Supporting Information

ECTS: An ultra-fast diffusion model for exploring chemical reactions with equivariant consistency

Mingyuan Xu[1], Bowen Li[2,3], Zhaojia Dong[4], Pavlo O. Dral[5,6,7], Tong Zhu[2,3,8*], Hongming Chen[1*]

[1]*Guangzhou National Laboratory, Guangzhou 510005, China*
[2]*Shanghai Innovation Institute, Shanghai 200003, China*
[3]*Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, 200062, China*
[4]*Institute of Atomic and Molecular Physics, Sichuan University, Chengdu 610065, China*
[5]*State Key Laboratory of Physical Chemistry of Solid Surfaces, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Department of Chemistry, and College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*
[6]*Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, ul. Grudziądzka 5, 87-100 Toruń, Poland*
[7]*Aitomistic, Shenzhen, China*
[8]*AI for Science Institute, Beijing 100080, China*
**Correspondence e-mail: tzhu@lps.ecnu.edu.cn; chen_hongming@gzlab.ac.cn.*

Theory and Methods

S1. Notations and problem definition

Here, we formulate the TS prediction and reaction path generation as a conditional generation problem given reactants and products. Both the reactants (R) and products (P) may consist of multiple molecules. A molecule can be represented as a three-dimensional graph $G = (H, E, X)$, including the node features $H = \{h_v\}$ (such as atomic numbers) for all nodes $\forall v \in V$, edge features $E = \{e_{v,w}\}$ (such as bond types) for all bonded pairs $\forall (v, w) \in B$, and a set of Cartesian coordinates $X = \{x_v\}$, $\forall v \in V$. Thus, the reaction can be expressed as an union of multiple three-dimensional molecular graphs $G_R = \cup\{G_i\}_{i\in R}$ (reactants) and $G_P = \cup\{G_j\}_{j\in P}$ (products), respectively.

We aim to design a generative model for the following three objectives:

1. Model conditional distribution $\rho(X_{\mathrm{TS}}|G_{\mathrm{R}}, G_{\mathrm{P}})$ of TS structures X_{TS} given the 3D graph of

reactants G_{R} and products G_{P}. This allows us to generate plausible TS structures for a specific elementary reaction.

2. Predict reaction energy barrier E_a by analyzing the reactants, products, and the generated TSs.
3. Generate potential reaction pathways for a given reaction.

S2. Background of consistency model.

Recently, diffusion processes based on Denoising Diffusion Probabilistic Models (DDPM), such as OA-ReactDiff[34] and TSDiff,[37] involve a systematic approach where noise from fixed posterior distributions $q(X^t|X^{t-1})$ is gradually introduced until the true TS X_{TS}^0 are effectively obscured over T steps. During the generative process, an initial state X^T is sampled from a standard Gaussian distribution. The TS is then progressively refined using the learned Markov kernels $p_\theta(X^{t-1}|G_{\mathrm{R}}, G_{\mathrm{P}}, X^t)$, which are conditioned on the given reactants G_{R} and products G_{P}. This iterative refinement allows the generation of plausible TS conformations from the Gaussian noise.

Song *et al*. have demonstrated that DDPM-based diffusion process can be described as a discretization process on the time and noise, formatted as a stochastic differential equation (SDE) as defined in Eq (S1):[52]

$$dC = f(X,t)dt + g(t)dw \tag{S1}$$

where the $f(X,t)$ and $g(t)$ are diffusion and drift functions of the SDE. As illustrated in Figure 1(a), the drift term causes the SDE trajectory to resemble Brownian motion as time progresses forward and backward. The rough denoising trajectory requires numerous iterations to achieve plausible transition structures, corresponding to relatively low efficiency.

Karras *et al*. simplified the Eq (1) by setting $f(X,t) = 0$ and $g(t) = \sqrt{2t}$.[53] In this case, the denoising trajectory can be solved using an empirical probability flow (PF) ordinary differential equation (ODE) as shown in Eq (S2):

$$\frac{dC_t}{dt} = -ts_\phi(C_t, t) \tag{S2}$$

Usually, the trajectories of ODEs are smoother than those of SDEs, as depicted in Figure 1b.

Recently, Song et al. introduced consistency models (CMs) that directly map noise points on PF-ODE trajectories to real data, bypassing the reverse-time SDE, which results in a smoother denoising trajectory.[54] CMs require only a few refinement steps (2 to 5) to generate high-quality images. Building upon this, we have extended the CMs to equivariant CM[55] (ie. Ec-Conf) for conformation generation. As a further extension to the chemical reaction field, we propose the ECTS model, an energy co-generated equivariant CM, for predicting TS conformation, and E_a, as well as generation of potential reaction path.

S3. Training and inference details of ECTS.

Consistency training of f_{Ts} and f_E. For a given transition state structure X_{Ts} sampled from the datasets, we use $X_{Ts} + t_{n+1} \cdot z$ and $X_{Ts} + t_n \cdot z$ to represent a pair of data (X_{Ts}^n, X_{Ts}^{n+1}) on ODE trajectory $\{X_{Ts}^t\}_{t\in[\epsilon,T]}$, and minimize the difference between the denoised points to achieve self-consistency of f_{Ts}.

$$\mathcal{L}(\theta_{Ts}) = MSE(f_{Ts}(X_{Ts} + t_{n+1} \cdot z, t_{n+1}), f_{Ts}(X_{Ts} + t_n \cdot z, t_n)) \quad \text{(S3)}$$

Supporting the f_{Ts} denoised transition structures from $(X_{Ts} + t_{n+1} \cdot z, X_{Ts} + t_n)$ are f_{Ts}^n and f_{Ts}^{n-1}, the training objective of f_E is as follows:

$$\mathcal{L}(\theta_E) = MSE\left(f_E(f_{Ts}^n, t_{n+1}), f_{\theta_E}(f_{Ts}^{n-1}, t_n)\right) + MSE\left(f_{\theta_E}(f_{Ts}^n, t_{n+1}), E_{Ts}^0\right) \quad \text{(S4)}$$

Similar to Ec-Conf, an exponential moving average (EMA) technique is also introduced to stabilize the training process and enhance the final performance of $f_{Ts}$and f_E. Here we introduce two additional functions, $f_{\overline{Ts}}$ and $f_{\bar{E}}$, which utilize the EMA of the original parameters θ_{Ts} and θ_E throughout the training process, denoted as $\theta_{\overline{Ts}}$ and $\theta_{\bar{E}}$. Here, f_{Ts} is referred as "TS online network" and $f_{\overline{Ts}}$ as the "TS target network". The function f_{Ts} is ultimately optimized by minimizing the difference between $f_{Ts}(X_{Ts} + t_{n+1} \cdot z, t_{n+1})$ and $f_{\overline{Ts}}(X_{Ts} + t_n \cdot z, t_n)$. Thus, the loss function of f_{Ts} is reformulated as following:

$$\mathcal{L}(\theta_{Ts}, \theta_{\overline{Ts}}) = MSE(f_{Ts}(X_{Ts} + t_{n+1} \cdot z, t_{n+1}), f_{\overline{Ts}}(X_{Ts} + t_n \cdot z, t_n)) \quad \text{(S5)}$$

Similarly, f_E is referred as "Energy online network" and $f_{\bar{E}}$ as the "Energy target network". The loss function of f_E is reformulated as shown in Eq (S6).

$$\mathcal{L}(\theta_E, \theta_{\bar{E}}) = MSE\left(f_E(f_{Ts}^n, t_{n+1}), f_{\bar{E}}\left(f_{\overline{Ts}}^{n-1}, t_n\right)\right) + MSE(f_E(f_{Ts}^n, t_{n+1}), E_{Ts}^0) \quad \text{(S6)}$$

Here, $f_{\overline{Ts}}^{n-1}$ is redefined as the output of $f_{\overline{Ts}}(X_{Ts} + t_n \cdot z, t_n)$. The parameters of θ_{Ts} and θ_E are updated using stochastic gradient descent, whereas $\theta_{\overline{Ts}}$ and $\theta_{\bar{E}}$ are updated with

an exponential moving average as shown in Eq. (S7), where μ is the decay rate predefined by EMA schedule.

$$\bar{\theta} = \mu\bar{\theta} + (1 - \mu)\theta \quad \text{(S7)}$$

By adopting this approach, we can perform equivariant consistency training to obtain the approximate functions f_{Ts} and f_E as shown on Fig.S1a.

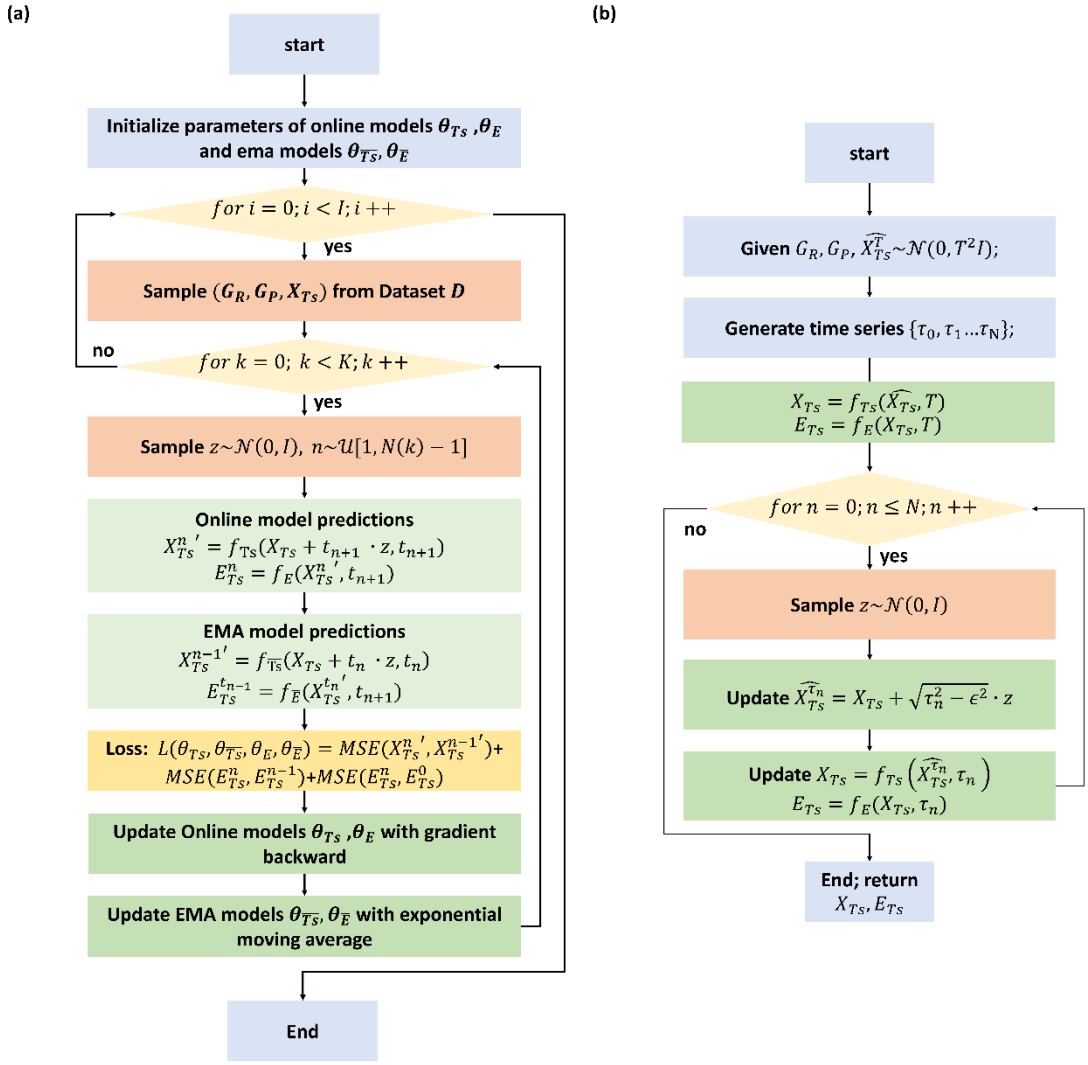

Figure S1. Detailed ECTS work of consistency training of f_{Ts} and f_E (a) and inference for TS generation and E_a prediction(b).

Inference for TS generation and E_a predictions. During the TS inference phase, samples are initially drawn from the initial distribution$\widehat{X_{Ts}} \sim \mathcal{N}(0, T^2 I)$. Subsequently, f_{Ts} generates TS structures $X_{Ts} = f_{Ts}(\widehat{X_{Ts}}, T)$. The conformers are refined with greedy algorithms that alternate between denoising and noise injection steps at a series of time points $\{\tau_0, \tau_2 \ldots \tau_N\}$. Concurrently with the TS generations, the f_E predicts the TS energy at each denoising step

until the max number of iterations is reached. Ultimately, the function ECTS returns the predicted TS structure X_{Ts} and its energy E_{Ts} as shown in Fig. S1b.

S4. Conditional interpolations with ECTS for reaction path generations.

Understanding chemical reaction pathways is crucial for deciphering the mechanisms and subtleties of chemical reactions, including the sequence of bond formation and cleavage, as well as the associated energy changes. In ECTS, an additional replica of f_{TS} is created, denoted as f_{Path}, to learn the conditional distribution of the reaction path relative to reactants and products by training on each point on the reaction path.

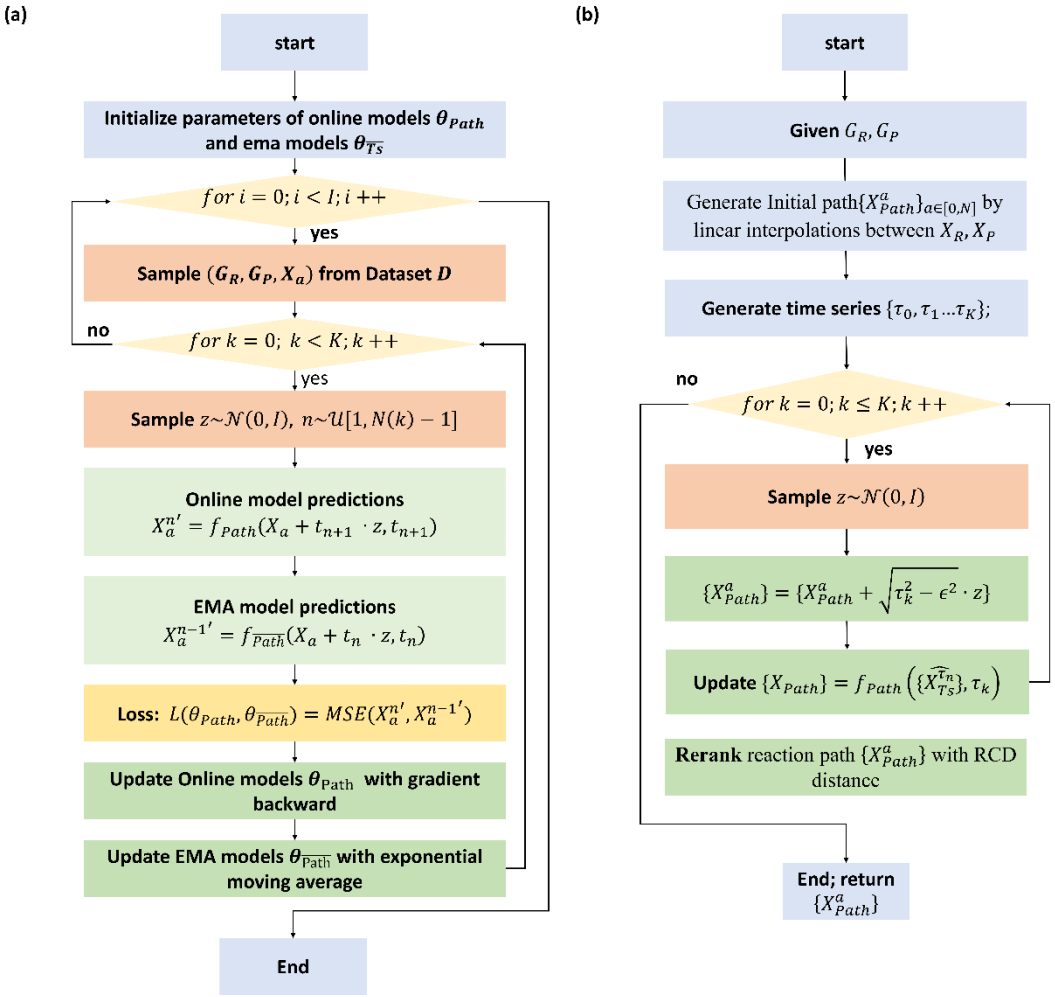

Figure S2. Detailed ECTS work of consistency training of f_{Path} and path inference (b).

Consistency training of $\mathbf{f_{Path}}$. For a given structure on the reaction path X_a sampled from the datasets, we use $X_a + t_{n+1} \cdot z$ and $X_a + t_n \cdot z$ to represent a pair of data (X_a^n, X_a^{n+1}) on ODE trajectory $\{X_a^t\}_{t\in[\epsilon,T]}$. Simialry to f_{Ts} and $f_{\overline{Ts}}$, $f_{\overline{Path}}$ are introduced, utilize the

EMA of the original parameters θ_{Path} throughout the training process, denoted as $\theta_{\overline{Path}}$. f_{Path} is referred as "Path online network" and $f_{\overline{Path}}$ as the "Path target network". The function f_{Path} is ultimately optimized by minimizing the difference between $f_{Path}(X_a + t_{n+1} \cdot z, t_{n+1})$and $f_{\overline{Path}}(X_a + t_n \cdot z, t_n)$ as shown in Eq (S8).

$$\mathcal{L}(\theta_{Path}) = MSE(f_{Path}(X_a + t_{n+1} \cdot z, t_{n+1}), f_{\overline{Path}}(X_a + t_n \cdot z, t_n)) \quad \text{(S8)}$$

The parameters of θ_{Path} are updated using stochastic gradient descent, whereas $\theta_{\overline{Path}}$ are updated with an exponential moving average as shown in Eq. (S7)

Inference for reaction path generation. Conditional interpolation is employed to generate reaction pathways as shown in Figure 6a. For a given reactant-product pair, the linear interpolations$\{X_a\}_{a\in[0,N]}$are taken as the initial path for f_{Path} to generate the reaction pathway. Given that f_{Path} can generate points on the reaction pathway from noise, these linear interpolations serve as intermediate structures in the denoising process as shown in Figure 6a. These intermediate structures are refined to the actual reaction pathways refined with greedy algorithms that alternate between denoising and noise injection steps at a series of time points as illustrated in Figure S2b. However, the order of points on the generated pathway may be disrupted as illustrated in Figure S3a. To address this issue, structures along the reaction pathway are ordered by a devised metrics composed based on reaction coordinates. As shown in Figure 1e, if the bond $B_{i,j}$ between atoms i and j is broken or formed in the reaction, the 3D distance $d_{i,j}$ between atoms i and j is defined as a reaction coordinate. The collection of reaction coordinates $\{d_{i,j}\}_{B^R_{i,j}\neq B^P_{i,j}}$ serves as a reaction coordinate descriptor D to describe the points on the generated reaction path.

For a given point on the reaction path, we define the RCD distance D_{RCD} to represent its positional relationship with respect to the reactants and products:

$$D^a_{\text{RCD}} = \|D_a - D_{\text{R}}\|_2 - \|D_a - D_{\text{P}}\|_2 \,, a \in [0, N] \quad \text{(S9)}$$

where D_{R} and D_{P} represent the reaction coordinate descriptors of the reactants and products, respectively, and N denotes the number of sampled immediate structures.

The smaller D^a_{RCD} values, the point a is closer to reactant and farther from the products. Then a series of generated structures on the reaction path $\{X_a\}_{a\in[0,N]}$, are sorted according to

Eq (S10).

$$\{X_a\} = sort([D^a_{\mathrm{RCD}}]), a \in [0, N] \quad \text{(S10)}$$

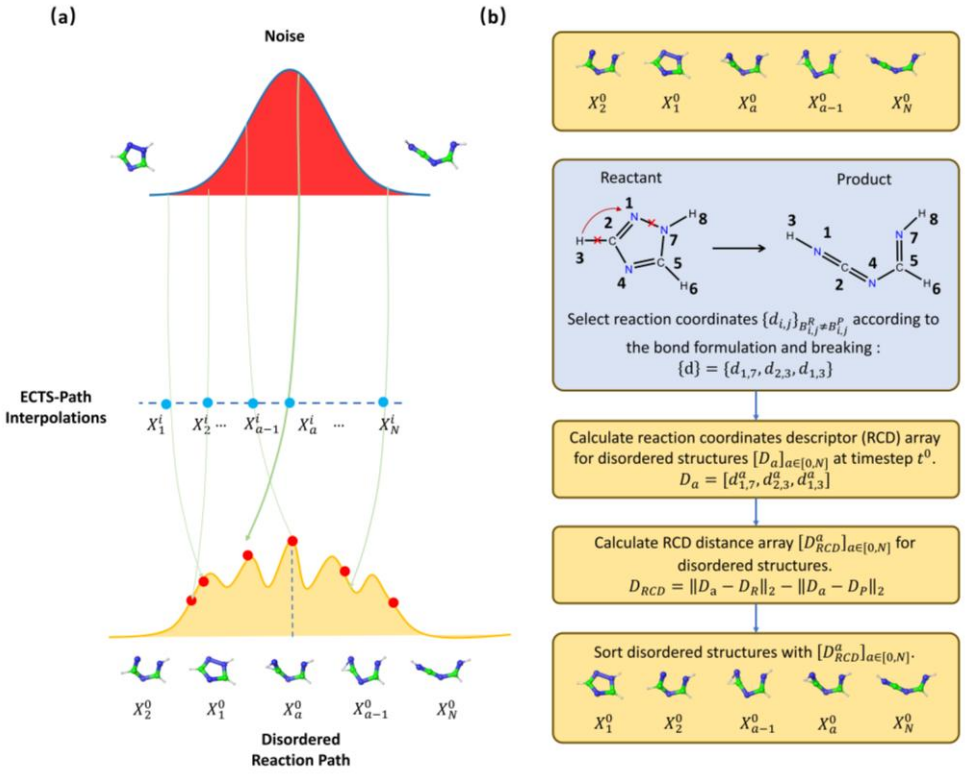

Figure S3. (a) The denoising process between linear interpolation and refined reaction path. The order of points on the generated pathway may be disrupted. (b) The workflow to reorder the points on the reaction path with RCD distance.

Concurrent generation of the TS structure, E_a and reaction path. In ECTS, f_{TS}, f_E, f_{Path} can collaborate in one diffusion process for the concurrent generation of the TS structure, E_a, and reaction path. Starting from the linear interpolation $\{X_a\}_{a\in[0,N]}$ between reactant and product and a series of denoising time points $\{\tau_0, \tau_1 \dots \tau_K\}$, f_{TS} first refines the middle structure $X^{N/2}$ in $\{X_a\}_{a\in[0,N]}$ as the initial guess of TSs. In the following iterations, f_{path} updates the reaction path. The structure closest to the TS guess from the previous denoising step are further refined by f_{TS}, then f_E updates the prediction of E_a. In this way, the reaction path and TS structure are generated by alternating denoising and noise injection steps, as illustrated in Fig. S4. Simultaneously, the reaction energy barrier is also iteratively predicted throughout this process.

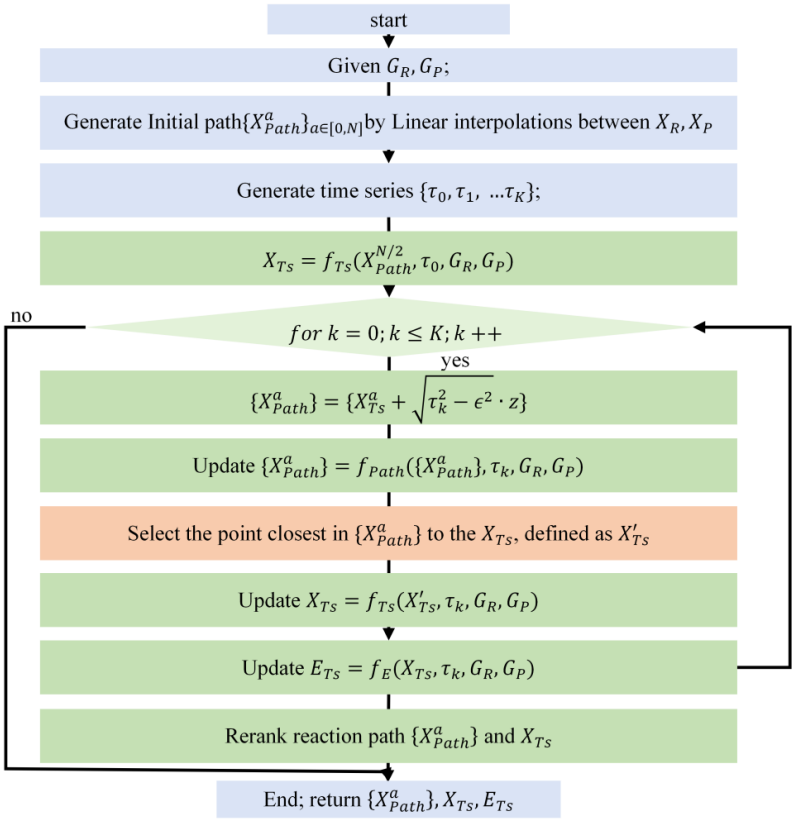

Figure S4. The workflow of ECTS-Path interpolations for reaction path generations.

S5. Detail architecture of DMP layer and E3NN layer.

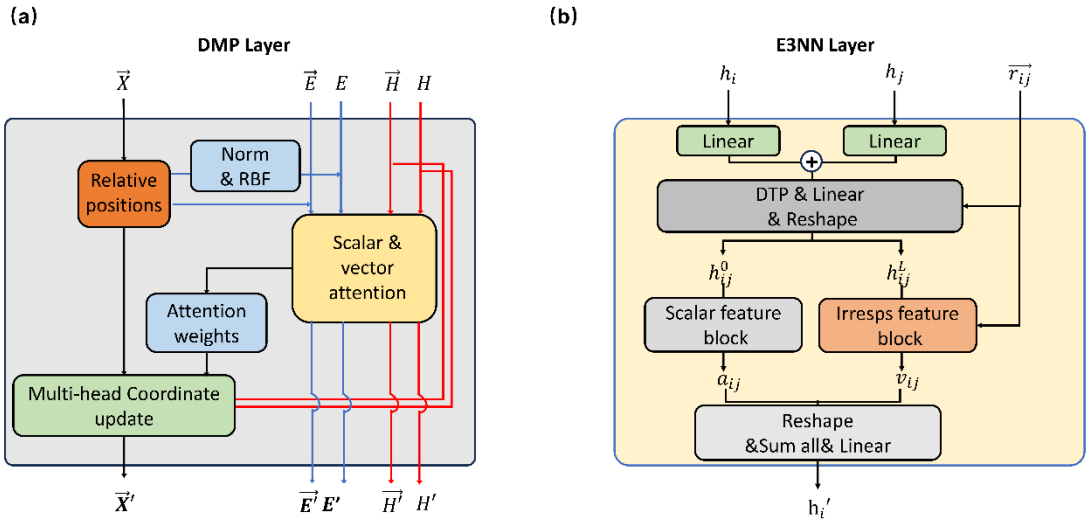

Figure S5. (a) the details of DMP layer. (b) the details of the Equiformer module (E3NN) utilized in energy predictor module.