analytical chemistry

pubs.acs.org/ac Article

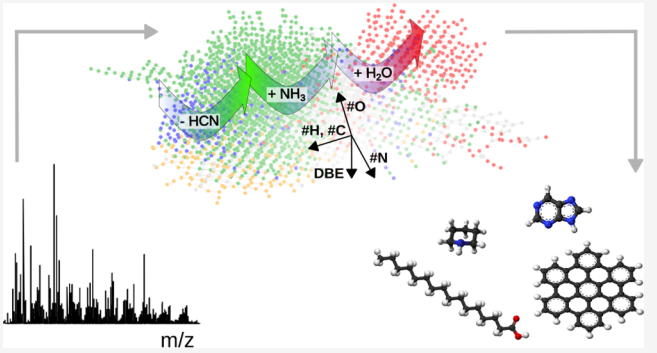
Network Analysis Reveals Spatial Clustering and Annotation of Complex Chemical Spaces: Application to Astrochemistry

Alexander Ruf* and Grégoire Danger

Cite This: *Anal. Chem.* 2022, 94, 14135−14142 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: How are molecules linked to each other in complex systems? In a proof-of-concept study, we have developed the method mol2net (https://zenodo.org/record/7025094) to generate and analyze the molecular network of complex astrochemical data (from high-resolution Orbitrap MS[1] analysis of H_2O:CH_3OH:NH_3 interstellar ice analogs) in a data-driven and unsupervised manner, without any prior knowledge about chemical reactions. The molecular network is clustered according to the initial NH_3 content and unlocked HCN, NH_3, and H_2O as spatially resolved key transformations. In comparison with the PubChem database, four subsets were annotated: (i) saturated C-backbone molecules without N, (ii) saturated N-backbone molecules, (iii) unsaturated C-backbone molecules without N, and (iv) unsaturated N-backbone molecules. These findings were validated with previous results (e.g., identifying the two major graph components as previously described N-poor and N-rich molecular groups) but with additional information about subclustering, key transformations, and molecular structures, and thus, the structural characterization of large complex organic molecules in interstellar ice analogs has been significantly refined.



INTRODUCTION

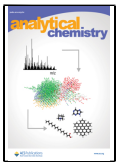
The complexity of chemical systems is based on species' interaction patterns.[1] This leads to emergent collective systems' properties such self-organization,[2] autocatalysis,[3] or global reactivity,[4] properties that are hidden on the level of individual species.

Astrochemical samples are here of specific relevance, as their outstanding molecular diversities exceed those of other chemical samples (e.g., biochemical or environmental data).[5−7] The evolution of organic material, from molecular clouds toward planetary systems, can be simulated by laboratory experiments.[8] The benefit of such experiments is two-fold: (i) controlled experimental parameter space and (ii) in-depth chemical analysis (e.g., high-resolving, high accurate and sensitive mass spectrometry).[6] The resulting high-dimensional data sets enable to study the complete diversity of astrochemical organic molecules from a system's chemistry perspective.[6]

The comprehensive and unbiased analysis of complex chemical systems is a huge challenge and implies constant development of analytical methodology with innovative solutions for their data analyses.[6,9] Network analysis represents an excellent tool to analyze complex systems as complete as possible, in general[10] or in (astro)chemical context.[6,11] This approach covers a wide range of chemical applications, including astrochemistry,[12] meteorites,[4] prebiotic chemistry,[13] origin(s) of Life,[14] astrobiology,[15] systems chemistry,[16] autocatalysis,[3] environmental science,[17] or biochemistry.[18] Networks allow one to visualize[19] and further quantify complex interactions as they unveil systemic patterns instead of addressing individual reactions only. They thus shed light into the organization and interconnections of chemical systems as a whole.[20] Even though reaction network analysis has powerfully demonstrated the characterization of chemical systems with partially unknown reaction rates,[21] current reaction network methods have two key drawbacks because of their supervised architecture: (i) individually addressed reactions (side reactions are ignored) and (ii) reaction rate dependence on physical environment (e.g., gas-phase vs solid-state-mediated reactions). In other words, current reaction network analysis depends on prior knowledge about the individual reactions.[22]

In this proof-of-concept study, we developed the method mol2net[24] that has analyzed the interactions of many tens of thousands of different chemical compounds. We applied data-driven graph-based network analysis to astrochemical data as

Received: March 22, 2022
Accepted: September 6, 2022
Published: October 9, 2022



https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

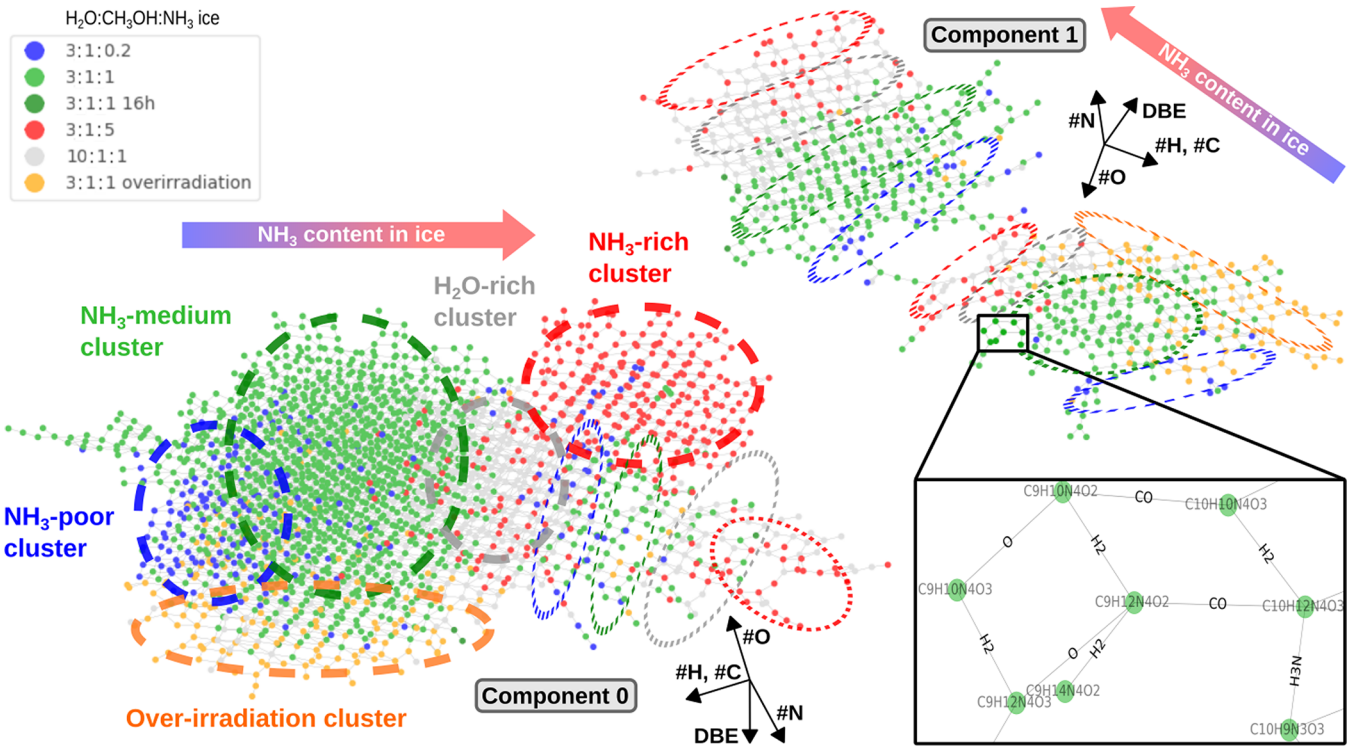


Figure 1. Clustering of the molecular network. The molecular network is grouped into two major graph components (components 0 and 1, with 2357 nodes/5608 edges and 752 nodes/1431 edges, respectively) and different clusters (blue, green, gray, and red clusters, illustrated by dashed-line zones) that are related to the NH_3 content in the initial ice. Fifteen different H_2O:CH_3OH:NH_3 ice samples that varied in their initial ice composition (3:1:0.2, 3:1:1, 3:1:5, and 10:1:1) were used. An additional physical process, namely, over-irradiation (organge cluster), shows chemical properties that are different to those of cold ice chemistry (at 77 K). Element gradients further organize the network.

an example for complex chemical systems. A previously well-described data set of interstellar ice analogs from high-resolution Orbitrap MS^1 data (without fragmentation information, no MS^n data required) was used[23−27] that allowed for validation of this novel method. This data set represents 15 different H_2O:CH_3OH:NH_3 ice samples (3:1:0.2, 3:1:1, 3:1:5, and 10:1:1) that are grouped into three categories: (i) NH_3 content, (ii) H_2O content, and (iii) additional physical processes (over-irradiation) of the formed residue. The molecular network was generated by mass difference matching[28−30] of 3270 CHNO molecular formulas (between 199 and 365 amu, atomic mass unit, nodes in the network) with a set of in silico transformations (edges in the network). We want to highlight that no prior knowledge about chemical reactions (e.g., reaction rates) was required. In contrast to the previous studies,[26,27] we have here not only characterized the compounds independently but we also put emphasis on their interconnectedness by discussing (i) the clustering of the molecular network, (ii) the key transformations, (iii) the subnetworks and PubChem database unlock structural information, and (iv) the annotation of complex chemical spaces.

RESULTS AND DISCUSSION

Clustering of the Molecular Network. Figure 1 illustrates the molecular network of interstellar ice analogs, as whole and zoom-in pictures. Using a transformation set of {H_2, O, CO, NH_3} (named as "minimal set"), 95% of all 3270 molecular formulas could be linked within two graph components (components 0 and 1, with 2357 nodes/5608 edges and 752 nodes/1431 edges, respectively, Supporting Figure 1). The two graph components represent two previously reported distinct N-poor and N-rich molecular groups (Supporting Figure 4).[26,27] Transformations were chosen in a data-driven and unsupervised manner as informed from prior distance matrix analysis. Our goal was to keep the set of transformations as minimal as possible to maximize interpretation of individual transformations. We have tested six other transformations {CHN, CH_3N, H_2O, CO_2, CH_2O, CHNO} and found that their step-by-step addition is invariant with respect to the network topology and addition sequence, as tested by (i) network clustering, (ii) shape of degree distributions, and (iii) component frequency (Supporting Figure 3). Thus, the four transformations from the "minimal set" explain the complete connectivity of the analyzed greater than 3000 compounds in the interstellar ice analogs. These transformations represent potential reactions in interstellar ices, based on previous laboratory experiments (H_2,[31,32] O,[33] NH_3[34−36]).

The molecular network is clustered according to the NH_3 content (arrow "NH_3 content in ice", NH_3 poor (blue), NH_3 medium (green), H_2O rich (gray), and NH_3 rich (red) Figure 1). The trend from NH_3 poor to NH_3 rich is intuitive but the H_2O-rich cluster in between highlights the importance of H_2O-mediated organic chemistry inside ices.[26,37,38] Molecules in the over-irradiation cluster (orange, additional irradiation of the resulting ice residue at 300 K[26,27,39,40]) are aligned orthogonal to the "NH_3 content gradient", accompanied by higher DBE (double bond equivalent) and lower #O (decarboxylation/formation of unsaturated, aromatic structures[26,27]). The same cluster gradient has been observed for component 1 as well and furthermore substructured in each component. This

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

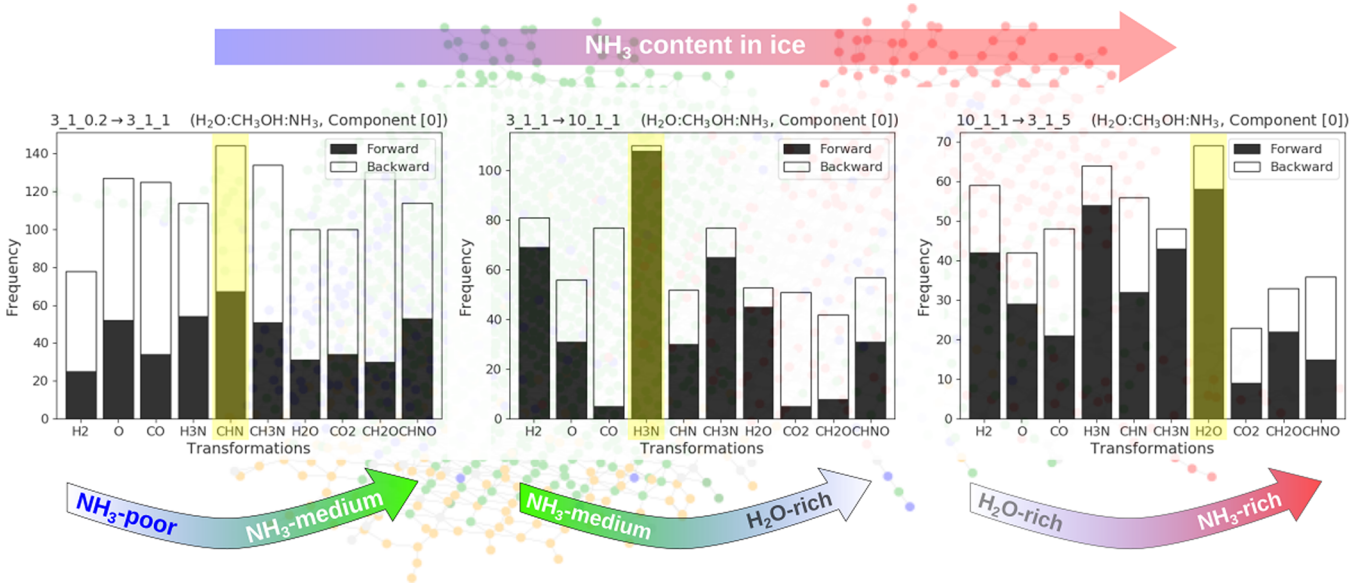


Figure 2. Key transformations. Cluster-specific transitions illustrate potential key transformations: CHN, NH_3, and H_2O (with directions) along the NH_3 content gradient in the initial ice (NH_3-poor → NH_3-medium cluster, NH_3-medium → H_2O-rich cluster and H_2O-rich → NH_3-rich cluster).

indicates that similar initial reactivities in ices result in two different types of chemical pathways (components 0 and 1 housing the minimal set H_2, O, CO, NH_3). The novelty of this type of clustering is that it is based on species' interactions and not on the species' (isolated) characteristics what result in highly resolved clustering structures (two gradients/substructures per graph component).

Element maps represent molecular cartography and further organize the complex molecular network (Figure 1, Supporting Figure 4). For both components 0 and 1, hydrogen and carbon counts (#H, #C) are correlated to each other but orthogonally correlated to nitrogen counts (#N). Molecules in the NH_3-poor cluster (blue) show higher #C and #H and lower #N, compared to the NH_3-rich cluster (red, competition between C and N in molecular backbone formation[26,27]). For component 0, oxygen counts (#O) and DBE are aligned orthogonal to #C and #H, with low #O and high DBE at the bottom (including the over-irradiation cluster). For component 1 (high DBE), #O is fairly constant.

The different clusters show different chemical characteristics[41] (Supporting Figure 5). H/C and N/C ratios increase from NH_3-poor to NH_3-rich ices. N/C ratios get very high for high NH_3 content (N/C up to 1.3) that may represent cyclic secondary amines (in accordance to high H/C and low DBE for component 0). The major chemical family is CHNO molecules, next to CHN ones. The number of CHN molecules slightly increases with higher NH_3 content. Molecules in the over-irradiation cluster show a depletion of CHNO molecules (relative to CHN ones, in agreement with lower #O).

Key Transformations. Figure 2 illustrates key transformations in the molecular network that connect molecules along different network clusters, for example, for the transition from the NH_3-poor toward NH_3-medium cluster.

Transitions from the NH_3-poor cluster to the NH_3-medium cluster are dominated by CHN and CH_3N subtractions (that potentially represent HCN and CH_2NH transformations with backward/forward transformations >1 (Figure 2). The NH_3-medium and H_2O-rich clusters are majorly linked by NH_3 additions, followed by the addition of H_2 (in agreement with increased H/C and decreased DBE, Figure 2). H_2O is added most for the transition from the H_2O-rich cluster toward the NH_3-rich cluster, followed by of NH_3 addition. This indicates that H_2O may not only passively contribute to the complex chemistry in ices[38] but acts as an active reagent[42−44] or catalyst.[37]

Transformation patterns are similar between components 0 and 1, on average, and for cluster−cluster transitions (Supporting Figure 6). In detail, there are enrichments in H_2 (majorly for NH_3-poor ices) and CHN as well as depletions in H_2O and CO_2 for component 1 (relative to component 0). Furthermore, the transition from the H_2O-rich cluster toward the NH_3-rich cluster in component 1 is governed by CHN transformations instead of H_2O, which indicates a more prominent role of nitrogen chemistry for component 1 molecules, compared to those of component 0.

Subnetworks and PubChem Database Unlock Structural Information. Figure 3 shows detailed information about molecular structures as inferred via two approaches: (i) splitting the minimal set of transformations and (ii) comparison to the PubChem database.[45] We want to remind here that structural information has been unlocked from mass spectrometric MS^1 data, without fragmentation information (no MS^n data required) that at has not been reported to date (to the best of our knowledge).

Subnetworks were generated by splitting the minimal set of transformations. Each element from {H_2, O, CO, NH_3} ("minimal set" of transformations) got iteratively removed to unveil detailed information about molecular subsets (Figure 3A). The resulting subnetworks consisted of both many small components and moderate-sized ones (Supporting Figure 7). This implies that removing each of the four transformations has a significant effect on the stability of the network (many small components). However, some local stable molecular groups remain connected (moderate-sized ones). In a next step, the key components have been identified based on (i) size of components and (ii) maximal variation of map

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

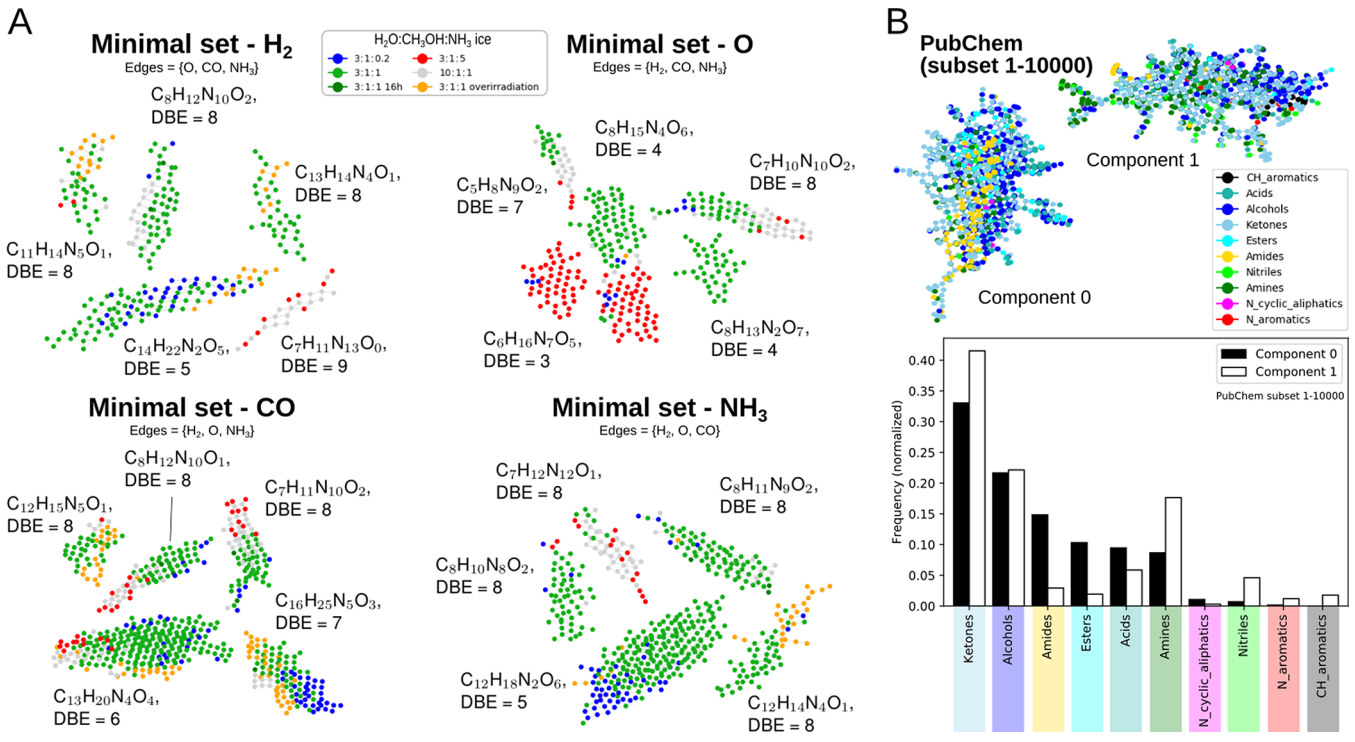


Figure 3. Subnetworks and PubChem database unlock structural information. (A) Splitted minimal set of transformations. Illustrated molecular formulas and DBEs represent average numbers (median). (B) Comparison to the PubChem database.[45] Similary to the ice molecular network, two major graph components were identified using the same set of transformations ({H_2, O, CO, NH_3}, "minimal set" of transformations) for the PubChem network (details about the PubChem data are given in the Supporting Information), unraveling their compound class characteristics.

parameters (e.g., #H, #C, or DBE). Each of the four groups for the selected five key components (minimal set - H_2, minimal set - O, minimal Set - CO, minimal set - NH_3) consist of 24−253 molecular formulas and show distinct chemical properties (Figure 3A, Supporting Figure 8). The "minimal set - H_2" and "minimal set - NH_3" subnetworks specifically show components that are distinct in #N (e.g., N_2 or N_{10} subsets), whereas "minimal set - O" and "minimal set - CO" subnetworks have distinct #C components (e.g., C_5 or C_{16} subsets). Furthermore, the components also cluster with respect to the original ice samples (e.g., NH_3-enriched component "minimal set - O"). All four splitted "minimal set" graphs commonly show distinct #O and in DBE subsets.

The comparison with the PubChem database[45] has revealed further insights into molecular structures and compound classes. We have applied the same workflow for network generation and analysis for the PubChem data as done for those of the ice (transformation set of {H_2, O, CO, NH_3}"minimal set", extraction of two largest graph components, Figure 3B). We have randomly sampled two PubChem subsets with 10,000 compounds each and compared their two major graph components with those of the ice molecular network, using element maps and transformation frequencies (Supporting Figure 9). Qualitatively, element maps and transformation frequencies match between the ice and PubChem data, but H maps showed significant variation for the PubChem subset 10,001−20,000. Thus, we continued with the detailed analysis using the PubChem subset 1−10,000 only (details about the PubChem data are given in the Supporting Information). Based on IUPAC names, compound classes have been defined (details are given in the Supporting Information), and the network has been color coded accordingly. Figure 3B illustrates that component 0 is enriched in the number of amides, esters, and acids, whereas component 1 shows higher counts in ketones, amines, nitriles, and aromatics. The enrichment of the number of aromatic compounds, N- and C-based ones, is in accordance with high DBE values as observed for the ice data (Supporting Figure 4).

Annotation of Complex Chemical Spaces. All combined results were mapped onto the original molecular network of interstellar ice analogs (Figure 4). This reveals spatial clustering and annotation of the network. The two major graph components were thus organized into different clusters that show distinct elemental characteristics (e.g., C or N counts, inferred from splitting the "minimal set" of transformation) and were assigned with compound class likelihoods (inferred from PubChem database). We further show that a gradient in N- vs C-backbone molecules in each component highlights the competitive two different types of chemistry, namely, the carbon- and nitrogen-driven ones (in relation to N/C ratios). Four subsets were identified: (i) saturated C-backbone molecules without N (e.g., open-chain compounds; component 0, left), (ii) saturated N-backbone molecules (e.g., cyclic secondary amines; component 0, right), (iii) unsaturated C-backbone molecules without N (e.g., polycyclic aromatic hydrocarbons, PAHs; component 1, bottom), and (iv) unsaturated N-backbone molecules (e.g., aromatic N-heterocycles; component 1, top). In short, component 0 represents aliphatic molecules and component 1 unsaturated (potentially aromatic) molecules, with both C- and N-backbones.

Component 0 represents saturated C-backbone molecules and saturated N-backbone ones. From left to right, nitrogen-based chemistry dominates over carbon chemistry, that is supported by N/C ratios. On the total left, a C-backbone cluster with C_{12} unsaturated molecules was identified (DBE =

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

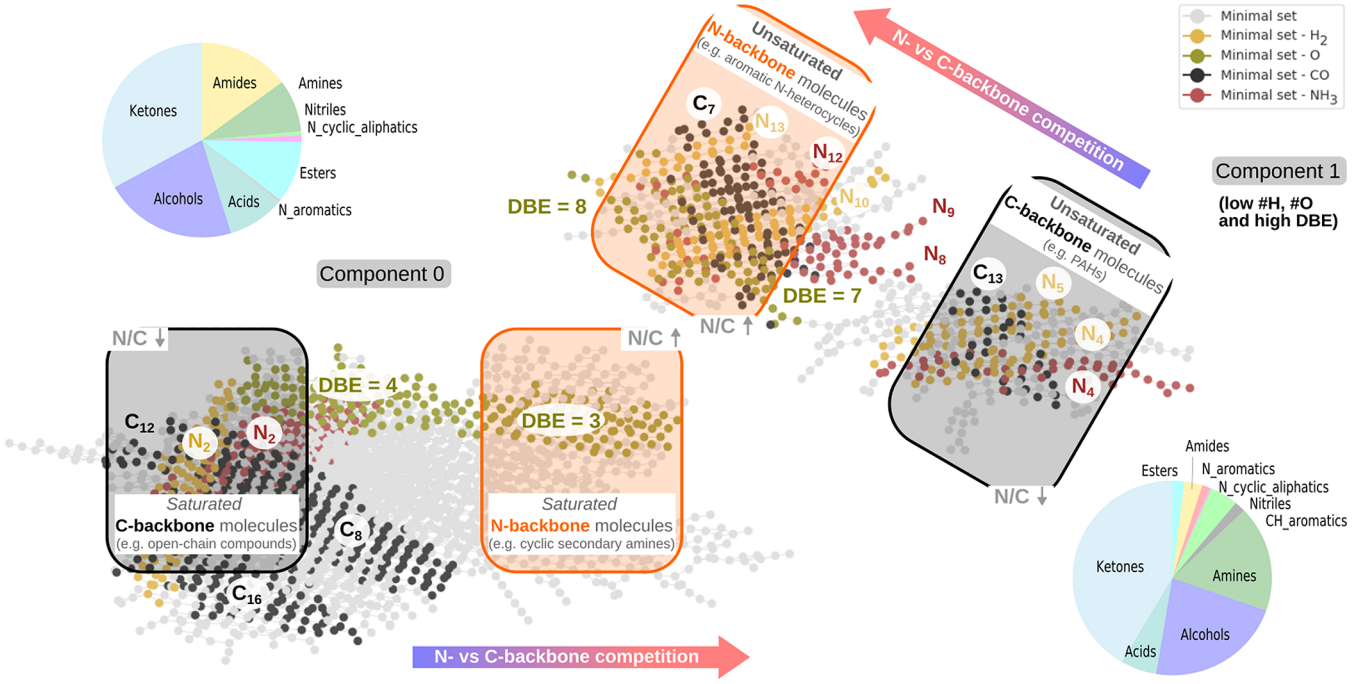


Figure 4. Annotation of complex chemical spaces. All combined results were mapped onto the original molecular network of interstellar ice analogs. The colors are related to the different splitted minimal sets of transformations (gray, minimal set; dark golden, minimal set - H_2; olive, minimal set - O; black, minimal set - CO; brown-red, minimal set - NH_3). Pie charts are derived from PubChem data.

8, N/C = 0), whereas on the right, a N-backbone cluster represents N_7 saturated molecules (DBE = 3, N/C = 2.5). Component 0 molecules might be specifically enriched in the number of amides (15%), esters (11%), and acids (9%).

Component 1 shows unsaturated C- and N-backbone molecules with low #H and #O. N-Based molecules are enriched from bottom to top. On the bottom, a C-backbone cluster with N_4 unsaturated molecules was identified (DBE = 8, N/C = 0.5), whereas on top, a N-backbone cluster shows N_{13} unsaturated molecules (DBE = 8, N/C = 3). Component 1 molecules might be specifically enriched in the number of ketones (42%), amines (17%), and nitriles (5%), as well as in CH- (e.g., polycyclic aromatic hydrocarbons, PAHs) and N-based aromatic compounds (e.g., N-heterocycles) where the latter were almost absent in component 0.

In comparison to previous studies by Fresneau et al. (based on average values for atomic parameters)[26] and Gautier et al. (based on ternary diagrams of atomic-ratios),[27] the here presented network analysis sets emphasis species' interactions, and both confirm but also refine current knowledge. The previously reported two distinct molecular groups were validated, but the network analysis significantly refined their characterizations because of enhanced resolution of the molecular space (multidimensional representation) that enabled the spatial differentiation between NH_3-medium and NH_3-poor/NH_3-rich ices (H_2O:CH_3OH:NH_3 = 3:1:1 from H_2O:CH_3OH:NH_3 = 3:1:0.2/3:1:5). Whereas Fresneau et al.[26] and Gautier et al.[27] described them as "carbon- and nitrogen-enriched distributions", we show here that the discriminant parameter is unsaturation (DBE, and not C or N). While Fresneau et al.[26] reported similar overall #O, the network's high molecular resolution enabled differentiation in #O for saturated compounds, whereas unsaturated molecules showed similar #O (Supporting Figure 4). Additionally, Fresneau et al.[26] and Gautier et al.[27] have reported that H_2O- and NH_3-rich ices (H_2O:CH_3OH:NH_3 = 10:1:1 and 3:1:5, respectively) show similar degrees in N incorporation, but the network analysis unlocked variation (Supporting Figure 4). Whereas Gautier et al.[27] reported that NH_3-rich ices "lead to a group of unsaturated molecules in the final residue, while H_2O rich ices lead to saturated ones", but networks show that there are both saturated and unsaturated molecules for both NH_3- and H_2O-rich ices.

Conclusions. In a proof-of-concept study, we have analyzed the network of complex astrochemical data (from high-resolution Orbitrap MS^1 analysis of H_2O:CH_3OH:NH_3 interstellar ice analogs). The molecular network is clustered according to the initial NH_3 content and unlocked HCN, NH_3, and H_2O as spatially resolved key transformations. In comparison with the PubChem database, four subsets were structurally annotated. This method was validated with previous results by Fresneau et al.[26] and Gautier et al.[27] Current knowledge was confirmed but also refined (e.g., characterization of two distinct molecular groups).

The implications of the presented network analysis are not specific to astrochemistry but manifold for complex chemical systems, including (i) the characterization of nondetected species in any sample, (ii) the comprehensive elucidation of chemical reactivity (e.g., radical vs thermal chemistry), or (iii) studying of emergence behavior. For astrochemistry, these findings provide the very first insights about transformation patterns and structures of large complex organic molecules in interstellar ice analogs (>200 amu, those of prebiotic interest). To date, information about large molecules' chemistries is hidden to traditional chemical models[46−49] as detailed information on their individual reaction rates and mechanisms is missing. Even though the here newly described graph-based network analysis does not directly address mechanistic insights, it acts as a screening tool to unveil all possible transformation pathways that can now be further investigated by upcoming

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

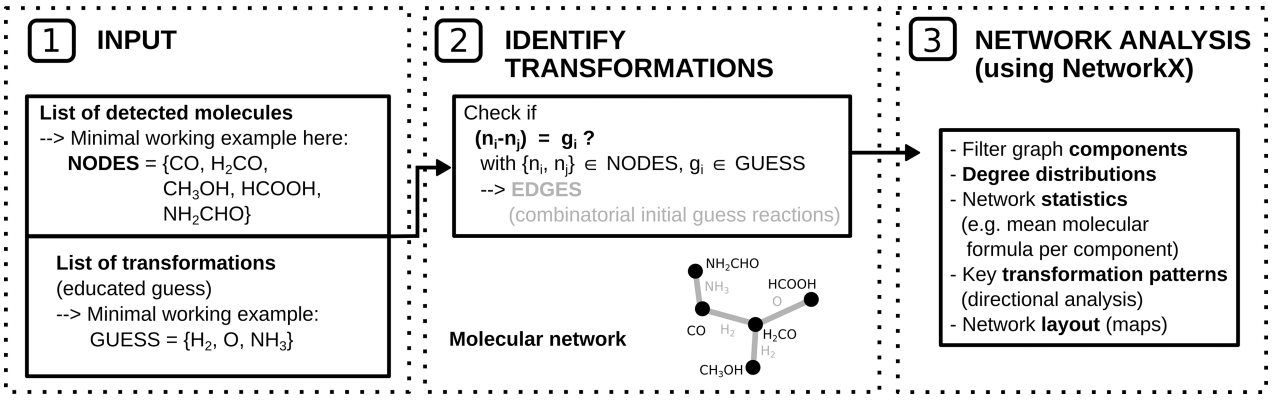


Figure 5. Schematic algorithmic workflow of the method mol2net:[24] (1) input, (2) identify transformations, and (3) network analysis (using NetworkX).

targeted experimental and/or computational studies. Furthermore, the identified transformation patterns might be generalized that would allow the characterization of non-detected species as well. This has strong implications for astrochemistry to probe the chemistries of large complex organic molecules that would otherwise be hidden to observations.

DATA SET AND METHODS

Data Set of Interstellar Ice Analogs. This data set represents 15 different H_2O:CH_3OH:NH_3 ice samples (3:1:0.2, blue, 3 replicates; 3:1:1, green, 3 replicates; 3:1:5, red, 2 replicates; 10:1:1, gray, replicates/over-irradiation, orange, 3 replicates). Ice deposition/irradiation time was 48 h except for the "3:1:1 16 h" ice with a deposition/irradiation time of 16 h. All details about the experiments and their high resolution Orbitrap mass spectrometric analysis have been described by Fresneau et al.[26] and Gautier et al.[27] Data will be made available on request.

Method: Molecular Network Generation and Analysis. The algorithmic workflow of the method mol2net[24] is organized as follows (Figure 5): (1) input, (2) identify transformations, and (3) network analysis (using NetworkX). (1) Input: Theoretical masses of prior assigned 3270 unique molecular formulas were used for network generation, as performed via Python.[50] (2) Identify transformations: Theoretical masses (nodes in the network) were connected by edges (transformations) if their mass differences match with the theoretical mass differences of the transformations (e.g., for the "minimal set" of transformations, used as basis in this study): {ΔH_2, ΔO, ΔCO, ΔNH_3}, with ΔH_2 = 2.01565 amu, ΔO = 15.994915 amu, ΔCO = 27.994915 amu, and ΔNH_3 = 17.026549 amu, with amu = atomic mass unit. Transformations were chosen in a data-driven and unsupervised manner as informed from prior distance matrix analysis. Our goal was to keep the set of transformations as minimal as possible to maximize interpretation of individual transformations. We have tested six other transformations {CHN, CH_3N, H_2O, CO_2, CH_2O, CHNO} and found that their step-by-step addition is invariant with respect to the network topology and addition sequence, as tested by (i) network clustering, (ii) shape of degree distributions, and (iii) component frequency (Supporting Figure 3). The term "clustering" is used here as a result of the color-coded network layout (by sample) that represents compound interactions and should not be mismatched by traditional, statistical clustering techniques. (3) Network analysis (using NetworkX): Molecular network analysis was performed via Python,[50] and specifically, NetworkX[51] was used for graph computations and network layouts (graphviz_layout, "nx.nx_pydot.graphviz_layout(G)"). The following parameters were majorly addressed: filtered graph components, shape of degree distributions, network statistics, key transformation patterns, and plotted network. All computations were performed on a stand-alone computer. The Python code of mol2net is available on Zenodo and GitHub.[24]

PubChem Database. The PubChem database was downloaded via the Python package pubchempy (https://pubchempy.readthedocs.io/en/latest/). We have filtered the downloaded data for the same chemical elements (CHNO) and mass range (199 and 365 amu) as present in the ice data set. Two subsets with 10,000 compounds each were randomly sampled (1−10,000 and 10,001−20,000; details about CID identifiers are given in the Supporting Information). Compound classes have been defined by data mining their IUPAC name endings as following: CH_aromatics = classes.str.endswith('benzene'), CH_aromatics = classes.str.endswith('naphthalene'), CH_aromatics = classes.str.endswith('pyrene'), Acids = classes.str.endswith('acid'), Alcohols = classes.str.endswith('ol'), Ketones = classes.str.endswith('one'), Esters = classes.str.endswith('oate'), Esters = classes.str.endswith('acetate'), Amides = classes.str.endswith('amide'), Amides = classes.str.endswith('carbamate'), Nitriles = classes.str.endswith('nitrile'), Amines = classes.str.endswith('amine'), N_cyclic_aliphatics = classes.str.endswith ('piperidine'), N_aromatics = classes.str.endswith('imidazole'), N_aromatics = classes.str.endswith('quinoline'), N_aromatics = classes.str.endswith('triazine'), N_aromatics = classes.str.endswith('indole'), N_aromatics = classes.str.endswith('piperazine'), and N_aromatics = classes.str.endswith('oxazole'). Based on qualitative element map comparison, detailed analyses were performed with the PubChem subset 1−10,000 only.

Data and Code Availability. The Python code of mol2net used in this study is available on Zenodo and GitHub.[24] Data will be made available on request.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c01271.

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142

Figures, as mentioned in the text, and details about the PubChem data set (PDF)

AUTHOR INFORMATION

Corresponding Author

Alexander Ruf − *Laboratoire de Physique des Interactions Ioniques et Moléculaires (PIIM), Université Aix-Marseille, CNRS, 13013 Marseille, France; Department of Chemistry and Pharmacy, Ludwig-Maximilians-University, 81377 Munich, Germany; Excellence Cluster ORIGINS, 85748 Garching, Germany;* orcid.org/0000-0002-0383-4292; Email: alexander.ruf@cup.uni-muenchen.de

Author

Grégoire Danger − *Laboratoire de Physique des Interactions Ioniques et Moléculaires (PIIM), Université Aix-Marseille, CNRS, 13013 Marseille, France; Aix-Marseille Université, CNRS, CNES, LAM, 13013 Marseille, France; Institut Universitaire de France (IUF), 75231 Paris, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c01271

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.R. and G.D. thank the Centre National d'Etudes Spatiales for funding (CNES Postdoctoral Fellowship 2020) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2094-390783311 (A.R.). We gratefully acknowledge Franco Moritz for discussions about Graph Theory and network analysis, Alexander Siegle for nitrogen chemistry discussions, Thomas Javelle for database searches, and Fabrice Duvernay for input about ice reactivity.

REFERENCES

(1) Lehn, J.-M. *Angew. Chem., Int. Ed.* **2013**, *52*, 2836−2850.
(2) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabási, A.-L. *Nature* **2000**, *407*, 651−654.
(3) Blokhuis, A.; Lacoste, D.; Nghe, P. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 25230−25236.
(4) Ruf, A.; Kanawati, B.; Hertkorn, N.; Yin, Q.-Z.; Moritz, F.; Harir, M.; Lucio, M.; Michalke, B.; Wimpenny, J.; Shilobreeva, S.; Bronsky, B.; Saraykin, V.; Gabelica, Z.; Gougeon, R. D.; Quirico, E.; Ralew, S.; Jakubowski, T.; Haack, H.; Gonsior, M.; Jenniskens, P.; Hinman, N. W.; Schmitt-Kopplin, P. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 2819−2824.
(5) Schmitt-Kopplin, P.; Gabelica, Z.; Gougeon, R. D.; Fekete, A.; Kanawati, B.; Harir, M.; Gebefuegi, I.; Eckel, G.; Hertkorn, N. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2763−2768.
(6) Ruf, A.; d'Hendecourt, L.; Schmitt-Kopplin, P. *Life* **2018**, *8*, 18.
(7) Schmitt-Kopplin, P.; Hemmler, D.; Moritz, F.; Gougeon, R.; Lucio, M.; Meringer, M.; Mueller, C.; Harir, M.; Hertkorn, N. *Faraday Discuss.* **2019**, *218*, 9.
(8) Schlemmer, S.; Giesen, T.; Mutschke, H. *Laboratory Astrochemistry*; Wiley, 2014.
(9) Li, D.; Liang, W.; Feng, X.; Ruan, T.; Jiang, G. *TrAC Trends in Analytical Chemistry* **2021**, *143*, 116409.
(10) Barabási, A.-L. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2013**, *371*, 20120375.
(11) Fialkowski, M.; Bishop, K. J.; Chubukov, V. A.; Campbell, C. J.; Grzybowski, B. A. *Angew. Chem.* **2005**, *117*, 7429−7435.
(12) Jolley, C. C.; Douglas, T. *Astrophysical Journal* **2010**, *722*, 1921.
(13) Wołos, A.; Roszak, R.; Zadło-Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; Spólnik, G.; Dygas, M.; Szymkuć, S.; Grzybowski, B. A. *Science* **2020**, *369*, na DOI: 10.1126/science.aaw1955.
(14) Adam, Z. R.; Fahrenbach, A. C.; Jacobson, S. M.; Kacar, B.; Zubarev, D. Y. *Sci. Rep.* **2021**, *11*, 1−10.
(15) Jolley, C.; Douglas, T. *Astrobiology* **2012**, *12*, 29−39.
(16) Andersen, J. L.; Flamm, C.; Merkle, D.; Stadler, P. F. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2017**, *375*, 20160354.
(17) Gonsior, M.; Powers, L. C.; Williams, E.; Place, A.; Chen, F.; Ruf, A.; Hertkorn, N.; Schmitt-Kopplin, P. *Water research* **2019**, *155*, 300−309.
(18) Kim, H.; Smith, H. B.; Mathis, C.; Raymond, J.; Walker, S. I. *Science advances* **2019**, *5*, No. eaau0149.
(19) Villaveces, J. M.; Koti, P.; Habermann, B. H. *Advances and applications in bioinformatics and chemistry: AABC* **2015**, *8*, 11−22.
(20) Dittrich, P.; Di Fenizio, P. S. *Bulletin of mathematical biology* **2007**, *69*, 1199−1231.
(21) Leopold, M.; Siebert, M.; Siegle, A. F.; Trapp, O. *ChemCatChem.* **2021**, *13*, 2746.
(22) Unsleber, J. P.; Reiher, M. *Annu. Rev. Phys. Chem.* **2020**, *71*, 121−142.
(23) Danger, G.; Orthous-Daunay, F.-R.; de Marcellus, P.; Modica, P.; Vuitton, V.; Duvernay, F.; Flandinet, L.; Le Sergeant d'Hendecourt, L.; Thissen, R.; Chiavassa, T. *Geochim. Cosmochim. Acta* **2013**, *118*, 184−201.
(24) *rufalexan/mol2net: v0.1.0*, 2022; https://zenodo.org/record/7025094 and https://github.com/rufalexan/mol2net.
(25) Danger, G.; Fresneau, A.; Mrad, N. A.; De Marcellus, P.; Orthous-Daunay, F.-R.; Duvernay, F.; Vuitton, V.; Le Sergeant d'Hendecourt, L.; Thissen, R.; Chiavassa, T. *Geochim. Cosmochim. Acta* **2016**, *189*, 184−196.
(26) Fresneau, A.; Mrad, N. A.; Le Sergeant d'Hendecourt, L. L.; Duvernay, F.; Flandinet, L.; Orthous-Daunay, F.-R.; Vuitton, V.; Thissen, R.; Chiavassa, T.; Danger, G. *Astrophysical Journal* **2017**, *837*, 168.
(27) Gautier, T.; Danger, G.; Mousis, O.; Duvernay, F.; Vuitton, V.; Flandinet, L.; Thissen, R.; Orthous-Daunay, F.-R.; Ruf, A.; Chiavassa, T.; Le Sergeant d'Hendecourt, L. *Earth and Planetary Science Letters* **2020**, *531*, 116011.
(28) Breitling, R.; Ritchie, S.; Goodenowe, D.; Stewart, M. L.; Barrett, M. P. *Metabolomics* **2006**, *2*, 155−164.
(29) Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. *European Journal of Mass Spectrometry* **2011**, *17*, 415−421.
(30) Moritz, F.; Hemmler, D.; Kanawati, B.; Schnitzler, J.-P.; Schmitt-Kopplin, P. *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Elsevier, 2019; pp 357−405.
(31) Watanabe, N.; Nagaoka, A.; Shiraki, T.; Kouchi, A. *Astrophysical Journal* **2004**, *616*, 638.
(32) Linnartz, H.; Ioppolo, S.; Fedoseev, G. *Int. Rev. Phys. Chem.* **2015**, *34*, 205−237.
(33) Bergner, J. B.; Öberg, K. I.; Rajappan, M. *Astrophysical Journal* **2019**, *874*, 115.
(34) Duvernay, F.; Dufauret, V.; Danger, G.; Theulé, P.; Borget, F.; Chiavassa, T. *Astronomy & Astrophysics* **2010**, *523*, A79.
(35) Noble, J. A.; Theule, P.; Borget, F.; Danger, G.; Chomat, M.; Duvernay, F.; Mispelaer, F.; Chiavassa, T. *Mon. Not. R. Astron. Soc.* **2013**, *428*, 3262−3273.
(36) Theulé, P.; Duvernay, F.; Danger, G.; Borget, F.; Bossa, J.; Vinogradoff, V.; Mispelaer, F.; Chiavassa, T. *Adv. Space Res.* **2013**, *52*, 1567−1579.
(37) Caro, G. M.; Schutte, W. *Astronomy & Astrophysics* **2003**, *412*, 121−132.
(38) Fresneau, A.; Danger, G.; Rimola, A.; Theule, P.; Duvernay, F.; Chiavassa, T. *Mon. Not. R. Astron. Soc.* **2014**, *443*, 2991−3000.
(39) de Marcellus, P.; Fresneau, A.; Brunetto, R.; Danger, G.; Duvernay, F.; Meinert, C.; Meierhenrich, U. J.; Borondics, F.;

Chiavassa, T.; Le Sergeant d'Hendecourt, L. *Mon. Not. R. Astron. Soc.* **2017**, *464*, 114−120.
(40) Ruf, A.; Bouquet, A.; Schmitt-Kopplin, P.; Boduch, P.; Mousis, O.; Danger, G. *Astronomy & Astrophysics* **2021**, *655*, A74.
(41) Van Krevelen, D. *Fuel* **1950**, *29*, 269−84.
(42) Rimola, A.; Skouteris, D.; Balucani, N.; Ceccarelli, C.; Enrique-Romero, J.; Taquet, V.; Ugliengo, P. *ACS Earth and Space Chemistry* **2018**, *2*, 720−734.
(43) Gerakines, P.; Moore, M. H.; Hudson, R. L. *Astron. Astrophys.* **2000**, *357*, 793−800.
(44) Zheng, W.; Kaiser, R. I. *Chem. Phys. Lett.* **2007**, *450*, 55−60.
(45) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213.
(46) Wakelam, V.; Smith, I. W. M.; Herbst, E.; Troe, J.; Geppert, W.; Linnartz, H.; Oberg, K.; Roueff, E.; Agundez, M.; Pernot, P.; Cuppen, H. M.; Loison, J. C.; Talbi, D. *Space science reviews* **2010**, *156*, 13−72.
(47) Wakelam, V.; Herbst, E.; Loison, J.-C.; Smith, I. W. M.; Chandrasekaran, V.; Pavone, B.; Adams, N. G.; Bacchus-Montabonel, M.-C.; Bergeat, A.; Beroff, K.; Bierbaum, V. M.; Chabot, M.; Dalgarno, A.; van Dishoeck, E. F.; Faure, A.; Geppert, W. D.; Gerlich, D.; Galli, D.; Hebrard, E.; Hersant, F.; Hickson, K. M.; Honvault, P.; Klippenstein, S. J.; Le Picard, S.; Nyman, G.; Pernot, P.; Schlemmer, S.; Selsis, F.; Sims, I. R.; Talbi, D.; Tennyson, J.; Troe, J.; Wester, R.; Wiesenfeld, L. *Astrophysical Journal Supplement Series* **2012**, *199*, 21.
(48) McElroy, D.; Walsh, C.; Markwick, A. J.; Cordiner, M. A.; Smith, K.; Millar, T. J. *Astronomy & Astrophysics* **2013**, *550*, A36.
(49) Cuppen, H. M.; Walsh, C.; Lamberts, T.; Semenov, D.; Garrod, R. T.; Penteado, E. M.; Ioppolo, S. *Space Science Reviews* **2017**, *212*, 1−58.
(50) Van Rossum, G.; Drake, F. L., Jr. *Python Tutorial*; Centrum voor Wiskunde en Informatica Amsterdam, 1995; Vol. *620*.
(51) Hagberg, A.; Swart, P.; S Chult, D. *Exploring Network Structure, Dynamics, and Function Using NetworkX*; U.S. Department of Energy, 2008.

https://doi.org/10.1021/acs.analchem.2c01271
Anal. Chem. 2022, 94, 14135−14142