1   **Validation of the identification reliability of known and assumed UDMH transformation products**
2   **using gas chromatographic retention indices and machine learning**

3   **Anastasia E. Karnaeva[1*], Anastasia Y. Sholokhova[1]**

4   [1] Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Moscow,

5   Russia

6   *Corresponding author

7   E-mail addresses: karnaevajun@gmail.com (A. Karnaeva), shonastya@yandex.ru (A. Sholokhova)

8   **Abstract**

9   Thirty two commercially available standards were used to determine chromatographic retention indices

10   (RIs) for three different stationary phases. The selected compounds were nitrogen-containing heterocycles

11   and amides, which are unsymmetrical dimethylhydrazine (UDMH) transformation products or its assumed

12   transformation products. UDMH is a highly toxic compound widely used in the space industry, that forms

13   a number of different compounds in the environment. Well-known transformation products may exceed

14   UDMH itself in their toxicity, but most of the products are poorly investigated, while posing a huge

15   environmental threat. Experimental RIs for the stationary phases, RIs from the NIST database, and

16   predicted RIs are presented in this paper. It is shown that there are virtually no RIs for UDMH

17   transformation products in the NIST database. In addition, even among those compounds for which RIs

18   were known, inconsistencies were identified. Adding RIs to the database and eliminating erroneous data

19   would allow for more reliable identification when standards are not available. The discrepancies identified

20   between experimental RI values and predicted values will allow for adjustments to the machine learning

21   models that are used for prediction. Previously proposed compounds as possible transformation products

22   without the use of standards and NMR method were confirmed.

23

24   **Key words**

25   Retention index; unsymmetrical dimethylhydrazine; identification; retention index prediction

26   **1. Introduction**

27   Due to the rising amount of new chemicals released to the environment non-targeted analysis using

28   mass spectrometry (MS) coupled to liquid/gas chromatography (LC/GC) have risen as a new approach in

29   analytical chemistry to detect, identify and eventually quantify chemicals in environmental samples [1,2].

30   Especially non-targeted analysis is prevalent when working with unsymmetrical dimethylhydrazine

31   (UDMH), which is a toxic component of rocket fuel. The particular complexity of UDMH as a research

32   object lies in its high reactivity, which leads to the formation of a large number of transformation products

33   [3,4]. Establishing the structure of new UDMH transformation products is a challenging task as standard

34   samples are in most cases unavailable. Also, various environmental factors, such as humidity, pH, matrix,

1

etc., influence the UDMH transformation, which leads to the fact that at the moment there is no precise understanding of what substances will be formed under certain conditions [5,6]. The expansive distribution of false transformation products should be mentioned, as often the identification of new compounds belonging to transformation products is superficially performed only by the mass spectra obtained without the use of additional methods [5,7]. In our work, we also observed that some compounds might not be available in databases and identification was done indirectly based on available published data [8].

Majority of studies identify UDMH transformation products by a library search of the MS-spectra database. At the same time, a database search in around 20% of times shows an incorrect result even if the mass spectrum is available in the database. Nevertheless, the structure elucidation of isomeric compounds based only on their fragmentation pattern, may sometimes not be feasible, since they produce common fragments and the reference standards are not always available [9]. At the moment many substances known as possible UDMH transformation products are not available in any databases [7,10]. Only several works use standards [11] and nuclear magnetic resonance (NMR) [10,12]. At the same time, even the use of high-resolution MS [3,13] does not allow to assume a possible structure, since tens of thousands of possible structures with different properties can correspond to the brutto-formula. GC retention index (RI) can be used as an additional factor to confirm or reject candidates during GC-MS identification [14]. However, reference RIs are available for less than half of the compounds in the NIST database, and reference RIs are not available for most UDMH transformation products [15]. This problem can be partially solved by predicting RI using machine learning [16]. During the last decade prediction of RIs with machine learning has been applied more and more widely to the problems of non-targeted analysis in environmental chemistry [17–22]

The work of our team demonstrated the successful use of a combination of high-performance liquid chromatography-high resolution mass spectrometry (HPLC-HRMS) and GC-MS with artificial intelligence techniques to cross-check the proposed structures of new UDMH transformation products [7,15]. However, here we would like to highlight a very important nuance that machine learning is conducted on known databases, including NIST, and the accuracy of prediction directly depends on the correctness of the available values. Khrisanfov et al. developed a novel approach to find potentially erroneous entries in a large-scale database with mostly unique entries [23]. The authors [14,24] also raise concerns that some RI values in the database may be erroneous. RI filtering in the NIST database is an important task because about 80% of the data have only one RI value per stationary phase (SP) [23]. Identifying incorrect entries will improve the prediction accuracy of machine learning models.

Therefore, the aim of this work was to compare experimental, reference (from the NIST database) and predicted RIs obtained for different SP for possible UDMH transformation products. The experimental RI and predicted RI were compared with those available in the literature. This approach is presented for the first time on the example of UDMH transformation identification. For this purpose, standards of substances were selected that are or are suspected to be UDMH transformation products or may be structurally similar to its transformation products. Standard SP for the analysis of UDMH transformation products were chosen

2

72 for the experiment. Different types of SP ranging from non-polar (RTx-5MS) [25,26], mid-polar (DB-1701)

73 [27,28] to polar (HP-INNOWax) [29,30] have been used for the analysis earlier. It is required to obtain

74 experimental GC RI for these compounds at different SP and compare the obtained values with those

75 available in the literature and predicted by machine learning methods. The obtained data will allow to enrich

76 the databases, which will subsequently improve the accuracy of GC RI prediction. Also, the data obtained

77 by us can be used to adjust the identification methods of unknown UDMH transformation products, which

78 will allow to make a more precise assessment of UDMH and its transformation product impact on the

79 environment.

80

81 **2. Experimental**

82 **2.1 Analytes, reagents and materials**

83 32 commercially available nitrogen-containing compounds, representing proved and assumed

84 UDMH transformation products were chosen as target analytes (Table 1). 1,3-dimethyl-1H-pyrazole-4-

85 carbonitrile, 1H-1,2,4-triazole, 1-methyl-1H-pyrazol-3-amine, 1H-1,2,4-triazol-3-amine, 3-methyl-1H-

86 1,2,4-triazole, 4H-1,2,4-triazol-4-amine, 5-ethyl-1H-pyrazole-3-amine, 1H-1,2,3-triazole, imidazole, 2-

87 methylimidazole, 1,2-dimethylimidazole, 2-ethyl-4-methyl-1H-imidazole, N,N-dimethylformamide,

88 pyrazole, pyridazine, pyrimidine, pyrazine were purchased from Sigma-Aldrich (St. Louis, MO, USA). 1-

89 methyl-1H-1,2,4-triazole, 4-hydroxypyridine (4-pyridinol), N,N-dimethylurea, 3-methylpyrazole, 1-

90 methylpyrazole, 4-hydroxybutyric acid hydrazide, acetamide, 3,5-dimethyl-1H-pyrazole, 1-

91 methylimidazole were purchased from Angene International Limited (Nanjing, China) and 2-methyl-1H-

92 imidazole-1-propanamine, 3,4-dimethyl-1H-pyrazole, 3-ethyl-1H-1,2,4-triazol-5-amine, 6-methyl-1,3,5-

93 triazine-2,4-diamine, 1-ethylimidazole were purchased from Makclin Biochemical Co (Shanghai, China).

94 2-formyl-1,1-dimethylhydrazine (N-(dimethylamino)formamide) was purchased from Northern (Arctic)

95 Federal University, Arkhangelsk [31].

96 HPLC grade methanol (EVA Science, St. Petersburg, Russia) and HPLC grade isopropanol

97 (PanReac, Barcelona, Spain) were used for preparing analyte solutions. Aqueous solutions of analytes were

98 prepared using ultrapure water (Milli-Q). Alkane standard solution C_8–C_{40} (0.5 mg/mL of each component

99 in dichloromethane, Sigma-Aldrich, St. Louis, MO, USA), alkane standard solution C_8–C_{20} (40 mg/L of

100 each component in hexane, Sigma-Aldrich, St. Louis, MO, USA), and alkane standard solution C_5–C_{24} in

101 iso-octane (ChromLab, Russia) were used for the determination of RI. Different types of columns ranging

102 from non-polar (RTx-5MS 5% Phenyl 95% dimethylpolysiloxane 30m×0.25mm×0.25μm, Restek,

103 Bellefonte, PA), mid-polar (DB-1701, 14% cyanopropylphenyl-86% dimethyl polysiloxane,

104 30m×0.32mm×0.25μm Agilent J&W Scientific) to polar column (HP-INNOWax, 30m×0.25mm×0.25μm

105 Polyethylene glycol Agilent J&W Scientific) were used for the analysis.

106 **2.2 Preparation of analyte solutions**

107   Stock solutions of individual analytes with the concentrations of 5-7 mg/mL were prepared from
108   accurately weighed portions in methanol, isopropanol and water (Milli-Q) and stored at –20 °C in a
109   laboratory freezer. The standard solutions of analyte mixtures used for the determination of RI were
110   prepared by mixing and then diluting stock solutions with methanol or isopropanol.

4

111   Table 1 Target analytes

№	Compound	CAS	Structure	Mol. weight, Da	Smiles
1	Acetamide	60-35-5		59.07	CC(N)=O
2	1H-Imidazole	288-32-4		68.08	C1=CN=CN1
3	1H-Pyrazole	288-13-1		68.08	C1=CNN=C1
4	1H-1,2,4-Triazole	288-88-0		69.07	C1=NC=NN1
5	1H-1,2,3-Triazole	288-36-8		69.07	C1=NNN=C1
6	N,N-Dimethylformamide	68-12-2		73.09	CN(C)C=O
7	Pyrimidine	289-95-2		80.09	C1=CN=CN=C1

5

8	Pyridazine	289-80-5		80.09	C1=CC=NN=C1
9	Pyrazine	290-37-9		80.09	C1=CN=CC=N1
10	3-Methyl-1H-pyrazole	1453-58-3		82.10	CC1=CC=NN1
11	1-Methyl-1H-pyrazole	930-36-9		82.10	CN1C=CC=N1
12	1-Methyl-1H-imidazole	616-47-7		82.10	CN1C=CN=C1
13	2-Methyl-1H-imidazole	693-98-1		82.10	CC1=NC=CN1
14	3-Methyl-1H-1,2,4-triazole	7170-01-6		83.09	CC1=NC=NN1
15	1-Methyl-1H-1,2,4-triazole	6086-21-1		83.09	CN1C=NC=N1
16	1H-1,2,4-Triazol-3-amine	61-82-5		84.08	C1=NNC(=N1)N

6

17	4H-1,2,4-Triazol-4-amine	584-13-4		84.08	C1=NN=CN1N
18	N,N-dimethylurea	598-94-7		88.11	CN(C)C(=O)N
19	N-(dimethylamino)formamide (2-formyl-1,1-dimethylhydrazine)	3298-49-5		88.11	CN(C)NC=O
20	4-Hydroxypyridine	626-64-2		95.10	C1=CNC=CC1=O
21	3,4-Dimethyl-1H-pyrazole	2820-37-3		96.13	CC1=C(NN=C1)C
22	1,2-Dimethyl-1H-imidazole	1739-84-0		96.13	CC1=NC=CN1C
23	1-Ethyl-1H-imidazole	7098-07-9		96.13	CCN1C=CN=C1
24	3,5-Dimethyl-1H-pyrazole	67-51-6		96.13	CC1=CC(=NN1)C

25	1-Methyl-1H-pyrazol-3-amine	1904-31-01		97.12	CN1C=CC(=N1)N
26	2-Ethyl-4-methyl-1H-imidazole	931-36-2		110.16	CCC1=NC=C(N1)C
27	5-Ethyl-1H-pyrazole-3-amine	1904-24-1		111.15	CCC1=CC(=NN1)N
28	3-Ethyl-1H-1,2,4-triazol-5-amine	22819-05-2		112.13	CCC1=NC(=NN1)N
29	4-Hydroxybutyric Acid Hydrazide	3879-08-1		118.13	C(CC(=O)NN)CO
30	1,3-Dimethyl-1H-pyrazole-4-carbonitrile	87412-96-2		121.14	CC1=NN(C=C1C#N)C
31	6-Methyl-1,3,5-triazine-2,4-diamine	542-02-9		125.13	CC1=NC(=NC(=N1)N)N
32	2-Methyl-1H-imidazole-1-propanamine	2258-21-1		139.20	CC1=NC=CN1CCCN

112

8

2.3 GC-MS

A Shimadzu GCMS-TQ8040 system (Shimadzu Corporation, Kyoto, Japan) was used to perform GC-MS analysis. GC-MS parameters for HP-INNOWax were used as follows: an injection volume of 0.7 μL, a sample injection temperature of 250 °C, a split injection mode with a ratio of 1:10, and a carrier gas flow rate of 0.9 mL/min. Temperature program: 3 min at 40 °C, ramping the temperature to 240 °C at a rate of 8 °C/min, then 51 min at 240 °C; ion source temperature 200 °C and scanning range 40–650 m/z.

The conditions used for RTx-5MS —flow rate: 1.0 mL/min; a split injection mode with a ratio of 1:5; program: 3 min at 50 °C, then ramping to 300 °C at a rate of 10 °C/min, then 3 min at 300 °C.

GC-MS parameters for DB-1701 were used as follows: an injection volume of 0.7 μL, a sample injection temperature of 250 °C, a split injection mode with a ratio of 1:10, and a carrier gas flow rate of 2.1 mL/min. Temperature program: 3 min at 40 °C, ramping the temperature to 260 °C at a rate of 8 °C/min, then 5 min at 260 °C; ion source temperature 200 °C and scanning range 40–600 m/z.

Each sample was injected two or three times. The NIST 17 (NIST, Gaithersburg, MD, USA) database was used for the library search. The injector needle was cleaned with solvent (isopropanol or hexane) five times before and five times after each injection.

2.4 Retention indices

Retention indices (RIs) for GC were first described by Kovats [32] for isothermal conditions and then modified by Van den Dool and Kratz [33] for the temperature-programmed GC. Linear RI and isothermal RI based on n-alkanes were used in this work. A linear RI was calculated based on the following equation [34]:

$$RI = 100 * \left(\frac{N + (t_x - t_N)}{t_{N+1} - t_N} \right),$$

where RI is the retention index, t_N, t_{N+1}– RT of two n-alkanes with n and n+1 carbon atoms respectively ($C_N H_{2N+2}$ and $C_{N+1} H_{2N+4}$), t_x – RT of the considered compound, $t_N < t_x < t_{N+1}$.

GC-MS RIs were predicted using a previously published multimodal machine learning method, more details can be found in previous works [14,15].

2.5 Analytical reproducibility

To control the obtained results of the experimental RIs calculation, the index calculations were performed 2 times with a 1-month interval for each SP. As a result, 4 to 6 values of experimental retention indices were obtained for each target analyte. The difference in measured RI values did not exceed ∼5-7 units in all cases. The uncertainties in the library RI were larger: the previously reported value of the RI median error in the library was ∼11 units, however, the library contained multiple incorrect values with errors greatly exceeding ∼11 [23,35].

3. Results and discussion

9

3.1 Non-polar stationary phase

The non-polar SP has the most data on RIs. Although the polar SP is more often used for nitrogen-containing UDMH transformation products [36], a number of studies used non-polar columns to analyze one of the most toxic UDMH transformation products, N-nitrosodimethylamine [26,37]. Furthermore, non-polar 5% phenyl polydimethylsiloxane columns were also used for UDMH transformation products before [38–40]. Therefore, obtaining experimental RIs for the non-polar SP is of interest. Table 2 summarizes the obtained experimental RIs available in the NIST database as well as the predicted RIs.

Table 2. RI for the non-polar stationary phase

№	Compound	RI$_{exp}$, mean value	Standard Deviation	RI predicted	Δ	NIST value
1	Pyrazine	738	1	730	8	718
2	Pyrimidine	756	0	785	-29	736
3	1-Methylpyrazole	764	1	749	15	743
4	Acetamide	777	2	736	41	741
5	N,N-Dimethylformamide	788	2	762	26	750
6	1H-1,2,3-Triazole	827	0	1041	-214	NA
7	Pyrazole	840	1	856	-16	864
8	1-Methyl-1H-1,2,4-triazole	855	1	823	32	NA
9	2-Formyl-1,1-dimethylhydrazine	878	3	897	-19	NA
10	Pyridazine	918	1	940	-21	915
11	3-Methylpyrazole	921	1	893	28	930
12	1-Methylimidazole	927	0	917	10	929
13	Imidazole	968	1	963	5	1069
14	1,2,4-Triazole	972	1	1057	-84	1116
15	1-Methyl-1H-pyrazol-3-amine	1002	1	1128	-125	NA
16	3,5-dimethylpyrazole	1004	1	1017	-13	1014
17	1-Ethylimidazole	1005	2	988	17	990
18	1,2-Dimethylimidazole	1014	0	990	24	1006
19	2-Methylimidazole	1016	2	981	35	1050
20	3-Methyl-1H-1,2,4-triazole	1025	1	1028	-2	NA
21	3,4-dimethyl-1H-pyrazole	1030	1	1072	-42	NA

10

22	N,N-dimethylurea	1037	2	943	93	NA
23	2-Ethyl-4-methylimidazole	1145	1	1050	95	NA
24	1,3-dimethyl-1H-Pyrazole-4-carbonitrile	1193	1	1138	55	1256
25	4-Hydroxypyridine	1195	4	1068	129	1154
26	1H-1,2,4-Triazol-3-amine	1225	3	1178	45	1312
27	5-Ethyl-1H-pyrazole-3-amine	1263	1	1263	0	NA
28	4-Hydroxybutyric Acid Hydrazide	1314	0	1337	-23	NA
29	6-Methyl-1,3,5-triazine-2,4-diamine	1337	1	1482	-145	NA
30	3-ethyl-1H-1,2,4-triazol-5-amine	1344	3	1296	48	NA
31	2-Methyl-1H-imidazole-1-propanamine	1419	1	1413	5	NA
32	4-Amino-4H-1,2,4-triazole	1441	2	1100	342	NA

NA – not available

Δ – RI_{exp}-RI_{pred}

Table 2 shows that, even for such a common stationary phase as non-polar, the RI values for UDMH transformation products in the NIST database are very limited. It is worth noting the discrepancies between RI_{exp} and NIST value for such substances as imidazole, 1,2,4-triazole, and 1H-1,2,4-triazol-3-amine. For these compounds there are single records of RI in the NIST database. It will be the object of a special research by the authors. It should be noted that for such compounds as 1H-1,2,4-triazol-3-amine and 4-amino-4H-1,2,4-triazole it was not possible to select the elution conditions (flow rate, concentration of analyte) so that the substances came out with a narrow chromatographic peak. The peaks corresponding to the retention of these components had a very blurred profile (Figure S1). In the study where the index for 1H-1,2,4-triazol-3-amine was calculated [41], there were no chromatograms to compare the obtained results.

It is interesting to compare the RI values obtained for the target analytes with previously suggested UDMH transformation products without using standards. Thus, the structures of possible UDMH transformation products were proposed earlier in studies using GC-MS, HPLC-HRMS and machine learning methods [7,15]. Among the proposed candidates as possible transformation products were the following: 1-methylpyrazole (766), 3-methylpyrazole (941), 1-methyl-1H-1,2,4-triazole (852), 3,5-dimethylpyrazole (1025), and 1,3-dimethyl-1H-pyrazole-4-carbonitrile (1187). Experimental retention indices for the proposed candidates from that paper are presented in parentheses [7]. Comparison with the values presented above (Table 2) showed the coincidence of the experimental RIs for the referenced standards and for the structures proposed by the authors. 1,3-Dimethyl-1H-pyrazole-4-carbonitrile was first mentioned in the work as a UDMH transformation product [15]. Authors using the presented algorithm rejected proposed candidates generated by the NIST database (more details in the work [15]) and, including the predicted RI, suggested 1,3-dimethyl-1H-pyrazole-4-carbonitrile as a possible candidate. One of candidates in this case was 2-amino-4-methyl-1H-pyrrole-3-carbonitrile,

11

183 which is an isomer of 1,3-dimethyl-1H-Pyrazole-4-carbonitrile and only the RI can distinguish them,
184 since the isomers have the same mass spectral characteristics but different chromatographic patterns. The
185 experimental RIs presented in Table 2 will be useful for further identification of the UDMH
186 transformation products as well as other nitrogen-containing mixtures, since RIs in the databases are
187 limited even for such simple molecules. Furthermore, the available indices in the database may be
188 erroneous and require additional verification, as noted above and earlier by other authors [23,35]. The
189 provided experimental RIs will not only identify new and/or confirm those proposed in the literature, but
190 can also be used to correct erroneous entries in the NIST database, thus correcting the retention index
191 prediction and improving the prediction accuracy.

192   In addition, there are 3 compounds from the listed substances with mass spectra not available in
193 the database. One of them, 2-formyl-1,1-dimethylhydrazine, which is a UDMH transformation product,
194 was described in several studies [8,12,42]. Substances such as 2-methyl-1H-imidazole-1-propanamine
195 and 5-ethyl-1H-pyrazole-3-amine may be transformation products, but because their mass spectra are not
196 available in Nist17, their identification may be extremely difficult. Their mass spectra are presented in
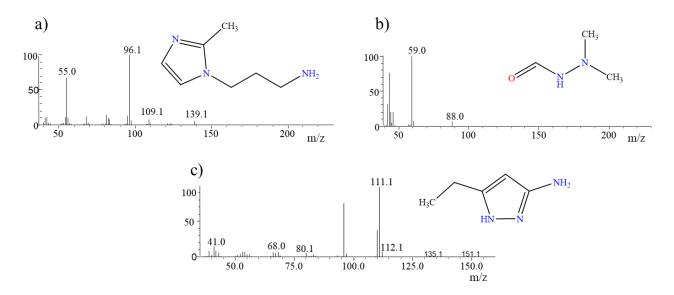197 Fig. 1.

198

199

200   Figure 1. Mass-spectrum of a) 2-methyl-1H-imidazole-1-propanamine; b) 2-formyl-1,1-
201   dimethylhydrazine; c) 5-ethyl-1H-pyrazole-3-amine

202   Adding mass spectra to the NIST database is also important for more accurate prediction of mass
203 spectra by machine learning methods. If there is no mass spectrum in the database, then predicting the
204 mass spectra of the candidate structure and comparing it with the experimental one can be very useful to
205 further identify and weed out the false candidate.

206 **3.2 Polar stationary phase**

12

One of the most common SPs for analyzing nitrogen-containing UDMH transformation products is the polar phase [8,29,30,39]. However, for polar SP, reference experimental RI is available for approximately 10,000 compounds, which is several times less than for non-polar SP [43]. This fact makes identification very difficult in the absence of standards, which makes RI prediction for polar SP an important task. Table 3 summarizes the experimentally obtained RI, predicted RI, and RI values published in NIST, when available, for the polar SP.

Table 3. RI for the polar stationary phase

№	Compound	RI exp, mean value	Standard Deviation	RI predicted	Δ	NIST value
1	Pyrazine	1218	0	1221	-4	1212
2	1-methylpyrazole	1257	2	1163	93	NA
3	Pyrimidine	1270	1	1292	-22	1267
4	N,N-Dimethylformamide	1337	1	1327	10	1326
5	1-Methyl-1H-1,2,4-triazole	1597	1	1354	243	NA
6	2-Formyl-1,1-dimethylhydrazine	1664	2	1804	-140	NA
7	1-methylimidazole	1695	1	1618	77	1681
8	1-Ethylimidazole	1733	1	1681	52	1714
9	Pyridazine	1736	2	1543	192	NA
10	1,2-Dimethylimidazole	1744	2	1565	178	1715
11	Acetamide	1776	0	1712	64	1764
12	Pyrazole	1822	2	1653	169	NA
13	3-methylpyrazole	1855	2	1695	160	1690
14	1-Methyl-1H-pyrazol-3-amine	1889	2	1922	-33	NA
15	3,5-dimethylpyrazole	1892	0	1678	214	1680
16	3,4-dimethyl-1H-pyrazole	1955	2	1795	159	NA
17	1H-1,2,3-Triazole	1975	1	2150	-175	NA

13

18	N,N-dimethylurea	2090	3	1952	136	NA
19	1H-Pyrazole-4-carbonitrile, 1,3-dimethyl-	2116	3	1885	230	NA
20	2-Methylimidazole	2165	1	1865	300	2146
21	2-Ethyl-4-methylimidazole	2181	1	1798	382	NA
22	Imidazole	2235	1	2149	86	2186
23	3-Methyl-1H-1,2,4-triazole	2311	3	1931	378	NA
24	1,2,4-Triazole	2374	4	2056	316	NA
25	2-Methyl-1H-imidazole-1-propanamine	2439	0	2265	174	NA
26	5-Ethyl-1H-pyrazole-3-amine	2601	0	2110	491	NA
27	6-Methyl-1,3,5-triazine-2,4-diamine	2862	3	2486	374	NA
28	4-Hydroxypyridine	2862	5	2203	657	NA
29	3-ethyl-1H-1,2,4-triazol-5-amine	3058	2	2189	869	NA
30	3-Amino-1,2,4-triazole	3121	1	2288	833	NA
31	4-Amino-4H-1,2,4-triazole	3528	2	2258	1270	NA
32	4-Hydroxybutyric Acid Hydrazide	-	-	-	-	NA

NA – not available

Δ – RI_{exp}-RI_{pred}

214

215   Table 3 shows that there are even fewer known retention indices on the polar phase for the UDMH
216 transformation products than for the nonpolar phase. We also found discrepancies in retention indices
217 with the available data for 3,5-dimethylpyrazole and 3-methylpyrazole. Zellner et al. [34] reported that
218 the reproducibility of retention indices between different laboratories was usually within 10 –50 index

14

219 units for polar phases. They stated that the large range of units for polar columns could be related to the

220 heterogeneity between columns manufactured by different companies, the misuse of the gas

221 chromatographic technique, a possible reactivity between the analyte and the stationary phase, and the

222 absence of an analytical reproducibility, as most publications do not report data on analysis performed in

223 duplicate or triplicate. The total absence of RI values for such classes of substances as triazoles, pyrazoles

224 and, to a lesser extent, imidazoles results in predicted RIs, even for simple molecules, being significantly

225 different from the experimental ones, although the patterns of their retention are preserved. This is

226 particularly evident for the substituted triazoles, 3-amino-1,2,4-triazole, and 4-amino-4H-1,2,4-triazole.

227 This result shows the necessity of obtaining correct RI data even for such simple structures, which will

228 significantly increase the reliability of retention index prediction.

229 **3.3 Mid-polar stationary phase**

230 Mid-polarity SP columns are rarely used to analyze UDMH transformation products, but mid-

231 polar columns (DB-1701) were used to analyze N-nitrosodimethylamine [27,28]. There are also very few

232 data on RIs for mid-polarity SP, which makes columns with this phase an interesting object for such

233 studies. Based on our results, we can conclude that the mid-polar column (DB-1701) has the best

234 separating ability in relation to a complex mixture of UDMH transformation products. Thus, substances

235 such as 3-amino-1,2,4-triazole and 4-amino-4H-1,2,4-triazole, which had blurred chromatographic

236 profiles on HP-INNOWax and RTx-5MS, came out with good chromatographic peaks on DB-1701

237 (Figure S2). Figure 2 shows the chromatogram of one of the selected mixtures of the target analytes with

238 a mixture of alkanes. In Table 2, the results of calculation and prediction of retention indices for the
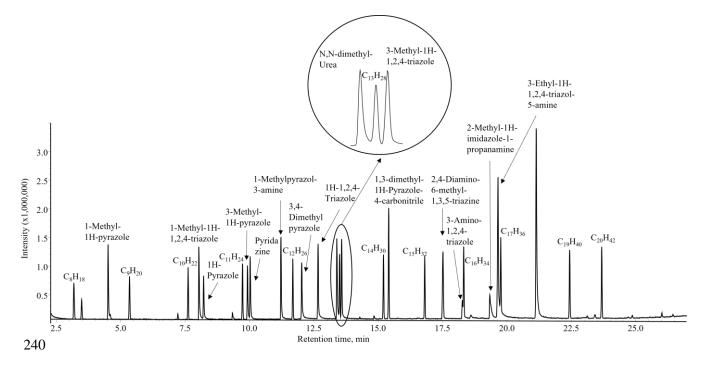
239 studied compounds are shown.

240

241 Figure 2. The GC-MS chromatogram of the mixture of target analytes with alkanes (DB-1701)

15

242   Table 4. RI for the mid-polar stationary phase

№	Compound	RI exp, mean value	Standard Deviation	RI predicted	Δ
1	Pyrazine	820	0	836	-16
2	Pyrimidine	848	0	852	-4
3	1-methylpyrazole	862	0	847	15
4	N,N-Dimethylformamide	944	0	865	79
5	Acetamide	1014	0	959	55
6	1-Methyl-1H-1,2,4-triazole	1019	0	917	102
7	Pyrazole	1028	0	982	46
8	1H-1,2,3-Triazole	1031	0	1220	-190
9	2-Formyl-1,1-dimethylhydrazine	1080	0	1136	-55
10	3-methylpyrazole	1110	0	1082	28
11	Pyridazine	1115	0	1026	88
12	1-methylimidazole	1118	0	1088	30
13	1-Methyl-1H-pyrazol-3-amine	1176	0	1287	-111
14	3,5-dimethylpyrazole	1193	0	1135	58
15	1-Ethylimidazole	1197	0	1172	25
16	1,2-Dimethylimidazole	1200	0	1103	97
17	3,4-dimethyl-1H-pyrazole	1219	0	1179	40
18	Imidazole	1252	0	1288	-37
19	1,2,4-Triazole	1253	0	1242	11
20	2-Methylimidazole	1292	0	1164	128
21	N,N-dimethylurea	1294	0	1186	108
22	3-Methyl-1H-1,2,4-triazole	1304	0	1166	138
23	2-Ethyl-4-methylimidazole	1406	0	1219	187
24	1,3-dimethyl-1H-Pyrazole-4-carbonitrile	1413	0	1386	26
25	5-Ethyl-1H-pyrazole-3-amine	1535	2	1422	114
26	6-Methyl-1,3,5-triazine-2,4-diamine	1546	1	1756	-210
27	4-Hydroxypyridine	1592	1	1351	241
28	3-Amino-1,2,4-triazole	1597	0	1454	143
29	4-Hydroxybutyric Acid Hydrazide	1652	0	1590	63

16

30	2-Methyl-1H-imidazole-1-propanamine	1670	1	1652	18
31	3-ethyl-1H-1,2,4-triazol-5-amine	1692	1	1480	212
32	4-Amino-4H-1,2,4-triazole	1997	0	1334	663

NA – not available

Δ – RI_{exp}-RI_{pred}

Table 4 shows that the predicted RIs for the selected analytes correlated well with the experimental data, which we did not predict. The better convergence for the RIs for the mid-polar SP compared to the polar stationary phase can probably be explained by the fact that for this SP there are fewer RIs values and therefore fewer erroneous values. Also, for 4-amino-4H-1,2,4-triazole, the largest discrepancy between the RI and the experimental one is observed for the nonpolar and polar phases as well.

4. Conclusions

For the first time, retention indices were calculated for some confirmed and assumed UDMH transformation products for 3 different stationary phases - polar (HP-INNOWax), non-polar (RTx-5MS) and medium polarity (DB-1701). In this work, for the first time, an approach was used where experimental retention indices and predicted retention indices were compared with those available in the literature. For a large number of confirmed UDMH transformation products there are no published values of retention indices even for the non-polar phase. However, it should be noted that in the course of this study some discrepancies were found with the records in the NIST database for the polar and nonpolar phases. Erroneous values greatly affect the accuracy of the prediction, especially when the database is small, as seen in the polar phase example. The predicted indices for the mid-polar stationary phase showed good convergence with the experimental data. It is also worth noting that this phase showed the best separation ability compared to the polar and non-polar phases, and these phases should be used to analyze UDMH transformation products. The obtained data will allow to enrich the databases, which will subsequently improve the accuracy of GC RI prediction. Our results can be used in non-target screening to adjust the identification of compounds of emerging concerns that are not available as reference standards. This approach would provide, by establishing the right structure, a more accurate assessment of the environmental impact of UDMH and its transformation products.

273

CRediT authorship contribution statement

A. Karnaeva: Methodology, Investigation, Validation, Conceptualization, Writing-Original draft preparation, **A. Sholokhova**: Investigation, Validation, Conceptualization, Writing-Original draft preparation

278

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

282

References

[1] B. González-Gaya, N. Lopez-Herguedas, D. Bilbao, L. Mijangos, A.M. Iker, N. Etxebarria, M. Irazola, A. Prieto, M. Olivares, O. Zuloaga, Suspect and non-target screening: the last frontier in environmental analysis, Analytical Methods 13 (2021) 1876–1904. https://doi.org/10.1039/D1AY00111F.

[2] B. Giannelli Moneta, S.E. Aita, E. Barbaro, A.L. Capriotti, A. Cerrato, A. Laganà, C.M. Montone, S. Piovesana, F. Scoto, C. Barbante, C. Cavaliere, Untargeted analysis of environmental contaminants in surface snow samples of Svalbard Islands by liquid chromatography-high resolution mass spectrometry, Science of The Total Environment 858 (2023) 159709. https://doi.org/10.1016/j.scitotenv.2022.159709.

[3] N. V. Ul'yanovskii, D.S. Kosyakov, M.S. Popov, I.S. Shavrina, A.D. Ivakhnov, B. Kenessov, A.T. Lebedev, Rapid quantification and screening of nitrogen-containing rocket fuel transformation products by vortex assisted liquid-liquid microextraction and gas chromatography – high-resolution Orbitrap mass spectrometry, Microchemical Journal 171 (2021) 106821. https://doi.org/10.1016/j.microc.2021.106821.

[4] S.D. Iartsev, I.S. Pytskii, A.E. Karnaeva, A.K. Buryak, Surface-assisted laser desorption/ionization mass spectrometry for the detection of low-molecular-weight and oligomeric products of 1,1-dimethylhydrazine transformation on the surfaces of construction materials, Russian Journal of Physical Chemistry B 11 (2017) 680–683. https://doi.org/10.1134/S199079311704025X.

[5] A.L. Milyushkin, A.E. Karnaeva, Unsymmetrical dimethylhydrazine transformation products: A review, Science of the Total Environment 891 (2023). https://doi.org/10.1016/j.scitotenv.2023.164367.

[6] Z.B. Khesina, A.E. Karnaeva, I.S. Pytskii, A.K. Buryak, The mysterious mass death of marine organisms on the Kamchatka Peninsula: A consequence of a technogenic impact on the environment or a natural phenomenon?, Mar Pollut Bull 166 (2021). https://doi.org/10.1016/j.marpolbul.2021.112175.

[7] A.Yu. Sholokhova, D.D. Matyushin, O.I. Grinevich, S.A. Borovikova, A.K. Buryak, Intelligent Workflow and Software for Non-Target Analysis of Complex Samples Using a Mixture of Toxic Transformation Products of Unsymmetrical Dimethylhydrazine as an Example, Molecules 28 (2023) 3409. https://doi.org/10.3390/molecules28083409.

314 [8] A.E. Karnaeva, A.L. Milyushkin, Z.B. Khesina, A.K. Buryak, 1-Methyl-1H-1,2,4-triazole as the
315   main marker of 1,1-dimethylhydrazine exposure in plants, Environmental Science and Pollution
316   Research 29 (2022) 64225–64231. https://doi.org/10.1007/s11356-022-22157-y.

317 [9] R. Aalizadeh, M.-C. Nika, N.S. Thomaidis, Development and application of retention time
318   prediction models in the suspect and non-target screening of emerging contaminants, J Hazard
319   Mater 363 (2019) 277–285. https://doi.org/10.1016/j.jhazmat.2018.09.047.

320 [10] A.L. Milyushkin, K.P. Birin, D.D. Matyushin, A. V. Semeikin, S.D. Iartsev, A.E. Karnaeva, A.
321   V. Uleanov, A.K. Buryak, Isomeric derivatives of triazoles as new toxic decomposition products
322   of 1,1-dimethylhydrazine, Chemosphere 217 (2019) 95–99.
323   https://doi.org/10.1016/j.chemosphere.2018.10.155.

324 [11] N. V. Ul'yanovskii, D.E. Lakhmanov, I.I. Pikovskoi, D.I. Falev, M.S. Popov, A.Yu.
325   Kozhevnikov, D.S. Kosyakov, Migration and transformation of 1,1-dimethylhydrazine in peat
326   bog soil of rocket stage fall site in Russian North, Science of The Total Environment 726 (2020)
327   138483. https://doi.org/10.1016/j.scitotenv.2020.138483.

328 [12] A.D. Smolenkov, I.A. Rodin, A. V. Shpak, O.A. Shpigun, 1-Formyl-2,2-dimethylhydrazine as a
329   new decomposition product of 1,1-dimethylhydrazine, Int J Environ Anal Chem 87 (2007) 351–
330   359. https://doi.org/10.1080/03067310601068882.

331 [13] N.V. Ul'yanovskii, D.S. Kosyakov, I.I. Pikovskoi, Yu.G. Khabarov, Characterisation of
332   oxidation products of 1,1-dimethylhydrazine by high-resolution orbitrap mass spectrometry,
333   Chemosphere 174 (2017) 66–75. https://doi.org/10.1016/j.chemosphere.2017.01.118.

334 [14] D.D. Matyushin, A.K. Buryak, Gas Chromatographic Retention Index Prediction Using
335   Multimodal Machine Learning, IEEE Access 8 (2020) 223140–223155.
336   https://doi.org/10.1109/ACCESS.2020.3045047.

337 [15] A.Yu. Sholokhova, O.I. Grinevich, D.D. Matyushin, A.K. Buryak, Machine learning-assisted
338   non-target analysis of a highly complex mixture of possible toxic unsymmetrical
339   dimethylhydrazine transformation products with chromatography-mass spectrometry,
340   Chemosphere 307 (2022) 135764. https://doi.org/10.1016/j.chemosphere.2022.135764.

341 [16] D.D. Matyushin, A.Yu. Sholokhova, A.K. Buryak, A deep convolutional neural network for the
342   estimation of gas chromatographic retention indices, J Chromatogr A 1607 (2019) 460395.
343   https://doi.org/10.1016/j.chroma.2019.460395.

344 [17] P.B. Joshi, Navigating with chemometrics and machine learning in chemistry, Artif Intell Rev 56
345   (2023) 9089–9114. https://doi.org/10.1007/s10462-023-10391-w.

346 [18] A. Kajtazi, G. Russo, K. Wicht, H. Eghbali, F. Lynen, Facilitating structural elucidation of small
347   environmental solutes in RPLC-HRMS by retention index prediction, Chemosphere 337 (2023)
348   139361. https://doi.org/10.1016/j.chemosphere.2023.139361.

349 [19] B. Sepehri, R. Ghavami, S. Farahbakhsh, R. Ahmadi, Machine learning-based quantitative
350   structure–retention relationship models for predicting the retention indices of volatile organic
351   pollutants, International Journal of Environmental Science and Technology 19 (2022) 1457–
352   1466. https://doi.org/10.1007/s13762-021-03271-9.

353 [20] C. Veenaas, A. Linusson, P. Haglund, Retention-time prediction in comprehensive two-
354   dimensional gas chromatography to aid identification of unknown contaminants, Anal Bioanal
355   Chem 410 (2018) 7931–7941. https://doi.org/10.1007/s00216-018-1415-x.

356 [21] S. Riahi, E. Pourbasheer, M.R. Ganjali, P. Norouzi, Investigation of different linear and
357   nonlinear chemometric methods for modeling of retention index of essential oil components:

19

358   Concerns to support vector machine, J Hazard Mater 166 (2009) 853–859.
359   https://doi.org/10.1016/j.jhazmat.2008.11.097.

360 [22] K. Ng, N. Alygizakis, A. Androulakakis, A. Galani, R. Aalizadeh, N.S. Thomaidis, J. Slobodnik,
361   Target and suspect screening of 4777 per- and polyfluoroalkyl substances (PFAS) in river water,
362   wastewater, groundwater and biota samples in the Danube River Basin, J Hazard Mater 436
363   (2022) 129276. https://doi.org/10.1016/j.jhazmat.2022.129276.

364 [23] M.D. Khrisanfov, D.D. Matyushin, A.S. Samokhin, A general procedure for finding potentially
365   erroneous entries in the database of retention indices, Anal Chim Acta 1297 (2024) 342375.
366   https://doi.org/10.1016/j.aca.2024.342375.

367 [24] T. Vrzal, M. Malečková, J. Olšovská, DeepReI: Deep learning-based gas chromatographic
368   retention index predictor, Anal Chim Acta 1147 (2021) 64–71.
369   https://doi.org/10.1016/j.aca.2020.12.043.

370 [25] N. V. Ul'yanovskii, D.S. Kosyakov, I.I. Pikovskoi, M.S. Popov, Study of the Products of
371   Oxidation of 1,1-Dimethylhydrazine by Nitrogen Dioxide in an Aqueous Solution by High-
372   Resolution Mass Spectrometry, Journal of Analytical Chemistry 73 (2018) 1223–1228.
373   https://doi.org/10.1134/S1061934818130130.

374 [26] H. Wang, P. Yang, Y. Song, D. Kong, J. Lu, Formation of N-nitrosodimethylamine from
375   dimethylamine compounds during UV disinfection in the presence of nitrite, Sep Purif Technol
376   338 (2024) 126571. https://doi.org/10.1016/j.seppur.2024.126571.

377 [27] R. Shen, S.A. Andrews, Demonstration of 20 pharmaceuticals and personal care products
378   (PPCPs) as nitrosamine precursors during chloramine disinfection, Water Res 45 (2011) 944–
379   952. https://doi.org/10.1016/j.watres.2010.09.036.

380 [28] A. Yahaya, D. Babatunde, L.W.B. Olaniyan, O. Agboola, Application of chromatographic
381   techniques in the analysis of total nitrosamines in water, Heliyon 6 (2020) e03447.
382   https://doi.org/10.1016/j.heliyon.2020.e03447.

383 [29] M.S. Popov, N. V. Ul'yanovskii, D.S. Kosyakov, Gas Chromatography–Mass Spectrometry
384   Quantification of 1,1-Dimethylhydrazine Transformation Products in Aqueous Solutions:
385   Accelerated Water Sample Preparation, Molecules 26 (2021) 5743.
386   https://doi.org/10.3390/molecules26195743.

387 [30] N. V. Ul'yanovskii, D.S. Kosyakov, S.A. Pokryshkin, K.G. Bogolitsyn, Determination of
388   transformation products of 1,1-dimethylhydrazine by gas chromatography–tandem mass
389   spectrometry, Journal of Analytical Chemistry 70 (2015) 1553–1560.
390   https://doi.org/10.1134/S1061934815130080.

391 [31] M.S. Popov, N. V. Ul'yanovskii, D.S. Kosyakov, Direct quantification of 1,1-dimethylhydrazine
392   transformation products in sandy soil by thermal desorption gas chromatography – Tandem mass
393   spectrometry, Microchemical Journal 197 (2024) 109833.
394   https://doi.org/10.1016/j.microc.2023.109833.

395 [32] E. Kováts, Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1:
396   Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone, Helv Chim Acta
397   41 (1958) 1915–1932. https://doi.org/10.1002/hlca.19580410703.

398 [33] H. van Den Dool, P. Dec. Kratz, A generalization of the retention index system including linear
399   temperature programmed gas—liquid partition chromatography, J Chromatogr A 11 (1963) 463–
400   471. https://doi.org/10.1016/S0021-9673(01)80947-X.

401 [34] B. d'Acampora Zellner, C. Bicchi, P. Dugo, P. Rubiolo, G. Dugo, L. Mondello, Linear retention
402 indices in gas chromatographic analysis: a review, Flavour Fragr J 23 (2008) 297–314.
403 https://doi.org/10.1002/ffj.1887.

404 [35] D.D. Matyushin, A.Yu. Sholokhova, A.E. Karnaeva, A.K. Buryak, Various aspects of retention
405 index usage for GC-MS library search: A statistical investigation using a diverse data set,
406 Chemometrics and Intelligent Laboratory Systems 202 (2020) 104042.
407 https://doi.org/10.1016/j.chemolab.2020.104042.

408 [36] C. Hu, Y. Zhang, Y. Zhou, Z. Liu, X. Feng, Unsymmetrical dimethylhydrazine and related
409 compounds in the environment: Recent updates on pretreatment, analysis, and removal
410 techniques, J Hazard Mater 432 (2022) 128708. https://doi.org/10.1016/j.jhazmat.2022.128708.

411 [37] B. Jurado-Sánchez, E. Ballesteros, M. Gallego, Comparison of microwave assisted, ultrasonic
412 assisted and Soxhlet extractions of N-nitrosamines and aromatic amines in sewage sludge, soils
413 and sediments, Science of The Total Environment 463–464 (2013) 293–301.
414 https://doi.org/10.1016/j.scitotenv.2013.06.002.

415 [38] A.K. Buryak, T.M. Serdyuk, A. V. Ul'yanov, Investigation of the reaction products of
416 unsymmetrical dimethylhydrazine with potassium permanganate by gas chromatography-mass
417 spectrometry, Theoretical Foundations of Chemical Engineering 45 (2011) 550–555.
418 https://doi.org/10.1134/S0040579510051057.

419 [39] B.N. Kenessov, J.A. Koziel, T. Grotenhuis, L. Carlsen, Screening of transformation products in
420 soils contaminated with unsymmetrical dimethylhydrazine using headspace SPME and GC–MS,
421 Anal Chim Acta 674 (2010) 32–39. https://doi.org/10.1016/j.aca.2010.05.040.

422 [40] B. Kenessov, S. Batyrbekova, M. Nauryzbayev, T. Bekbassov, M. Alimzhanova, L. Carlsen,
423 GC-MS Determination of 1-Methyl-1H-1,2,4-triazole in Soils Affected by Rocket Fuel Spills in
424 Central Kazakhstan, Chromatographia 67 (2008) 421–424. https://doi.org/10.1365/s10337-008-
425 0535-4.

426 [41] F. Erdmann, G. Rochholz, H. Schutz, Retention-indices on OV-1 of approximately 170
427 commonly used pesticides, Mikrochim Acta 106 (1992) 219–226.
428 https://doi.org/10.1007/BF01242093.

429 [42] D. V. Ovchinnikov, S.A. Vakhrameev, D.I. Falev, N. V. Ul'yanovskii, D.S. Kosyakov, Rapid
430 Simultaneous Quantification of 1-Formyl-2,2-Dimethylhydrazine and Dimethylurea Isomers in
431 Environmental Samples by Supercritical Fluid Chromatography–Tandem Mass Spectrometry,
432 Molecules 27 (2022) 5025. https://doi.org/10.3390/molecules27155025.

433 [43] D.D. Matyushin, A.Yu. Sholokhova, A.K. Buryak, Deep Learning Based Prediction of Gas
434 Chromatographic Retention Indices for a Wide Variety of Polar and Mid-Polar Liquid Stationary
435 Phases, Int J Mol Sci 22 (2021) 9194. https://doi.org/10.3390/ijms22179194.

436

437