Chemical Research in Toxicology

pubs.acs.org/crt

Article

(Q)SAR Approaches to Predict the Extent of Nitrosation in Pharmaceutical Compounds

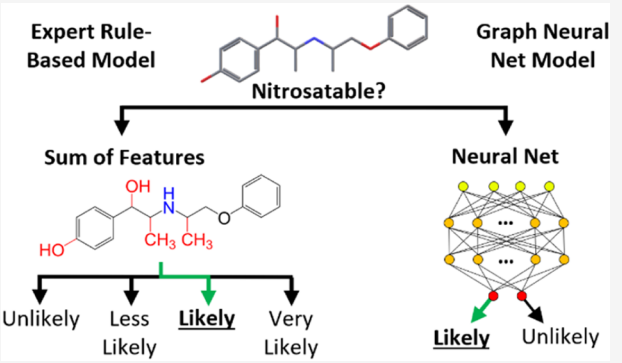
Krystle Reiss,* Roustem Saiakhov, and Suman Chakravarti

Cite This: *Chem. Res. Toxicol.* 2025, 38, 433−447

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Since their discovery as impurities in numerous pharmaceuticals beginning in 2018, there has been a strong push to predict and prevent the formation of mutagenic nitrosamines. Several experimental methods, particularly the Nitrosation Assay Procedure, have been developed to predict a molecule's susceptibility to nitrosation. Here, we have compiled the results of hundreds of these experiments from the literature to construct two structure−activity relationship models: a statistical model and an expert rule-based model. The statistical model has been built with graph neural networks and was trained on a dataset of 207 nitrogen-containing molecules. This model makes a binary call for each nitrogen center, predicting if it is likely to be nitrosated or not. Conversely, the rule-based model labels each possible nitrosamine product as one of four categories, ranging from "unlikely" to "very likely". It makes this determination based on 15 rules, which cover 12 deactivating (inhibit nitrosation) and 3 activating (favor nitrosation) features that have been drawn from the literature. Both models perform remarkably well, with accuracies of ∼80%. The rule-based model is generally biased toward favoring nitrosation while the statistical model is more likely to classify an amine as un-nitrosatable due to the makeup of the dataset. Using the models together can balance these biases and further improve the reliability of both.



INTRODUCTION

N-nitrosamines (NAs) are a class of molecules that have been suspected carcinogens since the 50s.[1−3] They are characterized by a nitroso group (N═O) bonded to an amine by a N−N single bond. Nitrosamines are typically stable until metabolically activated, making them long-lasting impurities.[4] *In vivo*, they are activated by cytochrome P450, resulting in a reactive carbocation that can bind to DNA, causing potentially carcinogenic mutations.[1,4] In addition to the typical amines, other nitrogen-containing groups can also undergo nitrosation, such as amides, ureas, guanidines, and carbamates. Due to polarization of the amide bond, these nitrosated groups do not require metabolic activation,[5] making them less long-lived than nitrosamines, although some can have half-lives of days to months.[6]

In 2018, the nitrosamine, NDMA, was found as an impurity in the drug valsartan.[7] This was quickly followed by the discovery of *N*-nitrosodimethylamine (NDMA), *N*-nitrosodiethylamine (NDEA), *N*-nitrosopiperidine (NPIP), and other small dialkyl *N*-nitrosamine impurities (Figure 1(a)) in other sartans[8] and other active pharmaceutical ingredients (APIs).[9−13] In 2021, varenicline (branded as Chantix and Champix) was recalled when *N*-nitrosovarenicline was found as an impurity.[14,15] Unlike the small NAs found previously, *N*-nitrosovarenicline is the *N*-nitrosated form of the API, varenicline. These larger nitrosamine impurities are known as Nitrosamine Drug Substance-Related Impurities (NDSRIs) and several examples are shown in Figure 1(b). In the following years, several more drugs (irbesartan,[16] propranolol,[17] orphenadrine,[18] quinapril,[19] sitagliptin,[20] and more) were recalled for similar impurities. An industrial survey followed in November 2022, which found that 90% of potential complex NDSRIs and 20% of small nitrosamines had been confirmed by analytical testing.[21] That same month, an *in silico* study of 12,000 small molecules found that 40% of drugs and 30% of impurities are potential nitrosamine precursors.[22] Similar to the industry survey, most of the potential nitrosamines were complex NDSRIs like nitrosovarenicline, although more potent small nitrosamines (*e.g.*, NDMA) were also found. Currently, NAs are controlled as mutagenic impurities according to the ICH M7[23] as a part of the Cohort of Concern.

Table 1 shows six example drugs with their daily dosage, nitrosamine yield in an excess of nitrite, theoretical maximum daily nitrosamine intake, and Acceptable Intake (AI). The AI is

Received: October 15, 2024
Revised: February 17, 2025
Accepted: February 19, 2025
Published: February 27, 2025

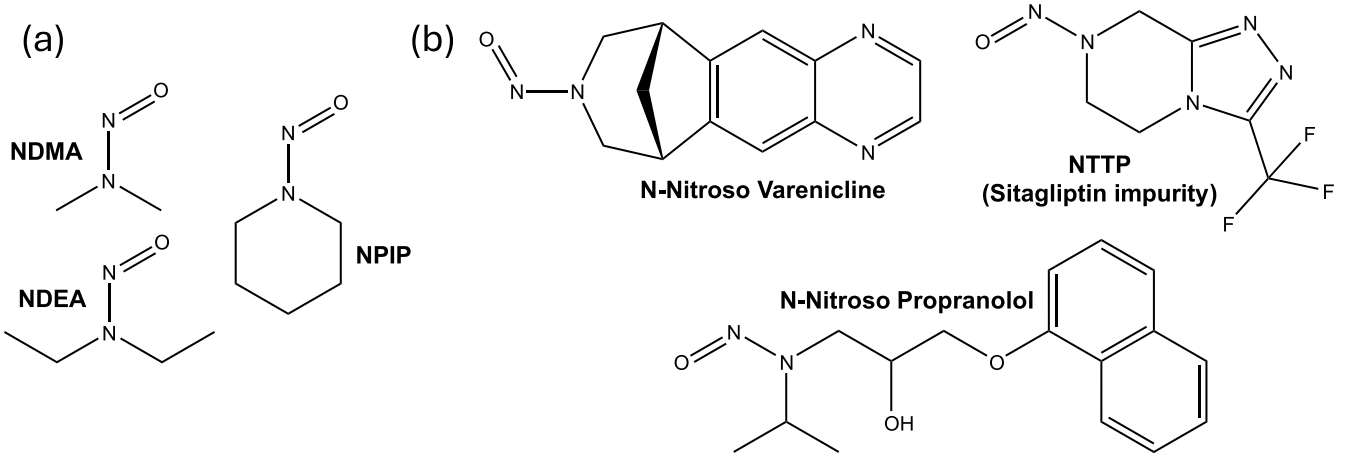


Figure 1. Examples of (a) small-molecule nitrosamine impurities and (b) Nitrosamine Drug Substance-Related Impurities (NDSRIs).

Table 1. Comparison of Theoretical Maximum Daily Intake of Nitrosamines versus the Maximum Acceptable Intake Determined by the CPCA[a]

Parent	Upper Dosage	Nitrosamine	Nitrosamine Yield (%)	Maximum Daily Intake of NA (ng/day)	CPCA AI (ng/day)	CPCA Category
Desipramine	200 mg/day[29]	*N*-nitrosodesipramine	2.8[26]	6,200,000	18 or 26.5	1
Ergometrine	6 mg/day[30]	*N*-nitrosonorergometrine	0.8[27]	50,000	1500	4
Indomethacin	150 mg/day[31]	NDMA	0.01[28]	3000	18 or 26.5	1
Prazosin	20 mg/day[32]	*N*-nitrosohexamethylenimine	2.1[26]	180,000	100	2
Thioproperazine	5 mg/day[33]	*N*-nitroso-*N*′-methylpiperazine	1.2[34]	17,000	400	3
Triprolidine HCl	2.5 mg/5 mL × 20 mL/day[35]	1-nitrosopyrrolidine	0.001[34]	32	1500	4

[a]Dosages are based on "Proper Use" as published by the Mayo Clinic or the NIH. Nitrosamines and their percent yields are based on experiments done in acidic conditions (pH 3-4) with an excess of nitrite. These results show that even low conversion to nitrosamines can far exceed the acceptable intake recommended by CPCA guidelines.

determined using the Carcinogenic Potency Categorization Approach (CPCA),[24,25] which categorizes nitrosamines into one of five potency categories, with Category 1 being the most potent. The CPCA's scope is limited to *N*-nitroso compounds that have a carbon atom bonded on both sides of the nitrogen and no heteroatoms double-bonded to those carbon atoms, which excludes *N*-nitrosamides and similar compounds from the CPCA. Each category has its own recommended AI, ranging from 26.5 or 18 ng/day (US FDA or EMA) for Category 1 to 1500 ng/day for Categories 4 and 5. The results in Table 1 show that even a small amount of nitrosation (1% or less) can lead to nitrosamine intake that is orders of magnitude above acceptable limits. The only case that falls below its respective nitrosamine AI is Triprolidine, which has a maximum dosage of 10 mg/day and only produces a 0.001% yield of Category 4 molecule. While it is worth noting that these are "worst case" scenarios, where the drug molecules are exposed to an excess of nitrite, there are other drugs with far higher nitrosamine yields under these same conditions, including oxytetracycline (63%),[26] varenicline (69%),[27] duloxetine (85%),[27] and ethambutol (>99%).[28] Easily nitrosated drugs like these could feasibly achieve >0.001% nitrosation under storage conditions or during oral intake, depending on the availability of nitrosating agents.

There are two common mechanisms for *N*-nitrosation of secondary and tertiary amines,[36] which are shown in Figure 2. Secondary amines can be nitrosated directly with a single equivalent of nitrite, which binds to the nitrogen center and displaces the proton. For tertiary amines, the process of

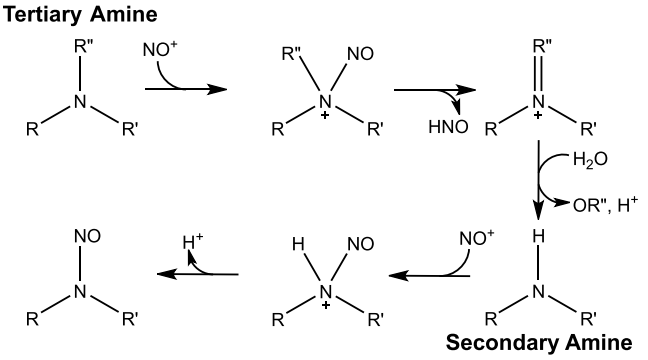


Figure 2. Typical mechanism for *N*-nitrosation. Tertiary amines are first dealkylated by one equivalent of nitrite. From here, nitrosation progresses identically for secondary and tertiary amines, with hydrogen being displaced by a final nitrite equivalent.

nitrosation is more complex and requires two equivalents of nitrite: the first to dealkylate the amine, converting it into a secondary amine, and the second to convert the secondary amine to a nitrosamine, as described above. Most often, the nitrosating agent is the limiting reagent in the formation of unwanted nitrosamines. This is partially why tertiary amines, requiring two equivalents of nitrite, are generally considered to be at a significantly lower risk for nitrosation and are nitrosated over 1000 times more slowly than secondary amines.[37] Multiple other factors will affect to what extent an amine will be nitrosated. These include environmental factors, such as

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

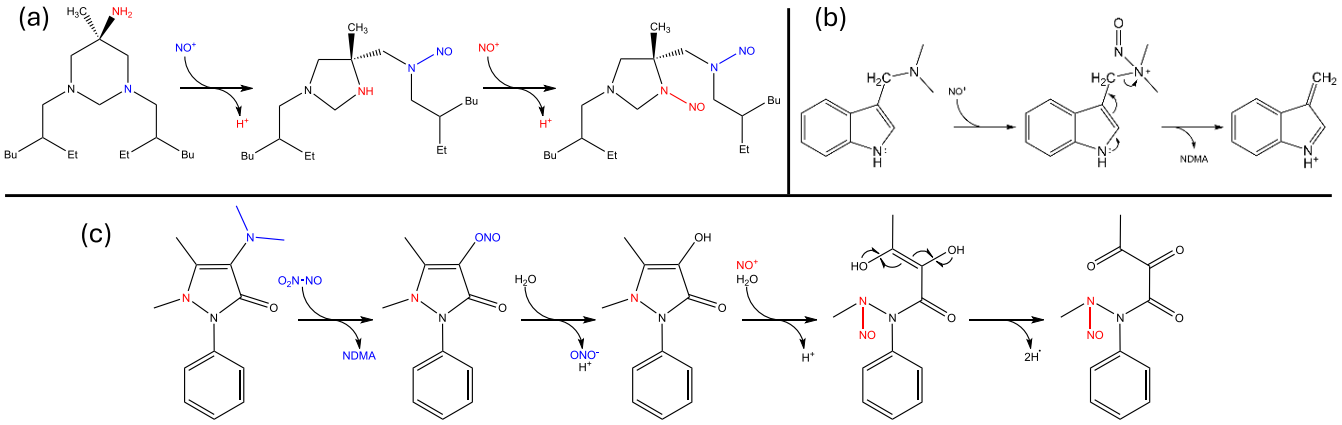


Figure 3. Less common nitrosation mechanisms. The tertiary amine of Hexetidine (a) can be nitrosated in one step (blue), and this leads to a rearrangement that deprotonates the distal primary amine to a secondary (red). The new secondary amine can be nitrosated as well, forming a dually N-nitrosated molecule. Gramine (b) also has a tertiary amine that can be dealkylated and nitrosated by one equivalent of nitrite due to a nearby double bond. Aminopyrine (c) reacts with nitrous anhydride to release NDMA (blue). The leftover nitrite ester is replaced by water. Nitrite and a second water react with the hydrazine (red) and open the pyrazole ring.

pH and temperature, and the makeup of the final drug product, such as water content and particle size.

Within the drug molecule itself, even subtle structural differences can have a drastic effect on the site and extent of nitrosation. A key factor in nitrosation is very often the availability of an amine's lone pair (of electrons) for nucleophilic attack by nitrite. The availability of the lone pair for electrophilic attack is what allows nitrite to bind. An aromatic amine's lone pair tends to be more delocalized and less reactive, and it may even direct nitrosation to another site in the conjugated system, as can happen in aniline-containing molecules.[38] Other structural differences can cause amines to nitrosate using a different mechanism than that discussed above. Most notably, some tertiary amines are able to bypass the typical requirement for two equivalents of nitrite, dealkylating and nitrosating in one step. This most often appears in molecules that have two amines in close proximity. Hexetidine,[39] gramine,[40,41] and aminopyrine[42] are prime examples and their mechanisms are shown in Figure 3. All three can nitrosate a tertiary amine with one nitrite equivalent, taking advantage of the lone pair of another nearby amine, a double bond, or both. Such atypical tertiary amines can sometimes nitrosate more quickly than even secondary amines. Despite being derived from tertiary amines, the production of NDMA from both aminopyrine and gramine is comparable to that of dimethylamine (a secondary amine attached to two methyls), with aminopyrine having the fastest rate of the three.[40]

While varied, the above nitrosation reactions are all initiated by nitrite and are typically catalyzed by an aqueous, acidic environment. However, there are other reagents, conditions, and mechanisms that can also lead to nitrosation. Fremy's salt,[43] peroxynitrite,[44] nitrosyl halides,[45] and nitrosonium salts[46] have all been shown to act as nitrosating agents, although their mechanisms are not all fully understood. Sources for stable *N*-nitrosamines are not limited to secondary and tertiary amines, either. Nitramines ($R_2N\text{-}NO_2$),[47] hydrazines ($R_2N{-}N(R/H)_2$),[48] hydrazones ($R_2N{-}N = R$),[49] and organometallics like Grignard reagents (R-Mg-X)[50−52] are all nitrosatable reaction centers that undergo alternative mechanisms. The potential for nitrosation *in vivo* has also been extensively studied as conditions in the digestive tract can open multiple pathways to nitrosation. The presence of nitrite along with the acidity of the stomach can catalyze nitrosation chemically, but several strains of bacteria found in saliva and the digestive tract possess enzymes that can act as catalysts, particularly at neutral-to-basic pH.[53,54] Ziebarth et al. incubated saliva, drug molecules, and nitrate at pH 7.2 for 24 h and found that almost half of the 23 tested molecules produced detectable levels of *N*-nitrosamines. Additionally, urinary infections have been linked to the presence of nitrosamines in urine,[55] which also correlated with an increased risk of bladder cancer.[56] As with the saliva bacteria, strains isolated from the urine of infected patients were shown to catalyze nitrosation of morpholine.[57]

With all the varying ways that amines in pharmaceuticals can become nitrosated, there was a need for a consistent method to test the susceptibility of amine-containing APIs and impurities. In 1978, the WHO put forward the Nitrosation Assay Procedure (NAP) test, which standardized nitrosation testing conditions.[58] Under the NAP test, the reaction must have a duration of 1−4 h, a pH of 3−4, a temperature of 37 °C, and nitrite and drug concentrations of 40 mM and 10 mM, respectively. These conditions—warm and acidic with an excess of nitrite—are very favorable for *N*-nitrosamine formation, meaning that if a drug is resistant to nitrosation in the NAP test, it is highly unlikely to be nitrosated under more typical conditions (*e.g.*, dry storage at room temperature). Hundreds of drug molecules have been evaluated with the NAP test with the results published in dozens of papers. These results are an invaluable source of data that can be used as the basis for *in silico* tools, (Q)SAR models in particular.

In general, there are two basic types of (Q)SAR models: expert rule-based and statistical.[59] Statistical methods use experimental data to train machine learning, neural networks, or other models to predict a particular endpoint. Expert rule-based models rely on human-written rules derived from available knowledge. An example of this would be the CPCA,[24,25] discussed above. An advantage of expert rule-based models is the ability to include a mechanistic explanation for structural alerts. Statistical models have no built-in mechanistic reasoning. They simply detect patterns in the data but do not rationalize why those patterns are there. Conversely, expert rule-based models use human knowledge as

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

the basis for their rules. That scientific understanding can be encoded into the model itself and included as part of the model's output. The disadvantage of expert rule-based models is that they are often more limited in scope than statistical models due to limits in existing knowledge.

Here, we have constructed both an expert model and a statistical model to take advantage of the strengths of each. By using the models together, expert rules can fill in missing or underrepresented data, and statistical models can catch trends missed by human analysis. The rule-based model assigns integer scores to select molecular features and structures known to affect nitrosation both as activators and deactivators. The 11 deactivating features include high pK_a (>9.5), aromaticity, amides (including sulfonamide, urea, and carbamate groups), steric hindrance near the amine, and the presence of other nitrosatable centers. For tertiary amines, the rule-based model also considers the size of the leaving group as a deactivating feature. Since there can be up to three unique leaving groups per tertiary amine, the model will generate individual scores for each unique nitrosamine product, making it possible to anticipate the exact nitrosamine that will be produced. There are only three activating features: nitro groups (which can cause intermolecular nitrosation), dialkyl aromatic tertiary amines (where conjugation stabilizes alternative reaction pathways), and tertiary amines with a R_2N-CH_2-Ar moiety like gramine, whose mechanism has been discussed above and shown in Figure 3(b). The sum of these scores determines the likelihood of nitrosation. Conversely, the statistical model is built entirely by algorithms interpreting experimental data (*i.e.*, molecular structures and nitrosamine yields). Our statistical models utilize graph neural networks to make a binary classification of whether nitrosation of an amine is expected. Because the learning algorithm is inherently ignorant of chemistry, the data used to train it must be both extensive and curated. We accomplished the latter by choosing data that abided by the NAP test guidelines, letting the model only consider the molecular structure and excluding experimental conditions such as pH, nitrite availability, or presence of catalytic enzymes.

METHODS

Rule-Based Model. *Generating Nitrosamines.* The algorithm generates *N*-nitrosamines by nitrosating secondary and tertiary nitrogen centers in the query compound, such as amines, amides, ureas, carbamates, and sulfonamides. For secondary nitrogen centers, only one *N*-nitrosamine is produced and for tertiary nitrogen centers, the bonds between the nitrogen and its substituents are sequentially broken, converting them to secondary nitrogen before final nitrosation (Figure 4). This can result in up to three nitrosamines, depending on the substituents. Constraints include not breaking bonds if part of a ring or if the connected atom lacks hydrogens, as tertiary amines only undergo *N*-nitrosation if they have α protons.[28,52] Nitrosation of primary and quaternary nitrogen centers were not covered due to the instability of nitrosamines produced by primary amines and the quaternary amines being uncommon in drug compounds and the literature.

Building the Expert Knowledge Rule Base. A comprehensive literature review was conducted to compile existing knowledge on the nitrosation potential of nitrogen centers in organic molecules under various conditions. Eleven deactivating and three activating features were identified; each was assigned a score based on its impact. These features were converted into simple rules to be used for calculating nitrosation likelihood. The rules were designed to be reaction-specific, *i.e.*, they could distinguish between the formation of different nitrosamine from a single center or from multiple nitrogen centers

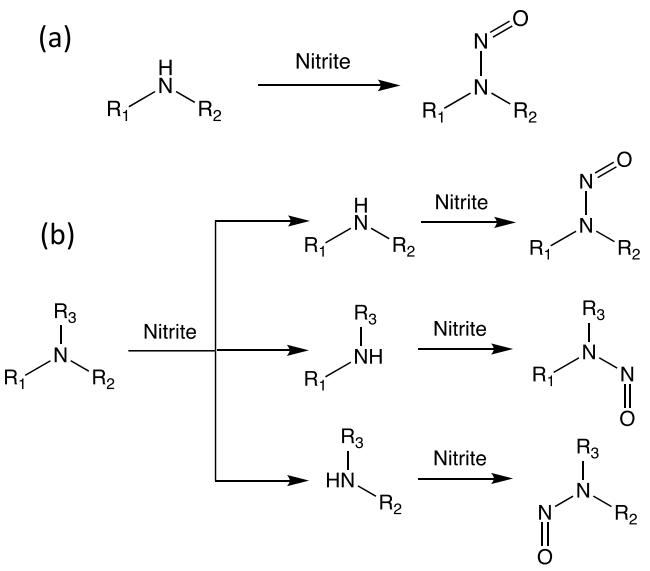


Figure 4. Nitrosation of (a) secondary and (b) tertiary nitrogen centers.

in a single molecule. Double nitrosations were not included for simplicity.

Calculating Likelihood Scores. The likelihood score indicates the probability of a particular *N*-nitrosamine forming from a compound. It is calculated by summing the individual scores of various deactivating and activating features coded in the expert rules. Scores can be positive or negative, with higher scores representing a greater likelihood. If no expert rules apply, a default score of zero is assigned indicating "very likely" formation. This method is similar to the group contribution method, where features collectively impact the final outcome in an additive manner.

Scores are classified into four categories for ease of interpretation:

i. −1 or higher: Very likely
ii. −2 or −3: Likely
iii. −4 or −5: Less likely
iv. −6 or lower: Unlikely

Searching Experimental Reaction Data. Once nitrosamines are generated, the predicted parent-nitrosamine pairs are searched in a nitrosation reaction dataset, which has been built for this study. If successful, a query to this dataset usually returns the following results:

i. Experimentally observed parent-nitrosamine pairs identical to the predicted pair (exact hits)
ii. Nitrosation reactions of compounds with similar structural features around the nitrosatable nitrogen as the query compound (similar hits)

The aim is to check the agreement of the predicted nitrosamines, including their formation likelihoods, with experimental data. This also provides chemists with insights for synthesizing predicted nitrosamines for analysis. Given the limited dataset of 695 reactions, finding exact matches for many nitrosation reactions of NDSRIs is unlikely. Therefore, retrieving similar hits is useful, as compounds with similar structural features around the nitrogen reaction center are expected to behave similarly during nitrosation. This approach offers more robust information than searching only for exact matches.

The query reaction is searched as a whole by generating two distinct fingerprints: one capturing structural features around the amine nitrogen of the reactant and the other around the >N−N=O moiety of the product. Features within three bonds from these centers were considered. Traditional fingerprints, often represented as sparse arrays dominated by zeros with occasional ones, were deemed unsuitable for capturing the nuanced structural details required for this task. Instead, we employed specialized fingerprints, previously developed by us,[60] consisting of 600 continuous values, providing a

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

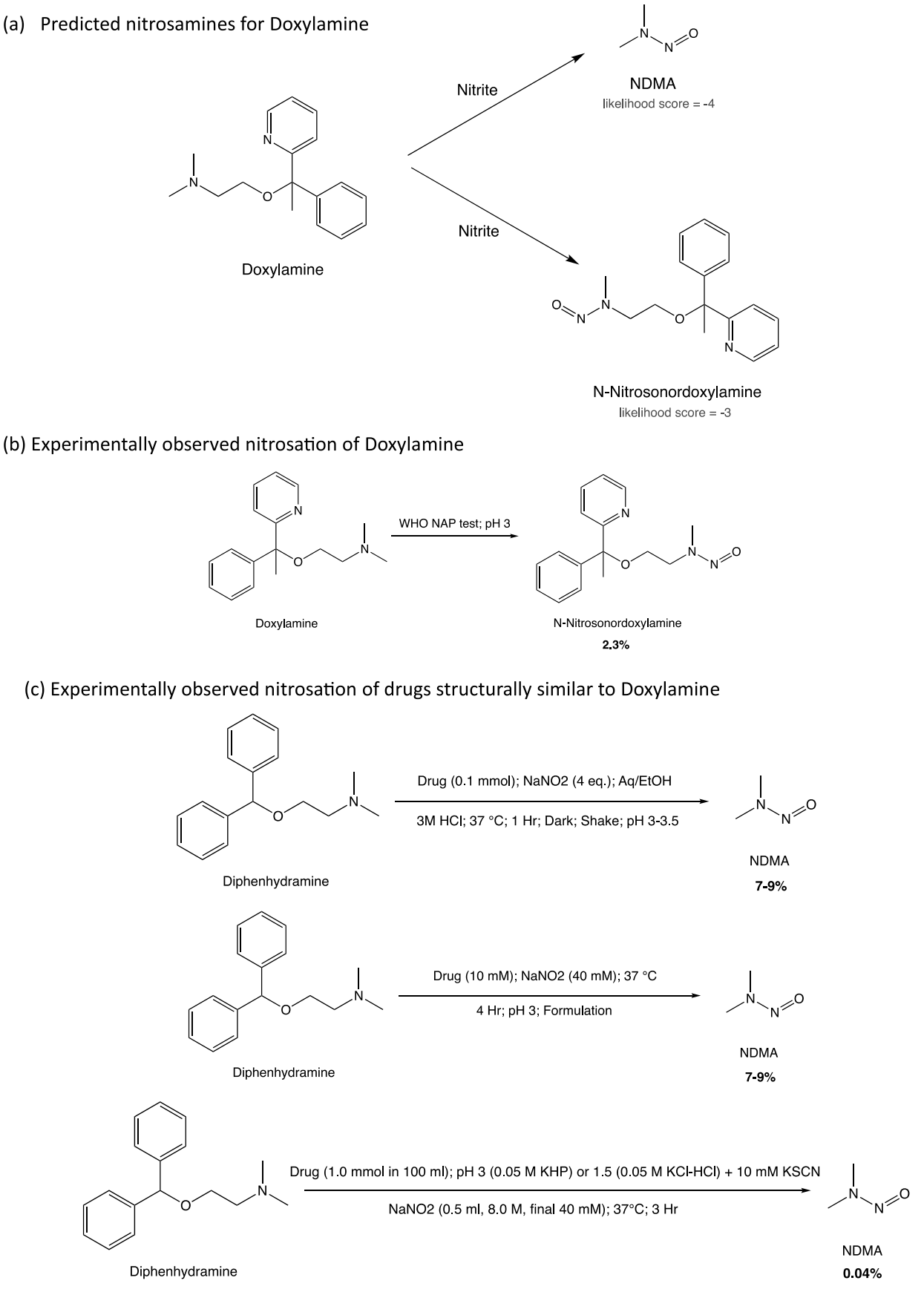


Figure 5. Nitrosation of doxylamine. (a) Predicted nitrosamines. (b) Exact hit received from the experimental nitrosation dataset. (c) Similar hits retrieved from the dataset belonging to the drug diphenhydramine which has similar structural features as doxylamine around the amine reaction center.

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

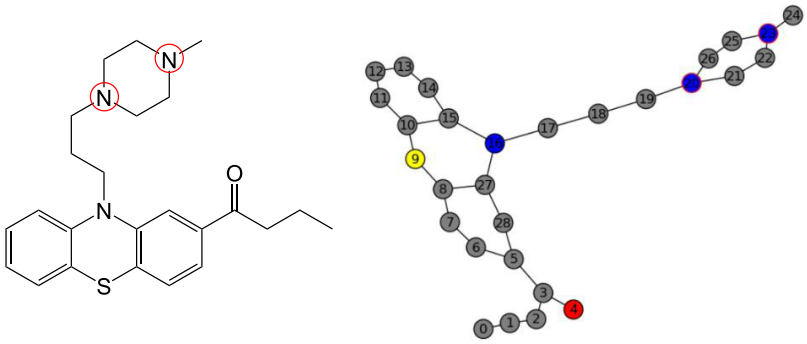


Figure 6. Butaperazine, shown as both a typical 2D chemical structure (left) and as a NetworkX graph (right). Graph neural networks do not consider coordinates or angles, only how the nodes are connected. As such, arrangement of the nodes is trivial and may not abide by conventional practices. Nitrosated centers are outlined in red. (Gray: carbon, Blue: nitrogen, Red: oxygen, Yellow: sulfur).

Table 2. Several Values Tested for Each Hyperparameter[a]

	Batch Size	Hidden Layers	Heads (GAT only)	Learning Rate	Weight Decay	Training Epochs
Tested Values	16, 32, 64	8, 16, 32, 64, 128	1, 2, 4, 8	1E(−4, −3, −2, −1)	0, 5×10^{-4}	5, 10, 25, 50, 75, 100, 150
GCN	16	8	N/A	1×10^{-2}	5×10^{-4}	50
GAT	16	16	4	1×10^{-2}	5×10^{-4}	50

[a]The best-performing values are given for both models

significantly richer informational profile. The transformation associated with the query reaction is then represented by subtracting the product fingerprint from the reactant fingerprint. The similarity is computed between the query and the target reaction using the cosine similarity function.[61]

This is illustrated in Figure 5 with the nitrosation of doxylamine (CAS # 469−21−6). Our method predicted two nitrosamines (Figure 5(a)): NDMA and *N*-nitrosonordoxylamine, with likelihood scores of −4 and −3, respectively. Searching the dataset for exact matches yielded one reaction forming *N*-nitrosonordoxylamine (Figure 5(b)) and none for NDMA. However, searching for similar reactions retrieved three nitrosation reactions of diphenhydramine, all yielding NDMA (Figure 5(c)). Unlike the exact reaction match, performed according to the WHO NAP test, these similar reactions were conducted under different conditions, resulting in NDMA formation. This suggests that NDMA formation is possible from doxylamine.

Statistical Model. To develop a statistical model, we turned to the power of neural networks. Each molecule was represented as a graph. An example of such a graph for Butaperazine is shown in Figure 6. Nodes and edges represented the individual atoms and bonds, respectively. Each node was given four attributes: atomic number, formal charge, hybridization, and implicit hydrogen count. Each edge had only one feature: bond order. These attributes were computed from the SMILES string for each molecule by the RDKit module[62] (SMILES strings were stripped of chirality and any explicit hydrogens beforehand). No coordinates were provided. Graphs were created using NetworkX[63] and then passed to PyTorch[64] to train and test node classification models. One of two classes was assigned to each node representing nitrosation. All atoms that did not have an atomic number of 7 (nitrogen) were automatically assigned to Class 0 (negative) as only *N*-nitrosation was being considered. For nitrogen atoms, if there was evidence of at least 1% nitrosation, the node was assigned to Class 1 (positive). Otherwise, they were also assigned to Class 0.

Two varieties of graph neural networks were tested: Graph Convolution Networks (GCN)[65] and Graph Attention networks (GAT).[66] Both of these are known to generally perform well in node classification. GCNs and GATs are in the same general family of "Convolution Neural Networks", but GATs have the added benefit of a "graph attention" layer. The attention layer allows the network to focus on what it determines to be the most important part of the input for each node, considering both the node's own features and those of its neighbors, and then generates a new set of features.[66] Multihead attention networks will try several of these attention computations simultaneously and aggregate the results, stabilizing the learning process.

Network model performance was tested by random subsampling. This method was chosen due to the small size of the dataset (207 molecules), where withholding 10% or more of the dataset for testing would have shrunk the training set too severely. After randomizing the dataset, 5 molecules were withheld as a test set, using the remaining 202 to train the model. This process was repeated 40 times, each time with a new set of 5 (repeating none), and the results averaged. Cross entropy loss was used as the training criterion. For both GCN and GAT models, several variations of each hyperparameter were tested (Table 2) and the combinations that yielded the best overall accuracy, predictive value, sensitivity, and specificity (eqs 1−4) were used as the final models. Predictive values were calculated for both positive and negative calls.

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{predictive value} = \frac{TP}{TP + FP} \text{ or } \frac{TN}{TN + FN} \quad (2)$$

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (4)$$

TP: true positive, TN: true negative, FP: false positive, FN: false negative

DATA

Rule-Based Model. For the reaction search tool, a dataset of 695 *N*-nitrosation reactions was constructed based on a comprehensive literature review, the sources for which are given in the Supporting Information. Each reaction was annotated with experimental conditions, observed yields, and literature references. The dataset spans diverse structural features and reaction conditions, including both forced nitrosation reactions such as the NAP test and reactions under simulated gastric conditions.

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

Statistical Model. For training and testing of the statistical model, a dataset of 207 unique molecules was built from 30 published articles (Table S1, literature references in Supporting Information). Unlike the reaction dataset, this collection was limited to reactions that generally followed the NAP guidelines: pH of 3−4, reaction duration of 1−4 h, 37 °C, and an excess of nitrite. This limitation was established to keep experimental conditions consistent across the training set. Some reasonable exceptions were allowed: a pH of up to 4.5, room temperature (25 °C) reactions, or a reaction time of up to 6 h. However, more extreme conditions were not permitted, such as a pH of less than 3, reaction temperatures of over 50 °C, or reactions performed *in vitro*. To prevent duplicate molecules in the dataset, only one reaction result for each unique molecule was included. In the case where one molecule had two or more reported nitrosamine yields, the highest yield was used, provided experimental conditions met those stated above. We did not consider the stability of the nitrosamine when using the reported yields, outside of disregarding primary amines. Nitrosamines that are too unstable to be reliably measured *in chemico* are unlikely to be of concern as mutagens *in vivo*. In the case of shorter reactions (1−3 h), the yields were extrapolated to 4 h by assuming a first-order reaction (eq 5) to enable direct comparison between the nitrosation rates of different molecules. The original drug molecule concentration, $[A]_0$, and the reaction time, t, were taken from the provided methods, and the ending drug concentration, $[A]_t$, was calculated from the nitrosamine percent yield. For example, if the percent yield is 30% after 1 h, then $[A]_{t=1h} = 0.7 \times [A]_0$. The rate, k, was calculated in eq 6. This rate was then used in eq 7 to find the expected yield at 4 h, $[A]_{t=4h}$.

$$-kt = \ln\left(\frac{[A]_t}{[A]_0}\right) \tag{5}$$

$$k = -\frac{1}{t}\ln\left(\frac{[A]_t}{[A]_0}\right) \tag{6}$$

$$[A]_t = [A]_0\, e^{-kt} \tag{7}$$

Much of the nitrosation literature included characterization of the nitrosamine product, allowing for easy assignment of the nitrosated center within the parent molecule. In the case where no definitive product was reported but there was only one amine in the reactant molecule, the reaction center was presumed to be that singular amine. When there were multiple amine groups, an online search was done to find known impurities of the drug molecule. If *N*-nitrosamine impurities were documented, these were assumed to be the detected nitrosamines, and the reactive amine was assigned accordingly. Molecules for which no nitrosamine structure could be found, despite a positive NAP test yield, were not included in the final dataset. Additionally, several parent molecules were reported to produce small-molecule nitrosamines (*e.g.*, NDMA). If there was more than one plausible source for the small nitrosamine within the parent molecule, the results were not included in the dataset. In total, the dataset consisted of 207 molecules with 325 secondary and tertiary amines. Of these nitrogen centers, just under one-third had evidence of at least a 1% nitrosamine yield.

RESULTS AND DISCUSSION

Rule-Based Model. *Structural Features Affecting N-Nitrosation.* A comprehensive survey of scientific literature identified a number of structural features influencing *N*-nitrosation, most of which are deactivating. Some sources also provided experimental data including nitrosamine yields, helping us in the assignment of preliminary scores for these features.

Deactivating features—These features inhibit nitrosation as explained below:

1. Tertiary nitrogen (−3): With few exceptions, tertiary amines are reported to be much less reactive toward nitrosation compared to secondary amines. This is due to the need for an additional slow, rate-limiting dealkylation step and an extra mole of the nitrosating agent.[28,38,52,67,68] Consequently, a tertiary nitrogen is considered a deactivating feature and is assigned a score of −3.
2. Tertiary nitrogen with a leaving group larger than ethyl (−1): When a tertiary amine is asymmetrical, bulkier groups separate less readily from the nitrogen during nitrosamine formation.[52] Therefore, we assigned a deactivating score of −1 for nitrosamines formed from tertiary nitrogen due to the cleavage of groups larger than an ethyl moiety.
3. Aromatic dialkyl tertiary amine with a strong electron-withdrawing group in the aromatic ring (−1): While aromatic dialkyl amines are generally more reactive toward nitrosation than trialkyl amines, the presence of a strong electron-withdrawing group in the benzene ring significantly reduces their reactivity.[69] Consequently, we assigned a deactivating score of −1 for this feature.
4. Presence of other nitrosation centers in the molecule (−1): Other functional groups can preferentially react with nitrosating agents, forming *O*- or *C*-nitrosamines or nitrites.[68] These reactions can compete with the formation of *N*-nitrosamines. Therefore, phenolic and aliphatic −OH, primary aromatic amines, and aromatic sulfonamides ($-SO_2NH_2$) were included in this list and assigned a deactivating score of −1.
5. Tertiary amine, Amine center pK_a more than 9.5 (−2): Ashworth et al.[37] reported that tertiary amines with a pK_a above 9.5 do not significantly generate nitrosamines through nitrosation. Therefore, we assigned a deactivating score of −2 to any tertiary amine center with a pK_a over 9.5.
6. Steric hindrance next to N (−1): Steric effects near the nitrogen atom significantly influence the nitrosation of secondary and tertiary amines.[52,70−72] For example, Jones et al.[70] found that the relative rate of nitrosation of 2-methylpiperidine was about 5 times slower than that of piperidine. This reduced rate is due to the decreased accessibility of the nitrogen's lone pair of electrons to the nitrosating agent. Therefore, we assigned a deactivating score of −1 for the presence of any isopropyl or *t*-butyl type sp^3 carbon atom bonded to the nitrogen center.
7. Secondary amines with pK_a more than 9.5 (−2): It has been reported that basic secondary amines with a pK_a greater than 9.5 typically do not produce significant levels of *N*-nitrosamines.[37] This is because strongly basic amines tend to remain in their protonated state even at lower pH, making them less likely to participate in

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

Table 3. Results of the Nitrosation Prediction of 50 Drug Compounds as Part of a Validation Exercise for the Rule-Based Model, Both for the Original and Modified Rules[a]

					Scores and Ranks Using the Original Rules		Scores and Ranks after the Modified Rules	
Query Compound	Nitrosamine #	*Exact Hit* Reaction Yields	*Similar Hit* Reaction Yields	Average Yield	Predicted Likelihood Score	(Predicted Rank, Scaled Observed Rank)	Predicted Likelihood Score	(Predicted Rank, Scaled Observed Rank)
Acetohexamide	1				−3		−6	
	2		0.59, 0, 43	14.53	−4	(5, 5.47)	−4	(7, 7.15)
Acetylspiramycin	1	0.005	0.001, 0.35, 0.003	0.09	−6	(3, 1.00)	−8	(3, 1.00)
	2		17.6	17.6	−7	***(2, 5.61)***	−8	***(3, 7.34)***
	3		17.6	17.6	−5	(4, 5.61)	−6	(5, 7.34)
Ajmaline	1		90, 80	85	−4	***(5, 8.73)***	−1	(10, 11.63)
Aminopyrine	1		85, 91, 84, 79, 80	83.8	−2	(7, 8.59)	1	(12, 11.44)
Amitriptyline	1		1.2, 2, 0.12, 1	1.08	−3	***(6, 3.17)***	−4	***(7, 3.98)***
	2	0.006, 0.06, 0.05	0.3, 0.6, 0.7, 0.01, 0.01, 1.2	0.29	−4	***(5, 1.68)***	−6	***(5, 1.93)***
Bephenium	1							
Betanidine	1		0.001, 90, 87	59	−1	(8, 8.19)	−1	(10, 10.88)
	2		0.001, 90, 87	59	−1	(8, 8.19)	−1	(10, 10.88)
Captodiamine	1	0.7	9, 0, 0.04, 0.3, 0.3	1.72	−4	(5, 3.98)	−6	(5, 5.10)
	2		2.3, 1.7, 1.2, 0.88, 0.23, 0.005	1.09	−3	***(6, 3.31)***	−4	(7, 4.17)
Cefalexin	1		5.5	5.5	−4	(5, 4.80)	−4	(7, 6.22)
Cefradine	1		5.5	5.5	−4	(5, 4.80)	−4	(7, 6.22)
Chlordiazepoxide	1	55, 70, 0.28, 0.01, 57.40		36.54	−4	(5, 7.10)	−6	***(5, 9.39)***
Cinnarizine	1				−5		−7	
	2				−4		−6	
Clomiphene	1		0, 0.01, 2.3, 1.7	1	−3	***(6, 3.03)***	−4	***(7, 3.80)***
	2	0, 0, 0.41, 0.008	0, 0.01, 0.15, 0.4	0.12	−4	***(5, 1.27)***	−6	***(5, 1.37)***
Desipramine	1		2, 1, 0.24, 0.24, 0.02	0.7	−4	***(5, 2.36)***	−6	(5, 2.86)
	2	2.8, 10, 1	3.4, 72, 99, 85.4, 95	46.08	−2	(7, 7.78)	−2	(9, 10.32)
Dimenhydrinate	1		2.3, 1.2, 2.0, 0.12, 1	1.32	−3	***(6, 3.44)***	−4	(7, 4.36)
	2	9, 0, 0.04	0.3, 0.6, 0.07, 0.01, 0.01, 1.2	1.40	−4	(5, 3.58)	−6	(5, 4.54)
Doxylamine	1	2.3	1.2, 2, 0.12, 1	1.32	−3	***(6, 3.44)***	−4	(7, 4.36)
	2		9, 0, 0.04, 0.3, 0.6	1.99	−4	(5, 4.25)	−6	(5, 5.47)
Duloxetine	1	85.4	99, 72, 3.4, 2.8, 10	45.43	−2	(7, 7.64)	−2	(9, 10.14)
Enalapril	1	67.8	0	33.9	−1	(8, 6.69)	−1	(10, 8.83)
Ephedrine	1	68, 30, 0.02	18.8, 9.25, 94, 50.7, 7.45	34.78	−4	(5, 6.83)	−4	(7, 9.02)
Ergometrine	1	0.8		0.8	−5	(4, 2.63)	−6	(5, 3.24)
	2		52.3, 34.3, 0, 14	25.15	−3	(6, 6.42)	−3	(8, 8.46)
	3		79	79	−5	***(4, 8.46)***	−5	***(6, 11.25)***
Ethambutol	1	0.26, 80, 42, 100, 2	76, 15.7, 0.001, 82, 71	46.9	−4	***(5, 7.92)***	−4	***(7, 10.51)***
Ethosuximide	1				−3		−6	
Felodipine	1	45.1		45.1	0	(9, 7.51)	0	(11, 9.95)
Flufenamic acid	1		100, 97, 98, 97, 90	96.4	0	(9, 9.00)	0	(11, 12.00)
Flupentixol	1		0.5, 0.2, 0.06	0.25	−5	***(4, 1.54)***	−7	(4, 1.75)
	2		0.5, 0.7	0.6	−5	(4, 2.08)	−7	(4, 2.49)
Fonazine	1		1−2, 1, 0.24, 0.02	0.42	−4	***(5, 1.81)***	−6	(5, 2.12)
	2				−3		−4	
	3	0.8	6, 0.27, 2, 2, 0.011, 0.011	1.52	−5	(4, 3.71)	−7	(4, 4.73)
	4		5.8, 14, 1.2, 2	5.75	−4	(5, 5.07)	−5	(6, 6.59)
Guanethidine	1	0, 0.015, 63		21	−6	***(3, 6. 15)***	−8	***(3, 8.08)***
Hydrochlorothiazide	1	19.2, 0.001, 8, 64	0.3	18.3	−1	(8, 5.88)	−1	(10, 7.71)
	2				−5		−3	

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

Table 3. continued

Query Compound	Nitrosamine #	*Exact Hit* Reaction Yields	*Similar Hit* Reaction Yields	Average Yield	Scores and Ranks Using the Original Rules: Predicted Likelihood Score	Scores and Ranks Using the Original Rules: (Predicted Rank, Scaled Observed Rank)	Scores and Ranks after the Modified Rules: Predicted Likelihood Score	Scores and Ranks after the Modified Rules: (Predicted Rank, Scaled Observed Rank)
Isoxsuprine	1		68, 30, 0.02, 18.8, 9.25	25.21	−2	(7, 6.56)	−2	(9, 8.64)
Melperone	1		1.8, 2.1	1.95	−6	(3, 4. 12)	−8	(3, 5.29)
Methadone	1	0.18	0.8, 6, 0.27, 2, 0.011	1.54	−5	(4, 3.85)	−7	(4, 4.92)
	2		1.7, 1.2, 0.23, 0.005	0.78	−4	***(5, 2.49)***	−5	(6, 3.05)
Minocycline	1				−7		−8	
	2	11, 34, 0.02, 2	0, 3.70, 3.70, 0.01, 0, 0	5.64	−8	***(1 , 4.93)***	−10	***(1, 6.41)***
	3		85, 90, 89, 87, 75	85.2	−3	***(6, 8.86)***	0	(11, 11.81)
Mirtazapine	1	0.8	0.03, 5.8, 14, 74	18.93	−3	(6, 6.02)	−4	(7, 7.90)
Nialamide	1	0, 0.41		0.2	0	***(9, 1.41)***	0	***(11, 1.56)***
	2				−3		−3	
	3	0, 0.41	100, 100, 100, 100, 66	66.63	−3	***(6, 8.32)***	−3	***(8, 11.07)***
Nortriptyline	1	0.012, 9.8	72, 3.4, 85.4, 99, 2.8	38.92	−2	(7, 7.37)	−2	(9, 9.76)
Oxytetracycline	1				−7		−8	
	2	1.3, 15, 0, 2.70, 2.70, 0.01	0, 3.70, 3.70, 0.01, 0, 0	2.27	−8	***(1, 4.39)***	−10	***(1, 5.66)***
Phenadoxone	1	2.5	0 p, 0, 0.41, 0.01, 9	2.38	−5	(4, 4.53)	−7	(4, 5.85)
Phenylephrine	1		0.008, 10, 23, 19.1, 18.8	14.18	−1	***(8, 5.34)***	−1	***(10, 6.97)***
Pipamperone	1		1.8, 2.1, 0.07, 0.01, 0.01	0.8	−4	***(5, 2.63)***	−6	(5, 3.24)
Pramipexole	1		17.1, 3.4, 72, 2.8, 10	21.06	−4	(5, 6.29)	−4	(7, 8.27)
Prenylamine	1		94, 18.8, 9.25, 48.9, 5.85	35.36	−1	(8, 6.97)	−1	(10, 9.20)
Promethazine	1		2, 1, 0.24	1.08	−4	(5, 3.17)	−6	(5, 3.98)
	2	6, 0.27, 2, 0.011, 0.011	0.8, 0.18, 0.005, 0.012, 0.16,	0.95	−5	(4, 2.90)	−7	(4, 3.61)
	3		17.6, 5.8, 14, 1.2, 2	8.12	−4	(5, 5.20)	−5	(6, 6.78)
Propranolol	1	15.7, 0.001, 82, 71, 94	50.7, 7.45, 0.012, 9.60, 48.9	37.94	−4	(5, 7.24)	−4	(7, 9.58)
Protriptyline	1	3.4	72, 2.8, 10, 1	17.84	0	***(9, 5.75)***	0	***(11, 7.53)***
Quetiapine	1	0.2	0.06, 0.5	0.25	−5	***(4, 1.54)***	−7	(4, 1.75)
Ranitidine	1				0		−2	
	2				−1		−6	
	3				−2		−2	
	4				−2		−2	
Roxithromycin	1		0.001, 0.35, 0.003, 0.005, 0.18	0.11	−6	(3, 1.14)	−8	(3, 1.19)
	2	17.6	1.2, 2, 0.12, 1	4.38	−5	(4, 4.66)	−6	(5, 6.03)
Thioproperazine	1		0.03, 0.8, 0.5,	0.44	−3	***(6, 1.95)***	−4	***(7, 2.31)***
	2	1.2	0.19	0.69	−4	***(5, 2.22)***	−6	(5, 2.68)
	3		2, 1, 0.24, 0.02	0.82	−4	(5, 2.76)	−6	(5, 3.42)
	4				−3		−4	
Varenicline	1	69.3	19.1, 23, 85.4, 99, 2.8	49.77	0	(9, 8.05)	0	(11, 10.69)
Venlafaxine	1		0.3, 0.6, 0.07, 0.01, 0.01, 1.2	0.44	−7	(2, 1.95)	−9	(2, 2.31)
	2		1.2, 2, 0.12, 1	1.08	−6	(3, 3.17)	−7	(4, 3.98)

[a]Bold italic indicates significant disagreement.

nitrosation reactions. Nitrosation is more likely to occur with secondary amines of weak basicity (*e.g.*, morpholine, piperazine, and *N*-methylaniline).[73] Consequently, we assigned a deactivating score of −2 for secondary amines with a pK_a greater than 9.5.

8. Guanidines (−1): According to Mirvish,[73] the nitrosation of guanidines yields nitrosocyanamides and

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

nitrosoureas. Both of these reactions are relatively slow. This is due to the weaker nucleophilic character of the NH moiety.[52] Therefore, we assigned a deactivating score of −1 for guanidines.

9. Amide Nitrogen in a Resonating Ring (−2): According to Mirvish,[74] forming NO-derivatives of amides is challenging when the nitrogen atom is part of a resonating ring system. This difficulty is likely due to the delocalization of the nitrogen's lone pair of electrons into the ring structure, reducing their availability for reaction. Consequently, we assigned a deactivating score of −2 for such amides.
10. Amides, Ureas, Carbamates, and Related Compounds (−3): The reactivity of amides, ureas, carbamates, and heteroamides is generally lower compared to secondary amines due to the weaker nucleophilic character of the NH moiety.[52] Consequently, we assigned a deactivating score of −3 for these compounds.
11. Sulfonamides (−4): Following the same logic as the amides, the sulfonamides were also assigned a deactivating score of −4.

Activating features—Presence of these features enhances nitrosation:

1. Tertiary amines bearing a $-CH_2-$ aryl moiety (+2): Although tertiary amines are typically less reactive than secondary amines, some can rapidly form *N*-nitroso compounds. Tertiary amines with an electron-rich benzylic group as an alkyl group are particularly prone to nitrosative dealkylation.[52] Specifically, dimethyl tertiary amines with a $-CH_2-$ aryl moiety (*e.g.*, benzyl, furfuryl) have been identified as having a high potential for NDMA formation.[37] Consequently, we assigned an activating score of +2 for these compounds.
2. Tertiary amine, aromatic dialkyl (+1): Aromatic dialkyl amines are significantly more reactive toward nitrosation than trialkyl amines.[36,69,75] These reactions produce arylalkyl nitrosamines. Therefore, we assigned an activating score of +1 for these compounds.
3. Tertiary Amines Containing a Nitro Group (+2): Compounds with both a tertiary amine moiety and a nitro group can form nitrosamines through an intermolecular reaction, where the nitro group of one molecule nitrosates the tertiary amine of another.[22,76] This mechanism, responsible for NDMA formation in some ranitidine products, occurs without external nitrosating agents. Therefore, we assigned an activating score of +2 for these compounds.

Assessing Predictive Performance and Enhancing the Expert Knowledge Rules. The method was validated by comparing predicted *N*-nitrosamines with experimental data. Subsequently, potential enhancements to the expert rules were identified and implemented.

Fifty drug compounds were randomly selected for this purpose, from which 83 nitrosamines were predicted. A search in the nitrosation reaction dataset was performed to identify exact or similar hits from experimentally observed chemical reactions. A manual inspection was performed to exclude reactions from the "similar hits" set that did not accurately represent the query chemical or lacked clearly specified yields. The rationale behind these exclusions is given in Table S2. Average yields from these reactions were used as observed experimental yields. Thirteen predicted nitrosamines from five drugs lacked experimental data, and the drug Bephenium did not have any predicted nitrosamines due to the absence of any secondary or tertiary nitrogen, leaving us with data for 69 nitrosamines (Supporting File 1).

We anticipated a correlation between the calculated likelihood scores and the experimental yields. Indeed, higher yields generally corresponded, though weakly, with higher likelihood scores (Figure S1(a)). However, due to the categorical nature of the likelihood scores and the continuous nature of the experimental yields, a correlation coefficient might not accurately capture their relationship. Therefore, we developed a measure to assess the correspondence between two different rankings of the nitrosamines: one based on the calculated likelihood scores and the other using the experimental yields. The more similar the two rankings, the higher the prediction accuracy of the expert rules was assumed. To achieve this, first, the experimental yields were ranked in ascending order, starting from 1, where lower yields received lower ranks and higher yields received higher ranks. Identical yields were assigned the same rank. This ranking is referred to as the *Observed Rank*, ranging from 1 to 60 across the 69 reactions used for validation. Similarly, the calculated nitrosation likelihood scores were ranked in ascending order, referred to as the *Predicted Rank*. Next, the *Observed Ranks* were scaled to match the range defined by the *Predicted Ranks* (resulting in the *Scaled Observed Rank*). A *Scaled Rank Difference* was then calculated using the formula

$$\text{Scaled Rank Difference} = \frac{|\,\text{Scaled Obs. Rank} - \text{Predicted Rank}\,|}{\text{Max. Observed Rank}}$$

If the *Scaled Rank Difference* is 0.25 or greater, the predicted score is considered to have significantly deviated from the experimentally observed yields. This cutoff balances the occurrence of too many discrepancies or too few. The predicted and observed ranks for each nitrosamine are given below in Table 3. The details of the rank calculations are also provided as part of Supporting Table S2.

We found the calculated scores for 27 out of 69 nitrosamines differed significantly from the experimental yields (with scaled rank difference of 0.25 or more), resulting in an accuracy of 61%. Examining the parent drugs of these nitrosamines revealed patterns and possible explanations for the discrepancies, leading to changes in the expert rules to improve predictive performance. We avoided making changes for every disagreement to prevent artificial overfitting. Notably, scores for three specific features were modified, while only one new rule was added:

1. For tertiary amines with a $-CH_2-$ aryl moiety (activating feature #1), NDMA formation is favored through C−N bond cleavage, while other bond cleavages are typically not observed. Consequently, these other cleavages were assigned a high deactivating score of −8. For example, in Gramine, NDMA formation from the tertiary nitrogen on the side chain is very likely, whereas des-methyl *N*-nitrosamine is not (Figure 7).
2. Based on multiple instances among the 27 nitrosamines that showed significant discrepancies, we found the deactivation score of −3 insufficient (deactivating feature #1). Therefore, we increased the degree of deactivation by lowering the score from −3 to −4.

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

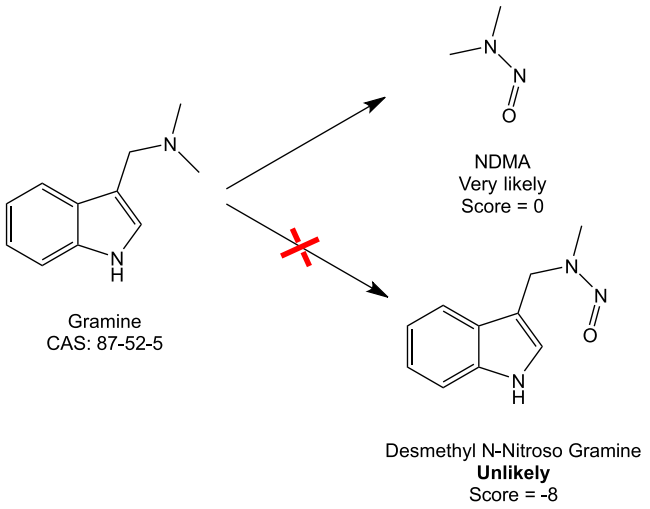


Figure 7. Nitrosation of gramine favors the formation of NDMA over desmethyl N-nitroso gramine.

3. The original expert rules gave a deactivating score of −4 to the secondary nitrogen of sulfonamides ($-SO_2-NH-$) (deactivating feature #11). However, experimental data showed that nitrosation of the −NH− group in sulfonamides is easier than initially thought. To remain conservative, we raised the score from −4 to −2.
4. We introduced a new deactivating feature for secondary −NH− groups flanked by two carbonyl groups (CO−NH−CO), two sulfonyl groups ($SO_2-NH-SO_2$), or one of each (CO-NH−SO_2). This feature was assigned a deactivating score of −5.

We applied the four modifications to the expert rules and reran the predictions. This significantly improved the alignment between the calculated likelihood scores and experimental yields, reducing the significant discrepancies to 16 nitrosamines and increasing the accuracy to 77%. How these changes affected each nitrosamine is detailed in Table 3. Figure S1(b) shows the elimination of several outliers after their scores were recalculated with the new rules. They were almost all assigned higher scores, more than doubling the correlation coefficient.

Statistical Model. With ideal hyperparameters, the GCN model accuracy was ∼80%, with the predictive value of positive and negative calls also being ∼80%. However, the sensitivity was <30% while the specificity was >96%. Testing across different parameters caused only minimal changes in accuracy, with the lowest accuracy being ∼74%. The sensitivity and predictive value of positive calls were most responsive to changing hyperparameters. The positive predictive value could climb as high as 86%, however, sensitivity in these cases dropped to under 20%. This means that the model was achieving a higher positive predictive value by being overly selective when calling a positive. GAT models performed a bit better, with up to 85% accuracy and a predictive value of ∼83% for positive and negative calls. Sensitivity was also improved to 55% (specificity remained the same as the GCN model at ∼95%). GAT models tended to be much less affected by changing parameters, with accuracy rarely dipping below 80% or sensitivity below 40%, meaning that the worst GAT models were still performing better than the best GCN model. Altering the number of hidden layers, attention heads, or batch size had negligible effect on the model. In both GCN and GAT models, optimizing the learning rate and number of training epochs were the key to maximizing model performance.

Using the parameters in Table 2, GCN and GAT models were able to correctly classify 76% and 84% of all amines, respectively. The predictive value of positive and negative calls was nearly equal to the overall accuracy in both models. Both models suffered from a low recall of nitrosatable amines, with the GCN model recalling only 1 in 4 positives and the GAT model doing a little better than 1 in 2 (Table 4). Specificity was nearly 100% in both models. These results point to the model prioritizing negative calls and being too conservative when making positive calls. Undoubtedly, this is due to the dataset used to train the model. Within the 207 molecules, there were 143 secondary amines and 182 tertiary amines. Of these, only 110−just under a third−were recorded as nitrosatable. There were also 93 other amines, primarily consisting of sp^2-hybridized nitrogens, none of which were nitrosatable. This biased the dataset toward calling nitrogen atoms as un-nitrosatable.

An attempt was made to balance the dataset by duplicating molecules that contained positive nodes, as well as by splitting the negative molecules between three models and having those models vote on the final result. In the model containing duplicate positive molecules, duplication was only done after the dataset was split into training and test sets to prevent a test molecule from appearing in the training set. This process was also repeated to triple the number of positive molecules. In the case of multiple-model voting, two methods were tried: a simple majority vote and allowing a positive vote from any model to result in a positive call overall ("Positive Overrules"). These results are given in Table 4. In all four cases, the sensitivity was increased to between 60 and 72%. However, it also resulted in a proportional decrease in the positive predictive value. The "Positive Overrules" model had the highest sensitivity at 72%, but it also had a positive predictive value of only 53%. With the current dataset, it does not seem possible to further improve the sensitivity rate without severely

Table 4. Statistics for the Best-Performing Models Shown in the First Two Rows[a]

Method	Note	Accuracy (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	Sensitivity (%)	Specificity (%)
GCN		76	79	77	26	96
GAT		84	83	85	55	95
GAT	duplicate positives	81	74	86	61	89
GAT	triplicate positives	79	63	88	70	82
GAT	majority vote	81	62	88	66	86
GAT	positive overrules	76	53	89	72	77

[a]Multiple modifications of the data were attempted in order to improve the models' sensitivity. However, all attempts either had little effect or resulted in a drastic decrease in the models' precision.

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

impacting precision. The only way to improve the model will likely be to add new nitrosatable molecules to the dataset.

Comparison with the Expert Rules Model. The best model (GAT) was tested against the same 50 molecules used to evaluate the expert rule-based model. These 50 parent molecules contained a total of 81 amines. The rule-based model made predictions for 70 of the 81, producing 83 unique nitrosamines. The results from the GAT model are recorded in Table S3. Predictions are only given for secondary and tertiary amines, as these are the amine types that the expert model is designed to predict, as well as the quaternary amine in Bephenium. The training set for the statistical model did include other nitrogen types, however, only secondary and tertiary amines showed any evidence of nitrosation within this dataset. Because of this, all other amine types were predicted to be negative by the statistical model. If the expert model produced more than one score for a tertiary amine due to multiple potential nitrosamine products, the highest score was used for comparison. Of the 50 parent molecules, 37 appear in the dataset with a total of 58 secondary and tertiary amines between them. In these cases, the "true class" and NAP test percent yield are also given in Table S3. Withholding a random combination of five compounds from the training set at a time for testing, the statistical model correctly predicts 45 of the secondary and tertiary amines (13 positives and 32 negatives), yielding an accuracy of 78%. The model missed 11 positive amines (54% sensitivity) and only 2 negatives (94% specificity).

Compared to the expert model scores in Table 5, the statistical model only called one positive for an amine with a score of −5 or less ("less likely" to "unlikely"). For expert scores of −3 and above (classified as "likely" or "very likely"), the statistical model called 40% of amines as positive. This would appear to be a significant drop in sensitivity compared to the cross-validated results (55%). However, the drop is easily explained by the opposing biases of the two models. The statistical model is far more likely to call a false negative than a false positive, with a sensitivity of 55% compared to a specificity of 95%. Conversely, the rule-based model is more likely to assign a high score to an amine with little-to-no nitrosamine yield than a low score to an amine with a high yield. These opposing biases make the models complementary, as when they agree, the call is more likely to be correct.

Table 5. Count of Positive and Negative Predictions by the Statistical Model for Each Expert-Rule Score

Expert Model	Statistical Model		
Score	Positive Calls	Negative Calls	Total
2	0	1	1
1	0	1	1
0	2	4	6
−1	2	5	7
−2	6	0	6
−3	0	4	4
−4	5	12	17
−5	0	4	4
−6	0	13	13
−7	0	6	6
−8	1	4	5

Model Limitations. *Both Models.*

1. They do not account for any degradation products of the query amine, which may undergo further nitrosation.
2. They only address *N*-nitrosation and do not cover *C*-nitrosation or *O*-nitrosation.

Rule-Based Model.

1. The rules assume generic nitrosation conditions (pH 3−4 and excess nitrites) and may not be effective under synthesis conditions that are very different.
2. The rules do not cover all nitrosation mechanisms, particularly for tertiary amines.

Statistical Model.

1. The dataset is limited to general NAP conditions: excess of nitrite, pH 3−4.5, and ∼25−37 °C
2. The knowledge of the model is limited to a relatively small dataset (207 molecules), which likely does not sufficiently cover less common functional groups.
3. The dataset does not include nitrosatable amines that are not secondary or tertiary.

CONCLUSIONS

While they have been suspected to be carcinogenic for decades, the discovery of nitrosamines in pharmaceuticals over the past several years has created an immediate need to both prevent and detect nitrosation of amine-containing drug molecules. Numerous experimental methods have been developed to test for nitrosamines during and after the synthetic process, but prevention is still the ultimate goal. Taking advantage of the wealth of nitrosation data generated over the past 50 years, we have built two (Q)SAR models to predict nitrosation based on a molecule's structure.

The expert rule model consists of 12 deactivating features and 3 activating features, all based on an in-depth literature review. The final version of the model assigns one of four categories, an ability to be nitrosated as "very likely", "unlikely", or somewhere in between with an accuracy of 77%. The second model was statistical, training graph neural networks to predict which amines would be nitrosated based on a dataset of 207 molecules. The Graph Attention networks models were shown to perform the best, with an accuracy of 84% and almost 100% specificity while the sensitivity figures were not so impressive. By using both models together, we can increase the confidence in the predictions being made. The expert rules have a strong, literature-backed rationale behind them while the statistical model is able to detect subtle patterns that may escape human notice. Additionally, the expert model is biased toward making positive calls while the statistical model leans toward negative. Assessing amines with both models provides a better-informed decision on their susceptibility to nitrosation and can find its use, for example, in designing synthetic routes with minimal risks of nitrosamines.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.4c00435.

Graphs of the correlation between experimental yields and rule-based model scores before and after optimization and dataset literature references (PDF)

Sample output from the rule-based nitrosation tool (PDF)

Statistical model dataset (XLSX)

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447

Rankings based on expert rules and a list of accepted and excluded nitrosation reactions for the validation set compounds (XLSX)

Comparison of the results for the 50-molecule validation set from both models (XLSX)

AUTHOR INFORMATION

Corresponding Author

Krystle Reiss − *MultiCASE Inc., Mayfield Heights, Ohio 44124, United States;* orcid.org/0000-0003-0997-7953; Email: krystle@multicase.com

Authors

Roustem Saiakhov − *MultiCASE Inc., Mayfield Heights, Ohio 44124, United States*

Suman Chakravarti − *MultiCASE Inc., Mayfield Heights, Ohio 44124, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.4c00435

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. CRediT: **Krystle Reiss** conceptualization, formal analysis, investigation, methodology, validation, visualization, writing - original draft; **Roustem Saiakhov** conceptualization, data curation, methodology, supervision, writing - review & editing; **Suman Chakravarti** conceptualization, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, writing - original draft, writing - review & editing.

Notes

The authors declare the following competing financial interest(s): The rule-based model has been incorporated into the software package, QSAR Flex, which is created and sold by MultiCASE, Inc. All authors are employed by MultiCASE.

ACKNOWLEDGMENTS

The authors would like to thank Mounika Girireddy for her assistance in creating the reaction dataset used by the expert rule-based model.

REFERENCES

(1) Magee, P.; Barnes, J. The Production of Malignant Primary Hepatic Tumours in the Rat by Feeding Dimethylnitrosmine. *Br. J. Cancer* **1956**, *10*, 114−122.

(2) Lijinsky, W.; Taylor, H. The Effect of Substituents on the Carcinogenicity of N-Nitrosopyrrolidine in Sprague-Dawley Rats. *Cancer Res.* **1976**, *36* (6), 1988−1990.

(3) Klopman, G.; Frierson, M. R.; Dauchot, P. Carcinogenicity of N-nitrosamines Derived from Reaction of Sodium Nitroprusside with Some Therapeutically Important Secondary Amines. *Anesthesiology* **1987**, *67* (2), 279−280.

(4) Guttenplan, J. B. "N-Nitrosamines" bacterial mutagenesis and in vitro metabolism. *Mutat. Res., Rev. Genet. Toxicol.* **1987**, *186*, 81−134.

(5) Loeppky, R. N. Advances and Perspectives. In *Nitrosamine and N-Nitroso Compound Chemistry and Biochemistry*; Loeppky, R. N.; Michejda, C. J., Eds.; American Chemical Society: Washington, DC, 1993; Vol. *553*, pp 1−18.

(6) Lijinsky, W. Structure-activity relations in carcinogenesis by N-nitroso compounds. *Cancer Metastasis Rev.* **1987**, *6* (3), 301−356.

(7) European Medicines Agency. EMA Reviewing Medicines Containing Valsartan from Zhejiang Huahai Following Detection of an Impurity: Some Valsartan Medicines Being Recalled across the EU 2018. https://www.ema.europa.eu/en/documents/press-release/ema-reviewing-medicines-containing-valsartan-zhejiang-huahai-following-detection-impurity_en.pdf. [Accessed July 22, 2024].

(8) U.S. Food and Drug Administration. FDA Provides Update on Its Ongoing Investigation into ARB Drug Products; Reports on Finding of a New Nitrosamine Impurity in Certain Lots of Losartan and Product Recall 2019. https://www.prnewswire.com/news-releases/fda-provides-update-on-its-ongoing-investigation-into-arb-drug-products-reports-on-finding-of-a-new-nitrosamine-impurity-in-certain-lots-of-losartan-and-product-recall-300804982.html. [Accessed July 22, 2024].

(9) European Medicines Agency. Suspension of Ranitidine Medicines in the EU 2020. https://www.ema.europa.eu/en/documents/press-release/suspension-ranitidine-medicines-eu_en.pdf. [Accessed July 22, 2024].

(10) U.S. Food and Drug Administration. FDA Requests Removal of All Ranitidine Products (Zantac) from the Market 2020. https://www.fda.gov/news-events/press-announcements/fda-requests-removal-all-ranitidine-products-zantac-market. [Accessed July 22, 2024].

(11) European Medicines Agency. EMA update on metformin diabetes medicines 2019 https://www.ema.europa.eu/en/documents/press-release/ema-update-metformin-diabetes-medicines_en.pdf. [Accessed July 22, 2024].

(12) U.S. Food and Drug Administration. FDA Alerts Patients and Health Care Professionals to Nitrosamine Impurity Findings in Certain Metformin Extended-Release Products 2020. https://www.fda.gov/news-events/press-announcements/fda-alerts-patients-and-health-care-professionals-nitrosamine-impurity-findings-certain-metformin. [Accessed July 22, 2024].

(13) U.S. Food and Drug Administration. Mylan Initiates Voluntary Nationwide Recall of Three Lots of Nizatidine Capsules, USP, Due to the Detection of Trace Amounts of NDMA (N-Nitrosodimethylamine) Impurity Found in the Active Pharmaceutical Ingredient Manufactured by Solara Active Pharma 2020. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/mylan-initiates-voluntary-nationwide-recall-three-lots-nizatidine-capsules-usp-due-detection-trace. [Accessed July 22, 2024].

(14) European Medicines Agency. CHAMPIX (varenicline)—Lots to be Recalled Due to Presence of Impurity N-nitroso-Varenicline Above the Pfizer Acceptable Daily Intake Limit 2021. https://www.ema.europa.eu/en/documents/dhpc/direct-healthcare-professional-communication-dhpc-champix-varenicline-lots-be-recalled-due-presence-impurity-n-nitroso-varenicline-above-pfizer-acceptable-daily-intake-limit_en.pdf. [Accessed July 22, 2024].

(15) U.S. Food and Drug Administration. Pfizer Expands Voluntary Nationwide Recall to include All Lots of CHANTIX (Varenicline) Tablets Due to N-Nitroso Varenicline Content 2021. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/pfizer-expands-voluntary-nationwide-recall-include-all-lots-chantixr-varenicline-tablets-due-n. [Accessed July 22, 2024].

(16) U.S. Food and Drug Administration. Lupin Pharmaceuticals, Inc. Issues Voluntarily Nationwide Recall of All Irbesartan Tablets and Irbesartan and Hydrochlorothiazide Tablets Due to Potential Presence of N-nitrosoirbesartan Impurity 2021. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/lupin-pharmaceuticals-inc-issues-voluntarily-nationwide-recall-all-irbesartan-tablets-and-irbesartan. [Accessed July 22, 2024].

(17) Health Canada. Pfizer Recalls Inderal-LA (Propranolol Hydrochloride) Capsules Due to a Nitrosamine Impurity 2022. https://recalls-rappels.canada.ca/en/alert-recall/pfizer-recalls-inderal-propranolol-hydrochloride-capsules-due-nitrosamine-impurity. [Accessed July 22, 2024].

(18) U.S. Food and Drug Administration. Sandoz, Inc. Issues Nationwide Recall of 13 Lots of Orphenadrine Citrate 100 mg Extended Release Tablets Due to Presence of a Nitrosamine Impurity 2022. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/sandoz-inc-issues-nationwide-recall-13-lots-orphenadrine-

citrate-100-mg-extended-release-tablets-due. [Accessed July 22, 2024].
(19) U.S. Food and Drug Administration. Pfizer Voluntary Nationwide Recall of Lots of ACCUPRIL (Quinapril HCl) Due to N-Nitroso-Quinapril Content 2022 https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/pfizer-voluntary-nationwide-recall-lots-accuprilr-quinapril-hcl-due-n-nitroso-quinapril-content. [Accessed July 22, 2024].
(20) U.S. Food and Drug Administration. FDA Works to Avoid Shortage of Sitagliptin Following Detection of Nitrosamine Impurity 2022. https://www.fda.gov/drugs/drug-safety-and-availability/fda-works-avoid-shortage-sitagliptin-following-detection-nitrosamine-impurity. [Accessed July 25, 2024].
(21) Medicines for Europe. NIOG Meeting 2022. https://www.ema.europa.eu/en/documents/presentation/industry-presentation-2-fourth-meeting-nitrosamine-implementation-oversight-group-niog-and-pharmaceutical-industry_en.pdf. [Accessed 18 July 2024].
(22) Schlingemann, J.; Burns, M. J.; Ponting, D. J.; Martins Avila, C.; Romero, N. E.; Jaywant, M. A.; Smith, G. F.; Ashworth, I. W.; Simon, S.; Saal, C.; Wilk, A. The Landscape of Potential Small and Drug Substance Related Nitrosamines in Pharmaceuticals. *J. Pharm. Sci.* **2023**, *112* (5), 1287−1304.
(23) European Medicines Agency. Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk M7 (R2) 2023. https://database.ich.org/sites/default/files/ICH_M7%28R2%29_Guideline_Step4_2023_0216_0.pdf. [Accessed July 22, 2024].
(24) Center for Drug Evaluation and Research. Recommended Acceptable Intake Limits for Nitrosamine Drug Substance-Related Impurities (NDSRIs), U.S. Food and Drug Administration 2023. https://www.fda.gov/media/170794/download. [Accessed July 23, 2024].
(25) Kruhlak, N. L.; Schmidt, M.; Froetschl, R.; Graber, S.; Haas, B.; Horne, I.; Horne, S.; King, S. T.; Koval, I. A.; Kumaran, G.; Langenkamp, A.; McGovern, T. J.; Peryea, T.; Sanh, A.; Ferreira, A. S.; van Aerts, L.; Vespa, A.; Whomsley, R. Determining recommended acceptable intake limits for N-nitrosamine impurities in pharmaceuticals: Development and application of the Carcinogenic Potency Categorization Approach (CPCA). *Regul. Toxicol. Pharmacol.* **2024**, *150*, No. 105640.
(26) Rao, G. S.; Krishna, G. Drug-Nitrite Interactions: Formation of N-Nitroso, C-Nitroso, and Nitro Compounds from Sodium Nitrite and Various Drugs under Physiological Conditions. *J. Pharm. Sci.* **1975**, *64* (9), 1579−1581.
(27) Schmidtsdorff, S.; Neumann, J.; Schmidt, A. H.; Parr, M. K. Risk assessment for nitrosated pharmaceuticals: A future perspective in drug development. *Arch. Pharm.* **2022**, *355* (4), No. e2100435.
(28) Gillatt, P.; Hart, R.; Walters, C.; Reed, P. Susceptibilities of Drugs to Nitrosation Under Standardized Chemical Conditions. *Food Chem. Toxicol.* **1984**, *22* (4), 269−274.
(29) Merative, Micromedex, "Desipramine (Oral Route)," Mayo Clinic 2024 https://www.mayoclinic.org/drugs-supplements/desipramine-oral-route/proper-use/drg-20071955. [Accessed September 27, 2024].
(30) Merative, Micromedex, "Ergotamine (Oral Route)," Mayo Clinic 2024. https://www.mayoclinic.org/drugs-supplements/ergotamine-oral-route/proper-use/drg-20526910. [Accessed September 27, 2024].
(31) Mayo Clinic. Indomethacin (Oral Route) 2024. https://www.mayoclinic.org/drugs-supplements/indomethacin-oral-route/proper-use/drg-20069700. [Accessed September 26, 2024].
(32) Mayo Clinic. Prazosin (Oral Route) 2024. https://www.mayoclinic.org/drugs-supplements/prazosin-oral-route/description/drg-20065617#drug-proper-use. [Accessed December 19, 2024].
(33) Merative, Micromedex, "Phenothiazine (Oral Route, Parenteral Route, Rectal Route) 2024. https://www.mayoclinic.org/drugs-supplements/phenothiazine-oral-route-parenteral-route-rectal-route/proper-use/drg-20070394.[Accessed August 27, 2024].
(34) Sakai, A.; Inoue, T.; Tanimura, A. Formation of volatile nitrosamines by drug-nitrite interactions under physiological conditions. *GANN Jpn. J. Cancer Res.* **1984**, *75* (3), 245−252.
(35) National Institute of Health. HISTEX Syrup 2024 https://dailymed.nlm.nih.gov/dailymed/fda/fdaDrugXsl.cfm?setid=cc466d33-cf3e-4f1c-a794-6d3d79903adf&type=display. [Accessed September 26, 2024].
(36) Williams, D. Nitrosation at Nitrogen Centers. In *Nitrosation Reactions and the Chemistry of Nitric Oxide*; Elsevier: Amsterdam, 2004; pp 35−56.
(37) Ashworth, I. W.; Dirat, O.; Teasdale, A.; Whiting, M. Potential for the Formation of N-Nitrosamines during the Manufacture of Active Pharmaceutical Ingredients: An Assessment of the Risk Posed by Trace Nitrite in Water. *Org. Process Res. Dev.* **2020**, *24* (9), 1629−1646.
(38) Ashworth, I. W.; Curran, T.; Dirat, O.; Zheng, J.; Whiting, M.; Lee, D. A Consideration of the Extent That Tertiary Amines Can Form N-Nitroso Dialkylamines in Pharmaceutical Products. *Org. Process Res. Dev.* **2023**, *27* (10), 1714−1718.
(39) Bae, J.-Y.; Mende, P.; Shevlin, G.; Spiegelhalder, B.; Preussmann, R.; Loeppky, R. The Nitrosation of Hexetidine and Hexedine: Characterization of the Major Nitrosamine from Common Antimicrobial Agents. *Chem. Res. Toxicol.* **1994**, *7* (6), 868−876.
(40) Mangino, M. M.; Scanlan, A. Nitrosation of the Alkaloids Hordenine and Gramine, Potential Precursors of N-Nitrosodimethylamine in Barley Malt. *J. Agric. Food Chem.* **1985**, *33* (4), 699−705.
(41) Ahmad, M.; Libbey, L.; Scanlan, R. Identification of Minor Nitrosation Products of the Alkaloid Gramine by Mass Spectroscopy. *Food Chem. Toxicol.* **1987**, *25* (11), 867−870.
(42) Mirvish, S.; Gold, B.; Eagen, M.; Arnold, S. Kinetics of the Nitrosation of Aminopyrine to give Dimethylnitrosamine. *Z. Krebsforsch. Klin. Onkol.* **1974**, *82*, 259−268.
(43) Castedo, L.; Riguera, R.; Vázquez, M. P. Fremy's salt (potassium nitrosodisulphonate): a nitrosating reagent for amines. *J. Chem. Soc., Chem. Commun.* **1983**, 301−302.
(44) Masuda, M.; Mower, H. F.; Pignatelli, B.; Celan, I.; Friesen, M. D.; Nishino, H.; Ohshima, H. Formation of N-Nitrosamines and N-Nitramines by the Reaction of Secondary Amines with Peroxynitrite and Other Reactive Nitrogen Species: Comparison with Nitrotyrosine Formation. *Chem. Res. Toxicol.* **2000**, *13* (4), 301−308.
(45) Miyahara, M.; Kamiya, S.; Nakadate, M. Nitrosation of 1, 3-Diarylureas with Nitrosyl Chloride, Dinitrogen Trioxide and Dinitrogen Tetroxide in Dimethylformamide. *Chem. Pharm. Bull.* **1983**, *31* (1), 41−44.
(46) Sather, A. C.; Berryman, O. B.; Ajami, D.; Rebek, J. Reactivity of N-nitrosoamides in confined spaces. *Tetrahedron Lett.* **2011**, *52* (17), 2100−2103.
(47) Kamiya, S. A Convenient Synthesis of 1-(2-Halogenoethyl)-3-nitro-1-nitrosoguanidines. *Chem. Pharm. Bull.* **1972**, *20* (11), 2497−2500.
(48) Lunn, G.; Sansone, E. B. Oxidation of 1,1-dimethylhydrazine (UDMH) in aqueous solution with air and hydrogen peroxide. *Chemosphere* **1994**, *29* (7), 1577−1590.
(49) Enders, D.; Eichenauer, H.; Baus, U.; Schubert, H.; Kremer, K. Asymmetric syntheses via metalated chiral hydrazones: Overall enantioselective α-alkylation of acyclic ketones. *Tetrahedron* **1984**, *40* (8), 1345−1359.
(50) Müller, E.; Metzger, H. Über Nitrosoverbindungen, V. Mitteil.: Einwirkung von Nitrosylchlorid oder Stickstoffmonoxyd auf Cyclohexylmagnesiumchlorid und auf Benzylmagnesiumchlorid. *Chem. Ber.* **1956**, *89* (2), 396−406.
(51) Polivin, Y. N.; Karakhanov, R. A.; Kelarev, V. I.; Sheveleva, T. S.; Ugrak, B. I. Reaction of Methylmagnesium Iodide with Dibutylcyanamide and Dibutylnitrosamine. *Russ. Chem. Bull.* **1993**, *42* (12), 1991−1994.
(52) López-Rodríguez, R.; McManus, J. A.; Murphy, N. S.; Ott, M. A.; Burns, M. J. Pathways for N-Nitroso Compound Formation: Secondary Amines and Beyond. *Org. Process Res. Dev.* **2020**, *24* (9), 1558−1585.

(53) Calmels, S.; Ohshima, H.; Bartsch, H. Nitrosamine Formation by Denitrifying and Non-denitrifying Bacteria: Implication of Nitrite Reductase and Nitrate Reductase in Nitrosation Catalysis. *Microbiology* **1988**, *134*, 221−226.
(54) Ziebarth, D.; Spiegelhalder, B.; Bartsch, H. N-nitrosation of medicinal drugs catalysed by bacteria from human saliva and gastrointestinal tract, including Helicobacter pylori. *Carcinogenesis* **1997**, *18* (2), 383−389.
(55) Stickler, D. J.; Chawla, J. C.; Tricker, A. R.; Preussmann, R. N-nitrosamine generation by urinary tract infections in spine injured patients. *Spinal Cord* **1992**, *30*, 855−863.
(56) Ibahim, A. S.; Elsebai, I. Epidemiology of Bladder Cancer. In *Bladder Cancer*; CRC Press: Boca Raton, 1983; pp 17−37.
(57) Calmels, S.; Ohshima, H.; Crespi, M.; Leclerc, H.; Cattoen, C.; Bartsch, H. In *Relevance of N-Nitroso Compounds to Human Cancer: Exposures and Mechanisms*, Relevance of N-Nitroso Compounds to Human Cancer: Exposures and Mechanisms; IARC: Lyon, 1987; pp 391−395.
(58) IARC Working Group. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Some Pharmaceutical Drugs*; World Health Organization: Lyon, 1980.
(59) Chakravarti, S.; Saiakhov, R. MultiCASE Platform for In Silico Toxicology. In *In Silico Methods for Predicting Drug Toxicity*; Humana Press: Totowa, 2022; Vol. *2425*, pp 497−518.
(60) Chakravarti, S. K. Distributed Representation of Chemical Fragments. *ACS Omega* **2018**, *3* (3), 2825−2836.
(61) Singhal, A. Modern Information Retrieval: A Brief Overview. *IEEE Data Eng. Bull.* **2001**, *24* (4), 35−43.
(62) RDKit: Open-Source Cheminformatics. https://www.rdkit.org.
(63) Hagberg, A. A.; Schult, D. A.; Swart, P. J. In *Exploring Network Structure, Dynamics, and Function Using NetworkX*; Proceedings of the 7th Python in Science Conference, vol. SciPy2008, 2008; pp 11−15.
(64) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E.; DeVito, Z.et al. In *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, Proceedings of the 33rd International Conference on Neural Information Processing Systems ACM Digital Library, 2019; pp 8026−8037.
(65) Kipf, T. N.; Welling, M. *Semi-Supervised Classification with Graph Convolutional Networks* ICLR: Toulon, 2017.
(66) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. *Graph Attention Networks*; ICLR: Vancouver, 2018.
(67) Ashworth, I. W.; Dey, D.; Dirat, O.; McDaid, P.; Lee, D.; Moser, J.; Nanda, K. K. Formation of Dialkyl-N-nitrosamines in Aqueous Solution: An Experimental Validation of a Conservative Predictive Model and a Comparison of the Rates of Dialkyl and Trialkylamine Nitrosation. *Org. Process Res. Dev.* **2023**, *27* (10), 1759−1766.
(68) Ashworth, I. W.; Blanazs, A.; Byrne, J. J.; Dirat, O.; Fennell, J. W.; Kuhl, N.; Wells, S. L.; Whiting, M. P. Approaches and Considerations for the Investigation and Synthesis of N-Nitrosamine Drug Substance-Related Impurities (NDSRIs). *Org. Process Res. Dev.* **2023**, *27* (10), 1784−1791.
(69) European Commission: Directorate-General for Health and Consumers. Opinion on Nitrosamines and Secondary Amines in Cosmetic Products 2012. https://data.europa.eu/doi/10.2772/4079.
(70) Jones, A. R.; Lijinsky, W.; Singer, G. M. Steric Effects in the Nitrosation of Piperidines. *Cancer Res.* **1974**, *34* (5), 1079−1081.
(71) Gowenlock, B. G.; Hutchison, R. J.; Little, J.; Pfab, J. Nitrosative dealkylation of some symmetrical tertiary amines. *J. Chem. Soc., Perkin Trans. 2* **1979**, *2* (8), 1110−1114.
(72) Shephard, S. E.; Schlatter, C.; Lutz, W. K. Assessment of the risk of formation of carcinogenic N-nitroso compounds from dietary precursors in the stomach. *Food Chem. Toxicol.* **1987**, *25* (1), 91−108.
(73) Mirvish, S. S. N-nitroso compounds: Their chemical and in vivo formation and possible importance as environmental carcinogens. *J. Toxicol. Environ. Health* **1977**, *2* (6), 1267−1277.
(74) Mirvish, S. S. Formation of N-nitroso compounds: Chemistry, kinetics, and in vivo occurrence. *Toxicol. Appl. Pharmacol.* **1975**, *31* (3), 325−351.
(75) Loeppky, R. N.; Singh, S. P.; Elomari, S.; Hastings, R.; Theiss, T. E. Evidence for Radical Cations in Linked Mechanisms of N,N-Dialkyl Aromatic Amine Nitration and Nitrosative Dealkylation. *J. Am. Chem. Soc.* **1998**, *120* (21), 5193−5202.
(76) King, F. J.; Searle, A. D.; Urquhart, M. W. Ranitidine—Investigations into the Root Cause for the Presence of N-Nitroso-N,N-dimethylamine in Ranitidine Hydrochloride Drug Substances and Associated Drug Products. *Org. Process Res. Dev.* **2020**, *24* (12), 2915−2926.

https://doi.org/10.1021/acs.chemrestox.4c00435
Chem. Res. Toxicol. 2025, 38, 433−447