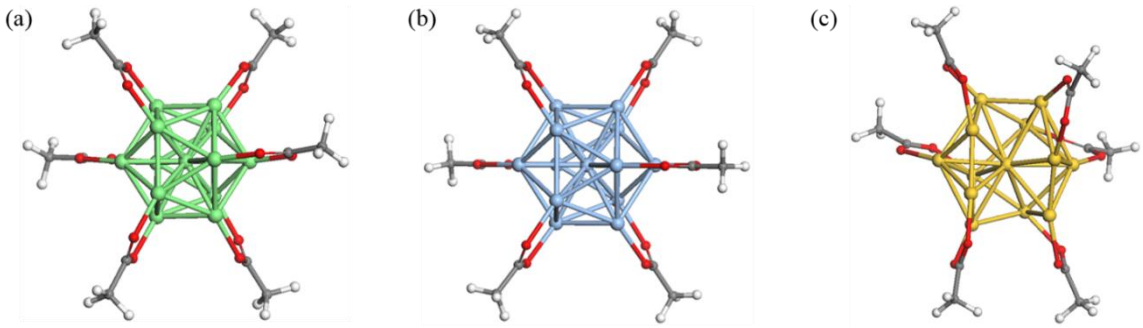


Figure 8-7. DFT-optimized structures of all-carboxylate-protected M_{13} nanoclusters: (a) $[Cu_{13}(CH_3COO)_6]^-$; (b) $[Ag_{13}(CH_3COO)_6]^-$; (c) $[Au_{13}(CH_3COO)_6]^-$.

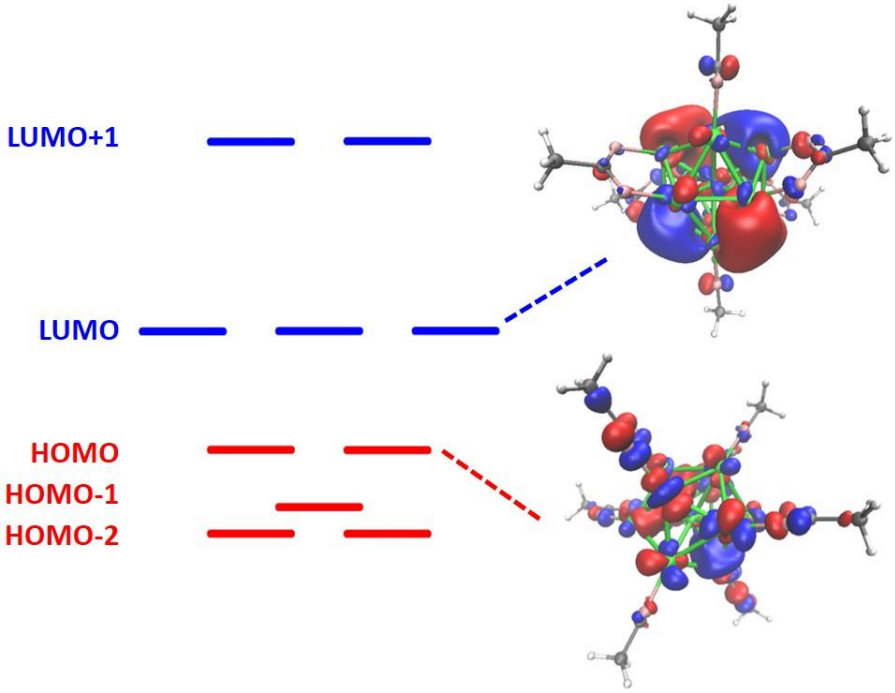


Figure 8-8. Orbital energy diagram and frontier orbitals of $[Cu_{13}(CH_3COO)_6]^-$.

In the structures of $[Cu_{13}(CH_3COO)_6]^-$ and $[Ag_{13}(CH_3COO)_6]^-$ in Figure 8-7, each surface metal atom is coordinated by one O atom from a CH_3COO group. In general, the coordination number of O atoms around a surface Ag or Cu atom depends on both the surface coverage of ligands and the electron count of the cluster (hence the oxidation state of the metal atoms). Higher coverages of ligands and higher oxidation states of surface metal atoms can lead to higher coordination numbers.[47] While keeping the electron count constant at 8, we tested the coverage effect by examining 7, 8, and 9 CH_3COO groups on Ag_{13}. We found that the AgO_2 coordination (two CH_3COO groups coordinate one Ag atom) is indeed present at these high coverages (Figure 8-9) but the AgO_3 coordination either changed to AgO_2 or was higher in energy. We expect that the AgO_3 coordination mode

may be stabilized if we lower the total electron count (in other words, making the surface Ag atoms more positively charged). Same can be expected for Cu as well and a detailed further study is warranted.

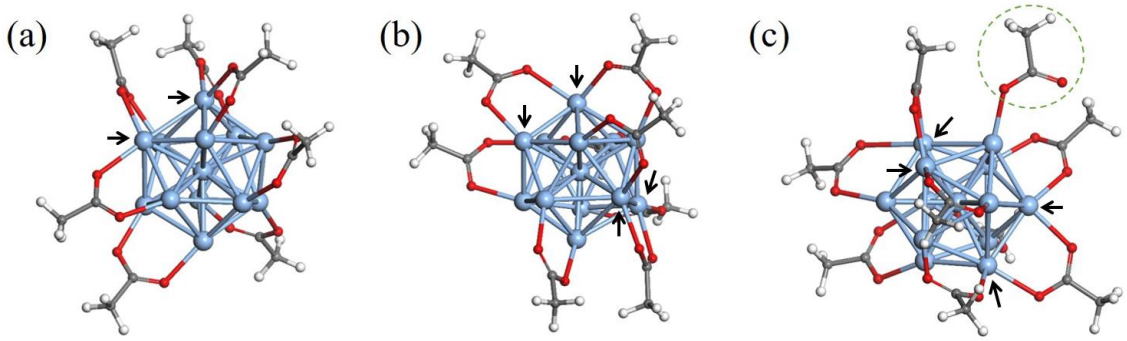
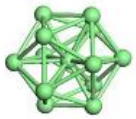
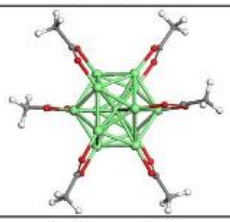
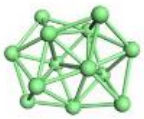
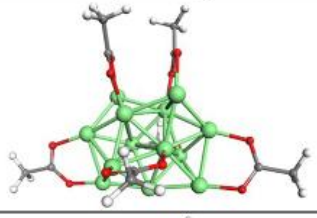
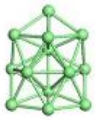
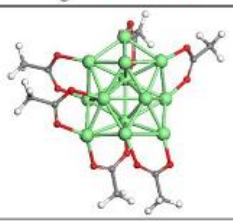


Figure 8-9. Most stable structures after DFT-optimization: (a) $[Ag_{13}(CH_3COO)_7]^{2-}$; (b) $[Ag_{13}(CH_3COO)_8]^{3-}$; (c) $[Ag_{13}(CH_3COO)_9]^{4-}$. Arrows indicate some μ^2-Ag atoms; the dashed circle shows the monodentate CH_3COO group.

Another point to note is that the focus of the present work is on the ligand-metal interface, so we have not explored the global minima of the $[M_{13}(CH_3COO)_6]^-$ clusters beyond the assumed icosahedral core. To assess if this assumption is sound, we have tested a couple of popular minima of bare M_{13} (Cs and C_2 symmetry) in addition to the icosahedral structure.[48] The comparison is shown in supplementary material Figure 8-10. One can see that the $[M_{13}(CH_3COO)_6]^-$ cluster with the icosahedral core is much more stable than the other two cores for both Cu and Ag. Given our focus on the interface and the results in Figure 8-10, we think that using the icosahedral M_{13} core is a good initial model to simulate carboxylate-protected Ag and Cu clusters.

(a)

Cu_{13} kernel	Symmetry	Optimized $[Cu_{13}(CH_3COO)_6]^-$	Relative Energy (eV)
	I_h		0.00
	C_2		+1.51
	C_s		+1.86

(b)

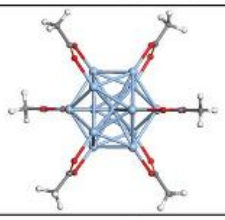
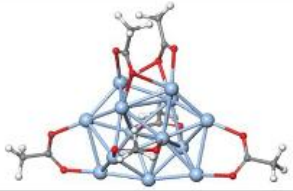
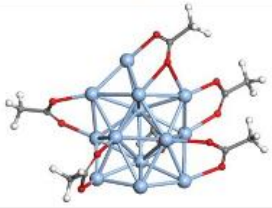
Ag_{13} kernel	Symmetry	Optimized $[Ag_{13}(CH_3COO)_6]^-$	Relative Energy (eV)
	I_h		0.00
	C_2		+1.20
	C_s		+1.75

Figure 8-10. Comparison of the stability of the $[M_{13}(CH_3COO)_6]^-$ clusters with different core structures: (a) $[Cu_{13}(CH_3COO)_6]^-$; (b) $[Ag_{13}(CH_3COO)_6]^-$.

8.5 Summary and conclusions

We have investigated the interfaces between a model carboxylate group, CH_3COO, and the coinage metal surfaces and clusters by first principles density functional theory. The μ_2-CH_3COO mode was found to be the most stable configuration on all three M(111) surfaces (M=Cu, Ag, Au), while the μ_3-CH_3COO is also stable on Cu(111) and Ag(111). The binding energy of CH_3COO correlates with the amount of charge transfer from the metal surface to the adsorbate; CH_3COO binds with Cu(111) the strongest with the shortest interfacial Cu-O bonds and the largest amount of negative charge (~ -0.65 e). A distinct and similar saturation coverage of about 7 CH_3COO groups per nm^2 was found on all three surfaces. At the saturation coverage on the Ag(111) and Au(111) surfaces, all surface metal atoms are coordinated by μ_2-CH_3COO, while μ_3-CH_3COO modes appear in some intermediate coverages on Cu(111) and Ag(111). Geometry optimization of monolayer-protected icosahedral M_{13} clusters with six CH_3COO ligands found that the eight-electron superatomic $[Cu_{13}(CH_3COO)_6]^-$ and $[Ag_{13}(CH_3COO)_6]^-$ nanoclusters are stable and maintain the icosahedral kernel, but a larger deformation was found in the case of Au. Our work provides insights into the interface between carboxylate groups and coinage metals that could help design and synthesis of new atomically precise coinage-metal clusters.

References

1. S. A. DiBenedetto, A. Facchetti, M. A. Ratner and T. J. Marks, *Adv. Mater.*, 2009, **21**, 1407-1433.
2. S. Bharathi, M. Nogami and O. Lev, *Langmuir*, 2001, **17**, 2602-2609.
3. M. A. Abbas, P. V. Kamat and J. H. Bang, *ACS Energy Letters*, 2018, **3**, 840-854.

4. A. Yahia-Ammar, D. Sierra, F. Mérola, N. Hildebrandt and X. Le Guével, *ACS nano*, 2016, **10**, 2591-2599.

5. S. S. Kumar, K. Kwak and D. Lee, *Electroanalysis*, 2011, **23**, 2116-2124.

6. S. Zeng, K.-T. Yong, I. Roy, X.-Q. Dinh, X. Yu and F. Luan, *Plasmonics*, 2011, **6**, 491.

7. G. Li and R. Jin, *Acc. Chem. Res.*, 2013, **46**, 1749-1758.

8. R. Jin, G. Li, S. Sharma, Y. Li and X. Du, *Chem. Rev.*, 2021, **121**, 567-648.

9. K. Konishi, M. Iwasaki and Y. Shichibu, *Acc. Chem. Res.*, 2018, **51**, 3125-3133.

10. D.-e. Jiang, M. L. Tiago, W. Luo and S. Dai, *J. Am. Chem. Soc.*, 2008, **130**, 2777-2779.

11. Y. Pei, Y. Gao and X. C. Zeng, *J. Am. Chem. Soc.*, 2008, **130**, 7830-7832.

12. H. Grönbeck and H. Häkkinen, *J. Phys. Chem. B*, 2007, **111**, 3325-3327.

13. H. Häkkinen, *Nature chemistry*, 2012, **4**, 443.

14. P. Maity, H. Tsunoyama, M. Yamauchi, S. Xie and T. Tsukuda, *J. Am. Chem. Soc.*, 2011, **133**, 20123-20125.

15. A. V. Zhukhovitskiy, M. J. MacLeod and J. A. Johnson, *Chem. Rev.*, 2015, **115**, 11503-11532.

16. M. R. Narouz, K. M. Osten, P. J. Unsworth, R. W. Y. Man, K. Salorinne, S. Takano, R. Tomihara, S. Kaappa, S. Malola, C.-T. Dinh, J. D. Padmos, K. Ayoo, P. J. Garrett, M. Nambo, J. H. Horton, E. H. Sargent, H. Häkkinen, T. Tsukuda and C. M. Crudden, *Nature Chemistry*, 2019, **11**, 419-425.

17. Q. Tang, G. Hu, V. Fung and D.-e. Jiang, *Acc. Chem. Res.*, 2018, **51**, 2793-2802.

18. C. M. Aikens, *Acc. Chem. Res.*, 2018, **51**, 3065-3073.

19. Y. Pei, P. Wang, Z. Ma and L. Xiong, *Acc. Chem. Res.*, 2019, **52**, 23-33.

20. Q. Tang and D.-e. Jiang, *Chem. Mater.*, 2017, **29**, 6908-6915.

21. P. D. Jadzinsky, G. Calero, C. J. Ackerson, D. A. Bushnell and R. D. Kornberg, *Science*, 2007, **318**, 430-433.

22. P. Maksymovych, D. C. Sorescu and J. T. Yates, *Phys. Rev. Lett.*, 2006, **97**, 146103.

23. Q. Tang and D.-e. Jiang, *The Journal of Physical Chemistry C*, 2015, **119**, 10804-10810.

24. X.-K. Wan, Q. Tang, S.-F. Yuan, D.-e. Jiang and Q.-M. Wang, *J. Am. Chem. Soc.*, 2015, **137**, 652-655.

25. Z. Lei, X.-K. Wan, S.-F. Yuan, Z.-J. Guan and Q.-M. Wang, *Acc. Chem. Res.*, 2018, **51**, 2465-2474.

26. J. Yan, B. K. Teo and N. Zheng, *Acc. Chem. Res.*, 2018, **51**, 3084-3093.

27. Y. T. Tao, *J. Am. Chem. Soc.*, 1993, **115**, 4350-4358.

28. S. J. Lee, S. W. Han, H. J. Choi and K. Kim, *J. Phys. Chem. B*, 2002, **106**, 2892-2900.

29. S.-Y. Lin, T.-K. Tsai, C.-M. Lin, C.-h. Chen, Y.-C. Chan and H.-W. Chen, *Langmuir*, 2002, **18**, 5473-5478.

30. J.-W. Park, *Particle & Particle Systems Characterization*, 2019, **36**, 1800329.

31. H. Yu, T. Szilvási, K. Wang, J. I. Gold, N. Bao, R. J. Twieg, M. Mavrikakis and N. L. Abbott, *J. Am. Chem. Soc.*, 2019, **141**, 16003-16013.

32. K.-G. Liu, X.-M. Gao, T. Liu, M.-L. Hu and D.-e. Jiang, *J. Am. Chem. Soc.*, 2020, **142**, 16905-16909.

33. W.-D. Liu, J.-Q. Wang, S.-F. Yuan, X. Chen and Q.-M. Wang, *Angew. Chem. Int. Ed.*, DOI: https://doi.org/10.1002/anie.202100972.

34. M. Walter, J. Akola, O. Lopez-Acevedo, P. D. Jadzinsky, G. Calero, C. J. Ackerson, R. L. Whetten, H. Grönbeck and H. Häkkinen, *Proceedings of the National Academy of Sciences*, 2008, **105**, 9157-9162.

35. G. Kress and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169-11186.

36. P. E. Blöchl, *Physical review B*, 1994, **50**, 17953.

37. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

38. G. Henkelman, A. Arnaldsson and H. Jónsson, *Computational Materials Science*, 2006, **36**, 354-360.

39. R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165-169.

40. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.

41. D. Andrae, U. Häußermann, M. Dolg, H. Stoll and H. Preuß, *Theor. Chim. Acta*, 1990, **77**, 123-141.

42. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, 154104.

43. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, 1983, **105**, 7512-7516.

44. M. Abyazisani, J. Bradford, N. Motta, J. Lipton-Duffin and J. MacLeod, *The Journal of Physical Chemistry C*, 2018, **122**, 17836-17845.

45. M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz and R. Jin, *J. Am. Chem. Soc.*, 2008, **130**, 5883-5885.

46. J.-J. Li, Z.-J. Guan, Z. Lei, F. Hu and Q.-M. Wang, *Angew. Chem. Int. Ed.*, 2019, **58**, 1083-1087.

47. S. Sharma, K. K. Chakrahari, J.-Y. Saillard and C. Liu, *Acc. Chem. Res.*, 2018, **51**, 2475-2483.

48. K. E. Batista, V. K. Ocampo-Restrepo, M. D. Soares, M. G. Quiles, M. J. Piotrowski and J. L. Da Silva, *J. Chem. Inf. Model.*, 2020, **60**, 537-545.

Chapter 9. Summary and Outlook

In this dissertation, we studied the selective ligands in solvent extraction to separate trivalent lanthanides (Ln) by using different computational approaches mainly including DFT and machine learning. Several novel ligands have been suggested as synthesis targets accessible for experimental study in the future, including but not limited to bis-lactam-1,10-phenanthroline derived N-oxide ligands, hexalkyl-nitrilotriacetamide, and dithiophosphinic acids. According to different types of ligands, different effects of several factors have been individually studied, such as conjugation size, charge distribution, donor distance, molecule size, substituent type, etc.

In Chapter 3, the relative aqueous La(III)/Ln(III) selectivity of a novel family of mixed N,O-donor ligands, generated from 2,9-bis-lactam-1,10-phenanthroline (BLPhen) and containing N-oxide functionalities, has been computationally assessed. In the selective separation of trivalent lanthanides, three new ligands were shown to be promising and empirically feasible targets. We discovered that key control factors that influence a ligand's selectivity for lanthanides include conjugation, O-O distance, planarity of the generated complex, and the electron density on the two O atoms.

In Chapter 4, we have trained deep neural networks on the experimental data of distribution coefficients measured for hundreds of ligands for 14 Ln(III) ions to accurately and quickly predict their distribution coefficients for a given ligand and the extraction conditions, advancing the solvent-extraction separation of rare-earth elements. We discovered that the training model performed most accurately on the validation set when molecular physicochemical descriptors and atomic extended-connectivity fingerprints

were combined to best describe the ligands. We have further investigated several hyperparameter combinations, which resulted in a set of ideal hyperparameters. On the validation set, the best trained model did well: $R^2 = 0.85$ and RMSE = 0.53. By modifying the diglycolamide (DGA) backbone and side chains when synthesizing four new ligands, we were able to further test our model by measuring the log D values of Ln(III) ions. We discovered that the predicted distribution coefficients from our trained neural network agree quite well with the measured values. Once ligands are synthesized, one can imagine that our neural network may be used to quickly forecast the log D values of Ln(III) ions for thousands to hundreds of thousands of them.

In Chapter 5, four amide-trizainyl-phenanthroline ligands and more than 60 hexalkyl-nitrilotriacetamide (NTA) ligands have a high potential to improve the selective separations of Ln(III), according to machine learning-based high-throughput screenings on thousands of ligands produced by Fast Assembly of SMILES Fragments (FASMIFRA). Two structure-activity connections have been confirmed by additional computational studies using hundreds of NTA ligands. For an NTA ligand with greater selectivity, one with a bigger size is desired. The other is that adding one or two dimethyl-hexyl groups to the terminals of the amide group substituents on NTA ligands with the same size will increase selectivity.

In Chapter 6, inspired by several successful examples of ligands effective in Eu(III)/Am(III) separation that are able to be furtherly utilized in Ln(III) selective separations, we improve the work discussed in Chapter 4. We perform another machine learning model to accurately and quickly predict the distribution coefficients of Ln(III) and

Am(III) for a given ligand and the extraction conditions. The experimental data on distribution coefficients that are currently available include measurements from over 2,000 extractions and 194 ligands. Atomic extended-connectivity fingerprints and molecular physicochemical descriptors are coupled to represent a ligand. After evaluating and contrasting many different combinations of hyperparameters, the best trained model, with R2 = 0.83 and MAE = 0.45, performed well on the validation set. The trained model analyzes the behavior of ligands used only for the Eu(III)/Am(III) separation in the database. Some bis-phosphine oxide phenanthrolines, dithiophosphinic acids, and bis-pyrazole phenanthrolines have constant high potentials in future Ln(III) selective separations.

In Chapters 7 and 8, in addition to rare-earth elements, several theoretical analyses based on the first principles for atomically precise ligand-protected nanoclusters were carried out to comprehend the interactions between metals and organic ligands. Our work provides insights into the interface between carboxylate groups and coinage metals that μ_2-CH_3COO is the most stable on all three (111) surfaces; μ_3-CH_3COO is only stable on Cu(111) and Ag(111). And $[Cu_{13}(CH_3COO)_6]^-$ and $[Ag_{13}(CH_3COO)_6]^-$ nanoclusters are stable and maintain the icosahedral kernel, but a larger deformation was found in the case of Au. Given the broad availability and variety of carboxylic acids including amino acids, all our work suggests that carboxylate groups could be the next-generation ligands to further expand the universe of atomically precise metal clusters, especially for Cu and Ag.

In conclusion, the most important, accurate and quick model for predicting 14 Ln(III) distribution coefficients for a given ligand and the extraction conditions have been

established. And it is feasible to combine this model and molecule generation tools to accomplish the discovery of new selective ligands in Ln(III) separations. Hence, we have solved the two imperative questions presented in the Introduction Chapter on how to relate the simulation to the extraction property and how to use simulated results to expand the diversity of selective extracts. More data may be added to the training data set when it becomes available to further enhance this model, especially for new ligand systems, but the core approach is similar, leading to simple training and high efficiency. The other topic for the lanthanide selective extraction we make some efforts into is the structure-selectivity relationship. The current machine learning model can make some conclusions based on a large number of data and statistics, but the simulations via density functional theory provide more chemical insights such as partial charge and orbital information although the efficiency is limited. The limited speed is caused by too many electrons in the lanthanide complexes, too many possible configurations caused by anions, etc. A machine learning model using an algorithm traceable to the structure information of inputs may provide more chemical understandings with high efficiency and effectiveness. All in all, future chemistry in all areas will benefit from the machine learning approach with different core algorithms, not only the extractants of lanthanides and actinides, but also all the interactions between organic ligands and metal cations. Even a general model accessible for all chemical bindings in the coordinated metal complex can be expected which can revolutionarily evolve the methodology and knowledge in coordination chemistry, separation chemistry and all related fields.

Appendix A. SMILES of Ligands in the Dataset

SMILES of 194 ligands used in Ln (III), Am(III), Cm(III), Bk(III), Cf(III), and Es(III) solvent extractions

O=C(c1cccc(n1)C(=O)N(C)c1ccccc1)N(C)c1ccccc1

C67=CC1=C(C2=C(C=C1)C=C3C(=N2)C(N(C(=C3C4=CC=CC=C4)C5=CC=CC=C5)CCCCCC)=O)N=C6C(N(C(=C7C8=CC=CC=C8)C9=CC=CC=C9)CCCCCC)=O

C12=NC5=C(C=C1C3C(N(C2=O)CCCCCC)C4CCC3C4)C=CC6=CC7=C(N=C56)C(N(C8C7C9CCC8C9)CCCCCC)=O

C12=CC=C(C(N(CCCCCCCC)CCCCCCCC)=O)N=C1C3=C(C=C2)C=CC(=N3)C(N(CCCCCCCC)CCCCCCCC)=O

C(C(=O)N(C)C1=CC=CC=C1)OCC(=O)N(C)C2=CC=CC=C2

C(C(=O)N(CCCC)CCCC)OCC(=O)N(CCCC)CCCC

CC(C)CN(CC(C)C)C(=O)COCC(=O)N(CC(C)C)CC(C)C

C(C(=O)N(CCCCCC)CCCCCC)OCC(=O)N(CCCCCC)CCCCCC

C(C(=O)N(CCCCCCCC)CCCCCCCC)OCC(=O)N(CCCCCCCC)CCCCCCCC

C(C(=O)N(C)CCCCCCCC)OCC(=O)N(C)CCCCCCCC

C(C(=O)N(C(CC)C)CCCCCCCC)OCC(=O)N(C(CC)C)CCCCCCCC

O=C([C@H](C)O[C@H](C)C(=O)N(CCCCCCCC)CCCCCCCC)N(CCCCCCCC)CCCCCCCC

O=C([C@H](C)O[C@@H](C)C(=O)N(CCCCCCCC)CCCCCCCC)N(CCCCCCCC)CCCCCCC

CN(C(=O)CC(=O)N(C)c1ccccc1)c1ccccc1

CN(C(=O)COCC(=O)N(C)c1ccccc1)c1ccccc1

CCCCCCOP(O)(=O)OCCCCCC

O=C(C(CCOCCCCCC)C(=O)N(C)CCCCCCCC)N(C)CCCCCCCC

O=C(COCC(=O)N(CCCCCCCCCC)CCCCCCCCCC)N(CCCCCC)CCCCCC

CC12CCC(c3nnc(nc31)c1cccc(n1)c1nc3c(nn1)C1CCC3(C)C1(C)C)C2(C)C

O=C(c1cccc(n1)c1cccc(n1)C(=O)N(CC)c1ccccc1)N(CC)c1ccccc1

O=C(c1cccc(n1)c1cccc(n1)C(=O)N(CC)c1cc(C)ccc1C)N(CC)c1cc(C)ccc1C

O=C(c1cccc(n1)c1cccc(n1)C(=O)N(CC)c1ccc(C)cc1C)N(CC)c1ccc(C)cc1C

O=C(c1cccc(n1)c1cccc(n1)C(=O)N(CC)c1cc(C)cc(C)c1)N(CC)c1cc(C)cc(C)c1

CCCCCCCCCN(CCCCCCCCC)C(=O)c1ccc2ccc3cccnc3c2n1

Cc1ccc(cc1)N(CCCCCCCCC)C(=O)c1ccc2ccc3cccnc3c2n1

Cc1cc(cc(C)c1)N(CCCCCCCCC)C(=O)c1ccc2ccc3cccnc3c2n1

Cc1cc(cc(C)c1)N(CCCCCCCCC)C(=O)c1nc2c(cc1)ccc1ccc(nc12)C(=O)N(CCCCCCCCC)c1cc(C)cc(C)c1

Cc1ccc(cc1)N(CCCCCCCCC)C(=O)c1nc2c(cc1)ccc1ccc(nc12)C(=O)N(CCCCCCCCC)c1ccc(C)cc1

CCCc1nnc(nc1CCC)c1cccc(n1)c1nc(CCC)c(CCC)nn1

CCCCCc1nc(nnc1CCCCC)c1cccc(n1)c1cccc(n1)c1nc(CCCCC)c(CCCCC)nn1

CCCCN(CCCC)C(=O)C(CC1=CC=CC=C1CC(C(=O)N(CCCC)CCCC)C(=O)N(CCCC)CCCC)C(=O)N(CCCC)CCCC

O=P(Cc1cccc(CP(=O)(CCCCCCCC)CCCCCCCC)[n+]1[O-])(CCCCCCCC)CCCCCCCC

CCCCN(CCCC)C(=O)CCC(=O)N(CCCC)CCCC

CCc1ccc(cc1)N(CC)C(=O)c1nc2c(cc1)ccc1ccc(nc12)C(=O)N(CC)c1ccc(CC)cc1

CCCCN(CCCC)C(=O)c1nc2c(cc1)ccc1ccc(nc12)C(=O)N(CCCC)CCCC

O=C(O)CN(CC(=O)O)CCN(CCO)CC(=O)O

O=C(O)CN(CC(=O)O)C1CCCCC1N(CC(=O)O)CC(=O)O

O=C(c1cccc(n1)C(=O)N(CC)c1cc(C)ccc1)N(CC)c1cc(C)ccc1

O=C(c1cccc(n1)C(=O)N(CC)c1ccc(C)cc1)N(CC)c1ccc(C)cc1

O=C(c1cccc(n1)C(=O)N(CC)c1ccccc1C)N(CC)c1ccccc1C

CCN(C(=O)c1cccc(c1)c1cccc(c1)C(=O)N(CC)c1ccccc1)c1ccccc1

O=C(COCC(=O)N(CCCC)C(C)CCCCC)N(CCCC)CC(C)CCCC

CN(CCCCCCCC)C(=O)COCC(=O)N(C)CCCCCCCC

CCCCCCCCN(CCCCCCCC)C(=O)CN(CC(=O)N(CCCCCCCC)CCCCCCCC)CC(=O)N(CCCCCCCC)CCCCCCCC

Oc1ccc(cc1O)C=1Oc2cc(O)cc(O)c2C(=O)C=1O

CN(CCCCCCCCCCCC)C(=O)COCC(=O)N(C)CCCCCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCCCCC)CC)N(CC)CCCCCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCCCCC)CCC)N(CCC)CCCCCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCCCCC)CCCC)N(CCCC)CCCCCCCCCCCC

CCCCCCCCP(=O)(CC(=O)N(CC(C)C)CC(C)C)C1=CC=CC=C1

O=C(C(CCOCCCCCC)C(=O)N(C)CCCC)N(C)CCCC

O=C(CC(=O)N(C)CCCC)N(C)CCCC

CN(CCCCCCCC)C(=O)C(CCCCCC)C(=O)N(C)CCCCCCCC

O=C(CC(=O)N(C)CCCCCCCC)N(C)CCCCCCCC

O=C(CC(=O)N(C)CCCCCC)N(C)CCCCCC

CCCCCCN(CCCCCC)C(=O)CC(=O)N(CCCCCC)CCCCCC

CCCCCCCCN(CCCCCCCC)C(=O)CC(=O)N(CCCCCCCC)CCCCCCCC

CCCCN(CCCC)C(=O)CC(=O)N(CCCC)CCCC

O=C(c1nc(ccc1)C(=O)N1CCCC1)N1CCCC1

O=C(c1nc(ccc1)c1cccc(n1)C(=O)N1CCCC1)N1CCCC1

O=C(c1nc2c(cc1)ccc1ccc(nc12)C(=O)N1CCCC1)N1CCCC1

O=P(c1ccccc1)(c1ccccc1)c1nc2c(cc1)ccc1ccc(nc12)P(=O)(c1ccccc1)c1ccccc1

O=P(OCCCC)(OCCCC)c1nc2c(cc1)ccc1ccc(nc12)P(=O)(OCCCC)OCCCC

CCCCc1cc(n[NH]1)c1ccc2ccc3ccc(nc3c2n1)c1cc(CCCC)[NH]n1

CC(C)Cc1cc(n[NH]1)c1ccc2ccc3ccc(nc3c2n1)c1cc(CC(C)C)[NH]n1

CCCCCCCCc1cc(n[NH]1)c1ccc2ccc3ccc(nc3c2n1)c1cc(CCCCCCCC)[NH]n1

O=C(c1nc2c(ccc3c(Cl)cc(nc32)C(=O)N2CCCC2)c(Cl)c1)N1CCCC1

O=C(c1cc(Cl)c2ccc3c(Cl)cc(nc3c2n1)C(=O)N1CCCCC1)N1CCCCC1

Clc1cc(nc2c1ccc1c(Cl)cc(nc12)C(=O)N1CCc2ccccc21)C(=O)N1CCc2ccccc21

Clc1cc(nc2c1ccc1c(Cl)cc(nc12)C(=O)N1CCCc2ccccc21)C(=O)N1CCCc2ccccc21

Clc1cc(nc2c1ccc1c(Cl)cc(nc12)C(=O)n1c2ccccc2c2ccccc21)C(=O)n1c2ccccc2c2ccccc21

CC1(C)CCC(C)(C)c2nnc(nc21)c1cccc(n1)c1cccc(n1)c1nc2c(nn1)C(C)(C)CCC2(C)C

O=P(CCCCCCCC)(CC(=O)N(CC(C)C)CC(C)C)c1ccccc1

O=C(COCC(=O)N(CCCC)CCCC)N(CCCC)CCCC

Cc1ccc(cc1)N(CC)C(=O)c1ccc2ccc3cccnc3c2n1

O=C(COCC(=O)N(CCCCCCCC)CCCCCCCC)N(CCCCCCCC)CCCCCCCC

O=C(COCC(=O)N(CCCCC)CCCCC)N(CCCCC)CCCCC

O=C(COCC(=O)N(CCCCCC)CCCCCC)N(CCCCCC)CCCCCC

O=C(COCC(=O)N(CC(CC)CCCC)CC(CC)CCCC)N(CC(CC)CCCC)CC(CC)CCCC

O=C(COCC(=O)N(CCCCCCCCCCCC)CCCCCCCCCCCC)N(CCCCCCCCCCCC)CCCCCCCCCC

O=P(OCC)(OCC)c1nc2c(cc1)ccc1ccc(nc12)P(=O)(OCC)OCC

CCCCCCCCN(CCCCCCCC)C(=O)COCC(=O)NCc1c(CC)c(NCC(=O)COCC(=O)N(CCCCCCC)CCCCCCCC)c(CC)c(NCC(=O)COCC(=O)N(CCCCCCCC)CCCCCCCC)c1CC

CCCCCCCCN(CCCCCCCC)C(=O)COCC(=O)NCCNC(=O)c1cc(cc(c1)C(=O)NCCNC(=O)COCC(=O)N(CCCCCCCC)CCCCCCCC)C(=O)NCCNC(=O)COCC(=O)N(CCCCCCCC)CCCCCCCC

CCCCCCCCN(CCCCCCCC)C(=O)COCC(=O)NCCCOc1cc(OCCCNC(=O)COCC(=O)N(CCCCCCCC)CCCCCCCC)cc(c1)OCCCNC(=O)COCC(=O)N(CCCCCCCC)CCCCCCCC

O=C(COCC(=O)N(CCCCCC)CCCCCC)N1CCCC1

O=C(COCC(=O)N(CCCCCC)CCCCCC)N1CCCCC1

O=C(COCC(=O)N(CCCCCC)CCCCCC)N1CCOCC1

O=C(COCC(=O)N(CCCCCCCC)CCCCCCCC)N1CCCC1

O=C(COCC(=O)N(CCCCCCCC)CCCCCCCC)N1CCCCC1

O=C(COCC(=O)N(CCCCCCCC)CCCCCCCC)N1CCOCC1

OC(=O)c1ccccc1C(=O)N(CCCC)CCCC

C1COCCOCCOCCOCCOCCO1

O1CCOCCOC2CCCCC2OCCOCCOC2CCCCC12

C1COc2ccccc2OCCOCCOc2ccccc2OCCO1

C1Oc2ccccc2OCCOCCOc2ccccc2OCc2cccc1n2

CCCCP(=O)(c1ccccc1)c1ccc2ccc3cccnc3c2n1

CCCCP(=O)(c1ccccc1)c1nc2c(cc1)ccc1ccc(nc12)P(=O)(CCCC)c1ccccc1

Cc1ccc(cc1)N(CC)C(=O)c1nc2c(cc1)ccc1ccc(nc12)C(=O)N(CC)c1ccc(C)cc1

O=C(COCC(=O)N(CC(CC)CCCC)CCCCCCCCC)N(CCCCCCCCC)CC(CC)CCCC

CN(CC(CC)CCCC)C(=O)COCC(=O)N(C)CC(CC)CCCC

CN(CCCC(CCCC)CCCCCC)C(=O)COCC(=O)N(C)CCCC(CCCC)CCCCCC

CN(CCCCC(C)CCCC(C)C)C(=O)COCC(=O)N(C)CCCCC(C)CCCC(C)C

O=C(COCC(=O)N(CCCCCCCCC)CC)N(CC)CCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCC)CCC)N(CCC)CCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCC)CCCCCCCCC)N(C)C

O=C(COCC(=O)N(CCCCCCCCC)CCCCCCCCC)N(C)CCCCCCCCC

O=C(COCC(=O)N(CCCCCCCCCCCCC)CCCCCCCCC)N(CCCCCCCCC)CCCCCCCCCCCC

O=C(COCC(=O)N(CCC(C)CC(C)(C)C)CCCCCCCCC)N(CCCCCCCCC)CCC(C)CC(C)(C)C

O=C(COCC(=O)N(CCCC(CCCC)CCCCCC)CCCCCCCCC)N(CCCCCCCCC)CCCC(CCCC)CCCCCC

O=C(COCC(=O)N(CCCCC(C)CCCC(C)C)CCCCCCCCC)N(CCCCCCCCC)CCCCC(C)CCCC(C)C

O=C(COCC(=O)N1CC(C)C(CSC(C)(C)CCCCCCCCC)C1)N1CC(C)C(CSC(C)(C)CCCCCCCCC)C1

O=C(P(OCCCC)(OCCCC)=O)N(CC)CC

O=P(OCCCC)(OCCCC)CC(N(CC)CC)=O

O=P(OCCCC)(OCCCC)CP(OCCCC)(OCCCC)=O

O=P(OCCCCCC)(OCCCCCC)CC(N(CC)CC)=O

O=P(OC1=CC=CC=C1)(OCCCCCC)CC(N(CC)CC)=O

O=P(OCCCCCC)(CC(N(CC)CC)=O)CCCCCC

O=P(CCCCCC)(CC(N(CC)CC)=O)CCCCCC

O=P(CC(CC)CC)(CC(N(CC)CC)=O)CC(CC)CC

O=P(CC(N(CCCC)CCCC)=O)(CCCCCC)CCCCCC

O=P(CC(N(CC(C)C)CC(C)C)=O)(CCCCCC)CCCCCC

O=P(CC(N(C(C)CC)C(C)CC)=O)(CCCCCC)CCCCCC

O=P(OCCCCCC)(CCC(N(CC)CC)=O)OCCCCCC

O=P(OCCCCCC)(CC(N(CC)CC)=O)OCCCCCC

O=P(CCCCCCCC)(CC(N(CC(C)C)CC(C)C)=O)C1=CC=CC=C1

O=P(CC(N(CC)CC)=O)(CC(CC)CC)CC(CC)CC

O=P(CCCCCCCC)(CCCCCCCC)CC(N(CC(C)C)CC(C)C)=O

CC(C)CN(CC(C)C)C(CP(C1CCCCC1)(C2CCCCC2)=O)=O

O=P(CCCCCC)(CC(CC(C)C)=O)CCCCCC

O=P(OCC)(OCC)C1=NC2=C(N=C(P(OCC)(OCC)=O)C=C3)C3=CC=C2C=C1

O=C1N(CCCCCC)[C@@H]2[C@@H](C3CCC2C3)C(C1=C4)=CC5=C4C(C=C(C(N(CCCCCC)[C@H]6C7CCC(C7)[C@@H]86)=O)C8=C9)=C9C=C5

O=C1N(CCCCCC)C(c2ccccc2)C(C3=CC=CC=C3)C(C1=C4)=CC5=C4C(C=C(C(N(CCCCCC)C(c6ccccc6)C7C8=CC=CC=C8)=O)C7=C9)=C9C=C5

O=C1N(CCCCCC)C(CCC)C(CCC)C(C1=C2)=CC3=C2C(C=C(C(N(CCCCCC)C(CCC)C4CCC)=O)C4=C5)=C5C=C3

O=C(COCC(N(CCCCCCCC)CCCCCCCC)=O)N(CCCCCCCC)CCCCCCCC

O=C(N(CCCCCCCC)CCCCCCCC)C1=NC(C2=CC=CC(C3=NC(C(N(CCCCCCCC)CCCCCCC)=O)=CC=C3)=N2)=CC=C1

O=C(N(CCCC)CCCC)C1=NC(C2=CC=CC(C3=NC(C(N(CCCC)CCCC)=O)=CC=C3)=N2)=CC=C1

O=C(N(C1=CC=CC=C1)CC)C2=NC(C3=CC=CC(C4=NC(C(N(CC)C5=CC=CC=C5)=O)=CC=C4)=N3)=CC=C2

CC1=NC(C2=NC(C3=NC(C)=C(C)N=N3)=CC=C2)=NN=C1C

CCC(N=N1)=C(CC)N=C1C2=CC=CC(C3=NN=C(CC)C(CC)=N3)=N2

O=C(NC1=CC=CC=C1)C2=NC3=C(N=C(C(NC4=CC=CC=C4)=O)C=C5)C5=CC=C3C=C2

O=C(N(C)C1=CC=CC=C1)C2=NC3=C(N=C(C(N(C)C4=CC=CC=C4)=O)C=C5)C5=CC=C3C=C2

O=C(N(CC)C1=CC=CC=C1)C2=NC3=C(N=C(C(N(CC)C4=CC=CC=C4)=O)C=C5)C5=CC=C3C=C2

O=C(N(CCCCCCCC)C1=CC=CC=C1)C2=NC3=C(N=C(C(N(CCCCCCCC)C4=CC=CC=C4)=O)C=C5)C5=CC=C3C=C2

O=P(OC(C)C)(OC(C)C)C1=CC=CC(P(OC(C)C)(OC(C)C)=O)=C1

O=P(OC1CCCCC1)(OC2CCCCC2)C3=CC=CC(P(OC4CCCCC4)(OC5CCCCC5)=O)=C3

O=P(OCCCCC)(OCCCCC)C1=CC=CC(P(OCCCCC)(OCCCCC)=O)=C1

O=P(OCC(CC)CCCC)(OCC(CC)CCCC)C1=CC=CC(P(OCC(CC)CCCC)(OCC(CC)CCCC)=O)=C1

O=C(N1CCCC2=C1C=CC=C2)C3=NC4=C(N=C(C(N5C(C=CC=C6)=C6CCC5)=O)C=C7)C7=CC=C4C=C3

O=C(N(CCCCCCCC)CCCCCCCC)CO

O=C(COCC(N(CCCCCCCCCCCC)CCCCCCCCCCCC)=O)N(CCCCCCCCCCCC)CCCCCCCCCC

O=C(COCC(N(CCCCCCCCCC)CCCCCCCCCC)=O)N(CCCCCCCCCC)CCCCCCCCCC

O=C(COCC(N(CCCCCC)CCCCCC)=O)N(CCCCCC)CCCCCC

O=C(COCC(N(CCCCC)CCCCC)=O)N(CCCCC)CCCCC

O=C(COCC(N(CCCC)CCCC)=O)N(CCCC)CCCC

O=C(COCC(N(CCC)CCC)=O)N(CCC)CCC

O=C(COCC(N(C1=CC=CC=C1)C2=CC=CC=C2)=O)N(C3=CC=CC=C3)C4=CC=CC=C4

O=C(COCC(N(C1=CC=CC=C1)C)=O)N(C2=CC=CC=C2)C

O=C(COCC(N(CC(CC)CCCC)CC(CC)CCCC)=O)N(CC(CC)CCCC)CC(CC)CCCC

CC1(C)[C@]2(C(N3CCCCC3)=O)C4=NC(C5=NC6=C(N=C(C7=NN=C([C@H]8CC[C@]9(C(N%10CCCCC%10)=O)C8(C)C)C9=N7)C=C%11)C%11=CC=C6C=C5)=NN=C4[C@H]1CC2

CCCCCC(C(CCCCC)=N1)=NN=C1C2=CC=CC(C3=CC=CC(C4=NC(CCCCC)=C(CCCCC)N=N4)=N3)=N2

O=C(C(CCCCCCCCCCCCCC)C(N(CCCC)C)=O)N(CCCC)C

O=C(C(CCCCCCCCCCCCCC)C(N(C1=CC=C(Cl)C=C1)C)=O)N(C2=CC=C(Cl)C=C2)C

O=C(C(CCOCCCCCCCCCCCC)C(N(CCCC)C)=O)N(CCCC)C

O=C(C(CCCOCCCCCCCCCCC)C(N(CCCC)C)=O)N(CCCC)C

O=C(C(CCOCCCCCC)C(N(CCCCCCCC)C)=O)N(CCCCCCCC)C

O=P(OCCCC)(OCCCC)C1=NC(C(N=C(P(OCCCC)(OCCCC)=O)C=C2)=C2C=C3)=C3C=C1

SP(C1=CC=CC=C1)(CCCCCCCC)=S

SP(C1=CC=CC=C1)(C2=CC=CC=C2C(F)(F)F)=S

SP(CCCC)(CCCCCCCC)=S

SP(C1=CC=CC=C1C(F)(F)F)(CCCCCCCC)=S

SP(C1=CC=CC=C1C(F)(F)F)(C2=CC=CC=C2C(F)(F)F)=S

CCCCCCCCC1=CC(C2=NC3=C(N=C(C4=NNC(CCCCCCCC)=C4)C=C5)C5=CC=C3C=C2)=NN1

CCCCC1=CC(C2=NC3=C(N=C(C4=NNC(CCCC)=C4)C=C5)C5=CC=C3C=C2)=NN1

CC(C)(C)C1=CC(C2=NC3=C(N=C(C4=NNC(C(C)(C)C)=C4)C=C5)C5=CC=C3C=C2)=NN1

O=C(N1CCCC1)C2=NC(C(N=C(C(N3CCCC3)=O)C=C4Cl)=C4C=C5)=C5C(Cl)=C2

O=P(C1=CC=CC=C1)(C2=CC=CC=C2)C3=NC(C(N=C(P(C4=CC=CC=C4)(C5=CC=CC=C5)=O)C=C6)=C6C=C7)=C7C=C3

O=C(C1=NC2=C(C=C1)C=CC3=C2N=C(C=C3)C(N4CCCC4)=O)N5CCCC5

O=C(N1CCCC1)C2=NC(C3=NC(C(N4CCCC4)=O)=CC=C3)=CC=C2

O=C(N1CCCC1)C2=CC=CC(C(N3CCCC3)=O)=N2

O=P(C1=CC=CC=C1)(CCCC)C2=NC(C(N=C(P(C3=CC=CC=C3)(CCCC)=O)C=C4)=C4C=C5)=C5C=C2

O=P(C1=CC=CC=C1)(CCCC)C2=NC(C(N=CC=C3)=C3C=C4)=C4C=C2

O=P(CC)(C1=NC2=C(C=C1)C=CC3=C2N=C(C=C3)P(C4=CC=CC=C4)(CC)=O)C5=CC=CC=C5

O=P(OCC)(C1=NC2=C(C=C1)C=CC3=C2N=C(C=C3)P(C4=CC=CC=C4)(OCC)=O)C5=CC=CC=C5

O=C(N1CCCC1)C2=NC(C(N=C(C(N3CCCC3)=O)C=C4)=C4C=C5)=C5C=C2

O=C(N1C(C)CCC1)C2=NC(C(N=C(C(N3CCCC3C)=O)C=C4)=C4C=C5)=C5C=C2

O=C(N1C(C)CCC1)C2=NC(C(N=C(C(N3CCCC3C)=O)C=C4Cl)=C4C=C5)=C5C(Cl)=C2

CCCCCCN1C2=C(C=CC=C2)N=C1C3=NC(C4=NC5=C(C=CC=C5)N4)=CC=C3

CCCCCCN1C2=C(C=CC=C2)N=C1C3=NC(C4=NC5=C(C=CC=C5)N4CCCCCC)=CC=C3

CCCCCCCCCCCCCOC1=CC(C2=NC(C=CC=C3)=C3N2)=NC(C4=NC5=C(C=CC=C5)N4)=C1

CCCCCCCCCCCCCOC1=CC(C2=NC(C=CC=C3)=C3S2)=NC(C4=NC5=C(C=CC=C5)S4)=C1

CCCCCCCCCCCCCC(CCCCCCCCCCC)COC1=CC(C2=NC(C=CC=C3)=C3O2)=NC(C4=NC5=C(C=CC=C5)O4)=C1

CCCCCCCCCCCCCOC1=CC(C2=NC(C=CC=C3)=C3O2)=NC(C4=NC5=C(C=CC=C5)O4)=C1

CC(CCC1(C)C)(C)C(C1=N2)=NN=C2C3=CC=CC(C4=CC=CC(C5=CC=CC(C6=NC(C(C)(C)CCC7(C)C)=C7N=N6)=N5)=N4)=N3

Appendix B. Publication List

1. T. Liu, K. R. Johnson, S. Jansone-Popova and D.-e. Jiang, *JACS Au*, 2022, **2**, 1428-1434.

2. T. Liu, A. S. Ivanov, I. Popovs, S. Jansone-Popova and D.-e. Jiang, *RSC Adv.*, under review.

3. T. Liu, and D.-e. Jiang, *J. Chem. Phys.*, 2021, **155**, 034301.

4. Y. Li, Y. Song, X. Zhang, T. Liu, T. Xu, H. Wang, D.-e. Jiang and R. Jin, *J. Am. Chem. Soc.*, 2022, **27**, 12381–12389.

5. S. Wang, T. Liu, and D.-e. Jiang, *ACS Appl. Mater. Interfaces*, 2021, **13**, 53468-53474.

6. Y. Cao, T. Liu, T. Chen, B. Zhang, D.-e. Jiang and J. Xie, *Nat. Commun.*, 2021, **12**, 1-7.

7. Z. Yang, T. Liu, S. Wang, H. Chen, X. Suo, T. Wang, B.P. Thapaliya, D.-e. Jiang, I. Popovs and S. Dai, *Chem. Mater.*, 2021, **33**, 3386-3393.

8. K. G. Liu*, X. M. Gao*, T. Liu*, M. L. Hu, and D.-e. Jiang. *J. Am. Chem. Soc.*, 2020, **142**, 16905-16909.

9. B. Liu, T. Liu, S. W. Luo, and L. Z. Gong, *Org. Lett.*, 2014, **16**, 6164-6167.