Updated: 3 July 2017, Stellan Fischer
Edits: 25 July 2017, Emma Schymanski

# **KEMI Market List: Organic chemicals potentially identified on the EU market**

Stellan Fischer, Swedish Chemicals Agency (KEMI).

The “KEMI market list” was created by Stellan Fischer, KEMI, to support the identification work of “unknowns” in non-target mass spectrometry and is documented in this file. It consists of structurally well-defined substances that are expected to be on the market. The major information sources are regulatory databases (see Table 1 below). The list covers all types of chemicals (industrial chemicals, pharmaceuticals, pesticides etc.).

For a majority of the chemicals put on the market, the molecular structure is not fully defined. This market list only contains the well-defined ones, obtained as detailed below (“Chemical identity”). Future updates of structure data in authority databases will make it necessary to update this market list progressively.

The use pattern on the market can be utilized to predict the release and occurrence of chemicals in the environment. However, such information are, to a large extent, not publicly available due to business confidentiality. Instead here the “use information” is transformed into an “exposure score”. A hazard score has also been developed and included to the list to support the identification of potentially emerging substances (see “Risk-based prioritization”).

This list is now available on the NORMAN Suspect Exchange: <http://www.norman-network.com/?q=node/236>

## **Chemical identity**

Chemicals in the list are compiled from different national/regional inventory lists, with a focus on the EU market. Chemicals on “the market” here are defined as chemicals occurring on authority’s inventory lists.

In the first version of this list (Aug. 2016), a different use-related inventory list was compiled. Well-defined substances were selected (if they had a SMILES code). The final selection of substances was risk based (expected exposure and/or expected hazard, see “Risk based prioritization” below).

In the first update of the list (May 2017), the number of substances in the list doubled (~24,000) due to the following improvements:

* The identification and categorization of the use data was improved.
* The categorization of hazard data was improved (by name-based structural similarity).
* A non-EU use related inventory list was integrated (CPCat from US EPA). It is assumed to contribute to exposure within EU via import of goods.

The current list (July 2017) now contains 30,000 substances. The main data sources are given in Table 1. Many of these lists contain only CAS numbers or names. The CAS number has been the main or primary identifier in retrieving the structural information while creating the market list. The EC number have been used to trace the identity in the REACH and hazard classification databases.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1:** Overview of the data sources for the list of substances on the market, and their contribution to (i) identification of substances (ii) calculation of the Exposure score (iii) calculation of the Hazard score.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data source** | **Date**  | **Data** | **Main** | **Chemical** | **Exposure** | **Hazard** |
| **compiled** | **year** | **ID** | **identity** | **score** | **score** |
| **National/regional inventory lists** |   |   |  |   |   |   |
| Registered or preregistered substances according to REACH | Aug. 2016 | 2016 | CAS | x | x |   |
| The REACH Annex III inventory. (Q)SAR Indications for hazardous (eco)toxicological properties for “preregistered substances”. | Dec. 2016 | 2016 | CAS ECno | x | x |   |
| The Product registers of chemical products in the Nordic countries (worst case data, SPIN database) | Aug. 2016 | 2014 | CAS | x | x |  x |
| The Swedish Product Register (detailed use data) | Apr. 2017 | 2015 | CAS | x | x | x |
| The US EPA DSSTOX database | Aug. 2016 | 2015 | CAS | x |   |   |
| The US EPA CPCat database, May 2014 (chemicals in market products) https://actor.epa.gov/cpcat/faces/home.xhtm | May 2014 | 2006 | CAS | x | x |  x |
| **Use specific inventory lists** |   |   |  |   |   |   |
| The inventory of PFAS on the world market (not published) | Aug. 2016 | 2016 | CAS | x | x |   |
| Printing inks in EU (EUPIA 2013) | Aug. 2016 | 2013 | CAS | x | x |   |
| Textile chemicals in EU (not published) | Aug. 2016 | 2014 | CAS | x | x |   |
| Chemicals in cosmetics (the COSING database and the INCHI list) | Aug. 2016 | 2016 | CAS | x | x |   |
| Ingredients in Swedish Pharmaceuticals | Aug. 2016 | 2015 | CAS | x | x |  x |
| Pesticides used in Sweden (KemI) | Aug. 2016 | 2014 | CAS | x | x |  x |
| **Other lists** |   |   |  |   |   |   |
| The hazard classification (CLP) database | Aug. 2016 | 2016 | CAS ECno | x |   | x |

 |

## **Risk based prioritization**

Release and hazard potentials have been used for screening potential emerging substances that may need more thorough investigations. The screening work requires high throughput computer-based technology. Input data needs to be simple and standardized. For that purpose the Swedish Chemicals Agency (KEMI) has developed different semi-quantitative hazard and exposure scores. The exposure and hazard scores applied for this market list have been optimized for chronic effects on aquatic organisms.

## **The Exposure score (version 2)**

A chemical use pattern can be used to estimate their general release potential. In cases when use information is available, it has been used for a semi-quantitative prediction of the possibility to be released into the environment.

An exposure score has been developed based on the use pattern. This Exposure score is based on the Exposure score version 1 which consists of three parts:

1. the degree of uncontrolled release during use
2. the quantity used
3. the wideness of the use on the market.

Version 1 was developed for the prioritization o the NORMAN list of emerging substances (<http://www.normandata.eu/?q=node/50> ), and was based on publicly available data from REACH registration (consumed tonnage) and the SPIN database (use pattern data). In version 1b the number of information sources have been increased, covering also non-published regulatory information (see Table 1). In version 1c some errors in the algorithm were corrected (which have excluded some contributions to the exposure scores).

The estimations are in many cases rather uncertain, since information on consumed amount is often missing or confidential.

## **The Hazard score (version 1b)**

The hazard score was developed for screening purposes only. The Hazard score is based on the hazard classification according to the CLP (harmonized **C**lassification, **L**abelling and **P**ackaging) EU regulation. For the most hazardous substances, the decision on the classification of a chemical is taken at Community level. These are usually carcinogenic, mutagenic, toxic for reproduction or respiratory sensitisers. It is mandatory for the suppliers of the respective substance or mixture to apply this harmonised classification and labelling. For most chemicals/mixtures, suppliers need to decide on the classification. This is called self-classification. For more information see <https://echa.europa.eu/regulations/clp/understanding-clp>.

In the hazard score, different (eco)toxicological CLP endpoints are weighted to a single score. Also ecotoxicological endpoints (aquatic chronic toxicity) are considered among the most hazardous chemicals. The most hazardous chemicals have the highest score. Less severe hazards result in a lower score. For the self-classified substances, the maximum hazard score is reduced compared to the level generated from a harmonized classification. A complimentary source of hazard information, used for the hazard score, is indications that a substance is used as a pesticide of pharmaceutical (only in version 1). Chemicals with no or limited hazard information but with a similar structure (e.g. different salts) to a hazardous chemical will get a similar but reduced hazard score. In version 1b some errors in the algorithm were corrected (which excluded some contributions to the hazard scores). Furthermore, the calculation module taking the use in pharmaceutical and pesticides into account has temporarily been removed for reconstruction.