

# Cluster #0 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 0, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 8.0 ( $\pm 2.5$ ) on the prediction outcome.

## Properties

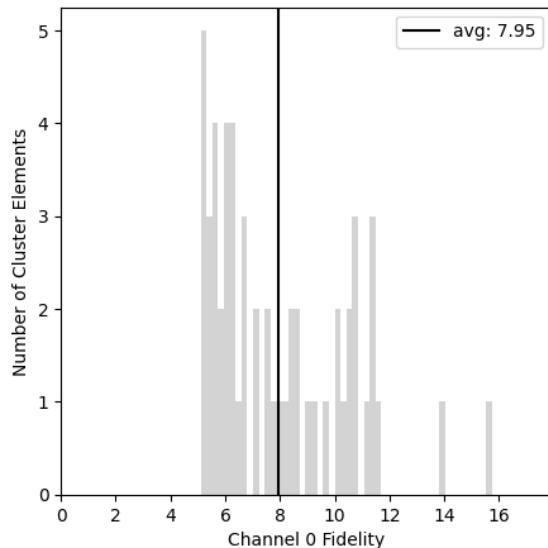
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	55
Channel Index	0.0 (0.0)

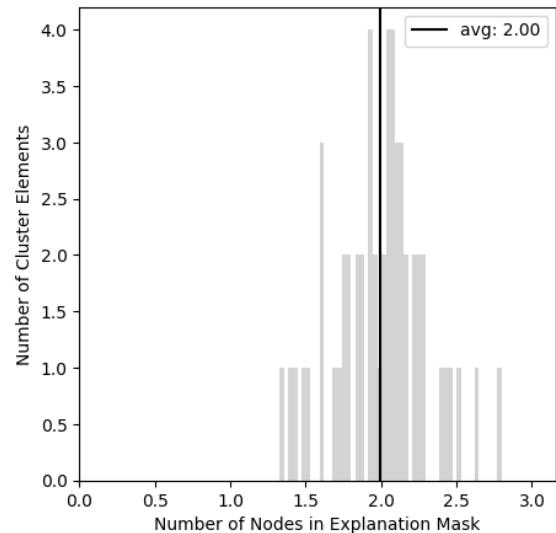
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

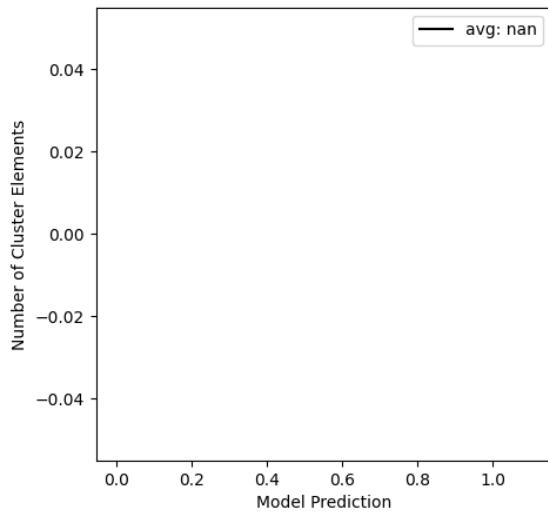
Prediction Impact Distribution



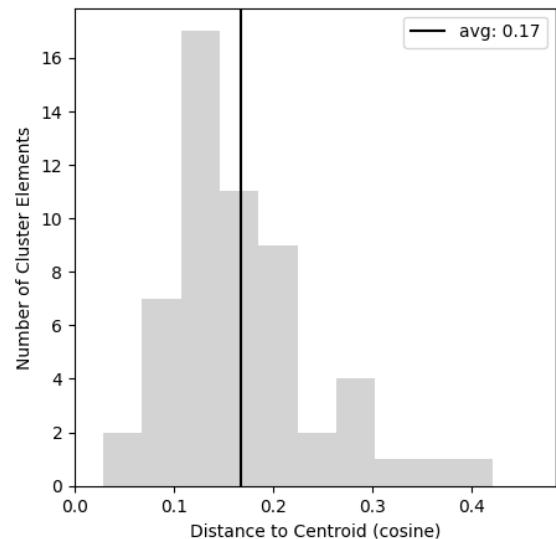
Mask Size Distribution



Prediction Output Distribution

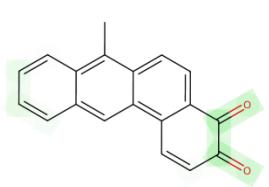
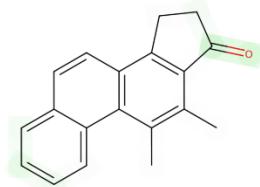
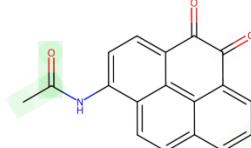
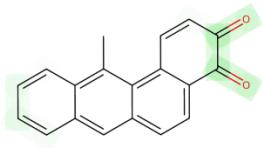
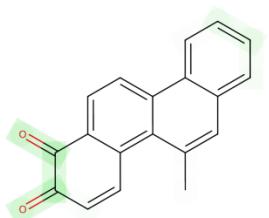
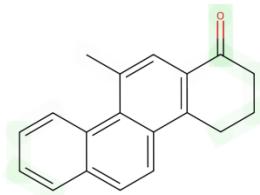
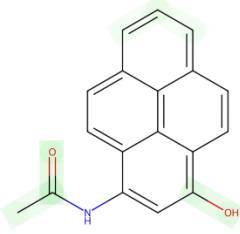
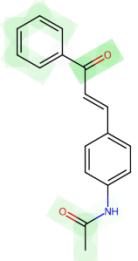
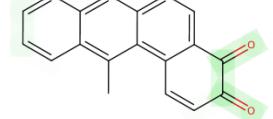
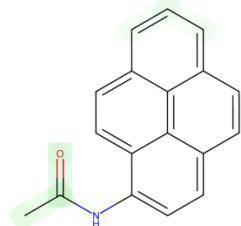
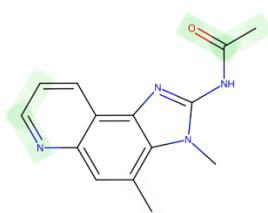
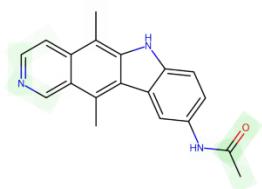
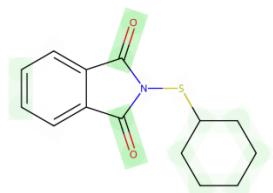
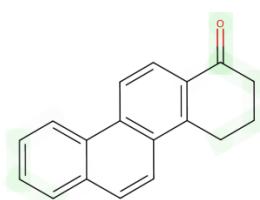
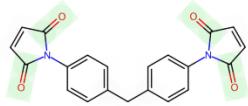
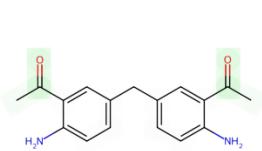


Distance to Centroid Distribution



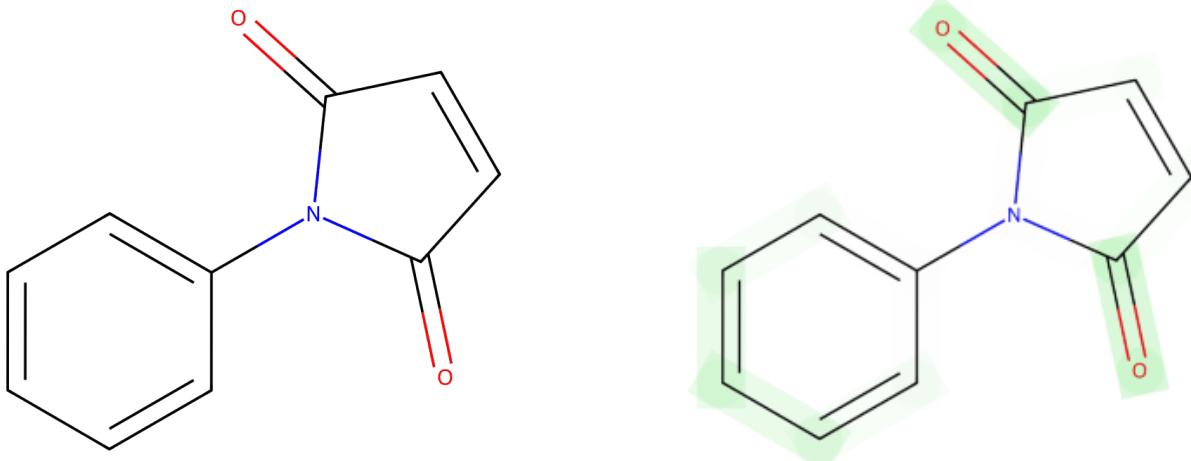
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure given by the SMILES representation "O=C1C=CC(=O)N1c1ccccc1" features a lactam ring (a cyclic amide) fused with a benzene ring. Lactams are known to have lower reactivity due to the resonance stabilization of the amide bond which disperses the electron density and makes the molecule less electrophilic. Additionally, the aromatic benzene ring is known for its stability due to delocalized pi electrons. The combined effect of these structural features may contribute to the molecule being less reactive with DNA, thereby exhibiting a lower tendency to cause mutations.

**Hypothesis:** Molecules containing the substructure represented by "O=C1C=CC(=O)N1c1ccccc1" are hypothesized to be non-mutagenic due to the inherent stability of the lactam ring and the benzene ring. The resonance stabilization in the lactam reduces its reactivity while the aromatic system in benzene provides stability by electron delocalization, both of which are likely to diminish the molecule's capability to interact with and cause mutations in genetic material.

# Cluster #1 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 1, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 8.7 ( $\pm 1.3$ ) on the prediction outcome.

## Properties

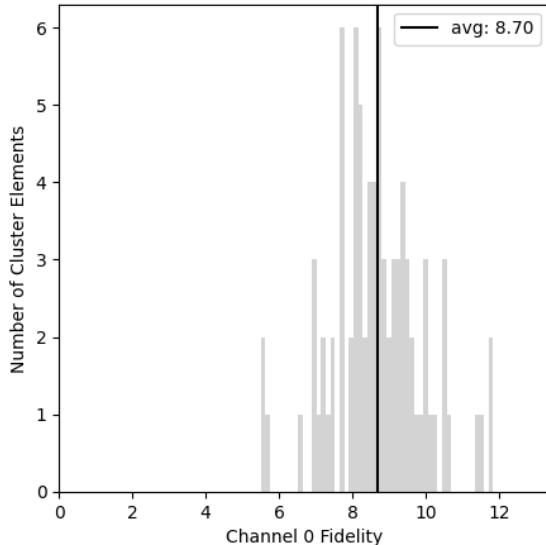
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	83
Channel Index	0.0 (0.0)

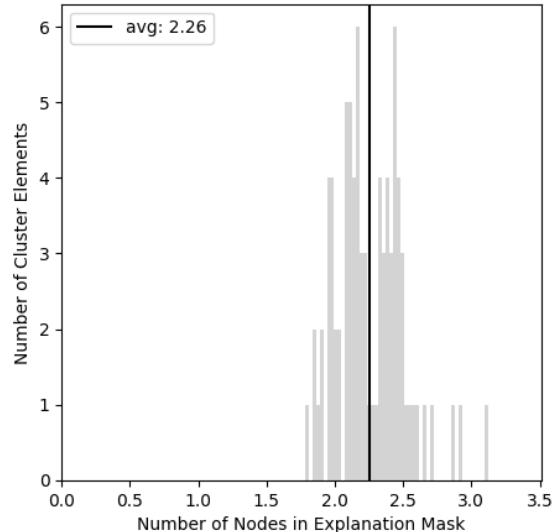
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

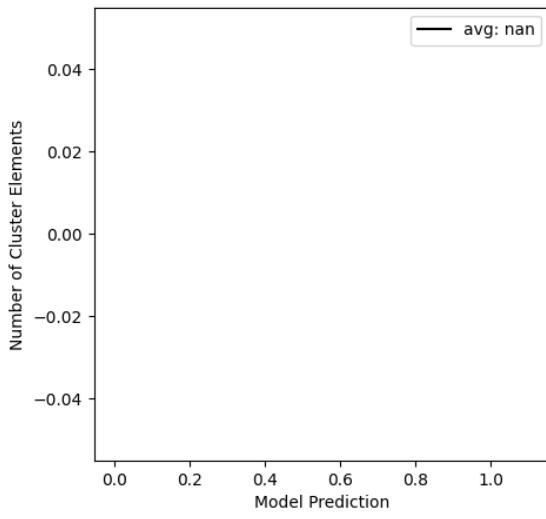
Prediction Impact Distribution



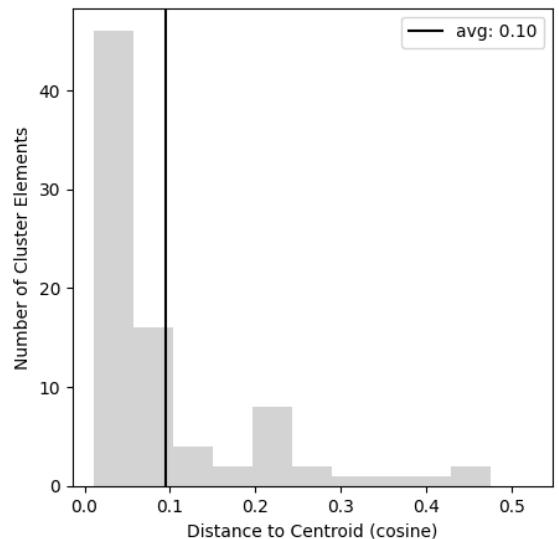
Mask Size Distribution



Prediction Output Distribution

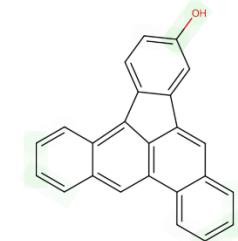
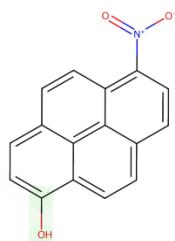
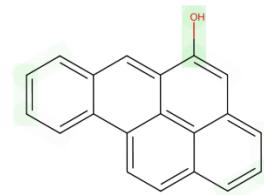
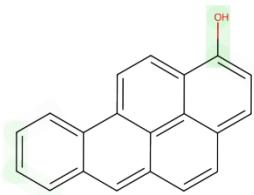
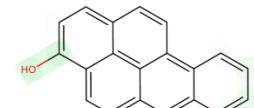
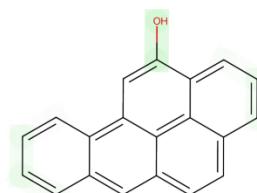
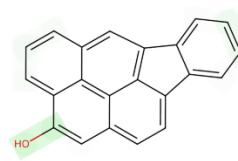
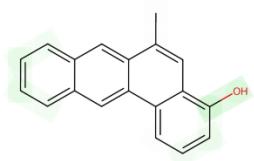
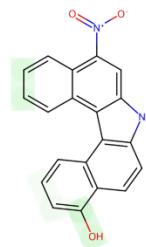
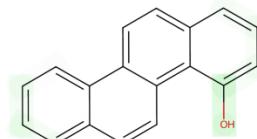
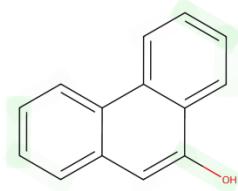
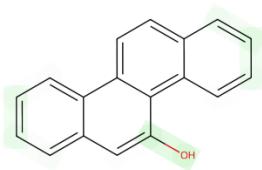
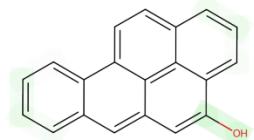
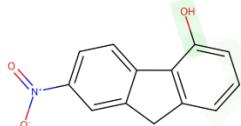
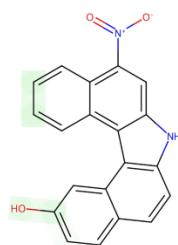
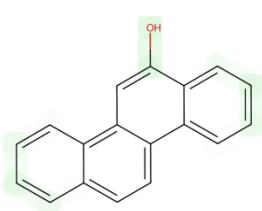


Distance to Centroid Distribution



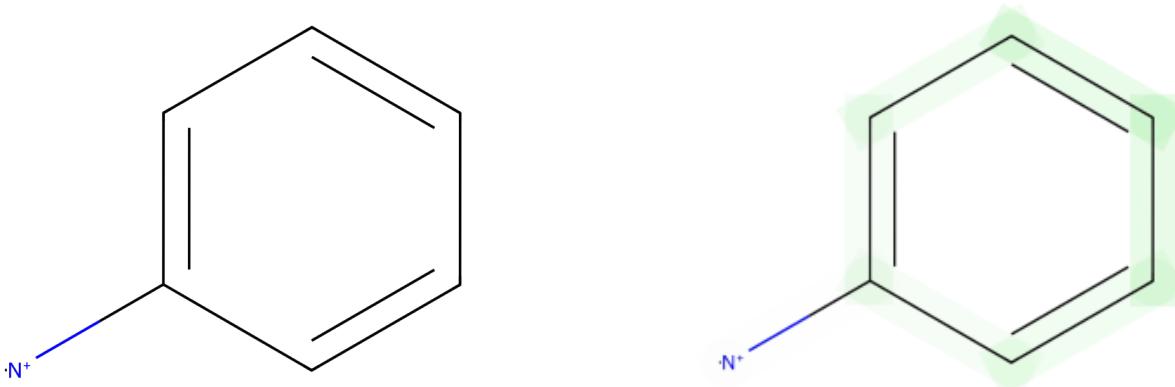
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The compound represented by the SMILES notation "[N+]-c1:c:c:c:c:1" includes a positively charged nitrogen atom (N+) attached to an aromatic benzene ring. Aromaticity typically confers stability to the molecular structure due to the delocalized electrons in the benzene ring. Additionally, the positive charge on the nitrogen atom may reduce the reactivity of the compound, since it will be seeking to acquire electrons to neutralize the charge rather than forming electrophilic interactions with DNA nucleotides, which could lead to mutagenicity.

**Hypothesis:** The presence of a positively charged nitrogen atom attached to a benzene ring is hypothesized to yield a structure with reduced mutagenic potential. This reduced mutagenicity may be due to the stabilizing effects of aromaticity and the electron-deficient nature of the charged nitrogen, which diminishes the compound's propensity to interact with and cause modifications to genetic material.

# Cluster #2 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 2, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 8.3 ( $\pm 1.4$ ) on the prediction outcome.

## Properties

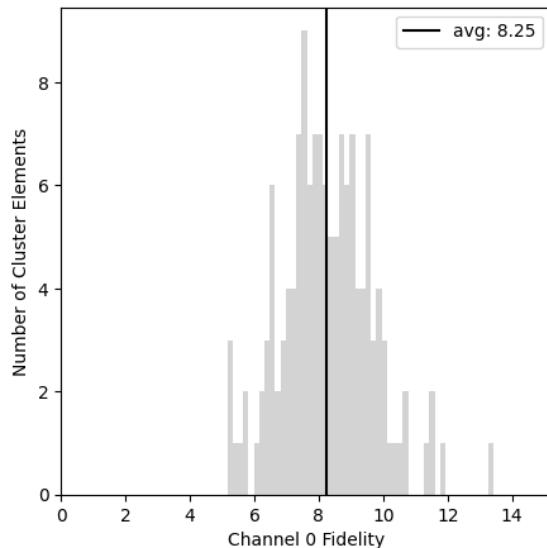
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	139
Channel Index	0.0 (0.0)

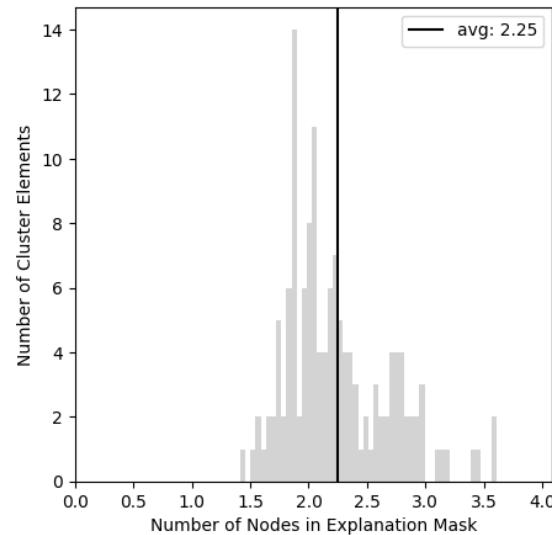
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

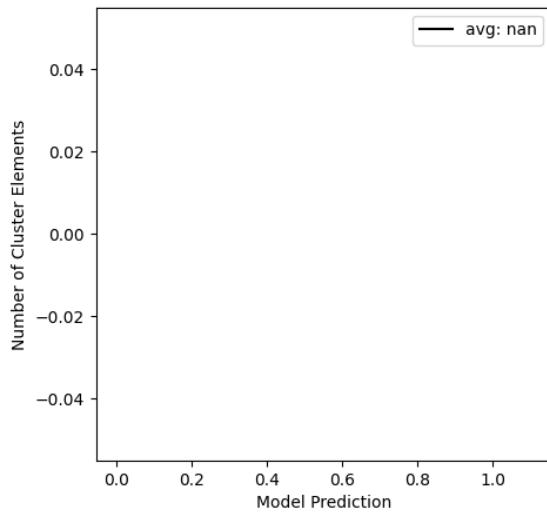
Prediction Impact Distribution



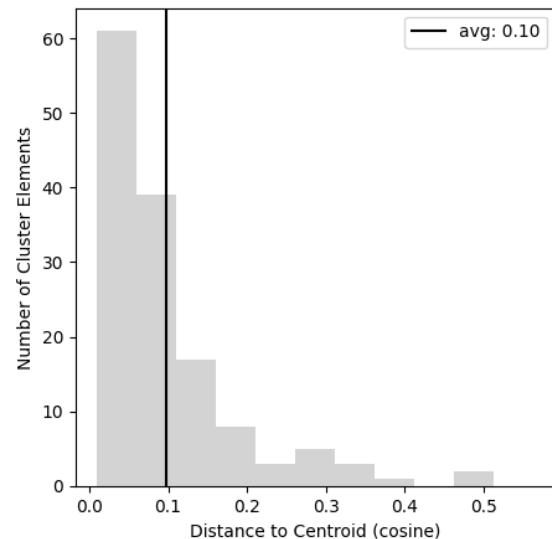
Mask Size Distribution



Prediction Output Distribution

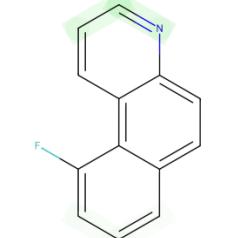
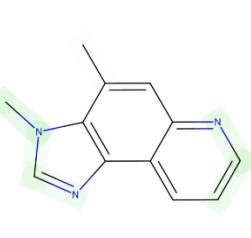
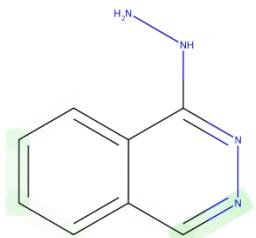
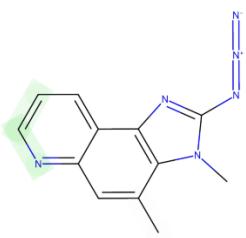
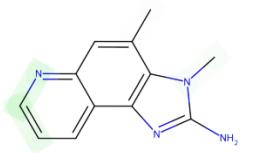
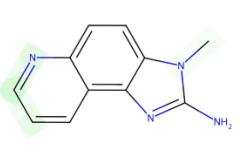
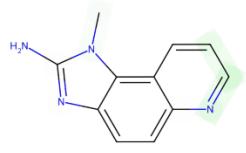
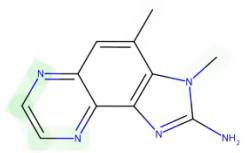
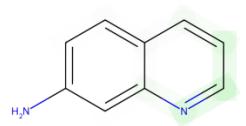
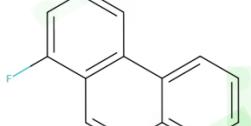
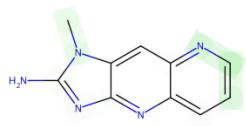
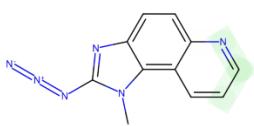
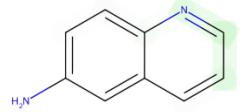
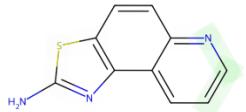
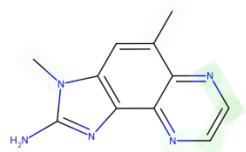
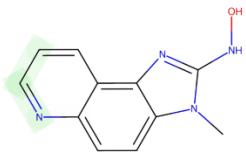


Distance to Centroid Distribution



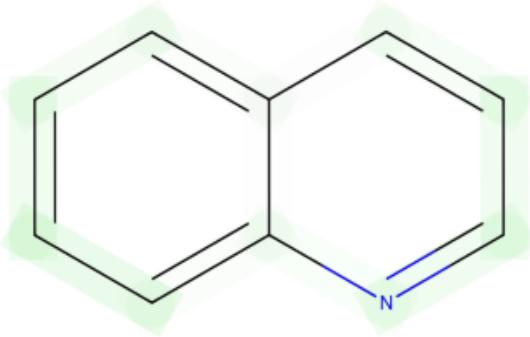
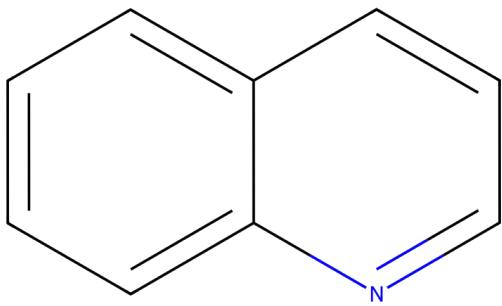
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation "c1:c:c:c:c2:n:c:c:c:c2:c:1" describes a polycyclic aromatic compound containing a nitrogen atom. The presence of aromatic systems generally suggests potential mutagenicity due to the ability of these systems to interact with DNA through intercalation or electron transfer. However, the inclusion of a nitrogen atom within the ring system can potentially disrupt the electron-rich nature of the aromatic system, reducing its ability to interact with DNA. Moreover, the nitrogen atom could also influence the spatial configuration of the molecule, which in turn might hinder its ability to cause mutations by not fitting properly into the DNA helix or not forming the necessary reactive metabolites.

**Hypothesis:** The presence of a nitrogen atom in a polycyclic aromatic system is hypothesized to confer a SMALL influence towards non-mutagenic properties. The nitrogen atom's electron-withdrawing effect may reduce the aromatic system's reactivity, and the alteration in the molecule's geometry could prevent effective interaction with the genetic material. Thus, while aromatic compounds are often associated with mutagenicity, this particular structure may be an exception due to its specific atomic composition and configuration.

# Cluster #3 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 3, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 7.7 ( $\pm 1.0$ ) on the prediction outcome.

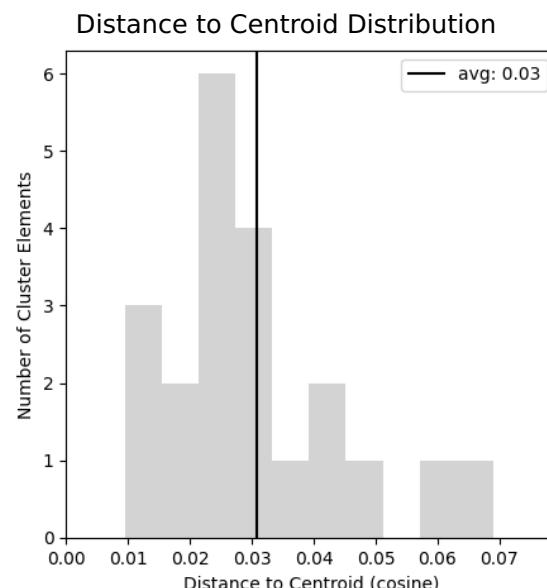
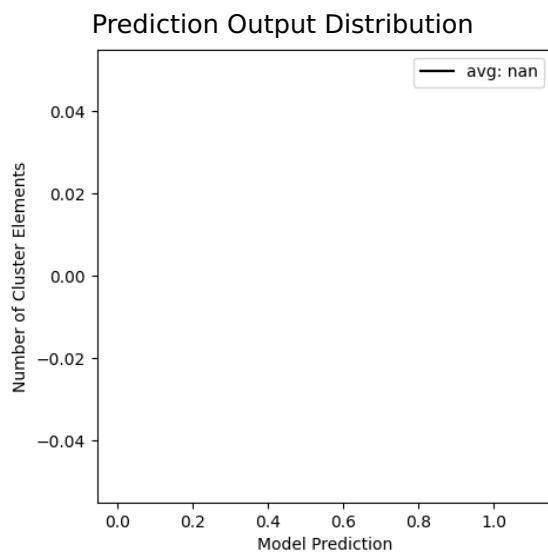
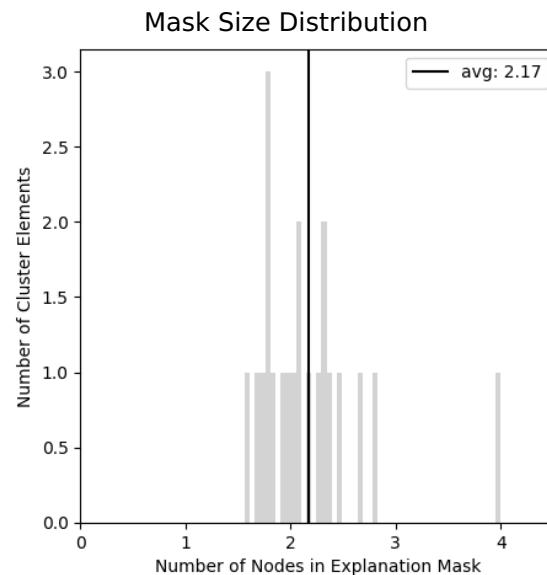
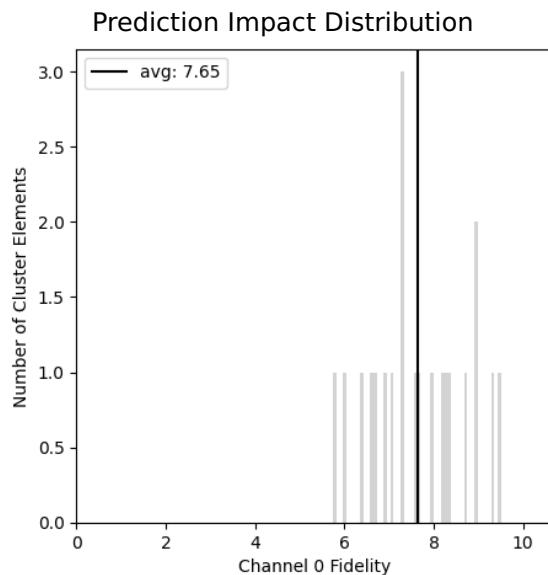
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	0.0 (0.0)

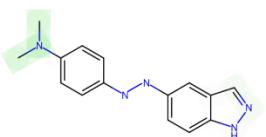
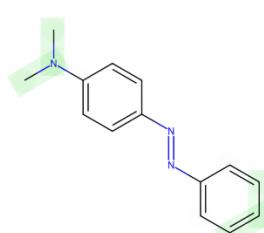
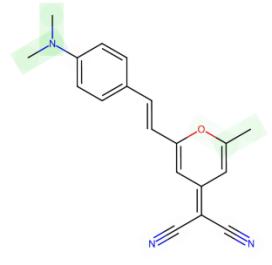
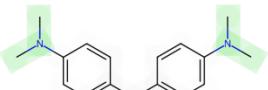
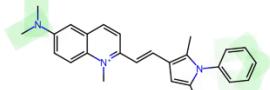
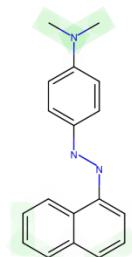
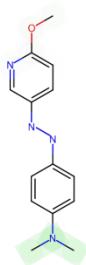
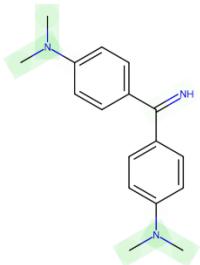
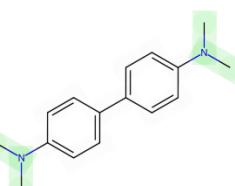
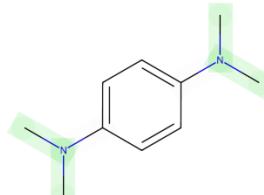
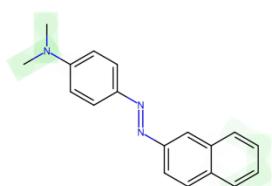
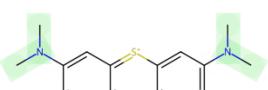
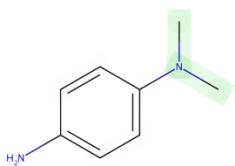
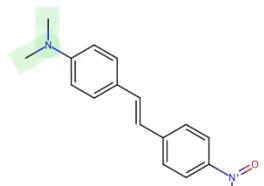
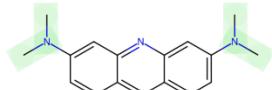
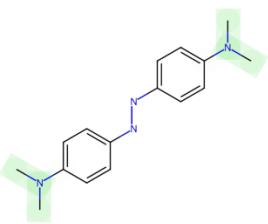
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the given SMILES "C=C-C(=C)-C=C-C" indicates a conjugated system with alternating single and double bonds. Conjugated systems are known for their stability due to the delocalization of electrons across the molecule, which can decrease the likelihood of the molecule reacting with DNA. The underlying hypothesis for this structure showing a **SMALL** influence towards being "non-mutagenic" is that the stability of the conjugated system may not provide the necessary reactivity to interact with and cause mutations in genetic material.

**Hypothesis:** The given conjugated molecular substructure is hypothesized to have a **SMALL** influence towards "non-mutagenic" properties due to electron delocalization imparting stability and reducing reactivity with DNA. This stability limits the molecule's potential to form DNA adducts or cause base pair substitutions, which are key steps in mutagenesis.

# Cluster #4 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 4, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 6.7 ( $\pm 1.1$ ) on the prediction outcome.

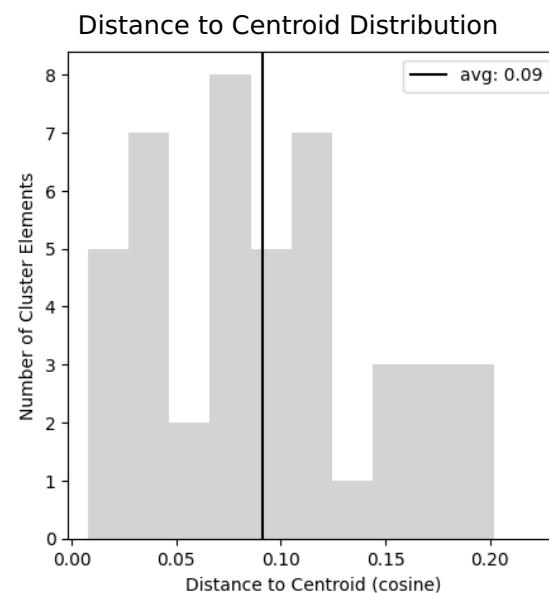
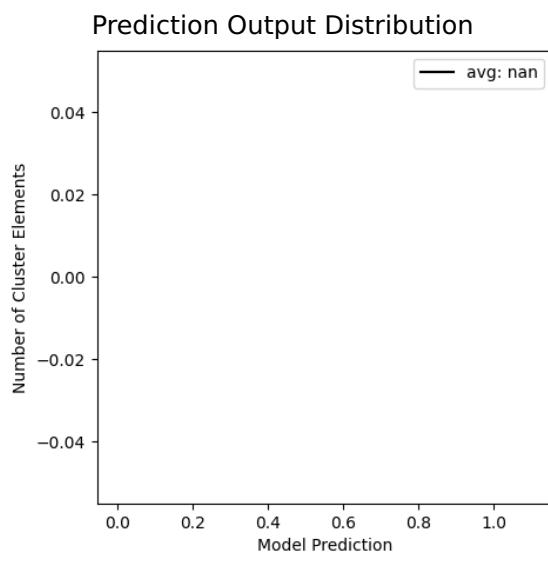
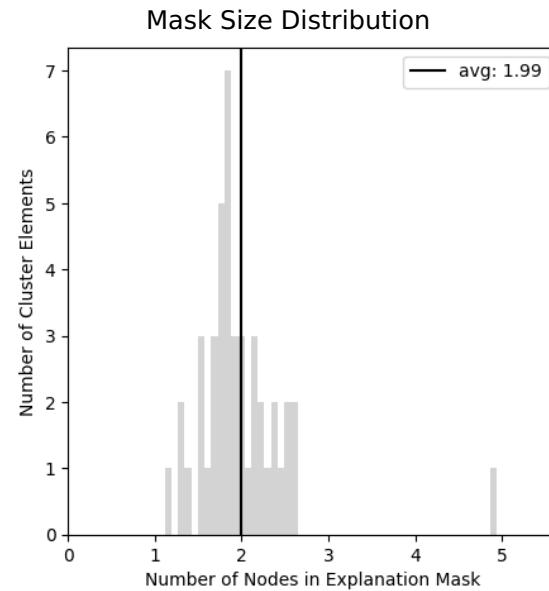
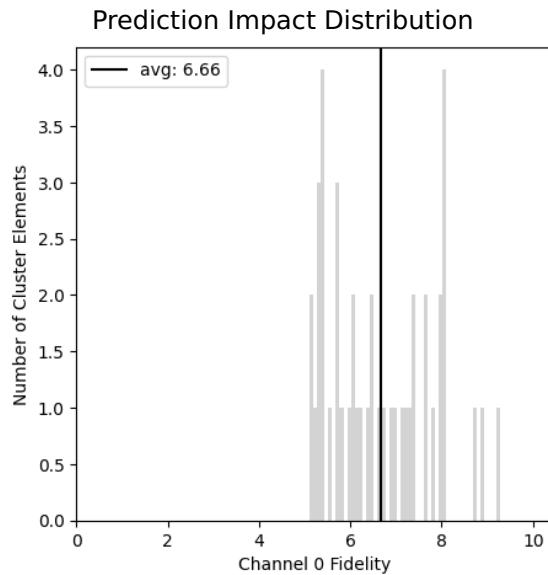
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	44
Channel Index	0.0 (0.0)

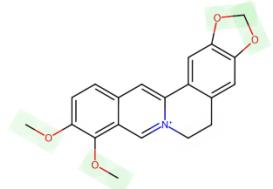
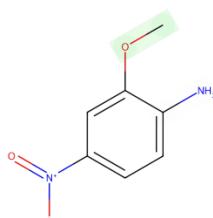
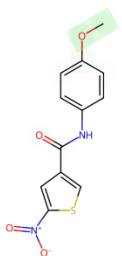
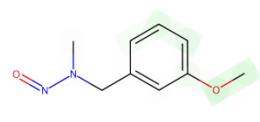
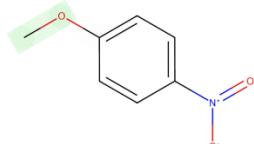
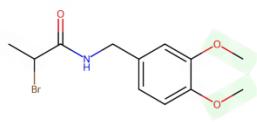
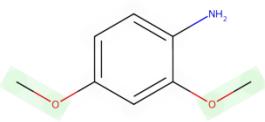
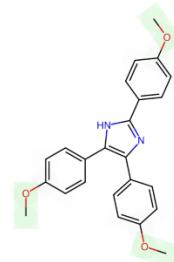
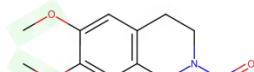
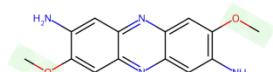
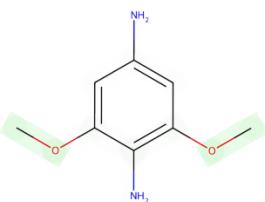
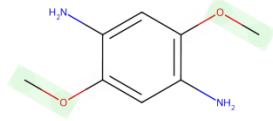
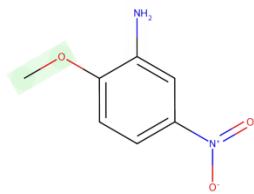
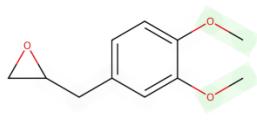
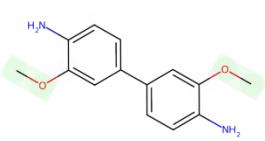
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



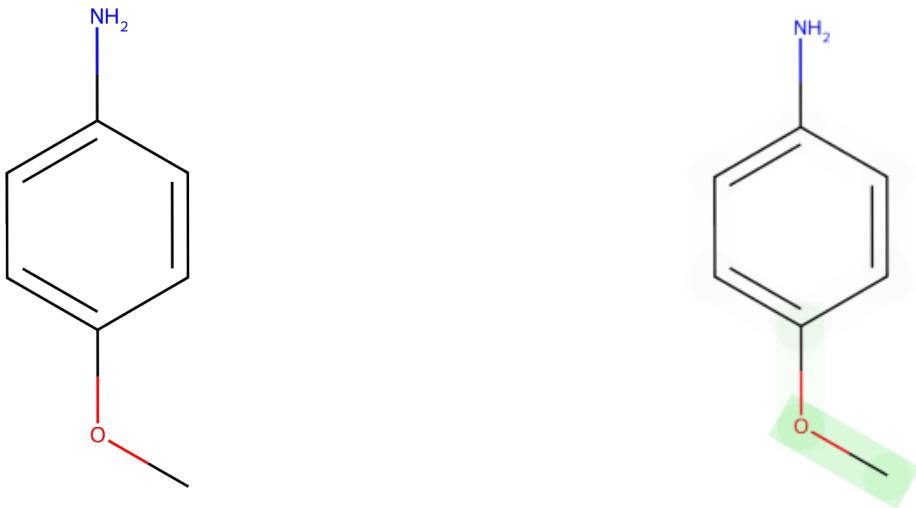
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a phenol-like aromatic system with a methoxy group (-OCH<sub>3</sub>) and an amino group (-NH<sub>2</sub>) substituted onto a benzene ring. Both methoxy and amino groups are electron-donating substituents, which can stabilize the benzene ring through resonance, and reduce the reactivity of the ring towards electrophilic aromatic substitution - a common mechanism through which some chemical mutagens act. Furthermore, the presence of the hydroxyl group provides hydrogen bonding capabilities, potentially increasing water solubility and facilitating excretion from the body before any potential DNA damage occurs.

**Hypothesis:** The structure "C-O-c1:c:c:c(-N):c:c:1" is hypothesized to exhibit non-mutagenic properties due to the electron-donating nature of the methoxy and amino groups and the potential for increased water solubility from the hydroxyl group. These characteristics may reduce the ability of the molecule to interact with DNA and participate in mutagenic reactions.

# Cluster #5 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 5, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.8$ ) nodes. The concept is generally associated with an impact of 8.8 ( $\pm 2.1$ ) on the prediction outcome.

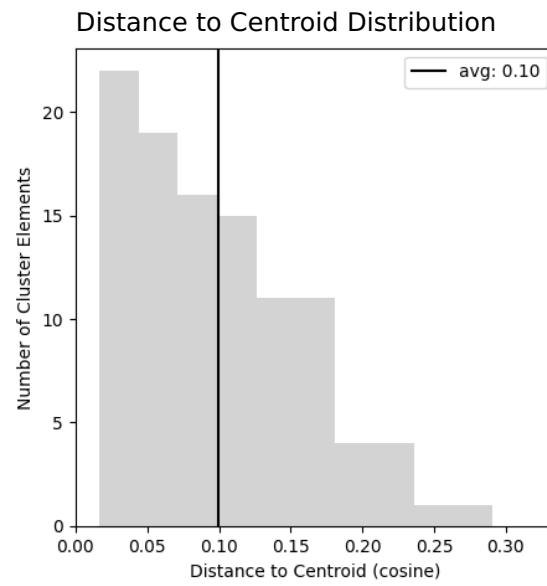
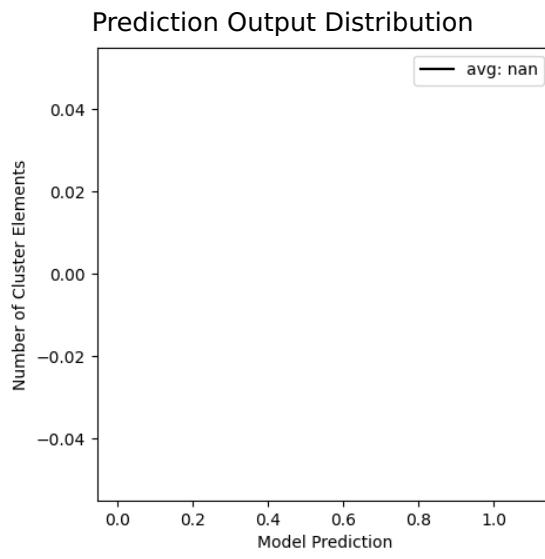
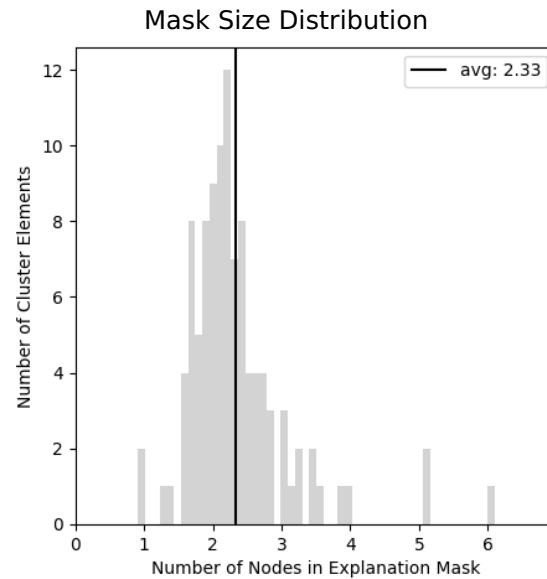
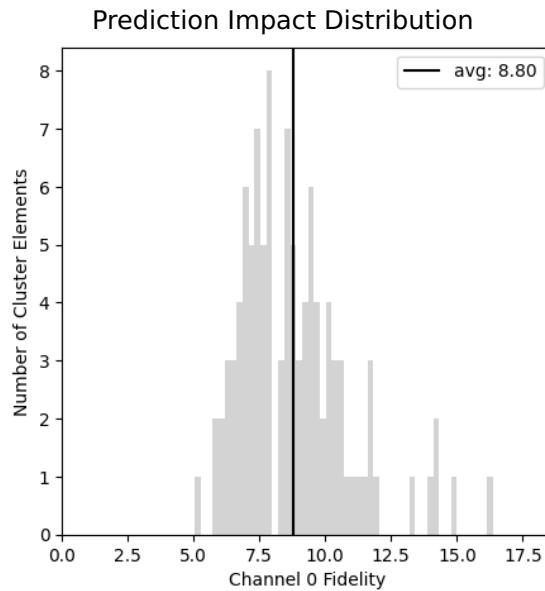
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	104
Channel Index	0.0 (0.0)

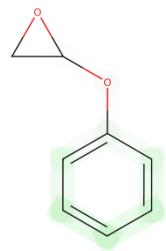
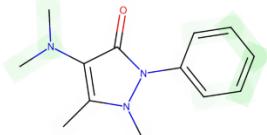
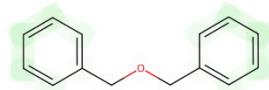
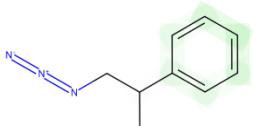
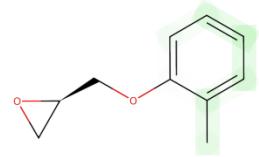
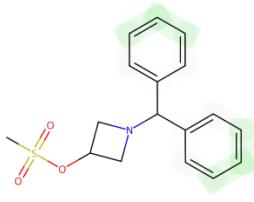
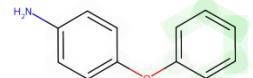
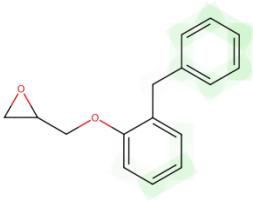
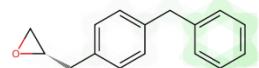
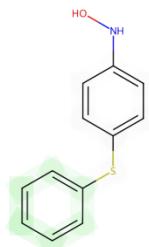
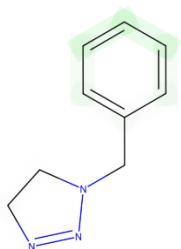
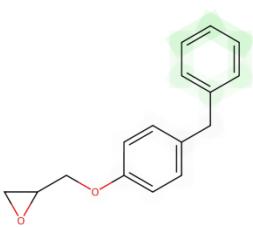
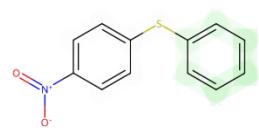
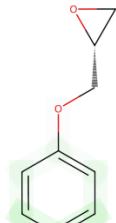
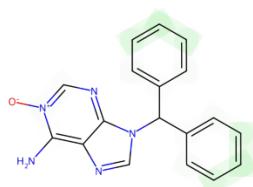
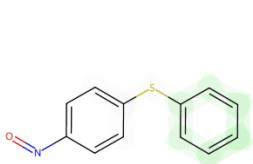
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



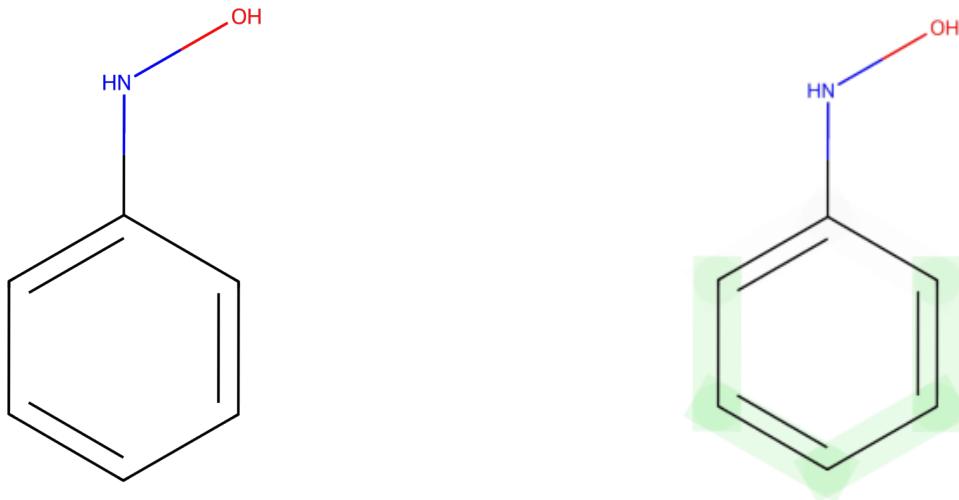
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES represents a molecular structure with an aromatic ring (benzene ring) attached to a nitroso group (O-N-) which is a well-known functional group. Typically, aromatic rings are considered to have certain stable resonance structures that enable them to absorb and redistribute energy throughout the ring, potentially limiting the reactivity of the molecule with DNA. The nitroso group, however, can be a cause for concern since it's known to form reactive species that could potentially damage DNA under certain conditions. Nevertheless, the term "SMALL influence towards non-mutagenic" indicates that the molecule's overall structure might somehow inhibit the nitroso group's reactivity or interfere with the interaction of the nitroso group with genetic material.

**Hypothesis:** Molecules with the substructure "O-N-c1:c:c:c:c:1" tend towards being non-mutagenic due to the potential stabilization of the nitroso group by the adjacent aromatic ring. It is hypothesized that the delocalization of electrons within the aromatic ring may reduce the reactivity of the nitroso group, or the steric environment created by the ring could hinder the interaction of the nitroso group with DNA.

# Cluster #6 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 6, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 7.4 ( $\pm 1.3$ ) on the prediction outcome.

## Properties

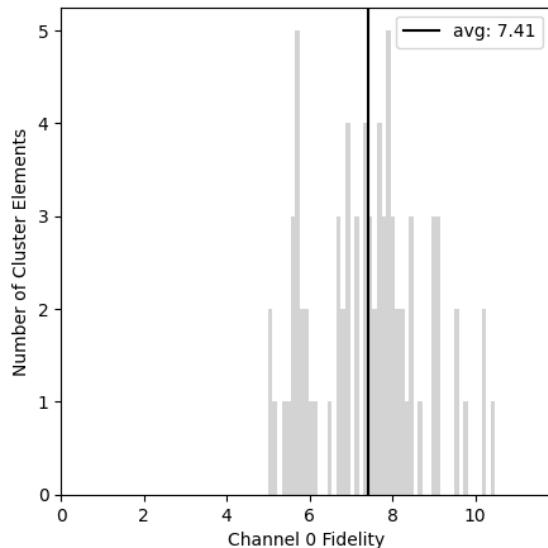
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	77
Channel Index	0.0 (0.0)

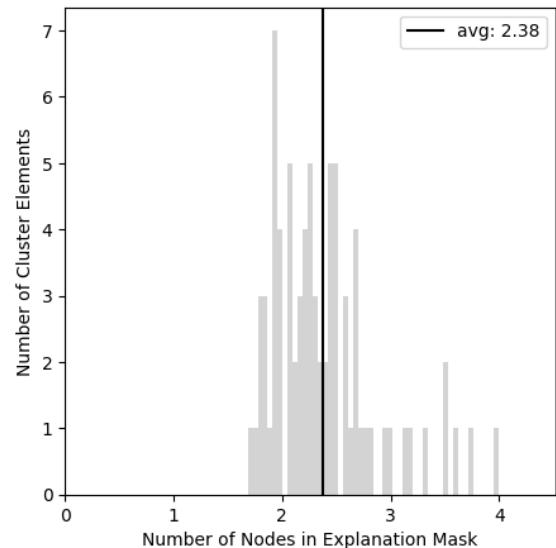
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

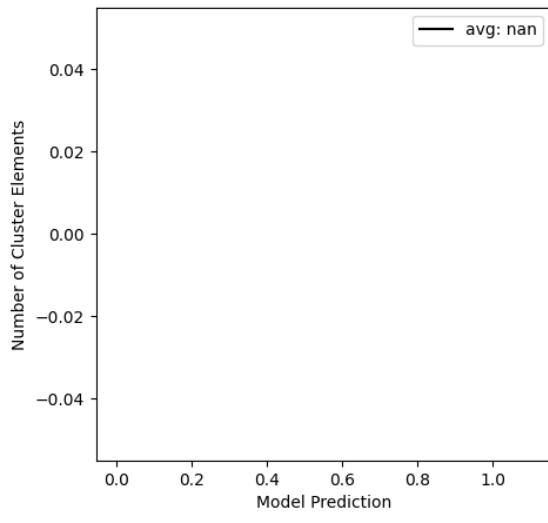
Prediction Impact Distribution



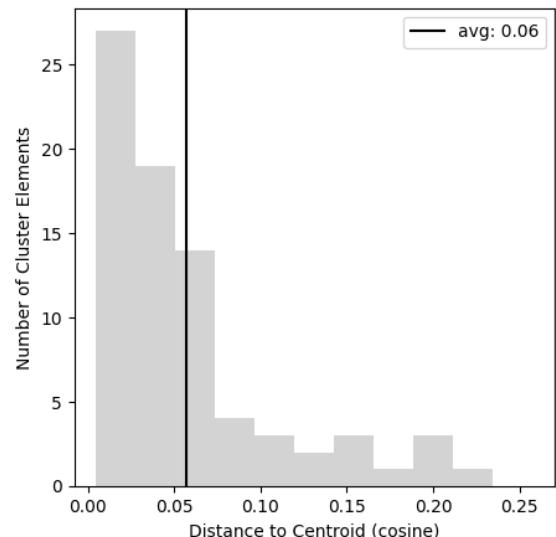
Mask Size Distribution



Prediction Output Distribution

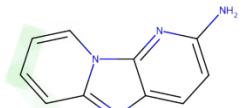
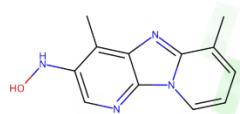
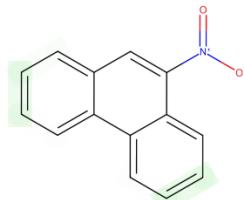
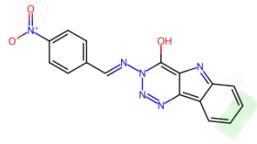
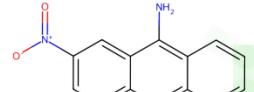
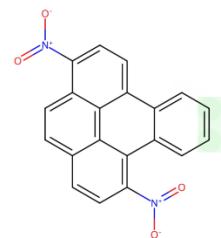
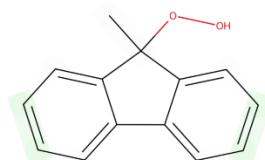
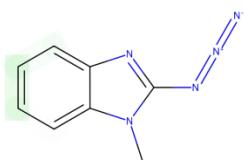
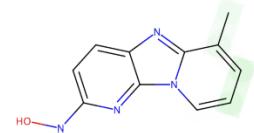
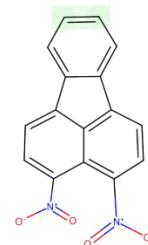
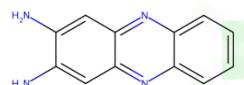
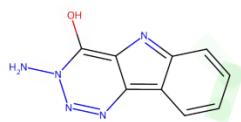
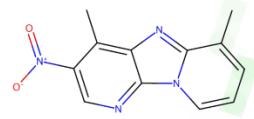
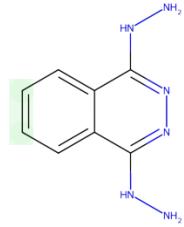
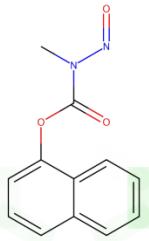
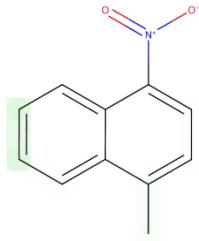


Distance to Centroid Distribution



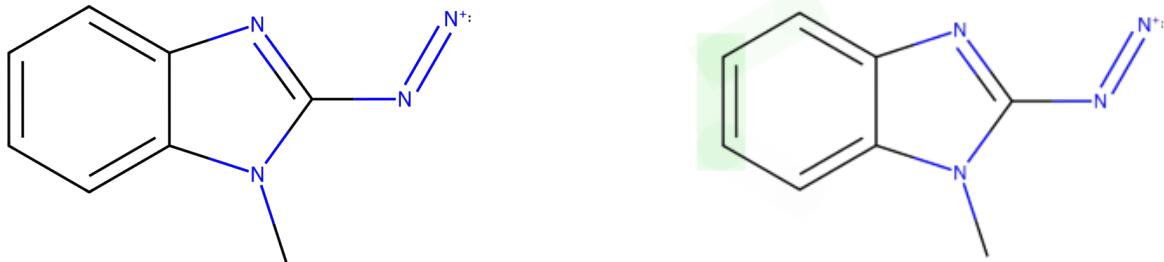
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a compound with an aromatic azo group ( $-N=N-$ ) connecting two aromatic rings with a cyano group ( $C\equiv N$ ) attached. Aromatic azo compounds are known to exhibit varying degrees of mutagenicity depending on their specific structure. However, in this structure, the azo group is adjacent to a cyano group, which is an electron-withdrawing group. This could lead to electron delocalization in the aromatic rings and the azo bond, potentially reducing the reactivity of the azo group. This reduction in reactivity could lower the chances of it forming covalent bonds with DNA, which is a key step in the process of mutagenesis. The presence of aromatic rings can contribute to planarity, making it easier for the molecule to slip between the DNA base pairs. Yet, the potential stabilization by the cyano group could still minimize its reactivity, thereby influencing its mutagenicity to a lesser extent.

**Hypothesis:** The molecular fragment "C-n1:c(-N=[N+]):n:c2:c:c:c:c:2:1" is associated with a relatively low mutagenic potential. The presence of an electron-withdrawing cyano group adjacent to the azo linkage could be responsible for stabilizing the molecule and reducing its reactivity with DNA. The implication is that this structural configuration could lead to decreased ability of the molecule to cause genetic mutations, despite the typical concerns associated with aromatic azo compounds.

# Cluster #7 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 7, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 9.9 ( $\pm 1.3$ ) on the prediction outcome.

## Properties

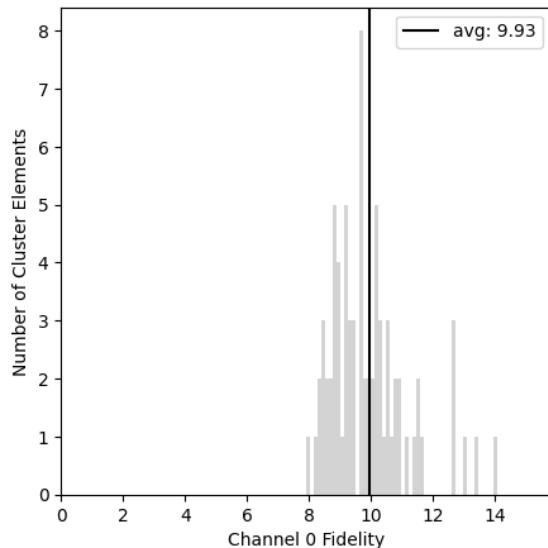
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	74
Channel Index	0.0 (0.0)

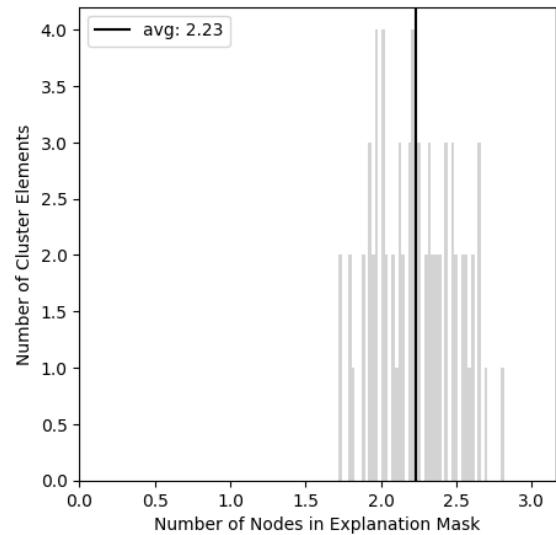
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

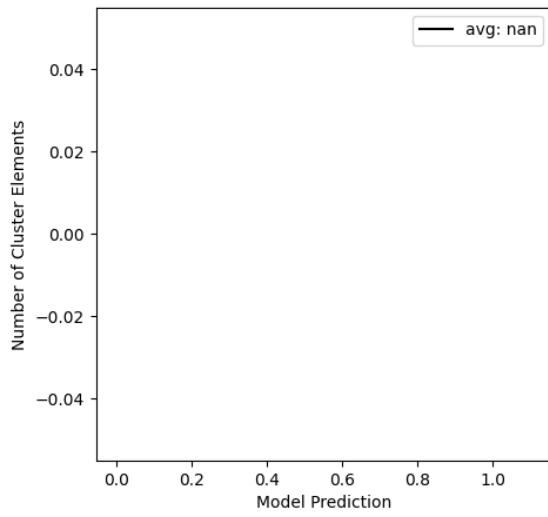
Prediction Impact Distribution



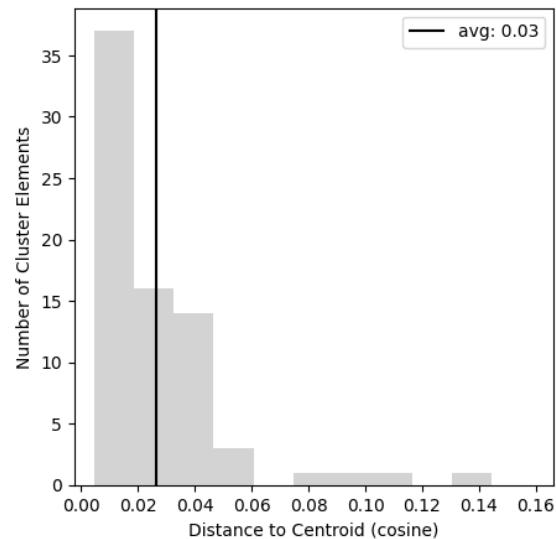
Mask Size Distribution



Prediction Output Distribution

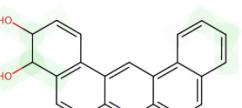
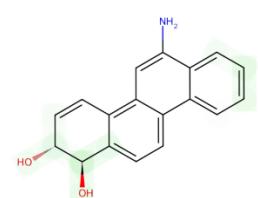
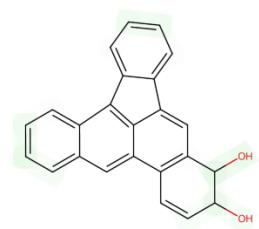
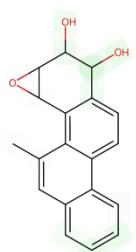
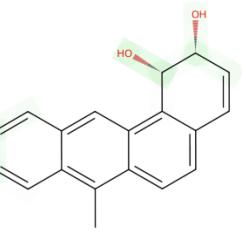
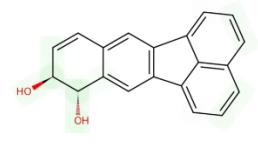
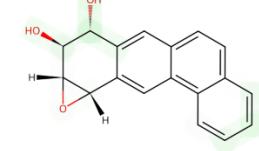
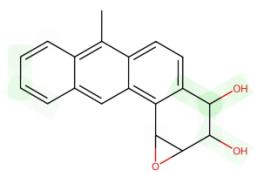
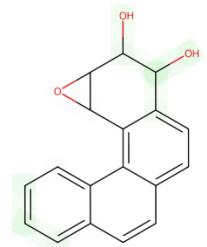
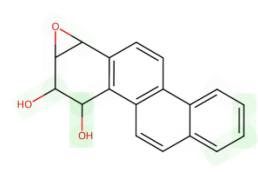
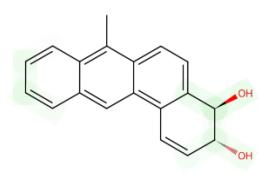
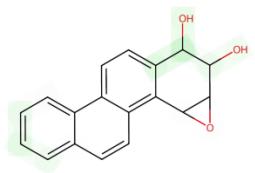
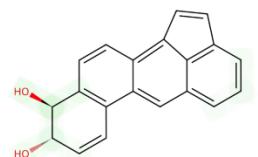
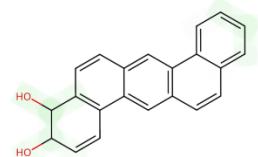
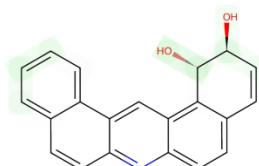
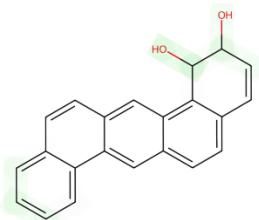


Distance to Centroid Distribution



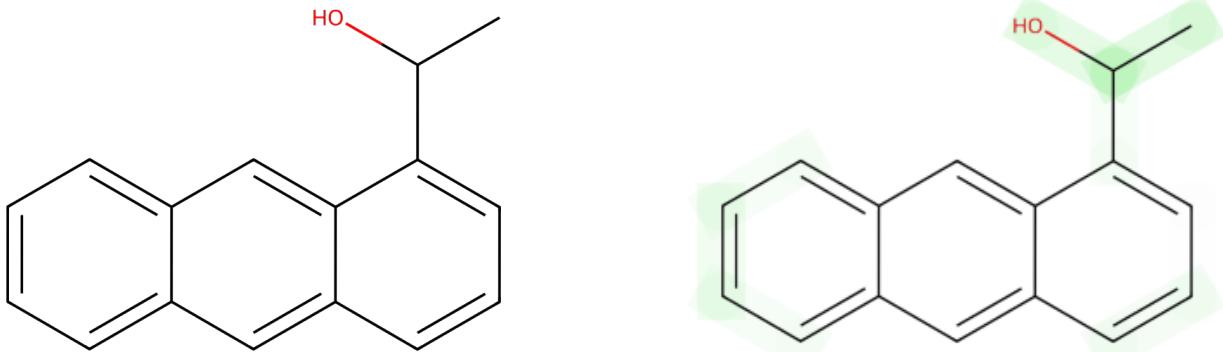
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a phenolic compound with a hydroxyl group (-OH) attached to a benzene ring that is connected to another benzene ring through a methylene bridge (-CH<sub>2</sub>-). Phenolic compounds often exhibit antioxidant properties due to their ability to donate hydrogen atoms from their hydroxyl groups, which can neutralize free radicals. Free radicals are known to cause DNA damage, leading to mutations. Therefore, the presence of phenolic substructures can provide a protective effect against DNA damage, resulting in a reduced mutagenic potential.

**Hypothesis:** Molecules containing the phenolic substructure represented by the given SMILES notation are likely to exhibit reduced mutagenicity due to their antioxidant capabilities. The hydroxyl group can donate a hydrogen atom, reducing the reactivity of free radicals, while the stabilizing effect of the aromatic system provides a structural context where these interactions can occur effectively.

# Cluster #8 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 8, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 10.7 ( $\pm 1.6$ ) on the prediction outcome.

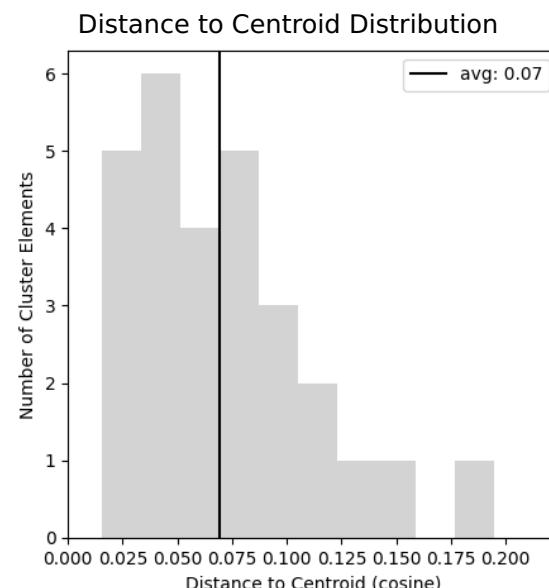
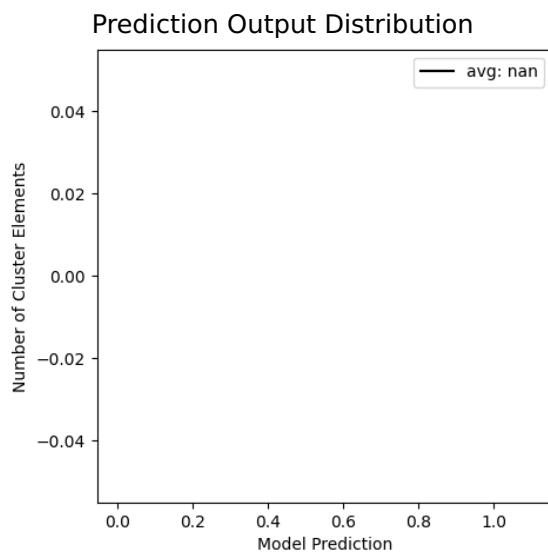
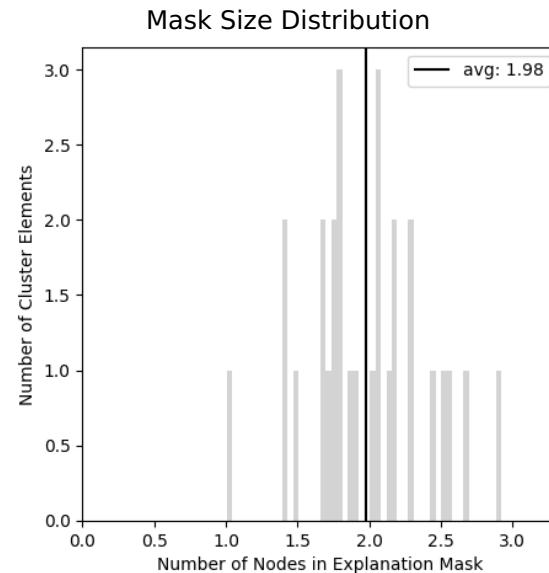
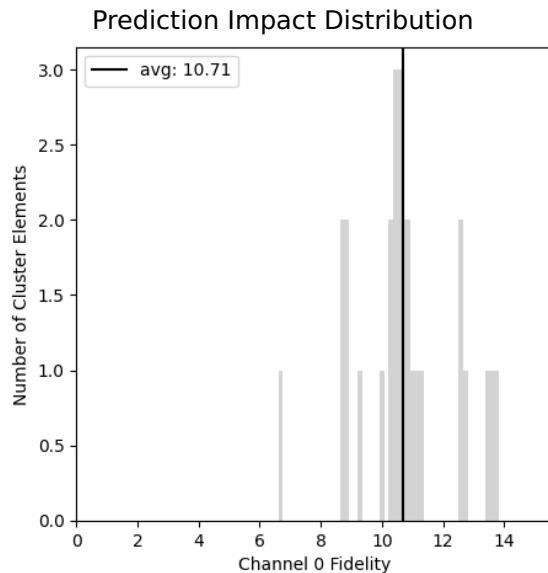
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	28
Channel Index	0.0 (0.0)

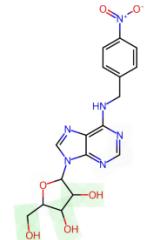
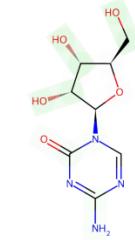
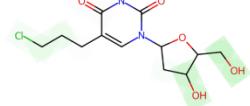
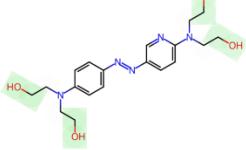
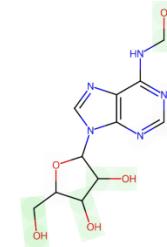
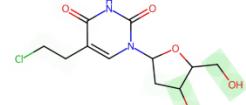
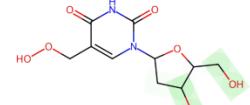
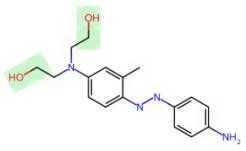
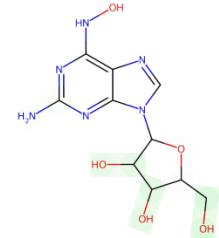
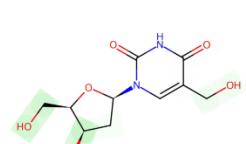
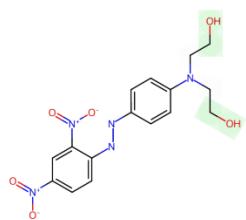
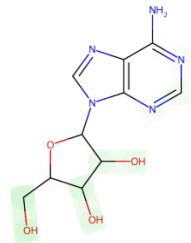
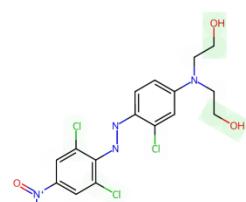
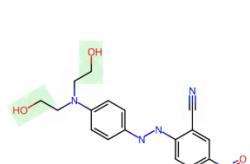
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



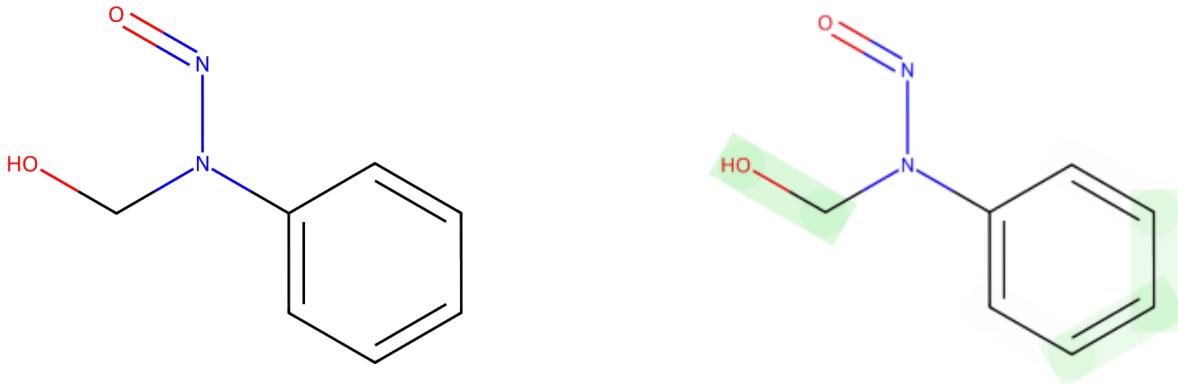
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure contains a nitroso group ( $\text{O}=\text{N}-$ ) and a nitro group ( $-\text{N}(\text{-C-O})-$ ), both of which are attached to a benzene ring through a nitrogen atom. Nitroso groups are known to be pro-mutagenic due to their ability to form DNA adducts, whereas nitro groups can undergo biotransformation to form reactive intermediates such as nitroso derivatives. However, when these groups are attached to a benzene ring, the stability of the ring can hinder the groups' reactivity, making the whole compound less likely to interact with DNA. Additionally, the presence of an oxygen atom bonded to a carbon atom ( $-\text{C-O}$ ) might suggest some electron-withdrawing effects which could reduce the reactivity of nitroso and nitro groups, thus lowering the mutagenicity.

**Hypothesis:** The molecular fragment " $\text{O}=\text{N}-\text{N}(\text{-C-O})-\text{c1:c:c:c:c:1}$ " is less likely to be mutagenic. The stabilizing influence of the benzene ring coupled with potential electron-withdrawing effects from the oxygen-carbon bond might attenuate the reactivity of nitroso and nitro groups, thus reducing the propensity of the molecule to cause genetic modifications.

# Cluster #9 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 9, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 12.1 ( $\pm 1.3$ ) on the prediction outcome.

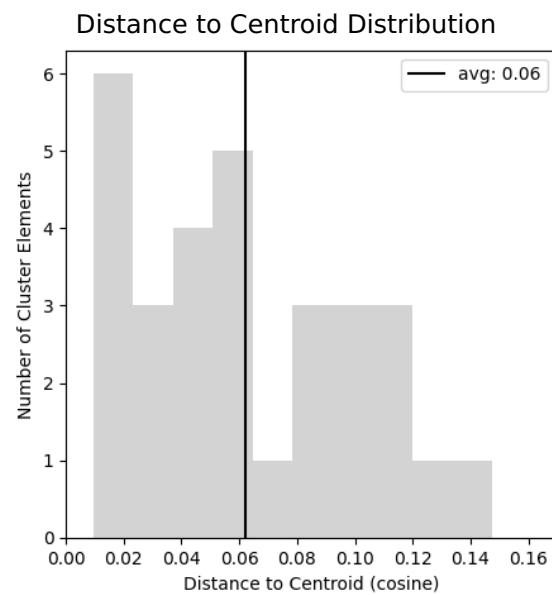
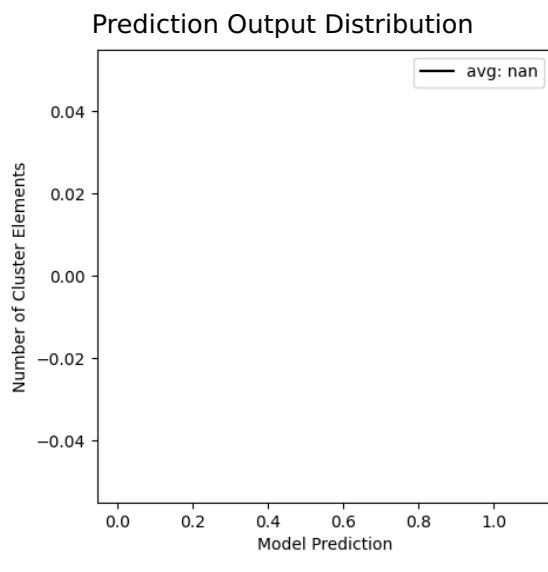
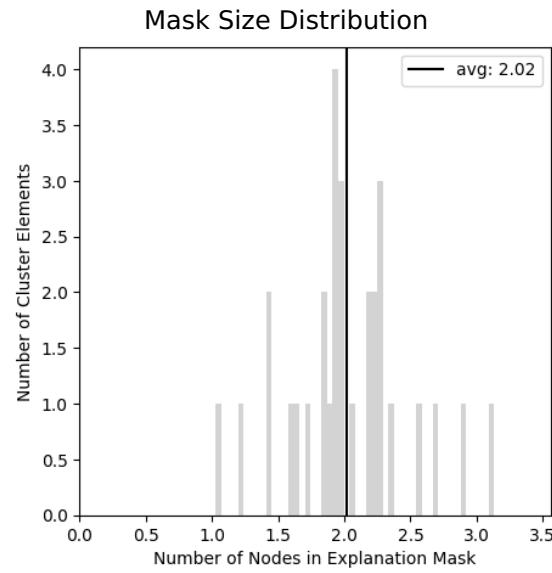
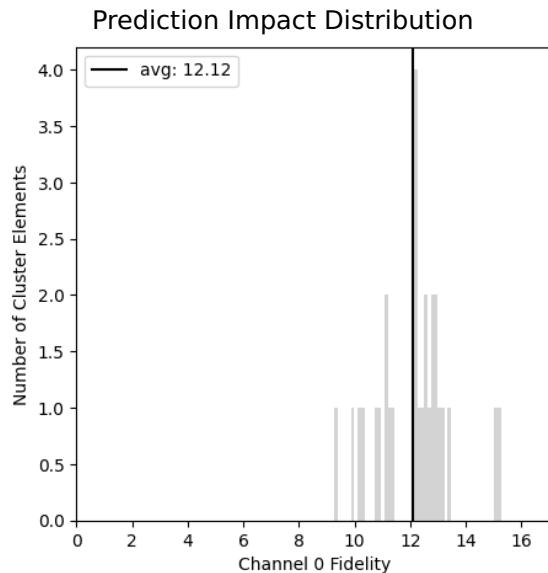
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	30
Channel Index	0.0 (0.0)

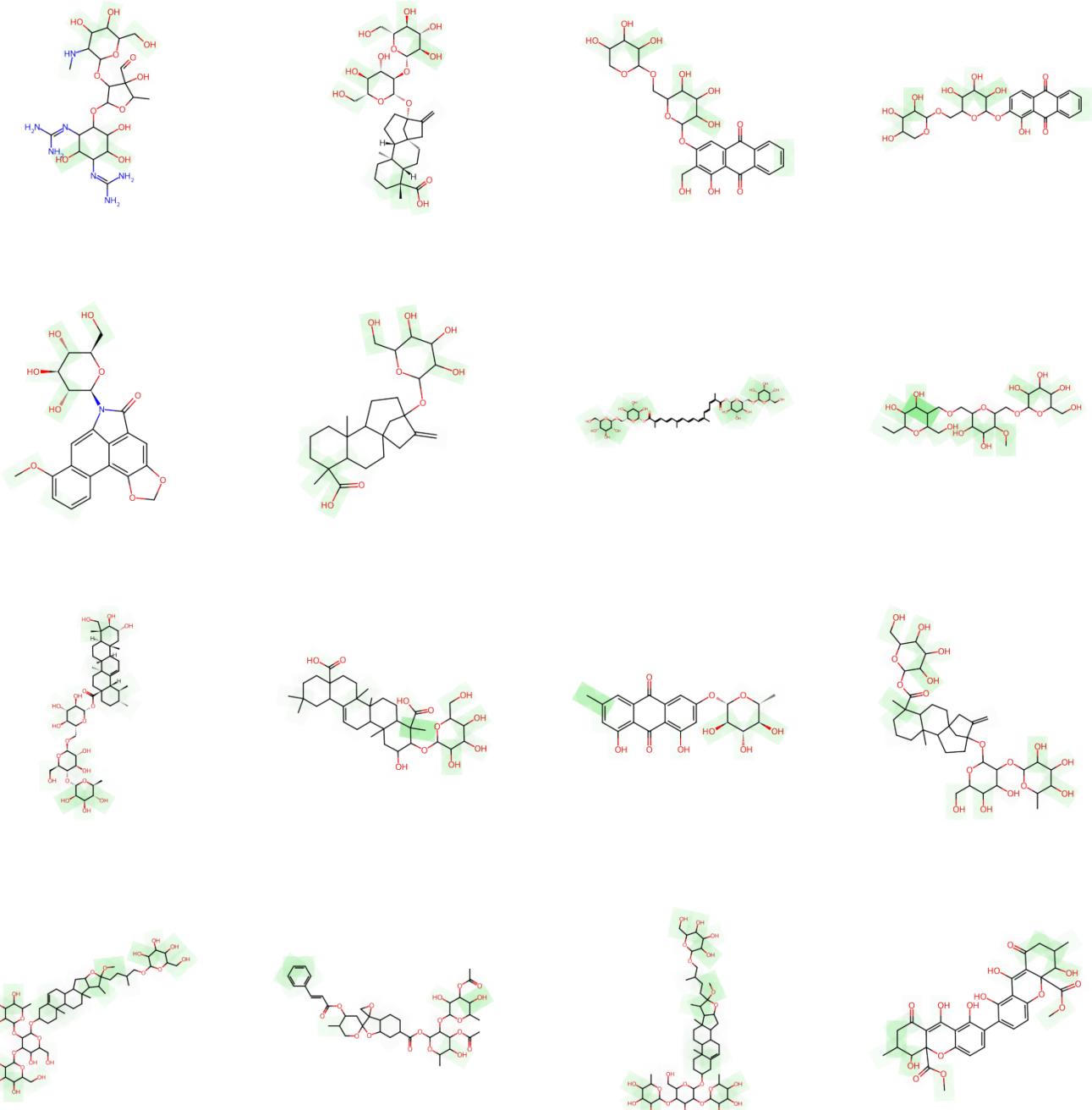
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



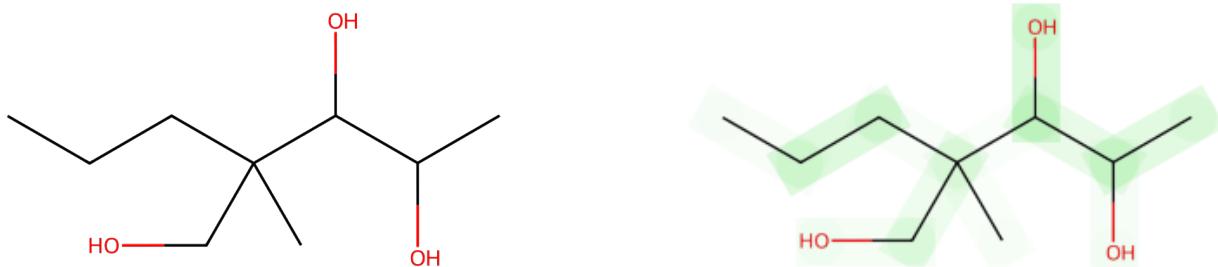
## Example Elements

i) This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure contains a long hydrocarbon chain with substitution of alkyl groups and hydroxyl groups (-OH). Hydroxyl groups are known to increase water solubility, which might reduce a compound's ability to penetrate cellular membranes and interact with DNA. Furthermore, the lack of highly reactive functional groups, like aromatic amines or nitro groups which are known for their mutagenic potential, suggests that the molecule may not readily form DNA adducts or induce mutations.

**Hypothesis:** The molecular fragment "C-C-C-C(-C)(-C-O)-C(-O)-C(-C)-O" is hypothesized to have a medium influence towards being non-mutagenic. The presence of hydroxyl groups may enhance solubility and reduce the ability of the molecule to interact with DNA, while the absence of highly reactive mutagenic functional groups decreases the likelihood of it causing genetic modifications.

# Cluster #10 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 10, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.9 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 15.8 ( $\pm 1.3$ ) on the prediction outcome.

## Properties

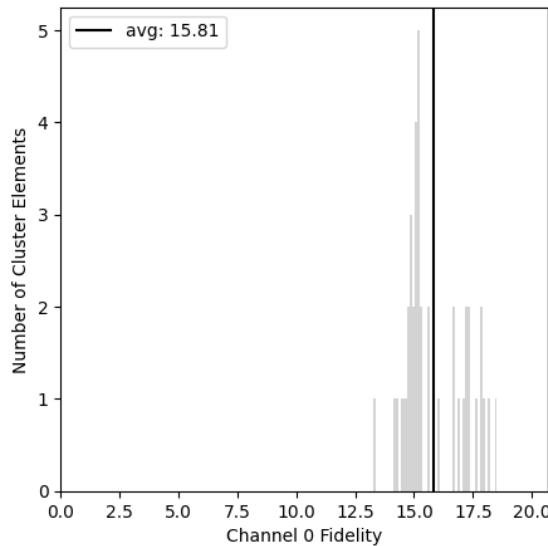
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	41
Channel Index	0.0 (0.0)

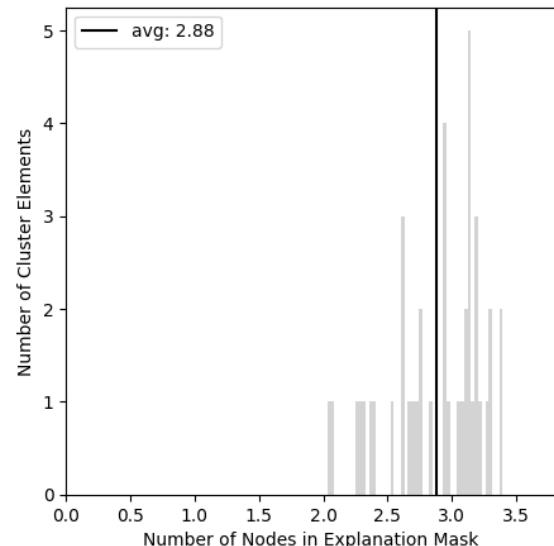
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

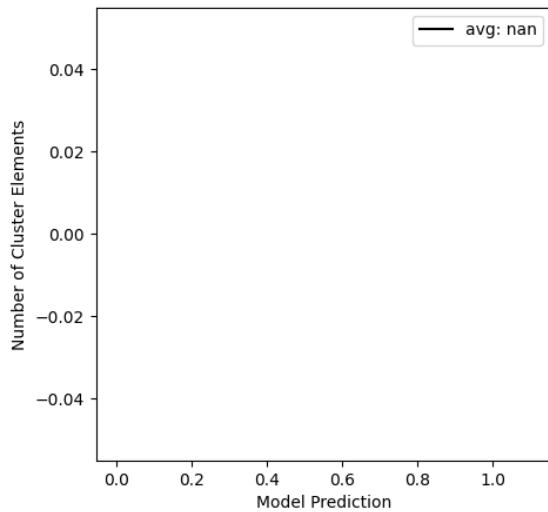
Prediction Impact Distribution



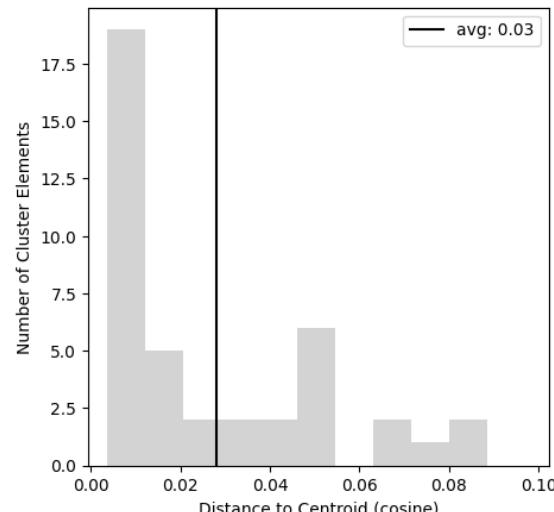
Mask Size Distribution



Prediction Output Distribution

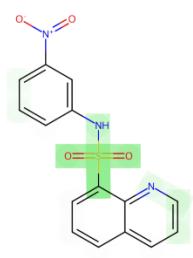
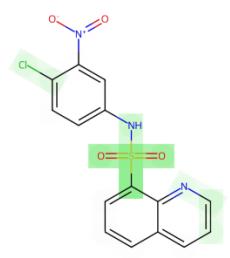
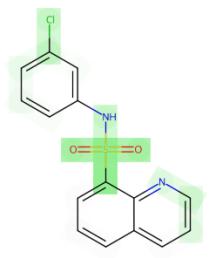
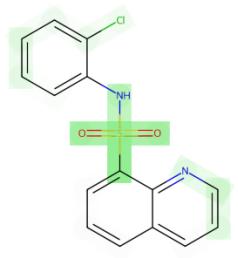
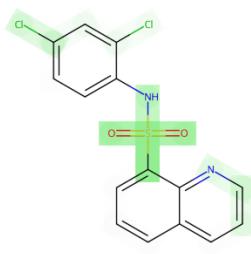
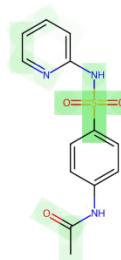
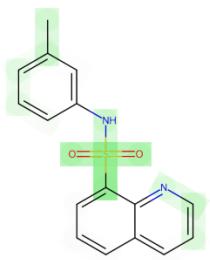
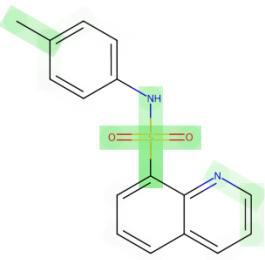
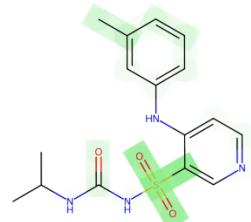
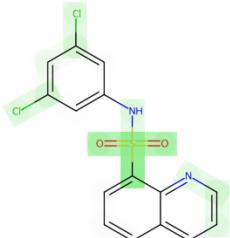
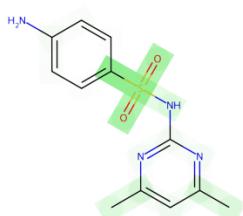
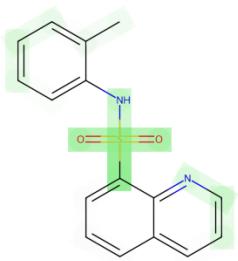
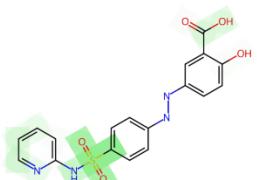
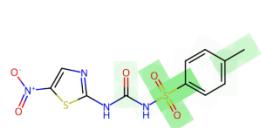
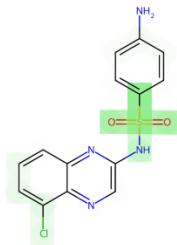
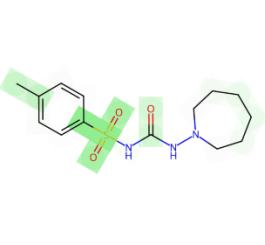


Distance to Centroid Distribution



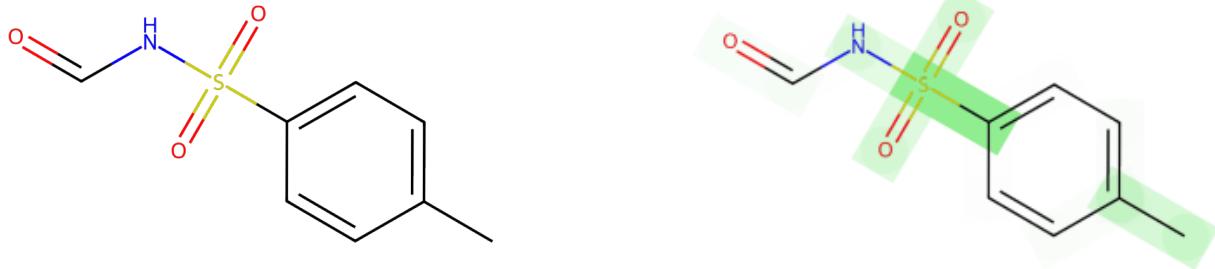
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure in question features a thiophene ring (sulfur-containing aromatic compound) with an attached sulfonamide group ( $-S(=O)(=O)-N-$ ) and an aldehyde group ( $-C=O$ ). The presence of the sulfonyl group ( $-SO_2-$ ) next to the amine ( $-NH-$ ) could potentially stabilize the electronic structure of the molecule due to resonance, reducing its reactivity with DNA. The aldehyde group, while potentially reactive, may not be in the correct configuration or electronic environment to interact with genetic material effectively. Therefore, the overall molecular structure may not fit or react easily with the nucleophilic sites on DNA necessary for mutagenesis.

**Hypothesis:** The molecule with the given substructure is deemed to have a medium influence towards being "non-mutagenic." This could be due to the stabilization provided by the electron-withdrawing sulfonyl group reducing the reactivity of the molecule. Additionally, the spatial configuration afforded by the connected aromatic ring may further hinder the molecule's ability to interact with DNA in a way that would cause a mutation.

# Cluster #11 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 11, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 3.0 ( $\pm 1.6$ ) nodes. The concept is generally associated with an impact of 15.1 ( $\pm 1.2$ ) on the prediction outcome.

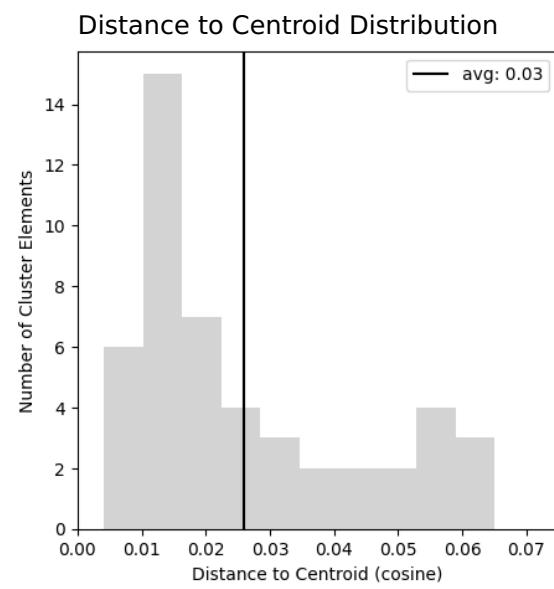
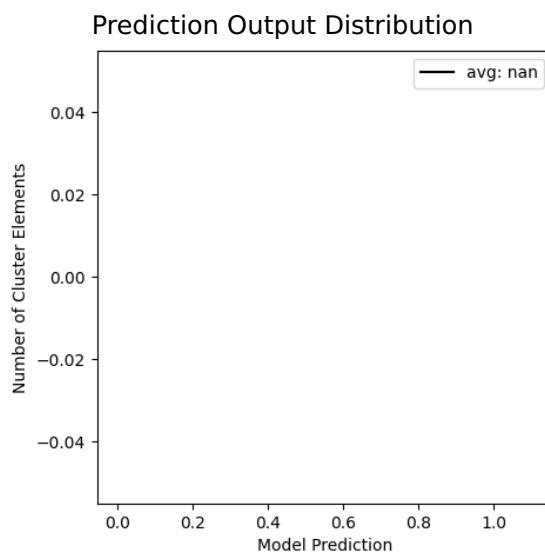
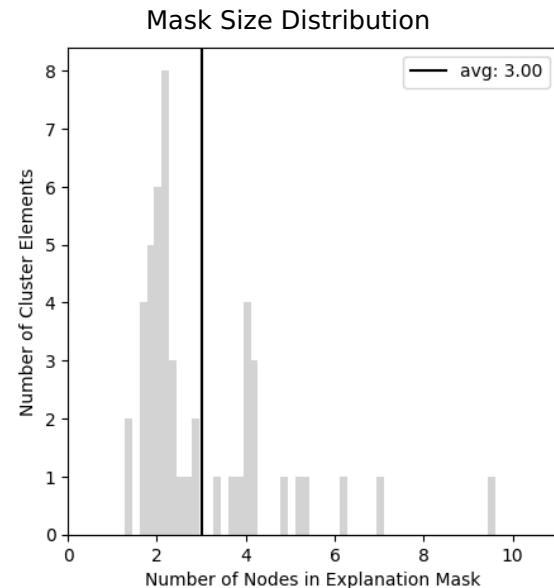
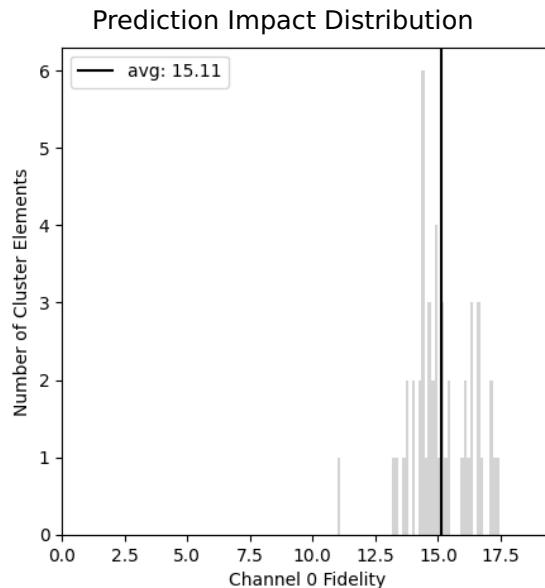
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	48
Channel Index	0.0 (0.0)

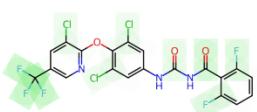
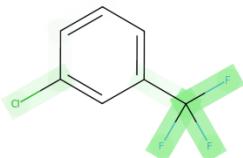
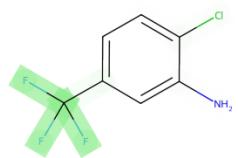
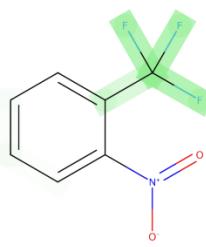
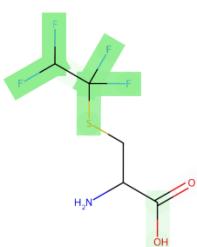
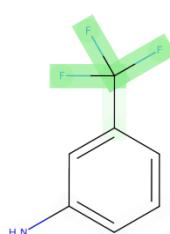
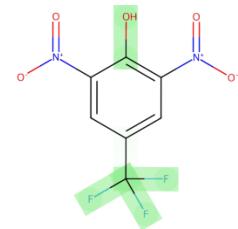
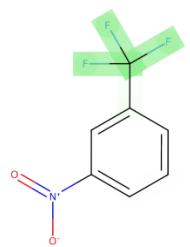
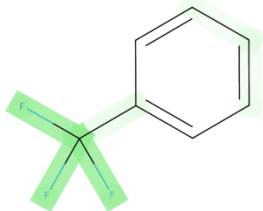
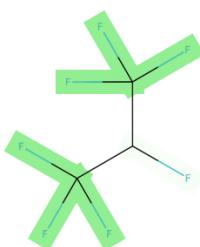
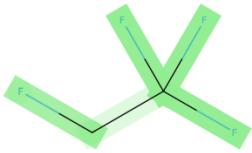
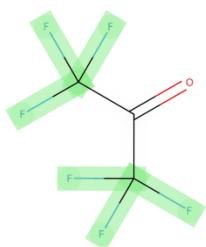
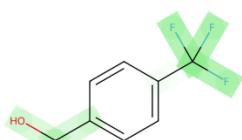
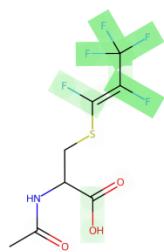
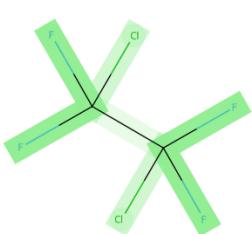
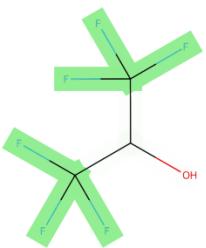
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "F-C-F" represents a carbon atom single bonded to two fluorine atoms. The fluorine atoms, being highly electronegative, will pull electron density away from the central carbon atom, creating a strong inductive effect which can stabilize the molecule and reduce its reactivity. A molecule with lower reactivity is less likely to interact with DNA or other cellular components to cause mutagenic changes.

**Hypothesis:** Molecules containing the "F-C-F" substructure are hypothesized to have a medium influence towards being non-mutagenic due to the inductive effect of the fluorine atoms which stabilizes the molecule and decreases its reactivity, thus reducing its potential to cause genetic modifications.

# Cluster #12 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 12, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 3.2 ( $\pm 1.7$ ) nodes. The concept is generally associated with an impact of 15.0 ( $\pm 2.0$ ) on the prediction outcome.

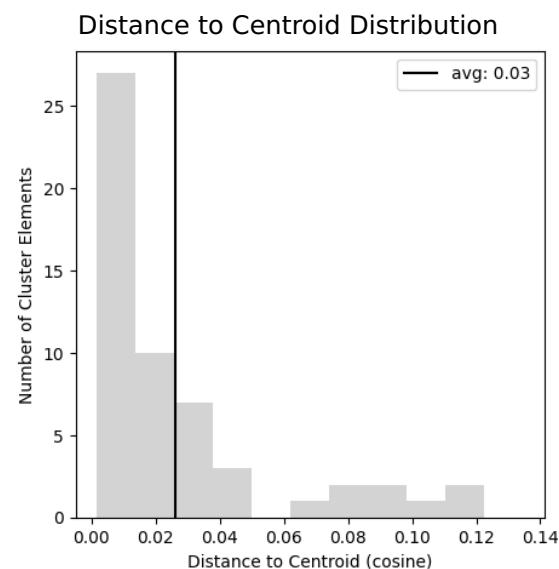
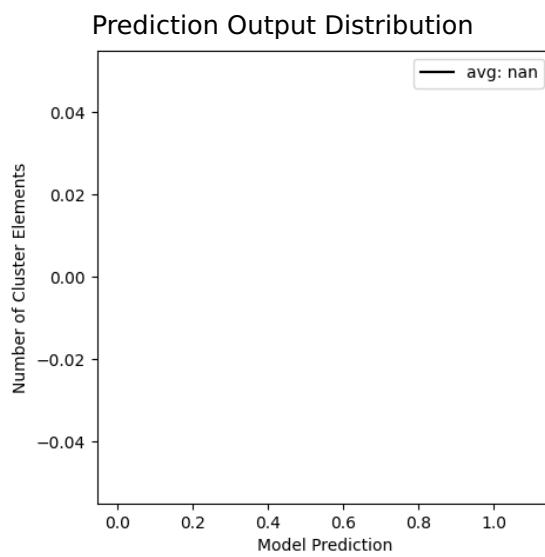
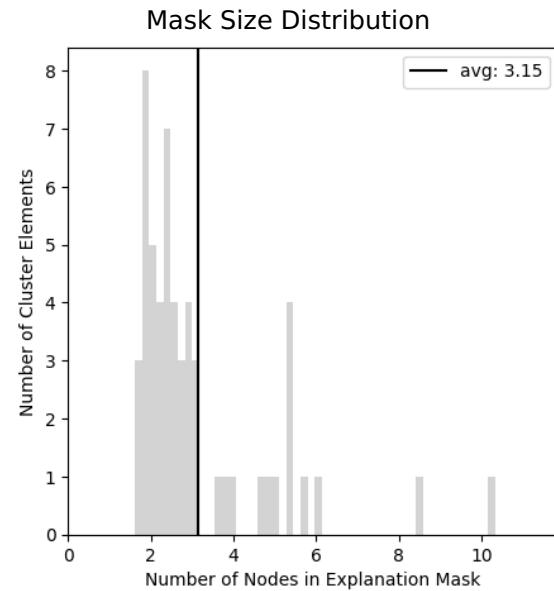
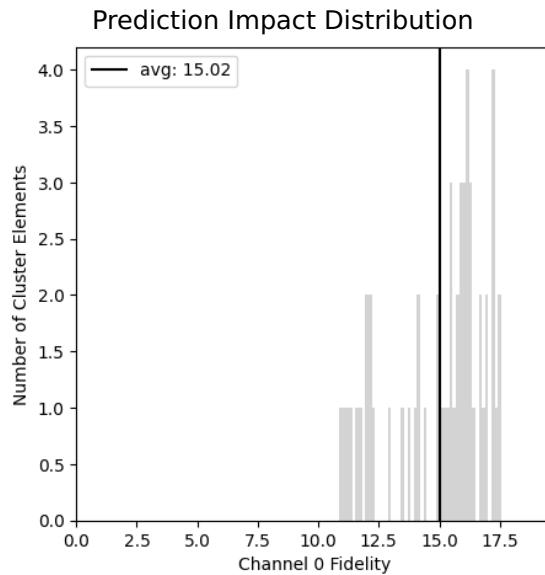
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	55
Channel Index	0.0 (0.0)

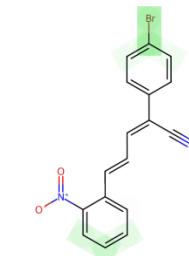
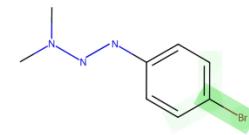
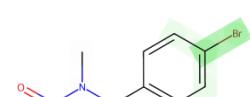
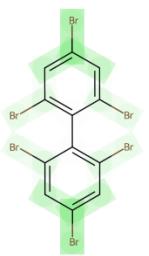
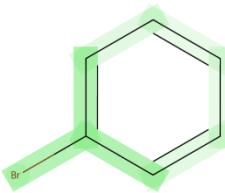
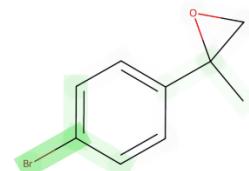
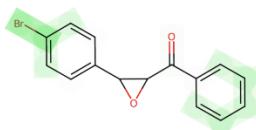
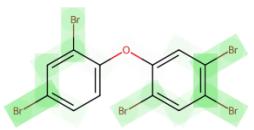
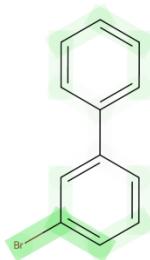
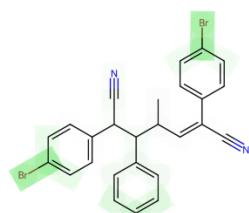
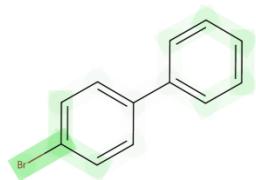
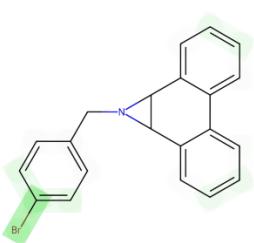
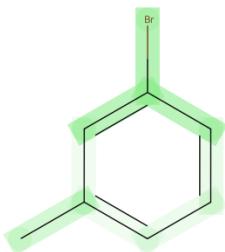
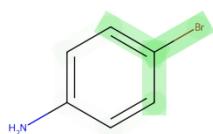
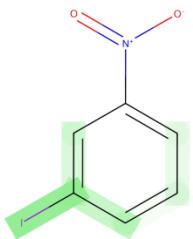
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



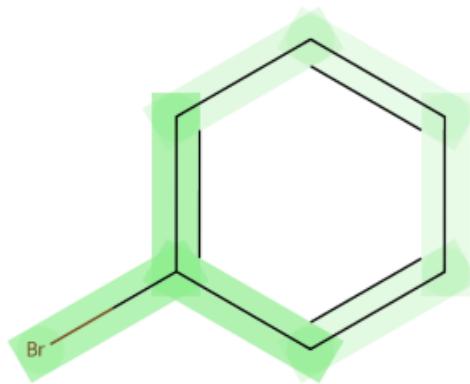
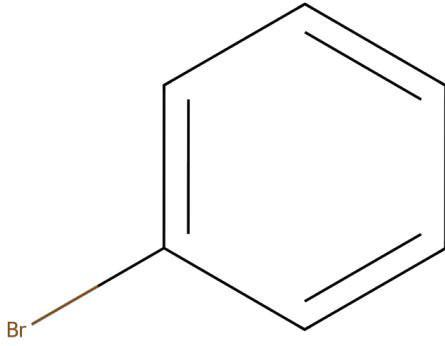
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation denotes a bromobenzene molecule, which consists of a benzene ring with a bromine atom attached to it. Since halogens such as bromine are electron-withdrawing groups, they can stabilize the electron-rich aromatic ring through inductive effects, reducing the reactivity of the ring towards electrophilic attack that could lead to mutagenic activity. Additionally, the steric bulk of the bromine atom may hinder interactions with molecular targets within the cell, further mitigating any potential mutagenic effect.

**Hypothesis:** Bromobenzene has a medium influence toward being non-mutagenic due to the electron-withdrawing and steric effects of the bromine substituent. This halogenated aromatic compound is less likely to engage in reactions that result in the modification of genetic material, compared to unsubstituted or more reactive aromatic compounds.

# Cluster #13 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 13, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 3.6 ( $\pm 1.5$ ) nodes. The concept is generally associated with an impact of 16.5 ( $\pm 1.4$ ) on the prediction outcome.

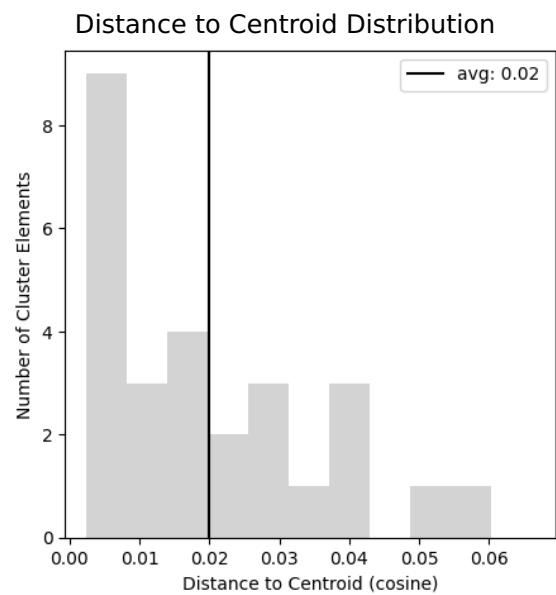
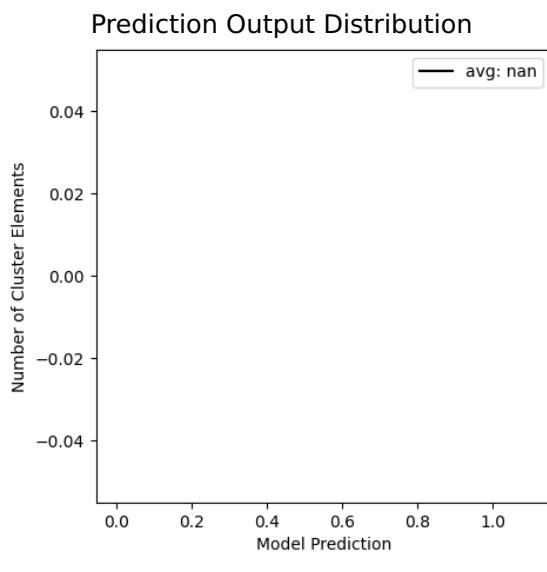
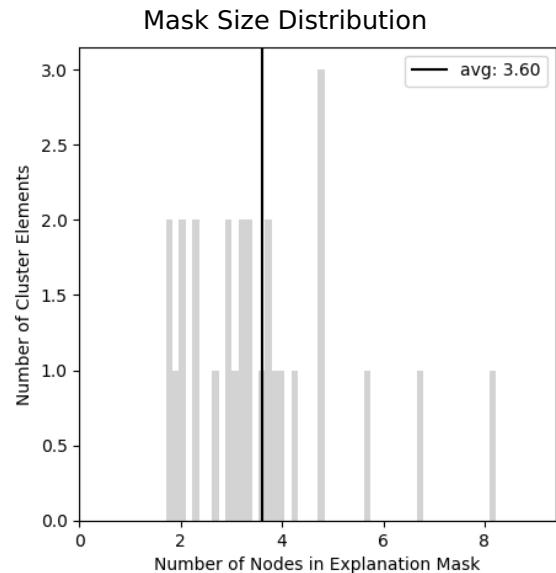
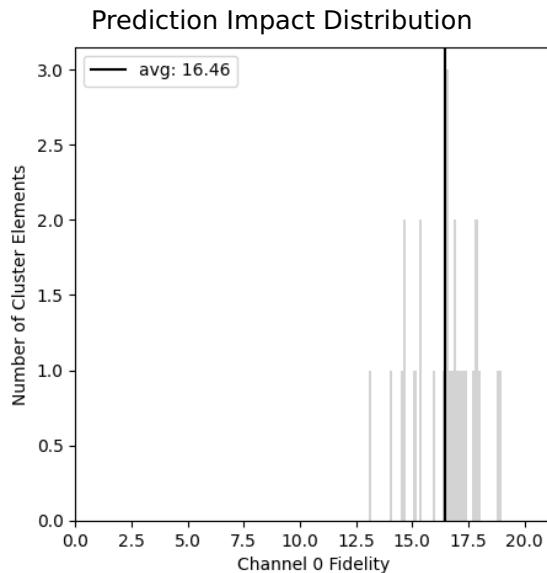
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	27
Channel Index	0.0 (0.0)

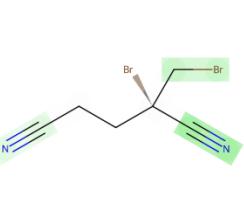
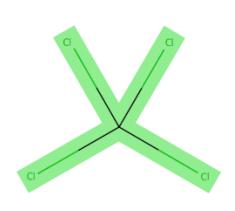
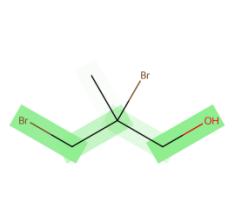
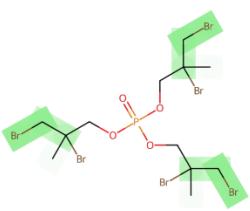
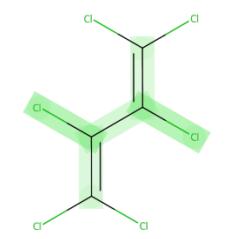
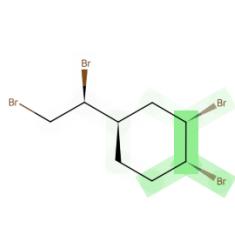
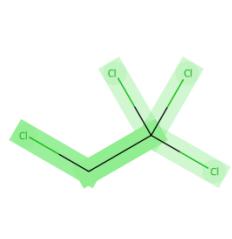
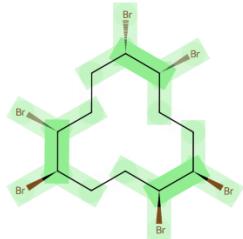
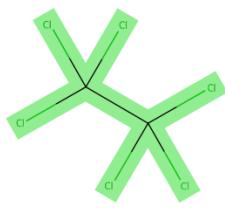
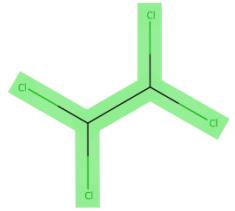
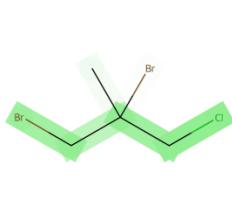
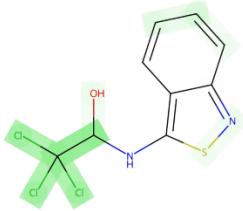
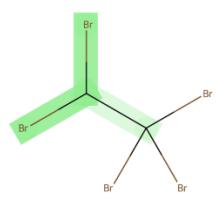
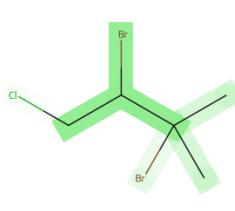
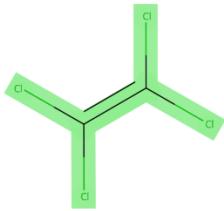
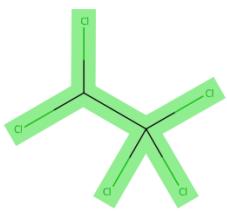
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular structure denoted by the SMILES representation "Cl-C=C-Cl" contains a vinyl chloride unit, which is a type of structure where a double bond connects a carbon atom to another carbon atom, with each carbon further bonded to a chlorine atom. The presence of a double bond usually implies potential reactivity, as the pi-electrons are available for reactions with biological molecules. However, in this case, the substitution of both carbons with chlorine atoms may hinder such reactivity due to electronic effects such as resonance and inductive effect, where the chlorine atoms withdraw electron density from the double bond, making it less reactive. Furthermore, the chlorines potentially increase the steric hindrance, preventing the necessary close approach of the DNA for a successful interaction leading to mutagenicity. Therefore, this particular molecular structure may not be sufficiently reactive to directly interact with and cause mutations in genetic material.

**Hypothesis:** Molecules containing the structure "Cl-C=C-Cl" are hypothesized to display a medium influence toward being non-mutagenic. The electron-withdrawing effect of the chlorine atoms coupled with increased steric hindrance is likely to reduce the reactivity of the double bond towards nucleophilic attack by DNA nucleotides, thus limiting the mutagenic potential of this molecular substructure.

# Cluster #14 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 14, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.9 ( $\pm 1.0$ ) nodes. The concept is generally associated with an impact of 13.6 ( $\pm 1.0$ ) on the prediction outcome.

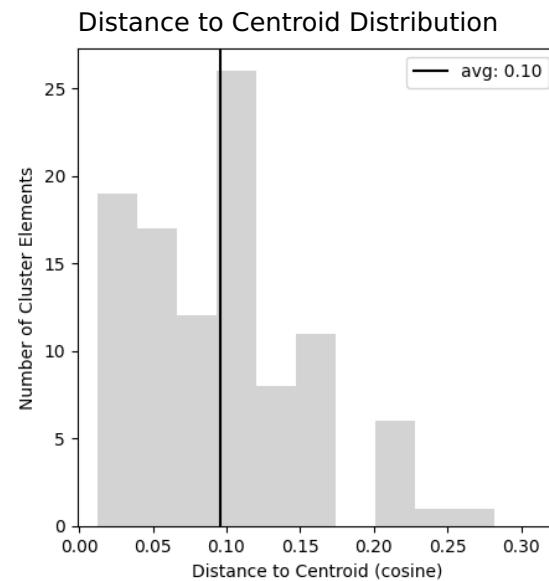
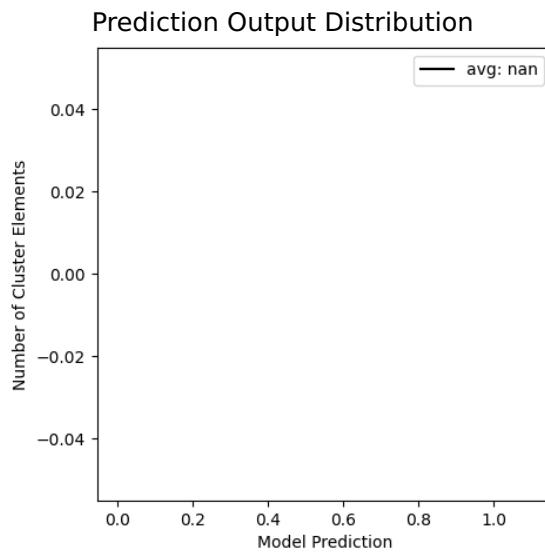
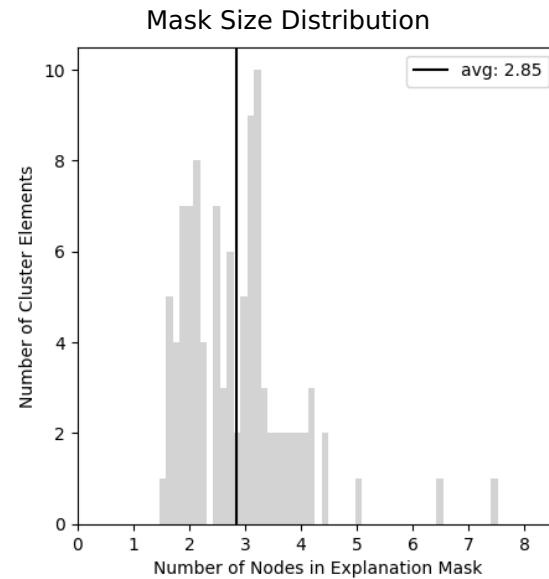
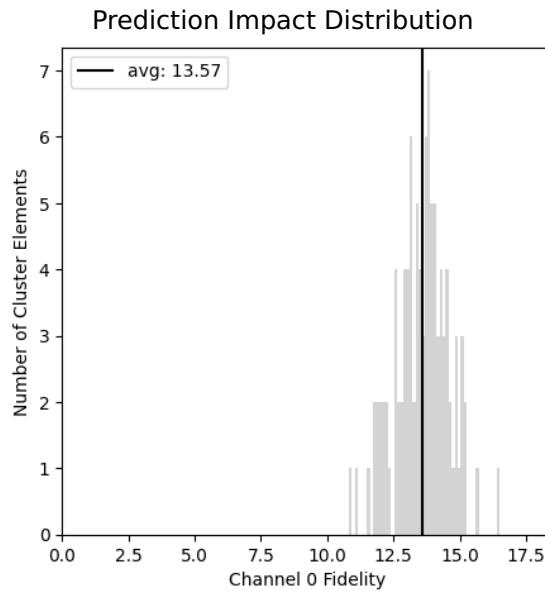
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	101
Channel Index	0.0 (0.0)

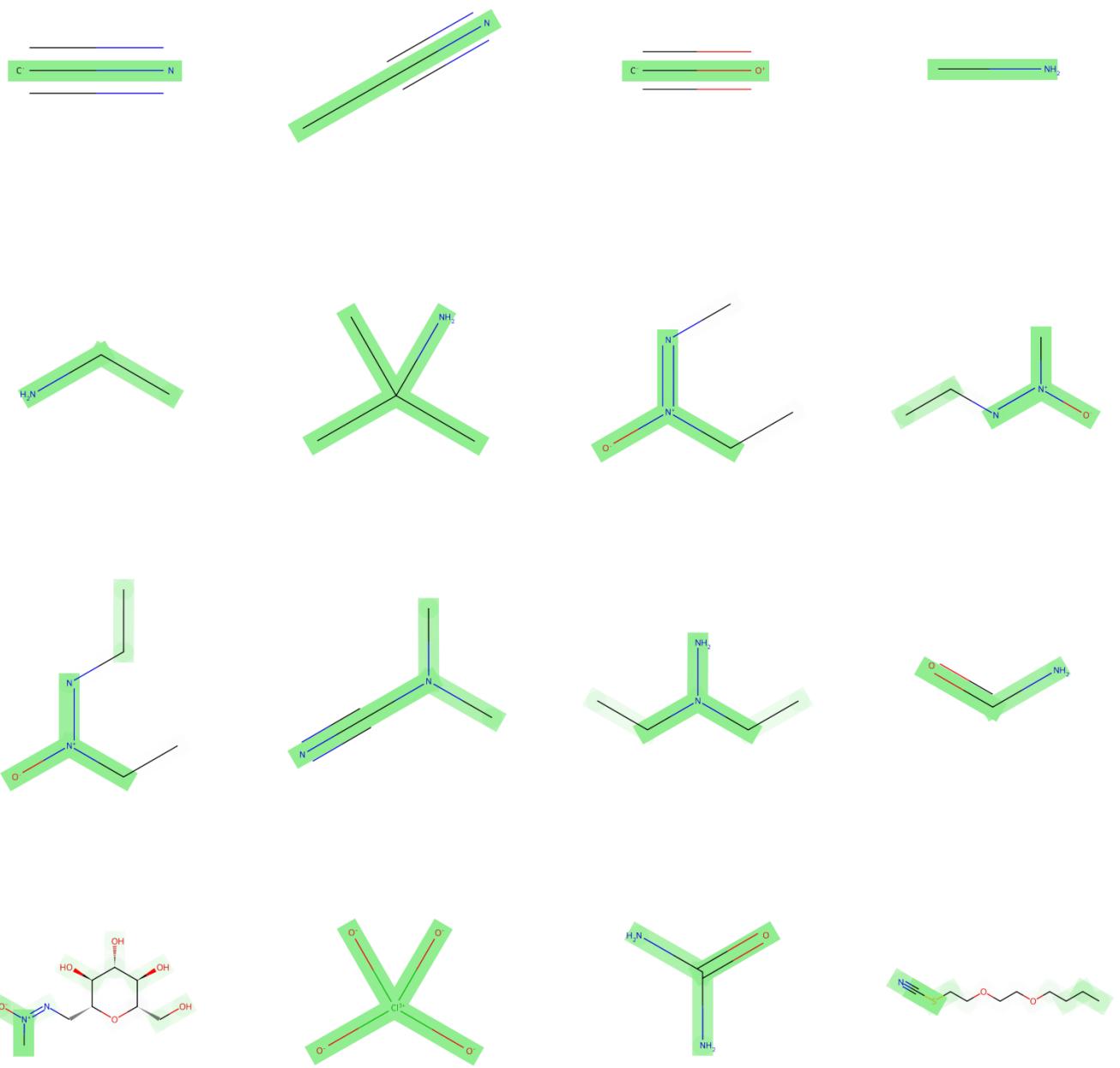
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a methylamine group, as indicated by the substructure "CN" in SMILES, might reduce the likelihood of a molecule being a mutagen. This could be due to the increased stability the methyl group confers to the amine, making it less reactive and less likely to interact with DNA. Moreover, a methylated nitrogen might be unable to form the necessary bonds with DNA bases that are required for mutagenic activity.

**Hypothesis:** Molecules containing the substructure "CN" are hypothesized to be less mutagenic. This could stem from the enhanced stability provided by the methyl group and the steric hindrance it poses, potentially preventing binding to DNA. The methyl group adjacent to the nitrogen may be critical in reducing the mutagenic potential of the molecule.

# Cluster #15 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 15, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.5 ( $\pm 0.9$ ) nodes. The concept is generally associated with an impact of 14.8 ( $\pm 2.3$ ) on the prediction outcome.

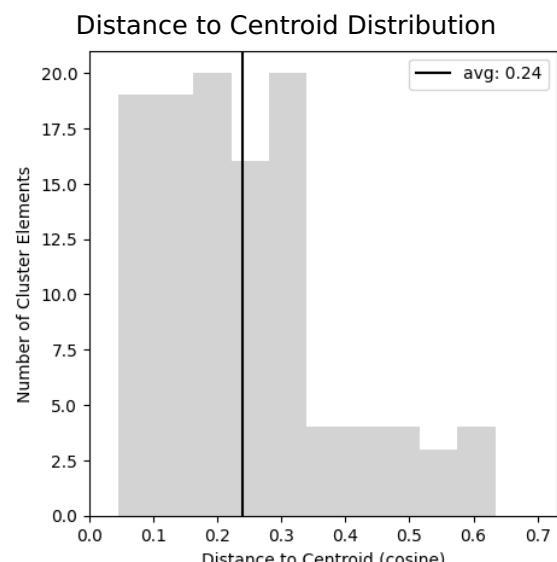
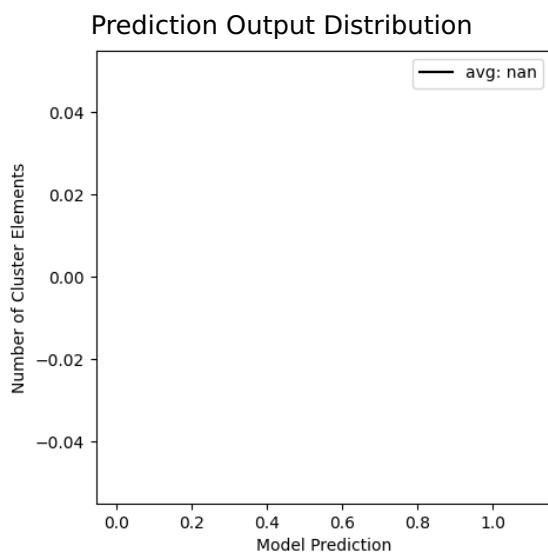
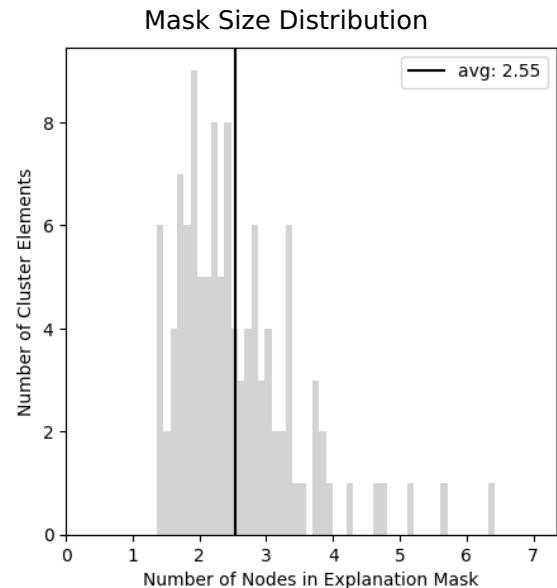
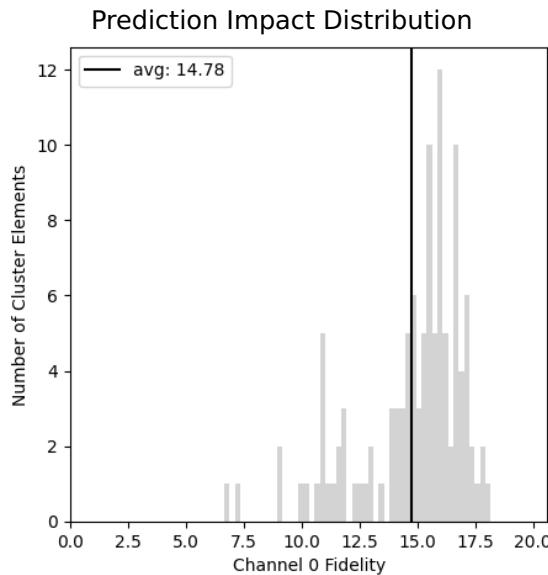
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	113
Channel Index	0.0 (0.0)

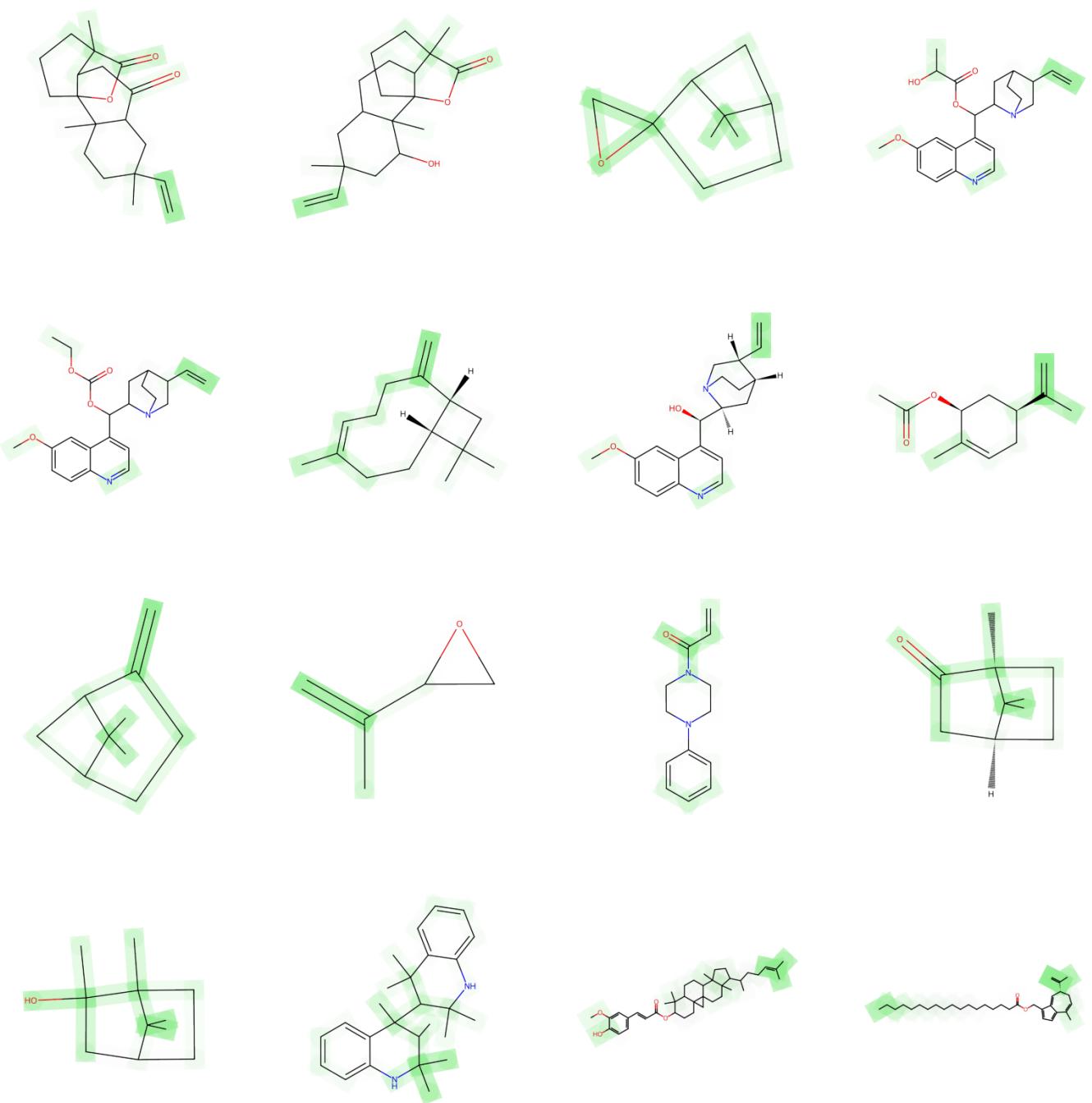
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



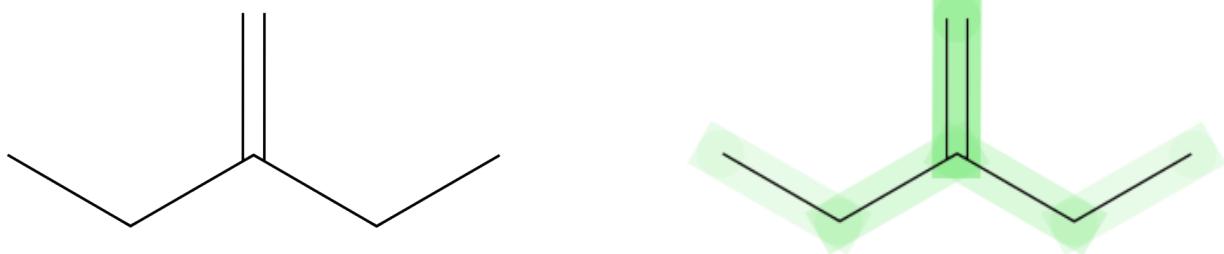
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents an aliphatic hydrocarbon chain with a double bond. Non-aromatic hydrocarbons, particularly those without extensive conjugation or reactive functional groups, tend to be less reactive with DNA. The structure lacks obvious electrophilic centers which are typically responsible for forming adducts with DNA bases, leading to mutations. The presence of saturated carbons adjacent to an isolated double bond may also reduce the likelihood of the compound undergoing metabolic activation to form a more reactive, mutagenic species.

**Hypothesis:** Molecules containing the substructure "C=C(-C-C)-C-C" exhibit a medium influence towards being non-mutagenic, likely due to the lack of highly reactive functional groups and the presence of stabilizing saturated carbon atoms adjacent to the double bond. The isolated double bond in a non-conjugated system does not significantly raise the mutagenic potential of the compound, while the surrounding saturated hydrocarbons may inhibit any minor reactive tendencies the double bond may possess.

# Cluster #16 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 16, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.5 ( $\pm 0.7$ ) nodes. The concept is generally associated with an impact of 14.8 ( $\pm 1.8$ ) on the prediction outcome.

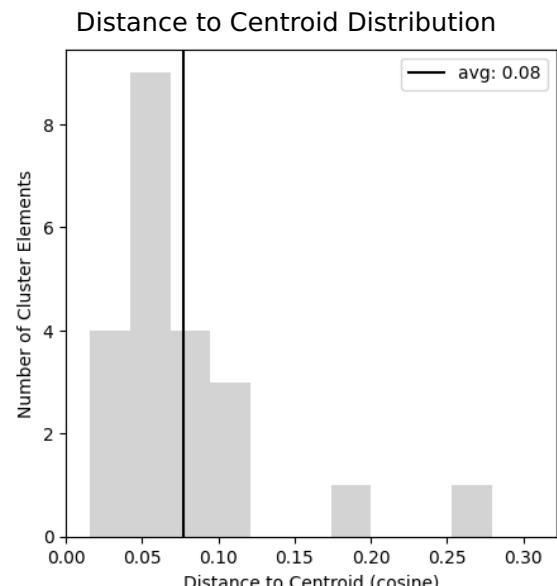
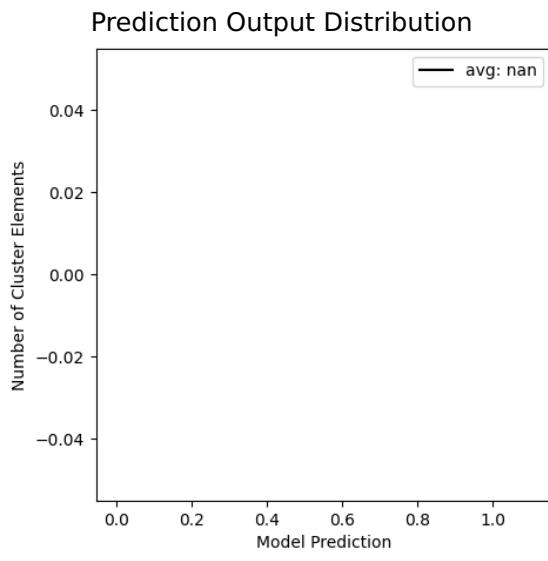
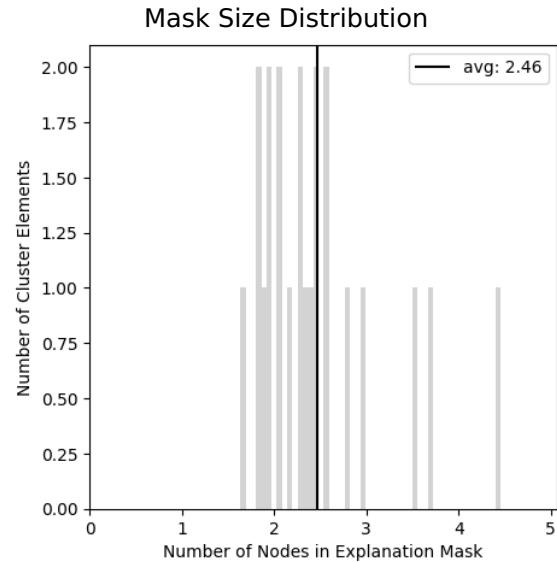
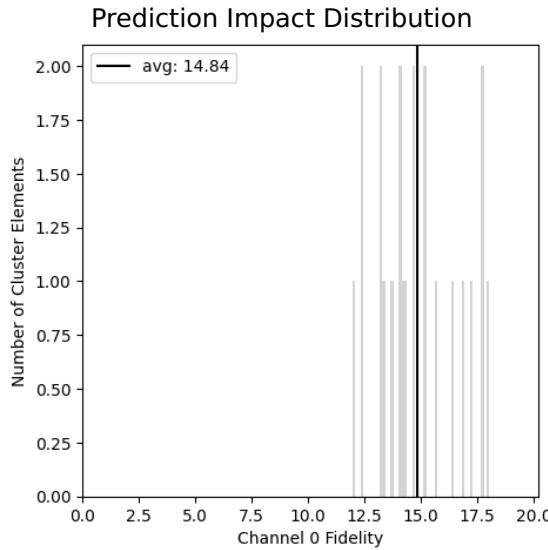
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	22
Channel Index	0.0 (0.0)

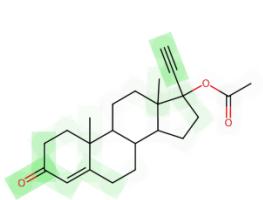
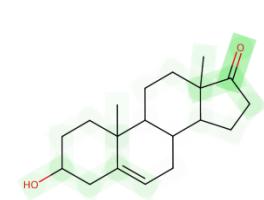
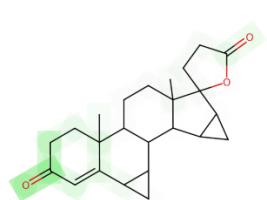
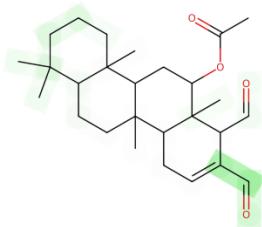
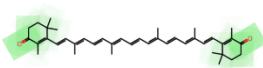
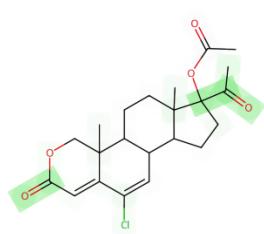
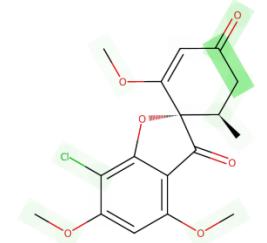
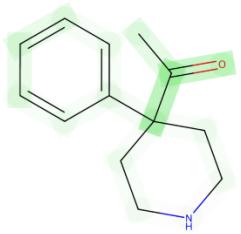
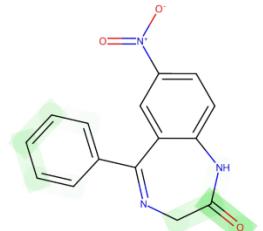
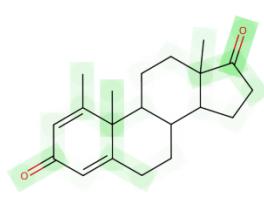
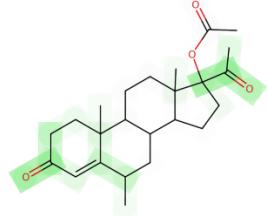
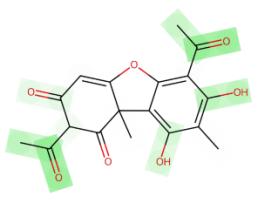
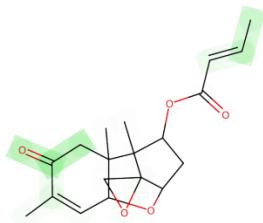
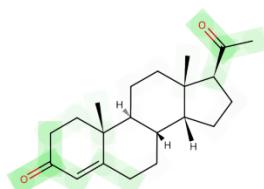
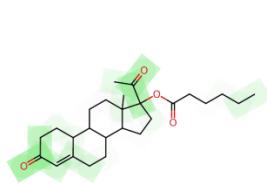
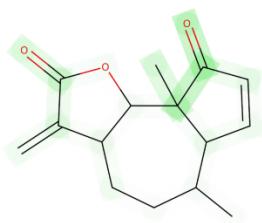
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



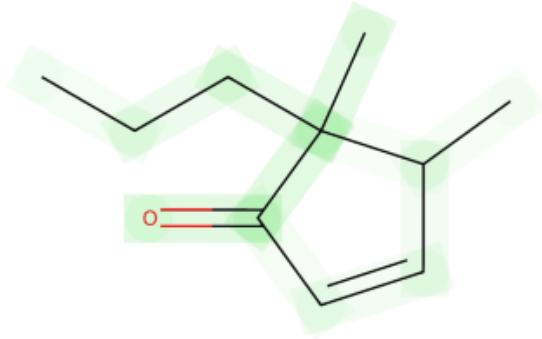
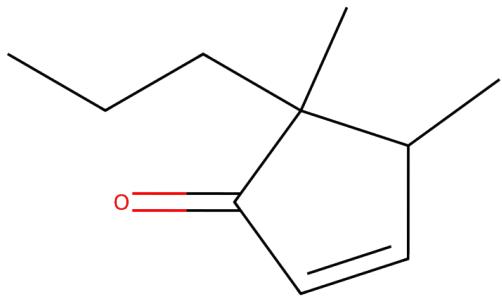
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure in question appears to contain a cyclohexane ring with a ketone functional group and an alkene. The cyclohexane ring is a relatively stable structure that does not readily react with DNA. Moreover, the ketone group is generally less reactive compared to aldehydes, which can form DNA adducts leading to mutations. Alkenes can participate in reactions, but the positioning of the alkene in this structure may be sterically hindered by the cyclohexane ring, reducing its reactivity.

**Hypothesis:** Molecules containing the substructure "C-C-C1(-C)-C(=O)-C=C-C-1-C" are likely to exhibit a medium influence towards being non-mutagenic. The stability of the cyclohexane ring and the lower reactivity of the ketone functional group may prevent interactions with genetic material. Additionally, the potential steric hindrance affecting the alkene's reactivity could further reduce the likelihood of mutagenic activity.

# Cluster #17 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 17, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 11.1 ( $\pm 1.7$ ) on the prediction outcome.

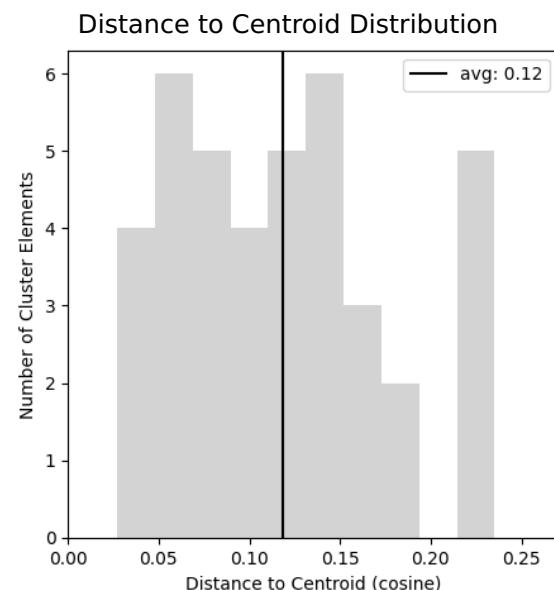
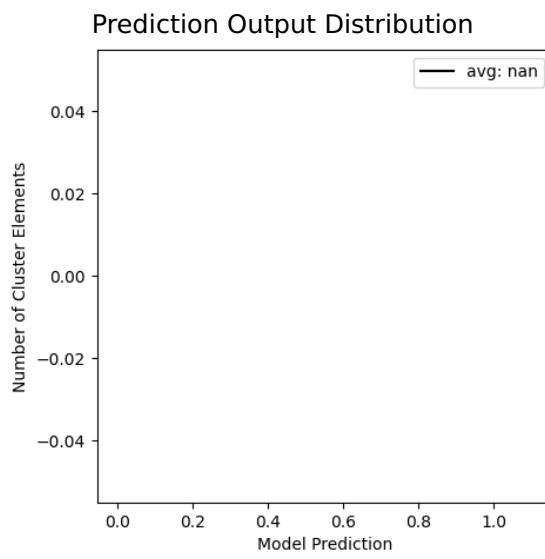
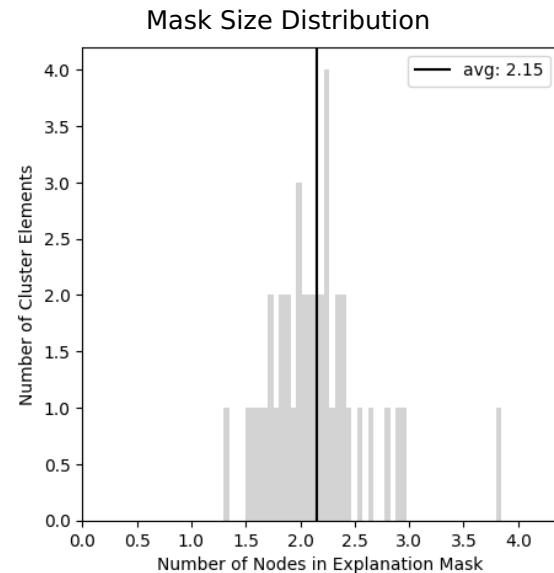
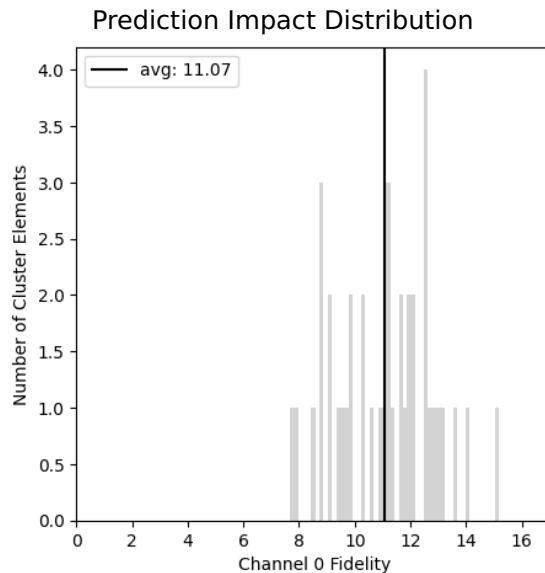
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	40
Channel Index	0.0 (0.0)

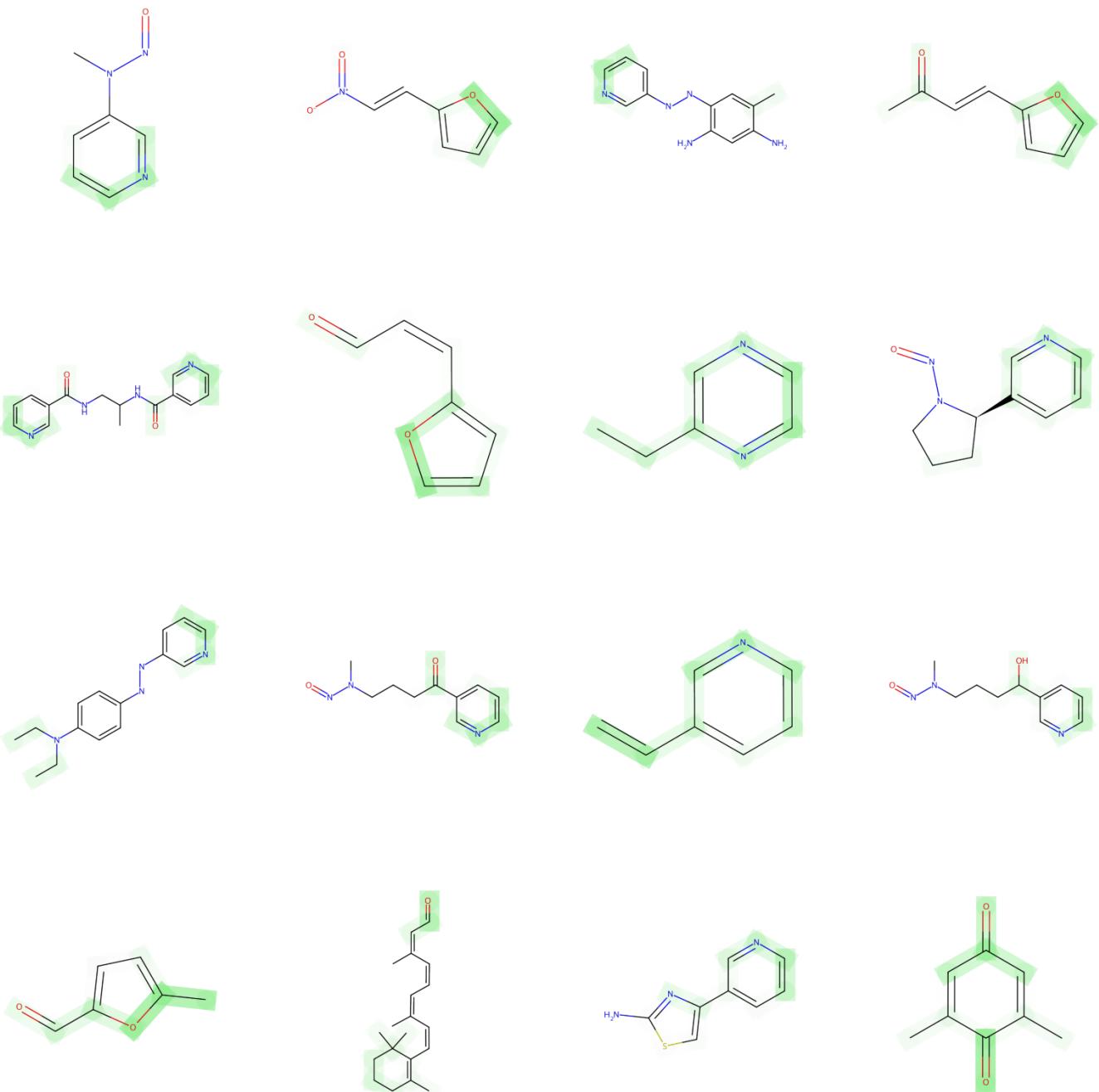
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES string represents a pyridine-like heteroaromatic ring containing nitrogen (N) atoms. Heteroaromatic compounds like pyridines often have a high degree of resonance stability due to the delocalization of  $\pi$ -electrons over the ring system, which may reduce their ability to interact with DNA. This stability could make them less likely to form reactive intermediates that could cause mutation. Additionally, the electronegative nitrogen atoms might result in a lower electron density on the compound, which can reduce its reactivity towards nucleophilic attack on genetic material.

**Hypothesis:** A pyridine-like aromatic ring has a medium influence towards being non-mutagenic, possibly due to the stability conferred by electron delocalization and the presence of electronegative nitrogen atoms that lower reactivity. The aromatic system's resonance stability diminishes its capacity to form DNA-reactive intermediates, while the nitrogen's electronegativity decreases the likelihood of nucleophilic attacks that could lead to mutagenesis.

# Cluster #18 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 18, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 1.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 9.6 ( $\pm 1.4$ ) on the prediction outcome.

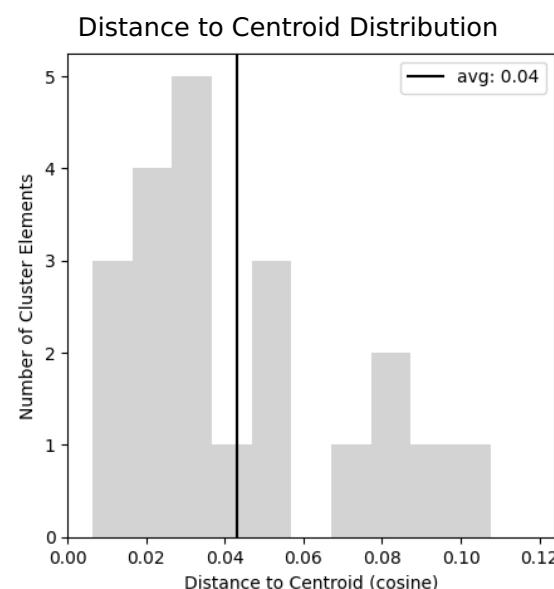
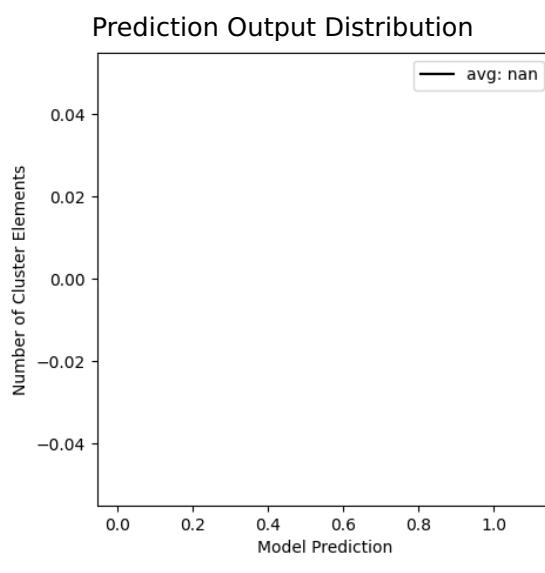
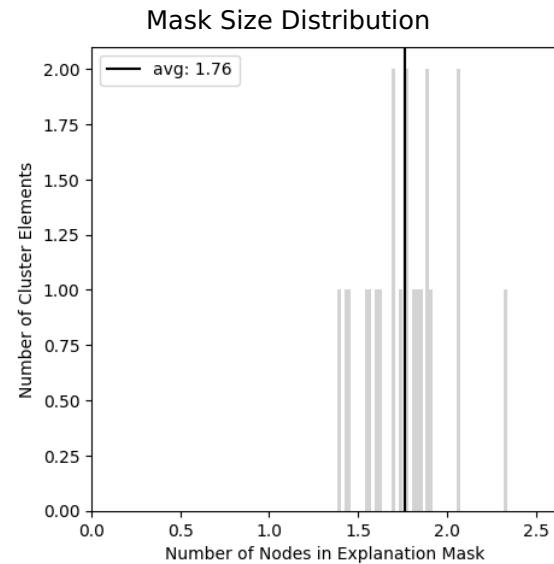
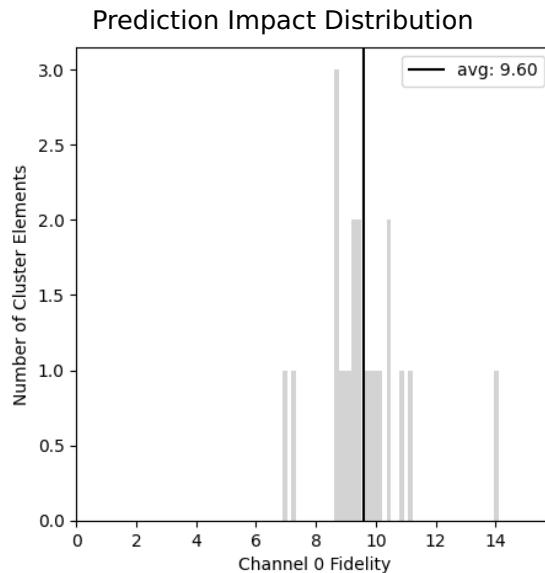
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	0.0 (0.0)

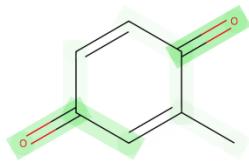
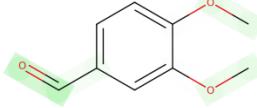
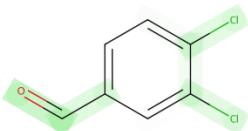
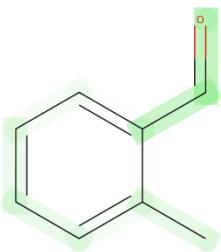
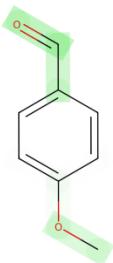
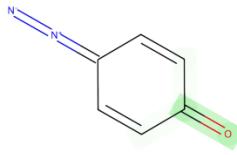
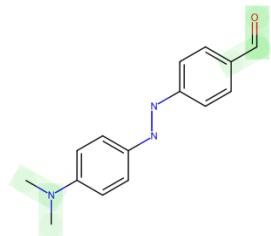
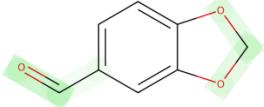
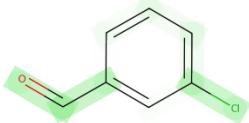
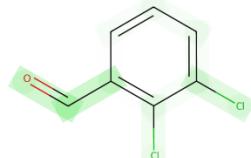
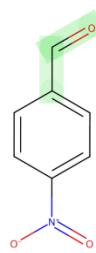
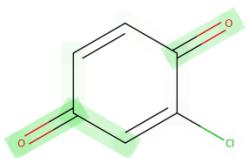
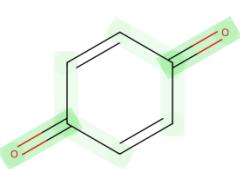
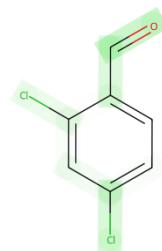
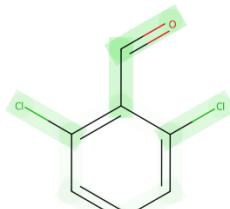
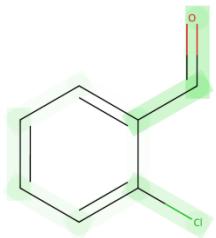
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



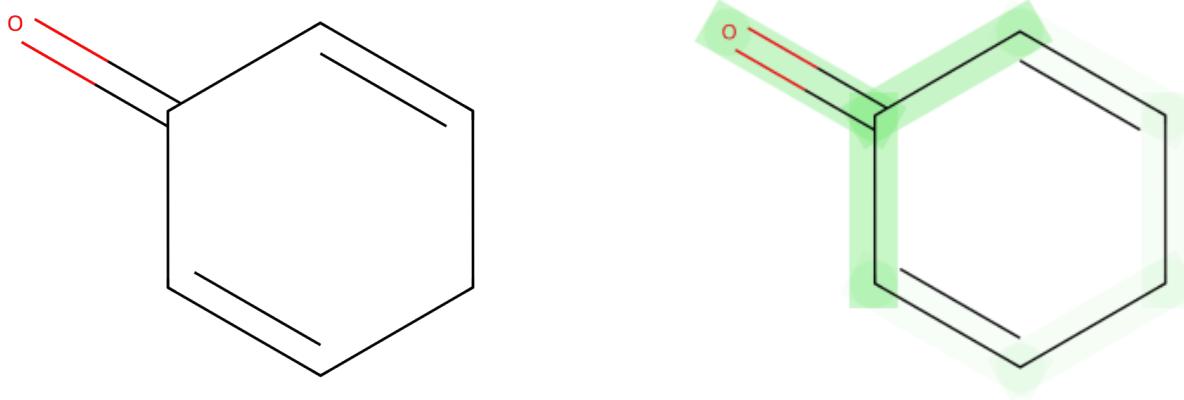
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure represents a cyclic ketone with alternating double and single bonds. The presence of conjugated double bonds could potentially allow for electron delocalization which generally promotes stability within the ring. Additionally, the ketone group ( $C=O$ ) is known for its electron withdrawing characteristics, which further contributes to the stability of the molecule by reducing the likelihood of electron-rich regions that could react with DNA. Stable compounds are less likely to interact with genetic material and therefore might be expected to exhibit lower mutagenicity.

**Hypothesis:** The cyclic ketone structure with alternating double and single bonds is associated with low mutagenicity due to the stabilization provided by electron delocalization and the electron-withdrawing effect of the ketone group. This stability reduces the molecule's reactivity with genetic material, therefore decreasing the likelihood of causing mutations.

# Cluster #19 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 19, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 1.7 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 11.0 ( $\pm 2.3$ ) on the prediction outcome.

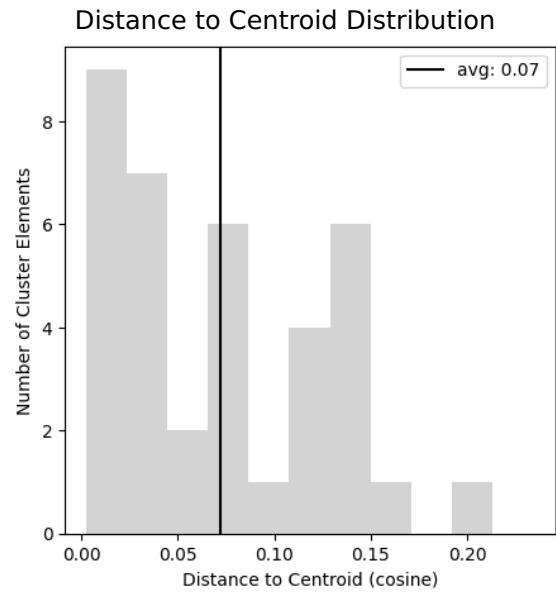
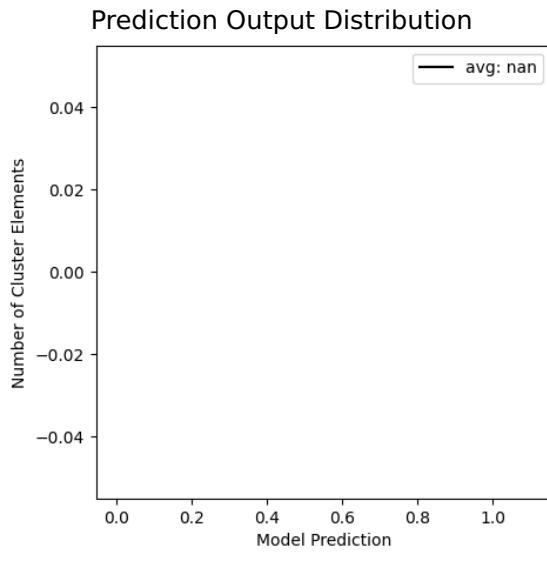
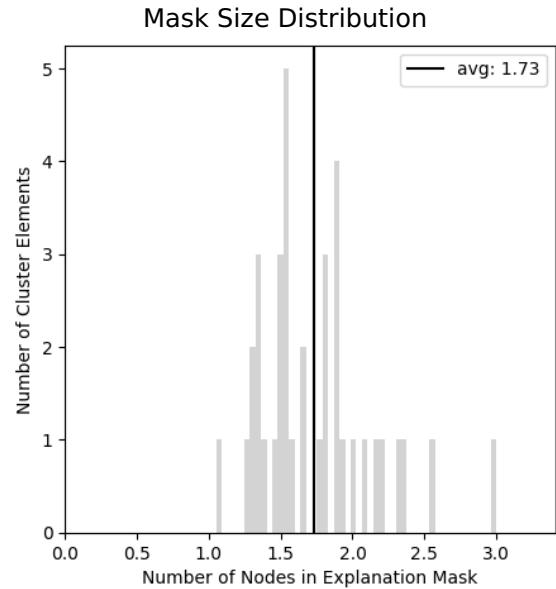
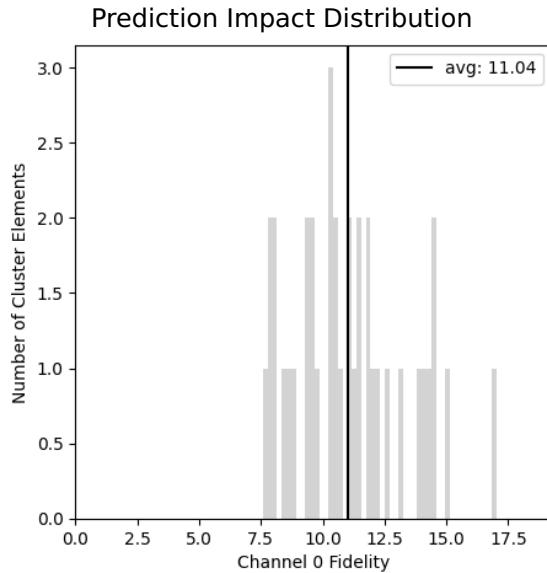
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	37
Channel Index	0.0 (0.0)

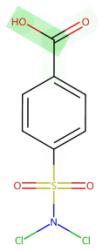
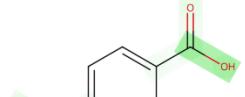
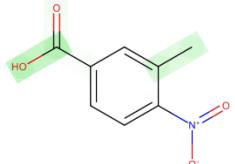
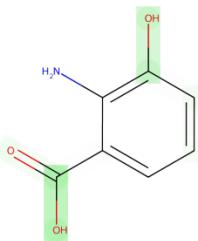
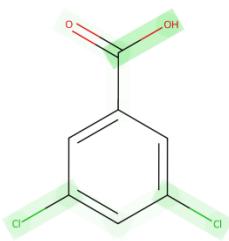
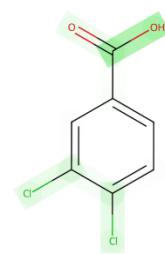
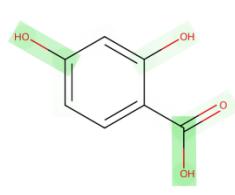
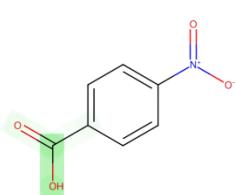
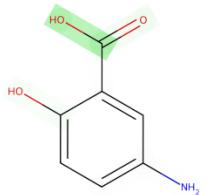
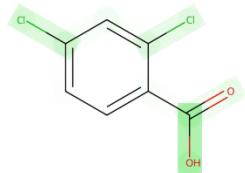
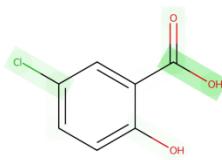
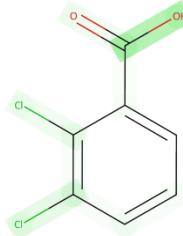
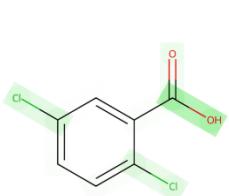
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



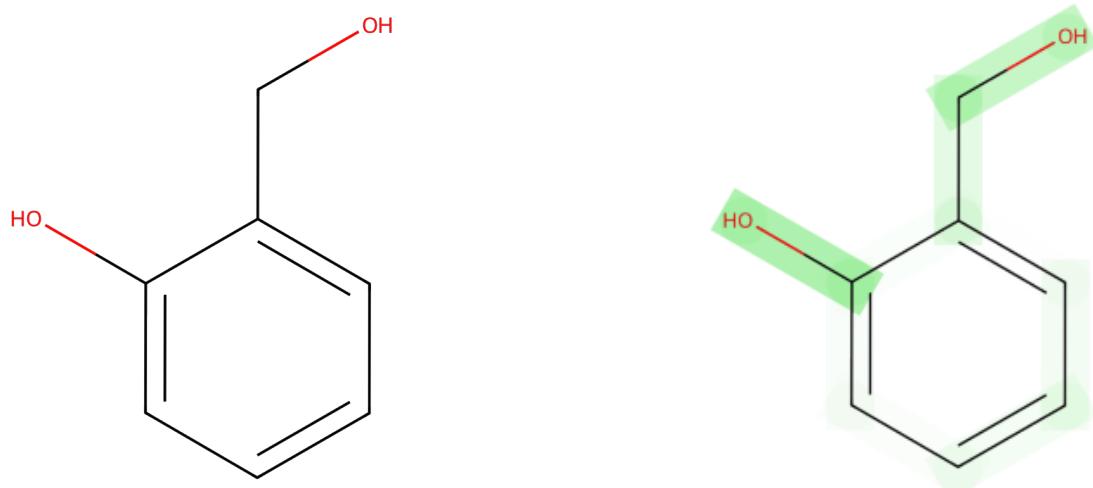
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecule structure described by the given SMILES representation is that of a benzene ring with two hydroxyl groups (-OH) attached via a methylene bridge (-CH<sub>2</sub>-), forming a bisphenol structure. Hydroxyl groups are known to be electron-donating through resonance, which increases the electron density on the benzene ring, making it less susceptible to electrophilic attack that might result in the formation of DNA adducts and lead to mutagenesis. Moreover, the methylene bridge might introduce steric hindrance, potentially impeding interactions with the genetic material.

**Hypothesis:** The given molecular substructure is associated with a medium tendency to be non-mutagenic due to the presence of hydroxyl groups which stabilize the structure and reduce its reactivity with genetic material. The bisphenol backbone along with steric hindrance from the methylene group likely contribute to the non-mutagenic character by limiting the molecule's ability to intercalate into DNA and thus reducing the chance of mutations.

# Cluster #20 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 20, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.8$ ) nodes. The concept is generally associated with an impact of 9.1 ( $\pm 0.9$ ) on the prediction outcome.

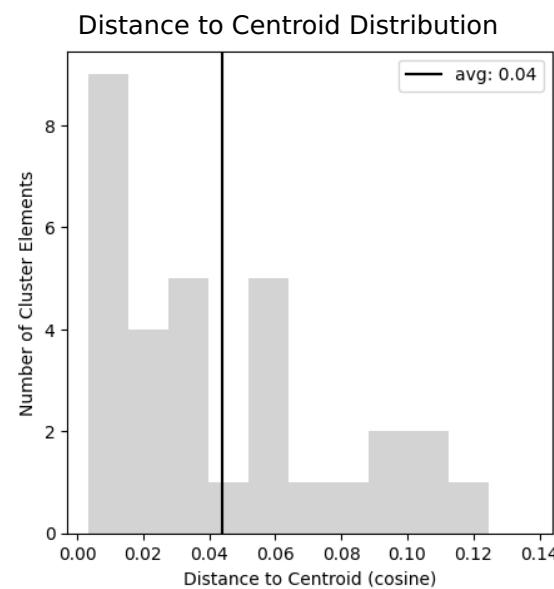
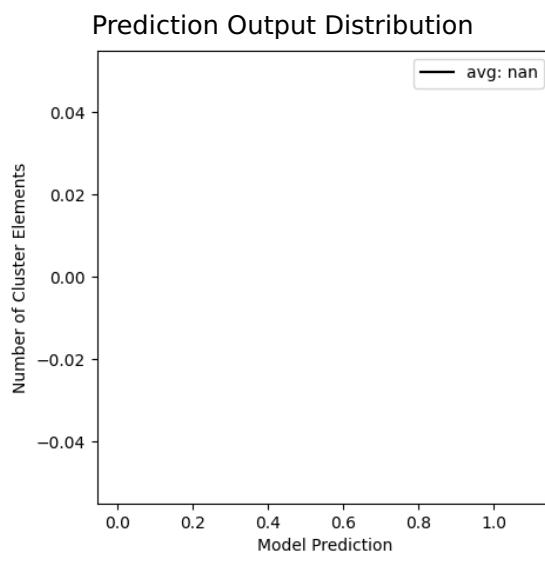
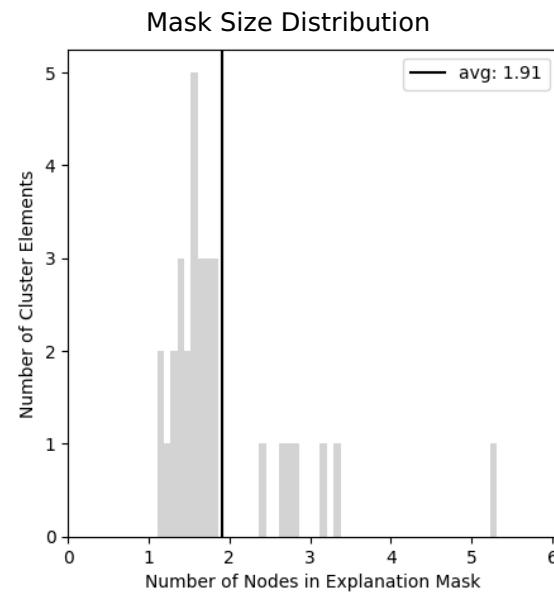
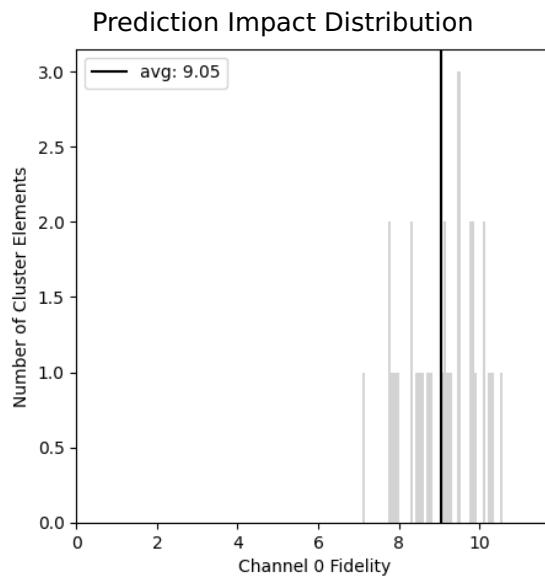
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	31
Channel Index	0.0 (0.0)

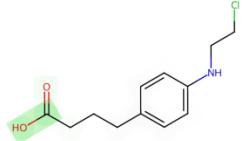
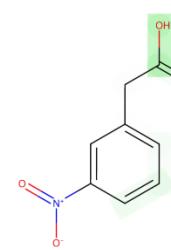
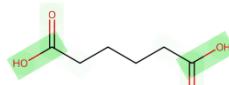
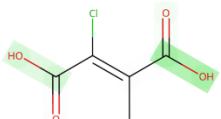
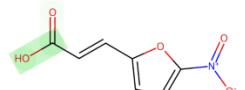
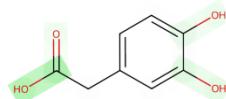
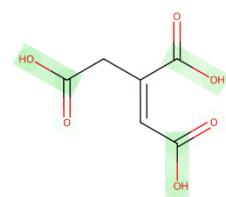
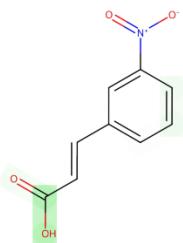
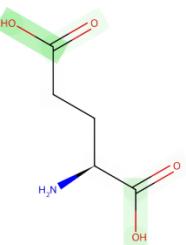
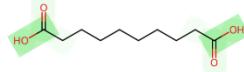
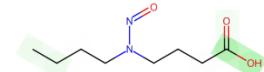
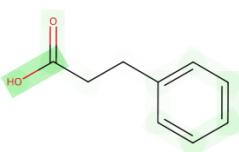
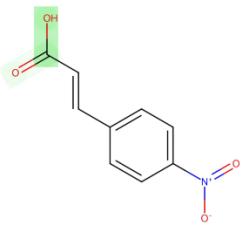
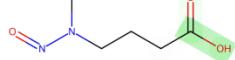
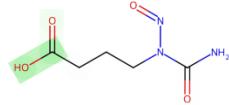
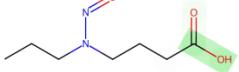
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure "N-C-C-C-C-O" represents a molecule with an amino group (-NH<sub>2</sub>) at one end and a hydroxyl group (-OH) at the other, connected by a four-carbon alkane chain. These functional groups are relatively stable and less likely to interact with DNA in a way that causes mutations. The length of the carbon chain could provide enough spatial separation to reduce the chance of the amino and hydroxyl groups coming into reactive contact with genetic material. Additionally, the absence of highly reactive or electrophilic groups in this structure could account for its lower mutagenicity.

**Hypothesis:** Molecules with the structure "N-C-C-C-C-O" have a **SMALL** influence on non-mutagenicity due to the stability and low-reactivity of the terminal amino and hydroxyl groups along with the length of the carbon chain providing spatial separation from genetic material. The structural components lack the necessary reactivity usually associated with mutagenic chemicals, hence reducing their potential to cause genetic alterations.

# Cluster #21 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 21, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.6 ( $\pm 1.0$ ) nodes. The concept is generally associated with an impact of 6.3 ( $\pm 0.8$ ) on the prediction outcome.

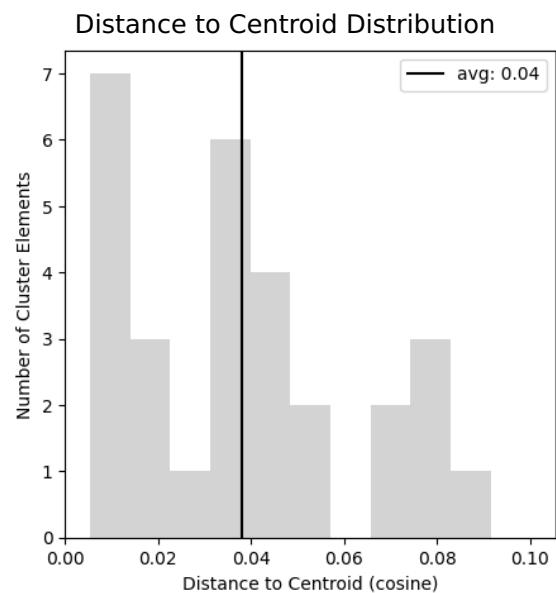
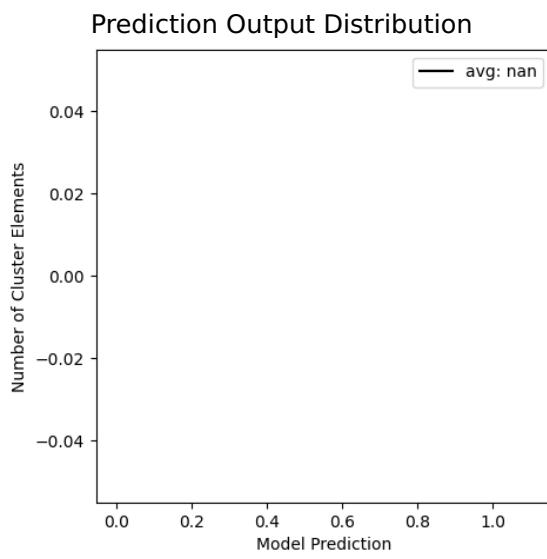
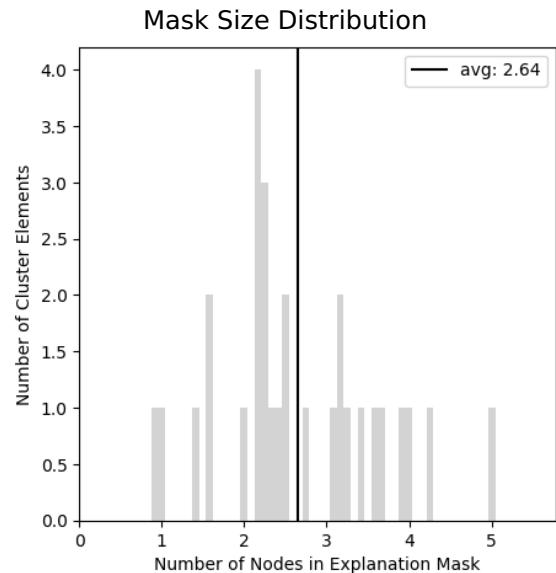
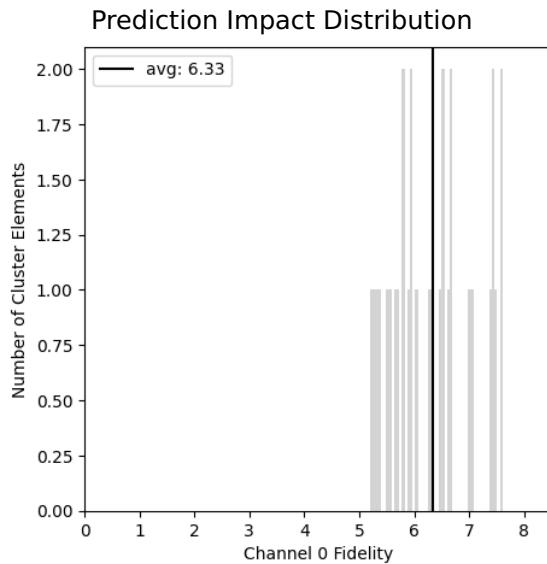
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	29
Channel Index	0.0 (0.0)

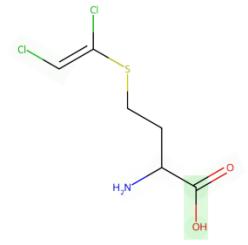
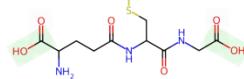
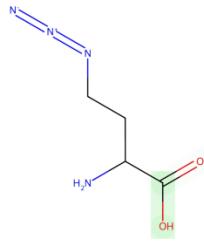
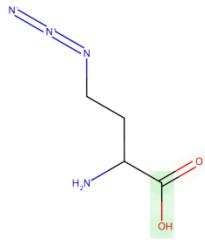
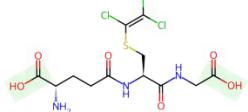
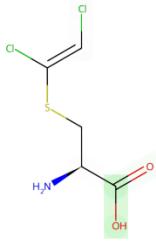
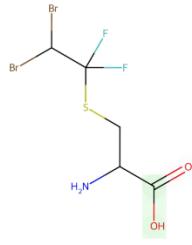
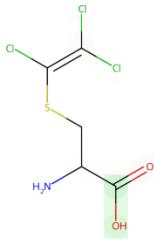
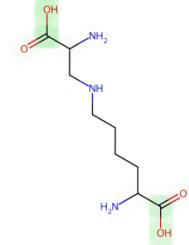
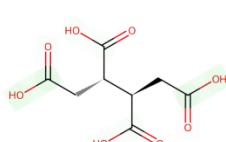
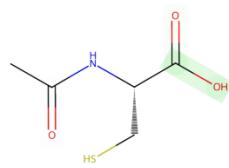
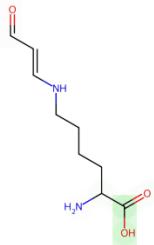
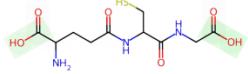
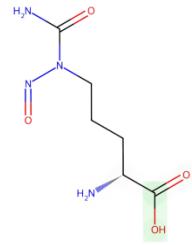
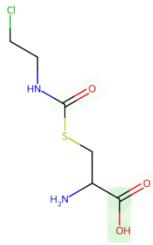
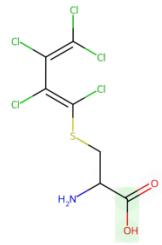
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



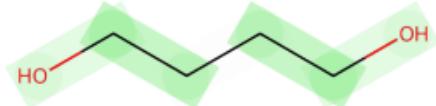
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by "O-C-C-C-C-O" indicates a linear chain of carbons with hydroxyl (OH) groups at either end. Typically, linear and less complex structures have a lower likelihood of interacting with DNA in ways that could cause mutations. The polar hydroxyl groups may also contribute to the molecule being less reactive, as they could be involved in hydrogen bonding with water, making the molecule less available to interact with genetic material.

**Hypothesis:** The molecular substructure "O-C-C-C-C-O" is associated with a small influence towards being "non-mutagenic". The linear and simple carbon chain minimizes complex interactions with DNA, and the presence of hydroxyl groups at either end may reduce the molecule's reactivity and potential to cause DNA modification.

# Cluster #22 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 22, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 12.8 ( $\pm 1.6$ ) on the prediction outcome.

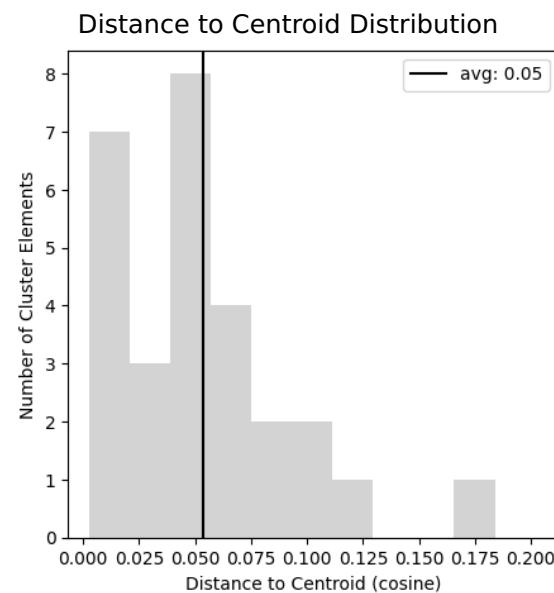
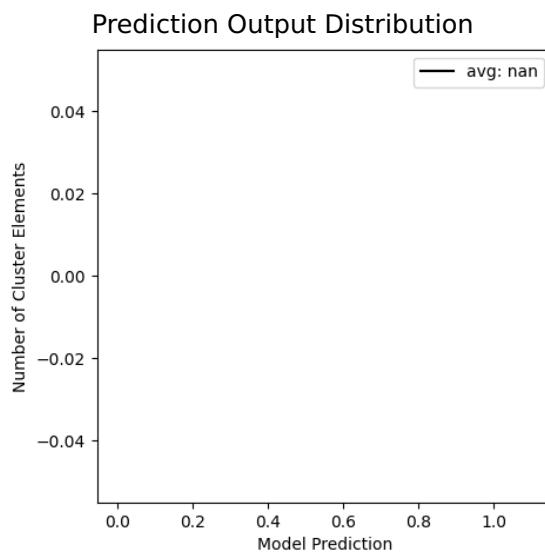
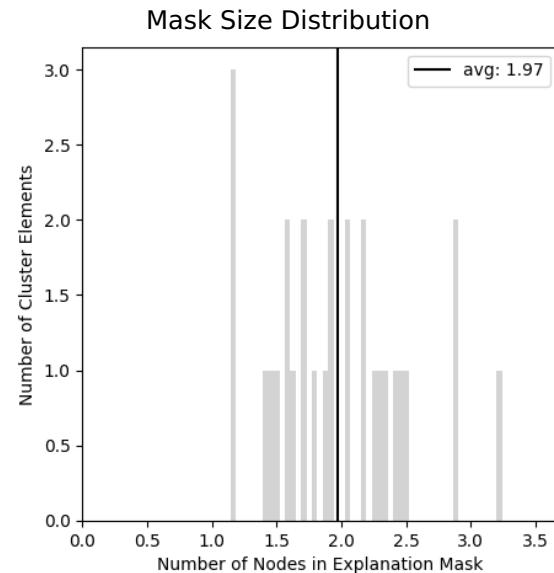
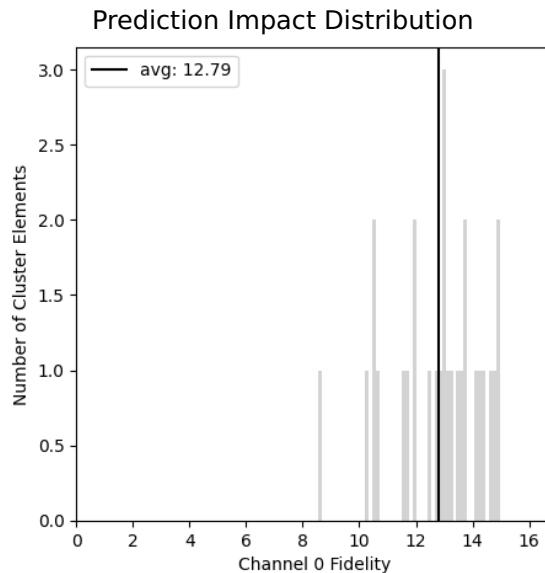
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	28
Channel Index	0.0 (0.0)

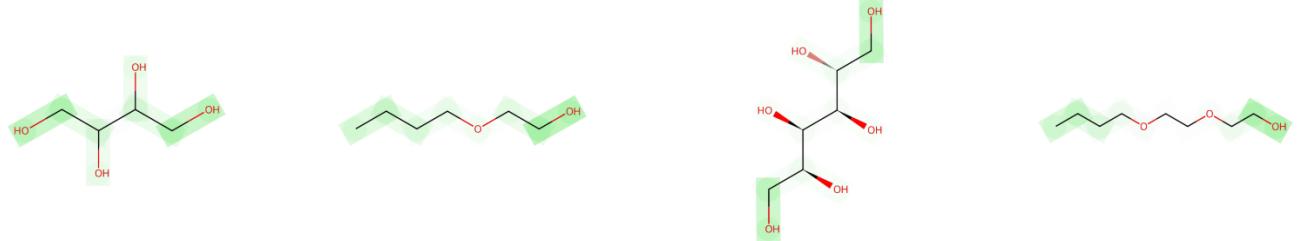
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation "C-C-C-C-O" corresponds to a linear alkyl chain ending with a hydroxyl group. Alkyl groups are typically considered non-reactive and are assumed not to interact significantly with DNA. The hydroxyl group, although it can be a reactive site, in the context of a long alkyl chain might not be sufficiently reactive to cause mutagenic effects. The presumed non-reactivity of this structure can result in a medium influence towards the molecule being non-mutagenic, as it does not contain any significant electrophilic sites or functional groups known to react with DNA bases or the phosphodiester backbone.

**Hypothesis:** Molecules containing the substructure "C-C-C-C-O" are hypothesized to exhibit a medium tendency towards being non-mutagenic. This is due to the inherent non-reactivity of the alkyl chain and the limited reactivity of the terminal hydroxyl group in this context, which reduces the likeliness of interaction with DNA.

# Cluster #23 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 23, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.5 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 14.3 ( $\pm 2.0$ ) on the prediction outcome.

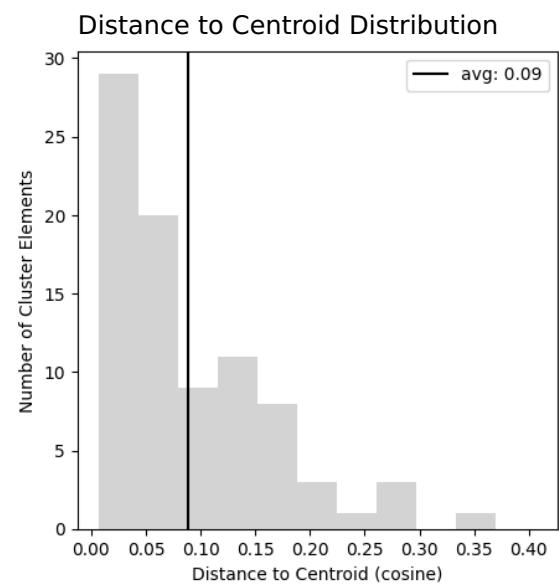
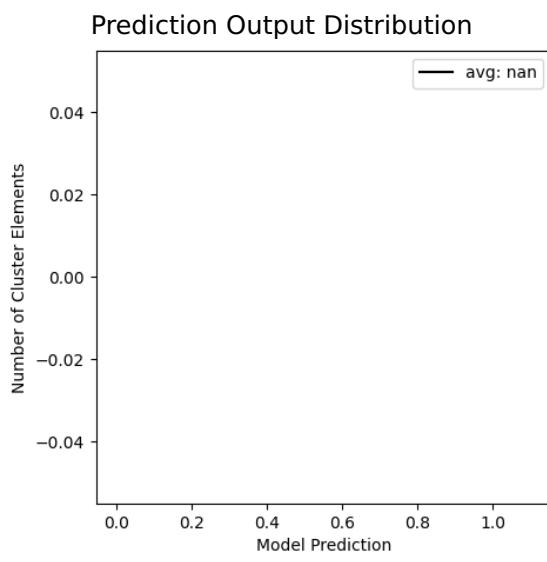
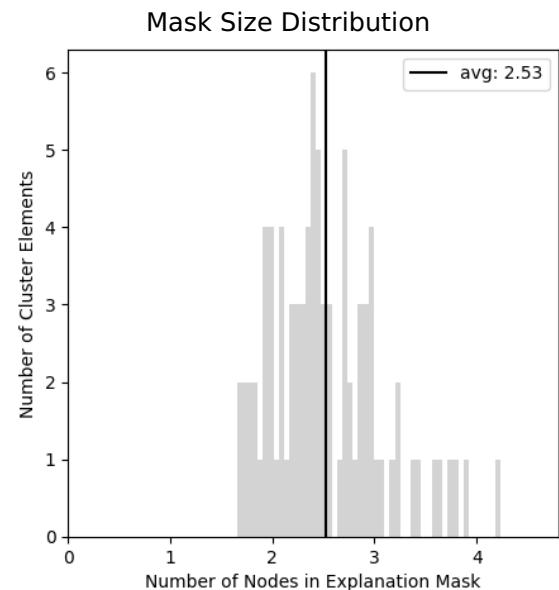
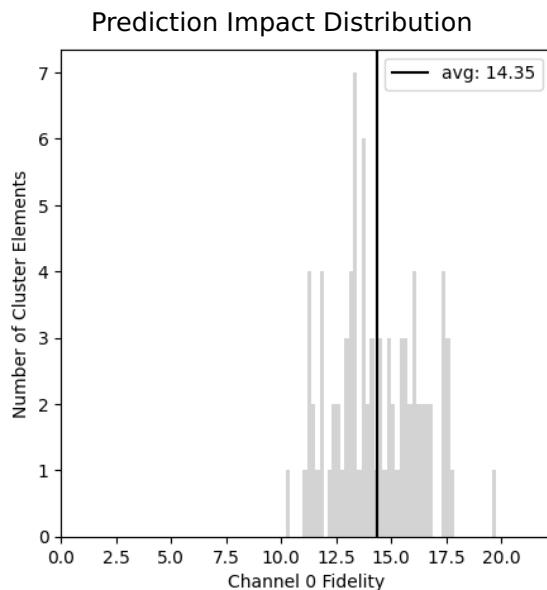
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	85
Channel Index	0.0 (0.0)

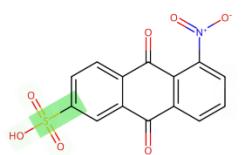
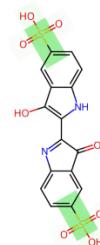
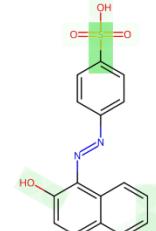
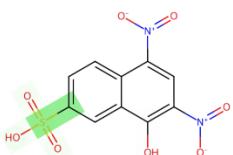
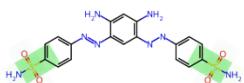
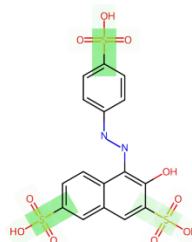
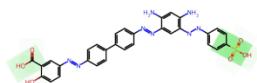
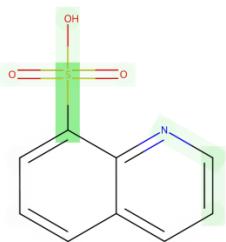
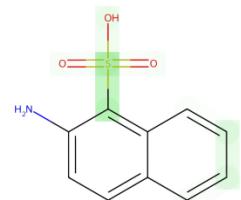
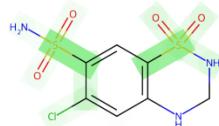
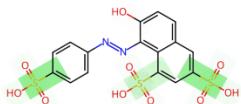
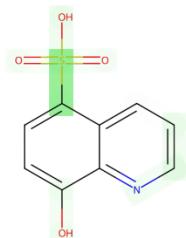
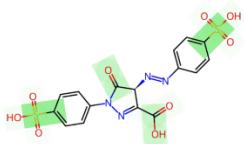
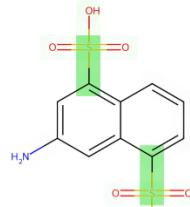
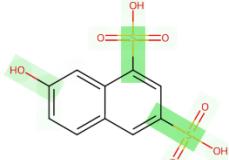
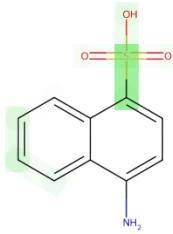
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the nitro group (N-c1) adjacent to a benzene ring in the molecule denotes an electron-withdrawing property that could impact the stability of the molecule. The sulfonyl group (-S(=O)(=O)O) provides further electron-withdrawing effect, which might lead to reduced reactivity and consequently less interaction with genetic material. A molecule that is less reactive is less likely to interact with DNA, reducing its mutagenic potential.

**Hypothesis:** Substructures containing nitro groups adjacent to a benzene ring as well as a sulfonyl side group are hypothesized to be medium influencers towards non-mutagenicity. The electron-withdrawing nature of these groups is hypothesized to stabilize the molecule and reduce its reactive interactions with DNA, thus decreasing the chance for mutagenic events.

# Cluster #24 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 24, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.1 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 13.1 ( $\pm 2.3$ ) on the prediction outcome.

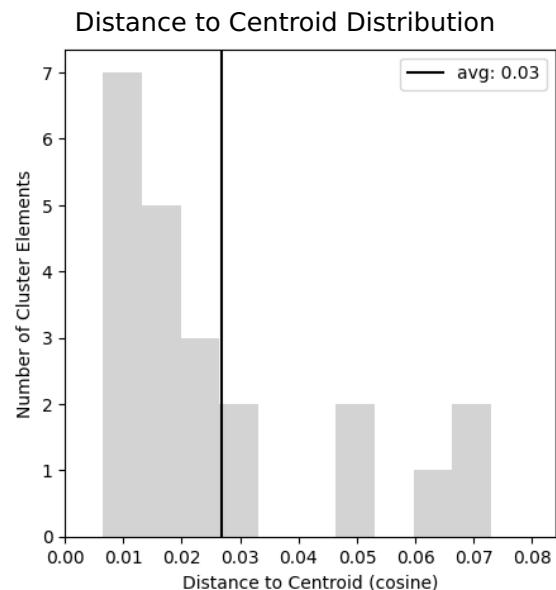
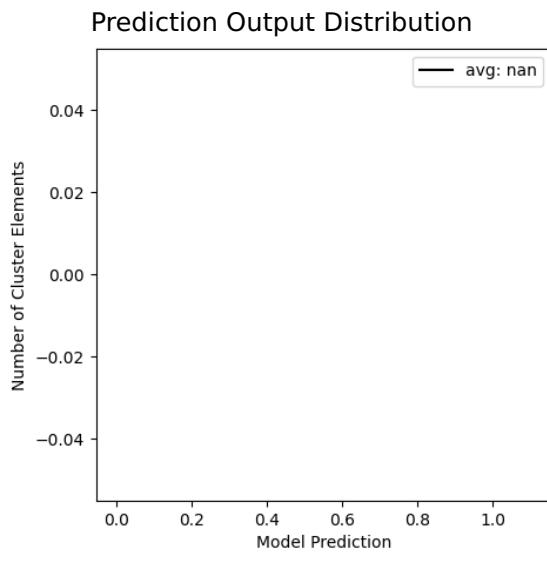
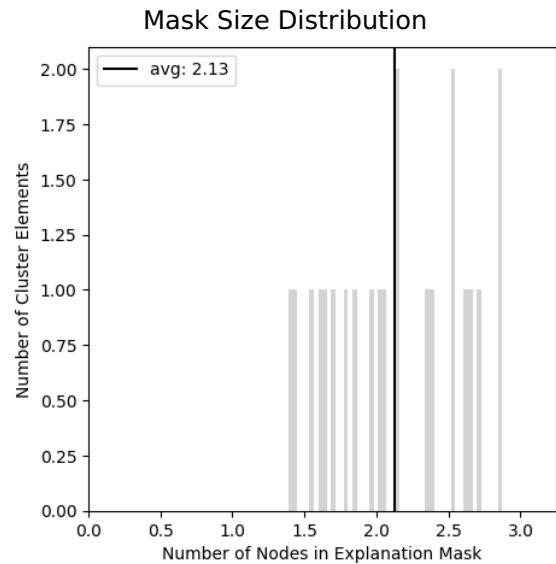
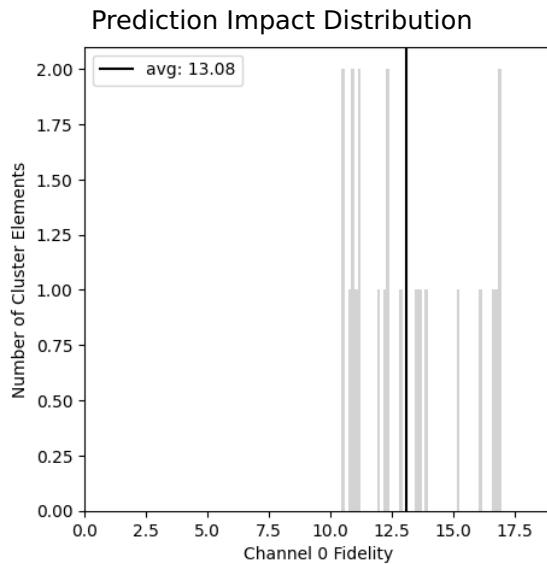
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	22
Channel Index	0.0 (0.0)

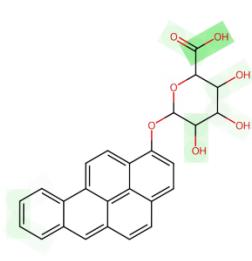
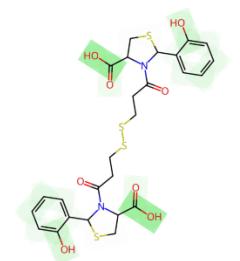
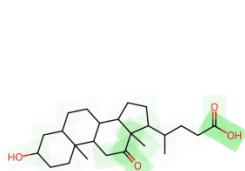
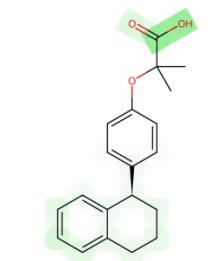
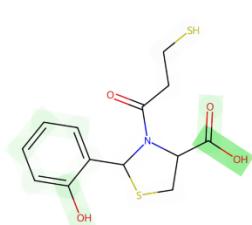
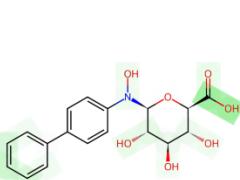
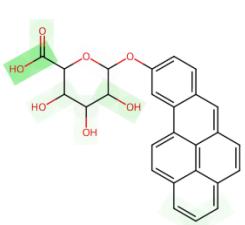
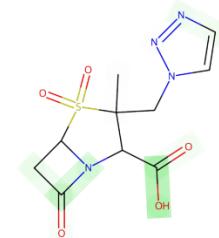
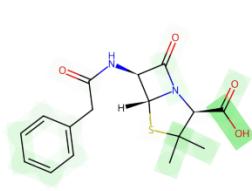
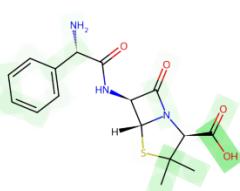
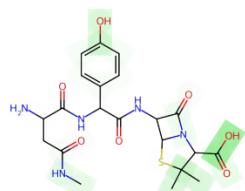
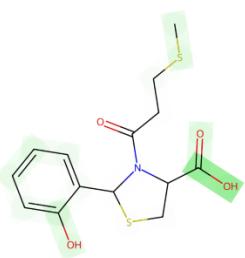
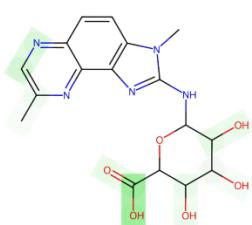
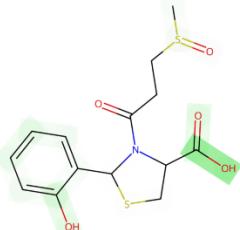
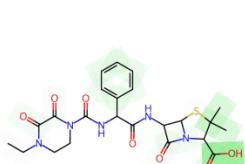
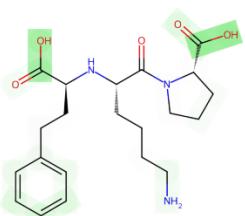
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



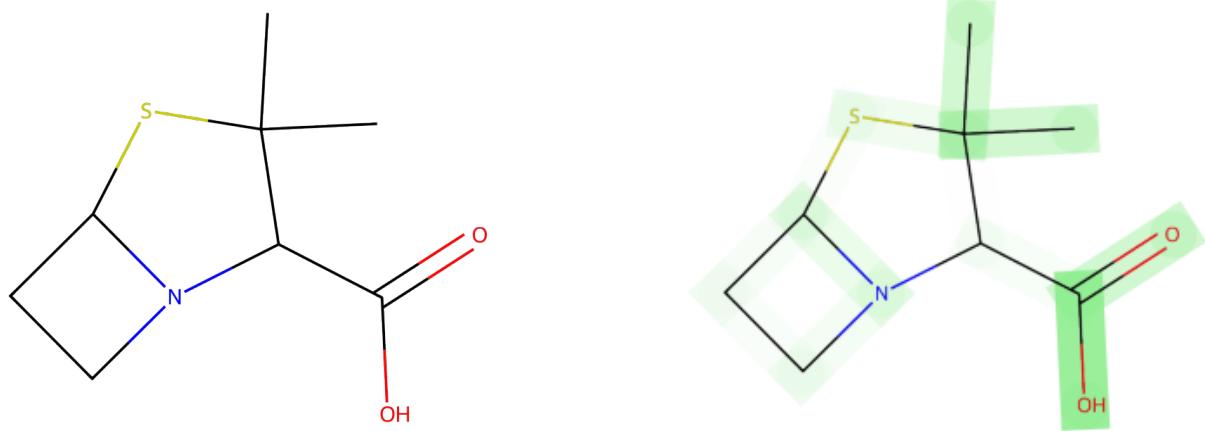
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure indicated by the provided SMILES contains a thioether group (-S-) linking two cyclic systems, and a carboxylic acid group (-C(=O)-O). Thioethers are relatively stable and have low reactivity, which means they are less likely to interact with DNA. The carboxylic acid group is commonly found in biomolecules and is generally not associated with mutagenicity. Their presence in a molecule could lead to increased solubility and better excretion from the body, reducing the exposure time of the genetic material to potentially mutagenic agents. In addition, the bicyclic (two-ring) system might provide some steric hindrance, potentially reducing the molecule's ability to intercalate into DNA.

**Hypothesis:** The presence of a thioether and a carboxylic acid group in a bicyclic structure confers a medium influence towards non-mutagenicity. The stability of the thioether and the biocompatibility of the carboxylic acid, as well as the steric effects of the bicyclic system, likely reduce the chances of the molecule causing genetic mutations.

# Cluster #25 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 25, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.1 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 11.1 ( $\pm 1.7$ ) on the prediction outcome.

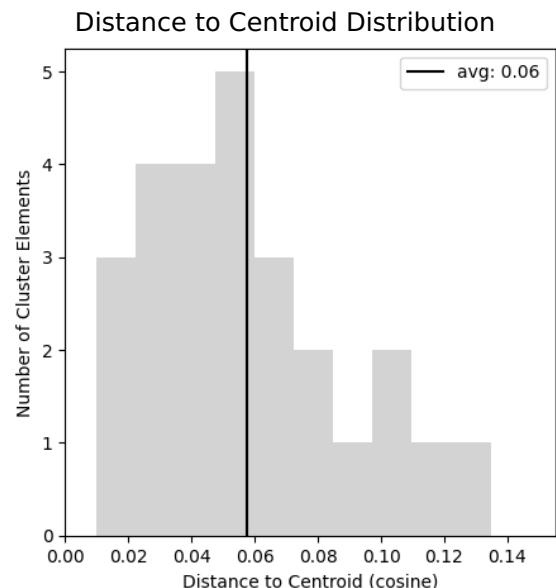
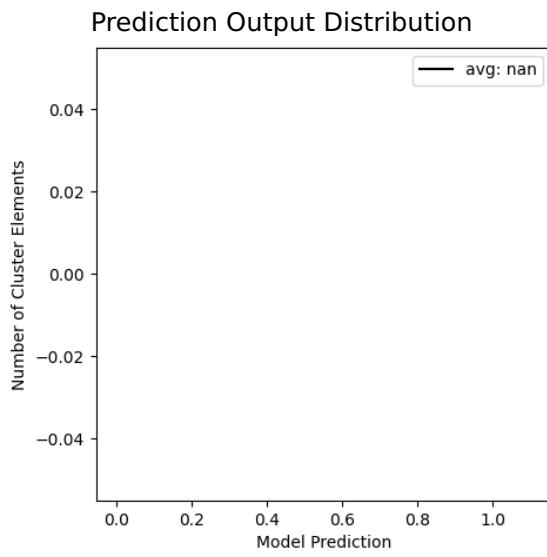
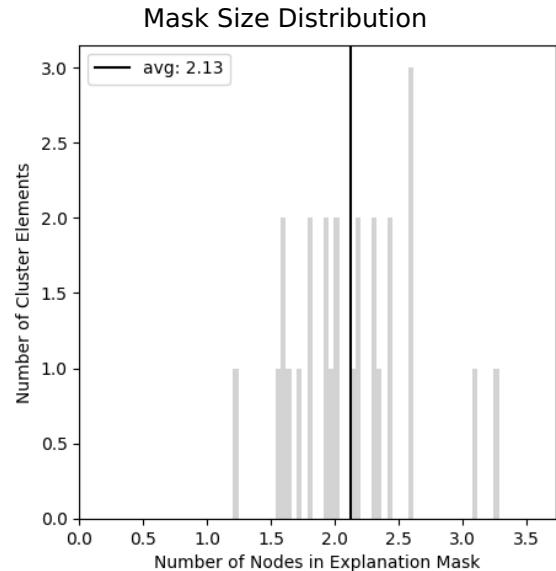
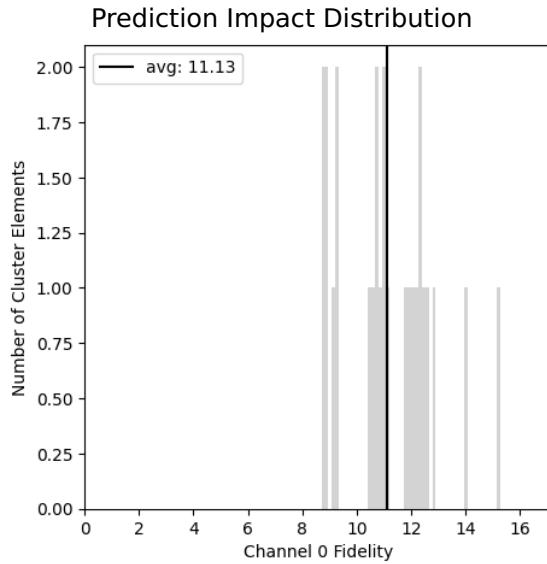
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	26
Channel Index	0.0 (0.0)

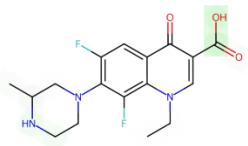
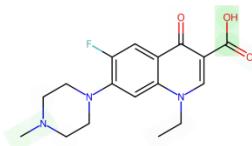
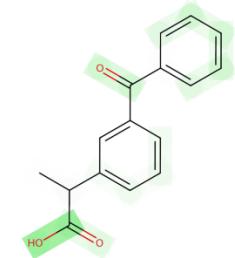
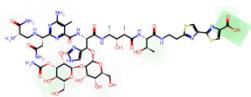
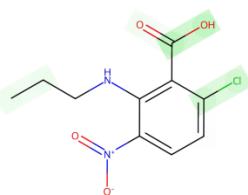
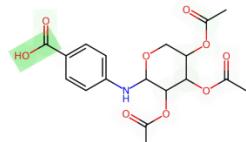
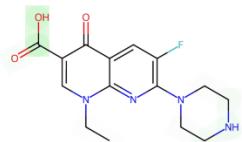
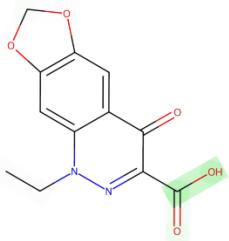
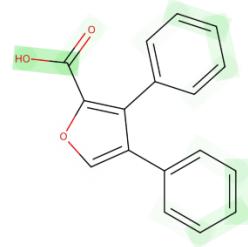
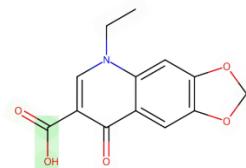
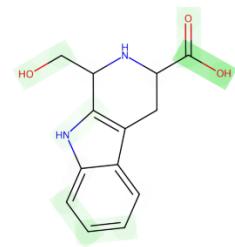
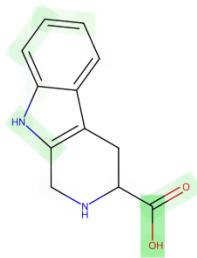
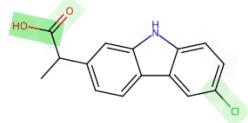
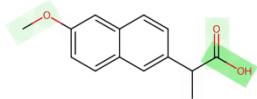
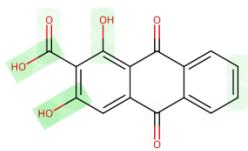
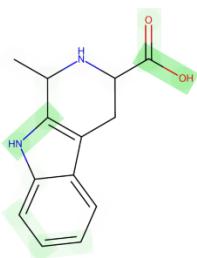
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



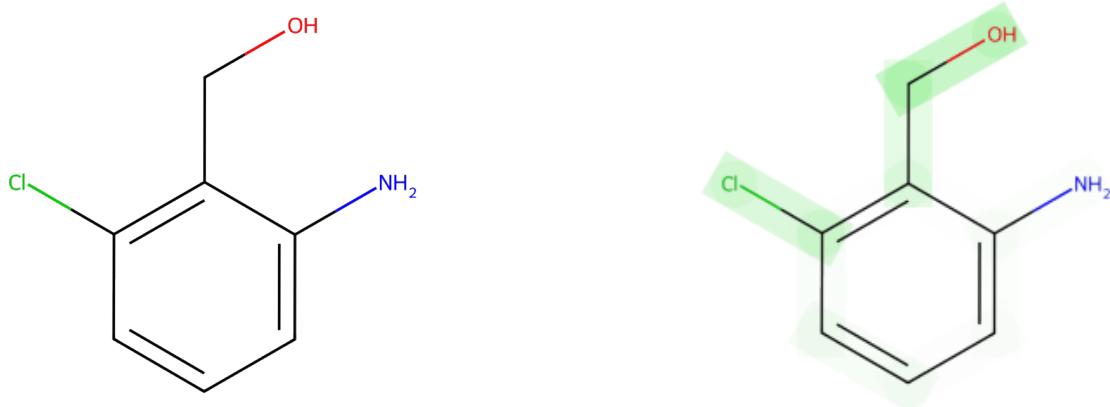
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation describes a benzene ring with a chlorine atom and a hydroxymethyl group attached on opposite ends, a nitro group (N=O) is also present. The presence of the chlorine atom could potentially make this molecule act as an electrophile which could form adducts with DNA, leading to mutagenicity. However, the hydroxymethyl group, being a donor of electron density, could stabilize the molecule and reduce its reactivity, decreasing its potential to interact with and alter DNA. Furthermore, the nitro group could contribute to reducing the mutagenic effect by drawing electron density away from the benzene ring, thus lessening its potential to form DNA adducts.

**Hypothesis:** The molecular substructure represented by "N-c1ccccc1Cc2cc(Cl)cc(O)c2" has a medium influence towards "non-mutagenic" properties. The electron-withdrawing effect of the nitro group along with the electron-donating influence of the hydroxymethyl group may work in tandem to reduce the overall reactivity of the molecule, potentially decreasing its mutagenic capability. The presence of the chlorine atom suggests a possible mutagenic effect, but the other substituents may mitigate this property.

# Cluster #26 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 26, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 10.7 ( $\pm 2.5$ ) on the prediction outcome.

## Properties

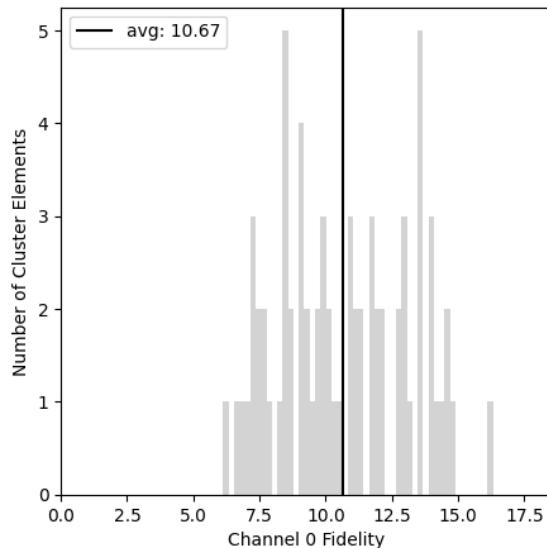
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	70
Channel Index	0.0 (0.0)

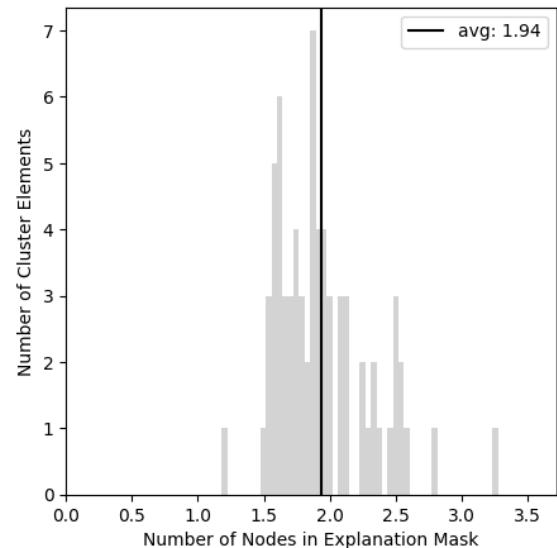
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

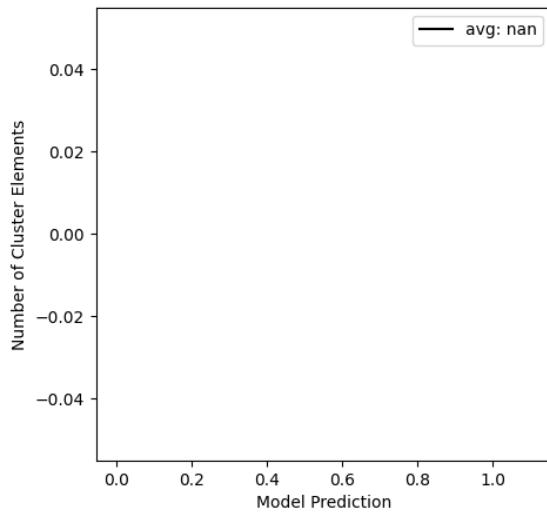
Prediction Impact Distribution



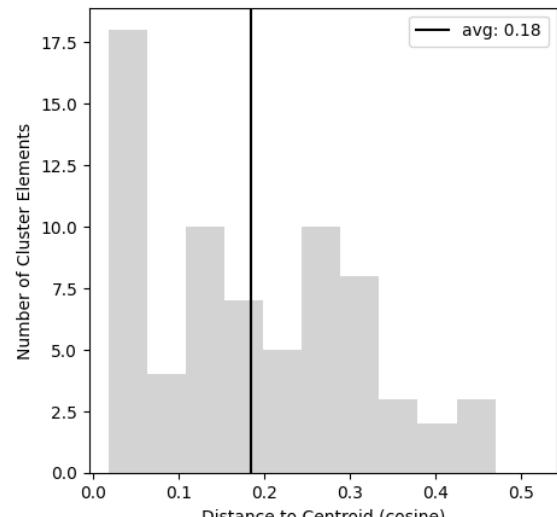
Mask Size Distribution



Prediction Output Distribution

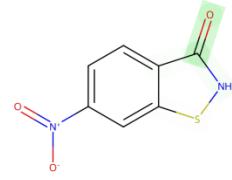
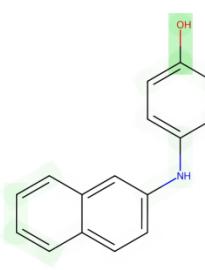
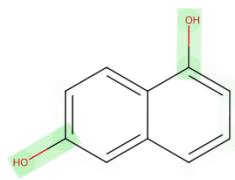
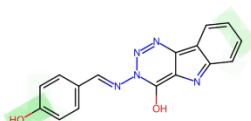
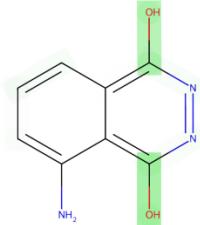
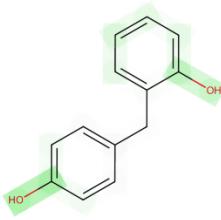
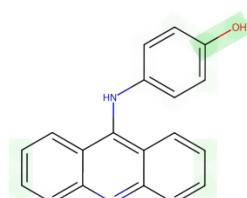
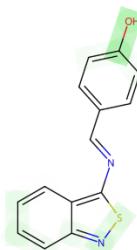
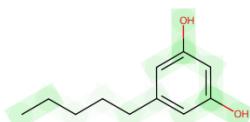
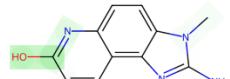
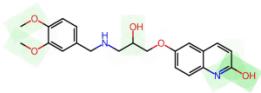
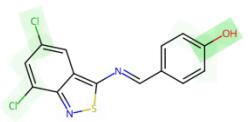
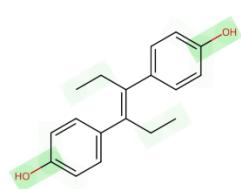
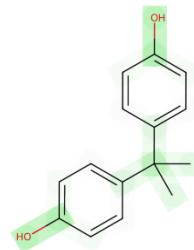
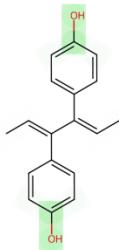
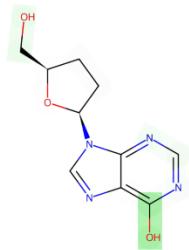


Distance to Centroid Distribution



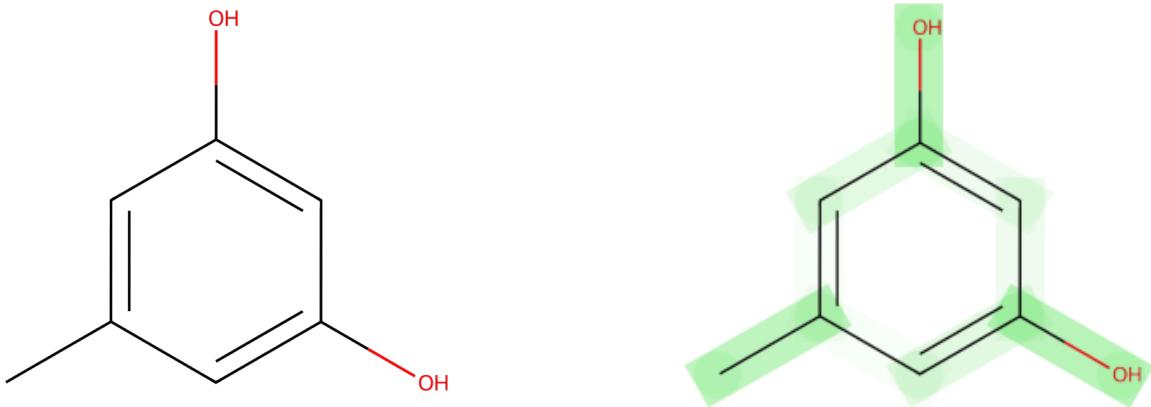
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a phenolic compound, which features a benzene ring with two hydroxyl (-OH) groups in the ortho position. The hydroxyl groups can participate in extensive hydrogen bonding, which could stabilize the DNA structure and reduce its susceptibility to mutagenic agents. Additionally, the electron-donating effect of the hydroxyl group ortho to each other could lead to delocalization of electrons, reducing the potential for the molecule to form reactive, mutagenic intermediates.

**Hypothesis:** The presence of two ortho hydroxyl groups on a benzene ring is associated with a medium influence towards non-mutagenicity. The enhanced hydrogen bonding capability and electron delocalization induced by these -OH groups may stabilize genetic material and decrease the propensity for mutagenicity.

# Cluster #27 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 27, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 13.3 ( $\pm 1.3$ ) on the prediction outcome.

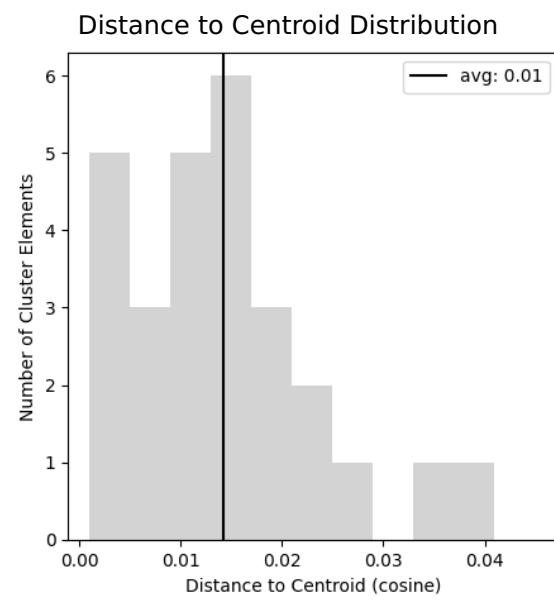
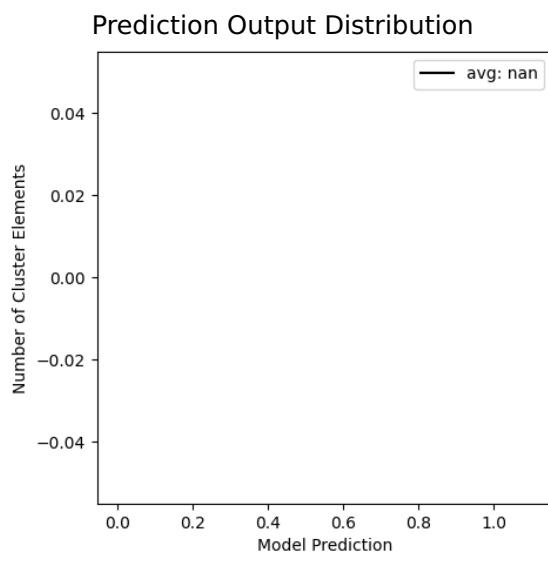
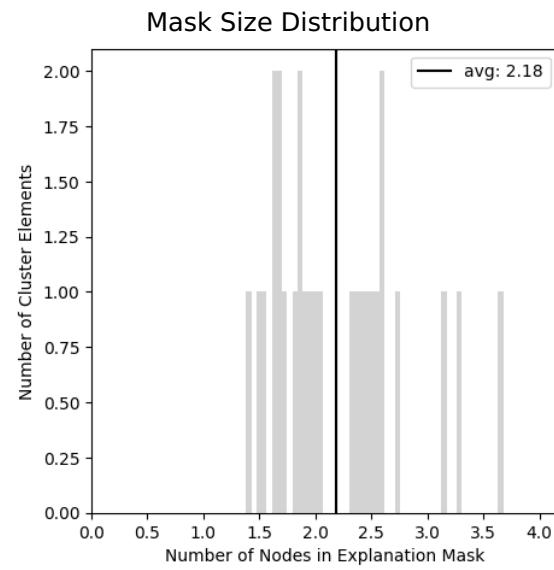
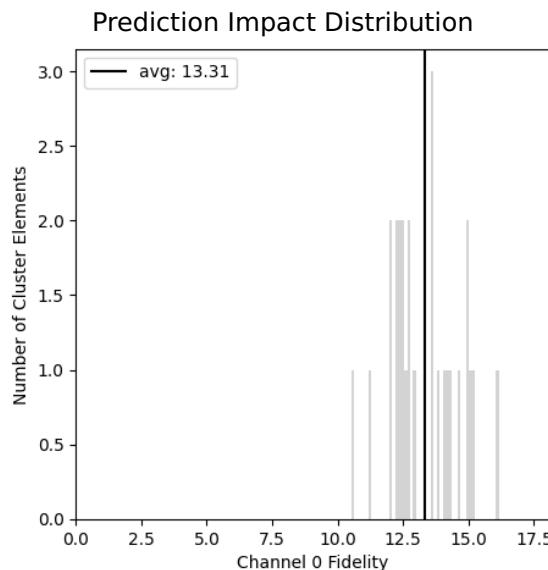
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	27
Channel Index	0.0 (0.0)

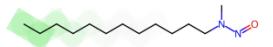
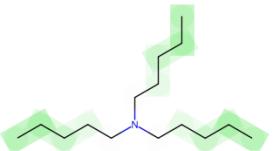
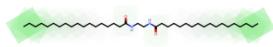
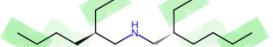
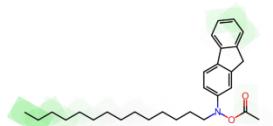
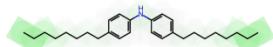
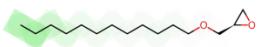
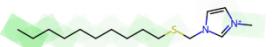
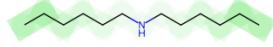
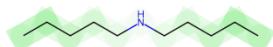
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



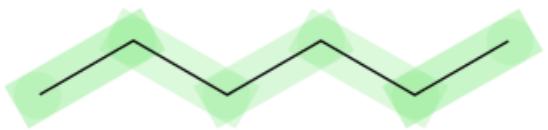
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES "C-C-C-C-C-C" represents a linear or straight chain alkane composed of six carbon atoms. Alkanes are typically non-reactive due to their strong C-C and C-H bonds and lack of functional groups that can interact with DNA or other genetic material. The absence of double or triple bonds, rings, or electronegative atoms in this structure makes it less likely to engage in reactions that could lead to genetic alterations. Hence, the molecular structure is believed to confer a certain level of stability, reducing its mutagenic potential.

**Hypothesis:** A straight-chain alkane with six carbon atoms tends to be non-mutagenic. This is likely because it does not contain any reactive functional groups or structures that can easily interact with or damage DNA. The strong and stable C-C and C-H bonds lower the probability of participating in reactions that could cause genetic modifications.

# Cluster #28 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 28, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.8$ ) nodes. The concept is generally associated with an impact of 8.4 ( $\pm 2.0$ ) on the prediction outcome.

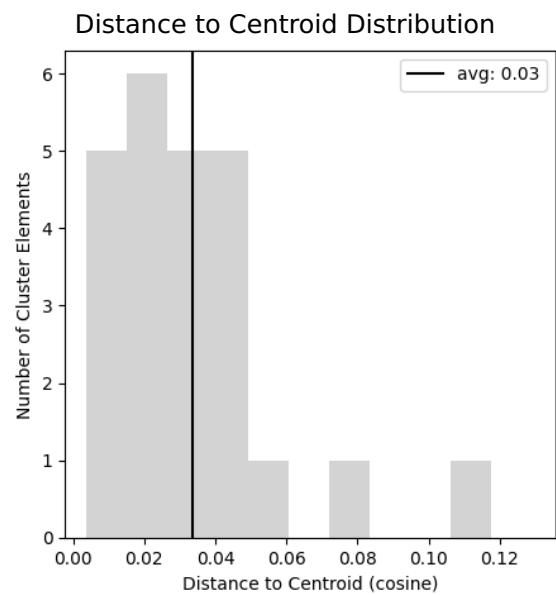
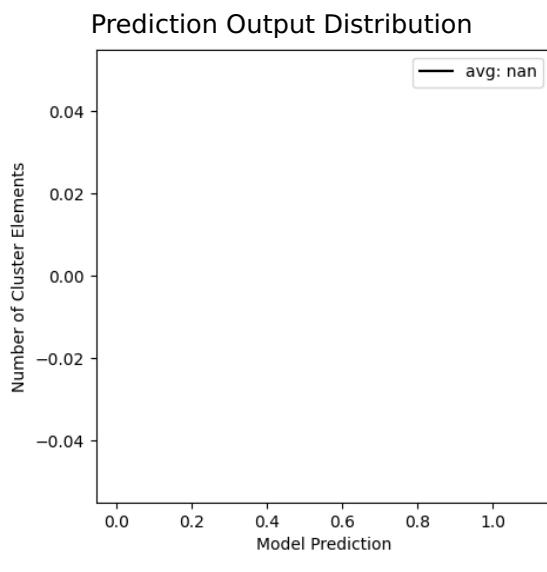
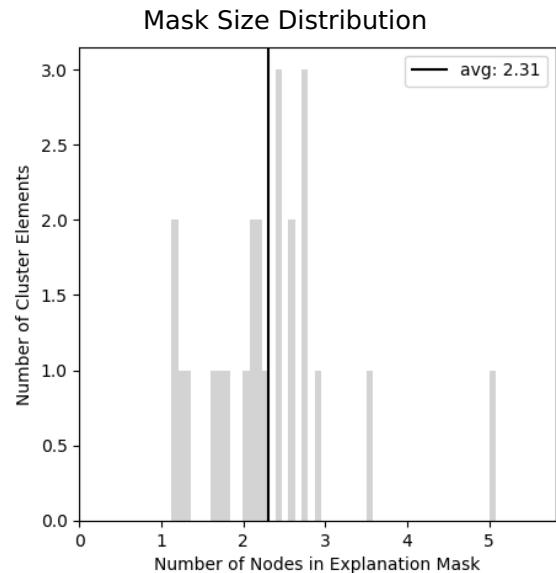
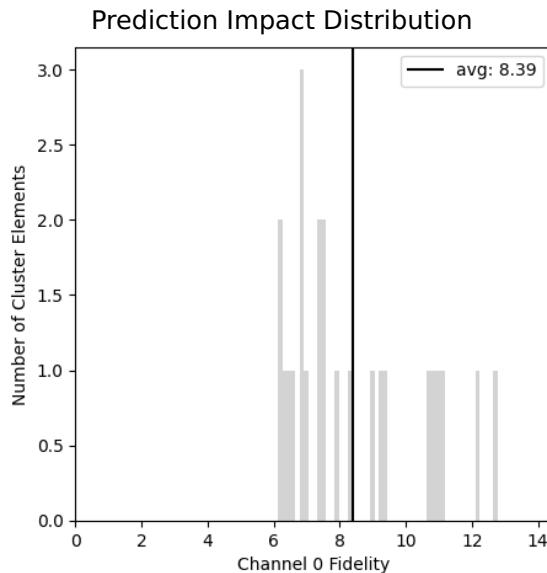
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	0.0 (0.0)

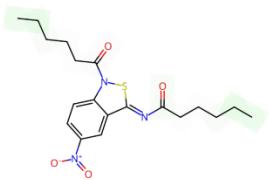
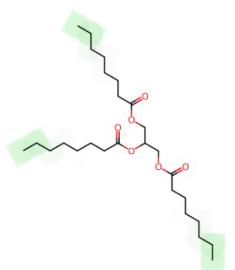
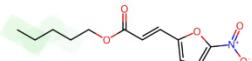
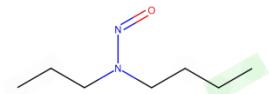
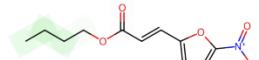
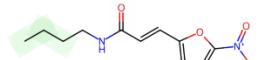
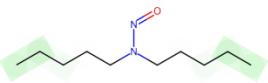
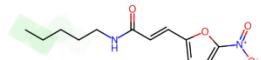
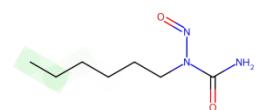
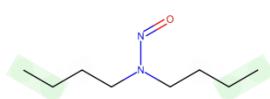
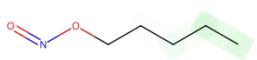
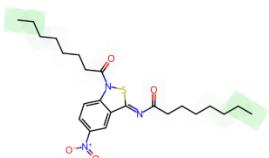
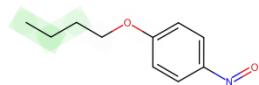
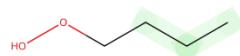
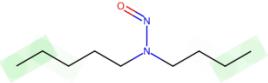
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure represented by the SMILES "C-C-C-O-O" consists of a linear chain of three carbon atoms followed by two oxygen atoms. The linear carbon chain without the presence of any functional group associated with reactivity suggests a lack of reactive sites that could interact with DNA to cause mutations. Additionally, the presence of two oxygen atoms at the end of the chain may imply the formation of a peroxide linkage, which can be relatively stable under physiological conditions unless other structural features that promote reactivity are present.

**Hypothesis:** The "C-C-C-O-O" substructure presents a **SMALL** influence towards "non-mutagenicity" due to the absence of highly reactive functional groups and the potential stability of the peroxide linkage under normal biological conditions.

# Cluster #29 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 29, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 6.6 ( $\pm 0.7$ ) on the prediction outcome.

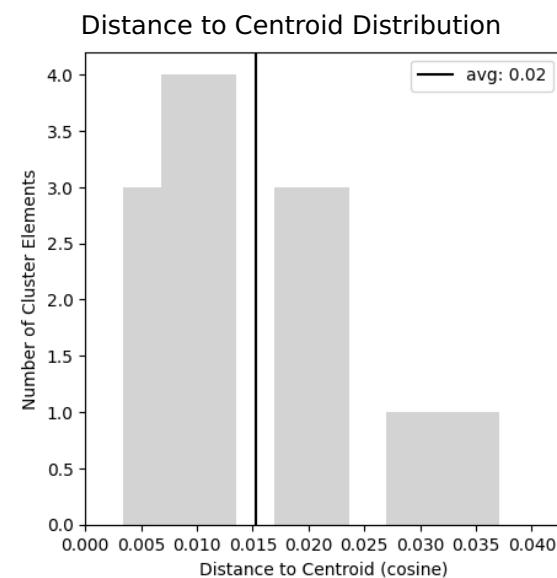
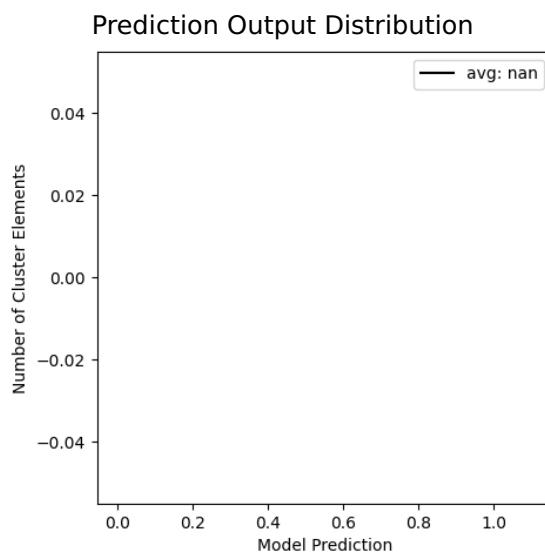
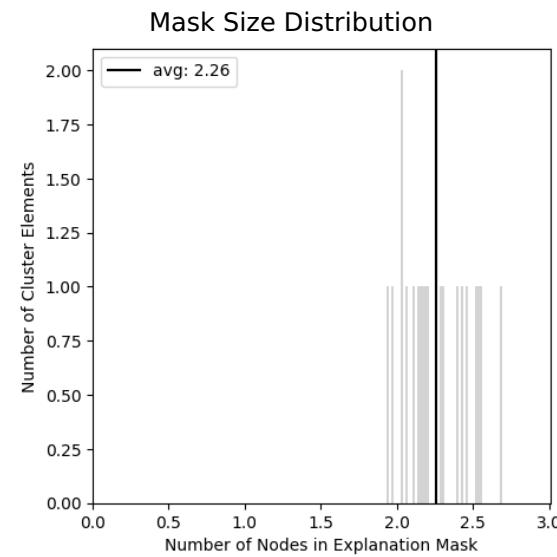
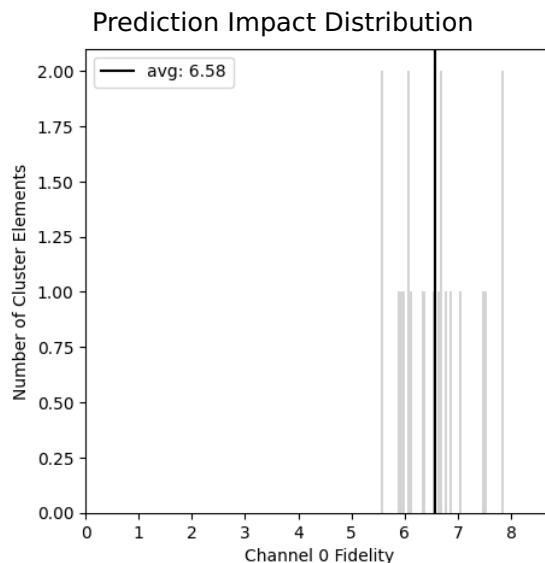
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	0.0 (0.0)

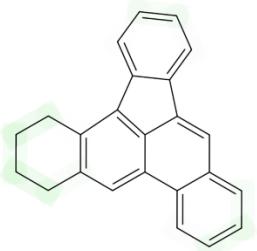
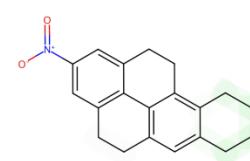
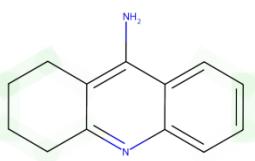
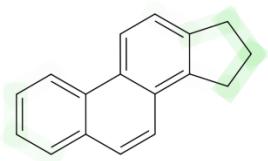
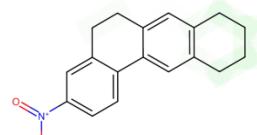
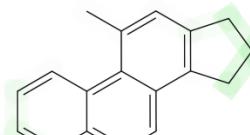
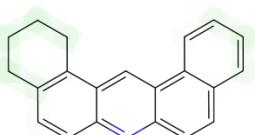
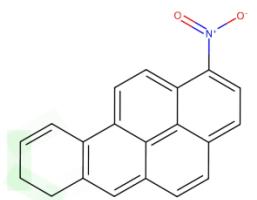
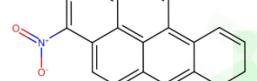
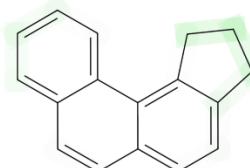
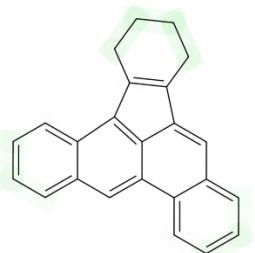
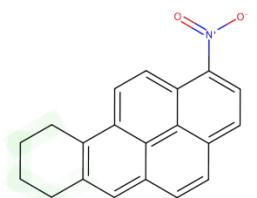
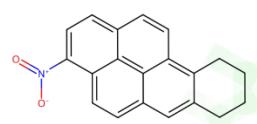
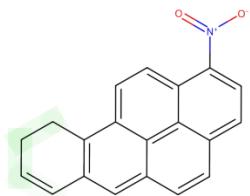
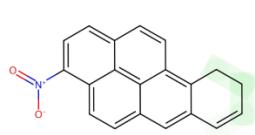
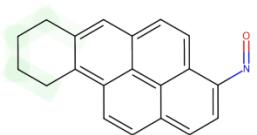
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a benzene ring with a propyl group (C-C-C) and an N-O group attached to it. The benzene ring itself is known to be stable and less reactive due to its aromaticity. The addition of a propyl group doesn't significantly disrupt this stability. Moreover, the presence of the -[N<sup>+</sup>]-[O<sup>-</sup>] group (nitroso group) suggests a potential for electron donation and withdrawal, respectively, which could potentially stabilize any reactive intermediates that might form and therefore reduce the likelihood of the molecule interacting with DNA.

**Hypothesis:** Molecules featuring the substructure "C-C-C-c1:c:c(-C):c:c(-[N+]-[O-]):c:1" tend to be non-mutagenic. The stability imparted by the aromatic benzene ring and the electron-donating and withdrawing effects of the propyl and nitroso groups, respectively, reduce the molecule's ability to interact with and cause mutations in DNA.

# Cluster #30 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 30, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.8 ( $\pm 1.1$ ) nodes. The concept is generally associated with an impact of 7.8 ( $\pm 1.5$ ) on the prediction outcome.

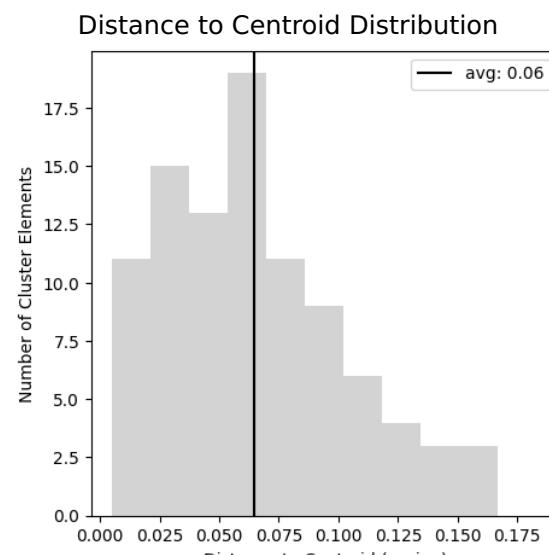
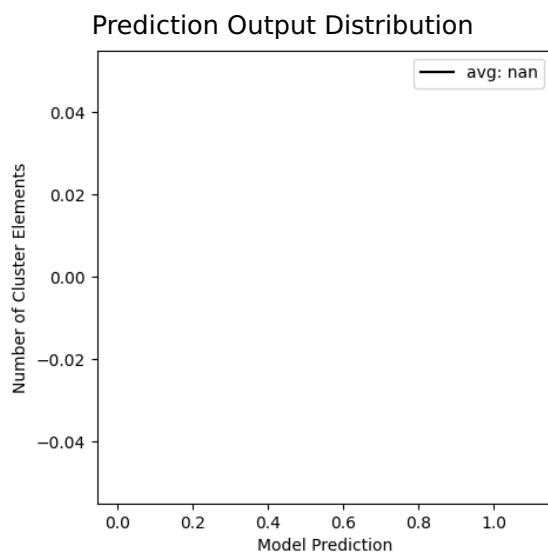
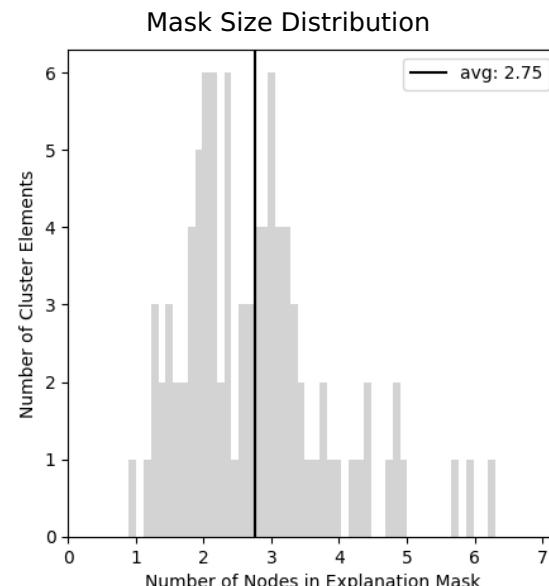
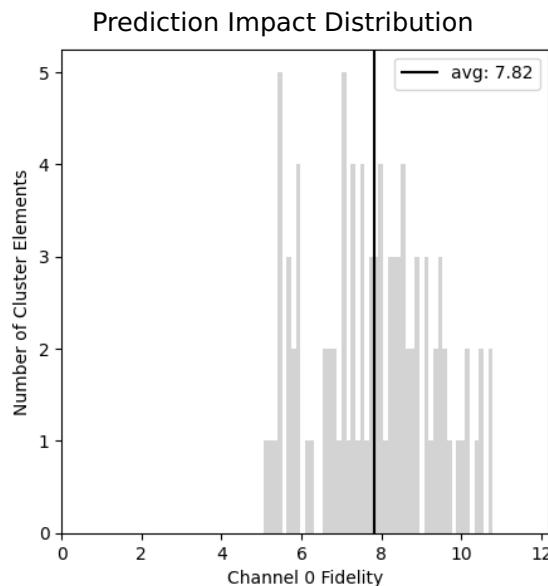
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	94
Channel Index	0.0 (0.0)

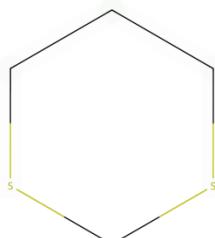
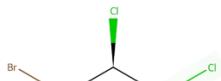
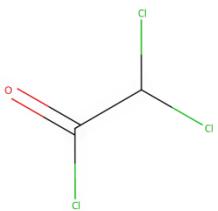
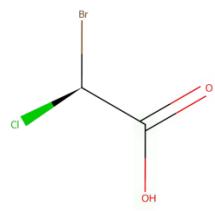
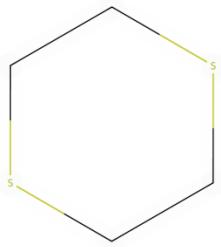
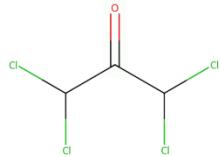
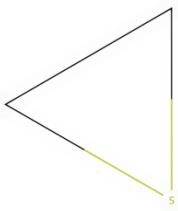
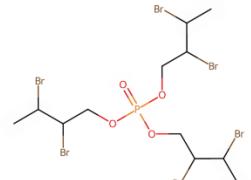
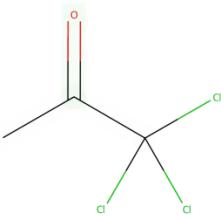
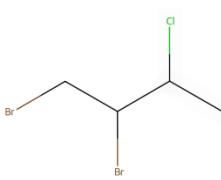
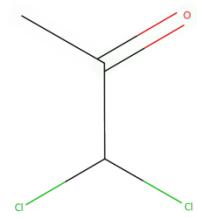
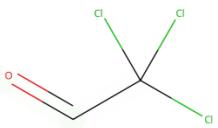
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



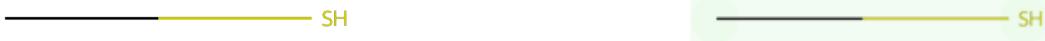
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a carbon-sulfur (C-S) bond suggests a thiol or sulfide linkage which generally lacks a highly reactive functionality that would be necessary to interact with DNA and cause mutations. Thiols and sulfides are known for their relative stability, and without additional functional groups or structural features that might compromise this stability, the C-S bond itself is less likely to participate in reactions that would lead to genetic modification. The lack of highly electronegative atoms or significant delocalization of electrons in this fragment means it is less likely to form electrophilic or nucleophilic species that can interact with DNA bases.

**Hypothesis:** Molecules containing the C-S substructure tend to exhibit minimal mutagenic activity due to the stability and low reactivity of this bond. The sulfur atom does not significantly distort the electron distribution in the structure, and without additional reactive sites, the C-S bond does not readily engage in the type of chemical interactions necessary to alter DNA.

# Cluster #31 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 31, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 10.4 ( $\pm 1.2$ ) on the prediction outcome.

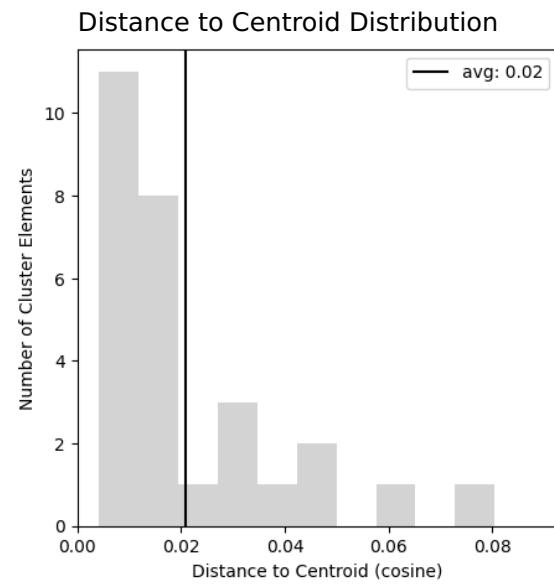
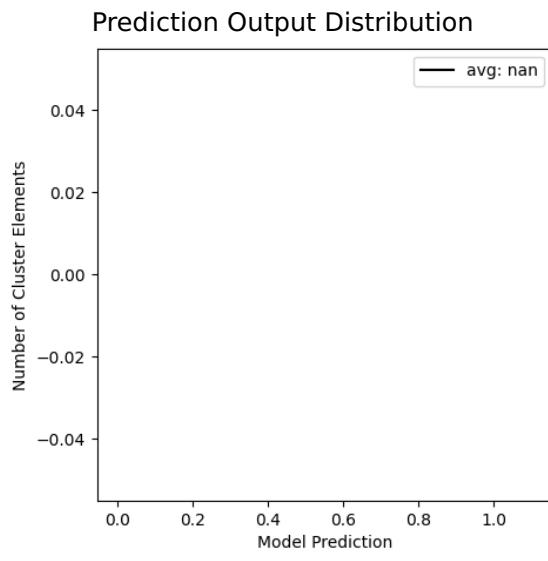
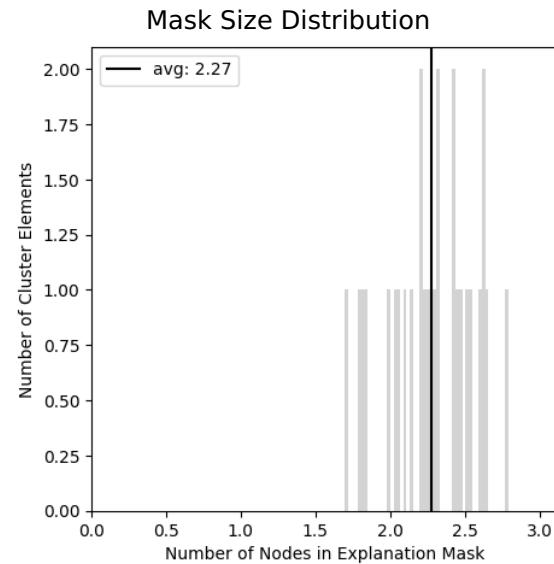
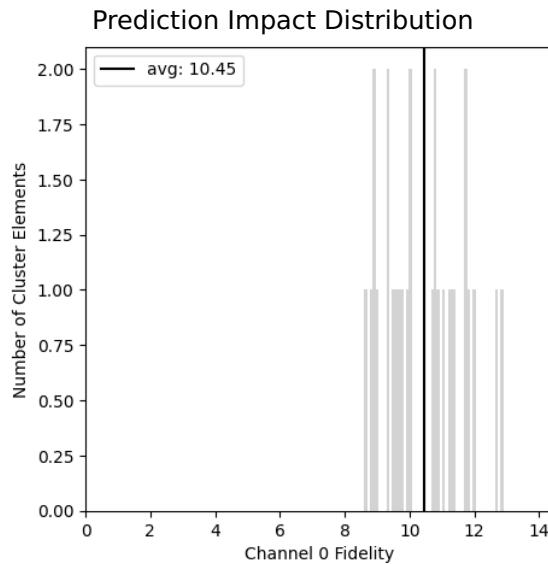
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	28
Channel Index	0.0 (0.0)

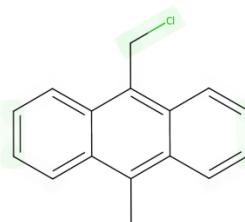
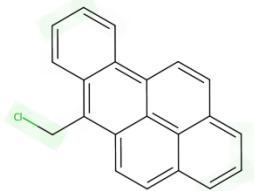
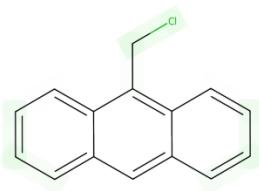
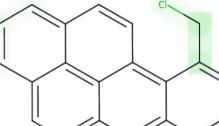
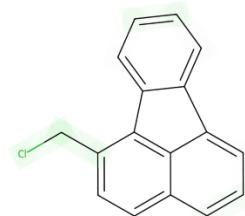
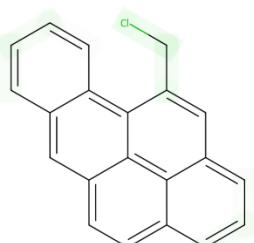
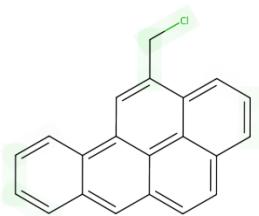
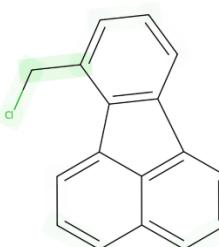
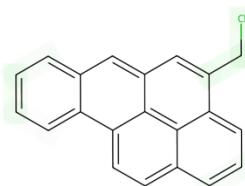
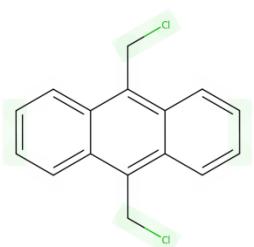
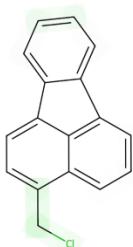
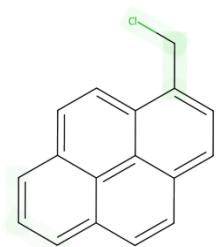
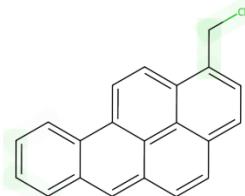
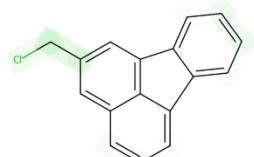
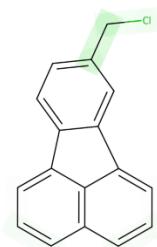
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



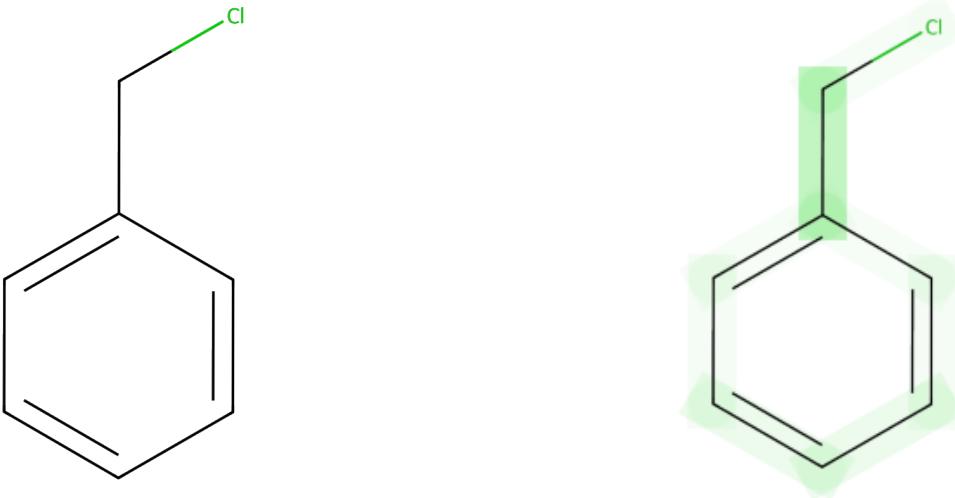
# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a molecule consisting of a benzene ring connected to a chloromethyl group. A benzene ring is a stable, non-reactive aromatic system that typically does not interact strongly with DNA unless it has been activated by a certain group. In this case, the chloromethyl group attached to the benzene ring could be hypothesized to reduce the reactivity of the system. Chlorine is an electronegative atom and may withdraw electron density from the benzene ring, thereby decreasing its potential to form electrophilic metabolites that are capable of DNA interaction, which could lead to mutations. Moreover, the chloromethyl group is not a large appendage, suggesting that steric hindrance is not significantly increased which could otherwise escalate the mutagenic potential through physical interaction with DNA.

**Hypothesis:** The presence of a chloromethyl group attached to a benzene ring leads to a medium influence towards "non-mutagenic" properties in the molecule. The electronegative chlorine atom could withdraw electron density from the benzene ring thus reducing its reactivity, while the small size of the chloromethyl group does not significantly increase the likelihood of physical interaction with DNA.

# Cluster #32 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 32, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.1 ( $\pm 0.7$ ) nodes. The concept is generally associated with an impact of 9.6 ( $\pm 2.2$ ) on the prediction outcome.

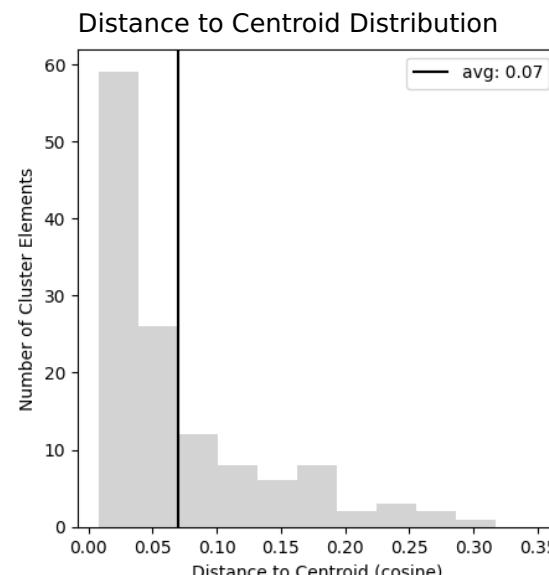
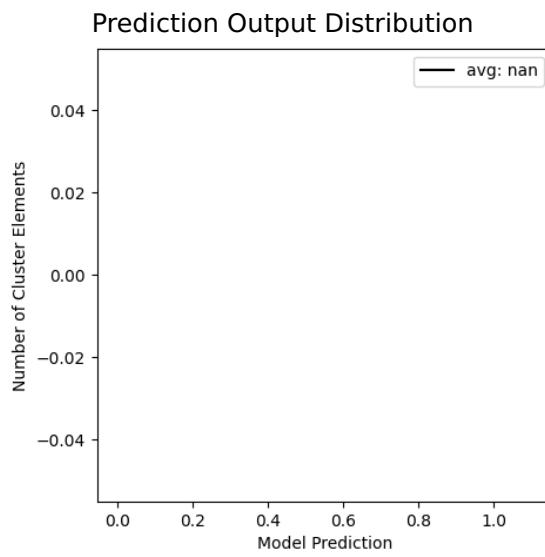
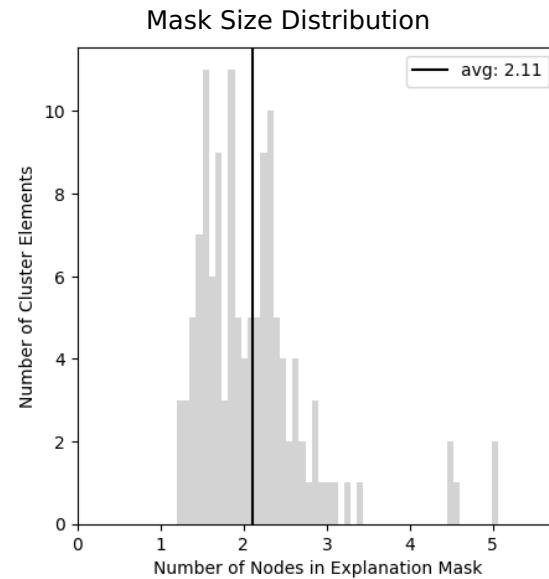
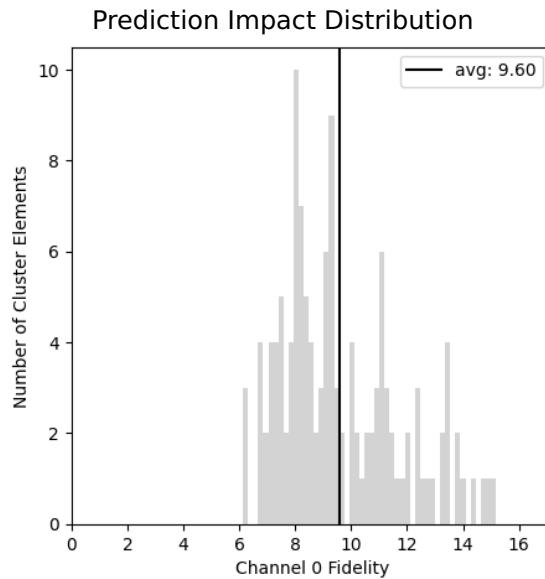
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	127
Channel Index	0.0 (0.0)

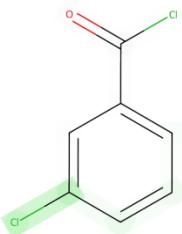
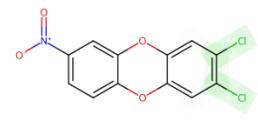
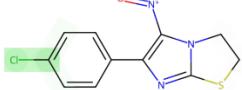
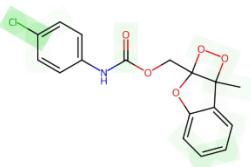
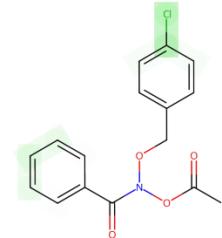
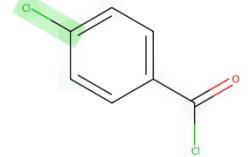
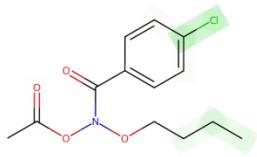
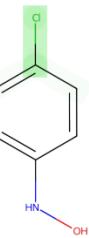
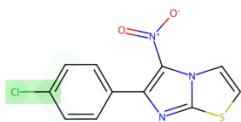
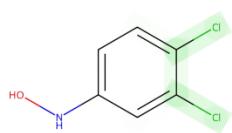
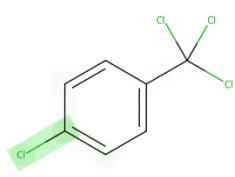
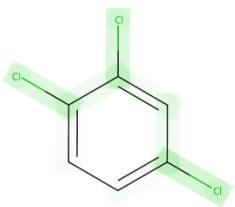
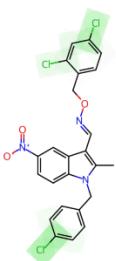
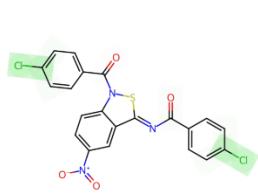
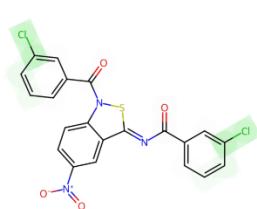
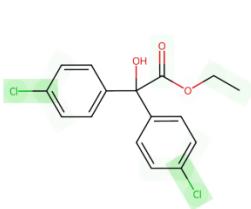
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a chlorinated aromatic amine. Aromatic amines are known to undergo oxidation to form hydroxylamines, which can be mutagenic. However, the presence of a chlorine atom on the aromatic ring can potentially withdraw electron density through inductive and resonance effects, leading to a decrease in the reactivity of the amine group. This diminished reactivity makes the formation of mutagenic metabolites less likely, thereby reducing mutagenicity.

**Hypothesis:** The presence of a chlorine atom on an aromatic amine reduces its mutagenicity. The electronegative chlorine decreases the electron density on the amine, making it less reactive and less likely to form mutagenic metabolites. This structural feature likely contributes to the observed low mutagenic activity.

# Cluster #33 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 33, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 13.3 ( $\pm 2.3$ ) on the prediction outcome.

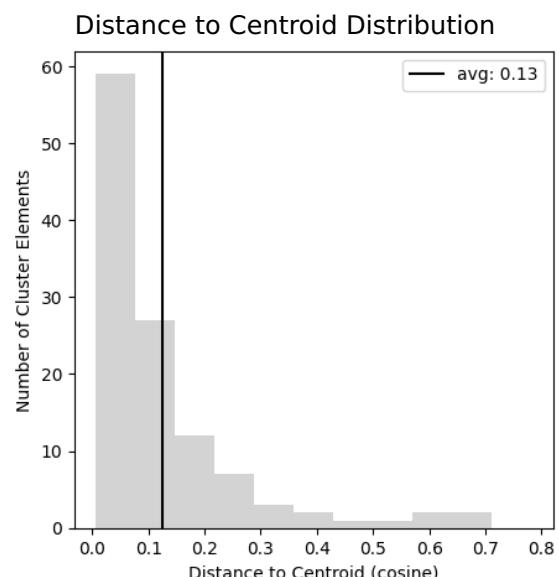
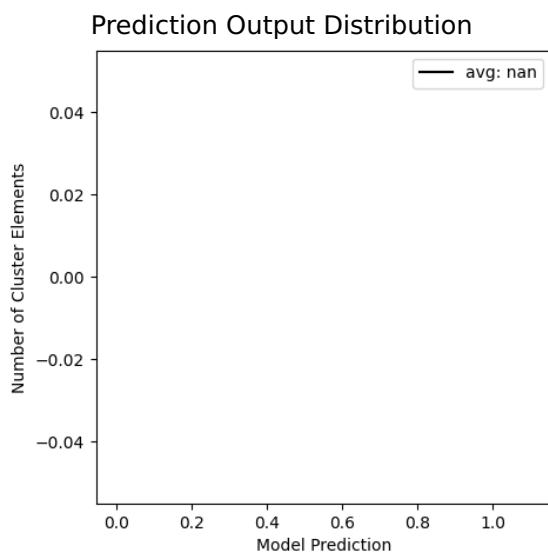
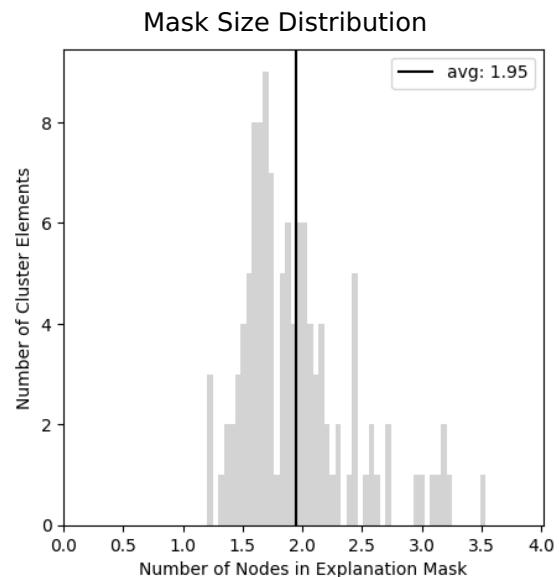
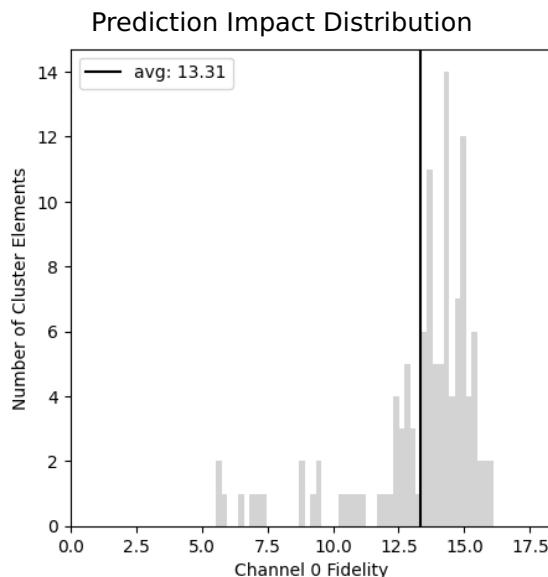
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	116
Channel Index	0.0 (0.0)

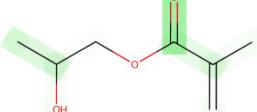
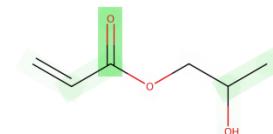
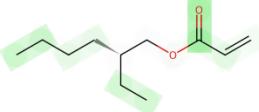
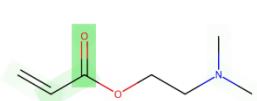
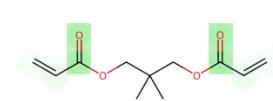
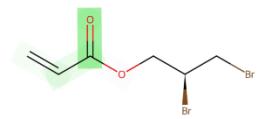
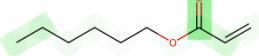
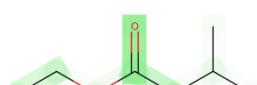
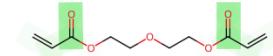
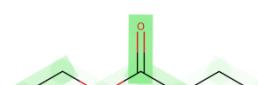
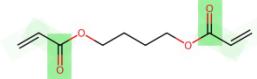
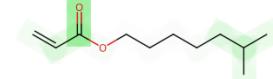
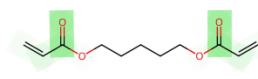
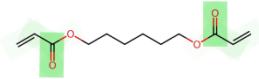
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The functional group delineated by the SMILES code "C-C-C-C=O" represents a butyraldehyde structure - a four-carbon chain with an aldehyde (carbonyl group bonded to a terminal carbon). Aldehydes generally can form reversible covalent bonds with nucleophilic sites in DNA or proteins but the medium-length carbon chain in butyraldehyde may confer sufficient steric hindrance, reducing the electrophilic character of the carbonyl carbon and thus its ability to interact with genetic material. Additionally, the lack of high energy or highly reactive functional groups in the given structure suggests it does not readily participate in reactions that could lead to mutagenic events.

**Hypothesis:** Butyraldehyde, as represented by "C-C-C-C=O," is hypothesized to have a medium influence towards being non-mutagenic due to its moderate carbon chain length and presence of an aldehyde group, which potentially limits its interaction with DNA. Steric hindrance and lower reactivity of the carbon chain are the likely reasons inhibiting the molecule's mutagenic capability.

# Cluster #34 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 34, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 15.1 ( $\pm 0.8$ ) on the prediction outcome.

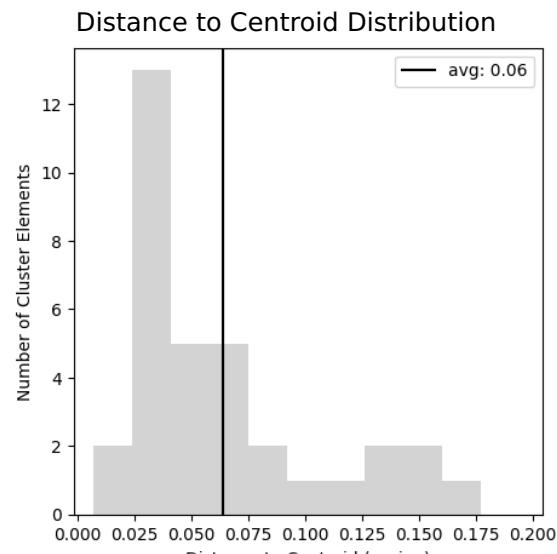
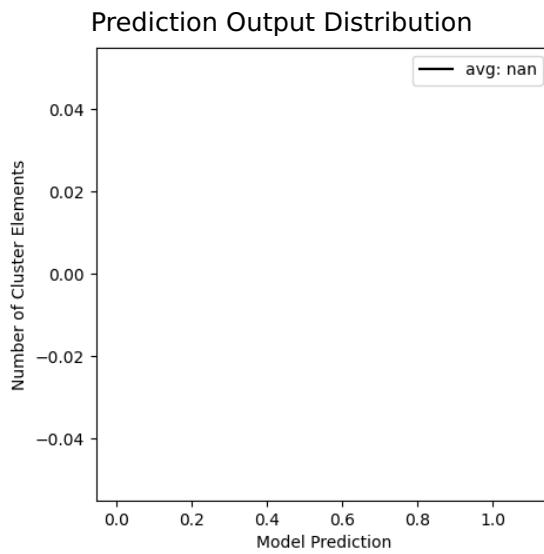
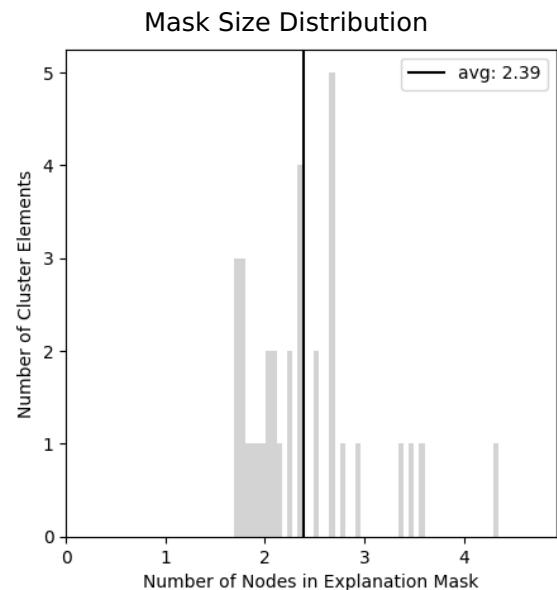
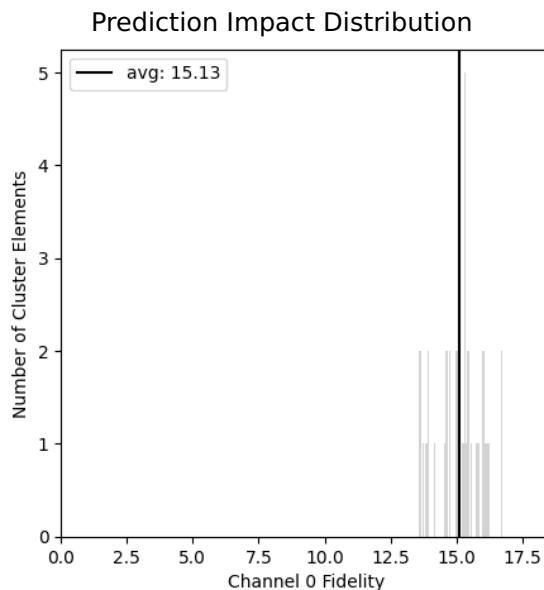
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	34
Channel Index	0.0 (0.0)

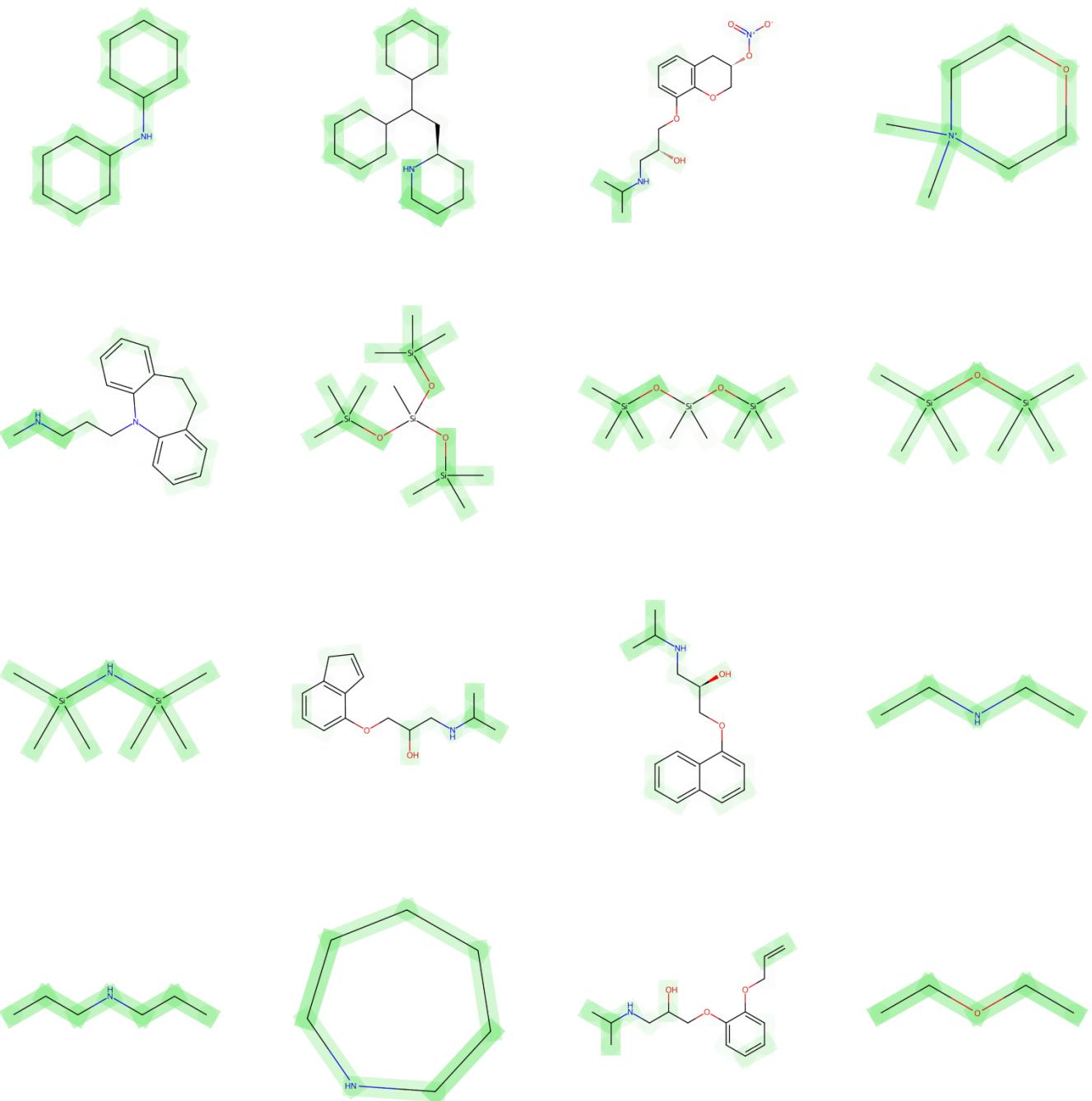
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The linear alkane structure represented by "C-C-C" in SMILES denotes a simple hydrocarbon chain with no apparent functional groups that typically participate in mutagenic activities, such as aromatic rings or reactive electrophilic sites. Alkanes are generally known to be less reactive due to their saturated carbon-carbon bonds, which makes them less likely to interact with DNA or other biomolecules in a way that causes mutagenesis. Alkanes do not easily form the necessary reactive intermediates that can attack genetic material and therefore, this structure is less likely to be mutagenic.

**Hypothesis:** Molecules containing the simple hydrocarbon chain substructure "C-C-C" are associated with a medium influence towards being non-mutagenic. This is likely due to the lack of functional groups in the structure that are capable of DNA interaction or damage, and because of the saturated nature of the hydrocarbon chain which results in lower reactivity with the genetic material of living organisms.

# Cluster #35 - non-mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 35, from importance channel 0 (*non-mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 7.3 ( $\pm 2.5$ ) on the prediction outcome.

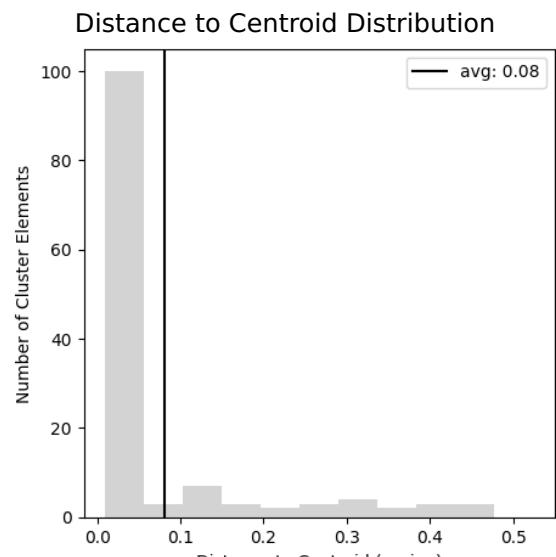
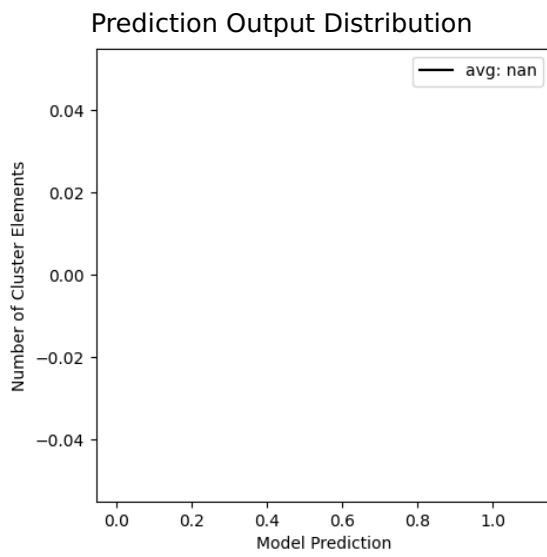
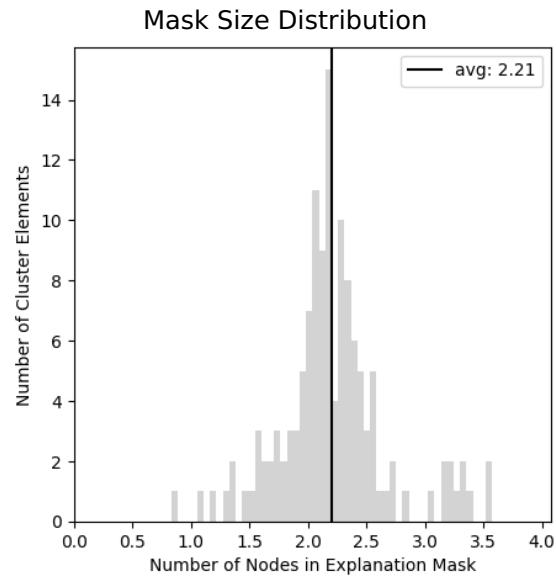
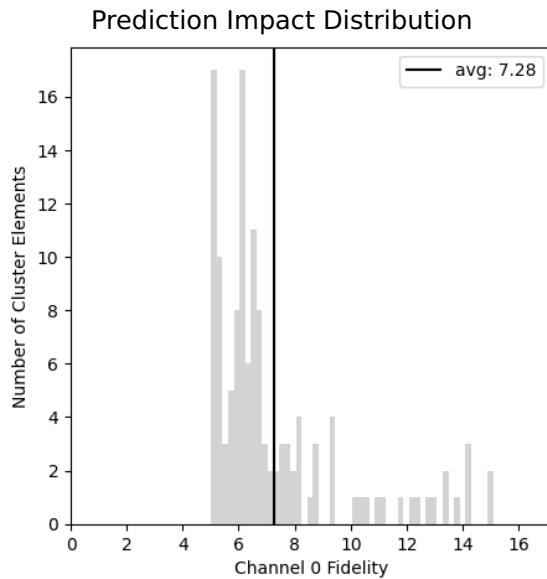
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	130
Channel Index	0.0 (0.0)

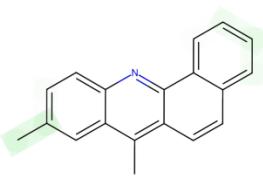
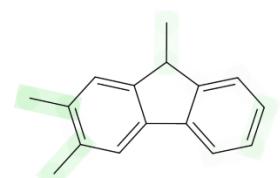
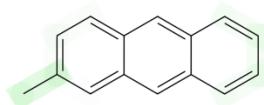
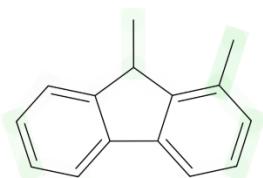
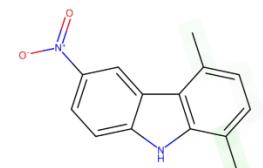
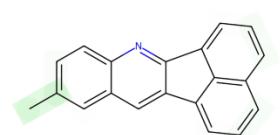
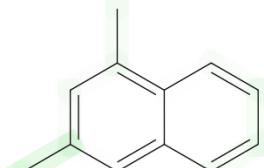
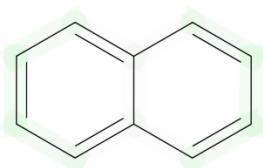
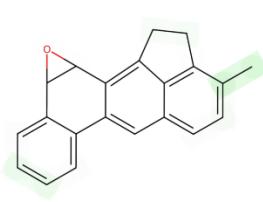
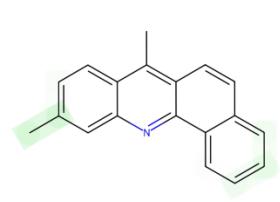
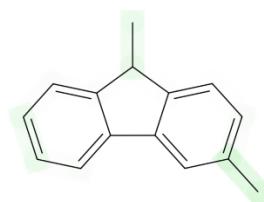
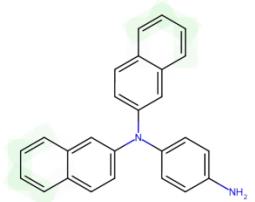
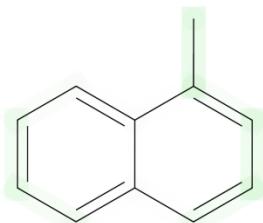
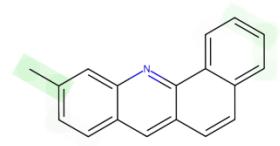
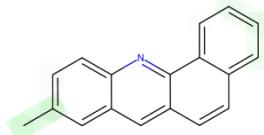
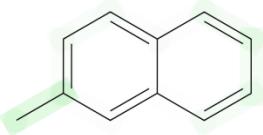
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



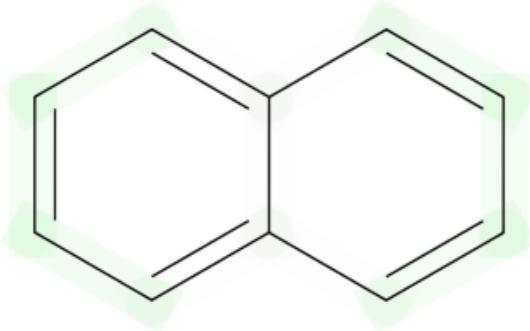
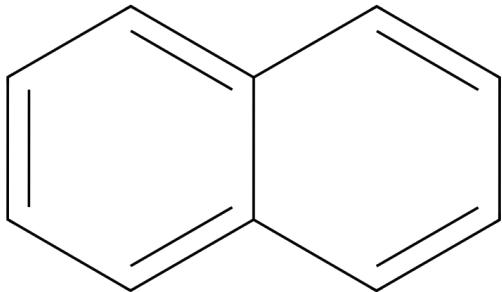
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation corresponds to the molecular structure of naphthalene, a polycyclic aromatic hydrocarbon (PAH) consisting of two fused benzene rings. PAHs are known to intercalate in DNA, leading to mutations; however, naphthalene has a relatively simple structure compared to other more complex and larger PAHs. The spatial configuration and electronic distribution allow it to stack between the DNA bases but may not be as reactive or fitting as other complex structures, thus having a smaller influence on causing mutations.

**Hypothesis:** Naphthalene's structure has a modest impact on its mutagenicity. It stacks between DNA bases without the heightened reactivity of larger PAHs, leading to fewer mutations. This hypothesis is corroborated by empirical evidence suggesting naphthalene's relatively lower mutagenic potential compared to more structurally complex PAHs.

# Cluster #36 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 36, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 8.3 ( $\pm 1.2$ ) on the prediction outcome.

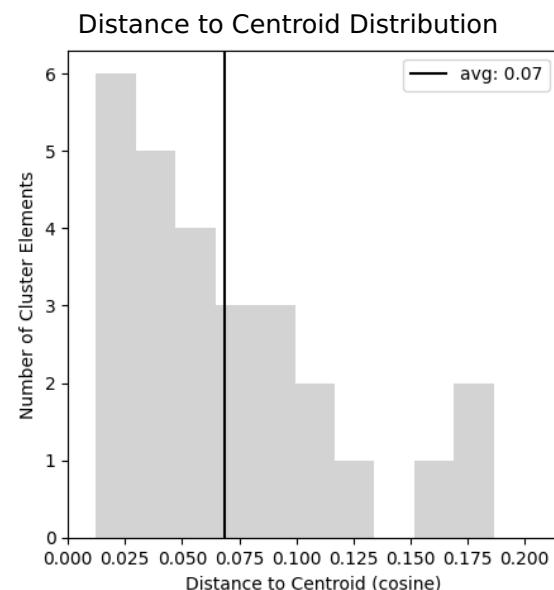
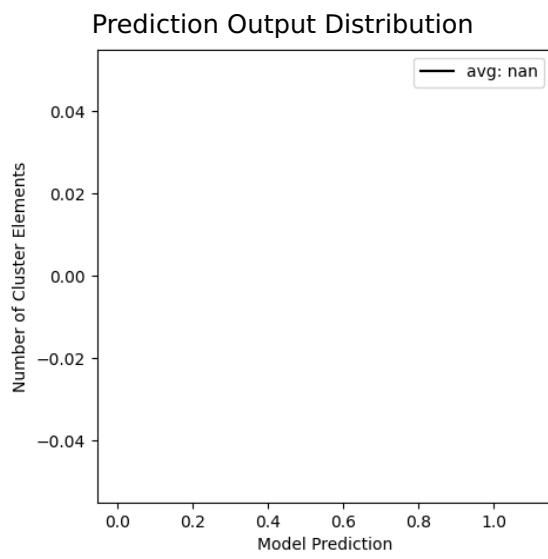
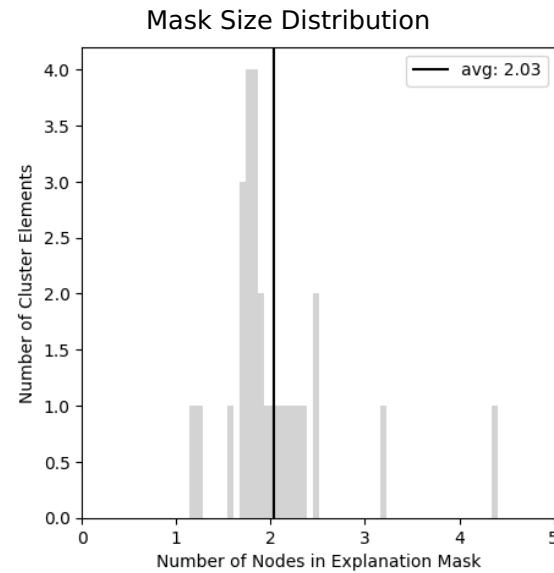
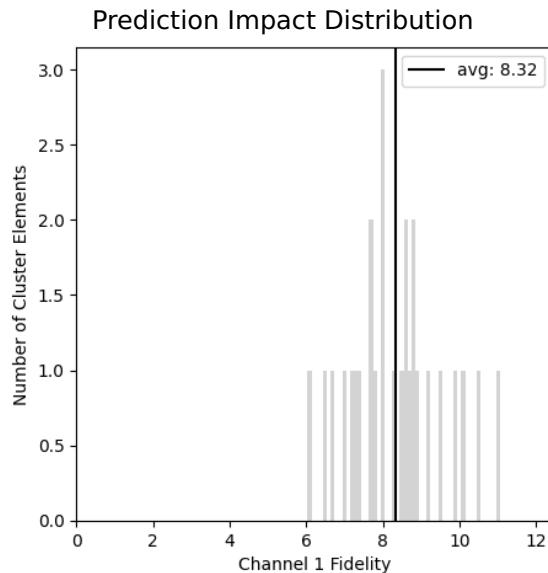
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	27
Channel Index	1.0 (0.0)

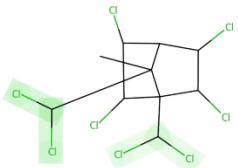
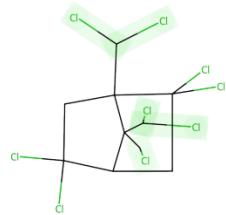
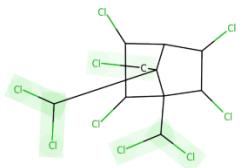
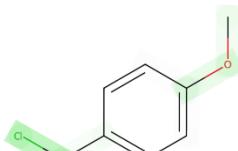
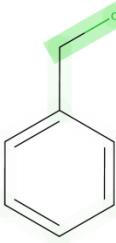
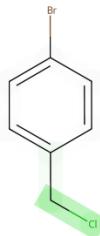
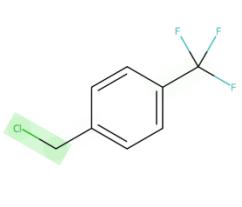
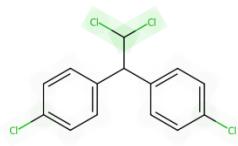
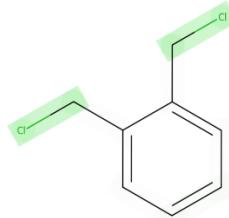
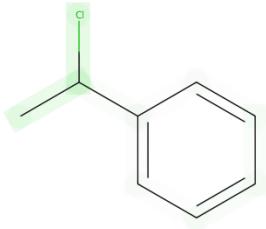
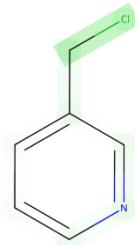
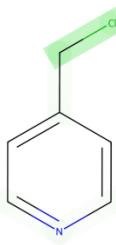
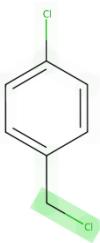
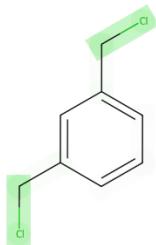
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure "F-C-F" represents a carbon atom single-bonded to two fluorine atoms. This is a simple structure that can be part of larger molecules. The mutagenic property is often associated with a molecule's ability to interact with DNA and cause changes in its structure. In this case, the presence of fluorine atoms may potentially play a role in such interactions due to their high electronegativity, which could disturb the electron distribution in adjacent molecules or DNA itself. However, the term "SMALL influence" suggests that this interaction is not strong, possibly due to the stabilizing effect of fluorine that may reduce the reactivity of this molecular fragment.

**Hypothesis:** The "F-C-F" structure has a small influence on mutagenicity, potentially due to the electronegative nature of fluorine that can perturb nearby electron distributions. However, the influence is limited, potentially because the fluorine atoms also confer stability to the structure, preventing it from easily interacting with or causing significant damage to DNA.

# Cluster #37 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 37, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 17.9 ( $\pm 2.7$ ) on the prediction outcome.

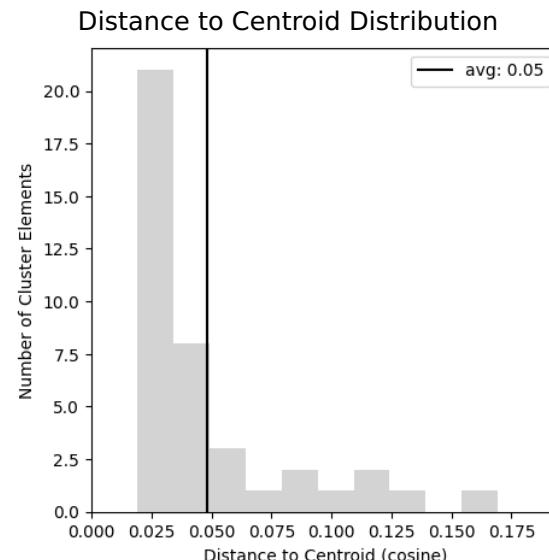
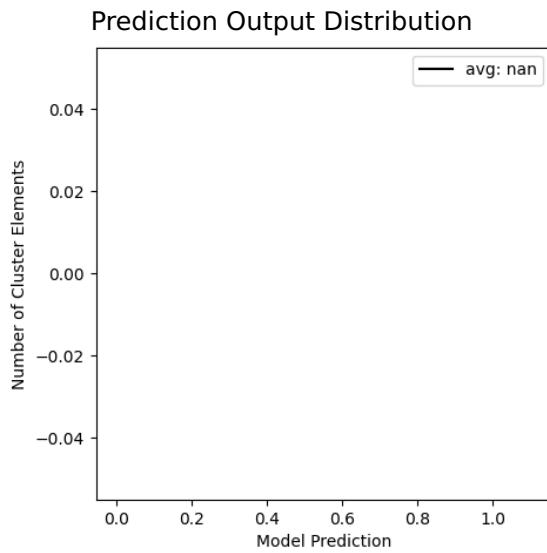
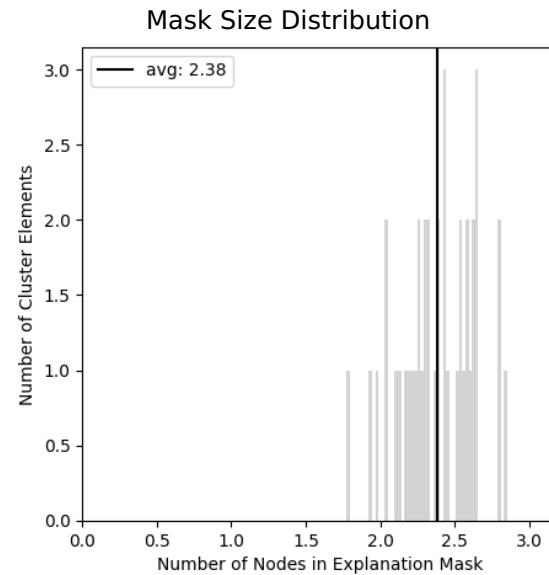
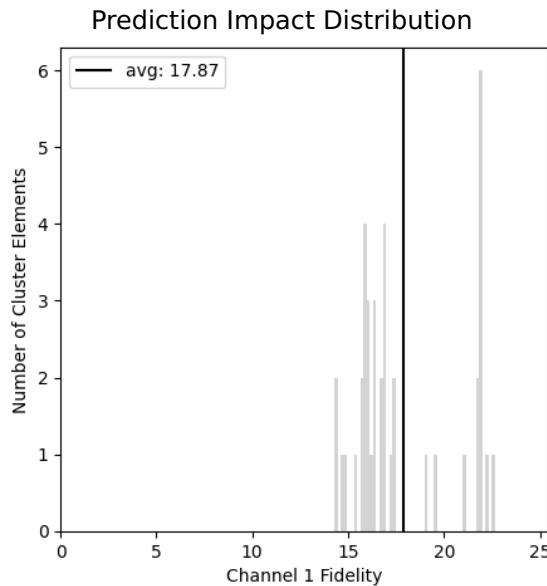
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	40
Channel Index	1.0 (0.0)

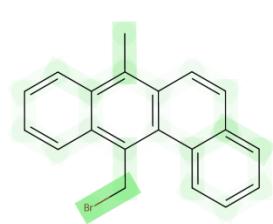
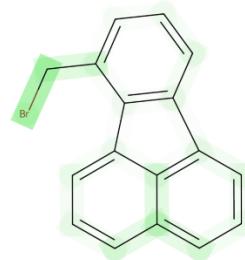
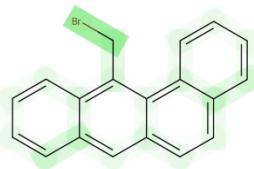
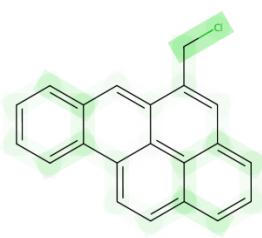
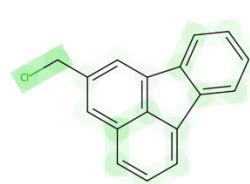
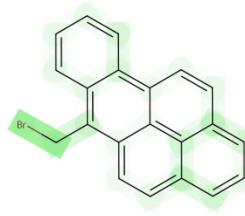
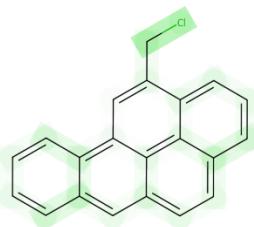
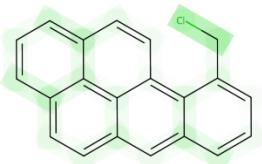
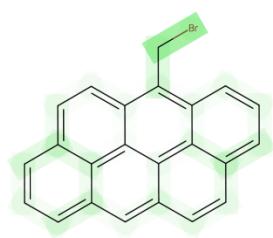
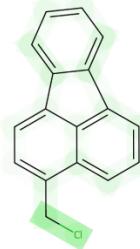
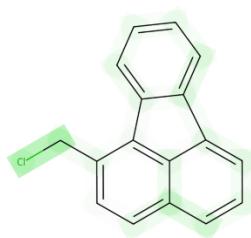
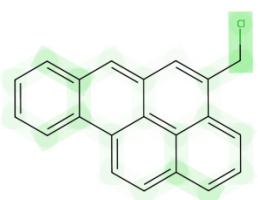
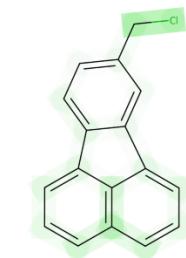
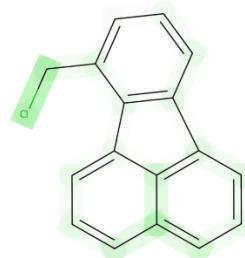
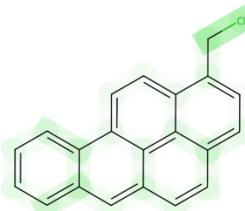
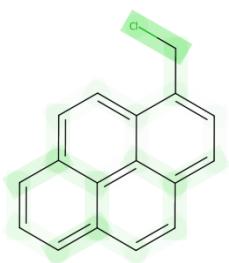
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



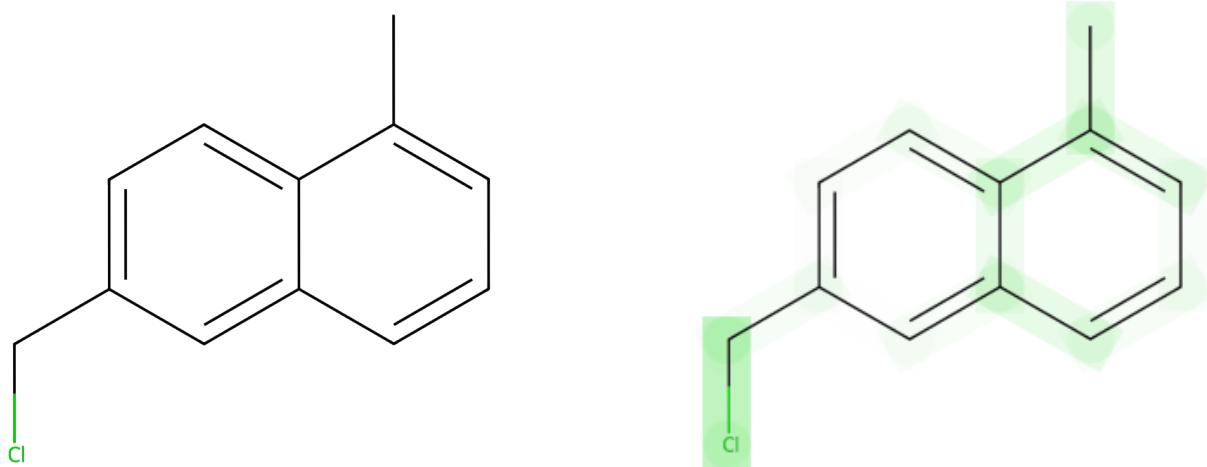
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the given SMILES notation indicates the presence of a chlorinated biphenyl group, which is a form of polychlorinated biphenyl (PCB). PCBs are known to be a group of chemicals that can potentially form DNA adducts or interfere with DNA replication, leading to mutations. The presence of the chlorine atom, specifically when bonded to an aromatic system as in a biphenyl group, can stabilize the formation of a positively charged intermediate during metabolic activation. This intermediate can then interact with DNA, forming covalent bonds that result in mutagenicity.

**Hypothesis:** The molecular structure consisting of a chlorinated biphenyl group has a medium influence towards mutagenicity. This is likely due to the ability of the chlorine substituent to stabilize charged intermediates during metabolism, which then leads to DNA interaction and potential mutations. The aromaticity of the biphenyl system also plays a role in the persistence of the compound within biological systems, potentially increasing exposure and mutagenic risk.

# Cluster #38 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 38, from importance channel 1 (*mutagenic*), represents a motif consisting of 4.7 ( $\pm 0.8$ ) nodes. The concept is generally associated with an impact of 18.7 ( $\pm 2.6$ ) on the prediction outcome.

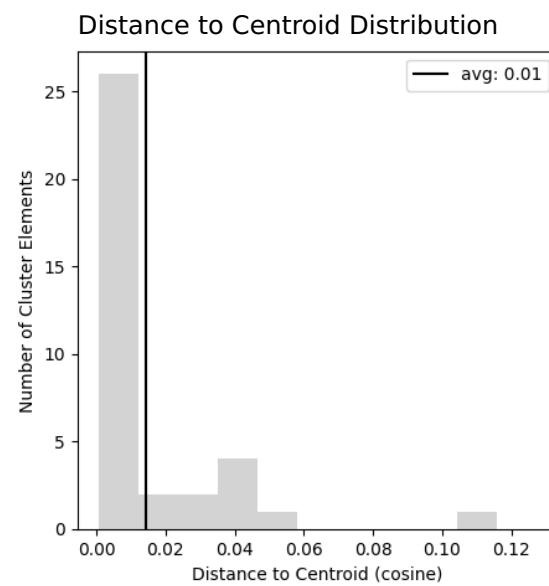
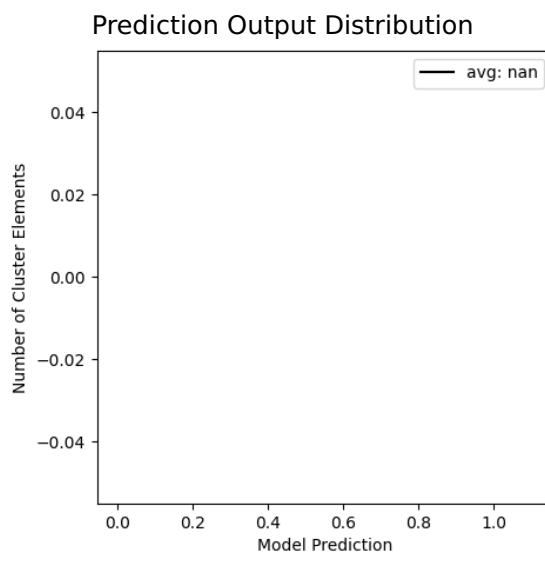
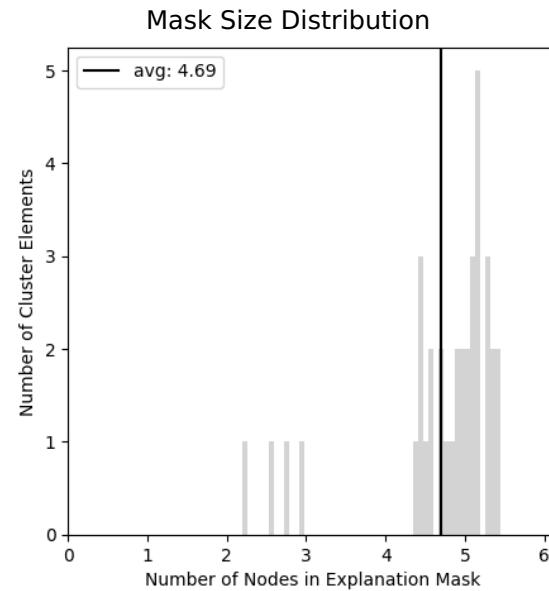
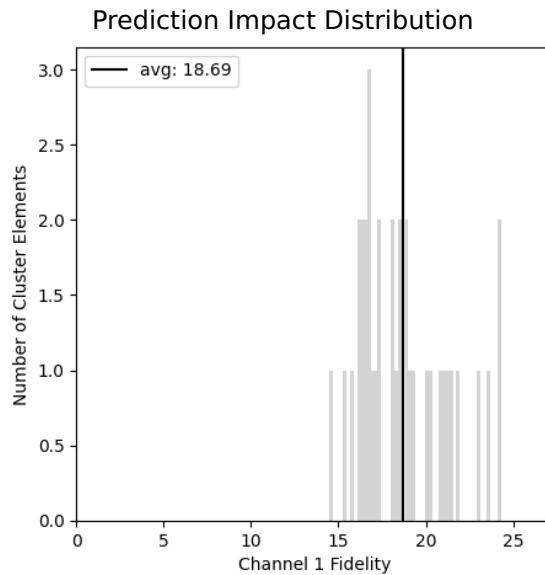
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	36
Channel Index	1.0 (0.0)

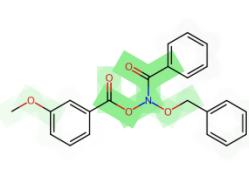
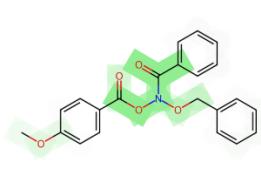
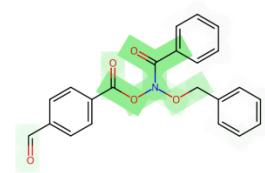
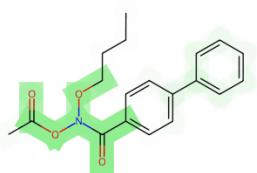
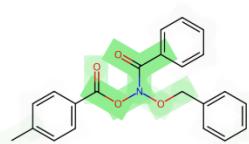
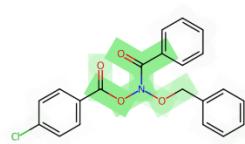
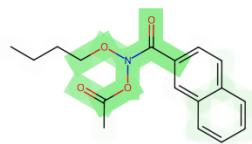
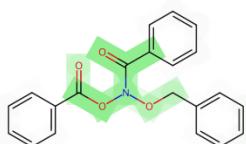
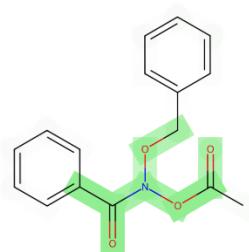
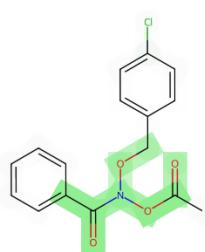
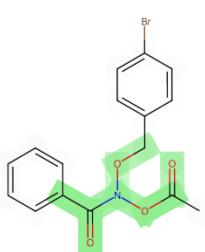
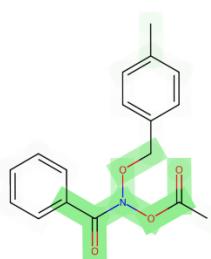
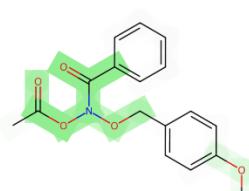
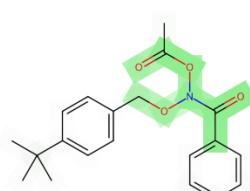
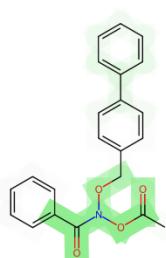
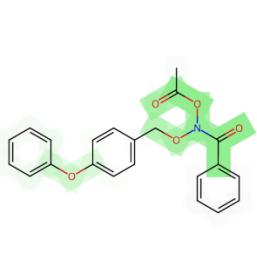
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



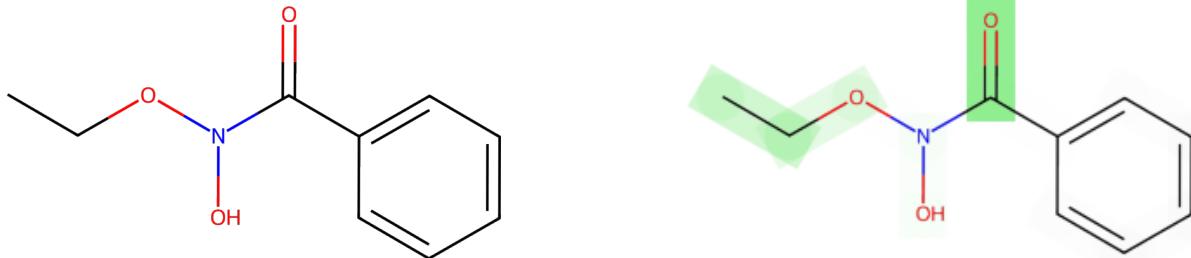
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a molecular fragment containing a benzene ring attached to an ester group and a nitrosoamine. The presence of the highly reactive nitroso group (N-O) adjacent to the amine (N-H or N-R groups) may facilitate the formation of nitrenium ions, which are highly electrophilic and can attack DNA, leading to mutations. The planar aromatic benzene ring can intercalate between DNA bases, thus bringing the nitrosoamine in proximity to genetic material and increasing the likelihood of a reaction. The ester functionality might influence the molecule's bioavailability and stability, thereby modulating the overall mutagenic potential.

**Hypothesis:** Molecules containing the substructure "C-C-O-N(-O)-C(=O)-c1:c:c:c:c:1" have a medium influence on mutagenicity. This may be due to the reactivity of the nitrosoamine group towards DNA and the intercalative properties of the benzene ring. Modulating factors such as stability and bioavailability introduced by the ester group can influence the degree of mutagenicity.

# Cluster #39 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 39, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 16.1 ( $\pm 1.8$ ) on the prediction outcome.

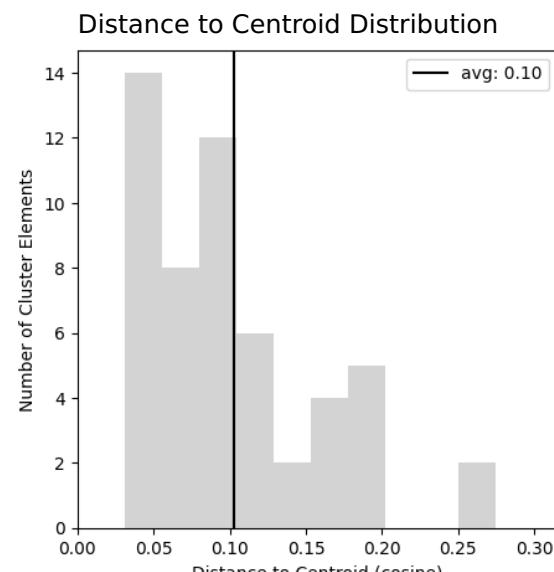
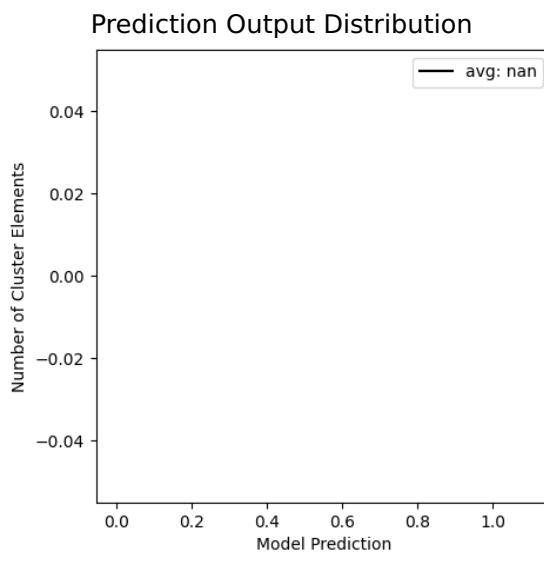
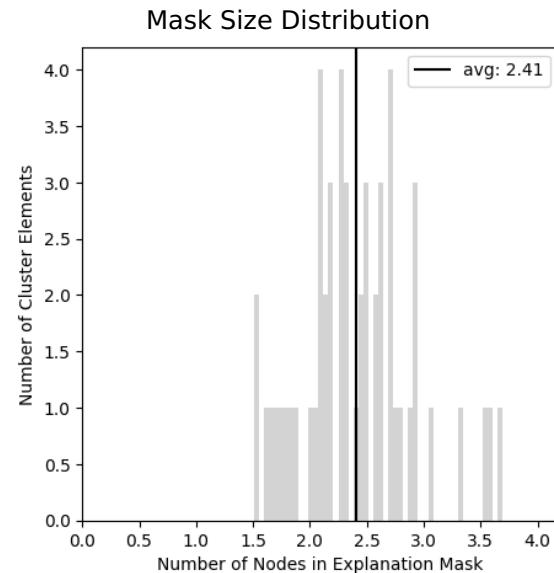
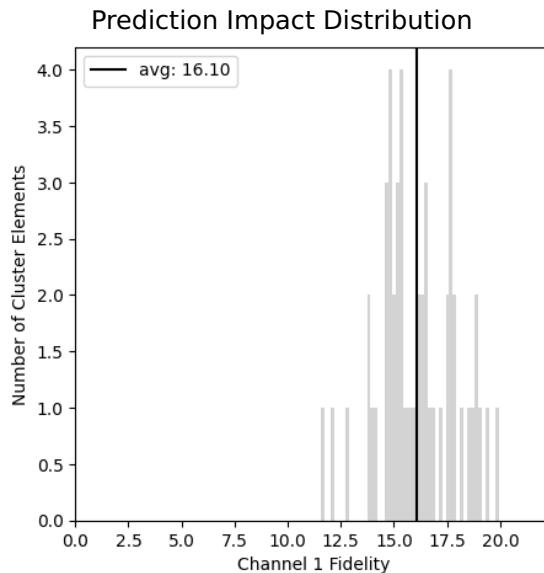
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	53
Channel Index	1.0 (0.0)

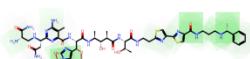
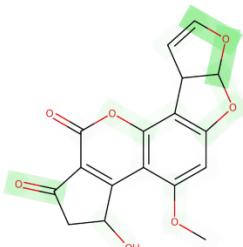
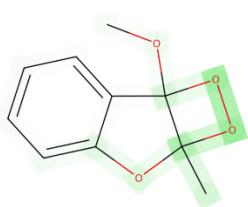
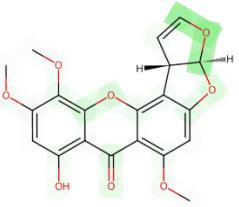
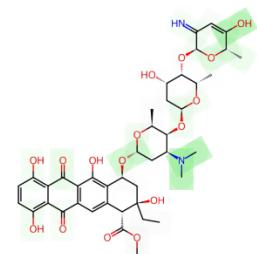
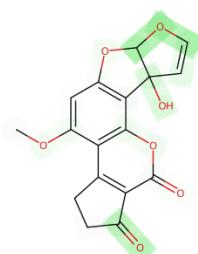
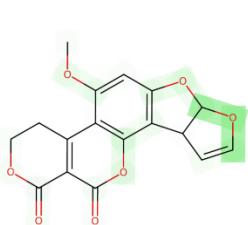
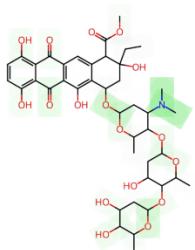
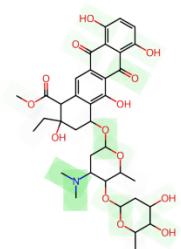
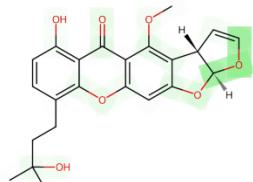
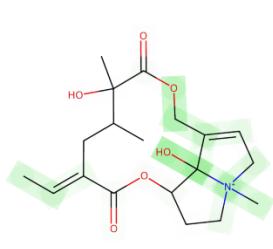
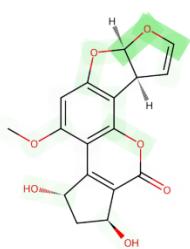
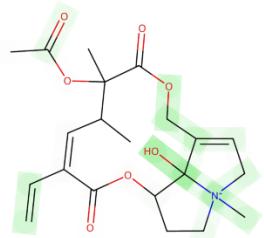
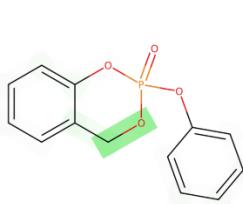
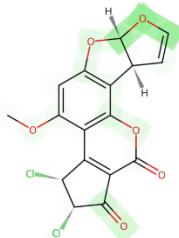
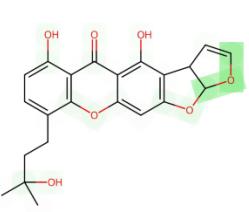
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



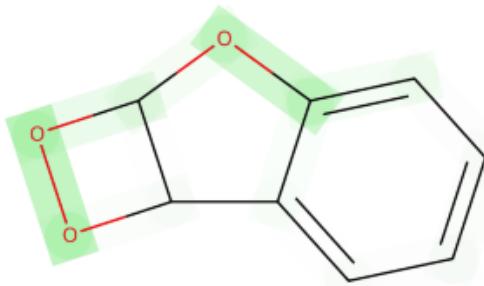
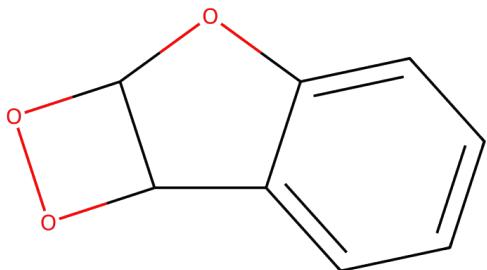
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure provided suggests the presence of a fused ring system where a benzene ring (an aromatic ring with alternating double bonds) is connected to a second ring that includes both oxygen and carbon atoms. The presence of oxygen atoms in cyclic structures (heterocyclic compounds) can significantly impact the electron distribution and reactivity. In this case, the -O-O- linkage (peroxide linkage) within the ring system could provide a point of vulnerability for the generation of free radicals under biological conditions. Free radicals are highly reactive and can readily interact with DNA, leading to possible mutagenic events. Additionally, the aromatic system may facilitate planar stacking interactions with the base pairs of the DNA helix, further increasing the likelihood of an interaction leading to mutagenic outcomes.

**Hypothesis:** The molecular substructure with fused benzene and heterocyclic rings containing peroxide linkages is hypothesized to have a medium influence on mutagenicity. The peroxide linkage could be a source of free radicals that damage DNA, while the planar aromatic system may interact with DNA structure. These factors may contribute to the observed mutagenic properties of molecules containing this substructure.

# Cluster #40 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 40, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.6 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 18.0 ( $\pm 1.7$ ) on the prediction outcome.

## Properties

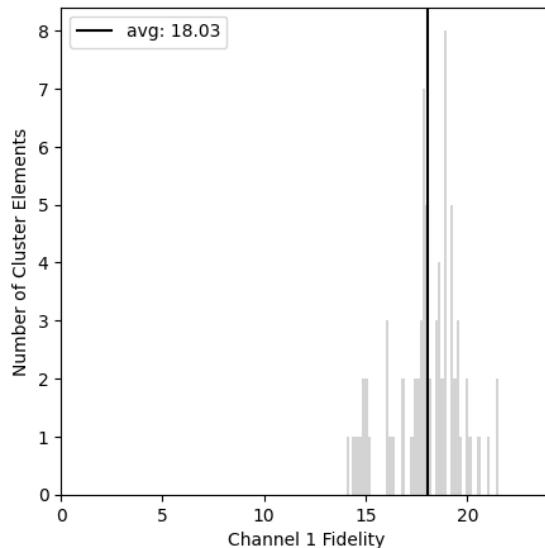
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	73
Channel Index	1.0 (0.0)

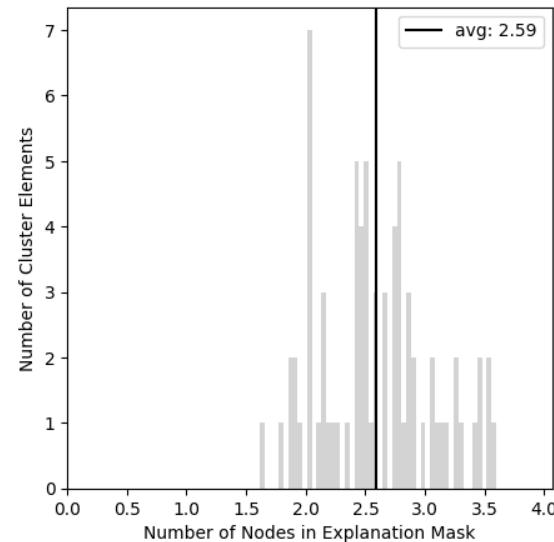
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

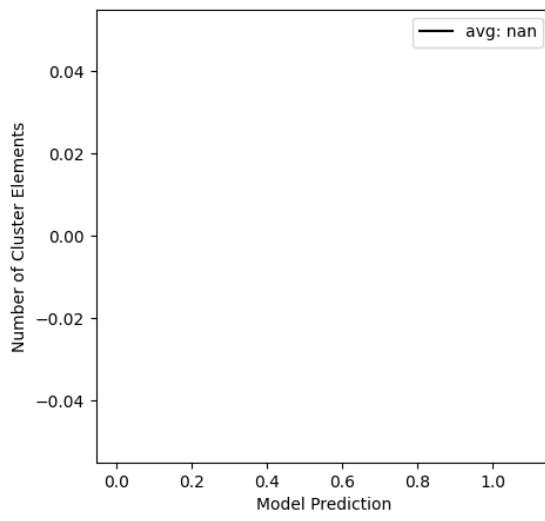
Prediction Impact Distribution



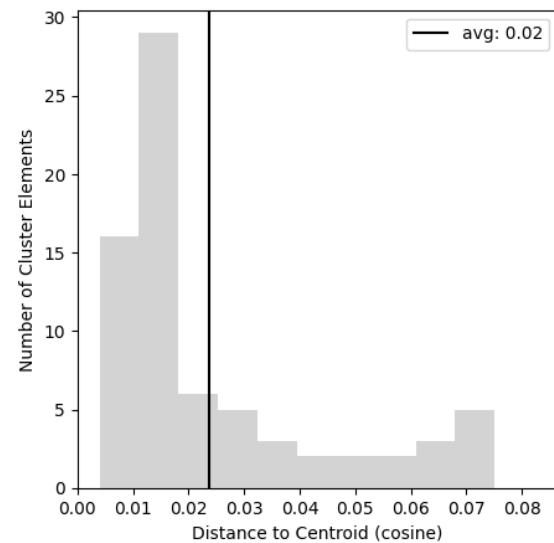
Mask Size Distribution



Prediction Output Distribution

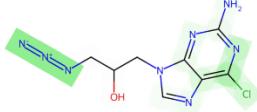
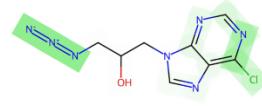
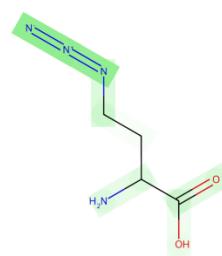
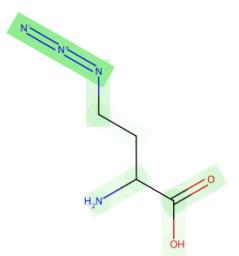
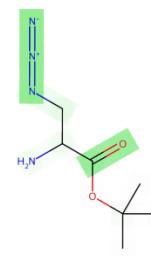
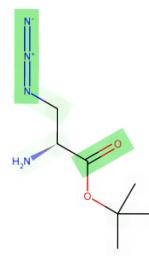
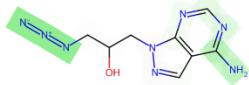
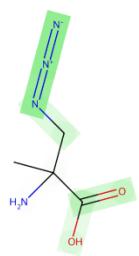
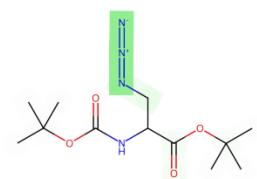
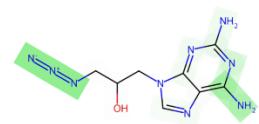
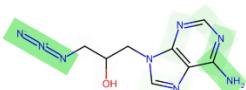
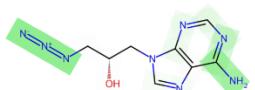
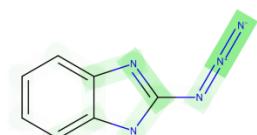
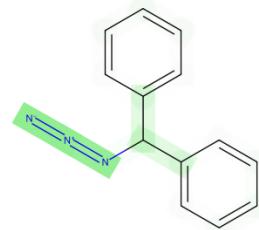
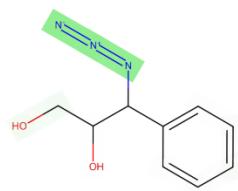
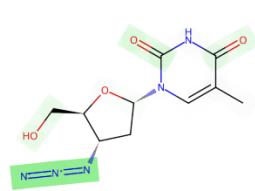


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES "[N-]=[N+]=N" represents an azide group, which is a highly reactive species known to interact with biomolecules, leading to potential mutations. The azide group's reactivity comes from its ability to release nitrogen gas ( $N_2$ ), a linear triatomic molecule. This process can result in the formation of reactive intermediates or can cause the alteration of DNA through the generation of radical species or direct interaction with the DNA, thus disrupting the genetic material structure.

**Hypothesis:** The presence of an azide group in a molecule is associated with a medium influence on mutagenicity. The high reactivity of the azide group may facilitate interactions that could damage DNA, while the release of nitrogen gas represents a potential mechanism for mutagenesis. The "medium" classification might suggest that while azides are reactive, their actual mutagenic potential could be moderated by other factors such as molecular context or stability of the azide group within the larger molecule.

# Cluster #41 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 41, from importance channel 1 (*mutagenic*), represents a motif consisting of 3.2 ( $\pm 0.7$ ) nodes. The concept is generally associated with an impact of 18.2 ( $\pm 1.6$ ) on the prediction outcome.

## Properties

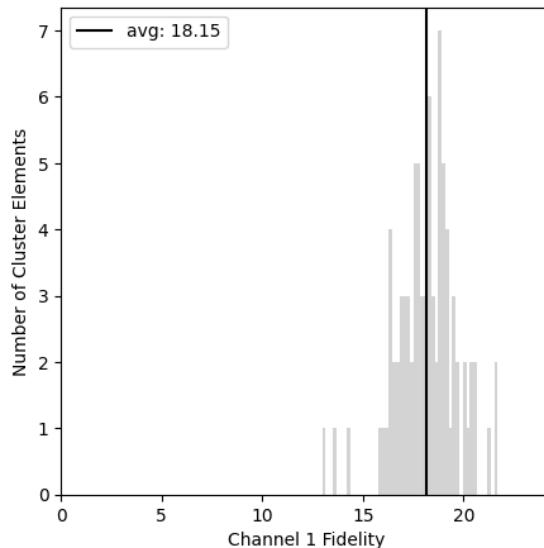
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	84
Channel Index	1.0 (0.0)

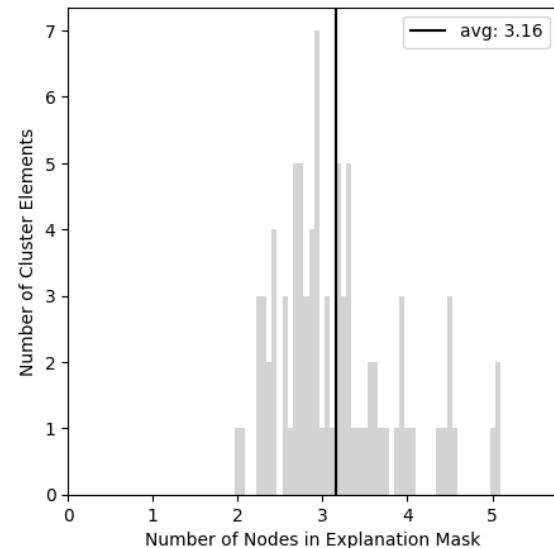
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

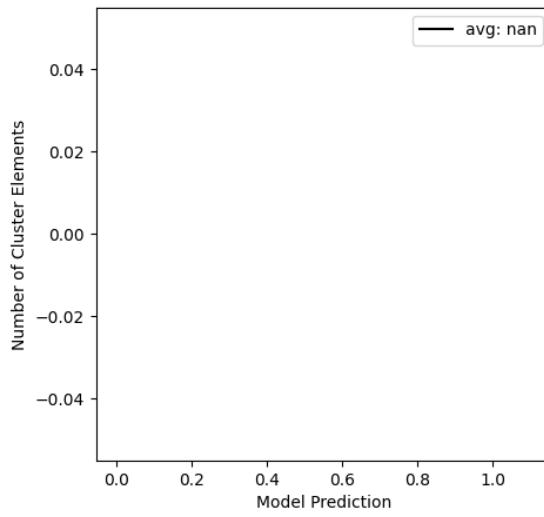
Prediction Impact Distribution



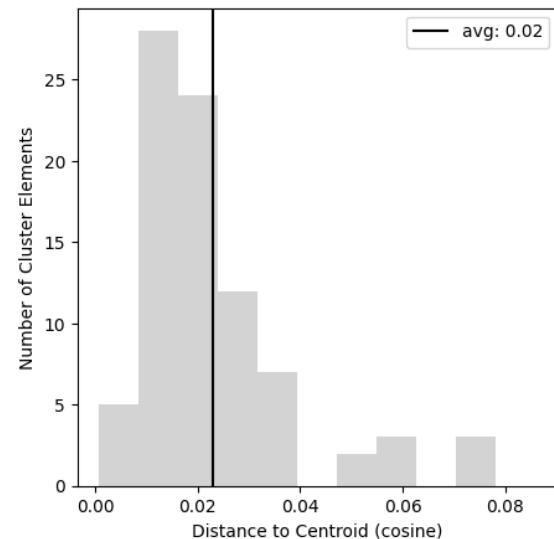
Mask Size Distribution



Prediction Output Distribution

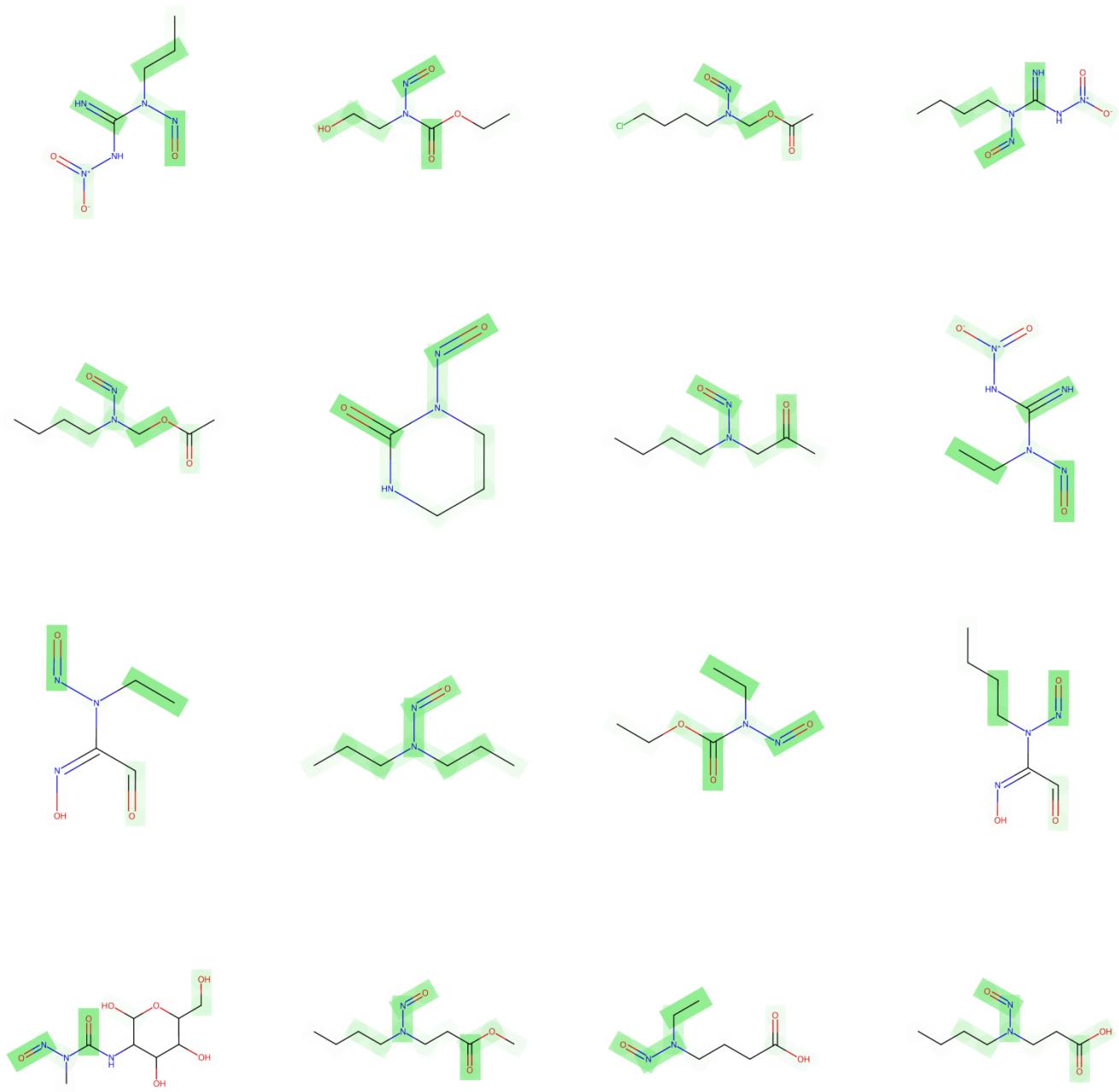


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structural subunit "C-C-N(-C-C)-N=O" contains two functional moieties that are of particular interest concerning mutagenicity: an amine group (-N(-C-C)-) and a nitroso group (-N=O). Amines are known to form diazonium ions which can interact with nucleic acids, leading to their potential mutagenic activity. A nitroso group can also be involved in the formation of reactive intermediates that can attack DNA, causing mutations. The combination of these two groups in one molecule could contribute to an overall medium mutagenicity because while each group has mutagenic potential, the placement within the molecule and other steric or electronic factors might limit their reactivity or interaction with genetic material to some extent.

**Hypothesis:** Molecules containing the "C-C-N(-C-C)-N=O" substructure are predicted to exhibit a medium level of mutagenic activity due to the reactive potential of the amine and nitroso groups. This intermediate mutagenicity could be attributed to the balance between the inherent reactivity of these functional groups and the mitigating influence of the rest of the molecule's structure on their ability to interact with DNA.

# Cluster #42 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 42, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 16.8 ( $\pm 1.6$ ) on the prediction outcome.

## Properties

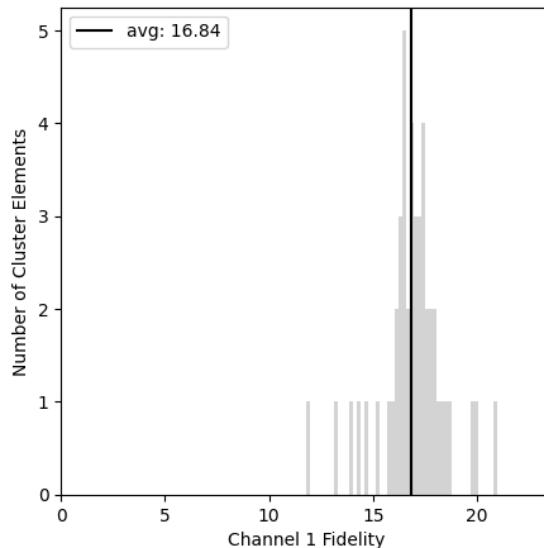
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	47
Channel Index	1.0 (0.0)

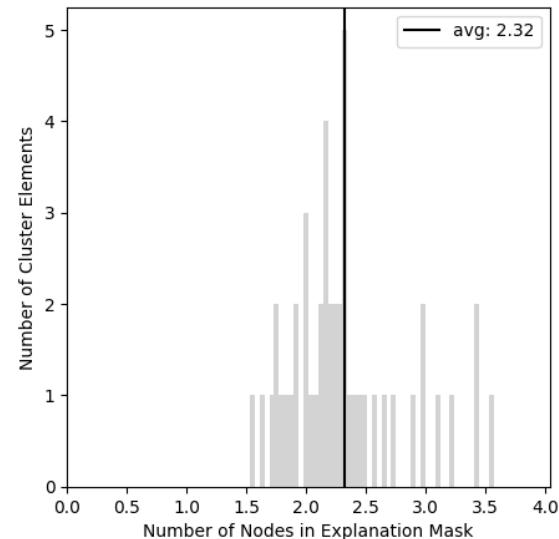
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

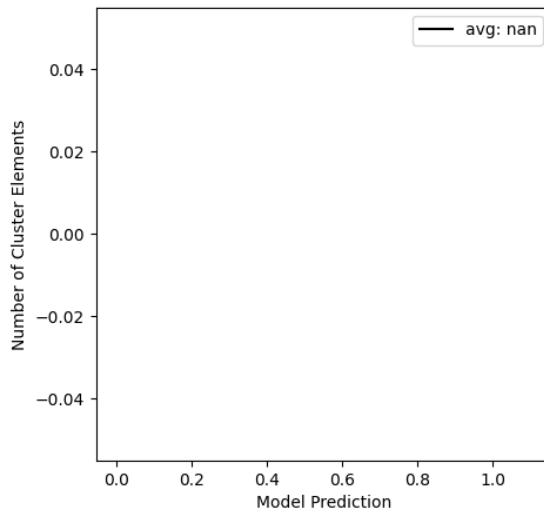
Prediction Impact Distribution



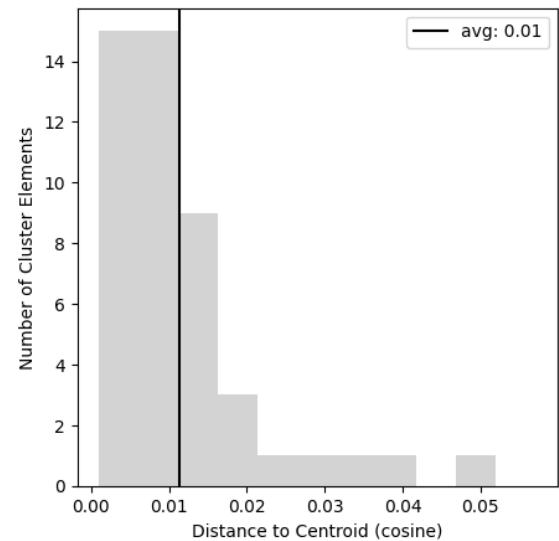
Mask Size Distribution



Prediction Output Distribution

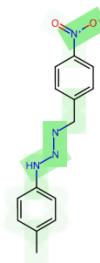
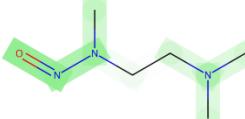
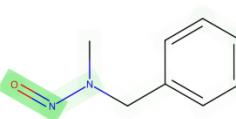
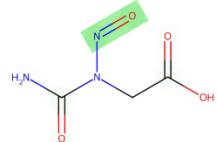
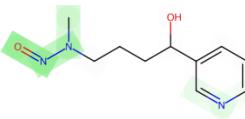
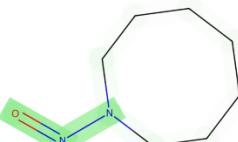
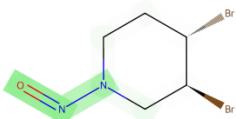
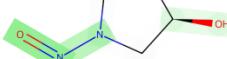
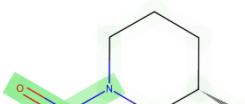
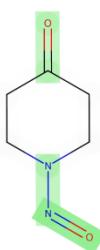
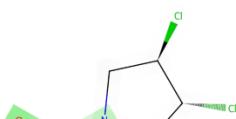
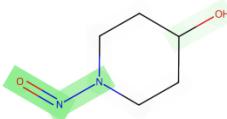
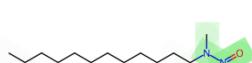
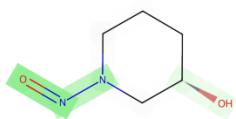


Distance to Centroid Distribution



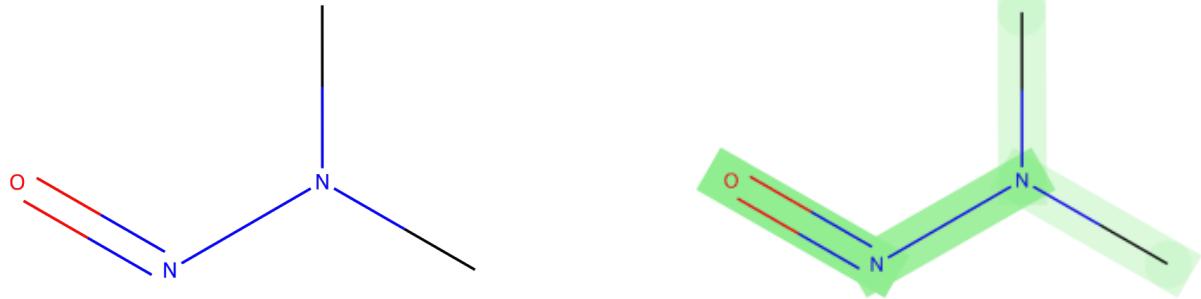
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure "C-N(-C)-N=O", known as a nitroso group attached to a secondary amine, suggests the presence of a functional group that is capable of undergoing redox cycling. The intermediate nitroso compounds and their reduced hydroxylamine derivatives can covalently modify nucleic acids, leading to mutations. The mutagenic potential of such structures is influenced by the ability to form reactive oxygen species (ROS) and DNA adducts through metabolic activation.

**Hypothesis:** Molecules containing the substructure "C-N(-C)-N=O" exhibit a medium level of mutagenicity due to their potential to undergo redox reactions, leading to the formation of DNA-damaging agents. This structure is hypothesized to be less mutagenic than nitroaromatics due to steric hindrance caused by the secondary amine, which may reduce its activation and subsequent interaction with genetic material.

# Cluster #43 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 43, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.8 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 18.6 ( $\pm 1.4$ ) on the prediction outcome.

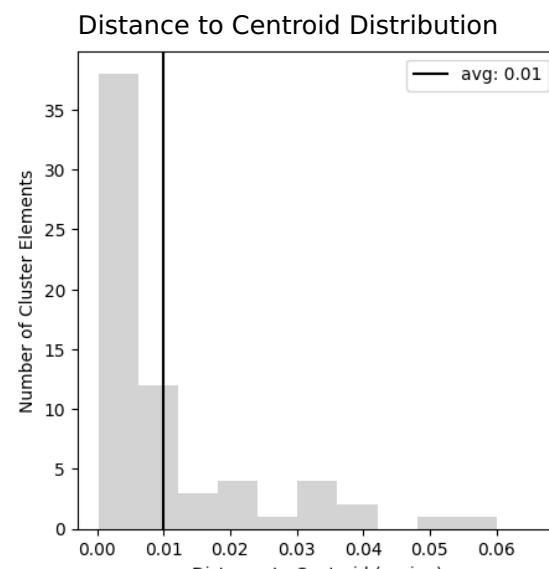
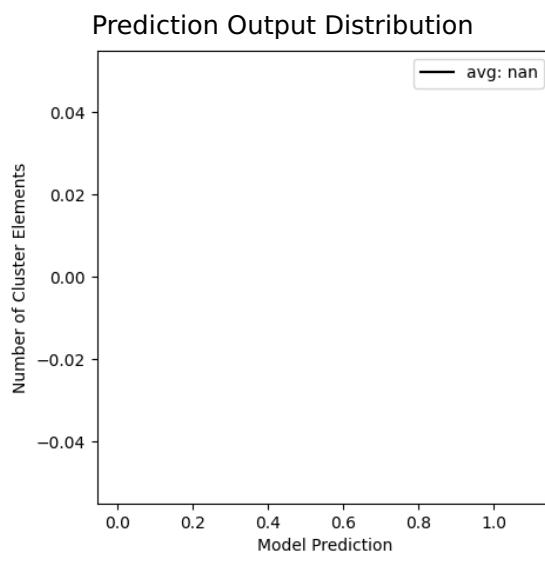
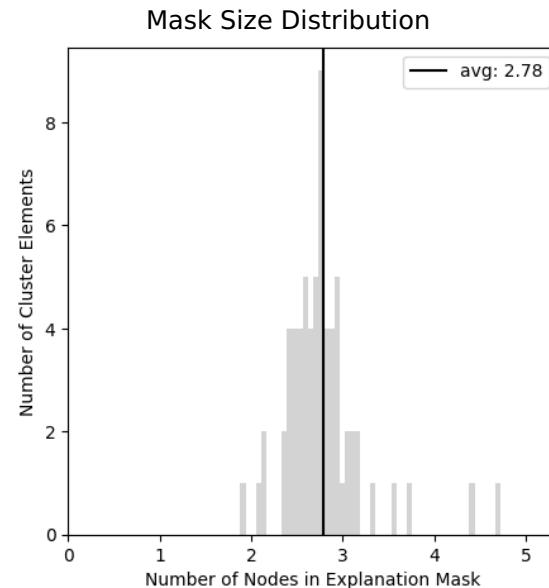
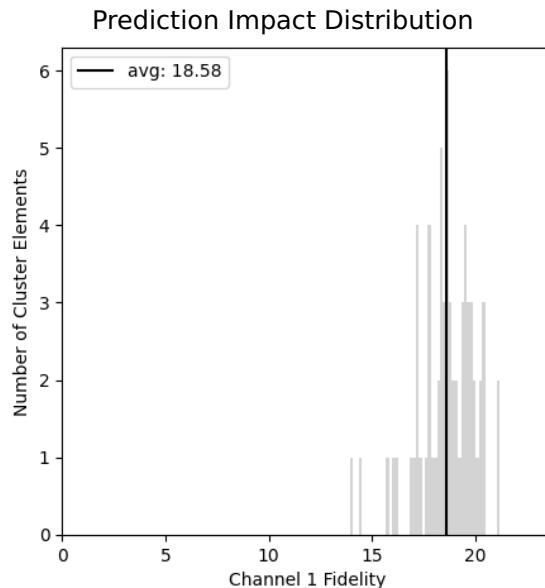
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	66
Channel Index	1.0 (0.0)

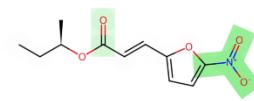
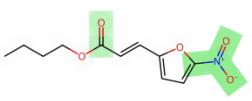
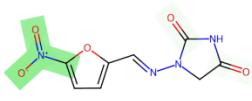
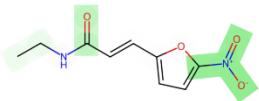
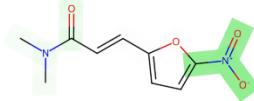
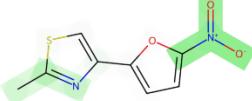
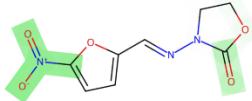
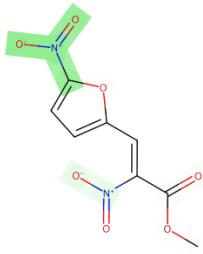
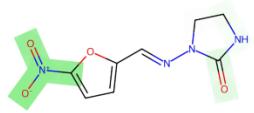
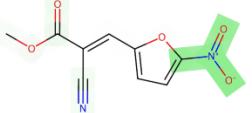
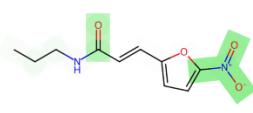
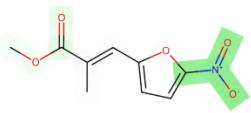
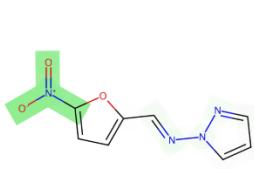
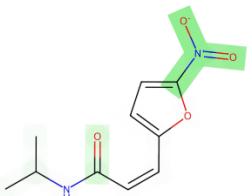
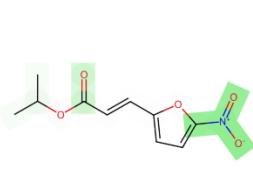
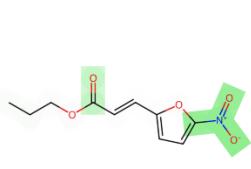
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



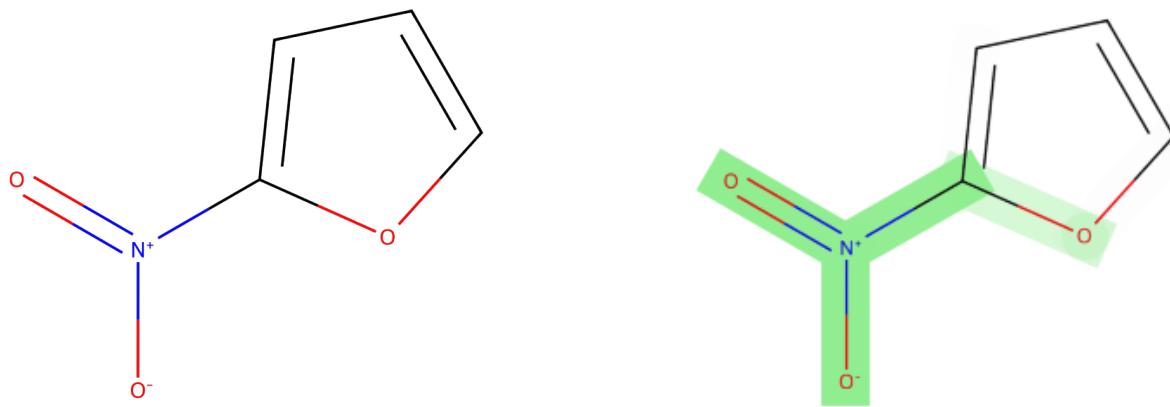
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "O=[N+](-[O-])-c1:c:c:o:1" corresponds to a nitro-group substituted aromatic compound, specifically a nitrophenol. The nitro group ( $\text{NO}_2$ ) is an electron-withdrawing group, which can make the aromatic system more reactive. The phenolic OH group, especially when ionized, can form hydrogen bonds or interact with other biomolecules. These interactions could lead to disruptions in the normal processes of DNA, such as replication or repair, possibly causing mutations.

**Hypothesis:** Compounds with a nitro group attached to an aromatic ring, especially in conjunction with a phenolic OH, have a medium mutagenic potential. The electron-withdrawing nature of the nitro group and the potential for the phenolic OH to interact with DNA components may lead to an increased likelihood of genetic material alteration.

# Cluster #44 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 44, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 17.2 ( $\pm 1.7$ ) on the prediction outcome.

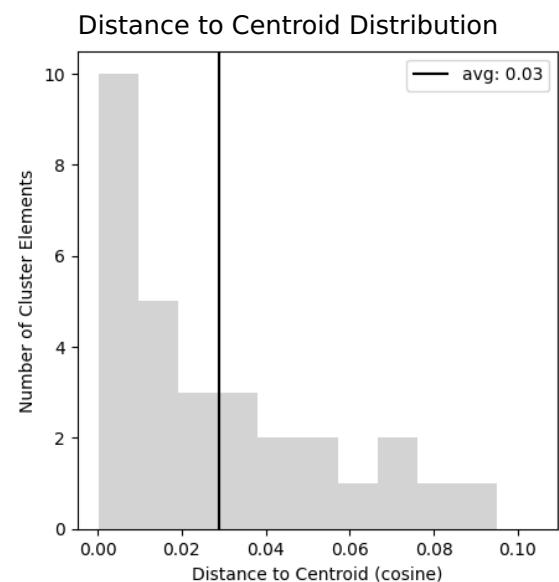
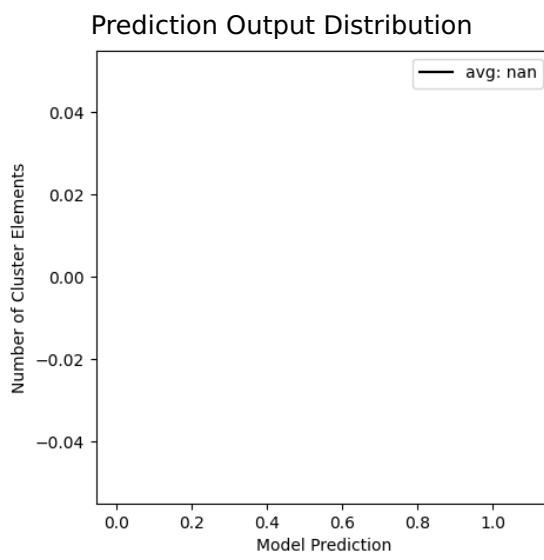
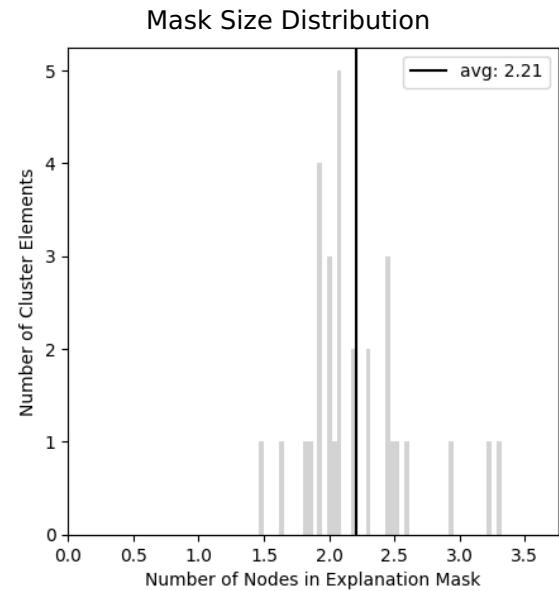
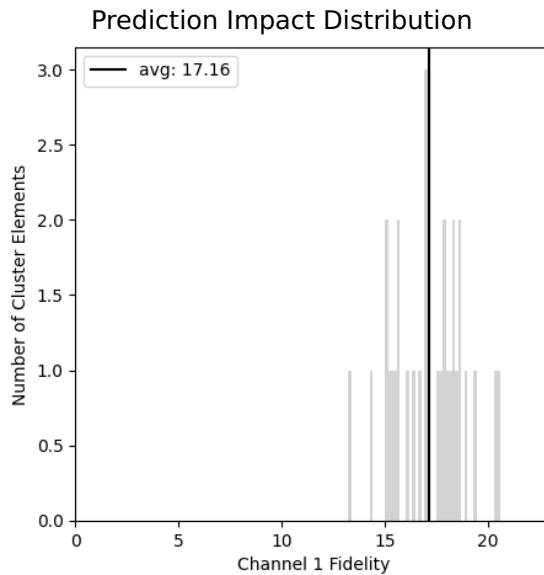
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	30
Channel Index	1.0 (0.0)

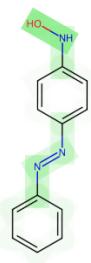
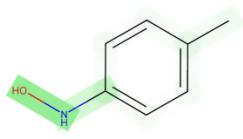
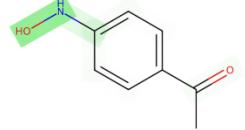
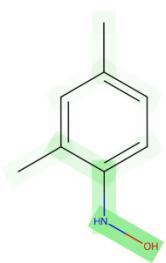
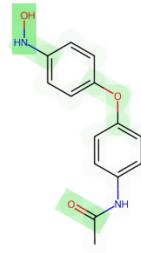
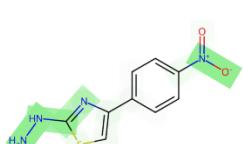
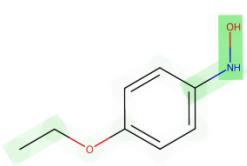
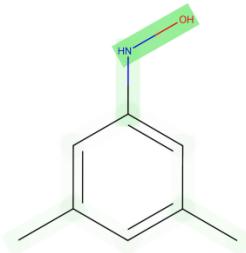
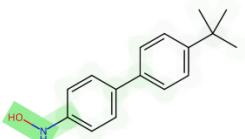
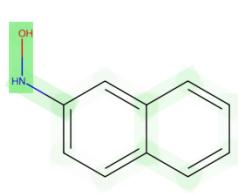
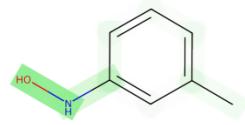
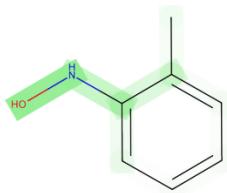
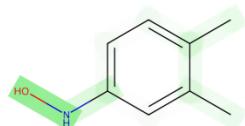
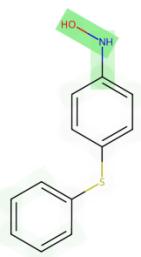
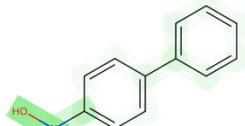
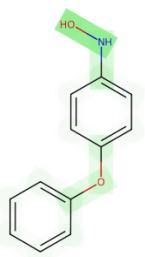
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The fragment represented by the SMILES string "[N+]-[O-]" denotes a nitro group, a functional group commonly found in various organic compounds, including some mutagens. The nitro group is known to be bioactivated within cells to form reactive nitroso derivatives and other reactive species. These reactive species can cause oxidative damage to DNA, leading to mutations. The positive charge on the nitrogen and the negative charge on the oxygen in the nitro group indicates a polarized bond, which can facilitate the bioactivation process.

**Hypothesis:** Molecules containing the nitro group substructure "[N+]-[O-]" are hypothesized to have a medium influence on mutagenicity due to the capability of the nitro group to be bioactivated into reactive species that damage DNA. The polarization of the nitro group is a key factor in this bioactivation process, anticipating a structure-property relationship that contributes to the mutagenicity of the molecule.

# Cluster #45 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 45, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 15.4 ( $\pm 2.0$ ) on the prediction outcome.

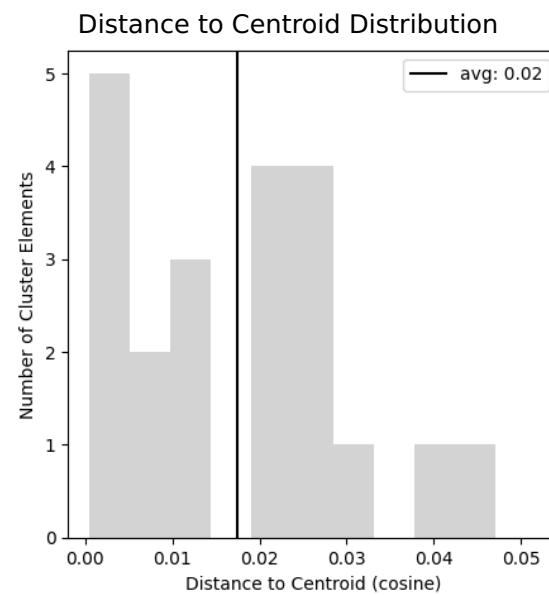
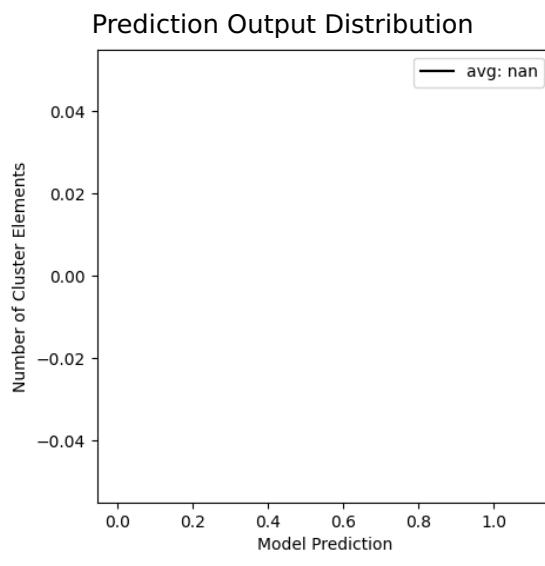
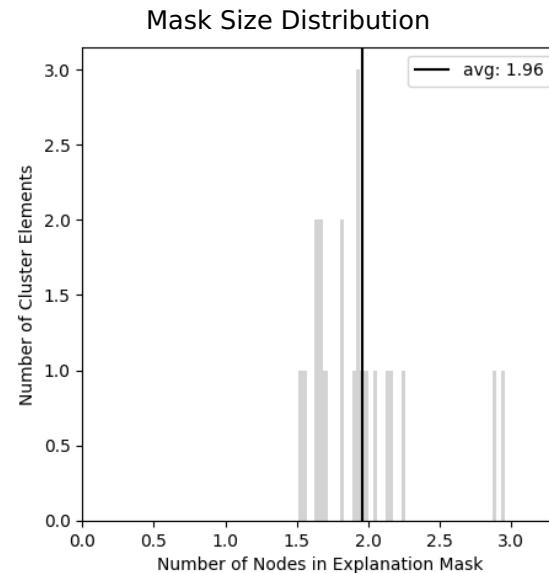
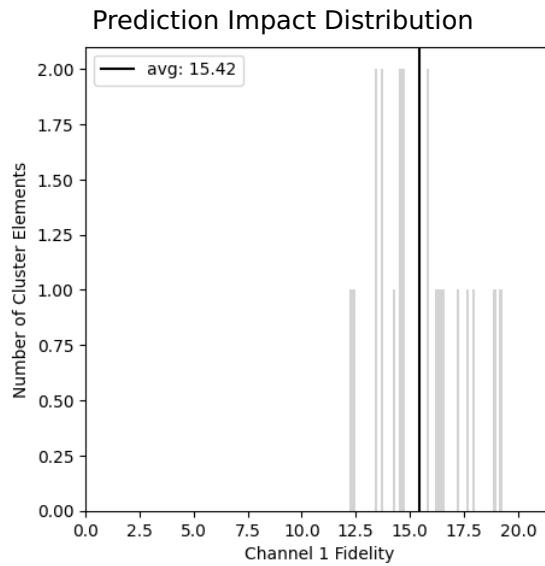
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

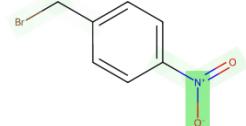
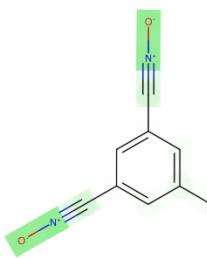
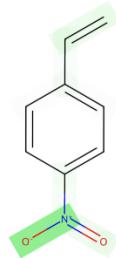
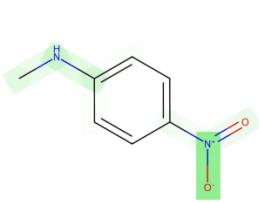
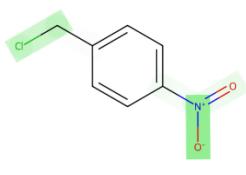
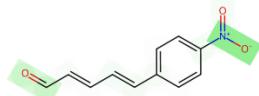
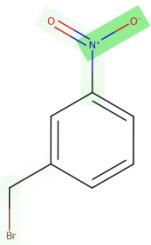
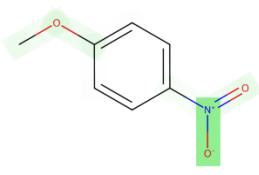
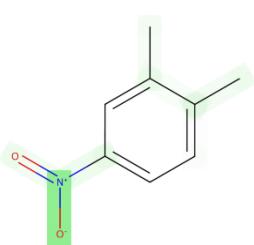
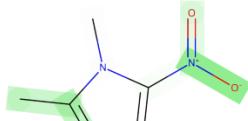
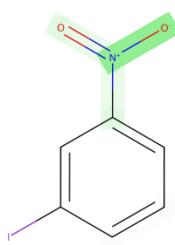
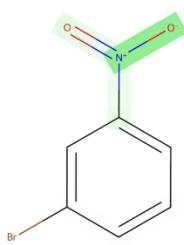
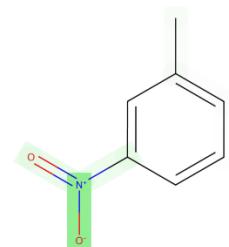
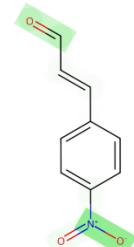
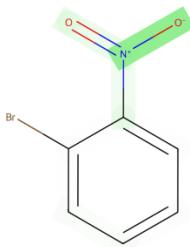
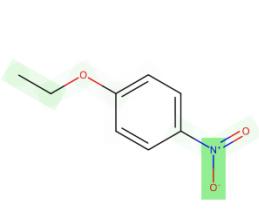
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a nitro group attached to a benzene ring, known as a nitroarene. Nitroarenes are well-documented mutagens. The principal reason is the nitro group ( $O=[N^+]$ ), which can be metabolically reduced *in vivo* to form reactive nitroso intermediates. These intermediates can further undergo enzymatic processes to produce highly reactive species that can cause DNA adducts, leading to mutations. Moreover, the delocalization of electrons in the benzene ring could stabilize the formation of these reactive species, making the molecule more likely to interact with genetic material.

**Hypothesis:** The presence of a nitro group attached to a benzene ring has a medium influence on mutagenicity due to the formation of reactive nitroso intermediates that can interact with DNA. The aromatic system stabilizes these intermediates, facilitating the mutagenic process.

# Cluster #46 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 46, from importance channel 1 (*mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 15.7 ( $\pm 1.7$ ) on the prediction outcome.

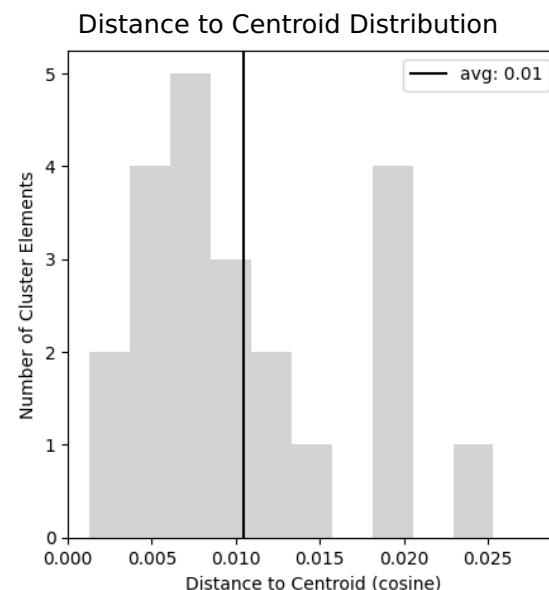
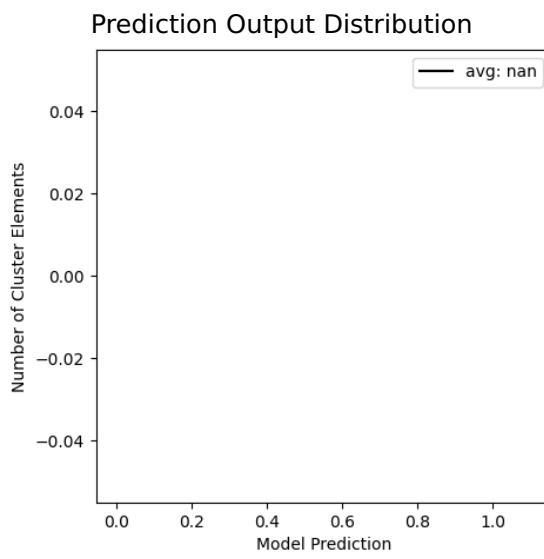
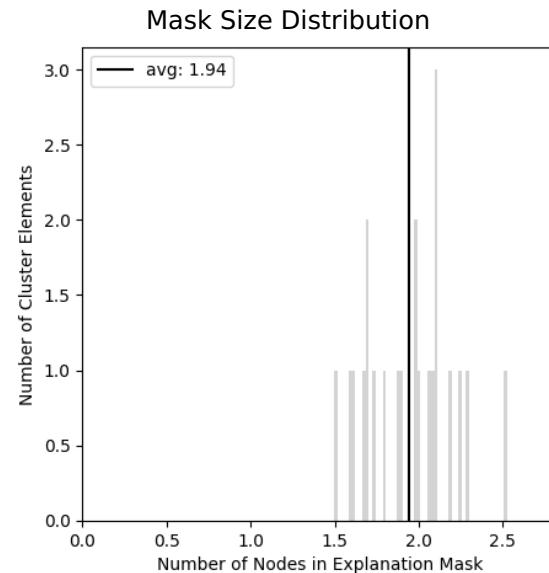
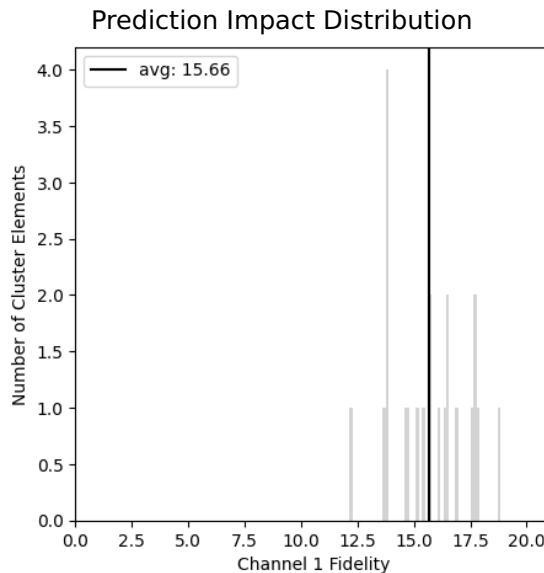
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	22
Channel Index	1.0 (0.0)

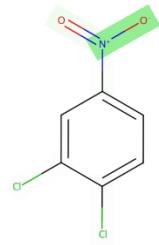
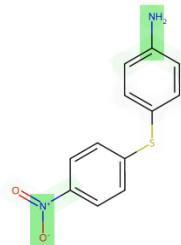
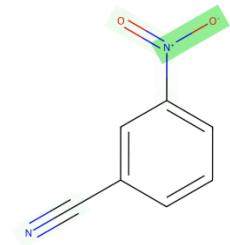
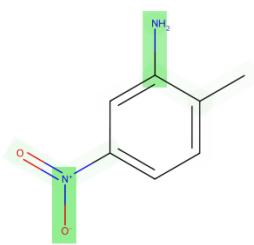
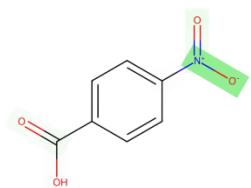
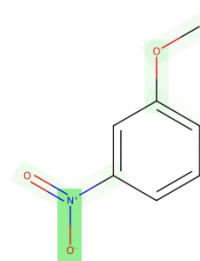
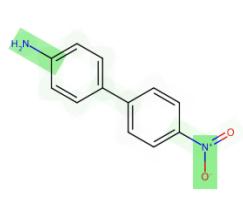
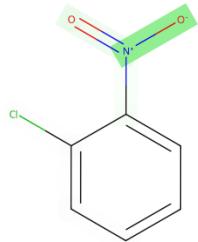
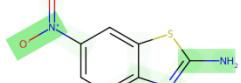
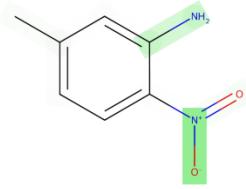
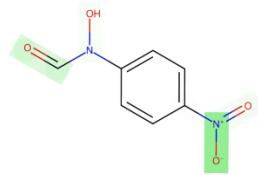
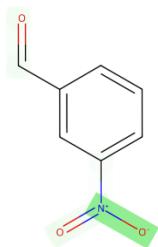
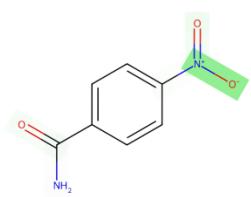
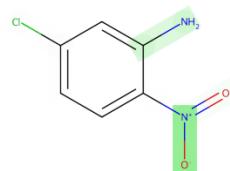
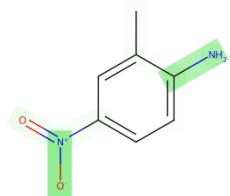
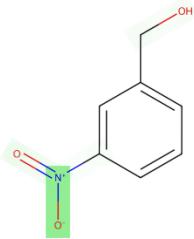
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



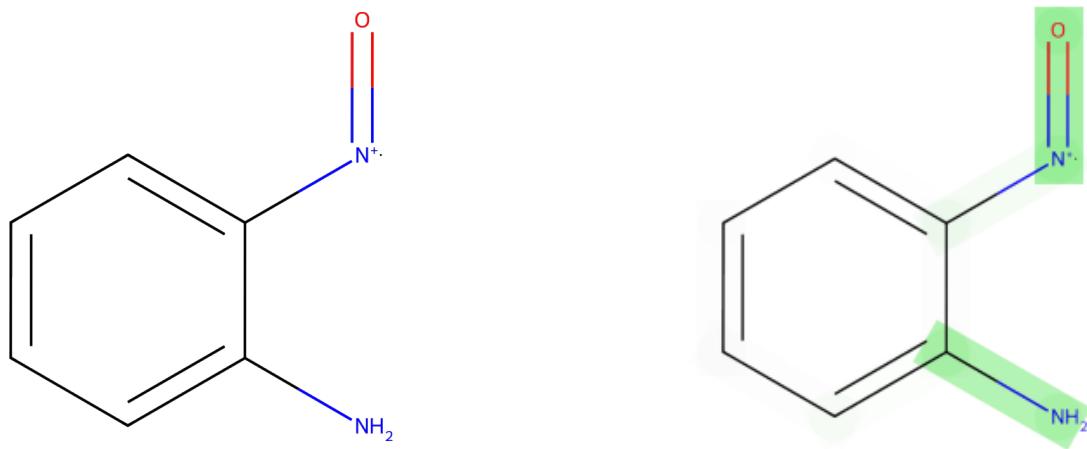
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The provided SMILES notation represents a nitrated aromatic compound, specifically a nitrobenzene derivative. Aromatic compounds are already known for their stable structures due to resonance, but when a nitro group is substituted onto the ring, as in the case of this SMILES structure, the mutagenic potential is altered. The nitro group (-NO<sub>2</sub>), particularly in this case where it is positively charged ([N+] = O), is an electron-withdrawing group. It increases the electron deficiency on the aromatic ring, making the ring more susceptible to electrophilic attack. This can increase the possibility of the molecule interacting with DNA, potentially causing mutations.

**Hypothesis:** Nitrated aromatic compounds have a medium influence on mutagenicity. The nitro group's electron-withdrawing character increases the reactivity of the aromatic ring, thus facilitating interactions with DNA that can result in genetic mutations. This is further supported by empirical evidence showing that aromatic compounds with electron-withdrawing groups like nitro groups are commonly associated with mutagenic activity.

# Cluster #47 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 47, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 15.0 ( $\pm 2.1$ ) on the prediction outcome.

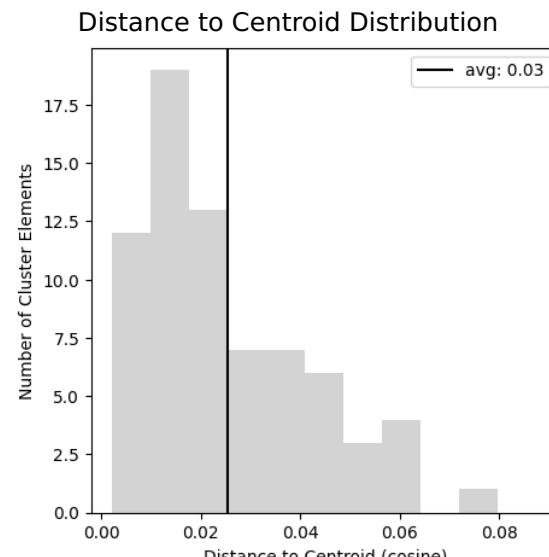
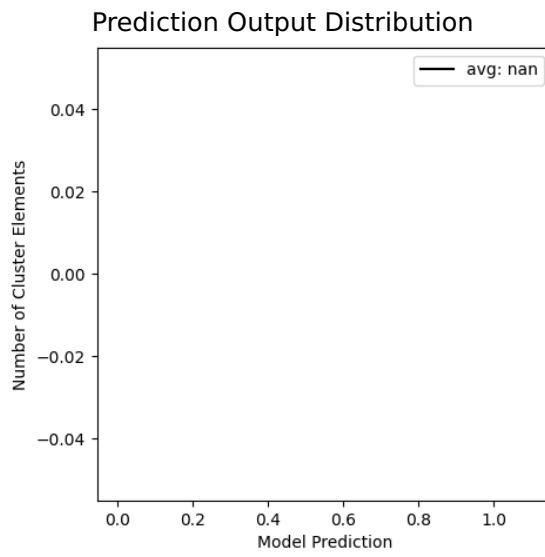
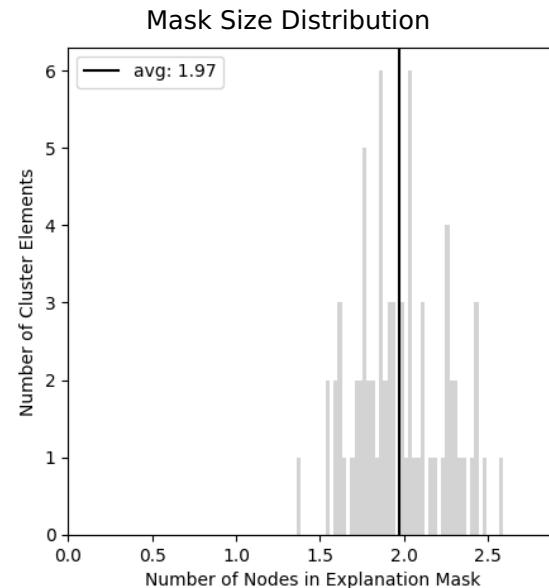
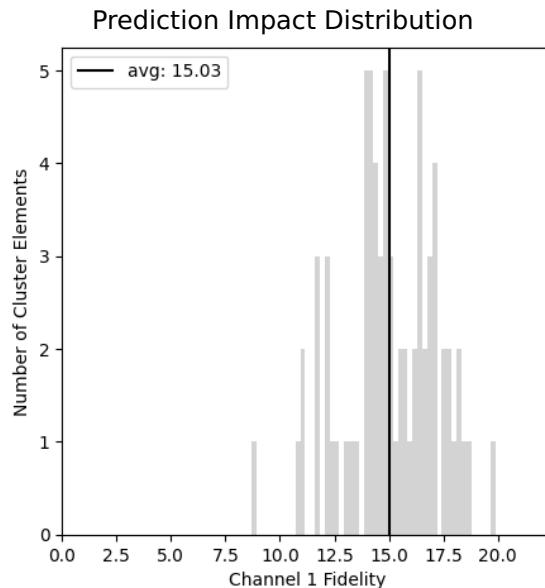
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	72
Channel Index	1.0 (0.0)

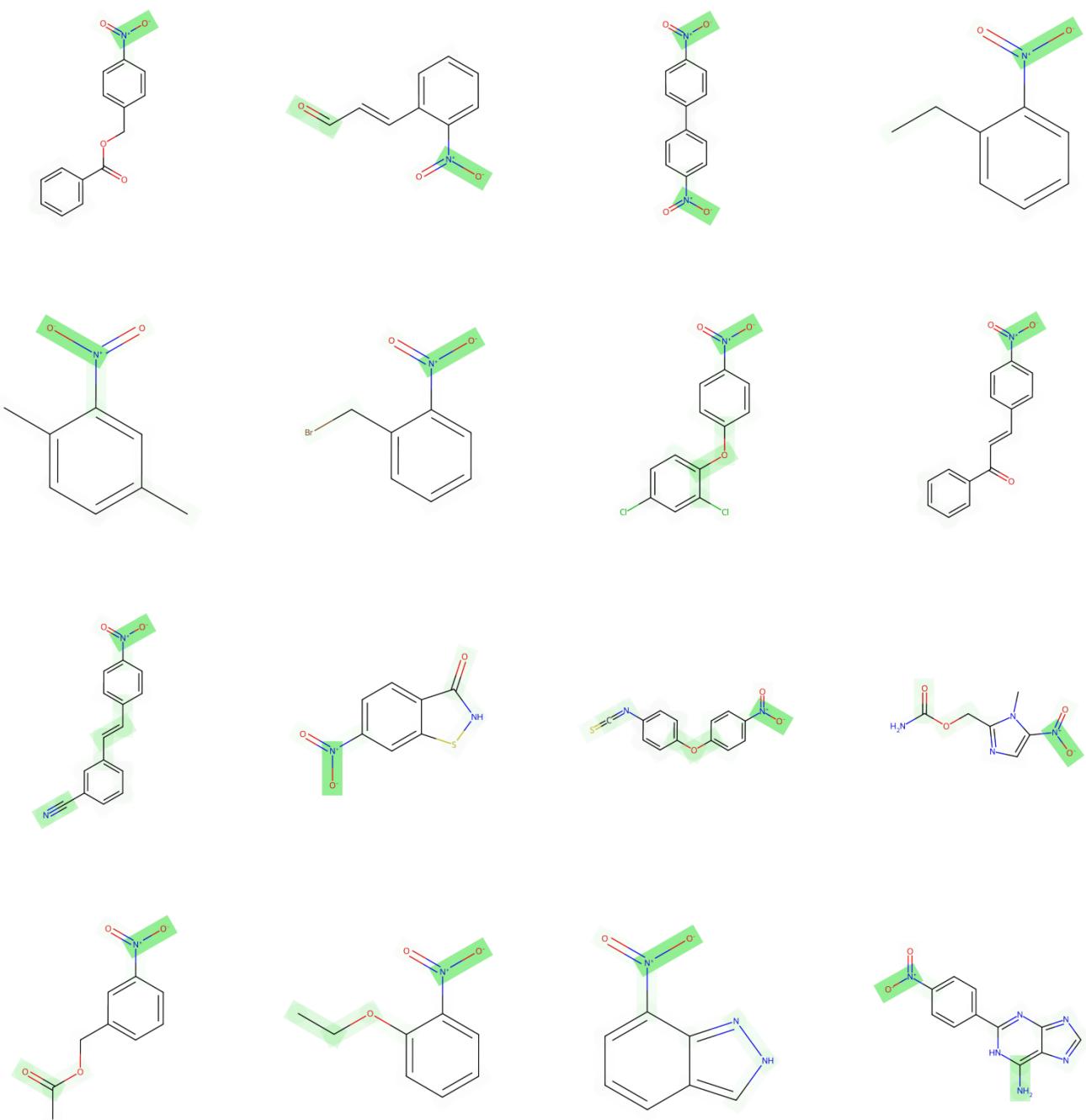
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



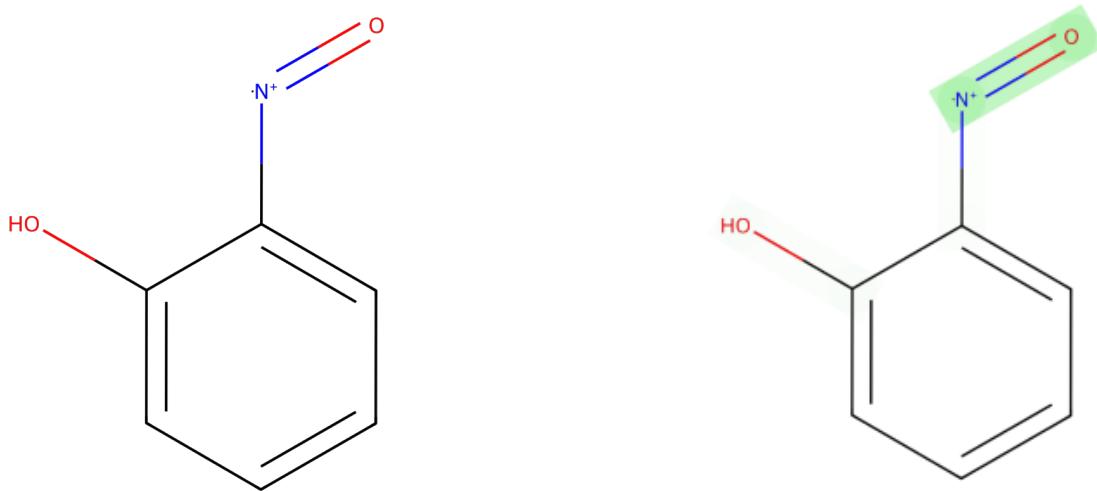
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a nitro-phenol, which includes a nitro group (-NO<sub>2</sub>) and a hydroxyl group attached to a benzene ring. The nitro group in the structure is known to undergo enzymatic reduction in vivo, forming reactive nitroso derivatives and reactive oxygen species that can cause DNA damage. Concurrently, the phenolic hydroxyl group can participate in electron transfer reactions, forming phenoxy radicals. These radicals can also interact with DNA, leading to potential mutations. These collective reactive intermediates, derived from both functional groups, can interact with genetic material, culminating in mutagenicity.

**Hypothesis:** Molecules with a nitro-phenol substructure have a medium propensity for mutagenicity. The nitro group becomes bioactivated to form reactive intermediates that can damage DNA, and the phenolic hydroxyl group can form radicals that further contribute to a molecule's mutagenic potential.

# Cluster #48 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 48, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 14.2 ( $\pm 1.8$ ) on the prediction outcome.

## Properties

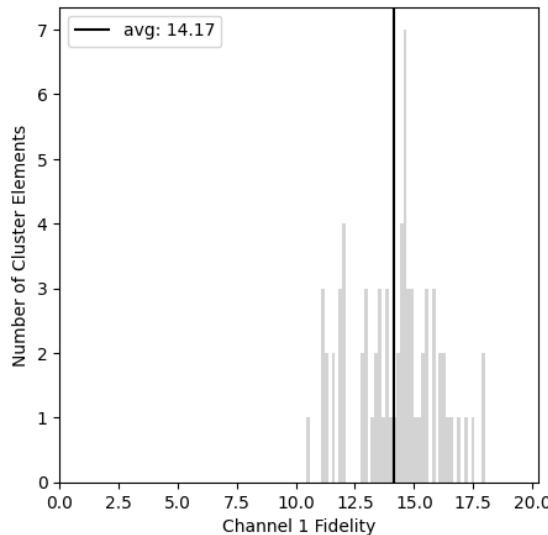
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	72
Channel Index	1.0 (0.0)

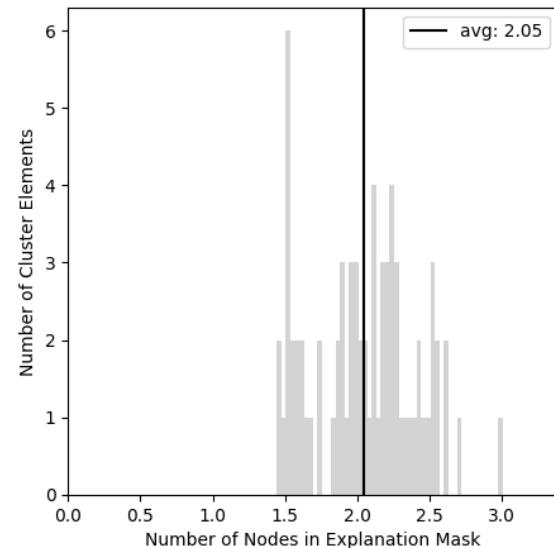
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

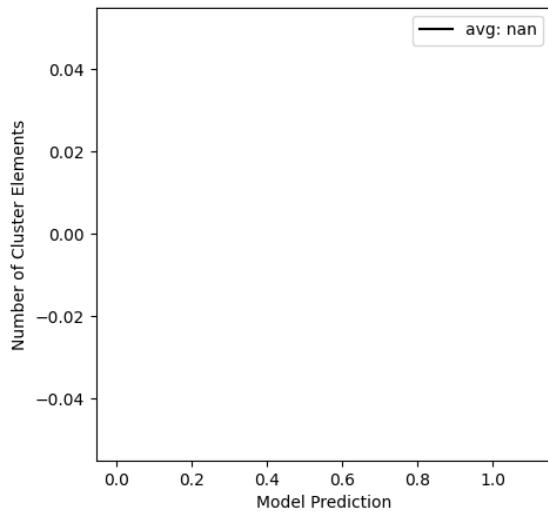
Prediction Impact Distribution



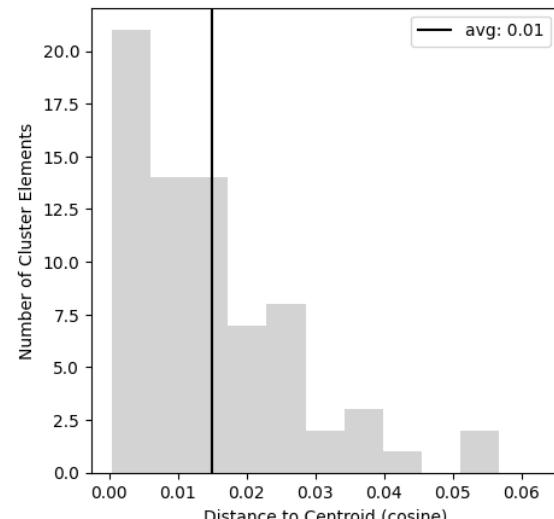
Mask Size Distribution



Prediction Output Distribution

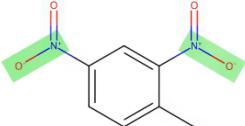
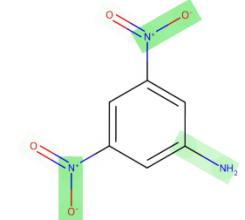
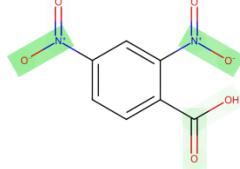
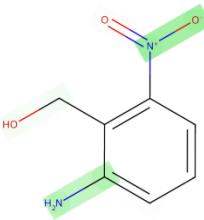
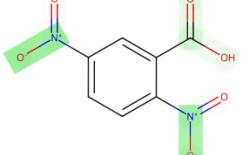
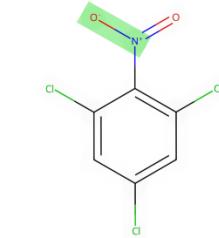
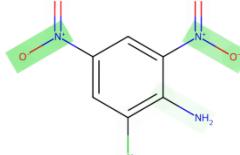
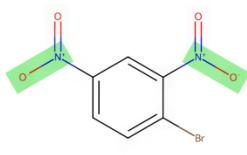
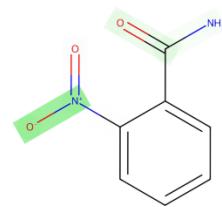
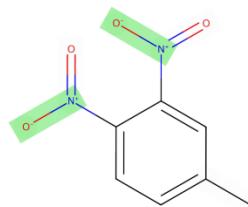
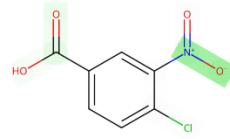
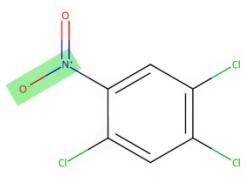
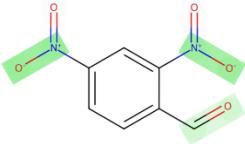
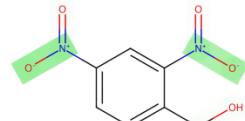
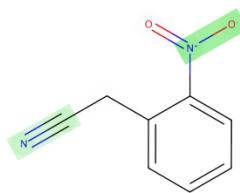
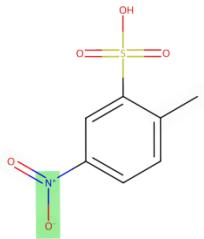


Distance to Centroid Distribution



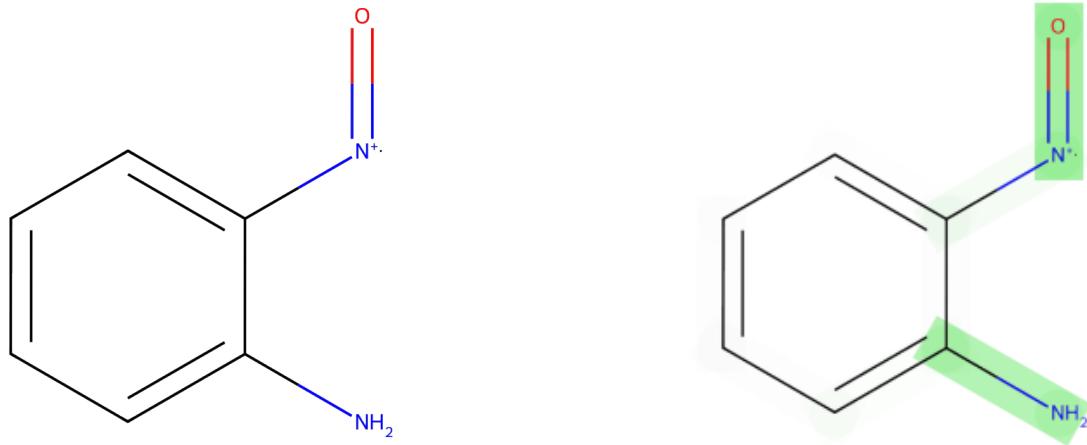
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure presented is a nitro-aromatic compound, specifically a nitro group (-NO<sub>2</sub>) attached to a benzene ring. Nitro-aromatic compounds are known to be metabolically activated *in vivo* to form reactive species that can interact with DNA, such as nitroso derivatives and reactive oxygen species (ROS). These reactive intermediates can cause oxidative damage to DNA bases, strand breaks, or covalent bonding with DNA, leading to mutagenic effects. The presence of the nitro group in the ortho-position relative to the amino group may facilitate electron transfer processes that generate these reactive species. The delocalized electrons of the benzene ring may also play a role in stabilizing the reactive intermediates, thus contributing to the mutagenic potential of the molecule.

**Hypothesis:** The presence of a nitro group attached to a benzene ring in the structure "N-c1:c:c:c:c:c1-[N+]=O" is associated with a medium influence on mutagenicity. This is likely due to the metabolic activation of the nitro group to form reactive intermediates that can damage DNA. The interaction between the nitro group and the aromatic system may further stabilize these intermediates, enhancing their ability to cause mutagenic effects.

# Cluster #49 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 49, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 16.2 ( $\pm 2.5$ ) on the prediction outcome.

## Properties

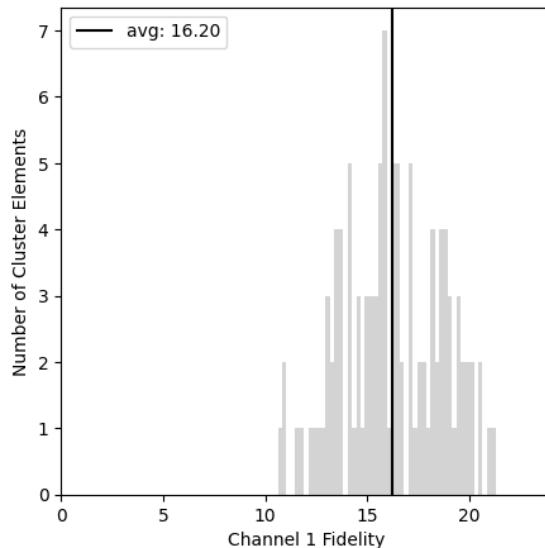
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	108
Channel Index	1.0 (0.0)

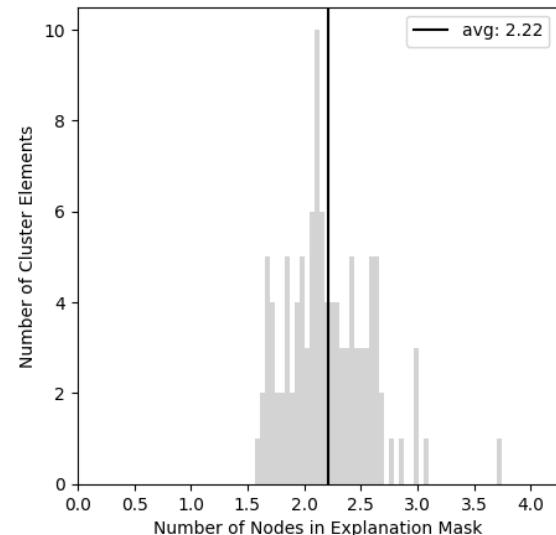
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

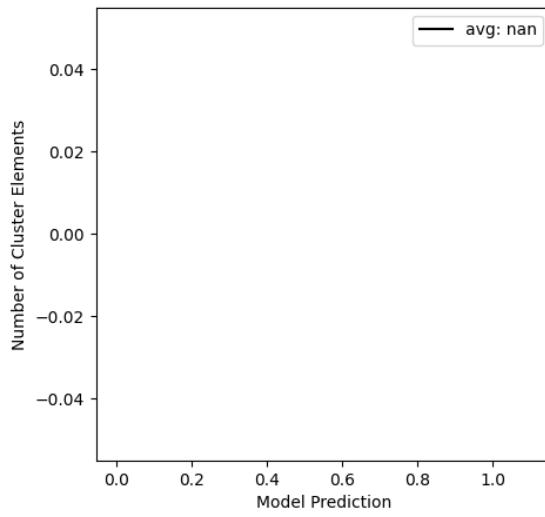
Prediction Impact Distribution



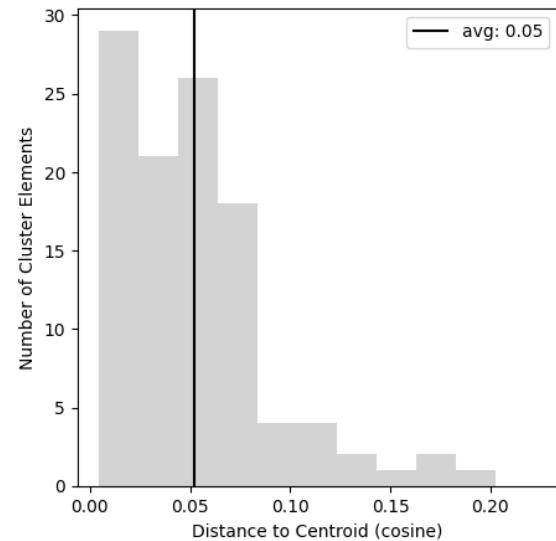
Mask Size Distribution



Prediction Output Distribution

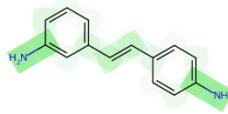
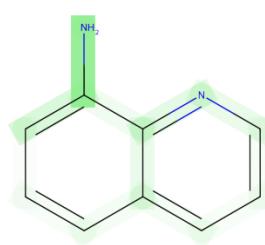
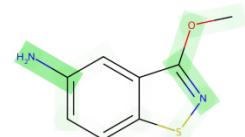
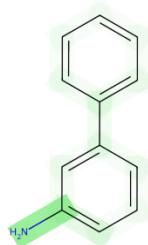
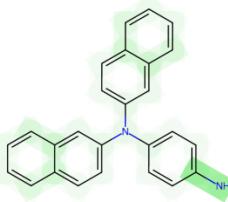
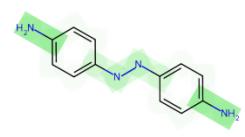
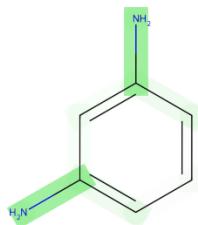
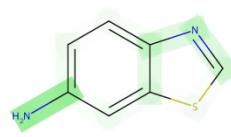
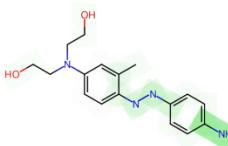
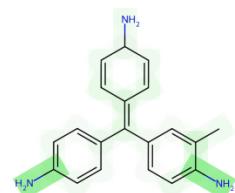
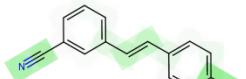
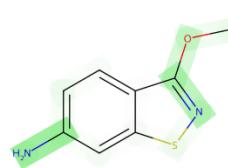
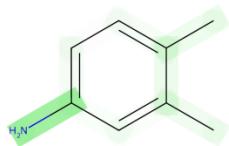
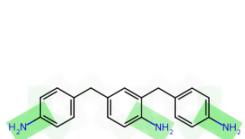
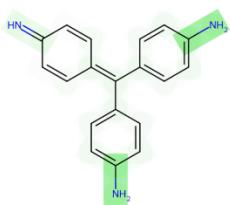


Distance to Centroid Distribution



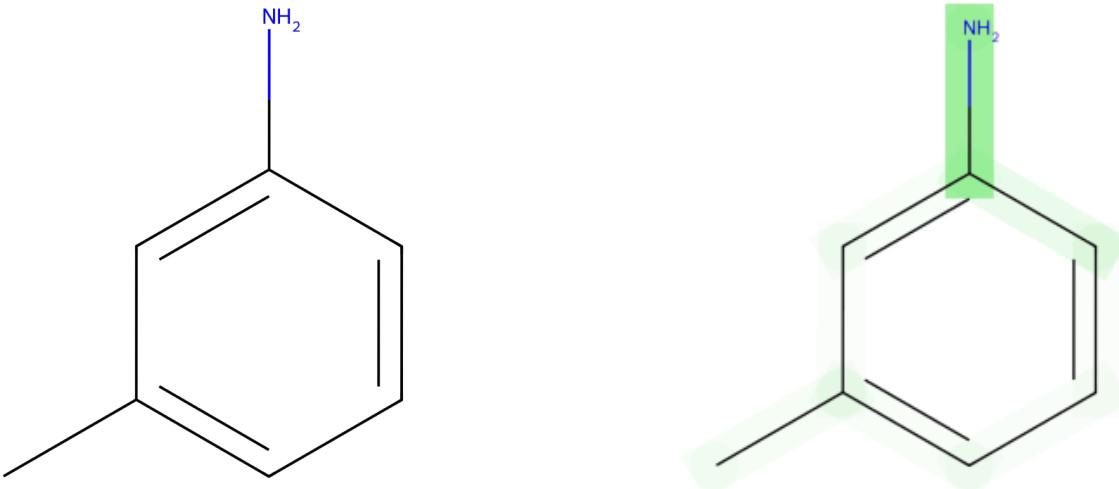
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure in the SMILES representation "C-c1:c:c:c:c(-N):c:1" depicts a benzene ring substituted with a methyl group (C-) and an amino group (-N). Aromatic amines, such as the substitution portrayed, are known to undergo metabolic activation in the body, often leading to the formation of electrophilic species that can interact with DNA, potentially causing mutations. The electron-donating effects of the amino group can increase the electron density on the benzene ring, making it more reactive towards electrophilic attack. This suggests that the particular arrangement of the substituents on the aromatic system contributes to the molecule's mutagenic potential.

**Hypothesis:** The presence of a methyl-substituted benzene ring with an amino group leads to a medium influence on mutagenicity. The electron-donating amino group increases the reactivity of the benzene ring towards electrophilic attack, which is likely to enhance the mutagenic potential. Metabolic activation can further transform this structure into an electrophilic species capable of DNA interaction, thus increasing the probability of mutation events.

# Cluster #50 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 50, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 11.0 ( $\pm 3.5$ ) on the prediction outcome.

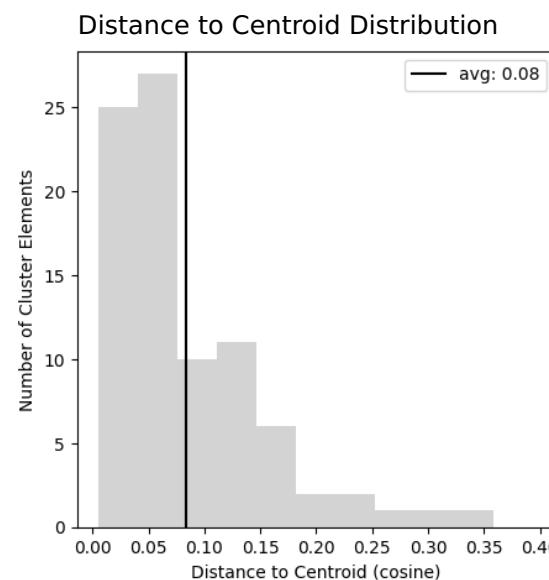
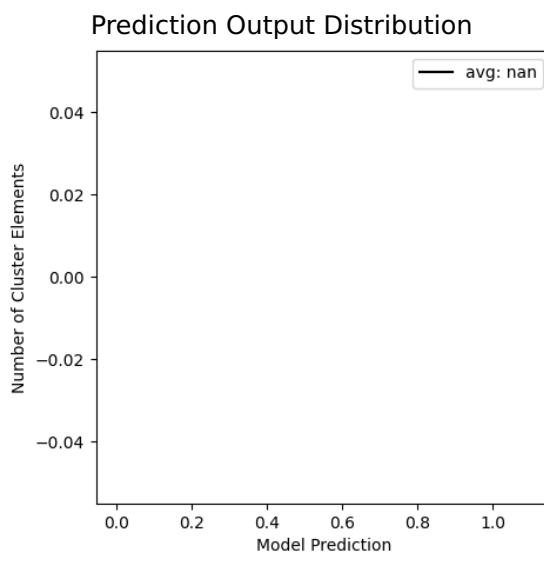
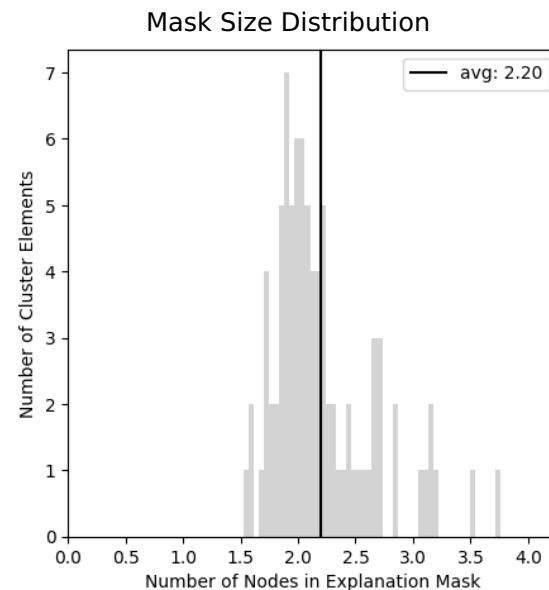
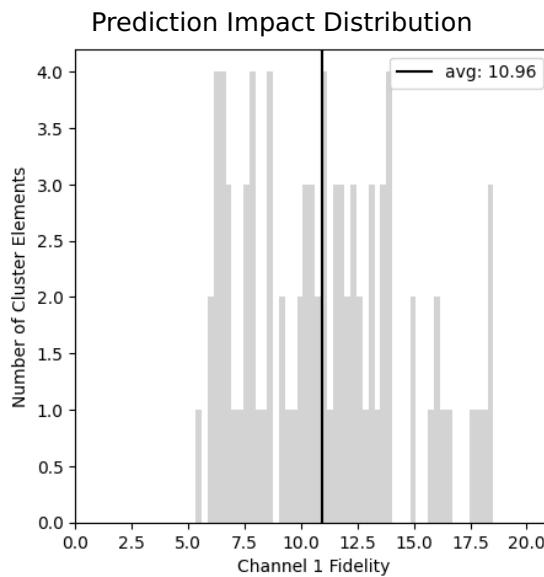
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	86
Channel Index	1.0 (0.0)

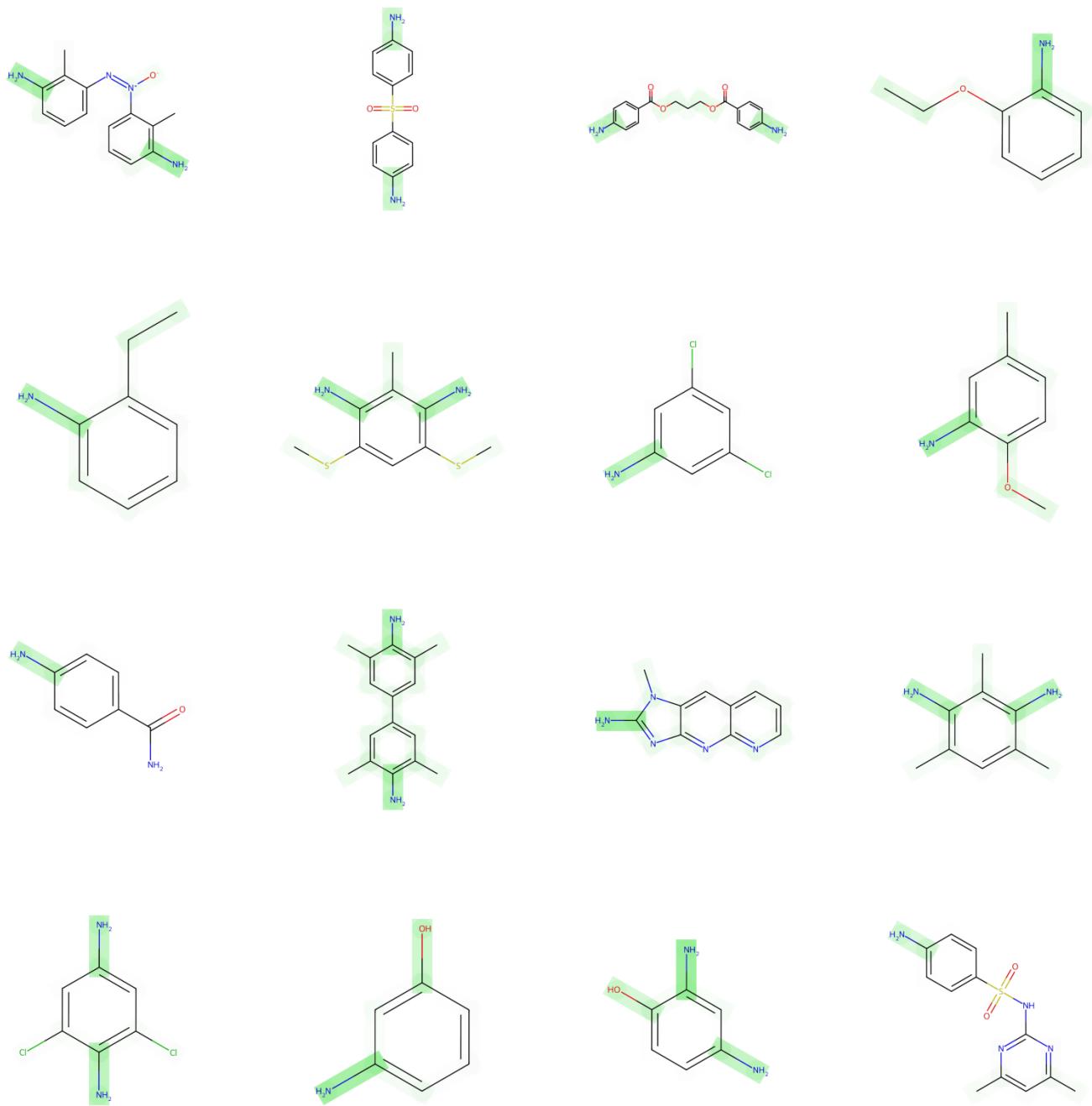
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by " $\text{N}-\text{c}1:\text{c}:\text{c}:\text{c}(-\text{O}):-\text{c}:\text{c}:1$ " is an aromatic amine with a hydroxyl group in the para position relative to the nitrogen. Aromatic amines are known intercalators, which means they can insert themselves between DNA base pairs, potentially causing mutations during DNA replication. The presence of the hydroxyl group in the para position can further stabilize the intercalating structure through hydrogen bonding, increasing the likelihood of interacting with DNA.

**Hypothesis:** The molecular substructure " $\text{N}-\text{c}1:\text{c}:\text{c}:\text{c}(-\text{O}):-\text{c}:\text{c}:1$ " exhibits a medium influence on mutagenicity. The aromatic amine can intercalate with DNA, and the hydroxyl group may enhance this interaction through additional hydrogen bonding. The inherent stability of the aromatic ring system, alongside the electron-donating effects of the nitrogen and oxygen, likely contributes to the observed mutagenic activity.

# Cluster #51 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 51, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.6 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 18.1 ( $\pm 2.4$ ) on the prediction outcome.

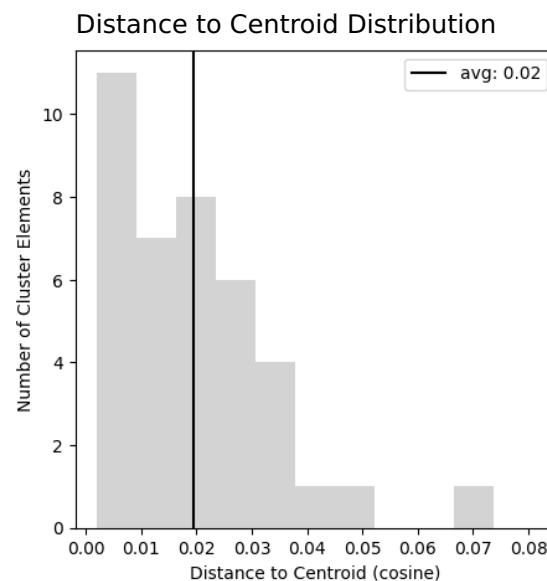
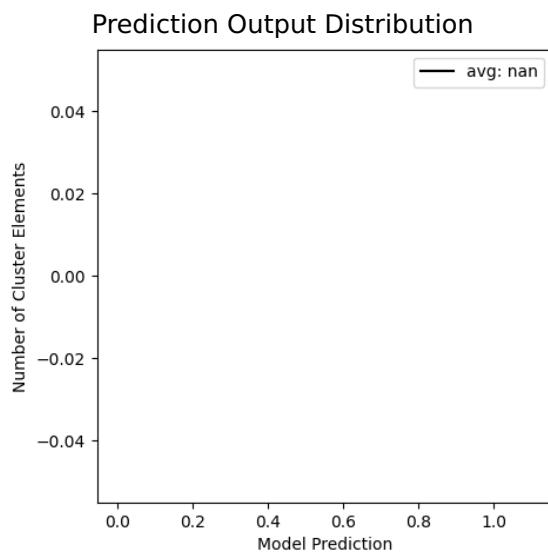
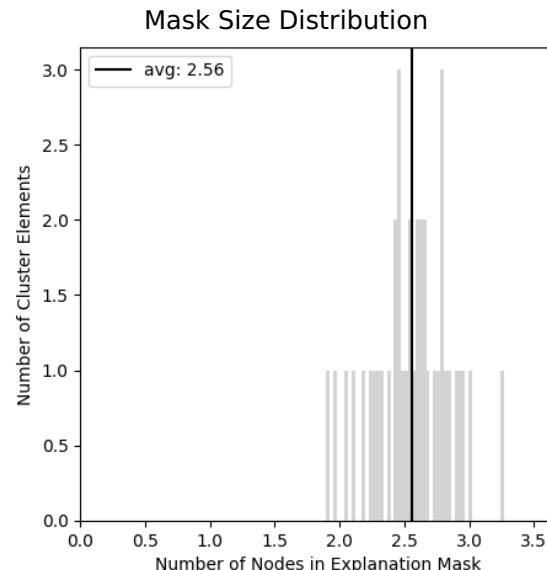
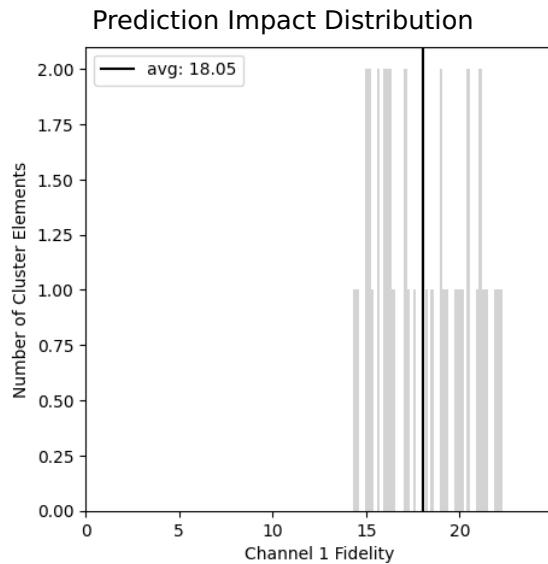
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	39
Channel Index	1.0 (0.0)

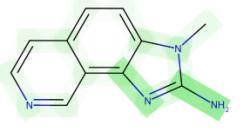
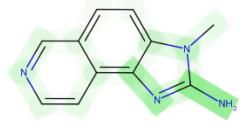
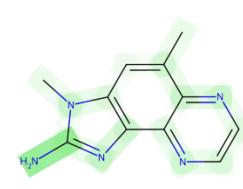
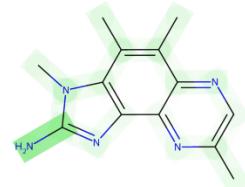
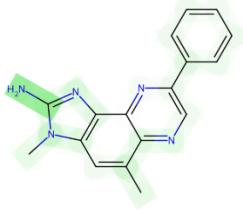
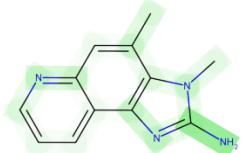
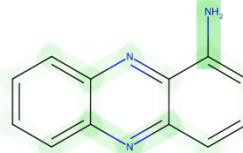
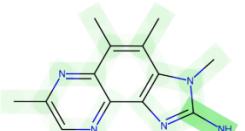
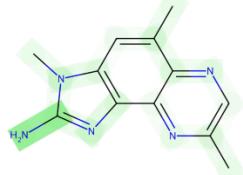
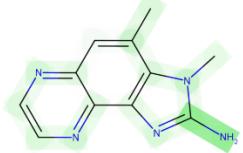
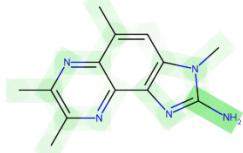
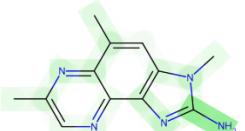
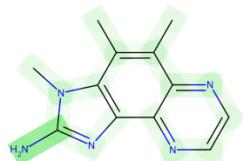
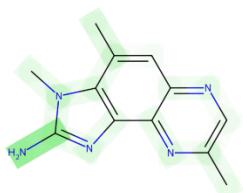
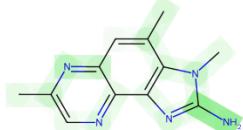
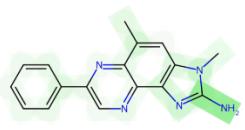
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



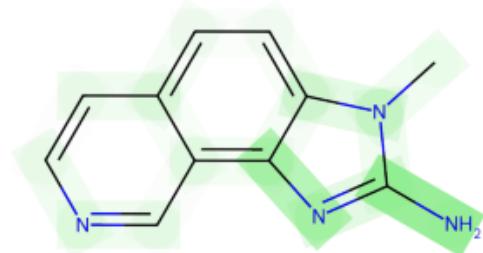
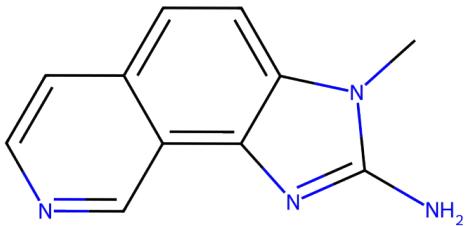
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation corresponds to a structure that appears to be a fused aromatic system containing a triazine ring and a pyridine ring. Aromatic structures, particularly those containing nitrogen atoms, are known to intercalate with DNA, leading to mutations. The presence of such ring systems increases the electron density, making the molecule more reactive towards nucleophilic sites in DNA. The triazine ring can act as an electron-deficient site, which can facilitate the formation of covalent bonds with the nucleophilic centers in DNA. This intercalation process can lead to insertions, deletions, or substitutions in the genetic material, thus affecting the mutagenicity of the molecule.

**Hypothesis:** The structure "CN1C(N)=NC2=C3C=NC=CC3=CC=C21" exhibits a medium influence on mutagenicity due to the presence of electron-rich fused aromatic rings, which can intercalate with DNA. The triazine component may increase the reactivity of the molecule, enabling it to form covalent bonds with DNA, resulting in modifications of the genetic material and contributing to the mutagenic potential.

# Cluster #52 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 52, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.7 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 16.1 ( $\pm 3.0$ ) on the prediction outcome.

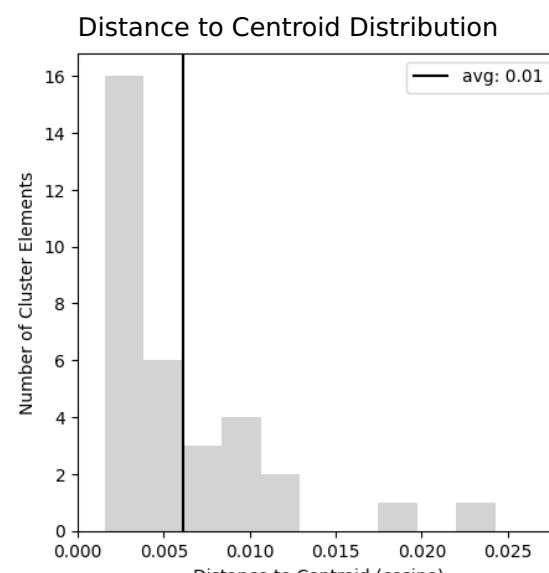
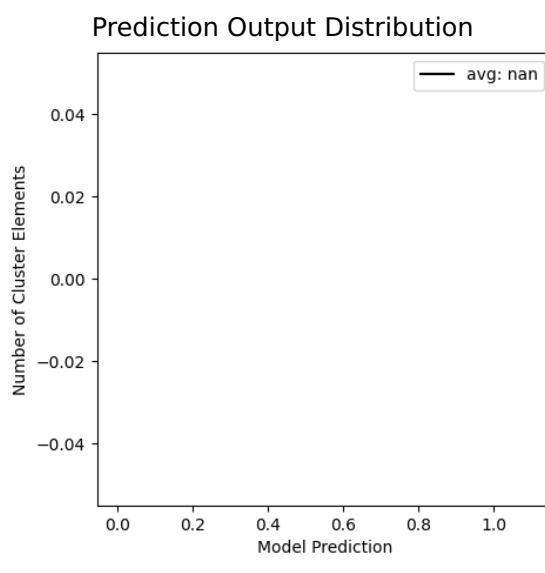
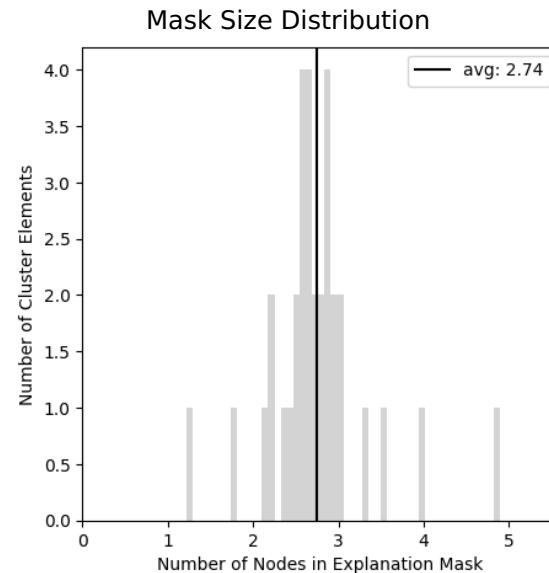
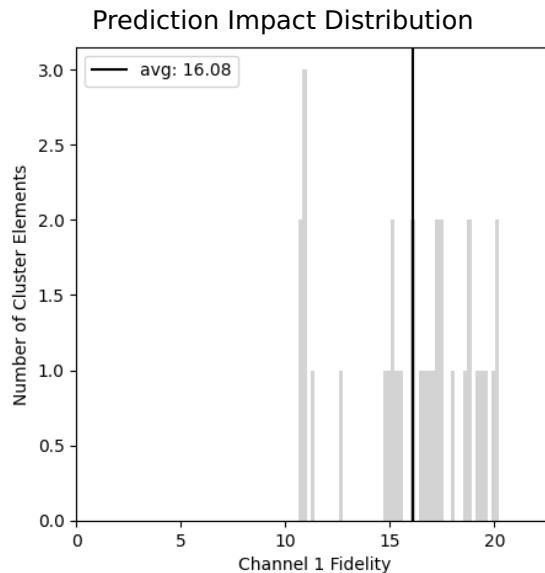
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	33
Channel Index	1.0 (0.0)

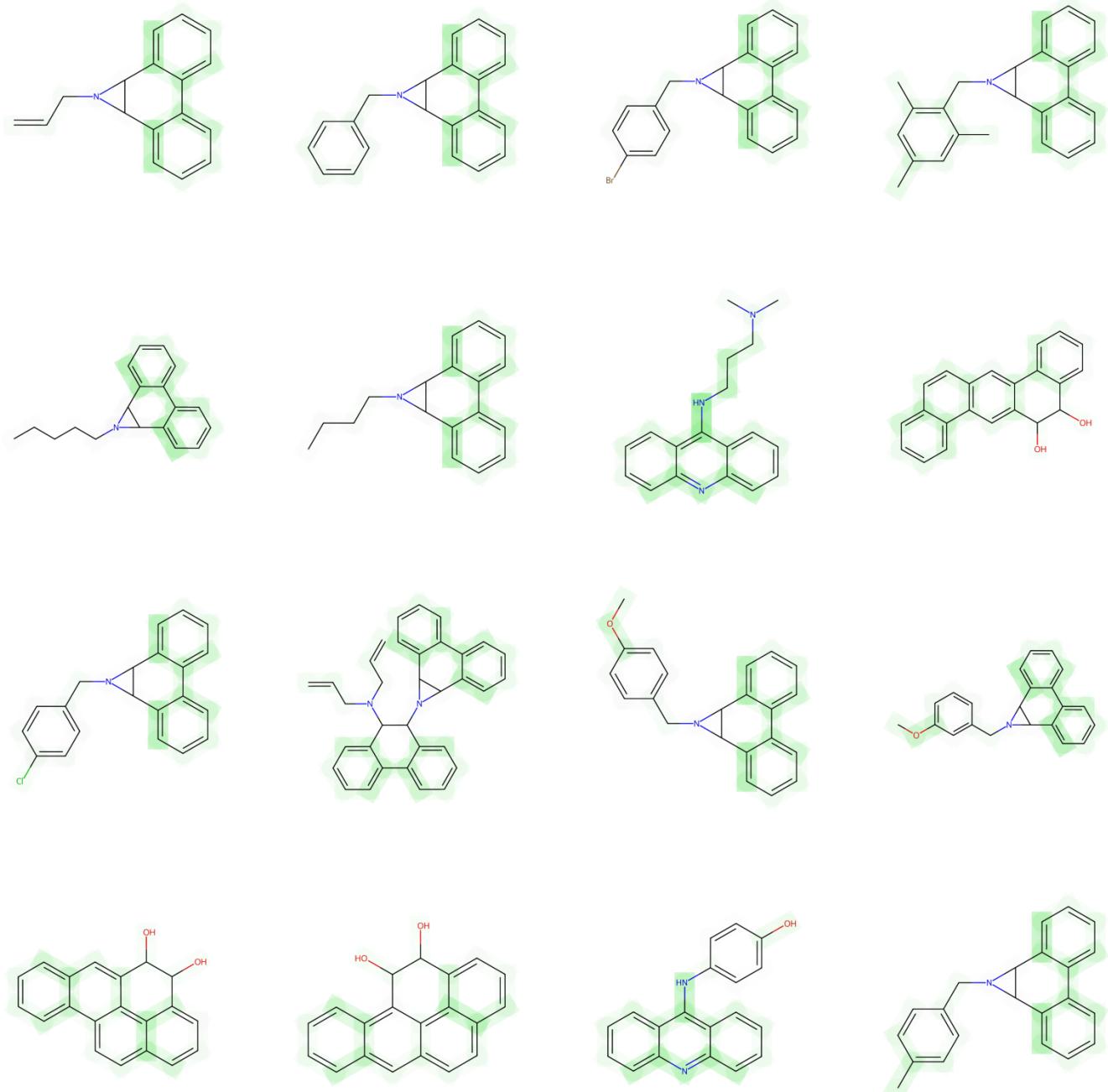
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



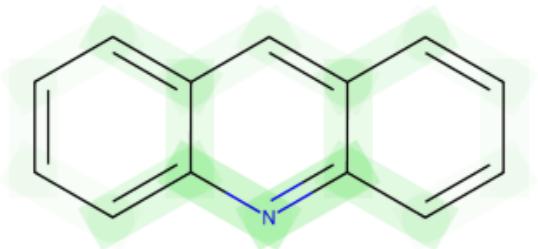
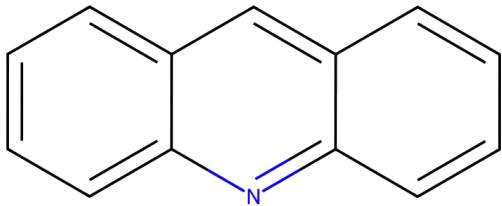
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the given SMILES notation is a polycyclic aromatic compound that features fused benzene rings forming an extended planar system. This type of structure is known for its potential to intercalate between DNA base pairs due to its planar geometry, which can lead to disruption of DNA replication and transcription, consequently increasing the risk of mutations. The presence of a nitrogen atom within the ring system can also contribute to reactivity, potentially forming adducts with DNA nucleophiles. Such interactions can lead to structural chromosomal aberrations or point mutations.

**Hypothesis:** The mutagenicity of the molecule related to the SMILES "c1:c:c:c2:n:c3:c:c:c:c3:c:c:2:c:1" is attributed to its polycyclic aromatic structure capable of DNA intercalation and a nitrogen heteroatom that can facilitate the formation of DNA adducts. This molecular arrangement disrupts normal DNA processes, which subsequently can lead to an increased occurrence of mutations.

# Cluster #53 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 53, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 15.3 ( $\pm 3.0$ ) on the prediction outcome.

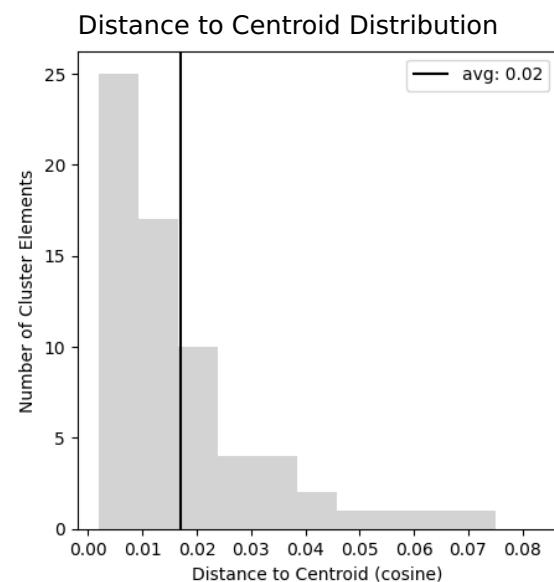
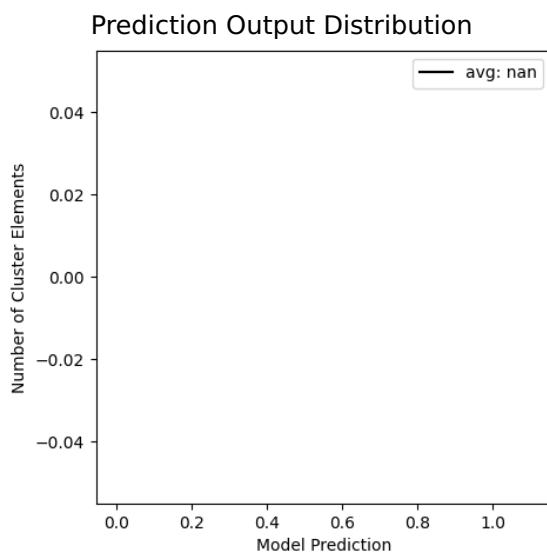
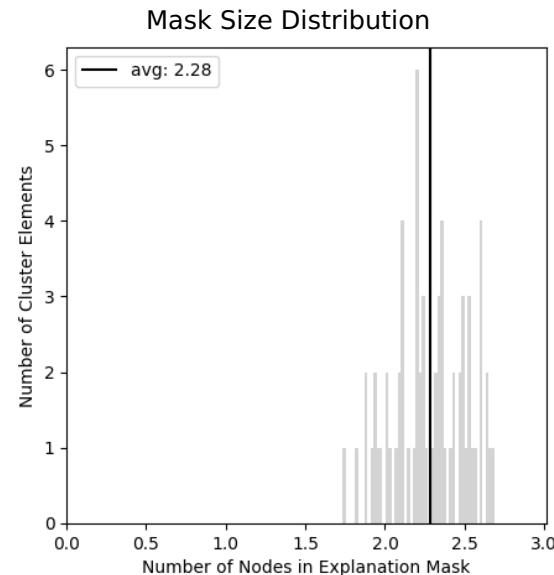
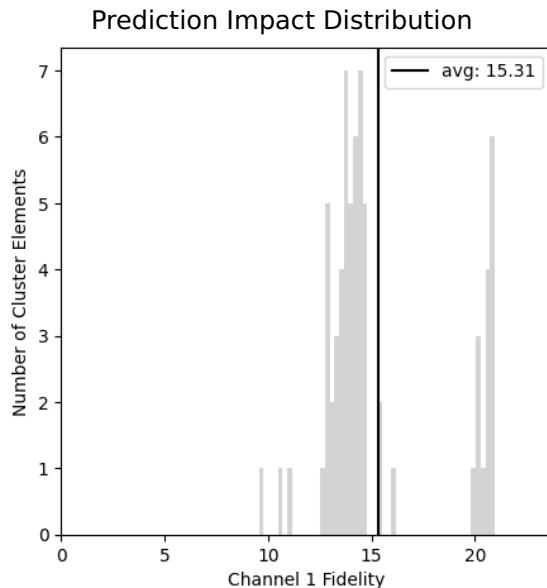
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	66
Channel Index	1.0 (0.0)

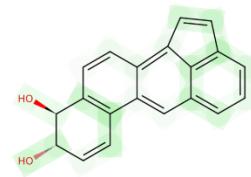
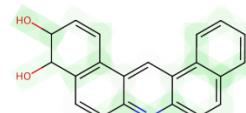
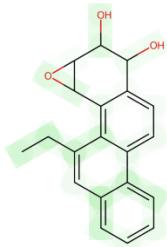
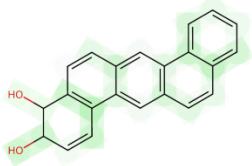
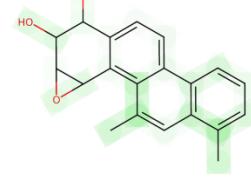
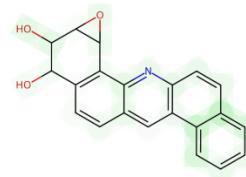
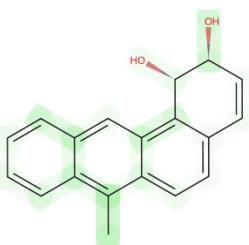
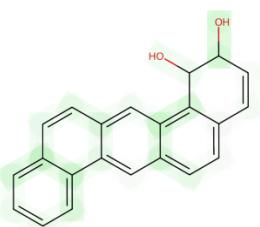
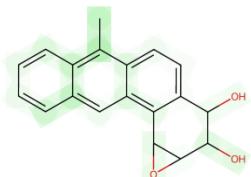
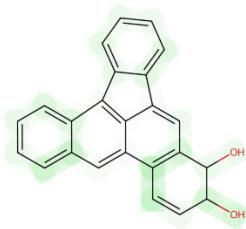
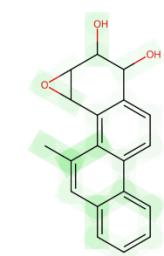
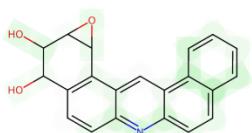
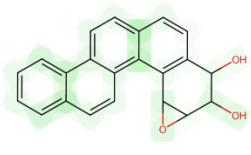
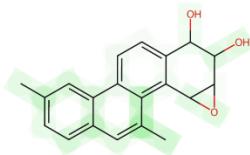
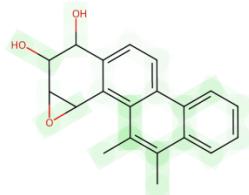
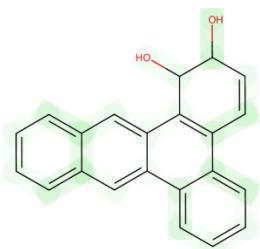
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



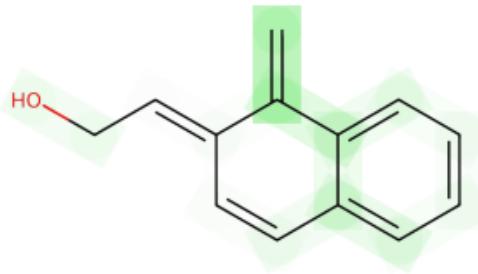
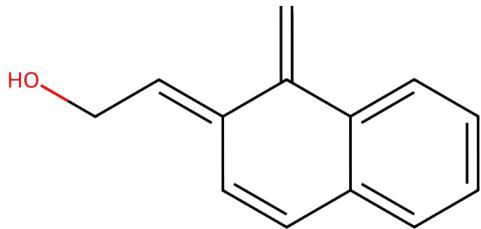
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a conjugated aromatic system with an enol ether moiety (-C-C-O group linked to an aromatic ring). Conjugated systems can allow for the delocalization of electrons, making the molecule more reactive and potentially able to interact with nucleic acids. Additionally, the presence of the double bond outside the aromatic system (enol ether) could act as an alkylating site, further enhancing the molecule's ability to react with DNA and cause mutations.

**Hypothesis:** The presence of a conjugated aromatic system with an enol ether moiety in a molecule is hypothesized to have a medium influence on the molecule's mutagenicity. The delocalization of electrons across the conjugated system may increase molecular reactivity, while the enol ether could serve as an alkylating site, both factors potentially leading to DNA interactions and subsequent mutations.

# Cluster #54 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 54, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 17.9 ( $\pm 2.5$ ) on the prediction outcome.

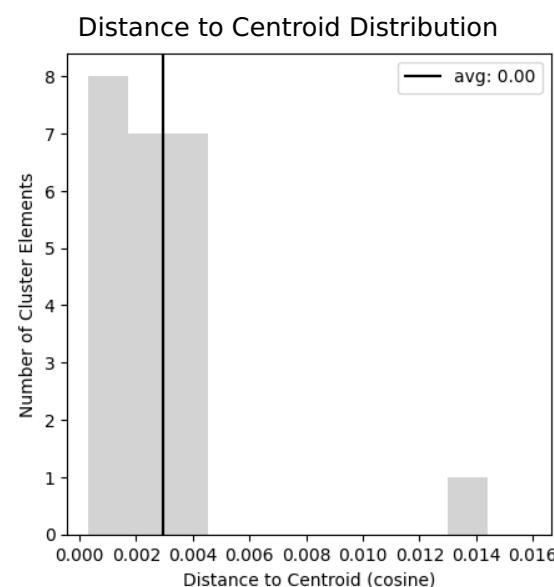
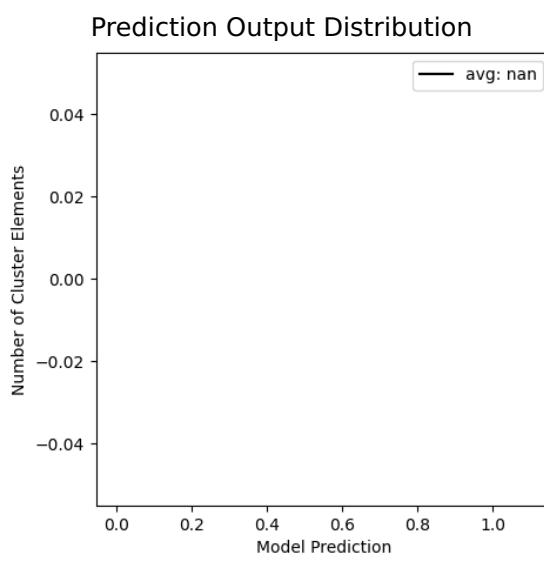
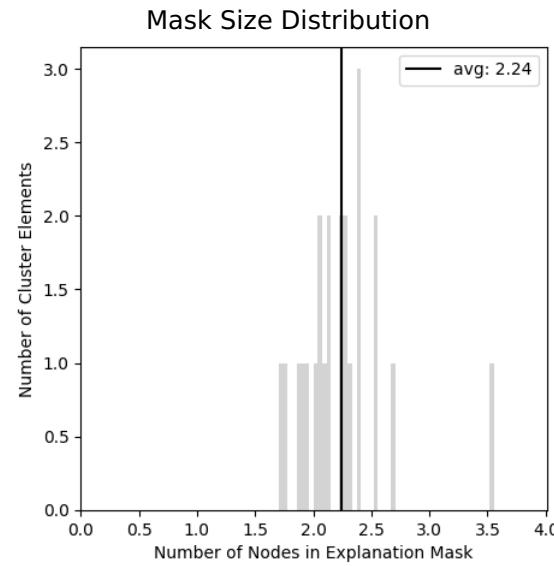
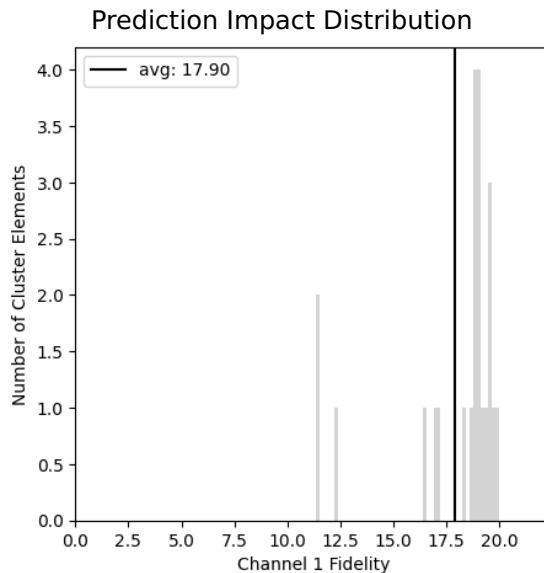
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	23
Channel Index	1.0 (0.0)

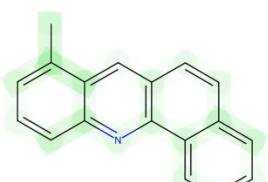
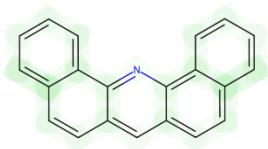
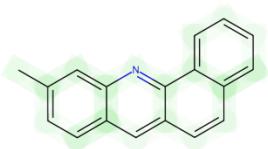
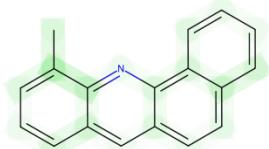
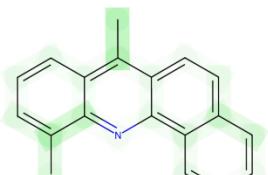
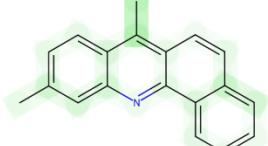
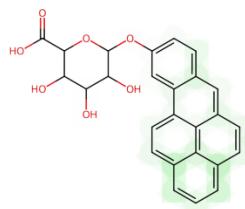
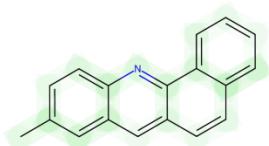
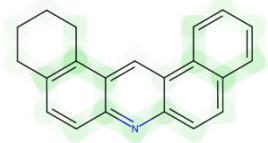
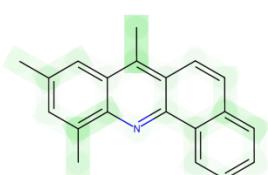
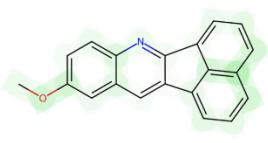
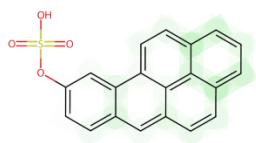
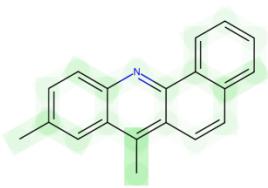
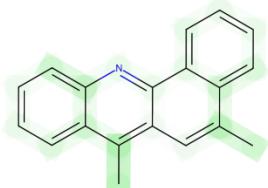
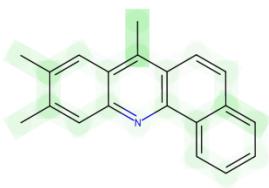
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



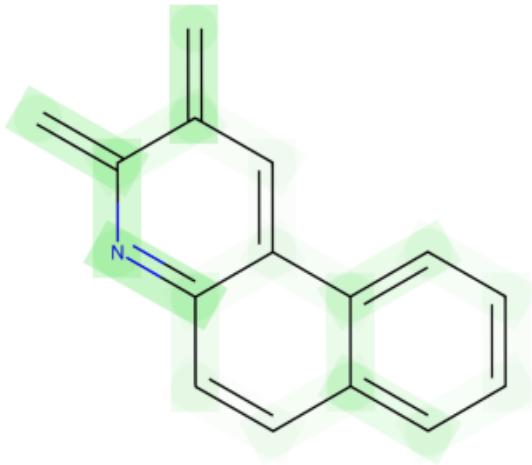
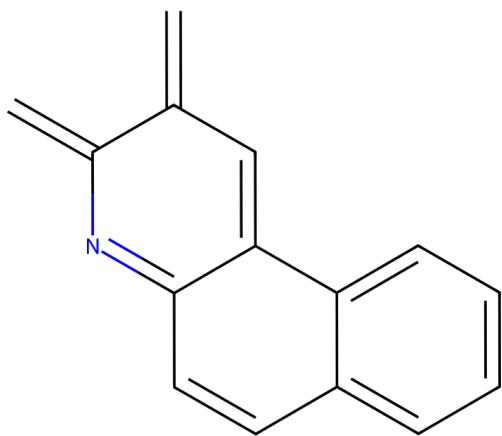
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure provided represents a polycyclic aromatic hydrocarbon containing conjugated double bonds and a nitrogen atom incorporated into the ring system. These types of structures are known to intercalate into DNA, inserting themselves between base pairs. This can disrupt the DNA structure and potentially cause errors during replication. The presence of the nitrogen atom can further increase the reactivity of the compound, making it more likely to form covalent bonds with the DNA, resulting in mutations.

**Hypothesis:** The given structure with multiple aromatic rings and a nitrogen atom has a medium influence on mutagenicity. The extended conjugation system allows the molecule to absorb in the UV-Vis region, and the molecule's planarity enables it to stack between DNA bases, leading to a medium likelihood of causing genetic mutations. The presence of the nitrogen atom may allow it to act as a nucleophile or an electrophile, contributing to its mutagenic potential.

# Cluster #55 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 55, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 19.1 ( $\pm 1.6$ ) on the prediction outcome.

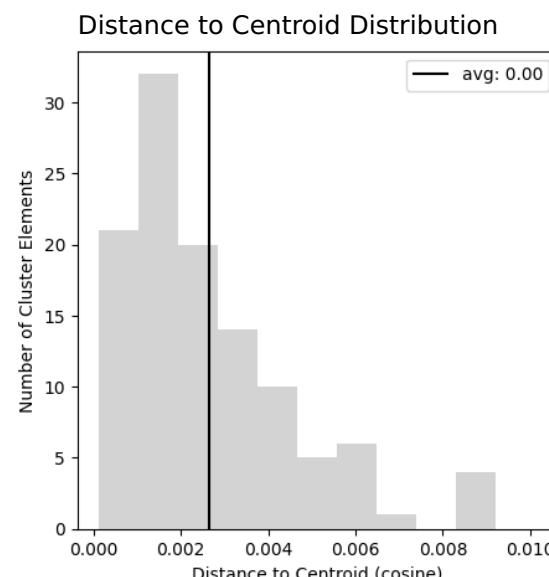
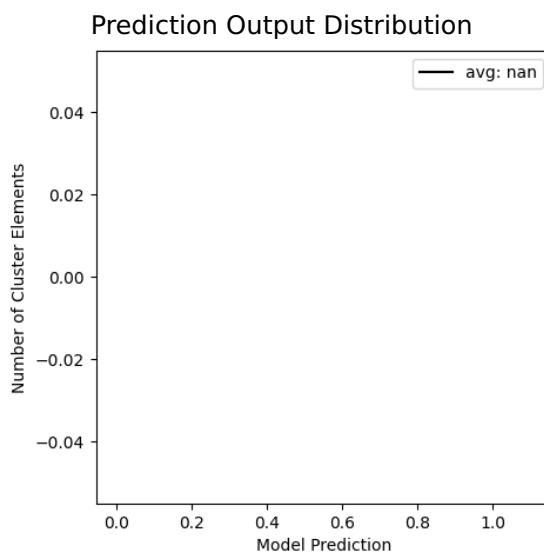
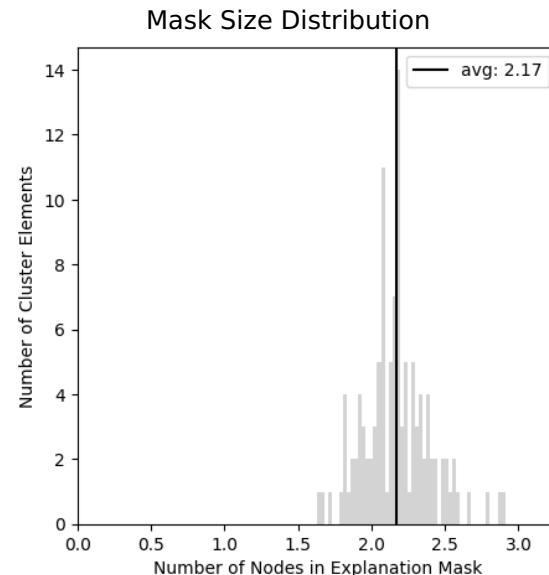
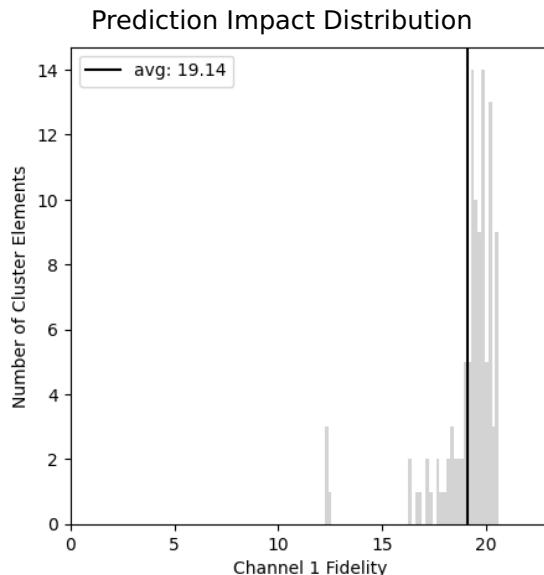
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	113
Channel Index	1.0 (0.0)

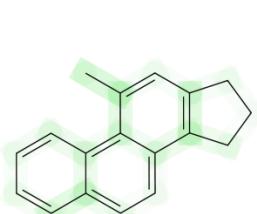
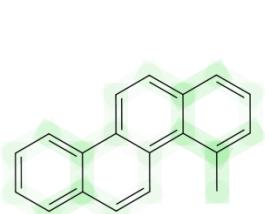
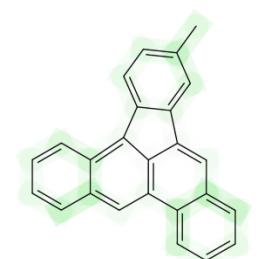
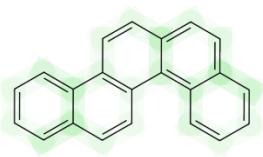
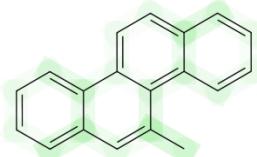
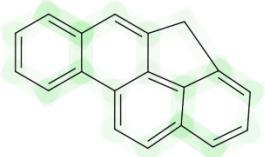
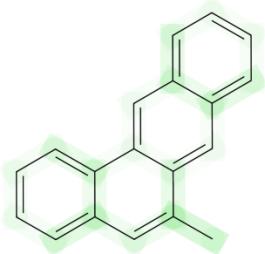
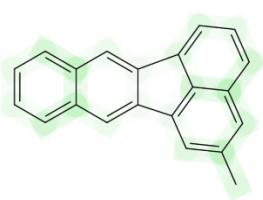
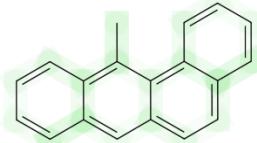
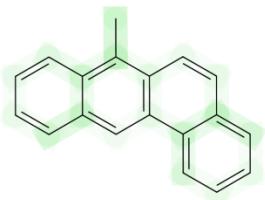
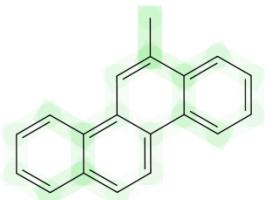
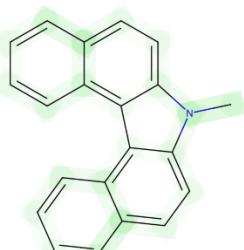
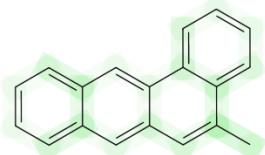
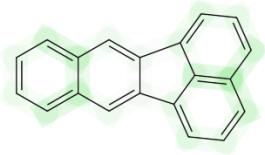
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



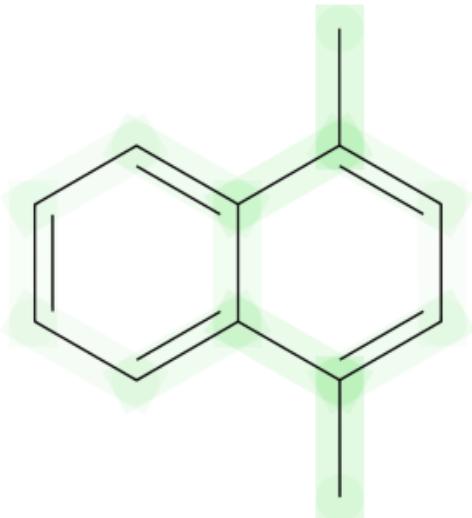
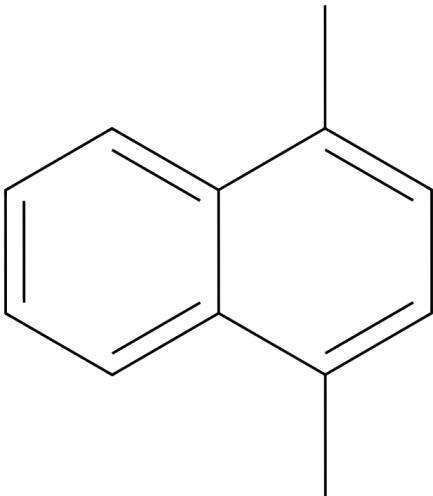
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented in the SMILES notation "C-c1:c:c:c(-C):c2:c:c:c:c:1:2" describes a polycyclic aromatic hydrocarbon (PAH) substructure, which is known for its planar and stable ring system allowing it to easily intercalate between DNA base pairs. This intercalation can disrupt the normal DNA duplication process, leading to errors in replication or transcription which might lead to mutations. Moreover, PAHs can be metabolically activated in the body to form reactive epoxides that covalently bind to DNA, forming adducts, which is another pathway to induce mutagenicity.

**Hypothesis:** Molecules with the substructure "C-c1:c:c:c(-C):c2:c:c:c:c:1:2" have a medium mutagenic influence. This is likely due to the structure's ability to intercalate into DNA or to be metabolically activated to reactive species that form DNA adducts, both of which can induce mutations during cell replication.

# Cluster #56 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 56, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 14.8 ( $\pm 0.9$ ) on the prediction outcome.

## Properties

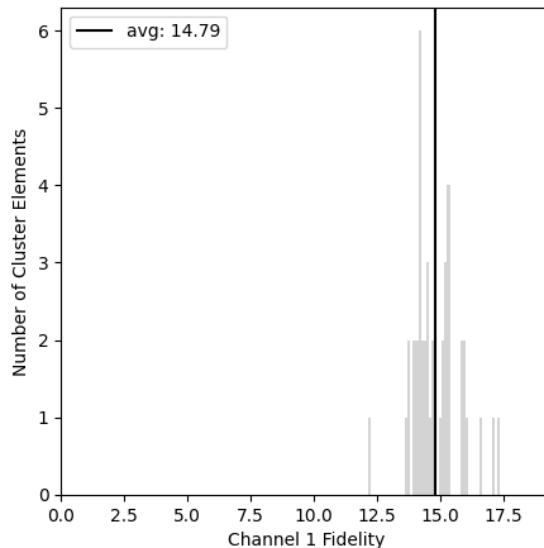
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	45
Channel Index	1.0 (0.0)

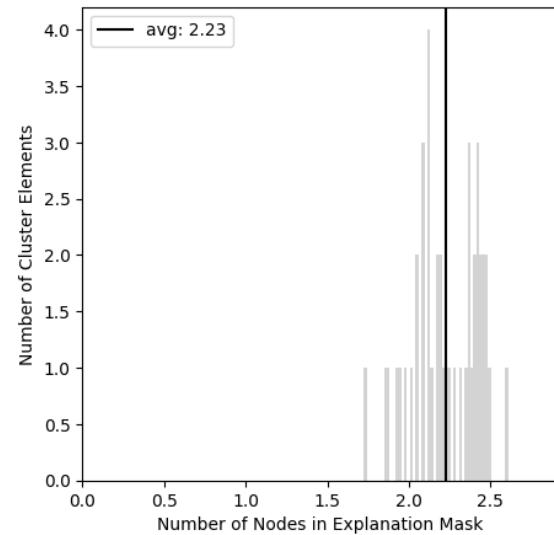
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

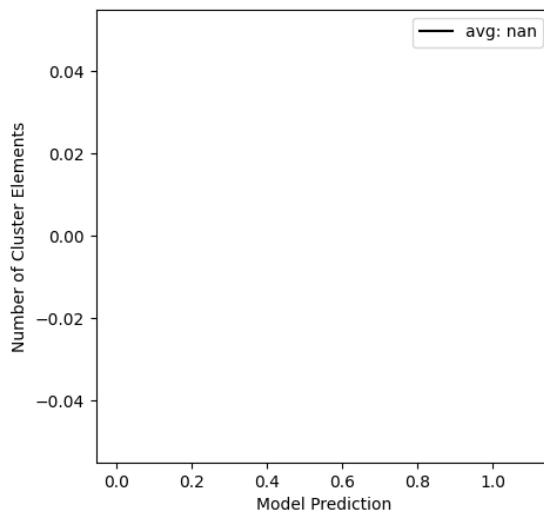
Prediction Impact Distribution



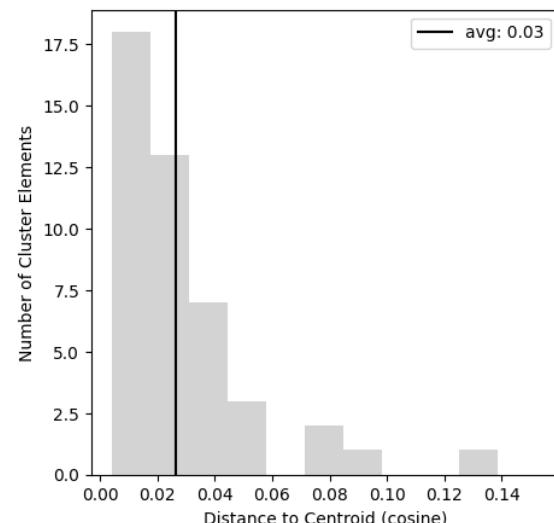
Mask Size Distribution



Prediction Output Distribution

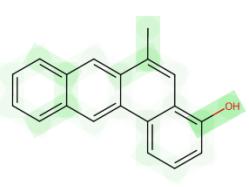
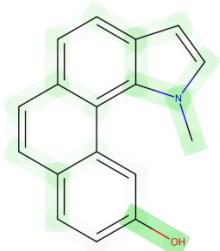
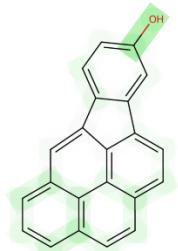
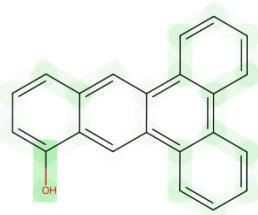
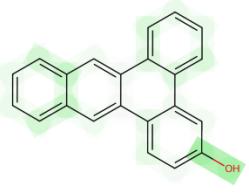
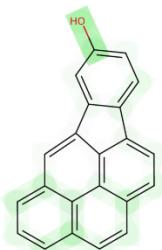
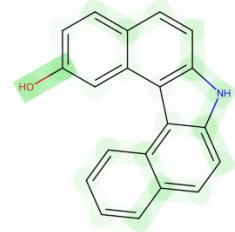
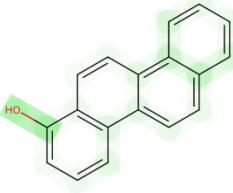
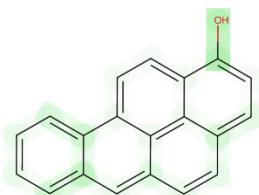
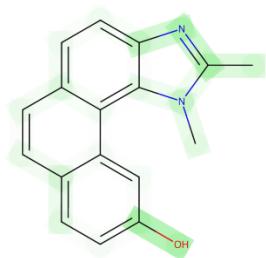
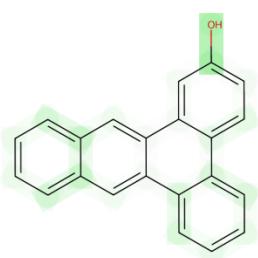
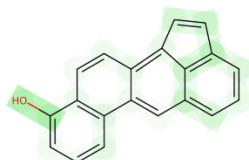
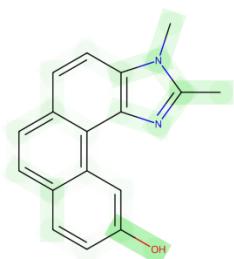
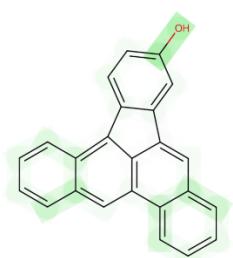


Distance to Centroid Distribution



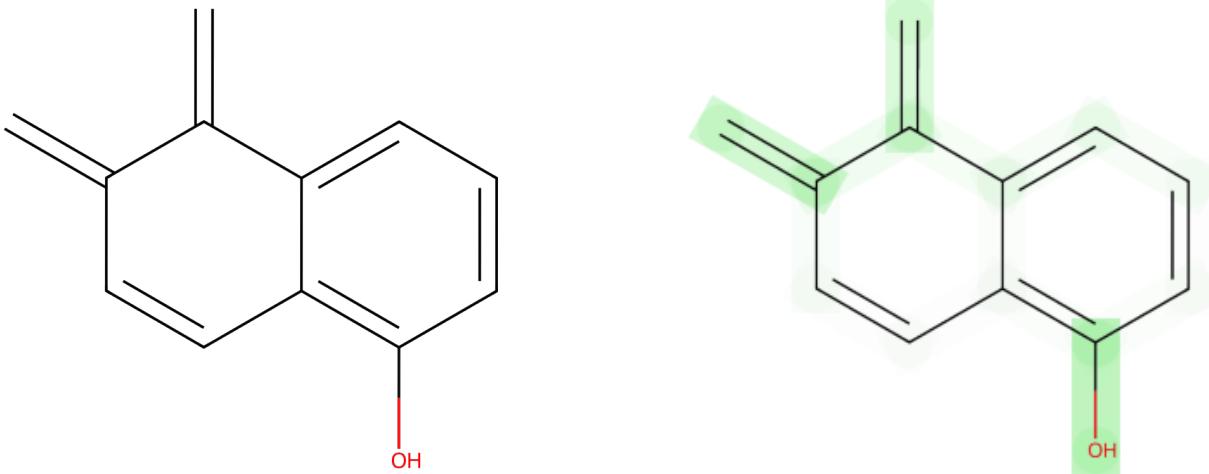
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure "C=c1:c:c:c2:c(-O):c:c:c:c2:c:1=C" represents a polycyclic aromatic compound with a hydroxyl group (-OH) attached. Polycyclic aromatic hydrocarbons (PAHs) are known to interact with DNA, causing mutations primarily through the formation of adducts after metabolic activation. The presence of the hydroxyl group could potentially influence the metabolic activation pathway, affecting the mutagenicity of the compound. In this case, the hydroxyl group might either stabilize reactive intermediates or alter the molecular conformation, which modifies the molecule's interaction with DNA.

**Hypothesis:** The given structure, a polycyclic aromatic system with conjugated double bonds and a hydroxyl substituent, is hypothesized to have a MEDIUM mutagenic influence. This is likely due to the hydroxyl group's impact on the molecule's reactivity and the overall polycyclic structure's ability to form stable DNA adducts, leading to mutations upon metabolic activation.

# Cluster #57 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 57, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of 13.3 ( $\pm 1.9$ ) on the prediction outcome.

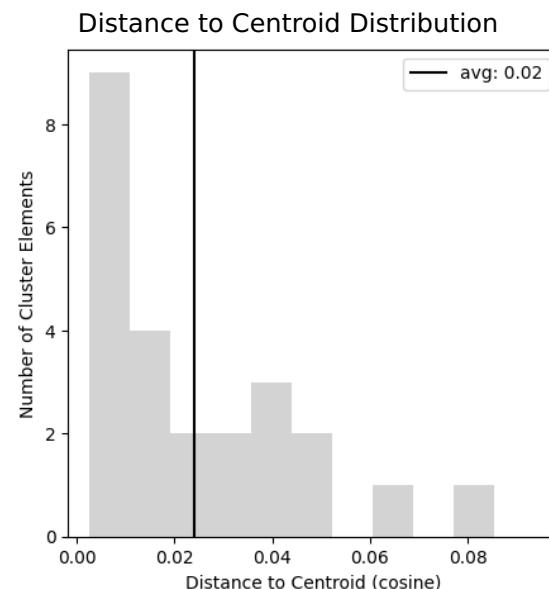
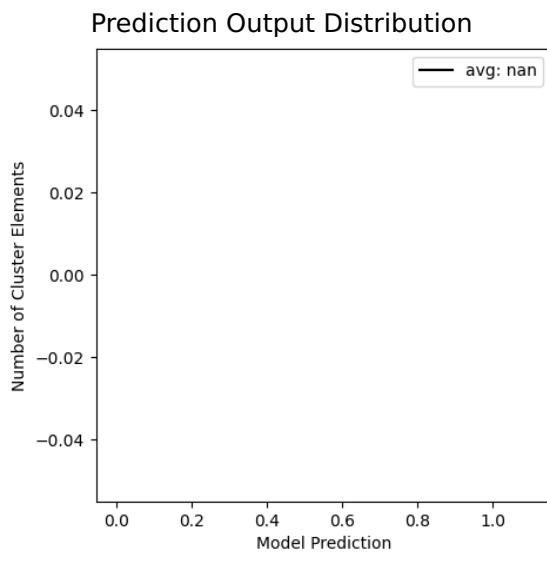
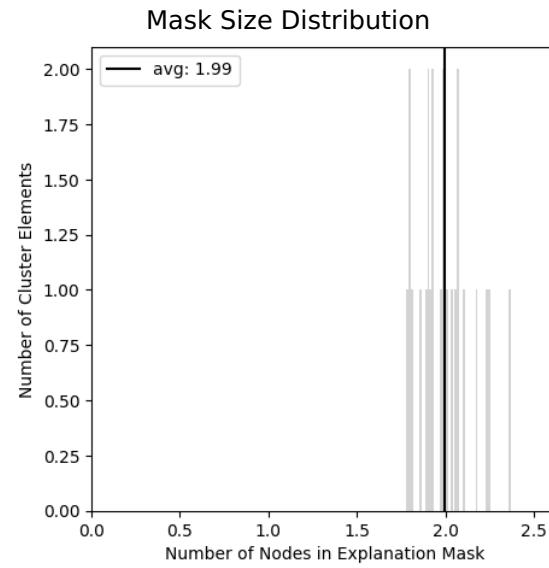
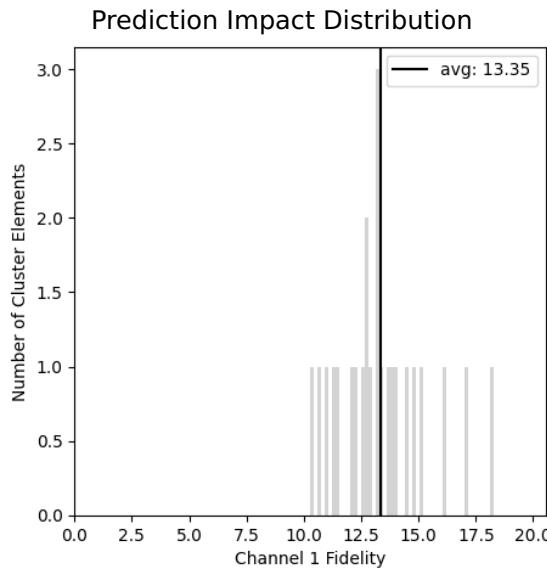
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	1.0 (0.0)

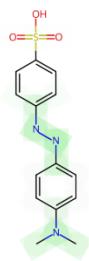
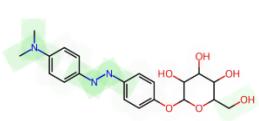
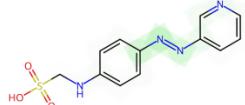
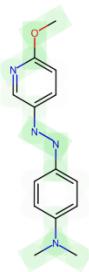
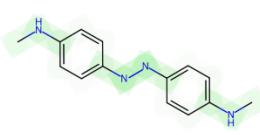
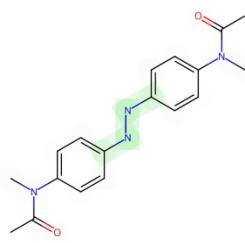
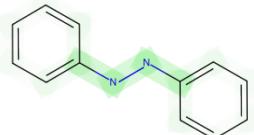
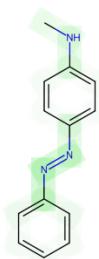
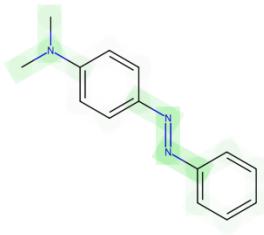
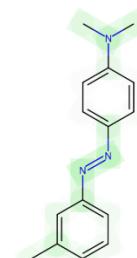
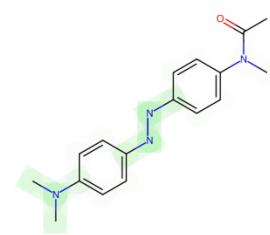
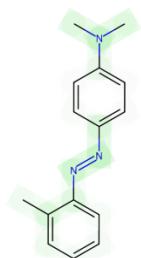
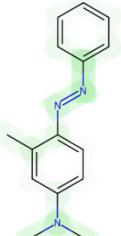
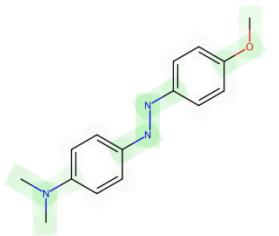
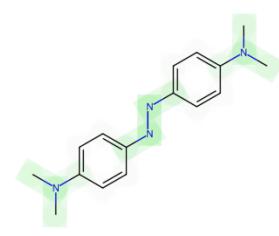
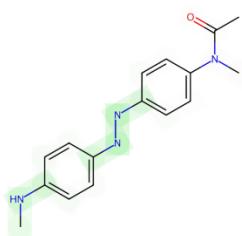
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



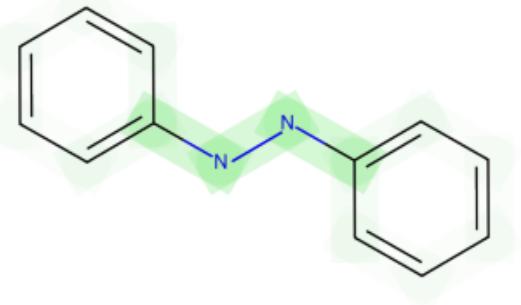
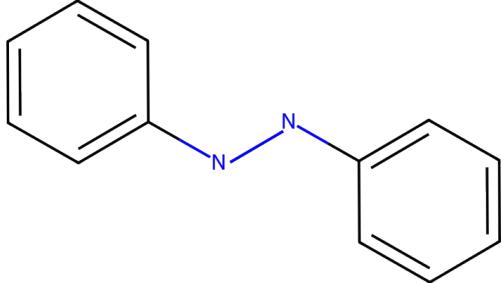
# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES represents a structure that contains an azo group (-N=N-) linked to an aromatic phenyl ring. Azo compounds are known for their mutagenic properties due to their ability to undergo reductive cleavage, producing highly reactive nitrenium ions that can interact with DNA. The phenyl ring, being aromatic, possesses a certain electron density that can stabilize the azo linkage and consequently impact the molecule's reactivity with the genetic material.

**Hypothesis:** Molecules containing the azo-linked phenyl substructure have a medium influence on mutagenicity. This is likely due to the azo group's ability to form reactive species capable of damaging DNA, while the electron-rich phenyl ring influences the stability and reactivity of the mutagenic azo group.

# Cluster #58 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 58, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 15.3 ( $\pm 1.4$ ) on the prediction outcome.

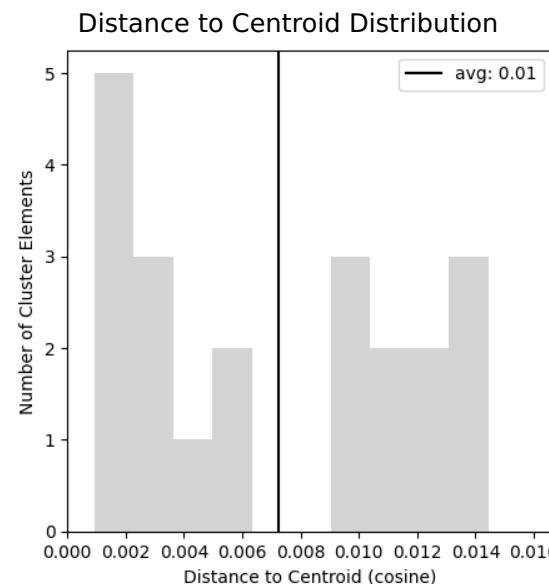
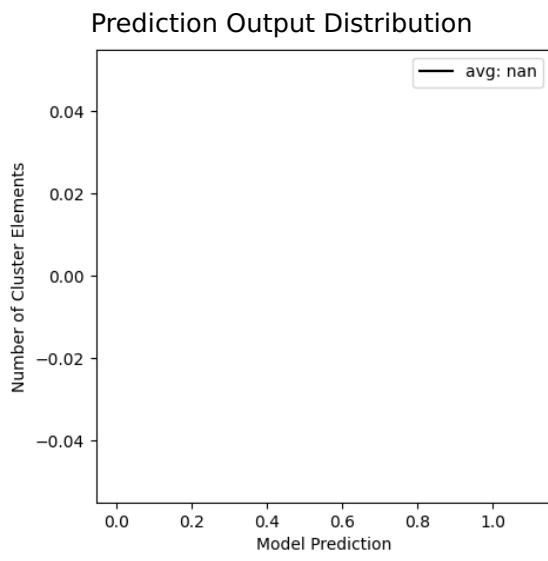
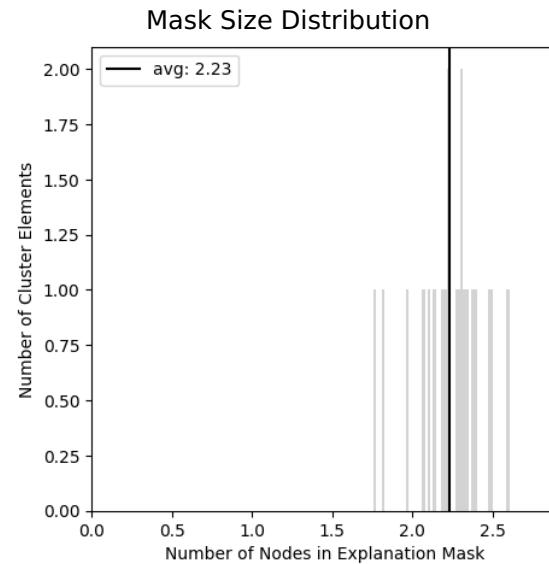
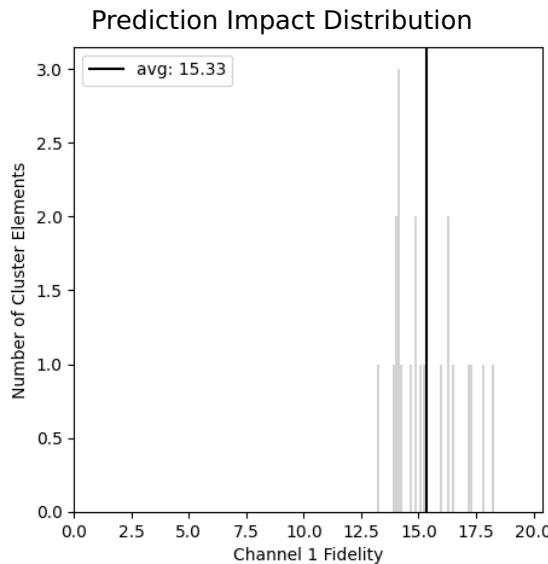
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

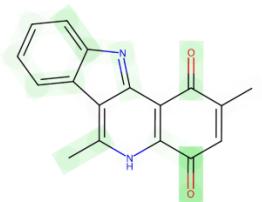
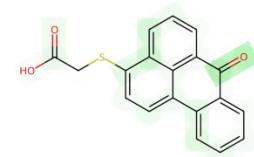
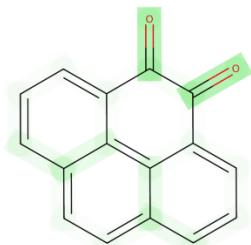
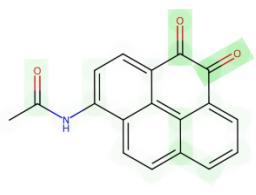
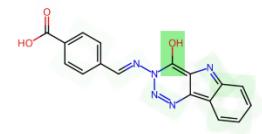
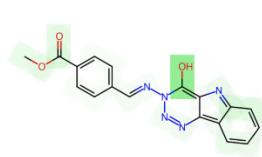
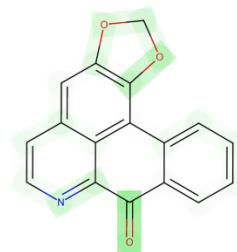
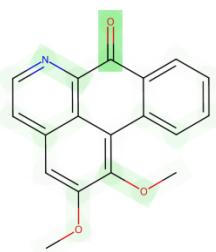
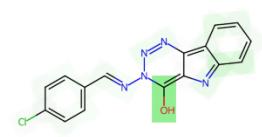
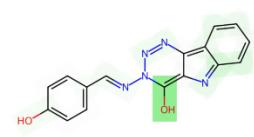
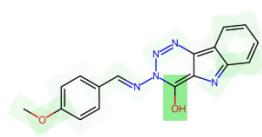
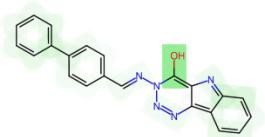
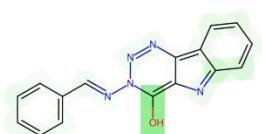
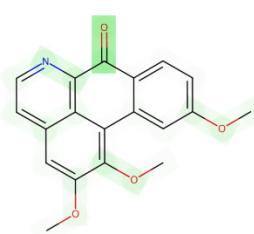
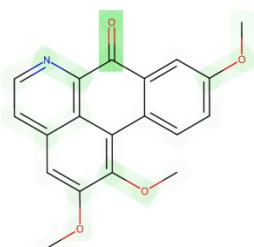
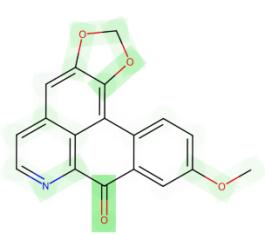
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



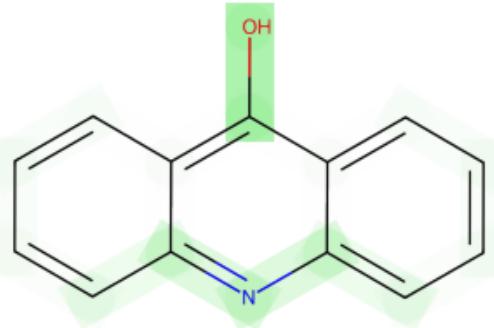
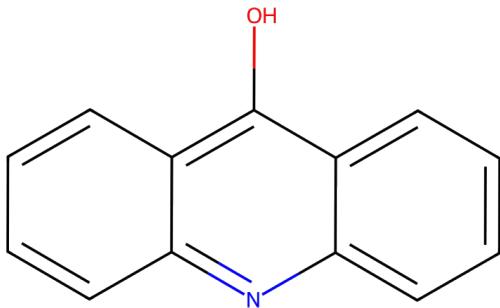
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure provided contains a fused heterocyclic compound, which is a polycyclic compound with at least one heteroatom in the rings, in this case nitrogen and oxygen. These heteroatoms, particularly in aromatic compounds, can participate in electron-rich interactions with DNA nucleobases, leading to potential mutagenic effects. Compounds that can intercalate into DNA or form covalent bonds with it can disrupt normal base-pairing, causing errors in DNA replication and transcription, ultimately leading to mutations.

**Hypothesis:** The presence of a fused heterocyclic substructure with oxygen and nitrogen heteroatoms is associated with a medium level of mutagenicity. This may be due to the electron-rich nature of the heteroatoms and the planar structure of the compound, which allow interactions with DNA that can disrupt its normal function. The medium influence on mutagenicity suggests that while it has the structural features to interact with DNA, other factors such as the size of the molecule, the specific arrangement of the rings, or the presence of other functional groups may moderate its mutagenic potential.

# Cluster #59 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 59, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 12.7 ( $\pm 2.0$ ) on the prediction outcome.

## Properties

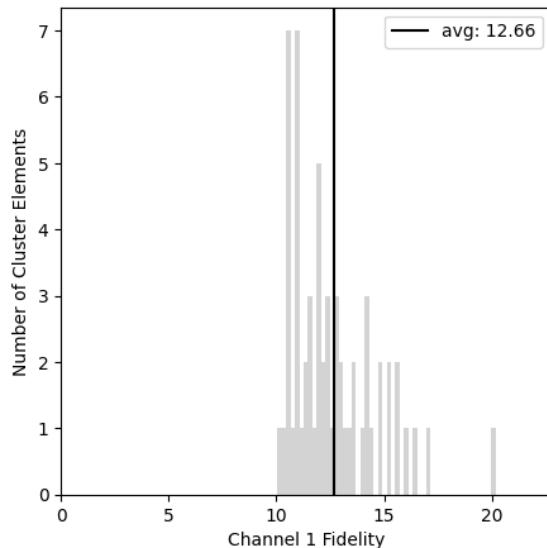
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	59
Channel Index	1.0 (0.0)

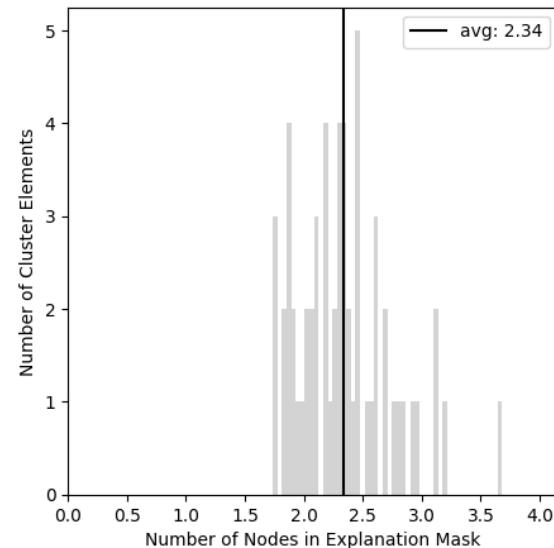
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

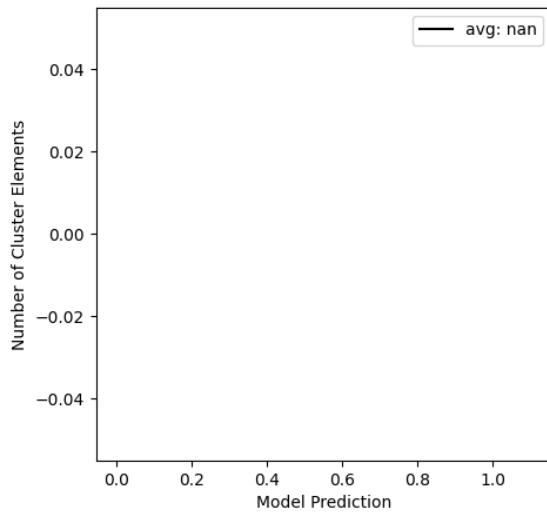
Prediction Impact Distribution



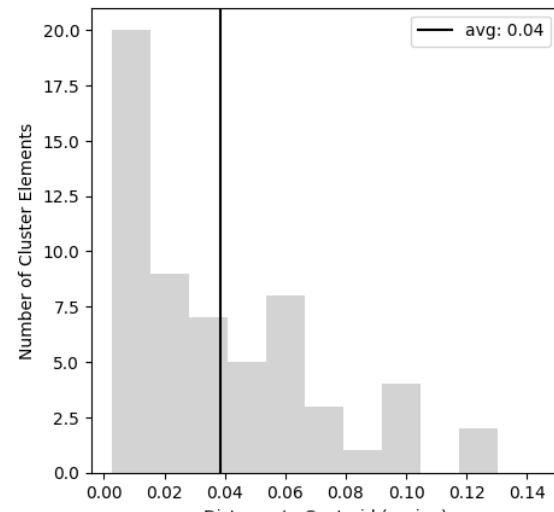
Mask Size Distribution



Prediction Output Distribution

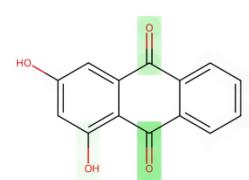
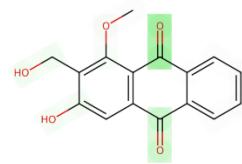
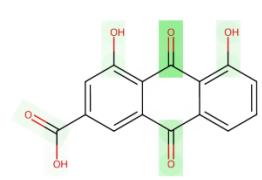
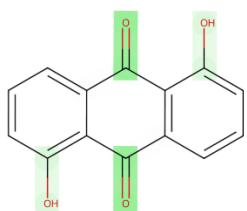
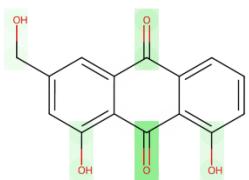
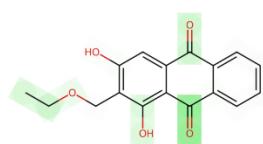
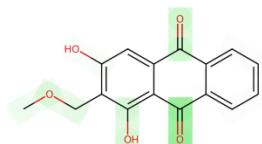
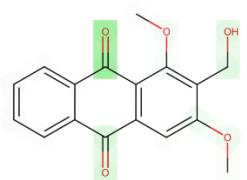
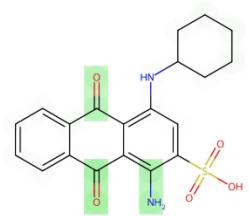
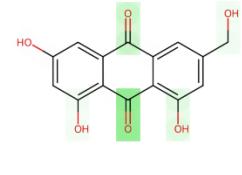
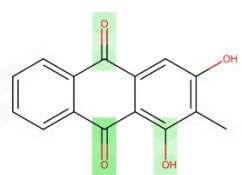
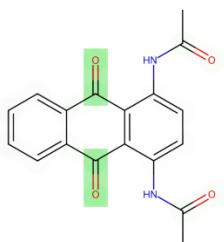
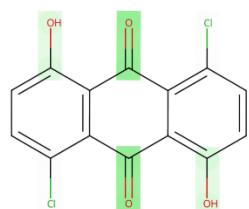
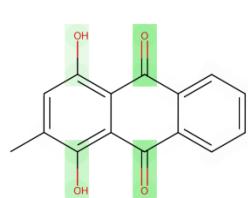
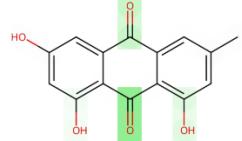
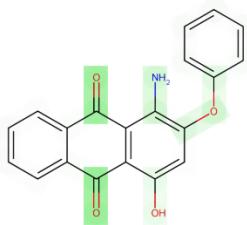


Distance to Centroid Distribution



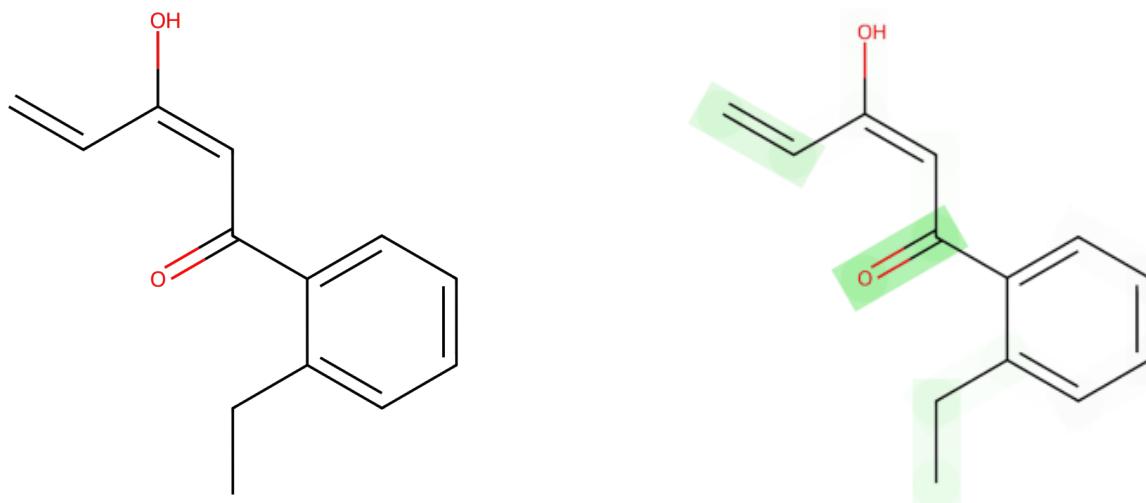
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure contains a styrene-like fragment ( $C=C-C(-O)=C-$ ) combined with a substituted phenyl ring and a  $-C(=O)-$  moiety. The presence of the styrene-like double bond suggests potential reactivity towards nucleophilic centers in genetic material, especially when in the vicinity of an electron-withdrawing group such as a carbonyl ( $-C(=O)-$ ). The conjugated system formed by the double bonds and the aromatic ring could facilitate electron delocalization, potentially forming reactive intermediates. Also, the substituted phenyl ring can interact with DNA base pairs through intercalation or stacking interactions, which may distort the DNA helix and affect DNA replication or repair.

**Hypothesis:** The present structure, with a conjugated system including a styrene-like double bond and a substituted phenyl ring, is linked to medium mutagenicity. The electron-rich areas can form reactive intermediates that might covalently bond with DNA, while the conjugated system's planar nature may facilitate interactions with DNA structures, leading to mutagenic outcomes.

# Cluster #60 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 60, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 13.8 ( $\pm 2.4$ ) on the prediction outcome.

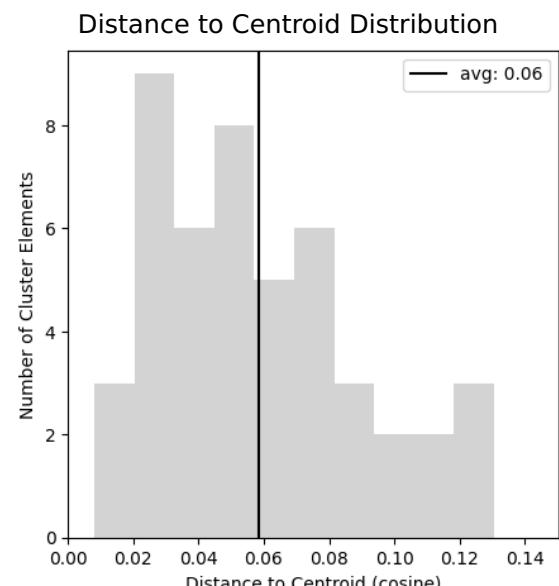
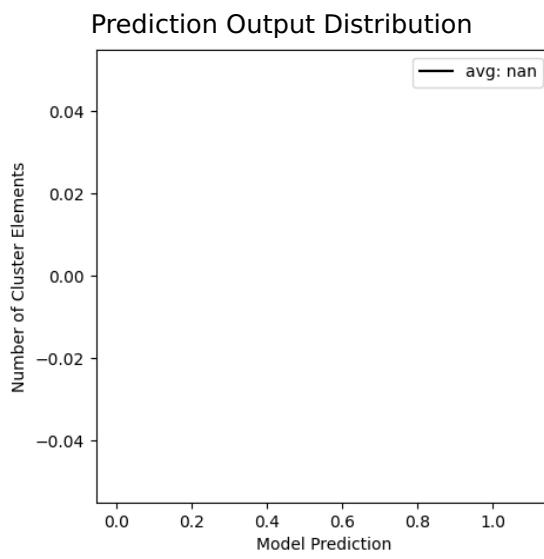
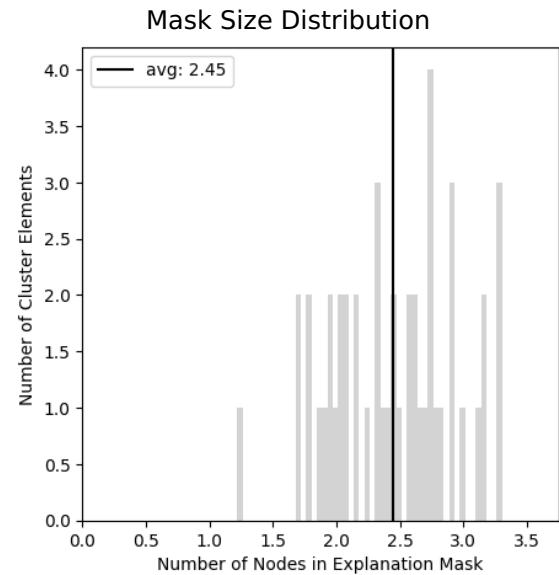
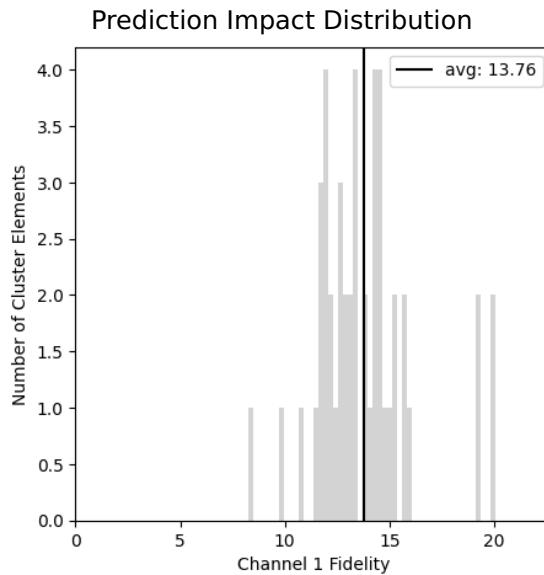
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	47
Channel Index	1.0 (0.0)

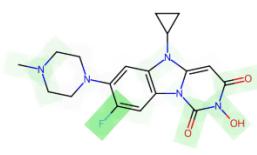
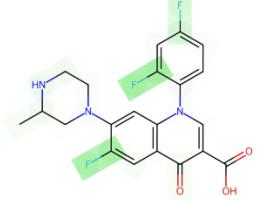
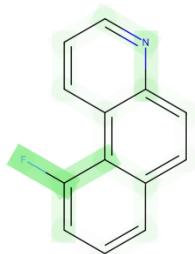
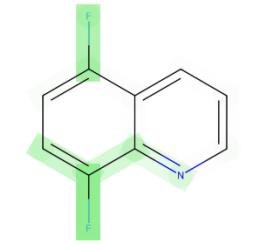
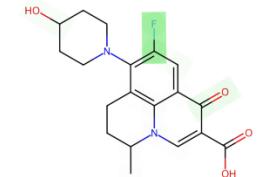
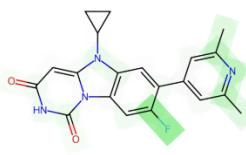
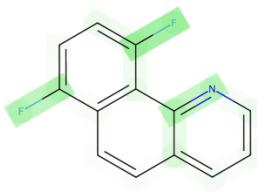
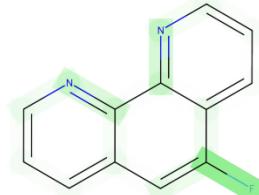
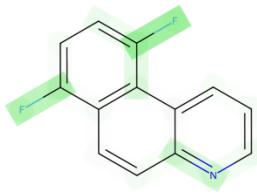
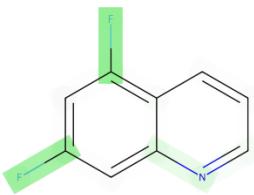
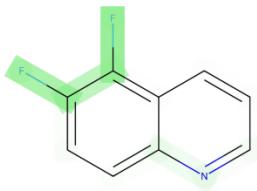
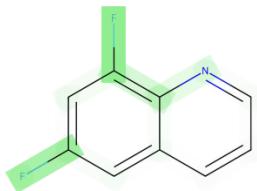
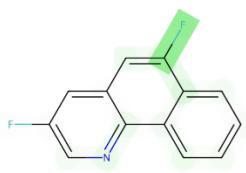
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



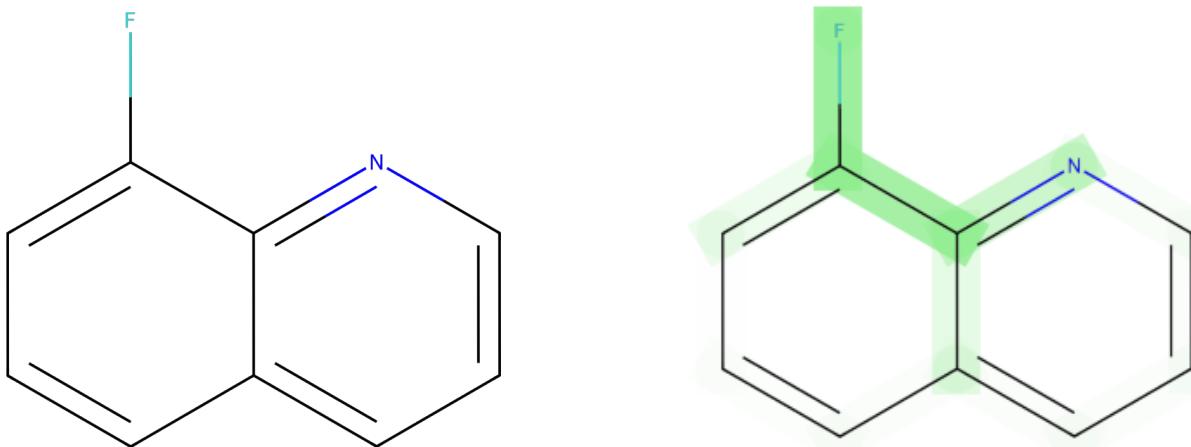
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular fragment represented by "F-c1:c:c:c:c2:c:c:n:c:1:2" appears to be a fluorinated aromatic ring with a fused pyridine-like nitrogen heterocycle. The aromatic system is known for its stability due to resonance which allows electrons to be delocalized over the entire structure, while the nitrogen atom in the ring provides an additional site for electron density and lone pair interactions. Fluorine is a highly electronegative atom that influences the electron distribution within the structure, potentially enhancing the reactivity of the molecule towards nucleophilic attack or electron transfer processes that could lead to DNA damage.

**Hypothesis:** The presence of a fluorinated aromatic ring fused with a nitrogen-containing heterocycle suggests a medium mutagenic potential. The fluorine atom may enhance the electrophilic nature of the molecule, making it more reactive with nucleophiles, including DNA bases. The aromatic system, stabilized by resonance, may be capable of interacting with DNA through intercalation or other non-covalent interactions, further contributing to mutagenicity.

# Cluster #61 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 61, from importance channel 1 (*mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 9.6 ( $\pm 2.4$ ) on the prediction outcome.

## Properties

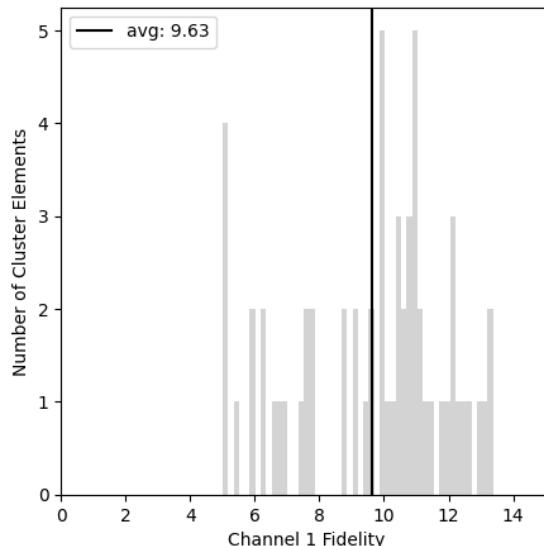
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	60
Channel Index	1.0 (0.0)

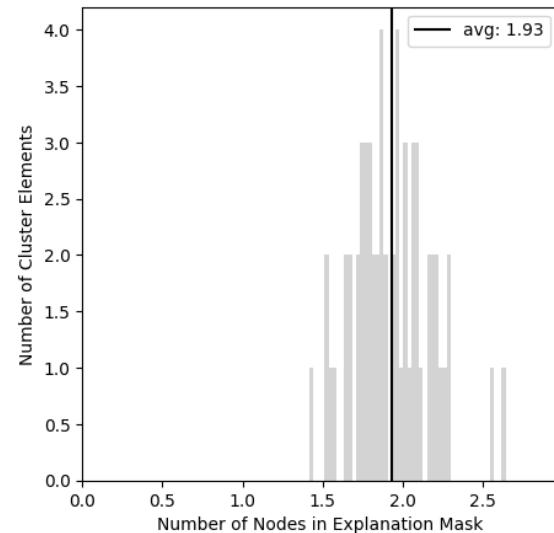
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

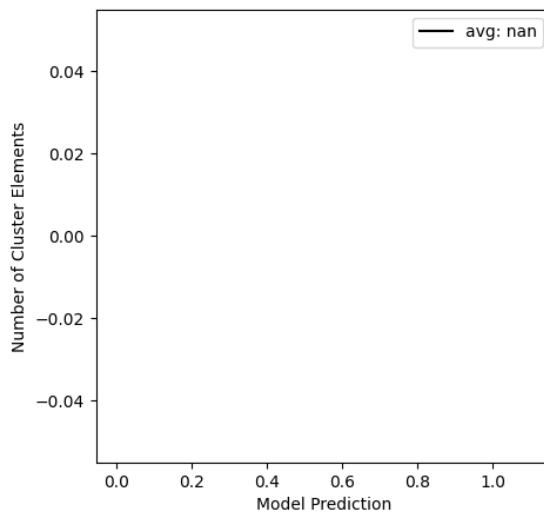
Prediction Impact Distribution



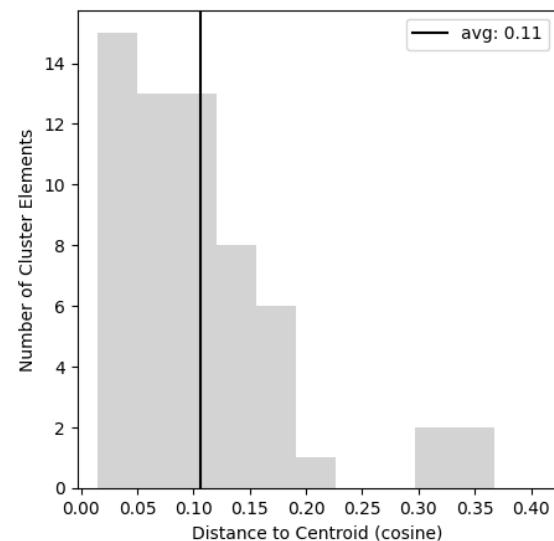
Mask Size Distribution



Prediction Output Distribution

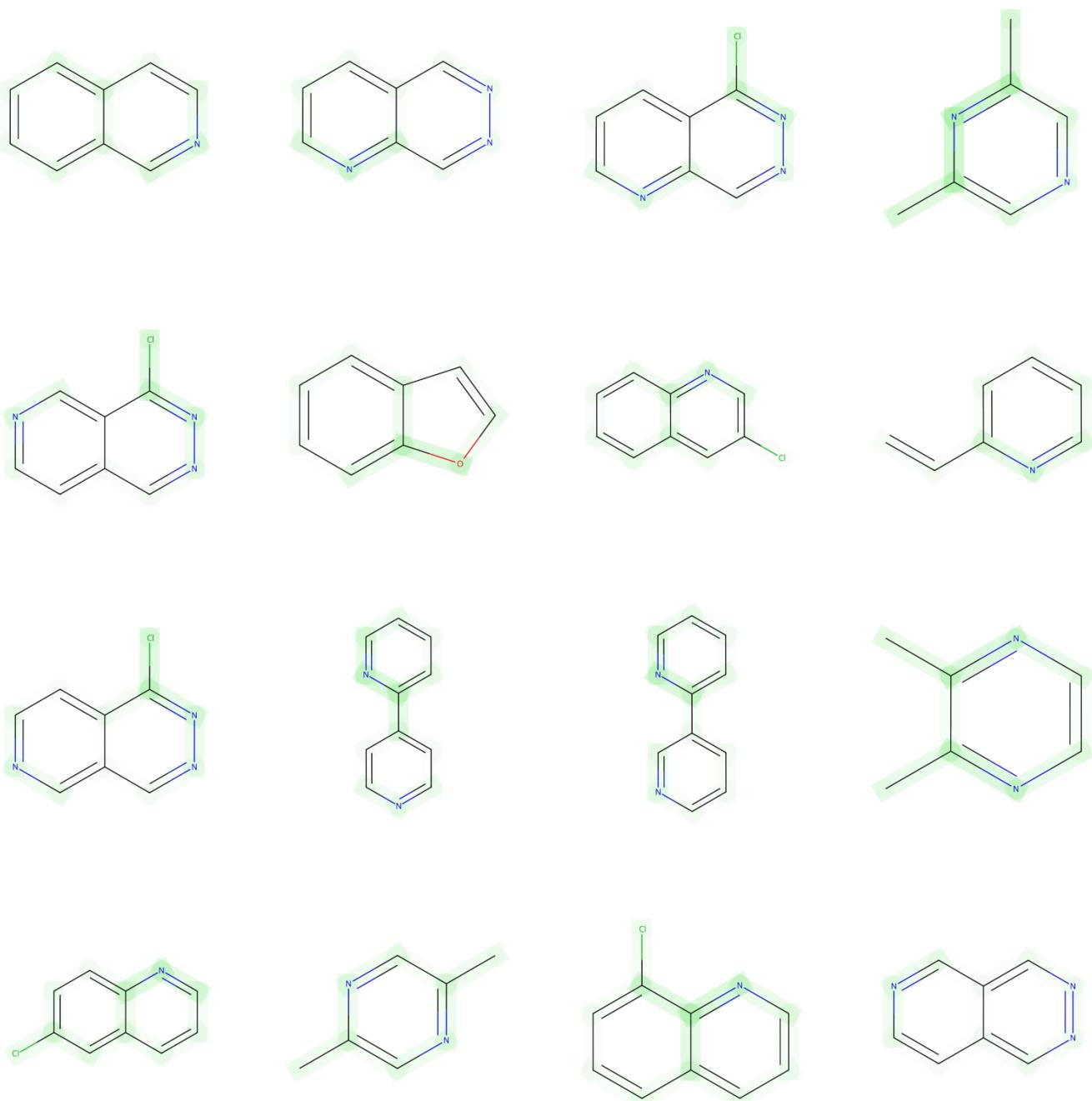


Distance to Centroid Distribution



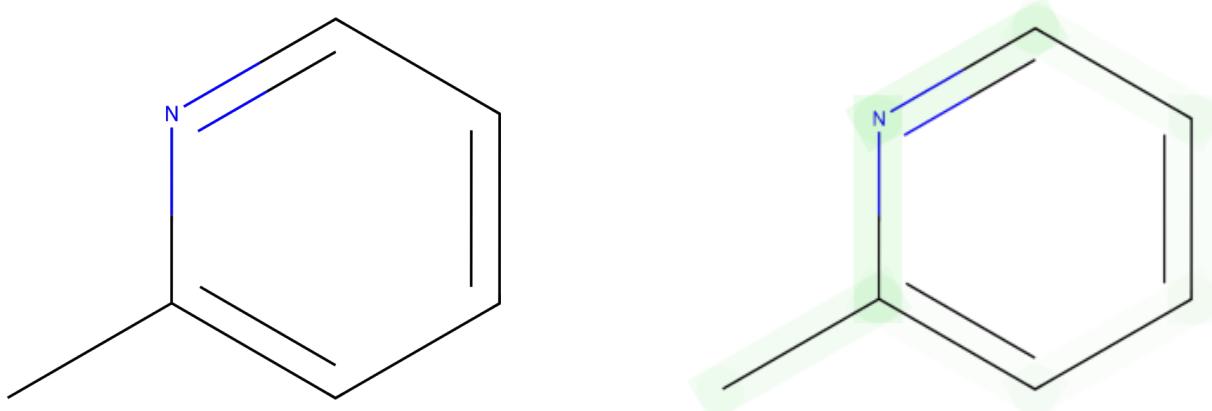
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents an aromatic compound with a heteroatom - nitrogen (indicated as 'n' in the SMILES) incorporated into an otherwise carbon-based ring. Aromatic compounds are known for their stability due to delocalized electrons within the ring. The presence of the nitrogen atom, however, introduces an electron lone pair into the electron system of the ring, which could potentially interact with DNA structures if the molecule were to intercalate or bind with DNA. However, the term "SMALL influence towards mutagenic" suggests that while there may be potential for interactions with genetic material, it is not particularly strong in this instance.

**Hypothesis:** The SMILES representation "C-c1:c:c:c:c:n:1" corresponds to a stable aromatic molecule with a single nitrogen heteroatom, indicating a small mutagenic influence. It is hypothesized that the aromatic stability reduces reactive tendencies, while the lone pair on the nitrogen has the potential to interact with DNA, but the extent of this interaction is limited, resulting in a low mutagenic potential.

# Cluster #62 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 62, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 8.6 ( $\pm 1.7$ ) on the prediction outcome.

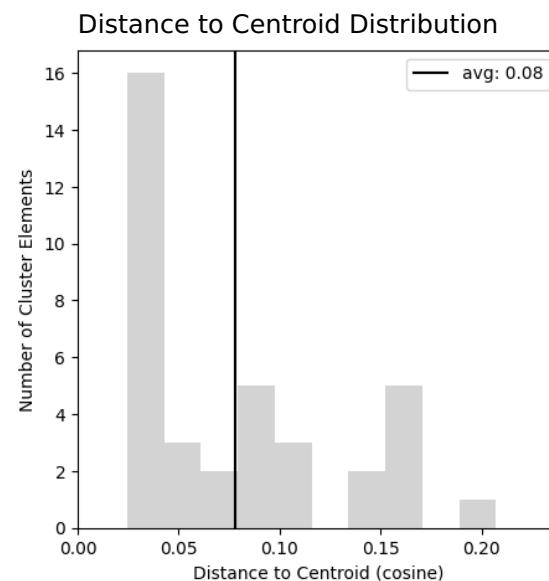
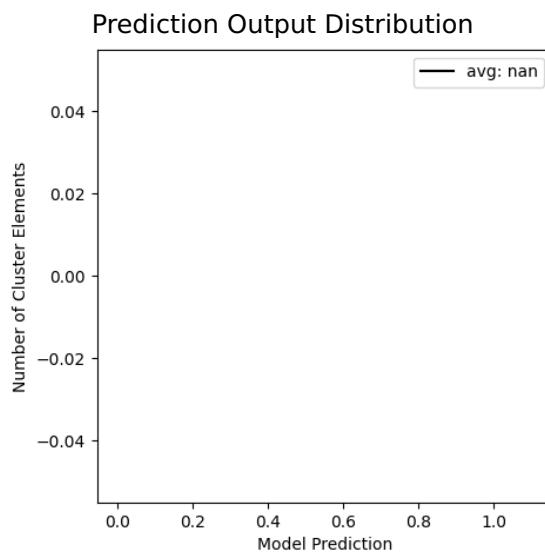
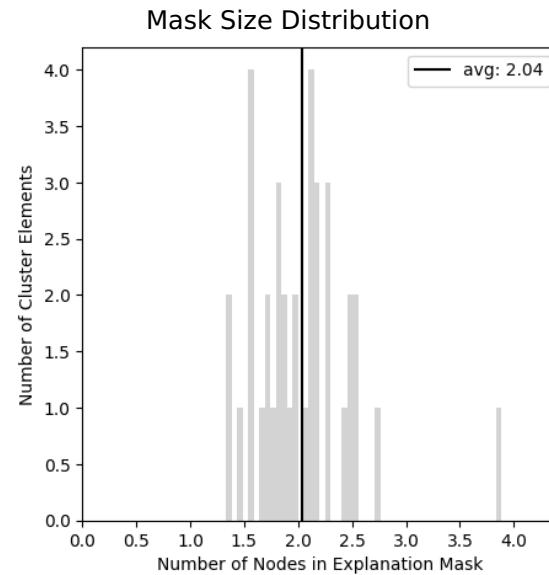
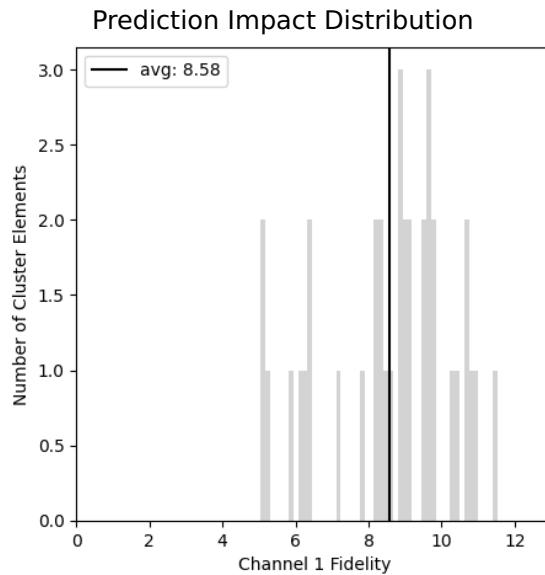
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	37
Channel Index	1.0 (0.0)

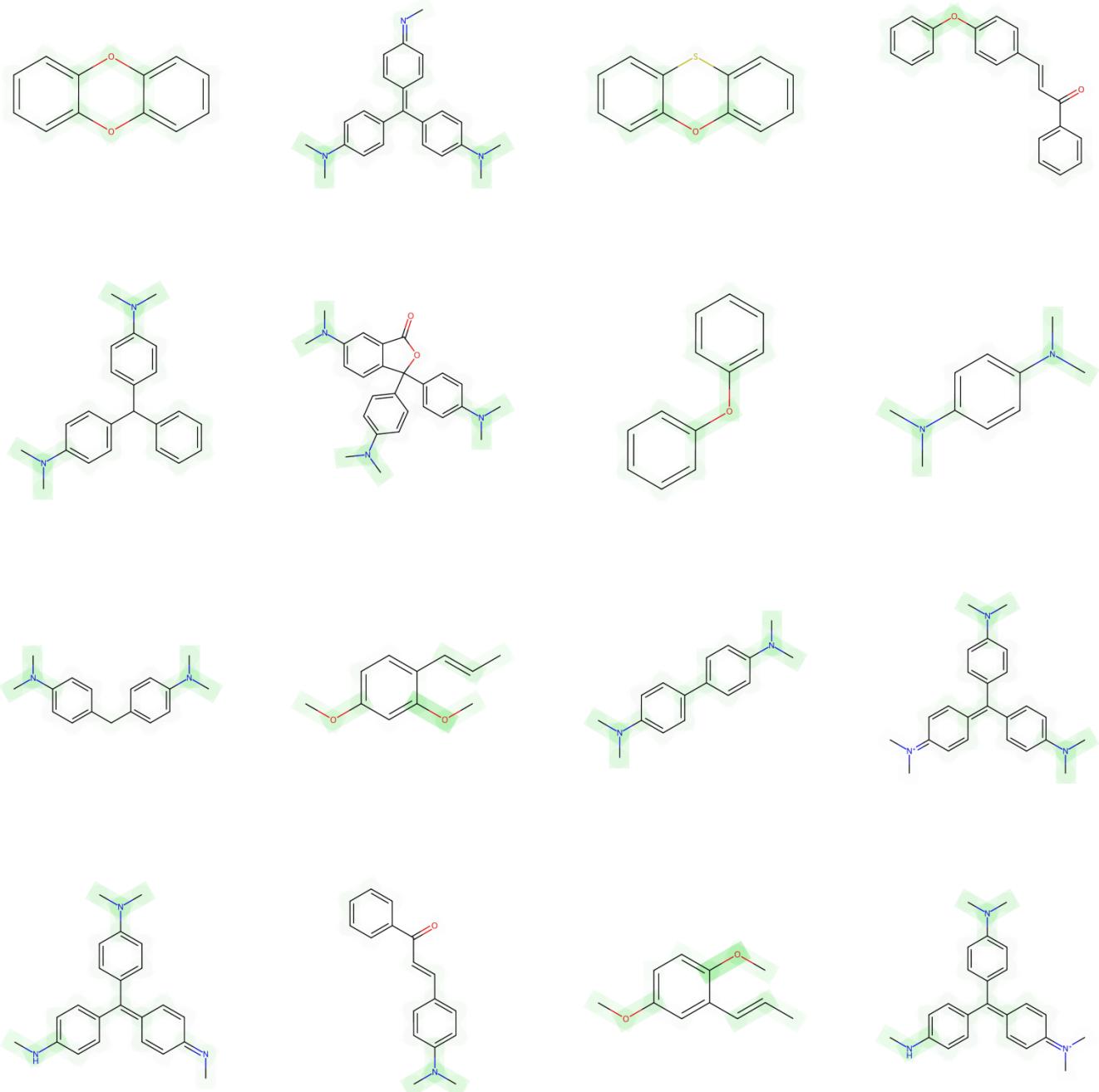
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



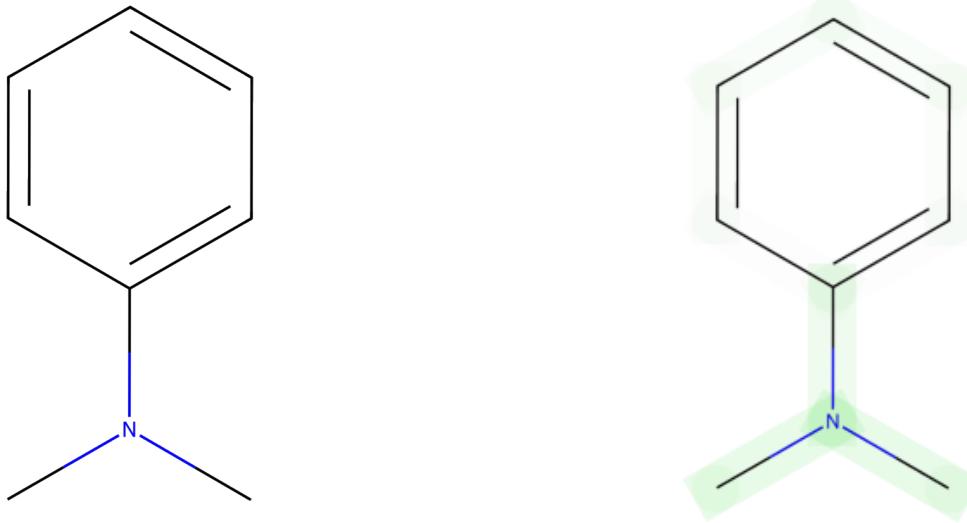
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure "C-N(-C)-c1:c:c:c:c:c:1" represents a structure with an aromatic ring connected to an amine (NH) group via a methylene (-CH<sub>2</sub>-) bridge. Aromatic rings are known for their stability due to resonance, but also for their ability to intercalate or stack between the base pairs in DNA, potentially causing physical disruption of the helical structure. The amine group can be a site for bioactivation, potentially forming reactive metabolites that can covalently bind to DNA. However, the linking methylene group may offer enough separation to reduce the likelihood of such interactions.

**Hypothesis:** The aromatic amine substructure poses a small mutagenic risk. Its aromatic ring could enable interactions with DNA, but the steric hindrance provided by the methylene group reduces its ability to come in close enough proximity to the DNA to cause significant damage. The amine group could undergo bioactivation to a less safe form, yet the overall structure does not greatly favor such a transformation.

# Cluster #63 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 63, from importance channel 1 (*mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 9.3 ( $\pm 2.4$ ) on the prediction outcome.

## Properties

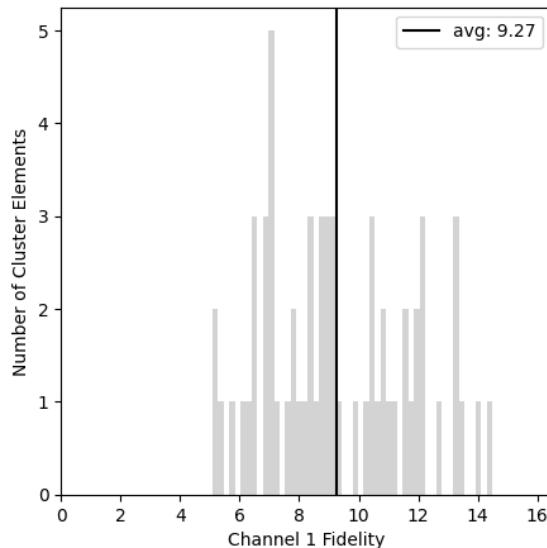
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	62
Channel Index	1.0 (0.0)

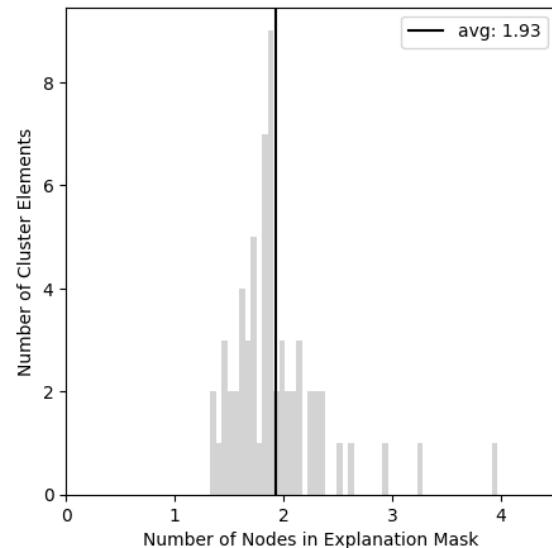
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

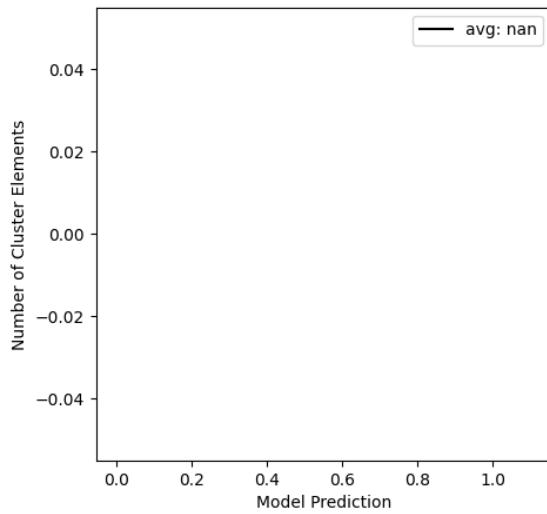
Prediction Impact Distribution



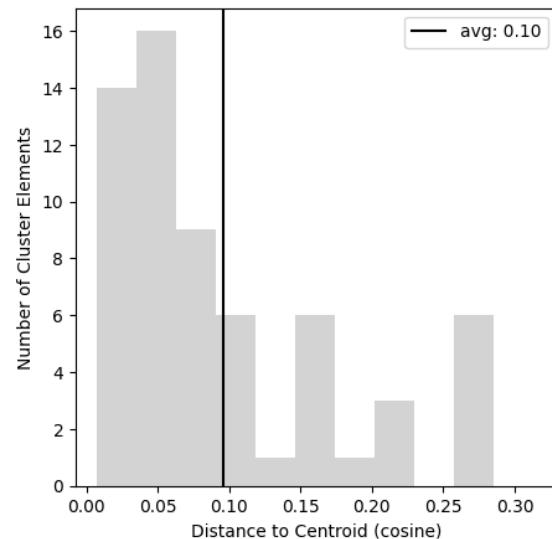
Mask Size Distribution



Prediction Output Distribution

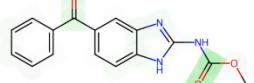
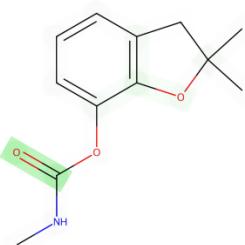
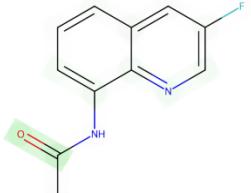
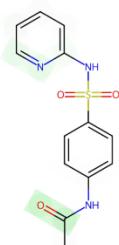
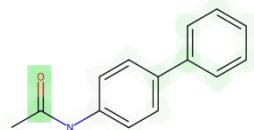
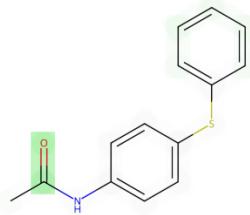
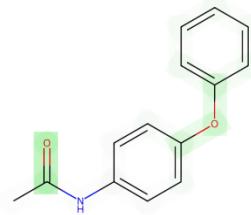
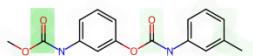
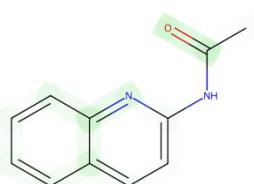
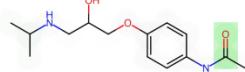
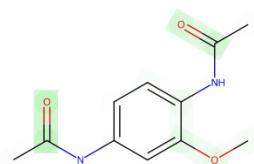
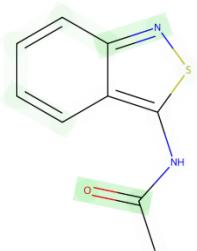
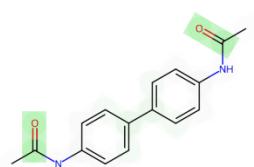
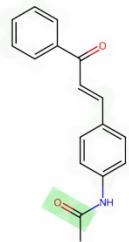
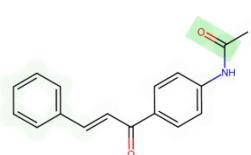
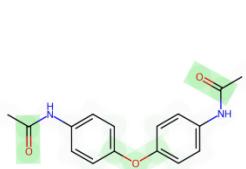


Distance to Centroid Distribution



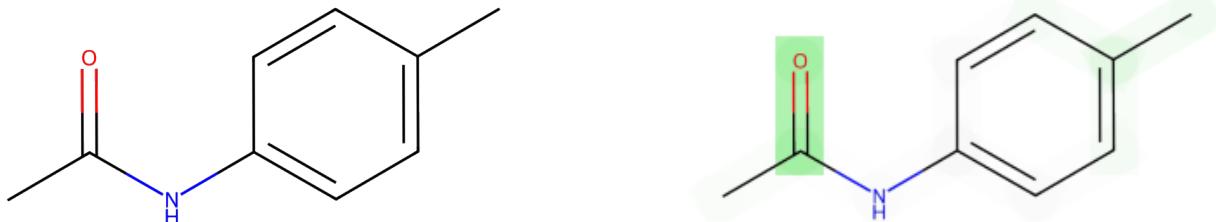
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented here is a benzene ring linked to an acetyl group with a nitrogen-containing substituent. This connection suggests an ability to form electrophilic metabolites that can interact with DNA. The presence of the acetyl group ( $\text{C}=\text{O}$ ) could potentially be transformed into an electrophilic species capable of covalent binding to DNA, leading to mutations. However, the descriptor "SMALL" influence indicates that although the structure has mutagenic components, its overall effect may be moderated by other factors such as steric hindrance or the presence of electron-donating groups that reduce its electrophilic potential.

**Hypothesis:** The molecular substructure in question has a small mutagenic potential, likely because the acetyl group can form reactive electrophiles, but the moderating effects from the rest of the molecule reduce this potential. Possible electron donation from the alkyl group on the benzene ring or steric factors may limit the accessibility of this structure to DNA, thus diminishing its mutagenicity.

# Cluster #64 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 64, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 13.6 ( $\pm 2.3$ ) on the prediction outcome.

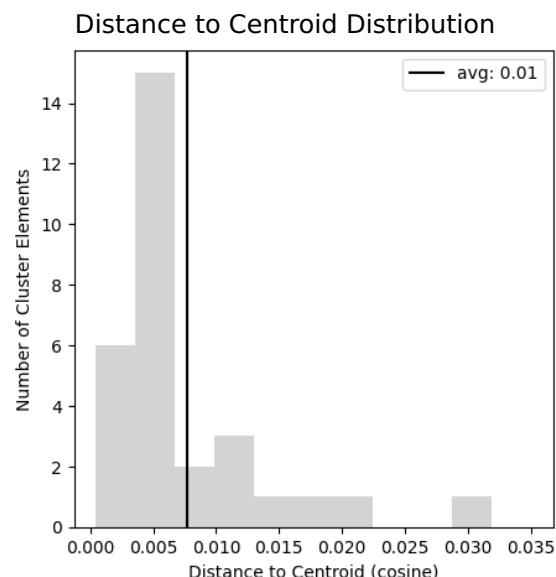
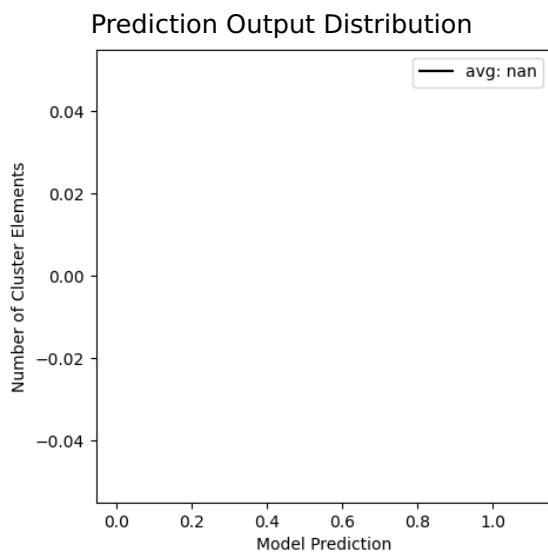
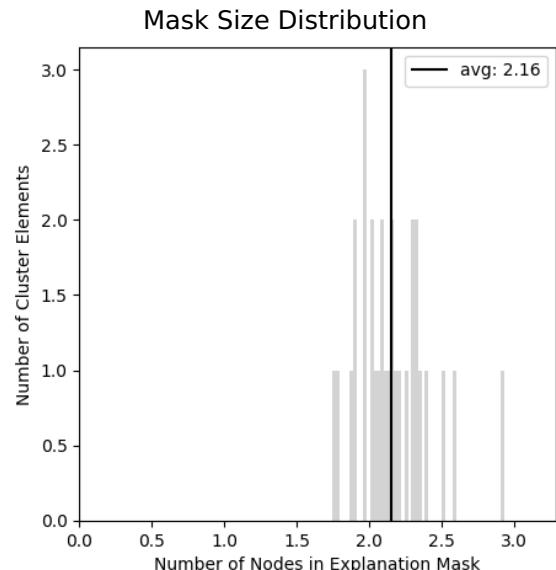
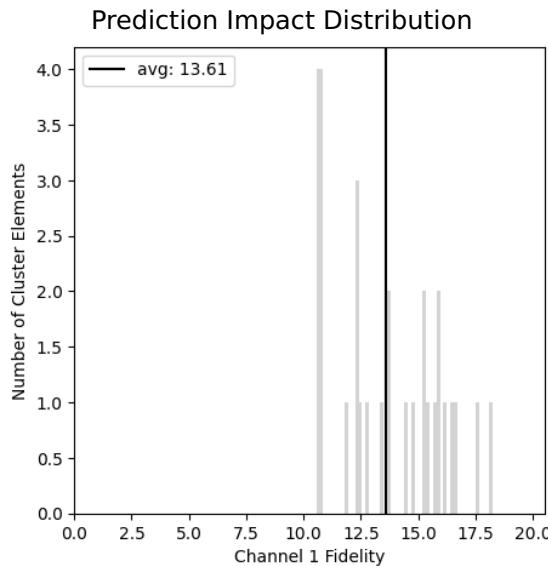
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	30
Channel Index	1.0 (0.0)

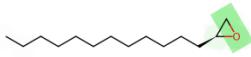
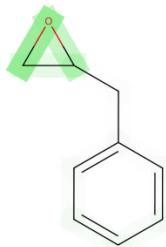
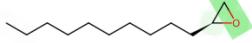
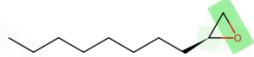
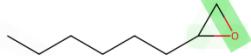
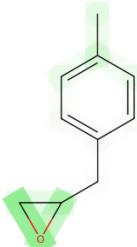
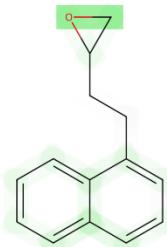
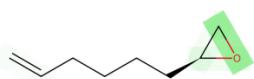
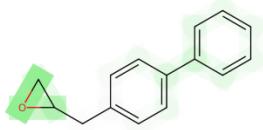
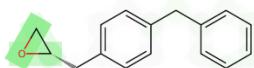
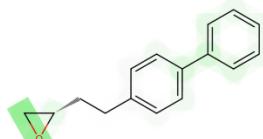
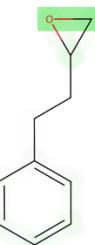
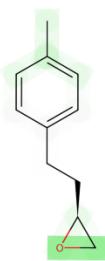
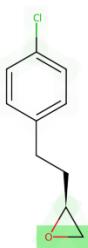
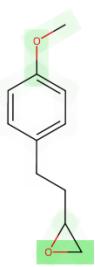
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



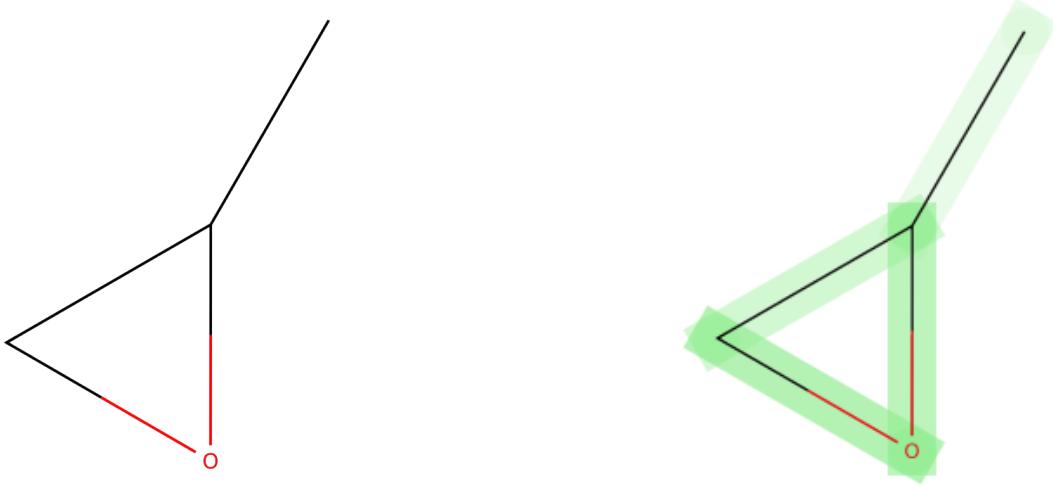
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure shown by "C-C1-C-O-1" represents a cycloalkane with a single oxygen atom, indicating the presence of an epoxide group. Epoxides are known to be reactive due to the strained three-membered ring which, upon exposure to nucleophilic sites in biological macromolecules like DNA, can lead to the formation of covalent adducts. These adducts can cause errors in replication and transcription processes, leading to mutations. Additionally, the reactivity of epoxides is influenced by the surrounding molecular context, which includes other adjacent functional groups and overall molecular size and shape, potentially affecting accessibility to the DNA.

**Hypothesis:** Molecules containing the "C-C1-C-O-1" substructure have a medium influence on mutagenicity due to the presence of the reactive epoxide group. This strained three-membered ring is capable of interacting with nucleophilic sites in DNA, promoting the formation of mutations, which is moderated by other structural factors that influence the molecule's reactivity and DNA accessibility.

# Cluster #65 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 65, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.6 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 16.2 ( $\pm 2.3$ ) on the prediction outcome.

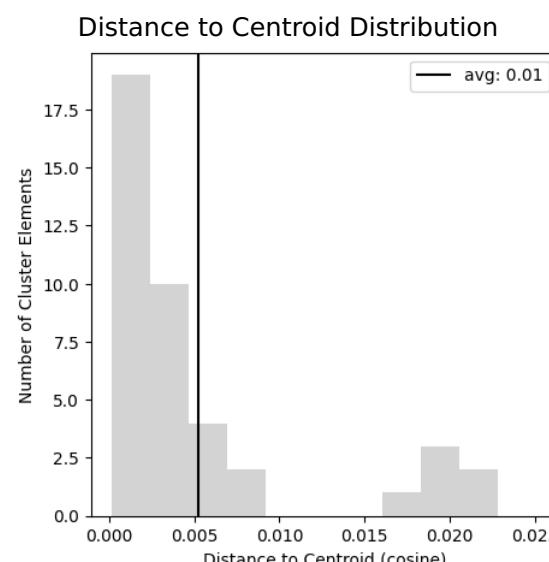
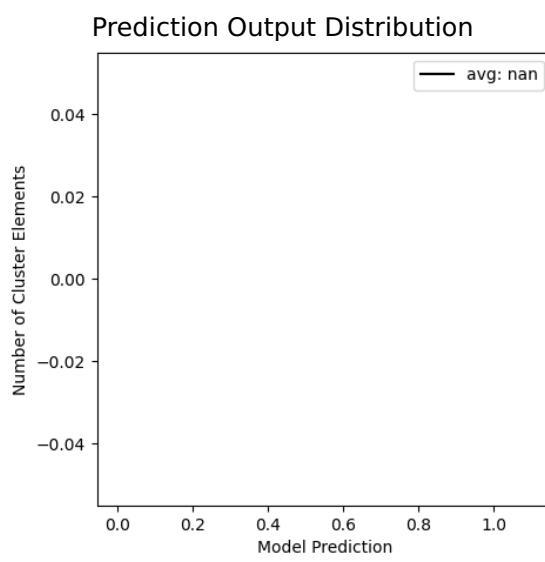
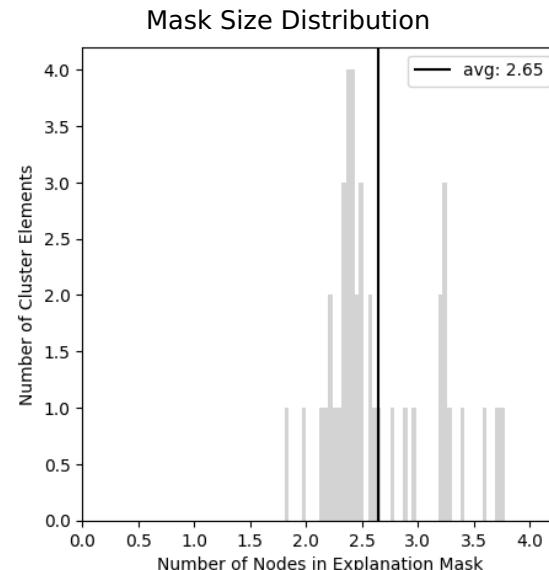
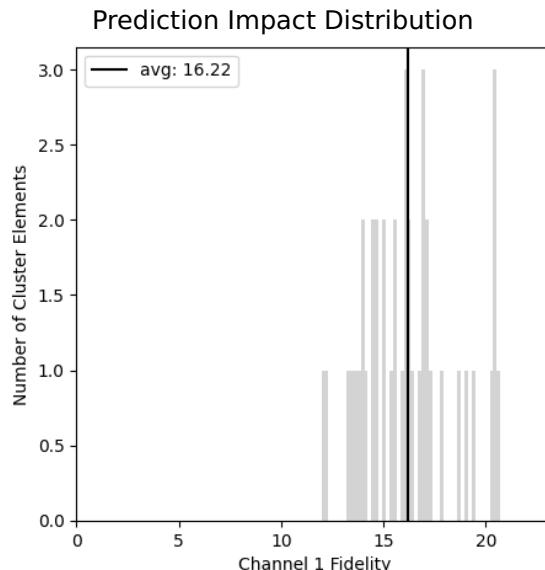
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	41
Channel Index	1.0 (0.0)

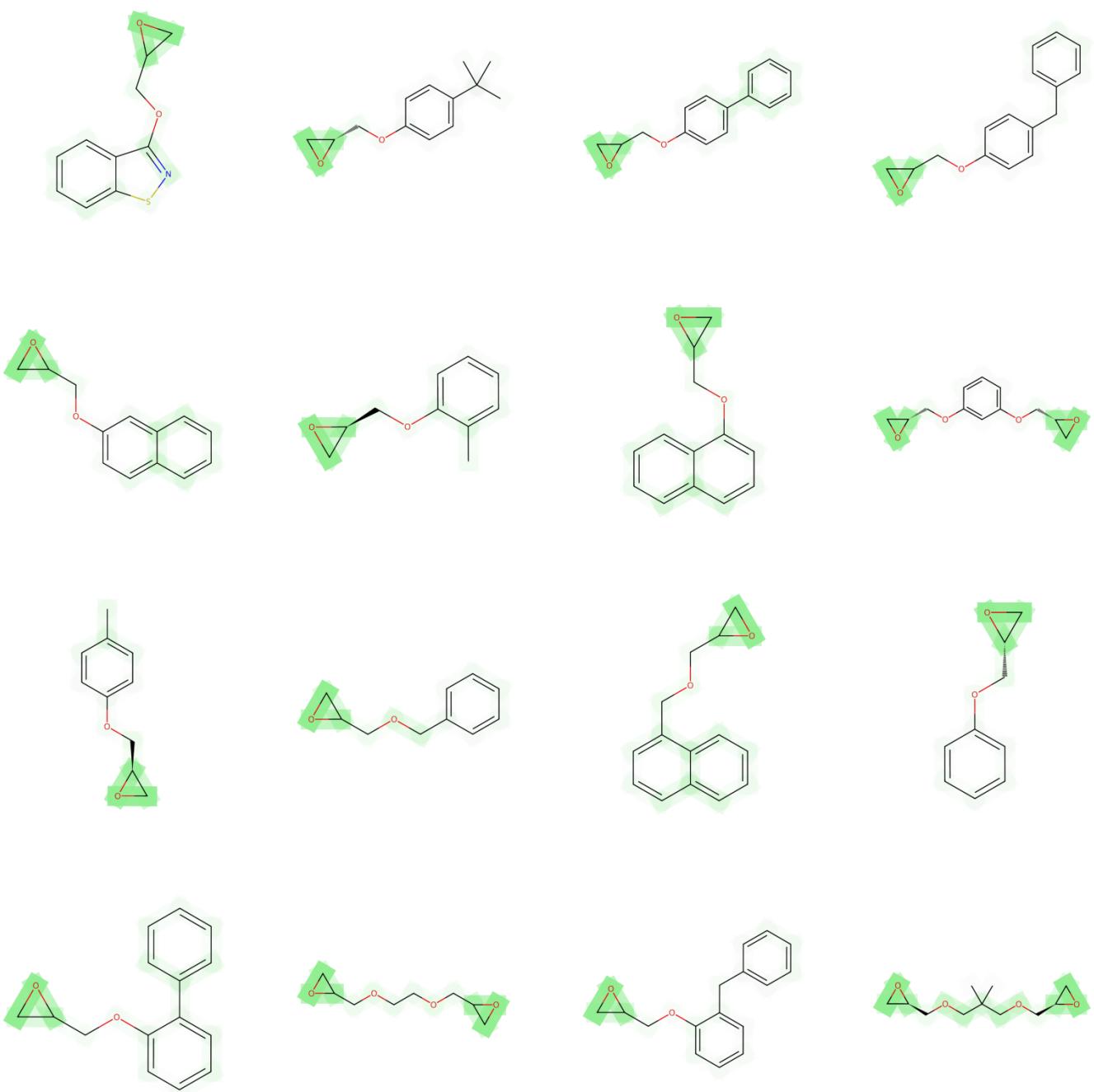
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



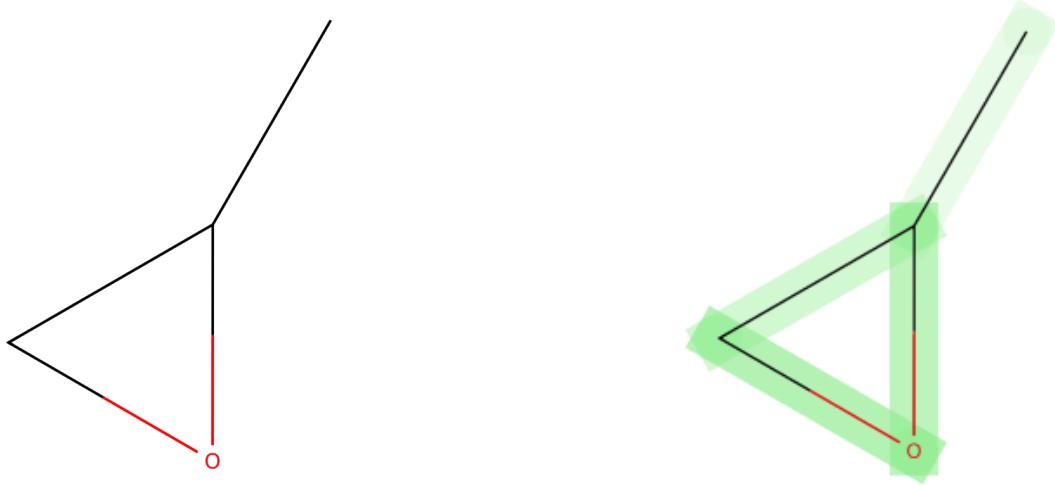
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation "C-C1-C-O-1" suggests a cyclic structure containing a three-membered ring with an oxygen atom (an epoxide). Epoxides are known to be highly reactive due to the strained ring structure, which imparts significant ring strain and thus leads to the reactivity of the epoxide group. The epoxide can react with DNA nucleophiles, such as the amine groups in DNA bases, leading to mutations as it forms covalent bonds with the genetic material, potentially causing frameshift mutations or point mutations.

**Hypothesis:** Molecules containing the substructure "C-C1-C-O-1" potentially possess medium mutagenicity due to the reactive nature of the epoxide group. The inherent ring strain of the three-membered epoxide ring enhances its electrophilic character, making it prone to attack nucleophilic sites in DNA, causing mutagenic events.

# Cluster #66 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 66, from importance channel 1 (*mutagenic*), represents a motif consisting of 3.2 ( $\pm 0.7$ ) nodes. The concept is generally associated with an impact of 17.1 ( $\pm 1.9$ ) on the prediction outcome.

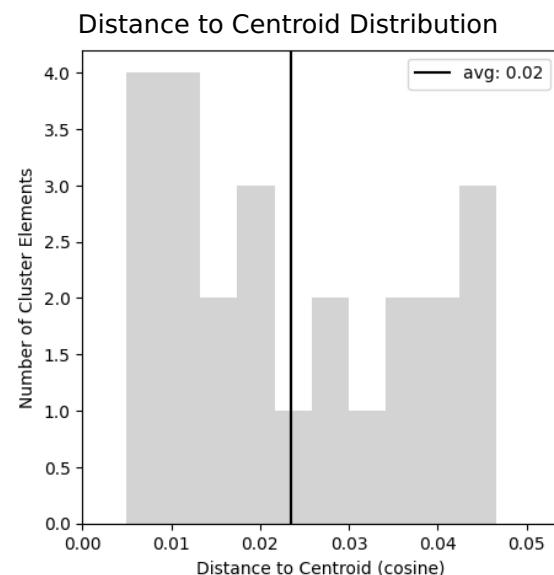
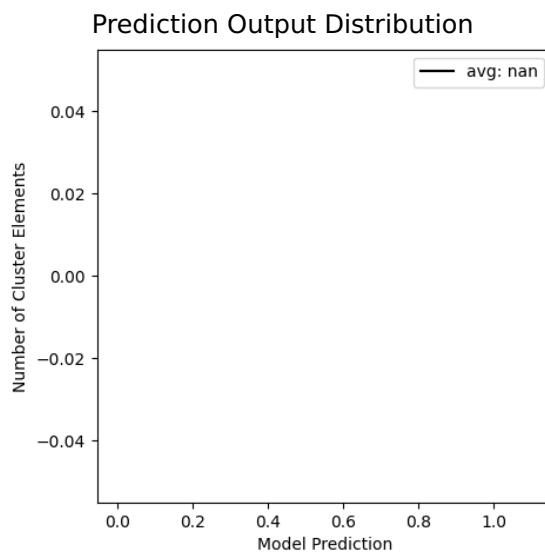
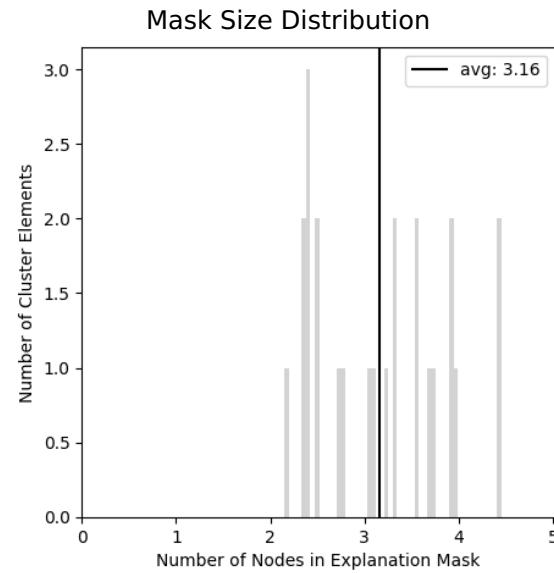
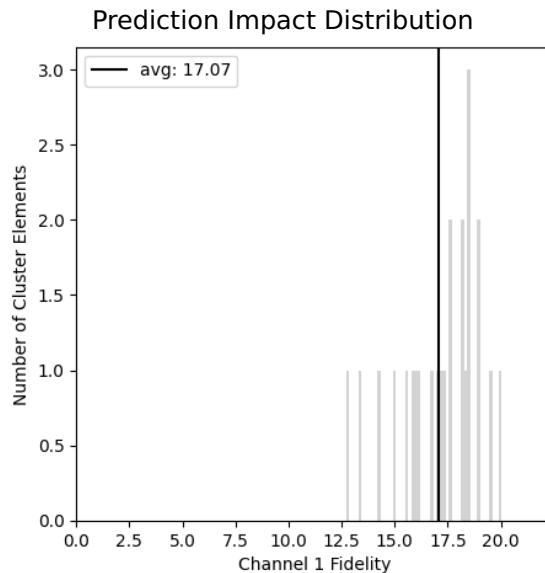
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	1.0 (0.0)

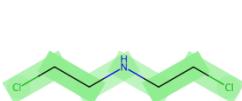
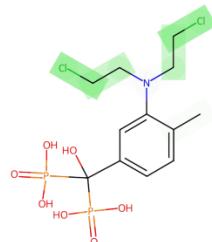
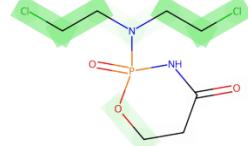
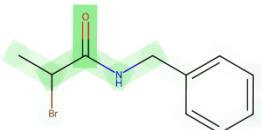
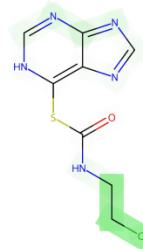
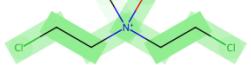
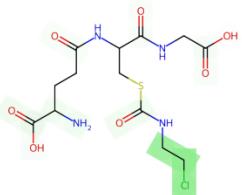
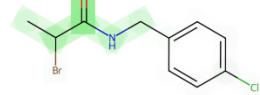
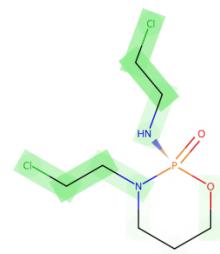
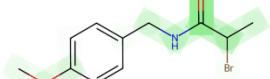
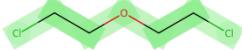
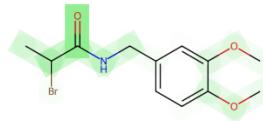
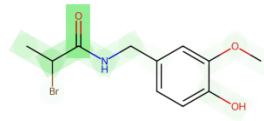
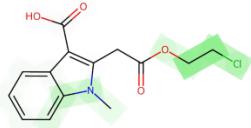
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given molecular substructure contains an alkyl chloride group, known for forming electrophilic species capable of binding to DNA, often leading to deleterious genetic mutations. The electron-withdrawing effect of the chlorine atom in the substructure can make the adjacent carbon more susceptible to nucleophilic attack, facilitating the formation of a covalent bond between the alkyl chloride and the genetic material. Additionally, the presence of oxygen atoms suggests potential for forming reactive oxygen species, further contributing to the mutagenic potential.

**Hypothesis:** When a molecule contains the substructure C-O-C-C-Cl, it has a medium level of mutagenicity. The chlorine atom provides a site for potential alkylation of DNA, while the oxygen atoms may contribute to oxidative stress, both of which increase the likelihood of genetic mutations.

# Cluster #67 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 67, from importance channel 1 (*mutagenic*), represents a motif consisting of 4.2 ( $\pm 1.2$ ) nodes. The concept is generally associated with an impact of 16.3 ( $\pm 1.0$ ) on the prediction outcome.

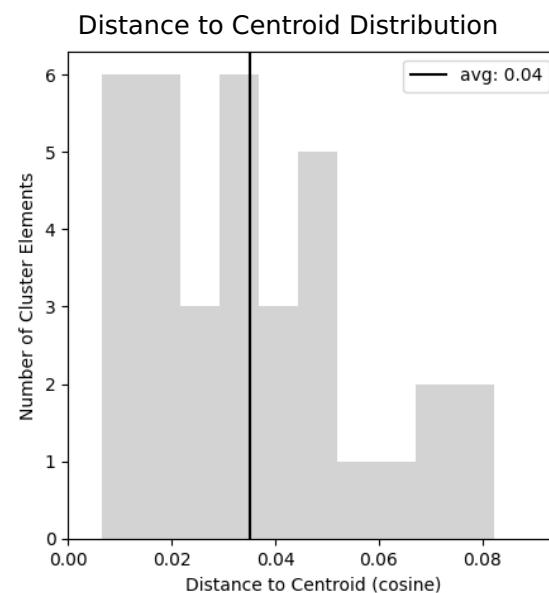
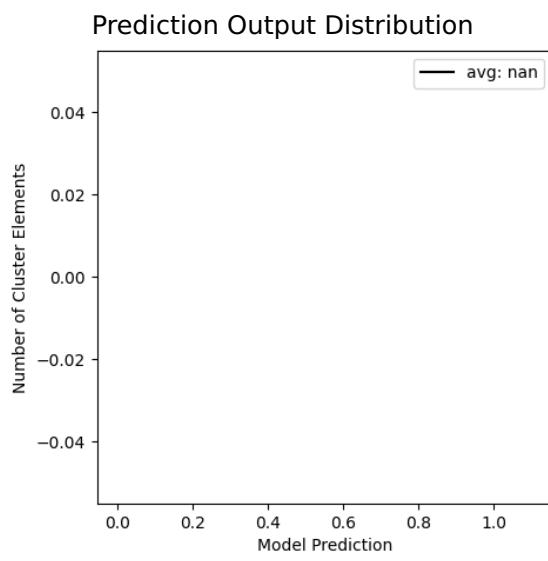
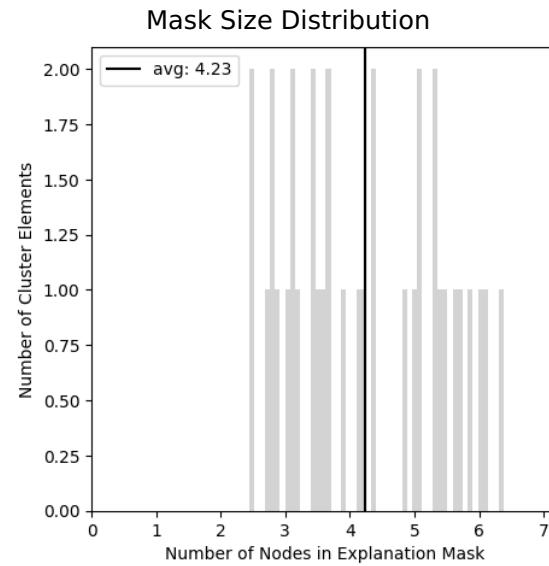
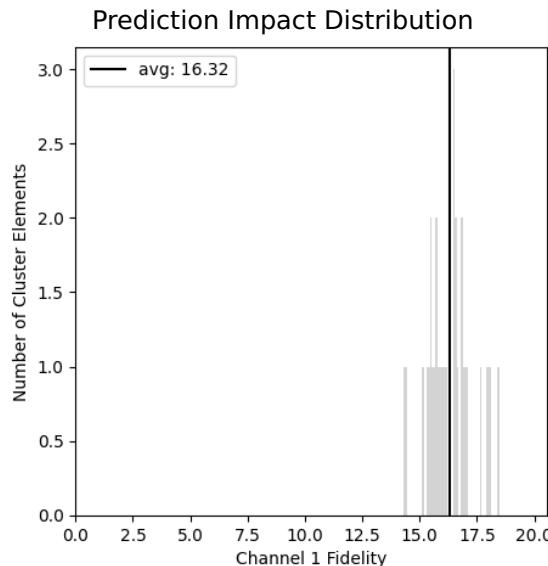
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	35
Channel Index	1.0 (0.0)

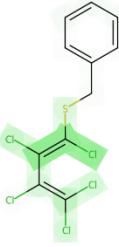
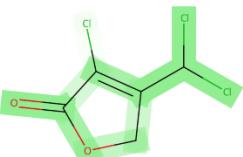
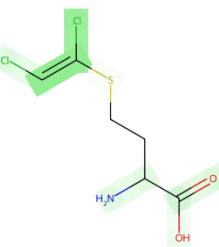
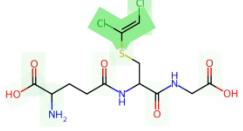
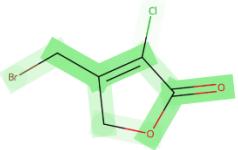
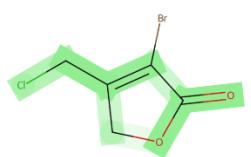
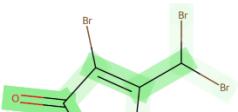
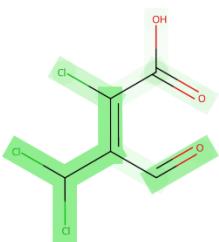
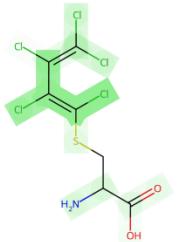
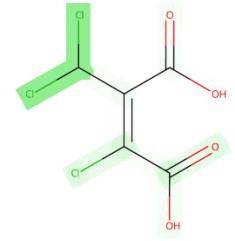
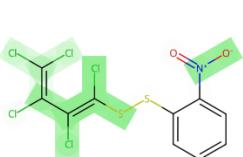
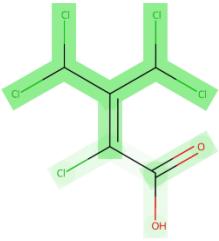
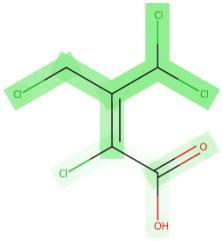
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



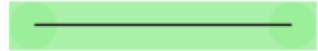
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The "C-C" substructure represents a simple alkane linkage, where two carbon atoms are single-bonded. This type of bond is typically stable and non-reactive under normal physiological conditions. However, the presence of additional functional groups attached to this basic carbon chain can significantly alter its reactivity and interaction with DNA. The medium mutagenic influence suggests that when present in certain molecular contexts, the alkane structure might facilitate a molecule's access to DNA, potentially through intercalation or by inducing a conformational change in the DNA structure itself. Factors such as the length of the carbon chain, branching, and the presence of nearby electron-donating or -withdrawing groups can modulate this effect.

**Hypothesis:** The simple alkane linkage "C-C" exhibits a medium influence on mutagenicity. This influence is likely contingent on the broader molecular context in which the alkane structure appears, emphasizing the role of additional functional groups and the overall 3D conformation of the molecule in determining its interaction with DNA.

# Cluster #68 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 68, from importance channel 1 (*mutagenic*), represents a motif consisting of 3.5 ( $\pm 1.2$ ) nodes. The concept is generally associated with an impact of 17.6 ( $\pm 1.1$ ) on the prediction outcome.

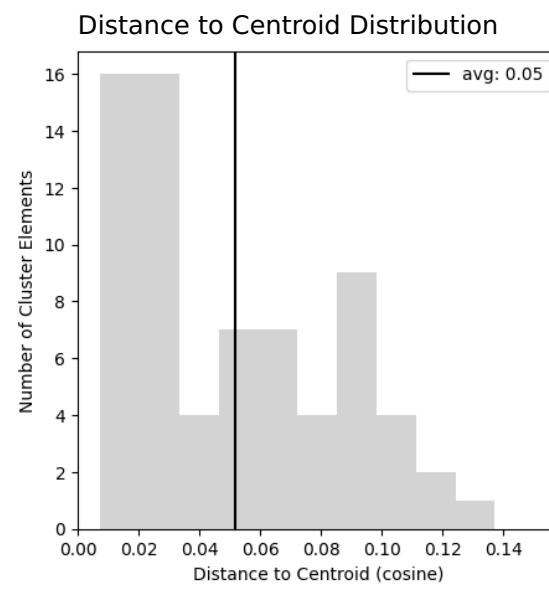
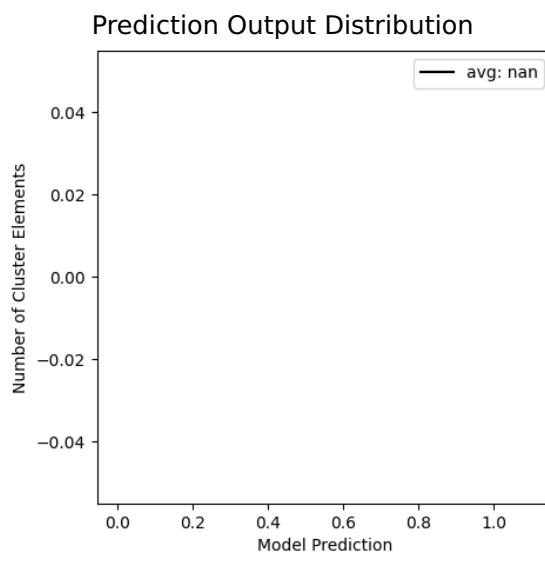
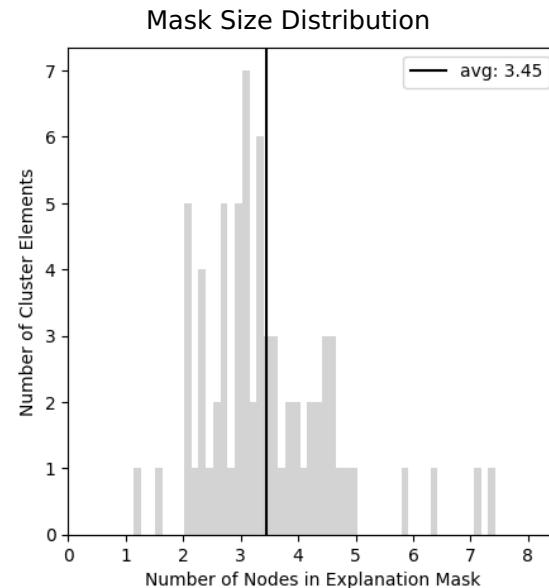
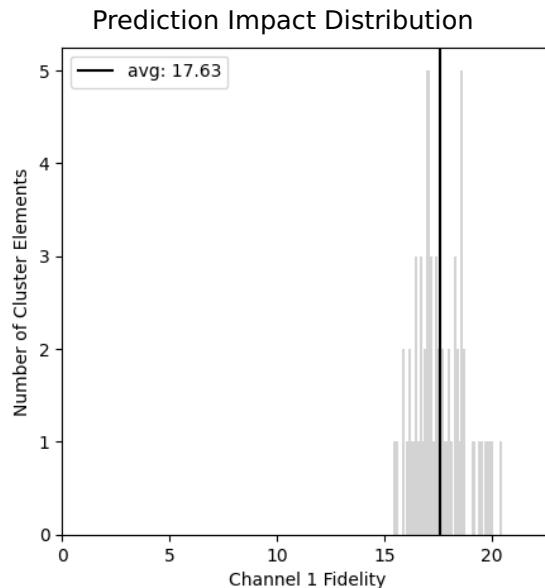
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	70
Channel Index	1.0 (0.0)

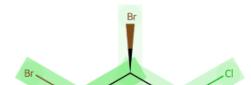
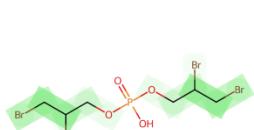
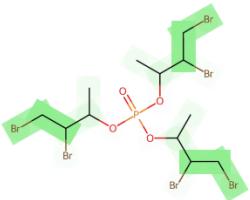
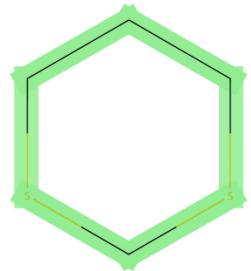
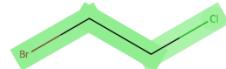
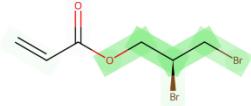
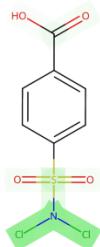
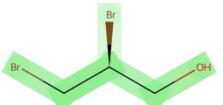
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



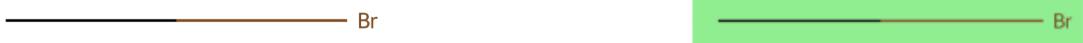
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a bromine atom (Br) bonded to a carbon atom (C-Br) might lead to the formation of reactive intermediates that are capable of interacting with DNA, thus potentially causing mutations. The bromine atom is a good leaving group, which means that it can easily depart from the carbon atom, leaving behind a positively charged carbon (carbonium ion) or other highly reactive species. These reactive intermediates are the ones that can interact with DNA, leading to mutagenicity.

**Hypothesis:** A C-Br bond in a molecular structure correlates with a medium level of mutagenic potential. The bromine atom's ability to leave and create reactive intermediates can interact with DNA, causing mutations. This structural feature is a noteworthy predictor but other factors may modulate the overall mutagenic potential of the molecule.

# Cluster #69 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 69, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of 19.4 ( $\pm 2.8$ ) on the prediction outcome.

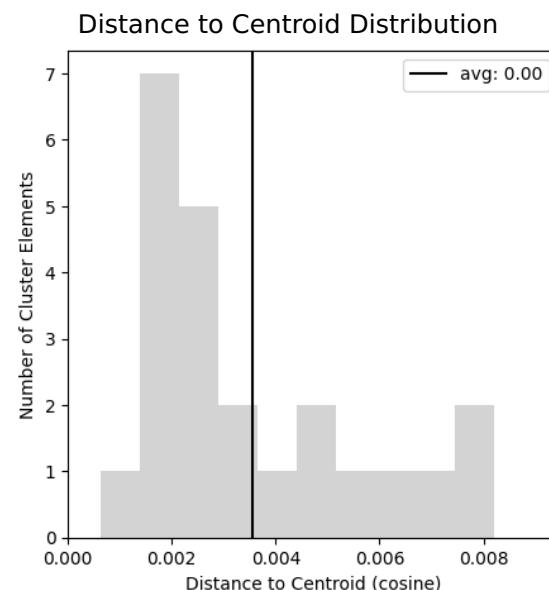
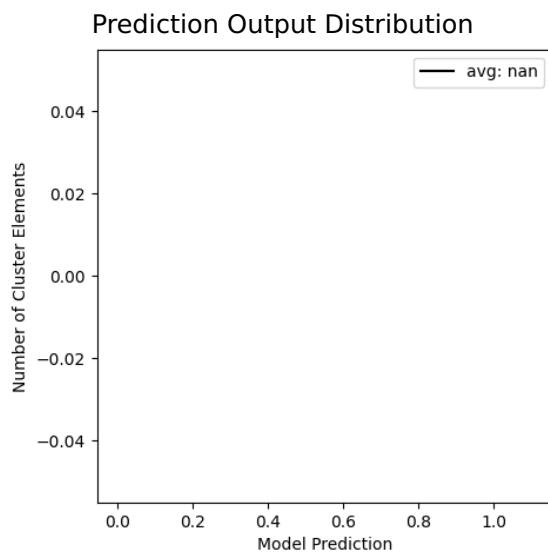
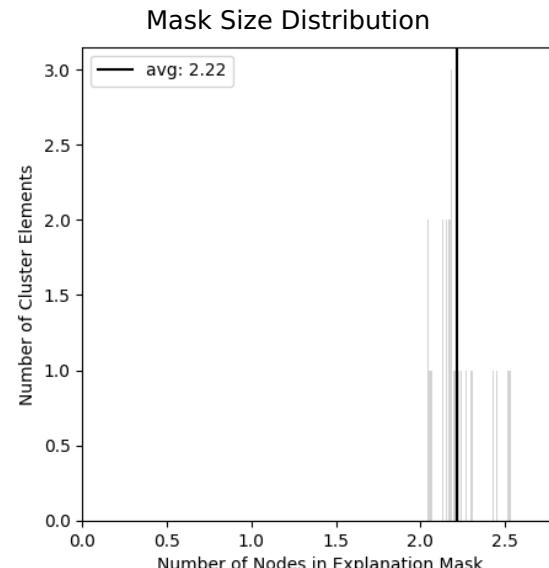
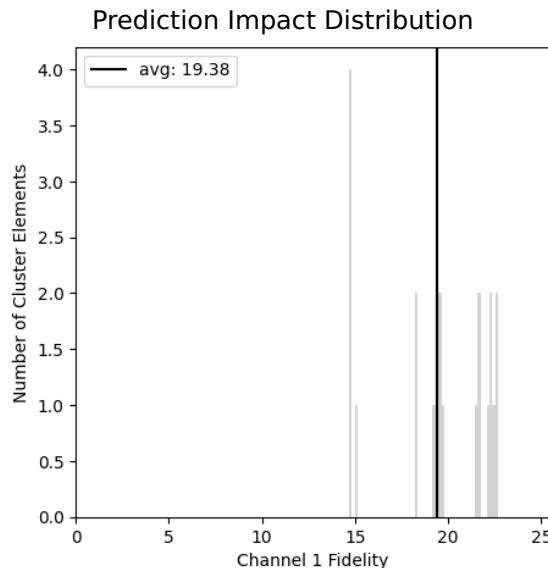
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	23
Channel Index	1.0 (0.0)

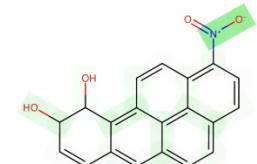
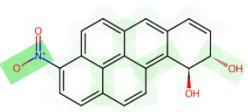
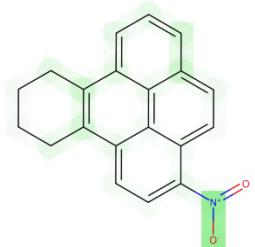
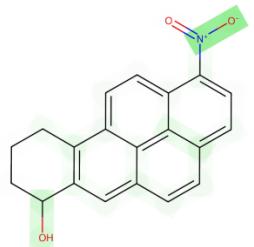
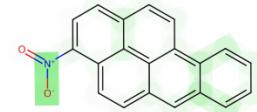
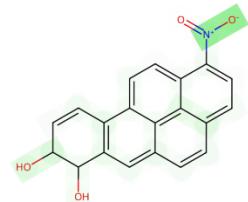
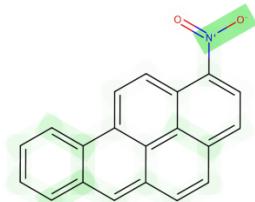
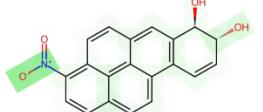
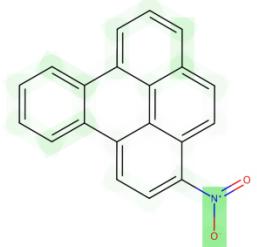
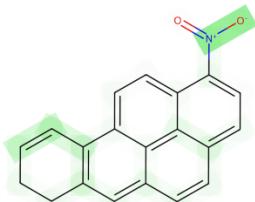
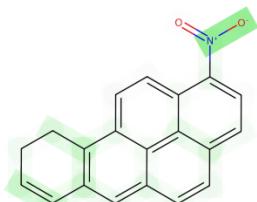
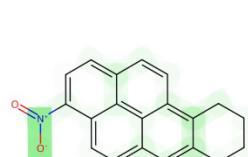
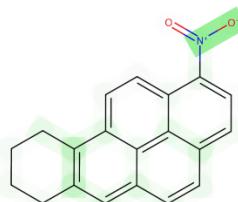
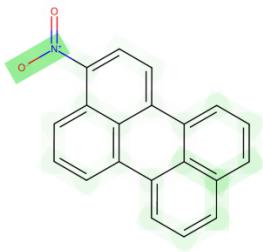
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



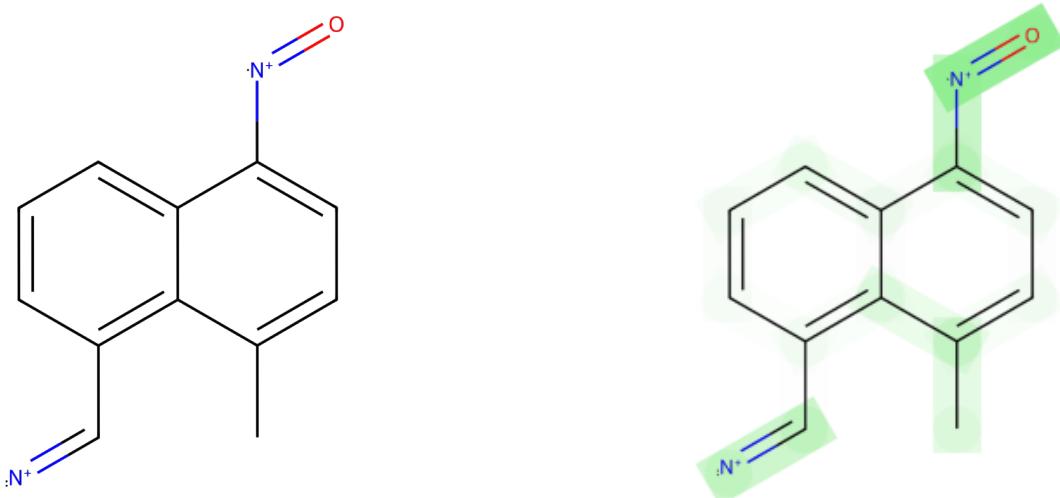
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the nitro group ( $-[N+]=O$ ) and the nitrile group ( $-C=[N+]$ ) in the given aromatic structure are both electron-withdrawing groups that can stabilize the positive charge on the molecule. This stabilization can make the molecule more reactive, hence, facilitate its interaction with DNA. The conjugation in the aromatic system might also lead to delocalization of electrons, which could increase the reactivity of the molecule. Reactive molecules are often involved in the modification of genetic material, leading to mutagenicity.

**Hypothesis:** The presence of the nitro and nitrile groups in the aromatic molecular fragment is hypothesized to contribute to its medium mutagenicity. The electron-withdrawing nature of these groups increases the molecule's reactivity, thereby enhancing its ability to interact and cause mutations in DNA. The delocalized electronic system of the aromatic rings could further enhance this reactivity, consolidating the link between structure and mutagenic property.

# Cluster #70 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 70, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 18.4 ( $\pm 2.5$ ) on the prediction outcome.

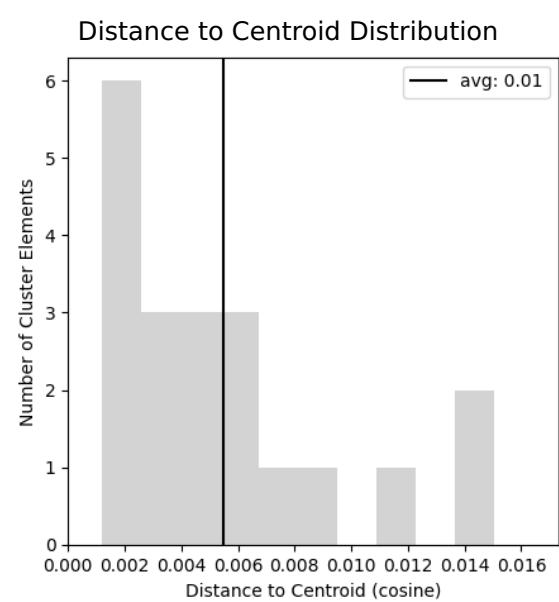
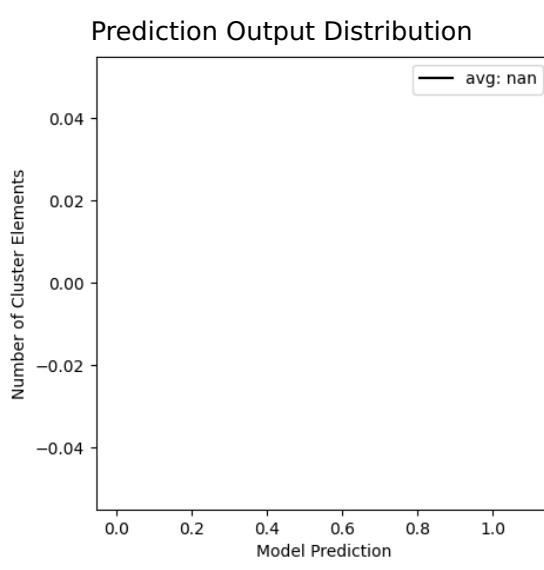
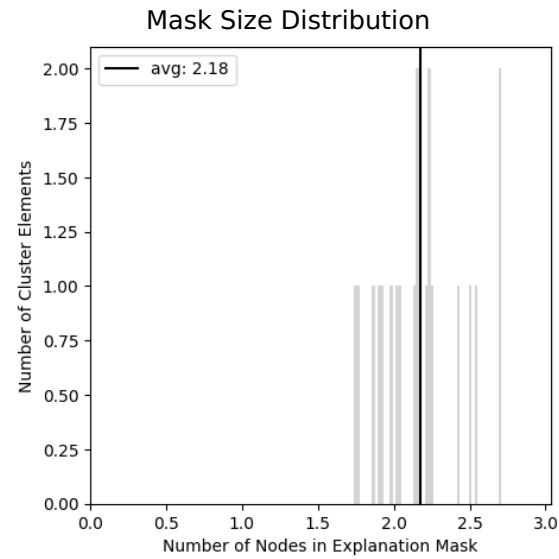
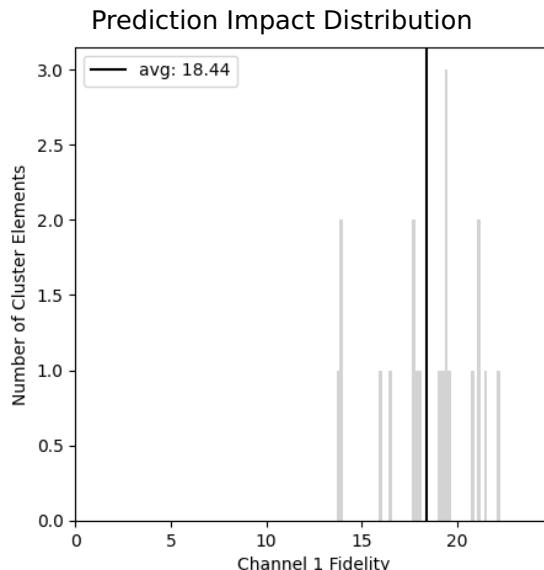
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	1.0 (0.0)

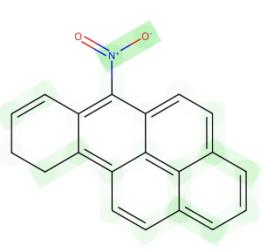
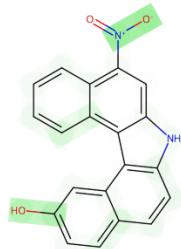
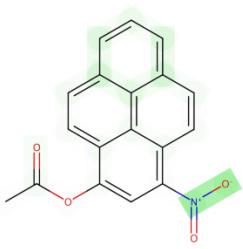
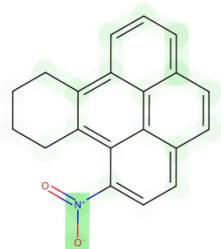
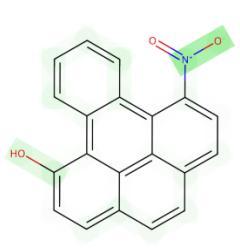
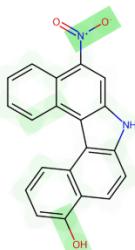
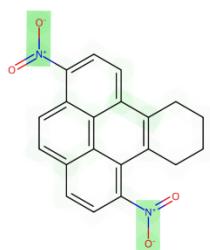
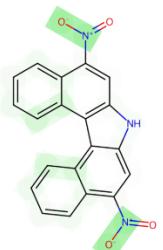
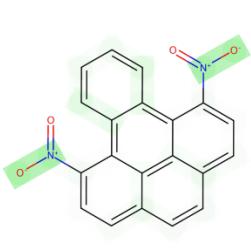
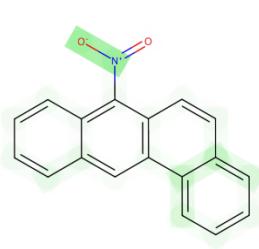
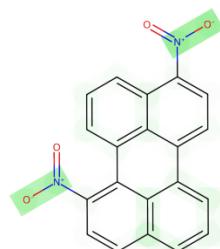
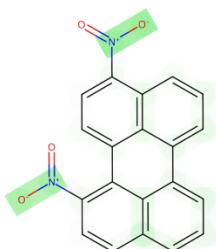
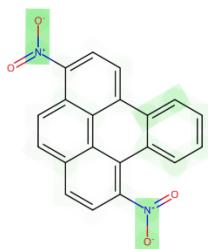
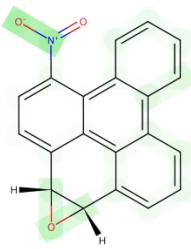
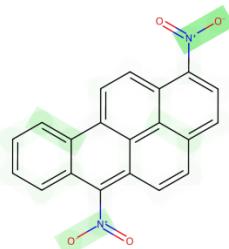
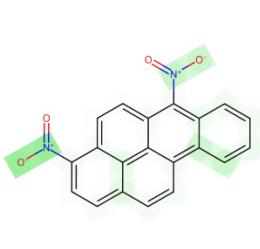
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



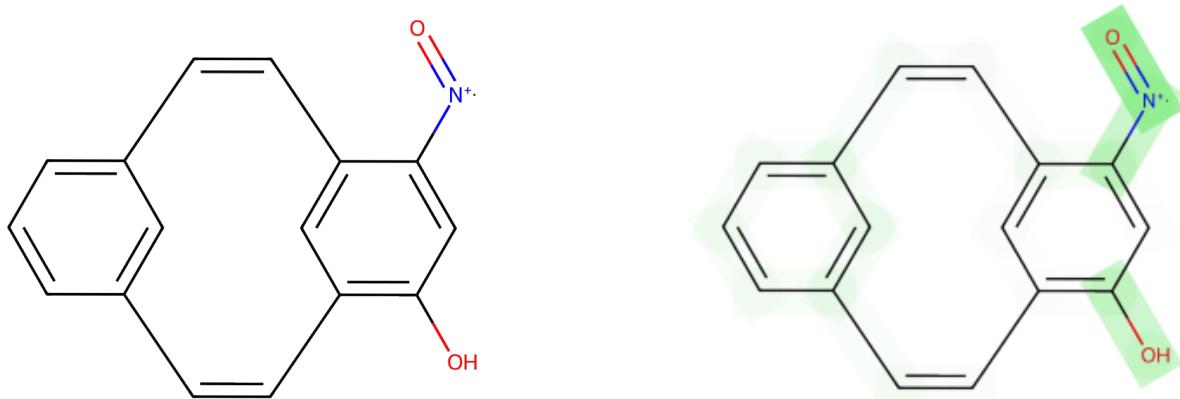
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a nitrophenol derivative characterized by a nitro group ( $O=[N+]$ ) attached to a phenolic ring ( $-c1:c:c(-O)c2\dots$ ) and an extensive conjugated system spanning three interconnected rings. Nitro groups are known for their electron-withdrawing properties, which can facilitate the formation of reactive intermediates capable of interacting with DNA. The extended conjugation allows for greater delocalization of charge, possibly stabilizing these intermediates. Phenolic  $-OH$  groups can form hydrogen bonds, enhancing the molecule's ability to interact with the genetic material.

**Hypothesis:** The presence of a nitro group and an extensive conjugated system in conjunction with a phenolic hydroxyl group in a molecule contributes to its medium mutagenicity. The nitro group increases the likelihood of forming reactive intermediates, while the extended conjugation stabilizes these species, and the hydroxyl group facilitates interactions with DNA, thereby increasing mutagenic potential.

# Cluster #71 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 71, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.4 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 17.2 ( $\pm 2.6$ ) on the prediction outcome.

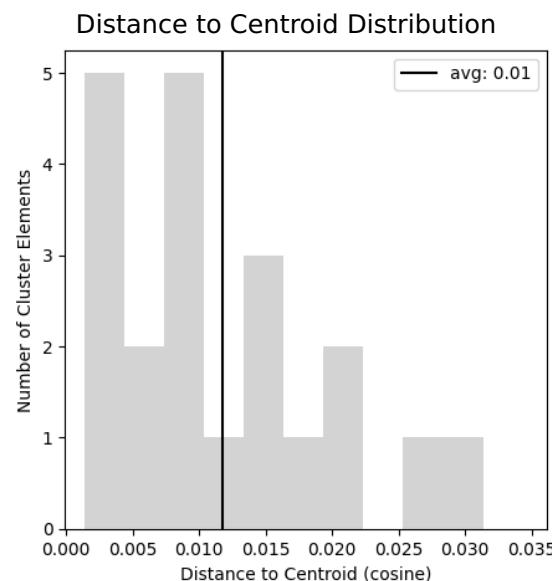
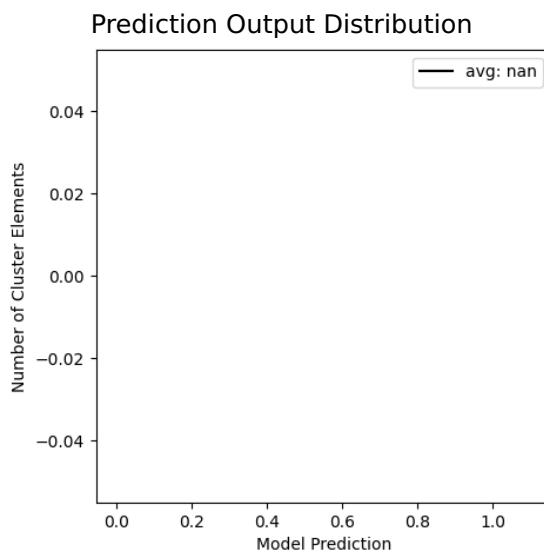
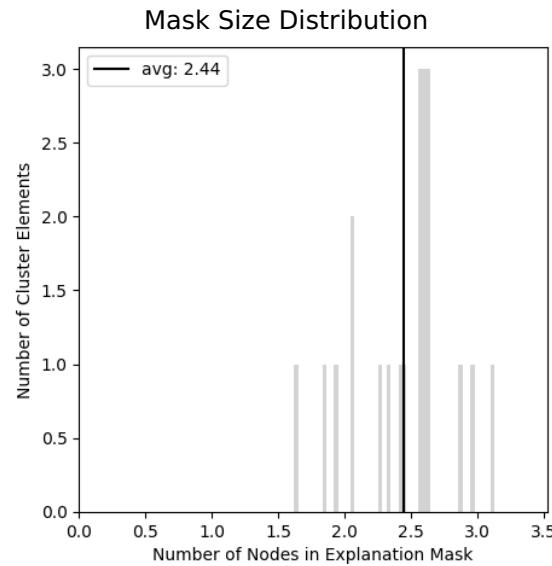
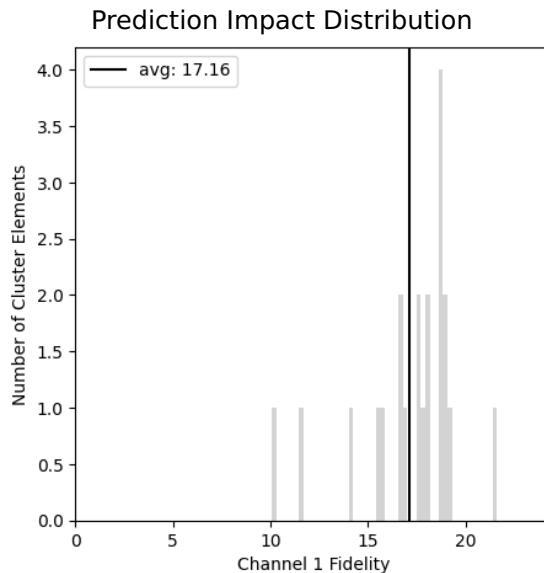
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

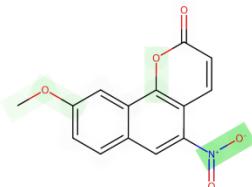
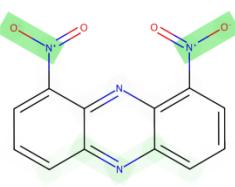
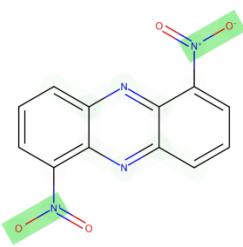
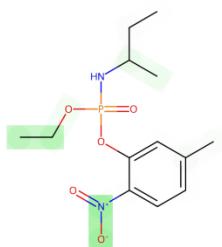
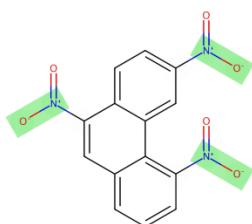
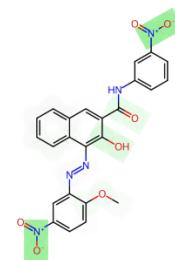
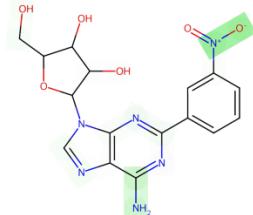
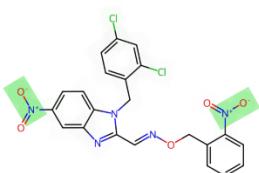
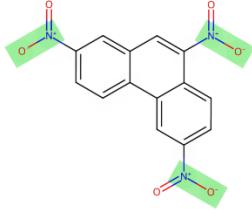
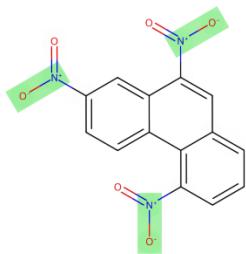
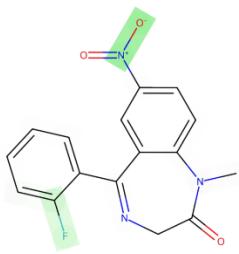
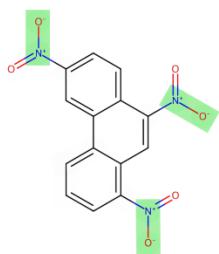
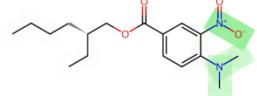
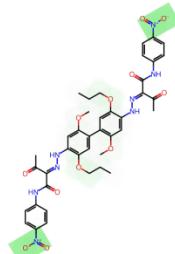
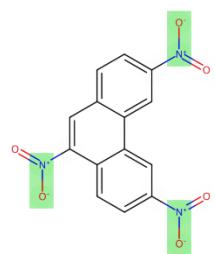
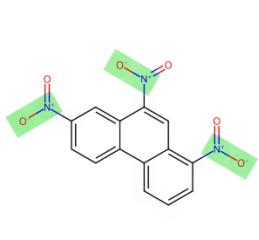
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



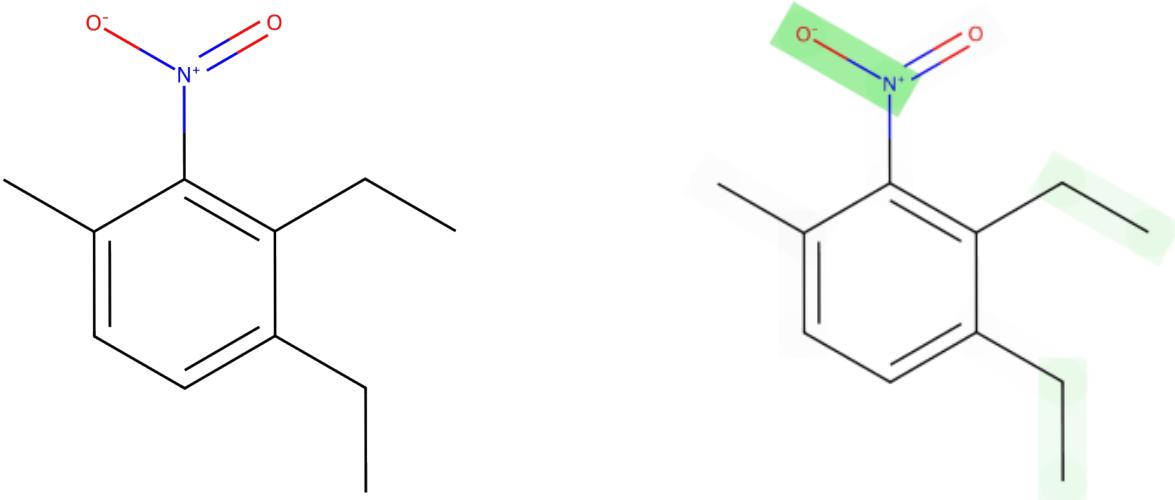
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure contains a nitro group (-[N+](=O)-[O-]), which is a known electrophile that can form adducts with nucleophilic sites in DNA, leading to mutagenicity. The molecule also features extended conjugation and aromatic rings, which can stabilize the interactions with genetic material. The presence of alkyl chains (C-C) might increase the lipophilicity of the compound, allowing it to cross cell membranes more easily and interact with DNA.

**Hypothesis:** The presence of the nitro group and aromaticity in the molecule increases its mutagenic potential. The nitro group can react with DNA, while the aromatic rings can promote planar interactions with DNA bases, both leading to a medium influence on mutagenicity. The alkyl chains potentially facilitate the transport of the molecule into the cell, thus influencing its mutagenic activity.

# Cluster #72 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 72, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 16.6 ( $\pm 3.2$ ) on the prediction outcome.

## Properties

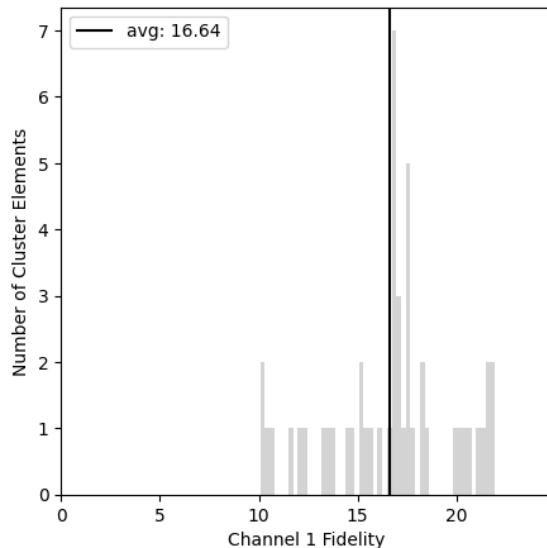
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	48
Channel Index	1.0 (0.0)

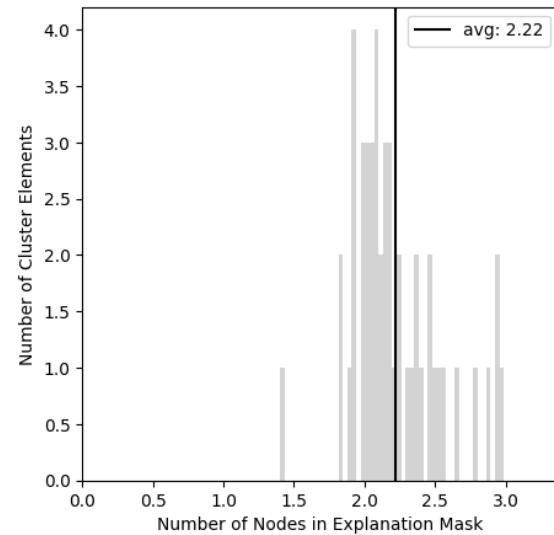
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

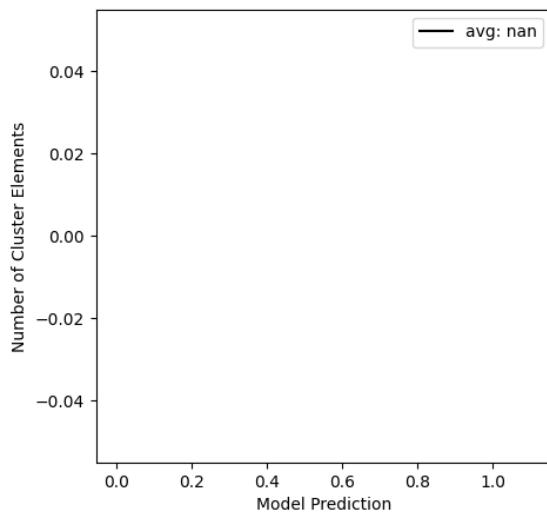
Prediction Impact Distribution



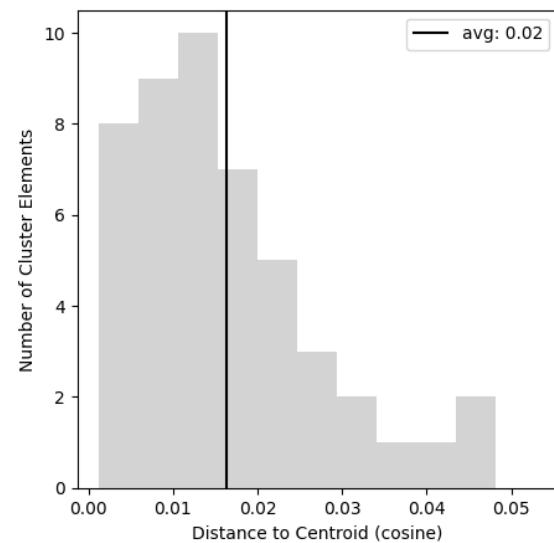
Mask Size Distribution



Prediction Output Distribution

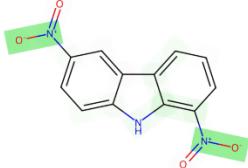
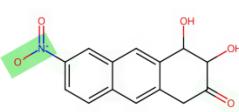
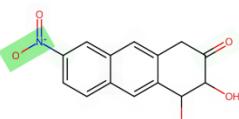
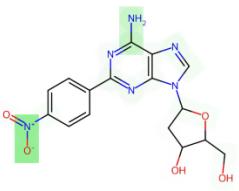
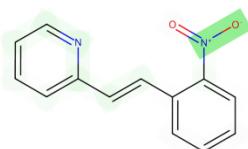
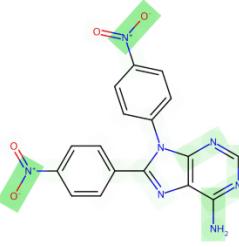
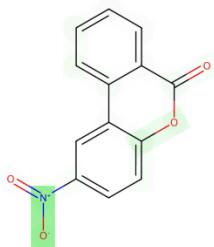
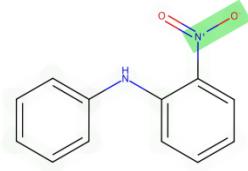
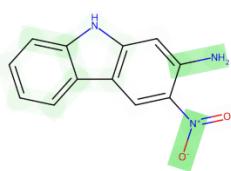
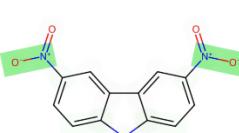
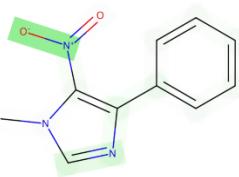
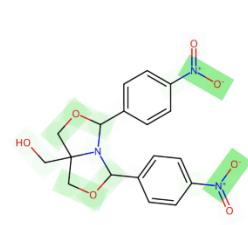
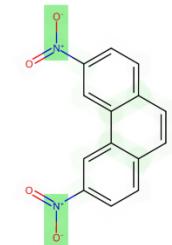
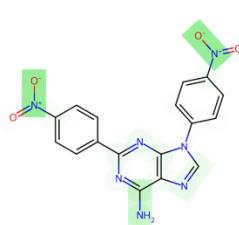
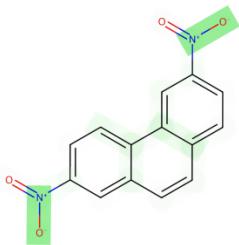


Distance to Centroid Distribution



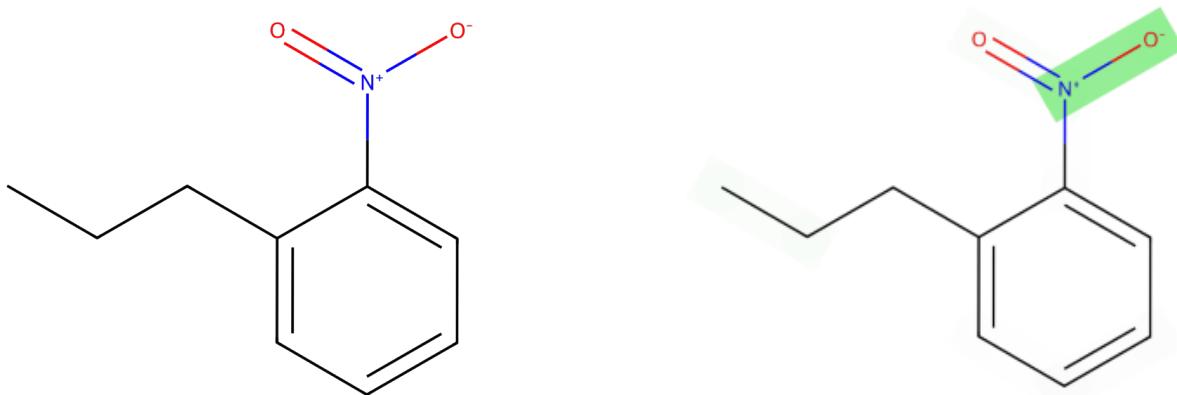
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure provided indicates the presence of an aromatic system next to a nitro group. The aromatic ring, typically stable, can be activated by the electron-withdrawing nitro group (-NO<sub>2</sub>), making the ring more reactive. The nitro group, being an electrophile, can form covalent bonds with nucleophilic sites in DNA, potentially causing mutations. Furthermore, the metabolism of aromatic nitro compounds in living organisms can lead to the formation of reactive intermediates that can directly interact with the genetic material.

**Hypothesis:** The given molecular fragment with an aromatic ring attached to a nitro group contributes to medium mutagenicity. The electron-withdrawing effect of the nitro group activates the aromatic ring towards electrophilic attack, facilitating the interaction with DNA. The metabolic activation of this nitroaromatic compound yields reactive species capable of inducing genetic alterations.

# Cluster #73 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 73, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.0 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 17.2 ( $\pm 1.9$ ) on the prediction outcome.

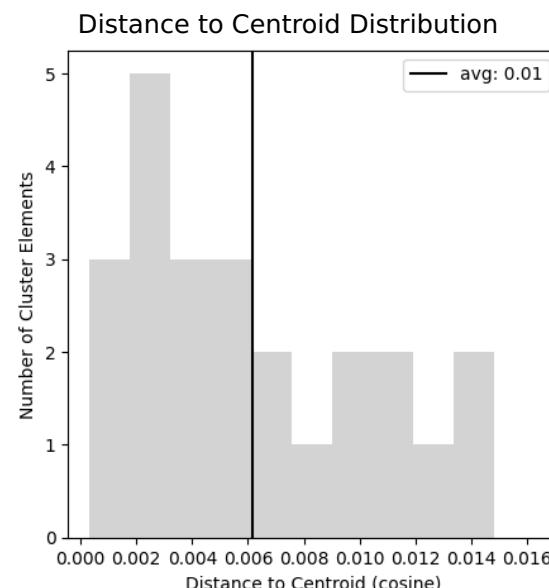
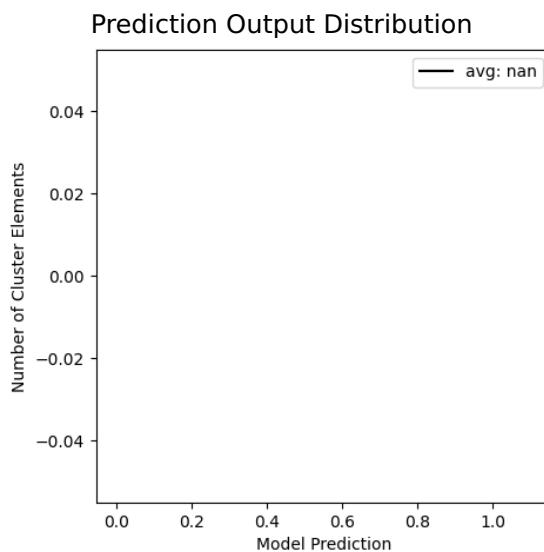
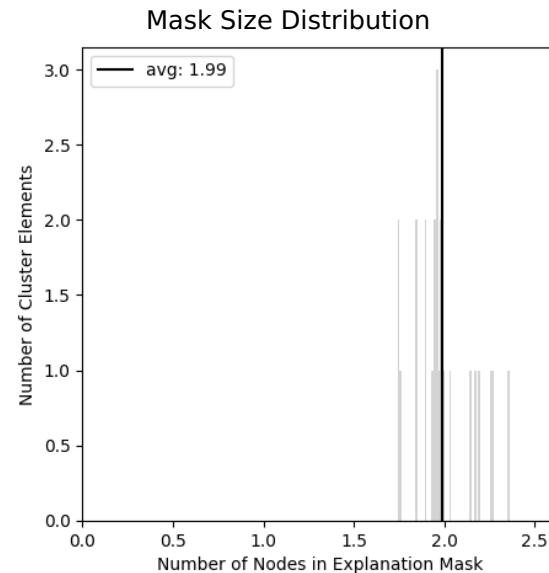
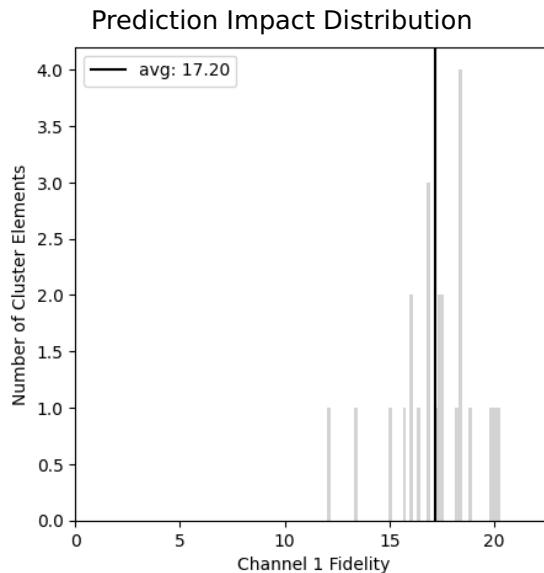
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	1.0 (0.0)

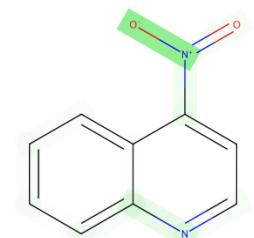
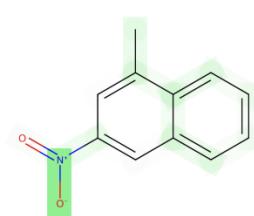
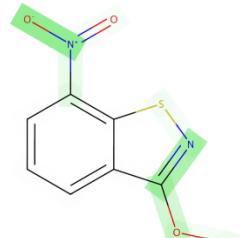
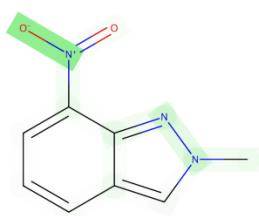
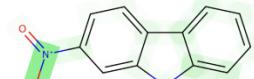
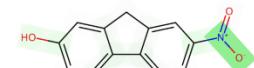
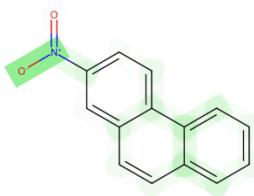
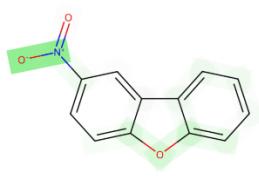
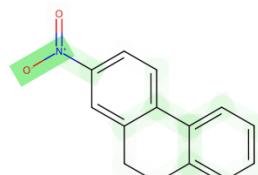
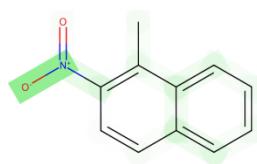
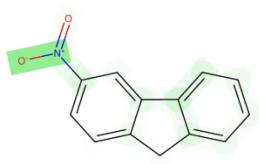
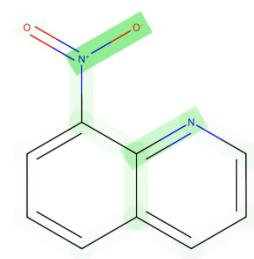
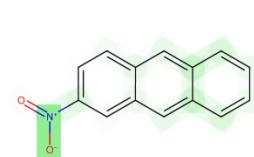
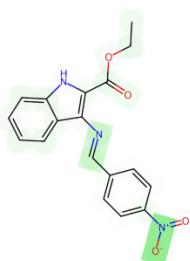
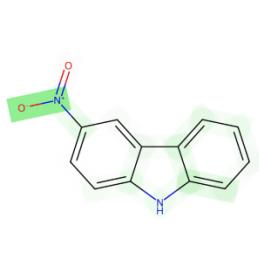
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



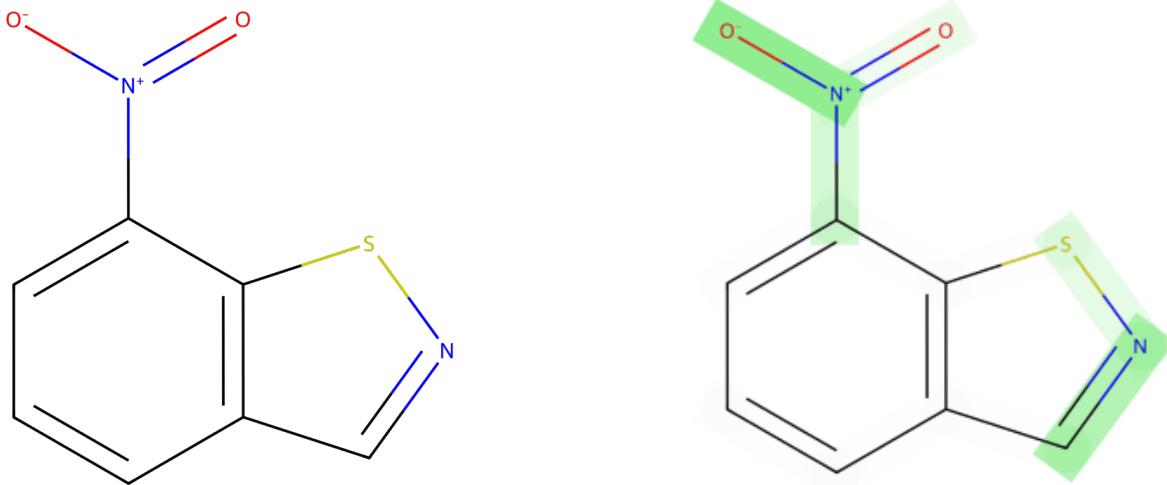
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure represented by the given SMILES code "O=[N+]([O-])c1c(c(c(c1)S)N)C(=O)c2ccccc2" includes a nitro group (O=[N+]([O-])) attached to an aromatic system, specifically a naphtho[1,2-d]thiazole moiety characterized by the two fused aromatic rings with nitrogen and sulfur atoms. Nitroaromatic compounds are known to be bioactivated in vivo by reduction of the nitro group to reactive nitroso derivatives, which can induce mutagenicity by forming adducts with DNA. The presence of heteroatoms such as nitrogen and sulfur in aromatic systems can also contribute to the electronic properties of the molecule, potentially stabilizing the reactive intermediates formed during metabolism, making the DNA adduct formation more likely.

**Hypothesis:** Molecules containing the substructure "O=[N+]([O-])c1c(c(c(c1)S)N)C(=O)c2ccccc2" have a MEDIUM influence towards mutagenicity due to the presence of the nitro group in conjugation with an aromatic system containing nitrogen and sulfur heteroatoms. The nitro group can be metabolically reduced to reactive intermediates, while the heteroatoms in the fused ring system can modulate the electronic structure of the molecule, both contributing to the potential formation of DNA adducts that can lead to mutation.

# Cluster #74 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 74, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 17.4 ( $\pm 2.6$ ) on the prediction outcome.

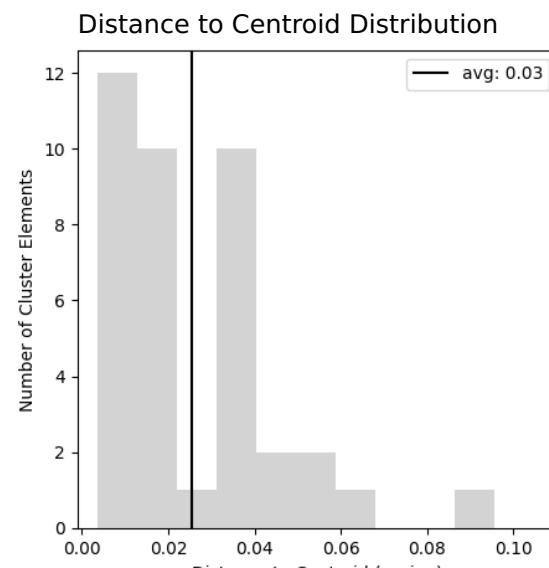
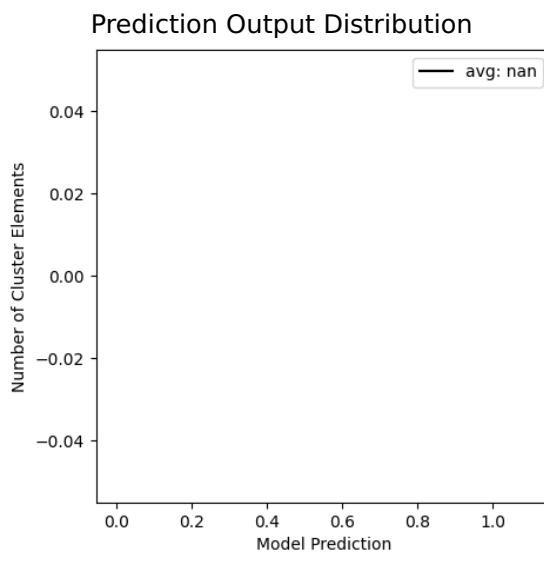
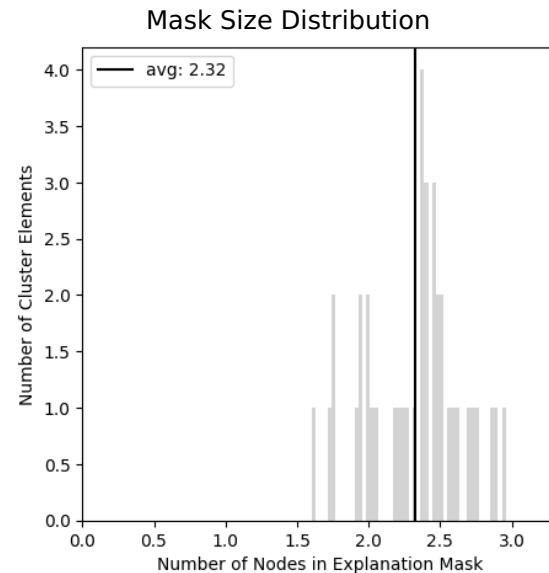
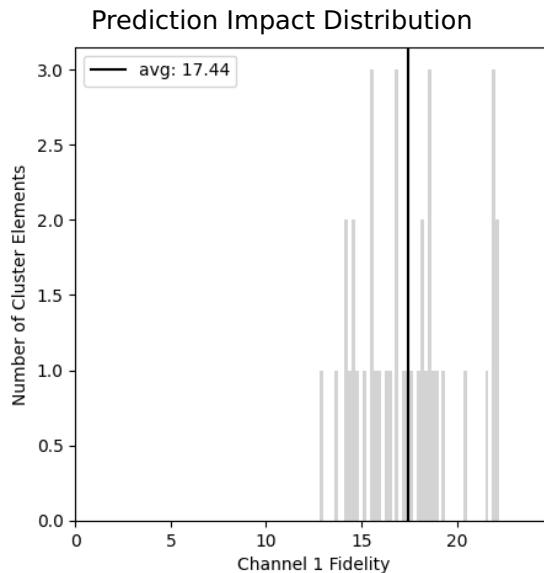
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	39
Channel Index	1.0 (0.0)

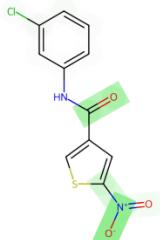
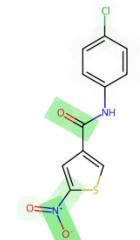
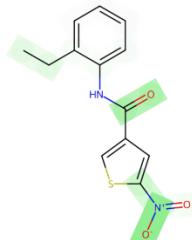
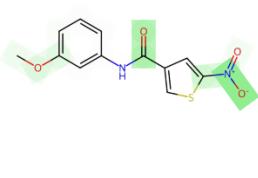
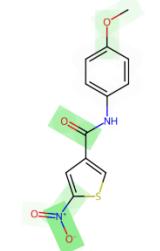
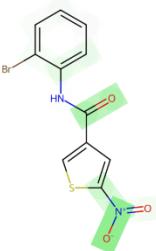
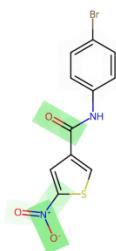
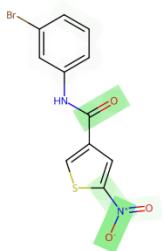
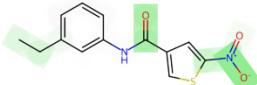
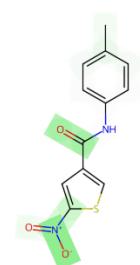
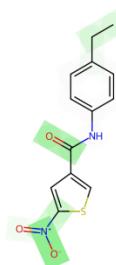
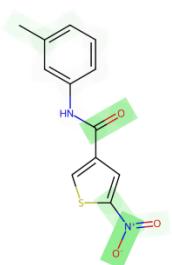
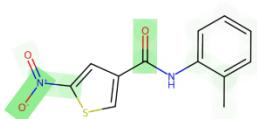
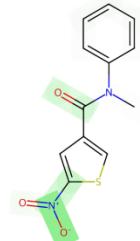
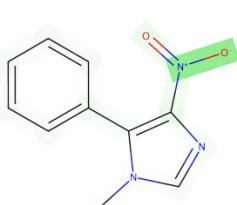
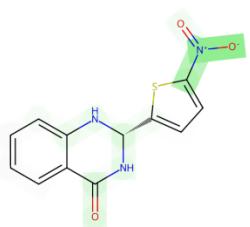
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



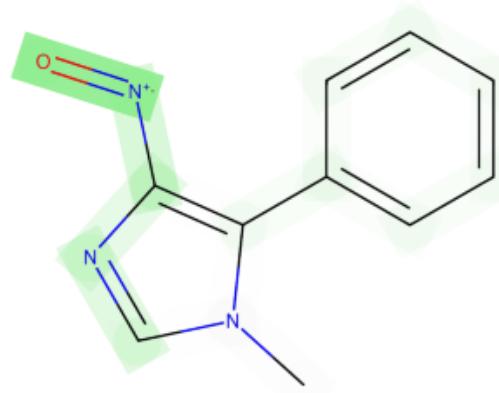
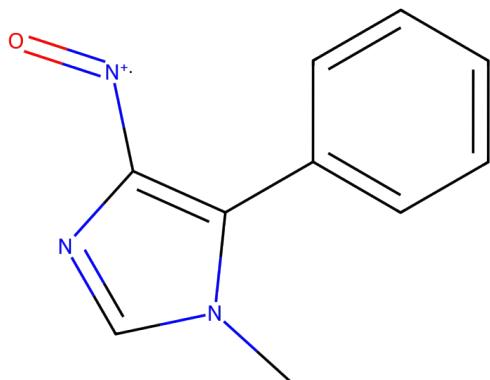
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation corresponds to a molecule containing a nitro group ( $-[N+]=O$ ) attached to a pyridine ring, which in turn is connected to a benzene ring. The nitro group is known to be an electron-withdrawing group, which could lead to the formation of reactive intermediates such as nitrenium ions that may attack the DNA, leading to mutations. Additionally, the planar structure of the aromatic rings allows easy intercalation between DNA base pairs, potentially causing errors in DNA replication.

**Hypothesis:** The molecular substructure with a nitro group attached to a conjugated aromatic system has a medium influence on mutagenicity. It is presumed that the electron-withdrawing effect of the nitro group increases the formation of reactive intermediates, while the planar aromatic system can intercalate with DNA, both of which could lead to mutagenic outcomes.

# Cluster #75 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 75, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 13.6 ( $\pm 1.6$ ) on the prediction outcome.

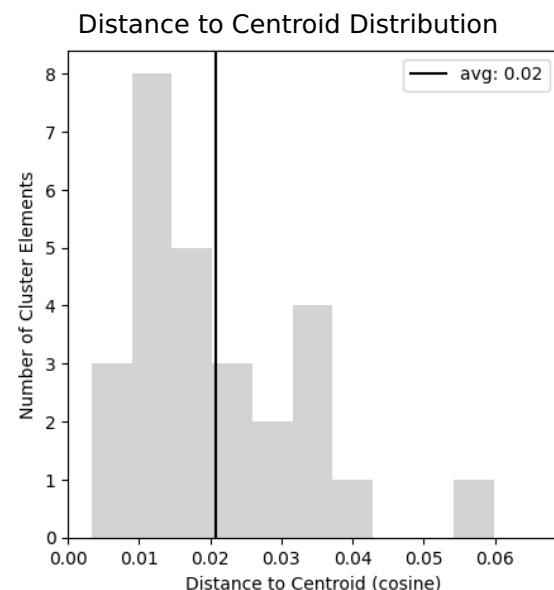
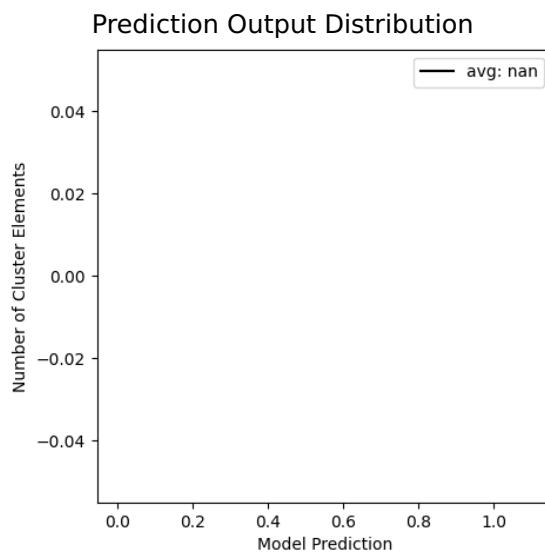
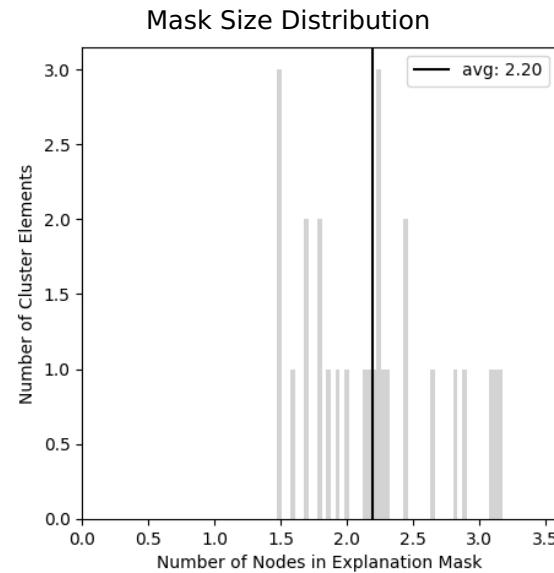
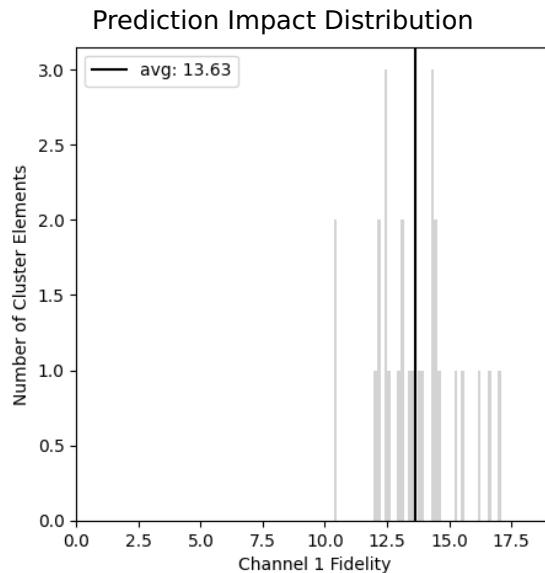
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	27
Channel Index	1.0 (0.0)

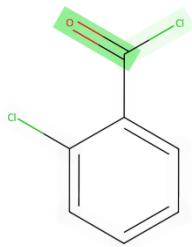
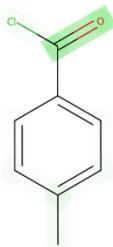
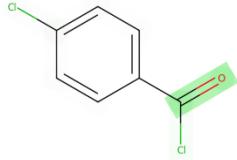
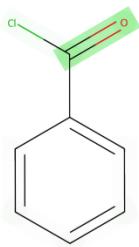
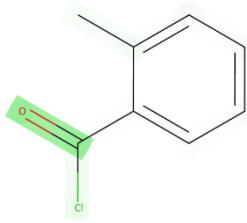
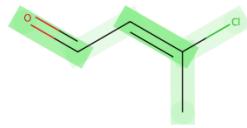
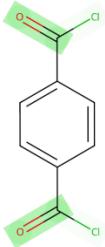
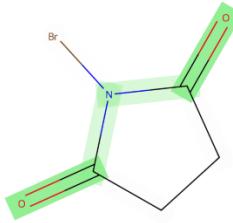
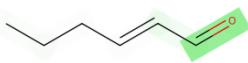
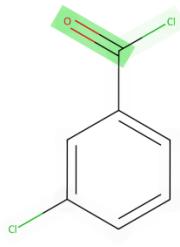
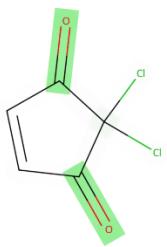
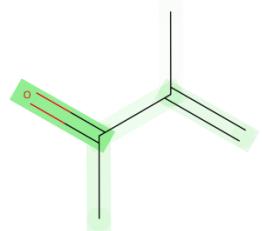
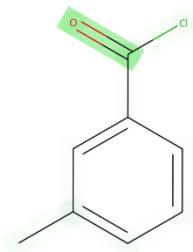
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



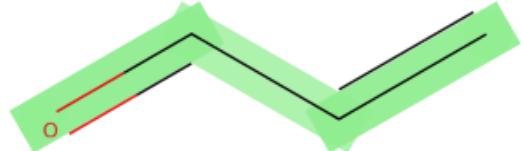
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the unsaturated bond ( $C=C$ ) adjacent to a carbonyl group ( $C=O$ ) is significant in mutagenicity due to its potential reactivity. The electron-rich double bond can interact with DNA nucleophilic sites, leading to undesirable modifications. The adjacent carbonyl group can stabilize the transition state of such reactions, making the entire substructure more reactive and thus more likely to result in mutagenic effects.

**Hypothesis:** Molecules containing the structure " $C=C-C=O$ " have a medium potential for mutagenicity due to the increased reactivity of the conjugated system. The double bond provides a site for interaction with genetic material, and the carbonyl group stabilizes reactive intermediates, both factors contributing to the probability of causing genetic modifications.

# Cluster #76 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 76, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.1 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 7.9 ( $\pm 1.4$ ) on the prediction outcome.

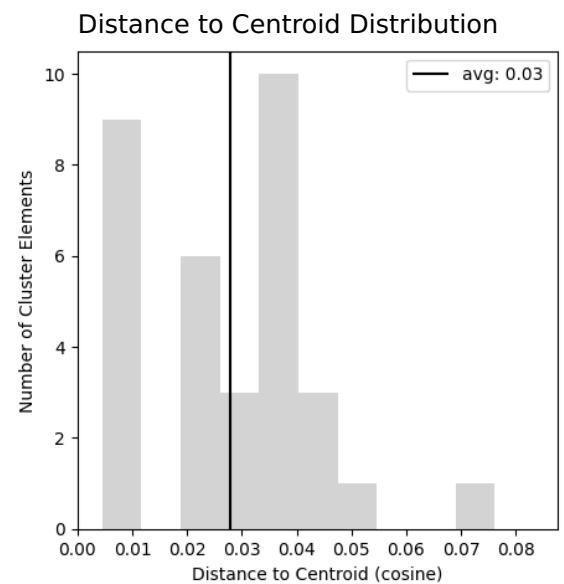
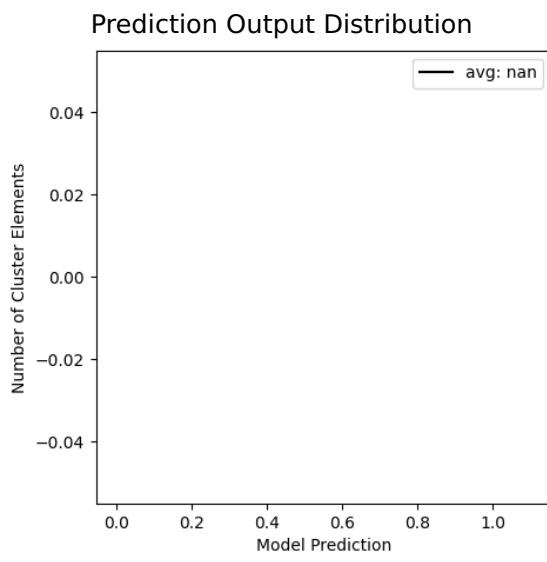
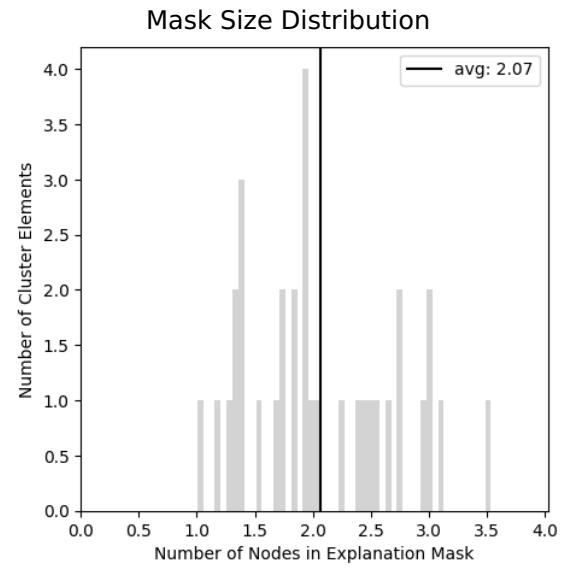
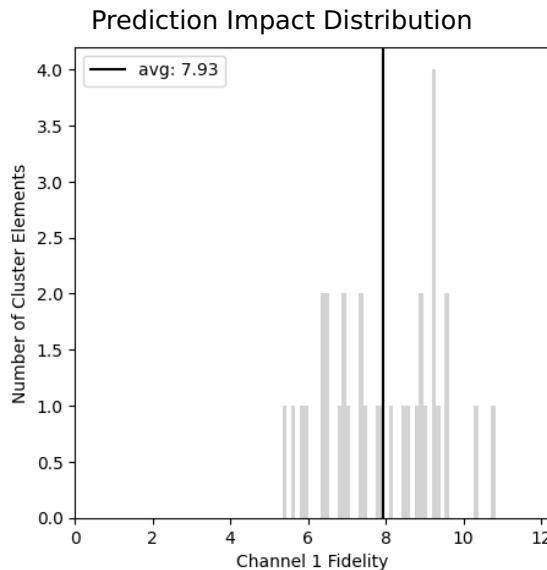
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	33
Channel Index	1.0 (0.0)

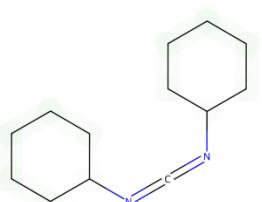
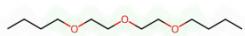
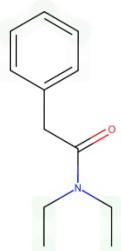
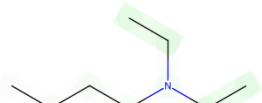
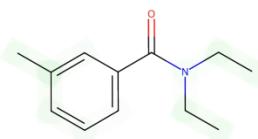
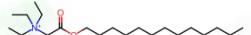
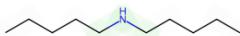
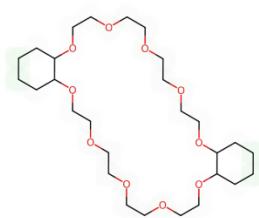
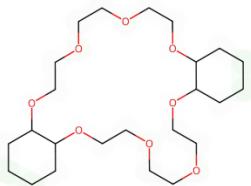
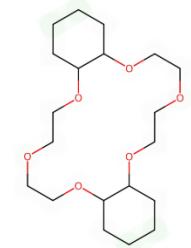
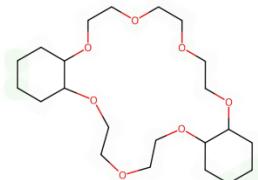
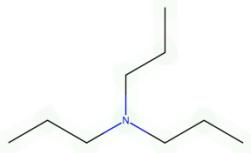
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure represented by "C-C-O-C-C" indicates a linear chain of carbon atoms with an ether linkage (oxygen atom connected to two carbon atoms). The presence of ether linkages in a molecule may be less likely to interact with DNA in a way that causes mutagenesis because it doesn't possess any significant electrophilic or reactive sites that typically interact with nucleic acids. This could lend them to be more stable and less reactive, meaning the likelihood of them causing mutations in DNA is lower. Substitution on the carbon atoms or the presence of additional functional groups could alter this basic fragment's properties.

**Hypothesis:** Molecules containing the "C-C-O-C-C" substructure have a **SMALL** influence on mutagenicity due to the stability and low reactivity of ether linkages. The carbon chain and ether oxygen do not readily form DNA adducts or cause DNA strand breaks, which are common mechanisms for mutagenesis, hence they are less likely to be implicated in mutagenic activity.

# Cluster #77 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 77, from importance channel 1 (*mutagenic*), represents a motif consisting of 1.9 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 12.3 ( $\pm 0.9$ ) on the prediction outcome.

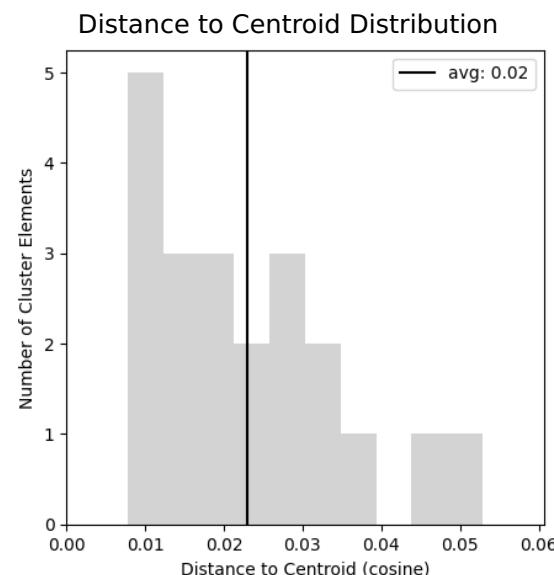
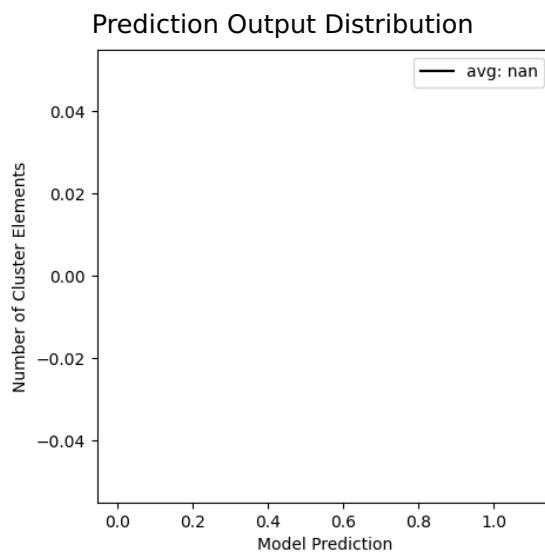
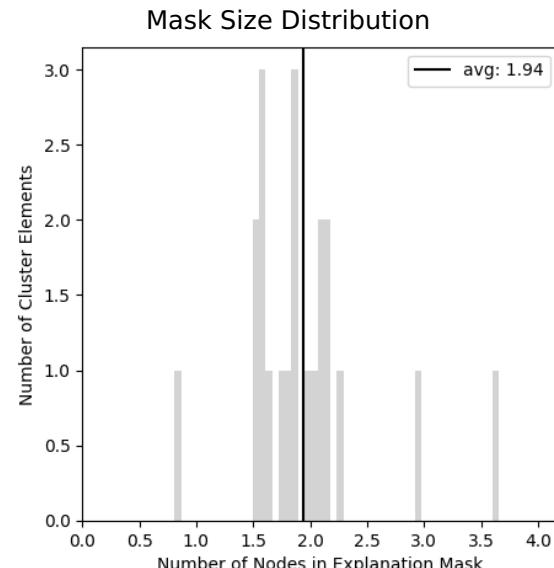
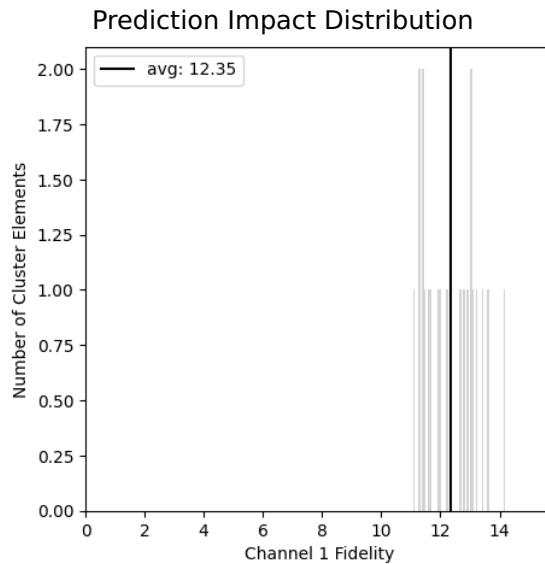
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

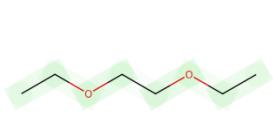
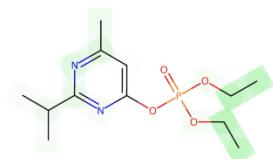
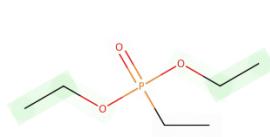
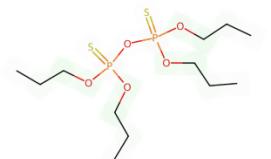
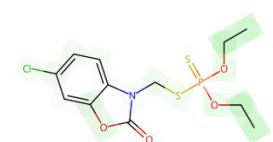
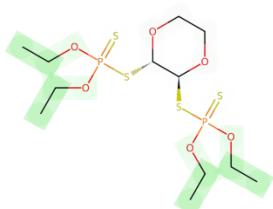
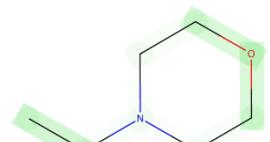
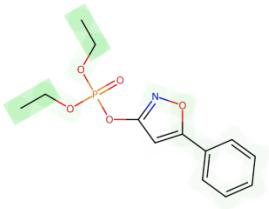
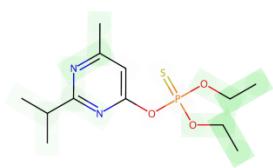
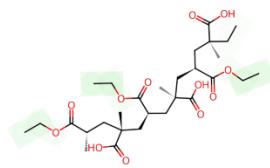
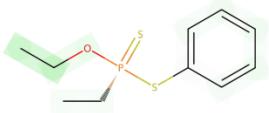
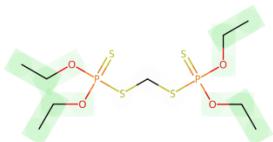
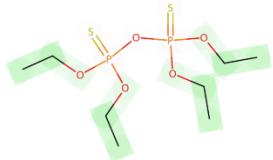
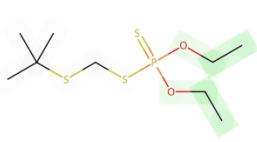
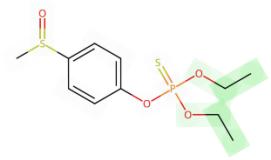
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "C-C-O-P-C-C" suggests the presence of an organophosphate group, which is commonly associated with mutagenic activity due to its ability to interfere with DNA synthesis and repair mechanisms. The phosphorus atom bonded to oxygen (O-P) can form reactive oxygen species or can interfere with the function of essential enzymes such as those involved in the process of DNA replication and repair, leading to mutations. Organophosphates can also alkylate DNA directly, causing structural alterations that result in mutagenesis.

**Hypothesis:** Molecules containing the substructure "C-C-O-P-C-C" have a medium mutagenic potential due to the organophosphate functional group's ability to disrupt DNA processes or directly alkylate DNA. The "C-C" linkages to the organophosphate group suggest a possible lipid solubility, facilitating the molecule's interaction with the lipid-rich cell membranes and potentially increasing its mutagenic potential by enhancing cellular uptake.

# Cluster #78 - mutagenic

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 78, from importance channel 1 (*mutagenic*), represents a motif consisting of 2.3 ( $\pm 0.9$ ) nodes. The concept is generally associated with an impact of 7.5 ( $\pm 1.6$ ) on the prediction outcome.

## Properties

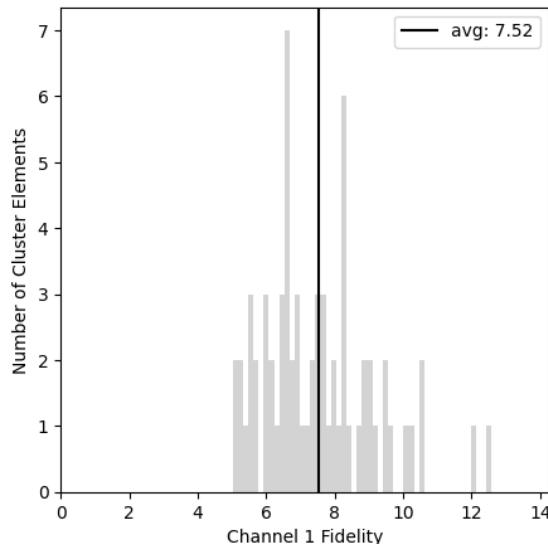
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	67
Channel Index	1.0 (0.0)

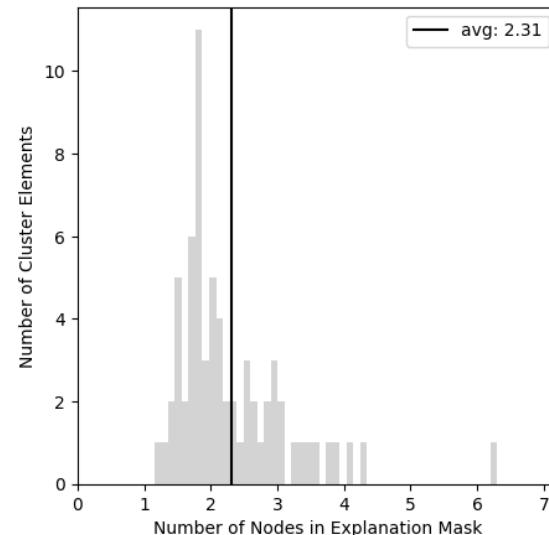
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

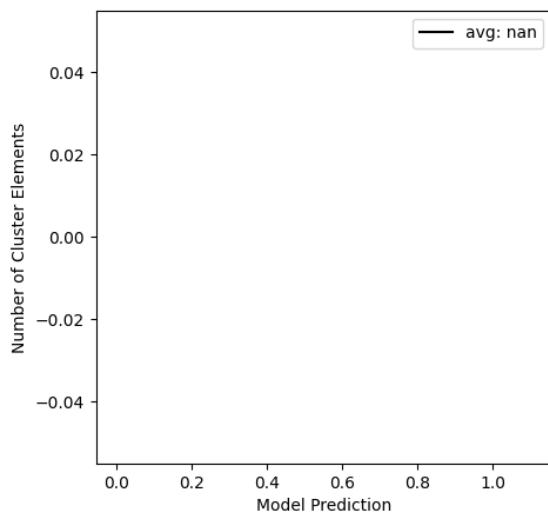
Prediction Impact Distribution



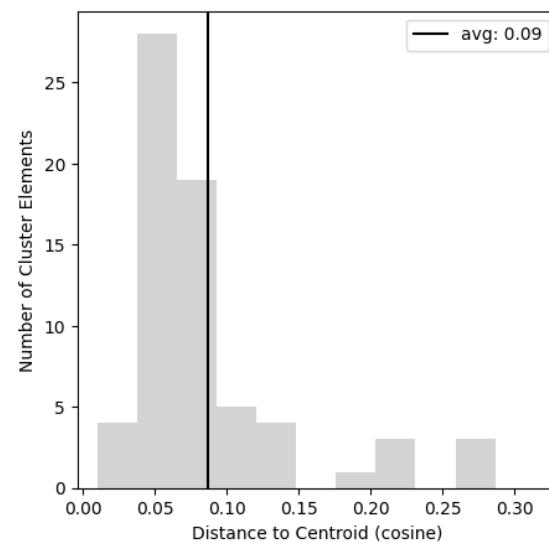
Mask Size Distribution



Prediction Output Distribution

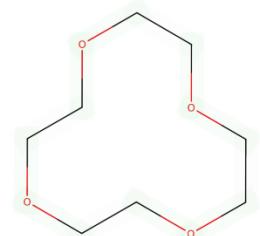
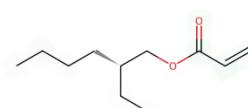
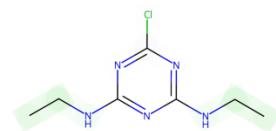
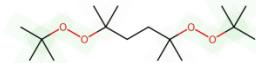
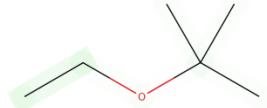
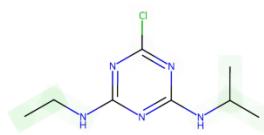
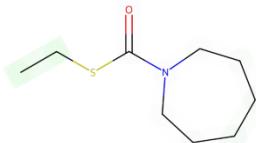
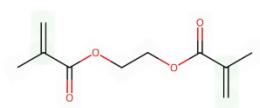
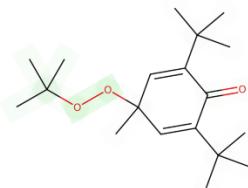
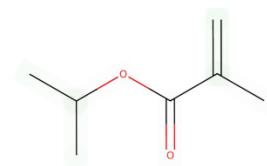
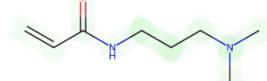
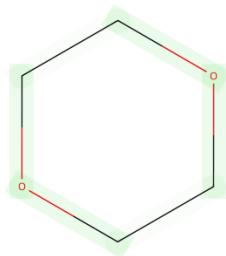
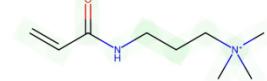
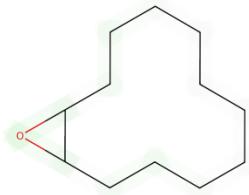


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure represented by "C-C-C" indicates a simple chain of three carbon atoms connected by single bonds, which is a common feature found in alkanes. Typically, simple alkanes have limited reactivity due to the strength of their C-C bonds and lack of functional groups that interact with DNA bases or cause DNA lesions, which are necessary for mutagenicity. Therefore, it's reasonable to hypothesize that the minimal structural complexity and reactivity of this substructure results in a lower potential to cause mutations.

**Hypothesis:** The "C-C-C" substructure is associated with a small influence on mutagenicity due to its simple and stable alkane-like nature. The lack of functional groups that could form adducts with DNA and the strong C-C single bonds, which are less likely to break and form reactive intermediates, contribute to the lower mutagenic potential.