

# 고분자 물성예측 및 역설계를 위한 인공지능·기계학습 연구 동향

## Recent Progresses in Artificial Intelligence and Machine Learning for Polymer Property Prediction and Inverse Design

하민영 | Min Young Ha

Department of Chemical Engineering, Kyung Hee University,  
Rm. 6038, Woojungwon, 1732 Deogyong-daero, Yongin, Gyeonggi 17104, Korea  
E-mail: myha@khu.ac.kr

### 1. 서론

인공지능(artificial intelligence, AI) 및 기계학습(machine learning, ML) 기법은 과학기술 전 분야에 걸쳐 패러다임 전환을 이끌고 있다. 단백질 3차원 구조 예측에서 정밀 실험 수준의 정확도를 달성한 AlphaFold<sup>1</sup>는 2024년 노벨화학상을 수상하며 AI 기반 과학·공학 연구의 가능성을 전세계에 각인시켰다. 최근에는 확산 모델(diffusion model)을 활용하여 원자 좌표, 원소 조성, 격자 구조를 동시에 생성하는 MatterGen<sup>2</sup>이 발표되어 신규 무기 결정 구조의 대규모 생성적 설계가 검증되기에 이르렀다.

소재과학 분야에서는 2011년 미국의 소재 게놈 이니셔티브(Materials Genome Initiative, MGI)<sup>3,4</sup>를 기점으로 소재 정보학(materials informatics)이 본격적으로 부상하였다. MGI에 힘입어 Materials Project 등 대규모 데이터베이스(DB) 구축 사업과 그에 기반한 ML 기반 물성예측이 활발히 수행되었으며, 특히 무기재료 분야에서는 범용 기계학습 힘장(universal machine learning force field)의 발전이 주목할 만하다. MACE,<sup>5</sup> CHGNet,<sup>6</sup> SevenNet<sup>7</sup> 등은 수백만 개의 제일원리 시뮬레이션 데이터로 훈련되어 주기율표상 대부분의 원소에 대응되는 범용적 원자간 퍼텐셜을 제공하며, 거대 언어 모델과 유사하게 훈련 데이터와 모델 크기가 증가할수록 예측 정확도가 체계적으로 향상되는 스케일링 거동(scaling behavior)을 나타내고 있다. 이를 편향-분산 상충관계(bias-variance tradeoff)의 관점에서 해석하면 무기재료 정보학은 방대한 표준 DB에 힘입어 저편향/고분산(low bias, high variance) 영역에 진입하고 있다고 볼 수 있다.

반면 고분자 정보학(polymer informatics)은 이와 대비되는 상황에 놓여 있다. 고분자의 화학적 공간은 단량체 구조, 중합 메커니즘, 선형·분지·가교 등 아키텍처, 분자량 분포, 전처리·후처리 조건, 복합소재 배합 등 연속형·범주형 변수가 혼재된 다차원·멀티스케일 구조를 갖기 때문에 무기 결정계에 비해 복잡도가 대단히 높고 표준화된 DB 관리체계를 정의하기 어렵다. 이러한 본질적 복잡성은 고분자 정보학에 있어 세 가지 핵심적 난점을 초래한다. 첫째, 고분자의 명명법·합성 이력·물성 측정 조건 등이 연구마다 상이한 탓에 데이터 통합이 용이하지 않으며, 결과적으로 표준화된 대규모 DB 구축이 어렵다. 둘째, 단량체 분자 구조 및 아키텍처 설계 등 상이한 그래프 자료 구조를 ML 모델이 인식할 수 있는 형태로 변환하는 표현 학습(representation learning) 문제가 존재한다. 셋째, 고분자의 최종 물성은 단량체·중합체 구조뿐 아니라 가공 온도, 경화 조건, 배합비 등 공정 변수에 크게 의존하므로 구조-물성 관계만으로 예측하는 데 한계가 있다.

Author



하민영

2016  
2022  
2025  
현재

서울대학교 화학생물공학부 (학사)  
서울대학교 화학생물공학부 (박사)  
삼성전자 SAIT 전문연구원  
경희대학교 화학공학과 조교수

결과적으로 고분자 정보학은 현재 세부 도메인별로 최적화된 소규모 모델에 의존하는 고편향/저분산(high bias, low variance) 단계에 머물러 있다고 진단할 수 있다. 본 원고에서는 상기한 난점에 대응하는 최근 연구 동향을 DB, 표현학습, 역설계 순으로 소개하고 향후 전망을 논의하고자 한다.

## 2. 본론

### 2.1 고분자 DB 현황

ML 기반 물성예측의 성능은 훈련 데이터의 품질과 규모에 크게 좌우된다. 본 섹션에서는 고분자 정보학의 기반이 되는 주요 공개 DB의 현황을 정리한다.

현존하는 가장 큰 실험 기반 고분자 물성 DB는 일본 물질·재료연구기구(NIMS)에서 운영하는 PoLyInfo<sup>8</sup>이다. PoLyInfo는 1995년 과학기술진흥기구(JST) 산하 프로젝트로 출범하여 2003년에 NIMS로 이관되었으며, 22,000편의 문헌에서 수집한 55만여 개의 데이터 포인트로 구성된다. 그러나 전체 DB의 30%가 10종의 고분자에 편중되어 있고 전체 물성치의 40%가 유리전이온도에 집중된 등 데이터 분포의 편중 문제가 제기된 바 있으며,<sup>9</sup> DB 운영 규약상 개별 데이터 포인트의 검색만 허용되고 대규모 다운로드가 금지되어 있어서 ML 모델 학습에 활용하기는 어렵다.

대규모 실험 DB의 구축에는 막대한 비용과 시간이 소요되며 접근할 수 있는 표현 영역이 제한되기 때문에 시뮬레이션 및 가상 고분자 구조로 이를 보강하려는 시도가 활발하다. Bradley Olsen 연구팀에서 발표한 블록공중합체 상거동 데이터베이스(BCDB)<sup>11</sup>는 문헌 및 장이론 계산 결과를 종합하여 5,400여 건의 상평형 데이터를 공개하였으며, Rampi Ramprasad 연구팀에서 제일원리 계산 결과 위주로 구성된 Khazana DB<sup>12</sup>는 고분자 물성예측을 수행하는 Polymer Genome<sup>13,14</sup> 웹 플랫폼의 학습 자료로 활용된 바 있다. 일본의 RadonPy 컨소시엄에서는 고분자의 분자동역학(MD) 시뮬레이션을 자동 수행하는 RadonPy 라이브러리<sup>15</sup>와 후가쿠(Fugaku) 슈퍼컴퓨터를 활용하여 700만 개 이상의 시뮬레이션 데이터로 구성된 PolyOmics DB<sup>10</sup>를 발표하였다(그림 1). 한편 Tengfei

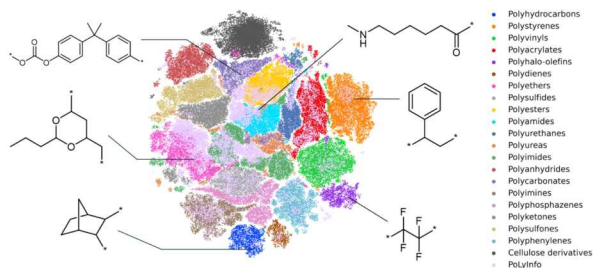


그림 1. PolyOmics DB의 2차원 평면상의 투영 결과. 다양한 고분자 카테고리들이 색깔별로 표현됨.<sup>10</sup>

Luo 연구팀에서는 12,000개의 고분자 구조를 학습한 분자 생성 모델을 활용하여 100만여 개의 가상 고분자 구조를 생성한 P11M DB<sup>16</sup>를 발표하였으며, 이는 TransPolymer<sup>17</sup> 등 후속 모델의 학습에 활용된 바 있다.

### 2.2 고분자 표현학습

ML 모델이 고분자의 구조-물성 관계를 학습하기 위해서는 고분자의 화학 구조를 수치적 입력으로 변환하는 표현학습(representation learning)이 선행되어야 한다. 이 과정은 크게 두 단계로 구분된다. 첫째, 고분자의 화학 구조를 기계가 읽을 수 있는 형태로 기술하는 문자열 표현을 정의하고, 둘째, 이 표현으로부터 ML 모델에 직접 입력되는 수치적 기술자(descriptor) 또는 지문(fingerprint)을 산출하는 것이다. 본 섹션에서는 문자열 표현, 분자 기술자, 그래프 신경망 기반 접근의 세 범주로 나누어 최근 연구 동향을 소개한다.

소분자 화학에서는 전통적으로 분자 구조를 SMILES(Simplified Molecular-Input Line-Entry System)으로 표기하는데, 이는 대부분의 ML 모델이 다루기 어려운 문자열 구조를 갖기 때문에 SMILES 문자열로부터 고정 길이의 수치 벡터를 산출하는 분자 기술자 및 지문이 널리 활용되었다. 화학 정보학에서 확립된 ECFP(Extended-Connectivity FingerPrint) 분자 지문이나 Mordred 기술자 등이 고분자에도 널리 적용되고 있으며, 특히 단일중합체(homopolymer)의 경우 단량체의 SMILES로부터 지문을 계산하여 ML 모델의 입력으로 사용하는 방식이 일반적이다. 2025년 2,000여 팀이 참여한 Kaggle Open Polymer Challenge<sup>18</sup>에서 상위 솔루션들이 ECFP 등 전통적 분자 지문과 의사결정 트리의 조합으로 심층학습 모델을 능가한 사실은, 현재 고분자 데이터 규모에서 수작업 기술자 기반 접근이 여전히 강력한 기준선임을 시사한다.

한편, 반복 단위의 연결 방식, 다분산성(polydispersity), 입체규칙성(tacticity), 공중합 시퀀스 등의 통계적 특성이나 구조적 복잡성을 표현해야 하는 고분자 소재에는 SMILES 표현법의 한계가 존재하기 때문에 고분자 자료형에 특화된 문자열 표기법 연구 역시 활발히 진행 중이다. MIT에서 제안한 BigSMILES<sup>19</sup>는 확률적 객체(stochastic object)를 도입하여 고분자의 반복단위 연결 패턴, 분자량 분포, 서열 통계를 기계가 읽을 수 있는 형태로 인코딩하였으며, 이후 생성형 워크플로우에 대응하기 위한 G-BigSMILES로 확장<sup>20</sup>하여 고분자 사슬 배열의 확률적 샘플링을 지원하였다. 국내에서는 허수미 교수 연구팀에서 고분자의 계층적 구조를 반영한 문자열 표현인 HAPPY(Hierarchically Abstracted rePeat unit of PolYmers)를 제안<sup>21</sup>하였는데, 고분자 표현에서 자주 등장하는 작용기에 단일 문자를 할당하여 SMILES 대비 문자 길이를 대폭 축소함으로써 학습 속도와 물성예측 성능을 함께 제고하였다(그림 2). 동 연구팀은 최근에

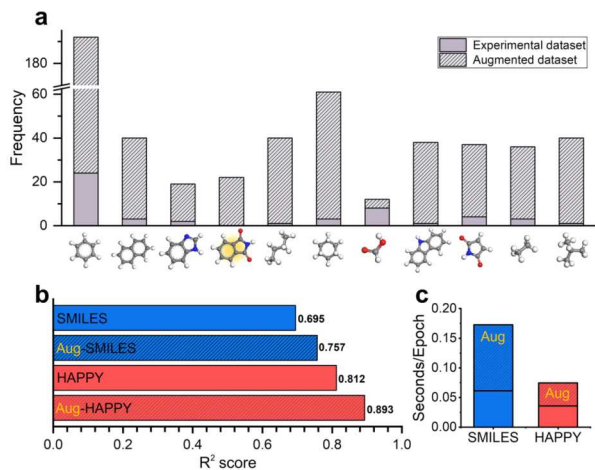


그림 2. 고분자 구조에 특화된 HAPPY 문자열 표현의 효과. (a) 학습 데이터 내의 작용기 분포. (b,c) HAPPY 표현 도입시 기존의 SMILES 표현에 비해 고분자 물성 예측의 결정계수( $R^2$ )가 증가하고 에포크 당 학습에 필요한 시간이 감소함.<sup>21</sup>

HAPPY 표현을 트랜스포머 아키텍처에 적용한 화학 정보 언어모델(CI-LLM)을 발표<sup>22</sup>하여 고분자에 특화된 문자열 표현이 생성형 역설계 단계에서의 스케폴드 유지율과 추론 성능 개선에 모두 기여할 수 있음을 시연하였다.

그래프 신경망(graph neural network, GNN)은 분자를 원자(node)와 결합(edge)으로 구성된 그래프로 표현하고, 결합을 통해 화학적 메시지를 전달함으로써 구조적 특징을 자동으로 학습하는 심층학습 아키텍처이다. 전술한 분자 지문과 달리, GNN은 사람이 설계한 기술자 없이도 목표 과업에 최적화된 분자 표현을 데이터로부터 직접 학습할 수 있다는 장점이 있다. 다만 고분자 전체 사슬을 그래프로 직접 표현하기는 어렵기 때문에 고분자에 GNN을 적용할 때는 중합체 수준의 연결망 정보를 효과적으로 인코딩하는 것이 관건이다.

삼성전자 연구팀에서는 그래프 합성곱 신경망(graph convolutional network, GCN)을 단량체 그래프에 적용하여 단일중합체의 유리전이온도( $T_g$ ), 녹는점( $T_m$ ), 탄성률(elastic modulus)을 예측<sup>23</sup>한 바 있다. 흥미롭게도 GCN은  $T_g$ 에 대해서 0.90의 높은 결정계수( $R^2$ )를 보였으나 탄성률 예측 결정계수는 0.54에 불과하여 기계적 물성 예측에는 분자 구조 정보를 넘어 공정·모폴로지 정보의 통합이 필요함을 시사하였다. 중합체 수준의 정보를 반영한 사례로 Connor Coley 연구팀에서는 사슬 아키텍처, 단량체 양론비, 중합도를 포괄적으로 인코딩하는 기법을 제시<sup>24</sup>하였으며, 다중작업 GNN을 활용하여 선형·분지형 폴리에스테르의 구조를 학습할 수 있음을 확인<sup>25</sup>한 사례도 보고되었다. 한편 고려대학교 김용주 교수 연구팀에서는 조립자 분자동역학(coarse-grained MD) 시뮬레이션 데이터로부터 병솔형 공중합체(bottlebrush copolymer)의 상거동을 학습·예측하는 GCN 모델을 구축<sup>26</sup>

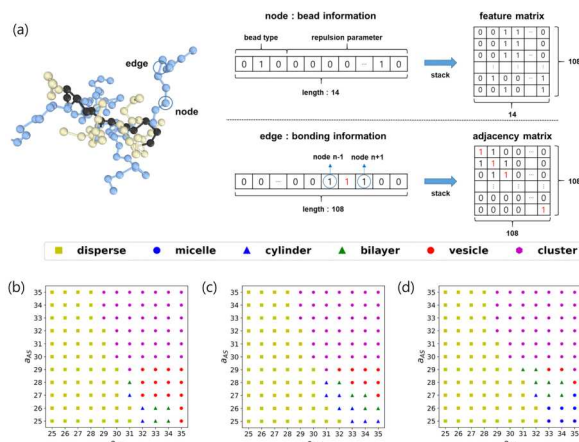


그림 3. (a) 병솔형 고분자의 상거동 예측을 위한 그래프신경망 구성 도해. (b-d) 측쇄 길이 변화에 따른 상평형도 천이.<sup>26</sup>

함으로써 그래프 신경망이 복잡한 분지 고분자의 위상 구조를 효과적으로 부호화할 수 있음을 입증하였다(그림 3).

각 접근법은 고유한 장단점을 지닌다. 수작업 기술자/지문 기반 접근은 고분자의 화학적 특성에 대한 귀납적 편향(inductive bias)을 부여하기 때문에 해석 가능성이 높으며 소규모 데이터에서도 안정적이나, 기술자를 정의했던 초기 도메인과 상이한 소재군에 적용할 때 성능이 열화될 우려가 있다. GNN 기반 접근은 귀납적 편향에서 자유롭기 때문에 목표 과업에 최적화된 구조적 특징을 유연하게 학습할 수 있으나 충분한 훈련 데이터를 필요로 한다. 궁극적으로 단량체 화학 구조, 서열 및 분지 패턴, 분자량 분포, 상위 구조 및 모폴로지 등 고분자의 다양한 계층 구조를 포괄적으로 인코딩할 수 있는 표현의 개발이 핵심 과제이다.

### 2.3 최적 고분자 역설계

고분자 정보학의 양대 축은 고분자 표현을 기반으로 물성을 예측하는 순방향 예측(forward prediction) 모델과, 목표 물성으로부터 최적 분자 구조 및 공정 조건을 탐색하는 역설계(inverse design) 모델로 구성된다. 현재 주목받는 역설계 방법론은 크게 베이저안 능동학습과 생성 모델 접근으로 나눌 수 있다.

베이저안 능동학습(Bayesian active learning) 혹은 베이저안 최적화(Bayesian optimization)는 고정된 후보 공간에서 예측 불확실성을 정량화함으로써 후보 소재·공정에서의 잠재적 정보 이득을 평가하고 후속 실험을 추천함으로써 최소한의 평가 횟수로 최적 조건을 탐색하는 순차적 최적화 기법이다(그림 4). 능동학습 접근은 특히 데이터가 부족한 연구 초기 상황에서 탐색 가속 효과가 뛰어난 것으로 알려져 있으며, 기존 연구는 대부분 무기소재 최적 탐색에 집중<sup>27-29</sup>되어 왔으나, Ramprasad 연구팀에서는 DB 내의 소규모 부분집합에서 출발하여  $T_g$ 가 높은 고분자를 탐색하는 능동학습 가상 실험을

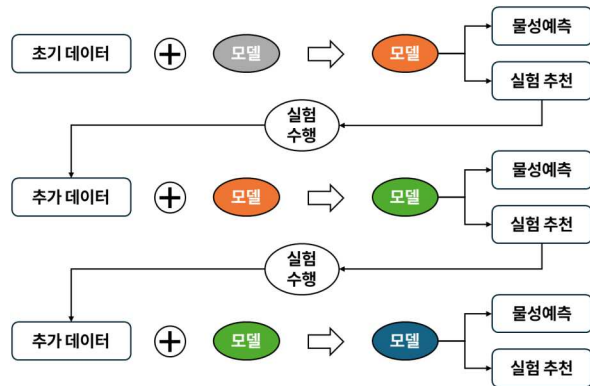


그림 4. 능동학습 반복 워크플로우의 도해. 불확실성 정량화에 근거하여 후속 실험을 추천하고 이를 수집하여 모델을 점진적으로 업데이트함.

수행하여 고분자 자료형에 대해서도 가우스과정 기반 능동 학습이 탐색 가속 효과가 있음을 실증<sup>30</sup>하였다. 한편 일본 NIMS에서는 분자량, 혼합비, 경화 온도에 따른 에폭시 접착제의 기계적 특성을 예측하고 후속 실험을 지정하는 능동학습 모델을 개발함으로써 접착 강도를 극대화하는 사례를 발표<sup>31</sup>한 바 있다.

생성 모델은 대규모 데이터를 학습한 심층신경망이 데이터 분포를 학습하여 기존 후보를 넘어선 신규 분자를 생성하는 접근법으로 최근 다양한 목적물성과 아키텍처에 대한 연구가 진행 중이다. Ramprasad 연구팀에서는 변분 오토인코더 (variational autoencoder) 모델을 활용하여 열적 안정성이 뛰어난 고분자를 생성<sup>32</sup>하였고, 도요타 연구팀은 GPT 및 확산 기반 생성 모델로 고분자전해질 구조를 생성하고 분자 동역학 시뮬레이션으로 검증한 결과 기존 벤치마크에 비해 우월한 이온 전도도를 갖는 고분자 17종을 생성함을 보고<sup>33</sup>하여 생성 모델 접근의 잠재력을 확인하였다.

### 3. 결론

이상에서 살펴본 바와 같이 고분자 정보학은 데이터베이스 구축, 표현학습, 물성예측 및 역설계의 전 분야에서 의미 있는 진전을 이루고 있으나, 방대한 고분자 설계공간에 대응할 수 있는 표준 스키마 및 DB가 부재하기 때문에 도메인·물성별로 개별 최적화된 소규모 모델에 의존하는 고풍향/저분산 단계에 머무르는 것으로 보인다. 이러한 한계를 극복하기 위한 전략으로 다음의 방향이 주목받고 있다.

첫째, 전이학습 (transfer learning)은 데이터가 풍부한 도메인에서 학습한 지식을 데이터가 부족한 고분자 물성 예측에 전달하는 접근으로, PolyOmics 등 시뮬레이션 기반 대규모 DB가 존재할 경우 활용이 용이한 기법이다. NIMS 연구팀은 XenonPy 라이브러리<sup>34</sup>를 통해 140,000개 이상의 사전학습 인공신경망을 제공하여 수십 개 수준의 소규모

데이터에서도 유효한 교차 도메인 전이학습이 가능함을 보였으며, 동 연구팀은 고분자 열전도도와  $T_g$  데이터에 대하여 보간 (interpolation)이 아닌 외삽 (extrapolation) 영역에서도 예측 성능을 향상시킬 수 있음을 시연<sup>35</sup>하였다.

둘째, 다중충실도 학습 (multi-fidelity learning)은 계산 비용이 낮은 저충실도 (low-fidelity) 데이터와 정확도가 높은 고충실도 (high-fidelity) 데이터를 체계적으로 결합하는 전략이다.<sup>36</sup> 현재 다중충실도 학습은 주로 계층적 시뮬레이션 데이터에서 시도되고 있는데, 고분자 밴드갭 예측 사례에서는 모든 시나리오에서 다중충실도 학습 성능이 우수함이 보고된 바 있다. 향후 고분자 분야의 저충실도 시뮬레이션과 고충실도 실험 데이터의 융합에 적용하여 소규모 실험 데이터에서의 예측 성능을 제고하는 연구가 필요할 것으로 보인다 (그림 5).

셋째, 물리 기반 사전지식의 통합 (physics-informed ML)은 알려진 물리 법칙이나 경험적 관계식을 모델 구조에 명시적으로 반영하는 접근이다. 고분자 유변학 분야에서는 점도 방정식 (viscosity equation)을 신경망 구조에 인코딩한 물리 강화 신경망 (physics-enforced neural network, PENN)을 제안하여 외삽 영역에서 일반화 성능이 현저히 개선됨을 확인<sup>37</sup>하였다. 한편 거대 언어모델을 물리 기반 고분자 물성예측 모델에 통합하는 시도도 제시되고 있는데, 풍부한 물리 기반 합성 데이터로 사전학습한 후 소량의 실험 데이터로 미세조정 (fine-tuning)하는 2단계 LLM이 순수 데이터 기반 접근 대비 유의한 성능 향상을 보임이 보고<sup>38</sup>되었다. 이러한 접근은 데이터가 부족한 상황에서 물리적 제약 조건이 귀납적 편향 (inductive bias)으로 작용하여 학습 효율을 높이는 데 기여할 수 있다.

종합하면, 고분자 정보학의 추가적인 발전을 위해서는 커뮤니티 차원의 표준화된 데이터 인프라 구축, 고분자 고유의 계층적 복잡성을 반영하는 표현학습 방법론의 발전, 그리고 물리 기반 사전지식과 다중충실도 학습을 적극 활용한 소규모 데이터 학습 전략의 고도화가 필요하다. 무엇보다

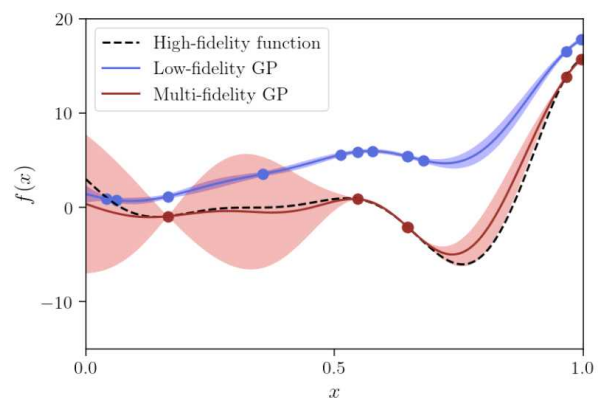


그림 5. 목적함수의 충실도 수준에 따른 편향을 반영하는 다중충실도 가우스과정 (GP) 모델링의 도해.<sup>36</sup>

고분자 도메인 전문가가 ML/AI 모델 개발에 적극적으로 참여하여 도메인 지식을 모델에 효과적으로 반영하는 것이 중요하다. 궁극적으로는 데이터와 AI 기반으로 설계된 고분자 소재가 실제 고성능을 발현하고 산업에 기여함으로써 실험실에서의 발견이 산업적 가치로 전환되는 단계를 지향해야 할 것이다.

### 참고문헌

1. J. Jumper *et al.*, *Nature*, **596**, 583 (2021).
2. C. Zeni *et al.*, *Nature*, **639**, 624 (2025).
3. J. J. de Pablo, B. Jones, C. L. Kovacs, V. Ozolins, and A. P. Ramirez, *Curr. Opin. Solid State Mater. Sci.*, **18**, 99 (2014).
4. J. J. de Pablo, N. E. Jackson, M. A. Webb, L.-Q. Chen, J. E. Moore, D. Morgan, R. Jacobs, T. Pollock, D. G. Schlom, E. S. Toberer, J. Analytis, I. Dabo, D. M. DeLongchamp, G. A. Fiete, G. M. Grason, G. Hautier, Y. Mo, K. Rajan, E. J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, and J.-C. Zhao, *Npj Comput. Mater.*, **5**, 41 (2019).
5. I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, *Adv. Neural Inf. Process. Syst.*, **35**, 11423 (2022).
6. B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, *Nat. Mach. Intell.*, **5**, 1031 (2023).
7. Y. Park, J. Kim, S. Hwang, and S. Han, *J. Chem. Theory Comput.*, **20**, 4857 (2024).
8. M. Ishii, T. Ito, H. Sado, and I. Kuwajima, *Sci. Technol. Adv. Mater. Methods*, **4**, 2354649 (2024).
9. S. Wu, H. Yamada, Y. Hayashi, M. Zamengo, and R. Yoshida, DOI: 10.48550/arXiv:2010.07683 (2020).
10. R. Yoshida *et al.*, DOI: 10.48550/arXiv:2511.11626 (2025).
11. N. J. Rebello, A. Arora, H. Mochigase, T.-S. Lin, J. Shi, D. J. Audus, E. S. Muckley, A. Osmani, and B. D. Olsen, *J. Chem. Inf. Model.*, **64**, 6464 (2024).
12. T. D. Huan, A. Mannodi-Kanakithodi, C. Kim, V. Sharma, G. Pilania, and R. Ramprasad, *Sci. Data*, **3**, 160012 (2016).
13. C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, and R. Ramprasad, *J. Phys. Chem. C*, **122**, 17575 (2018).
14. H. Doan Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J. P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton, and R. Ramprasad, *J. Appl. Phys.*, **128**, 171104 (2020).
15. Y. Hayashi, J. Shiomi, J. Morikawa, and R. Yoshida, *Npj Comput. Mater.*, **8**, 222 (2022).
16. R. Ma and T. Luo, *J. Chem. Inf. Model.*, **60**, 4684 (2020).
17. C. Xu, Y. Wang, and A. Barati Farimani, *Npj Comput. Mater.*, **9**, 64 (2023).
18. G. Liu, S. Alosious, S. Mahajan, E. Inae, Y. Zhu, Y. Liu, R. Zhang, J. Xu, A. Howard, Y. Li, T. Luo, and M. Jiang, DOI: 10.48550/arXiv:2512.08896 (2025).
19. T.-S. Lin, C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, K. F. Jensen, and B. D. Olsen, *ACS Cent. Sci.*, **5**, 1523 (2019).
20. L. Schneider, D. Walsh, B. Olsen, and J. J. de Pablo, *Digit. Discov.*, **3**, 51 (2024).
21. J. Ahn, G. P. Irianti, Y. Choe, and S.-M. Hur, *Npj Comput. Mater.*, **10**, 110 (2024).
22. J. Ahn, G. P. Irianti, V. Thapar, and S.-M. Hur, DOI: 10.48550/arXiv:2512.06301 (2025).
23. J. Park, Y. Shim, F. Lee, A. Rammohan, S. Goyal, M. Shim, C. Jeong, and D. S. Kim, *ACS Polym. Au*, **2**, 213 (2022).
24. M. Aldeghi and C. W. Coley, *Chem. Sci.*, **13**, 10486 (2022).
25. O. Queen, G. A. McCarver, S. Thatigotla, B. P. Abolins, C. L. Brown, V. Maroulas, and K. D. Vogiatzis, *Npj Comput. Mater.*, **9**, 90 (2023).
26. W. Hwang, S. Kwon, W. B. Lee, and Y. Kim, *Soft Matter*, **20**, 4905 (2024).
27. M. Kim, M. Y. Ha, W.-B. Jung, J. Yoon, E. Shin, I. Kim, W. B. Lee, Y. Kim, and H. Jung, *Adv. Mater.*, **34**, 2108900 (2022).
28. M. Kim, Y. Kim, M. Y. Ha, E. Shin, S. J. Kwak, M. Park, I.-D. Kim, W.-B. Jung, W. B. Lee, Y. Kim, and H.-T. Jung, *Adv. Mater.*, **35**, 2211497 (2023).
29. M. Park, M. Kim, Y. Kim, S. J. Kwak, W.-B. Jung, H.-T. Jung, W. B. Lee, and Y. Kim, *Chem. Eng. J.*, **518**, 164693 (2025).
30. C. Kim, A. Chandrasekaran, A. Jha, and R. Ramprasad, *MRS Commun.*, **9**, 860 (2019).
31. S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama, and M. Naito, *Sci. Technol. Adv. Mater.*, **20**, 1010 (2019).
32. R. Batra, H. Dai, T. D. Huan, L. Chen, C. Kim, W. R. Gutekunst, L. Song, and R. Ramprasad, *Chem. Mater.*, **32**, 10489 (2020).
33. Z. Yang, W. Ye, X. Lei, D. Schweigert, H.-K. Kwon, and A. Khajeh, *Npj Comput. Mater.*, **10**, 296 (2024).
34. H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, and R. Yoshida, *ACS Cent. Sci.*, **5**, 1717 (2019).
35. K. Noda, A. Wakiuchi, Y. Hayashi, and R. Yoshida, *Commun. Mater.*, **6**, 36 (2025).
36. A. Manoj, A. S. Georgiou, D. G. Giovanis, T. P. Sapsis, and I. G. Kevrekidis, DOI: 10.48550/arXiv:2508.01013 (2025).
37. A. Jain, R. Gurnani, A. Rajan, H. J. Qi, and R. Ramprasad, *Npj Comput. Mater.*, **11**, 42 (2025).
38. N. Liu, S. Jafarzadeh, B. Y. Lattimer, S. Ni, J. Lua, and Y. Yu, *Nat. Comput. Sci.*, **5**, 245 (2025).