香港中文大學
The Chinese University of Hong Kong

Journal Club

nature computational science

Brief Communication

https://doi.org/10.1038/s43588-025-00783-z

Toward a unified benchmark and framework for deep learning-based prediction of nuclear magnetic resonance chemical shifts

Zihan Li

2nd May 2025

Nat. Comput. Sci. **2025**, *5*, 292–300.

Content

1. Background for Research
2. NMRNet
3. Conclusion and Methods

1. Background for Research

Importance

Nuclear magnetic resonance (NMR) is a particularly powerful technique in chemistry, biology and materials science

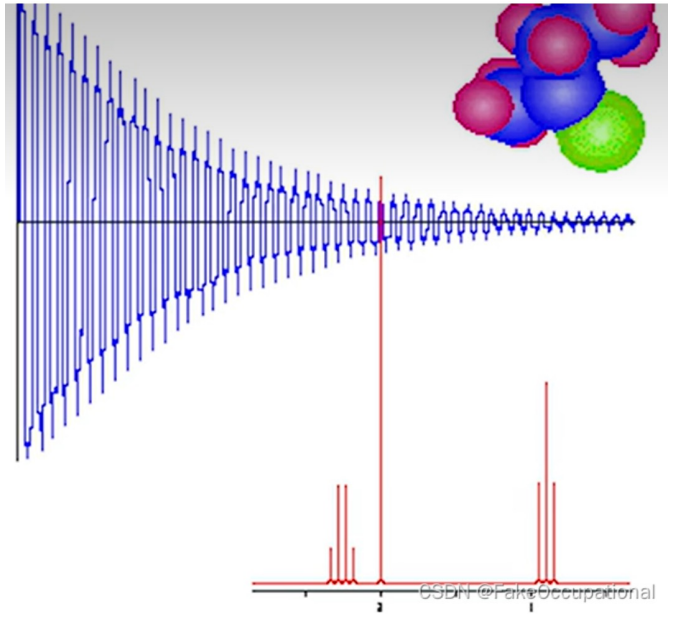
Accurate prediction of NMR chemical shifts aids in spectrum interpretation, structure revision and configuration determination

Traditional methods for NMR chemical shift prediction often struggle to balance accuracy and efficiency, particularly for complex molecular architectures

NMR

Definition

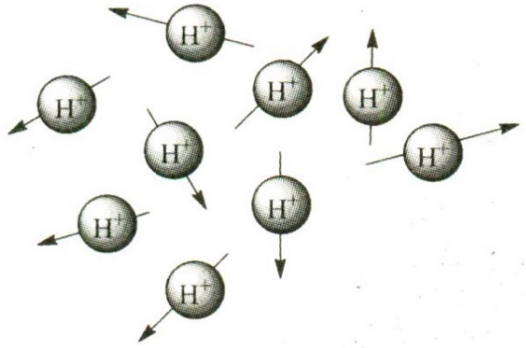
A physical process in which the **atomic nuclei** with **non-zero magnetic moments** undergo splitting of their **spin energy levels** under the action of an **external magnetic field** and **resonantly absorb** the radio-frequency radiation of a certain frequency.

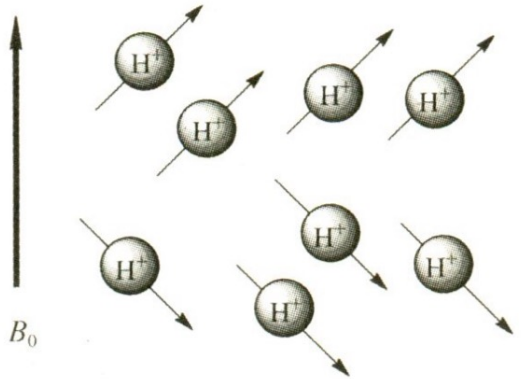


Atoms with magnetic moments

Spin angular momentum ≠ 0

eg. ^{1}H, ^{13}C, ^{19}F, ^{31}P





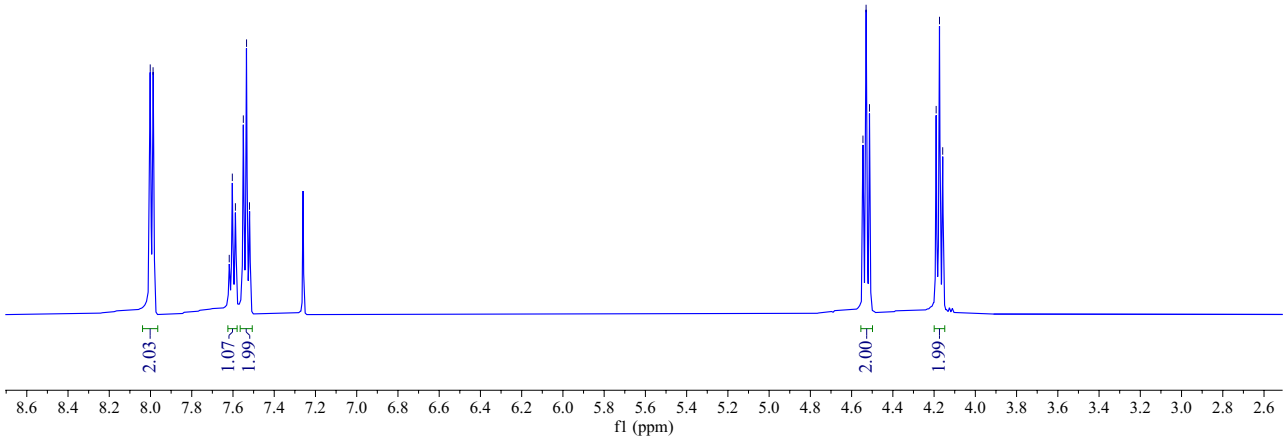
NMR

Chemical shift δ (ppm)

Electron density of each atom is different

the magnetic field intensity required for resonance also varies

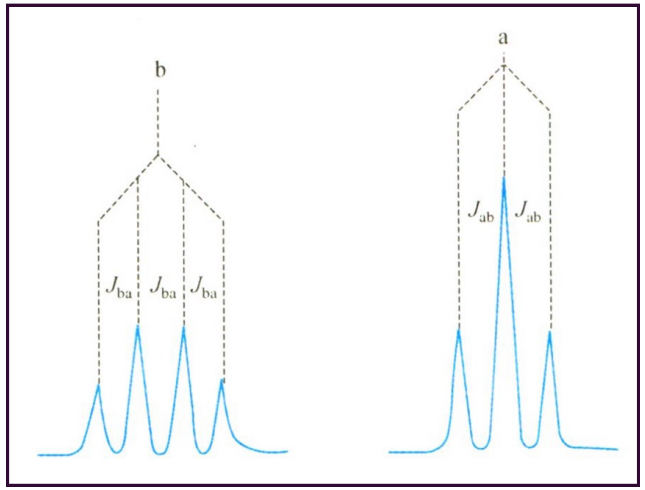
which consequently leads to different chemical shifts.



NMR

Coupling constant *J* (Hz)

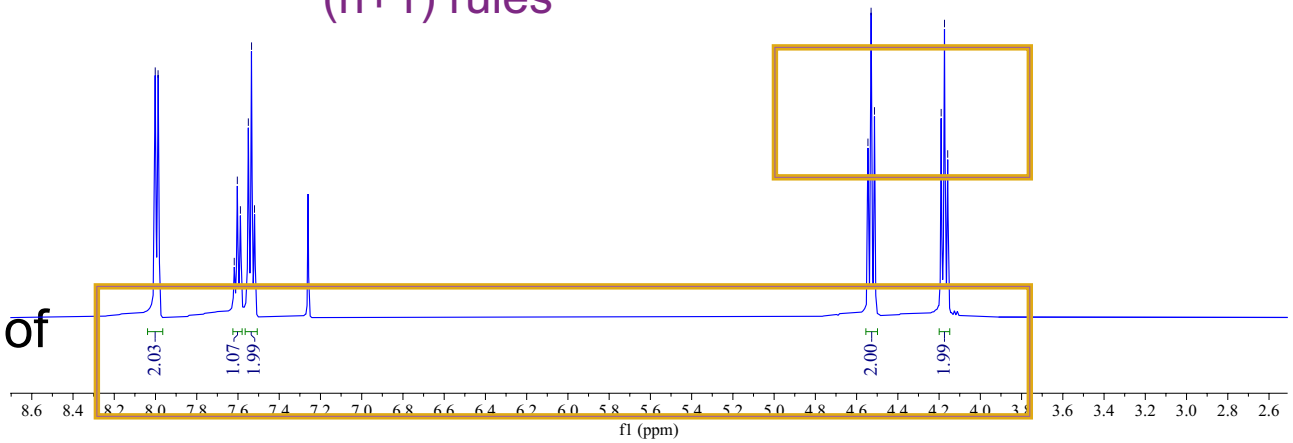
Atoms that are close to each other will influence one another, causing a certain peak to split into multiple peaks.
(like adding another magnetic field)



(n+1) rules

Peak area

In the ^{1}H NMR, it is proportional to the number of corresponding H.



NMR

Solid-state NMR

Test using solid samples. It is used to study the orientation and crystallinity of **polymer materials**, the structure of **crystals**, as well as the structure and function of **biological macromolecules**.

Liquid-state NMR

Dissolve the sample in a deuterated solvent for testing. It is used to study the structures of **small molecules.**

Research Progress

liquid-state NMR

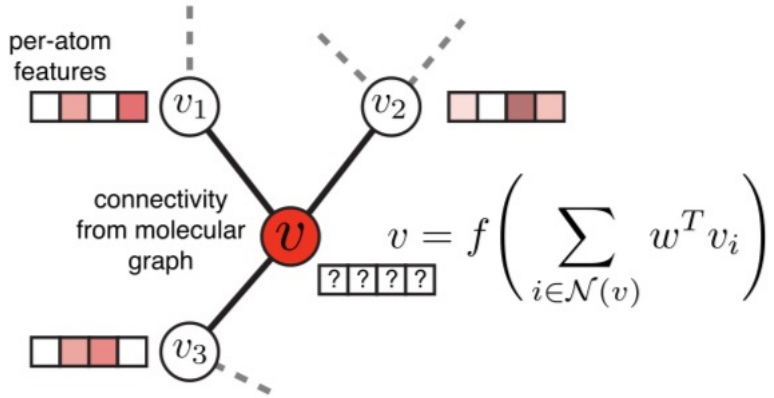
nmrshiftdb2 & QM9-NMR databases

***via* high-throughput framework to solve solute–solvent interactions**

Graph Convolutional Network

&

Message Passing Neural Network



Research Progress

The model must account for **Periodic Boundary Conditions** to accurately represent the system.

This Work

NMRNet Framework

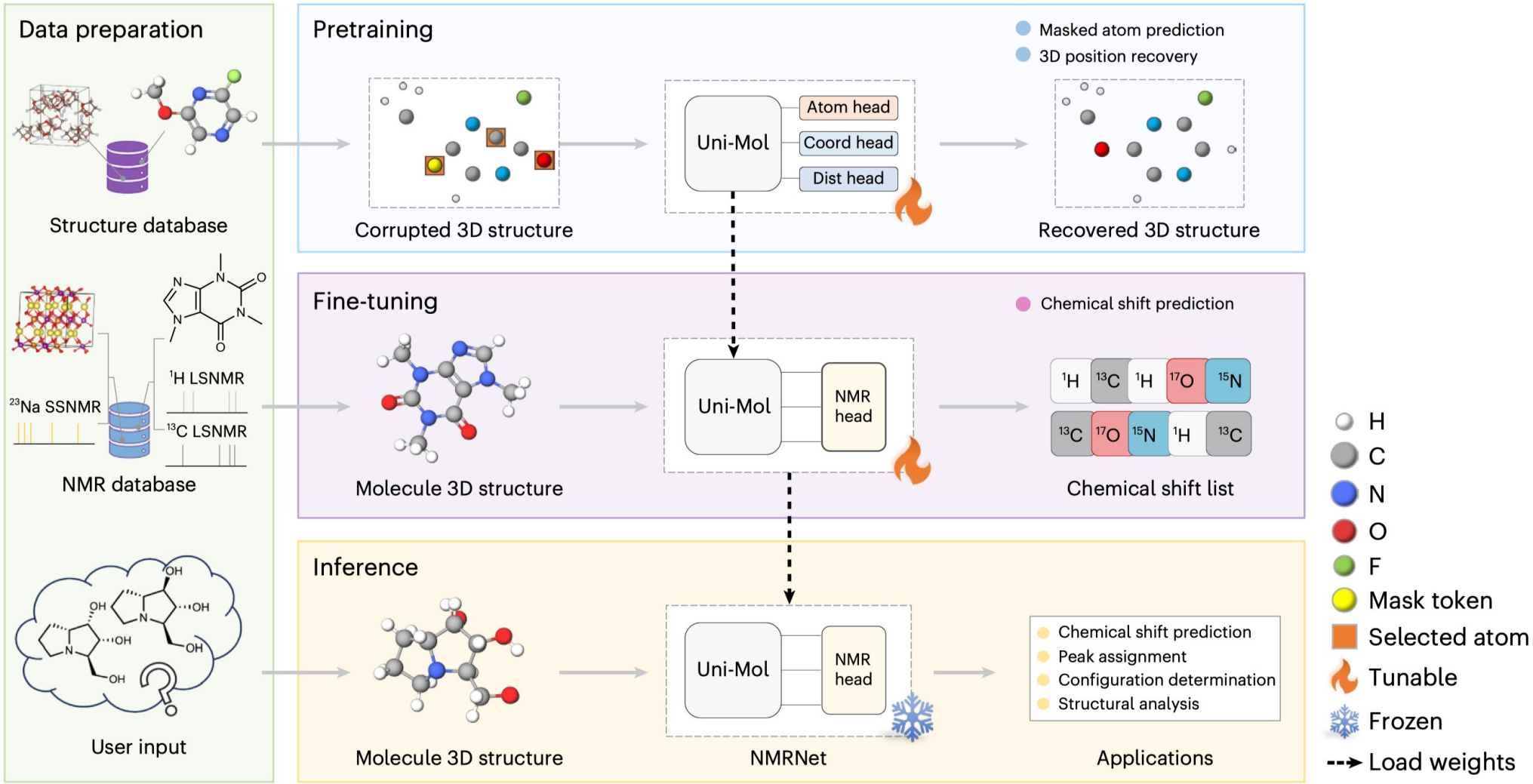
- **Unified framework**
- **Pre-training + fine-tuning**
- **Based on SE(3) Transformer architecture**

  A special GCN with coordinates

- **A standardized benchmark dataset**

2. NMRNet

Schematic diagram of the NMRNet framework



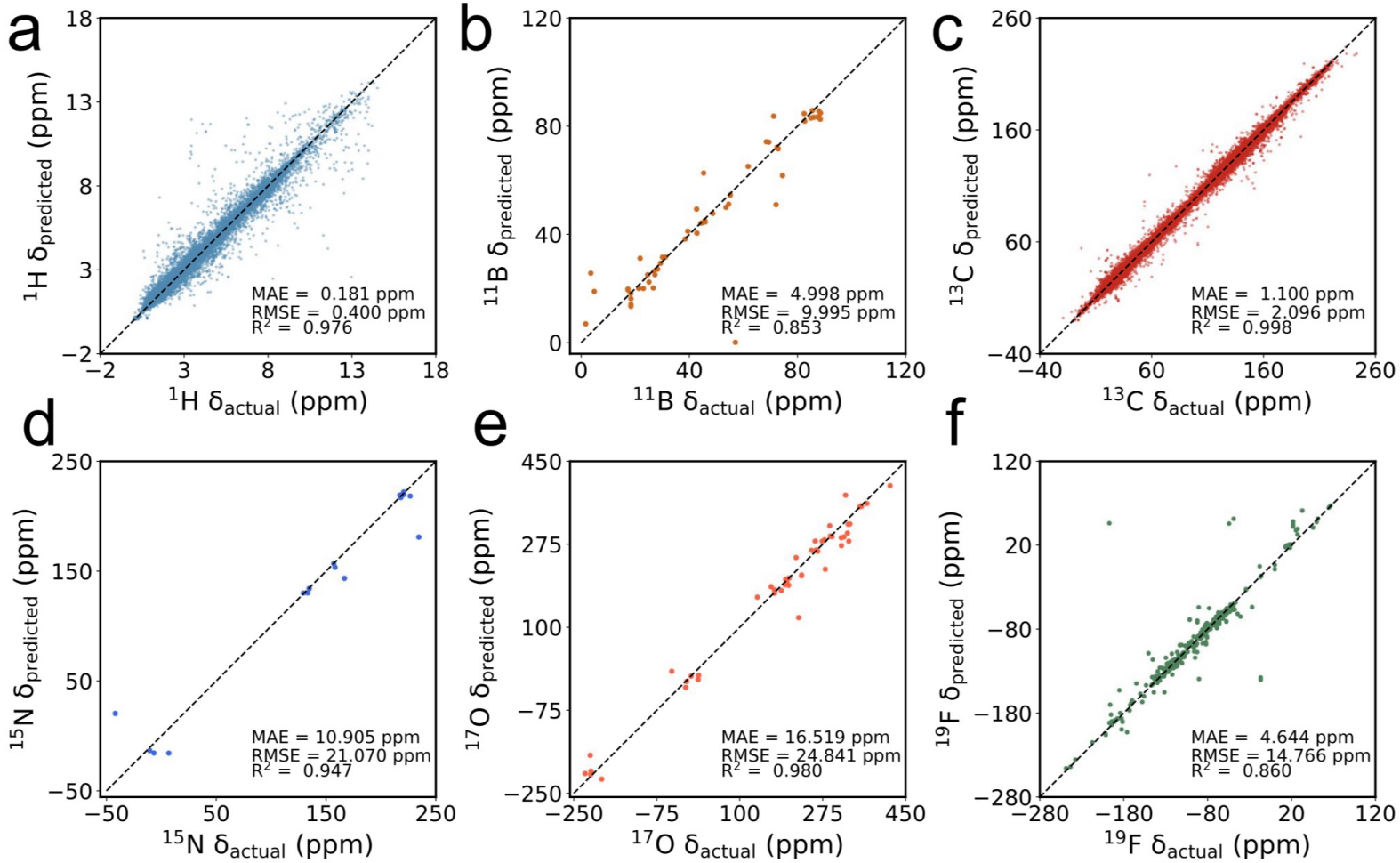
Fine-tuning NMRNet with liquid-state NMR data

Using manually modified nmrshiftdb2-2024 (Fine-tuning)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

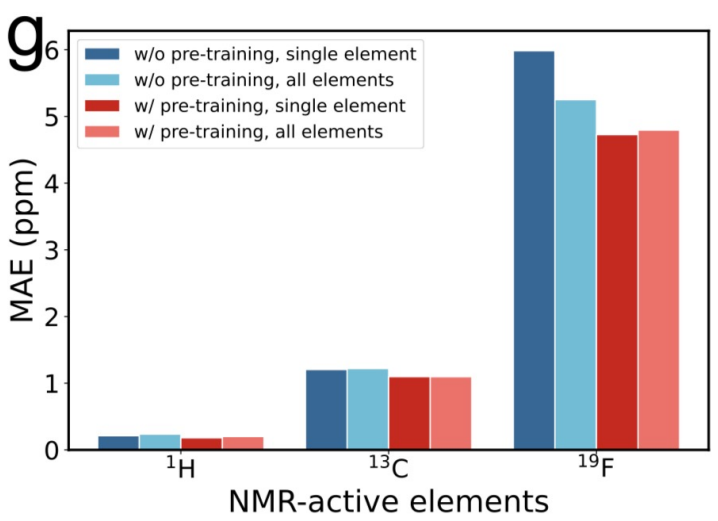
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

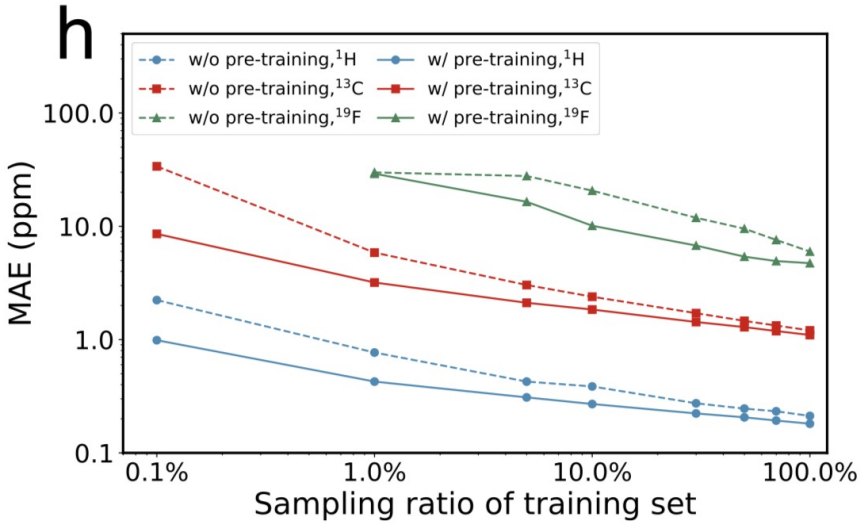
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$



Fine-tuning NMRNet with liquid-state NMR data

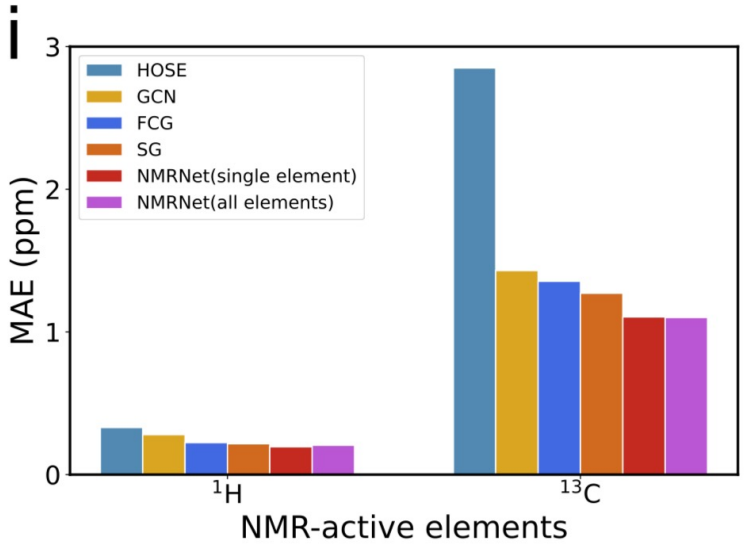
Pre-training

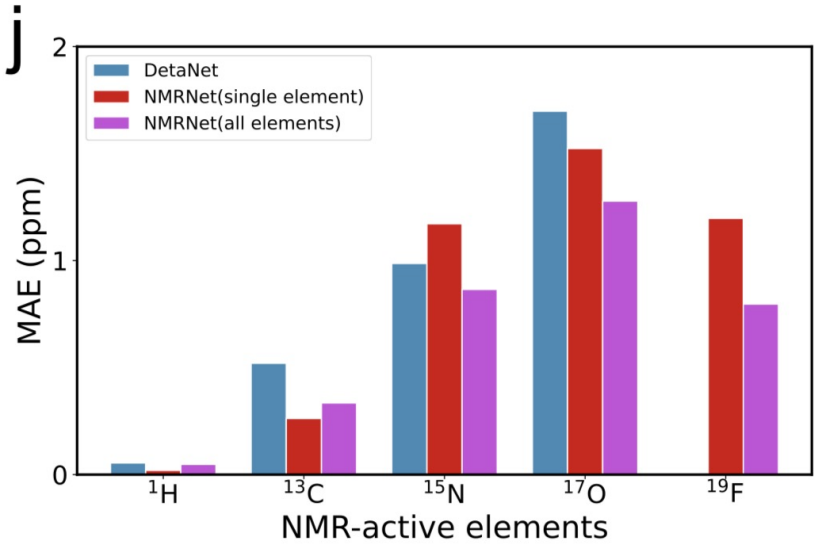




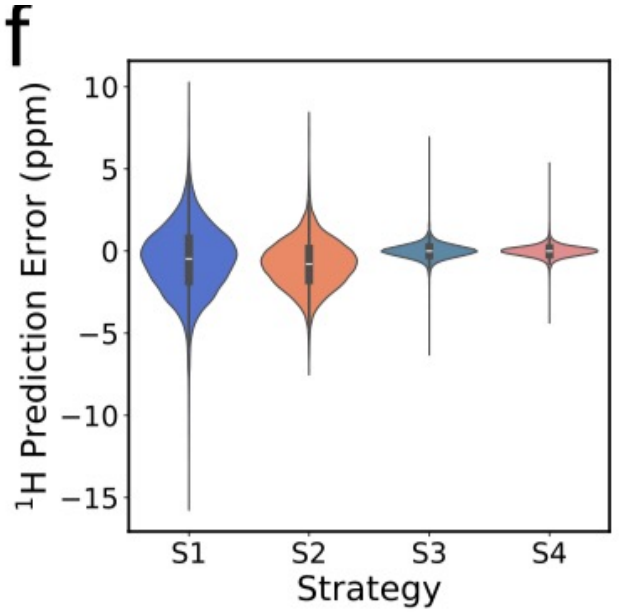
Data volume

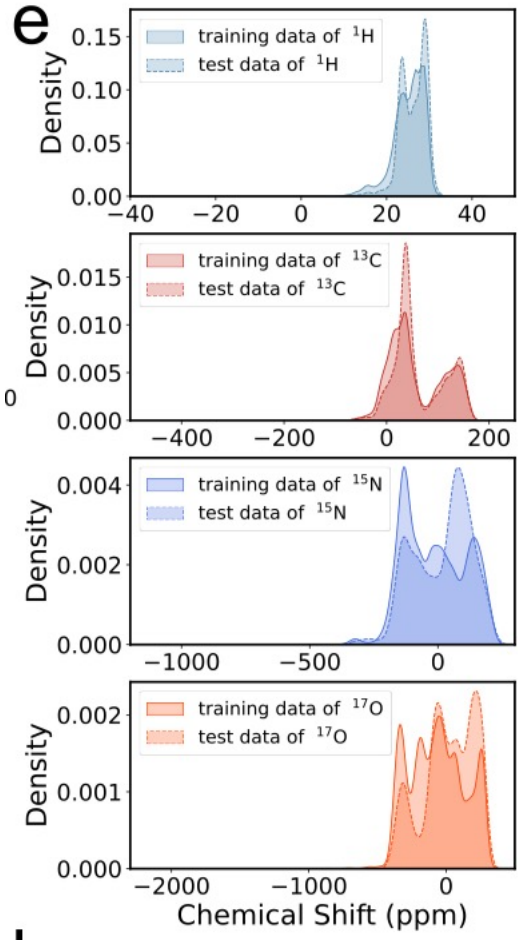
Comparisons

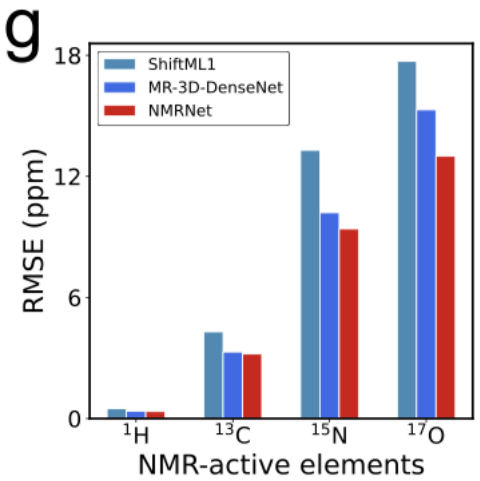




Fine-tuning NMRNet with solid-state NMR data

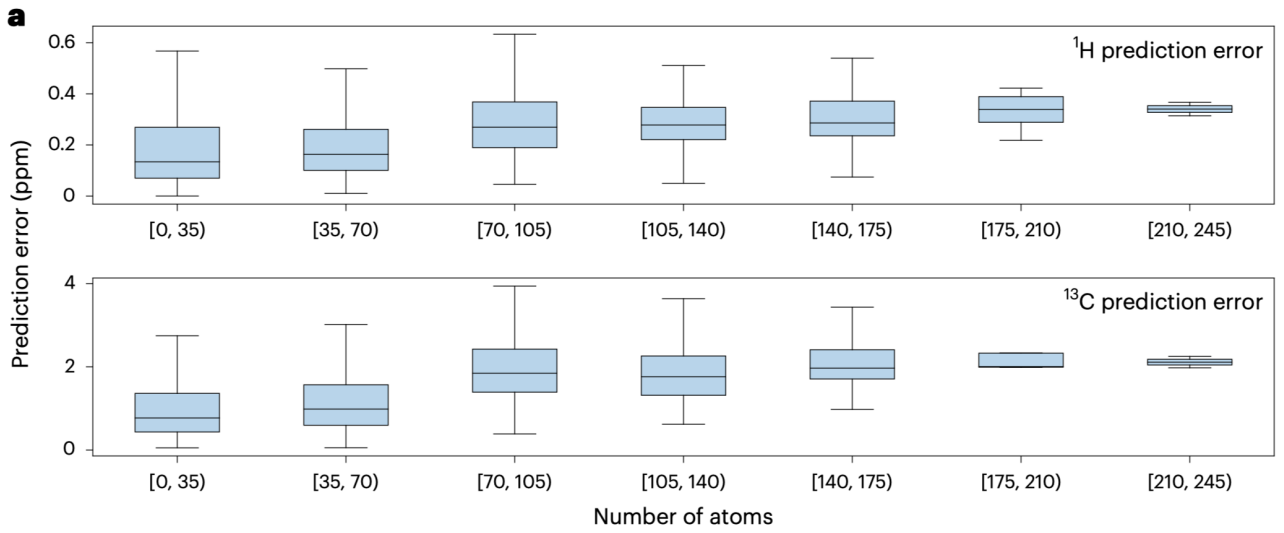


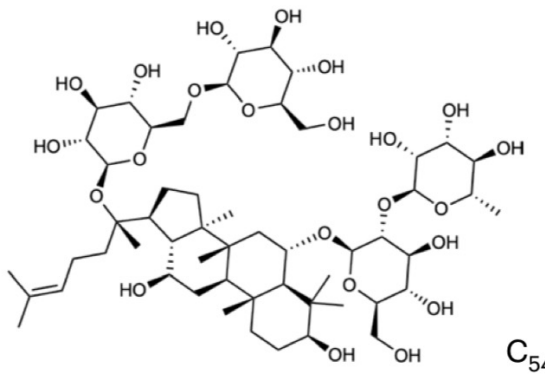




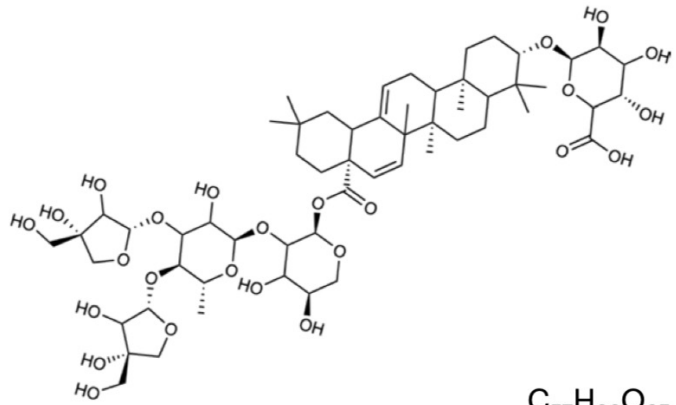
Applications of the fine-tuned NMRNet

Generalization of the NMRNet





$C_{54}H_{92}O_{23}$



$C_{57}H_{88}O_{25}$

Applications of the fine-tuned NMRNet

Peak assignment

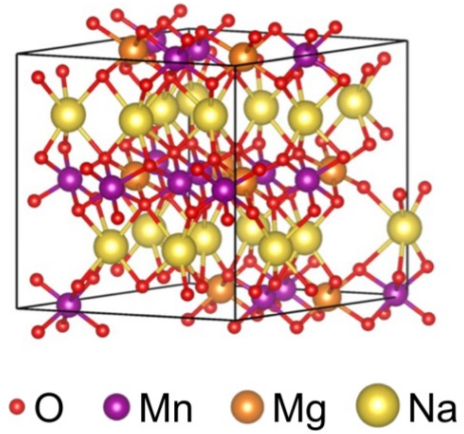
The accuracy is insufficient

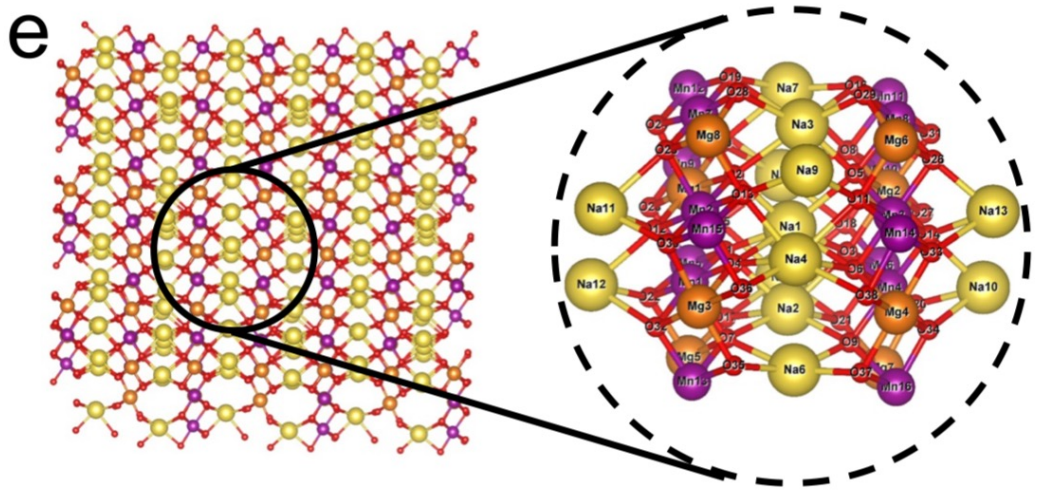
Configuration determination

Comparing the experimental chemical shift and the predicted shift (without coupling constant)

Correlation between NMR and the local environment

Infer the local structure through the chemical shifts of solid-state NMR.





3. Conclusion and Methods

Conclusion

- Hard to handle complex structures
- Highly Depend on databases
- Need to improve the accuracy ← Take solvent and temp. into consideration
- Hard to do Peak assignment ← Achieve the prediction of *J* at the same time
- Structures to Spectra Only

Methods

Data preparation process

RDKit
Atomic Simulation Environment
Python Materials Genomics

→ Abstract 3D information
→ InChIKeys are used as identifier for each molecule
→ Manually screen and remove the errors

NMRNet framework

The molecular domain is introduced by Uni-Mol (open-access)

Thank You