Data-Driven Atom-Level Explanation of Polymer Properties

Yuhao Liu, Jiahui Yang

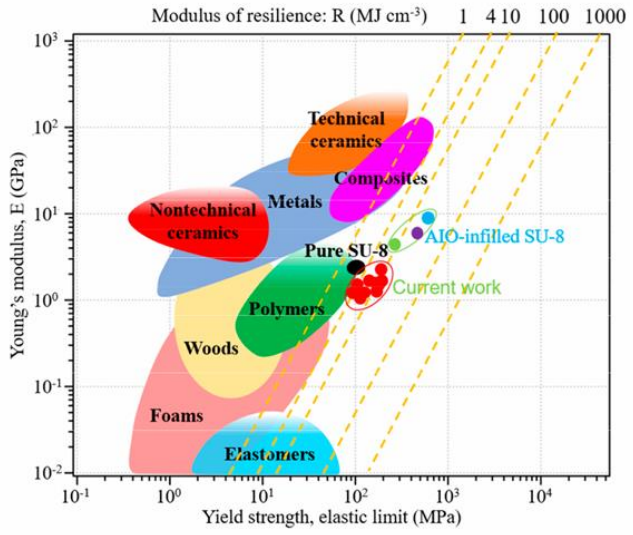
Department of Mechanical Engineering

University of Wisconsin – Madison

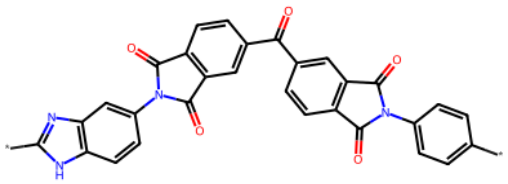
December 01, 2024

Contents

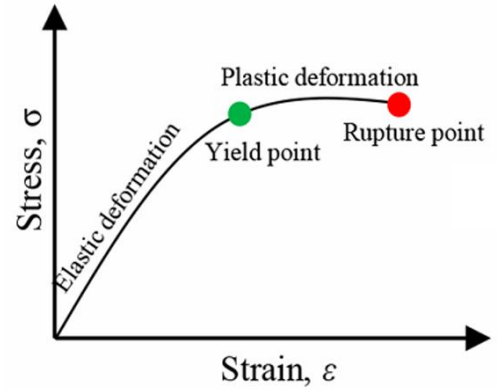
Introduction



- **Widespread Applications of Polymers**
- **Challenges:**
  - Complex molecular structures & chain configurations
  - Microstructure influences their mechanical properties

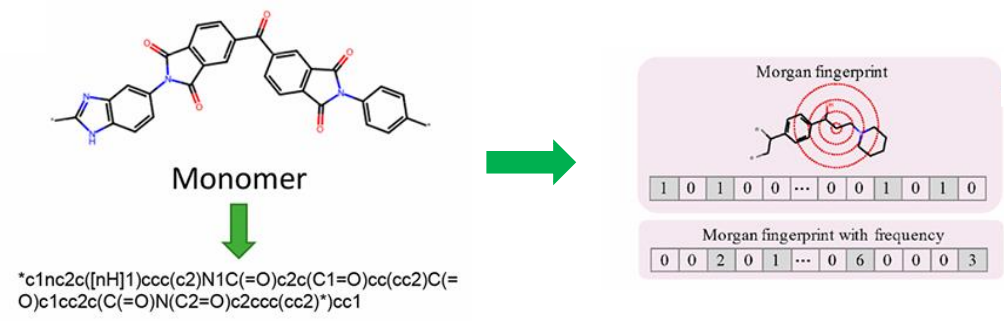


Monomer

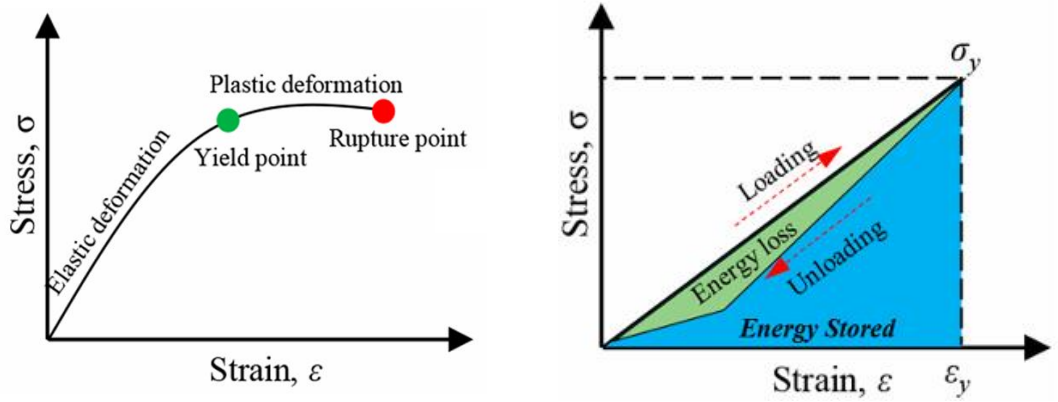


[1] Yue, Tianle, et al. Journal of Chemical Theory and Computation 19.14 (2023): 4641-4653.

Problem Statement

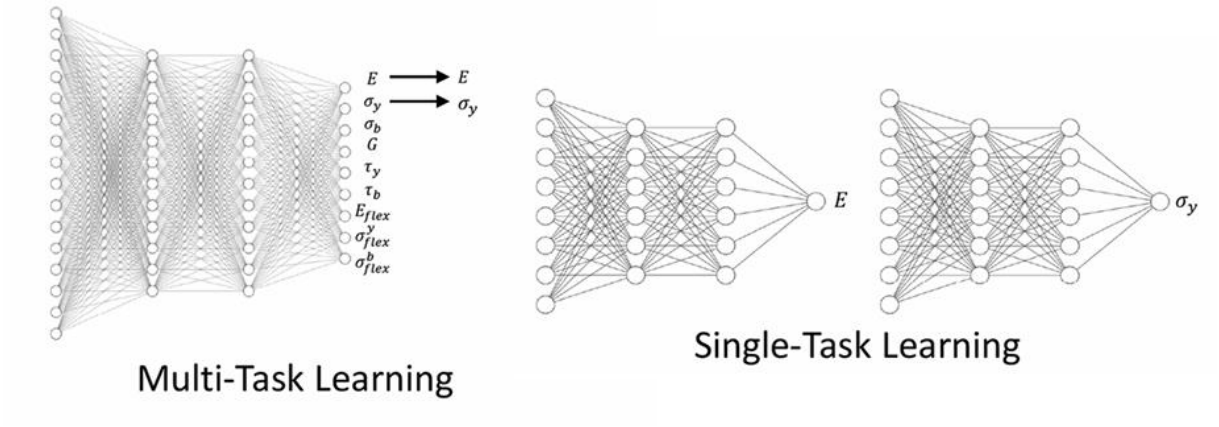


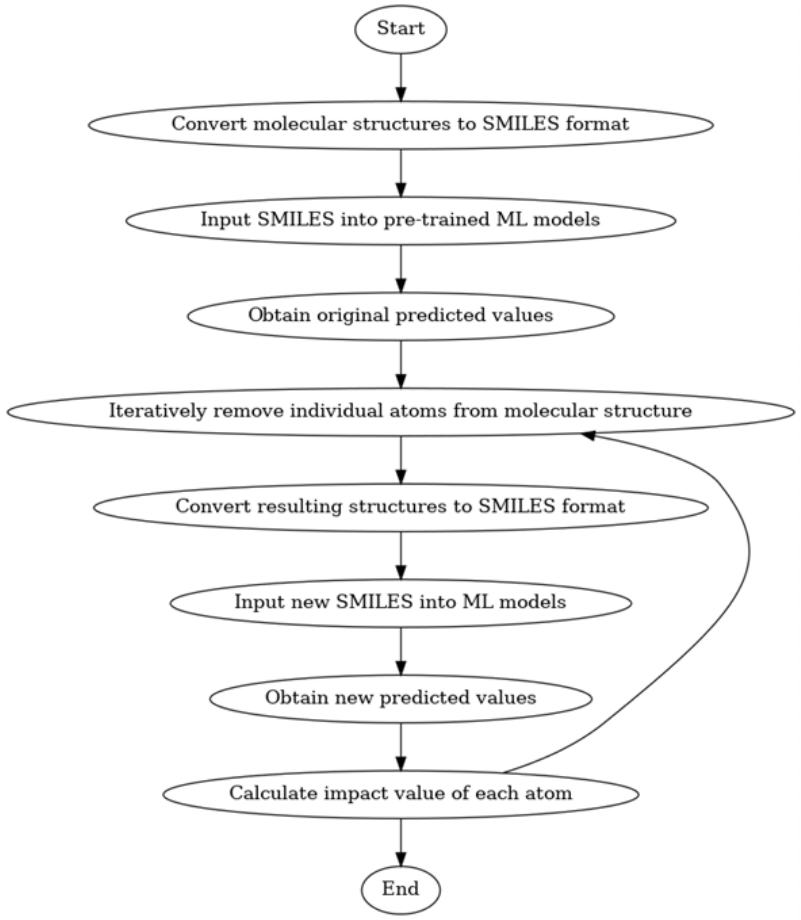
- **Machine Learning in Materials Science**
  - An essential tool for property prediction
  - Fail to capture the intricate details of polymer chains (topological information), such as SMILES
  - Require validation through other methods, such as MD simulations



Research Methods & Model Settings

A **single-task** machine learning model is employed to explore the relationship between polymer microstructures and macroscopic properties.



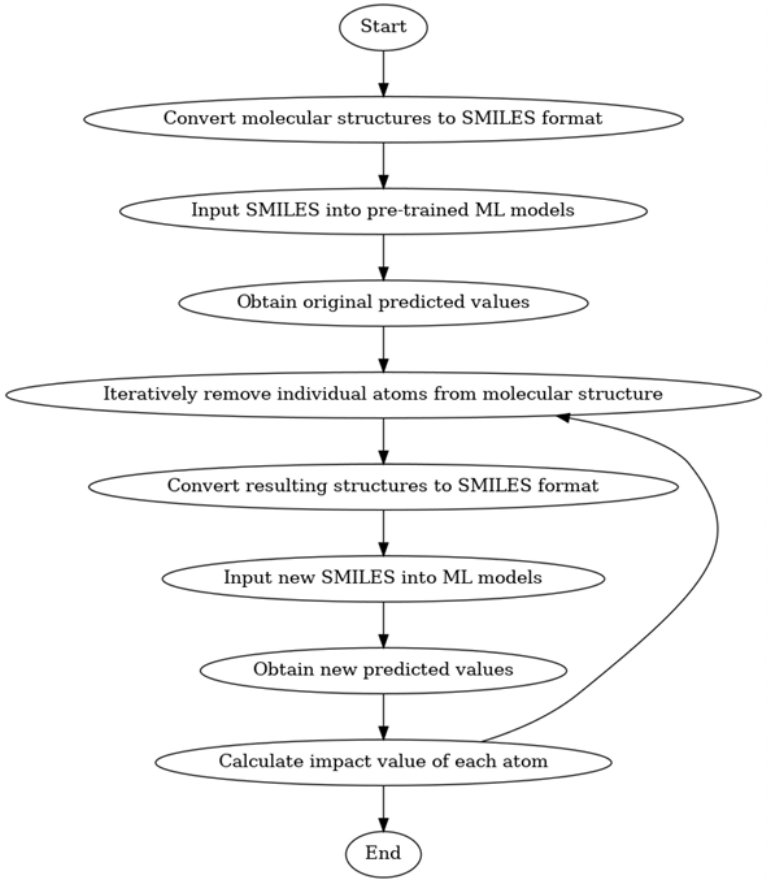


Research Methods & Model Settings

Data Source:

Data was sourced from **PoLyInfo** [2], a comprehensive database containing information on over 18,000 polymers, including 923 homopolymers with documented Young's modulus values.

SMILES [3] strings were used as input to the ML model to represent the chemical structure of polymers.

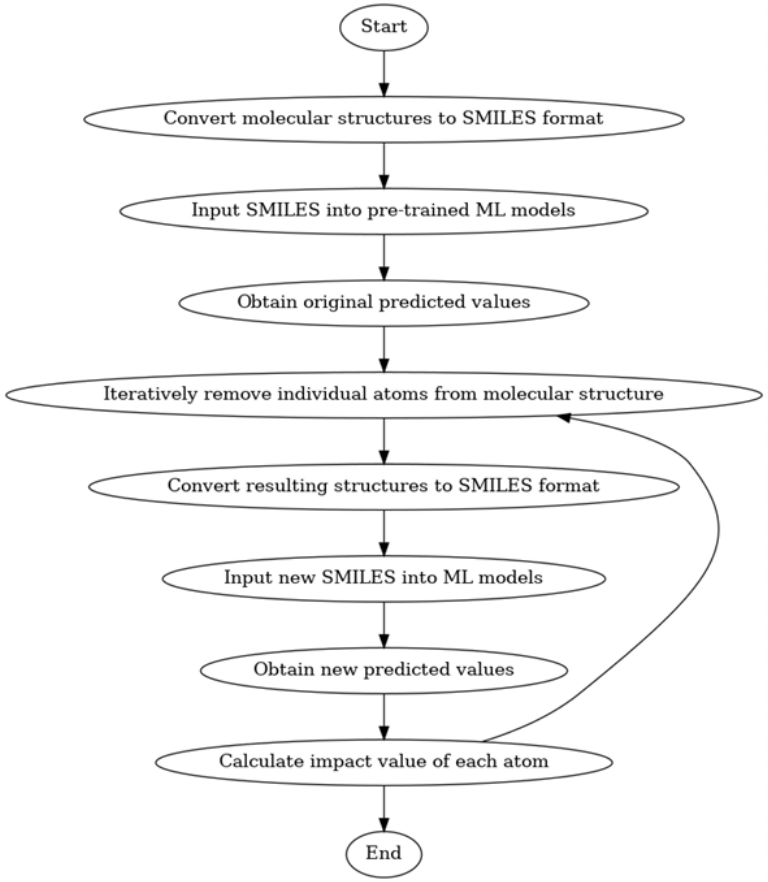


[2] Otsuka, Shingo, et al. 2011 International Conference on Emerging Intelligent Data and Web Technologies. IEEE, 2011.

[3] Weininger, David. Journal of chemical information and computer sciences 28.1 (1988): 31-36.

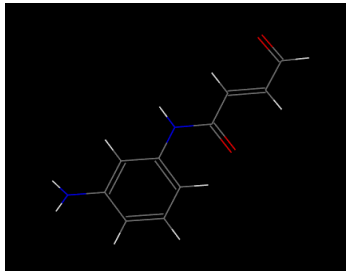
Research Methods & Model Settings

- Convert molecular structures into **SMILES**
- Train the ML model using **SMILES** and corresponding **Young's modulus values**
- Iteratively **remove atoms** and **recalculate predictions** to evaluate atomic-level contributions
- **SHAP** (Shapley Additive Explanations) values reveal substructure contributions to ML predictions
- **Atom-level explanations** map prediction changes from atom removal

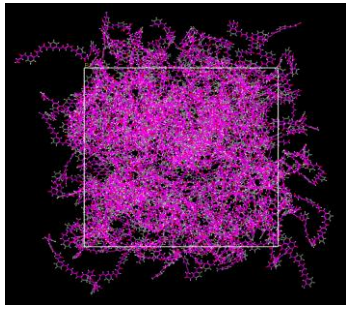


Research Methods & Model Settings

Step 1 Model Establishment (Materials Studio)



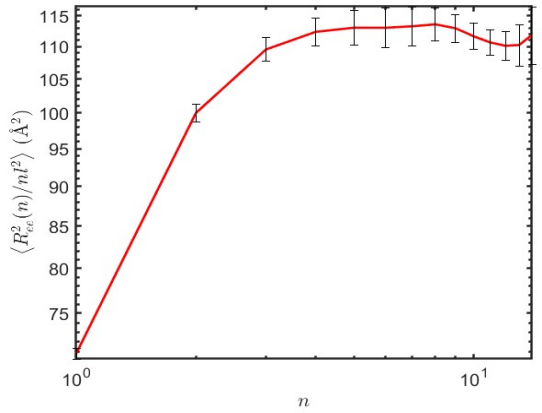
Monomer



Number of chains 80
Length of chains 15

Step 2 Relaxation (LAMMPS)

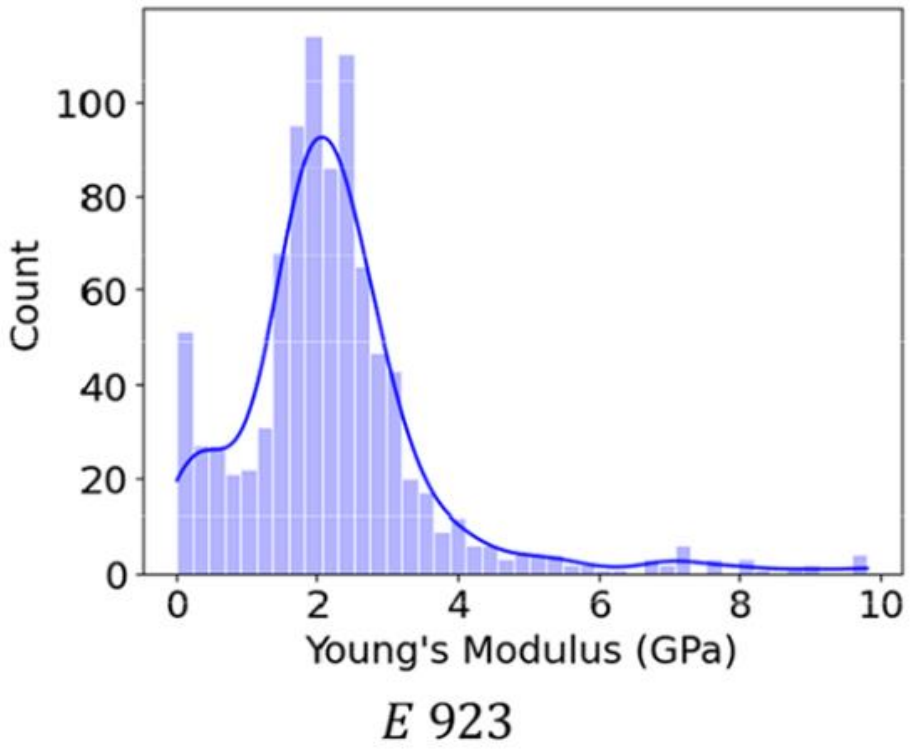
1. Relaxation
Assemble: npt
Temperature: 650K, Time step: 0.5 fs
2. Anneal
Assemble: npt
Temperature: 650K → 300K, Time step: 0.5 fs
3. Relaxation
Assemble: npt
Temperature: 300K, Time step: 0.5 fs

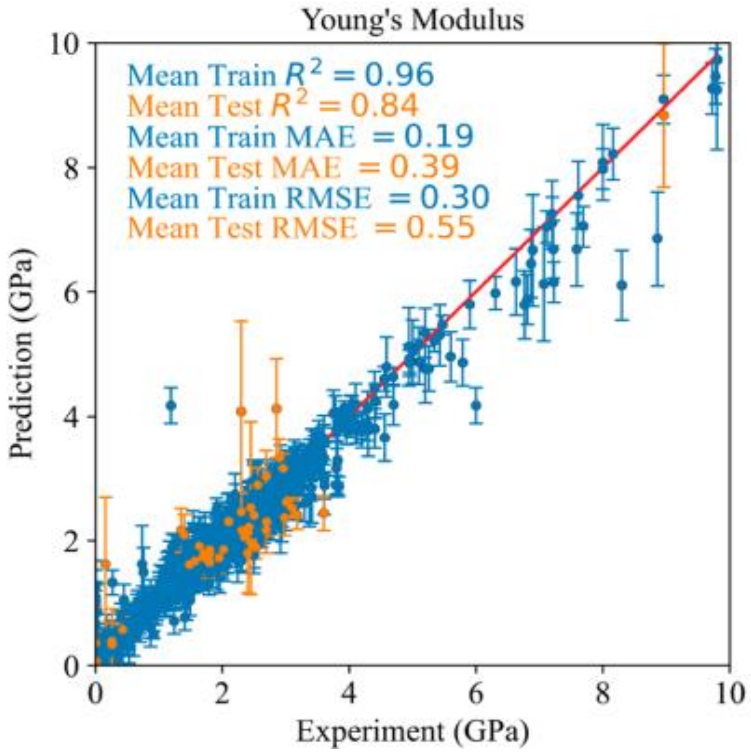


Step 3 Tension (LAMMPS)

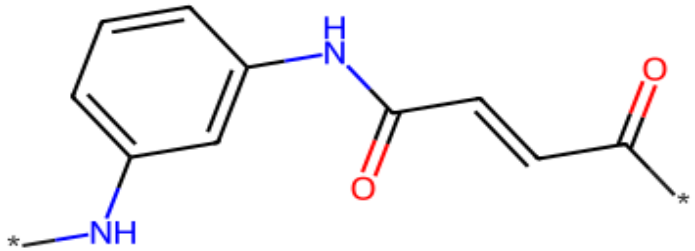
Assemble: nvt
Temperature: 300K
Time Step: 0.5fs
Potential : PCFF, LJ (9/6 types)
Engineering strain rate: $10^9 s^{-1}$

Results & Discussion





Results & Discussion



Sample ID:	**0041326-002-001-001**
Polymer ID:	**P402387**
Name:	**poly(imino-1,3-phenyleneiminobut-2-ene-1,4-dioyl)**
Polymer type:	**Homopolymer**
Polymer Class:	**Polyamides/thioamides**
CU formula:	**C10H8N2O2**
Formula weight(FW):	**188.18**
Characteristics of material:	**Shape and constructional features: linear**
Material type:	**Neat resin**
Polymerization informations:	**Reactant: maleic acid,m-phenylenediamine** **Types: polycondensation**

Thermal decompositions	
Thermal decomposition weight loss	ca5[%]
Thermal decomposition temperature	ca179[C]
Measurement conditions	Heating rate;10C/min ! gas;air
Measurement methods	TG
Remarks	Taken from the Fig.3

[4] Rabie, S. T.; Eid, M. A. M.,Journal of Vinyl & Additive Technology,13,4,206-214,2007,10.1002/vnl.20129

Results & Discussion

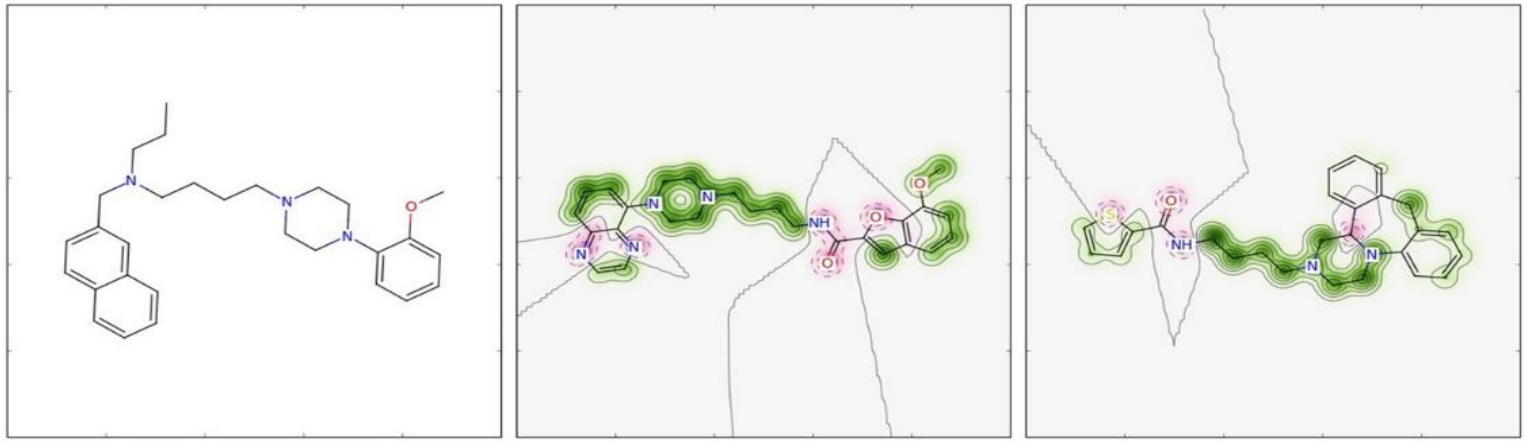


Figure 2 Similarity maps for atom-pairs (AP) fingerprint. Similarity map of molecule **2** (middle) and molecule **3** (right) using AP. The reference compound is molecule **1** (left). Color scheme: removing bits decreases similarity (i.e. positive difference) (green), no change in similarity (gray), removing bits increases similarity (i.e. negative difference) (pink). The default maximum path length of 30 was used for AP.

ref_fp = this_fp

Prediction property

weight=ML(ref_fp)- ML(new_fp)

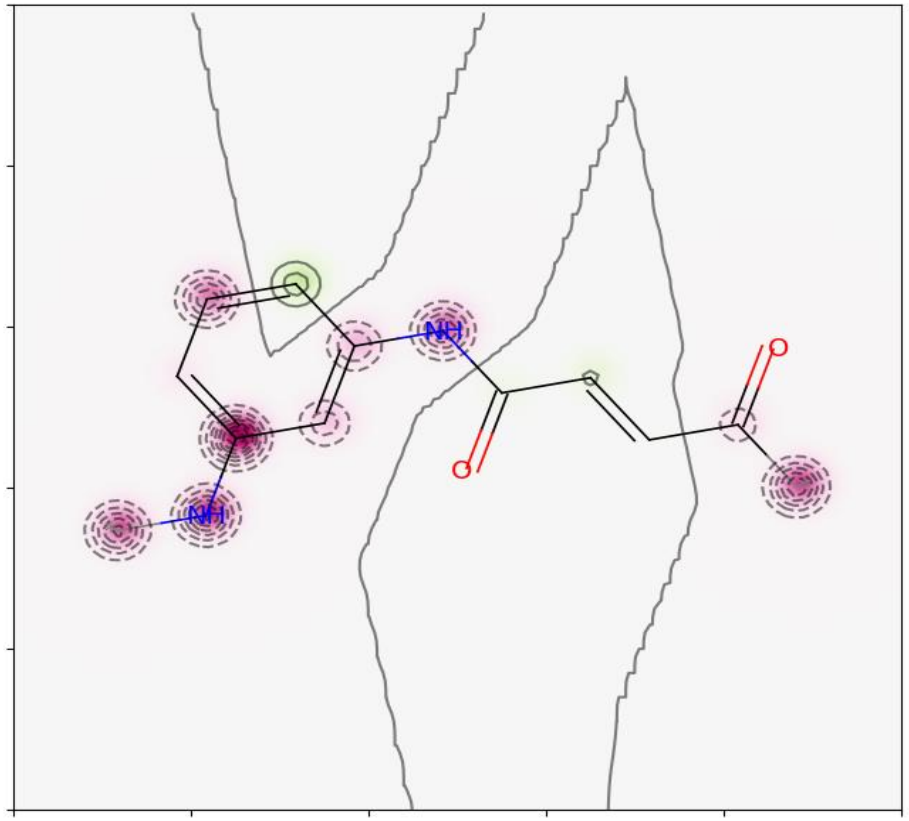
```
ref_fp = calculate_fingerprint(ref_mol)
this_fp = calculate_fingerprint(this_mol)
weights = []
orig_simil = dice_similarity(ref_fp, this_fp)
for atom in this_mol.get_atoms():
    new_fp = calculate_fingerprint_without_atom
             (this_mol, atom)
    new_simil = dice_similarity(ref_fp, new_fp)
    weight = orig_simil - new_simil
    weights.append(weight)
```

→ **single-task model**

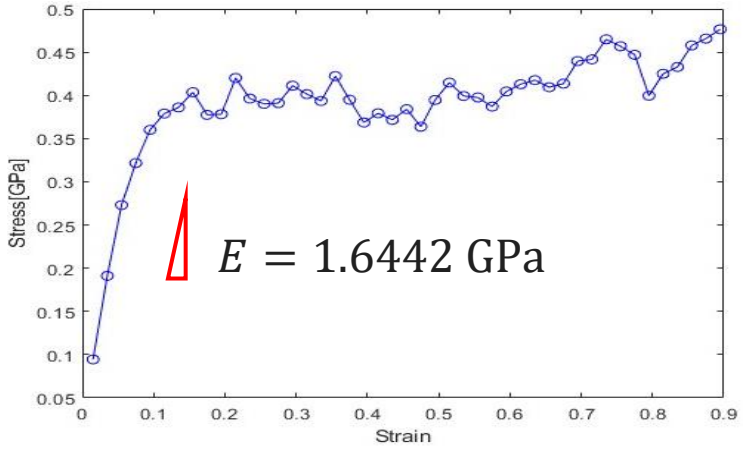
[5] Riniker, Sereina, and Gregory A. Landrum. Journal of cheminformatics 5 (2013): 1-7.

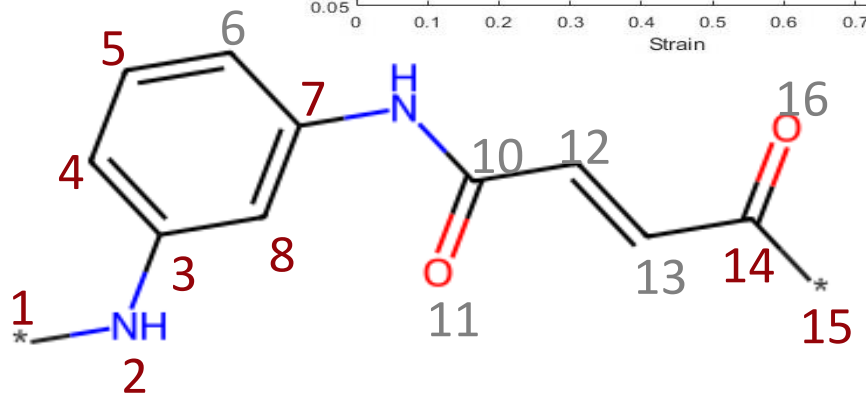
Results & Discussion

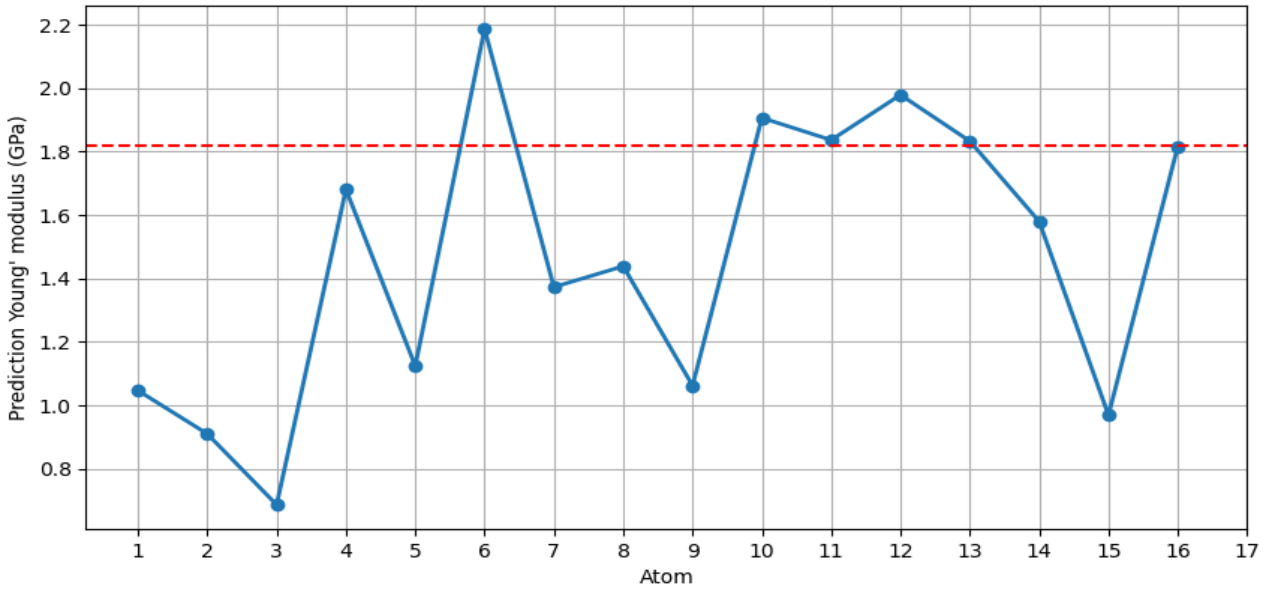
(a)



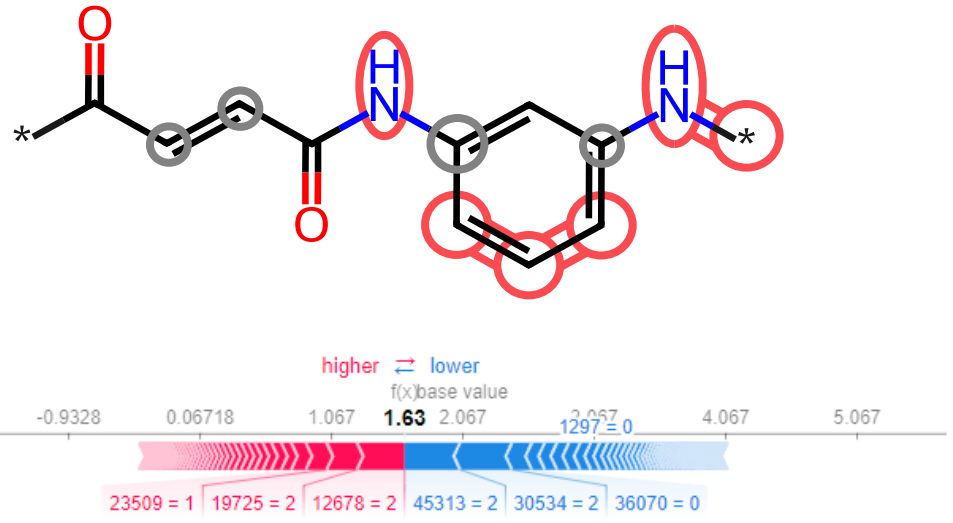
(b)

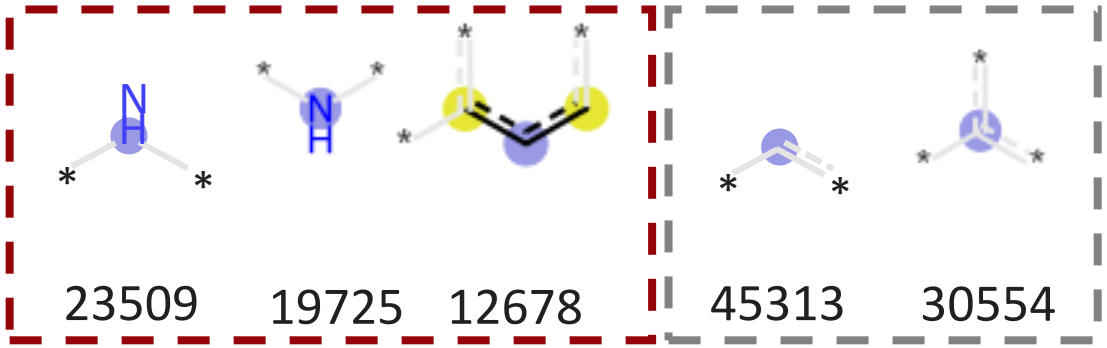




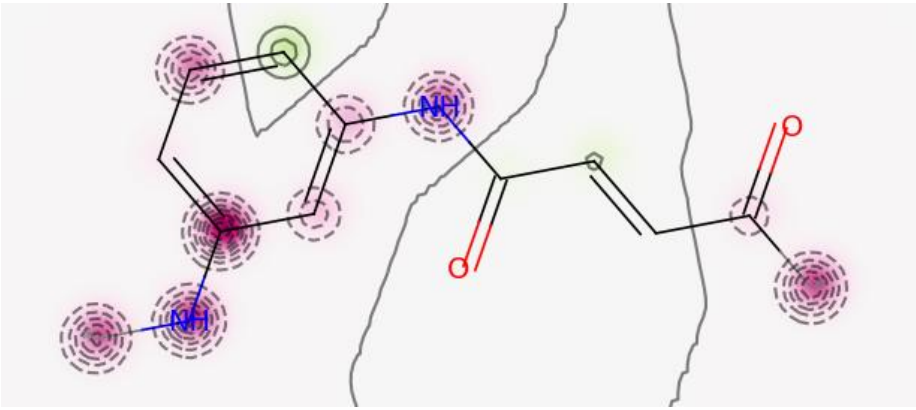


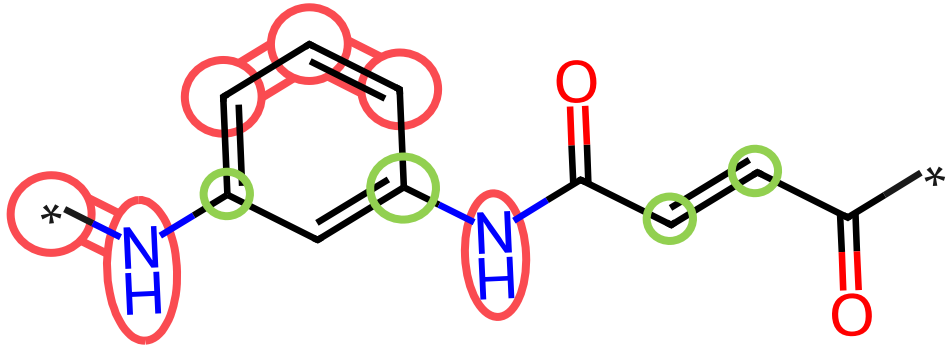
Results & Discussion(SHPB value)





Results & Discussion





Conclusion

- The **single-task neural network model** outperforms molecular dynamics simulation in predicting elastic modulus, with only an **8% prediction error**.
- **SHAP values** calculated using the single-task machine learning model reveal key impacts of polymer substructures on performance, though some findings are inconsistent with atom-level explanations.


WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON