

Supporting Information

Providing direction for mechanistic inferences in radical cascade cyclization using Transformer model

Jiangcheng Xu, Yun Zhang, Jiale Han, An Su, Haoran Qiao, Chengyun Zhang, Jing Tang, Xi Shen, Bin Sun, Wenbo Yu, Silong Zhai, Xinqiao Wang, Yejian Wu, Weike Su*, Hongliang Duan*

*Corresponding author. Email: Pharmlab@zjut.edu.cn (W.S.); hduan@zjut.edu.cn (H.D.)

Content

1. The definition of key intermediate.....	1
2. Radical cascade cyclization is widely used in the life science and pharmaceutical industry.	2
3. Attention weight interpretation for examples from randomly chosen reactions.....	3
4. Cases randomly selected from the self-built dataset.....	5
5. The Transformer model	17
6. The impact of different text representations	19
7. Details of cross-validations experiments	22
8. Analysis of “wrong” predictions.....	26
9. Computational studies.....	28
10. References.....	32

1. The definition of key intermediate

Radical cascade cyclization reaction proceeds in four steps:

- (i) Radical formation via single-electron transfer (SET).
- (ii) Radical addition via a radical attack on unsaturated bonds to generate radical intermediates.
- (iii) Radical cyclization, resulting in carbon-carbon/heteroatom bonds.
- (iv) Radical or radical intermediate quenching by another radical donor or by hydrogen abstraction.

To simplify the reaction mechanism, we define the intermediate just after the first radical addition as “Key intermediate I”, which can show the site for the first addition.

The intermediate after the ring generated was defined as “Key intermediate II” if the cascade reaction constructs only single carbo-/heterocycles. If construct multiple cycles, the intermediate before the last ring closed was defined as “Key intermediate II” (Fig. S1a). Key intermediate II can reflect important information whether rearrangement, hydrogen migration, or aryl migration, and other transformations that occurred during the reaction.

Key intermediate I & II can effectively guide the judgment of the reaction path. The methodology for speculating the intermediates in radical cascade cyclization is analogous to a game of traversing a maze. In this game, if there is an appropriate “road sign”, the players can quickly walk through the maze. Moreover, if there is the possibility of multiple routes, the particular “road sign” can indicate the most reasonable route. The “key intermediate” serves as the “road sign” in the mechanism inference methodology (Fig. S1b).

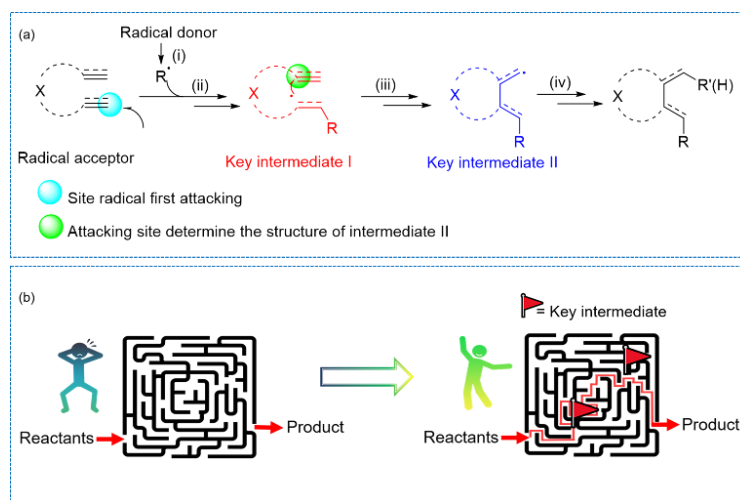


Fig. S1 The definition of key intermediate.

2. Radical cascade cyclization is widely used in the life science and pharmaceutical industry.

Radical cascade cyclization is highly valuable for the preparation of cyclic compounds since more than two bonds can be formed by a single preparation step (Fig S2 Upper). Pharmaceuticals and complex natural products can be synthesized via radical cascade cyclization¹⁻³ (Fig S2 Lower).

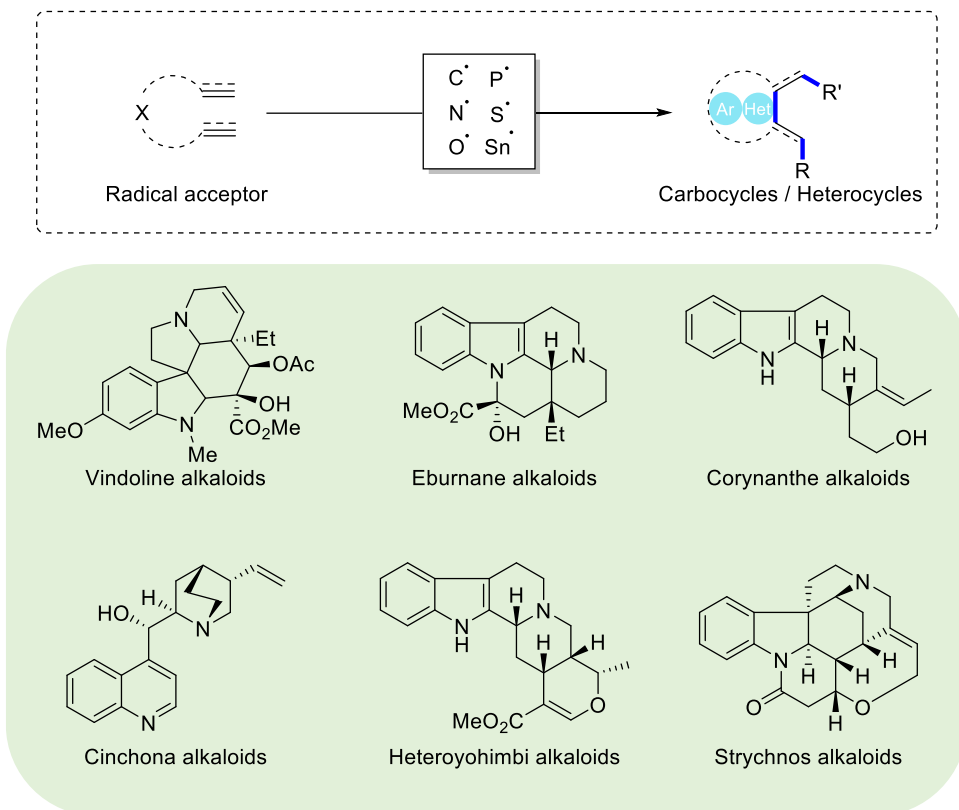


Fig. S2 Pharmaceuticals and complex natural products can be synthesized via radical cascade cyclization.

3. Attention weight interpretation for examples from randomly chosen reactions

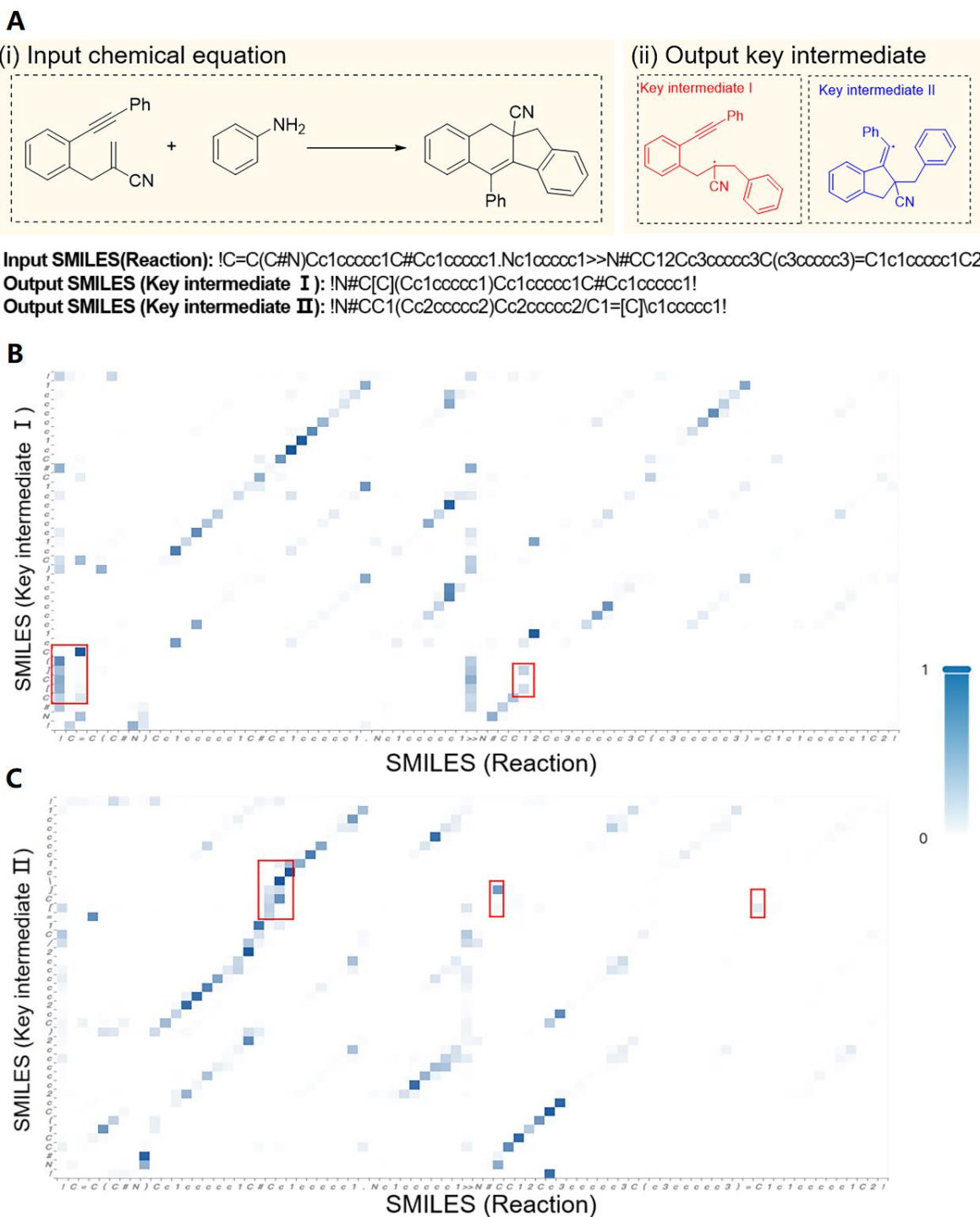


Fig. S3 Attention weight interpretation for example from the randomly chosen reaction (Entry=371). (A) Example from randomly chosen reaction in self-built data set (Entry=371), The visualization of attention weights for (B) key Intermediate I and (C) key Intermediate II.

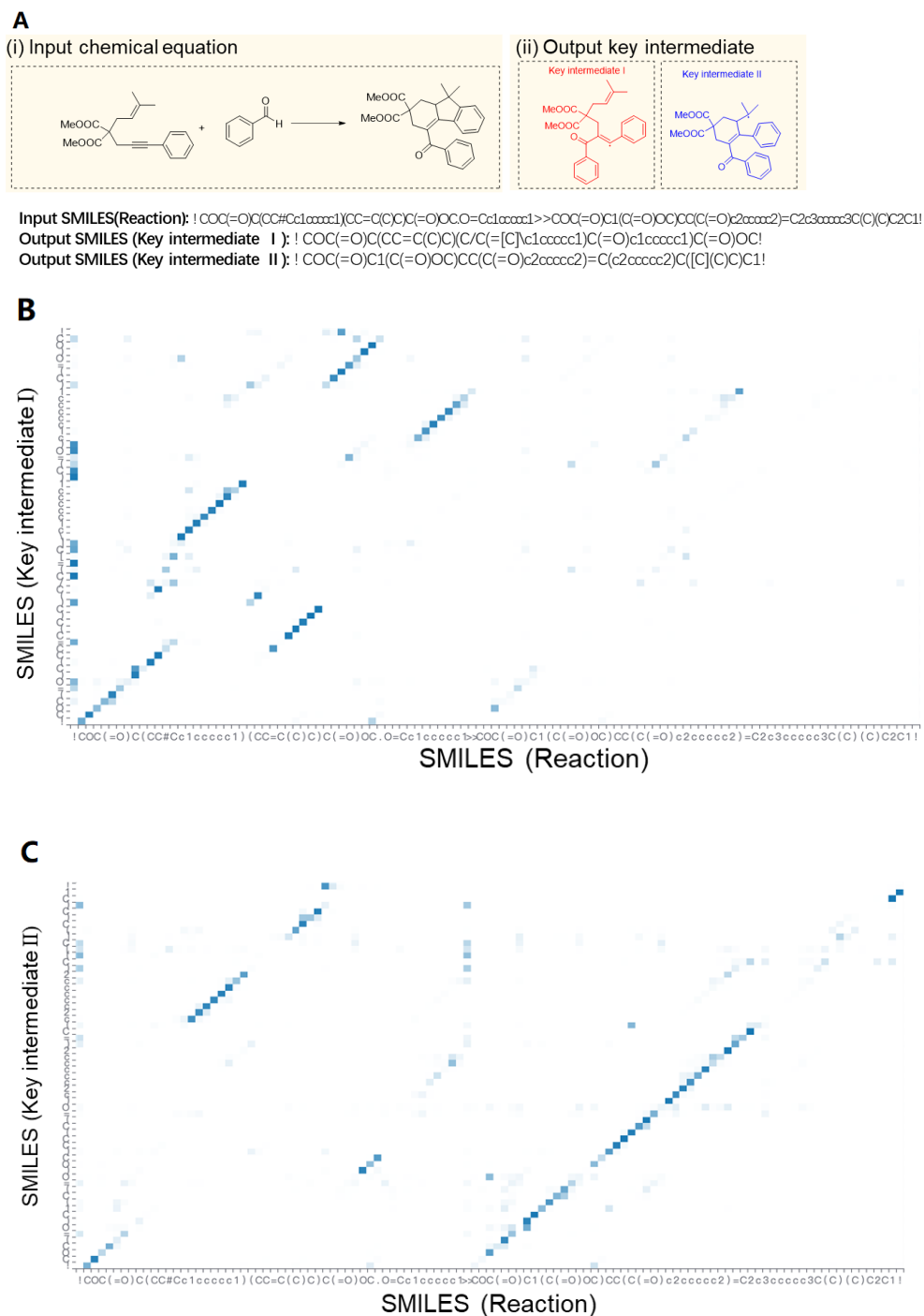


Fig. S4 Attention weight interpretation for example from randomly chosen reaction (Entry=23). (A) Example from the randomly chosen reaction in self-built data set (Entry=23). The visualization of attention weights for (B) key Intermediate I and (C) key Intermediate II. There is an interesting phenomenon that the key intermediate I has a large correlation with the reactants and intermediate II with the product. This was reflected in the number and depth of the weight squares on the left and right sides. More and darker squares occur in the left means key intermediate I have a kinship with the reactant, intermediate II has a closer relationship with the product. It is maybe related to the definition of intermediate II.

4. Cases randomly selected from the self-built dataset.

The training, fine-tuning, validation, and testing datasets used in our study are available from <https://github.com/hongliangduan/transRCC>. We randomly selected 12 cases (Example I - VII), each case's entry, reference, basic information, reaction formula, proposed mechanism, the corresponding key intermediate, the relevant data in the database was shown in table S1-S12.

Table S1. Example I randomly selected from the self-built dataset.

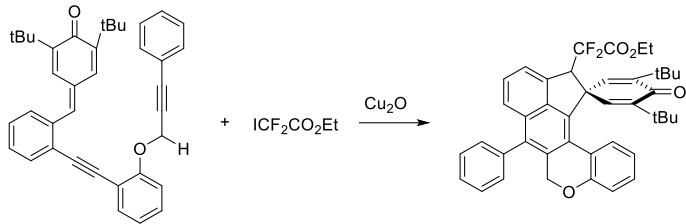
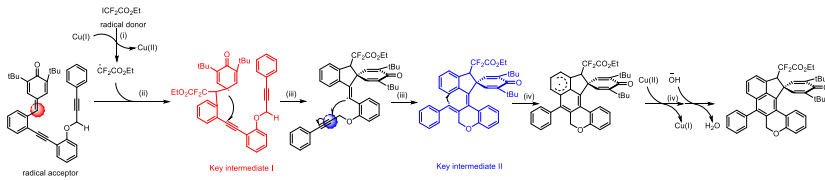

Entry	41
Source	H. D. Zuo <i>et al.</i> , Cu-Catalyzed radical-triggered spirotricyclization of enediynes and enyne-nitriles for the synthesis of pentacyclic spiroindenes. <i>Org. Chem. Front.</i> 8 , 1496-1502 (2021).
Information	3 new rings constructed; C- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>CC(C)(C)C1=CC(=Cc2ccccc2C#Cc2ccccc2OCC#Cc2ccccc2)C=C(C(C)(C)C)C1=O.CCOC(=O)C(F)(F)I>>CCOC(=O)C(F)(F)C1c2cccc3c(-c4ccccc4)c4c(c(c23)C12C=C(C(C)(C)C)C(=O)C(C(C)(C)C)=C2)-c1cccc1OC4 CCOC(=O)C(F)(F)C([C]1C=C(C(C)(C)C)C(=O)C(C(C)(C)C)=C1)c1cccc1C#Cc1cccc1OCC#Cc1cccc1 CCOC(=O)C(F)(F)C1c2ccccc2/C(=C2C(=[C]c3ccccc3)\COc3ccccc3\2)C12C=C(C(C)(C)C)C(=O)C(C(C)(C)C)=C2</chem>

Table S2. Example II randomly selected from the self-built dataset.

Entry	82
Source	M. Hu <i>et al.</i> , Metal-Free Radical [2+2+1] Carbocyclization of Benzene-Linked 1,n-Enynes: Dual C(sp ³)-H Functionalization Adjacent to a Heteroatom. <i>Angew. Chem. Int. Ed.</i> 54 , 9577-9580 (2015).
Information	2 new rings constructed; C- center radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C1COCCO1.C=C(C)C(=O)N(C)c1cccc1C#Cc1ccc(C)cc1>>Cc1ccc(C2=C3c4ccccc4N(C)C(=O)C3(C)CC23COCCO3)cc1 C[C](CC1COCCO1)C(=O)N(C)c1cccc1C#Cc1ccc(C)cc1 Cc1ccc(/C=C2\c3ccccc3N(C)C(=O)C2(C)C[C]2COCCO2)c1</chem>

Table S3. Example III randomly selected from the self-built dataset.

Entry	138
Source	W. Wu <i>et al.</i> , Synthesis of Sulfonylated Lactones via Ag-Catalyzed Cascade Sulfonylation/Cyclization of 1,6-Enynes with Sodium Sulfinates. <i>J. Org. Chem.</i> 82 , 1224-1230 (2017).
Information	1 new ring constructed; S- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(C)COC(=O)C#Cc1ccccc1.Cc1ccc(S(=O)(=O)[Na])cc1>>C c1ccc(S(=O)(=O)CC2(C)COC(=O)/C2=C/c2ccccc2)cc1 C[C](C OC(=O)C#Cc1ccccc1)CS(=O)(=O)c1ccc(C)cc1 Cc1ccc(S(=O)(=O)CC2(C)COC(=O)/C2=[C]/c2ccccc2)cc1</chem>

Table S4. Example IV randomly selected from the self-built dataset.

Entry	161
Source	C. Li <i>et al.</i> , Synthesis of Fused Pyran Derivatives via Visible-Light-Induced Cascade Cyclization of 1,7-Enynes with Acyl Chlorides. <i>Org. Lett.</i> 19 , 512-515 (2017).
Information	2 new rings construct; C- center radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(C)C(=O)N(C)c1ccccc1C#Cc1ccc(Br)cc1.O=C(Cl)c1ccccc1>>CN1C(=O)C2(C)C=C(c3ccccc3)OC(c3ccc(Br)cc3)=C2c2ccc1 C[C](CC(=O)c1ccccc1)C(=O)N(C)c1ccccc1C#Cc1ccc(Br)cc1 CN1C(=O)C(C)(CC(=O)c2ccccc2)/C(=[C]/c2ccc(Br)cc2)c2ccccc21</chem>

Table S5. Example V randomly selected from the self-built dataset.

Entry	228
Source	J. Xuan <i>et al.</i> , Radical Cascade Cyclization: Reaction of 1,6-Enynes with Aryl Radicals by Electron Catalysis. <i>E. J. Org. Chem.</i> 2016 , 4961-4964 (2016).
Information	2 new rings constructed; C- center radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(Cc1cccc1C#Cc1cccc1)C(=O)OC.COc1ccc(N)cc1>>COC(=O)C12Cc3ccc(OC)cc3C(c3cccc3)=C1c1cccc1C2 COC(=O)[C](Cc1ccc(OC)cc1)Cc1cccc1C#Cc1cccc1 COC(=O)C1(Cc2ccc(OC)cc2)Cc2cccc2/C1=[C]c1cccc1</chem>

Table S6. Example VI randomly selected from the self-built dataset.

Entry	241
Source	Y. An <i>et al.</i> , Synthesis of trifluoromethylated 3,4-dihydroquinolin-2(1H)-ones via a photo-induced radical cyclization of benzene-tethered 1,7-enynes with Togni reagent. <i>Org. Chem. Front.</i> 3 , 994-998 (2016).
Information	2 new rings constructed; C- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(C)C(=O)N(C)c1ccccc1C#Cc1ccccc1.O=C1O[IH2](C(F)(F)F)c2ccccc21>>CN1C(=O)C(C)(CC(F)(F)F)/C(=C(/I)c2ccccc2)c2ccccc21 C[C](CC(F)(F)F)C(=O)N(C)c1ccccc1C#Cc1ccccc1 CN1C(=O)C(C)(CC(F)(F)F)/C(=[C]\c2ccccc2)c2ccccc21</chem>

Table S7. Example VII randomly selected from the self-built dataset.

Entry	247
Source	Y. Zhou <i>et al.</i> , Radical Tandem Bicyclization Triggered by the alpha-Position of alpha,beta-Unsaturated Ketones Bearing Nonterminal 1,6-Enynes: Synthesis of the 5H-Benzo[a]fluoren-5-one Skeleton. <i>Org. Lett.</i> 22 , 8359-8364 (2020).
Information	2 new rings construct; C- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>Cc1ccc(C#Cc2ccccc2C(=O)/C=C/c2ccccc2)cc1.Cc1ccc(C=O)cc1>>Cc1ccc(C(=O)C2C(=O)c3ccccc3C3=C(c4ccc(C)cc4)c4ccccc4C32)cc1 Cc1ccc(C#Cc2ccccc2C(=O)C([C]c2ccccc2)C(=O)c2ccc(C)cc2)cc1 Cc1ccc(/[C]=C2\c3ccccc3C(=O)C(C(=O)c3ccc(C)cc3)C2c2ccccc2)cc1</chem>

Table S8. Example VIII randomly selected from the self-built dataset.

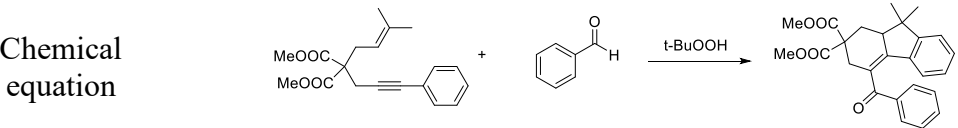
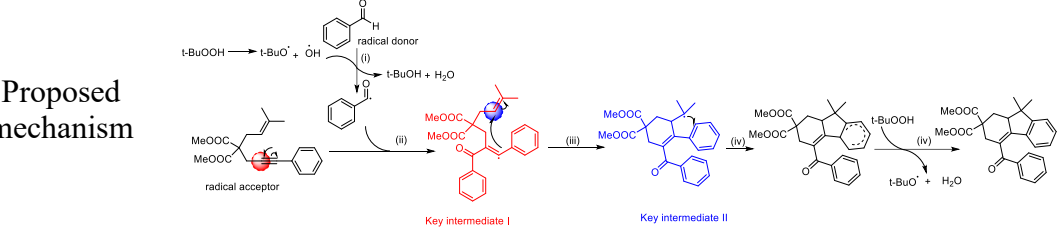
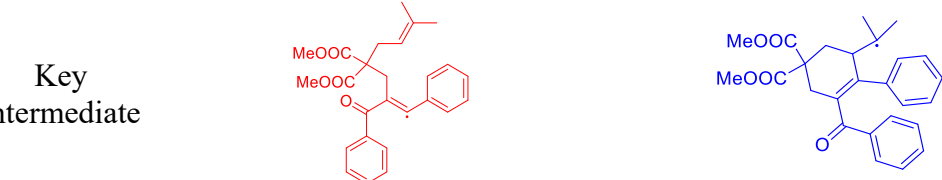
Entry	266
Source	J. Y. Luo <i>et al.</i> , Metal-free cascade radical cyclization of 1,6-enynes with aldehydes. <i>Chem. Commun.</i> 50 , 1564-1566 (2014).
Information	2 new rings constructed; C- center radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>COC(=O)C(CC#Cc1ccccc1)(CC=C(C)C)C(=O)OC.O=Cc1ccccc1>>COC(=O)C1(C(=O)OC)CC(C(=O)c2ccccc2)=C2c3ccccc3C(C)(C)C2C1 COC(=O)C(CC=C(C)C)(C/C(=[C])\c1ccccc1)C(=O)c1ccccc1)C(=O)OC COC(=O)C1(C(=O)OC)CC(C(=O)c2ccccc2)=C(c2ccccc2)C([C])(C)C1</chem>

Table S9. Example IX randomly selected from the self-built dataset.

Entry	273
Source	Y. Li <i>et al.</i> , Visible Light-Initiated C(sp ³) Br/C(sp ³) H Functionalization of α -Carbonyl Alkyl Bromides through Hydride Radical Shift. <i>Adv. Syn. Catal.</i> 358 , 1219-1228 (2016).
Information	2 new rings constructed; C- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(Cc1ccccc1C#Cc1ccccc1)C(=O)OC.CC(Br)C(=O)Oc1ccccc1>>COC(=O)C12Cc3ccccc3C1=C(c1ccccc1)C(C)(C(=O)Oc1ccccc1)C2 COC(=O)[C](Cc1ccccc1C#Cc1ccccc1)CC(C)C(=O)Oc1ccccc1 COC(=O)C1(CCC(=O)Oc2ccccc2)C2ccccc2/C1=C\c1ccccc1</chem>

Table S10. Example X randomly selected from the self-built dataset.

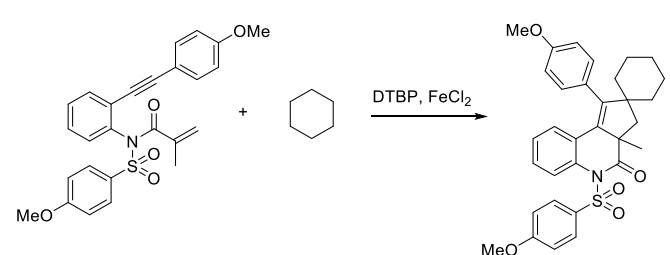
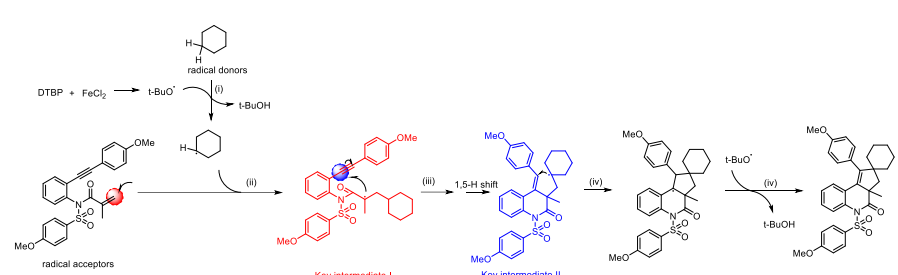
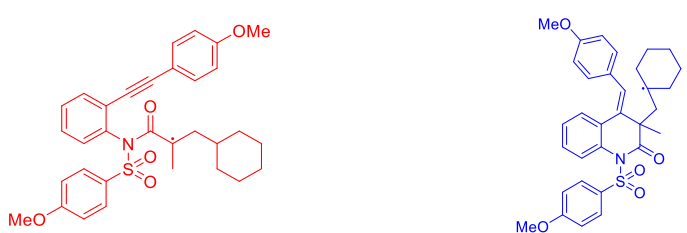
Entry	438
Source	J. K. Qiu <i>et al.</i> , Catalytic Dual 1,1-H-Abstraction/Insertion for Domino Spirocyclizations. <i>J. Am. Chem. Soc.</i> 137 , 8928-8931 (2015).
Information	2 new rings construct; C- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C1CCCCC1.C=C(C)C(=O)N(c1ccccc1C#Cc1ccc(OC)cc1)S(=O)(=O)c1ccc(OC)cc1>>COc1ccc(C2=C3c4ccccc4N(S(=O)(=O)c4cc(OC)cc4)C(=O)C3(C)CC23CCCCC3)cc1 COc1ccc(C#Cc2cccc2N(C(=O)CCC2CCCC2)S(=O)(=O)c2ccc(OC)cc2)cc1 COc1ccc(/C=C2\c3ccccc3N(S(=O)(=O)c3ccc(OC)cc3)C(=O)C2(C)C[C]2CCCCC2)cc1</chem>

Table S11. Example XI randomly selected from the self-built dataset.

Entry	449
Source	J. Xuan <i>et al.</i> , Construction of Polycyclic gamma-Lactams and Related Heterocycles via Electron Catalysis. <i>Org. Lett.</i> 18 , 6372-6375 (2016).
Information	2 new rings constructed; C- centre radical
chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<chem>C=C(CN(C(=O)C#Cc1cccc1)S(=O)(=O)c1ccc(C)cc1)C(=O)O</chem> <chem>C.Nc1ccc(Cl)cc1>>COC(=O)C12Cc3ccc(Cl)cc3C(c3ccccc3)=C</chem> <chem>1C(=O)N(S(=O)(=O)c1ccc(C)cc1)C2 COC(=O)[C](Cc1ccc(Cl)</chem> <chem>cc1)CN(C(=O)C#Cc1cccc1)S(=O)(=O)c1ccc(C)cc1 COC(=O)</chem> <chem>C1(Cc2ccc(Cl)cc2)CN(S(=O)(=O)c2ccc(C)cc2)C(=O)/C1=[C]c</chem> <chem>1cccc1</chem>

Table S12. Example XII randomly selected from the self-built dataset.

Entry	688
Source	K. Pati <i>et al.</i> , Traceless directing groups in radical cascades: from oligoalkynes to fused helicenes without tethered initiators. <i>J. Am. Chem. Soc.</i> 137 , 1165-1180 (2015).
Information	2 new rings constructed; Sn- centre radical
Chemical equation	
Proposed mechanism	
Key intermediate	
Relevant data in the database	<pre>CCCC[SnH](CCCC)CCCC.COc1ccc(C#Cc2ccccc2)c(C(C#Cc2c cccc2)OC)c1>>COc1ccc2c3c(ccc2c1)c1cccc1C3c1cccc1 CC CC[Sn](CCCC)(CCCC)/C(=[C]\c1cccc1)C(OC)c1cc(OC)ccc1 C#Cc1cccc1 CCCC[Sn](CCCC)(CCCC)C1=C(c2ccccc2)/C(=[C]/c2ccccc2)c2ccc(OC)cc2C1OC</pre>

5. The Transformer model

The key component of the Transformer is Multi-Head Attention (MHA). For each head, the equation is defined as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q^T K}{\sqrt{d_k}}\right) V$$

where Q , K , V are input embedding matrices and d_k is the embedding dimension. With the MHA, the information can be handled parallelly on different subspaces. At present, the Transformer remains a popular architecture for a wide variety of problems, including NLP, computer vision (CV), and reaction prediction.

The Transformer model based on the work from Duan *et al.*^{4,5} was applied in this study to predict the reaction intermediates (Fig S5). The model is based on an encoder-decoder architecture. The encoder is composed of several same layers and each layer contains two sub-layers: a multi-head self-attention mechanism composed of several parallel scaled dot-product attention layers, and a simple position feedforward network. Before layers normalization, the model introduces a residual connection to each of the two sub-layers. The decoder is also composed of the same layers, each layer is composed of three sub-layers. In addition to the same two sub-layers as each encoder layer, a masked multi-head self-attention mechanism is added to the decoder. Similar to the encoder process, we also use residual connections around each sub-layer and then perform layers normalization.

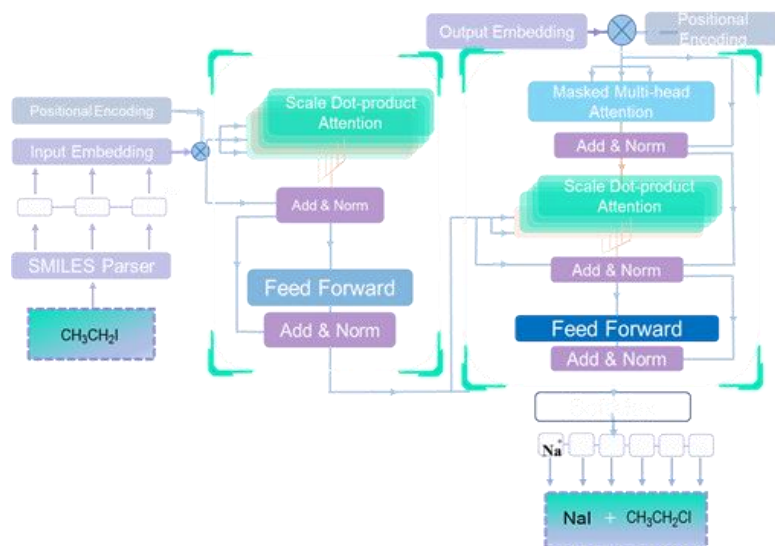


Fig. S5 The Transformer model for radical intermediate prediction.

Atom-based tokenization strategy, which reduces the vocab size to relatively small and constant size for most representation of organic molecules, was adopted. This tokenization strategy can perfectly avoid correcting vocabulary size for different models and Out Of Vocabulary (OOV) problems. Because of the small vocabulary size, the results turn out to increase the accuracy and robustness of models, meanwhile, the confidence of the model increases in test/valid procedure. The vocabulary comprises chemical elements in the periodic table and special symbols defined in SMILES. The vocab file is available at <https://github.com/hongliangduan/transRCC>

The following hyperparameters were used for key intermediate prediction:

approx_num: 800

optimize adam:

- beta1: 0.9

- beta2: 0.997

epsilon: 1e-9

n_heads: 8

emb_dim: 256

num_layers: 6

FFN_inner_units: 2048

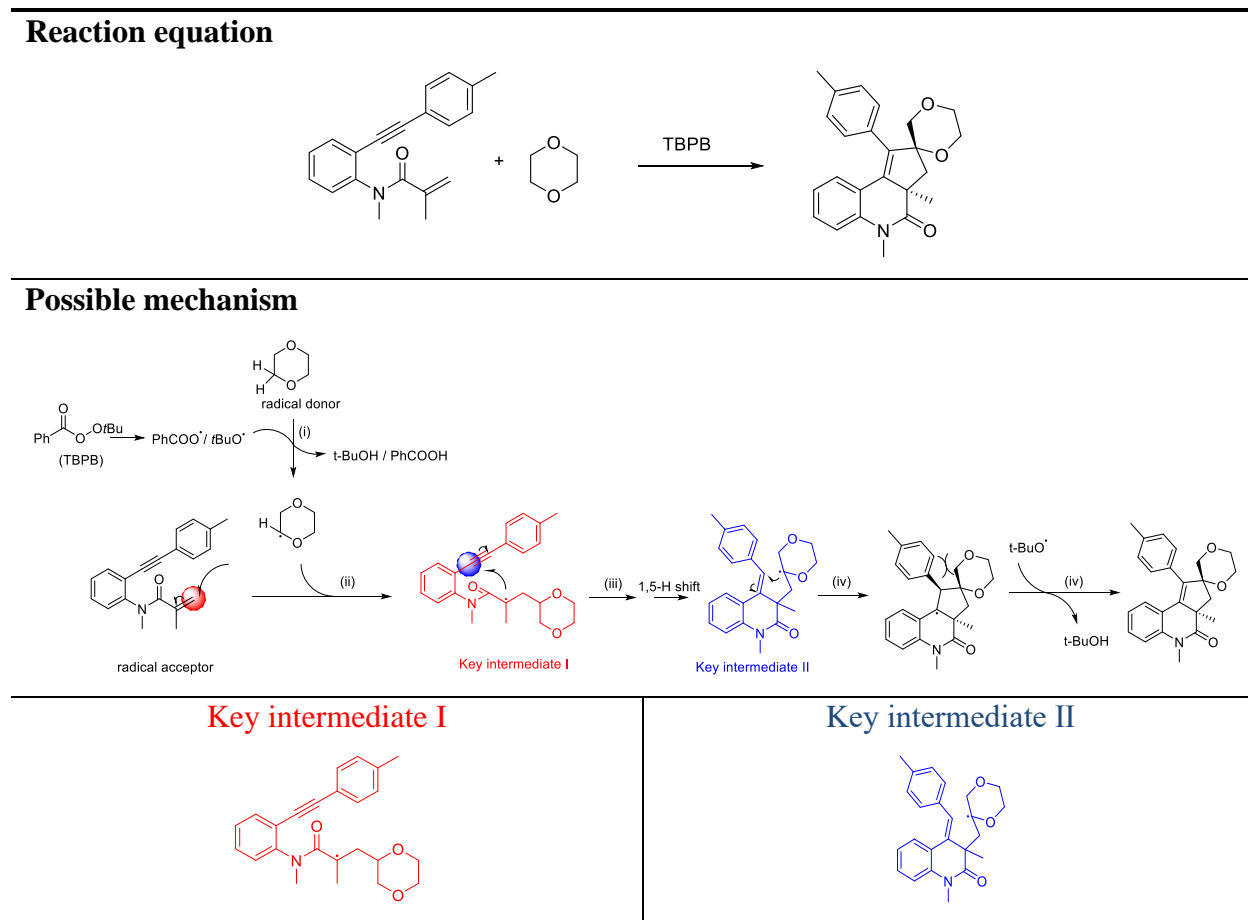
dropout: 0.3

6. The impact of different text representations

To compare the impact of different text representations on intermediate prediction accuracy, we compared SMILES with DeepSMILES, SELFIES, and SMARTS in this study. The experiment process is as follows.

(i) The records in the general chemical reaction dataset and self-built dataset were described by respective text representation, standard with RDKit subsequently. We randomly chose an example from self-dataset (entry=82), the reaction equation with corresponding two key intermediate were represented by different text representation (SMILES, DeepSMILES, SELFIES, SMARTS) and shown in Table S13 (ii) The pretrain process and fine-tuning process were taken. The process is as same as the process mentioned above. The training process is stopped when the reaching a steady state. (iii) Text process was taken, and Experiment 1 (refers to entry 1 in Table S18, 99 reactions) was selected as a representative example for analysis.

Table S13 a randomly chosen record from the self-built dataset (entry = 82) represented by different text representation



SMILES (canonical):

C1COCCO1.C=C(C)C(=O)N(C)c1ccccc1C#Cc1ccc(C)cc1>>Cc1ccc(C2=C3c4ccccc4N(C)C(=O)C3(C)CC23COCCO3)cc1|||C[C](CC1COCCO1)C(=O)N(C)c1ccccc1C#Cc1ccc(C)cc1|||Cc1ccc(/C=C2\c3ccccc3N(C)C(=O)C2(C)C[C]2COCCO2)cc1

DeepSMILES (canonical):

CCOCCO6.C=CC)C=O)NC)cccccc6C#CccccC)cc6>>CccccC=Ccccccc6NC)C=O)C%10C)CC%13COCCO6))))))))))))))))))cc6|||C[C]CCCCOCCO6))))))C=O)NC)cccccc6C#CccccC)cc6| ||Ccccc/C=C\cccccc6NC)C=O)C%10C)C[C]COCCO6))))))))))))))))))cc6

SELFIES (canonical):

[C][C][O][C][C][O][Ring1][=Branch1].[C][=C][Branch1][C][C][C][=Branch1][C][=O][N][Branch1][C][C][C][=C][C][=C][C][=C][Ring1][=Branch1][C][#C][C][=C][C][=C][Branch1][C][C][C][=C][Ring1][#Branch1]>>[C][C][=C][C][=C][Branch2][Ring2][Ring2][C][=C][C][=C][C][=C][C][=C][Ring1][=Branch1][N][Branch1][C][C][C][=Branch1][C][=O][C][Ring1][N][Branch1][C][C][C][C][Ring1][S][C][O][C][C][O][Ring1][=Branch1][C][=C][Ring2][Ring1][O]|||C[C][CH0][Branch1][#Branch2][C][C][C][O][C][C][O][Ring1][=Branch1][C][=Branch1][C][=O][N][Branch1][C][C][C][=C][C][=C][C][=C][Ring1][=Branch1][C][#C][C][=C][C][=C][Branch1][C][C][C][=C][Ring1][#Branch1]|||C[C][=C][C][=C][Branch2][Ring2][C][C][=C][C][=C][C][=C][Ring1][=Branch1][N][Branch1][C][C][C][=Branch1][C][=O][C][Ring1][N][Branch1][C][C][C][CH0][C][O][C][C][O][Ring1][=Branch1][C][=C][Ring2][Ring1][O]

SMARTS (canonical):

[#6]1-[#6]-[#8]-[#6]-[#6]-[#8]-1.[#6]=[#6](-[#6])-[#6](=[#8])-[#7](-[#6])-[#6]1:[#6]:[#6]:[#6]:[#6]:[#6]:1-[#6]#[#6]-[#6]1:[#6]:[#6]:[#6]:[#6](-[#6]):[#6]:[#6]:1>>[#6]-[#6]1:[#6]:[#6]:[#6](-[#6]2=[#6]3-[#6]4:[#6]:[#6]:[#6]:[#6]:[#6]:4-[#7](-[#6])-[#6](=[#8])-[#6]-3(-[#6])-[#6]-[#6]-23-[#6]-[#8]-[#6]-[#6]-[#8]-3):[#6]:[#6]:1|||[#6]-[#6](-[#6]-[#6]1-[#6]-[#8]-[#6]-[#6]-[#8]-1)-[#6](=[#8])-[#7](-[#6])-[#6]1:[#6]:[#6]:[#6]:[#6]:[#6]:1-[#6]#[#6]-[#6]1:[#6]:[#6]:[#6](-[#6]):[#6]:[#6]:1|||[#6]-[#6]1:[#6]:[#6]:[#6](/[#6]=[#6]2[#6]3:[#6]:[#6]:[#6]:[#6]:3-[#7](-[#6])-[#6](=[#8])-[#6]-2(-[#6])-[#6]-[#6]2-[#6]-[#8]-[#6]-[#6]-[#8]-2):[#6]:[#6]:1

The result shows that SMILES fits the model with the highest accuracy (Table S14), while other text representations are not suitable for our model. The Transformer model exists poor performance in avoiding invalid SELFIES in this task. One probable reason is that the scale of the dataset is too small, and the second is that the model was initially designed for the representation of SMILES.

The dataset expressed by different representations (pretraining and self-built) are available from: <https://github.com/hongliangduan/transRCC>.

Table S14 comparison of prediction result with different text representations

Entry	text representation	Top-1 accuracy (%)	
		Key intermediate I	Key intermediate II
1	SMILES	92.9	88.8
2	DeepSMILES	74.5	43.9
3	SELFIES	0	0
4	SMARTS	52.9	49.4

7. Details of cross-validations experiments

Table S15. Details of cross-validations experiments: Prediction work via Transformer-transfer learning model. The Transformer-transfer model exhibits an accuracy of 94.5% for key intermediate I and 92.5% for key intermediate II. With the aid of the general chemical reactivity rules and knowledge obtained in the pretraining process, the Transformer-transfer learning model was more accurate and be well used to cope with radical cascade cyclization's intermediate prediction.

Entry	Accuracy (%)	
	Key intermediate I	Key intermediate II
1	92.9	88.8
2	95.7	94.6
3	92.3	93.4
4	95.5	93.3
5	92.9	92.9
6	98.8	97.5
7	96.3	91.5
8	92.2	87.0
9	94.6	93.2
10	94.3	92.9
Average	94.5	92.5

Table S16. Details of crossvalidations experiments: Prediction work via Transformer baseline model. Transformer- baseline model exhibits the accuracy of 29.7% for key intermediate I and 26.6% for key intermediate II.

Entry	Accuracy (%)	
	Key intermediate I	Key intermediate II
1	29.6	29.6
2	26.9	24.7
3	31.8	27.4
4	33.7	29.2
5	30.6	25.9
6	39.8	31.3
7	21.3	23.8
8	20.8	22.1
9	31.1	24.3
10	31.4	28.6
Average	29.7	26.6

Table S17 Details of cross-validations experiments: Prediction of two key intermediates simultaneously. The Transformer-transfer learning model was trained to predict two key intermediates simultaneously. Although the difficulty increased, the results showed no significant decrease in the performance with the prediction accuracy rate.

Entry	Accuracy (%)		
	Key intermediate I	Key intermediate II	Key intermediate I&II ^a
1	90.8	85.7	81.6
2	97.8	93.5	93.5
3	93.4	94.5	90.1
4	93.3	88.8	87.6
5	94.1	93.0	91.8
6	92.8	90.4	89.2
7	97.5	92.5	91.3
8	88.3	87.0	81.8
9	94.6	93.2	93.2
10	91.4	88.6	87.1
Average	93.4	90.72	88.72

^a Predict two key intermediates simultaneously.

Table S18 The detailed accuracies of the Transformer-transfer learning of different number of newly constructed rings. For the convenience of analysis and understanding, reactions were classified according to the number of newly constructed rings. Experiment 1 (refers to entry 1 in table S13) was selected as a representative example for analysis (99 reactions). The detailed accuracies of the Transformer-transfer learning of different number of newly constructed rings are described. With the increasing of the constructed rings, the accuracy decreases. For key intermediate I, as the number of rings increases, the prediction accuracy rates could be got with 94.2%, 92.5% and 83.3%, respectively. In some degree, the accuracy depends on the complexity of regio- or stereo-selectivity, and the training data decrease as the number of newly construct rings increases. A similar result was obtained with respect to the intermediate II, and the accuracy was 93.8%, 82.5%, and 66.7% as the number of rings increase. The accuracy of key intermediate I is slightly higher than intermediate II with accuracy of 92.9% and 88.8%, respectively.

Type of reaction ^a	Accuracy (%)		
	Key intermediate I	Key intermediate II	Key intermediate I&II ^b
1	94.2	93.8	90.4
2	92.5	82.5	77.5
≥3	83.3	66.7	66.7
Total	92.9	88.8	81.6

^aBased on the number of newly constructed rings. ^bPredict two key intermediates simultaneously.

8. Analysis of “wrong” predictions

With introducing transfer learning methods, the Transformer-transfer learning model achieves a higher accuracy than the Transformer-baseline model. However, there are several error types that continuously accompany the Transformer-transfer learning model to tackle the intermediate prediction task: carbon number error, group migration error, SMILES error and other errors.

Table S19 Analysis of “wrong” predictions: error of carbon number. According to original documented literature, this kind of error occurs when the reactants have a large and complex chemical structure. Reactants in the radical cascade cyclization have large and complex chemical structures, and the probability of group migration error in this case also increases accordingly. Because of the lack of mathematical knowledge, it sometimes makes mistakes when it counts carbon atoms (a), Groups at sites with strong steric hindrance (b), and less common groups such as cyanide (c), and halogen (d) are sometimes overlooked. Although there are local additions or deletions of atoms or groups, the overall structure is consistent with the fact of the reaction, which proves that the model has a deep understanding of the radical cascade cyclization.

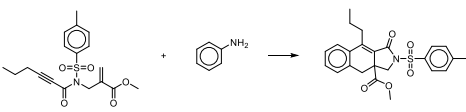
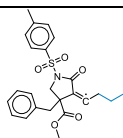
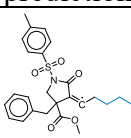
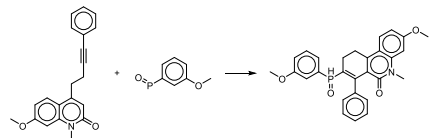
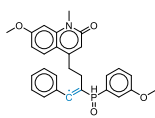
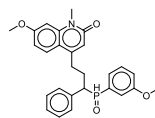
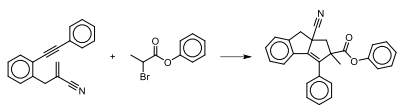
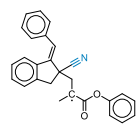
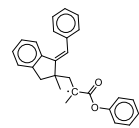
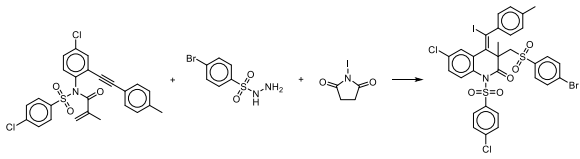
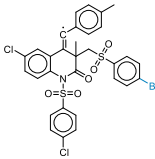
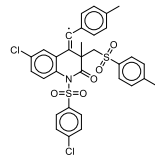
Entry	Chemical equation (input)	Ground truth	Wrong prediction
a			
b			
c			
d			

Table S20 Analysis of “wrong” predictions: error of group migration. The Transformer model was misled by the structure of the substrate, but the attack site was still correct, further showing that the model had grasped the mechanism of radical cascade cyclization.

Entry	Chemical equation (input)	Ground truth	Wrong prediction
e			
f			

Table S21 Analysis of “wrong” predictions: error of SMILES. The mis-prediction of SMILES is a flaw in the Transformer model, which is prevalent in a variety of prediction tasks. This error is mainly reflected in the misuse of slashes and backslashes in this intermediate prediction task, and (g) shows some representative examples of SMILES errors. Another manifestation of this error is the double bond cis-trans isomerism error (h).

<p>target <chem>CCCC[Sn](CCCC)(CCCC)C1=C(c2ccccc2)C([C]c2ccccc2)c2ccccc21</chem></p> <p>prediction <chem>CCCC[Sn](CCCC)(CCCC)C1=C(c2ccccc2)/C(=[C]/c2ccccc2)c2ccccc21</chem></p> <p>g</p>	<p>target <chem>[C]C1(C)C(=O)N(S(=O)(=O)c2ccc(C)cc2)c2ccc(Cl)cc2/C1=C(c1ccccc1)c1ccc(OC)cc1</chem></p> <p>prediction <chem>[C]C1(C)C(=O)N(S(=O)(=O)c2ccc(C)cc2)c2ccc(Cl)cc2C1=C(c1ccccc1)c1ccc(OC)cc1</chem></p> <p>h</p>
---	---

9. Computational studies

The radical cascade cyclization of 1,6-enynes with aryl sulfonyl chlorides by using visible-light-induced catalysis is shown in Fig. 8a.

A mechanism Li *et al.*⁶ proposed is shown in Fig S6 (Path I). They described that an aryl radical is firstly formed by a single-electron transfer from the excited state $[\text{Ru}(\text{bpy})_3]^{2+}$ to an aryl sulfonyl chloride, and subsequent addition of the aryl radical to the triple bond result in radical intermediate I. Then intermediate I undergo the cyclization reaction with the alkene to yield intermediate II. After intramolecular cyclization process of intermediate II, the cyclic radical is oxidized to the corresponding cyclic cation and subsequently transformed into the target product after deprotonation, accompanied by regeneration of the active $[\text{Ru}(\text{bpy})_3]^{2+}$ species.

The Transformer model provides additional information by identifying key intermediates I and II, implying that the reaction may have proceeded via a pathway (Fig. S7, PathII) that is differs from the mechanism mentioned above. Here, the excited $[\text{Ru}(\text{bpy})_3]^{2+}$, sulfonyl radicals are formed by selective cleavage of the S–Cl bond. Subsequently, the sulfonyl radicals add to the terminal alkene to generate tertiary alkyl radicals. The 5-exo-cyclization leads to the formation of a vinyl radical that further undergoes 1,5-aryl migration and the subsequent release of SO_2 to give rise to the primary alkyl radical intermediate. The following process is the same as that inferred by Li *et al.*

Computational studies indicated that aryl radicals are not easily generated in the visible-light-induced cyclization of 1,6-enynes with aryl sulfonyl ($\Delta G = -44.0$ vs -65.4 kcal/mol). according to the Curtin-Hammett principle, as long as equilibration of such radicals proceeds faster than any of their subsequent reactions, the reaction outcome is determined by the lowest energy transition state. Further computational studies (Fig S8) reveal that the pathway proposed by Transformer model is more facile than the reported one. All the density functional theory (DFT) calculations were performed using a Dmol3 Module.^{7,8} The generalized gradient approximation (GGA) method with Perdew-Burke-Ernzerhof (PBE) function was executed to simulate the molecule. A global orbital cutoff of 4.5 \AA was used. And the force and energy convergence criterion were set to 0.002 Ha/\AA and 10^{-5} Ha , respectively.

The details are as follows:

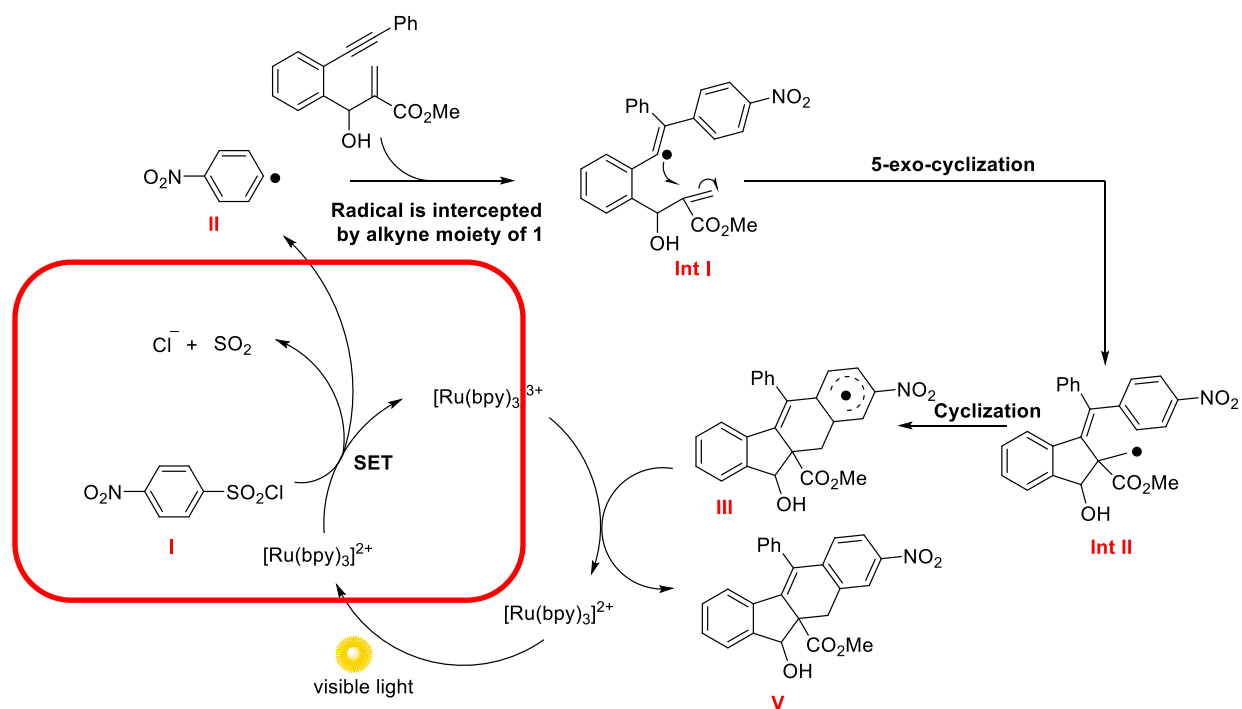
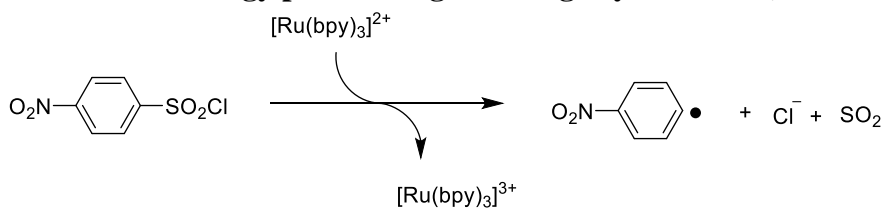


Fig. S6 Path I (proposed by Li *et al.*)

Table S22 Calculated free energy profile for generating aryl radicals (circled in red)



	E total (Ha) (kcal/mol)	$G_{\text{Total}}^{298\text{K}}$ (Ha) (kcal/mol)	$E_{\text{Tcorr}}^{298\text{K}}$ (Ha) (kcal/mol)
Ru(ii)*	-5927.76215	0.40364	-5926.85851
ArSO ₂ Cl	-1444.21755	0.05872	-1444.21755
Aryl radical	-435.74653	0.05545	-435.74653
SO ₂	-548.39698	-0.01855	-548.41553
Cl ⁻	-460.08451	0	-460.08451
Ru(iii)	-5927.36191	0.40683	-5926.95508

$$\Delta G_{\text{reaction}}^{298\text{K}} = 627.5 \left[\sum E_{\text{Tcorr}}^{298\text{K}}(\text{product}) - \sum E_{\text{Tcorr}}^{298\text{K}}(\text{reaction}) \right] = -44.0 \text{ kcal/mol}$$

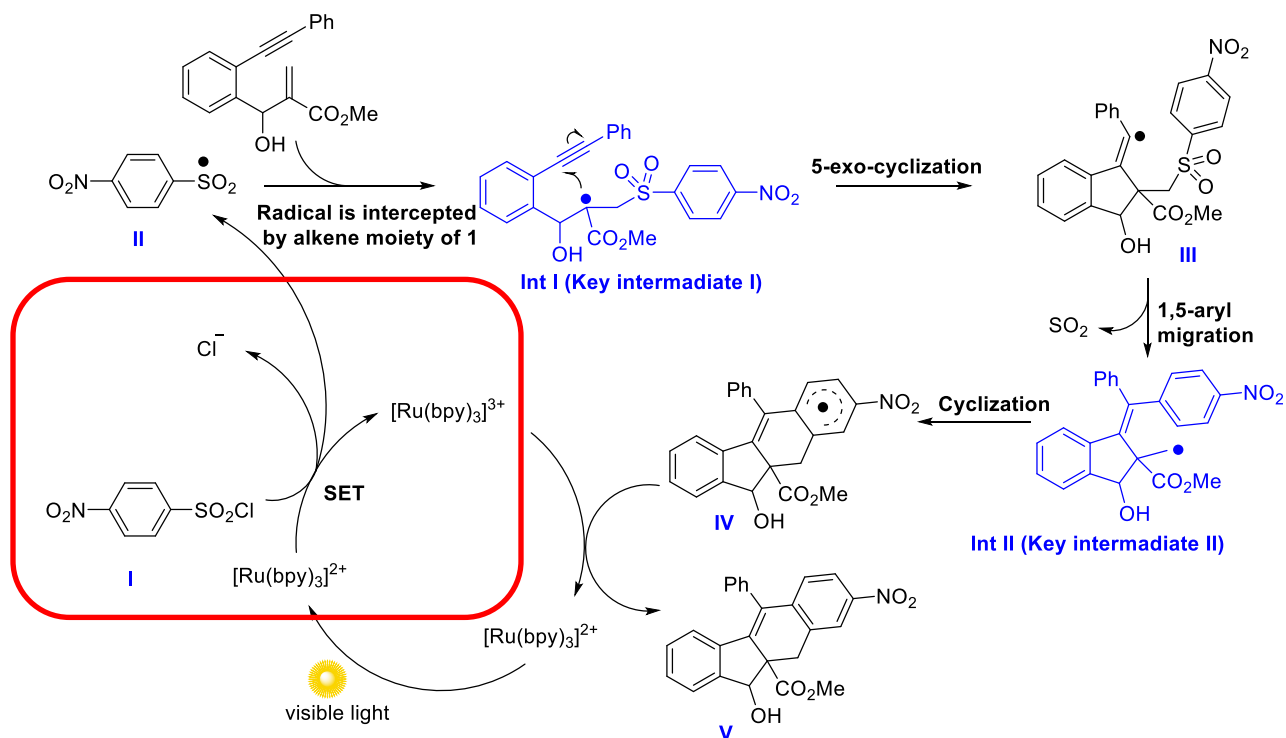
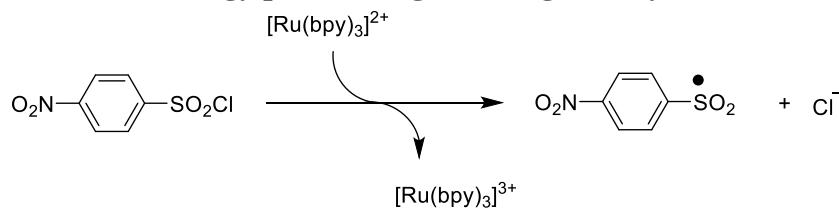


Fig. S7 Path II (proposed by Transformer model)

Table S23 Calculated free energy profile for generating sulfonyl radicals (circled in red)



	E total (Ha) (kcal/mol)	$G_{\text{Total}}^{298\text{K}}$ (Ha) (kcal/mol)	$E_{\text{Tcorr}}^{298\text{K}}$ (Ha) (kcal/mol)
Ru(ii)*	-5927.76215	0.40364	-5926.85851
ArSO ₂ Cl	-1444.21755	0.05872	-1444.21755
Sulfonyl radical	-984.19995	0.05931	-984.14064
Cl ⁻	-460.08451	0	-460.08451
Ru(iii)	-5927.36191	0.40683	-5926.95508

$$\Delta G_{\text{reaction}}^{298\text{K}} = 627.5 \left[\sum E_{\text{Tcorr}}^{298\text{K}}(\text{product}) - \sum E_{\text{Tcorr}}^{298\text{K}}(\text{reaction}) \right] = -65.4 \text{ kcal/mol}$$

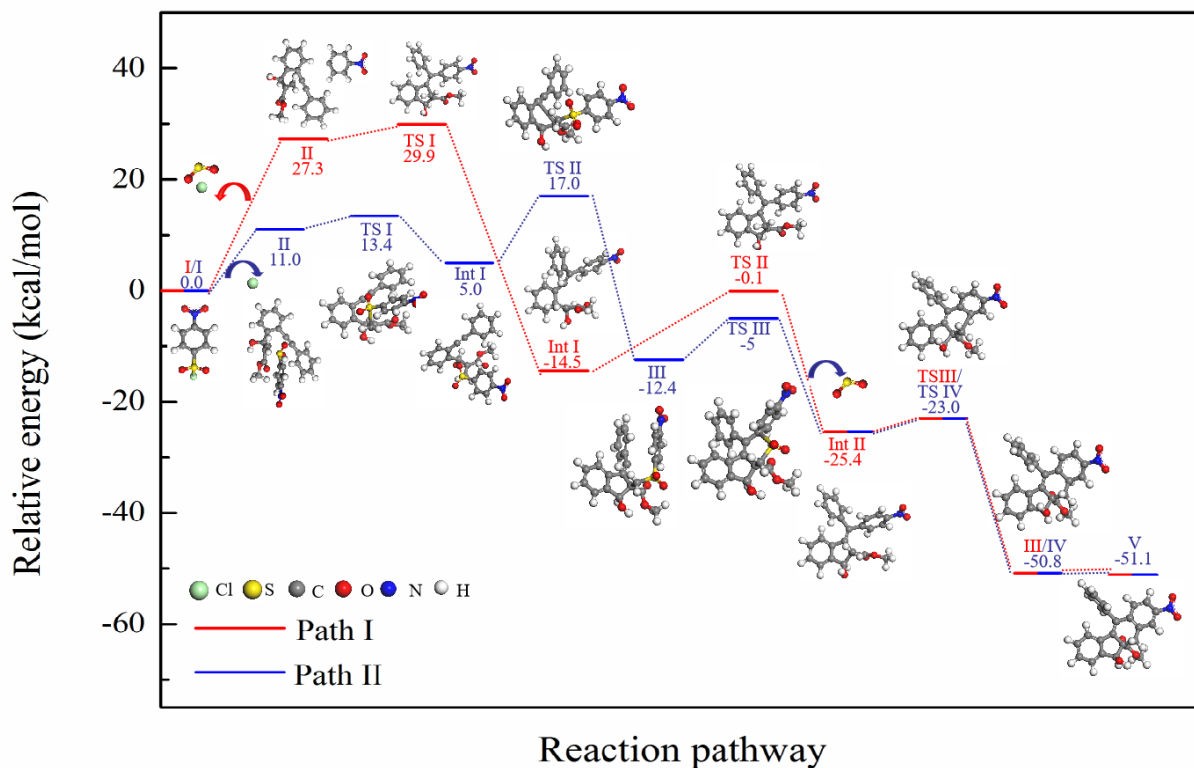


Fig. S8 Calculated free energy profile for the two possible mechanisms. Path I (proposed by Li *et al.*) Vs Path II (proposed by Transformer model)

10. References

- 1 J. K. Qiu, B. Jiang, Y. L. Zhu, W. J. Hao, D. C. Wang, J. Sun, P. Wei, S. J. Tu, G. Li, *J. Am. Chem. Soc.*, 2015, **137**, 8928-8931.
- 2 Y. Li, B. Liu, R. J. Song, Q. A. Wang, J. H. Li, *Adv. Synth. Catal.*, 2016, **358**, 1219-1228.
- 3 S. Zhou, S. Bommezzjin, and John A. M., *Org. Lett.*, 2002, **4**, 443-445.
- 4 Y. Zhang, L. Wang, X. Wang, C. Zhang, J. Ge, J. Tang, A. Su, H. Duan, *Org. Chem. Front.*, 2021, **8**, 1415-1423.
- 5 L. Wang, C. Zhang, R. Bai, J. Li, H. Duan, *Chem. Commun.*, 2020, **56**, 9368-9371.
- 6 G. B. Deng, Z. Q. Wang, J. D. Xia, P. C. Qian, R. J. Song, M. Hu, L. B. Gong, J. H. Li, *Angew. Chem. Int. Ed.*, 2013, **52**, 1535-1538.
- 7 B. Delley, *J. Chem. Phys.*, 2000, **113**, 7756-7764.
- 8 T. Kar, S. Scheiner, *J. Am. Chem. Soc.*, 1995, **117**, 1344-1351.