Leveraging Cheminformatics to Bolster the Control of Chemical Warfare Agents and their Precursors

STEFANO COSTANZI, GREGORY D. KOBLENTZ, RICHARD T.CUPITT[1]

Abstract

International frameworks and national legislation contain lists of controlled chemicals that can be employed as chemical warfare agents or precursors for their synthesis. The development and wide adoption of a cheminformatics tool could overcome several practical problems inherent to the way in which the identification of such chemicals is currently conducted, namely: the same chemical can be identified with a multitude of synonyms; different versions of the same chemical, for instance isotopically labeled versions, have different registry numbers; some lists define whole families of related chemicals of concern in a single entry, thus complicating certification of compliance; and lists of controlled chemicals are subject to change and must be kept current. Composed of an up-to-date database of relevant lists of controlled chemicals with an associated easy-to-use software, this tool would help address these problems by converting any entered chemical name or registry number into a chemical structure, and automatically checking whether that structure matches any entry of the database. Efforts by the Pistoia Alliance for the control of regulated narcotic and psychotropic substances has led to the development of commercial software that can be used as a starting point for the development of the proposed cheminformatics tool for nonproliferation purposes. By helping frontline officers and chemical

1 Stefano Costanzi is an Associate Professor of Chemistry at American University with an educational background in both the chemical sciences and international affairs. His current scholarly work is based on two research lines: the computational modeling of the interactions between chemicals and living organisms; the analysis of gaps in current policies and practices that allow the proliferation of chemical weapons and the conceptualization of solutions and tools to narrow them. Gregory D. Koblentz is an Associate Professor and Director of the Biodefense Graduate Program at George Mason University's Schar School of Policy and Government. He is also an Associate Faculty at the Center for Security Policy Studies at George Mason and a member of the Scientist Working Group on Biological and Chemical Security at the Center for Arms Control and Nonproliferation. Dr. Richard T. Cupitt is Senior Fellow and Director of the Partnerships in Proliferation Prevention program at the Stimson Center. He has served with the U.S. State and Commerce Departments, the United Nations, and in a variety of academic posts and consultancies. He has several books and more than 20 peer reviewed articles on nonproliferation export controls, along with dozens of other security or trade-oriented publications.

industry to handle families of chemicals, this cheminformatics tool could facilitate the inclusion of families of chemicals in control lists, thus closing potential proliferation loopholes. Finally, beyond chemical warfare agents and precursors, this cheminformatics tool could be generally used to handle any list of controlled chemicals.

Keywords

Australia Group, border controls, chemical database, chemical security, chemical weapons, cheminformatics, compliance, Chemical Weapons Convention, export controls, Markush structures

1. Introduction: International Lists for the Control of Chemical Warfare Agents and Precursors

A number of international frameworks such as the nearly universal Chemical Weapons Convention (CWC), the Australia Group (AG), and the Wassenaar Arrangement, contribute to a coordinated multilateral effort to stem the proliferation of chemical weapons (CW). To serve this purpose, these frameworks contain lists of chemicals that can be employed as chemical warfare agents (i.e. the toxic chemicals on which chemical weapons are based) or precursors for their synthesis. Some of the lists exclusively comprise individual chemicals, with each entry in the list identifying a specific chemical. Other lists comprise both individual chemicals as well as families of related chemicals, defined as a central chemical scaffold bearing a number of attached variable chemical groups. Both types of entries consist of one chemical name, with no synonyms given, which identifies either an individual chemical or a family of chemicals. Moreover, for individual chemicals, the lists also report the associated Chemical Abstract Service (CAS) registry number.[2] In the paragraphs below, the authors give a brief survey of the lists contained in the above-mentioned international frameworks. Subsequently, the remainder of the article is dedicated to discussion regarding why dealing with lists composed of chemical names and registry numbers makes it impractical to identify controlled chemicals and discuss how such problems could be overcome by the development and wide adoption of a suitable cheminformatics tool.

In its Annex on Chemicals, the CWC contains a tiered system of three schedules of chemicals meant to support the verification regime imposed by the convention.[3] Importantly, beyond supporting the Convention's declaration requirements and verification regime, the CWC schedules are also incorporated by various countries in their export control lists. Each schedule is divided into a part A, which lists chemical warfare agents, and a part B, which lists chemical precursors. Schedule 1 comprises chemicals that are regarded exclusively or almost exclusively

2 The Chemical Abstract Service (CAS, <https://www.cas.org>) Registry of the American Chemical Society is a collection of the disclosed information available on chemical substances published in the scientific literature. Each substance in the registry is assigned a unique numeric identifier, the CAS Registry Number.

3 The 1993 CWC is an international treaty that poses a complete ban on chemical weapons and has currently been ratified or acceded to by 193 State Parties; Organisation for the Prohibition of Chemical Weapons (OPCW), Chemical Weapons Convention, <https://www.opcw.org/chemical-weapons-convention>.

as chemical warfare agents or precursors for their synthesis. Conversely, Schedules 2 and 3 comprise dual-use chemicals that also have legitimate, non-military commercial applications, on a smaller scale for Schedule 2, and on a larger scale for Schedule 3.[4] While some of the entries in the CWC schedules identify individual chemicals, others identify families of related chemicals. Including families of chemicals in the CWC schedules is important as it allows a single entry to cover a large group of structurally related chemicals without the need to enumerate them.[5] To avoid confusion, it should be noted that the CWC schedules do not need to be taken as an exhaustive list of chemical warfare agents. Indeed, the CWC does not limit the definition of chemical weapons to those in the schedules, but includes all "toxic chemicals and their precursors, except where intended for purposes not prohibited under this Convention, as long as the types and quantities are consistent with such purposes."[6]

The Australia Group, a group of more than 40 countries that coordinate their export control regulations on dual-use equipment and material that could be used to produce chemical and biological weapons, has compiled a list of chemicals that can be used as precursors for the synthesis of chemical warfare agents.[7] The Australia Group's "Chemical Weapons Precursors" list comprises a total of 65 chemicals. Of these, 40 are listed in part B of one of the three CWC schedules (either as individual chemicals, or by virtue of being included in one of the families of chemicals listed in the CWC schedules). Conversely, the remaining 25 chemicals are not covered by the CWC schedules. The Australia Group's Chemical Weapons Precursors list, however, does not include families of chemicals. Instead, the 65 chemicals that it covers are explicitly enumerated as individual compounds.

Similarly, the Wassenaar Arrangement, an international framework that was established in 1996 to provide greater transparency of transfers of conventional weapons and dual-use items, includes on its Munitions List 7 (ML7), *inter alia*, a range of chemical agents with military applications.[8] In particular, the chemical agents listed in ML7 include the CWC Schedule 1 chemicals, the incapacitating agent 3-quinuclidinyl benzilate (BZ), which is a CWC Schedule 2 chemical, and lists of chemical defoliants and riot control agents, neither of which is included in the CWC schedules. With the exception of the families of chemicals listed in CWC Schedule 1, all the remaining CW-related entries in the Wassenaar Arrangement ML7 are enumerated as individual chemicals. Beyond CW proliferation, two other export control arrangements, the Nuclear Suppliers Group (NSG) and the Missile Technology Control Regime (MTCR), include chemicals that are regarded as a threat for nuclear and missile proliferation in their control lists, explicitly enumerated as single chemicals and annotated with CAS registry numbers.

---

4 CWC Annex on Chemicals, <https://www.opcw.org/chemical-weapons-convention/annexes/annex-chemicals/annex-chemicals>.

5 Stefano Costanzi and Gregory D. Koblentz, "Controlling Novichoks After Salisbury: Revising the Chemical Weapons Convention Schedules," *The Nonproliferation Review* (2019), <https://www.tandfonline.com/doi/full/10.1080/10736700.2019.1662618>.

6 In CWC Article II, Paragraph 1 (a), <https://www.opcw.org/chemical-weapons-convention/articles/article-ii-definitions-and-criteria>.

7 Australia Group, <https://australiagroup.net/en/>.

8 The Wassenaar Arrangement, <https://www.wassenaar.org>.

Other lists of controlled chemicals, often based on the lists in the above-mentioned international frameworks, have been crafted by international and domestic bodies to support the implementation of the CWC mandates, export controls, and homeland security. For example, the World Customs Organization (WCO), through its Strategic Trade Control Enforcement (STCE) Implementation Guide, provides, *inter alia*, information to customs administrations on how to implement trade controls on strategic chemicals, and contains an annex (Annex V) that compiles many of the individual chemicals (but not families of chemicals) listed by the CWC, Australia Group, NSG, MTCR, and the Wassenaar Arrangement.[9] As another example, the Chemical Facility Anti-Terrorism Standards (CFATS) of the U.S. Department of Homeland Security (DHS) contains an appendix (Appendix A) that provides a list of more than 300 chemicals of interest, including, *inter alia*, chemical warfare agents and precursors for their synthesis.[10]

2. A Cheminformatics Solution to Help Navigate Lists of Chemical Warfare Agents and Precursors

The way in which frontline officers working in the areas of border security, customs, homeland security, and export controls currently conduct the identification of chemical warfare agents and precursors falling within the scope of international verification, export control, and domestic security control lists is affected by two main problems.

The first problem faced by government officials and employees of chemical companies tasked with confirming whether a given chemical is part of a control list is that such lists identify chemicals of concern through names and registry numbers rather than through chemical structures. Identifying whether a chemical is covered by a control list is not straightforward, as chemicals can be identified through a multitude of synonyms.[11] Hence, as illustrated in Section 3 (Example 1), the fact that a chemical name is not found in a list does not guarantee that the name under scrutiny is not a synonym of one of the listed names. The inclusion of CAS registry numbers in control lists addresses the synonyms problems, as it is straightforward for frontline operators to verify whether a given registry number is included in a list of

9 World Customs Organization, "Strategic Trade Control Enforcement Implementation Guide," 2019, <http://www.wcoomd.org/-/media/wco/public/global/pdf/topics/enforcement-and-compliance/tools-and-instruments/stce-implementation-guide/stce-implementation-guide_en.pdf>. Annex V, which lists more than 250 chemicals by CAS number, appears on pages 143-145.

10 U.S. Department of Homeland Security, "CFATS, Appendix A: Chemicals of Interest (COI)," <https://www.dhs.gov/cisa/appendix-chemicals-interest>.

11 In chemical nomenclature, names provide a description of the structure of a molecule, defining the molecule's central scaffold, the chemical groups attached to the central scaffold, and the positions on the central scaffold where the groups are attached. In turn, each chemical group can be formed by its own scaffold with chemical groups attached to it. The International Union of Pure and Applied Chemistry (IUPAC) has developed recommendations for a standardized chemical nomenclature. See <https://iupac.org/what-we-do/nomenclature/>. However, as the complexity of a molecule increases, so does the number of synonymous names that unambiguously describe its structure.

controlled chemicals.[12] However, different salts, tautomers, stereoisomers, and isotopically labeled variants of the same chemical have different registry numbers.[13,14,15] As a result, as also illustrated in Section 3 (Example 1), the fact that a registry number is not included in a list does not guarantee that the chemical identified by the number is not just a different form of a controlled chemical. As discussed in Section 3, controlling less frequent variants of a chemical, such as rare stereoisomers or isotopically labelled variants, is gaining increasing importance due to the changing chemical weapons proliferation landscape.

A second problem lies in the fact that several of the entries on chemical weapon verification and export control lists, such as the CWC schedules, refer to whole families of chemicals rather than individual chemicals. Covering whole chemical families is very important as closely related chemicals often interact with living organisms in a similar manner. Hence, as discussed in Section 4 of this article, the inclusion of whole families of chemicals in export control lists avoids dangerous proliferation loopholes. However, the inclusion of families of chemicals complicates the work of customs officers and export control officials, as inferring whether a given chemical under scrutiny is a member of one the listed families is not a straightforward task for non-chemists. Notably, as illustrated in Section 4 (Example 3), chemical names and CAS registry numbers are not useful for identifying whether a chemical is a member of one of the families of chemicals included on a control list.

To surmount these shortcomings, the authors propose the development and wide adoption of a customized, easy-to-use cheminformatics tool, comprised of a database capable of storing structures of individual chemicals as well as whole families of related chemicals, and an associated user-friendly software tool to query that database. The proposed cheminformatics tool would make it easier for frontline officers and chemical industry employees to automatically assess whether a given chemical under scrutiny falls within the scope of a control list by: a) seamlessly converting any entered chemical name or registry number into a chemical structure and automatically establishing the equivalence of all its salts, tautomers, stereoisomers, and isotopically labeled variants, prior to querying the database, thus addressing the current overreliance on chemical names and registry numbers; and b) automatically checking whether any entered chemical fits the definition of one of the chemical families stored in the associated database, thus addressing the difficulties that non-chemists would face when trying to manually

12 CAS registry numbers are included on lists of controlled chemicals promulgated by, among others, the CWC Schedules, the Australia Group export control list, the Wassenaar Arrangement Munitions List, the WCO Strategic Trade Control Enforcement Implementation Guide, national legislation and regulations from various countries, as well as domestic security lists such as the chemicals of interests listed by the Chemical Facility Anti-Terrorism Standards (CFATS) of the U.S. Department of Homeland Security (DHS).

13 Tautomers are molecules that have the same formula but different connectivity (structural isomers, particularly due to changes in the position of a hydrogen – or more in general a single atom or group – within the molecule) and are rapidly interconverted into each other.

14 Stereoisomers are molecules that have the same formula and the same connectivity between their constituent atoms, but differ for the spatial orientation of their atoms.

15 Isotopically labeled molecules are molecules in which one or more of the constituent atoms is substituted with a less common isotope of the same atom (isotopes are atoms with the same number of protons but different number of neutrons).

fulfill this task.

In addition to being used by government agencies responsible for export control, customs enforcement, domestic security, and verification of compliance with the CWC declaration requirements, the cheminformatics tool could also be used by chemical manufacturers and shippers, including those that produce chemicals on a small scale and offer custom synthesis services, to ensure compliance with export control regulations as well as for the verification requirements mandated by the CWC. Lastly, the availability of a suitable cheminformatics tool could support the expansion of current export control lists, by allowing a wider inclusion of whole families of chemicals.

The proposed effort to develop a cheminformatics tool to bolster the control of chemical weapons and precursors has some parallels to the Substance Compliance Service Project, a project organized by a group of pharmaceutical and technology companies, the Pistoia Alliance. The Substance Compliance Service Project is meant to provide organizations working in the pharmaceutical and life sciences domains with tools to identify controlled substances, intended as narcotic and psychotropic drugs, that are regulated under national and international legislation. As the authors will discuss in Section 6.2, this project has led to the development of two commercial cheminformatics tools that would constitute an excellent starting point for the implementation of our proposed tool.[16]

3. Challenges to Relying on Chemical Names and Registry Numbers to Identify Controlled Chemicals

As discussed in Section 2, the fact that a chemical name or a CAS registry number is not included in a list of controlled chemicals is not sufficient to conclude that the list does not cover the chemical in question, or one of its variants, including salts, tautomers, stereoisomers, and isotopically labeled variants. This point is well-illustrated by Example 1, which demonstrates how a multitude of chemical names and registry numbers can be associated with the same chemical, as chemical names have a multitude of synonyms and different variants of the same chemical are assigned different CAS registry numbers.

Notably, subjecting to strict controls all variants of the chemicals of concern, including those that are only synthesized in small volumes, is of particular importance in the current chemical weapons proliferation landscape, given that, as illustrated in Example 2, chemical weapons are used not only in large quantities as Weapons of Mass Destruction (WMDs), but also in smaller quantities for counterinsurgency or even very small quantities for assassination purposes.[17] Consequently, it would be advisable to remove exemptions from export controls for chemicals

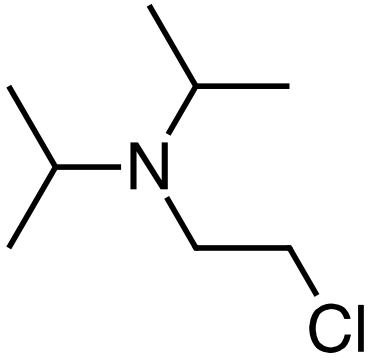
16 Daniel Taylor, Stuart G. Bowden, Reinhard Knorr, Derek R. Wilson, John Proudfoot, and Anne E. Dunlop, "The Pistoia Alliance Controlled Substance Compliance Service Project: From Start to Finish," *Drug Discovery Today, Vol*. 20, No. 2 (2015), pp. 175-180 , <https://www.sciencedirect.com/science/article/pii/S1359644614003870>.

17 Rebecca Hersman and Suzanne Claeys, "Rigid Structures, Evolving Threat: Preventing the Proliferation and Use of Chemical Weapons," *CSIS Briefs* (2019), <https://www.csis.org/analysis/rigid-structures-evolving-threat-preventing-proliferation-and-use-chemical-weapons>.

that are typically produced in small quantities or sold for low dollar values.[18] Indeed, the need to control less frequent variants of controlled chemicals was recently underscored by the Scientific Advisory Board (SAB) of the Organization for the Prohibition of Chemical Weapons (OPCW), the international organization charged with ensuring the implementation of the CWC, which recommended that "isotopically labelled compound or stereoisomer related to the parent chemical specified in the schedule should be interpreted as belonging to the same schedule."[19]

3.1. Example 1: Many names and CAS registry numbers are associated with a VX precursor

The Australia Group lists a chloroethylamine precursor for the synthesis of the nerve agent VX with the name N,N-diisopropyl-(beta)-aminoethyl chloride and the CAS registry number 96-79-7. The same chemical, however, can be identified with other synonyms, including, among others, N,N-diisopropyl-2-aminoethyl chloride and N,N-Diisopropyl-2-chloroethylamine (Figure 1).



N,N-Diisopropyl-(beta)-aminoethyl chloride
N,N-Diisopropyl-2-aminoethyl chloride
N,N-Diisopropyl-2-chloroethylamine
N-(2-Chloroethyl)-N-(1-methylethyl)-2-propanamine

Figure 1. The same chloroethylamine precursor for the synthesis of the nerve agent VX can be identified with different synonyms, four of which are shown in the figure.

18 For instance, for ECCN 1C350 items, "a license is not required for sample shipments when the cumulative total of these shipments does not exceed a 55-gallon container or 200 kg of a single chemical to any one consignee during a calendar year," U.S. Commerce Control List (CCL). Category 1 - Special Materials and Related Equipment, Chemicals, "Microorganisms," and "Toxins," <https://www.bis.doc.gov/index.php/documents/regulations-docs/2332-ccl1-10-24-18/file>.

19 OPCW, "Response to the Director-General's Request to the Scientific Advisory Board to Provide Further Advice on Scheduled Chemicals," SAB-23/WP.1, April 28, 2016, <https://www.opcw.org/sites/default/files/documents/SAB/en/sab-23-wp01_e_.pdf>; and OPCW, "Report of The Scientific Advisory Board on Developments in Science and Technology for the Fourth Special Session of the Conference of the States Parties to Review Figure the Operation of the Chemical Weapons Convention," RC-4/DG.1, April 30, 2018, <https://www.opcw.org/sites/default/files/documents/CSP/RC-4/en/rc4dg01_e_.pdf>.

The SciFinder tool of the American Chemical Society identifies the same chemical with yet another synonym, N-(2-chloroethyl)-N-(1-methylethyl)-2-propanamine (Table 1). Moreover, a SciFinder search for this chemical retrieves nine entries, each with a different CAS registry number, which are different salts, complexes, or isotopically labelled versions of the same chemical (Table 1) — each of these variants could be used to synthesize VX.[20] This example demonstrates that it is not possible to comprehensively identify a chemical only through its name or its CAS registry number.

Name	CAS number
N-(2-Chloroethyl)-*N*-(1-methylethyl)-2-propanamine	96-79-7
N-(2-Chloroethyl)-*N*-(1-methylethyl)-2-propanamine hydrochloride	4261-68-1
N-(2-Chloroethyl)-*N*-[1-(methyl-d_3)ethyl-*1*,*2*,*2*,*2*-d_4]-2-propan-*1*,*1*,*1*,*2*,*3*,*3*,*3*-d_7-amine hydrochloride	75174-04-8
N-(2-Chloroethyl-*1*,*1*,*2*,*2*-d_4)-*N*-(1-methylethyl)-2-propanamine hydrochloride	69277-06-1
N-(2-Chloroethyl-*2*-^{14}C)-*N*-(1-methylethyl)-2-propanamine hydrochloride	69277-09-4
Tetrachloro-(*T*-4)-cobaltate(2-) dihydrogen, compd. with *N*-(2-chloroethyl)-*N*-(1-methylethyl)-2-propanamine	252230-07-2
N-(2-Chloroethyl-*2*-^{14}C)-*N*-(1-methylethyl)- 2-propanamine	741623-74-5
N-(2-Chloroethyl-*1*,*1*,*2*,*2*-d_4)-*N*-(1-methylethyl)-2-propanamine	774486-54-3
N-(2-Chloroethyl)-*N*-[1-(methyl-d_3)ethyl-*1*,*2*,*2*,*2*-d_4]-2-propan-*1*,*1*,*1*,*2*,*3*,*3*,*3*-d_7-amine	774488-89-0

Table 1. CAS Registry entries for the chemical N,N-diisopropyl-(beta)-aminoethyl chloride, a chloroethylamine precursor for the synthesis of the nerve agent VX (source: SciFinder).

3.2. Example 2: Small quantities of chemicals are being employed for counterinsurgency and assassinations

For counterinsurgency purposes, a single munition filled with grams or kilograms of nerve agent could be sufficient. As reported by the French government, an unexploded grenade found in Syria was found to contain about 100 ml of sarin at an estimated purity of 60%, equivalent to about 60 grams of the nerve agent.[21] Each of the eight to twelve Volcano rockets used in the 2013 Ghouta attack in Syria carried approximately 50-60 liters of sarin, equivalent

20 In analogy with the OPCW SAB recommendation that "isotopically labelled compound or stereoisomer related to the parent chemical specified in the [CWC] schedule should be interpreted as belonging to the same schedule", it would be advisable that entries in lists of controlled chemicals, including the Australia Group list, be interpreted as comprising all variants of the same parent structure.

21 France, "National Executive Summary of Declassified Intelligence: Assessment of Syria's Chemical Warfare Program," September 2, 2013, <https://www.diplomatie.gouv.fr/en/country-files/syria/events/article/chemical-attack-in-syria-national-evaluation-presented-by-jean-marc-ayrault>.

at most to 50-60 kg of the agent in the unlikely scenario of 100% purity.[22]

Milligrams of some of the most toxic nerve agents are sufficient for assassination purposes. The assassination of Kim Jong-nam at the Kuala Lumpur airport, Malaysia in February 2018 and the attempted assassination of Sergei and Julia Skripal in Salisbury, United Kingdom in March 2018 exemplify the security threat posed by small quantities of chemical warfare agents.

Kim Jong-nam was killed with a binary form of the nerve agent VX.[23] Two VX precursors were applied in sequence by two different individuals on the face of the victim where the precursors reacted together giving rise to the deadly nerve agent. Given the toxicity of the nerve agent VX, less than 10 mg are sufficient to kill a 70 kg person through dermal contact.

The attempted assassination of Sergei and Julia Skripal, which also led to the subsequent poisoning of two police officers and two persons in the nearby town of Amesbury, was carried out with a Novichok agent—a chemical belonging to a family of nerve agents developed by the Soviet Union under the Foliant program.[24] According to the OPCW, the amount of Novichok agent associated with the Salisbury incident "should probably be characterised in milligrams."[25] This is consistent with an estimated lethal dose comparable to that of the nerve agent VR, for which the dose that would kill 50% of the individuals exposed through dermal contact (LD_{50}) is around 0.1 mg/kg or lower.[26] The *faux* perfume bottle recovered from the home of the victims in Amesbury contained at most 5.5 ml of Novichok agent, which is probably equivalent to five to six grams.[27]

---

22 "Attacks on Ghouta: Analysis of Alleged Use of Chemical Weapons in Syria," Human Rights Watch, September 10, 2013, <https://www.hrw.org/report/2013/09/10/attacks-ghouta/analysis-alleged-use-chemical-weapons-syria>.

23 Tomomasa Nakagawa and Anthony T. Tu, "Murders with VX: Aum Shinrikyo in Japan and the assassination of Kim Jong-Nam in Malaysia," *Forensic Toxicology*, Vol. 36, No. 2 (2018), pp. 542-544, <https://link.springer.com/article/10.1007/s11419-018-0426-9>.

24 Costanzi and Koblentz, "Controlling Novichoks After Salisbury"; Lev A. Fedorov, Chemical Weapons in Russia: History, Ecology, Politics (Moscow, Russia: Center of Ecology Policy of Russia, 27 July 1994); and Marcin Kloske and Zygfryd Witkiewicz, "Novichoks – The A Group of Organophosphorus Chemical Warfare Agents," *Chemosphere*, Vol. 221 (2019), pp. 672–682 < https://www.sciencedirect.com/science/article/pii/S0045653519300542>.

25 OPCW, "OPCW Spokesperson's Statement on Amount of Nerve Agent Used in Salisbury," May 4, 2018, <https://www.opcw.org/media-centre/news/2018/05/opcw-spokespersons-statement-amount-nerve-agent-used-salisbury>.

26 According to Mirzayanov, the most potent Novichok agents have a toxicity comparable to that of the nerve agent VR or higher. See Vil S. Mirzayanov, *State Secrets: An Insider's Chronicle of the Russian Chemical Weapons Program*, (Denver: Outskirts Press, 2008), pp 143-145. It should be noted, however, that the results of recent computational studies are not in agreement with Mirzayanov's claim. In particular, see Lars Carlsen, "After Salisbury Nerve Agents Revisited," *Molecular informatics*, Vol. 38 (2019), p.1800106, <https://onlinelibrary.wiley.com/doi/abs/10.1002/minf.201800106>; and Hanusha Bhakhoa, Lydia Rhyman, and Ponnadurai Ramasami, "Theoretical Study of the Molecular Aspect of the Suspected Novichok Agent A234 of the Skripal Poisoning," *Royal Society Open Science*, Vol. 6, No. 2 (2019), p. 181831, <https://royalsocietypublishing.org/doi/full/10.1098/rsos.181831>.

27 Martin Robinson, "Novichok James Bond-Style," *The Daily Mail*, September 6, 2018, <https://www.dailymail.co.uk/news/article-6138471/Novichok-James-Bond-style-Perfume-bottle-used-Salisbury-assassins-scientists.html>.

This example illustrates how chemicals that are not typically produced in large quantities pose a security threat, given that chemical warfare agents can be employed for counterinsurgency and assassination purposes.

4. Importance of Including Whole Families of Chemicals in Control Lists

Structurally related chemicals often have a similar biological activity, in other words, they interact similarly with living organisms. For this reason, claims in pharmaceutical patents very commonly cover large families of structurally related chemicals.[28] Such practice allows pharmaceutical companies to protect not only the specific chemical that is being evaluated as a drug candidate but also its structural analogs, thus preventing the development of similar drugs by competitors. Similarly, the CWC Schedules list several families of related chemicals along with individual chemicals of concern.[29] This approach allows the CWC Schedules to cover a large chemical space. However, other lists relevant to chemical weapon proliferation, for example the list of precursor chemicals compiled by the Australia Group, are composed exclusively of individual chemicals. As a result, even if such lists are long and detailed, they are not comprehensive and may not include many of the precursors that can be employed in the manufacturing of chemical weapons. Moreover, lists that encompass individual chemicals only cover chemicals of current concern, leaving analogues that could be synthesized in the future unregulated.

Enumerating individual chemicals makes it easier for companies and government officials to verify whether a compound is subject to export controls. However, this approach also makes it easier for proliferators to find loopholes. In particular, as illustrated by Example 3, this approach may allow proliferators to work around the constraints posed by export control lists by seeking analogs of the controlled compounds that are not covered by the lists. Moreover, as illustrated by Example 4, this approach makes it impractical to cover large families of chemical warfare agents and their precursors, despite the responsibilities of states under the CWC. Conversely, listing whole families of chemicals of concern rather than individual chemicals grants a broad coverage of related chemicals that are likely to have similar toxicological properties since they are based on the same structural scaffold.

Despite the advantages that it offers, listing families of chemicals also poses some practical problems for those who are tasked with ensuring compliance with verification, export control, customs, and domestic security regulations. Specifically, this requires the ability to infer whether a given chemical of interest falls within the definition of one of the listed families, which is not

---

28 David A. Cosgrove, "Markush Structures and Chemical Patents," in Nathan Brown, ed., *Scaffold Hopping in Medicinal Chemistry* (Weinheim, Germany: Wiley, 2013), pp. 15–38, Michael F. Lynch, "Generic Chemical Structures in Patents (Markush Structures): The Research Project at the University of Sheffield," *World Patent Information*, Vol. 8, No. 2 (1986), pp. 85-91; and Szabolcs Csepregi, Nóra Máté, Róbert Wágner, Tamás Csizmazia, Szilárd Dóránt, Erika Bíró, Tim Dudgeon, Ali Baharev, and Ferenc Csizmadia, "Representation of Markush Structures: From Molecules Toward Patents," *Journal of Cheminformatics*, Vol. 3, No. S-1 (2011), p. 7.

29 For an example of an entry of the CWC schedules that identifies a family of related chemicals, see Figure 3.

a straightforward task for non-chemists. This problem is well illustrated by Example 5.

4.1. Example 3: Listing families of related chemicals allows a broad coverage of chemicals of proliferation concern

As discussed in Example 1, the Australia Group lists a key chloroethylamine precursor for the synthesis of the nerve agent VX, namely N,N-diisopropyl-(beta)-aminoethyl chloride (Figure 1). However, the Australia Group does not list N,N-diethyl-(beta)-aminoethyl chloride, the analogous precursor for the synthesis of VM and VR, both of which are very lethal nerve agents (Figure 2). Percutaneous exposure studies conducted with different animal models indicate that VM's LD_{50} (the dose that causes lethality in 50% of the exposed population) is comparable to that of VX (although in some species it is up to three times higher).[30] VR's LD_{50} is estimated to be up to 10 times lower than that of VX.[31] When Syria acceded to the CWC in 2013, it declared the possession, among other chemicals, of 25 tons of the chloroethylamine precursor for the synthesis of VM or VR, possibly acquired or produced as a consequence of the fact that, in 2002, it encountered difficulties in procuring precursors for the preparation of VX.[32,33]

Unlike the Australia Group, which covers the mentioned VX precursor but not the precursor for VM or VR, the CWC covers both precursors as part of entry 2B10 of Schedule 2, which features a general family of 2-chloroethylamines (Figure 2).

30 Department of the Army, *Chemical Agent Data Sheets,* Vol. 1, Edgewood Arsenal Special Report EO-SR-74001 (Aberdeen Proving Ground, MD: Edgewood Arsenal 1974), <http://www.dtic.mil/dtic/tr/fulltext/u2/b028222.pdf>.

31 Stefano Costanzi, John-Hanson Machado, and Moriah Mitchell, "Nerve Agents: What They Are, How They Work, How to Counter Them," *ACS Chemical Neuroscience*, Vol. 9, No. 5 (2018), pp. 873–85, <https://pubs.acs.org/doi/10.1021/acschemneuro.8b00148>.

32 Helen Rice, Christopher H. Dalton, Matthew E. Price, Stuart J. Graham, A. Christopher Green, John Jenner, Helen J. Groombridge, and Christopher M. Timperley, "Toxicity and Medical Countermeasure Studies on the Organophosphorus Nerve Agents VM and VX," *Proceedings of the Royal Society A*, Vol. 471 (2015), p. 20140891, <https://royalsocietypublishing.org/doi/full/10.1098/rspa.2014.0891>.

33 Rene Backmann, "How the Assad Regime Built a Chemical Arsenal with the Aid of Germany and Other Countries," *Mediapart*, June 2, 2017, <https://www.mediapart.fr/en/journal/international/020617/how-assad-regime-built-chemical-arsenal-aid-germany-and-other-countries>.

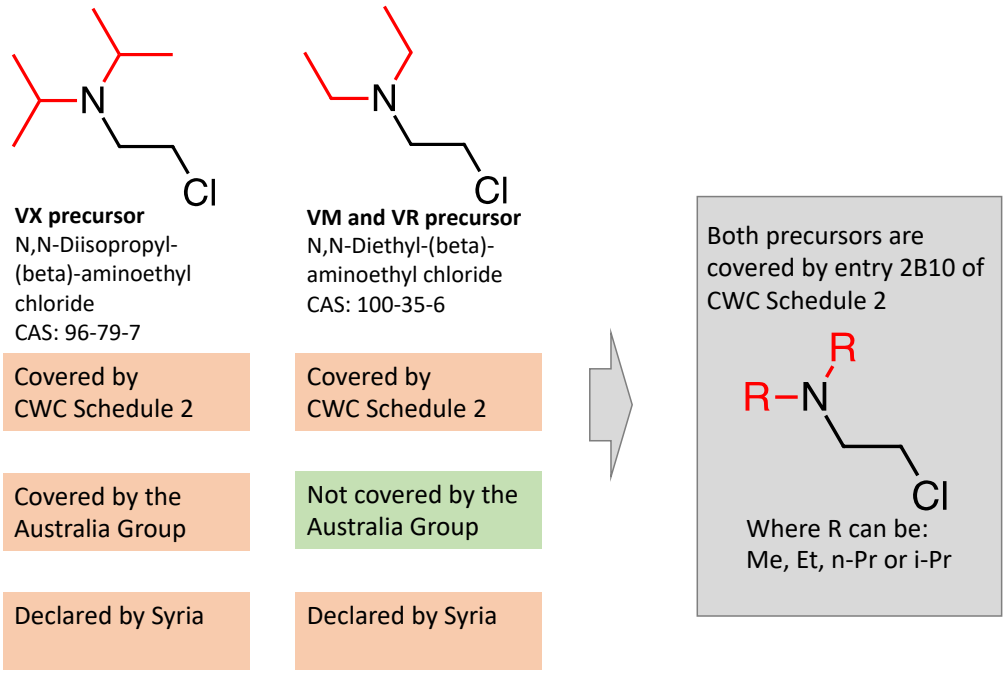


Figure 2. N,N-Diisopropyl-(beta)-aminoethyl chloride, a precursor for the synthesis of the nerve agent VX is covered by the Australia Group export control list. However, analogous precursors for the synthesis of the nerve agent VM, N,N-Diethyl-(beta)- aminoethyl chloride is not. Both chemicals are covered by Schedule 2 of the Chemical Weapons Convention (CWC) as members of a family of 2-chloroethylamine compounds identified by entry 2B10. The chemical structure on the right is a "Markush" structure, a common way of representing families of chemicals by drawing a shared scaffold with variable chemical groups (R groups) attached. The variable substituents in the three precursors and the corresponding R groups in the Markush structure are marked in red.

Hence, export control legislation that incorporates the CWC schedules, such as that in force in the United States and the European Union, does cover the chloroethylamine precursor for the synthesis of VM or VR.[34] As this example illustrates, adopting a family-based approach to chemical control lists allows for a much wider coverage than listing individual chemicals, thus covering dangerous loopholes that would remain open by listing exclusively individual chemicals.

4.2. Example 4: Enumerating all chemicals comprised in a family of chemicals requires chemistry skills

The VX precursor discussed in Examples 1 and 3 is not enumerated explicitly in the CWC schedules. Conversely, as illustrated in Figure 3, entry 2B10 of Schedule 2 identifies the family

34 For U.S. legislation, see U.S. Commerce Control List (CCL). Category 1 - Special Materials and Related Equipment, Chemicals, "Microorganisms," and "Toxins," October 24, 2018, <https://www.bis.doc.gov/index.php/documents/regulations-docs/2332-ccl1-10-24-18/file>. For EU regulations, see EU Commission Delegated Regulation EC 1382/2014, October 22, 2014, <https://trade.ec.europa.eu/doclib/docs/2015/january/tradoc_152996.pdf>.

of chemicals depicted on the left. Whenever an entry of a control list refers to a whole family of chemicals, as in the case of entry 2B10 of the CWC schedules, individual members are not explicitly enumerated, perhaps with the exception of some examples. Conversely, the entry defines the family as a central scaffold (N,N-Dialkyl aminoethyl-2-chloride) with variable substituents (methyl, ethyl, normal propyl or isopropyl) attached to it (Figure 3).

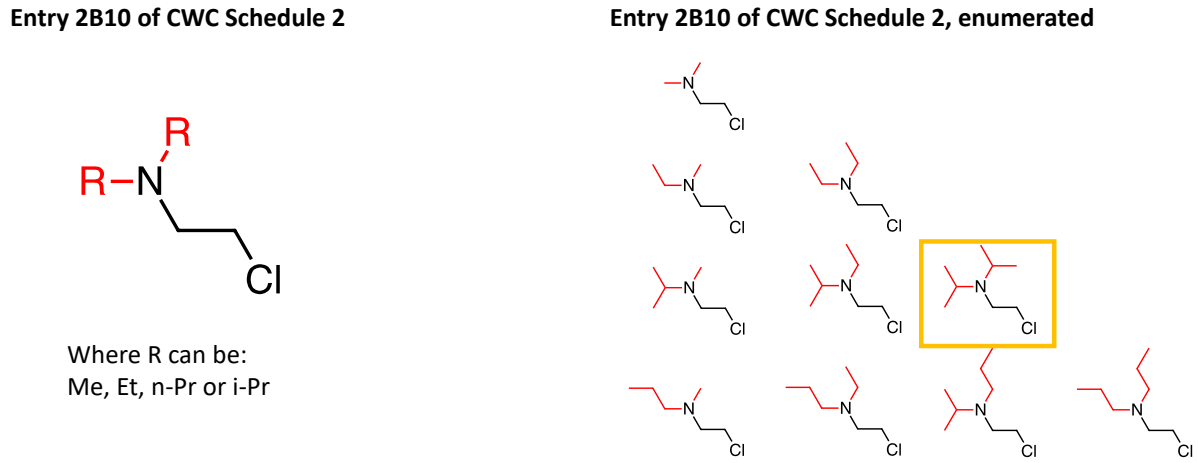


Figure 3. Entry 2B10 of CWC Schedule 2 reads as follows "N,N-Dialkyl (Me, Et, n-Pr or i-Pr) aminoethyl-2-chlorides and corresponding protonated salts." In the left panel, the entry is represented as a Markush structure, while the right panel enumerates all of the chemicals that fall within the scope of the entry's definition. The variable substituents and the corresponding R groups in the Markush structure are marked in red. The only member of entry 2B10 listed by the Australia Group is enclosed in a yellow box.

Any chemical that fits the definition is a member of the family. In the case of entry 2B10 of the CWC schedules, the family of chemicals defined by this entry comprises ten members, one of which is the VX precursor discussed in Examples 1 and 3 (encased in a yellow rectangle). The ten chemicals that comprise the family are not explicitly enumerated. Thus, their chemical names or CAS registry numbers are not included in the entry definition, and, consequently, cannot be used to identify the ten chemicals as members of the family. As a result, the identification of the ten chemicals as members of entry 2B10 of the CWC schedules requires the ability to infer that their chemical structure fits the definition of the chemical family provided in the entry. As this example demonstrates, it would not be straightforward for a non-chemist to identify all of the individual chemicals that are members of the same family.

4.3. Example 5: New entries added to the CWC Schedules define large families of chemicals

In the aftermath of the Skripal incident with Novichok described in Example 2, the Director General of the OPCW called on the SAB to conduct a literature review on "next generation toxic chemicals capable of acting as nerve agents" and assess whether such chemicals and their

precursors fall under the CWC's scheduled chemicals.[35] As the SAB has long noted, the CWC's list of scheduled chemicals were finalized twenty-five years ago and should be reviewed in light of subsequent developments.[36] In November 2019, the 24th Conference of State Parties of the CWC adopted two proposals to add toxic chemicals that were developed or researched as chemical warfare agents during the Cold War, namely Novichok agents, including the agent used in Salisbury, and certain carbamates, to CWC Schedule 1.[37,38] The addition of these toxic chemicals to CWC Schedule 1 will enter into force 180 days after the notification of the decision by the OPCW Director-General to all State Parties and the United Nations Secretary-General.[39]

The first proposal, jointly submitted by the United States, Canada, and the Netherlands, is composed of two elements, each of which covers one large family of related Novichok agents (Figure 4).

Figure 4. General structures of Novichok agents covered by the joint proposal, shown as Markush structures, with the variable R groups marked in red.

The other proposal, submitted by the Russian Federation, is composed of four elements. The first two elements of the Russian proposal cover a subset of the two Novichok families listed by the joint proposal. The third element covers an additional Novichok agent, listed as an individual chemical. The fourth element covers two families of carbamates, which are chemically unrelated to the Novichok agents, and were researched as chemical warfare agents

35 OPCW, "Note by the Director-General: Request for Information From States Parties on New Types of Nerve Agents," S/1621/2018, May 2, 2018, <https://www.opcw.org/fileadmin/OPCW/S_series/2018/en/s-1621-2018_e_.pdf>.

36 OPCW, "Report of The Scientific Advisory Board on Developments in Science and Technology for the Fourth Special Session of the Conference of the States Parties to Review the Operation of the Chemical Weapons Convention," RC-4/DG.1, April 30, 2018, <http://www.opcw.org/sites/default/files/documents/CSP/RC-4/en/rc4dg01_e_.pdf>.

37 OPCW, "Technical Change to Schedule 1(A) of the Annex on Chemicals to the Chemical Weapons Convention," C-24/DEC.4, November 27, 2019, <https://www.opcw.org/sites/default/files/documents/2019/11/c24dec04%28e%29.pdf>.

38 OPCW, "Changes to Schedule 1 of the Annex on Chemicals to the Chemical Weapons Convention," C-24/DEC.5, November 27, 2019, <https://www.opcw.org/sites/default/files/documents/2019/11/c24dec05%28e%29.pdf>.

39 OPCW, "Conference of the States Parties Adopts Decisions to Amend Chemical Weapons Convention Annex," November 27, 2019, <https://www.opcw.org/media-centre/news/2019/11/conference-states-parties-adopts-decisions-amend-chemical-weapons>.

by the United States and other countries during the Cold War.[40]

Although the joint proposal and the Russian proposal differ in multiple, important ways, they both contain entries that define large families of chemicals.[41] Adopting a family-based approach is crucial given the large number of structurally-related chemicals that were reportedly synthesized and researched (and, in the case of the Soviet Foliant program, developed into chemical weapons) by the chemical weapons programs of various countries during the Cold War.

Following the recent decision to include of the above-mentioned chemicals in CWC Schedule 1, it is advisable that a thorough review of the precursors for their synthesis be conducted with the goal of including those that constitute a high proliferation threat in arms control verification and export control lists. This appears particularly important in light of the fact that, as revealed by Vil Mirzayanov, a chemist formerly involved in the Soviet chemical weapons program, one of the goals of the Foliant program was to devise chemical warfare agents based on precursors that were projected not to be included in the schedules of the upcoming CWC.[42] Given the large scope of the families of chemicals that are about to be added to CWC Schedule 1, it seems inevitable that a family-based approach will have to be adopted to ensure effective coverage of such precursors.

This example illustrates how it would be impossible for verification and export control lists to effectively cover Novichok-related toxic chemicals and precursors without listing entire families of related chemicals.

5. Keeping Verification and Export Control Lists Current is Paramount

In a rapidly changing chemical proliferation landscape, it is vitally important for government authorities and chemical companies to have easy access to current and complete lists of all chemicals of concern. Failures to update such lists, or to ensure the wide promulgation of such lists, can create loopholes in the export control system and fuel the proliferation of chemical weapons. A study of Japanese enforcement of import-export regulations found that 27 percent of violations had been committed because the culprit was not aware of the relevant laws and regulations and another 13 percent because they had improperly interpreted the rules or had made operational errors in their application.[43] This problem is well described by Example 6.

---

40 OPCW, "Recommendation for a Change to Schedule 1 of the Annex on Chemicals to the Chemical Weapons Convention," EC-M-62/DEC.1, January 14, 2019, <www.opcw.org/sites/default/files/documents/2019/01/ecm62dec01%2B%28e%29.pdf>.

41 Costanzi and Koblentz, "Controlling Novichoks After Salisbury,"

42 Vil S. Mirzayanov, "State Secrets: An Insider's Chronicle of the Russian Chemical Weapons Program," (Denver: Outskirts Press, 2008), pp 148.

43 Philippe Mauger and Raymond A. Zilinskas, "Current Challenges to Export Controls Related to Chemical Warfare Interdiction," in Michael Crowley, Malcolm Dando and Lijun Shang, eds., *Preventing Chemical Weapons: Arms Control and Disarmament as the Sciences Converge* (London: The Royal Society of Chemistry, 2018), <https://pubs.rsc.org/en/content/ebook/978-1-78262-649-7>, p. 425.

5.1. Example 6: Export of chemicals to Syria

There have been several cases of European companies shipping controlled chemicals to Syria without the proper export licenses. These cases demonstrate the criminal, reputational, and financial risks that companies face if they run afoul of export control regulations in this sensitive area, even if the alleged violation is not intentional or is not prosecuted.

In June 2012, the Australia Group created a Syria-specific control list, beyond those chemicals already included on the common control list, to be implemented on a national basis by all of its members. In June 2012, the European Union promulgated new export control regulations, subsequently updated with various amendments, which, beyond posing additional restrictions on the chemicals already present in the general EU export control lists, requires prior authorization for the export to Syria of a list of chemicals that were not part of the list of items already controlled under the EU's dual-use export regulations. These are chemicals that, although widely used in the chemical industry for non-military purposes, are of concern for their possible role as precursors for the synthesis of chemical warfare agents. Among others, the EU included dichloromethane and diethylamine on its Syria-specific control list in 2012 and added isopropanol of 95% purity or higher, acetone, and methanol in 2013.[44] Isopropanol is one of the key precursors that Syria has used to produce the nerve agent sarin. In 2013, Syria declared to the OPCW that it possessed 133 tons of the chemical and it had to destroy this stockpile under OPCW supervision.[45]

In May 2018, Belgium Customs initiated a lawsuit against three Flemish companies for the shipment of controlled chemicals to Syria without receiving proper export licenses. AAE Chemie Trading, the wholesaler that supplied the chemicals, Anex Customs which provided administrative services to AAE Chemie, and Danmar Logistics which transported the chemicals, and two managing directors of these companies, were accused of exporting 168 tons of highly pure isopropanol, 219 tons of acetone, 77 tons of methanol, and 21 tons of dichloromethane to Syria between May 2014 and December 2016.[46]

During the trial, the companies argued that they were not aware of this licensing obligation. They claimed that at the time, the Customs tool TARWEB did not mention that an export license was required, and that was the reason why AAE Chemie had not applied for the required

---

44 Council of the European Union, Council Regulation (EU) No 509/2012 of 15 June 2012 amending Regulation (EU) No 36/2012 concerning restrictive measures in view of the situation in Syria, June 15, 2012, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:156:0010:0037:EN:PDF>; Council Regulation (EU) No 697/2013 of 22 July 2013 amending Regulation (EU) No 36/2012 concerning restrictive measures in view of the situation in Syria, Official Journal of the European Union, L198, July 23, 2013, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:198:0028:0034:EN:PDF>; and Ian Anthony, "Exports of Dual-Use Chemicals to Syria: An Assessment of European Union Export Controls," Non-Proliferation Paper No. 35, EU Non-Proliferation Consortium, January 2014, <https://www.sipri.org/sites/default/files/EUNPC_no-35.pdf>.

45 Dan Kaszeta, "Isopropyl Alcohol and Sarin: One Is Needed For the Other," *Bellingcat*, January 22, 2019, <https://www.bellingcat.com/news/2019/01/22/isopropyl-alcohol-and-sarin-one-is-needed-for-the-other/>.

46 "Antwerp Court Convicts Three Flemish Firms for Shipping 168 tonnes of Isopropanol to Syria," *Syria Archives*, February 7, 2019, <https://syrianarchive.org/en/investigations/belgium-isopropanol/sentencing>.

license. According to a lawyer for the defendants, the licensing requirement for shipments of isopropanol to Syria was not added to the tool until April 19, 2018, one day after a Belgian newspaper published an article about the shipments.[47] An audit conducted by Belgian Customs of their handling of this case also revealed numerous weaknesses in their internal procedures and capabilities. According to the audit, "most control weaknesses are in the first line." Customs officers reported that they were under time pressure, that it was difficult for them to consult European legislation using the internal Customs system, they received insufficient training, and Internet access was problematic.[48]

On February 7, 2019, the Antwerp Criminal court convicted the Flemish firms and company officials of violating Belgian export control regulations. The court imposed conditional fines on the three companies totaling EUR €921,443, fined the two managers a total of EUR €846,443 EUR, and imposed prison sentences of 12 months for one manager and four months for the other. By this time, AAE Chemie had declared bankruptcy, reportedly as a direct consequence of this case.[49]

In June 2019, a coalition of non-governmental groups asked prosecutors in Belgium and Germany to investigate whether the Belgium chemical producer and distributor BASF Antwerpen NV, the German chemical manufacturer Sasol Germany GmbH, and the German chemical wholesaler Brenntag AG and its Swiss subsidiary, Brenntag Schweizerhall, exported isopropanol and diethylamine to Syria in 2014 without the proper authorizations.[50] Both chemicals, beyond their legitimate uses in the chemical and pharmaceutical industries, can be employed as precursors for the synthesis of nerve agents. The shares of Brenntag declined 5.8 percent as a result of the alleged export control violations.[51] The companies claimed that the customer was a pharmaceutical company in Syria that produces medicine under license from the Swiss firm Novartis.[52] In August 2019, German prosecutors decided there were insufficient grounds to prosecute Brenntag for export control violations.[53]

This example illustrates the importance of ensuring that informatics systems and databases of

---

47 Ibid.

48 Ibid.

49 Ibid.

50 "German and Belgian Prosecutors Urged to Investigate Chemical Shipments to Syria," *Open Society Justice Initiative*, June 3, 2019, <https://www.justiceinitiative.org/newsroom/german-and-belgian-prosecutors-urged-to-investigate-chemical-shipments-to-syria>; and Rebecca Staudenmaier, "German Firms Sent Weapons-Grade Chemicals to Syria Despite Sanctions," *Deutsche Welle*, June 25, 2019, <https://p.dw.com/p/3L5VP>.

51 "Shares in Germany's Brenntag Drop on Dual-Use Chemicals Sale to Syria," *Reuters*, June 26, 2019, <https://www.reuters.com/article/us-brenntag-syria/shares-in-germanys-brenntag-drop-on-dual-use-chemicals-sale-to-syria-idUSKCN1TR0UZ>.

52 Rebecca Staudenmaier, "German Firms Sent Weapons-Grade Chemicals to Syria Despite Sanctions," *Deutsche Welle*, June 25, 2019, <https://p.dw.com/p/3L5VP>.

53 Anneli Palmen and Ludwig Burger, "German prosecutors Won't Probe Brenntag Over Chemicals Sale to Syria," *Reuters,* August 13, 2019, <https://www.reuters.com/article/us-brenntag-syria/german-prosecutors-wont-probe-brenntag-over-chemicals-sale-to-syria-idUSKCN1V311W>.

controlled chemicals are promptly updated to reflect changes in legislation. Indeed, it could be argued that TARWEB would have been able to identify the chemicals listed on the Syria-specific legislation if only it had been updated more quickly.

6. A Cheminformatics Tool to Bolster the Control of Chemical Warfare Agents and their Precursors

To solve the various issues described above, the authors recommend the development and wide adoption of an easy-to-use cheminformatics tool that would allow government agencies responsible for customs enforcement, export controls, domestic security, and arms control verification, as well as employees of companies that produce and ship chemicals, to automatically assess whether a given chemical under scrutiny falls within the scope of a control list. In particular, the proposed tool would allow frontline operators to enter a chemical of concern through a variety of means and would subsequently:

i. Automatically convert any entered chemical name or registry number to a chemical structure and automatically establish the equivalence of all of its salts, tautomers, stereoisomers, and isotopically labeled variants, prior to querying the associated database, thus addressing the ineffectiveness of the current practice of verifying compliance with export control lists solely on the basis of chemical names and registry numbers;

ii. Automatically check whether any entered chemical fits the definition of one of the chemical families stored in the associated database, thus addressing the difficulties that non-chemists would face when attempting to manually fulfill this task.

The proposed cheminformatics tool should consist of (1) an annotated database of chemical structures and (2) an associated, easy-to-use software interface intended to query the database to automatically assess whether a given chemical is included in a control list. To achieve scalability and facilitate the installation and maintenance of the system, the back end of the cheminformatics tool should run on a centrally managed remote server. This approach would ensure that the tool had the necessary scalability, reliability, and accessibility. To facilitate usage, the front end should be available via the web as well through dedicated apps for a variety of operating systems, including those that run on mobile devices. As illustrated by Example 6, it will be important to ensure that all the lists of controlled chemicals featured in the database are promptly updated to reflect the most current legislation, thus addressing the difficulties of controlling chemicals in a rapidly changing proliferation landscape. A remotely administered, central database would take the burden of updating the lists of controlled chemicals off of the user. The next section provides a detailed description of the desirable features of the database and the associated interface, while the following section addresses existing technology on which the cheminformatics tools can be based.

6.1. An annotated database of chemicals that can handle Markush structures and an associated software with an easy-use interface

The recommended database of chemical structures should comprise all the chemicals listed by international frameworks for the control of chemical warfare agents, precursors for their

synthesis, and strategic chemicals in general (see Section 1), as well as all relevant national export control lists and lists of chemicals of domestic security concern identified by individual countries. Importantly, it should be capable of handling both individual chemicals as well as families of chemicals. Specifically, the tool should be able to handle families of chemicals through the use of "Markush" structures, in other words, structures that identify families of related chemicals by defining a central scaffold with one or more variable substituents attached to it, also known as R groups (for an example of a Markush structure, see Figure 1).[54] The database should also be capable of handling exceptions for compounds that, although falling within the scope of a listed chemical family, are not covered by these control lists. This approach, which is already adopted by the CWC Schedules, is essential to avoid practical problems deriving from the inclusion in the control lists of extremely common chemicals as well as industrial chemicals produced in large quantities. Lastly, the database should be divided into sections, each pertaining to a specific jurisdiction, in order to allow users to query only specific sections rather than the entire database, if desired.

Cheminformatics software tools can be used to query databases of chemical structures. A cheminformatics software intended to bolster the control of chemical weapons and their precursors should leverage a cheminformatics back end to query the above-mentioned annotated database of controlled chemicals and automatically assess whether a given chemical, entered through an easy-to-use front end, is subject to export control, domestic security, and/or arms control verification requirements (Figure 5), tailored to the regulatory regimes in different jurisdictions.

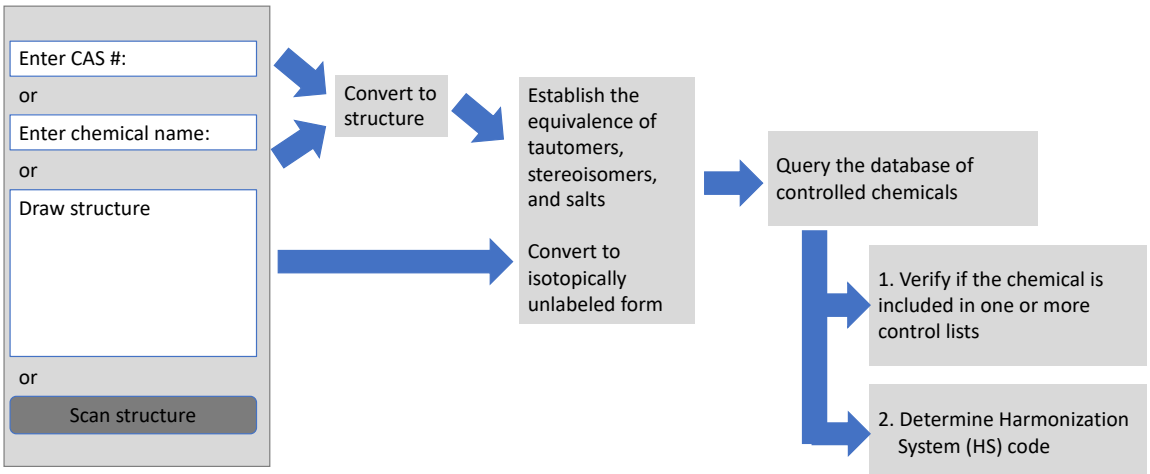


Figure 5. Simplified, schematic representation of the cheminformatics tool. In addition to checking whether a chemical is included in one or more control lists, the tool would also determine the Harmonization System (HS) code of the query chemical.

54 David A. Cosgrove, "Markush Structures and Chemical Patents," in Nathan Brown, ed., *Scaffold Hopping in Medicinal Chemistry* (Weinheim, Germany: Wiley, 2013), pp. 15–38, Michael F. Lynch, "Generic Chemical Structures in Patents (Markush Structures): The Research Project at the University of Sheffield," *World Patent Information*, Vol. 8, No. 2 (1986), pp. 85-91; and Szabolcs Csepregi, Nóra Máté, Róbert Wágner, Tamás Csizmazia, Szilárd Dóránt, Erika Bíró, Tim Dudgeon, Ali Baharev, and Ferenc Csizmadia, "Representation of Markush Structures: From Molecules Toward Patents," *Journal of Cheminformatics*, Vol. 3, No. S-1 (2011), p. 7.

To further assist agencies responsible for export control and customs enforcement, as well as chemical manufacturers and traders, the software should also automatically determine the Harmonized System (HS) code associated with the chemicals listed in these control lists.[55]

Very importantly, a cheminformatics software intended to bolster the control of chemical weapons and their precursors should be endowed with an easy-to-use interface that can be easily operated by users without extensive training in chemistry. For example, to facilitate the work of frontline officers, it will be useful to include in the software the ability to enter chemicals of interest by directly scanning a printed figure of the structure. This would be particularly beneficial for chemicals that do not have an associated CAS registry number (e.g. a member of a controlled family of chemical without a CAS registry number). As the authors have mentioned, identifying chemicals through names is not straightforward. Drawing structures would be more effective but could prove a difficult task for non-chemists. Scanning printed structures would provide a simple way to input chemical structures without the need to draw them.

6.2. Existing technology for the development of the cheminformatics tool

As discussed in Section 4, claims in pharmaceutical patents often cover large families of structurally related chemicals to allow companies to protect a wide portion of the chemical space around a given drug, thus preventing the development of similar drugs by competitors.[56] As a result, Markush structures are very common in pharmaceutical patents and databases that can handle Markush structures have been developed for the pharmaceutical sector. In particular, the American Chemical Society produces and maintains MARPAT, a database of more than 1.1 million searchable Markush structures found in chemical patents.[57] Similarly, the software companies ChemAxon and Biovia offer JChem Base and Biovia Direct, respectively, which are both database systems capable of handling Markush structures.[58] In parallel with the development of databases that can handle Markush structures as well as individual chemicals, software tools have been developed that allow a user to enter a specific chemical and use it to query said databases. These tools include the SciFinder and STN tools of the ACS, which can query the MARPAT database, and the ChemAxon and Biovia tools, which can query the

55 The HS system was established in 1983 to provide an international standard to categorize goods for tariff purposes. HS Code Section 06 (Chapters 28-38) pertains to "Products of the Chemicals and Allied Industries." See <http://www.wcoomd.org/-/media/wco/public/global/pdf/topics/nomenclature/activities-and-programmes/30-years-hs/hs-compendium.pdf>.

56 David A. Cosgrove, "Markush Structures and Chemical Patents," in Nathan Brown, ed., *Scaffold Hopping in Medicinal Chemistry* (Weinheim, Germany: Wiley, 2013), pp. 15–38, Michael F. Lynch, "Generic Chemical Structures in Patents (Markush Structures): The Research Project at the University of Sheffield," *World Patent Information*, Vol. 8, No. 2 (1986), pp. 85-91; and Szabolcs Csepregi, Nóra Máté, Róbert Wágner, Tamás Csizmazia, Szilárd Dóránt, Erika Bíró, Tim Dudgeon, Ali Baharev, and Ferenc Csizmadia, "Representation of Markush Structures: From Molecules Toward Patents," *Journal of Cheminformatics*, Vol. 3, No. S-1 (2011), p. 7.

57 American Chemical Society, "Markush–MARPAT," <https://www.cas.org/support/documentation/markush>.

58 ChemAxon, "JChem Chemical Database Concepts," <https://docs.chemaxon.com/display/docs/JChem+Chemical+Database+Concepts>; and Biovia, "Biovia Direct,". <https://www.3dsbiovia.com/products/collaborative-science/biovia-direct/>.

JChem Base and Biovia Direct databases.[59,60,61] These tools allow users to enter a chemical of interest in a variety of ways, including by typing chemical names or registry numbers or drawing chemical structures, and subsequently use the entered chemical to query the database. These software tools have very robust engines to automatically convert any entered chemical name or CAS registry number into a chemical structure prior to launching the query and to automatically associate all salts, tautomers, stereoisomers, and isotopically labeled variants of the same chemical species. Such an approach avoids the problems associated with the non-uniqueness of chemical names and the fact that different variants of the same chemicals have different CAS registry numbers as illustrated in Example 1.

This technology is ideally suited to support the development of a database of chemical warfare agents and precursors listed in relevant national and international legislation and regulations. In fact, based on the JChem Base and the Biovia Direct technology, the software companies ChemAxon and Scitegrity have already developed and offer cheminformatics tools composed of a database with an associated software for controlled substances chiefly intended as prescription and illegal narcotic and psychotropic drugs. The ChemAxon and Scitegrity tools named Compliance Checker and Controlled Substances Squared (CS2) were improved or developed within the scope of the Substance Compliance Service Project of the Pistoia Alliance to provide pharmaceutical companies with tools to identify substances that are controlled under national and international legislation.[62]

Although primarily focused on controlled narcotic and psychotropic substances, Compliance Checker and Controlled Substances Squared also offer some coverage of control lists relevant to chemical weapons and precursors. In particular, both tools cover the three CWC Schedules and the EU regime for the control of exports, transfer, brokering, and transit of dual-use items.[63] As of January 2020, however, neither tool covers the 2012 EU regulation mentioned in Example 6, which requires prior authorization for export to Syria of a list of chemicals, including isopropanol, not covered under the EU's dual-use export regulations. In addition, Compliance Checker also covers German control lists related to CWC implementation and the export of military weapons,while Controlled Substances Squared covers the Australia Group's Chemical Weapons Precursors List.[64,65] As a result, Compliance Checker and Controlled Substances Squared are both well suited for developing a cheminformatics tool that can bolster the control

---

59 CAS, "SciFinder," <https://www.cas.org/products/scifinder>.

60 CAS, "STN," <https://www.cas.org/products/stn>.

61 ChemAxon, "Markush tools," <https://chemaxon.com/products/markush-tools>.

62 Taylor, et al., "The Pistoia Alliance Controlled Substance Compliance Service Project;" ChemAxon, "Compliance Checker," <https://chemaxon.com/products/compliance-checker>; and Scitegrity, "Controlled Substances Squared," <https://www.scitegrity.co.uk/index.php?page=cs2>.

63 European Commission, "Dual-use Trade Controls," <http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual-use-controls/>.

64 Ausführungsverordnung zum Chemiewaffenübereinkommen, Bundesministerium der Justiz und für Verbraucherschutz, <http://www.gesetze-im-internet.de/cw_v/>.

65 Ausführungsgesetz zu Artikel 26 Abs. 2 des Grundgesetzes (Gesetz über die Kontrolle von Kriegswaffen), Bundesministerium der Justiz und für Verbraucherschutz, <https://www.gesetze-im-internet.de/krwaffkontrg/BJNR004440961.html>.

of chemical weapons and their precursors through the addition of all relevant national and international lists of controlled chemicals. Of note, both tools are already structured in a way that allows users to select one or more jurisdictions of interest prior to launching a query. Moreover, Scitegrity already offers a tool, named ExpediChem, that determines the HS code of a chemical from its chemical structure.[66]

As discussed, it may be useful to include a feature that allows users to input query chemicals by scanned printed molecular structures. Neither Compliance Checker nor Controlled Substances Squared are endowed with such a feature. However, technology to fulfill this task has been developed.[67] Among others, the National Cancer Institute has developed an open source application to fulfill this task called OSRA (Optical Structure Recognition Application).[68] The scanning feature of the software tool that the authors propose could be based on these existing or similar algorithms.

It is worth noting that although the front end of a cheminformatics tool intended to support the control of chemical weapons and precursors should be designed in a way that makes it easy to use for non-chemists, frontline officers might have to redirect more complex cases to officials with additional training.

7. Conclusions

To bolster the control of chemical warfare agents and their precursors, the authors propose the development and adoption of a cheminformatics tool composed of a database and an associated software. This tool is intended to help frontline officers and chemical companies overcome major problems that currently make it cumbersome to thoroughly check whether a chemical of interest is included on a list of controlled chemicals.

First, by converting any entered chemical name or registry number into a chemical structure, the proposed tool would allow a thorough checking of control lists, unhindered by the fact that the same chemical can be identified with a multitude of synonyms and that different versions of the same chemical will have different registry numbers associated with them.

Second, the proposed cheminformatics tool would make it easier to handle control lists that encompass whole families of related chemicals of concern in addition to individual chemicals. In particular, the proposed tool will allow operators to automatically confirm whether a chemical belongs to one of the listed families of chemicals.

---

66 Scitegrity, "Harmonized Tariff Schedule Classification," <https://www.scitegrity.co.uk/index.php?page=Tariff-coding>.

67 Igor V. Filippov, Mihai Lupu, and Alan P. Sexton, "Modern Approaches to Chemical Image Recognition," in Mihai Lupu, Katja Mayer, Noriko Kando, and Anthony J. Trippe, eds., *Current Challenges in Patent Information Retrieval* (Berlin, Heidelberg: Springer, 2017), pp. 369-389.

68 National Cancer Institute, "OSRA: Optical Structure Recognition Application,". <https://cactus.nci.nih.gov/osra/>.

Although the technology for its implementation largely exists, a cheminformatics tool must be designed to have an easy-to-use interface geared toward non-chemists and be tailored to the regulatory requirements of different jurisdictions. Such an interface should allow users to input chemicals in several different ways and should produce an easy-to-interpret output that clearly indicates whether the queried chemical falls within the scope of that jurisdiction's control lists. Also, very importantly, a thorough study will have to be conducted in order to identify relevant national and international control lists and add them to the database, which will have to be promptly updated as the lists are amended.

By helping frontline officers as well as chemical and shipping companies to handle families of chemicals, the adoption of a cheminformatics tool could also foster, in the long term, expansion of control lists that currently exclusively list single chemicals, such as the list of chemical precursors produced by the Australia Group or the WCO Strategic Trade Control Enforcement Implementation Guide, to include families of chemicals. The addition of families of chemicals to control lists would make these lists more comprehensive and more resilient to changes in the proliferation landscape, thereby closing loopholes that would allow proliferators to seek chemical warfare agents or precursors closely related to controlled chemicals but not encompassed by the control lists.

Finally, although the focus of this paper is on chemical weapons nonproliferation, the proposed cheminformatics tool can be generally outfitted with datasets relative to any list of controlled chemicals. For instance, datasets could include, among others, the explosive precursors listed by Programme Global Shield, the chemicals that pose a threat for nuclear or missile proliferation listed in the NSG and MTCR, as well as narcotic and psychotropic precursors covered by the Pistoia Alliance project.[69] This would provide frontline officers with a single tool to handle all lists of controlled chemicals, independent of the reasons for their control.

69 World Customs Organization, "Improvised Explosive Devices (IEDs) Programme Global Shield". <http://www.wcoomd.org/en/topics/enforcement-and-compliance/activities-and-programmes/security-programme/programme-global-shield.aspx>.