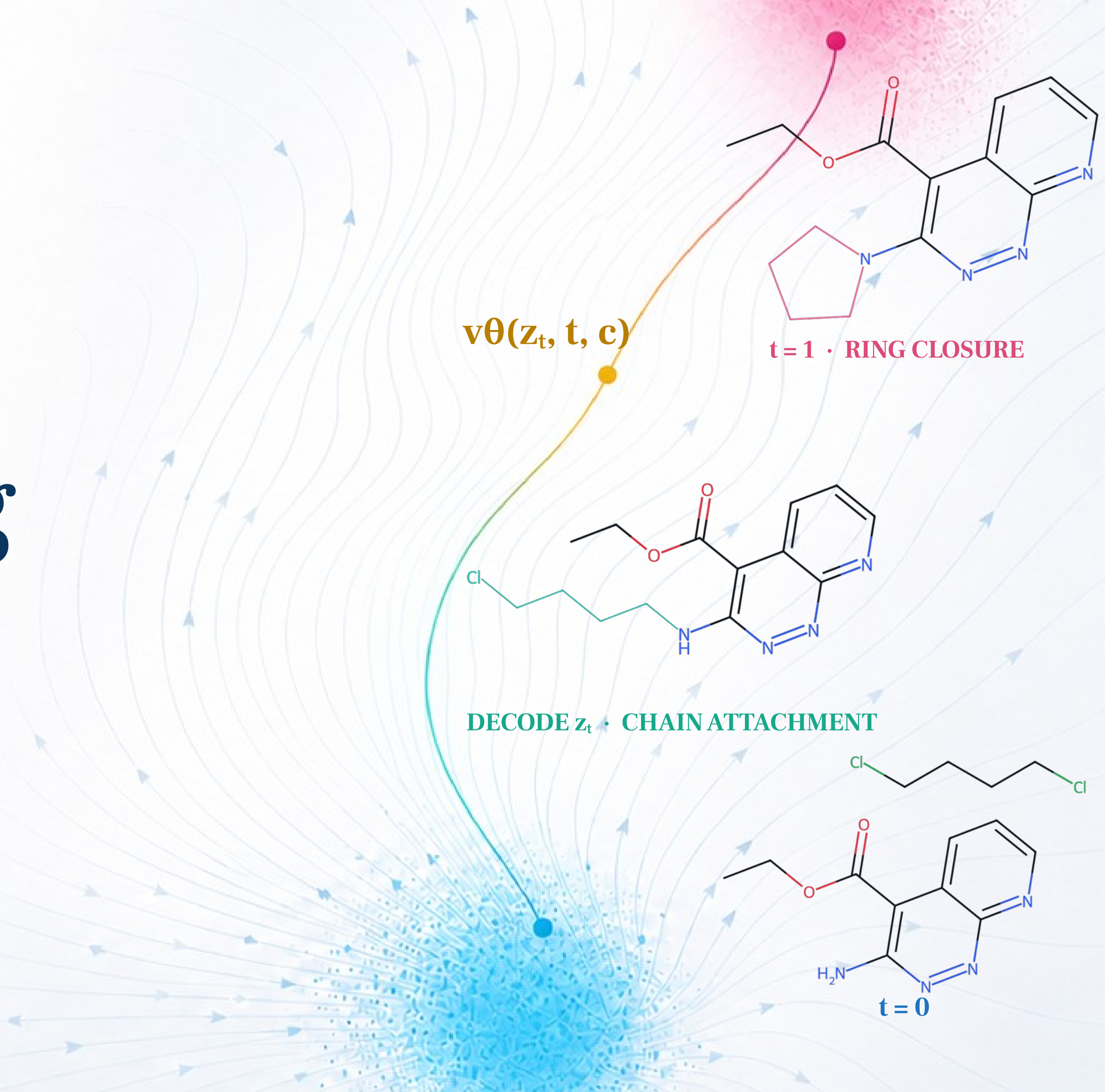
UNIVERSITY OF NOTRE DAME

LATENTRXNFLOW / AI FOR SCIENCES

Driving Reaction Trajectories via Latent Flow Matching

Accurate endpoints. Observable paths.

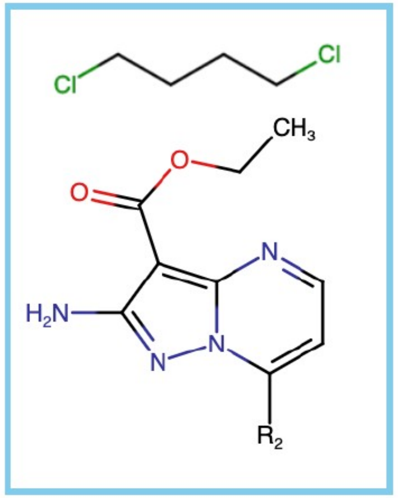


Yili Shen · Xiangliang Zhang

KDD 2026 · AI4Sciences Oral

August 13, 2026 · Jeju, Korea

Can a model solve this reaction?



Reactants R

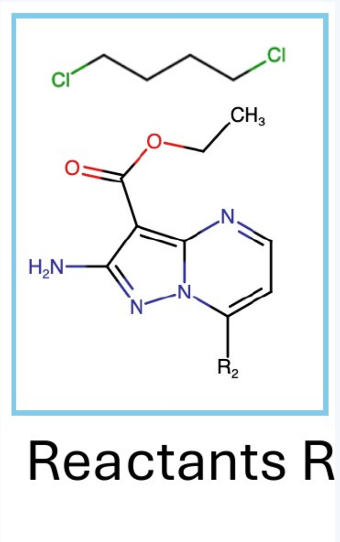
Amine + 1,4-dichlorobutane

?

Predict the product

Given reactants and conditions, output the major product.

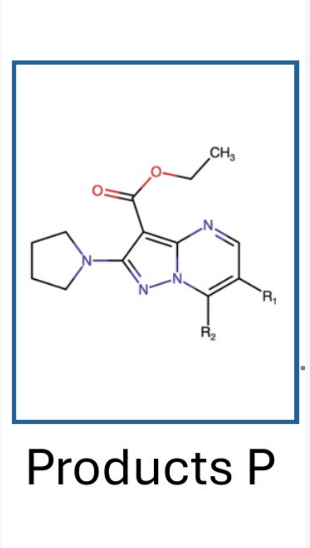
Prediction became translation



Reactants

TRANSLATE

SMILES / graph in
product out



Product

A valid conditional mapping — and a remarkably effective one.

But the transformation itself remains hidden.

The answer hides the path

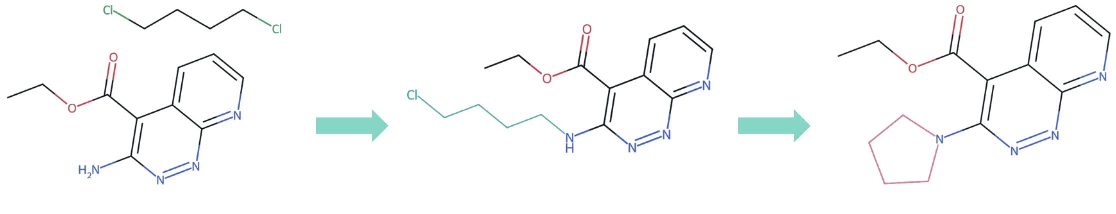
Endpoint-only model

t = 0 · REACTANTS

t = 1 · PRODUCT

Where did the model hesitate?
Did it pass through a better answer?
Why should we trust the endpoint?

Observable trajectory



t = 0

decode z_t · chain attachment

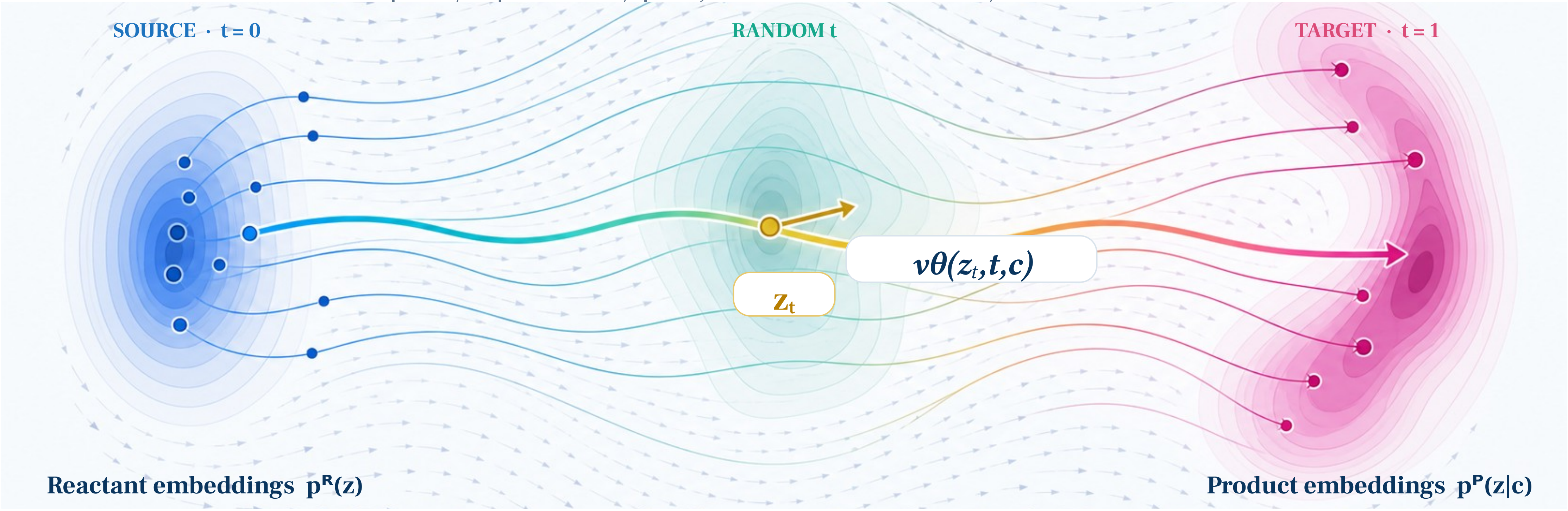
t = 1 · ring closure

Inspect decoded states
Measure trajectory geometry
Intervene before failure

Endpoint models give an answer. Trajectories expose how the model gets there.

Flow matching learns a transport field

Specify a probability path; learn the local velocity that realizes it.



TRAIN Pair endpoints → sample z_t → regress velocity

$$v\theta(z_t,t,c) \approx z^P - z^R$$

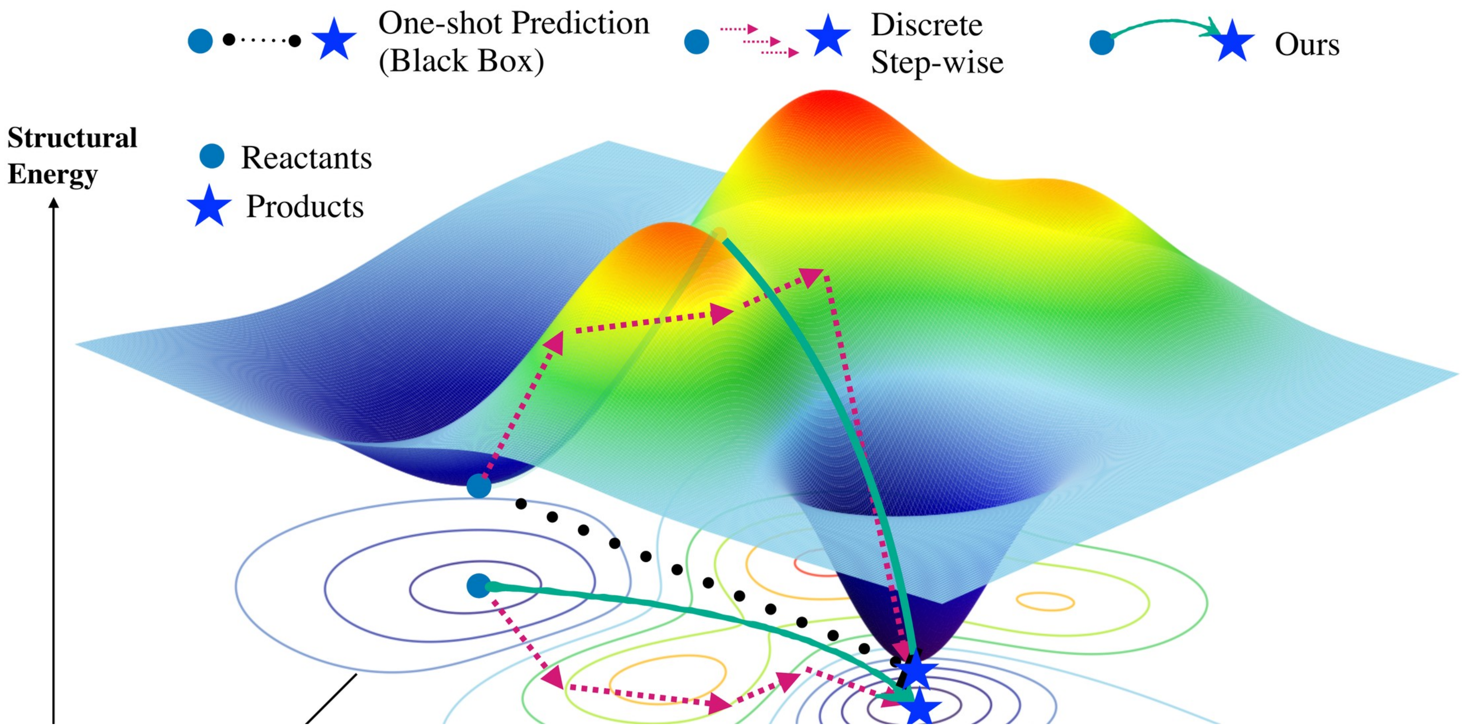
PREDICT Start at z^R → integrate small moves → reach z^P

$$dz_t/dt = v\theta(z_t,t,c)$$

Why reactions fit: paired endpoints define the transport; latent space makes every point on the path continuous and decodable.

Reactions flow in latent space

Encode molecules, then integrate a learned vector field from reactants to products.

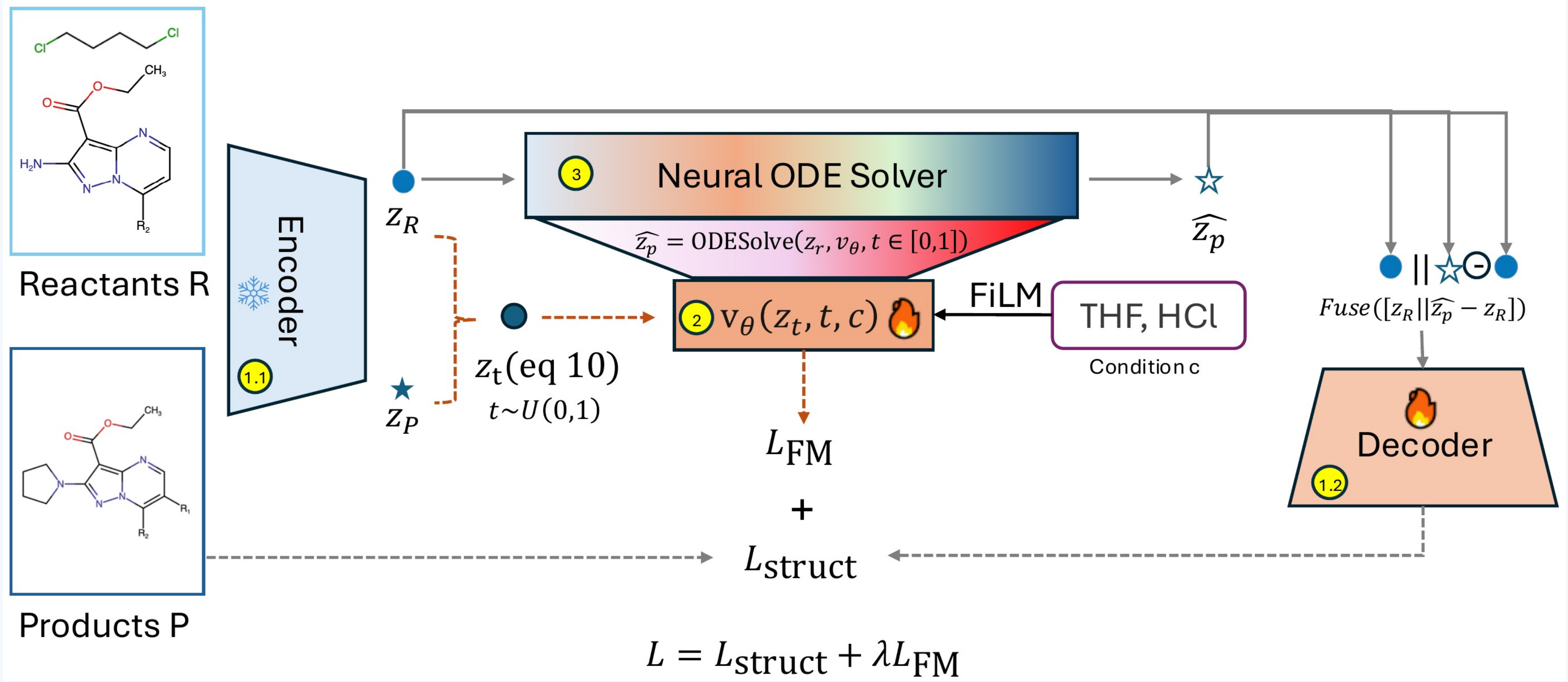


The trajectory is a model-internal reaction coordinate — observable and decodable, but not physical time.

Encode → flow → decode

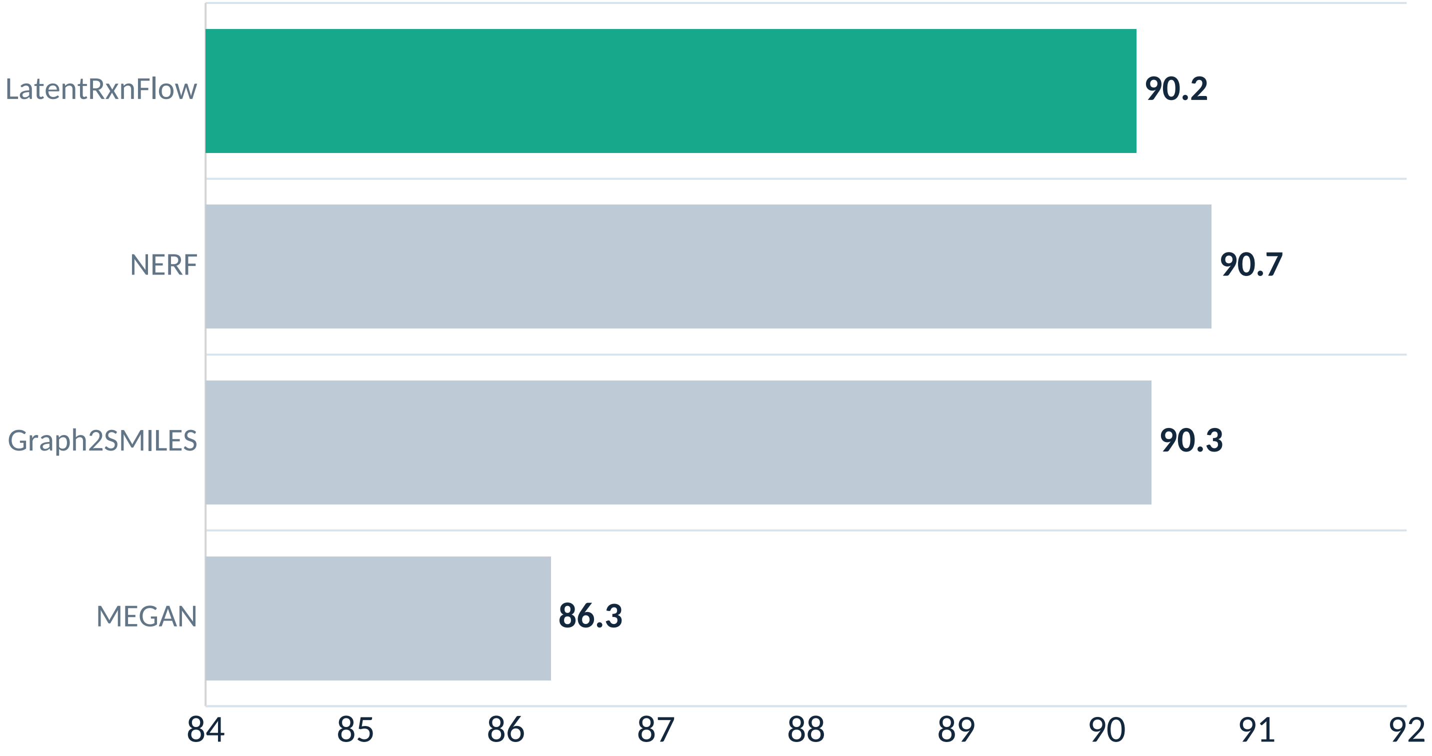
Three coupled components; the ODE trajectory can be decoded at any time.

Training: endpoint pairs only



Trajectories stay competitive at the endpoint

Top-1 exact match (%)



90.2%

Top-1 exact match

94.8%

Top-10 exact match

8.5 ms

per reaction · RK4

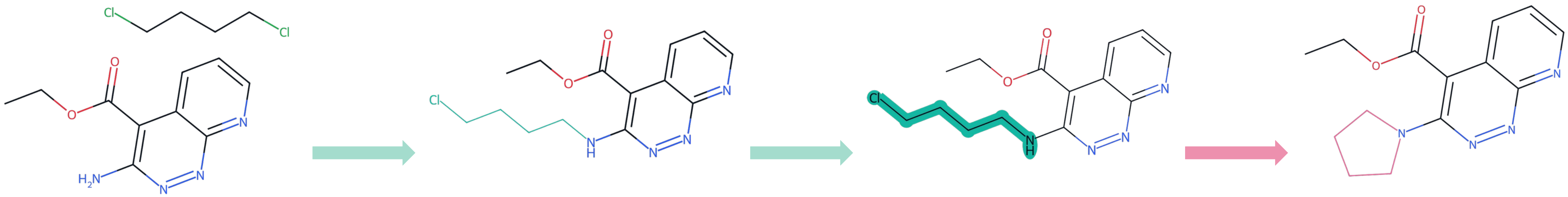
The continuous path does not require an endpoint-accuracy trade-off.

† Baseline values reported in the original papers; FlowER omitted because it uses a different dataset.

Meaningful chemistry emerges without labels

Only the endpoints carry molecular supervision.

SUPERVISED — NO INTERMEDIATE LABELS — SUPERVISED



t = 0	decode z_t	decode $z_{t+\Delta}$	t = 1
reactants	**chain attachment**	**attached state persists**	**ring closure**

No intermediate molecules were provided as targets.

Yet the decoded path organizes into **chain attachment → ring closure**.

DECODED FAILURE TAXONOMY

Some errors reach the product — then lose it

Decode 11 points along each trajectory, then classify what the model actually visited.

ALL 39,984 TEST REACTIONS

36,057 correct · 90.2%

3,927 errors · 9.8%

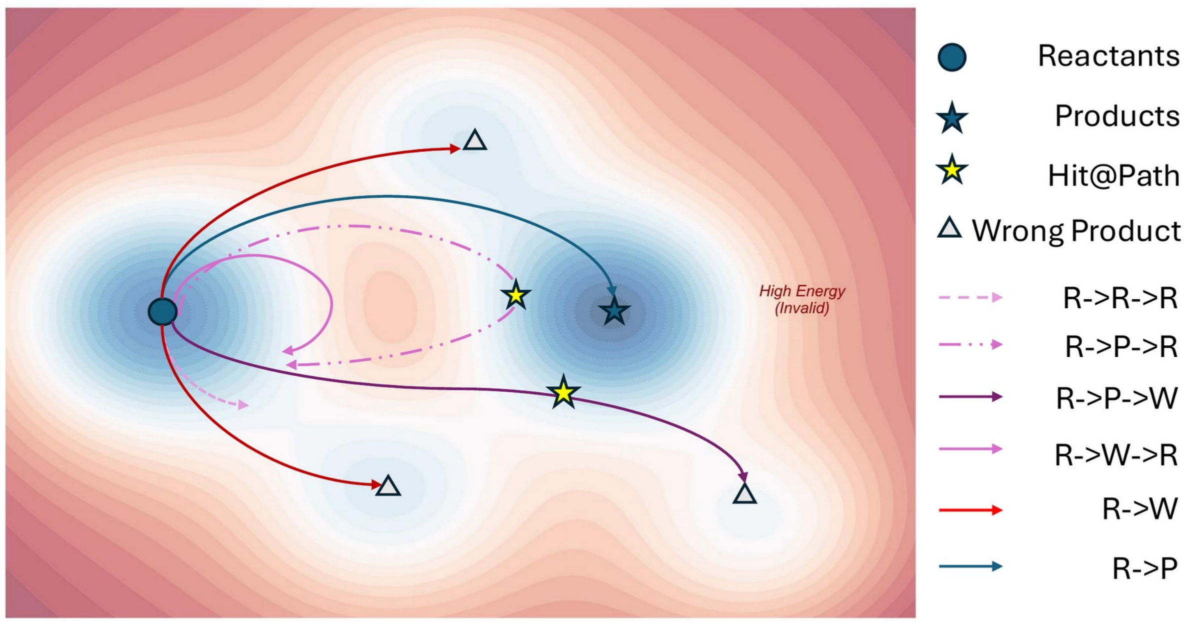
WITHIN THE 3,927 ERRORS

2,995 · 76.2%
never reach product

687 · 17.5%
reach product, then leave

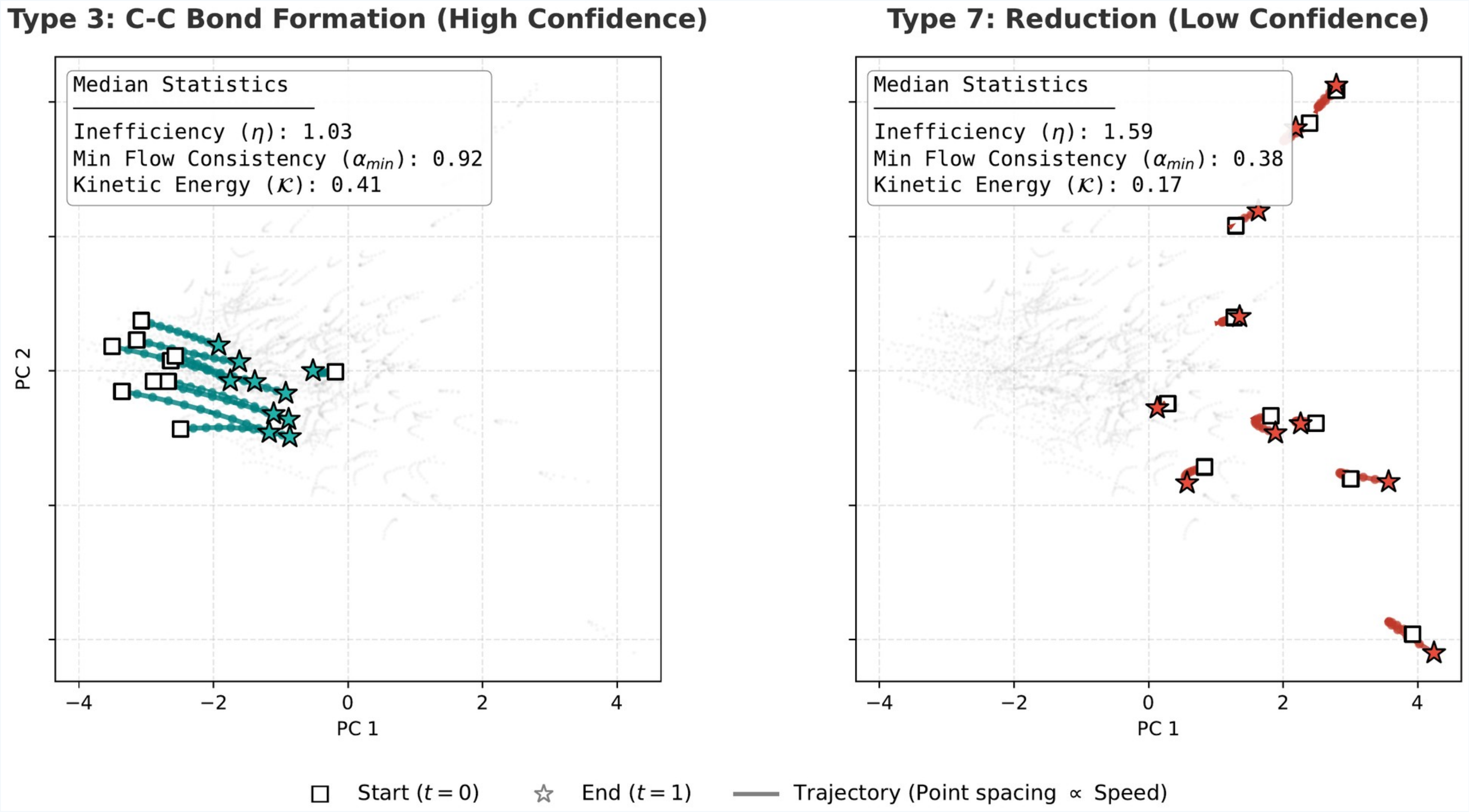
245 · 6.2%
low-reactivity / return

Two potentially recoverable modes
stopping or return can intervene

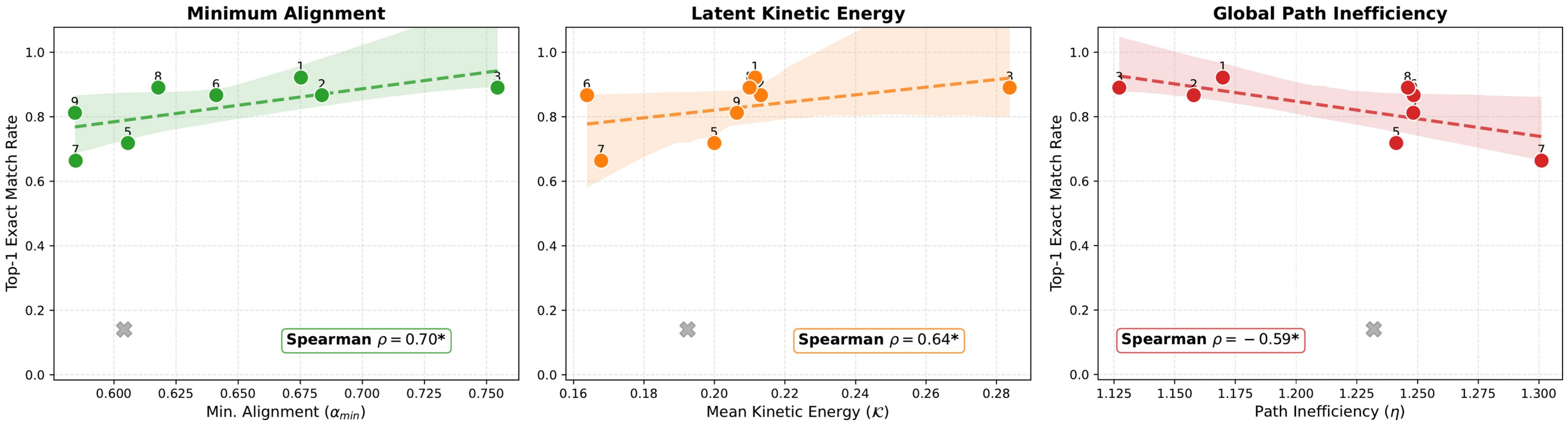


Geometry reveals uncertainty

Confident predictions move **directly and consistently**; uncertain ones detour or stagnate.



Trajectory signals predict reliability



ALIGNED MOTION ↑

steps point in a shared direction

SUSTAINED MOTION ↑

the trajectory keeps moving

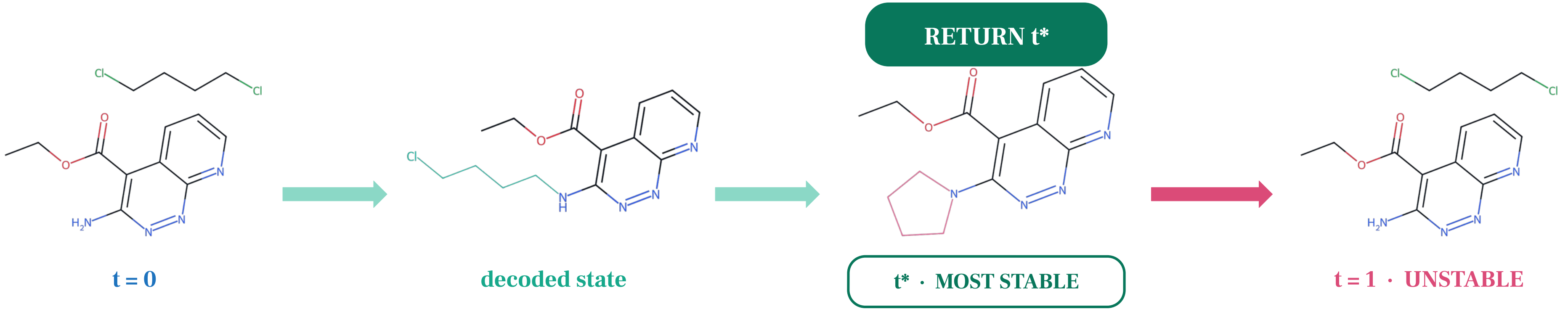
PATH INEFFICIENCY ↓

fewer detours along the path

Reliability is visible in how the path moves — not only where it ends.

Detect overshoot, then rewind

The path itself tells us when the last decoded state should not be trusted.



DETECT OVERSHOOT

$\bar{v}$end $\le$ 0.15 AND final plateau $\ge$ 3 decoded steps

Also trigger if the endpoint is invalid or equals the reactant.

RETURN t*

Candidates: valid, non-reactant decoded states

prefer a longer plateau and lower speed

84.4%

oscillation errors rescued

27 / 32

18.9%

overshoot errors rescued

124 / 655

0.23%

correct endpoints corrupted

82 / 36,057

151 rescued − 82 corrupted = +69 correct reactions, without retraining

KDD2026 August 9–13, 2026 Jeju, Korea

Let reaction models show their work

Latent trajectories turn a prediction into an observable process.

1 Observe

Decode any point along the path

2 Diagnose

Failure modes and confidence from motion

3 Act

Repair or abstain without retraining

NEXT

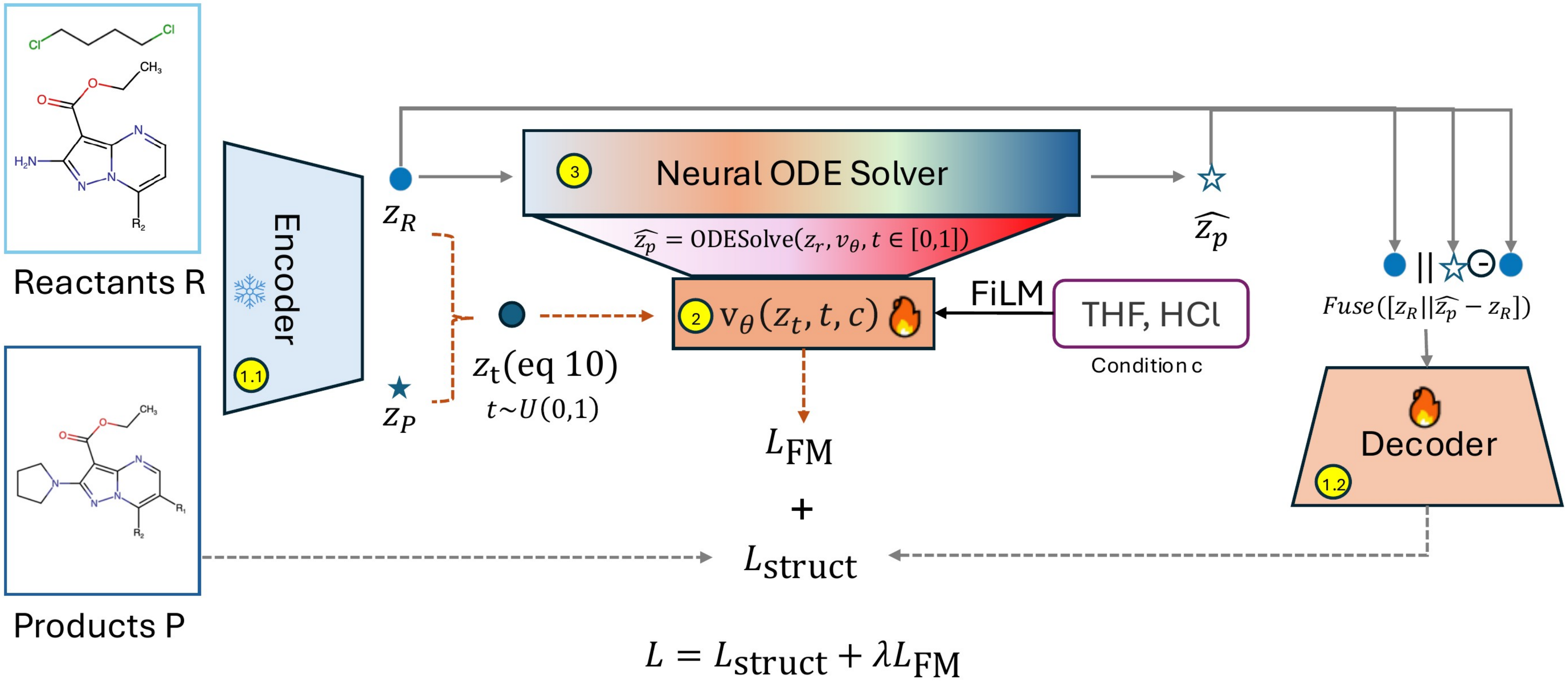
Build an intermediate-state benchmark to evaluate the reasoning path itself.

A conventional model gives the answer.
LatentRxnFlow begins to reveal how it got there.

SCAN & CONNECT

GitHub | Homepage | WeChat | arXiv

Model components and objectives



Objective: L = Lstruct + λLFM · residual bond decoding · FiLM conditioning · RK4 integration

Endpoint accuracy and inference speed

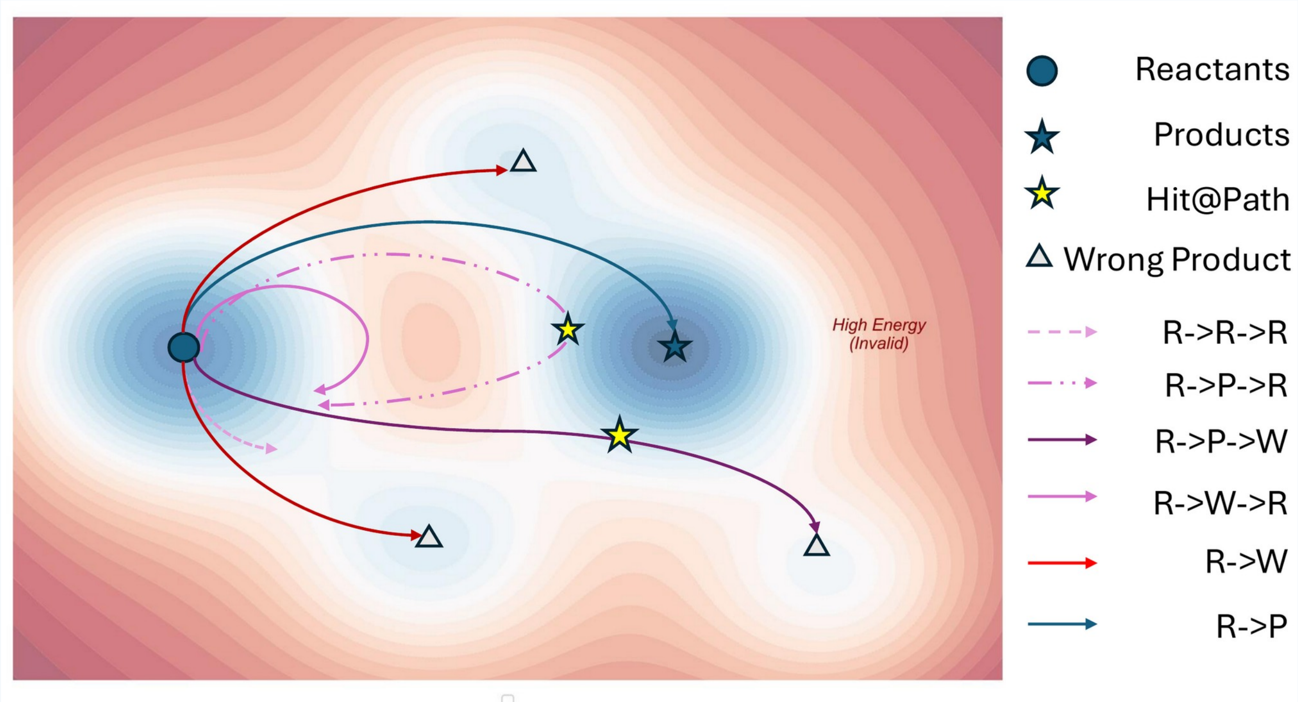
Model	Paradigm	Top-1	Top-3	Top-5	Top-10
Chemformer Large†	Sequence / beam	91.3	–	93.7	94.0
Graph2SMILES†	Hybrid / beam	90.3	94.0	94.6	95.2
MEGAN†	Graph edits	86.3	92.4	94.0	95.4
NERF†	Latent one-shot	90.7	93.3	93.7	–
Ours — no flow	Latent one-shot	89.6	91.8	92.5	92.8
LatentRxnFlow	**Latent ODE flow**	**90.2**	**92.9**	**94.0**	**94.8**

Inference latency on NVIDIA RTX 3090

Model	Solver	Steps	Latency
Chemformer	Beam	≈50	280.8 ms
LatentRxnFlow	RK4	10	8.5 ms

† Reported in original papers

Trajectory taxonomy



R → P

Converges and stays

R → W

Never reaches product

R → P → W/R

Hits product, then leaves

R → R/W → R

Fails to escape reactants

Gating performance by trajectory type

Trajectory type	Count	Trigger	Rescue / corruption	Net
R → R → R Stagnation	114	100.0%	0.0%	0
R → P → R Oscillation	**32**	**100.0%**	**84.4% rescued**	**+27**
R → P → W Overshooting	**655**	**26.0%**	**18.9% rescued**	**+124**
R → W → R Divergence	131	100.0%	0.0%	0
R → W Drift	2,995	13.1%	0.0%	0
R → P Stable	36,057	8.2%	0.23% corrupted	−82

+151
wrong → right

-82
right → wrong

+0.16
Top-1 points

What does trajectory geometry measure?

α

Minimum alignment

Does every local step point toward the global destination?

K

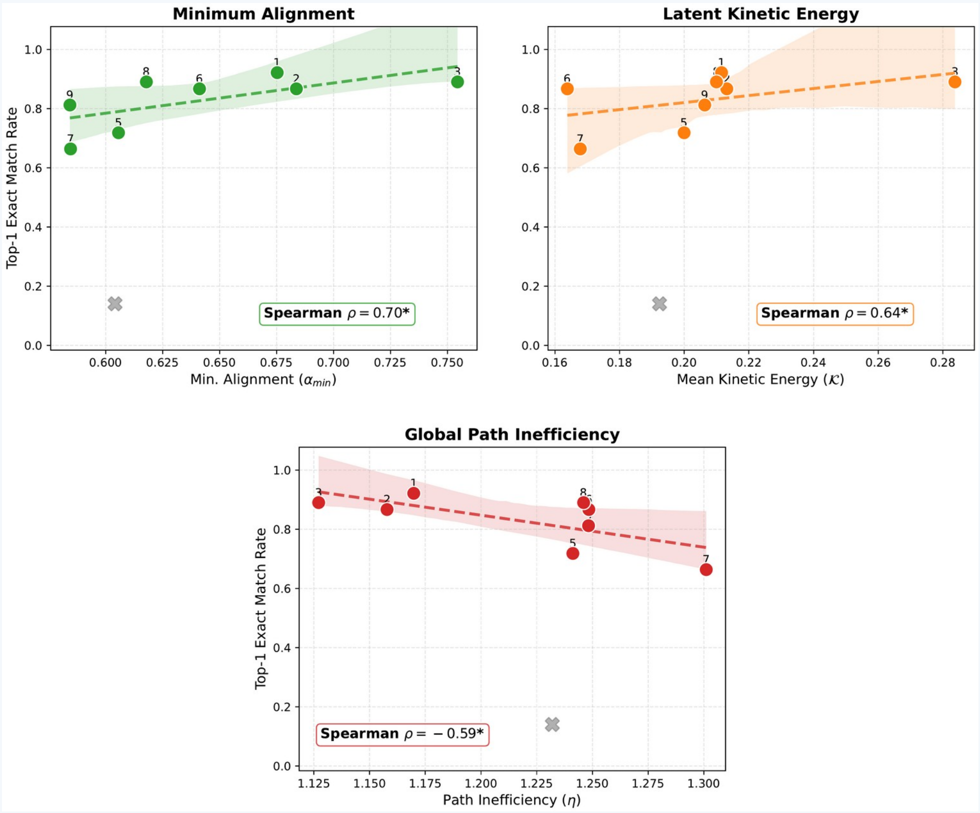
Latent kinetic energy

Does the path maintain decisive forward motion?

η

Path inefficiency

How much longer is the path than the direct displacement?



Straight, aligned, sustained motion indicates higher confidence.