This article can be cited before page numbers have been issued, to do this please use: T. Bond, A. Simperler, N. Graham, L. Li, W. Gan, X. Yang and M. Templeton, Environ. Sci.: Water Res. Technol., 2017, DOI: 10.1039/C7EW00068E.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the author guidelines.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the ethical guidelines, outlined in our author and reviewer resource centre, still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Water impact statement

Computational chemistry methods were used to define the molecular properties of reactive NDMA precursors. The knowledge gained informed a search for previously unknown NDMA precursors. These included rivastigmine, a therapeutic, and conessine, a naturally occurring species, whose experimental NDMA yields were found to be 83.3±0.5% and 42.3±1.8% mol/mol, respectively.
Computational descriptors were used to identify previously unknown NDMA precursors
Defining the molecular properties of N-nitrosodimethylamine (NDMA) precursors using computational chemistry

Tom Bond\textsuperscript{1,2}*, Alexandra Simperler\textsuperscript{3}, Nigel Graham\textsuperscript{2}, Li Ling\textsuperscript{4}, Wenhui Gan\textsuperscript{5}, Xin Yang\textsuperscript{5} and Michael R. Templeton\textsuperscript{3}

1. Department of Civil and Environmental Engineering, University of Surrey, Guildford, GU2 7XH, UK
2. Department of Civil and Environmental Engineering, Imperial College London, South Kensington campus, London, SW7 2AZ, UK
3. NSCCS, Department of Chemistry, Imperial College London, South Kensington campus, London, SW7 2AZ, UK
4. Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
5. School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China

*Corresponding author: email: t.bond@surrey.ac.uk. Phone: +44(0)7709565170
Abstract

*N*-nitrosodimethylamine (NDMA) is a potent carcinogen and can be produced during chloramination of drinking water and wastewater. Computational chemistry methods were used for the first time to calculate molecular descriptors for 64 NDMA precursors containing a dimethylamine (DMA) moiety. Descriptors were partial charge, bond length and $pK_a$ of the DMA nitrogen and planarity of the DMA group. Precursors classified on the basis of chemical functionality showed distinct relationships between partial charge and NDMA formation. Quaternary amines and tertiary amines with the DMA bonded to –COR and –CSR groups had a combination of low NDMA formation and high partial charge. The most potent NDMA precursors are tertiary amines with an acidic hydrogen and electron-donating group α and β to the DMA respectively. They also have comparable molecular descriptors: relatively negative partial charges, low planarity values, high bond lengths and $pK_a$ values from ~8.3-10.1. A literature search identified 233 potential NDMA precursors that have never been tested experimentally. Of these chemicals 60% are therapeutics, 13% veterinary therapeutics and 10% natural products. Analysis combining qualitative assessment of chemical functionality and computational calculation of molecular descriptors successfully identified rivastigmine, a therapeutic, and conessine, a naturally occurring species, whose NDMA yields were determined experimentally to be 83.3±0.5% and 42.3±1.8% mol/mol, respectively. This study defines the molecular properties associated with reactive NDMA precursors and the origin and identity of those amines which contribute to NDMA formation in drinking water.

**KEYWORDS** nitrosamine; precursor; disinfection; chloramine; drinking water; quantum chemical descriptors
1. Introduction

The nitrosamines are a family of potent human carcinogens. \(^1\) \(N\)-nitrosodimethylamine (NDMA, \((\text{H}_3\text{C})_2\text{NNO})\) is the most frequently encountered species in drinking water and municipal wastewater and California’s Office of Environmental Health Hazard Assessment (OEHHA) has set a 3 ng·L\(^{-1}\) public health goal for the compound in drinking water. \(^2\) Early work on the chemistry of nitrosamine formation concentrated on the acid-catalysed nitrosation of amines. \(^1\) Subsequently, in the 1980s and 1990s, NDMA was detected in Canadian \(^3\) and Californian \(^4\) drinking water sources. These occurrences were linked to industrial contamination and it was the discovery, by Choi and Valentine in 2002 \(^5\), that NDMA could be formed from reactions between monochloramine (\(\text{NH}_2\text{Cl}\)) and the secondary amine dimethylamine (DMA, \((\text{H}_3\text{C})_2\text{NH}\)) during water treatment that stimulated a wealth of research on this subject.

From analysis of water and wastewater samples, it has become apparent that NDMA formation is associated with chloramination, rather than chlorination, as a disinfection method, as well as wastewater influence in drinking water sources. \(^6\), \(^7\) Initial studies on NDMA formation in drinking water focused on reactions between DMA and monochloramine. While DMA is the most obvious NDMA precursor, its transformation to NDMA is rather inefficient, with molar yields of 0.5–2.7% mol/mol reported. \(^8\)-\(^12\) However, studies with a variety of tertiary amines containing a DMA moiety have identified a range of NDMA precursors which are far more reactive than DMA. For example, the pharmaceutical ranitidine is converted to NDMA by chloramination with yields of 63–94% mol/mol. \(^10\), \(^12\)-\(^16\)

This in itself is interesting, because it implies there is an alternative formation pathway which does not involve DMA. The study of Le Roux et al. \(^17\) suggested that ranitidine is a better nucleophile than DMA and hence NDMA formation proceeds via nucleophilic attack. Similarly, Selbes et al. \(^13\) reported that amines with electron-donating substituents
preferentially formed NDMA relative to those with electron-withdrawing substituents. However, in neither case were these ideas quantitatively supported. A number of different mechanisms have been proposed to explain NDMA formation during chloramination\textsuperscript{5, 12, 17-19}. This signifies that the chemistry of NDMA formation is complex, with multiple mechanisms potentially operative during chloramination of water and wastewater. There is also a high degree of uncertainty regarding why certain tertiary amines are far more reactive than DMA in generating NDMA. The first objective of this study was to employ computational chemistry to quantify the properties of the DMA moiety in reactive NDMA precursors. Computational methods have been used to propose a new NDMA formation mechanism from tertiary amines involving direct nitrosation\textsuperscript{19}. However, computational techniques have not been deployed to provide data about the molecular properties of NDMA precursors. Since many reactive NDMA precursors are complex synthetic compounds, descriptive molecular parameters, such as Hammett and Taft constants, are unavailable\textsuperscript{20}; therefore approaches which allow properties to be calculated are beneficial. The second objective was to validate the approach taken by measuring experimental NDMA yields of potential/untested precursors.

2. Materials and methods

2.1. Precursor Selection

NDMA formation data from 64 precursors were collated from 12 studies\textsuperscript{9-16, 21-25} which reported the monochloramination of a variety of organic amines containing a DMA moiety (Tables 1 and S2). The full dataset is shown in the supplementary information, with data from selected precursors presented in Table 1. Analysed studies used formation potential methodologies, the principle of which is to maximize NDMA yields. Thus, excess monochloramine doses (0.1 – 4.2 mM) and long contact times (1 – 10 days) were employed,
while pH values ranged from 6.8 – 8.5 (see Table S1). Where data on NDMA formation from
the same precursor was reported from multiple studies, the yield reported by Sacher et al\textsuperscript{10}
was taken wherever possible, as this study quantified yields from 25 amines under consistent
conditions (0.4 mM chloramine, pH 7, 20 °C, 168 h). Otherwise the yield measured at
conditions most similar to those of Sacher et al\textsuperscript{10} was selected. For amines studied on
multiple occasions the range of reported NDMA yields are shown in Table 1 and Table S2,
with the value used here given in brackets. Precursors with multiple DMA groups, for
example, DMP3\textsuperscript{21}, minocycline\textsuperscript{14}, 2,4,6-tris(dimethylaminomethyl)phenol\textsuperscript{12} and four
diamines studied by Liu et al\textsuperscript{19}, were excluded. These amines potentially have NDMA yields
in excess of 100% mol/mol, so their yields cannot be directly compared with precursors
containing a single DMA group. Similarly, benzalkonium chloride\textsuperscript{16} is a mixture of
compounds so it was excluded, as was data regarding NDMA formation during reactions of
model amines with ozone\textsuperscript{26} and chlorine dioxide\textsuperscript{16}.

2.2. Computational methods

The following molecular descriptors were calculated: partial charge (of the DMA nitrogen),
planarity (of the DMA moiety), bond length (of (CH\textsubscript{3})\textsubscript{2}NwR) and pK\textsubscript{a} (of the DMA nitrogen)
(Tables 1 and Table S3). The partial charge was selected because the charge density on the
DMA nitrogen is one property suggested as potentially explaining why ranitidine is far more
reactive than other pharmaceuticals or personal care products.\textsuperscript{10, 15} The bond length was
included as it is also related to the presence of electron donating and electron withdrawing
groups. The planarity was selected since it relates to the geometry of the DMA group and
hence the accessibility of the nitrogen lone pair for reaction. It is defined as the angle θ
between the planes C\textsubscript{1}-N-R and C\textsubscript{2}-N-R. It serves as a measure for the planarity of the
N(CH\textsubscript{3})\textsubscript{2} group, whereby 180° means completely planar and values < 180° define the extent
of aplanarity. Finally, \( pK_a \) is a key property for aqueous reactions involving amines since it reflects proportions of the neutral and cationic forms. Furthermore, experimental values are unavailable for many NDMA precursors.

All starting geometries for each structure were generated using GaussView 5.0.9.\(^{27}\) All calculations except those for \( pK_a \) were performed using the Gaussian09 Rev. D.01 suite\(^{28}\) of programmes. Calculations were performed on the unprotonated amine, as the protonated form is unlikely to take part in nucleophilic or electrophilic reactions with chloramines. Geometry optimizations for all compounds were performed in the presence of water, i.e. the molecules (solutestes) are placed in a cavity within the solvent reaction field. The default option in Gaussian 09 was used, which is the Polarizable Continuum Model (PCM), using the integral equation formalism variant (IEFPCM).\(^{29}\) Thereby, the solute cavity is created via a set of overlapping spheres. Gaussian calculates the electron density of the solute using the chosen functional and basis set. An external iteration procedure then calculates the energy in solution. Geometry optimisations for all structures in water were undertaken using the M062X\(^{30}\) functional and 6-311+g(2d,2p) basis set. Frequency calculations were performed on all geometries obtained by the M062X calculations, to characterize them as stationary points (minima) on their potential energy surfaces.

The partial charges on the DMA nitrogen atoms were calculated using the CHELPG (CHarges from ELectrostatic Potentials using a Grid based method) procedure.\(^{31}\) CHELPG belongs to the family of electrostatic potential (ESP) derived charges, which are very reliable in predicting partial charges of polar molecules\(^{32}\), including those with nitrogen atoms.\(^{33}\)

The bond lengths for \((\text{CH}_3)_2\text{N-R}\) and planarity of the DMA moiety (\(\vartheta\)) were obtained from the corresponding geometry-optimized structures with measuring tools in GaussView 5.0.9. Molecular descriptors for benzyldimethyldodecylamine\(^{23}\) were calculated with the
alkyl tail capped after three carbons. This enabled us to produce reliable descriptors but reduce the computational effort.

The pKₐ values were predicted using Jaguar 8.2\textsuperscript{34}, which uses non-empirical DFT calculations post-processed with empirical corrections. Its automatic workflow is based on a simple four-step thermodynamic cycle, as described by Bochevarov et al.\textsuperscript{35}. We used the default, whereby three-steps of this cycle are modelled using the B3LYP functional and 6-31G* \textsuperscript{36} and cc-pVTZ(+)\textsuperscript{37} basis sets, respectively. All calculations were setup via Maestro 9.6\textsuperscript{38}, the graphical user interface provided by Schrödinger. The calculation of pKₐ values was performed in water and with an accurate self-consistent field (SCF) setting. Since the calculation cycle generates a cation and Jaguar will not deal with cationic amines, pKₐ values for Category 5 amines and toluylene blue are unavailable. Finally, correlation analyses were undertaken using SPSS.

2.3 Experimental methods

Experimental NDMA formation potentials were measured using a UPLC/MS/MS system (Agilent 1290 Infinity solvent delivery module and autosampler, with an Agilent G6410B triple quadrupole mass spectrometer (Agilent, USA) with a ZORBAX Eclipse Plus C18 column (50 × 2.1 mm, (1.8 mm)). Samples were injected directly (without solid phase extraction), the method detection limit was 10 µg·L\textsuperscript{-1} and NDMA-d6 was the internal standard. Precursor and monochloramine concentrations were 0.1 mM and 2 mM, respectively. The contact time was 5 days at pH 8.0. Further details are given in Gan et al.\textsuperscript{16}

3. Results

3.1. Partial charge of DMA nitrogen: precursor categories

Of the 64 analysed NDMA precursors only 13 had a NDMA formation potential yield over 5% mol/mol and only seven a yield over 20% mol/mol (Table 1). The majority of precursors,
had NDMA yields under 2% mol/mol (Tables 1 and S2). Thus, only a limited proportion of the tested amines form significant yields of NDMA. The amines analysed were classified into five categories (Table 1). The three Category 5 (quaternary) amines had positive partial charges: 0.27, 0.071, 0.329 for benzylidimethyldecylamine, tetramethylamine and choline respectively, as expected given their cationic nature. This is associated with low NDMA formation: 0.28, 0.0 and 0.16% mol/mol respectively (Tables 1 and S2). The eight Category 4 amines, where the DMA nitrogen is bonded to a heteroatom (either S or N) had partial charges from -0.445 to 0.295 (mean -0.189) and, once again, NDMA formation was low, at less than 1.6% mol/mol (Tables 1 and S2). The 10 compounds in Category 3 ((CH₃)₂N=C= R₁(R₂)) form less than 12.0% mol/mol NDMA (Tables 1 and S2). The most potent precursor is 4-dimethylaminoantipyrine, which is the only amine in Category 3 where the DMA group is bonded to a heterocycle (antipyrine or phenazone) and which has the lowest partial charge of -0.286. In contrast, the eight precursors where the DMA is adjacent to electron-withdrawing -COR or -CSR substituents, have higher partial charges, from -0.225 to 0.220, versus a mean partial charge of -0.300 for the complete dataset. This is associated with low NDMA yields of ≤2.5% mol/mol (Table S2-S3). The 11 Category 2 amines, in which the DMA nitrogen is bonded directly to an aryl (aromatic) group, were relatively unreactive at generating NDMA with yields from 0.06 – 9.4% mol/mol, and the most potent precursor being methyl yellow. This category has a range of negative partial charge values, from -0.18 to -0.557 (mean = -0.368).

The 31 Category 1 amines contain the majority of the more negatively charged compounds, as well the seven most potent NDMA precursors. The mean partial charge is -0.412, the lowest (most negative) of the five categories. It is defined by having the DMA moiety connected via an alkyl group to an R group (Table 1). The seven most potent precursors are N,N-dimethyl-1-(1H-pyrrol-2-yl)methanamine (DMPMA), methadone, N,N-
dimethylthiophene-2-methylamine (or DMthiophene), 5-(dimethylaminomethyl) furfuryl alcohol (DMfururyl), ranitidine, \( N,N \)-dimethylbenzylamine (DMBzA) and \( N,N \)-dimethylisopropylamine (DMiPA), which had NDMA formation potential yields of 25%, 70%, 49.5%, 50.3-74.9%, 62.9-94.2%, 83.8% and 83.9% mol/mol, respectively (Table 1).

For four of these seven precursors – DMPMA, DMthiophene, DMfururyl and ranitidine - the R group is an aromatic, heterocyclic, five-membered ring - a furan, thiophene or pyrrole - bonded in the \( \beta \)-position to the DMA nitrogen. Sacher and co-workers\(^10\) noted that reactive precursors which were identified had this functionality. However, it is not simply the case that such compounds will always act as potent NDMA precursors, as nizatidine and gramine share these features and have NDMA yields under 5% (Table S1). Moreover, the two most reactive precursors - DMBzA and DMiPA - have, respectively, a benzene group and two methyl groups \( \beta \) to the DMA nitrogen, rather than a five-membered heterocyclic ring. These two amines are interesting because they are structurally and functionally very similar to two precursors which are far less reactive, namely: DMBzA to \( N,N \)-dimethylaniline (DMAN, NDMA yield = 0.2%) and DMiPA to \( N,N \)-dimethyltertbutylamine (DMtBA, NDMA yield = 6.2%) (Table 1). Finally, DMA stands alone, not only because it is the only secondary amine studied, but because it has distinct molecular properties, with the lowest negative charge of all precursors at -0.802 (Figure 1).

### 3.2. Partial charge of DMA nitrogen: individual precursors

It can be seen that the more reactive NDMA precursors all have relatively negative partial charge when compared with the complete dataset (Figure 1). The seven most potent precursors mentioned above had partial charges between -0.252 and -0.522, respectively (Table 1), whereas the mean partial charge across all precursors was -0.300. However, the mean partial charge for the Category 1 amines was -0.412 and the reactive precursors have
intermediate negative partial charge amongst this category (Figure 1). The most striking example of a compound with low negative partial charge and low NDMA yield is DMA, with its respective values being -0.802 and 0.8% mol/mol (Table 2). The low partial charge of DMA is explained by the greater electronegativity of carbon (Pauling electronegativity = 2.55) relative to hydrogen (Pauling electronegativity = 2.20). Comparing DMBzA with DMAN and DMiPA with DMtBA reinforces this message, as their respective M062X partial charges were -0.252, -0.462, -0.522 and -0.543. Hence, for each of these two pairs, the precursor with the lowest partial charge had the lower NDMA formation. Overall this indicates that, while the most reactive NDMA precursors belong to the category with the lowest partial charge, high NDMA formation is not simply a function of low partial charge. In other words the presence of electron-donating groups, rather than electron-withdrawing groups, is a prerequisite for a reactive precursor, but other factors also play a role.

3.3. Role of \( pK_a \) in NDMA formation

For comparison with the theoretical values presented here, literature \( pK_a \) values for DMA, amitriptyline, chlorphenamine, ranitidine, TMA and DMform were reported as 10.73 (10.7), 9.4 (9.0), 9.13 (8.9), 8.2 (8.4), 9.8 (10.3) and -0.3 (-1.1) respectively\(^{14,39}\). The numbers shown in brackets are theoretical \( pK_a \) values from Jaguar 8.2 and show good agreement with literature values.

By plotting calculated \( pK_a \) values against NDMA formation another interesting pattern is revealed (Figure 1). The seven most reactive NDMA precursors were among the compounds with the highest \( pK_a \) on the DMA nitrogen. Respectively DMthiophene, DMfururyl, ranitidine, DMBzA and DMiPA had \( pK_a \) values clustered close together between 8.3 and 10.1, whereas the range among the whole list of investigated precursors was from -6.1 to 12.7 (Figure 1). There were only five amines with \( pK_a \) values above 10.1, and amongst them was...
DMA (10.7). Tetracycline had the highest $pK_a$ of 12.7. DMAN had a $pK_a$ of 4.9, which was far lower than that for DMBzA (9.3). Category 3 and 4 precursors had $pK_a$ values from -2.1 to 6.5 and from -6.1 to 7.3 respectively. The equivalent range for Category 1 precursors was 4.3 to 12.7 and, excepting DMA, the precursors with the highest $pK_a$ values belong to this category (Figure 1). High $pK_a$ values are a manifestation of having electron-donating substituents around the amine nitrogen, which stabilise the quaternary form. Conversely, electron-withdrawing substituents, such as present in Category 3 and 4 amines, will increase acidity and produce lower $pK_a$ values.

If NDMA formation depended on nucleophilic attack of an organic amine by chloramine, then, other factors being equal, it would be expected that compounds with the lowest $pK_a$ values would exhibit the highest conversion to NDMA, as these will have the highest proportion of unprotonated amine at circumneutral pH. However, this is the reverse of the observed pattern. Similarly, it would be expected that NDMA yields would increase with pH, being highest under alkaline conditions. In fact, it has been demonstrated for a range of precursors, namely DMA$^{40}$, ranitidine$^{14,41}$, amitriptyline$^{14}$, mifepristone$^{14}$ and sumatriptan$^{41}$, that NDMA formation peaks around pH 7-8. This indicates that the rate determining step/s in NDMA formation is/are either acid-catalysed or most efficient around neutral pH. The described pattern is consistent with the involvement of dichloramine in NDMA formation, since disproportionation of monochloramine to dichloramine is acid-catalysed$^{42}$.

3.4. Planarity and bond length of NDMA precursors

The seven most potent NDMA precursors had planarity ($\theta$) values from 122.1 – 127.3°, which places them amongst the amines with lowest planarity (Figure 2), as the mean $\theta$ value for all studied precursors was 137.8°. Unsymmetrical dimethylhydrazine (UDMH) and DMA had the two lowest $\theta$ values, of 115 and 119.4°, while the three quaternary Category 5
amines, which have four substituents around the amine nitrogen, also have low θ values, from
120° - 123.2° (Figure 2); i.e. they have a geometry approaching tetrahedral (bond angles
109.5° in a perfect tetrahedron). For the seven most reactive NDMA precursors, the reason
they possess low θ values is different: because they favour pyramidal sp³ hybridization on the
amino nitrogen and have a nitrogen lone pair which is more accessible for reaction with
chloramines and other aqueous reagents. It can be seen from Figures 1 and 2 that precursors
with low planarity also have R-groups with electron-donating properties.

DMtBA had a higher planarity than DMiPA, with the respective values being 127.8 and
123.0°, while DMAN had a higher planarity than DMBzA, and the respective values being
149.3 and 127.3°. Thus, for both of these pairs, the amine with higher NDMA formation had
lower planarity (Table 1). This is the converse of the situation with partial charge, indicating
planarity is more closely related to NDMA formation.

The N-R bond length in DMA was 1.012 Å, distinctly shorter than the other precursors
(Figure 2). Bond lengths of the precursors segregate into clear groups based on the categories
defined in Table 1, with the exception of Category 4. All of the Category 2 and 3 precursors
have lower bond lengths than all but one of the Category 1 precursors (oxytetracycline). In
turn Category 1 precursors have lower bond lengths than the Category 5 precursors (Figure
2). The bond lengths of the Category 4 precursors were more variable, consistent with the
differing identity of the heteroatom in these compounds. The four amines with the highest
bond lengths were cyazofamid, N,N-dimethyl-N-(4-methylphenyl)-sulfamide (AR-
DM sulfamide), N,N-Dimethylsulfamide (DMS) and thiotixene, with values ranging from
1.587 to 1.652 Å. These compounds are the only four considered which contain an electron-
withdrawing sulfonyl functionality bonded to the DMA nitrogen. This illustrates that the
described pattern in Figure 2 can be rationalised by noting that the presence of electron-
withdrawing groups will reduce bond length, whereas electron-donating groups increase bond
length.

4. Properties of reactive NDMA precursors

From consideration of the complete dataset, there were no strong correlations between any of
the individual molecular properties and NDMA formation, as indicated by calculation of
Spearman rank correlation coefficients ($r_s$) which were between -0.228 and 0.305 (Table 2).
The property with the lowest negative correlation coefficient was partial charge, while $pK_a$
had the highest positive correlation coefficient (Table 2). As was alluded to previously, the
explanation for these weak correlations is that reactive NDMA precursors are characterised
by their low partial charge and high $pK_a$ values.

In terms of relationships between the molecular properties, there were stronger negative
correlation coefficients between $pK_a$ and partial charge, -0.700, and also between $pK_a$ and
planarity, -0.743 (Table 2). These correlations illustrate the covariance between $pK_a$, partial
charge and planarity: amines with a high $pK_a$ value tend to also have a more negative partial
charge and a low planarity on the DMA nitrogen.

One explanation for the absence of any strong correlations involving NDMA formation is
that reactions between organic amines and chloramines can proceed via multiple mechanisms
and/or rate-determining-steps. Thus, molecular properties may not have a uniform influence
on NDMA formation. However, when correlations were calculated for just the seven most
reactive precursors, a coefficient of 0.847 between planarity and NDMA yield is evident
(Table 2), which disappears as more precursors are added to the analysis. This indicates that
the accessibility of the DMA lone pair, as quantified by planarity, is more important than the
presence of electron-donating groups, as quantified by partial charge, in promoting NDMA
formation amongst the reactive subset of precursors.
The properties of DMBzA versus DMAN, and of DMiPA versus DMtBA, have been compared in the proceeding sections. Each pair has similar molecular descriptors, but dissimilar NDMA formation. There is a noteworthy functional difference between these two pairs not well captured by the molecular descriptors: both DMBzA and DMiPA (the reactive precursors) have an acidic hydrogen in the α-position to the DMA nitrogen. In contrast DMAN and DMtBA (the less reactive precursors) do not have this structure (Table 1). N,N-dimethylphenethylamine (DMPhA) also makes a pertinent comparison with DMBzA and DMAN, due to their functional similarities. Once again, the similar chemical functionality of these three amines translates into similar molecular descriptors, except that DMPhA has lower planarity (Table 1). However, DMPhA differs in that it has a low NDMA yield, of 0.4% mol/mol, and a benzene ring in the γ-position to the DMA nitrogen (Table 1). A likely explanation for its very low conversion into NDMA is that the benzene ring is too distant from the α-hydrogen to effectively influence the rate-determining steps in NDMA formation.

5. Searching for currently unknown precursors

The preceding sections define the properties of reactive NDMA precursors using both quantitative computational descriptors and qualitative functional information. Both were combined when searching for additional compounds with the requisite characteristics to generate NDMA upon chloramination, i.e. currently unknown precursors. To achieve this, the Merck index\textsuperscript{43, 44}, which contains over 10,000 monographs about compounds of particular significance with respect to research, commerce and environmental impact, was explored. Use of the search term “dimethylamino” returned 233 potential precursors, once compounds whose NDMA formation potential have already been quantified experimentally were excluded (Table S4). Of this number, 60% are therapeutics, 13% veterinary therapeutics, 10% natural products, and 6-8% belonging to each of five other categories: antibiotics, controlled
substances, biological stains or dyes, chemical reagents or pesticides. Since these selected chemicals can belong to more than one of these categories, the totals appear to exceed 100%.

These numbers are consistent with the established association between the presence of wastewater effluent reuse in drinking water sources and NDMA formation. Nonetheless, they also highlight other potential sources of NDMA precursors: agricultural run-off, industrial wastewater and biological compounds. NDMA formation from a number of veterinary antibiotics has only recently been reported\textsuperscript{25} and is emphasised by these numbers. In order to select a smaller shortlist of compounds which are potentially reactive NDMA precursors the following steps were used:

1. Chemicals which (i) belong to Category 1 (\((\text{CH}_3)_2\text{N-Alkyl-R})\), (ii) have a hydrogen alpha to the DMA group, and (iii) an electron-donating substituent/s beta to the DMA group (defined as two alkyl groups, 1+ aromatic ring or 1+ heterocyclic ring) were selected. Obsolete chemicals were also excluded at this point. This left 68 compounds with attributes enabling them to be reactive NDMA precursors, based on this purely qualitative assessment (Table S5). Of this number, 62%, 20%, 18% and 7% are therapeutics, veterinary therapeutics, natural products and controlled substances, respectively (Table S5).

2. Representatives of classes containing multiple amines of similar functionality in Table S5 were selected. These chemical classes are the macrolides, tetracyclines, phenothiozine derivatives, steroid alkaloids and opioids/opiates, which are represented by 20, 12, 5, 3 and 4 compounds in Table S5, respectively. All of the tetracycline derivatives were excluded at this stage, since there were already three representatives in the analysed dataset, namely chlortetracycline, oxytetracycline and tetracycline, all with relatively low NDMA yields of 0.9, 1.4 and 1.2 %, respectively (Table S2). Of the four opioids/opiates, dimepheptanol, levomethadyl acetate and methadyl acetate are structurally similar to methadone and thus likely also to be reactive NDMA precursors. However, they were excluded as the aim was to
elucidate currently unknown types of NDMA structures, while the fourth opioid/opiate, tilidine, was retained as it has a different chemical functionality. In addition, four macrolides, two phenothiazine derivatives and two steroid alkaloids were taken forward into the list of 25 selected potential NDMA precursors (Table S6).

3. Molecular descriptors for the 25 compounds selected in Step 2 were calculated, using the same methods as for experimentally-tested precursors (Table S6).

4. Select compounds which have molecular descriptors within a range of values defined by the seven most reactive NDMA precursors in Table 1, i.e. DMfurfuryl, DMBzA, DMiPA, DMPMA, DMthiophene, ranitidine and methadone. The properties of these seven amines are listed in Table S7, with the 14 potential precursors whose molecular descriptors fall within the limits defined by the known precursors shown in bold in Table S6.

5. The NDMA formation potential of a selection of the 14 compounds left by following Steps 1-4 was tested experimentally. Only five of these compounds were found to be commercially-available and their experimental NDMA yields are shown in Table 3 and discussed below.

6. Previously unknown NDMA precursors

The five previously untested precursors have experimental NDMA yields ranging from 2.0±0.0 to 83.3±0.5% mol/mol (Table 3). The most potent of these is rivastigmine, a therapeutic drug used for the treatment of dementia. This amine has an aromatic functionality beta to a dimethylamine group, as do DMBzA, DMfurfuryl, DMPMA, ranitidine and DMthiophene, which are five of the seven most reactive precursors which had previously been tested (Table 1). Rivastigmine also has a methyl alpha to the dimethylamine group, a feature shared with methadone and DMiPA, which are also amongst the seven most reactive precursors tested prior to this study. Other amines in Table S5 with an aromatic ring beta to a
dimethylamine group, and an alpha proton, are adinazolam, ciramadol, dimefline, taxine and vetrabutine. All of these except ciramadol were selected in the list shown in Table S6 as being potentially-reactive NDMA precursors. Thus, an aromatic ring beta to a dimethylamine group, combined with an alpha proton, is likely to manifest itself in the molecular descriptors associated with high NDMA yields.

The second most potent previously unknown precursor is conessine, which forms 42.3±1.8% mol/mol of NDMA upon chloramination (Table 3). This is of particular interest as it is a steroid alkaloid known from several plant species and which highlights that NDMA precursors may occur amongst components of aquatic natural organic matter and not just in wastewater-impacted drinking water sources. To the best of our knowledge, this is the first time such a high NDMA yield has been reported from a naturally-occurring amine.

The three other previously unknown precursors all formed ≤4.7% mol/mol NDMA (Table 3). Promethazine is one of five phenothiozine derivatives in Table S5, all of which are therapeutics, while three are also veterinary therapeutics. As four of the five amines in Table 4, including promethazine, are used as therapeutics, these are potentially amongst the pool of NDMA precursors found in wastewater-impacted drinking water sources.

Finally, this study illustrates how computational chemistry can complement traditional experimental research in the area of water quality engineering by providing insight into the reactivity and fate of specific aquatic contaminants. With respect to NDMA formation, approaches involving computational chemistry could, for example, be applied to further elucidate the pathways leading to NDMA formation from different precursors under water treatment conditions.

6. Conclusions
This study is the first to use computational methods to define the molecular properties of amines which form NDMA during chloramination. Its key findings are as follows:

- Precursors classified on the basis of chemical functionality showed distinct relationships between partial charge and NDMA formation. Quaternary amines and tertiary amines with the DMA bonded to –COR and –CSR groups had a combination of low NDMA formation and high partial charge. Tertiary amines with the DMA bonded to an alkyl group, which includes the most potent NDMA precursors, had relatively negative partial charges.

- The most potent NDMA precursors also possessed relatively high pK\textsubscript{a} values and low planarity values. This demonstrates that electron-donating substituents and accessibility of the DMA lone pair are important in promoting NDMA formation, whereas a high partial charge on the DMA nitrogen suppresses NDMA formation.

- A literature search identified 233 compounds which are potential NDMA precursors, but have never been tested experimentally. Of these chemicals 60% are therapeutics, 13% veterinary therapeutics and 10% natural products. This highlights the origin and identity of those amines which contribute to NDMA formation in drinking water.

- Analysis combining qualitative assessment of chemical functionality and computational calculation of molecular descriptors successfully identified rivastigmine, a therapeutic compound, and conessine, found in several plant species, as potent NDMA precursors with experimentally determined yields of 83.3±0.5% and 42.3±1.8% mol/mol, respectively. The latter is the first time such high NDMA formation has been reported from a naturally-occurring precursor.

6. Acknowledgements
The role of the EPSRC UK National Service for Computational Chemistry Software (NSCCS) in supplying resources for this project is gratefully acknowledged. The authors are also grateful to the Imperial College High Performance Computing Service for supplying resources:

http://www.imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing

The first author also acknowledges the support of the Imperial College London Junior Research Fellowship Scheme.

7. References


44. Merck, The Merck Index online, [https://www.rsc.org/merck-index](https://www.rsc.org/merck-index), (accessed 11 November 2016, 2016).
<table>
<thead>
<tr>
<th>Precursor</th>
<th>Abbreviation</th>
<th>NDMA yield (% mol/mol)</th>
<th>Reference/s</th>
<th>Structure</th>
<th>Partial charge</th>
<th>Planarity (°)</th>
<th>Bond length (Å)</th>
<th>pK_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimethylamine</td>
<td>DMA</td>
<td>0.5-2.7 (0.8)</td>
<td>8-12</td>
<td><img src="image" alt="Structure" /></td>
<td>-0.802</td>
<td>119.8</td>
<td>1.012</td>
<td>10.7</td>
</tr>
</tbody>
</table>

**Category 1: (CH₃)₂N-Alkyl-R**

<p>| 5-(dimethylaminomethyl) furfuryl alcohol | DMfurfuryl | 50.3-84.6 (50.3) | 10, 12, 21 | <img src="image" alt="Structure" /> | -0.384 | 122.6 | 1.454 | 8.7 |
| N,N-dimethylbenzylamine | DMBzA | 82.5-83.8 (83.5) | 12, 13, 16 | <img src="image" alt="Structure" /> | -0.252 | 127.3 | 1.467 | 9.3 |
| N,N-Dimethylisopropylamine | DMiPA | 83.9 | 13 | <img src="image" alt="Structure" /> | -0.522 | 123.0 | 1.471 | 10.1 |
| N,N-dimethylphenethylamine | DMPhA | 0.4 | 13 | <img src="image" alt="Structure" /> | -0.469 | 121.6 | 1.460 | 9.4 |
| N,N-dimethyl-1-(1H-pyrrol-2-yl)methanamine | DMPMA | 25 | 13 | <img src="image" alt="Structure" /> | -0.424 | 121.3 | 1.465 | 9.6 |
| N,N-dimethyl-1-(1H-pyrrol-2-yl)methanamine | DMtBA | 6.2 | 13 | <img src="image" alt="Structure" /> | -0.543 | 127.8 | 1.489 | 10.4 |
| Methadone | 70 | 24 | <img src="image" alt="Structure" /> | -0.385 | 127.7 | 1.473 | 8.3 |
| Ranitidine |  | 62.9-94.2 (62.9) | 10, 12-16 | <img src="image" alt="Structure" /> | -0.439 | 122.6 | 1.455 | 8.4 |
| N,N-dimethylthiophene-2-methylamine | DMthiophene | 49.5-77.6 (49.5) | 10, 13 | <img src="image" alt="Structure" /> | -0.379 | 122.1 | 1.461 | 8.9 |</p>
<table>
<thead>
<tr>
<th>Category 2: (CH₃)₂N-Aryl</th>
<th>![Chemical Structure]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N,N-dimethylaniline</td>
<td>DMAN</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>-0.462</td>
</tr>
<tr>
<td></td>
<td>149.3</td>
</tr>
<tr>
<td></td>
<td>1.386</td>
</tr>
<tr>
<td></td>
<td>4.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 3: (CH₃)₂N-C=R₁(R₂)</th>
<th>![Chemical Structure]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-dimethylaminoantipyrine</td>
<td>4-DMAantipyrine</td>
</tr>
<tr>
<td>12.0</td>
<td>16</td>
</tr>
<tr>
<td>-0.286</td>
<td>126.9</td>
</tr>
<tr>
<td>1.414</td>
<td>6.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 4: (CH₃)₂N-heteroatom</th>
<th>![Chemical Structure]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsymmetrical dimethylhydrazine</td>
<td>UDMH</td>
</tr>
<tr>
<td>0.5</td>
<td>16</td>
</tr>
<tr>
<td>-0.322</td>
<td>115.1</td>
</tr>
<tr>
<td>1.458</td>
<td>7.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 5: (CH₃)₂N⁺-R₂</th>
<th>![Chemical Structure]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetramethylamine</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>23</td>
</tr>
<tr>
<td>0.329</td>
<td>120.0</td>
</tr>
<tr>
<td>1.497</td>
<td><strong>cation</strong></td>
</tr>
</tbody>
</table>

*na = not available. The full dataset is shown in the supporting information.*
Table 2: Spearman rank correlation coefficients ($r_s$) between NDMA formation and molecular properties

<table>
<thead>
<tr>
<th></th>
<th>All precursors (n = 64)</th>
<th>Most potent precursors (n = 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDMA yield</td>
<td>Partial charge</td>
</tr>
<tr>
<td>Partial charge</td>
<td>-0.228</td>
<td></td>
</tr>
<tr>
<td>Planarity</td>
<td>-0.157</td>
<td>0.448</td>
</tr>
<tr>
<td>Bond length</td>
<td>0.026</td>
<td>-0.198</td>
</tr>
<tr>
<td>$pK_a$</td>
<td>0.305</td>
<td>-0.700</td>
</tr>
</tbody>
</table>
Table 3: Experimental NDMA yields for previously untested precursors. Precursor and monochloramine concentrations: 0.1 mM and 2 mM, respectively. Contact time: 5 days at pH 8.0.

<table>
<thead>
<tr>
<th>Precursor</th>
<th>Comments</th>
<th>Structure</th>
<th>NDMA yield (% mol/mol)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almotriptan</td>
<td>Drug prescribed for acute headaches. Brand names include Axert, Almogran, Almotrex and Amignul.</td>
<td></td>
<td>2.0±0.0</td>
</tr>
<tr>
<td>Conessine</td>
<td>Steroid alkaloid found in several plant species.</td>
<td></td>
<td>42.3±1.8</td>
</tr>
<tr>
<td>Inosine</td>
<td>Antiviral drug. A combination of three different molecules. The dimethyl-containing one is dimethylaminoisopropanol.</td>
<td></td>
<td>4.7±0.0*</td>
</tr>
<tr>
<td>Pranobex</td>
<td>Used for treatment of dementia. Brand name Exelon.</td>
<td></td>
<td>83.3±0.5</td>
</tr>
<tr>
<td>Rivastigmine</td>
<td>Neuroleptic drug prescribed for various medical conditions.</td>
<td></td>
<td>2.8±0.0</td>
</tr>
<tr>
<td>Promethazine</td>
<td>Calculated based on molecular weight of dimethylamine-containing molecule only.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Relationships between partial charge and NDMA formation (above) and between $pK_a$ and NDMA formation (below). Each data point represents the descriptor and NDMA yield for a single precursor. Precursor categories are shown in Table 1.
Figure 2: Relationships between planarity and NDMA formation (above) and between bond length and NDMA formation (below). Each data point represents the descriptor and NDMA yield for a single precursor. Precursor categories are shown in Table 1.