# Exploring the thermodynamic landscape and kinetic pathways for metastable materials synthesis

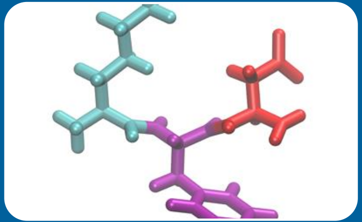
**S. Banik, R. Batra, T. Loeffler, H. Chan, S. Srinivasan, S. Manna, A. Chandra, G. Chen, Chris Fry, Pierre Darancet,**

**Subramanian Sankaranarayanan**

Center for Nanoscale Materials
Argonne National Laboratory

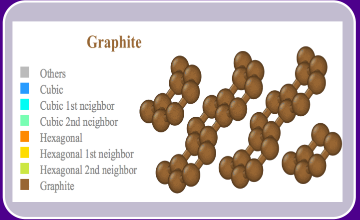
Department of Mechanical and Industrial Engineering,
University of Illinois Chicago

# Overview - High-dimensional Search Problems in Materials Synthesis



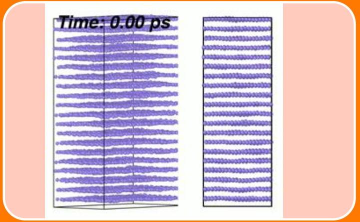
## Sequence problem in peptide synthesis

- Large and discrete search space
- Overcoming human bias in peptide synthesis




## Explore thermodynamic landscape of metastable materials

- Large and Continuous search space
- A phase diagram for metastable materials



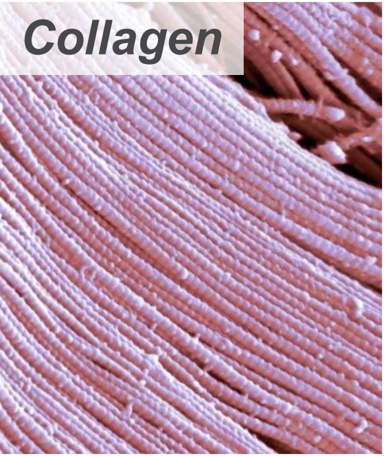

## Control kinetic pathways for synthesizing metastable materials

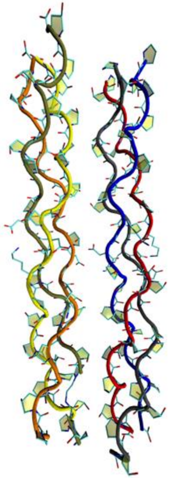
- Learn time-dependent synthesis protocols
- Controlled selection of metastable polymorphs

# Sequence problem in peptide synthesis

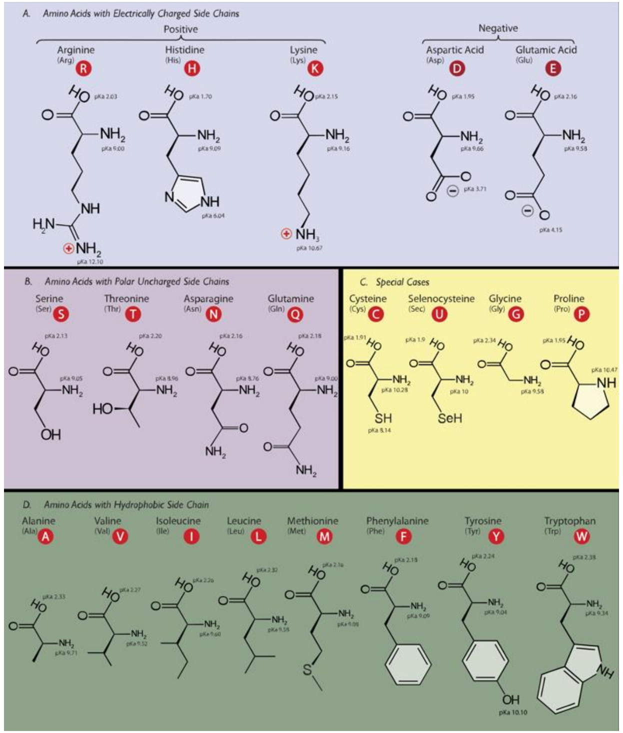
# Amino acid sequence controls the self-assembly of peptides

## Examples

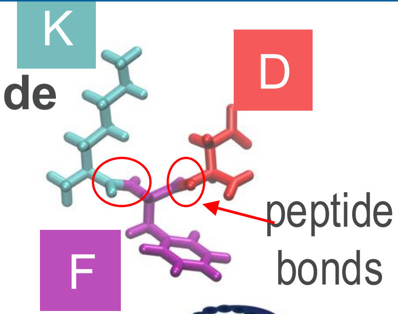



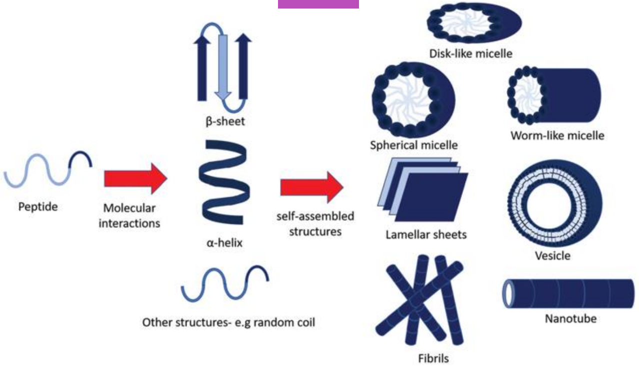



## 20 Amino Acids




## E.g. Tripeptide KFD




- Useful for catalysis, sensing, tissue engineering
- Easy tunable properties
- Huge diversity

# Traditional Approaches to Peptide Design and Synthesis

**Sequence Challenge in Peptide Discovery and Synthesis**

# of possibilities: $20^n$

| Sequence length (n) | # of candidates | Simulation time |
|---|---|---|
| 3 | 8000 | 3 weeks |
| 5 | 3.2 M | 20 years |
| 8 | 25.6 B | Many years |

**Empirical Rational Design**

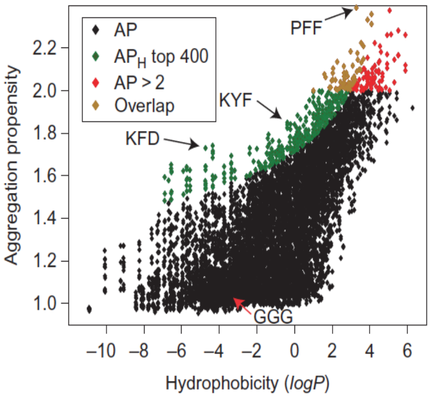
- Patterning (npnpn)
- Hydrophobicity scales
- Human bias; based on limited data

**Computational Design**

- MD simulation to estimate peptide properties
- Hydrophobicity scales
- Brute-force search; non-scalable

**Can we do better?**

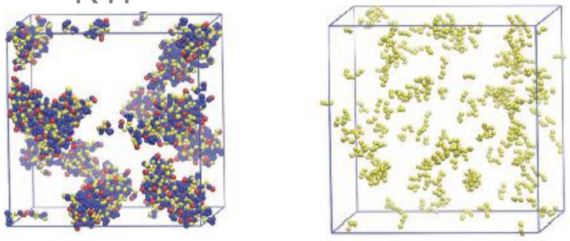
**Exploring self-assembling tripeptides**



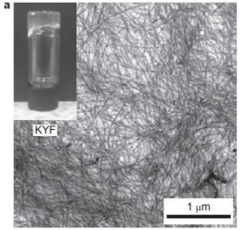

(MARTINI force-field)

KYF

GGG



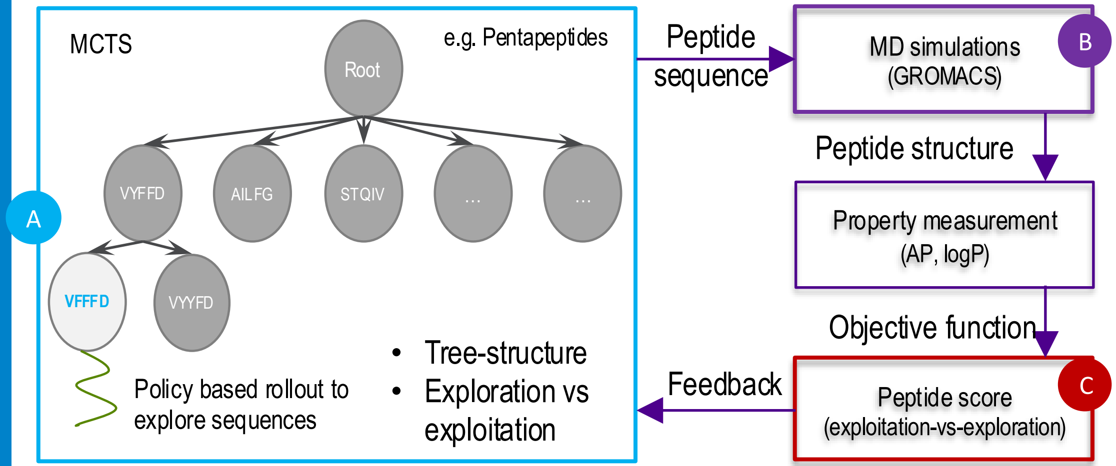
KYF




P. W. J. M. Frederix et al., Nature Chemistry, 7, 30 (2015)

# AI-Expert for Peptide Discovery

## AI-guided autonomous computational search




**Scoring function**

(AP)*(logP)

**Aggregation propensity**

**Hydrophilicity**

### Key components

A. Monte Carlo tree search (MCTS) → Promising sequence generation

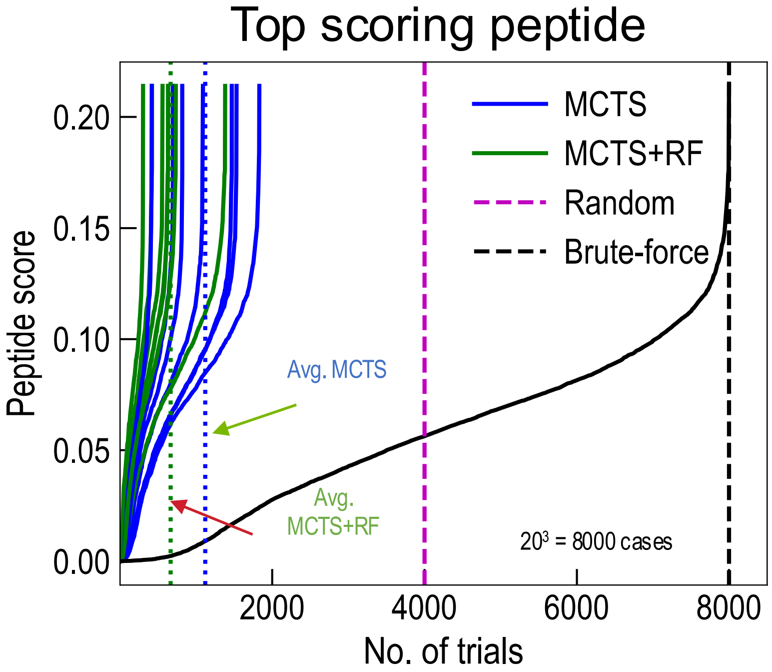
B. MD simulations → Model structure

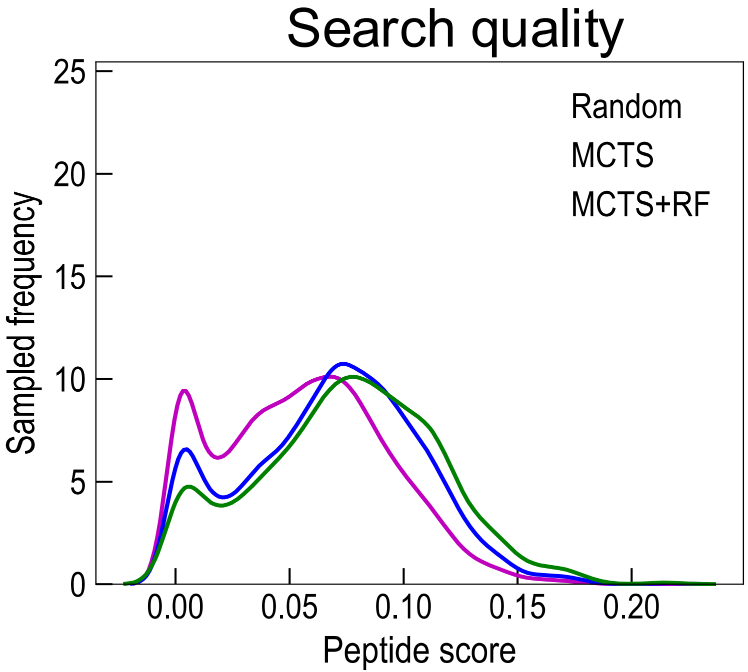
C. Scoring function → Sequence evaluation

**Search acceleration:** Random forest ML model for efficient rollouts **(MCTS+RF scheme)**

Batra, R, et al. "Self-assembling Peptide Discovery: Overcoming Human Bias With Machine Learning." Nature Chem (2022) (12):1427-1435

# AI-Expert Results – Validation on Tripeptides



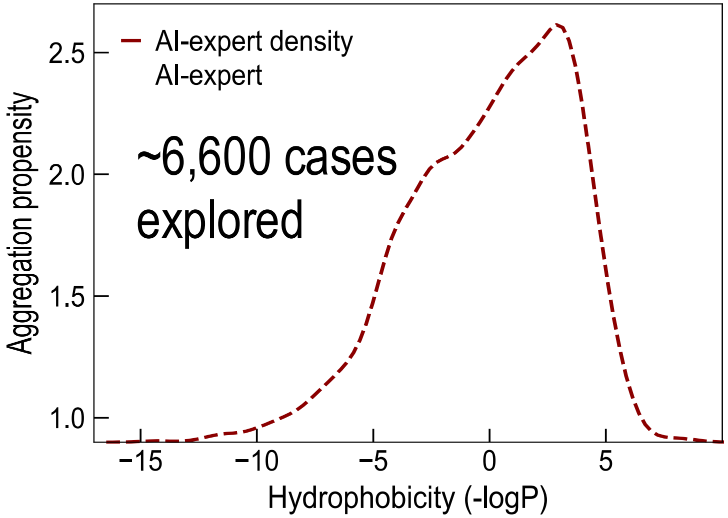

- Tripeptides offer a manageable space (8000 cases) for validation
- Superior performance of MCTS+RF, MCTS methodology

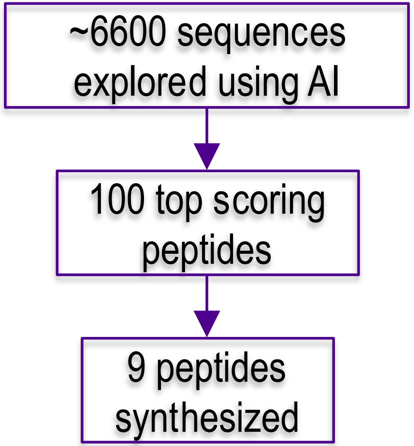
**AI-expert with MCTS+RF scheme performs best!**



Nature Chem (2022) (12):1427-1435

# AI-Expert Results – Screening of Pentapeptides



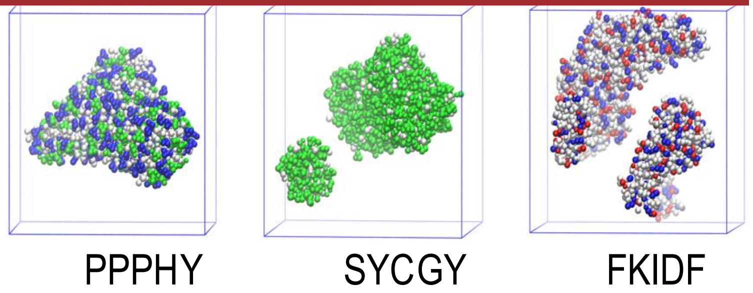

Example top candidates from AI-expert

| pep | AP | logP | Score |
|---|---|---|---|
| PTPCY | 2.54 | -0.2 | 0.49 |
| PPPHY | 2.46 | -0.18 | 0.44 |
| SYCGY | 2.42 | 0.17 | 0.42 |
| FFEKF | 2.24 | 1.30 | 0.34 |
| KWEFY | 2.20 | 1.92 | 0.33 |
| FKIDF | 2.17 | 1.90 | 0.31 |
| WKPYY | 2.11 | -0.57 | 0.26 |

- Pentapeptides cannot be explored exhaustively (3.2 M cases)
- AI-expert candidates have **high scores** and **high diversity**

**Are these candidates any good?**

Computationally visible aggregation






Nature Chem (2022) (12):1427-1435

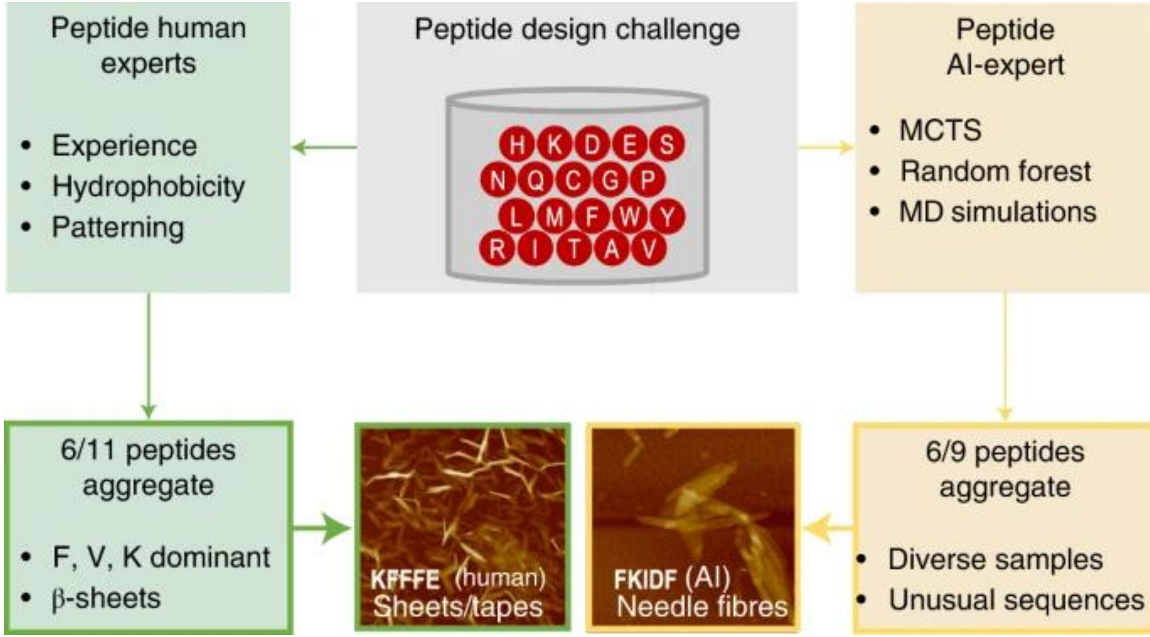
# Comparison of AI-expert & Human Experts

## Requested sequences from six Human experts

- Leaders in peptide research
- Years of synthesis experience
- 11 (out of 23) sequences synthesized

Top scoring pentapeptides: Computational score

| Rank | Peptide | Expert | Score |
|---|---|---|---|
| 1 | PPPHY | AI | 0.44 |
| 2 | SYCGY | AI | 0.43 |
| 3 | WKPYY | AI | 0.34 |
| 4 | KWEFY | AI | 0.33 |
| 5 | FFEKF | AI | 0.31 |
| 6 | FKFEF | Human | 0.28 |
| 7 | FKIDF | AI | 0.25 |
| 8 | VVVVV | Human | 0.21 |




- **For aggregation, AI-expert >= Human-experts**
- Based on just computational scores, AI-expert performs better
- Low computational score = no aggregation (e.g. KVKVK)
- High computational score ≠ aggregation (e.g. PPPHY)

Nature Chem (2022) (12):1427-1435

# Key conclusions

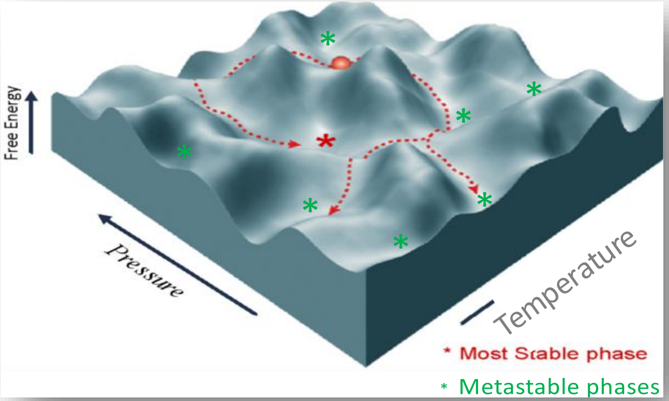
- Typically, peptide discovery research yields fewer than ten candidates per study and is subject to human bias.
- A fully autonomous computational search engine has been developed to reveal non-intuitive peptide sequences with high potential for self-assembly.
- The AI expert combines Monte Carlo tree search, molecular dynamics simulations, and a scoring function closely paired with experimental synthesis.
- AI-Expert screens millions of pentapeptide to efficiently determine 6,600 candidates for self-assembly
- While success percentages are similar between AI-Expert and Human Experts, the sequences that self-assemble are non-intuitive with high self-assembly propensity

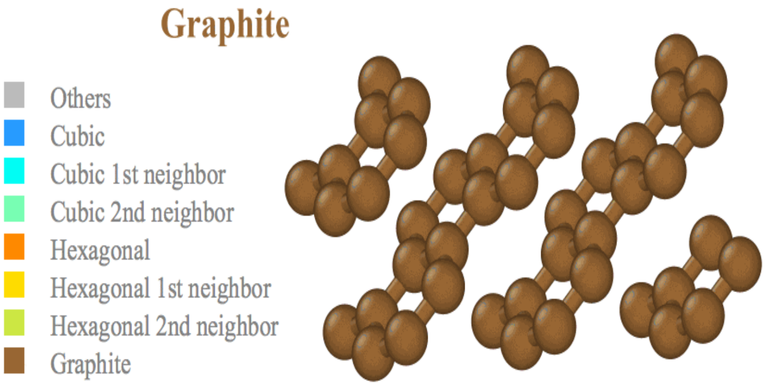
# Exploring Thermodynamic Landscape of Metastable Phases

Discrete $\longrightarrow$ Continuous Search Space

# Synthesis of Metastable Materials

**Opportunity:** **Metastable materials often have exotic properties but their synthesis pathways remain poorly understood**



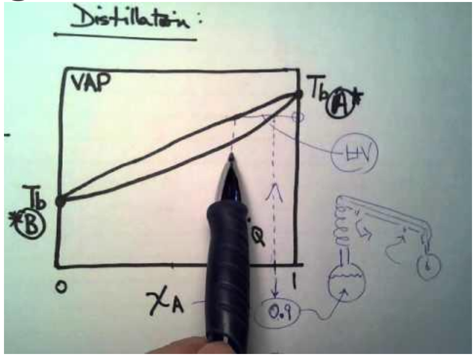

**Materials Science Challenge:** **Thermodynamics & Kinetics**

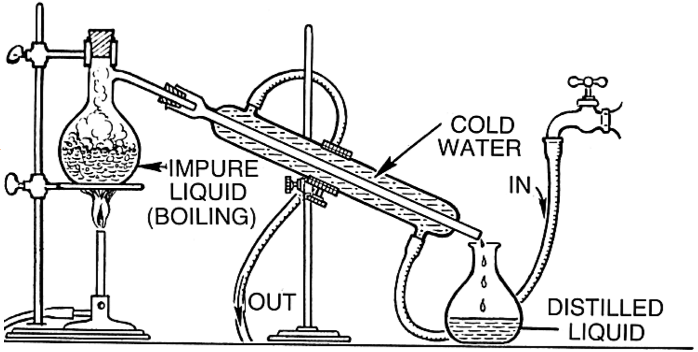
**AI Challenge:** **Continuous space search, high-dimensional, multiple objectives/rewards, physical/chemical constraints to name a few**

**Grand Challenge and Beyond Exascale but Game Changer for Materials Science**

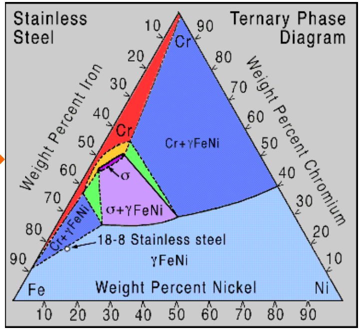
# PHASE DIAGRAMS AND SYNTHESIS

**1800's**



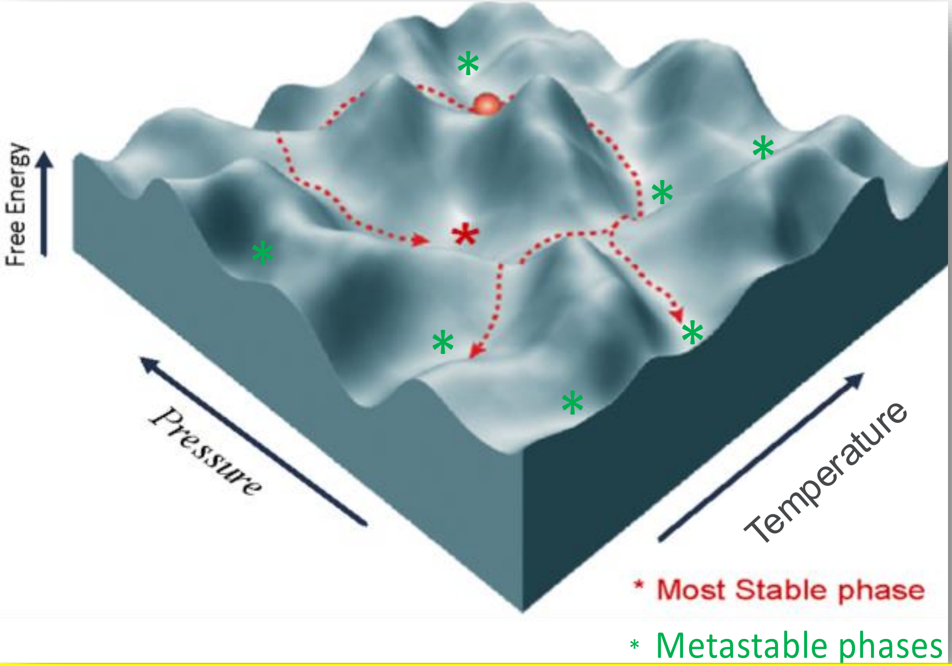

Thermo-Calc Software




**Present**

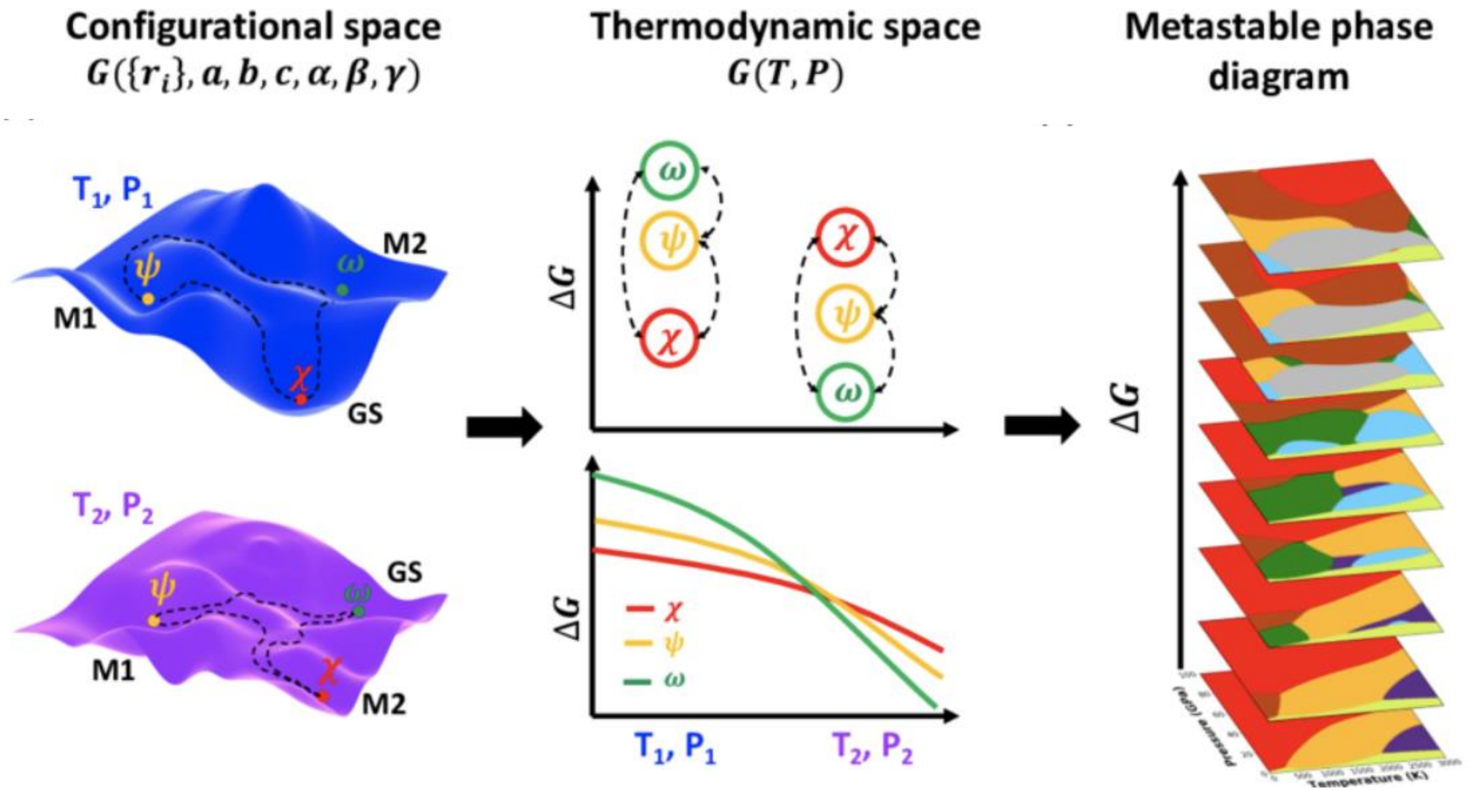
**Phase diagram → Equilibrium**

# Equilibrium and metastable states during synthesis



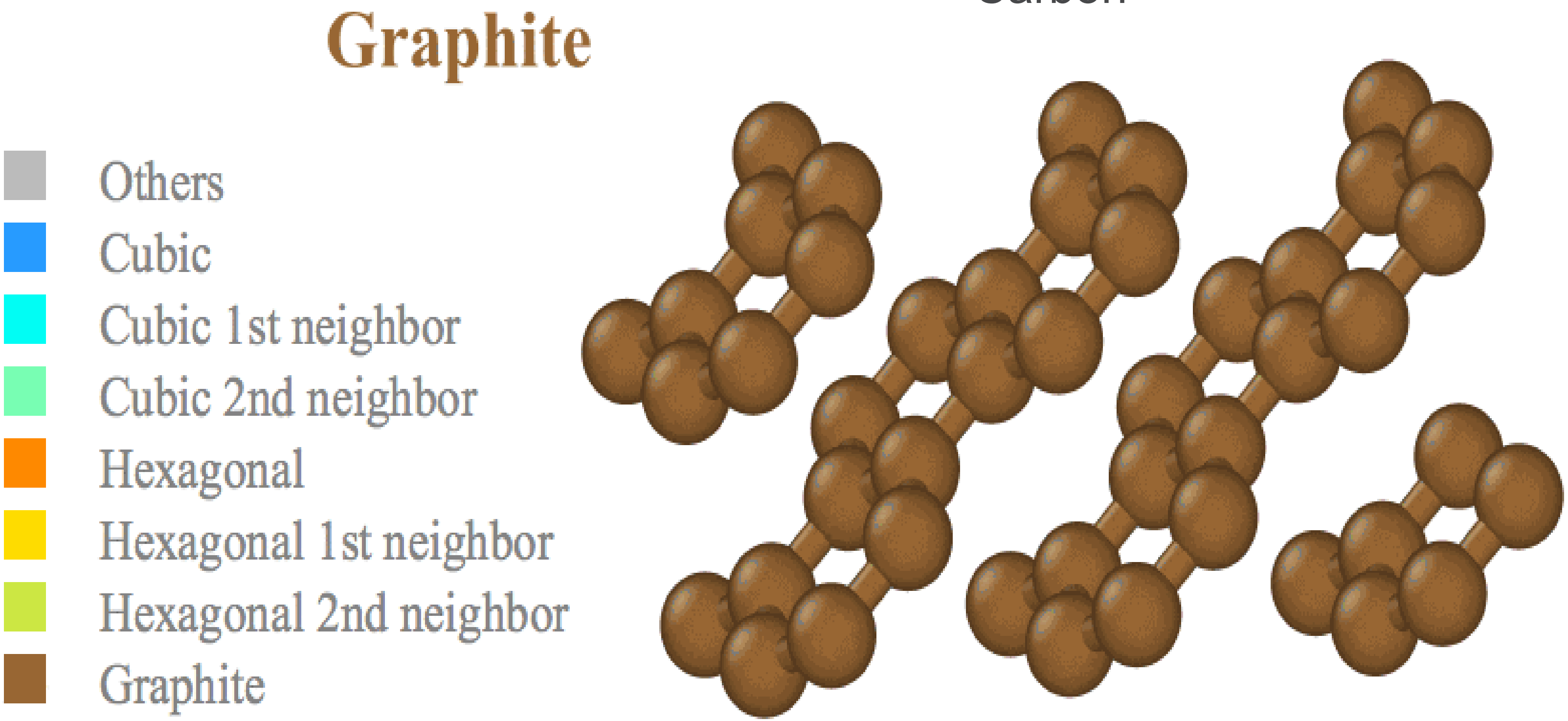

**Metastable phase diagrams are data-intensive and offer infinite possibilities**

# Metastable Phase Diagram



# Sample and Discover Metastable Phases

Carbon




## Continuous Search Space

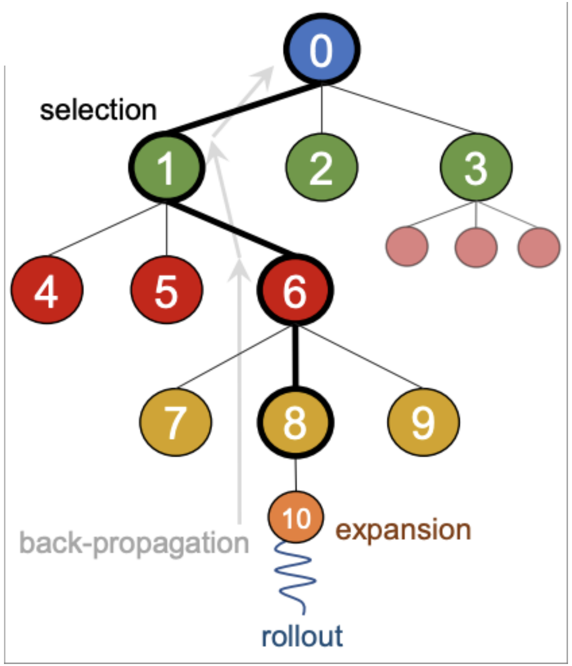
# High Dimensional Continuous Search Space

## Important Modifications for Continuous MCTS

(1) A uniqueness function to avoid degeneracy

(2) Correlating tree depth to action space

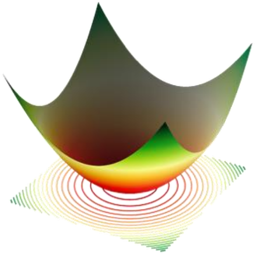
(3) Implementing an adaptive sampling of playouts



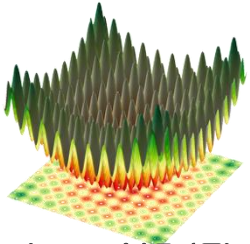

*Nature Communications* **volume 13**, Article number: 368 (2022)

# Performance on Standard Test Functions

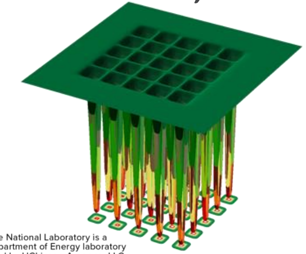
## Sphere (Unimodal)

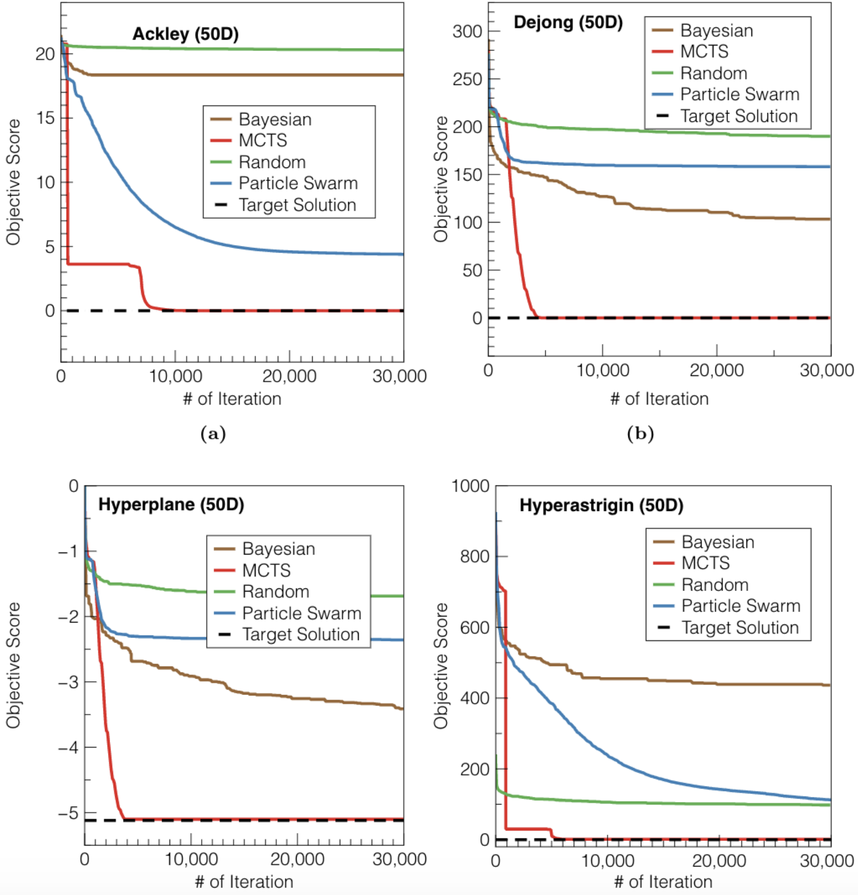


## Rastrigin (Multimodal)

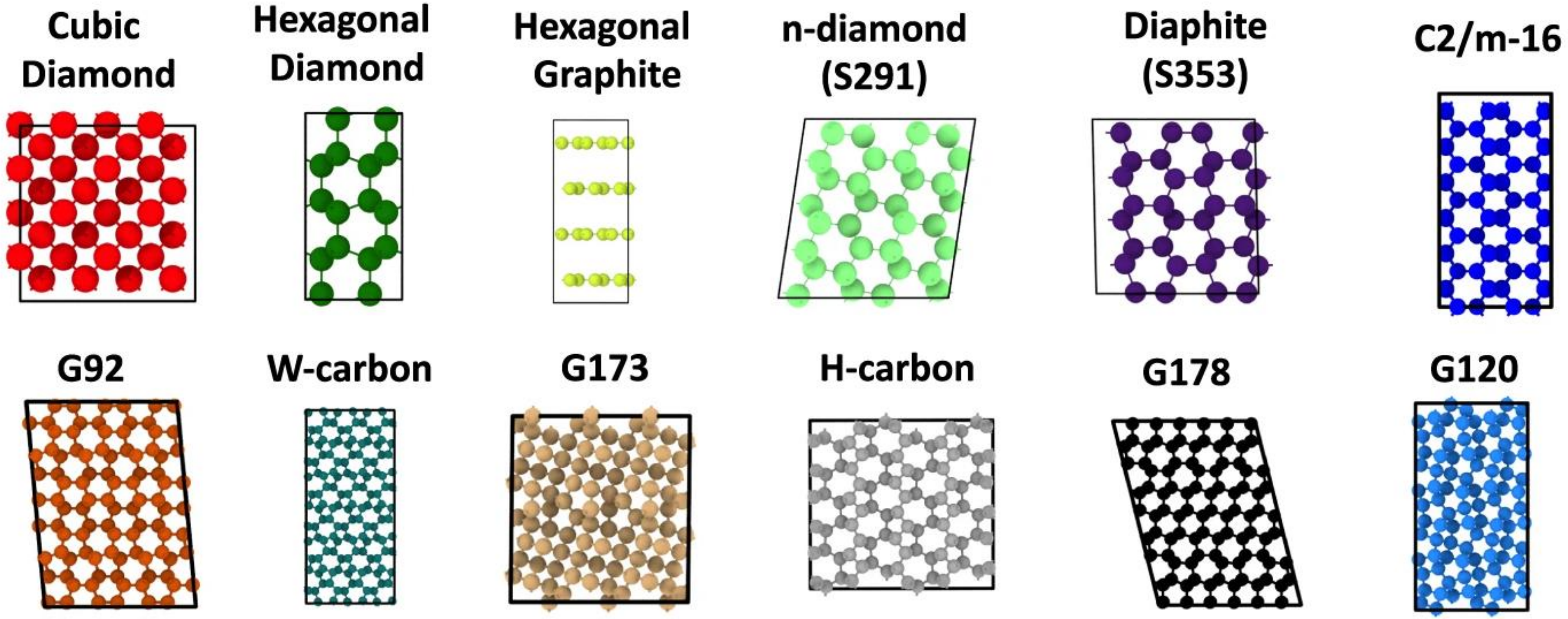


## Dejong N5 (Fixed Dimension)





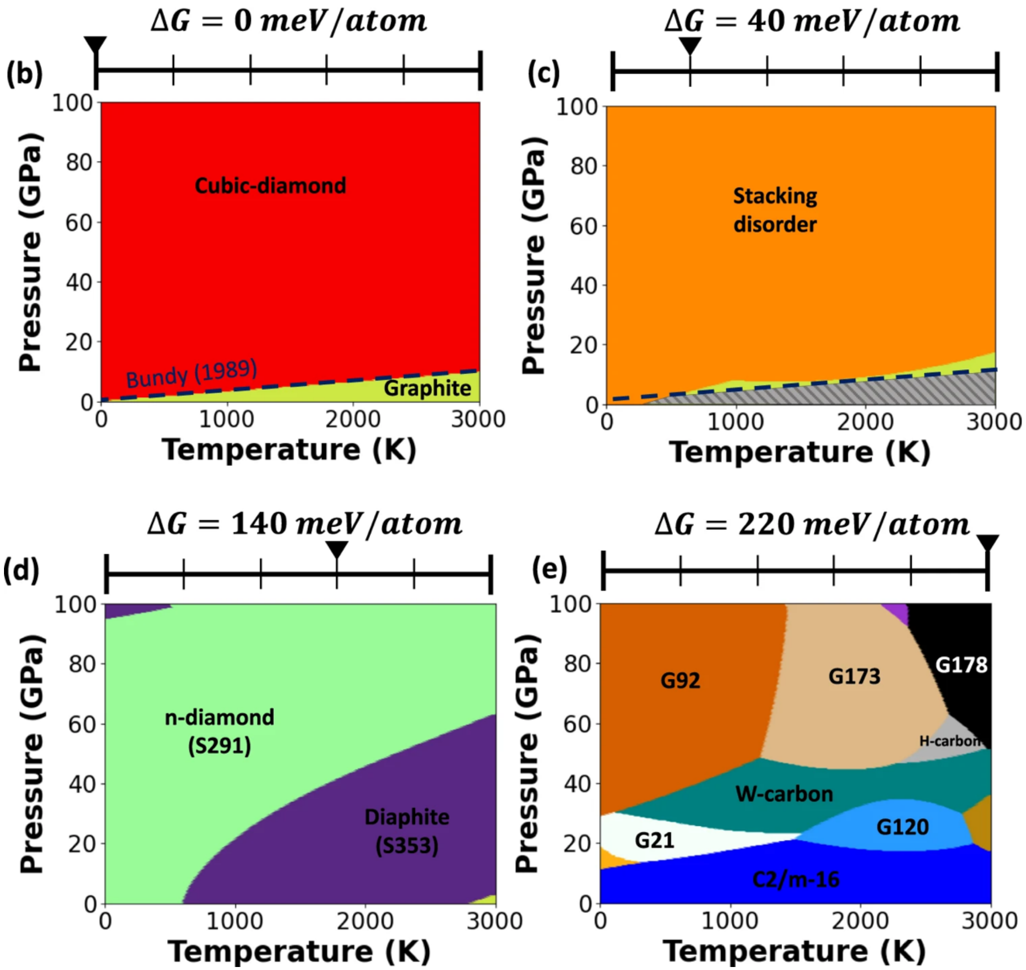
# Discover metastable phases




**~1000 metastable phases of carbon identified within 0.5 eV per atom**

<u>*Nature Communications*</u> **volume 13**, Article number: 3251 (2022)

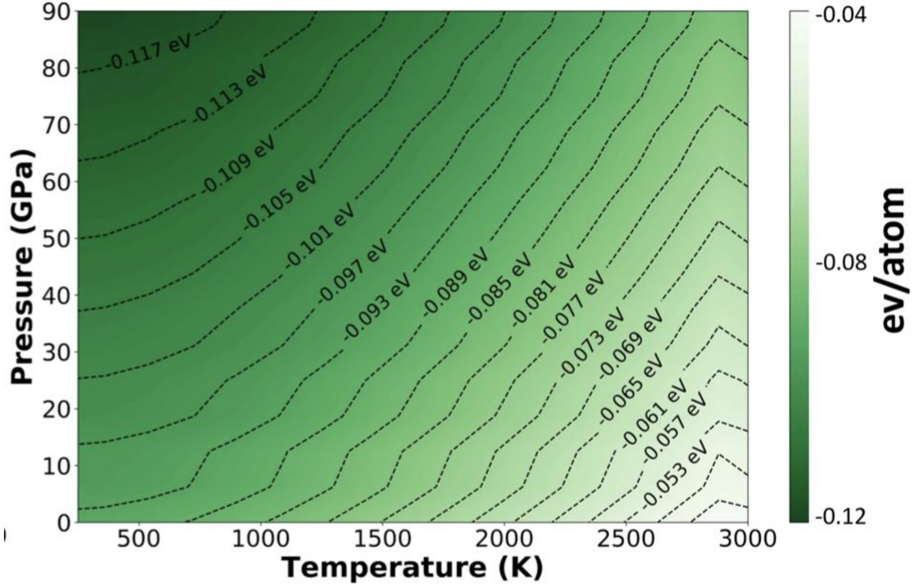
# Construct Metastable Phase Diagram



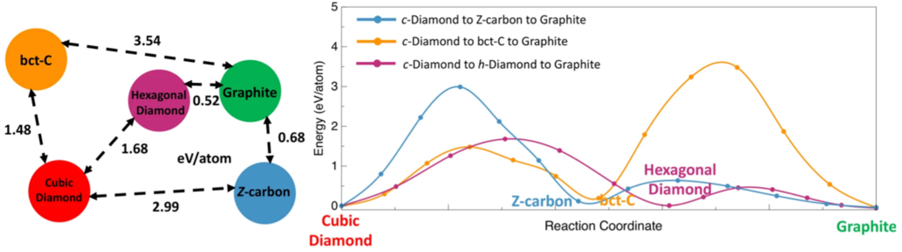
# Post Processing – Synthesizability and Domains of relative phase stability

*Relative stability of metastable phases*

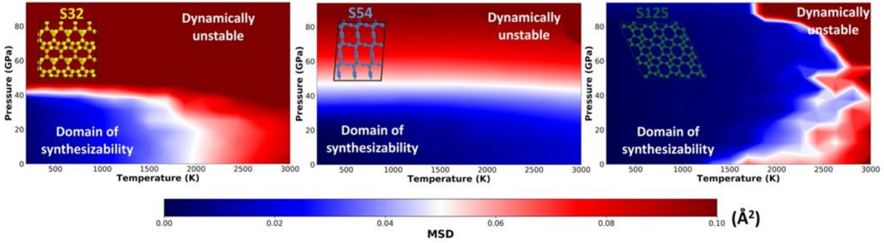
$$\Delta G = G_{hexagonal-diamond} - G_{diaphite}$$




*Transformation barriers to metastable phases*

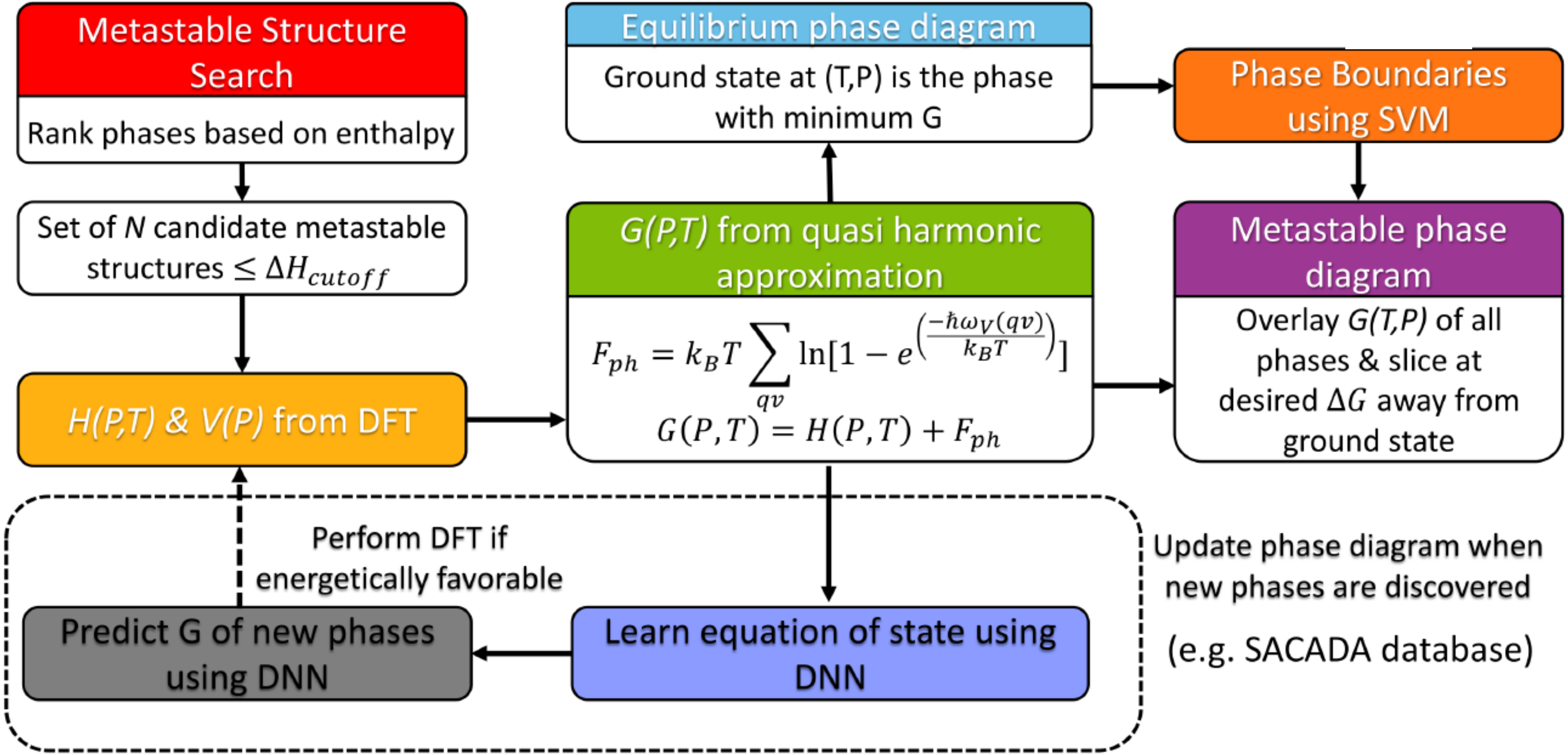



*Domains of synthesizability*




<u>*Nature Communications*</u> **volume 13,** Article number: 3251 (2022)

# A workflow for metastable phase diagram construction



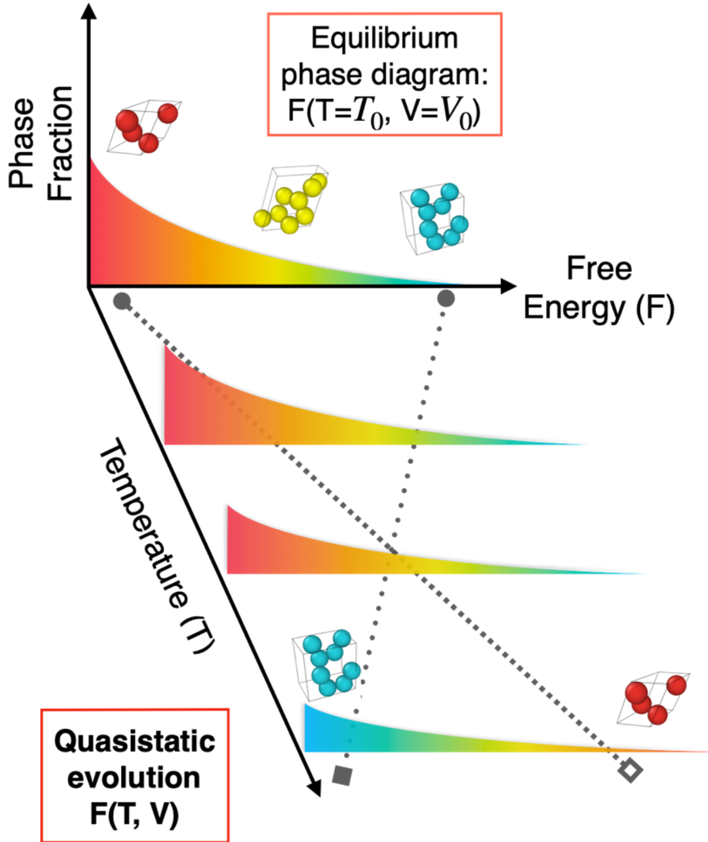



*Nature Communications* **volume 13**, Article number: 3251 (2022)

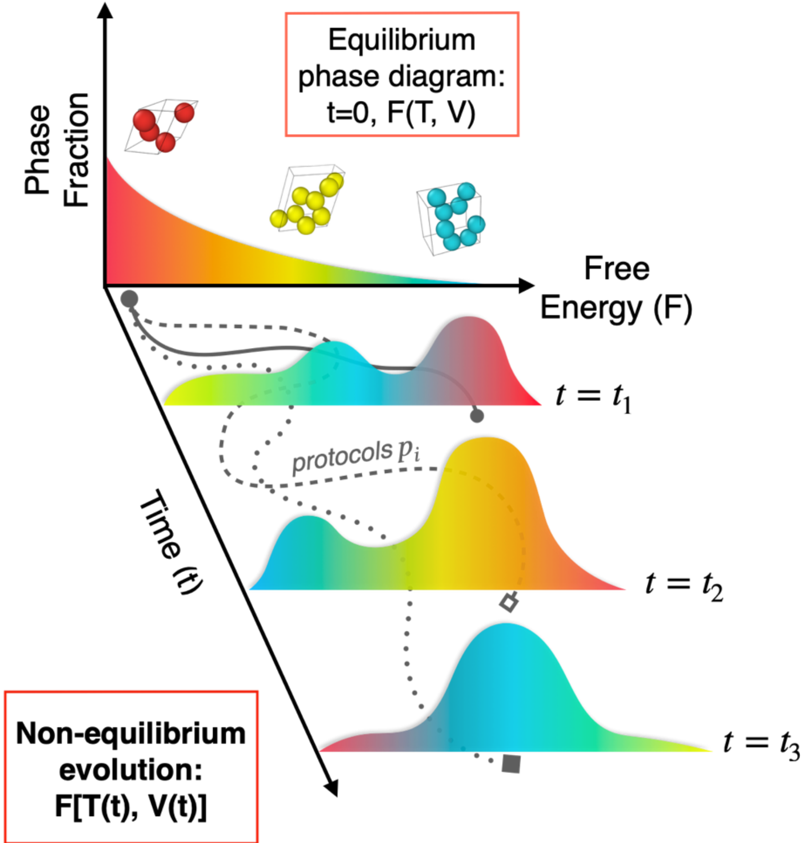
# Learning kinetic pathways for synthesizing metastable materials

# What about kinetics?

## Thermodynamics




## Kinetics



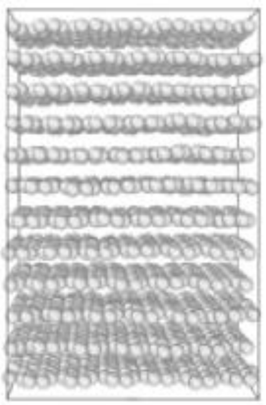
# Can we learn time-dependent protocols for polymorph selection?

Graphite (hexagonal)

T(t) - ? ; P(t)?; Lx(t), Ly(t), Lz(t)?

Desired Polymorph (cubic, hexagonal or any other)

Create molecular or atomic models



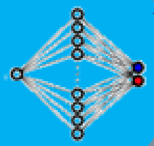
## Neuroevolutionary workflow

Molecular Dynamics

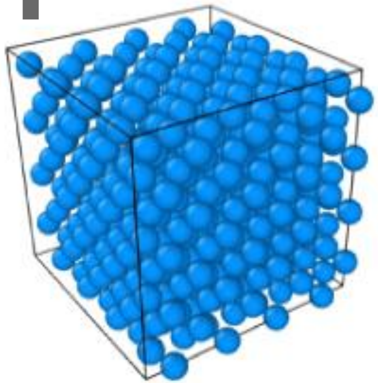
T, P. L

time, structure

$\Delta\phi_i(t) \equiv$



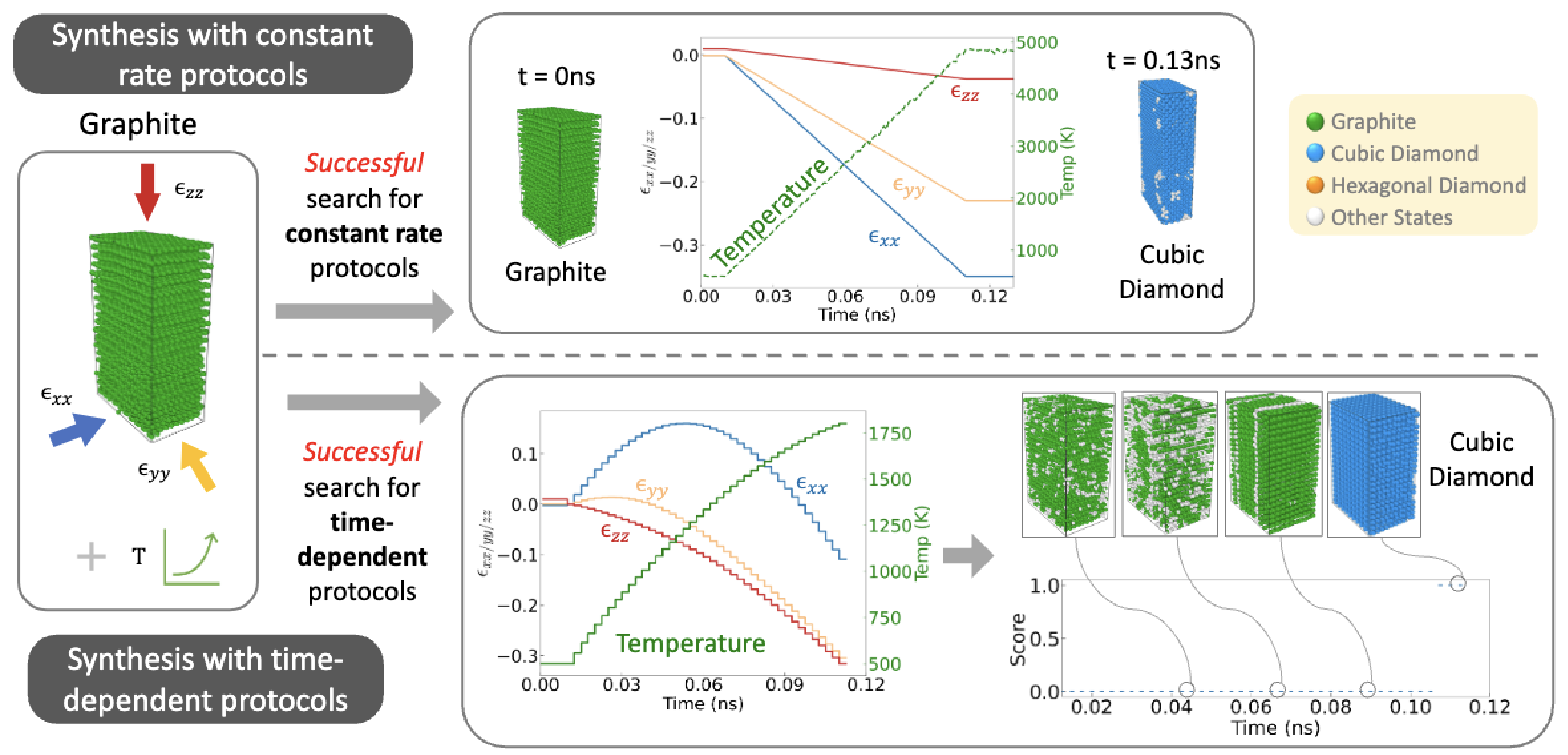
Use learned protocol [T(t), P(t), L(t)] in experiments



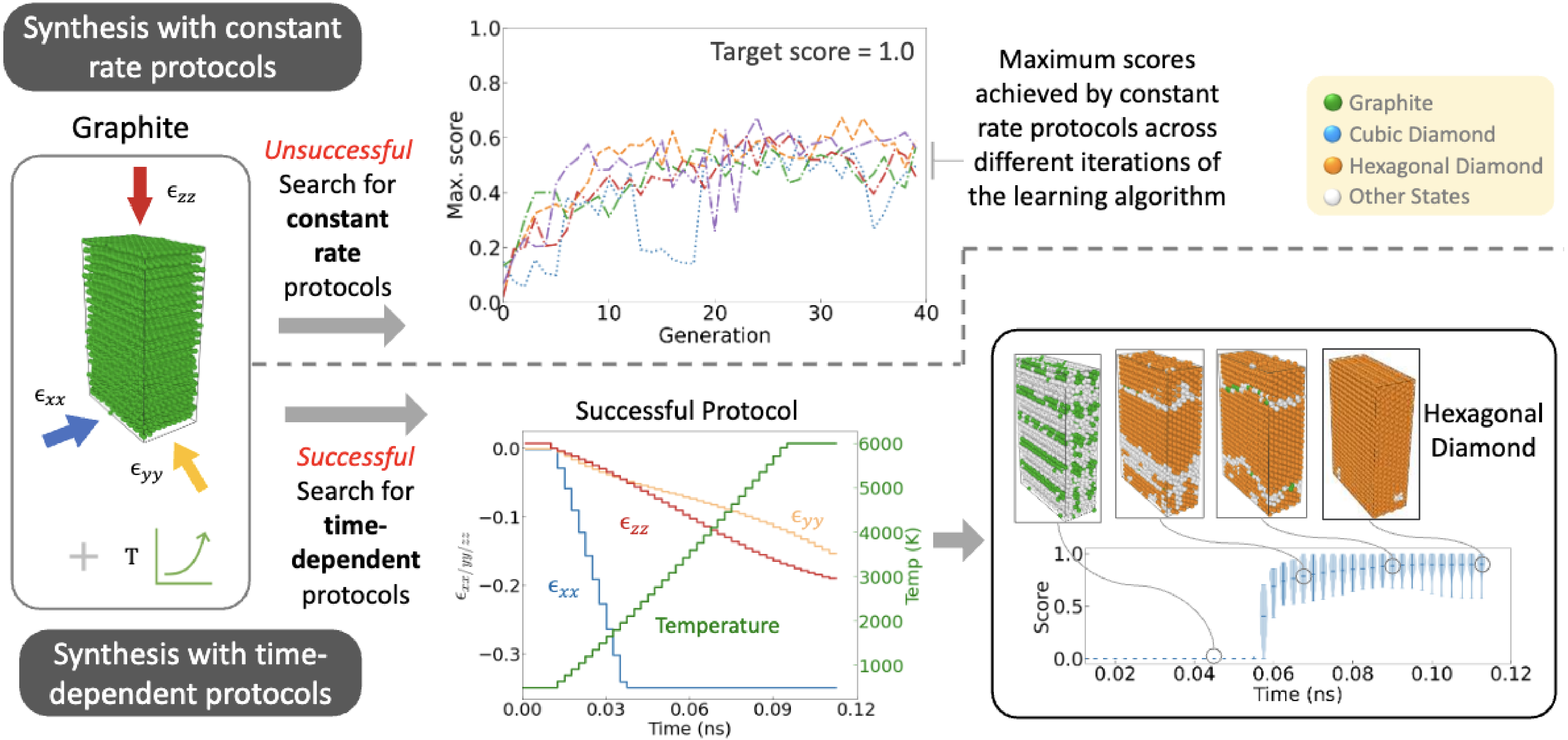
Repeat for several Generations to obtain optimized NN weights and desired T(t), P(t)

Collaborator – Stephen Whitelam (L2G workflow)

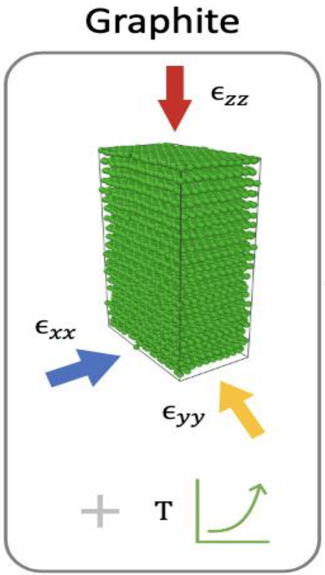
# Learn time-dependent protocols to form Cubic Diamond



# Learn time-dependent protocols to form Hexagonal Diamond



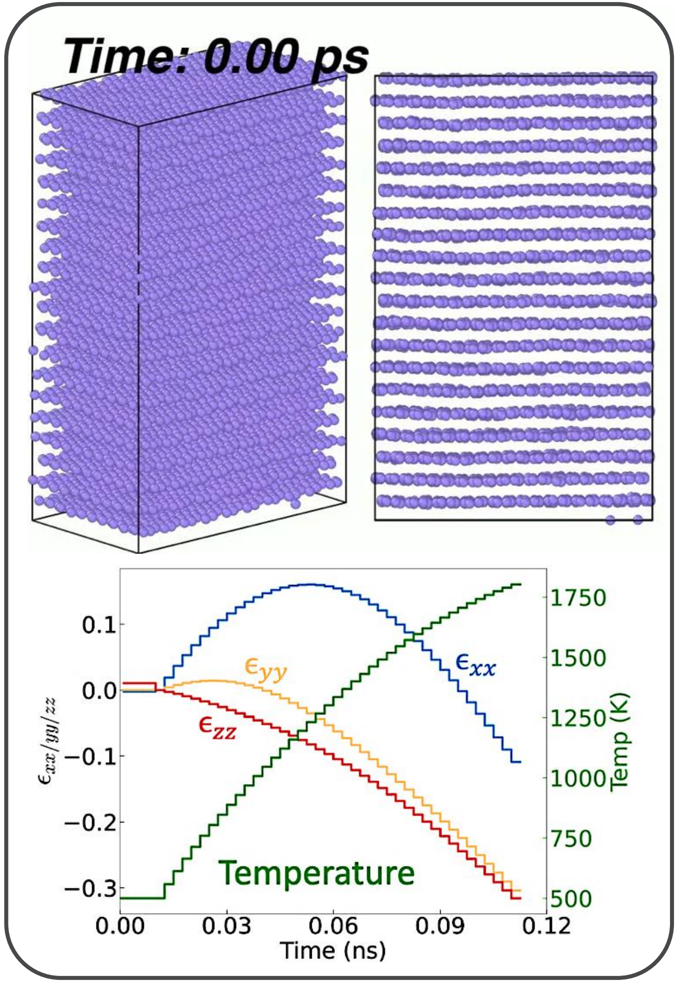
# Mechanisms for transformation to Cubic vs. Hexagonal Diamond

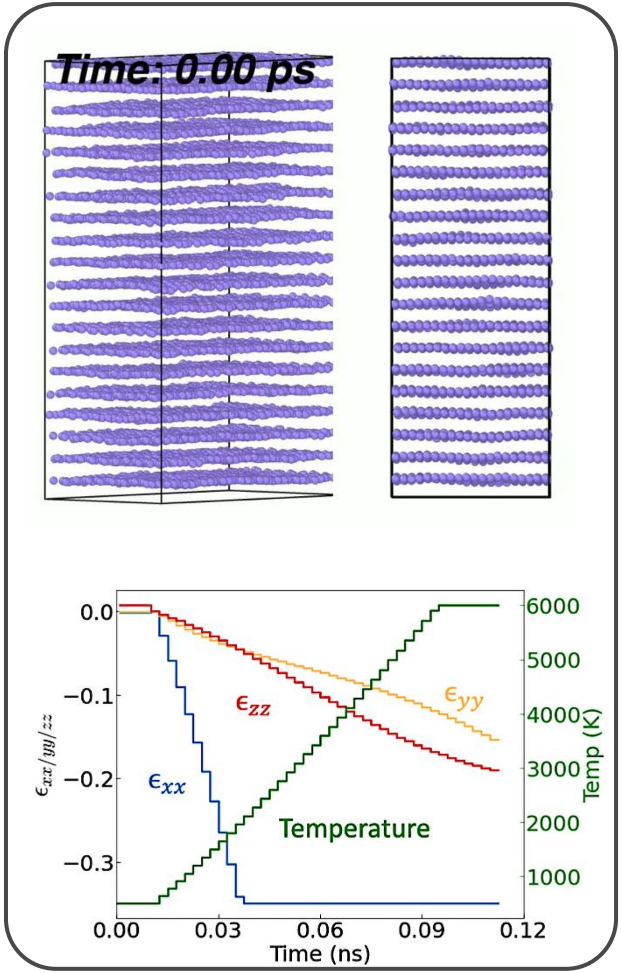



Graphitic

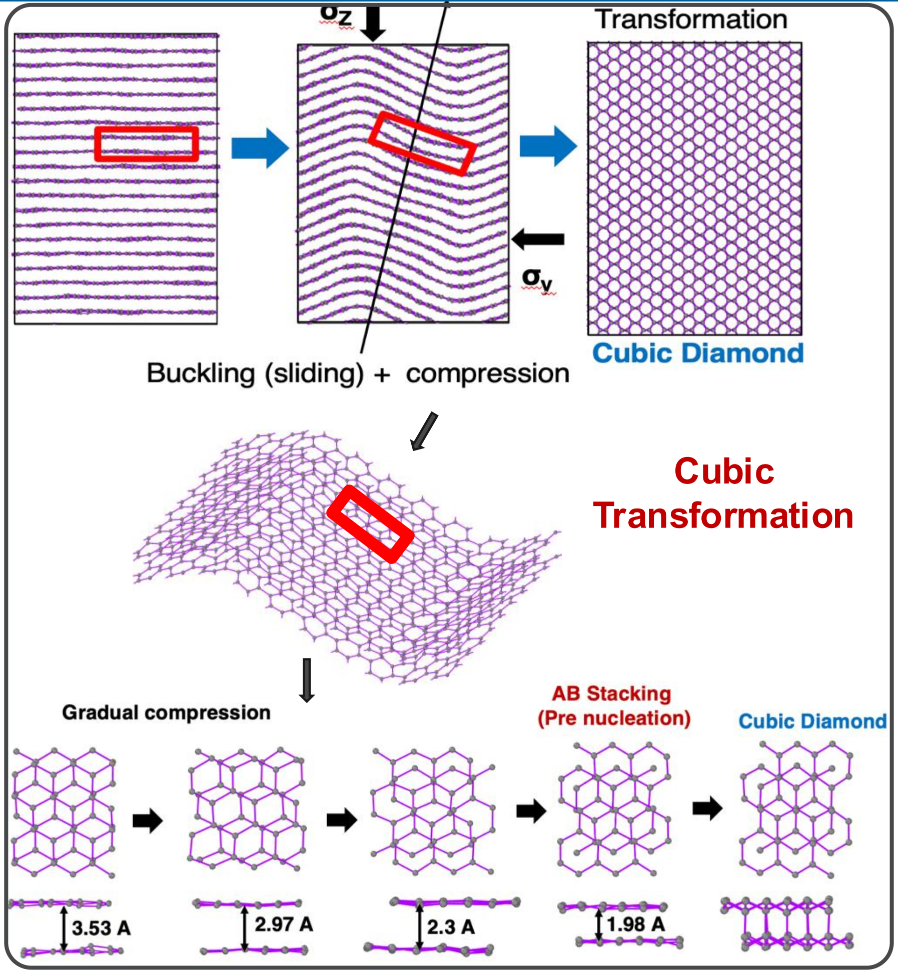
Cubic

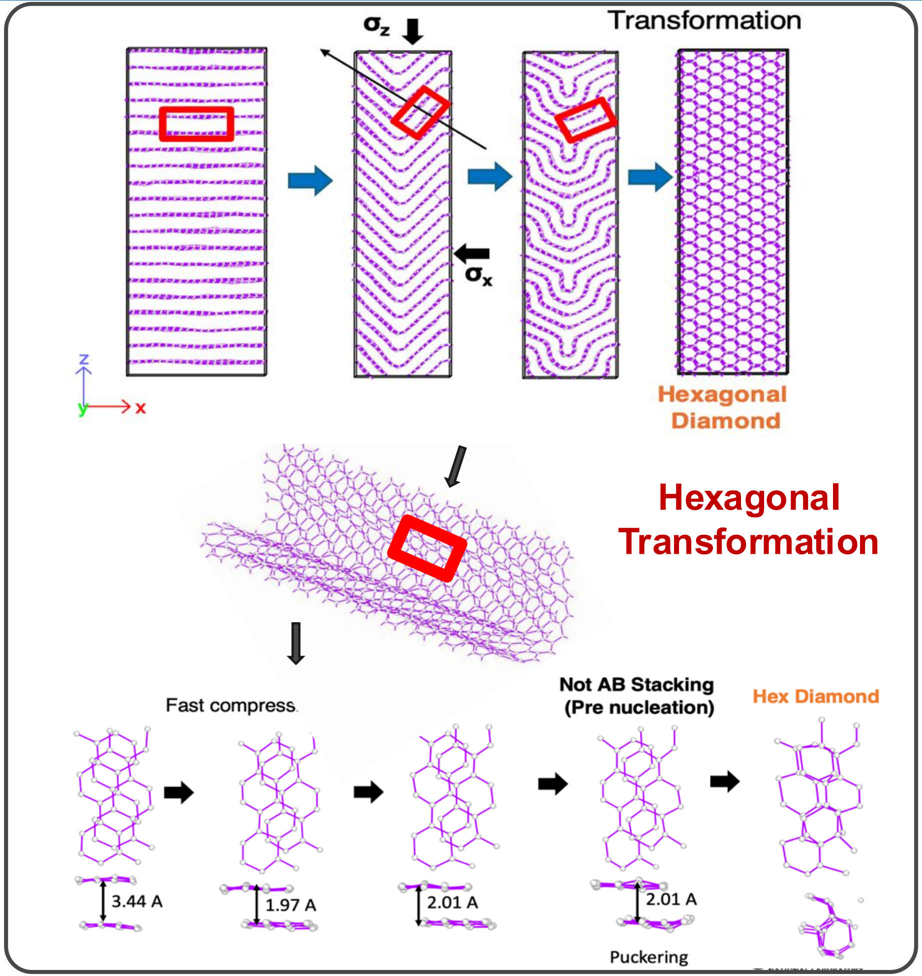
Hexagonal

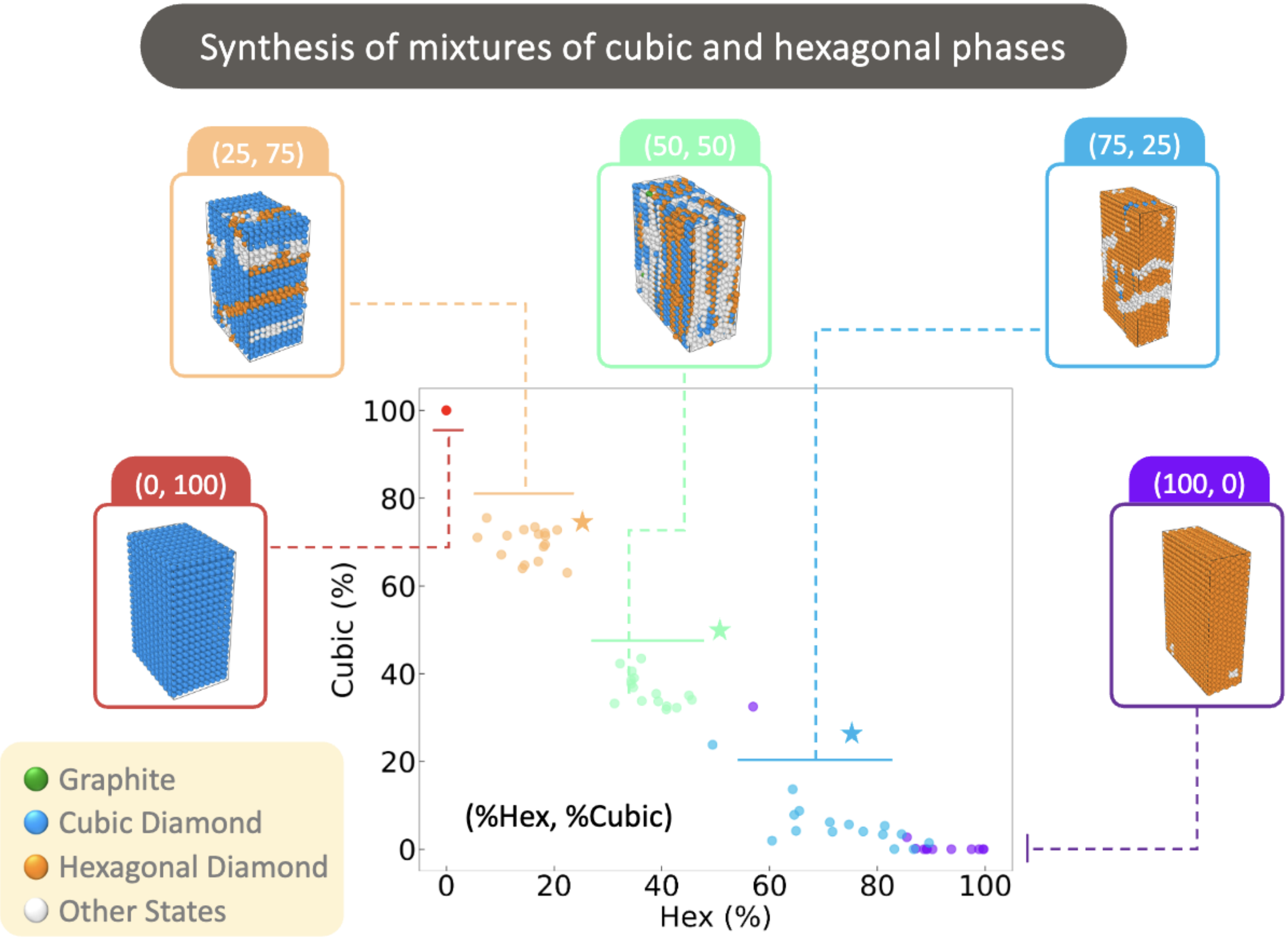


# Mechanisms for transformation to Cubic vs. Hexagonal Diamond





Synthesis of mixtures of cubic and hexagonal phases
(25, 75)
(50, 50)
(75, 25)
(0, 100)
(100, 0)
Cubic (%)
Hex (%)
(%Hex, %Cubic)
Graphite
Cubic Diamond
Hexagonal Diamond
Other States

# Future Outlook

- Extension to multicomponent systems
- Fast accurate models to represent atomistic and molecular interactions
- Fingerprinting or scoring function to identify emerging order
- Information extraction from multimodal experiments
- Digital Twins for synthesis of metastable materials
- Autonomous synthesis of materials beyond equilibrium

# THANK YOU

*Work performed at the Center for Nanoscale Materials, a U.S. Department of Energy Office of Science User Facility, was supported by the U.S. DOE, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.*