Microsoft

Distributional Graphormer (DiG): From Structure Prediction to Distribution Prediction

Chang Liu
Microsoft Research AI for Science
on behalf of the DiG team

AI for Scientific Computation

End-to-end prediction:

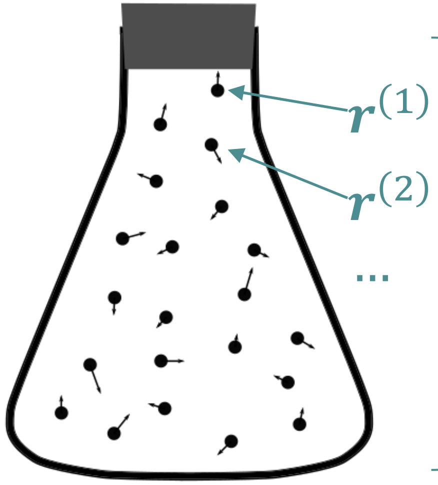
- Molecular-system descriptor
→ Equilibrium structure, molecular property, …
- Molecule structure
→ Potential energy (for molecular dynamics simulation)

Protein Structure Prediction

Distributions in Thermodynamics

But the world is *probabilistic* when zooming in:

- Molecular-system descriptor
  → Equilibrium structure $\boldsymbol{r}^*$
  → **Structure distribution** $p(\boldsymbol{r})$

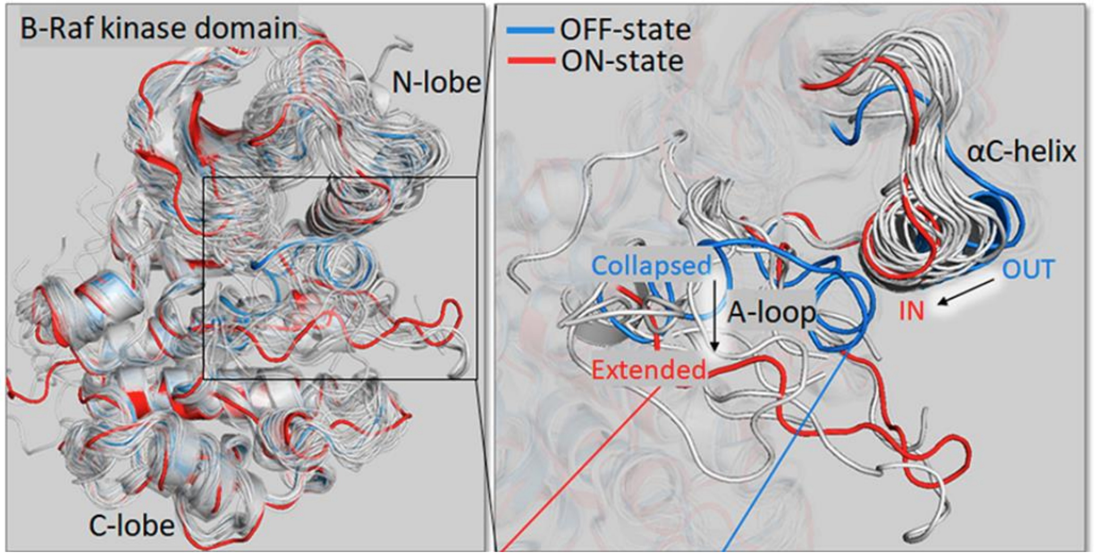


$$p(\boldsymbol{r}) \propto \exp\left(-\frac{E(\boldsymbol{r})}{kT}\right)$$

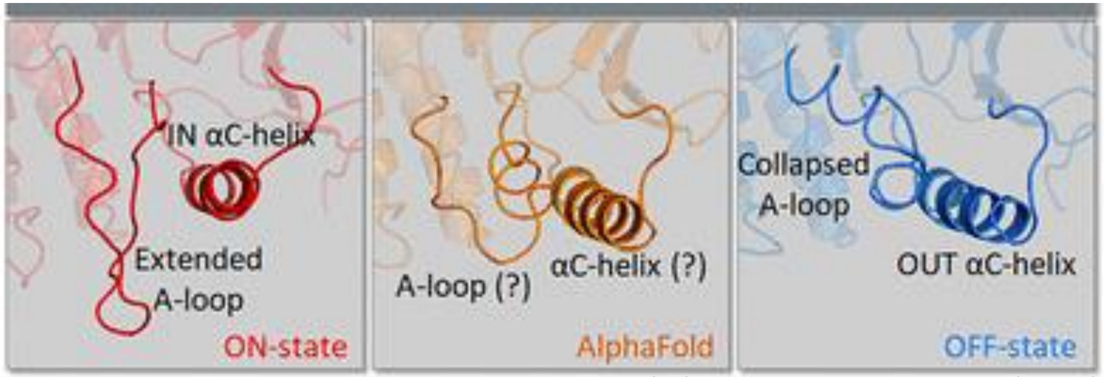
Canonical (const. NVT)

https://en.wikipedia.org/wiki/Ensemble_(mathematical_physics)

Proteins function by multiple structures:



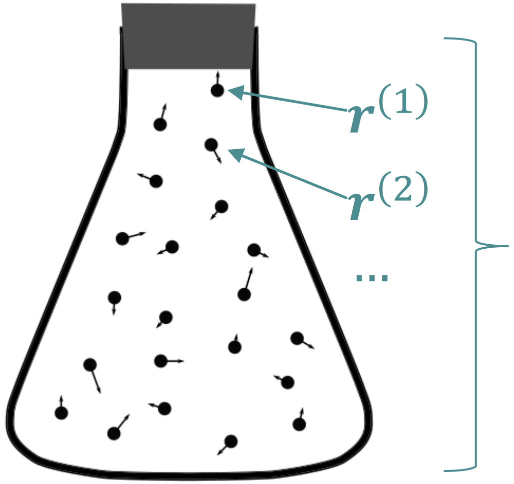
AlphaFold cannot predict any of them:



"AlphaFold, Artificial Intelligence (AI), and Allostery," JPCB (2022)

Distributions in Thermodynamics

The states of molecules $\boldsymbol{r}$ in real world follow a distribution.

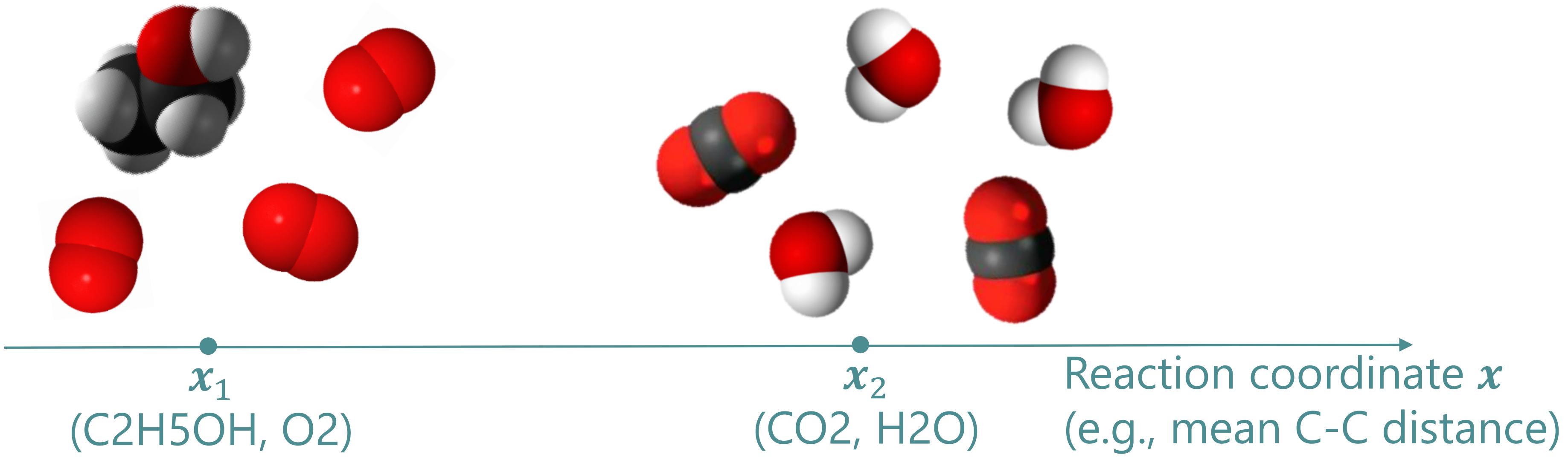


Canonical (const. NVT)

https://en.wikipedia.org/wiki/Ensemble_(mathematical_physics)

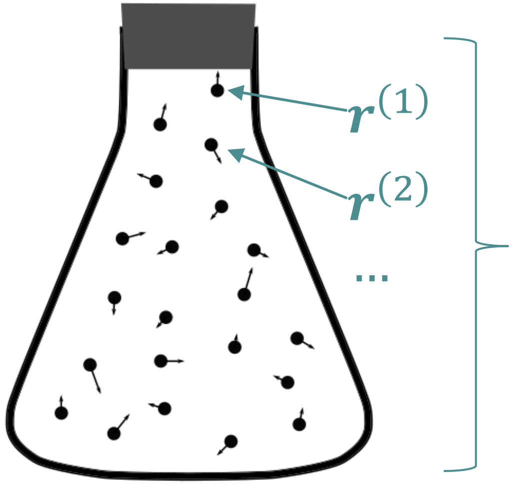
Detailed Description:

- In equilibrium: Boltzmann distribution $p(\boldsymbol{r}) = \exp\left(-\frac{E(\boldsymbol{r})}{kT}\right)/Z$.
- Non-equilibrium: $q(\boldsymbol{r})$.
  - $q(\boldsymbol{r}) = p(\boldsymbol{z}|\boldsymbol{x}_1)\delta_{\boldsymbol{x}_1}(\boldsymbol{x})$,
    where $\boldsymbol{r} \leftrightarrow (\text{reaction coord } \boldsymbol{x}, \text{other coord } \boldsymbol{z})$.



Distributions in Thermodynamics

The states of molecules $\boldsymbol{r}$ in real world follow a distribution.



Canonical (const. NVT)

https://en.wikipedia.org/wiki/Ensemble_(mathematical_physics)

Detailed Physics:

- Macroscopic property: $\mathbb{E}_{q(\boldsymbol{r})}[f(\boldsymbol{r})]$.
- (Helmholtz) Free energy: $F[q] = \mathbb{E}_{q(\boldsymbol{r})}[E(\boldsymbol{r})] - Tk\underbrace{\mathbb{E}_{q(\boldsymbol{r})}[-\log q(\boldsymbol{r})]}_{\text{Entropy } S[q]}$.

  If $q_{\boldsymbol{x}_1}(\boldsymbol{r}) = q(\boldsymbol{z}|\boldsymbol{x}_1)\delta_{\boldsymbol{x}_1}(\boldsymbol{x})$, then

  $$F[q_{\boldsymbol{x}_1}] = -kT(\mathrm{ELBO}[q(\boldsymbol{z}|\boldsymbol{x}_1), p(\boldsymbol{r})] - \log Z)$$
  $$= -kT\big(-\mathrm{KL}\big(q(\boldsymbol{z}|\boldsymbol{x}_1)\|p(\boldsymbol{z}|\boldsymbol{x}_1)\big) + \log Z_{\boldsymbol{x}_1}\big).$$

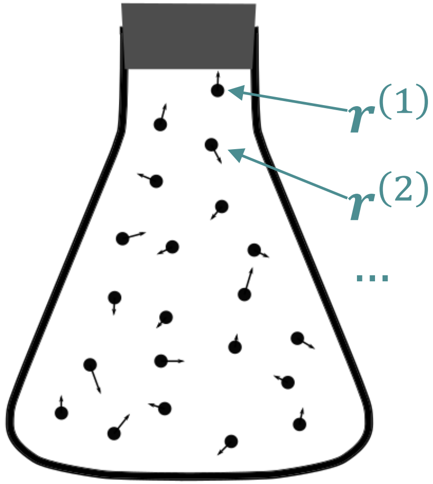
  If in partial equilibrium (meta-stable state), $q(\boldsymbol{z}|\boldsymbol{x}_1) = p(\boldsymbol{z}|\boldsymbol{x}_1)$, then

  $$F[q_{\boldsymbol{x}_1}] = -kT \log \int \exp\left(-\frac{E(\boldsymbol{x}_1, \boldsymbol{z})}{kT}\right) \mathrm{d}\boldsymbol{z},$$

  and $p(\boldsymbol{x}) = \exp\left(-\frac{F[q_{\boldsymbol{x}}]}{kT}\right)/Z$ is preserved.

Distributions in Thermodynamics

The states of molecules $\boldsymbol{r}$ in real world follow a distribution.



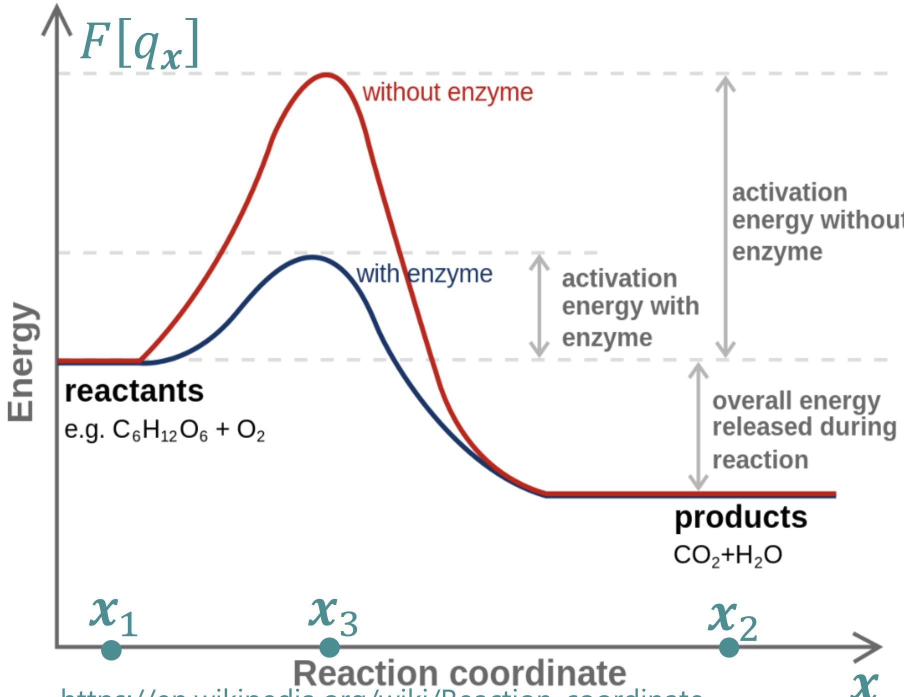
Canonical (const. NVT)

https://en.wikipedia.org/wiki/Ensemble_(mathematical_physics)

Detailed Physics:

- Macroscopic property: $\mathbb{E}_{q(\boldsymbol{r})}[f(\boldsymbol{r})]$.
- (Helmholtz) Free energy: $F[q] = \mathbb{E}_{q(\boldsymbol{r})}[E(\boldsymbol{r})] - Tk\underbrace{\mathbb{E}_{q(\boldsymbol{r})}[-\log q(\boldsymbol{r})]}_{\text{Entropy } S[q]}$.

  If $q_{\boldsymbol{x}_1}(\boldsymbol{r}) = p(\boldsymbol{z}|\boldsymbol{x}_1)\delta_{\boldsymbol{x}_1}(\boldsymbol{x})$, then
  - $F[q_{\boldsymbol{x}_2}] - F[q_{\boldsymbol{x}_1}]$: Reaction free energy ➔ **stability/concentration**.
  - $F[q_{\boldsymbol{x}_3}] - F[q_{\boldsymbol{x}_1}]$: Activation energy ➔ **reaction rate**.



https://en.wikipedia.org/wiki/Reaction_coordinate

Distributions in Thermodynamics

Querying detailed physics:

- Macroscopic property: $\mathbb{E}_{q(\boldsymbol{r})}[f(\boldsymbol{r})]$.
- (Helmholtz) Free energy: $F[q] = \mathbb{E}_{q(\boldsymbol{r})}[E(\boldsymbol{r})] - Tk\underbrace{\mathbb{E}_{q(\boldsymbol{r})}[-\log q(\boldsymbol{r})]}_{\text{Entropy } S[q]}$.

Traditional computation:

- $\mathbb{E}_{q(\boldsymbol{r})}[\cdot]$ or $\mathbb{E}_{p(\boldsymbol{r})}[\cdot]$: MD/MCMC / umbrella (importance) sampling.
  ➔ convergence issue (gap between state transition time scale (μs to ms) and affordable simulation time (ps)), auto-correlation / particle degeneracy.
- $\log q(\boldsymbol{r})$: harmonic approximation / direct free energy estimation.
  ➔ coarse approximation (locally second-order) / need to know reaction coord. $\boldsymbol{x}$.

Distributions in Thermodynamics

Querying detailed physics:

- Macroscopic property: $\mathbb{E}_{q(\boldsymbol{r})}[f(\boldsymbol{r})]$.
- (Helmholtz) Free energy: $F[q] = \mathbb{E}_{q(\boldsymbol{r})}[E(\boldsymbol{r})] - Tk\underbrace{\mathbb{E}_{q(\boldsymbol{r})}[-\log q(\boldsymbol{r})]}_{\text{Entropy } S[q]}$.

Deep generative models:

➔ IID sampling: most sample-efficient way.

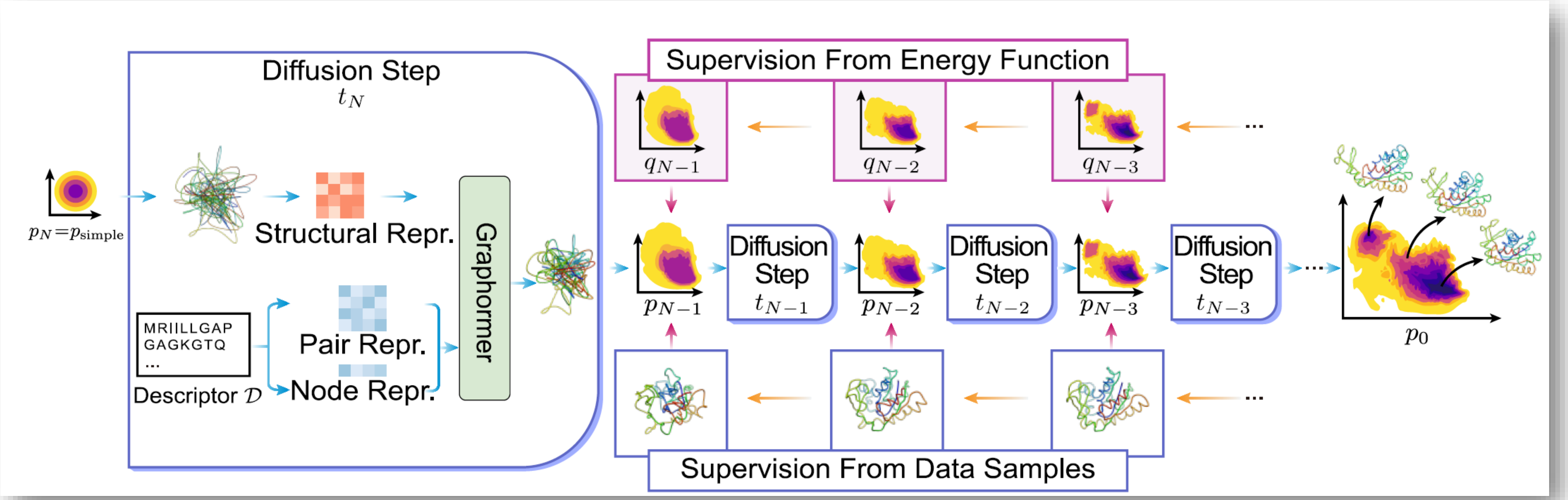
➔ Capability to approximate the complicated distribution.

- Boltzmann generator [Noé et al., 2019, *Science*]:

  ➔ Only applicable to systems with MD data.
- **Distributional Graphormer (DiG)**:

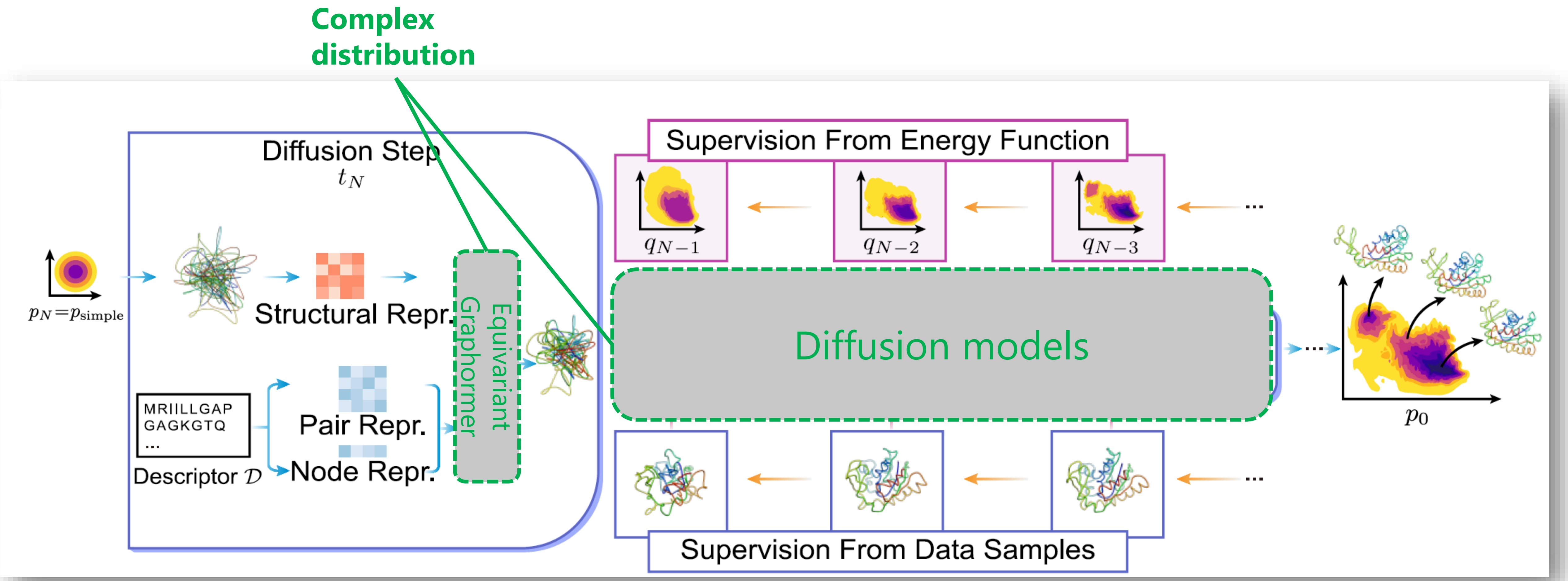
  ➔ Transferable to a range of systems.

Distributional Graphormer



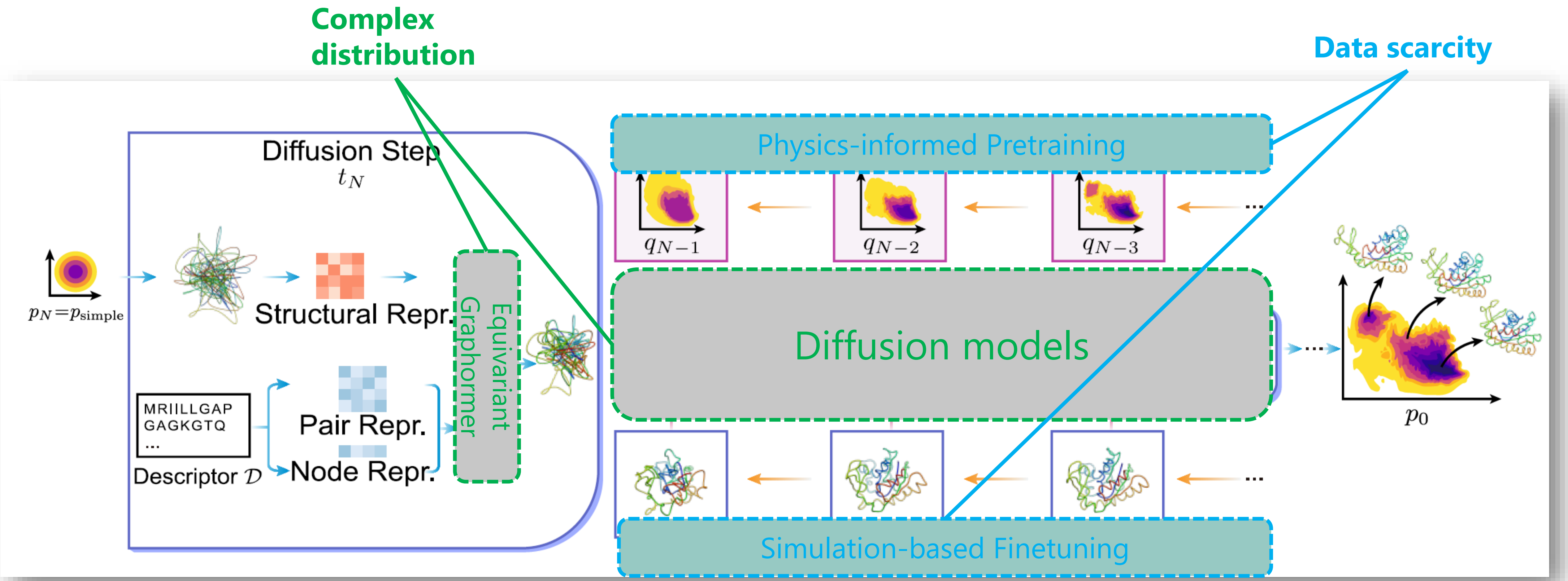
Project Homepage: <https://distributionalgraphormer.github.io/>

Distributional Graphormer



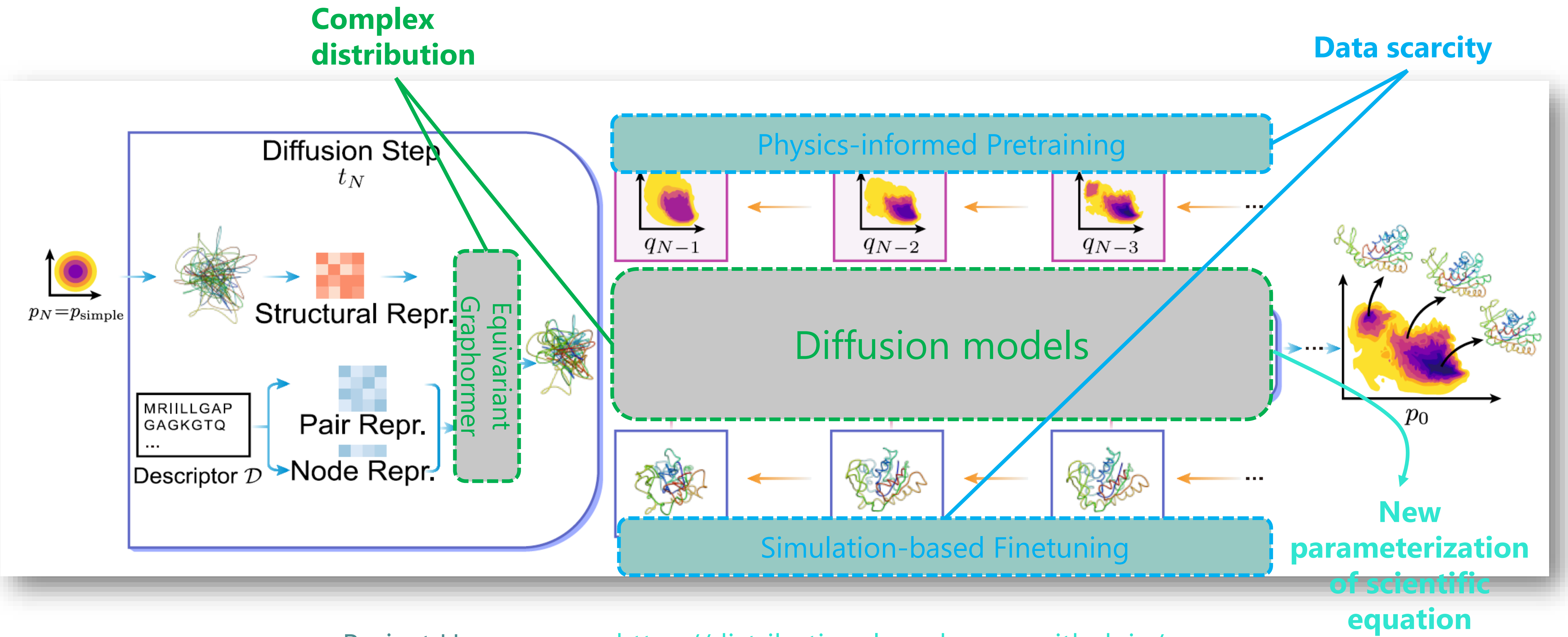
Project Homepage: <https://distributionalgraphormer.github.io/>

Distributional Graphormer



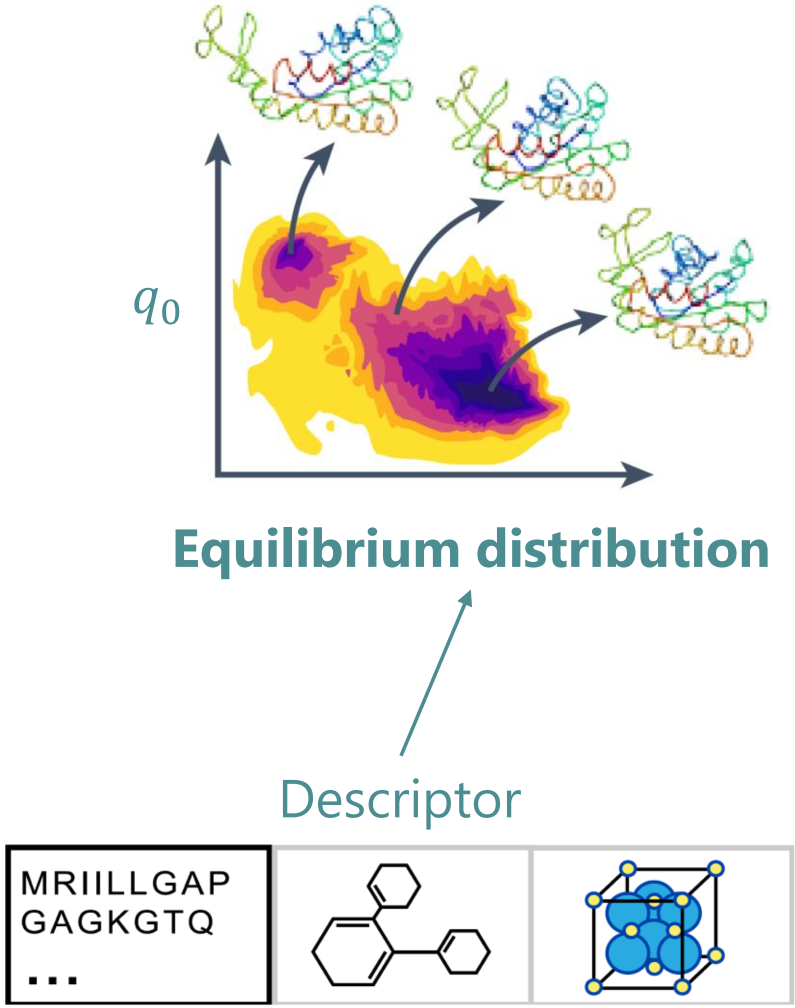
Project Homepage: <https://distributionalgraphormer.github.io/>

Distributional Graphormer

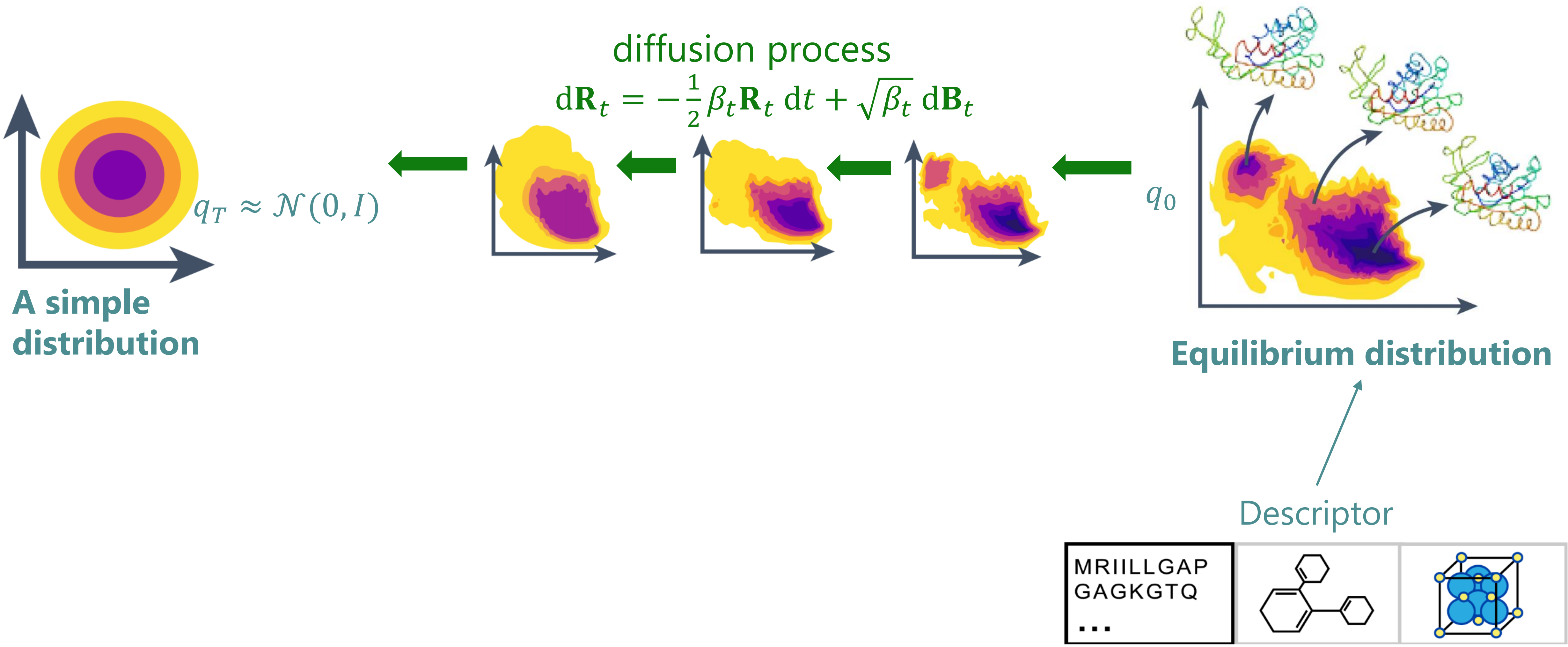


Project Homepage: <https://distributionalgraphormer.github.io/>

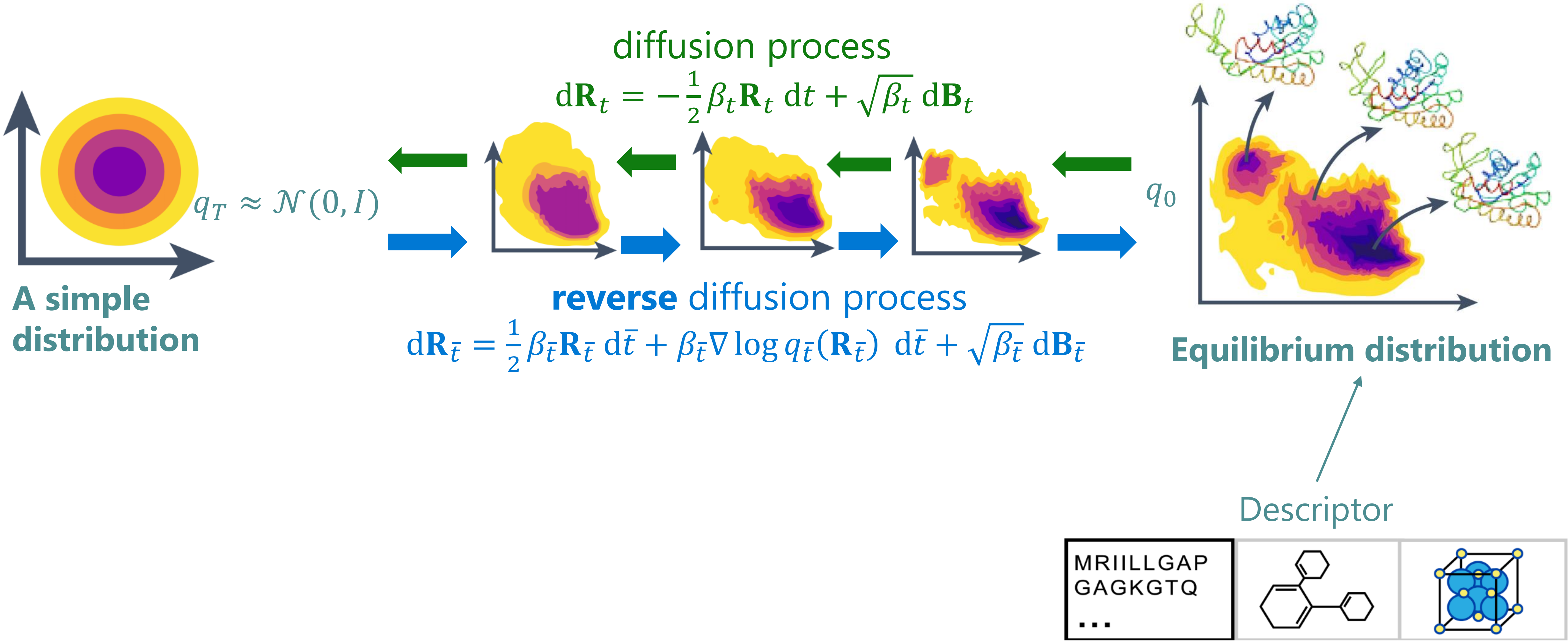
Diffusion Model



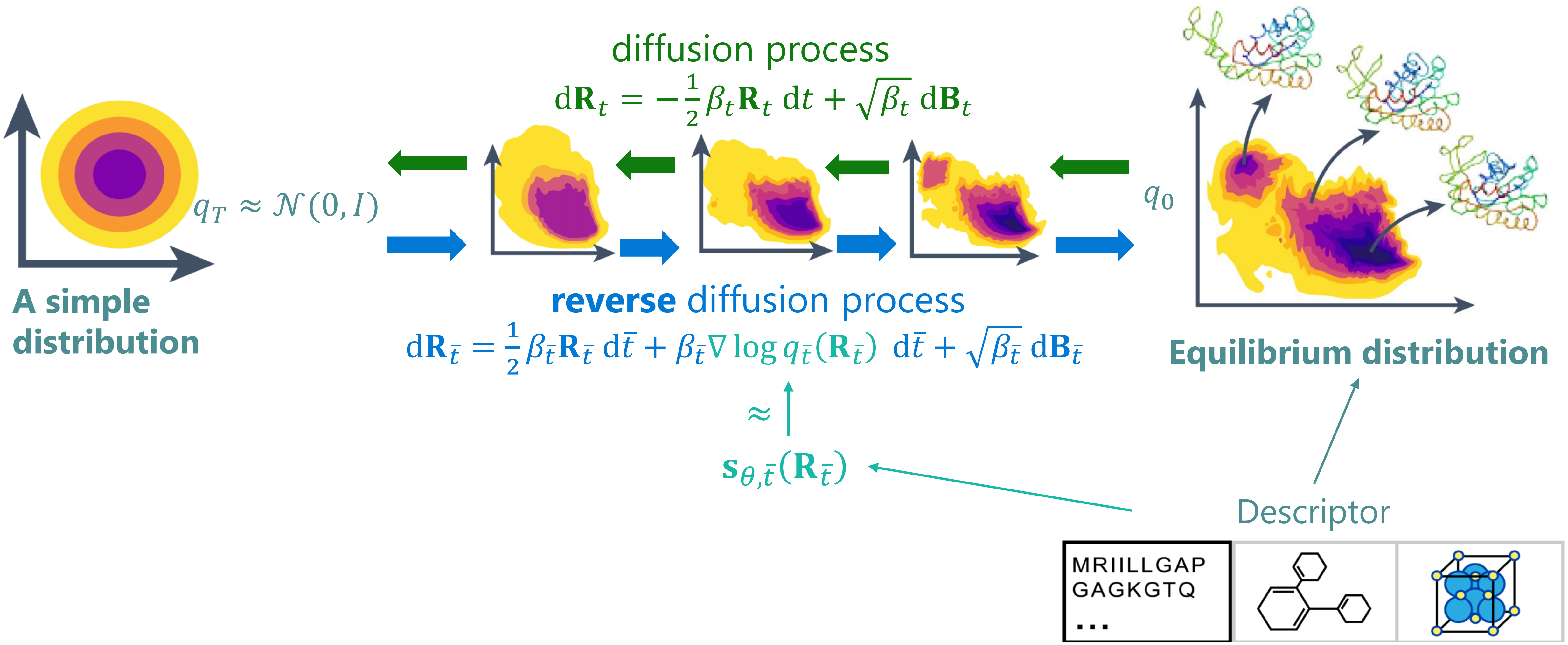
Diffusion Model



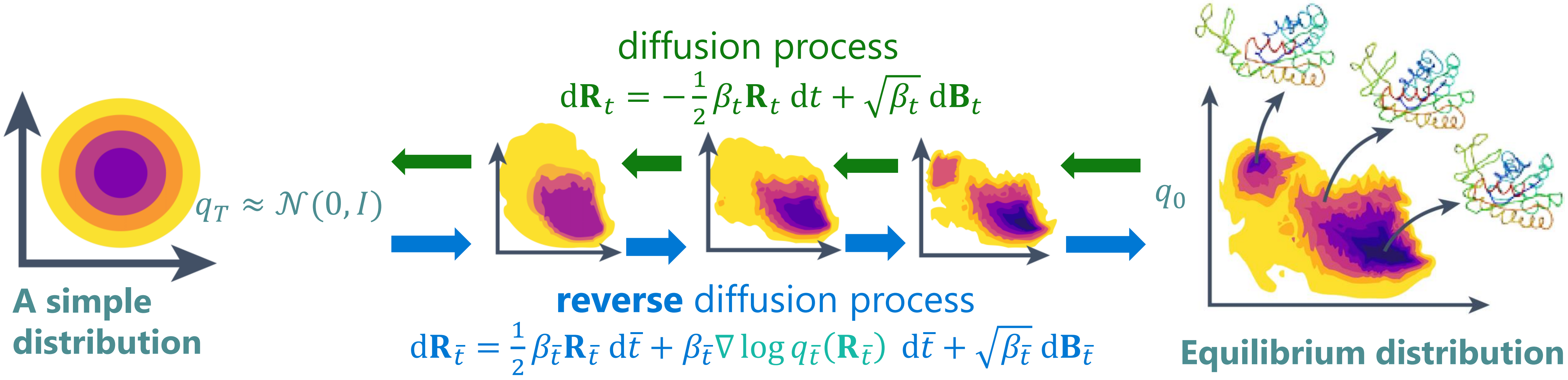
Diffusion Model



Diffusion Model



Diffusion Model



Structure generation → Density calculation → Conditional generation

Capabilities

Structure generation

$$\mathbf{R}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \xrightarrow{\mathrm{d}\mathbf{R}_{\bar{t}} = \frac{1}{2}\beta_{T-\bar{t}}\mathbf{R}_{\bar{t}}\,\mathrm{d}\bar{t} + \beta_{T-\bar{t}}\,\mathbf{s}_{\theta,\bar{t}}(\mathbf{R}_{\bar{t}})\,\mathrm{d}\bar{t} + \sqrt{\beta_{T-\bar{t}}}\,\mathrm{d}\mathbf{B}_{\bar{t}}} \mathbf{R}_0$$

Density calculation

$$\log p_\theta(\mathbf{R}) = \underbrace{\log \mathcal{N}(\mathbf{R}_T; \mathbf{0}, \mathbf{I})}_{\text{density of the simple distribution}} - \underbrace{\int_0^T \frac{\beta_t}{2} \nabla \cdot \left(\mathbf{R}_t + \mathbf{s}_{\theta,t}(\mathbf{R}_t)\right) \mathrm{d}t}_{\text{track and integrate the instantaneous change of density}},$$

where $\mathrm{d}\mathbf{R}_t = -\frac{\beta_t}{2}\left(\mathbf{R}_t + \mathbf{s}_{\theta,t}(\mathbf{R}_t)\right)\mathrm{d}t$ from $\mathbf{R}_0 = \mathbf{R}$.

Conditional generation

$$\mathbf{R}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \xrightarrow{\mathrm{d}\mathbf{R}_{\bar{t}} = \frac{1}{2}\beta_{\bar{t}}\mathbf{R}_{\bar{t}}\,\mathrm{d}\bar{t} + \beta_{\bar{t}}\Big(\mathbf{s}_{\theta,\bar{t}}(\mathbf{R}_{\bar{t}}) + \overbrace{\nabla_{\mathbf{R}_{\bar{t}}} \log p(c|\mathbf{R}_{\bar{t}})}^{\text{property predictor/classifier}}\Big)\,\mathrm{d}\bar{t} + \sqrt{\beta_{\bar{t}}}\,\mathrm{d}\mathbf{B}_{\bar{t}}} \mathbf{R}_0$$

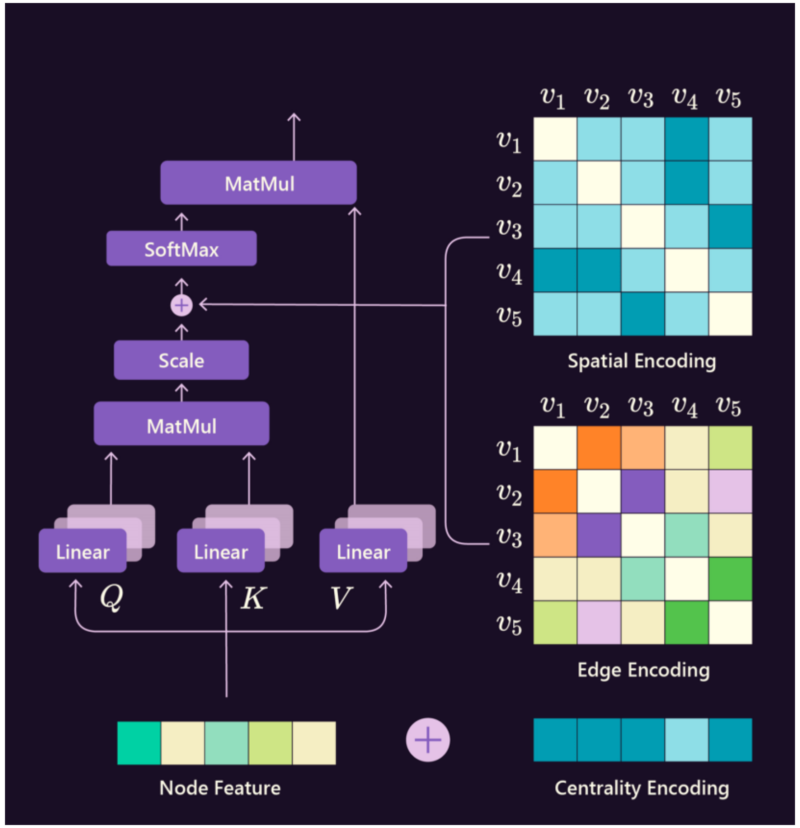
conditional score

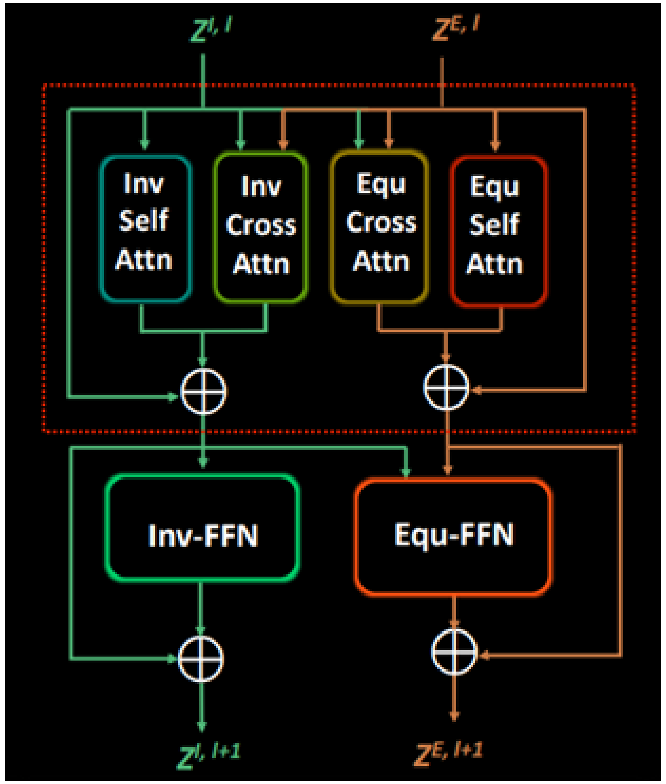
$$\nabla_{\mathbf{R}} \log q(\mathbf{R}|c) = \nabla_{\mathbf{R}} \log q(\mathbf{R}) + \nabla_{\mathbf{R}} \log p(c|\mathbf{R})$$

Equivariant Graphormer

Invariant distribution = invariant prior + equivariant score model.

- Graphormer to encode complex structure
- Minimum inductive bias to ensure equivariance





Challenges in Model Training

Data scarcity

- Hard to collect sufficient experimental or simulation data to well characterize the equilibrium distribution for various systems.

Need stepwise training signals

- Supervision is only available at the end of the diffusion process.
- Backprop through the whole diffusion-process simulation is very costly.

Training from Energy Function & Simulation Data

- Physically-Informed Pre-training

$$\underbrace{\sum_m \left\| \mathbf{s}_{\theta,0}\left(\mathbf{R}^{(m)}\right) + \nabla E\left(\mathbf{R}^{(m)}\right) \right\|^2}_{\text{Energy-function supervision}} + \underbrace{\sum_t \sum_m \left\| \frac{\beta_t}{2} \nabla \left(\left(\mathbf{R}_t^{(m)} + \nabla\right) \cdot \mathbf{s}_{\theta,t}\left(\mathbf{R}_t^{(m)}\right) + \left\| \mathbf{s}_{\theta,t}\left(\mathbf{R}_t^{(m)}\right) \right\|^2 \right) - \partial_t \mathbf{s}_{\theta,t}\left(\mathbf{R}_t^{(m)}\right) \right\|^2}.$$

Energy-function supervision

- Propagate the supervision by consistency with Fokker-Planck equation
- Stepwise loss function

- Training with Simulation Data

Data supervision

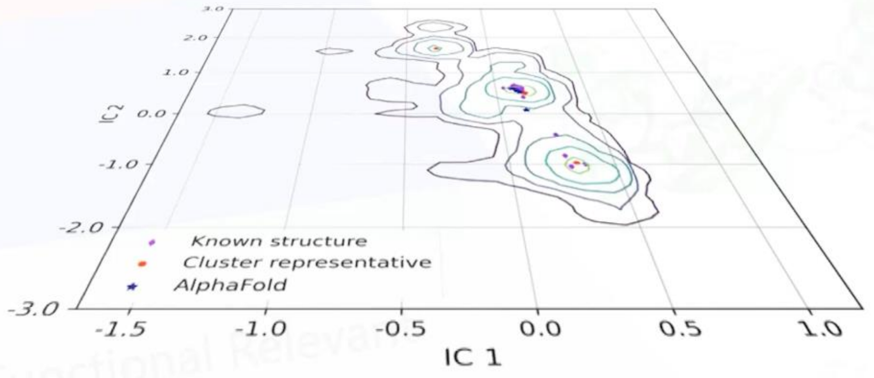
$$\sum_t \sum_n \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\mathbf{I})} \left\| \sigma_t^2 \mathbf{s}_{\theta,t}\left(\alpha_t \overbrace{\mathbf{R}^{(n)}}^{\text{Data supervision}} + \sigma_t \boldsymbol{\epsilon}\right) + \boldsymbol{\epsilon} \right\|^2.$$

Protein Conformation Sampling

SARS-CoV-2 main protease

(PDB id: 6lu7, 306 residues)

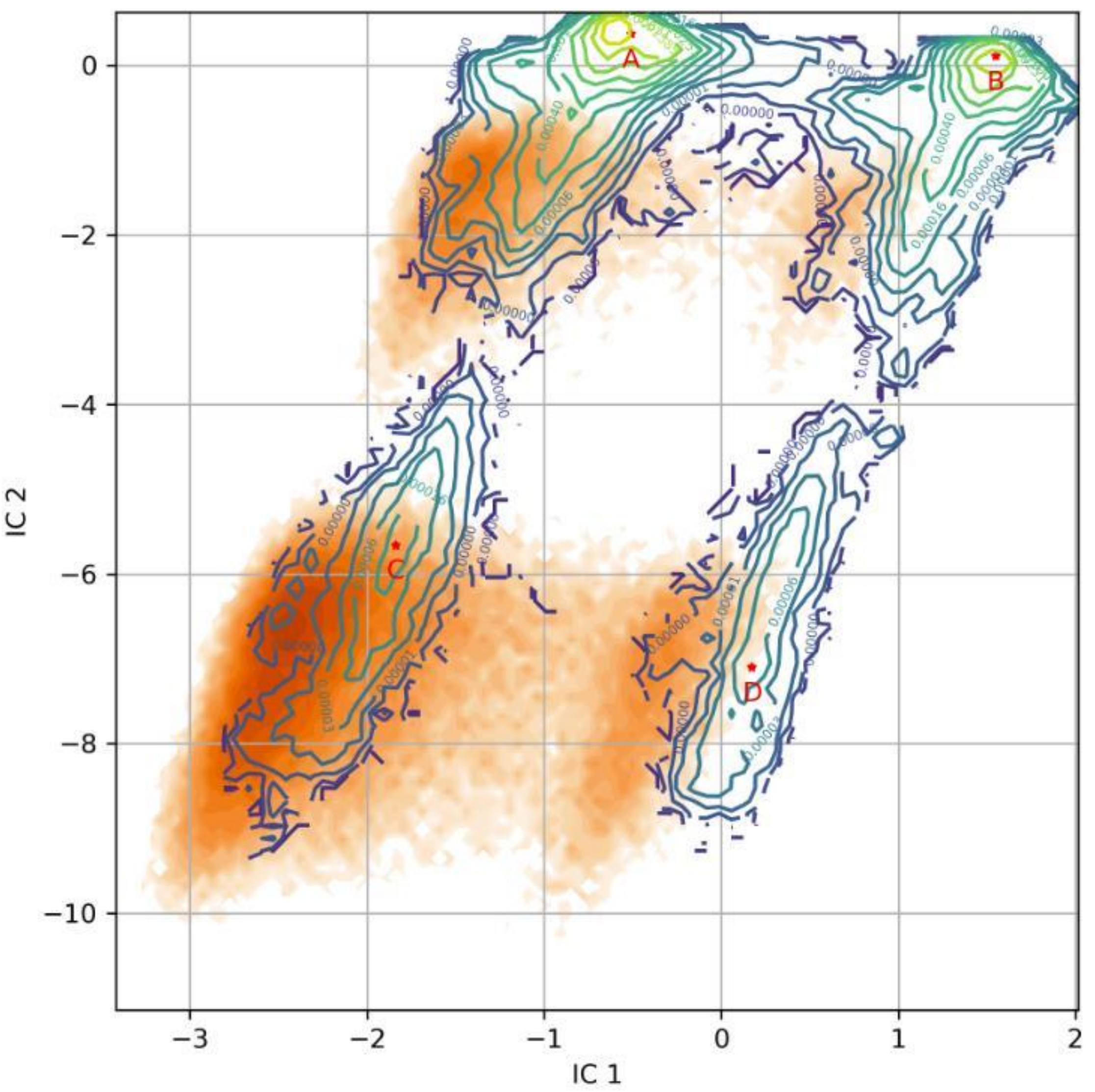
- Ground-truth:
  MD Simulation from
  Folding@home, 2.6 ms.
  Est. **26k GPU days.**
- DiG sampled structures:
  ~40k structures,
  ~18 GPU days.

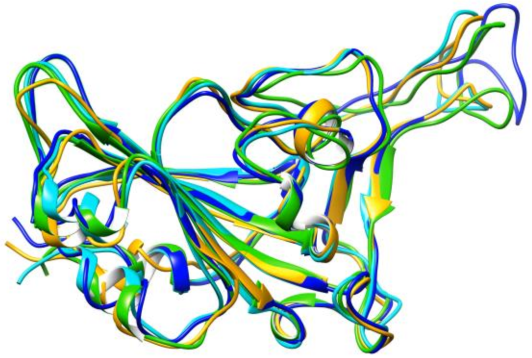


Protein Conformation Sampling

SARS-CoV-2 spike receptor-binding domain
(PDB id: 6m0j, 229 residues)

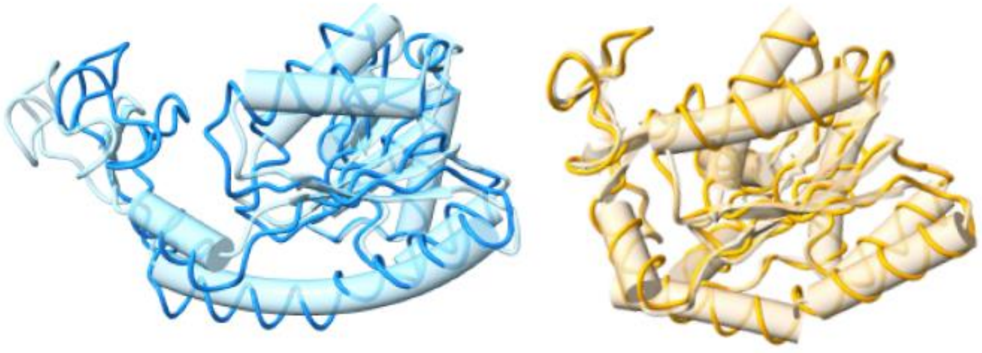
- Ground-truth:
  MD Simulation from Folding@home, 1.8 ms.
  Est. **20k GPU days**.
- EDP sampled structures:
  ~80k structures,
  ~**18 GPU days**.





Sampling Meta-Stable Structures

- Adenylate kinase 腺苷酸激酶

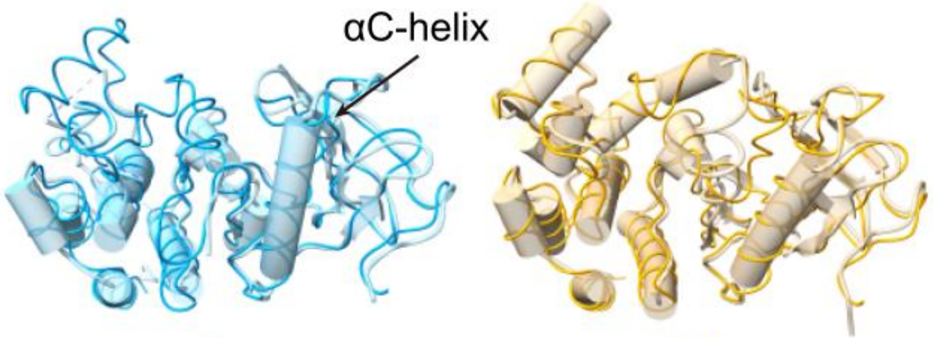


4ake 1ake

- LmrP membrane protein LmrP 膜蛋白

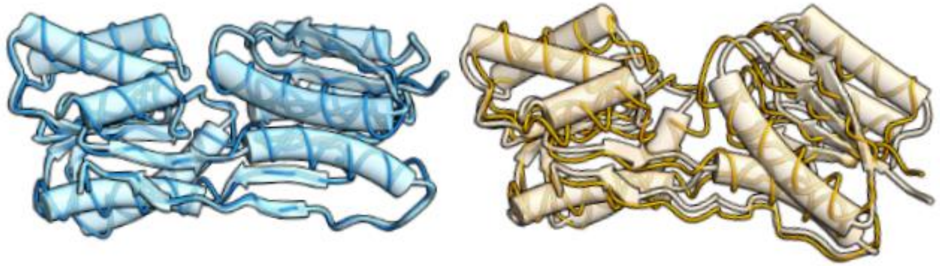
DEER-AF 6t1z

- Human B-Raf kinase 人类 B-Raf 激酶



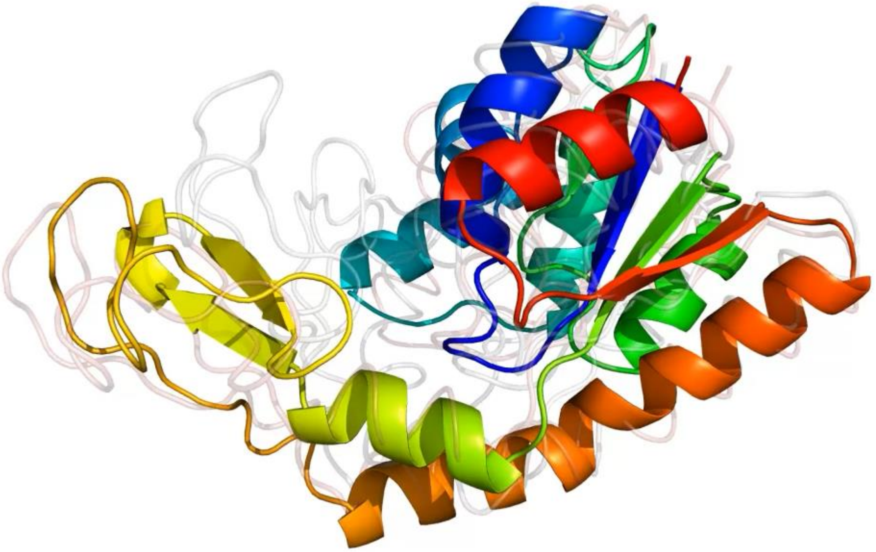
6uan 3skc

- D-Ribose binding protein 大肠杆菌 D-核糖结合蛋白



2dri 1urp

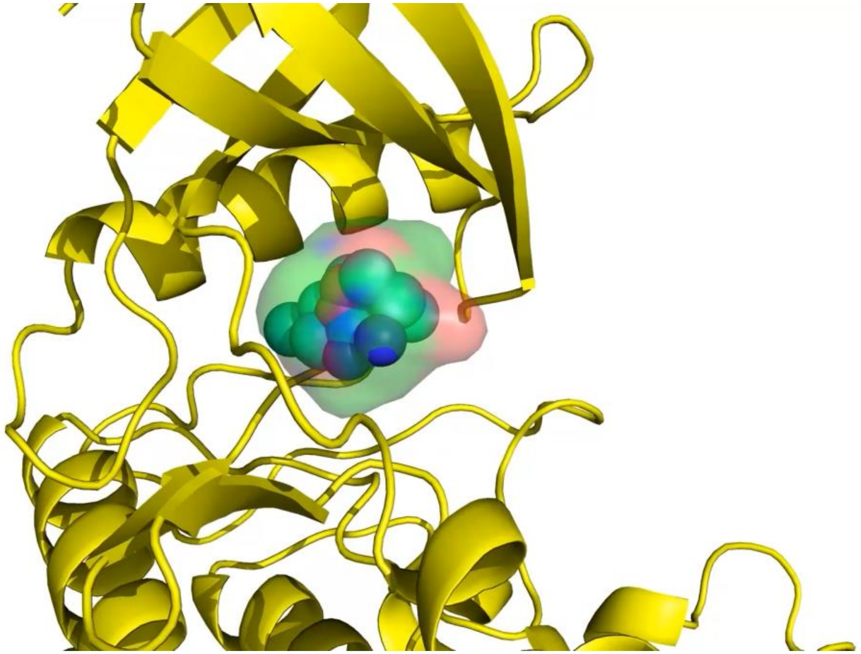
Conformation Transition Pathway Prediction

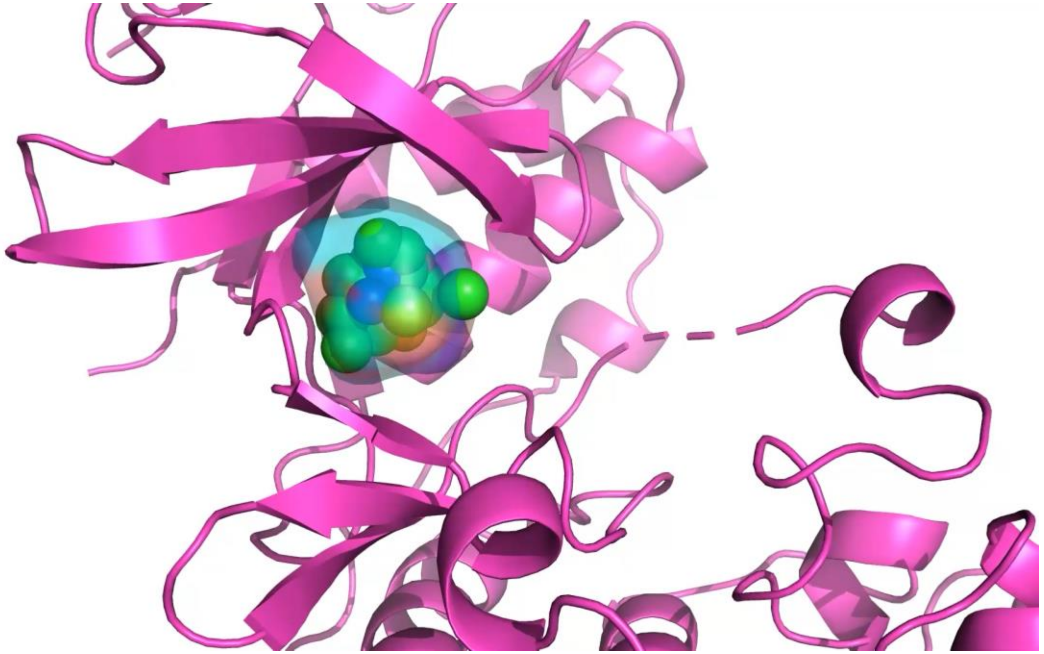


Adenylate Kinase (open ↔ close)
腺苷酸激酶

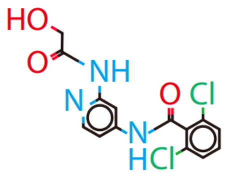
LmrP membrane protein (open ↔ close)
LmrP 膜蛋白

Protein-Ligand Binding Sampling



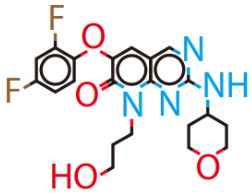


Tyk2 binding with



酪氨酸激酶

p38 binding with



p38 丝裂原活化蛋白激酶

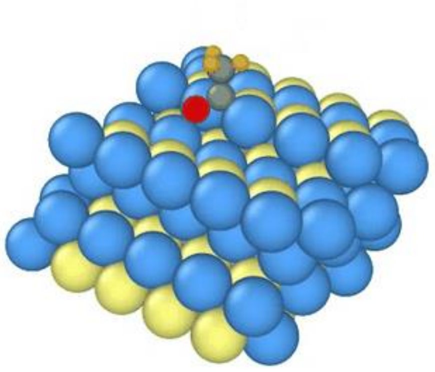
与红细胞和血小板的生成相关，在一些免疫过程中起关键作用

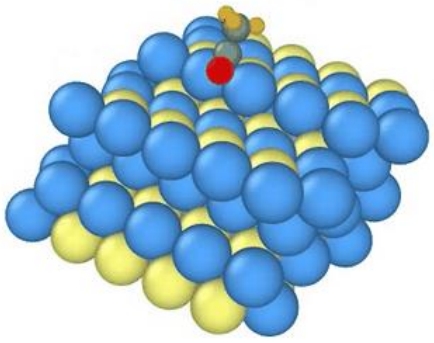
参与细胞生理和病理过程，包括凋亡、应激、进入细胞周期、炎症反应

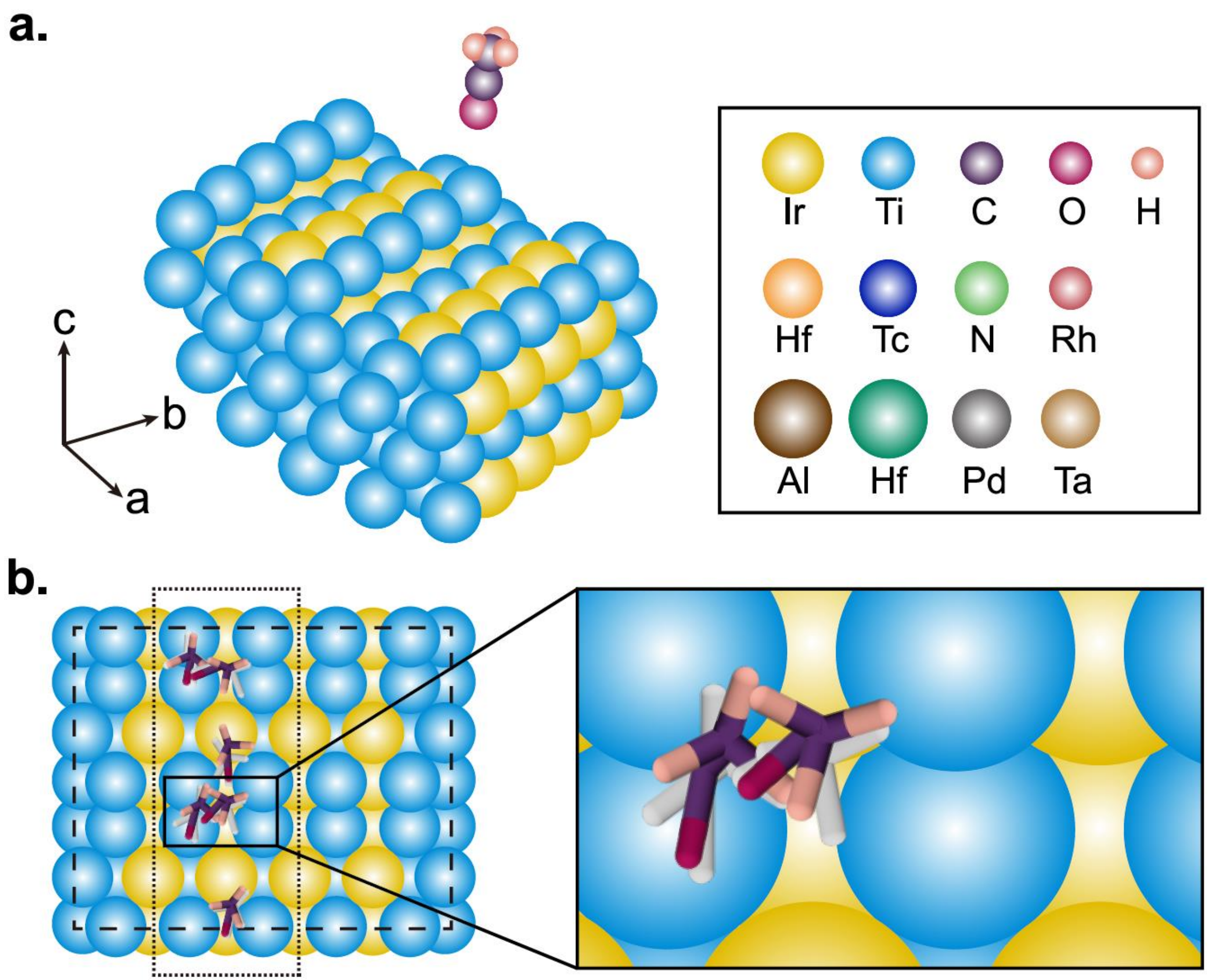
Catalyst Adsorption Sampling

Highlights:

- Generalizable to unseen systems.
- Discovered new adsorption sites (verified by DFT).
- Speed-up:
  - Classical (DFT MD/Relaxation): **days**.
  - DiG: **seconds**.

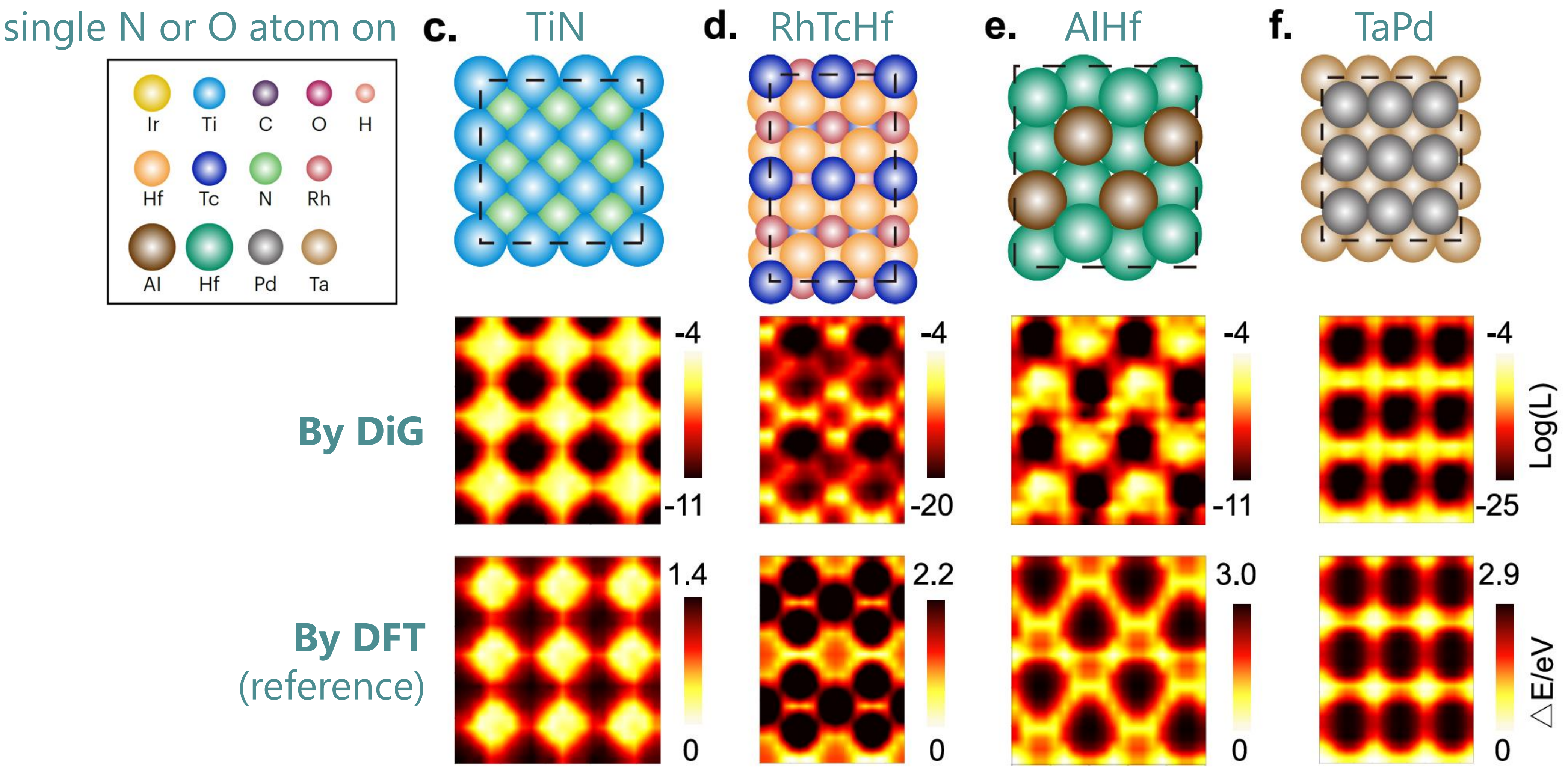






CH3-C-O- on **Ti-Ir** surface

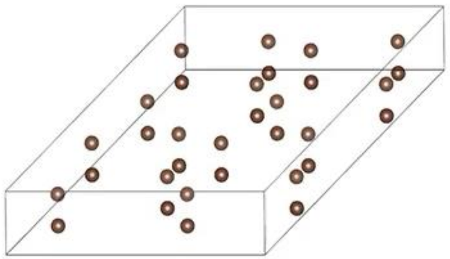
Density Estimation

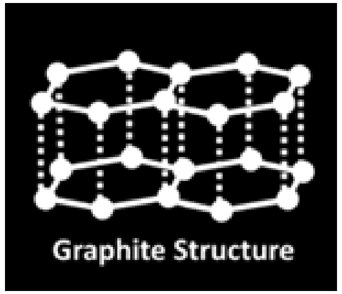


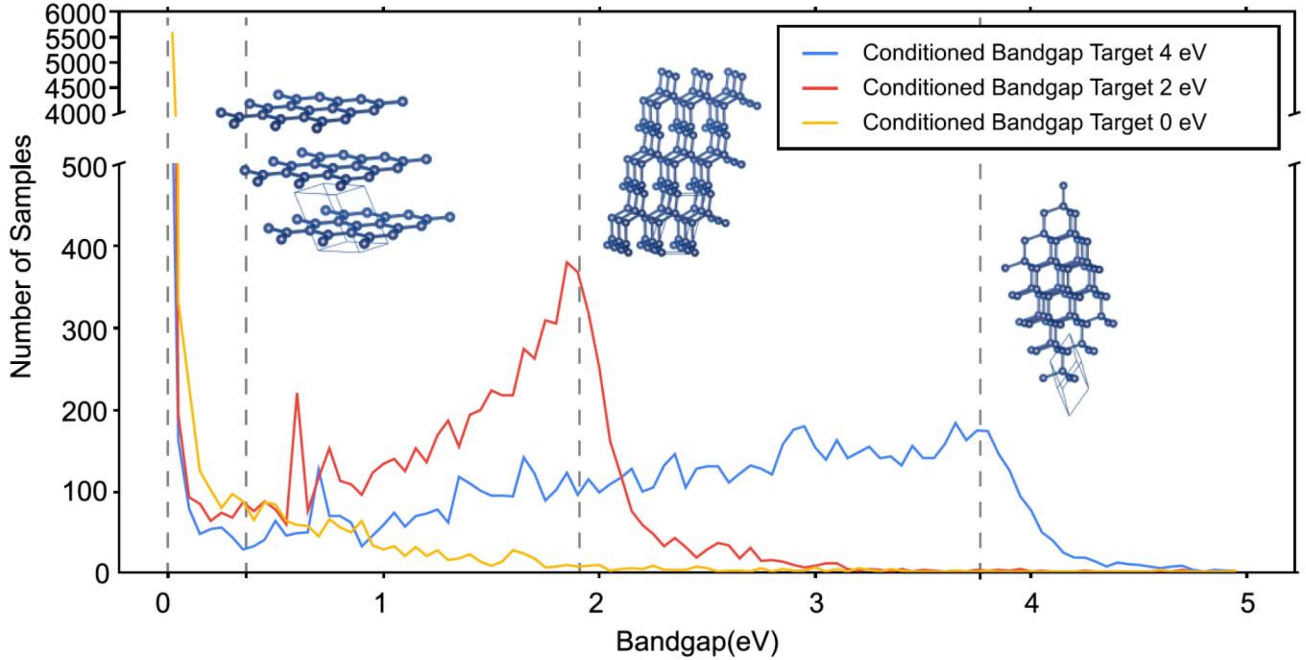
Inverse Design

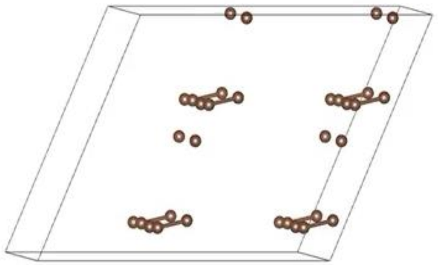
Band Gap = 0 eV → **Graphite**

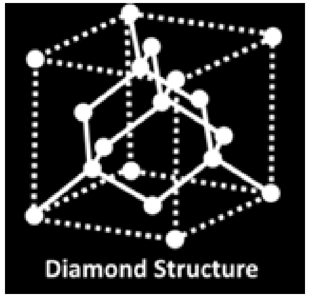
Band Gap = 4 eV → **Diamond**











Microsoft

Thank You