Electronic Supplementary Material (ESI) for ChemComm.

Electronic Supplementary Information

Predicting and Analyzing Organic Reaction Pathways by Combining Machine Learning and Reaction Network Approaches

Tomonori Ida*[a], Honoka Kojima[a] and Yuta Hori[b]

[a] Division of Material Chemistry, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 920-1192, Japan.
[b] Center for Computational Sciences, University of Tsukuba, Tsukuba 305-8577, Japan.
*E-mail: ida@se.kanazawa-u.ac.jp

Table of contents

1.Computational Methods

Our program consists of two parts: Generator and Analyzer. The computational details are described separately in Sections 1.1 and 1.2. The prediction accuracy is explained in Sections 1.3. Using codes and parameters are also explained in Section 1.4.

1.1 Construction of the reaction network (Generator part)

A reaction network was constructed based on our previous work,[24] and the procedure used to construct this network is described as follows. Initially, the chemical formula of the desired reactant is converted into a molecular graph, and then each atom repeatedly dissociates and bonds according to the octet (duet) rule. Notably, radical dissociation is not considered in this study. Repetition of the operation generates numerous molecular graphs of the reaction intermediates containing ionic states, with the molecular graphs considered as nodes in the reaction network. Finally, an organic reaction network is constructed by connecting these nodes with edges that assume a single dissociation or bond between nodes. The construction of this network is shown schematically in Fig. 1(A).

For atoms and quasiparticles, such as Ph and CH_3, the numbers of valence electrons, possible charge numbers, and availabilities of the electron-deficient states are shown in Table S1. The aim of this system is to predict pathways rather than products, and thus, halogens with 7 valence electrons, such as Cl and Br, are collectively referred to as X atoms in this study. Similarly, single-valence metals, such as Li and Na, are not distinguished and are denoted as M atoms. In addition, the ionic fragments are limited to four atoms to prevent the cleavage of the molecule into smaller fragments. Only molecular graphs that satisfy the conditions for each atom of the molecular graph are retained in the generated network as nodes, and the other nodes are deleted. In the training data, the generation of a novel molecular graph is halted when the number of nodes is >800.

The obtained network comprises numerous molecular graphs with stable structures, aside from the reactant, with a stable structure defined as a molecular graph with no ionic states or electron-deficient components. In a representative reaction, one stable molecular graph is included as the correct product, and the others are classified as incorrect products, such as the corresponding structural isomers. All stable molecular graphs are defined as product candidates. Subsequently, all shortcuts from the reactant node to the product candidate nodes are scanned. The shortcuts that lead to the correct product are defined as the correct reaction pathways, whereas those leading to the incorrect products are defined as incorrect pathways. Notably, there may be more than one correct pathway, although there are generally numerous incorrect pathways, and the dashed routes shown in Fig. 1(B) indicate the correct and incorrect pathways. Subsequently, to examine the features of a reaction pathway trained by machine learning, the fragment structures of the molecular graphs of the correct and incorrect reaction pathways were analyzed. These fragment structures are used as descriptors in machine learning, and the value of the descriptor is based on the appearance frequency (average appearance rate) of each fragment structure in the pathway, as shown in Fig. 1(C). The average appearance rates of all fragment structures are used as the feature of each reaction pathway. In this study, 46 fragment structures were used as descriptors, as shown in Table S2. The features of the correct and incorrect reaction pathways in the organic reaction network were obtained.

1.2 Training of the reaction pathways (Analyzer part)

As training data, 50 basic reactions were selected from chemistry textbooks,[27-30] including 19 addition, 5 elimination, and 26 substitution reactions, as shown in Supporting Figure S2. All organic reaction networks were generated using the chemical formulae of the reactants. The total number of nodes (molecular graphs) in the generated networks was 51,793, with 941 product candidates, of which 50 were correct products. Moreover, there were 53,753 reaction pathways for all product candidates, of which 364 were correct. For each reaction, a correct and an incorrect pathway were randomly selected, and two features were learned using pairwise logistic regression, where “pairwise” refers to a correct and an incorrect pathway for each reaction. The advantages of the pairwise method were twofold: (1) it corrected the imbalance in the numbers of correct and incorrect pathways and (2) suppressed the influences of structures that were not directly related to the reaction. The learning model was trained such that the features of the correct pathway yielded higher points than those obtained using the features of the incorrect pathway. For one training reaction, a hundred of the correct and incorrect pair was selected, and total 5,000 batches for the 50 training reactions were learned in one epoch. After learning, all pathways for each reaction were ranked by the number of points (highest first), and learning was restarted until the correct pathways

were in the top 5. The loss function was a cross-entropy function, which was optimized using Adam[31] with an average loss of 0.03. The average loss was set to 0.03, and after several calculations with different initial values, the learning model converged to an accuracy of 88% for the top 1 and 100% for the top 5 on the training set and 45.7% for the top 1 on the test set. For loss values less than 0.03, the model overfitted and the top 1 accuracy of the test decreased. On the other hand, when the loss value exceeds 0.03, the model fails to achieve the top 5 of all correct pathways on the training set. The points for each descriptor are shown in Table S3.

1.3 Computational method for prediction accuracy

The prediction accuracy was scored using prediction data within the top 5 as follows: 5 points if the top 1 was correct, 4 points if the top 2 was correct, etc., e.g., in a reaction, all correct predictions from the top 1 to top 5 yield a maximum score of 15 (= 5 (top 1) + 4 (top 2) + … + 1 (top 5)), and all incorrect predictions yield a score of 0. The score of the prediction with all descriptors was 573 points out of 750 (= 15 points x 50 reactions).

1.4 Code and parameter

Program codes and all input and output data used in this work are available on GitHub (https://github.com/ida-rnet/RNet/). The program codes were written in Python 3.8.3 using Numpy 1.18.5, Torch 1.7.1 and NetworkX 2.4 modules, and the standard modules (OS and Pickle) were also imported. In the learning process, the parameters of the Adam optimizer were fixed, where the learning rate (lr) = 0.01, β_1 = 0.9 and β_2 = 0.999.

2.Limitations of This Model

The limitations of the proposed model are shown as follows. One is that the top 1 selection rate in the test data is lower than those reported in other studies. This can be improved by the incorporation of other machine-learning algorithms, such as deep neural networks. However, because complex learning models reduce the degree of explainability and induce a prediction black box, it defeats the purpose of this study. In contrast, we are interested in the extent to which our approaches can improve the top 1 selection rate, and such improvements are the subject of future work.

In addition, three other limitations of the reaction networks were identified. First, as the number of atoms handled increases, the size of the network increases rapidly and becomes unanalyzable. The figure below shows the relationship between the number of particles and the computation time on test data. In the calculations with the number of particles of 10 or more, the average and the longest computational time are plotted for the same number of particles because of the existence of different reactions. The longest calculation (#13) in the test data took about one hour (3718 seconds). The practical limitation of our approach is 10 particles at this time, although most previous reaction prediction studies can be done in a few seconds or less. For reactions with more than 15 particles, the computation time for a single reaction may exceed several hours, because the number of possible reactive points increases too much. All calculations were performed on Xeon Silver 4214 processor and no GPU.

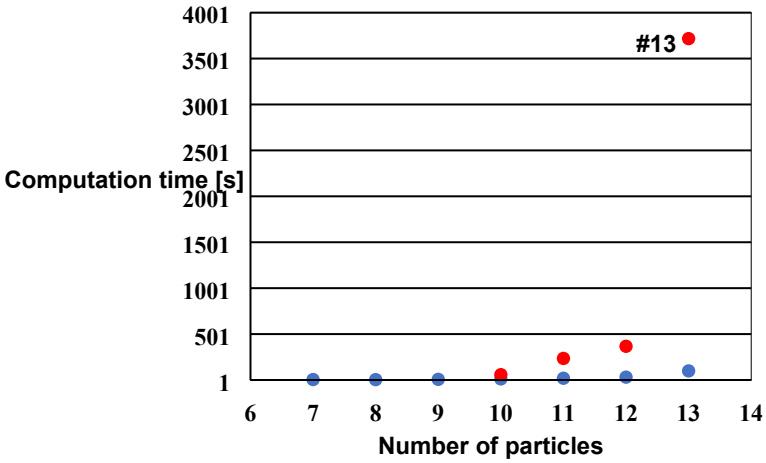


Figure S1. Relationship between the number of particles and the computation time of the prediction in test reactions. The red plots show the longest calculation time among the same number of particles.

Second, although resonance structures exist in the reaction network, the proposed model cannot evaluate resonance; e.g., when a halogen is added to butadiene (Figure 2(B)), the learning model could have predicted the 1,4-product if the stability of the resonance had been considered. This may be accounted for by only analyzing the shortcut from the reactant to the product; thus, the reciprocating resonance structures and shortcut branching are ignored. The third problem is related to the second, which prevents the treatment of catalysts using the model. Reaction pathways involving catalysts are always present in the network, and if a search for detour pathways is permitted, the number of candidate pathways becomes excessive. As observed with the test dataset, even with only 35 reactions, the number of candidate pathways is ~4 million. Therefore, considering detour pathways at the current stage is challenging. Despite these problems, if a network can be constructed to resolve these issues, or if they can be overcome using novel descriptors, the proposed learning model should be a strong scientific predictor of organic reaction mechanisms.

3.Table Contents

Table S1. Numbers of valence electrons, maximum and minimum formal charges, maximum numbers of shared electrons, and permitted electron-deficient states of the employed atoms and quasiparticles. The octet (duet) rule states that the maximum number of shared electrons should not be exceeded.

Atom	Valence	Max. Charge	Min. Charge	Max. Electron	Electron-Deficient
Ph	3	+1	0	4	4
X	7	0	−1	8	8
M	1	+1	0	2	0
MgX	1	+1	0	2	0
O	6	+1	−1	8	8
N	5	+1	−1	8	6
CH_3	1	+1	0	2	0
CH_2	2	+1	0	4	2
C	4	+1	−1	8	6
H	1	+1	−1	2	0

Table S2. Forty-six selected fragment structures used as descriptors in machine learning, wherein "+" and "−" represent positive and negative formal charges and "≡," "=," and "–" indicate triple, double, and single bonds, respectively.

One-particle structures			
H^+	H^-	Ph^+	M^+
O^+	O^-	N^+	N^-
C^-	X^-	=C=	=N=
Two-particle structures			
C≡C	C=C	O≡C	O=C
O≡N	O=N	N≡C^-	N=C
Ph=C	Ph=N	Ph–C	Ph–N
X–C	M–C	O–C	N–C
Ph–M	Ph–H	M–N	N–H
X–N	O–N	M–O	O–H
X–M	X–H	M–H	H–H
O^-–C^+			
Other structures			
C^+H_3	C^+H_2R	C^+HR_2	C^+R_3
C^-– (CO)			

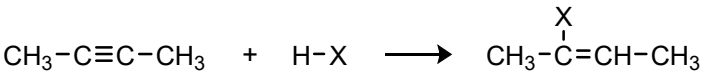
Table S3. Points of the fragment structures as descriptors. The group numbers of the structures are also shown.

Positive points			Negative points		
Fragment structure	Points	Group	Fragment structure	Points	Group
X–M	4.32	3	M–C	−5.50	3
H^+	3.64	3	C^+H_3	−5.27	3
$N{\equiv}C^-$	3.22	2	X–N	−4.76	3
N^+	2.84	2	H_2	−4.40	3
M^+	2.51	2	C≡C	−4.22	3
O–H	2.47	2	Ph^+	−3.95	3
C^+R_3	2.29	2	Ph–H	−3.66	3
C^+HR_2	2.18	2	C^-	−3.36	3
N–H	1.88	2	C=C	−3.30	3
O^+	1.87	2	H^-	−3.19	2
C^-–(CO)	1.84	2	M–N	−3.14	2
X–H	1.61	1	C^+H_2R	−2.99	2
O=C	1.55	1	M–H	−2.84	2
O=N	1.51	1	O–N	−2.63	2
O^-	0.14	1	X–C	−2.28	2
N^-	0.13	1	N=C	−2.10	2
=N=	0.11	1	M–O	−2.03	2
Ph=N	0.10	1	=C=	−1.41	1
O≡N	0.04	1	Ph–C	−1.25	1
N–C	0.01	1	O–C	−0.97	1
Ph=C	0.00	1	O^-–C^+	−0.44	1
O≡C	0.00	1	Ph–M	−0.42	1
			Ph–N	−0.17	1
			X^-	−0.14	1

4.Figure Contents

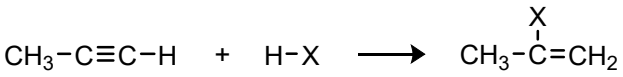
Figure S2. Learning reactions with their prediction accuracies used as the 50 training datasets (#1–#50). The "top" number represents the predicted ranking of the correct pathway of each reaction.

#1



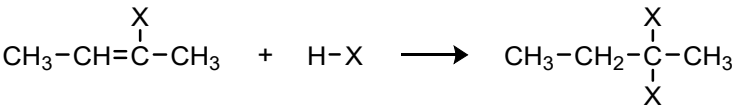
Top 1, 2, 3

#2



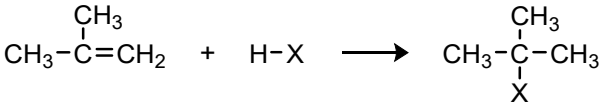
Top 1, 2, 3, 4

#3



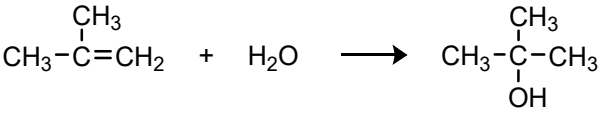
Top 5

#4



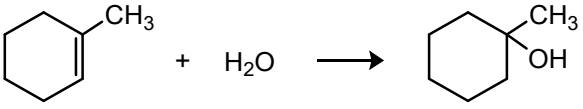
Top 1, 2

#5



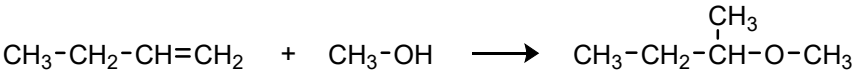
Top 1, 2, 3, 4, 5

#6



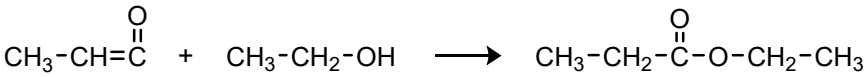
Top 1, 3

#7



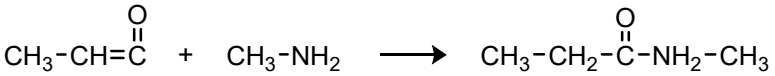
Top 5

#8



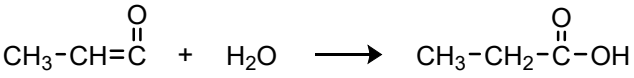
Top 2, 5

#9



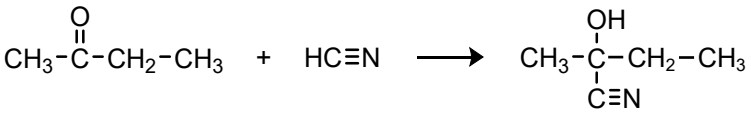
Top 1, 2, 3, 5

#10



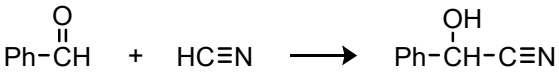
Top 1, 2, 3, 4

#11



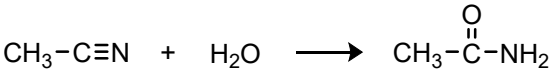
Top 1, 2, 3, 4

#12



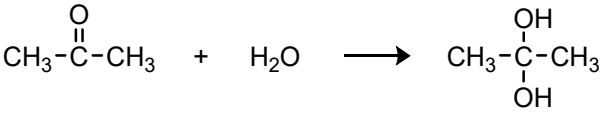
Top 1, 2, 3, 4, 5

#13



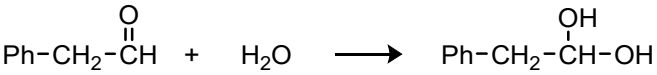
Top 1, 2, 3, 4, 5

#14



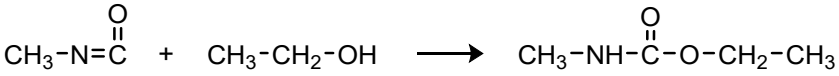
Top 3, 4

#15



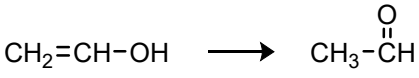
Top 1, 2, 3, 4, 5

#16



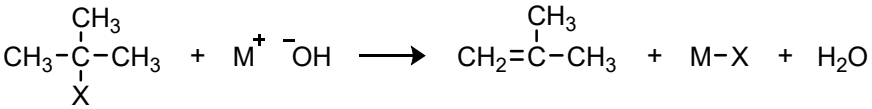
Top 1, 3, 4, 5

#17



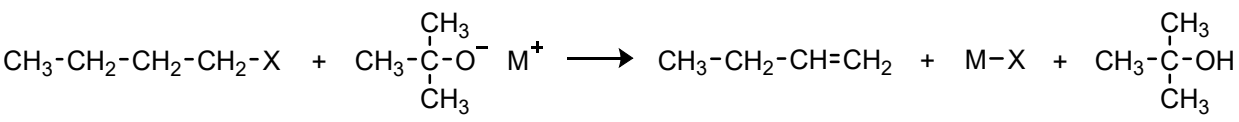
Top 1, 2, 3, 4, 5

#18



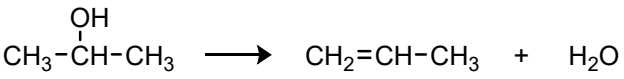
Top 1, 2

#19



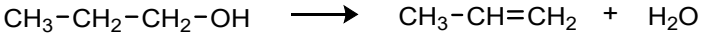
Top 2, 5

#20



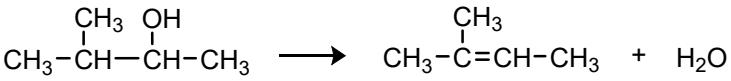
Top 1, 2, 3, 4, 5

#21



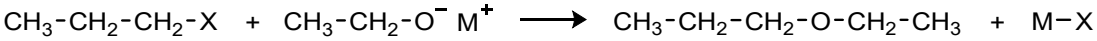
Top 1, 2

#22



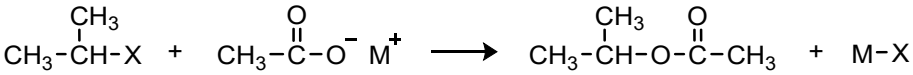
Top 1, 3, 4, 5

#23



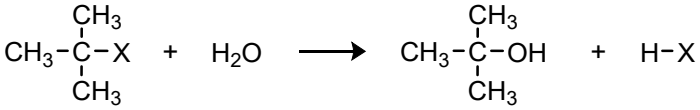
Top 3, 4

#24



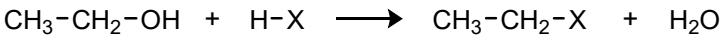
Top 1, 4

#25



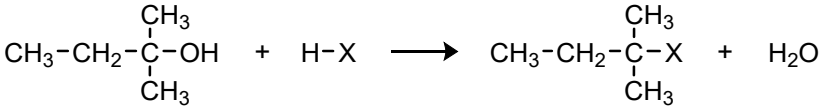
Top 1, 2, 4

#26



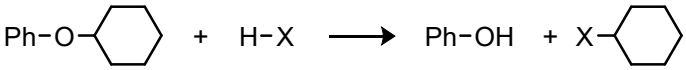
Top 1, 2, 3, 4, 5

#27



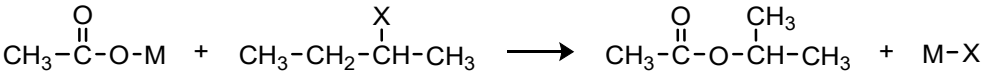
Top 1, 2, 3, 4, 5

#28



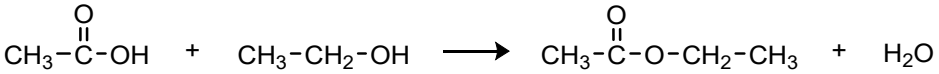
Top 1, 2, 3, 4, 5

#29



Top 1, 2, 3, 4, 5

#30

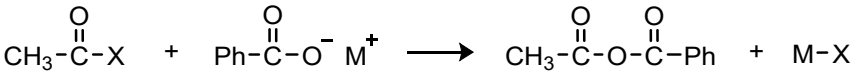


Top 1, 2, 3, 4, 5

#31

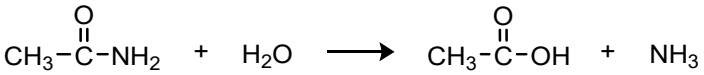
Top 1, 2, 3, 4, 5

#32



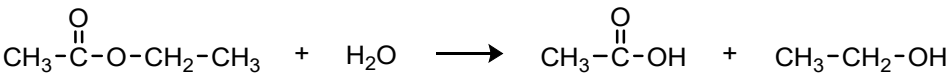
Top 1

#33



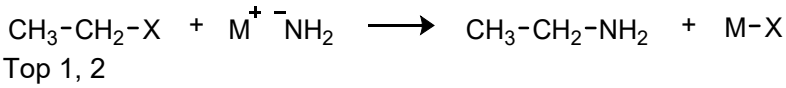
Top 1, 2, 3, 4, 5

#34



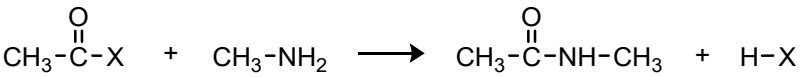
Top 1, 2, 3, 4, 5

#35



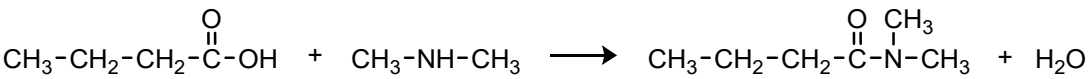
Top 1, 2

#36



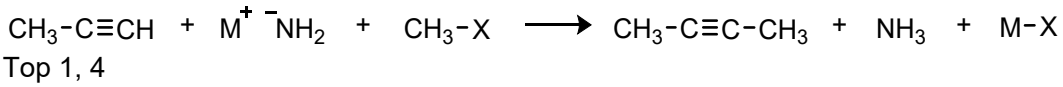
Top 1, 2, 3, 4, 5

#37



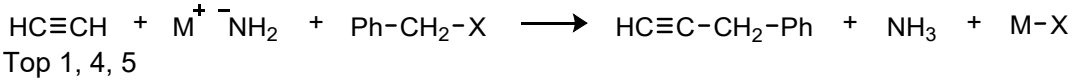
Top 1, 2, 3, 4, 5

#38



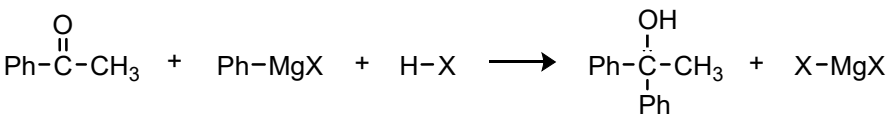
Top 1, 4

#39



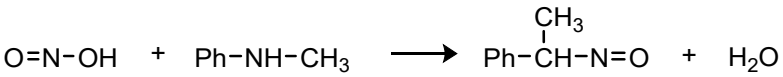
Top 1, 4, 5

#40



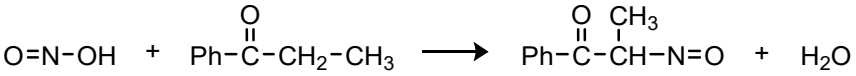
Top 1, 3, 4

#41



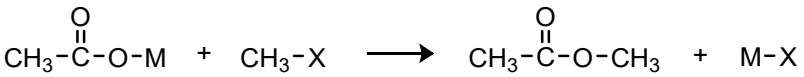
Top 1, 2, 3, 4, 5

#42



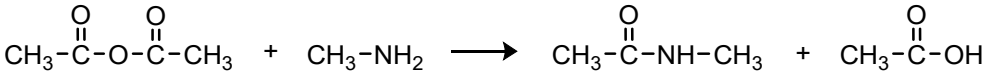
Top 1, 2, 3, 4

#43



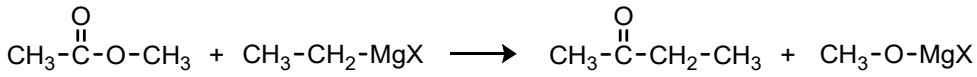
Top 1, 2, 4, 5

#44



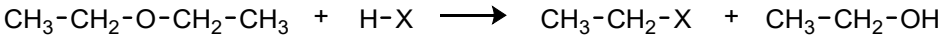
Top 1, 2, 3, 4, 5

#45



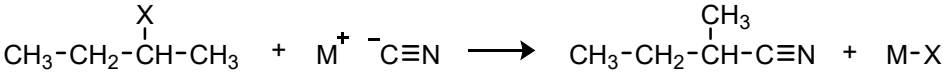
Top 1, 2, 3, 4

#46



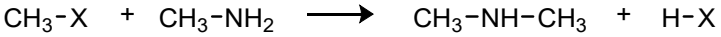
Top 1, 2, 5

#47



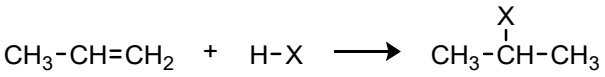
Top 1, 2

#48



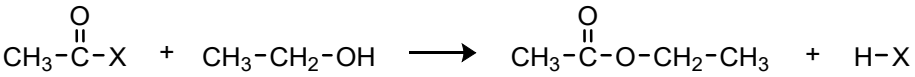
Top 1, 2, 3, 4

#49



Top 1, 2, 3, 4

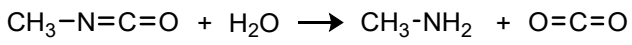
#50



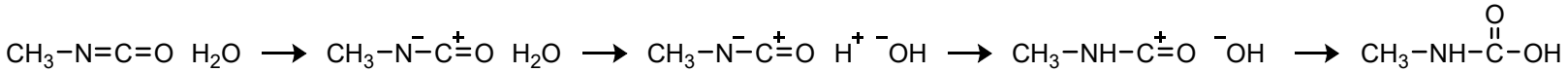
Top 1, 2, 3, 5

Figure S3. Top 5 predicted reaction pathways of the 35 test datasets (#1–#35). In each example, “All graphs” represents the number of nodes in the reaction network, and “Total candidates” is the number of stable structures at the nodes. “All paths” represents the number of shortcuts from the reactant node to the candidate nodes, and “Correct paths” is the number of shortcuts that lead to the correct product.

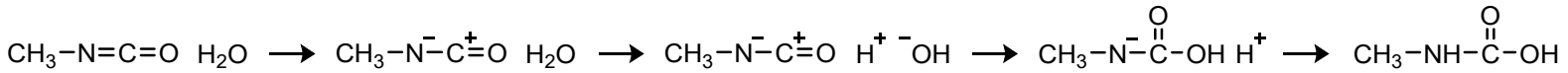
#1



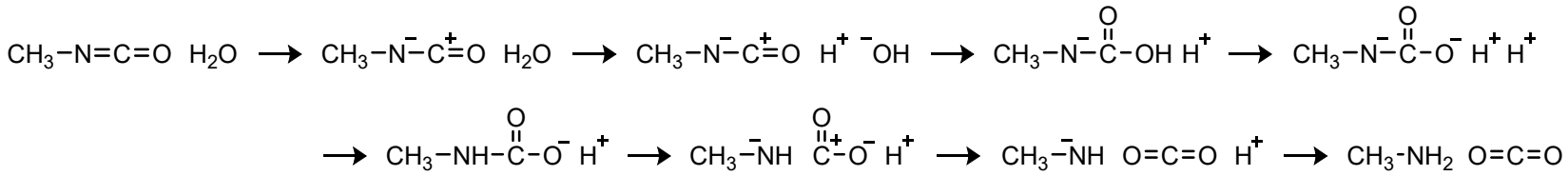
Top 1



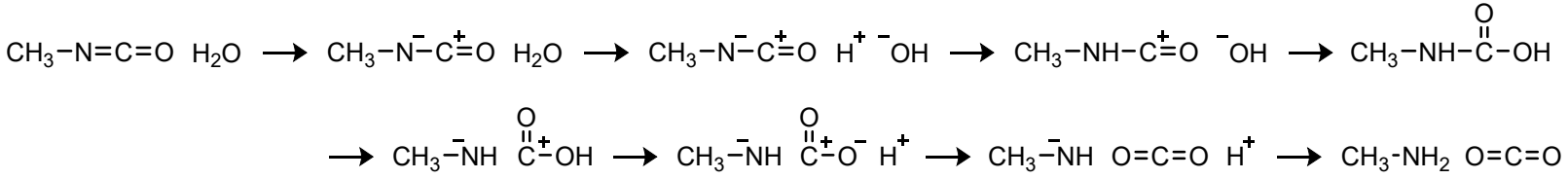
Top 2



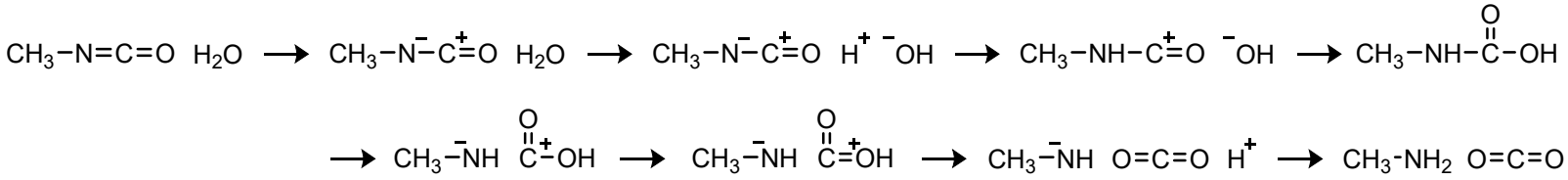
Top 3



Top 4

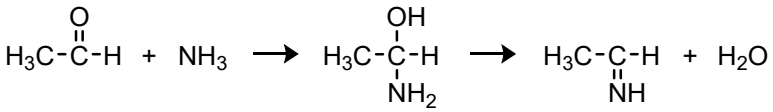


Top 5

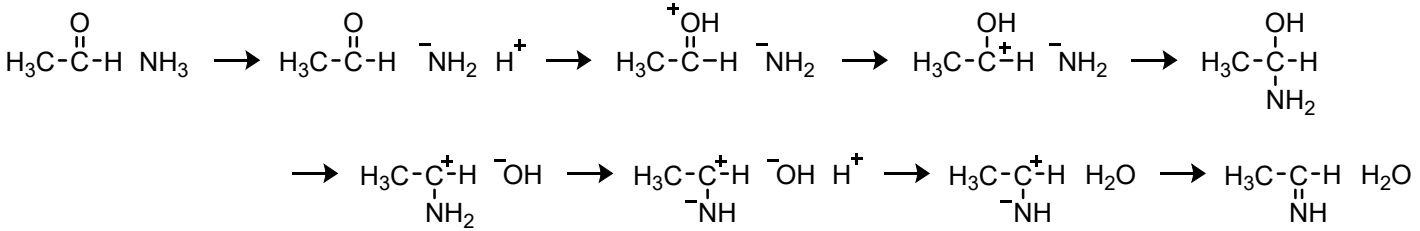


All graphs 2889 / Total candidates 40
All paths 1 477 555 / Correct paths 45

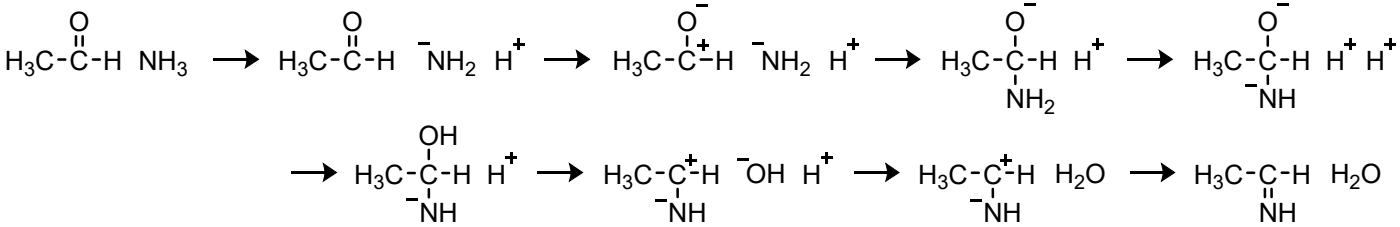
#2



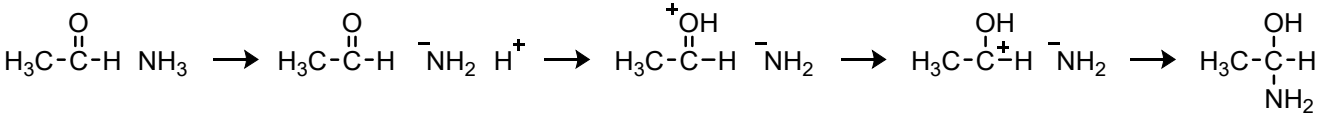
Top 1



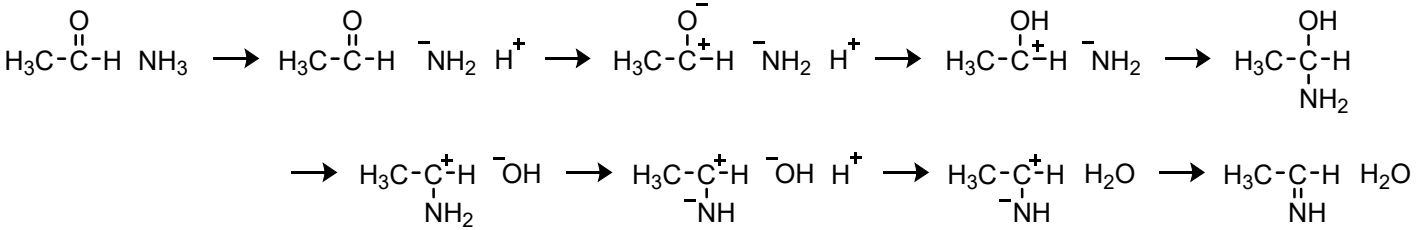
Top 2



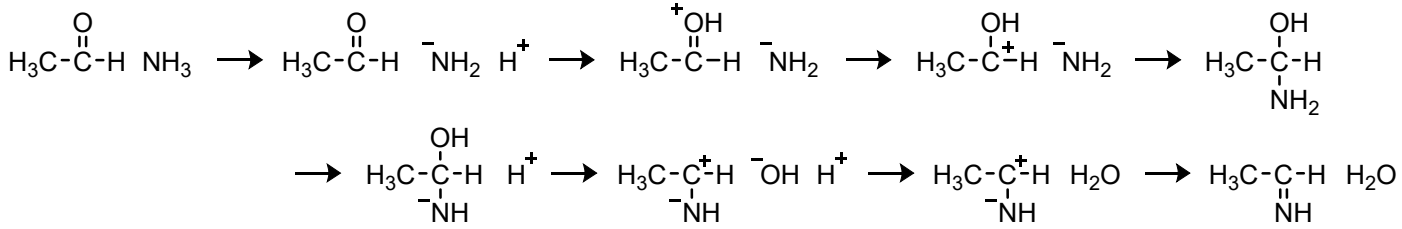
Top 3



Top 4

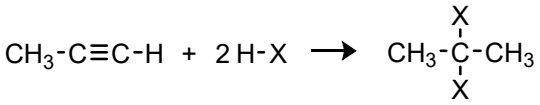


Top 5

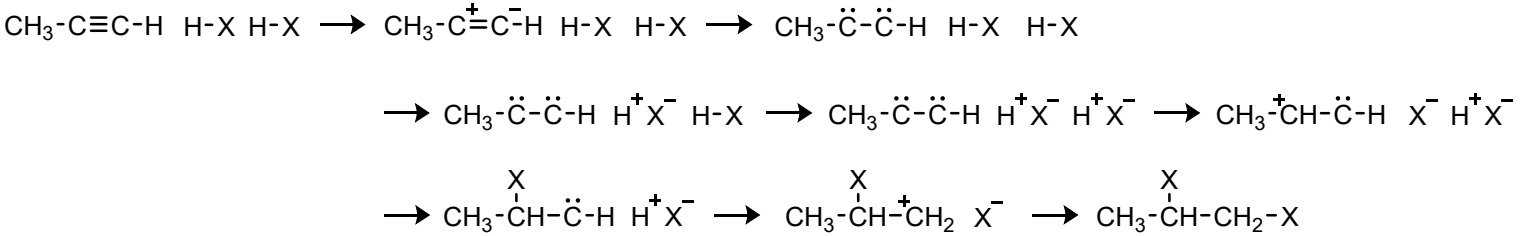


All graphs 2429 / Total candidates 38
All paths 1 734 834 / Correct paths 45

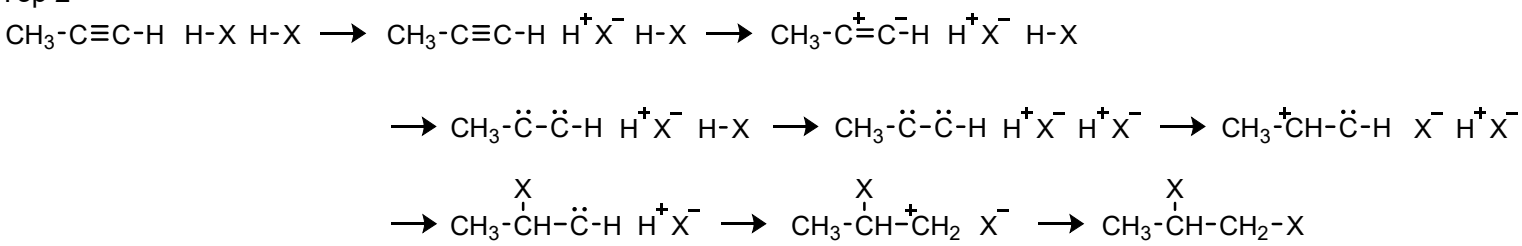
#3



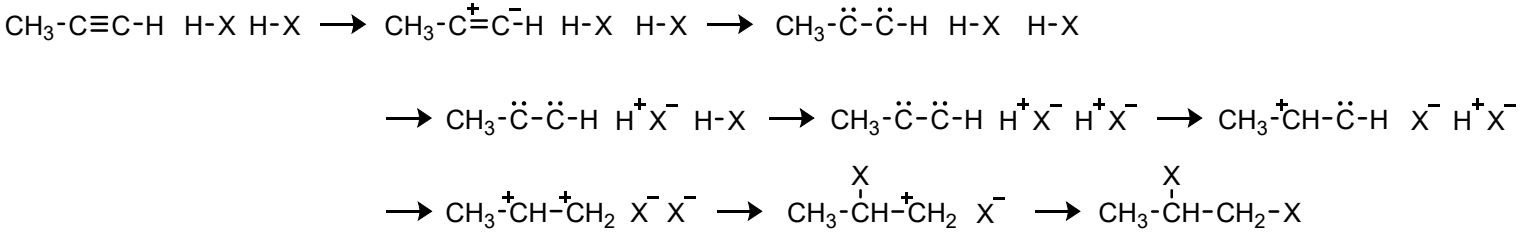
Top 1



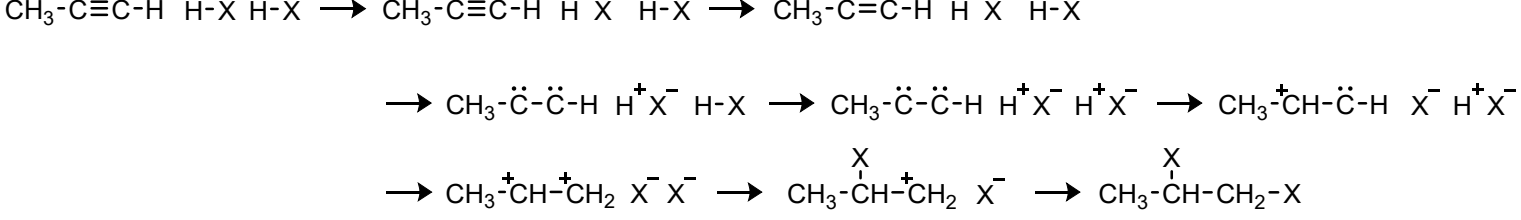
Top 2



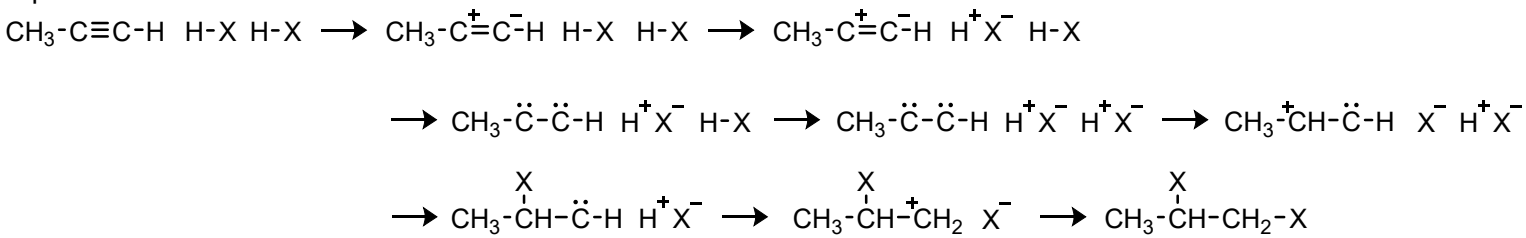
Top 3



Top 4

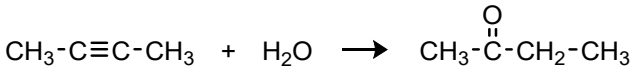


Top 5

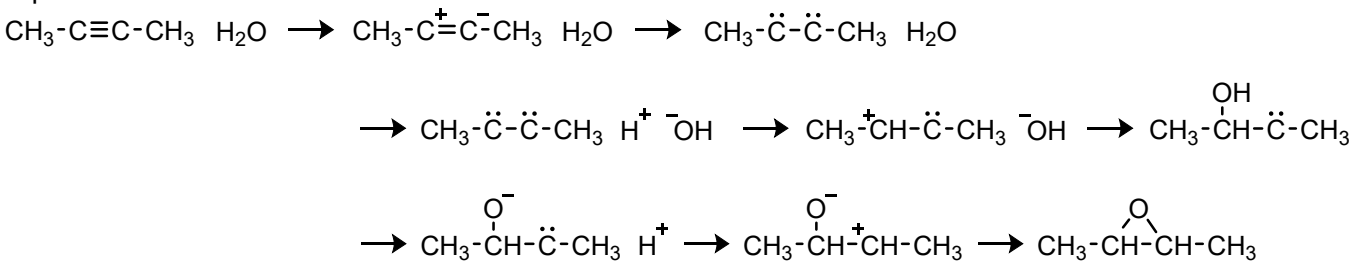


All graphs 531 / Total candidates 15
All paths 5728 / Correct paths 24

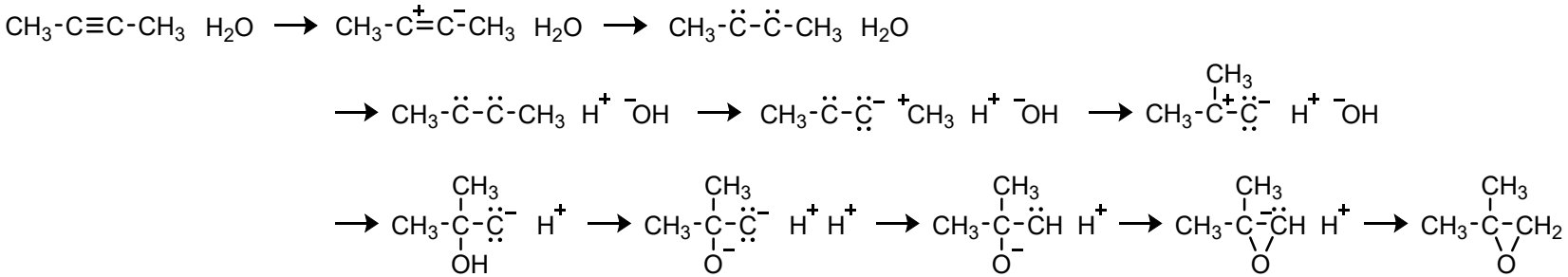
#4



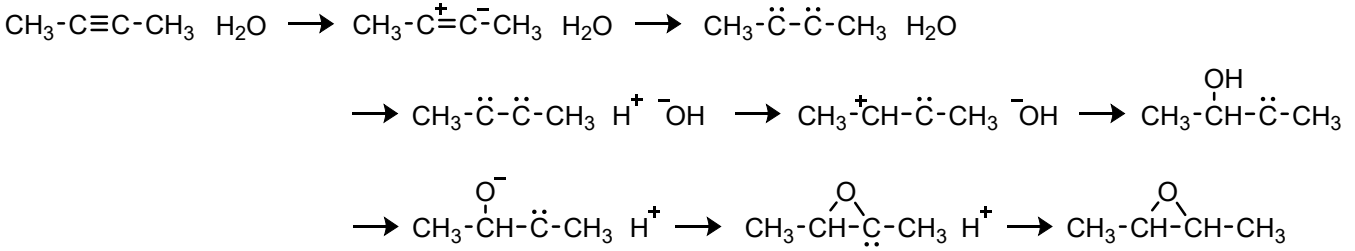
Top 1



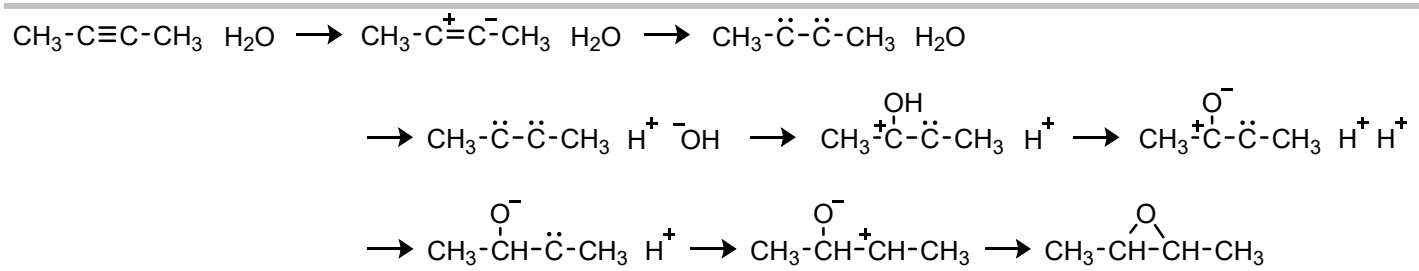
Top 2



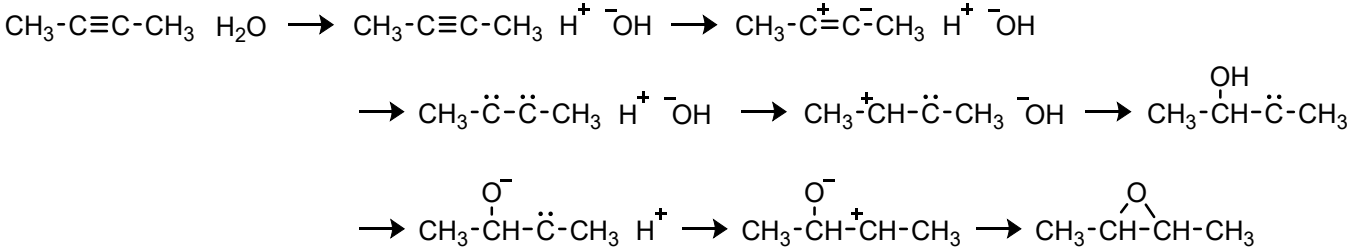
Top 3



Top 4

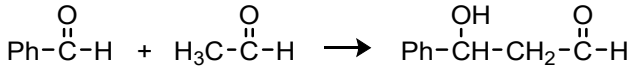


Top 5

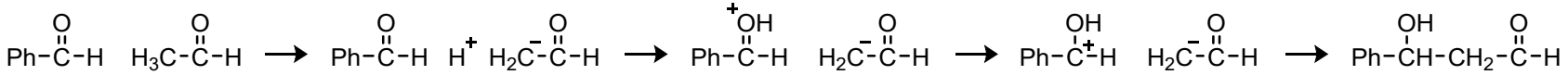


All graphs 858 / Total candidates 18
All paths 8943 / Correct paths 31

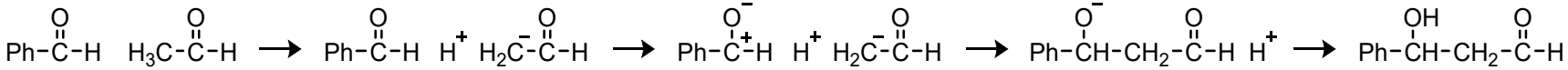
#5



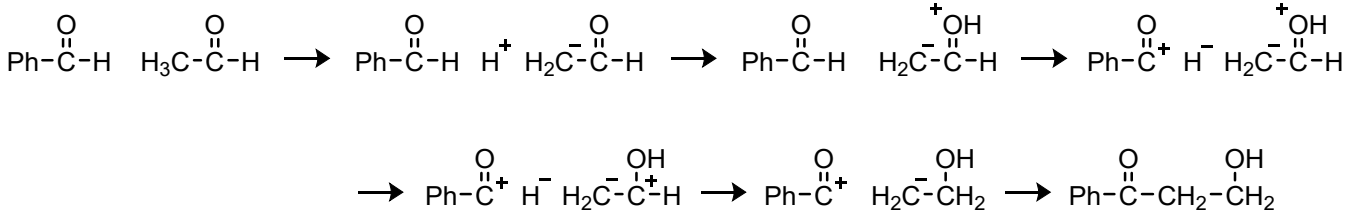
Top 1



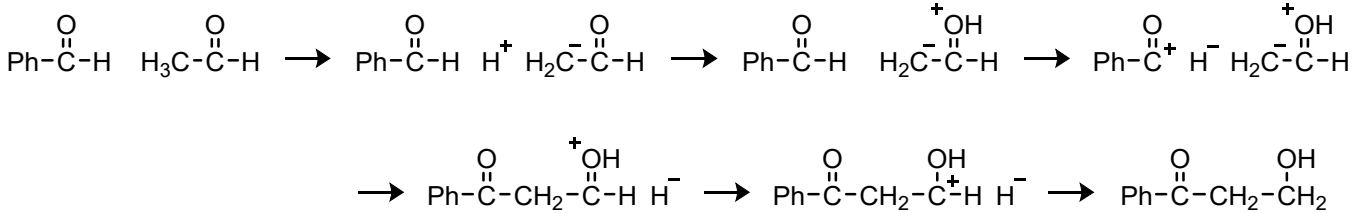
Top 2



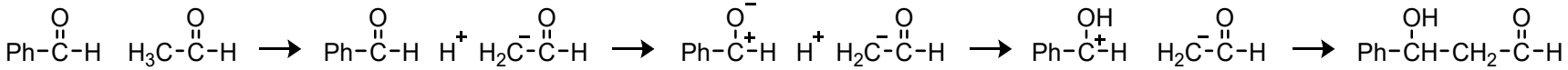
Top 3



Top 4

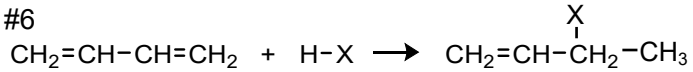


Top 5

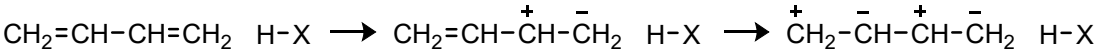


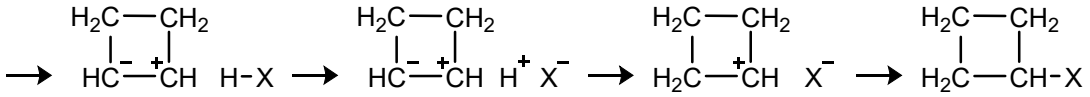
All graphs 6394 / Total candidates 45
All paths 1415 / Correct paths 5

#6

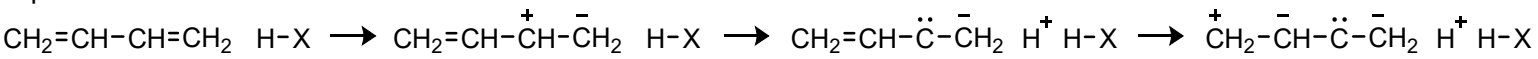


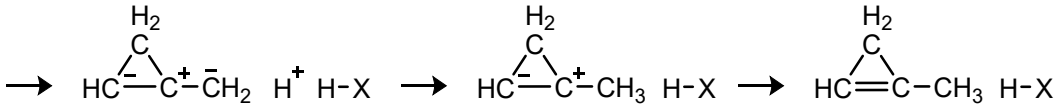
Top 1



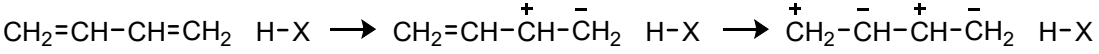


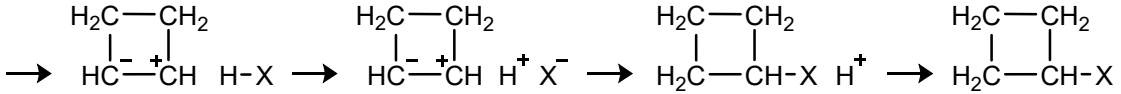
Top 2



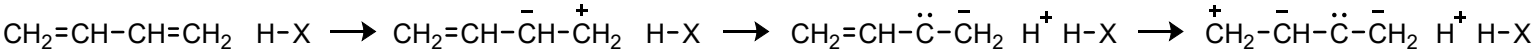


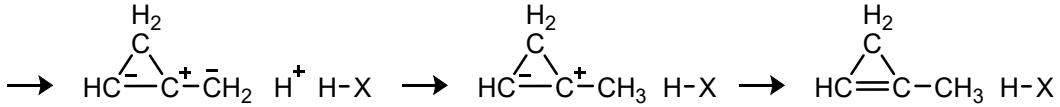
Top 3



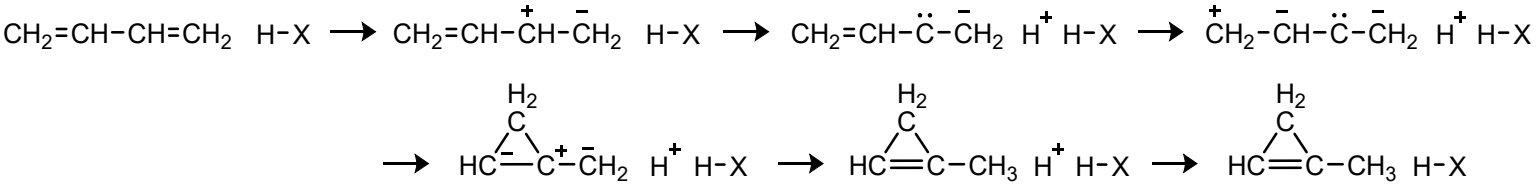


Top 4



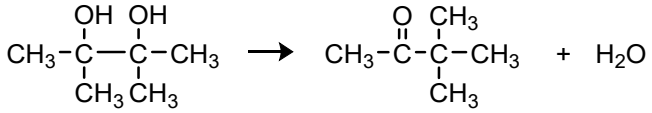


Top 5

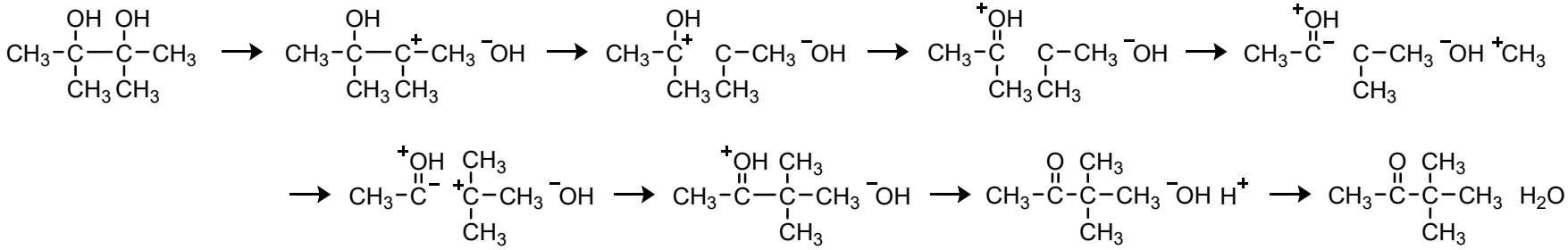


All graphs 4642 / Total candidates 38
All paths 649 / Correct paths 57

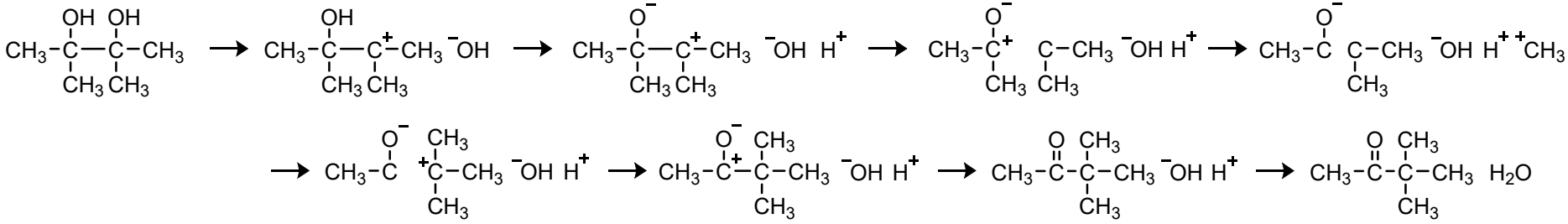
#7



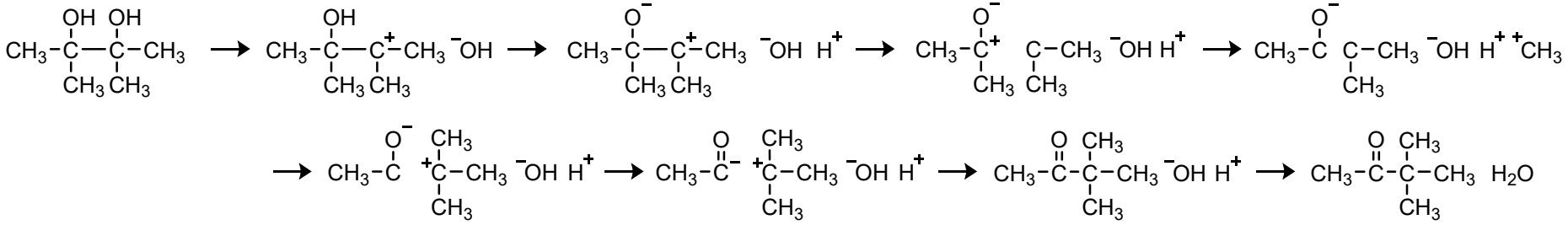
Top 1



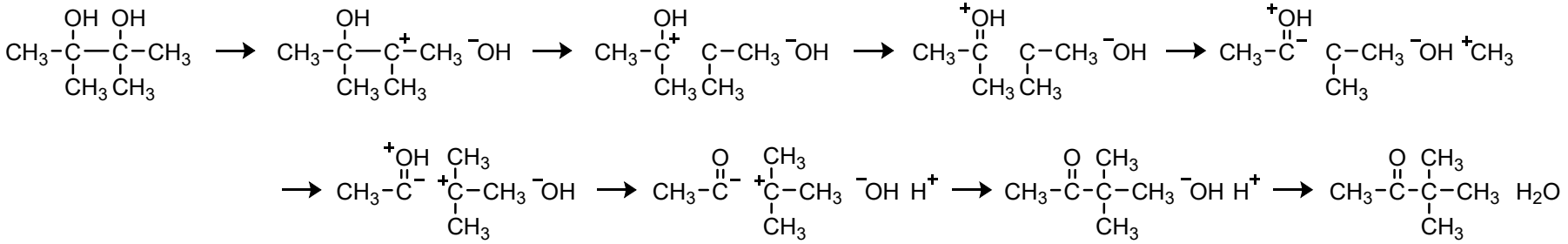
Top 2



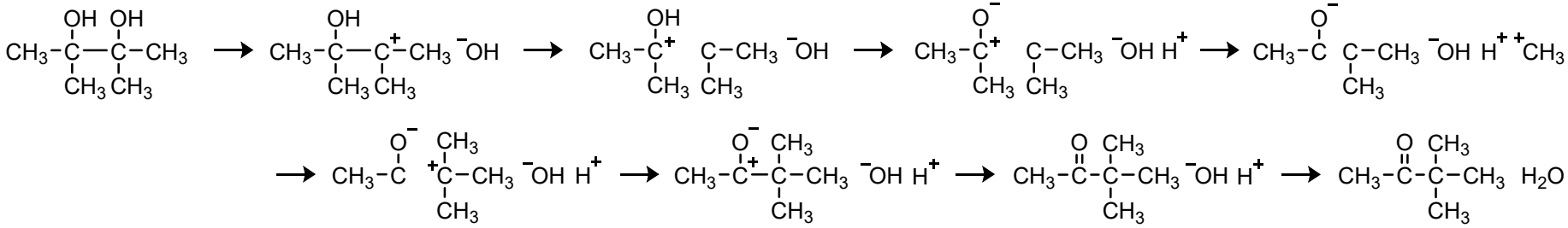
Top 3



Top 4

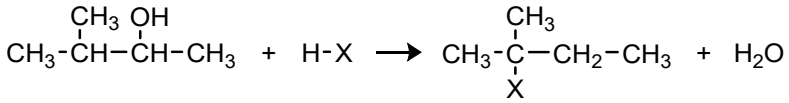


Top 5

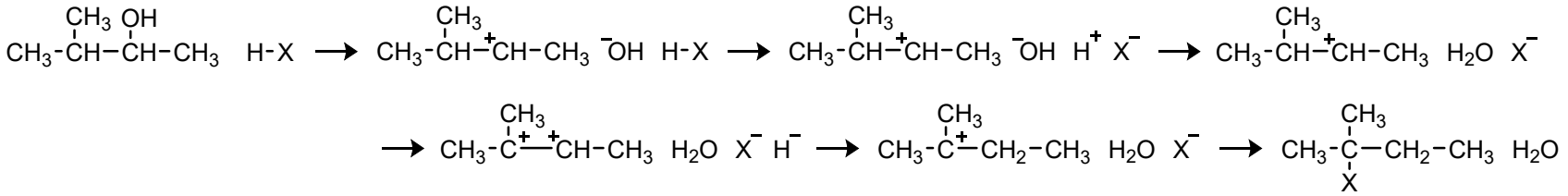


All graphs 3378 / Total candidates 52
All paths 94 395 / Correct paths 1,300

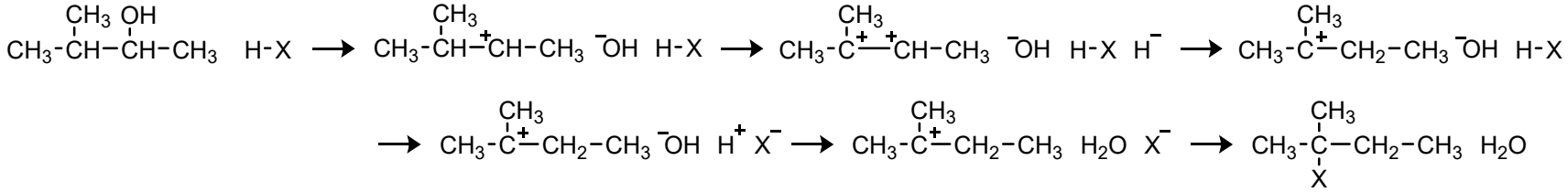
#8



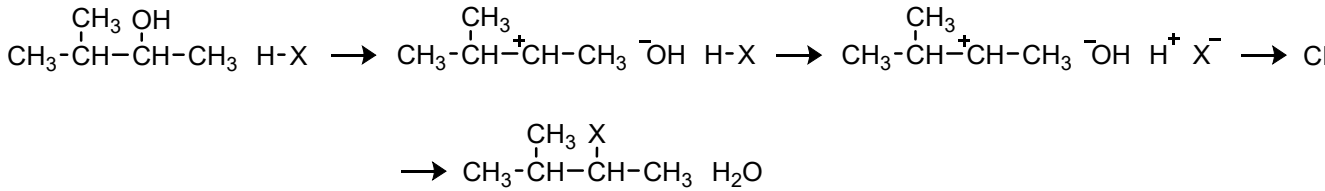
Top 1



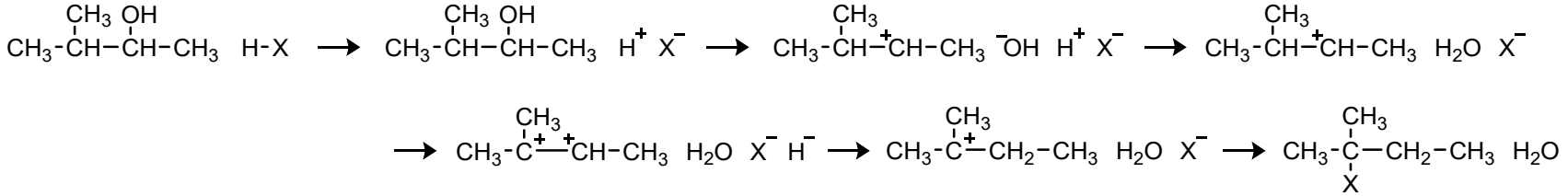
Top 2



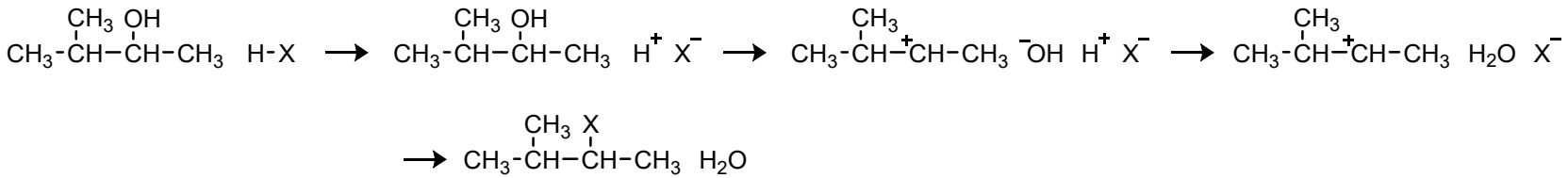
Top 3



Top 4

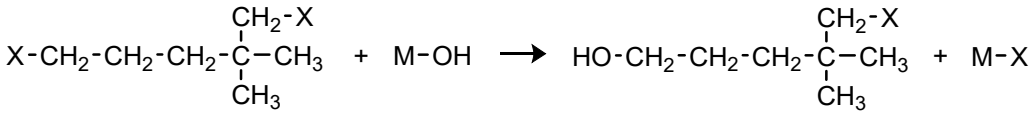


Top 5

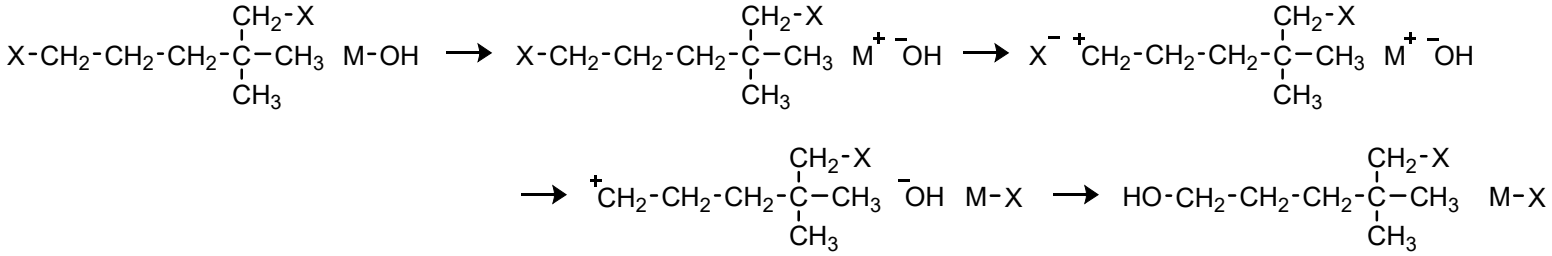


All graphs 4120 / Total candidates 41
All paths 965 / Correct paths 25

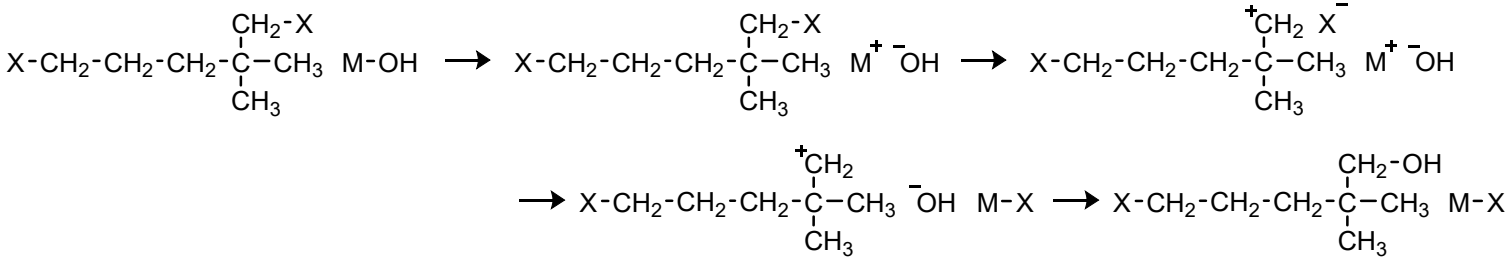
#9



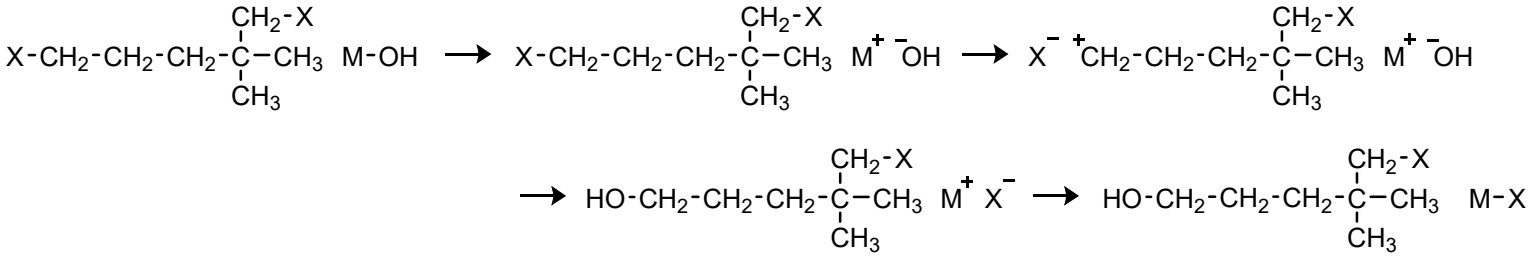
Top 1



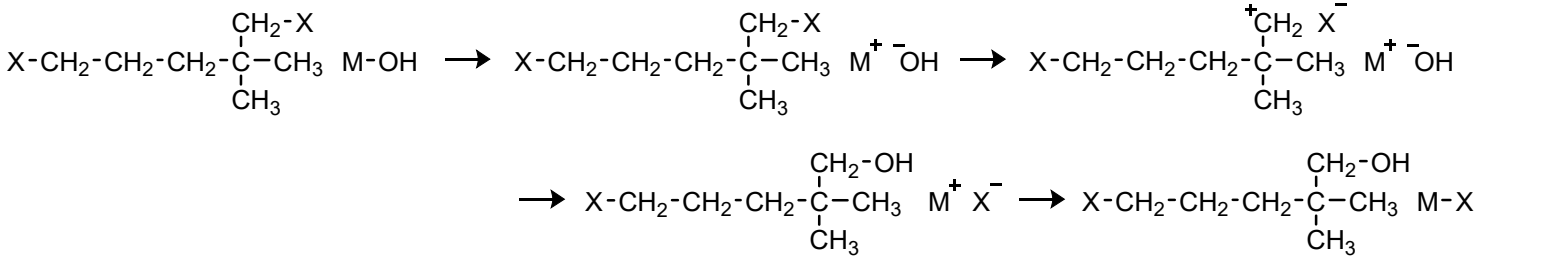
Top 2



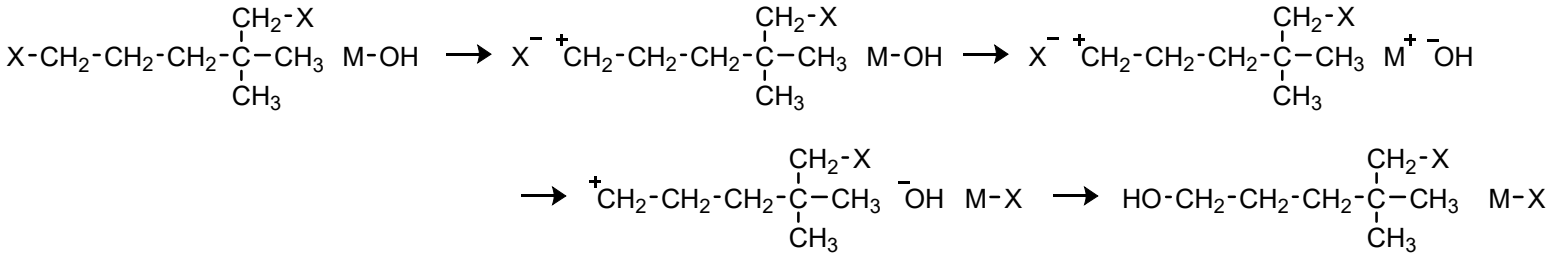
Top 3



Top 4

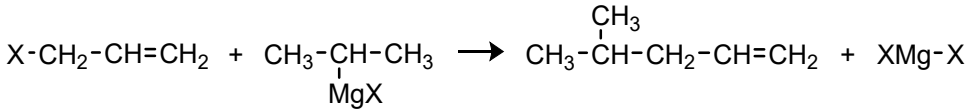


Top 5

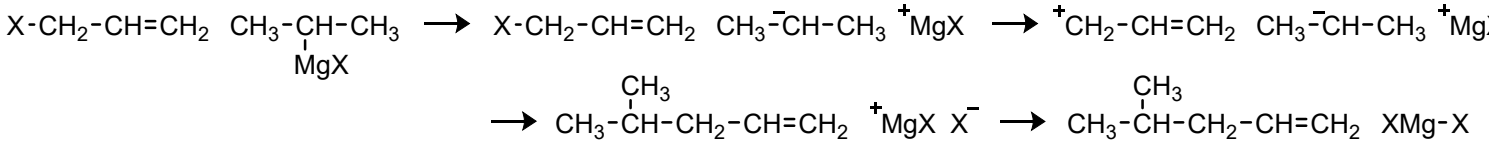


All graphs 3146 / Total candidates 34
All paths 165 / Correct paths 5

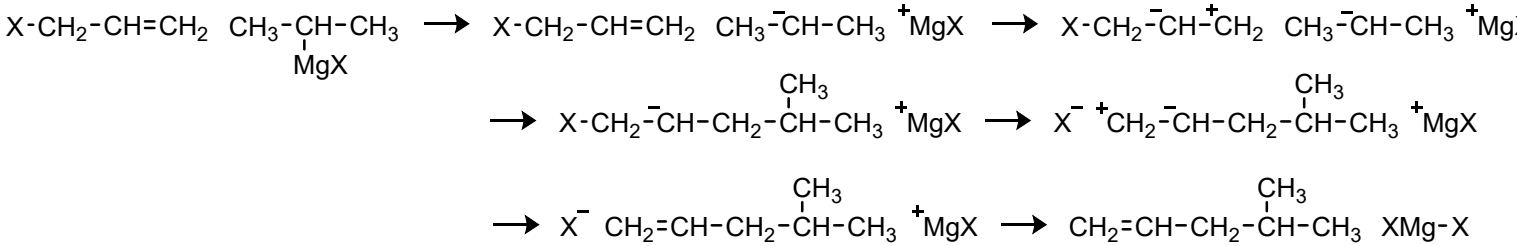
#10



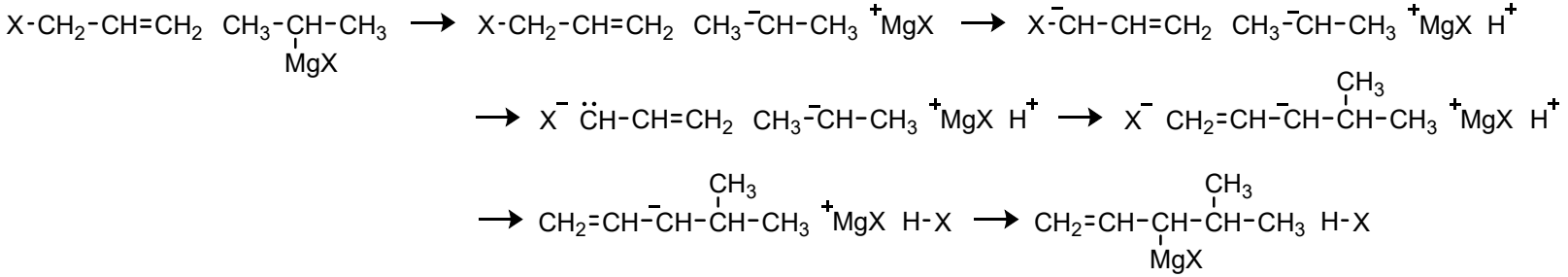
Top 1



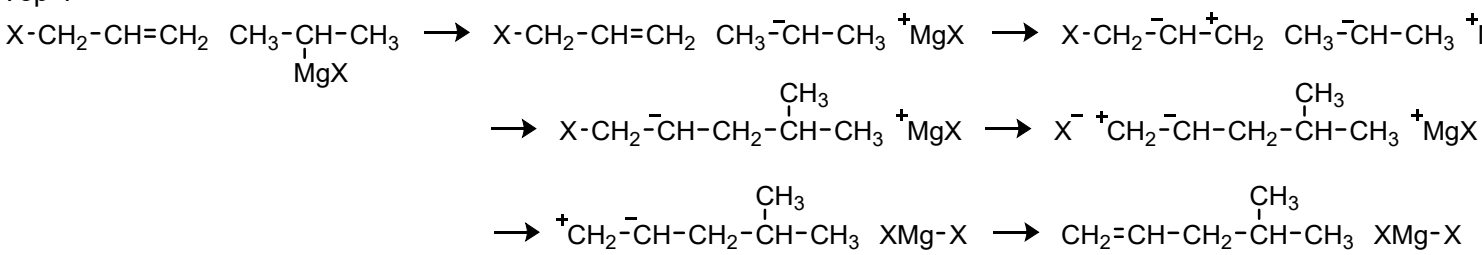
Top 2



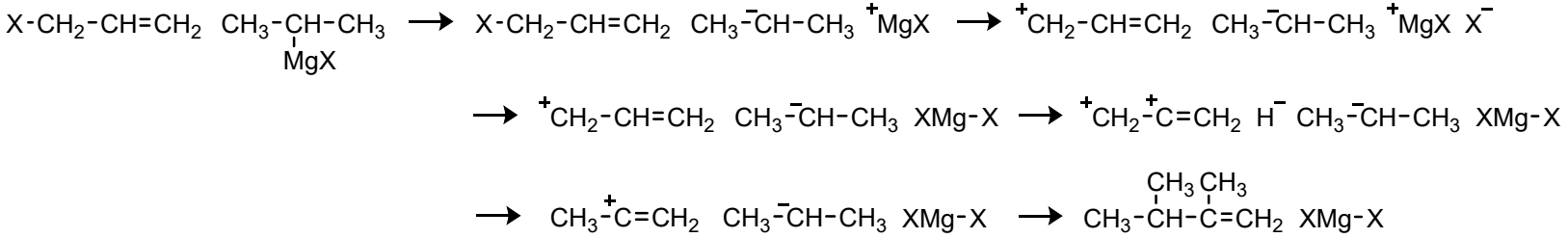
Top 3



Top 4

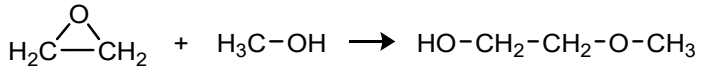


Top 5

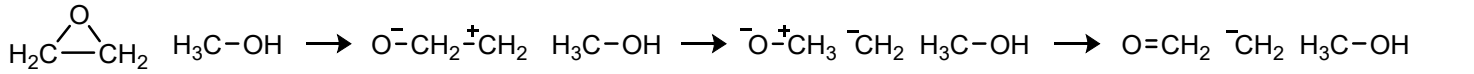


All graphs 6256 / Total candidates 95
All paths 2230 / Correct paths 4

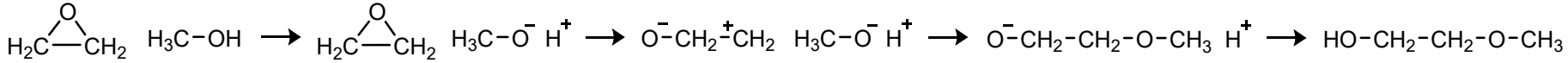
#11



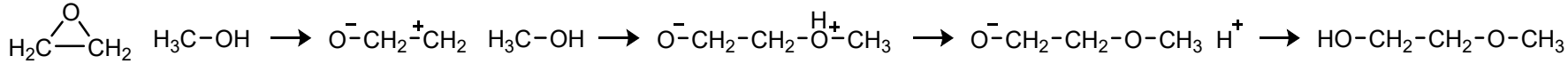
Top 1



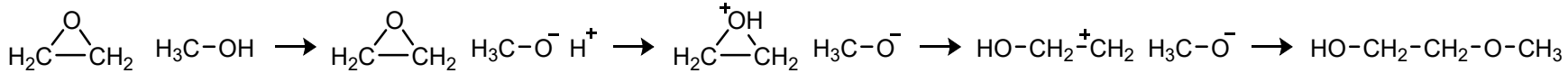
Top 2



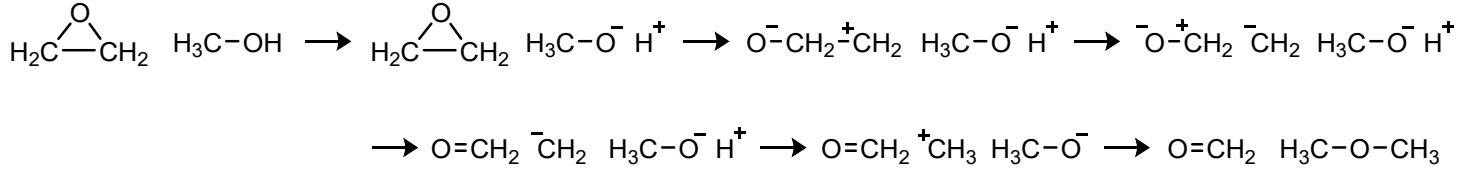
Top 3



Top 4

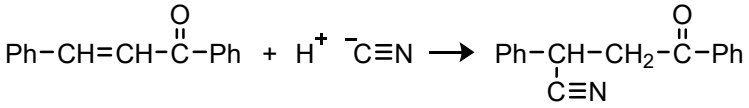


Top 5

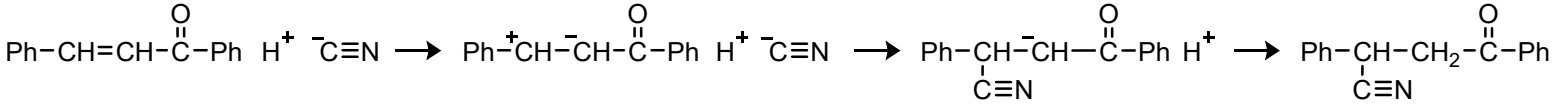


All graphs 82 / Total candidates 7
All paths 289 / Correct paths 12

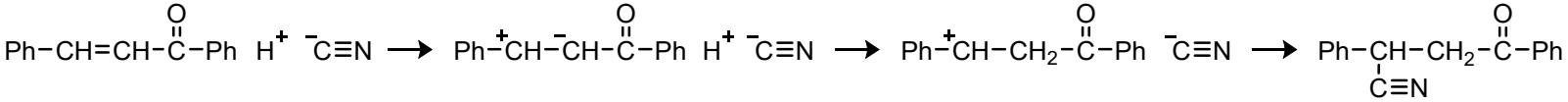
#12



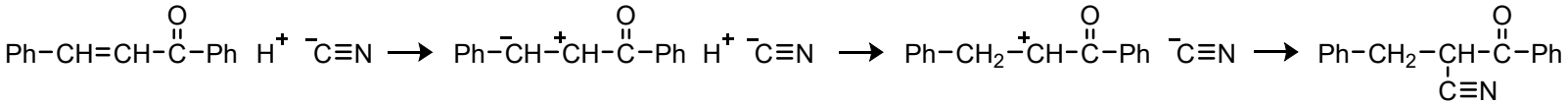
Top 1



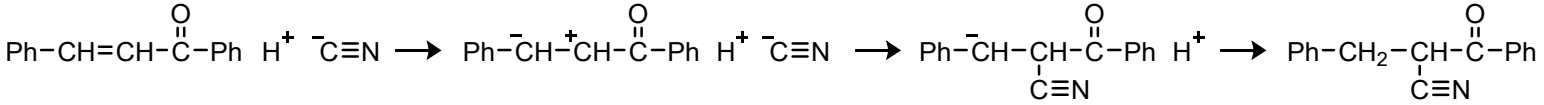
Top 2



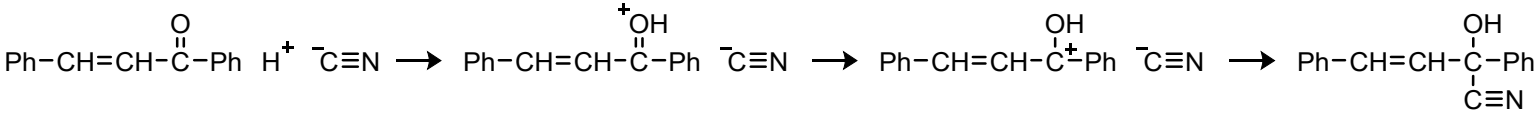
Top 3



Top 4

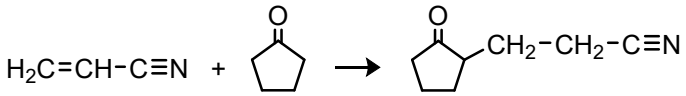


Top 5

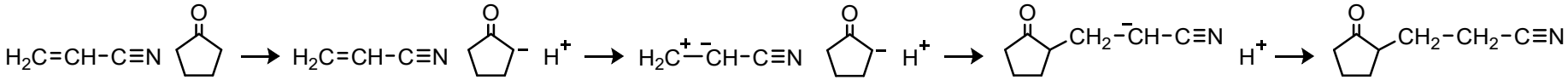


All graphs 3216 / Total candidates 10
All paths 18 / Correct paths 2

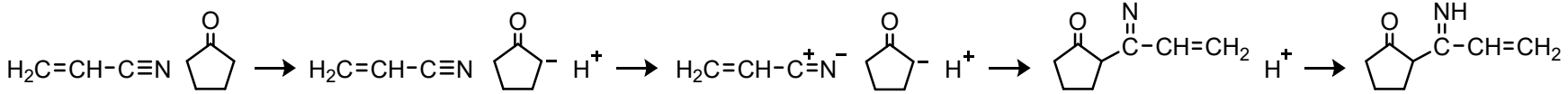
#13



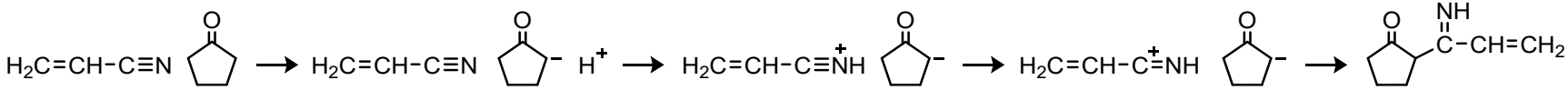
Top 1



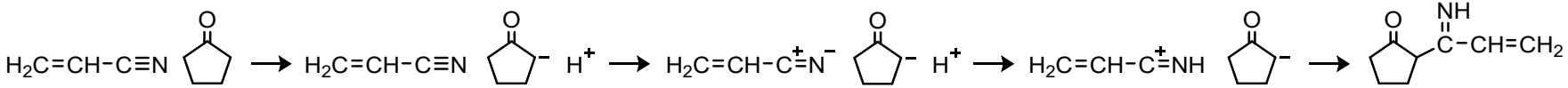
Top 2



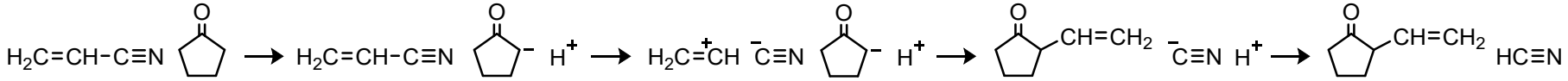
Top 3



Top 4

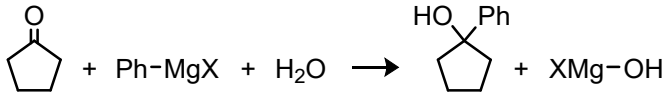


Top 5

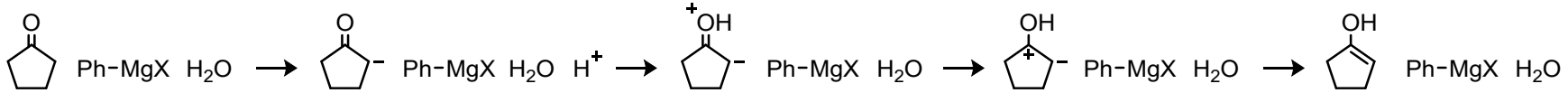


All graphs 12 375 / Total candidates 45
All paths 252 / Correct paths 4

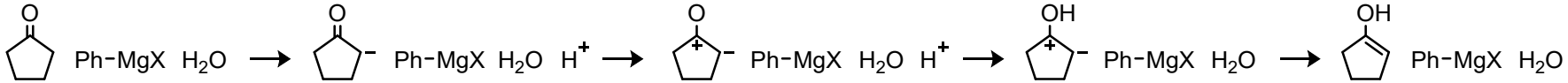
#14



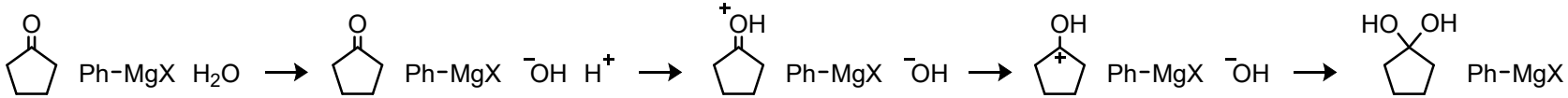
Top 1



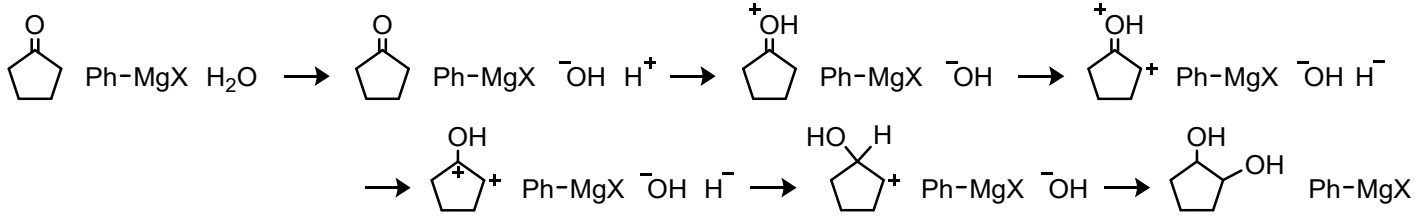
Top 2



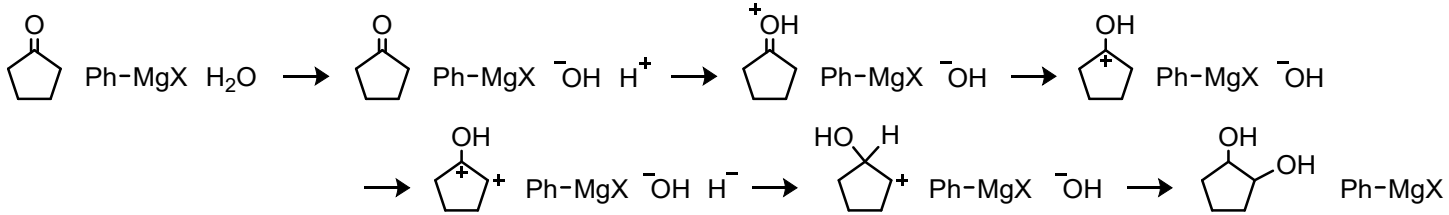
Top 3



Top 4

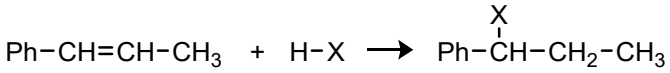


Top 5

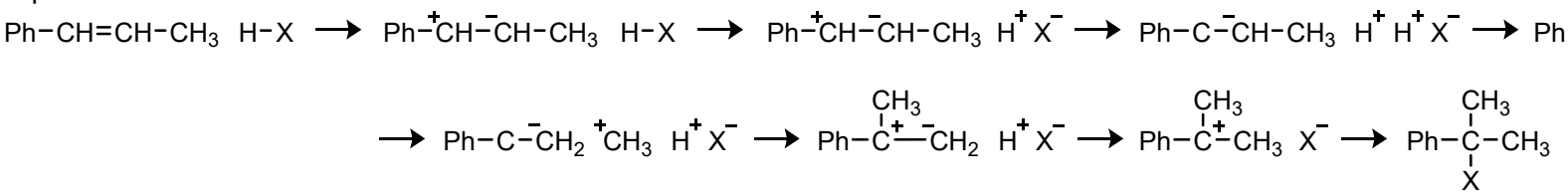


All graphs 7349 / Total candidates 80
All paths 1550 / Correct paths 15

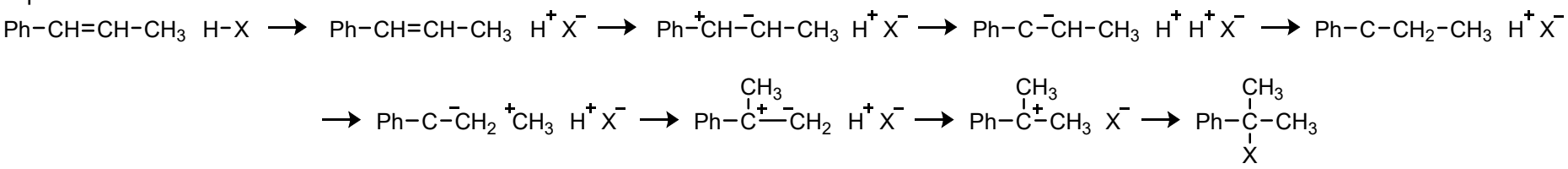
#15



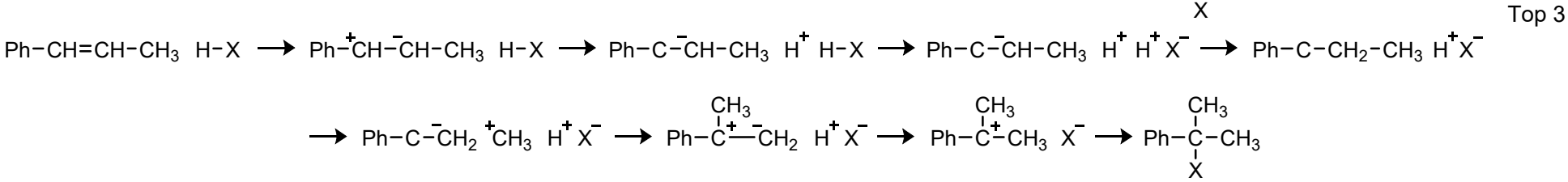
Top 1



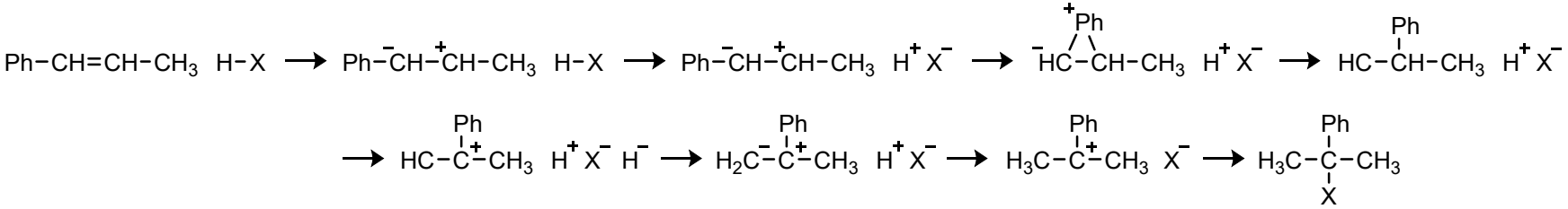
Top 2



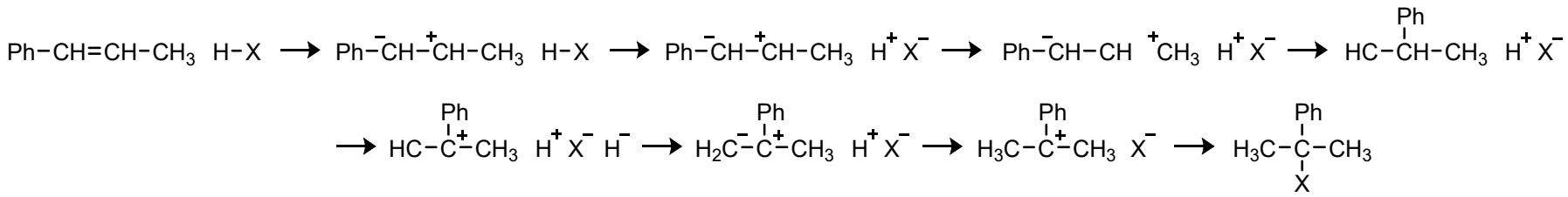
Top 3



Top 4

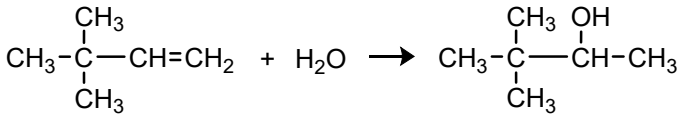


Top 5

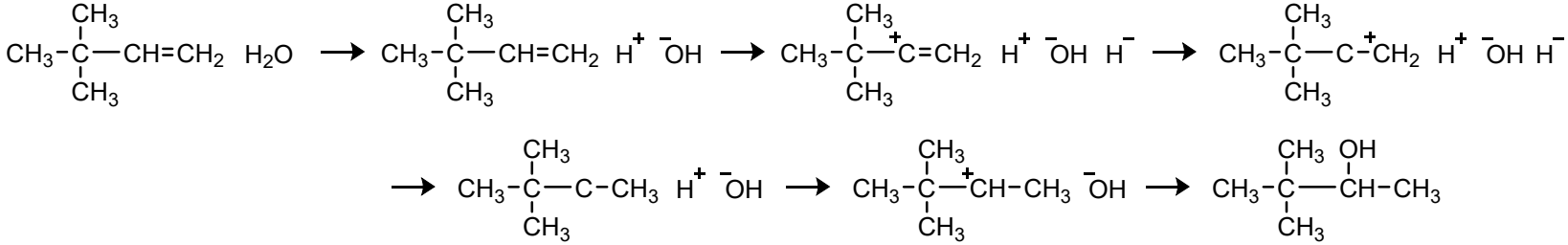


All graphs 1624 / Total candidates 24
All paths 2804 / Correct paths 4

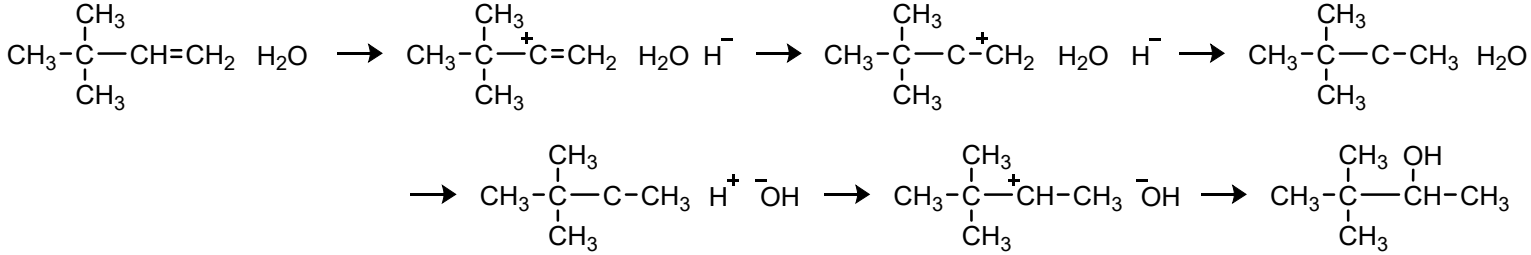
#16



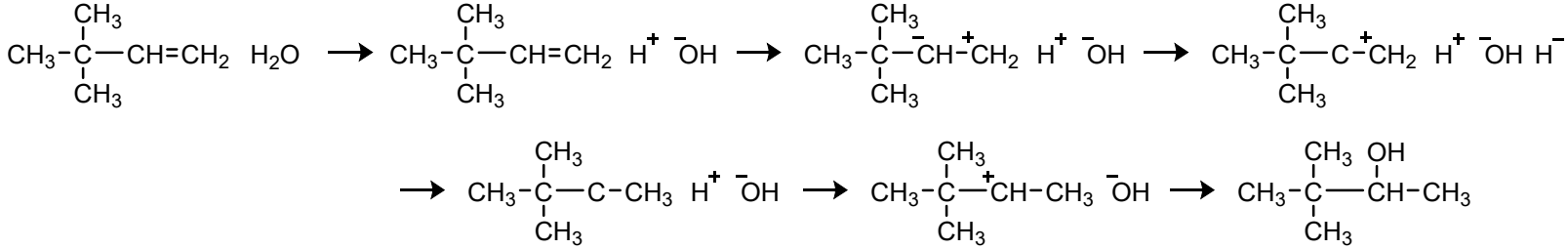
Top 1



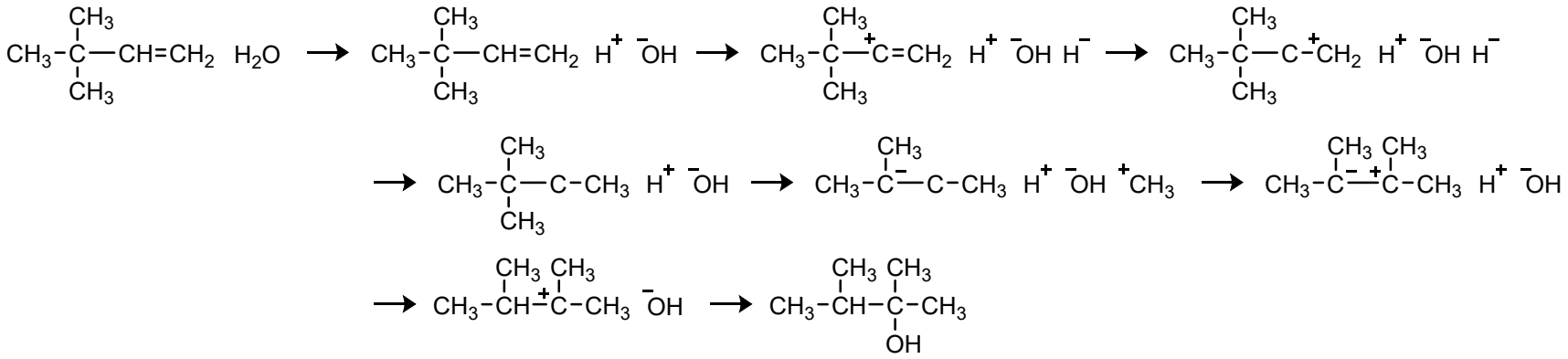
Top 2



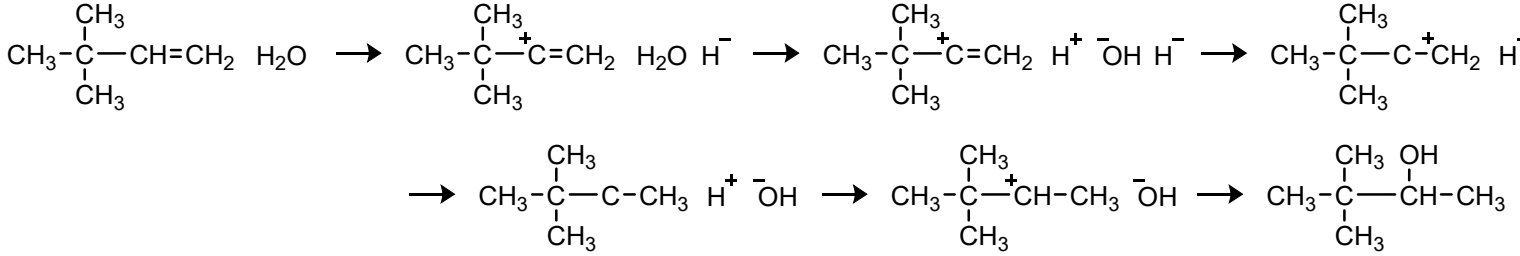
Top 3



Top 4

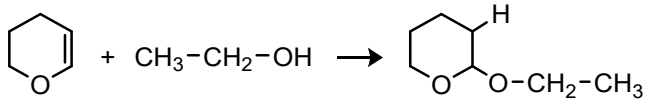


Top 5

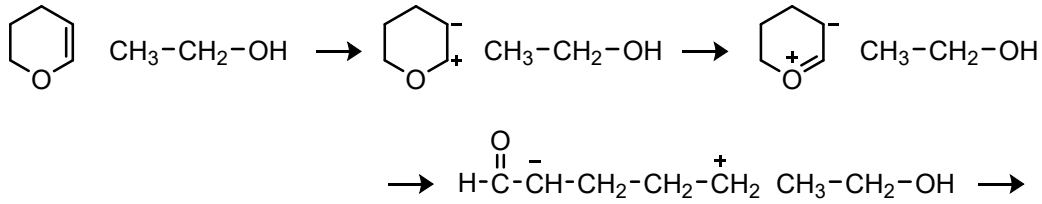


All graphs 4289 / Total candidates 53
All paths 10 619 / Correct paths 71

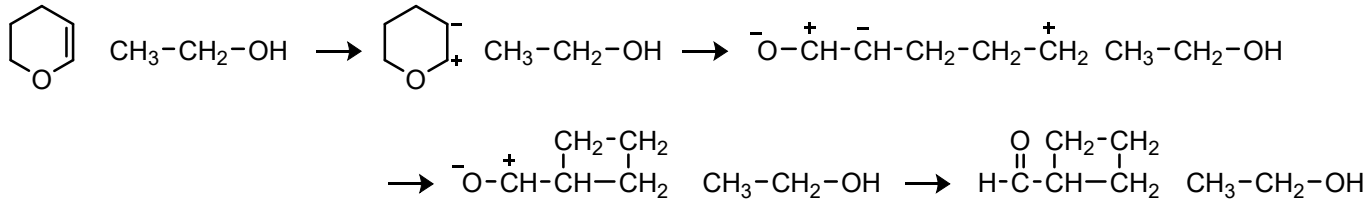
#17



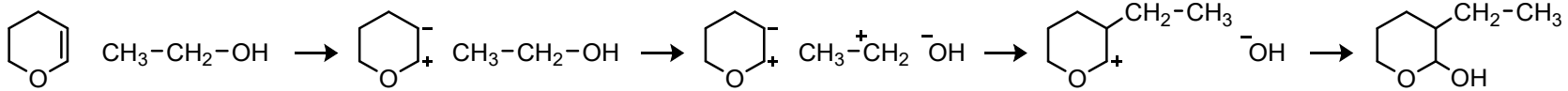
Top 1



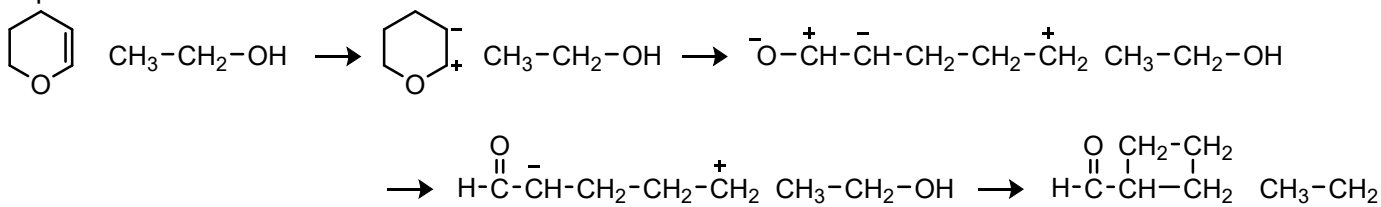
Top 2



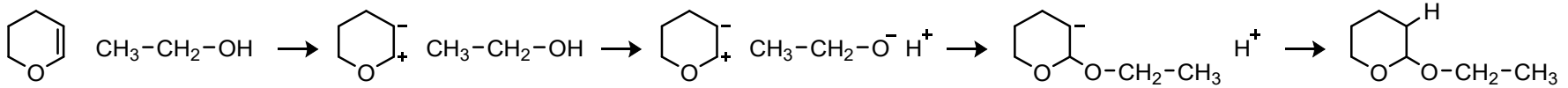
Top 3



Top 4

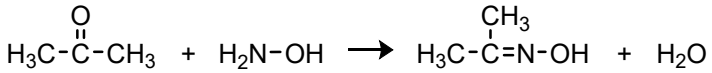


Top 5

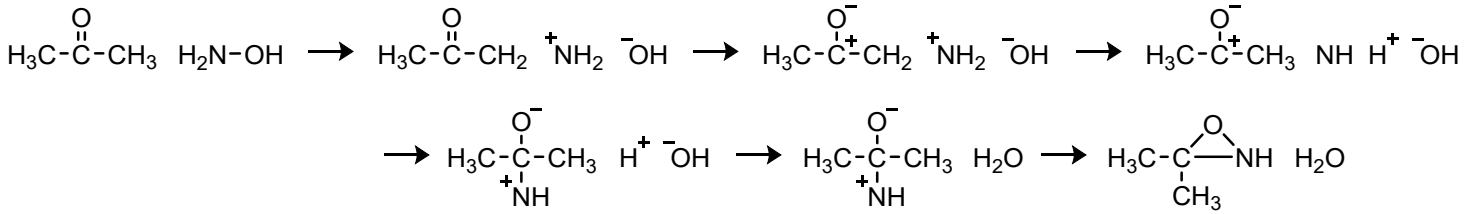


All graphs 3320 / Total candidates 25
All paths 145 / Correct paths 5

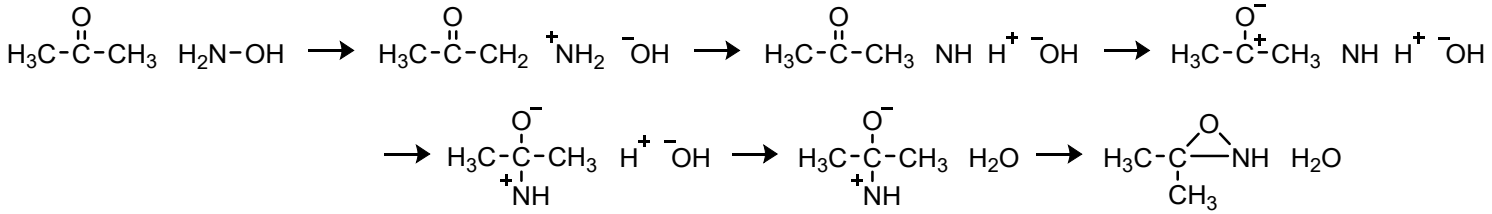
#18



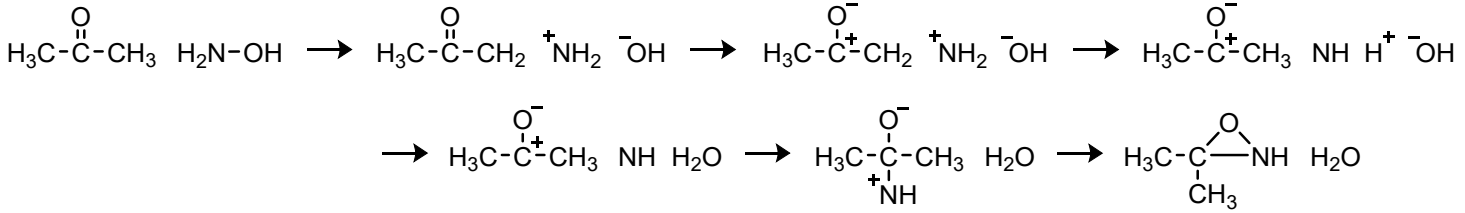
Top 1



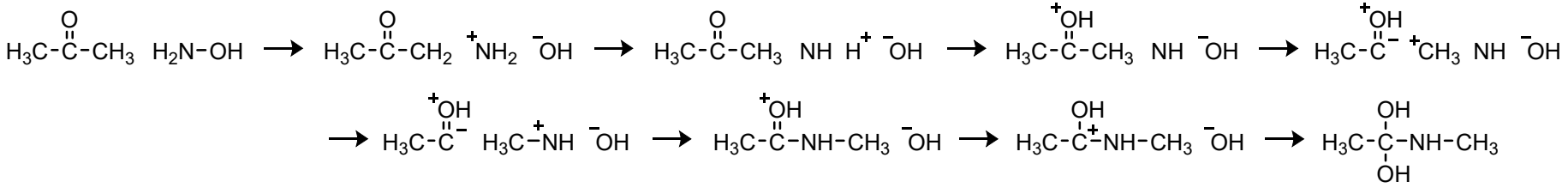
Top 2



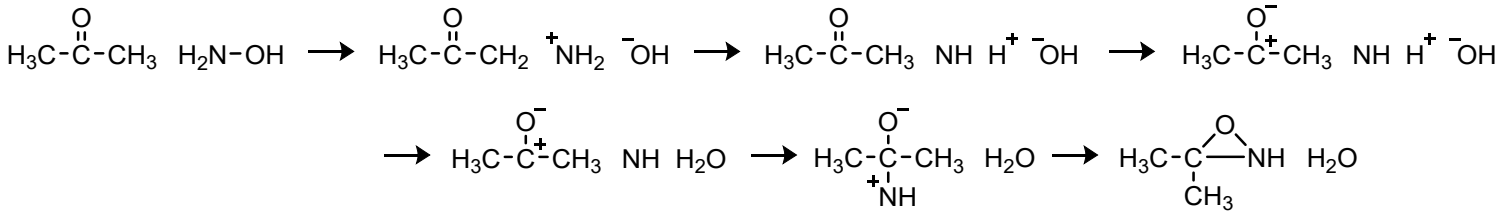
Top 3



Top 4

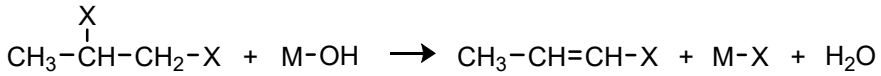


Top 5

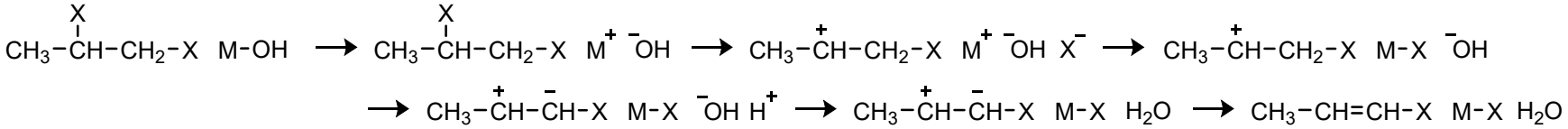


All graphs 4054 / Total candidates 54
All paths 25 692 / Correct paths 45

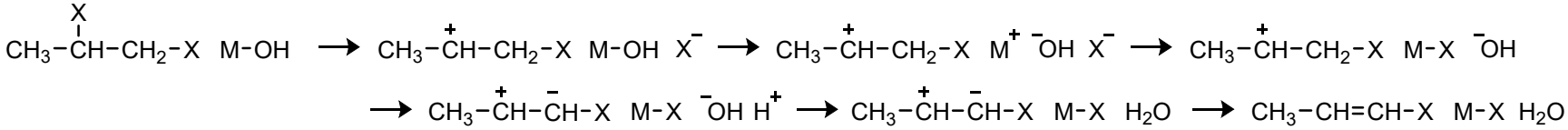
#19



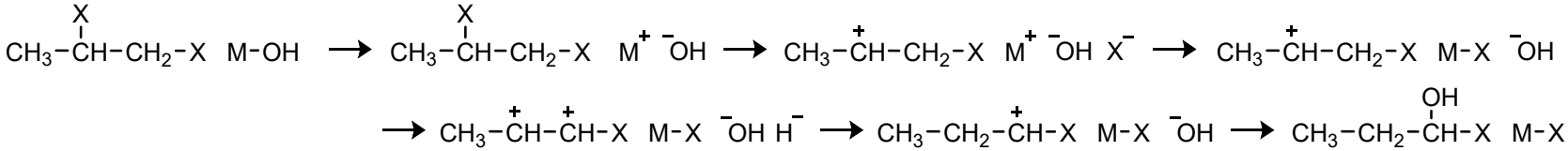
Top 1



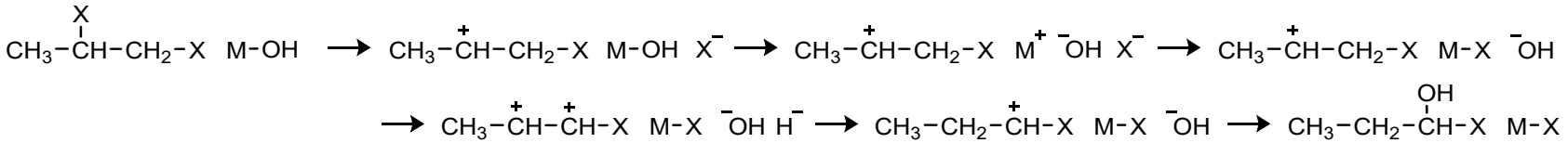
Top 2



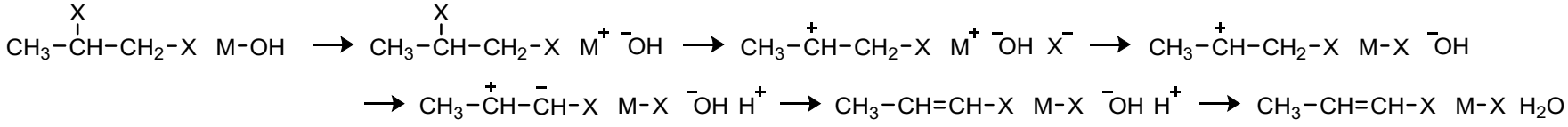
Top 3



Top 4

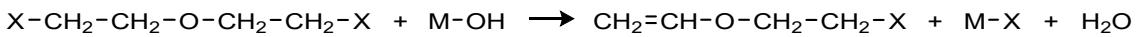


Top 5

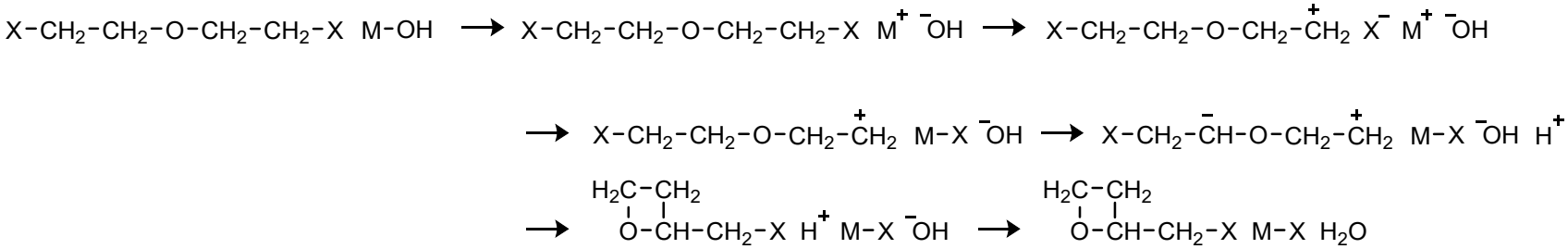


All graphs 6237 / Total candidates 96
All paths 2614 / Correct paths 25

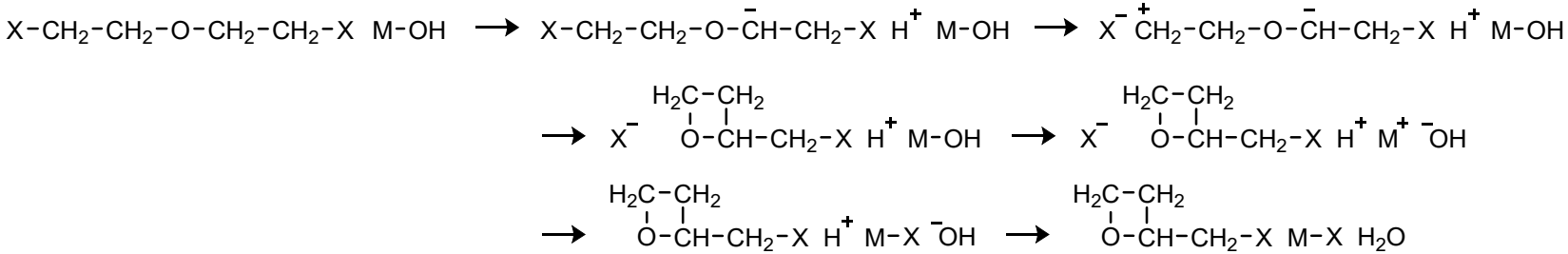
#20



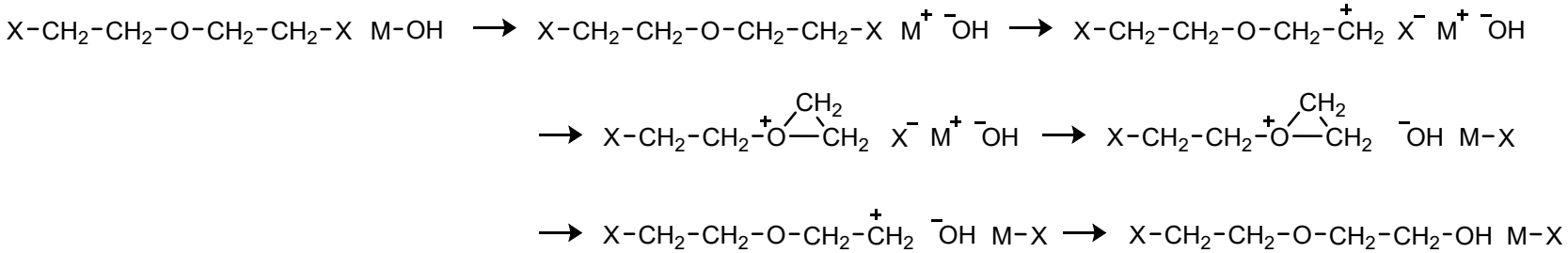
Top 1



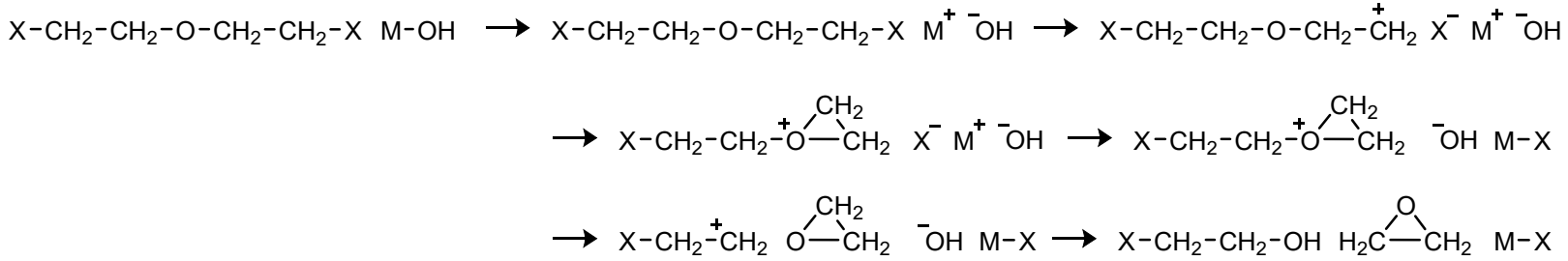
Top 2



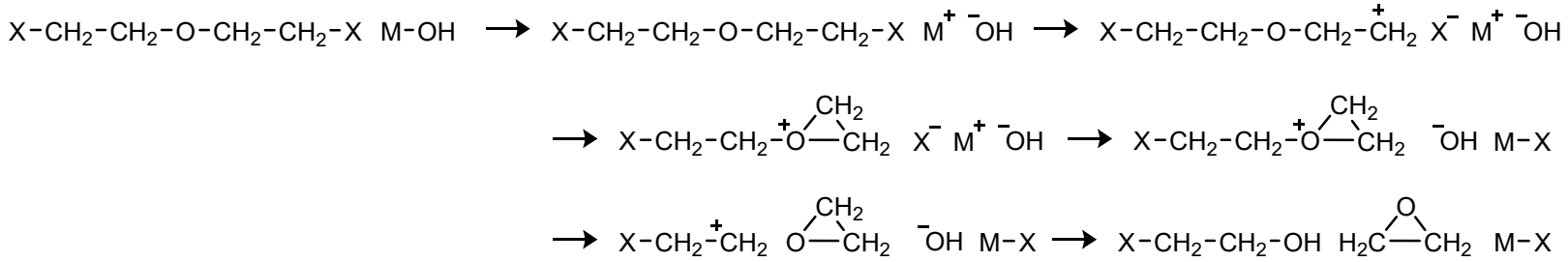
Top 3



Top 4

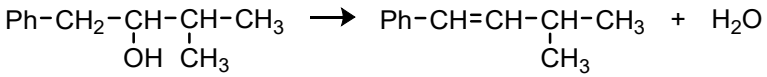


Top 5

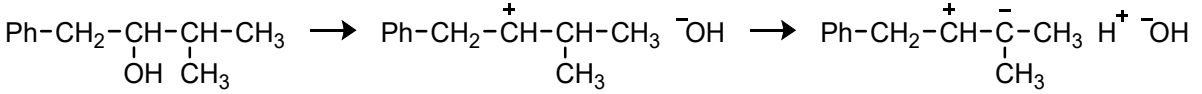


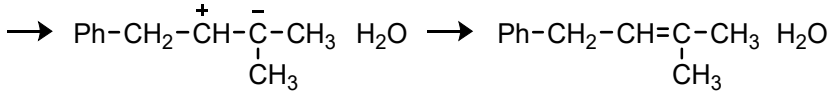
All graphs 5793 / Total candidates 91
All paths 2581 / Correct paths 25

#21

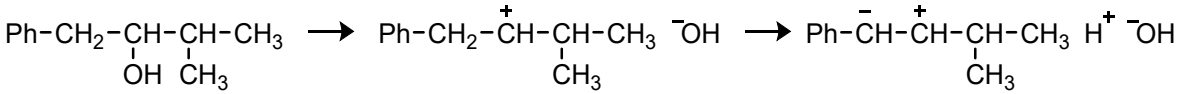


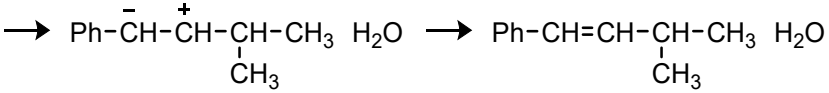
Top 1



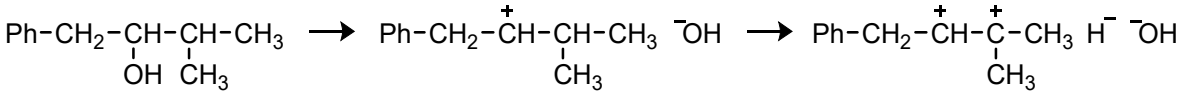


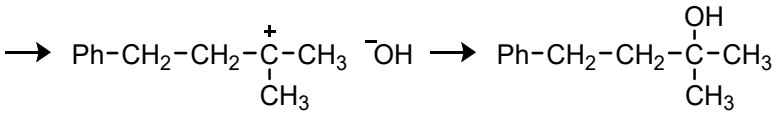
Top 2



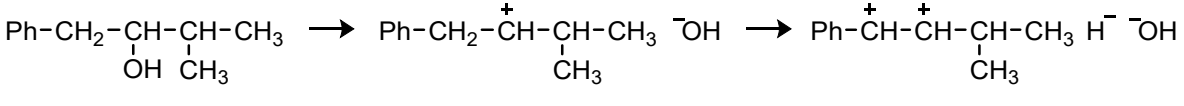


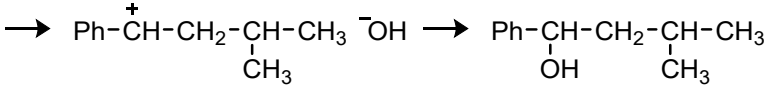
Top 3



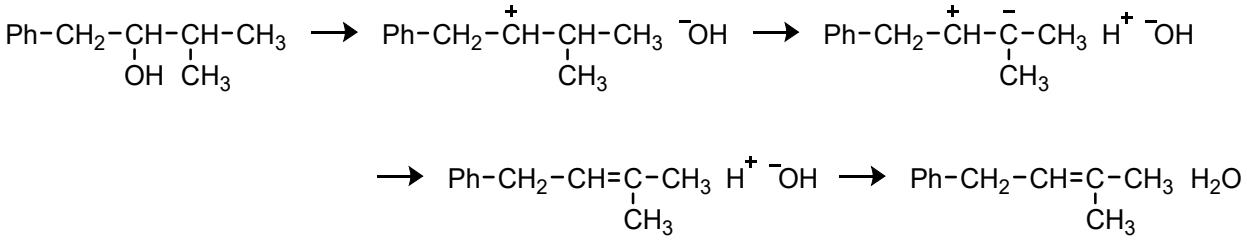


Top 4



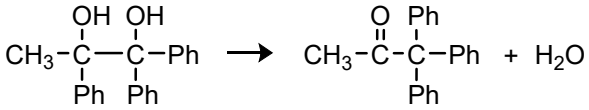


Top 5

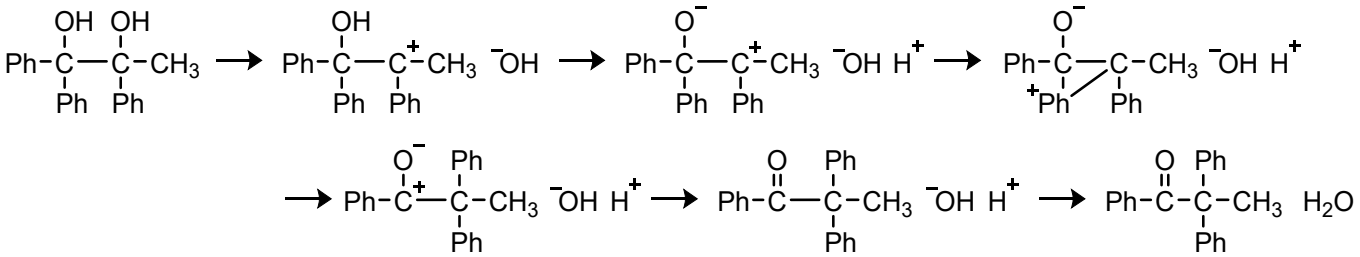


All graphs 4871 / Total candidates 33
All paths 154 / Correct paths 5

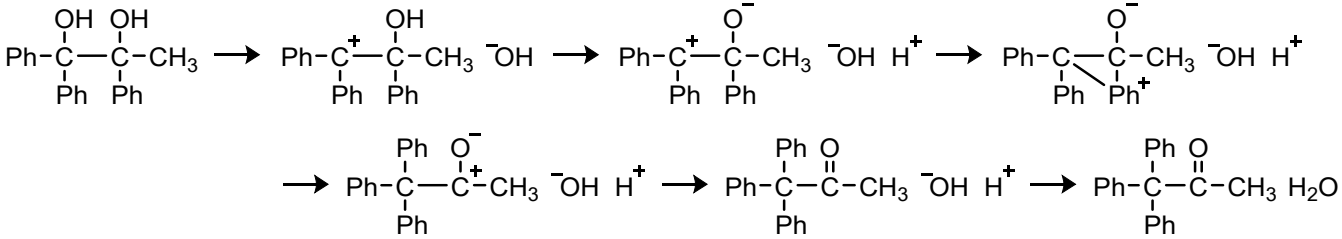
#22



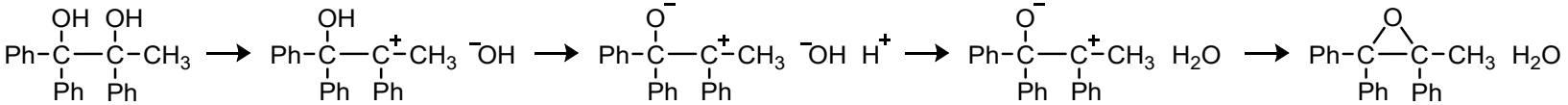
Top 1



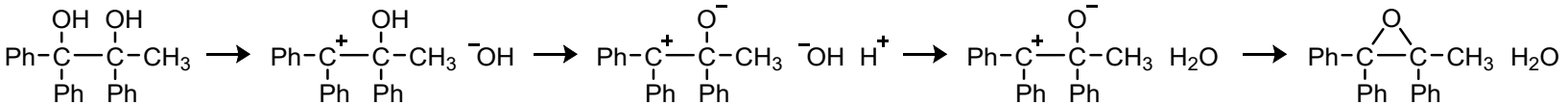
Top 2



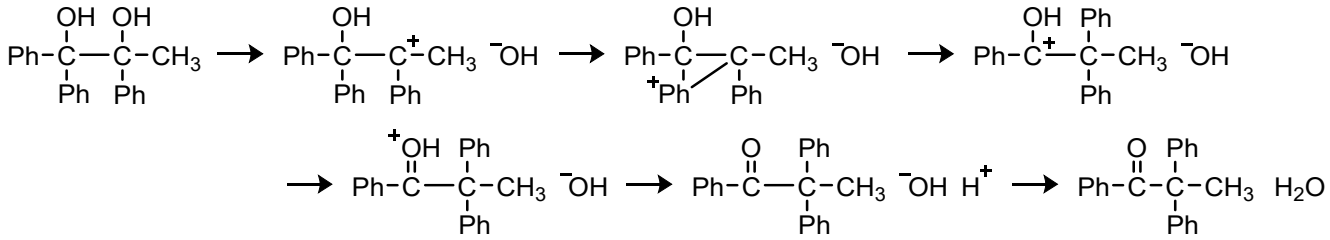
Top 3



Top 4

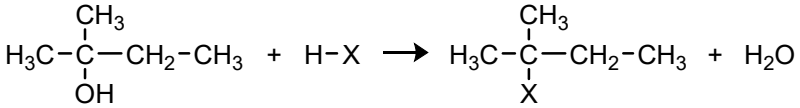


Top 5

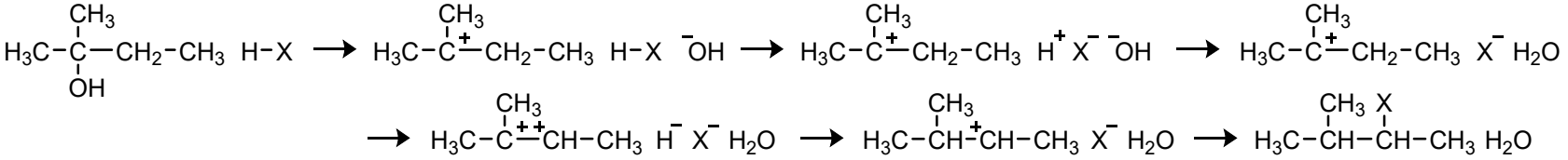


All graphs 3648 / Total candidates 31
All paths 521 / Correct paths 15

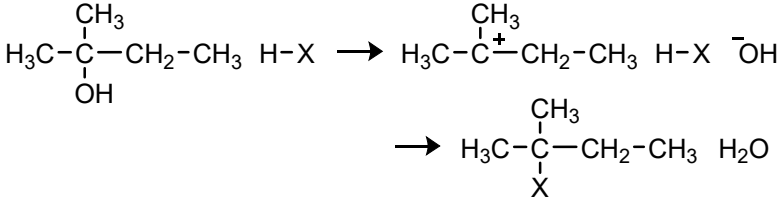
#23



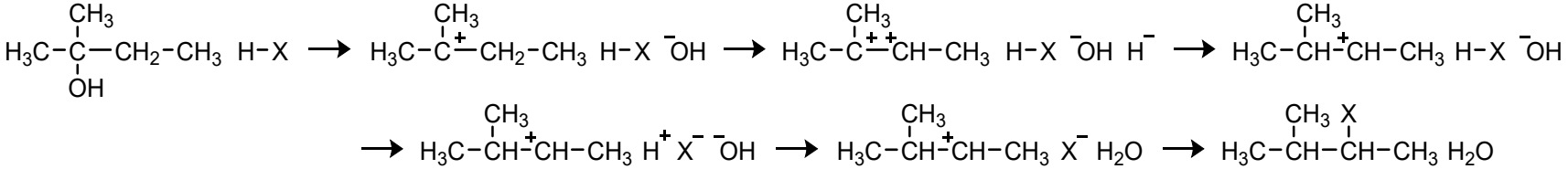
Top 1



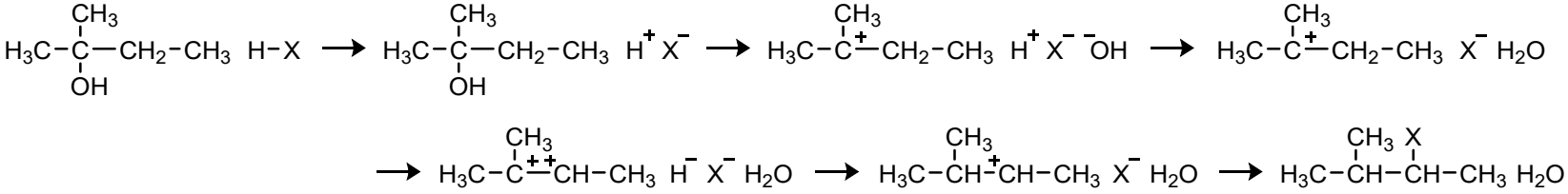
Top 2



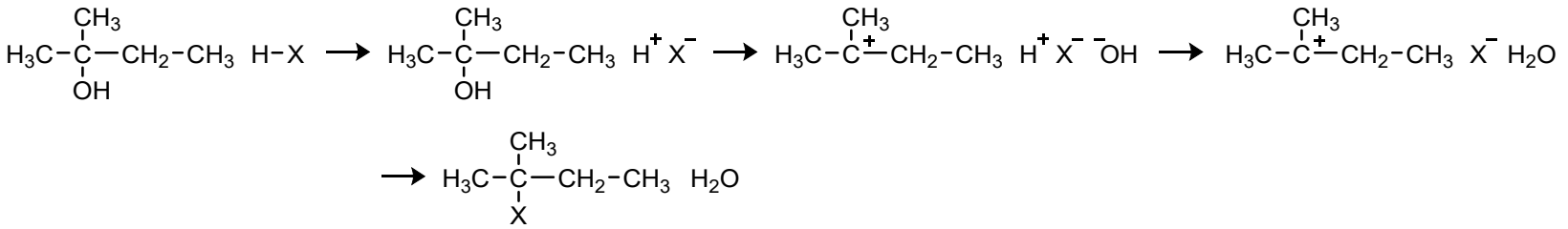
Top 3



Top 4

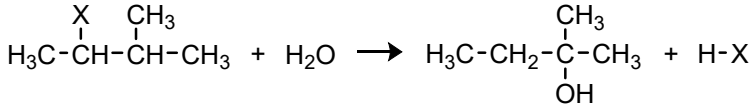


Top 5

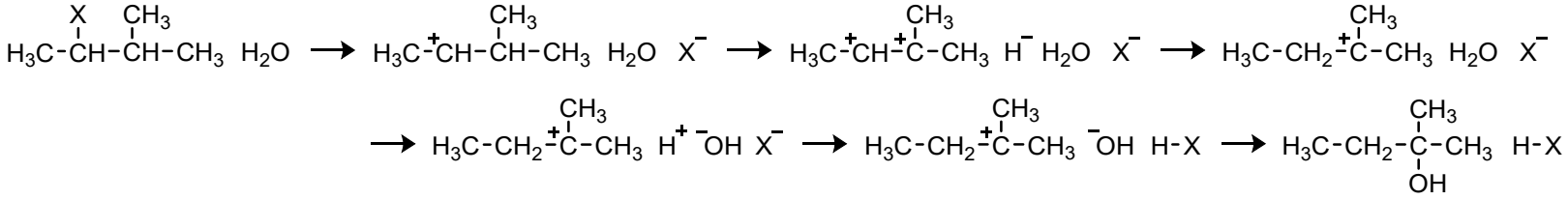


All graphs 6368 / Total candidates 98
All paths 2487 / Correct paths 5

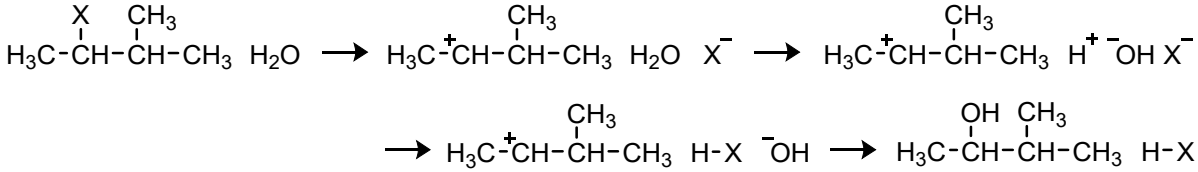
#24



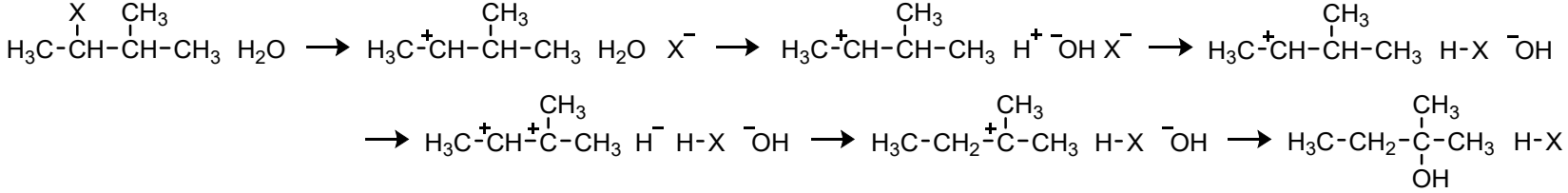
Top 1



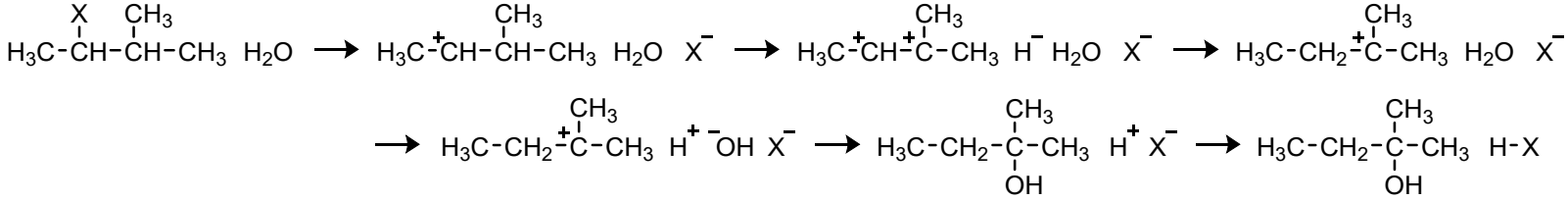
Top 2



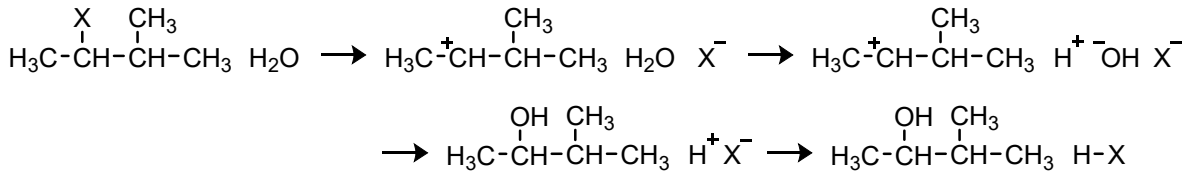
Top 3



Top 4

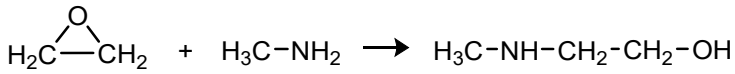


Top 5

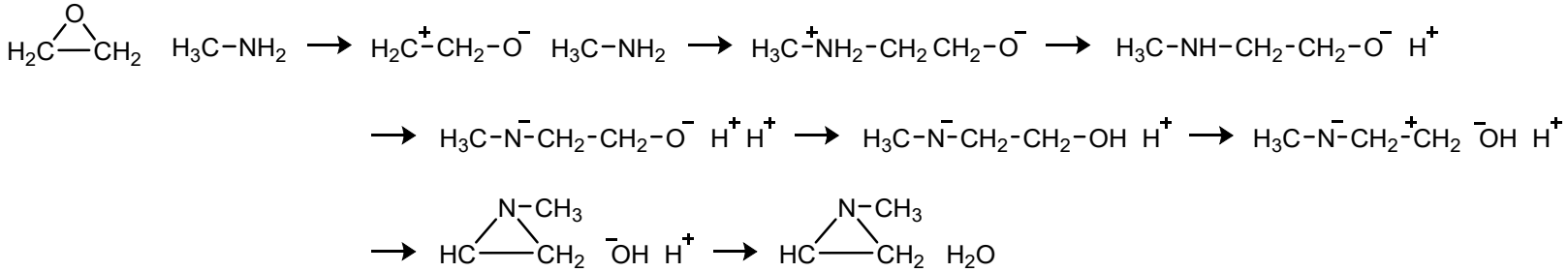


All graphs 4843 / Total candidates 64
All paths 1529 / Correct paths 25

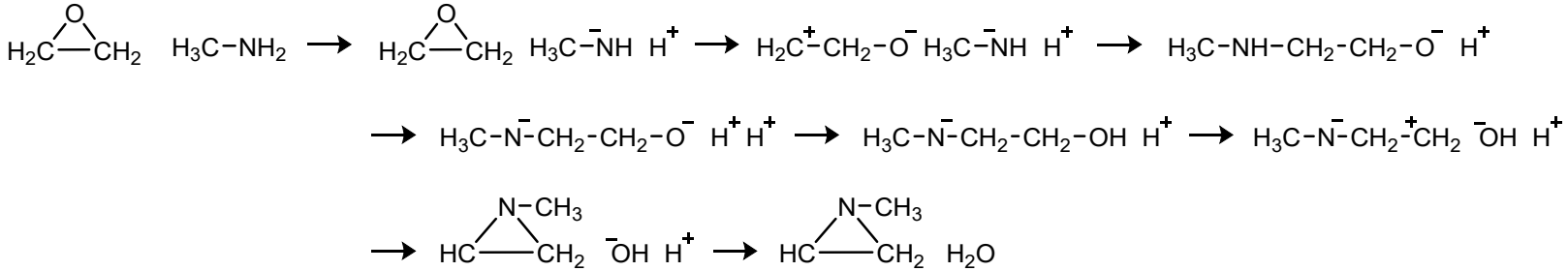
#25



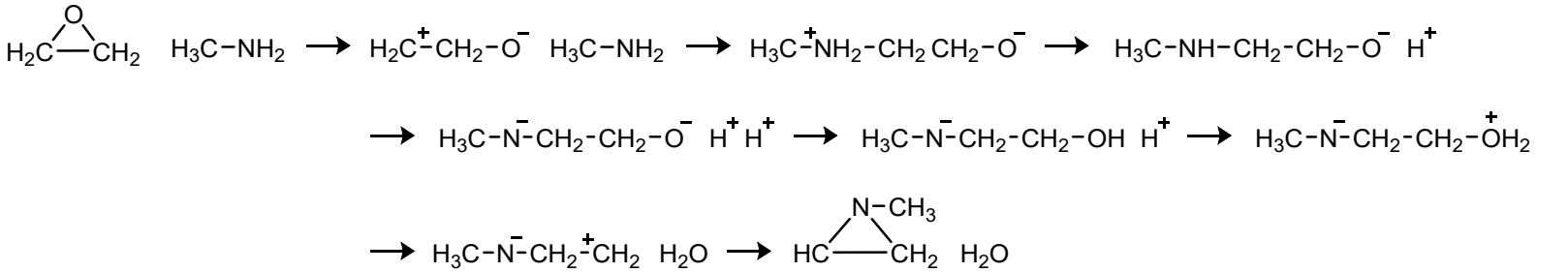
Top 1



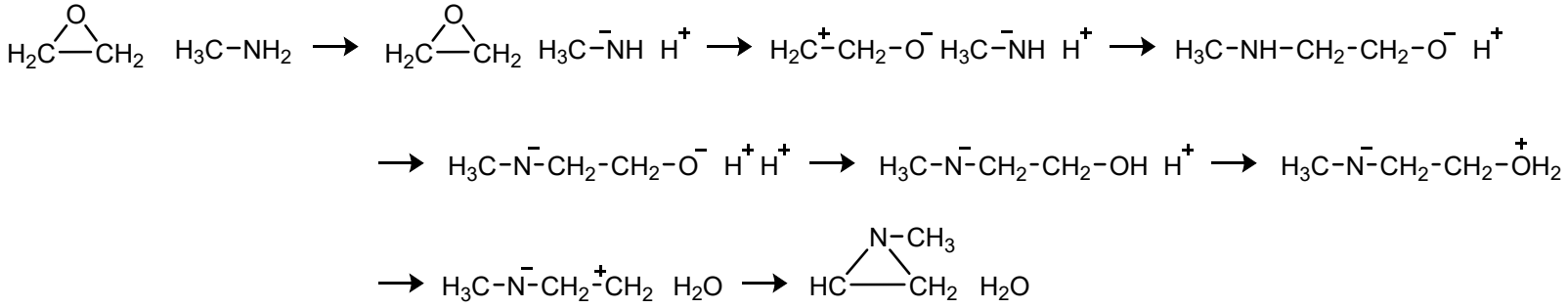
Top 2



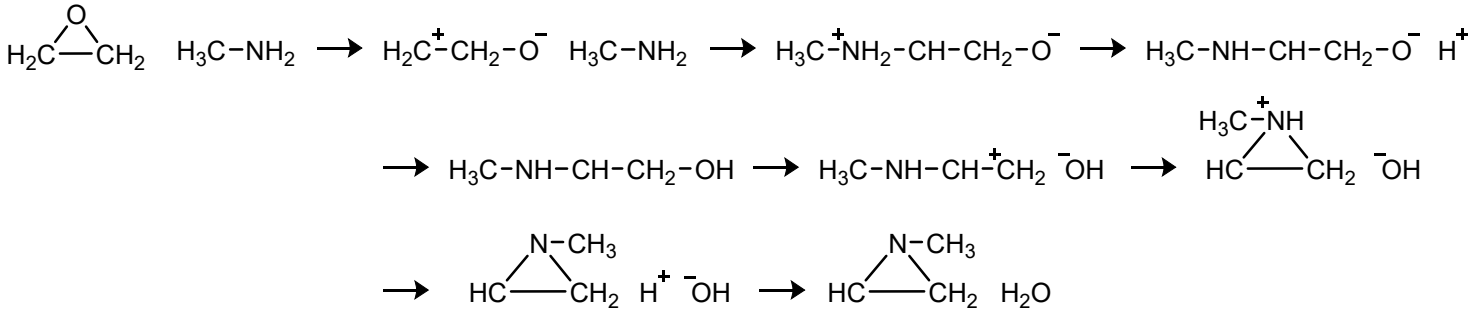
Top 3\



Top 4



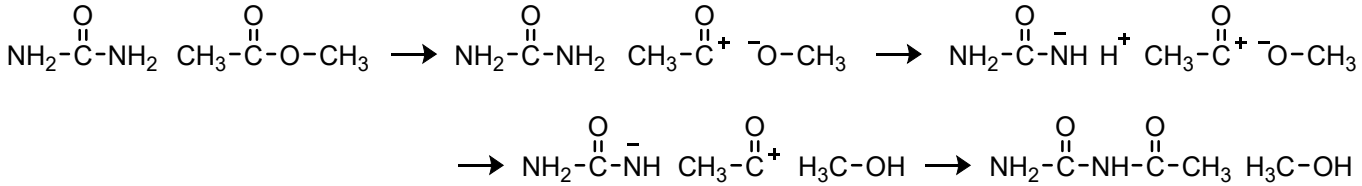
Top 5



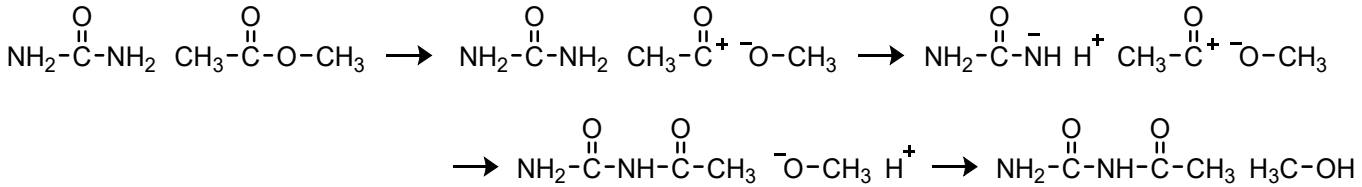
All graphs 1270 / Total candidates 50
All paths 364 154 / Correct paths 6

#26

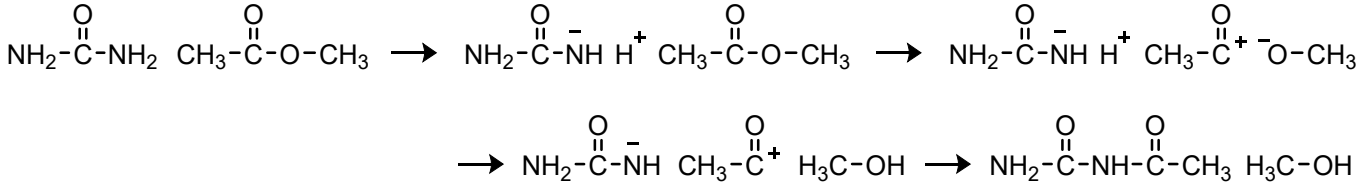
Top 1



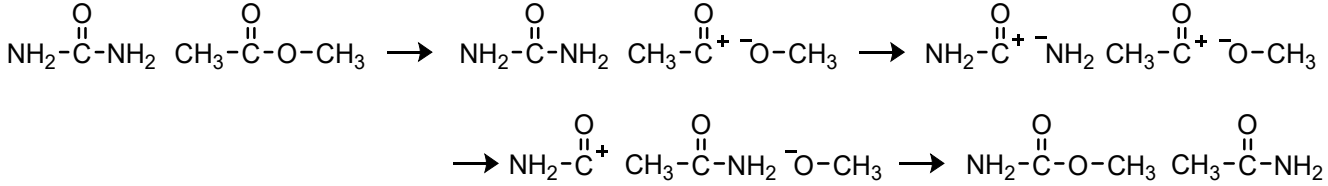
Top 2



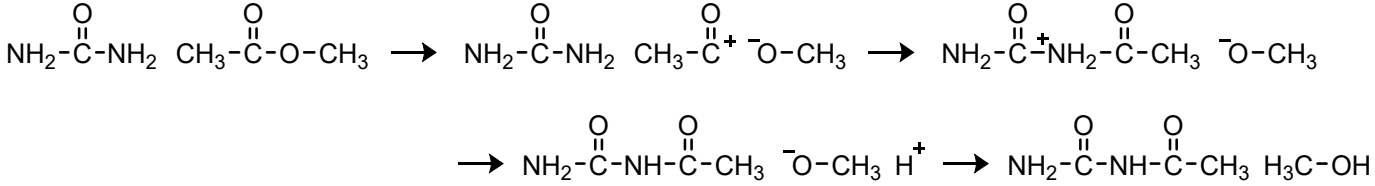
Top 3



Top 4

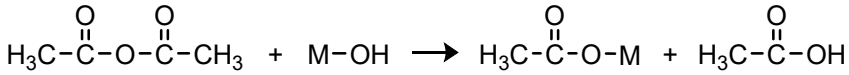


Top 5

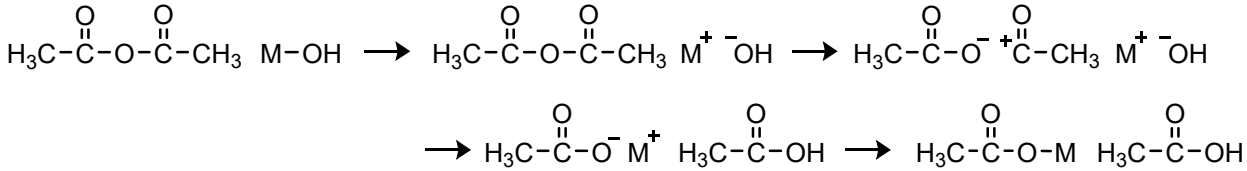


All graphs 6306 / Total candidates 23
All paths 143 / Correct paths 6

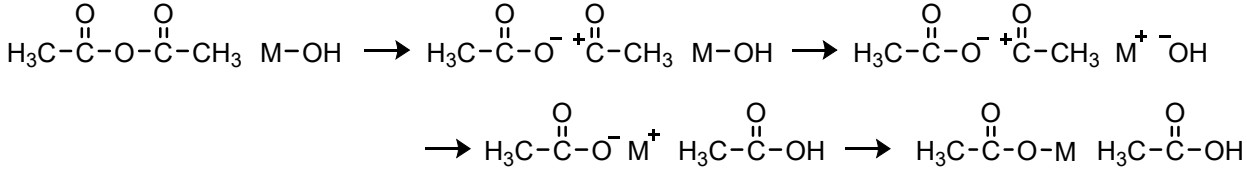
#27



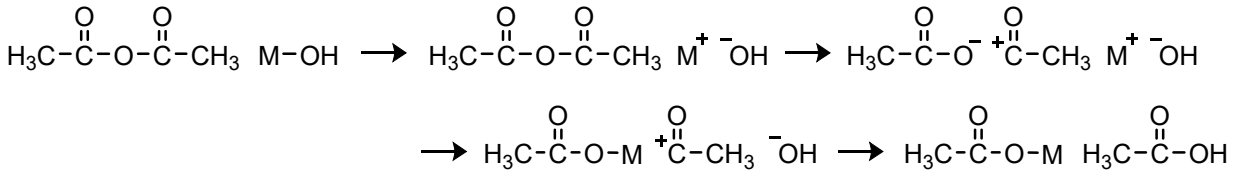
Top 1



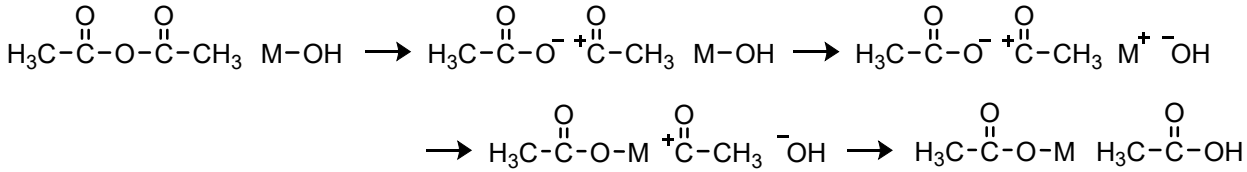
Top 2



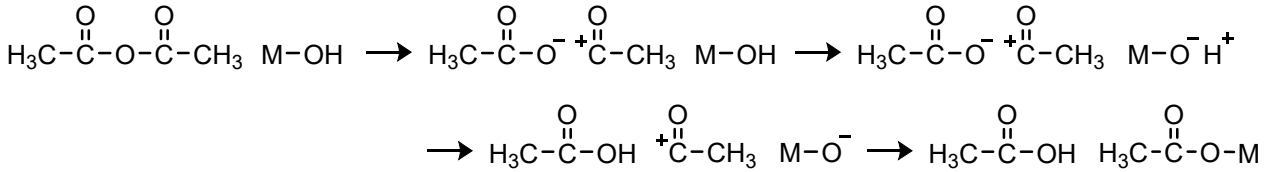
Top 3



Top 4



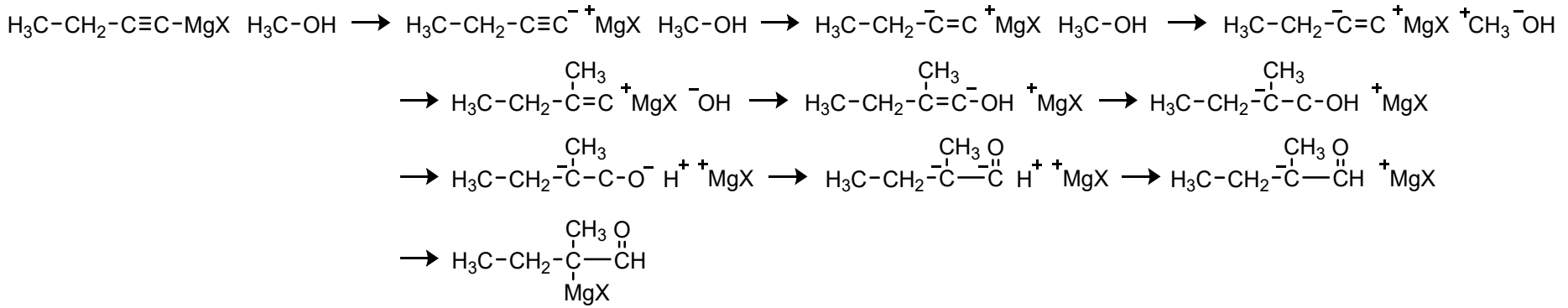
Top 5



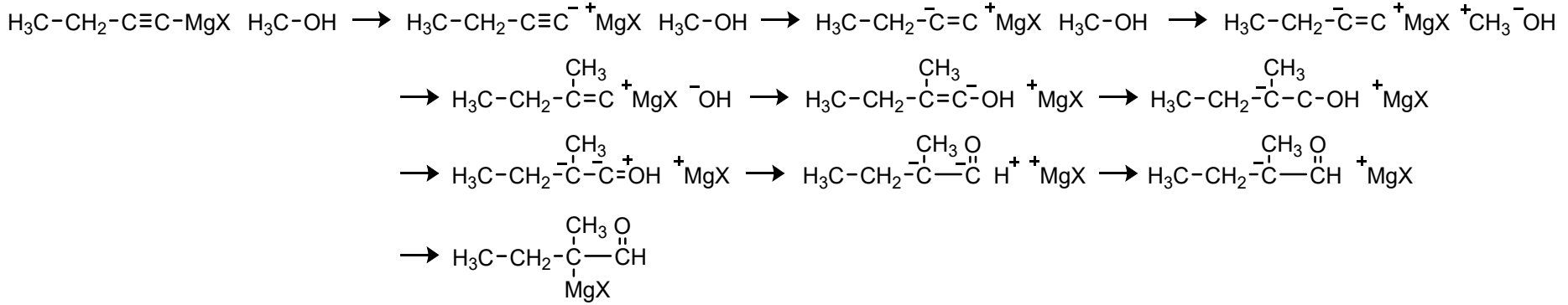
All graphs 3408 / Total candidates 24
All paths 833 / Correct paths 12

#28

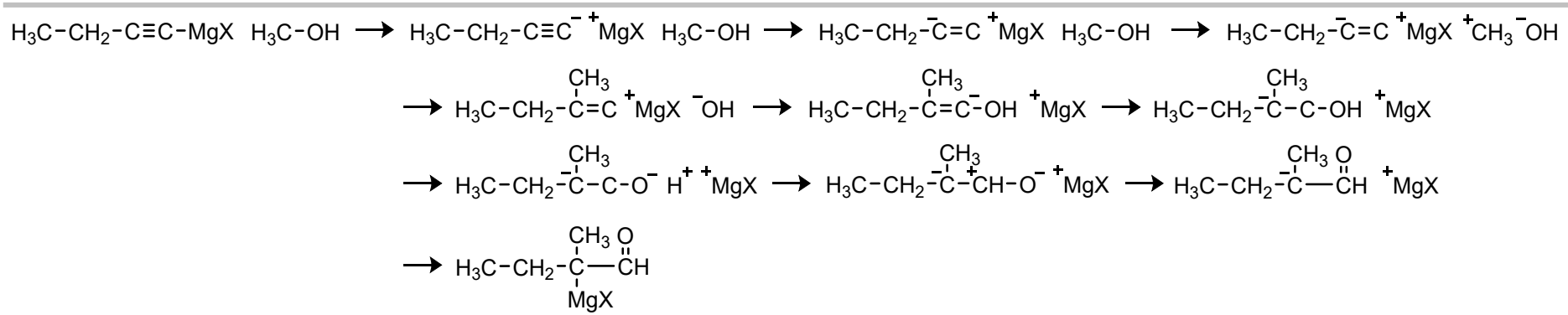
Top 1



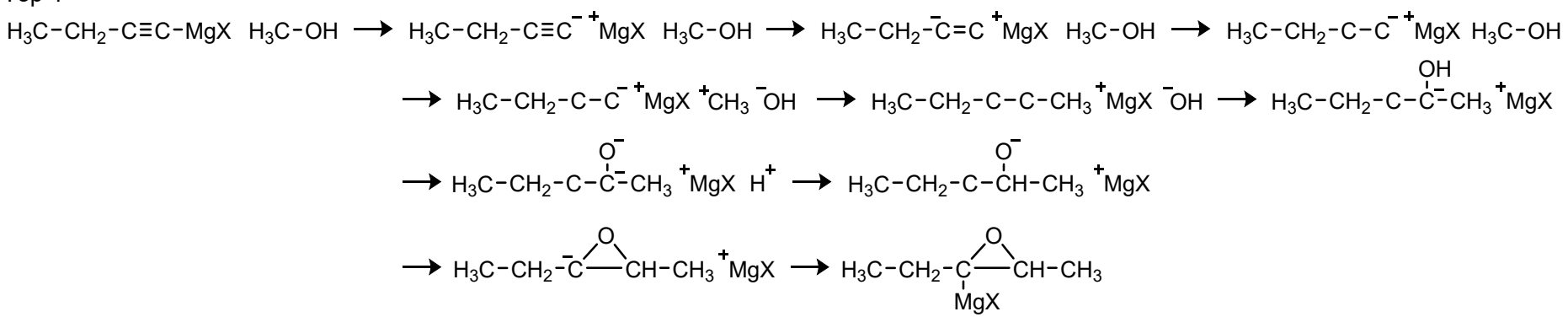
Top 2



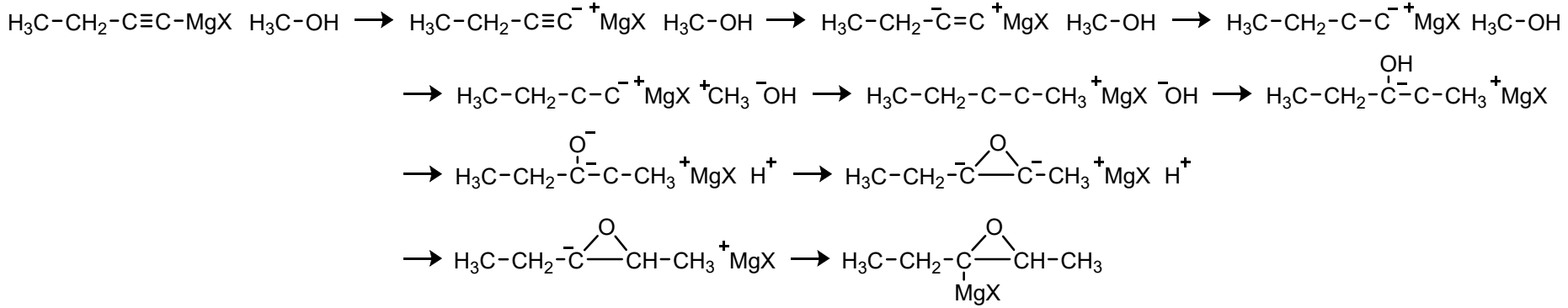
Top 3



Top 4

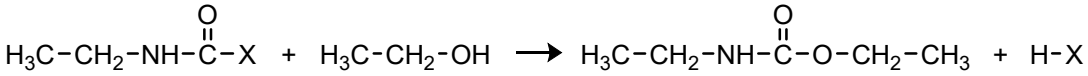


Top 5

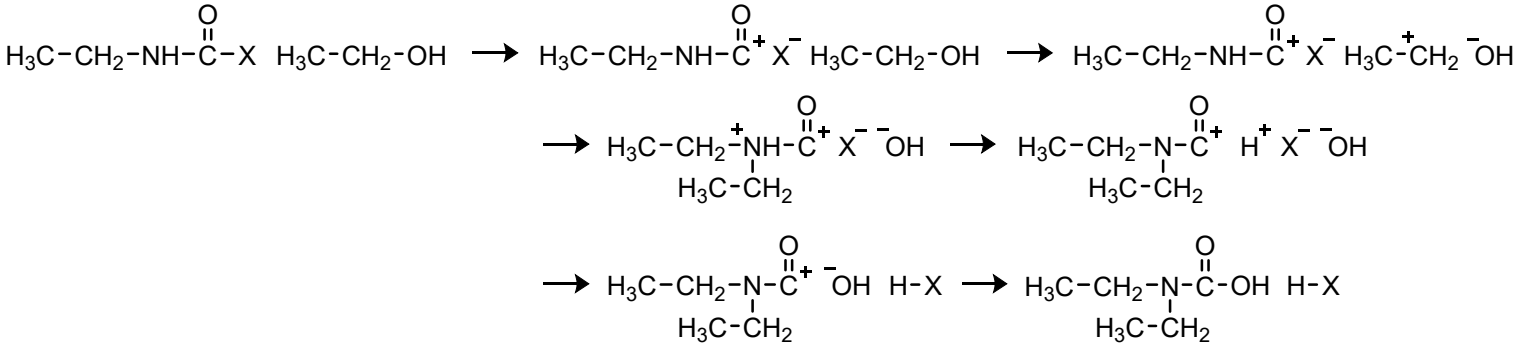


All graphs 3771 / Total candidates 88
All paths 105 205 / Correct paths 5

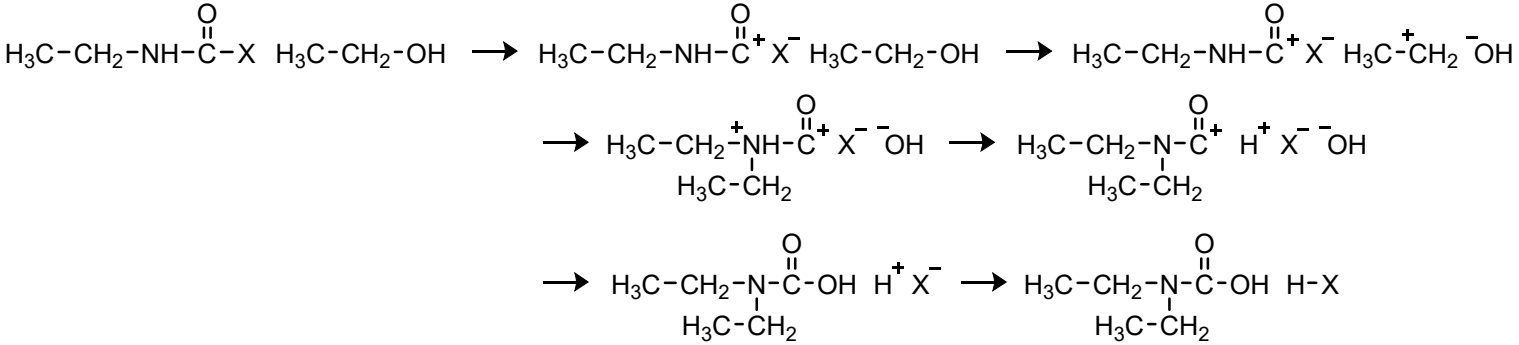
#29



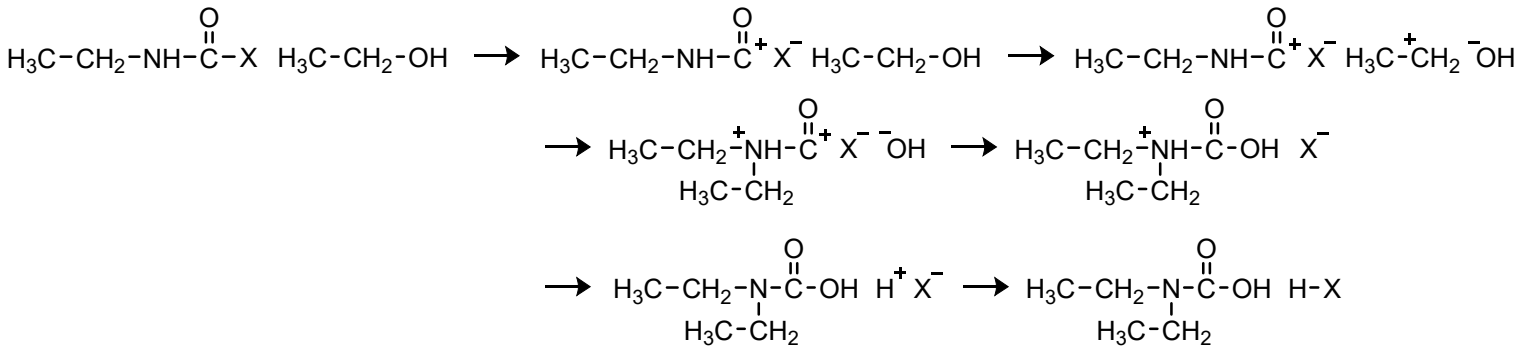
Top 1



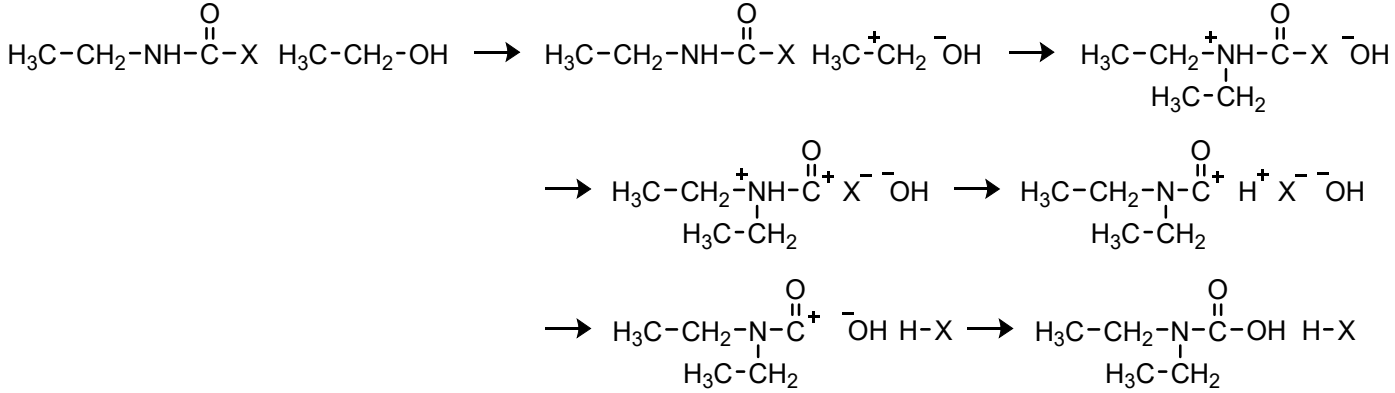
Top 2



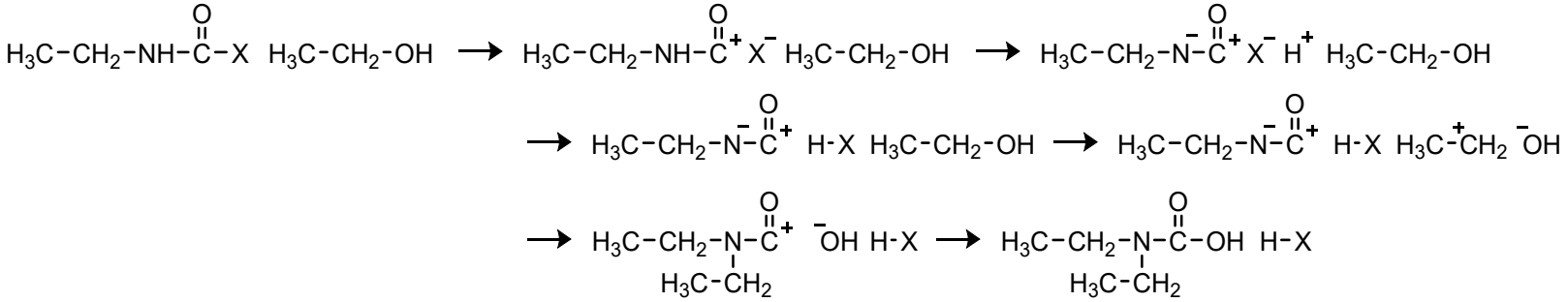
Top 3



Top 4



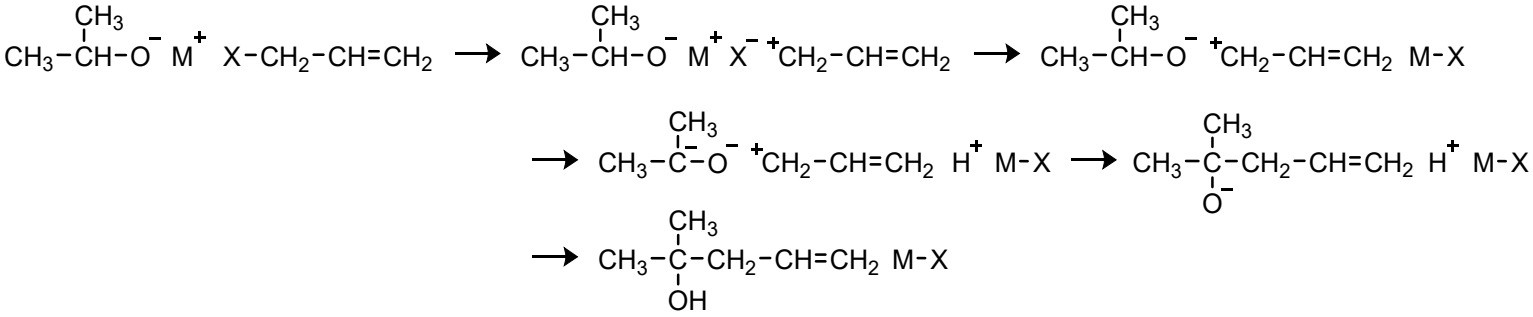
Top 5



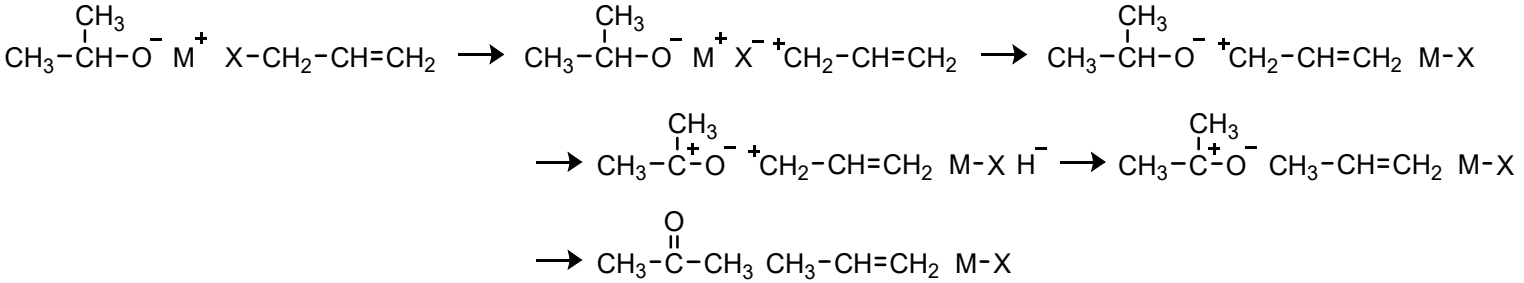
All graphs 4379 / Total candidates 67
All paths 1742 / Correct paths 5

#30

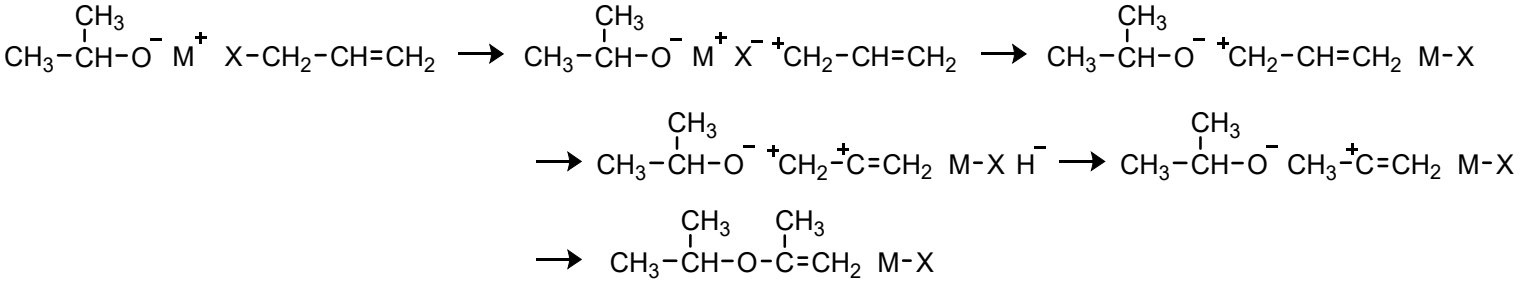
Top 1



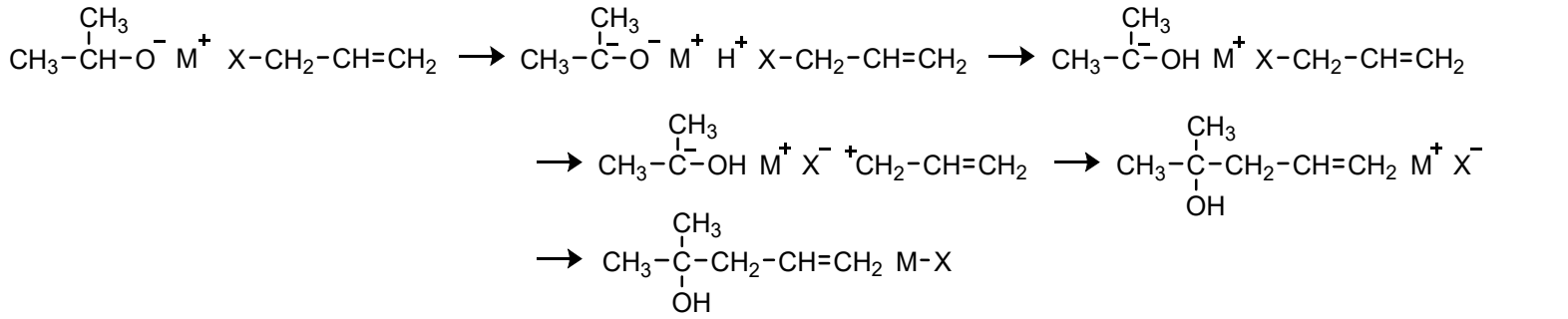
Top 2



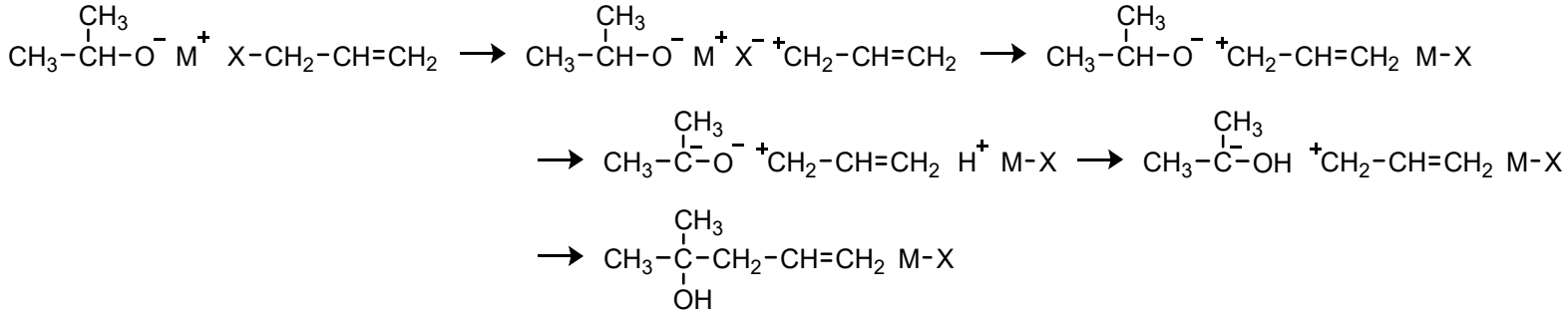
Top 3



Top 4

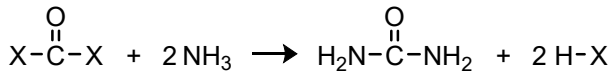


Top 5

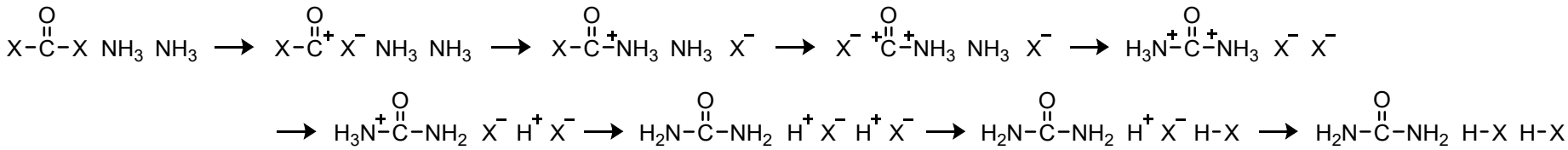


All graphs 6268 / Total candidates 63
All paths 369 / Correct paths 2

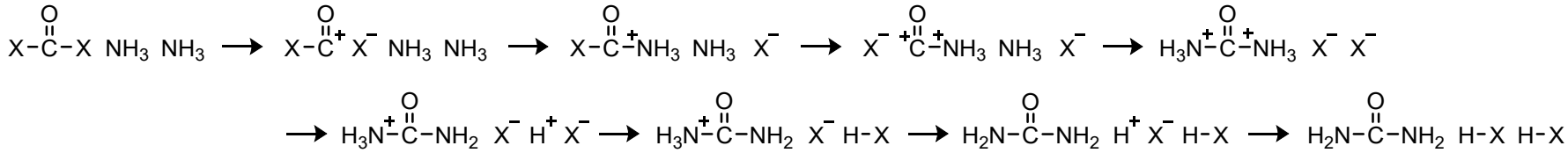
#31



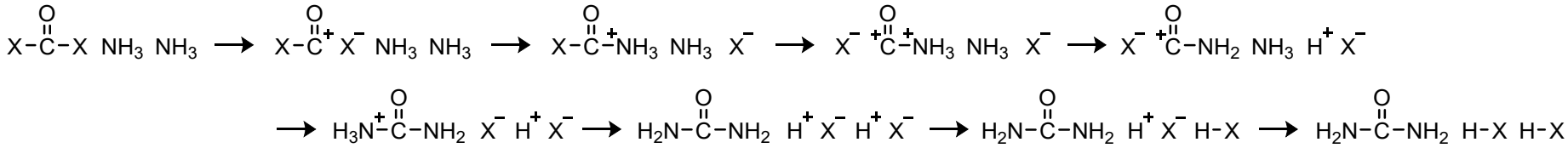
Top 1



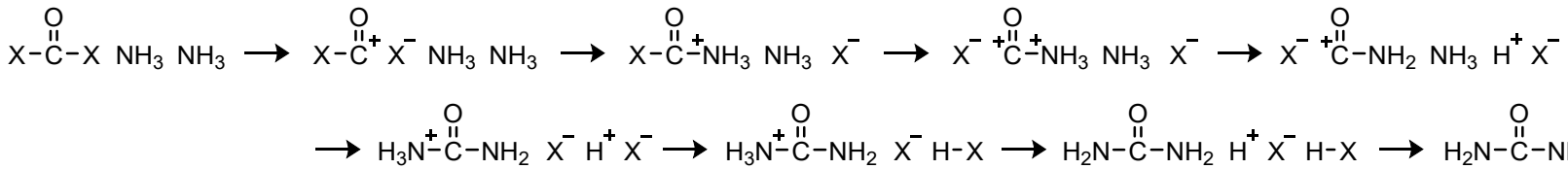
Top 2



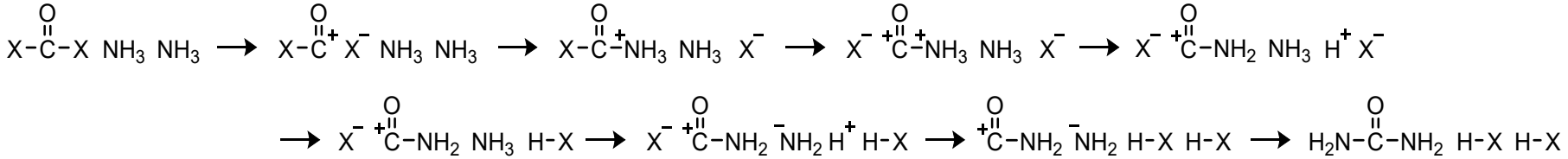
Top 3



Top 4

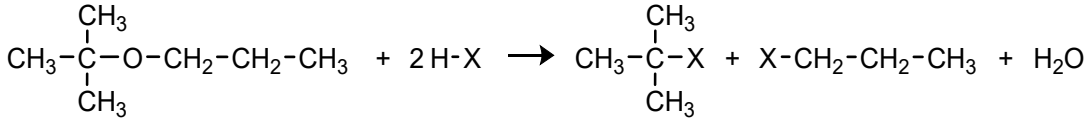


Top 5

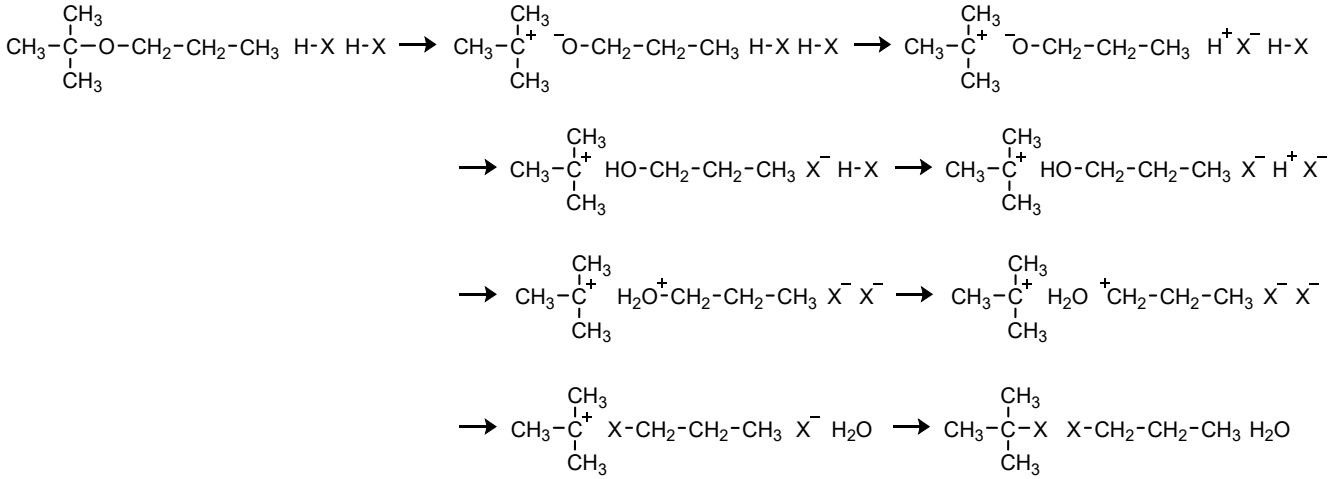


All graphs 5181 / Total candidates 36
All paths 7327 / Correct paths 88

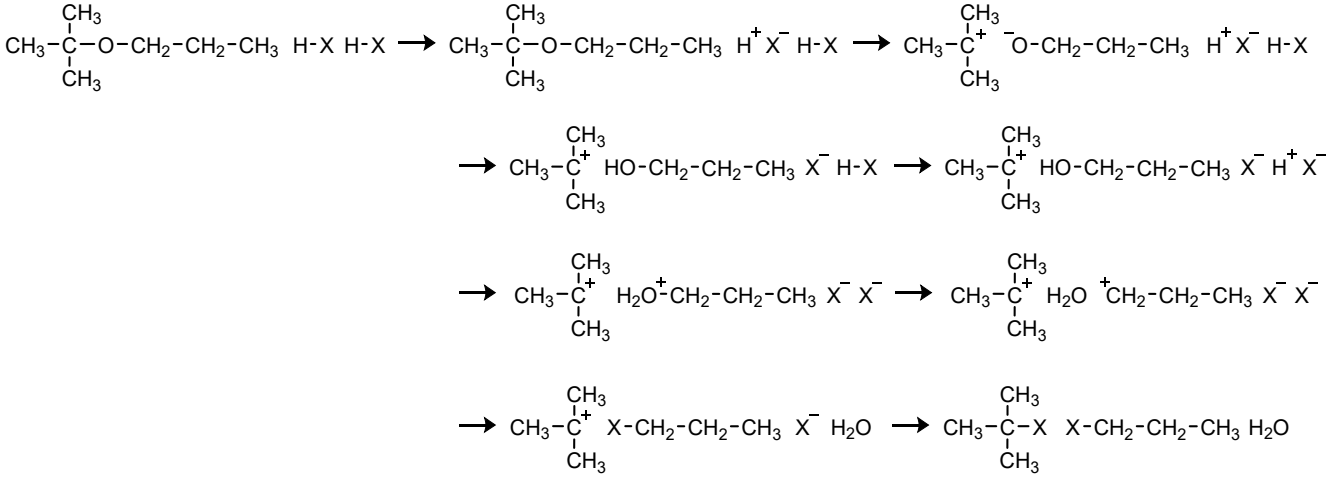
#32



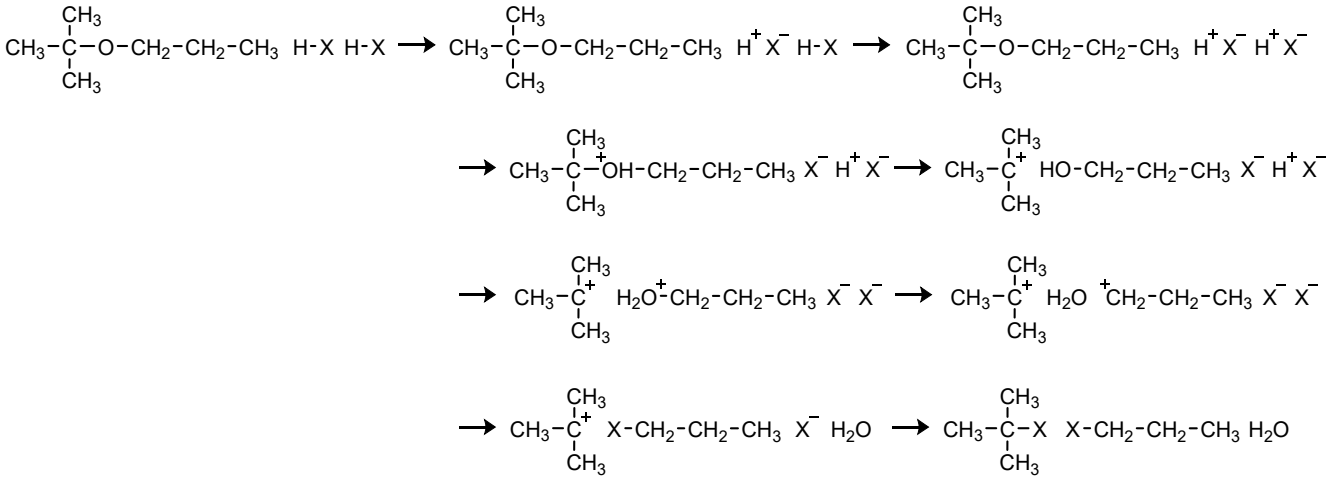
Top 1



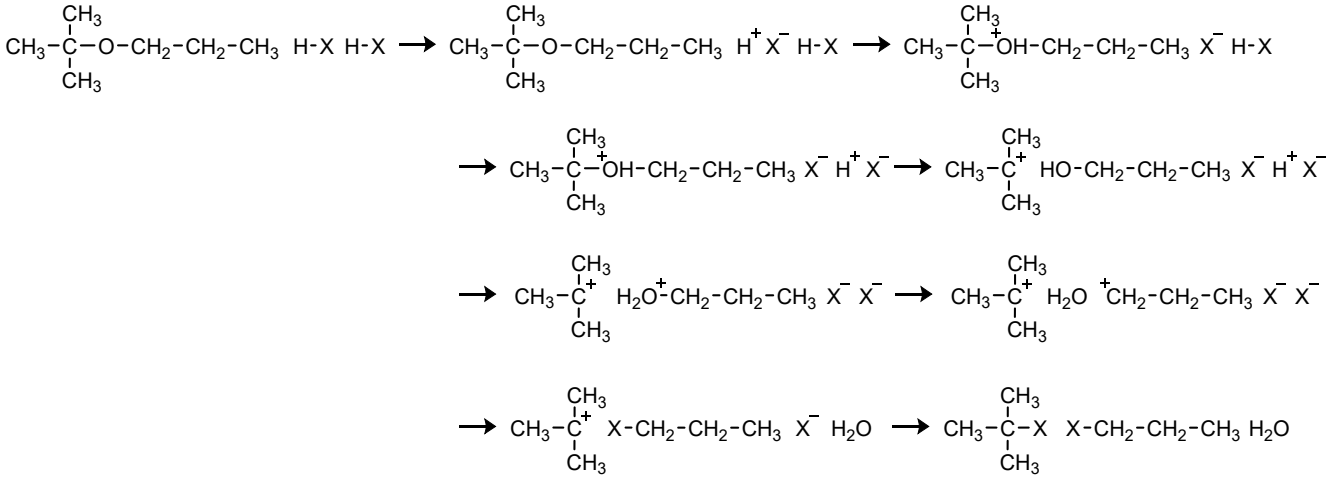
Top 2



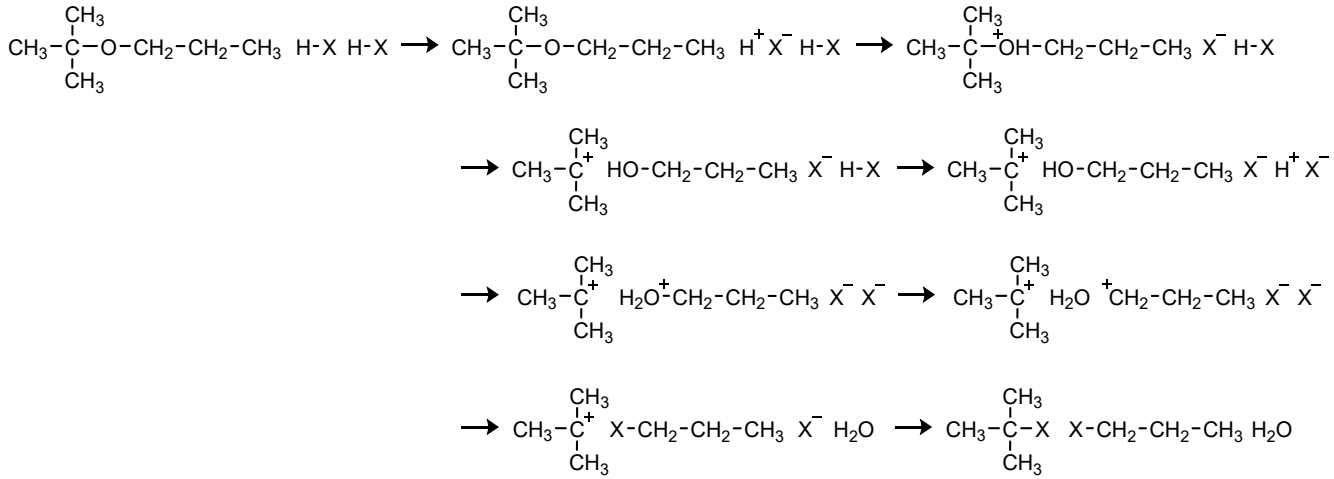
Top 3



Top 4

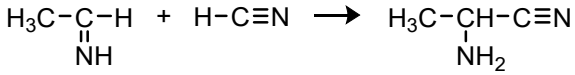


Top 5

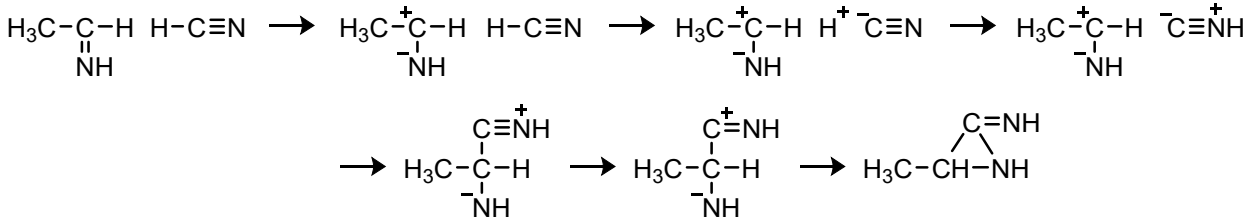


All graphs 3492 / Total candidates 106
All paths 40 610 / Correct paths 146

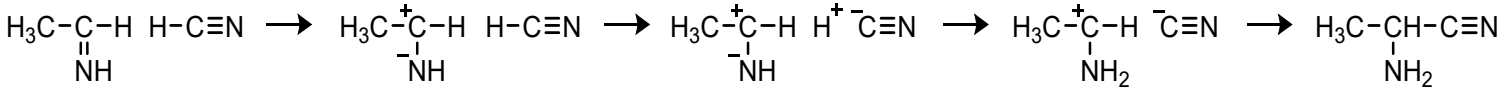
#33



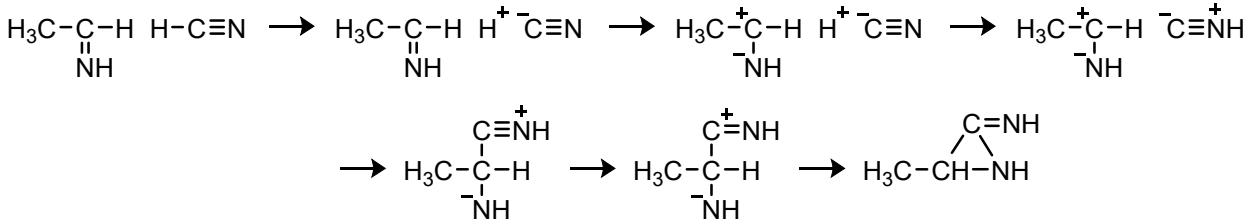
Top 1



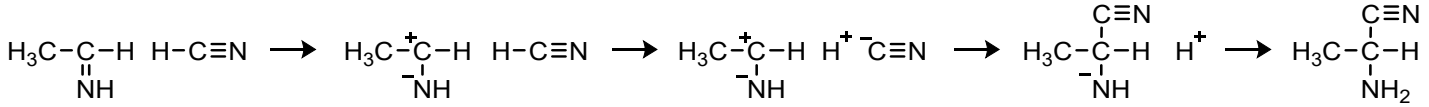
Top 2



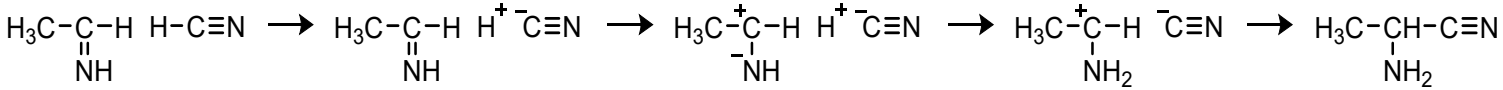
Top 3



Top 4

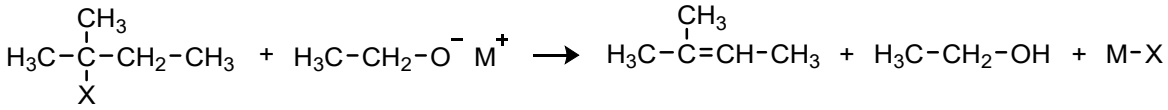


Top 5

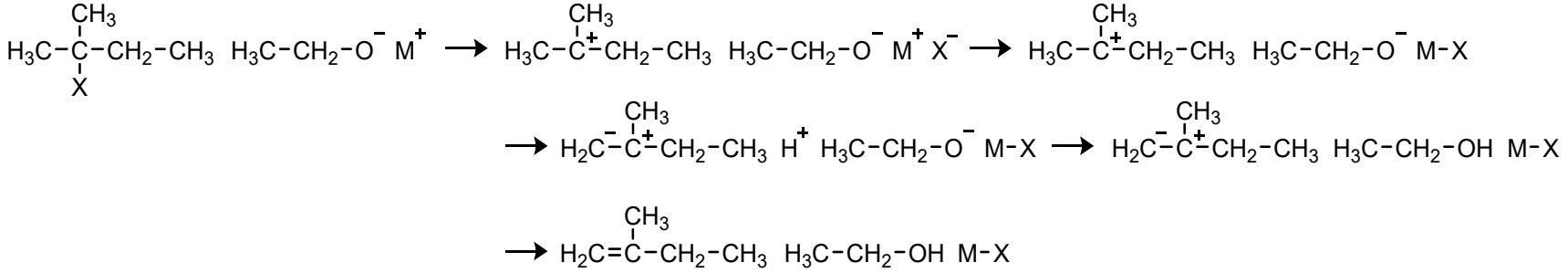


All graphs 3793 / Total candidates 29
All paths 1483 / Correct paths 9

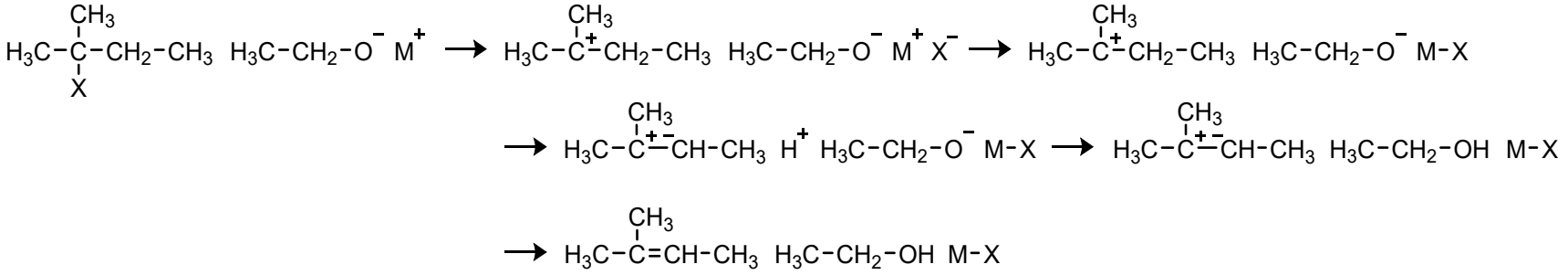
#34



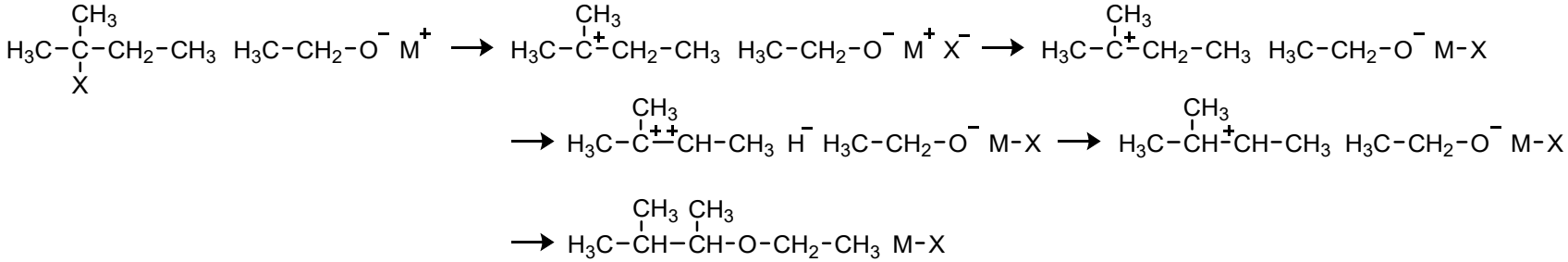
Top 1



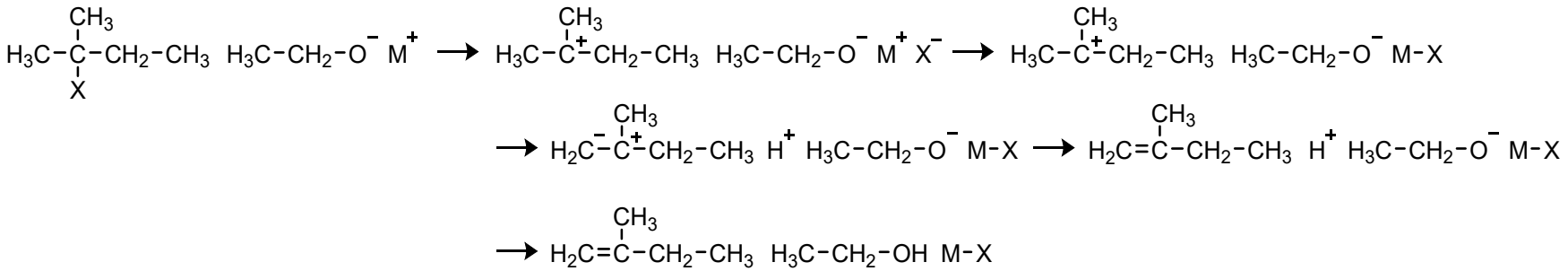
Top 2



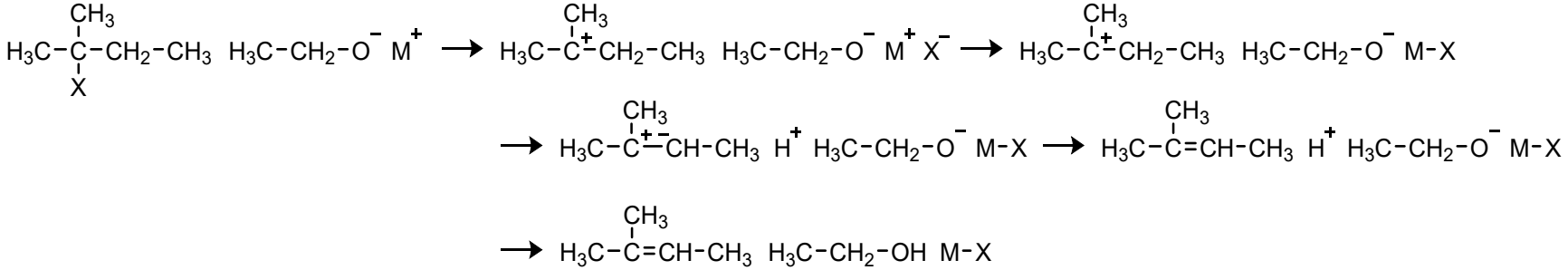
Top 3



Top 4

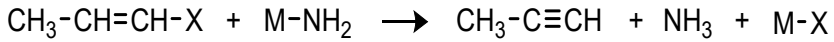


Top 5

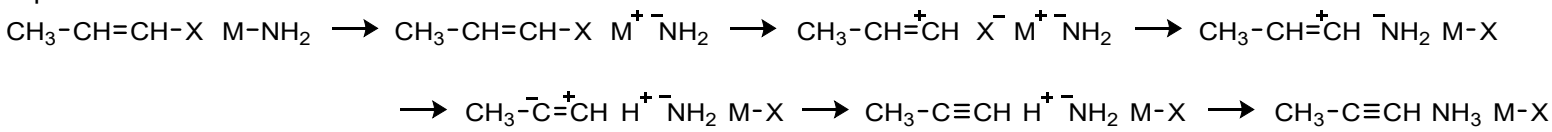


All graphs 5235 / Total candidates 92
All paths 736 / Correct paths 4

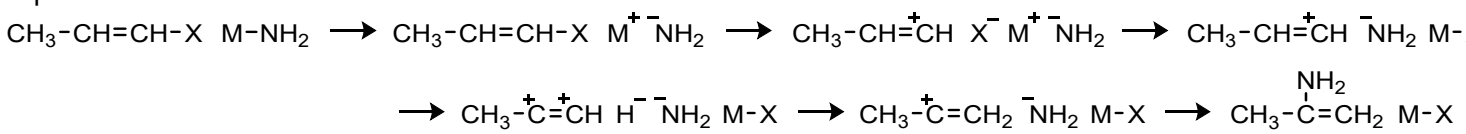
#35



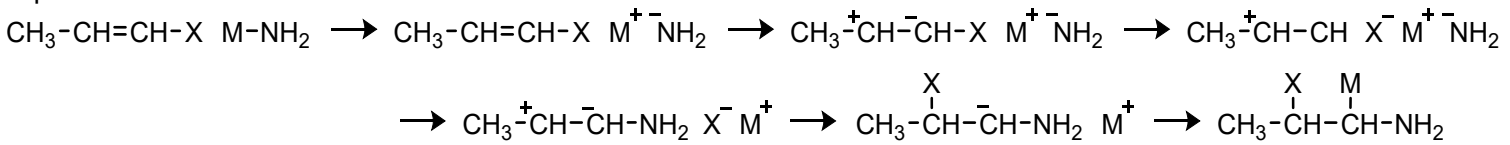
Top 1



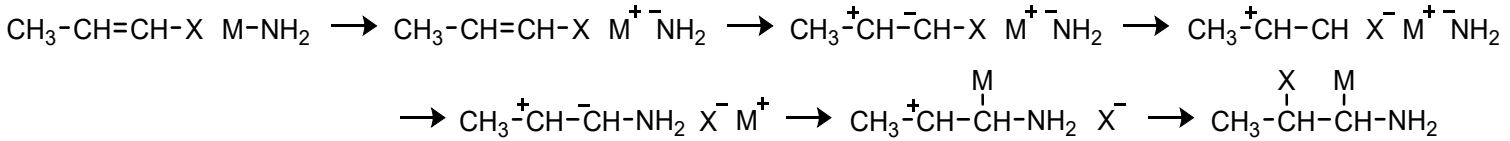
Top 2



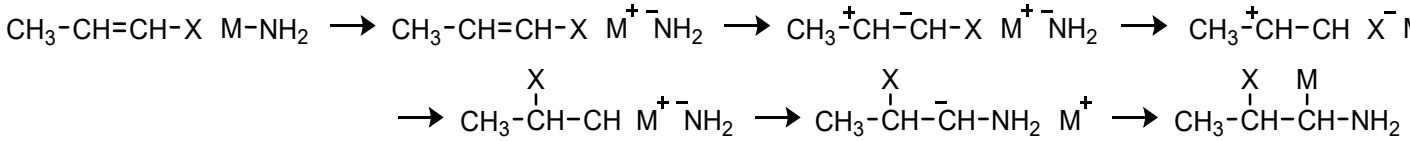
Top 3



Top 4



Top 5



All graphs 3764 / Total candidates 55
All paths 1822 / Correct paths 25

5.References

[24] T. Ida, M. Nishida, Y. Hori, *J. Phys. Chem. A* **2019**, *123*, 9579.
[27] M. Yano, *Organic Chemistry 1000 Nkocks for Reaction Mechanics*, Kagaku-Dojin, Tokyo, **2019**.
[28] The Pharmaceutical Society of Japan, *Essential Organic Reactions*, Kagaku-Dojin, Tokyo, **2019**.
[29] H. Meislich, H. Nechamkin, J. Sharefkin, G. Hademenos, *Schaum's Outline of Organic Chemistry*, McGraw Hill, New York, **2013**.
[30] K. P. C. Vollhardt, N. E. Schore, *Organic Chemistry: Structure and Function*, W. H. Freeman, New York, **2014**.
[31] D. P. Kingma, J. Ba, **2017**, arXiv preprint arXiv:1412.6980v9.