

Gebruikershandleiding voor de ZZS Similarity Tool

De ZZS similarity tool voorspelt of een stof structureel vergelijkbaar is met een ZZS (Zeer Zorgwekkende Stof). Structurele gelijkenis kan een indicatie zijn voor vergelijkbare Zeer Zorgwekkende eigenschappen. Het model is gebaseerd op de methodologie zoals beschreven en geëvalueerd door Wassenaar et al. (2019¹, 2021² en 2022³).

De ZZS similarity tool berekent op basis van een structuur van een molecuul in hoeverre het molecuul lijkt op bestaande ZZS afkomstig van de ZZS lijst. De uitkomsten van het tool laten zien op welke ZZS de stof het meest lijkt en welke mogelijke ZZS gevaarseigenschappen daarbij horen.

Het instrument is beschikbaar op: <https://rvszoekstelsysteem.rivm.nl/ZzsSimilarityTool> (zie Figuur 1).

Deze gebruikershandleiding beschrijft de technische procedure voor het gebruik van de tool en beschrijft hoe de resultaten worden weergegeven. Deze handleiding geeft geen informatie en toelichting hoe de resultaten van de tool verder gedeut kunnen worden of welke vervolgacties wenselijk zijn om de ZZS gevaarscriteria verder te specificeren. De volgende secties komen aan bod: Toepassingsgebied, Input, Output, Methodologie en Technische opmerkingen over de (implementatie van de) methodologie.



Figuur 1. Link naar de ZZS similarity tool.

¹ <https://doi.org/10.1016/j.comtox.2019.100110>

² <https://doi.org/10.1016/j.yrtph.2020.104834>

³ <https://doi.org/10.1002/jcc.26859>

Nieuw in deze versie

In januari 2020 is een eerste versie van de tool gelanceerd. In januari 2022 is de tool geüpdatet. In deze nieuwe versie zijn de volgende elementen toegevoegd en bijgewerkt:

- **Zoeken op CAS-nummer:** Het is nu mogelijk om een stof in te voeren op basis van een CAS-nummer. In de vorige versie was het alleen mogelijk om een stof in te voeren op basis van SMILES.
- **Zoeken in Batch-modus:** Het is nu mogelijk om de resultaten van de similarity tool voor een lijst van stoffen in één keer te berekenen en op te halen (met een maximum van 200 stoffen).
- **Update van de onderliggende dataset:** Alle tot 25 januari 2021 geïdentificeerde ZZS zijn toegevoegd aan de database van de ZZS similarity tool. In de vorige versie waren enkel de ZZS opgenomen die tot 1 maart 2018 geïdentificeerd waren.
- **Her-optimalisatie van de modellen:** De ZZS zijn onderverdeeld in vijf categorieën om de toxische eigenschappen en bijbehorende status beter weer te geven (CM, R, PBT/vPzB, ED en 'Other'), in plaats van drie categorieën (CMR, PBT/vPzB en ED). Omdat er nieuwe ZZS zijn toegevoegd en deze onderverdeeld in nieuwe categorieën, zijn ook de similarity modellen opnieuw geoptimaliseerd.
- **Verbeterde interpretatie van de uitkomsten:** Een kwantitatieve betrouwbaarheidsscore is toegevoegd aan de voorspellingen. Deze score is een toevoeging op de al bestaande kwalitatieve conclusie (voldoende vergelijkbaar: ja/nee), en draagt bij aan een betere interpretatie van de uitkomsten.

Toepassingsgebied van de tool

De methodologie zoals geïmplementeerd in de ZZS similarity tool is gebaseerd op het 'similar property principle', dat stelt dat stoffen met een vergelijkbare structuur mogelijk vergelijkbare toxische eigenschappen hebben. Daarom kan structurele gelijkenis tussen een stof en een bekende ZZS een indicatie zijn van vergelijkbare effecten, en een trigger zijn voor verdere inspectie en analyse.

De tool geeft aan of een stof mogelijk ZZS gevaarseigenschappen kan hebben. De ZZS-gevaarseigenschappen worden gedefinieerd in artikel 57 van REACH:

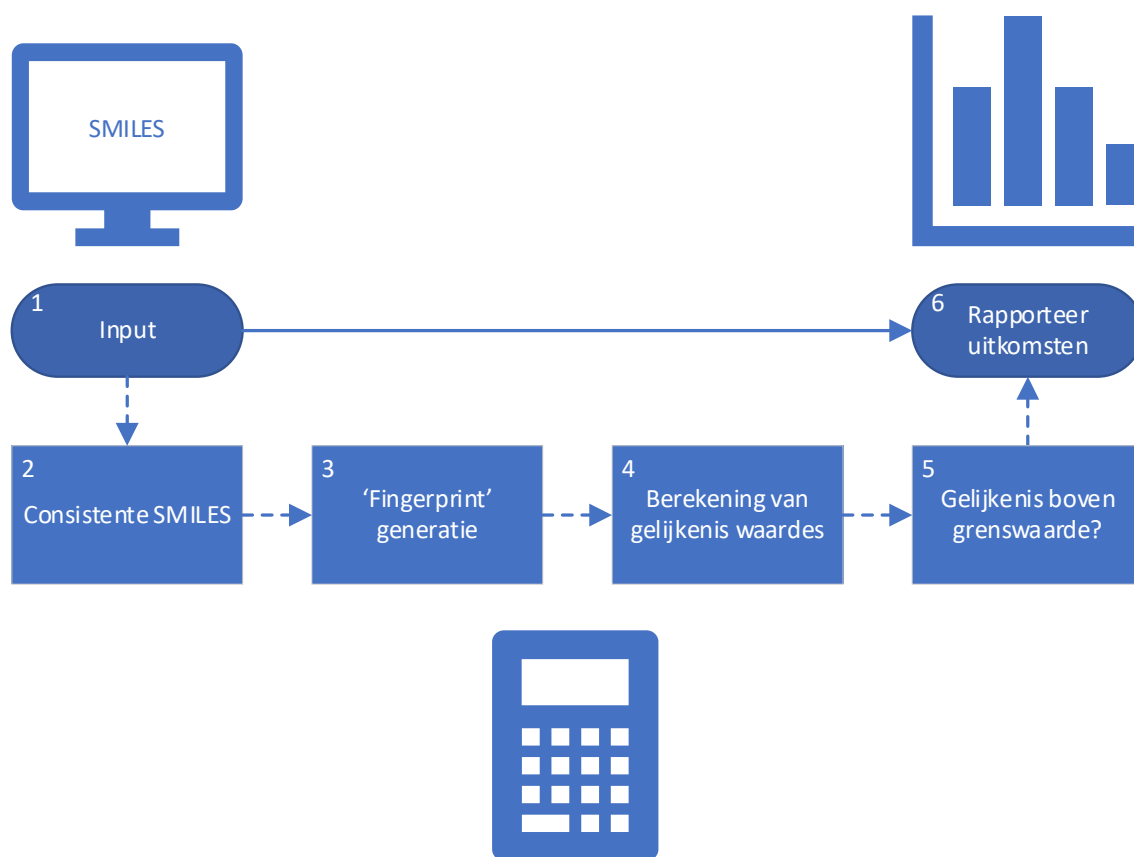
- a-c: CMR (ook bekend als: kankerverwekkend, mutageen en reprotoxisch).
- d-e: PBT/vPvB (ook bekend als: Persistent, Bioaccumulerend en Toxisch/ zeer Persistent en zeer Bioaccumulerend)
- f: Gelijkwaardig niveau van zorg. Dit zijn bijvoorbeeld stoffen met hormoonverstorende (ED) of sensibiliserende eigenschappen.

De ZZS similarity tool berekent of een stof structureel vergelijkbaar is met een ZZS met CM, R, PBT/vPvB en/of ED-eigenschappen overeenkomstig met de hierboven beschreven criteria. We noemen deze specifieke gevaarseigenschappen in deze handleiding ook wel 'ZZS-categorieën'. Daarnaast is er een categorie toegevoegd genaamd 'Other'. Deze categorie bestaat uit ZZS-stoffen die niet tot één van bovengenoemde categorieën behoren. Dit betreffen bijvoorbeeld stoffen die op de ZZS-lijst staan vanwege persistente, mobiele en toxische (PMT) eigenschappen, vanwege specifieke doelorgaantoxiciteit na herhaalde blootstelling (STOT-RE) of vanwege sensibiliserende eigenschappen. Daarnaast zijn stoffen ook opgenomen in de categorie 'Other' als ze niet op basis van opname in REACH, CLP of de POP-verordening op de ZZS-lijst staan. Voor stoffen die via andere routes op de ZZS-lijst staan zijn de gevaarseigenschappen niet altijd specifiek benoemd of eenduidig beschreven. Meer informatie over de onderliggende methodologie die is toegepast in de similarity tool is beschreven in de sectie 'Methodologie'.

Bij het gebruik van de tool moet het worden opgemerkt dat de tool niet instaat is samengestelde stoffen of mengsels te beoordelen (zie ook de sectie 'Input'). Wel kan de tool, indien de chemische structuur bekend is van alle afzonderlijke bestanddelen, de mogelijke ZZS gevaarscriteria berekenen voor deze bestanddelen afzonderlijk. Dit similarity tool kan niet gebruikt worden op stoffen die arseen, beryllium, cadmium, chroom, lood, kwik, nikkel en kobalt-metaalderivaten bevatten. Voor deze stoffen zijn met name de metaalatomen (of ionen) de oorzaak van de zorg, ongeacht de (organische) groepen die aanwezig zijn. Voor deze op metaal-gebaseerde complexen, wordt per definitie voorspeld dat ze ZZS zijn. De ZZS similarity tool kan wel worden gebruikt om een eerste voorspelling te genereren voor niet-dissociërende metalen (bijv. Organotin stoffen).

Daarnaast is het belangrijk – met betrekking tot de interpretatie van de resultaten – dat de afwezigheid van structurele gelijkenis met een ZZS niet per definitie geen ZZS-zorg betekent. Omgekeerd betekent een gelijkenis met een ZZS niet per definitie dat een stof een specifiek ZZS-effect heeft. Het is een trigger voor verdere inspectie en analyse. De ZZS similarity tool is bedoeld om te worden toegepast als een screeningsmodel.

Een vereenvoudigde workflow van het instrument is weergegeven in Figuur 2. Stap 1 is beschreven in de sectie 'Input' en stap 6 in de sectie 'Output'. De tussenliggende stappen zijn verder toegelicht in de sectie 'Methodologie'.



Figuur 2. Workflow van de ZZS similarity tool. Stap 1 en 6 beschrijven de in- en uitvoer zoals getoond door het model. Stap 2-5 betreffen processen die automatisch plaatsvinden wanneer u op "Calculate" drukt. Stap 1: definieer een stof als input in de vorm van een CAS-nummer of SMILES (zie paragraaf "Input"). Stap 2: De computer standaardiseert de SMILES om ervoor te zorgen dat er een gelijkwaardige vergelijking plaatsvindt met de ZZS-structuren. Stap 3: Twee verschillende vingerafdrukken worden gegenereerd voor de consistente/gestandaardiseerde SMILES (zie ³ voor meer informatie). Stap 4: De vingerafdrukken van de invoerstof worden vergeleken met de vingerafdrukken van respectievelijk alle CM, R, PBT / vPvB, ED en 'Other' ZZS. Verschillende similarity coëfficiënten worden toegepast voor de verschillende categorieën (zie ³ voor meer informatie). Stap 5: De computer analyseert of de gelijkens van de input-stof met de ZZS boven of onder de drempelwaarde liggen. Verschillende drempelwaardes worden toegepast voor de verschillende categorieën (zie ³ voor meer informatie). Stap 6: De resultaten van de ZZS similarity tool worden gerapporteerd (zie paragraaf "Output").

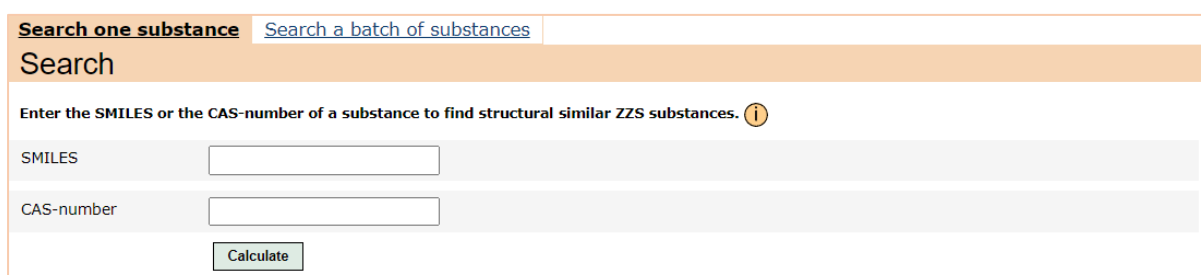
Input

Als invoer moet een SMILES-code of een CAS-nummer van een chemische structuur worden aan geleverd (zie Figuur 3).

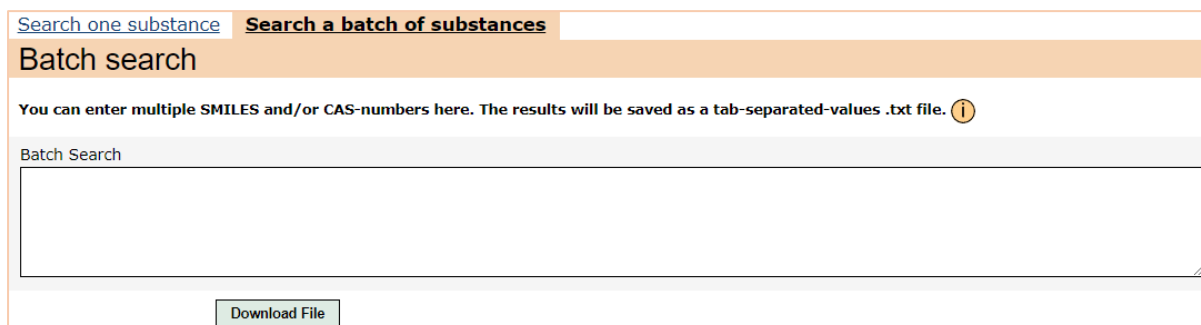
In principe gaat de voorkeur uit naar het gebruik van een CAS-nummer (indien mogelijk). Er is namelijk een lijst van 700.000 stoffen verbonden aan de dataset, en de structuren van deze stoffen zijn al voorgeprogrammeerd met betrekking tot SMILES neutralisatie en standaardisatie.

Als een CAS-nummer niet beschikbaar is, kan een SMILES (d.w.z. Simplified Molecular Input Line Entry System) gebruikt worden als input. Een SMILES-code vertegenwoordigt een chemische structuur door een lijnnotatie. Deze lijnnotatie kan door computersystemen worden geïnterpreteerd. Meer informatie over SMILES-expressie kan elders worden gevonden (bijvoorbeeld, ^{4,5,6}).

Voor het zoeken in Batch-modus kunnen zowel SMILES als CAS-nummers in combinatie gebruikt worden, zolang elke ingevoerde stof/notatie elke wordt gescheiden door een spatie, een komma, een verticale lijn (|) of door een nieuwe regel (zie Figuur 4).



Figuur 3. Input scherm "Search one substance"..



Figuur 4. Input scherm "Search a batch of substances".

Als u de SMILES van uw stof niet kent, kunt u de SMILES op verschillende websites zoeken, zoals:

- EPA's chemistry dashboard/DSSTox (<https://comptox.epa.gov/dashboard>)
Zoek uw stof op CAS of Naam. SMILES kunnen worden gevonden in de sectie 'structural identifiers'.
- Cactus (<https://cactus.nci.nih.gov/chemical/structure>)
Zoek uw stof op CAS of Naam and converter naar SMILES.
- PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)
Zoek uw stof op CAS of Naam. SMILES kunnen worden gevonden in sectie 2.1.4.

⁴ <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>

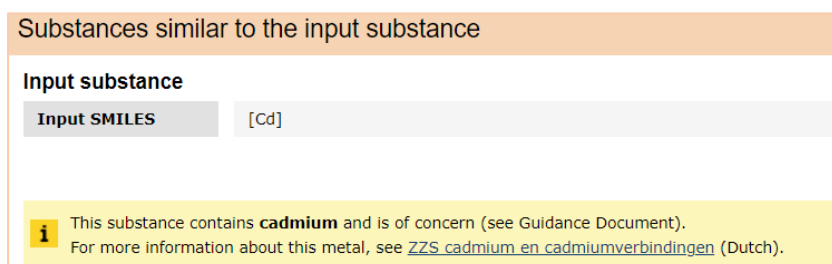
⁵ https://www.daylight.com/dayhtml_tutorials/languages/smiles/index.html

⁶ https://www.daylight.com/dayhtml_tutorials/languages/smiles/smiles_examples.html

Let op dat de ZZS similarity tool een SMILES-notatie vereist die één specifieke structuur vertegenwoordigt. Dit betekent dat mengsels of UVCB's ⁷ niet als geheel kunnen worden beoordeeld. Daarom is het gebruik van een punt (".") in de SMILES-notatie niet toegestaan (punten scheiden namelijk meerdere structuren binnen één SMILES-notatie). Voor mengsels of UVCB's wordt daarom geadviseerd om alle structuren – of een aantal representatieve structuren – gescheiden aan te bieden aan de ZZS similarity tool. Wanneer atomen in de SMILES-notatie met een lading worden uitgedrukt, zal de tool deze atomen waar mogelijk naar een neutrale vorm transformeren (zie ook ³).

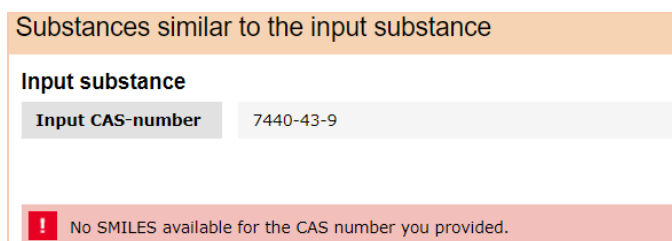
Het wordt aangeraden om altijd te evalueren of de ingevoerde structuur overeenkomt met uw chemische stof. Dit kan worden gedaan door de invoerstructuur te bekijken in het uitvoerscherm (zie de volgende paragraaf "Output").

Tijdens de invoer kunnen verschillende (fout)meldingen verschijnen. Voor metaal gebaseerde complexen – op basis van arseen, beryllium, cadmium, chroom, lood, kwik, nikkel of kobalt – wordt per definitie voorspeld dat ze ZZS zijn (Figuur 5). Als er een CAS-nummer wordt aangeleverd die niet in de onderliggende dataset zit verschijnt de melding zoals in Figuur 6. Wanneer u een ongeldige SMILES of CAS-nummer invoert, of wanneer de SMILES of CAS-nummer niet door het systeem als geldig wordt herkend, verschijnen de fouten zoals weergegeven in Figuur 7. Wanneer de ingevoerde SMILES te lang is of niet door het systeem kon worden verwerkt, verschijnen de foutmeldingen zoals weergegeven in Figuur 8. Daarnaast is het niet toegestaan om beide zoekvelden tegelijk te gebruiken (zie Figuur 9). Er mag enkel een CAS-nummer of een SMILES-code worden aangeleverd.



The screenshot shows a web interface titled "Substances similar to the input substance". Under the heading "Input substance", there is a button labeled "Input SMILES" and a text input field containing "[Cd]". Below this, a yellow warning box contains an information icon and the text: "This substance contains **cadmium** and is of concern (see Guidance Document). For more information about this metal, see [ZZS cadmium en cadmiumverbindingen](#) (Dutch)."

Figuur 5. Voorbeeld als een arseen, beryllium, cadmium, chroom, lood, kwik, nikkel of kobalt-metaalderivaat als invoer wordt opgegeven.



The screenshot shows a web interface titled "Substances similar to the input substance". Under the heading "Input substance", there is a button labeled "Input CAS-number" and a text input field containing "7440-43-9". Below this, a red error box contains an exclamation mark icon and the text: "No SMILES available for the CAS number you provided."

Figuur 6. Melding als het CAS-nummer niet beschikbaar is in de onderliggende dataset.

⁷ Substances of unknown variable composition, complex reaction products or biological materials.

Search

Enter the SMILES or the CAS-number of a substance to find structural similar ZZS substances. ⓘ

SMILES **The input is not a valid SMILES code.**

CAS-number **The input is not a valid CAS-number.**

Figuur 7. Fouten wanneer de invoer geen geldige SMILES-code is of niet wordt herkend als een geldige SMILES.

Search

Enter the SMILES or the CAS-number of a substance to find structural similar ZZS substances. ⓘ

SMILES **The field SMILES must be a string or array type with a maximum length of '750'.**

CAS-number

! An error occurred while processing the SMILES you provided.

! The generation of fingerprints took longer than the maximum allowed duration of 30 seconds.

Figuur 8. Fouten wanneer de invoer SMILES te lang is (meer dan 750 tekens) of wanneer de SMILES niet kon worden verwerkt.

! > You cannot search for a SMILES and a CAS-number at the same time.

Enter the SMILES or the CAS-number of a substance to find structural similar ZZS substances. ⓘ

SMILES

CAS-number

Figuur 9. Foutmelding wanneer beide zoekvelden tegelijk gebruikt worden.

Output

Het resultatenschermbestaat uit informatie over de ingevoerde stof en informatie over de resultaten.

Input ('Input substance')

In het resultatenschermbestaat ook informatie over de ingevoerde structuur verstrekt (zie Figuur 10). Deze informatie kan worden gebruikt om te controleren of de ingevoerde structuur overeenkomt met de stof die u bedoeld:

- **Input SMILES/CAS-number:** Toont de SMILES-notatie of CAS-nummer die u heeft opgegeven.
- **Consistent SMILES:** Toont de SMILES-notatie die door het model wordt gebruikt om de chemische gelijkheid te schatten. Deze stap wordt gebruikt om uniformiteit in SMILES-notatie voor de invoerstructuur en ZZS SMILES te waarborgen. Standaardisatie omvat bijvoorbeeld de neutrale expressie van atomen, waar mogelijk (zie ook ³ voor meer informatie).
- **Molecular structure:** Toont de structuur van de ingevoerde stof.

Input substance	
Input SMILES	<chem>c1ccccc1CC</chem>
Consistent SMILES	<chem>c1ccc(cc1)CC</chem>
Molecular structure	

Figuur 10. Voorbeeld van het uitvoerschermbestaat, met informatie over de ingevoerde stof.

Gelijkenis met ZZS

De resultaten worden geprioriteerd (d.w.z. de meeste vergelijkbare ZZS bovenaan) en gecategoriseerd in overeenstemming met ZZS-categorieën CM, R, PBT/vPvB, ED en 'Other'.

Zoals geïllustreerd in Figuur 11, heeft elke categorie (d.w.z. CM, R, PBT/vPvB, ED en 'Other') zijn eigen uitkomstentabel, en toont deze de meest vergelijkbare ZZS (bovenaan). Wanneer ten minste één ZZS binnen een categorie wordt beschouwd als structureel vergelijkbaar met de ingevoerde stof, worden de top drie meest vergelijkbare ZZS in die categorie weergegeven. Structureel vergelijkbare stoffen zijn oranje gemarkeerd (in plaats van grijs; zie Figuur 11). In het geval dat meer dan drie ZZS-stoffen als structureel vergelijkbaar worden beschouwd met de ingevoerde stof, kan additionele informatie worden verkregen door op de knop "Show all xxx similar substances" te drukken (zie Figuur 12). Als er in een categorie geen structurele vergelijkbare stoffen worden geïdentificeerd, wordt dit getoond (zie Figuur 13). De top drie meest vergelijkbare ZZS van deze categorie kunnen nog steeds worden weergegeven door op de knop "Show best matches" te drukken (zie Figuur 13).

Zoals hierboven gezegd, kan structurele gelijkheid een indicatie zijn van vergelijkbare toxiciteitsprofielen. Het moet worden opgemerkt dat de afwezigheid van structurele gelijkheid met een ZZS niet per definitie geen ZZS-zorg betekent. Omgekeerd betekent gelijkheid niet per definitie dat een stof een specifiek ZZS-effect heeft. Het geeft alleen een trigger voor verdere inspectie en analyse.

Similarity to CM substances									
Substance	CAS number(s)	EC number(s)	SMILES	Similar	Confidence in structural similarity	Details	Molecular structure	Possible toxicity	
styrene oxide	96-09-3	202-476-7	O1CC1c2ccccc2	Yes	64 %		Compare	☰	
chlorotoluene	100-44-7	202-853-6	c1ccc(cc1)CCl	Yes	51 %		Compare	☰	
5-nitroacenaphthene	602-87-9	210-025-0	O=[N+] ([O-])c3ccc2c1c3(cccc1CC2)	No	39 %		Compare	☰	

Figuur 11. Voorbeeld van het uitvoerscherm, met twee stoffen die structureel vergelijkbaar zijn met de ingevoerde stof (in oranje) en één stof die niet structureel vergelijkbaar is (in grijs).

Similarity to ED substances									
Substance	CAS number(s)	EC number(s)	SMILES	Similar	Confidence in structural similarity	Details	Molecular structure		
Representing 4-nonylphenol, branched	84852-15-3	284-325-5	Oc1ccc(cc1)CCCC(C)C(C)C	Yes	100 %		Compare		
Representing 4-heptylphenol, branched and linear			Oc1ccc(cc1)CCCCCCC	Yes	100 %	View	Compare		
Representing 4-nonylphenol, branched	84852-15-3	284-325-5	Oc1ccc(cc1)CCCC(C)C(C)C	Yes	99 %		Compare		

[Show all 22 similar substances](#)

Figuur 12. Voorbeeld van het uitvoerscherm wanneer meer dan 3 ZZS-stoffen van een categorie structureel vergelijkbaar zijn met de ingevoerde stof. .

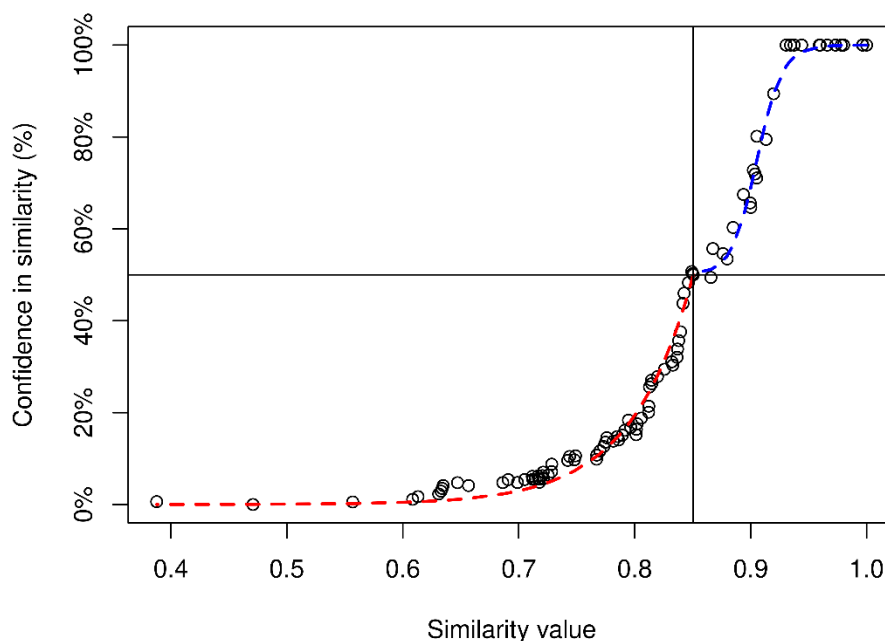
Similarity to PBT/vPvB substances									
<p>i No similar PBT/vPvB substances found.</p>									
<p>Show best matches</p>									

Figuur 13. Voorbeeld van het uitvoerscherm wanneer geen structureel vergelijkbare stoffen worden geïdentificeerd.

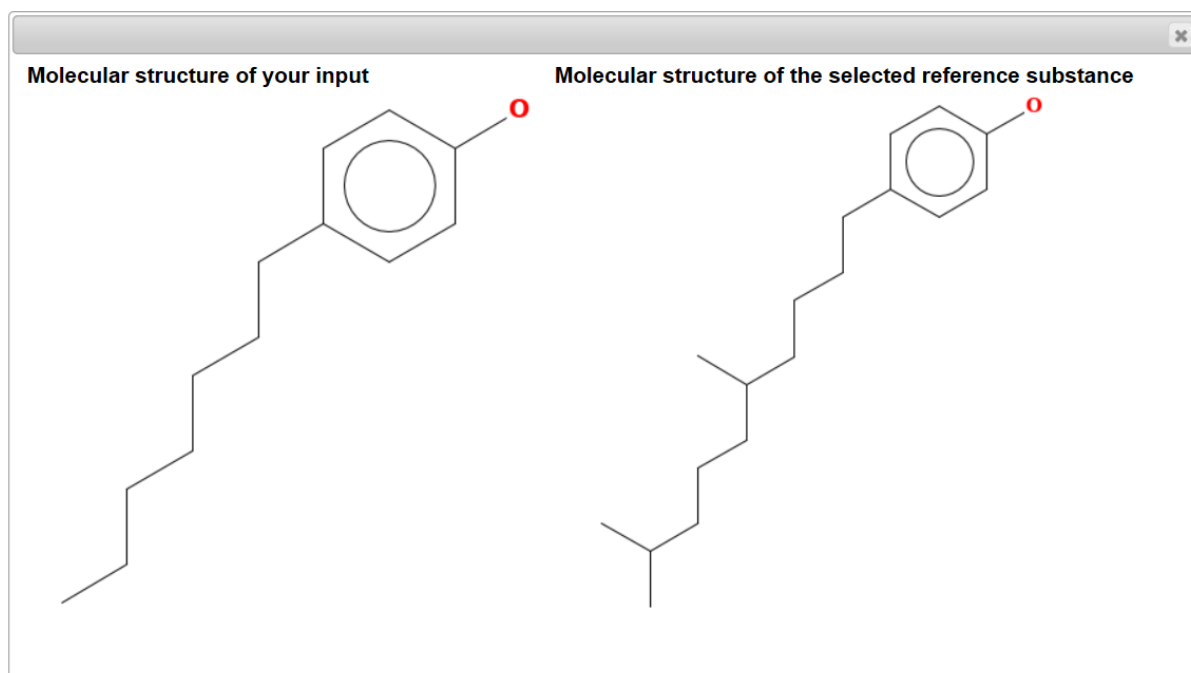
De volgende informatie is opgenomen in de resultaten-tabellen:

- **Substance:** Naam van de ZZS waarmee de ingevoerde stof wordt vergeleken.
- **CAS number(s):** CAS-nummer(s) van de ZZS waarmee de ingevoerde stof wordt vergeleken. Het CAS-nummer is een hyperlink naar de specifieke ZZS-stoffenpagina.
- **EC number(s):** EC-nummer(s) van de ZZS waarmee de ingevoerde stof wordt vergeleken. Het EC-nummer is een hyperlink naar de specifieke ZZS-stoffenpagina.
- **SMILES:** SMILES van de ZZS waarmee de ingevoerde stof wordt vergeleken.
- **Similar:** Ja of Nee. Wanneer de ingevoerde stof structureel vergelijkbaar wordt geacht met een ZZS, is de uitkomst 'Ja', anders is de uitkomst 'Nee'. Naast deze verklaring wordt de kleur van de rij bepaald door de gelijkenis (d.w.z. Ja = oranje; Nee = grijs).
- **Confidence in structural similarity:** Dit betreft een score die de mate van vertrouwen in de structurele gelijkenis weergeeft. De betrouwbaarheidsscore is gelijk aan of hoger dan 50% wanneer het model de ingevoerde stof als structureel vergelijkbaar beoordeelt met een ZZS. De verstreekte betrouwbaarheidsscores zijn gerelateerd aan de berekende gelijkeniscores tussen twee chemische stoffen, en de onderliggende verdeling is verschillend per model/ZZS-categorie (zie ³ voor meer details en Figuur 14 voor een voorbeeld).
- **Details:** Als er geen CAS- of EC-nummer beschikbaar is voor de ZZS-stof of groep ZZS-stoffen, wordt een link naar de ZZS-stofpagina verstrekt (zie Figuur 12).
- **Molecular structure:** Door op de knop "Compare" te drukken, worden de chemische structuren van de ingevoerde stof en de ZZS getoond (zie Figuur 15).
- **Possible toxicity:** ZZS-stoffen in de CM-, PBT/vPvB-, en 'Other'-dataset zijn opgenomen op basis van een specifieke zorg. Voor CMR kan dit kankerverwekkend of mutageen zijn. Voor

PBT/vPvB kunnen dit Persistente, Bioaccumulerende en Toxische eigenschappen zijn, of zeer Persistente en zeer Bioaccumulerende eigenschappen. Voor de 'Other-ZZS' kan dit gerelateerd zijn aan C-, M-, R-, PBT-, PBT/vPvB- of ED-gerelateerde eigenschappen, of aan andere type eigenschappen zoals PMT en sensibiliserend. In deze kolom wordt de mogelijke toxiciteit van de ZZS vermeld (er kunnen meerdere eigenschappen van toepassing zijn). Dit kan aanvullende informatie opleveren over mogelijke zorg voor de ingevoerde stof.



Figuur 14. De relatie tussen de structurele gelijkheid van een chemische stof en een ZZS en de betrouwbaarheidsscore. Elk gelijkheidsmodel heeft zijn eigen geoptimaliseerde distributie.

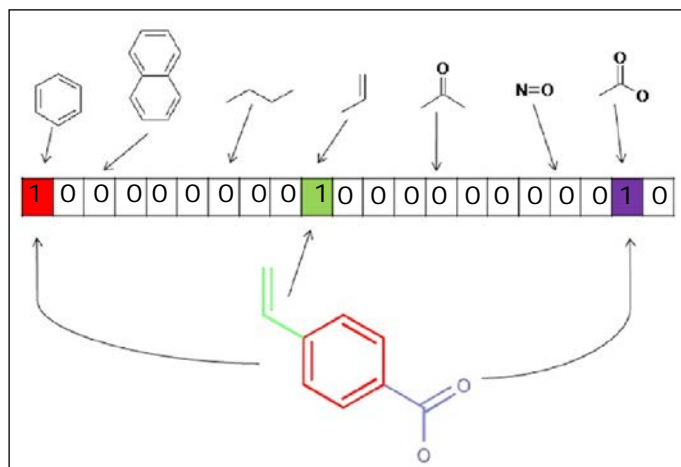


Figuur 15. Voorbeeld van het uitvoerscherm wanneer u op de knop "Compare" in de kolom "Molecular structure" drukt.

Methodologie

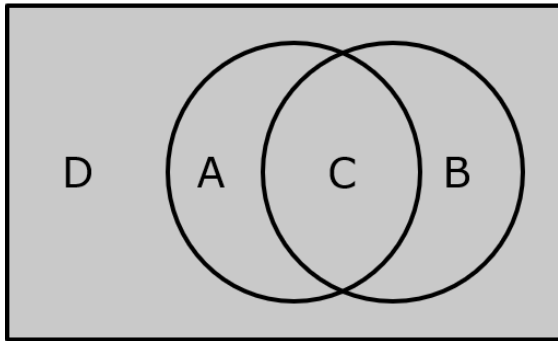
In deze paragraaf beschrijven we kort de onderliggende methodologie⁸. Structurele gelijkheid wordt uitgedrukt op basis van een similarity measure (maat voor vergelijkbaarheid) die bestaat uit een binaire vingerafdruk en een similarity coëfficiënt. Een vingerafdruk is een reeks binaire waarden die aan een structuur kan worden toegewezen en kan worden berekend op basis van de SMILES van een structuur (d.w.z. een specifieke chemische identificatie; zie ook paragraaf 'Input'). Deze vingerafdrucken hebben een specifieke lengte (bijvoorbeeld 1024) en bestaan uit 1'en en 0'en (zie Figuur 16 voor een illustratie). De vingerafdrucken van een stof en een ZZS-stof kunnen worden vergeleken met een similarity coëfficiënt die een gelijkheid geeft tussen 0 en 1. Waarin 0 aangeeft dat twee stoffen totaal verschillend zijn en een score van 1 aangeeft dat twee stoffen identiek zijn (zie Figuur 17 voor een illustratie).

Er zijn veel verschillende vingerafdrucken en similarity coëfficiënten in de literatuur beschikbaar en op basis van de analyse zoals beschreven in ³, is de beste combinatie geselecteerd voor stoffen van de verschillende ZZS-categorieën: CM, R, PBT/vPvB, ED en 'Other'. Vervolgens hebben we een optimale drempelwaarde afgeleid (d.w.z. een waarde tussen 0-1) voor de verschillende categorieën. Wanneer een nieuwe chemische stof met onbekende eigenschappen wordt vergeleken met een ZZS en de gelijkheid boven deze drempel ligt, dan wordt deze als structureel vergelijkbaar gezien met de ZZS. Dit betekent dat deze stof mogelijk zorgwekkend is voor de betreffende ZZS-categorie. Deze best presterende modellen (met bijbehorende drempelwaardes) zijn geïmplementeerd in de ZZS similarity tool (zie ook paragraaf 'Technische opmerkingen over de implementatie van de methodologie'). Er moet worden opgemerkt dat er nog meer details/specificaties van toepassing zijn op de ingebouwde modellen voor CM, R, PBT/vPvB, ED en 'Other', die u kunt vinden in ³. De volledige workflow van het instrument is weergegeven in Figuur 2.



Figuur 16: Illustratie van één type vingerafdruk, een numerieke uitdrukking van een chemische structuur. Elk bit in een bitstring is gerelateerd aan een specifiek fragment. Als dit fragment aanwezig is in een stof, wordt het bit ingesteld op "Aan" en wordt een score van 1 gegeven, als een fragment niet aanwezig is in een stof, wordt het bit ingesteld op "Uit" en wordt een score van 0 gegeven. Een vingerafdruk van een structuur bestaat dus uit 1'en en 0'en. Vingerafdrucken van twee stoffen kunnen worden vergeleken met behulp van een similarity coëfficiënt om de structurele gelijkheid uit te drukken (zie Figuur 17). Deze afbeelding is afkomstig van Cao et al. [Analytica Chimica Acta 752 (2012): 1-10].

⁸ Meer details zijn te vinden in ³; waarin we een groot aantal methoden hebben geanalyseerd die in de open literatuur beschikbaar zijn om de structurele gelijkheid tussen stoffen te beoordelen.



Substance a: 1 0 0 1 1 1 0 1 0 1
Substance b: 1 0 1 0 0 1 1 1 0 0
 C D B A A C B C D A

$$S_{ab} = \frac{N_C}{N_A + N_B + N_C}$$

$$S_{ab} = \frac{3}{3 + 2 + 3} = 0.375$$

Figuur 17: Illustratie van een similarity coëfficiënt, een formule om de overeenkomst tussen twee vingerafdrukken uit te drukken (met een waarde tussen 0 en 1). Een score van 0 geeft aan dat twee stoffen totaal verschillend zijn en een score van 1 geeft aan dat twee stoffen identiek zijn. Rechtsboven worden twee vingerafdrukken weergegeven. Als in beide stoffen een fragment aanwezig is, wordt dit een "C"-fragment genoemd; als een fragment niet in beide stoffen aanwezig is, wordt dit een "D"-fragment genoemd; als het fragment alleen in stof A aanwezig is, wordt het een "A"-fragment genoemd; en als het fragment alleen aanwezig is in stof B, wordt het een "B"-fragment genoemd (zie ook de linker afbeelding). Het aantal A-, B-, C- en D-fragmenten wordt geteld en kan worden gebruikt in een similarity coëfficiënt om de overeenkomst tussen stof A en B uit te drukken (zie rechtsonder voor een voorbeeld; er bestaat een grote variatie aan similarity coëfficiënten, dit is slechts een voorbeeld).

Technische opmerkingen over de (implementatie van de) methodologie

Het model is gebaseerd op de methodologie zoals geanalyseerd en geëvalueerd door Wassenaar et al. (2022)⁹ en maakt gebruik van de PaDEL-functionaliteiten (gebaseerd op de Chemistry Development Kit libraries)⁹.

Tijdens de implementatie van de CDK-vingerafdrukken en bijbehorende PaDEL-libraries zijn de vingerafdrukken van 19 ZS-stoffen niet identiek aan de vingerafdrukken zoals toegepast in Wassenaar et al. (2022) (waarschijnlijk vanwege kleine verschillen in de integratie van de onderliggende PaDEL-libraries voor SMILES-standaardisatie). Daarom zijn voor deze 19 ZS-stoffen de vingerafdrukken aangepast in de onderliggende dataset van het de ZS similarity tool.

Het model is geoptimaliseerd met de ZS lijst zoals beschikbaar op 25-01-2021. Nieuwe ZS-stoffen zullen worden toegevoegd tijdens een update van de ZS similarity tool.

⁹ PaDEL descriptor: <http://www.yapcwsoft.com/dd/padeldescriptor/>