SPECREACT: ROLE-AWARE SPECTRUM-TO-SPECTRUM REACTION MODELING WITH UNBALANCED OPTIMAL TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Chemical reactions are fundamental to functional molecule synthesis and play a central role in drug discovery, materials design, and chemical manufacturing. Spectra provide experimentally accessible and machine-readable observations of chemical transformations, motivating direct reaction modeling in the spectral domain without intermediate molecular-structure reconstruction. However, product spectrum prediction remains challenging because reactions contain an unordered and variable number of inputs with distinct chemical roles, while reactant-to-product spectra exhibit non-conservative changes such as peak shifts, intensity variations, and peak emergence or disappearance. We propose **SpecReact**, a spectrum-to-spectrum framework for predicting product spectra from reactant and reagent spectra. SpecReact employs a role-aware set encoder for variable-sized inputs and introduces an unbalanced optimal transport (UOT)-based spectral transformation module to model spectral correspondences and provide structured supervision for reaction-induced changes. We also construct a new spectrum-based reaction dataset from USPTO reaction records paired with simulated IR spectra. Experiments on simulated and experimentally measured spectra show that SpecReact consistently outperforms representative reaction baselines and remains robust to spectral perturbations. On experimental spectra dataset, SpecReact achieves a CosSim of 0.561 and a peak F1 of 0.517, demonstrating its effectiveness in practical spectroscopic scenarios. Our code can be found at https://anonymous.4open.science/r/SpecReact-A30C/.

1 INTRODUCTION

Chemical reactions are fundamental to the synthesis of functional molecules and play a central role in drug discovery, materials design, and chemical manufacturing Cooper et al. (2014); Lei et al. (2024); Andreo et al. (2022). In practical scenarios, spectra are important tools for molecular structure elucidation and are widely used to monitor chemical reaction processes, providing experimentally accessible and machine-readable signals that reflect molecular vibrations and local chemical environments Xue et al. (2023); Milman (2011). This capability is particularly important in autonomous laboratories, where robotic systems must autonomously assess reaction outcomes and continue experimental workflows with minimal human intervention, especially in experiments where direct human involvement is undesirable for safety reasons Bayley et al. (2024); Seifrid et al. (2022). As illustrated in Figure 1(a), spectra can therefore serve as direct experimental feedback connecting reaction execution and outcome assessment.

Machine learning has achieved substantial progress in chemical reaction prediction, with most existing approaches operating on molecular graphs or SMILES strings Schwaller et al. (2019); Sacha et al. (2021); Irwin et al. (2022). However, when reactions are observed through spectra, directly resolving the structures of reactants may introduce structural errors or incomplete functional group recovery, which can propagate through subsequent prediction stages and lead to substantial error accumulation, as illustrated in Figure 1(b). Direct spectrum-to-spectrum modeling instead establishes an end-to-end connection between reactant and reagent spectra and the corresponding product spectrum, avoiding unnecessary intermediate representation conversion. This motivates exploring chemical reaction modeling directly in the spectral domain.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

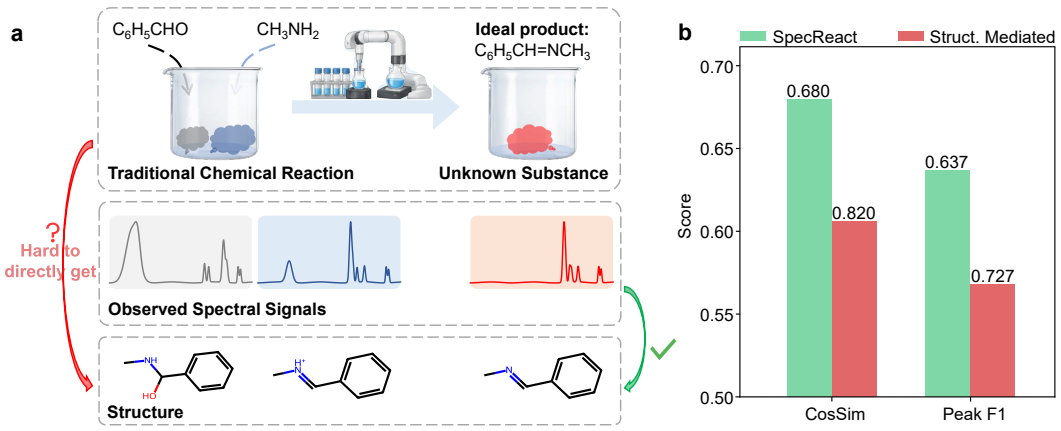

Figure 1: (a) Spectroscopic feedback in an autonomous reaction workflow. (b) Comparison between SpecReact and a structure-mediated (Struct. Mediated) reference on simulated spectra.

Despite these strengths, direct spectrum-to-spectrum reaction prediction mainly introduces two key challenges. First, a reaction may contain a variable number of reactants and reagents. Components within the same role have no natural ordering, whereas reactants and reagents make chemically distinct contributions to the reaction. A suitable model must therefore support variable-sized inputs and remain permutation invariant within each role while preserving differences across reaction roles Tu & Coley (2022). Second, reactant and product spectra can exhibit substantial changes during a reaction, including peak shifts, intensity variations, and the appearance or disappearance of spectral bands McFarlane et al. (1998); Mezzetti et al. (2022). We characterize such changes as non-conservative spectral transformations. These changes may occur simultaneously within a reaction, making the transformation difficult to describe with a single local correspondence. As a result, the reactant and product spectra do not admit a simple point-to-point correspondence Seifert et al. (2021), making these transformations difficult to capture through conventional point-wise reconstruction. A suitable model therefore needs to explicitly model such structured spectral changes.

To address these challenges, we propose **SpecReact**, a spectrum-to-spectrum framework for predicting product spectra from reactant and reagent spectra. To handle the unordered and variable-sized reaction inputs while preserving the distinct chemical roles of reactants and reagents, SpecReact employs a role-aware set encoder that separately models the two role-specific sets. This design preserves permutation invariance within each role while capturing their different contributions to the reaction context. To model the non-conservative spectral transformations that are difficult to characterize through point-wise reconstruction alone, we further introduce an unbalanced optimal transport (UOT) based spectral transformation module. By relaxing the strict marginal constraints of balanced OT, UOT can accommodate disappearing and emerging spectral components while establishing soft correspondences for preserved and displaced components. The resulting transport plan is further used to construct a spectral change map that provides structured supervision for peak persistence, disappearance, emergence, and displacement. In summary, our main contributions are as follows:

- We introduce **SpecReact**, a reaction modeling framework that directly generates product spectra from reactant and reagent spectra, without requiring resolved molecular structures as inputs.
- To address the unordered and variable number of reaction inputs while preserving the distinct chemical roles of reactants and reagents, we develop a role-aware set encoder that is permutation invariant within each role-specific set and models their dependent contributions to the reaction.
- To model the transformations between input and product spectra, we propose an UOT based spectral transformation module. It constructs soft correspondences between reactant and product spectra and provides structured and interpretable supervision for product spectrum prediction.
- We construct a new spectrum-based reaction dataset by converting molecular reaction records into paired reactant, reagent, and product IR spectra, enabling systematic evaluation of direct spectrum-to-spectrum reaction modeling.

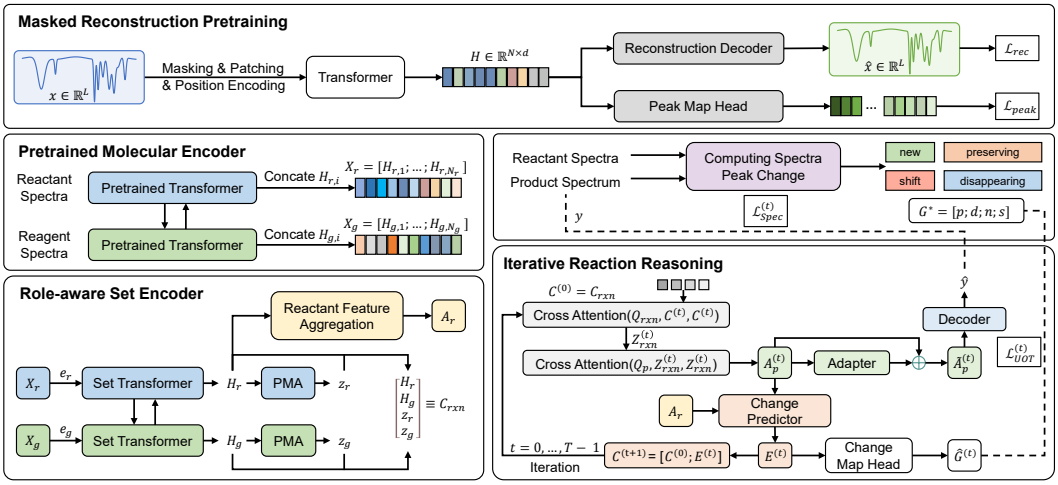

Figure 2: Overview of SpecReact. A role-aware set encoder aggregates reactant and reagent spectra, while UOT models reactant–product spectral changes. Iterative reaction reasoning integrates both representations to predict the product spectrum.

2 METHOD

2.1 PROBLEM STATEMENT

For the n-th reaction, the input consists of a set of reactant spectra $\mathcal{X}_r^{(n)} = \{x_{r,i}^{(n)}\}_{i=1}^{N_r^{(n)}}$ and a set of reagent spectra $\mathcal{X}_g^{(n)} = \{x_{g,j}^{(n)}\}_{j=1}^{N_g^{(n)}}$. Each spectrum satisfies $x_{r,i}^{(n)}, x_{g,j}^{(n)} \in \mathbb{R}^L$, where L denotes the spectrum length. Given a reaction dataset $\mathcal{D} = \{(\mathcal{X}_r^{(n)}, \mathcal{X}_g^{(n)}, y^{(n)})\}_{n=1}^{N}$, we learn a model F_Θ for direct reaction prediction in spectral space, such that $\hat{y}^{(n)} = F_\Theta(\mathcal{X}_r^{(n)}, \mathcal{X}_g^{(n)})$. All spectra are interpolated to the same length and independently normalized in intensity. This preprocessing preserves the relative peak intensities and peak shapes within each spectrum, but does not interpret intensity differences across spectra as physically calibrated concentration or stoichiometric information.

2.2 PRETRAINING

We first pretrain the spectrum encoder using molecule-resolved IR spectra from the training split, before learning the reaction mapping transformations. The pretraining objective is designed to capture generic spectral structures, such as local peak patterns and spectral continuity, independently of the downstream reaction prediction objective.

Given a complete spectrum $x = [x_1, x_2, \ldots, x_L] \in \mathbb{R}^L$, we randomly mask a subset of contiguous wavenumber regions and require the model to reconstruct the missing signal from the remaining observation $\widetilde{x}$. We partition $\widetilde{x}$ into $N = L/P$ non-overlapping patches, where the i-th patch is denoted by $\widetilde{p}_i \in \mathbb{R}^P$. After patching and positional encoding, the corresponding token is denoted by $h_{p,i}^0$. The complete token sequence $H^0 \in \mathbb{R}^{N\times d}$ is processed by a three-layer Transformer encoder

$$H = \mathrm{Encoder}(H^0). \tag{1}$$

We reconstruct the spectrum by expanding H along the sampling dimension to length L, yielding the point-level latent representation $\Phi \in \mathbb{R}^{L\times d}$. To reduce computation while preserving local spectral structure, we apply windowed self-attention, $\Phi_{\mathrm{loc}} = \mathrm{LocalEnc}(\Phi)$. At sampling position l, attention is restricted to a local neighborhood $\mathcal{N}(l)$, giving $\phi_l' = \sum_{j\in\mathcal{N}(l)} \alpha_{lj} W_V \phi_j$. This operation captures local peak continuity, overlap between neighboring peaks, and smooth transitions between masked and observed regions.

The local Transformer output is subsequently processed by a stack of one-dimensional residual convolutional blocks. Let $\Phi^0_{\mathrm{conv}} = \Phi_{\mathrm{loc}}$. The s-th block is defined as

$$\Phi^s_{\mathrm{conv}} = \Phi^{s-1}_{\mathrm{conv}} + \mathrm{DConv}\big(\Phi^{s-1}_{\mathrm{conv}}\big)\,. \tag{2}$$

Finally, LayerNorm and a pointwise linear projection produce the reconstructed spectrum

$$\hat{x} = W_o\,\mathrm{LN}(\Phi_{\mathrm{conv}}) + b_o, \hat{x} \in \mathbb{R}^L. \tag{3}$$

To explicitly preserve peak-related information, we attach a peak-map head to the encoder output. An offline peak extractor first constructs a target peak map $g \in [0,1]^L$ from the complete spectrum x. Because each patch corresponds to P sampling positions in the original spectrum, the peak-map head predicts P peak values from each patch token

$$\hat{g} = \sigma(\mathrm{Reshape}(W_g H + b_g))\,, \tag{4}$$

where $W_g : \mathbb{R}^d \to \mathbb{R}^P$ and σ denotes the sigmoid function, giving $\hat{g} \in [0,1]^L$.

To emphasize peak regions during reconstruction, we reweight the loss using the target peak map. Specifically, let $w_l = m_l(1 + \lambda_w g_l)$, where m_l selects masked positions and λ_w is a learnable peak-emphasis weight. The resulting peak-weighted reconstruction loss is

$$\mathcal{L}_{\mathrm{rec}} = \frac{\sum_{l=1}^{L} w_l(\hat{x}_l - x_l)^2}{\sum_{l=1}^{L} w_l + \epsilon}. \tag{5}$$

This objective applies standard reconstruction supervision to non-peak regions while increasing the contribution of peak regions. To alleviate the class imbalance between peak and non-peak positions, the peak-map head is optimized using a positive-class-weighted binary cross-entropy loss

$$\mathcal{L}_{\mathrm{peak}} = -\frac{1}{L}\sum_{l=1}^{L}\big[\alpha g_l \log \hat{g}_l + (1-g_l)\log(1-\hat{g}_l)\big]. \tag{6}$$

The complete pretraining objective is

$$\mathcal{L}_{\mathrm{pre}} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{peak}}\mathcal{L}_{\mathrm{peak}}. \tag{7}$$

2.3 Reaction Training

During reaction training, reactant and reagent spectra are encoded to predict the product spectrum $y \in \mathbb{R}^L$. Each input spectrum x_i is processed by the pretrained encoder to obtain H_i.

2.3.1 Role-Aware Set Encoder

Reactants and reagents have no natural ordering within their respective roles, while the two roles serve distinct functions in a reaction. We respectively concatenate the spectral representations belonging to the same role to form the reactant and reagent collections

$$X_r = [H_{r,1}; \ldots; H_{r,N_r}] \in \mathbb{R}^{(N_r N)\times d}, \quad X_g = [H_{g,1}; \ldots; H_{g,N_g}] \in \mathbb{R}^{(N_g N)\times d}. \tag{8}$$

We then add learnable role embeddings $e_r, e_g \in \mathbb{R}^d$ to the corresponding collections. The two role-specific sets are processed by a Set Transformer Lee et al. (2019) with shared model parameters

$$H_r = \mathrm{SetEnc}(X_r + e_r) \in \mathbb{R}^{(N_r N)\times d}, \quad H_g = \mathrm{SetEnc}(X_g + e_g) \in \mathbb{R}^{(N_g N)\times d}. \tag{9}$$

We use N role-specific learnable PMA seeds and then produce the role summaries $z_r = \mathrm{PMA}(s_r, H_r)$ and $z_g = \mathrm{PMA}(s_g, H_g)$. We stack the two role summaries as $z_{\mathrm{rxn}} = [z_r; z_g] \in \mathbb{R}^{2N\times d}$. The initial reaction context is constructed as

$$\mathcal{C}^{(0)}_{\mathrm{rxn}} = [H_r; H_g; z_{\mathrm{rxn}}] \in \mathbb{R}^{((N_r+N_g+2)N)\times d}. \tag{10}$$

For each wavenumber token position i and reactant molecular m, we compute attention weights across the reactants and aggregate their features

$$\alpha_{m,i} = \mathrm{softmax}_m\big(w^\top \mathrm{LN}(H_{r,m,i})\big),\ A_{r,i} = \sum_{m=1}^{N_r}\alpha_{m,i}H_{r,m,i},\ A_r = [A_{r,i}]_{i=1}^{N} \in \mathbb{R}^{N\times d}. \tag{11}$$

2.3.2 ITERATIVE REACTION REASONING

We perform multi-round reaction reasoning with parameters shared across iterations. Let $Q_{\mathrm{rxn}} \in \mathbb{R}^{N\times d}$ and $Q_p \in \mathbb{R}^{N\times d}$ denote learnable reaction and product queries, respectively. At iteration t, the reaction queries attend to the current context to obtain

$$Z_{\mathrm{rxn}}^{(t)} = \mathrm{CrossAttention}\big(Q_{\mathrm{rxn}}, \mathcal{C}_{\mathrm{rxn}}^{(t)}, \mathcal{C}_{\mathrm{rxn}}^{(t)}\big). \tag{12}$$

The resulting representation serves as both key and value for the product queries

$$A_p^{(t)} = \mathrm{CrossAttention}\big(Q_p, Z_{\mathrm{rxn}}^{(t)}, Z_{\mathrm{rxn}}^{(t)}\big). \tag{13}$$

A change predictor compares the reactant representation A_r with the candidate product representation $A_p^{(t)}$ and produces a latent spectral-change representation $E^{(t)} = \mathrm{ChangePredictor}(A_r, A_p^{(t)}) \in \mathbb{R}^{N\times d}$. A change-map head further maps this representation to a four-channel spectral change map, $\hat{G}^{(t)} = \mathrm{ChangeMapHead}(E^{(t)}) \in \mathbb{R}^{N\times 4}$. Except at the final iteration, $E^{(t)}$ is concatenated with the initial reaction context to construct the input to the next round, i.e., $\mathcal{C}_{\mathrm{rxn}}^{(t+1)} = [\mathcal{C}_{\mathrm{rxn}}^{(0)}; E^{(t)}]$. At every iteration, the candidate product representation is passed through a residual adapter, $\widetilde{A}_p^{(t)} = A_p^{(t)} + \mathrm{Adapter}(A_p^{(t)})$, and the pretrained spectrum decoder generates the corresponding product spectrum prediction $\hat{y}^{(t)}$. Detailed pseudocode of the iterative reasoning procedure is provided in Appendix B.

2.3.3 UOT-BASED SPECTRAL CHANGE SUPERVISION

During training, we construct a UOT problem between the ground-truth reactant and product spectra to encourage $E^{(t)}$ to capture structured spectral changes.

We first aggregate the reactant spectra and the ground-truth product spectrum into the same representation space as the spectral change map, obtaining $a \in \mathbb{R}_+^N$ and $b \in \mathbb{R}_+^N$, respectively. We define a transport matrix $T \in \mathbb{R}_+^{N\times N}$, where T_{ij} denotes the amount of spectral salience at source position i in the reactant spectrum that is softly matched to product position j.

Consequently, the UOT optimization objective is

$$\mathcal{J}_{\mathrm{UOT}} = \mathcal{J}_{\mathrm{move}} + \mathcal{J}_{\mathrm{src}} + \mathcal{J}_{\mathrm{tgt}}. \tag{14}$$

The transport cost and entropy regularization are given by

$$\mathcal{J}_{\mathrm{move}} = \langle C, T\rangle - \varepsilon\mathcal{H}(T). \tag{15}$$

The unbalanced source and target marginal penalties are

$$\mathcal{J}_{\mathrm{src}} = \tau_r D_{\mathrm{KL}}(T\mathbf{1} \,\|\, a)\,, \quad \mathcal{J}_{\mathrm{tgt}} = \tau_p D_{\mathrm{KL}}\big(T^\top\mathbf{1} \,\big\|\, b\big)\,. \tag{16}$$

where C_{ij} is constructed from the wavenumber distance between source position i and target position j, and $\mathcal{H}(\cdot)$ denotes the entropy regularizer. The unbalanced marginal penalties allow part of the reactant signal to remain untransported and part of the product signal to remain unexplained by the reactant, thereby providing representational capacity for disappearing and emerging peaks. The optimal transport solution is denoted by $T^* = \arg\min_T \mathcal{J}_{\mathrm{UOT}}$. In the definitions below, T denotes this optimized transport plan.

We use the optimal transport plan to construct a target spectral change map with four channels: persistence, disappearance, emergence, and displacement. The total mass transported from source position i to the product is defined as the persistent mass $p_i = \sum_j T_{ij}$. The untransported source mass is defined as the disappearing mass $d_i = [a_i - \sum_j T_{ij}]_+$. The product mass at position j that is not explained by the reactant is defined as the emerging mass $n_j = [b_j - \sum_i T_{ij}]_+$. For a source position with nonzero transported mass, the average wavenumber displacement is written as $s_i = \sum_j T_{ij}(\nu_j - \nu_i)/(p_i + \epsilon)$. The resulting target change map is

$$G^* = [p; d; n; s] \in \mathbb{R}^{N\times 4}. \tag{17}$$

This map describes the endpoint transformation between the reactants and the final product; it should not be interpreted as a complete reaction-kinetics trajectory. The prediction $\hat{G}^{(t)}$ from every reasoning iteration is supervised using the same UOT-derived target map. The corresponding loss is

$$\mathcal{L}_{\mathrm{UOT}}^{(t)} = \sum_{c\in\{p,d,n,s\}} \mathrm{SmoothL1}\Big(\hat{G}_c^{(t)}, G_c^*\Big)\,. \tag{18}$$

Table 1: Performance on simulated and experimental spectra across aromatic only (ARO), aromatic and long chain (ARO+LC), long chain only (LC), long chain and aliphatic ring (LC+ALR), and aliphatic ring only (ALR) reactants. **Bold** indicates the best result that is statistically significant under a t-test at the 95% confidence level. Standard deviations are reported in parentheses.

	ARO		ARO+LC		LC		LC+ALR		ALR	
Method	CosSim	F1	CosSim	F1	CosSim	F1	CosSim	F1	CosSim	F1
Simulated Spectra										
MT-Spec	$0.632_{(.008)}$	$0.587_{(.009)}$	$0.616_{(.011)}$	$0.574_{(.008)}$	$0.592_{(.013)}$	$0.536_{(.012)}$	$0.648_{(.009)}$	$0.558_{(.011)}$	$0.621_{(.012)}$	$0.544_{(.011)}$
G2S-Spec	$0.649_{(.013)}$	$0.607_{(.007)}$	$0.635_{(.010)}$	$0.598_{(.009)}$	$0.620_{(.012)}$	$0.563_{(.015)}$	$0.669_{(.008)}$	$0.581_{(.010)}$	$0.646_{(.010)}$	$0.568_{(.013)}$
MEGAN-Spec	$0.655_{(.012)}$	$0.619_{(.010)}$	$0.631_{(.008)}$	$0.614_{(.012)}$	$0.616_{(.013)}$	$0.575_{(.014)}$	$0.666_{(.010)}$	$0.594_{(.012)}$	$0.648_{(.009)}$	$0.576_{(.011)}$
ChemF-Spec	$0.660_{(.009)}$	$0.618_{(.010)}$	$0.638_{(.011)}$	$0.612_{(.009)}$	$0.623_{(.011)}$	$0.572_{(.008)}$	$0.667_{(.012)}$	$0.589_{(.010)}$	$0.651_{(.010)}$	$0.573_{(.012)}$
SpecReact	$\mathbf{0.691}_{(.008)}$	$\mathbf{0.642}_{(.009)}$	$\mathbf{0.667}_{(.007)}$	$\mathbf{0.641}_{(.009)}$	$\mathbf{0.649}_{(.009)}$	$\mathbf{0.598}_{(.006)}$	$\mathbf{0.713}_{(.011)}$	$\mathbf{0.611}_{(.007)}$	$\mathbf{0.684}_{(.008)}$	$\mathbf{0.593}_{(.008)}$
SpecReact+M	$0.770_{(.004)}$	$0.694_{(.006)}$	$0.747_{(.006)}$	$0.687_{(.005)}$	$0.736_{(.008)}$	$0.644_{(.009)}$	$0.784_{(.005)}$	$0.664_{(.007)}$	$0.760_{(.007)}$	$0.639_{(.008)}$
Experimental Spectra										
MT-Spec	$0.477_{(.017)}$	$0.437_{(.015)}$	$0.451_{(.019)}$	$0.423_{(.018)}$	$0.422_{(.019)}$	$0.370_{(.021)}$	$0.488_{(.020)}$	$0.396_{(.017)}$	$0.457_{(.022)}$	$0.376_{(.020)}$
G2S-Spec	$0.502_{(.011)}$	$0.458_{(.014)}$	$0.484_{(.019)}$	$0.444_{(.019)}$	$0.462_{(.010)}$	$0.400_{(.019)}$	$0.518_{(.017)}$	$0.423_{(.013)}$	$0.492_{(.013)}$	$0.402_{(.015)}$
MEGAN-Spec	$0.512_{(.012)}$	$0.468_{(.011)}$	$0.482_{(.013)}$	$0.458_{(.018)}$	$0.463_{(.011)}$	$0.413_{(.014)}$	$0.520_{(.019)}$	$0.435_{(.017)}$	$0.498_{(.021)}$	$0.410_{(.015)}$
ChemF-Spec	$0.515_{(.007)}$	$0.470_{(.016)}$	$0.492_{(.010)}$	$0.461_{(.017)}$	$0.468_{(.015)}$	$0.414_{(.012)}$	$0.522_{(.017)}$	$0.436_{(.015)}$	$0.503_{(.012)}$	$0.413_{(.013)}$
SpecReact	$\mathbf{0.567}_{(.013)}$	$\mathbf{0.526}_{(.017)}$	$\mathbf{0.539}_{(.009)}$	$\mathbf{0.519}_{(.014)}$	$\mathbf{0.514}_{(.010)}$	$\mathbf{0.462}_{(.015)}$	$\mathbf{0.579}_{(.018)}$	$\mathbf{0.489}_{(.015)}$	$\mathbf{0.551}_{(.017)}$	$\mathbf{0.454}_{(.013)}$
SpecReact+M	$0.667_{(.009)}$	$0.597_{(.011)}$	$0.639_{(.011)}$	$0.583_{(.014)}$	$0.615_{(.013)}$	$0.534_{(.019)}$	$0.683_{(.008)}$	$0.561_{(.010)}$	$0.651_{(.013)}$	$0.528_{(.012)}$

2.3.4 Optimization Objective

At iteration t, the product spectrum reconstruction loss is defined as $\mathcal{L}_{\text{spec}}^{(t)} = \|\hat{y}^{(t)} - y\|_2^2/L$. The complete reaction-training objective is

$$\mathcal{L}_{\text{rxn}} = \sum_{t=0}^{T-1} \omega_t \Big[\mathcal{L}_{\text{spec}}^{(t)} + \lambda_{\text{UOT}} \mathcal{L}_{\text{UOT}}^{(t)}\Big]. \tag{19}$$

Here, T is the number of reasoning iterations, ω_t controls the supervision weight assigned to iteration t, and λ_{UOT} balances final spectrum prediction against intermediate spectral-change learning. At reasoning time, the final-iteration prediction $\hat{y}^{(T-1)}$ is used as the model output.

3 Experiment

3.1 Experimental Setup

Datasets. We conduct experiments in both simulated and real-world scenarios. For the simulated setting, we construct a spectrum-based reaction dataset from the USPTO reaction benchmark Jin et al. (2017). Specifically, we preserve the original reactant, reagent, and product assignments and calculate molecule-resolved IR spectra for each reaction component, resulting in approximately 940K reaction samples covering approximately 1.14M unique molecules (Details in Appendix C). For the real-world setting, we collect experimental IR spectra from the NIST Chemistry WebBook [1] and match them to the molecular components in the USPTO reactions. We retain only reactions for which all required reactant, reagent, and product spectra are available.

Baselines. Our baselines include four representative reaction models, namely Molecular Transformer Schwaller et al. (2019), MEGAN Sacha et al. (2021), Chemformer Irwin et al. (2022), and Graph2SMILES Tu & Coley (2022). Their input and output modules are replaced with the pretrained spectrum encoder and decoder used in SpecReact. The resulting variants are denoted as MT-Spec, MEGAN-Spec, ChemF-Spec, and G2S-Spec, respectively. We further include SpecReact+M as a structure-assisted reference to assess how much reaction-relevant predictive information can already be captured directly from spectra. Detailed construction are provided in Appendix D.

Evaluation Metrics. We evaluate the performance of these models on cosine similarity (CosSim) and peak F1 score. Cosine similarity measures the global agreement between the predicted and

[1] https://webbook.nist.gov

Table 2: Performance of spectral reactions prediction with different numbers of reactants on simulated spectra. **Bold** indicates the best result that is statistically significant under a t-test at the 95% confidence level. Standard deviations are reported in parentheses.

	1 Reactant		2 Reactants		3 Reactants		≥ 4 Reactants	
Method	CosSim	F1	CosSim	F1	CosSim	F1	CosSim	F1
MT-Spec	$0.646_{(.009)}$	$0.596_{(.010)}$	$0.628_{(.008)}$	$0.583_{(.012)}$	$0.541_{(.017)}$	$0.510_{(.018)}$	$0.505_{(.023)}$	$0.482_{(.020)}$
G2S-Spec	$0.655_{(.007)}$	$0.611_{(.012)}$	$0.648_{(.008)}$	$0.605_{(.010)}$	$0.579_{(.014)}$	$0.552_{(.015)}$	$0.543_{(.021)}$	$0.526_{(.016)}$
MEGAN-Spec	$0.652_{(.006)}$	$0.620_{(.010)}$	$0.657_{(.007)}$	$0.617_{(.009)}$	$0.574_{(.016)}$	$0.578_{(.015)}$	$0.548_{(.019)}$	$0.559_{(.017)}$
ChemF-Spec	$0.661_{(.009)}$	$0.618_{(.008)}$	$0.662_{(.006)}$	$0.619_{(.010)}$	$0.582_{(.012)}$	$0.574_{(.014)}$	$0.550_{(.018)}$	$0.556_{(.017)}$
SpecReact	$\mathbf{0.683}_{(.005)}$	$\mathbf{0.640}_{(.006)}$	$\mathbf{0.687}_{(.004)}$	$\mathbf{0.638}_{(.007)}$	$\mathbf{0.618}_{(.010)}$	$\mathbf{0.616}_{(.012)}$	$\mathbf{0.590}_{(.015)}$	$\mathbf{0.594}_{(.016)}$
SpecReact+M	$0.756_{(.004)}$	$0.690_{(.005)}$	$0.766_{(.005)}$	$0.690_{(.006)}$	$0.703_{(.009)}$	$0.663_{(.011)}$	$0.685_{(.014)}$	$0.650_{(.016)}$

ground-truth spectral intensity profiles, whereas peak F1 evaluates the recovery of product spectrum peaks under a fixed peak-matching tolerance (Detailed evaluation protocol are provided in Appendix E). All experiments are repeated five times independently.

3.2 Main Results

3.2.1 Performance on Different Reactant Types.

We evaluate the models across different reactant categories to examine how molecular topology affects spectrum-to-spectrum reaction prediction. The results are reported in Table 1.

Obs. 1: SpecReact achieves the best performance on simulated spectra. As shown in Table 1, SpecReact outperforms the reaction baselines across all five reactant categories. Compared with the best reaction baseline in each category, SpecReact improves CosSim by 0.026–0.044 and peak F1 by 0.017–0.027. The competing baselines exhibit different strengths across metrics, whereas SpecReact maintains consistently strong performance across both global spectral similarity and peak recovery. These results validate role-aware representation and explicit transformation modeling.

Obs. 2: SpecReact exhibits metric-dependent sensitivity to reactant composition. SpecReact achieves lowest CosSim on the LC (0.649), while the LC+ALR reach the highest CosSim (0.713, + 9.9%), with peak F1 increasing more moderately from 0.598 to 0.611 (+2.2%). In contrast, the ARO+LC show lower CosSim than ARO reactions (0.667 vs. 0.691), while peak F1 remains nearly unchanged (0.641 vs. 0.642). These contrasting trends suggest that reactant composition is reflected more strongly in global spectral reconstruction than in prominent peak recovery, rather than exerting a uniform effect through any single topology.

Obs. 3: SpecReact shows stronger generalization to experimentally measured spectra. Under experimental setting, all methods exhibit lower performance due to instrumental noise, baseline drift, and other measurement variations. Nevertheless, SpecReact not only remains the best performing method across all reactant categories but also exhibits a smaller degradation from simulated to experimental spectra. Its CosSim decreases by only 0.124–0.135, compared with 0.143–0.170 for the baselines, while the corresponding peak F1 decrease is 0.116–0.139 versus 0.148–0.168. On experimental spectra, SpecReact achieves CosSim values of 0.514–0.579 and peak F1 values of 0.454–0.526, demonstrating stronger robustness to real-world spectral variations.

3.2.2 Reaction Prediction Results under Various Scenarios

To further assess SpecReact's robustness in diverse scenarios, we evaluate its performance across varying numbers of reactants, structural priors, and levels of spectral noise.

Obs. 4: SpecReact remains effective across varying reactant counts. As shown in Table 2, SpecReact consistently achieves the best performance among the main baselines across all reactant-count groups. Compared with the strongest baseline under each setting, it improves CosSim by 0.022–0.040 and peak F1 by 0.019–0.038, demonstrating its effectiveness in modeling variable-sized reaction inputs. Figure 3 further shows that SpecReact accurately recovers the overall spectral profiles and major product peaks across all settings. Even as the number of reactants increases and

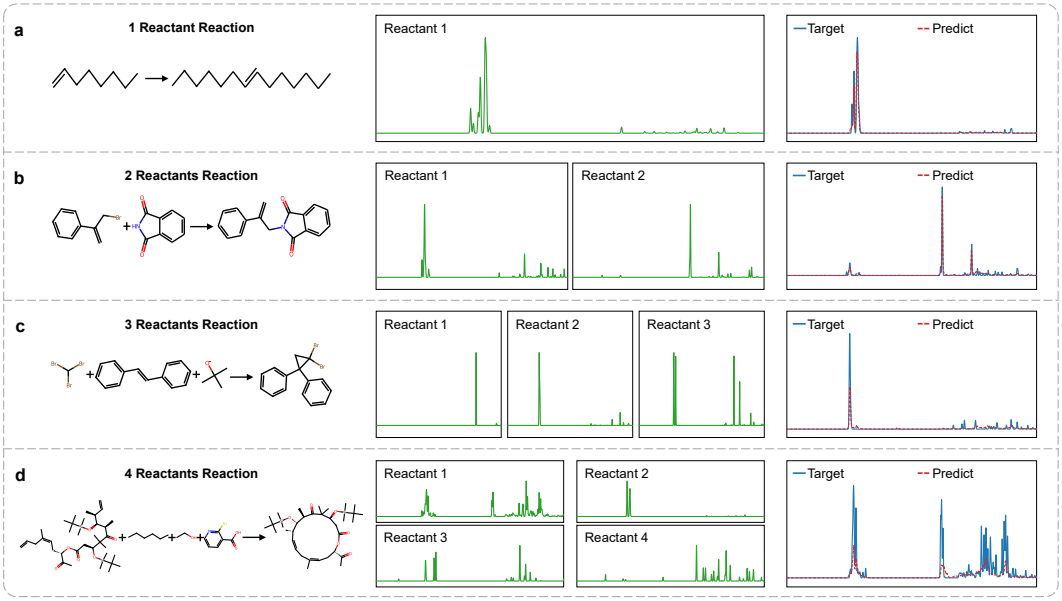

Figure 3: Qualitative predictions for reactions with one to four reactants. Each row shows an example, with frequency (cm^{-1}) and intensity on the horizontal and vertical axes, respectively.

the reaction inputs become more complex, the predicted spectra continue to preserve the principal spectral characteristics of the targets. These results demonstrate that SpecReact can effectively integrate variable-sized reactant sets and maintain strong product-spectrum prediction performance under increasing reaction complexity.

Obs. 5: Spectra alone capture substantial reaction-relevant information. The results across Tables 1 and 2 demonstrate that SpecReact consistently achieves strong product-spectrum prediction across diverse reactant categories and reactant-count settings using only spectral inputs. When SpecReact+M is introduced as an additional structure-assisted reference, SpecReact retains 86.1%–90.9% of its CosSim and 91.4%–93.3% of its peak F1 across the reported simulated settings. These results indicate that spectra alone contain substantial information required to model reaction outcomes, highlighting the effectiveness of direct spectrum-to-spectrum prediction while avoiding unnecessary intermediate structure reconstruction.

Obs. 6: SpecReact exhibits strong robustness to spectral noise. As shown in Figure 4(a–b), SpecReact maintains consistently strong CosSim and peak F1 as the Gaussian noise level increases, with both metrics degrading only gradually under increasingly severe perturbations. Its advantage over the competing models becomes more pronounced at higher noise levels, indicating that SpecReact is less sensitive to spectral corruption and better preserves both global spectral similarity and prominent product peaks. Even at $\sigma = 0.10$, SpecReact remains clearly ahead of the other methods on both metrics. Across the full noise range, its CosSim and peak F1 decrease by only 0.062 and 0.072, respectively, demonstrating stable product-spectrum prediction under noisy spectral inputs.

3.3 Ablation Studies and Hyperparameter Analysis

We further analyze the contributions of the role-aware encoder and UOT-based spectral transformation module, together with the sensitivity to the UOT loss weight.

Obs. 7: Role-aware encoding improves reaction representation. As shown in Figure 4(c), replacing the role-aware encoder with a shared SetFormer decreases CosSim from 0.680 to 0.653 and peak F1 from 0.637 to 0.614. This demonstrates that explicitly distinguishing reactants and reagents provides more informative reaction context than treating all molecular inputs with a shared set representation. The consistent degradation in both metrics further indicates that permutation-invariant set modeling alone is insufficient when chemically distinct reaction roles are collapsed.

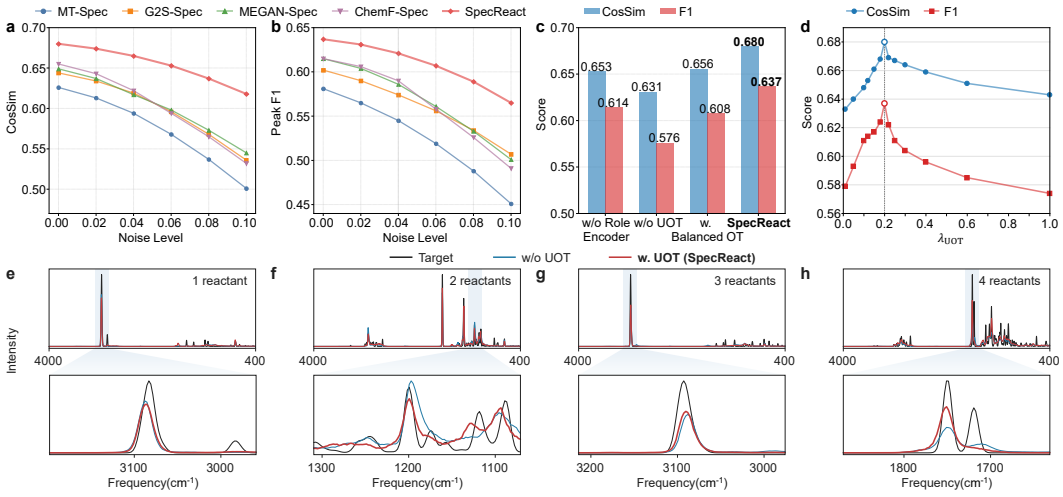

Figure 4: Robustness, ablation, and sensitivity analysis of SpecReact. (a–b) CosSim and peak F1 under increasing input noise. (c) Ablation of the role-aware encoder and UOT module. (d) Sensitivity to λ_{UOT}. (e–h) Product spectrum predictions w. UOT (SpecReact) and w/o UOT across one to four reactants, with enlarged regions shown below.

Obs. 8: UOT effectively models non-conservative spectral transformations. As shown in Figure 4(c), removing the UOT-based spectral transformation module reduces CosSim from 0.680 to 0.631 and peak F1 from 0.637 to 0.576. Replacing UOT with balanced OT improves the results to 0.656 and 0.608, respectively, but still underperforms the full model, demonstrating the benefit of relaxing marginal constraints when modeling disappearing and emerging spectral components. Figure 4(d) further shows that a moderate UOT loss weight yields the strongest performance, with both metrics reaching their optimum at $\lambda_{\mathrm{UOT}} = 0.20$. Together, these results demonstrate that UOT provides effective structured supervision for reaction-induced spectral transformations, while an appropriate loss weight balances transformation modeling with product-spectrum reconstruction.

Obs. 9: UOT improves fine-grained spectral reconstruction, particularly for more complex reactions. Figure 4(e–h) compares product spectrum predictions with and without UOT for reactions containing one to four reactants. For reactions with one or two reactants, both recover the overall spectral profiles and major peaks well, indicating that relatively simple reaction inputs can already be modeled effectively. As the number of reactants increases, however, the benefit of UOT becomes more evident. For the three and four reactant examples, SpecReact more closely follows the target spectrum in the enlarged regions, recovering local peak positions, shapes, and intensity variations more accurately than the variant. These results further demonstrate that UOT provides useful transformation guidance for spectral changes in increasingly complex reactions.

4 Conclusion

We proposed SpecReact, a spectrum-to-spectrum framework for directly predicting product IR from reactant and reagent spectra without intermediate molecular-structure reconstruction. SpecReact combines a role-aware set encoder with a UOT-based spectral transformation module to handle variable-sized reaction inputs, preserve distinct reaction roles, and model non-conservative spectral changes. We further construct a spectrum-based reaction dataset from the USPTO benchmark by pairing reaction components with simulated IR spectra. Experiments on simulated and experimentally measured spectra show that SpecReact consistently outperforms reaction baselines across diverse reactant settings and remains robust to spectral perturbations. Ablation, sensitivity, and qualitative analyses further demonstrate the effectiveness of role-aware encoding and UOT-based spectral transformation modeling. These results highlight the potential of direct reaction modeling from spectra and may help assess reaction outcomes and guide whether experiments should proceed, repeat, or adjust within closed-loop workflows in autonomous laboratories.

AI USE STATEMENT

Generative AI tools were used to assist with translation, language editing, refinement of the presentation and readability of the manuscript and editing portions of the software code. They were also used to provide suggestions on the organization and wording of portions of the paper. All AI-assisted content was carefully reviewed and revised by the authors and all AI-assisted code was reviewed and tested for correctness. The authors take full responsibility for the final content of this work, including all text, claims, results, and artifacts produced with the assistance of generative AI.

REFERENCES

Saleh Abdul Al and Abdul-Rahman Allouche. Neural network approach for predicting infrared spectra from 3d molecular structure. *Chemical Physics Letters*, 856:141603, 2024.

Marvin Alberts, Teodoro Laino, and Alain C Vaucher. Leveraging infrared spectroscopy for automated structure elucidation. *Communications Chemistry*, 7(1):268, 2024.

Jacopo Andreo, Romy Ettlinger, Orysia Zaremba, Quim Peña, Ulrich Lächelt, Roberto Fernández De Luis, Ralph Freund, Stefano Canossa, Evelyn Ploetz, Wei Zhu, et al. Reticular nanoscience: bottom-up assembly nanotechnology. *Journal of the American Chemical Society*, 144(17):7531–7550, 2022.

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15 (3):1652–1671, 2019.

Christoph Bannwarth, Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme. Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(2):e1493, 2021.

Oliver Bayley, Elia Savino, Aidan Slattery, and Timothy Noël. Autonomous chemistry: Navigating self-driving labs in chemical and material sciences. *Matter*, 7(7):2382–2398, 2024.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of computation*, 87(314):2563–2609, 2018.

Melanie M Cooper, Michael W Klymkowsky, and Nicole M Becker. Energy in chemical systems: An integrated approach. In *Teaching and learning of energy in K–12 education*, pp. 301–316. Springer, 2014.

Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.

Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3 (1):015022, 2022.

Wengong Jin, Connor W Coley, Regina Barzilay, and Tommi Jaakkola. Predicting organic reaction outcomes with weisfeiler-lehman network. *arXiv preprint arXiv:1709.04555*, 2017.

Greg Landrum, Paolo Tosco, Brian Kelley, Ricardo Rodriguez, David Cosgrove, Riccardo Vianello, et al. RDKit: Open-Source Cheminformatics Software, 2026. URL `https://doi.org/10.5281/zenodo.21741729`. Version 2026.03.5.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pp. 3744–3753. PMLR, 2019.

Zhigang Lei, Chengna Dai, Jason Hallett, and Mark Shiflett. Introduction: ionic liquids for diverse applications, 2024.

Xinyu Lu, Hao Ma, Hui Li, Jia Li, Yi Rong, Yuqiang Li, Tong Zhu, Guokun Liu, and Bin Ren. Vib2mol: from vibrational spectra to molecular structures-a unified deep learning framework. *arXiv preprint arXiv:2503.07014*, 2025.

Karen McFarlane, Brian Lee, Jon Bridgewater, and Peter C Ford. Time resolved infrared spectroscopy as a technique to study reactive organometallic intermediates. *Journal of organometallic chemistry*, 554(1):49–61, 1998.

Alberto Mezzetti, Josefine Schnee, Andrea Lapini, and Mariangela Di Donato. Time-resolved infrared absorption spectroscopy applied to photoinduced reactions: how and why. *Photochemical & Photobiological Sciences*, 21(4):557–584, 2022.

Boris L Milman. *Chemical identification and its quality assurance*. Springer, 2011.

Anthony K Rappé, Carla J Casewit, Kent S Colwell, William A Goddard III, and W Mason Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.

Sereina Riniker and Gregory A Landrum. Better informed distance geometry: using what we know to improve conformation generation. *Journal of chemical information and modeling*, 55(12): 2562–2574, 2015.

Mikołaj Sacha, Mikołaj Błaz, Piotr Byrski, Paweł Dabrowski-Tumanski, Mikołaj Chrominski, Rafał Loska, Paweł Włodarczyk-Pruszynski, and Stanisław Jastrzebski. Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling*, 61(7):3273–3284, 2021.

Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A Hunter, Costas Bekas, and Alpha A Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science*, 5(9):1572–1583, 2019.

Nathan A Seifert, Kirill Prozument, and Michael J Davis. Computational optimal transport for molecular spectra: The fully discrete case. *The Journal of Chemical Physics*, 155(18), 2021.

Martin Seifrid, Robert Pollice, Andres Aguilar-Granda, Zamyla Morgan Chan, Kazuhiro Hotta, Cher Tian Ser, Jenya Vestfrid, Tony C Wu, and Alan Aspuru-Guzik. Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab. *Accounts of Chemical Research*, 55(17):2454–2466, 2022.

Zhengkai Tu and Connor W Coley. Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *Journal of chemical information and modeling*, 62 (15):3503–3513, 2022.

Shuzhe Wang, Jagna Witek, Gregory A Landrum, and Sereina Riniker. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *Journal of chemical information and modeling*, 60(4):2044–2058, 2020.

Xi Xue, Hanyu Sun, Minjian Yang, Xue Liu, Hai-Yu Hu, Yafeng Deng, and Xiaojian Wang. Advances in the application of artificial intelligence-based spectral data interpretation: a perspective. *Analytical chemistry*, 95(37):13733–13745, 2023.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Zihan Zou, Yujin Zhang, Lijun Liang, Mingzhi Wei, Jiancai Leng, Jun Jiang, Yi Luo, and Wei Hu. A deep learning model for predicting selected organic molecular spectra. *Nature Computational Science*, 3(11):957–964, 2023.

A Related Work

Molecular reaction prediction. Data-driven forward reaction prediction commonly maps reactant and reagent structures to product structures. Molecular Transformer formulates reaction prediction as translation between SMILES sequences using an attention-based architecture Schwaller et al. (2019). Chemformer adopts a BART-style pretrained sequence-to-sequence framework for molecular and reaction modeling Irwin et al. (2022). Graph2SMILES replaces sequential reactant representations with a permutation-invariant molecular graph encoder while retaining an autoregressive product decoder Tu & Coley (2022). MEGAN instead represents chemical reactions as sequences of graph modifications Sacha et al. (2021). These methods provide effective architectures for learning chemical transformations but operate on resolved molecular structures. SpecReact instead performs reaction prediction directly from reactant and reagent spectra to the product spectrum.

Deep learning for molecular spectroscopy. Recent deep learning methods have advanced both structure-to-spectrum and spectrum-to-structure modeling. Structure-to-spectrum approaches predict IR and other molecular spectra from two- or three-dimensional molecular representations, substantially reducing the computational cost of conventional quantum-chemical calculations Zou et al. (2023); Al & Allouche (2024). In the reverse direction, transformer-based models have been developed to infer molecular structures from IR spectra Alberts et al. (2024). Vib2Mol further unifies spectrum–structure retrieval and molecular generation for vibrational spectra and demonstrates applications to reaction-product elucidation Lu et al. (2025). These studies primarily model spectrum–structure mappings for individual molecules, whereas SpecReact directly models reactions in the spectral domain from multiple reactant and reagent spectra to a product spectrum.

Set-structured and transport-based modeling. Reaction inputs naturally form variable-sized sets whose predictions should not depend on the ordering of molecules within the same reaction role. Deep Sets establishes a general framework for permutation-invariant learning Zaheer et al. (2017), while Set Transformer introduces attention-based interactions among set elements Lee et al. (2019). Generic set encoders, however, do not explicitly distinguish the different contributions of reactants and reagents. SpecReact therefore uses a role-aware set encoder that preserves permutation invariance within each role-specific set while maintaining the distinction between reaction roles. Optimal transport has also been applied to molecular spectra because it accounts for displacement of spectral intensity along the spectral axis Seifert et al. (2021). UOT extends this formulation to positive measures with unequal total mass Chizat et al. (2018), making it suitable for transformations in which spectral components may persist, disappear, emerge, or shift. SpecReact incorporates this property through a UOT-based spectral transformation module that provides structured supervision for product-spectrum prediction.

B Detailed Iterative Reaction Reasoning Procedure

Algorithm 1 summarizes the complete iterative reaction reasoning procedure. At each iteration, reaction queries extract information from the current reaction context, followed by product queries that construct the candidate product representation. The predicted spectral-change representation is used to update the reaction context for the next iteration. All parameters are shared across iterations, and the final-iteration spectrum is used as the model output during inference.

C Construction of the Spectrum-Based Reaction Dataset

Reaction Data Processing. We construct the simulated spectrum-based reaction dataset from the USPTO reaction benchmark used in our experiments. For each reaction, we retain the original separation between reactants, reagents, and products. The tokenized SMILES strings are first restored by removing token-level whitespace, without performing additional canonicalization. We then split each reaction side into its individual molecular components according to the dot-separated SMILES representation. Molecules with the same SMILES are indexed once within each dataset split so that their spectra can be reused across different reactions. We preserve the original 8:1:1 train/validation/test split and repeat all experiments five times with different random seeds.

Molecular Geometry Generation. To calculate the IR spectrum of each molecule, we first use RDKit Landrum et al. (2026) to convert its SMILES representation into a three-dimensional molec-

Algorithm 1 Iterative Reaction Reasoning

Require: Initial reaction context $\mathcal{C}_{\text{rxn}}^{(0)}$, reactant representation A_r, reaction queries Q_{rxn}, and product queries Q_p

Ensure: Product spectrum predictions $\{\hat{y}^{(t)}\}$ and spectral change maps $\{\hat{G}^{(t)}\}$

$\mathcal{C}^{(0)} \leftarrow \mathcal{C}_{\text{rxn}}^{(0)}$
for $t = 0, \ldots, T-1$ **do**
  $Z_{\text{rxn}}^{(t)} \leftarrow \text{CrossAttention}(Q_{\text{rxn}}, \mathcal{C}^{(t)}, \mathcal{C}^{(t)})$
  $A_p^{(t)} \leftarrow \text{CrossAttention}(Q_p, Z_{\text{rxn}}^{(t)}, Z_{\text{rxn}}^{(t)})$
  $E^{(t)} \leftarrow \text{ChangePredictor}(A_r, A_p^{(t)})$
  $\hat{G}^{(t)} \leftarrow \text{ChangeMapHead}(E^{(t)})$
  $\widetilde{A}_p^{(t)} \leftarrow A_p^{(t)} + \text{Adapter}(A_p^{(t)})$
  $\hat{y}^{(t)} \leftarrow \text{SpectrumDecoder}(\widetilde{A}_p^{(t)})$
  if $t < T-1$ **then**
    $\mathcal{C}^{(t+1)} \leftarrow [\mathcal{C}^{(0)}; E^{(t)}]$
  end if
end for
return $\{\hat{y}^{(t)}\}_{t=0}^{T-1}, \{\hat{G}^{(t)}\}_{t=0}^{T-1}$

ular geometry. Explicit hydrogen atoms are added, and three candidate conformers are generated using ETKDGv3 Riniker & Landrum (2015); Wang et al. (2020) with a fixed molecule-dependent random seed. Conformers are pruned using an RMS threshold of 0.5 Å. If standard embedding fails, ETKDGv3 is repeated with random-coordinate initialization. The generated conformers are subsequently optimized using MMFF94s Halgren (1996) when its parameters are available and UFF Rappé et al. (1992) otherwise. After a maximum of 500 optimization iterations, the conformers are ranked according to their force-field energies, and up to three lowest-energy candidates are retained for subsequent spectrum calculations.

IR Spectrum Calculation. For each selected molecular geometry, we perform a GFN2-xTB calculation Bannwarth et al. (2019; 2021) to obtain its vibrational frequencies and corresponding IR intensities. The molecular formal charge and number of unpaired electrons derived from the input molecule are explicitly provided to the calculation. Candidate conformers are evaluated in increasing force-field energy order until a valid vibrational calculation is obtained. For calculations that do not converge under the standard configuration, we additionally retry the lowest-energy conformer with an elevated electronic temperature.

The resulting vibrational calculation produces a set of discrete vibrational modes $\{(\nu_k, I_k)\}_{k=1}^{K}$, where ν_k is the vibrational frequency in cm^{-1} and I_k is its corresponding IR intensity. We convert these discrete modes into a continuous spectrum on a uniform wavenumber grid ranging from 400 to $4000\ \text{cm}^{-1}$ with an interval of $2\ \text{cm}^{-1}$. Only positive-frequency modes are included. Each vibrational mode is broadened using a Gaussian kernel

$$S(\nu) = \sum_{k:\nu_k>0} I_k^+ \exp\left(-\frac{(\nu-\nu_k)^2}{2\sigma^2}\right), \tag{20}$$

where $I_k^+ = \max(I_k, 0)$ and

$$\sigma = \frac{\text{FWHM}}{2\sqrt{2\ln 2}}. \tag{21}$$

We set the full width at half maximum to $\text{FWHM} = 12\ \text{cm}^{-1}$. No additional frequency scaling is applied. Finally, each spectrum is independently normalized by its maximum intensity

$$\widetilde{S}(\nu) = \frac{S(\nu)}{\max_{\nu'} S(\nu')}, \tag{22}$$

yielding a fixed-length IR spectrum with intensities in $[0, 1]$.

Quality Control and Reaction Assembly. We apply molecule-level quality control before assembling the final reaction dataset. Molecular components are excluded if molecular preparation or the

GFN2-xTB vibrational calculation fails, or if the resulting frequency–intensity records are missing, inconsistent, or contain non-finite values. Because isolated shallow imaginary modes can arise from incomplete relaxation of soft molecular degrees of freedom and do not necessarily invalidate the informative positive-frequency region of the simulated spectrum, we do not reject a molecule solely because of a single small imaginary frequency. Instead, we remove molecules exhibiting pronounced vibrational instabilities, defined as either a minimum vibrational frequency below $-100\ \mathrm{cm}^{-1}$ or at least two vibrational modes below $-20\ \mathrm{cm}^{-1}$.

Monatomic reagents or reactants are omitted from the reactant and reagent spectral inputs because they possess no internal vibrational degrees of freedom and therefore do not provide informative IR spectra. Reactions with monatomic products are excluded. Isotope-labeled molecules are also excluded because isotope-specific atomic masses are not retained in the XYZ-based spectrum-generation workflow. No explicit filtering is imposed based on molecular size or formal charge.

After quality control, the retained spectra are mapped back to their original reactant, reagent, and product roles. A reaction is included only when all remaining reactant and reagent components have valid spectra and a valid product spectrum is available. The resulting records are assembled into the final spectrum-based reaction dataset used for model training and evaluation.

D Spectrum-Based Baseline Adaptation

We replace their original molecular structure or SMILES interfaces with the pretrained spectrum encoder and spectrum decoder used in our framework. The reaction-specific modeling components of each baseline are retained whenever they can be transferred to continuous spectral representations.

MT-Spec. Molecular Transformer originally formulates reaction prediction as sequence-to-sequence translation over molecular tokens. In MT-Spec, we replace the original molecular token embeddings and output head with the shared spectrum encoder and spectrum decoder. The intermediate Transformer layers are retained to model the transformation from reactant and reagent spectral representations to the corresponding product representation.

Chemformer-Spec. Chemformer employs a BART-style encoder–decoder architecture for molecular sequence modeling. We use the shared spectrum encoder to provide continuous spectral representations and the shared spectrum decoder to generate the product spectrum, while retaining its encoder–decoder reaction transformation blocks. To avoid introducing additional molecular structure information, Chemformer-Spec is trained without the original SMILES-pretrained checkpoint.

G2S-Spec. Graph2SMILES originally encodes reactants as molecular graphs and predicts the product as an autoregressive molecular sequence. In G2S-Spec, the graph-specific input encoder is replaced with the shared spectrum encoder, while the molecular sequence output interface is replaced with the shared spectrum decoder. Its intermediate reaction transformation layers are retained to map the encoded reactant and reagent spectra to the product representation.

MEGAN-Spec. MEGAN originally models chemical reactions through iterative graph-state updates and discrete graph-edit prediction. For this adaptation, we remove the graph-specific molecular representation and graph-edit output head while retaining its iterative attention and reaction-state update blocks. These retained components operate on continuous spectral representations, which are subsequently converted into the product spectrum by the shared spectrum decoder.

Structure-Assisted Reference. SpecReact+M serves as a structure-assisted reference for contextualizing how much reaction-relevant information can be captured directly from spectra. We compute a 4096-bit Morgan fingerprint for each input molecule, project it into the reaction representation space, and combine it with the corresponding spectral representation while keeping the remaining SpecReact architecture unchanged throughout the evaluation.

Structure-Mediated Motivating Reference. For the motivating comparison in Figure 1(b), we construct a structure-mediated reaction reference. Each input spectrum is first converted into a molecular structure using an independently trained spectrum-to-structure model. The predicted reactant and reagent structures are then provided to Molecular Transformer Schwaller et al. (2019) to predict the product structure. Finally, an independently trained structure-to-IR model converts the predicted product structure into the corresponding product spectrum. This reference therefore

performs reaction prediction through an intermediate molecular-structure representation, whereas SpecReact directly maps the input spectra to the product spectrum.

E Peak F1 Evaluation Protocol

Peak F1 evaluates the agreement between the predicted and target spectra at the level of local spectral peaks. Given a predicted spectrum and its corresponding target spectrum, we first clip all negative intensities to zero. Each spectrum is then smoothed using a centered moving average with a window size of 20 grid points, where endpoint values are replicated for boundary padding.

Local maxima are detected on the smoothed spectra. A candidate peak is retained only if its intensity is at least 0.005 times the maximum intensity of the corresponding smoothed spectrum. We further enforce a minimum distance of 15 grid points between neighboring peaks; when two detected peaks violate this constraint, the higher-intensity peak is retained.

Let $\mathcal{P}$ and $\mathcal{T}$ denote the detected peak sets of the predicted and target spectra, respectively. A predicted peak and a target peak are considered a candidate match if their positions differ by at most 15 grid points. We then perform one-to-one maximum-cardinality matching over all candidate pairs. If N_{match} denotes the number of matched peak pairs, peak precision and recall are defined as

$$P_{\mathrm{peak}} = \frac{N_{\mathrm{match}}}{|\mathcal{P}|}, \qquad R_{\mathrm{peak}} = \frac{N_{\mathrm{match}}}{|\mathcal{T}|}. \tag{23}$$

The final peak F1 score is

$$\mathrm{F1}_{\mathrm{peak}} = \frac{2P_{\mathrm{peak}}R_{\mathrm{peak}}}{P_{\mathrm{peak}} + R_{\mathrm{peak}}}. \tag{24}$$

F Noise Robustness Evaluation Protocol

To evaluate robustness to spectral perturbations, we add Gaussian noise to the reactant and reagent spectra at test time. Following the common relative-noise setting for normalized spectral signals, the perturbed spectrum is defined as

$$\widetilde{x} = x + \epsilon, \qquad \epsilon_l \sim \mathcal{N}\Big(0,\, (\sigma I_{\max})^2\Big), \tag{25}$$

where $I_{\max} = \max_l x_l$ is the maximum intensity of the input spectrum and σ controls the noise level. We evaluate $\sigma \in \{0, 0.02, 0.04, 0.06, 0.08, 0.10\}$. Since each spectrum is independently normalized before being provided to the model, $I_{\max} = 1$, and σ therefore directly specifies the standard deviation of the added noise relative to the maximum peak intensity. For example, $\sigma = 0.10$ corresponds to Gaussian noise with a standard deviation equal to 10% of the maximum spectral intensity. Noise is applied only to the reactant and reagent spectra, while the target product spectrum remains unchanged. All models are evaluated under the same perturbation levels.

Table 3: Effect of spectrum pretraining on product-spectrum prediction using simulated spectra.

Setting	CosSim	F1
w/o Pretraining	0.641	0.588
Reconstruction Only	0.666	0.616
Full Pretraining	**0.680**	**0.637**

G Effect of Spectrum Pretraining

We evaluate the contribution of spectrum encoder pretraining using three settings. In *w/o Pretraining*, the spectrum encoder is randomly initialized and jointly trained with the reaction model. *Reconstruction Only* uses masked spectral reconstruction without peak-map supervision. *Full Pretraining* uses both masked reconstruction and peak-map prediction. All variants share the same reaction-modeling architecture and training configuration.

Spectrum pretraining consistently improves product-spectrum prediction. Compared with training the spectrum encoder from random initialization, masked reconstruction increases CosSim from 0.641 to 0.666 and peak F1 from 0.588 to 0.616, showing that reconstruction pretraining provides transferable representations of spectral shapes and local peak patterns. Incorporating peak-map prediction further improves CosSim to 0.680 and peak F1 to 0.637. The larger improvement in peak F1 indicates that explicit peak-aware supervision provides additional benefits for preserving localized spectral structures during subsequent reaction modeling.

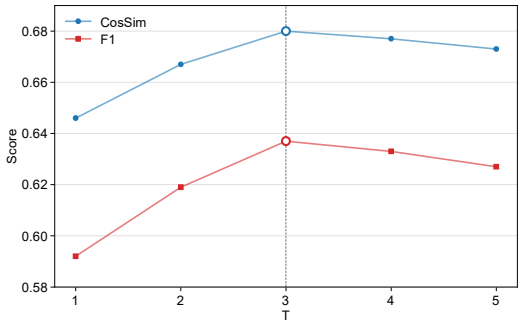

Figure 5: Effect of the number of reaction-reasoning iterations on simulated spectra.

H Effect of the Number of Reasoning Iterations

We investigate the effect of the number of reaction-reasoning iterations by varying T from 1 to 5. Each variant is trained and evaluated using the corresponding number of iterations. As shown in Figure 5, increasing T from 1 to 3 progressively improves both metrics. CosSim increases from 0.646 at $T = 1$ to 0.680 at $T = 3$, while peak F1 improves from 0.592 to 0.637. Further increasing the number of iterations provides no additional benefit, with both metrics decreasing slightly at $T = 4$ and $T = 5$. These results indicate that several rounds of reasoning are beneficial for refining the product representation, whereas excessive iterations may introduce redundant refinement or accumulated prediction errors. We therefore set $T = 3$ in the main experiments.

To further examine the prediction dynamics across reasoning iterations, Figures 6 and 7 present six representative reactions with product spectra predicted from $t = 1$ to $t = 5$. Across these examples, the largest improvements generally occur during the first few iterations, after which the predicted spectra become largely stable with only minor variations in spectral intensity and local peak reconstruction. Although the iteration yielding the highest CosSim varies slightly across individual reactions, the limited changes after $t = 3$ are consistent with the aggregate trend in Figure 5.

I Additional Qualitative Results across Reactant Counts

We provide additional product-spectrum prediction examples for reactions with one to four reactants. Figures 8 and 9 present four additional examples for each reactant count. Figure 8 shows reactions with one and two reactants, while Figure 9 shows reactions with three and four reactants. In each row, the reaction scheme and input reactant spectra are presented together with the target and predicted product spectra.

For reactions containing one or two reactants, the predicted spectra generally reproduce both the dominant absorption peaks and the overall spectral profiles, with only minor deviations in relative peak intensities. As the number of reactants increases to three or four, SpecReact continues to recover the main spectral regions and characteristic product peaks, while discrepancies become more visible in relative peak intensities and densely populated fingerprint regions. These examples further support the trend observed in Figure 3 and Table 2: prediction remains comparatively stable for simpler reaction inputs, whereas integrating more reactant spectra introduces greater ambiguity in product-spectrum reconstruction.

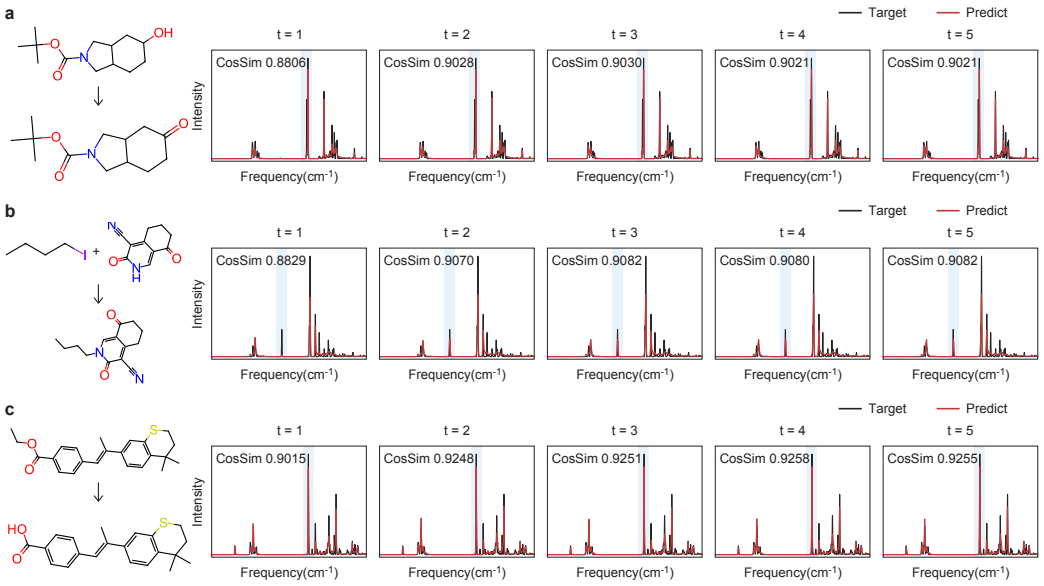

Figure 6: Representative examples of product-spectrum refinement across reasoning iterations. Examples (a–c) show the reaction transformation and target and predicted product spectra from $t = 1$ to $t = 5$, with CosSim reported for each iteration.

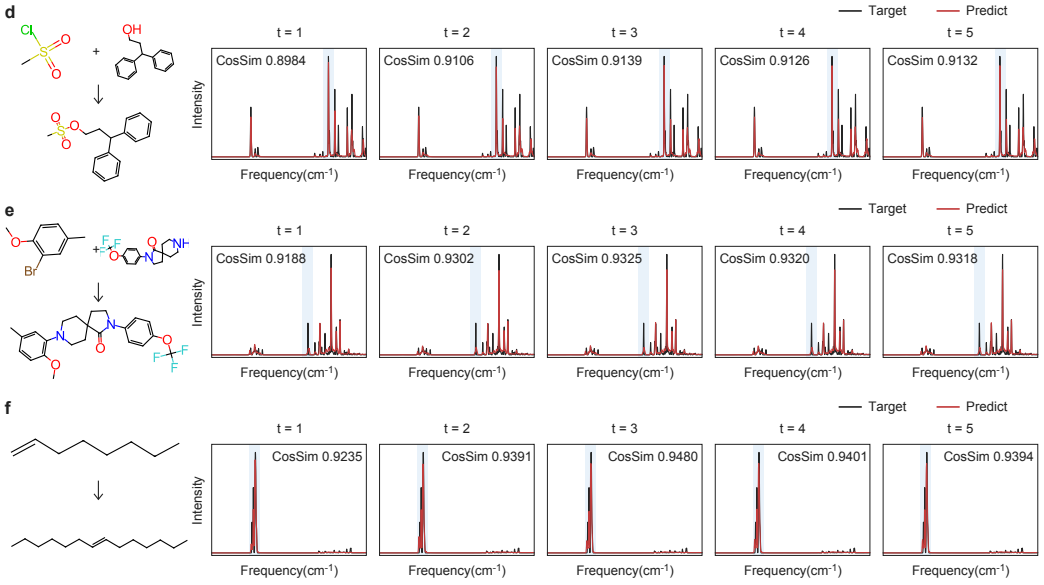

Figure 7: Additional examples of product-spectrum refinement across reasoning iterations. Examples (d–f) follow the same visualization protocol as Figure 6.

J Broader Impacts and Limitations

Broader Impacts. SpecReact enables product-spectrum prediction directly from experimentally accessible spectra and may support automated reaction assessment in autonomous laboratories. By providing machine-readable spectral predictions, the framework could help determine whether an experiment should proceed, be repeated, or be adjusted within a closed-loop workflow. The con-

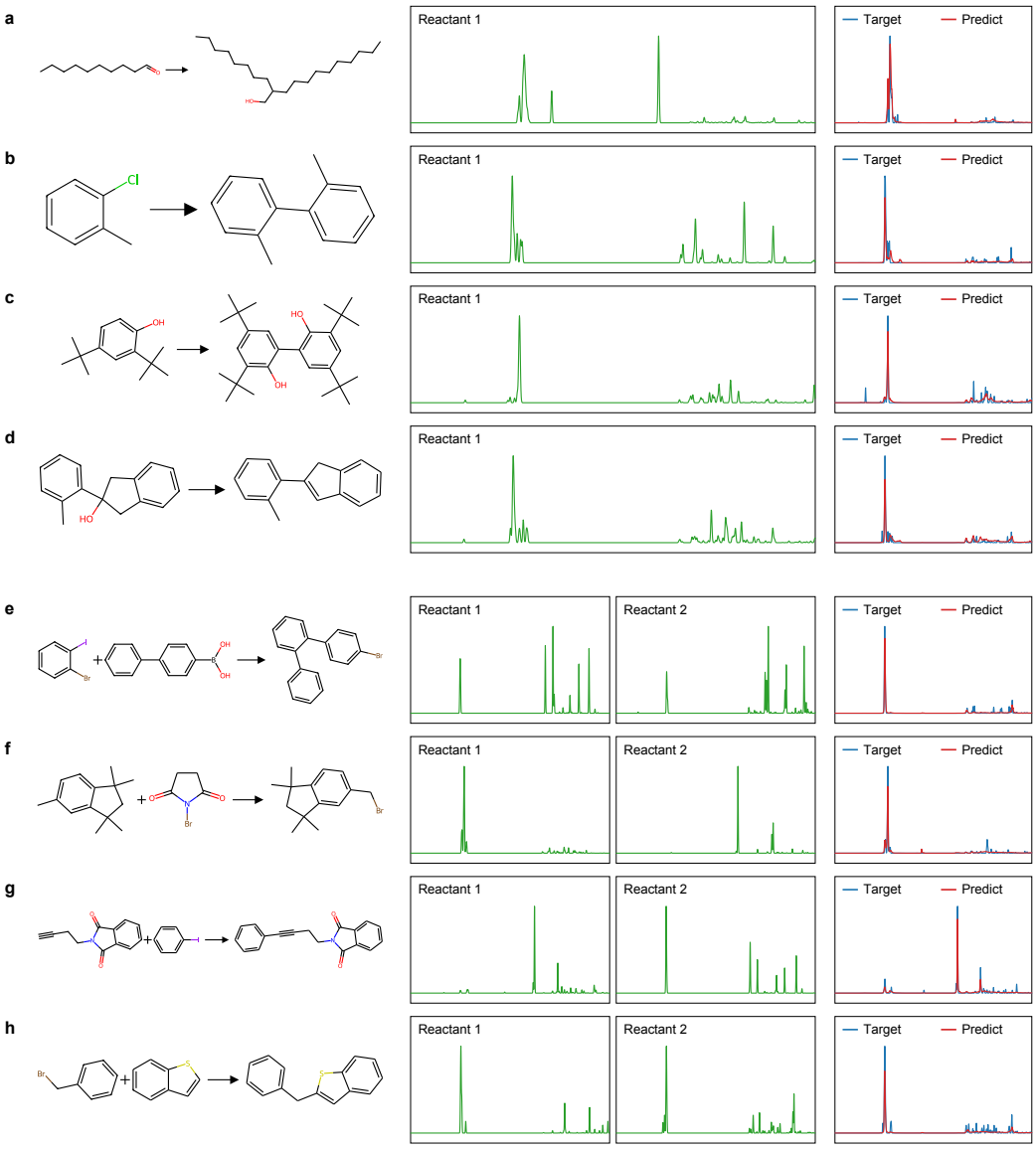

Figure 8: Additional qualitative examples for reactions with one and two reactants. Panels (a)–(d) correspond to single reactant reactions, and (e)–(h) correspond to two reactant reactions. Each row shows an example, with frequency (cm^{-1}) and intensity on the horizontal and vertical axes.

structed spectrum-based reaction dataset may further support research on reaction modeling directly in the spectral domain. Predicted spectra should nevertheless be treated as computational guidance rather than conclusive evidence of molecular identity or reaction outcome, and practical use should remain subject to expert review and experimental validation. Potential dual-use applications should also be managed under appropriate institutional, legal, and laboratory-safety controls.

Limitations. SpecReact models the endpoint mapping from reactant and reagent spectra to a product spectrum and does not recover reaction kinetics, transient intermediates, or mechanistic pathways. Because spectra are independently normalized, the current formulation does not explicitly represent calibrated concentration or stoichiometric information. Due to the scarcity of real-world reaction datasets containing matched reactant, reagent, and product spectra measured within the same reaction system, our experimental evaluation uses molecular IR spectra collected from the NIST

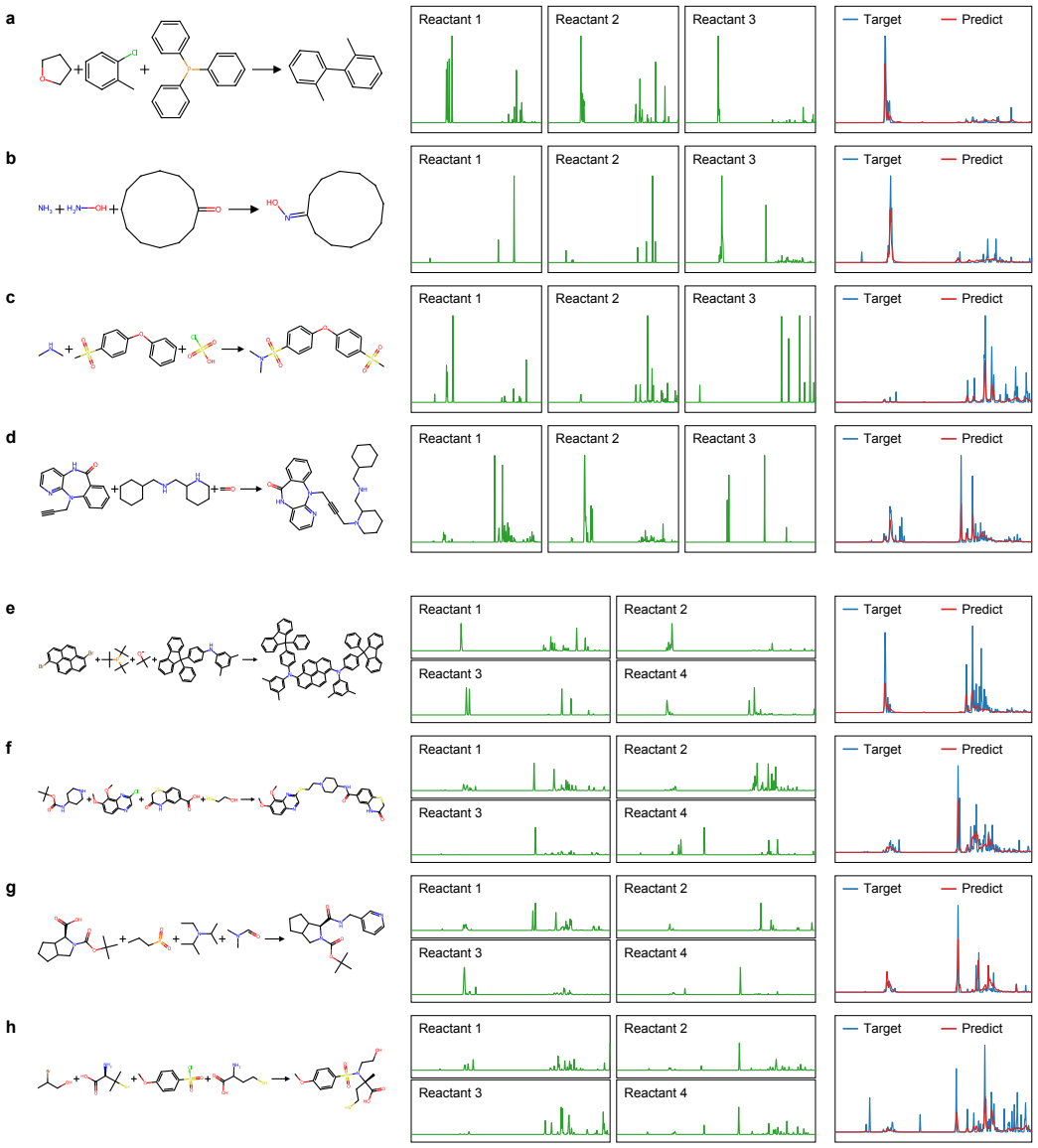

Figure 9: Additional qualitative examples for reactions with three and four reactants. Panels (a)–(d) correspond to three reactant reactions, and (e)–(h) correspond to four reactant reactions. Each row shows an example, with frequency (cm^{-1}) and intensity on the horizontal and vertical axes.

Chemistry WebBook and matched to the corresponding components in USPTO reactions. Although this construction enables evaluation with experimentally measured spectra, it does not fully capture reaction-specific measurement conditions or spectral consistency within the same experimental process. Constructing larger datasets with jointly measured reactant, reagent, and product spectra under controlled conditions is therefore an important direction for future work. Performance also decreases as the number of reactants increases, particularly for reactions containing three or more reactants. Finally, UOT-derived correspondences should be interpreted as soft spectral associations rather than specific vibrational-mode or mechanistic assignments, and predicted spectra should not be regarded as definitive evidence of reaction outcome or molecular identity.