

# WHAT IS THE LANGUAGE OF CHEMISTRY?

By Giovanna Pontes

The language of chemistry is written in connected atoms to illustrate molecular structures. The structure of a molecule (the chemical) is the most informative way to describe a chemical substance and understand its chemistry. Chemicals are given names and numerical identifiers, however relying solely on information that does not illustrate the chemical structure may fail to identify when a chemical substance falls under the Schedules of the Chemical Weapons Convention. To understand why, we explore the chemical language of the nerve agent sarin, a Schedule 1A01 chemical.



Explore "The Science for Diplomats Annex on Chemicals"

## WHAT IS IN A CHEMICAL ABSTRACTS SERVICE (CAS) REGISTRY NUMBER?

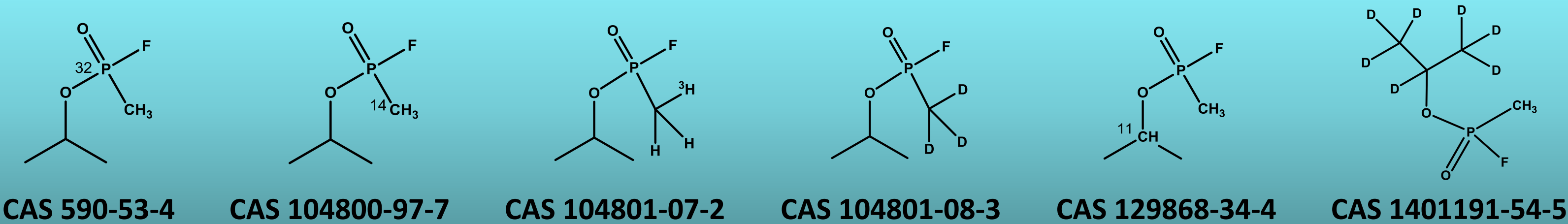
Chemical Abstract Service Registry numbers (CAS numbers) are a unique numerical identifier that are assigned to every chemical compound reported in scientific literature. CAS numbers contain up to 10 digits, divided into three parts by hyphens. For sarin, **CAS 107-44-8** is listed in the Annex on Chemicals of the Chemical Weapons Convention.

A unique CAS number is assigned to a specific chemical substance, however there is not necessarily a one-to-one relationship between CAS numbers and molecular structures. Different CAS numbers will be assigned, as illustrated below with sarin, for molecular structures indicating stereochemical configuration, containing isotopic labels, existing as hydrates and forming inclusion complexes. Defined compositions of chemicals in mixtures also receive unique CAS numbers.

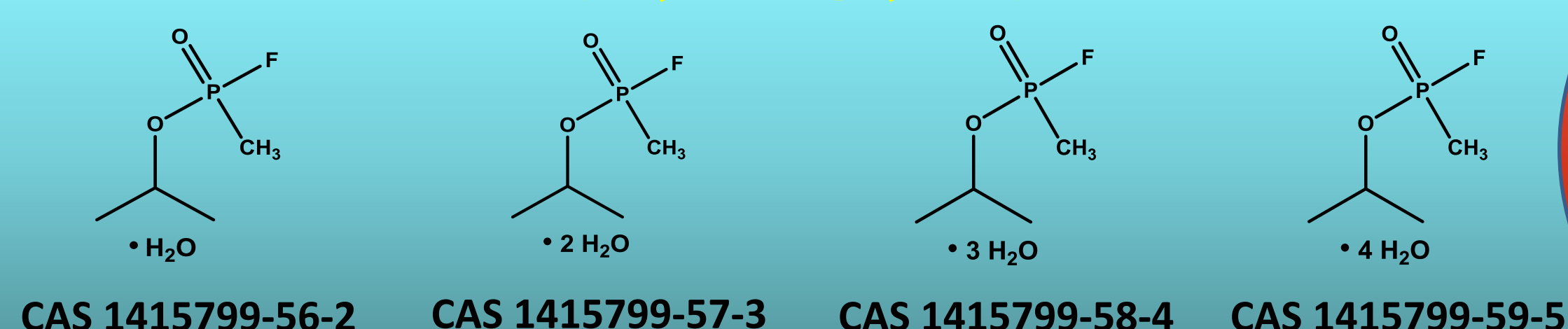


Learn more about CAS numbers

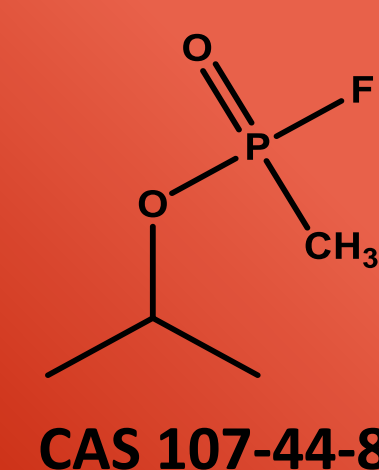
### ISOTOPICALLY LABELLED SARIN



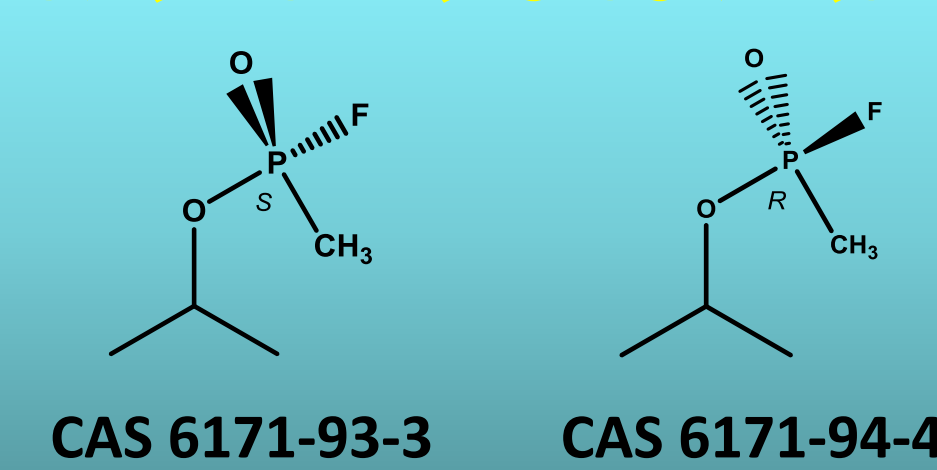
### SARIN HYDRATES



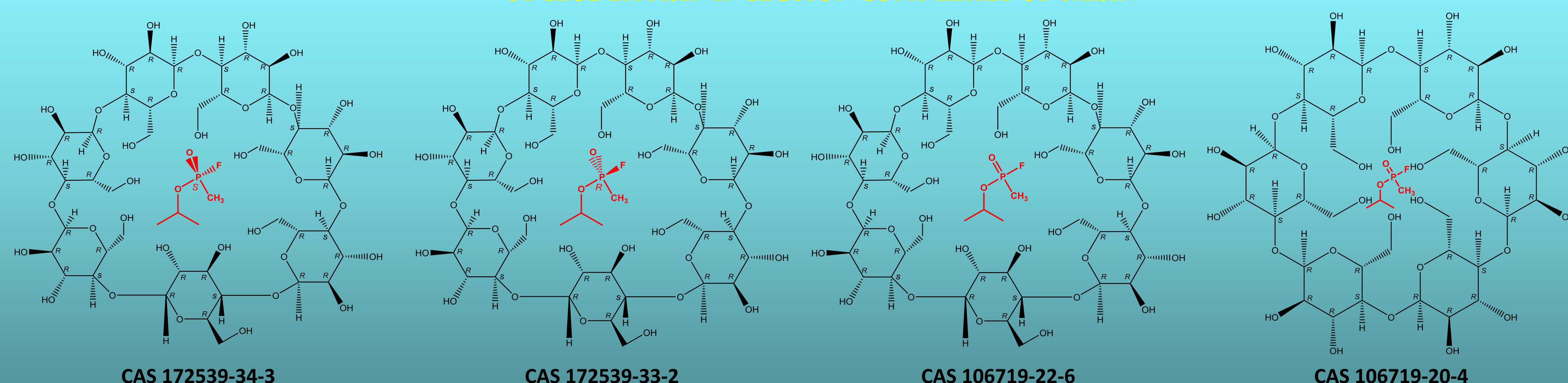
### SARIN



### SARIN STEREOISOMERS



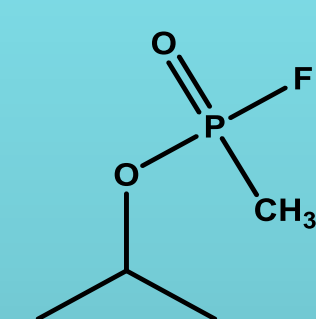
### CYCLODEXTRIN INCLUSION COMPLEXES OF SARIN



## OTHER WAYS TO DESCRIBE SARIN

### The IUPAC International Chemical Identifier (InChI):

An InChI is a chemical structure that has been converted into a machine-readable string of information. The string is unique to the compound it describes, and can encode absolute stereochemistry and isotopic labels. An InChI can be thought of as a bar-code for chemistry and chemical structures. The InChI enables efficient linking of diverse data compilations from printed and electronic sources. There is also an "InChI Key" designed for use with internet search engines to more easily find links to a given InChI.



The **InChI** for the structure on the left is **InChI=1S/C4H10FO2P/c1-4(2)7-8(3,5)6/h4H,1-3H3.**

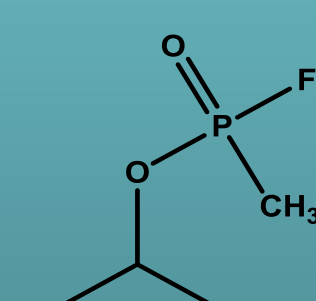
The **InChI Key** for the structure on the left is **DYAHQFWOVKZOOV-UHFFFAOYSA-N**



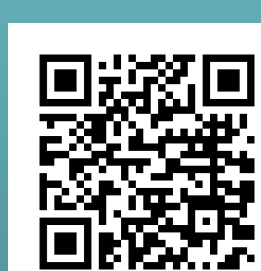
Learn more about InChI

### Canonical SMILES:

The simplified molecular-input line-entry system (SMILES) is a line notation describing a chemical structure in short ASCII strings. SMILES strings are designed to be readable by software that processes molecular structure information. SMILES are easily converted back into 2D or 3D molecular representations.



The **canonical SMILES** for the structure on the left is **CC(C)OP(=O)(C)F**



Learn more about canonical SMILES

## WHAT DOES ALL THIS MEAN?

"... there is a great need to help decision makers more effectively comprehend chemical information. Annotating the Annex on Chemicals with chemical structures should be considered"  
- Report of the Twenty-Eighth Session of the Scientific Advisory Board, paragraph 8.13 (SAB-28/1, dated 14 June 2019)



Read SAB-28/1

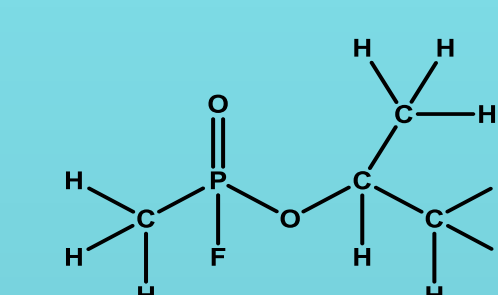
## WHAT IS IN A MOLECULAR FORMULA?

Molecular formulas describe the atomic composition of a chemical structure. For example, the molecular formula of sarin is **C<sub>4</sub>H<sub>10</sub>FO<sub>2</sub>P**, which indicates that a molecule of sarin contains:

- 4 carbon (C) atoms
- 10 hydrogen (H) atoms
- 1 fluorine (F) atom
- 2 oxygen (O) atoms
- 1 phosphorus (P) atom

Molecular formulas are compact and easy to communicate; however, unlike a molecular structure they do not illustrate the connections between the atoms.

### MOLECULAR STRUCTURE OF SARIN

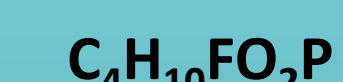
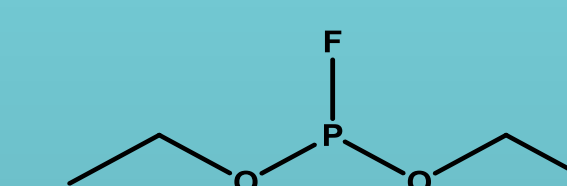


### MOLECULAR FORMULA OF SARIN

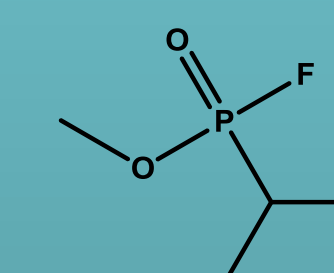


Molecular formulas may not be unique for a specific chemical substance. For instance, the following chemicals share an identical molecular formula with sarin, yet are different chemicals:

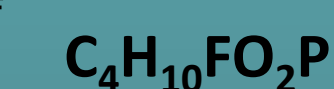
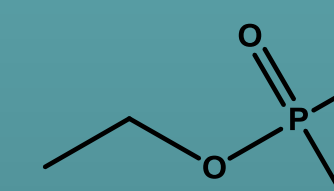
### DIETHOXY(FLUORO)PHOSPHANE



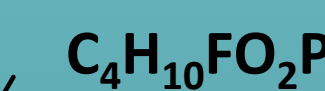
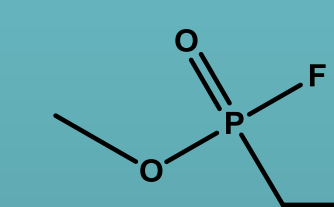
### METHYL ISOPROPYLPHOSPHONO FLUORIDATE



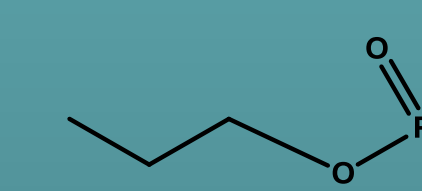
### ETHYL ETHYLPHOSPHONO FLUORIDATE



### METHYL PROPYLPHOSPHONO FLUORIDATE



### PROPYL METHYLPHOSPHONO FLUORIDATE



## SARIN

### SCIENTIFIC ADVISORY BOARD ADVICE

With reference to isotopic labels and stereoisomers, OPCW's Scientific Advisory Board has advised against relying solely on CAS numbers to define chemicals covered by the Schedules.



Review the SAB advice and its impact



## WHAT IS IN A CHEMICAL NAME?

The universal standard for chemical nomenclature ("the naming of chemicals") is defined by the International Union of Pure and Applied Chemistry (IUPAC). Yet, a chemical can have many names for historical and common use reasons, or simply to make it easier to talk about! The names listed below can be found within a broad set of databases and lists of chemical information. These are all synonyms for sarin. The list is NOT comprehensive, and all these names are reproduced exactly as formulated in the original source.

- (±)-Isopropyl methylphosphonofluoridate
- (±)-sarin
- (±)-Sarin
- (R<sub>S</sub>)-isopropyl methylphosphonofluoridate
- (R<sub>S</sub>)-O-isopropyl methylphosphonofluoridate
- (R<sub>S</sub>)-sarin
- 1-Methylethyl (±)-methylphosphonofluoridate
- 1-Methylethyl methylphosphonofluoridate
- 2-(FLUORO-METHYL-PHOSPHORYL)OXYPROPANE
- GB
- Glaucarubin
- IMPF
- ISOPROPOXY(METHYL)PHOSPHINOYL FLUORIDE
- Isopropoxymethylphosphoryl fluoride
- Isopropoxymethylphosphoryl fluoride
- Isopropoxymethylphosphoryl fluoride
- Isopropyl methanefluorophosphonate
- Isopropyl methylfluorophosphate
- Isopropyl methylfluorophosphonate
- Isopropyl methylfluorophosphonate

- Isopropyl methylphosphonofluoridate [*IUPAC Name*]
- Isopropylester kyseliny methylfluorofosfonove [*Czech*]
- Isopropyl-methylphosphonofluoridat [*German*]
- ISOPROPYL-METHYL-PHOSPHORYL FLUORIDE
- Metilfosfonofluoridato de O-isopropilo [*Spanish*]
- Metilfosfonofluoridato de O-isopropilo [*Spanish*]
- Methylfluorophosphonic acid isopropyl ester
- Methylfluorophosphoric acid isopropyl ester
- Methylfluorophosphorsaeureisopropyl ester
- Methylfluorophosphorsaeureisopropylester [*German*]
- méthylphosphonofluoridate de O-isopropyle
- Methylphosphonofluoridate d'isopropyle [*French*]
- Methylphosphonofluoridate acid, isopropyl ester
- Methylphosphonofluoridate acid 1-methylethyl ester
- Methylphosphonofluoridate acid isopropyl ester
- Isopropylmethyl Phosphonofluoridate
- O-Isopropyl methylfluorophosphonate
- o-Isopropyl methylphosphonofluoridate
- o-Isopropylmethyl Phosphonofluoridate
- O-изопропилметилфторфосфонат [*Russian*]

- ortho Isopropylmethyl Phosphonofluoridate
- ortho-Isopropylmethyl Phosphonofluoridate
- Phosphine oxide, fluoroisopropoxymethyl-
- Phosphonofluoridate, o-Isopropylmethyl
- Phosphonofluoridate, ortho-Isopropylmethyl
- Phosphonofluoridic acid, methyl-, 1-methylethyl ester
- Phosphonofluoridic acid, methyl-, isopropyl ester
- Phosphonofluoridic acid, methyl-, isopropyl ester, (±)
- Phosphonofluoridic acid, P-methyl-, 1-methylethyl ester
- Phosphoric acid, methylfluoro-, isopropyl ester
- PROPAN-2-YL FLUORO(METHYL)PHOSPHINATE
- propan-2-yl methylphosphonofluoridate
- racemic sarin
- rac-isopropyl methylphosphonofluoridate
- rac-propan-2-yl methylphosphonofluoridate
- Sarin
- Sarin II
- SARIN, (-)-
- SARIN, (+)-
- Зарин [*Russian*]

•Зарин [*Russian*]

•Sarin [*Spanish*]

•S-GB

•Zarin

•[*Arabic*] السارين

•[*Arabic*] مثيل فوسفونو فلوريدات أ-ايسوبروبيل

• 沙林 [*Chinese*]

• 甲基氟膦酸异丙酯 [*Chinese*]

WHAT IS YOUR FAVOURITE NAME FOR SARIN?