Generating skeleton reaction network for reactions of large-scale ReaxFF MD pyrolysis simulations based on machine learning predicted reaction class

Shanwen Yang, a,b Xiaoxia Li, a,b,c,* Mo Zheng a,b,c Chunxing Ren, a,b,c and Li Guo a,b,c

aState Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, P. R. China

bSchool of Chemical Engineering, University of Chinese Academy of Sciences, Beijing 100049, P. R. China

cInnovation Academy for Green Manufacture, Chinese Academy of Sciences, Beijing 100190, P. R. China

E-mail: xxia@ipe.ac.cn

SUPPORTING INFORMATION

Supplementary materials S1

Table S1.1 Reactions classes (RxC) of hydrocarbon pyrolysis defined for the classification of ReaxFF MD simulated reactions in SRG-Reax

<table>
<thead>
<tr>
<th>RxC</th>
<th>RxC name</th>
<th>Reaction instances</th>
<th>Reaction depiction &amp; Classification priority note</th>
<th>Reaction class group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C-C bond homolysis</td>
<td>CH₂CH₂ → CH₂⁻ + H₂CH⁻</td>
<td>Homolysis of C-C at varied positions</td>
<td>I. Homolysis</td>
</tr>
<tr>
<td>2</td>
<td>H-H bond homolysis</td>
<td>H → H⁺ + H⁻</td>
<td>H₂ formed at high temperatures is unstable, which may cleavage into H</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>β-scission</td>
<td>C⁻H⁻ → H⁻CH⁺ + H₂⁻</td>
<td>β-scission of radicals at varied positions</td>
<td>II. Carbon chain breaks induced by radicals, including chain branching</td>
</tr>
<tr>
<td>4</td>
<td>α-scission</td>
<td>H₂⁻C⁻CH₃ → CH₂⁻ + CH₃⁺</td>
<td>Infrequent reaction at high temperature</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Intra-molecular chain isomerization</td>
<td>H⁻C⁻CH₂CH₃ → H⁻CH₂⁺ + CH₃⁻</td>
<td>Intra-molecular isomerization may lead to branching</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Intra-molecular H-shift</td>
<td>OH⁻CH⁺ → H⁻ + CH⁻</td>
<td>H-shift may occur to its neighbor C atom or other C atoms</td>
<td>III. Intra-molecular H-shift on a</td>
</tr>
</tbody>
</table>
IV. Dehydrogenation of a carbon chain

V.

Inter-molecular H-abstraction by C

H-abstraction of H₂ by C

High-temperature reaction due to H₂ unstableness, where H₂ is considered as C₀

VI.

H radical addition to C

Recombination of C radicals

Combination of H radicals

Isopropyl detachment induced by C chain radical

Isopropyl detachment induced by isopropyl radical

VII.

α-branched bond scission of C ring

β-branched bond scission of C ring

Branch shift of C ring

Reconnection of the dropped side chain to other C of same ring

VIII.

Polycyclic bridge cleavage into a large ring

The type of polycyclic bridge cleavage has higher priority than ring opening

IX.

Adjacent C-C ring bond cleavage of

Ring-opening: relevant to the bridge bond of polycyclics
### Polycyclic Bridge Bond

#### β-ring opening of ring carbon radical

![β-ring opening of ring carbon radical](image1)

#### β-ring opening of branched carbon radical

![β-ring opening of branched carbon radical](image2)

#### α-ring opening of ring carbon radical

![α-ring opening of ring carbon radical](image3)

#### Infrequent reaction (α-ring opening of ring carbon radical tends to its recombination with other C radical)

#### α-ring bond scission of ring branch

![α-ring bond scission of ring branch](image4)

#### β-ring bond scission of ring branch

![β-ring bond scission of ring branch](image5)

#### Ring bond scission except for α and β ring bond of ring branch

![Ring bond scission except for α and β ring bond of ring branch](image6)

### Ring-Opening: Induced by Radical

#### Ring-opening at ring C neighboring to branching C

![Ring-opening at ring C neighboring to branching C](image7)

### Ring-Opening: Relative to the Position of the Branched Chain

#### Ring bond scission lower than the reaction types of Groups IX and X.

![Ring bond scission lower than the reaction types of Groups IX and X.](image8)

### Ring-Dehydrogenation

#### Dehydrogenation of C-H on a ring

![Dehydrogenation of C-H on a ring](image9)

### H₂ formation via dehydrogenation of ring and acyclic fragment

![H₂ formation via dehydrogenation of ring and acyclic fragment](image10)

### H₂ formation via dehydrogenation of ring and α-C of ring branch

![H₂ formation via dehydrogenation of ring and α-C of ring branch](image11)

### H₂ formation of ring H with H from acyclic dehydrogenation

![H₂ formation of ring H with H from acyclic dehydrogenation](image12)

### XII. Cyclic-relevant dehydrogenation

This type focuses on ring branch dehydrogenation of α-C, which tends to consequent ring opening.

Ring branch dehydrogenation of β/γ-C will be classified into Type 29.
31. H₂ formation via dehydrogenation of α-C of ring branch and acyclic fragment

32. Cyclic H-abstraction by H

33. H-abstraction of α-C on ring branch by H

34. Cyclic H-abstraction by acyclic radical

35. Acyclic H-abstraction by ring radical

36. Intra-ring H-shift

37. Ring branch H-shift to ring

38. Ring H-shift to ring branch

39. Ring increment via branch linked to host ring

40. Ring increment via intra-ring carbons connecting

41. Ring opening and recombination

42. Chain cyclization

XIII. Cyclic-relevant H-abstraction

XIV. Cyclic-relevant H-shift

XV. Formation and isomerization of ring structures, a possible pathway for aryl ring opening

Similar to chain isomerization

Chain isomerization into a ring via connecting of intra-chain carbon
Combination of aromatic carbon and C radical fragment

Combination of aromatic carbon and H

H-abstraction by aromatic carbon

H-shift from branched-chain to host aromatic ring

XVI. Aromatic ring π structure broken by aromatic carbon joining other structural fragments or atoms, a possible pathway for aryl ring opening

Table S1.2 Fuel models, ReaxFF MD pyrolysis simulation conditions covered, and reaction analysis parameters of VARxMD for preparing reaction data set

<table>
<thead>
<tr>
<th>Fuel model</th>
<th>Simulation condition</th>
<th>Reaction duration time (ps)</th>
<th>Sampling interval of ReaxFF MD simulation / Output frame interval for VARxMD analysis (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Isothermal simulation using NVT ensemble, simulation time-step = 0.1 fs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-dodecane</td>
<td>Simulation temperature (K) 2000</td>
<td>250</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simulation temperature (K) 2500</td>
<td>250</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simulation temperature (K) 3000</td>
<td>250</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-component surrogate model of RP-3</td>
<td>Simulation temperature (K) 2800</td>
<td>270</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-component surrogate model of RP-3</td>
<td>Simulation temperature (K) 2800</td>
<td>270</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-component baseline model of RP-3</td>
<td>Simulation temperature (K) 2800</td>
<td>270</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-component baseline model of RP-1</td>
<td>Simulation temperature (K) 2800</td>
<td>250</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Fig. S2 Manual check and labeling of high uncertainty predicted reactions in the active learning of SRG-Reax.

Supplementary materials S3

<table>
<thead>
<tr>
<th>Reaction feature name</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Features of broken or formed bond (Level 1)</strong></td>
<td></td>
</tr>
<tr>
<td>BondType</td>
<td>The type of broken or formed bond</td>
</tr>
<tr>
<td>IsAromatic</td>
<td>Whether the broken or formed bond is an aromatic bond</td>
</tr>
<tr>
<td>NumBridgedRings</td>
<td>How many rings does the broken or formed bond belong to</td>
</tr>
<tr>
<td>BondOrderReaxFF</td>
<td>The bond order of broken or formed bond</td>
</tr>
<tr>
<td>FunctionalGroups</td>
<td>The functional group types that the formed or broken bond belongs to</td>
</tr>
<tr>
<td><strong>Atom features of reaction site (Level 2)</strong></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>Relative atomic mass of a reaction site</td>
</tr>
<tr>
<td>NumRadicalElectrons</td>
<td>The number of radical electrons a reaction site has</td>
</tr>
<tr>
<td>Degree</td>
<td>The number of atoms connected to a reaction site (excluding those connected to H atoms)</td>
</tr>
<tr>
<td>TotalValence</td>
<td>Total valence of a reaction site</td>
</tr>
<tr>
<td>IsAromatic</td>
<td>Whether the reaction site is an aromatic atom</td>
</tr>
<tr>
<td>Reaction feature name</td>
<td>Feature description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>NumReactantFragments</td>
<td>Number of reactant fragments</td>
</tr>
<tr>
<td>NumProductFragments</td>
<td>Number of product fragments</td>
</tr>
<tr>
<td>NumBrokenBonds</td>
<td>Number of broken bonds</td>
</tr>
<tr>
<td>NumFormedBonds</td>
<td>Number of formed bonds</td>
</tr>
<tr>
<td>NumReactiveAtomsInReactants</td>
<td>Number of non-H reaction sites in reactants</td>
</tr>
<tr>
<td>NumReactiveAtomsInProducts</td>
<td>Number of non-H reaction sites in products</td>
</tr>
<tr>
<td>NumReactantFragments</td>
<td>Number of reactant fragments</td>
</tr>
<tr>
<td>NumProductFragments</td>
<td>Number of product fragments</td>
</tr>
<tr>
<td>NumBrokenBonds</td>
<td>Number of broken bonds</td>
</tr>
<tr>
<td>NumFormedBonds</td>
<td>Number of formed bonds</td>
</tr>
<tr>
<td>NumReactiveAtomsInReactants</td>
<td>Number of non-H reaction sites in reactants</td>
</tr>
<tr>
<td>NumReactiveAtomsInProducts</td>
<td>Number of non-H reaction sites in products</td>
</tr>
<tr>
<td>NumReactantFragments</td>
<td>Number of reactant fragments</td>
</tr>
<tr>
<td>NumProductFragments</td>
<td>Number of product fragments</td>
</tr>
<tr>
<td>NumBrokenBonds</td>
<td>Number of broken bonds</td>
</tr>
<tr>
<td>NumFormedBonds</td>
<td>Number of formed bonds</td>
</tr>
<tr>
<td>NumReactiveAtomsInReactants</td>
<td>Number of non-H reaction sites in reactants</td>
</tr>
<tr>
<td>NumReactiveAtomsInProducts</td>
<td>Number of non-H reaction sites in products</td>
</tr>
</tbody>
</table>

Table S3.2 The definitions of the 18 full reaction features in Input 3 of the tri-training classifier
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NumAromaticBrokenBonds</td>
<td>Number of aromatic broken bonds</td>
</tr>
<tr>
<td>NumAromaticFormedBonds</td>
<td>Number of aromatic formed bonds</td>
</tr>
<tr>
<td>NumBrokenRingBonds</td>
<td>Number of broken bonds of ring structures</td>
</tr>
<tr>
<td>NumFormedRingBonds</td>
<td>Number of formed bonds of ring structures</td>
</tr>
<tr>
<td>NumRadicalElectronsOfRxnSitesInReact</td>
<td>The total number of radical electrons of reaction site in reactants</td>
</tr>
<tr>
<td>NumRadicalElectronsOfRxnSitesInProd</td>
<td>The total number of radical electrons of reaction site in products</td>
</tr>
<tr>
<td>NumAromaticAtomsOfRxnSitesInReact</td>
<td>The total number of aromatic reaction site in reactants</td>
</tr>
<tr>
<td>NumAromaticAtomsOfRxnSitesInProd</td>
<td>The total number of aromatic reaction site in products</td>
</tr>
<tr>
<td>NumRingAtomsOfRxnSitesInReact</td>
<td>Number of reaction site of ring structures in reactants</td>
</tr>
<tr>
<td>NumRingAtomsOfRxnSitesInProd</td>
<td>Number of reaction site of ring structures in products</td>
</tr>
<tr>
<td>NumPiElectronsOfRxnSitesInReact</td>
<td>The total number of π electrons of reaction sites in reactants</td>
</tr>
<tr>
<td>NumPiElectronsOfRxnSitesInProd</td>
<td>The total number of π electrons of reaction sites in products</td>
</tr>
</tbody>
</table>

### Table S3.3 Details of 196 reaction features in vector of Input 1 highlighted in Fig. 4

#### # 0–5 full reaction features of Level 4

- 'NumReactantFragments', 'NumProductFragments', 'NumBrokenBonds', 'NumFormedBonds', 'NumReactiveAtomsInReactants', 'NumReactiveAtomsInProducts'

#### # Bond breaking

#### # 6–17 1st broken bond (BB1)

- **# bond (B) features of Level 1**
  - 'BB1_B_BondType', 'BB1_B_IsAromatic', 'BB1_B_NumBridgedRings', 'BB1_B_BondOrderReaxFF', 'BB1_B_FunctionalGroup'

- **# its environment (E) structure features of Level 3**
  - 'BB1_E_NumRadicalElectrons', 'BB1_E_NumPiElectrons', 'BB1_E_NumRingAtoms', 'BB1_E_NumNonRingRadicals', 'BB1_E_FunctionalGroupType1', 'BB1_E_FunctionalGroupType2', 'BB1_E_FunctionalGroupType3'

#### # 18–29 2nd broken bond (BB2)

- **# bond (B) features of Level 1**
  - 'BB2_B_BondType', 'BB2_B_IsAromatic', 'BB2_B_NumBridgedRings', 'BB2_B_BondOrderReaxFF', 'BB2_B_FunctionalGroup'

- **# its environment (E) structure features of Level 3**
  - 'BB2_E_NumRadicalElectrons', 'BB2_E_NumPiElectrons', 'BB2_E_NumRingAtoms', 'BB2_E_NumNonRingRadicals', 'BB2_E_FunctionalGroupType1', 'BB2_E_FunctionalGroupType2', 'BB2_E_FunctionalGroupType3'

#### # Bond formation

#### # 30–37 1st formed bond (FB1)

- **# bond (B) features of Level 1**
  - 'FB1_B_BondType', 'FB1_B_IsAromatic', 'FB1_B_NumBondRings', 'FB1_B_BondOrder'

- **# its environment (E) structure features of Level 3**
  - 'FB1_E_NumRadicalElectrons', 'FB1_E_NumPiElectrons', 'FB1_E_NumAtomRings', 'FB1_E_NumNonRingRadicals'

#### # 38–45 2nd formed bond (FB2)

- **# bond (B) features of Level 1**
  - 'FB2_B_BondType', 'FB2_B_IsAromatic', 'FB2_B_NumBondRings', 'FB2_B_BondOrder'

- **# its environment (E) structure features of Level 3**
  - 'FB2_E_NumRadicalElectrons', 'FB2_E_NumPiElectrons', 'FB2_E_NumAtomRings', 'FB2_E_NumNonRingRadicals'
# Atoms of reaction sites in reactants

### # 46–60 atom features of the 1st reaction site (BA1) in reactants of Level 2

- BA1_Mass
- BA1_NumRadicalElectrons
- BA1_Degree
- BA1_TotalValence
- BA1_IsAromatic
- BA1_NumBridgedRings
- BA1_TotalNumHs
- BA1_GasteigerCharges
- BA1_CrippenContribs
- BA1_LabuteASAContribs
- BA1_TPSAContribs
- BA1_NumPiElectrons
- BA1_NumLonePairElectronsReaxFF
- BA1_NetChargeReaxFF
- BA1_TotalBondOrderReaxFF

### # 61–75 atom features of the 2nd reaction site (BA2) in reactants of Level 2

- BA2_Mass
- BA2_NumRadicalElectrons
- BA2_Degree
- BA2_TotalValence
- BA2_IsAromatic
- BA2_NumBridgedRings
- BA2_TotalNumHs
- BA2_GasteigerCharges
- BA2_CrippenContribs
- BA2_LabuteASAContribs
- BA2_TPSAContribs
- BA2_NumPiElectrons
- BA2_NumLonePairElectronsReaxFF
- BA2_NetChargeReaxFF
- BA2_TotalBondOrderReaxFF

### # 76–90 atom features of the 3rd reaction site (BA3) in reactants of Level 2

- BA3_Mass
- BA3_NumRadicalElectrons
- BA3_Degree
- BA3_TotalValence
- BA3_IsAromatic
- BA3_NumBridgedRings
- BA3_TotalNumHs
- BA3_GasteigerCharges
- BA3_CrippenContribs
- BA3_LabuteASAContribs
- BA3_TPSAContribs
- BA3_NumPiElectrons
- BA3_NumLonePairElectronsReaxFF
- BA3_NetChargeReaxFF
- BA3_TotalBondOrderReaxFF

### # 91–105 atom features of the 4th reaction site (BA4) in reactants of Level 2

- BA4_Mass
- BA4_NumRadicalElectrons
- BA4_Degree
- BA4_TotalValence
- BA4_IsAromatic
- BA4_NumBridgedRings
- BA4_TotalNumHs
- BA4_GasteigerCharges
- BA4_CrippenContribs
- BA4_LabuteASAContribs
- BA4_TPSAContribs
- BA4_NumPiElectrons
- BA4_NumLonePairElectronsReaxFF
- BA4_NetChargeReaxFF
- BA4_TotalBondOrderReaxFF

### # 106–120 atom features of the 5th reaction site (BA5) in reactants of Level 2

- BA5_Mass
- BA5_NumRadicalElectrons
- BA5_Degree
- BA5_TotalValence
- BA5_IsAromatic
- BA5_NumBridgedRings
- BA5_TotalNumHs
- BA5_GasteigerCharges
- BA5_CrippenContribs
- BA5_LabuteASAContribs
- BA5_TPSAContribs
- BA5_NumPiElectrons
- BA5_NumLonePairElectronsReaxFF
- BA5_NetChargeReaxFF
- BA5_TotalBondOrderReaxFF

# Atoms of reaction sites in products

### # 121–135 atom features of the 1st reaction site (FA1) in products of Level 2

- FA1_Mass
- FA1_NumRadicalElectrons
- FA1_Degree
- FA1_TotalValence
- FA1_IsAromatic
- FA1_NumBridgedRings
- FA1_TotalNumHs
- FA1_GasteigerCharges
- FA1_CrippenContribs
- FA1_LabuteASAContribs
- FA1_TPSAContribs
- FA1_NumPiElectrons
- FA1_NumLonePairElectronsReaxFF
- FA1_NetChargeReaxFF
- FA1_TotalBondOrderReaxFF

### # 136–150 atom features of the 2nd reaction site (FA2) in products of Level 2

- FA2_Mass
- FA2_NumRadicalElectrons
- FA2_Degree
- FA2_TotalValence
- FA2_IsAromatic
- FA2_NumBridgedRings
- FA2_TotalNumHs
- FA2_GasteigerCharges
- FA2_CrippenContribs
- FA2_LabuteASAContribs
- FA2_TPSAContribs
- FA2_NumPiElectrons
- FA2_NumLonePairElectronsReaxFF
- FA2_NetChargeReaxFF
- FA2_TotalBondOrderReaxFF

### # 151–165 atom features of the 3rd reaction site (FA3) in products of Level 2

- FA3_Mass
- FA3_NumRadicalElectrons
- FA3_Degree
- FA3_TotalValence
- FA3_IsAromatic
- FA3_NumBridgedRings
- FA3_TotalNumHs
- FA3_GasteigerCharges
- FA3_CrippenContribs
- FA3_LabuteASAContribs
- FA3_TPSAContribs
- FA3_NumPiElectrons
- FA3_NumLonePairElectronsReaxFF
- FA3_NetChargeReaxFF
- FA3_TotalBondOrderReaxFF

### # 166–180 atom features of the 4th reaction site (FA4) in products of Level 2

- FA4_Mass
- FA4_NumRadicalElectrons
- FA4_Degree
- FA4_TotalValence
- FA4_IsAromatic
- FA4_NumBridgedRings
- FA4_TotalNumHs
- FA4_GasteigerCharges
- FA4_CrippenContribs
- FA4_LabuteASAContribs
- FA4_TPSAContribs
- FA4_NumPiElectrons
- FA4_NumLonePairElectronsReaxFF
- FA4_NetChargeReaxFF
- FA4_TotalBondOrderReaxFF

### # 181–195 atom features of the 5th reaction site (FA5) in products of Level 2
Table S3.4 Sample data of the 18 types of reaction descriptors

<table>
<thead>
<tr>
<th>RxC</th>
<th>Sample reactions</th>
<th>18 reaction features of Input 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 0 0 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 0 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 0 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5</td>
<td><img src="image4" alt="Image" /></td>
<td>1 1 1 1 3 3 0 0 0 0 1 1 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td><img src="image5" alt="Image" /></td>
<td>2 3 2 1 2 3 0 0 0 0 0 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9</td>
<td><img src="image6" alt="Image" /></td>
<td>2 2 1 1 2 2 0 0 0 0 1 0 0 0 0 0 0 1</td>
</tr>
<tr>
<td>10</td>
<td><img src="image7" alt="Image" /></td>
<td>2 2 1 1 2 2 0 0 0 0 2 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>11</td>
<td><img src="image8" alt="Image" /></td>
<td>2 2 1 1 3 2 0 0 0 0 1 1 0 0 0 0 0 0</td>
</tr>
<tr>
<td>12</td>
<td><img src="image9" alt="Image" /></td>
<td>2 1 0 1 2 1 0 0 0 0 1 1 0 0 0 1 1 1 0</td>
</tr>
<tr>
<td>13</td>
<td><img src="image10" alt="Image" /></td>
<td>2 1 0 1 2 2 0 0 0 0 2 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>14</td>
<td><img src="image11" alt="Image" /></td>
<td>2 1 0 1 2 3 0 0 0 0 2 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>15</td>
<td><img src="image12" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 0 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>16</td>
<td><img src="image13" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 1 3 0 0 0 0 0 0</td>
</tr>
<tr>
<td>17</td>
<td><img src="image14" alt="Image" /></td>
<td>1 2 1 0 2 2 0 0 0 0 0 2 0 0 0 1 1 0 0</td>
</tr>
</tbody>
</table>
Supplementary materials S4

Fig. S4 Feature importance of the top 50 among the 196 features of Input 1 of the single random forest classifier (RFC) adopted in the tri-training classifier.
Table S4 Evaluation results of reaction fingerprint (FP) candidates for single classifier of tri-training based on random forest

<table>
<thead>
<tr>
<th>Reaction descriptions</th>
<th>Micro-F1 score with different FP bit size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128</td>
</tr>
<tr>
<td>Input type</td>
<td></td>
</tr>
<tr>
<td><strong>Input 2</strong></td>
<td></td>
</tr>
<tr>
<td>(\sum) FP of products</td>
<td></td>
</tr>
<tr>
<td>- (\sum) FP of reactants</td>
<td></td>
</tr>
<tr>
<td><strong>Input 3</strong></td>
<td></td>
</tr>
<tr>
<td>(\sum) ERC FP of products</td>
<td></td>
</tr>
<tr>
<td>- (\sum) ERC FP of reactants</td>
<td></td>
</tr>
<tr>
<td>+ 18 reaction features</td>
<td></td>
</tr>
<tr>
<td><strong>Input 3 without reaction features</strong></td>
<td></td>
</tr>
<tr>
<td>(\sum) ERC FP of products</td>
<td></td>
</tr>
<tr>
<td>- (\sum) ERC FP of reactants</td>
<td></td>
</tr>
</tbody>
</table>

Supplementary materials S5
Fig. S5 Confusion matrix of all 46 classes after reaction classifier refinement.