

## 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. (2019)<sup>1</sup>.

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://rvszoekstysteem.rivm.nl/ZzsSimilarityTool> (zie Figuur 1).

Deze gebruikershandleiding beschrijft de technische procedure die moet worden gevolgd en beschrijft hoe de resultaten worden weergegeven. Deze handleiding geeft geen informatie en toelichting hoe de resultaten van de tool verder geduid 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.

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

## 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:

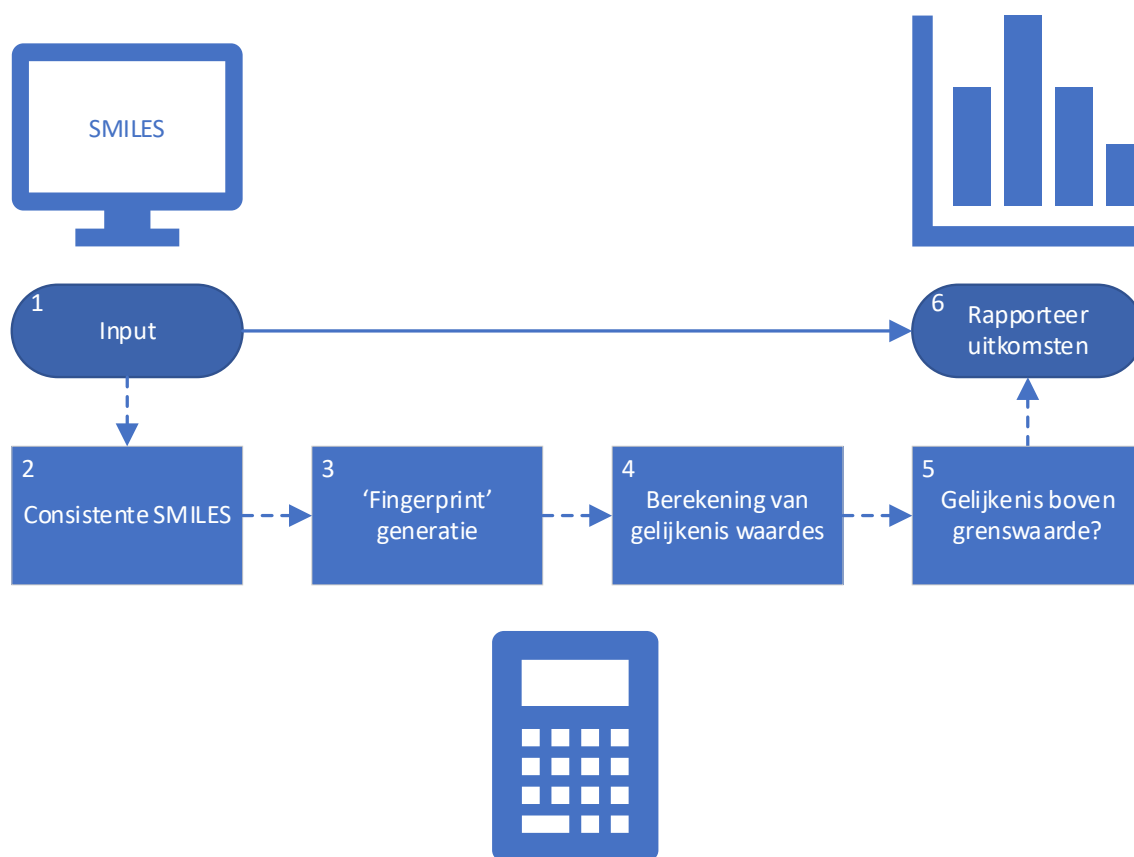
- a-c: CMR (ook bekend als: kankerverwekkend, mutageen en reprotoxisch).
- d-e: PBT/vPvB (ook bekend als: Persistent, Bioaccumulerend en Toxisch/ zeer Persistent en zeer Bioaccumulerend)
- f: Gelijkwaardig niveau van zorg. Dit zijn bijvoorbeeld hormoonverstorende of sensibiliserende eigenschappen.

De ZZS Similarity Tool berekent of een stof structureel vergelijkbaar is met een ZZS met CMR en/of, PBT/vPvB eigenschappen overeenkomstig met a-e criteria. Voor wat betreft artikel 57f kan de ZZS Similarity Tool enkel op mogelijke ED-eigenschappen screenen (ook bekend als: hormoonverstoorders). We noemen deze specifieke gevaarseigenschappen in deze handleiding ook wel 'ZZS-categorieën'. Meer informatie over de onderliggende methodologie is beschreven in de sectie 'Methodologie'.

Merk op dat de tool niet instaat is samengestelde stoffen of mengsels te beoordelen (zie ook de sectie 'Input'). Wel kan de tool, indien van de bestanddelen de chemische structuur bekend is, voor deze bestanddelen afzonderlijk de mogelijke ZZS gevaarscriteria berekenen. Dit model is niet van toepassing op stoffen met arseen, beryllium, cadmium, chroom, lood, kwik, nikkel en kobaltmetaalderivaten. 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).

Opgemerkt moet worden - 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 gelijkenis 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.

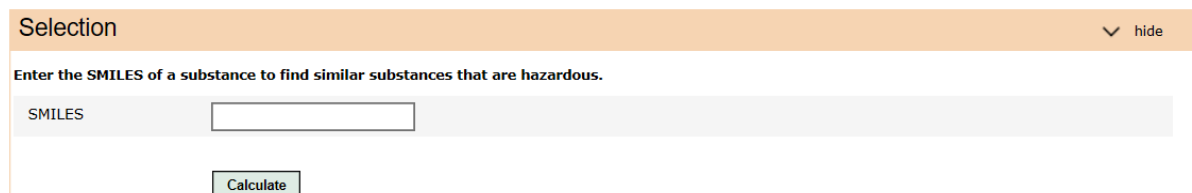
De volledige workflow van het instrument is weergegeven in Figuur 2. Stap 1 is beschreven in de sectie 'Input' en stap 6 in de sectie 'Output'. De tussenliggende stappen zijn verder toegelicht in de sectie 'Methodologie'.



Figuur 2. Workflow van de ZS 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 "berekenen" drukt. Stap 1: definieer een stof als input in de vorm van een SMILES (zie paragraaf "Input"). Stap 2: De computer standaardiseert de geleverde SMILES om een gelijke vergelijking met ZS-structuren te garanderen. Stap 3: Drie verschillende vingerafdrukken worden gegenereerd voor de consistente SMILES. De 'Extended'-vingerafdruk voor vergelijking met CMR ZS, de MACCS-vingerafdruk voor vergelijking met PBT/vPvB ZS en de PubChem-vingerafdruk voor vergelijking met ED ZS (zie <sup>1</sup> voor meer informatie). Stap 4: De vingerafdrukken van de invoerstof worden vergeleken met de vingerafdrukken van respectievelijk alle CMR, PBT / vPvB en ED ZS. Verschillende similarity coëfficiënten worden toegepast voor de verschillende categorieën (zie <sup>1</sup> voor meer informatie). Stap 5: De computer analyseert of de gelijkens van de input-stof met de ZS boven of onder de drempelwaarde liggen. Verschillende drempelwaardes worden toegepast voor de verschillende categorieën (zie <sup>1</sup> voor meer informatie). Stap 6: De resultaten van de ZS Similarity Tool worden gerapporteerd (zie paragraaf "Output").

## Input

Als invoer moet een SMILES-code van een chemische structuur worden aan geleverd (zie Figuur 3). Een SMILES (d.w.z. Simplified Molecular Input Line Entry System) vertegenwoordigt een chemische structuur door een lijnnotatie. Deze lijnnotatie kan door computersystemen worden geïnterpreteerd. Meer informatie over SMILES-expressie kan elders worden gevonden (bijvoorbeeld, <sup>2,3,4</sup>).



Figuur 3. Input scherm.

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.*
- ChemID Plus (<https://chem.nlm.nih.gov/chemidplus/>)  
*Zoek uw stof op CAS of Naam. SMILES kunnen worden gevonden in de sectie 'structure descriptors'.*

Let op dat de ZYS Similarity Tool een SMILES-notatie vereist die één specifieke structuur vertegenwoordigt. Dit betekent dat mengsels of UVCB's <sup>5</sup> 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 ZYS 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 <sup>1</sup>).

Het wordt aangeraden om altijd te evalueren of de input-structuur overeenkomt met uw chemische stof. Dit kan worden gedaan door de invoerstructuur zichtbaar te maken 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 ZYS zijn (Figuur 4). Wanneer u een ongeldige SMILES invoert of wanneer de SMILES niet door het systeem als geldig wordt herkend, verschijnen de fouten zoals weergegeven in Figuur 5. Wanneer de ingevoerde SMILES te lang is of niet door het systeem kon worden verwerkt, verschijnen de foutmeldingen zoals weergegeven in Figuur 6.

<sup>2</sup> <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>

<sup>3</sup> [https://www.daylight.com/dayhtml\\_tutorials/languages/smiles/index.html](https://www.daylight.com/dayhtml_tutorials/languages/smiles/index.html)

<sup>4</sup> [https://www.daylight.com/dayhtml\\_tutorials/languages/smiles/smiles\\_examples.html](https://www.daylight.com/dayhtml_tutorials/languages/smiles/smiles_examples.html)

<sup>5</sup> Substances of unknown variable composition, complex reaction products or biological materials.

Substances similar to the input substance

**Input substance**

Input SMILES

**i** This substance contains **cadmium** and is of concern (see Guidance Document).  
For more information about this metal, see [ZZS cadmium en cadmiumverbindingen](#) (Dutch).

Figuur 4. Voorbeeld als een arseen, beryllium, cadmium, chroom, lood, kwik, nikkel of kobalt-metaalderivaat als invoer wordt opgegeven.

Selection ▼ hide

Enter the SMILES of a substance to find similar substances that are hazardous.

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

**!** The SMILES you entered was not recognized as a valid SMILES.

Figuur 5. Fouten wanneer de invoer geen geldige SMILES-code is of niet wordt herkend als een geldige SMILES.

Selection ▼ hide

Enter the SMILES of a substance to find similar substances that are hazardous.

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

**!** An error occurred while processing the SMILES you provided.

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

Figuur 6. Fouten wanneer de invoer SMILES te lang is (meer dan 750 tekens) of wanneer de SMILES niet kon worden verwerkt.

## Output

Het uitvoerscherm bestaat uit informatie over de invoerstof en informatie over de resultaten.

### *Input ('Input substance')*

Binnen het uitvoerscherm wordt ook informatie over de invoerstructuur verstrekt (zie Figuur 7). Deze informatie kan worden gebruikt om te controleren of de invoerstructuur lijkt op uw chemische stof:

- **Input SMILES:** Toont de SMILES-notatie 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. Standardisatie omvat bijvoorbeeld de neutrale expressie van atomen, waar mogelijk (zie ook <sup>1</sup> 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 7. Voorbeeld van het uitvoerscherm, 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 CMR, PBT/vPvB en ED.

Zoals geïllustreerd in Figuur 8, heeft elke categorie (d.w.z. CMR, PBT/vPvB en ED) 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 8). In het geval dat meer dan drie ZZS-stoffen worden beschouwd als structureel vergelijkbaar met de ingevoerde stof, kan additionele informatie worden verkregen door op de knop "Show all xxx similar substances" te drukken (zie Figuur 9). Als er in een categorie geen structurele vergelijkbare stoffen worden geïdentificeerd, wordt dit getoond (zie Figuur 10). 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 10).

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 CMR substances							
Substance	CAS number(s)	EC number(s)	SMILES	Similar	Details	Molecular structure	Possible toxicity
styrene oxide	96-09-3	202-476-7	O1CC1c2ccccc2	Yes		<a href="#">Compare</a>	C
2-chlorotoluene	100-44-7	202-853-6	c1ccc(cc1)C(Cl)	Yes		<a href="#">Compare</a>	C
4-tert-butylbenzoic acid	98-73-7	202-696-3	O=C(O)c1ccc(cc1)C(C)(C)C	No		<a href="#">Compare</a>	R

Figuur 8. 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	Details	Molecular structure	
Representing nonylphenol	25154-52-3	246-672-0	Oc1ccc(cc1)CCCCCCCC	Yes		<a href="#">Compare</a>	
Representing 4-heptylphenol, branched and linear			Oc1ccc(cc1)CCCCCCC	Yes	<a href="#">View</a>	<a href="#">Compare</a>	
Representing 4-nonylphenol, branched	84852-15-3	284-325-5	Oc1ccc(cc1)CCCCCCC(C)C	Yes		<a href="#">Compare</a>	

[Show all 18 similar substances](#)

Figuur 9. Voorbeeld van het uitvoerscherm wanneer meer dan 3 ZZS-stoffen van een categorie structureel vergelijkbaar zijn met de ingevoerde stof. .

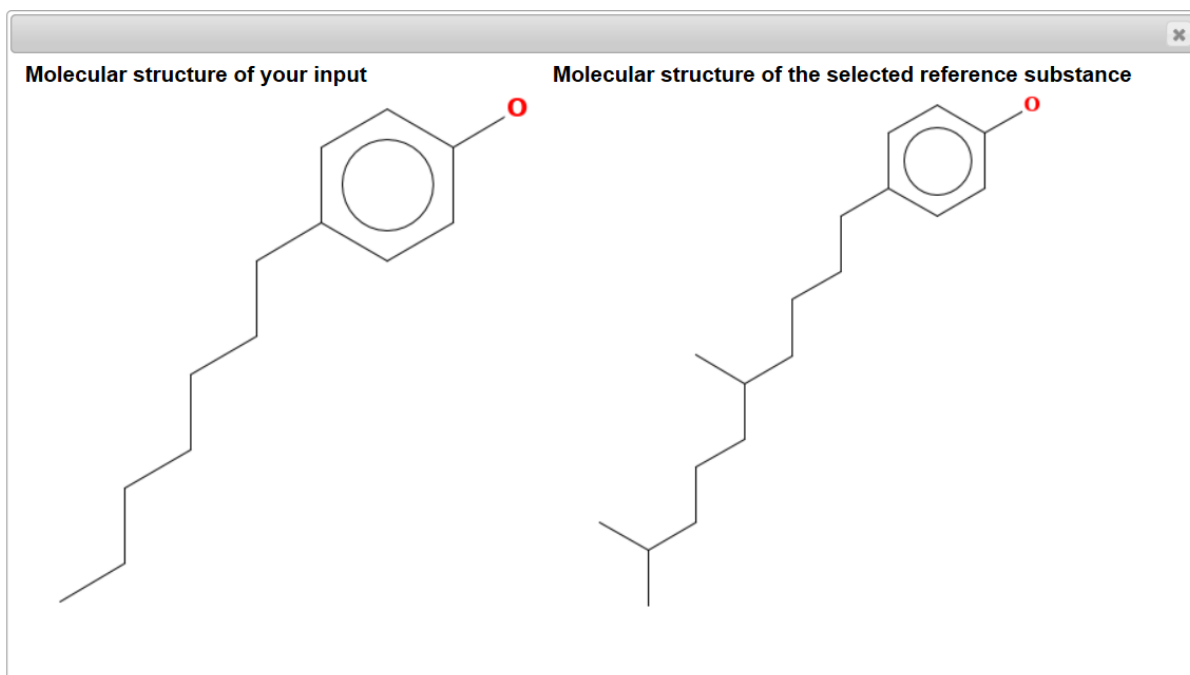
Similarity to PBT/vPvB substances	
<b>i</b>	No similar PBT/vPvB substances found.
<a href="#">Show best matches</a>	

Figuur 10. 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 gelijkens (d.w.z. Ja = oranje; Nee = grijs).
- **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 9).
- **Molecular structure:** Door op de knop "Compare" te drukken, worden de chemische structuren van de ingevoerde stof en de ZZS getoond (zie Figuur 11).
- **Possible toxicity:** ZZS-stoffen zoals opgenomen in de CMR- en PBT/vPvB-dataset zijn opgenomen op basis van een specifieke zorg. Voor CMR kan dit kankerverwekkend, mutageen of reprotoxisch zijn. Voor PBT/vPvB kan dit Persistente, Bioaccumulerende en Toxische eigenschappen zijn, of zeer Persistente en zeer Bioaccumulerende eigenschappen. 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.



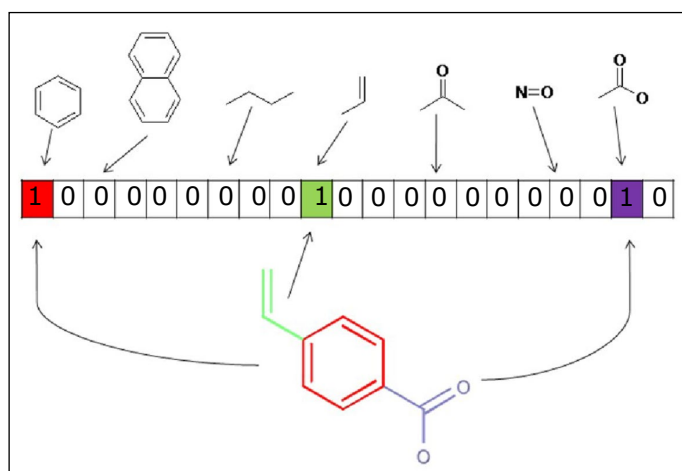
Figuur 11. Voorbeeld van het uitvoerscherm wanneer u op de knop "Compare" in de kolom "Molecular structure" drukt.



## Methodologie

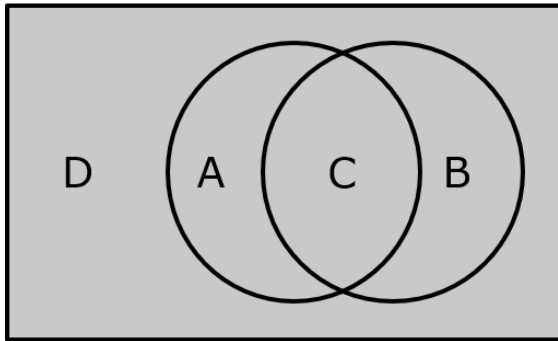
In deze paragraaf beschrijven we kort de onderliggende methodologie<sup>6</sup>. Structurele gelijkheid wordt uitgedrukt op basis van een similarity measure (maat voor vergelijkbaarheid) die bestaat uit een binaire vingerafdruk en een similarity coefficient. Een vingerafdruk is een reeks binaire waarden 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 12 voor een illustratie). De vingerafdrucken van een stof en een ZS-stof kunnen worden vergeleken met een similarity coefficient die een gelijkheid geeft tussen 0 en 1. Waarin 0 aangeeft dat twee stoffen totaal verschillend zijn en 1 aangeeft dat twee stoffen identiek zijn (zie Figuur 13 voor een illustratie).

Er zijn veel verschillende vingerafdrucken en similarity coëfficiënten in de literatuur beschikbaar en op basis van de analyse zoals beschreven in <sup>1</sup>, is de beste combinatie geselecteerd voor stoffen van de verschillende ZS-categorieën: CMR, PBT/vPvB en ED. 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 ZS en de gelijkheid boven deze drempel ligt, wordt deze als structureel vergelijkbaar gezien met de ZS. Dit betekent dat deze stof mogelijk zorgwekkend is voor de betreffende ZS-categorie. Deze best presterende modellen (met bijbehorende drempelwaardes) zijn geïmplementeerd in de ZS Similarity Tool (zie ook paragraaf 'Technische opmerkingen over de implementatie van de methodologie'). De volledige workflow van het instrument is weergegeven in Figuur 2.



Figuur 12: 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 13). Deze afbeelding is afkomstig van Cao et al. [Analytica Chimica Acta 752 (2012): 1-10].

<sup>6</sup> Meer details zijn te vinden in <sup>1</sup>; 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 13: Illustratie van een similarity coefficient, 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 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 coefficient om de overeenkomst tussen stof A en B uit te drukken (zie rechtsonder voor een voorbeeld; er zijn bestaan meerdere similarity coëfficiënten).

### Technische opmerkingen over de (implementatie van de) methodologie

Het model is gebaseerd op de methodologie zoals geanalyseerd en geëvalueerd door Wassenaar et al. (2019)<sup>1</sup> en maakt gebruik van de PaDEL-functionaliteiten (gebaseerd op de Chemistry Development Kit libraries)<sup>7</sup>. In tegenstelling tot Wassenaar et al. (2019), is het ED-model in deze online tool gebaseerd op de PubChem-vingerafdruk (2e beste model) in plaats van de FCFP4-vingerafdruk. Deze keuze is gemaakt om de implementatie te vereenvoudigen (alle vingerafdrukken in de ZZS-tool zijn allemaal gebaseerd op één programmeertaal), en kan kleine verschillen tot gevolg hebben tijdens het toepassen van het model.

Daarnaast konden tijdens de implementatie van de CDK-vingerafdrukken en bijbehorende PaDEL-libraries de vingerafdrukken van 14 ZZS-stoffen niet worden gereproduceerd zoals toegepast in Wassenaar et al. (2019) (waarschijnlijk vanwege kleine verschillen in de integratie van de onderliggende PaDEL-libraries voor SMILES-standaardisatie). Daarom zijn voor deze 14 ZZS-stoffen de vingerafdrukken aangepast in de onderliggende dataset van het model.

Het model is geoptimaliseerd met de ZZS lijst zoals beschikbaar op 01-03-2018. Nieuwe ZZS-stoffen worden toegevoegd tijdens een update van de ZZS Similarity Tool. Dit betekent ook dat er op dit moment geen overeenstemming wordt berekend met stoffen die zijn geclassificeerd als ZZS op basis van PMT-gerelateerde eigenschappen (persistent, mobiel en toxisch; die kunnen vallen onder artikel 57 f van REACH).

---

<sup>7</sup> PaDEL descriptor: <http://www.yapcwsoft.com/dd/padeldescriptor/>