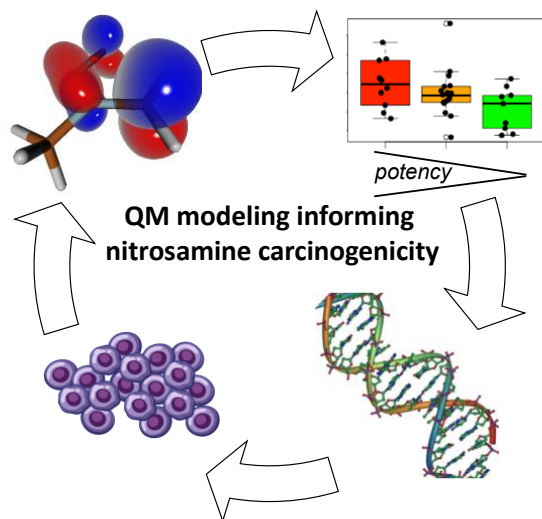


A Quantum-Mechanical Approach to Predicting Carcinogenic Potency of *N*-nitrosamine Impurities in Pharmaceuticals

Jakub Kostal,^{1,2} Adelina Voutchkova-Kostal^{1,2}*

¹Designing Out Toxicity (DOT) Consulting LLC, 2121 Eisenhower Avenue, Alexandria, VA
22314, United States; ²The George Washington University, 800 22nd St. NW, Washington, DC
20052, United States

For Table of Contents Only



KEYWORDS computational toxicology, QSAR, quantum mechanics, carcinogenicity, mutagenicity, nitrosamines

Abstract. *N*-nitrosamine contaminants in medicinal products are of concern due to their high carcinogenic potency; however, not all nitrosamines are created equal, and some are relatively benign chemicals. Understanding the structure-activity relationships (SARs) that drive hazard in one molecule versus another is key to both protecting human health and alleviating costly and sometimes inaccurate animal testing. Here, we report on an extension of the CADRE (Computer-Aided Discovery and REdesign) platform, used broadly by the pharmaceutical and personal care industries to assess environmental and human health endpoints, to predict carcinogenic potency of *N*-nitrosamines. The model distinguishes compounds in three potency categories with 78% accuracy in external testing, which surpasses reproducibility of rodent cancer bioassays and constraints imposed by limited (quality) data. Robustness of predictions for more complex pharmaceutical nitrosamines is maximized by capturing key SARs using quantum mechanics., i.e., by hinging the model on the underlying chemistry vs. chemicals in the

training set. To this end, the present approach can be leveraged in a quantitative hazard assessment and to offer qualitative guidance using electronic-structure comparison between well-studied analogs and unknown contaminants.

Introduction

The potential hazard posed by nitrosamine contaminants in human medicine was first recognized in 2018, when NDMA (*N*-nitrosodimethylamine), a potent carcinogen in rodents, was found in Valsatran.¹ Subsequently, *N*-nitrosamines have been detected at low levels in other pharmaceuticals, where a reaction of a secondary or tertiary amine with a nitrosating agent (e.g., sodium nitrite) generates these compounds in the manufacturing of the active pharmaceutical ingredient (API).² Currently, US and EU regulations require *N*-nitrosamine risk assessments to be carried out on all commercial medical products,³ by deriving acceptable intake (AI) limits from rodent TD₅₀ values. This poses a challenge due to the lack of reliable carcinogenicity data for many chemicals in this class.¹ Animal tests are economically and ethically costly, time-consuming and susceptible to supply-chain disruptions (e.g., during the Covid-19 pandemic), which can affect availability of animals. While regulations permit read across with ‘close’ analogs, these often lead to questionable outcomes given the intricacy of the underlying biochemistry and the structural complexity nitrosamine impurities can adopt in pharmaceutical manufacturing.⁴ Indeed, many detected *N*-nitrosamines are completely unlike the small, well-characterized (reference) compounds, and the current landscape of *N*-nitrosamine carcinogenic potency spans over 4 log units of TD₅₀ values.^{5,6}

Similar to a read-across, classification models and QSARs (Quantitative Structure-Activity Relationships) have shown limited ability to predict carcinogenic potency of *N*-nitrosamines.⁷ With dubious relation to the underlying mechanism of action (MOA), these tools are hindered by the lack of a broad and diverse dataset, inconsistently measured biological data and the apparent imbalance between carcinogenic and non-carcinogenic chemicals.⁷ For example, support vector

machines (SVM) and linear discriminant analysis (LDA) used by Luan et al to classify 148 *N*-nitroso compounds yielded models with ca. 95% and 90% accuracy, respectively; however, the models' predictive power for non-carcinogens in the test set was considerably lower (57% and 71% for SVM and LDA, respectively).⁸ Similarly, using a large number of physicochemical and structural descriptors, Harju et al relied on multiple linear regression to develop a well-correlated model (R^2 of ca. 0.72), which could not, however, be validated.⁷ Because of differences in the organ-specificity of tumors and the dependency of TD_{50} on species, sex and administration route, Helguera et al steered away from a global prediction model and developed multiple QSARs using the Topological Substructural Molecular Design (TOPS-MODE) approach.⁹⁻¹¹ While these models achieved good performance, it is difficult to assess their external predictivity given the relatively few compounds in each model's training set, narrowing their respective applicability domains. Lastly, a commercial software, MultiCASE (QSAR Flex), states that it predict nitrosamines' carcinogenic potency based on 'novel analog search methods';¹² however, without a corresponding peer-reviewed publication outlining its methodology and (external) validation, we cannot gauge its robustness and practical utility.¹³

Fortunately, today's *in silico* approaches are not limited to structure-based methods to infer toxicity for untested chemicals; computational chemistry has developed a broad repertoire of proven techniques that can be translated to predictive toxicology to build robust models.^{14, 15} For *N*-nitrosamines, quantum-mechanics (QM) modeling of molecular interactions in key initiating events (KIEs) is the logical next step in predicting their reactivity, and in turn carcinogenicity.¹⁶ This line of reasoning is consistent with the recent report by Cross and Ponting, which emphasized the need for SAR improvements based on *N*-nitrosamine reaction mechanisms.⁴ Furthermore, because mutagenicity has shown high sensitivity in Ames tests in predicting rodent carcinogenicity,^{5, 17} the KIEs can be related to the metabolic activation mechanisms by Cytochrome P450 (CYP), which control DNA alkylation rates of *N*-nitroso compounds.¹⁸ While QM approaches to study *N*-nitrosamine reactivity and the P450 metabolism are not new,^{16, 19-21} to

the best of our knowledge, no study has incorporated modern density functional theory (DFT) methods into a robust tool to predict carcinogenicity potency in support of substance-specific AI calculations as per ICH M7(R1) guidelines.²² This effort requires careful deconstruction of QM-based reaction-pathway modeling¹⁶ into computationally less-demanding steric and electronic factors that drive the formation of mutagens (and thus carcinogens), and can afford hazard assessments in reasonable timeframes competitive with other nonanimal approaches.

In dissecting the key molecular events in the bioactivation of *N*-nitrosamines in support of reduced QM modeling (Figure 1), the following observations have been made. Differential availability of CYP enzymes across species and their ability to metabolize smaller vs. larger *N*-nitrosamines suggest that a single-species approach in model training and size-related metrics is important in otherwise promiscuous CYPs.^{21, 23, 24} Size metrics should be derived from sampling methods, e.g., molecular dynamics or Monte Carlo (MC) simulations, to obtain reasonable view of larger *N*-nitrosamines' conformational landscapes and their molecular volume.

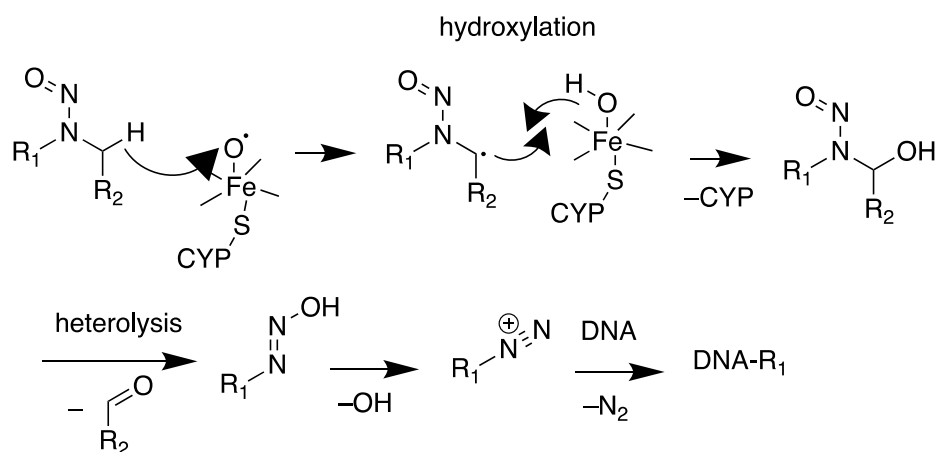


Figure 1. Proposed main pathway for metabolic activation of *N*-nitrosamines to DNA-alkylating agents by Cytochrome P450.

From Figure 1, hydroxylation at the α -C position triggers the biochemical cascade leading to mutations in DNA and carcinogenesis. Local steric and electrostatic effects do play a role, as does the number of α -C hydrogens on either side of the nitrosamine moiety (i.e., unsubstituted α -C positions increase mutagenicity/carcinogenicity), competing hydroxylations at β and, to a lesser extent, γ positions (i.e., hydroxylation at those positions decreases mutagenicity/carcinogenicity by hindering α -C activation, viz. Figure 1) and other enzymatic reaction mechanisms, which can both increase toxicity (e.g., oxidation to aldehydes) and decrease it (e.g., dinitrosation, glucuronidation, etc.).^{4, 18, 25-27} The interplay of these factors, especially for complex pharmaceuticals, undermines the utility of structural alerts to inform *N*-nitrosamine carcinogenicity. Correspondingly, Cross and Ponting found that none of the 39 structural features examined in their study were uniquely (and meaningfully) linked to *N*-nitrosamines of either high or low potency.⁴

In dealing with mechanistic complexity and underlying uncertainty, a computational toxicologist can integrate explicit modeling of 'known' events with statistical methods (assuming a large and chemically diverse dataset) and/or nonspecific reactivity approaches.^{14, 28} In our previous work, we have shown that robust toxicological models can be developed using the modular CADRE (Computer-Aided Discovery and REdesign) platform.^{15, 29} CADRE relies on a tiered approach to bioavailability, metabolic activation and covalent haptentation of biological targets by integrating an expert system with molecular simulations and QM calculations (Figure 2). Because uncertainty is of concern even for the well-characterized toxicity pathways (e.g., dermal sensitization), CADRE balances site-specific QM calculations of known transformations with descriptors derived from Frontier Molecular Orbital (FMO) Theory, which capture reactivity broadly.^{28, 30} Paired with transparent Linear Discriminant Analysis (LDA) and Multivariate Linear Regressions (MLR) modeling, this strategy has delivered robust predictions across multiple toxic endpoints and a diverse chemical space.^{15, 31-34} In particular, CADRE has outperformed other tools

when considering larger, heavily functionalized and biologically active APIs and their synthetic intermediates.^{14, 31}

Here we report an extension to the CADRE platform for *N*-nitrosamines, describing our approach to developing a predictive model for their carcinogenic potency. The current model relies on a combination of physicochemical properties, obtained from linear-response calculations in mixed quantum and classical mechanics (QM/MM) simulations, electronic parameters and reactivity indices, derived from DFT calculations. Our tool distinguishes 3 potency categories based on the 'cohort-of-concern' (COC) definition for *N*-nitrosamines with 77% accuracy in external validation.

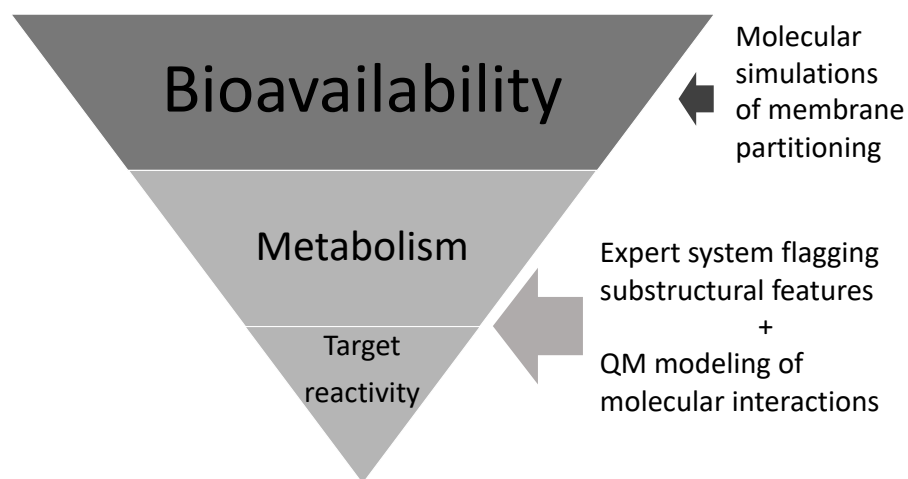


Figure 2. Outline of the modular CADRE (Computer-Aided Discovery and REdesign) platform for toxicity predictions and design of safer chemicals.

Methods

Dataset and data quality. A dataset of 96 compounds containing the nitrosamine moiety ([OX1]=[NX2][NX3]([#6,#1])[#6,#1])³⁵ was obtained from the Lhasa Carcinogenicity Database (LCDB, carcdb.lhasalimited.org), which draws data from the now-retired Carcinogenicity Potency

Database (CPDB)^{36, 37} (Table S1). Molecular weight of the collected *N*-nitrosamines ranged from 74 to 314 g/mol with an average value of 157 g/mol. The dataset spanned 6 orders of magnitude in calculated TD₅₀ values, from ca. 0.008 to 167 mg/kg, with an arithmetic and geometric means of 7.5 and 0.97 mg/kg, respectively. Due to species-induced variability in TD₅₀ data, only the prevalent rat model was selected. However, we noted that standard deviation (s) across species was as high as 400 mg/kg for some *N*-nitrosamines, while mean s per chemical across the entire dataset was ca. 8.5 mg/kg. The dataset was subsequently split into a training set (60) and a test set (36). Due to limited data, our dataset incorporated data from both sources available in LCDB: those with harmonic mean TD₅₀ values derived by Lhasa using a curated subset of tumor sites (41) and those obtained directly from CPDB (55). The former is considered to be of higher quality, by omitting outcomes lacking a dose-response effect, single-concentration measurements and TD₅₀'s obtained using the 'lifetable' vs. the terminal sacrifice method.³⁸

Computational modeling. The tiered structure of the *N*-nitrosamine model mimics that of the CADRE skin and respiratory sensitization models,^{15, 34} given our continued reliance on modeling of molecular interactions and the relevance of covalent binding in KIEs across all three endpoints. For all compounds here, ionization, an important factor in nitroso carcinogenicity,^{39, 40} was assessed at biological pH (7.4). Aqueous Monte Carlo (MC) simulations were used in conjunction with mixed quantum and classical mechanics calculations (QM/MM) to compute physicochemical properties related to toxicokinetics (e.g., molecular volume, solvent accessible area, dipole, polarizability, globularity etc.). Transport-related properties, such as the octanol-water partition coefficient ($\log P_{o/w}$), were predicted in linear-response calculations, based on solute-solvent energetics (i.e., Coulomb and van der Waals interactions) obtained from QM/MM/MC simulations.

In the second tier, mechanistic alerts for *N*-nitrosamines were used to flag in-domain chemicals using the SMARTS language.⁴¹ The subsequent third tier is based on a QM-FMO approach, relying on global and atom-based steric factors and parameters derived from

nitrosamines' electronic structure. Here, electronic structures were computed using density functional theory (DFT); the mPW1PW91/MIDIX+ method was selected based on performance in Lewis-acid/base and hydrogen-abstraction reactions, which are key to nitrosamines' metabolic pathways (Figure 1).⁴²

Statistical modeling. The R language and environment for statistical computing (version 4.1.2) was used for data analysis, linear regressions and linear discriminant analyses (LDAs).⁴³ Multivariate normality (MVN) of descriptors was determined using functions in the MVN library, and, where necessary, variables were transformed to a logarithmic scale. The following additional R packages were included for data analysis: ggplots2, YaleToolkit, calibrate, caret, MVN and MASS. Descriptor selection was carried out using genetic algorithm, implemented in the library genalg, using 100 iterations with a mutation probability of 0.05 and the Bayesian Information Criterion (BIC) to avoid overfitting. Internal performance was estimated using the leave-one-out (LOO) cross validation. Final model selection was based on performance of the top 5 models in external validation.

Results and Discussion

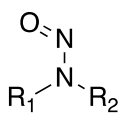
Electronic structure readacross. Readacross is one of the pillars of predictive toxicology, a technique commonly used along (Q)SARs and expert systems to infer toxicity of untested chemicals based on close structural analogs.^{28, 44} The difficulty with readacross is to ascertain the validity of structural similarity given the complex relationship between structural features and mechanistic requirements of KIEs in most adverse outcome pathways (AOPs).^{14, 45} This issue is exemplified by the (incorrect) use of simple alkyl *N*-nitrosamines to 'predict' carcinogenicity of more complex pharmaceuticals. Nonetheless, there is utility in readacross when considered at the electronic (vs. atomistic) level. To that end, Frontier Molecular Orbital (FMO) Theory can be used to derive both kinetic and thermodynamic stability of molecules.^{46, 47} Pairing FMO with

Fukui-function calculations allows us to develop electronic parameters that can inform both global and local propensity to accept or donate electron density, which are traits key to acid, base and radical chemistries.^{15, 28, 48, 49} These parameters can then be used to compare reactivity of a compound (and its specific moiety, such as the *N*-nitrosamine) to a 'known' quantity in an electronic-structure readacross (ESR). So long as the underlying mechanism is the same, and is reasonably-well characterized, there is no need for 'structural similarity' in ESR because this method hinges on fundamental principles of chemical reactivity, and because MO-derived parameters, assuming prudent functional and basis-set selection,⁵⁰ can accurately condense effects of the entire molecule. Relying on existing knowledge, the added benefit of an ESR analysis is in discovering the most meaningful QM parameters for developing a predictive model, and to validate, further elucidate or challenge our understanding of the KIE's molecular mechanism.

ESR Case Study. To demonstrate how ESR can be used to gauge reactivity (and by extension carcinogenicity) of *N*-nitrosamines, we considered the well-studied *N*-nitrosodiethylamine (NDEA, CAS 55-18-5) and *N*-nitrosodimethylamine (NDMA, CAS 62-75-9) compounds, for which we have reliable TD₅₀ data. We then developed a series of electronic (and steric) parameters relevant to KIEs in Figure 1, and applied them to DIPNA (*N*-nitrosodiisopropylamine, CAS 601-77-4), NDBA (Dibutyl nitrosamine, CAS 924-16-3), EIPNA (*N*-nitrosoethylisopropylamine, CAS 16339-04-1), NMBA (*N*-Nitrosomethylaminobutyric Acid, CAS 61445-55-4) and MeNP (1-methyl-4-nitrosopiperazine, CAS 16339-07-4), using NDEA and NDMA as reference compounds (Table 1). All of these compounds are the most commonly found *N*-nitrosamines in medicinal products, and have derived daily acceptable intake (AI) limits using SAR/read-across.^{2, 3} As a result, there are discrepancies between AI limits (column 3) and applied TD₅₀ values (column 2) in Table 1. Specifically, AI limits for DIPNA, NDNA, EIPNA and MeNP are based on SAR/read-across to NDEA's TD₅₀ (obtained from CPDB), whereas the limit

for NMBA was derived using NDMA's TD₅₀ (obtained from CPDB).^{2, 3} For NDEA and NDMA, the European Medicines Agency (EMA) uses AI limits based on a harmonic mean of TD₅₀ values for all available tumor sites for the rat model in CPDB, whereas our reference values correspond to the higher-quality Lhasa TD₅₀'s (viz. Methods). For NDBA and NMBA, only lower-quality CPDB TD₅₀ values were available.

Table 1. Steric and electronic parameters support hazard assessment of 5 nitrosamines using electronic structure readacross (ESR) to the well-studied *N*-nitrosodiethylamine (NDEA, CAS 55-18-5) and *N*-nitrosodimethylamine (NDMA, CAS 62-75-9). AI = daily acceptable intake limit; SAVA = solvent-accessible volume area; Log D_{7.4} = octanol-water distribution coefficient (pH = 7.4).

Compound (CAS number)		TD ₅₀ (mg/kg)	AI (ng/day)	mean SAVA, α-C (Å ³)	Log D _{7.4}	Radical susceptibility (α-carbon, eV)	Chemical potential, Δμ (eV)	Electrophilic susceptibility (α-carbon, eV)
DIPNA (601-77-4)	R ₁ ,R ₂ = iPr	Lacking data	26.5	12.4	1.59	0.0121	-0.219	0.0140
NDBA (924-16-3)	R ₁ ,R ₂ = nBu	0.691*	26.5	18.5	2.69	0.0258	-0.125	0.0351
EIPNA (16339-04-1)	R ₁ = Et, R ₂ = iPr	Lacking data	26.5	17.4	1.17	0.0193	-0.240	0.0311
NMBA (61445-55-4)	R ₁ = Me, R ₂ = aminobutyric acid	0.982*	96.0	26.0	-2.99	0.0262	-0.427 (-0.234)**	0.0208
MeNP (16339-07-4)	R ₁ = Me, R ₂ = pip	Lacking data	26.5	21.7	-0.71	0.0231	-0.479 (-0.111)**	0.0259
NDEA (55-18-5)	R ₁ ,R ₂ = Et	0.0177*	26.5	19.6	0.75	0.0318	-0.510	0.0411
NDMA (62-75-9)	R ₁ ,R ₂ = Me	0.177*	96.0	34.1	0.04	0.0461	-0.428	0.0492

*NDEA/NDMA use Lhasa (LCDB) TD₅₀ harmonic mean, whereas NDBA/NMBA rely on CPDB (Gold) TD₅₀ mean.

**These species are ionized at physiological pH of 7.4.

In considering data-poor nitrosamines in Table 1 (i.e., rows 2-6), we first focused on steric hindrance in the hydroxylation step, calculated as the solvent-accessible volume area

(SAVA) on the α -carbon(s). Here, DIPNA, NDBA, EIPNA, NMBA and MeNP all showed greater steric hindrance than NDMA, implying lower reactivity. However, NMBA and MeNP were both on average more accessible at the α -carbon-position than NDEA, which is a potent COC ($TD_{50} = 0.018$ mg/kg), indicating that sterics (i.e., structural features) alone cannot explain nitrosamine potency. In comparing electronic structures, we considered both global (i.e., molecular) properties as well as local (i.e., atom-based) reactivity indices. Global electronic properties were derived from the Frontier Molecular Orbital Theory (FMOT), while local properties were expressed via the Fukui function, $f(r) = \left[\frac{\partial \rho(r)}{\partial N} \right]_{v(r)}$, which measures propensity of an atom (or molecule) to accept or donate electron density. Here, we computed susceptibility to undergo radical chemistry at the α -carbon position, $f^0(C_\alpha) = [\rho_{N+1}(C_\alpha) + \rho_{N-1}(C_\alpha)]/2$, which reflects the initial C-H alkylation step, and where maxima in the susceptibility index correspond to greater propensity for this process (Table 1, Column 5). For asymmetric *N*-nitrosamines, values in Table 1 reflect the more reactive position. In our analysis, all 5 nitrosamines showed lower reactivity at α -C than either NDEA or NDMA, which is consistent with the limited experimental data available.

In deriving useful global properties, we computed the change in chemical potential for the hydroxylation step (i.e., $\Delta\mu$). Chemical potential is a well-established quantity that reflects the escaping tendency of electrons, $\left(\frac{\partial G}{\partial N}\right)$, and its change along the reaction pathway was shown to control uphill chemical processes (while downhill chemical reactions are typically controlled by a change in molecular hardness).⁵¹ Since the initial C-H activation step is endergonic,⁵² we compared $\Delta\mu$ across all 7 nitrosamines (Table 1, Column 6), noting that $\Delta\mu$ is capable of distinguishing between the greater carcinogenicity of NDEA ($\Delta\mu = -0.510$, $TD_{50} = 0.018$ mg/kg/day) over NDMA ($\Delta\mu = -0.428$, $TD_{50} = 0.177$ mg/kg). For EIPNA and NMBA, which yield two different hydroxylated metabolites (viz. Figure 1), $\Delta\mu$ was averaged assuming Boltzmann

distribution of states, i.e., $\Delta\mu = -RT \ln \sum_{i=1}^2 e^{-\frac{\Delta\mu_i}{RT}} + RT \ln 2$. From Table 1, all five *N*-nitrosamine ‘unknowns’ showed less favorable $\Delta\mu$ than NDEA or NDMA, indicating lesser propensity to form the hydroxylated metabolite. When ionization of NMBA and MeNP at physiological pH was considered, the difference in reactivity between NDEA/NDMA and the remaining compounds increased further (Table 1, Column 6, values in ellipses). We should note that ionization affects toxicity further by decreasing bioavailability. To that end, lesser relative toxicities of NMBA and MeNP are supported by their respective octanol-water distribution coefficients ($\log D$ at pH = 7.4), which are much lower than for the remaining nitrosamines (Table 1, Column 4), indicating limited partitioning across cell membranes.

Lastly, we briefly considered the electrophilicity of all 7 nitrosamines, which plays an important role downstream in the covalent binding of DNA by alkylating metabolites (Figure 1). It is reasonable to propose that sufficient electronic ‘signature’ is present in the parent molecules to elucidate differential electrophilic reactivity of their metabolites.⁵³ To that end, calculated maxima in electrophilic susceptibility on the α -C, $f^+(C_\alpha) = [\rho_{N+1}(C_\alpha) - \rho_N(C_\alpha)]$, showed that both NDEA and NDMA are more susceptible to a nucleophilic attack (Table 1, Column 7) than the remaining nitrosamines. This finding is consistent with our analysis above, indicating lower reactivity (and carcinogenicity) of the selected 5 nitrosamines than the well-characterized NDEA and NDMA compounds. Interestingly, while electrophilicity correctly distinguished the minute difference between NDBA and NMBA TD_{50} ’s, it did not reflect experimental differences between NDEA and NDMA, which was better predicted by $\Delta\mu$, highlighting the limitations of using a single variable (or a single KIE) in a readacross.

Importantly, Table 1 outcomes suggest that the daily acceptable intake (AI) limits derived using SAR and (structure-based) readacross are overly conservative. In considering available experimental data and computed reactivity parameters for the 5 compounds in question, none exceed metrics calculated for NDMA, for which AI is set at 96.0 ng/day (vs. 26.5 ng/day for NDEA).

In fact, all data indicate substantially lower reactivity and toxicity, with values closer to NMBA (AI = 96.0 ng/day) than either NDMA or NDEA.

Structure Activity Relationships. While ESR alleviates personal bias and applicability-domain issues associated with structural readacross, for large and complex molecules, it may be difficult to lean on a single variable. Furthermore, a quantitative (vs. qualitative) assessment of toxicity may be required by the industry or regulators. To that end, a statistical model, predictive of N-nitrosamine carcinogenic potency and constructed as a composite of several key electronic and steric parameters informed by toxicity mechanism, should be leveraged instead to deliver an accurate *in silico* assessment. Prior to developing such models, it is vital to perform analysis of key trends in the dataset, which could help elucidate mechanistic data clusters (i.e., analog series), and pinpoint (true) outliers.^{13, 54}

Here, we considered several 'known' SARs and showcased how underlying trends can be captured using QM (Figure 3). Since ionization increases water solubility and decreases bioavailability, we noted that ionizable nitrosamines were generally less potent carcinogens than non-ionizable compounds (viz. median TD₅₀ of 6.04 mg/kg for ionizables vs. 0.982 mg/kg for non-ionizables, as visualized in Figure 3, first row). This effect is captured well by aqueous solubility (log S), as predicted here from QM/MM/MC simulations. In comparing two representative analogs, **nid** and **noe**, computed log S (-1.2 and -0.065 for **nid** and **noe**, respectively) was consistent with calculated TD₅₀'s (68.7 and 0.18 mg/kg, respectively). It has been proposed that β hydrogens (Hs) decrease toxicity, owing to competing hydroxylation processes that lower the probability of heterolysis and subsequent DNA alkylation.^{4, 55} We found this rule to be misleading, as nitrosamines with no β-Hs were overall less toxic than those with β-Hs in the dataset (Figure 3, second row). As shown by structures **ndi** and **ndt**, (TD₅₀ = 0.11 and 31.2, respectively), β-Hs may imply lack of a more complex, sterically hindering moiety at

this position, thus promoting reactivity at the (carcinogenic) α position. In computations, this can be captured by the solvent accessible surface area (SASA) on α -C's, which was calculated to be 4.1 \AA^2 for *ndi* and 1.9 \AA^2 for *ndt*, supporting the difference in TD_{50} values.

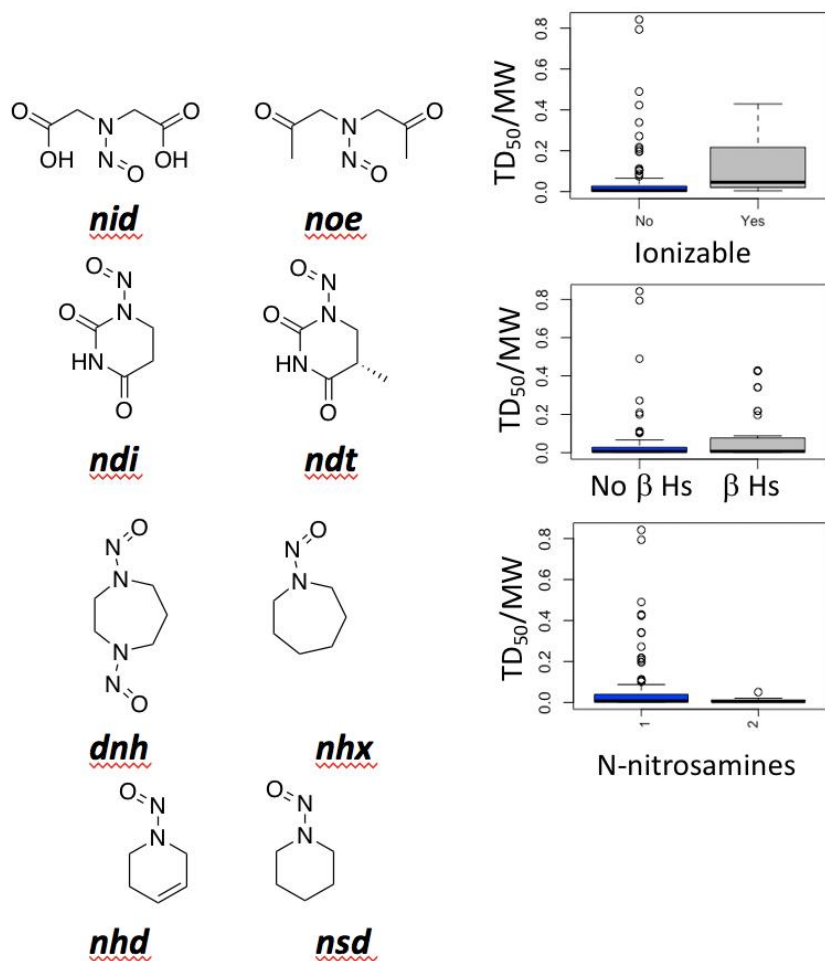


Figure 3. Structural analogs (left) highlighting corresponding key trends across the present N-nitrosamine dataset (right).

The data further shows that multiple nitrosamine groups increase potency, as showcased by *dnh* vs. *nhx*, and captured by the lower median TD_{50} value of compounds with two (0.70 mg/kg) vs. one (1.26 mg/kg) N-nitrosamine groups (Figure 3, third row). We can readily capture this property using QM, for example, by the change in chemical potential for the hydroxylation step

(i.e., $\Delta\mu$), which was calculated to be -4.21 eV for **dnh** and -3.67 eV for **nhx** (i.e., considerably less favorable for the latter). Lastly, sp^2 C's proximal to the α -C appear to increase carcinogenic potency (Figure 3, fourth row), as radical stabilization of the α -position by electron-hole delocalization promotes hydroxylation. Here, this effect is demonstrated by **nhd** vs. **nsd**, where radical stabilization by π -resonance in the former leads to greater toxicity ($TD_{50} = 0.059$ mg/kg) over the latter ($TD_{50} = 1.12$ mg/kg). Orbital-mixing effects can be quantified in QM calculations via Second Order Perturbation Theory (SOPT) by estimating the magnitude of the π - p donor-acceptor contributions or, indirectly, by population analysis that gauges electron-density partitioning across atoms in the molecule. To that end, for the case of **nhd** vs. **nsd**, Hirshfeld population analysis (HPA) showed that α -C in **nhd** is more electropositive (+0.09) than α -C's in **nsd** (+0.07), owing to the σ -electron withdrawing π bond.

For a few documented SARs, the present dataset afforded a comparison across a multiple-analog series (Figure 4). In Figure 4A, we examined the effect of increasing (symmetrical) alkyl chains around the nitrosamine group. Consistent with past studies, longer chains generally translate to lower carcinogenic potency (i.e., increasing TD_{50} 's), owing to increasing steric bulk on the α -C (captured here using SASA). The important exception is NDEA (**den**), especially as it is commonly used by the EMA and FDA in structure-based readacross to inform carcinogenicity of nitrosamine impurities in medicinal products. NDEA (**den**) is an order-of-magnitude more carcinogenic than NDMA (**dmn**). This activity 'peak' in the overall trend can be rationalized using SOPT (of Fock Matrix in Natural Bond Orbital, NBO, basis),⁵⁶ based on the metabolic pathway of *N*-nitrosamines (Figure 1). In contrast to NDMA (**dmn**), NDEA (**den**) benefits from orbital mixing (i.e., hyperconjugation), which stabilizes the radical intermediate prior to C_{α} -hydroxylation by $\sigma(C-H) \rightarrow p(C\bullet)$ donor-acceptor effects (Figure 5A). From Figure 5B, hyperconjugation is observed in Molecular Orbital Theory (MOT) as well; here, orbital mixing lowers energy of the MO consisting of the lone-pair atomic orbital (AO). Crucially, this

interaction is absent in NDMA (*dmn*), and while it exists in *dpn*, *ndb* and *dna*, diminishing SASA with increasingly alkyl-chain length likely compensates for this effect, leading to an overall decrease in carcinogenicity. The magnitude of $\sigma(\text{C-H}) \rightarrow \text{p}(\text{C}\bullet)$ orbital mixing in the NBO basis was estimated to be ca. 60-80 kcal/mol across the present alkyl-nitrosamine series, with slight increase in proceeding from shorter to longer alkyl chains.


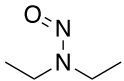
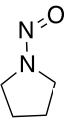
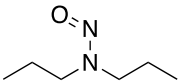
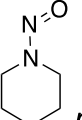
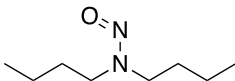
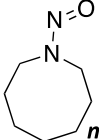
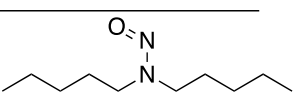
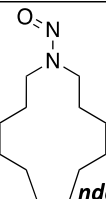
A.		B.			
TD ₅₀ (mg/kg)	SASA (Å ²)		TD ₅₀ (mg/kg)	SASA (Å ²)	
0.18	13.7	<i>dmn</i> 			
0.018	2.2	<i>den</i> 	<i>npy</i> 	2.0	22.7
0.19	1.70	<i>dpn</i> 	<i>nsd</i> 	1.1	22.4
0.69	1.77	<i>ndb</i> 	<i>nhm</i> 	0.038	20.1
4.1	1.02	<i>dna</i> 	<i>ndc</i> 	10.9	18.2

Figure 4. Series of acyclic (**A**) and cyclic (**B**) alkyl nitrosamine analogs with reliable rat TD₅₀ values in the present dataset and corresponding SASA values, computed at C_α. **A** and **B** columns are aligned so as to compare cyclic vs acyclic series' analogs.

Next, we were interested in comparing activities of acyclic and cyclic alkyl nitrosamines (Figure 4A vs 4B). We noted that smaller cyclic analogs were less potent than their acyclic counterparts (e.g., *den* vs. *npy* or *dpn* vs. *nsd*), despite greater accessibility of the α-C (owing

to the *syn* relationship of heteroatoms in a cycle vs. *anti* in a linear chain). However, because TD₅₀ values decrease with increasing ring size, *nhm* ends up being more potent than either *dpn* or *ndb*. Concurrently, SASA decreases with increasing ring size (Figure 4B); thus, it cannot explain the trend of increasing potency when proceeding from *npy* to *nhm*. Here, conformational sampling in conjunction with QM calculations is necessary, showing that as the ring size increases, the distance between the hydrogen on α -C and oxygen of the nitrosamine group decreases (Figure 5B). This effect, which stems from greater rigidity of smaller rings vs. conformational flexibility of larger rings, is relevant because the C $_{\alpha}$ -H position is hydroxylated in *N*-nitrosamine metabolism, and in the heterolysis (i.e., aldehyde cleavage) step, proton is transferred from the OH group at C $_{\alpha}$ to the O of the nitrosamine (Figure 5B, top – transferred proton outlined in red color). Proton transfers become more facile with decreasing distance between the donor and the acceptor.⁵⁷

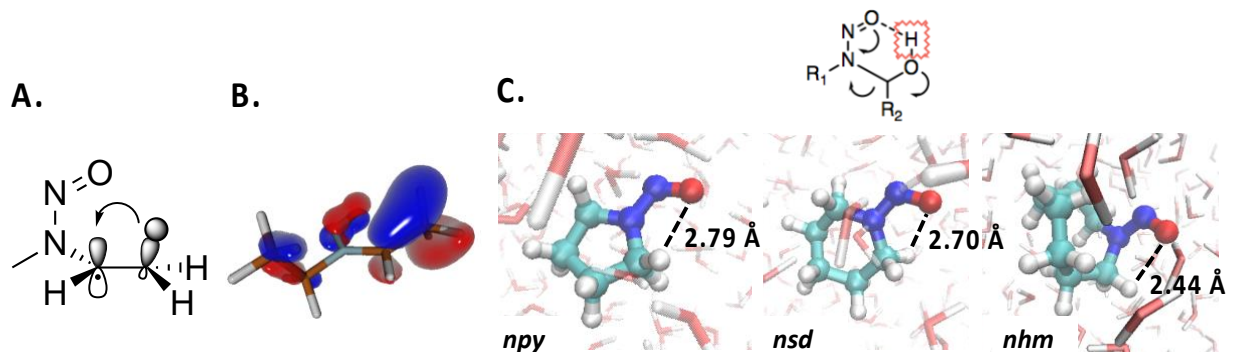


Figure 5. Orbital mixing in the NBO (Natural Bond Orbital) basis (A) and MOT (Molecular Orbital Theory) basis (B), capturing hyperconjugation effects between C• and proximal C-H σ bond. C: Mechanism of proton transfer in the heterolysis step of nitrosamine metabolism (top) as affected by the distance between the C $_{\alpha}$ -H position and the nitrosamine oxygen (bottom), obtained from aqueous QM/MM/MC simulations.

Physicochemical properties. Consistent with past studies,^{4, 58} the above analysis illustrates the difficulties in using structural alerts to reliably inform *N*-nitrosamine carcinogenic potential

(and even less so potency), and the important role that QM plays in modeling nitrosamine metabolism. Since physicochemical properties are commonly used in QSAR models,⁵⁹ both to capture bioavailability and, with limited success, as proxies for the more computationally demanding electronic-structure (i.e., reactivity) descriptors,²⁸ we carried out a comprehensive analysis to determine their relevance in *N*-nitrosamine carcinogenicity. For the sake of this exercise, we considered the entire dataset (i.e., 96 compounds), and assigned 4 categories of potency based on the 1.5 mg/kg/day TTC (Threshold of Toxicological Concern): Category 1 ($TD_{50} < 0.15$ mg/kg), Category 2 (TD_{50} : 0.15 – 1.5 mg/kg), Category 3 (TD_{50} : 1.5 – 15 mg/kg) and Category 4 ($TD_{50} > 15$ mg/kg). Here, we highlight the most significant properties that could be rationalized in the context of *N*-nitrosamine metabolism, as computed from aqueous QM/MM/MC simulations using linear response calculations (viz. Methods). From Figure 6, no single property could reliably distinguish compounds in different potency categories; however, we noted an informative trend to guide our predictive models. From left to right, globularity is an expression of molecular shape, and is highly conformation-dependent; previous studies have shown that globularity can be a useful descriptor for inhibition of enzymatic (and cellular) activity.^{60, 61} Here, nitrosamine COCs (Cat 1-2) were noted to be on average more spherical than non-COCs (Cat 3-4), though this distinction was statistically weak, likely due to the promiscuity of Cytochrome P450. Polarizability reflects the ability of a molecule's electron cloud to distort in response to an external field, inducing a dipole. As such, polarizability is a useful proxy for reactivity (e.g., in describing substitution and elimination reactions on carbon-halogen bonds) as well as bioavailability (i.e., highly polarizable compounds tend to be more lipophilic). To this end, past studies have exploited polarizability as a useful descriptor of neurotoxicity.⁶² Here, polarizability showed that potent COCs (Cat 1) were less polarizable than Cat 2 COCs; however, no useful trend was observed across all 4 categories. Aqueous solubility ($\log S$) and the octanol-water partition coefficient ($\log P_{o/w}$) are common descriptors of membrane permeability and general bioavailability,^{28, 30} where higher lipophilicity typically indicates greater

toxicity. Here, the trend is reversed, as COCs (Cat 1-2) appear to be more hydrophilic than non-COCs (Cat 3-4). Trends across all properties above are consistent with the landscape of studied *N*-nitrosamines, where the most potent carcinogens are smaller molecules (i.e., less sterically hindered, more globular, less polarizable and thus more water-soluble). This is further supported by predicted Caco-2 (i.e., apparent Caco-2 cell permeability in nm/sec), which is a model for non-active transport across the gut-blood barrier. Here, COC nitrosamines were found to be less permeable than non-COCs (Figure 6). Our findings are best summarized by the last graph in Figure 6, which shows distributions of computed van der Waals (i.e., Lennard Jones or LJ) interaction energies, where the trend is that of increasing interactions (favoring large and polarizable substituents) with decreasing potency. Overall, our analysis shows that while physicochemical properties are important descriptors of bioavailability, overreliance on these metrics in predicting nitrosamine carcinogenicity is likely to be misleading; we can infer from the present outcomes that reactivity, and not bioavailability, is the driver of toxicity across the known chemical space.

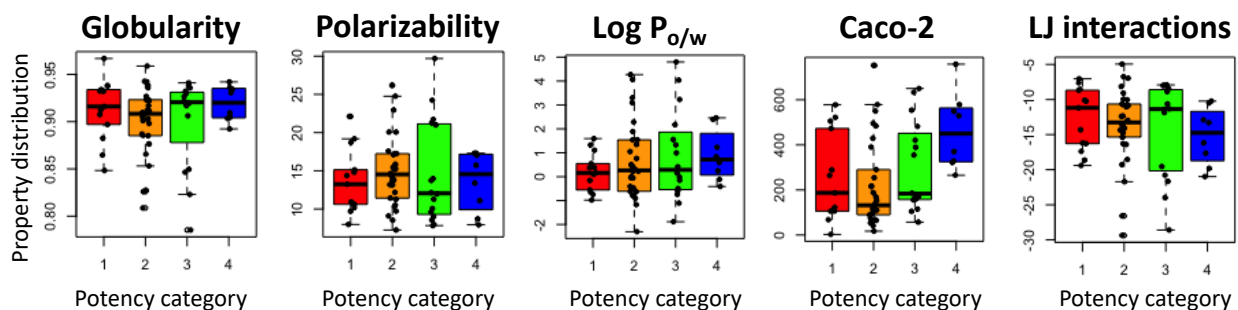


Figure 6. Representative physicochemical properties and their distributions calculated for potency categories of studied nitrosamines; Category 1 ($TD_{50} < 0.15$ mg/kg), Category 2 (TD_{50} : 0.15 – 1.5 mg/kg/day), Category 3 (TD_{50} : 1.5 – 15 mg/kg) and Category 4 ($TD_{50} > 15$ mg/kg). Log $P_{o/w}$ = octanol-water partition coefficient; Caco-2 = apparent Caco-2 cell permeability; LJ interactions = Lennard-Jones interaction energies between nitrosamines and aqueous medium.

Model Performance. We investigated the use of linear discriminant methods to classify *N*-nitrosamine contaminants into 3 potency categories: low potency (Cat 1, $TD_{50} \leq 0.15$ mg/kg), COC compounds (Cat 2, $0.15 < TD_{50} \leq 1.5$ mg/kg) and potent COCs (Cat 3, $TD_{50} > 1.5$ mg/kg). Linear discriminant analysis (LDA) identifies a discriminant function by dividing an *n*-dimensional descriptor space into regions separated by a hyperplane. We also explored a multivariate regression (MLR) analysis as means to predicting TD_{50} values. Both approaches were applied with variable selection based on a genetic algorithm. In both cases, the analyses generated posterior probabilities via a cross-validation scheme. A comprehensive set of 92 QM variables and physicochemical properties was tested to find the most robust descriptors. In light of the mechanistic complexity and to prevent overfitting, LDA models were developed separately for *N*-nitrosamines with and without β -Hs (Table 2).

The LDA 'no β -Hs' model afforded 92% accuracy and 83% LOO accuracy (Table 2). The performance of the LDA model with β -Hs was comparable at 94% accuracy and 89% LOO accuracy. In external validation, both models fared reasonably well with 71% and 79% accuracy, respectively. Indeed, these values can be deemed perfectly satisfactory given the notoriously low reproducibility of rodent cancer bioassays, which was estimated at 57% by comparing carcinogenicity classifications from the National Cancer Institute/National Toxicology Program and from open literature (both sourced from CPDB).⁶³ Given the relatively small size of the no β -Hs model, combining the two COC categories resulted in improved external performance (86%). Overall performance was very good at 92% (training set), 87% (LOO) and 78% (external). We should note that our training/test set ratios for both LDA models were considerably more stringent (1.7) than is the accepted industry standard (4)⁶⁴ to showcase the broad applicability of QM descriptors and to promote confidence in the models' robustness.

Table 2. Performance statistics for LDA models developed for carcinogenic potency of *N*-nitrosamines.

Model	Accuracy (%)	LOO Accuracy (%)	External Validation (%)
LDA (no β -Hs)	92 (12)*	83	71/86** (7)
LDA (β -Hs)	94 (48)	89	79 (29)
Total	93 (60)	87	78 (36)

*Number of compounds provided in ellipses.

**2-category split ($TD_{50} \leq 0.15$ mg/kg and $TD_{50} > 1.5$ mg/kg)

Lastly, we briefly explored MLRs to predict TD_{50} values. While a well-performing model was developed for potent nitrosamines without β -Hs ($TD_{50} < 1.5$ mg/kg), linear trends would break down above this range, owing to the high inherent variability of carcinogenesis as a multi-stage probabilistic process, and likely because high TD_{50} values are deemed less reliable.⁶⁵ From Figure 7, the best MLR model afforded R^2 and R^2_{adj} of 85% and 82%, respectively, across 13 compounds and using 3 descriptors (global electrophilicity index, electrophilic reactivity on the α -C and maximal radical susceptibility).

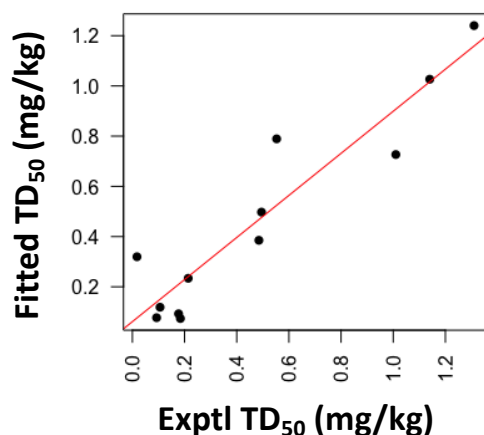


Figure 7. The best MLR model for *N*-nitrosamines without β -Hs and with $TD_{50} < 1.5$ mg/kg. $N = 13$, $R^2 = 85\%$, $R^2_{adj} = 82\%$.

Applicability domain. Transparency about model's applicability is key to trusting its real-world performance. For QSARs, applicability domain can be defined in myriad different ways,⁶⁶⁻⁶⁸ but it is always tied to the distribution of descriptor (and response) values in the training set. Thus, if a prediction is supported by value(s) outside this (multivariate) space, it should be flagged as out of domain. Practices then differ, based on the model and context of the hazard assessment, as to how to treat chemicals predicted this way.⁶⁹ In CADRE models (and models based on fundamentals of chemical interactions in general), the concept of applicability domain is less obvious.⁷⁰ Presently, we report out-of-domain predictions (based on out-of-range descriptor values) using a confidence score,^{15, 29, 34} but we have repeatedly found that even lower-confidence predictions are of high quality and aligned with experiment.^{14, 31} To this end, it can be proposed that models anchored in the underlying chemistry (vs. chemicals in the training set) have a 'softer' edge in their applicability domain, with fewer inter- and extrapolation concerns, as long as the molecular transformations are modeled appropriately and can be assumed to initiate the toxic endpoint in question.

In elucidating the applicability domain for the present models, we considered value ranges of descriptors in external validation (Figure 8). Specifically, we were curious as to whether significant deviations from training-set distribution impacted performance, hoping to boost confidence in applying our LDA models to a broad range of nitrosamine contaminants. This information is particularly important for endpoints with sparse data, where the use of *in silico* models in hazard assessment may be necessary to satisfy regulatory needs, limited access to test animals and general economic realities of new chemical development. As previously discussed, each QM descriptor in our LDA models (Figure 8) bears mechanistic relevance to the N-nitrosamine mechanism: global and local electrophilicity metrics gauge the compound's propensity to accept electron density based on FMO Theory and the Fukui function; C_{α}/C_{β} reactivity reflects orbital overlap with a proxy target (we rely on an alkylamine to simulate nucleic acid base); radical susceptibility captures the initial C-H alkylation step; and LJ

interactions represent van der Waals energies with surrounding water molecules, computed from QM/MM simulations. Surface energy, estimated here via the sum of molecular dipole, quadrupole and octopole moments, reflects electrostatic interactions with a medium as well as transport energies of holes and electrons in organic and biochemical reactions.^{71, 72} To this end, this metric relates to both toxicokinetic phenomena of membrane permeability and covalent events in the *N*-nitrosamine metabolic-activation pathway. Figure 8 clearly shows is that in both LDA models for *N*-nitrosamines without β -Hs (Figure 8A) and with β -Hs (Figure 8B), exceeding the distribution of descriptor values in the training set did not negatively impact external predictivity. In fact, there was no correlation between value distance from the training set and the rate of mispredictions, indicating robustness beyond current knowledge.

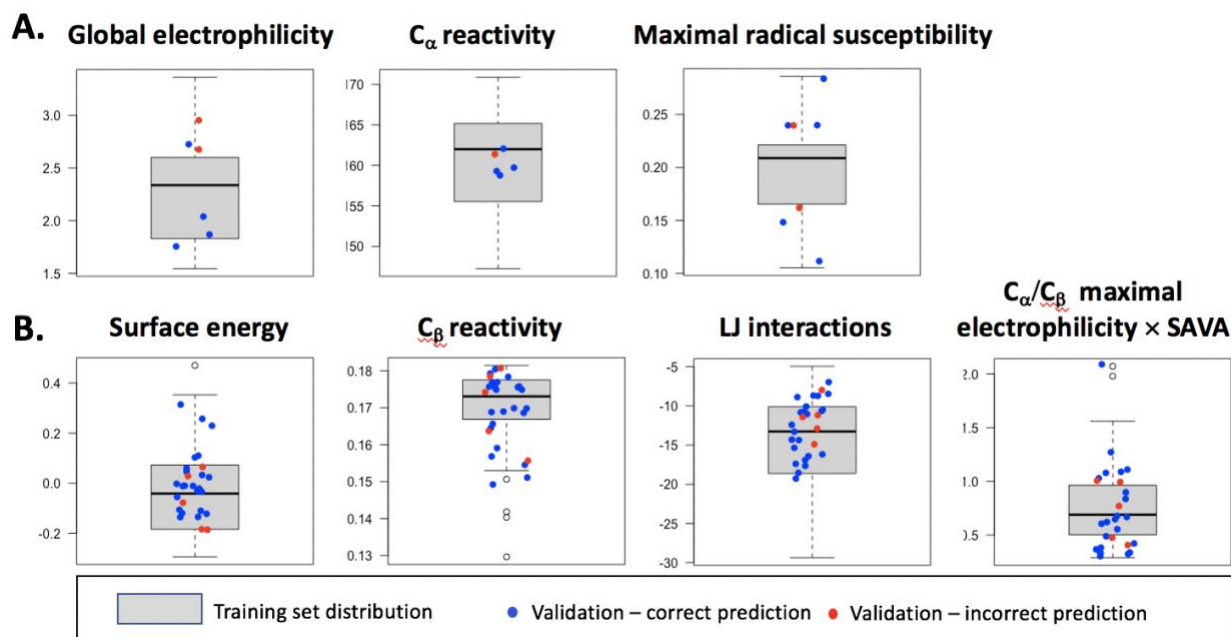


Figure 8. Descriptor values for correct predictions vs. mispredictions in external testing relative to value distributions in the training sets of the top LDA model for *N*-nitrosamines without β -Hs (A) and with β -Hs (B).

Model outliers and limitations. Understanding model outliers is important in building confidence for real-world applications. Here, all mispredictions were based on a single category under- or over-prediction, with the exception of *N*-nitrosoheptamethyleneimine (Figure 9B, *nhm*). This is a notable outcome, as the difference between Cat 1/2 and Cat 2/3 cutoffs is a mere 1 mg/kg/day, while standard deviation in TD₅₀'s across the entire dataset is ca. 8.5 mg/kg/day. Nonetheless, to better understand the model's limitations, we considered chemical structures of outliers in the context of the known SARs and the mechanism to form mutagens. From Figure 9A, *ntm*, *nte*, and *npe* are *N*-nitrosamines mispredicted in the training set, where *ntm* (TD₅₀ = 3.54 mg/kg) and *nte* (TD₅₀ = 2.55 mg/kg) were overpredicted as Cat 2 COCs (0.15 < TD₅₀ ≤ 1.5 mg/kg), and *npe* (TD₅₀ = 0.00797 mg/kg) was underpredicted as Cat 2 COC. The sulfur in *Ntm* makes it a softer analog of *nsm* (*N*-nitrosomorpholine, CAS 59-89-2, Table S1), which has a rat TD₅₀ of 0.135 mg/kg (i.e., Cat 1 COC). It is unlikely that the reactivity of the nitrosamine moiety is significantly affected by the swap of oxygen for sulfur; however, it could be proposed that a competing β-hydroxylation is more favorable for *ntm* over *nsm* owing to better stabilization of the radical intermediate by the electron-diffuse sulfur, leading to lower toxicity of *ntm*.⁷³ *N*-nitrosotrifluoroethylethylamine (*nte*) is an interesting case of a molecule that is σ-electron-withdrawing by the trifluoro group (CF₃ is roughly on par with Cl in electronegativity), while it is donating via σ(C–F)→*p*(C•) orbital mixing to stabilize the intermediate radical. Concurrently, the trifluoro group has an electron cloud that shields the α-C position from hydroxylation. Assuming robustness of experimental data, destabilizing and shielding interactions appear to 'win out', as *nte* is one and two orders of magnitude less potent than *dpm* and *den*, respectively. Lastly, *npe* is underpredicted by our LDA model. Here, the aromatic ring can undergo oxidation by P450 enzymes to an arene oxide, generating an additional mutagenic site in the molecule. Because of the lack of reliable animal data for complex nitrosamines, the present model's assessment of carcinogenic potency focuses only on the *N*-nitrosamine moiety.

In external testing, we identified 6 outliers (Figure 9B). The 1-[(2,3-dihydroxypropyl)nitrosoamino]-2-propanone (**CAS 92177-50-9**, $TD_{50} = 0.0352$ mg/kg) was underpredicted as Cat 2 COC. This toxicant is complicated by the hyperconjugation effects involving the β -carbonyl and β -OH groups, which can stabilize the radical intermediate in the hydroxylation step; a compound effect that appears to be underestimated in the model. Additionally, diols can cyclize to form mutagenic epoxides, which, though less likely under mild conditions or *in vivo*, is a transformation not captured by the model. Ethylnitrosocyanamide (**ecy**, $TD_{50} = 3.68$ mg/kg) was overpredicted as Cat 2 COC. While cyanamides are generally not mutagenic or carcinogenic,⁷⁴ the conjugation effects across the cyano and nitrosamine groups in **ecy** might be further decreasing the α -hydroxylation potential. Nitrosoheptamethyleneimine (**nhm**, $TD_{50} = 0.0383$ mg/kg) was underpredicted as Cat 3 COC (the only large misprediction in the dataset). We postulated that in the series of cyclic alkyl *N*-nitrosamines, the conformational flexibility of the larger analogs may result in closer positioning of the β -OH group to the nitrosamine for proton transfer in the heterolysis step (Figure 5C). However, this effect is likely underestimated by the model given the concurrent trend of decreasing potency in larger (and more polarizable) alkyl nitrosamines (Figure 4). *N*-nitrosotetrahydropyridine (**nhd**, $TD_{50} = 0.0599$ mg/kg) was underpredicted as Cat 2 COC, possibly due to the epoxidation on the double bond,⁷⁵ which generates an additional mutagenic site (viz. discussion of **npe** above). **Nde** ($TD_{50} = 5.98$ mg/kg) is a close analog of **CAS 92177-50-9** ($TD_{50} = 0.0352$ mg/kg) but of much lower potency. The difference in reactivity of the two compounds can be explained by the effect of the α -carbonyl group on the hydroxylation step via π - p orbital mixing. Lastly, 1-Nitroso-1-(2-chloroethyl)-3-(2-hydroxypropyl)urea (**CAS 106612-15-1**, $TD_{50} = 0.124$ mg/kg) was underpredicted as Cat 2 COC, though the measured TD_{50} value is very close (0.03 mg/kg) to the Cat 1/2 cutoff. A possible explanation could be the mutagenic site of alkyl chloride, increasing the potency of this compound.⁷⁶

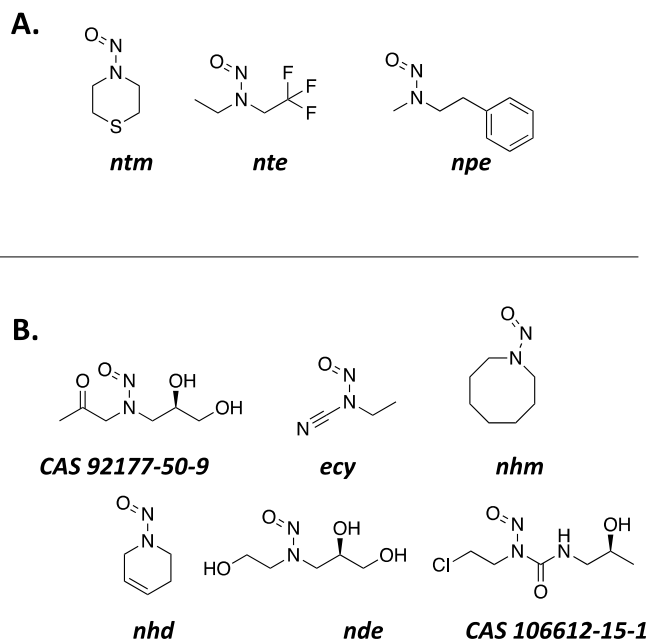


Figure 9. *N*-nitrosamines under- and over-predicted by the current LDA model in the training set (A) and the test set (B).

Application to EMA compounds. To offer immediate utility, we applied the present models to the complete list of nitrosamines in medicinal products, as compiled by EMA (Table 3).^{2,3} From Table 3, predicted potency categories are generally aligned with measured values where available, and are by and large less conservative than the applied AI limits. The latter is simply a consequence of the current AI limits, which are derived mostly from NDEA (CAS 55-18-5) or NDMA (CAS 62-75-9), despite these compounds being (bio)chemically (i.e., mechanistically) different from much of the rest of the dataset. In some cases, calculated TD₅₀'s or structurally relevant analogs were used toward AI limits for compounds in Table 3. For example, *N*-nitrosomorpholine (CAS 59-89-2) has an AI limit of 127 ng/day, which corresponds to both computed TD₅₀ from LCDB (0.135 mg/kg) and our predicted category of potency (Cat 1, TD₅₀ ≤ 0.15 mg/kg). Similarly, *N*-nitroso-*N*-methylaniline (CAS 614-00-6) has AI limit that corresponds to both measured and predicted potency categories (Cat 1).

Table 3. A list of *N*-nitrosamines commonly found in medicinal products, with AI limits provided by the European Medicines Agency (EMA) and corresponding predictions of carcinogenic potency determined with the CADRE program. AI = daily acceptable intake limit; SAVA = solvent-accessible volume area; LCDB = Lhasa Carcinogenicity Database, CPDB (Gold Carcinogenic Potency Database); Potency categories: Cat 1 ($TD_{50} \leq 0.15$ mg/kg), Cat 2 ($0.15 < TD_{50} \leq 1.5$ mg/kg) and Cat 3 ($TD_{50} > 1.5$ mg/kg).

<i>N</i>-nitrosamine (CAS number)	AI (ng/day)	TD₅₀ (mg/kg/day)	Source	Measured potency	Predicted potency
N-Nitrosodimethylamine (62-75-9)	96.0	0.177	LCDB	2	2
N-nitrosodiethylamine (55-18-5)	26.5	0.0177	LCDB	1	1
N-nitrosoethylisopropylamine (16339-04-1)	26.5	N/A	-	-	1
N-nitrosodiisopropylamine (601-77-4)	26.5	N/A	-	-	2
N-nitroso-N-methyl-4-aminobutyric acid (61445-55-4)	96.0	0.982	CPDB	2	2
1-Methyl-4-nitrosopiperazine (16339-07-4)	26.5	N/A	-	-	2
N-Nitroso-di-n-butylamine (924-16-3)	26.5	0.691	CPDB	2	2
N-nitroso-N-methylaniline (614-00-6)	34.3	0.106	LCDB	1	1
N-nitroso-morpholine (59-89-2)	127	0.135	LCDB	1	1
N-nitroso-varenicline (2755871-02-2)	37.0	N/A	-	-	3
N-nitrosodipropylamine (621-64-7)	26.5	0.186	CPDB	2	2
N-nitrosomethylphenidate (55557-03-4)	1300	N/A	-	-	3
N-nitrosopiperidine (100-75-4)	1300	1.12	LCDB	2	2
N-nitrosorasagilene (2470278-90-9)	18	N/A	-	-	2
7-Nitroso-3-(trifluoromethyl)-5,6,7,8-tetrahydro[1,2,4]triazolo[4,3- <i>a</i>]pyrazine	37	N/A	-	-	3
N-nitroso-1,2,3,6-tetrahydropyridine (55556-92-8)	37	0.0599	LCDB	1	2
N-nitrosonortriptyline (55855-42-0)	8		-	-	3
N-methyl-N-nitrosophenethylamine, (13256-11-6)	8	0.00797	LCDB	1	2

For SAR/read-across using structurally similar compounds, the AI limit of *N*-nitroso-varenicline (CAS 2755871-02-2), 37 ng/day, was determined using *N*-nitroso-1,2,3,6-tetrahydropyridine (CAS 55556-92-8). Though these compounds may appear ‘similar’, they are chemically different due to the presence of the double bond in the latter. We discussed an analogous case in Figure 3, contrasting *N*-nitroso-1,2,3,6-tetrahydropyridine (*nhd*) vs. *nsd*, where radical stabilization by π -resonance in the former leads to greater toxicity ($TD_{50} = 0.059$ mg/kg) over the latter ($TD_{50} =$

1.12 mg/kg). For the same reason, CADRE predicts *N*-nitroso-varenicline to be less potent than *N*-nitroso-1,2,3,6-tetrahydropyridine (Table 3). Comparably, while *N*-methyl-*N*-nitrosophenethylamine (CAS 13256-11-6) may appear to be a close structural analog of *N*-nitrosonortriptyline (55855-42-0), driving the AI limit for the latter (8 ng/day), the two compounds behave differently. In CADRE's Monte Carlo simulations, we observed that the ring π -conjugation is responsible for different conformational preference of the CH₃-N-CH₂-CH₂ dihedral angle in *N*-methyl-*N*-nitrosophenethylamine (*anti*) vs. *N*-nitrosonortriptyline (*syn*), resulting in the latter α -C being less accessible to P450 hydroxylation (Figure 10). Consequently, CADRE predicts the latter compound as a less potent carcinogen. In sum, assessments of structural similarity are not enough to ensure useful derivations of AI limits; we need to understand and capture the underlying chemistry of N-nitrosamine metabolism.

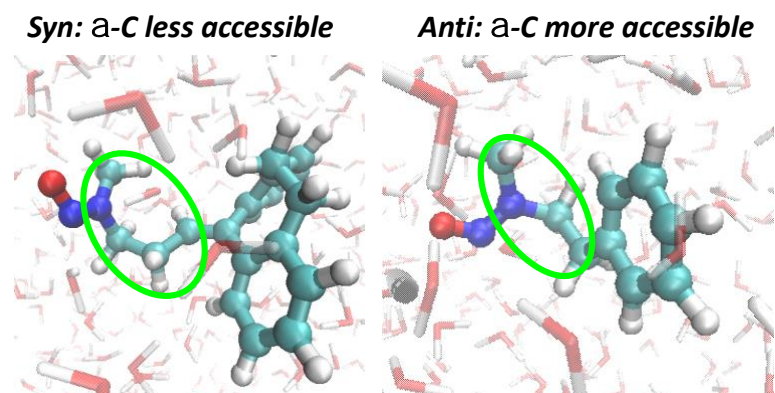


Figure 10. QM/MM/MC simulations used to capture the conformational difference in the CH₃-N-CH₂-CH₂ dihedral angle (marked in neon green color) for *N*-nitrosonortriptyline (*syn*) and *N*-methyl-*N*-nitrosophenethylamine (*anti*), resulting in the latter α -C being more accessible to P450 hydroxylation and thus more carcinogenic.

Future development. Despite indicated robustness of the present CADRE approach, any models that describe complex biochemical phenomena using limited data will have inherent drawbacks. While we can assess QM factors to 'correct' for outliers (viz. complex

hyperconjugation effects), we are constrained by the training-set size in terms of the number of descriptors we can implement to prevent overfitting. To improve predictivity in the short term, we envision a tiered approach for complex nitrosamines that uses the current LDA models in tandem with electronic-structure readacross, which can provide refinement based on expert evaluation of additional QM descriptors. In the long term, models can be advanced if more (reliable) TD₅₀ data, especially for complex nitrosamines, becomes available. This is not necessarily a call for more animal testing; rather, the scientific community should strive to improve data-sharing practices, allowing modelers to securely tap into existing proprietary data held by pharmaceutical companies to address both regulatory and new chemical development needs.^{14, 77} While the appetite for data 'openness' is growing,³¹ we need a systemic change to truly take advantage of our collective resources.⁷⁷

Conclusion. The use of SARs has been instrumental to our understanding of the steric and electronic factors contributing to the activation of *N*-nitrosamine pro-mutagens to carcinogenic compounds.⁴ However, direct use of SARs in predicting potency of *N*-nitrosamines is impractical due to the many confounding factors, the sheer complexity of the biochemical process and the limited dataset of reliable TD₅₀'s to train such models.^{4, 78} On the other hand, the use of QSARs based on physicochemical properties can be grossly misleading given a dubious link between these descriptors and the mechanism of action. Since covalent interactions dominate *N*-nitrosamine metabolism, the solution appears to be in the development of robust quantum-mechanical approaches based on electronic structure. Underpinned and validated using SARs, these methods can alleviate concerns about limited data by capturing the underlying chemistry without over-reliance on chemical structures.

Here, we outlined the extension of the QM-based CADRE platform to *N*-nitrosamines, with the goal of creating a predictive tool that is both robust and practically useful. We accomplished the latter by reducing reaction-pathway modeling, which is unfeasible in routine

hazard assessments, into key steric and electronic factors, and by recognizing that the low reproducibility of rodent carcinogenicity bioassays poses natural limitations on what a predictive model can accomplish. To this end, CADRE only offers categorical predictions, aimed at distinguishing the most hazardous nitrosamines from less potent COC and from no-concern compounds. External validation showed this can be accomplished with ca. 78% accuracy, though real-world metrics are expected to be somewhat lower. Our survey of model outliers and the relationship between out-of-domain descriptors and prediction accuracy offers insight into future development, while boosting confidence in applying the current model to complex *N*-nitrosamine impurities found in medicinal products. To this end, we should emphasize that the present model focuses on the contributions of the *N*-nitrosamine group to the compound's carcinogenic potency. For chemicals containing other mutagenic toxicophores, CADRE should be paired with other tools that assess this endpoint broadly to determine their overall hazard.

ASSOCIATED CONTENT

Supporting Information. Supporting Information Available: a complete list of compounds, data source, SMILES strings, experimental carcinogenic potency (TD₅₀) and predicted categories of carcinogenic potency. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Corresponding author

*Address corresponding to Jakub Kostal, Designing Out Toxicity (DOT) Consulting LLC, 2121 Eisenhower Avenue, Alexandria, VA 22314, United States. Email: jakub@toxfix.com

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

References

1. Ruepp, R.; Froetschl, R.; Bream, R.; Filancia, M.; Girard, T.; Spinei, A.; Weise, M.; Whomsley, R., The EU Response to the Presence of Nitrosamine Impurities in Medicines. *Front Med-Lausanne* **2021**, *8*.
2. European Medicines Agency. *Nitrosamine Impurities Report 2020*.
3. European Medicines Agency. *Temporary interim limits for NMBA, DIPNA, EIPNA, impurities in sartan blood pressure medicines; 2019*.
4. Cross, K. P.; Ponting, D. J., Developing Structure-Activity Relationships for N-Nitrosamine Activity. *Comput Toxicol* **2021**, *20*.
5. Thresher, A.; Foster, R.; Ponting, D. J.; Stalford, S. A.; Tennant, R. E.; Thomas, R., Are all nitrosamines concerning? A review of mutagenicity and carcinogenicity data. *Regul Toxicol Pharm* **2020**, *116*.
6. Thomas, R.; Thresher, A.; Ponting, D. J., Utilisation of parametric methods to improve percentile-based estimates for the carcinogenic potency of nitrosamines. *Regul Toxicol Pharm* **2021**, *121*.
7. Buist, H. E.; Devito, S.; Goldbohm, R. A.; Stierum, R. H.; Venhorst, J.; Kroese, E. D., Hazard assessment of nitrosamine and nitramine by-products of amine-based CCS: Alternative approaches. *Regul Toxicol Pharm* **2015**, *71* (3), 601-623.
8. Luan, F.; Zhang, R. S.; Zhao, C. Y.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T., Classification of the carcinogenicity of N-nitroso compounds based on support vector machines and linear discriminant analysis. *Chemical research in toxicology* **2005**, *18* (2), 198-203.
9. Helguera, A. M.; Gonzalez, M. P.; Cordeiro, M. N. D. S.; Perez, M. A. C., Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. *Toxicol Appl Pharm* **2007**, *221* (2), 189-202.
10. Helguera, A. M.; Cordeiro, M. N. D. S.; Perez, M. A. C.; Combes, R. D.; Gonzalez, M. P., Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds - Species: Rat; Sex: Male; Route of administration: Water. *Toxicol Appl Pharm* **2008**, *231* (2), 197-207.
11. Helguera, A. M.; Perez-Machado, G.; Cordeiro, M. N. D. S.; Combes, R. D., Quantitative structure-activity relationship modelling of the carcinogenic risk of nitroso compounds using regression analysis and the TOPS-MODE approach. *Sar and Qsar in Environmental Research* **2010**, *21* (3-4), 277-304.
12. MultiCASE QSAR Flex. *A novel software tool for in silico risk assessment of chemicals*. https://www.multicase.com/qsar-flex#mu_features_som_pred.
13. Tropsha, A., Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics* **2010**, *29* (6-7), 476-488.

14. Kostal, J.; Voutchkova-Kostal, A., Going All In: A Strategic Investment in In Silico Toxicology. *Chemical research in toxicology* **2020**.
15. Kostal, J.; Voutchkova-Kostal, A., CADRE-SS, an in Silico Tool for Predicting Skin Sensitization Potential Based on Modeling of Molecular Interactions. *Chemical research in toxicology* **2016**, *29* (1), 58-64.
16. Wenzel, J.; Schmidt, F.; Blumrich, M.; Amberg, A.; Czich, A., Predicting DNA-Reactivity of N-Nitrosamines: A Quantum Chemical Approach. *Chemical research in toxicology* **2022**, ASAP.
17. Li, K. T.; Ricker, K.; Tsai, F. C.; Hsieh, C. Y. J.; Osborne, G.; Sun, M.; Marder, M. E.; Elmore, S.; Schmitz, R.; Sandy, M. S., Estimated Cancer Risks Associated with Nitrosamine Contamination in Commonly Used Medications. *Int J Env Res Pub He* **2021**, *18* (18).
18. Lijinsky, W., Structure-Activity Relations in Carcinogenesis by N-Nitroso Compounds. *Cancer Metast Rev* **1987**, *6* (3), 301-356.
19. Andreozzi, P.; Klopman, G.; Hopfinger, A. J., Theoretical-Study of N-Nitrosamines and Their Presumed Proximate Carcinogens. *Cancer Biochem Bioph* **1980**, *4* (4), 209-220.
20. Poulsen, M.; Spangler, D.; Loew, G. H., Nitrosamine carcinogen activation pathway determined by quantum chemical methods. *Mol Toxicol* **1987**, *1* (1), 35-47.
21. Lonsdale, R.; Houghton, K. T.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J., Quantum mechanics/molecular mechanics modeling of regioselectivity of drug metabolism in cytochrome P450 2C9. *J Am Chem Soc* **2013**, *135* (21), 8001-15.
22. Kroes, R.; Renwick, A. G.; Cheeseman, M.; Kleiner, J.; Mangelsdorf, I.; Piersma, A.; Schilter, B.; Schlatter, J.; van Schothorst, F.; Vos, J. G.; Wurtzen, G., Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food and Chemical Toxicology* **2004**, *42* (1), 65-83.
23. Phillipson, C. E.; Ioannides, C., A Comparative-Study of the Bioactivation of Nitrosamines to Mutagens by Various Animal Species Including Man. *Carcinogenesis* **1984**, *5* (8), 1091-1094.
24. Lewis, D. F. V.; Ito, Y., Human CYPs involved in drug metabolism: structures, substrates and binding affinities. *Expert Opin Drug Met* **2010**, *6* (6), 661-674.
25. Mochizuki, M.; Anjo, T.; Okada, M., Isolation and Characterization of N-Alkyl-N-(Hydroxymethyl)Nitrosamines from N-Alkyl-N-(Hydroperoxymethyl)Nitrosamines by Deoxygenation. *Tetrahedron Lett* **1980**, *21* (38), 3693-3696.
26. Hecht, S. S., N-Nitroso-2-Hydroxymorpholine, a Mutagenic Metabolite of N-Nitrosodiethanolamine. *Carcinogenesis* **1984**, *5* (12), 1745-1747.
27. Michejda, C. J.; Koepke, S. R.; Kroeger-Koepke, M. B.; Bosan, W., Recent findings on the metabolism of beta-hydroxyalkylnitrosamines. *IARC Sci Publ* **1987**, (84), 77-82.
28. Kostal, J., Computational Chemistry in Predictive Toxicology: Status Quo et Quo Vadis? In *Advances in Molecular Toxicology* Fishbein, J., Ed. Elsevier: San Diego, 2016; Vol. 10, pp 139-186.
29. Melnikov, F.; Kostal, J.; Voutchkova-Kostal, A.; Zimmerman, J. B.; Anastas, P. T., Assessment of predictive models for estimating the acute aquatic toxicity of organic chemicals. *Green Chem* **2016**, *18*, 4432-4445.

30. Kostal, J.; Voutchkova-Kostal, A.; Anastas, P. T.; Zimmerman, J. B., Identifying and designing chemicals with minimal acute aquatic toxicity. *P Natl Acad Sci USA* **2015**, *112* (20), 6289-6294.
31. Graham, J. C.; Trejo-Martin, A.; Chilton, M. L.; Kostal, J.; Bercu, J.; Beutner, G. L.; Bruen, U. S.; Dolan, D. G.; Gomez, S.; Hillegass, J.; Nicolette, J.; Schmitz, M., An Evaluation of the Occupational Health Hazards of Peptide Couplers. *Chemical research in toxicology* **2022**.
32. Roland, C. D.; Moore, C. M.; Leal, J. H.; Semelsberger, T. A.; Snyder, C.; Kostal, J.; Sutton, A. D., Fully Recyclable Polycarbonates from Simple, Bio-Derived Building Blocks. *ACS Appl Polym Mater* **2021**, *3* (2), 730-736.
33. Thakore, R. R.; Takale, B. S.; Hu, Y. T.; Ramer, S.; Kostal, J.; Gallou, F.; Lipshutz, B. H., "TPG-lite": A new, simplified "designer" surfactant for general use in synthesis under micellar catalysis conditions in recyclable water. *Tetrahedron* **2021**, *87*.
34. Voutchkova-Kostal, A.; Vaccaro, S.; Kostal, J., Computer-Aided Discovery and Redesign (CADRE) for Respiratory Sensitization: A Tiered Mechanistic Model to Deliver Robust Performance across a Diverse Chemical Space. *Chem Res Toxicol*, 2022.
35. Kazius, J.; McGuire, R.; Bursi, R., Derivation and validation of toxicophores for mutagenicity prediction. *Journal of Medicinal Chemistry* **2005**, *48* (1), 312-320.
36. Gold, L. S.; Sawyer, C. B.; Magaw, R.; Backman, G. M.; Deveciana, M.; Levinson, R.; Hooper, N. K.; Havender, W. R.; Bernstein, L.; Peto, R.; Pike, M. C.; Ames, B. N., A Carcinogenic Potency Database of the Standardized Results of Animal Bioassays. *Environ Health Persp* **1984**, *58* (Dec), 9-319.
37. Gold, L. S.; Slone, T. H.; Manley, N. B.; Garfinkel, G. B.; Hudes, E. S.; Rohrbach, L.; Ames, B. N., The Carcinogenic Potency Database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/National Toxicology Program. *Environ Health Perspect* **1991**, *96*, 11-5.
38. Thresher, A.; Gosling, J. P.; Williams, R., Generation of TD50 values for carcinogenicity study data. *Toxicol Res-Uk* **2019**, *8* (5), 696-703.
39. Chu, C.; Magee, P. N., Metabolic-Fate of Nitrosoproline in the Rat. *Cancer research* **1981**, *41* (9), 3653-3657.
40. Greenblatt, M.; Lijinsky, W., Failure to induce tumors in Swiss mice after concurrent administration of amino acids and sodium nitrite. *J Natl Cancer Inst* **1972**, *48* (5), 1389-92.
41. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *J Cheminformatics* **2011**, *3*.
42. Lynch, B. J.; Truhlar, D. G., Small basis sets for calculations of barrier heights, energies of reaction, electron affinities, geometries, and dipole moments. *Theor Chem Acc* **2004**, *111* (2), 335-344.
43. Team, R. D. C. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2009.
44. Myatt, G. J.; Ahlberg, E.; Akahori, Y.; Allen, D.; Amberg, A.; Anger, L. T.; Aptula, A.; Auerbach, S.; Beilke, L.; Bellion, P.; Benigni, R.; Bercu, J.; Booth, E. D.; Bower, D.; Brigo, A.; Burden, N.; Cammerer, Z.; Cronin, M. T. D.; Cross, K. P.; Custer, L.; Dettwiler, M.; Dobo, K.; Ford, K. A.; Fortin, M. C.; Gad-McDonald, S. E.; Gellatly, N.; Gervais, V.; Glover, K. P.; Glowienke, S.; Van Gompel, J.; Gutsell, S.; Hardy, B.; Harvey, J. S.; Hillegass, J.; Honma, M.; Hsieh, J. H.; Hsu, C. W.; Hughes, K.; Johnson, C.; Jolly, R.; Jones, D.; Kemper, R.; Kenyon, M.

- O.; Kim, M. T.; Kruhlak, N. L.; Kulkarni, S. A.; Kummerer, K.; Leavitt, P.; Majer, B.; Masten, S.; Miller, S.; Moser, J.; Mumtaz, M.; Muster, W.; Neilson, L.; Oprea, T. I.; Patlewicz, G.; Paulino, A.; Lo Piparo, E.; Powley, M.; Quigley, D. P.; Reddy, M. V.; Richarz, A. N.; Ruiz, P.; Schilter, B.; Serafimova, R.; Simpson, W.; Stavitskaya, L.; Stidl, R.; Suarez-Rodriguez, D.; Szabo, D. T.; Teasdale, A.; Trejo-Martin, A.; Valentin, J. P.; Vuorinen, A.; Wall, B. A.; Watts, P.; White, A. T.; Wichard, J.; Witt, K. L.; Woolley, A.; Woolley, D.; Zwickl, C.; Hasselgren, C., In silico toxicology protocols. *Regul Toxicol Pharm* **2018**, *96*, 1-17.
45. Cronin, M. T. D.; Enoch, S. J.; Hewitt, M.; Madden, J. C., Formation of Mechanistic Categories and Local Models to Facilitate the Prediction of Toxicity. *Altex-Altern Anim Ex* **2011**, *28* (1), 45-49.
46. Fukui, K.; Yonezawa, T.; Shingu, H., A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *The Journal of Chemical Physics* **1952**, *20* (4), 722-725.
47. Lewis, D. F. V., Frontier orbitals in chemical and biological activity: Quantitative relationships and mechanistic implications. *Drug Metab Rev* **1999**, *31* (3), 755-816.
48. Lopachin, R. M.; Gavin, T.; Decaprio, A.; Barber, D. S., Application of the Hard and Soft, Acids and Bases (HSAB) theory to toxicant--target interactions. *Chemical research in toxicology* **2012**, *25* (2), 239-51.
49. Bulat, F. A.; Chamorro, E.; Fuentealba, P.; Toro-Labbe, A., Condensation of frontier molecular orbital Fukui functions. *J Phys Chem A* **2004**, *108* (2), 342-349.
50. Goerigk, L.; Grimme, S., A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys Chem Chem Phys* **2011**, *13* (14), 6670-6688.
51. Toro-Labbe, A., Characterization of chemical reactions from the profiles of energy, chemical potential and hardness. *J Phys Chem A* **1999**, *103* (22), 4398-4403.
52. Goetz, M. K.; Schneider, J. E.; Filatov, A. S.; Jesse, K. A.; Anderson, J. S., Enzyme-Like Hydroxylation of Aliphatic C-H Bonds From an Isolable Co-Oxo Complex. *J Am Chem Soc* **2021**, *143* (49), 20849-20862.
53. de Lomana, M. G.; Svensson, F.; Volkamer, A.; Mathea, M.; Kirchmair, J., Consideration of predicted small-molecule metabolites in computational toxicology. *Digital Discovery* **2022**, *1*, 158-172.
54. Maggiora, G. M., On outliers and activity cliffs - Why QSAR often disappoints. *J Chem Inf Model* **2006**, *46* (4), 1535-1535.
55. Li, Y. P.; Hecht, S. S., Metabolic Activation and DNA Interactions of Carcinogenic N-Nitrosamines to Which Humans Are Commonly Exposed. *International Journal of Molecular Sciences* **2022**, *23* (9).
56. Moharana, M.; Sahu, S. N.; Pattanayak, S. K., Natural Bond Orbital Analysis of creatinine: A DFT Approach Study. *Aip Conf Proc* **2019**, 2142.
57. Krishtalik, L. I., The mechanism of the proton transfer: an outline. *Bba-Bioenergetics* **2000**, *1458* (1), 6-27.
58. Helguera, A. M.; Gonzalez, M. P.; Cordeiro, M. N. D. S.; Perez, M. A. C., Quantitative structure - Carcinogenicity relationship for detecting structural alerts in nitroso compounds: Species, rat; Sex, female; Route of administration, Gavage. *Chemical research in toxicology* **2008**, *21* (3), 633-642.

59. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C. H.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A., QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry* **2014**, *57* (12), 4977-5010.
60. Wang, D. F.; Wiest, O.; Helquist, P.; Lan-Hargest, H. Y.; Wiech, N. L., QSAR Studies of PC-3 cell line inhibition activity of TSA and SAHA-like hydroxamic acids. *Bioorg Med Chem Lett* **2004**, *14* (3), 707-711.
61. Praba, G. O.; Velmurugan, D., Quantitative structure-activity relationship of some pesticides. *Indian J Biochem Bio* **2007**, *44* (6), 470-476.
62. Hansch, C.; Kurup, A., QSAR of chemical polarizability and nerve toxicity. 2. *J Chem Inf Comp Sci* **2003**, *43* (5), 1647-1651.
63. Gottmann, E.; Kramer, S.; Pfahringer, B.; Helma, C., Data quality in predictive toxicology: Reproducibility of rodent carcinogenicity experiments. *Environ Health Persp* **2001**, *109* (5), 509-514.
64. Puzyn, T.; Mostrag-Szlichtyng, A.; Gajewicz, A.; Skrzynski, M.; Worth, A. P., Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. *Struct Chem* **2011**, *22* (4), 795-804.
65. Krewski, D.; Gaylor, D. W.; Soms, A. P.; Szyszkowicz, M., An Overview of the Report - Correlation between Carcinogenic Potency and the Maximum Tolerated Dose - Implications for Risk Assessment. *Risk Anal* **1993**, *13* (4), 383-398.
66. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R., Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* **2012**, *17* (5), 4791-4810.
67. Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T., QSAR applicability domain estimation by projection of the training set descriptor space: a review. *Altern Lab Anim* **2005**, *33* (5), 445-59.
68. Kar, S.; Roy, K.; Leszczynski, J., Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling. *Methods Mol Biol* **2018**, *1800*, 141-169.
69. Amberg, A.; Andaya, R. V.; Anger, L. T.; Barber, C.; Beilke, L.; Bercu, J.; Bower, D.; Brigo, A.; Cammerer, Z.; Cross, K. P.; Custer, L.; Dobo, K.; Gerets, H.; Gervais, V.; Glowienke, S.; Gomez, S.; Gompel, J.; Harvey, J.; Hasselgren, C.; Honma, M.; Johnson, C.; Jolly, R.; Kemper, R.; Kenyon, M.; Kruhlak, N.; Leavitt, P.; Miller, S.; Muster, W.; Naven, R.; Nicolette, J.; Parenty, A.; Powley, M.; Quigley, D. P.; Reddy, M. V.; Sasaki, J. C.; Stavitskaya, L.; Teasdale, A.; Trejo-Martin, A.; Weiner, S.; Welch, D. S.; White, A.; Wichard, J.; Woolley, D.; Myatt, G. J., Principles and procedures for handling out-of-domain and indeterminate results as part of ICH M7 recommended (Q)SAR analyses. *Regul Toxicol Pharm* **2019**, *102*, 53-64.
70. Rakhimbekova, A.; Madzhidov, T. I.; Nugmanov, R. I.; Gimadiev, T. R.; Baskin, I. I.; Varnek, A., Comprehensive Analysis of Applicability Domains of QSPR Models for Chemical Reactions. *International Journal of Molecular Sciences* **2020**, *21* (15).
71. Schwarze, M.; Schellhammer, K. S.; Ortstein, K.; Benduhn, J.; Gaul, C.; Hinderhofer, A.; Toro, L. P.; Scholz, R.; Kublitski, J.; Roland, S.; Lau, M.; Poelking, C.; Andrienko, D.; Cuniberti, G.; Schreiber, F.; Neher, D.; Vandewal, K.; Ortman, F.; Leo, K., Impact of molecular

quadrupole moments on the energy levels at organic heterojunctions. *Nature Communications* **2019**, *10*.

72. Boča, R., *Theoretical Foundations of Molecular Magnetism*. Elsevier Science: 1999; Vol. 1, p 873.

73. Pitoj, M. M.; Patterson, B. M.; Furness, A. J.; Bastow, T. P.; McKinley, A. J., Fate of N-nitrosomorpholine in an anaerobic aquifer used for managed aquifer recharge: A column study. *Water Res* **2011**, *45* (8), 2550-2560.

74. Menargues, A.; Obach, R.; Valles, J. M., An Evaluation of the Mutagenic Potential of Cyanamide Using the Micronucleus Test. *Mutation research* **1984**, *136* (2), 127-129.

75. Kostal, J.; Voutchkova-Kostal, A.; Weeks, B.; Zimmerman, J. B.; Anastas, P. T., A free energy approach to the prediction of olefin and epoxide mutagenicity and carcinogenicity. *Chemical research in toxicology* **2012**, *25* (12), 2780-7.

76. Haranosono, Y.; Ueoka, H.; Kito, G.; Nemoto, S.; Kurata, M.; Sakaki, H., A reaction mechanism-based prediction of mutagenicity: alpha-halo carbonyl compounds adduct with DNA by S(N)2 reaction. *J Toxicol Sci* **2018**, *43* (1-3), 203-211.

77. Kostal, J.; Brooks, B. W.; Smith, C. A.; Devineni, G., O data, where art thou? Revolutionizing data sharing to advance our sustainability goals through smart chemical innovation. *Isience* **2022**, *25* (18), 105256.

78. Thomas, R.; Tennant, R. E.; Oliveira, A. A. F.; Ponting, D. J., What Makes a Potent Nitrosamine? Statistical Validation of Expert-Derived Structure–Activity Relationships. *Chemical research in toxicology* **2022**, ASAP.